Analyse intégrée des données omiques dans l'impact de l'alimentation sur la santé cardiométabolique

Thèse

Bénédicte L Tremblay

Doctorat en nutrition
Philosophiæ doctor (Ph. D.)

Québec, Canada

Analyse intégrée des données omiques dans l'impact de l'alimentation sur la santé cardiométabolique

Thèse

Bénédicte L. Tremblay

Sous la direction de :
Marie-Claude Vohl, directrice de recherche

Résumé

Au Canada, les maladies cardiovasculaires (MCV) sont la deuxième cause de mortalité après le cancer, et l'une des principales causes d'hospitalisation. La prise en charge des individus souffrant de MCV repose sur l'évaluation et le traitement de plusieurs facteurs de risque cardiométabolique, lesquels comprennent le syndrome métabolique, l'activité physique et l'alimentation. L'adoption de saines habitudes de vie, incluant notamment une alimentation équilibrée, demeure la pierre angulaire de la prévention des MCV. En effet, une alimentation riche en fruits et légumes est inversement reliée à l'incidence de MCV. Les biomarqueurs d'exposition à la diète permettent par ailleurs d'étudier l'impact des facteurs alimentaires sur le développement des MCV. Les caroténoïdes plasmatiques, qui sont des biomarqueurs de la consommation de fruits et de légumes, sont associés à la santé cardiométabolique. L'alimentation influence en plus une multitude de facteurs omiques, modulant ainsi le risque de MCV.

Les sciences omiques étudient l'ensemble complexe des molécules qui composent le corps. Parmi ces sciences, la génomique, l'épigénomique, la transcriptomique et la métabolomique s'intéressent respectivement à l'étude à grande échelle des gènes, de la méthylation de l'ADN, de l'expression génique et des métabolites. Étant donné qu'un seul type de données omiques ne permet généralement pas de saisir la complexité des processus biologiques, une approche intégrative combinant plusieurs données omiques s'avère idéale afin de déchiffrer la physiopathologie des traits complexes. La biologie des systèmes étudie les interactions complexes des différentes données omiques entre elles, et avec l'environnement ainsi que leur influence sur un trait d'intérêt, tel que la santé. Il existe plusieurs méthodes pour analyser et intégrer des données omiques. La génétique quantitative permet d'estimer les contributions des effets génétiques et environnementaux dans la variance de traits complexes. L'analyse de réseaux de corrélations pondérées permet de mettre en relation un grand nombre de données omiques interreliées avec un trait, comme par exemple un ensemble de facteurs de risque de maladies complexes.

L'objectif général de cette thèse est d'étudier l'impact des déterminants omiques sur la relation entre l'alimentation et la santé cardiométabolique. Le premier objectif spécifique, utilisant une approche de la génétique quantitative, est de caractériser l'héritabilité des données omiques et des caroténoïdes plasmatiques ainsi que de vérifier si le lien avec des facteurs de risque cardiométabolique peut être expliqué par des facteurs génétiques et environnementaux. Le deuxième objectif spécifique, utilisant une approche de réseaux de corrélations pondérées, est d'évaluer le rôle des données omiques individuelles et combinées dans la relation entre les caroténoïdes plasmatiques et le profil lipidique.

Ce projet de doctorat repose sur l'étude observationnelle GENERATION qui comprend 48 sujets en bonne santé répartis en 16 familles. Toutes les données omiques étudiées et les caroténoïdes plasmatiques ont démontré

des ressemblances familiales dues, à des degrés divers, à l'effet de la génétique et de l'environnement partagé. La génétique et l'environnement sont également impliqués dans le lien entre la méthylation de l'ADN et l'expression génique ainsi qu'entre les métabolites, les caroténoïdes et les facteurs de risque cardiométabolique. L'utilisation de réseaux de corrélations pondérées a en outre permis de mieux comprendre le système moléculaire interactif qui relie les caroténoïdes, la méthylation de l'ADN, l'expression génique et le profil lipidique. En conclusion, ces travaux basés sur des données omiques individuelles et combinées analysées dans des approches de la génétique quantitative et de réseaux de corrélations pondérées ont mis en lumière la relation entre l'alimentation et la santé cardiométabolique.

Abstract

After cancer, cardiovascular disease (CVD) is the second leading cause of death and one of the leading causes of hospitalization in Canada. CVD management is based on the assessment and treatment of several cardiometabolic risk factors, which include metabolic syndrome, physical activity, and diet. A healthy lifestyle, including a balanced diet, remains the key to prevent CVD. A diet rich in fruits and vegetables is inversely associated with CVD incidence. Biomarkers of exposure to diet are used to study the impact of dietary factors on the development of CVD. Plasma carotenoids, a biomarker of fruit and vegetable consumption, are associated with cardiometabolic health. Diet also influences a myriad of omics factors, thus modulating CVD risk.

Omics sciences study the complex set of molecules that make up the body. Among these sciences, genomics, epigenomics, transcriptomics, and metabolomics consider the large-scale study of genes, DNA methylation, gene expression, and metabolites, respectively. Given that a single type of omics data usually does not capture the complexity of biological processes, an integrative approach combining multiple omics data proves ideal to elucidate the pathophysiology of diseases. Systems biology studies the complex interactions of different omics data among themselves and with the environment on a trait such as health. There are several methods for analyzing and integrating omics data. Quantitative genetics estimates the contributions of genetic and environmental effects to the variance of complex traits such as omics data. Weighted correlation network analysis allows the association of a large number of omics data with a trait such as risk factors for diseases.

The general objective of this thesis is to study the impact of omics determinants in the link between diet and cardiometabolic health. The first specific objective, using a quantitative genetics approach, is to characterize the heritability of omics data and plasma carotenoids as well as to check if their link with cardiometabolic risk factors can be explained by genetic and environmental factors. The second specific objective, using a weighted correlation network approach, is to assess the role of individual and combined omics data in the relationship between plasma carotenoids and lipid profile.

This project is based on the *GENERATION* observational study, which includes 48 healthy subjects from 16 families. All omics data studied showed familial resemblances due, to varying degrees, to genetic and common environmental effects. Genetics and environment are also involved in the link between DNA methylation and gene expression, as well as between metabolites, carotenoids, and cardiometabolic risk factors. Moreover, weighted correlation network analysis has provided insight into the interactive molecular system that links carotenoids, DNA methylation, gene expression, and lipid profile. In conclusion, the present study, using approaches from quantitative genetics and weighted correlation network analysis, brought to light the impact of some individual and combined omics data in the link between diet and cardiometabolic health.

Table des matières

Liste des figures

Chapitre 1

Chapitre 3

Chapitre 4

Chapitre 5

Chapitre 7

Chapitre 8

Chapitre 9

Discussion et conclusion

Liste des tableaux

Chapitre 1

Chapitre 3

Chapitre 4

Chapitre 5

Chapitre 6

Chapitre 7

Chapitre 8

Chapitre 9

Liste des abréviations

ACC : *American College of Cardiology*
ADN : acide désoxyribonucléique
AGn-3 : acide gras oméga-3
ARN : acide ribonucléique
ARNm : ARN messager
AHA : *American Heart Association*
Apo B100 : apolipoprotéine B100
Apo E : apolipoprotéine E
CD36 : *cluster determinant 36*
CM : cardiométabolique
C-HDL : cholestérol HDL
C-LDL : cholestérol LDL
CpG : cytosine-phosphate-guanine
CRP : *C-reactive protein*
CT : cholestérol total
CVD : *cardiovascular disease*
DB2 : diabète de type 2
DBP : *diastolic blood pressure*
eQTL : *expression quantitative trait loci*
EWAS : étude d'association à l'échelle de l'épigénome (*epigenome-wide association study*)
FAS : *Fatty Acid Sensor*
FAV *: fruits and vegetables*
FDR *: false discovery rate*
GTEx *: Genotype-Tissue Expression*
GO : *gene ontology*
GS *: gene significance*
GWAS : étude d'association à l'échelle du génome (*genome-wide association study*)
IDL : lipoprotéine de densité intermédiaire
IMC : indice de masse corporelle
HDL : lipoprotéine de haute densité
LDL : lipoprotéine de faible densité
LINE-1 : *long-interspersed nucleotide repetitive elements-1*
LC-MS : chromatographie liquide couplée en ligne à la spectrométrie de masse
MCV : maladie cardiovasculaire
mQTL : *metabolite quantitative trait loci*
MHC : *major histocompatibility complex*
ME : *module eigengene*
meQTL : *methylation quantitative trait loci*
MM : *module membership*
NCEP ATP : *National Cholesterol Education Program Adult Treatment Panel*
NPC1L1 *: niemann-pick C1-like1*
PPAR *: peroxisome proliferator-activated receptor*

SNP : *single nucleotide polymorphism*
SR-B1 *: scavenger receptor class B type 1*
TOM *: topological overlap matrix*
TG : triglycéride
V_C : variance attribuable à l'environnement partagé
V_E : variance attribuable à l'environnement unique ou résiduel à chaque individu
V_G : variance attribuable à la génétique
VLDL : lipoprotéine de très faible densité
WGCNA : *weighted gene correlation network analysis*

À Benjamin

Remerciements

La réalisation de cette thèse n'aurait pas été possible sans la contribution, la collaboration et le soutien de nombreuses personnes. Je tiens donc à leur adresser mes plus sincères remerciements.

Mes premières pensées sont pour ma directrice de thèse, la Dre Marie-Claude Vohl, qui m'a offert un encadrement exemplaire tout au long de mes études graduées. Mon intérêt pour la recherche est né lors du stage de recherche que j'ai réalisé au sein de son équipe à l'Institut sur la nutrition et les aliments fonctionnels (INAF) en 2013. Au moment de déposer cette thèse, cela fera sept ans que nous travaillons ensemble. Merci Marie-Claude d'avoir cru en moi et de m'avoir fait confiance dès le début. Ce fut un privilège de faire partie de ton équipe et je te suis reconnaissante de m'avoir permis de réaliser mon plein potentiel. Grâce à ta grande disponibilité, à ton écoute et à ton soutien, j'ai beaucoup évolué sur les plans professionnel et personnel. J'espère sincèrement pouvoir continuer à travailler avec toi dans l'avenir.

Je tiens aussi à remercier Frédéric Guénard pour tout le soutien et l'aide qu'il m'a apportés pendant mon doctorat. Merci Fred pour ton implication dans mes projets, ton expertise en analyses statistiques, pangénomiques et bio-informatiques, et ta grande patience. Je n'aurais pas pu réaliser la moitié de ces travaux sans ton aide.

Je remercie également tous les membres de l'équipe de la Dre Vohl que j'ai côtoyés au cours des dernières années et plus particulièrement Bastien Vallée-Marcotte, Bénédicte Allam-Ndoul, Juan De Toro Martin, Maximilien Franck, Véronique Garneau, Michèle Kearney, Catherine Raymond et Alain Houde. Je garde de précieux souvenirs de nos discussions scientifiques, mais aussi de nos « bières et chips » et de nos activités en congrès. Ce fut un plaisir de partager votre quotidien.

Mes remerciements vont aussi au Dr Benoît Lamarche, coauteur des articles de cette thèse, pour ses précieux conseils. Merci Benoît d'avoir pris le temps de m'écouter et de me conseiller, notamment à propos de mon parcours académique. Nos échanges ont fait une grande différence dans ma perception du domaine de la recherche et dans mes choix de carrière.

Merci aux Drs Julie Robitaille, Francine Durocher et Ahmed El-Sohemy qui ont accepté d'évaluer ma thèse et d'agir à titre de jury lors de ma soutenance de thèse, deux étapes importantes de mon parcours.

Je remercie également les Instituts de recherche en santé du Canada et le Fonds de recherche du Québec – Santé pour m'avoir octroyé une bourse de recherche pour la durée de mon doctorat. Cela m'a permis de m'investir pleinement dans mes études en plus de participer à de nombreux congrès nationaux et internationaux.

Merci à mes collègues, qui sont devenus des amis : Élise Carbonneau, Audrée-Anne Dumas, Raphaëlle Jacob, Janie Allaire et Jean-Philippe Drouin-Chartier. Mon parcours à l'INAF a été tellement plus enrichissant et mémorable grâce à vous, avec qui j'ai pu partager tant d'événements marquants dans nos vies professionnelles et personnelles. Merci à tous les collègues étudiants, professionnels de recherche et professeurs que j'ai eu l'occasion de côtoyer et qui font de l'INAF un milieu de travail extrêmement accueillant et agréable.

En terminant, je tiens à remercier ma famille. Mes beaux-parents, Lyne et Marcel, pour leur soutien et le gardiennage de mon garçon pendant la rédaction de ma thèse. Mes frères, Frédéric et David, et ma Mémé qui sont fiers de moi et qui m'ont toujours encouragée. Mes parents, Chantal et Serge, pour leur indéfectible soutien et pour m'inciter à toujours aller plus loin dans la vie. Merci d'être mon meilleur public, de m'écouter et de me conseiller dans toutes les sphères de ma vie. Un merci tout particulier à mon père pour la correction de mes articles, de mes demandes de bourse et de cette thèse. Papa, ta contribution a fait une énorme différence dans la qualité de mes travaux. Un gros merci à mon mari, Mathieu, pour son soutien inconditionnel, sa grande patience et ses encouragements. Mathieu, ton enthousiasme pour mon parcours académique est une source importante de motivation pour moi. Enfin, merci à mon garçon, Benjamin, pour tout le bonheur qu'il m'apporte et pour l'équilibre qu'il me permet de préserver dans ma vie.

Avant-propos

Cette thèse comprend sept articles scientifiques publiés que j'ai rédigés à titre de premier auteur. Les quatre premiers (chapitres 3 à 6) portent sur l'héritabilité de la méthylation de l'ADN, de l'expression génique, des métabolites plasmatiques et des caroténoïdes plasmatiques. Les trois suivants (chapitres 7 à 9) portent sur l'analyse de réseaux de corrélations pondérées dans le lien entre les caroténoïdes plasmatiques et le profil lipidique.

L'ensemble des articles présentés dans cette thèse reposent sur l'étude observationnelle GENERATION. Ce projet de recherche est financé par la Chaire de recherche du Canada sur la génomique appliquée à la nutrition et la santé métabolique dont la Dre Marie-Claude Vohl est titulaire. L'objectif général de cette étude était d'évaluer, à l'aide de biomarqueurs d'exposition à la diète, l'impact de l'alimentation sur les données omiques de même que sur les facteurs de risque cardiométabolique à différents stades de la vie. L'étude a été réalisée à l'Institut sur la nutrition et les aliments fonctionnels (INAF) de l'Université Laval à Québec. Le recrutement des sujets participants s'est déroulé de mai 2011 à novembre 2014. Le devis de l'étude est détaillé dans l'article présenté au chapitre 3.

Je me dois tout d'abord de remercier les parents et leurs enfants d'avoir participé à l'étude. Grâce à eux, nous avons obtenu une énorme quantité d'informations pour étudier les données omiques dans le contexte de l'alimentation et de la santé. Des remerciements sont également dus au personnel infirmier de l'Unité d'investigation clinique de l'INAF, **Steve Larouche** et **Christiane Landry**, pour les soins prodigués aux participants. De même, je veux adresser un merci tout spécial aux professionnelles de recherche du projet : **Véronique Garneau** pour le recrutement des participants et la gestion des bases de données cliniques et nutritionnelles, et **Catherine Raymond**, pour l'extraction d'ADN et d'ARN ainsi que pour la préparation des échantillons au laboratoire. Je remercie également **Véronique Richard**, responsable du dosage des caroténoïdes plasmatiques. Enfin, je remercie encore une fois la **Dre Marie-Claude Vohl**, à titre de chercheure principale du projet et de l'ensemble des articles présentés dans cette thèse.

Les Drs Frédéric Guénard, Benoît Lamarche, Louis Pérusse et Marie-Claude Vohl sont coauteurs des sept articles de cette thèse. Par souci de concision, j'ai regroupé ma contribution et celle de mes coauteurs aux différents articles selon les deux approches d'analyse utilisées.

Pour les quatre articles utilisant l'approche de la génétique quantitative, j'étais responsable de la conceptualisation du projet, de la planification des analyses (pangénomiques et bio-informatiques) et de l'interprétation des résultats. J'étais en plus responsable de coordonner le travail de collaboration avec l'équipe du Dr Louis Pérusse pour les analyses d'héritabilité. J'ai aussi assumé la rédaction, la soumission et la révision

des articles pour leur publication dans des journaux avec comité de révision par les pairs. J'ai enfin eu la chance d'en présenter les résultats lors de 11 congrès régionaux, nationaux et internationaux.

Dre Marie-Claude Vohl et **Dr Benoît Lamarche**, tous deux professeurs à l'École de Nutrition de l'Université Laval, ont conçu le projet de recherche. La Dre Marie-Claude Vohl m'a conseillé et aidé à toutes les étapes du projet. Le **Dr Louis Pérusse**, professeur au département de kinésiologie de l'Université Laval, qui possède une expertise importante en génétique quantitative des traits complexes, a été très impliqué dans la planification et la réalisation des analyses d'héritabilité et de génétique bivariée. Le **Dr Frédéric Guénard**, bio-informaticien de l'équipe à l'époque et maintenant Conseiller en développement à la recherche de la Faculté de médecine de l'Université Laval, a contribué de façon importante à toutes les étapes du projet, mais plus spécifiquement à la planification et à la réalisation des analyses ainsi qu'à l'interprétation des résultats. C'est grâce à son aide précieuse que je me suis familiarisée avec un nouveau domaine de recherche et de nouvelles méthodes d'analyse, tout en progressant rapidement dans le projet. Les Drs Frédéric Guénard, Louis Pérusse, Benoît Lamarche et Marie-Claude Vohl ont tous révisé les manuscrits et en ont approuvé les versions finales. Pour terminer, je remercie **Christian Couture**, professionnel de recherche dans l'équipe du Dr Pérusse, pour les analyses d'héritabilité. Il a réalisé les différents modèles d'héritabilité en plus de prendre le temps de répondre à toutes nos questions.

Chapitre 3 : Tremblay BL, Guénard F, Lamarche B, Pérusse L, Vohl MC. Familial resemblances in blood leukocyte DNA methylation levels. *Epigenetics* (2016) 11(11):831-838.

Chapitre 4 : Tremblay BL, Guénard F, Lamarche B, Pérusse L, Vohl MC. Familial resemblances in human whole blood transcriptome. *BMC Genomics* (2018);19(1):300.

Chapitre 5 : Tremblay BL, Guénard F, Lamarche B, Pérusse L, Vohl MC. Familial resemblances in human plasma metabolites are attributable to both genetic and common environmental effects. *Nutrition Research* (2018) 10.1016.

Chapitre 6 : Tremblay BL, Guénard F, Lamarche B, Pérusse L, Vohl MC. Genetic and common environmental contributions to familial resemblances in plasma carotenoid concentrations in healthy families. *Nutrients* (2018);10(8).

Pour les trois articles utilisant l'approche de réseaux de corrélations pondérées, j'étais responsable de la conceptualisation du projet, de la planification et de la réalisation des analyses, ainsi que de l'interprétation des résultats. J'ai aussi rédigé, soumis et révisé les articles pour publication dans des journaux avec comité de révision par les pairs. **Dre Marie-Claude Vohl** et **Dr Benoît Lamarche** ont conçu le projet de recherche. La Dre

Marie-Claude Vohl m'a conseillé et aidé à toutes les étapes du projet. Le **Dr Frédéric Guénard** a réalisé les analyses de réseaux de corrélations pondérées. Le **Dr Louis Pérusse** a commenté et révisé la méthode d'analyse et les résultats obtenus. Les Drs Frédéric Guénard, Louis Pérusse, Benoît Lamarche et Marie-Claude Vohl ont révisé les manuscrits et en ont approuvé les versions finales.

Chapitre 7 : Tremblay BL, Guénard F, Lamarche B, Pérusse L, Vohl MC. Weighted gene co-expression network analysis to explain the relationship between plasma total carotenoids and lipid profile. *Genes and Nutrition* (2019) 14:16.

Chapitre 8 : Tremblay BL, Guénard F, Lamarche B, Pérusse L, Vohl MC. Network analysis of the potential role of DNA methylation in the relationship between plasma carotenoids and lipid profile. *Nutrients* (2019) 4;11(6).

Chapitre 9 : Tremblay BL, Guénard F, Lamarche B, Pérusse L, Vohl MC. Integrative network analysis of multi-omics data in the link between plasma carotenoid concentrations and lipid profile. *Lifestyle Genomics* (2020) 13 (1):11-19.

Notes concernant les publications non incluses dans cette thèse

Au cours de mes études doctorales, j'ai eu l'occasion de publier deux autres articles scientifiques à titre de premier auteur :

Tremblay BL, Guénard F, Rudkowska I, Lemieux S, Couture P, Vohl MC. Epigenetic changes in blood leukocytes following an omega-3 fatty acid supplementation. *Clinical Epigenetics* (2017) 9:43.

Cet article porte sur les changements de la méthylation de l'ADN suite à une supplémentation en acides gras oméga-3 dans l'étude clinique *FAS*. J'étais responsable de la planification et de la réalisation des analyses pangénomiques et bio-informatiques ainsi que de l'interprétation des résultats. J'étais aussi responsable de la rédaction, de la soumission et de la révision de l'article pour sa publication dans le journal *Clinical Epigenetics*. Le **Dr Frédéric Guénard** a contribué de façon importante à la planification et à la réalisation des analyses. Les Dres **Iwona Rudkowska** (Centre de recherche CHU de Québec), **Simone Lemieux** (École de Nutrition de l'Université Laval) et **Marie-Claude Vohl** ont conçu le projet de recherche. Le **Dr Patrick Couture,** chercheur-clinicien au Centre de recherche CHU de Québec, était responsable du suivi médical des participants. Tous les coauteurs ont révisé et approuvé la version finale du manuscrit.

Tremblay BL, Rudkowska I. Nutrigenomic point of view on effects and mechanisms of action of ruminant trans fatty acids on insulin resistance and type 2 diabetes. *Nutrition Reviews* (2017) 75(3):214-223.

Cette revue de littérature, qui traite des effets des acides gras *trans* laitiers sur la résistance à l'insuline et le diabète de type 2, a été rédigée en collaboration avec la **Dre Iwona Rudkowska**. Je la remercie de m'avoir donné l'occasion d'écrire cette revue de littérature et de la publier dans l'un des meilleurs journaux en nutrition. J'étais responsable de la planification et de la réalisation de la revue de littérature ainsi que de l'interprétation des résultats. J'étais aussi chargée de la rédaction, de la soumission et de la révision de l'article pour publication dans le journal *Nutrition Reviews*. La Dre Rudkowska a contribué à la planification de la revue de littérature, à l'interprétation des résultats et à la révision de l'article.

Introduction

Au Canada, les maladies cardiovasculaires (MCV) sont la deuxième plus importante cause de mortalité après le cancer et l'une des principales causes d'hospitalisation (1). Elles engendrent dans le système de santé canadien des coûts directs de 11,7 milliards de dollars annuellement (2). La prise en charge des individus souffrant de MCV repose sur l'évaluation et le traitement de plusieurs facteurs de risque cardiométabolique (CM), lesquels comprennent le syndrome métabolique, le tabagisme, l'activité physique et l'alimentation (3). L'adoption de saines habitudes de vie, incluant notamment une alimentation équilibrée, demeure la pierre angulaire de la prévention des maladies chroniques, dont les MCV (4, 5). En effet, une alimentation riche en fruits et légumes est inversement reliée à l'incidence de MCV (6). Les biomarqueurs d'exposition à la diète permettent par ailleurs d'étudier l'impact des facteurs alimentaires sur le développement des MCV (7). Les caroténoïdes plasmatiques, qui sont des biomarqueurs de la consommation de fruits et de légumes, sont associés à la santé CM (8, 9). De plus, l'alimentation a une incidence sur la quasi-totalité des processus cellulaires, allant de l'expression des gènes, à la synthèse et à la dégradation des protéines (10). L'alimentation affecte une multitude de données omiques, modulant ainsi le risque de maladies chroniques complexes, dont les MCV (11-13).

Les sciences omiques étudient l'ensemble complexe des molécules qui composent le corps. Parmi ces sciences, la génomique, l'épigénomique, la transcriptomique et la métabolomique s'intéressent respectivement à l'étude à grande échelle des gènes, de la méthylation de l'ADN, de l'expression génique et des métabolites. L'essor des sciences omiques a mené à des percées scientifiques majeures dans la compréhension de la physiopathologie et du traitement des maladies complexes. Étant donné qu'un seul type de données omiques ne permet généralement pas de saisir la complexité des processus biologiques, une approche intégrative combinant plusieurs données omiques s'avère idéale pour mieux comprendre les maladies (14). Jusqu'à maintenant, la plupart des études multi-omiques se sont concentrées sur l'effet de variations génétiques sur d'autres déterminants omiques. Toutefois, ces études incluent rarement des facteurs environnementaux, tels que l'alimentation.

La biologie des systèmes vise une compréhension plus holistique de la biologie en intégrant plusieurs types de données omiques en plus de l'environnement (15, 16). Les processus biologiques, tels que le développement de maladies, impliquent un système dynamique et interactif de données omiques qui sont influencées par l'environnement. La biologie des systèmes étudie les interactions complexes des différentes données omiques entre elles et avec l'environnement ainsi que leur influence sur un trait d'intérêt, tel que la santé (17). Cette science a le potentiel d'améliorer la prédiction, le diagnostic, le traitement et le pronostic des maladies (14). Toutefois, l'utilisation et l'intégration de données omiques dans la recherche en santé présentent plusieurs défis.

Tout d'abord, les maladies chroniques complexes sont multifactorielles, c'est-à-dire qu'elles sont influencées autant par des facteurs génétiques que par des facteurs environnementaux (18). La génétique quantitative permet d'estimer les contributions des effets génétiques et environnementaux dans la variance de traits complexes (19). De plus, l'analyse des données omiques génère des masses importantes de données qui sont fortement interreliées (20). L'analyse de réseaux de corrélations pondérées permet de tenir compte de larges échantillons de données omiques et de leur haut degré d'interconnexion en plus de les mettre en relation avec un trait d'intérêt (21).

À ce jour, aucune étude n'a utilisé une approche de la biologie des systèmes pour analyser de façon intégrée les données omiques dans l'impact de l'alimentation sur la santé CM au sein de familles en santé. Une meilleure compréhension des voies métaboliques régulées par l'alimentation et de leurs impacts sur la santé permettrait d'aider à prédire et à prévenir plus efficacement le risque de maladies chroniques complexes et d'émettre des recommandations nutritionnelles qui tiennent compte des profils omiques, des habitudes de vie et de l'environnement des individus. L'objectif général de cette thèse est d'étudier l'impact des déterminants omiques sur la relation entre l'alimentation et la santé CM. Le premier objectif spécifique, utilisant une approche de la génétique quantitative, est de caractériser l'héritabilité des données omiques et des caroténoïdes plasmatiques ainsi que de vérifier si le lien avec des facteurs de risque CM peut être expliqué par des facteurs génétiques et environnementaux. Le deuxième objectif spécifique, utilisant une approche de réseaux de corrélations pondérées, est d'évaluer le rôle des données omiques individuelles et combinées dans la relation entre les caroténoïdes plasmatiques et le profil lipidique, qui est un facteur de risque CM.

La suite de cette thèse est divisée en 10 chapitres. Le premier chapitre introduit la problématique de recherche. Il présente tout d'abord un portrait global de la santé CM, de l'alimentation et de leur relation. Sont ensuite abordées les données omiques (génomique, épigénomique, transcriptomique et métabolomique) ainsi que leurs relations avec la santé CM et l'alimentation. Finalement, les principes de la biologie des systèmes et de différentes méthodes d'analyse de données omiques sont expliqués. Le deuxième chapitre présente les objectifs et les hypothèses de recherche et introduit les sept articles scientifiques répondant aux objectifs de recherche. Ces articles constituent les chapitres 3 à 9. Le dernier chapitre comprend la discussion et la conclusion générale de la thèse.

Chapitre 1 Problématique générale

1.1 Santé cardiométabolique

Les maladies cardiovasculaires (MCV) sont un trouble du cœur et des vaisseaux sanguins qui inclut, entre autres, les cardiopathies ischémiques, les maladies cérébrovasculaires, la maladie vasculaire périphérique et l'insuffisance cardiaque (22). Les événements cardiovasculaires réfèrent à un problème de santé d'ordre cardiovasculaire qui a mené à une consultation médicale ou à une hospitalisation. Les événements cardiovasculaires de type hémorragique se produisent lors de la rupture d'un anévrisme dans la paroi d'un vaisseau sanguin. Les événements cardiovasculaires ischémiques se produisent lors de l'obstruction partielle ou totale d'un vaisseau sanguin par la plaque d'athérosclérose ou un thrombus (23). La plaque d'athérosclérose, qui est directement associée aux MCV, est caractérisée par une inflammation chronique et par l'accumulation de cholestérol dans la paroi interne des artères coronaires (24). Lorsque l'apport de sang au cœur est bloqué, il y a infarctus aigu du myocarde (23).

Les MCV sont la principale cause de décès dans le monde : environ 8,9 millions d'individus sont décédés d'une cardiopathie en 2015, ce qui représente 45 % du total des décès attribuables à des maladies non transmissibles (25). Au Canada, les MCV demeurent la deuxième cause de décès après le cancer et l'une des principales causes d'hospitalisation (1). Les MCV engendrent des coûts directs annuels de 11,7 milliards de dollars dans le système de santé canadien (2). La prise en charge des individus souffrant de MCV repose sur l'évaluation et le traitement de plusieurs facteurs de risque cardiométabolique (CM). Ces facteurs comprennent le syndrome métabolique, le tabagisme, l'activité physique et l'alimentation (3).

Le syndrome métabolique regroupe une constellation de facteurs de risque lipidiques et non lipidiques d'origine métabolique (26). Il est caractérisé par une résistance à l'insuline associée à l'obésité abdominale, elle-même associée à une panoplie d'anomalies athérogènes, dont la dyslipidémie, la dysglycémie, la tension artérielle élevée, un profil pro-thrombotique et un état inflammatoire (27, 28). Plusieurs facteurs prédisposent au syndrome métabolique dont, entre autres, la génétique, l'alimentation, la sédentarité, le tabagisme et le stress (29). Plusieurs organisations de santé à travers le monde ont proposé des valeurs seuils pour identifier les individus susceptibles d'être résistants à l'insuline et de présenter aussi des anomalies athérogènes et diabétogènes. L'*American Heart Association* (AHA), le *Lung and Blood Institute*, la *World Health Federation* et l'*International Atherosclerosis Society* ont harmonisé les critères diagnostiques du syndrome métabolique (30). Il y a syndrome métabolique lorsqu'au moins trois des facteurs de risque énumérés dans le tableau qui suit sont présents.

Tableau 1.1 Critères du syndrome métabolique

Facteurs de risque	Niveaux seuils
Tour de taille	
- Hommes	> 102 cm
- Femmes	> 88 cm
Triglycérides	≥ 1,7 mmol/L
Cholestérol HDL	
- Hommes	< 1,0 mmol/L
- Femmes	< 1,3 mmol/L
Tension artérielle	≥ 130/85 mmHg
Glycémie	≥ 5,6 mmol/L ou diabète type 2

Tableau tiré de l'*International Diabetes Federation* 2006 (31).

Le syndrome métabolique augmente de 1,5 à 2 fois le risque relatif de MCV et de 5 fois le risque de diabète de type 2 (DB2) (30, 32). Le risque CM représente le risque global de MCV résultant de la présence du syndrome métabolique, mais également de facteurs de risque traditionnels tels l'âge, le sexe, la dyslipidémie, la dysglycémie, la tension artérielle élevée, le tabagisme et des facteurs génétiques (29).

1.1.1 Profil lipidique

Les lipoprotéines sont des complexes macromoléculaires qui permettent le transport des lipides (cholestérol et triglycérides [TG]) dans le corps. Les apolipoprotéines sont des protéines constitutives des lipoprotéines qui sont essentielles à leur structure et à leur fonction. Il existe cinq types de lipoprotéines : les chylomicrons, les lipoprotéines de très faible densité (VLDL), les lipoprotéines de densité intermédiaire (IDL), les lipoprotéines de faible densité (LDL) et les lipoprotéines de haute densité (HDL). Toutes ces lipoprotéines contiennent à la fois du cholestérol et des TG en quantités variables (33).

La dyslipidémie est un facteur de risque CM. Elle se caractérise par des valeurs anormales de lipides sanguins et sa complication la plus fréquente est la MCV (34). Il existe plusieurs facteurs de risque CM de type lipidique. En effet, on peut associer au risque d'événements cardiovasculaires une augmentation des concentrations de cholestérol total (CT) (35), de cholestérol LDL (C-LDL) (36), d'apolipoprotéines B100 (apo B100) (37) et de TG (38). Par contre, une augmentation des concentrations de cholestérol HDL (C-HDL) est associée à un risque plus faible de MCV (36). La mesure du CT, qui comprend le cholestérol des cinq lipoprotéines, est considérée dans le calcul du risque absolu de maladie coronarienne de Framingham (39). Les particules LDL transportent la majorité du cholestérol en circulation (40). Le C-LDL est donc un facteur de risque important des MCV. En effet, une augmentation des concentrations de C-LDL de 1 mmol/L a été associée à une augmentation de 41 % et de 24 % du risque relatif d'événements coronariens chez les hommes et les femmes, respectivement (36). Les particules HDL effectuent le transport inverse du cholestérol, c'est-à-dire qu'ils recueillent le cholestérol

excédentaire dans les tissus périphériques et les macrophages et le transportent vers le foie (40). Le C-HDL est donc inversement associé aux MCV. Une augmentation dans les concentrations de C-HDL de 0,4 mmol/L été associée à une diminution de 34 % et de 24 % du risque d'événements coronariens chez les hommes et les femmes, respectivement (36). Des concentrations élevées de TG sont aussi associées au risque d'événements cardiovasculaires (41). Enfin, les concentrations d'apo B100 reflètent le nombre de particules athérogènes en circulation. Sa mesure est recommandée, en plus de la mesure des concentrations de C-LDL, pour l'évaluation du risque cardiovasculaire (34).

Plusieurs organisations à travers le monde proposent des lignes directrices pour le diagnostic, le dépistage et le traitement des dyslipidémies (39, 42-44). Le *National Cholesterol Education Program (NCEP) Adult Treatment Panel (ATP)* publie les lignes directrices pour le traitement des dyslipidémies en considérant les modifications alimentaires comme le fondement de la prévention primaire (45). L'AHA et *l'American College of Cardiology* (ACC) ont également émis des recommandations alimentaires afin de réduire le risque d'événements cardiovasculaires par une diminution du C-LDL (4, 44).

1.2 Alimentation

Selon des organismes de santé au Canada et aux États-Unis, l'adoption de saines habitudes de vie, incluant notamment une alimentation équilibrée, demeure la pierre angulaire de la prévention des maladies chroniques, dont les MCV (4, 5, 44). Le NCEP ATP III a émis des recommandations nutritionnelles afin de réduire le risque de MCV par une diminution des concentrations de C-LDL. Il est recommandé de diminuer la consommation de gras saturés et de cholestérol, et d'augmenter la consommation de fibres solubles et de stérols végétaux (42). D'après l'étude INTERHEART, qui compte près de 30 000 participants de 52 pays, neuf facteurs de risque modifiables (tabagisme, tour de taille, DB2, tension artérielle, apport alimentaire en fruits et légumes, consommation d'alcool, rapport des concentrations sanguines d'apo B et d'apo A1, stress et niveau d'activité physique) permettent de prédire environ 90 % du risque de développer des MCV (46). D'après cette même étude, le fait d'adopter une saine alimentation, d'augmenter le niveau d'activité physique et de ne pas fumer diminue de 80 % le risque d'infarctus du myocarde (46). Une alimentation riche en fruits et légumes est d'ailleurs inversement associée à l'incidence de MCV (6). En ce sens, l'AHA et ACC émettent également plusieurs recommandations nutritionnelles pour réduire le risque d'événements cardiovasculaires par une diminution des concentrations de C-LDL. Elle recommande de consommer une variété de fruits, de légumes, de grains entiers, de produits laitiers faibles en gras, de volailles, de poissons, de légumineuses, d'huiles polyinsaturées et de noix. Elle recommande aussi de limiter la consommation de viandes rouges, de boissons sucrées et de sucreries en plus de diminuer la quantité d'énergie provenant des acides gras saturés et *trans* (4, 44). En harmonie avec ces recommandations, la diète Portfolio est un régime alimentaire principalement à base de végétaux

comprenant quatre aliments ou nutriments hypocholestérolémiants (noix, protéines végétales, fibres solubles et stérols végétaux) qui a été élaboré pour réduire le risque de MCV (47). Dans une étude d'intervention contrôlée où les aliments sont fournis, la réduction du C-LDL avec la diète Portfolio était similaire à celle observée avec 20 mg de lovastatine, un médicament hypocholestérolémiant (-28,6 % pour la diète Portfolio vs -30,9 % pour le lovastatine) (48). Les preuves connues démontrent que la diète Portfolio conduit à des améliorations cliniquement significatives des concentrations de C-LDL, des autres facteurs de risque CM et du risque de maladie coronarienne sur 10 ans (47).

1.2.1 Caroténoïdes

Il existe plusieurs méthodes pour évaluer les apports alimentaires, dont les journaux alimentaires, les questionnaires de fréquence alimentaire et les rappels de 24 heures. Toutefois, ces méthodes sont basées sur l'auto déclaration et sont donc sujettes à plusieurs biais (49). Des études ont démontré un haut niveau de fausses déclarations lors de l'utilisation de ces méthodes autant chez les adultes que chez les enfants et les adolescents (50-52). Malgré que de nombreuses stratégies aient été élaborées pour mieux gérer le caractère aléatoire et atténuer les erreurs systématiques de ces mesures, les biais associés aux données alimentaires auto rapportées ne peuvent pas être éliminés complètement (53). La mesure des biomarqueurs d'exposition à la diète dans le sang, l'urine et les tissus s'avère plus objective pour caractériser les apports et les profils alimentaires (54, 55). Plus de 100 biomarqueurs d'exposition à la diète ont été identifiés dans des études de cohorte (7, 56). Ces biomarqueurs présentent un réel potentiel pour aider à mieux caractériser le risque CM (54). Ils pourraient également améliorer notre compréhension du lien entre l'alimentation et la génétique dans le développement des maladies chroniques (57). Enfin, l'étude de ces biomarqueurs peut mener à la découverte de nouvelles molécules bioactives ou de facteurs alimentaires associés au développement des maladies complexes (7).

Parmi la panoplie de biomarqueurs, on retrouve les caroténoïdes, qui ont souvent été étudiés pour leurs propriétés antioxydantes dans le contexte du cancer et de la santé cardiovasculaire (9). Ils constituent des biomarqueurs fiables de la consommation de fruits et légumes (7, 8, 58). En effet, la consommation de fruits et légumes fournit plus de 90 % de l'apport quotidien en caroténoïdes (59). Les caroténoïdes forment une famille de plus de 700 pigments liposolubles, parmi lesquels six caroténoïdes principaux (α-carotène, β-carotène, β-cryptoxanthine, lycopène, lutéine et zéaxanthine) représentent plus de 95 % des caroténoïdes circulant dans le plasma ou le sérum humain (59, 60). Les caroténoïdes sont divisés en deux grandes familles en fonction de leur composition chimique : les xanthophylles et les carotènes. Les xanthophylles (lutéine, zéaxanthine et β-cryptoxanthine) contiennent un ou plusieurs atomes d'oxygène. Les carotènes (α-carotène, β-carotène et lycopène) sont des hydrocarbones qui ne contiennent pas d'oxygène (61). Les caroténoïdes doivent être libérés

de la matrice alimentaire afin d'être acheminés aux cellules intestinales pour leur absorption. La transformation mécanique est essentielle à cette libération. L'exposition à la chaleur peut également améliorer la libération en ramollissant les structures de la paroi cellulaire (62). Les caroténoïdes libérés de la matrice végétale rejoignent les composés lipidiques du repas pour être intégrés aux micelles. La présence de lipides alimentaires dans la lumière intestinale est donc un élément essentiel à l'absorption des caroténoïdes (62). Les micelles transportent ensuite les caroténoïdes jusqu'à la bordure en brosse de l'intestin. Une partie de la lutéine, de la zéaxanthine et de la β-cryptoxanthine voyage spontanément de la micelle vers les cellules intestinales alors que l'action d'une lipase est essentielle pour détacher les carotènes (63). Plusieurs transporteurs, dont le *scavenger receptor class B type 1* (SR-B1), le *cluster determinant 36* (CD36) et le *niemann-pick C1-like1* (NPC1L1), sont impliqués dans l'absorption des caroténoïdes. La grande variabilité interindividuelle observée dans l'absorption des caroténoïdes peut s'expliquer en partie par des variations génétiques causant des différences d'expression des gènes codants pour ces transporteurs (64). Suite à leur absorption, les caroténoïdes suivent la voie des autres composés lipidiques en rejoignant la circulation sanguine dans les chylomicrons où ils pourront être transportés jusqu'aux tissus cibles (63). Environ 90 % des caroténoïdes sont entreposés dans les tissus et 10 % sont en circulation dans le plasma. Les caroténoïdes sont entreposés à 80-85 % dans le tissu adipeux, 8-12 % dans le foie et 2-3 % dans les muscles (65). Malgré les différents facteurs qui affectent la biodisponibilité des caroténoïdes, les caroténoïdes circulants sont corrélés aux apports en fruits et légumes. Les apports en β-carotène ($r = 0,45$, $P = 0,002$), lycopène ($r = 0,47$, $P < 0,001$) et lutéine ($r = 0,639$, $P < 0,001$) corrèlent avec les taux plasmatiques mesurés (66). Dans une étude regroupant six études d'interventions alimentaires contrôlées, les concentrations plasmatiques de β-cryptoxanthine, lutéine et zéaxanthine étaient positivement associées aux apports alimentaires de ces caroténoïdes et aussi à l'apport en fruits et légumes (58).

Une grande variabilité interindividuelle dans les niveaux circulants de caroténoïdes est observée et peut être attribuée à plusieurs facteurs, dont l'âge, le sexe, le poids corporel, l'activité physique, la génétique et le profil lipidique (58). En effet, le lien entre les concentrations plasmatiques de caroténoïdes et les lipides a été établi dans plusieurs études, entre autres par celle de NHANES 2003–2006 (67). Ce lien semble plausible puisque la majorité des caroténoïdes en circulation sont associés à des lipoprotéines (68, 69). Par exemple, l'α-carotène, le β-carotène et le lycopène sont présents dans le noyau hydrophobe des VLDL (10-16 %), LDL (58-73 %) et HDL (17-26 %) (68). D'autre part, la lutéine et la zéaxanthine sont présents à la surface des lipoprotéines, plus particulièrement des VLDL (16 %), LDL (31 %) et HDL (53 %) tandis que la β-cryptoxanthine est distribuée dans les VLDL (16 %) et de façon égale dans les LDL (40 %) et les HDL (40 %) (68). Les caroténoïdes alimentaires sont inversement associés au risque de MCV et de certains cancers (70), possiblement en raison de leurs propriétés antioxydantes, mais aussi à cause d'autres mécanismes, tels que la régulation de la croissance cellulaire, la réponse immunitaire et la modulation de l'expression des gènes (71, 72).

L'alimentation et les habitudes de vie en général ont une incidence sur la quasi-totalité des processus cellulaires, allant de l'expression des gènes, à la synthèse et à la dégradation des protéines (10). L'alimentation affecte la méthylation de l'ADN, l'expression des gènes ainsi que les concentrations de métabolites et module ainsi le risque de maladies chroniques complexes, dont les MCV, le DB2, le cancer et l'obésité (11-13, 73). Ces éléments seront détaillés dans les paragraphes qui suivent.

1.3 Données omiques

Les progrès dans le développement des technologies à haut débit ont permis de mesurer un grand nombre de molécules biologiques au sein d'une cellule ou d'un tissu (13). Les données moléculaires ainsi que les domaines scientifiques issus de ces technologies à haut débit sont dits omiques. Il existe de nombreux domaines de recherche omiques dont la génomique, l'épigénomique, la transcriptomique, la métabolomique et la protéomique qui correspondent respectivement à l'étude à grande échelle des gènes, de la méthylation de l'ADN, de l'expression des gènes, des métabolites et des protéines. Les différents types de données omiques étudiées dans cette thèse sont détaillés dans les paragraphes qui suivent.

1.3.1 Génomique

La génomique, qui étudie le génome entier, vise à comprendre le rôle des gènes dans différents phénomènes biologiques, notamment leur influence sur la santé et les maladies. Le génome humain est composé de 23 paires de chromosomes qui regroupent toute l'information génétique d'un individu. Les chromosomes, que l'on retrouve dans le noyau des cellules, sont une organisation du matériel génétique appelée acide désoxyribonucléique (ADN) (74). L'ADN (schématisé dans la figure 1.1 ci-dessous) est une macromolécule formée d'une double hélice de deux chaînes polynucléotidiques composées de quatre bases nucléotidiques : adénine, cytosine, guanine et thymine. Les bases nucléotidiques s'apparient de façon déterminée : l'adénine avec la thymine et la guanine avec la cytosine. L'ADN est utilisé pour la formation d'acide ribonucléique (ARN), qui agit comme intermédiaire dans la synthèse des protéines. Dans la séquence de l'ARN, les thymines qui sont remplacées par des uraciles (74).

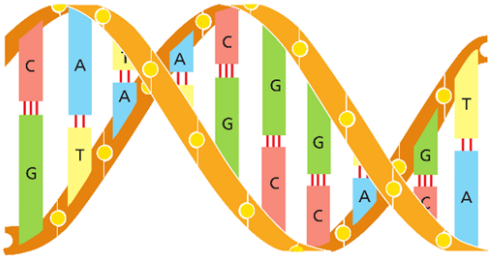


Figure 1.1 Structure de l'ADN
Figure adaptée de *Molecular Biology of the Cell 6th edition, 2015* (75).

Un gène correspond à une séquence de nucléotides qui occupent une position précise sur un chromosome. Les gènes sont composés de bases nucléotidiques et peuvent varier en longueur, de quelques centaines à plus de deux millions de bases. Le génome humain contient environ 20 000 gènes (76). Les gènes sont principalement composés d'exons et d'introns. Les exons servent à la production d'ARN messager (ARNm) et de protéines. Les introns sont des segments d'ADN non codants qui seront éliminés dans le processus d'épissage. Ils sont toutefois impliqués dans la régulation de l'expression génique et dans la stabilité de l'ADN (77). Parmi les 3,2 milliards de paires de bases du génome humain, seulement une petite proportion (~1 %) représente de l'ADN codant (78). Les étapes de la transcription et de la traduction sont nécessaires pour former une protéine à partir de l'ADN. La transcription consiste à synthétiser l'ARNm à partir d'un brin d'ADN. Lors du processus d'épissage, les exons sont conservés pour former l'ARNm. Par la suite, l'ARNm est traduit en une séquence d'acides aminés. Le ribosome, qui est l'organite de la cellule responsable de la synthèse protéique, analyse les séquences de trois nucléotides consécutifs appelés codon pour sélectionner le bon acide aminé. Tout ce processus est schématisé à la figure 1.2 ci-dessous.

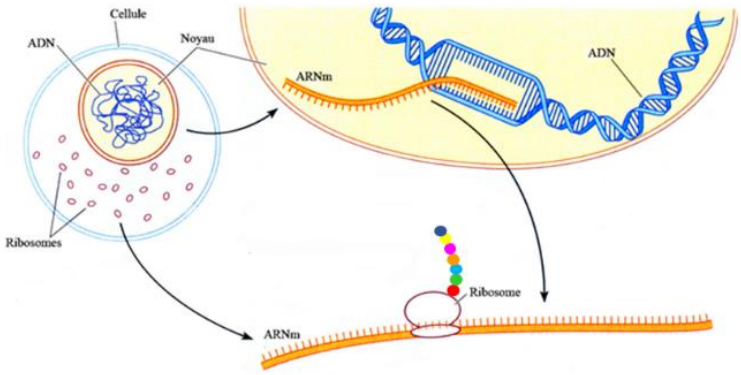


Figure 1.2 Synthèse protéique à partir d'un brin d'ADN
Figure adaptée de *Understanding Normal and Clinical Nutrition 10th edition, 2015* (79). L'ADN dans le noyau de la cellule est transcrit en ARNm. L'ARNm est ensuite traduit en une chaîne d'acides aminés.

Chez l'humain, les 64 codons possibles déterminent ou codent pour un total de 20 acides aminés et trois codons-stop (80). Certaines variations dans le code génétique peuvent avoir des impacts sur la traduction en acides aminés. Les variations dans un gène s'appellent polymorphismes. Les polymorphismes les plus connus, appelés *single nucleotide polymorphisms* (SNPs), sont ceux qui engendrent le changement d'une seule base nucléotidique de la séquence de l'ADN (81). Ce changement peut engendrer une modification du codon et de la protéine produite. Une mutation peut être silencieuse, c'est-à-dire que le changement d'une base de l'ADN n'engendre pas de changement dans la séquence des acides aminés de la protéine. À l'inverse, une mutation non silencieuse altère la séquence des acides aminés de la protéine. Une mutation non-sens cause quant à elle

l'interruption de la traduction, car le codon d'acide aminé a été remplacé par un codon-stop. Enfin, une mutation faux-sens est causée par la substitution d'un codon par un autre désignant un acide aminé différent (74).

Suite à la complétion du *Human Genome Project* en 2003, la génétique a connu un essor fulgurant (82). Les études d'association à l'échelle du génome (*genome-wide association studies* ou GWAS) ont mené à l'identification de milliers de SNPs associés à des maladies. Les GWAS permettent de vérifier l'association entre plusieurs centaines de milliers de polymorphismes et un phénotype précis (83). De nombreux GWAS ont identifié 175 variations génétiques qui affectent les concentrations de lipides dans la population (84). Ces variations expliquent entre 10 à 20 % de la variation totale du C-LDL, du C-HDL et des TG (85). Du fait de leur association avec un profil lipidique altéré, certaines variations sont également associées au risque de MCV. Par exemple, un SNP sur le gène de l'apolipoprotéine E (apo E), qui joue un rôle important dans le métabolisme du cholestérol et des TG, pourrait expliquer de 2 à 11 % de la variation totale des concentrations de cholestérol sanguin (86). De plus, une méta-analyse regroupant 48 études a démontré que l'allèle E4 de l'apo E est un facteur de risque des maladies coronariennes (87), puisqu'il est associé à une augmentation des concentrations de C-LDL (88). Des GWAS ont aussi identifié des variations génétiques associées au développement des maladies coronariennes (89). Parmi les 58 variations associées aux maladies coronariennes, le quart étaient aussi associées aux lipides (90).

L'alimentation est un important facteur environnemental influençant les phénotypes des maladies monogéniques et multifactorielles (91). Il existe de nombreuses interactions entre la génétique, l'alimentation et la santé (92). Les interactions gène-diète reflètent l'effet combiné d'un nutriment et d'un polymorphisme génétique, lequel s'exprime par un phénotype spécifique (91). Ces interactions se produisent donc lorsque l'effet de l'alimentation sur la santé dépend d'un génotype spécifique. Les interactions gène-diète ont d'abord été étudiées dans le contexte des maladies monogéniques, qui sont causées par une altération sur un seul gène, comme par exemple la phénylcétonurie, une maladie causée par une ou des mutations sur le gène phénylalanine hydroxylase (93). Cette maladie mène à une déficience mentale importante résultant d'une accumulation de l'acide aminé phénylalanine au cerveau. Des modifications alimentaires entraînant l'exclusion de la phénylalanine de l'alimentation sont à la base de la prise en charge de cette maladie (93). L'expression phénotypique de cette maladie est donc grandement influencée par l'alimentation. Les interactions gène-diète sont également étudiées dans le contexte de maladies chroniques complexes comme les MCV et de leurs facteurs de risque. Dans l'étude INTERHEART, l'influence des interactions entre le SNP rs2383206, situé dans la région chromosomique 9p21, et les profils alimentaires sur le risque d'infarctus du myocarde a été étudiée (94). Plus précisément, les individus homozygotes pour l'allèle de risque et présentant un faible score au profil alimentaire de type Prudent (caractérisé par un apport élevé en fruits et légumes) avaient un risque d'infarctus du myocarde 1,6 à 2 fois plus élevé (94). Les interactions gène-diète ont aussi été largement étudiées pour

tenter d'expliquer la grande variabilité des réponses par suite de l'adoption de recommandations nutritionnelles. Par exemple, les concentrations de TG à la suite d'une supplémentation en acides gras oméga-3 (AGn-3) sont très variables d'un individu à l'autre. Dans l'étude FINGEN, 40 % des participants n'ont pas obtenu une diminution de leur concentration de TG suite à une supplémentation en AGn-3 d'une durée de huit semaines (95). De même, 29 % des participants à l'étude *Fatty Acid Sensor* (FAS), menée par notre groupe de recherche, n'ont pas obtenu de diminution de leurs concentrations de TG après une supplémentation quotidienne de 3 g d'AGn-3 pendant six semaines (96). L'étiologie de la variabilité de la réponse hypotriglycéridémique aux AGn-3 est multifactorielle et s'explique en partie par des facteurs génétiques. Un score de risque génétique généré par notre équipe permet d'expliquer 49,7 % de la variance de la réponse des TG suite à la supplémentation en AGn-3 (97). Enfin, des variations génétiques interagissent non seulement avec les aliments, mais influencent aussi les préférences alimentaires, l'appétit, la satiété, l'absorption et le métabolisme des nutriments (92).

Les effets de la génétique sur d'autres nutriments spécifiques ont aussi été beaucoup étudiés. Plusieurs GWAS ont identifié des variations génétiques qui affectent les concentrations circulantes de caroténoïdes (98-101). Les variations génétiques peuvent causer des différences au niveau de l'absorption, de l'assimilation, de la distribution, du métabolisme et de l'excrétion des caroténoïdes (98, 102, 103). Dans une méta-analyse combinant les résultats de trois études sur près de 4 000 individus, l'allèle G du SNP rs6564851 était associée à des concentrations plus élevées de β- et α-carotène ainsi qu'à des concentrations plus faibles de lycopène, zéaxanthine et lutéine (100). Ce SNP est intéressant parce qu'il est situé près du gène *β-carotene 15,15'-monooxygenase 1* (*BCMO1*), qui catalyse la première étape de la conversion de la provitamine A des caroténoïdes en vitamine A au niveau de l'intestin grêle (104). De plus, des SNPs situés dans le gène *SCARB1* codant pour le SR-B1 étaient associés aux concentrations de lycopène chez des femmes post-ménopausées (101). Tel que mentionné précédemment, SR-B1 est un transporteur membranaire du cholestérol qui est aussi impliqué dans l'absorption des caroténoïdes (64).

1.3.2 Épigénomique

L'épigénomique est l'étude à l'échelle du génome des altérations moléculaires affectant l'expression des gènes indépendamment de la séquence de l'ADN (105). Il existe plusieurs modifications épigénétiques, dont la méthylation de l'ADN, les modifications des histones et les ARN non-codants, tels que les microARNs (106). Ces modifications sont responsables de la régulation étroite des profils d'expression génique spécifiques aux cellules et aux tissus, déterminant ainsi leurs structures et fonctions (107). Elles peuvent réguler l'expression des gènes et avoir des effets sur le métabolisme et le développement des maladies chroniques complexes (108). La méthylation de l'ADN est la modification épigénétique la mieux caractérisée (109). Elle consiste en l'addition d'un groupement méthyle (symbolisé par un point rouge dans le schéma 1.3 ci-dessous)

principalement au carbone de la cytosine adjacent à la guanine, formant ainsi le dinucléotide cytosine-phosphate-guanine (CpG) (108). Il existe environ 30 millions de sites CpG sur le génome (107).

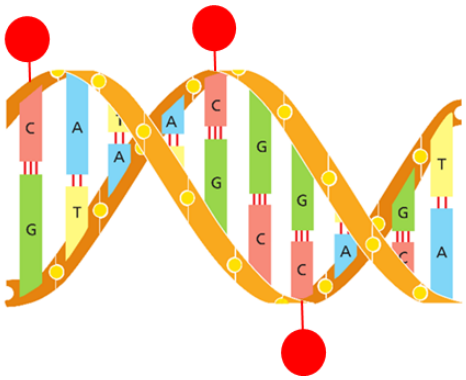


Figure 1.3 Méthylation de l'ADN
Figure adaptée de *Molecular Biology of the Cell 6th edition, 2015* (75). Ajout de groupements méthyles, en rouge, aux cytosines de l'ADN.

Des plateformes d'analyse ont été développées pour mesurer la méthylation des sites CpG à l'échelle du génome. La plateforme 450K d'Illumina permet de mesurer simultanément plus de 485 000 sites CpG. Malgré le nombre impressionnant de sites CpG mesurés, cela ne représente que 2 % de tous les sites CpG du génome humain et les régions géniques sont sur-représentées (75 % des sites) (110). La méthylation de l'ADN évolue durant la vie d'un individu. Lors de la fécondation, les signatures de méthylation des ovocytes et du sperme sont effacées (111). Par la suite, le processus de reprogrammation permettra d'assigner un profil de méthylation spécifique à chaque gène. Toutefois, un certain nombre de sites CpG méthylés sont protégés et transmis à la génération suivante par les cellules germinales primordiales (112). De plus, un changement dans la méthylation globale se produit avec l'âge. De façon générale la méthylation diminue à l'échelle du génome, mais la méthylation de certains promoteurs de gènes augmente (107). Plusieurs facteurs environnementaux, tels que l'alimentation, le tabagisme, l'alcool, le stress et l'exposition aux polluants, influencent la méthylation de l'ADN au courant de la vie (113).

La méthylation de l'ADN permet donc de relier les effets génétiques et les facteurs environnementaux, ce qui en fait un bon outil pour mieux comprendre les mécanismes des maladies multifactorielles. Les changements dans la méthylation de l'ADN contribuent à la régulation des processus biologiques sous-jacents aux MCV, tels que l'athérosclérose, l'hypertension et l'inflammation (114). Par exemple, des différences de méthylation dans les gènes liés à l'angiogenèse ont été rapportées entre des tissus cardiaques normaux et d'autres atteints d'insuffisance cardiaque (115). Certaines études épidémiologiques ont mis en relation les MCV avec la méthylation globale mesurée à l'aide de *long-interspersed nucleotide repetitive elements-1* (*LINE-1*). En ce sens,

des niveaux faibles de méthylation de *LINE-1* étaient associés à un risque plus élevé de cardiopathie ischémique et d'accidents vasculaires cérébraux (116). De plus, l'implication des mécanismes épigénétiques dans la régulation des concentrations de lipides est de plus en plus étudiée. Des études d'association à l'échelle de l'épigénome (*epigenome-wide association studies* ou EWAS) ont identifié des centaines de sites CpG associés aux lipides, dont des sites situés dans le gène *CPT1A*, codant pour la carnitine palmitotransférase responsable de la formation d'acylcarnitines (117, 118). De nombreux sites CpG associés aux lipides sont aussi associés à la dyslipidémie, au syndrome métabolique et au DB2 (119). Par exemple, cg18181703, situé dans le gène *SOCS3*, est associé aux concentrations plasmatiques de C-HDL et de TG en plus d'être associé à certaines caractéristiques du syndrome métabolique, dont l'obésité abdominale et la réponse à l'insuline (120).

Les facteurs environnementaux, comme l'alimentation, ont le potentiel de modifier la méthylation de l'ADN et ainsi moduler le risque de maladies, dont les MCV, le DB2, le cancer et l'obésité (11, 12, 121). Les études réalisées sur la *Dutch Famine Birth Cohort* ont suggéré l'implication des changements de méthylation induits par l'alimentation dans le développement des MCV (122). L'exposition *in utero* à la famine augmentait le risque de développer plus tard dans la vie de nombreuses conditions cardiovasculaires comme la maladie coronarienne, l'obésité et des concentrations plasmatiques de lipides élevées (122). Une récente EWAS a révélé que la méthylation de l'ADN de 30 sites CpG était associée à l'alimentation, évaluée à l'aide des scores de l'alimentation méditerranéenne et de l'*Alternative Healthy Eating Index* (123). Parmi ces sites CpG, plusieurs étaient aussi associés à des facteurs de risque cardiovasculaire et à la mortalité toutes causes confondues (123). De plus, plusieurs études ont considéré l'effet d'aliments ou de nutriments spécifiques sur la méthylation de l'ADN. La disponibilité du donneur de méthyle universel (S-adénosylméthionine) est déterminée par le métabolisme du carbone, impliquant les vitamines B6 et B12, le folate, la bétaïne et la choline, ainsi que les acides aminés méthionine, cystéine, sérine et glycine. La méthylation de l'ADN est alors altérée lorsque l'un de ses composés est déficient (107). Par ailleurs, la plasticité de la méthylation de l'ADN en réponse à une intervention alimentaire a également été démontrée dans de nombreuses études. Par exemple, une supplémentation en AGn-3 engendre des modifications des niveaux de méthylation de plusieurs gènes impliqués dans le métabolisme des lipides (124, 125). D'autres études ont également démontré l'effet de diètes comme la diète méditerranéenne ou de profils alimentaires de type Western ou Prudent sur la méthylation de l'ADN (126, 127). Dans la *North Texas Heath Study*, les sujets qui présentaient un profil alimentaire de type Prudent (caractérisé par un apport élevé en fruits et légumes) avaient une méthylation globale de *LINE-1* plus élevée que les sujets présentant un profil alimentaire de type Western (127).

Seules quelques études ont rapporté un lien entre les caroténoïdes et la méthylation de l'ADN. Dans les études cellulaires, le lycopène a eu des effets modestes à nuls sur la méthylation de *GSTP1*, lequel est impliqué dans les cancers de la prostate et du sein (128, 129). De plus, une étude chez 165 sujets en surpoids et obèses a

révélé une association entre la méthylation de *HERV*-w et de *TNFα* dans les leucocytes sanguins et les apports alimentaires en β-carotène et caroténoïdes (130). Un mécanisme possible serait l'effet inhibiteur des caroténoïdes sur l'activité de l'ADN méthyltransférase, une enzyme qui permet l'ajout du groupement méthyle sur l'ADN (131).

1.3.3 Transcriptomique

La transcriptomique est l'étude du profil d'expression des gènes à l'échelle du génome. La plupart des gènes exercent leurs effets biologiques par leur transcription (expression génique) en ARNm dans un tissu d'intérêt. L'unité de transcription d'un gène est l'ARNm. La transcriptomique est donc l'étude quantitative des quelques 20 000 gènes codants dans le génome. L'expression génique contrôle la quantité et le type de protéines qui sont exprimées dans une cellule, ce qui lui donne ses fonctions (132). Il existe une énorme variabilité dans l'expression des gènes dans les différents tissus et en réponse à des stimuli. Les études de transcriptomique sont réalisées à l'aide de puces d'expression ou du séquençage de l'ARN pour quantifier l'abondance de tous les transcrits exprimés dans des échantillons donnés (132).

La régulation de l'expression génique est complexe et est influencée par plusieurs facteurs, dont des facteurs épigénétiques. Les facteurs de transcription sont des protéines nécessaires à l'initiation de la transcription d'un gène. Ils peuvent se lier directement à l'ADN afin de recruter l'ARN polymérase, qui est l'enzyme responsable de la synthèse de l'ARN à partir de l'ADN (132). La régulation de l'expression génique sous-tend de nombreux processus biologiques fondamentaux, tels que la croissance, la différenciation en organes et tissus et le développement de maladies (132). De nombreuses variations génétiques influencent des traits complexes en modulant l'expression des gènes, modifiant ainsi l'abondance d'une ou de plusieurs protéines (133). En effet, certaines variations génétiques appelées *expression quantitative trait loci* (eQTLs) affectent les niveaux d'expression des gènes et, potentiellement, la susceptibilité aux maladies (134). Récemment, la création du consortium *Genotype-Tissue Expression (GTEx*) a permis de comprendre les fondements génétiques des variations de l'expression génique à travers plus de 40 tissus humains (135). Les eQTLs sont catégorisés selon la distance entre la variation génétique et le gène associé dont on mesure l'expression. Généralement, si la variation génétique est située près du gène associé (<1 mégabase [Mb]), on parle alors d'un *cis*-eQTL, alors que s'ils sont plus éloignés (>1 Mb) ou sur des chromosomes différents, on parle d'un *trans*-eQTL (136). L'étude des eQTLs a permis d'identifier des gènes candidats importants dans plusieurs maladies, dont les MCV, les maladies mentales et le cancer (137, 138).

L'expression génique est de plus en plus étudiée dans le contexte des MCV. En effet, des études ont démontré les changements d'expression de certains gènes chez des sujets souffrant de MCV (139). L'analyse différentielle de l'expression des gènes dans des plaques athéromateuses a permis d'identifier plusieurs voies métaboliques

associées à l'inflammation permettant de renforcer le rôle de l'inflammation dans le développement de l'athérosclérose (140). De nombreux gènes, dont plusieurs impliqués dans le métabolisme des lipides, étaient différentiellement exprimés lors d'un infarctus du myocarde (141). De plus, l'expression de gènes dans le tissu adipeux a été associée à des traits CM, dont les lipides plasmatiques (142). L'expression de centaines de gènes dans des leucocytes sanguins a également été associée aux concentrations de lipides plasmatiques, dont le C-LDL, le C-HDL, le CT et les TG (143). Plus précisément, les niveaux d'expression des gènes impliqués dans le métabolisme des lipides et la chaîne de transport d'électrons étaient corrélés positivement aux concentrations de TG et de C-LDL et négativement aux concentrations de C-HDL (143).

L'alimentation influence l'expression des gènes par différents mécanismes. Certains nutriments sont des ligands naturels de facteurs de transcription qui affectent par la suite l'expression de plusieurs gènes. Ainsi, les AGn-3 sont des ligands naturels des gènes *peroxisome proliferator-activated receptors* (PPARs), qui sont impliqués dans différentes voies métaboliques, dont le métabolisme des lipides, le métabolisme des glucides et les réponses inflammatoires (144). Les caroténoïdes et leurs dérivés (rétinoïdes) peuvent aussi exercer leurs effets sur l'expression des gènes par l'entremise de plusieurs systèmes de transcription, tels que les récepteurs rétinoïdes, le *nuclear factor-kappa B* et les PPARs (145). Seules quelques études ont été réalisées sur l'effet des profils alimentaires sur les niveaux d'expression des gènes. Une revue de littérature comprenant 14 études a démontré que la diète méditerranéenne était associée à une diminution de l'expression des gènes pro-inflammatoires, comparativement à une diète riche en acides gras saturés (146). Une étude de la cohorte PREDIMED a aussi montré que les diètes méditerranéennes enrichies en huile d'olive ou en noix modifiaient l'expression des gènes impliqués dans les voies métaboliques de l'athérosclérose et de l'hypertension (147). De plus, d'autres études ont par ailleurs considéré l'effet des différents types d'acides gras sur l'expression des gènes. Une diète riche en acides gras saturés a induit un profil d'expression pro-inflammatoire comparativement à une diète riche en acides gras monoinsaturés (148). Des études réalisées par notre groupe de recherche ont également révélé que les profils d'expression des gènes impliqués dans la réponse inflammatoire, l'immunité et l'athérosclérose étaient différents selon les profils alimentaires Prudent et Western (149). Finalement, la modification du profil d'expression des gènes engendre aussi des changements dans les profils de métabolites (150).

1.3.4 Métabolomique

La métabolomique est l'étude de tous les métabolites de faible poids moléculaire produits au cours du métabolisme et contenus dans des cellules, des tissus ou des organes (18). Les métabolites reflètent les effets exogènes de facteurs environnementaux et les effets endogènes de facteurs génétiques et pathologiques. Ils représentent des biomarqueurs de processus biologiques normaux, de pathologies et de réponse à des

interventions thérapeutiques ou pharmacologiques. Il y a généralement une distinction entre les métabolites endogènes et les biomarqueurs d'exposition à la diète, ces derniers étant issus directement de la digestion, de l'absorption et de la biotransformation des aliments (7). Parmi le large éventail de méthodes de dosage des métabolites, la chromatographie liquide couplée en ligne à la spectrométrie de masse (LC-MS) et la résonance magnétique nucléaire sont les plus fréquentes (151). La LC-MS est la méthode la plus utilisée dans les études métabolomiques en raison de sa sensibilité, de sa spécificité et de sa capacité à couvrir un large éventail de métabolites (152). Les analyses métabolomiques peuvent être non ciblées (tous les métabolites détectés sont mesurés) ou ciblées (pour certains groupes prédéfinis de métabolites). Avec l'approche non ciblée, il est possible d'identifier et de quantifier des centaines de métabolites (153). Avec l'approche ciblée, les classes de métabolites les plus fréquemment mesurés sont les lipides, les acylcarnitines et les acides aminés. Les lipides, dont le cholestérol, les phospholipides, les lysophospholipides et les sphingolipides, représentent une classe complexe de métabolites aux propriétés diverses (154). Les acylcarnitines formés d'un ester d'acide gras et d'une molécule de carnitine représentent les produits finaux du catabolisme de divers acides gras libres et d'acides aminés à chaîne ramifiée.

La métabolomique permet de mieux comprendre les mécanismes sous-jacents aux maladies complexes, telles que les MCV, le DB2 et le cancer, pour lesquelles les facteurs environnementaux ont un impact considérable (151). Des métabolites ont été associés à l'obésité, au DB2 et à l'athérosclérose (155-157). Une signature métabolomique dans des plaques athérosclérotiques a été associée au contenu de médiateurs inflammatoires dans la plaque et à la présence de symptômes (156). En ce sens, une étude a révélé que 89 métabolites permettaient de stratifier divers sous-types de MCV chez 2324 patients (158). Plusieurs lipides, dont les phospholipides et les lysophospholipides, sont de plus en plus étudiés dans le contexte des MCV (151). Un ensemble de métabolites, composé de l'histidine, de la phénylalanine, de la spermidine et de la phosphatidylcholine 34:4, a permis de diagnostiquer l'insuffisance cardiaque (159). Les métabolites ont également été associés à la santé CM (160, 161). En effet, les concentrations d'acylcarnitines ont été associées à l'obésité et à la résistance à l'insuline (162, 163). Plusieurs études rapportent des associations entre, d'une part, les acides aminés ramifiés, leurs dérivés (les acylcarnitines C3 et C5), et les acides aminés aromatiques et, d'autre part, le risque de DB2 ou de résistance à l'insuline (164-167). Plusieurs acides aminés, notamment la glutamine, le glutamate, la proline, l'alanine et la tyrosine, ont été associés à de multiples facteurs de risque CM, incluant la tension artérielle et les concentrations d'insuline, de TG et de C-HDL dans la *Framingham Heart Study* et la *Malmö Diet and Cancer Study* (161).

L'alimentation a un impact important sur le profil des métabolites. Une étude a démontré qu'une diète riche en acide palmitique, un acide gras saturé, a augmenté les concentrations d'acylcarnitines dans le sérum (168). Les concentrations plasmatiques et urinaires de plusieurs métabolites, dont des acylcarnitines (C2 et C3) et des

acides aminés (3-méthylhistidine), ont également été associées à la consommation de viande ou de poisson (169). Une consommation élevée de viande rouge et de matières grasses a été associée à un profil de métabolites composé d'acides aminés ramifiés, d'acides aminés aromatiques et d'acylcarnitine C3, lequel profil a été associé à un risque plus élevé de DB2 (155). Dans une étude publiée par notre groupe de recherche, le profil alimentaire de type Western a par ailleurs été associé à des concentrations augmentées d'acylcarnitines et d'acides aminés ramifiés (13).

1.4 Biologie des systèmes

Au cours des deux dernières décennies, le génotypage à haut débit sur des cohortes de milliers de patients a permis d'identifier des milliers de variations génétiques associées à plusieurs maladies (83, 170). Il est cependant difficile d'expliquer les fonctions des variations identifiées et de comprendre les mécanismes sous-jacents aux associations. En ce sens, seulement 7 % des variations identifiées sont situées dans des régions codantes pour des protéines (171). De plus, les variations identifiées n'expliquent généralement qu'une fraction de la composante héréditaire des maladies.

Étant donné qu'un seul type de données omiques ne permet généralement pas de saisir la complexité des événements moléculaires à l'origine des maladies, une approche intégrative combinant plusieurs données omiques s'avère idéale pour déchiffrer la physiopathologie des traits complexes (14). Jusqu'à maintenant, la plupart des études qui ont combiné des données omiques se sont concentrées sur l'effet de variations génétiques sur les déterminants omiques, notamment la méthylation de l'ADN (*methylation quantitative trait loci* [meQTL]), l'expression des gènes (eQTL) et les métabolites (*metabolite quantitative trait loci* [mQTL]) (18). Des GWAS ont permis d'identifier de nombreuses variations génétiques associées à la méthylation de sites spécifiques (meQTLs) (172). Étant donné la spécificité tissulaire et cellulaire de la méthylation de l'ADN, la distribution des meQTLs varie en fonction du tissu et du type cellulaire (173). De plus, les niveaux de certains métabolites sont fortement associés à des variations génétiques (mQTL) dont certaines sont également associées à des maladies (174). En somme, les études ont répertorié un grand nombre de QTLs dans différents volets omiques en vue de mieux comprendre la régulation génétique à travers différents tissus et types de cellules ainsi que son impact sur le développement de maladies. Toutefois, ces études incluent rarement des facteurs environnementaux comme l'alimentation. Le défi réside dans l'obtention d'une compréhension globale du système moléculaire interactif qui tienne compte simultanément de plusieurs types de données omiques (18). Étant donné les différences inhérentes des données issues de différentes approches omiques, leur l'intégration demeure une entreprise ardue. De plus, les outils analytiques spécifiques utilisés pour les données omiques individuelles ne sont pas adaptés pour permettre des comparaisons appropriées ou une intégration adéquate de plusieurs données omiques (16).

La biologie des systèmes vise une compréhension plus holistique de la biologie en intégrant plusieurs types de données omiques à l'aide de modèles mathématiques (15, 16). Malgré le large consensus sur l'importance de la biologie des systèmes, on ne s'entend pas sur sa définition (15). En gros, la biologie des systèmes est un domaine de recherche interdisciplinaire qui vise à mieux comprendre la biologie complexe des cellules et des organismes, ce qui inclut leur croissance, leur développement et leur évolution vers diverses maladies, en combinant des données générées par de multiples approches omiques (16). Les processus biologiques, tels que le développement de maladies, impliquent un système dynamique et interactif de couches moléculaires (génome, épigénome, transcriptome et métabolome) qui sont influencées par l'environnement. Dans le cadre de cette thèse, le phénotype, soit l'ensemble des traits observables d'un organisme, est fonction des interactions du génome, de l'épigénome, du transcriptome et du métabolome, entre eux et avec l'environnement, comme illustré par le schéma ci-dessous (17).

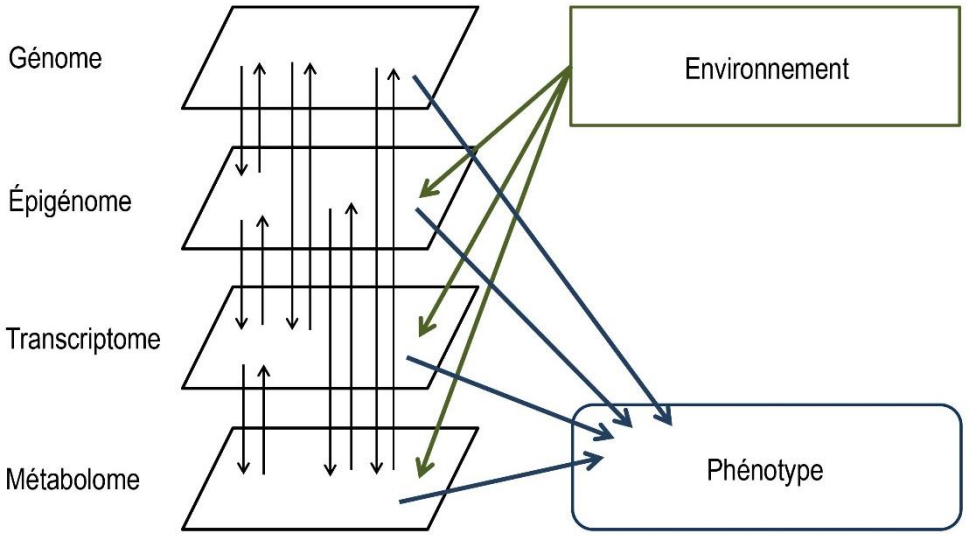


Figure 1.4 Schéma représentant différentes interactions de la biologie des systèmes
Figure adaptée de Dunn *et al*. *Chemical Society Reviews* 2011 (17).

Les approches multi-omiques suscitent beaucoup d'intérêt, considérant leurs applications potentielles en recherche en santé (175). Par exemple, le *Cancer Genome Atlas* a permis de grandement affiner les classes de cancer en croisant des données de génomique, d'épigénomique, de transcriptomique et de protéomique (176). Également dans le domaine de la cancérologie, plusieurs études ont conduit au développement de nouveaux algorithmes d'intégration de données multi-omiques afin de prédire des résultats phénotypiques (177-179). L'intégration multi-omique peut permettre de mieux comprendre les interactions biologiques entre les différents niveaux de données omiques et même d'identifier de nouveaux biomarqueurs pour stratifier les patients en sous-types de maladies (180). Plusieurs études ont utilisé des approches issues de la biologie des systèmes pour mieux comprendre la physiopathologie des MCV (181). La biologie des systèmes a aussi le

potentiel de classifier les individus en fonction de leur risque de développer certaines maladies, comme les MCV, en utilisant un score de risque polygénique (182). En plus de son potentiel dans la prédiction et le diagnostic précoce de maladies, la biologie des systèmes devrait devenir de plus en plus puissante pour améliorer le traitement et le pronostic des maladies (14).

Les différentes composantes de la biologie des systèmes varient en fonction du système étudié. Dans cette thèse, l'environnement au sens large et aussi l'alimentation, représentée par les caroténoïdes plasmatiques, seront étudiés. Les volets omiques de la génomique, de l'épigénomique, de la transcriptomique et de la métabolomique seront aussi pris en compte en plus des caractéristiques biochimiques sanguines reliées à la santé CM.

1.5 Méthodes d'analyse de données omiques

Il existe de nombreuses méthodes pour analyser les données omiques, lesquelles peuvent être classées en approches ciblées et non ciblées (183). Les approches ciblées permettent de mettre en relation des variations génétiques ainsi que des niveaux de méthylation, d'expression génique ou de métabolites spécifiques avec un trait d'intérêt. Tel que mentionné précédemment, un nombre important de données omiques spécifiques ont été associées à divers facteurs de risque CM et à l'alimentation. Dans les approches non ciblées, l'ensemble des données omiques est mis en relation avec un trait. Les GWAS, qui permettent l'identification de biomarqueurs omiques de maladies, en sont un bon exemple. L'intégration de plusieurs données omiques a permis de mieux comprendre la physiopathologie de plusieurs traits complexes. Ces études se sont principalement concentrées sur l'effet de variations génétiques sur les déterminants omiques (meQTL, eQTL et mQTL) (18).

L'utilisation et l'intégration de données omiques dans la recherche en santé présentent plusieurs défis. Tout d'abord, les maladies chroniques sont multifactorielles et sont influencées autant par des facteurs génétiques que par des facteurs environnementaux (18). La génétique quantitative permet de tenir compte de la variabilité multifactorielle de traits complexes (184). Elle permet d'estimer les contributions de la génétique et de l'environnement dans la variance d'un trait (184). Cependant, l'analyse des différentes données omiques génère une quantité phénoménale de données. Le nombre de données omiques dépasse largement le nombre de sujets ou d'échantillons dans les études. Ce nombre impressionnant de données représente une sérieuse difficulté pour l'analyse, l'intégration et l'association de ces données à des traits environnementaux comme l'alimentation ou des traits phénotypiques comme la santé CM. Il devient difficile de tenir compte de l'ensemble des données à l'échelle du génome sans subir les effets négatifs des comparaisons multiples inhérents à ce type de données (18). De plus, considérant le haut degré d'interconnexion entre les gènes, les données omiques sont aussi fortement interreliées (20). En effet, tel que mentionné, il existe plusieurs interactions non seulement entre les données omiques, mais aussi avec l'environnement et les traits phénotypiques (17). Des méthodes

telles que l'analyse de réseaux de corrélations pondérées ont été développées pour tenir compte de ces masses de données omiques interreliées. Dans cette thèse, deux méthodes d'analyse de données omiques sont abordées : la génétique quantitative et l'analyse de réseaux de corrélations pondérées.

1.5.1 Génétique quantitative

Des phénotypes tels que le poids, la taille ou l'indice de masse corporel varient considérablement d'un individu à l'autre et ne montrent pas un héritage mendélien simple (19). La génétique quantitative, également appelée génétique des caractères complexes, étudie ces phénotypes dans lesquels de nombreux gènes et facteurs non génétiques sont impliqués (19). Les traits complexes sont déterminés par l'action conjointe de plusieurs gènes et facteurs environnementaux (185). Pour un phénotype donné, des causes multiples, à la fois génétiques et non génétiques, ainsi que leurs interactions, contribuent à sa variation. La génétique quantitative se penche sur des phénotypes dont la variabilité est multifactorielle. Elle est basée sur des fondements statistiques, dont la décomposition de la variance d'un trait proposée par Fisher (186) et développée par Cockerham (187) et Kempthorne (188). Le degré d'expression phénotypique (variance phénotypique [V_P]) d'un trait varie en fonction d'effets génétiques (G) et d'effets environnementaux uniques à chaque individu ou résiduels (E) :

$$V_P = V_G + V_E$$

La variable G se décline en différentes composantes : valeur génétique additive (A), valeur de dominance (D) et valeur d'épistasie (I) (189). La valeur génétique additive (A) réfère au total des effets individuels des allèles transmis. La valeur de dominance (D) réfère à la mesure dans laquelle un trait apparaît dans une population ou un individu, résultant d'interactions alléliques à un locus spécifique. La valeur d'épistasie (I) réfère à la partie de la variance phénotypique et génotypique due à l'épistasie, c'est-à-dire aux interactions alléliques sur différents loci :

$$V_G = V_A + V_D + V_I$$

Au sens large, l'héritabilité désigne la proportion de variation phénotypique qui est due à des facteurs génétiques (additif, dominant et épistasie) (190). Au sens étroit, l'héritabilité désigne la proportion de la variation phénotypique qui est due à des facteurs génétiques additifs (190). Souvent, il n'y a pas de distinction faite entre ces deux termes. Le terme héritabilité est aussi utilisé en référence à la ressemblance entre les parents et leur(s) enfant(s). Une héritabilité élevée signifie qu'il y a une forte ressemblance entre les parents et leur(s) enfant(s) par rapport à un trait spécifique.

Il est également possible de tenir compte d'autres composantes dans le modèle, comme par exemple l'effet de l'environnement partagé par les membres d'une famille (C) et l'effet de l'environnement partagé par des jumeaux

(T). L'effet de l'environnement partagé par les membres d'une famille illustre l'effet de tous les éléments qui sont partagés au sein d'une famille, comme l'alimentation, le mode de vie, le statut socio-économique, la situation géographique, l'exposition à des polluants, etc. La variance phénotypique du trait est donc fonction des effets génétiques, de l'environnement partagé et des effets environnementaux uniques à chaque individu (189) :

$$V_P = V_G + V_C + V_E$$

La variance phénotypique globale est répartie en plusieurs composantes en utilisant la covariance observée dans le trait entre les membres de la famille et des matrices de structure qui prédisent la covariance entre les individus qui serait attribuable à ces composantes (189). Prenons l'exemple d'un modèle où la variance phénotypique de la lutéine, un caroténoïde, est décomposée en effets de la génétique (V_G), de l'environnement partagé (V_C) et de l'environnement unique ou résiduel à chaque individu (V_E). Pour simplifier le concept, l'exemple comprend deux familles : la première composée d'un père, d'une mère et d'un enfant et la seconde composée d'une mère et d'un enfant (Figure 1.5). Chaque composante de la variance (V_G, V_C et V_E) est accompagnée d'une matrice de structure qui prédit la covariance, entre tous les individus, attribuable à cette composante. La matrice de structure de l'effet génétique comprend des valeurs qui représentent les pourcentages d'effets génétiques partagés entre deux individus, soit:

- 1 : 100 % de la génétique est partagée entre les individus (lorsqu'un individu est comparé à lui-même ou à son jumeau monozygote);
- ½ : 50 % de la génétique est partagée entre les individus (entre le père et l'enfant, entre la mère et l'enfant et entre les enfants);
- 0 : 0 % de la génétique est partagée entre les individus (entre les parents ou les membres de familles différentes).

Pour ce qui est de la matrice de structure de l'effet de l'environnement partagé, elle illustre les familles. Des valeurs de 1 sont attribuées aux membres d'une même famille et des valeurs de 0 aux membres de familles différentes. Finalement, la matrice de structure de l'effet unique ou résiduel représente les effets qui sont uniques à chaque individu. Des valeurs de 1 sont donc attribuées lorsque l'individu est comparé à lui-même et des valeurs de 0 sont attribuées partout ailleurs (189).

Père 1, mère 1 et enfant 1

Mère 2 et enfant 2

Génétique

	P1	M1	E1	M2	E2
P1	1	0	1/2	0	0
M1	0	1	1/2	0	0
E1	1/2	1/2	1	0	0
M2	0	0	0	1	1/2
E2	0	0	0	1/2	1

Environnement partagé

	P1	M1	E1	M2	E2
P1	1	1	1	0	0
M1	1	1	1	0	0
E1	1	1	1	0	0
M2	0	0	0	1	1
E2	0	0	0	1	1

Résiduel ou unique

	P1	M1	E1	M2	E2
P1	1	0	0	0	0
M1	0	1	0	0	0
E1	0	0	1	0	0
M2	0	0	0	1	0
E2	0	0	0	0	1

Figure 1.5 Matrices de structure des effets génétiques, de l'environnement partagé et de l'effet résiduel ou unique à chaque individu

En utilisant ces trois matrices de covariance, il est possible de connaître la portion de la variance attribuable à ces composantes. Ces estimés de variance sont utilisés pour calculer les différents types d'héritabilité. L'héritabilité génétique représente la proportion de la variance attribuable aux effets génétiques sur la variance totale du modèle [V_G / (V_G + V_C + V_E)]. L'effet de l'environnement partagé représente la proportion de la variance attribuable aux effets de l'environnement partagé sur la variance totale [V_C / (V_G +V_C + V_E)]. L'héritabilité maximale représente la proportion de la variance attribuable aux effets génétiques et de l'environnement partagé sur la variance totale [(V_G + V_C) / V_G + V_C + V_E)]. Dans l'exemple de la lutéine cité à l'article du chapitre 6, l'héritabilité génétique est de 43,8 %, l'effet de l'environnement partagé de 44,5 % et l'héritabilité maximale de 88,3 % (43,8 % + 44,5%) (191). Cela signifie que 88,3 % de la variance des niveaux de lutéine est dû à des facteurs génétiques et environnementaux dans notre projet (191). Dans les travaux de cette thèse, ces matrices de structure ont été réalisées pour 48 individus, et ce, sur 472 494 niveaux de méthylation, 18 160 niveaux d'expression, 147 niveaux de métabolites et sept niveaux de caroténoïdes.

Il existe dans l'ensemble deux types de devis d'étude pour l'évaluation de l'héritabilité d'un trait. Il y a les études de jumeaux (monozygotes ou hétérozygotes) et les études de familles (192). Dans les études de jumeaux monozygotes, il est possible d'estimer avec plus de précision l'effet unique ou résiduel étant donné qu'ils partagent presque 100 % de leur bagage génétique de même que leur environnement. Dans les études de familles, les enfants biologiques partagent 50 % de leur bagage génétique avec leurs parents (184). Les ressemblances familiales surviennent lorsque les membres d'une famille qui partagent des facteurs génétiques et environnementaux sont plus similaires que des sujets non apparentés.

La méthode de décomposition de la variance utilisée dans les analyses d'héritabilité peut être étendue à une analyse génétique bivariée (deux traits). Cette analyse permet de tester la corrélation entre deux traits, appelée corrélation phénotypique. Cette corrélation est ensuite décomposée en corrélation génétique et corrélation environnementale (193). La corrélation génétique permet d'estimer l'effet génétique additif partagé entre les deux traits. Une corrélation génétique positive signifie que les facteurs génétiques affectent les deux traits de la même façon (augmentent ou diminuent le niveau des traits). Une corrélation génétique négative signifie que les facteurs génétiques affectent les deux traits de façon inverse. La corrélation environnementale permet d'estimer l'effet de facteurs environnementaux partagés entre les deux traits. Cette approche est utilisée dans plusieurs domaines, dont celui des maladies chroniques étant donné la multitude de facteurs de risque qui y sont interreliés (194).

Les ressemblances familiales de nombreux traits et maladies, dont les MCV, ont été étudiées. Selon des études de familles et de jumeaux, l'héritabilité des maladies coronariennes est d'environ 40 à 60 % (195). Une étude réalisée sur le *Swedish Twin Registry* auprès de 21 000 sujets suivis sur plus de 35 ans a rapporté que les estimés d'héritabilité du décès attribuable à la maladie coronarienne étaient de 57 % et 38 % chez les hommes et les femmes, respectivement (196). L'héritabilité des facteurs de risque cardiovasculaire, dont les concentrations de lipides a aussi été largement évaluée. Les estimés d'héritabilité des taux sériques de lipides varient de 28 % à 78 % (197). Les moyennes et les intervalles d'estimés d'héritabilité sont d'environ 55 % (39 % à 71 %) pour le CT, 52 % (28 % à 80 %) pour le C-HDL, 49 % (23 % à 72 %) pour le C-LDL et 32 % (13 % à 61 %) pour les TG (198).

Étant donné l'implication des déterminants omiques dans la physiopathologie des maladies chroniques, leurs ressemblances familiales ont été évaluées dans de nombreuses études. L'héritabilité de la méthylation de l'ADN est étroitement liée aux effets génétiques puisque les variations génétiques situées sur ou près d'un site CpG modifient les niveaux de méthylation de ce site (199). En effet, les études de familles et de jumeaux ont démontré que les facteurs génétiques sont les principaux déterminants des ressemblances familiales des niveaux de méthylation à l'échelle du génome (200, 201). Les ressemblances familiales de l'expression génique à l'échelle du génome sont aussi principalement attribuables à des facteurs génétiques (202). Les concentrations de métabolites sanguins présentent une grande variabilité interindividuelle due à des facteurs génétiques et environnementaux (203). Les ressemblances familiales des niveaux de métabolites sanguins sont dues à des facteurs génétiques et de l'environnement partagé (204). En somme, la littérature indique que ce sont les facteurs génétiques qui contribuent le plus aux ressemblances familiales des données omiques. Une attention particulière doit être portée aux composantes des modèles d'héritabilité puisque la majorité des études n'estiment pas les effets de l'environnement partagé dans les ressemblances familiales des données omiques.

1.5.2 Réseaux de corrélations pondérées

L'analyse de réseaux permet d'avoir une représentation simple des interactions entre différentes données (protéines, cellules, etc.) (21). L'approche de réseaux de corrélations pondérées aussi appelée *weighted gene correlation network analysis* (WGCNA) est une méthode d'analyse statistique de la biologie des systèmes. Elle a été élaborée pour tirer profit des nombreuses interactions dans les différentes données omiques, qui n'ont pas été beaucoup explorées jusqu'à maintenant. Plus spécifiquement, elle permet d'identifier des regroupements de gènes interreliés, appelés modules, et de les associer à différents traits phénotypiques. La méthode WGCNA a été élaborée par Steve Horvath et Peter Langfelder de l'Université de la Californie (21, 205). Elle est beaucoup utilisée avec des données omiques, dont l'épigénomique et la transcriptomique (205). Elle comprend quatre grandes étapes distinctes. La première est celle de la construction d'une matrice de similarité basée sur des corrélations entre les données omiques, ce qui permet d'illustrer les interactions entre les gènes. La deuxième étape consiste à identifier des modules, qui sont en fait des regroupements de gènes intercorrélés. Cette étape représente le fondement de la méthode de WGCNA, qui est de considérer des modules ou des regroupements de gènes intercorrélés plutôt que des gènes pris individuellement. Les modules sont souvent enrichis avec des gènes qui partagent des fonctions biologiques similaires. Dans le dendrogramme de modules reproduit ci-dessous, chaque ligne représente une valeur de méthylation; plus les niveaux de méthylation sont corrélés, plus les lignes sont rapprochées. Les niveaux de méthylation fortement corrélés sont alors assignés au même module, identifié par une couleur.

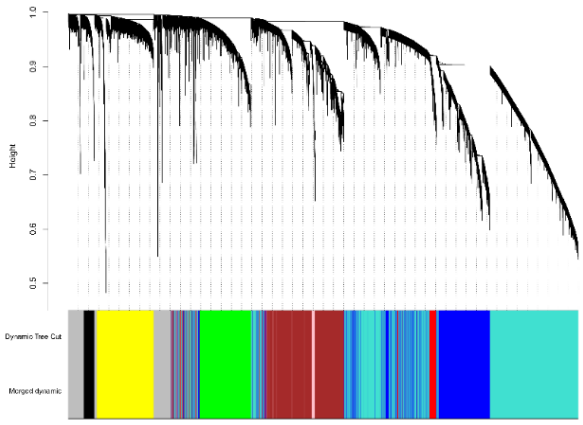


Figure 1.6 Dendrogramme représentant des modules de co-méthylation
Figure tirée de Tremblay *et al*. *Nutrients* 2019 (206). Le dendrogramme représente les niveaux de méthylation des 20 687 sites CpG et les neuf modules correspondants.

À la troisième étape, on associe les modules à des traits phénotypiques d'intérêt dans le but d'identifier des modules biologiquement pertinents. Pour y arriver, le *module eigengene* (ME) est corrélé aux traits. Il représente la composante principale du module, soit le niveau global de méthylation ou d'expression des gènes regroupés dans un module. Lors de la quatrième et dernière étape, on identifie les gènes clés, appelés *hub genes*, qui sont

fortement corrélés à la fois à l'ensemble d'un module d'intérêt et au trait phénotypique (205). Pour ce faire, deux critères sont utilisés soit celui du *gene significance* (GS) et celui du *module membership* (MM). Le critère GS représente la corrélation entre la méthylation ou l'expression du gène et le trait. Le MM représente la corrélation entre la méthylation ou l'expression d'un gène et le ME d'un module. Ainsi, plus le MM est élevé, plus la méthylation ou l'expression du gène représente bien la méthylation ou l'expression globale du module. Les *hub genes* sont donc les gènes possédant les valeurs de MM et de GS les plus élevées (207).

Cette méthode comporte plusieurs avantages. Tout d'abord, elle permet de regrouper les données afin de limiter le nombre de corrélations avec des traits phénotypiques. Par exemple, si on voulait associer la méthylation de l'ADN à l'échelle du génome au tour de taille, il faudrait faire des centaines de milliers de corrélations ou de régressions linéaires entre les niveaux de méthylation des différents sites CpG et le tour de taille. En utilisant WGCNA, on restreint les corrélations aux quelques modules (généralement moins de 10) (205) de sites CpG avec des niveaux de méthylation intercorrélés. De plus, WGCNA permet non seulement d'intégrer différents types de données omiques, mais aussi des données omiques provenant de différents tissus (208, 209). WGCNA a été appliquée avec succès dans différents domaines de recherche, dont la cancérologie, l'obésité et les MCV (210, 211). Une étude a rapporté des associations entre des modules de co-expression, l'indice de masse corporelle (IMC) et le calcium des artères coronaires (212). La méthode a également été appliquée pour identifier un module de co-expression associé à la cardiomyopathie hypertrophique chez des patients avec des maladies coronariennes (213). Elle a été utilisée autant pour des facteurs de risque CM que pour l'alimentation. WGCNA a ainsi permis d'identifier des *hub genes* impliqués dans l'hyperlipidémie (214). Enfin, cette méthode a été utilisée pour associer des modules d'expression génique avec des profils alimentaires chez des adultes en santé (215).

Chapitre 2 Objectifs et hypothèses

La prise en charge des individus souffrant de MCV repose sur l'évaluation et le traitement de plusieurs facteurs de risque CM. L'adoption de saines habitudes de vie, incluant notamment une alimentation équilibrée, demeure la pierre angulaire de la prévention des maladies chroniques, dont les MCV. L'alimentation peut en plus affecter la méthylation de l'ADN, l'expression des gènes et les concentrations de métabolites et ainsi moduler le risque de MCV. Une approche intégrative combinant les données issues de plusieurs approches omiques permet de mieux saisir la complexité des processus biologiques à l'origine des maladies. La biologie des systèmes étudie les interactions complexes entre les différentes données omiques les unes avec les autres et avec l'environnement, ainsi que leur influence sur la santé. Jusqu'à maintenant, la plupart des études qui ont combiné des données omiques se sont limitées à l'effet de variations génétiques sur les déterminants omiques. De plus, ces études incluent rarement des facteurs environnementaux, tels que l'alimentation. À ce jour, aucune étude n'a utilisé une approche de la biologie des systèmes pour analyser de façon intégrée les données omiques dans l'impact de l'alimentation sur la santé CM au sein de familles en santé.

Les travaux de cette thèse reposent sur l'étude observationnelle GENERATION, qui visait à évaluer, à l'aide de biomarqueurs d'exposition à la diète, l'impact de l'alimentation sur les données omiques de même que sur les facteurs de risque CM à différents stades de la vie. L'étude a été réalisée à l'Institut sur la nutrition et les aliments fonctionnels (INAF) de l'Université Laval à Québec. Le recrutement des sujets s'est déroulé de mai 2011 à novembre 2014. L'étude comprend 48 sujets répartis en 16 familles composées au minimum de la mère et d'un enfant. Les sujets devaient être en bonne santé et avoir un IMC entre 18 et 35 kg/m^2. Des prises de sang ont été effectuées pour tous les sujets afin de mesurer des données omiques (méthylation de l'ADN, expression génique et métabolites), des biomarqueurs d'exposition à la diète (caroténoïdes) et des facteurs de risque CM, dont le profil lipidique. Le devis de l'étude est détaillé dans l'article présenté au chapitre 3.

L'**objectif général** du présent projet est d'étudier l'impact des déterminants omiques sur la relation entre l'alimentation et la santé CM. Les **objectifs spécifiques** sont les suivants :

1. En utilisant la génétique quantitative,
    - Caractériser l'héritabilité de la méthylation de l'ADN, de l'expression génique, des métabolites plasmatiques et des caroténoïdes plasmatiques;
    - Vérifier si les relations des données omiques entre elles et avec des facteurs de risque CM peuvent être expliquées par des facteurs génétiques et environnementaux.

2. En utilisant les réseaux de corrélations pondérées,
    - Évaluer le rôle des données omiques individuelles et combinées dans la relation entre les caroténoïdes et le profil lipidique.

L'objectif de l'article présenté au **chapitre 3** était d'estimer les contributions des effets génétiques et de l'environnement partagé dans les ressemblances familiales des niveaux de méthylation. Comme c'est généralement le cas dans ce domaine, aucune hypothèse n'avait été avancée avant de réaliser cette étude.

Les objectifs de l'article présenté au **chapitre 4** étaient, dans un premier temps, d'estimer les contributions des effets génétiques et de l'environnement partagé dans les ressemblances familiales des niveaux d'expression génique à l'échelle du génome. Dans un deuxième temps et dans le but d'étudier plus en détail les mécanismes biologiques, nous avons également comparé les résultats des ressemblances familiales de l'expression génique avec ceux des ressemblances familiales des niveaux de méthylation de l'ADN des leucocytes sanguins dans la même cohorte. Aucune hypothèse n'avait été avancée avant de réaliser cette étude puisque ces approches sont basées sur le fait qu'aucune hypothèse *a priori* n'est formulée.

L'hypothèse de l'article présenté au **chapitre 5** était que les effets génétiques et de l'environnement partagé contribuent à la variance des concentrations plasmatiques de métabolites et que des effets génétiques et environnementaux partagés expliquent leurs corrélations avec des facteurs de risque CM. Sur la base des connaissances dans le domaine, il était ici possible de travailler avec des hypothèses formulées *a priori*.

L'objectif de l'article présenté au **chapitre 6** était d'estimer les contributions des effets génétiques et de l'environnement partagé dans la variance des concentrations plasmatiques de caroténoïdes (α-carotène, β-carotène, β-cryptoxanthine, lutéine, lycopène, zéaxanthine et caroténoïdes totaux) et de vérifier si leurs corrélations phénotypiques avec les facteurs de risque CM pourraient être expliquées par des effets génétiques et environnementaux partagés. L'hypothèse posée était que les effets génétiques et de l'environnement partagé contribuent à la variance des concentrations plasmatiques de caroténoïdes et que les effets génétiques et environnementaux partagés expliquent leurs corrélations phénotypiques avec les facteurs de risque CM.

L'objectif de l'article présenté au **chapitre 7** était d'examiner le rôle de l'expression génique dans la relation entre les concentrations plasmatiques de caroténoïdes totaux et le profil lipidique en utilisant l'analyse de réseaux de corrélations pondérées (WGCNA). L'hypothèse posée était que les niveaux d'expression génique à l'échelle du génome sont associés aux caroténoïdes totaux et que les modules de gènes associés aux caroténoïdes sont aussi corrélés au profil lipidique.

L'objectif de l'article présenté au **chapitre 8** était d'examiner le rôle de la méthylation de l'ADN dans la relation entre les concentrations plasmatiques de caroténoïdes totaux et le profil lipidique en utilisant WGCNA. L'hypothèse posée était que les niveaux de méthylation à l'échelle du génome sont associés aux caroténoïdes totaux et que les modules de gènes associés aux caroténoïdes sont aussi corrélés au profil lipidique.

L'objectif de l'article présenté au **chapitre 9** était d'examiner le rôle potentiel des données multi-omiques (méthylation de l'ADN et expression génique) dans l'interconnexion entre l'alimentation, représentée par les caroténoïdes totaux, et le profil lipidique en utilisant WGCNA. Aucune hypothèse n'avait été avancée avant de réaliser cette étude.

Chapitre 3 Ressemblances familiales des niveaux de méthylation de l'ADN des leucocytes sanguins

Bénédicte L. Tremblay, Frédéric Guénard, Benoît Lamarche, Louis Pérusse, Marie-Claude Vohl

L'article présenté dans ce chapitre s'intitule : *Familial resemblances in blood leukocyte DNA methylation levels.*

Cet article est publié dans la revue : *Epigenetics*. 2016 Nov;11(11):831-838.

Résumé

Introduction : Des facteurs épigénétiques tels que la méthylation de l'ADN sont des altérations de l'ADN affectant l'expression génique qui peuvent transmettre les effets de l'environnement à travers les générations. Seulement quelques études ont démontré le concept d'héritabilité épigénétique chez l'humain. Notre objectif est de quantifier les déterminants, génétiques et de l'environnement partagé, des ressemblances familiales des niveaux de méthylation de l'ADN en utilisant une cohorte de familles.

Méthodes : La méthylation de l'ADN a été mesurée chez 48 Canadiens français dans le cadre de l'étude GENERATION. La puce de méthylation *HumanMethylation450* (Illumina Inc., USA) a été utilisée pour mesurer les niveaux de méthylation de 485 577 sites CpG dans les leucocytes sanguins. L'héritabilité a été évaluée à l'aide de la méthode de la décomposition de la variance mis en œuvre dans le logiciel QTDT, qui partitionne la variance en effets génétiques (G), de l'environnement partagé (C) et de l'environnement non-partagé (E).

Résultats : Nous avons calculé l'héritabilité maximale, l'héritabilité génétique et l'effet de l'environnement partagé pour tous les sites CpG (12,7 %, 8,2 % et 4,5 %, respectivement) et pour les sites CpG statistiquement significatifs (81,8 %, 26,9 %, 54,9 %, respectivement). Une héritabilité maximale plus élevée a été observée dans le complexe majeur d'histocompatibilité sur le chromosome 6.

Conclusions : En conclusion, les ressemblances familiales des niveaux de méthylation étaient principalement attribuables à des facteurs génétiques lorsque l'on considère la moyenne à travers le génome, mais l'effet de l'environnement partagé joue un rôle important lorsque l'on considère les sites CpG statistiquement significatifs. D'autres études à l'échelle du génome sur des échantillons plus grands combinées à des études d'associations à l'échelle du génome sont nécessaires pour mieux comprendre les mécanismes sous-jacents à l'héritabilité de la méthylation de l'ADN.

Abstract

Introduction: Epigenetic factors such as DNA methylation are DNA alterations affecting gene expression that can convey environmental information through generations. Only a few studies have demonstrated epigenetic inheritance in humans. Our objective is to quantify genetic and common environmental determinants of familial resemblances in DNA methylation levels, using a family-based sample.

Methods: DNA methylation was measured in 48 French Canadians from 16 families as part of the GENERATION Study. We used *HumanMethylation450* array (Illumina Inc., USA) to measure DNA methylation levels in blood leukocytes on 485 577 CpG sites. The heritability was assessed using the variance components method implemented in the QTDT software, which partitions the variance into polygenic (G), common environmental (C) and non-shared environmental (E) effects.

Results: We computed maximal heritability, genetic heritability, and common environmental effect for all probes (12.7%, 8.2%, and 4.5% respectively) and for statistically significant probes (81.8%, 26.9%, and 54.9% respectively). Higher maximal heritability was observed in the Major Histocompatibility Complex region on chromosome 6.

Conclusions: In conclusion, familial resemblances in DNA methylation levels are mainly attributable to genetic factors when considering the average across the genome, but common environmental effect plays an important role when considering statistically significant probes. Further genome-wide studies on larger samples combined with GWAS are needed to better understand the underlying mechanisms of DNA methylation heritability.

Title page

Familial resemblances in blood leukocyte DNA methylation levels

Bénédicte L. Tremblay[1], Frédéric Guénard[1], Benoît Lamarche[1], Louis Pérusse[2], Marie-Claude Vohl[1, 2*]

1. Institute of Nutrition and Functional Foods (INAF), Laval University, 2440 Hochelaga Blvd, Quebec City, QC, G1V 0A6, Canada
2. CHU de Québec Research Center – Endocrinology and Nephrology, 2705 Laurier Blvd, Quebec City, QC, G1V 4G2, Canada

Corresponding author*:
Marie-Claude Vohl, Ph.D.
Institute of Nutrition and Functional Foods (INAF), Laval University
Pavillon des Services**,** 2440 Hochelaga Blvd, Quebec City, QC G1V 0A6, Canada
Phone: 418 656-2131 ext. 4676; Fax: 418 656-5877
E-mail: marie-claude.vohl@fsaa.ulaval.ca

Keywords: Familial resemblances; DNA methylation; Metabolic pathways; Blood leukocytes; Microarray.

Introduction

Epigenetics is the study of mitotically and meiotically heritable DNA alterations affecting gene expression that are not mediated by changes in DNA sequence (1). There are several epigenetic modifications, namely DNA methylation, histone modification, occupancy of chromatin factors, and changes in chromatin structure (2). DNA methylation is the best-characterized epigenetic modification and involves the methylation of cytosine residues, mainly at cytosine-phosphate-guanine (CpG) dinucleotides (3). DNA methylation, as other epigenetic modifications, can be altered by environmental conditions such as diet and smoking habits and thus affecting expression of various genes. The modification of DNA methylation by environment may be implicated in a wide range of cardiovascular risk factors including hypertension, atherosclerosis, and inflammation (4-6). Epigenetic modifications including DNA methylation have also the potential to convey the effects of environmental exposures transgenerationally (7).

There are two main forms of inheritance of epigenetic modifications: genetic inheritance and epigenetic inheritance via the gametes (3). Genetic inheritance refers to the effect of DNA sequence on the epigenetic marks with the example of a single nucleotide polymorphism (SNP) located at, or nearby, a CpG site that can disrupt DNA methylation at this site (8). Epigenetic inheritance via the gametes is the transmission of epigenetic marks, such as DNA methylation, across generations, which can occur during epigenetic reprogramming events (9, 10). However, demonstrations of epigenetic inheritance in humans remain rare.

Some epidemiological studies have demonstrated that a prenatal exposure to the Dutch Famine from December 1944 to April 1945 increased the risk of developing cardiovascular disease later in life (11). Some other transgenerational studies have reported that epigenetic modifications may persist for longer than a single generation. Indeed, a study by Kaati et al. reported the effect of grandparental food supply on the risk of grandchildren mortality due to type 2 diabetes (12).

Studies with twins have also demonstrated a greater similarity in DNA methylation levels in monozygotic twins compared to dizygotic twins (13). Twin studies also reported an average estimated genetic heritability of DNA methylation (proportion of variance explained by additive genetic factors) between 12 and 18% in whole blood (14, 15), 5% in placenta (15), and 7% in human umbilical vascular endothelial cells (15). These studies using promoter-specific genome-wide DNA methylation array (Illumina 27K) examined only a fraction of methylation variations. More recently, studies using genome-wide DNA methylation array (Illumina 450K) have reported an average genetic heritability of 19% in peripheral blood leukocytes and 19% in adipose tissue (9, 16).

Transgenerational similarity in DNA methylation levels has also been reported (13). Several genome-wide association studies with methylation quantitative trait loci (mQTL) in both *cis* and *trans* locations have confirmed

the genetic heritability of DNA methylation (14, 17, 18). However, it is difficult to understand to what extent the observed familial resemblance in DNA methylation is due to genetic heritability and/or common environmental effect.

Our objective is thus to quantify the contribution of genetic and common environmental effects in the familial resemblances in DNA methylation levels using a family-based sample of 48 French Canadians from 16 families.

Methods

Patients and design

A total of 48 subjects from 16 families were recruited in the Greater Quebec City metropolitan area, in Canada, as part of the GENERATION Study. Families were composed of 16 mothers, 6 fathers, and 26 children. To be eligible parents had to be the biological parents of their child (or children). Families living under the same roof comprised at least the mother and one child aged between 8 and 18. Parents had to be in good general health, non-smokers, with body mass index (BMI) ranging between 18 and 35 kg/m^2, and free of any metabolic conditions requiring treatment although the use of Synthroid® (levothyroxine) or oral contraceptive was tolerated. Children also had to be non-smokers and in good general health. They were not eligible if using psycho-stimulators [Ritalin® (methylphenidate), Concerta® (methylphenidate), and Strattera® (atomoxetine)]. Parents and children were asked to complete several dietary, physical activity, medical history, and pregnancy questionnaires under the supervision of a registered dietitian during their visit at the Institute of Nutrition and Functional Foods (INAF). The experimental protocol was approved by the Ethics Committees of Laval University Hospital Research Center and Laval University.

Anthropometric and metabolic measurements

Body weight, waist girth, and height were measured according to the procedures recommended by the Airlie Conference. (19) Blood samples were collected from an antecubital vein into vacutainer tubes containing EDTA after 12-hour overnight fast and 48-hour alcohol abstinence. Plasma was separated by centrifugation (2 500g for 10 min at 4 °C) and blood leukocytes were collected.

DNA extraction and methylation analysis

Genomic DNA was extracted from blood leukocytes using the GenElute Blood Genomic DNA kit (Sigma-Aldrich, St. Louis, MO, USA) for all 48 samples. DNA was quantified using both NanoDrop Spectrophotometer (Thermo Scientific, Wilmington, DE, USA) and PicoGreen DNA methods. Methylation levels were measured using Infinium Human Methylation 450 array (Illumina, San Diego, CA, USA). Bisulfite conversion and quantitative DNA methylation analysis were processed at the McGill University and Genome Quebec Innovation Center (Montreal, Canada). Illumina GenomeStudio software v2011.1 and the Methylation Module were used to analyze methylation data on 485 577 CpG sites. All samples were retained after quality control steps (bisulfite conversion, extension, staining, hybridization, target removal, negative and nonpolymorphic control probes). Methylation levels (beta values [β] vary from 0 to 1) were estimated as the proportion of total signal intensity from methylated-specific probe. Global normalisation using control probes was performed in GenomeStudio. Probes with a detection *P*-value > 0.01 in more than 5 subjects (> 10% of all subjects) were removed, as well as probes on the X and Y chromosomes (to eliminate gender bias) and probes mapped to multiple chromosomes (20). Thus, 472 494 probes were considered in the analysis.

Statistical analysis

R software v2.14.1 (R Foundation for Statistical Computing; http://www.r-project.org) (21) was used to compute the average absolute correlation of DNA methylation between relative pairs across all 472 494 probes. For heritability analysis, corrections were made for the effects of microarray, position on microarray, sex, age, age^2, sex*age, and sex*age^2, using a standard least squares model in JMP software v12. We used residuals from this model to compute heritability estimates using the variance component method implemented in QTDT v2.6.1 (22). We used a full general model in which the variance in methylation levels of each probe was partitioned into polygenic effects (Vg), common environmental effects shared by family members (Vc), and non-shared environmental effects unique to each individual (Ve). We tested this full general model against a null model of no familial resemblance in which Vg = Vc = 0. We then computed average maximal heritability, as the proportion of variance accounted by genetic and common environmental effects (Vg+Vc/Vg+Vc+Ve), average genetic heritability, as the proportion of variance accounted by genetics effects (Vg/Vg+Vc+Ve), and the average proportion of variance accounted by common environmental effects (Vc/Vg+Vc+Ve). Additionally, we computed heritability estimates in probes showing a significant familial effect (Vg and Vc significantly different from zero ($P \leq 0.05$)). We computed FDR-corrected *P*-values to account for multiple testing. We also used an alternative genetic model in which the variance in methylation levels of each probe was partitioned into Vg and Ve. We then computed average genetic heritability, as the proportion of variance accounted by genetic effects (Vg/Vg+Ve). We used RCircos package in R software to make the Circos plot (23). Finally, we used the IPA system (Ingenuity® Systems, www.ingenuity.com) to analyze pathways overrepresented among genes of significant and FDR-corrected significant probes. Using a right-tailed Fisher's exact test, IPA measured the likelihood that genes belong to specific overrepresented pathways. Pathway analysis was conducted with genes of significant probes (n=13 621), significant probes with a maximal heritability of 100% (n=1186), and FDR-corrected significant probes (n=105).

Results

Correlations of methylation levels between relative pairs

Average absolute correlations across all 472 494 probes were calculated between pairs from normalized methylation levels (Table 1). The average siblings correlation was of 0.299 ± 0.211 (n=13), the mother-offspring correlation was of 0.267 ± 0.183 (n=26), the parent-offspring correlation was of 0.243 ± 0.170 (n=37), whereas the unrelated individuals correlation was of 0.029 ±0.028 (n=1178). Comparisons of genome-wide methylation levels within pairs indicated that mother-offspring and parent-offspring pairs had more similar DNA methylation levels compared to unrelated individuals.

DNA methylation heritability analyses

We conducted heritability analyses to better characterize the similarity in genome-wide methylation levels between related individuals. Table 2 shows estimates of maximal heritability, genetic heritability, and common environmental effect from the full general model. When considering all probes (n=472 494), we obtained an average maximal heritability of 12.7%, a genetic heritability of 8.2%, and a common environmental effect of 4.5% (Table 2). Figure 1a shows the distribution of maximal heritability estimates for DNA methylation levels of all probes (n=472 494) with heritability from 0 to 100% on the x-axis and the count (number of probes) on the y-axis. A total of 215 136 probes (45.5% of all probes) and 1507 probes (0.32% of all probes) had an estimated maximal heritability of 0% and 100%, respectively.

When considering the 13 621 probes showing a significant ($P \leq 0.05$) familial effect, we obtained an average maximal heritability of 63.9%, a genetic heritability of 39.3%, and a common environmental effect of 24.6% (Table 2). When maximal heritability estimates of significant probes are plotted, the distribution shifts to the right, in accordance with the higher maximal heritability estimates ranging from 29.6 to 100% (Figure 1b). A total of 1186 probes (8.71% of all significant probes) gave an estimated maximal heritability of 100%.

Lastly, the same analysis was repeated for probes showing a significant familial effect after False Discovery Rate (FDR) correction ($P \leq 0.05$). We obtained 105 FDR-corrected significant probes giving an average maximal heritability of 81.8%, a genetic heritability of 26.9%, and a common environmental effect of 54.9% (Table 2). Figure 1c shows the distribution of maximal heritability estimates for these probes with maximal heritability ranging from 60 to 100%. A total of 30 probes (28.6% of all FDR-corrected significant probes) had an estimated maximal heritability of 100%. The distribution of genetic heritability and common environmental effect for the 105 most significant probes is depicted in the Figure 2. Genetic heritability ranged from 0 to 91.9% while common environmental effect ranged from 8 to 77.3%. A total of 48 probes had an estimated genetic heritability of 0%. On the other hand, majority of probes with a genetic component (51 out of 57 probes) were also found significant in the alternative genetic model (data not shown).

Moreover, Figure 3 shows Circos plot depicting the distribution of maximal heritability estimates across the genome. There is a visible higher maximal heritability on chromosome 6, which corresponds to the Major Histocompatibility Complex (MHC) region (24). A total of six probes out of 105 were located in the MHC, and five were associated with genes (*TAP2, TRIM10*, and *PSMB8*). These 105 FDR-corrected significant probes were assigned to 69 genes as illustrated in Figure 3.

Pathways analyses

Ingenuity Pathway Analysis (IPA) revealed that 218 pathways were significantly overrepresented among genes (n=5625) of the 13 621 significant probes (Supplementary Table 1). Among these pathways, 15 were also overrepresented among genes (n=69) of the 105 FDR-corrected significant probes (Table 3). These latter were related to signaling pathways (Antigen Presentation Pathway , Epithelial Adherens Junction Signaling, Cell Cycle: G1/S Checkpoint Regulation , Glioma Signaling, Pancreatic Adenocarcinoma Signaling, and Protein Ubiquitination Pathway), cholesterol metabolism (Superpathway of Cholesterol Biosynthesis, Cholesterol Biosynthesis I, Mevalonate Pathway I, Cholesterol Biosynthesis II [via 24,25-dihydrolanosterol], and Cholesterol Biosynthesis III [via Desmosterol]), alanine metabolism (Alanine Degradation III and Alanine Biosynthesis II), sugar nucleotides metabolism (GDP-mannose Biosynthesis), and cell structure biosynthesis (Colanic Acid Building Blocks Biosynthesis) (Table 3). Among those 15 pathways, only 2 were also overrepresented among genes (n=680) of the 1186 significant probes with a maximal heritability of 100% (Antigen Presentation Pathway and Cell Cycle: G1/S Checkpoint Regulation) (Supplementary Table 2).

Discussion

The aim of this study was to quantify contribution of genetic and common environmental effects in the familial resemblance of DNA methylation levels in a family-based sample of 48 subjects from 16 families. Average absolute correlations between relative pairs revealed that DNA methylation levels of related individuals are more similar than DNA methylation levels of unrelated individuals. Similar absolute correlations were reported by McRae et al., suggesting that correlations of DNA methylation levels within family members is caused by an underlying genetic similarity (9).

In accordance with other studies, genetic factors across the genome seem to be the major determinant of familial resemblances in DNA methylation levels. Indeed, we obtained a ratio of two-thirds for genetic heritability and one-third for common environmental effect for all probes (n=472 494) and significant probes (n=13 621). However, when considering only the most significant probes (n=105), the ratio is reversed. In most significant probes, genetic component had an important contribution to familial resemblances in DNA methylation levels only in some probes. In the same manner, the majority of most significant probes with a genetic component in the full general model are found in the alternative genetic model. According to this alternative model, genetic heritability estimate for all probes was of 14.2% which is slightly lower than estimates reported in other published studies. Indeed, McRae et al. using 614 individuals from 117 families of European descent reported a genetic heritability of 19.9% in peripheral blood leukocytes (9). Grunberg et al. using 648 twins of European descent also reported a genetic heritability of 19% in adipose tissue (16). Our genetic heritability estimate is more similar to the one reported by Gervin et al. that ranges from 2 to 16% in MHC region in CD4+ lymphocytes of healthy monozygotic and dizygotic twins (25). In this study, genetic heritability also varies by gene region types (CpG islands, 5´ ends of genes, conserved noncoding regions, and randomly selected regions) (25). Similarly, we also observed a variation in maximal heritability depending on the chromosomic and gene position, as depicted in Figure 3. However, our results should be compared with caution considering that Gervin et al. only analysed 1760 CpG sites in the MHC region while we considered 472 494 CpG sites across the genome (25). The difference in the types of cells (all leucocytes vs CD4+ lymphocytes) should also be considered.

Regarding common environmental effect in all probes, we reported a higher percentage (4.5%) than McRae et al. (2.3%) (9). We can postulate that the homogenous environment to which subjects of our cohort are exposed may allow us to detect more significant common environmental effect. Indeed, all French-Canadian subjects are Caucasians, lived in the same city (Quebec City), and have similar socio-economic characteristics and dietary patterns (data not shown). Moreover, dietary habits may affect DNA methylation levels of highly heritable probes. This association may be plausible since pathways related to cholesterol and alanine metabolism were significantly overrepresented among genes of the most significant probes (n=105). Most importantly, the effect of common environmental component in those probes is non-negligible since it contributes to different extent (8

to 77.3%) to maximal heritability in all 105 probes. The inclusion of the common environmental effect in the full general model thus allows us to detect more significant probes that contribute to better explain familial resemblances in DNA methylation levels.

Overall, we reported a maximal heritability of 12.7%, which is relatively low, but maximal heritability varies a lot from one CpG site to another (ranging from 0 to 100%) and several CpG sites (13 621 sites) show evidence of familial effect (maximal heritability > 29.6%). Some CpG sites (105 sites) also show strong evidence of familial effect (maximal heritability > 60%) even after FDR correction for multiple testing. We also observed highly heritable probes in the MHC region on chromosome 6 as reported by McRae et al. (9). Interestingly, they also reported that these probes located in the MHC had highly significant *cis* mQTL.

Finally, using pathways analysis, we identified overrepresented pathways among genes of significant and FDR-corrected significant probes. A total of 15 pathways were still overrepresented among genes of FDR-corrected probes thus suggesting a strong overrepresentation of these pathways. Interestingly, among the 15 pathways, five were related to cholesterol metabolism. Studies in twins and family studies have demonstrated the strong heritability of cholesterol levels. Indeed, high-density lipoprotein cholesterol has an estimated heritability ranging from 40 to 60% (26-28). In addition, the heritability of low-density lipoprotein peak particle diameter is 52% when considering age, BMI, and plasma triglyceride concentrations (29). Thus, this may suggest a potential link between strong DNA methylation heritability of some CpG sites within genes involved in cholesterol metabolism and the reported heritability of cholesterol levels. Obviously, we do not suggest a direct association, but it sheds light on specific epigenetic factors that may contribute to the heritability of cholesterol levels. Moreover, most overrepresented pathways among genes represented by the 1186 probes with a maximal heritability of 100% were related to cellular immune response (Supplementary Table 2). This finding is in line with the fact that we observed a higher maximal heritability in the MHC region, which is known to contain genes involved in immune function (30).

The present study has its own limitations. The main one is the small sample size, which limits the statistical power required to detect significant genetic heritability and common environmental effect. Given the family design used in this study, the effects of common environment and epigenetic inheritance are highly intertwined and consequently difficult to estimate separately. McRae et al. using 117 families failed to distinguish these two effects and reported that estimated common environmental effects may be inflated by potential epigenetic inheritance (9). Therefore, separating common environmental effect and epigenetic inheritance will require a very large sample size (9). In counterpart, as mentioned earlier, the homogeneity of the present study sample may allow us to detect more significant common environmental effect. Our small sample size may also explain why we observed such a large proportion of probes (45.5% of all probes) giving an estimated maximal heritability

of 0%. In fact, a study by Grunberg et al. stated that a sample of 648 twins was not sufficient to obtain reliable heritability estimates of less than 10% (16). Thus, our small sample size may affect the reliability of low values of heritability estimates and partly explain the large number of probes with an estimated maximal heritability of 0%. In addition, a SNP found in a probe location or nearby may cause methylation differences and impact the estimation of DNA methylation heritability (8). A study by Lemire et al. demonstrated that there is extensive long-range regulation of CpG methylation associated with genetic variation (31). The fact that CpG sites associated with a SNP were not excluded in the present study may affect heritability analysis and explain why we obtain 1186 significant probes that give an estimated maximal heritability of 100%. To verify that, we conducted further analysis to investigate SNPs position in these probe locations using the 1000 Genomes Project Phase 1 Version 3 (32). We found that 1014 probes (85.5% of significant probes with a maximal heritability of 100%) have one or more SNPs within their location (defined as ± 25 base pairs from the CpG site). This suggests that the presence of SNPs at, or nearby, a CpG site may contribute to increase genetic heritability and therefore influence maximal heritability. However, we cannot confirm that SNPs ascertained in the 1000 Genomes Project are present in our population. In addition, McRae et al. reported that the effect of SNPs within a probe location have limited effect on average genetic heritability across the genome (9). Finally, DNA methylation is specific to the type of cell, tissue, locus, and developmental stages, which affects DNA methylation heritability estimates (13). However, McRae et al. reported little effect of the cellular composition of blood samples on the heritability estimates for the majority of probes (genetic heritability reduced from 19.9% to 17.6%) (9). Other studies have also reported that DNA methylation patterns are globally correlated between 11 different somatic tissues (33, 34). However, it is important to consider cellular composition when interpreting and comparing DNA methylation heritability results.

In conclusion, in a family-based sample of 48 French Canadians, familial resemblances in blood leukocytes DNA methylation levels are mainly due to genetic factors when considering the average across the genome, but common environmental effect plays an important role when considering most significant probes. It would be interesting to verify possible associations between common environment effect such as dietary habits and methylation levels in highly heritable probes. Considering the small sample size upon which the present study is based, results should be viewed with caution. Further epigenome genome-wide studies on larger samples combined with GWAS are needed to better understand the underlying mechanisms of DNA methylation heritability.

Acknowledgments

We would like to thank Véronique Garneau and Christian Couture, who contributed to the success of this study. We also thank Catherine Raymond for the laboratory work.

Funding

This work was supported by the Canada Research Chair in Genomics Applied to Nutrition and Metabolic Health. MCV is Tier 1 Canada Research Chair in Genomics Applied to Nutrition and Metabolic Health. BLT is a recipient of a scholarship from Fonds de recherche du Québec – Santé (FRQS).

Disclosure of potential conflicts of interest

Authors have no potential conflicts of interest.

References

1. Waterland RA. Epigenetic mechanisms and gastrointestinal development. The Journal of pediatrics. 2006;149(5 Suppl):S137-42.

2. Feinberg AP. Epigenetics at the epicenter of modern medicine. JAMA. 2008;299(11):1345-50.

3. Daxinger L, Whitelaw E. Understanding transgenerational epigenetic inheritance via the gametes in mammals. Nature reviews Genetics. 2012;13(3):153-62.

4. Friso S, Pizzolo F, Choi SW, Guarini P, Castagna A, Ravagnani V, et al. Epigenetic control of 11 beta-hydroxysteroid dehydrogenase 2 gene promoter is related to human hypertension. Atherosclerosis. 2008;199(2):323-7.

5. Turunen MP, Aavik E, Yla-Herttuala S. Epigenetics and atherosclerosis. Biochimica et biophysica acta. 2009;1790(9):886-91.

6. Makar KW, Wilson CB. DNA methylation is a nonredundant repressor of the Th2 effector program. Journal of immunology. 2004;173(7):4402-6.

7. Pembrey ME, Bygren LO, Kaati G, Edvinsson S, Northstone K, Sjostrom M, et al. Sex-specific, male-line transgenerational responses in humans. European journal of human genetics : EJHG. 2006;14(2):159-66.

8. Shoemaker R, Deng J, Wang W, Zhang K. Allele-specific methylation is prevalent and is contributed by CpG-SNPs in the human genome. Genome research. 2010;20(7):883-9.

9. McRae AF, Powell JE, Henders AK, Bowdler L, Hemani G, Shah S, et al. Contribution of genetic variation to transgenerational inheritance of DNA methylation. Genome biology. 2014;15(5):R73.

10. Guibert S, Forne T, Weber M. Global profiling of DNA methylation erasure in mouse primordial germ cells. Genome research. 2012;22(4):633-41.

11. Roseboom TJ, van der Meulen JH, Osmond C, Barker DJ, Ravelli AC, Schroeder-Tanka JM, et al. Coronary heart disease after prenatal exposure to the Dutch famine, 1944-45. Heart. 2000;84(6):595-8.

12. Kaati G, Bygren LO, Pembrey M, Sjostrom M. Transgenerational response to nutrition, early life circumstances and longevity. European journal of human genetics : EJHG. 2007;15(7):784-90.

13. Bell JT, Spector TD. DNA methylation studies using twins: what are they telling us? Genome biology. 2012;13(10):172.

14. Bell JT, Tsai PC, Yang TP, Pidsley R, Nisbet J, Glass D, et al. Epigenome-wide scans identify differentially methylated regions for age and age-related phenotypes in a healthy ageing population. PLoS genetics. 2012;8(4):e1002629.

15. Gordon L, Joo JE, Powell JE, Ollikainen M, Novakovic B, Li X, et al. Neonatal DNA methylation profile in human twins is specified by a complex interplay between intrauterine environmental and genetic factors, subject to tissue-specific influence. Genome research. 2012;22(8):1395-406.

16. Grundberg E, Meduri E, Sandling JK, Hedman AK, Keildson S, Buil A, et al. Global analysis of DNA methylation variation in adipose tissue from twins reveals links to disease-associated variants in distal regulatory elements. American journal of human genetics. 2013;93(5):876-90.

17. Gibbs JR, van der Brug MP, Hernandez DG, Traynor BJ, Nalls MA, Lai SL, et al. Abundant quantitative trait loci exist for DNA methylation and gene expression in human brain. PLoS genetics. 2010;6(5):e1000952.

18. Drong AW, Nicholson G, Hedman AK, Meduri E, Grundberg E, Small KS, et al. The presence of methylation quantitative trait loci indicates a direct genetic influence on the level of DNA methylation in adipose tissue. PloS one. 2013;8(2):e55923.

19. Callaway C, Chumlea W, Bouchard C, Himes J, Lohman T, Martin A, et al. Standardization of anthropomeric measurements : The Airlie (VA) Consensus Conference. Champaign, IR, USA; 1988.

20. Price ME, Cotton AM, Lam LL, Farre P, Emberly E, Brown CJ, et al. Additional annotation enhances potential for biologically-relevant analysis of the Illumina Infinium HumanMethylation450 BeadChip array. Epigenetics & chromatin. 2013;6(1):4.

21. Team RC. R: A language and environment for statistical computing. R Foundation for Statistical Computing. Vienna, Austria. 2013.

22. Abecasis GR, Cardon LR, Cookson WO. A general test of association for quantitative traits in nuclear families. American journal of human genetics. 2000;66(1):279-92.

23. Zhang H, Meltzer P, Davis S. RCircos: an R package for Circos 2D track plots. BMC bioinformatics. 2013;14:244.

24. Trowsdale J, Knight JC. Major histocompatibility complex genomics and human disease. Annual review of genomics and human genetics. 2013;14:301-23.

25. Gervin K, Hammero M, Akselsen HE, Moe R, Nygard H, Brandt I, et al. Extensive variation and low heritability of DNA methylation identified in a twin study. Genome research. 2011;21(11):1813-21.

26. Weissglas-Volkov D, Pajukanta P. Genetic causes of high and low serum HDL-cholesterol. Journal of lipid research. 2010;51(8):2032-57.

27. Qasim A, Rader DJ. Human genetics of variation in high-density lipoprotein cholesterol. Current atherosclerosis reports. 2006;8(3):198-205.

28. Lusis AJ, Mar R, Pajukanta P. Genetics of atherosclerosis. Annual review of genomics and human genetics. 2004;5:189-218.

29. Bosse Y, Vohl MC, Despres JP, Lamarche B, Rice T, Rao DC, et al. Heritability of LDL peak particle diameter in the Quebec Family Study. Genetic epidemiology. 2003;25(4):375-81.

30. Traherne JA. Human MHC architecture and evolution: implications for disease association studies. International journal of immunogenetics. 2008;35(3):179-92.

31. Lemire M, Zaidi SH, Ban M, Ge B, Aissi D, Germain M, et al. Long-range epigenetic regulation is conferred by genetic variation located at thousands of independent loci. Nature communications. 2015;6:6326.

32. Genomes Project C, Abecasis GR, Auton A, Brooks LD, DePristo MA, Durbin RM, et al. An integrated map of genetic variation from 1,092 human genomes. Nature. 2012;491(7422):56-65.

33. Byun HM, Siegmund KD, Pan F, Weisenberger DJ, Kanel G, Laird PW, et al. Epigenetic profiling of somatic tissues from human autopsy specimens identifies tissue- and individual-specific DNA methylation patterns. Human molecular genetics. 2009;18(24):4808-17.

34. Fan S, Zhang X. CpG island methylation pattern in different human tissues and its correlation with gene expression. Biochemical and biophysical research communications. 2009;383(4):421-5.

Tables

Table 3.1 Average absolute correlations across all probes of normalised methylation levels between relative pairs

Relative pairs	n	Correlation ± SD
Siblings	13	0.299 ± 0.211
Mother-Offspring	26	0.267 ± 0.183
Parent-Offspring	37	0.243 ± 0.170
Unrelated	1178	0.029 ± 0.028

Abbreviation : Standard deviation (SD)

Table 3.2 Heritability estimates

Type of heritability estimates (%)	All probes (n=472 494)	Significant probes (n=13 621)	FDR-corrected significant probes (n=105)
Maximal heritability	12.7	63.9	81.8
Genetic heritability	8.2	39.3	26.9
Common environmental effect	4.5	24.6	54.9

Table 3.3 Overrepresented pathways identified among genes of FDR-corrected significant probes (n=105)

IPA Canonical Pathways	*P* value	Annotated genes
Superpathway of Cholesterol Biosynthesis[b]	0.00324	*DHCR7, MVK*
Antigen Presentation Pathway[a]	0.00562	*PSMB8, TAP2*
Alanine Degradation III[c]	0.00603	*GPT*
Alanine Biosynthesis II[c]	0.00603	*GPT*
Epithelial Adherens Junction Signaling[a]	0.00977	*YES1, MAGI1, NOTCH1*
Cell Cycle: G1/S Checkpoint Regulation[a]	0.01622	*NRG1, RB1*
GDP-mannose Biosynthesis[d]	0.01820	*PMM1*
Glioma Signaling[a]	0.03548	*RB1, IGF2R*
Cholesterol Biosynthesis I[b]	0.03890	*DHCR7*
Mevalonate Pathway I[b]	0.03890	*MVK*
Cholesterol Biosynthesis II (via 24,25-dihydrolanosterol)[b]	0.03890	*DHCR7*
Cholesterol Biosynthesis III (via Desmosterol)[b]	0.03890	*DHCR7*
Pancreatic Adenocarcinoma Signaling[a]	0.04074	*RB1, NOTCH1*
Colanic Acid Building Blocks Biosynthesis[e]	0.04169	*PMM1*
Protein Ubiquitination Pathway[a]	0.04169	*UBE2E3, PSMB8, TAP2*

Pathways related to: a: signaling pathways (n=6), b: cholesterol metabolism (n=5), c: alanine metabolism (n=2), d: sugar nucleotides metabolism (n=1), and e: cell structure biosynthesis (n=1)

Figures

Figure 3.1 Distribution of maximal heritability estimates for DNA methylation of A) all probes (n=472 494), B) significant probes (n=13 621), C) FDR-corrected significant probes (n=105)

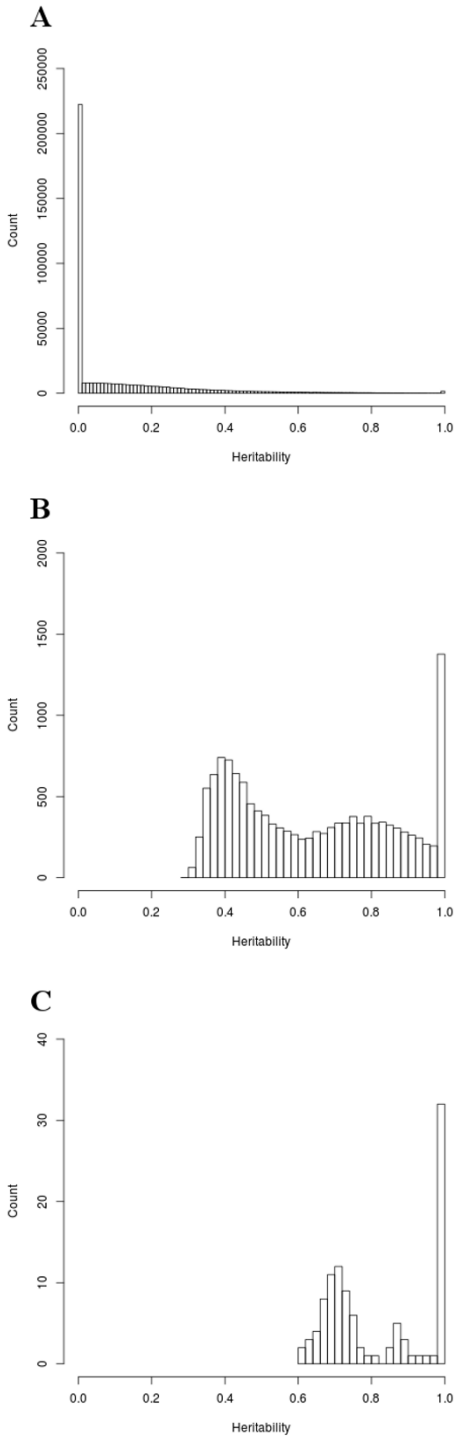


Figure 3.2 Distribution of genetic heritability and common environmental effect of FDR-corrected significant probes (n=105)

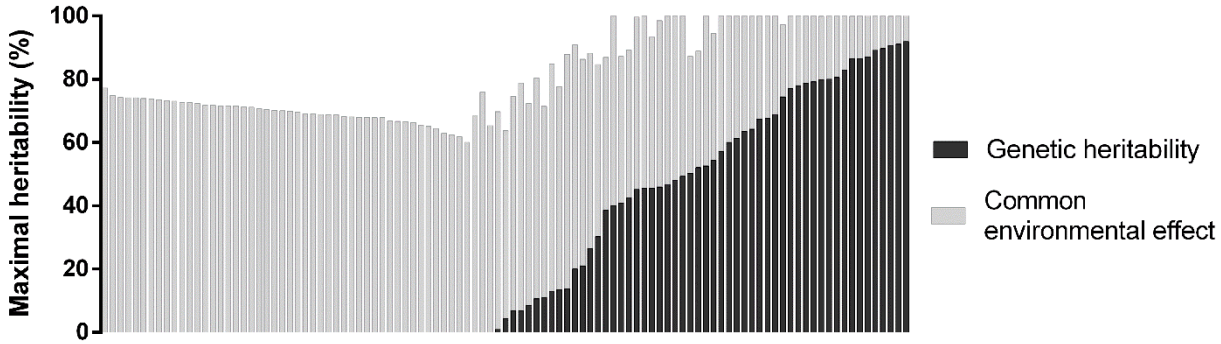
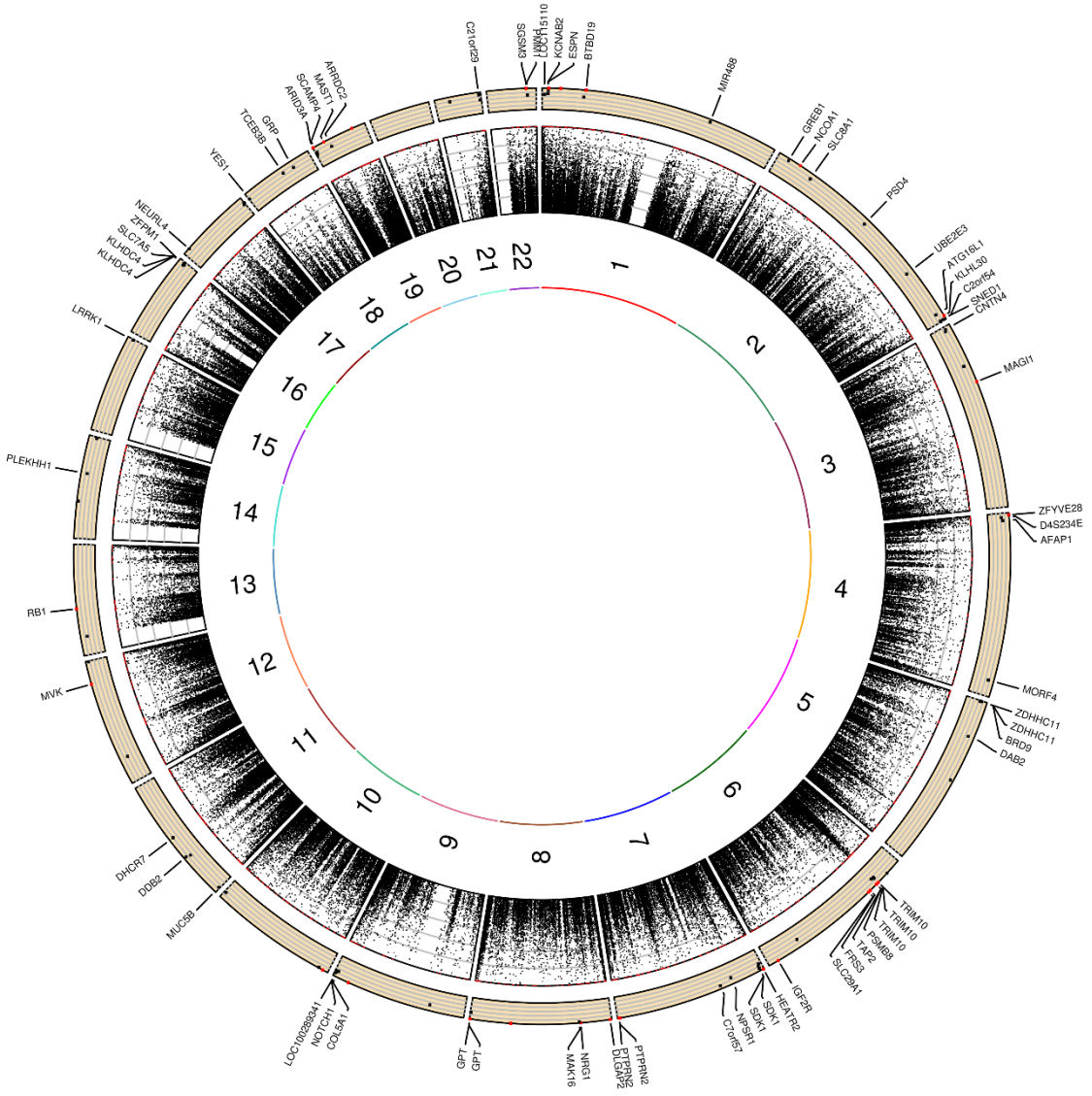


Figure 3.3 Circos plot depicting the distribution of maximal heritability estimates across the genome



Legend: Moving from inner to outer circles, first circle represents chromosomes. Maximal heritability of all 472 494 probes has been represented in second circle as scatter plot (values ranging from 0 to 100%). Third circle represents maximal heritability of the 105 FDR-corrected significant probes as scatter plot with genes name in which probes are located. Note: only 4 of the 5 genes (*ARID3A*, *ARRDC2*, *MAST1*, *SCAMP4*, and *ZNF235*) on chromosome 19 are shown due to place limitation.

Supplemental material

Supplemental material is available at https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5221602/

Chapitre 4 Ressemblances familiales des niveaux d'expression génique du sang

Bénédicte L. Tremblay, Frédéric Guénard, Benoît Lamarche, Louis Pérusse, Marie-Claude Vohl

L'article présenté dans ce chapitre s'intitule : *Familial resemblances in human whole blood transcriptome.*

Cet article est publié dans la revue : *BMC Genomics. 2018 Apr 27;19(1):300.*

Résumé

Introduction : Considérant le rôle de l'expression génique dans la prédisposition aux maladies chroniques et l'agrégation familiale des maladies chroniques, l'étude des ressemblances familiales des niveaux d'expression génique est très pertinente. Peu d'études ont évalué la contribution des effets génétiques et de l'environnement partagé dans les ressemblances familiales des niveaux d'expression génique du sang. L'objectif est de quantifier la contribution des effets génétiques et de l'environnement partagé dans les ressemblances familiales des niveaux d'expression génique à l'échelle du génome dans le sang. Nous réalisons également des comparaisons avec les ressemblances familiales des niveaux de méthylation de l'ADN à l'échelle du génome des leucocytes sanguins dans la même cohorte afin d'étudier plus en détail les mécanismes biologiques.

Résultats : L'héritabilité maximale, l'héritabilité génétique et l'effet de l'environnement partagé des niveaux d'expression génique ont été calculés pour toutes les sondes (20,6 %, 15,6 % et 5,0 %, respectivement) et pour les sondes présentant un effet familial significatif (78,1 %, 60,1 % et 18,0 %, respectivement). Des corrélations phénotypiques entre les niveaux d'expression génique et de méthylation de l'ADN, ajustés pour l'hétérogénéité des cellules sanguines, ont été calculées pour les sondes avec un effet familial significatif. Un total de 78 paires de sondes parmi les 7 618 401 paires de sondes possibles étaient significatives après la correction de Bonferroni (valeur de *P* corrigée = $6,56 \times 10^{-9}$). Des corrélations génétiques significatives entre les niveaux d'expression et de méthylation ont été identifiées pour 25 paires de sondes (corrélation génétique absolue de 0,97).

Conclusions : Les ressemblances familiales des niveaux d'expression génique étaient principalement attribuables à des facteurs génétiques, mais l'effet de l'environnement partagé a également joué un rôle en particulier pour les sondes démontrant un effet familial significatif. Les sondes et les sites CpG avec un effet familial semblent être sous un fort contrôle génétique partagé.

Abstract

Introduction: Considering the implication of gene expression in the susceptibility of chronic diseases and the familial clustering of chronic diseases, the study of familial resemblances in gene expression levels is then highly relevant. Few studies have considered the contribution of both genetic and common environmental effects to familial resemblances in whole blood gene expression levels. The objective is to quantify the contribution of genetic and common environmental effects in the familial resemblances of whole blood genome-wide gene expression levels. We also make comparisons with familial resemblances in blood leukocytes genome-wide DNA methylation levels in the same cohort in order to further investigate biological mechanisms.

Results: Maximal heritability, genetic heritability, and common environmental effect were computed for all probes (20.6%, 15.6%, and 5.0% respectively) and for probes showing a significant familial effect (78.1%, 60.1%, and 18.0% respectively). Pairwise phenotypic correlations between gene expression and DNA methylation levels adjusted for blood cell heterogeneity were computed for probes showing significant familial effect. A total of 78 probe pairs among the 7 618 401 possible pairs passed Bonferroni correction (corrected *P*-value= 6.56×10^{-9}). Significant genetic correlations between gene expression and DNA methylation levels were found for 25 probe pairs (absolute genetic correlation of 0.97).

Conclusions: Familial resemblances in gene expression levels were mainly attributable to genetic factors, but common environmental effect also played a role especially in probes showing a significant familial effect. Probes and CpG sites with familial effect seem to be under a strong shared genetic control.

Title page

Familial resemblances in human whole blood transcriptome

Bénédicte L. Tremblay[1], Frédéric Guénard[1], Benoît Lamarche[1], Louis Pérusse[2], Marie-Claude Vohl[1, 2*]

1. Institute of Nutrition and Functional Foods (INAF), Laval University, 2440 Hochelaga Blvd, Quebec City, QC, G1V 0A6, Canada

2. CHU de Québec Research Center – Endocrinology and Nephrology, 2705 Laurier Blvd, Quebec City, QC, G1V 4G2, Canada

Corresponding author*:
Marie-Claude Vohl, Ph.D.
Institute of Nutrition and Functional Foods (INAF), Laval University
Pavillon des Services**,** 2440 Hochelaga Blvd, Quebec City, QC G1V 0A6, Canada
Phone: 418 656-2131 ext. 4676; Fax: 418 656-5877
E-mail: marie-claude.vohl@fsaa.ulaval.ca

Keywords: Gene expression; Familial resemblances; DNA methylation; Genetic correlations; Metabolic pathways.

Introduction

Gene expression is increasingly studied in the context of chronic diseases due to the increasing evidence of its implication in disease susceptibility (1-3). The nature of genetic variance for chronic diseases may result from hereditary changes impacting gene regulation rather than protein structure and function (3). The study of the inheritance patterns and familial resemblances of gene expression is then highly relevant to better understand the origin of these diseases. Studies, including The Brisbane Systems Genetics Study (GSGS) have investigated the heritability of gene expression in whole blood (4-8). However, few studies have considered the contribution of both genetic and common environmental effects to familial resemblances in gene expression levels (4, 6-8).

The relationship between gene expression and DNA methylation is well established (9). Several studies have assessed the role of genetic factors on gene expression and DNA methylation (10-12). Previous studies have also reported genetic heritability of DNA methylation levels ranging from 8 to 18% in whole blood (13-15). A study by our group in French Canadians showed that familial resemblances in blood leukocytes genome-wide DNA methylation levels are mainly attributable to genetic factors, but that common environmental effect also plays a role (16). A study by Shakhbazov et al. also demonstrated a strong, shared genetic control for DNA methylation and gene expression (17).

Our objective is thus to quantify the contribution of genetic and common environmental effects in familial resemblances of gene expression using whole blood genome-wide expression levels in a family-based sample of 48 French Canadians from 16 families. To further investigate biological mechanisms, we also compared results on gene expression with familial resemblances in blood leukocytes genome-wide DNA methylation levels in the same cohort.

Methods

Patients and design

A total of 48 subjects from 16 families were recruited in the Greater Quebec City metropolitan area, in Canada, as part of the GENERATION Study. Families were composed of 16 mothers, 6 fathers, and 26 children. To be eligible parents had to be the biological parents of their child (or children). Parents had to be in good general health, non-smokers, with body mass index ranging between 18 and 35 kg/m^2, and free of any metabolic conditions requiring treatment although the use of Synthroid® (levothyroxine) or oral contraceptive was tolerated. Families living under the same roof comprised at least the mother and one child aged between 8 and 18. Children also had to be non-smokers and in good general health. They were not eligible if using psycho-stimulators [Ritalin® (methylphenidate), Concerta® (methylphenidate), and Strattera® (atomoxetine)]. Parents and children were asked to complete several dietary, physical activity, medical history, and pregnancy questionnaires under the supervision of a registered dietitian during their visit at the Institute of Nutrition and Functional Foods (INAF). The experimental protocol was approved by the Ethics Committees of Laval University Hospital Research Center and Laval University. All participants (adults and children) signed an informed consent document. Parental consent was also obtained by signing the child consent document.

Anthropometric and metabolic measurements

Body weight, waist girth, and height were measured according to the procedures recommended by the Airlie Conference (18). Blood samples were collected from an antecubital vein into PAXgeneTM tubes after 12-hour overnight fast and 48-hour alcohol abstinence.

RNA extraction and gene expression analysis

Total RNA was isolated and purified from whole blood using PAXgene Blood RNA Kit (QIAGEN). Quantification and verification of the purified RNA was assessed using both the NanoDrop (Thermo Scientific, Wilmington, DE, USA) and the 2100 Bioanalyzer (Agilent Technologies, Cedar Creek, TX, USA). The HumanHT-12 v4 Expression BeadChip (Illumina Inc., San Diego, CA) was used to measure expression levels of ~ 47 000 probes (> 31 000 annotated genes). This was performed at the McGill University and Genome Quebec Innovation Centre (Montreal, Canada). The FlexArray software (version 1.6) (19) and the lumi R package were used to analyze and normalize gene expression levels. More specifically, robust multi-array analysis, variance stabilizing transformation, and quantile normalization were used for background correction, variance stabilization, and normalization, respectively. Probes with a detection *P*-value ≤ 0.05 in at least 25% of all subjects were considered in analysis. A total of 18 160 probes among the 47 323 probes on the microarray (38.4%) showed significant gene expression in blood and were used for heritability analysis. This result is similar to the one reported in the BSGS (17 926 probes with a detection *P*-value ≤ 0.05 in 10% of all 862 subjects) (3). Details about DNA extraction and genome-wide DNA methylation analysis of the autosomal 472 494 CpG sites detected

in more than 90% of all subjects (*P*-value > 0.01) and mapped to unique location are presented in a previous paper on the same cohort (16).

Adjustment for blood cell heterogeneity

In methylation analysis, the cell proportions were predicted with the Jaffe et al. algorithm for all 48 subjects (20). This algorithm has been chosen because it was adjusted for the Illumina HumanMethylation450k (20). We obtained estimated cell counts for six different cell types: CD8+ T cells, CD4+ T cells, NK cells, B-cells, monocytes, and granulocytes. In expression analysis, the cell proportions were predicted with the Abbas et al. algorithm for all subjects (21). This algorithm has been chosen because it was developed specifically for blood (21). We obtained estimated cell counts for 17 different cell types (resting helper T cells, activated helper T cells, resting cytotoxic T cells, activated cytotoxic T cells, resting B cells, activated B cells, BCR-ligated B cells, IgM memory B cells, IgA/IgG memory B cells, plasma cells, resting NK cells, activated NK cells, monocytes, activated monocytes, resting dendritic cells, activated dendritic cells, and neutrophils) (21) that have been grouped in three classes: lymphocytes, neutrophils, and monocytes.

Statistical analysis

R software v2.14.1 (R Foundation for Statistical Computing; http://www.r-project.org) (22) was used to compute the mean absolute correlation of raw gene expression levels between relative pairs across all 18 160 probes. For heritability analysis of gene expression and DNA methylation levels, corrections were made for the effects of microarray, position on microarray, sex, age, age^2, sex*age, sex*age^2, and blood cell counts (estimated cell counts for six cell types in methylation analyses and three cell types in gene expression analyses), using a standard least squares model in JMP software v12. We used residuals from this model to compute heritability estimates using the variance component method implemented in QTDT v2.6.1 (23). We used full general models in which the variance in gene expression levels of each probe and DNA methylation levels of each CpG site were partitioned into polygenic effects (Vg), common environmental effects shared by family members (Vc), and non-shared environmental effects unique to each individual (Ve). We tested this full general model against a null model of no familial resemblance in which Vg = Vc = 0. We then computed mean maximal heritability, as the proportion of variance accounted by genetic and common environmental effects (Vg+Vc/Vg+Vc+Ve), mean genetic heritability, as the proportion of variance accounted by genetics effects (Vg/Vg+Vc+Ve), and the mean common environmental effect as the proportion of variance accounted by common environmental effects (Vc/Vg+Vc+Ve). Additionally, we computed heritability estimates in probes and CpG sites showing a significant familial effect (Vg and Vc significantly different from zero ($P \leq 0.05$)). We computed FDR-corrected *P*-values to account for multiple testing. For comparison purposes, we computed alternative genetic models in which the variance in gene expression and DNA methylation levels was partitioned into Vg and Ve. We then computed mean genetic heritability, as the proportion of variance accounted by genetic effects (Vg/Vg+Ve).

Phenotypic Pearson correlations between DNA methylation and gene expression levels were computed using R software v2.14.1 (22). Assuming bivariate normal distribution, asymptomatic *P*-values were computed based on Fisher Z transformation as in Shakhbazov et al. for comparison purposes (17). Genetic correlations between DNA methylation and gene expression were calculated using a bivariate analysis in SOLAR Eclipse version 7.6.4. Bonferroni corrections were used to account for multiple testing in phenotypic and genetic correlations. We used RCircos package and a modified version of the qqman package in R software to make the Circos plot and the chromosomic representations. (24). Finally, we used the IPA system (Ingenuity® Systems, www.ingenuity.com) to analyze pathways overrepresented among genes of significant probes (n=1211).

Results

Correlations of gene expression levels between relative pairs

Mean absolute correlations across all 18 160 detected probes were calculated between pairs from normalized gene expression levels (Table 1). The mean correlation coefficients were 0.31 ± 0.21 among siblings (n=13), 0.24 ± 0.17 for mother-offspring pairs (n=26), 0.23 ± 0.16 for parent-offspring pairs (n=37), and 0.021 ± 0.015 among unrelated individuals (n=1078). Related individuals (siblings, mother-offspring, and parent-offspring pairs) had more similar genome-wide gene expression levels compared to unrelated individuals according to these correlations.

Gene expression heritability analyses

Table 2 shows estimates of maximal heritability, genetic heritability, and common environmental effect from the full general model. When considering gene expression levels adjusted for blood cell heterogeneity of all probes (n=18 160), we obtained a mean maximal heritability of 20.6%, a genetic heritability of 15.6%, and a common environmental effect of 5.0% (Table 2). We have previously quantified contribution of genetic and common environmental effects in familial resemblances in genome-wide DNA methylation levels, but analyses did not account for blood cell composition (16). The analyses were reconducted taking into account the cell composition. We observed slightly lower values when computing maximal heritability, genetic heritability, and common environmental effect for all 472 494 cytosine-phosphate-guanine (CpG) sites (8.1%, 5.5%, and 2.6% respectively). DNA methylation heritability estimates adjusted for blood cell composition were similar but slightly lower than the ones previously reported, with maximal heritability of 12.7%, genetic heritability of 8.2%, and common environmental effect of 4.5% (16). The distribution of maximal heritability estimates of gene expression levels of all probes (n=18 160) with heritability ranging from 0 to 100% on the x-axis and the count (number of probes) on the y-axis is presented in Figure 1A. A total of 7182 probes (39.5% of all probes) and 318 probes (1.8% of all probes) had an estimated maximal heritability of 0% and 100%, respectively.

Moreover, when considering the 1211 probes showing a significant ($P \leq 0.05$) familial effect, a mean maximal heritability of 78.1%, a genetic heritability of 60.1%, and a common environmental effect of 18.0% were observed (Table 2). We observed similar estimates in DNA methylation analysis when computing maximal heritability, genetic heritability, and common environmental effect for the 6291 CpG sites showing a significant familial effect (66.4%, 39.6%, and 26.8% respectively). Once again, methylation heritability estimates adjusted for blood cell heterogeneity were very similar to the ones previously reported, with maximal heritability of 63.9%, genetic heritability of 39.3%, and common environmental effect of 24.6% (16). When maximal heritability estimates of significant probes are plotted, the distribution shifts to the right, in accordance with the higher maximal heritability estimates ranging from 32.8 to 100% (Figure 1B). A total of 255 probes (21.1% of all significant probes) gave an estimated maximal heritability of 100% (genetic heritability ranges from 0 to 100% and common

environmental effect ranges from 0 to 72.1%). Distributions of genetic heritability and common environmental effect for all probes (n=18 160) and significant probes (n=1211) are illustrated in Supplemental figures 1 and 2. Chromosomal representation of probes (n=1211) and CpG sites (n=6291) showing significant familial effect on gene expression and DNA methylation levels respectively are presented in Figure 2. Heritable probes and CpG sites are distributed across the genome. The visible peak on chromosome 6 corresponds to the Major Histocompatibility Complex (MHC) (25). Figure 3 depicted the number of significant CpG sites overlapping significant transcript regions (± 2kb and ± 5kb). CpG sites and probes showing significant familial effect were not very closely located.

Pairwise phenotypic correlations were computed for gene expression levels of the 1211 probes and DNA methylation levels of the 6291 CpG sites showing significant familial effect. Gene expression and DNA methylation levels adjusted for blood cell heterogeneity were significantly correlated in 78 probe pairs among the 7 618 401 possible pairwise tests after Bonferroni correction (corrected P= 0.05/ (1211 x 6291) = 6.56 x 10^{-9}) (Supplemental table 1). Further, genetic correlations between gene expression and DNA methylation levels for all 78 probe pairs passing Bonferroni threshold were estimated and genetic correlations were obtained for 39 probe pairs. Genetic correlations estimated the additive genetic effect that is shared between gene expression and DNA methylation levels. A total of 25 probe pairs had a significant genetic correlation ($P \leq 0.05$) and the mean absolute value of genetic correlation was 0.97 (Supplemental table 2). Only two probe pairs had a genetic correlation passing Bonferroni correction (corrected P=0.05/ 39 = 0.00128) (Table 3). The first probe pair was composed of cg22561794 located within the gene *BTNL8* on chromosome 5 and transcript ID_3520128 (NM_016437) located on chromosome 17. The second probe pair comprised cg02797144 located in an intergenic region on chromosome 16 and transcript ID_3990435 (BX282075) on chromosome 7.

Lastly, heritability analysis was repeated for probes showing a significant familial effect after False Discovery Rate (FDR) correction ($P \leq 0.05$). We obtained 12 FDR-corrected significant probes, assigned to 12 genes (*SPAG7, TMEM141, ZCCHC11, NDUFA2, CBL, HNRNPM, OXT, POLE4, MGC4677, DYNLT1, SOS1*, and *TBCA*), giving a mean maximal heritability of 94.6% (range 72.1 to 100%), a genetic heritability of 50.9% (range 0 to 74.1%), and a common environmental effect of 43.8% (ranges 25.9 to 72.1%) (Table 4). Six probes (*SPAG7, NDUFA2, CBL, HNRNPM, POLE4, MGC4677*) had an estimated maximal heritability of 100%. No CpG site in DNA methylation analysis passed the Bonferroni correction. Without the adjustment for blood cell heterogeneity, a total of 105 FDR-corrected significant CpG sites showed a mean maximal heritability of 81.8%, a genetic heritability of 26.9%, and a common environmental effect of 54.9% (16).

Moreover, Figure 4 shows Circos plot depicting the distribution of maximal heritability estimates of gene expression across the genome. These 12 FDR-corrected significant probes were assigned to 12 genes as

illustrated in Figure 3. There is also a visible higher maximal heritability in the MHC region on chromosome 6. A total of 14 probes out of 1211 were located in the MHC region, and were associated with the *HLA-F, HLA-H, HLA-A, HCG2P7, PPP1R11, HCP5, LST1, BAT2, C6ORF25, C6ORF27, LOC401252, HLA-DRB1, PSMB9, TAP2* genes.

Pathways analyses

Ingenuity Pathway Analysis (IPA) revealed that 140 pathways were significantly ($P \leq 0.05$) overrepresented among genes of the 1211 probes with a significant familial effect (Supplemental table 3). These pathways were related to inflammatory and immune response (n=52), cell cycle regulation (n=35), cancer (n=17), DNA and RNA regulation (n=11), intracellular and second messenger signaling (n=10), cardiovascular signaling (n=7), disease specific pathways (n=5), and nuclear receptor signaling (n=3). Interestingly, 22 out of the 140 pathways were in common with the 75 significantly overrepresented pathways among genes of the 6291 CpG sites with significant familial effect (Supplemental table 4). The detailed list of those 22 common pathways is presented in Supplemental table 5.

Discussion

The first aim of this study was to quantify the contribution of genetic and common environmental effects in familial resemblances of gene expression levels in a family-based sample of 48 French Canadians from 16 families. Mean absolute correlations between relative pairs suggest an underlying genetic similarity. Similar absolute correlations for DNA methylation levels were observed in the same cohort (16). In accordance with other studies, genetic factors seem to be the major determinant of familial resemblances in gene expression levels (4, 5, 17, 26). Indeed, for all probes (n=18 160), significant probes (n=1211), and FDR-corrected significant probes (n=12), familial effect was mainly due to genetic heritability. FDR-corrected significant probes were assigned to 12 genes (*SPAG7, TMEM141, ZCCHC11, NDUFA2, CBL, HNRNPM, OXT, POLE4, MGC4677, DYNLT1, SOS1, and TBCA*).

Using an alternative genetic model for comparison purposes, we obtained a genetic heritability estimate for all probes of 22.8%, which is higher than estimate reported in Huan et al. (5). Using the same alternative model, a total of 7682 probes (42.3% of all probes) had a genetic heritability > 0 compared to 14 753 probes (82% of all probes) in Powell et al. (4). In the same manner, we obtained 2097 probes with a genetic heritability > 0 ($P \leq 0.05$) compared to 7161 genes in Huan et al. (5). Overall, we reported higher genetic heritability estimates but lower number of probes with genetic heritability > 0. This may be explained by the small sample size of the present study that limits the statistical power to detect low heritability estimates. In fact, in a sample of 648 twins, Grundberg et al. argued that this number was not sufficient to obtain reliable heritability estimates of less than 10% (8, 27). Thus, the small sample size upon which the present study is based on may affect the reliability of low values of heritability estimates and partly explain the higher overall genetic heritability estimate and lower number of probes with genetic heritability > 0 compared to other studies.

Moreover, a total of 6555 probes had a common environmental effect > 0 in the full general model compared to 3373 probes in Powell et al. (4). These results should however be compared with caution considering that Powell et al. used a different general model which includes additive, non-additive and common environmental effects and did not account for blood cell heterogeneity (4). Grundberg et al. also reported a common environmental component in 32% of expressed lymphoblastoid cell line transcripts (8). Interestingly, Nédélec et al. demonstrated that expression of several genes in macrophages is different between subjects of European and African ancestry (28). All subjects in the present study are French Canadian Caucasians of French-Canadian descent living in the same city (Quebec City). Thus, we can hypothesize that gene expression levels are more homogeneous in a founder population of European descent (29).

The secondary aim was to compare results of the present study with results on familial resemblances in genome-wide DNA methylation levels using the same cohort in order to further investigate biological mechanisms. We

first reported that familial resemblances in genome-wide gene expression and DNA methylation levels are mainly due to genetic effects with a contribution of common environmental effect. CpG sites and probes with a strong familial effect were distributed across the genome and were not very closely located (± 2 and 5 kb). This is in accordance with a study by Van Eijk KR et al. that found that expression and methylation modules (clusters of interconnected genes) exhibit relatively few overlapping genes, although some of the overlaps were statistically significant (12). Grundberg et al. also showed that over 60% of gene expression heritability is *trans* to the structural gene (8). Shakhbazov et al. demonstrated that gene expression and DNA methylation probe pairs with shared QTL(s) have larger genetic correlations in contrast with the same chromosome probe pairs without shared QTL (17). A study by Price et al. using 722 Icelanders from family cohorts demonstrated that the proportion of gene expression heritability attributable to *cis* regulation was 37% in blood (30). The proportion of heritability of gene expression attributable to *cis* regulation is also expected to increase as a function of the number of different cell types (30). We could therefore hypothesise that the adjustment for blood cell composition attenuated the proportion of heritability attributable to *cis* regulation.

Regarding the 78 significant phenotypic correlations between DNA methylation and gene expression levels, only three probe pairs were located on the same chromosome. This suggests that *cis* regulation of single nucleotide polymorphisms (SNPs) may not be responsible for genetic heritability of probes and phenotypic correlations with methylation levels of CpG sites. We reported 25 significant genetic correlations between gene expression and DNA methylation levels adjusted for blood cell heterogeneity, thus suggesting a shared genetic control. We reported higher genetic correlation (-0.97/0.97, for negative and positive genetic correlations, respectively) than Shakhbazov et al. (-0.69/0.68) (17). This could be explained by the fact that they calculated correlations between gene expression and DNA methylation of probe pairs across the genome ($5x10^9$), while we restrained analyses to probes and CpG sites with a significant familial effect ($7x10^6$). Shakhbazov et al. also demonstrated that correction for cell heterogeneity greatly impacts correlations between genome-wide DNA methylation and gene expression levels with a 300 times reduction in number of probe pairs passing Bonferroni correction. Accordingly, we also observed a 48 times reduction (from 1211 to 25 significant genetic correlations) after the correction for blood cell composition in our subset of significant probes and CpG sites. A total of two out of 25 probe pair correlations remained significant after Bonferroni correction. The first probe pair comprised cg22561794 on the *BTNL8* gene encoding for butyrophilin like 8 and NM_016437 on *TUBG2* gene encoding for the tubulin gamma 2. *BTNL8* gene is involved in immune response as it stimulates cytokine production and is also altered in intestinal inflammation and colon cancer (31, 32). *TUBG2* gene is primarily detected in the brain and its expression seems to be closely related to oncogenesis (33, 34). The second probe pair comprised cg02797144 located in an intergenic region on chromosome 16 and BX282075 expression probe measuring *HS.511718* on chromosome 7.

We also observed similarities in overrepresented pathways of significant probes and CpG sites. Indeed, 22 pathways were in common between overrepresented pathways (n= 140) of significant probes and significant CpG sites (n=75) (16). Thus, some CpG sites and probes with a familial effect seem to be implicated in the same metabolic pathways. The majority of these 22 common pathways were related to inflammatory and immune response, which is in line with the 14 highly heritable probes (mean maximal heritability of 82.0%) found in the MHC region on chromosome 6 (25). Highly heritable CpG sites in the MHC region were also reported by us (16) and others (13). This result is also in line with the reported ancestry-associated differences in the gene regulatory response to infection (28).

This study has its own strengths and limitations. The main strength is the comparison of results on familial resemblances in gene expression with results on familial resemblances in DNA methylation levels in the same cohort. The use of a cell type predictor both in methylation and expression analysis allow to greatly attenuate the bias associated with blood cell heterogeneity, especially in phenotypic and genetic correlations. The calculation of genetic correlations adds important information about the additive genetic effect that is shared between gene expression and DNA methylation levels. To the best of our knowledge this is the first study on familial resemblances in genome-wide gene expression levels carried out on healthy French Canadians, a founder population of European descent. Regarding statistical analysis, a filter was applied to exclude lowly expressed genes and thus avoiding bias in estimates of gene expression heritability. Yang et al. reported that genes with very low expression levels have a reduced statistical power to be detected for a significant genetic component (35). Adjustments were also made to account for multiple testing.

Regarding limitations of the study, the main one is the small sample size which has been discussed earlier. The fact that we do not have genotypes for study subjects represents another limitation considering the impact of SNPs on gene expression levels. Indeed, an eQTL analysis using 862 subjects from the GSGS revealed 15 000 associations of SNPs with gene expression levels (3). We also used a cell type predictor in methylation and expression analysis to reduce the effect of blood cell heterogeneity. However, Shakhbazov et al. demonstrated that correction for predicted cell counts is not sufficient to remove the effect of blood cell heterogeneity on the correlations between DNA methylation and gene expression levels (17). Using the observed cell proportions remains the best way to remove this effect as much as possible (17). Moreover, certain methodological considerations in pathway analysis including the fact that annotation of genetic variants is inconsistent across databases, incomplete and biased toward known genes must be noted (36). The gold standard to establish validity of findings from pathway analysis remains the replication of results in independent studies.

Conclusion

In conclusion, familial resemblances in gene expression levels are mainly attributable to genetic factors but common environmental effect also plays a role especially in probes showing a significant familial effect. Heritability estimates of genome-wide gene expression are similar but higher than those of genome-wide DNA methylation. Finally, pairwise phenotypic correlations between gene expression and DNA methylation of probes with familial effect are mainly attributable to a shared genetic control as showed by high genetic correlations.

Ethics approval and consent to participate

All participants (adults and children) signed an informed consent document. Parental consent was also obtained by signing the child consent document. The experimental protocol was approved by the Ethics Committees of Laval University Hospital Research Center and Laval University.

Consent for publication

Not applicable

Availability of data and material

Expression datasets supporting the conclusion of this article are available in the Gene Expression Omnibus (GEO) database, www.ncbi.nlm.nih.gov/geo (GSE114620).

Competing interests

Authors have no potential conflicts of interest.

Funding

This work was supported by the Canada Research Chair in Genomics Applied to Nutrition and Metabolic Health. MCV is Tier 1 Canada Research Chair in Genomics Applied to Nutrition and Metabolic Health. BLT is a recipient of a scholarship from Fonds de recherche du Québec – Santé (FRQS).

Authors' contributions

Each author contribution to work: BL and MCV designed research; BLT, FG, and LP conducted research and performed statistical analyses; BLT wrote the paper; BLT and MCV have primary responsibility for final content. All authors read and approved the final manuscript.

Acknowledgments

We would like to thank Véronique Garneau and Christian Couture, who contributed to the success of this study. We also thank Catherine Raymond for the laboratory work.

References

1. Cookson W, Liang L, Abecasis G, Moffatt M, Lathrop M. Mapping complex disease traits with global gene expression. Nature reviews Genetics. 2009;10(3):184-94.

2. Goring HH, Curran JE, Johnson MP, Dyer TD, Charlesworth J, Cole SA, et al. Discovery of expression QTLs using large-scale transcriptional profiling in human lymphocytes. Nature genetics. 2007;39(10):1208-16.

3. Powell JE, Henders AK, McRae AF, Caracella A, Smith S, Wright MJ, et al. The Brisbane Systems Genetics Study: genetical genomics meets complex trait genetics. PloS one. 2012;7(4):e35430.

4. Powell JE, Henders AK, McRae AF, Kim J, Hemani G, Martin NG, et al. Congruence of additive and non-additive effects on gene expression estimated from pedigree and SNP data. PLoS genetics. 2013;9(5):e1003502.

5. Huan T, Liu C, Joehanes R, Zhang X, Chen BH, Johnson AD, et al. A systematic heritability analysis of the human whole blood transcriptome. Human genetics. 2015;134(3):343-58.

6. Bryois J, Buil A, Ferreira PG, Panousis NI, Brown AA, Vinuela A, et al. Time-dependent genetic effects on gene expression implicate aging processes. Genome research. 2017;27(4):545-52.

7. Wright FA, Sullivan PF, Brooks AI, Zou F, Sun W, Xia K, et al. Heritability and genomics of gene expression in peripheral blood. Nature genetics. 2014;46(5):430-7.

8. Grundberg E, Small KS, Hedman AK, Nica AC, Buil A, Keildson S, et al. Mapping cis- and trans-regulatory effects across multiple tissues in twins. Nature genetics. 2012;44(10):1084-9.

9. Jones PA. Functions of DNA methylation: islands, start sites, gene bodies and beyond. Nature reviews Genetics. 2012;13(7):484-92.

10. Bell JT, Pai AA, Pickrell JK, Gaffney DJ, Pique-Regi R, Degner JF, et al. DNA methylation patterns associate with genetic and gene expression variation in HapMap cell lines. Genome biology. 2011;12(1):R10.

11. Moen EL, Zhang X, Mu W, Delaney SM, Wing C, McQuade J, et al. Genome-wide variation of cytosine modifications between European and African populations and the implications for complex traits. Genetics. 2013;194(4):987-96.

12. van Eijk KR, de Jong S, Boks MP, Langeveld T, Colas F, Veldink JH, et al. Genetic analysis of DNA methylation and gene expression levels in whole blood of healthy human subjects. BMC genomics. 2012;13:636.

13. McRae AF, Powell JE, Henders AK, Bowdler L, Hemani G, Shah S, et al. Contribution of genetic variation to transgenerational inheritance of DNA methylation. Genome biology. 2014;15(5):R73.

14. Gordon L, Joo JE, Powell JE, Ollikainen M, Novakovic B, Li X, et al. Neonatal DNA methylation profile in human twins is specified by a complex interplay between intrauterine environmental and genetic factors, subject to tissue-specific influence. Genome research. 2012;22(8):1395-406.

15. Bell JT, Tsai PC, Yang TP, Pidsley R, Nisbet J, Glass D, et al. Epigenome-wide scans identify differentially methylated regions for age and age-related phenotypes in a healthy ageing population. PLoS genetics. 2012;8(4):e1002629.

16. Tremblay BL, Guenard F, Lamarche B, Perusse L, Vohl MC. Familial resemblances in blood leukocyte DNA methylation levels. Epigenetics. 2016;11(11):831-8.

17. Shakhbazov K, Powell JE, Hemani G, Henders AK, Martin NG, Visscher PM, et al. Shared genetic control of expression and methylation in peripheral blood. BMC genomics. 2016;17:278.

18. Callaway C, Chumlea W, Bouchard C, Himes J, Lohman T, Martin A, et al. Standardization of anthropomeric measurements : The Airlie (VA) Consensus Conference. Champaign, IR, USA; 1988.

19. Blazejczyk M MM, Nadon R. FlexArray: A statistical data analysis software for gene expression microarrays. Genome Quebec, Montreal, Canada. 2007.

20. Jaffe AE, Irizarry RA. Accounting for cellular heterogeneity is critical in epigenome-wide association studies. Genome biology. 2014;15(2):R31.

21. Abbas AR, Wolslegel K, Seshasayee D, Modrusan Z, Clark HF. Deconvolution of blood microarray data identifies cellular activation patterns in systemic lupus erythematosus. PloS one. 2009;4(7):e6098.

22. Team RC. R: A language and environment for statistical computing. R Foundation for Statistical Computing. Vienna, Austria. 2013.

23. Abecasis GR, Cardon LR, Cookson WO. A general test of association for quantitative traits in nuclear families. American journal of human genetics. 2000;66(1):279-92.

24. Zhang H, Meltzer P, Davis S. RCircos: an R package for Circos 2D track plots. BMC bioinformatics. 2013;14:244.

25. Trowsdale J, Knight JC. Major histocompatibility complex genomics and human disease. Annual review of genomics and human genetics. 2013;14:301-23.

26. Emilsson V, Thorleifsson G, Zhang B, Leonardson AS, Zink F, Zhu J, et al. Genetics of gene expression and its effect on disease. Nature. 2008;452(7186):423-8.

27. Grundberg E, Meduri E, Sandling JK, Hedman AK, Keildson S, Buil A, et al. Global analysis of DNA methylation variation in adipose tissue from twins reveals links to disease-associated variants in distal regulatory elements. American journal of human genetics. 2013;93(5):876-90.

28. Nedelec Y, Sanz J, Baharian G, Szpiech ZA, Pacis A, Dumaine A, et al. Genetic Ancestry and Natural Selection Drive Population Differences in Immune Responses to Pathogens. Cell. 2016;167(3):657-69 e21.

29. Roy-Gagnon MH, Moreau C, Bherer C, St-Onge P, Sinnett D, Laprise C, et al. Genomic and genealogical investigation of the French Canadian founder population structure. Human genetics. 2011;129(5):521-31.

30. Price AL, Helgason A, Thorleifsson G, McCarroll SA, Kong A, Stefansson K. Single-tissue and cross-tissue heritability of gene expression via identity-by-descent in related or unrelated individuals. PLoS genetics. 2011;7(2):e1001317.

31. Chapoval AI, Smithson G, Brunick L, Mesri M, Boldog FL, Andrew D, et al. BTNL8, a butyrophilin-like molecule that costimulates the primary immune response. Molecular immunology. 2013;56(4):819-28.

32. Lebrero-Fernandez C, Wenzel UA, Akeus P, Wang Y, Strid H, Simren M, et al. Altered expression of Butyrophilin (BTN) and BTN-like (BTNL) genes in intestinal inflammation and colon cancer. Immunity, inflammation and disease. 2016;4(2):191-200.

33. Draberova E, Sulimenko V, Vinopal S, Sulimenko T, Sladkova V, D'Agostino L, et al. Differential expression of human gamma-tubulin isotypes during neuronal development and oxidative stress points to a gamma-tubulin-2 prosurvival function. FASEB journal : official publication of the Federation of American Societies for Experimental Biology. 2017;31(5):1828-46.

34. Niu Y, Liu T, Tse GM, Sun B, Niu R, Li HM, et al. Increased expression of centrosomal alpha, gamma-tubulin in atypical ductal hyperplasia and carcinoma of the breast. Cancer science. 2009;100(4):580-7.

35. Yang S, Liu Y, Jiang N, Chen J, Leach L, Luo Z, et al. Genome-wide eQTLs and heritability for gene expression traits in unrelated individuals. BMC genomics. 2014;15:13.

36. Aslibekyan S, Almeida M, Tintle N. Pathway analysis approaches for rare and common variants: insights from Genetic Analysis Workshop 18. Genetic epidemiology. 2014;38 Suppl 1:S86-

Tables

Table 4.1 Mean absolute correlations across all probes of normalised gene expression levels between relative pairs

Relative pairs	n	Correlation ± SD
Siblings	13	0.31 ± 0.21
Mother-Offspring	26	0.24 ± 0.17
Parent-Offspring	37	0.23 ± 0.16
Unrelated	1078	0.021 ± 0.015

Abbreviation: Standard deviation (SD)

Table 4.2 Heritability estimates of gene expression levels

Type of heritability estimates (% ± SD)	All probes (n=18 160)	Significant probes (n=1211)	FDR-corrected significant probes (n=12)
Maximal heritability	20.62 ± 27.09	78.13 ± 20.70	94.63 ± 9.66
Genetic heritability	15.60 ± 26.49	60.11 ± 37.90	50.87 ± 23.77
Common env. effect	5.02 ± 10.07	18.01 ± 19.59	43.75 ± 14.50

Abbreviation: Standard deviation (SD)

Table 4.3 Top two probe pairs showing a significant genetic correlation

Transcript	CpG site	h2 expression ± SE	h2 methylation ± SE	Correlation coefficient	*P*-value
ID_3520128 (NM_016437, *TUBG2*, Chr17)	cg22561794 (*BTNL8*, Chr5)	0.63 ±0.22	0.71 ± 0.20	-0.96	0.00116
ID_3990435 (BX282075, *HS.511718*, Chr7)	cg02797144 (Intergenic, Chr16)	0.64 ± 0.22	0.65 ± 0.19	-1	0.00118

Abbreviations: Chromosome (Chr), Heritability (h2), Standard error (SE). All positions are from de Genome Build 37.

Table 4.4 Heritability estimates of false discovery rate-corrected significant probes

Transcript	Maximal heritability	Genetic heritability	Common env. effect	*P* (corrected *P*)*
ID_3440070 (NM_004890, *SPAG7*, Chr17)	100.0	57.86	42.14	3×10^{-6} (0.029)
ID_5130113 (NM_032928, *TMEM141*, Chr9)	83.64	26.04	57.60	4×10^{-6} (0.029)
ID_4260093 (NM_001009882, *ZCCHC11*, Chr1)	97.74	52.07	45.67	6×10^{-6} (0.029)
ID_6840189 (NM_002488, *NDUFA2*, Chr5)	100.0	74.11	25.89	8×10^{-6} (0.029)
ID_6960209 (NM_005188, *CBL*, Chr11)	100.0	66.75	33.25	8×10^{-6} (0.029)
ID_2360669 (NM_031203, *HNRNPM*, Chr19)	100.0	58.45	41.55	1×10^{-5} (0.030)
ID_1850372 (NM_000915, *OXT*, Chr20)	72.08	0	72.08	2×10^{-5} (0.036)
ID_2120286 (NM_019896, *POLE4*, Chr2)	100.0	67.63	32.37	2×10^{-5} (0.036)
ID_6350189 (NM_052871, *MGC4677*, Chr2)	100.0	64.83	35.17	2×10^{-5} (0.036)
ID_6350634 (NM_006519, *DYNLT1*, Chr6)	100.0	67.11	32.89	2×10^{-5} (0.036)
ID_2140519 (NM_005633, *SOS1*, Chr2)	82.07	15.11	66.96	3×10^{-5} (0.045)
ID_5960093 (NM_004607, *TBCA*, Chr5)	100.0	60.54	39.46	3×10^{-5} (0.045)

Abbreviations: Chromosome (Chr), Environmental (env), False discovery rate (FDR). *P*-value of familial effect (Vg and Vc significantly different from 0). * False discovery rate-corrected *P* values. All positions are from de Genome Build 37.

Figures

Figure 4.1 Distribution of maximal heritability estimates for gene expression levels of A) all probes (n=18 160), B) significant probes (n=1211)

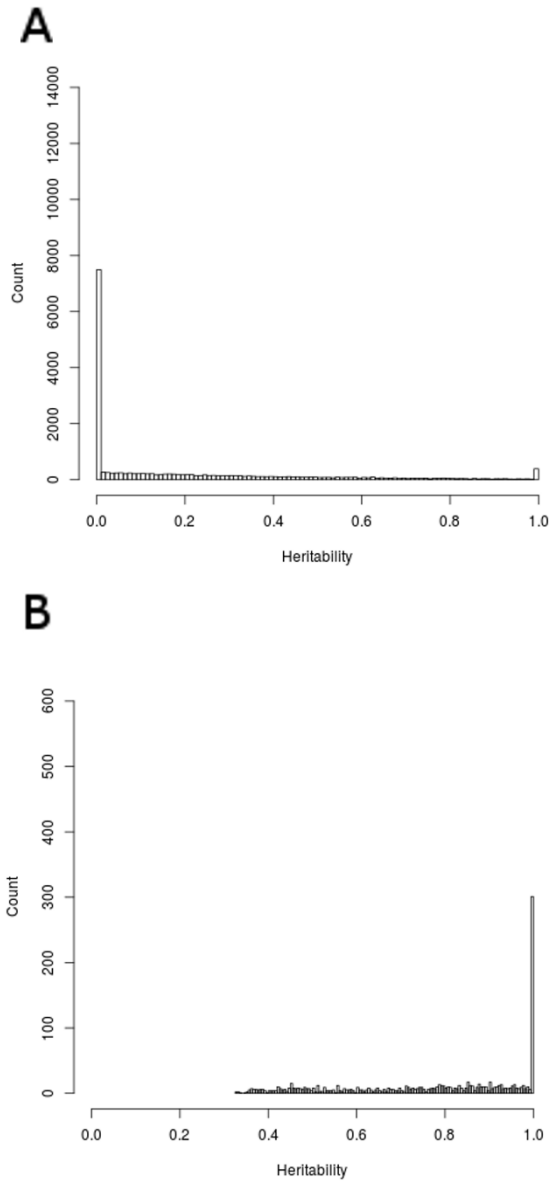
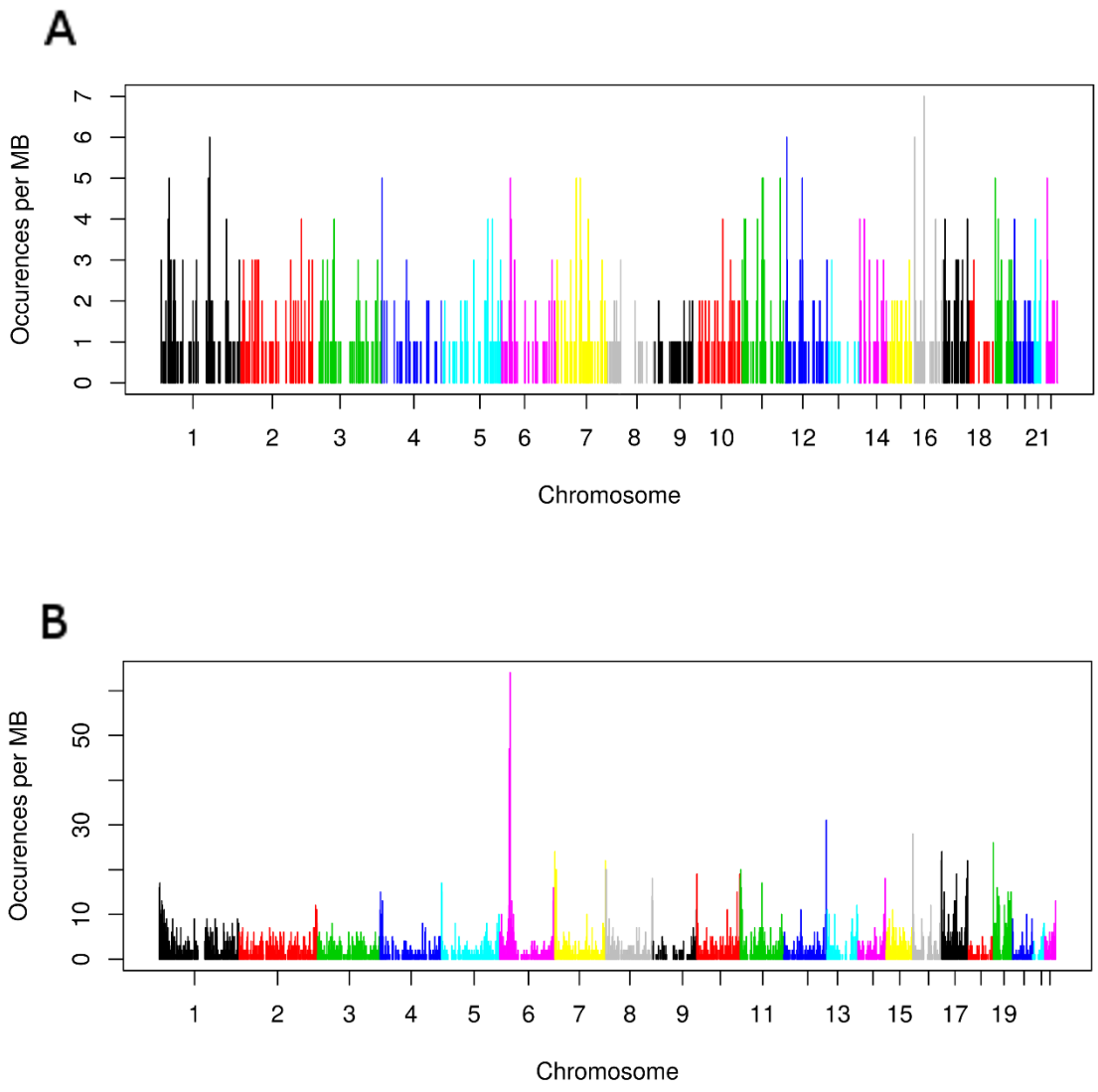
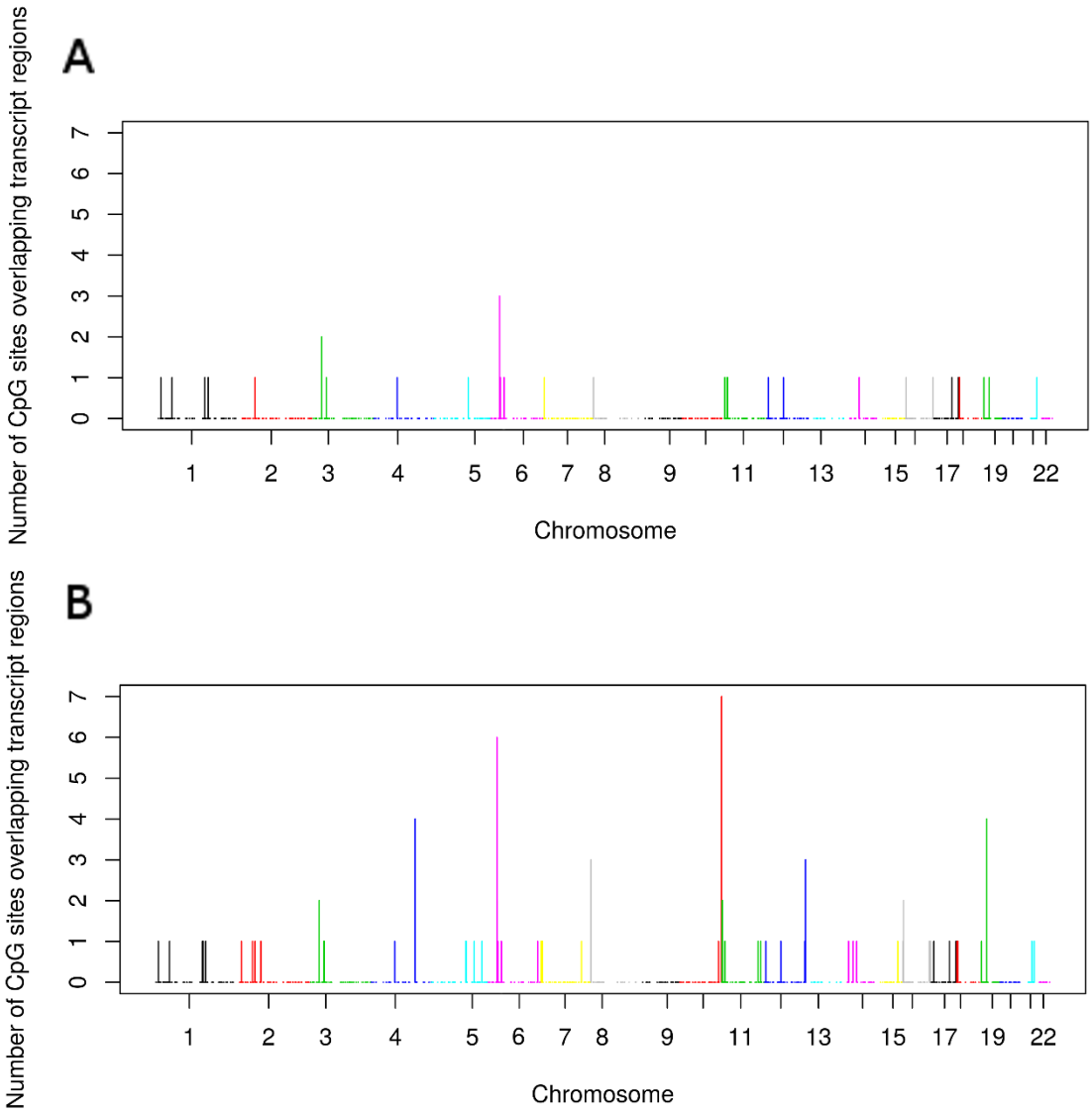


Figure 4.2 Chromosomal representation of annotated A) probes and B) CpG sites showing significant familial effect in gene expression and DNA methylation levels, respectively



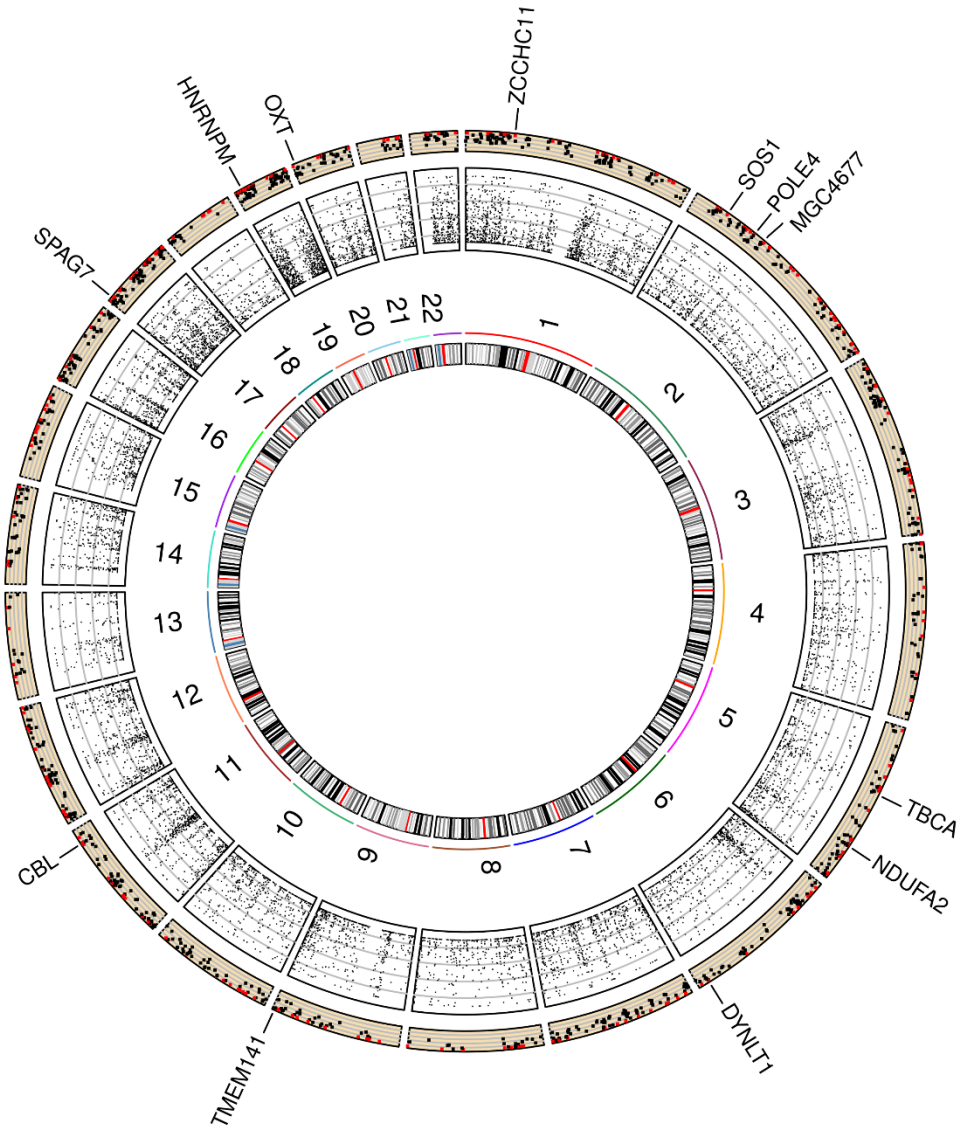
Legend: All positions are from the Genome Build 37.

Figure 4.3 Chromosomal representation of number of significant CpG sites overlapping significant transcript regions A) ± 2kb and B) ± 5kb



Legend: All positions are from de Genome Build 37.

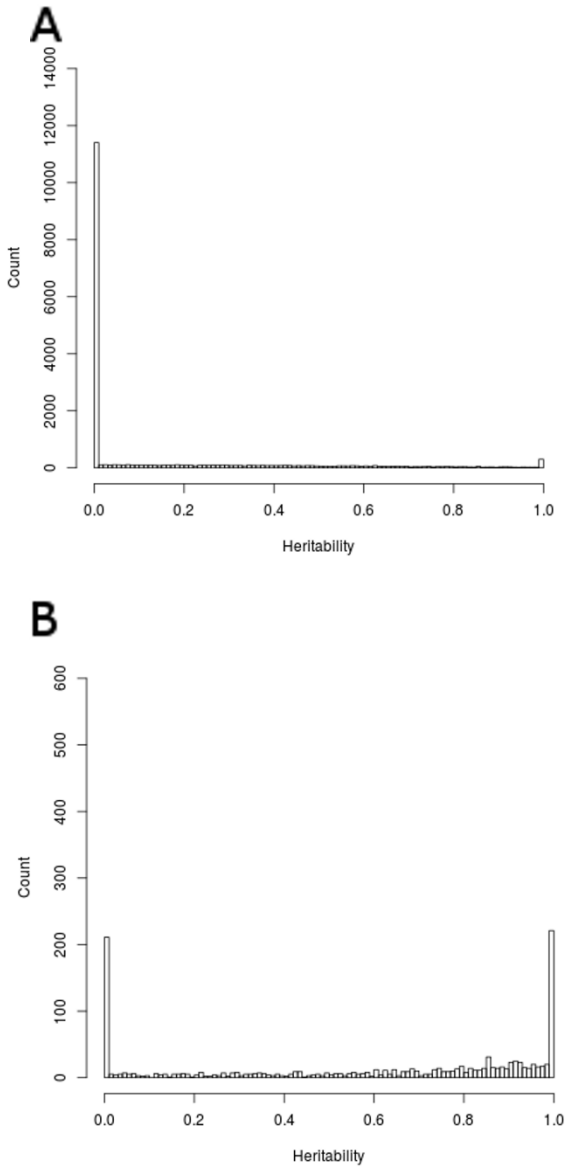
Figure 4.4 Circos plot depicting the distribution of maximal heritability estimates of gene expression across the genome



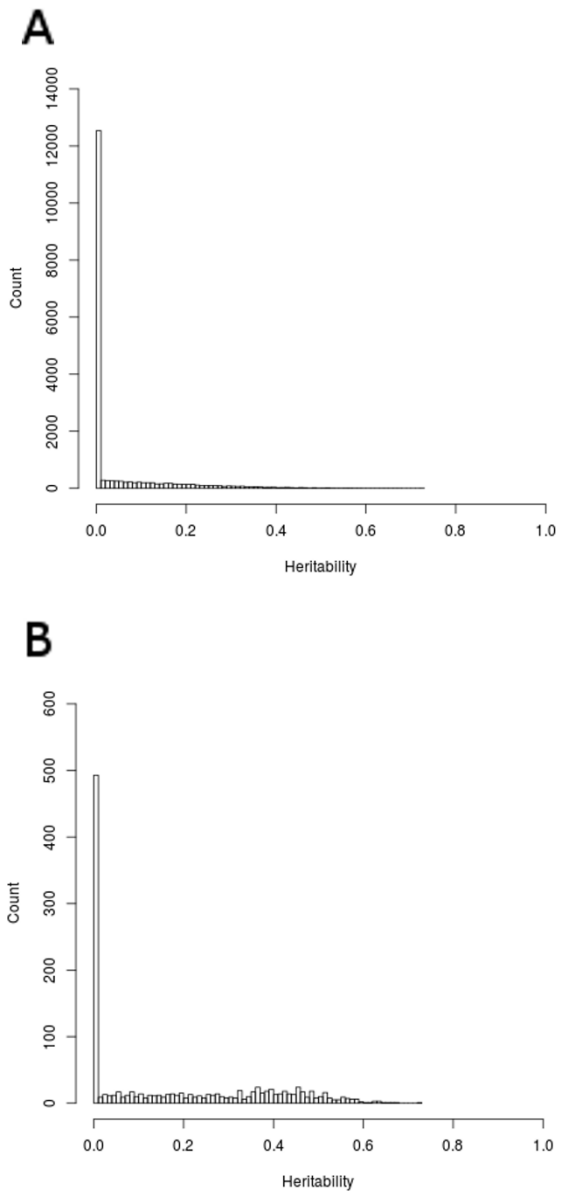
Legend: Moving from inner to outer circles, first circle represents chromosomes. Maximal heritability of all 18 160 probes has been represented in second circle as scatter plot (values ranging from 0 to 100%). Third circle represents maximal heritability of the 1211 probes showing a significant familial effect as scatter plot. Genes name of the 12 probes that passed FDR correction are also represented.

Supplemental material

Supplemental figure 4.1 Distribution of genetic heritability estimates for gene expression levels of A) all probes (n=18 160), B) significant probes (n=1211)



Supplemental figure 4.2 Distribution of common environmental effect estimates for gene expression levels of A) all probes (n=18 160), B) significant probes (n=1211)



Supplemental table 4.1 Significant phenotypic correlations between DNA methylation and gene expression levels (n=78)

Transcript	Gene	Chr	Position	CpG	Gene	Chr	Position	Correlation coefficient	*P*-value
ID_1780647	*ADCK2*	7	140041178-140041227	cg13692482	*SLC2A8*	9	130168651	-0,788306125	2.91×10^{-11}
ID_5910390	*SF3A1*	22	30728343-30728392	cg08146708		1	167393634	-0,760022938	3.73×10^{-10}
ID_3990435	*HS.511718*	7	141981986-141982035	cg22283959	*NR2F2*	15	96880782	0,759424981	3.92×10^{-10}
ID_5910390	*SF3A1*	22	30728343-30728392	cg12133451		1	227746453	-0,754439646	5.93×10^{-10}
ID_3990435	*HS.511718*	7	141981986-141982035	cg13220357	*SGSM1*	22	25202163	0,754175408	6.06×10^{-10}
ID_1710132	*MED4*	13	48650431-48650480	cg12505184	*LAP3*	4	17585801	-0,749089326	9.13×10^{-10}
ID_3400438	*HLA-A*	6	29913300-29913349	cg14095101	*TRPM6*	9	77502747	0,748942827	9.24×10^{-10}
ID_620376	*PPP1R11*	6	30037919-30037968	cg17140673	*ELOVL6*	4	111117400	-0,748831801	9.32×10^{-10}
ID_3990435	*HS.511718*	7	141981986-141982035	cg07088771	*TNXB*	6	32057846	-0,748422786	9.63×10^{-10}
ID_1710132	*MED4*	13	48650431-48650480	cg22309983	*TRPV1*	17	3497580	-0,748285012	9.73×10^{-10}
ID_2510523	*RBCK1*	20	359499-359548	cg07167185	*LYPLA2*	1	24120017	-0,746056672	1.16×10^{-09}
ID_3990435	*HS.511718*	7	141981986-141982035	cg05265020	*NRN1*	6	6002865	0,746037744	1.16×10^{-09}
ID_5360243	*ITGB5*	3	125965004-125965053	cg16289175	*AP2M1*	3	183898788	-0,745859168	1.18×10^{-09}
ID_1780647	*ADCK2*	7	140041178-140041227	cg25649188	*CASKIN2*	17	73499917	-0,745255878	1.24×10^{-09}
ID_3990435	*HS.511718*	7	141981986-141982035	cg20594316	*ZFYVE28*	4	2389156	-0,745125618	1.25×10^{-09}
ID_2510523	*RBCK1*	20	359499-359548	cg19033444	*PRR5*	22	45065754	-0,744997149	1.26×10^{-09}
ID_1780647	*ADCK2*	7	140041178-140041227	cg07490142	*FLJ39582*	22	21357269	0,74494403	1.27×10^{-09}
ID_4180176	*KLHDC4*	16	87741498-87741547	cg05564340	*U2AF1*	21	44518156	-0,744832715	1.28×10^{-09}
ID_1710132	*MED4*	13	48650431-48650480	cg21285564	*MIR1306*	22	20072475	-0,743903909	1.37×10^{-09}
ID_3990435	*HS.511718*	7	141981986-141982035	cg07191524		16	1077624	-0,74383539	1.38×10^{-09}
ID_5360243	*ITGB5*	3	125965004-125965053	cg12505184	*LAP3*	4	17585801	-0,743316501	1.44×10^{-09}
ID_3990435	*HS.511718*	7	141981986-141982035	cg04187403	*NRN1*	6	5999377	0,742837132	1.49×10^{-09}
ID_2340059	*C16ORF53*	16	29738678-29738727	cg17763566	*HLA-DPB2*	6	33083457	0,742715121	1.51×10^{-09}
ID_3990435	*HS.511718*	7	141981986-141982035	cg00122779	*TNXB*	6	32033006	-0,748015313	1.51×10^{-09}
ID_5360243	*ITGB5*	3	125965004-125965053	cg06251181	*CORO1C*	12	109046825	-0,742239663	1.56×10^{-09}
ID_3990435	*HS.511718*	7	141981986-141982035	cg14386951	*NRN1*	6	5999186	0,740578699	1.78×10^{-09}
ID_6420168	*DBNDD2*	20	43472577-43472626	cg26993038		14	104686639	0,739762945	1.89×10^{-09}

ID_1780647	*ADCK2*	7	140041178-140041227	cg15174906		6	156919703	0,73852253	2.08×10^{-09}
ID_3990435	*HS.511718*	7	141981986-141982035	cg02448934	*BTNL2*	6	32367970	-0,737912511	2.18×10^{-09}
ID_3520128	*TUBG2*	17	38072433-38072482	cg22561794	*BTNL8*	5	180335743	-0,743205616	2.19×10^{-09}
ID_3400438	*HLA-A*	6	29913300-29913349	cg21758140		2	161083993	0,737534638	2.24×10^{-09}
ID_3990435	*HS.511718*	7	141981986-141982035	cg00570469		19	36602113	-0,737374341	2.27×10^{-09}
ID_5360243	*ITGB5*	3	125965004-125965053	cg11063328	*GTF2H1*	11	18343203	0,737064234	2.32×10^{-09}
ID_7050470	*TAX1BP1*	7	27806189-27806234	cg24994593	*LDLRAD3*	11	35966604	0,736252393	2.47×10^{-09}
ID_5360243	*ITGB5*	3	125965004-125965053	cg22309983	*TRPV1*	17	3497580	-0,735113424	2.69×10^{-09}
ID_3420068	*HAT1*	2	172556438-172556487	cg24994593	*LDLRAD3*	11	35966604	0,735092272	2.69×10^{-09}
ID_3130220	*TMEM158*	3	45266007-45266056	cg20095851	*SCNN1A*	12	6486701	0,734263063	$2.86 \times 10^{--09}$
ID_3990435	*HS.511718*	7	141981986-141982035	cg09081941		12	132338464	-0,734131802	2.89×10^{-09}
ID_2450064	*NFE2L2*	2	177803791-177803840	cg08931968	*NCRNA00171*	6	30005372	-0,733595263	3.01×10^{-09}
ID_3990435	*HS.511718*	7	141981986-141982035	cg07944221		16	29272727	0,733084374	3.13×10^{-09}
ID_1710132	*MED4*	13	48650431-48650480	cg08931968	*NCRNA00171*	6	30005372	-0,732828331	3.19×10^{-09}
ID_3990435	*HS.511718*	7	141981986-141982035	cg05368730		12	132657879	0,732457774	3.27×10^{-09}
ID_6100762	*PSMC6*	14	52257383-52257432	cg24994593	*LDLRAD3*	11	35966604	0,732132728	3.35×10^{-09}
ID_3520128	*TUBG2*	17	38072433-38072482	cg03118417	*RANBP3*	19	5974092	-0,731011092	3.64×10^{-09}
ID_2450064	*NFE2L2*	2	177803791-177803840	cg21285564	*MIR1306*	22	20072475	-0,73069189	3.73×10^{-09}
ID_1710132	*MED4*	13	48650431-48650480	cg16289175	*AP2M1*	3	183898788	-0,730671931	3.74×10^{-09}
ID_3990435	*HS.511718*	7	141981986-141982035	cg22561794	*BTNL8*	5	180335743	-0,735798862	3.81×10^{-09}
ID_620376	*PPP1R11*	6	30037919-30037968	cg05832509	*TRIM4*	7	99515080	-0,730127256	3.89×10^{-09}
ID_3990435	*HS.511718*	7	141981986-141982035	cg12527097		1	5540501	-0,730101683	3.89×10^{-09}
ID_620376	*PPP1R11*	6	30037919-30037968	cg15118514	*ATP11A*	13	113499073	0,729048314	4.21×10^{-09}
ID_5360243	*ITGB5*	3	125965004-125965053	cg21285564	*MIR1306*	22	20072475	-0,729030181	4.21×10^{-09}
ID_2450064	*NFE2L2*	2	177803791-177803840	cg22309983	*TRPV1*	17	3497580	-0,728984839	4.23×10^{-09}
ID_3890326	*SOD2*	6	160103621-160103670	cg18041642		17	62062565	-0,728633188	4.34×10^{-09}
ID_1710132	*MED4*	13	48650431-48650480	cg11063328	*GTF2H1*	11	18343203	0,728080067	4.51×10^{-09}
ID_3990435	*HS.511718*	7	141981986-141982035	cg14225450	*THBS2*	6	169637832	-0,727436149	4.73×10^{-09}
ID_3990435	*HS.511718*	7	141981986-141982035	cg15920379	*TCF7L2*	10	114886678	0,727366484	4.75×10^{-09}
ID_3990435	*HS.511718*	7	141981986-141982035	cg24390732	*WDR27*	6	170060563	-0,727053754	4.86×10^{-09}

ID_5360243	*ITGB5*	3	125965004-125965053	cg27406802	*NCRNA00171*	6	29995967	-0,726881391	4.92×10^{-09}
ID_1090692	*GPBP1*	5	56594244-56594293	cg14830846	*ADAMTS2*	5	178763293	-0,726745683	4.97×10^{-09}
ID_3990435	*HS.511718*	7	141981986-141982035	cg06378617	*KRT23*	17	39095141	-0,726441285	5.08×10^{-09}
ID_3990435	*HS.511718*	7	141981986-141982035	cg20780302	*CCDC57*	17	80062106	-0,726391211	5.10×10^{-09}
ID_5360243	*ITGB5*	3	125965004-125965053	cg01797265	*LOC100128288*	17	8265333	-0,726281844	5.14×10^{-09}
ID_3520128	*TUBG2*	17	38072433-38072482	cg05368730		12	132657879	0,726152774	5.19×10^{-09}
ID_3990435	*HS.511718*	7	141981986-141982035	cg15931850		4	1550089	-0,726104847	5.20×10^{-09}
ID_5360243	*ITGB5*	3	125965004-125965053	cg20821566		17	81014667	-0,726089838	5.21×10^{-09}
ID_3990435	*HS.511718*	7	141981986-141982035	cg19952303	*SLC6A15*	12	85306323	0,725787256	5.33×10^{-09}
ID_1710026	*CCDC112*	5	114632442-114632458	cg00951302		6	166131985	-0,725744579	5.34×10^{-09}
ID_6330092	*USP16*	21	29348268-29348317	cg19033444	*PRR5*	22	45065754	0,725679389	5.37×10^{-09}
ID_1710402	*TYW1B*	7	71372479-71372528	cg13488501	*NSUN4*	1	46805584	0,725118083	5.59×10^{-09}
ID_6550315	*LYRM1*	16	20843714-20843763	cg04477431	*BAT4*	6	31631716	-0,724702839	5.76×10^{-09}
ID_2510523	*RBCK1*	20	359499-359548	cg24994593	*LDLRAD3*	11	35966604	-0,724477188	5.85×10^{-09}
ID_1940180	*PSMB9*	6	32825066-32825115	cg07088771	*TNXB*	6	32057846	-0,724025523	6.04×10^{-09}
ID_1710132	*MED4*	13	48650431-48650480	cg20821566		17	81014667	-0,723932398	6.08×10^{-09}
ID_3990435	*HS.511718*	7	141981986-141982035	cg13979407	*BTNL2*	6	32366096	-0,723915866	6.09×10^{-09}
ID_3990435	*HS.511718*	7	141981986-141982035	cg14476852		8	8506370	-0,723905945	6.09×10^{-09}
ID_3990435	*HS.511718*	7	141981986-141982035	cg02797144		16	3245099	-0,723539758	6.25×10^{-09}
ID_130164	*SPAG9*	17	46414954-46415003	cg26033526	*TAP1*	6	32819858	0,723209158	6.40×10^{-09}
ID_3990435	*HS.511718*	7	141981986-141982035	cg02071853		3	154748651	-0,7228658	6.56×10^{-09}

All positions are from the Genome Build 37

Supplemental table 4.2 Probe pairs showing a significant genetic correlation (n=25)

Transcript	CpG site	Correlation coefficient	Standard error	*P*-value
ID_3520128	cg22561794	-0,964333574	0,164671121	**0,0011625**
ID_3990435	cg02797144	-1	0	**0,0011794**
ID_3990435	cg00570469	-1	0	0,0013769
ID_3420068	cg24994593	1	0	0,0017468
ID_1780647	cg07490142	1	0	0,0018385
ID_3990435	cg15920379	1	0	0,0023268
ID_6100762	cg24994593	1	0	0,0026442
ID_7050470	cg24994593	1	0	0,0030249
ID_3990435	cg02071853	-0,949564788	0,181976145	0,0030539
ID_2510523	cg19033444	-1	0	0,0035549
ID_3990435	cg13979407	-0,978738186	0,173453373	0,0043278
ID_3990435	cg12527097	-1	0	0,0054747
ID_2510523	cg24994593	-1	0	0,0073416
ID_6330092	cg19033444	0,986137987	0,202581869	0,0090022
ID_1710132	cg21285564	-1	0	0,0100532
ID_1780647	cg13692482	-1	0	0,0114569
ID_2450064	cg21285564	-0,969743327	0,165129151	0,0128644
ID_3990435	cg14225450	-0,829976097	0,164848254	0,0137982
ID_2450064	cg22309983	-0,982426682	0,16980186	0,0155302
ID_3990435	cg19952303	0,877729037	0,137487344	0,0155373
ID_2450064	cg08931968	-1	0	0,01559
ID_2340059	cg17763566	0,858766655	0,144042657	0,0184768
ID_1090692	cg14830846	-0,829227794	0,213461147	0,0265416
ID_5360243	cg20821566	-1	0	0,0338085
ID_1710132	cg20821566	-1	0	0,0476369

Supplemental table 4.3 Overrepresented pathways identified among genes of probes with a familial effect (n=1211)

Supplemental table 4.4 Overrepresented pathways identified among genes of CpG with a familial effect (n=6291)

Supplemental table 4.5 Common overrepresented pathways identified among genes with significant familial effect in gene expression levels (n=1211) and DNA methylation (n=6291)

Supplemental tables 4.3 to 4.5 are available at https://bmcgenomics.biomedcentral.com/articles/10.1186/s12864-018-4698-6.

Chapitre 5 Ressemblances familiales des niveaux de métabolites plasmatiques

Bénédicte L. Tremblay, Frédéric Guénard, Benoît Lamarche, Louis Pérusse, Marie-Claude Vohl

L'article présenté dans ce chapitre s'intitule : *Familial resemblances in human plasma metabolites are attributable to both genetic and common environmental effects.*

Cet article est publié dans la revue : *Nutrition Research. 2019 Jan;61:22-30.*

Résumé

Introduction : Les métabolites sont d'une grande importance pour comprendre la pathogenèse de plusieurs maladies. L'étude des facteurs génétiques affectant les concentrations de métabolites permet de mieux comprendre les mécanismes sous-jacents des maladies complexes. Plusieurs études ont évalué l'héritabilité des métabolites, mais aucune n'a étudié les influences potentielles des facteurs génétiques et environnementaux dans la relation entre les métabolites et les facteurs de risque cardiométabolique (CM). Nous avons testé l'hypothèse que les effets génétiques et de l'environnement partagé contribuent à la variance des concentrations plasmatiques de métabolites et que des effets génétiques et environnementaux partagés expliquent leurs corrélations phénotypiques avec des facteurs de risque CM.

Méthodes : Pour tester cette hypothèse, la méthode de décomposition de la variance et l'analyse génétique bivariée ont été utilisées chez 48 Canadiens français provenant de 16 familles.

Résultats : Les ressemblances familiales ont été estimées pour les 147 métabolites détectés et neuf métabolites (acétylornithine, acylcarnitine C9, arginine, phosphatidylcholine acyl-alkyl C36:4, sérotonine, lysophosphatidylcholine acyl C20:4, citrulline, diméthylarginine asymétrique, phosphatidylcholine acyl-alkyl C36 :5) ont démontré un effet familial significatif (55,7 %, 18,7 % et 37,0 % pour l'héritabilité maximale, l'héritabilité génétique et l'effet de l'environnement partagé, respectivement). La citrulline, la phosphatidylcholine acyl-alkyl C36:4, la phosphatidylcholine acyl-alkyl C36:5 et la sérotonine avaient des corrélations phénotypiques significatives avec les facteurs de risque CM. La citrulline avait une corrélation génétique positive avec l'apolipoprotéine B-100, tandis que la phosphatidylcholine acyl-alkyl C36:5 avait une corrélation environnementale positive avec le cholestérol total.

Conclusions : En conclusion, les ressemblances familiales des concentrations de métabolites étaient principalement attribuables à l'effet de l'environnement partagé pour les métabolites avec un effet familial significatif. Des facteurs génétiques et environnementaux partagés peuvent également influencer la relation entre les métabolites et les facteurs de risque CM.

Abstract

Introduction: Metabolites are of great importance for understanding the pathogenesis of several diseases. Understanding the genetic contribution to metabolite concentrations may provide insights into mechanisms of complex diseases. Several studies have investigated heritability of metabolites, but none investigated potential influences of genetic and environmental factors on the relationship between metabolites and cardiometabolic (CM) risk factors. Thus, we tested the hypothesis that both genetic and common environmental effects contribute to the variance of plasma metabolite concentrations and that shared genetic and environmental effects explain their phenotypic correlations with CM risk factors.

Methods: To test this hypothesis, variance component method and bivariate genetic analysis were performed in a family-based sample of 48 French Canadians from 16 families.

Results: Familial resemblances were computed for all 147 detected metabolites and nine (acetylornithine, acylcarnitine C9, arginine, phosphatidylcholine acyl-alkyl C36:4, serotonin, lysophosphatidylcholine acyl C20:4, citrulline, asymmetric dimethylarginine, phosphatidylcholine acyl-alkyl C36:5) showed a significant familial effect (55.7%, 18.7%, and 37.0% for maximal heritability, genetic heritability, and common environmental effect, respectively). Citrulline, phosphatidylcholine acyl-alkyl C36:4, phosphatidylcholine acyl-alkyl C36:5, and serotonin had significant phenotypic correlations with CM risk factors. Citrulline had a positive genetic correlation with apolipoprotein B100, while phosphatidylcholine acyl-alkyl C36:5 had a positive environmental correlation with total cholesterol.

Conclusions: In conclusion, familial resemblances in metabolite concentrations were mainly attributable to common environmental effect when considering metabolites with a significant familial effect. Common genetic and environmental factors may also influence the relationship between metabolites and CM risk factors.

Title page

Familial resemblances in human plasma metabolites are attributable to both genetic and common environmental effects

Bénédicte L. Tremblay[1], Frédéric Guénard[1], Benoît Lamarche[1], Louis Pérusse[1, 2], Marie-Claude Vohl[1, 2*]

1. Institute of Nutrition and Functional Foods (INAF), Laval University, 2440 Hochelaga Blvd, Quebec City, QC, G1V 0A6, Canada

2. CHU de Québec Research Center – Endocrinology and Nephrology, 2705 Laurier Blvd, Quebec City, QC, G1V 4G2, Canada

Corresponding author*:
Marie-Claude Vohl, Ph.D.
Institute of Nutrition and Functional Foods (INAF), Laval University
Pavillon des Services, 2440 Hochelaga Blvd, Quebec City, QC G1V 0A6, Canada
Phone: 418 656-2131 ext. 4676; Fax: 418 656-5877
E-mail: marie-claude.vohl@fsaa.ulaval.ca

Keywords: Metabolomics; Familial resemblances; Bivariate genetic analysis; Cardiometabolic risk factors; Human.

Introduction

Metabolomics is the study of the metabolome, which is the complete set of substrates, intermediates, and products of cellular metabolism (1). Metabolomics may help further understanding of the pathogenesis of several diseases (2). Indeed, metabolites have been associated with several conditions and diseases including obesity, type 2 diabetes, and atherosclerosis (3-5). They have also been associated to traditional cardiometabolic (CM) risk factors (6, 7). Indeed, the branched-chain amino acids/aromatic amino acids related pattern and the glutamine/glycine/serine/asparagine related pattern were significantly associated with metabolic syndrome in middle-aged men (6). Cheng S et al. also reported that several amino acids including glutamine, glutamate, proline, alanine, and tyrosine were associated with multiple CM risk factor including insulin, blood pressure, triglyceride and HDL cholesterol concentrations in *the Framingham Heart Study* and the *Malmö Diet and Cancer Study* (7).

Blood metabolite concentrations showed inter-individual variability due to variations in genetic and environmental exposure (8). Genome-wide association studies (GWAS) have discovered numerous genetic loci associated with metabolites from human adult blood and/or urine samples (9-20). Several studies have investigated correlations of metabolites concentrations between monozygotic twins (16, 21, 22), while other estimated genetic heritability or the contribution of both genetic and shared environmental effects (16, 18, 21, 23-25). However, these studies did not investigate potential influences of genetic and environmental factors on the relationship between metabolites and CM risk factors. Understanding the genetic and environmental contributions to metabolite concentrations may provide insights into mechanisms of action of complex diseases.

Quantitative genetic approaches estimate the contribution of genetic and environmental influences on a phenotype using a multifactorial model based on family pedigree, which posits that a phenotype is caused by genetic and environmental effects, usually expressed as variance components (26). The genetic component (G) consists of several genes having a small, linear additive effect. The environmental component can be further partitioned into factors that are in common (C) among family members and those that are not shared or residual (E). Thus, the total phenotypic variance of a trait may be expressed as $V_P = V_G + V_C + V_E$. The most commonly used study designs to estimate heritability are nuclear families (parents and their biological offspring) and twins (monozygotic and dizygotic twins) (27). The variance component method described above can be extended to a bivariate (two traits) genetic analysis. In the context of the present study, the phenotypic correlation (ρP) between two traits (plasma metabolite and CM risk factor) can be portioned into genetic (ρG) and random environmental (ρE) correlations (28).

Thus, we tested the hypothesis that both genetic and common environmental effects contribute to the variance of plasma metabolite concentrations and that shared genetic and environmental effects explain their phenotypic

correlations with CM risk factors. To test this hypothesis, variance component method and bivariate genetic analysis were performed in a family-based sample of 48 French Canadians from 16 families.

Methods

Patients and design

A total of 48 Caucasian French Canadian from 16 families were recruited in the Greater Quebec City metropolitan area, in Canada, as part of the GENERATION Study. Families were composed of 16 mothers, 6 fathers, and 26 children. Most parents had a university education and a family income > 80 000$ CAN. Parents had to be the biological parents of their child (or children), in good general health, non-smokers, with body mass index (BMI) ranging between 18 and 35 kg/m^2, and free of any metabolic conditions requiring treatment although the use of Synthroid® (levothyroxine) or oral contraceptive was tolerated. Families living under the same roof comprised at least the mother and one child aged between 8 and 18. Children also had to be non-smokers, in good general health and not using psycho-stimulators [Ritalin® (methylphenidate), Concerta® (methylphenidate), and Strattera® (atomoxetine)]. Parents and children were asked to complete several dietary, physical activity, medical history, and pregnancy questionnaires under the supervision of a registered dietitian during their visit at the Institute of Nutrition and Functional Foods (INAF). The experimental protocol was approved by the Ethics Committees of Laval University Hospital Research Center and Laval University. All participants signed an informed consent document. Parental consent was also obtained by signing the child assent document. Exploratory approach with no *a priori* power analysis was used in the present study.

Anthropometric and metabolic measurements

Body weight, waist girth, and height were measured according to the procedures recommended by the Airlie Conference (29). Blood samples were collected from an antecubital vein into vacutainer tubes containing EDTA after 12-hour overnight fast and 48-hour alcohol abstinence. Plasma was separated by centrifugation (2 500xg for 10 min at 4 °C), aliquoted, and stored at -80 °C. Plasma samples were thawed once and were kept on ice during metabolite profiling procedure.

Metabolite measurements

Most studies carried out heritability analysis using targeted metabolomics approach and some studies considered a wide range of serum metabolites measured by the Biocrates AbsoluteIDQ p150kit (Biocrates Life Sciences, AG, Austria) (21, 22). For comparison purposes, targeted metabolomics approach using the Biocrates AbsoluteIDQp180 kit (Biocrates Life Sciences, AG, Austria) was used (30). Among the 188 metabolites interrogated by The Biocrates Absolute IDQ p180 (Biocrates Life Sciences AG, Austria) mass spectrometry method, concentrations were detected for 159 metabolites from six compound classes including: 22 acylcarnitines (AC), 21 amino acids (AA), 13 biogenic amines (BA), hexose (sum of hexoses – about 90–95% glucose), 87 glycerophospholipids [13 lysophosphatidylcholines acyl (LysoPC a) and 74 phosphatidylcholines (PC diacyl (aa) and acyl-alkyl (ae)], and 15 sphingolipids (SM). This is in accordance with the numbers of metabolites detected in other studies using the same Biocrates kit (31, 32). The metabolite profiling was carried

out using 10 µL of plasma and according to the manufacturer's instructions at CHENOMX (Edmonton, AL, Canada). For all analyzed metabolites, the concentrations are reported in µM. Values considered outliers, defined as values falling outside of the means ± 4 standard deviations (12 outliers for all metabolites) were excluded from heritability analyses. Metabolite concentrations below the lower limits of detection (LOD) were given a value of LOD/2. Moreover, 12 metabolites having > 25% of samples (13 subjects) below the LOD were excluded (C10:1, C12, C14:1-OH, C3-DC (C4-OH), C5-OH (C3-DC-M), C5-DC (C6-OH), C7-DC, LysoPC a C24:0, PC aa C24:0, PC aa C40:1, spermidine, and spermine). All metabolites were natural log-transformed before analysis to ensure normal distribution. Thus, a total of 147 metabolites were considered in analysis.

Statistical analyses

Statistical Analysis Software (SAS) was used to compute differences in CM parameters between fathers and mothers, and between daughters and sons using an unpaired t-test. Normality was assessed using the Univariate command in SAS and skewness and kurtosis values between -2 and 2 and -4 and 4, respectively were used as cut-offs. R software v2.14.1 (R Foundation for Statistical Computing; http://www.r-project.org) (33) was used to compute average absolute correlations of 147 natural log-transformed metabolite concentrations between relative pairs. Unrelated pairs comprise all possible comparisons (n=1078) between unrelated subjects (from different families) and spouses. For heritability analysis, adjustments were made for the effects of sex, age, and classes of BMI (underweight, normal, overweight, and obese) using a standard least squares model in JMP software v12. Classes of BMI were used to compare BMI among parents with BMI percentile among children. Cut-offs for BMI and BMI percentile were both from the World Health Organization (34, 35). Residuals from this model were used to compute heritability estimates using the variance component method implemented in QTDT v2.6.1 (36). The variance component method (full general model) was used in the present study for comparison purposes with other studies. We used a full general model in which the variance of each metabolite level was partitioned into polygenic effects (Vg), common environmental effects shared by family members (Vc), and non-shared environmental effects unique to each individual (Ve). We tested and computed *P*-value for familial effect by comparing this full general model against a null model of no familial resemblance in which Vg = Vc = 0. We then computed maximal heritability, as the proportion of variance accounted by genetic and common environmental effects (Vg+Vc/Vg+Vc+Ve), genetic heritability, as the proportion of variance accounted by genetics effects (Vg/Vg+Vc+Ve), and common environmental effect as the proportion of variance accounted by common environmental effects (Vc/Vg+Vc+Ve). Additionally, we computed heritability estimates in metabolites showing a significant familial effect (Vg and Vc $\neq 0$, $P \leq 0.05$). However, the small sample size may limit the statistical power to detect low heritability estimates. We also computed reduced models in which the variance in metabolite levels was partitioned into Vg and Ve [genetic heritability = (Vg/(Vg+Ve)] and Vc and Ve [common environmental effects = (Vc/(Vc+Ve)]. The best fit model was also chosen by comparing the fit statistics

[Akaike's Information Criterion (AIC)]. The best fit model was defined as the model with the smallest AIC value (37).

Polar histogram is used to display a large number of histograms in a small space. The polar histogram was constructed with the R package "phenotypicForest" (https://github.com/chrislad/phenotypicForest, accessed July 23, 2018) using graphical "ggplot2" package. The heat map was also constructed with the R package "ggplot2". Phenotypic Pearson correlations (ρP) between all 147 metabolites and nine CM risk factors were calculated for all 48 subjects. Once again metabolite concentrations were adjusted for the effects of sex, age, and classes of BMI. Only the nine metabolites with significant familial effect were further considered for bivariate genetic analysis. Significant phenotypic correlations (ρP) ($P \leq 0.05$) between pairs of traits have been portioned into genetic (ρG) and environmental (ρE) correlations using a bivariate genetic analysis in SOLAR Eclipse version 7.6.4. Bonferroni corrections were used to account for multiple testing in phenotypic, genetic, and environmental correlations (0.05/9 CM risk factors = 0.0056).

Results

Characteristics of study participants

Characteristic of participants within each subgroup are given in Table 1. High-density lipoprotein cholesterol (HDL-C) and total cholesterol/HDL-C (TC/HDL-C) were significantly different between fathers and mothers, whereas glucose concentrations were significantly different between boys and girls. Concentrations of all 147 metabolites measured in the fasting state are presented in Supplemental Table S1.

Correlations of metabolite concentrations between relative pairs

Average absolute correlations across all 147 detected metabolites were calculated between pairs from natural log-transformed metabolite concentrations. The average correlation coefficients were 0.28 ± 0.18 among siblings (n=13), 0.23 ± 0.17 for mother-offspring pairs (n=26), 0.20 ± 0.15 for parent-offspring pairs (n=37), and 0.02 ± 0.01 among unrelated individuals (n=1078). According to these correlations, metabolite concentrations are more similar in related individuals (siblings, mother-offspring, and parent-offspring pairs) than in unrelated individuals.

Heritability analysis of metabolite concentrations

Heritability analyses were conducted to better characterize the similarity in metabolite concentrations of related individuals. When considering all detected metabolites (n=147), we obtained an average maximal heritability of 17.8% (ranges 0 to 82.5%), a genetic heritability of 11.6% (ranges 0 to 82.5%), and a common environmental effect of 6.2% (ranges 0 to 58.1%). A total of 44 metabolites (30.0% of all metabolites) had an estimated maximal heritability of 0%. Polar histogram of the distribution of genetic heritability, common environmental effect, and non-shared effect for all 147 metabolites is depicted in Figure 1. Detailed heritability estimates with standard errors of all 147 metabolites are also provided in Supplemental Table S2. Reduced and best fit models are also available in Supplemental Table S3.

A total of nine metabolites showed a significant ($P \leq 0.05$) familial effect: acetylornithine, AC C9, arginine, PC ae C36:4, serotonin, LysoPC a C20:4, citrulline, asymmetric dimethylarginine (ADMA), and PC ae C36:5. When considering these nine metabolites, we obtained an average maximal heritability of 55.7% (ranges 38.6 to 82.5%), a genetic heritability of 18.7% (ranges 0 to 82.5%), and a common environmental effect of 37.0% (ranges 0 to 58.1%). Detailed heritability estimates of all nine metabolites with significant familial effect are presented in Table 2.

Bivariate genetic analysis with cardiometabolic risk factors

Phenotypic correlations between all 147 metabolites and nine CM risk factors (apolipoprotein B100 (ApoB100), CRP, glucose, HDL-C, insulin, LDL-C, TC, TC/HDL-C, and TG) were calculated for all 48 subjects (Supplemental

Table S4). Metabolites within a compound class were inter-correlated as shown in the heat map in Supplemental Figure S1. The hypothesis was that shared genetic and environmental effects explain the phenotypic correlations of metabolites showing significant familial effect with CM risk factors. In order to test this hypothesis, only the nine metabolites with a significant familial effect were further considered for bivariate genetic analysis (Supplemental Table S5). Thus, significant phenotypic correlations ($P \leq 0.05$) were obtained for citrulline, PC ae C36:4, PC ae C36:5, and serotonin (Table 3). Positive phenotypic correlation between PC ae C36:5 and HDL-C (r=0.42) remained significant after Bonferroni correction (adjusted P=0.05/9=0.0056). Phenotypic correlations were partitioned into genetic and environmental correlations using bivariate genetic analysis. Citrulline had a significant positive genetic correlation with ApoB100 (r=1), while PC ae C36:5 had a significant positive environmental correlation with TC (r=1) (Table 3). These genetic and environmental correlations had a standard error of 0 suggesting that for each individual, the whole phenotypic correlation can be fully explained by the genetic or environmental correlation. Genetic and environmental correlations also remained significant after Bonferroni correction (adjusted P for citrulline=0.05 and adjusted P for PC ae C36:5=0.05/3=0.017).

Discussion

The first part of the hypothesis that both genetic and common environmental effects contribute to the variance of plasma metabolite concentrations is accepted. Higher average absolute correlations of metabolite concentrations between relative pairs compared to unrelated individuals suggest an underlying genetic similarity. Correlations in parent-offspring and spouse pairs were similar to those reported by Draisma et al. in 123 metabolites (22). However, the number of spouse pairs (n=6) in the present study greatly limits interpretation of the results. Nevertheless, a correlation between spouses may suggest a contribution of the common environment shared by family members to the familial resemblances in metabolite concentrations (22).

Among the 147 detected metabolites nine showed a significant familial effect. Two-thirds of the familial effect were due to common environmental effect and one third was attributable to genetic heritability for these nine metabolites. A strong genetic component in familial resemblance of metabolites concentrations is reported in several studies (22-25). Homogeneous environment of the cohort subjects, with limited confounding environmental factors, may allow the detection of more significant common environmental effect in the present study. Indeed, all participants are of French-Canadian descent, live in the same city (Quebec City), and have similar socio-economic characteristics and dietary patterns (data not shown).

Four metabolites out of the nine showing a significant familial effect were related to arginine biosynthesis and metabolism. Acetylornithine and citrulline are acetylated amino acid intermediates of the biosynthesis of arginine from glutamate (38). Moreover, ADMA, which is synthesized from proteins containing L-arginine, plays a major role in endothelial dysfunction (39, 40). Elevated concentrations of ADMA promote atherosclerosis by inhibiting nitric oxide (NO) synthesis (40). It is thus considered as an independent risk factor for progression of atherosclerosis, cardiovascular death, and all-cause mortality (41-43). To the best of our knowledge, this is the first study that estimated familial resemblances in plasma ADMA concentrations in healthy families. One study in a large European population reported a genetic heritability of 0% for the combination of symmetric dimethylarginine and ADMA (18). Similarly, ADMA concentrations had no genetic heritability but a considerable common environmental effect (38.6%) in the present study. The contribution of common environmental effect to familial resemblances in ADMA concentrations may suggest a potential link between familial environment shared by family members and the susceptibility to endothelial dysfunction and atherosclerosis. Arginine and BA implicated in arginine metabolism (acetylornithine and citrulline) also depicted high common environmental effect. Arginine had the highest heritability estimate among AA in families heavily burdened with premature coronary artery disease (CAD) (25). However, it is difficult to compare results because the study by Shah et al. was conducted on families heavily burdened by premature CAD. Shin et al. estimated heritability of more than 400 metabolites using an ACE model in the TwinsUK cohort (18). Arginine, acetylornithine, and citrulline had a genetic heritability of 24.4%, 55.3%, and 47.1%, respectively in this cohort (18). In the present study, arginine,

acetylornithine, and citrulline had low genetic heritability varying from 0 to 7.3%. This difference may be explained by the homogeneous environment of the cohort subjects but also the small sample size that can affect the reliability of low values of heritability estimates.

Several studies have investigated heritability of AC concentrations in adults and newborns (24, 25). The genetic heritability of long-chain AC was of 39% in Shah et al. Once again, it is difficult to compare results because the study by Shah et al. was conducted on families heavily burdened with CAD (25). In female twins, AC C9 had a genetic heritability of 62% and a common environmental effect of 0% (21). We observed an almost null genetic heritability for AC C9 while there was an important common environmental effect (49.1%). In line with our results, a stronger environmental component for the variation in medium-chain and long-chain AC concentrations was reported in newborn twins (24).

PC and LysoPC are major classes of phospholipids that play an important role in several key processes such as cell survival and inflammation and are included in the structure of lipoproteins and platelets (44, 45). A GWAS by Dermikan et al. identified loci associated with plasma phospholipid concentrations (13). This is concordant with the fact that PC ae C36:4 and PC ae C36:5 had a strong genetic heritability in this present study. This may also be in line with the high heritability of fasting triglycerides (22, 46).

Serotonin is an amine neurotransmitter that has essential role in regulation of emotion and anxiety (47). In Hutterites, a founder population, the narrow and broad sense heritability of serotonin concentrations were of 51% and 100%, respectively (48). In the present study, serotonin concentrations had an important common environmental effect and a null genetic heritability. However, results should be compared with caution considering that Abney et al. used a different model in which the variance was portioned into additive, dominance, and environmental effects (48). Shin et al. reported a genetic heritability of serotonin concentrations of 33.1% in an ACE model (18).

The second part of the hypothesis that shared genetic and environmental effects explain phenotypic correlations between metabolites and CM risk factors is also accepted. Citrulline had a significant positive phenotypic and genetic correlations with ApoB100, indicating that genetic factors affect traits in the same way (increase or decrease concentrations). Citrulline supplementation increased NO production and reduced lipoproteins oxidation (LDL, ApoB) in subjects with angina (49). Citrulline has been shown to increase expression of *APOBEC2* encoding for the apolipoprotein B mRNA editing enzyme catalytic subunit 2 (50). However, little is known about the function and editing activity of APOBEC2. Moreover, PC ae C36:5 had a significant positive environmental correlation with TC, indicating that environmental factors affect traits in the same way. Enzymes of PC metabolism have been shown to potentially affect cholesterol trafficking, storage, and homeostasis (51).

Accordingly, PC, which were highly inter-correlated in the present study, have also been associated with plasma lipoprotein homeostasis (52).

The present study has strengths and limitations. The main strength lies in the inclusion of bivariate genetic analysis that allows a better characterization of factors influencing the relationship between plasma metabolites and CM risk factors. This is also the first study to evaluate familial resemblances in plasma metabolite concentrations measured by the Biocrates AbsoluteIDQ p180 kit (Biocrates Life Sciences, AG, Austria). The use of a general model that included both genetic and common environmental effects allows a better characterization of familial resemblances. The study of a founder population with relatively homogeneous genetics and shared environment is a new aspect in this field (53). Regarding limitations, the main one is the small sample size that limits the statistical power to detect low heritability estimates. This may explain the considerable number of metabolites with a maximal heritability of 0% (30% of all metabolites). Moreover, the small sample size may limit the refinement of correlation estimates in bivariate genetic analysis. Low or null heritability estimates may also be explained by the fact that plasma samples were collected after an overnight fast. Indeed, some metabolite concentrations such as hexose are decreased in fasting state thus reducing the inter-individual variation (54). Moreover, the full GCE model was not the best fit model as defined by the model with the smallest AIC value. This may limit the interpretation of the results. However, for the majority of metabolites the difference in AIC was small (< 2), which suggests that models have approximately equal weight in the data (37). For comparison purposes, we analysed results of full GCE models for all metabolites (18, 21, 23, 24). However, reduced and best fit models are available in the present study. Plasma metabolites concentrations can be influenced by several factors including diet, BMI, and gut microbiome (21, 55-57). We have taken into account the effect of BMI (classes of BMI) but not the potential confounding effects of diet and gut microbiome. Despite that we have data on diet we decided not to use them considering the potential bias related to only one assessment with a self-reported dietary record (58). Moreover, we considered a limited number of objective confounders (age, sex, and BMI) in order to prevent the reduction of the statistical power to detect significant heritability estimates and correlations. Finally, since paternity was not genetically confirmed in our study, non-paternity may impact heritability estimates more specifically genetic heritability estimates. However, the probability that non-paternity influenced our findings is very small considering the low non-paternity rate in the general population (59), and the small number of fathers (n=6) included in the present study.

In conclusion, familial resemblances in metabolite concentrations were mainly attributable to common environmental effect when considering metabolites with a significant familial effect. To the best of our knowledge, this is the first study to report impact of common environmental effect in familial resemblances in plasma ADMA concentrations, a novel biomarker of risk for cardiovascular diseases. Common genetic and environmental

factors also seem to influence metabolites and CM risk factors. Further studies are needed to better understand these possible influences and their potential impact on disease development.

Acknowledgements

We would like to thank Christian Couture, Véronique Garneau, and Catherine Raymond who contributed to the success of this study. This work was supported by the Canada Research Chair in Genomics Applied to Nutrition and Metabolic Health, which is not involved in the conduct of the research and preparation of the article. MCV is Tier 1 Canada Research Chair in Genomics Applied to Nutrition and Metabolic Health. BLT is a recipient of a scholarship from the Canadian Institutes of Health Research (CIHR). Authors have no potential conflicts of interest.

References

1. Griffin JL, Atherton H, Shockcor J, Atzori L. Metabolomics as a tool for cardiac research. Nature reviews Cardiology. 2011;8(11):630-43.

2. Gowda GA, Zhang S, Gu H, Asiago V, Shanaiah N, Raftery D. Metabolomics-based methods for early disease diagnostics. Expert review of molecular diagnostics. 2008;8(5):617-33.

3. Floegel A, Stefan N, Yu Z, Muhlenbruch K, Drogan D, Joost HG, et al. Identification of serum metabolites associated with risk of type 2 diabetes using a targeted metabolomic approach. Diabetes. 2013;62(2):639-48.

4. Tomas L, Edsfeldt A, Mollet IG, Perisic Matic L, Prehn C, Adamski J, et al. Altered metabolism distinguishes high-risk from stable carotid atherosclerotic plaques. European heart journal. 2018.

5. Jourdan C, Petersen AK, Gieger C, Doring A, Illig T, Wang-Sattler R, et al. Body fat free mass is associated with the serum metabolite profile in a population-based study. PloS one. 2012;7(6):e40009.

6. Ntzouvani A, Nomikos T, Panagiotakos D, Fragopoulou E, Pitsavos C, McCann A, et al. Amino acid profile and metabolic syndrome in a male Mediterranean population: A cross-sectional study. Nutrition, metabolism, and cardiovascular diseases : NMCD. 2017;27(11):1021-30.

7. Cheng S, Rhee EP, Larson MG, Lewis GD, McCabe EL, Shen D, et al. Metabolite profiling identifies pathways associated with metabolic risk in humans. Circulation. 2012;125(18):2222-31.

8. Suhre K, Gieger C. Genetic variation in metabolic phenotypes: study designs and applications. Nature reviews Genetics. 2012;13(11):759-69.

9. Suhre K, Shin SY, Petersen AK, Mohney RP, Meredith D, Wagele B, et al. Human metabolic individuality in biomedical and pharmaceutical research. Nature. 2011;477(7362):54-60.

10. Nicholson G, Rantalainen M, Li JV, Maher AD, Malmodin D, Ahmadi KR, et al. A genome-wide metabolic QTL analysis in Europeans implicates two loci shaped by recent positive selection. PLoS genetics. 2011;7(9):e1002270.

11. Illig T, Gieger C, Zhai G, Romisch-Margl W, Wang-Sattler R, Prehn C, et al. A genome-wide perspective of genetic variation in human metabolism. Nature genetics. 2010;42(2):137-41.

12. Gieger C, Geistlinger L, Altmaier E, Hrabe de Angelis M, Kronenberg F, Meitinger T, et al. Genetics meets metabolomics: a genome-wide association study of metabolite profiles in human serum. PLoS genetics. 2008;4(11):e1000282.

13. Demirkan A, van Duijn CM, Ugocsai P, Isaacs A, Pramstaller PP, Liebisch G, et al. Genome-wide association study identifies novel loci associated with circulating phospho- and sphingolipid concentrations. PLoS genetics. 2012;8(2):e1002490.

14. Hicks AA, Pramstaller PP, Johansson A, Vitart V, Rudan I, Ugocsai P, et al. Genetic determinants of circulating sphingolipid concentrations in European populations. PLoS genetics. 2009;5(10):e1000672.

15. Tukiainen T, Kettunen J, Kangas AJ, Lyytikainen LP, Soininen P, Sarin AP, et al. Detailed metabolic and genetic characterization reveals new associations for 30 known lipid loci. Human molecular genetics. 2012;21(6):1444-55.

16. Kettunen J, Tukiainen T, Sarin AP, Ortega-Alonso A, Tikkanen E, Lyytikainen LP, et al. Genome-wide association study identifies multiple loci influencing human serum metabolite levels. Nature genetics. 2012;44(3):269-76.

17. Rhee EP, Ho JE, Chen MH, Shen D, Cheng S, Larson MG, et al. A genome-wide association study of the human metabolome in a community-based cohort. Cell metabolism. 2013;18(1):130-43.

18. Shin SY, Fauman EB, Petersen AK, Krumsiek J, Santos R, Huang J, et al. An atlas of genetic influences on human blood metabolites. Nature genetics. 2014;46(6):543-50.

19. Yu B, Zheng Y, Alexander D, Morrison AC, Coresh J, Boerwinkle E. Genetic determinants influencing human serum metabolome among African Americans. PLoS genetics. 2014;10(3):e1004212.

20. Draisma HHM, Pool R, Kobl M, Jansen R, Petersen AK, Vaarhorst AAM, et al. Genome-wide association study identifies novel genetic variants contributing to variation in blood metabolite levels. Nature communications. 2015;6:7208.

21. Menni C, Zhai G, Macgregor A, Prehn C, Romisch-Margl W, Suhre K, et al. Targeted metabolomics profiles are strongly correlated with nutritional patterns in women. Metabolomics : Official journal of the Metabolomic Society. 2013;9(2):506-14.

22. Draisma HH, Beekman M, Pool R, van Ommen GJ, Adamski J, Prehn C, et al. Familial resemblance for serum metabolite concentrations. Twin research and human genetics : the official journal of the International Society for Twin Studies. 2013;16(5):948-61.

23. Nicholson G, Rantalainen M, Maher AD, Li JV, Malmodin D, Ahmadi KR, et al. Human metabolic profiles are stably controlled by genetic and environmental variation. Molecular systems biology. 2011;7:525.

24. Alul FY, Cook DE, Shchelochkov OA, Fleener LG, Berberich SL, Murray JC, et al. The heritability of metabolic profiles in newborn twins. Heredity. 2013;110(3):253-8.

25. Shah SH, Hauser ER, Bain JR, Muehlbauer MJ, Haynes C, Stevens RD, et al. High heritability of metabolomic profiles in families burdened with premature cardiovascular disease. Molecular systems biology. 2009;5:258.

26. Pérusse LR, T.K. Bouchard, C. Evidence of a genetic component to obesity from genetic epidemiology. In: In: Bray GA BC, eds, editor. Handbook of obesity - Epidemiology, Etiology, and Physiopathology; Boca Raton, FL: CRC Press; 2013. p. 91-104.

27. Rice TK. Familial resemblance and heritability. Advances in genetics. 2008;60:35-49.

28. Heath AC, Kessler RC, Neale MC, Hewitt JK, Eaves LJ, Kendler KS. Testing hypotheses about direction of causation using cross-sectional family data. Behavior genetics. 1993;23(1):29-50.

29. Callaway C, Chumlea W, Bouchard C, Himes J, Lohman T, Martin A, et al. Standardization of anthropomeric measurements : The Airlie (VA) Consensus Conference. Champaign, IR, USA; 1988.

30. Sciences BL. Biocrates Life Sciences, The Deep Phenotyping Company. Austria; 2017. Contract No.: October 23.

31. Gonzalez-Freire M, Moaddel R, Sun K, Fabbri E, Zhang P, Khadeer M, et al. Targeted Metabolomics Shows Low Plasma Lysophosphatidylcholine 18:2 Predicts Greater Decline of Gait Speed in Older Adults: The Baltimore Longitudinal Study of Aging. The journals of gerontology Series A, Biological sciences and medical sciences. 2018.

32. Carayol M, Licaj I, Achaintre D, Sacerdote C, Vineis P, Key TJ, et al. Reliability of Serum Metabolites over a Two-Year Period: A Targeted Metabolomic Approach in Fasting and Non-Fasting Samples from EPIC. PloS one. 2015;10(8):e0135437.

33. Team RC. R: A language and environment for statistical computing. R Foundation for Statistical Computing. Vienna, Austria. 2013.

34. World Health Organization G. Obesity: Preventing and managing the global epidemic. Report of a WHO Consultation. Geneva: World Health Organization; 2000.

35. de Onis M, Lobstein T. Defining obesity risk status in the general childhood population: which cut-offs should we use? International journal of pediatric obesity : IJPO : an official journal of the International Association for the Study of Obesity. 2010;5(6):458-60.

36. Abecasis GR, Cardon LR, Cookson WO. A general test of association for quantitative traits in nuclear families. American journal of human genetics. 2000;66(1):279-92.

37. Burnham KP A, DR. Model Selection and Multimodel Inference: A Practical Information-Theoretic Approach. edn. n, editor. New York: Springer; 2002.

38. Morizono H, Cabrera-Luque J, Shi D, Gallegos R, Yamaguchi S, Yu X, et al. Acetylornithine transcarbamylase: a novel enzyme in arginine biosynthesis. Journal of bacteriology. 2006;188(8):2974-82.

39. Vallance P, Leone A, Calver A, Collier J, Moncada S. Accumulation of an endogenous inhibitor of nitric oxide synthesis in chronic renal failure. Lancet. 1992;339(8793):572-5.

40. Sibal L, Agarwal SC, Home PD, Boger RH. The Role of Asymmetric Dimethylarginine (ADMA) in Endothelial Dysfunction and Cardiovascular Disease. Current cardiology reviews. 2010;6(2):82-90.

41. Lu TM, Ding YA, Lin SJ, Lee WS, Tai HC. Plasma levels of asymmetrical dimethylarginine and adverse cardiovascular events after percutaneous coronary intervention. European heart journal. 2003;24(21):1912-9.

42. Valkonen VP, Paiva H, Salonen JT, Lakka TA, Lehtimaki T, Laakso J, et al. Risk of acute coronary events and serum concentration of asymmetrical dimethylarginine. Lancet. 2001;358(9299):2127-8.

43. Meinitzer A, Seelhorst U, Wellnitz B, Halwachs-Baumann G, Boehm BO, Winkelmann BR, et al. Asymmetrical dimethylarginine independently predicts total and cardiovascular mortality in individuals with angiographic coronary artery disease (the Ludwigshafen Risk and Cardiovascular Health study). Clinical chemistry. 2007;53(2):273-83.

44. van Meer G, Voelker DR, Feigenson GW. Membrane lipids: where they are and how they behave. Nature reviews Molecular cell biology. 2008;9(2):112-24.

45. Engelmann B, Kogl C, Kulschar R, Schaipp B. Transfer of phosphatidylcholine, phosphatidylethanolamine and sphingomyelin from low- and high-density lipoprotein to human platelets. The Biochemical journal. 1996;315 (Pt 3):781-9.

46. Knoblauch H, Busjahn A, Munter S, Nagy Z, Faulhaber HD, Schuster H, et al. Heritability analysis of lipids and three gene loci in twins link the macrophage scavenger receptor to HDL cholesterol concentrations. Arteriosclerosis, thrombosis, and vascular biology. 1997;17(10):2054-60.

47. Marcinkiewcz CA, Mazzone CM, D'Agostino G, Halladay LR, Hardaway JA, DiBerto JF, et al. Serotonin engages an anxiety and fear-promoting circuit in the extended amygdala. Nature. 2016;537(7618):97-101.

48. Abney M, McPeek MS, Ober C. Broad and narrow heritabilities of quantitative traits in a founder population. American journal of human genetics. 2001;68(5):1302-7.

49. Morita M, Sakurada M, Watanabe F, Yamasaki T, Doi H, Ezaki H, et al. Effects of Oral L-Citrulline Supplementation on Lipoprotein Oxidation and Endothelial Dysfunction in Humans with Vasospastic Angina. Immunology, endocrine & metabolic agents in medicinal chemistry. 2013;13(3):214-20.

50. Sato Y, Probst HC, Tatsumi R, Ikeuchi Y, Neuberger MS, Rada C. Deficiency in APOBEC2 leads to a shift in muscle fiber type, diminished body mass, and myopathy. The Journal of biological chemistry. 2010;285(10):7111-8.

51. Lagace TA. Phosphatidylcholine: Greasing the Cholesterol Transport Machinery. Lipid insights. 2015;8(Suppl 1):65-73.

52. Vance DE. Role of phosphatidylcholine biosynthesis in the regulation of lipoprotein homeostasis. Current opinion in lipidology. 2008;19(3):229-34.

53. Moreau C, Lefebvre JF, Jomphe M, Bherer C, Ruiz-Linares A, Vezina H, et al. Native American admixture in the Quebec founder population. PloS one. 2013;8(6):e65507.

54. Krug S, Kastenmuller G, Stuckler F, Rist MJ, Skurk T, Sailer M, et al. The dynamic range of the human metabolome revealed by challenges. FASEB journal : official publication of the Federation of American Societies for Experimental Biology. 2012;26(6):2607-19.

55. Playdon MC, Moore SC, Derkach A, Reedy J, Subar AF, Sampson JN, et al. Identifying biomarkers of dietary patterns by using metabolomics. The American journal of clinical nutrition. 2017;105(2):450-65.

56. Scalbert A, Brennan L, Manach C, Andres-Lacueva C, Dragsted LO, Draper J, et al. The food metabolome: a window over dietary exposure. The American journal of clinical nutrition. 2014;99(6):1286-308.

57. Zhao H, Shen J, Djukovic D, Daniel-MacDougall C, Gu H, Wu X, et al. Metabolomics-identified metabolites associated with body mass index and prospective weight gain among Mexican American women. Obesity science & practice. 2016;2(3):309-17.

58. Archer E, Blair SN. Implausible data, false memories, and the status quo in dietary assessment. Advances in nutrition. 2015;6(2):229-30.

59. Greeff JM, Erasmus JC. Three hundred years of low non-paternity in a human population. Heredity. 2015;115(5):396-404.

Tables

Table 5.1 Biochemical parameters of parents and children

Biochemical parameters	Fathers (n=6)	Mothers (n=16)	Boys (n=18)	Girls (n=8)
Age (years)	42.0 ± 2.8	42.4 ± 6.0	11.9 ± 3.8	10.1 ± 2.0
BMI (kg/m^2)	24.8 ± 1.3	23.5 ± 3.4	-	-
BMI percentile	-	-	48.9 ± 32.7	52.1 ± 29.8
TC (mmol/L)	4.70 ± 0.61	4.67 ± 0.55	4.35 ± 0.53	4.15 ± 0.48
HDL-C (mmol/L)	1.36 ± 0.33[a]	1.73 ± 0.35[a]	1.56 ± 0.28	1.54 ± 0.31
LDL-C (mmol/L)	2.83 ± 0.67	2.53 ± 0.49	2.36 ± 0.42	2.20 ± 0.52
TC/HDL-C	3.62 ± 0.88[a]	2.79 ± 0.63[a]	2.84 ± 0.42	2.79 ± 0.60
ApoB100 (g/L)	0.89 ± 0.22	0.77 ± 0.11	0.72 ± 0.12	0.67 ± 0.15
TG (mmol/L)	1.13 ± 0.35	0.88 ± 0.34	0.94 ± 0.39	0.89 ± 0.41
Glucose (mmol/L)	5.25 ± 0.38	5.19 ± 0.34	4.92 ± 0.24[b]	4.73 ± 0.14[b]
Insulin (pmol/L)	56.33 ± 7.99	70.44 ± 38.68	64.44 ± 22.50	81.00 ± 40.75
CRP (mg/L)	0.45 ± 0.39	0.89 ± 0.93	0.47 ± 0.79	0.33 ± 0.22

All values are means ± SD. [a] Means are significantly different ($P \leq 0.05$) between fathers and mothers. [b] Means are significantly different ($P \leq 0.05$) between boys and girls. *P*-values are provided by an unpaired t-test. Abbreviations: Apolipoprotein B100 (ApoB100), C-reactive protein (CRP), High-density lipoprotein cholesterol (HDL-C), Low-density lipoprotein cholesterol (LDL-C), Total cholesterol (TC), Total cholesterol/HDL-C (TC/HDL-C), Triglycerides (TG).

Table 5.2 Heritability estimates of nine metabolites with significant familial effect

Metabolites	Maximal heritability (%)	Genetic heritability (%)	Common env. effect (%)	*P*-value familial effect
Acetylornithine	65.4	7.3	58.1	0.0005
Acylcarnitine C9	50.5	1.4	49.1	0.0024
Arginine	53.8	0	53.8	0.0057
PC ae C36:4	76.7	76.7	0	0.0081
Serotonin	51.5	0	51.5	0.0091
LysoPC a C20:4	39.5	0	39.5	0.012
Citrulline	42.3	0	42.3	0.015
ADMA	38.6	0	38.6	0.029
PC ae C36:5	82.5	82.5	0	0.034

Heritability estimates are obtained from a full general GCE model. Abbrevations: Acyl (a), Acyl-alkyl (ae), Asymmetric dimethylarginine (ADMA), Environmental (env.), Lysophosphatidylcholine (LysoPC), Phosphatidylcholine (PC).

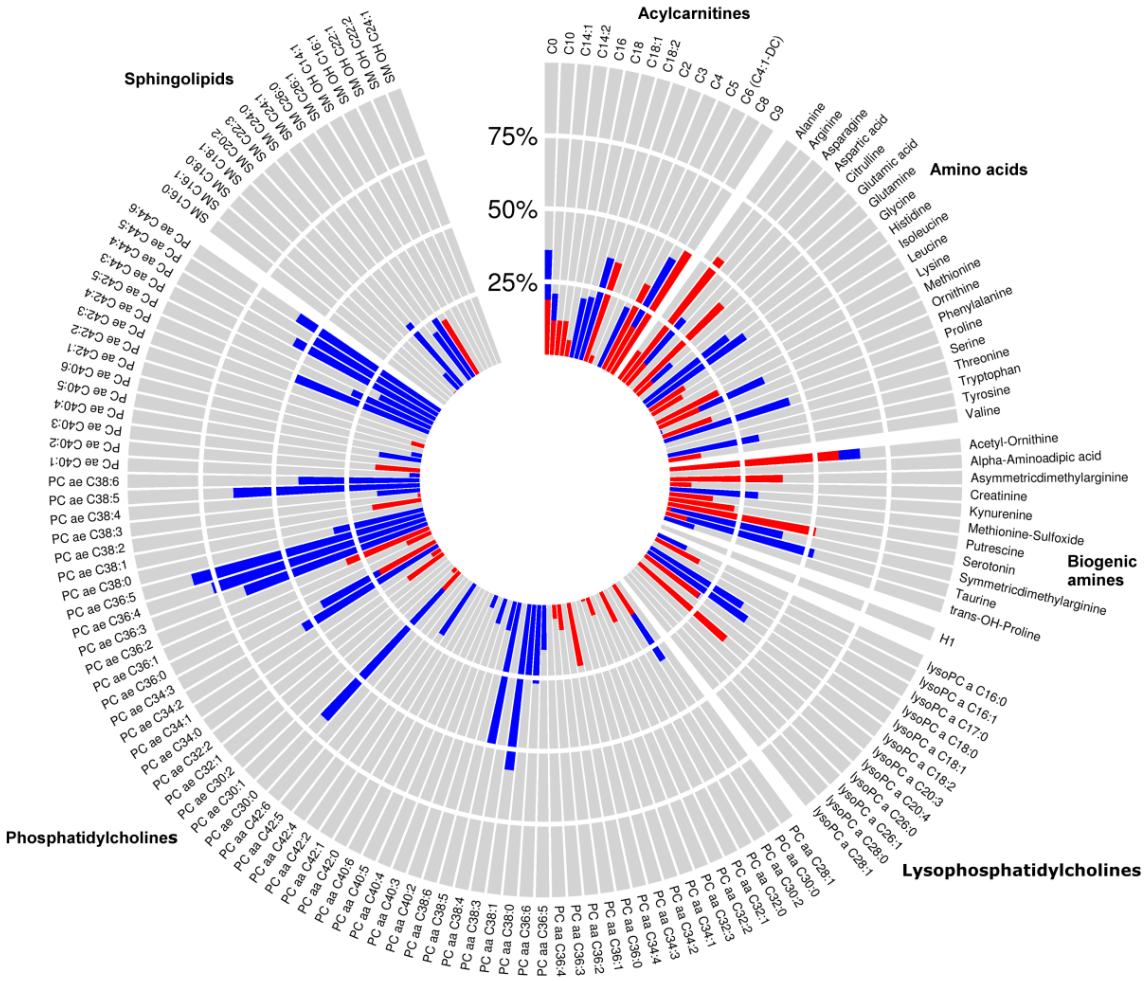
Table 5.3 Significant phenotypic, genetic, and environmental correlations between metabolite concentrations and cardiometabolic risk factors

Cardiometabolic risk factor	$\rho_P \pm SE$	*P*	$\rho_G \pm SE$	*P*	$\rho_E \pm SE$	*P*
			Citrulline			
ApoB100	0.29 ± 0.14	0.046*	1 ± 0	0.027**	-0.65 ± 1.32	0.33
			PC ae C36 :4			
HDL-C	0.32 ± 0.13	0.025*	1 ± 0	0.16	0.02 ± 0.52	0.97
TC	0.30 ± 0.14	0.047*	-0.32 ± 0.87	0.65	0.87 ± 0.26	0.083
			PC ae C36 :5			
HDL-C	0.42 ± 0.12	0.0020**	1 ± 0	0.13	0.22 ± 0.34	0.60
Insuline	-0.34 ± 0.14	0.021*	-0.63 ± 0.41	0.22	-0.05 ± 0.52	0.92
TC	0.39 ± 0.13	0.0070*	-0.98 ± 2.03	0.36	1 ± 0	4.29×10^{-6}**
			Serotonin			
CRP[1]	0.36 ± 0.12	0.0079*	1 ± 0	0.21	0.29 ± 0.33	0.42

Phenotypic Pearson correlation (ρ_P) is partitioned into genetic (ρ_G) and environmental (ρ_E) correlations. * Significant correlation *P*-value ≤ 0.05. ** Significant Bonferroni adjusted *P*-value.[1] Values are $\log_{10}$ transformed. Abbreviations: Acyl-alkyl (ae), Apolipoprotein B100 (ApoB100), C-reactive protein (CRP), High-density lipoprotein cholesterol (HDL-C), Phosphatidylcholine (PC), Standard error (SE), Total cholesterol (TC).

Figure

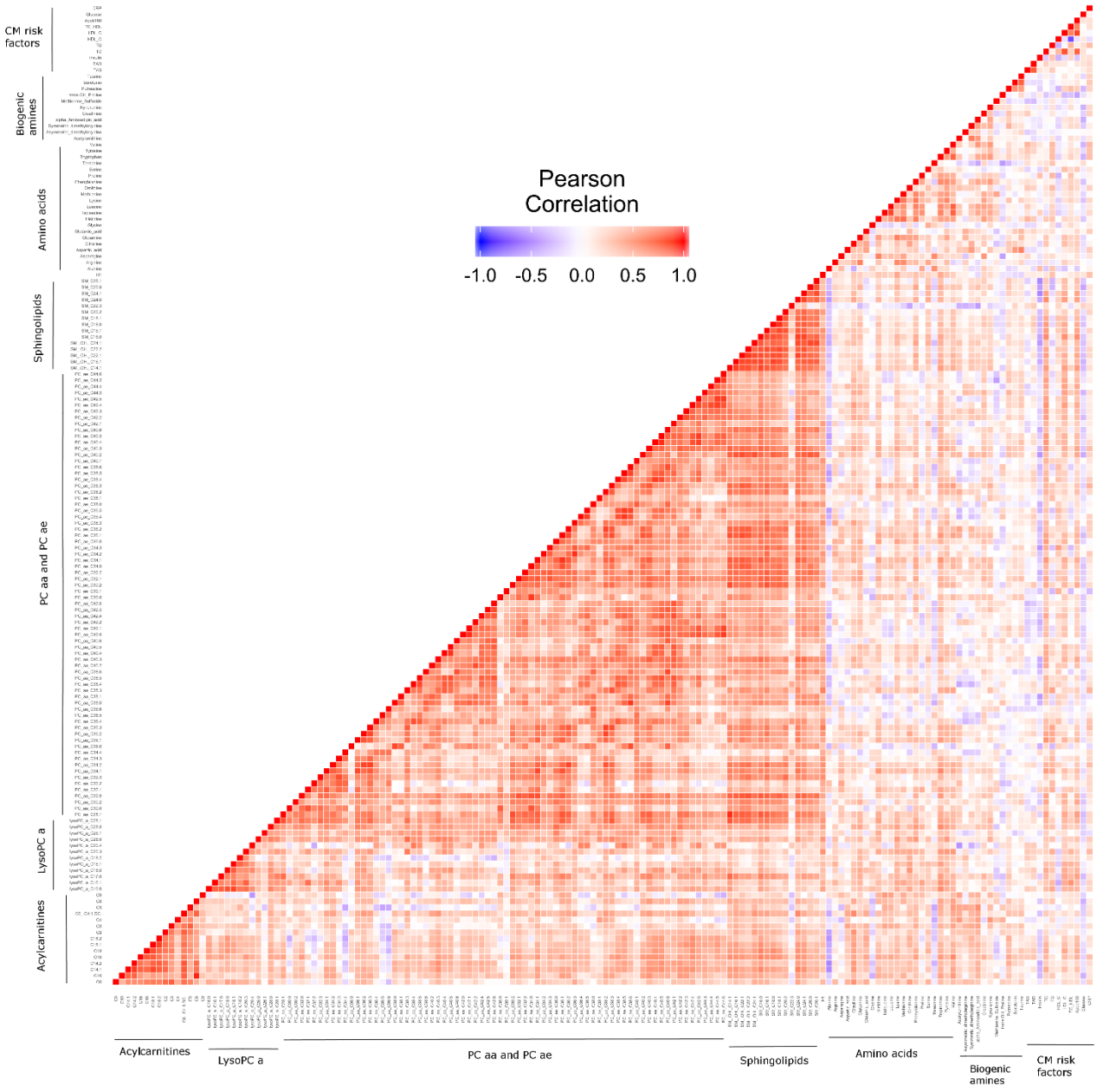
Figure 5.1 Polar histogram depicting genetic heritability, common environmental effect, and non-genetic effect of all 147 detected metabolites concentrations



Legend: Each bar represents a metabolite and is identified by its name. Metabolites are grouped according to compound classes (acylcarnitines, amino acids, biogenic amines, lysophosphatidylcholines, phosphatidylcholines, and sphingolipids). Each bar representing heritability estimates (%) is normalised to 100%. A white guide helps reading the values across the plot. It shows the 25%, 50%, and 75% marks. Genetic heritability is shown in blue, common environment effect in red, and non-shared effect in grey.

Supplemental material

Supplemental figure 5.1 Heat map of correlations between all 147 metabolite concentrations and nine cardiometabolic risk factors



Supplemental table 5.1 Plasma metabolite concentrations (μM) of all subjects (n=48)

Supplemental table 5.2 Heritability estimates of all 147 plasma metabolite concentrations from the full GCE model

Supplemental table 5.3 Reduced, full, and best fit models for all 147 metabolites

Supplemental table 5.4 Phenotypic correlations between all metabolite concentrations and cardiometabolic risk factors

Supplemental tables 5.1 to 5.4 are available at https://www.sciencedirect.com/science/article/pii/S0271531718304287.

Supplemental table 5.5 Phenotypic correlations between nine metabolite concentrations showing significant familial effect and nine cardiometabolic risk factors

Metabolites	CM risk factors	ρ_P	SE	*P*-value
Acetylornithine	Apob100	0.0722983	0.1423405	0.6085745
	TC	-0.163004	0.147727	0.2756465
	TC/HDL-C	0.1490176	0.1435544	0.3076067
	Glucose	-0.093791	0.1428945	0.5132212
	HDL-C	-0.183025	0.1403678	0.2007656
	Insulin	-0.123183	0.1508224	0.4159717
	CRP[1]	0.1492036	0.1442966	0.3140634
	LDL-C	-0.08774	0.1526005	0.5673551
	TG	0.0851733	0.1518015	0.5772669
Acylcarnitine C9	Apob100	0.1726208	0.1381148	0.2211607
	TC	0.1365959	0.1456885	0.3532371
	TC/HDL-C	0.1895406	0.1412407	0.1919217
	Glucose	-0.185055	0.141776	0.1988408
	HDL-C	-0.065813	0.1415598	0.6432707
	Insulin	-0.036034	0.1538918	0.8145277
	CRP[1]	0.0865489	0.1442288	0.5526698
	LDL-C	0.1446493	0.1502774	0.3448883
	TG	0.0736586	0.1519633	0.6296512
Arginine	Apob100	0.1987	0.1391175	0.1594633
	TC	0.1750995	0.1449961	0.2371087
	TC/HDL-C	0.1493842	0.1478887	0.3243208
	Glucose	0.0116742	0.1464255	0.9364165
	HDL-C	-0.061744	0.1483962	0.6797885
	Insulin	-0.136272	0.1600194	0.4275821
	CRP[1]	-0.031086	0	1
	LDL-C	0.2468178	0.1467952	0.1315148
	TG	-0.101027	0.1494208	0.501525
Asymmetric dimethyarginine	Apob100	0.0284998	0.142924	0.8420918
	TC	-0.02325	0.1487564	0.8758866
	TC/HDL-C	0.0213794	0.1478997	0.8852157
	Glucose	0.0686906	0.1434141	0.6326725
	HDL-C	-0.102768	0.1441231	0.4817363
	Insulin	-0.049837	0.1505738	0.7403469
	CRP[1]	0.0398973	0.1430346	0.7807171
	LDL-C	0.0190308	0.1533063	0.901194
	TG	-0.087534	0.1503881	0.5613582

Citrulline	Apob100	0.2901913	0.1409932	0.0456589*
	TC/HDL-C	0.2797867	0.1369528	0.0595066
	Glucose	-0.130614	0.1406995	0.3587381
	CRP[1]	0.0892467	0.1419404	0.5309568
	TG	-0.005004	0.157858	0.9746973
LysoPC a C20:4	Apob100	0.1951507	0.1391704	0.1845133
	TC	0.2625116	0.146913	0.0990106
	TC/HDL-C	0.1662271	0.1438035	0.2602197
	Glucose	-0.04697	0.1450677	0.7472058
	HDL-C	0.0221996	0.1426457	0.8762454
	Insulin	-0.065273	0.1562955	0.6794408
	CRP[1]	0.025616	0.1424817	0.8572442
	LDL-C	0.2638093	0.1485983	0.0984942
	TG	0.1010284	0.154645	0.5185371
PC ae C36:4	Apob100	0.113509	0.1508835	0.4606833
	TC	0.2962386	0.1370693	0.0467156*
	TC/HDL-C	-0.125063	0.1511259	0.4070174
	Glucose	-0.068477	0.1453043	0.6400919
	HDL-C	0.3165884	0.1325974	0.0246872*
	Insulin	-0.12385	0.1518338	0.4195369
	CRP[1]	0.1659485	0.1413926	0.2227886
	LDL-C	0.1353653	0.1553047	0.3905882
	TG	-0.077708	0.1549469	0.617433
PC ae C36:5	Apob100	0.1412063	0.1509593	0.3641746
	TC	0.3947486	0.1271087	0.0069362*
	TC/HDL-C	-0.20537	0.1438991	0.1606875
	Glucose	-0.07495	0.1447997	0.6075446
	HDL-C	0.4247471	0.1199526	0.0019727**
	Insulin	-0.339778	0.1362328	0.021106*
	LDL-C	0.2163063	0.1470725	0.159312
	TG	-0.232787	0.1497823	0.1277675
Serotonin	Apob100	0.1851871	0.1373761	0.1879075
	TC	0.036985	0.1572613	0.8132289
	TC/HDL-C	0.150447	0.1434043	0.3052704
	Glucose	-0.015071	0.1427112	0.9156872
	HDL-C	-0.074282	0.1430302	0.6061794
	Insulin	-0.194223	0.141863	0.1804804
	CRP[1]	0.3638891	0.1238703	0.0078692*
	LDL-C	0.1310939	0.1525779	0.3954716

	TG	-0.078637	0.1477009	0.5954364

* Significant correlation *P*-value ≤ 0.05. ** Significant Bonferroni adjusted *P*-value (0.05/9=0.0056).[1] Values are log10 transformed.

Abbreviations: Apolipoprotein B100 (ApoB100), Cardiometabolic (CM), C-reactive protein (CRP), High-density lipoprotein cholesterol (HDL-C), Low-density lipoprotein cholesterol (LDL-C), Phenotypic correlation (ρ_P), Standard error (SE), Total cholesterol (TC), Total cholesterol/HDL-C (TC/HDL-C), Triglycerides (TG).

Chapitre 6 Ressemblances familiales des caroténoïdes plasmatiques

Bénédicte L. Tremblay, Frédéric Guénard, Benoît Lamarche, Louis Pérusse, Marie-Claude Vohl

L'article présenté dans ce chapitre s'intitule : *Genetic and common environmental contributions to the familial resemblances in plasma carotenoid concentrations in healthy families.*

Cet article est publié dans la revue : *Nutrients. 2018 Jul 31;10(8):1002.*

Résumé

Introduction : Les caroténoïdes présentent une variabilité interindividuelle qui peut être due à des facteurs génétiques. Seule une étude a rapporté l'héritabilité de l'α-carotène et du β-carotène sériques, sans toutefois tenir compte de la composante environnementale. La présente étude vise à estimer la contribution des effets génétiques et de l'environnement partagé à la variance des concentrations de caroténoïdes, et à tester si leurs corrélations phénotypiques avec des facteurs de risque cardiométabolique s'expliquent par des effets génétiques et environnementaux partagés.

Méthodes : Les concentrations plasmatiques de caroténoïdes (α-carotène, β-carotène, β-cryptoxanthine, lutéine, lycopène, zéaxanthine et caroténoïdes totaux) ont été mesurées chez 48 sujets en santé. Les estimés d'héritabilité ont été calculés avec la méthode de décomposition de la variance.

Résultats : La lutéine et le lycopène ont démontré un effet familial significatif (p = 6 x 10^{-6} et 0,0043, respectivement). L'héritabilité maximale, l'héritabilité génétique et l'effet de l'environnement partagé ont été estimés pour la lutéine (88,3 %, 43,8 % et 44,5 %, respectivement) et le lycopène (45,2 %, 0 % et 45,2 %, respectivement). Des corrélations phénotypiques significatives entre les concentrations de caroténoïdes et les facteurs de risque cardiométabolique ont été obtenues pour la β-cryptoxanthine, le lycopène et la zéaxanthine.

Conclusions : Les ressemblances familiales des concentrations de lycopène étaient principalement attribuables à des effets de l'environnement partagé alors que pour les concentrations de lutéine, elles étaient attribuables à des effets génétiques et de l'environnement partagé. Des facteurs génétiques et environnementaux partagés peuvent influencer les caroténoïdes et les facteurs de risque cardiométabolique, mais des études supplémentaires sont nécessaires pour mieux comprendre l'impact potentiel sur le développement de la maladie.

Abstract

Introduction: Carotenoids have shown an interindividual variability that may be due to genetic factors. The only study that has reported heritability of serum α- and β-carotene has not considered the environmental component. This study aimed to estimate the contribution of both genetic and common environmental effects to the variance of carotenoid concentrations and to test whether their phenotypic correlations with cardiometabolic risk factors are explained by shared genetic and environmental effects.

Methods: Plasma carotenoid concentrations (α-carotene, β-carotene, β-cryptoxanthin, lutein, lycopene, zeaxanthin, and total carotenoids) of 48 healthy subjects were measured. Heritability estimates of carotenoid concentrations were calculated using the variance component method.

Results: Lutein and lycopene showed a significant familial effect ($p = 6 \times 10^{-6}$ and 0.0043, respectively). Maximal heritability, genetic heritability, and common environmental effect were computed for lutein (88.3%, 43.8%, and 44.5%, respectively) and lycopene (45.2%, 0%, and 45.2%, respectively). Significant phenotypic correlations between carotenoid concentrations and cardiometabolic risk factors were obtained for β-cryptoxanthin, lycopene, and zeaxanthin.

Conclusions: Familial resemblances in lycopene concentrations were mainly attributable to common environmental effects, while for lutein concentrations they were attributable to genetic and common environmental effects. Common genetic and environmental factors may influence carotenoids and cardiometabolic risk factors, but further studies are needed to better understand the potential impact on disease development.

Title page

Genetic and common environmental contributions to the familial resemblances in plasma carotenoid concentrations in healthy families

Bénédicte L. Tremblay, Frédéric Guénard, Benoît Lamarche, Louis Pérusse, and Marie-Claude Vohl*

Institute of Nutrition and Functional Foods (INAF), Laval University, 2440 Hochelaga Blvd, Quebec City, QC, G1V 0A6, Canada

Corresponding author*:

Marie-Claude Vohl, Ph.D.

Institute of Nutrition and Functional Foods (INAF), Laval University

Pavillon des Services, 2440 Hochelaga Blvd, Quebec City, QC G1V 0A6, Canada

Phone: 418 656-2131 ext. 4676; Fax: 418 656-5877

E-mail: marie-claude.vohl@fsaa.ulaval.ca

Keywords: β-cryptoxanthin; Bivariate analysis; Cardiometabolic risk factors; Carotenoids; Familial resemblances; French Canadians; Heritability; Lutein; Lycopene.

Introduction

Adoption of healthy lifestyle habits, including healthy eating, is the cornerstone of chronic disease prevention (1, 2). Fruits and vegetables (FAV) consumption is inversely associated with chronic diseases (3, 4). Methods used to evaluate food consumption such as 24-hour recalls, food frequency questionnaires, and food diaries rely on self-reported data, which can distort estimation of certain food intakes (5, 6). Although strategies have been developed to better manage random and mitigate systematic errors, it is not possible to completely eliminate bias from self-reporting dietary assessment methods (7).

In that regard, measuring biomarkers of food consumption in blood, urine and tissue may represent a more objective method to evaluate dietary intakes and patterns (8, 9). Carotenoids are a reliable biomarker of FAV consumption (10-12). Indeed, FAV provide over 90% of the daily carotenoid intake (13). Carotenoids are a family of more than 700 fat-soluble pigments, but α-carotene, β-carotene, β-cryptoxanthin, lutein, lycopene, and zeaxanthin represent over 95% of total circulating carotenoids in human plasma or serum (13, 14).

Inter-individual variability in circulating carotenoids has been observed and may be partly attributable to genetic factors that cause differences in the absorption, assimilation, distribution, metabolism, and excretion of carotenoids (15-17). Moreover, several genome-wide studies have identified genetic variants that influence circulating carotenoid concentrations (15, 18-20). Until now, only one study has reported heritability of serum carotenoids. The SAFARI study investigated genetics of serum carotenoids in Mexican American children (21). More specifically, it reported high genetic heritability of α-carotene and β-carotene along with phenotypic and genetic correlations of α-carotene and β-carotene with obesity-related traits (21). However, it did not account for the environmental component in its heritability estimates (21).

The aim of the present study was to estimate the contribution of both genetic and common environmental effects to the variance of plasma carotenoid concentrations (α-carotene, β-carotene, β-cryptoxanthin, lutein, lycopene, zeaxanthin, and total carotenoids) and to test whether their phenotypic correlations with traditional cardiometabolic (CM) risk factors could be explained by shared genetic and environmental effects. Thus, we tested the hypothesis that both genetic and common environmental effects contribute to the variance of plasma carotenoid concentrations and that shared genetic and environmental effects explain their phenotypic correlations with CM risk factors. To test this hypothesis, variance component method and bivariate genetic analysis were performed in a family-based sample of 48 French Canadians from 16 families. Our results highlighted that familial resemblances in lycopene concentrations were mainly attributable to common environmental effect, while for lutein concentrations, they were attributable to both genetic and common environmental effects. Common genetic and environmental factors seem to influence carotenoids and CM risk

factors but further studies are needed to better understand these possible influences and their potential impact on disease development.

Methods

Patients and design

A total of 48 Caucasian French-Canadian subjects from 16 families were recruited in the Greater Quebec City metropolitan area, in Canada, as part of the GENERATION Study, whose recruitment began in May 2011. The GENERATION Study was designed to evaluate familial resemblances in omics (DNA methylation (22) and gene expression (23)) and metabolic (metabolites and carotenoids) profiles in healthy families. Families were composed of 16 mothers, 6 fathers, and 26 children. The majority of parents had a university education and a family income > 80 000$ CAN. Inclusion criteria were that families living under the same roof comprise at least the mother and one child aged between 8 and 18. Parents had to be the biological parents of their child (or children), in good general health, with body mass index (BMI) ranging between 18 and 35 kg/m^2. The use of Synthroid® (levothyroxine) or oral contraceptive was tolerated. Children also had to be in good general health. Exclusion criteria were smoking, self-reported history of metabolic conditions requiring treatment, and use of psycho-stimulators [Ritalin® (methylphenidate), Concerta® (methylphenidate), and Strattera® (atomoxetine)]. Parents and children were asked to complete several dietary, physical activity, medical history, and pregnancy questionnaires under the supervision of a registered dietitian during their visit at the Institute of Nutrition and Functional Foods (INAF). The experimental protocol was approved by the Ethics Committees of Laval University Hospital Research Center and Laval University. All participants (adults and children) signed an informed consent document. Parental consent was also obtained by signing the child consent document.

Anthropometric and cardiometabolic measurements

Body weight, waist girth, and height were measured according to the procedures recommended by the Airlie Conference (24). Blood samples were collected from an antecubital vein into vacutainer tubes containing EDTA after 12-hour overnight fast and 48-hour alcohol abstinence. Plasma was separated by centrifugation (2500 g for 10 min at 4°C), and samples were aliquoted and frozen (-80°C) for subsequent analyses. Enzymatic assays were used to measure plasma total cholesterol (TC) and triglyceride (TG) concentrations (25, 26). Precipitation of very-low density lipoprotein (VLDL) and low-density lipoprotein (LDL) particles in the infranatant with heparin manganese chloride generated the high-density lipoprotein cholesterol (HDL-C) fraction (27). LDL cholesterol (LDL-C) was calculated with the Friedewald formula (28). Fasting glucose levels were enzymatically measured (29). Using a sensitive assay, plasma C-reactive protein (CRP) was measured by nephelometry (Behring Latex-Enhanced on the Behring Nephelometer BN-100; Prospec equipment, Behring Diagnostic, Westwood, MA, USA) (30).

Carotenoid measurements

Carotenoid standards were purchased from Sigma (Oakville, Ontario, Canada). Stock solutions for each carotenoid were prepared (1 mg in 100 mL of solvent) in ethanol for β-cryptoxanthin, lutein, and zeaxanthin and

in hexane for β-carotene and lycopene. The concentration of each stock solution was determined using a UV spectrophotometer and the specific molecular extinction coefficient of the molecule (31). Carotenoid standards were solubilized with methanol/dichloromethane (65/35, v/v) for a final concentration of 2 µM. These solutions were then used to perform calibration curves. Retinol acetate (15 µM) was used as an internal standard.

Plasma samples were thawed a few hours before analysis. A total of 100 µL of plasma, 20 µL of 2-propanol, and 20 µL of internal standard were transferred in Eppendorf tubes. Samples were transferred on a 400µL fixed well plate (ISOLUTE® SLE+, Biotage, Charlotte, NC) and eluted with 1800 µL of hexane:isopropanol (90/10, *v/v*) in each well. Each eluted sample was evaporated under nitrogen and reconstituted with 300 µL of methanol:dichloromethane (65/35, *v/v*). Plates were shaken for 10 minutes and samples were transferred into high performance liquid chromatography glass vials to be analyzed.

High performance liquid chromatography (HPLC)-UV analysis was performed using an Agilent 1260 liquid handling system (Agilent, Mississauga, Ontario, Canada) equipped with a binary pump system and a C30 reversed phase column (YMC America Inc. Allentown, PA) kept at constant temperature (35°C). Carotenoids were separated with a mobile phase consisting of methanol:water (98/2, *v/v*; Eluent A) and methyl-tert-butyl ether (MTBE; Eluent B; VWR, Mississauga, Ontario, Canada). Injection volume was set at 40µL and the flow-rate was set at 1 mL/min and the gradient elution was as follows: 2% Eluent B (initial), 2.0-80% Eluent B (0.0-27.0 min), isocratic 80% Eluent B (27.0-31.0 min), 80.0-2.0% Eluent B (31.0-31.1 min), and isocratic 2% Eluent B (31.1-34.0 min). UV detector was set at 450 nm and identification of each compound was confirmed using retention time and UV spectra (190-640nm) of the pure compounds. Data acquisition was carried out with the Chemstation software (Agilent, Mississauga, Ontario, Canada). For all carotenoids the concentrations are reported in µmol/L of plasma. Value considered outlier, defined as value falling outside of the mean ± 4 standard deviations (1 outlier in β-crypoxanthin) was excluded from heritability analyses.

Statistical analysis

Statistical Analysis Software (SAS) was used to compute differences in CM parameters [TC, LDL-C, HDL-C, TC/HDL-C, TG, apolipoprotein B100 (apoB100), glucose, insulin, systolic blood pressure (SBP), diastolic blood pressure (DBP), and CRP] between fathers and mothers, and between daughters and sons using an unpaired t-test. Variables not normally distributed were log10 transformed before analyses. For heritability analysis, adjustments were made for the effects of sex, age, and categories of BMI (underweight, normal, overweight, and obese) using a standard least squares model in JMP software v12. Categories of BMI were used to compare BMI among parents with BMI percentile among children. Cut-offs for BMI and BMI percentile were both from the World Health Organization (32, 33). Residuals from this model were used to compute heritability estimates using the variance component method implemented in QTDT v2.6.1 (34). We used a full general model in which the

variance in concentration of each carotenoid was partitioned into polygenic effects (Vg), common environmental effects shared by family members (Vc), and non-shared environmental effects unique to each individual (Ve). We tested this full general model against a null model of no familial resemblance in which Vg = Vc = 0. We then computed average maximal heritability, as the proportion of variance accounted by genetic and common environmental effects ((Vg+Vc)/(Vg+Vc+Ve)), average genetic heritability, as the proportion of variance accounted by genetics effects (Vg/(Vg+Vc+Ve)), and common environmental effect, as the proportion of variance accounted by common environmental effects (Vc/(Vg+Vc+Ve)). For comparison purposes, we computed an alternative genetic model in which the variance in carotenoid concentrations was partitioned into Vg and Ve. We then computed average genetic heritability, as the proportion of variance accounted by genetic effects (Vg/(Vg+Ve)). Phenotypic Pearson correlations (ρP) between seven carotenoid concentrations and 11 CM risk factors were calculated. Once again carotenoids concentrations were adjusted for the effects of sex, age, and categories of BMI. Significant phenotypic correlations (ρP) between pairs of traits have been portioned into genetic (ρG) and environmental (ρE) correlations using a bivariate genetic analysis. Analyses were performed using SOLAR Eclipse version 7.6.4. Bonferroni corrections were used to account for multiple testing in phenotypic, genetic, and environmental correlations.

Results

Characteristics of study participants

Characteristic of study participants are presented in Table 1. Fathers and mothers had significant differences in HDL-C, TC/HDL-C, and SBP, whereas glucose concentrations were significantly different between boys and girls. Concentrations of all six carotenoids (α-carotene, β-carotene, β-cryptoxanthin, lutein, lycopene, and zeaxanthin) and total carotenoids measured in the fasting state are presented in Supplementary Table 1.

Heritability analysis of plasma carotenoid concentrations

Maximal heritability, genetic heritability, and common environmental effect from the full general model for the seven measurements of carotenoids are presented in Table 2. A significant familial effect was observed for lutein and lycopene (*P*-values of $6x10^{-5}$ and $4.3x10^{-3}$, respectively). Plasma lutein concentrations had a maximal heritability of 88.32%, a genetic heritability of 43.82%, and a common environmental effect of 44.50%, whereas for plasma lycopene concentrations, the maximal heritability was of 45.23%, the genetic heritability of 0%, and the common environmental effect of 45.23%.

Bivariate genetic analysis between carotenoids and cardiometabolic risk factors

Phenotypic correlations between seven measurements of carotenoids (α-carotene, β-carotene, β-cryptoxanthin, lutein, lycopene, zeaxanthin, and total carotenoids) and 11 CM risk factors (TC, LDL-C, HDL-C, TC/HDL-C, TG, apolipoprotein B100 (apoB100), glucose, insulin, SBP, DBP, and CRP) were calculated (Supplementary Table 2). Significant phenotypic correlations ($P \leq 0.05$) were obtained for β-cryptoxanthin, lycopene, and zeaxanthin (Table 3). Positive phenotypic correlation between lycopene and DBP (r=0.44) remained significant after Bonferroni correction (adjusted P=0.05/11=0.0045). Phenotypic correlations were partitioned into genetic and environmental correlations using bivariate genetic analysis. Significant genetic and environmental correlations ($P \leq 0.05$) were obtained for lycopene and β-cryptoxanthin, respectively (Table 3). Indeed, lycopene had a significant genetic correlation with DBP (r=1), while β-cryptoxanthin had a significant environmental correlation with CRP (r=0.61) (Table 3). Genetic and environmental correlations also remained significant after Bonferroni correction (adjusted P=0.05/2=0.025).

Discussion

As stated previously, the aim of the present study was, first, to estimate the contribution of both genetic and common environmental effects to the variance of plasma carotenoid concentrations (α-carotene, β-carotene, β-cryptoxanthin, lutein, lycopene, zeaxanthin, and total carotenoids). Familial resemblances in lutein concentrations were due to both genetic and common environmental effects, while only common environmental effects contributed to familial resemblances in lycopene concentrations. Lutein is a robust biomarker of FAV consumption (12). Several studies, mostly based on self-reported data, also linked FAV intakes and plasma lutein concentrations (11, 35, 36), which is interesting considering that lutein had the strongest maximal heritability in the present study. This suggests that plasma lutein, a biomarker of consumption of FAV, is heritable. Lycopene may have potential benefits against chronic diseases such as cardiovascular diseases due to its anti-atherogenic effect (13, 37-39). Not all studies have reported the relationship between circulating lycopene and FAV consumption (35, 40, 41). Nevertheless, a considerable maximal heritability of lycopene concentrations was reported in the present study. This suggests a significant familial effect in plasma lycopene concentrations of healthy subjects, which may be related to chronic diseases.

Farook et al. reported highly significant heritability estimates of α-carotene and β-carotene in children (21). They used an alternative genetic model that determines which proportion of the phenotypic variance is attributable to additive genetic effects (genetic heritability). Using the same model in the present study, all genetic heritability estimates increased (Supplementary Table 3). The genetic heritability of α-carotene and β-carotene were barely significant, but lutein and lycopene still had highly significant genetic effects. The inflation in heritability estimates in the alternative genetic model may be due to the inclusion of variance due to common environmental effect with the variance due to additive genetic effects, which has been acknowledged in Farook et al (21). Moreover, Gueguen et al. reported a significant familial effect for serum retinol and α-tocopherol concentrations in 387 healthy French families (42). Additive genetic and shared common environment effects for retinol (30.5% and 14.2%, respectively) and α-tocopherol were reported (22.1% and 18.7%, respectively) (42). The results of the present study also shed light on the contribution of common environmental effect to the familial resemblances in plasma lutein and lycopene concentrations, but the estimates are slightly higher than what was found by Gueguen et al. (42).

The aim of the present study was also, secondly, to test phenotypic, genetic, and environmental correlations between carotenoid concentrations and 11 traditional CM risk factors. There were significant phenotypic correlations between β-cryptoxanthin, lycopene, as well as zeaxanthin and CM risk factors. β-cryptoxanthin had a significant positive environmental correlation with CRP, indicating that common environmental factors affect traits in the same way (increase or decrease concentrations). Epidemiologic studies have reported an inverse correlation between β-cryptoxanthin and CRP concentrations (43-46). However, studies including intervention

trials reported positive or no association between carotenoids and CRP (47-49). A systematic review and meta-analysis by Cheng et al. reported weak evidence that interventions with supplements of lycopene had an effect on plasma CRP concentrations (50). Accordingly, a carotenoid component score accounted for only 3.3% of the total variance in serum CRP concentrations (51). Interestingly, the relationship between β-cryptoxanthin and CRP concentrations could also be explained by TG concentrations, which are correlated with β-cryptoxanthin in the present study. Other studies have reported a correlation between CRP and TG concentrations (52, 53) and a significant correlation was also observed in the present study (r=0.34 *P*=0.02). Thus, the correlation between β-cryptoxanthin and CRP concentrations may be explained through β-cryptoxanthin correlation with TG concentrations, which is concordant with the environmental correlation reported in bivariate analysis.

Moreover, lycopene had a strong significant positive genetic correlation with DBP, suggesting that shared genetic factors affecting both lycopene concentrations and DBP act in the same way. Studies have reported beneficial (54-56), neutral (57, 58), and even detrimental (59, 60) effects of lycopene on blood pressure. According to Cheng et al., there was only weak evidence that lycopene intervention had an effect on DBP (50). According to a recent review by Thies et al., there are not enough studies to draw substantial conclusions on the effects of lycopene on blood pressure (39). Interestingly, the strong genetic correlation suggesting shared genetic factors affecting both lycopene concentrations and DBP may be plausible since *SETD7* has been associated with both traits (15, 61). Indeed, a genetic variant (rs7680948) within *SETD7*, encoding for SET domain containing lysine methyltransferase 7, was associated with serum lycopene concentrations in Amish adults (15). *SETD7* was also associated with DBP response and may play a role in high glucose-induced vascular dysfunction (61, 62). Genetic variants and changes in expression of genes involved in lycopene absorption and metabolism may also impact CM risk factors including blood pressure but this area still needs to be further studied (63).

The present study has strengths, but also some limitations. The main strength results from the study of six predominant plasma carotenoids. To the best of our knowledge, this is the first study that computed heritability and bivariate genetic analysis of so many carotenoids. The calculation of genetic and environmental correlations adds important information about the additive genetic and environmental effects that are shared between carotenoids and CM risk factors. Another strength derives from the adjustments for age, sex, and categories of BMI that were made in the heritability and correlation analyses. Several studies have reported sex differences in lutein and β-cryptoxanthin concentrations (12, 64, 65). Body weight may also interfere with circulating carotenoid concentrations since these may be accumulated in adipose tissue (66, 67). Age may also impacts circulating carotenoids (17, 68). Moreover, Bonferroni corrections were made to account for multiple testing. On the other hand, the study's main limitation resides in the small sample size that limits the statistical power to detect significant heritability estimates and significant correlations between traits. This may also explain why we

observed strong correlation of r=1 in the bivariate analyses. The number of subjects may limit the accurate quantification of correlations between traits. The results should therefore be interpreted with caution. The study of a founder population with relatively homogeneous genetics and shared environment is a new aspect in this field (69). However, this limits the generalization of results to other populations. Moreover, circulating carotenoids are also associated with blood lipid profile: lower TC, LDL-C, and HDL-C concentrations are associated with lower circulating carotenoids (49, 70). We did not adjust carotenoid concentrations for lipid profile but adjustments for sex, body weight, and age may partly account for this effect since men and women depict different lipid profiles. Finally, our study did not account for diet, physical activity, smoking, and alcohol consumption of participants, all of which may affect circulating carotenoid concentrations (71, 72). Despite that we have data on diet and physical activity, we decided not to use them considering the bias related to self-reporting dietary assessment methods (6) and the fact that there is a lack of data on the link between physical activity and plasma carotenoids, especially in children (17). Moreover, we took into account a limited number of objective confounders (age, sex, and BMI) in order to prevent the reduction of the statistical power to detect significant heritability estimates and correlations.

In conclusion, familial resemblances in lycopene concentrations were mainly attributable to common environmental effect. Regarding lutein concentrations, familial resemblances were attributable to both genetic and common environmental effects. To the best of our knowledge, this is the first study to report on the contribution of both genetic and common environmental effects to the variance of six predominant plasma carotenoids in healthy families. Common genetic and environmental factors seem to influence carotenoids and CM risk factors, but further studies are needed to better understand these possible influences and their potential impact on disease development.

Author Contributions

Conceptualization, Benoît Lamarche and Marie-Claude Vohl; Data curation, Bénédicte L. Tremblay, Frédéric Guénard and Louis Pérusse; Formal analysis, Bénédicte L. Tremblay, Frédéric Guénard and Louis Pérusse; Funding acquisition, Marie-Claude Vohl; Investigation, Benoît Lamarche and Marie-Claude Vohl; Methodology, Bénédicte L. Tremblay, Frédéric Guénard and Marie-Claude Vohl; Project administration, Marie-Claude Vohl; Resources, Louis Pérusse and Marie-Claude Vohl; Software, Bénédicte L. Tremblay, Frédéric Guénard and Louis Pérusse; Supervision, Marie-Claude Vohl; Visualization, Bénédicte L. Tremblay; Writing – original draft, Bénédicte L. Tremblay; Writing – review & editing, Bénédicte L. Tremblay, Frédéric Guénard, Benoît Lamarche, Louis Pérusse and Marie-Claude Vohl.

Funding

This work was supported by the Canada Research Chair in Genomics Applied to Nutrition and Metabolic Health.

Acknowledgments

We would like to acknowledge the contribution of Véronique Richard for the measurements of plasma carotenoid concentrations and the contribution of Christian Couture for statistical analysis. We also thank Véronique Garneau who contributed to the success of this study.

Conflicts of Interest

The authors declare no conflict of interest.

References

1. Eckel RH, Jakicic JM, Ard JD, de Jesus JM, Houston Miller N, Hubbard VS, et al. 2013 AHA/ACC guideline on lifestyle management to reduce cardiovascular risk: a report of the American College of Cardiology/American Heart Association Task Force on Practice Guidelines. Circulation. 2014;129(25 Suppl 2):S76-99.

2. Anderson TJ, Gregoire J, Hegele RA, Couture P, Mancini GB, McPherson R, et al. 2012 update of the Canadian Cardiovascular Society guidelines for the diagnosis and treatment of dyslipidemia for the prevention of cardiovascular disease in the adult. The Canadian journal of cardiology. 2013;29(2):151-67.

3. Lim SS, Vos T, Flaxman AD, Danaei G, Shibuya K, Adair-Rohani H, et al. A comparative risk assessment of burden of disease and injury attributable to 67 risk factors and risk factor clusters in 21 regions, 1990-2010: a systematic analysis for the Global Burden of Disease Study 2010. Lancet. 2012;380(9859):2224-60.

4. Boeing H, Bechthold A, Bub A, Ellinger S, Haller D, Kroke A, et al. Critical review: vegetables and fruit in the prevention of chronic diseases. European journal of nutrition. 2012;51(6):637-63.

5. Subar AF. Developing dietary assessment tools. Journal of the American Dietetic Association. 2004;104(5):769-70.

6. Archer E, Blair SN. Implausible data, false memories, and the status quo in dietary assessment. Advances in nutrition. 2015;6(2):229-30.

7. Labonte ME, Kirkpatrick SI, Bell RC, Boucher BA, Csizmadi I, Koushik A, et al. Dietary assessment is a critical element of health research - Perspective from the Partnership for Advancing Nutritional and Dietary Assessment in Canada. Applied physiology, nutrition, and metabolism = Physiologie appliquee, nutrition et metabolisme. 2016:1-4.

8. Heinzmann SS, Brown IJ, Chan Q, Bictash M, Dumas ME, Kochhar S, et al. Metabolic profiling strategy for discovery of nutritional biomarkers: proline betaine as a marker of citrus consumption. The American journal of clinical nutrition. 2010;92(2):436-43.

9. Baldrick FR, Woodside JV, Elborn JS, Young IS, McKinley MC. Biomarkers of fruit and vegetable intake in human intervention studies: a systematic review. Critical reviews in food science and nutrition. 2011;51(9):795-815.

10. Scalbert A, Brennan L, Manach C, Andres-Lacueva C, Dragsted LO, Draper J, et al. The food metabolome: a window over dietary exposure. The American journal of clinical nutrition. 2014;99(6):1286-308.

11. Souverein OW, de Vries JH, Freese R, Watzl B, Bub A, Miller ER, et al. Prediction of fruit and vegetable intake from biomarkers using individual participant data of diet-controlled intervention studies. The British journal of nutrition. 2015;113(9):1396-409.

12. Couillard C, Lemieux S, Vohl MC, Couture P, Lamarche B. Carotenoids as biomarkers of fruit and vegetable intake in men and women. The British journal of nutrition. 2016;116(7):1206-15.

13. Maiani G, Caston MJ, Catasta G, Toti E, Cambrodon IG, Bysted A, et al. Carotenoids: actual knowledge on food sources, intakes, stability and bioavailability and their protective role in humans. Molecular nutrition & food research. 2009;53 Suppl 2:S194-218.

14. Eroglu A, Harrison EH. Carotenoid metabolism in mammals, including man: formation, occurrence, and function of apocarotenoids. Journal of lipid research. 2013;54(7):1719-30.

15. D'Adamo CR, D'Urso A, Ryan KA, Yerges-Armstrong LM, Semba RD, Steinle NI, et al. A Common Variant in the SETD7 Gene Predicts Serum Lycopene Concentrations. Nutrients. 2016;8(2):82.

16. Borel P, de Edelenyi FS, Vincent-Baudry S, Malezet-Desmoulin C, Margotat A, Lyan B, et al. Genetic variants in BCMO1 and CD36 are associated with plasma lutein concentrations and macular pigment optical density in humans. Annals of medicine. 2011;43(1):47-59.

17. Bohn T, Desmarchelier C, Dragsted LO, Nielsen CS, Stahl W, Ruhl R, et al. Host-related factors explaining interindividual variability of carotenoid bioavailability and tissue concentrations in humans. Molecular nutrition & food research. 2017;61(6).

18. Beydoun MA, Nalls MA, Canas JA, Evans MK, Zonderman AB. Gene polymorphisms and gene scores linked to low serum carotenoid status and their associations with metabolic disturbance and depressive symptoms in African-American adults. The British journal of nutrition. 2014;112(6):992-1003.

19. Ferrucci L, Perry JR, Matteini A, Perola M, Tanaka T, Silander K, et al. Common variation in the beta-carotene 15,15'-monooxygenase 1 gene affects circulating levels of carotenoids: a genome-wide association study. American journal of human genetics. 2009;84(2):123-33.

20. Zubair N, Kooperberg C, Liu J, Di C, Peters U, Neuhouser ML. Genetic variation predicts serum lycopene concentrations in a multiethnic population of postmenopausal women. The Journal of nutrition. 2015;145(2):187-92.

21. Farook VS, Reddivari L, Mummidi S, Puppala S, Arya R, Lopez-Alvarenga JC, et al. Genetics of serum carotenoid concentrations and their correlation with obesity-related traits in Mexican American children. The American journal of clinical nutrition. 2017;106(1):52-8.

22. Tremblay BL, Guenard F, Lamarche B, Perusse L, Vohl MC. Familial resemblances in blood leukocyte DNA methylation levels. Epigenetics. 2016;11(11):831-8.

23. Tremblay BL, Guenard F, Lamarche B, Perusse L, Vohl MC. Familial resemblances in human whole blood transcriptome. BMC genomics. 2018;19(1):300.

24. Callaway C, Chumlea W, Bouchard C, Himes J, Lohman T, Martin A, et al. Standardization of anthropomeric measurements : The Airlie (VA) Consensus Conference. Champaign, IR, USA; 1988.

25. McNamara JR, Schaefer EJ. Automated enzymatic standardized lipid analyses for plasma and lipoprotein fractions. ClinChimActa. 1987;166(1):1-8.

26. Burstein M, Samaille J. [On a rapid determination of the cholesterol bound to the serum alpha- and beta-lipoproteins]. ClinChimActa. 1960;5:609.

27. Albers JJ, Warnick GR, Wiebe D, King P, Steiner P, Smith L, et al. Multi-laboratory comparison of three heparin-Mn2+ precipitation procedures for estimating cholesterol in high-density lipoprotein. ClinChem. 1978;24(6):853-6.

28. Friedewald WT, Levy RI, Fredrickson DS. Estimation of the concentration of low-density lipoprotein cholesterol in plasma, without use of the preparative ultracentrifuge. ClinChem. 1972;18(6):499-502.

29. Richterich R, Dauwalder H. [Determination of plasma glucose by hexokinase-glucose-6-phosphate dehydrogenase method]. Schweizerische medizinische Wochenschrift. 1971;101(17):615-8.

30. Pirro M, Bergeron J, Dagenais GR, Bernard PM, Cantin B, Despres JP, et al. Age and duration of follow-up as modulators of the risk for ischemic heart disease associated with high plasma C-reactive protein levels in men. ArchInternMed. 2001;161(20):2474-80.

31. Gleize B, Steib M, Andre M, Reboul E. Simple and fast HPLC method for simultaneous determination of retinol, tocopherols, coenzyme Q(10) and carotenoids in complex samples. Food chemistry. 2012;134(4):2560-4.

32. World Health Organization G. Obesity: Preventing and managing the global epidemic. Report of a WHO Consultation. Geneva: World Health Organization; 2000.

33. de Onis M, Lobstein T. Defining obesity risk status in the general childhood population: which cut-offs should we use? International journal of pediatric obesity : IJPO : an official journal of the International Association for the Study of Obesity. 2010;5(6):458-60.

34. Abecasis GR, Cardon LR, Cookson WO. A general test of association for quantitative traits in nuclear families. American journal of human genetics. 2000;66(1):279-92.

35. Daniels JA, Mulligan C, McCance D, Woodside JV, Patterson C, Young IS, et al. A randomised controlled trial of increasing fruit and vegetable intake and how this influences the carotenoid concentration and activities of PON-1 and LCAT in HDL from subjects with type 2 diabetes. Cardiovascular diabetology. 2014;13:16.

36. Brevik A, Andersen LF, Karlsen A, Trygg KU, Blomhoff R, Drevon CA. Six carotenoids in plasma used to assess recommended intake of fruits and vegetables in a controlled feeding study. European journal of clinical nutrition. 2004;58(8):1166-73.

37. Wang XD. Lycopene metabolism and its biological significance. The American journal of clinical nutrition. 2012;96(5):1214S-22S.

38. Agudo A, Cabrera L, Amiano P, Ardanaz E, Barricarte A, Berenguer T, et al. Fruit and vegetable intakes, dietary antioxidant nutrients, and total mortality in Spanish adults: findings from the Spanish cohort of the European Prospective Investigation into Cancer and Nutrition (EPIC-Spain). The American journal of clinical nutrition. 2007;85(6):1634-42.

39. Thies F, Mills LM, Moir S, Masson LF. Cardiovascular benefits of lycopene: fantasy or reality? The Proceedings of the Nutrition Society. 2017;76(2):122-9.

40. Crane TE, Kubota C, West JL, Kroggel MA, Wertheim BC, Thomson CA. Increasing the vegetable intake dose is associated with a rise in plasma carotenoids without modifying oxidative stress or inflammation in overweight or obese postmenopausal women. The Journal of nutrition. 2011;141(10):1827-33.

41. Carlsen MH, Karlsen A, Lillegaard IT, Gran JM, Drevon CA, Blomhoff R, et al. Relative validity of fruit and vegetable intake estimated from an FFQ, using carotenoid and flavonoid biomarkers and the method of triads. The British journal of nutrition. 2011;105(10):1530-8.

42. Gueguen S, Leroy P, Gueguen R, Siest G, Visvikis S, Herbeth B. Genetic and environmental contributions to serum retinol and alpha-tocopherol concentrations: the Stanislas Family Study. The American journal of clinical nutrition. 2005;81(5):1034-44.

43. Kritchevsky SB, Bush AJ, Pahor M, Gross MD. Serum carotenoids and markers of inflammation in nonsmokers. American journal of epidemiology. 2000;152(11):1065-71.

44. Ford ES, Liu S, Mannino DM, Giles WH, Smith SJ. C-reactive protein concentration and concentrations of blood vitamins, carotenoids, and selenium among United States adults. European journal of clinical nutrition. 2003;57(9):1157-63.

45. Eroglu A, Schulze KJ, Yager J, Cole RN, Christian P, Nonyane BAS, et al. Plasma proteins associated with circulating carotenoids in Nepalese school-aged children. Archives of biochemistry and biophysics. 2018;646:153-60.

46. Schulze KJ, Christian P, Wu LS, Arguello M, Cui H, Nanayakkara-Bind A, et al. Micronutrient deficiencies are common in 6- to 8-year-old children of rural Nepal, with prevalence estimates modestly affected by inflammation. The Journal of nutrition. 2014;144(6):979-87.

47. Williams EJ, Baines KJ, Berthon BS, Wood LG. Effects of an Encapsulated Fruit and Vegetable Juice Concentrate on Obesity-Induced Systemic Inflammation: A Randomised Controlled Trial. Nutrients. 2017;9(2).

48. Scheurig AC, Thorand B, Fischer B, Heier M, Koenig W. Association between the intake of vitamins and trace elements from supplements and C-reactive protein: results of the MONICA/KORA Augsburg study. European journal of clinical nutrition. 2008;62(1):127-37.

49. Wang L, Gaziano JM, Norkus EP, Buring JE, Sesso HD. Associations of plasma carotenoids with risk factors and biomarkers related to cardiovascular disease in middle-aged and older women. The American journal of clinical nutrition. 2008;88(3):747-54.

50. Cheng HM, Koutsidis G, Lodge JK, Ashor A, Siervo M, Lara J. Tomato and lycopene supplementation and cardiovascular risk factors: A systematic review and meta-analysis. Atherosclerosis. 2017;257:100-8.

51. Wood AD, Strachan AA, Thies F, Aucott LS, Reid DM, Hardcastle AC, et al. Patterns of dietary intake and serum carotenoid and tocopherol status are associated with biomarkers of chronic low-grade systemic inflammation and cardiovascular risk. The British journal of nutrition. 2014;112(8):1341-52.

52. Firdous S. Correlation of CRP, fasting serum triglycerides and obesity as cardiovascular risk factors. Journal of the College of Physicians and Surgeons--Pakistan : JCPSP. 2014;24(5):308-13.

53. Löwbeer C, Mårtensson, L., Berg, E. and Wallinder, H. . Associations between C-Reactive Protein and Apolipoproteins, Lipoprotein (a) and Conventional Serum Lipids in Outpatients: Correlations and Time Trends. Open Journal of Clinical Diagnostics. 2015;5:33-40.

54. John JH, Ziebland S, Yudkin P, Roe LS, Neil HA, Oxford F, et al. Effects of fruit and vegetable consumption on plasma antioxidant concentrations and blood pressure: a randomised controlled trial. Lancet. 2002;359(9322):1969-74.

55. Engelhard YN, Gazer B, Paran E. Natural antioxidants from tomato extract reduce blood pressure in patients with grade-1 hypertension: a double-blind, placebo-controlled pilot study. American heart journal. 2006;151(1):100.

56. Kim JY, Paik JK, Kim OY, Park HW, Lee JH, Jang Y, et al. Effects of lycopene supplementation on oxidative stress and markers of endothelial function in healthy men. Atherosclerosis. 2011;215(1):189-95.

57. Hozawa A, Jacobs DR, Jr., Steffes MW, Gross MD, Steffen LM, Lee DH. Circulating carotenoid concentrations and incident hypertension: the Coronary Artery Risk Development in Young Adults (CARDIA) study. Journal of hypertension. 2009;27(2):237-42.

58. Itsiopoulos C, Brazionis L, Kaimakamis M, Cameron M, Best JD, O'Dea K, et al. Can the Mediterranean diet lower HbA1c in type 2 diabetes? Results from a randomized cross-over study. Nutrition, metabolism, and cardiovascular diseases : NMCD. 2011;21(9):740-7.

59. Paterson E, Gordon MH, Niwat C, George TW, Parr L, Waroonphan S, et al. Supplementation with fruit and vegetable soups and beverages increases plasma carotenoid concentrations but does not alter markers of oxidative stress or cardiovascular risk factors. The Journal of nutrition. 2006;136(11):2849-55.

60. Street DA, Comstock GW, Salkeld RM, Schuep W, Klag MJ. Serum antioxidants and myocardial infarction. Are low levels of carotenoids and alpha-tocopherol risk factors for myocardial infarction? Circulation. 1994;90(3):1154-61.

61. Magvanjav O, Gong Y, McDonough CW, Chapman AB, Turner ST, Gums JG, et al. Genetic Variants Associated With Uncontrolled Blood Pressure on Thiazide Diuretic/beta-Blocker Combination Therapy in the PEAR (Pharmacogenomic Evaluation of Antihypertensive Responses) and INVEST (International Verapamil-SR Trandolapril Study) Trials. Journal of the American Heart Association. 2017;6(11).

62. Paneni F, Costantino S, Battista R, Castello L, Capretti G, Chiandotto S, et al. Adverse epigenetic signatures by histone methyltransferase Set7 contribute to vascular dysfunction in patients with type 2 diabetes mellitus. Circulation Cardiovascular genetics. 2015;8(1):150-8.

63. Moran NE, Erdman JW, Jr., Clinton SK. Complex interactions between dietary and genetic factors impact lycopene metabolism and distribution. Archives of biochemistry and biophysics. 2013;539(2):171-80.

64. Jansen MC, Van Kappel AL, Ocke MC, Van 't Veer P, Boshuizen HC, Riboli E, et al. Plasma carotenoid levels in Dutch men and women, and the relation with vegetable and fruit consumption. European journal of clinical nutrition. 2004;58(10):1386-95.

65. Olmedilla B, Granado F, Blanco I, Rojas-Hidalgo E. Seasonal and sex-related variations in six serum carotenoids, retinol, and alpha-tocopherol. The American journal of clinical nutrition. 1994;60(1):106-10.

66. El-Sohemy A, Baylin A, Kabagambe E, Ascherio A, Spiegelman D, Campos H. Individual carotenoid concentrations in adipose tissue and plasma as biomarkers of dietary intake. The American journal of clinical nutrition. 2002;76(1):172-9.

67. Vioque J, Weinbrenner T, Asensio L, Castello A, Young IS, Fletcher A. Plasma concentrations of carotenoids and vitamin C are better correlated with dietary intake in normal weight than overweight and obese elderly subjects. The British journal of nutrition. 2007;97(5):977-86.

68. Woodside JV, Young IS, Gilchrist SE, Vioque J, Chakravarthy U, de Jong PT, et al. Factors associated with serum/plasma concentrations of vitamins A, C, E and carotenoids in older people throughout Europe: the EUREYE study. European journal of nutrition. 2013;52(5):1493-501.

69. Moreau C, Lefebvre JF, Jomphe M, Bherer C, Ruiz-Linares A, Vezina H, et al. Native American admixture in the Quebec founder population. PloS one. 2013;8(6):e65507.

70. Moran R, Nolan JM, Stack J, O'Halloran AM, Feeney J, Akuffo KO, et al. Non-Dietary Correlates and Determinants of Plasma Lutein and Zeaxanthin Concentrations in the Irish Population. The journal of nutrition, health & aging. 2017;21(3):254-61.

71. Gruber M, Chappell R, Millen A, LaRowe T, Moeller SM, Iannaccone A, et al. Correlates of serum lutein + zeaxanthin: findings from the Third National Health and Nutrition Examination Survey. The Journal of nutrition. 2004;134(9):2387-94.

72. Al-Delaimy WK, van Kappel AL, Ferrari P, Slimani N, Steghens JP, Bingham S, et al. Plasma levels of six carotenoids in nine European countries: report from the European Prospective Investigation into Cancer and Nutrition (EPIC). Public health nutrition. 2004;7(6):713-22.

Tables

Table 6.1 Characteristic and biochemical parameters of study subjects

Biochemical parameters	Fathers (n=6)	Mothers (n=16)	Boys (n=18)	Girls (n=8)
Age (years)	42.0 ± 2.8	42.4 ± 6.0	11.9 ± 3.8	10.1 ± 2.0
BMI (kg/m^2)	24.8 ± 1.3	23.5 ± 3.4	-	-
BMI percentile	-	-	48.9 ± 32.7	52.1 ± 29.8
TC (mmol/L)	4.70 ± 0.61	4.67 ± 0.55	4.35 ± 0.53	4.15 ± 0.48
HDL-C (mmol/L) [1]	1.36 ± 0.33	1.73 ± 0.35	1.56 ± 0.28	1.54 ± 0.31
LDL-C (mmol/L)	2.83 ± 0.67	2.53 ± 0.49	2.36 ± 0.42	2.20 ± 0.52
TC/HDL-C [1]	3.62 ± 0.88	2.79 ± 0.63	2.84 ± 0.42	2.79 ± 0.60
ApoB100 (g/L)	0.89 ± 0.22	0.77 ± 0.11	0.72 ± 0.12	0.67 ± 0.15
TG (mmol/L)	1.13 ± 0.35	0.88 ± 0.34	0.94 ± 0.39	0.89 ± 0.41
Glucose (mmol/L) [2]	5.25 ± 0.38	5.19 ± 0.34	4.92 ± 0.24	4.73 ± 0.14
Insulin (pmol/L)	56.3 ± 8.0	70.4 ± 38.7	69.4 ± 22.5	81.0 ± 40.8
SBP (mm Hg) [1]	117.5 ± 12.5	103.7 ± 8.9	103.2 ± 9.9	108.0 ± 11.3
DBP (mm Hg)	70.7 ± 13.6	63.3 ± 8.2	58.3 ± 8.1	65.6 ± 8.5
CRP (mg/L)	0.45 ± 0.39	0.89 ± 0.93	0.47 ± 0.79	0.33 ± 0.22

All values are means ± SD. [1] Means are significantly different ($P \leq 0.05$) between fathers and mothers. [2] Means are significantly different ($P \leq 0.05$) between boys and girls. Abbreviations: Apolipoprotein B100 (ApoB100), C-reactive protein (CRP), Diastolic blood pressure (DBP), High-density lipoprotein cholesterol (HDL-C), Low-density lipoprotein cholesterol (LDL-C), Systolic blodd pressure (SBP), Standard deviation (SD), Total cholesterol (TC), Total cholesterol/HDL-C (TC/HDL-C), Triglycerides (TG).

Table 6.2 Heritability estimates of carotenoid concentrations

Carotenoids	Maximal heritability (%)	Genetic heritability (%)	Common env. effect (%)	Familial effect *P*-value
α-carotene	43.40	28.54	14.87	0.18
β-carotene	51.29	41.68	9.61	0.14
β-cryptoxanthin	23.89	0.0043	23.88	0.49
Lutein	88.32	43.82	44.50	0.000006*
Lycopene	45.23	0.000095	45.23	0.0043*
Zeaxanthin	44.67	44.67	0	0.45
Total carotenoids	43.38	25.81	17.57	0.13

Heritability estimates from the full ACE general model. Familial effect (genetic and common environmental effects) significantly different from 0 ($P \leq 0.05$).

Table 6.3 Significant phenotypic, genetic, and environmental correlations between carotenoid concentrations and cardiometabolic risk factors

CM risk factor	$\rho_P \pm SE$	*P*	$\rho_G \pm SE$	*P*	$\rho_E \pm SE$	*P*
			β-cryptoxanthin			
TG	0.31 ± 0.14	0.035*	0.35 ± 0.42	0.50	0.28 ± 0.33	0.44
CRP[1]	0.28 ± 0.14	0.046*	-1 ± 0	0.51	0.61 ± 0.31	0.012**
			Lycopene			
SBP	0.31 ± 0.14	0.042*	0.62 ± 0.41	0.16	-0.12 ± 0.48	0.80
DBP	0.44 ± 0.12	0.0017**	1 ± 0	0.009**	-0.53 ± 0.55	0.23
			Zeaxanthin			
TG	0.40 ± 0.13	0.021*	0.61 ± 0.42	0.21	0.18 ± 0.39	0.65
Glucose	0.33 ± 0	0.011*	0.24 ± 0.76	1	0.51 ± 0.23	0.13

Phenotypic correlation (ρ_P) is partitioned into genetic (ρ_G) and environmental (ρ_E) correlations. * Significant correlation *P*-value ≤ 0.05. ** Significant Bonferroni adjusted *P*-value.[1] Values are log10 transformed. Abbreviations: Apolipoprotein B100 (ApoB100), Cardiometabolic (CM), C-reactive protein (CRP), Diastolic blood pressure (DBP), High-density lipoprotein cholesterol (HDL-C), Low-density lipoprotein cholesterol (LDL-C), Standard error (SE), Systolic blood pressure (SBP), Total cholesterol (TC), Total cholesterol/HDL-C (TC/HDL-C), Triglycerides (TG).

Supplemental material

Supplemental table 6.1 Concentrations of plasma carotenoids (µmol/L of plasma)

Subjects	Lutein	Zeaxanthin	β-Cryptoxanthin	α-Carotene	β-Carotene	Lycopene	Total carotenoids
Gen01	0.36689606	0.0678172	0.48942068	0.94815424	1.96626358	1.90477202	5.74332378
Gen02	0.19040767	0.05211724	0.2671793	0.39046788	1.63413354	2.97212935	5.50643497
Gen03	0.20500873	0.04703513	0.33127708	1.3088189	1.97765103	2.14359446	6.01338535
Gen05	0.24382521	0.05322967	0.42173556	0.80394106	2.58467386	2.80014081	6.90754618
Gen06	0.21138746	0.03997405	0.47272927	0.87987086	2.33627976	2.28340529	6.2236467
Gen07	0.15361658	0.03836336	0.20938586	1.01423133	1.62869228	1.21588093	4.26017034
Gen08	0.13950888	0.03245329	0.2880089	0.56440081	1.07092427	1.52477859	3.62007474
Gen09	0.18935861	0.06023487	0.24183803	0.36889688	1.46894869	1.84769435	4.17697143
Gen10	0.45531524	0.05366406	0.68650319	1.86816097	5.09892831	1.60505358	9.76762535
Gen11	0.13301695	0.0379158	0.40775388	1.17280733	1.78868368	1.15260911	4.69278675
Gen12	0.1495655	0.03782232	0.52583487	1.20447	2.13195549	1.51354258	5.56319076
Gen13	0.22080196	0.04605841	0.40370279	1.95125405	4.36470233	1.77749248	8.76401203
Gen14	0.19439964	0.0346023	0.47240977	1.69428919	2.77308628	1.17814005	6.34692723
Gen15	0.31456871	0.07906164	0.64207259	0.585099	1.26299619	2.07400924	4.95780737
Gen16	0.20941464	0.0590733	0.66509735	0.43477479	1.2177428	1.90799437	4.49409726
Gen17	0.31659735	0.05824728	0.43726324	3.07504388	4.90271428	1.52411709	10.3139831
Gen18	0.40720388	0.06879612	0.52897983	1.77242233	2.73849116	1.40131048	6.9172038

Gen19	0.21406127	0.04334581	0.21302691	1.46073827	2.3907969	0.81416873	5.13613789
Gen20	0.38953299	0.11518446	0.54511458	0.58599723	1.94027507	1.46287634	5.03898068
Gen21	0.25823509	0.05614234	0.5515715	3.29451982	4.1048887	1.59654332	9.86190078
Gen22	0.19637546	0.05178252	0.52029086	1.43217409	2.81824867	1.43588238	6.45475399
Gen23	0.19796017	0.03400027	0.36306339	0.6257273	1.36047476	1.52337096	4.10459685
Gen24	0.23360289	0.05076413	0.42062581	1.11325177	1.89522389	1.31461194	5.02808043
Gen25	0.19012552	0.06455947	0.73073545	0.68144593	1.40988861	0.8309107	3.90766568
Gen26	0.20673794	0.04472159	0.11642864	0.1033121	0.30102258	0.5048136	1.27703645
Gen27	0.23324584	0.07133053	0.53723883	0.38361577	0.77381218	0.77033291	2.76957605
Gen28	0.22858608	0.06703344	0.40164353	0.34477048	0.57881226	0.69589933	2.31674512
Gen29	0.22569969	0.06288219	0.19197018	0.79944354	1.42605149	1.67944118	4.38548827
Gen30	0.18846852	0.05111517	0.18449405	1.76125731	3.75436001	1.36074407	7.30043914
Gen31	0.1034949	0.03639765	0.23772312	0.55324807	1.51238711	1.61846353	4.06171437
Gen32	0.1379356	0.05037577	0.58562149	1.28098545	2.66908747	1.96597652	6.6899823
Gen33	0.34863636	0.06872228	0.59045342	1.05182969	2.8442326	1.16548687	6.06936121
Gen34	0.34080966	0.08200735	0.53132439	0.87811067	3.02914672	1.32128546	6.18268426
Gen35	0.18558526	0.05637978	0.69489283	1.20488219	2.06864071	1.10169917	5.31207994
Gen36	0.21559215	0.1016761	0.82701707	2.65607443	4.85041133	1.40830847	10.0590796
Gen38	0.26706147	0.09209016	.	3.55025825	6.33803969	1.79779113	12.0452407
Gen39	0.26395015	0.06463881	0.53649644	1.52282902	3.89619121	1.72663359	8.01073921

Gen40	0.22008198	0.04334325	0.45042163	1.29773283	3.99433146	2.71061293	8.71652407
Gen41	0.33789169	0.08950376	0.40316064	2.05426427	3.12184275	1.53282075	7.53948386
Gen42	0.16596505	0.05910571	0.43107599	0.74232185	1.36939518	1.38768457	4.15554834
Gen43	0.18579005	0.05817576	0.6537702	0.9184596	2.08512821	1.28347111	5.18479494
Gen44	0.26494297	0.09297957	0.90039749	1.52573514	3.79402374	2.5744044	9.15248331
Gen45	0.49521561	0.05468624	0.35857692	1.28192912	3.42957551	1.58930011	7.2092835
Gen46	0.72653971	0.09196268	0.47041018	1.0846667	1.94897283	1.42895167	5.75150378
Gen47	0.5848228	0.06066084	0.40081868	1.18038703	2.54237772	1.59877117	6.36783824
Gen48	0.55555079	0.07599738	0.5545489	0.91123368	1.78155749	1.2136722	5.09256043
Gen49	0.21725674	0.05761664	0.4287788	0.67397551	2.02502444	1.58549025	4.98814238
Gen50	0.22004146	0.03996355	0.286557	0.66712712	1.28660567	1.18069892	3.68099372

Supplemental table 6.2 Phenotypic correlations between carotenoid levels and cardiometabolic risk factors

CM risk factor	ρ_P ± SE	*P*-value
α-carotene		
TC	0.20 ± 0.14	0.17
LDL-C	0.092 ± 0.15	0.55
HDL-C	0.23 ± 0	0.065
TC/HDL-C	-0.10 ± 0.17	0.49
TG	0.087 ± 0.15	0.57
ApoB100	0.22 ± 0.14	0.10
Glucose	0.0094 ± 0	0.94
Insulin	-0.067 ± 0.15	0.65
SBP	-0.19 ± 0.15	0.22
DBP	-0.087 ± 0.16	0.59
CRP[1]	0.15 ± 0.14	0.30
β-carotene		
TC	0.23 ± 0.14	0.14
LDL-C	0.13 ± 0.15	0.38
HDL-C	0.18 ± 0	0.15
TC/HDL-C	-0.068 ± 0.15	0.64
TG	0.017 ± 0.15	0.91
ApoB100	0.27± 0.13	0.052
Glucose	0.027 ± 0.14	0.85
Insulin	-0.15 ± 0.15	0.32
SBP	-0.25 ± 0.15	0.11
DBP	-0.048 ± 0.16	0.77
CRP[1]	0.12 ± 0.14	0.42
β-cryptoxanthin		
TC	-0.094 ± 0.15	0.53
LDL-C	-0.22 ± 0.15	0.15
HDL-C	0.027 ± 0.15	0.86
TC/HDL-C	-0.092 ± 0.15	0.54
TG	0.31 ± 0.14	0.035*
ApoB100	-0.041 ± 0.15	0.79
Glucose	0.12± 0.14	0.41
Insulin	0.18 ± 0.14	0.22
SBP	-0.017 ± 0.16	0.91
DBP	0.057 ± 0.16	0.71
CRP [1]	0.28 ± 0.14	0.046*
Lutein		
TC	0.19 ± 0.15	0.22
LDL-C	0.083 ± 0.15	0.58
HDL-C	0.26 ± 0.17	0.077
TC/HDL-C	-0.17 ± 0.15	0.25
TG	-0.015 ± 0.15	0.92
ApoB100	0.14 ± 0.14	0.32
Glucose	0.095 ± 0.14	0.51
Insulin	-0.16 ± 0.15	0.30
SBP	-0.26± 0.15	0.092
DBP	-0.23 ± 0.15	0.14
CRP[1]	0.00038 ± 0.14	1.00

Lycopene		
TC	0.16 ± 0.15	0.30
LDL-C	0.11 ± 0.16	0.48
HDL-C	-	-
TC/HDL-C	-0.11 ± 0.14	0.44
TG	0.012 ± 0.15	0.94
ApoB100	0.12 ± 0.15	0.43
Glucose	0.12 ± 0.14	0.41
Insulin	0.052 ± 0.15	0.73
SBP	0.31 ± 0.14	0.042*
DBP	0.44 ± 0.12	0.0017**
CRP[1]	-0.024 ± 0.14	0.87
Zeaxanthin		
TC	0.17 ± 0.16	0.29
LDL-C	0.092 ± 0.16	0.56
HDL-C	0.042 ± 0.15	0.78
TC/HDL-C	0.048 ± 0.15	0.75
TG	0.40 ± 0.13	0.021*
ApoB100	0.22± 0.14	0.13
Glucose	0.33 ± 0	0.011*
Insulin	0.054 ± 0.15	0.72
SBP	0.042 ± 0.16	0.79
DBP	0.074± 0.16	0.65
CRP[1]	0.089 ± 0.15	0.55
Total carotenoids		
TC	0.24 ± 0.14	0.12
LDL-C	0.12 ± 0.15	0.43
HDL-C	0.21 ± 0.14	0.14
TC/HDL-C	-0.094 ± 0.14	0.52
TG	0.079 ± 0.15	0.60
ApoB100	0.27 ± 0.13	0.058
Glucose	0.058 ± 0.14	0.69
Insulin	-0.091 ± 0.15	0.54
SBP	-0.14 ± 0.15	0.35
DBP	0.037 ± 0.16	0.82
CRP[1]	0.14 ± 0.14	0.28

* Significant correlation *P*-value ≤ 0.05. ** Significant Bonferroni adjusted *P*-value (0.05/11=0.0045).[1] Values are log10 transformed. Abbreviations: Apolipoprotein B100 (ApoB100), Cardiometabolic (CM), C-reactive protein (CRP), Diastolic blood pressure (DBP), High-density lipoprotein cholesterol (HDL-C), Low-density lipoprotein cholesterol (LDL-C), Phenotypic correlation (ρ_P), Standard error (SE), Systolic blood pressure (SBP), Total cholesterol (TC), Total cholesterol/HDL-C (TC/HDL-C), Triglycerides (TG).

Supplemental table 6.3 Heritability estimates of carotenoid levels

Carotenoids	Genetic heritability (%)	Genetic effect *P*-value
α-carotene	48.38	0.0793
β-carotene	56.23	0.0504
β-cryptoxanthin	31.29	0.3578
Lutein	99.81	0.000009*
Lycopene	76.61	0.0023*
Zeaxanthin	44.65	0.208
Total carotenoids	51.68	0.0541

Heritability estimates from the alternative AE model. * Genetic effect significantly different from 0 ($P \leq 0.05$).

Chapitre 7 Analyse de réseaux de co-expression génique dans la relation entre les caroténoïdes plasmatiques et le profil lipidique

Bénédicte L. Tremblay, Frédéric Guénard, Benoît Lamarche, Louis Pérusse, Marie-Claude Vohl

L'article présenté dans ce chapitre s'intitule : *Weighted gene co-expression network analysis to explain the relationship between plasma total carotenoids and lipid profile*.

Cet article est publié dans la revue : *Genes and Nutrition. 2019 May 8;14:16.*

Résumé

Introduction : La variabilité des caroténoïdes circulants est attribuable à plusieurs facteurs, dont les variations génétiques et le profil lipidique. Toutefois, peu d'études ont examiné l'impact de l'expression génique sur la variabilité interindividuelle des caroténoïdes circulants. La plupart des études ont considéré l'expression des gènes individuellement sans tenir compte de leur degré élevé d'interconnexion. L'analyse de réseaux de corrélations pondérées (WGCNA) est une méthode de la biologie des systèmes utilisée pour identifier des groupes de gènes avec des niveaux d'expression hautement corrélés pour ensuite les associer à des traits phénotypiques. L'objectif de cette étude observationnelle est d'examiner la relation entre les concentrations plasmatiques de caroténoïdes totaux et le profil lipidique en utilisant WGCNA.

Résultats : Les niveaux d'expression dans le sang de 533 sondes étaient associés aux caroténoïdes totaux plasmatiques. Parmi les quatre modules identifiés avec WGCNA, les modules turquoise, bleu et brun étaient corrélés au cholestérol HDL et au cholestérol total. Les sondes fortement corrélées au cholestérol HDL et au cholestérol total étaient également les sondes les plus importantes des modules brun et bleu. Un total de quatre et de 29 *hub genes* associés aux caroténoïdes totaux étaient également associés au cholestérol HDL et au cholestérol total, respectivement.

Conclusions : Les niveaux d'expression de 533 sondes étaient associés aux concentrations plasmatiques de caroténoïdes totaux. À l'aide de WGCNA, quatre modules et plusieurs *hub genes* du métabolisme des lipides et des caroténoïdes ont été identifiés. Cette analyse intégrative fournit des preuves du rôle potentiel de la co-expression génique dans la relation entre les caroténoïdes et les lipides sanguins. D'autres études et la validation des *hub genes* sont nécessaires.

Abstract

Introduction: Variability in circulating carotenoids may be attributable to several factors including, among others, genetic variants, and lipid profile. However, relatively few studies have considered the impact of gene expression in the inter-individual variability in circulating carotenoids. Most studies considered expression of genes individually and ignored their high degree of interconnection. Weighted gene co-expression network analysis (WGCNA) is a systems biology method used for finding gene clusters with highly correlated expression levels and for relating them to phenotypic traits. The objective of the present observational study is to examine the relationship between plasma total carotenoid concentrations and lipid profile using WGCNA.

Results: Whole blood expression levels of 533 probes were associated with plasma total carotenoids. Among the four WGCNA distinct modules identified, turquoise, blue and brown modules correlated with plasma HDL-C and total cholesterol. Probes showing a strong association with HDL-C and total cholesterol were also the most important elements of the brown and blue modules. A total of four and 29 hub genes associated with total carotenoids were potentially related to HDL-C and total cholesterol, respectively.

Conclusions: Expression levels of 533 probes were associated with plasma total carotenoid concentrations. Using WGCNA, four modules and several hub genes related to lipid and carotenoid metabolism were identified. This integrative analysis provides evidence for the potential role of gene co-expression in the relationship between carotenoids and lipid concentrations. Further studies and validation of the hub genes are needed.

Title page

Weighted gene co-expression network analysis to explain the relationship between plasma total carotenoids and lipid profile

Bénédicte L. Tremblay[1, 2], Frédéric Guénard[1, 2], Benoît Lamarche[1, 2], Louis Pérusse[1, 3], Marie-Claude Vohl[1,2*]

1. Institute of Nutrition and Functional Foods (INAF), Laval University, 2440 Hochelaga Blvd, Quebec City, QC, G1V 0A6, Canada.
2. School of Nutrition, Laval University, 2425 rue de l'Agriculture, Quebec City, QC, G1V 0A6, Canada.
3. Department of Kinesiology, Laval University, 2300 rue de la Terrasse, Quebec City, QC, G1V 0A6, Canada.

*Corresponding author:
Marie-Claude Vohl, Ph.D.
Institute of Nutrition and Functional Foods (INAF), Laval University
2440 Hochelaga Blvd, Quebec City, QC, G1V 0A6, Canada
Phone: 418 656-2131 ext. 404676; Fax: 418 656-5877
E-mail: marie-claude.vohl@fsaa.ulaval.ca

Keywords: Carotenoids; French Canadians; Hub genes; Lipid profile; WGCNA.

Introduction

Carotenoids are a reliable biomarker of fruit and vegetable consumption (1, 2). Over 90% of the daily carotenoid intakes are provided by fruits and vegetables (3). Carotenoids are a family composed of more than 700 fat-soluble pigments, but α-carotene, β-carotene, β-cryptoxanthin, lutein, lycopene, and zeaxanthin represent over 95% of total circulating carotenoids in human plasma or serum (3, 4). Inter-individual variability in circulating carotenoids has been observed and may be attributable to several factors including age, sex, body weight, physical activity, genetic, and lipid profile (2). Accordingly, lower total cholesterol (TC), LDL-cholesterol (LDL-C), and HDL-cholesterol (HDL-C) concentrations were associated with lower circulating carotenoids (5). Plasma HDL-C concentrations also mediated the observed sex difference in serum carotenoids in Caucasian individuals (5). Correlations were observed between α- and β-carotene and HDL-C and triglyceride (TG) concentrations (6) and between β-crytoxanthin and zeaxanthin and TG in a previous study by our group (7). The NHANES 2003–2006 study also reported that several serum carotenoids positively correlated with HDL-C concentrations (8). Thus, the link between circulating carotenoids and plasma lipid profile has been observed in several studies and is plausible considering that the majority of circulating carotenoids are associated with lipoproteins (9).

Carotenoid intakes are inversely associated with the risk of cardiovascular diseases and certain cancers (10). They may mediate their effects mainly via antioxidant properties but also via other mechanisms such as gap junction communication, cell growth regulation, immune response, and modulation of gene expression (11, 12). Several genome-wide association studies have identified genetic variants that influence circulating carotenoid concentrations (13-16). Genetic variants may cause differences in the absorption, assimilation, distribution, metabolism, and excretion of carotenoids (13, 17, 18). Carotenoids have also been shown to regulate the expression of genes protective against carcinogenesis and inflammation (19). Carotenoids and their derivatives (e.g., retinoid) may exert their effect on gene expression via several transcriptional systems such as the retinoid receptors, the nuclear factor-kappa B, and the peroxisome proliferator-activated receptors (20). However, relatively few studies have considered the impact of gene expression in the inter-individual variability in circulating carotenoid concentrations. Gene expression may represent a potential mechanism that links carotenoids to lipid profile. In addition, no previous study has examined the interconnection between plasma carotenoids, lipid profile, and genome-wide gene expression levels. Most studies considered expression of genes individually and ignored the high degree of interconnection between genes. Weighted gene co-expression network analysis (WGCNA) is a widely used systems biology approach used to identify gene clusters (modules) with highly correlated expression levels, to relate the modules to phenotypic traits, and to identify key hub genes within modules that are related to the phenotypic traits (21).

Thus, the objective of the present observational study is to examine the relationship between plasma total carotenoid concentrations and lipid profile using WGCNA. The hypothesis was that genome-wide gene

expression levels are associated with plasma total carotenoids and that clusters of genes associated with carotenoids are correlated to lipid profile. To test this hypothesis, genes associated with total carotenoids were identified using regressions and WGCNA was used to identify specific modules and key hub genes related to lipid profile.

Methods

Patients and design

A total of 48 Caucasian French-Canadian subjects from 16 families were recruited in the Greater Quebec City metropolitan area, in Canada, as part of the GENERATION Study. The GENERATION Study was designed to evaluate familial resemblances in omics (DNA methylation (55) and gene expression (56)) and metabolic (metabolites (57) and carotenoids (7)) profiles in healthy families and to test impact of these profiles on cardiometabolic health. Families were composed of 16 mothers, 6 fathers, and 26 children. Families living under the same roof comprised at least the mother and one child aged between 8 and 18. Parents had to be the biological parents of their child (or children), in good general health, non-smokers, with body mass index (BMI) ranging between 18 and 35 kg/m^2, and free of any metabolic conditions requiring treatment, although the use of Synthroid® (levothyroxine) or oral contraceptive was tolerated. Children also had to be non-smokers, in good general health and not using psycho-stimulators [Ritalin® (methylphenidate), Concerta® (methylphenidate), and Strattera® (atomoxetine)]. Blood samples were taken from both parents and children during their visit at the Institute of Nutrition and Functional Foods (INAF). The experimental protocol was approved by the Ethics Committees of Laval University Hospital Research Center and Laval University. All participants (adults and children) signed an informed consent document. Parental consent was also obtained by signing the child consent document.

Anthropometric and cardiometabolic measurements

Body weight, waist girth, and height were measured according to the procedures recommended by the Airlie Conference (58). Blood samples were collected from an antecubital vein into vacutainer tubes containing EDTA after 12-hour overnight fast and 48-hour alcohol abstinence. Plasma was separated by centrifugation (2 500g for 10 min at 4°C), and samples were aliquoted and frozen (-80°C) for subsequent analyses. Enzymatic assays were used to measure plasma TC and TG concentrations (59, 60). Precipitation of very-low density lipoprotein (VLDL) and LDL particles in the infranatant with heparin manganese chloride generated the HDL-C fraction (61). LDL-C was calculated with the Friedewald formula (62). Apolipoprotein B-100 (ApoB100) concentrations were measured in plasma by the rocket immunoelectrophoretic method (63). Using a sensitive assay, plasma C-reactive protein (CRP) was measured by nephelometry (Prospec equipment Behring) (64)

RNA extraction and gene expression analysis

Total RNA was isolated and purified from whole blood using PAXgene Blood RNA Kit (QIAGEN). Quantification and verification of the purified RNA was assessed using both the NanoDrop (Thermo Scientific, Wilmington, DE, USA) and the 2100 Bioanalyzer (Agilent Technologies, Cedar Creek, TX, USA). The HumanHT-12 v4 Expression BeadChip (Illumina Inc., San Diego, CA) was used to measure expression levels of ~ 47 000 probes (> 31 000 annotated genes). This was performed at the McGill University and Genome Quebec Innovation

Centre (Montreal, Canada). The FlexArray software (version 1.6) (65) and the lumi R package were used to analyze and normalize gene expression levels. Probes with a detection *P*-value ≤ 0.05 in at least 25% of all subjects were considered in analysis. A total of 18 160 probes among the 47 323 probes on the microarray (38.4%) showed significant gene expression in blood.

Plasma carotenoid measurements

Samples and standards used for the measurement of carotenoid concentrations were prepared as reported previously (7). Briefly, 100 µL of fasting plasma samples were thawed a few hours before analysis. Plasma samples were transferred to Eppendorf tubes with 20 µL of 2-propanol and 20 µL of carotenoid standard. Samples were transferred on a 400µL fixed well plate (ISOLUTE® SLE+, Biotage, Charlotte, NC) with 900 µL of hexane:isopropanol (90/10, *v/v*) in each well. Each extracted sample was evaporated under nitrogen and reconstituted with 300 µL of methanol:dichloromethane (65/35, *v/v*). Plates were shaken for 10 minutes and samples were transferred into high performance liquid chromatography glass vials for analysis.

High performance liquid chromatography (HPLC)-UV analysis was performed using an Agilent 1260 liquid handling system (Agilent, Mississauga, Ontario, Canada) as described previously (7). Carotenoids were separated with a mobile phase consisting of methanol:water (98/2, *v/v*; Eluent A) and methyl-tert-butyl ether (MTBE; Eluent B; VWR, Mississauga, Ontario, Canada). Flow-rate was set at 1 mL/min and the gradient elution was as follows: 2% Eluent B (initial), 2.0-80% Eluent B (0.0-27.0 min), isocratic 80% Eluent B (27.0-31.0 min), 80.0-2.0% Eluent B (31.0-31.1 min), and isocratic 2% Eluent B (31.1-34.0 min). UV detector was set at 450 nm and identification of each compound was confirmed using retention time and UV spectra (190-640nm) of the pure compounds. Data acquisition was carried out with the Chemstation software (Agilent, Mississauga, Ontario, Canada). For all carotenoids the concentrations are reported in µmol/L of plasma. One outlier in β-crypoxanthin, defined as value falling outside of the mean ± 4 standard deviations, was excluded from analyses.

Association between total carotenoids and gene expression levels

Total plasma carotenoid (µmol/L of plasma) concentrations were calculated as the sum of α-carotene, β-carotene, β-cryptoxanthin, lutein, lycopene, and zeaxanthin concentrations. Concentrations of plasma carotenoids are available in Additional Table 1. R software v2.14.1 (R Foundation for Statistical Computing; http://www.r-project.org) (66) was used to compute regressions between normalised gene expression levels of all 18 160 probes and total carotenoids adjusted for the family ID. Weighted gene co-expression network analysis (WGCNA) was performed with gene expression levels of 533 probes showing a significant association (*P*-value ≤ 0.05, obtained from the linear model function) with total carotenoids.

Weighted gene co-expression network analysis (WGCNA)

WGCNA was performed with the WGCNA package (21, 67) in R software (66). First, co-expression was measured between each gene pair using Pearson correlation coefficients (varying from -1 to 1). Considering the small sample size of the present study, Pearson correlations measuring linear relationships were selected to avoid the pitfall of overfitting (68). To transform the correlation coefficients into a weighted adjacency matrix (values ranging from 0 to 1), the co-expression similarity was raised to a power $\beta = 6$ (68). The adjacency matrix allows measuring the connection strengths between nodes. From this matrix, we can build the topological overlap matrix (TOM) that considers the topological similarity. The TOM was then used to calculate the corresponding dissimilarity (1-TOM) in order to form clusters. Average linkage hierarchical clustering coupled with the TOM-based dissimilarity was used to group genes with coherent expression profiles into modules (68). More specifically, the dynamic tree cutting algorithm (deep split = 2, minimum number of genes per modules = 30, cut height = 0.25) was used to detect gene modules (clusters of densely interconnected genes in terms of co-expression). The Partitioning Around Medoids (PAM) method was also used to allow the assignment of outlying genes to modules. Colors are randomly assigned to modules except for grey color, which is reserved for the module with unassigned genes. To identify modules that were significantly associated with lipid profile traits (TC, LDL-C, HDL-C, TG, ApoB100), correlations between module eigengenes (MEs) (i.e., the first principal component of the module, which represents the overall expression level of the module) (69) and traits were computed. Gene significance (GS), defined as the absolute correlation between the gene and the trait, was used to quantify associations of individual genes with lipid profile traits. To quantify the similarity of all genes to every module, a quantitative measure of module membership (MM) was defined as the correlation of the ME and the gene expression profile. Genes with the highest MM and highest GS were those with high significance (hub genes) (70). The hub genes within a module were chosen based on GS >0.2 and MM >0.6, with a *P*-value ≤0.05. All these analyses were computed using commands implemented in the WGCNA package.

Functional enrichment analysis and hub genes classification

Reactome and Gene ontology (GO) enrichment analysis was conducted with KEGG Orthology-Based Annotation System (KOBAS) on the genes of modules of interest (71). Hypergeometric test / Fisher's exact test was used to identify significant pathways. FDR correction method (Benjamini and Hochberg) was used to account for multiple testing in enrichment analysis. The most significant Reactome terms were used to characterize brown and blue modules. TFs enrichment analysis was conducted using Enrichr (http://amp.pharm.mssm.edu/Enrichr/), an online biological information database that integrates several biological databases (72). The most significant TFs in each module were retained. Moreover, analyses were carried out to suggest potential implications of hub genes in the relationship between carotenoids and lipids. First, regressions between expression levels of all hub genes and plasma lipid concentrations adjusted for the family ID were computed in Statistical Analysis Software (SAS) version 9.4. More specifically, the seven transcripts of the brown module were tested for associations with HDL-C and the 32 transcripts of the blue

module were tested for associations with TC. Significant associations ($P \leq 0.05$) are presented in Additional Table 5. Considering that all hub genes initially showed associations with total carotenoids, potential causal relationships of hub genes were classified as followed: 1) Unknown: genes that showed association with carotenoids but not with lipids; 2) Likely regulated: genes that showed association with both carotenoids and lipids; 3) Likely regulator: genes that showed association with carotenoids and either have been associated with lipids in literature or showed association with lipids in addition to being shown to influence lipid levels in the literature. Literature of associations of genes with lipids is detailed in the discussion. These classification terms were used to quickly have an idea of the potential implications of hub genes in the relationship between carotenoids and lipids.

Statistical analysis

SAS version 9.4 was used to compute sex differences in cardiometabolic parameters between fathers and mothers, and between daughters and sons using an unpaired t-test.

Results

Characteristics of study participants

Characteristics and biochemical parameters of participants are presented in Table 1. Fathers and mothers had significant differences in HDL-C concentrations. Concentrations of all six main carotenoids (α-carotene, β-carotene, β-cryptoxanthin, lutein, lycopene, and zeaxanthin) and total carotenoids measured in the fasting state are presented in Additional Table 1.

Associations between total carotenoids and genome-wide gene expression levels

Normalized gene expression levels of 18 160 probes were tested for associations with plasma total carotenoid concentrations. A total of 533 probes were associated ($P \leq 0.05$) with total carotenoids (Additional Table 2). WGCNA was then conducted on this subset of probes, which were associated with total carotenoids in order to put them in relationship with lipid profile.

Weighted gene co-expression network analysis (WGCNA)

A total of four distinct modules were identified from gene expression levels of the 533 probes using a dynamic tree cutting algorithm (Figure 1). The blue, brown, turquoise, and grey modules contained 117, 64, 184, and 168 genes, respectively. The 168 uncorrelated genes were assigned to the grey module, which was excluded from further analysis. Using a cut-off height of 0.25 for the clustering of module eigengenes (MEs), none of the four modules merged. Thus, the merged dynamic yielded same modules as the dynamic tree cut (Figure 1). To find modules of interest, correlations between MEs and lipid profile traits (TC, LDL, HDL, TG, ApoB100) were computed. According to the heatmap of module-trait correlations (Figure 2), ME of the turquoise module (184 genes) correlated inversely with HDL-C (r=-0.32 p=0.03) and TC (r=-0.35 p=0.01), while ME of the blue module (117 genes) correlated inversely with HDL-C (r=-0.36 p=0.01) and TC (r=-0.41 p=0.004), and ME of the brown module (64 genes) correlated positively with HDL-C (r=0.31 p=0.03).

Gene significance (GS), defined as the correlation between genes expression and lipid profile traits, was put in relation with module membership (MM), defined as the correlation of the ME and the gene expression profile. GS for HDL-C correlated with MM in the brown module (r=0.29 p=0.02), while GS for TC correlated with MM in the blue module (r=0.42 $p=2.4\text{x}10^{-6}$) (Figure 3). This suggests that genes highly significantly associated with HDL-C and TC were also the most important elements of the brown and blue modules, respectively. Thus, brown and blue modules were selected as modules of interest in subsequent analyses.

Enrichment analysis

Enrichment analysis was performed for genes in brown and blue modules in order to elucidate biological mechanisms. In the GO database, a total of 73 GO terms from all three main categories (biological process,

cellular component, and molecular function) were significantly enriched (corrected $p \leq 0.05$) in the brown module (Additional Table 3). In the GO database, a total of 253 GO terms also from all three main categories were significantly enriched (corrected $p \leq 0.05$) in the blue module (Additional Table 4). In the Reactome database, the most significant terms were used to better define modules of interest. Thus, the function representing the brown module was "apoptosis", while the function representing the blue module was "mRNA splicing" (Additional Tables 3 and 4). Moreover, transcription factor (TF) enrichment analysis was conducted to identify TFs that are enriched in genes of the brown and blue modules. In the brown module, *NFE2L2* was the more significantly enriched TF (p=0.016) with target genes *MAF* and *MAFF*. In the blue module, *ESR1* was the most significantly enriched TF (p=0.0047) with target genes *NOP56, ATF1, LDHA, SMARCB1, HNRNPA2B1, HNRNPR, SLC3A2, DNM1L, CCT7, HNRNPA1, DBN1*, and *MARK2*.

Hub gene analysis

Hub gene analysis in brown and blue modules was conducted in order to refine the analysis of potential mechanisms. A total of seven transcripts (four hub genes: *CISH, FAM123C, ZSWIM4, FAM13A*) were identified in the brown module (Table 2). In the blue module, 32 transcripts (29 hub genes: *FEZ1, MYBPC3, OPRL1, SIRPB1, SLC16A3, SLC4A1, GPR89C, PSMD6, RANBP1, AP3M1, LOC221710, SIRT5, DNM1L, CR2, LBH, HNRNPA1, BIVM, NOP56, LOC728026, FBXO5, SNRPN, CCT7, TIAL1, SLC25A4, PCNA, PNN, S100PBP, UBE1DC1*, and *ZCCHC11*) were identified (Table 3). Classification terms (unknown, likely regulated, and likely regulator) were added in the Tables 2 and 3 to suggest potential causal relationships of the hub genes in the interconnection between total carotenoids and plasma lipid concentrations.

Discussion

We first tested the association between whole blood genome-wide gene expression levels and plasma total carotenoid concentrations in order to identify probes influenced by carotenoids. A total of 533 probes were significantly associated with total carotenoids. To the best of our knowledge, this is the first study that considers the impact of plasma total carotenoids on genome-wide gene expression levels. Up to now, studies focused mainly on the effect of genetic variants on the inter-individual variability in β-carotene, lycopene, and lutein bioavailability in response to dietary interventions (22-24). SNPs were located in genes mainly related to fatty acids and cholesterol absorption including *SR-B1, CD36, NPC1L1*, and *ABCA1*. Since several proteins are involved in the absorption and metabolism of carotenoids, several papers have suggested that variations in the expression of the genes encoding for these proteins might also impact carotenoid bioavailability (25). Moreover, carotenoids modulate the mechanisms of cell proliferation, growth factor signaling, gap junction communication, and lead to changes in the expression of many proteins involved in these processes (20). The effect of carotenoids on gene expression may result from direct interaction with ligand-activated nuclear receptors such as retinoid receptors, the nuclear factor-kappa B, and the peroxisome proliferator-activated receptors (20, 26, 27). In accordance, two important TFs were enriched among genes associated with total carotenoids. *NFE2L2*, which encodes a TF regulating genes with antioxidant response elements, was enriched in the brown module. Interestingly, some carotenoids have been showed to protect against light-induced cell damage through *NFE2L2* activation (28). *ESR1*, a ligand-activated TF, was enriched in the blue module. Genetic variations in this gene were associated with serum TC and LDL-C concentrations (29).

Second, WGCNA was used to identify modules and key genes related to lipid profile in the subset of 533 probes associated with carotenoids. MEs of the turquoise and blue modules had a significant correlation with HDL-C and TC, while ME of the brown module had a significant correlation with HDL-C. This suggests that highly co-expressed genes in these modules had potential interaction or consistent biological effects on HDL-C and TC. Only the brown and the blue modules showed a significant correlation between the GS for HDL-C and TC, respectively, and the MM of the module. Functional enrichment analysis revealed several enrichments of categories related to cell metabolism and processes. However, functional enrichment analysis yields large amount of GO terms and Reactome terms non-specific to lipid profile. Thus, it was not precise enough to identify specific pathways or genes related to lipid or carotenoid metabolism.

The analysis with hub genes allowed refining the analysis of potential mechanisms linking carotenoids and lipid profile. A total of four hub genes were identified in the brown module: *FAM123C, ZSWIM4, FAM13A*, and *CISH*. *FAM13A* and *CISH* are of interest in the context of lipid profile. *FAM13A* encodes for a family with sequence similarity 13 member A. *FAM13A* variants have been associated with HDL-C in individuals from various descents (30, 31). *FAM13A* promoted fatty acid oxidation, possibly by interacting with an activating sirutin 1, and

increasing expression of *CPT1A* (32). Interestingly, another study using WGCNA identified *FAM13A* as a hub gene related to hyperlipidemia (33). *CISH* encodes for a cytokine inducible SH2 containing protein. It controls the signaling of a variety of cytokines, in particular interleukin-2, and seems to be critical for T-cell proliferation and survival in response to infection (34). Despite the fact that expression of *CISH* was not inhibited in the presence of delipidated HDL lipoproteins in a cell study, the link is plausible since HDL specifically inhibits the production of pro-inflammatory cytokines, which is also controlled by *CISH* (35). A total of 29 hub genes were identified in the blue module. Relevant genes related to lipid and carotenoid metabolism include *MYBPC3, OPRL1, SLC16A3, SIRT5, HNRNPA1, SLC25A4*, and *PCNA*. *MYBPC3* encodes for a myosin binding protein C, cardiac. Mutation in this gene (*MYBPC3-Q1061X*) results in hypertrophic cardiomyopathy and cardiac oxidative stress with elevated TG and branched chain amino acid levels (36, 37). *OPRL1* encodes for an opioid related nociceptin receptor 1 involved in many biological functions including stress, inflammatory and immune responses. A study by our group identified interaction effects between 29 SNPs, including rs2229205 in *OPRL1*, and total fat intake on LDL peak particle diameter (38). *SLC16A3* encodes for a proton-linked monocarboxylate transporter designated solute carrier family 16 member 3. Monocarboxylate transporters are involved in the transport of short chain fatty acids and may also be involved in the transport of cholesterol-lowering agents (39). *SLC16A3* was significantly upregulated in 102 men receiving an antioxidant-rich diet compared to controls (40). *SIRT5* encodes for sirtuin 5, which is involved in the regulation of mitochondrial metabolism, oncogenesis, and oxidative stress (41, 42). Obese Sirt5−/− mice showed increased serum cholesterol concentrations compared to Sirt5+/+ mice (43). *HNRNPA1* encodes for a heterogeneous nuclear ribonucleoprotein A1, which has been shown to reduce HMGCR enzyme activity and increase LDL-C uptake (44). *SLC25A4* encodes for a solute carrier family 25 member 4. A variation in this gene is associated with hypertrophic cardiomyopathy (45). Linoleic acid also increased *ANT1* (*SLC25A1*) expression as a compensatory response to an increase in intracellular ROS (46). Finally, *PCNA* encodes for a proliferating cell nuclear antigen, which has high expression in tumor tissues (47). Interestingly, carotenoids present various suppressive abilities against *PCNA* expressions in cell proliferation (48). Lycopene also decreased the positive rate of PCNA and protein expressions of PCNA in lung tissue (49). In summary, several hub genes were related to lipid metabolism, oxidative stress, or antioxidant action of carotenoids. Thus, the plausible link between carotenoids and lipid profile does not seem to be entirely due to the fact that carotenoids are transported by lipoproteins. Indeed, circulating carotenoids have been showed to impact TFs involved in lipid metabolism (50, 51). Accordingly, in the present study, plasma carotenoids modulated expression of several genes related, among others, to lipid metabolism. Moreover, classification terms were used to suggest potential causal relationships of the hub genes in the interconnection between total carotenoids and plasma lipid concentrations. A total of seven transcripts were classified as “likely regulator” as they have been associated with or shown to influence lipid levels in the literature and thus represent potential regulators of lipid concentrations. For example, *FAM13A*, which is associated with total carotenoids,

also demonstrated a regulatory effect on HDL-C via its genetic variations and effect on fatty acid oxidation (30-32). A total of 20 transcripts were classified as “likely regulated” meaning that they were associated with both total carotenoids and lipid concentrations in the present study. Finally, a total of 12 transcripts were classified as “unknown” considering they showed only associations with total carotenoids.

The present study has strengths, but also some limitations. The main strength results from the study of the combination of both genome-wide gene expression levels and total carotenoids, representing six predominant plasma carotenoids. To the best of our knowledge, this is the first study that considers the impact of plasma total carotenoids on genome-wide expression levels. Another strength is the use of WGCNA that adds important information about the effect of co-expression network of genes, which is useful to detect biological pathways related to lipid profile. On the other hand, the study’s main limitation resides in the small sample size, which limits statistical power. However, the study of a founder population with relatively homogeneous genetics and shared environment is a new aspect in this field (52). Nonetheless, this limits the generalization of results to other populations. Finally, our study did not account for diet, physical activity, smoking, and alcohol consumption of participants, all of which may affect circulating carotenoid concentrations (53, 54).

Conclusions

In conclusion, whole blood expression levels of 533 probes were associated with plasma total carotenoid concentrations. Using WGCNA, four modules were identified. A total of four and 29 hub genes associated with total carotenoids were potentially related to HDL-C and TC, respectively. This integrative analysis provides evidence for the potential role of gene co-expression in the relationship between carotenoids and lipid concentrations. Further studies and validation of the hub genes are needed. Finally, this article may also serve as an example of how to include a wide range of omics data in nutrition studies, using systems biology methods.

Ethics approval and consent to participate

All participants (adults and children) signed an informed consent document. Parental consent was also obtained by signing the child consent document. The experimental protocol was approved by the Ethics Committees of Laval University Hospital Research Center and Laval University.

Consent for publication

Not applicable

Availability of data and material

Expression datasets supporting the conclusion of this article are available in the Gene Expression Omnibus (GEO) database, www.ncbi.nlm.nih.gov/geo (GSE114620).

Competing interests

Authors have no potential conflicts of interest.

Funding

This work was supported by the Canada Research Chair in Genomics Applied to Nutrition and Metabolic Health. MCV is Tier 1 Canada Research Chair in Genomics Applied to Nutrition and Metabolic Health. BLT is a recipient of a scholarship from Canadian Institutes of Health Research (CIHR).

Authors' contributions

Each author contribution to work: BL and MCV designed research; BLT, FG, and LP conducted research and performed statistical analyses; BLT wrote the paper; BLT and MCV have primary responsibility for final content. All authors read and approved the final manuscript.

Acknowledgments

We would like to thank Christian Couture, Véronique Garneau, Catherine Raymond, and Véronique Richard who contributed to the success of this study.

References

1. Souverein OW, de Vries JH, Freese R, Watzl B, Bub A, Miller ER, et al. Prediction of fruit and vegetable intake from biomarkers using individual participant data of diet-controlled intervention studies. The British journal of nutrition. 2015;113(9):1396-409.

2. Couillard C, Lemieux S, Vohl MC, Couture P, Lamarche B. Carotenoids as biomarkers of fruit and vegetable intake in men and women. The British journal of nutrition. 2016;116(7):1206-15.

3. Maiani G, Caston MJ, Catasta G, Toti E, Cambrodon IG, Bysted A, et al. Carotenoids: actual knowledge on food sources, intakes, stability and bioavailability and their protective role in humans. Molecular nutrition & food research. 2009;53 Suppl 2:S194-218.

4. Eroglu A, Harrison EH. Carotenoid metabolism in mammals, including man: formation, occurrence, and function of apocarotenoids. Journal of lipid research. 2013;54(7):1719-30.

5. Allore T, Lemieux S, Vohl MC, Couture P, Lamarche B, Couillard C. Correlates of the difference in plasma carotenoid concentrations between men and women. The British journal of nutrition. 2018:1-10.

6. Farook VS, Reddivari L, Mummidi S, Puppala S, Arya R, Lopez-Alvarenga JC, et al. Genetics of serum carotenoid concentrations and their correlation with obesity-related traits in Mexican American children. The American journal of clinical nutrition. 2017;106(1):52-8.

7. Tremblay BL, Guenard F, Lamarche B, Perusse L, Vohl MC. Genetic and Common Environmental Contributions to Familial Resemblances in Plasma Carotenoid Concentrations in Healthy Families. Nutrients. 2018;10(8).

8. Wang Y, Chung SJ, McCullough ML, Song WO, Fernandez ML, Koo SI, et al. Dietary carotenoids are associated with cardiovascular disease risk biomarkers mediated by serum carotenoid concentrations. The Journal of nutrition. 2014;144(7):1067-74.

9. Clevidence BA, Bieri JG. Association of carotenoids with human plasma lipoproteins. Methods in enzymology. 1993;214:33-46.

10. Rao AV, Rao LG. Carotenoids and human health. Pharmacological research. 2007;55(3):207-16.

11. Paiva SA, Russell RM. Beta-carotene and other carotenoids as antioxidants. Journal of the American College of Nutrition. 1999;18(5):426-33.

12. Bertram JS. Carotenoids and gene regulation. Nutrition reviews. 1999;57(6):182-91.

13. D'Adamo CR, D'Urso A, Ryan KA, Yerges-Armstrong LM, Semba RD, Steinle NI, et al. A Common Variant in the SETD7 Gene Predicts Serum Lycopene Concentrations. Nutrients. 2016;8(2):82.

14. Beydoun MA, Nalls MA, Canas JA, Evans MK, Zonderman AB. Gene polymorphisms and gene scores linked to low serum carotenoid status and their associations with metabolic disturbance and depressive symptoms in African-American adults. The British journal of nutrition. 2014;112(6):992-1003.

15. Ferrucci L, Perry JR, Matteini A, Perola M, Tanaka T, Silander K, et al. Common variation in the beta-carotene 15,15'-monooxygenase 1 gene affects circulating levels of carotenoids: a genome-wide association study. American journal of human genetics. 2009;84(2):123-33.

16. Zubair N, Kooperberg C, Liu J, Di C, Peters U, Neuhouser ML. Genetic variation predicts serum lycopene concentrations in a multiethnic population of postmenopausal women. The Journal of nutrition. 2015;145(2):187-92.

17. Borel P, de Edelenyi FS, Vincent-Baudry S, Malezet-Desmoulin C, Margotat A, Lyan B, et al. Genetic variants in BCMO1 and CD36 are associated with plasma lutein concentrations and macular pigment optical density in humans. Annals of medicine. 2011;43(1):47-59.

18. Bohn T, Desmarchelier C, Dragsted LO, Nielsen CS, Stahl W, Ruhl R, et al. Host-related factors explaining interindividual variability of carotenoid bioavailability and tissue concentrations in humans. Molecular nutrition & food research. 2017;61(6).

19. Hix LM, Lockwood SF, Bertram JS. Bioactive carotenoids: potent antioxidants and regulators of gene expression. Redox report : communications in free radical research. 2004;9(4):181-91.

20. Sharoni Y, Danilenko M, Dubi N, Ben-Dor A, Levy J. Carotenoids and transcription. Archives of biochemistry and biophysics. 2004;430(1):89-96.

21. Langfelder P, Horvath S. WGCNA: an R package for weighted correlation network analysis. BMC bioinformatics. 2008;9:559.

22. Tremblay BL, Guenard F, Lamarche B, Perusse L, Vohl MC. Familial resemblances in blood leukocyte DNA methylation levels. Epigenetics. 2016;11(11):831-8.

23. Tremblay BL, Guenard F, Lamarche B, Perusse L, Vohl MC. Familial resemblances in human whole blood transcriptome. BMC genomics. 2018;19(1):300.

24. Tremblay BL, Guenard F, Lamarche B, Perusse L, Vohl MC. Familial resemblances in human plasma metabolites are attributable to both genetic and common environmental effects. Nutrition research. 2019;61:22-30.

25. Callaway C, Chumlea W, Bouchard C, Himes J, Lohman T, Martin A, et al. Standardization of anthropomeric measurements : The Airlie (VA) Consensus Conference. Champaign, IR, USA; 1988.

26. McNamara JR, Schaefer EJ. Automated enzymatic standardized lipid analyses for plasma and lipoprotein fractions. ClinChimActa. 1987;166(1):1-8.

27. Burstein M, Samaille J. [On a rapid determination of the cholesterol bound to the serum alpha- and beta-lipoproteins]. ClinChimActa. 1960;5:609.

28. Albers JJ, Warnick GR, Wiebe D, King P, Steiner P, Smith L, et al. Multi-laboratory comparison of three heparin-Mn2+ precipitation procedures for estimating cholesterol in high-density lipoprotein. ClinChem. 1978;24(6):853-6.

29. Friedewald WT, Levy RI, Fredrickson DS. Estimation of the concentration of low-density lipoprotein cholesterol in plasma, without use of the preparative ultracentrifuge. ClinChem. 1972;18(6):499-502.

30. Laurell CB. Quantitative estimation of proteins by electrophoresis in agarose gel containing antibodies. AnalBiochem. 1966;15(1):45-52.

31. Pirro M, Bergeron J, Dagenais GR, Bernard PM, Cantin B, Despres JP, et al. Age and duration of follow-up as modulators of the risk for ischemic heart disease associated with high plasma C-reactive protein levels in men. ArchInternMed. 2001;161(20):2474-80.

32. Blazejczyk M MM, Nadon R. FlexArray: A statistical data analysis software for gene expression microarrays. Genome Quebec, Montreal, Canada. 2007.

33. Team RC. R: A language and environment for statistical computing. R Foundation for Statistical Computing. Vienna, Austria. 2013.

34. Langfelder P, Horvath S. Fast R Functions for Robust Correlations and Hierarchical Clustering. Journal of statistical software. 2012;46(11).

35. Zhang B, Horvath S. A general framework for weighted gene co-expression network analysis. Statistical applications in genetics and molecular biology. 2005;4:Article17.

36. Langfelder P, Horvath S. Eigengene networks for studying the relationships between co-expression modules. BMC systems biology. 2007;1:54.

37. Horvath S, Zhang B, Carlson M, Lu KV, Zhu S, Felciano RM, et al. Analysis of oncogenic signaling networks in glioblastoma identifies ASPM as a molecular target. Proceedings of the National Academy of Sciences of the United States of America. 2006;103(46):17402-7.

38. Xie C, Mao X, Huang J, Ding Y, Wu J, Dong S, et al. KOBAS 2.0: a web server for annotation and identification of enriched pathways and diseases. Nucleic acids research. 2011;39(Web Server issue):W316-22.

39. Kuleshov MV, Jones MR, Rouillard AD, Fernandez NF, Duan Q, Wang Z, et al. Enrichr: a comprehensive gene set enrichment analysis web server 2016 update. Nucleic acids research. 2016;44(W1):W90-7.

40. Borel P, Desmarchelier C, Nowicki M, Bott R. A Combination of Single-Nucleotide Polymorphisms Is Associated with Interindividual Variability in Dietary beta-Carotene Bioavailability in Healthy Men. The Journal of nutrition. 2015;145(8):1740-7.

41. Borel P, Desmarchelier C, Nowicki M, Bott R. Lycopene bioavailability is associated with a combination of genetic variants. Free radical biology & medicine. 2015;83:238-44.

42. Borel P, Desmarchelier C, Nowicki M, Bott R, Morange S, Lesavre N. Interindividual variability of lutein bioavailability in healthy men: characterization, genetic variants involved, and relation with fasting plasma lutein concentration. The American journal of clinical nutrition. 2014;100(1):168-75.

43. Borel P, Desmarchelier C. Bioavailability of Fat-Soluble Vitamins and Phytochemicals in Humans: Effects of Genetic Variation. Annual review of nutrition. 2018;38:69-96.

44. Seo JY, Kim H, Seo JT, Kim KH. Oxidative stress induced cytokine production in isolated rat pancreatic acinar cells: effects of small-molecule antioxidants. Pharmacology. 2002;64(2):63-70.

45. Mueller E, Smith M, Sarraf P, Kroll T, Aiyer A, Kaufman DS, et al. Effects of ligand activation of peroxisome proliferator-activated receptor gamma in human prostate cancer. Proceedings of the National Academy of Sciences of the United States of America. 2000;97(20):10990-5.

46. Inoue Y, Shimazawa M, Nagano R, Kuse Y, Takahashi K, Tsuruma K, et al. Astaxanthin analogs, adonixanthin and lycopene, activate Nrf2 to prevent light-induced photoreceptor degeneration. Journal of pharmacological sciences. 2017;134(3):147-57.

47. Huang Q, Wang TH, Lu WS, Mu PW, Yang YF, Liang WW, et al. Estrogen receptor alpha gene polymorphism associated with type 2 diabetes mellitus and the serum lipid concentration in Chinese women in Guangzhou. Chinese medical journal. 2006;119(21):1794-801.

48. Willer CJ, Schmidt EM, Sengupta S, Peloso GM, Gustafsson S, Kanoni S, et al. Discovery and refinement of loci associated with lipid levels. Nature genetics. 2013;45(11):1274-83.

49. Spracklen CN, Chen P, Kim YJ, Wang X, Cai H, Li S, et al. Association analyses of East Asian individuals and trans-ancestry analyses with European individuals reveal new loci associated with cholesterol and triglyceride levels. Human molecular genetics. 2017;26(9):1770-84.

50. Jiang Z, Knudsen NH, Wang G, Qiu W, Naing ZZC, Bai Y, et al. Genetic Control of Fatty Acid beta-Oxidation in Chronic Obstructive Pulmonary Disease. American journal of respiratory cell and molecular biology. 2017;56(6):738-48.

51. Miao L, Yin RX, Pan SL, Yang S, Yang DZ, Lin WX. Weighted Gene Co-Expression Network Analysis Identifies Specific Modules and Hub Genes Related to Hyperlipidemia. Cellular physiology and biochemistry : international journal of experimental cellular physiology, biochemistry, and pharmacology. 2018;48(3):1151-63.

52. Khor CC, Vannberg FO, Chapman SJ, Guo H, Wong SH, Walley AJ, et al. CISH and susceptibility to infectious diseases. The New England journal of medicine. 2010;362(22):2092-101.

53. Gruaz L, Delucinge-Vivier C, Descombes P, Dayer JM, Burger D. Blockade of T cell contact-activation of human monocytes by high-density lipoproteins reveals a new pattern of cytokine and inflammatory genes. PloS one. 2010;5(2):e9418.

54. Lynch TLt, Sivaguru M, Velayutham M, Cardounel AJ, Michels M, Barefield D, et al. Oxidative Stress in Dilated Cardiomyopathy Caused by MYBPC3 Mutation. Oxidative medicine and cellular longevity. 2015;2015:424751.

55. Jorgenrud B, Jalanko M, Helio T, Jaaskelainen P, Laine M, Hilvo M, et al. The Metabolome in Finnish Carriers of the MYBPC3-Q1061X Mutation for Hypertrophic Cardiomyopathy. PloS one. 2015;10(8):e0134184.

56. Rudkowska I, Perusse L, Bellis C, Blangero J, Despres JP, Bouchard C, et al. Interaction between Common Genetic Variants and Total Fat Intake on Low-Density Lipoprotein Peak Particle Diameter: A Genome-Wide Association Study. Journal of nutrigenetics and nutrigenomics. 2015;8(1):44-53.

57. Lean CB, Lee EJ. Genetic variations of the MCT4 (SLC16A3) gene in the Chinese and Indian populations of Singapore. Drug metabolism and pharmacokinetics. 2012;27(4):456-64.

58. Bohn SK, Myhrstad MC, Thoresen M, Holden M, Karlsen A, Tunheim SH, et al. Blood cell gene expression associated with cellular stress defense is modulated by antioxidant-rich food in a randomised controlled clinical trial of male smokers. BMC medicine. 2010;8:54.

59. Bringman-Rodenbarger LR, Guo AH, Lyssiotis CA, Lombard DB. Emerging Roles for SIRT5 in Metabolism and Cancer. Antioxidants & redox signaling. 2018;28(8):677-90.

60. Liang F, Wang X, Ow SH, Chen W, Ong WC. Sirtuin 5 is Anti-apoptotic and Anti-oxidative in Cultured SH-EP Neuroblastoma Cells. Neurotoxicity research. 2017;31(1):63-76.

61. Yu J, Sadhukhan S, Noriega LG, Moullan N, He B, Weiss RS, et al. Metabolic characterization of a Sirt5 deficient mouse model. Scientific reports. 2013;3:2806.

62. Yu CY, Theusch E, Lo K, Mangravite LM, Naidoo D, Kutilova M, et al. HNRNPA1 regulates HMGCR alternative splicing and modulates cellular cholesterol metabolism. Human molecular genetics. 2014;23(2):319-32.

63. Echaniz-Laguna A, Chassagne M, Ceresuela J, Rouvet I, Padet S, Acquaviva C, et al. Complete loss of expression of the ANT1 gene causing cardiomyopathy and myopathy. Journal of medical genetics. 2012;49(2):146-50.

64. Won JC, Park JY, Kim YM, Koh EH, Seol S, Jeon BH, et al. Peroxisome proliferator-activated receptor-gamma coactivator 1-alpha overexpression prevents endothelial apoptosis by increasing ATP/ADP translocase activity. Arteriosclerosis, thrombosis, and vascular biology. 2010;30(2):290-7.

65. Wang X, Wang D, Yuan N, Liu F, Wang F, Wang B, et al. The prognostic value of PCNA expression in patients with osteosarcoma: A meta-analysis of 16 studies. Medicine. 2017;96(41):e8254.

66. Cheng HC, Chien H, Liao CH, Yang YY, Huang SY. Carotenoids suppress proliferating cell nuclear antigen and cyclin D1 expression in oral carcinogenic models. The Journal of nutritional biochemistry. 2007;18(10):667-75.

67. Donaldson MS. A carotenoid health index based on plasma carotenoids and health outcomes. Nutrients. 2011;3(12):1003-22.

68. Gervois P, Torra IP, Fruchart JC, Staels B. Regulation of lipid and lipoprotein metabolism by PPAR activators. Clinical chemistry and laboratory medicine. 2000;38(1):3-11.

69. Keller H, Dreyer C, Medin J, Mahfoudi A, Ozato K, Wahli W. Fatty acids and retinoids control lipid metabolism through activation of peroxisome proliferator-activated receptor-retinoid X receptor heterodimers. Proceedings of the National Academy of Sciences of the United States of America. 1993;90(6):2160-4.

70. Moreau C, Lefebvre JF, Jomphe M, Bherer C, Ruiz-Linares A, Vezina H, et al. Native American admixture in the Quebec founder population. PloS one. 2013;8(6):e65507.

71. Gruber M, Chappell R, Millen A, LaRowe T, Moeller SM, Iannaccone A, et al. Correlates of serum lutein + zeaxanthin: findings from the Third National Health and Nutrition Examination Survey. The Journal of nutrition. 2004;134(9):2387-94.

72. Al-Delaimy WK, van Kappel AL, Ferrari P, Slimani N, Steghens JP, Bingham S, et al. Plasma levels of six carotenoids in nine European countries: report from the European Prospective Investigation into Cancer and Nutrition (EPIC). Public health nutrition. 2004;7(6):713-22.

Tables

Table 7.1 Characteristic and biochemical parameters of study subjects

Biochemical parameters	Fathers (n=6)	Mothers (n=16)	Boys (n=18)	Girls (n=8)
Age (years)	42.0 ± 2.8	42.4 ± 6.0	11.9 ± 3.8	10.1 ± 2.0
BMI (kg/m^2)	24.8 ± 1.3	23.5 ± 3.4	-	-
BMI percentile	-	-	48.9 ± 32.7	52.1 ± 29.8
TC (mmol/L)	4.70 ± 0.61	4.67 ± 0.55	4.35 ± 0.53	4.15 ± 0.48
HDL-C (mmol/L) [1]	1.36 ± 0.33	1.73 ± 0.35	1.56 ± 0.28	1.54 ± 0.31
LDL-C (mmol/L)	2.83 ± 0.67	2.53 ± 0.49	2.36 ± 0.42	2.20 ± 0.52
ApoB100 (g/L)	0.89 ± 0.22	0.77 ± 0.11	0.72 ± 0.12	0.67 ± 0.15
TG (mmol/L)	1.13 ± 0.35	0.88 ± 0.34	0.94 ± 0.39	0.89 ± 0.41
Total carotenoids (µmol/L)	5.77 ± 2.93	6.57 ± 2.23	5.88 ± 2.37	5.30 ± 1.09

All values are means ± SD. [1] Means are significantly different ($P \leq 0.05$) between fathers and mothers. Abbreviations: Apolipoprotein B-100 (ApoB100), body mass index (BMI), high-density lipoprotein cholesterol (HDL-C), low-density lipoprotein cholesterol (LDL-C), standard deviation (SD), total cholesterol (TC), triglycerides (TG).

Table 7.2 Hub genes identified in the brown module

Probe ID	Gene (Chr, Accession number)	GS HDL-C	*P*-value GS	MM	*P*-value MM	Potential causal relationships of the hub genes[1]
ID_4920064	*CISH* (Chr3, NM_145071)	0.29	0.049	0.67	1.44×10^{-7}	Likely regulator
ID_6480731	*AMER3* (Chr2, NM_ 001105193)	0.36	0.011	0.68	1.12×10^{-7}	Likely regulated
ID_3870603	*ZSWIM4* (Chr19, NM_ 023072)	0.29	0.047	0.69	5.12×10^{-8}	Unknown
ID_290093	*LOC647322* (Chr2, withdrawn)	0.32	0.028	0.74	2.39×10^{-9}	Unknown
ID_4540376	*FAM13A* (Chr4, NM_ 001015045)	0.29	0.047	0.76	2.88×10^{-10}	Likely regulator
ID_2260315	*HS.572752*	0.29	0.047	0.77	1.48×10^{-10}	Unknown
ID_3420706	*HS.566359*	0.29	0.049	0.83	2.03×10^{-13}	Unknown

[1]Classification terms have been added to suggest potential causal relationships of the hub genes in the interconnection between total carotenoids and plasma lipid concentrations. Abbreviations: Chromosome (Chr), gene significance (GS), high-density lipoprotein cholesterol (HDL-C), module membership (MM).

Table 7.3 Hub genes identified in the blue module

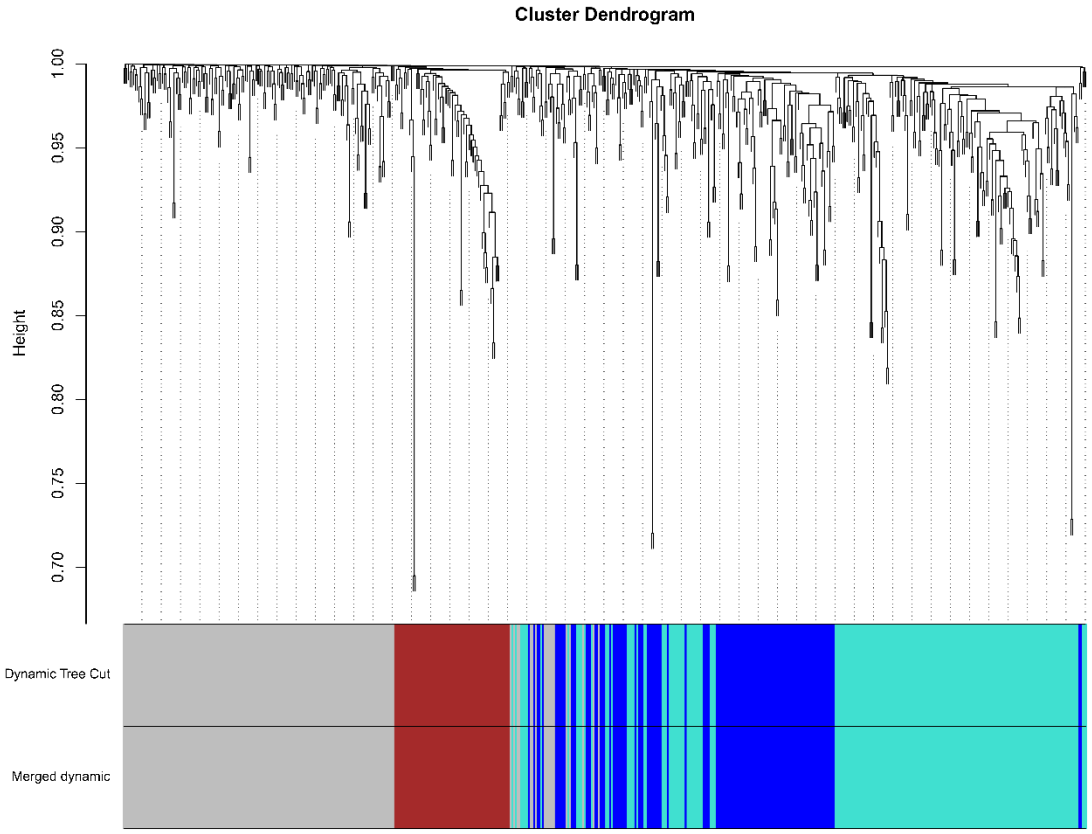
Probe ID	Gene (Chr, Accession number)	GS TC	*P*-value GS	MM	*P*-value MM	Potential causal relationships of the hub genes[1]
ID_2350019	*FEZ1* (Chr11, NM_ 022549)	0.31	0.032	-0.81	4.55×10^{-12}	Unknown
ID_6940768	*MYBPC3* (Chr11, NM_000256)	0.34	0.019	-0.80	1.06×10^{-11}	Likely regulator
ID_1440754	*OPRL1* (Chr20, NM_000913)	0.30	0.037	-0.70	3.76×10^{-8}	Likely regulator
ID_5870341	*LOC642691* (Chr1, withdrawn)	0.30	0.040	-0.69	5.85×10^{-8}	Unknown
ID_3780017	*SIRPB1* (Chr20, NM_006065)	0.40	0.0049	-0.68	1.01×10^{-7}	Likely regulated
ID_110719	*SLC16A3* (Chr17, NM_004207)	0.30	0.037	-0.62	2.45×10^{-6}	Likely regulator
ID_2120152	*SLC4A1* (Chr17, NM_000342)	0.31	0.030	-0.60	6.55×10^{-6}	Likely regulated
ID_2630739	*HS.154948*	-0.42	0.0032	0.60	6.37×10^{-6}	Likely regulated
ID_6040072	*GPR89C* (Chr1, NM_001097616)	-0.46	0.0012	0.61	4.10×10^{-6}	Likely regulated
ID_5810070	*PSMD6* (Chr3, NM_014814)	-0.36	0.012	0.62	3.19×10^{-6}	Likely regulated
ID_7000735	*RANBP1* (Chr22, NM_002882)	-0.38	0.0072	0.62	2.63×10^{-6}	Likely regulated
ID_6200437	*AP3M1* (Chr10, NM_207012)	-0.47	0.00076	0.62	2.35×10^{-6}	Likely regulated
ID_4670703	*SMIM13* (Chr6, XM_927609)	-0.35	0.013	0.62	2.29×10^{-6}	Likely regulated
ID_1510270	*SIRT5* (Chr6, NM_012241)	-0.29	0.042	0.63	1.48×10^{-6}	Likely regulator
ID_3890632	*DNM1L* (Chr12, NM_012063)	-0.32	0.028	0.63	1.46×10^{-6}	Unknown
ID_4810292	*CR2* (Chr1, NM_001877)	-0.42	0.0030	0.64	1.03×10^{-6}	Likely regulated

ID_2810246	*LBH* (Chr2, NM_030915)	-0.46	0.0011	0.64	9.22×10^{-7}	Likely regulated
ID_10209	*HNRNPA1* (Chr12, NM_031157)	-0.53	8.78×10^{-5}	0.64	7.83×10^{-7}	Likely regulator
ID_1090367	*BIVM* (Chr13, NM_017693)	-0.38	0.0085	0.66	3.64×10^{-7}	Likely regulated
ID_5560465	*NOP56* (Chr20, NM_006392)	-0.38	0.0079	0.67	1.70×10^{-7}	Likely regulated
ID_1990154	*PTMAP12* (Chr9, XM_001126659)	-0.31	0.035	0.67	1.51×10^{-7}	Unknown
ID_4010097	*FBXO5* (Chr6, NM_012177)	-0.42	0.0027	0.68	1.37×10^{-7}	Likely regulated
ID_7100367	*HS.571502*	-0.39	0.0057	0.69	7.78×10^{-8}	Likely regulated
ID_4850133	*SNRPN* (Chr15, NM_022806)	-0.37	0.0096	0.69	6.75×10^{-8}	Likely regulated
ID_7160719	*CCT7* (Chr2, NM_006429)	-0.47	0.00069	0.69	4.71×10^{-8}	Likely regulated
ID_5560504	*TIAL1* (Chr10, NM_003252)	-0.29	0.042	0.70	2.85×10^{-8}	Unknown
ID_6840747	*SLC25A4* (Chr4, NM_001151)	-0.32	0.028	0.72	7.82×10^{-9}	Unknown
ID_6900079	*PCNA* (Chr20, NM_182649)	-0.42	0.0032	0.74	2.27×10^{-9}	Likely regulated
ID_7150132	*PNN* (Chr14, NM_002687)	-0.31	0.031	0.74	2.23×10^{-9}	Unknown
ID_3170189	*S100PBP* (Chr1, NM_022753)	-0.30	0.038	0.76	4.91×10^{-10}	Unknown
ID_110059	*UBA5* (Chr3, NM_024818)	-0.39	0.0061	0.80	1.12×10^{-11}	Likely regulated
ID_4260093	*TUT4* (Chr1, NM_001009882)	-0.46	0.00096	0.81	2.10×10^{-12}	Likely regulated

[1]Classification terms have been added to suggest potential causal relationships of the hub genes in the interconnection between total carotenoids and plasma lipid concentrations Abbreviations: Chromosome (Chr), gene significance (GS), module membership (MM), total cholesterol (TC).

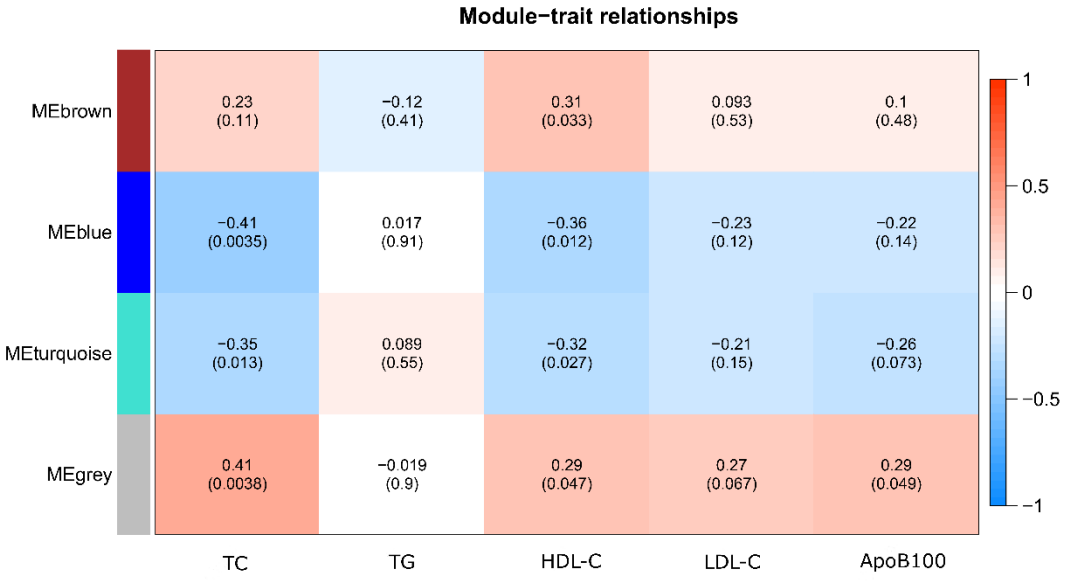
Figures

Figure 7.1 Gene dendrogram obtained using average linkage hierarchical clustering



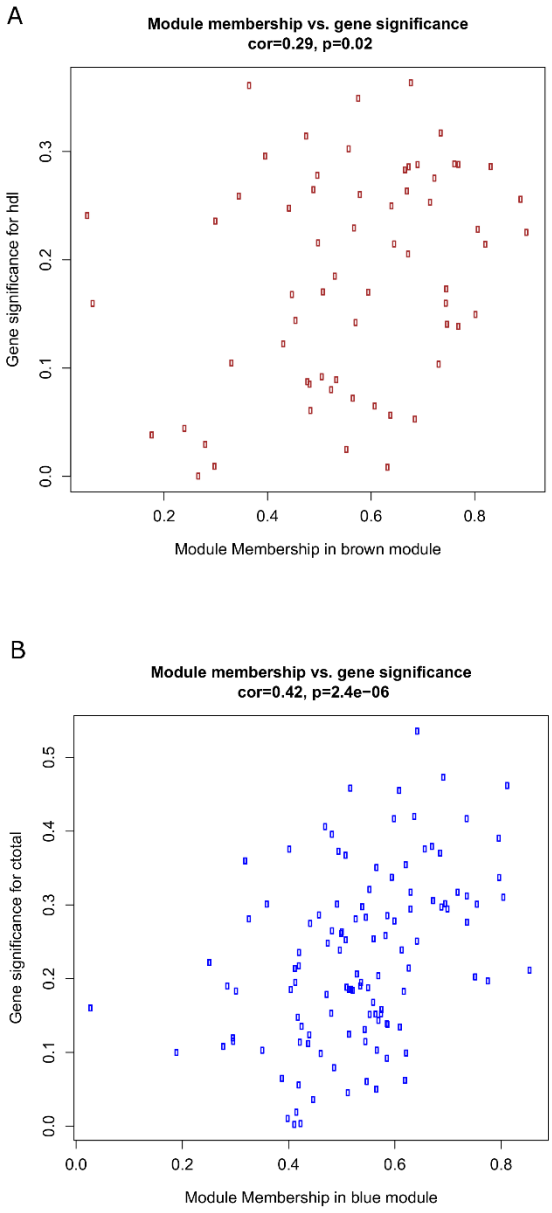
Four module colors are shown correspondingly. The merged dynamic yielded same modules as the dynamic tree cut using a cut-off of 0.25.

Figure 7.2 Heatmap of module-trait relationships depicting correlations between module eigengenes and lipid profile traits



Numbers in the table correspond to the correlation r and the *P*-value in parentheses. The degree of correlation is illustrated with the color legend.

Figure 7.3 Scatterplots of gene significance for lipid profile traits and module membership in the A) brown and B) blue modules



Supplemental material

Supplemental table 7.1 Concentrations of plasma carotenoids (µmol/L of plasma)

Subjects	Lutein	Zeaxanthin	β-Cryptoxanthin	α-Carotene	β-Carotene	Lycopene	Total carotenoids
Gen01	0.36689606	0.0678172	0.48942068	0.94815424	1.96626358	1.90477202	5.74332378
Gen02	0.19040767	0.05211724	0.2671793	0.39046788	1.63413354	2.97212935	5.50643497
Gen03	0.20500873	0.04703513	0.33127708	1.3088189	1.97765103	2.14359446	6.01338535
Gen05	0.24382521	0.05322967	0.42173556	0.80394106	2.58467386	2.80014081	6.90754618
Gen06	0.21138746	0.03997405	0.47272927	0.87987086	2.33627976	2.28340529	6.2236467
Gen07	0.15361658	0.03836336	0.20938586	1.01423133	1.62869228	1.21588093	4.26017034
Gen08	0.13950888	0.03245329	0.2880089	0.56440081	1.07092427	1.52477859	3.62007474
Gen09	0.18935861	0.06023487	0.24183803	0.36889688	1.46894869	1.84769435	4.17697143
Gen10	0.45531524	0.05366406	0.68650319	1.86816097	5.09892831	1.60505358	9.76762535
Gen11	0.13301695	0.0379158	0.40775388	1.17280733	1.78868368	1.15260911	4.69278675
Gen12	0.1495655	0.03782232	0.52583487	1.20447	2.13195549	1.51354258	5.56319076
Gen13	0.22080196	0.04605841	0.40370279	1.95125405	4.36470233	1.77749248	8.76401203
Gen14	0.19439964	0.0346023	0.47240977	1.69428919	2.77308628	1.17814005	6.34692723
Gen15	0.31456871	0.07906164	0.64207259	0.585099	1.26299619	2.07400924	4.95780737
Gen16	0.20941464	0.0590733	0.66509735	0.43477479	1.2177428	1.90799437	4.49409726
Gen17	0.31659735	0.05824728	0.43726324	3.07504388	4.90271428	1.52411709	10.3139831
Gen18	0.40720388	0.06879612	0.52897983	1.77242233	2.73849116	1.40131048	6.9172038
Gen19	0.21406127	0.04334581	0.21302691	1.46073827	2.3907969	0.81416873	5.13613789
Gen20	0.38953299	0.11518446	0.54511458	0.58599723	1.94027507	1.46287634	5.03898068
Gen21	0.25823509	0.05614234	0.5515715	3.29451982	4.1048887	1.59654332	9.86190078
Gen22	0.19637546	0.05178252	0.52029086	1.43217409	2.81824867	1.43588238	6.45475399
Gen23	0.19796017	0.03400027	0.36306339	0.6257273	1.36047476	1.52337096	4.10459685
Gen24	0.23360289	0.05076413	0.42062581	1.11325177	1.89522389	1.31461194	5.02808043
Gen25	0.19012552	0.06455947	0.73073545	0.68144593	1.40988861	0.8309107	3.90766568
Gen26	0.20673794	0.04472159	0.11642864	0.1033121	0.30102258	0.5048136	1.27703645

Gen27	0.23324584	0.07133053	0.53723883	0.38361577	0.77381218	0.77033291	2.76957605
Gen28	0.22858608	0.06703344	0.40164353	0.34477048	0.57881226	0.69589933	2.31674512
Gen29	0.22569969	0.06288219	0.19197018	0.79944354	1.42605149	1.67944118	4.38548827
Gen30	0.18846852	0.05111517	0.18449405	1.76125731	3.75436001	1.36074407	7.30043914
Gen31	0.1034949	0.03639765	0.23772312	0.55324807	1.51238711	1.61846353	4.06171437
Gen32	0.1379356	0.05037577	0.58562149	1.28098545	2.66908747	1.96597652	6.6899823
Gen33	0.34863636	0.06872228	0.59045342	1.05182969	2.8442326	1.16548687	6.06936121
Gen34	0.34080966	0.08200735	0.53132439	0.87811067	3.02914672	1.32128546	6.18268426
Gen35	0.18558526	0.05637978	0.69489283	1.20488219	2.06864071	1.10169917	5.31207994
Gen36	0.21559215	0.1016761	0.82701707	2.65607443	4.85041133	1.40830847	10.0590796
Gen38	0.26706147	0.09209016	.	3.55025825	6.33803969	1.79779113	12.0452407
Gen39	0.26395015	0.06463881	0.53649644	1.52282902	3.89619121	1.72663359	8.01073921
Gen40	0.22008198	0.04334325	0.45042163	1.29773283	3.99433146	2.71061293	8.71652407
Gen41	0.33789169	0.08950376	0.40316064	2.05426427	3.12184275	1.53282075	7.53948386
Gen42	0.16596505	0.05910571	0.43107599	0.74232185	1.36939518	1.38768457	4.15554834
Gen43	0.18579005	0.05817576	0.6537702	0.9184596	2.08512821	1.28347111	5.18479494
Gen44	0.26494297	0.09297957	0.90039749	1.52573514	3.79402374	2.5744044	9.15248331
Gen45	0.49521561	0.05468624	0.35857692	1.28192912	3.42957551	1.58930011	7.2092835
Gen46	0.72653971	0.09196268	0.47041018	1.0846667	1.94897283	1.42895167	5.75150378
Gen47	0.5848228	0.06066084	0.40081868	1.18038703	2.54237772	1.59877117	6.36783824
Gen48	0.55555079	0.07599738	0.5545489	0.91123368	1.78155749	1.2136722	5.09256043
Gen49	0.21725674	0.05761664	0.4287788	0.67397551	2.02502444	1.58549025	4.98814238
Gen50	0.22004146	0.03996355	0.286557	0.66712712	1.28660567	1.18069892	3.68099372

Supplemental table 7.2 Probes showing significant association with plasma total carotenoids (n=533)

Supplemental table 7.3 Functional enrichment analysis in the brown module using GO and Reactome databases

Supplemental table 7.4 Functional enrichment analysis in the blue module using GO and Reactome databases

Supplemental tables 7.2 to 7.4 are available at https://genesandnutrition.biomedcentral.com/articles/10.1186/s12263-019-0639-5.

Supplemental table 7.5 Significant *P*-values of regressions between expression levels of hub genes and plasma lipid concentrations

Probe ID	TC	HDL-C
Brown module		
ID_4920064		0.0506
ID_6480731		0.0106
ID_3870603		0.0561
ID_290093		0.0568
ID_4540376		0.0465
ID_2260315		0.0689
ID_3420706		0.0897
Blue module		
ID_2350019	0.1002	
ID_6940768	0.0587	
ID_1440754	0.1083	
ID_5870341	0.1202	
ID_3780017	0.0184	
ID_110719	0.1007	
ID_2120152	0.0933	
ID_2630739	0.0134	
ID_6040072	0.0053	
ID_5810070	0.0426	
ID_7000735	0.0215	
ID_6200437	0.0024	
ID_4670703	0.0485	
ID_1510270	0.1286	
ID_3890632	0.0921	
ID_4810292	0.0127	
ID_2810246	0.0052	
ID_10209	0.0005	
ID_1090367	0.0311	

ID_5560465	0.0279	
ID_1990154	0.1097	
ID_4010097	0.0117	
ID_7100367	0.0228	
ID_4850133	0.0303	
ID_7160719	0.0032	
ID_5560504	0.122	
ID_6840747	0.0827	
ID_6900079	0.0121	
ID_7150132	0.0948	
ID_3170189	0.1178	
ID_110059	0.0243	
ID_4260093	0.004	

Abbreviations: High-density lipoprotein cholesterol (HDL-C), Total cholesterol (TC).

Chapitre 8 Analyse de réseaux de co-méthylation dans la relation entre les caroténoïdes plasmatiques et le profil lipidique

Bénédicte L. Tremblay, Frédéric Guénard, Benoît Lamarche, Louis Pérusse, Marie-Claude Vohl

L'article présenté dans ce chapitre s'intitule : *Network analysis of the potential role of DNA methylation in the relationship between plasma carotenoids and lipid profile*.

Cet article est publié dans la revue : *Nutrients. 2019 Jun 4;11(6):1265*

Résumé

Introduction : La variabilité des caroténoïdes plasmatiques peut être attribuable à plusieurs facteurs, dont les variations génétiques et le profil lipidique. Jusqu'à présent, l'impact de la méthylation de l'ADN sur cette variabilité n'a pas été largement étudié. L'analyse de réseaux de corrélation pondérées (WGCNA) est une méthode de la biologie des systèmes utilisée pour identifier des groupes de gènes (modules) avec des niveaux de méthylation hautement corrélés pour ensuite les associer à des traits phénotypiques. L'objectif de la présente étude était d'examiner le rôle de la méthylation de l'ADN dans la relation entre les concentrations plasmatiques de caroténoïdes totaux et le profil lipidique en utilisant WGCNA chez 48 sujets en santé.

Résultats : Les niveaux de méthylation de l'ADN à l'échelle du génome dans les leucocytes sanguins de 20 687 sur 472 245 sites CpG étaient associés aux concentrations de caroténoïdes totaux. En utilisant WGCNA, neuf modules de co-méthylation ont été identifiés. Un total de 2734 *hub genes* (17 top *hub genes*) étaient potentiellement associés au profil lipidique.

Conclusions : Cette étude fournit des preuves de l'implication potentielle de la co-méthylation des gènes dans la relation entre les caroténoïdes plasmatiques et le profil lipidique. D'autres études et la validation des *hub genes* sont nécessaires.

Abstract

Introduction: Variability in plasma carotenoids may be attributable to several factors including genetic variants and lipid profile. Until now, the impact of DNA methylation on this variability has not been widely studied. Weighted gene correlation network analysis (WGCNA) is a systems biology method used for finding gene clusters (modules) with highly correlated methylation levels and for relating them to phenotypic traits. The objective of the present study was to examine the role of DNA methylation in the relationship between plasma total carotenoid concentrations and lipid profile using WGCNA in 48 healthy subjects.

Results: Genome-wide DNA methylation levels of 20 687 out of 472 245 CpG sites in blood leukocytes were associated with total carotenoid concentrations. Using WGCNA, nine co-methylation modules were identified. A total of 2734 hub genes (17 unique top hub genes) were potentially related to lipid profile.

Conclusions: This study provides evidence for the potential implication of gene co-methylation in the relationship between plasma carotenoids and lipid profile. Further studies and validation of the hub genes are needed.

Title page

Network analysis of the potential role of DNA methylation in the relationship between plasma carotenoids and lipid profile

Bénédicte L. Tremblay[1, 2], Frédéric Guénard[1, 2], Benoît Lamarche[1, 2], Louis Pérusse[1, 3], Marie-Claude Vohl[1,2]*

1. Institute of Nutrition and Functional Foods (INAF), Laval University, 2440 Hochelaga Blvd, Quebec City, QC, G1V 0A6, Canada.
2. School of Nutrition, Laval University, 2425 rue de l'Agriculture, Quebec City, QC, G1V 0A6, Canada.
3. Department of Kinesiology, Laval University, 2300 rue de la Terrasse, Quebec City, QC, G1V 0A6, Canada.

Corresponding author*:
Marie-Claude Vohl, Ph.D.
Institute of Nutrition and Functional Foods (INAF), Laval University
Pavillon des Services, 2440 Hochelaga Blvd, Quebec City, QC G1V 0A6, Canada
Phone: 418 656-2131 ext. 4676; Fax: 418 656-5877
E-mail: marie-claude.vohl@fsaa.ulaval.ca

Keywords**:** Carotenoids; DNA methylation; French Canadians; Hub genes; Lipid profile; WGCNA.

Introduction

Cardiometabolic (CM) diseases comprise conditions ranging from insulin resistance, metabolic syndrome to cardiovascular disease and type 2 diabetes (1). Healthy eating, including consumption of fruits and vegetables, is associated with a favorable CM health (2). Carotenoids, which are reliable biomarkers of fruit and vegetable intakes, are composed of hundreds of fat-soluble pigments (3). However, six main carotenoids (α-carotene, β-carotene, β-cryptoxanthin, lutein, lycopene, and zeaxanthin) represent over 95% of total circulating carotenoids in human (4, 5). Variability among individuals in circulating carotenoids may be due to several factors, which include age, sex, body weight, genetics, and lipid profile (3). Lower circulating carotenoid concentrations are associated with lower plasma total cholesterol (TC), LDL-cholesterol (LDL-C), and HDL-cholesterol (HDL-C) concentrations (6). Moreover, α- and β-carotene correlated with HDL-C and triglyceride (TG) concentrations (7), while β-crytoxanthin and zeaxanthin correlated with TG concentrations in a previous study by our group (8). Thus, several studies have observed associations between circulating carotenoids and plasma lipid concentrations, which are both transported by lipoproteins (9).

Genetic variants have been shown to influence circulating carotenoid concentrations, by causing differences in the absorption, assimilation, distribution, metabolism, and excretion of carotenoids (10-15). Carotenoids and their derivatives (e.g., retinoid) have also been shown to modulate gene expression via several transcriptional systems (16). However, very few studies have documented the relationship between carotenoids and DNA methylation, at least from a genome-wide viewpoint. In cell studies, lycopene had modest to null effects on DNA methylation of *GSTP1*, which is involved in prostate and breast cancers (17, 18). Moreover, a study of 165 overweight and obese subjects revealed an association between DNA methylation of *HERV-w* and *TNFα* in blood leukocytes and dietary intakes of β-carotene and carotenoids (19). Interestingly, the reported protective effects of high fruit and vegetable intake on age-related diseases (coronary heart disease, stroke, type 2 diabetes) may be mediated by the association between blood carotenoids levels and extrinsic epigenetic age acceleration, which is a measure of epigenetic age (20). Thus, circulating carotenoids and dietary intakes of carotenoids seem to impact DNA methylation. Moreover, there is more and more evidence that DNA methylation plays a role in the regulation of blood lipid levels and lipid metabolism-linked phenotypes and diseases (21). No previous study has considered the involvement of genome-wide DNA methylation levels in the association between plasma carotenoids and lipid profile. Weighted gene correlation network analysis (WGCNA) is a widely used systems biology approach designed for high dimensional data (i.e., gene expression, DNA methylation, metabolites etc.) (22). It allows finding gene clusters (modules) with highly correlated DNA methylation levels, relating these modules to phenotypic traits, and identifying key hub genes within modules that are related to phenotypic traits (22).

The objective of the present study was to examine the role of DNA methylation in the relationship between plasma carotenoid concentrations and lipid profile using WGCNA in 48 healthy subjects. First, CpG sites

whose DNA methylation levels are associated with carotenoid concentrations were identified using linear regressions. Second, WGCNA was used to identify specific modules and key hub genes related to lipid profile traits. The hypotheses were that genome-wide DNA methylation levels in blood leukocytes are associated with plasma total carotenoid concentrations and that clusters of genes associated with carotenoids are also correlated to lipid profile traits. Our results highlighted the potential implication of gene co-methylation in the relationship between plasma carotenoids and lipid profile.

Methods

Patients and design

The GENERATION Study aimed at evaluating familial resemblances in omics (DNA methylation (23) and gene expression (24)) and metabolic (metabolites (25) and carotenoids (8)) profiles in healthy families. The GENERATION Study comprises a total of 48 French-Canadian subjects from 16 families of Quebec City (Canada). Families composed of 16 mothers, six fathers, and 26 children lived under the same roof. Families comprised at least the mother and one child (8-18 years old). Parents had to be the biological parents of their child (or children), non-smokers, with body mass index (BMI) ranging between 18 and 35 kg/m^2, and free of any metabolic conditions requiring treatment (Synthroid® (levothyroxine) and oral contraceptive were tolerated). Children also had to be non-smokers, in good general health and not using psycho-stimulators [Ritalin® (methylphenidate), Concerta® (methylphenidate), and Strattera® (atomoxetine)]. Blood samples were taken from both parents and children at the Institute of Nutrition and Functional Foods (INAF). The experimental protocol was approved by the Ethics Committees of Laval University Hospital Research Center and Laval University. All participants signed an informed consent document. Parental consent was also obtained by signing the child consent document.

Anthropometric and CM measurements

Body weight and height were measured (26). Blood samples were collected from an antecubital vein into vacutainer tubes containing EDTA after 12-hour fast and 48-hour alcohol abstinence. Plasma was separated by centrifugation (2 500g for 10 min at 4°C), and samples were aliquoted and frozen (-80°C). Plasma TC and TG concentrations were measured using enzymatic assays (27, 28). Precipitation of very-low density lipoprotein and LDL particles in the infranatant with heparin manganese chloride generated the HDL-C fraction for measurements of HDL levels (29). LDL-C was estimated using the Friedewald formula (30). The rocket immunoelectrophoretic method was used to measure plasma apolipoprotein B-100 (ApoB100) concentrations (31). Plasma C-reactive protein (CRP) was measured by nephelometry using a sensitive assay (Prospec equipment Behring) (32). Fasting plasma glucose concentrations were enzymatically measured (33). Radioimmunoassay with polyethylene glycol separation was used to measure fasting plasma insulin concentrations (34).

DNA extraction and methylation analysis

Genomic DNA was extracted from blood leukocytes using the GenElute Blood Genomic DNA kit (Sigma-Aldrich, St. Louis, MO, USA) in all 48 subjects. NanoDrop Spectrophotometer (Thermo Scientific, Wilmington, DE, USA) and PicoGreen DNA methods were used to quantify DNA. Infinium Human Methylation 450 array (Illumina, San Diego, CA, USA) was used to measure DNA methylation levels. McGill University and Genome Quebec Innovation Center (Montreal, Canada) proceeded to the bisulfite conversion and quantitative DNA methylation analysis. Methylation data on all 485 577 CpG sites were analyzed using Illumina GenomeStudio software v2011.1 and the Methylation Module. All samples were retained after

quality control steps (23). GenomeStudio was used to perform global normalisation using control probes. Probes with a detection *P*-value > 0.01 in more than 5 subjects (>10% of all subjects) were removed. We restricted our analysis to autosomes and multi-mapped probes were also excluded (35). Thus, 472 245 probes were considered in the analysis.

Carotenoid measurements

Samples and standards preparation have been reported previously (8). Briefly, 100 µL of fasting plasma samples were thawed a few hours before analysis and transferred to Eppendorf tubes with 20 µL of 2-propanol, and 20 µL of carotenoid standard. Samples were transferred on a 400µL fixed well plate (ISOLUTE® SLE+, Biotage, Charlotte, NC) with 900 µL of hexane:isopropanol (90/10, *v/v*) in each well. Each extracted sample was evaporated under nitrogen and reconstituted with 300 µL of methanol:dichloromethane (65/35, *v/v*). Plates were shaken for 10 minutes and samples were transferred into high performance liquid chromatography glass vials.

Agilent 1260 liquid handling system (Agilent, Mississauga, Ontario, Canada) was used to perform high performance liquid chromatography (HPLC)-UV analysis as previously described (8). Carotenoids were separated with a mobile phase consisting of methanol:water (98/2, *v/v*; Eluent A) and methyl-tert-butyl ether (MTBE; Eluent B; VWR, Mississauga, Ontario, Canada). Flow-rate and gradient elution are detailed in a previous paper (8). UV detector was set at 450 nm and identification of each compound was confirmed using retention time and UV spectra (190-640nm) of the pure compounds. Chemstation software (Agilent, Mississauga, Ontario, Canada) was used to carry out data acquisition. The carotenoid concentrations are reported in µmol/L of plasma. One value for β-crypoxanthin was considered an outlier (defined as a value falling outside of the mean ± 4 standard deviations) and was excluded from analyses. The exclusion of this outlier is likely to have very little impact on further analysis considering that it affected only one subject and the mean of total carotenoids remained very similar (6.00 µmol/L without the outlier and 6.04 µmol/L including the outlier).

Association between DNA methylation levels and total carotenoid concentrations

Concentrations of plasma carotenoids are available in Supplementary Table S1. Plasma total carotenoid concentrations (µmol/L of plasma) were defined as the sum of α-carotene, β-carotene, β-cryptoxanthin, lutein, lycopene, and zeaxanthin concentrations. Regressions between normalised DNA methylation levels of all 472 245 CpG sites and total carotenoid concentrations adjusted for the family ID were computed using R software v2.14.1 (R Foundation for Statistical Computing; http://www.r-project.org) (36). Plasma total carotenoid concentrations (independent variable) were used to predict DNA methylation levels (dependent variable). A *P*-value ≤ 0.05 was used to identify significant associations. Regressions were adjusted for family ID (fixed effect) to account for the familial structure. In order to maintain a more exploratory approach, the regressions were not adjusted for other confounding factors. Choices of linear model and confounding factors were made for comparison purposes with similar study by our group (37). Weighted gene correlation

network analysis (WGCNA) was performed with methylation levels of 20 687 CpG sites showing a significant association (*P*-value ≤ 0.05) with total carotenoid concentrations. This allowed evaluating co-methylation similarities only in CpG sites that are associated with carotenoids.

Weighted gene correlation network analysis (WGCNA)

WGCNA was performed with the WGCNA package (22, 38) in R software using default parameters (36). A weighted adjacency matrix was established by calculating Pearson correlations between each CpG site pair. The co-methylation similarity matrix was raised to a power β = 5 (R^2 of 0.9570) to calculate weighted adjacency matrix (39) (Supplementary Figure 1). From this matrix, values were used to construct a topological overlap matrix (TOM), which provided a similarity measure. The TOM is then used to calculate the corresponding dissimilarity (1-TOM). Next, CpG sites with coherent methylation profiles were grouped into modules using the average linkage hierarchical clustering applied to the TOM-based dissimilarity (39). The dynamic tree cutting algorithm (function = cutreeDynamic, deep split = 0, minimum number of genes per module = 40, cut height = 0.25) was used to detect methylation modules (clusters of densely interconnected CpG sites in terms of co-methylation). The assignment of outlying genes to modules was performed using the Partitioning Around Medoids (PAM) method. DNA methylation levels of CpG sites within a module were summarized with the module eigengene (ME) value, which is the overall methylation level of CpG sites clustering in a module. To identify modules that were significantly associated with lipid profile traits (TC, LDL-C, HDL-C, TG, ApoB100), correlation between MEs (40) and traits were computed. Associations of individual genes with lipid profile traits was quantified using gene significance (GS), defined as the absolute correlation between the gene and the trait. Module membership (MM), defined as the correlation of the ME and the gene methylation profile, was used to quantify the similarity of all genes to each module. Correlations between MM and GS were computed to identify modules of interest. We considered CpG sites with the highest MM and GS to be those of highest biological relevance (41). CpG sites with a GS >0.2, a MM >0.8, and a *P*-value ≤ 0.05 were retained as hub genes. Due to the high number of hub genes obtained (n=2734), genes were subsequently ordered according to their GS with the trait. A total of three top annotated hub genes were selected in each six module-lipid association, for a total of 18 top hub genes. VisANT 5.0 was used to construct and visualize the topological interaction network in the black, turquoise, and pink modules (42). VisANT is a software framework for mining, analyzing and visualizing hierarchical organization of biological networks (43). Weighted correlation cut-offs of 0.23, 0.44, and 0.27 were used for the black, turquoise, and pink modules, respectively. The cut-offs were chosen to obtain visually interpretable networks, without being too heavily loaded. Central genes were those with the most gene-gene connections. Enrichr (http://amp.pharm.mssm.edu/Enrichr/) was used to obtain KEGG pathways 2019 and Gene Ontology molecular functions 2018 associated with genes in the black, turquoise, and pink modules (44).

Results

Characteristics and biochemical parameters

Characteristics and biochemical parameters including plasma total carotenoid concentrations of the parents and children are presented in Table 1. Plasma concentrations of the six main carotenoids (α-carotene, β-carotene, β-cryptoxanthin, lutein, lycopene, and zeaxanthin) and of total carotenoids concentrations measured in each participant are presented in Supplementary Table S1.

Weighted gene correlation network analysis (WGCNA)

Normalised methylation levels of all 472 245 CpG sites were tested for associations with plasma total carotenoid concentrations using regressions. A total of 20 687 CpG sites were significantly associated ($P \leq 0.05$) with total carotenoids (Supplementary Table S2). This subset of CpG sites associated with total carotenoids was then related to lipid profile traits using WGCNA.

A total of nine distinct modules were identified from methylation levels of the 20 687 CpG sites associated with total carotenoid concentrations using a dynamic tree cutting algorithm (Figure 1). Using a cut-off height of 0.25 for the clustering of module eigengenes (MEs) (merged dynamic), none of the modules has merged (Figure 1). The black, blue, brown, green, grey, pink, red, turquoise, and yellow modules contained 431, 3308, 3235, 2096, 1479, 243, 482, 7119, and 2294 CpG sites, respectively. The 1479 uncorrelated CpG sites were assigned to the grey module, which was excluded from further analysis. Correlations between MEs and lipid profile traits (TC, LDL-C, HDL-C, TG, ApoB100) were computed to find modules of interest. According to the heatmap of module-trait correlations (Figure 2), ME of the black module (431 CpG sites) had a negative correlation with HDL-C (r=-0.34 *P*=0.017). The ME of the turquoise module (7119 CpG sites) had negative correlations with TC (r=-0.46 *P*=0.0011), HDL-C (r=-0.30 *P*=0.036), LDL-C (r=-0.32 *P*=0.028), and ApoB100 (r=-0.34 *P*=0.018). The ME of the pink module (243 CpG sites) had a negative correlation with HDL-C (r=-0.29 *P*=0.048).

Gene significance (GS) (i.e., correlation between CpG site methylation and lipid profile traits) was correlated to module membership (MM) (i.e., correlation of the ME and the CpG site methylation profile). In the black module, GS for HDL-C had a significant correlation with the MM (r=0.25 $P=1.5 \times 10^{-7}$) (Supplementary Figure 2). In the turquoise module, GS for TC, HDL-C, LDL-C, and ApoB100 had significant correlations with the MM (r=0.46 $P < 1 \times 10^{-200}$, r=0.17 $P=2.6 \times 10^{-47}$, r=0.41 $P < 1 \times 10^{-200}$, and r=0.43 $P < 1 \times 10^{-200}$, respectively) (Supplementary Figure 2). In the pink module, GS for HDL-C had a significant correlation with the MM (r=0.37 $P=2.7 \times 10^{-9}$) (Supplementary Figure 2). Hence, the black, turquoise, and pink modules were considered as modules of interest for subsequent analyses.

Hub gene analysis

Hub gene analysis in black, turquoise, and pink modules was conducted to explore the underlying mechanism behind the associations observed. A total of 17, 2700, and 17 hub genes were identified in the

black, turquoise, and pink modules, respectively (Supplementary Table S3). Due to their high number, hub genes were subsequently ordered according to their GS with the trait. Top three most significant hub genes were selected in each module-lipid association, for a total of 18 top hub genes (Table 2). In the black module, *PAX4, TBC1D16*, and *PGA5* were the top three hub genes associated with HDL-C. In the turquoise module, *RBL2, GRIN3A*, and *TEX2* were the top three hub genes associated with TC, while *LST1, GOSR1*, and *GBGT1* were associated with HDL-C, *C7orf50, PDPK1, SYT17* with LDL-C, and finally *GABBR1, RAG2*, and *C7orf50* were associated with ApoB100 (Table 2). In the pink module, *RNASE11, TRIM68*, and *DEPDC1* were the top three hub genes associated with HDL-C (Table 2).

Topological interaction networks

Topological interaction networks were used to illustrate gene-gene interactions within the black, turquoise, and pink modules. In order not to overload networks, weighted correlation cut-offs were used. In the black module containing 431 CpG sites, only annotated CpG sites with a weighted correlation of 0.23 were included (n=14 unique genes) (Figure 3). In the turquoise module containing 7119 CpG sites, only annotated CpG sites with a weighted correlation of 0.44 were included (n=20 unique genes) (Figure 4). In the pink module containing 243 CpG sites, only annotated CpG sites with a weighted correlation of 0.27 were included (n=13 unique genes) (Figure 5). Gene Ontology molecular functions and KEGG pathways of genes in the black, turquoise, and pink modules are presented in Supplementary Table S4.

Discussion

First, to identify CpG sites associated with carotenoids, associations between genome-wide DNA methylation levels and plasma total carotenoid concentrations were tested. A total of 20 687 CpG sites were significantly associated with total carotenoid concentrations. To the best of our knowledge, this is the first study that considered the association of plasma total carotenoid concentrations with genome-wide DNA methylation levels. Until now, studies have been mainly focusing on the effect of genetic variants (SNPs) on the variability in carotenoid concentrations (45-47). Circulating carotenoids have also been shown to have an impact on gene expression via transcription factors (16). However, very few studies have considered the impact of carotenoids on DNA methylation. In two cell studies, lycopene had modest to null effects on DNA methylation of *GSTP1*, which is involved in prostate and breast cancers (17, 18). In addition, a study of 165 overweight and obese subjects observed positive associations between *HERV-w* methylation and dietary intakes of β-carotene and carotenoids, while *TNFα* methylation showed negative associations with dietary intakes of β-carotene and carotenoids (19). Moreover, blood carotenoids levels were also associated with extrinsic epigenetic age acceleration (20). Thus, the relationship between carotenoids and DNA methylation has not been widely studied using pan-genomic tool. A potential mechanism underlying the action of carotenoids on DNA methylation is the inhibition of the DNA methyltransferase (DNMT). Studies have shown that carotenoids (lycopene and astaxanthin) decreased DNMT activity and protein level in vitro (48, 49).

Secondly, modules and key hub genes related to lipid profile traits in the subset of 20 687 CpG sites associated with carotenoid concentrations were identified using WGCNA. Among nine modules identified, MEs of the black and pink modules had a significant correlation with HDL-C, while ME of the turquoise module had a significant correlation with TC, HDL-C, LDL-C, and ApoB100. Thus, highly co-methylated genes in the black, turquoise, and pink modules may have potential interactions or biological effects on lipid concentrations. Moreover, these modules also showed positive correlations between GS for several lipid profile traits and the MM of the respective module. This suggests that genes showing a strong association with lipid profile traits were also the most important elements of their respective module. The analysis with hub genes allowed getting insights on the potential mechanisms linking carotenoid concentrations and lipid profile. Three top hub genes were selected in each of the six module-lipid associations. Among these 18 top hub genes (17 unique top hub genes), several were related to lipid profile. First, cg07665923 is located in *C7orf50*. Several SNPs within this gene have been associated with TC. Indeed, SNP rs1997243 was associated with TC in a genome-wide association study on 188 577 individuals (50). Moreover, another SNP (rs6951245) has been shown to be a novel pleiotropic locus for both CRP and TC (51). SNP rs11763835 is also a peak *cis*-microRNA-eQTL associated with TC (52). Second, cg02500883 is located in *GRIN3A*, encoding for a glutamate ionotropic receptor NMDA type subunit 3A. SNPs within this gene have been associated with circulating lipid concentrations (53, 54). Interestingly, treatment with astaxanthin, a xanthophyll carotenoid, in rat neurons reduced *GRIN3A* gene expression (55). Third, cg21671607 is located in *RAG2*, encoding for the recombination activating 2. Variations in this gene are associated with several

clinical phenotypes going from severe, early-onset infections to inflammation and autoimmunity (56). Interestingly ApoE and RAG2-deficent mice have lower plasma TC levels than do immune-competent apoE mice suggesting an effect of the immune system on plasma lipid homeostasis (57). Finally, cg24240870 is located in *GOSR1* (*GS28*), encoding for the Golgi SNAP receptor complex member 1, which transports proteins among the endoplasmic reticulum and the Golgi, and between Golgi compartments. In murine melanocytes, 25-hydroxycholesterol reduced protein levels of GS28 and may be linked to cholesterol homeostasis through an OSBP-related protein (58).

Moreover, topological interaction network visualization represented another method to identify novel central genes in co-methylation modules. In the black module, five central genes were identified (*ANKRD5, HCFC1R1, MUC2, NAP1L4*, and *SGCD*). A genetic variant (rs7935422) in *NAP1L4*, encoding for a nucleosome assembly protein 1 like 4, has been associated with HDL-C concentrations in various populations (59). Moreover, genetic variants in *SGCD*, encoding for a sarcoglycan delta, have been associated with circulating lipid concentrations (54) and HDL particle size (60). In the turquoise module, two central genes were identified (*PACS2* and *LGALS12*). Interestingly, *PACS2* impacts lipid metabolism by controlling formation of endoplasmic reticulum lipid-synthesizing centers, which are found on mitochondria-associated membranes (61). Moreover, *LGALS12* is encoding for the galectin 12. Expression of *LGALS12* in adipocytes is up-regulated by PPAR-γ agonists suggesting its role in insulin signaling and type 2 diabetes (62). In the pink module, two central genes were identified (*CNNM2* and *FAM188A*). A genetic variant in *CNNM2*, encoding for a cyclin and CBS domain divalent metal cation transport mediator 2, has been associated with coronary artery disease and blood pressure (63-65). In summary, several hub genes and central genes were related to lipid metabolism, immune and inflammatory response, and the sensible connection between the two.

The present study has some limitations. The main limitation resides in the small sample size, which may reduce variability in DNA methylation levels, as well as limiting the statistical power to detect CpG sites associated with total carotenoid concentrations. We did not account for multiple testing since we used an exploratory approach and did not want this study to be overly constrained. However, WGCNA allowed reducing the impact of multiple testing by grouping the 20 687 CpG sites into nine modules. Even though DNA methylation in blood leukocytes is representative of DNA methylation in other tissues (66), it is influenced by cell composition. However, the impact of cell composition was not taking into account in the present analyses. In a previous paper by our group, cell heterogeneity was predicted using DNA methylation levels and it estimated proportions of six different cell types (24). These six additional cofounders were not included in the present analysis in order to prevent the reduction of the statistical power to detect significant associations and correlations. Finally, our study did not account for diet, physical activity, smoking, and alcohol consumption of the participants, all of which may affect circulating carotenoid concentrations (67, 68).

In conclusion, DNA methylation levels of 20 687 CpG sites were associated with plasma total carotenoid concentrations. Using WGCNA, nine co-methylation modules were identified. A total of 2734 hub genes, and more specifically 17 unique top hub genes, associated with total carotenoid concentrations were potentially related to lipid profile traits. Even though further studies and validation of the hub genes are needed, this provides evidence for the potential role of gene co-methylation in the relationship between carotenoids, lipid profile and ultimately CM health.

Author Contributions

Conceptualization, Benoît Lamarche and Marie-Claude Vohl; Data curation, Bénédicte L. Tremblay, Frédéric Guénard and Louis Pérusse; Formal analysis, Bénédicte L. Tremblay and Frédéric Guénard; Funding acquisition, Marie-Claude Vohl; Investigation, Benoît Lamarche and Marie-Claude Vohl; Methodology, Bénédicte L. Tremblay, Frédéric Guénard and Marie-Claude Vohl; Project administration, Marie-Claude Vohl; Resources, Louis Pérusse and Marie-Claude Vohl; Software, Bénédicte L. Tremblay, and Frédéric Guénard; Supervision, Marie-Claude Vohl; Visualization, Bénédicte L. Tremblay; Writing – original draft, Bénédicte L. Tremblay; Writing – review & editing, Bénédicte L. Tremblay, Frédéric Guénard, Benoît Lamarche, Louis Pérusse and Marie-Claude Vohl.

Funding

This work was supported by the Canada Research Chair in Genomics Applied to Nutrition and Metabolic Health.

Acknowledgments

We would like to thank Christian Couture, Véronique Garneau, and Catherine Raymond who contributed to the success of this study. This work was supported by the Canada Research Chair in Genomics Applied to Nutrition and Metabolic Health. MCV is Tier 1 Canada Research Chair in Genomics Applied to Nutrition and Metabolic Health. BLT is a recipient of a scholarship from the Canadian Institutes of Health Research (CIHR).

Conflicts of Interest

The authors declare no conflict of interest.

References

1. Guo F, Moellering DR, Garvey WT. The progression of cardiometabolic disease: validation of a new cardiometabolic disease staging system applicable to obesity. Obesity. 2014;22(1):110-8.

2. Martinez-Gonzalez MA, de la Fuente-Arrillaga C, Lopez-Del-Burgo C, Vazquez-Ruiz Z, Benito S, Ruiz-Canela M. Low consumption of fruit and vegetables and risk of chronic disease: a review of the epidemiological evidence and temporal trends among Spanish graduates. Public health nutrition. 2011;14(12A):2309-15.

3. Couillard C, Lemieux S, Vohl MC, Couture P, Lamarche B. Carotenoids as biomarkers of fruit and vegetable intake in men and women. The British journal of nutrition. 2016;116(7):1206-15.

4. Maiani G, Caston MJ, Catasta G, Toti E, Cambrodon IG, Bysted A, et al. Carotenoids: actual knowledge on food sources, intakes, stability and bioavailability and their protective role in humans. Molecular nutrition & food research. 2009;53 Suppl 2:S194-218.

5. Eroglu A, Harrison EH. Carotenoid metabolism in mammals, including man: formation, occurrence, and function of apocarotenoids. Journal of lipid research. 2013;54(7):1719-30.

6. Allore T, Lemieux S, Vohl MC, Couture P, Lamarche B, Couillard C. Correlates of the difference in plasma carotenoid concentrations between men and women. The British journal of nutrition. 2018:1-10.

7. Farook VS, Reddivari L, Mummidi S, Puppala S, Arya R, Lopez-Alvarenga JC, et al. Genetics of serum carotenoid concentrations and their correlation with obesity-related traits in Mexican American children. The American journal of clinical nutrition. 2017;106(1):52-8.

8. Tremblay BL, Guenard F, Lamarche B, Perusse L, Vohl MC. Genetic and Common Environmental Contributions to Familial Resemblances in Plasma Carotenoid Concentrations in Healthy Families. Nutrients. 2018;10(8).

9. Clevidence BA, Bieri JG. Association of carotenoids with human plasma lipoproteins. Methods in enzymology. 1993;214:33-46.

10. D'Adamo CR, D'Urso A, Ryan KA, Yerges-Armstrong LM, Semba RD, Steinle NI, et al. A Common Variant in the SETD7 Gene Predicts Serum Lycopene Concentrations. Nutrients. 2016;8(2):82.

11. Beydoun MA, Nalls MA, Canas JA, Evans MK, Zonderman AB. Gene polymorphisms and gene scores linked to low serum carotenoid status and their associations with metabolic disturbance and depressive symptoms in African-American adults. The British journal of nutrition. 2014;112(6):992-1003.

12. Ferrucci L, Perry JR, Matteini A, Perola M, Tanaka T, Silander K, et al. Common variation in the beta-carotene 15,15'-monooxygenase 1 gene affects circulating levels of carotenoids: a genome-wide association study. American journal of human genetics. 2009;84(2):123-33.

13. Zubair N, Kooperberg C, Liu J, Di C, Peters U, Neuhouser ML. Genetic variation predicts serum lycopene concentrations in a multiethnic population of postmenopausal women. The Journal of nutrition. 2015;145(2):187-92.

14. Borel P, de Edelenyi FS, Vincent-Baudry S, Malezet-Desmoulin C, Margotat A, Lyan B, et al. Genetic variants in BCMO1 and CD36 are associated with plasma lutein concentrations and macular pigment optical density in humans. Annals of medicine. 2011;43(1):47-59.

15. Bohn T, Desmarchelier C, Dragsted LO, Nielsen CS, Stahl W, Ruhl R, et al. Host-related factors explaining interindividual variability of carotenoid bioavailability and tissue concentrations in humans. Molecular nutrition & food research. 2017;61(6).

16. Sharoni Y, Danilenko M, Dubi N, Ben-Dor A, Levy J. Carotenoids and transcription. Archives of biochemistry and biophysics. 2004;430(1):89-96.

17. Liu AG, Erdman JW, Jr. Lycopene and apo-10'-lycopenal do not alter DNA methylation of GSTP1 in LNCaP cells. Biochemical and biophysical research communications. 2011;412(3):479-82.

18. King-Batoon A, Leszczynska JM, Klein CB. Modulation of gene methylation by genistein or lycopene in breast cancer cells. Environmental and molecular mutagenesis. 2008;49(1):36-45.

19. Bollati V, Favero C, Albetti B, Tarantini L, Moroni A, Byun HM, et al. Nutrients intake is associated with DNA methylation of candidate inflammatory genes in a population of obese subjects. Nutrients. 2014;6(10):4625-39.

20. Quach A, Levine ME, Tanaka T, Lu AT, Chen BH, Ferrucci L, et al. Epigenetic clock analysis of diet, exercise, education, and lifestyle factors. Aging. 2017;9(2):419-46.

21. Mittelstrass K, Waldenberger M. DNA methylation in human lipid metabolism and related diseases. Current opinion in lipidology. 2018;29(2):116-24.

22. Langfelder P, Horvath S. WGCNA: an R package for weighted correlation network analysis. BMC bioinformatics. 2008;9:559.

23. Tremblay BL, Guenard F, Lamarche B, Perusse L, Vohl MC. Familial resemblances in blood leukocyte DNA methylation levels. Epigenetics. 2016;11(11):831-8.

24. Tremblay BL, Guenard F, Lamarche B, Perusse L, Vohl MC. Familial resemblances in human whole blood transcriptome. BMC genomics. 2018;19(1):300.

25. Tremblay BL, Guenard F, Lamarche B, Perusse L, Vohl MC. Familial resemblances in human plasma metabolites are attributable to both genetic and common environmental effects. Nutrition research. 2019;61:22-30.

26. Callaway C, Chumlea W, Bouchard C, Himes J, Lohman T, Martin A, et al. Standardization of anthropomeric measurements : The Airlie (VA) Consensus Conference. Champaign, IR, USA; 1988.

27. McNamara JR, Schaefer EJ. Automated enzymatic standardized lipid analyses for plasma and lipoprotein fractions. ClinChimActa. 1987;166(1):1-8.

28. Burstein M, Samaille J. [On a rapid determination of the cholesterol bound to the serum alpha- and beta-lipoproteins]. ClinChimActa. 1960;5:609.

29. Albers JJ, Warnick GR, Wiebe D, King P, Steiner P, Smith L, et al. Multi-laboratory comparison of three heparin-Mn2+ precipitation procedures for estimating cholesterol in high-density lipoprotein. ClinChem. 1978;24(6):853-6.

30. Friedewald WT, Levy RI, Fredrickson DS. Estimation of the concentration of low-density lipoprotein cholesterol in plasma, without use of the preparative ultracentrifuge. ClinChem. 1972;18(6):499-502.

31. Laurell CB. Quantitative estimation of proteins by electrophoresis in agarose gel containing antibodies. AnalBiochem. 1966;15(1):45-52.

32. Pirro M, Bergeron J, Dagenais GR, Bernard PM, Cantin B, Despres JP, et al. Age and duration of follow-up as modulators of the risk for ischemic heart disease associated with high plasma C-reactive protein levels in men. ArchInternMed. 2001;161(20):2474-80.

33. Richterich R, Dauwalder H. [Determination of plasma glucose by hexokinase-glucose-6-phosphate dehydrogenase method]. Schweizerische medizinische Wochenschrift. 1971;101(17):615-8.

34. Desbuquois B, Aurbach GD. Use of polyethylene glycol to separate free and antibody-bound peptide hormones in radioimmunoassays. The Journal of clinical endocrinology and metabolism. 1971;33(5):732-8.

35. Price ME, Cotton AM, Lam LL, Farre P, Emberly E, Brown CJ, et al. Additional annotation enhances potential for biologically-relevant analysis of the Illumina Infinium HumanMethylation450 BeadChip array. Epigenetics & chromatin. 2013;6(1):4.

36. Team RC. R: A language and environment for statistical computing. R Foundation for Statistical Computing. Vienna, Austria. 2013.

37. Tremblay BL, Guenard F, Lamarche B, Perusse L, Vohl MC. Weighted gene co-expression network analysis to explain the relationship between plasma total carotenoids and lipid profile. Genes & nutrition. 2019;14:16.

38. Langfelder P, Horvath S. Fast R Functions for Robust Correlations and Hierarchical Clustering. Journal of statistical software. 2012;46(11).

39. Zhang B, Horvath S. A general framework for weighted gene co-expression network analysis. Statistical applications in genetics and molecular biology. 2005;4:Article17.

40. Langfelder P, Horvath S. Eigengene networks for studying the relationships between co-expression modules. BMC systems biology. 2007;1:54.

41. Horvath S, Zhang B, Carlson M, Lu KV, Zhu S, Felciano RM, et al. Analysis of oncogenic signaling networks in glioblastoma identifies ASPM as a molecular target. Proceedings of the National Academy of Sciences of the United States of America. 2006;103(46):17402-7.

42. VisANT 5.0. Visual analyses of metabolic networks in cells and ecosystems.

http://visant.bu.edu. Accessed 18 December 18. 2018 [Available from: http://visant.bu.edu.

43. Hu Z, Snitkin ES, DeLisi C. VisANT: an integrative framework for networks in systems biology. Briefings in bioinformatics. 2008;9(4):317-25.

44. Kuleshov MV, Jones MR, Rouillard AD, Fernandez NF, Duan Q, Wang Z, et al. Enrichr: a comprehensive gene set enrichment analysis web server 2016 update. Nucleic acids research. 2016;44(W1):W90-7.

45. Borel P, Desmarchelier C, Nowicki M, Bott R. A Combination of Single-Nucleotide Polymorphisms Is Associated with Interindividual Variability in Dietary beta-Carotene Bioavailability in Healthy Men. The Journal of nutrition. 2015;145(8):1740-7.

46. Borel P, Desmarchelier C, Nowicki M, Bott R. Lycopene bioavailability is associated with a combination of genetic variants. Free radical biology & medicine. 2015;83:238-44.

47. Borel P, Desmarchelier C, Nowicki M, Bott R, Morange S, Lesavre N. Interindividual variability of lutein bioavailability in healthy men: characterization, genetic variants involved, and relation with fasting plasma lutein concentration. The American journal of clinical nutrition. 2014;100(1):168-75.

48. Fu LJ, Ding YB, Wu LX, Wen CJ, Qu Q, Zhang X, et al. The Effects of Lycopene on the Methylation of the GSTP1 Promoter and Global Methylation in Prostatic Cancer Cell Lines PC3 and LNCaP. International journal of endocrinology. 2014;2014:620165.

49. Yang Y, Fuentes F, Shu L, Wang C, Pung D, Li W, et al. Epigenetic CpG Methylation of the Promoter and Reactivation of the Expression of GSTP1 by Astaxanthin in Human Prostate LNCaP Cells. The AAPS journal. 2017;19(2):421-30.

50. Willer CJ, Schmidt EM, Sengupta S, Peloso GM, Gustafsson S, Kanoni S, et al. Discovery and refinement of loci associated with lipid levels. Nature genetics. 2013;45(11):1274-83.

51. Ligthart S, Vaez A, Hsu YH, Inflammation Working Group of the CC, Pmi Wg XCP, LifeLines Cohort S, et al. Bivariate genome-wide association study identifies novel pleiotropic loci for lipids and inflammation. BMC genomics. 2016;17:443.

52. Huan T, Rong J, Liu C, Zhang X, Tanriverdi K, Joehanes R, et al. Genome-wide identification of microRNA expression quantitative trait loci. Nature communications. 2015;6:6601.

53. Willer CJ, Sanna S, Jackson AU, Scuteri A, Bonnycastle LL, Clarke R, et al. Newly identified loci that influence lipid concentrations and risk of coronary artery disease. Nature genetics. 2008;40(2):161-9.

54. Comuzzie AG, Cole SA, Laston SL, Voruganti VS, Haack K, Gibbs RA, et al. Novel genetic loci identified for the pathophysiology of childhood obesity in the Hispanic population. PloS one. 2012;7(12):e51954.

55. Altunrende ME, Gezen-Ak D, Atasoy IL, Candas E, Dursun E. The Role of Astaxanthin on Transcriptional Regulation of NMDA Receptors Voltage Sensitive Calcium Channels and Calcium Binding Proteins in Primary Cortical Neurons. Noro psikiyatri arsivi. 2018;55(4):295-300.

56. Notarangelo LD, Kim MS, Walter JE, Lee YN. Human RAG mutations: biochemistry and clinical implications. Nature reviews Immunology. 2016;16(4):234-46.

57. Reardon CA, Blachowicz L, White T, Cabana V, Wang Y, Lukens J, et al. Effect of immune deficiency on lipoproteins and atherosclerosis in male apolipoprotein E-deficient mice. Arteriosclerosis, thrombosis, and vascular biology. 2001;21(6):1011-6.

58. Hall AM, Krishnamoorthy L, Orlow SJ. 25-hydroxycholesterol acts in the Golgi compartment to induce degradation of tyrosinase. Pigment cell research. 2004;17(4):396-406.

59. Hoffmann TJ, Theusch E, Haldar T, Ranatunga DK, Jorgenson E, Medina MW, et al. A large electronic-health-record-based genome-wide study of serum lipids. Nature genetics. 2018;50(3):401-13.

60. Kaess BM, Tomaszewski M, Braund PS, Stark K, Rafelt S, Fischer M, et al. Large-scale candidate gene analysis of HDL particle features. PloS one. 2011;6(1):e14529.

61. Simmen T, Aslan JE, Blagoveshchenskaya AD, Thomas L, Wan L, Xiang Y, et al. PACS-2 controls endoplasmic reticulum-mitochondria communication and Bid-mediated apoptosis. The EMBO journal. 2005;24(4):717-29.

62. Yang RY, Hsu DK, Yu L, Chen HY, Liu FT. Galectin-12 is required for adipogenic signaling and adipocyte differentiation. The Journal of biological chemistry. 2004;279(28):29761-6.

63. Coronary Artery Disease Genetics C. A genome-wide association study in Europeans and South Asians identifies five new loci for coronary artery disease. Nature genetics. 2011;43(4):339-44.

64. Dichgans M, Malik R, Konig IR, Rosand J, Clarke R, Gretarsdottir S, et al. Shared genetic susceptibility to ischemic stroke and coronary artery disease: a genome-wide analysis of common variants. Stroke. 2014;45(1):24-36.

65. 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.

66. Guénard F DY, Hould FS, Lebel S, Tchernof A, Marceau P and Vohl MC. Use of Blood as a Surrogate Model for the Assessment of Visceral Adipose Tissue Methylation Profiles Associated with the Metabolic Syndrome in Men. J Mol Genet Med. 2016;10(1).

67. Gruber M, Chappell R, Millen A, LaRowe T, Moeller SM, Iannaccone A, et al. Correlates of serum lutein + zeaxanthin: findings from the Third National Health and Nutrition Examination Survey. The Journal of nutrition. 2004;134(9):2387-94.

68. Al-Delaimy WK, van Kappel AL, Ferrari P, Slimani N, Steghens JP, Bingham S, et al. Plasma levels of six carotenoids in nine European countries: report from the European Prospective Investigation into Cancer and Nutrition (EPIC). Public health nutrition. 2004;7(6):713-22.

Tables

Table 8.1 Characteristic and biochemical parameters

Biochemical parameters	Parents (n=22)	Children (n=26)
Age (years)	42.3 ± 5.3	11.3 ± 3.4
BMI (kg/m^2)	23.9 ± 3.0	-
BMI percentile	-	50 ± 31.1
TC (mmol/L)	4.68 ± 0.55	4.28 ± 0.51
HDL-C (mmol/L)	1.63 ± 0.38	1.55 ± 0.29
LDL-C (mmol/L)	2.61 ± 0.55	2.31 ± 0.45
TC/HDL-C	3.02 ± 0.78	2.83 ± 0.47
ApoB100 (g/L)	0.80 ± 0.15	0.70 ± 0.13
TG (mmol/L)	0.95 ± 0.35	0.93 ± 0.39
Glucose (mmol/L)	5.20 ± 0.34	4.86 ± 0.23
Insulin (pmol/L)	66.59 ± 33.54	73.00 ± 28.96
CRP (mg/L)	0.77 ± 0.84	0.43 ± 0.66
Total carotenoids (µmol/L)	6.35 ± 2.39	5.70 ± 2.05

All values are means ± SD. Abbreviations: Apolipoprotein B100 (ApoB100), Body mass index (BMI), C-reactive protein (CRP), High-density lipoprotein cholesterol (HDL-C), Low-density lipoprotein cholesterol (LDL-C), Standard deviation (SD), Total cholesterol (TC), Total cholesterol/HDL-C (TC/HDL-C), Triglycerides (TG).

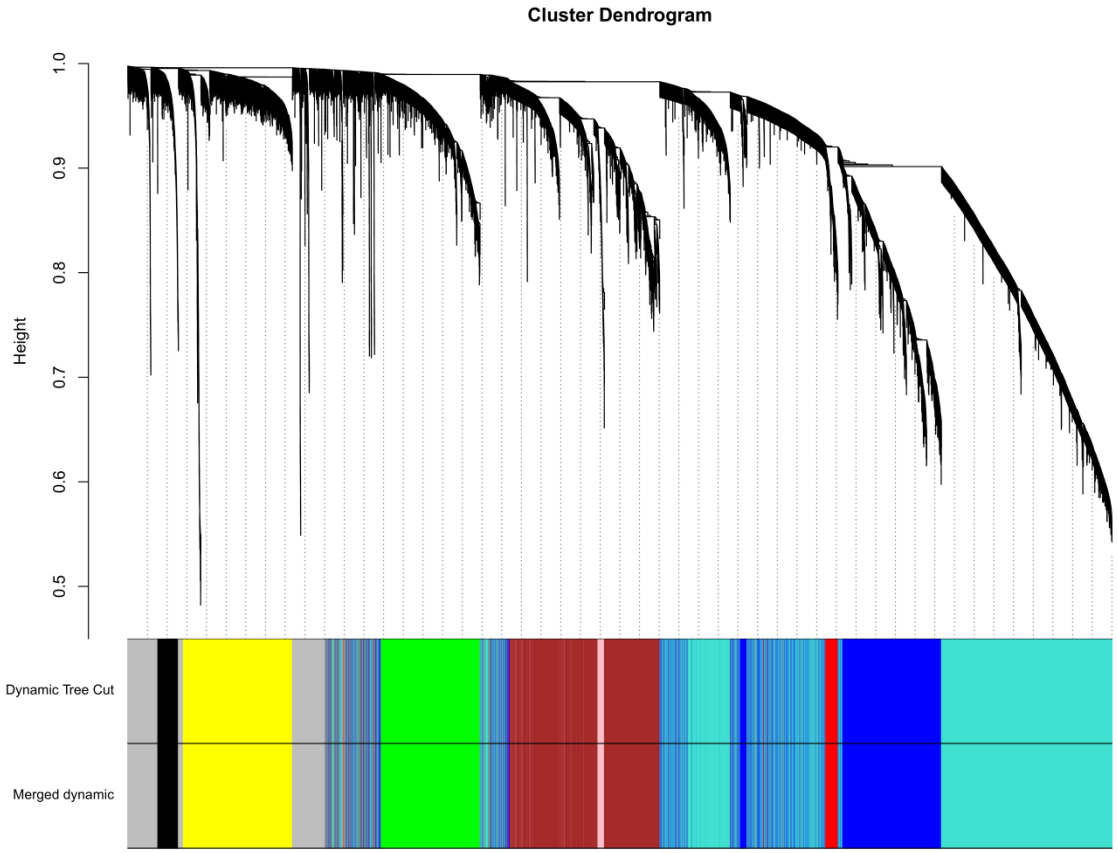
Table 8.2 Top hub genes identified from WGCNA analysis

Module-lipid	CpG site	Gene (Chr, Accession number)	GS	*P*-value GS	MM	*P*-value MM
Black HDL-C	cg01999908	*PAX4* (Chr7, NM_006193)	-0.37	0.0092	0.85	2.10×10^{-14}
	cg08731068	*TBC1D16* (Chr17, NM_019020)	-0.37	0.011	0.89	1.70×10^{-17}
	cg04010296	*PGA5* (Chr11, NM_014224)	-0.36	0.011	0.83	4.08×10^{-13}
Turquoise TC	cg00000029	*RBL2* (Chr16, NM_005611)	-0.59	8.49×10^{-6}	0.83	3.45×10^{-13}
	cg02500883	*GRIN3A* (Chr9, NM_133445)	-0.57	1.98×10^{-5}	0.81	2.52×10^{-12}
	cg00864012	*TEX2* (Chr17, NM_018469)	-0.56	3.39×10^{-5}	0.85	2.39×10^{-14}
Turquoise HDL-C	cg27616007	*LST1* (Chr6, NR_029461)	-0.42	0.0027	0.81	3.82×10^{-12}
	cg24240870	*GOSR1* (Chr17, NM_004871)	-0.42	0.0030	0.81	4.55×10^{-12}
	cg18089000	*GBGT1* (Chr9, NM_021996)	-0.42	0.0031	0.86	3.88×10^{-15}
Turquoise LDL-C	cg07665923	*C7orf50* (Chr7, NM_001134396)	-0.45	0.0012	0.83	2.22×10^{-13}
	cg02119755	*PDPK1* (Chr16, NM_031268)	-0.44	0.0017	0.87	1.62×10^{-15}
	cg06142108	*SYT17* (Chr16, NM_016524)	-0.43	0.0022	0.88	1.75×10^{-16}
Turquoise ApoB100	cg17053201	*GABBR1* (Chr6, NM_021904)	-0.46	0.00096	0.82	8.40×10^{-13}
	cg21671607	*RAG2* (Chr11, NM_000536)	-0.46	0.0011	0.84	7.35×10^{-14}
	cg07665923	*C7orf50* (Chr7, NM_001134396)	-0.46	0.0011	0.83	2.22×10^{-13}
Pink HDL-C	cg15505294	*RNASE11* (Chr14, NM_145250)	0.36	0.012	-0.95	1.87×10^{-24}
	cg01719157	*TRIM68* (Chr11, NM_018073)	-0.35	0.013	0.81	4.10×10^{-12}
	cg09113768	*DEPDC1* (Chr1, NM_017779)	-0.35	0.015	0.83	2.22×10^{-13}

Abbreviations: Apolipoprotein B100 (ApoB100), chromosome (Chr), gene significance (GS), high-density lipoprotein cholesterol (HDL-C), low-density lipoprotein cholesterol (LDL-C), module membership (MM), total cholesterol (TC)

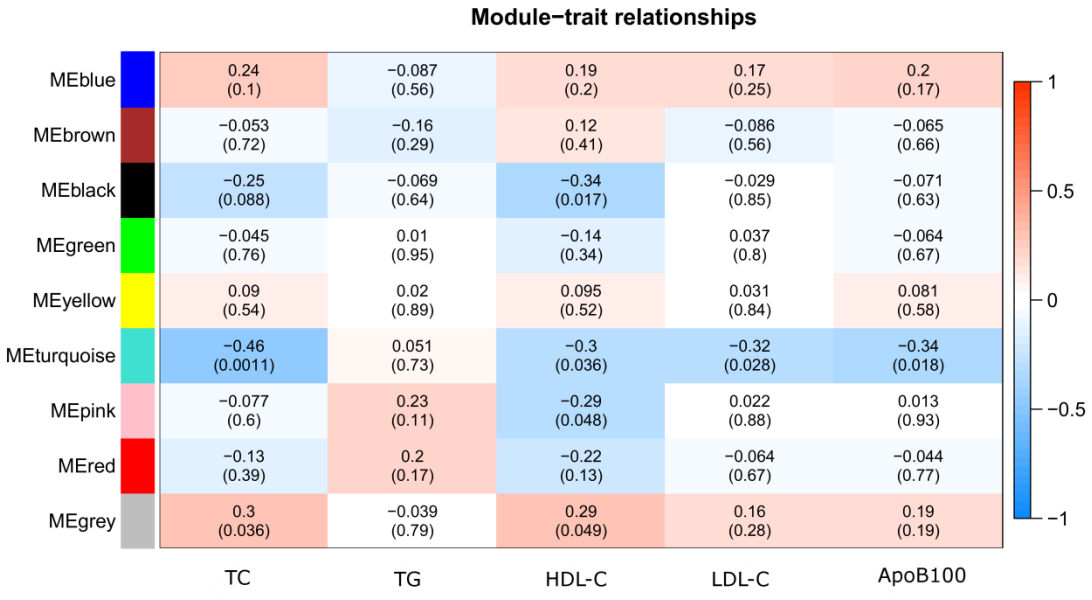
Figures

Figure 8.1 Gene dendrogram obtained using average linkage hierarchical clustering



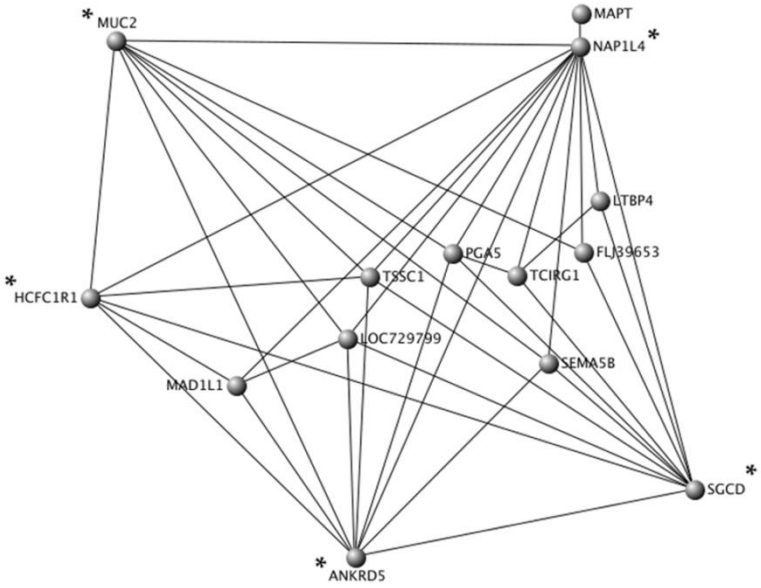
The dynamic tree cut yielded nine modules. The merged dynamic yielded same modules as the dynamic tree cut using a cut-off of 0.25. Module colors are shown correspondingly.

Figure 8.2 Heatmap of module-trait relationships depicting correlations between module eigengenes and lipid profile traits



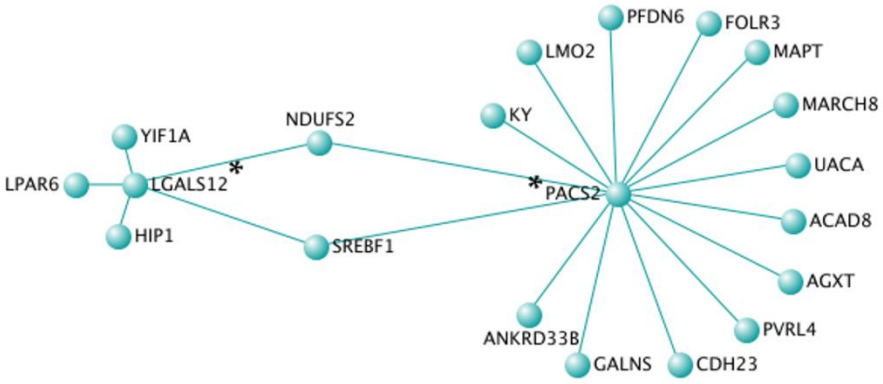
Numbers in the table correspond to the correlation r and the *P*-value in parentheses. The degree of correlation is illustrated with the color legend.

Figure 8.3 Topological interaction network of 14 unique genes in the black module



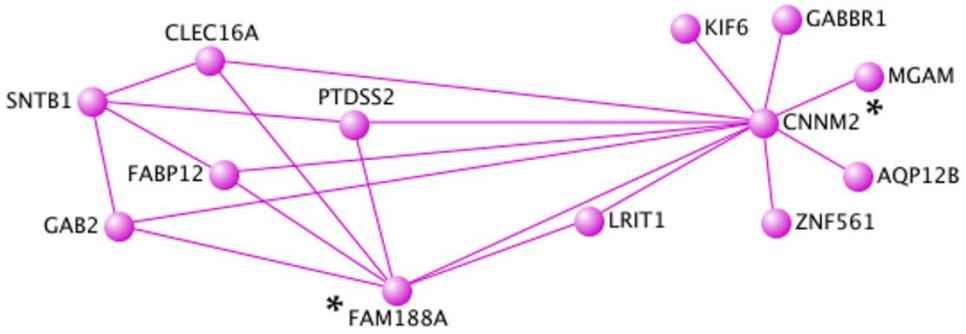
Each gene is represented by a node and edge number is proportional to connection strength. Gene-gene interaction network was constructed and visualized using VisANT 5.0. Central genes are identified using asterisks.

Figure 8.4 Topological interaction network of 20 unique genes in the turquoise module



Each gene is represented by a node and edge number is proportional to connection strength. Gene-gene interaction network was constructed and visualized using VisANT 5.0. Central genes are identified using asterisks.

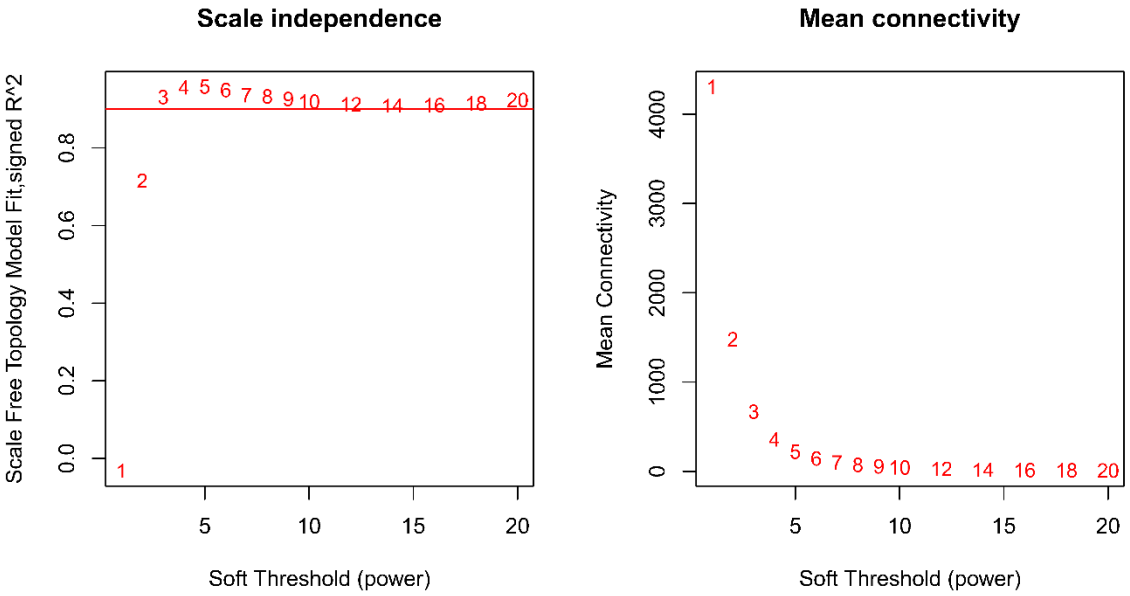
Figure 8.5 Topological interaction network of 13 unique genes in the pink module



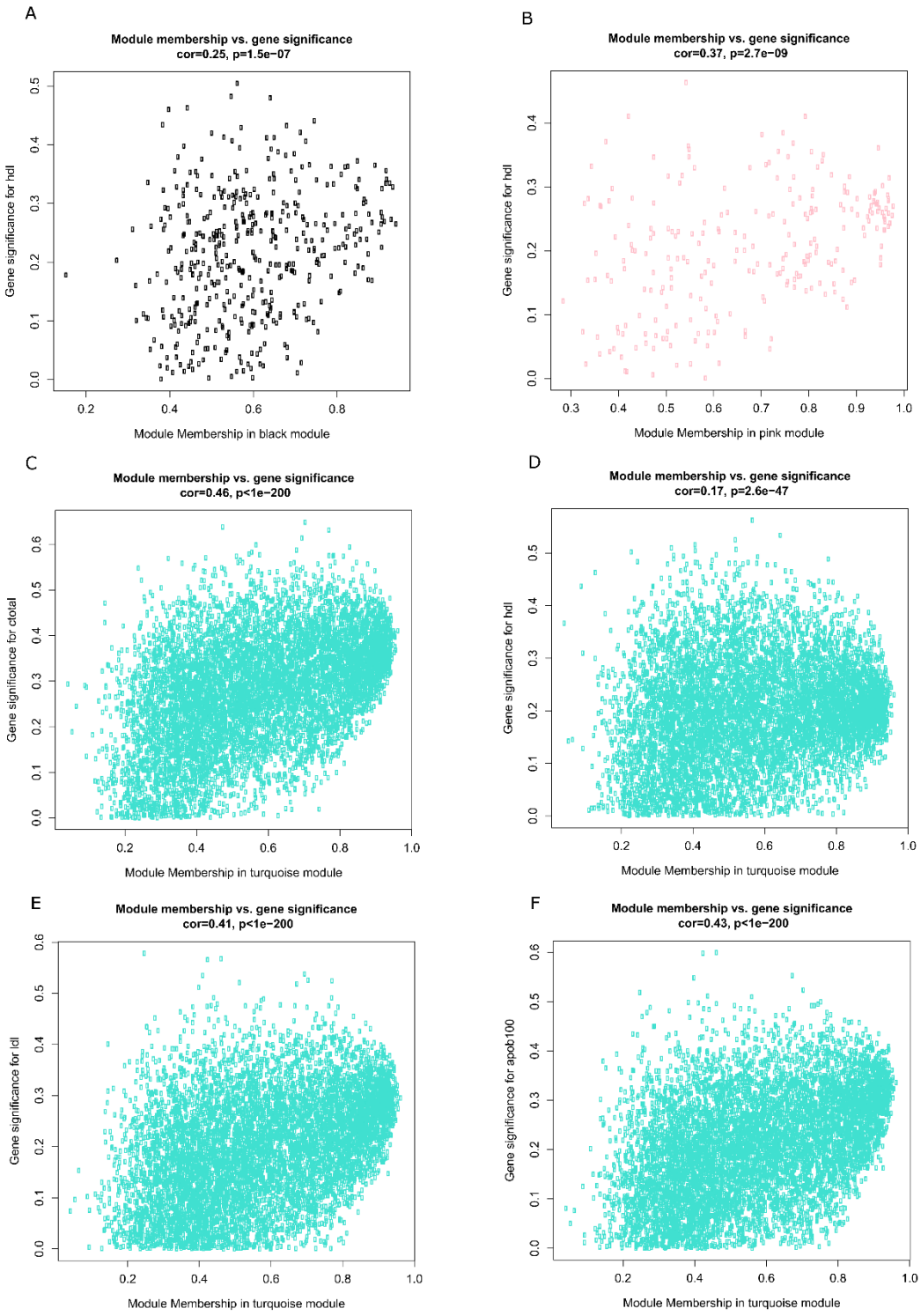
Each gene is represented by a node and edge number is proportional to connection strength. Gene-gene interaction network was constructed and visualized using VisANT 5.0. Central genes are identified using asterisks.

Supplemental material

Supplemental figure 8.1 Scale independence and mean connectivity



Supplemental figure 8.2 Scatterplots of gene significance for lipid profile traits and module membership in the A) black-HDL-C, B) pink-HDL-C, C) turquoise-TC, D) turquoise-HDL-C, E) turquoise-LDL-C, and F) turquoise-ApoB100



Supplemental table 8.1 Concentrations of plasma carotenoids (μmol/L of plasma)

Subjects	Lutein	Zeaxanthin	β-Cryptoxanthin	α-Carotene	β-Carotene	Lycopene	Total carotenoids
Gen01	0.36689606	0.0678172	0.48942068	0.94815424	1.96626358	1.90477202	5.74332378
Gen02	0.19040767	0.05211724	0.2671793	0.39046788	1.63413354	2.97212935	5.50643497
Gen03	0.20500873	0.04703513	0.33127708	1.3088189	1.97765103	2.14359446	6.01338535
Gen05	0.24382521	0.05322967	0.42173556	0.80394106	2.58467386	2.80014081	6.90754618
Gen06	0.21138746	0.03997405	0.47272927	0.87987086	2.33627976	2.28340529	6.2236467
Gen07	0.15361658	0.03836336	0.20938586	1.01423133	1.62869228	1.21588093	4.26017034
Gen08	0.13950888	0.03245329	0.2880089	0.56440081	1.07092427	1.52477859	3.62007474
Gen09	0.18935861	0.06023487	0.24183803	0.36889688	1.46894869	1.84769435	4.17697143
Gen10	0.45531524	0.05366406	0.68650319	1.86816097	5.09892831	1.60505358	9.76762535
Gen11	0.13301695	0.0379158	0.40775388	1.17280733	1.78868368	1.15260911	4.69278675
Gen12	0.1495655	0.03782232	0.52583487	1.20447	2.13195549	1.51354258	5.56319076
Gen13	0.22080196	0.04605841	0.40370279	1.95125405	4.36470233	1.77749248	8.76401203
Gen14	0.19439964	0.0346023	0.47240977	1.69428919	2.77308628	1.17814005	6.34692723
Gen15	0.31456871	0.07906164	0.64207259	0.585099	1.26299619	2.07400924	4.95780737
Gen16	0.20941464	0.0590733	0.66509735	0.43477479	1.2177428	1.90799437	4.49409726
Gen17	0.31659735	0.05824728	0.43726324	3.07504388	4.90271428	1.52411709	10.3139831
Gen18	0.40720388	0.06879612	0.52897983	1.77242233	2.73849116	1.40131048	6.9172038
Gen19	0.21406127	0.04334581	0.21302691	1.46073827	2.3907969	0.81416873	5.13613789
Gen20	0.38953299	0.11518446	0.54511458	0.58599723	1.94027507	1.46287634	5.03898068
Gen21	0.25823509	0.05614234	0.5515715	3.29451982	4.1048887	1.59654332	9.86190078
Gen22	0.19637546	0.05178252	0.52029086	1.43217409	2.81824867	1.43588238	6.45475399
Gen23	0.19796017	0.03400027	0.36306339	0.6257273	1.36047476	1.52337096	4.10459685
Gen24	0.23360289	0.05076413	0.42062581	1.11325177	1.89522389	1.31461194	5.02808043
Gen25	0.19012552	0.06455947	0.73073545	0.68144593	1.40988861	0.8309107	3.90766568
Gen26	0.20673794	0.04472159	0.11642864	0.1033121	0.30102258	0.5048136	1.27703645
Gen27	0.23324584	0.07133053	0.53723883	0.38361577	0.77381218	0.77033291	2.76957605
Gen28	0.22858608	0.06703344	0.40164353	0.34477048	0.57881226	0.69589933	2.31674512

Gen29	0.22569969	0.06288219	0.19197018	0.79944354	1.42605149	1.67944118	4.38548827
Gen30	0.18846852	0.05111517	0.18449405	1.76125731	3.75436001	1.36074407	7.30043914
Gen31	0.1034949	0.03639765	0.23772312	0.55324807	1.51238711	1.61846353	4.06171437
Gen32	0.1379356	0.05037577	0.58562149	1.28098545	2.66908747	1.96597652	6.6899823
Gen33	0.34863636	0.06872228	0.59045342	1.05182969	2.8442326	1.16548687	6.06936121
Gen34	0.34080966	0.08200735	0.53132439	0.87811067	3.02914672	1.32128546	6.18268426
Gen35	0.18558526	0.05637978	0.69489283	1.20488219	2.06864071	1.10169917	5.31207994
Gen36	0.21559215	0.1016761	0.82701707	2.65607443	4.85041133	1.40830847	10.0590796
Gen38	0.26706147	0.09209016	.	3.55025825	6.33803969	1.79779113	12.0452407
Gen39	0.26395015	0.06463881	0.53649644	1.52282902	3.89619121	1.72663359	8.01073921
Gen40	0.22008198	0.04334325	0.45042163	1.29773283	3.99433146	2.71061293	8.71652407
Gen41	0.33789169	0.08950376	0.40316064	2.05426427	3.12184275	1.53282075	7.53948386
Gen42	0.16596505	0.05910571	0.43107599	0.74232185	1.36939518	1.38768457	4.15554834
Gen43	0.18579005	0.05817576	0.6537702	0.9184596	2.08512821	1.28347111	5.18479494
Gen44	0.26494297	0.09297957	0.90039749	1.52573514	3.79402374	2.5744044	9.15248331
Gen45	0.49521561	0.05468624	0.35857692	1.28192912	3.42957551	1.58930011	7.2092835
Gen46	0.72653971	0.09196268	0.47041018	1.0846667	1.94897283	1.42895167	5.75150378
Gen47	0.5848228	0.06066084	0.40081868	1.18038703	2.54237772	1.59877117	6.36783824
Gen48	0.55555079	0.07599738	0.5545489	0.91123368	1.78155749	1.2136722	5.09256043
Gen49	0.21725674	0.05761664	0.4287788	0.67397551	2.02502444	1.58549025	4.98814238
Gen50	0.22004146	0.03996355	0.286557	0.66712712	1.28660567	1.18069892	3.68099372

Supplemental table 8.2 CpG sites showing significant correlation with plasma total carotenoids (n=20 687)

Supplemental table 8.3 Hub genes identified from WGCNA analysis (n=2734)

Supplemental table 8.4 Molecular functions and pathways in the black, turquoise, and pink modules

Supplemental tables 8.2 to 8.4 are available online at https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6628241/.

Chapitre 9 Analyse intégrée de réseaux de co-omiques dans la relation entre les caroténoïdes plasmatiques et le profil lipidique

Bénédicte L. Tremblay, Frédéric Guénard, Benoît Lamarche, Louis Pérusse, Marie-Claude Vohl

L'article présenté dans ce chapitre s'intitule : *Integrative network analysis of multi-omics data in the link between plasma carotenoid concentrations and lipid profile.*

Cet article est publié dans la revue : *Lifestyle Genomics. 2020;13(1):11-19.*

Résumé

Introduction: Les caroténoïdes, qui sont des biomarqueurs fiables de la consommation de fruits et légumes, sont positivement associés au profil lipidique. Les concentrations de caroténoïdes en circulation peuvent interagir avec plusieurs profils omiques, dont le génome, le transcriptome et l'épigénome. Peu d'études ont utilisé des approches multi-omiques et elles incluent rarement des facteurs environnementaux, tels que l'alimentation. L'objectif de cette étude observationnelle était d'examiner le rôle potentiel des déterminants provenant de différentes approches omiques dans l'interconnexion entre l'alimentation, représentée par les caroténoïdes totaux, et le profil lipidique en utilisant l'analyse de réseaux de corrélations pondérées (WGCNA).

Méthodes: Les niveaux de méthylation de l'ADN des leucocytes sanguins de 472 245 sites CpG et les niveaux d'expression génique du sang de 18 160 transcrits ont été associés aux concentrations de caroténoïdes totaux en utilisant des régressions chez 48 sujets en santé. WGCNA a été utilisée pour identifier les modules de données omiques et les *hub genes* associés au profil lipidique.

Résultats: Parmi les gènes associés aux concentrations de caroténoïdes totaux, un total de 236 gènes avaient des niveaux de méthylation et d'expression associés aux caroténoïdes totaux. En utilisant WGCNA, six modules, constitués de groupes de gènes hautement corrélés et représentés par des couleurs, ont été identifiés et associés au profil lipidique. Les sondes des modules turquoise et vert étaient corrélées aux concentrations plasmatiques de lipides. Un total de 28 *hub genes* ont été identifiés.

Conclusions: La méthylation de l'ADN et l'expression génique à l'échelle du génome étaient tous deux associés aux concentrations plasmatiques de caroténoïdes totaux. Cette analyse intégrative a identifié plusieurs *hub genes* impliqués principalement dans le métabolisme des lipides et la réponse inflammatoire et qui présentent plusieurs variations génétiques associées aux concentrations plasmatiques de lipides. Cela permet une compréhension plus globale du système moléculaire interactif entre les caroténoïdes, les données omiques et le profil lipidique plasmatique.

Abstract

Introduction: Carotenoids which are a reliable biomarker of fruit and vegetable consumption are positively associated with lipid profile. Circulating carotenoid concentrations may interact with several omics profiles including genome, transcriptome, and epigenome. Few studies have used multi-omics approaches, and they rarely include environmental factors, such as diet. The objective of this observational study was to examine the potential role of multi-omics data in the interconnection between diet, represented by total carotenoids, and lipid profile using weighted gene correlation network analysis (WGCNA).

Methods: Blood leukocytes DNA methylation levels of 472 245 CpG sites and whole blood gene expression levels of 18 160 transcripts were tested for associations with total carotenoid concentrations using regressions in 48 healthy subjects. WGCNA was used to identify co-omics modules and hub genes related to lipid profile.

Results: Among genes associated with total carotenoid concentrations, a total of 236 genes were identified at both DNA methylation and gene expression levels. Using WGCNA, six modules, consisting in groups of highly correlated genes represented by colors, were identified and linked to lipid profile. Probes clustered in the turquoise and green modules correlated with plasma lipid concentrations. A total of 28 hub genes were identified.

Conclusions: Genome-wide DNA methylation and gene expression levels were both associated with plasma total carotenoid concentrations. Several hub genes, mostly involved in lipid metabolism and inflammatory response with several genetic variants associated with plasma lipid concentrations, came out of the integrative analysis. This provides a comprehensive understanding of the interactive molecular system between carotenoids, omics, and plasma lipid profile.

Title page

Integrative network analysis of multi-omics data in the link between plasma carotenoid concentrations and lipid profile

Bénédicte L. Tremblay[1, 2], Frédéric Guénard[1, 2], Benoît Lamarche[1, 2], Louis Pérusse[1, 3], Marie-Claude Vohl[1,2*]

[1.] Institute of Nutrition and Functional Foods (INAF), Laval University, Quebec City (QC), Canada

[2.] School of Nutrition, Laval University, Quebec City (QC), Canada

[3.] Department of Kinesiology, Laval University, Quebec City (QC), Canada

Corresponding author*:
Marie-Claude Vohl, Ph.D.
Institute of Nutrition and Functional Foods (INAF), Laval University
Pavillon des Services, 2440 Hochelaga Blvd, Quebec City, QC G1V 0A6, Canada
Phone: 418 656-2131 ext. 4676; Fax: 418 656-5877
E-mail: marie-claude.vohl@fsaa.ulaval.ca

Keywords: Carotenoids; DNA methylation; Gene expression; Lipids; WGCNA.

Introduction

Diet is the cornerstone for the prevention and treatment of chronic diseases. A diet rich in fruits and vegetables is inversely associated with the incidence of cardiovascular diseases (1). Plasma carotenoid concentrations are a reliable biomarker of fruit and vegetable consumption (2, 3). Variability in circulating carotenoids may be due to several factors including age, sex, body weight, physical activity, genetic, and lipid profile (3). Accordingly, the majority of circulating carotenoids are associated with several lipoproteins and plasma lipids (4, 5).

Several genome-wide association studies have identified genetic variants that influence circulating carotenoid concentrations (6, 7). Moreover, carotenoids may exert their effect on gene expression, via several transcriptional systems (8), and on DNA methylation of specific genes (9). Thus, omics profiles, including genome, transcriptome, and epigenome may interact with circulating carotenoids. Relating omics data to a specific trait requires the integration of a very large amount of data. Weighted gene correlation network analysis (WGCNA) is a widely used systems biology approach designed for high dimensional data such as omics data (10). It then becomes possible to relate highly correlated genes (modules) to a phenotypic trait, and to identify key hub genes within modules that are related to a phenotypic trait (10).

Studies usually include only one type of omics data. However, omics data are not only interconnected with each other, but also with the environment and phenotypic traits (11). An integrative multi-omics approach may provide a holistic understanding of an interactive molecular system (12). Only a few studies have used multi-omics approaches and most focused on the effect of genetics on omics markers, including DNA methylation (meQTLS) and gene expression (eQTLs) (12). Moreover, studies rarely include environmental factors, such as diet. A study in healthy overweight men reported the effects of specific dietary components on low-grade inflammation using a multi-omics approach (13).

In the present study, the objective was to examine the potential role of multi-omics data in the interconnection between diet, represented by total carotenoids, and lipid profile using WGCNA. First, multi-omics data (DNA methylation and gene expression levels) were tested for associations with plasma total carotenoid concentrations using linear regressions. Second, WGCNA was used to link carotenoids-associated multi-omics data to lipid profile.

Methods

Patients and design

A total of 48 Caucasian French-Canadian subjects from 16 families were recruited in the Greater Quebec City metropolitan area, in Canada, as part of the GENERATION Study, whose recruitment began in May 2011. The GENERATION Study was designed to evaluate familial resemblances in omics (DNA methylation (14) and gene expression (15)) and metabolic (metabolites (16) and carotenoids (17)) profiles in healthy families and to test impact of these profiles on cardiometabolic (CM) health. Families living under the same roof comprised at least the mother and one child aged between 8 and 18. Parents had to be the biological parents of their child (or children), in good general health, non-smokers, with body mass index (BMI) ranging between 18 and 35 kg/m^2, and free of any metabolic conditions requiring treatment, although the use of Synthroid® (levothyroxine) or oral contraceptive was tolerated. Children also had to be non-smokers, in good general health and not using psycho-stimulators [Ritalin® (methylphenidate), Concerta® (methylphenidate), and Strattera® (atomoxetine)]. Families were composed of 16 mothers, 6 fathers, and 26 children. Blood samples were taken from both parents and children during their visit at the Institute of Nutrition and Functional Foods (INAF). The experimental protocol was approved by the Ethics Committees of Laval University Hospital Research Center and Laval University. All participants (adults and children) signed an informed consent form. Parental consent was also obtained by signing the child consent document.

Anthropometric and lipid parameter measurements

Body weight and height were measured according to the procedures recommended by the Airlie Conference (18). Blood samples were collected from an antecubital vein into vacutainer tubes containing EDTA after 12-hour overnight fast and 48-hour alcohol abstinence. Plasma was separated by centrifugation (2 500 g for 10 min at 4°C), and samples were aliquoted and frozen (-80°C) for subsequent analyses. Enzymatic assays were used to measure plasma total cholesterol (TC) and triglyceride (TG) concentrations (19, 20). Precipitation of very low-density lipoprotein (VLDL) and low-density lipoprotein (LDL) particles in the infranatant with heparin manganese chloride generated the high-density lipoprotein cholesterol (HDL-C) fraction (21). LDL-C was calculated with the Friedewald formula (22). Apolipoprotein B-100 (ApoB100) concentrations were measured in plasma by the rocket immunoelectrophoretic method (23).

RNA extraction and gene expression analysis

Total RNA was isolated and purified from whole blood using PAXgene Blood RNA Kit (QIAGEN) as previously described (15). The HumanHT-12 v4 Expression BeadChip (Illumina Inc., San Diego, CA) was used to measure expression levels of ~ 47 000 probes (> 31 000 annotated genes). This was performed at the McGill University and Genome Quebec Innovation Centre (Montreal, Canada). The FlexArray software (version 1.6) (24) and the lumi R package were used to analyze and normalize gene expression levels. Probes with a detection *P*-value ≤ 0.05 in at least 25% of all subjects were considered in analysis. A total of

18 160 probes among the 47 323 probes on the microarray (38.4%) showed significant gene expression in blood.

DNA extraction and methylation analysis

Genomic DNA was extracted from blood leukocytes using the GenElute Blood Genomic DNA kit (Sigma-Aldrich, St. Louis, MO, USA) as previously described (14). Methylation levels were measured using Infinium Human Methylation 450 array (Illumina, San Diego, CA, USA). Bisulfite conversion and quantitative DNA methylation analysis were processed at the McGill University and Genome Quebec Innovation Center (Montreal, Canada). Illumina GenomeStudio software v2011.1 and the Methylation Module were used to analyze methylation data on 485 577 CpG sites. Global normalization using control probes was performed in GenomeStudio. Probes with a detection *P*-value > 0.01 in more than 5 subjects (> 10% of all subjects) were removed, as well as probes on the X and Y chromosomes (to eliminate gender bias), and probes mapped to multiple chromosomes (25). Thus, 472 245 probes were considered in the analysis.

Carotenoid measurements

Samples and standards used for the measurement of carotenoid concentrations were prepared as reported previously (17). Briefly, carotenoid standards were purchased from Sigma (Oakville, Ontario, Canada). A total of 100 µL of plasma, 20 µL of 2-propanol, and 20 µL of carotenoid standard were transferred in Eppendorf tubes. Samples were transferred on a 400µL fixed well plate (ISOLUTE® SLE+, Biotage, Charlotte, NC) with 900 µL of hexane:isopropanol (90/10, v/v) in each well. Each extracted sample was evaporated under nitrogen and reconstituted with 300 µL of methanol:dichloromethane (65/35, v/v). Plates were shaken for 10 minutes and samples were transferred into high performance liquid chromatography glass vials to be analyzed.

High performance liquid chromatography (HPLC)-UV analysis was performed using an Agilent 1260 liquid handling system (Agilent, Mississauga, Ontario, Canada) as previously described (17). Carotenoids were separated with a mobile phase consisting of methanol:water (98/2, v/v; Eluent A) and methyl-tert-butyl ether (MTBE; Eluent B; VWR, Mississauga, Ontario, Canada) with a flow-rate of 1 mL/min. UV detector was set at 450 nm and identification of each compound was confirmed using retention time and UV spectra (190-640nm) of the pure compounds. Data acquisition was carried out with the Chemstation software (Agilent, Mississauga, Ontario, Canada). For all carotenoids, the concentrations are reported in µmol/L of plasma. One outlier in β-crypoxanthin, defined as value falling outside of the mean ± 4 standard deviations, was excluded from analyses.

Associations between multi-omics data and plasma total carotenoids

Total plasma carotenoid (µmol/L of plasma) concentrations were calculated as the sum of α-carotene, β-carotene, β-cryptoxanthin, lutein, lycopene, and zeaxanthin concentrations. Concentrations of plasma carotenoids are available in a previous study (17). As previously described, R software v2.14.1 (R

Foundation for Statistical Computing; http://www.r-project.org) (26) was used to compute linear regressions between gene expression levels of all 18 160 transcripts, DNA methylation levels of all 472 245 CpG sites and total carotenoids adjusted for the family ID (27, 28).

Weighted gene correlation network analysis (WGCNA)

WGCNA was performed with the WGCNA package (10, 29) in R software (26). A weighted adjacency matrix was established by calculating Pearson correlations between each probe pair. The co-omics similarity was raised to a power $\beta = 3$ to calculate adjacency (30). Correlations between probes were used to construct a topological overlap matrix (TOM), which provided a similarity measure. The TOM was then used to calculate the corresponding dissimilarity (1-TOM). Average linkage hierarchical clustering coupled with the TOM-based dissimilarity was used to group probes with coherent omics profiles into modules (30). More specifically, the dynamic tree cutting algorithm (deep split = 3, minimum number of probes per module = 30, cut height = 0.25) was used to detect modules. The assignment of outlying probes to modules was made using the Partitioning Around Medoids (PAM) method. Colors are randomly assigned to modules except for grey color, which is reserved for the module with unassigned probes. To identify modules associated with lipid profile traits (TC, LDL-C, HDL-C, TG, ApoB100), correlations between module eigengenes (MEs) (i.e., the first principal component of the module) (31) and traits were computed. Gene significance (GS), defined as the absolute correlation between the probe and the trait, was used to quantify associations of individual probes with lipid profile traits. To quantify the similarity of all probes to every module, a quantitative measure of module membership (MM) was defined as the correlation of the ME and the probe DNA methylation or gene expression levels. Probes with the highest MM and highest GS were those with high significance (hub genes) (32). The hub genes within a module were chosen based on $GS > 0.2$ and $MM > 0.8$, with a *P*-value ≤ 0.05. Topological interaction network in the turquoise module was constructed and visualized using VisANT 5.0, a software framework for hierarchical organization of biological networks (33). Moreover, to obtain visually interpretable network, a weighted correlation cut-off of 0.42 was used in the turquoise module. Genes with four connections or more were selected as central genes.

Statistical analysis

Statistical Analysis Software (SAS) version 9.4 was used to compute means and SD of biochemical parameters in parents and children. CV was computed as the ratio of the SD to the mean for each 240 expression transcripts and 466 CpG sites. The non-parametric 1way Wilcoxon procedure was used to test difference in CV (non-normally distributed), while an unpaired t-test was used to test difference in MM of the turquoise module (normally distributed).

Results

Characteristics of study participants

Characteristics including plasma total carotenoid and lipid concentrations of healthy participants (parents and children) are presented in Table 1. Concentrations of all six main plasma carotenoids (α-carotene, β-carotene, β-cryptoxanthin, lutein, lycopene, and zeaxanthin) and total carotenoids are presented in a previous study (17).

Associations between multi-omics data and plasma total carotenoids

As previously reported in studies by our group, genome-wide DNA methylation levels of 20 687 out of 472 245 CpG sites (28) and genome-wide expression levels of 533 out of 18 160 transcripts (27) were associated ($P \leq 0.05$) with plasma total carotenoid concentrations. To perform an integrative analysis, subsets of CpG sites and transcripts associated with total carotenoids were collapsed to gene level to retain only common genes. A total of 236 genes (represented by 466 CpG sites and 240 transcripts) were identified at both DNA methylation and gene expression levels (Supplemental Table S1). WGCNA was then conducted on these subsets of probes as illustrated in the study schematic (Supplemental Figure S1).

Weighted gene correlation network analysis (WGCNA)

A total of six distinct modules were identified from DNA methylation levels of the 466 CpG sites and gene expression levels of the 240 transcripts using a dynamic tree cutting algorithm (Figure 1). The blue, brown, green, turquoise, yellow, and grey modules contained 133, 96, 34, 381, 38, and 24 probes, respectively. Grey module contained the 24 uncorrelated probes, which were excluded from further analysis. None of the modules were merged using the merged dynamic algorithm (cut-off height of 0.25) (Figure 1). Correlations between module eigengenes (MEs) and lipid profile traits (TC, LDL, HDL, TG, ApoB100) were computed to find lipid profile-correlated modules. ME representing the 381 probes clustered in the turquoise module correlated inversely with TC (r=-0.46 p=0.001), HDL-C (r=-0.31 p=0.03), LDL-C (r=-0.32 p=0.03), and ApoB100 (r=-0.34 p=0.02) (Figure 2). ME representing the 34 probes clustered in the green module correlated positively with HDL-C (r=0.34 p=0.02) (Figure 2). Gene significance (GS) of each plasma lipids was correlated with module membership (MM) of each module. MM in the turquoise module correlated with GS for TC (r=0.42 p=1 x10^{-17}), HDL-C (r=0.11 p=0.032), LDL-C (r=0.39 p=2.7 x10^{-15}), and ApoB100 (r=0.43 p=1.4 x 10^{-18}). MM in the green module correlated with GS for HDL-C (r=0.35 p=0.042). This suggests that probes highly significantly associated with these lipid profile traits were also the most important elements of the turquoise and green modules. Thus, the turquoise and green modules were selected as modules of interest in subsequent analyses.

Hub gene analysis

Hub gene analysis, identifying genes with the highest MM and highest GS, was conducted for the turquoise and green modules in order to refine the analysis of potential mechanisms linking carotenoids to lipid profile. A total of 35 probes (28 unique genes) were identified in the turquoise module: *AFF1, AMICA1, ARHGEF10,*

AZU1, C10orf105, C19orf76, CDH23, ENSA, FES, GABBR1, HCCA2, ICAM4, IFRD1, IKZF3, LBH, LCOR, LGALS2, MBNL1, MPO, NLK, PRKCZ, PTPRJ, RPUSD3, SAMD3, SCRN1, SLC24A4, SPTLC2, and *TRIB1*. None of the probes in the green module passed pre-determined cut-offs. Detailed gene-trait associations with GS and MM of the turquoise module are presented in the Supplemental Table S2. Interestingly, hub genes were represented by 34 CpG sites and only one transcript. Since one of the criteria for selection of hub genes is the MM (> 0.8), we tested potential difference in the MM of expression transcripts and methylation CpG sites in the turquoise module. The mean MM of expression transcripts (0.064 ± 0.40) was significantly lower (p=0.01) than the mean MM of methylation CpG sites (0.14 ± 0.39). MM was defined as the correlation of the MEs and the gene expression or DNA methylation levels. Thus, the variability in gene expression and DNA methylation levels may impact this correlation (i.e., the MM). Across all 240 transcripts, expression levels had a mean coefficient of variation (CV) significantly lower than the mean CV of methylation levels of all 466 CpG sites (0.013 ± 0.001 vs 0.16 ± 0.18, p<0.0001). Thus, the fact that CpG sites represent 34 of the 35 hub probes seems to be explained by their higher MM due to a greater variability in methylation levels as compared to expression levels.

Topological network analysis

In order to better define gene-gene interactions in the turquoise and green modules, topological network analysis was performed. In the turquoise module (381 probes), only annotated probes with a weighted correlation cut-off of 0.42 were included. Thus, a total of 15 unique genes were included. Three genes with more than four connections (*AFF1, ARHGEF10*, and *CDH23*) were central to the network (Figure 3). Considering that genes in the green module were weakly inter-connected (highest weighted correlation of 0.14), the analysis was not done in this module.

Discussion

The aim of the present study was to examine the potential role of multi-omics data in the interconnection between diet, represented by total carotenoids, and lipid profile using WGCNA. First, association between multi-omics data and plasma total carotenoid concentrations was assessed. Based on previous studies by our group (27, 28), transcripts and CpG sites associated with total carotenoid concentrations were identified. This is concordant with the previously reported effects of carotenoids on gene expression (8) and DNA methylation (9). For the purpose of the multi-omics analysis, genes identified at both gene expression and DNA methylation levels were retained to study their combined effects on lipid profile.

Second, WGCNA was used to link multi-omics data, associated with carotenoids, to lipid profile. WGCNA yielded six modules including the turquoise and green modules, which were relevant considering their associations with the lipid profile. Interestingly, in the turquoise module, 72 out of 228 unique genes were in common between expression transcripts and methylation CpG sites, suggesting a concordance in the potential biological effects of some genes. The turquoise module seems biologically plausible since its ME is correlated with several plasma lipids. In addition, as shown by their GS, genes clustering in this module are also correlated with lipids, and have biological functions as well as associations reported with lipids in the literature. Indeed, among the 28 unique hub genes identified, the following nine hub genes were of interest in the context of lipid profile: *AFF1, ARHGEF10, LCOR, LGALS2, MBNL1, MPO, SLC24A4, SPTLC2*, and *TRIB1*. The correlation graphs between methylation levels of CpG sites within these genes, lipid concentrations, and ME of the turquoise module are presented in Supplemental Figures S2 to S5. *AFF1* encodes for the AF4/FMR2 family member 1 implicated in lymphocyte development and autoimmune disease (34). Single nucleotide polymorphisms (SNPs) within this gene have been associated with plasma TG and HDL-C concentrations (35, 36). *ARHGEF10* encodes for the Rho guanine nucleotide exchange factor 10. Interestingly, a study by our group reported that a SNP within this gene alters the activity of the delta-6 desaturase and influences susceptibility to hypertriglyceridemia (37). A genetic variant within this gene was also associated with atherosclerotic stroke in Chinese population (38). *LCOR* encodes for the ligand dependent nuclear receptor corepressor that regulates lipogenesis and may be a potential target for treating hepatic steatosis (39). *LGALS2* encodes for the galectin 2. Variations in this gene were associated with inflammatory biomarkers, cellular adhesion molecules, risk of coronary heart disease, and also with insulin-glucose profile (40, 41). *MBNL1*, encoding for the muscle blind like splicing regulator 1, modulates alternative splicing of pre-mRNAs. A variation in this gene has been associated with HDL-C levels (35). *MPO* gene encodes for a myeloperoxidase, a pro-inflammatory enzyme stored in granulocytes (42). Myeloperoxidase also promotes oxidation of HDL particles potentially involved in the development of atherosclerosis. An increase in MPO concentration leads to a decrease in ApoAI and HDL-C levels and disturbs HDL function (43). Moreover, the level of myeloperoxidase was significantly lower in rats following a supplementation with lycopene (44). *SLC24A4* encodes for the solute carrier family 24 member 4. Variations within this gene have been associated with lipoprotein concentrations and size measurements of

more than 1000 subjects of the GOLDN study (45). However, the role of *SLC24A4* in the lipid metabolism remains unclear. *SPTLC2* encodes for the serine palmitoyltransferase long chain base subunit 2. *SPTLC2*+/– macrophages have significantly lower sphingomyelin levels in plasma membrane and lipid rafts. This reduction impaired inflammatory responses and enhanced reverse cholesterol transport mediated by ABC transporters (46). Liver *SPTLC2* deficiency also decreased plasma sphingomyelin levels in both HDL and non-HDL fractions (47). *TRIB1* encodes for the tribbles pseudokinase 1. Genetic variants within *TRIB1* have been associated with plasma lipid concentrations (TG, HDL-C and TC) (48) and with an increased risk of ischemic heart disease (48, 49). Genetic variant in *TRIB1* have an effect on alpha-tocopherol levels (6). Even though the ME representing probes in the green module correlated with HDL-C, none of the probes passed pre-determined cut-offs of GS and MM. Moreover, topological network analysis was performed in the turquoise module to investigate gene-gene interactions independently of the GS and MM. A total of three central genes (*AFF1, ARHGEF10*, and *CDH23*) were identified. These genes were also highlighted as hub genes in the previous analysis.

This multi-omics approach allowed to better understand the potential underlying mechanisms of action of plasma carotenoids on lipid profile. It provided a more global understanding of the role of genome-wide DNA methylation and gene expression in this association. Indeed, WGCNA was computed on both DNA methylation and gene expression levels to evaluate their combined effects on the plasma lipid profile. We compared results obtained in the combined WGCNA with results obtained in WGCNA independently performed on methylation and expression. Only a small proportion of genes (n=62) within the modules of interest (i.e., that showed correlations with plasma lipid concentrations) were in common between methylation and expression analysis in previous studies. Interestingly, vast majority of probes found in the module of interest, obtained in WGCNA independently performed on methylation and expression, were also identified in the present combined WGCNA. These results suggest that a combined WGCNA can highlight genes of important modules of interest observed in WGCNA independently performed on methylation and expression data.

The present study has several strengths. To the best of our knowledge, this is the first study to examine the potential role of multi-omics data in the interconnection between plasma carotenoids and the lipid profile. This approach allowed evaluating the combined effect of DNA methylation and gene expression on lipid profile. WGCNA also allowed relating large omics data sets to phenotypic traits, while reducing multiple comparison burdens by clustering correlated genes into modules (10). The study also considered genome-wide omics data and six predominant plasma carotenoid concentrations to obtain a more complete picture of the association. However, this study has some limitations. The main one is the small sample size. This may limit the statistical power to detect significant associations between total carotenoids and DNA methylation and gene expression levels. It may also reduce variability in DNA methylation and gene expression levels, and thus limit the identification of hub genes, based on the MM, in the WGCNA. Finally,

our study did not account for dietary profiles, physical activity, smoking, and alcohol consumption, which may affect circulating carotenoid concentrations (50, 51).

In conclusion, genome-wide DNA methylation and gene expression levels were both associated with plasma total carotenoid concentrations. WGCNA clustered multi-omics data into modules that were further linked to lipid profile to reveal key hub genes involved in lipid metabolism, plasma lipid concentrations and carotenoids. This integrative analysis provides a comprehensive understanding of the interactive molecular system between carotenoids, DNA methylation, gene expression, and the plasma lipid profile.

Acknowledgement

We would like to thank Christian Couture, Véronique Garneau, Catherine Raymond, and Véronique Richard who contributed to the success of this study.

Statement of Ethics

All participants (adults and children) signed an informed consent document. Parental consent was also obtained by signing the child consent document. The experimental protocol was approved by the Ethics Committees of Laval University Hospital Research Center and Laval University.

Disclosure Statement

The authors have no conflicts of interest to declare.

Funding Sources

This work was supported by the Canada Research Chair in Genomics Applied to Nutrition and Metabolic Health. MCV is Tier 1 Canada Research Chair in Genomics Applied to Nutrition and Metabolic Health. BLT is a recipient of a scholarship from Canadian Institutes of Health Research (CIHR).

Author Contributions

Each author contribution to work: BL and MCV designed research; BLT, FG, and LP conducted research and performed statistical analysis; BLT wrote the paper; BLT and MCV have primary responsibility for final content. All authors read and approved the final manuscript.

References

1. Martinez-Gonzalez MA, de la Fuente-Arrillaga C, Lopez-Del-Burgo C, Vazquez-Ruiz Z, Benito S, Ruiz-Canela M. Low consumption of fruit and vegetables and risk of chronic disease: a review of the epidemiological evidence and temporal trends among Spanish graduates. Public health nutrition. 2011;14(12A):2309-15.

2. Souverein OW, de Vries JH, Freese R, Watzl B, Bub A, Miller ER, et al. Prediction of fruit and vegetable intake from biomarkers using individual participant data of diet-controlled intervention studies. The British journal of nutrition. 2015;113(9):1396-409.

3. Couillard C, Lemieux S, Vohl MC, Couture P, Lamarche B. Carotenoids as biomarkers of fruit and vegetable intake in men and women. The British journal of nutrition. 2016;116(7):1206-15.

4. Allore T, Lemieux S, Vohl MC, Couture P, Lamarche B, Couillard C. Correlates of the difference in plasma carotenoid concentrations between men and women. The British journal of nutrition. 2018:1-10.

5. Clevidence BA, Bieri JG. Association of carotenoids with human plasma lipoproteins. Methods in enzymology. 1993;214:33-46.

6. Ferrucci L, Perry JR, Matteini A, Perola M, Tanaka T, Silander K, et al. Common variation in the beta-carotene 15,15'-monooxygenase 1 gene affects circulating levels of carotenoids: a genome-wide association study. American journal of human genetics. 2009;84(2):123-33.

7. Zubair N, Kooperberg C, Liu J, Di C, Peters U, Neuhouser ML. Genetic variation predicts serum lycopene concentrations in a multiethnic population of postmenopausal women. The Journal of nutrition. 2015;145(2):187-92.

8. Sharoni Y, Danilenko M, Dubi N, Ben-Dor A, Levy J. Carotenoids and transcription. Archives of biochemistry and biophysics. 2004;430(1):89-96.

9. Bollati V, Favero C, Albetti B, Tarantini L, Moroni A, Byun HM, et al. Nutrients intake is associated with DNA methylation of candidate inflammatory genes in a population of obese subjects. Nutrients. 2014;6(10):4625-39.

10. Langfelder P, Horvath S. WGCNA: an R package for weighted correlation network analysis. BMC bioinformatics. 2008;9:559.

11. Dunn WB, Broadhurst DI, Atherton HJ, Goodacre R, Griffin JL. Systems level studies of mammalian metabolomes: the roles of mass spectrometry and nuclear magnetic resonance spectroscopy. Chemical Society reviews. 2011;40(1):387-426.

12. Sun YV, Hu YJ. Integrative Analysis of Multi-omics Data for Discovery and Functional Studies of Complex Human Diseases. Advances in genetics. 2016;93:147-90.

13. Bakker GC, van Erk MJ, Pellis L, Wopereis S, Rubingh CM, Cnubben NH, et al. An antiinflammatory dietary mix modulates inflammation and oxidative and metabolic stress in overweight men: a nutrigenomics approach. The American journal of clinical nutrition. 2010;91(4):1044-59.

14. Tremblay BL, Guenard F, Lamarche B, Perusse L, Vohl MC. Familial resemblances in blood leukocyte DNA methylation levels. Epigenetics. 2016;11(11):831-8.

15. Tremblay BL, Guenard F, Lamarche B, Perusse L, Vohl MC. Familial resemblances in human whole blood transcriptome. BMC genomics. 2018;19(1):300.

16. Tremblay BL, Guenard F, Lamarche B, Perusse L, Vohl MC. Familial resemblances in human plasma metabolites are attributable to both genetic and common environmental effects. Nutrition research. 2019;61:22-30.

17. Tremblay BL, Guenard F, Lamarche B, Perusse L, Vohl MC. Genetic and Common Environmental Contributions to Familial Resemblances in Plasma Carotenoid Concentrations in Healthy Families. Nutrients. 2018;10(8).

18. Callaway C, Chumlea W, Bouchard C, Himes J, Lohman T, Martin A, et al. Standardization of anthropomeric measurements : The Airlie (VA) Consensus Conference. Champaign, IR, USA; 1988.

19. McNamara JR, Schaefer EJ. Automated enzymatic standardized lipid analyses for plasma and lipoprotein fractions. ClinChimActa. 1987;166(1):1-8.

20. Burstein M, Samaille J. [On a rapid determination of the cholesterol bound to the serum alpha- and beta-lipoproteins]. ClinChimActa. 1960;5:609.

21. Albers JJ, Warnick GR, Wiebe D, King P, Steiner P, Smith L, et al. Multi-laboratory comparison of three heparin-Mn2+ precipitation procedures for estimating cholesterol in high-density lipoprotein. ClinChem. 1978;24(6):853-6.

22. Friedewald WT, Levy RI, Fredrickson DS. Estimation of the concentration of low-density lipoprotein cholesterol in plasma, without use of the preparative ultracentrifuge. ClinChem. 1972;18(6):499-502.

23. Laurell CB. Quantitative estimation of proteins by electrophoresis in agarose gel containing antibodies. AnalBiochem. 1966;15(1):45-52.

24. Blazejczyk M MM, Nadon R. FlexArray: A statistical data analysis software for gene expression microarrays. Genome Quebec, Montreal, Canada. 2007.

25. Price ME, Cotton AM, Lam LL, Farre P, Emberly E, Brown CJ, et al. Additional annotation enhances potential for biologically-relevant analysis of the Illumina Infinium HumanMethylation450 BeadChip array. Epigenetics & chromatin. 2013;6(1):4.

26. Team RC. R: A language and environment for statistical computing. R Foundation for Statistical Computing. Vienna, Austria. 2013.

27. Tremblay BL, Guenard F, Lamarche B, Perusse L, Vohl MC. Weighted gene co-expression network analysis to explain the relationship between plasma total carotenoids and lipid profile. Genes & nutrition. 2019;14:16.

28. Tremblay BL, Guenard F, Lamarche B, Perusse L, Vohl MC. Network Analysis of the Potential Role of DNA Methylation in the Relationship between Plasma Carotenoids and Lipid Profile. Nutrients. 2019;11(6):1265.

29. Langfelder P, Horvath S. Fast R Functions for Robust Correlations and Hierarchical Clustering. Journal of statistical software. 2012;46(11).

30. Zhang B, Horvath S. A general framework for weighted gene co-expression network analysis. Statistical applications in genetics and molecular biology. 2005;4:Article17.

31. Langfelder P, Horvath S. Eigengene networks for studying the relationships between co-expression modules. BMC systems biology. 2007;1:54.

32. Horvath S, Zhang B, Carlson M, Lu KV, Zhu S, Felciano RM, et al. Analysis of oncogenic signaling networks in glioblastoma identifies ASPM as a molecular target. Proceedings of the National Academy of Sciences of the United States of America. 2006;103(46):17402-7.

33. VisANT 5.0. Visual analyses of metabolic networks in cells and ecosystems.

http://visant.bu.edu. Accessed 18 December 18. 2018 [Available from: http://visant.bu.edu.

34. Sun QQ, Hua DJ, Huang SC, Cen H, Zhou L, Shao S. Association study of AFF1 rs340630 polymorphism with genetic susceptibility to rheumatoid arthritis in Chinese population. Brazilian journal of medical and biological research = Revista brasileira de pesquisas medicas e biologicas. 2018;51(7):e7126.

35. Klarin D, Damrauer SM, Cho K, Sun YV, Teslovich TM, Honerlaw J, et al. Genetics of blood lipids among ~300,000 multi-ethnic participants of the Million Veteran Program. Nature genetics. 2018;50(11):1514-23.

36. Spracklen CN, Chen P, Kim YJ, Wang X, Cai H, Li S, et al. Association analyses of East Asian individuals and trans-ancestry analyses with European individuals reveal new loci associated with cholesterol and triglyceride levels. Human molecular genetics. 2017;26(9):1770-84.

37. de Toro-Martin J, Guenard F, Rudkowska I, Lemieux S, Couture P, Vohl MC. A common variant in ARHGEF10 alters delta-6 desaturase activity and influence susceptibility to hypertriglyceridemia. Journal of clinical lipidology. 2018;12(2):311-20 e3.

38. Li H, Yu S, Wang R, Sun Z, Zhou X, Zheng L, et al. ARHGEF10 gene polymorphism is closely associated with the risk of ischemic stroke in Northern Han Chinese population. Neurological research. 2017;39(2):158-64.

39. Song Y, Shan S, Zhang Y, Liu W, Ding W, Ren W, et al. Ligand-dependent corepressor acts as a novel corepressor of thyroid hormone receptor and represses hepatic lipogenesis in mice. Journal of hepatology. 2012;56(1):248-54.

40. Asselbergs FW, Pai JK, Rexrode KM, Hunter DJ, Rimm EB. Effects of lymphotoxin-alpha gene and galectin-2 gene polymorphisms on inflammatory biomarkers, cellular adhesion molecules and risk of coronary heart disease. Clinical science. 2007;112(5):291-8.

41. Christensen MB, Lawlor DA, Gaunt TR, Howell WM, Davey Smith G, Ebrahim S, et al. Genotype of galectin 2 (LGALS2) is associated with insulin-glucose profile in the British Women's Heart and Health Study. Diabetologia. 2006;49(4):673-7.

42. Zou S, Pan X, Hua C, Wu M, He B, Chen Z. Myeloperoxidase -463 G/A polymorphism is associated with lung cancer risk: A meta-analysis with 7420 cases and 9132 controls. Journal of cancer research and therapeutics. 2018;14(Supplement):S282-S7.

43. Kimak E, Zieba B, Duma D, Solski J. Myeloperoxidase level and inflammatory markers and lipid and lipoprotein parameters in stable coronary artery disease. Lipids Health Dis. 2018;17(1):71.

44. Reifen R, Nur T, Matas Z, Halpern Z. Lycopene supplementation attenuates the inflammatory status of colitis in a rat model. International journal for vitamin and nutrition research Internationale Zeitschrift fur Vitamin- und Ernahrungsforschung Journal international de vitaminologie et de nutrition. 2001;71(6):347-51.

45. Kraja AT, Borecki IB, Tsai MY, Ordovas JM, Hopkins PN, Lai CQ, et al. Genetic analysis of 16 NMR-lipoprotein fractions in humans, the GOLDN study. Lipids. 2013;48(2):155-65.

46. Chakraborty M, Lou C, Huan C, Kuo MS, Park TS, Cao G, et al. Myeloid cell-specific serine palmitoyltransferase subunit 2 haploinsufficiency reduces murine atherosclerosis. The Journal of clinical investigation. 2013;123(4):1784-97.

47. Li Z, Li Y, Chakraborty M, Fan Y, Bui HH, Peake DA, et al. Liver-specific deficiency of serine palmitoyltransferase subunit 2 decreases plasma sphingomyelin and increases apolipoprotein E levels. The Journal of biological chemistry. 2009;284(39):27010-9.

48. Willer CJ, Sanna S, Jackson AU, Scuteri A, Bonnycastle LL, Clarke R, et al. Newly identified loci that influence lipid concentrations and risk of coronary artery disease. Nature genetics. 2008;40(2):161-9.

49. Varbo A, Benn M, Tybjaerg-Hansen A, Grande P, Nordestgaard BG. TRIB1 and GCKR polymorphisms, lipid levels, and risk of ischemic heart disease in the general population. Arteriosclerosis, thrombosis, and vascular biology. 2011;31(2):451-7.

50. Gruber M, Chappell R, Millen A, LaRowe T, Moeller SM, Iannaccone A, et al. Correlates of serum lutein + zeaxanthin: findings from the Third National Health and Nutrition Examination Survey. The Journal of nutrition. 2004;134(9):2387-94.

51. Al-Delaimy WK, van Kappel AL, Ferrari P, Slimani N, Steghens JP, Bingham S, et al. Plasma levels of six carotenoids in nine European countries: report from the European Prospective Investigation into Cancer and Nutrition (EPIC). Public health nutrition. 2004;7(6):713-22.

Table

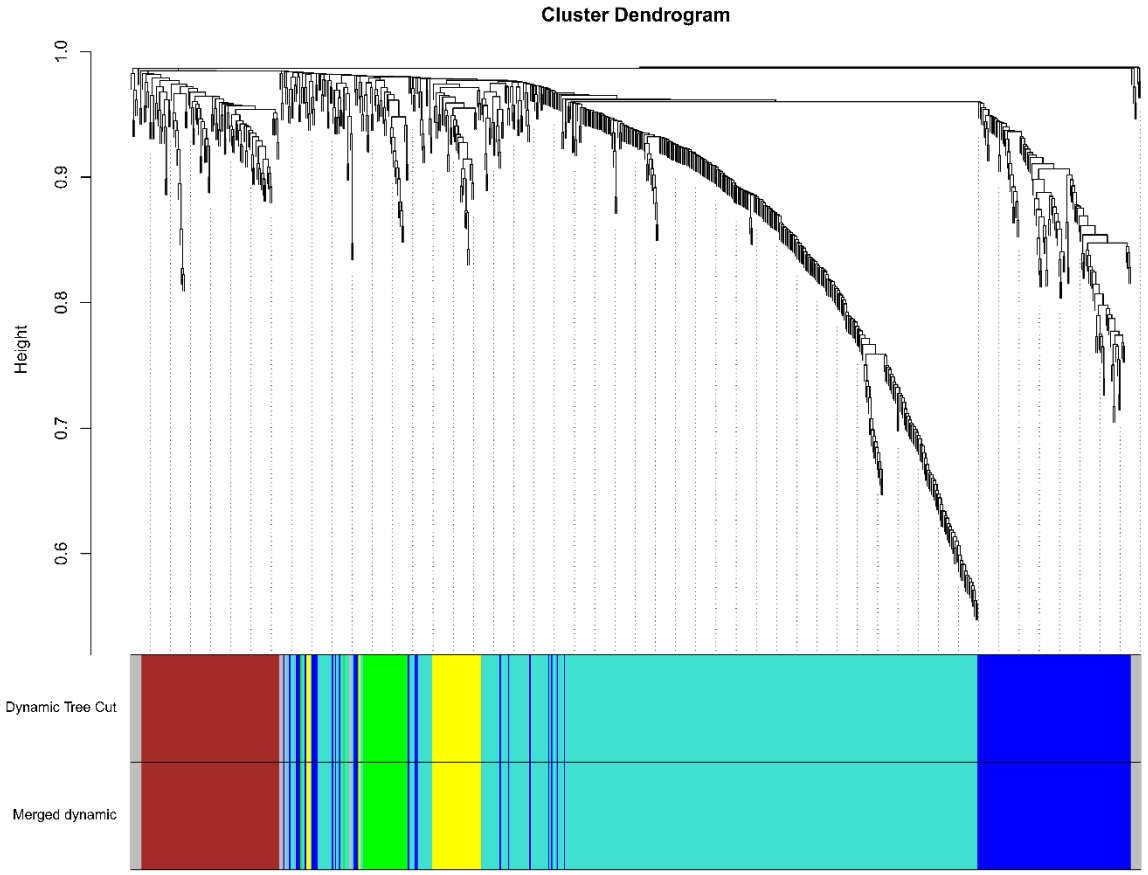
Table 9.1 Characteristic and lipid profile parameters of study subjects

Biochemical parameters	Parents (n=22)	Children (n=26)
Age (years)	42.3 ± 5.3	11.3 ± 3.4
BMI (kg/m^2)	23.9 ± 3.0	-
BMI percentile	-	50 ± 31.1
TC (mmol/L)	4.68 ± 0.55	4.28 ± 0.51
HDL-C (mmol/L)	1.63 ± 0.38	1.55 ± 0.29
LDL-C (mmol/L)	2.61 ± 0.55	2.31 ± 0.45
ApoB100 (g/L)	0.80 ± 0.15	0.70 ± 0.13
TG (mmol/L)	0.95 ± 0.35	0.93 ± 0.39
Total carotenoids	6.35 ± 2.39	5.70 ± 2.05

All values are means ± SD. Abbreviations: Apolipoprotein B100 (ApoB100), High-density lipoprotein cholesterol (HDL-C), Low-density lipoprotein cholesterol (LDL-C), Standard deviation (SD), Total cholesterol (TC), Triglycerides (TG).

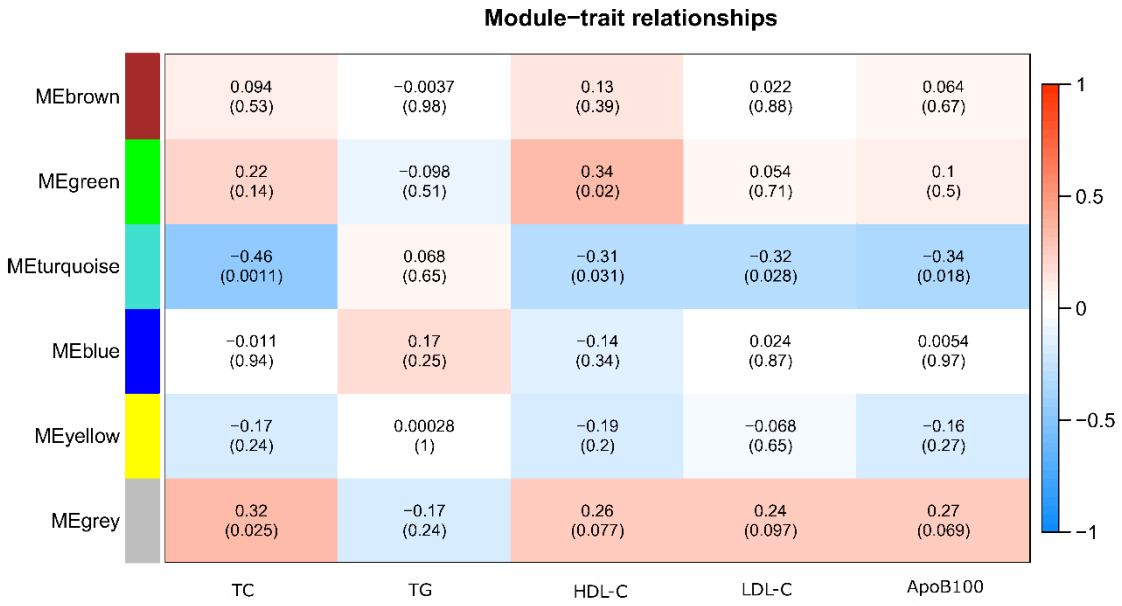
Figures

Figure 9.1 Gene dendrogram obtained using average linkage hierarchical clustering



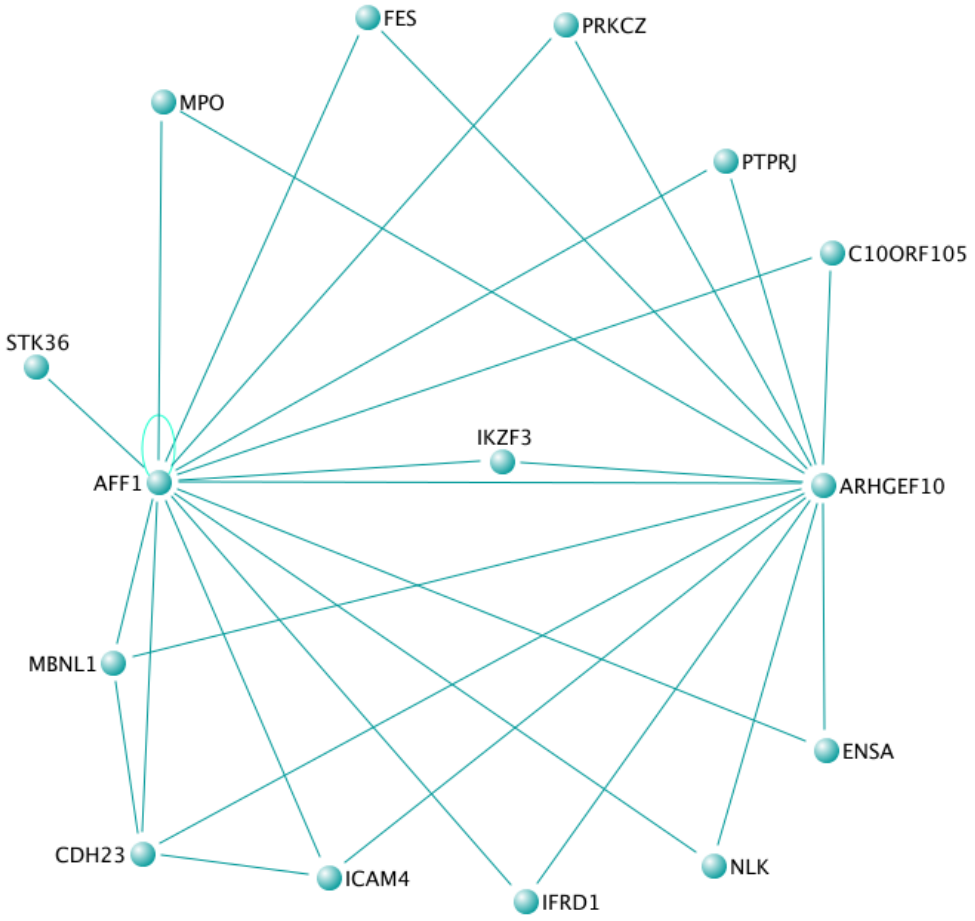
Six module colors are shown correspondingly. The merged dynamic yielded same modules as the dynamic tree cut using a cut-off of 0.25.

Figure 9.2 Heatmap of module-trait relationships depicting correlations between module eigengenes and lipid profile traits



Numbers in the table correspond to the correlation r and the *P*-value in parentheses. The degree of correlation is illustrated with the color legend. Abbreviations: Apolipoprotein B100 (ApoB100), High-density lipoprotein cholesterol (HDL-C), Low-density lipoprotein cholesterol (LDL-C), Module eigengene (ME), Total cholesterol (TC), Triglycerides (TG).

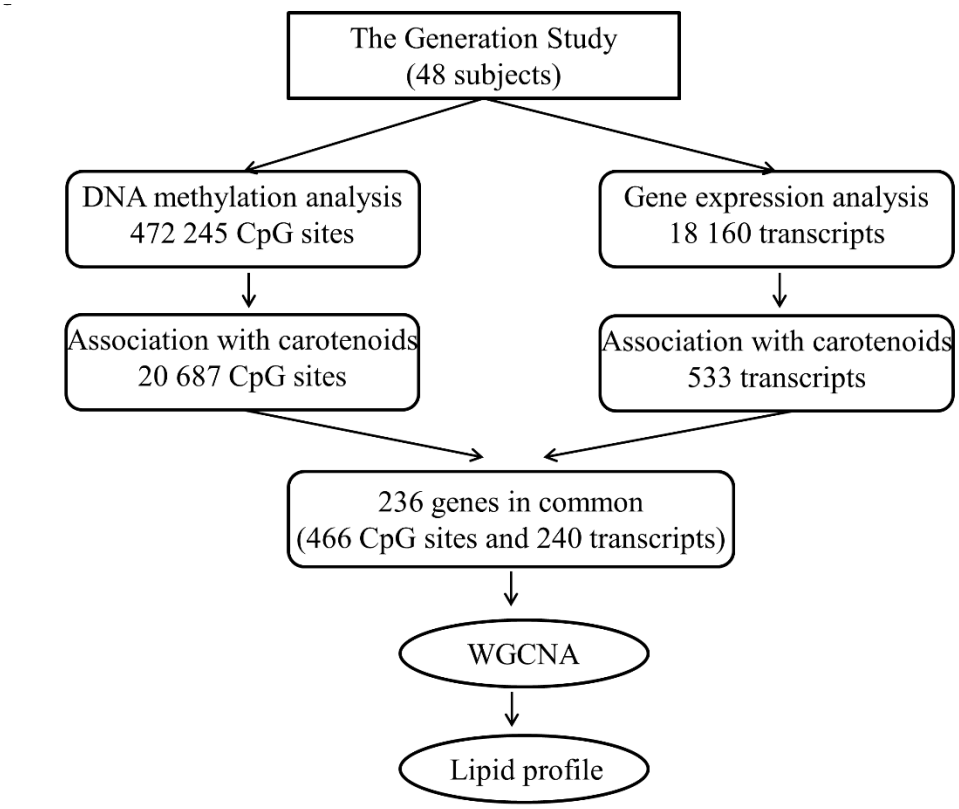
Figure 9.3 Topological interaction network of 15 unique genes in the turquoise module



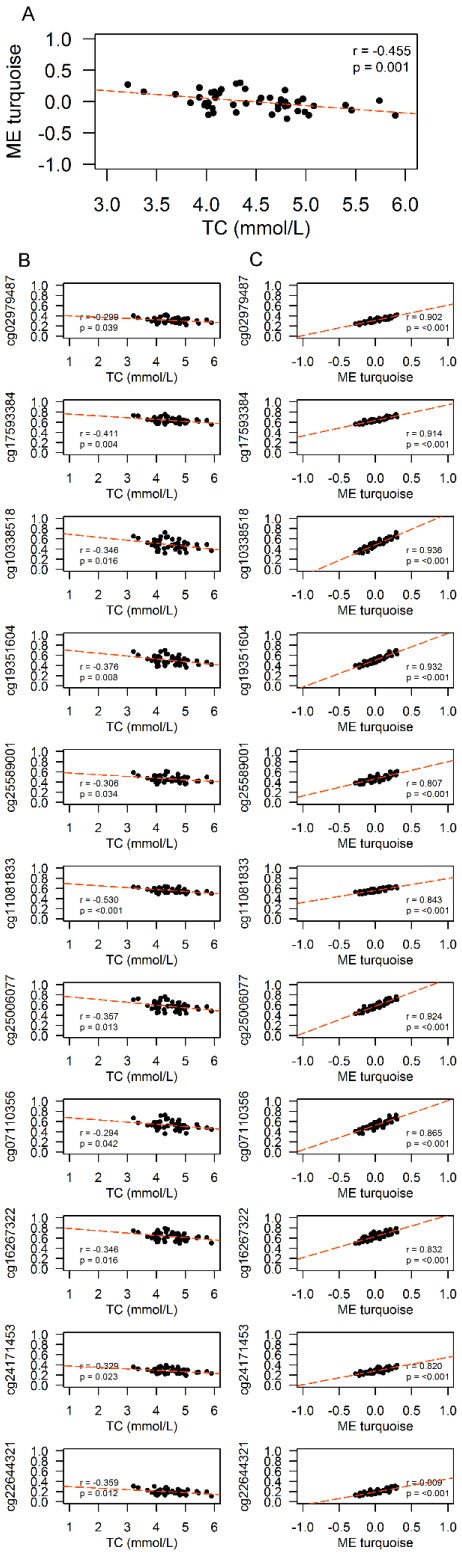
Each gene is represented by a node and edge number is proportional to connection strength. Gene-gene interaction network was constructed and visualized using VisANT 5.0.

Supplemental material

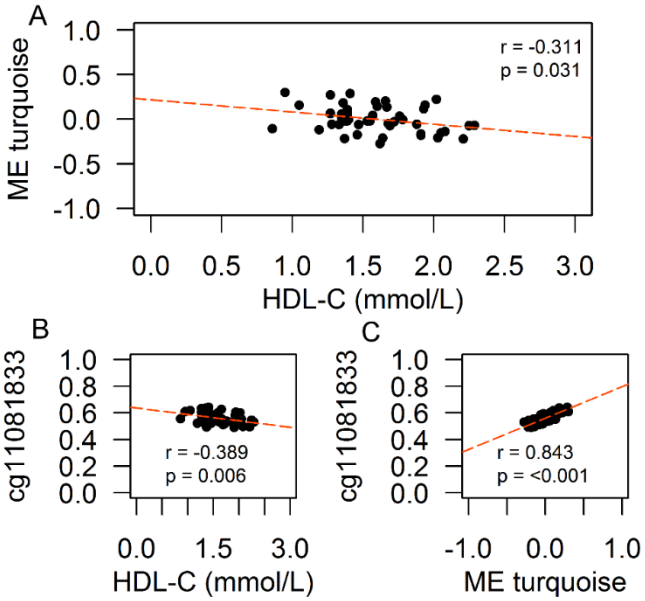
Supplemental figure 9.1 Study schematic illustrating different steps of analysis leading to WGCNA



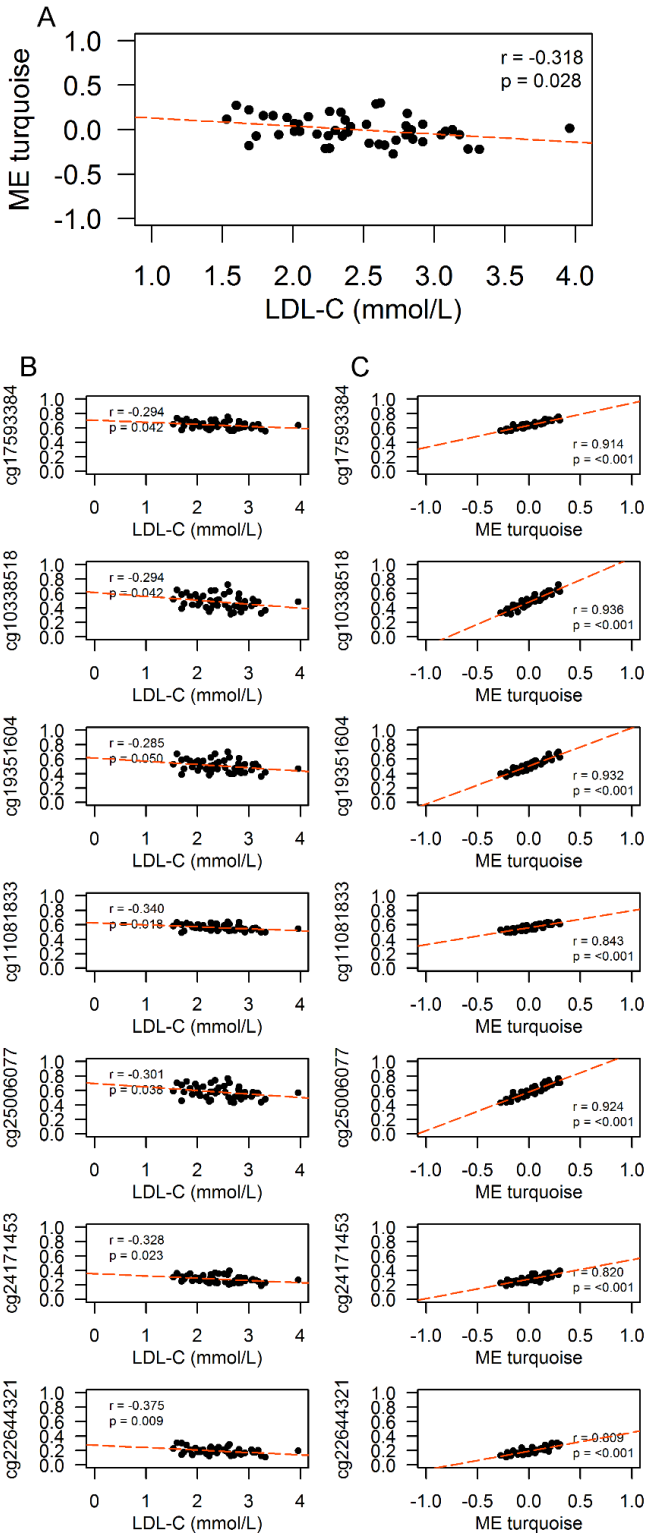
Supplemental figure 9.2 Correlation graphs of A) total cholesterol and ME of the turquoise module, B) methylation levels of CpG sites within hub genes and total cholesterol, C) methylation levels of CpG sites within hub genes and ME of the turquoise module



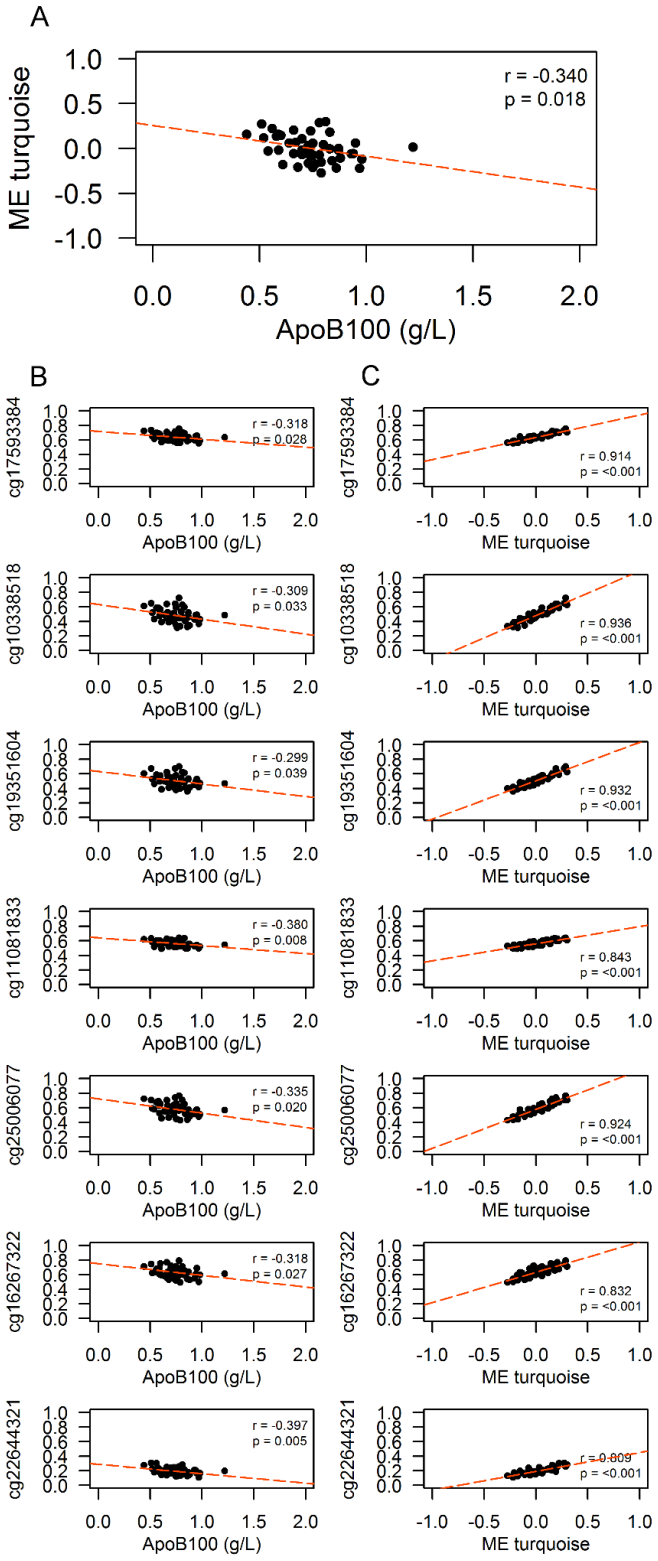
Supplemental figure 9.3 Correlation graphs of A) HDL-cholesterol and ME of the turquoise module, B) methylation levels of CpG site within *LGALS2* and HDL-cholesterol, C) methylation levels of CpG site within *LGALS2* and ME of the turquoise module



Supplemental figure 9.4 Correlation graphs of A) LDL-cholesterol and ME of the turquoise module, B) methylation levels of CpG sites within hub genes and LDL-cholesterol, C) methylation levels of CpG sites within hub genes and ME of the turquoise module



Supplemental figure 9.5 Correlation graphs of A) ApoB100 and ME of the turquoise module, B) methylation levels of CpG sites within hub genes and ApoB100, C) methylation levels of CpG sites within hub genes and ME of the turquoise module



Supplemental table 9.1 CpG sites and transcripts associated with total carotenoid concentrations (n=706)

Supplemental table 9.2 Hub genes identified from WGCNA analysis in the turquoise module

Supplemental tables are available at https://www.karger.com/Article/FullText/503828

Discussion et conclusion

Retour sur les objectifs

L'objectif général de cette thèse était d'étudier l'impact des déterminants omiques sur la relation entre l'alimentation et la santé CM. Le premier objectif spécifique, utilisant une approche de la génétique quantitative, était de caractériser l'héritabilité de la méthylation de l'ADN, de l'expression génique, des métabolites plasmatiques et des caroténoïdes plasmatiques ainsi que de vérifier si le lien avec des facteurs de risque CM pouvait être expliqué par des facteurs génétiques et environnementaux. Le deuxième objectif spécifique, utilisant une approche de réseaux de corrélations pondérées, était d'évaluer le rôle des données omiques individuelles et combinées dans la relation entre les caroténoïdes plasmatiques et le profil lipidique.

Dans un premier temps, nous avons utilisé l'approche de la génétique quantitative, tel que rapporté dans les quatre articles présentés aux chapitres 3 à 6. Le schéma ci-dessous (Figure 1) illustre les interactions étudiées à l'aide de cette approche, soit l'héritabilité génétique en rouge, l'effet de l'environnement partagé en vert et l'analyse génétique bivariée en bleu.

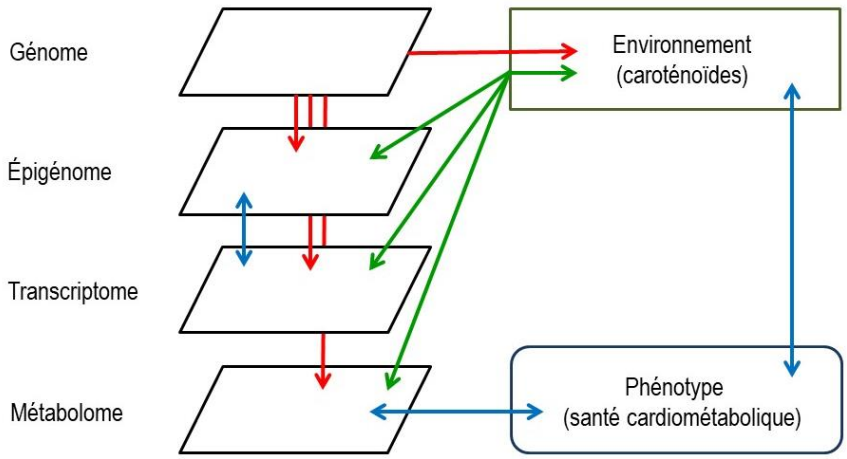


Figure 1 Schéma représentant les différentes interactions de la biologie des systèmes étudiées à l'aide de l'approche de la génétique quantitative
Figure adaptée de Dunn *et al*. *Chemical Society Reviews* 2011 (17). L'héritabilité génétique est représentée en rouge, l'effet de l'environnement partagé en vert et l'analyse génétique bivariée en bleu.

Cette approche a fait ressortir que toutes les données omiques étudiées et les caroténoïdes plasmatiques démontraient des ressemblances familiales dues, à des degrés divers, à l'effet de la génétique et de l'environnement partagé. L'analyse génétique bivariée a démontré que des facteurs génétiques et environnementaux étaient également impliqués dans le lien entre la méthylation et l'expression ainsi qu'entre les métabolites, les caroténoïdes et les facteurs de risque CM.

Dans un deuxième temps, nous avons eu recours à l'approche de réseaux de corrélations pondérées, tel que décrit dans les trois articles présentés aux chapitres 7 à 9. Ce faisant, nous avons pu mettre en relation les caroténoïdes, la méthylation de l'ADN, l'expression des gènes et le profil lipidique, comme l'illustre le schéma ci-dessous (Figure 2). La méthylation et l'expression à l'échelle du génome étaient associées aux concentrations plasmatiques de caroténoïdes totaux. Plusieurs *hub genes* associés aux caroténoïdes étaient aussi potentiellement reliés au profil lipidique. Ainsi, nous avons été à même de mieux comprendre le système moléculaire par l'entremise duquel interagissent les caroténoïdes, la méthylation, l'expression et les lipides.

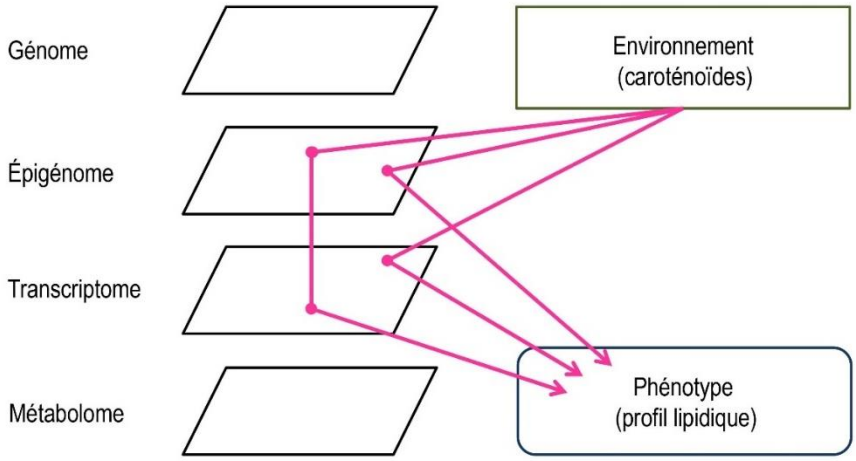


Figure 2 Schéma représentant les différentes interactions de la biologie des systèmes étudiées à l'aide de l'approche de réseaux de corrélations pondérées
Figure adaptée de Dunn *et al*. *Chemical Society Reviews* 2011 (17).

Conformément à nos objectifs, ces travaux ont mis en lumière l'impact des déterminants omiques, pris individuellement et de manière combinée, sur la relation entre l'alimentation et la santé CM. Toutefois, force est de constater que le système moléculaire interactif qui relie l'alimentation et la santé CM est extrêmement complexe et que plusieurs autres interactions doivent être élucidées. Des approches menant à une compréhension globale de l'ensemble des interactions sont donc requises.

Nos travaux ont tout de même un potentiel d'application en nutrition de précision. Ils peuvent servir à bâtir un outil de prédiction du risque CM. Les données omiques pourraient compléter les facteurs de risque CM traditionnels afin d'aider à prédire de façon plus précise le risque CM. Cet outil de prédiction (score de risque CM) inclurait les données omiques, validées dans d'autres cohortes, qui démontrent une forte héritabilité et qui sont associées à l'alimentation ainsi qu'aux facteurs de risque CM. En incluant l'héritabilité, on pourrait identifier des biomarqueurs omiques du risque CM qui ont le potentiel d'exercer leurs effets à travers les générations. Ce score pourrait être combiné, par exemple, avec un score des habitudes de vie comprenant la qualité

nutritionnelle, l'activité physique et le tabagisme. L'inclusion de ces facteurs environnementaux modifiables pourrait amoindrir le déterminisme perçu souvent associé aux scores de risque polygénique (182). Le but ultime, de façon générale, serait de mieux prévenir les maladies CM en caractérisant plus précisément le risque des individus et des familles, en ciblant les individus à haut risque et en leur fournissant les outils et le suivi nécessaires pour qu'ils modifient leurs habitudes de vie et se protègent davantage contre les complications CM.

Forces et limites des travaux

Le projet à l'origine de cette thèse est novateur étant donné l'utilisation de données omiques pour étudier la relation entre l'alimentation et la santé. Peu d'études considèrent les déterminants omiques dans cette relation et encore moins de façon intégrée. L'utilisation de deux approches d'analyse complémentaires représente également un aspect novateur. Les paragraphes qui suivent constituent une discussion des forces et des limites des travaux, présentée selon les deux approches d'analyse utilisées.

Avec l'approche de la génétique quantitative, l'une des principales forces des travaux est l'étude de données omiques à l'échelle du génome. En effet, l'héritabilité a été estimée pour 472 494 sites CpG et 18 160 transcrits. Les logiciels d'héritabilité couramment utilisés ont été conçus pour évaluer l'héritabilité d'un trait qui présente une ou quelques valeurs par individu, comme par exemple la taille. Dans notre cas, le calcul d'héritabilité d'une grande masse de données, soit près de 23 millions de valeurs (472 494 sites CpG x 48 individus) dans le cas de la méthylation, met rudement à l'épreuve la capacité de calcul des ordinateurs. Une autre force de nos travaux réside dans l'inclusion de l'effet de l'environnement partagé dans le modèle d'héritabilité. La majorité des études rapportent l'héritabilité génétique sans tenir compte de l'effet de l'environnement. Considérant l'effet important de facteurs environnementaux comme l'alimentation sur les déterminants omiques, l'inclusion de cette composante dans le modèle est plus que pertinente. Nous sommes d'avis que cette inclusion a permis de mieux quantifier les ressemblances familiales dans notre cohorte. L'utilisation de l'analyse génétique bivariée, dérivée de la méthode de la décomposition de la variance, s'est aussi avérée être une force, en nous permettant d'évaluer l'implication des facteurs génétiques et environnementaux dans le lien entre les données omiques et les facteurs de risque CM. De la même façon, les différentes analyses de voies métaboliques ont permis de mieux comprendre les voies métaboliques des gènes présentant des niveaux de méthylation et d'expression héritables. Enfin, de nombreux ajustements ont été faits pour tenir compte des facteurs confondants, dont un ajustement pour la composition cellulaire du sang. La méthylation et l'expression sont influencées par la composition cellulaire du sang (216). Cette dernière, idéalement, est évaluée directement à l'aide du compte des cellules du sang (216). Dans notre cas, cette option n'étant pas possible, nous avons tout de même estimé la composition cellulaire du sang à l'aide d'algorithmes de prédiction. Ces algorithmes utilisent une méthode de déconvolution à partir des niveaux normalisés de méthylation (217) et d'expression (218) afin d'identifier les sites CpG et les transcrits dont les niveaux sont représentatifs du compte cellulaire.

L'approche de la génétique quantitative présente toutefois certaines limites. La plus importante, dans notre cas, est la taille de notre échantillon. En effet, le petit nombre de sujets (n = 48) limite la puissance statistique pour estimer les niveaux d'héritabilité. Selon Grundberg *et al*. un échantillon de 648 jumeaux n'était pas suffisant pour obtenir des estimés fiables d'héritabilité de moins de 10 % (173). Rijsdijk et Sham rapportent qu'au moins 200 paires de traits sont nécessaires pour obtenir des estimés d'héritabilité fiables pour un trait dont l'héritabilité est élevée, alors que le nombre de paires devrait être 10 à 20 fois plus grand pour les traits d'héritabilité modérée ou faible (219). La taille de l'échantillon de notre projet affecte donc la fiabilité des estimés d'héritabilité faible et pourrait expliquer qu'un grand nombre de sites CpG et de transcrits avaient une héritabilité estimée de 0%. De plus, la taille réduite de l'échantillon a peut-être nui lorsqu'il s'est agi de faire la distinction entre les effets génétiques et ceux de l'environnement partagé. En effet, les concepts de l'environnement partagé et de l'héritabilité épigénétique sont étroitement interreliés, et donc difficiles à séparer. Selon McRae *et al*. un grand échantillon est nécessaire pour distinguer ces deux effets (200). Une autre limite est le fait que nous n'avions pas les données de génotypage pour les sujets. Nous n'avons donc pas pu évaluer l'effet du génotype sur les diverses données omiques et leurs impacts sur les estimés d'héritabilité. En ce sens, nous n'avons pas pu éliminer les sites CpG associés à des SNPs, ce qui peut expliquer le grand nombre de sites CpG présentant des estimés d'héritabilité maximale de 100 %. En effet, des analyses complémentaires avec des SNPs du *1000 Genomes Project* ont révélé que près de 86 % de ces sites CpG présentaient un ou plusieurs SNPs à leur localisation (220). Enfin, une autre limite résulte du fait que nous n'avons pas ajusté les concentrations plasmatiques de caroténoïdes pour les apports alimentaires en caroténoïdes. En effet, les outils d'évaluation alimentaire utilisés dans notre projet (un rappel de 24 heures pour les parents et les enfants et un questionnaire de fréquence alimentaire pour les parents) n'étaient pas suffisants pour fournir une estimation fiable des apports alimentaires.

Avec l'approche de réseaux de corrélations pondérées, nous considérons comme une force importante qu'elle nous ait permis de tenir compte des données omiques à l'échelle du génome de façon individuelle et combinée. En ce sens, la formation de modules de gènes interreliés permet de réduire considérablement le nombre de variables pour les associations. De plus, l'approche a permis d'identifier des *hub genes* associés autant aux caroténoïdes, aux données omiques qu'au profil lipidique. Notons en passant que cette approche est très rarement utilisée dans le domaine de la nutrition. Dans le but de mieux comprendre les mécanismes sous-jacents aux associations observées, des analyses d'enrichissement fonctionnel et des réseaux topologiques d'interaction ont été utilisés. Les analyses d'enrichissement fonctionnel ont permis d'identifier des voies métaboliques surreprésentées dans les modules de données omiques associés à la fois aux caroténoïdes et au profil lipidique. L'analyse de réseaux topologiques d'interaction a aussi permis d'illustrer les interactions entre les gènes les plus fortement corrélés au sein des modules d'intérêt.

Les limites de l'approche de réseaux de corrélations pondérées recoupent en partie celles de l'approche de la génétique quantitative. En effet, la taille de l'échantillon représente également la principale limite de cette approche. Cela a pu limiter la puissance statistique pour détecter des associations significatives entre les caroténoïdes et les données omiques, en plus de réduire la variabilité dans les niveaux de données omiques. Une autre limite est due au fait que nous n'avons pas validé les *hub genes* dans une autre cohorte. Pour terminer, nous n'avons pas fait d'ajustement pour divers facteurs (alimentation et activité physique) qui peuvent influencer les concentrations de caroténoïdes plasmatiques.

Perspectives

À la lumière de nos résultats, la prochaine étape serait de conduire des études supplémentaires pour valider les résultats et investiguer plus en profondeur certains aspects. Tout d'abord, dans une potentielle étude de génétique quantitative, il serait nécessaire de disposer d'un grand échantillon de familles afin d'augmenter la puissance statistique pour détecter l'héritabilité, surtout pour les traits d'héritabilité faible ou modérée. Un grand échantillon permettrait aussi de mieux cerner les sources de variance des données omiques dans la population (175). Considérant l'impact important de la génétique dans les analyses d'héritabilité, il serait impératif d'obtenir les données génomiques des sujets pour les associer avec les autres données omiques. En ce sens, l'ajout de données de protéomique et de métabolomique issues d'approches non-ciblées permettrait de mieux comprendre le système dynamique et interactif de couches moléculaires reliant l'alimentation et la santé CM. L'évaluation adéquate des apports alimentaires (à l'aide d'un questionnaire de fréquence et de quatre à six rappels de 24 heures (221)) et de l'activité physique permettrait de tenir compte de ces facteurs confondants. Finalement, il serait intéressant que la cohorte soit plus variée du point de vue de la santé CM en incluant des sujets qui présentent des IMC plus élevés et des facteurs de risque de MCV (p. ex. hypertension artérielle, dyslipidémies ou résistance à l'insuline). Cette étude pourrait utiliser le calcul de l'héritabilité et l'analyse génétique bivariée pour évaluer les ressemblances familiales des données omiques et l'implication de la génétique et de l'environnement dans leurs liens avec des profils alimentaires et des facteurs de risque CM. Il serait intéressant de vérifier les associations entre des profils alimentaires ou des nutriments spécifiques et les niveaux de méthylation et d'expression de sondes héritables avec une forte composante de l'environnement partagé. Enfin, plusieurs études dans le domaine de la cancérologie ont utilisé des approches multi-omiques pour affiner les classes de cancers et identifier de nouveaux biomarqueurs de manière à stratifier les patients en sous-types de cancers (176, 180). L'intégration de données multi-omiques permet aussi de stratifier les individus selon leur risque cardiovasculaire. Khera *et al*. ont estimé le risque de développer des maladies coronariennes selon un score polygénique et le niveau d'adhésion à de saines habitudes de vie (alimentation équilibrée, non-fumeur, IMC normal, pratique régulière d'activité physique) (182). De plus, le suivi de l'état de santé à l'aide d'approches multi-omiques pourrait permettre d'identifier des biomarqueurs précoces des MCV et de les prévenir grâce à la modification des habitudes de vie (222). En somme, l'analyse intégrée des données

omiques associées à l'alimentation a le potentiel de permettre l'identification de cibles moléculaires pour la prévention et le traitement des maladies chroniques (123).

De même, dans une potentielle étude de réseaux de corrélations pondérées, un plus grand échantillon de sujets non apparentés permettrait d'augmenter la variabilité des données omiques et des traits, en plus d'accroître la puissance statistique pour détecter des associations significatives. La validation des *hub genes* dans une autre cohorte serait primordiale. L'étude d'une cohorte plus variée d'un point de vue CM permettrait peut-être d'identifier plus d'associations entre les facteurs de risque CM et les données omiques. Enfin, une estimation adéquate des apports alimentaires et de l'activité physique permettrait d'ajuster les concentrations de caroténoïdes pour ces facteurs confondants. Cette étude pourrait permettre l'identification de potentiels biomarqueurs omiques de l'impact de l'alimentation, représentée par des profils alimentaires, sur la santé CM. L'impact de l'alimentation serait alors évalué avec des biomarqueurs omiques souvent modifiés en amont des facteurs de risque CM. Toujours en utilisant l'analyse de réseaux de corrélations pondérées, il serait très intéressant de croiser les résultats de plusieurs analyses dans différents tissus d'intérêt, dont le sang et le tissu adipeux, pour mieux comprendre l'influence globale de l'alimentation sur la santé CM. L'approche de réseaux de corrélations pondérées offre une multitude d'applications cliniques. Dans le domaine de la cancérologie, elle permet d'identifier des gènes qui ont le potentiel d'être utilisés comme biomarqueurs pour améliorer le diagnostic et le pronostic des patients (223). Cette approche pourrait permettre de mieux comprendre les mécanismes sous-jacents aux MCV en plus d'identifier des biomarqueurs pour stratifier le risque et améliorer les décisions thérapeutiques et le pronostic des patients (224). À titre d'exemple, une étude a révélé que la combinaison des niveaux d'expression de trois gènes identifiés avec cette approche serait un biomarqueur potentiel pour le diagnostic de l'infarctus du myocarde (225).

Les travaux dont cette thèse fait état ont également mis en lumière l'importance du défi que représente l'intégration de masses de données omiques en santé. Le système moléculaire interactif entre l'alimentation et la santé CM est extrêmement complexe et d'autres approches d'analyse sont nécessaires pour mieux le comprendre. À mon avis, l'utilisation de modèles issus de l'intelligence artificielle (apprentissage machine et apprentissage profond) est plus que prometteuse pour intégrer les données omiques dans le contexte de la santé afin d'augmenter la précision dans la prédiction du risque. Ces modèles utilisent la capacité des ordinateurs d'apprendre à réaliser une tâche à partir de données sans avoir été, au préalable, spécifiquement programmés pour cette tâche. Les modèles peuvent ainsi révéler des informations cliniquement pertinentes, mais enfouies dans une masse de données, lesquelles informations peuvent être utilisées pour la prise de décision ainsi que le traitement, le contrôle et la prévention des maladies (226). De plus, ces modèles ont l'avantage, comparativement aux méthodes traditionnelles, de pouvoir tenir compte des interactions complexes entre les variables en plus d'identifier des effets non linéaires. Des modèles issus de l'intelligence artificielle ont

déjà été développés pour intégrer des données omiques dans le contexte de la santé. Ainsi, le logiciel *Analysis Tool for Heritable and Environmental Network Associations* (ATHENA) combine des modèles d'apprentissage automatique pour analyser des variables catégorielles (p. ex. des SNPs) et quantitatives (p. ex. des niveaux d'expression génique) afin de générer des modèles prédisant des résultats catégoriques (p. ex. l'état de la maladie) ou quantitatifs (p. ex. les taux de cholestérol) (227). Ces méthodes représentent un outil prometteur pour intégrer non seulement plusieurs types de données omiques, mais aussi des facteurs environnementaux dans des questions de santé. Je crois que les domaines de la santé, des sciences omiques et de l'intelligence artificielle sont non seulement complémentaires, mais de plus en plus indissociables pour mieux comprendre et traiter les maladies. Ces méthodes nous permettraient d'élaborer des algorithmes de prédiction de la santé basés sur les profils omiques, les habitudes de vie et l'environnement des individus afin de mieux prévenir et traiter les maladies chroniques complexes.

Conclusion générale

À ce jour, aucune étude n'avait encore utilisé des approches de la biologie des systèmes pour analyser de façon intégrée les données omiques à l'échelle du génome dans l'impact de l'alimentation sur la santé CM au sein de familles en santé. Les travaux décrits dans cette thèse ont permis d'étudier l'impact des déterminants omiques, pris individuellement et de manière combinée, sur la relation entre l'alimentation et la santé CM à l'aide des approches de la génétique quantitative et des réseaux de corrélations pondérées. Ces travaux illustrent également une façon d'inclure un large éventail de données omiques dans la recherche en santé et en nutrition à l'aide d'approches de la biologie des systèmes. Une meilleure compréhension des voies métaboliques régulées par l'alimentation et de leurs impacts sur la santé permettrait d'aider à prédire et à prévenir plus efficacement le risque de maladies chroniques complexes et d'émettre des recommandations nutritionnelles qui tiennent compte des profils omiques, des habitudes de vie et de l'environnement des individus. En conclusion, le système moléculaire interactif qui relie l'alimentation et la santé CM est extrêmement complexe et des approches permettant de tenir compte de l'ensemble de ces interactions sont nécessaires.

Bibliographie

1. Canada. S. Tableau 102-0561 - Principales causes de décès, population totale, selon le groupe d'âge et le sexe, Canada. CANSIM (base de données sur les décès) [En ligne]. Ottawa: Statistique Canada2019 [Available from: http://www5.statcan.gc.ca/cansim/a26?lang=fra&id=1020561&retrLang=fra.

2. Adlspd C. Le fardeau économique de la maladie au Canada, 2005-2008. 2014.

3. Cannon CP. Cardiovascular disease and modifiable cardiometabolic risk factors. Clinical cornerstone. 2008;9(2):24-38; discussion 9-41.

4. Eckel RH, Jakicic JM, Ard JD, de Jesus JM, Houston Miller N, Hubbard VS, et al. 2013 AHA/ACC guideline on lifestyle management to reduce cardiovascular risk: a report of the American College of Cardiology/American Heart Association Task Force on Practice Guidelines. Circulation. 2014;129(25 Suppl 2):S76-99.

5. Anderson TJ, Gregoire J, Hegele RA, Couture P, Mancini GB, McPherson R, et al. 2012 update of the Canadian Cardiovascular Society guidelines for the diagnosis and treatment of dyslipidemia for the prevention of cardiovascular disease in the adult. The Canadian journal of cardiology. 2013;29(2):151-67.

6. Martinez-Gonzalez MA, de la Fuente-Arrillaga C, Lopez-Del-Burgo C, Vazquez-Ruiz Z, Benito S, Ruiz-Canela M. Low consumption of fruit and vegetables and risk of chronic disease: a review of the epidemiological evidence and temporal trends among Spanish graduates. Public health nutrition. 2011;14(12A):2309-15.

7. Scalbert A, Brennan L, Manach C, Andres-Lacueva C, Dragsted LO, Draper J, et al. The food metabolome: a window over dietary exposure. The American journal of clinical nutrition. 2014;99(6):1286-308.

8. Souverein OW, de Vries JH, Freese R, Watzl B, Bub A, Miller ER, et al. Prediction of fruit and vegetable intake from biomarkers using individual participant data of diet-controlled intervention studies. The British journal of nutrition. 2015;113(9):1396-409.

9. Sies H, Stahl W, Sundquist AR. Antioxidant functions of vitamins. Vitamins E and C, beta-carotene, and other carotenoids. Annals of the New York Academy of Sciences. 1992;669:7-20.

10. Panagiotou G, Nielsen J. Nutritional systems biology: definitions and approaches. Annual review of nutrition. 2009;29:329-39.

11. Waterland RA, Michels KB. Epigenetic epidemiology of the developmental origins hypothesis. Annual review of nutrition. 2007;27:363-88.

12. Barros SP, Offenbacher S. Epigenetics: connecting environment and genotype to phenotype and disease. Journal of dental research. 2009;88(5):400-8.

13. Bouchard-Mercier A, Rudkowska I, Lemieux S, Couture P, Vohl MC. The metabolic signature associated with the Western dietary pattern: a cross-sectional study. Nutrition journal. 2013;12:158.

14. Karczewski KJ, Snyder MP. Integrative omics for health and disease. Nature reviews Genetics. 2018;19(5):299-310.

15. Breitling R. What is systems biology? Frontiers in physiology. 2010;1:9.

16. Pinu FR, Beale DJ, Paten AM, Kouremenos K, Swarup S, Schirra HJ, et al. Systems Biology and Multi-Omics Integration: Viewpoints from the Metabolomics Research Community. Metabolites. 2019;9(4).

17. Dunn WB, Broadhurst DI, Atherton HJ, Goodacre R, Griffin JL. Systems level studies of mammalian metabolomes: the roles of mass spectrometry and nuclear magnetic resonance spectroscopy. Chemical Society reviews. 2011;40(1):387-426.

18. Sun YV, Hu YJ. Integrative Analysis of Multi-omics Data for Discovery and Functional Studies of Complex Human Diseases. Advances in genetics. 2016;93:147-90.

19. Hill WG. Understanding and using quantitative genetic variation. Philosophical transactions of the Royal Society of London Series B, Biological sciences. 2010;365(1537):73-85.

20. Collins FS, Green ED, Guttmacher AE, Guyer MS, Institute USNHGR. A vision for the future of genomics research. Nature. 2003;422(6934):835-47.

21. Zhang B, Horvath S. A general framework for weighted gene co-expression network analysis. Statistical applications in genetics and molecular biology. 2005;4:Article17.

22. OMS. Cardiovascular diseases 2015 [updated 2015.

23. Canada Adlspd. Rapport du système canadien de surveillance des maladies chroniques: Les maladies du coeur au Canada, 2018. Ottawa; 2018.

24. Lusis AJ. Atherosclerosis. Nature. 2000;407(6801):233-41.

25. Mortality GBD, Causes of Death C. Global, regional, and national life expectancy, all-cause mortality, and cause-specific mortality for 249 causes of death, 1980-2015: a systematic analysis for the Global Burden of Disease Study 2015. Lancet. 2016;388(10053):1459-544.

26. Expert Panel on Detection E, Treatment of High Blood Cholesterol in A. Executive Summary of The Third Report of The National Cholesterol Education Program (NCEP) Expert Panel on Detection, Evaluation, And Treatment of High Blood Cholesterol In Adults (Adult Treatment Panel III). JAMA. 2001;285(19):2486-97.

27. Despres JP, Cartier A, Cote M, Arsenault BJ. The concept of cardiometabolic risk: Bridging the fields of diabetology and cardiology. Annals of medicine. 2008;40(7):514-23.

28. Despres JP. Is visceral obesity the cause of the metabolic syndrome? Annals of medicine. 2006;38(1):52-63.

29. Despres JP, Lemieux I, Bergeron J, Pibarot P, Mathieu P, Larose E, et al. Abdominal obesity and the metabolic syndrome: contribution to global cardiometabolic risk. Arteriosclerosis, thrombosis, and vascular biology. 2008;28(6):1039-49.

30. Alberti KG, Eckel RH, Grundy SM, Zimmet PZ, Cleeman JI, Donato KA, et al. Harmonizing the metabolic syndrome: a joint interim statement of the International Diabetes Federation Task Force on Epidemiology and Prevention; National Heart, Lung, and Blood Institute; American Heart Association; World Heart Federation; International Atherosclerosis Society; and International Association for the Study of Obesity. Circulation. 2009;120(16):1640-5.

31. Federation ID. The IDF consensus worldwide definition of the metabolic syndrome. 2006 2006.

32. Galassi A, Reynolds K, He J. Metabolic syndrome and risk of cardiovascular disease: a meta-analysis. The American journal of medicine. 2006;119(10):812-9.

33. Cox RA, Garcia-Palmieri MR. Cholesterol, Triglycerides, and Associated Lipoproteins. In: rd, Walker HK, Hall WD, Hurst JW, editors. Clinical Methods: The History, Physical, and Laboratory Examinations. Boston1990.

34. Anderson TJ, Gregoire J, Pearson GJ, Barry AR, Couture P, Dawes M, et al. 2016 Canadian Cardiovascular Society Guidelines for the Management of Dyslipidemia for the Prevention of Cardiovascular Disease in the Adult. The Canadian journal of cardiology. 2016;32(11):1263-82.

35. Stamler J, Wentworth D, Neaton JD. Is relationship between serum cholesterol and risk of premature death from coronary heart disease continuous and graded? Findings in 356,222 primary screenees of the Multiple Risk Factor Intervention Trial (MRFIT). JAMA. 1986;256(20):2823-8.

36. Sharrett AR, Ballantyne CM, Coady SA, Heiss G, Sorlie PD, Catellier D, et al. Coronary heart disease prediction from lipoprotein cholesterol levels, triglycerides, lipoprotein(a), apolipoproteins A-I and B, and HDL density subfractions: The Atherosclerosis Risk in Communities (ARIC) Study. Circulation. 2001;104(10):1108-13.

37. Thanassoulis G, Williams K, Ye K, Brook R, Couture P, Lawler PR, et al. Relations of change in plasma levels of LDL-C, non-HDL-C and apoB with risk reduction from statin therapy: a meta-analysis of randomized trials. Journal of the American Heart Association. 2014;3(2):e000759.

38. Emerging Risk Factors C, Di Angelantonio E, Sarwar N, Perry P, Kaptoge S, Ray KK, et al. Major lipids, apolipoproteins, and risk of vascular disease. JAMA. 2009;302(18):1993-2000.

39. Sytkowski PA, Kannel WB, D'Agostino RB. Changes in risk factors and the decline in mortality from cardiovascular disease. The Framingham Heart Study. The New England journal of medicine. 1990;322(23):1635-41.

40. Ramasamy I. Recent advances in physiological lipoprotein metabolism. Clinical chemistry and laboratory medicine. 2014;52(12):1695-727.

41. Liu J, Zeng FF, Liu ZM, Zhang CX, Ling WH, Chen YM. Effects of blood triglycerides on cardiovascular and all-cause mortality: a systematic review and meta-analysis of 61 prospective studies. Lipids Health Dis. 2013;12:159.

42. Third Report of the National Cholesterol Education Program (NCEP) Expert Panel on Detection, Evaluation and Treatment of High Blood Cholesterol in Adults (2015 2015.

43. Grundy SM, Stone NJ, Bailey AL, Beam C, Birtcher KK, Blumenthal RS, et al. 2018 AHA/ACC/AACVPR/AAPA/ABC/ACPM/ADA/AGS/APhA/ASPC/NLA/PCNA Guideline on the Management of Blood Cholesterol: A Report of the American College of Cardiology/American Heart Association Task Force on Clinical Practice Guidelines. Journal of the American College of Cardiology. 2019;73(24):e285-e350.

44. Arnett DK, Blumenthal RS, Albert MA, Buroker AB, Goldberger ZD, Hahn EJ, et al. 2019 ACC/AHA Guideline on the Primary Prevention of Cardiovascular Disease: A Report of the American College of Cardiology/American Heart Association Task Force on Clinical Practice Guidelines. Circulation. 2019;140(11):e596-e646.

45. National Cholesterol Education Program Expert Panel on Detection E, Treatment of High Blood Cholesterol in A. Third Report of the National Cholesterol Education Program (NCEP) Expert Panel on Detection, Evaluation, and Treatment of High Blood Cholesterol in Adults (Adult Treatment Panel III) final report. Circulation. 2002;106(25):3143-421.

46. Yusuf S, Hawken S, Ounpuu S, Dans T, Avezum A, Lanas F, et al. Effect of potentially modifiable risk factors associated with myocardial infarction in 52 countries (the INTERHEART study): case-control study. Lancet. 2004;364(9438):937-52.

47. Chiavaroli L, Nishi SK, Khan TA, Braunstein CR, Glenn AJ, Mejia SB, et al. Portfolio Dietary Pattern and Cardiovascular Disease: A Systematic Review and Meta-analysis of Controlled Trials. Progress in cardiovascular diseases. 2018;61(1):43-53.

48. Jenkins DJ, Kendall CW, Marchie A, Faulkner DA, Wong JM, de Souza R, et al. Effects of a dietary portfolio of cholesterol-lowering foods vs lovastatin on serum lipids and C-reactive protein. JAMA. 2003;290(4):502-10.

49. Bingham SA. Limitations of the various methods for collecting dietary intake data. Annals of nutrition & metabolism. 1991;35(3):117-27.

50. Forrestal SG. Energy intake misreporting among children and adolescents: a literature review. Maternal & child nutrition. 2011;7(2):112-27.

51. Archer E, Pavela G, Lavie CJ. The Inadmissibility of What We Eat in America and NHANES Dietary Data in Nutrition and Obesity Research and the Scientific Formulation of National Dietary Guidelines. Mayo Clinic proceedings. 2015;90(7):911-26.

52. Westerterp KR, Goris AH. Validity of the assessment of dietary intake: problems of misreporting. Current opinion in clinical nutrition and metabolic care. 2002;5(5):489-93.

53. Labonte ME, Kirkpatrick SI, Bell RC, Boucher BA, Csizmadi I, Koushik A, et al. Dietary assessment is a critical element of health research - Perspective from the Partnership for Advancing Nutritional and Dietary Assessment in Canada. Applied physiology, nutrition, and metabolism = Physiologie appliquee, nutrition et metabolisme. 2016:1-4.

54. Heinzmann SS, Brown IJ, Chan Q, Bictash M, Dumas ME, Kochhar S, et al. Metabolic profiling strategy for discovery of nutritional biomarkers: proline betaine as a marker of citrus consumption. The American journal of clinical nutrition. 2010;92(2):436-43.

55. Baldrick FR, Woodside JV, Elborn JS, Young IS, McKinley MC. Biomarkers of fruit and vegetable intake in human intervention studies: a systematic review. Critical reviews in food science and nutrition. 2011;51(9):795-815.

56. Zamora-Ros R, Touillaud M, Rothwell JA, Romieu I, Scalbert A. Measuring exposure to the polyphenol metabolome in observational epidemiologic studies: current tools and applications and their limits. The American journal of clinical nutrition. 2014;100(1):11-26.

57. Holmes E, Loo RL, Stamler J, Bictash M, Yap IK, Chan Q, et al. Human metabolic phenotype diversity and its association with diet and blood pressure. Nature. 2008;453(7193):396-400.

58. Couillard C, Lemieux S, Vohl MC, Couture P, Lamarche B. Carotenoids as biomarkers of fruit and vegetable intake in men and women. The British journal of nutrition. 2016;116(7):1206-15.

59. Maiani G, Caston MJ, Catasta G, Toti E, Cambrodon IG, Bysted A, et al. Carotenoids: actual knowledge on food sources, intakes, stability and bioavailability and their protective role in humans. Molecular nutrition & food research. 2009;53 Suppl 2:S194-218.

60. Eroglu A, Harrison EH. Carotenoid metabolism in mammals, including man: formation, occurrence, and function of apocarotenoids. Journal of lipid research. 2013;54(7):1719-30.

61. Castenmiller JJ, West CE. Bioavailability and bioconversion of carotenoids. Annual review of nutrition. 1998;18:19-38.

62. Hedren E, Diaz V, Svanberg U. Estimation of carotenoid accessibility from carrots determined by an in vitro digestion method. European journal of clinical nutrition. 2002;56(5):425-30.

63. Borel P, Grolier P, Armand M, Partier A, Lafont H, Lairon D, et al. Carotenoids in biological emulsions: solubility, surface-to-core distribution, and release from lipid droplets. Journal of lipid research. 1996;37(2):250-61.

64. Reboul E. Absorption of vitamin A and carotenoids by the enterocyte: focus on transport proteins. Nutrients. 2013;5(9):3563-81.

65. Bendich A, Olson JA. Biological actions of carotenoids. FASEB journal : official publication of the Federation of American Societies for Experimental Biology. 1989;3(8):1927-32.

66. Scott KJ, Thurnham DI, Hart DJ, Bingham SA, Day K. The correlation between the intake of lutein, lycopene and beta-carotene from vegetables and fruits, and blood plasma concentrations in a group of women aged 50-65 years in the UK. The British journal of nutrition. 1996;75(3):409-18.

67. Wang Y, Chung SJ, McCullough ML, Song WO, Fernandez ML, Koo SI, et al. Dietary carotenoids are associated with cardiovascular disease risk biomarkers mediated by serum carotenoid concentrations. The Journal of nutrition. 2014;144(7):1067-74.

68. Allore T, Lemieux S, Vohl MC, Couture P, Lamarche B, Couillard C. Correlates of the difference in plasma carotenoid concentrations between men and women. The British journal of nutrition. 2018:1-10.

69. Clevidence BA, Bieri JG. Association of carotenoids with human plasma lipoproteins. Methods in enzymology. 1993;214:33-46.

70. Rao AV, Rao LG. Carotenoids and human health. Pharmacological research. 2007;55(3):207-16.

71. Paiva SA, Russell RM. Beta-carotene and other carotenoids as antioxidants. Journal of the American College of Nutrition. 1999;18(5):426-33.

72. Bertram JS. Carotenoids and gene regulation. Nutrition reviews. 1999;57(6):182-91.

73. Pico C, Serra F, Rodriguez AM, Keijer J, Palou A. Biomarkers of Nutrition and Health: New Tools for New Approaches. Nutrients. 2019;11(5).

74. Strachan T, Read AP. Human Molecular Genetics. Publisher BS, editor. New York1996.

75. Alberts B, Johnson AD, Lewis J, Morgan D, Raff M, Robert K, et al. Molecular Biology of the Cell. Sixth edition ed: W.W. Norton and Company; 2015 November 18, 2014.

76. Ezkurdia I, Juan D, Rodriguez JM, Frankish A, Diekhans M, Harrow J, et al. Multiple evidence strands suggest that there may be as few as 19,000 human protein-coding genes. Human molecular genetics. 2014;23(22):5866-78.

77. Clancy S. RNA splicing: introns, exons and spliceosome. Nature Education. 2008;1(1):31.

78. Harrow J, Nagy A, Reymond A, Alioto T, Patthy L, Antonarakis SE, et al. Identifying protein-coding genes in genomic sequences. Genome biology. 2009;10(1):201.

79. Rolfes SR, Pinna K, Whitney E. Understanding Normal and Clinical Nutrition. 10th edition ed: Cengage Learning; 2015 January 31, 2014.

80. D'Onofrio DJ, Abel DL. Redundancy of the genetic code enables translational pausing. Frontiers in genetics. 2014;5:140.

81. Genomes Project C, Auton A, Brooks LD, Durbin RM, Garrison EP, Kang HM, et al. A global reference for human genetic variation. Nature. 2015;526(7571):68-74.

82. International Human Genome Sequencing C. Finishing the euchromatic sequence of the human genome. Nature. 2004;431(7011):931-45.

83. Manolio TA. Genomewide association studies and assessment of the risk of disease. The New England journal of medicine. 2010;363(2):166-76.

84. Liu DJ, Peloso GM, Yu H, Butterworth AS, Wang X, Mahajan A, et al. Exome-wide association study of plasma lipids in >300,000 individuals. Nature genetics. 2017;49(12):1758-66.

85. Willer CJ, Schmidt EM, Sengupta S, Peloso GM, Gustafsson S, Kanoni S, et al. Discovery and refinement of loci associated with lipid levels. Nature genetics. 2013;45(11):1274-83.

86. Davignon J, Gregg RE, Sing CF. Apolipoprotein E polymorphism and atherosclerosis. Arteriosclerosis. 1988;8(1):1-21.

87. Song Y, Stampfer MJ, Liu S. Meta-analysis: apolipoprotein E genotypes and risk for coronary heart disease. Annals of internal medicine. 2004;141(2):137-47.

88. Bennet AM, Di Angelantonio E, Ye Z, Wensley F, Dahlin A, Ahlbom A, et al. Association of apolipoprotein E genotypes with lipid levels and coronary risk. JAMA. 2007;298(11):1300-11.

89. Kessler T, Vilne B, Schunkert H. The impact of genome-wide association studies on the pathophysiology and therapy of cardiovascular disease. EMBO molecular medicine. 2016;8(7):688-701.

90. McPherson R, Tybjaerg-Hansen A. Genetics of Coronary Artery Disease. Circulation research. 2016;118(4):564-78.

91. Ordovas JM, Corella D. Nutritional genomics. AnnuRevGenomics HumGenet. 2004;5:71-118.

92. Minihane AM. The genetic contribution to disease risk and variability in response to diet: where is the hidden heritability? ProcNutrSoc. 2013;72(1):40-7.

93. BICKEL H, GERRARD J, HICKMANS EM. Influence of phenylalanine intake on phenylketonuria. Lancet. 1953;265(6790):812-3.

94. Do R, Xie C, Zhang X, Mannisto S, Harald K, Islam S, et al. The effect of chromosome 9p21 variants on cardiovascular disease may be modified by dietary intake: evidence from a case/control and a prospective study. PLoS Med. 2011;8(10):e1001106.

95. Caslake MJ, Miles EA, Kofler BM, Lietz G, Curtis P, Armah CK, et al. Effect of sex and genotype on cardiovascular biomarker response to fish oils: the FINGEN Study. AmJClinNutr. 2008;88(3):618-29.

96. Rudkowska I, Paradis AM, Thifault E, Julien P, Barbier O, Couture P, et al. Differences in metabolomic and transcriptomic profiles between responders and non-responders to an n-3 polyunsaturated fatty acids (PUFAs) supplementation. Genes Nutr. 2013;8(4):411-23.

97. Vallee Marcotte B, Guenard F, Lemieux S, Couture P, Rudkowska I, Calder PC, et al. Fine mapping of genome-wide association study signals to identify genetic markers of the plasma triglyceride response to an omega-3 fatty acid supplementation. The American journal of clinical nutrition. 2019;109(1):176-85.

98. D'Adamo CR, D'Urso A, Ryan KA, Yerges-Armstrong LM, Semba RD, Steinle NI, et al. A Common Variant in the SETD7 Gene Predicts Serum Lycopene Concentrations. Nutrients. 2016;8(2):82.

99. Beydoun MA, Nalls MA, Canas JA, Evans MK, Zonderman AB. Gene polymorphisms and gene scores linked to low serum carotenoid status and their associations with metabolic disturbance and depressive symptoms in African-American adults. The British journal of nutrition. 2014;112(6):992-1003.

100. Ferrucci L, Perry JR, Matteini A, Perola M, Tanaka T, Silander K, et al. Common variation in the beta-carotene 15,15'-monooxygenase 1 gene affects circulating levels of carotenoids: a genome-wide association study. American journal of human genetics. 2009;84(2):123-33.

101. Zubair N, Kooperberg C, Liu J, Di C, Peters U, Neuhouser ML. Genetic variation predicts serum lycopene concentrations in a multiethnic population of postmenopausal women. The Journal of nutrition. 2015;145(2):187-92.

102. Borel P, de Edelenyi FS, Vincent-Baudry S, Malezet-Desmoulin C, Margotat A, Lyan B, et al. Genetic variants in BCMO1 and CD36 are associated with plasma lutein concentrations and macular pigment optical density in humans. Annals of medicine. 2011;43(1):47-59.

103. Bohn T, Desmarchelier C, Dragsted LO, Nielsen CS, Stahl W, Ruhl R, et al. Host-related factors explaining interindividual variability of carotenoid bioavailability and tissue concentrations in humans. Molecular nutrition & food research. 2017;61(6).

104. Goodman DS, Huang HS, Shiratori T. Mechanism of the biosynthesis of vitamin A from beta-carotene. The Journal of biological chemistry. 1966;241(9):1929-32.

105. Bird A. Perceptions of epigenetics. Nature. 2007;447(7143):396-8.

106. Milagro FI, Mansego ML, De Miguel C, Martinez JA. Dietary factors, epigenetic modifications and obesity outcomes: progresses and perspectives. Molecular aspects of medicine. 2013;34(4):782-812.

107. Tammen SA, Friso S, Choi SW. Epigenetics: the link between nature and nurture. Molecular aspects of medicine. 2013;34(4):753-64.

108. Sharma S, Kelly TK, Jones PA. Epigenetics in cancer. Carcinogenesis. 2010;31(1):27-36.

109. Daxinger L, Whitelaw E. Understanding transgenerational epigenetic inheritance via the gametes in mammals. Nature reviews Genetics. 2012;13(3):153-62.

110. Sandoval J, Heyn H, Moran S, Serra-Musach J, Pujana MA, Bibikova M, et al. Validation of a DNA methylation microarray for 450,000 CpG sites in the human genome. Epigenetics. 2011;6(6):692-702.

111. Gannon JR, Emery BR, Jenkins TG, Carrell DT. The sperm epigenome: implications for the embryo. Adv Exp Med Biol. 2014;791:53-66.

112. Feng S, Cokus SJ, Zhang X, Chen PY, Bostick M, Goll MG, et al. Conservation and divergence of methylation patterning in plants and animals. Proceedings of the National Academy of Sciences of the United States of America. 2010;107(19):8689-94.

113. Kader F, Ghai M. DNA methylation-based variation between human populations. Mol Genet Genomics. 2017;292(1):5-35.

114. Zhong J, Agha G, Baccarelli AA. The Role of DNA Methylation in Cardiovascular Risk and Disease: Methodological Aspects, Study Design, and Data Analysis for Epidemiological Studies. Circulation research. 2016;118(1):119-31.

115. Movassagh M, Choy MK, Goddard M, Bennett MR, Down TA, Foo RS. Differential DNA methylation correlates with differential expression of angiogenic factors in human heart failure. PloS one. 2010;5(1):e8564.

116. Baccarelli A, Wright R, Bollati V, Litonjua A, Zanobetti A, Tarantini L, et al. Ischemic heart disease and stroke in relation to blood DNA methylation. Epidemiology. 2010;21(6):819-28.

117. Hedman AK, Mendelson MM, Marioni RE, Gustafsson S, Joehanes R, Irvin MR, et al. Epigenetic Patterns in Blood Associated With Lipid Traits Predict Incident Coronary Heart Disease Events and Are Enriched for Results From Genome-Wide Association Studies. Circulation Cardiovascular genetics. 2017;10(1).

118. Dekkers KF, van Iterson M, Slieker RC, Moed MH, Bonder MJ, van Galen M, et al. Blood lipids influence DNA methylation in circulating cells. Genome biology. 2016;17(1):138.

119. Mittelstrass K, Waldenberger M. DNA methylation in human lipid metabolism and related diseases. Current opinion in lipidology. 2018;29(2):116-24.

120. Ali O, Cerjak D, Kent JW, Jr., James R, Blangero J, Carless MA, et al. Methylation of SOCS3 is inversely associated with metabolic syndrome in an epigenome-wide association study of obesity. Epigenetics. 2016;11(9):699-707.

121. Choi SW, Friso S. Epigenetics: A New Bridge between Nutrition and Health. Advances in nutrition. 2010;1(1):8-16.

122. Roseboom TJ, van der Meulen JH, Osmond C, Barker DJ, Ravelli AC, Schroeder-Tanka JM, et al. Coronary heart disease after prenatal exposure to the Dutch famine, 1944-45. Heart. 2000;84(6):595-8.

123. Ma J, Rebholz CM, Braun KVE, Reynolds LM, Aslibekyan S, Xia R, et al. Whole Blood DNA Methylation Signatures of Diet Are Associated With Cardiovascular Disease Risk Factors and All-Cause Mortality. Circ Genom Precis Med. 2020;13(4):e002766.

124. Hoile SP, Clarke-Harris R, Huang RC, Calder PC, Mori TA, Beilin LJ, et al. Supplementation with N-3 long-chain polyunsaturated fatty acids or olive oil in men and women with renal disease induces differential

changes in the DNA methylation of FADS2 and ELOVL5 in peripheral blood mononuclear cells. PloS one. 2014;9(10):e109896.

125. Tremblay BL, Guenard F, Rudkowska I, Lemieux S, Couture P, Vohl MC. Epigenetic changes in blood leukocytes following an omega-3 fatty acid supplementation. Clinical epigenetics. 2017;9:43.

126. Arpon A, Riezu-Boj JI, Milagro FI, Razquin C, Martinez-Gonzalez MA, Corella D, et al. Adherence to Mediterranean diet is associated with methylation changes in inflammation-related genes in peripheral blood cells. Journal of physiology and biochemistry. 2017.

127. Zhang FF, Morabia A, Carroll J, Gonzalez K, Fulda K, Kaur M, et al. Dietary patterns are associated with levels of global genomic DNA methylation in a cancer-free population. The Journal of nutrition. 2011;141(6):1165-71.

128. Liu AG, Erdman JW, Jr. Lycopene and apo-10'-lycopenal do not alter DNA methylation of GSTP1 in LNCaP cells. Biochemical and biophysical research communications. 2011;412(3):479-82.

129. King-Batoon A, Leszczynska JM, Klein CB. Modulation of gene methylation by genistein or lycopene in breast cancer cells. Environmental and molecular mutagenesis. 2008;49(1):36-45.

130. Bollati V, Favero C, Albetti B, Tarantini L, Moroni A, Byun HM, et al. Nutrients intake is associated with DNA methylation of candidate inflammatory genes in a population of obese subjects. Nutrients. 2014;6(10):4625-39.

131. Fu LJ, Ding YB, Wu LX, Wen CJ, Qu Q, Zhang X, et al. The Effects of Lycopene on the Methylation of the GSTP1 Promoter and Global Methylation in Prostatic Cancer Cell Lines PC3 and LNCaP. International journal of endocrinology. 2014;2014:620165.

132. Pedrotty DM, Morley MP, Cappola TP. Transcriptomic biomarkers of cardiovascular disease. Progress in cardiovascular diseases. 2012;55(1):64-9.

133. Lappalainen T, Sammeth M, Friedlander MR, t Hoen PA, Monlong J, Rivas MA, et al. Transcriptome and genome sequencing uncovers functional variation in humans. Nature. 2013;501(7468):506-11.

134. Powell JE, Henders AK, McRae AF, Caracella A, Smith S, Wright MJ, et al. The Brisbane Systems Genetics Study: genetical genomics meets complex trait genetics. PloS one. 2012;7(4):e35430.

135. Consortium GT. The Genotype-Tissue Expression (GTEx) project. Nature genetics. 2013;45(6):580-5.

136. Schadt EE, Molony C, Chudin E, Hao K, Yang X, Lum PY, et al. Mapping the genetic architecture of gene expression in human liver. PLoS Biol. 2008;6(5):e107.

137. Ross SA. Evidence for the relationship between diet and cancer. Experimental oncology. 2010;32(3):137-42.

138. Schnabel RB, Baccarelli A, Lin H, Ellinor PT, Benjamin EJ. Next steps in cardiovascular disease genomic research--sequencing, epigenetics, and transcriptomics. Clinical chemistry. 2012;58(1):113-26.

139. Joehanes R, Ying S, Huan T, Johnson AD, Raghavachari N, Wang R, et al. Gene expression signatures of coronary heart disease. Arteriosclerosis, thrombosis, and vascular biology. 2013;33(6):1418-26.

140. Sulkava M, Raitoharju E, Mennander A, Levula M, Seppala I, Lyytikainen LP, et al. Differentially expressed genes and canonical pathways in the ascending thoracic aortic aneurysm - The Tampere Vascular Study. Scientific reports. 2017;7(1):12127.

141. Kontou P, Pavlopoulou A, Braliou G, Bogiatzi S, Dimou N, Bangalore S, et al. Identification of gene expression profiles in myocardial infarction: a systematic review and meta-analysis. BMC medical genomics. 2018;11(1):109.

142. Raulerson CK, Ko A, Kidd JC, Currin KW, Brotman SM, Cannon ME, et al. Adipose Tissue Gene Expression Associations Reveal Hundreds of Candidate Genes for Cardiometabolic Traits. American journal of human genetics. 2019;105(4):773-87.

143. Ma J, Dempsey AA, Stamatiou D, Marshall KW, Liew CC. Identifying leukocyte gene expression patterns associated with plasma lipid levels in human subjects. Atherosclerosis. 2007;191(1):63-72.

144. Contreras AV, Torres N, Tovar AR. PPAR-alpha as a key nutritional and environmental sensor for metabolic adaptation. Advances in nutrition. 2013;4(4):439-52.

145. Sharoni Y, Danilenko M, Dubi N, Ben-Dor A, Levy J. Carotenoids and transcription. Archives of biochemistry and biophysics. 2004;430(1):89-96.

146. Konstantinidou V, Covas MI, Sola R, Fito M. Up-to date knowledge on the in vivo transcriptomic effect of the Mediterranean diet in humans. Molecular nutrition & food research. 2013;57(5):772-83.

147. Castaner O, Corella D, Covas MI, Sorli JV, Subirana I, Flores-Mateo G, et al. In vivo transcriptomic profile after a Mediterranean diet in high-cardiovascular risk patients: a randomized controlled trial. The American journal of clinical nutrition. 2013;98(3):845-53.

148. van Dijk SJ, Feskens EJ, Bos MB, Hoelen DW, Heijligenberg R, Bromhaar MG, et al. A saturated fatty acid-rich diet induces an obesity-linked proinflammatory gene expression profile in adipose tissue of subjects at risk of metabolic syndrome. The American journal of clinical nutrition. 2009;90(6):1656-64.

149. Bouchard-Mercier A, Paradis AM, Rudkowska I, Lemieux S, Couture P, Vohl MC. Associations between dietary patterns and gene expression profiles of healthy men and women: a cross-sectional study. Nutrition journal. 2013;12:24.

150. Zheng Y, Yu B, Alexander D, Steffen LM, Boerwinkle E. Human metabolome associates with dietary intake habits among African Americans in the atherosclerosis risk in communities study. American journal of epidemiology. 2014;179(12):1424-33.

151. Lam SM, Wang Y, Li B, Du J, Shui G. Metabolomics through the lens of precision cardiovascular medicine. J Genet Genomics. 2017;44(3):127-38.

152. Baig F, Pechlaner R, Mayr M. Caveats of Untargeted Metabolomics for Biomarker Discovery. Journal of the American College of Cardiology. 2016;68(12):1294-6.

153. Wishart DS, Knox C, Guo AC, Eisner R, Young N, Gautam B, et al. HMDB: a knowledgebase for the human metabolome. Nucleic acids research. 2009;37(Database issue):D603-10.

154. Lam SM, Shui G. Lipidomics as a principal tool for advancing biomedical research. J Genet Genomics. 2013;40(8):375-90.

155. Floegel A, Stefan N, Yu Z, Muhlenbruch K, Drogan D, Joost HG, et al. Identification of serum metabolites associated with risk of type 2 diabetes using a targeted metabolomic approach. Diabetes. 2013;62(2):639-48.

156. Tomas L, Edsfeldt A, Mollet IG, Perisic Matic L, Prehn C, Adamski J, et al. Altered metabolism distinguishes high-risk from stable carotid atherosclerotic plaques. European heart journal. 2018.

157. Jourdan C, Petersen AK, Gieger C, Doring A, Illig T, Wang-Sattler R, et al. Body fat free mass is associated with the serum metabolite profile in a population-based study. PloS one. 2012;7(6):e40009.

158. Fan Y, Li Y, Chen Y, Zhao YJ, Liu LW, Li J, et al. Comprehensive Metabolomic Characterization of Coronary Artery Diseases. Journal of the American College of Cardiology. 2016;68(12):1281-93.

159. de Couto G, Ouzounian M, Liu PP. Early detection of myocardial dysfunction and heart failure. Nature reviews Cardiology. 2010;7(6):334-44.

160. Ntzouvani A, Nomikos T, Panagiotakos D, Fragopoulou E, Pitsavos C, McCann A, et al. Amino acid profile and metabolic syndrome in a male Mediterranean population: A cross-sectional study. Nutrition, metabolism, and cardiovascular diseases : NMCD. 2017;27(11):1021-30.

161. Cheng S, Rhee EP, Larson MG, Lewis GD, McCabe EL, Shen D, et al. Metabolite profiling identifies pathways associated with metabolic risk in humans. Circulation. 2012;125(18):2222-31.

162. Floegel A, Wientzek A, Bachlechner U, Jacobs S, Drogan D, Prehn C, et al. Linking diet, physical activity, cardiorespiratory fitness and obesity to serum metabolite networks: findings from a population-based study. International journal of obesity. 2014;38(11):1388-96.

163. Schooneman MG, Vaz FM, Houten SM, Soeters MR. Acylcarnitines: reflecting or inflicting insulin resistance? Diabetes. 2013;62(1):1-8.

164. Newgard CB, An J, Bain JR, Muehlbauer MJ, Stevens RD, Lien LF, et al. A branched-chain amino acid-related metabolic signature that differentiates obese and lean humans and contributes to insulin resistance. Cell metabolism. 2009;9(4):311-26.

165. McCormack SE, Shaham O, McCarthy MA, Deik AA, Wang TJ, Gerszten RE, et al. Circulating branched-chain amino acid concentrations are associated with obesity and future insulin resistance in children and adolescents. Pediatric obesity. 2013;8(1):52-61.

166. Tai ES, Tan ML, Stevens RD, Low YL, Muehlbauer MJ, Goh DL, et al. Insulin resistance is associated with a metabolic profile of altered protein metabolism in Chinese and Asian-Indian men. Diabetologia. 2010;53(4):757-67.

167. Lu J, Xie G, Jia W, Jia W. Insulin resistance and the metabolism of branched-chain amino acids. Frontiers of medicine. 2013;7(1):53-9.

168. Kien CL, Bunn JY, Stevens R, Bain J, Ikayeva O, Crain K, et al. Dietary intake of palmitate and oleate has broad impact on systemic and tissue lipid profiles in humans. The American journal of clinical nutrition. 2014;99(3):436-45.

169. Cheung W, Keski-Rahkonen P, Assi N, Ferrari P, Freisling H, Rinaldi S, et al. A metabolomic study of biomarkers of meat and fish intake. The American journal of clinical nutrition. 2017;105(3):600-8.

170. Welter D, MacArthur J, Morales J, Burdett T, Hall P, Junkins H, et al. The NHGRI GWAS Catalog, a curated resource of SNP-trait associations. Nucleic acids research. 2014;42(Database issue):D1001-6.

171. Pennisi E. The Biology of Genomes. Disease risk links to gene regulation. Science. 2011;332(6033):1031.

172. Lemire M, Zaidi SH, Ban M, Ge B, Aissi D, Germain M, et al. Long-range epigenetic regulation is conferred by genetic variation located at thousands of independent loci. Nature communications. 2015;6:6326.

173. Grundberg E, Meduri E, Sandling JK, Hedman AK, Keildson S, Buil A, et al. Global analysis of DNA methylation variation in adipose tissue from twins reveals links to disease-associated variants in distal regulatory elements. American journal of human genetics. 2013;93(5):876-90.

174. Kettunen J, Tukiainen T, Sarin AP, Ortega-Alonso A, Tikkanen E, Lyytikainen LP, et al. Genome-wide association study identifies multiple loci influencing human serum metabolite levels. Nature genetics. 2012;44(3):269-76.

175. Hasin Y, Seldin M, Lusis A. Multi-omics approaches to disease. Genome biology. 2017;18(1):83.

176. Cancer Genome Atlas N. Comprehensive molecular portraits of human breast tumours. Nature. 2012;490(7418):61-70.

177. Kim SY, Kim TR, Jeong HH, Sohn KA. Integrative pathway-based survival prediction utilizing the interaction between gene expression and DNA methylation in breast cancer. BMC medical genomics. 2018;11(Suppl 3):68.

178. Liu W, Bai X, Liu Y, Wang W, Han J, Wang Q, et al. Topologically inferring pathway activity toward precise cancer classification via integrating genomic and metabolomic data: prostate cancer as a case. Scientific reports. 2015;5:13192.

179. Liu W, Li C, Xu Y, Yang H, Yao Q, Han J, et al. Topologically inferring risk-active pathways toward precise cancer classification by directed random walk. Bioinformatics. 2013;29(17):2169-77.

180. Wirapati P, Sotiriou C, Kunkel S, Farmer P, Pradervand S, Haibe-Kains B, et al. Meta-analysis of gene expression profiles in breast cancer: toward a unified understanding of breast cancer subtyping and prognosis signatures. Breast cancer research : BCR. 2008;10(4):R65.

181. Joshi A, Rienks M, Theofilatos K, Mayr M. Systems biology in cardiovascular disease: a multiomics approach. Nature reviews Cardiology. 2020.

182. Khera AV, Emdin CA, Drake I, Natarajan P, Bick AG, Cook NR, et al. Genetic Risk, Adherence to a Healthy Lifestyle, and Coronary Disease. The New England journal of medicine. 2016;375(24):2349-58.

183. Amos W, Driscoll E, Hoffman JI. Candidate genes versus genome-wide associations: which are better for detecting genetic susceptibility to infectious disease? Proc Biol Sci. 2011;278(1709):1183-8.

184. Rice TK. Familial resemblance and heritability. Advances in genetics. 2008;60:35-49.

185. Rao DC. Genetic dissection of complex traits: an overview. Advances in genetics. 2001;42:13-34.

186. RA F. The correlation between relatives on the supposition of Mendelian inheritance. Trans R Soc Edin. 1918;52:399-433.

187. Cockerham CC. An Extension of the Concept of Partitioning Hereditary Variance for Analysis of Covariances among Relatives When Epistasis Is Present. Genetics. 1954;39(6):859-82.

188. Kempthorne O. The correlation between relatives in a random mating population. Proceedings of the Royal Society of London Series B, Biological sciences. 1954;143(910):102-13.

189. Almasy L, Blangero J. Variance component methods for analysis of complex phenotypes. Cold Spring Harb Protoc. 2010;2010(5):pdb top77.

190. Wray N, Visscher P. Estimating trait heritability. Nature Education. 2008;1(1):29.

191. Tremblay BL, Guenard F, Lamarche B, Perusse L, Vohl MC. Genetic and Common Environmental Contributions to Familial Resemblances in Plasma Carotenoid Concentrations in Healthy Families. Nutrients. 2018;10(8).

192. Mayhew AJ, Meyre D. Assessing the Heritability of Complex Traits in Humans: Methodological Challenges and Opportunities. Curr Genomics. 2017;18(4):332-40.

193. Heath AC, Kessler RC, Neale MC, Hewitt JK, Eaves LJ, Kendler KS. Testing hypotheses about direction of causation using cross-sectional family data. Behavior genetics. 1993;23(1):29-50.

194. Yang F, Tang Z, Deng H. Bivariate association analysis for quantitative traits using generalized estimation equation. J Genet Genomics. 2009;36(12):733-43.

195. Vinkhuyzen AA, Wray NR, Yang J, Goddard ME, Visscher PM. Estimation and partition of heritability in human populations using whole-genome analysis methods. Annual review of genetics. 2013;47:75-95.

196. Zdravkovic S, Wienke A, Pedersen NL, Marenberg ME, Yashin AI, De Faire U. Heritability of death from coronary heart disease: a 36-year follow-up of 20 966 Swedish twins. J Intern Med. 2002;252(3):247-54.

197. Heller DA, de Faire U, Pedersen NL, Dahlen G, McClearn GE. Genetic and environmental influences on serum lipid levels in twins. The New England journal of medicine. 1993;328(16):1150-6.

198. Perusse L, Rice T, Despres JP, Bergeron J, Province MA, Gagnon J, et al. Familial resemblance of plasma lipids, lipoproteins and postheparin lipoprotein and hepatic lipases in the HERITAGE Family Study. Arteriosclerosis, thrombosis, and vascular biology. 1997;17(11):3263-9.

199. Shoemaker R, Deng J, Wang W, Zhang K. Allele-specific methylation is prevalent and is contributed by CpG-SNPs in the human genome. Genome research. 2010;20(7):883-9.

200. McRae AF, Powell JE, Henders AK, Bowdler L, Hemani G, Shah S, et al. Contribution of genetic variation to transgenerational inheritance of DNA methylation. Genome biology. 2014;15(5):R73.

201. Bell JT, Tsai PC, Yang TP, Pidsley R, Nisbet J, Glass D, et al. Epigenome-wide scans identify differentially methylated regions for age and age-related phenotypes in a healthy ageing population. PLoS genetics. 2012;8(4):e1002629.

202. Powell JE, Henders AK, McRae AF, Kim J, Hemani G, Martin NG, et al. Congruence of additive and non-additive effects on gene expression estimated from pedigree and SNP data. PLoS genetics. 2013;9(5):e1003502.

203. Suhre K, Gieger C. Genetic variation in metabolic phenotypes: study designs and applications. Nature reviews Genetics. 2012;13(11):759-69.

204. Draisma HH, Beekman M, Pool R, van Ommen GJ, Adamski J, Prehn C, et al. Familial resemblance for serum metabolite concentrations. Twin research and human genetics : the official journal of the International Society for Twin Studies. 2013;16(5):948-61.

205. Langfelder P, Horvath S. WGCNA: an R package for weighted correlation network analysis. BMC bioinformatics. 2008;9:559.

206. Tremblay BL, Guenard F, Lamarche B, Perusse L, Vohl MC. Network Analysis of the Potential Role of DNA Methylation in the Relationship between Plasma Carotenoids and Lipid Profile. Nutrients. 2019;11(6):1265.

207. Langfelder P, Horvath S. Eigengene networks for studying the relationships between co-expression modules. BMC systems biology. 2007;1:54.

208. Tremblay BL, Guenard F, Lamarche B, Perusse L, Vohl MC. Integrative Network Analysis of Multi-Omics Data in the Link between Plasma Carotenoid Concentrations and Lipid Profile. Lifestyle Genom. 2020;13(1):11-9.

209. Qiu J, Du Z, Wang Y, Zhou Y, Zhang Y, Xie Y, et al. Weighted gene co-expression network analysis reveals modules and hub genes associated with the development of breast cancer. Medicine. 2019;98(6):e14345.

210. Horvath S, Zhang B, Carlson M, Lu KV, Zhu S, Felciano RM, et al. Analysis of oncogenic signaling networks in glioblastoma identifies ASPM as a molecular target. Proceedings of the National Academy of Sciences of the United States of America. 2006;103(46):17402-7.

211. Wang W, Jiang W, Hou L, Duan H, Wu Y, Xu C, et al. Weighted gene co-expression network analysis of expression data of monozygotic twins identifies specific modules and hub genes related to BMI. BMC genomics. 2017;18(1):872.

212. Ding J, Reynolds LM, Zeller T, Muller C, Lohman K, Nicklas BJ, et al. Alterations of a Cellular Cholesterol Metabolism Network Are a Molecular Feature of Obesity-Related Type 2 Diabetes and Cardiovascular Disease. Diabetes. 2015;64(10):3464-74.

213. Liu J, Jing L, Tu X. Weighted gene co-expression network analysis identifies specific modules and hub genes related to coronary artery disease. BMC cardiovascular disorders. 2016;16:54.

214. Miao L, Yin RX, Pan SL, Yang S, Yang DZ, Lin WX. Weighted Gene Co-Expression Network Analysis Identifies Specific Modules and Hub Genes Related to Hyperlipidemia. Cellular physiology and biochemistry : international journal of experimental cellular physiology, biochemistry, and pharmacology. 2018;48(3):1151-63.

215. Christensen JJ, Ulven SM, Thoresen M, Westerman K, Holven KB, Andersen LF. Associations between dietary patterns and gene expression pattern in peripheral blood mononuclear cells: A cross-sectional study. Nutrition, metabolism, and cardiovascular diseases : NMCD. 2020.

216. Shakhbazov K, Powell JE, Hemani G, Henders AK, Martin NG, Visscher PM, et al. Shared genetic control of expression and methylation in peripheral blood. BMC genomics. 2016;17:278.

217. Jaffe AE, Irizarry RA. Accounting for cellular heterogeneity is critical in epigenome-wide association studies. Genome biology. 2014;15(2):R31.

218. Abbas AR, Wolslegel K, Seshasayee D, Modrusan Z, Clark HF. Deconvolution of blood microarray data identifies cellular activation patterns in systemic lupus erythematosus. PloS one. 2009;4(7):e6098.

219. Rijsdijk FV, Sham PC. Analytic approaches to twin data using structural equation models. Briefings in bioinformatics. 2002;3(2):119-33.

220. Tremblay BL, Guenard F, Lamarche B, Perusse L, Vohl MC. Familial resemblances in human whole blood transcriptome. BMC genomics. 2018;19(1):300.

221. Carroll RJ, Midthune D, Subar AF, Shumakovich M, Freedman LS, Thompson FE, et al. Taking advantage of the strengths of 2 different dietary assessment instruments to improve intake estimates for nutritional epidemiology. American journal of epidemiology. 2012;175(4):340-7.

222. Chen R, Mias GI, Li-Pook-Than J, Jiang L, Lam HY, Chen R, et al. Personal omics profiling reveals dynamic molecular and medical phenotypes. Cell. 2012;148(6):1293-307.

223. Wu J, Liu XJ, Hu JN, Liao XH, Lin FF. Transcriptomics and Prognosis Analysis to Identify Critical Biomarkers in Invasive Breast Carcinoma. Technol Cancer Res Treat. 2020;19:1533033820957011.

224. Niu X, Zhang J, Zhang L, Hou Y, Pu S, Chu A, et al. Weighted Gene Co-Expression Network Analysis Identifies Critical Genes in the Development of Heart Failure After Acute Myocardial Infarction. Frontiers in genetics. 2019;10:1214.

225. Liu Z, Ma C, Gu J, Yu M. Potential biomarkers of acute myocardial infarction based on weighted gene co-expression network analysis. Biomed Eng Online. 2019;18(1):9.

226. Tobore I, Li J, Yuhang L, Al-Handarish Y, Kandwal A, Nie Z, et al. Deep Learning Intervention for Health Care Challenges: Some Biomedical Domain Considerations. JMIR Mhealth Uhealth. 2019;7(8):e11966.

227. Holzinger ER, Dudek SM, Frase AT, Pendergrass SA, Ritchie MD. ATHENA: the analysis tool for heritable and environmental network associations. Bioinformatics. 2014;30(5):698-705.