AVERTISSEMENT

Ce document est le fruit d'un long travail approuvé par le jury de soutenance et mis à disposition de l'ensemble de la communauté universitaire élargie.

Il est soumis à la propriété intellectuelle de l'auteur. Ceci implique une obligation de citation et de référencement lors de l'utilisation de ce document.

D'autre part, toute contrefaçon, plagiat, reproduction illicite encourt une poursuite pénale.

Contact : ddoc-theses-contact@univ-lorraine.fr

LIENS

Code de la Propriété Intellectuelle. articles L 122. 4
Code de la Propriété Intellectuelle. articles L 335.2- L 335.10
http://www.cfcopies.com/V2/leg/leg_droi.php
http://www.culture.gouv.fr/culture/infos-pratiques/droits/protection.htm

Nancy-Université
Université Henri Poincaré

paul verlaine université - metz

École Doctorale BioSE (Biologie – Santé – Environnement)

THÈSE

Présentée et soutenue publiquement pour obtenir le titre de

DOCTEUR DE L'UNIVERSITÉ HENRI POINCARÉ

Mention : 'Sciences de la Vie et de la Santé'

Par **Ndeye Coumba NDIAYE**

APPROCHES MÉTHODOLOGIQUES DES ÉTUDES D'ASSOCIATIONS PANGÉNOMIQUES DES FACTEURS DE RISQUE DES PATHOLOGIES CARDIOVASCULAIRES

le 20 Décembre 2010

MEMBRES DU JURY :

Rapporteurs	Monsieur Steve E. Humphries	Professeur, Centre for Cardiovascular Genetics Royal Free University College London Medical School (Londres, Angleterre)
	Monsieur David A. Tregouet	DR2 INSERM, INSERM UMR S 525 (D.A.T.) Université Pierre et Marie Curie (Paris, France)
Examinateurs	Madame Sophie Visvikis-Siest	DR2 INSERM, EA4373 'Génétique Cardiovasculaire' Faculté de Pharmacie (Nancy, France) *Directeur de thèse*
	Madame Éliane Albuisson	Professeur, EA4373 'Génétique Cardiovasculaire' Faculté de Pharmacie (Nancy, France) *Co-directeur de thèse*
	Madame Sudha Seshadri	Professeur, Boston University School of Medicine (Framingham, USA)
	Monsieur Jean-Louis Guéant	Professeur, INSERM U954 Exposition aux Risques génétiques et environnementaux Faculté de Médecine (Nancy, France)
Membre invité	Madame Nathalie Thilly	Département d'Épidémiologie Clinique et d'Évaluation CEC-CIE6 INSERM, EA 4360 Faculté de Médecine (Nancy, France)

Il faut travailler, sinon par goût, au moins par désespoir,

puisse que, tout bien vérifié, travailler est moins ennuyeux que s'amuser

Charles Baudelaire, Journaux intimes

_____

Do or do not, … there is no try

Master Yoda

À ma Trinité : Aïssatou, Mandiogou, Cheikh.

Tchi barké Mame Bathie ak Mame Cheikh.

Remerciements

à Madame le Docteur Sophie Visvikis-Siest, pour la confiance que vous m'avez témoignée dès le début de notre collaboration il y a de cela quatre ans et pour avoir accepté de diriger ma thèse. Votre soutien indéfectible, votre aide précieuse sur le plan scientifique, et vos grandes qualités humaines se sont avérés déterminants dans la réalisation de nos projets. C'est avec votre accompagnement constant tout au long de ces trois années de thèse que j'ai pu m'initier à des méthodologies qui me passionnent aujourd'hui.

À nos collaborations futures…

à Madame le Professeur Éliane Albuisson, pour m'avoir encadré à l'occasion de mon Initiation à la Recherche et pour avoir accepté de co-encadrer ma thèse. Merci de l'intérêt constant que vous avez porté à mes travaux.

aux membres de mon jury, ***Madame le Professeur Sudha Seshadri***, ***Madame le Docteur Nathalie Thilly***, ***Monsieur le Professeur Steve E. Humphries***, ***Monsieur le Professeur Jean-Louis Guéant***, et ***Monsieur le Docteur David A. Tregouet***, je souhaiterai exprimer ma plus sincère reconnaissance. Vous m'honorez en acceptant de juger mes travaux.

à Monsieur le Professeur Philippe Froguel, pour m'avoir donné l'opportunité de collaborer avec votre équipe dans la réalisation de ce projet qui me tient à cœur.

à Monsieur le Docteur David Meyre, pour votre accueil, votre ouverture d'esprit et le sens de l'humour que nous partageons.

à Monsieur le Professeur Gérard Siest, pour vous être constamment intéressé à mon travail et m'avoir permis d'élargir mes connaissances.

à l'INSERM et à la Région Loraine, pour m'avoir alloué une bourse de thèse de 3 ans ; ce qui m'a permis de mener ce projet avec sérénité.

Je souhaiterai également remercier toutes les personnes avec lesquelles j'ai vivement collaboré tout au long de ce projet : le Docteur Rosa-Maria Guéant-Rodriguez, Cécile Lecoeur, le Docteur Nabila-Bouatia-Naji, le Professeur Georges Dedoussis, le Docteur Stéphanie Debette et le Docteur Ming Hue Chen.

Ainsi que tous les membres de l'EA4373 'Génétique Cardiovasculaire' présents et passés, pour les échanges scientifiques et humains : Anastasia, Bérangère, Bernard, Brigitte, Christine, Daniel, Élise, Hermine, Hind, Michèle, Mohsen, Patricia, Rania, Said, Sandy et Shu-Min.

Et à tous ceux que j'aurais oublié, milles excuses et milles mercis.

Enfin, je tiens à conclure ces remerciements avec mes proches qui m'ont soutenue [et supportée dans tous les sens du terme] … et qui se reconnaîtront.

RÉSUMÉ

Les affections cardiovasculaires sont des pathologies complexes résultant des relations et interconnexions d'un grand nombre de facteurs de risque génétiques, constitutionnels et environnementaux. Si l'étiologie génétique de ces affections a largement été étudiée ces dernières années, de nombreuses zones d'ombre persistent. En effet, les variants fonctionnels à l'origine de la large l'héritabilité génétique des facteurs de risque cardiovasculaire et les mécanismes d'interactions gène-gène-environnement masquant ou modulant l'effet de la génétique restent encore à élucider. Le développement d'outils plus performants que l'approche classique gène candidat et la mise en place d'études épidémiologiques axées sur des populations pédiatriques dans lesquelles les interactions gène-environnement sont minimisées pourraient apporter une information déterminante. C'est le postulat principal que nous avons poursuivi au cours de cette thèse.

Ce mémoire présente l'étude de la génétique de certains facteurs de risque des pathologies cardiovasculaires : les lipides, la pression artérielle, l'haptoglobine et le facteur de croissance de l'endothélium vasculaire (VEGF), avec comme outil de base, une population de familles Françaises supposées saines suivies pendant dix ans à l'occasion de trois examens de santé : la cohorte STANISLAS.

Nous avons dans un premier temps quantifié l'héritabilité de ces facteurs de risque cardiovasculaire de façon à souligner l'intérêt de les étudier dans des populations pédiatriques. Ainsi, nous avons mis en évidence la forte héritabilité génétique des traits lipidiques (44% pour le cholestérol total et 32,8% pour le HDL-cholestérol ; $p<0,001$) ; et de l'haptoglobine (35,3%, $p<0,001$) dans la cohorte STANISLAS. Nous avons également constaté une absence d'héritabilité génétique de la pression artérielle, résultat contradictoire avec une étude antérieure que notre équipe a mené sur la même population.

Après avoir décrit l'influence des facteurs de risque cardiovasculaire des parents STANISLAS sur les phénotypes de leurs enfants, nous avons réalisé un Genome-Wide Scan (GWS) sur un échantillon de 631 enfants non apparentés issus de la cohorte STANISLAS de façon à mener des études d'association pangénomique ou Genome-Wide Association Studies (GWAS) selon deux approches distinctes :

- Une approche exploratoire destinée à rechercher des variants génétiques associés à un facteur de risque émergent des pathologies cardiovasculaires : l'haptoglobine ;
- Et une approche réplicative chez l'enfant axée sur des facteurs de risque largement étudiés par la communauté scientifique : les lipides et la pression artérielle.

Les études pangénomiques nécessitant des réplications dans de larges populations indépendantes, nous avons intégré à cette occasion des Consortia Internationaux nous donnant accès à près de 9 000 enfants appartenant à quatre cohortes pédiatriques (OBE, NFBC86, GENDAI et Verona) ainsi qu'à plus de 13 300 adultes provenant de trois cohortes distinctes (DESIR, Framingham et PIVUS).

Nous avons ainsi pu démontrer que certaines associations génétiques mises en évidence chez l'adulte avaient un effet observable dès l'enfance et avons découvert 11 variants génétiques inédits, dont 6 non répliqués chez les adultes : rs12591800, rs4918082, rs2607919, rs933193, rs724815 et rs314672.

Dans le cas de l'haptoglobine, marqueur récemment proposé en prévention cardiovasculaire, nous avons découvert et répliqué un variant expliquant 42% de l'héritabilité génétique globale: le rs2000999 ($p_{méta\text{-}analyse}=8,1\times10^{-59}$). Ce variant a été très récemment associé au cholestérol total et au LDL-cholestérol; association que nous avons répliqué dans nos populations pédiatriques ($p_{cholestérol\ total}=0,002$, $p_{LDL}=0,0008$). Ainsi, nous avons proposé un biomarqueur liant l'inflammation au risque cardiovasculaire et agissant dès l'enfance.

Les discordances que nous avions précédemment constaté en ce qui concerne l'héritabilité de la pression artérielle nous ont mené à expérimenter l'influence d'un facteur tiers : l'obésité, sur l'effet des variants associés à ce trait et ceci nous a permis de mieux souligner l'importance de la prise en compte des facteurs d'environnement.

Ce cheminement faisant, nous avons pu discuter des outils bioinformatiques, biostatistiques et méthodologiques manquants encore aujourd'hui à l'approche pangénomique. Nous avons également traité de l'opportunité de l'utilisation de clusters génétiques en prévention et des interactions gène-gène, problématique classique de l'épidémiologie génétique, en nous appuyant sur l'approche gène-candidat étant donné le manque de méthodologie adéquate dans les études pangénomiques.

Une autre lacune de ces études est le fait qu'elles se limitent généralement à l'identification de variants inédits sans investiguer leurs éventuelles fonctionnalités. Ainsi, nous avons démontré l'intérêt de l'intégration d'une approche transcriptomique aux résultats pangénomiques du VEGF, molécule dont la variabilité interindividuelle est pour plus de la moitié sous contrôle génétique (60,6%). Nous avons pu mettre en évidence quatre variants génétiques originaux indépendamment associés au VEGF : rs6921438 ($p=6,11\text{x}10^{-506}$), rs4416670 ($p=1,47\text{x}10^{-12}$), rs6993770 ($p=2,50\text{x}10^{-16}$) et rs10738760 ($p=1,96\text{x}10^{-34}$) et permettant d'expliquer 48% de l'héritabilité génétique de ce trait. Parmi eux, la fonctionnalité du SNP rs6993770 a pu être démontrée par une approche transcriptomique.

En conclusion, lors de notre thèse, nous avons eu à la fois une démarche épidémiologique et une démarche de proposition de designs intégrés permettant de mener plus avant les investigations sur l'étiologie génétique des pathologies cardiovasculaires.

ABSTRACT

Complex synergies of genetic, constitutional and environmental risk factors are at the origin of cardiovascular pathologies. Their genetic etiology has been extensively investigated these last decades but many questions remain. Indeed, functional variants explaining the large genetic heritability of cardiovascular risk factors and gene-gene-environment interactions hiding or modulating genetics underpinning cardiovascular physiopathology still have to be elucidated. Above the classic candidate gene approach, developing more efficient tools and designing epidemiologic studies focused on pediatric populations in which gene-environment interactions are minimized could be determinant. This is the major postulate followed during our thesis.

Herein we describe our study of the genetics of several cardiovascular risk factors: lipids, blood pressure, haptoglobin and vascular endothelial growth factor (VEGF), on a French family cohort of supposedly healthy subjects followed during ten years: the STANISLAS Family Study.

In a first step, we assessed the heritability of these cardiovascular risk factors and stressed the interest of studying them in pediatric populations. We highlighted the large genetic heritability of lipid traits (44% for total cholesterol and 32.8% for HDL-cholesterol; $p<0.001$); and haptoglobin (35.3%, $p<0.001$) in the STANISLAS cohort. In contradiction with previous studies of our team performed on the same population, we did not observe any significant genetic heritability for blood pressure.

After describing the parental influence of cardiovascular risk factors in the children of the STANISLAS cohort, we performed a Genome-Wide Scan (GWS) on a subsample of 631 unrelated children selected in the STANISLAS cohort and followed two approaches of Genome-Wide Association Studies (GWAS) focused on pediatric populations:

- An exploratory approach seeking for genetic variants associated to an emerging cardiovascular risk factor: haptoglobin;
- And a replication approach focused on widely studied risk factors: lipids and blood pressure.

Since replications in large independent populations are needed in GWAS, we integrated International Consortia giving us access to almost 9,000 children from four pediatric cohorts (OBE, NFBC86, GENDAI and Verona) and 13,300 adults from three different cohorts (DESIR, Framingham and PIVUS).

We demonstrated that genetic associations highlighted in adults were observable in childhood and found 11 novel genetic variants associated to lipid traits and among which 6 were not replicated in adults: rs12591800, rs4918082, rs2607919, rs933193, rs724815 and rs314672.

Concerning haptoglobin, a recently proposed cardiovascular risk marker, we discovered and replicated a variant explaining 42% of its genetic heritability: rs2000999 ($p_{meta\text{-}analysis} = 8.1\times10^{-59}$). This variant was recently associated to total and LDL-cholesterol; an association we replicated in our pediatric populations ($p_{total\ cholesterol} = 0.002$, $p_{LDL} = 0.0008$). SNP rs2000999 thus present a putative link between inflammation and cardiovascular risk beginning in childhood.

The discrepancies previously observed in the heritability assessment of blood pressure led us to evaluate the effect of obesity on variants associated with this trait. We thus stressed the importance of taking into account environmental factors.

This journey gave us the opportunity to address the tools missing in bioinformatics, biostatistics and methodologies used in GWAS approaches. We also discussed the opportunity of using genetic clusters in prevention strategies and gene-gene interactions, a classic problematic of genetic epidemiology, using a candidate gene approach since there is no such methodology available in GWAS yet.

Another major gap was the fact that GWAS generally identify novel variants without seeking for their eventual functionality. We thus demonstrated the interest to integrate transciptomic approaches to GWAS using VEGF, a molecule having more than the half of its interindividual variability under genetic control (60.6%). We highlighted four novel genetic variants having an effect on VEGF: rs6921438 ($p = 6.11\text{x}10^{-506}$), rs4416670 ($p = 1.47\text{x}10^{-12}$), rs6993770 ($p = 2.50\text{x}10^{-16}$) and rs10738760 ($p = 1.96\text{x}10^{-34}$) and explaining 48% of its genetic heritability. Among them, we demonstrated the functionality of rs6993770 using a transcriptomic approach.

In summary, during our thesis, we followed an epidemiologic approach and proposed integrated designs in order to upgrade the investigations on genetic epidemiology of cardiovascular diseases.

Liste des abréviations

A2I	Agence d'innovation industrielle
ABCA1	ATP-binding cassette transporter
ABCG1	ATP-binding cassette sub-family G member 1
ADN	Acide desoxyribonucléique
AHA	American Heart Association
AHR	Aryl hydrocarbon receptor
ALT	Alanine aminotransférase
AMORIS	Apolipoprotein Mortality RISk study
ANAES	Agence nationale d'accréditation et d'évaluation en santé
ANOVA	Analyse de variance
ApoA-I	Apolipoprotéine A-I
ApoB	Apolipoprotéine B
ApoC-III	Apolipoprotéine C-III
ApoE	Apolipoprotéine E
ARN	Acide ribonucléique
AST	Aspartate aminotransférase
AVC	Accident vasculaire cérébral
B12	Vitamine B12 / cobalamine
B6	Vitamine B6 / pyroxidine
BSA	Sérum albumine bovine
CBS	Cystathionine β-synthase
CCPPRB	Comité Consultatif de Protection des Personnes dans la Recherche Biomédicale
CépiDC	Centre d'épidémiologie sur les causes médicales de décès
CETP	Protéine de transfert du cholestérol estérifié
CHARGE	Cohorts for Heart and Aging Research in Genomic Epidemiology
CIM	Classification internationale des maladies
CMP	Centre de médecine préventive
CNIL	Commission Nationale de l'Informatique et des Libertés
CNRS	Centre national de la recherche scientifique
CPAM	Caisse primaire d'assurance maladie
CRB	Centre de ressources biologiques
CRP	Protéine C-réactive
CV	Coefficient de variation
DESIR	Data from an Epidemiological Study on the Insulin Resistance syndrome
EDTA	Ethylenediaminetetraacetic
EGF	Epidermal growth factor
EGIR	European Group for the Study of Insulin Resistance
EIM	Epaisseur intima-média
ELISA	Enzyme-Linked Immunosorbent Assay
EPIC	European Prospective Investigation of Cancer
EUA	Excrétion urinaire d’albumine

EUROSPAN	European Special Populations Research Network
FGF	Fibroblast Growth Factor
FHS	Framingham heart study
FTO	Fat mass and obesity associated gene
FVII	Facteur VII
G2h	Glycémie deux heures après charge glucosée
GAJ	Glycémie à jeun
GENDAI	GeNe and Diet Attica Investigation
GGT	Gamma-glutamyltransférase
GWAS	Genome-wide association study
GWS	Genome-wide scan
HAS	Haute autorité de santé
HDL-C	High-density lipoprotein cholesterol
HHP	Honolulu Heart Program
HP	Haptoglobin gene
HPR	Haptoglobin related protein gene
HTA	hypertension
HWE	Equilibre de Hardy Weinberg
IAJ	Isulinémie à jeun
ICAM-I	Intercellular adhesion molecule-I
IDF	International Diabetes Federation
IGE-PCV	Interactions gène-environnement en physiopathologies cardiovasculaires
IL-6	Interleukine-6
IL-8	Interleukine-8
IMC	Indice de masse corporelle
INSERM	Institut national de la santé et de la recherche médicale
LCAT	Lécithine cholestérol acyl transférase
LD	Déséquilibre de liaison
LDL-C	Low-density lipoprotein cholesterol
MA	Méta-analyse
MAF	Fréquence de l'allèle mineur
MCP-1	Monocyte chemotactic protein-1
MCV	Maladies cardiovasculaires
MONICA	MONItoring trends and determinants in CArdiovascular disease
MTHFR	Methylenetetrahydrofolate reductase
NCETP-ATPIII	National Cholesterol Education Program’s Adult Treatment Panel III
NFBC86	Northern Finland 1986 Birth Cohort
OLA	Oligonucleotide Ligation Assay
OMS	Organisation mondiale de la santé
PAD	Pression artérielle diastolique
PAI-1	Inhibiteur de l'activateur du plasminogène
PAL	Phosphatase alcaline
PAS	Pression artérielle systolique
PCR	Polymerase chain reaction

PIVUS	Prospective Investigation of the Vasculature in Upssala Seniors
PP	Pression pulsée
PRI	Protéines de la Réaction Inflammatoire
RCVA	Risque cardiovasculaire absolu
ROS	Espèces d'oxygène réactives
SAH	S-adenosyl-l-*homocysteine*
SAM	S-adenosyl methionine
SM	Syndrome métabolique
SNP	Single nucleotide polymorphism
SR-B1	Sterol regulatory-binding protein
STANISLAS	Suivi Temporaire Annuel Non Invasif de la Santé des Lorrains Assurés Sociaux
TE	Tris-EDTA
TG	Triglycérides
TNFα	Tunor necrosis factor-α
UMR	Unité mixte de recherche
VEGF	Facteur de croissance de l'endothélium vasculaire
VIH	Virus de l'immunodéficience humaine

TABLE DES MATIÈRES

Liste des tableaux 2

Liste des figures 3

AVANT-PROPOS 6

LISTE DES PUBLICATIONS DE THÈSE 10

CHAPITRE 1 – HYPOTHÈSES ET OBJECTIFS 13

CHAPITRE 2 – SITUATION DU SUJET 17
1. Épidémiologie des maladies cardiovasculaires 18
2. Physiopathologie des maladies cardiovasculaires 20
   2.1. Les facteurs de risque classiques des maladies cardiovasculaires 22
      2.1.1. L'âge et le sexe 22
      2.1.2. Les antécédents familiaux 23
      2.1.3. Le Tabagisme 23
      2.1.4. L'hypertension artérielle 24
      2.1.5. Les dyslipidémies 25
         Le cholestérol total et le LDL-cholestérol 25
         Le HDL-cholestérol 25
         Les triglycérides 27
      2.1.6. Le diabète 27
      2.1.7. L'obésité 28
      2.1.8. Les facteurs nutritionnels 29
      2.1.9. La sédentarité 30
      2.1.10. Le syndrome métabolique 31
   2.2. Les facteurs de risque émergents des maladies cardiovasculaires 34
      2.2.1. L'hyperhomocystéinémie 34
      2.2.2. Les facteurs psychosociaux et environnementaux 35
      2.2.3. Les facteurs thrombogéniques 36
      2.2.4. Les marqueurs de l'inflammation 36
      2.2.5. Les molécules d'adhésion 39
   2.3. La composante génétique des maladies cardiovasculaires 41

CHAPITRE 3 – POPULATIONS ET MÉTHODES 46
1. Populations 47
   1.1. La cohorte STANISLAS 47
      1.1.1. Questionnaire alimentaire et de mode de vie 48
      1.1.2. Données cliniques et anthropométriques 49
      1.1.3. Impédance bioélectrique 49
      1.1.4. Dosages biologiques 50
      1.1.5. Les données génétiques 52
         La technique Multiplex en collaboration avec Roche Molecular Systems 52
         Le typage pangénomique ou genome-wide scan 53
         Le typage spécifique des variants génétiques issus de la phase initiale de genome-wide association study en vue des analyses de réplication 54
      1.1.6. Les données transcriptomiques 56
   1.2. La population OBE 57

1.3. La Northern Finland 1986 Birth Cohort (NFBC86) 57
1.4. La GeNe Diet Attica Investigation (GENDAI) 57
1.5. La cohorte VERONA 58
1.6. La cohorte Data from an Epidemiological Study on the Insulin Resistance syndrome (DESIR) 58
1.7. La cohorte FRAMINGHAM 58
1.8. La Prospective Investigation of the Vasculature in Upssala Seniors (PIVUS) study 59
2. Analyses statistiques 59
2.1. Test de l'équilibre de Hardy-Weinberg 59
2.2. Déséquilibre de laison 61
2.3. Test d'ajustement : loi de distribution des variables 61
2.4. Statistiques descriptives 61
2.5. Études d'association 61
2.6. Corrélations familiales 62

CHAPITRE 4 – RÉSULTATS ET DISCUSSION 64
1. Étude de l'effet précoce de l'héritabilité génétique des facteurs de risque des pathologies cardiovasculaires dans la cohorte STANISLAS 64
1.1. L'héritabilité des traits lipidiques, de la pression artérielle, de l'haptoglobine et du facteur de croissance de l'endothélium vasculaire dans la Cohorte STANISLAS 65
1.2. Étude exploratoire de l'effet phénotypique de l'héritabilité des facteurs de risque des maladies cardiovasculaire chez les enfants de la Cohorte STANISLAS 73
2. Étude d'associations pangénomiques de facteurs de risque des maladies cardiovasculaire dans des populations pédiatriques : l'haptoglobine, la pression artérielle et les lipides 80
2.1. Genome-Wide Association Study de l'Haptoglobine dans des populations pédiatriques 82
2.2. Genome-Wide Association Study de la pression artérielle dans des populations pédiatriques. Effet de l'obésité 86
2.3. Genome-Wide Association Study de traits lipidiques dans des populations pédiatriques. Identification de variants inédits 92
3. Identification de profils génétiques chez les sujets supposés sains de la cohorte STANISLAS : répercussions phénotypiques 94
4. Quels sont les challenges méthodologiques qu'attendent les GWAS ? 100
4.1. Approche épistasique de la variabilité génétique de la pression artérielle 102
4.2. Étude d'association pangénomique du VEGF et identification des variants responsable de sa forte héritabilité génétique. Vers une approche intégrée en transcriptomique 105

CHAPITRE 5 – CONCLUSIONS ET PERSPECTIVES 112

RÉFÉRENCES BIBLIOGRAPHIQUES 119

PUBLICATIONS 132

Liste des tableaux

- Tableau 1 : Mortalité par maladies de l'appareil circulatoire (CIM10) en France en 2008 19
- Tableau 2 : Surmortalité masculine par maladies de l'appareil circulatoire (CIM10) en France 22
- Tableau 3 : Classification de l'indice de masse corporelle 28
- Tableau 4 : Classification du tour de taille 29
- Tableau 5 : Définitions du syndrome métabolique 33
- Tableau 6 : Estimation de l'héritabilité des accidents vasculaires cérébraux, des affections coronariennes et des principaux facteurs de risque cardiovasculaire 42
- Tableau 7 : Caractéristiques basales des 2 471 sujets inclus dans l'étude transversale - Données issues de la cohorte STANISLAS à t_0 : 1994-1995 68
- Tableau 8 : Évolution des caractéristiques des 1 208 sujets inclus dans l'étude longitudinale - Données issues de la cohorte STANISLAS à t_0 (1994-1995) et t_5 (1999-2000) 70
- Tableau 9 : Composante de variance dans les 667 familles inclues à l'étude transversale et des 353 familles inclues dans l'étude longitudinale 72
- Tableau 10 : Caractéristiques phénotypiques des familles étudiées et variations au cours des examens de santé 76
- Tableau 11 : Phases de mise en évidence et réplication de l'association du rs2000999 avec le taux d'haptoglobine dans les populations pédiatriques 84
- Tableau 12 : Estimation de l'héritabilité de l'haptoglobine dans la cohorte STANISLAS – composante de variance de Fisher 85
- Tableau 13 : Étude de réplication chez les enfants minces et obèses 88
- Tableau 14 : Évaluation clinique et mesures biologiques au sein des clusters au premier examen de santé (t_0) 96
- Tableau 15 : Évaluation clinique et mesures biologiques au sein des clusters au second examen de santé (t_5) 98
- Tableau 16 : Caractéristiques phénotypiques de la population étudiée 104
- Tableau 17 : SNPs 'genome-wide significants' à la suite de la réplication 108
- Tableau 18 : Corrélations significatives entre les variants génotypés dans la cohorte STANISLAS et les transcripts VEGF 110

Liste des figures

- Figure 1 : Évolution par sexe des taux de décès par maladies de l'appareil circulatoire (CIM10) standardisés par âge en France 18
- Figure 2 : Transport reverse du cholestérol et autres propriétés du HDL-cholestérol 26
- Figure 3 : Physiopatholgie du Syndrome Métabolique 32
- Figure 4 : Métabolisme de l'homocystéine 34
- Figure 5 : Profils de variation de concentration plasmatique des protéines de la phase aiguë de l'inflammation à la suite d'un stimulus modéré 37
- Figure 6 : Stades d'évolution de la plaque d'athérosclérose. De la strie lipidique à la rupture de plaque 38
- Figure 7 : Variants du VEGF. Description de l'utilisation des exons dans les différents variants du VEGF, et de la fonction des séquences qu'ils codent 40
- Figure 8 : Modèle d'interactions gène-gène-environnement impliquées dans la physiopathologie des maladies cardiovasculaires 41
- Figure 9 : Phases de recrutement de la cohorte STANISLAS 47
- Figure 10 : Modalités du typage pangénomique utilisant la puce Illumina Human CNV370-Duo 54
- Figure 11 : Design classique des Genome-wide association studies 55
- Figure 12 : Groupes familiaux en fonction du niveau de risque des parents STANISLAS 74
- Figure 13 : Niveaux des facteurs de risque cardiovasculaire des enfants Stanislas en fonction des groupes familiaux au premier examen de santé 78
- Figure 14 : Niveaux des facteurs de risque cardiovasculaire des enfants Stanislas en fonction des groupes familiaux au second examen de santé 79
- Figure 15 : Signature des clusters génétiques identifiés 95
- Figure 16 : organisation génomique du gène VEGF 106
- Figure 17 : Stratégie d'étude intégrative de l'étiologie génétique des Maladies Cardiovasculaires 115

AVANT-PROPOS

En 2030, les maladies cardiovasculaires (MCV) seront à l'origine de 23,56 millions de décès et représenterons toujours la principale cause de mortalité au niveau mondial selon les prévisions de l'Organisation Mondiale de la Santé (OMS).

De nombreuses études épidémiologiques ont été à l'origine d'une meilleure compréhension de l'étiologie de ces pathologies complexes et à la mise en place de mesures préventives basées sur la minimisation des facteurs de risque identifiés. Ainsi, sur le plan évolutif, 'l'épidémie cardiovasculaire' a régressé dans les pays industrialisés depuis le début des années 1980 … mais pour atteindre un plateau.

Cette inflexion de l'effet des mesures préventives est bien évidemment due aux limites de l'efficacité des interventions mises en place, notamment à l'adhésion des populations cibles (réponses aux campagnes de prévention visant le tabagisme, la consommation d'alcool, … etc.), mais surtout au fait que la communauté scientifique n'ait pas encore élucidé la totalité des phénomènes menant au développement des pathologies cardiovasculaires. Notamment la génétique sous-jacente.

Le challenge de ces prochaines années est donc de 'réinventer' les techniques et méthodologies d'épidémiologie génétique dont nous disposons.

Historiquement, ces études épidémiologiques étaient basées sur une approche 'gène candidat' définie selon une connaissance *a priori* de la physiopathologie de la maladie d'intérêt et la communauté scientifique s'est donc logiquement heurté aux limites de ses propres connaissances. D'où le plateau observé.

Avec l'avènement des typages pangénomiques ou Genome-Wide Scan (GWS), une approche s'affranchissant de l'hypothèse préalable a vu le jour et a permis de mettre en évidence des cheminements physiopathologiques non soupçonnés et par la même occasion, des biomarqueurs inédits. Les études d'associations pangénomiques ou Genome-Wide Association Studies (GWAS) en sont encore à leurs préludes et n'intègrent pas encore les techniques statistiques complexes qui permettraient, comme pour l'approche gène-candidat, d'incorporer les interactions gène-gène (épistasie) et les interactions gène-environnement. Elles ont néanmoins permis de révolutionner la connaissance étiologique des MCV et des facteurs de risque associés.

L'équipe EA4373 'Génétique Cardiovasculaire' dispose d'un outil permettant de mener à bien ces investigations : la Cohorte STANISLAS regroupant à l'inclusion 1 006 familles nucléaires d'origine Lorraine depuis plus de deux générations. Ces familles, composées de deux parents ayant au moins deux enfants, ont fait l'objet de trois bilans de santé à cinq ans d'intervalle au Centre de Médecine Préventive (CMP) de Vandœuvre-lès-Nancy. Au cours de ces bilans de santé, des données à la fois socio-économiques, biologiques et cliniques ont été recueillies. Au total, 600 variables différentes ont été obtenues pour chaque participant et des banques ADN / ARN / et lymphocytaires ont été constituées et conservées dans un Centre de Ressources Biologiques (CRB) mis en place à cette occasion : le CRB Interactions Gènes-Environnement en Physiopathologies CardioVasculaires (IGE-PCV).

Cet outil nous a permis de mettre en place une stratégie d'étude de divers facteurs de risque des MCV avec une approche pangénomique et de proposer des outils méthodologiques et logistiques acquis tout au long de nos travaux de recherche.

La nécessité, dans le design de telles études, d'analyses confirmatoires dans des populations indépendantes, a mené notre équipe à intégrer des Consortia Internationaux de façon à avoir accès à de larges populations de réplication de nos résultats. Nous avons fait appel en priorité à des populations pédiatriques afin de mieux explorer les interactions gène-environnement du fait de la complexité de telles interactions et du manque de standardisation de l'information concernant ces facteurs dans les populations adultes existantes : les facteurs d'environnement ont un poids moindre chez l'enfant.

Parallèlement, nous avons traité de questions importantes en épidémiologie génétique comme l'utilité des clusters dans la classification des sujets à risque cardiovasculaire ainsi que la nécessité d'intégrer la transcriptomique aux études pangénomiques de façon à déterminer la fonctionnalité éventuelle des variants génétiques mis en évidence.

Notez que lors de la rédaction de ce mémoire, nous avons choisi d'utiliser le 'nous' singulier. Notre contribution effective dans chaque étude est signalée dans le chapitre 'PUBLICATIONS' à la fin de la page de résumé de chaque article.

En règle générale, nous avons participé à la conception des designs des études, à la veille bibliographique et méthodologique, aux débriefings hebdomadaires fédérant l'ensemble des équipes collaboratrices, aux analyses statistiques (en ayant préalablement suivi des formations spécifiques pour les analyses pangénomiques), à la rédaction des articles, ainsi qu'à leur suivi après soumission.

Nous avons également participé à la gestion des échantillons biologiques au sein du CRB IGE-PCV : aliquotages et gestion de la traçabilité des échantillons biologiques.

LISTE DES PUBLICATIONS DE THÈSE

Avertissements

Certains des articles présentés dans ce rapport de thèse sont en cours de publication. Par conséquent, les résultats qu'ils exposent doivent rester confidentiels. Ces cas de figures sont signalés en grisé dans la liste ci-dessous, dans les résumés ainsi que dans le mémoire.

Les auteurs signalés par un '*' ont contribué de façon équivalente au travail de recherche et de rédaction des articles concernés.

REVUES

Ndeye Coumba Ndiaye*, Mohsen Azimi Nehzad*, Said El Shamieh*, Sophie Visvikis-Siest*. Cardiovascular Diseases and Related Quantitative Traits in the Era of Genome-Wide Association Studies: Where have we Been and What should we Expect? 2010 [soumis]

ARTICLES

Bernard Herbeth, Anastasia Samara, **Ndeye Coumba Ndiaye**, Jean-Brice Marteau, Hind Berrahmoune, Gérard Siest, Sophie Visvikis-Siest. Metabolic syndrome-related composite factors over 5 years in the STANISLAS Family Study: Genetic heritability and common environmental influences. Clinica Chimica Acta. 2010; 411: 833–9.

Ndeye Coumba Ndiaye*, Éliane Albuisson*, Isabelle Colombet, Gilles Chatellier, Jean-Brice Marteau, Sébastien Dadé, Gérard Siest, Sophie Visvikis-Siest. Parental precocious influences on offspring cardiovascular risk markers: an exploratory study in the STANISLAS cohort. Pers Med. 2009;6(3):343-52.

Philippe Froguel*, **Ndeye Coumba Ndiaye***, Nabila Bouatia-Naji, Cécile Lecoeur, Gérard Siest, Bernard Herbeth, Jean-Brice Marteau, Rosa-Maria Rodriguez-Guéant, Marion Marchand, Éliane Albuisson, Anita Morandi, Yann Labrune, Vincent Vatin, Claudio Maffeis, Paul Elliott, Marjo-Riita Jarvelin, George V Dedoussis, David Meyre, Sophie Visvikis-Siest. GWAS LDL and total cholesterol associated rs2000999 in *Haptoglobin Related Protein*

explains 42% of the Haptoglobin Levels Genetic Variability in children: a putative biological link between inflammation and cardio-vascular risk factors. 2010 [soumis]

Éliane Albuisson, Sandy Maumus, **Ndeye Coumba Ndiaye**, Bérangère Marie, Nicolas Jay, François Kholer, Gérard Siest, Sophie Visvikis-Siest. Genetic profiling in healthy subjects from the STANISLAS cohort based on 24 polymorphisms: effects on biological variables. Clin Chem Lab Med. 2008;46(1):64–72.

Ndeye Coumba Ndiaye, Said EL Shamieh, Jean Brice Marteau, Gérard Siest, Micheal Y Tsai, Sophie Visvikis-Siest. Functional epistases regulate blood pressure. 2010 [soumis]

Stéphanie Debette*, Sophie Visvikis-Siest*, Ming Hue Chen, **Ndeye Coumba Ndiaye**, C Song, AL DeStefano, R Safa, Mohsen Azimi Nezhad, D Sawyer, Jean-Brice Marteau, V Xanthakis, Gérard Siest, L Sullivan, Michèle Pfister, H Smith, SH Choi, J Lamont, L Lind, Q Yang, P Fitzgerald, E Ingelsson, Ramachandran S Vasan, Sudha Seshadri. Identification of Cis- and Trans-acting Genetic Variants Explaining up to Half the Heritability of Circulating VEGF levels. 2010 [soumis]

AUTRES PUBLICATIONS

Gérard Siest, Hamanou Benachour, Daniel Lambert, Jean-Brice Marteau, Christine Masson, **Ndeye Coumba Ndiaye**, Sophie Visvikis-Siest. Functional genomics towards personalized healthcare. 4th 'Biologie Prospective' Santorini Conference, Santorini Island, Greece, September 21-23, 2008. Pers Med. 2009;6(1):19-32.

Adam Barker, Stephen J. Sharp, Nicholas J. Timpson, Nabila Bouatia-Naji, Nicole M. Warrington, Stavroula Kanoni, Lawrie J. Beilin, Soren Brage, Panos Deloukas, David Evans, Anders Grontved, Neelam Hassanali, Deborah A. Lawlor, Cecile Lecoeur, Ruth J. F. Loos, Stephen J. Lye, Mark I. McCarthy, Trevor Mori, **Ndeye Coumba Ndiaye**, John P. Newnham, Ioanna Ntalla, Craig E. Pennell15, Beate St Pourcain, Inga Prokopenko, Sue M. Ring, Naveed Sattar, Sophie Visvikis-Siest, George V. Dedoussis, Lyle J. Palmer, Philippe Froguel, George Davey Smith, Ulf Ekelund, Nicholas J. Wareham, Claudia Langenberg. Association of genetic loci with glucose levels in early life: a meta-analysis of over 6,000 children. 2010 [soumis]

CHAPITRE 1 – HYPOTHÈSES ET OBJECTIFS

Le développement des MCV est conditionné par des interrelations complexes existant entre facteurs de risque génétiques, constitutionnels et environnementaux que les nombreuses études d'épidémiologie étiologique ont permis d'élucider en partie. En partie seulement, notamment en ce qui concerne l'épidémiologie génétique, car l'approche classique était basée sur la présélection de variants génétiques impliqués dans des voies métaboliques connues.
Or la physiopathologie des MCV empreinte également des voies méconnues que seul l'avènement des études pangénomiques 'hypothesis free' a permis d'entrevoir.

Ainsi, grâce à cette révolution technologique, depuis 2007, de nombreux variants inédits associés aux pathologies cardiovasculaires ou à leurs facteurs de risque ont été mis en évidence et de nouvelles voies métaboliques identifiées. Néanmoins, des zones d'ombres persistent, notamment l'effet cumulatif des variants identifiées qui reste pour le moment très en deçà de l'héritabilité génétique estimée des facteurs de risque cardiovasculaire. Nous pensons que ce fossé provient de carences dans l'information et la modélisation génétiques. Côté modélisation, l'héritabilité manquante pourrait s'expliquer par des phénomènes épistasiques, c'est-à-dire d'interaction entre gènes. En effet, leur influence n'est pas simplement additive, ou dominante/récessive, mais aussi interactive et seule une approche systémique par réseaux de gènes permettra de l'évaluer. De plus, les interactions gène-environnement ne sont pas encore intégrées aux approches pangénomiques alors qu'il est établi que dans certains cas, la présence de gènes parfaitement fonctionnels ne suffit pas à l'émergence d'un trait chez l'individu en l'absence de certaines stimulations environnementales. Enfin, la fonctionnalité de l'immense majorité des variants issus des GWAS reste méconnue.

Ainsi, dans cette thèse, nous avons émis l'hypothèse que la compréhension des pathologies multifactorielles par l'approche pangénomique est conditionnée par la nécessité de mettre en œuvre des designs intégrés permettant d'appréhender les 'carences' méthodologiques qu'elle comporte. Notamment, l'étude de la fonctionnalité des variants inédits, des phénomènes épistasiques et des interactions gène-environnement qu'il est difficile d'appréhender dans le cadre d'études de réplication intégrant des populations d'horizons divers et avec un recueil de l'information non standardisé. Pour ce dernier point, notre postulat était de nous focaliser sur des populations pédiatriques où les facteurs d'environnement (tabagisme, consommation d'alcool, prise de la pilule, … etc.) sont moins importants pour la recherche de nouveaux variants de quatre voies métaboliques associées au risque cardiovasculaire : les lipides, la pression artérielle, l'inflammation via l'haptoglobine et l'adhésion via le facteur de croissance de l'endothélium vasculaire (VEGF).

Nous avons donc poursuivi quatre objectifs majeurs :

(1) Acquérir les compétences méthodologiques et logistiques que nécessite une approche pangénomique.
(2) Mettre en évidence des variants inédits associés aux facteurs de risque des pathologies cardiovasculaires précédemment cités.
(3) Appréhender les possibilités et limitations méthodologiques des études d'associations pangénomiques dans le domaine du risque cardiovasculaire.
(4) Expérimenter une stratégie globale exploratoire intégrant l'étude des interactions gène-gène et gène-environnement en nous basant sur des gènes-candidats ainsi que sur la recherche de fonctionnalité par la transcriptomique.

CHAPITRE 2 – SITUATION DU SUJET

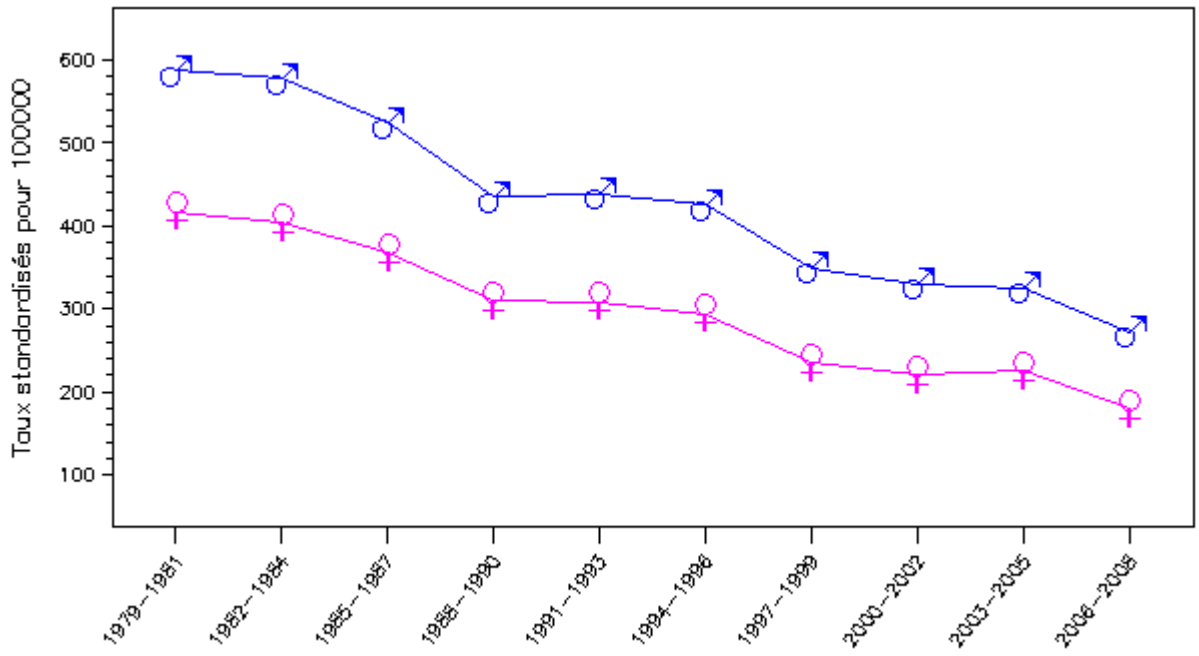
1. Épidémiologie des maladies cardiovasculaires

Les MCV sont à l'origine de 17,1 millions décès chaque année dans le monde, ce qui représente 29% de la mortalité globale selon les données de l'OMS et elles atteindrons 23,56 millions de décès au niveau mondial en 2030.

En France, malgré un ralentissement de la mortalité cardiovasculaire depuis ces trente dernières années (**Figure 1**), ces affections ont été à l'origine de plus de 146 000 décès en 2008 (53,5% de femmes) soit 27,5% de la totalité des décès toutes causes confondues (**Tableau 1**) selon les données du 'Centre d'Épidémiologie sur les causes médicales de Décès' (CépiDC – INSERM, http://www.cepidc.vesinet.inserm.fr/).

Les MCV sont donc toujours une préoccupation majeure de santé publique, dont les principales manifestations sont les cardiopathies ischémiques et les maladies cérébrovasculaires. Elles représentaient respectivement 25,7 et 22% de décès en 2008.

Figure 1 : Évolution par sexe des taux de décès par maladies de l'appareil circulatoire (CIM10) standardisés par âge en France



Source : CépiDC – INSERM (http://www.cepidc.vesinet.inserm.fr/)

Tableau 1 : Mortalité par maladies de l'appareil circulatoire (CIM10) en France en 2008

Code CIM	Libellé	Sexe	Total	<1	1-4	5-14	15-24	25-34	35-44	45-54	55-64	65-74	75-84	85-94	95+
I00-I99	**Maladie de l'appareil circulatoire**	M	67923	31	17	12	90	225	1010	2966	6284	10175	24402	19719	2992
		F	78204	28	19	20	53	135	414	1040	2014	4951	22233	35216	12081
		T	146127	59	36	32	143	360	1424	4006	8298	15126	46635	54935	15073
I20 -I25	Cardiopathies ischémiques	M	21348	2	0	0	6	45	389	1256	2534	3650	7581	5257	628
		F	16178	1	3	1	1	9	84	243	440	1185	4896	7211	2104
		T	37526	3	3	1	7	54	473	1499	2974	4835	12477	12468	2732
I30-I33, I39-I52	Autres cardiopathies	M	19279	13	11	10	48	89	263	686	1407	2242	6458	6726	1326
		F	25719	17	12	10	20	38	83	224	516	1184	6289	12066	5260
		T	44998	30	23	20	68	127	346	910	1923	3426	12747	18792	6586
I60 -I69	Maladies cérébrovasculaires	M	13452	5	4	1	24	52	177	500	1011	2069	5369	3756	484
		F	18729	4	2	6	10	46	133	310	516	1373	5920	8187	2222
		T	32181	9	6	7	34	98	310	810	1527	3442	11289	11943	2706
I00-I99 nc preced	Autres maladie de l'appareil circulatoire	M	13844	11	2	1	12	39	181	524	1332	2214	4994	3980	554
		F	17578	6	2	3	22	42	114	263	542	1209	5128	7752	2495
		T	31422	17	4	4	34	81	295	787	1874	3423	10122	11732	3049

Source : CépiDC – INSERM (http://www.cepidc.vesinet.inserm.fr/)

2. Physiopathologie des maladies cardiovasculaires

Sous la terminologie de MCV, la littérature médicale fait référence à différentes pathologies chroniques ou évènements ayant en commun une physiopathologie liée à l'athérosclérose infraclinique et responsable de mort prématurée : les maladies coronariennes, les accidents vasculaires cérébraux (AVC), les pathologies vasculaires périphériques et l'insuffisance cardiaque. Ces affections sont dites multifactorielles car elles sont la résultante de l'association synergique de plusieurs facteurs de risque 'se rapportant à des attributs, caractéristiques ou expositions d'un individu, associés à une incidence accrue de l'affection' (lien de causalité entre le facteur et la maladie). Ces facteurs de risque sont retenus en fonction de différents critères :

- La force de leur association avec la maladie exprimée par le risque relatif observé chez les sujets exposés par rapport aux non exposés ;
- Le caractère graduel de cette association quand le risque de développer la maladie augmente parallèlement au niveau du facteur de risque ;
- La temporalité c'est-à-dire le fait que le facteur de risque précède bien la maladie et non l'inverse ;
- Le caractère universel de l'association vérifié par la cohérence entre différentes études portant sur des populations différentes et des groupes vivant dans des conditions différentes ;
- L'indépendance, c'est-à-dire la persistance de l'association entre le facteur et la maladie même lorsque sont pris en considération les effets d'autres facteurs de risque dans le cadre d'une analyse multivariée ;
- Le caractère plausible de l'association vis-à-vis des résultats d'autres approches (sciences fondamentales, expérimentations animales ou in vitro, études cliniques, … etc.) ;
- Et enfin la réversibilité quand la correction du facteur de risque permet de réduire l'incidence de la maladie. Il s'agit là de l'argument le plus fort en termes de causalité nécessitant la réalisation d'essais cliniques d'intervention pour être affirmé.

Classiquement, les recommandations identifient trois groupes de facteurs de risque cardiovasculaire en fonction de leur degré d'imputabilité et de l'état actuel des connaissances scientifiques [Agence nationale d'accréditation et d'évaluation en santé (ANAES) Juin 2004. Méthodes d'évaluation du risque cardiovasculaire global. Service Évaluation en Santé Publique ; Grundy et al[a]. 1999 ; Grundy et al[b]. 1999] :

- Les facteurs de risque majeurs qui ont un effet multiplicateur sur le risque cardiovasculaire indépendamment des autres facteurs de risque ;
- Les facteurs de risque prédisposants qui ont un effet potentialiseur lorsqu'ils sont associés aux facteurs de risque majeurs ;
- Et les facteurs de risque émergents (ou discutés) des MCV associés à un risque augmenté mais dont le degré d'imputabilité et/ou le mécanisme d'action sont méconnus à ce jour.

Ces facteurs de risque sont soit constitutionnels ou génétiques donc non modifiables, soit environnementaux et donc modifiables.

Notez que le terme 'marqueur de risque' est préférable dans les cas où le facteur est fortement associé au risque de survenue de la MCV tout en n'ayant pas l'un des critères précédemment cités. Notamment la séquence temporelle ou le caractère indépendant. Par exemple, l'impact de l'obésité sur le risque coronarien est en bonne partie dépendant de son impact propre sur l'hypertension artérielle (HTA), les dyslipidémies et le diabète.
Dans la pratique toutefois de tels marqueurs sont très utiles pour identifier certains individus particulièrement à risque, car ils peuvent intégrer l'impact de plusieurs facteurs de risque. Il peut s'agir de paramètres cliniques comme la pression pulsée (PP) ; ou de mesures para-cliniques comme l'épaisseur intima-média (EIM) mesurée au niveau carotidien et qui prédit le risque ultérieur d'évènement cardiovasculaire.

2.1. Les facteurs de risque classiques des maladies cardiovasculaires

Les facteurs de risque classiques et donc validés des MCV sont l'âge et le sexe, les antécédents familiaux, le tabagisme, l'HTA, les dyslipidémies, le diabète, l'obésité, les facteurs nutritionnels et la sédentarité.

2.1.1. L'âge et le sexe

Il est bien établi que le risque cardiovasculaire augmente avec l'âge et que l'homme est nettement plus exposé aux accidents cardiovasculaires que la femme en période d'activité génitale. Cette disparité est d'ailleurs visible sur les courbes d'évolution des taux de décès (**Figure 1**). Néanmoins, la surmortalité masculine est plus marquée avant 65 ans (**Tableau 2**), avec des proportions quasiment inchangées depuis 30 ans ; puis le niveau de risque de la femme rejoint très progressivement celui de l'homme après la ménopause.

Tableau 2 : Surmortalité masculine par maladies de l'appareil circulatoire (CIM10) en France

	Tous âges			Moins de 65ans		
période	**Masculin**	**Féminin**	**M/F***	**Masculin**	**Féminin**	**M/F***
1979-1981	587,6	417,8	1,4	91,1	30,1	3,0
1982-1984	577,9	403,6	1,4	95,3	30,7	3,1
1985-1987	524,3	368,4	1,4	94,7	29,8	3,2
1988-1990	435,2	309,5	1,4	79,4	25,2	3,2
1991-1993	439,4	309,0	1,4	71,6	23,0	3,1
1994-1996	425,4	294,6	1,4	65,2	20,6	3,2
1997-1999	349,6	234,8	1,5	60,8	19,3	3,2
2000-2002	330,8	221,0	1,5	52,6	16,7	3,1
2003-2005	325,1	225,1	1,4	44,0	14,1	3,1
2006-2008	272,2	179,0	1,5	41,1	13,4	3,1

Taux standardisés par âge pour 100 000
Source : CépiDC – INSERM (http://www.cepidc.vesinet.inserm.fr/)

Lors de ses recommandations en 2000, L'ANAES, devenue la Haute Autorité de Santé (HAS) en 2004, proposait de considérer l'âge comme un facteur de risque des MCV à partir de 45 ans chez l'homme et de 55 ans chez la femme.

2.1.2. Les antécédents familiaux

Il s'agit des accidents cardiovasculaires précoces survenus chez les parents du premier degré (avant 55 ans chez l'homme et 65 ans chez la femme).

Il a été démontré, notamment par une étude prospective au sein de la cohorte Framingham [Schildkraut et al. 1989], que la survenue d'un décès d'origine coronarienne chez un parent augmentait de 30% le risque de maladie coronarienne chez ses enfants, hommes ou femmes, avec une probabilité plus forte qu'elle s'installe avant 60 ans. Ce facteur de risque est d'ailleurs l'un des composants des équations de calcul du risque cardiovasculaire absolu (RCVA) [Anderson et al[a]. 1991 ; Anderson et al[b]. 1991].

Néanmoins, sa prise en compte comporte une certaine imprécision car l'âge réel de survenue de l'évènement cardiovasculaire est souvent difficile à déterminer par le sujet investigué. De plus, cette prédisposition familiale englobe également les anomalies métaboliques familiales (dyslipidémies, diabète, … etc.) déjà incluses dans les équations de risque.

2.1.3. Le tabagisme

Avec de nombreux composants délétères favorisant les complications de l'athérosclérose [Pipe et al. 2010] et un effet d'autant plus élevé que le sujet est jeune [Gooding & Ferranti. 2010], le tabagisme est un facteur de risque cardiovasculaire modifiable exposant tout particulièrement au risque d'infarctus du myocarde et de mort subite : les gros fumeurs inhalant la fumée voyaient leur risque de mort subite multiplié par 5 dans l'Étude prospective parisienne [Jouven et al. 1999]. Il est également un facteur de risque majeur de l'artériopathie oblitérante des membres inférieurs.

Son caractère réversible est bien établi : en prévention primaire, le risque coronarien des sujets en sevrage tabagique depuis 2 à 3 ans ne diffère pas significativement de celui de sujets n'ayant jamais fumé. En prévention secondaire le risque de récidive d'infarctus ou de décès diminue de près de 50% dès la première année pour rejoindre en quelques années le risque de sujets n'ayant jamais fumé.

2.1.4. L'hypertension artérielle

Il n'existe pas de seuil de pression artérielle individualisé en deçà duquel le risque cardiovasculaire que constitue ce trait peut être considéré comme nul.
Le rôle délétère de la pression artérielle diastolique (PAD) a été le premier à avoir été mis en évidence mais c'est la pression artérielle systolique (PAS) qui a la signification pronostique la plus impactante. Quant à la PP (PAS-PAD) elle traduit une altération de la compliance (ou fonction d'amortissement) des gros vaisseaux.

L'HTA est également un composant permettant la détermination du RCVA [Anderson et al[a]. 1991 ; Anderson et al[b]. 1991] mais la pression artérielle est mesurée en clinique, au cabinet du médecin, ce qui altère souvent sa signification pronostique.

Au niveau circulatoire, le processus d'artériosclérose, se traduisant par une atteinte pathologique de la média au niveau des gros et des petits vaisseaux, est un phénomène spécifiquement lié à l'HTA et au vieillissement à l'inverse du processus d'athérosclérose, dans lequel l'HTA n'intervient qu'en tant que facteur de risque.
Au niveau cardiaque, l'HTA favorise la formation de plaques athéromateuses au niveau des gros troncs coronariens, générant ainsi une authentique insuffisance coronarienne organique. Elle contribue également, en collaboration avec divers facteurs neurohormonaux, à l'apparition d'une hypertrophie ventriculaire gauche associant hypertrophie myocytaire et expansion du tissu de soutien.

Les traitements antihypertenseurs ont un effet direct sur la morbidité et la mortalité cardiovasculaire en réduisant nettement le taux de survenue des AVC (de 30 à 40% suivant les populations considérées) et des accidents coronariens (10 à 15 %). Ainsi, l'HAS [Prise en charge des patients adultes atteints d'hypertension artérielle essentielle – Actualisation 2005] préconisait les objectifs tensionnels suivants :

- PAS/PAD < 140/90 mmHg chez l'adulte d'âge moyen ;
- PAS < 150 mmHg chez le sujet âgé avec hypertension systolique ;
- PAS/PAD < 140/80 mm Hg chez le diabétique de type 2 ;
- PAS/PAD < 130/85 mm Hg chez l'insuffisant rénal sans protéinurie ;
- PAS/PAD < 125/75 mm Hg chez l'insuffisant rénal avec protéinurie > 1g/j.

2.1.5. Les dyslipidémies

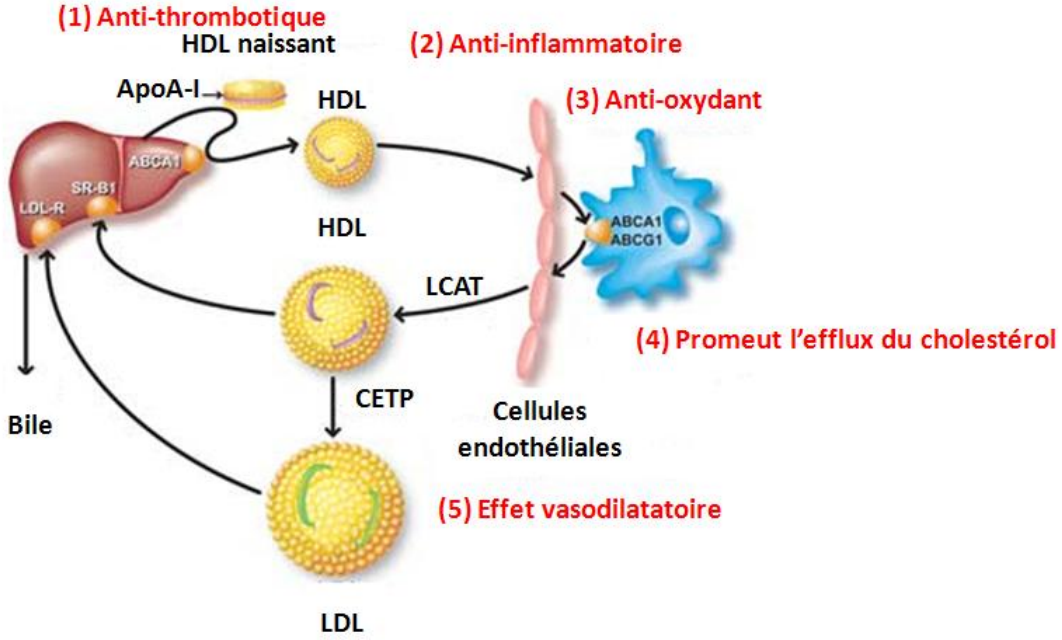
Le cholestérol total et le LDL-cholestérol

Le lien entre hypercholestérolémie et athérosclérose est particulièrement bien établi, essentiellement pour les pathologies coronariennes. L'élévation du cholestérol total et de la low-density lipoprotein cholesterol (LDL-C, forme principale de transport du cholestérol dans l'organisme) est associée à une augmentation curvilinéaire du risque coronarien [Stamler et al. 2008]. Néanmoins, cette relation est très atténuée chez les femmes en raison du poids important que joue la high-density lipoprotein cholesterol (HDL-C) ainsi que chez les sujets âgés de plus de 65 ans comme cela sera exposé par la suite.

Le HDL-cholestérol

Le HDL-C est impliqué dans le transport reverse du cholestérol vers les voies d'élimination (**Figure 2**). C'est donc sa diminution qui constitue un facteur de risque cardiovasculaire comme cela a été démontré dans de nombreuses études épidémiologiques : une augmentation du HDL-C de 0,026 mmol/L s'accompagne d'une diminution du risque coronarien de 2% chez l'homme et de 3% chez la femme [Gordon et al. 1989].

Figure 2 : Transport reverse du cholestérol et autres propriétés du HDL-cholestérol



Source : HDL-C, the changing testing paradigm
Remaley & Warnick, American Association for Clinical Chemistry (AACC)

Il s'agit également d'un facteur dont les interconnexions avec les facteurs de risque classiques des MCV sont connues ; notamment le fait que sa concentration est abaissée par le tabagisme, l'hypertriglycéridémie, et à l'inverse augmentée par la consommation d'alcool et les œstrogènes. Cette dernière caractéristique explique en partie la protection féminine vis à vis du risque coronarien évoquée précédemment et le fait que cette protection féminine disparaisse progressivement avec la ménopause.

Selon les recommandations françaises, une concentration en HDL-C inférieure à 0,9 mmol/L (inférieure à 1,03 mmol/L selon les recommandations américaines) peut être considérée comme un facteur de risque supplémentaire. Il a d'ailleurs été proposé, notamment par l'étude Framingham [Anderson et al[a]. 1991 ; Anderson et al[b]. 1991], d'utiliser les ratios [cholestérol total / HDL-C] ou [LDL-C / HDL-C] pour la prédiction du risque coronarien.

Les Triglycérides

L'élévation des triglycérides (TG), et notamment l'hypertriglycéridémie (> 1,32-1,75 mmol/L) est associée à une augmentation du risque coronarien. Cependant, la relation entre TG et risque coronarien est largement dépendante de l'effet d'autres facteurs de risque qui s'accompagnent souvent d'une élévation de la triglycéridémie : l'obésité, le diabète, l'hypercholestérolémie, la baisse du HDL-C, l'HTA, … etc.

Récemment, le phénotype 'hypertriglycéridémie-fort tour de taille' (hypertriglyceridemic-waist phenotype) a été associé à une détérioration du profil de risque cardiométabolique et à une accentuation du lien entre hypertriglycéridémie et pathologies coronariennes [Arsenault et al. 2010] chez 21 787 participants âgés de 45 à 79 ans issus de la 'European Prospective Investigation of Cancer (EPIC)-Norfolk Prospective Population Study' et suivis sur une dizaine d'années.

2.1.6. Le diabète

Le diabète est fortement lié au risque de maladie coronarienne qu'il multiplie par un facteur 3 chez la femme, et 2 chez l'homme. Ainsi, le risque cardiovasculaire des femmes diabétiques rejoint celui des hommes non diabétiques.

L'étude MONItoring trends and determinants in CArdiovascular disease' (MONICA) menée par l'OMS a pu observer que la mortalité coronarienne était aussi élevée chez un diabétique n'ayant pas fait d'infarctus, que chez un non diabétique ayant déjà fait un infarctus dans une population Finnoise [Miettinen et al. 1998] : la quantification du RVCA donne donc un poids équivalent au diabète et aux antécédents d'infarctus.

Une attention toute particulière devrait être portée à la prévention primo-secondaire du diabète comme cela a récemment été discuté, toujours au sein de populations Finnoises [Barengo en al. 2008] : quatre études de population menées dans les années 70 (1972 et 1977) puis 80 (1982 et 1987) et réunissant un total de 16 779 hommes et 18 235 femmes suivies pendant dix ans ont permis d'observer une augmentation, en une décennie, du

risque relatif de mortalité coronarienne chez les hommes diabétiques comparés aux hommes non diabétiques :

- De 1,67 à 1,75 chez les hommes de 25 à 49 ans ayant un diabète antérieur à leur inclusion dans une cohorte ;
- Et de 1 à 1,92 chez les hommes développant un diabète au cours de leur période de suivi.

2.1.7. L'obésité

Un excès ou une insuffisance de poids est chiffrable par le calcul de l'indice de masse corporelles (IMC) ou rapport [poids / taille²] selon la formule de Quetelet (**Tableau 3**).

Tableau 3 : Classification de l'indice de masse corporelle

Classification	**IMC (kg/m²)**	**Risque de MCV, diabète et HTA**
Maigreur extrême	< 16	Élevé
Maigreur	< 18,5	Accru
Poids normal	18,5 - 24,9	Faible
Embonpoint	25,0 - 29,9	Accru
Obésité, classe 1	30,0 - 34,9	Élevé
Obésité, classe 2	35,0 - 39,9	Très élevé
Obésité, classe 3 (obésité morbide)	≥ 40	Extrêmement élevé

Source : Lignes directrices canadiennes pour la classification du poids chez les adultes
http://www.hc-sc.gc.ca/fn-an/alt_formats/hpfb-dgpsa/pdf/nutrition/weight_book-livres_des_poids-fra.pdf

L'obésité et la surcharge pondérale sont associées à un risque coronarien nettement accru mais ce facteur est en partie dépendant de son impact sur les autres facteurs de risque cardiovasculaire : l'HTA (75% en partie dues à la surcharge pondérale), les dyslipidémies ou le diabète.

Néanmoins, la surcharge pondérale a un effet propre car après ajustement sur les facteurs associés, le risque ne s'annule pas. En effet, Rimm et ses collaborateurs ont démontré que le risque relatif de maladie coronarienne chez les hommes âgés de 40 à 65 ans après cet

ajustement était de 1,72 pour un IMC compris entre 25 et 29 ; de 2,61 entre 29 et 33 ; et de 3,44 pour les obèses ayant un IMC > 33 kg/m², en comparaison avec les hommes ayant un IMC < 23 kg/m² [Rimm et al. 1995]. Au delà de 65 ans l'association entre surcharge pondérale et risque coronarien était cependant très atténuée.

La répartition de l'adiposité est également un facteur de risque en soi. En effet, une répartition de type centrale ou androïde de l'adiposité (au niveau de l'abdomen) majore significativement le risque. Ainsi, le tour de taille (**Tableau 4**), ou le ratio [tour de taille / tour de hanche] constituent des facteurs de risque importants et l'excès d'adiposité abdominale est fortement associé à plusieurs perturbations des facteurs de risque cardiométaboliques [Liu et al. 2010].

Tableau 4 : Classification du tour de taille

Risque de MCV, diabète et HTA	Hommes	Femmes
Faible	< 94 cm	< 80 cm
Accru	> 94 cm	> 80 cm
Considérablement accru	> 102 cm	> 88 cm

Source : Lignes directrices canadiennes pour la classification du poids chez les adultes
http://www.hc-sc.gc.ca/fn-an/alt_formats/hpfb-dgpsa/pdf/nutrition/weight_book-livres_des_poids-fra.pdf

2.1.8. Les facteurs nutritionnels

L'influence majeure des facteurs nutritionnels sur le risque cardiovasculaire, et en particulier coronarien, est démontrée par de nombreux travaux épidémiologiques.

En 1952, une étude comparative portant sur les habitudes alimentaires et les risques de MCV visant à suivre de façon précise l'état de santé de 13 000 individus répartis dans sept pays (les États-Unis, la Finlande, la Hollande, le Japon, l'Italie, la Yougoslavie, et la Grèce) sur plus de vingt ans : la 'Seven Countries Study' [Keys et al. 1984 ; Kromhout et al. 1989], a montré d'importantes différences d'incidence des maladies coronariennes entre

populations, le minimum étant observé au Japon et dans les populations méditerranéennes avec des incidences plus de 10 fois inférieures à celle observée en Finlande. Ces disparités étaient expliquées en partie par une cholestérolémie plus basse. Cependant à cholestérolémie égale, le risque coronarien était beaucoup plus faible dans les pays à faible incidence, soulignant l'influence d'autres facteurs environnementaux dont les facteurs nutritionnels. En particulier, l'étroite corrélation entre consommation en graisses saturées et risque coronarien [Kromhout et al. 1999].

Mais l'importance des facteurs nutritionnels est également observable dans le cas des études de migration comme cela a été le cas pour la 'Ni Hon San Study' dans laquelle il a été démontré qu'une consommation plus importante en protéines animales et en graisses saturées des Japonais ayant immigré à Hawaï était à l'origine de l'incidence moindre des AVC observée dans cette communauté par rapport à celle demeurée au Japon [Takeya et al. 1984]. De telles disparités chez des sujets de différentes ethnicités vivant dans des zones géographiques (ou culturelles) différentes a également été observée par le 'Honolulu Heart Program' (HHP) [Benfante et al. 1992].

2.1.9. La sédentarité

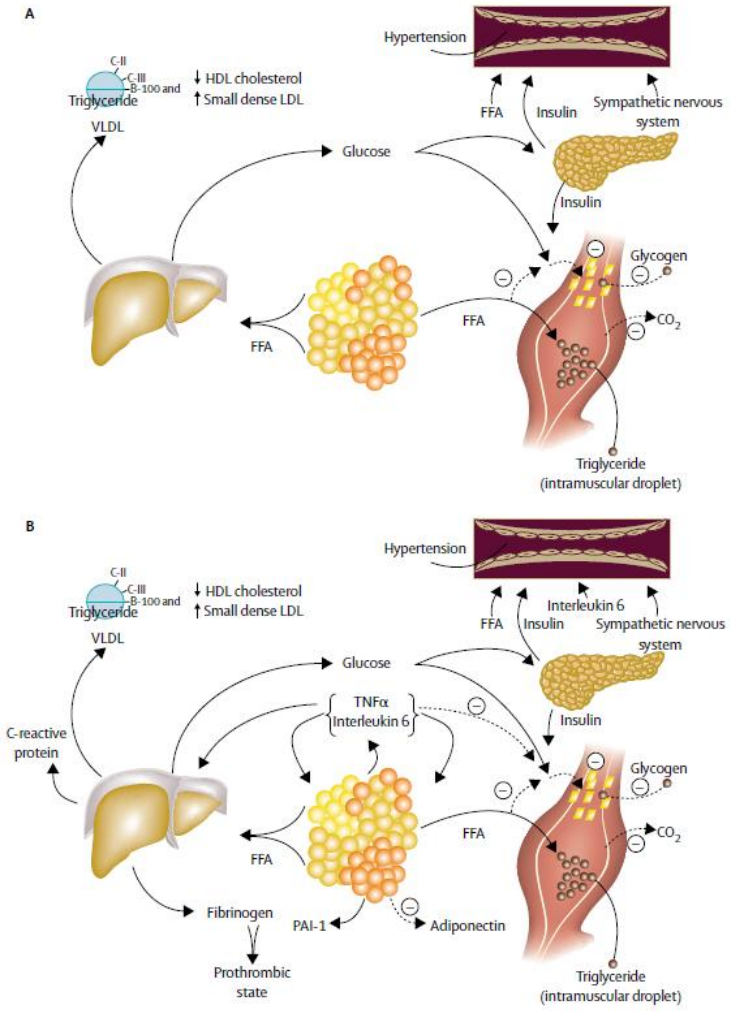
Une méta-analyse de plusieurs études de cohorte a démontré que la sédentarité multipliait par 1,9 le risque de décès d'origine coronarienne après ajustement sur les autres facteurs de risque [Berlin & Goldtiz. 1990]. De même, dans les suites d'un infarctus du myocarde, l'absence d'activité physique est associée à une plus forte mortalité totale et coronaire, par rapport aux patients qui bénéficient d'une réadaptation cardiovasculaire.
En prévention primaire, une activité physique régulière diminue le risque cardiovasculaire tout en modifiant de nombreux facteurs de risque associés [Aijaz et al. 2010] et en post infarctus, un programme de réadaptation cardiovasculaire est préconisé par la Société Française de Cardiologie.

2.1.10. Le syndrome métabolique

Le syndrome métabolique (SM) ou syndrome X est un cluster regroupant diverses anomalies cliniques et biologiques souvent modérées mais dont l'association présente un risque cardiovasculaire majeur [Alberti et al. 2005 ; Eckel et al. 2010] (**Figure 3**). Ce syndrome, qui a été répertorié dans la 10ème révision de la classification internationale des maladies (CIM10) en 1990, n'est cependant pas une pathologie proprement dite mais un état pré-pathologique ou 'bombe à retardement cardiovasculaire' augmentant le risque de diabète de type 2 et de maladie coronarienne [Eckel et al. 2005].

Il en existe plusieurs définitions, proposées notamment par l'OMS et l'European Group for the Study of Insulin Resistance (EGIR) en 1999, le National Cholesterol Education Program's Adult Treatment Panel III (NCEP-ATPIII) en 2001 et l'International Diabetes Federation (IDF) en 2005. Les différenciations de ces définitions portaient sur les seuils de classification des composantes du SM mais un consensus relatif existait sur les facteurs à prendre en compte : une insulinorésistance, une dyslipidémie, une hypertension, et une obésité abdominale (**Tableau 5**).

Figure 3 : Physiopatholgie du Syndrome Métabolique



[Eckel et al. 2005]

A : Les acides gras libres (FFA) sont émis en abondance par une masse étendue de tissu adipeux. Dans le foie, les FFA conduisent à une augmentation de la production du glucose et, des triglycérides et de la sécrétion des VLDLs. Les anomalies des lipides et lipoprotéines associés incluent la diminution de la concentration en HDL-C et une augmentation de la densité des LDLs. Les FFA réduisent également l'insulinosensibilité au niveau musculaire en inhibant l'entrée du glucose médiée par l'insuline. Les défauts associés incluent une diminution de la transformation du glucose en glycogène et une augmentation de l'accumulation des lipides sous forme de triglycérides. L'élévation de la concentration circulante en glucose et dans une certaine mesure des FFA augmentent l'insulinosécrétion qui conduit à une hyperinsulinémie. L'hyperinsulinémie peut conduire à une augmentation de la réabsorption rénale sodique et à une augmentation de l'activité du système nerveus symathique (SNS) et contribuer ainsi au développement de l'hypertension. L'augmentation des FFA circulants peut aussi conduire à l'hypertension.

B: L'effet endocrine et paracrine de l'état proinflammatoire se superpose et contribue à l'IR produite par des FFA en excès. Produites par différentes cellules dans le tissu adipeux qui incluent les adipocytes et les macrophages, l'augmentation de la sécrétion d'IL-6 et de TNF-α, parmi d'autres cytokines, conduit à une aggravation de l'IR et à la lipolyse des stocks de triglycérides qui sont transformés en FFA circulants. L'IL 6 et d'autres cytokines sont également augmentées et peuvent conduire à des productions hépatiques élevées de glucose et des VLDL ainsi qu'à une augmentation de l'IR au niveau musculaire. Les cytokines et les FFA augmentent aussi la production hépatique de fibrinogène et de PAI-1 qui complète la surproduction de PAI-1 par le tissu adipeux. Ceci conduit à un état pro-thrombotique. Des diminutions dans la production de l'adiponectine, une cytokine anti-inflammatoire et insulinosensibilisante, sont également associées au SM et pourraient contribuer à la physiopathologie du SM.

Tableau 5 : Définitions du syndrome métabolique

Phénotypes	OMS (1999)	EGIR (1999)	NCEP-ATPIII (2001)	IDF (2005)
Insulinorésistance	Clamp < Q1	IAJ > Q4	-	-
Intolérance au glucose	GAJ ≥ 6,1 et/ou G2h ≥ 6,1	GAJ ≥ 6,1	GAJ ≥ 6,1	GAJ ≥ 5,6 ou DT2
Dyslipidémie	TG ≥ 1,7 et/ou HDL-C < 0,9 H, 1 F	TG ≥ 2 et/ou HDL-C < 1	TG ≥ 1,7 et HDL-C < 0,9 H, 1 F	TG ≥ 1,7 et HDL-C < 1,03 H, 1,29 F
Hypertension	PAS/PAD ≥ 140/90	PAS/PAD ≥ 140/90	PAS/PAD ≥ 130/85	PAS/PAD ≥ 130/85
Obésité	RTH > 0,90 H, 0,85 F et/ou IMC > 30	TT ≥ 94 H, 80 F	TT ≥ 102 H, 88 F	TT ≥ 94 H, 80 F
Microalbuminurie	EUA > 20 ou Alb/créat ≥ 30	-	-	-
Syndrome Métabolique	IR et/ou ITG + au moins 2 critères	IAJ + au moins 2 critères	Au moins 3 critères	Obésité + au moins 2 critères

IAJ : insulinémie à jeun ; GAJ : glycémie à jeun (mmol/L) ; G2h : glycémie 2 heures après charge glucosée ; TG (mmol/L) HDL-C (mmol/L) ; PAS/PAD (mmHg) ; EUA : excrétion urinaire d'albumine (µg/min)

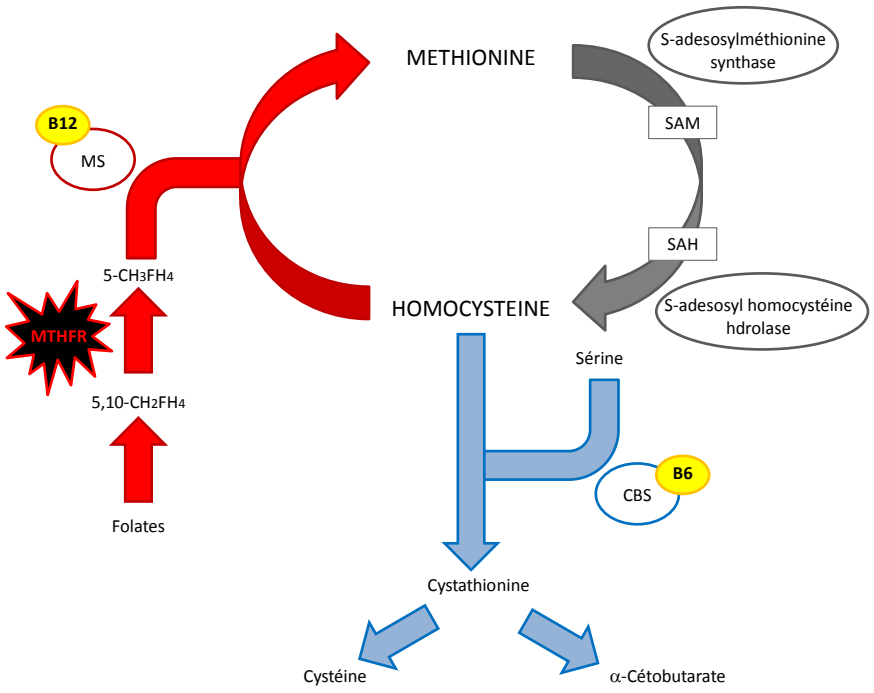
2.2. Les facteurs de risque émergents des maladies cardiovasculaires

Il s'agit de facteurs de risque en cours de validation : l'hyperhomocystéinémie, les facteurs psychosociaux, les facteurs thrombogéniques, les marqueurs de l'inflammation, les molécules d'adhésion, les traitements hormonaux, la fréquence cardiaque, … etc.
Cette liste est donc par définition non exhaustive.

2.2.1. L'hyperhomocystéinémie

L'homocystéine est un acide aminé intervenant dans le métabolisme de la méthionine et des folates. Elle se situe au carrefour de 2 voies métaboliques : la transsulfuration qui dégrade l'homocystéine en cystéine avec l'intervention de la vitamine B6 comme cofacteur ; et la reméthylation qui la retransforme en méthionine avec pour cofacteur la vitamine B12 et les folates (**Figure 4**).

Figure 4 : Métabolisme de l'homocystéine



Physiologiquement, l'homocystéinémie est comprise entre 5 et 15 mmol/L. L'homocystinurie, maladie autosomique récessive rare, associe une hyperhomocystéinémie majeure (concentration 20 fois supérieure à la normale), un retard mental, des anomalies osseuses, une luxation du cristallin et des thromboses artérielles et veineuses précoces. Une hyperhomocystéinémie modérée a également été associée à une élévation du risque athérothrombotique mais sa quantification n'a pas pour autant été préconisée en prévention primaire [Ducros et al. 2009].

Des études d'intervention visant à supplémenter l'alimentation en folates seuls ou en association aux vitamines B6 ou B12 ont pourtant montré leur efficacité à abaisser la concentration sérique en homocystéine et en cholestérol total chez les enfants [Papandreou et al. 2010] et à réduire le risque coronarien des sujets âgés ayant un déficit en vitamines B12 [Kurt et al. 2010].

2.2.2. Les facteurs psychosociaux et environnementaux

Le risque cardiovasculaire des catégories sociales défavorisées est plus élevé pour des raisons d'ordre à la fois culturel, économique et nutritionnel ; ce qui explique la complexité de mener de telles études car elles sont soumises à de nombreux facteurs de confusion. Les bénéfices éventuels d'une prise en charge de facteurs psychosociaux sont donc difficiles à évaluer.

Du point de vue environnemental, il a été observé, notamment par l'étude MONICA, que la température journalière moyenne était inversement proportionnelle au taux d'infarctus et que la pollution atmosphérique avait également un impact cardiovasculaire [Bener et al. 2009].

2.2.3. Les facteurs thrombogéniques

Un grand nombre de facteurs prothrombotiques ont été individualisés au cours de ces dernières années : en particulier le fibrinogène et le facteur VII (FVII) ont été désignés comme des facteurs de risque d'infarctus du myocarde indépendants du cholestérol et du tabac, la viscosité sanguine, le taux de globules blancs et la concentration en facteur de von Willebrand, et l'inhibiteur de l'activateur du plasminogène 1 (PAI-1).

2.2.4. Les marqueurs de l'inflammation

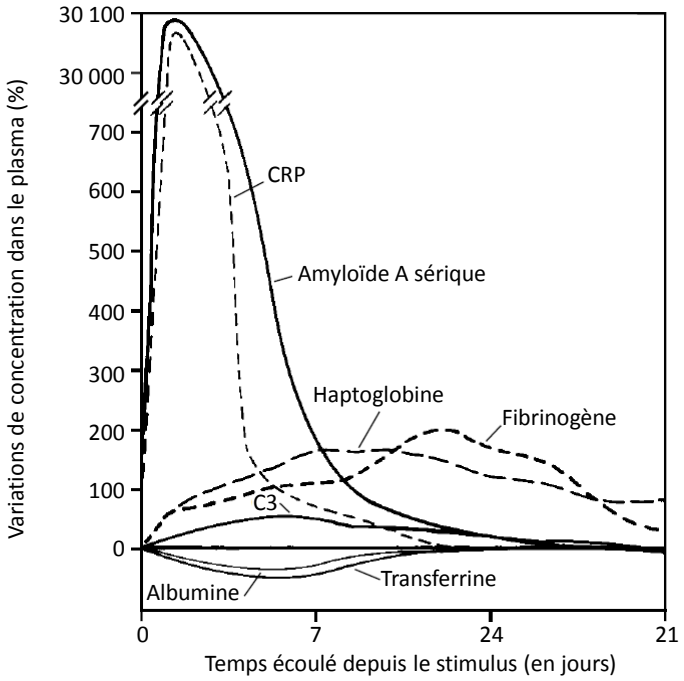
Il a clairement été établi que le processus inflammatoire se situait au cœur de nombreuses pathologies et l'opportunité d'utiliser la quantification des marqueurs de l'inflammation, notamment les protéines de la phase aiguë ou Protéines de la Réaction Inflammatoire (PRI) (**Figure 5**), pour identifier les sujets à risque cardiovasculaire en prévention primaire a été discutée. Les recommandations de la 'National Cholesterol Education Program Adult Treatment Panel III' [Expert Panel on Detection, Evaluation, and Treatment of High Blood Cholesterol in Adults. 2001] exposaient notamment (1) le manque de standardisation de leur quantification ; (2) de résultats cohérents issus des études épidémiologique prospectives et (3) le fait que la prise en compte de tels marqueurs ne modifiait pas les scores prédictifs préexistants. La Vème conférence de 'l'American Heart Association' (AHA) [Pearson et al. 2004] déplorait quant à elle en 2004 le manque de consensus professionnel ou gouvernemental sur la manière d'intégrer la quantification de tels marqueurs dans la pratique clinique.

Néanmoins, l'implication des marqueurs de l'inflammation (protéine C-réactive ou CRP, quantification des leucocytes, … etc.) dans la médiation de tous les stades du processus d'athérosclérose (**Figure 6**) [Libby et al. 2002] en commençant par le stade infraclinique et surtout leur forte implication dans les effets pléiotropiques des thérapies hypolipémiantes [Albert et al. 2001 ; Liao. 2002 ; Topol. 2004] suggèrent un effet bénéfique de leur incorporation dans les stratégies de prise en compte du risque cardiovasculaire.

Traditionnellement, le fait qu'une augmentation de la CRP majore le risque cardiovasculaire est connu : il s'agit d'un marqueur et d'un activateur de l'inflammation qui interagit avec les

autres cytokines au niveau de l'endothélium et favorise le développement des plaques athéromateuses. Elle est également un marqueur de la présence de nombreux éléments inflammatoires au sein de la plaque, caractéristique connue pour favoriser sa rupture. C'est donc un facteur de risque entré depuis déjà quelques temps dans la pratique courante.

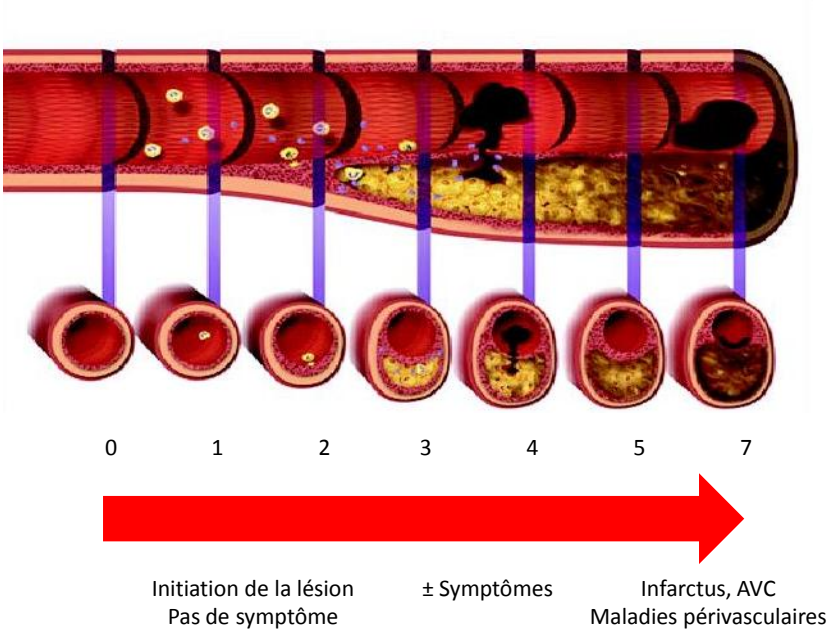
Figure 5 : Profils de variation de concentration plasmatique des protéines de la phase aiguë de l'inflammation à la suite d'un stimulus modéré



[Gabay & Kushner. 1999]

Les protéines de la phase aiguë de l'inflammation sont définies empiriquement comme des protéines dont la concentration plasmatique évolue au cours d'une réaction inflammatoire. Les Protéines dont la concentration plasmatique augmente sont appelées Protéines de la Réaction Inflammatoire positives (PRI+) et celles dont la concentration plasmatique diminue Protéines de la Réaction Inflammatoire négative (PRI-). Chez l'Homme, les PRI+ sont l'α1 glycoprotéine acide, l'α1 protéase inhibiteur, l'α1 antichymotrypsine, l'haptoglobine, la céruloplasmine, la CRP et l'Amyloïde A protéine. Les principales fonctions des PRI résultent d'une interaction des protéines avec des ligands d'origine diverse, et de la formation de complexes dont l'épuration est assurée soit par le système réticulo-endothélial, soit par l'hépatocyte. Il s'ensuivra l'inactivation des protéases, la neutralisation de molécules toxiques tels l'hémoglobine et les ions superoxides, l'évacuation de débris membranaires, leucocytaires ou cellulaires et de débris nucléaires telle la chromatine. La chute de la concentration plasmatique des PRI au cours d'une réaction inflammatoire, entraîne une augmentation de la fraction libre de certains ligands (acides gras, hormones, vitamines, oligoéléments).

Figure 6 : Stades d'évolution de la plaque d'athérosclérose. De la strie lipidique à la rupture de plaque



[Libby et al. 2002]

Mais d'autres marqueurs de l'inflammation ont récemment été proposés en prévention cardiovasculaire. L'haptoglobine notamment est une molécule de la phase aiguë de l'inflammation (**Figure 4**) synthétisée par le foie et dont une diminution drastique est un indice d'hémolyse pathologique, d'insuffisance hépatique, ou de maladie rénale. Cette protéine, comme l'ensemble des protéines de la phase aiguë, est régulée par des cytokines [Gabay & Kushner. 1999] et a pour propriété principale de former un complexe avec l'hémoglobine lors de son relargage par les érythrocytes en vue de son élimination dans le foie. Le taux d'haptoglobine a été lié au développement de maladies infectieuses telles que la malaria, la tuberculose et l'infection au virus de l'immunodéficience humaine (VIH) [Carter et al. 2007]. L'haptoglobine est composée de deux paires de chaînes α et β et est présente sous trois formes phénotypiques associées à diverses pathologies : Hp1-1, Hp2-2 et Hp2-1 [Koda et al. 1998]. Il a notamment été démontré que des sujets diabétiques porteurs du

phénotype Hp2-2 avaient un risque cardiovasculaire quatre à cinq fois plus élevé [Asleh et al. 2008].

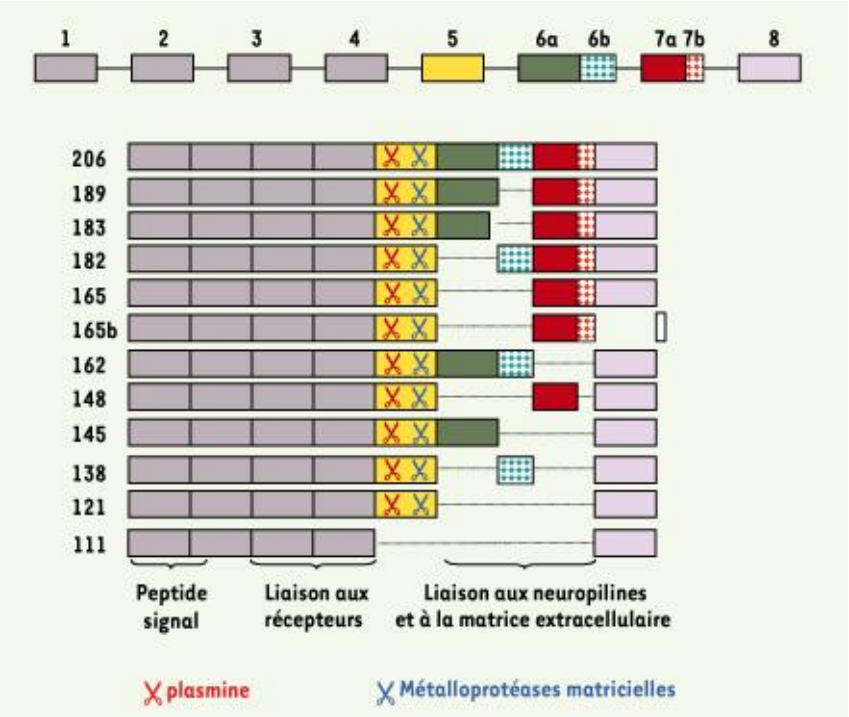
Récemment, l'haptoglobine a été liée, à la suite d'une étude prospective sur près de 12 ans, à l'infarctus du myocarde, aux AVC et à l'insuffisance cardiaque chez 342 125 sujets issus de l'Apolipoprotein Mortality RISk study (AMORIS) [Holme et al. 2010]. Dans cette même population, l'importance de l'incorporation des marqueurs communs d'inflammation, dont l'haptoglobine, dans les stratégies de prédiction d'événements cardiovasculaires [Holme et al. 2010] a permis de rouvrir le débat sur l'opportunité d'utiliser de tels marqueurs en prévention.

2.2.5. Les molécules d'adhésion

Le VEGF est l'un des principaux facteurs contrôlant le développement et le maintien de l'arbre vasculaire chez l'homme (angiogenèse) avec le Fibroblast Growth Factor (FGF). Ce facteur est principalement associé au développement tumoral et métastatique [Carmeliet & Jain. 2000] mais il est également au centre de nombreuses pathologies [Ferrara et al. 2003] dont les MCV. En effet, des taux élevés de VEGF circulants ont été observés dans les cas de pathologies ischémiques [Hojo et al. 2000 ; Blann et al. 2002] et des thérapies pro-angiogènes ciblées entre autre sur ce type de pathologie et utilisant le VEGF sont en phase d'essais cliniques [Buysschaert et al. 2007]. Des taux élevés de VEGF ont également été associés à l'insuffisance cardiaque [Chin et al. 2002], aux AVC [Slevin et al. 2000] ainsi qu'au diabète [Wirostko et al. 2008].

Le gène codant pour le VEGF contient 8 exons dont un épissage alternatif aboutit à l'existence de plusieurs isoformes (**Figure 7**). Les variants $VEGF_{206}$ et $VEGF_{189}$ sont peu biodisponibles en raison de leur liaison élevée à la matrice extracellulaire, alors que le variant $VEGF_{121}$ est actif et librement diffusible. Le variant $VEGF_{165}$, également diffusible, a des propriétés de liaison intermédiaires.

Figure 7 : Variants du VEGF. Description de l'utilisation des exons dans les différents variants du VEGF, et de la fonction des séquences qu'ils codent



[Lambert et al. 2008]

Les exons 1 et 2 codent pour le peptide signal, impliqué dans la sécrétion de la protéine dans le milieu extracellulaire, les exons 3 et 4 pour les sites de liaison aux récepteurs et pour un site de glycosylation, et les exons 6 et 7 pour les séquences de liaison à la matrice extracellulaire et aux neuropilines.

2.3. La composante génétique des maladies cardiovasculaires

PUBLICATION N°1

La physiopathologie des MCV est donc extrêmement complexe et implique des interconnexions entre facteurs de risque très divers, dont certains sont encore méconnus. Mais à cette complexité phénotypique s'ajoute une composante génétique très importante et c'est ces interactions gène-gène-environnement qui sont à l'origine de la survenue de pathologies cardiovasculaires (**Figure 8**).

La physiopathologie cardiovasculaire comporte une part d'héritabilité génétique très importante (**Tableau 6**) et la recherche des variants impliqués dans cette héritabilité est dans un premier temps passée par une approche gène candidat [Casas et al. 2006].

Figure 8 : Modèle d'interactions gène-gène-environnement impliquées dans la physiopathologie des maladies cardiovasculaires

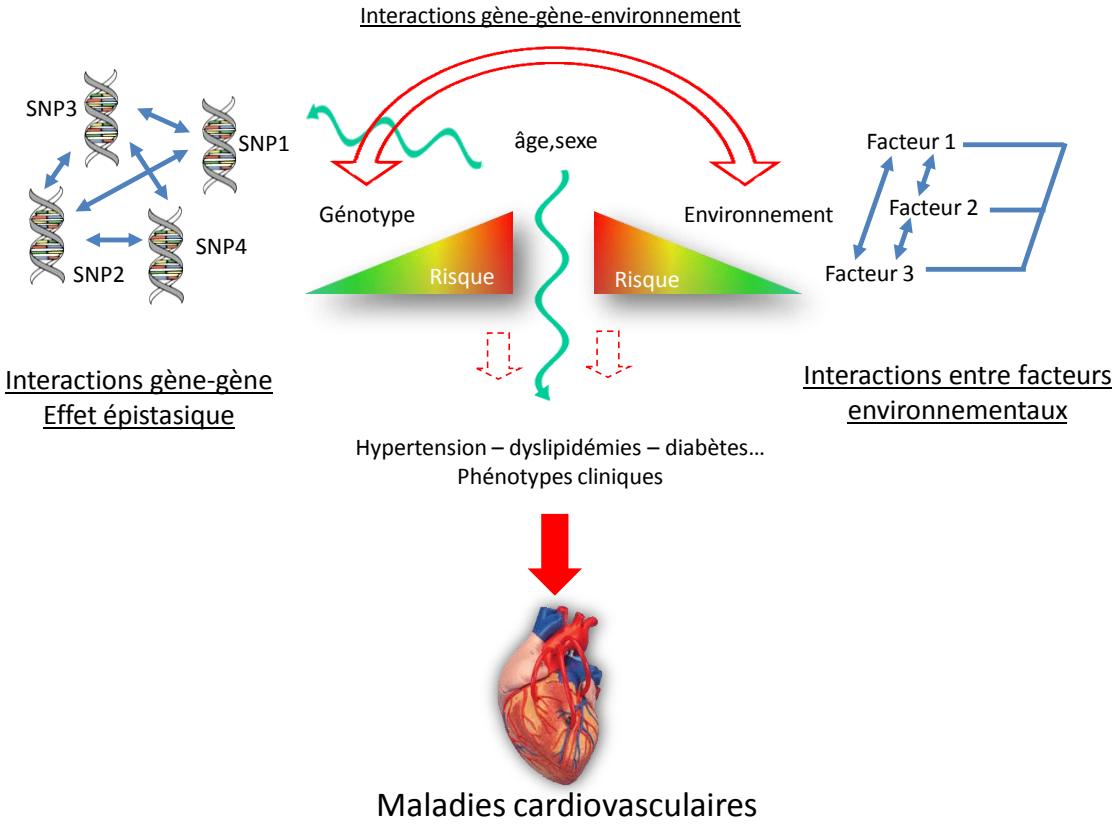


Tableau 6 : Estimation de l'héritabilité des accidents vasculaires cérébraux, des affections coronariennes et des principaux facteurs de risque cardiovasculaire

Traits	Héritabilité	Références
Diabète de type 2 Glycémie Insulinémie	24-85%	Poulsen et al. 1999 Hanson et al. 2001 McQueen et al. 2003 Mills et al. 2004 Souren et al. 2007
Syndrome métabolique	38-61%	McQueen et al. 2003 Bayoumi et al. 2007
Obésité IMC	31-94%	O'Connell et al. 1988 Brown et al. 2003 McQueen et al. 2003 Mills et al. 2004 Souren et al. 2007
Hypertension Pression artérielle	17-49%	Brown et al. 2003 McQueen et al. 2003 Mills et al. 2004
Cholestérol total	32-75%	O'Connell et al. 1988 Heller et al. 1993 Mills et al. 2004 Souren et al. 2007
Triglycérides	28-72%	Heller et al. 1993 McQueen et al. 2003 Mills et al. 2004
LDL-cholestérol	43-78%	O'Connell et al. 1988 Mills et al. 2004 Souren et al. 2007
HDL-cholestérol	42-79%	O'Connell et al. 1988 Heller et al. 1993 McQueen et al. 2003 Mills et al. 2004 Souren et al. 2007
Apolipoprotéine B	51-78%	Heller et al. 1993
Homocystéine	0-66%	Reed et al. 1991 Berg et al. 1992 Cesari et al. 2000 Siva et al. 2007 Bathum et al. 2007 Nilsson et al. 2009
FVII	38-40%	Freeman et al. 2002 Vossen et al. 2007
PAI-1	28%	Freeman et al. 2002
Affections coronariennes Athérosclérose	5-84%	Zdravkovic et al. 2002 Fischer et al. 2005 Wienke et al. 2005 Cassidy-Bushrow et al. 2007
Accidents vasculaires cérébraux	17-32%	Bak et al. 2002

Au sein de l'EA4373, l'héritabilité génétique de nombreux traits liés à ces affections a pu être quantifiée :

- La molécule d'adhèsion intracellulaire 1 (ICAM-I), la E-Sélectine et la P-Sélectine [Pontheiux et al. 2003 ; Ponthieux et al. 2004 ; Marteau et al. 2008] ;
- L'interleukine 6 (IL-6) et le facteur de nécrose tumorale alpha (TNFα) [Haddy et al. 2003 ; Haddy et al. 2005] ;
- L'IL-8, la protéine chimiotactique 1 de Monocyte (MCP-1), le facteur de croissance épidermique (EGF) et le VEGF [Berrahmoune et al. 2006 ; Berrahmoune et al. 2007[b]] ;
- L'insuline [Maumus et al. 2007] ;
- Et la CRP [Berrahmoune et al. 2007[c]].

De nombreux variants génétiques associés, au stade infraclinique, à la variabilité interindividuelle de divers facteurs de risque cardiovasculaire ont également pu être identifiés. Notamment les gènes *APOE* [Schiele et al. 2000 ; Siest et al. 2000], *ESELE* et *PSELE* [Pontheiux et al. 2004, Marteau et al. 2008], *LPL* [Sass et al. 2000 ; Herbeth et al 2007], *PON* [Vincent Viry et al. 2003], *CETP* [Pallaud et al. 2001], *ACE* [Tiret et al. 1992], *MPO* [Leninger-Muller et al. 2003], *ICAM1*, *IL6* et *TNFA* [Haddy et al. 2003], *IL1* [Berrahmoune et al. 2007[a]], *FGB* [Maumus et al. 2007], *VEGF* et *EGF* [Berrahmoune et al. 2007[b]], *CYP2C19* [Bertrand-Thiebault et al. 2008] et *F7* [Sass et al. 2004].
Cette approche conventionnelle a aussi permis de mettre en évidence les interactions gène-environnement, notamment l'effet des allèles *ε2* et *ε4* sur la concentration de l'Apo-Lp-non-B chez les femmes sous contraceptifs oraux [Starck et al. 2001] ; de l'allèle *APOC3-1100C* sur la concentration en apolipoprotéines exclusivement chez les femmes non-fumeuses et sur l'augmentation des taux d'Apolipoprtéine A-I (ApoA-I) et d'HDL-C uniquement chez les hommes [Pallaud et al. 2001] ; l'influence du polymorphisme *L/F554* de la *SELE* sur la PAS dépendante de l'IMC [Marteau et al. 2004] ; et l'interaction entre le polymorphisme *T3801C* du cytochrome 1A1 et le *G1661A* du AHR avec le tabac sur la pression artérielle [Gambier et al. 2006].

Néanmoins, à l'heure actuelle, il est admis que cette approche basée sur une hypothèse préalable ne permettra pas à elle seule d'élucider la génétique du risque cardiovasculaire.

L'approche pangénomique, qui a vu le jour en 2007, a permis d'étendre la détermination des variants associés au risque cardiovasculaire à des voies métaboliques inédites [Manolio et al. 2008 ; Hardy et al. 2009 ; Manolio et al. 2010]

Ainsi, plus de 650 GWAS ont été menées sur diverses pathologies dont 16 sur la survenue d'évènements cardiovasculaires (http://www.genome.gov/gwastudies). Ces dernières ont permis de mettre en évidence 29 variants génétiques sur 20 gènes. Néanmoins, différentes voies métaboliques peuvent mener à des pathologies cardiovasculaires communes d'où l'opportunité d'effectuer des recherche sur des traits quantitatifs associés à ces voies métaboliques (lipides, pression artérielle, … etc.) [Ouwehand et al. 2009].

C'est ainsi que le gène *FTO* a été fortement associé à l'obésité et au diabète dans les populations à la fois adultes et pédiatriques [Meyre et al. 2009 ; Meyre et al. 2010]. Ce dernier phénotype a également été associé à 22 variants localisés à proximité des gènes *CDKN2A, CDKN2B, HHEX, IGF2BP2, KCNJ11, CDKAL1, TCF7L2, THADA, TSPAN8, LGR5, VEGF, NOTCH2, ADAM30, PPARG, SLC30A8, JAZF1, SYN2, CDC, ADAMTS9, CDC123,* et *CAMK1D* [Zeggini et al. 2007 ; Zeggini et al. 2008].

Les études pangénomiques effectuées sur les lipides et la pression artérielle représentent respectivement 26 et 10% de la totalité des études. Elles ont permis d'associer 97 variants dans 54 loci au métabolisme lipidique à ce jour et le gène *CETP*, précédemment connu par l'approche gène candidat, est robustement répliqué en association à la fois avec le cholestérol total, le LDL-C et les TG.

À l'inverse, les études sur l'HTA ce sont avérées peu conclusives avant la mise en place de très larges études. Les Consortia 'Cohorts for Heart and Aging Research in Genomic Epidemiology' (CHARGE) [Newton-Cheh et al. 2009] et GlobalBPgen [Levy et al. 2009] ont récemment permis de mettre en évidence 14 variants associés aux traits quantitatifs pression artérielle ou à l'HTA.

Ainsi, les GWAS ont permis d'étudier les prédispositions aux MCV sous un jour nouveau, mais l'énigme de l'héritabilité manquante perdure [Maher et al. 2008 ; Manolio et al. 2009] et de nombreuses limitations méthodologiques restent à solutionner, voire à contourner.

CHAPITRE 3 – POPULATIONS ET MÉTHODES

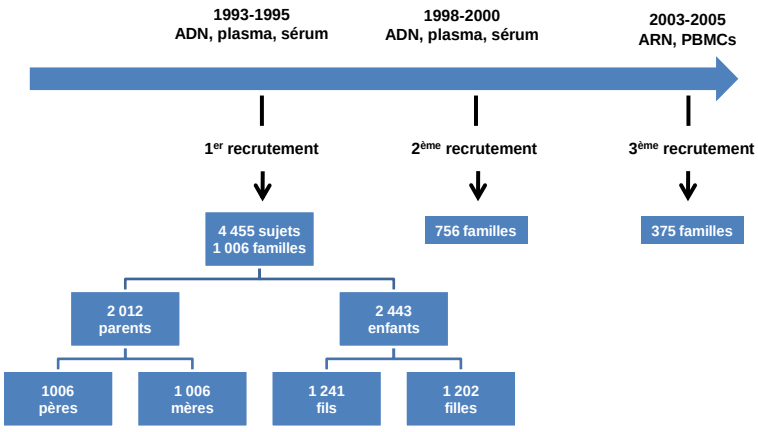
1. Populations

1.1. La cohorte STANISLAS

Initiée au CMP de Vandœuvre-lès-Nancy en 1993, la cohorte STANISLAS (Suivi Temporaire Annuel Non Invasif de la Santé des Lorrains Assurés Sociaux) [Siest et al. 1998 ; Visvikis-Siest et al. 2008] est une cohorte familiale visant à l'origine à étudier les composantes génétiques et environnementales des facteurs de risque des MCV. Les familles STANISLAS, originaires de la Meurthe-et-Moselle, ont été suivies de 1993 à 2005 à l'occasion de 3 visites médicales tous les cinq ans au CMP sur invitation de la Caisse primaire d'assurance maladie (CPAM).

Le recrutement initial a eu lieu de 1993 à 1995 (t_0) et a permis d'inclure 1 006 familles biparentales (4 455 sujets) d'origine française avec au moins deux enfants de plus de six ans. Tous les individus inclus étaient supposées sains, c'est-à-dire exempts de maladies aiguës et/ou chroniques. 75% des familles initialement recrutées ont participé à la seconde visite (1998-2000, t_5) et 375 familles à la troisième (2003-2005, t_{10}) (**Figure 9**).

Figure 9 : Phases de recrutement de la cohorte STANISLAS



Selon les lois bioéthiques en vigueur, chaque participant a donné par écrit son consentement éclairé, approuvé par le Comité local d'éthique de Nancy.
L'ensemble des données recueillies permettaient donc d'effectuer des études transversales ou longitudinales sur une période de 10 ans.

Lors des examens de santé, un personnel qualifié a effectué le recueil des données :

- Des médecins cardiologues pour les échographies cardiaques ;
- Des médecins du CMP pour la détermination du développement pubertaire ;
- Des infirmières du CMP pour les mesures de la pression artérielle, de la fréquence cardiaque, du tour de taille, du tour de hanches, de l'épaisseur des plis cutanés, et le l'ostéodensitométrie ;
- Des opératrices du CMP pour la mesure du poids, de la taille et de l'impédance.

1.1.1. Questionnaires alimentaires et de mode de vie

Des questionnaires standardisés en fonction de l'âge des participants (adultes/enfants/adolescents) ont été administrés lors des examens de santé. Ils ont notamment concernés les antécédents familiaux de MCV, la consommation médicamenteuse (répartie par classes de médicaments), les loisirs et l'activité physique, la consommation d'alcool et de tabac, … etc.
Selon le type de questionnaire, ils ont été auto-administrés ou renseignés avec l'aide de personnel qualifié.

1.1.2. Données cliniques et anthropométriques

Les données anthropométriques : poids (balance électrique d'une précision de 200g), taille (toises d'une précision de 0,1 cm), tour de taille, tour de hanche ont été mesurées sur les sujets légèrement habillés et déchaussés. Elles ont permis de calculer l'IMC selon la formule de Quetelet et le ratio [taille / hanche].
Ces mesures ont été réalisées par des infirmières habilitées et la fiabilité des appareils utilisés a été périodiquement contrôlée tout au long de l'étude.
Les différents stades pubertaires ont été évalués chez les adolescents par des médecins selon la classification de Tanner datant de 1962.

La pression artérielle a été mesurée à trois reprises avec 5 minutes d'intervalle sur le bras gauche en position couchée avec un spygmomanomètre ou tensiomètre à mercure. La valeur retenue était la moyenne des trois mesures consécutives.

1.1.3. Impédance bioélectrique

L'impédance bioélectrique a été mesurée dans un sous-échantillon à jeun sans activité physique préalable (à la suite d'une position couchée pendant 10 minutes). La résistance totale du corps a été mesurée à l'aide de l'analyseur d'impédance BIA-101 (RJL Systems, Detroit, MI, USA) par une personne qualifiée.
Les équations établies par Segal et collaborateurs [Segal et al. 1988] ont été utilisées pour calculer la masse maigre et la masse grasse a été estimée selon la formule suivante :

Masse grasse (kg) = Poids (kg) – Masse maigre (kg)

1.1.4. Dosages biologiques

Les prélèvements sanguins ont été effectués sur les participants à jeun depuis 12 heures au niveau d'une veine antécubitale, en position allongée. Deux types de tubes de sang modèle Vacutainer® (Becton Dickinson, Grenoble, France) ont été prélevés :

- Des tubes secs (sans anticoagulant) contenant un gel séparateur de sérum pour l'extraction de molécules contenues dans le sérum. Ces tubes ont été centrifugés à 1000×g pendant 15 minutes à température ambiante. Une partie du sérum ainsi récupéré était immédiatement analysée pour les dosages biochimiques et une autre partie stockée sous forme de paillettes dans l'azote liquide à – 196°C. Ce tube surnuméraire a servi au dosage ultérieur de l'insuline, de l'Apolipoprotéine C-III (ApoC-III) et des molécules d'adhésion.
- Des tubes d'acide Ethylenediaminetetraacetic (EDTA) pour ADN et plasma (où l'IL-6 et le TNFα sont dosés). Après avoir centrifugé le sang frais total recueilli dans des tubes EDTA à 1000×g à température ambiante pendant 10 minutes, le buffy coat formé entre le plasma et les globules rouges a été prélevé puis fractionné en paillettes stockées à – 196°C. Le plasma a également été conservé dans des paillettes à – 196°C.

Les concentrations sériques en cholestérol total, TG et glucose ont été enzymatiquement déterminées grâce à un automate AU 640 (Olympus, Rungis, France).

La concentration sérique en HDL-C a été mesurée dans le surnageant par la même technique enzymatique que pour le cholestérol total sur un automate COBAS-Mira analyser (Roche Diagnostics, Bâle, Suisse) après précipitation des lipoprotéines contenant l'ApoB (chylomicrons, VLDL et LDL) par phosphotungstate et chlorure de magnésium (méthode PTA).

Les concentrations en LDL-C ont été calculées selon la formule de Friedewald [Friedewald & Fredrickson. 1972] lorsque la concentration en TG était inférieure à 4,56 mmol/L et en l'absence de chylomicrons :

En g/L	LDL-C = (cholestérol total) – (HDL-C) – TG/5
En mmol/L	LDL-C = (cholestérol total) – (HDL-C) – TG/2,18

Les concentrations sériques en ApoA-I et en apolipoprotéine B (ApoB) ont été dosées par immunonéphélométrie sur un automate Behring Nephelometer Analyser II ou BNII (Dade Behring, Marburg, Allemagne).

Les concentrations sériques en ApoC-III et en apolipoprotéine E (ApoE) ont été mesurées par immunoturbidimétrie sur un automate COBAS-Mira (Roche Diagnostics, Bâle, Suisse).

Les activités enzymatiques de la gamma-glutamyltransférase (GGT), de la phosphatase alcaline (PAL), de l'aspartate aminotransférase (AST) et de l'alanine aminotransférase (ALT) ont été enzymatiquement déterminées avec l'automate AU 640 (Olympus Merck, Darmstadt, Allemagne).

Les concentrations sériques en CRP, haptoglobine, orosomucoïde ont été mesurées par immunonéphélométrie sur un automate BNII (Dade Behring, Marburg, Allemagne).

Les concentrations sériques en E, L et P-sélectines et en ICAM-I ont été dosées grâce à des kits commercialisés (R&D System, Abington, Royaume-Uni) par la technique ELISA (Enzyme-Linked Immunosorbent Assay) qui utilise une méthode immuno-enzymatique de type 'sandwich'.

Les concentrations plasmatiques en IL-6 et en TNFα ont été déterminées grâce à des kits ELISA (R&D System, Abington, Royaume-Uni).

Les concentrations sériques en insuline ont été dosées grâce à un kit de dosage d'immuno-enzymologie microparticulaire (Microparticular Enzymatic ImmunoAssay ou MEIA) sur un automate IMx (Abbott laboratories, Abbott Park, IL, USA). Il s'agit d'une méthode de type 'double sandwich' permettant d'éviter d'éventuelles réactions croisées avec la proinsuline, le peptide C et le glucagon. La réaction est basée sur la formation d'un complexe entre l'insuline humaine contenue dans le sérum et des microparticules recouvertes d'anticorps anti-insuline. Par la suite, un conjugué d'anticorps anti-insuline couplé à la PAL est ajouté et se lie sur le complexe antigène-anticorps. Enfin, l'addition d'un substrat, le phosphate de méthyl-4-ombelliférone, permet d'observer l'apparition d'un produit fluorescent mesuré 448 nm. Des lavages de la matrice sont effectués entre chaque étape afin d'éliminer tout ce qui n'est pas fixé.

Les caractéristiques du test étaient les suivantes :

- Concentration moyenne en insuline : 7,19 ± 4,98 µU/ml (51,6 ± 35,7 pmol/l) ;
- Imprécision déterminée sur un pool de sérum de 20 sujets : 2,72% ;
- Coefficients de variation (CV) = écart-type/moyenne×100 :
    - Concentrations basses d'insuline (7,58±0,37 µU/ml) : CV=4,88% ;
    - Concentrations moyennes d'insuline (38,38±1,97 µU/ml) : CV=5,13% ;
    - Concentrations élevées d'insuline (120,23±5,38 µU/ml) : CV=4,47%.

1.1.5. Les données génétiques

L'ADN a été extrait à partir du buffy coat congelé à – 196°C par la méthode de 'salting out' [Miller et al. 1988] : seul le lysat cellulaire obtenu par une RNAse (la protéine K) a été traité. L'ADN extrait a ensuite été solubilisé dans un tampon TE (Tris-EDTA).

Plusieurs techniques ont permis la détermination des polymorphismes au sein de la cohorte STANISLAS : des techniques Multiplex et un GWS.

La technique Multiplex en collaboration avec Roche Molecular Systems :

Cette technique Multiplex a été développée en collaboration avec Roche Molecular Systems (Alameda, CA, USA) et a permis la mise au point de trois kits ciblant les gènes candidats impliqués dans différentes voies métaboliques liées à la physiopathologie des MCV [Cheng et al. 1998] : le métabolisme lipidique, la pression artérielle, la coagulation, l'adhésion cellulaire, le métabolisme de l'homocystéine, et pour le Multiplex III, l'inflammation uniquement.

Il s'agit d'une technique où plusieurs séquences cibles de gènes d'intérêt sont amplifiées simultanément, au cours d'une même réaction de PCR au moyen d'amorces biotinylées en 5' spécifiques aux séquences des gènes. L'étape de détection fait appel à un dot-blot inverse. Les produits de PCR sont hybridés à des sondes oligonucléotidiques complémentaires, conjuguées en 5' à la sérum albumine bovine (BSA), ce qui permet leur immobilisation en rangées sur une bandelette (ou strip) recouverte de nylon. La détection colorimétrique des

polymorphismes fait intervenir le système biotine/streptavidine-peroxydase de raifort. La concentration des sondes est définie de telle sorte que les génotypes homozygotes et hétérozygotes apparaissent avec des intensités différentes lors de la réaction colorimétrique.

Les kits Multiplex développés avec Roche contiennent le matériel suivant :

- Thermocycleur : Gene Amp PCR System 9600 (Perkin-Elmer Cetus);
- Plaques à PCR 96 puits;
- Bain-Marie rotatif : Hot Shaker plus, Bellco, Vineland
- Réactifs : SSPE 20X, SDS 20%, Citrate 20X, SA-HRP (streptavidine – horseradish peroxidase), substrat A (H_2O_2 0,01% dans une solution de citrate), substrat B (3,3',5,5' tétraméthylbenzidine 0,1% dans du diméthylformamide 4X), eau stérile, eau ultra-pure, huile de paraffine
- Appareil photo MP-4
- Réactifs de PCR :
    - Multiplex II : Tampon Pré-mix, Primer A, Primer B
    - Multiplex III : Tampon Pré-Mix, Primer

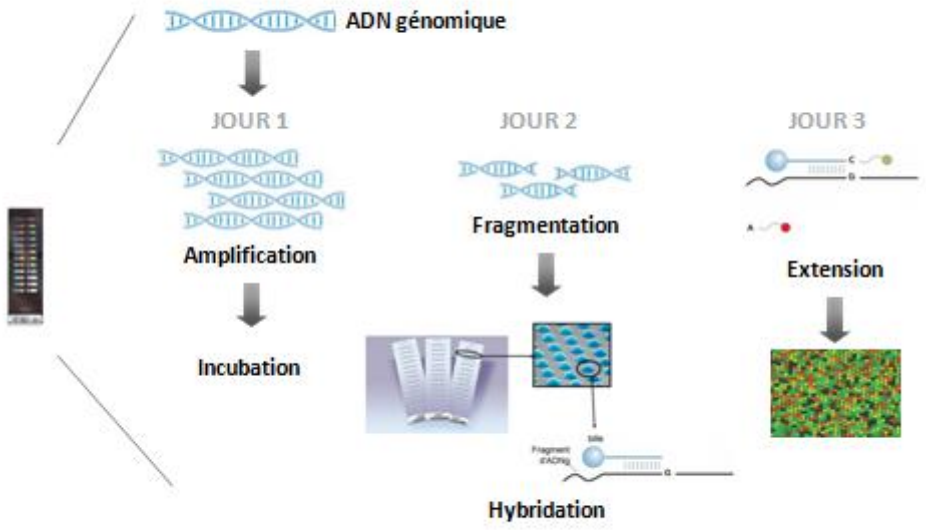
Le typage pangénomique ou genome-wide scan :

Le GWS a été effectué sur un échantillon de 631 enfants non apparentés de la cohorte STANISLAS par une puce Illumina Human CNV370-Duo selon les protocoles prévus par Illumina (**Figure 10**) [Bouatia-Naji et al. 2009].

2 552 SNPs ayant un fort déséquilibre de liaison (LD, $p < 0,001$), un faible taux de génotypage (call rate < 95%) ou une fréquence de l'allèle mineur (MAF) insuffisante (< 1%) on été exclus lors du contrôle qualité. Nous avions au final 318 237 variants génétiques (SNPs).

Ces typages ont été effectués dans le cadre d'une collaboration avec l'UMR CNRS 9199, Institut Pasteur de Lille dirigé par le Professeur Philippe Froguel.

Figure 10 : Modalités du typage pangénomique utilisant la puce Illumina Human CNV370-Duo

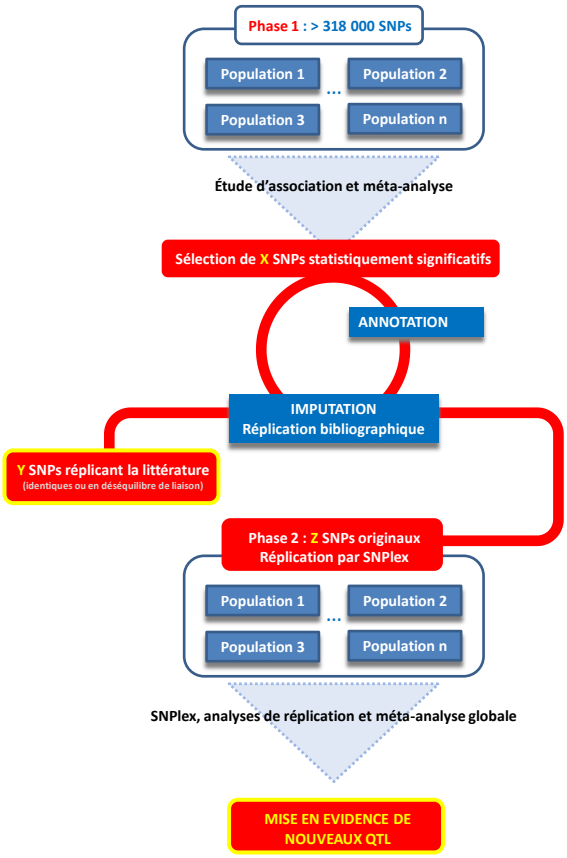


http://snpcenter.grcf.jhmi.edu/downloads/HumanCNV370_DataSheet.pdf

Le typage spécifique des variants génétiques issus de la phase initiale de genome-wide association study en vue des analyses de réplication :

La spécificité des GWAS réside dans le fait que de telles études nécessitent plusieurs phases d'expérimentation (**Figure 11**) : une phase d'analyse pangénomique proprement dite où la ou les cohortes disposant d'un GWS, que nous appellerons la ou les 'discovery cohorts', sont analysées. Il s'agit d'une phase où un grand nombre de tests multiples sont effectués et où les significations statistiques des résultats générés sont très faibles (le seuil 'genome-wide significant' est fixé à $5x10^{-8}$). Il convient donc de confirmer ou d'infirmer les associations les plus significatives obtenues lors d'une ou de plusieurs phases de réplication dans des populations indépendantes [Chanock et al. 2007 ; Hoover. 2007] de façon à éliminer les faux positifs [Wang et al. 2005]. Seuls les SNPs d'intérêt (du fait de leur signification statistique ou de leur plausibilité) sont donc secondairement génotypés dans ces populations indépendantes de façon à minimiser les coûts.

Figure 11 : Design classique des Genome-wide association studies



Pour effectuer ces génotypages de réplication, nous avons utilisé la technologie Applied Biosystems SNPlex™ basée sur une Oligonucleotide Ligation Assay (OLA) combinée avec une technique d'amplification Multiplex.

Les variants génétiques passant le stade de contrôle qualité avaient un 'call rate' de plus de 95% et leur équilibre de Hardy-Weinberg (HWE) était testé ($p > 0,001$).

(http://www3.appliedbiosystems.com/cms/groups/mcb_support/documents/generaldocuments/cms_042019.pdf).

1.1.6. Les données transcriptomiques

Ce chapitre concerne le VEGF.

L'ARN a été extrait de Cellules mononucléaire de sang périphérique (PBMCs) précédemment isolées par centrifugation sur gradient de densité (Ficoll-Paque^{TM} PLUS; Amersham BioSciences), puis stockées à - 80°C [Marteau et al. 2005] avec un automate MagNaPure utilisant le kit d'isolation MagNA Pure LC RNA HP et le protocole de lyse externe RNA HP Blood (Roche Diagnostics, France).

Sa transcription reverse a été effectuée en utilisant 200 unités de M-MuLV Reverse Transcriptase et 0,25 µg d'oligos(dt) (Promega, France) [Marteau et al. 2005] et la quantification des transcrits codants pour les isoformes $VEGF_{121}$ et $VEGF_{165}$ et le gène de contrôle *β-2- microglobulin* (*B2M*) a été effectuée en utilisant une technologie TaqMan® (LC TaqMan Master kitRoche Diagnostics, France) et un LightCycler. Toutes les procédures ont été effectuées en double et le niveau de détection de chaque transcrit était compris entre 1 et 10 copies.

Les spécificités et optimisation de cette RT-PCR ont été contrôlées à l'aide de la technologie SYBR® Green (LC FastStart DNA Master^{PLUS} SYBR Green I kit, Roche Diagnostic, France).

Les primers et amorces ont été 'designées' pour amplifier spécifiquement les isoformes de VEGF en se basant sur leurs sites d'épissage. Pour toutes les expérimentations, la variabilité intra et inter-test étaient respectivement de 11% et 5%. Les produits de PCR de chaque isoforme ont été purifiés avec un kit QiaQuick (Qiagen, France) et les concentrations déterminées avec un spectrophotomètre.

La quantification protéique du VEGF (121 et 165) dans les PBMCs a été effectuée avec un appareillage Randox Ltd (Crumlin, UK) [Fitzgerald et al. 2005].

La principale population de nos investigations et la 'discovery cohort' de nos études pangénomiques dans les cohortes pédiatriques était la cohorte STANISLAS. Ainsi, de façon à valider les résultats que nous avons obtenus selon le design des GGWAS précédemment schématisé, il a fallu que nous intégrions des Consortia Internationaux nous donnant accès à diverses populations. Un bref descriptif de ces dernières est exposé ci-dessous.

1.2. La population OBE

Il s'agit d'une population d'enfants obèses (définis comme ayant un IMC supérieur au 97ème percentile d'une cohorte Française type en fonction de l'âge et du sexe) [Rolland-Cachera et al. 1991] et appartenant à 449 familles nucléaires ayant au moins un enfant obèse. Ces enfants ont été recrutés dans l'unité endocrinienne pédiatrique de l'Hôpital Jeanne de Flandres, Lille, France ou par l'intermédiaire d'une campagne médiatique.

1.3. La Northern Finland 1986 Birth Cohort (NFBC86)

La NFBC86 est une cohorte prospective ayant inclus des mères Finnoises d'origine Européenne attendant un enfant à naître entre le 1er Juillet 1985 et le 30 Juin 1986 dans deux provinces de Finlande [Jarvelin et al. 1997].
Entre Août 2001 et Juin 2002, 9 215 membres de la cohorte vivant en Finlande et d'adresse connue ont été invités à subir des examens cliniques. 6 798 sujets ont accepté de participer à cette étude au final (74% de taux de participation).

1.4. La GeNe and Diet Attica Investigation (GENDAI)

La cohorte pédiatrique GENDAI est composée d'enfants vivant dans la région d'Attica en Grèce [Papoutsakis et al. 2007]. De Novembre 2005 à Juin 2006, 1 138 péri-adolescents (53% de filles; moyenne d'âge de 11,2±0,7 ans) ont été aléatoirement recrutés dans les écoles élémentaires d'Attica avec l'accord de l'Institutional Review Board de l'université

d'Harokopio, du Ministère Grec de l'Éducation et du comité d'éthique de l'Université d'Harokopio.

1.5. La cohorte VERONA

La cohorte Verona est composée d'enfants italiens (43% de filles ; âge moyen de 10,75 ± 1,22 ans) recrutés dans la population générale de Vérone. Les familles ciblée ont été aléatoirement sélectionnées à partir des registres municipaux de la ville et contactés par voie postale.

1.6. La cohorte Data from an Epidemiological Study on the Insulin Resistance syndrome (DESIR)

Il s'agit de 2 566 hommes et de 2 633 femmes Français âgés de 30 à 64 ans et recrutés de 1994 à 1996 parmi les assurés sociaux. Des examens de santé périodiques gratuits leurs ont été proposés. Cette étude a été approuvée par le Comité Consultatif de Protection des Personnes dans la Recherche Biomédicale (CCPPRB) de l'Hôpital Bicêtre, Paris, France et la Commission Nationale de l'Informatique et des Libertés (CNIL). Tous les participants ont signé un formulaire de consentement éclairé [Balkau et al. 1997].

1.7. La cohorte FRAMINGHAM

La Framingham Heart Study (FHS) est une étude prospective qui s'est poursuivie sur 3 générations dans la ville de Framingham, USA. Elle a été initiée en 1948 en vue d'étudier les facteurs de risque des MCV et est composée aujourd'hui de la génération originale suivie depuis 1948 [Dawber et al. 1966], de leurs enfants et conjoints de leurs enfants depuis 1971 (cohorte Offspring) [Feinleib et al. 1975] et de leurs petits-enfants depuis 2002 [Splansky et al. 2007] soit 4 095 participants avec au moins un parents issus de la cohorte Offspring. Seuls les participants de la 3ème génération ont été dosés pour le VEGF dès leur premier examen de

santé en 2002-2005. L'Institutional Review Board de la faculté de Médecine de Boston a approuvé cette étude.

1.8. La Prospective Investigation of the Vasculature in Upssala Seniors (PIVUS) study

Tous les individus âgés de 70 ans et vivant dans la communauté d'Uppsala en Suède entre Avril 2001 et Juin 2004 étaient éligibles et ont été aléatoirement sélectionnés [Lind et al. 2005]. Tous les individus incorporés à l'étude ont été examinés dans le mois de leur 70ène anniversaire de façon à standardiser leurs âges. Sur les 2 025 individus conviés, 1 016 ont été intégrés à l'étude soit un taux de participation de 50% et 999 ont fourni de l'ADN. L'étude PIVUS a reçu l'accord du comité d'éthique de l'université d'Uppsala.

La concentration en haptoglobine a été quantifiée pour les populations pédiatriques OBE et GENDAI par l'unité INSERM U954 'Exposition aux risques génétiques et environnementaux' de la Faculté de Médecine Henri Poincaré Nancy 1 dans le cadre de ce projet collaboratif.

2. Analyses statistiques

Les analyses statistiques ont été effectuées avec SPSS versions 14.0 à 17.0 (SPSS Inc., Chicago, IL, USA), SAS (SAS Institute, Cary, NC, USA), R, et Plink.

2.1. Test de l'équilibre de Hardy-Weinberg

L'HWE ou équilibre panmictique, est le modèle théorique central de la génétique des populations. La notion d'équilibre dans le modèle de Hardy-Weinberg est soumise à des hypothèses définissant une population idéale :
1/ Une population panmictique c'est-à-dire où les couples se forment au hasard (panmixie), et leurs gamètes se rencontrent au hasard (pangamie) ;

2/ Une population ‘infinie’ minimisant les variations d'échantillonnage ;

3/ Pas de perte ou de gain d’allèle : ni sélection, ni mutation, ni migration ;

4/ Et pas de croisement entre générations différentes : générations successives discrètes.

Ceci permet de maintenir la diversité génétique de la population et tend vers un équilibre stable de la distribution génotypique. Les fréquences alléliques restent également stables de génération en génération dans une population diploïde idéale et ne dépendent que des fréquences de la génération initiale. De plus, les fréquences génotypiques ne dépendent que des fréquences alléliques.

Ainsi, dan une population idéale de taille n et avec comme allèles du même locus A et a, p la fréquence de l’allèle A et q=1-p la fréquence de l’allèle a :

Phénotypes	[A]	[Aa]	[aa]
Génotypes	AA	Aa	aa
Effectifs génotypiques	n_{AA}	n_{Aa}	n_{aa}

Les fréquences des génotypes en HWE sont :

$f(AA) = p^2$

$f(aa) = q^2 = (1-p)^2$

$f(Aa) = 2pq = 2p(1-p)$

et elles sont constantes au fil des générations.

Ainsi, la formule de calcul de la fréquence allélique p dans une population donnée est :

$$p = (2\, n_{AA} + n_{Aa}) / 2n$$

Au cours de notre étude, nous nous sommes assurés de l’HWE par le test du Chi-deux (χ^2) avec le logiciel SPSS pour l’approche gène candidat, ou avec le logiciel plink pour l’approche pangénomique.

2.2. Déséquilibre de liaison

L'hypothèse de déséquilibre de liaison entre les paires de loci a été vérifiée par plink.

2.3. Test d'ajustement : loi de distribution des variables

Nous avons vérifié la normalité et la symétrie de la distribution des variables étudiées par le test de Kolmogorov-Smirnov qui étudie l'écart entre la fonction des probabilités cumulées de l'échantillon et la fonction des probabilités cumulées de la loi théorique et en calculant les coefficients de dissymétrie et d'aplatissement (skewness-kurtosis).

2.4. Statistiques descriptives

Les moyennes et écart-types ont été calculées pour les variables quantitatives et les effectifs et fréquences pour les variables qualitatives. Des comparaisons d'effectifs de différents types ont été menées en utilisant le test du χ^2.

2.5. Études d'association

Des comparaisons de moyennes de variables biologiques et cliniques ont été effectuées ainsi que des analyses de régression. Différents tests ont été effectués selon la nature et la distribution des valeurs de chaque variable.

- Des tests t de Student et analyse de variance (ANOVA) avec correction de Bonferroni le cas échéant ;
- Des régressions multiples utilisant la méthode des moindres carrés ;
- Des régressions logistiques dans le cas de variables qualitatives ;
- Des tests non paramétriques de Mann-Whitney, Kruskal-Wallis et Friedman qui ne font aucune hypothèse sur la forme des distributions sous-jacentes

2.6. Corrélations familiales

La ressemblance familiale est la ressemblance entre individus apparentés dont l'origine est le partage de gènes communs, de facteurs environnementaux, induit des corrélations familiales de deux types :

- intra-génération (entre enfants ou entre parents) ;
- et inter-génération (parents-enfants) [Tregouet et al. 1999].

CHAPITRE 4 – RÉSULTATS ET DISCUSSION

1. Étude de l'effet précoce de l'héritabilité génétique des facteurs de risque des pathologies cardiovasculaires dans la cohorte STANISLAS.

PUBLICATIONS N° 2 & 3

Comme cela a été précédemment évoqué, les divers facteurs de risque impliqués dans le développement des pathologies cardiovasculaires présentent une composante génétique dont l'amplitude et les mécanismes d'action restent à élucider. Nous avons choisi, au sein de la cohorte STANISLAS, d'en étudier certains classiques : les traits lipidiques et la pression artérielle ; et d'autres émergents : l'haptoglobine, molécule d'inflammation, et le VEGF, molécule d'adhésion cellulaire. Une telle approche visait à mettre en exergue l'intérêt de mener des études d'épidémiologie étiologique sur les variants génétiques impliqués dans l'héritabilité de ces traits au sein de populations pédiatriques. Les résultats obtenus nous ont confortés dans cette voie en nous permettant de visualiser les modulations que pouvaient exercer les facteurs environnementaux sur l'héritabilité génétique ; ou du moins sur sa quantification.

De plus, nous avons pu démontrer que sur un point de vue strictement phénotypique, les effets de cette héritabilité étaient mesurables en pratique clinique chez les enfants.

Ceci souligne donc l'intérêt, pour mesurer les effets propres de la génétique, de se focaliser sur ces populations pour lesquelles les effets de l'environnement sont encore minimes.

1.1. L'héritabilité des traits lipidiques, de la pression artérielle, de l'haptoglobine et du facteur de croissance de l'endothélium vasculaire dans la Cohorte STANISLAS. *PUBLICATION N°2*

Les dyslipidémies, l'HTA, et les marqueurs de l'inflammation sont des facteurs de risque des MCV impliqués dans la détermination du SM au même titre que l'obésité centrale, l'acide urique et les enzymes hépatiques.

Dans un échantillon regroupant 667 familles nucléaires de la cohorte STANISLAS composées de deux parents et d'au moins un enfant biologique : 1 334 parents âgés de 30 à 64 ans et leurs 1 137 enfants âgés de 12 à 25 ans (**Tableau 7**), nous avons étudié de manière transversale dans un premier temps l'héritabilité génétique des divers composants du SM.

La cohorte STANISLAS ayant été suivie de manière prospective sur 10 ans, nous avons également souhaité effectuer une étude longitudinale à court terme (5 ans) et avons sélectionné, dans l'échantillon précédent, les familles ayant subi les deux examens de santé consécutifs de 1994-95 et de 1999-2000. Ceci représentait 353 familles : 706 parents et leurs 502 enfants (**Tableau 8**).
Le modèle d'estimation de l'héritabilité que nous avons utilisé prenait en compte trois composantes indépendantes : la composante polygénique, l'environnement partagé au sein de la cellule familiale, et l'environnement non partagé. Le **Tableau 9** réuni les deux premières composantes étudiées.

Avec la première étude transversale, nous avons pu mettre en évidence une grande héritabilité génétique des certains variants d'intérêt :

- Les traits lipidiques : 44% pour le cholestérol total, 32,8% pour le HDL-C, 30% pour l'ApoA-I, 38,5% pour l'ApoB, et 34,7% pour l'ApoE avec une importante signification statistique ($p<0,001$) ;
- Et un marqueur d'inflammation : l'haptoglobine avec une héritabilité génétique représentant plus du tiers de la variance globale dans cet échantillon (35,3% ; $p<0,001$).

À l'opposé, nous n'avons obtenu aucune héritabilité génétique significative pour la pression artérielle. Ce dernier résultat était en contradiction avec les études antérieurement menées sur un échantillon différent de la cohorte STANISLAS : de Andrade et ses collaborateurs [de Andrade et al. 2002] observaient en 2002 une héritabilité génétique de la PAS de 32%. Une telle disparité au sein d'une même cohorte prouve que la quantification de l'héritabilité d'un trait peut varier en fonction de la spécificité des échantillons sélectionnés et notamment en fonction de facteurs d'environnement ayant un effet modulateur ou masquant plus ou moins appuyé. C'est un aspect que nous avons choisi d'investiguer plus avant et que nous exposerons dans les chapitres à venir.

Il est à noter que lors de l'étude longitudinale des variations à 5 ans des composants du SM, aucune héritabilité génétique ne s'est avérée significative. À l'inverse, l'effet de l'environnement partagé était très élevé dans tous les cas de figure. Ceci nous permet d'insister sur le fait que la notion d'héritabilité génétique, ou du moins sa quantification, est une 'photo' à l'instant t de l'effet génétique réel, et qu'à court terme, cet effet pourrait être totalement masqué par l'environnement.

Enfin, notez que le VEGF n'étant pas un composant du SM, nous ne l'avons pas intégré à cette étude. Néanmoins, son héritabilité génétique a été précédemment évaluée dans la cohorte STANISLAS [Berrahmoune et al. 2007]. Dans le sous-ensemble de la cohorte dans lequel ce trait avait été quantifié, sa variabilité interindividuelle était pour plus de la moitié (60.6%) sous contrôle génétique. Néanmoins, les variants responsables de cette héritabilité étaient restés indéterminés à l'époque.

Nous avons donc, dans la suite de nos travaux, cherché à identifier les variants concernés.

Tableau 7 : Caractéristiques basales des 2 471 sujets inclus dans l'étude transversale - Données issues de la cohorte STANISLAS à t_0 : 1994-1995

	Adultes		Enfants	
	Pères (n=667)	Mères (n=667)	Fils (n=575)	Filles (n=562)
Age, années	42,5 (4,6)	40,4 (4,5)	15,5 (2,9)	15,9 (3,3)
Poids, kg	77,1 (11,1)	61,4 (10,2)	57,1 (13,3)	53,7 (9,7)
IMC, kg/m²	25,5 (3,2)	23,7 (3,9)	20,1 (3,0)	20,5 (3,1)
IPC, %	102,1 (12,8)	103,0 (17,5)	103,3 (13,1)	104,5 (15,1)
Tour de taille, cm	88,9 (9,0)	75,1 (9,0)	71,1 (7,4)	66,4 (6,6)
Ratio taille/hanche, 1	0,91 (0,06)	0,77 (0,06)	0,81 (0,04)c	0,73 (0,05)
Consom alcool, g/j	23,9 (25,7)	4,7 (8,9)	1,1 (5,7)	0,3 (1,5)
Consom tabac, cig/j	5,4 (9,4)	2,6 (6,4)	0,8 (3,3)	1,2 (4,2)
Traitement anti-HT, %	3,3	1,7	0	0
Traitement anti-diab, %	0,4	0	0	0
Hypolipémiants, %	7	1,7	0,5	0,5
Anti-inflam, %	3,8	2,2	2,6	2,8
PAS, mmHg	128,3 (12,5)	120,3 (13,6)	121,2 (11,7)	115,2 (9,6)
PAD, mmHg	78,6 (9,7)	72,7 (10,3)	62,1 (11,1)	61,9 (10,2)
Triglycérides, mool/L	1,21 (0,71–2,06)	0,84 (0,55–1,29)	0,74 (0,48–1,14)	0,82 (0,54–1,23)
Cholestérol total, mmol/L	5,99 (1,10)	5,52 (0,91)	4,49 (0,82)	4,85 (0,84)
HDL-cholestérol, mmol/L	1,39 (0,35)	1,73 (0,42)	1,43 (0,36)	1,55 (0,35)c
ApoB, g/L	1,15 (0,26)	0,98 (0,22)	0,79 (0,19)	0,85 (0,19)
ApopAI, g/L	1,55 (0,23)	1,71 (0,26)	1,44 (0,21)	1,53 (0,24)
ApoE, mg/L	41,3 (14,4)	37,0 (10,6)	36,2 (9,8)	39,1 (10,7)
Glycémie, mmol/L	5,19 (0,56)	4,88 (0,40)	4,92 (0,34)	4,79 (0,39)
Acide urique, µmom/L	319,5 (67,7)	230,1 (48,6)	293,0 (61,2)	242,6 (45,7)
hs-CRP, mg/L	1,00 (0,35–2,87)	0,97 (0,31–3,02)	0,45 (0,15–1,34)	0,47 (0,16–1,38)
Haptoglobine, g/L	1,09 (0,49)	1,07 (0,42)	0,72 (0,41)	0,89 (0,41)
Orosomucoïde, g/L	0,79 (0,19)	0,71 (0,18)	0,71 (0,21)	0,69 (0,20)

ASAT, U/l	23,4 (16,6–33,0)	17,3 (13,0–23,0)	23,3 (16,7–32,8)	18,7 (14,3–24,4)
ALAT, U/l	30,2 (18,7–48,8)	17,6 (10,9–28,1)	19,2 (12,6–29,3)	15,5 (10,6–22,7)
GGT, U/l	30,4 (16,6–55,7)	16,2 (9,7–27,1)	13,9 (10,1–19,2)	11,3 (8,4–15,2)
Globules blancs, 109/l	6,92 (1,86)	7,05 (1,84)	6,58 (1,66)	7,12 (1,79)
Syndrome métabolique, %	14,8	6,4	0,5	1,1

Tableau 8 : Évolution des caractéristiques des 1 208 sujets inclus dans l'étude longitudinale - Données issues de la cohorte STANISLAS à t_0 (1994-1995) et t_5 (1999-2000)

	Adultes		Enfants	
	Pères (n=667)	Mères (n=667)	Fils (n=575)	Filles (n=562)
Δ Poids, kg	2,41 (4,07)	2,10 (4,44)	13,99 (10,11)	6,22 (7,52)
Δ IMC, kg/m²	0,85 (1,33)	0,86 (1,70)	2,15 (1,80)	1,53 (2,35)
Δ IPC, %	2,22 (5,23)	0,98 (7,26)	-0,72 (8,03)	0,21 (10,83)
Δ Tour de taille, cm	2,40 (4,70)	1,75 (4,65)	6,58 (5,19)	2,87 (5,58)
Δ Ratio taille/hanche, 1	0,02 (0,04)	0,02 (0,04)	0,01 (0,05)	-0,01 (0,05)
Δ Consom alcool, g/j	-0,70 (16,51)	0,89 (7,63)	3,70 (8,71)	1,06 (3,06)
Δ Consom tabac, cig/j	-0,52 (5,63)	-0,03 (3,35)	3,28 (6,41)	2,39 (5,49)
Δ PAS, mmHg	-0,07 (12,72)	0,27 (13,01)	3,62 (12,05)	0,70 (9,98)
Δ PAD, mmHg	-1,17 (11,25)	-0,63 (11,70)	4,86 (13,03)	4,37 (12,46)
Δ Triglycérides, mool/L	0,11 (0,93)	0,03 (0,41)	0,08 (0,41)	0,01 (0,48)
Δ Cholestérol total, mmol/L	0,04 (0,89)	0,08 (0,76)	-0,11 (0,68)	0,10 (0,78)
Δ HDL-cholestérol, mmol/L	0,02 (0,28)	0,05 (0,33)	-0,15 (0,31)	0,09 (0,34)
Δ ApoB, g/L	-0,03 (0,21)	-0,03 (0,17)	-0,03 (0,14)	0,00 (0,18)
Δ ApopAI, g/L	-0,01 (0,18)	-0,01 (0,23)	-0,08 (0,21)	0,07 (0,27)
Δ ApoE, mg/L	1,78 (12,08)	2,59 (8,76)	-2,62 (7,40)	-3,16 (9,69)
Δ Glycémie, mmol/L	0,22 (0,86)	0,11 (0,53)	-0,04 (0,92)	-0,06 (0,51)
Δ Acide urique, mmom/L	3,64 (55,06)	-3,21 (40,43)	30,84 (63,54)	-16,50 (43,92)
Δ hs-CRP, mg/L	0,01 (7,39)	0,15 (5,53)	0,83 (5,87)	0,66 (2,66)
Δ Haptoglobine, g/L	0,05 (0,35)	0,07 (0,34)	0,29 (0,40)	0,18 (0,39)
Δ Orosomucoïde, g/L	0,00 (0,16)	0,02 (0,17)	0,04 (0,23)	-0,01 (0,24)
Δ ASAT, U/l	0,61 (13,88)	1,24 (7,99)	-3,55 (12,09)	-1,20 (6,52)
Δ ALAT, U/l	0,22 (19,67)	0,59 (15,96)	-0,47 (18,55)	0,25 (10,61)
Δ GGT, U/l	4,75 (22,94)	2,69 (19,12)	2,96 (6,13)	2,14 (5,61)
Δ Globules blancs, 109/l	0,01 (1,53)	0,10 (1,67)	0,24 (1,81)	0,25 (1,96)
Δ Syndrome métabolique, %				

	t0	14,2	7,1	0	1,2
	t5	20,7	7,9	0,8	0,4

Tableau 9 : Composante de variance dans les 667 familles inclues à l'étude transversale et des 353familles inclues dans l'étude longitudinale

	t0		t5	
	σ^2G (variance polygénique, %)	σ^2C (enironnement partagé, %)	σ^2G (variance polygénique, %)	σ^2C (enironnement partagé, %)
Cholestérol total	44,0***	7,8*	0	12,9***
ApoB	38,5***	12,6***	0	16,5***
Triglycérides	11,7	16,8***	3,7	5,9
ApoE	34,7***	10,6**	0	22,4***
Cluster risque lipidique	40,5***	8,9**	0	10,2***
Tour de taille	4,1	25,8***	0	15,8***
IPC	7	22,6***	0	11,1***
PAS	10,9	8,8***	0	13,4***
PAD	0	12,2***	7,1	16,1**
Acide urique	37,8***	12,0***	8,1	23,7***
Cluster adiposité/PA	6,8	21,0***	0	14,3***
Glycémie	20,8***	17,6***	0	24,0***
HDL-cholestérol	32,8***	16,3***	0	27,2***
ApoAI	30***	23,4***	0	25,5***
Cluster lipides protectrices	35,0***	14,9***	0	18,3***
ALAT	18,8**	9,0**	0	5,7*
ASAT	13,6	14,3***	0	2,6
GGT	8,4	11,9***	0	10,4***
Cluster enzymes hépatiques	18,4**	9,8**	0	2,9
Haptoglobine	35,3***	10,0**	7,3	9,9*
hs-CRP	5,9	11,0**	0	5,6
Orosomucoïde	8,9	16,7***	0	7,4**
Globules blancs	4,4	16,8***	0	8,3**
Cluster inflammation	7,6	14,2***	0	9,0**

1.2. Étude exploratoire de l'effet phénotypique de l'héritabilité des facteurs de risque des maladies cardiovasculaire chez les enfants de la Cohorte STANISLAS.

PUBLICATION N°3

Avant de poursuivre nos investigations génétiques, nous avons choisi de faire un descriptif simple de l'effet familial visualisable des facteurs de risque classiques des MCV chez les enfants de la cohorte STANISLAS : la pression artérielle (systolique et diastolique), les lipides (cholestérol total, HDL-C, LDL-C, TG), la glycémie et les données anthropométriques (IMC et tour de taille). Le but était de démontrer que dans les populations pédiatriques ayant des données infracliniques, il était quand même possible de repérer les individus à risque en fonction des variables de leurs parents.

Nous avons donc utilisé la totalité des familles STANISLAS pour lesquelles nous avions les données investiguées à disposition pour les deux premiers bilans de santé : 693 familles biparentales et leurs enfants soit 1 386 parents et 1 194 enfants au total (**Tableau 10**).

Dans un premier temps, nous avons classé les parents considérés 'à risque' selon le niveau individuel de leurs facteurs de risque en fonction des recommandations Nationales et Internationales :

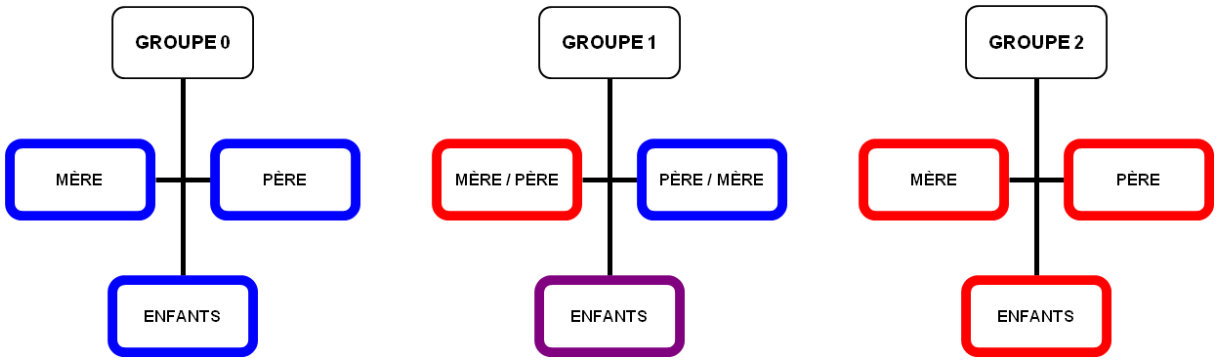
- PAS > 130mmHg pour la pré-HTA et > 140 mmHg pour l'HTA systolique ;
- PAD > 85mmHg pour la pré-HTA et > 90 mmHg pour l'HTA diastolique ;
- Cholestérol total ≥ 6 mmol/L ;
- LDL-C ≥ 5,16 mmol/L ;
- HDL-C ≤ 1 mmol/L ;
- TG ≥ 2 mmol/L ;
- Glucose > 6,2 mmol/L ;
- IMC ≥ 25 kg/m² ;
- Et tour de taille > 80 cm pour les pères et 94 cm pour les mères.

Cette première classification nous a permis d'identifier les familles considérées à risque pour chaque trait étudié :

- Les familles non à risque constituaient les groupes 0 lorsqu'aucun parent n'était à risque ;
- Les familles à risque modéré ou groupes 1 lorsqu'un parent (père ou mère) était à risque pour le trait considéré ;
- Et enfin, les familles à risque élevé lorsque les deux parents avaient leurs facteurs de risque au-delà des valeurs seuil listées plus haut.

  (**Figure 12**)

Figure 12 : Groupes familiaux en fonction du niveau de risque des parents STANISLAS



Ces groupes ainsi constitués, nous avons pu constater qu'à l'exception de la glycémie, tous les facteurs de risque cardiovasculaire des enfants appartenant aux familles à risque élevé avaient des niveaux supérieurs à ceux des enfants appartenant aux familles à risque modéré ; qui étaient eux même supérieurs à ceux des familles non à risque (**Figures 13 et 14**). Il y avait là ce que l'on pourrait qualifier 'd'effet dose' ou caractère graduel : une caractéristique requise des facteurs de risque cardiovasculaire comme cela a été précédemment expliqué.

Dans le cas de la glycémie, le faible effectif des groupes à risque était à notre sens à l'origine de la non signification des résultats. En effet, la cohorte STANISLAS est composée d'individus

a priori sains et leur glycémie est donc en moyenne en deçà des seuils préconisés en prévention.
Ainsi, même sur un plan strictement phénotypique, les effets de l'héritabilité d'un facteur de risque cardiovasculaire sont mesurables en pratique clinique chez les enfants. Il suffit pour cela d'évaluer le 'niveau de risque' de leurs parents.

Ce fait, associé à la difficulté que représente la quantification de l'effet propre de la génétique sur la variabilité interindividuelle d'un trait donné, souligne à notre sens l'intérêt que représente les études d'épidémiologie génétique se focalisant sur des populations pédiatriques pour lesquelles les effets de l'environnement sont encore minimes.

Tableau 10 : Caractéristiques phénotypiques des familles étudiées et variations au cours des examens de santé

	Examen	Pères (n=693)*	Mères (n=693)*	p-values	Test de Levène	Fils (n=585)*	Filles (n=609)*	p-values	Test de Levène
Age, années	t0	42,85 ± 5,08	40,7 ± 4,8	< 0,001	0,254	13,3 ± 3,6	13,6 ± 3,9	0,29	0,03
	t5	47,56 ± 5,12	45,5 ± 4,8	< 0,001	0,257	18 ± 3,6	18,3 ± 3,9	0,257	0,045
PAS, mmHg	t0	127,6 ± 12,3	119,2 ± 12,5	< 0,001	0,008	115,2 ± 12,3	111 ± 9,5	< 0,001	< 0,001
	t5	128,4 ± 13,4	121,1 ± 14,2	< 0,001	0,904	122,4 ± 10,8	115,4 ± 9,1	< 0,001	< 0,001
	Δ	1,2 ± 12,2	1,9 ± 12,3	0,254	0,168	7,2 ± 11,6	4,4 ± 10,2	< 0,001	0,001
PAD, mmHg	t0	78,1 ± 9,7	72,5 ± 9,7	< 0,001	0,001	57,8 ± 11	57,2 ± 10,7	0,668	0,441
	t5	77,6 ± 10,2	72,2 ± 9,9	< 0,001	0,369	64 ± 10,5	62,9 ± 1	0,074	0,236
	Δ	-0,5 ± 11	-0,3 ± 10,5	0,703	1,157	6,2 ± 12,5	5,4 ± 12,1	0,299	0,949
PP, mmHg	t0	49,5 ± 9,3	46,7 ± 8,8	< 0,001	6,641	57,4 ± 12,1	53,5 ± 10,1	< 0,001	< 0,001
	t5	51,1 ± 9,4	48,9 ± 9	< 0,001	0,823	58,5 ± 12,5	52,5 ± 10,9	< 0,001	0,001
	Δ	1,7 ± 10,6	2,2 ± 10,4	0,348	0,034	1,1 ± 15,1	-1 ± 13,3	0,01	0,001
PAM, mmHg	t0	94,6 ± 9,7	88,1 ± 9,9	< 0,001	0,38	76,9 ± 9,9	75,3 ± 9,1	0,005	0,026
	t5	94,6 ± 10,5	88,5 ± 10,7	< 0,001	0,212	83,5 ± 8,8	80,4 ± 8,2	< 0,001	0,123
	Δ	0,1 ± 10,3	0,5 ± 10	0,467	0,319	6,5 ± 9,9	5,1 ± 9,7	0,011	0,98
Cholestérol total, mmol/L	t0	5,99 ± 1,12	5,53 ± 0,9	< 0,001	21,268	4,63 ± 0,82	4,87 ± 0,8	< 0,001	0,465
	t5	5,98 ± 1,03	5,62 ± 0,93	< 0,001	5,355	4,37 ± 0,8	4,8 ± 0,92	< 0,001	0,062
	Δ	-0,02 ± 0,943	0,08 ± 0,78	0,031	8,204	-0,27 ± 0,69	-0,07 ± 0,73	< 0,001	0,601
LDL-cholestérol, mmol/L	t0	1,06 ± 0,98	3,54 ± 0,86	< 0,001	11,277	2,88 ± 0,76	3,06 ± 0,75	< 0,001	0,704
	t5	3,81 ± 0,9	3,34 ± 0,85	< 0,001	1,856	2,51 ± 0,75	2,71 ± 0,83	< 0,001	0,44
	Δ	-0,25 ± 0,76	-0,21 ± 0,66	0,223	9,741	-0,37 ± 0,61	-0,35 ± 0,61	0,589	0,48
HDL-cholestérol, mmol/L	t0	1,27 ± 0,35	1,56 ± 0,39	< 0,001	10,778	1,4 ± 0,36	1,43 ± 0,34	0,146	0,126
	t5	1,42 ± 0,4	1,75 ± 0,44	< 0,001	11,25	1,39 ± 0,34	1,63 ± 0,39	< 0,001	0,003

	Δ	0,15 ± 0,28	0,19 ± 0,33	0,008	17,119	-0,01 ± 0,32	0,2 ± 0,33	< 0,001	0,345
Triglycérides, mmol/L	t0	1,39 ± 1,47	0,88 ± 0,53	< 0,001	51,105	0,7 ± 0,4	0,76 ± 0,36	0,005	758
	t5	1,54 ± 1,06	1,1 ± 1,49	< 0,001	20,511	0,92 ± 0,42	0,93 ± 0,44	0,654	0,656
	Δ	0,15 ± 1,22	0,23 ± 1,45	0,303	11,33	0,22 ± 0,42	0,17 ± 0,46	0,054	0,195
Glucose, mmol/L	t0	5,22 ± 0,7	4,92 ± 0,42	< 0,001	9,01	4,94 ± 0,37	4,84 ± 0,37	< 0,001	0,623
	t5	5,3 ± 0,91	4,93 ± 0,54	< 0,001	12,101	4,84 ± 0,7	4,7 ± 0,48	< 0,001	0,233
	Δ	0,07 ± 0,81	0,01 ± 0,56	0,08	13,406	-0,1 ± 0,76	-0,15 ± 0,56	0,204	0,243
IMC, kg/m²	t0	25,5 ± 3,3	23,8 ± 4	< 0,001	21,516	18,8 ± 3,3	19 ± 3,2	0,284	0,967
	t5	26,3 ± 3,6	24,8 ± 4,5	< 0,001	29,076	21,2 ± 3,3	21,1 ± 3,2	0,562	0,602
	Δ	0,9 ± 1,3	1 ± 1,6	0,114	21,529	2,4 ± 1,8	2,1 ± 2,3	0,011	< 0,001
tour de taille, cm	t0	88,5 ± 8,8	75 ± 8,8	< 0,001	0,001	66,5 ± 8,7	62,5 ± 7,2	< 0,001	< 0,001
	t5	90,7 ± 9,2	90,7 ± 9,2	< 0,001	3,608	74,4 ± 8,1	67,4 ± 6,8	< 0,001	< 0,001
	Δ	2,4 ± 4,7	2,4 ± 4,7	0,423	0,017	7,7 ± 5,3	4,9 ± 5,8	< 0,001	0,121

Figure 13 : Niveaux des facteurs de risque cardiovasculaire des enfants Stanislas en fonction des groupes familiaux au premier examen de santé

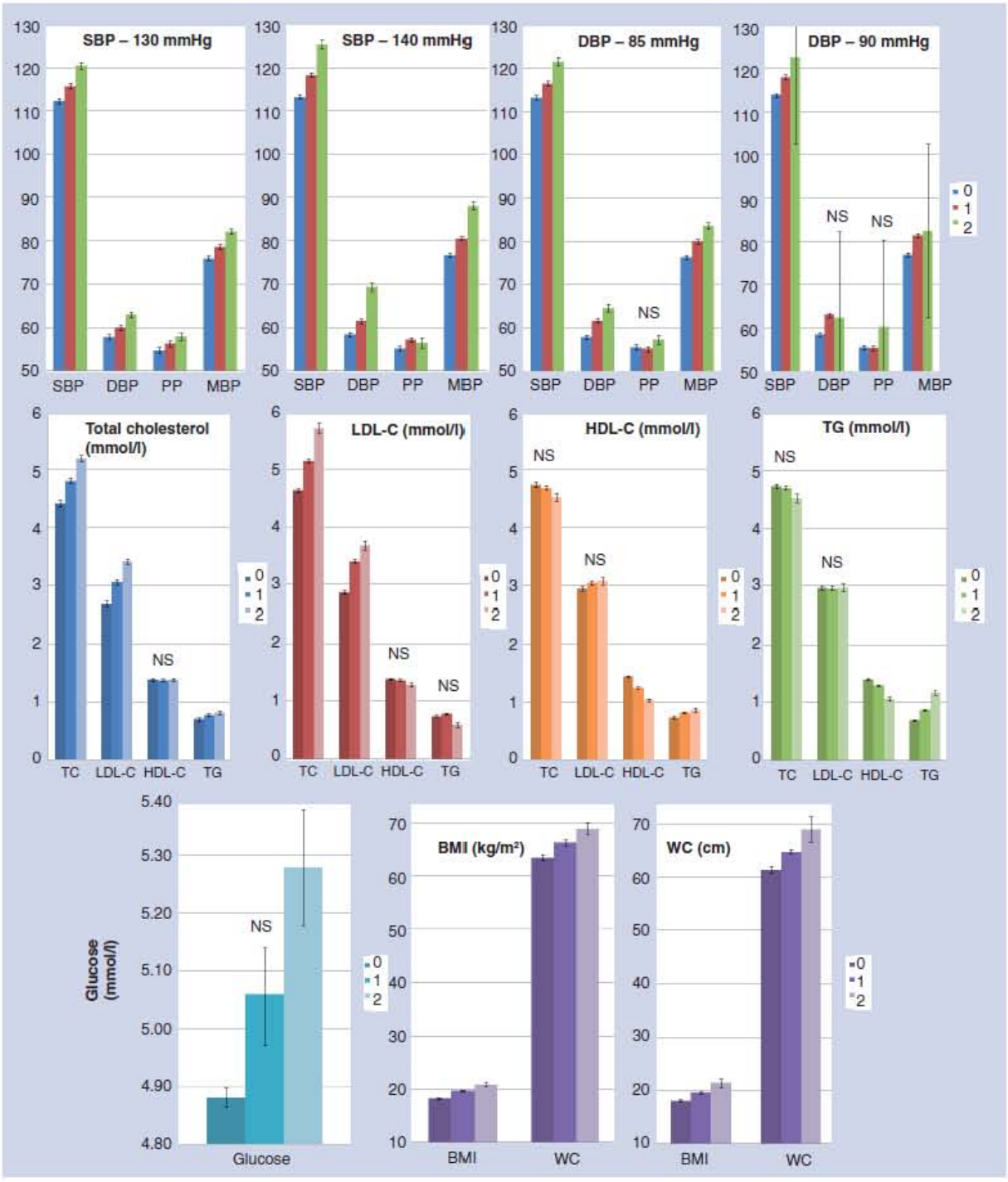
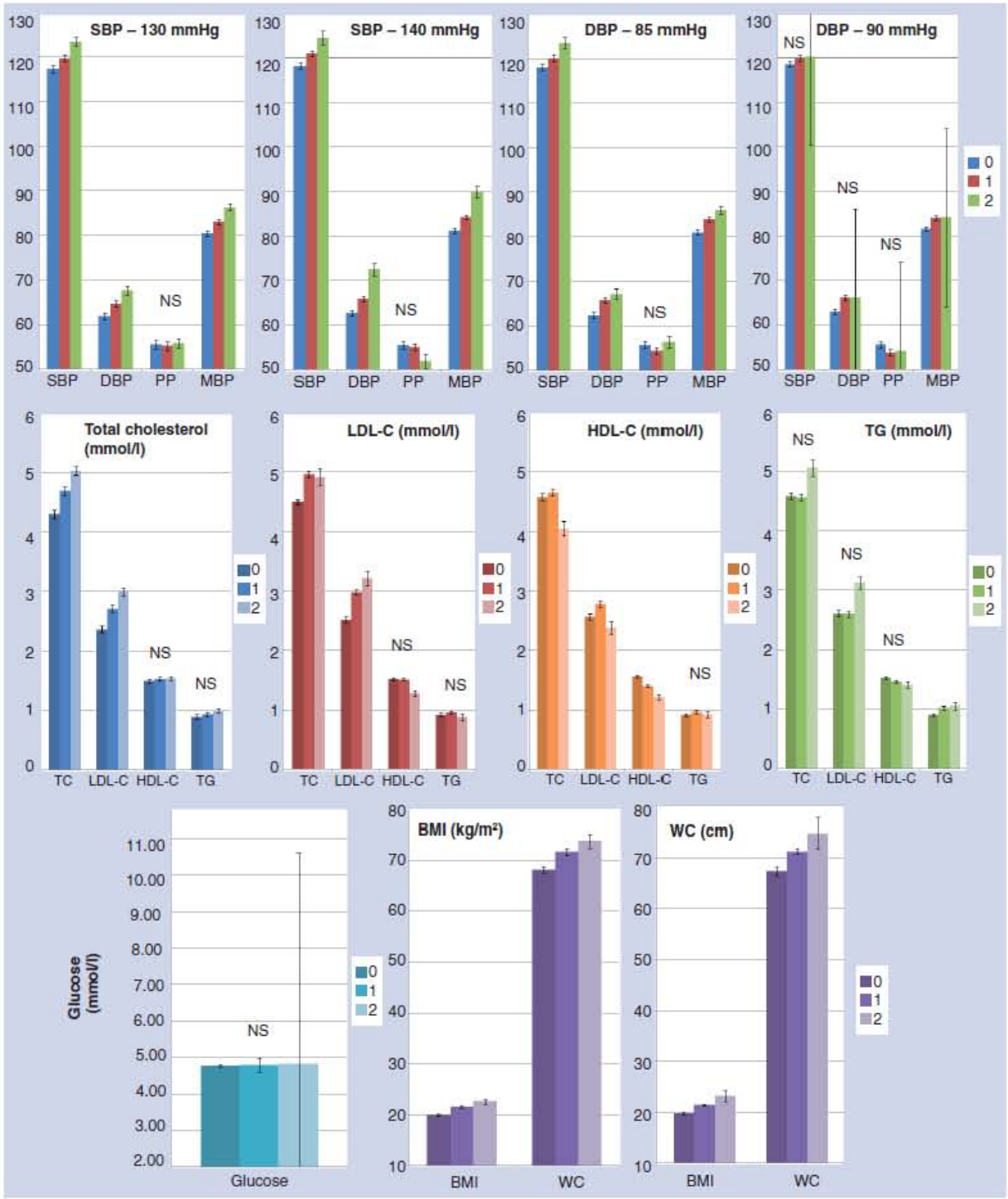


Figure 14 : Niveaux des facteurs de risque cardiovasculaire des enfants Stanislas en fonction des groupes familiaux au second examen de santé



2. Étude d'associations pangénomiques de facteurs de risque des maladies cardiovasculaire dans des populations pédiatriques : l'haptoglobine, la pression artérielle et les lipides.

Avec l'avènement des études d'associations pangénomiques ou GWAS, l'identification des variants génétiques impliqués dans la variabilité interindividuelle des facteurs de risque des pathologies cardiovasculaires a pris un tournant décisif. En effet, l'approche gène-candidat classique, utilisée antérieurement, était basée sur une connaissance *a priori* des mécanismes et des interconnexions menant au développement de la pathologie. La communauté scientifique étudiait des variants génétiques ciblés selon une hypothèse physiopathologique préalable (identification d'une voie enzymatique … etc.). Par conséquent, les découvertes fulgurantes des premières années de l'épidémiologie génétique des pathologies cardiovasculaire ont fini par s'essouffler.

De plus, nous avons évoqué précédemment le cas de traits ayant une forte héritabilité génétique non expliquée par les gènes candidats. Ceci souligne bien la nécessité, pour dépasser le pallier atteint, de développer une approche s'affranchissant d'une hypothèse préalable et permettant par conséquent de mettre en évidence des variants et cheminements physiopathologiques inédits : l'approche pangénomique.

Nous avons effectué le GWS d'un échantillon de 631 enfants minces non apparentés issus de la cohorte STANISLAS avec pour objectif de mener des études d'association pangénomique selon deux approches distinctes :

- Une approche exploratoire et innovante visant à rechercher des variants génétiques associés à un facteur de risque émergent des MCV pour lequel nous avons observé dans la cohorte STANISLAS une forte héritabilité génétique : l'haptoglobine [Herbeth et al. 2010] ;
- Et une approche 'réplicative' des résultats antérieurement obtenus lors de l'étude de facteurs de risque classiques des pathologies cardiovasculaires dans des populations pédiatrique : la pression artérielle et les lipides.

Les GWAS ayant un design en plusieurs stades hiérarchiques consécutifs et nécessitant donc une réplication des résultats des stades antérieurs dans de larges populations indépendantes, nous avons intégré des Consortia Internationaux et initié des projets collaboratifs permettant d'avoir accès à de telles populations.

Au terme de ce volet de notre thèse, nous avons pu démontrer que les variants génétiques connus chez l'adulte pouvaient être détectés chez l'enfant et que surtout, certains variants inédits, dont l'effet pouvait être masqué par des facteurs environnementaux, étaient observables dans des populations pour lesquelles ces facteurs environnementaux étaient minimisés : les populations pédiatriques.

Nous avons également mis en évidence des variants génétiques associés à des traits jusque là non investigués et donnant un éclairage nouveau aux interconnexions existant entre divers facteurs de risque des pathologies cardiovasculaires.

2.1. Genome-Wide Association Study de l'Haptoglobine dans des populations pédiatriques.

PUBLICATION N°4

L'haptoglobine est une molécule inflammatoire qui a la propriété de former un complexe avec l'hémoglobine lors de son relargage par les érythrocytes en vue de son élimination dans le foie. Elle a donc pour propriété avérée de prévenir la génération d'espèces d'oxygène réactives ou reactive oxygen species (ROS) [Okazaki et al. 1997].
Comme cela a été précédemment évoqué, Holme et ses collaborateurs [Holme et al. 2010] ont récemment établi un lien entre cette molécule et le risque d'infarctus du myocarde, d'AVC et d'insuffisance cardiaque dans l'étude AMORIS. L'haptoglobine est ainsi un facteur de prédiction majeur des évènements cardiovasculaires [Holme et al. 2009]. Mais quel est le mécanisme sous-jacent ?

Les 631 enfants STANISLAS ont constitué l'échantillon de phase 1 ou 'discovery set'. Nous y avons effectué une GWAS et identifié une forte association de variants présents dans un block en déséquilibre de liaison de 218 kilo bases incluant le gène *Haptoglobin* (*HP*) et le gène *Haptoglobin Related Protein* (*HPR*) sur le chromosome 16 avec les taux d'haptoglobine. Seuls les variants rs2000999 (β=-0,123; p=6,32x10^{-13}) et rs10492825 (β=-0,093; p=5,50$\times 10^{-08}$) atteignaient le seuil 'genome-wide significant'. Leur proximité physique (en déséquilibre de liaison, r^2=0, 48 – HapMap CEU release 27) pouvait laisser présumer que nous ne visualisions qu'un unique signal, et c'est ce que nous avons vérifié par une régression conditionnée de chaque variant sur l'autre. Lors de cette régression conditionnée, nous avons également testé tous les variants présents dans la zone de recombinaison de ce locus. Au terme de cette procédure, seul le rs2000999 dans le gène *HPR* conservait une signification élevée : $p_{ajusté\ sur\ rs10492825}$=1,9x$10^{-7}$. Il était donc le seul ayant un effet indépendant sur la variabilité interindividuelle de l'haptoglobine (**Tableau 11**).

Pour cette étude spécifique, nous avons réévalué l'héritabilité génétique de l'haptoglobine dans un échantillon indépendant de celui dans lequel cette procédure avait précédemment été faite. Nous avons donc sélectionné 656 familles nucléaires (2 680 sujets) dans la cohorte STANISLAS pour lesquels les variants rs2000999 et rs10492825 avaient été génotypés. La

variance polygénique de ce trait était également élevée dans cet échantillon (24,9%) et un ajustement avec le rs2000999 diminuait ce résultat de 10,5% alors qu'un ajustement sur le rs10492825 ne changeait rien. Le variant rs2000999 participait donc à hauteur de 42% à l'héritabilité génétique globale de l'haptoglobine (**Tableau 12**).

L'association du rs2000999 a été répliquée et confirmée lors d'une phase 2 dans un échantillon familial de la cohorte STANISLAS, et les cohortes GENDAI et OBE (N=4 579, $p_{réplication}=3,5x10^{-41}$, $p_{méta-analyse}=8,09x10^{-59}$, **Tableau 11**).

Le rs2000999 a été très récemment associé au cholestérol total et au LDL-C dans des populations adultes : 4 200 sujets issus du consortium 'European Special Populations Research Network' (EUROSPAN, http://eurospan.gen-info.hr/publications.html) [Igl et al. 2010] et plus de 100 000 sujets Européens et non Européens réunis dans une large méta-analyse ayant permis de confirmer 95 variants génétiques fortement associés aux lipides [Teslovich et al. 2010]. Nous avons réaffirmé cette association dans 5 cohortes pédiatriques réunissant 8 946 enfants ($p_{cholestérol\ total}=0,002$, $p_{LDL}=0,0008$).

Au terme de cette étude, nous avons donc identifié un unique variant génétique, le rs2000999, expliquant quasiment la moitié de l'héritabilité génétique de l'haptoglobine (42%) et permettant de faire le lien entre inflammation et des facteurs de risque cardiovasculaire connus : les lipides. En pratique clinique, le fait que cette association soit si forte à un jeune âge souligne l'intérêt potentiel du génotypage du rs2000999 pour permettre l'identification précoce des sujets à risque concernant toutes les pathologies dans lesquelles l'haptoglobine est impliquée, et dans notre cas les MCV.

Tableau 11 : Phases de mise en évidence et réplication de l'association du rs2000999 avec le taux d'haptoglobine dans les populations pédiatriques

SNP	Gene	Population	Stage	N	MAF	11	12	22	β	SE	p brute	p ajustée*
rs2000999	HPR	STANISLAS	Discovery	601	0,21	0,736 ± 0,406	0,532 ± 0,321	0,361 ± 0,180	-0,123	0,017	6,32E-13	1,95E-07
		STANISLAS	Replication	3145	0,20	1,056 ± 0,467	0,819 ± 0,445	0,549 ± 0,330	-0,138	0,007	1,17E-75	
		OBE	Replication	1 015	0,22	1,261 ± 0,492	1,064 ± 0,455	1,002 ± 0,534	-0,088	0,013	1,20E-11	
		GENDAI	Replication	419	0,25	0,811 ± 0,450	0,813 ± 0,558	0,731 ± 0,319	-0,021	0,021	0,306	
		REPLICATION	Replication	4 579	0,22						3,49E-41	
		MA	Replication	5 180							8,09E-59	

* signification statistique après ajustement sur le SNP rs10492825

Tableau 12 : Estimation de l'héritabilité de l'haptoglobine dans la cohorte STANISLAS – composante de variance de Fisher

Modèle	σ^2G Variance polygénique (%)	σ^2C Envonnement partagé (%)	σ^2E Environnement spécifique (%)	Variance totale	Modèle alternatif χ^2 (df) P
1*	0,0478 ± 0,0120	0,0250 ± 0,0065	0,1194 ± 0,0078	0,1922	
	24,9 %	**13,0 %**	**62,1 %**		
2**	0,0253 ± 0,0112	0,0275 ± 0,0059	0,1224 ± 0,0076	0,1752	Modèle 1: 213,86 (4)
	14,4 %	**15,7 %**	**69,9 %**		p<0,001

* Sans ajustement sur le rs2000999

** Ajustement sur le rs2000999

2.2. Genome-Wide Association Study de la pression artérielle dans des populations pédiatriques. Effet de l'obésité.

Nous avons vu dans le chapitre précédent que l'identification de l'effet d'un variant pouvait être conditionnée par le type de population et surtout par l'effet de l'environnement sur cette population. Dans une étude non encore publiée, nous avons donc voulu mettre en évidence ce fait en illustrant la relation bien connue entre pression artérielle et obésité [Muntner et al. 2004].

Nous avons donc mené deux GWAS simultanées chez les 631 enfants minces de la cohorte STANISLAS et chez 1 015 enfants obèses issus de la population OBE. Dans les deux populations, aucun variant n'a atteint le seuil 'genome-wide significant' mais ceci n'avait rien de surprenant car la pression artérielle est un trait qui a toujours été difficile à investiguer en GWAS [Levy et al. 2007]. Ce n'est que très récemment que des variants génétiques ont été identifiés dans de très larges populations Européennes [Levy et al. 2009 ; Newton-Cheh et al. 2010] et Japonaises [Takeuchi et al. 2009].

Nous avons par la suite entrepris de répliquer *in silico*, dans nos populations pédiatriques, les résultats obtenus par les GWAS précédemment citées. Cette procédure visait à minimiser le nombre de tests multiples et donc à augmenter le seuil de signification.

À l'exception des gènes *CYP17A1* et *ITGA9*, tous les loci mis en évidence dans des populations adultes ont pu être répliqués dans nos deux populations pédiatriques, notamment *CCL20*, *STK39*, *CNTN4*, *CAMK4*, *C6orf166-SPACA1*, *TRPA1-KCNB2*, *C14orf118*, *CDH13, MECOM*, *C10orf107*, *PLEKHA7*, *TBX3-BX5*, *SH2B3*, *CSK-ULK3*, *ZNF652*, *CASZ1*, *MDS1*, *ATP2B1*, PLCD3, *MTHFR* et *ATP2B1* (**Tableau 13**). Néanmoins, nous avons observé que les SNPs spécifiques à une population ne l'étaient pas dans l'autre et ceci expliquait le fait que la méta-analyse regroupant l'échantillon STANISLAS et l'échantillon OBE ne permettaient d'observer aucune signification nominale. Ainsi, au sein d'un même loci, le variant impliqué dans la variabilité interindividuelle de la pression artérielle (systolique ou diastolique) était dépendant de la corpulence de la population considérée.

Le cas particulier du rs7556897 localisé sur le gène *CCL20* a également soulevé notre attention. En effet, ce variant était le seul à être associé à la PAD dans les deux populations

simultanément mais avec un effet opposé : une diminution de la pression chez les enfants minces (β=-1,27 ± 0,56 ; p=0,024) et une augmentation de la pression chez les enfants obèses (β=+2,18 ± 0,71 ; p=0,002) (**Tableau 13**). Cette constatation venait renforcer de précédentes études *in vitro* impliquant ce gène dans la physiopathologie de l'obésité : *CCL20* code pour une chémokine de 96 acides amines exprimée dans le foie, les lymphocytes et les adipocytes [Duffaut et al. 2009] et dont l'expression a été positivement corrélée à l'IMC. Une augmentation des niveaux d'ARN messager et de la protéine chez les enfants obèses peut donc expliquer en partie une augmentation parallèle de leur pression artérielle.

Du point de vue de la fonctionnalité, le locus *SUFU* a également démontré un intérêt car il a récemment été associé à l'adipogenèse [Pospisilik et al. 2010].

Ainsi, cette étude nous a permis de prouver que les variants génétiques liés à la pression artérielle avaient un effet continu tout au long de la vie ; mais surtout que cet effet pouvait être modulé par les répercussions que pouvait avoir l'environnement sur les individus concernés : en l'occurrence ici, sur l'obésité.
Il serait opportun de répliquer ces résultats dans de plus larges échantillons, de préférence pédiatriques, et d'effectuer des études de transcriptomique de façon à nous assurer que les associations et surtout que les disparités observées sont robustes.

Tableau 13 : Étude de réplication chez les enfants minces et obèses

SNP biblio	CHR	Gène à proximité	SNP en LD	*Gène à proximité(LD)*	p(STANISLAS)	P (OBE)	trait
rs10493340	1	N/A	rs1498937	*ATG4C -- FOXD3*	0,014	0,885	PAD
			rs953653	*ATG4C -- FOXD3*	0,030	0,229	PAD
rs7591163	2	*CCL20*	rs9288627	*SLC19A3 -- CCL20*	0,020	0,117	PAD
			rs7556897	*SLC19A3 -- CCL20*	0,024	0,002	PAD
			rs940339	*CCL20 -- WDR69*	0,857	0,026	PAS
			rs4973239	*SLC19A3 -- CCL20*	0,309	0,019	PAD
rs6749447	2	*STK39*	rs7595310	*STK39*	0,023	0,757	PAS
		serine threonine kinase 39	rs7597374	*STK39 -- LASS6*	0,030	0,853	PAS
			rs9646776	*STK39 -- LASS6*	0,863	0,045	PAD
rs4370013	3	*CNTN4 contactin 4*	rs6442742	*CNTN4*	0,033	0,869	PAD
			rs10510238	*CNTN4*	0,047	0,246	PAD
			rs4685527	*CNTN4*	0,639	0,046	PAS
			rs6768322	*CNTN4*	0,756	0,035	PAD
rs10491334	5	*CAMK4*		*CAMK4*	0,025	0,606	PAS
		calcium/calmodulin-dependent	rs25923	*CAMK4*	0,008	0,288	PAS
		protein kinase IV	rs3733995	*CAMK4*	0,047	0,963	PAS
			rs389402	*CAMK4*	0,867	0,010	PAD
			rs10515427	*CAMK4*	0,250	0,014	PAD
rs2509458	6	AL356970.1	rs9344757	*CNR1 66 RNGTT*	0,044	0,128	PAS
			rs1535987	*C6orf166 -- SPACA1*	0,013	0,392	PAD
			rs2787942	*C6orf166 -- SPACA1*	0,647	0,004	PAS
rs1963982	8	RP11-656G20.1	rs276648	*TRPA1 -- KCNB2*	0,019	0,577	PAD
			rs2701453	*TRPA1 -- KCNB2*	0,036	0,572	PAD
			rs276648	*TRPA1 -- KCNB2*	0,509	0,032	PAS
			rs4738223	*TRPA1 -- KCNB2*	0,207	0,045	PAS

rs2121070	14	C14orf118	rs8008940	*C14orf118 -- ESRRB*	0,016	0,934	PAS
			rs2360982	*C14orf118 -- ESRRB*	0,920	0,036	PAD
rs935334	14	ESRRB	rs8008940	*C14orf118 -- ESRRB*	0,016	0,934	PAS
			rs2360982	*C14orf118 -- ESRRB*	0,920	0,036	PAD
rs3096277	16	*CDH13*	rs12446894	*CDH13*	0,010	0,184	PAS
		cadherin 13, H-cadherin (heart)	rs9940922	*CDH13*	0,026	0,898	PAS
			rs11149606	*HSBP1 -- MLYCD*	0,547	0,028	PAS
			rs1035533	*CDH13*	0,955	0,041	PAS
rs9815354	3	*ULK4*	rs9847688	*TRAK1*	0,231	0,021	PAD
		unc-51-like kinase 4	rs17066958	*TRAK1*	0,515	0,039	PAD
rs1918974	3	*MECOM*	rs9859831	*MDS1*	0,003	0,389	PAD
			rs7622799	*MDS1*	0,618	0,013	PAS
rs1530440	10	*C10orf107*	rs10740055	*ARID5B*	0,016	0,880	PAS
			rs1992626	*C10orf107 -- ARID5B*	0,027	0,891	PAD
			rs1530439	*C10orf107 -- ARID5B*	0,738	0,004	PAS
			rs7090027	*C10orf107 -- ARID5B*	0,616	0,010	PAS
rs11024074	11	*PLEKHA7*	rs217752	*PLEKHA7*	0,544	0,042	PAS
		pleckstrin homology	rs2433750	*TBX3*	0,005	0,367	PAS
		domain containing,	rs1863723	*TBX3*	0,045	0,422	PAD
		family A member 7	rs10850391	*TBX3*	0,233	0,007	PAS
			rs10850391	*TBX3*	0,391	0,013	PAD
			rs10850375	*TBX3*	0,580	0,022	PAD
rs2384550	12	*TBX3 -- TBX5*	rs2433750	*TBX3*	0,005	0,367	PAS
			rs1863723	*TBX3*	0,045	0,422	PAD
			rs10850391	*TBX3*	0,233	0,007	PAS
			rs10850391	*TBX3*	0,391	0,013	PAD
			rs10850375	*TBX3*	0,580	0,022	PAD

rs653178	12	*SH2B3*	rs3847953	*CUTL2*	0,029	0,638	PAS
		SH2B adaptor protein 3	rs2239194	*SH2B3*	0,511	0,006	PAS
rs6495122	15	*CSK --ULK3*	rs11856413	*C15orf17 -- COX5A*	0,048	0,365	PAS
			rs8025170	*SCAMP2 -- MPI*	0,004	0,627	PAS
			rs936227	*CPLX3 -- SCAMP2*	0,446	0,037	PAS
rs16948048	17	*ZNF652*	rs4987082	*PHB*	0,034	0,205	PAS
			rs2293215	*GNGT2*	0,013	0,751	PAS
			rs935129	*PHB*	0,787	0,021	PAD
			rs6504599	*B4GALNT2*	0,508	0,043	PAD
rs12046278	1	*CASZ1*	rs284254	*CASZ1*	0,027	0,340	PAS
			rs880315	*CASZ1*	0,023	0,857	PAS
			rs284254	*CASZ1*	0,012	0,213	PAD
			rs4845948	*CASZ1*	0,915	0,045	PAD
			rs284262	*CASZ1*	0,280	0,048	PAD
rs448378	3	*MDS1*	rs9859831	*MDS1*	0,003	0,389	PAD
			rs7622799	*MDS1*	0,618	0,013	PAS
rs1004467	10	*CYP17A1*	rs7096169	*C10orf32*	0,030	0,890	PAS
			rs7086205	*CUEDC2 -- C10orf95*	0,329	0,011	PAS
			rs1163073	*INA*	0,885	0,035	PAS
			rs11191537	*CNNM2*	0,605	0,050	PAD
			rs4917978	*SUFU*	0,630	0,016	PAD
			rs11594179	*SUFU*	0,221	0,044	PAD
rs11191548	10	*CYP17A1*	rs7096169	*C10orf32*	0,030	0,890	PAS
			rs729211	*FAM26C*	0,017	0,443	PAS
			rs216497	*PLEKHA7*	0,044	0,502	PAS
			rs2863733	*NEURL*	0,009	0,682	PAD
			rs1163073	*INA*	0,885	0,035	PAS
			rs6584555	*NEURL*	0,923	0,030	PAS
			rs4917978	*SUFU*	0,630	0,016	PAD
			rs11594179	*SUFU*	0,221	0,044	PAD

rs381815	11	*PLEKHA7*	rs216497	*PLEKHA7*	0,044	0,502	PAS
			rs217752	*PLEKHA7*	0,544	0,042	PAS
rs2681492	12	*ATP2B1*	rs10506979	*ATP2B1*	0,038	0,058	PAS
rs12946454	17		rs1989229	*FMNL1*	0,001	0,889	PAD
			rs7223320	*PLCD3*	0,028	0,129	PAD
			rs12450321	*HEXIM2*	0,803	0,015	PAS
			rs11657325	*HEXIM2*	0,521	0,034	PAD
rs11014166	10	*ACBD4*	rs2482100	*CACNB2*	0,004	0,549	PAD
		acyl-CoA binding domain containing 4	rs10508561	*CACNB2*	0,959	0,021	PAS
			rs7895709	*NUSUN6*	0,823	0,030	PAD
rs3184504	12	*SH2B3*	rs3847953	*CUTL2*	0,029	0,638	PAS
			rs2239194	*SH2B3*	0,511	0,006	PAS
rs17367504	1	*MTHFR*	rs4845877	*AGTRAP -- MTHFR*	0,014	0,796	PAD
			rs1133398	*AGTRAP -- MTHFR*	0,024	0,551	PAD
			rs6676300	*NPPB -- KIAA2013*	0,613	0,028	PAS
			rs1801133	*MTHFR*	0,350	0,039	PAD
rs880315	1	*CASZ1*		*CASZ1*	0,023	0,857	PAS
			rs284254	*CASZ1*	0,027	0,340	PAS
rs155524	3	*ITGA9*					
rs16998073	4	*FGF5*	rs7657506	*ANTXR2 -- PRDM8*	0,025	0,984	PAS
			rs7667228	*ANTXR2*	0,015	0,741	PAD
			rs2290755	*FGF5 -- C4orf22*	0,524	0,004	PAD
			rs10518233	*ANTXR2 -- PRDM8*	0,067	0,026	PAD
rs12413409	10	*CYP17A1*	rs729211	*FAM26C*	0,017	0,443	PAS
			rs7096169	*C10orf32*	0,030	0,890	PAS
			rs4917978	*SUFU*	0,630	0,016	PAD
rs2681472	12	ATP2B1	rs10506979	*ATP2B1*	0,038	0,058	PAS
rs1378942	15	*CSK -- ULK3*	rs881536	*CPLX3 -- SCAMP2*	0,033	0,176	PAS

2.3. Genome-Wide Association Study de traits lipidiques dans des populations pédiatriques. Identification de variants inédits.

Les traits lipidiques ont largement été investigués dans le cadre des GWAS sur des populations adultes. Néanmoins, il n'existe à notre connaissance aucune étude ciblée sur des populations pédiatriques alors que ces traits subissent une influence très importante des facteurs environnementaux. Notre postulat était donc le suivant : il serait intéressant de s'assurer que les associations observées chez les adultes sont visualisables dès l'enfance mais aussi qu'il existe une sorte d'effet cumulatif des facteurs environnementaux qui masqueraient avec le temps l'effet de certains variants génétiques.

Nous avons donc mis en place une étude d'association pangénomique réunissant 9 800 enfants d'origine Européenne dans deux phases consécutives et 8 886 adultes dans une troisième phase de réplication.

La 'discovery cohort' s'est composée des enfants STANISLAS et OBE pour lesquels nous avons mené deux GWAS sur le cholestérol total, les HDL et LDL-C et les TG. À la suite d'une méta-analyse, nous avons effectué un SNPlex dans quatre populations pédiatriques : STANISLAS (échantillon non concerné par le GWS), NFBC86, GENDAI et Verona pour une phase 2. Nous avons ainsi pu démontrer que certaines associations génétiques mises en évidence chez l'adulte avaient un effet observable dès l'enfance comme dans le cas de la pression artérielle. Le rs780094 dans le locus *GCKR* ($p=3,70x10^{-14}$) a été associé aux TG [Willer et al. 2008, Aulchenko et al. 2009, Kathiresan et al. 2009]; et le gène *PCSK9* lié dans la littérature au LDL-C [Willer et al. 2008, Kathiresan et al. 2009] associé chez les enfants au cholestérol total ($p=3,96x10^{-05}$), aux TG ($p=1,04x10^{-03}$) et au HDL-C ($p=1,15x10^{-02}$) par l'intermédiaire du rs207136. Enfin, nous avons confirmé l'association du locus *TOMM40* proche du gène *APOE* avec le LDL-C ($p=3,25x10^{-13}$) et le cholestérol total ($p=7,39x10^{-10}$) [Aulchenko et al. 2009] via le rs2075650.

Mais nous avons surtout mis en évidence 11 variants inédits : rs10518422, rs2082163, rs11075365, rs4818278, rs2823986, rs4918082, rs2607919, rs933193, rs724815, rs12591800 et rs314672.

C'est ces 11 variants que nous avons recherché à répliquer dans une phase 3 réunissant deux populations adultes issues des cohortes STANISLAS et DESIR. Nous avons ainsi pu confirmer que 6 variants étaient spécifiques à l'enfance : le rs12591800 (*ATPBD4*, $p_{HDL}=7,41x10^{-05}$ et $p_{cholestérol\ total}=1,26x10^{-02}$); le rs4918082 (*PLCE1*, $p_{HDL}=2,62x10^{-03}$ et $p_{TG}=1,01x10^{-02}$); le rs2607919 (*NINJ2*, $p_{cholestérol\ total}=1,25x10^{-03}$ et $p_{HDL}=3,80x10^{-02}$); le rs933193 (*DAD1*, $p_{HDL}=3,77x10^{-03}$); le rs724815 (*NOVA1*, $p_{HDL}=1,10x10^{-02}$); et le rs314672 (*PRKD2*, $p_{TG}=3,18x10^{-02}$).

Ce derniers cas de figure peut s'expliquer par le fait que la croissance de l'effet de l'environnement sur un sujet donnée (avec l'âge) masque peu à peu l'effet de certains variants génétiques.

Cette étude en toujours en cours.

3. Identification de profils génétiques chez les sujets supposés sains de la cohorte STANISLAS : répercussions phénotypiques.

PUBLICATION N°5

Mais pour quelle raison est-il si important de mettre en évidence les variants génétiques des facteurs de risque des MCV chez des sujets encore sains … et jeunes ?

À terme, il serait profitable de mettre en évidence des profils génétiques permettant de sélectionner des sujets à risque sur lesquels cibler plus particulièrement des stratégies préventives adaptées.

Ainsi, nous avons recherché chez 1 551 sujets de la cohorte STANISLAS des profils génétiques ou clusters explicatifs des niveaux de facteurs de risque divers : anthropométrie, pression artérielle, lipides, consommation de tabac et d'alcool. Pour se faire, nous avons utilisé une approche gène candidat au sein de la cohorte STANISLAS.

À l'aide d'une combinaison factorielle sur 27 variants génétiques connus, nous avons isolé six clusters génétiques (**Figure 15**) ayant un potentiel discriminant indépendant de l'âge sur les facteurs de risque étudiés en relation (**Tableaux 14 et 15**). Ces clusters ou profils génétiques ont été nommés C_{1453}, C_{44}, C_{16}, C_{14}, C_{12} et C_{10} en fonction de leurs effectifs.

- C_{10} était très corrélé à une PAD et des concentrations en cholestérol total, TG, ApoB, ApoE (et LDL-C uniquement à t_0) élevées et à des concentrations en ApoA-I et HDL-C faibles ;
- Au contraire les sujets appartenant à C_{16} avaient des concentrations en cholestérol total, LDL-C (à t0), TG (à t5), ApoB, ApoE faibles et des concentrations en ApoA-I et HDL-C élevées ;
- C_{44} était associé à de faibles consommations d'alcool et de tabac ;
- Et les individus du profil C_{12} étaient plus minces (tour de taille, tour de hanche, IMC, poids faibles) et avaient de faibles concentrations en glucose, cholestérol total et ApoB à t_5.

Ainsi, la détermination de clusters génétiques 'à risque' cardiovasculaire permet en pratique d'identifier un groupe d'individus pour lesquels les phénotypes concernés seront plus élevés.

Il serait intéressant de poursuivre cette approche en combinant les résultats GWAS avec ceux des études gène-candidat antérieures.

Figure 15 : Signature des clusters génétiques identifiés

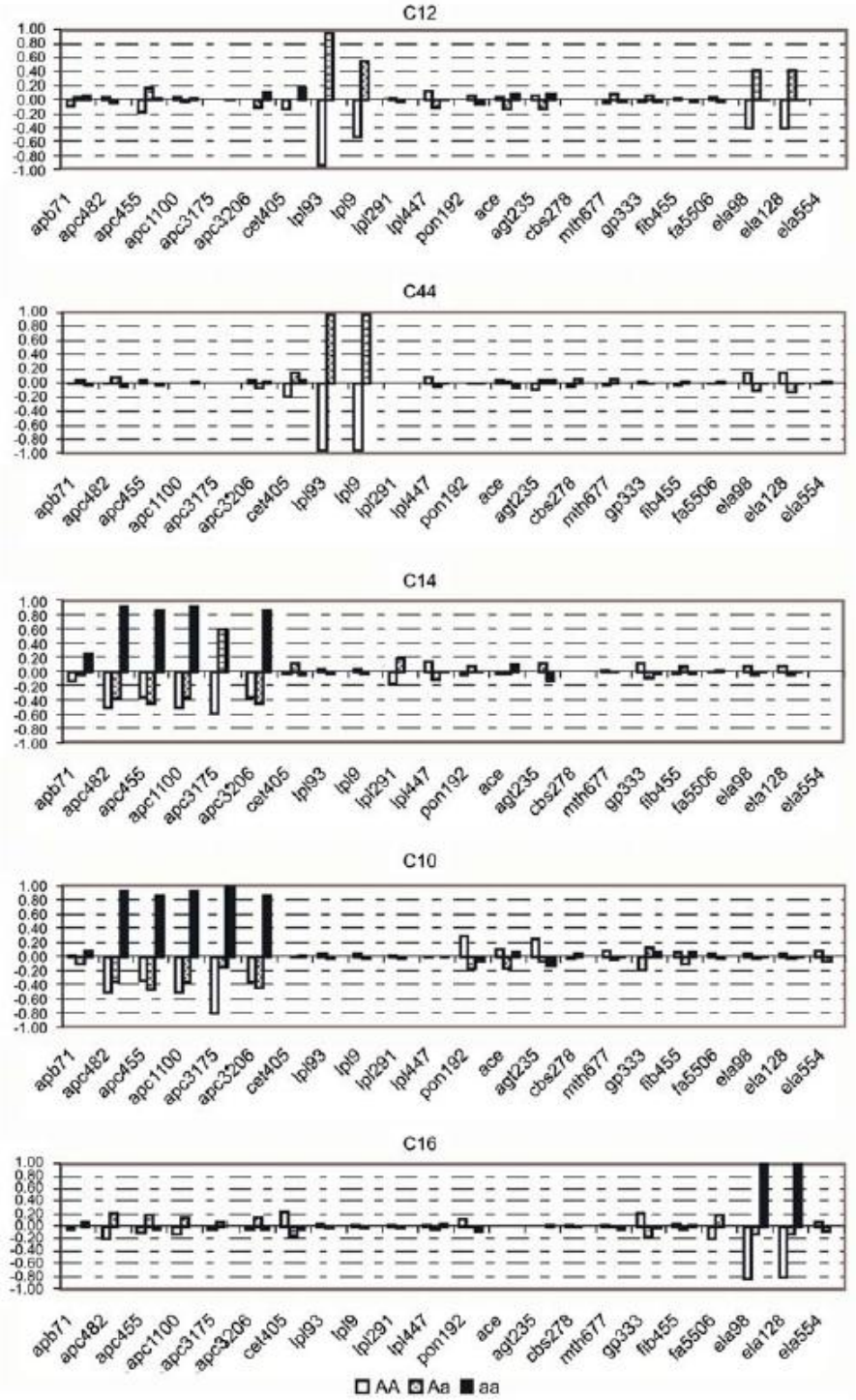


Tableau 14 : Évaluation clinique et mesures biologiques au sein des clusters au premier examen de santé (t_0)

t0, moyenne (SD)	C1549 (n=1549)	C1553 (n=1553)	C12 (n=12)	C44 (n=44)	C14 (n=14)	C10 (n=10)	C16 (n=16)	p-value
Age, années	40,50 (4,60)	40,44 (4,57)	40,08 (4,03)	42,14 (5,32)	39,57 a (4,26)	43,50 b (4,88)	41,31 (4,32)	0,101
Taille, cm	167,72 (8,99)	167,68 (9,01)	165,53 (7,97)	169,66 (8,17)	163,56 a (8,48)	169,57 (9,91)	169,99 b (8,47)	0,238
Poids, kg	68,88 (13,03)	68,76 (12,94)	64,01 a (9,16)	70,81 (12,65)	70,60 (22,83)	74,30 b (15,65)	72,45 (10,87)	0,319
IMC, kg/m²	24,37 (3,51)	24,35 (3,48)	23,46 a (3,62)	24,45 (2,91)	25,97 b (6,39)	25,62 (3,49)	25,08 (3,40)	0,639
Tour de taille, cm	81,25 (10,98)	81,12 (10,93)	78,17 a (10,66)	83,86 (10,45)	83,86 (18,19)	86,10 b (11,49)	83,19 (8,70)	0,233
Tour de hanche, cm	97,73 (6,71)	97,73 (6,67)	95,25 a (5,46)	96,82 (5,89)	98,64 (11,93)	99,20 (7,91)	99,81 b (6,88)	0,499
Ratio taille/hanche, 1	0,83 (0,09)	0,83 (0,09)	0,82 a (0,10)	0,86 (0,08)	0,85 (0,12)	0,87 b (0,08)	0,83 (0,07)	0,12
PAS, mmHg	122,67 (12,64)	122,54 (12,47)	123,42 (12,83)	125,02 (11,65)	122,57 (8,20)	135,00 b (32,01)	119,69 a (11,28)	0,626
PAD, mmHg	74,81 (9,78)	74,79 (9,69)	71,83 a (9,72)	75,95 (9,29)	72,21 (8,47)	84,30 b (18,79)	71,94 (10,24)	0,020**
Consoma lcool, g/jour	13,96 (21,91)	13,55 a (21,40)	23,96 (31,43)	14,99 (19,74)	14,64 (21,48)	23,44 (29,05)	34,21 b (43,47)	0,122
Consom tabac act, cig/jour	4,25 (8,39)	4,25 (8,36)	4,25 (11,51)	3,00 (6,57)	5,57 (10,56)	2,90 a (4,86)	7,19 b (11,96)	0,805
Consom tabac pas, cig/jour	4,27 (9,43)	4,14 (9,19)	6,58 (12,20)	5,86 (9,70)	11,07 b (21,97)	9,10 (13,88)	1,31 a (4,51)	0,144
Glycémie, mmol/L	5,07 (0,512)	5,07 (0,51)	4,93 a (0,26)	5,08 (0,54)	5,16 (0,73)	5,29 (0,45)	5,31 b (0,40)	0,103
Triglycérides, mmol/L	1,06 (0,67)	1,04 a (0,65)	1,19 (0,54)	1,31 (0,86)	1,36 (1,10)	1,77 b (0,96)	1,12 (0,92)	0,009***
Cholestérol total, mmol/L	5,73 (1,00)	5,72 (1,00)	5,96 (1,02)	5,82 (0,97)	5,65 (0,91)	6,64 b (0,82)	5,61 a (1,01)	0,059*
LDL-cholestérol, mmol/L	3,83 (0,95)	3,82 (0,95)	4,02 (0,95)	3,95 (0,92)	3,75 (0,82)	4,65 b (1,01)	3,58 a (1,07)	0,059*
HDL-cholestérol, mmol/L	1,41 (0,40)	1,42 (0,40)	1,39 (0,30)	1,27 (0,43)	1,27 (0,39)	1,18 a (0,26)	1,53 b (0,61)	0,014**
ApoB, g/L	1,06 (0,25)	1,05 (0,25)	1,14 (0,21)	1,10 (0,27)	1,06 (0,27)	1,31 b (0,27)	1,02 a (0,28)	0,042**
ApoAI, g/L	1,62 (0,26)	1,63 (0,26)	1,60 (0,21)	1,54 (0,25)	1,53 (0,24)	1,51 a (0,17)	1,65 b (0,25)	0,111

APOE, mg/L	38,41 (10,96)	38,24 a (10,80)	38,80 (5,48)	41,67 (15,08)	40,35 (15,11)	42,16 b (13,13)	40,12 (9,04)	0,708

Tableau 15 : Évaluation clinique et mesures biologiques au sein des clusters au second examen de santé (t_5)

t5, moyenne (SD)	C1549 (n=1128)	C1553 (n=1052)	C12 (n=10)	C44 (n=34)	C14 (n=11)	C10 (n=8)	C16 (n=13)	p-value
Age, années	45,55 (4,64)	45,50 (4,63)	43,70 a (3,47)	47,50 b (4,87)	44,18 (4,75)	47,50 b (4,72)	46,62 (4,61)	0,081*
Taille, cm	167,66 (9,04)	167,62 (9,05)	166,17 (8,01)	169,36 (8,71)	162,45 a (8,61)	171,14 b (10,01)	169,85 (9,38)	0,274
Poids, kg	71,15 (13,67)	71,07 (13,65)	64,39 a (7,65)	72,69 (13,25)	68,05 (15,61)	78,93 b (19,20)	76,96 (12,25)	0,148
IMC, kg/m²	25,20 (3,82)	25,19 (3,84)	23,39 a (2,96)	25,19 (3,14)	25,53 (4,20)	26,61 (4,29)	26,72 b (4,06)	0,39
Tour de taille, cm	83,10 (11,52)	83,01 (11,54)	77,10 a (7,33)	85,29 (11,32)	82,64 (12,50)	87,88 b (12,91)	86,42 (10,20)	0,212
Tour de hanche, cm	97,89 (7,20)	97,88 (7,25)	93,90 a (5,38)	97,71 (5,94)	97,91 (7,84)	97,25 (5,95)	102,25 b (6,65)	0,118
Ratio taille/hanche, 1	0,85 (0,09)	0,85 (0,09)	0,83 a (0,093)	0,87 (0,09)	0,84 (0,11)	0,90 b (0,09)	0,84 (0,08)	0,397
PAS, mmHg	124,19 (13,73)	123,97 (13,66)	129,10 (12,56)	126,79 (12,70)	122,48 a (10,26)	137,79 b (24,75)	124,67 (13,26)	0,161
PAD, mmHg	74,65 (10,16)	74,51 a (10,14)	75,60 (12,35)	76,19 (8,84)	74,61 (9,51)	84,67 b (11,47)	75,05 (11,22)	0,205
Consoma lcool, g/jour	13,60 (19,61)	13,39 (19,50)	15,85 (19,72)	12,79 a (16,94)	17,04 (20,94)	27,62 b (29,33)	19,24 (26,11)	0,494
Consom tabac act, cig/jour	3,28 (7,79)	3,20 (7,72)	5,30 (12,60)	2,74 a (6,21)	8,73 b (12,36)	6,13 (9,11)	3,38 (6,64)	0,429
Consom tabac pas, cig/jour	5,31 (10,25)	5,40 (10,31)	4,00 (8,43)	4,15 (8,69)	4,09 (8,01)	7,50 b (17,53)	1,69 a (5,53)	0,862
Glycémie, mmol/L	5,08 (0,75)	5,07 (0,74)	4,97 a (0,74)	5,34 (0,98)	5,32 (1,04)	5,50 b (0,59)	5,05 (0,48)	0,133
Triglycérides, mmol/L	1,27 (1,31)	1,25 (1,33)	1,24 (0,72)	1,55 (0,97)	1,56 (1,20)	2,43 b (1,04)	1,06 a (0,43)	0,005***
Cholestérol total, mmol/L	5,77 (0,96)	5,77 (0,97)	5,51 a (0,88)	5,83 (0,90)	5,95 (0,97)	6,49 b (0,90)	5,69 (0,70)	0,337
LDL-cholestérol, mmol/L	3,61 (0,95)	3,60 (0,95)	3,47 (0,72)	3,72 (0,92)	3,78 (0,95)	4,18 b (0,93)	3,39 a (0,63)	0,408
HDL-cholestérol, mmol/L	1,58 (0,45)	1,59 (0,44)	1,48 (0,50)	1,39 (0,47)	1,45 (0,37)	1,20 a (0,34)	1,81 b (0,68)	0,003***
ApoB, g/L	1,026 (0,237)	1,02 (0,24)	0,99 a (0,22)	1,06 (0,25)	1,09 (0,26)	1,29 b (0,25)	1,00 (0,21)	0,080*
ApoAI, g/L	1,620 (0,250)	1,62 (0,25)	1,55 (0,23)	1,52 (0,22)	1,61 (0,24)	1,51 a (0,22)	1,75 b (0,31)	0,054*

APOE, mg/L	40,62 (13,82)	40,50 (13,94)	39,30 (7,86)	41,57 (12,55)	39,06 a (10,41)	48,86 b (17,72)	44,41 (10,10)	0,511

4. Quels sont les challenges méthodologiques qu'attendent les GWAS ?

PUBLICATIONS 1, 6 & 7

Les GWAS ont incontestablement révolutionné la manière d'appréhender l'épidémiologie génétique. Néanmoins, cette approche en est encore à ses balbutiements et de nombreux challenges méthodologiques restent à relever.

Comme évoqué précédemment, les facteurs environnementaux ont un effet plus ou moins important sur l'héritabilité génétique des facteurs de risque des pathologies cardiovasculaires. Ces interactions gène-environnement ne sont cependant que très rarement investiguées. Pourtant, elles sont en grande partie responsables des discordances observées. La non réplication dans une population lambda d'une association trouvée ailleurs peut être le reflet de l'effet de facteurs environnementaux différents.

Une autre zone d'ombre méthodologique de telles études est la prise en compte des interactions épistasiques (interactions gène-gène). Nous avons conscience sur ce dernier point de la grande difficulté méthodologique que pourrait représenter la prise en compte de tels effets dans une perspective de GWAS. Lors de la 5ème conférence de Santorini 'Functional Genomics Towards Personalized Health Care' qui a eu lieu en Grèce en Septembre dernier, le Professeur Ioannidis exposait que la prise en compte de toutes les combinaisons possibles dans 12 000 000 variants génétiques correspondait à mener 10^{2085} analyses et porterait donc le seuil de genome-wide significance à $p=10^{-2087}$. Dans ces conditions, force est de constater que les approches gène-candidat ne font pas encore partie du passé.

Enfin, les GWAS permettent d'identifier des variants génétiques associés à des traits ou à des pathologies mais à ce jour, seulement 12% des SNPs identifiés sont localisés à proximité de régions codant pour des protéines [Hindorff et al. 2009]. Quarante pourcent se trouvent dans des régions intergéniques et 40 autres pourcent dans des introns. Ce phénomène peut bien sûr signifier que ces zones introniques et intergéniques pourraient avoir un rôle régulateur de l'expression de certains gènes impliqués [Hardy et al. 2009].

Ceci souligne la nécessité de placer les GWAS dans une démarche intégrée permettant de passer à une approche gène-candidat lorsqu'un design plus complexe est nécessaire à une meilleure compréhension des associations observées (et ce en attendant les avancées technologiques à venir), et allant jusqu'à une approche transcriptomique comme cela a été le cas pour notre étude pangénomique du VEGF.

4.1. Approche épistasique de la variabilité génétique de la pression artérielle.

PUBLICATION N°6

Comme nous l'avons expliqué dans le paragraphe introductif de ce chapitre, une GWAS prenant en compte les effets épistasiques serait irréalisable à l'état actuel des connaissances et des outils à notre disposition. Néanmoins, pour illustrer le fait que la prise en compte de ces interactions gène-gène peux permettre de mettre en évidence des associations non significatives en individuel, nous avons entrepris d'étudier les interactions épistasiques de 100 gènes candidats de diverses voies métaboliques liées aux facteurs de risque cardiovasculaires : la régulation de la pression artérielle, l'inflammation, le métabolisme lipidique, le métabolisme de l'homocystéine, l'adhésion cellulaire et la thrombose en rapport avec la pression artérielle chez 1 912 individus d'âge moyen et d'origine française issus de notre CRB IGE-PCV (**Tableau 16**). Ces sujets étaient soit non hypertendus ne prenant donc aucune médication antihypertensive, soit hypertendus et dans ce dernier cas de figure, leur données cliniques étaient déterminées avant l'administration de médication. Ce sont ces données sans effet de la thérapie qui ont été utilisées dans notre étude.

Dans un premier temps, nous avons mené des études d'association individuelle de chaque variant ayant une MAF supérieure ou égale à 2% et ayant un HWE conforme. La puissance statistique de chaque analyse (1-β) a été calculée *a posteriori* et seules les associations observes pour des puissances supérieures ou égales à 80% ont été considérées significatives. Nous avons ainsi pu mettre en évidence neuf SNPs significativement associés à la pression artérielle dont 5 fonctionnels: les rs5742910 et rs6046 dans le gène *FVII* ($p \leq 0,024$), le rs1800469 dans *TGFB1* ($p < 0,001$), le rs1800590 dans le *LPL* ($p \leq 0,003$) et le rs2228570 dans *VDR* ($P < 0.001$).

Par la suite, nous avons analysé les interactions épistasiques de la totalité des variants quelques soient leur MAF. Toutes les combinaisons deux à deux possibles ont été testées et ceci nous a permis de mettre en évidence les interactions épistasiques fonctionnelles suivantes : *LPL* (rs1801177)**PON1* (rs662) ($p \leq 0,032$), *VCAM1* (rs1041163)**APOB* (rs1367117) ($p \leq 0,04$), *SCGB1A1* (rs3741240)**LPL* (rs1800590 et rs1801177) ($p \leq 0,031$). Il était à noter que les variants rs3741240 et rs1801177 n'étaient pas significativement associés à la pression

artérielle lors des analyses d'association individuelles et que c'est leurs combinaisons épistasique qui ont permis de mettre en évidence leur rôle sur la pression systolique et diastolique.

Cette étude a ainsi pu démontrer les interconnexions existant entre différentes voies métaboliques et de ce fait mettre en évidence de nouvelles cibles thérapeutiques impliquées dans la physiopathologie de l'HTA.

Tableau 16 : Caractéristiques phénotypiques de la population étudiée

Characteristiques	Total (n = 1 912)	Men (n = 932)	Women (n = 980)	p
Age (années)	51,13 ± 10,02	51,21 ± 9,31	51,06 ± 10,66	0,740
IMC (kg/m²)	26,01 ± 3,75	26,70 ± 3,26	25,35 ± 4,06	<0,001
PAS (mmHg)	142,86 ± 21,47	144,62 ± 19,69	141,18 ± 22,92	<0,001
PAD (mmHg)	87,49 ± 14,26	89,10 ± 13,10	85,96 ± 15,14	<0,001
PP (mmHg)	62,18 ± 15,11	61,89 ± 14,34	62,46 ± 15,81	0,414
Classes tensionnelles (%)				
< 120/80	18,5	12,4	24,3	
120-139/80-89	21,1	25,3	17,1	
140-159/90-99	25,9	26,5	25,3	
≥ 160/100	34,5	35,7	33,3	
Tabagisme (%)				
Actuel	18,3	22,8	14,1	
Ancien	20,6	30,6	11,0	
Jamais	61,1	46,6	74,9	
Cholestérol total (mmol/L)	5,89 ± 1,00	5,93 ± 0,97	5,85 ± 1,01	0,082
LDL-C (mmol/L)	3,60 ± 0,88	3,70 ± 0,88	3,52 ± 0,88	<0,001
HDL-C (mmol/L)	1,64 ± 0,43	1,50 ± 0,40	1,77 ± 0,41	<0,001
Triglycérides (mmol/L)	1,37 ± 1,20	1,55 ± 1,00	1,20 ± 1,34	<0,001
Glucose (mmol/L)	5,25 ± 0,78	5,41 ± 0,83	5,10 ± 0,71	<0,001

4.2. Étude d'association pangénomique du VEGF et identification des variants responsable de sa forte héritabilité génétique. Vers une approche intégrée en transcriptomique.

PUBLICATION N°7

Nous avons vu précédemment que le VEGF, qui a été dosé dans un échantillon restreint de la cohorte STANISLAS, avait une très forte héritabilité génétique (60,6%). Cette héritabilité ne s'expliquait pas par les gènes candidats dont nous disposions à l'époque et nous avons donc choisi de les identifier par une approche 'hypothesis free'.

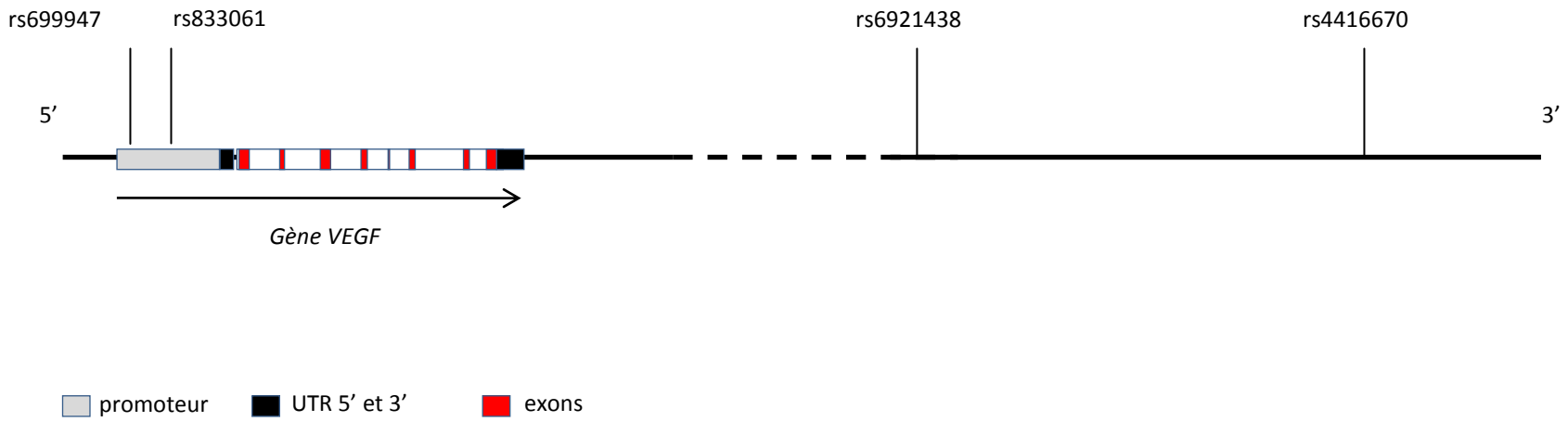
Nous avons ainsi entrepris une GWAS chez 3 546 sujets issus de la Framingham Heart Study (discovery set). Cent quarante SNPs localisés dans les régions 6p21, 8q23.1 et 9p24.2 ont atteint la 'genome-wide significance', dont le plus significatif : le rs6921438 sur le chromosome 6p21.1 ($p=6,11x10^{-506}$) localisé à 171 kilo bases du gène *VEGF* et proche du gène de la *Mitochondrial Ribosomal Protein L14* (*MRPL14*) et de *MCG45491* (*C6orf223*) (**Figure 16**).

Les régressions conditionnées ont permis d'identifier trois SNPs supplémentaires ayant un effet propre sur la variabilité interindividuelle du VEGF : rs4416670 ($p=1,47x10^{-12}$), rs6993770 ($p=2,50x10^{-16}$) et rs10738760 ($p=1,96x10^{-34}$).

Ces quatre variants (rs6921438, rs4416670, rs6993770 et rs10738760) ont permis d'expliquer 47,6% de l'héritabilité génétique globale du VEGF.

Lors d'une seconde phase, nous avons mené une étude de réplication dans deux échantillons indépendants : 859 adultes issus de la cohorte STANISLAS et 868 adultes Suédois issus de la cohorte PIVUS (**Tableau 17**). À la suite des méta-analyses, 19 SNPs restaient associés à la concentration en VEGF ($p<0,05$). Ces associations étaient inchangées après ajustement sur l'HTA, le tabagisme, l'obésité centrale et le SM dans les 3 populations.

Figure 16 : Organisation génomique du gène VEGF



Nous avons précédemment déploré le manque d'information sur la fonctionnalité des variants identifiés dans le cadre des GWAS. Nous avons donc intégré à notre approche une étude transcriptomique et quantifié l'expression de deux variants d'epissage de l'ARNm correspondants à deux isoformes du VEGF : $VEGF_{121}$ et $VEGF_{165}$. Cette expression a été quantifiée dans les PBMCs de 220 sujets issus de la cohorte STANISLAS et la corrélation des variants $VEGF_{121}$ et $VEGF_{165}$ a donc été testée avec les 24 variants génotypés avec succès dans la cohorte STANISLAS (**Tableau 18**). Ainsi, le rs910611 localisé sur le chromosome 6p21.1, les rs16873365, rs16873402, rs6993770 (indépendamment associé au VEGF, $p=2,50x10^{-16}$) et rs16873291 sur le chromosome 8q23.1 et le rs2375980 sur le chromosome 9p24.2 ont été significativement associés au variant $VEGF_{121}$.

Ces résultats soulignent donc la nécessité d'investiguer plus avant les facteurs de risque cardiovasculaires dont l'héritabilité, connue, ne s'explique pas par l'approche gène-candidat mais aussi la nécessité d'intégrer les approches pangénomiques dans des démarches globales permettant de déterminer la fonctionnalité de chaque variant génétique identifié.

Tableau 17 : SNPs 'genome-wide significants' à la suite de la réplication

SNP	Chr	Position*	MAF	MA	brin	beta† (FHS)	SE (FHS)	p (FHS)	p (STAN)	p (PIVUS)	Meta-p (replication)	Meta-p (tot)	Gene1	Distance (kb)	Gene2	Distance (kb)
rs6921438	6	44033585	0.51	G	+	0.72	0.01	$6.11x10^{-506}$	$2.93x10^{-39}$	NA	$2.93x10^{-39}$	$1.38x10^{-527}$	MGC45491	42.7	MRPL14	155.8
rs4513773	6	44033504	0.47	G	+	-0.70	0.01	$2.08x10^{-482}$	NA	$6.23x10^{-139}$	$6.23x10^{-139}$	$4.45x10^{-619}$	MGC45491	42.8	MRPL14	155.8
rs9472159	6	44027673	0.50	C	+	0.76	0.02	$4.30x10^{-452}$	$4.11x10^{-35}$	$3.27x10^{-109}$	$3.99x10^{-134}$	$6.54x10^{-570}$	MGC45491	48.6	MRPL14	161.7
rs9369434	6	44026385	0.53	C	+	0.84	0.02	$2.15x10^{-442}$	$5.31x10^{-28}$	$1.43x10^{-63}$	$2.71x10^{-86}$	$3.91x10^{-490}$	MGC45491	49.9	MRPL14	163.0
rs1776717	6	44059314	0.21	A	+	-0.23	0.02	$8.10x10^{-20}$	$9.75x10^{-6}$	$3.74x10^{-4}$	$1.44x10^{-8}$	$2.29x10^{-26}$	MGC45491	17.0	MRPL14	130.0
rs1776721	6	43998961	0.31	T	+	-0.18	0.02	$1.52x10^{-19}$	0.02	$3.43x10^{-8}$	$5.17x10^{-8}$	$6.70x10^{-26}$	MGC45491	77.3	VEGF	136.8
rs1886979	6	44012879	0.41	G	+	0.17	0.02	$3.71x10^{-19}$	0.01	$3.23x10^{-6}$	$6.73x10^{-7}$	$3.77x10^{-24}$	MGC45491	63.4	VEGF	150.7
rs9472155	6	44005705	0.22	T	+	-0.20	0.02	$4.45x10^{-19}$	0.01	$3.93x10^{-9}$	$7.52x10^{-9}$	$2.58x10^{-26}$	MGC45491	70.6	VEGF	143.5
rs844294	6	44008685	0.52	C	+	-0.15	0.02	$1.19x10^{-14}$	0.09	$2.25x10^{-5}$	$4.25x10^{-5}$	$5.92x10^{-18}$	MGC45491	67.6	VEGF	146.5
rs4416670	6	44058431	0.55	T	+	0.13	0.02	$1.47x10^{-12}$	$2.87x10^{-4}$	0.10	$1.41x10^{-4}$	$1.80x10^{-15}$	MGC45491	17.9	MRPL14	130.9
rs910611	6	44058829	0.08	C	+	-0.26	0.04	$2.61x10^{-10}$	0.11	$6.36x10^{-6}$	$2.83x10^{-5}$	$6.58x10^{-14}$	MGC45491	17.5	MRPL14	130.5
rs6993770	8	106650704	0.32	T	+	-0.17	0.02	$2.50x10^{-16}$	0.02	$3.99x10^{-8}$	$9.96x10^{-8}$	$1.51x10^{-22}$	ZFPM2	0	LRP12	980.4
rs16873402	8	106658423	0.33	T	+	-0.15	0.02	$1.97x10^{-14}$	0.16	$9.49x10^{-9}$	$1.86x10^{-6}$	$2.14x10^{-19}$	ZFPM2	0	LRP12	988.1
rs16873365	8	106627411	0.22	T	+	-0.16	0.02	$5.65x10^{-12}$	0.37	$2.09x10^{-6}$	$1.57x10^{-4}$	$4.06x10^{-15}$	ZFPM2	0	LRP12	957.1
rs7013321	8	106662734	0.49	A	+	-0.14	0.02	$6.75x10^{-12}$	0.01	NA	0.0144	$9.51x10^{-13}$	ZFPM2	0	LRP12	992.4
rs6993696	8	106650460	0.46	A	+	-0.13	0.02	$8.54x10^{-12}$	0.05	$1.49x10^{-4}$	$6.18x10^{-5}$	$3.33x10^{-15}$	ZFPM2	0	LRP12	980.1
rs16873291	8	106597206	0.31	T	+	-0.13	0.02	$5.30x10^{-11}$	0.07	$7.65x10^{-7}$	$4.08x10^{-6}$	$1.07x10^{-15}$	ZFPM2	0	LRP12	926.9

rs1349319	8	106625810	0.39	A	+	0.11	0.02	$3.59x10^{-8}$	0.05	$1.32x10^{-3}$	$3.03x10^{-4}$	$4.86x10^{-11}$	ZFPM2	0	LRP12	955.5
rs10738760	9	2681186	0.49	A	+	0.28	0.02	$1.96x10^{-34}$	0.03	$1.12x10^{-8}$	$5.69x10^{-8}$	$3.91x10^{-37}$	KCNV2	26.3	VLDLR	36.7
rs6475920	9	2663933	0.36	A	+	-0.24	0.02	$3.76x10^{-32}$	0.02	$2.40x10^{-8}$	$5.94x10^{-8}$	$6.55x10^{-37}$	VLDLR	19.4	KCNV2	43.6
rs4741756	9	2658187	0.28	C	+	-0.25	0.02	$2.95x10^{-31}$	0.09	$8.64x10^{-5}$	$1.24x10^{-4}$	$4.08x10^{-31}$	VLDLR	13.7	KCNV2	49.3
rs2375980	9	2682622	0.42	G	+	-0.25	0.02	$1.30x10^{-27}$	0.02	$2.25x10^{-8}$	$4.67x10^{-8}$	$5.50x10^{-6}$	KCNV2	24.9	VLDLR	38.1
rs10122587	9	2681951	0.28	T	+	-0.22	0.02	$3.02x10^{-24}$	0.02	NA	0.0243	$1.81x10^{-23}$	KCNV2	25.6	VLDLR	37.5
rs10967492	9	2671175	0.21	A	+	-0.22	0.02	$1.02x10^{-21}$	0.10	NA	0.104	$6.07x10^{-20}$	VLDLR	26.7	KCNV2	36.3
rs10967470	9	2665698	0.24	G	+	-0.22	0.02	$1.17x10^{-21}$	0.04	NA	0.0442	$1.56x10^{-20}$	VLDLR	21.2	KCNV2	41.8

Tableau 18 : Corrélations significatives entre les variants génotypés dans la cohorte STANISLAS et les transcripts VEGF

phénotype	SNP	Chr	position	MA	MAF	beta †	SE	p	h2q (%)
ARNm_121	rs16873365	8	106627411	T	0,22	22,71	7,22	0,002	4,73
ARNm_121	rs16873402	8	106658423	T	0,33	12,15	5,10	0,017	2,84
ARNm_121	rs6993770	8	106650704	T	0,32	12,06	5,23	0,021	2,82
ARNm_121	rs16873291	8	106597206	T	0,31	11,95	5,37	0,026	2,47
ARNm_121	rs2375980	9	2682622	G	0,42	10,15	4,75	0,032	2,03
ARNm_121	rs910611	6	44058829	C	0,08	-19,47	9,49	0,040	2,13

CHAPITRE 5 – CONCLUSIONS ET PERSPECTIVES

Tout au long des trois années de nos travaux de recherche, nous nous sommes attachées à acquérir les compétences méthodologiques et logistiques nous permettant d'intégrer l'ère des études pangénomiques.

Notre équipe d'accueil, l'EA4373 'Génétique Cardiovasculaire', a une longue expérience de l'épidémiologie étiologique des MCV et a pu mettre en évidence de nombreux variants génétiques associés aux facteurs de risque de ces pathologies avec comme principal outil d'investigation la cohorte STANISLAS.

Jusqu'en 2008, l'approche gène-candidat était la principale méthodologie utilisée et notre intégration dans cette équipe visait donc à initier une approche pangénomique. Cette approche serait à intégrer dans une stratégie globale de l'équipe réunissant l'ensemble des compétences de nos chercheurs au sein du projet BioIntelligence soutenu par l'Agence d'Innovation Industrielle (A2I) et signé en Avril 2010 avec l'INSERM transfert. Nous sommes tenus à la confidentialité en ce qui concerne les modalités de ce projet, néanmoins il est important de noter qu'il s'agit d'un projet collaboratif, financé par OSEO, et visant à proposer aux industriels des sciences de la vie, et en particulier aux industriels pharmaceutiques, des biomarqueurs pertinents en pharmacogénétique et pharmacogénomique à intégrer dans un environnement logiciel.

Dans un premier temps, nous nous sommes attachées à identifier les manquements méthodologiques que comportaient de telles approches : l'absence de prise en compte des interactions gène-gène, pour des raisons de logistique principalement, mais surtout l'absence de prise en compte des interactions gène-environnement. Ce dernier point étant complexifié par le fait que les GWAS nécessitent, pour valider les variants mis en évidence, des phases de réplication dans de très larges populations indépendantes les unes des autres et que cette diversité implique une 'non standardisation' des méthodologies de quantification de ces facteurs d'environnement. Notre premier postulat était donc qu'il serait nécessaire de mener des GWAS dans des populations minimisant l'influence de ces facteurs environnementaux : les populations pédiatriques. En effet, de telles GWAS menées sur les facteurs de risque des MCV sont minoritaires et principalement axées sur l'étude des marqueurs de l'obésité et du diabète de type 2 [Meyre et al. 2009 ; Bouatia-Naji et al. 2009 ; Qu et al. 2010 ; Scherag et al. 2010 ; Speliotes et al. 2010 ; Dupuis et al. 2010 ; Meyre et al. 2010].

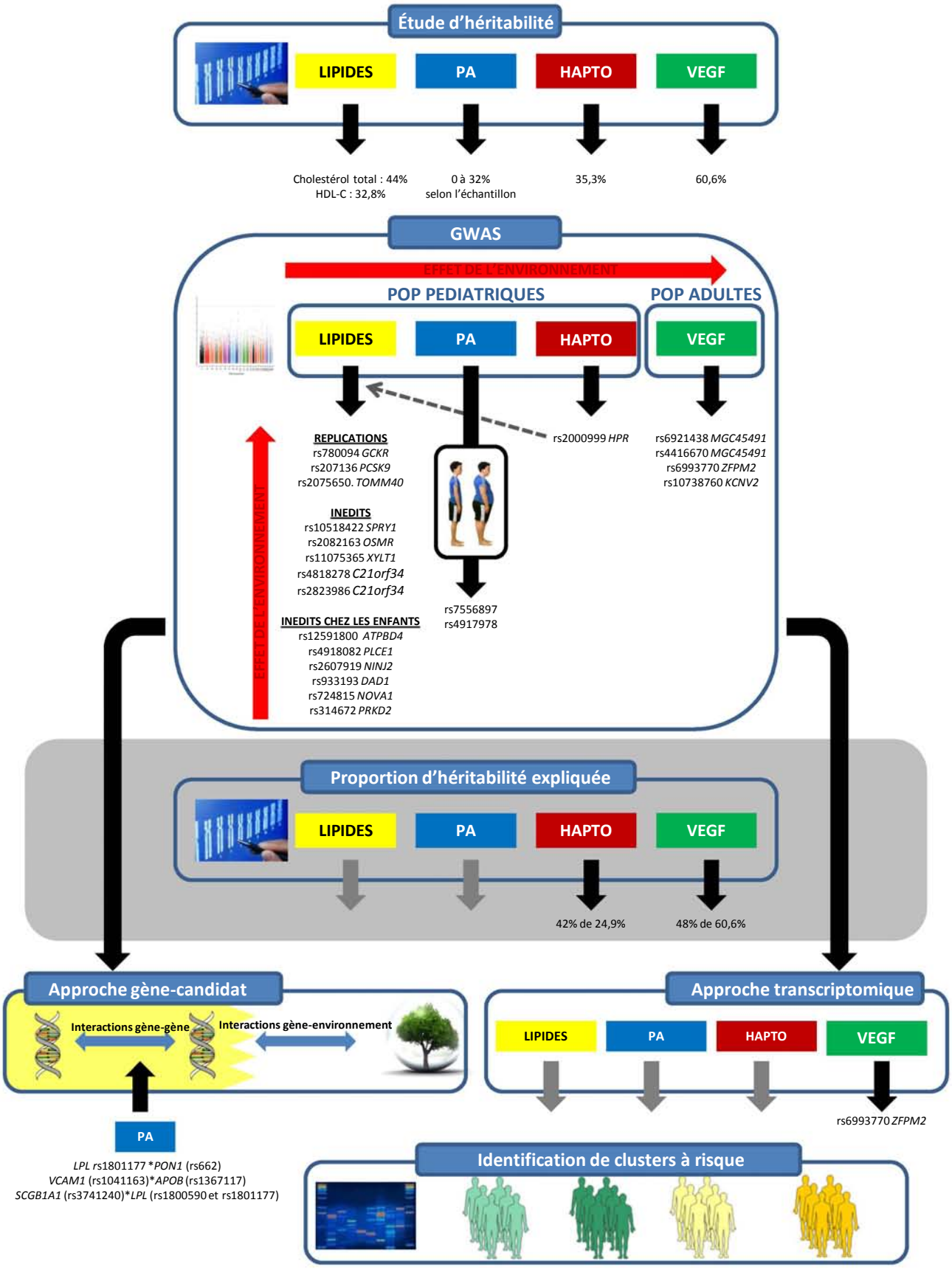
Ainsi, sur le plan pratique, nous avons entrepris d'expérimenter les possibilités d'une telle stratégie en ciblant à la fois des facteurs de risque classiques et largement étudiés des MCV : les lipides et la pression artérielle, mais aussi des facteurs de risque émergents ou discutés : l'haptoglobine et le VEGF.

Ceci nous a permis de 'designer' et proposer une stratégie d'étude globale en trois phases intégrant diverses méthodologies et approches expérimentales schématisée dans la **Figure 17**.

1) Une phase de détermination de la part d'héritabilité génétique du trait concerné.
   Nous avons ainsi quantifié l'héritabilité génétique (ou variance polygénique) du cholestérol total : 44%, du HDL-C : 32,8%, de l'haptoglobine : 35,3% et du VEGF dans une étude antérieure : 60,6%. Nous avons également mis en évidence la nécessité de déterminer cette héritabilité dans diverses populations de façon à identifier les traits sous forte influence environnementale. En effet, nous n'avons observé aucune héritabilité génétique pour la pression artérielle dans l'échantillon de la cohorte STANISLAS que nous avons étudié pour cette étape, résultat contradictoire avec une étude antérieure menée sur un échantillon différent de la même population.

2) Une phase de GWAS dans des populations pédiatriques.
   Lors de ces travaux, nous avons pu prouver que les variants génétiques antérieurement mis en évidence dans des populations adultes, en particulier pour les lipides, avaient un effet observable dès l'enfance et surtout que de nouveaux variants pouvaient être découverts. Certains semblaient d'ailleurs spécifiques à l'enfance et des études confirmatoires sont en cours.
   Dans le cas de la pression artérielle, pour laquelle nous avions constaté une forte variation de l'héritabilité, nous avons testé l'effet de l'obésité et observé une modulation de l'effet de loci connus sur la pression en fonction de la corpulence des sujets.

Figure 17 : Stratégie d'étude intégrative de l'étiologie génétique des MCV



Cette phase pangénomique nous a également permis de mettre en évidence un variant génétique semblant être à la jonction de deux voies métaboliques car ayant une action à la fois sur l'haptoglobine et sur les lipides (cholestérol total et HDL-C).
Notez que nous avons effectué une GWAS du VEGF sur des populations adultes étant donné que son héritabilité génétique de 60,6% impliquerait une faible action des facteurs environnementaux et qu'il était donc moins primordial de les minimiser. Nous avons d'ailleurs confirmé cette forte héritabilité dans la cohorte Framingham.

Une étape pivot de nos travaux a été la détermination d'une portion importante de l'héritabilité génétique expliquée par nos GWAS : elle était de 42% pour l'haptoglobine, et de 48% pour le VEGF.

3) Enfin, nous proposons une phase explicative en deux temps par :
    a. La détermination de l'influence des phénomènes épistasiques et des interactions gène-environnement qui doivent être intégrés lors des analyses futures des variants mis en évidence par les GWAS.
    En effet, en utilisant une approche gène-candidat préalable, nous avons pu démontrer l'existence de phénomènes épistasiques régulant la pression artérielle et qui n'étaient pas observables lors des analyses d'association individuelles.
    b. La recherche d'une fonctionnalité par une approche transcriptomique.
    Nous avons pu démontrer l'intérêt de cette approche en identifiant un variant, parmi les 4 indépendamment associés au VEGF, lié à l'ARNm $VEGF_{121}$ et présentant ainsi une éventuelle cible thérapeutique.

Un tel design en strates serait à notre sens un bon outil d'élucidation de l'étiologie génétique des MCV même si certaines problématiques ne sont pas encore résolues (*PUBLICATION N°1*) et permettrait la détermination de clusters à risque en prévention primo-secondaire.

Une quatrième étape serait l'expérimentation pharmacogénomique et pharmacogénétique de notre cheminement et de nos résultats dans le cadre du projet BioIntelligence. Même si nous convenons qu'il est prématuré d'imaginer avoir des retombées thérapeutiques dans un futur proche, nous sommes convaincus que de telles expérimentations doivent avoir lieu largement en amont.

C'est pour cette raison que nous allons poursuivre la mise en œuvre de ce design intégré par la détermination de la proportion d'héritabilité expliquée par les SNPs identifiés dans le cas des lipides et de la pression artérielle. Nous poursuivrons également notre approche transcriptomique sur ces deux traits mais aussi sur l'haptoglobine de façon à compléter le cheminement analytique que nous avons conçu et mis en place au cours de notre thèse.

Bien évidemment, des adaptations sont à prévoir en fonction des avancées méthodologiques et logistiques constantes, qui sont la caractéristique de l'ère des GWAS, et que nous continuerons à acquérir.

RÉFÉRENCES BIBLIOGRAPHIQUES

[1] 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 May 16;285(19):2486-97.

[2] Aijaz B, Brown TM, Sanderson BK, Bittner V. Exercise as a treatment for the risk of cardiovascular disease. Curr Treat Options Cardiovasc Med. 2010 Aug;12(4):329-41.

[3] Albert MA, Danielson E, Rifai N, Ridker PM. Effect of statin therapy on C-reactive protein levels: the pravastatin inflammation/CRP evaluation (PRINCE): a randomized trial and cohort study. Jama. 2001 Jul 4;286(1):64-70.

[4] Alberti KG, Zimmet P, Shaw J. The metabolic syndrome--a new worldwide definition. Lancet. 2005 Sep 24-30;366(9491):1059-62.

[5] Anderson KM. A nonproportional hazards Weibull accelerated failure time regression model. Biometrics. 1991 Mar;47(1):281-8.

[6] Anderson KM, Odell PM, Wilson PW, Kannel WB. Cardiovascular disease risk profiles. Am Heart J. 1991 Jan;121(1 Pt 2):293-8.

[7] Arsenault BJ, Lemieux I, Despres JP, Wareham NJ, Kastelein JJ, Khaw KT, et al. The hypertriglyceridemic-waist phenotype and the risk of coronary artery disease: results from the EPIC-Norfolk prospective population study. Cmaj. 2010 Sep 21;182(13):1427-32.

[8] Asleh R, Blum S, Kalet-Litman S, Alshiek J, Miller-Lotan R, Asaf R, et al. Correction of HDL dysfunction in individuals with diabetes and the haptoglobin 2-2 genotype. Diabetes. 2008 Oct;57(10):2794-800.

[9] Aulchenko YS, Ripatti S, Lindqvist I, Boomsma D, Heid IM, Pramstaller PP, et al. Loci influencing lipid levels and coronary heart disease risk in 16 European population cohorts. Nat Genet. 2009 Jan;41(1):47-55.

[10] Bak S, Gaist D, Sindrup SH, Skytthe A, Christensen K. Genetic liability in stroke: a long-term follow-up study of Danish twins. Stroke. 2002 Mar;33(3):769-74.

[11] Balkau B, Eschwege E, Tichet J, Marre M. Proposed criteria for the diagnosis of diabetes: evidence from a French epidemiological study (D.E.S.I.R.). Diabetes Metab. 1997 Nov;23(5):428-34.

[12] Barengo NC, Katoh S, Moltchanov V, Tajima N, Tuomilehto J. The diabetes-cardiovascular risk paradox: results from a Finnish population-based prospective study. Eur Heart J. 2008 Aug;29(15):1889-95.

[13] Bathum L, Petersen I, Christiansen L, Konieczna A, Sorensen TI, Kyvik KO. Genetic and environmental influences on plasma homocysteine: results from a Danish twin study. Clin Chem. 2007 May;53(5):971-9.

[14] Bayoumi RA, Al-Yahyaee SA, Albarwani SA, Rizvi SG, Al-Hadabi S, Al-Ubaidi FF, et al. Heritability of determinants of the metabolic syndrome among healthy Arabs of the Oman family study. Obesity (Silver Spring). 2007 Mar;15(3):551-6.

[15] Bener A, Dogan M, Ehlayel MS, Shanks NJ, Sabbah A. The impact of air pollution on hospital admission for respiratory and cardiovascular diseases in an oil and gas-rich country. Eur Ann Allergy Clin Immunol. 2009 Jun;41(3):80-4.

[16] Benfante R. Studies of cardiovascular disease and cause-specific mortality trends in Japanese-American men living in Hawaii and risk factor comparisons with other Japanese populations in the Pacific region: a review. Hum Biol. 1992 Dec;64(6):791-805.

[17] Berg K, Malinow MR, Kierulf P, Upson B. Population variation and genetics of plasma homocyst(e)ine level. Clin Genet. 1992 Jun;41(6):315-21.

[18] Berlin JA, Colditz GA. A meta-analysis of physical activity in the prevention of coronary heart disease. Am J Epidemiol. 1990 Oct;132(4):612-28.

[19] Berrahmoune H, Herbeth B, Lamont JV, Fitzgerald PS, Visvikis-Siest S. Association between TNF and IL-1 bloc polymorphisms and plasma MCP-1 concentration. Atherosclerosis. 2007 Jun;192(2):348-53.

[20] Berrahmoune H, Herbeth B, Lamont JV, Masson C, Fitzgerald PS, Visvikis-Siest S. Heritability for plasma VEGF concentration in the Stanislas family study. Ann Hum Genet. 2007 Jan;71(Pt 1):54-63.

[21] Berrahmoune H, Herbeth B, Siest G, Visvikis-Siest S. Heritability of serum hs-CRP concentration and 5-year changes in the Stanislas family study: association with apolipoprotein E alleles. Genes Immun. 2007 Jun;8(4):352-9.

[22] Berrahmoune H, Lamont JV, Herbeth B, FitzGerald PS, Visvikis-Siest S. Biological determinants of and reference values for plasma interleukin-8, monocyte chemoattractant protein-1, epidermal growth factor, and vascular endothelial growth factor: Results from the STANISLAS cohort. Clin Chem. 2006 Mar;52(3):504-10.

[23] Bertrand-Thiebault C, Berrahmoune H, Thompson A, Marie B, Droesch S, Siest G, et al. Genetic Polymorphism of CYP2C19 gene in the Stanislas cohort. A link with inflammation. Ann Hum Genet. 2008 Mar;72(Pt 2):178-83.

[24] Blann AD, Belgore FM, McCollum CN, Silverman S, Lip PL, Lip GY. Vascular endothelial growth factor and its receptor, Flt-1, in the plasma of patients with coronary or peripheral atherosclerosis, or Type II diabetes. Clin Sci (Lond). 2002 Feb;102(2):187-94.

[25] Bouatia-Naji N, Bonnefond A, Cavalcanti-Proenca C, Sparso T, Holmkvist J, Marchand M, et al. A variant near MTNR1B is associated with increased fasting plasma glucose levels and type 2 diabetes risk. Nat Genet. 2009 Jan;41(1):89-94.

[26] Brown WM, Beck SR, Lange EM, Davis CC, Kay CM, Langefeld CD, et al. Age-stratified heritability estimation in the Framingham Heart Study families. BMC Genet. 2003;4 Suppl 1:S32.

[27] Buysschaert I, Carmeliet P, Dewerchin M. Clinical and fundamental aspects of angiogenesis and anti-angiogenesis. Acta Clin Belg. 2007 May-Jun;62(3):162-9.

[28] Carmeliet P, Jain RK. Angiogenesis in cancer and other diseases. Nature. 2000 Sep 14;407(6801):249-57.

[29] Carter K, Worwood M. Haptoglobin: a review of the major allele frequencies worldwide and their association with diseases. Int J Lab Hematol. 2007 Apr;29(2):92-110.

[30] Casas JP, Cooper J, Miller GJ, Hingorani AD, Humphries SE. Investigating the genetic determinants of cardiovascular disease using candidate genes and meta-analysis of association studies. Ann Hum Genet. 2006 Mar;70(Pt 2):145-69.

[31] Cassidy-Bushrow AE, Bielak LF, Sheedy PF, 2nd, Turner ST, Kullo IJ, Lin X, et al. Coronary artery calcification progression is heritable. Circulation. 2007 Jul 3;116(1):25-31.

[32] Cesari M, Burlina AB, Narkiewicz K, Sartori MT, Sacchetto A, Rossi GP. Are fasting plasma homocyst(e)ine levels heritable? A study of normotensive twins. J Investig Med. 2000 Sep;48(5):351-8.

[33] Chanock SJ, Manolio T, Boehnke M, Boerwinkle E, Hunter DJ, Thomas G, et al. Replicating genotype-phenotype associations. Nature. 2007 Jun 7;447(7145):655-60.

[34] Cheng S, Pallaud C, Grow MA, Scharf SJ, Erlich HA, Klitz W, et al. A multilocus genotyping assay for cardiovascular disease. Clin Chem Lab Med. 1998 Aug;36(8):561-6.

[35] Chin BS, Chung NA, Gibbs CR, Blann AD, Lip GY. Vascular endothelial growth factor and soluble P-selectin in acute and chronic congestive heart failure. Am J Cardiol. 2002 Dec 1;90(11):1258-60.

[36] Dawber TR, Kannel WB. The Framingham study. An epidemiological approach to coronary heart disease. Circulation. 1966 Oct;34(4):553-5.

[37] de Andrade M, Gueguen R, Visvikis S, Sass C, Siest G, Amos CI. Extension of variance components approach to incorporate temporal trends and longitudinal pedigree data analysis. Genet Epidemiol. 2002 Mar;22(3):221-32.

[38] Ducros V, Barro C, Yver J, Pernod G, Polack B, Carpentier P, et al. Should plasma homocysteine be used as a biomarker of venous thromboembolism? A case-control study. Clin Appl Thromb Hemost. 2009 Oct;15(5):517-22.

[39] Duffaut C, Zakaroff-Girard A, Bourlier V, Decaunes P, Maumus M, Chiotasso P, et al. Interplay between human adipocytes and T lymphocytes in obesity: CCL20 as an adipochemokine and T lymphocytes as lipogenic modulators. Arterioscler Thromb Vasc Biol. 2009 Oct;29(10):1608-14.

[40] Dupuis J, Langenberg C, Prokopenko I, Saxena R, Soranzo N, Jackson AU, et al. New genetic loci implicated in fasting glucose homeostasis and their impact on type 2 diabetes risk. Nat Genet. 2010 Feb;42(2):105-16.

[41] Eckel RH, Alberti KG, Grundy SM, Zimmet PZ. The metabolic syndrome. Lancet. 2010 Jan 16;375(9710):181-3.

[42] Eckel RH, Grundy SM, Zimmet PZ. The metabolic syndrome. Lancet. 2005 Apr 16-22;365(9468):1415-28.

[43] Feinleib M, Kannel WB, Garrison RJ, McNamara PM, Castelli WP. The Framingham Offspring Study. Design and preliminary data. Prev Med. 1975 Dec;4(4):518-25.

[44] Ferrara N, Gerber HP, LeCouter J. The biology of VEGF and its receptors. Nat Med. 2003 Jun;9(6):669-76.

[45] Fischer M, Broeckel U, Holmer S, Baessler A, Hengstenberg C, Mayer B, et al. Distinct heritable patterns of angiographic coronary artery disease in families with myocardial infarction. Circulation. 2005 Feb 22;111(7):855-62.

[46] Fitzgerald SP, Lamont JV, McConnell RI, Benchikh el O. Development of a high-throughput automated analyzer using biochip array technology. Clin Chem. 2005 Jul;51(7):1165-76.

[47] Freeman MS, Mansfield MW, Barrett JH, Grant PJ. Genetic contribution to circulating levels of hemostatic factors in healthy families with effects of known genetic polymorphisms on heritability. Arterioscler Thromb Vasc Biol. 2002 Mar 1;22(3):506-10.

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

[49] Gabay C, Kushner I. Acute-phase proteins and other systemic responses to inflammation. N Engl J Med. 1999 Feb 11;340(6):448-54.

[50] Gambier N, Marteau JB, Batt AM, Marie B, Thompson A, Siest G, et al. Interaction between CYP1A1 T3801C and AHR G1661A polymorphisms according to smoking status on blood pressure in the Stanislas cohort. J Hypertens. 2006 Nov;24(11):2199-205.

[51] Gooding HC, de Ferranti SD. Cardiovascular risk assessment and cholesterol management in adolescents: getting to the heart of the matter. Curr Opin Pediatr. 2010 Aug;22(4):398-404.

[52] Gordon DJ, Probstfield JL, Garrison RJ, Neaton JD, Castelli WP, Knoke JD, et al. High-density lipoprotein cholesterol and cardiovascular disease. Four prospective American studies. Circulation. 1989 Jan;79(1):8-15.

[53] Grundy SM, Pasternak R, Greenland P, Smith S, Jr., Fuster V. Assessment of cardiovascular risk by use of multiple-risk-factor assessment equations: a statement for healthcare professionals from the American Heart Association and the American College of Cardiology. Circulation. 1999 Sep 28;100(13):1481-92.

[54] Grundy SM, Pasternak R, Greenland P, Smith S, Jr., Fuster V. AHA/ACC scientific statement: Assessment of cardiovascular risk by use of multiple-risk-factor assessment equations: a statement for healthcare professionals from the American Heart Association and the American College of Cardiology. J Am Coll Cardiol. 1999 Oct;34(4):1348-59.

[55] Haddy N, Sass C, Droesch S, Zaiou M, Siest G, Ponthieux A, et al. IL-6, TNF-alpha and atherosclerosis risk indicators in a healthy family population: the STANISLAS cohort. Atherosclerosis. 2003 Oct;170(2):277-83.

[56] Haddy N, Sass C, Maumus S, Marie B, Droesch S, Siest G, et al. Biological variations, genetic polymorphisms and familial resemblance of TNF-alpha and IL-6 concentrations: STANISLAS cohort. Eur J Hum Genet. 2005 Jan;13(1):109-17.

[57] Hanson RL, Imperatore G, Narayan KM, Roumain J, Fagot-Campagna A, Pettitt DJ, et al. Family and genetic studies of indices of insulin sensitivity and insulin secretion in Pima Indians. Diabetes Metab Res Rev. 2001 Jul-Aug;17(4):296-303.

[58] Hardy J, Singleton A. Genomewide association studies and human disease. N Engl J Med. 2009 Apr 23;360(17):1759-68.

[59] Heller DA, de Faire U, Pedersen NL, Dahlen G, McClearn GE. Genetic and environmental influences on serum lipid levels in twins. N Engl J Med. 1993 Apr 22;328(16):1150-6.

[60] Herbeth B, Gueguen S, Leroy P, Siest G, Visvikis-Siest S. The lipoprotein lipase serine 447 stop polymorphism is associated with altered serum carotenoid concentrations in the Stanislas Family Study. J Am Coll Nutr. 2007 Dec;26(6):655-62.

[61] Herbeth B, Samara A, Ndiaye C, Marteau JB, Berrahmoune H, Siest G, et al. Metabolic syndrome-related composite factors over 5 years in the STANISLAS Family Study: genetic heritability and common environmental influences. Clin Chim Acta. 2010 Jun 3;411(11-12):833-9.

[62] Hindorff LA, Sethupathy P, Junkins HA, Ramos EM, Mehta JP, Collins FS, et al. Potential etiologic and functional implications of genome-wide association loci for human diseases and traits. Proc Natl Acad Sci U S A. 2009 Jun 9;106(23):9362-7.

[63] Hojo Y, Ikeda U, Zhu Y, Okada M, Ueno S, Arakawa H, et al. Expression of vascular endothelial growth factor in patients with acute myocardial infarction. J Am Coll Cardiol. 2000 Mar 15;35(4):968-73.

[64] Holme I, Aastveit AH, Hammar N, Jungner I, Walldius G. Haptoglobin and risk of myocardial infarction, stroke, and congestive heart failure in 342,125 men and women in the Apolipoprotein MOrtality RISk study (AMORIS). Ann Med. 2009;41(7):522-32.

[65] Holme I, Aastveit AH, Hammar N, Jungner I, Walldius G. Inflammatory markers, lipoprotein components and risk of major cardiovascular events in 65,005 men and women in the Apolipoprotein MOrtality RISk study (AMORIS). Atherosclerosis. 2010 Nov;213(1):299-305.

[66] Hoover RN. The evolution of epidemiologic research: from cottage industry to "big" science. Epidemiology. 2007 Jan;18(1):13-7.

[67] Igl W, Johansson A, Wilson JF, Wild SH, Polasek O, Hayward C, et al. Modeling of environmental effects in genome-wide association studies identifies SLC2A2 and HP as novel loci influencing serum cholesterol levels. PLoS Genet. 2010 Jan;6(1):e1000798.

[68] Jarvelin MR, Elliott P, Kleinschmidt I, Martuzzi M, Grundy C, Hartikainen AL, et al. Ecological and individual predictors of birthweight in a northern Finland birth cohort 1986. Paediatr Perinat Epidemiol. 1997 Jul;11(3):298-312.

[69] Jouven X, Desnos M, Guerot C, Ducimetiere P. Predicting sudden death in the population: the Paris Prospective Study I. Circulation. 1999 Apr 20;99(15):1978-83.

[70] Kathiresan S, Willer CJ, Peloso GM, Demissie S, Musunuru K, Schadt EE, et al. Common variants at 30 loci contribute to polygenic dyslipidemia. Nat Genet. 2009 Jan;41(1):56-65.

[71] Keys A, Menotti A, Aravanis C, Blackburn H, Djordevic BS, Buzina R, et al. The seven countries study: 2,289 deaths in 15 years. Prev Med. 1984 Mar;13(2):141-54.

[72] Koda Y, Soejima M, Yoshioka N, Kimura H. The haptoglobin-gene deletion responsible for anhaptoglobinemia. Am J Hum Genet. 1998 Feb;62(2):245-52.

[73] Kromhout D. Fatty acids, antioxidants, and coronary heart disease from an epidemiological perspective. Lipids. 1999;34 Suppl:S27-31.

[74] Kromhout D, Keys A, Aravanis C, Buzina R, Fidanza F, Giampaoli S, et al. Food consumption patterns in the 1960s in seven countries. Am J Clin Nutr. 1989 May;49(5):889-94.

[75] Kurt R, Yilmaz Y, Ermis F, Kalayoglu Besisik S, Polat N, Elitok A, et al. Folic Acid and vitamin B12 supplementation improves coronary flow reserve in elderly subjects with vitamin B12 deficiency. Arch Med Res. 2010 Jul;41(5):369-72.

[76] Lambert CA, Mineur P, Nusgens BV. [VEGF111: Dr Jekyll and Mr Hyde?]. Med Sci (Paris). 2008 Jun-Jul;24(6-7):579-80.

[77] Leininger-Muller B, Hoy A, Herbeth B, Pfister M, Serot JM, Stavljenic-Rukavina M, et al. Myeloperoxidase G-463A polymorphism and Alzheimer's disease in the ApoEurope study. Neurosci Lett. 2003 Oct 2;349(2):95-8.

[78] Levy D, Ehret GB, Rice K, Verwoert GC, Launer LJ, Dehghan A, et al. Genome-wide association study of blood pressure and hypertension. Nat Genet. 2009 Jun;41(6):677-87.

[79] Levy D, Larson MG, Benjamin EJ, Newton-Cheh C, Wang TJ, Hwang SJ, et al. Framingham Heart Study 100K Project: genome-wide associations for blood pressure and arterial stiffness. BMC Med Genet. 2007;8 Suppl 1:S3.

[80] Liao JK. Isoprenoids as mediators of the biological effects of statins. J Clin Invest. 2002 Aug;110(3):285-8.

[81] Libby P, Ridker PM, Maseri A. Inflammation and atherosclerosis. Circulation. 2002 Mar 5;105(9):1135-43.

[82] Lind L, Fors N, Hall J, Marttala K, Stenborg A. A comparison of three different methods to evaluate endothelium-dependent vasodilation in the elderly: the Prospective Investigation of the Vasculature in Uppsala Seniors (PIVUS) study. Arterioscler Thromb Vasc Biol. 2005 Nov;25(11):2368-75.

[83] Liu J, Fox CS, Hickson DA, May WD, Hairston KG, Carr JJ, et al. Impact of Abdominal Visceral and Subcutaneous Adipose Tissue on Cardiometabolic Risk Factors: The Jackson Heart Study. J Clin Endocrinol Metab. 2010 Sep 15.

[84] Maher B. Personal genomes: The case of the missing heritability. Nature. 2008 Nov 6;456(7218):18-21.

[85] Manolio TA. Genomewide association studies and assessment of the risk of disease. N Engl J Med. 2010 Jul 8;363(2):166-76.

[86] Manolio TA, Brooks LD, Collins FS. A HapMap harvest of insights into the genetics of common disease. J Clin Invest. 2008 May;118(5):1590-605.

[87] Manolio TA, Collins FS, Cox NJ, Goldstein DB, Hindorff LA, Hunter DJ, et al. Finding the missing heritability of complex diseases. Nature. 2009 Oct 8;461(7265):747-53.

[88] Marteau JB, Lambert D, Herbeth B, Marie B, Droesch S, Tregouet DA, et al. P-selectin polymorphisms' influences on P-selectin serum concentrations and on their familial correlation: the STANISLAS family study. J Thromb Haemost. 2008 Jun;6(6):920-7.

[89] Marteau JB, Mohr S, Pfister M, Visvikis-Siest S. Collection and storage of human blood cells for mRNA expression profiling: a 15-month stability study. Clin Chem. 2005 Jul;51(7):1250-2.

[90] Marteau JB, Sass C, Pfister M, Lambert D, Noyer-Weidner M, Visvikis S. The Leu554Phe polymorphism in the E-selectin gene is associated with blood pressure in overweight people. J Hypertens. 2004 Feb;22(2):305-11.

[91] Maumus S, Marie B, Vincent-Viry M, Siest G, Visvikis-Siest S. Analysis of the effect of multiple genetic variants of cardiovascular disease risk on insulin concentration variability in healthy adults of the STANISLAS cohort. The role of FGB-455 G/A polymorphism. Atherosclerosis. 2007 Apr;191(2):369-76.

[92] McQueen MB, Bertram L, Rimm EB, Blacker D, Santangelo SL. A QTL genome scan of the metabolic syndrome and its component traits. BMC Genet. 2003;4 Suppl 1:S96.

[93] Meyre D, Delplanque J, Chevre JC, Lecoeur C, Lobbens S, Gallina S, et al. Genome-wide association study for early-onset and morbid adult obesity identifies three new risk loci in European populations. Nat Genet. 2009 Feb;41(2):157-9.

[94] Meyre D, Proulx K, Kawagoe-Takaki H, Vatin V, Gutierrez-Aguilar R, Lyon D, et al. Prevalence of loss-of-function FTO mutations in lean and obese individuals. Diabetes. 2010 Jan;59(1):311-8.

[95] Miettinen H, Lehto S, Salomaa V, Mahonen M, Niemela M, Haffner SM, et al. Impact of diabetes on mortality after the first myocardial infarction. The FINMONICA Myocardial Infarction Register Study Group. Diabetes Care. 1998 Jan;21(1):69-75.

[96] Miller SA, Dykes DD, Polesky HF. A simple salting out procedure for extracting DNA from human nucleated cells. Nucleic Acids Res. 1988 Feb 11;16(3):1215.

[97] Mills GW, Avery PJ, McCarthy MI, Hattersley AT, Levy JC, Hitman GA, et al. Heritability estimates for beta cell function and features of the insulin resistance syndrome in UK families with an increased susceptibility to type 2 diabetes. Diabetologia. 2004 Apr;47(4):732-8.

[98] Muntner P, He J, Cutler JA, Wildman RP, Whelton PK. Trends in blood pressure among children and adolescents. Jama. 2004 May 5;291(17):2107-13.

[99] Newton-Cheh C, Johnson T, Gateva V, Tobin MD, Bochud M, Coin L, et al. Genome-wide association study identifies eight loci associated with blood pressure. Nat Genet. 2009 Jun;41(6):666-76.

[100] Nilsson SE, Read S, Berg S, Johansson B. Heritabilities for fifteen routine biochemical values: findings in 215 Swedish twin pairs 82 years of age or older. Scand J Clin Lab Invest. 2009;69(5):562-9.

[101] O'Connell DL, Heller RF, Roberts DC, Allen JR, Knapp JC, Steele PL, et al. Twin study of genetic and environmental effects on lipid levels. Genet Epidemiol. 1988;5(5):323-41.

[102] Okazaki T, Yanagisawa Y, Nagai T. Analysis of the affinity of each haptoglobin polymer for hemoglobin by two-dimensional affinity electrophoresis. Clin Chim Acta. 1997 Feb 17;258(2):137-44.

[103] Ouwehand WH. The discovery of genes implicated in myocardial infarction. J Thromb Haemost. 2009 Jul;7 Suppl 1:305-7.

[104] Pallaud C, Gueguen R, Sass C, Grow M, Cheng S, Siest G, et al. Genetic influences on lipid metabolism trait variability within the Stanislas Cohort. J Lipid Res. 2001 Nov;42(11):1879-90.

[105] Pallaud C, Sass C, Zannad F, Siest G, Visvikis S. APOC3, CETP, fibrinogen, and MTHFR are genetic determinants of carotid intima-media thickness in healthy men (the Stanislas cohort). Clin Genet. 2001 May;59(5):316-24.

[106] Papandreou D, Rousso I, Malindretos P, Makedou A, Arvanitidou M. Effects of oral folate supplementation on serum total homocysteine and cholesterol levels in hyperhomocysteinemic children. Nutr Clin Pract. 2010 Aug;25(4):390-3.

[107] Papoutsakis C, Vidra NV, Hatzopoulou I, Tzirkalli M, Farmaki AE, Evagelidaki E, et al. The Gene-Diet Attica investigation on childhood obesity (GENDAI): overview of the study design. Clin Chem Lab Med. 2007;45(3):309-15.

[108] Pearson TA, Mensah GA, Hong Y, Smith SC, Jr. CDC/AHA Workshop on Markers of Inflammation and Cardiovascular Disease: Application to Clinical and Public Health Practice: overview. Circulation. 2004 Dec 21;110(25):e543-4.

[109] Pipe AL, Papadakis S, Reid RD. The role of smoking cessation in the prevention of coronary artery disease. Curr Atheroscler Rep. 2010 Mar;12(2):145-50.

[110] Ponthieux A, Herbeth B, Droesch S, Haddy N, Lambert D, Visvikis S. Biological determinants of serum ICAM-1, E-selectin, P-selectin and L-selectin levels in healthy subjects: the Stanislas study. Atherosclerosis. 2004 Feb;172(2):299-308.

[111] Ponthieux A, Herbeth B, Droesch S, Lambert D, Visvikis S. Age- and sex-related reference values for serum adhesion molecule concentrations in healthy individuals: intercellular adhesion molecule-1 and E-, P-, and L-selectin. Clin Chem. 2003 Sep;49(9):1544-6.

[112] Pospisilik JA, Schramek D, Schnidar H, Cronin SJ, Nehme NT, Zhang X, et al. Drosophila genome-wide obesity screen reveals hedgehog as a determinant of brown versus white adipose cell fate. Cell. 2010 Jan 8;140(1):148-60.

[113] Poulsen P, Kyvik KO, Vaag A, Beck-Nielsen H. Heritability of type II (non-insulin-dependent) diabetes mellitus and abnormal glucose tolerance--a population-based twin study. Diabetologia. 1999 Feb;42(2):139-45.

[114] Qu HQ, Bradfield JP, Li Q, Kim C, Frackelton E, Grant SF, et al. In silico replication of the genome-wide association results of the Type 1 Diabetes Genetics Consortium. Hum Mol Genet. 2010 Jun 15;19(12):2534-8.

[115] Reed T, Malinow MR, Christian JC, Upson B. Estimates of heritability of plasma homocyst(e)ine levels in aging adult male twins. Clin Genet. 1991 Jun;39(6):425-8.

[116] Rimm EB, Chan J, Stampfer MJ, Colditz GA, Willett WC. Prospective study of cigarette smoking, alcohol use, and the risk of diabetes in men. Bmj. 1995 Mar 4;310(6979):555-9.

[117] Rolland-Cachera MF, Cole TJ, Sempe M, Tichet J, Rossignol C, Charraud A. Body Mass Index variations: centiles from birth to 87 years. Eur J Clin Nutr. 1991 Jan;45(1):13-21.

[118] Ryndel M, Behre CJ, Brohall G, Prahl U, Schmidt C, Bergstrom G, et al. The haptoglobin 2-2 genotype is associated with carotid atherosclerosis in 64-year old women with established diabetes. Clin Chim Acta. 2010 Apr 2;411(7-8):500-4.

[119] Sass C, Blanquart C, Morange PE, Pfister M, Visvikis-Siest S. Association between factor VII polymorphisms and blood pressure: the Stanislas Cohort. Hypertension. 2004 Nov;44(5):674-80.

[120] Sass C, Herbeth B, Siest G, Visvikis S. Lipoprotein lipase (C/G)447 polymorphism and blood pressure in the Stanislas Cohort. J Hypertens. 2000 Dec;18(12):1775-81.

[121] Scherag A, Dina C, Hinney A, Vatin V, Scherag S, Vogel CI, et al. Two new Loci for body-weight regulation identified in a joint analysis of genome-wide association studies for early-onset extreme obesity in French and german study groups. PLoS Genet. 2010;6(4):e1000916.

[122] Schiele F, De Bacquer D, Vincent-Viry M, Beisiegel U, Ehnholm C, Evans A, et al. Apolipoprotein E serum concentration and polymorphism in six European countries: the ApoEurope Project. Atherosclerosis. 2000 Oct;152(2):475-88.

[123] Schildkraut JM, Myers RH, Cupples LA, Kiely DK, Kannel WB. Coronary risk associated with age and sex of parental heart disease in the Framingham Study. Am J Cardiol. 1989 Sep 15;64(10):555-9.

[124] Segal KR, Van Loan M, Fitzgerald PI, Hodgdon JA, Van Itallie TB. Lean body mass estimation by bioelectrical impedance analysis: a four-site cross-validation study. Am J Clin Nutr. 1988 Jan;47(1):7-14.

[125] Siest G, Bertrand P, Herbeth B, Vincent-Viry M, Schiele F, Sass C, et al. Apolipoprotein E polymorphisms and concentration in chronic diseases and drug responses. Clin Chem Lab Med. 2000 Sep;38(9):841-52.

[126] Siest G, Visvikis S, Herbeth B, Gueguen R, Vincent-Viry M, Sass C, et al. Objectives, design and recruitment of a familial and longitudinal cohort for studying gene-environment interactions in the field of cardiovascular risk: the Stanislas cohort. Clin Chem Lab Med. 1998 Jan;36(1):35-42.

[127] Siva A, De Lange M, Clayton D, Monteith S, Spector T, Brown MJ. The heritability of plasma homocysteine, and the influence of genetic variation in the homocysteine methylation pathway. Qjm. 2007 Aug;100(8):495-9.

[128] Slevin M, Krupinski J, Slowik A, Kumar P, Szczudlik A, Gaffney J. Serial measurement of vascular endothelial growth factor and transforming growth factor-beta1 in serum of patients with acute ischemic stroke. Stroke. 2000 Aug;31(8):1863-70.

[129] Souren NY, Paulussen AD, Loos RJ, Gielen M, Beunen G, Fagard R, et al. Anthropometry, carbohydrate and lipid metabolism in the East Flanders Prospective Twin Survey: heritabilities. Diabetologia. 2007 Oct;50(10):2107-16.

[130] Speliotes EK, Willer CJ, Berndt SI, Monda KL, Thorleifsson G, Jackson AU, et al. Association analyses of 249,796 individuals reveal 18 new loci associated with body mass index. Nat Genet. 2010 Nov;42(11):937-48.

[131] Splansky GL, Corey D, Yang Q, Atwood LD, Cupples LA, Benjamin EJ, et al. The Third Generation Cohort of the National Heart, Lung, and Blood Institute's Framingham Heart Study: design, recruitment, and initial examination. Am J Epidemiol. 2007 Jun 1;165(11):1328-35.

[132] Stamler J, Neaton JD. The Multiple Risk Factor Intervention Trial (MRFIT)--importance then and now. Jama. 2008 Sep 17;300(11):1343-5.

[133] Starck M, Schiele F, Herbeth B, Vincent-Viry M, Beaud B, Siest G, et al. Apolipoproteins E and C-III in apo B- and non-apo B-containing lipoproteins in middle-aged women from the Stanislas cohort: effect of oral contraceptive use and common apolipoprotein E polymorphism. Atherosclerosis. 2001 Apr;155(2):509-16.

[134] Takeuchi F, Isono M, Katsuya T, Yamamoto K, Yokota M, Sugiyama T, et al. Blood pressure and hypertension are associated with 7 loci in the Japanese population. Circulation. 2010 Jun 1;121(21):2302-9.

[135] Takeya Y, Popper JS, Shimizu Y, Kato H, Rhoads GG, Kagan A. Epidemiologic studies of coronary heart disease and stroke in Japanese men living in Japan, Hawaii and California: incidence of stroke in Japan and Hawaii. Stroke. 1984 Jan-Feb;15(1):15-23.

[136] Teslovich TM, Musunuru K, Smith AV, Edmondson AC, Stylianou IM, Koseki M, et al. Biological, clinical and population relevance of 95 loci for blood lipids. Nature. 2010 Aug 5;466(7307):707-13.

[137] Tiret L, Rigat B, Visvikis S, Breda C, Corvol P, Cambien F, et al. Evidence, from combined segregation and linkage analysis, that a variant of the angiotensin I-converting enzyme (ACE) gene controls plasma ACE levels. Am J Hum Genet. 1992 Jul;51(1):197-205.

[138] Topol EJ. Intensive statin therapy--a sea change in cardiovascular prevention. N Engl J Med. 2004 Apr 8;350(15):1562-4.

[139] Tregouet DA, Herbeth B, Juhan-Vague I, Siest G, Ducimetiere P, Tiret L. Bivariate familial correlation analysis of quantitative traits by use of estimating equations: application to a familial analysis of the insulin resistance syndrome. Genet Epidemiol. 1999;16(1):69-83.

[140] Vincent-Viry M, Sass C, Bastien S, Aguillon D, Siest G, Visvikis S. PON1-192 phenotype and genotype assessments in 918 subjects of the Stanislas cohort study. Clin Chem Lab Med. 2003 Apr;41(4):535-40.

[141] Visvikis-Siest S, Siest G. The STANISLAS Cohort: a 10-year follow-up of supposed healthy families. Gene-environment interactions, reference values and evaluation of biomarkers in prevention of cardiovascular diseases. Clin Chem Lab Med. 2008;46(6):733-47.

[142] Vossen CY, Callas PW, Hasstedt SJ, Long GL, Rosendaal FR, Bovill EG. A genetic basis for the interrelation of coagulation factors. J Thromb Haemost. 2007 Sep;5(9):1930-5.

[143] Wang WY, Barratt BJ, Clayton DG, Todd JA. Genome-wide association studies: theoretical and practical concerns. Nat Rev Genet. 2005 Feb;6(2):109-18.

[144] Wienke A, Herskind AM, Christensen K, Skytthe A, Yashin AI. The heritability of CHD mortality in danish twins after controlling for smoking and BMI. Twin Res Hum Genet. 2005 Feb;8(1):53-9.

[145] 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. Nat Genet. 2008 Feb;40(2):161-9.

[146] Wirostko B, Wong TY, Simo R. Vascular endothelial growth factor and diabetic complications. Prog Retin Eye Res. 2008 Nov;27(6):608-21.

[147] 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 Sep;252(3):247-54.

[148] Zeggini E, Scott LJ, Saxena R, Voight BF, Marchini JL, Hu T, et al. Meta-analysis of genome-wide association data and large-scale replication identifies additional susceptibility loci for type 2 diabetes. Nat Genet. 2008 May;40(5):638-45.

[149] Zeggini E, Weedon MN, Lindgren CM, Frayling TM, Elliott KS, Lango H, et al. Replication of genome-wide association signals in UK samples reveals risk loci for type 2 diabetes. Science. 2007 Jun 1;316(5829):1336-41.

PUBLICATIONS

PUBLICATION N°1

Ndeye Coumba Ndiaye*, Mohsen Azimi Nehzad*, Said El Shamieh*, Sophie Visvikis-Siest*. Cardiovascular Diseases and Related Quantitative Traits in the Era of Genome-Wide Association Studies: Where have we Been and What should we Expect? 2010 [soumis]

Du fait de l'hétérogénéité des phénotypes cardiovasculaires cliniques et des populations investiguées, les études pangénomique ou Genome-Wide Association Studies des pathologies cardiovasculaires et des facteurs de risque associés ne sont que partiellement répliquées, et ce malgré le grands nombre de variants génétiques identifiées ces dernières années.

De nouveaux mécanismes physiopathologiques ont certes été mis en évidence mais l'identification du variant causal reste problématique.

Dans les prochaines années, les approches pangénomiques devraient donc intégrer les interactions gène-gène-environnement au sein de stratégies globales impliquant des approches complémentaires telles que la transcriptomique et la protéomique.

L'ère des GWAS est plein de promesses mais nous n'en avons vu que les préludes.

Contribution :

Il s'agit d'une revue de la littérature pour laquelle nous avons effectué des recherches bibliographiques portant notamment sur les méthodologies en vigueur dans le cadre des études pangénomiques ainsi que sur les limitations rencontrées par la communauté scientifique.

Nous faisons également partie de l'équipe rédactionnelle et avons révisé et complété la version finale de l'article avant transmission au Dr Visvikis-Siest.

Cardiovascular diseases and related quantitative traits in the era of Genome-Wide Association Studies. Where have we been and what should we expect?

Ndeye Coumba NDIAYE[1], Mohsen AZIMI NEHZAD[1], Said EL SHAMIEH[1], Sophie VISVIKIS-SIEST[1*]

All authors contributed equally to this work

Authors' affiliation:

[1] "Cardiovascular Genetics" Research Unit EA4373, 30 rue Lionnois, Université Henri Poincaré – Nancy 1, F-54000 Nancy, France

* Corresponding author

Authors' e-mail addresses:

Ndeye Coumba NDIAYE coumba.ndiaye@inserm.fr

Mohsen AZIMI NEHZAD mo_azmepid@yahoo.com

Said EL SHAMIEH said.shamieh@gmail.com

Sophie VISVIKIS-SIEST Sophie.Visvikis-Siest@inserm.fr

Abstract

Genome-Wide Association Studies on cardiovascular diseases and related quantitative traits revealed numerous genetic variants only partly replicated, probably due to the heterogeneity of the clinical phenotypes and populations studied. Even if new mechanisms have been identified there is still a long way to go to identifying the causal variants. Genome-Wide Association Studies methodologies should, in the coming years, integrate gene-gene and gene-environment analyses in a global scientific strategy involving subsequent transcriptomic and proteomic investigations. The Genome-Wide Association Studies era holds many promises but we are just at the beginning.

Review

Cardiovascular diseases: common disorders with important genetic component

Cardiovascular diseases (CVD) are the world's largest killers claiming 17.1 million lives each year i.e. 29% of global deaths according to the World Health Organisation (WHO). In 2004, 7.2 million deaths were due to coronary heart disease (CHD) and 5.7 million to stroke, and their prominence is expected to increase: CVD are projected to remain the leading cause of death reaching 23.56 million by 2030. Risk algorithms such as the Framingham Risk Score for absolute risk[1], incorporating classical risk factors (age, gender, blood lipid concentrations, blood pressure, body mass index, family history and smoking habit), have been constructed for primary prevention strategies the efficacy of which has reached a plateau. It is thus necessary to integrate new risk factors and DNA-based tests assessing the genetic predisposition to CVD[2] in public health policies. This is much more important given the high genetic component of these diseases and of their related quantitative traits[2].

The aetiology of CVD results from a complex combination of interrelated environmental and constitutive risk factors that have been intensively studied in the last decades[3]. This complex architecture and its genetic background[4, 5] is still poorly understood and substantial discrepancies remain in estimating the heritability of numerous CVD related traits[6].

However, for many years now, developments in genetic analyses have been instrumental in highlighting important novel biological pathways in CVD susceptibilities and have provided a far better understanding of the fundamental inherent biological pathway leading to the development of the pathologies being considered.

Before the Genome-Wide Association Study (GWAS) era, genetics of phenotypes of interest (diseases or traits) were assessed by candidate gene approaches. Twins, families and population-based studies were used to estimate heritability, and even to explore interactions between genetic and environmental factors. Candidate genes were defined when the physiopathology of a disease was known (e.g. an enzyme deficiency) and all the results which followed relied on theoretical premises. Innovative discoveries needed a hypothesis-free approach. As far as this is concerned, since 2005, GWAS designs have revolutionized the genetic understanding of CVD and their associated quantitative traits[7-9]. In fact, recent technological advances enabling rapid and accurate genotyping of millions of single-nucleotide polymorphisms (SNPs) in large populations have led to conducting GWAS for different CVD phenotypes and related traits (Figure 1) and to the identification of hundreds of genetic factors, some robustly replicated (NHGRI Catalog of published genome-wide association studies, accessed September, 2010, http://www.genome.gov/gwastudies: common disease-common variant hypothesis)[10, 11] with small individual effects. Moreover, GWAS have revealed previously unsuspected pathological pathways[4, 12]. However, finding causal variants explaining the high heritability of the different CVD phenotypes has not yet achieved satisfactory results[13].

In this work we review GWAS findings concerning CVD and their related quantitative traits, missing heritability and future challenges.

Brief overview of Genome-Wide Association Studies on cardiovascular diseases and their related quantitative traits

To date, at least 650 GWAS focusing on about 150 distinct diseases and traits have reported more than 800 significant SNP-trait associations (genome-wide significance for $P < 5x10^{-8}$, http://www.genome.gov/gwastudies). At least, 16 have been conducted on cardiovascular events and have reported the association of 29 genetic variants found in 20 genes (supplementary Table 1). However, the scientific community pointed out that some caution is needed when interpreting data found in GWAS for CVD, given the weak sample size of the majority of these studies. The genetic variants are very rarely robustly linked to the cardiovascular events. Similarly, the lack of suggestive associations stressed the value of having accurate clinical subtype classification and standardised methodology for multicentre studies.

GWAS of the quantitative traits related to CVD could provide an alternative approach for better comprehending the phenotypic diversity of a given CVD as stated by the Wellcome Trust Case Control Consortium (WTCCC)[14].

Sixty percent of CVD and related trait GWAS are focused on obesity and have reported the association of about 60 variants in the vicinity of 35 genes (Figure 1). Less than two thirds of them, including some appreciable paediatric studies, have been replicated. Interestingly, the *FTO* gene was identified as the first locus harbouring common variants with an unequivocal impact on obesity predisposition and fat mass in population studies and being significantly associated with diabetes. The most important reports are summarised in Supplementary Table 1.

More than a quarter, at least, of all GWAS for CVD and related traits have studied fasting plasma glucose and diabetes mellitus (Supplementary Table 1). Up to ninety percent of them have been replicated in European, American and Asian populations. In two distinct studies, Zeggini et al.[15, 16] reported the association of type 2 diabetes mellitus with 22 SNPs next to the following genes: *CDKN2A, CDKN2B, FTO, HHEX, IGF2BP2, KCNJ11, CDKAL1, TCF7L2, THADA, TSPAN8, LGR5, VEGF, NOTCH2, ADAM30, PPARG, SLC30A8, JAZF1, SYN2, CDC, ADAMTS9, CDC123, CAMK1D*. Of these, two SNPS in the vicinity of *TCF7L2* had the greatest association with type 2 diabetes (Supplementary Table 1).

Lipid profile and blood pressure (BP) studies represent a considerable part of the GWAS undertaken (26% and 10% respectively, Figure 1). Concerning lipid profiles, 9 GWAS revealed 97 SNPs in 54 loci ($p< 10^{-6}$). About two-thirds of SNPs have been successfully replicated, notably 2 in the vicinity of *CETP* (Supplementary Table 1).

It is important to mention that although rs3764261 in *CETP* was strongly associated with total cholesterol, low-density lipoprotein cholesterol (LDL-C) and triglycerides (TG) ($p= 7x10^{-380}$)[17], this SNP was not related to coronary artery diseases.

It should be noted that caution needs to be exercised with the meta-analyses on hyperlipidaemic and diabetic individuals[17-19], since they have included some common subsamples: similar association results will not be surprising[20].

In addition to lipid profiles, replication studies have been relatively successful in confirming BP and hypertension (HTN) key loci[21-23]. In the largest joint meta-analysis published in May 2009, by CHARGE[24] and GlobakBPgen[4] consortia, researchers discovered 14 loci associated with systolic BP, diastolic BP and/or HTN. Recently, Takeuchi et al.[21] succeeded in replicating 7 loci in Japanese subjects, whereas Liu et al.[22] and Hong et al.[23] have confirmed some of them in Chinese Hans and Koreans respectively. These main findings constitute a milestone in bringing genome knowledge into BP regulation in diverse populations, as will be discussed below[25].

GWAS strengths and limitations

Current Genome-Wide Association Studies designs

GWAS involve various tiered designs including case-control and cohort studies for the identification of trait/disease-SNP associations. These tiered strategies allow affordable procedures to be carried out by conducting a genome-wide scan in a discovery set in order to select a core of significantly associated SNPs that will be subsequently genotyped in larger replication sets for confirmation[26]. Theses intricate designs involve more and more tiers according to the complexity and originality of the trait of interest[27] and allow the elimination of previous spurious associations (false positives)[11]. Traditionally, genetic associations having causal functionality and appearing in multiple populations and studies are considered reliable[11, 26].

Testing the association for millions of SNPs with a specific trait or disease implies a large number of statistical tests (at least one per SNP) making nominal significance (0.05) inappropriate for originally selecting associated variants. Such designs thus require large populations[28] and have led to large international consortia being set up in which substantial pooled GWAS have been conducted. Indeed, significant progress in the management of the methodological weaknesses of GWAS has been made under the auspices of large consortia providing numerous independent populations for pooling analyses and replications in specific fields, such as the Glucose and Insulin Related Traits Consortium (MAGIC)[29]. These pooling strategies required controls for differences in allele frequencies[30] to minimise false positive findings and reinforce the validity of genetic variants found to be associated with specific traits or pathologies.

Only 12% of SNPs discovered by all the GWAS performed to date and associated with traits are located in the vicinity of protein-coding regions of genes, although SNPs in protein-coding regions are heavily over-represented on genotyping arrays[31]. Forty percent fall in intergenic regions, and another 40% are located in introns. Initially raising concern in the scientific community, the repeated replication of signals falling consecutively in so-called gene deserts sharpened the focus on the potential roles of intronic, and particularly intergenic regions in regulating gene expression[7].

Future challenges

Over the past few years, GWAS have reported a large number of novel genetic variants associated with CVD[32, 33]. We now need to perform GWAS on populations with diverse geographic ancestries that have undergone more mutations and greater recombination events. This type of study can give greater degrees of genetic variation and shorter stretches of linkage disequilibrium allowing better localisation of genome-wide association signals[13, 15]. Several different ethnicities such as Japanese[21], African Americans[34] and mainly Europeans[4, 24] were examined. European populations are closely related[25], their genomes have undergone less recombination events than those of Africans for example[35], an advantage that has increased the choice of this specific descendent. Will we be able to replicate initial findings in non-European populations?

It might be asked why numerous worldwide replication studies confirmed loci but not SNPs. In fact, genetic variants having high occurrence in Europeans can differ in their frequencies across populations; many SNPs are rare in non-Europeans[36]. Moreover, allelic frequencies differ even if the same SNP is found in diverse populations[37].

GWAS expansion will ultimately allow us to identify “worldwide genetic variants” that are capable of influencing CVD and will further define the most suitable ethnicity to reveal a genotype-phenotype effect.

In this context, GWAS involving African populations are potentially of interest. Low linkage disequilibrium between genetic variants is one of their main characteristics, due in part to a long history of recombination and mutations[25]. When SNPs extend over short genomic regions, further sequencing may be very effective in defining the true risk variant[25].

What will be the application of these findings in clinical practice and for predicting CVD risk? Kathiresan and colleagues integrated genetics into cardiovascular risk scores on the basis of nine SNPs associated with cholesterol levels[38]. But does it add anything more than algorithms previously constructed on the basis of phenotypes[39]? Even strongly associated with the risk of CVD, after adjustment for standard risk factors, the new score developed by Kathiresan et al. did not change the receiver-operating-characteristic curve (AUC)[38]. Similar

attempts in prostate cancer gave the same result[40]. Finding the causal variants implicated in these pathologies could overcome this problem.

Missing heritability in cardiovascular diseases, an Achilles' heel

Recent GWAS had an enormous sample size (≈100,000 individuals) but explained only about 10% of heritability[17], the "dark matter" of GWAS.

GWAS have focused on common variants (allelic frequency >5% in the general population)[35] and ignored low minor allele frequency (MAF) (0.5% <MAF<5% in the general population) and rare ones (MAF<0.5%) that are predicted to have a greater risk effect[35]. In contrast to common variants thought to be old in the history of humanity, rare variants are more recent and therefore not geographically extended[25]. Detecting associations with rare variants will be the first step to better understanding of their role in CVD. When focusing on rare genetic variants, GWAS require very large populations and different statistical methods that jointly analyse variants in a locus instead of testing each variant individually. The next generation of dense arrays with more genome-wide covering are making this goal feasible.

Structural large variants (>500Kb) such as deletions and duplications (Copy number variants) represent rare variants that may affect gene expression and pathways in humans[5]. These genomic imbalances occur at an allele frequency of <0.05% and are present in about 8% of the total population[5], making them rare but collectively common. Results available suggest that copy number variants strongly affect gene expression, thereby affecting mRNA splicing and transcriptional activity[41].

Gene-environment interactions (GxE) are also issues that blind many GWAS. Integrating gene-environment interactions could help in highlighting different loci effects according to modifiable risk factors, particularly in CVD. This strategy needs large consortia, methodological efforts and stratified designs in order to improve the existing tools for GWAS that do not yet incorporate complex statistics. Although, environmental pathways acting through epigenetic mechanisms to modify gene expression[42] and DNA methylation[41] have given some fruitful results in other chronic diseases[43] and exposures[44], they are yet to be studied in CVD. In addition to the usual challenges reported for genetic association studies, a successful GxE study should take into account sample size, exposure assessment and heterogeneity, described in full elsewhere[42]. GxE in GWAS requires enormous populations[42]. Smith and Day[45] explained that detecting an interaction needed a sample size at least four times greater than that required for detecting a main effect of comparable magnitude. Although difficult, studying GxE interactions could help in understanding discrepancies due to heterogeneous exposure.

Gene-gene (epistatic) interactions may also play an important role in discovering genes that have not yet been found by the consensual single-locus approach. This statement has been

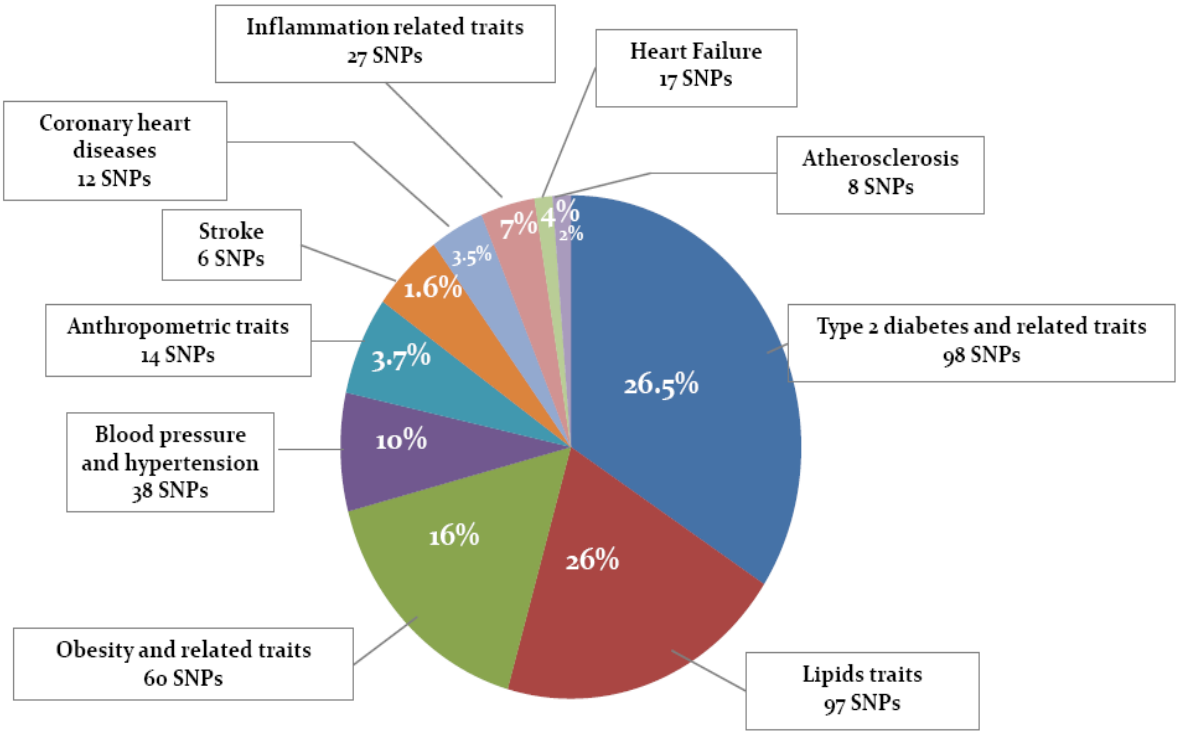
extensively reviewed[8, 46, 47] and both parametric and nonparametric multi-locus methods have been developed to detect such interactions[48] in the last years. Epistatic interactions have been documented for susceptibility to cancer[49], morphology[50] and autoimmune conditions[51]. However, current epistatic designs do not have genome-wide coverage, so that their application to high-dimensional genome-wide data including all imaginable SNPs remains a crucial challenge[52].

Furthermore, we advise that all loci associated with CVD and related traits should be replicated in paediatric populations, as a common variant could have a continuous impact throughout life. Epidemiological and pathophysiological evidence suggests that the precursors of CVD originate in childhood[53, 54]: the atherosclerotic process starts in childhood[55] and numerous studies have shown that risk factors such as increased BP levels during childhood strongly predict HTN in adults[56]. Isolating genetic variations that may influence a given risk factor for CVD at childhood, where many environmental factors such as alcohol intake, stress and smoking are absent, might therefore have major implications for public health and could be a challenge in designing primary preventive strategies.

Concluding remarks and perspectives

GWAS have revealed many novel genetic risk variants in CVD, making it an auspicious era for understanding genetics better. For the time being we need to grasp the unexpected involvement of certain functional and mechanistic pathways identified in CVD processes involving many quantitative traits. These have to be extensively investigated in parallel with the pathology itself. Future GWAS must also detect low MAF and rare variants, and include GxG, GxE and well-designed candidate gene studies (in term of statistical power, homogeneity, and exposure assessment) and involve subsequent transcriptomic and proteomic investigations. It is time to pause and think, instead of digging down and down for more genetic loci that will have even smaller phenotypic effects as the number of individuals studied rises exponentially.

Figure 1: Published Genome Wide Association Studies on different cardiovascular diseases and related traits (http://hugenavigator.net/HuGENavigator/gWAHitStartPage.do).



Inflammation related traits: IL-18, C-reactive protein, adiponectin and homocystein.
Type 2 diabetes and related traits: type 2 diabetes, fasting plasma glucose and 2 hours glucose challenge.
Lipid traits: Low density lipoprotein cholesterol, High density lipoprotein cholesterol, triglyceride and total cholesterol.
Obesity and related traits: obesity and body mass index.
Blood pressure and hypertension : Diastolic and systolic blood pressure.
All SNPs are found in Huge navigator site[57].

References

1 Wilson, P. W., D'Agostino, R. B., Levy, D., Belanger, A. M., Silbershatz, H. and Kannel, W. B., Prediction of coronary heart disease using risk factor categories, Circulation, 1998, 97: 1837-1847.

2 Casas, J. P., Cooper, J., Miller, G. J., Hingorani, A. D. and Humphries, S. E., Investigating the genetic determinants of cardiovascular disease using candidate genes and meta-analysis of association studies, Ann Hum Genet, 2006, 70: 145-169.

3 Cooney, M. T., Dudina, A., D'Agostino, R. and Graham, I. M., Cardiovascular risk-estimation systems in primary prevention: do they differ? Do they make a difference? Can we see the future?, Circulation, 2010, 122: 300-310.

4 Levy, D., Ehret, G. B., Rice, K., Verwoert, G. C., Launer, L. J., Dehghan, A., Glazer, N. L., Morrison, A. C., Johnson, A. D., Aspelund, T., Aulchenko, Y., Lumley, T., Kottgen, A., Vasan, R. S., Rivadeneira, F., Eiriksdottir, G., Guo, X., Arking, D. E., Mitchell, G. F., Mattace-Raso, F. U., Smith, A. V., Taylor, K., Scharpf, R. B., Hwang, S. J., Sijbrands, E. J., Bis, J., Harris, T. B., Ganesh, S. K., O'Donnell, C. J., Hofman, A., Rotter, J. I., Coresh, J., Benjamin, E. J., Uitterlinden, A. G., Heiss, G., Fox, C. S., Witteman, J. C., Boerwinkle, E., Wang, T. J., Gudnason, V., Larson, M. G., Chakravarti, A., Psaty, B. M. and van Duijn, C. M., Genome-wide association study of blood pressure and hypertension, Nat Genet, 2009, 41: 677-687.

5 Eichler, E. E., Flint, J., Gibson, G., Kong, A., Leal, S. M., Moore, J. H. and Nadeau, J. H., Missing heritability and strategies for finding the underlying causes of complex disease, Nat Rev Genet, 2010, 11: 446-450.

6 Levy, D., DeStefano, A. L., Larson, M. G., O'Donnell, C. J., Lifton, R. P., Gavras, H., Cupples, L. A. and Myers, R. H., Evidence for a gene influencing blood pressure on chromosome 17. Genome scan linkage results for longitudinal blood pressure phenotypes in subjects from the framingham heart study, Hypertension, 2000, 36: 477-483.

7 Hardy, J. and Singleton, A., Genomewide association studies and human disease, N Engl J Med, 2009, 360: 1759-1768.

8 Manolio, T. A., Genomewide association studies and assessment of the risk of disease, N Engl J Med, 2010, 363: 166-176.

9 Manolio, T. A., Brooks, L. D. and Collins, F. S., A HapMap harvest of insights into the genetics of common disease, J Clin Invest, 2008, 118: 1590-1605.

10 Reich, D. E. and Lander, E. S., On the allelic spectrum of human disease, Trends Genet, 2001, 17: 502-510.

11 Wang, W. Y., Barratt, B. J., Clayton, D. G. and Todd, J. A., Genome-wide association studies: theoretical and practical concerns, Nat Rev Genet, 2005, 6: 109-118.

12 Willer, C. J., Speliotes, E. K., Loos, R. J., Li, S., Lindgren, C. M., Heid, I. M., Berndt, S. I., Elliott, A. L., Jackson, A. U., Lamina, C., Lettre, G., Lim, N., Lyon, H. N., McCarroll, S. A., Papadakis, K., Qi, L., Randall, J. C., Roccasecca, R. M., Sanna, S., Scheet, P., Weedon, M. N., Wheeler, E., Zhao, J. H., Jacobs, L. C., Prokopenko, I., Soranzo, N., Tanaka, T., Timpson, N. J., Almgren, P., Bennett, A., Bergman, R. N., Bingham, S. A., Bonnycastle, L. L., Brown, M., Burtt, N. P., Chines, P., Coin, L., Collins, F. S., Connell, J. M., Cooper, C., Smith, G. D., Dennison, E. M., Deodhar, P., Elliott, P., Erdos, M. R., Estrada, K., Evans, D. M., Gianniny, L., Gieger, C., Gillson, C. J., Guiducci, C., Hackett, R., Hadley, D., Hall, A. S., Havulinna, A. S., Hebebrand, J., Hofman, A., Isomaa, B., Jacobs, K. B., Johnson, T., Jousilahti, P., Jovanovic, Z., Khaw, K. T., Kraft, P., Kuokkanen, M., Kuusisto, J., Laitinen, J., Lakatta, E. G., Luan, J., Luben, R. N., Mangino, M., McArdle, W. L., Meitinger, T., Mulas, A., Munroe, P. B., Narisu, N., Ness, A. R., Northstone, K., O'Rahilly, S., Purmann, C., Rees, M. G., Ridderstrale, M., Ring, S. M., Rivadeneira, F., Ruokonen, A., Sandhu, M. S., Saramies, J., Scott, L. J., Scuteri, A., Silander, K., Sims, M. A., Song, K., Stephens, J., Stevens, S., Stringham, H. M., Tung, Y. C., Valle, T. T., Van Duijn, C. M., Vimaleswaran, K. S., Vollenweider, P., Waeber, G., Wallace, C., Watanabe, R. M., Waterworth, D. M., Watkins, N., Witteman, J. C., Zeggini, E., Zhai, G., Zillikens, M. C., Altshuler, D., Caulfield, M. J., Chanock, S. J.,

Farooqi, I. S., Ferrucci, L., Guralnik, J. M., Hattersley, A. T., Hu, F. B., Jarvelin, M. R., Laakso, M., Mooser, V., Ong, K. K., Ouwehand, W. H., Salomaa, V., Samani, N. J., Spector, T. D., Tuomi, T., Tuomilehto, J., Uda, M., Uitterlinden, A. G., Wareham, N. J., Deloukas, P., Frayling, T. M., Groop, L. C., Hayes, R. B., Hunter, D. J., Mohlke, K. L., Peltonen, L., Schlessinger, D., Strachan, D. P., Wichmann, H. E., McCarthy, M. I., Boehnke, M., Barroso, I., Abecasis, G. R. and Hirschhorn, J. N., Six new loci associated with body mass index highlight a neuronal influence on body weight regulation, Nat Genet, 2009, 41: 25-34.

13 Manolio, T. A., Collins, F. S., Cox, N. J., Goldstein, D. B., Hindorff, L. A., Hunter, D. J., McCarthy, M. I., Ramos, E. M., Cardon, L. R., Chakravarti, A., Cho, J. H., Guttmacher, A. E., Kong, A., Kruglyak, L., Mardis, E., Rotimi, C. N., Slatkin, M., Valle, D., Whittemore, A. S., Boehnke, M., Clark, A. G., Eichler, E. E., Gibson, G., Haines, J. L., Mackay, T. F., McCarroll, S. A. and Visscher, P. M., Finding the missing heritability of complex diseases, Nature, 2009, 461: 747-753.

14 Ouwehand, W. H., The discovery of genes implicated in myocardial infarction, J Thromb Haemost, 2009, 7 Suppl 1: 305-307.

15 Zeggini, E., Weedon, M. N., Lindgren, C. M., Frayling, T. M., Elliott, K. S., Lango, H., Timpson, N. J., Perry, J. R., Rayner, N. W., Freathy, R. M., Barrett, J. C., Shields, B., Morris, A. P., Ellard, S., Groves, C. J., Harries, L. W., Marchini, J. L., Owen, K. R., Knight, B., Cardon, L. R., Walker, M., Hitman, G. A., Morris, A. D., Doney, A. S., McCarthy, M. I. and Hattersley, A. T., Replication of genome-wide association signals in UK samples reveals risk loci for type 2 diabetes, Science, 2007, 316: 1336-1341.

16 Zeggini, E., Scott, L. J., Saxena, R., Voight, B. F., Marchini, J. L., Hu, T., de Bakker, P. I., Abecasis, G. R., Almgren, P., Andersen, G., Ardlie, K., Bostrom, K. B., Bergman, R. N., Bonnycastle, L. L., Borch-Johnsen, K., Burtt, N. P., Chen, H., Chines, P. S., Daly, M. J., Deodhar, P., Ding, C. J., Doney, A. S., Duren, W. L., Elliott, K. S., Erdos, M. R., Frayling, T. M., Freathy, R. M., Gianniny, L., Grallert, H., Grarup, N., Groves, C. J., Guiducci, C., Hansen, T., Herder, C., Hitman, G. A., Hughes, T. E., Isomaa, B., Jackson, A. U., Jorgensen, T., Kong, A., Kubalanza, K., Kuruvilla, F. G., Kuusisto, J., Langenberg, C., Lango, H., Lauritzen, T., Li, Y., Lindgren, C. M., Lyssenko, V., Marvelle, A. F., Meisinger, C., Midthjell, K., Mohlke, K. L., Morken, M. A., Morris, A. D., Narisu, N., Nilsson, P., Owen, K. R., Palmer, C. N., Payne, F., Perry, J. R., Pettersen, E., Platou, C., Prokopenko, I., Qi, L., Qin, L., Rayner, N. W., Rees, M., Roix, J. J., Sandbaek, A., Shields, B., Sjogren, M., Steinthorsdottir, V., Stringham, H. M., Swift, A. J., Thorleifsson, G., Thorsteinsdottir, U., Timpson, N. J., Tuomi, T., Tuomilehto, J., Walker, M., Watanabe, R. M., Weedon, M. N., Willer, C. J., Illig, T., Hveem, K., Hu, F. B., Laakso, M., Stefansson, K., Pedersen, O., Wareham, N. J., Barroso, I., Hattersley, A. T., Collins, F. S., Groop, L., McCarthy, M. I., Boehnke, M. and Altshuler, D., Meta-analysis of genome-wide association data and large-scale replication identifies additional susceptibility loci for type 2 diabetes, Nat Genet, 2008, 40: 638-645.

17 Teslovich, T. M., Musunuru, K., Smith, A. V., Edmondson, A. C., Stylianou, I. M., Koseki, M., Pirruccello, J. P., Ripatti, S., Chasman, D. I., Willer, C. J., Johansen, C. T., Fouchier, S. W., Isaacs, A., Peloso, G. M., Barbalic, M., Ricketts, S. L., Bis, J. C., Aulchenko, Y. S., Thorleifsson, G., Feitosa, M. F., Chambers, J., Orho-Melander, M., Melander, O., Johnson, T., Li, X., Guo, X., Li, M., Shin Cho, Y., Jin Go, M., Jin Kim, Y., Lee, J. Y., Park, T., Kim, K., Sim, X., Twee-Hee Ong, R., Croteau-Chonka, D. C., Lange, L. A., Smith, J. D., Song, K., Hua Zhao, J., Yuan, X., Luan, J., Lamina, C., Ziegler, A., Zhang, W., Zee, R. Y., Wright, A. F., Witteman, J. C., Wilson, J. F., Willemsen, G., Wichmann, H. E., Whitfield, J. B., Waterworth, D. M., Wareham, N. J., Waeber, G., Vollenweider, P., Voight, B. F., Vitart, V., Uitterlinden, A. G., Uda, M., Tuomilehto, J., Thompson, J. R., Tanaka, T., Surakka, I., Stringham, H. M., Spector, T. D., Soranzo, N., Smit, J. H., Sinisalo, J., Silander, K., Sijbrands, E. J., Scuteri, A., Scott, J., Schlessinger, D., Sanna, S., Salomaa, V., Saharinen, J., Sabatti, C., Ruokonen, A., Rudan, I., Rose, L. M., Roberts, R., Rieder, M., Psaty, B. M., Pramstaller, P. P., Pichler, I., Perola, M., Penninx, B. W., Pedersen, N. L., Pattaro, C., Parker, A. N., Pare, G., Oostra, B. A., O'Donnell, C. J., Nieminen, M. S., Nickerson, D. A., Montgomery, G. W., Meitinger, T., McPherson, R., McCarthy, M. I., McArdle, W., Masson, D., Martin, N. G., Marroni, F., Mangino, M., Magnusson, P. K., Lucas, G., Luben, R., Loos, R. J., Lokki, M. L., Lettre, G., Langenberg, C., Launer, L. J., Lakatta, E. G., Laaksonen, R., Kyvik, K. O., Kronenberg, F., Konig, I. R., Khaw, K. T., Kaprio, J., Kaplan, L. M., Johansson, A., Jarvelin, M. R., Janssens, A. C., Ingelsson, E., Igl, W., Kees Hovingh, G., Hottenga, J. J., Hofman, A., Hicks, A. A.,

Hengstenberg, C., Heid, I. M., Hayward, C., Havulinna, A. S., Hastie, N. D., Harris, T. B., Haritunians, T., Hall, A. S., Gyllensten, U., Guiducci, C., Groop, L. C., Gonzalez, E., Gieger, C., Freimer, N. B., Ferrucci, L., Erdmann, J., Elliott, P., Ejebe, K. G., Doring, A., Dominiczak, A. F., Demissie, S., Deloukas, P., de Geus, E. J., de Faire, U., Crawford, G., Collins, F. S., Chen, Y. D., Caulfield, M. J., Campbell, H., Burtt, N. P., Bonnycastle, L. L., Boomsma, D. I., Boekholdt, S. M., Bergman, R. N., Barroso, I., Bandinelli, S., Ballantyne, C. M., Assimes, T. L., Quertermous, T., Altshuler, D., Seielstad, M., Wong, T. Y., Tai, E. S., Feranil, A. B., Kuzawa, C. W., Adair, L. S., Taylor, H. A., Jr., Borecki, I. B., Gabriel, S. B., Wilson, J. G., Holm, H., Thorsteinsdottir, U., Gudnason, V., Krauss, R. M., Mohlke, K. L., Ordovas, J. M., Munroe, P. B., Kooner, J. S., Tall, A. R., Hegele, R. A., Kastelein, J. J., Schadt, E. E., Rotter, J. I., Boerwinkle, E., Strachan, D. P., Mooser, V., Stefansson, K., Reilly, M. P., Samani, N. J., Schunkert, H., Cupples, L. A., Sandhu, M. S., Ridker, P. M., Rader, D. J., van Duijn, C. M., Peltonen, L., Abecasis, G. R., Boehnke, M. and Kathiresan, S., Biological, clinical and population relevance of 95 loci for blood lipids, Nature, 2010, 466: 707-713.

18 Kathiresan, S., Melander, O., Guiducci, C., Surti, A., Burtt, N. P., Rieder, M. J., Cooper, G. M., Roos, C., Voight, B. F., Havulinna, A. S., Wahlstrand, B., Hedner, T., Corella, D., Tai, E. S., Ordovas, J. M., Berglund, G., Vartiainen, E., Jousilahti, P., Hedblad, B., Taskinen, M. R., Newton-Cheh, C., Salomaa, V., Peltonen, L., Groop, L., Altshuler, D. M. and Orho-Melander, M., Six new loci associated with blood low-density lipoprotein cholesterol, high-density lipoprotein cholesterol or triglycerides in humans, Nat Genet, 2008, 40: 189-197.

19 Willer, C. J., Sanna, S., Jackson, A. U., Scuteri, A., Bonnycastle, L. L., Clarke, R., Heath, S. C., Timpson, N. J., Najjar, S. S., Stringham, H. M., Strait, J., Duren, W. L., Maschio, A., Busonero, F., Mulas, A., Albai, G., Swift, A. J., Morken, M. A., Narisu, N., Bennett, D., Parish, S., Shen, H., Galan, P., Meneton, P., Hercberg, S., Zelenika, D., Chen, W. M., Li, Y., Scott, L. J., Scheet, P. A., Sundvall, J., Watanabe, R. M., Nagaraja, R., Ebrahim, S., Lawlor, D. A., Ben-Shlomo, Y., Davey-Smith, G., Shuldiner, A. R., Collins, R., Bergman, R. N., Uda, M., Tuomilehto, J., Cao, A., Collins, F. S., Lakatta, E., Lathrop, G. M., Boehnke, M., Schlessinger, D., Mohlke, K. L. and Abecasis, G. R., Newly identified loci that influence lipid concentrations and risk of coronary artery disease, Nat Genet, 2008, 40: 161-169.

20 Hegele, R. A., Plasma lipoproteins: genetic influences and clinical implications, Nat Rev Genet, 2009, 10: 109-121.

21 Takeuchi, F., Isono, M., Katsuya, T., Yamamoto, K., Yokota, M., Sugiyama, T., Nabika, T., Fujioka, A., Ohnaka, K., Asano, H., Yamori, Y., Yamaguchi, S., Kobayashi, S., Takayanagi, R., Ogihara, T. and Kato, N., Blood pressure and hypertension are associated with 7 loci in the Japanese population, Circulation, 2010, 121: 2302-2309.

22 Liu, C., Li, H., Qi, Q., Lu, L., Gan, W., Loos, R. J. and Lin, X., Common variants in or near FGF5, CYP17A1 and MTHFR genes are associated with blood pressure and hypertension in Chinese Hans, J Hypertens, 2010.

23 Hong, K. W., Jin, H. S., Lim, J. E., Kim, S., Go, M. J. and Oh, B., Recapitulation of two genomewide association studies on blood pressure and essential hypertension in the Korean population, J Hum Genet, 2010, 55: 336-341.

24 Newton-Cheh, C., Johnson, T., Gateva, V., Tobin, M. D., Bochud, M., Coin, L., Najjar, S. S., Zhao, J. H., Heath, S. C., Eyheramendy, S., Papadakis, K., Voight, B. F., Scott, L. J., Zhang, F., Farrall, M., Tanaka, T., Wallace, C., Chambers, J. C., Khaw, K. T., Nilsson, P., van der Harst, P., Polidoro, S., Grobbee, D. E., Onland-Moret, N. C., Bots, M. L., Wain, L. V., Elliott, K. S., Teumer, A., Luan, J., Lucas, G., Kuusisto, J., Burton, P. R., Hadley, D., McArdle, W. L., Brown, M., Dominiczak, A., Newhouse, S. J., Samani, N. J., Webster, J., Zeggini, E., Beckmann, J. S., Bergmann, S., Lim, N., Song, K., Vollenweider, P., Waeber, G., Waterworth, D. M., Yuan, X., Groop, L., Orho-Melander, M., Allione, A., Di Gregorio, A., Guarrera, S., Panico, S., Ricceri, F., Romanazzi, V., Sacerdote, C., Vineis, P., Barroso, I., Sandhu, M. S., Luben, R. N., Crawford, G. J., Jousilahti, P., Perola, M., Boehnke, M., Bonnycastle, L. L., Collins, F. S., Jackson, A. U., Mohlke, K. L., Stringham, H. M., Valle, T. T., Willer, C. J., Bergman, R. N., Morken, M. A., Doring, A., Gieger, C., Illig, T., Meitinger, T., Org, E., Pfeufer, A., Wichmann, H. E., Kathiresan, S., Marrugat, J., O'Donnell, C. J., Schwartz, S. M., Siscovick, D. S., Subirana, I., Freimer, N. B., Hartikainen, A. L., McCarthy, M. I., O'Reilly, P. F., Peltonen, L., Pouta, A., de Jong, P. E., Snieder, H., van Gilst, W.

H., Clarke, R., Goel, A., Hamsten, A., Peden, J. F., Seedorf, U., Syvanen, A. C., Tognoni, G., Lakatta, E. G., Sanna, S., Scheet, P., Schlessinger, D., Scuteri, A., Dorr, M., Ernst, F., Felix, S. B., Homuth, G., Lorbeer, R., Reffelmann, T., Rettig, R., Volker, U., Galan, P., Gut, I. G., Hercberg, S., Lathrop, G. M., Zelenika, D., Deloukas, P., Soranzo, N., Williams, F. M., Zhai, G., Salomaa, V., Laakso, M., Elosua, R., Forouhi, N. G., Volzke, H., Uiterwaal, C. S., van der Schouw, Y. T., Numans, M. E., Matullo, G., Navis, G., Berglund, G., Bingham, S. A., Kooner, J. S., Connell, J. M., Bandinelli, S., Ferrucci, L., Watkins, H., Spector, T. D., Tuomilehto, J., Altshuler, D., Strachan, D. P., Laan, M., Meneton, P., Wareham, N. J., Uda, M., Jarvelin, M. R., Mooser, V., Melander, O., Loos, R. J., Elliott, P., Abecasis, G. R., Caulfield, M. and Munroe, P. B., Genome-wide association study identifies eight loci associated with blood pressure, Nat Genet, 2009, 41: 666-676.

25 Rosenberg, N. A., Huang, L., Jewett, E. M., Szpiech, Z. A., Jankovic, I. and Boehnke, M., Genome-wide association studies in diverse populations, Nat Rev Genet, 2010, 11: 356-366.

26 Chanock, S.J., Manolio, T., Boehnke, M., Boerwinkle, E., Hunter, D. J., Thomas, G., Hirschhorn, J. N., Abecasis, G., Altshuler, D., Bailey-Wilson, J. E., Brooks, L. D., Cardon, L. R., Daly, M., Donnelly, P., Fraumeni, J. F., Jr., Freimer, N. B., Gerhard, D. S., Gunter, C., Guttmacher, A. E., Guyer, M. S., Harris, E. L., Hoh, J., Hoover, R., Kong, C. A., Merikangas, K. R., Morton, C. C., Palmer, L. J., Phimister, E. G., Rice, J. P., Roberts, J., Rotimi, C., Tucker, M. A., Vogan, K. J., Wacholder, S., Wijsman, E. M., Winn, D. M. and Collins, F. S., Replicating genotype-phenotype associations. Nature, 2007, 447:655-60

27 Hoover, R. N., The evolution of epidemiologic research: from cottage industry to "big" science, Epidemiology, 2007, 18: 13-17.

28 Hunter, D. J. and Kraft, P., Drinking from the fire hose--statistical issues in genomewide association studies, N Engl J Med, 2007, 357: 436-439.

29 Dupuis, J., Langenberg, C., Prokopenko, I., Saxena, R., Soranzo, N., Jackson, A. U., Wheeler, E., Glazer, N. L., Bouatia-Naji, N., Gloyn, A. L., Lindgren, C. M., Magi, R., Morris, A. P., Randall, J., Johnson, T., Elliott, P., Rybin, D., Thorleifsson, G., Steinthorsdottir, V., Henneman, P., Grallert, H., Dehghan, A., Hottenga, J. J., Franklin, C. S., Navarro, P., Song, K., Goel, A., Perry, J. R. B., Egan, J. M., Lajunen, T., Grarup, N., Sparso, T., Doney, A., Voight, B. F., Stringham, H. M., Li, M., Kanoni, S., Shrader, P., Cavalcanti-Proenca, C., Kumari, M., Qi, L., Timpson, N. J., Gieger, C., Zabena, C., Rocheleau, G., Ingelsson, E., An, P., O'Connell, J., Luan, J. a., Elliott, A., McCarroll, S. A., Payne, F., Roccasecca, R. M., Pattou, F., Sethupathy, P., Ardlie, K., Ariyurek, Y., Balkau, B., Barter, P., Beilby, J. P., Ben-Shlomo, Y., Benediktsson, R., Bennett, A. J., Bergmann, S., Bochud, M., Boerwinkle, E., Bonnefond, A., Bonnycastle, L. L., Borch-Johnsen, K., Bottcher, Y., Brunner, E., Bumpstead, S. J., Charpentier, G., Chen, Y.-D. I., Chines, P., Clarke, R., Coin, L. J. M., Cooper, M. N., Cornelis, M., Crawford, G., Crisponi, L., Day, I. N. M., de Geus, E. J. C., Delplanque, J., Dina, C., Erdos, M. R., Fedson, A. C., Fischer-Rosinsky, A., Forouhi, N. G., Fox, C. S., Frants, R., Franzosi, M. G., Galan, P., Goodarzi, M. O., Graessler, J., Groves, C. J., Grundy, S., Gwilliam, R., Gyllensten, U., Hadjadj, S., Hallmans, G., Hammond, N., Han, X., Hartikainen, A.-L., Hassanali, N., Hayward, C., Heath, S. C., Hercberg, S., Herder, C., Hicks, A. A., Hillman, D. R., Hingorani, A. D., Hofman, A., Hui, J., Hung, J., Isomaa, B., Johnson, P. R. V., Jorgensen, T., Jula, A., Kaakinen, M., Kaprio, J., Kesaniemi, Y. A., Kivimaki, M., Knight, B., Koskinen, S., Kovacs, P., Kyvik, K. O., Lathrop, G. M., Lawlor, D. A., Le Bacquer, O., Lecoeur, C., Li, Y., Lyssenko, V., Mahley, R., Mangino, M., Manning, A. K., Martinez-Larrad, M. T., McAteer, J. B., McCulloch, L. J., McPherson, R., Meisinger, C., Melzer, D., Meyre, D., Mitchell, B. D., Morken, M. A., Mukherjee, S., Naitza, S., Narisu, N., Neville, M. J., Oostra, B. A., Orru, M., Pakyz, R., Palmer, C. N. A., Paolisso, G., Pattaro, C., Pearson, D., Peden, J. F., Pedersen, N. L., Perola, M., Pfeiffer, A. F. H., Pichler, I., Polasek, O., Posthuma, D., Potter, S. C., Pouta, A., Province, M. A., Psaty, B. M., Rathmann, W., Rayner, N. W., Rice, K., Ripatti, S., Rivadeneira, F., Roden, M., Rolandsson, O., Sandbaek, A., Sandhu, M., Sanna, S., Sayer, A. A., Scheet, P., Scott, L. J., Seedorf, U., Sharp, S. J., Shields, B., Sigurethsson, G., Sijbrands, E. J. G., Silveira, A., Simpson, L., Singleton, A., Smith, N. L., Sovio, U., Swift, A., Syddall, H., Syvanen, A.-C., Tanaka, T., Thorand, B., Tichet, J., Tonjes, A., Tuomi, T., Uitterlinden, A. G., van Dijk, K. W., van Hoek, M., Varma, D., Visvikis-Siest, S., Vitart, V., Vogelzangs, N., Waeber, G., Wagner, P. J., Walley, A., Walters, G. B., Ward, K. L., Watkins, H., Weedon, M. N., Wild, S. H., Willemsen, G., Witteman, J. C. M., Yarnell, J. W. G., Zeggini, E., Zelenika, D., Zethelius, B., Zhai, G.,

Zhao, J. H., Zillikens, M. C., Borecki, I.B., Loos, R. J. F., Meneton, P., Magnusson, P. K. E., Nathan, D. M., Williams, G. H., Hattersley, A. T., Silander, K., Salomaa, V., Smith, G. D., Bornstein, S. R., Schwarz, P., Spranger, J., Karpe, F., Shuldiner, A. R., Cooper, C., Dedoussis, G. V., Serrano-Rios, M., Morris, A. D., Lind, L., Palmer, L. J., Hu, F. B., Franks, P. W., Ebrahim, S., Marmot, M., Kao, W. H. L., Pankow, J. S., Sampson, M. J., Kuusisto, J., Laakso, M., Hansen, T., Pedersen, O., Pramstaller, P. P., Wichmann, H. E., Illig, T., Rudan, I., Wright, A. F., Stumvoll, M., Campbell, H., Wilson, J. F., Bergman, R.N., Buchanan, T. A., Collins, F. S., Mohlke, K. L., Tuomilehto, J., Valle, T. T., Altshuler, D., Rotter, J. I., Siscovick, D. S., Penninx, B. W. J. H., Boomsma, D. I., Deloukas, P., Spector, T. D., Frayling, T. M., Ferrucci, L., Kong, A., Thorsteinsdottir, U., Stefansson, K., van Duijn, C. M., Aulchenko, Y. S., Cao, A., Scuteri, A., Schlessinger, D., Uda, M., Ruokonen, A., Jarvelin, M.-R., Waterworth, D. M., Vollenweider, P., Peltonen, L., Mooser, V., Abecasis, G. R., Wareham, N. J., Sladek, R., Froguel, P., Watanabe, R. M., Meigs, J. B., Groop, L., Boehnke, M., McCarthy, M. I., Florez, J. C. and Barroso, I., New genetic loci implicated in fasting glucose homeostasis and their impact on type 2 diabetes risk. Nat Genet, 2010, 42:105-16.

30 Pearson, T.A. and Manolio, T. A., How to interpret a genome-wide association study, JAMA, 2008, 299:1335-44.

31 Hindorff, L.A., Sethupathy, P., Junkins, H. A., Ramos, E. M., Mehta, J. P., Collins, F. S. and Manolio, T. A., Potential etiologic and functional implications of genome-wide association loci for human diseases and traits, Proc Nat Acad Sci U S A, 2009, 106:9362-7.

32 Preuss, M., Konig, I. R., Thompson, J. R., Erdmann, J., Absher, D., Assimes, T. L., Blankenberg, S., Boerwinkle, E., Chen, L., Cupples, L. A., Hall, A. S., Halperin, E., Hengstenberg, C., Holm, H., Laaksonen, R., Li, M., Marz, W., McPherson, R., Musunuru, K., Nelson, C. P., Burnett, M. S., Epstein, S. E., O'Donnell, C. J., Quertermous, T., Rader, D. J., Roberts, R., Schillert, A., Stefansson, K., Stewart, A. F., Thorleifsson, G., Thorsteinsdottir, U., Voight, B. F., Wells, G. A., Ziegler, A., Kathiresan, S., Reilly, M. P., Samani, N. J. and Schunkert, H., Design of the Coronary Artery Disease Genome-Wide Replication and Meta-Analysis (CARDIoGRAM) Study -- A Genome-Wide Association Meta-Analysis Involving More than 22,000 Cases and 60,000 Controls, Circ Cardiovasc Genet, 2010, 3:475-83.

33 Yang, Q., Kottgen, A., Dehghan, A., Smith, A. V., Glazer, N. L., Chen, M. H., Chasman, D. I., Aspelund, T., Eiriksdottir, G., Harris, T. B., Launer, L., Nalls, M., Hernandez, D., Arking, D. E., Boerwinkle, E., Grove, M. L., Li, M., Kao, W. L., Chonchol, M., Haritunians, T., Li, G., Lumley, T., Psaty, B. M., Shlipak, M., Hwang, S. J., Larson, M. G., O'Donnell, C. J., Upadhyay, A., van Duijn, C. M., Hofman, A., Rivadeneira, F., Stricker, B., Uitterlinden, A. G., Pare, G., Parker, A. N., Ridker, P. M., Siscovick, D. S., Gudnason, V., Witteman, J. C., Fox, C. S. and Coresh, J., Multiple Genetic Loci Influence Serum Urate And Their Relationship With Gout and Cardiovascular Disease Risk Factors, Circ Cardiovasc Genet, 2010.

34 Adeyemo, A., Gerry, N., Chen, G., Herbert, A., Doumatey, A., Huang, H., Zhou, J., Lashley, K., Chen, Y., Christman, M. and Rotimi, C., A genome-wide association study of hypertension and blood pressure in African Americans, PLoS Genet, 2009, 5: e1000564.

35 Ehret, G. B., Genome-wide association studies: contribution of genomics to understanding blood pressure and essential hypertension, Curr Hypertens Rep, 2010, 12: 17-25.

36 Dhandapany, P. S., Sadayappan, S., Xue, Y., Powell, G. T., Rani, D. S., Nallari, P., Rai, T. S., Khullar, M., Soares, P., Bahl, A., Tharkan, J. M., Vaideeswar, P., Rathinavel, A., Narasimhan, C., Ayapati, D. R., Ayub, Q., Mehdi, S. Q., Oppenheimer, S., Richards, M. B., Price, A. L., Patterson, N., Reich, D., Singh, L., Tyler-Smith, C. and Thangaraj, K., A common MYBPC3 (cardiac myosin binding protein C) variant associated with cardiomyopathies in South Asia, Nat Genet, 2009, 41: 187-191.

37 Myles, S., Davison, D., Barrett, J., Stoneking, M. and Timpson, N., Worldwide population differentiation at disease-associated SNPs, BMC Med Genomics, 2008, 1: 22.

38 Kathiresan, S., Melander, O., Anevski, D., Guiducci, C., Burtt, N. P., Roos, C., Hirschhorn, J. N., Berglund, G., Hedblad, B., Groop, L., Altshuler, D. M., Newton-Cheh, C. and Orho-Melander, M., Polymorphisms associated with cholesterol and risk of cardiovascular events, 2008, 358:1240-9.

39 Anderson, K.M., Odell, P. M., Wilson, P. W. and Kannel, W. B, Cardiovascular disease risk profiles, Am Heart J, 1991, 121:293-8.

40 Zheng, S.L., Sun, J., Wiklund, F., Smith, S., Stattin, P., Li, G., Adami, H.-O., Hsu, F.-C., Zhu, Y., Balter, K., Kader, A. K., Turner, A. R., Liu, W., Bleecker, E. R., Meyers, D. A., Duggan, D., Carpten, J. D., Chang, B.-L., Isaacs, W. B., Xu, J. and Gronberg, H., Cumulative association of five genetic variants with prostate cancer, N Engl J Med, 2008, 358:910-9.
41 Henrichsen, C. N., Vinckenbosch, N., Zollner, S., Chaignat, E., Pradervand, S., Schutz, F., Ruedi, M., Kaessmann, H. and Reymond, A., Segmental copy number variation shapes tissue transcriptomes, Nat Genet, 2009, 41: 424-429.
42 Thomas, D., Gene-environment-wide association studies: emerging approaches, Nat Rev Genet, 2010, 11: 259-272.
43 Perera, F., Tang, W. Y., Herbstman, J., Tang, D., Levin, L., Miller, R. and Ho, S. M., Relation of DNA methylation of 5'-CpG island of ACSL3 to transplacental exposure to airborne polycyclic aromatic hydrocarbons and childhood asthma, PLoS One, 2009, 4: e4488.
44 Baccarelli, A., Wright, R. O., Bollati, V., Tarantini, L., Litonjua, A. A., Suh, H. H., Zanobetti, A., Sparrow, D., Vokonas, P. S. and Schwartz, J., Rapid DNA methylation changes after exposure to traffic particles, Am J Respir Crit Care Med, 2009, 179: 572-578.
45 Smith, P. G. and Day, N. E., The design of case-control studies: the influence of confounding and interaction effects, Int J Epidemiol, 1984, 13: 356-365.
46 Elbers, C. C., van Eijk, K. R., Franke, L., Mulder, F., van der Schouw, Y. T., Wijmenga, C. and Onland-Moret, N. C., Using genome-wide pathway analysis to unravel the etiology of complex diseases, Genet Epidemiol, 2009, 33: 419-431.
47 Li, J., A novel strategy for detecting multiple loci in Genome-Wide Association Studies of complex diseases, Int J Bioinform Res Appl, 2008, 4: 150-163.
48 Jiang, X., Barmada, M. M. and Visweswaran, S., Identifying genetic interactions in genome-wide data using Bayesian networks, Genet Epidemiol, 2010, 34: 575-581.
49 Fijneman, R. J., de Vries, S. S., Jansen, R. C. and Demant, P., Complex interactions of new quantitative trait loci, Sluc1, Sluc2, Sluc3, and Sluc4, that influence the susceptibility to lung cancer in the mouse, Nat Genet, 1996, 14: 465-467.
50 Leamy, L. J., Routman, E. J. and Cheverud, J. M., An epistatic genetic basis for fluctuating asymmetry of mandible size in mice, Evolution, 2002, 56: 642-653.
51 Wandstrat, A. and Wakeland, E., The genetics of complex autoimmune diseases: non-MHC susceptibility genes, Nat Immunol, 2001, 2: 802-809.
52 Ziegler, A., Konig, I. R. and Thompson, J. R., Biostatistical aspects of genome-wide association studies, Biom J, 2008, 50: 8-28.
53 Raitakari, O. T., Juonala, M., Kahonen, M., Taittonen, L., Laitinen, T., Maki-Torkko, N., Jarvisalo, M. J., Uhari, M., Jokinen, E., Ronnemaa, T., Akerblom, H. K. and Viikari, J. S., Cardiovascular risk factors in childhood and carotid artery intima-media thickness in adulthood: the Cardiovascular Risk in Young Finns Study, Jama, 2003, 290: 2277-2283.
54 Mahoney, L. T., Burns, T. L., Stanford, W., Thompson, B. H., Witt, J. D., Rost, C. A. and Lauer, R. M., Coronary risk factors measured in childhood and young adult life are associated with coronary artery calcification in young adults: the Muscatine Study, J Am Coll Cardiol, 1996, 27: 277-284.
55 Oliveira, F. L., Patin, R. V. and Escrivao, M. A., Atherosclerosis prevention and treatment in children and adolescents, Expert Rev Cardiovasc Ther, 2010, 8: 513-528.
56 Bao, W., Threefoot, S. A., Srinivasan, S. R. and Berenson, G. S., Essential hypertension predicted by tracking of elevated blood pressure from childhood to adulthood: the Bogalusa Heart Study, Am J Hypertens, 1995, 8: 657-665.
57 Yu, W., Gwinn, M., Clyne, M., Yesupriya, A. and Khoury, M. J., A navigator for human genome epidemiology, Nat Genet, 2008, 40: 124-125.

Supplementary Table 1: Published Genome Wide Association Studies on cardiovascular diseases traits (http://hugenavigator.net/HuGENavigator/gWAHitStartPage.do).

CVD trait	Variant(s)	Near gene(s)/ clusters	Region	Trait(s)	Reference(s)	Initial/ replication sample size	P value	OR or beta-coefficient and [95% CI] or effect	Variant-Risk Allele
Cardiovascular events	rs11206510	PCSK9	1p32.3	MI (early onset)	Kathiresan et al, 2009 Nat Genet [1]	2,967 cases, 3,075 controls /9,746 cases, 9,746 controls	9.6x10-9	1.15 [1.10-1.21]	rs11206510-T
	rs4977574	CDKN2A-CDKN2B	9p21.3	MI (early onset)	Kathiresan et al, 2009 Nat Genet [1]	2,967 cases, 3,075 controls /9,746 cases, 9,746 controls	3x10-44	1.29 [1.25-1.34]	rs4977574-G
	rs1333049	CDKN2A-CDKN2B	9p21.3	CHD	Samani et al, 2007 N Engl J Med [2]	1,926 cases, 2,938 controls /875 cases, 1,644 controls	3x10-19	1.36 [1.27-1.46]	rs1333049-C
					WTCCC et al, 2007 Nature [3]	-	1.8x10-14	1.47 [1.27-1.70]	rs1333049-C
	rs599839	PSRC1	1p13.3	CHD	Samani et al, 2007 N Engl J Med [2]	1,926 cases, 2,938 controls /875 cases, 1,644 controls	4x10-9	1.29 [1.18-1.40]	rs599839-A
	rs9818870	MRAS	3q22.3	CHD	Erdmann et al, 2009, Nat Genet [4]	6,990 cases, 8,955 controls/12,417 cases, 12,411 controls	7 x 10-13	1.15 [1.11-1.19]	rs9818870-T
	4 SNPs (haplotype)	SLC22A3, LPAL2 LPA	6q25.3	CHD	Tregouet et al, 2009 Nat Genet [5]	1,926 cases, 2,938 controls/7,073 cases, 7,325 controls	4 x 10-15	1.82 [1.57-2.12]	-
Lipids profile				HDL-C, LDL-C TC, TG	Teslovich et al, 2010 Nature [6]	Up to 100,184 European descent individuals /9,705 South Asians, 15,046 East Asians, 8,061 African Americans, 7,063 Europeans	7x10-380	3.39 mg/dl	
	rs3764261	CETP	16q13	HDL-C	Willer et al, 2008 Nat Genet [7]	8,656 individuals /11,399 individuals	2x10-57	3.47 mg/dl	rs3764261-A
				LDL-C	Hiura et al, 2009 Circ J [8]	900 Japanese individuals /1,810 Japanese individuals	3x10-12	6.2 mg/dl	
				HDL-C	Ridker et al, 2009 Circ Cardiovasc Genet [9]	18,425 European ancestry women /NR	4x10-93	3.09 mg/dL	rs1800775-A
	rs1800775	CETP	16q13	HDL-C	Kathiresan et al, 2008 Nat Genet [10]	2,758 individuals /18,544 individuals	1x10-73	-0.18 [0.16-0.20]	rs1800775-C
				HDL-C	Saxena et al, 2007 Science [11]	up to 5,217 individuals /NR	3x10-13	-	rs1800775-C

				LDL-C	Kathiresan et al, 2008 Nat Genen [10]	2,758 individuals /18,544 individuals,	1x10-60	0.19 [0.15-0.23]	
				Lp-PLA2 activity and mass	Suchindran et al, 2010 PLoS Genet [12]	6,668 Caucasian individuals /NR	6x10-24	8 nmol/ml/min	
	rs4420638	APOC1	19q13.32	TG	Saxena et al, 2007 Science [11]	up to 5,217 individuals /NR	3x10-13 (LDL)	-	rs4420638-G
	rs693	APOB	TG	LDL-C	Kathiresan et al, 2008 Nat Genet [10]	2,758 individuals /18,544 individuals	1x10-21	0.12 [0.10-0.14]	rs693-A
				TC	Aulchenko et al ,2008 Nat Genet [13]	22,562 individuals /NR	9x10-23	-0.1	rs693-G
				TG	Saxena et al, 2007 Science [11]	up to 5,217 individuals /NR	7x10-7(LDL)	-	rs693-A
				LDL-C			5x10-13	-0.18	
	rs12272004	APOA1 APOA4 APOA5 APOC3	11q23.3		Aulchenko et al ,2008 Nat Genet [13]	22,562 individuals /NR			rs12272004-C
				TC			7x10-7	-0.11	
				TG			5x10-13	-0.18	
Blood glucose	rs560887	G6PC2	2q24.3	FGRT	Dupuis et al,2010 Nat Genet [14]	up to 46,186 European descent individuals /up to 76,558 European ancestry individuals	9x10-218 (FBG) 2x10-66 (HOMA-B)	-	rs560887-C
	rs7903146	TCF7L2	10q25.2	T2D	Saxena et al, 2007 Science [11]	1,464 cases, 1,467 controls /5,065 cases, 5,785 controls (also includes meta-analysis from DGI, FUSION and WTCCC)	2.3x10-31	1.38 [1.31-1.46]	rs7903146-T
					Sladeck, 2007 Nature [15]	661 cases, 614 controls /2,617 cases, 2,894 controls	2x10-34	1.65 [1.28, 2.02]	rs7903146-T
	rs10830963	MTNR1B	11q14.3	FGRT	Dupuis et al, 2010 Nat Genet [14]	up to 46,186 European descent individuals / up to 76,558 European ancestry individuals	6x10-175 (FPG)	-	rs10830963-G
	rs9272346	HLA-E	6p21.32	T1D	WTCCC,2007 Nature [3]	1,963 cases, 2,938 controls 2997 trios, 4,000 cases, 5,000 controls	5x10-134	5.49 [4.83-6.24]	rs9272346-G

	rs7901695	TCF7L2	10q25.2	T2D	Zeggini et al, 2007 Science [16]	1,924 cases, 2,938 controls /3,757 cases, 5,346 controls (also includes meta-analysis from DGI, FUSION and WTCCC)	1x10-48	1.37 [1.31-1.43]	rs7901695-C
	rs2681472	ATP2B1	12q21.33	HTN	Levy et al, 2009 Nat Genet [17]	29,136 individuals /34,433 individuals	2x10-11	0.15 [0.11-0.19]	rs2681472-A
				DBP			1x10-9	0.5 [0.34-0.66] mm Hg	
Blood pressure	rs3184504	SH2B3	12q24.12	DBP	Levy et al, 2009 Nat Genet [17]	29,136 individuals /34,433 individuals	3x10-14	0.48 [0.36-0.60] mm Hg	rs3184504-T
				SBP			5x10-9	0.58 [0.38-0.78] mm Hg	
	rs653178	ATXN2 SH2B3	12q24.12	DBP	Newton-Cheh et al, 2009 Nat Genet [18]	34,433 individuals / Up to 100,347 white individuals, up to 12,889 Indian Asian individuals	3x10-18	-0.46 [0.36-0.56] mm Hg	rs653178-T
	rs1378942	CSK	15q24.1	DBP	Newton-Cheh et al, 2009 Nat Genet [18]	34,433 individuals / Up to 100,347 white individuals, up to 12,889 Indian Asian individuals	1x10-23	0.43 [0.35-0.51] mm Hg	rs1378942-C
	rs16998073	FGF5	4q21.21	DBP	Newton-Cheh et al, 2009 Nat Genet [18]	34,433 individuals /Up to 100,347 white individuals, up to 12,889 Indian Asian individuals	1x10-21	0.5 [0.40-0.60] mm Hg	rs16998073-T
Hypertension in young	rs6729869 rs10495809 rs9308945 rs6711736	FAM98A	2p22.3	HTN	Yang et al, 2009 PLoS ONE [19]	175 Han Chinese cases, 175 Han Chinese controls /833 Han Chinese cases, 833 Han Chinese controls	3x10-10	-	4 SNPs (Haplotype)
	rs8050136	FTO	16q12.2	BMI, T2D, weight	Thorleifsson et al, 2008 Nat Gene [20]	80,969 individuals /11,036 individuals	1x10-47	8.04 [6.96-9.12] % SD	rs8050136-A
	rs9939609	FTO	16q12.2	BMI	Willer et al, 2008 Nat Genet [7]	32,387 individuals /59,092 individuals	4x10-51	0.33 [0.27-0.39] kg/m2	rs9939609-A
Obesity					Frayling et al, 2007 Science [21]	10,657 adults /19,424 adults, 10,172 children	2x10-20	0.36 kg/m2	rs9939609-A
	rs1424233	MAF	16q23.2	obesity	Meyre et al, 2009 Nat Genet [22]	695 obese adults, 685 obese children, 731 lean adults, 685 lean children / 1,171 obese adults, 896 obese children, 1,114 lean adults, 1,297 lean children, 4,417 adults, 5,291 children	4x10-13	1.12 [1.00-1.24]	rs1424233-A
	rs1421085	FTO	16q12.2	obesity	Meyre et al, 2009 Nat Genet [22]	695 obese adults, 685 obese children, 731 lean adults, 685 lean children / 1,171 obese adults, 896 obese children, 1,114 lean adults, 1,297 lean children, 4,417 adults, 5,291 children	1x10-28	1.39 [1.27-1.51]	rs1421085-C

	rs1558902	FTO	16q12.2	WC	Heard-Costa et al, 2009 PLoS Genet [23]	31,373 Caucasian individuals /38,641 Caucasian individuals	5x10-19	-	rs1558902-C
				Extreme obesity	Scherag et al, 2010 PLoS Genet [24]	1,138 French and German extremely obese children, 1,120 French and German normal or underweight children /1,892 European ancestry obese or overweight children and adolescents, 3,763 European ancestry normal or underweight controls, 715 families with obese offspring	7x10-13	1.37 [1.26-1.50]	rs1558902-A
				BMI	Willer et al, 2008 Nat Genet [7]	32,387 individuals /59,092 individuals	5x10-18	0.2 [0.12-0.28] kg/m2	rs17782313-C
	rs17782313	MC4R	18q21.32	Obesity	Meyre et al, 2009 Nat Genet [22]	695 obese adults, 685 obese children, 731 lean adults, 685 lean children /1,171 obese adults, 896 obese children, 1,114 lean adults, 1,297 lean children, 4,417 adults, 5,291 children	5x10-15	1.22 [1.05-1.40]	rs17782313-C
				BMI	Loos et al, 2008 Nat Genet [25]	16,876 individuals /60,352 individuals	3x10-15	0.05 [0.04-0.06]	rs17782313-C
	rs651007	ABO	9q34.2	SELE	Qi et al, 2010 Hum mol genetic [26]	1,005 European ancestry women /518 European ancestry women	2x10-82	-17.23	rs651007-T
	rs6136	SELP	1q24.2	SELP	Barbalic, 2010 Human mol genetic [27]	Up to 9,813 European descent individuals /NR	4x10-61	22.6 [19.86-25.34]	rs6136-T
	rs1799969	ICAM1	19p13.2	ICAM1	Pare et al, 2008 plos genetic [28]	4,570 women /2,008 women	4x10-47	-28.19 umol/L	rs1799969-G
Circulatory inflammatory markers	rs7553007	CRP	1q23.2	CRP	Elliott et al, 2009 JAMA [29]	17,967 European and Indian Asian men and women /13,615 European and Indian Asian men and women	8x10-44	-20.7 [18.9-23.4]	rs7553007-A
				CRP	Elliott et al, 2009 JAMA [29]	17,967 European and Indian Asian men and women /13,615 European and Indian Asian men and women	5x10-27	-21.8 [18.1-25.3]	rs4420638-G
	rs4420638	APOC1	19q13.32	LDL-C	Kathiresan et al,2008 Nat Genet [10]	2,758 individuals /18,544 individuals	1x10-60	0.19 [0.15-0.23]	rs4420638-G
				Lp-PLA2 activity and mass	Suchindran et al,2010 PLoS Genet [12]	6,668 Caucasian individuals /NR	6x10-24	8 nmol/ml/min	rs4420638-G
	rs1183910	HNF1A	12q24.31	CRP	Elliott et al,2009 JAMA [29]	17,967 European and Indian Asian men and women /13,615 European and Indian Asian men and women	1x10-30	-13.8 [10.9-16.6]	rs1183910-T

MI : myocardial infarction, CHD : coronary heart disease, HDL-C: high density lipoprotein cholesterol, LDL-C: low density lipoprotein cholesterol, TC: total cholesterol, TG: triglycerides, NR: not replicated, Lp-PLA2 : lipoprotein-associated phospholipase A2, FGRT :fasting glucose-related traits, FBG: fasting blood glucose, HOMA-B: homeostasis model for assessment beta cell function, T2D: type 2 diabetes, T1D: type 1 diabetes, HTN: hypertension, DBP: diastolic blood presure, SBP: systolic blood pressure, BMI: body mass index, WC:waist circumference, SELE: E selectin, SELP:P selectin, ICAM1: inter cellular adhesion molecule 1, CRP: C-reactive protein.

1) Kathiresan S et al. Common variants at 30 loci contribute to polygenic dyslipidemia. Nat Genet. 2009 Jan;41(1):56-65. Epub 2008 Dec 7.
2) Samani NJ et al.Genomewide association analysis of coronary artery disease. N Engl J Med. 2007 Aug 2;357(5):443-53. Epub 2007 Jul 18.
3) Wellcome Trust Case Control Consortium.Genome-wide association study of 14,000 cases of seven common diseases and 3,000 shared controls. Nature. 2007 Jun 7;447(7145):661-78.
4) Erdmann J et al. New susceptibility locus for coronary artery disease on chromosome 3q22.3. Nat Genet. 2009 Mar;41(3):280-2. Epub 2009 Feb 8.
5) Tregouet DA et al. Genome-wide haplotype association study identifies the SLC22A3-LPAL2-LPA gene cluster as a risk locus for coronary artery disease. Nat Genet. 2009 Mar;41(3):283-5. Epub 2009 Feb 8.
6) Teslovich TM et al.Biological, clinical and population relevance of 95 loci for blood lipids. Nature. 2010 Aug 5;466(7307):707-13.
7) Willer CJ et al. Newly identified loci that influence lipid concentrations and risk of coronary artery disease. Nat Genet. 2008 Feb;40(2):161-9. Epub 2008 Jan 13.
8) Hiura Y et al. Identification of genetic markers associated with high-density lipoprotein-cholesterol by genome-wide screening in a Japanese population: the Suita study. Circ J. 2009 Jun;73(6):1119-26. Epub 2009 Apr 10
9) Ridker PM et al. Polymorphism in the CETP gene region, HDL cholesterol, and risk of future myocardial infarction: Genomewide analysis among 18 245 initially healthy women from the Women's Genome Health Study. Circ Cardiovasc Genet. 2009 Feb;2(1):26-33. Epub 2009 Jan 23.
10) Kathiresan S et al. Six new loci associated with blood low-density lipoprotein cholesterol, high-density lipoprotein cholesterol or triglycerides in humans. Nat Genet. 2008 Feb;40(2):189-97. Epub 2008 Jan 13.
11) Saxena R et al. Genome-wide association analysis identifies loci for type 2 diabetes and triglyceride levels. Science. 2007 Jun 1;316(5829):1331-6. Epub 2007 Apr 26
12) Suchindran S et al. Genome-wide association study of Lp-PLA(2) activity and mass in the Framingham Heart Study. PLoS Genet. 2010 Apr 29;6(4):e1000928.
13) Aulchenko YS et al. Loci influencing lipid levels and coronary heart disease risk in 16 European population cohorts. Nat Genet. 2009 Jan;41(1):47-55. Epub 2008 Dec 7.
14) Dupuis J et al. New genetic loci implicated in fasting glucose homeostasis and their impact on type 2 diabetes risk. Nat Genet. 2010 Feb;42(2):105-16. Epub 2010 Jan 17.
15) Sladek R,A genome-wide association study identifies novel risk loci for type 2 diabetes. Nature. 2007 Feb 22;445(7130):881-5. Epub 2007 Feb 11.
16) Zeggini E et al. Replication of genome-wide association signals in UK samples reveals risk loci for type 2 diabetes. Science. 2007 Jun 1;316(5829):1336-41. Epub 2007 Apr 26.
17) Levy D et al. Genome-wide association study of blood pressure and hypertension. Nat Genet. 2009 May 10. [Epub ahead of print]
18) Newton-Cheh C et al. Genome-wide association study identifies eight loci associated with blood pressure. Nat Genet. 2009 May 10. [Epub ahead of print]
19) Yang HC et al.Genome-wide association study of young-onset hypertension in the Han Chinese population of Taiwan. PLoS One. 2009;4(5):e5459. Epub 2009 May 7.
20) Thorleifsson G et al. Genome-wide association yields new sequence variants at seven loci that associate with measures of obesity. Nat Genet. 2009 Jan;41(1):18-24. Epub 2008 Dec 14.
21) Frayling TM, A common variant in the FTO gene is associated with body mass index and predisposes to childhood and adult obesity. Science. 2007 May 11;316(5826):889-94. Epub 2007 Apr 12.
22) Meyre D et al. Genome-wide association study for early-onset and morbid adult obesity identifies three new risk loci in European populations. Nat Genet. 2009 Feb;41(2):157-9. Epub 2009 Jan 18.
23) Heard-Costa NL et al. NRXN3 is a novel locus for waist circumference: a genome-wide association study from the CHARGE Consortium. PLoS Genet. 2009 Jun;5(6):e1000539. Epub 2009 Jun 26.

24) Scherag A et al.Two new Loci for body-weight regulation identified in a joint analysis of genome-wide association studies for early-onset extreme obesity in French and german study groups. PLoS Genet. 2010 Apr 22; 6(4):e1000916.
25) Loos RJ.Common variants near MC4R are associated with fat mass, weight and risk of obesity. Nat Genet. 2008 Jun;40(6):768-75. Epub 2008 May 4.
26) Qi L et al Genetic variants in ABO blood group region, plasma soluble E-selectin levels and risk of type 2 diabetes. Hum Mol Genet. 2010 May 1;19(9):1856-62. Epub 2010 Feb 10.
27) Barbalic M. Large-scale genomic studies reveal central role of ABO in sP-selectin and sICAM-1 levels. Hum Mol Genet. 2010 May 1;19(9):1863-72. Epub 2010 Feb 18.
28) Paré G et al. Novel association of ABO histo-blood group antigen with soluble ICAM-1: results of a genome-wide association study of 6,578 women. PLoS Genet. 2008 Jul 4;4(7):e1000118.
29) Elliott P et al. Genetic Loci associated with C-reactive protein levels and risk of coronary heart disease. JAMA. 2009 Jul 1;302(1):37-48

PUBLICATION N°2

Bernard Herbeth, Anastasia Samara, **Ndeye Coumba Ndiaye**, Jean-Brice Marteau, Hind Berrahmoune, Gérard Siest, Sophie Visvikis-Siest. Metabolic syndrome-related composite factors over 5 years in the STANISLAS Family Study: Genetic heritability and common environmental influences. Clinica Chimica Acta. 2010; 411: 833–9.

Dans cette étude nous avons estimé l'héritabilité génétique et l'environnement partagé de différents traits liés au syndrome métabolique dans de jeunes familles Françaises issues de la cohorte STANISLAS par une approche transversale puis longitudinale.

Dix neuf traits assemblés en cinq clusters ont été investigués par une analyse factorielle et nous avons observé de fortes héritabilités génétiques (20 à 44%) pour les lipides, les lipoprotéines, l'acide urique, le glucose et les clusters associés ; et des héritabilités génétiques modérées (inférieures à 20%) pour les triglycérides, l'adiposité, la pression artérielle, les enzymes hépatiques et les marqueurs de l'inflammation en transversal (analyse lors du premier examen de santé). L'environnement partagé avait un effet sur toutes les variables investiguées.

En longitudinal à court terme (5 ans), les variances polygéniques étaient au contraire toutes faibles et non significatives et l'environnement partagé très élevé.

Ainsi, nous avons observé des héritabilités génétiques plus faibles que celles généralement constatées pour les traits associés au syndrome métabolique tandis que l'effet de l'environnement partagé était plus élevé.

Contribution :

Lors de cette étude nous avons participé à la recherche bibliographique et à l'interprétation des résultats. Nous avons également fait partie de l'équipe rédactionnelle.

Clinica Chimica Acta 411 (2010) 833–839

Contents lists available at ScienceDirect

Clinica Chimica Acta

journal homepage: www.elsevier.com/locate/clinchim

Metabolic syndrome-related composite factors over 5 years in the STANISLAS Family Study: Genetic heritability and common environmental influences

Bernard Herbeth *, Anastasia Samara, Coumba Ndiaye, Jean-Brice Marteau, Hind Berrahmoune, Gérard Siest, Sophie Visvikis-Siest

EA 4373, "Cardiovascular Genetics" research team, Université Henri Poincaré, Nancy Université, Faculté de Pharmacie, 30 rue Lionnois, F-54000 Nancy, France

ARTICLE INFO

Article history:
Received 7 January 2010
Received in revised form 23 February 2010
Accepted 23 February 2010
Available online 25 February 2010

Keywords:
Metabolic syndrome
Factor analysis
Heritability

ABSTRACT

Background: We estimated genetic heritability and common environmental influences for various traits related to metabolic syndrome in young families from France.
Methods: At entrance and after 5 years, nineteen traits related to metabolic syndrome were measured in a sample of families drawn from the STANISLAS study. In addition, 5 aggregates of these traits were identified using factor analysis.
Results: At entrance, genetic heritability was high (20 to 44%) for plasma lipids and lipoproteins, uric acid, fasting glucose, and the related clusters "risk lipids" and "protective lipids". Intermediate or low genetic heritability (less than 20%) was shown for triglycerides, adiposity indices, blood pressure, hepatic enzyme activity, inflammatory makers and the related clusters: "liver enzymes", "adiposity/blood pressure" and "inflammation". Moreover, common environmental influences were significant for all the parameters. With regard to 5-year changes, polygenic variance was low and not statistically significant for any of the individual variables or clusters whereas shared environment influence was significant.
Conclusions: In these young families, genetic heritability of metabolic syndrome-related traits was generally lower than previously reported while the common environmental influences were greater. In addition, only shared environment contributed to short-term changes of these traits.
© 2010 Elsevier B.V. All rights reserved.

1. Introduction

The metabolic syndrome is a complex combination of traditional cardiovascular risk factors(central obesity, hyperglycemia, insulin resistance, dyslipidemia and hypertension) [1], and of some novel additional factors that include inflammation markers (CRP, haptoglobin, orosomucoid), white blood cells count [2–4], serum uric acid [5,6] and liver enzymes [gamma-glutamyl transferase (GGT), alanine aminotransferase (ALAT), and aspartate aminotransferase (ASAT)] that are important markers of non-alcoholic fatty liver disease [7]. Several approaches exist for studying simultaneously numerous factors of the metabolic syndrome. One is the use of a bivariate genetic/environmental analysis for pairs of metabolic syndrome-related traits in different family members [8]. Another one is the use of a dimension reduction technique, such as factor analysis, that allows the study of aggregates of phenotypes (clusters) instead of the original phenotypes in variance component analysis. This method has been used in family studies for common metabolic syndrome-related factors [9–12]. However, although some studies have included novel factors of the metabolic syndrome for studying their heritability (CRP, WBC, uric acid) [8,9,12–14], no data exists about clusters of these factors. Likewise, no study has examined whether genetic or environmental susceptibility may underlie clusters of short-term changes in these metabolic syndrome-related factors.

Therefore, we aimed to estimate additive genetic heritability and common environmental effects for clusters and individual quantitative traits of both common and novel metabolic syndrome risk factors at baseline and over 5 years. Two subsets of data of the STANISLAS Family Cohort were used: a sample of 667 families ($n=2471$ individuals) for the cross-sectional study and a sample of 353 families ($n=1208$ individuals) for the longitudinal study.

2. Methods

2.1. Study subjects

This work is based on the STANISLAS Family Study, a 10-year longitudinal study conducted since 1994 on 1006 families selected at the Center for Preventive Medicine of Vandoeuvre-lès-Nancy (east of France) [15,16]. In this work, we present 2 subsets of data, firstly on metabolic syndrome-related variables measured during the initial

Abbreviations: ALAT, alanine aminotransferase; apo, apolipoprotein; ASAT, aspartate aminotransferase; BMI, body mass index; BWI, body weight index; DBP, diastolic blood pressure; GGT, gamma-glutamyl transferase; HDL, high-density lipoprotein; hs-CRP, high sensitivity C-reactive protein; SBP, systolic blood pressure; WC, waist circumference; WHR, waist-to-hip ratio.

* Corresponding author. Tel.: +33 3 83 68 21 72; fax: +33 3 83 32 13 22.
E-mail address: Bernard.herbeth@nancy.insem.fr (B. Herbeth).

0009-8981/$ – see front matter © 2010 Elsevier B.V. All rights reserved.
doi:10.1016/j.cca.2010.02.070

examination in 1994–95 and secondly on 5-year changes of metabolic syndrome-related variables. Due to the design of the STANISLAS Family Study, subjects were of French origin and were free from acute or chronic diseases such as stroke, myocardial infarction or cancer. Children at entrance who were younger than 11 years were not included in this study.

A total of 667 families with two biological parents and at least one child (nuclear families) were eligible for the study based on the baseline check-up in 1994–95 (1334 parents aged 30–64 years at entrance and 1137 children aged 12–25 years). The response rate of individuals for the second check-up was 73%. Consequently, 353 families attending a check-up both in 1994–95 and in 1999–2000, composed of two parents and at least one child, were eligible for the longitudinal study (706 parents and 502 children).

Each subject gave written informed consent for participating in this study, which was approved by the local ethics committee (Comité de Protection des Personnes dans la Recherche Biomédicale de Lorraine, France).

2.2. Blood samples and data collection

Blood samples were collected after an overnight fast between 8.00 and 9.00 a.m. or between 11.00 and 12.00 a.m. Data were collected by using relevant questionnaires including, information about lifestyle such as tobacco, alcohol and drug consumption and personal medical history. Physical examinations and functional tests were performed and basic blood constituents were measured as described previously [15]. Serum concentrations of fasting glucose, total cholesterol, triglycerides, uric acid and activities of ASAT, ALAT, and GGT were measured with commercially available kits on an AU5021 apparatus (all from Merck, Darmstadt, Germany) on fresh aliquots, and within 2 h. Serum apolipoproteins AI and B, hs-CRP, haptoglobin and orosomucoid were measured by immunonephelometry on a Behring Nephelometer Analyser (BN II, Siemens Healthcare Diagnostics, Deerfield IL) within 2 h after sampling. Apolipoprotein E was measured by turbidimetry and HDL-cholesterol by the precipitation by phosphotungstate, both on a Cobas-Mira analyzer (Roche).

Metabolic syndrome in adults was defined according to the International Diabetes Federation (IDF) criteria [17]: waist circumference (WC) $\geq$94 cm in men or $\geq$80 cm in women plus any two of the four following criteria (1) triglycerides $\geq$1.7 mmol/l or drug treatment for elevated triglycerides, (2) HDL-C <1.03 mmol/l in men or <1.3 mmol/l in women or drug treatment for reduced HDL-C, (3) SBP $\geq$130 mm Hg or DBP $\geq$85 mm Hg or anti-hypertensive drug treatment, (4) fasting glucose $\geq$5.6 mmol/l or drug treatment for elevated glucose. For adolescents, specific IDF criteria were used: the age- and gender-specific cut-points for each metabolic syndrome component proposed in the literature [18].

For the whole sample, body weight index (BWI) that is as an appropriate measure of adiposity in children, independently of growth phenomenon and sexual maturation, was used in analysis (BWI − [weight/weight reference values for sex and age]/100). Weight reference values are derived from Rolland-Cachera et al. [19].

2.3. Statistical analysis

Statistical analyses were performed by using the SAS software package version 9.1 (SAS Institute, Inc., Cary, NC). Before statistical analyses, serum concentrations of the above mentioned variables were adjusted for the effect of time of blood sampling (between 8.00 and 9.00 a.m. or between 11.00 and 12.00 a.m.) and for between-month variability to remove the effect of daily, monthly (seasonal) and analytical variations. Briefly, for each variable, levels were regressed on both time of blood sampling (at the beginning of morning or at the end) and on monthly truncated means obtained on selected healthy males aged 20–40 years attending the health check-up concurrently with the sample population of this study; the variable used being the sum: residual + crude mean of the overall sample.

Since the distributions of triglycerides and hs-CRP concentrations, and ASAT, ALAT and GGT activities exhibited a long-tailed positive skewness and kurtosis, $\log_{10}$ transformation was used. The Gaussian distribution of transformed variables was verified by using normal probability plots.

Factor analysis was undertaken to avoid the problem of multiple comparisons and redundancy of information. Prior analysis, the nineteen initial metabolic syndrome-related factors measured at entrance (data from cross-sectional study): BWI, WC, systolic and diastolic blood pressures; serum concentrations of triglycerides, total cholesterol, HDL-cholesterol, apo A1, apo B, apo E, fasting glucose, uric acid, hs-CRP, orosomucoid, and haptoglobin; activities of ALAT, ASAT and GGT; and white blood cells count were adjusted for age and specific drug use separately for fathers, mothers, sons and daughters. Likewise, the 5-year changes in variables (data from longitudinal study) were adjusted for age, specific drug use and values at entrance, separately for fathers, mothers, sons and daughters. Factor analysis was followed by orthogonal (varimax) rotation to assist in interpretation of the factors and to ensure that the factors were uncorrelated. We determined the number of factors to retain by using both the proportion of common variance to be accounted for by the retained factors ($\approx$100%) and the Scree test. The Scree plot is a plot of the eigenvalues of derived factors. To simplify interpretations, only variables with factor loadings having absolute values greater than 0.29 were shown. The factor score for each pattern was calculated by summing levels of metabolic syndrome-related variables weighted by their factor loadings. Factor scores were standardized to have a mean of 0 and a standard deviation of 1. The scores reflect how closely participant's metabolic syndrome-related variables resemble each identified pattern, with higher scores representing closer resemblance. At baseline and for 5-year changes, each individual received a factor score for each identified cluster.

Variance component analysis was applied to assess the relative contributions of genetic and environment to metabolic syndrome-related factors and to clusters (at entrance and for 5-year changes). Under a pure polygenic model, a phenotype (P) is a function of genetic (G) and environment (E) effects (i.e., $P = G + E$), usually expressed in terms of variance components ($\sigma_P^2 = \sigma_G^2 + \sigma_E^2$). The genetic heritability (usually named broad-sense genetic) is defined as the proportion of the total phenotype variance that is due to genetic effects ($h_G^2 = \sigma_G^2/\sigma_P^2$), mainly due to a large number of genes, each with a small, linear, and additive effect.

The environmental component can be decomposed into common familial (due to the effect of sharing the same household at the time of measurement, C) and random nonfamilial (due to individual-specific factors, R) environmental effects (i.e., $\sigma_P^2 = \sigma_G^2 + \sigma_C^2 + \sigma_R^2$). Thus, analogous to genetic heritability, the cultural heritability or common environmental variance component may be defined as ($c^2 = \sigma_C^2/\sigma_P^2$), which is assumed to be due to a large number of linear and additive familial environmental effects [20].

The analysis was conducted by using a multivariate normal model for pedigree analysis as described by Lange et al. [21,22] with the software FISHER, which also performed goodness-of-fit test of the underlying multi-normal distribution. These three components were assumed to be normally distributed with mean equal to 0 and variance equal to σ_G^2, σ_C^2 and σ_R^2, respectively. For all tests, statistical significance was taken at $p \leq 0.05$.

3. Results

Baseline characteristics of the 667 families (2471 individuals) included in the cross-sectional study are presented in Table 1. Five-year changes in characteristics of the 353 families (1208 individuals) included in the longitudinal study are shown in Table 2.

Table 1
Baseline characteristics of the 2471 individuals included in the cross-sectional study.
Data from the STANISLAS Family Study: entrance screening of 1994–1995.

Variables	Adults		Children	
	Father ($n=667$)	Mother ($n=667$)	Son ($n=575$)	Daughter ($n=562$)
Age (years)	42.5 (4.6)	40.4 (4.5)	15.5 (2.9)	15.9 (3.3)
Weight (kg)	77.1 (11.1)	61.4 (10.2)	57.1 (13.3)	53.7 (9.7)
Body mass index (kg/m^2)	25.5 (3.2)	23.7 (3.9)	20.1 (3.0)	20.5 (3.1)
Body weight index (%)	102.1 (12.8)	103.0 (17.5)	103.3 (13.1)	104.5 (15.1)
Waist circumference (cm)	88.9 (9.0)	75.1 (9.0)	71.1 (7.4)	66.4 (6.6)
Waist-to-hip ratio	0.91 (0.06)	0.77 (0.06)	0.81 (0.04)[c]	0.73 (0.05)
Alcohol consumption (g/day)	23.9 (25.7)	4.7 (8.9)	1.1 (5.7)	0.3 (1.5)
Tobacco consumption (cig/day)	5.4 (9.4)	2.6 (6.4)	0.8 (3.3)	1.2 (4.2)
Anti-hypertensive use (%)	3.3	1.7	0.0	0.0
Anti-diabetic use (%)	0.4	0.0	0.0	0.0
Lipid lowering agent use (%)	7.0	1.7	0.5	0.5
Anti-inflammatory use (%)	3.8	2.2	2.6	2.8
Systolic blood pressure (mm Hg)	128.3 (12.5)	120.3 (13.6)	121.2 (11.7)	115.2 (9.6)
Diastolic blood pressure (mm Hg)	78.6 (9.7)	72.7 (10.3)	62.1 (11.1)	61.9 (10.2)
Triglycerides (mmol/l)	1.21 (0.71–2.06)	0.84 (0.55–1.29)	0.74 (0.48–1.14)	0.82 (0.54–1.23)
Total cholesterol (mmol/l)	5.99 (1.10)	5.52 (0.91)	4.49 (0.82)	4.85 (0.84)
HDL-cholesterol (mmol/l)	1.39 (0.35)	1.73 (0.42)	1.43 (0.36)	1.55 (0.35)[c]
Apolipoprotein B (g/l)	1.15 (0.26)	0.98 (0.22)	0.79 (0.19)	0.85 (0.19)
Apolipoprotein A1 (g/l)	1.55 (0.23)	1.71 (0.26)	1.44 (0.21)	1.53 (0.24)
Apolipoprotein E (mg/l)	41.3 (14.4)	37.0 (10.6)	36.2 (9.8)	39.1 (10.7)
Fasting glucose (mmol/l)	5.19 (0.56)	4.88 (0.40)	4.92 (0.34)	4.79 (0.39)
Uric acid (μmol/l)	319.5 (67.7)	230.1 (48.6)	293.0 (61.2)	242.6 (45.7)
Hs-CRP (mg/l)	1.00 (0.35–2.87)	0.97 (0.31–3.02)	0.45 (0.15–1.34)	0.47 (0.16–1.38)
Haptoglobin (g/l)	1.09 (0.49)	1.07 (0.42)	0.72 (0.41)	0.89 (0.41)
Orosomucoid (g/l)	0.79 (0.19)	0.71 (0.18)	0.71 (0.21)	0.69 (0.20)
ASAT (U/l)	23.4 (16.6–33.0)	17.3 (13.0–23.0)	23.3 (16.7–32.8)	18.7 (14.3–24.4)
ALAT (U/l)	30.2 (18.7–48.8)	17.6 (10.9–28.1)	19.2 (12.6–29.3)	15.5 (10.6–22.7)
GGT (U/l)	30.4 (16.6–55.7)	16.2 (9.7–27.1)	13.9 (10.1–19.2)	11.3 (8.4–15.2)
White blood cells (10^9/l)	6.92 (1.86)	7.05 (1.84)	6.58 (1.66)	7.12 (1.79)
Metabolic syndrome (%)	14.8	6.4	0.5	1.1

Values are arithmetic mean of crude data (SD), geometric mean (range of 1SD) or proportion (%).
Metabolic syndrome was defined according to IDF criteria (see Methods).
ALAT alanine aminotransferase, ASAT aspartate aminotransferase, GGT gamma-glutamyl transferase, HDL high-density lipoprotein, hs-CRP high-sensitivity C-reactive protein.

Table 2
Five-year changes in characteristics of the 1208 individuals included in the longitudinal study.
Data from the STANISLAS Family Study: check-up of 1994–95 and 1999–2000).

Variables	Adults		Children	
	Father ($n=353$)	Mother ($n=353$)	Son ($n=248$)	Daughter ($n=254$)
Δ Weight (kg)	2.41 (4.07)	2.10 (4.44)	13.99 (10.11)	6.22 (7.52)
Δ Body mass index (kg/m^2)	0.85 (1.33)	0.86 (1.70)	2.15 (1.80)	1.53 (2.35)
Δ Body weight index	2.22 (5.23)	0.98 (7.26)	−0.72 (8.03)	0.21 (10.83)
Δ Waist circumference (cm)	2.40 (4.70)	1.75 (4.65)	6.58 (5.19)	2.87 (5.58)
Δ Waist-to-hip ratio	0.02 (0.04)	0.02 (0.04)	0.01 (0.05)	−0.01 (0.05)
Δ Alcohol consumption (g/day)	−0.70 (16.51)	0.89 (7.63)	3.70 (8.71)	1.06 (3.06)
Δ Cigarette consumption (cig/day)	−0.52 (5.63)	−0.03 (3.35)	3.28 (6.41)	2.39 (5.49)
Δ Systolic blood pressure (mm Hg)	−0.07 (12.72)	0.27 (13.01)	3.62 (12.05)	0.70 (9.98)
Δ Diastolic blood pressure (mm Hg)	−1.17 (11.25)	−0.63 (11.70)	4.86 (13.03)	4.37 (12.46)
Δ Triglycerides (mmol/l)	0.11 (0.93)	0.03 (0.41)	0.08 (0.41)	0.01 (0.48)
Δ Total cholesterol (mmol/l)	0.04 (0.89)	0.08 (0.76)	−0.11 (0.68)	0.10 (0.78)
Δ HDL-cholesterol (mmol/l)	0.02 (0.28)	0.05 (0.33)	−0.15 (0.31)	0.09 (0.34)
Δ Apolipoprotein B (g/l)	−0.03 (0.21)	−0.03 (0.17)	−0.03 (0.14)	0.00 (0.18)
Δ Apolipoprotein A1 (g/l)	−0.01 (0.18)	−0.01 (0.23)	−0.08 (0.21)	0.07 (0.27)
Δ Apolipoprotein E (mg/l)	1.78 (12.08)	2.59 (8.76)	−2.62 (7.40)	−3.16 (9.69)
Δ Fasting glucose (mmol/l)	0.22 (0.86)	0.11 (0.53)	−0.04 (0.92)	−0.06 (0.51)
Δ Uric acid (μmol/l)	3.64 (55.06)	−3.21 (40.43)	30.84 (63.54)	−16.50 (43.92)
Δ Hs-CRP (mg/l)	0.01 (7.39)	0.15 (5.53)	0.83 (5.87)	0.66 (2.66)
Δ Haptoglobin (g/l)	0.05 (0.35)	0.07 (0.34)	0.29 (0.40)	0.18 (0.39)
Δ Orosomucoid (g/l)	0.00 (0.16)	0.02 (0.17)	0.04 (0.23)	−0.01 (0.24)
Δ ASAT (U/l)	0.61 (13.88)	1.24 (7.99)	−3.55 (12.09)	−1.20 (6.52)
Δ ALAT (U/l)	0.22 (19.67)	0.59 (15.96)	−0.47 (18.55)	0.25 (10.61)
Δ GGT (U/l)	4.75 (22.94)	2.69 (19.12)	2.96 (6.13)	2.14 (5.61)
Δ White blood cells (10^9/l)	0.01 (1.53)	0.10 (1.67)	0.24 (1.81)	0.25 (1.96)
Metabolic syndrome				
Inclusion	14.2%	7.1%	0.0%	1.2%
5-year later	20.7%	7.9%	0.8%	0.4%

Values are arithmetic mean of crude data (SD) or proportion (%).
Metabolic syndrome was defined according to IDF criteria (see Methods).
Δ, change over 5 years; ALAT, alanine aminotransferase; ASAT, aspartate aminotransferase; GGT, gamma-glutamyl transferase; HDL, high-density lipids; hs-CRP, high-sensitivity C-reactive protein.

Parents who did not attend the follow-up examination after five years were significantly younger than responders. In couples who did not respond, fathers had significantly higher mean values for waist-to-hip ratio (WHR), cigarette consumption, fasting glucose, triglycerides, uric acid, haptoglobin, orosomucoid, and activities of ALAT and GGT; mothers had significantly higher weight and BMI. Children of both sexes not attending the second examination were older than those who responded, and consequently had significantly higher weight, BMI, waist circumference (WC), alcohol intake and cigarette consumption. In addition, sons had higher SBP and levels of uric acid and haptoglobin and lower level of apolipoprotein A1 and high-density lipids (HDL)-cholesterol and daughter had higher DBP and GGT activity.

3.1. Factor analysis

At baseline, factor analysis identified five major factors ("risk lipids", "adiposity/blood pressure", "protective lipids", "liver enzymes" and "inflammation") explaining 39.9%, 22.6%, 15.6%, 12.5%, and 9.4% of variance, respectively (Table 3). The first five largest positive eigenvalues accounted for 100.0% of the common variance. The Scree plot displayed a sharp bend at the sixth eigenvalue, reinforcing the preceding conclusion. The "risk lipids" factor included total cholesterol, apo B, triglycerides and apo E; the "adiposity/blood pressure" factor included WC, BWI, uric acid, and diastolic and systolic blood pressures; the "protective lipids" factor included HDL-cholesterol, and apo A1; the "liver enzymes" factor included ALAT, ASAT and GGT activities and "inflammation" factor hs-CRP, haptoglobin, orosomucoid, and WBC. Fasting glucose did not attain threshold for factor loading (>0.29).

With regard to the 5-year changes in metabolic syndrome-related variables, the factor analysis (Table 3) identified the same precedent five major clusters: "risk lipids", "body mass/blood pressure", "liver enzymes", "inflammation", and "protective lipids" accounting for 36.6%, 21.8%, 17.5%, 13.0% and 10.5% of the total variance in the data set, respectively (99.4% of the common variance). The clusters at entrance and for 5-year changes included the same variables, except

Table 3
Results of factor analysis with metabolic syndrome-related variables and factor loadings.[a]

Factor-loading patterns	Risk lipids	Adiposity/blood pressure	Protective lipids	Liver enzymes	Inflammation
A. Cross-sectional study: 2471 individuals					
Total cholesterol	0.917	–[b]	–	–	–
Apolipoprotein B	0.895	–	–	–	–
Triglycerides	0.444	–	–	–	–
Apolipoprotein E	0.402	–	–	–	–
Waist circumference	–	0.786	–	–	–
Body weight index	–	0.774	–	–	–
Systolic blood pressure	–	0.435	–	–	–
Diastolic blood pressure	–	0.404	–	–	–
Uric acid	–	0.297	–	–	–
Fasting glucose	–	–	–	–	–
HDL-cholesterol	–	–	0.888	–	–
Apolipoprotein A1	–	–	0.813	–	–
ALAT	–	–	–	0.802	–
ASAT	–	–	–	0.742	–
GGT	–	–	–	0.508	–
Haptoglobin	–	–	–	–	0.641
Hs-CRP	–	–	–	–	0.618
Orosomucoid	–	–	–	–	0.609
White blood cells	–	–	–	–	0.392
% of explained variance	39.9	22.6	15.6	12.5	9.4
Factor-loading patterns	Risk lipids	Adiposity/blood pressure	Liver enzymes	Inflammation	Protective lipids
B. Longitudinal study: 1208 individuals					
Δ Total cholesterol	0.814	–	–	–	0.308
Δ Apolipoprotein B	0.787	–	–	–	–
Δ Apolipoprotein E	0.612	–	–	–	–
Δ Triglycerides	0.576	–	–	–	–
Δ Body weight index	–	0.800	–	–	–
Δ Waist circumference	–	0.789	–	–	–
Δ Systolic blood pressure	–	0.381	–	–	–
Δ Diastolic blood pressure	–	0.344	–	–	–
Δ Fasting Glucose	–	–	–	–	–
Δ Uric acid	–	–	–	–	–
Δ ALAT	–	–	0.841	–	–
Δ ASAT	–	–	0.783	–	–
Δ GGT	–	–	0.411	–	–
Δ Orosomucoid	–	–	–	0.726	–
Δ Haptoglobine	–	–	–	0.722	–
Δ Hs–CRP	–	–	–	0.518	–
Δ White blood cells	–	–	–	0.344	–
Δ HDL-cholesterol	–	–	–	–	0.779
Δ Apolipoprotein A1	–	–	–	–	0.696
% of explained variance	36.6	21.8	17.5	13.0	10.5

All variables were adjusted for age and specific drug use, separately for fathers, mothers, sons and daughters for the cross-sectional study and for age, specific drug use and values at entrance for the longitudinal study. Log_{10} transformed values were used for triglycerides, hs-CRP, ALAT, ASAT and GGT.
Δ, change over 5 years; ALAT, alanine aminotransferase; ASAT, aspartate aminotransferase; GGT, gamma-glutamyl transferase; HDL, high-density lipids; hs-CRP, high-sensitivity C-reactive protein.
[a] Factor loadings represent the correlations between the variables and the factors.
[b] Factor loading <0.29.

uric acid that was no longer included in the 'adiposity/blood pressure' factor in the longitudinal study.

3.2. Genetic heritability and common environmental variance

We used the model that assumed three independent components (polygenic, common (familial) environmental and random non familial) with no differences in parts of component variances according to generation for all clusters and individual traits. Only the estimates of genetic heritability and common environmental variance are shown in Table 4 for baseline and 5-year changes, respectively. At entrance, polygenic variance was statistically significant for total cholesterol, apolipoprotein B, apolipoprotein E, uric acid, fasting glucose, HDL-cholesterol, apolipoprotein A1, ALAT, haptoglobin and three clusters ("risk lipids", "protective lipids" and "liver enzymes"); values varied from 18.4% to 44.0%. Conversely, for triglycerides, waist circumference, body weight index, systolic and diastolic blood pressure, activities of ALAT and GGT, hs-CRP, orosomucoid, WBC and two clusters ("adiposity/blood pressure" and "inflammation"), polygenic variance was not statistically significant. At entrance, common environmental variances were statistically significant from zero for all individual variables and clusters with a range of 7.8–25.8%. With regard to 5-year changes, polygenic variance was no longer statistically significant for any of the individual variables or factors (Table 4). Common environmental variance was statistically significant for all variables except for triglycerides, hs-CRP, ASAT activity and "liver enzyme" cluster.

4. Discussion

The "STANISLAS Family Study" was planned in order to estimate genetic heritability and common environmental influence for risk factors of complex diseases such as the metabolic syndrome, taking advantage of specific features of this unique sample of French nuclear families, including parents of middle age and young children in good health. In this paper, with regard to metabolic syndrome-related factors: 1) we added to the literature with our specific data about components of familial aggregation of traditional traits and clusters; 2) we estimated heritability of novel associated traits and clusters (to our knowledge, no data exists for these traits); 3) we estimated components of familial aggregation for short-term changes of these associated traits and clusters. In addition, due to the design of the STANISLAS Family Cohort, subjects of this study were not excluded on the basis of elevated risk factors or drug use. Including such subjects make the results about familial aggregation potentially applicable to general population of the east of France.

4.1. Factor analysis

The factor analysis approach has been used in previous studies of the metabolic syndrome [23–28], including family aggregation [9–12] and has yielded a variety of possible models. We identified five summary factors named as "risk lipids", "adiposity/blood pressure", "protective lipids", "inflammation" including inflammatory markers and "liver enzymes" including three important markers of liver metabolism. Noteworthy, these two last clusters were reported only in the STANISLAS Study. Likewise, when using patterns of changes of these metabolic variables, the same five summary clusters were revealed and confirmed, suggesting a relationship between these variables and their changes over time.

In our population, blood pressure was loaded with BWI and WC, contrary to most previous studies. Our results suggest a strong relationship of adiposity with blood pressure regulation, or the existence of a common causal factor that underlies the different components of the metabolic syndrome. These findings are in agreement with a familial study on metabolic syndrome on supposed healthy individuals [10]. Interrelationships between adiposity and

Table 4
Variance components for individual variables and the 5 factor scores in the 667 families included in the cross-sectional study (entrance variables) and in the 353 families included in the longitudinal study (5-year changes).

	Entrance variables		5-year changes	
	σ_G^2 (polygenic variance, %)	σ_C^2 (common environmental variance, %)	σ_G^2 (polygenic variance, %)	σ_C^2 (common environmental variance, %)
Total cholesterol	44.0***	7.8*	0.0	12.9***
Apolipoprotein B	38.5***	12.6***	0.0	16.5***
Triglycerides	11.7	16.8***	3.7	5.9
Apolipoprotein E	34.7***	10.6**	0.0	22.4***
Risk lipid cluster	40.5***	8.9**	0.0	10.2***
Waist circumference	4.1	25.8***	0.0	15.8***
Body weight index	7.0	22.6***	0.0	11.1***
Systolic blood pressure	10.9	8.8***	0.0	13.4***
Diastolic blood pressure	0.0	12.2***	7.1	16.1**
Uric acid	37.8***	12.0***	8.1	23.7***
Adiposity/blood pressure cluster	6.8	21.0***	0.0	14.3***
Fasting glucose	20.8***	17.6***	0.0	24.0***
HDL-cholesterol	32.8***	16.3***	0.0	27.2***
Apolipoprotein A1	30.0***	23.4***	0.0	25.5***
Protective lipid cluster	35.0***	14.9***	0.0	18.3***
ALAT	18.8**	9.0**	0.0	5.7*
ASAT	13.6	14.3***	0.0	2.6
GGT	8.4	11.9***	0.0	10.4***
Liver enzyme cluster	18.4**	9.8**	0.0	2.9
Haptoglobin	35.3***	10.0**	7.3	9.9*
Hs-CRP	5.9	11.0**	0.0	5.6
Orosomucoid	8.9	16.7***	0.0	7.4**
White blood cells	4.4	16.8***	0.0	8.3**
Inflammation cluster	7.6	14.2***	0.0	9.0**

*$p<0.05$, **$p<0.01$, ***$p<0.001$, else: non significant.
Variance components were expressed as percentage.
All variables were adjusted for age and specific drug use for the cross-sectional study and, for age, specific drug use and values at entrance for the longitudinal study, separately for fathers, mothers, sons and daughters except for clusters. Log_{10} transformed values were used for triglycerides, hs-CRP, ALAT, ASAT and GGT.
ALAT, alanine aminotransferase; ASAT, aspartate aminotransferase; GGT, gamma-glutamyl transferase; HDL, high-density lipids; Hs-CRP, high-sensitivity C-reactive protein.

blood pressure could be different in healthy individuals as compared to patients with insulin resistance or diabetes. In contrast, we did not find the common cluster named "glucose/insulin resistance" [9,10,25–28]; this is probably due to lack of biological parameters related to glucose metabolism such as fasting insulin, as well as to the healthy state of our sample of population.

4.2. Heritability

In order to compare our data, we collected the main worldwide results from twin and family studies (nuclear or of complex pedigree) where genetic and common environmental variances were estimated (see online Supplementary material). Values of genetic heritability in this table are not absolute measurements, but rather reflect the amount of genotypic variation as compared to environmental variation. The estimates are specific to a given population and are generally larger when the genetic background is more diverse (increasing of the genetic variance) or when environmental effects are minimized (decreasing of the environmental variance). Consequently, heritability for the same trait may vary quite widely across populations where the relative importance of these components is very different [20]. Comparison of data of the literature should take into account important criteria such as: design (twins, nuclear families or more complex pedigrees), origin and characteristics of the population (age, sex, lifestyle, genetic background, genetic isolate, with or without risk factors, acute or chronic diseases), number and type of traits studied, adjustment for covariates, method used to estimate heritability (path analysis or analysis of variance). As an example, the online table (see Supplementary material) shows that genetic heritabilities from twin studies are higher than those estimated from nuclear family studies or complex pedigrees.

With regard to genetic heritability our results are high (20 to 44%) and close to the estimates or within the ranges described by other family studies for plasma lipids and lipoproteins (total cholesterol, apo B, apo E, apo A1, HDL-cholesterol), uric acid and fasting glucose. Conversely, genetic heritability for triglycerides, adiposity indices (waist circumference and body weight index), blood pressure (both diastolic and systolic), hepatic enzyme activity (ALAT, ASAT, GGT) and inflammatory markers (hs-CRP, orosomucoid and white blood cells) is low (less than 19%), smaller than published data and often not statistically significant. Genetic heritability of haptoglobin level represents about a third of global variance and seems to be an exception. However, in spite of no data exists about familial resemblance of this protein, haptoglobin has been described as polymorphic with three major phenotypes that are associated with differences in serum concentrations [29].

Estimates of genetic heritability at entrance for the five clusters are consistent with those of the related individual trait: high for "risk lipids" and "protective lipids" (40% and 35%, respectively), intermediate for "liver enzymes" (18%) and low for "adiposity/blood pressure" and "inflammation" (about 7%). However, triglycerides and uric acid appear as exceptions, probably due to low factor loadings with their respective clusters. Discordance between heritability estimates of "inflammation" cluster and haptoglobin remains unexplained. Higher genetic heritability for "risk lipids", "protective lipids" and related individual traits is consistent with the success to date in identifying genetic polymorphism associated with raised cholesterol, apo- and lipoproteins in general population [30]. On the contrary, small genetic components for "adiposity/blood pressure" and "inflammation factors" could reflect the difficulties to identify genomic regions involved in the regulation of these related factors in the general population (for instance, for body weight and obesity, [31]). Generally, the genetic component is lower and environmental influence is higher in nuclear families including young parents and children than those including older subjects [32,33].

When comparing common environmental influence with previous studies (see online Supplementary material), our estimates are often higher or close to the upper range and more than 14% for triglycerides, waist circumference, body mass index, adiposity/blood pressure cluster, glucose, HDL-cholesterol, protective lipid cluster and apo A1; risk factors that are generally influenced by long-term lifestyle interventions [34]. Common environmental influence observed in our study for orosomucoid, white blood cells count and the related cluster could be due to familial aggregation of various benign infections. Unfortunately, for numerous factors, no data exists about common environmental influence for comparison with our results.

With regard to 5-year changes, polygenic variance was low and not statistically significant for any of the individual variables or factors whereas shared environment influence was significant for most factors and clusters except for triglycerides, hs-CRP, ASAT and related cluster. Our findings highlight the impact of the common familial environment and lifestyle in the short-term change of metabolic syndrome components. While genetic factors have only an effect at baseline, only environmental factors seem to influence substantially changes of metabolic syndrome-related variables over short time. Data about change of these factors are very scarce and hard to compare with ours. For instance, a recent family study showed that changes in BMI over 8–10 year period appear to be genetically influenced (heritability of 36%), while heritability estimates for changes in waist was lower (21%) [35]. Similarly, a 12-year follow-up in the Quebec Family Study suggested that 36% of the phenotypic variability of BMI was transmissible [36]. Discrepancies with our data could be due to difference in methodology: the first study used complex pedigree (80 families and 379 subjects) with about 73% of subject aged 45 years and more [35]; in the second study, only global heritability due to both genetic factors and the shared common familial environment was estimated [36].

4.3. Strengths and limitations

The major strengths of our present study were that the participants were young (including children and adolescents), stemmed from nuclear families (without complex pedigrees), at low-risk of cardiovascular diseases, drawn from the general population of the east of France without genetic isolation phenomenon. No study about familial aggregation of MS factors was ever done in this area. Moreover, both cross-sectional and short-term longitudinal approaches were studied and both individual traits and clusters were analyzed. For these reasons, our data specific of young, healthy and unselected French families are unique.

Several limitations of our study have also to be addressed. First, estimates of genetic heritability didn't take into account major gene effects that are largely or entirely nonadditive, gene–gene interactions (epistasis) and gene–environment interactions. Second, as in our study in samples of nuclear families, consisting of parents and offspring, the genetic and familial environmental effects could be resolvable with difficulties because these relatives share both genes and environments. Third, various factors could affect the estimation and interpretability of heritabilities, for example, assumptions regarding linearity and additivity, assortative mating, and the underlying distribution of the data (non Gaussian distribution of the data can lead to errors in hypothesis testing) [20]. Nevertheless, the software FISHER performed goodness-of-fit test of the underlying multinormal distribution. Lastly, genetic heritability estimates from clusters could be under or overestimate due to individual traits included in there computation (for instance triglycerides).

In conclusion, in these young families from the east of France, estimates of genetic heritability at entrance were generally lower than previously reported while the common environmental influences were greater. Only common environmental factors are related to short-term changes. Our data highlight the importance of lifestyle, at least for short-time periods, on the evolution of the metabolic syndrome-related traits.

Acknowledgments

We are deeply grateful for the cooperation of the families participating in the STANISLAS Cohort. We acknowledge the management, reception, preclinical, laboratory and medical staff of the Center for Preventive Medicine of Vandoeuvre-lès-Nancy (France). We also thank David Trégouët (INSERM U 525, Paris, France) for providing the program package of familial correlation computation. The STANISLAS Cohort Study is supported by the Caisse Nationale d'Assurance Maladies des Travailleurs Salariés (CNAM), the Institut National de la Santé et de la Recherche Médicale (INSERM), the Région Lorraine, the Communauté Urbaine du Grand Nancy (CUGN) and the Université Henri Poincaré (UHP).

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at doi:10.1016/j.cca.2010.02.070.

References

[1] Reaven GM. Banting lecture 1988. Role of insulin resistance in human disease. Diabetes 1988;37:1595–607.
[2] Engstrom G, Stavenow L, Hedblad B, et al. Inflammation-sensitive plasma proteins and incidence of myocardial infarction in men with low cardiovascular risk. Arterioscler Thromb Vasc Biol 2003;23:2247–51.
[3] Rutter MK, Meigs JB, Sullivan LM, D'Agostino Sr RB, Wilson PW. C-reactive protein, the metabolic syndrome, and prediction of cardiovascular events in the Framingham Offspring Study. Circulation 2004;110:380–5.
[4] Vozarova B, Weyer C, Lindsay RS, Pratley RE, Bogardus C, Tataranni PA. High white blood cell count is associated with a worsening of insulin sensitivity and predicts the development of type 2 diabetes. Diabetes 2002;51:455–61.
[5] Ford ES, Li C, Cook S, Choi HK. Serum concentrations of uric acid and the metabolic syndrome among US children and adolescents. Circulation 2007;115:2526–32.
[6] Onat A, Uyarel H, Hergenc G, et al. Serum uric acid is a determinant of metabolic syndrome in a population-based study. Am J Hypertens 2006;19:1055–62.
[7] Rector RS, Thyfault JP, Wei Y, Ibdah JA. Non-alcoholic fatty liver disease and the metabolic syndrome: an update. World J Gastroenterol 2008;14:185–92.
[8] Tang W, Hong Y, Province MA, et al. Familial clustering for features of the metabolic syndrome: the National Heart, Lung, and Blood Institute (NHLBI) Family Heart Study. Diabetes Care 2006;29:631–6.
[9] Austin MA, Edwards KL, McNeely MJ, et al. Heritability of multivariate factors of the metabolic syndrome in nondiabetic Japanese Americans. Diabetes 2004;53:1166–9.
[10] Chien KL, Hsu HC, Chen WJ, Chen MF, Su TC, Lee YT. Familial aggregation of metabolic syndrome among the Chinese: report from the Chin-Shan community family study. Diabetes Res Clin Pract 2007;76:418–24.
[11] Lin HF, Boden-Albala B, Juo SH, Park N, Rundek T, Sacco RL. Heritabilities of the metabolic syndrome and its components in the Northern Manhattan Family Study. Diabetologia 2005;48:2006–12.
[12] Stirnadel H, Lin X, Ling H, et al. Genetic and phenotypic architecture of metabolic syndrome-associated components in dyslipidemic and normolipidemic subjects: the GEMS Study. Atherosclerosis 2008;197:868–76.
[13] Best LG, North KE, Tracy RP, et al. Genetic determination of acute phase reactant levels: the strong heart study. Hum Hered 2004;58:112–6.
[14] de Maat MP, Bladbjerg EM, Hjelmborg JB, Bathum L, Jespersen J, Christensen K. Genetic influence on inflammation variables in the elderly. Arterioscler Thromb Vasc Biol 2004;24:2168–73.
[15] Siest G, Visvikis S, Herbeth B, et al. Objectives, design and recruitment of a familial and longitudinal cohort for studying gene–environment interactions in the field of cardiovascular risk: the Stanislas cohort. Clin Chem Lab Med 1998;36:35–42.
[16] Visvikis-Siest S, Siest G. Cohort: a 10-year follow-up of supposed healthy families. Gene–environment interactions, reference values and evaluation of biomarkers in prevention of cardiovascular diseases. Clin Chem Lab Med 2008;46:733–47.
[17] Alberti KG, Zimmet P, Shaw J. The metabolic syndrome—a new worldwide definition. Lancet 2005;366:1059–62.
[18] Jolliffe CJ, Janssen I. Development of age-specific adolescent metabolic syndrome criteria that are linked to the Adult Treatment Panel III and International Diabetes Federation criteria. J Am Coll Cardiol 2007;49:891–8.
[19] Rolland-Cachera MF, Cole TJ, Sempe M, Tichet J, Rossignol C, Charraud A. Body Mass Index variations: centiles from birth to 87 years. Eur J Clin Nutr 1991;45:13–21.
[20] Rice TK. Familial resemblance and heritability. Adv Genet 2008;60:35–49.
[21] Lange K, Westlake J, Spence MA. Extensions to pedigree analysis. III. Variance components by the scoring method. Ann Hum Genet 1976;39:485–91.
[22] Lange K, Weeks D, Boehnke M. Programs for pedigree analysis: MENDEL, FISHER, and dGENE. Genet Epidemiol 1988;5:471–2.
[23] Hanley AJ, Karter AJ, Festa A, et al. Factor analysis of metabolic syndrome using directly measured insulin sensitivity: the Insulin Resistance Atherosclerosis Study. Diabetes 2002;51:2642–7.
[24] Kraja AT, Hunt SC, Pankow JS, et al. An evaluation of the metabolic syndrome in the HyperGEN study. Nutr Metab (Lond) 2005;2:2–11.
[25] Maison P, Byrne CD, Hales CN, Day NE, Wareham NJ. Do different dimensions of the metabolic syndrome change together over time? Evidence supporting obesity as the central feature. Diabetes Care 2001;24:1758–63.
[26] Meigs JB, D'Agostino Sr RB, Wilson PW, Cupples LA, Nathan DM, Singer DE. Risk variable clustering in the insulin resistance syndrome. The Framingham Offspring Study. Diabetes 1997;46:1594–600.
[27] Sakkinen PA, Wahl P, Cushman M, Lewis MR, Tracy RP. Clustering of procoagulation, inflammation, and fibrinolysis variables with metabolic factors in insulin resistance syndrome. Am J Epidemiol 2000;152:897–907.
[28] Shen BJ, Todaro JF, Niaura R, et al. Are metabolic risk factors one unified syndrome? Modeling the structure of the metabolic syndrome X. Am J Epidemiol 2003;157:701–11.
[29] Carter K, Worwood M. Haptoglobin: a review of the major allele frequencies worldwide and their association with diseases. Int J Lab Hematol 2007;29:92–110.
[30] Sviridov D, Nestel PJ. Genetic factors affecting HDL levels, structure, metabolism and function. Curr Opin Lipidol 2007;18:157–63.
[31] Rankinen T, Zuberi A, Chagnon YC, et al. The human obesity gene map: the 2005 update. Obesity (Silver Spring) 2006;14:529–644.
[32] Saunders CL, Gulliford MC. Heritabilities and shared environmental effects were estimated from household clustering in national health survey data. J Clin Epidemiol 2006;59:1191–8.
[33] Pilia G, Chen WM, Scuteri A, et al. Heritability of cardiovascular and personality traits in 6,148 Sardinians. PLoS Genet 2006;2:1207–23.
[34] Brown T, Avenell A, Edmunds LD, et al. Systematic review of long-term lifestyle interventions to prevent weight gain and morbidity in adults. Obes Rev 2009;10:627–38.
[35] Friedlander Y, Meiner V, Sharon N, Siscovick DS, Miserez AR. Leptin, insulin, and obesity-related phenotypes: genetic influences on levels and longitudinal changes. Obesity (Silver Spring) 2009;17:1458–60.
[36] Rice T, Perusse L, Bouchard C, Rao DC. Familial aggregation of body mass index and subcutaneous fat measures in the longitudinal Quebec family study. Genet Epidemiol 1999;16:316–34.

PUBLICATION N°3

Ndeye Coumba Ndiaye*, Éliane Albuisson*, Isabelle Colombet, Gilles Chatellier, Jean-Brice Marteau, Sébastien Dadé, Gérard Siest, Sophie Visvikis-Siest. Parental precocious influences on offspring cardiovascular risk markers: an exploratory study in the STANISLAS cohort. Pers Med. 2009;6(3):343-52.

Les antécédents familiaux des pathologies cardiovasculaires sont des facteurs de risque connus des maladies cardiovasculaires. Le but de notre étude était donc d'évaluer l'influence des niveaux de risque parentaux sur le niveau de ces mêmes facteurs de risque chez leurs enfants dans une cohorte familiale d'origine Française : la cohorte STANISLAS dans laquelle des familles biparentales avec au moins deux enfants biologique avaient été suivies pendant cinq ans à l'occasion de deux examens de santé. Les facteurs de risque des enfants (pression artérielle, traits lipidiques, glycémie, indice de masse corporelle et tour de taille), étaient comparés en fonction du niveau de risque des parents dans 693 familles. À l'exception de la glycémie, tous ces traits, bien qu'infracliniques, étaient plus élevés dans les familles ayant les deux parents à risque pour le trait considéré à t_0 et à t_5.
Ainsi, il serait opportun d'envisager des stratégies préventives familiales pour mieux repérer les enfants à risque cardiovasculaire et ainsi mener des interventions précoces.

Contribution :
Lors de cette étude exploratoire, nous avons conçu le design, effectué la totalité des analyses statistiques, construit les tables et figures ainsi qu'interprété les résultats. Nous avons également rédigé la première version de l'article.

For reprint orders, please contact: reprints@futuremedicine.com

Parental precocious influences on offspring cardiovascular risk markers: an exploratory study in the STANISLAS Cohort

Familial history of cardiovascular disease is acknowledged as a risk indicator in offspring. The aim of this study was to assess whether the cardiovascular risk factors in parents predicted the risk of their children developing cardiovascular disease in a French population: the STANISLAS Cohort, in which Caucasian biparental families with at least two siblings were followed for 5 years. Silent risk factors (blood pressure, lipid traits, glycemia, BMI and waist circumference) of children were compared according to their parents' risk status in a subsample of 693 families. All of these traits, with the exception of glucose, were significantly higher in children who had parents at a high risk than in children with parents at a low risk at the first health examination, and these results were confirmed again 5 years later at the second health examination. Thus, silent cardio–metabolic risk factors can be screened in children according to the risk status of their parents for early prevention. The influence of parents' variants on their offspring underlined the need to initiate familial prevention strategies, with a particular follow-up of young individuals between childhood and adolescence.

KEYWORDS: cardiovascular precocious risk personalized prevention silent risk factor

Ndeye Coumba Ndiaye[1,2], Éliane Albuisson[1,2*], Isabelle Colombet[3,4], Gilles Chatellier[3,4], Jean-Brice Marteau[1], Sébastien Dadé[1], Gérard Siest[1] & Sophie Visvikis-Siest[1]

*Author for correspondence: [1]Université Henri Poincaré Nancy I – Faculté de Pharmacie, Équipe 'Génétique Cardiovasculaire' – 30 rue Lionnois, F-54000, Nancy, France Tel.: +33 383 682 184, Fax: +33 383 321 322, sophie.visvikis-siest@nancy.inserm.fr [2]Centre Hospitalier Universitaire, Nancy, France [3]Université Paris V, Paris, France [4]Hôpital Européen Georges Pompidou, Paris, France [2]These authors contributed equally to this work

Cardiovascular disorders are still the major cause of mortality and morbidity in Western countries, and the installation of its risk factors begin during childhood [1]. Moreover, familial influence in cardiovascular diseases can begin early, even in healthy families, and target-organ damage may occur in children with elevated risk factors [2]. This highlights the importance of studying both genetic and environmental factors in order to improve primary prevention strategies focused on young populations at high risk. Indeed, it seems important to identify silent cardio–metabolic risk factors in childhood.

In this study, we examined whether cardiovascular risk factors in parents predicted the risk of their children developing cardiovascular disease.

Methods

Study design

The STANISLAS Family Study consists of 1006 biparental families with at least two siblings over the age of 4 years. They were recruited between 1993 and 1995 at the Centre of Preventive Medicine of Vandoeuvre-lès-Nancy (eastern France) and underwent health examinations every 5 years [3]. Individuals suffering from acute or chronic disorders or with a previous personal history of cardiovascular disease were excluded. All participants gave their written informed consent and the study was approved by local ethics committees. This cohort was designed to assess, in a healthy population, the role of genetics and environmental factors in the variability of cardiovascular risk factors.

Initially, we performed a preliminary unpublished study including the adults and adolescents of the STANISLAS Cohort using the Framingham prediction equation for coronary disease at 10 years [4]. However, since the Framingham equation was developed with an American population, it overestimates risk assessment in European populations [5]; in particular, the influence of age was too strong in the model, making the classification of individuals biased. Therefore, we chose to study the following cardiovascular risk factors in children: systolic and diastolic blood pressure (SBP; DBP), total cholesterol (TC), low-density lipoprotein cholesterol (LDL-C), high-density lipoprotein cholesterol (HDL-C), triglycerides (TG), glucose, BMI and waist circumference (WC) according to their parental risk status. Thus, the metabolic pathways highlighted in this study will allow the future integration of genetic and nongenetic information into medical decisions [6].

Study population

The current study includes cross-sectional and longitudinal analyses of data collected between 1993–1995 and 1999–2000 (corresponding to two health examinations) on

future medicine part of fsg

2580 individuals belonging to 693 biparental families (693 fathers and mothers, 585 sons and 609 daughters).

Data collection

Blood pressure was measured under standardized conditions (supine position) with a sphygmomanometer. Three readings were taken and the mean value was used for analysis.

Blood samples were collected between 8:00 and 9:00 am or 11:30 am and 12:30 pm by venipuncture after overnight fasting. TC, HDL-C, TG and glucose were assayed using enzymatic methods (AU640 [Olympus, Watford, UK]) and LDL-C was calculated using the Friedewald formula.

Weight and height were measured while the participants were standing in light clothing without shoes. BMI was calculated as weight in kg divided by height in m^2, and the WC was measured at the level of the umbilicus.

Parental risk status definition

In the first step, parents at risk were determined according to thresholds defined by international guidelines and recommendations:

- SBP greater than 130 mmHg for prehypertension and above 140 mmHg for hypertension [7];
- DBP greater than 85 mmHg for prehypertension and above 90 mmHg for hypertension [7];
- TC of 6 mmol/l or more [8];
- LDL-C of 5.16 mmol/l or more [101];
- HDL-C of 1 mmol/l or less [101];
- TG of 2 mmol/l or more [101];
- Glucose greater than 6.2 mmol/l [9];
- BMI of 25 kg/m^2 or more;
- WC greater than 80 cm for fathers and 94 cm for mothers [10].

Child risk status

Children's blood pressure, lipid traits, glycemia, BMI and WC were analyzed in a second step according to their parent's risk status. Three groups of children were studied:

- Children with neither parent at risk – groups 0;
- Children with only one parent (father or mother) at risk – groups 1;
- Children with both parents at risk – groups 2.

Statistical analyses

Risk factors among children and parents were described by the mean ± standard deviation.

Differences between groups were tested at the first health examination (t0) and 5 years later at the second health examination (t5) using the Student's t-test, and analysis of variance adjusted results for sex and age, taking into account possible interactions between sexes and groups. Variation between the two health examinations was also assessed among groups. Continuous variables were log-transformed to approach normality and the Levene test established the homogeneity of variance between groups. The SPSS® statistical software version 16.0 (SPSS Inc., IL, USA) was used and the level of significance (p-value) was 0.01.

Results

Phenotypic characteristics of the population studied at t0 and t5 and longitudinal variations are summarized in TABLE 1 (mean values ± standard deviation of all variables assessed by family members). The mean age of fathers was 42.85 years, mothers 40.75 years, sons 13.35 years and daughters 13.58 years of age at t0. Fathers compared with mothers had significantly higher levels for all measurements with the exception of HDL-C. Among children, SBP, pulse pressure, mean blood pressure, TC, LDL-C, TG, glucose and WC were significantly higher in male versus female siblings at t0. A total of 5 years later, HDL-C became significantly lower in daughters and TG stabilized among genders.

Comparisons between children's groups

Children's parameters significantly differed between groups ($p < 0.001$) at t0 and t5, except for DBP (hypertension threshold) and glucose (TABLE 2). Lower values for all parameters were observed in groups 0 and higher values in groups 2, except for HDL-C (FIGURES 1 & 2). No interactions between gender and groups were observed. Differences in blood pressure within groups were influenced by age and gender for SBP at t0 and t5. The gender effect was significant for TC and HDL-C at t5 and the age effect became apparent at the second examination for TC and LDL-C, but not for HDL-C. TG levels were modulated by age but not by gender. The age effect was also significant in BMI and WC, although gender modulated this last variable only. Longitudinal variations differed among groups for SBP and TG.

Table 1. Phenotypic characteristics of family members.

Phenotypes	Health examination	Fathers (n = 693)*	Mothers (n = 693)*	Student t-test p-value	Levene test p-value	Sons (n = 585)*	Daughters (n = 609)*	Student t-test p-value	Levene test p-value
Age (years)	t0	42.85 ± 5.08	40.7 ± 4.8	0.000	0.254	13.3 ± 3.6	13.6 ± 3.9	0.290	0.030
	t5	47.56 ± 5.12	45.5 ± 4.8	0.000	0.257	18.0 ± 3.6	18.3 ± 3.9	0.257	0.045
SBP (mmHg)	t0	127.6 ± 12.3	119.2 ± 12.5	0.000	0.008	115.2 ± 12.3	111.0 ± 9.5	0.000	0.000
	t5	128.4 ± 13.4	121.1 ± 14.2	0.000	0.904	122.4 ± 10.8	115.4 ± 9.1	0.000	0.000
	Δ	1.2 ± 12.2	1.9 ± 12.3	0.254	0.168	7.2 ± 11.6	4.4 ± 10.2	0.000	0.001
DBP (mmHg)	t0	78.1 ± 9.7	72.5 ± 9.7	0.000	0.001	57.8 ± 11.0	57.5 ± 10.7	0.668	0.441
	t5	77.6 ± 10.2	72.2 ± 9.9	0.000	0.369	64.0 ± 10.5	62.9 ± 1.0	0.074	0.236
	Δ	-0.5 ± 11.0	-0.3 ± 10.5	0.703	1.157	6.2 ± 12.5	5.4 ± 12.1	0.299	0.949
PP (mmHg)	t0	49.5 ± 9.3	46.7 ± 8.8	0.000	6.641	57.4 ± 12.1	53.5 ± 10.1	0.000	0.000
	t5	51.1 ± 9.4	48.9 ± 9.0	0.000	0.823	58.5 ± 12.5	52.5 ± 10.9	0.000	0.001
	Δ	1.7 ± 10.6	2.2 ± 10.4	0.348	0.034	1.1 ± 15.1	-1.0 ± 13.3	0.010	0.001
MPB (mmHg)	t0	94.6 ± 9.7	88.1 ± 9.9	0.000	0.380	76.9 ± 9.9	75.3 ± 9.1	0.005	0.026
	t5	94.6 ± 10.5	88.5 ± 10.7	0.000	0.212	83.5 ± 8.8	80.4 ± 8.2	0.000	0.123
	Δ	0.1 ± 10.3	0.5 ± 10.0	0.467	0.319	6.5 ± 9.9	5.1 ± 9.7	0.011	0.980
TC (mmol/l)	t0	5.99 ± 1.12	5.53 ± 0.90	0.000	21.268	4.63 ± 0.82	4.87 ± 0.80	0.000	0.465
	t5	5.98 ± 1.03	5.62 ± 0.93	0.000	5.355	4.37 ± 0.80	4.80 ± 0.92	0.000	0.062
	Δ	-0.02 ± 0.94	0.08 ± 0.78	0.031	8.204	-0.27 ± 0.69	-0.07 ± 0.73	0.000	0.601
LDL-C (mmol/l)	t0	4.06 ± 0.98	3.54 ± 0.86	0.000	11.277	2.88 ± 0.76	3.06 ± 0.75	0.000	0.704
	t5	3.81 ± 0.90	3.34 ± 0.85	0.000	1.856	2.51 ± 0.75	2.71 ± 0.83	0.000	0.440
	Δ	-0.25 ± 0.76	-0.21 ± 0.66	0.223	9.741	-0.37 ± 0.61	-0.35 ± 0.61	0.589	0.480
HDL-C (mmol/l)	t0	1.27 ± 0.35	1.56 ± 0.39	0.000	10.778	1.40 ± 0.36	1.43 ± 0.34	0.146	0.126
	t5	1.42 ± 0.40	1.75 ± 0.44	0.000	11.250	1.39 ± 0.34	1.63 ± 0.39	0.000	0.003
	Δ	0.15 ± 0.28	0.19 ± 0.33	0.008	17.119	-0.01 ± 0.32	0.20 ± 0.33	0.000	0.345
TG (mmol/l)	t0	1.39 ± 1.47	0.88 ± 0.53	0.000	51.105	0.70 ± 0.40	0.76 ± 0.36	0.005	0.758
	t5	1.54 ± 1.06	1.10 ± 1.49	0.000	20.511	0.92 ± 0.42	0.93 ± 0.44	0.654	0.656
	Δ	0.15 ± 1.22	0.23 ± 1.45	0.303	11.330	0.22 ± 0.42	0.17 ± 0.46	0.054	0.195
Glucose (mmol/l)	t0	5.22 ± 0.70	4.92 ± 0.42	0.000	9.010	4.94 ± 0.37	4.84 ± 0.37	0.000	0.623
	t5	5.30 ± 0.91	4.93 ± 0.54	0.000	12.101	4.84 ± 0.70	4.70 ± 0.48	0.000	0.233
	Δ	0.07 ± 0.81	0.01 ± 0.56	0.080	13.406	-0.10 ± 0.76	-0.15 ± 0.56	0.204	0.243

Mean ± standard deviation.

Δ: Variation between two health examinations; DBP: Diastolic blood pressure; HDL: High-density lipoprotein; LDL: Low-density lipoprotein; MBP: Mean blood pressure; n: Sample size; PP: Pulse pressure; SBP: Systolic blood pressure; t0: Values at first health examination; t5: Values at second health examination; TC: Total cholesterol; TG: Triglycerides; WC: Waist circumference.

Table 1. Phenotypic characteristics of family members.

Phenotypes	Health examination	Fathers (n = 693)*	Mothers (n = 693)*	Student t-test p-value	Levene test p-value	Sons (n = 585)*	Daughters (n = 609)*	Student t-test p-value	Levene test p-value
BMI (kg/m²)	t0	25.5 ± 3.3	23.8 ± 4.0	0.000	21.516	18.8 ± 3.3	19.0 ± 3.2	0.284	0.967
	t5	26.3 ± 3.6	24.8 ± 4.5	0.000	29.076	21.2 ± 3.3	21.1 ± 3.2	0.565	0.602
	Δ	0.9 ± 1.3	1.0 ± 1.6	0.114	21.529	2.4 ± 1.8	2.1 ± 2.3	0.011	0.000
WC (cm)	t0	88.5 ± 8.8	75.0 ± 8.8	0.000	0.001	66.5 ± 8.7	62.5 ± 7.2	0.000	0.000
	t5	90.7 ± 9.2	77.2 ± 10.2	0.000	3.608	74.4 ± 8.1	67.4 ± 6.8	0.000	0.000
	Δ	2.4 ± 4.7	2.2 ± 4.7	0.423	0.017	7.7 ± 5.3	4.9 ± 5.8	0.000	0.121

**Mean ± standard deviation.*
Δ: Variation between two health examinations; DBP: Diastolic blood pressure; HDL: High-density lipoprotein; LDL: Low-density lipoprotein; MBP: Mean blood pressure; n: Sample size; PP: Pulse pressure; SBP: Systolic blood pressure; t0: Values at first health examination; t5: Values at second health examination; TC: Total cholesterol; TG: Triglycerides; WC: Waist circumference.

Discussion

It is now well established that cardiovascular disorders have roots in childhood [2], with both genetic and environmental components.

Thus, prevention efforts should begin precociously with the early detection of silent cardiovascular risk factors. In this respect, our results demonstrated associations between cardiovascular risk factors in children and their parental risk status in a healthy French population. These relationships were maintained through time and concerned almost all the cardiovascular risk markers studied: blood pressure, lipid traits, BMI and WC. This enables us to confirm that the different metabolic pathways implicated in the genesis of cardiovascular pathophysiology take asset early on through the combination of parental genetics and a shared environment, even in families without any personal history of cardiovascular disease.

The effect of gender on physiological traits related to the genesis of cardiovascular disorders was also confirmed, with a higher risk in male than in female adults [11]. Indeed, males and females represented two different targets for prevention strategies.

Differences between groups observed for blood pressure were in the same range as results generally obtained by prevention strategies using antihypertensive therapy [8]. Obviously, genetic variations also determine our phenotypic observations, since some evidence suggests that genetic contribution to blood pressure is approximately 30–50% [12].

The evolution of lipid traits over 5 years suggests that those traits are dependant on hormonal differences between sons and daughters [13]. We observed that the gender effect became significant for TC and LDL-C.

Since BMI and WC traits are strongly influenced by lifestyle, results on BMI and WC may blur the boundaries between genetics and environmental factors. However, the *FTO* polymorphism, previously associated with early-onset and morbid adult obesity [14,15], was recently linked to the interindividual variation of BMI within a subgroup of 391 adults aged 30–54 years and 430 children and adolescents aged 5–20 years extracted from the STANISLAS Cohort [102]. Other studies have also been conducted on this subject: the Quebec Family Study demonstrated that genes influenced the propensity to store fat in the abdominal area [16], and the cardiovascular component of the European Prospective Investigation into Cancer and Nutrition [17],

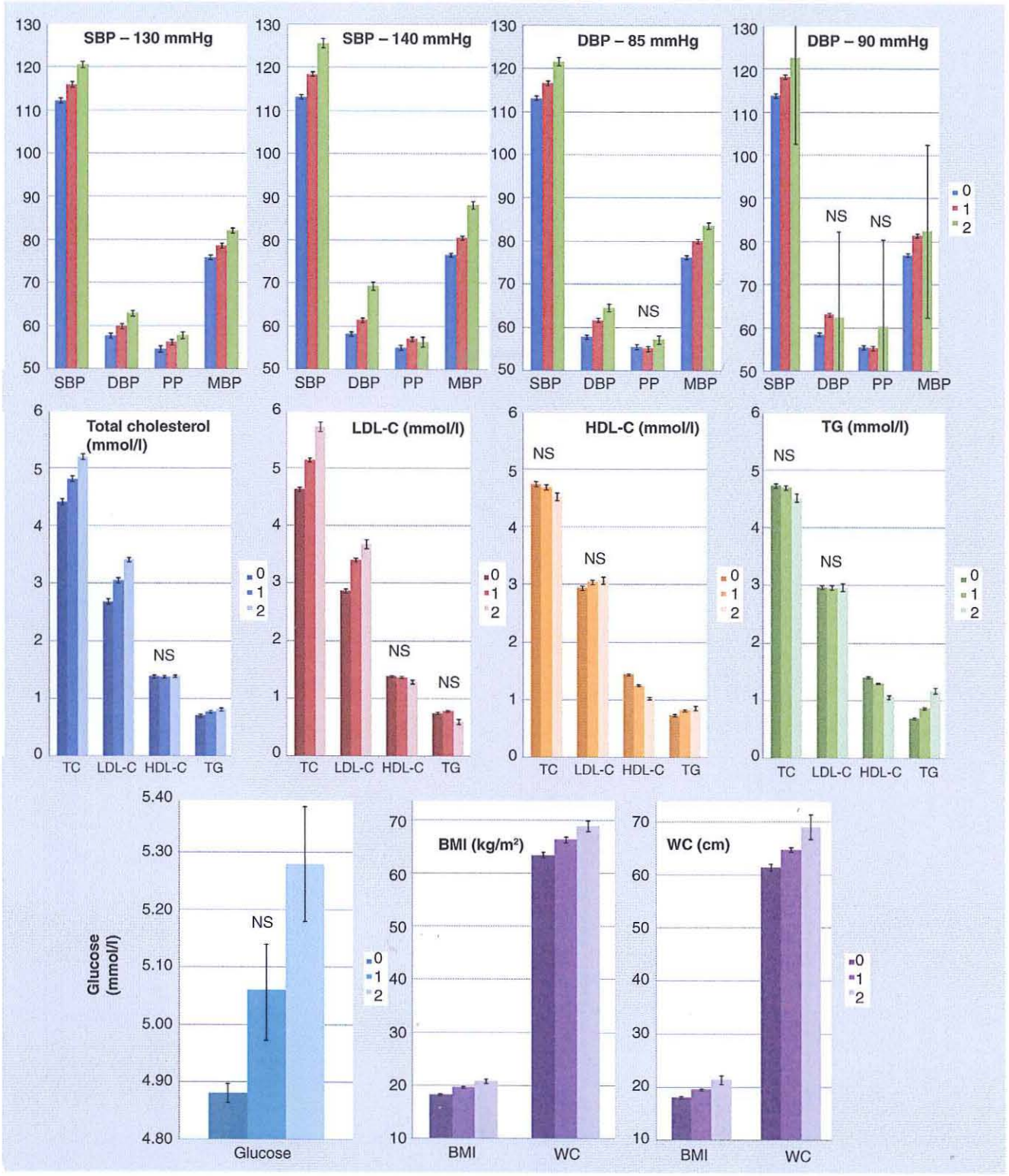


Figure 1. Investigation of 11 parental clinical parameters between children groups (0, 1 and 2) at t0. Comparisons between the three groups were all significant with $p \leq 0.001$, unless reported on the histograms as NS. Mean values in each group with 95% CI are represented.
DBP: Diastolic blood pressure; HDL-C: High-density lipoprotein cholesterol; LDL-C: Low-density lipoprotein cholesterol; MBP: Mean blood pressure; NS: Nonsignificant; PP: Pulse pressure; SBP: Systolic blood pressure; t0: First health examination; TC: Total cholesterol; TG: Triglycerides; WC: Waist circumference.

Table 2. Variations of children's parameters among groups.

Risk factors	Health examination	N of children per group			Mean ± SD of children per group			p-value group	p-value sex	p-value age	Sex × group interaction p-value	Power of analysis
		0	*1*	*2*	*0*	*1*	*2*					
SBP – preHT (mmHg)	t0	633	477	82	110.9 ± 10.6	114.6 ± 11.0	120.9 ± 11.6	0.000	0.000	0.000	0.057	1.000
	t5				117.5 ± 9.8	119.9 ± 11.1	123.6 ± 11.0	0.000	0.000	0.000	0.042	0.995
	Δ				6.5 ± 10.5	5.32 ± 11.3	2.7 ± 12.2	0.005	0.018	0.046	0.145	0.842
SBP – HT (mmHg)	t0	961	214	17	111.9 ± 10.8	117.5 ± 11.3	124.6 ± 10.9	0.000	0.002	0.000	0.449	1.000
	t5				118.3 ± 10.1	121.0 ± 11.9	124.5 ± 11.2	0.002	0.000	0.000	0.130	0.896
	Δ				6.4 ± 10.8	3.5 ± 11.7	-0.1 ± 7.4	0.000	0.029	0.047	0.503	0.965
DBP – preHT (mmHg)	t0	840	313	39	56.5 ± 10.6	60.1 ± 10.8	62.6 ± 12.8	0.001	0.027	0.000	0.111	0.934
	t5				62.4 ± 10.2	65.8 ± 1.0	67.2 ± 10.2	0.000	0.342	0.000	0.551	0.959
	Δ				5.9 ± 12.3	5.7 ± 12.4	4.6 ± 12.5	0.908	0.287	0.311	0.070	0.751
DBP – HT (mmHg)	t0	1.053	132	7	57.2 ± 10.8	61.3 ± 10.2	61.1 ± 16.2	0.021	0.001	0.000	0.006	0.703
	t5				63.1 ± 10.3	66.2 ± 9.8	66.2 ± 9.5	0.108	0.735	0.000	0.590	0.745
	Δ				5.9 ± 12.4	5.0 ± 11.3	5.1 ± 17.7	0.584	0.004	0.393	0.004	0.139
TC (mmol/l)	t0	471	565	158	4.46 ± 0.78	4.87 ± 0.76	5.21 ± 0.79	0.000	0.000	0.124	0.905	1.000
	t5				4.31 ± 0.78	4.70 ± 0.90	5.04 ± 0.87	0.000	0.000	0.000	0.925	1.000
	Δ				-0.15 ± 0.68	-0.17 ± 0.73	-0.18 ± 0.80	0.808	0.000	0.000	0.984	0.832
LDL-C (mmol/l)	t0	965	202	7	2.87 ± 0.72	3.41 ± 0.77	3.68 ± 0.96	0.000	0.121	0.781	0.774	1.000
	t5				2.52 ± 0.74	2.98 ± 0.89	3.22 ± 0.78	0.000	0.017	0.000	0.293	1.000
	Δ				-0.35 ± 0.60	-0.43 ± 0.69	-0.47 ± 0.45	0.192	0.370	0.000	0.620	0.849
HDL-C (mmol/l)	t0	880	278	30	1.48 ± 0.35	1.27 ± 0.30	1.05 ± 0.26	0.000	0.735	0.000	0.396	1.000
	t5				1.56 ± 0.39	1.40 ± 0.34	1.21 ± 0.22	0.000	0.000	0.544	0.480	1.000
	Δ				0.08 ± 0.36	0.13 ± 0.31	0.17 ± 0.22	0.023	0.000	0.646	0.985	0.962
TG (mmol/l)	t0	975	209	10	0.69 ± 0.34	0.87 ± 0.49	1.17 ± 0.58	0.000	0.119	0.000	0.019	1.000
	t5				0.90 ± 0.42	1.02 ± 0.46	1.05 ± 0.33	0.002	0.729	0.000	0.849	0.884
	Δ				0.21 ± 0.43	0.15 ± 0.50	-0.12 ± 0.60	0.004	0.101	0.001	0.019	0.850
Glucose (mmol/l)	t0	1.154	38	2	4.89 ± 0.37	5.05 ± 0.38	5.28 ± 0.41	0.013	0.008	0.447	0.252	0.758

Δ: Variation between two health examinations; DBP: Diastolic blood pressure; HDL: High-density lipoprotein; HT: Hypertension thresholds used to gather groups; LDL: Low-density lipoprotein; n: Sample size; preHT: Pre-hypertension thresholds used to gather groups; SBP: Systolic blood pressure; t0: Values at first health examination; t5: Values at second health examination; TC: Total cholesterol; TG: Triglycerides; WC: Waist circumference.

Table 2. Variations of children's parameters among groups (cont.).

Risk factors	Health examination	N of children per group			Mean ± SD of children per group			p-value group	p-value sex	p-value age	Sex × group interaction p-value	Power of analysis
		0	*1*	*2*	*0*	*1*	*2*					
	t5				4.77 ± 0.60	4.79 ± 0.60	4.83 ± 0.64	0.954	0.219	0.955	0.790	0.757
	Δ				-0.12 ± 0.67	-0.26 ± 0.64	-0.46 ± 0.23	0.366	0.628	0.000	0.307	0.726
BMI (mmol/l)	t0	441	537	215	17.6 ± 2.5	19.4 ± 3.2	20.2 ± 3.7	0.000	0.690	0.000	0.694	1.000
	t5				19.9 ± 2.4	21.6 ± 3.3	22.6 ± 4.0	0.000	0.143	0.000	0.164	1.000
	Δ				2.3 ± 1.6	2.2 ± 2.1	2.4 ± 2.6	0.700	0.002	0.001	0.158	0.808
WC (mmol/l)	t0	196	858	38	61.5 ± 6.3	64.8 ± 8.3	69.1 ± 8.5	0.000	0.000	0.000	0.265	0.999
	t5				67.4 ± 6.2	71.3 ± 8.3	74.8 ± 9.3	0.000	0.000	0.000	0.302	1.000
	D				5.7 ± 5.0	6.4 ± 5.7	5.6 ± 8.5	0.163	0.000	0.010	0.053	0.803

Δ: Variation between two health examinations; DBP: Diastolic blood pressure; HDL: High-density lipoprotein; HT: Hypertension thresholds used to gather groups; LDL: Low-density lipoprotein; n: Sample size; preHT: Pre-hypertension thresholds used to gather groups; SBP: Systolic blood pressure; t0: Values at first health examination; t5: Values at second health examination; TC: Total cholesterol; TG: Triglycerides; WC: Waist circumference.

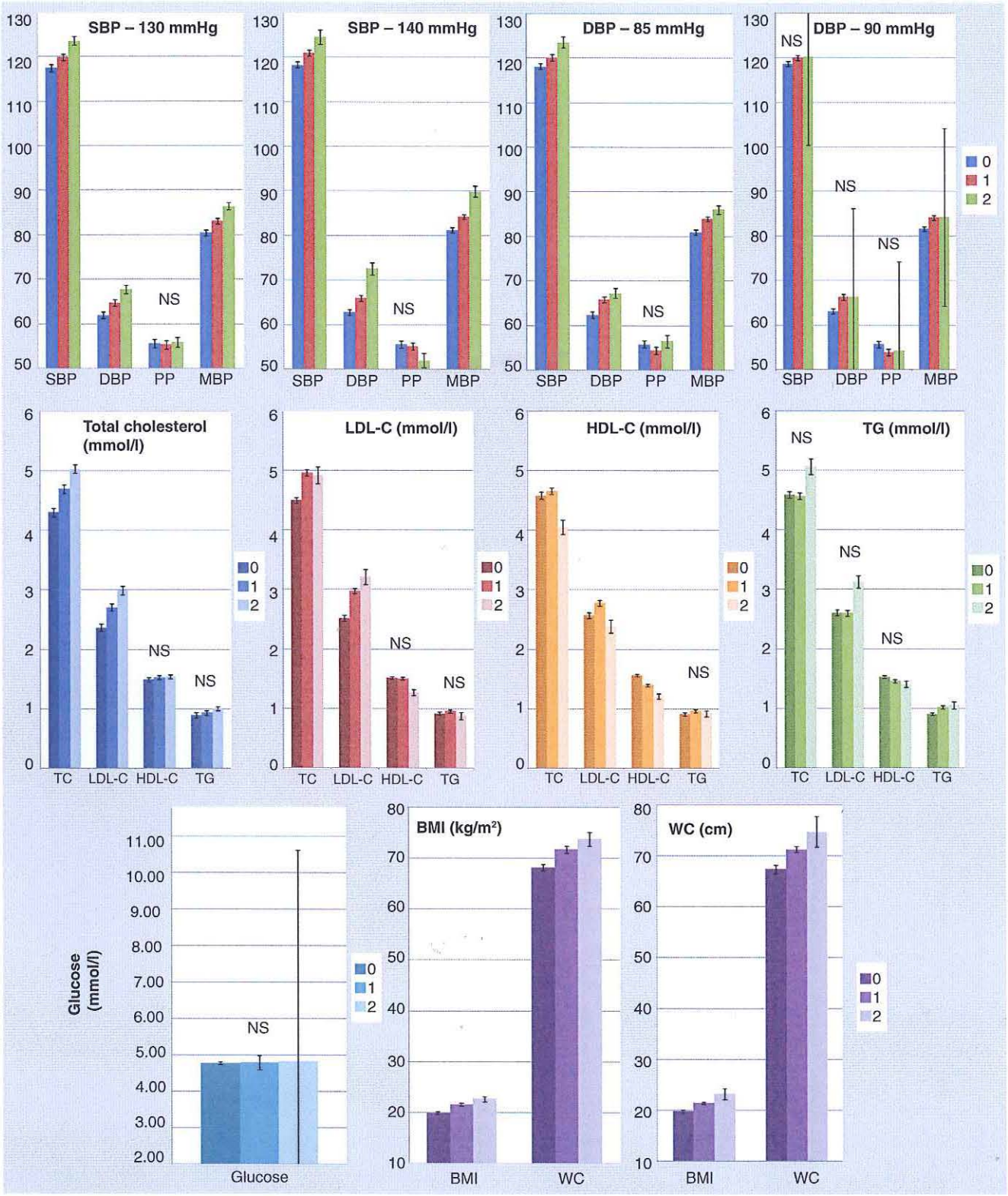


Figure 2. Investigation of 11 parental clinical parameters between children groups (0, 1 and 2) at t5. Comparisons between the three groups were all significant with $p \leq 0.001$, unless reported on the histograms as NS. Mean values in each group with 95% CI are represented.

DBP: Diastolic blood pressure; HDL-C: High-density lipoprotein cholesterol; LDL-C: Low-density lipoprotein cholesterol; MBP: Mean blood pressure; NS: Nonsignificant; PP: Pulse pressure; SBP: Systolic blood pressure; t5: Second health examination; TC: Total cholesterol; TG: Triglycerides; WC: Waist circumference.

which is in progress and includes ten European countries, attempts to assess coronary mortality using several prominent dietary hypotheses.

It is also important to note that the validity of BMI, reflecting composite phenotype, varies across the life curve according to gender as a relevant cardiovascular risk marker. Some recent evidence suggests that adiposity is a better indicator of at-risk obesity [18]. Moreover, this trait should be closely monitored since BMI changes dynamically during childhood and adolescence [19].

Conclusion

Current prevention strategies are primarily based on adult populations and there is a relative consensus that individuals should be targeted at 20, 50 and 70 years of age [20,21]. Indeed, international guidelines did not address prevention in children, whereas pathological studies have demonstrated that both the presence and extent of atherosclerotic lesions at autopsy after the unexpected death of children and young adults correlate positively and significantly with LDL-C, TG, SBP, DBP and BMI [22,23]. The Bogalusa Heart Study underlined the familial nature of the insulin resistance syndrome [24] and the early natural history of Type 2 diabetes [25].

Since the cardiovascular disorder process begins in childhood and adolescence, early prevention in childhood is critically important. Thus, prevention strategies should be focused on 'at risk' families and interventions should be undertaken in young individuals, a group for whom primary prevention should hold the most promise [26].

A recent study performed on 2720 German child–parent pairs affirmed that risk screening in children may serve for early prevention in high-risk families by targeting young parents [27]. On the contrary, our methodology supposes that children requiring early-risk screening should be targeted before any clinical manifestations of cardiovascular risk factors according to their parents' parameters.

Future perspective

With regard to public health practice, screening silent risk factors using the familial environment could be a cost-effective method of identifying children and adolescents needing intensive and early preventive care in the ambition to reduce the future burden of adult cardiovascular disease.

Financial & competing interests disclosure

The authors have no relevant affiliations or financial involvement with any organization or entity with a financial interest in or financial conflict with the subject matter or materials discussed in the manuscript. This includes employment, consultancies, honoraria, stock ownership or options, expert testimony, grants or patents received or pending, or royalties.

No writing assistance was utilized in the production of this manuscript.

Executive summary

- The persistent progression of cardiovascular disease morbidity and mortality in developing countries, in spite of the general effort in prevention, implies a better focalization is required in different target populations.
- Cardiovascular disease processes begin early in childhood through a synergy of genetic factors and environmental exposure, all depending on the parents' risk status.
- Parental variations of most cardiovascular risk markers (blood pressure, lipids and corpulence) influence their children's cardiovascular disease risk.
- Silent risk factors in children should be screened according to their parental risk status in order to initiate early prevention strategies.

Bibliography

Papers of special note have been highlighted as:
- ▪ of interest
- ▪▪ of considerable interest

1 Li S, Chen W, Srinivasan SR *et al.*: Childhood cardiovascular risk factors and carotid vascular changes in adulthood: the Bogalusa Heart Study. *JAMA* 290, 2271–2276 (2003).

▪▪ **Introduces the major hypothesis on which we base our study: prevention strategies should focus on children in order to avoid the development of cardiovascular disorders in adulthood.**

2 Burke GL, Cresanta JL, Shear CL, Miner MH, Berenson GS: Cardiovascular risk factors and their modification in children. *Cardiol. Clin.* 4(1), 33–46 (1986).

3 Visvikis-Siest S, Siest G: The STANISLAS Cohort: a 10-year follow-up of supposed healthy families. Gene-environment interactions, reference values and evaluation of biomarkers in prevention of cardiovascular diseases. *Clin. Chem. Lab. Med.* 46(6), 733–747 (2008).

4 Odell PM, Anderson KM, Kannel WB: New models for predicting cardiovascular events. *J. Clin. Epidemiol.* 47(6), 583–592 (1994).

5 Bastuji-Garin S, Deverly A, Moyse D *et al.*: The Framingham prediction rule is not valid in a European population of treated hypertensive patients. *J. Hypertens.* 20(10), 1973–1980 (2002).

6 Cambien F, Tiret L: Genetics of cardiovascular diseases: from single mutations to the whole genome. *Circulation* 116(15), 1714–1724 (2007).

7 Chobanian AV, Bakris GL, Black HR *et al.*: The seventh report of the Joint National Committee on prevention, detection, evaluation, and treatment of hight blood pressure: the JNC7 report. *J. Am. Med. Assoc.* 289, 2560–2572 (2003).

▪ **We used this article to define parental risk status for blood pressure.**

8 Stamler J, Daviglus ML, Garside DB, Dyer AR, Greenland P, Neaton JD: Relationship of baseline serum cholesterol levels in 3 large cohorts of younger men to long-term coronary, cardiovascular, and all-cause mortality to the longevity. *JAMA* 284, 311–318 (2000).

▪ **We used this article to define parental risk status for lipid traits.**

9 Expert Panel on Detection, Evaluation, and Treatment of High Blood Cholesterol in Adults: 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* 285, 2486–2497 (2001).

▪ **We used this report to define parental risk status for glucose.**

10 Wierzbicki AS, Ganotakis ES, Mikhailidis DP: Shape of the nations survey and attitudes tocardiometabolic risk. *Curr. Med. Res. Opin.* 23(1), 25–28 (2007).

11 Largo-Jenssen T, Lo Fo Wong S, van den Muijsenbergh M: The importance of gender in health problems. *Eur. J. Gen. Pract.* 14(Suppl. 1), 33–37 (2008).

12 Marteau JB, Zaiou M, Siest G, Visvikis-Siest S: Genetic determinants of blood pressure regulation. *J. Hypertens.* 23, 2127–2143 (2005).

13 Lawson EA, Miller KK, Mathur VA *et al.*: Hormonal and nutritional effects on cardiovascular risk markers in young women. *J. Clin. Endocrinol. Metab.* 92(8), 3089–3094 (2007).

14 Meyre D, Delplanque J, Chèvre JC *et al.*: Genome-wide association study for early-onset and morbid adult obesity identifies three new risk loci in European populations. *Nat. Genet.* 41(2), 157–159 (2009).

15 Scuteri A, Sanna S, Chen WM *et al.*: Genome-wide association scan shows genetic variants in the *FTO* gene are associated with obesity-related traits. *PloS Genet.* 3, E115 (2007).

16 Perusse L, Rice T, Chagnon YC *et al.*: A genome-wide scan for abdominal fat assessed by computed tomography in the Quebec Family Study. *Diabetes* 50, 614–621 (2001).

17 Danesh J, Saracci R, Berglund G *et al.*: EPIC-Heart: the cardiovascular component of a prospective study of nutritional, lifestyle and biological factors in 520,000 middle-aged participants from 10 European countries. *Eur. J. Epidemiol.* 22, 129–141 (2007).

18 Poirier P: Adiposity and cardiovascular disease. what is the best indicator of 'at risk' obesity? *Med. Sci. (Paris)* 24, 1096–1098 (2008).

19 Funatogawa I, Funatogawa T, Yano E: Do overweight children necessarily make overweight adults? Repeated cross sectional annual nationwide survey of Japanese girls and women over nearly six decades. *BMJ* 337, A802 (2008).

20 Pearson TA, Blair SN, Daniels SR *et al.*: AHA Guidelines for primary prevention of cardiovascular disease and stroke: 2002 update: consensus panel guide to comprehensive risk reduction for adult patients without coronary or other atherosclerotic vascular diseases. American Heart Association Science Advisory and Coordinating Committee. *Circulation* 106, 388–391 (2002).

21 Wensing M, Ludt S, Campbell S, Van Lieshout J, Volbracht E, Grol R: European Practice Assessment of Cardiovascular Risk Management (EPA Cardio): protocol of an international observational study in primary care. *Implement. Sci.* 4(1), 3 (2009).

▪ **Exposes current practice in primary prevention strategies and the lack of interest in children.**

22 Fernandez-Britto JE, Wong R, Contreras D, Nordet P, Sternby NH: Pathomorphometrical characteristics of atherosclerosis in youth. A multinational investigation of WHO/World Heart Federation (1986–1996), using atherometric system. *Nutr. Metab. Cardiovasc. Dis.* 9, 210–219 (1999).

23 McGill HC Jr, McMahan CA, Herderick EE, Malcom GT, Tracy RE, Strong J: Origin of atherosclerosis in childhood and adolescence. *Am. J. Clin. Nutr.* 72, 1307–1315 (2000).

24 Chen W, Srinivasan SR, Elkasabany A, Berenson GS: The association of cardiovascular risk factor clustering related to insulin resistance syndrome (syndrome X) between young parents and their offspring: the Bogalusa Heart Study. *Atherosclerosis* 145, 197–205 (1999).

25 Nguyen QM, Srinivasan SR, Xu JH, Chen W, Berenson GS: Changes in risk variables of metabolic syndrome since childhood in pre-diabetic and Type 2 diabetic subjects: the Bogalusa Heart Study. *Diabetes Care* 31(10), 2044–2049 (2008).

26 Jenssen BP, Jacobson MS: Screening adolescents for lipid disorders: what is the best approach? *Adolesc. Med. State Art Rev.* 19(3), 507–520 (2008).

27 Schwandt P, Bischoff-Ferrari HA, Staehelin HB, Haas GM: Cardiovascular risk screening in school children predicts risk in parents. *Atherosclerosis* (2009) (Epub ahead of print).

▪▪ **Authors used an opposite methodology – screening at-risk families using the childrens' risk status. We both conclude that there is a need for precocious primary prevention strategies, but screening parents would be a more effective early prevention measure to target silent factors risk increasing throughout childhood.**

Website

101 AFSSAPS : Prise en charge thérapeutique du patient dyslipidémique – recommandations (2005) www.afssaps.fr/Infos-de-securite/Recommandations-de-bonne-pratique/Prise-en-charge-du-patient-dyslipidemique

▪ **We used this report to define parental risk status for lipid traits.**

102 Visvikis-Siest S, Samara A, Meyre D, Herbeth B, Froguel P: Significant relationship between *FTO* variants and 5-year BMI changes in children and adolescents of the STANISLAS Cohort www.plosgenetics.org/annotation/listThread.action?inReplyTo=info%3Adoi%2F10.1371%2Fannotation%2F916bc243-b8fb-4b9c-9ac4-c89f3501c55c&root=info%3Adoi%2F10.1371%2Fannotation%2F916bc243-b8fb-4b9c-9ac4-c89f3501c55c

PUBLICATION N°4

Philippe Froguel*, **Ndeye Coumba Ndiaye***, Nabila Bouatia-Naji, Cécile Lecoeur, Gérard Siest, Bernard Herbeth, Jean-Brice Marteau, Rosa-Maria Rodriguez-Guéant, Marion Marchand, Éliane Albuisson, Anita Morandi, Yann Labrune, Vincent Vatin, Claudio Maffeis, Paul Elliott, Marjo-Riita Jarvelin, George V Dedoussis, David Meyre, Sophie Visvikis-Siest. GWAS LDL and total cholesterol associated rs2000999 in *Haptoglobin Related Protein* explains 42% of the Haptoglobin Levels Genetic Variability in children: a putative biological link between inflammation and cardio-vascular risk factors. 2010 [soumis]

Une étude d'association pangénomique initialement effectuées sur 631 enfants non apparentés issus de la cohorte STANISLAS et suivie par une réplication chez trois cohortes Européennes a permis d'identifier le SNP rs2000999 dans le gène *Haptoglobin Related Protein* expliquant une grande part de l'héritabilité génétique de l'haptoglobine ($P_{méta-analyse}=8.1\times10^{-59}$, 42 % de l'héritabilité génétique).

Ce même variant a récemment été associé chez l'adulte au cholestérol total et au LDL-C dans de larges GWAS. Nous avons répliqué ces résultats chez 8 946 enfants Européens et ainsi mis en évidence un lien putatif entre inflammation et ces facteurs de risque cardiovasculaire connus.

Contribution :

Cette étude pangénomique porte sur l'haptoglobine dont l'investigation a été initiée par l'EA4373. Nous avons effectué les analyses au sein de la 'discovery cohort' STANISLAS et sélectionné les variants génétiques à répliquer par SNPlex. Nous avons également effectué les analyses au sein des populations de réplication en double aveugle avec un collaborateur statisticien de Lille. Deux méthodologies distinctes et deux logiciels différents avaient été utilisés pour ce faire. La méta-analyse a été effectuée à Lille et nous avons réalisé les régressions conditionnées nous permettant de déterminer l'indépendance des variants génétiques retenus.

Nous avons également construit les tables et rédigé la première version du manuscrit et les 'supplementary data'.

Notez que l'haptoglobine est un trait rare qu'il a donc fallu doser dans les populations de réplication sélectionnées. Nous avons participé à la préparation des échantillons à doser ainsi que de ceux à génotyper à Lille.

GWAS LDL and total cholesterol associated rs2000999 in *Haptoglobin Related Protein* explains 42% of the Haptoglobin Levels Genetic Variability in children: a putative biological link between inflammation and cardio-vascular risk factors.

Philippe Froguel[1,2,3]*¥, Ndeye Coumba Ndiaye[4]*, Nabila Bouatia-Naji[1,2], Cécile Lecoeur[1,2], Gérard Siest[4], Bernard Herbeth[4], Jean-Brice Marteau[4], Rosa-Maria Guéant-Rodriguez[5], Marion Marchand[1,2], Éliane Albuisson[4], Anita Morandi[1,2,6], Yann Labrune[1,2], Vincent Vatin[1,2], Claudio Maffeis[6], Paul Elliott[7], Marjo-Riita Jarvelin[7,8,9], George V Dedoussis[10], David Meyre[1,2], Sophie Visvikis-Siest[4¥]

1. UMR 8199, CNRS, Institut Pasteur de Lille, F-59019, France
2. University Lille Nord de France, F-59000, France.
3. Department of Genomics of Common Disease, School of Public Health, Imperial College London, Hammersmith Hospital, W12 0NN, London, UK.
4. "Cardio-vascular Genetics" Research Unit, EA4373, Université Henri Poincaré - Nancy 1, F-54000 Nancy, France.
5. Inserm, U954, Genetics and Environmental Risk Exposure, Faculty of Medicine, Nancy-Université Vandoeuvre-les-Nancy, F-54511, France.
6. Regional Centre for Juvenile Diabetes, Obesity and Clinical Nutrition, Verona, Italy
7. Department of Epidemiology and Public Health, Imperial College London, W2 1PG London, UK.
8. Institute of Health Sciences, University of Oulu, Finland
9. National Institute for Health and Welfare, Oulu, Finland, Biocenter Oulu, University of Oulu, Finland.
10. Department of Nutrition–Dietetics, Harokopio University, Athens, Greece.

* Equal first authors

¥ Corresponding authors

Philippe Froguel

Genomic Medicine, Imperial College London, Hammersmith Hospital, Room E306, Burlington-Danes building, Du Cane Road, London, W12 0NN , United Kingdom

Tel.:+442075946520

p.froguel@imperial.ac.uk

and

Sophie Visvikis-Siest

"Cardiovascular Genetics" Research Unit, EA4373, Université Henri Poincaré - Nancy 1, F-54000 Nancy, France.

Tel: +33383682184 / +33607602569

Fax: +33383321322

Sophie.Visvikis-Siest@inserm.fr

Abstract

A GWAS initially conducted in 631 French children followed by replication in three European pediatric samples, identified rs2000999 in the *Haptoglobin Related Protein* gene as a strong genetic determinant of Haptoglobin levels ($P_{overall}=8.1\times10^{-59}$, 42 % of its genetic variability).
Recent GWAS meta-analyses have concluded that rs2000999 associates with cholesterol levels in adults, a finding we replicated in 8,946 European children. This study provides a putative biological link between inflammation and cardio-vascular risk factors.

Human haptoglobin (Hp) is an acute inflammatory protein, essentially synthesized by the liver, which is up-regulated by cytokines as the other acute phase proteins[1]. Hp binds hemoglobin (Hb) released from the erythrocytes[2]. Hp-Hb complexes are further eliminated from the circulation, mainly by the liver Kupffer's cells, preventing the generation of oxygen reactive species[2]. Hp plasma levels can increase up to 200% during the ten days following acute inflammation onset[1].

In a prospective study (11.8 years follow-up), Holme et al[3] demonstrated that Hp is a significant risk factor of acute myocardial infarction, stroke and heart failure in 342,125 individuals from the Apolipoprotein Mortality RISk study (AMORIS). Very recently, the same authors[4] stressed the importance of a combination of commonly available inflammatory markers, Hp included, as predictors of major cardio-vascular events in a subset of 65,050 individuals, suggesting that these routinely assessed inflammation markers should be incorporated in daily medical practice to evaluate cardio-vascular risk.

Hp is composed of two pairs of an α and a β chains. A duplication of 1.7 kb in the Hp gene modifies Hp α-chain polypeptides causing three common gel migration protein profiles: Hp1-1, Hp2-2 and Hp2-1[5]. Several studies have shown that diabetic individuals with the Hp2-2 phenotype were at four to five times greater risk of cardio-vascular events[6] and prone to carotid atherosclerosis[7]. The Hp2-2 phenotype was associated with reduced clearance of macrophage-Hp-Hb complex, favoring iron deposition, oxidative stress, and active macrophage accumulation in the atherosclerotic plaque[8]. A role for this Hp polymorphism was also proposed in the development of many other human diseases such as malaria, tuberculosis and HIV[9].

We conducted a genome-wide association study (GWAS) in 631 children of the STANISLAS Family Study (SFS) cohort of French European origin, to identify DNA variations contributing to the inter-individual variability of Hp levels and to assess their putative associations with cardio-vascular risk factors. The GWAS strongest signal was located in a 218-kb linkage disequilibrium (LD) block on chromosome 16 that includes both the *Haptoglobin* (*Hp)* and the *Haptoglobin Related Protein (HPR)* genes. *HPR* shares more than 90% of nucleotide-sequence homology with *HP* and forms a 30-kb cluster with this latter gene on chromosome 16q22.3[10].

With the squared root transformed haptoglobin measurement adjusted for gender, age and lnBMI, under the additive model, we identified in this region two genome-wide significant association signals distant of 90-kb: rs2000999 (β=0.123 with G as effect allele; P=6.32x10^{-13}) and rs10492825 (β=0.0876 with A as effect allele; P=5.50$\times 10^{-08}$) (Table 1). rs2000999 and rs10492825 polymorphisms display a moderate LD (r^2=0.48, HapMap CEU release 27). In order to assess the redundancy between these two signals, we run conditional regression analyses for both SNPs adjusted for each other and

found that rs2000999 is driving alone the association observed on the *HP/HPR* locus (rs2000999 $P_{adjusted\ by\ rs10492825}=1.9\times10^{-7}$, rs10492825 $P_{adjust\ by\ 200099}=0.91$).

We further replicated the association between rs2000999 and Hp levels in three additional European cohorts including 2 pediatric cohorts: the GENDAI study from Greece and a subset of obese children from East of France (N_{tot}=1,435); and a familial subset (N_{tot}=3,167) of the SFS ($P_{replication}=3.5\times10^{-41}$, $P_{overall}= 8.09x10^{-59}$, Table 1). We estimated that adjusted squared root transformed Hp levels increase by 0.09 g/L per rs2000999 allele A in average.

We determined that 24.9% of the Hp variance components was under polygenic control in 656 nuclear families (2,680 individuals) from the STANISLAS cohort and that rs2000999 is the major genetic determinant of Hp levels explaining alone 10.5% of its global variance and thus controlling 42% of the polygenic variability of this trait (Supplementary Table 2).

SNP rs2000999 is located in the third intron of the *HPR* gene, 2.2 kb downstream of the *HP* gene. It is in the vicinity of the well known and frequent Hp1 polymorphism, causing three protein gel migration profiles (Hp1-1, Hp1-2 and Hp2-2). This polymorphism generates a short (Hp1:83 amino acids) and a long (Hp2:142 amino acids) α chain. Hp2 is caused by a duplication of 59 α chain amino acid residues resulting to an intragenic duplication of 1.7 kb[5]. To assess the degree of independency of rs2000999 from the Hp1 polymorphism, we genotyped this latter in a subset of 341 children from the GWAS first stage children sample. We found no evidence for high LD neither between Hp1 and rs2000999 (r^2=0.135) nor with the other SNPs that were genotyped by the Illumina array within the 218 kb LD block that includes both Hp1 and rs2000999 ($0.001<r^2<0.137$; N=31 SNPs). In this subsample, we found that Hp1 polymorphism (MAF=0.46) and rs2000999 (MAF=0.20) are highly associated with Hp levels, as expected ($P=4x10^{-7}$ and $P=1x10^{-7}$ respectively). When both variants were included in the same regression model, we found that both variants are significantly and independently contributing to the association (P_{HP1}=0.001 and $P_{rs2000999}=5x10^{-5}$). These data indicate that the association of rs2000999 we identified here is novel and not redundant with the well known Hp1 polymorphism.

In addition to Hp1, the homozygosity for a very rare deletion in Caucasians (~0.1%)[11] of more than 20 kb in this region causes the Hp0 anhaptoglobinemia phenotype with complete absence of serum Hp[5]. Importantly, rs2000999 is located in the region of the *HPR* covered by this deletion which suggests that the DNA sequence around this SNP is critical for Hp levels.

SNP rs2000999 has been recently reported to associate with total cholesterol in 4,200 adults from the EUROSPAN consortium[12] and to total cholesterol and LDL-cholesterol in more than 100,000 adults of European and non-European ancestry[13], with a p-value that reaches genome-wide significant levels ($P=3x10^{-24}$). This study did not assess the effect of rs2000999 in pediatric cohorts but

prompted us to do so in five independent pediatric cohorts gathering 8,946 European children. Total cholesterol was ln-transformed and we normalized the LDL cholesterol by computing the square root. Both measurements were adjusted for gender, age and ln-BMI and we tested the association under an additive model. Our data showed that rs2000999 was associated with total cholesterol.($P_{\text{total cholesterol}}$=0.002, Table 2) and LDL cholesterol (p=0.0008, Table2).

Consistent and compelling evidence has already demonstrated the association between raised plasma levels of inflammatory markers with incidence of cardiovascular diseases, impaired cardiac function and worse prognosis. Among these inflammatory markers Hp has already been consistently linked to cardiovascular diseases and very recently it was demonstrated to give additive predictive information for risk of major cardiovascular events[4]. SNP rs2000999 is located on the *HP/HPR* cluster and influences lipid levels. This SNP thus represents a major biomarker linking inflammation via Hp and cardio-vascular risk. In the light of our findings, we hypothesize that rs2000999 effect on LDL and total cholesterol may be mediated by the regulatory function of *HPR* on circulating Hp levels. Hp binds apolipoprotein A-I[14] and protects the Apo A-I effector domain for lecithin-cholesterol acyltransferase against oxidative stress, and consequently, impairs the high-density lipoprotein function. Hp also binds the Apo E and influences cholesterol esterification[15]. These structural characteristics confer to Hp a possible role in the reverse transport of cholesterol between peripheral cells and the liver for degradation which, with endothelial repair by progenitor cells and plaque neovascularization, is a preservation system of vessel wall integrity.

In summary, we identified rs2000999 as a strong contributor to Hp genetic heritability (explaining 42% of the global genetic heritability of Hp). We also demonstrated in childhood its role in the modulation of lipid levels, extending adults data. This study provides therefore a putative biomarker linking inflammation to major cardio-vascular risk factors. In clinical practice, rs2000999 genotyping could be used to correct Hp values for assessment of risk in different pathologies in which Hp is implicated, especially at younger age for early identification of individuals at cardio-vascular risk.

Acknowledgements

The authors would like to thank the study participants and their families for their time and effort to help create the present study. They would like to thank also all the field investigators for the recruitment and examinations of the populations involved in this study: the STANISLAS Family Study supported by the “Caisse Nationale d’Assurance Maladies des Travailleurs Salariés” (CNAM), the “Institut National de la Santé et de la Recherche Médicale” (INSERM), the “Région Lorraine”, the “Communauté Urbaine du Grand Nancy”, the Henri Poincaré University of Nancy 1, and the “BioIntelligence” project; the Verona and French Obese children cohorts supported by funding from

“Contrat Plan Etat Région” entitled “Genetic dissection of polygenic obesity using GWA approach” and the “Fond Européen pour le Développement Régional”; the NFBC86 which received financial support from the Academy of Finland (project grants 104781, 120315 and Center of Excellence in Complex Disease Genetics), the University Hospital Oulu, Biocenter, University of Oulu, Finland, the European Commission (EURO-BLCS, Framework 5 award QLG1-CT-2000-01643), the NHLBI grant 5R01HL087679-02 through the STAMPEED program (1RL1MH083268-01), NIH/NIMH (5R01MH63706:02), the ENGAGE project and grant agreement HEALTH-F4-2007-201413, and the Medical Research Council (studentship grant G0500539); and the GENDAI cohort funded by a research grant from Coca-Cola Hellas.

We thank Sidonie Vivequin for the double lecture of all gels that were generated for the genotyping of Hp1, Jean-Claude Chèvre and Franck De Greave for bionformatics support, Marianne Deweirder and Michèle Pfister for DNA preparation. We are also grateful to SIEMENS for supporting the Hp levels measurements in the replication populations.

We acknowledge the INSERM, SV-S, BH, DM and NB-N employer.

Table 1 – Discovery and replication of rs2000999 association data for Hp levels

SNP	Gene	Population	Stage	N	MAF	11			12			22			β	SE	raw P	P adjusted with rs10492825	
rs2000999	HPR	SFS - unrelated children	Discovery	601	0.21	0.736	±	0.406	0.532	±	0.321	0.361	±	0.180	-0.123	0.017	6.32E-13	1.95E-07	*
		SFS - families	Replication	3,145	0.20	1.056	±	0.467	0.819	±	0.445	0.549	±	0.330	-0.138	0.007	1.17E-75		
		Obese children	Replication	1,015	0.22	1.261	±	0.492	1.064	±	0.455	1.002	±	0.534	-0.088	0.013	1.20E-11		
		GENDAI	Replication	419	0.25	0.811	±	0.450	0.813	±	0.558	0.731	±	0.319	-0.021	0.021	0.306		
		Replication sample	Replication	4,579	0.22												3.49E-41		
		Overall meta-analysis	Replication	5,180													8.09E-59		

Table 2 – Association of rs2000999 with lipid traits

			Total cholesterol			LDL-cholesterol		
Lipid traits	N	MAF	β	SE	p-value	β	SE	p-value
SFS	1,644	0.202	-0.003	0.008	0.73	0.0004	0.010	0.967
Obese Children	1,015	0.216	-0.008	0.011	0.439	-0.009	0.013	0.507
NFBC86	5,310	0.186	-0.013	0.004	0.002	-0.022	0.006	8.96E-05
GENDAI	409	0.253	-0.023	0.012	0.055	-0.031	0.013	0.022
Verona Cohort	401	0.200	-0.011	0.014	0.446	-0.011	0.016	0.515
Overall meta-analysis	8,779		-0.011	0.003	0.002	-0.017	0.004	0.0008

References

1. Gabay C, Kushner I (1999) Acute-phase proteins and other systemic responses to inflammation. N Engl J Med 340:448-454
2. Okazaki T, Yanagisawa Y, Nagai T (1997) Analysis of the affinity of each haptoglobin polymer for hemoglobin by two-dimensional affinity electrophoresis. Clin Chim Acta 258:137-144
3. Holme I, Aastveit AH, Hammar N, Jungner I, Walldius G, (2009) Haptoglobin and risk of myocardial infarction, stroke, and congestive heart failure in 342,125 men and women in the Apolipoprotein MOrtality RISk study (AMORIS). Ann Med 41:522-532
4. Holme I, Aastveit AH, Hammar N, Jungner I, Walldius G (2010) Inflammatory markers, lipoprotein components and risk of major cardiovascular events in 65,005 men and women in the Apolipoprotein MOrtality RISk study (AMORIS). Atherosclerosis
5. Koda Y, Soejima M, Yoshioka N, Kimura H (1998) The haptoglobin-gene deletion responsible for anhaptoglobinemia. Am J Hum Genet 62:245-252
6. Asleh R, Blum S, Kalet-Litman S, Alshiek J, Miller-Lotan R, Asaf R, Rock W, Aviram M, Milman U, Shapira C, Abassi Z, Levy AP (2008) Correction of HDL dysfunction in individuals with diabetes and the haptoglobin 2-2 genotype. Diabetes 57:2794-2800
7. Ryndel M, Behre CJ, Brohall G, Prahl U, Schmidt C, Bergström G, Fagerberg B, Olson FJ (2010) The haptoglobin 2-2 genotype is associated with carotid atherosclerosis in 64-year old women with established diabetes. Clin Chim Acta 411:500-504
8. Levy AP, Levy JE, Kalet-Litman S, Miller-Lotan R, Levy NS, Asaf R, Guetta J, Yang C, Purushothaman KR, Fuster V, Moreno PR (2007) Haptoglobin genotype is a determinant of iron, lipid peroxidation, and macrophage accumulation in the atherosclerotic plaque. Arterioscler Thromb Vasc Biol 27:134-140
9. Carter K, Worwood M (2007) Haptoglobin: a review of the major allele frequencies worldwide and their association with diseases. Int J Lab Hematol 29:92-110
10. Maeda N (1985) Nucleotide sequence of the haptoglobin and haptoglobin-related gene pair. The haptoglobin-related gene contains a retrovirus-like element. J Biol Chem 260:6698-6709
11. Peretti de Albuquerque Wobeto V, Zaccariotto TR, de Fátima Sonati M (2008) Polymorphism of human haptoglobin and its clinical importance. Genet Mol Biol 31:602-620
12. Igl W, Johansson A, Wilson JF, Wild SH, Polasek O, Hayward C, Vitart V, Hastie N, Rudan P, Gnewuch C, Schmitz G, Meitinger T, Pramstaller PP, Hicks AA, Oostra BA, van Duijn CM, Rudan I, Wright A, Campbell H, Gyllensten U (2010) Modeling of environmental effects in genome-wide association studies identifies SLC2A2 and HP as novel loci influencing serum cholesterol levels. PLoS Genet 6:e1000798
13. Teslovich TM, Musunuru K, Smith AV, Edmondson AC, Stylianou IM, Koseki M, Pirruccello JP, et al. (2010) Biological, clinical and population relevance of 95 loci for blood lipids. Nature 466:707-713
14. Braeckman L, De Bacquer D, Delanghe J, Claeys L, De Backer G (1999) Associations between haptoglobin polymorphism, lipids, lipoproteins and inflammatory variables. Atherosclerosis 143:383-388
15. Salvatore A, Cigliano L, Carlucci A, Bucci EM, Abrescia P (2009) Haptoglobin binds apolipoprotein E and influences cholesterol esterification in the cerebrospinal fluid. J Neurochem 110:255-263

Supplementary information

GWAS LDL and total cholesterol associated rs2000999 in *Haptoglobin Related Protein* explains 42% of the Haptoglobin Levels Genetic Variability in children: a putative biological link between inflammation and cardio-vascular risk factors.

Philippe Froguel[1,2,3]*¥, Ndeye Coumba Ndiaye[4]*, Nabila Bouatia-Naji[1,2], Cécile Lecoeur[1,2], Gérard Siest[4], Bernard Herbeth[4], Jean-Brice Marteau[4], Rosa-Maria Guéant-Rodriguez[5], Marion Marchand[1,2], Éliane Albuisson[4], Anita Morandi[1,2,6], Yann Labrune[1,2], Vincent Vatin[1,2], Claudio Maffeis[6], Paul Elliott[7], Marjo-Riita Jarvelin[7,8,9], George V Dedoussis[10], David Meyre[1,2], Sophie Visvikis-Siest[4¥]

1. UMR 8199, CNRS, Institut Pasteur de Lille, F-59019, France
2. University Lille Nord de France, F-59000, France.
3. Department of Genomics of Common Disease, School of Public Health, Imperial College London, Hammersmith Hospital, W12 0NN, London, UK.
4. "Cardio-vascular Genetics" Research Unit, EA4373, Université Henri Poincaré - Nancy 1, F-54000 Nancy, France.
5. Inserm, U954, Genetics and Environmental Risk Exposure, Faculty of Medicine, Nancy-Université Vandoeuvre-les-Nancy, F-54511, France.
6. Regional Centre for Juvenile Diabetes, Obesity and Clinical Nutrition, Verona, Italy
7. Department of Epidemiology and Public Health, Imperial College London, W2 1PG London, UK.
8. Institute of Health Sciences, University of Oulu, Finland
9. National Institute for Health and Welfare, Oulu, Finland, Biocenter Oulu, University of Oulu, Finland.
10. Department of Nutrition–Dietetics, Harokopio University, Athens, Greece.

* Equal first authors

¥ Corresponding authors

Philippe Froguel

Genomic Medicine, Imperial College London, Hammersmith Hospital, Room E306, Burlington-Danes building, Du Cane Road, London, W12 0NN , United Kingdom

Tel.:+442075946520

p.froguel@imperial.ac.uk

and

Sophie Visvikis-Siest

"Cardiovascular Genetics" Research Unit, EA4373, Université Henri Poincaré - Nancy 1, F-54000 Nancy, France.

Tel: +33383682184 / +33607602569

Fax: +33383321322

Sophie.Visvikis-Siest@inserm.fr

Study populations

The STANISLAS Family Study (SFS)

The SFS is a 10-year longitudinal survey involving, since 1993-1995, 1,006 volunteer families of European ancestry whose members were free of chronic disease (cardio-vascular or cancer)[1]. The SFS samples and data are part of the Biological Resources Bank (BRC) "Interactions Gène-Environnement en Physiopathologie CardioVasculaire" (IGE-PCV) in Nancy, France. Genome-wide genotyping was performed on a subset of 685 unrelated lean children constituting the discovery cohort as described previously[2]. The 3,145 remaining individuals after quality control (1,937 adults and 1,208 children) were involved in the replication studies. Hp levels,BMI and cardio-vascular risk traits were available for all participants.

Obese Children

We studied obese children (defined as BMI > 97th percentile for age and sex according to a French cohort [3]) ascertained from 449 nuclear families with at least one obese offspring, recruited in the Paediatric Endocrine Unit of Jeanne de Flandres Hospital, Lille or through national media campaign. We analyzed 922 children with Hp, BMI and cardio-vascular risk related traits available.

The GeNe and Diet Attica Investigation (GENDAI)

The GENDAI pediatric cohort comprised children living in the Attica region of Greece [4]. From November 2005 to June 2006, 1,138 peri-adolescent children (53% girls; mean age: 11.2±0.7 years) were recruited from randomly selected elementary schools of Attica. This research study was approved by the Institutional Review Board of Harokopio University, the Greek Ministry of Education and the Ethics Committee of Harokopio University. We analyzed 401 children with Hp, BMI and cardio-vascular risk related traits available.

The Northern Finland 1986 Birth Cohort (*NFBC86)*

The NFBC86 is a prospective birth cohort including all Finnish mothers of European ancestry with children whose expected date of birth fell between July 1, 1985 and June 30, 1986 in the two northernmost provinces in Finland [5]. Clinical examination at 15–16 years follow-up was conducted between August 2001 and June 2002. All cohort members living in Finland with known address (n=9,215) were invited, and 6,798 participated (74%). We analyzed 5,285 adolescents successfully genotyped in the NFBC86 cohort with BMI and cardio-vascular risk related traits available.

The Verona cohort

Verona cohort are Italian children (43% girls; mean age: 10.75±1.22 years) recruited from the general population of Verona, Italy, whose families were randomly extracted from the registry office database of the town, and contacted by post. We analysed 401 children successfully genotyped with at least BMI total and LDL-cholesterol available.

All the populations involved were recruited in accordance with the latest version of the Declaration of Helsinki. All participants and their parents gave a written informed consent. The study protocols were approved by the local ethics committees for the protection of subjects for biomedical research.

Genotyping

Genome-wide genotypes were generated in the 685 STANISLAS children using the Illumina Human CNV370-Duo array as described previously [2]. Briefly; we used 750 ng of genomic DNA following Illumina's protocol and BeadStation genotyping solutions (Illumina), based on the GenCall software application (Illumina) to automatically cluster, call genotypes, and assign confidence scores using the GenTrain clustering algorithm (Illumina). We discarded 2,552 SNPs that had extreme Hardy-Weinberg HW disequilibrium ($P < 0.001$), low genotyping calling rates (<95%) or low minor-allele frequencies (<1%). We retained 318,237 SNPs for analysis. Genomic control λ_{GC} was 1.01.

We used the Applied Biosystems SNPlex™ technology to replicate the association of rs200099 in the SFS replication set, obese children, GENDAI, NFBC1986 and Verona cohort DNAs.

SNPlex is based on the Oligonucleotide Ligation Assay (OLA) combined with multiplex PCR target amplification (http://www.appliedbiosystems.com). Allelic discrimination was performed by capillary electrophoresis analysis using an Applied Biosystems 3730xl DNA Analyzer and GeneMapper 3.7 software. Genotyping call rate was above 95% in all populations studied and rs2000999 was in HW equilibrium ($p>0.001$).

We used the previously reported PCR based method [6] to genotype Hp1 polymorphism in a sub-sample of the 685 children (N=341) to allow LD assessment between Hp1 and common variants at the *HP/HPR* locus. Only genotypes that were concordant following a double blind genotyping call by two independent readers were retained for statistical analyses.

Haptoglobin dosage

Hp was measured by high sensitivity immunophotometry analyses with BN™II Siemens analyzer and Siemens reagents. Levels in children were low (0.65 g/L ± 0.39 in the SFS unrelated children) in accordance with reference distribution for Hp reporting significant fell of its levels during the first decade of life[7].

Statistical analyses

Heritability estimate of Hp levels in the SFS

The relative contribution of genetic (G) and environmental (E) factors to the inter-individual variability of Hp (P) was assessed on a subsample of 656 nuclear families (2,680 individuals) ascertained from the SFS. This contribution is expressed, under a pure polygenic model, in terms of variance components: $\sigma^2P = \sigma^2G + \sigma^2E$. This yields to the determination of the genetic heritability of Hp defined as the proportion of the total phenotype variance due to additive polygenic effects ($h^2G = \sigma^2G / \sigma^2P$). The environmental component is decomposed into common familial (household sharing, C) and random non familial (individual-specific factors, R) environmental effects (i.e., $\sigma^2P = \sigma^2G + \sigma^2C + \sigma^2R$). Thus, the environmental heritability of Hp is defined as $c^2=\sigma^2C / \sigma^2P$. The analysis was conducted using a multivariate normal model for pedigree analysis [8; 9] with the FISHER software and the variance components were assumed to be normally distributed with mean equal to 0 and variance equal to σ^2G, σ^2C, and σ^2R respectively. Heritability estimate was assessed unadjusted, then adjusted for age and gender. The contribution of rs2000999 and Hp1-1 polymorphism to the genetic heritability of Hp levels were assessed using the same methodology aforementioned.

The Hp polygenic variance represented 24.9% ($P < 0.001$) of total variance. Shared (within families) and random environmental variances were 13 and 62.1% respectively (supplementary table 1).

Genome-wide association and replication analyses

We carried out genome-wide association (GWA) (Hp levels) and replication for rs2000999 (Hp and all metabolic quantitative traits) analyses using linear regression models under the additive genetic model with one degree of freedom, with adjustment for age, gender and BMI using PLINK [10]. The summary statistics were combined in the meta-analyses using the inverse normal method with equal weight for each population. In this method, P values of each study are transformed into their inverse normal z score and the weighted sum, over all studies, is compared to a normal N(0,1), provided the sum of squared weights equals 1. The estimates of variants effect on Hp and their standard errors for each separate analysis were combined in the meta-analysis using the weighted inverse normal method, and the overall effect and its confidence interval were estimated using the inverse variance method implemented in 'meta.summaries' function of R RMETA package. No major heterogeneity in effects was observed ($P < 0.02$). We used mixed model to analyse the association of genetic variants with Hp levels and metabolic traits in related individuals in the SFS and the obese children.

Softwares

For general statistical analysis, we used SPSS 14.0, R statistical environment version 2.11.1 or PLINK1.07. To draw the LD map, we used Haploview software.

URLs.

PLINK1.07, http://pngu.mgh.harvard.edu/~purcell/plink/; R package 2.11.1, http://www.r-project.org/; Haploview, http://www.broadinstitute.org/haploview/haploview.

Supplementary Table 1 – Phenotypic characteristics of the studied populations

Population	Sample size male/female	Age (y) Mean [95% CI]	BMI (kg/m^2) Mean [95% CI]	Haptoglobin (g/L) Mean [95% CI]	LDL cholesterol (mmo/L) Mean [95% CI]	Total cholesterol (mmol/L) Mean [95% CI]
SFS	**890/842**	**12.95 [12.82-13.08]**	**18.82 [18.67-18.97]**	**0.71 [0.69-0.73]**	**2.96 [2.92-2.99]**	**4.69 [4.65-4.73]**
Obese Children	**489/537**	**11.07 [10.86-11.27]**	**28.24 [27.84-28.64]**	**1.18 [1.15-1.22]**	**2.74 [2.70-2.79]**	**4.47 [4.42-4.52]**
NFBC86	**2613/2697**	**16**	**21.27 [21.17-21.37]**	**NA**	**2.54 [2.52-2.56]**	**4.27 [4.24-4.29]**
VERONE	**246/189**	**10.90 [10.75-11.04]**	**18.01 [17.78-18.23]**	**NA**	**2.41 [2.40-2.47]**	**4.21 [4.15-4.28]**
GENDAI	**201/228**	**11.16 [11.10-11.23]**	**19.76 [19.46-20.07]**	**0.81 [0.76-0.85]**	**3.12 [3.06-3.17]**	**4.79 [4.73-4.86]**

Supplementary Table 2 – Heritability estimate of Hp levels in the STANISLAS Family Study – Fisher's variance component

Model	σ^2G Polygenic variance (%)	σ^2C Common environmental variance (%)	σ^2E Specific environmental variance (%)	Total variance	Alternate model χ^2 (df) P
Model 1*	0.0478 ± 0.0120	0.0250 ± 0.0065	0.1194 ± 0.0078	0.1922	
	24.9 %	**13.0 %**	**62.1 %**		
Model 2**	0.0253 ± 0.0112	0.0275 ± 0.0059	0.1224 ± 0.0076	0.1752	Model 1: 213.86 (4)
	14.4 %	**15.7 %**	**69.9 %**		p<0.001

* Unadjusted

** Adjusted on rs2000999

References

1. Visvikis-Siest S, Siest G (2008) The STANISLAS Cohort: a 10-year follow-up of supposed healthy families. Gene-environment interactions, reference values and evaluation of biomarkers in prevention of cardiovascular diseases. Clin Chem Lab Med 46:733-747
2. Bouatia-Naji N, Bonnefond A, Cavalcanti-Proenca C, Sparso T, Holmkvist J, Marchand M, Delplanque J, et al. (2009) A variant near MTNR1B is associated with increased fasting plasma glucose levels and type 2 diabetes risk. Nat Genet 41:89-94
3. Rolland-Cachera MF, Cole TJ, Sempe M, Tichet J, Rossignol C, Charraud A (1991) Body Mass Index variations: centiles from birth to 87 years. Eur J Clin Nutr 45:13-21
4. Papoutsakis C, Vidra NV, Hatzopoulou I, Tzirkalli M, Farmaki A-E, Evagelidaki E, Kapravelou G, Kontele IG, Skenderi KP, Yannakoulia M, Dedoussis GV (2007) The Gene-Diet Attica investigation on childhood obesity (GENDAI): overview of the study design. Clin Chem Lab Med 45:309-315
5. Jarvelin MR, Elliott P, Kleinschmidt I, Martuzzi M, Grundy C, Hartikainen AL, Rantakallio P (1997) Ecological and individual predictors of birthweight in a northern Finland birth cohort 1986. Paediatr Perinat Epidemiol 11:298-312
6. Koch W, Latz W, Eichinger M, Roguin A, Levy AP, Schomig A, Kastrati A (2002) Genotyping of the common haptoglobin Hp 1/2 polymorphism based on PCR. Clin Chem 48:1377-1382
7. Ritchie RF, Palomaki GE, Neveux LM, Navolotskaia O, Ledue TB, Craig WY (2000) Reference distributions for the positive acute phase serum proteins, alpha1-acid glycoprotein (orosomucoid), alpha1-antitrypsin, and haptoglobin: a practical, simple, and clinically relevant approach in a large cohort. J Clin Lab Anal 14:284-292
8. Lange K, Weeks D, Boehnke M (1988) Programs for Pedigree Analysis: MENDEL, FISHER, and dGENE. Genet Epidemiol 5:471-472
9. Lange K, Westlake J, Spence MA (1976) Extensions to pedigree analysis. III. Variance components by the scoring method. Ann Hum Genet 39:485-491
10. Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MA, Bender D, Maller J, Sklar P, de Bakker PI, Daly MJ, Sham PC (2007) PLINK: a tool set for whole-genome association and population-based linkage analyses. Am J Hum Genet 81:559-575

PUBLICATION N°5

Éliane Albuisson, Sandy Maumus, **Ndeye Coumba Ndiaye**, Bérangère Marie, Nicolas Jay, François Kholer, Gérard Siest, Sophie Visvikis-Siest. Genetic profiling in healthy subjects from the STANISLAS cohort based on 24 polymorphisms: effects on biological variables. Clin Chem Lab Med. 2008;46(1):64–72.

Les associations entre profils génétiques et données biologiques ou cliniques ne sont pas clairement établies, surtout en ce qui concerne les sujets supposés sains. Nous avons ainsi utilisé une analyse multivariée de 24 polymorphismes associés aux principales voies métaboliques des pathologies cardiovasculaires au sein de 1 551 sujets sains issus de la cohorte STANISLAS pour déterminer des profils génétiques à t_0 et t_5.
Nous avons identifié six clusters génétiques permettant une assignation correcte de 99,6% des sujets dont un cluster impliquant le gène *SELE* qui pourrait avoir un rôle protecteur et le gène *APOC3* associé à une augmentation du risque cardiovasculaire.

Contribution :
Lors de cette étude nous avons participé à la recherche bibliographique et à l'interprétation des résultats. Nous avons également fait partie de l'équipe rédactionnelle et fait le suivi des modifications demandées par les reviewers qui concernaient en particulier des analyses statistiques supplémentaires.

Clin Chem Lab Med 2008;46(1):64–72 © 2008 by Walter de Gruyter • Berlin • New York. DOI 10.1515/CCLM.2008.007

Genetic profiling in healthy subjects from the Stanislas cohort based on 24 polymorphisms: effects on biological variables

Eliane Albuisson, Sandy Maumus, Ndeye-Coumba Ndiaye, Bérangère Marie, Nicolas Jay, François Kohler, Gérard Siest and Sophie Visvikis-Siest*

INSERM U525, Nancy University, Nancy, France

Abstract

Background: The association of genetic profiles with biological or clinical assessments is not clearly established especially among apparently healthy subjects.
Methods: A multivariate statistical analysis was performed on 24 polymorphisms related to the main metabolic pathways involved in cardiovascular diseases (CVDs). They were collected among 1551 healthy subjects of the Stanislas cohort to obtain genetic profiles. Association with biological variables was then studied at baseline (t_0) and 5 years later (t_5).
Results: Six genetic clusters were identified with relevant profiles and five polymorphisms from the selectin, apolipoprotein C3 and lipoprotein lipase genes (*SELE*-98G/T, *APOC3*–3175C/G, *APOC3*–482C/T, *APOC3*–1100C/T, *LPL*–93T/G) were sufficiently characteristic to associate 99.6% of the subjects with their corresponding cluster. A 5-year follow-up showed that clinical and biological measurements in relation to CVD risk factors already differ with triglyceride ($p=0.009$ for t_0 and $p=0.005$ for t_5) and high-density lipoprotein cholesterol ($p=0.014$ for t_0 and $p=0.003$ for t_5) for these previous genetic clusters.
Conclusions: This study presents the hypothesis that *SELE* could be protective, whereas *APOC3* could be associated with risk. It remains to be seen whether these polymorphisms will be predictive of CVD events among the selected clusters of different metabolic subtypes after a 10-year follow-up.
Clin Chem Lab Med 2008;46:64–72.

Keywords: cardiovascular diseases; clusters; multivariate analysis; polymorphisms; predisposition.

Introduction

For many years, we have investigated gene-environment interactions in healthy subjects, searching for relevant genetic markers for cardiovascular risk prediction (1–3). A large number of single nucleotide polymorphisms, primarily based on case control studies, are identified in the literature, and this list must be updated regularly (4–6). Thanks to the Human Genome Project (7) and the International HapMap Consortium, comprehensive websites provide a wealth of readily accessible sequence and polymorphism data. Proteomics investigators are employing a similar approach. The simultaneous development of systems biology (8) and the accompanying systems literature analysis (9) has opened up a new era that is critical to the success of preventive medicine (10). Above all, cardiovascular diseases (CVDs) are especially complex, with multifactorial determinants in occurrence and aetiology. Factors may or may not correlate, and there is still no clear understanding of the role of genetic polymorphisms as a basis for differences and contradictions in outcome in genetic-based studies (11, 12).

The aim of this study was to identify genetic profiles among healthy subjects from the Stanislas cohort and then to study the usefulness of this profiling as a marker of clinic and/or biological differences or tendencies in the CVD domain.

A total of 24 polymorphic genes were selected from the main biological pathways involved in cardiovascular risk and disease: lipid metabolism, thrombosis, blood pressure, inflammation, adhesion and homocysteine cycle. These genes were measured in a large cohort of healthy families (13) constituted in 1993 and composed of initially apparently healthy subjects. Statistics included a combination of multivariate analysis performed using a similar strategy as described in a previous study (14).

Materials and methods

Study population

The Stanislas cohort was constituted in 1993 and designed to investigate factors related to CVD. Details of the Stanislas cohort have been previously described (1). Individuals were recruited during a detailed check-up examination, and informed consent was obtained from all participants. A total of 1551 individuals (772 men and 779 women), not under treatment with lipid-lowering, antihypertensive or anti-inflammatory drugs, were selected from the cohort. In this relatively young population, with a mean age of 40.5 years (SD 4.6 years), all women were premenopausal.

Data collection

Personal and medical history, medication, smoking habits and alcohol consumption were recorded using a standard-

*Corresponding author: Sophie Visvikis-Siest, INSERM U525, Université Henri Poincaré (Nancy 1), Faculté de Pharmacie, 30 rue Lionnois, 54000 Nancy, France
Phone: +33-3-83682184, Fax: +33-3-83321322,
E-mail: Sophie.Visvikis-Siest@nancy.inserm.fr
Received May 11, 2007; accepted July 31, 2007; previously published online November 18, 2007

ised questionnaire. Smokers were defined as those who were current smokers. Blood pressure (calculated as the mean of three measurements taken under standardised conditions, in a supine position with a sphygmomanometer) was also measured. All the biochemical measurements have been described in full in a previous paper (13). Blood samples were collected after an overnight fast. Serum glucose, cholesterol and triglyceride (TG) levels were measured using standard enzymatic methods (Merck, Darmstadt, Germany), automated on au5021 (Olympus, Rungis, France). Serum apolipoproteins AI (APOAI) and B (APOB) were determined by immunonephelometry on a Dade-Behring Nephelometer Analyser, using Dade-Behring reagents (Rueil-Malmaison, France). Serum APOE, which is a therapeutic target (15), was determined by turbidimetry and serum high-density lipoprotein cholesterol (HDL-C) by phosphotungstate precipitation on a Cobas-Mira (Roche Diagnostics, Meylan, France). Low-density lipoprotein cholesterol (LDL-C) was calculated by the Friedwald formula: LDL-C = TC–HDL-C–TG/2.18, where TC is total cholesterol. Genomic DNA was extracted from peripheral blood samples by the salting-out method (16), and then genotyping for 27 polymorphisms (Table 1) was performed by a multiplex assay, which has been described previously (17, 18).

Statistical strategy and methods

Multivariate analyses were used to avoid univariate multiple testing, to improve the statistical power and to take into account the genetic information as a whole. The ratio of the number of subjects/number of studied genetic polymorphisms (1551/27 = 57.4) gave statistical power to the analysis and made it possible to use a multivariate strategy. Given our aim, the study of genetic data is essentially descriptive and is not looking to give a general model by building an equation or suggesting coefficients. This is why the whole sample is analysed at one time and not by successive subsamples selected at random. The pertinence of the clusters was tested by a logistic regression able, in addition, to select the most discriminating genetic data. As only genetic information was considered to obtain the genetic clusters, results of multivariate analysis are not shaped by the biological results. The whole procedure took six steps: description and selection of genetic data, computation of factorial continuous data, identification of the clusters, evaluation of the pertinence and discrimination among variables, characterisation of the clusters, and biological comparisons among clusters. Analysis was performed using SPSS statistical software version 14.0 (SPSS Inc., Chicago, IL, USA).

Step 1: Description and selection of genetic data A descriptive analysis was performed on the 27 genetic polymorphisms obtained for the 1551 subjects. The upper limit of 1549 subjects out of 1551 (99.87%) represented by only one occurrence was chosen to exclude the corresponding variable. Data with a very low variability were then excluded from the study. We did not exclude genotypes that had a high probability of correlation with one another, because the choice of a threshold to identify these correlations, and consequently the selection of the corresponding data would have been arbitrary.

Step 2: Computation of factorial continuous data To gather genetic variables that were closely linked together, factorial combinations were obtained by multiple correspondence analysis [HOMALS-SPSS: homogeneity analysis by means of Alternating Least Squares. HOMALS Version 1.0 by Data Theory Scaling System Group (DTSS), Faculty of Social and Behavioural Sciences, Leiden University, The Netherlands.]. These factorial combinations decrease the dimension of the information's subspace from its initial value (equal to the initial number of variables), and consequently reduce the number of data to be used in the following step of the analysis. As each variant had three possible genotypes, except for the *APOE* with six possible genotypes, a three-dimensional and then a six-dimensional multiple correspondence analysis was performed. Therefore, we obtained two sets of results, one with three factorial combinations and one with six factorial combinations, for each of the 1551 subjects. HOMALS was performed on initial nominal variables, but generated continuous factorial data enabling a hierarchical classification.

Step 3: Identification of the clusters An ascendant hierarchical classification using the squared Euclidean distance was performed on the factorial combinations previously obtained. This classification gave a dendogram reflecting the proximities of the 1551 subjects. The optimum number of clusters was determined by the level of the cut-off in the tree structure of the dendogram. A cut-off of six clusters was chosen for each dendogram according to the main dimension of the factorial data injected in the classification. The two sets of factorial data previously obtained were separately

Table 1 Allele frequencies for the polymorphisms at the beginning of the study.

Gene	Allele (A/a)	a %	Gene	Allele (A/a)	a %
APOB	Thr71Ile	32.8	*PON1*	Gln192Arg	29.9
APOB	Arg3500Gln*	0.05	*ACE*	I/D	56.4
APOC3	–482C→T	27.1	*AGT*	Met235Thr	41
APOC3	–455T→C	37.7	*CBS*	Ile278Thr	7.7
APOC3	1100C→T	26.7	*CBS*	Gly307SER*	0.05
APOC3	3175C→G	9.2	*MTHFR*	677C→T	36.7
APOC3	3206T→G	36.8	*GPIIIa*	Leu33Pro	15.7
CETP	Ile405Val	29.1	*FGB*	–455G→A	20.6
CETP	Asp442Gly*	0.05	*FV*	Arg506Gln	2.2
LPL	–93T→G	2	*SELE*	98G→T	8.1
LPL	–39T→C*	0	*SELE*	Ser128Arg	8.1
LPL	Asp9Asn	1.8	*SELE*	Leu554Phe	4.3
LPL	Asn291Ser	1.6			
LPL	Ser447Ter	12.1			
APOE	codons (112/158)	ε2, 8.6	ε3, 78.5	ε4, 12.5	

Data are presented as percentages. *Excluded from the analysis.

used and produced two separate dendograms. The coherence between the number and the composition of the clusters obtained for these two dendograms was a condition for the continuation of the study. A reasonable number of clusters with firm aggregations of subjects were necessary to enter the following step.

Step 4: Evaluation of the pertinence of the profiling and discrimination among variables Based on the identified clusters and using the initial nominative genetic variables, a multiple logistic regression was performed to identify among the genotypes which were the most useful for discriminating between clusters. The redundant information, difficult to select a priori, due to variables that were closely linked together was then removed to test the pertinence of this profiling. The total percentage of individuals correctly classified accepted here was not less than 95%.

Step 5: Characterisation of the clusters To label each cluster, confidence intervals (CIs) calculated on the whole sample were used as references and a graphical representation was proposed. The relative frequencies of each genetic polymorphism among the 1551 subjects and the corresponding 95% CIs were calculated for each allele. The sample size was large enough to obtain precision in the estimates and to provide ranges of plausible values likely to include the unknown population parameters. Width and confidence limits of these CIs calculated on the 1551 subjects were used as references. For each cluster, relative frequencies were calculated for each genetic allele and compared to the lower and the upper boundaries of the corresponding CI previously obtained from the whole sample. If the CI contained the relative frequency, a zero value was associated with the allele (called y-value here and equal to 0). If the relative frequency was greater than the upper bound, it showed an excess of subjects, and the difference between these two values was considered as positive (positive y-value). If the relative frequency was lower than the lower bound, it showed a lack of subjects, and the difference between these two values was considered as negative (negative y-value). A chart on two axes was then created for each cluster. On the x-axis, without scale, the 22 genetic variables with their distributions are indicated as follows: homozygous for the common allele, heterozygous, homozygous for the less frequent allele. On the y-axis, in percentages, are the y-values previously computed. The higher the absolute y-value, the greater the distance from the baseline (above or below according to the y-value sign). To characterise the profile/signature of each cluster, only the genotypes for which the absolute y-value of at least one allele was 20% or more in this cluster were taken into account. Thus, a list of genotypes was established for each cluster.

Step 6: Study of clinical and biological measurements between the clusters at t_0 and t_{+5} Clinical assessments and biological measurements were compared among the clusters at inclusion and 5 years later. According to the nature and the distribution of each variable, we performed parametric (Student's t-test or analysis of variance) or non-parametric (Mann-Whitney test, Kruskal-Wallis test or Friedman's test) comparisons. A p-value <0.01 was considered statistically significant, because many variables were potentially linked.

Results

Step 1: Description and selection of genetic data

Table 1 presents the genetic polymorphisms selected at the beginning of the study. *APOB*-Arg3500Gln, *CETP*-Asp442Gly, *LPL*–39T/C and *CBS*-Gly307Ser polymorphisms with a frequency of the rare allele $\leq 1\%$ were excluded.

Steps 2 and 3: Computation of factorial continuous data and identification of the clusters

The six clusters of the three-dimension solution were of the following sizes: 1465 subjects, 44 subjects, 24 subjects, 16 subjects, 1 subject and 1 subject. The six clusters of the six-dimension solution were of the following sizes: 1467 subjects, 56 subjects, 10 subjects, 16 subjects, 1 subject and 1 subject. Results are coherent (Table 2) and eight clusters can be clearly individualised. Only the six containing more than 1 subject each were analysed, and in reference to their respective sizes they will be called C_{1453}, C_{44}, C_{16}, C_{14}, C_{12} and C_{10}.

Step 4: Evaluation of the pertinence of the profiling and discrimination among variables

The multiple logistic regression selected *SELE*–98G/T ($p<0.0001$), *APOC3*–3175C/G ($p<0.0001$), *APOC3*–482C/T ($p<0.0001$), *APOC3*–1100C/T ($p<0.0001$) and *LPL*–93T/G ($p<0.0001$). These five genetic polymorphisms were sufficient to classify 99.6% of the 1552 subjects in their group and, more importantly, 95.8% of the subjects in the smaller groups, (1451/1453 or 99.9% for C_{1453}, 8/12 or 66.7% for C_{12}, 44/44 or 100% for C_{44}, 14/14 or 100% for C_{14}, 10/10 or 100% for C_{10}, and 16/16 or 100% for C_{16}).

Table 2 Crossing of the six clusters ascendant hierarchical classification with the HOMALS three-dimensional solution and the HOMALS six-dimensional solution.

HOMALS 6 dimensions/ HOMALS 3 dimensions	1	2	3	4	5	6	
1	1453	12					1465
2		44					44
3	14		10				24
4				16			16
5					1		1
6						1	1
	1467	56	10	16	1	1	1551

Results are presented as frequencies.

Step 5: Characterisation of the clusters

The CIs computed among the whole population (1551 subjects) provide ranges of plausible values in agreement with the Hardy-Weinberg equilibrium. In the graphical representation (Figure 1), only C_{12}, C_{44}, C_{14}, C_{10} and C_{16} are shown, because all of the allele relative frequencies of C_{1453} had zero y-values. As the crossing of the results predicted (Table 2), C_{12} and C_{44} are similar, as are C_{14} and C_{10}, C_{16} is definitely distinct. The numbers and the percentages of individuals observed in a cluster for the main genotypes of its list are presented below.

C_{12} is characterised by Heterozygosity for *LPL* and *SELE*: 100% (12/12) for the *LPL* polymorphism –93T/G; 58% (7/12) for *LPL*-Asp9Asn, *SELE*- 98G/T and Ser128Arg.

C_{44} is characterised by 1/Heterozygosity for *LPL* and *CETP*: 100% (44/44) for *LPL* –93T/G and *LPL* Asp9Asn; 59% (26/44) for *CETP* Ile405Val. 2/Homozygosity for the common allele for SELE: 100% (44/44) for SELE 98G/T and Ser128Arg.

C_{14} is characterised by 1/Homozygosity for the less frequent allele for *APOC3* (except the 3175C/G polymorphism) and *APOB*: 100% (14/14) of *APOC3* –482C/T, –455T/C, –1100C/T and –3206T/G polymorphisms; 35.7% (5/14) for the *APOB* Thr71Ile polymorphism. 2/Heterozygosity for the *APOC3* gene –3175C/G polymorphism: 78.6% (11/14). Additionally, a lack of homozygosity for the less frequent allele (0/14 or 0%) for the *AGT* Met235Thr polymorphism can be noted.

C_{10} is characterised by 1/Homozygosity for the less frequent allele for *APOC3*: 100% (10/10) for –482C/T, –455T/C, –1100C/T, –3175C/G and –3206T/G. 2/Homozygosity for the common allele for the Gln192Arg polymorphism of the *PON1* gene: 80% (8/10); and for the *AGT* Met235Thr polymorphism: 60% (6/10).

C_{16} is characterised by 1/Homozygosity for the less frequent allele of the *SELE* gene: 100% (16/16) of individuals for the 98G/T and for the Ser128Arg polymorphisms. 2/Homozygosity for the common allele of the *Ile405Val* polymorphism of the *CETP* gene: 75% (12/16); and for *GPIIIa* Leu33Pro: 93.8% (15/16); 3/Heterozygosity for *APOC3* polymorphisms.

As a whole, C_{12} and C_{44} have similarities (heterozygous) for the *LPL* gene and dissimilarity for the SELE gene. C_{14} and C_{10} have similarities (homozygous for the less frequent allele) for APOC3 polymorphisms except for APOC3–3175C/G. C_{16} is characterised by individuals homozygous for the less frequent allele of SELE polymorphisms.

Step 6: Study of clinical and biological measurements between the clusters at t_0 and t_{+5}

Results are summarised in Tables 3 and 4. Table 3 displays the results obtained at t_0, whereas Table 4 displays results from 5 years later (t_{+5}). The given p-values are associated with statistical tests carried out on the six clusters. There was no difference for gender among the clusters ($p=0.22$, ns). At t_0, the six clusters differ significantly for TG ($p=0.009$), HDL-C ($p=0.014$) and are nearly significant for diastolic blood pressure (DBP, $p=0.020$), APOB ($p=0.042$), TC ($p=0.059$) and LDL-C ($p=0.059$). At t_{+5}, the six clusters differ significantly for TG ($p=0.005$), HDL-C ($p=0.003$) and are nearly significant for APOAI ($p=0.054$). Overall, the results obtained at t_0 are in coherence with the results at t_{+5}. C_{10} presents the highest mean values for DBP and concentrations of TGs, TC, APOB and APOE (LDL-C at t_0) and the lowest mean values for APOAI and HDL-C concentrations. C_{16} presents the highest mean values for HDL-C and APOAI concentrations and the lowest mean values for concentrations of TC, APOB and LDL-C at t_0 and TG at t_{+5}. C_{44} is the least individualised cluster. It associates with the less addictive habits (alcohol consumption and smoking) at t_{+5}. C_{12} associates with the lowest mean values for waist, hip, body mass index (BMI), weight, and glucose level. The C_{12} group members are thinner, with the lowest glucose, TC and APOB concentrations at t_{+5}. The association of clusters with biological variables is independent of age.

Discussion

Considering the homogeneity resulting from inclusion and non-inclusion criteria of the 1551 initially apparently healthy subjects from the Stanislas cohort, we obtained a genetic homogeneity for most of the individuals (C_{1453}), but five other groups have clearly emerged at the same time. Despite the fact that these groups are relatively small, their genetic profiles can be clearly individualised and five genetic polymorphisms were sufficient to separate them. Clinical and/or biological differences already exist among these clusters. Some results are significant and others are of borderline significance, but it is important to note that most of them are coherent among the clusters at t_0 and 5 years later.

Considering the CVD marker values, two groups are especially interesting: C_{16} and C_{10}. The genetic signature of C_{16} consists of the E-selectin gene (*SELE* 98G/T and Ser128Arg) and the *CETP* gene (Ile405Val polymorphism). The genetic profile of C_{10} is characterised by the *APOC3* gene (–482C/T, –455T/C, 1100C/T, 3175C/G and 3206T/G), the *PON1* gene (Gln192Arg polymorphism) and the *AGT* gene (Met235Thr polymorphism). On the one hand, traditional risk factors for CVD are highest in C_{10} and the least high in C_{16} (significant for TC, APOB and LDL-C at t_0 and for TGs at t_{+5}). On the other hand, the biological parameters that are linked to a protective cardiovascular profile are observed in the reverse order for these clusters (HDL-C and APOAI concentrations at t_0 and t_{+5}). Therefore, individuals that belong to C_{10} present a risk profile for CVD, whereas individuals of C_{16} cluster seem to have a protective profile. In the C_{16} cluster, all the individuals are *SELE*-98TT and Arg128Arg and 75% are *CETP* 405VV. Information in the literature is sparse concerning the *SELE*- 98G/T polymorphism. Here, we confirm results from a previous study in

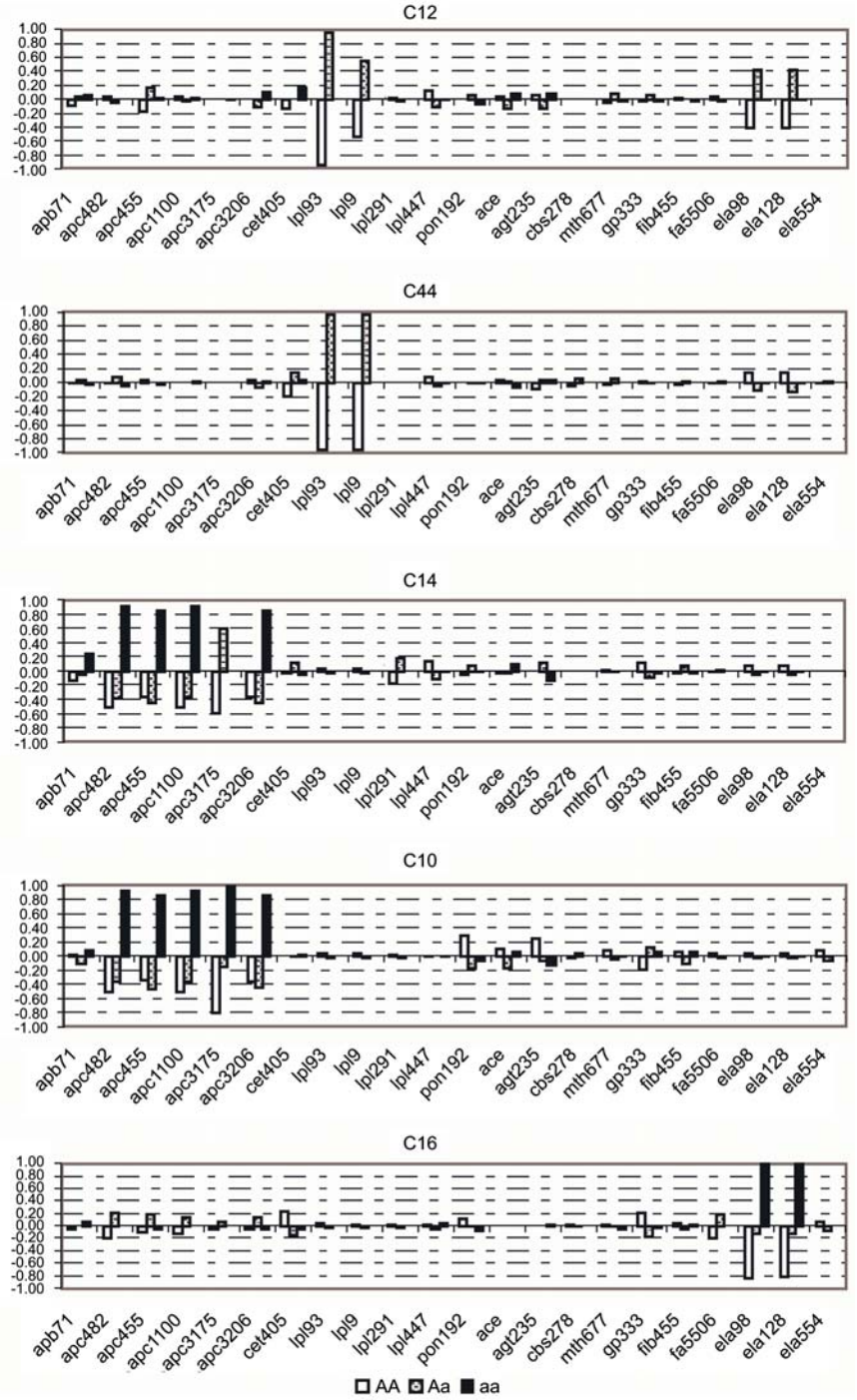


Figure 1 Genetic signatures/profiles of the clusters.
For each of the five clusters named C_{12}, C_{44}, C_{14}, C_{10} and C_{16}, the 22 genetic variables on the x-axis are represented. apb71: *APOB* Thr71Ile; apc482: *APOC3*–482C/T; apc455: *APOC3*–455T/C; apc1100: *APOC3* 1100C/T; apc3175: *APOC3* 3175C/G; apc3206: *APOC3* 3206T/G; cet405: *CETP* Ile405Val; lpl93: *LPL* –93T/G; lpl9: *LPL* Asp9Asn; lpl291: *LPL* Asn291Ser; lpl447: *LPL* Ser447Ter; pon192: *PON1* Gln192Arg; ace: *ACE I/D*; agt235: *AGT* Met235Thr; cbs278: *CBS* Ile278Thr; mth677: *MTHFR* 677C/T; gp333: *GPIIIa* Leu33Pro; fib455: *FGB* –455G/A; fa5506: *FV* Arg506Gln; ela98: *SELE* 98G/T; ela128: *SELE* Ser128Arg; ela554: *SELE* Leu554Phe. The distribution of each genetic variable is represented on the y-axis (plus and minus 100%) by three potential bars in the following order: homozygous for the common allele, AA (white bar); heterozygous, Aa (stippled bar); homozygous for the less frequent allele, aa (black bar). The dimension and the sign of each bar depend, respectively, on the value and on the sign of its y-value. The cluster C_{1453} is not represented, because all its y-values are equal to 0. See Materials and methods for the calculation of y-values.

which the 98T (and by extension 128Arg, because it is in strong linkage disequilibrium with the 98G/T polymorphism) allele was associated with higher HDL-C concentration in women only (13, 19). Moreover, our findings are in agreement with results from the EARS II study, which was performed on European

Table 3 Clinical assessments and biological measurements among the clusters.

t_0, mean (SD)	C_{1549} (n = 1549)	C_{1453} (n = 1453)	C_{12} (n = 12)	C_{44} (n = 44)	C_{14} (n = 14)	C_{10} (n = 10)	C_{16} (n = 16)	p-Value
Age, years	40.50 (4.60)	40.44 (4.57)	40.08 (4.03)	42.14 (5.32)	39.57[a] (4.26)	43.50[b] (4.88)	41.31 (4.32)	0.101
Height, cm	167.72 (8.99)	167.68 (9.01)	165.53 (7.97)	169.66 (8.17)	163.56[a] (8.48)	169.57 (9.91)	169.99[b] (8.47)	0.238
Weight, kg	68.88 (13.03)	68.76 (12.94)	64.01[a] (9.16)	70.81 (12.65)	70.60 (22.83)	74.30[b] (15.65)	72.45 (10.87)	0.319
BMI, kg/m^2	24.37 (3.51)	24.35 (3.48)	23.46[a] (3.62)	24.45 (2.91)	25.97[b] (6.39)	25.62 (3.49)	25.08 (3.40)	0.639
Waist, cm	81.25 (10.98)	81.12 (10.93)	78.17[a] (10.66)	83.86 (10.45)	83.86 (18.19)	86.10[b] (11.49)	83.19 (8.70)	0.233
Hip, cm	97.73 (6.71)	97.73 (6.67)	95.25[a] (5.46)	96.82 (5.89)	98.64 (11.93)	99.20 (7.91)	99.81[b] (6.88)	0.499
WHR, 1	0.83 (0.09)	0.83 (0.09)	0.82[a] (0.10)	0.86 (0.08)	0.85 (0.12)	0.87[b] (0.08)	0.83 (0.07)	0.120
SBP, mm Hg	122.67 (12.64)	122.54 (12.47)	123.42 (12.83)	125.02 (11.65)	122.57 (8.20)	135.00[b] (32.01)	119.69[a] (11.28)	0.626
DBP, mm Hg	74.81 (9.78)	74.79 (9.69)	71.83[a] (9.72)	75.95 (9.29)	72.21 (8.47)	84.30[b] (18.79)	71.94 (10.24)	0.020**
Alcohol, g/day	13.96 (21.91)	13.55[a] (21.40)	23.96 (31.43)	14.99 (19.74)	14.64 (21.48)	23.44 (29.05)	34.21[b] (43.47)	0.122
Current smoking, cig/day	4.25 (8.39)	4.25 (8.36)	4.25 (11.51)	3.00 (6.57)	5.57 (10.56)	2.90[a] (4.86)	7.19[b] (11.96)	0.805
Past smoking, cig/day	4.27 (9.43)	4.14 (9.19)	6.58 (12.20)	5.86 (9.70)	11.07[b] (21.97)	9.10 (13.88)	1.31[a] (4.51)	0.144
Glucose, mmol/L	5.07 (0.512)	5.07 (0.51)	4.93[a] (0.26)	5.08 (0.54)	5.16 (0.73)	5.29 (0.45)	5.31[b] (0.40)	0.103
TG, mmol/L	1.06 (0.67)	1.04[a] (0.65)	1.19 (0.54)	1.31 (0.86)	1.36 (1.10)	1.77[b] (0.96)	1.12 (0.92)	0.009***
TC, mmol/L	5.73 (1.00)	5.72 (1.00)	5.96 (1.02)	5.82 (0.97)	5.65 (0.91)	6.64[b] (0.82)	5.61[a] (1.01)	0.059*
LDL-C, mmol/L	3.83 (0.95)	3.82 (0.95)	4.02 (0.95)	3.95 (0.92)	3.75 (0.82)	4.65[b] (1.01)	3.58[a] (1.07)	0.059*
HDL-C, mmol/L	1.41 (0.40)	1.42 (0.40)	1.39 (0.30)	1.27 (0.43)	1.27 (0.39)	1.18[a] (0.26)	1.53[b] (0.61)	0.014**
APOB, g/L	1.06 (0.25)	1.05 (0.25)	1.14 (0.21)	1.10 (0.27)	1.06 (0.27)	1.31[b] (0.27)	1.02[a] (0.28)	0.042**
APOAI, g/L	1.62 (0.26)	1.63 (0.26)	1.60 (0.21)	1.54 (0.25)	1.53 (0.24)	1.51[a] (0.17)	1.65[b] (0.25)	0.111
APOE, mg/L	38.41 (10.96)	38.24[a] (10.80)	38.80 (5.48)	41.67 (15.08)	40.35 (15.11)	42.16[b] (13.13)	40.12 (9.04)	0.708

Results at t_0. *$p < 0.10$; **$p < 0.05$; ***$p < 0.01$; [a]lower value among clusters; [b]higher value among clusters. BMI, body mass index; WHR, waist-to-hip ratio; SBP, systolic blood pressure; DBP, diastolic blood pressure; TG, triglyceride; TC, total cholesterol; LDL-C, low-density lipoprotein cholesterol; HDL-C, high-density lipoprotein cholesterol; APOAI, apolipoprotein AI; APOB, apolipoprotein B; APOE, apolipoprotein E.

Table 4 Clinical assessments and biological measurements among the clusters.

t_{+5}, mean (SD)	C_{1549} (n = 1128)	C_{1453} (n = 1052)	C_{12} (n = 10)	C_{44} (n = 34)	C_{14} (n = 11)	C_{10} (n = 8)	C_{16} (n = 13)	p-Value
Age, years	45.55 (4.64)	45.50 (4.63)	43.70[a] (3.47)	47.50[b] (4.87)	44.18 (4.75)	47.50[b] (4.72)	46.62 (4.61)	0.081*
Height, cm	167.66 (9.04)	167.62 (9.05)	166.17 (8.01)	169.36 (8.71)	162.45[a] (8.61)	171.14[b] (10.01)	169.85 (9.38)	0.274
Weight, kg	71.15 (13.67)	71.07 (13.65)	64.39[a] (7.65)	72.69 (13.25)	68.05 (15.61)	78.93[b] (19.20)	76.96 (12.25)	0.148
BMI, kg/m^2	25.20 (3.82)	25.19 (3.84)	23.39[a] (2.96)	25.19 (3.14)	25.53 (4.20)	26.61 (4.29)	26.72[b] (4.06)	0.390
Waist, cm	83.10 (11.52)	83.01 (11.54)	77.10[a] (7.33)	85.29 (11.32)	82.64 (12.50)	87.88[b] (12.91)	86.42 (10.20)	0.212
Hip, cm	97.89 (7.20)	97.88 (7.25)	93.90[a] (5.38)	97.71 (5.94)	97.91 (7.84)	97.25 (5.95)	102.25[b] (6.65)	0.118
WHR, 1	0.85 (0.09)	0.85 (0.09)	0.83[a] (0.093)	0.87 (0.09)	0.84 (0.11)	0.90[b] (0.09)	0.84 (0.08)	0.397
SBP, mm Hg	124.19 (13.73)	123.97 (13.66)	129.10 (12.56)	126.79 (12.70)	122.48[a] (10.26)	137.79[b] (24.75)	124.67 (13.26)	0.161
DBP, mm Hg	74.65 (10.16)	74.51[a] (10.14)	75.60 (12.35)	76.19 (8.84)	74.61 (9.51)	84.67[b] (11.47)	75.05 (11.22)	0.205
Alcohol, g/day	13.60 (19.61)	13.39 (19.50)	15.85 (19.72)	12.79[a] (16.94)	17.04 (20.94)	27.62[b] (29.33)	19.24 (26.11)	0.494
Current smoking, cig/day	3.28 (7.79)	3.20 (7.72)	5.30 (12.60)	2.74[a] (6.21)	8.73[b] (12.36)	6.13 (9.11)	3.38 (6.64)	0.429
Past smoking, cig/day	5.31 (10.25)	5.40 (10.31)	4.00 (8.43)	4.15 (8.69)	4.09 (8.01)	7.50[b] (17.53)	1.69[a] (5.53)	0.862
Glucose, mmol/L	5.08 (0.75)	5.07 (0.74)	4.97[a] (0.74)	5.34 (0.98)	5.32 (1.04)	5.50[b] (0.59)	5.05 (0.48)	0.133
TG, mmol/L	1.27 (1.31)	1.25 (1.33)	1.24 (0.72)	1.55 (0.97)	1.56 (1.20)	2.43[b] (1.04)	1.06[a] (0.43)	0.005***
TC, mmol/L	5.77 (0.96)	5.77 (0.97)	5.51[a] (0.88)	5.83 (0.90)	5.95 (0.97)	6.49[b] (0.90)	5.69 (0.70)	0.337
LDL-C, mmol/L	3.61 (0.95)	3.60 (0.95)	3.47 (0.72)	3.72 (0.92)	3.78 (0.95)	4.18[b] (0.93)	3.39[a] (0.63)	0.408
HDL-C, mmol/L	1.58 (0.45)	1.59 (0.44)	1.48 (0.50)	1.39 (0.47)	1.45 (0.37)	1.20[a] (0.34)	1.81[b] (0.68)	0.003***
APOB, g/L	1.026 (0.237)	1.02 (0.24)	0.99[a] (0.22)	1.06 (0.25)	1.09 (0.26)	1.29[b] (0.25)	1.00 (0.21)	0.080*
APOAI, g/L	1.620 (0.250)	1.62 (0.25)	1.55 (0.23)	1.52 (0.22)	1.61 (0.24)	1.51[a] (0.22)	1.75[b] (0.31)	0.054*
APOE, mg/L	40.62 (13.82)	40.50 (13.94)	39.30 (7.86)	41.57 (12.55)	39.06[a] (10.41)	48.86[b] (17.72)	44.41 (10.10)	0.511

Results at t_{+5}. *$p<0.10$; **$p<0.05$; ***$p<0.01$; [a]lower value among clusters; [b]higher value among clusters. BMI, body mass index; WHR, waist-to-hip ratio; SBP, systolic blood pressure; DBP, diastolic blood pressure; TG, triglyceride; TC, total cholesterol; LDL-C, low-density lipoprotein cholesterol; HDL-C, high-density lipoprotein cholesterol; APOAI, apolipoprotein AI; APOB, apolipoprotein B; APOE, apolipoprotein E.

men and showed that individuals homozygous for the rare V405 allele had the highest HDL levels (20). Concerning the C_{10} cluster, all the individuals are homozygous for the less frequent alleles of the *APOC3* polymorphisms studied. We demonstrated previously that APOB, LDL-C and TC concentrations were higher in women with the *APOC3* 1100TT genotype who do not smoke. Additionally, the T allele was related to higher TG, APOCIII and APOB-related traits, i.e., TC, LDL-C and APOB in men. The TT genotype was related to lower levels of APOAI and HDL-C only in male non-smokers (13). The genetic signature of this cluster was very close to that of C_{14}. The difference between the two groups came especially from an over-representation of *APOC3* 3175GG in C_{10}. Therefore, this polymorphism, in addition, but to a lesser extent to the other *CETP* polymorphisms, may be important in the pathophysiology of CVD. Previous studies have shown that the rare allele of the 3175C/G (also called SstI) polymorphism of the *APOC3* gene is associated with elevated TG and APOCIII concentrations (21, 22), higher blood pressure (23) and increased risk of myocardial infarction (24). In the Framingham Offspring Study, the 3175G allele was associated with elevated fasting insulin concentrations in men only (25). We observed that most of the subjects in C_{10} are homozygous for the common allele of the *PON1* Gln192Arg polymorphism. The Gln192 allele is associated with low paraoxon activity (26). Moreover, low PON1 activity was suggested to be associated with coronary heart disease (27, 28). Conversely, subjects with the Arg192 allele were found to be at higher risk for CVD than subjects with the GlnGln genotype (29, 30). Other investigations, involving French and Irish (31), Finnish (32) and Japanese subjects (33) did not show any significant association between the *PON1* Gln192Arg polymorphism and CVD. Moreover, Gln homozygotes were found to be more advanced in age compared to Arg allele carriers in the Stanislas cohort (34), and being homozygous for 192 ArgArg was linked to enhanced survival (35). One possible explanation for the observed discrepancies could come from the incomplete concordance between PON1 phenotypes and PON1 genotypes (36). Finally, concerning the *AGT* Met235Thr polymorphism, results have suggested that the ThrThr genotype could be a genetic marker for arterial stiffness in patients with never-treated hypertension (37). The risk of developing hypertension is higher for MetThr and ThrThr genotypes compared to MetMet genotypes (38). Moreover, homozygosity for the angiotensinogen 235T variant was found to increase the risk of myocardial infarction in patients with multi-vessel coronary artery disease (39). On the other hand, results from a study by Gardemann et al. showed that blood pressure levels and the prevalence of high blood pressure did not vary according to M235T genotypes, and the M235T genotype distributions did not differ between survivors of myocardial infarction and controls (40). Other studies did not find that the 235T allele was a risk factor for coronary heart disease (41, 42). Our previous results on the AGT Met235Thr polymorphism did not show any association between this variant and carotid intima-media thickness (43).

In summary, the results of this study are in agreement with the literature, especially with studies involving confirmed CVD. Of course, this work concerns the CVD risk factor-related measurement and not the disease itself. We cannot show today, whether the gene clusters identified do really contribute to the confirmed disease, but we have produced strong evidence, according to biological results, that *SELE* may be linked to a cardiovascular protective profile, while *APOC3* polymorphisms, and 3175C/G in particular, may increase the CVD risk. This profiling allows accuracy in the description and in the follow-up of the cohort: analysis of future events according to previously identified genetic groups.

Acknowledgements

This research was funded by the French Institut National de la Santé Et de la Recherche Médicale (INSERM), by the Région Lorraine and the Communauté Urbaine du Grand Nancy. We also wish to thank the Centre de Médecine Préventive of Vandoeuvre-Lès-Nancy, France for the recruitment of the Stanislas cohort and the families who have accepted to participate. The Stanislas cohort is supported by the Caisse Nationale d'Assurance Maladies des Travailleurs Salariés (CNAM) and samples and data are taking part of the CRB (Centre of Biological Resources). Finally, we thank Roche Molecular Systems for providing us with genotyping reagents under research collaboration.

References

1. Siest G, Visvikis S, Herberth B, Gueguen R, Vincent-Viry M, Sass C, et al. Objectives, design and recruitment of a familial and longitudinal cohort for studying gene-environment interactions in the field of cardiovascular risk: the Stanislas cohort. Clin Chem Lab Med 1998;36:35–42.
2. Visvikis S, Sass C, Pallaud C, Grow MA, Zannad F, Siest G, et al. Familial studies on the genetics of cardiovascular diseases: the Stanislas cohort. Clin Chem Lab Med 2000; 38:827–32.
3. Mansour-Chemaly M, Haddy N, Siest G, Visvikis S. Family studies: their role in the evaluation of genetic cardiovascular risk factors. Clin Chem Lab Med 2002;40: 1085–96.
4. Pallaud C, Stranieri C, Sass C, Siest G, Pignatti F, Visvikis S. Candidate gene polymorphisms in cardiovascular disease: a comparative study of frequencies between a French and an Italian population. Clin Chem Lab Med 2001;39:146–54.
5. Xia Y, Baum L, Pang CP, Siest G, Visvikis S. Cardiovascular risk-associated allele frequencies for 15 genes in healthy elderly French and Chinese. Clin Chem Lab Med 2005;43:817–22.
6. Visvikis-Siest S, Marteau JB. Genetic variants predisposing to CVD. Curr Opin Lipidol 2006;17:139–51.
7. Lander ES, Linton LM, Birren B, Nusbaum C, Zody MC, Baldwin J, et al. Initial sequencing and analysis of the human genome. Nature 2001;409:860–921.
8. Kitano H. Computational systems biology. Nature 2002; 420:206–10.
9. Persidis A, Deftereos S, Persidis A. Systems literature analysis. Pharmacogenomics 2004;5:943–7.

10. Siest G. Study of reference values and biological variation: a necessity and a model for preventive medicine centers. Clin Chem Lab Med 2004;42:810–6.
11. Colhoun HM, McKeigue PM, Davey Smith G. Problems of reporting genetic associations with complex outcomes. Lancet 2003;361:865–72.
12. Selinger-Leneman H, Genin E, Norris JM, Khlat M. Does accounting for gene-environment (GxE) interaction increase the power to detect the effect of a gene in a multifactorial disease? Genet Epidemiol 2003;24:200–7.
13. Pallaud C, Gueguen R, Sass C, Grow M, Chen S, Siest G, et al. Genetic influences on lipid metabolism trait variability within the Stanislas cohort. J Lipid Res 2001; 42:1879–90.
14. Woda A, Tubert-Jeannin S, Bouhassira D, Attal N, Fleiter B, Goulet JP, et al. Towards a new taxonomy of idiopathic orofacial pain. Pain 2005;116:396–406.
15. Siest G, Bertrand P, Herberth B, Vincent-Viry M, Schiele F, Sass C, et al. Apolipoprotein E polymorphisms and concentration in chronic diseases and drug responses. Clin Chem Lab Med 2000;38:841–52.
16. Miller SA, Dykes DD, Polesky HF. A simple salting out procedure for extracting DNA from human nucleated cells. Nucleic Acids Res 1988;16:1215.
17. Cheng S, Pallaud C, Grow MA, Scharf SJ, Erlich HA, Klitz W, et al. A multilocus genotyping assay for cardiovascular disease. Clin Chem Lab Med 1998;36:561–6.
18. Hoppe C, Klitz W, Cheng S, Apple R, Steiner L, Robles L, et al. Gene interactions and stroke risk in children with sickle cell anemia. Blood 2004;103:2391–6.
19. Tilly P, Sass C, Vincent-Viry M, Aguillon D, Siest G, Visvikis S. Biological and genetic determinants of serum APOC-III concentration: reference limits from the Stanislas cohort. J Lipid Res 2003;44:43043–6.
20. Gudnason V, Kakko S, Nicaud V, Savolainen MJ, Kesäniemi YA, Tahvanainen E, et al. Cholesteryl ester transfer protein gene effect on CETP activity and plasma high-density lipoprotein in European populations. The EARS group. Eur J Clin Invest 1999;29:116–28.
21. Shoulders CC, Harry PJ, Lagrost L, White SE, Shah NF, North JD, et al. Variation at the APO AI/CIII/AIV gene complex is associated with elevated plasma levels of APO CIII. Atherosclerosis 1991;87:239–47.
22. Talmud PJ, Humphries SE Apolipoprotein C-III gene variation and dyslipidaemia. Curr Opin Lipidol 1997;8:154–8.
23. Tas S, Abdella NA. Blood pressure, coronary artery disease, and glycaemic control in type 2 diabetes mellitus: relation to apolipoprotein-CIII gene polymorphism. Lancet 1994;343:1194–5.
24. Ferns GA, Stocks J, Ritchie C, Galton DJ. Genetic polymorphisms of apolipoprotein C-III and insulin in survivors of myocardial infarction. Lancet 1985;2:300–3.
25. Russo GT, Meigs JB, Cupples LA, Demissie S, Otvos JD, Wilson PW, et al. Association of the Sst-I polymorphism at the APOc3 gene locus with variations in lipid levels, lipoprotein subclass profiles and coronary heart disease risk: the Framingham offspring study. Atherosclerosis 2001;158:173–81.
26. Mackness MI, Mackness B, Durrington PN, Connelly PW, Hegele RA. Paraoxonase: biochemistry, genetics and relationship to plasma lipoproteins. Curr Opin Lipidol 1996;7:69–76.
27. McElveen J, Mackness MI, Colley CM, Peard T, Warner S, Walker CH. Distribution of paraoxon hydrolytic activity in the serum of patients after myocardial infarction. Clin Chem 1986;32:671–3.
28. Ayub A, Mackness MI, Arrol S, Mackness B, Patel J, Durrington PN. Serum paraoxonase after myocardial infarction. Arterioscler Thromb Vasc Biol 1999;19:330–5.
29. Saha N, Roy AC, Teo SH, Tay JS, Ratnam SS. Influence of serum paraoxonase polymorphism on serum lipids and apolipoproteins. Clin Genet 1991;40:277–82.
30. Hegele RA, Brunt JH, Connelly PW. Multiple genetic determinants of variation of plasma lipoproteins in Alberta Hutterites. Arterioscler Thromb Vasc Biol 1995; 15:861–71.
31. Herrmann SM, Blanc H, Poirier O, Arveiler D, Luc G, Evans A, et al. The Gln/Arg polymorphism of human paraoxonase (PON 192) is not related to myocardial infarction in the ECTIM Study. Atherosclerosis 1996; 126:299–303.
32. Antikainen M, Murtomäki S, Syvänne M, Pahlman R, Tahvanainen E, Jauhiainen M, et al. The Gln-Arg191 polymorphism of the human paraoxonase gene (HUMPonA) is not associated with the risk of coronary artery disease in Finns. J Clin Invest 1996;98:883–5.
33. Suehiro T, Nakauchi Y, Yamamoto M, Arii K, Itoh H, Hamashige N, et al. Paraoxonase gene polymorphism in Japanese subjects with coronary heart disease. Int J Cardiol 1996;57:69–73.
34. Xia Y, Gueguen R, Vincent-Viry M, Siest G, Visvikis S. Effect of six candidate genes on early aging in a French population. Aging Clin Exp Res 2003;15:111–6.
35. Rea IM, McKeown PP, McMaster D, Young IS, Patterson C, Savage MJ, et al. Paraoxonase polymorphisms PON1 192 and 55 and longevity in Italian centenarians and Irish nonagenarians. A pooled analysis. Exp Gerontol 2004; 39:629–35.
36. Vincent-Viry M, Sass C, Bastien S, Aguillon D, Siest G, Visvikis S. PON1-192 phenotype and genotype assessments in 918 subjects of the Stanislas cohort study. Clin Chem Lab Med 2003;41:535–40.
37. Bozec E, Lacolley P, Bergaya S, Boutouyvrie P, Meneton P, Herissé-Legrand M, et al. Arterial stiffness and angiotensinogen gene in hypertensive patients and mutant mice. J Hypertens 2004;22:1299–307.
38. Kishimoto T, Suyama A, Igarashi A, Osaki Y, Okamoto M, Yamamoto T, et al. Angiotensinogen gene variation and hypertension in a cohort study in Japanese. J Epidemiol 2001;11:115–9.
39. Olivieri O, Stranieri C, Girelli D, Pizzolo F, Grazioli S, Russo C, et al. Homozygosity for angiotensinogen 235T variant increases the risk of myocardial infarction in patients with multi-vessel coronary artery disease. J Hypertens 2001;19:879–84.
40. Gardemann A, Stricker J, Humme J, Nguyen QD, Katz N, Philipp M, et al. Angiotensinogen T174M and M235T gene polymorphisms are associated with the extent of coronary atherosclerosis. Atherosclerosis 1999;145:309–14.
41. Ko YL, Ko YS, Wang SM, Chu PH, Teng MS, Cheng NJ, et al. Angiotensinogen and angiotensin-I converting enzyme gene polymorphisms and the risk of coronary artery disease in Chinese. Hum Genet 1997;100:210–4.
42. Yamakawa-Kobayashi K, Arinami T, Hamaguchi H. Absence of association of angiotensinogen gene T235 allele with increased risk of coronary heart disease in Japanese. Lancet 1995;346:515.
43. Pallaud C, Sass C, Zannad F, Siest G, Visvikis S. APOc3, CETP, fibrinogen, and Mthfr are genetic determinants of carotid intima-media thickness in healthy men (the Stanislas cohort). Clin Genet 2001;59:316–24.

PUBLICATION N°6

Ndeye Coumba Ndiaye, Said EL Shamieh, Jean Brice Marteau, Gérard Siest, Micheal Y Tsai, Sophie Visvikis-Siest. Functional epistases regulate blood pressure. 2010 [soumis]

Notre objectif était d'étudier les interactions épistasiques de 100 variants génétiques appartenant à 65 gènes candidats de différentes voies métaboliques (pression artérielle, inflammation, métabolisme lipidique, métabolisme de l'homocystéine, adhésion cellulaire, thrombose).

Nous avons donc génotypé 1 912 adultes de façon à effectuer des études d'association individuelles.

Neuf SNPs ont montré une association significative avec la pression artérielle systolique, diastolique ou la pression pulsée. Cinq étaient fonctionnels : rs5742910 et rs6046 dans le *FVII* (P≤0,024), rs1800469 au sein du gène *TGFB1* (P<0,001) et potentiellement responsables d'une diminution de la pression artérielle. Le rs1800590 dans le gène *LPL* (P≤0,003*)* et le rs2228570 dans le *VDR* (P<0,001) avaient quant à eux une potentielle fonction d'altération de la structure protéique.

Nous avons également observé les interactions épistatiques suivantes : *LPL* (rs1801177)**PON1* (rs662) (P≤0,032) pour tous les traits de pression artérielle mais aussi des interactions spécifiques à la pression systolique: *VCAM1* (rs1041163)**APOB* (rs1367117) (P≤0,04), ou diastolique: *SCGB1A1* (rs3741240)**LPL* (rs1800590 et rs1801177) (P≤0,031). Une étude *in silico* nous a permis de déduire la fonctionnalité de ces interactions.

Contribution :

Lors de cette étude, nous avons participé à l'élaboration du design. Nous avons effectué la totalité des analyses statistiques et pris en charge la rédaction du premier manuscrit.

Functional Epistases Regulate Blood Pressure

Ndeye Coumba Ndiaye[1], PhD Student, Said EL Shamieh[1], PhD Student, Jean Brice Marteau[1], PhD, Gérard Siest[1],PharmD, PhD, Micheal Y Tsai[2],MD, PhD, Sophie Visvikis-Siest[1], Pharm D, PhD.

5,651 words

Running Title: Epistasis on Blood Pressure Regulation

Authors' affiliations

1 "Cardiovascular Genetics" Research Unit, EA-4373, Université Henri Poincaré - Nancy 1, 54000 Nancy, France

2 Department of Laboratory Medicine & Pathology, University of Minnesota, Minneapolis, MN 55455-0392, USA

The authors have no relationships with industry

Address for correspondence: Dr. Sophie Visvikis-Siest

"Cardiovascular Genetics" Research Unit, EA-4373, Université Henri Poincaré - Nancy 1

30 rue Lionnois – 54000 Nancy, France

Tel: +33607602569 / Fax:+33383682163

Sophie.Visvikis-Siest@inserm.fr

Abstract

Objectives: To investigate epistatic interactions of 100 candidate genes from various biological pathways in regard to blood pressure.

Background: In spite of candidate gene and genome-wide association studies, only a small number of genetic variants regulating blood pressure have been identified. This could be partly due to gene-gene interactions: epistasis.

Methods: One hundred genetic variants in 65 candidate genes from various biological pathways (blood pressure regulation, inflammation, lipid metabolism, homocysteine metabolism, cellular adhesion, and thrombosis) were genotyped in 1,912 French middle-aged individuals in order to study their combined effects on blood pressure regulation.

Results: Among 9 SNPs significantly associated with systolic, diastolic blood pressure and pulse pressure, 5 were functional. rs5742910 and rs6046 in *FVII* (P≤0.024), rs1800469 in *TGFB1* (P<0.001) may be responsible for blood pressure defects modifying their plasma protein levels, whereas rs1800590 in *LPL* (P≤0.003*)* and rs2228570 in *VDR* (P<0.001) may act via altering protein structure. When combinatory analyses were performed, one epistatic interaction was common for the 3 blood pressure traits investigated: *LPL* (rs1801177)***PON1* (rs662) (P≤0.032). In addition, specific interactions were found for systolic blood pressure: *VCAM1* (rs1041163)***APOB* (rs1367117) (P≤0.04), as well as diastolic blood pressure: *SCGB1A1* (rs3741240)***LPL* (rs1800590 and rs1801177) (P≤0.031). An *in silico* search yielded functional roles for these epistatic interactions.

Conclusions: The present findings provide insight into pathways such as lipid metabolism which may act synergistically with genes in the same or other pathways to modify hypertensive risk. This

strategy reveals novel pathophysiologic mechanisms underlying hypertension and potential therapeutic targets.

A Table listing all the investigated genetic variants is provided in supplementary material.

Key words: blood pressure, epistasis, single nucleotide polymorphism, epidemiology

Introduction

Blood pressure (BP) is a continuous, consistent and modifiable risk factor for cardiovascular diseases (CVD) such as ischemic heart disease (IHD) and stroke which are major causes of mortality and morbidity in the world (1). Population-based studies showed that 972 million adults worldwide were hypertensive in 2005 and it is predicted to increase by about 60% to 1.56 billion in 2025 (2).

Existing evidence suggests that BP heritability ranges from 30 to 60% (3). Despite the identification of specific causal genes involved in the regulatory pathways of rare familial forms of hypertension (4), both BP and hypertension are still considered polygenic traits involving a number of metabolic pathways (5,6).

The complex architecture of BP regulation may account for the numerous discrepancies found among studies of genetics underlying hypertension. Inflammation, thrombosis, cellular adhesion molecules and lipid metabolism all appear to play a role (7-9). While aspects of BP regulation may be the product of an inflammation stimulus (7,10,11), thrombin is also known to induce several intracellular pathways (8), modulate vascular tone (12) and have pro-inflammatory effects. Mounting evidence also suggests that cellular adhesion molecules and apolipoproteins are closely related to hypertension (9,13). All of these properties may be involved in initiating and maintaining elevated BP levels (8). Overall, however, genetic factors associated with BP are poorly characterized and these pathways remain unclear.

In the Genome-wide association studies (GWAS) era, technological advances facilitated the genotyping of more than two million single nucleotide polymorphisms (SNPs). Researchers have since attempted to identify susceptibility loci of BP (5,14) but limited progress has been made thus far (15,16). A possible explanation is the significant genetic and phenotypic heterogeneity for this phenotype and the modest effect size of risk alleles (17). Besides, GWAS have not incorporated epistatic interactions, though examining such interactions may be essential (18) in identifying the genes involved in BP regulation.

To our knowledge, no previous study has been conducted on European individuals in assessing the epistasis of cardiovascular candidate genes on BP regulation. Thus, the purpose of this study was to investigate 100 genetic variants in 65 cardiovascular candidate genes from various biological pathways in regard to BP traits, and their epistatic interactions. Based on previous findings (15,16,18), we hypothesize that epistatic interactions among candidate SNPs may be more informative than individual SNPs analyses in influencing BP regulation.

Methods

Ethics statement

The participants are part of a human sample storage platform: the Biological Resources Bank (BRC) "Interactions Gène-Environnement en Physiopathologie CardioVasculaire" (BRC IGE-PCV) in Nancy, East of France. Each individual involved gave his written informed consent and the project's populations were approved by the local ethics committee (*Comité Consultatif de Protection des Personnes se prêtant à une Recherche Biomédicale – Lorraine, France*).

Study population

The present study enrolled 1,912 middle-aged French subjects (51.3 % women) selected from the BRC IGE-PCV on the basis of the following criteria: (1) subjects were Caucasians born in France since three generations; (2) no antihypertensive drugs administration or, for hypertensive patients, the BP phenotypes were measured at baseline before drug administration; and (3) complete clinical and genotypic data available.

Clinical and biological data collection

Systolic blood pressure (SBP) and diastolic blood pressure (DBP) was measured under constant temperature (19°C to 21°C) and standardized conditions (supine position) with a manual sphygmomanometer by expert nurses. The values reported were the means of 3 readings at each examination on 20 min intervals. Pulse pressure (3) were calculated as the difference between SBP

and DBP. All individuals underwent complete medical examination including anthropometric and biochemical measurements collected with standardized methods(19).

Genotyping

Genomic DNA was extracted from peripheral blood samples by the salting out method (Miller SA, 1988). Genotyping was performed by using multilocus assay with an immobilized probe approach previously described(20) (Roche Molecular Systems, Pleasanton, California, USA).

After PCR amplification using pooled biotinylated primers and hybridation to sequence-specific oligonucleotide probes, genotype assignments were performed by two independent observers using proprietary Roche Molecular Systems image processing software.

Discordant results (< 3% of all scoring) were resolved by a third observer and if necessary, by a joint reading.

Blinded replication of the fragment analysis was performed with 10% of the selected samples on an automated high-performance liquid chromatography instrument (Wave system™; Transgenomics Inc. NC, USA), both sequencing and restriction analyses fully matched.

PolyPhen analysis of nonsynonymous SNPs

The prediction of possible impact of nonsynonymous (ns) SNPs on the structure and function of their specific protein was performed using PolyPhen(21).

Statistical analyses

Statistical analyses were performed using the SPSS® statistical software version 16.0 (SPSS, Inc, Chicago, Illinois).

Polymorphisms with minor allele frequencies (MAF) less than 2% were excluded from individual analyses. However, for interaction analyses we included all SNPs as low MAF may identify an epistatic phenomenon. HWE was tested using an exact method.

Separate models were constructed to determine the independent effect of each genetic variant on SBP, DBP, and PP assuming additive models with the common wild type (WT) as the reference group. Age, sex and body mass index (BMI)-adjusted linear regressions were performed for individual association analyses after ensuring that the gender did not interact with the genotype.

The variants individually associated in our study were compared to previous findings in the literature, notably on previous GWAS, by imputation analyses using plink(22).

Epistasis of variants that were independently associated with BP was established by interaction terms existing between genetic variants on linear additive models.

Power of all analyses (1-β) was determined a posterior and we considered significant associations only if it was greater than 80%. Type 1 error coefficient (α) was fixed à 5% and statistical significance was reached where p-values were less than 0.05 as we performed individual tests on each genetic variant.

Results

Phenotypic characteristics of the studied population

Phenotypic characteristics of the 1,912 French middle aged participants (51.13 years ± 10.02) are shown in Table 1. SBP, DBP, total cholesterol, low density lipoprotein-cholesterol (LDL-C) and glucose levels were higher in men compared to women (P<0.001). Women had higher PP and high density lipoprotein-cholesterol (HDL-C) levels (P<0.001). 18.5% of participants had normal BP (SBP<120 and DBP<80mmHg), 21.1% were prehypertensive (SBP and/or DBP between 120 and 139mmHg and/or 80 and 89mmHg), 25.9% had a stage 1 hypertension (SBP and/or DBP between 140 and 159mmHg and/or 90 and 99mmHg) and 34.5% were at stage 2 (≥160/100mmHg) according to the Seventh Report of the Joint National Committee (23).

Investigated genetic variants

Genotype and allele frequencies of all investigated genes of MAF quality control were consistent with Hardy-Weinberg equilibrium (HWE, data available upon request).The full details of investigated genetic variants are listed in Supplementary Table 1: the analysis included 100 SNPs that belong to 65 cardiovascular candidate genes selected according to our cardiovascular genetics background (19). Among the included SNPs, 50 were nonsynonymous, and according to Polyphen, 5 were damaging for their protein structure.

Individual association analysis

The results of individual association analyses of genetic variants with BP traits are shown in Table 2. Thirty SNPs in 25 genes including various metabolic pathways were significantly associated with SBP,

DBP or PP: *APOA4, APOB, APOC3, LIPC, LPL, PON1, PON2* and *LDLR* (lipid metabolism), *VCAM1* and *SELP* (cellular adhesion), *FVII* (coagulation factor), *GNB3*, *ATR1, AGT* and *NOS3* (BP regulation), *TNF*, *IL10*, *IL1B*, *C5*, *CSF2*, *TGFB1, CCL11*, *C3*, *LTA* and *VDR* (inflammatory pathways). Fourteen SNPs were associated with at least 2 or the 3 BP traits investigated. rs854560 in *PON1* (P=0.047) was associated with only SBP (Table 2A), and rs675 in *APOA4* (P=0.024), rs1367117 in *APOB* (P=0.024), rs2542052 and rs2854116 in *APOC3* (P≤0.024) respectively were specific for DBP (Table 2B). Nine SNPs rs1800590 in *LPL* (P≤0.003), rs5742911 in *LDLR* (P≤0.009), rs5742910 and rs6046 in *FVII* (P≤0.007 and P≤0.024 respectively), rs1800469 in *TGFB1* (P<0.001), rs1143634 in *IL1B* (P≤0.005), rs25882 in *CSF2* (<0.001), rsrs1129844 in *CCL11* (P<0.001) and rs2228570 in *VDR* (P<0.001) were associated with the 3 BP traits.

All genetic variants associated with a BP trait had an age and BMI cumulative effect (P<0.001, Table 2).

We performed imputation analyses in order to compare our results with the literature, but did not find any genetic variant in linkage disequilibrium with previous reports (all correlation coefficients <80%).

Epistatic interaction with BP

Epistatic interactions with BP are shown in Table 3. All genetic markers previously associated with at least one BP trait showed epistatic interactions influencing BP. One epistasis was common for the 3 BP traits: *LPL* (rs1801177)***PON1* (rs662), and 2 were specific to SBP and DBP: *VCAM1* (rs1041163)***APOB* (rs1367117) and *SCGB1A1* (rs3741240)***LPL* (rs1800590 and rs1801177) (Tables 3A and 3B). Interestingly, rs3741240 in *SCGB1A1* and rs1801177 in *LPL* were not individually associated with SBP and DBP (LPL was initially excluded from individual analyses tables as its MAF

was 1.82 but was associated with PP), but their epistatic combinations significantly influenced SBP and DBP (P = 0.013 and 0.026 respectively, Tables 3A and 3B).

Discussion

Even at early stages, hypertension is a major cause of disabilities and death for millions throughout the world (1). Observational studies involving more than 1 million adults from 61 prospective studies have indicated that death from hypertension associated diseases (both IHD and stroke) increases linearly from levels as low as 115/75mmHg among middle and elderly age individuals (23). Therefore, isolating genetic variations that may influence BP at middle age might have major implications for public health; however, individual genes have so far shown a weak genetic effect (11,24). It is therefore essential that epistatic gene interactions be studied to identify gene combinations that synergistically influence BP regulation and, by extension, dysregulation.

In the current study composed of 1,912 French middle aged individuals, we identify 30 SNPs in 25 genes involved in various biological pathways including lipid metabolism (*LIPC*, *LPL*, *PON1*, PON2, *APOA4*, *APOB* and *LDLR*), cellular adhesion molecules (*VCAM1* and *SELP*), coagulation factors *(FVII),* BP regulation (*AGTR1, AGT and NOS3*) and inflammation (*IL10*, *IL1B*, *C5*, *CCR3*, *CSF2*, *CCL11*, TNF, *TGFB1*, *C3*, *LTA* and *VDR*) that were associated with at least one of the 3 BP trait investigated (P<0.05, Table 2). Whereas 14 SNPs were associated with at least 2 BP traits, rs854560 in *PON1* (P=0.047) was associated with SBP only (Table 2A) and rs675 in *APOA4* (P=0.024), rs1367117 in A*POB* (P=0.024), rs2542052 and rs2854116 in *APOC3* (P≤0.024), and rs361525 in *TNF* (P=0.01) were specific for DBP (Table 2B). Interestingly, the following 9 SNPs were associated with the 3 BP traits (Table 2): rs1800590 in *LPL* (P≤0.003), rs5742911 in *LDLR* (P≤0.009), rs5742910 and rs6046 in *FVII* (P≤0.007 and P≤0.024 respectively), rs1800469 in *TGFB1* (P<0.001), rs1143634 in *IL1B* (P≤0.005), rs25882 in *CSF2* (P<0.001) rsrs1129844 in *CCL11* (P<0.001) and rs2228570 in *VDR* (P<0.001).

It is noteworthy that all 25 genes associated with at least one of the 3 BP traits overlapped with loci previously reported in BP GWAS (25-30). For example, in a genome-wide linkage analysis study conducted on 2,259 Nigerian individuals (26), Cooper et al associated chromosome regions 5q and 7q

with DBP, whereas others showed the association of 19q with SBP (26). In addition, Camp et al (25) reported the importance of loci on chromosome 11q and 12q in PP regulation in American individuals (25). All these chromosome regions contain the genes identified here (Supplementary Table 1).

Genetic variation may be responsible for defective BP phenotypes by altering either the structure of encoded proteins or gene expression (15). Among genetic variants commonly associated with SBP, DBP and PP, 5 were functional SNPs. For rs1800469 in *TGFB1*, there is evidence to suggest altered gene expression (31,32) as this SNP has been reported to be associated with increased *TGFB1* gene transcription and plasma protein concentration (32). For rs5742910 and nsSNP rs6046 in *FVII*, we have previously reported that both are associated with a one third decrease in FVII levels (8). It is important to emphasize that for both rs6041 and rs5742910 in FVII, we have previously shown similar associations in another sample population of the BCR (8). Together with previous data, our finding that *FVII* and *TGFB1* polymorphisms are associated with BP may be due to modifications of gene expression and resulting alterations in plasma protein levels.

Another mechanism through which polymorphisms may affect gene expression is by disrupting microRNA binding (miRNA). MicroRNAs inhibit gene expression by base pairing with complementary mRNA sequences. SNPs located within such targets can affect this base pairing. For example in the present study, rs5186 (1166 A/C) in *AGTR1* is located in the 3'UTR, a region associated with both SBP and DBP ($P \leq 0.05$). The C allele polymorphism of *AGTR1* may disrupt binding of miR155 to its complementary *AGTR1* mRNA (33)—resulting in an overall increase in gene expression. In agreement with our finding, Wang et al (34) and Kobashi et al (35) reported this allele in case-control studies to be significantly increased in hypertensive patients compared to normotensive ones (34,35).

In contrast to the above *AGTR1, FVII* and *TGFB1* polymorphisms, other polymorphisms may also influence genetic variation through altered protein structure and resulting protein activity. For

example, it is reported that the rs1800590 polymorhism in *LPL* affects lipoprotein lipase enzyme function (36,37). Similarly, the coding nsSNP rs2228570 in *VDR* introduces an amino acid substitution (Met1Thr) that may be damaging for the structure of its encoded protein, according to Polyphen (21). Thus, the influences of *LPL* and *VDR* polymorphisms on BP may be the result of modifications in their protein structures and resulting changes in protein activities.

Epistatic interactions may also play pivotal roles in the discrepancies found across genetic studies. Numerous studies reported that these interactions may identify genetic markers that are not captured by individual marker analysis tests and/or revealed by the combinatory effect of loci in other pathways (14,38). Herein we showed that all genetic markers individually associated with at least one BP trait also had epistatic interactions influencing BP phenotypes (Table 3). Interestingly, among epistatic interactions, one was common for all BP traits: *LPL* (rs1801177)**PON1* (rs662) (P≤0.032, Table 3) and 2 were specific for SBP and DBP: *VCAM1* (rs1041163)**APOB* (rs1367117) and *SCGB1A1* (rs3741240)**LPL* (rs1800590 and rs1801177) (Table 3A and 3B), respectively.

A number of potential mechanisms mediating the observed genetic associations were identified with an *in silico* search. First, the *PON1* gene encodes the 355 amino acid protein paraoxonase, an enzyme that hydrolyzes paraoxon. The genetic variant rs662 (Arg192Gln) of *PON1* results from the replacement of Arg by Gln of amino acid 192; the protein has been shown to have lower paraoxonase activity (39) resulting in an increase of oxidative stresss (40)—a characteristic closely linked to elevated BP (41).

The genetic variant rs1801177 in *LPL* is also a functional nsSNP, resulting from the conversion of aspartic acid to asparagine of amino acid 9 in the mature protein,. This SNP has been shown to partially affect LPL enzymatic activity (42). Putatively, oxidative stress coupled with partially inactivated LPL would result in elevated BP levels. The positive β coefficients for *LPL*

(rs1801177)*_PON1_ (rs662) (0.072 and 0.057 respectively) strengthened our hypothesis (data available on demand).

Another epistatic interaction was found between _SCGB1A1_ (rs3741240) and _LPL_ (rs1800590). rs3741240 in _SCGB1A1_ is a nonfunctional SNP that has been shown to inhibit phospholipase A2 (43) and decrease plasma levels of HDL-C (44). Taken together, _LPL_ and _SCGB1A1_ may interact through HDL-C to influence BP. The present study also showed epistatic interactions between _VCAM1_ (rs1041163)*_APOB_ (rs1367117). Whereas rs1041163 in _VCAM1_ is a nonfunctional SNP, rs1367117 in _APOB_ is possibly damaging for its protein structure. Human studies have reported significant correlations between APOB, the main component of LDL-C, and VCAM1 in the context of hypertension. Hubel et al (45) have reported significant correlations between soluble levels of APOB and VCAM1 in women with endothelial dysfunction and hypertension as a result of preeclampsia.

Importantly, previous candidate gene studies have revealed other epistatic interactions in modifying the BP phenotype (38,46). In a cross-sectional study conducted on 402 middle aged and elderly Japanese individuals, Misono et al (38) demonstrated that individuals carrying a combination of _ADRB2_ Gly/Gly and _NOS3_ Glu/Glu genotypes had higher hypertensive risk. Moreover, Kohara et al (46) replicated a case-control study conducted on 9,700 participants and reported that epistatic interactions are associated with increased risk of developing hypertension. Combined with our findings, there is clear evidence that numerous epistatic interactions involving multiple physiological pathways influence hypertensive risk.

Strengths and limitations

The present findings are strengthened by the fact that, compared with previous epistatic interaction studies, that the current investigation gives a biological validity in an ethnically homogenous population (French origin). However some limitations should be mentioned: (1) caution must be

taken in interpreting the identified epistatic interactions, as they were observed in the same population in which the genetic markers were discovered and results in different populations might be replicated; (2) consistency of our results should be underwent on a Genome-wide association study design on a large population; and (3) as the participants in this study are of French origin, our findings must be confirmed in different ethnic groups.

Conclusions and perspectives

In summary, we identify (1) susceptibility genes directly influencing BP including *LIPC*, *LPL*, *PON1*, PON2, *APOA4*, *APOB* and *LDLR* (lipid metabolism), *VCAM1* and *SELP* (cellular adhesion molecules), *FVII* (coagulation factors), *AGTR1 and AGT* and *NOS3* (BP regulation), *IL10*, *IL1B*, *C5*, *CCR3*, *CSF2*, *CCL11*, TNF, *TGFB1*, *C3*, *LTA* and *VDR* (inflammation); (2) biologically validated pathways involved in lipid metabolism such as *LPL* (rs1801177) and *PON1* (rs662), *SCGB1A1* (rs3741240) which may modify the risk of hypertension through epistatic interactions. Our findings may provide insight into novel pathophysiologic mechanisms underlying hypertension and potential therapeutic targets. Confirmation in different ethnic groups is currently underway.

Acknowledgments

The data used here are part of the BRC “Interactions Gène-Environnement en Physiopathologie CardioVasculaire” in Nancy, France. The Populations involved, was supported by the “Caisse Nationale d'Assurance Maladies des Travailleurs Salariés”, the “Institut National de la Santé et de la Recherche Médicale” (INSERM), the “Région Lorraine”, the “Communauté Urbaine du Grand Nancy,” and the “Henri Poincaré” University of Nancy I. We are deeply grateful to the cooperation of the individuals participating. We are grateful to Roche molecular system Alameda (California, USA) for their support in the Mutiplex essay genotyping and we thank Chistine Masson for her participation in the performance of the Multiplex genotyping.

References

1. Hua H, Zhou S, Liu Y, et al. Relationship between the regulatory region polymorphism of human tissue kallikrein gene and essential hypertension. J Hum Hypertens 2005;19:715-21.
2. Kearney PM, Whelton M, Reynolds K, Muntner P, Whelton PK, He J. Global burden of hypertension: analysis of worldwide data. Lancet 2005;365:217-23.
3. Levy D, DeStefano AL, Larson MG, et al. Evidence for a gene influencing blood pressure on chromosome 17. Genome scan linkage results for longitudinal blood pressure phenotypes in subjects from the framingham heart study. Hypertension 2000;36:477-83.
4. Lifton RP, Gharavi AG, Geller DS. Molecular mechanisms of human hypertension. Cell 2001;104:545-56.
5. Ehret GB. Genome-wide association studies: contribution of genomics to understanding blood pressure and essential hypertension. Curr Hypertens Rep;12:17-25.
6. Pickering GW, Wright AD, Heptinstall RH. The reversibility of malignant hypertension. Lancet 1952;2:952-6.
7. Chae CU, Lee RT, Rifai N, Ridker PM. Blood pressure and inflammation in apparently healthy men. Hypertension 2001;38:399-403.
8. Sass C, Blanquart C, Morange PE, Pfister M, Visvikis-Siest S. Association between factor VII polymorphisms and blood pressure: the Stanislas Cohort. Hypertension 2004;44:674-80.
9. Savoia C, Volpe M, Alonzo A, Rossi C, Rubattu S. Natriuretic peptides and cardiovascular damage in the metabolic syndrome: molecular mechanisms and clinical implications. Clin Sci (Lond);118:231-40.
10. Androulakis ES, Tousoulis D, Papageorgiou N, Tsioufis C, Kallikazaros I, Stefanadis C. Essential hypertension: is there a role for inflammatory mechanisms? Cardiol Rev 2009;17:216-21.
11. Marteau JB, Zaiou M, Siest G, Visvikis-Siest S. Genetic determinants of blood pressure regulation. J Hypertens 2005;23:2127-43.
12. Sicker T, Wuchold F, Kaiser B, Glusa E. Systemic vascular effects of thrombin and thrombin receptor activating peptide in rats. Thromb Res 2001;101:467-75.
13. Niu W, Qi Y, Qian Y, Gao P, Zhu D. The relationship between apolipoprotein E epsilon2/epsilon3/epsilon4 polymorphisms and hypertension: a meta-analysis of six studies comprising 1812 cases and 1762 controls. Hypertens Res 2009;32:1060-6.
14. Harrap SB. Blood pressure genetics: time to focus. J Am Soc Hypertens 2009;3:231-7.
15. Levy D, Ehret GB, Rice K, et al. Genome-wide association study of blood pressure and hypertension. Nat Genet 2009;41:677-87.
16. Newton-Cheh C, Johnson T, Gateva V, et al. Genome-wide association study identifies eight loci associated with blood pressure. Nat Genet 2009;41:666-76.
17. Ding K, Kullo IJ. Genome-wide association studies for atherosclerotic vascular disease and its risk factors. Circ Cardiovasc Genet 2009;2:63-72.
18. Thomas D. Gene-environment-wide association studies: emerging approaches. Nat Rev Genet;11:259-272.
19. Visvikis-Siest S, Siest G. The STANISLAS Cohort: a 10-year follow-up of supposed healthy families. Gene-environment interactions, reference values and evaluation of biomarkers in prevention of cardiovascular diseases. Clin Chem Lab Med 2008;46:733-47.
20. Altshuler D, Pollara VJ, Cowles CR, et al. An SNP map of the human genome generated by reduced representation shotgun sequencing. Nature 2000;407:513-6.
21. Ramensky V, Bork P, Sunyaev S. Human non-synonymous SNPs: server and survey. Nucleic Acids Res 2002;30:3894-900.

22. Purcell S, Neale B, Todd-Brown K, et al. PLINK: a tool set for whole-genome association and population-based linkage analyses. Am J Hum Genet 2007;81:559-75.
23. Chobanian AV, Bakris GL, Black HR, et al. The Seventh Report of the Joint National Committee on Prevention, Detection, Evaluation, and Treatment of High Blood Pressure: the JNC 7 report. Jama 2003;289:2560-72.
24. Gong M, Hubner N. Molecular genetics of human hypertension. Clin Sci (Lond) 2006;110:315-26.
25. Camp NJ, Hopkins PN, Hasstedt SJ, et al. Genome-wide multipoint parametric linkage analysis of pulse pressure in large, extended utah pedigrees. Hypertension 2003;42:322-8.
26. Cooper RS, Luke A, Zhu X, et al. Genome scan among Nigerians linking blood pressure to chromosomes 2, 3, and 19. Hypertension 2002;40:629-33.
27. Harrap SB. Where are all the blood-pressure genes? Lancet 2003;361:2149-51.
28. Montasser ME, Shimmin LC, Hanis CL, Boerwinkle E, Hixson JE. Gene by smoking interaction in hypertension: identification of a major quantitative trait locus on chromosome 15q for systolic blood pressure in Mexican-Americans. J Hypertens 2009;27:491-501.
29. North KE, Rose KM, Borecki IB, et al. Evidence for a gene on chromosome 13 influencing postural systolic blood pressure change and body mass index. Hypertension 2004;43:780-4.
30. Rice T, Rankinen T, Province MA, et al. Genome-wide linkage analysis of systolic and diastolic blood pressure: the Quebec Family Study. Circulation 2000;102:1956-63.
31. Salam MT, Gauderman WJ, McConnell R, Lin PC, Gilliland FD. Transforming growth factor- 1 C-509T polymorphism, oxidant stress, and early-onset childhood asthma. Am J Respir Crit Care Med 2007;176:1192-9.
32. Silverman ES, Palmer LJ, Subramaniam V, et al. Transforming growth factor-beta1 promoter polymorphism C-509T is associated with asthma. Am J Respir Crit Care Med 2004;169:214-9.
33. Martin MM, Buckenberger JA, Jiang J, et al. The human angiotensin II type 1 receptor +1166 A/C polymorphism attenuates microrna-155 binding. J Biol Chem 2007;282:24262-9.
34. Wang WY, Zee RY, Morris BJ. Association of angiotensin II type 1 receptor gene polymorphism with essential hypertension. Clin Genet 1997;51:31-4.
35. Kobashi G, Hata A, Ohta K, et al. A1166C variant of angiotensin II type 1 receptor gene is associated with severe hypertension in pregnancy independently of T235 variant of angiotensinogen gene. J Hum Genet 2004;49:182-6.
36. Mailly F, Fisher RM, Nicaud V, et al. Association between the LPL-D9N mutation in the lipoprotein lipase gene and plasma lipid traits in myocardial infarction survivors from the ECTIM Study. Atherosclerosis 1996;122:21-8.
37. Mailly F, Tugrul Y, Reymer PW, et al. A common variant in the gene for lipoprotein lipase (Asp9-->Asn). Functional implications and prevalence in normal and hyperlipidemic subjects. Arterioscler Thromb Vasc Biol 1995;15:468-78.
38. Misono M, Maeda S, Iemitsu M, et al. Combination of polymorphisms in the beta2-adrenergic receptor and nitric oxide synthase 3 genes increases the risk for hypertension. J Hypertens 2009;27:1377-83.
39. Nevin DN, Zambon A, Furlong CE, et al. Paraoxonase genotypes, lipoprotein lipase activity, and HDL. Arterioscler Thromb Vasc Biol 1996;16:1243-9.
40. Marsillach J, Checa MA, Pedro-Botet J, Carreras R, Joven J, Camps J. Paraoxonase-1 in female infertility: a possible role against oxidative stress-induced inflammation. Fertil Steril.
41. Gkaliagkousi E, Douma S, Zamboulis C, Ferro A. Nitric oxide dysfunction in vascular endothelium and platelets: role in essential hypertension. J Hypertens 2009;27:2310-20.
42. Rouis M, Lohse P, Dugi KA, et al. Homozygosity for two point mutations in the lipoprotein lipase (LPL) gene in a patient with familial LPL deficiency: LPL(Asp9-->Asn, Tyr262-->His). J Lipid Res 1996;37:651-61.

43. Chowdhury B, Mantile-Selvaggi G, Miele L, Cordella-Miele E, Zhang Z, Mukherjee AB. Lys 43 and Asp 46 in alpha-helix 3 of uteroglobin are essential for its phospholipase A2 inhibitory activity. Biochem Biophys Res Commun 2002;295:877-83.
44. Tietge UJ, Maugeais C, Lund-Katz S, Grass D, deBeer FC, Rader DJ. Human secretory phospholipase A2 mediates decreased plasma levels of HDL cholesterol and apoA-I in response to inflammation in human apoA-I transgenic mice. Arterioscler Thromb Vasc Biol 2002;22:1213-8.
45. Hubel CA, Lyall F, Weissfeld L, Gandley RE, Roberts JM. Small low-density lipoproteins and vascular cell adhesion molecule-1 are increased in association with hyperlipidemia in preeclampsia. Metabolism 1998;47:1281-8.
46. Kohara K, Tabara Y, Nakura J, et al. Identification of hypertension-susceptibility genes and pathways by a systemic multiple candidate gene approach: the millennium genome project for hypertension. Hypertens Res 2008;31:203-12.

Table 1: Phenotypic characteristics of the studied population.

Characteristics	Total (n = 1,912)	Men (n = 932)	Women (n = 980)	p-value
Age (years)	51.13 ± 10.02	51.21 ± 9.31	51.06 ± 10.66	0.740
BMI (kg/m²)	26.01 ± 3.75	26.70 ± 3.26	25.35 ± 4.06	<0.001
SBP (mmHg)	142.86 ± 21.47	144.62 ± 19.69	141.18 ± 22.92	<0.001
DBP (mmHg)	87.49 ± 14.26	89.10 ± 13.10	85.96 ± 15.14	<0.001
PP (mmHg)	62.18 ± 15.11	61.89 ± 14.34	62.46 ± 15.81	0.414
Baseline BP category (%)				
< 120/80	18.5	12.4	24.3	
120-139/80-89	21.1	25.3	17.1	
140-159/90-99	25.9	26.5	25.3	
≥ 160/100	34.5	35.7	33.3	
Smoking (%)				
Current	18.3	22.8	14.1	
Former	20.6	30.6	11.0	
Never	61.1	46.6	74.9	
Total cholesterol (mmol/L)	5.89 ± 1.00	5.93 ± 0.97	5.85 ± 1.01	0.082
LDL-cholesterol (mmol/L)	3.60 ± 0.88	3.70 ± 0.88	3.52 ± 0.88	<0.001
HDL-cholesterol (mmol/L)	1.64 ± 0.43	1.50 ± 0.40	1.77 ± 0.41	<0.001
Triglycérides (mmol/L)	1.37 ± 1.20	1.55 ± 1.00	1.20 ± 1.34	<0.001
Glucose (mmol/L)	5.25 ± 0.78	5.41 ± 0.83	5.10 ± 0.71	<0.001

BMI: body mass index, SBP: systolic blood pressure, DBP: diastolic blood pressure, PP: pulse pressure

Table 2: Genetic variants associated with blood pressure.

2A – Systolic Blood Pressure

Gene	SNP ID	MAF	p-value (snp)	bêta	p-value (10)	p-value (sex)	p-value (bmi)
LIPC	rs1800588	20.958	0.010	0.055	<0.001	0.182	<0.001
LPL	rs1800590	2.586	<0.001	0.076	<0.001	0.198	<0.001
PON1	rs854560	36.391	0.047	0.044	<0.001	0.096	<0.001
PON1	rs662	30.294	0.008	0.059	<0.001	0.520	<0.001
PON2	rs6954345 / rs7493	23.462	0.056	0.043	<0.001	0.043	<0.001
LDLR	rs5742911	32.308	0.009	0.06	<0.001	0.158	<0.001
AGT1R	rs5186	28.933	0.026	-0.051	<0.001	0.513	<0.001
FVII	rs5742910	15.024	0.003	-0.028	<0.001	0.137	<0.001
FVII	rs6046	12.874	0.001	-0.028	<0.001	0.254	<0.001
VCAM1	rs1041163	18.006	0.006	0.062	<0.001	0.044	<0.001
SELP	rs6133	13.943	0.050	0.024	<0.001	0.867	<0.001
TGFB1	rs1800469	29.048	0.000	-0.026	<0.001	0.051	<0.001
IL1B	rs1143634	19.397	0.002	-0.044	<0.001	0.842	<0.001
C5	rs17611	40.759	0.055	0.049	<0.001	0.150	<0.001
CSF2	rs25882	25.596	<0.001	0.124	<0.001	0.300	<0.001
CCL11	rsrs1129844	13.778	<0.001	-0.119	<0.001	0.818	<0.001
LTA	rs909253	31.859	0.053	0.037	<0.001	0.331	<0.001
VDR	rs2228570	40.139	<0.001	-0.012	<0.001	0.155	<0.001

2B – Diastolic Blood Pressure

Gene	SNP ID	MAF	p-value (snp)	bêta	p-value (10)	p-value (sex)	p-value (bmi)
APOA4	rs675	18.536	0.024	-0.048	<0.001	0.237	<0.001
APOB	rs1367117	31.301	0.024	-0.029	<0.001	0.031	<0.001
APOC3	rs2542052	37.970	0.014	-0.045	<0.001	0.001	<0.001
APOC3	rs2854116	37.017	0.024	-0.038	<0.001	0.001	<0.001
LIPC	rs1800588	20.958	0.058	0.047	<0.001	0.032	<0.001
LPL	rs1800590	2.586	0.003	0.076	<0.001	0.565	<0.001
PON1	rs662	30.294	0.020	0.056	<0.001	0.032	<0.001
LDLR	rs5742911	32.308	0.002	0.071	<0.001	0.008	<0.001
AGT1R	rs5186	28.933	0.050	-0.045	<0.001	0.254	<0.001
FVII	rs5742910	15.024	0.001	-0.042	<0.001	0.068	<0.001
FVII	rs6046	12.874	<0.001	-0.045	<0.001	0.110	<0.001
TGFB1	rs1800469	29.048	<0.001	-0.038	<0.001	0.002	<0.001
IL1B	rs1143634	19.397	0.005	-0.052	<0.001	0.948	<0.001
CSF2	rs25882	25.596	<0.001	0.148	<0.001	0.013	<0.001
CCL11	rsrs1129844	13.778	<0.001	-0.142	<0.001	0.807	<0.001
TNF	rs361525	6.830	0.010	0.053	<0.001	1.000	<0.001
VDR	rs2228570	40.139	<0.001	-0.009	<0.001	0.037	<0.001

2C – Pulse Pressure

Gene	SNP ID	MAF	p-value (snp)	bêta	p-value (10)	p-value (sex)	p-value (bmi)
LPL	rs1800590	2.586	<0.001	0.06	<0.001	0.311	<0.001
LDLR	rs5742911	32.308	0.007	0.068	<0.001	0.004	<0.001
AGT	rs699	42.978	0.007	0.069	<0.001	0.007	<0.001
GNB3	rs5443	33.533	0.034	0.006	<0.001	0.017	<0.001
FVII	rs5742910	15.024	0.007	-0.029	<0.001	0.098	<0.001
FVII	rs6046	12.874	0.024	-0.032	<0.001	0.309	<0.001
TGFB1	rs1800469	29.048	<0.001	-0.016	<0.001	0.032	<0.001
IL1B	rs1143634	19.397	0.004	-0.05	<0.001	0.269	<0.001
C3	rs2230199	22.708	0.031	-0.014	<0.001	0.232	<0.001
CSF2	rs25882	25.596	<0.001	0.094	<0.001	0.146	<0.001
NOS3	rs1799983	35.775	0.051	0.046	<0.001	0.011	<0.001
CCL11	rsrs1129844	13.778	<0.001	-0.118	<0.001	0.028	<0.001
LTA	rs909253	31.859	0.005	0.05	<0.001	0.005	<0.001
TNF	rs1800629	13.316	0.013	0.026	<0.001	0.245	<0.001
TNF	rs361525	6.830	0.002	0.056	<0.001	0.129	<0.001
VDR	rs2228570	40.139	<0.001	-0.034	<0.001	0.014	<0.001

Table 3: Epistatic interactions between Genetic variants associated with blood pressure.

3A – Systolic Blood Pressure

			APOA4	*APOB*	*LIPC*	*LPL*	*LPL*	*PON1*	*AGTR1*	*AGT*	*GNB3*	*FVII*	*FVII*	*TNF*	*ICAM1*	*VCAM1*	*CSF2*
			rs5110	rs1367117	rs1800588	rs1800590	rs1801177	rs662	rs5186	rs699	rs5443	rs5742910	rs6046	rs1800629	rs5491	rs1041163	rs25882
		MAF	8.01	31.3	21.1	2.6	1.82	30.3	28.9	42.9	33.5	15.1	12.87	13.1	0.56	18.01	25.6
PON1	rs662	30.3				0.03	0.017										
GNB3	rs5443	33.5			0.05					0.039							
FVII	rs6046	12.87							0.039								
ICAM1	rs5491	0.56		0.034													
VCAM1	rs1041163	18.01		0.04													
SCGB1A1	rs3741240	46.9				0.02	0.013						0.033				
TGFB1	rs1800469	29.1													0.05	0.03	
CSF2	rs25882	25.6									0.003						
LTC4S	rs730012	44.8				0.03											
TNF	rs1800629	13.3					0.02		0.04			0.05		0.002			
TNF	rs361525	6.8															0.02
VDR	rs2228570	40.1	0.01					0.039			0.03						

3B – Diastolic Blood Pressure

			APOB	*APOC3*	*APOC3*	*LPL*	*LPL*	*PON1*	*AGT*	*GNB3*	*FVII*	*CCL11*	*TNF*	*ICAM1*	*VCAM1*
			rs1367117	rs2542052	rs2854116	rs1800590	rs1801177	rs662	rs699	rs5443	rs5742910	rsrs1129844	rs1800629	rs5491	rs1041163
		MAF	31.3	37.9	37.01	2.6	1.82	30.3	42.9	33.5	15.02	28.1	13.1	0.56	18.01
PON1	rs662	30.3					0.032								
GNB3	rs5443	33.5							0.044						
FVII	rs5742910	15.02		0.022	0.035										
VCAM1	rs1041163	18.01	0.027						0.032					0.048	
SCGB1A1	rs3741240	46.9				0.03	0.026								
TGFB1	rs1800469	29.1													0.004
TNF	rs1800629	13.3									0.021		0.003		
VDR	rs2228570	40.1						0.044				0.036			

3C – Pulse Pressure

			APOA4	*APOB*	*APOC3*	*LPL*	*LPL*	*LDLR*	*GNB3*	*FVII*	*TNF*	*VCAM1*
			rs675	rs1367117	rs2854116	rs1800590	rs1801177	rs5742911	rs5443	rs6046	rs1800629	rs1041163
		MAF	18.5	31.3	37.01	2,600	1.82	32.3	33.5	12.87	13.1	18.01
PON1	rs662	30.3				0.041	0.007					
PON2	rs6954345	23.5	0.048									
AGTR1	rs5186	28.9		0.026								
FVII	rs6046	12.87										
VCAM1	rs1041163	18.01							0.039			
SCGB1A1	rs3741240	46.9								0.01		0.024
CSF2	rs25882	25.6			0.047							
LTC4S	rs730012	44.7						0.009				
TNF	rs1800629	13.1									<0.001	
VDR	rs2228570	40.1										

PUBLICATION N°7

Stéphanie Debette*, Sophie Visvikis-Siest*, Ming Hue Chen, **Ndeye Coumba Ndiaye**, C Song, AL DeStefano, R Safa, Mohsen Azimi Nezhad, D Sawyer, Jean-Brice Marteau, V Xanthakis, Gérard Siest, L Sullivan, Michèle Pfister, H Smith, SH Choi, J Lamont, L Lind, Q Yang, P Fitzgerald, E Ingelsson, Ramachandran S Vasan, Sudha Seshadri. Identification of Cis- and Trans-acting Genetic Variants Explaining up to Half the Heritability of Circulating VEGF levels. 2010 [soumis]

Le VEGF a un impact sur l'angiogenèse, l'athérosclérose et le cancer. Malgré sa forte héritabilité, sa génétique est peu connue. Notre objectif était donc d'identifier les variants génétiques associés aux niveaux circulants du VEGF par une approche pangénomique.

Dans un premier temps, nous avons effectué une GWAS sur 3 546 sujets issus de la troisième génération de la cohorte Framingham puis une réplication dans 1 727 sujets issus de deux populations indépendantes : la cohorte STANISLAS et la Prospective Investigation of the Vasculature in Uppsala Seniors study (PIVUS). Quatre variants génétiques situés dans trois régions chromosomiques distinctes avaient une association indépendante avec le VEGF au terme de la régression conditionnée : rs6921438 et rs4416670 (6p21.1, $p=6,11x10^{-506}$ et $p=1,47x10^{-12}$), rs6993770 (8q23.1, $p=2,50x10^{-16}$) et rs10738760 (9p24.2, $p=1,96x10^{-34}$). Ces variants expliquaient 48% de la variabilité génétique du VEGF.

Dans une étude transcriptomique, six variants genome-wide significant ont été associés aux ARN messager du VEGF dans les cellules mononucléaire de sang périphérique.

Ainsi, nous avons pu mettre en évidence des variants génétiques expliquant prés de la moitié de l'héritabilité génétique du VEGF et donnant des pistes importantes sur les voies métaboliques régulant ce trait.

Contribution :

Cette étude a été effectuée en collaboration avec l'équipe Framingham et la cohorte STANISLAS était une population de réplication. Nous avons aidé à la sélection et à la

préparation des échantillons de cette population pour les réplications. Nous avons également préparé le fichier de transcriptomique.

Nous avons participé à la recherche bibliographique, aux différentes conférences téléphoniques et échanges donnant la direction du manuscrit ainsi qu'à la rédaction de ce papier pour lequel nous sommes troisième auteur.

Identification of Cis- and Trans-acting Genetic Variants Explaining up to Half the Heritability of Circulating VEGF levels

Debette et al.

Short Title: Genetic correlates of circulating VEGF levels

Debette S (MD, PhD)[1,2]*, Visvikis-Siest S (PhD)[3]*, Chen MH (PhD)[1,2], Ndiaye NC (MS)[3], Song C (MS)[4], DeStefano AL (PhD)[1,2,5], Safa R (PhD)[6], Azimi Nezhad M (MS)[3], Sawyer D (MD, PhD)[7], Marteau JB (PhD)[3], Xanthakis V (MS)[5], Siest G (PhD)[3], Sullivan L (PhD)[5], Pfister M (PhD)[3], Smith H (MS)[7], Choi SH (BA)[1,5], Lamont J (MSc)[8], Lind L (MD, PhD)[9], Yang Q (PhD)[1,2,5], Fitzgerald P (PhD, DSc)[8], Ingelsson E (MD, PhD)[4], Vasan RS (MD)[1,2,10]**, Seshadri S (MD)[1,2]**

* Drs. Debette and Visvikis-Siest contributed equally as first authors, ** Drs. Vasan and Seshadri contributed equally as last authors.

Correspondence to:

Sudha Seshadri, MD

Associate Professor, Department of Neurology

Boston University School of Medicine; B602, 72 East Concord Street, Boston, MA 02118

Phone: +1-617-414-1337; fax: +1-617-638-8086; email: suseshad@bu.edu

Authors Affiliations

[1] Department of Neurology, Boston University School of Medicine, MA, USA; [2] National Heart Lung and Blood Association's Framingham Heart Study, [3] "Cardiovascular Genetics" Research Unit, EA4373,

Université Henri Poincaré - Nancy 1, 54000 Nancy, France; [4] Department of Medical Epidemiology and Biostatistics, Karolinska Institutet Box 281 SE-171 77, Stockholm, Sweden; [5] Department of Biostatistics, Boston University, School of Public Health, Boston, MA, USA; [6] Division of Graduate Medical Sciences, Boston University, Boston, MA, USA; [7]Cardiovascular Division, Vanderbilt University, Nashville, TN, USA; [8] Randox Laboratories Ltd., Crumlin, Antrim, UK; [9]Department of Medical Sciences, Uppsala University Hospital, SE-751 85 Uppsala, Sweden; [10] Sections of Preventive Medicine and Cardiology, Evans Memorial Department of Medicine, Boston University, School of Medicine, Boston, MA, USA

Abstract word count: 231

Manuscript word count: 5998

Background: VEGF impacts angiogenesis, atherosclerosis and cancer. Although the heritability of circulating VEGF levels is high, little is known about its genetic underpinnings. Our aim was to identify genetic variants associated with circulating VEGF levels using an unbiased genome-wide approach.

Methods and results: We undertook a genome-wide association study (GWAS) of serum VEGF levels in 3,546 participants of the Framingham Heart Study (FHS), with pre-planned replication in 1,727 participants from two independent samples, the STANISLAS Family Study (SFS) and the Prospective Investigation of the Vasculature in Uppsala Seniors study (PIVUS). One hundred and forty SNPs reached genome-wide significance ($p<5x10^{-8}$). We found evidence of replication for the most significant associations in both replication datasets. In a conditional GWAS 4 SNPs mapping to 3 chromosomal regions were independently associated with circulating VEGF levels: rs6921438 and rs4416670 (6p21.1, $p=6.11x10^{-506}$ and $p=1.47x10^{-12}$), rs6993770 (8q23.1, $p=2.50x10^{-16}$) and rs10738760 (9p24.2, $p=1.96x10^{-34}$). A genetic liability score including these four SNPs explained 48% of the heritability of serum VEGF levels. Six of the SNPs that reached genome-wide significance in the GWAS were significantly associated with VEGF mRNA levels in PBMCs.

Conclusions: Genetic variants explaining up to half the heritability of serum VEGF levels were identified. These new insights provide important clues to the pathways regulating circulating VEGF levels and could be used to study whether emerging VEGF and anti-VEGF based therapies should be tailored according to the genetic background.

Vascular endothelial growth factor (VEGF, also referred to as VEGFA in contrast to other members of the VEGF family) is pivotal in many physiological and pathological processes.[1] It is primarily known for its key role in the stimulation of angiogenesis, with a potent mitogenic effect on vascular endothelial cells from arteries, veins and lymphatics.[2] VEGF also promotes vasodilatation by inducing the production of nitric oxide and prostacyclin by endothelial cells.[3] In addition, VEGF is involved in hematopoietic development and chemotaxis of monocytes, regulation of osteoclast differentiation, stimulation of surfactant production,[1] and has neurotrophic and neuroprotective effects on neuronal and glial cells.[4] Elevated circulating VEGF levels have been observed in vascular disease (ischemic heart disease,[5, 6] heart failure,[7] stroke[8]), and in various other disorders, including diabetes,[9] cognitive decline and dementia,[10, 11] reproductive,[12-14] immune-inflammatory disorders,[1516] and neoplastic diseases.[17, 18] Administration of VEGF promotes angiogenesis in patients with critical leg ischemia, as well as in animal models of coronary and limb ischemia.[19] VEGF inhibitors such as bevacizumab and sorafenib have been successfully used to inhibit angiogenesis in several tumors,[20, 21] in macular degeneration[22] and in rheumatoid arthritis.[23] However, despite the considerable toxicity associated with VEGF inhibitor drugs,[24] there have been no pharmacogenomic studies to identify potential subgroups of responders partly because the genetic determinants of VEGF concentrations remain poorly understood.
Indeed, although the heritability of circulating VEGF levels is very high, ranging between 60 and 80%,[25-27] few studies have assessed the relation between circulating VEGF levels and genetic variants, yielding inconsistent results. The aim of the present study was to identify genetic variants associated with circulating VEGF levels using an unbiased genome-wide approach in a large community-based sample.

Methods

Study populations

The Framingham Heart Study (FHS)

The FHS, initiated in 1948, is a three-generation, community-based, prospective cohort study conducted in Framingham, MA, USA.[33-35] Serum VEGF levels were measured in third generation cohort participants (2002-2005) and genome-wide genotyping was performed on these individuals at Affymetrix (Santa Clara, CA) through an NHLBI funded SNP-Health Association Resource (SHARe) project. After excluding participants who had prevalent cardiovascular disease or failed to meet quality control standards, 3,546 participants were enrolled.

The STANISLAS Family Study (SFS)

The SFS is a 10-year longitudinal survey involving 1,006 volunteer families from Vandoeuvre-lès-Nancy, France, whose members were free of chronic disease (cardiovascular or cancer) between 1993-1995.[28] Plasma VEGF levels were measured at the second examination cycle (1998-2000) in a randomly selected subsample; of these 859 persons from 217 families, who also had DNA and met genotyping quality control criteria, were included.

Prospective Investigation of the Vasculature in Uppsala Seniors (PIVUS) study

The PIVUS study is a population-based study that enrolled 1,016 70-year old individuals living in the community of Uppsala, Sweden (2001-2004).[29] Of these, 999 persons provided DNA for genetic studies and after exclusions for prevalent cardiovascular disease and inadequate genotyping quality, 868 participants were eligible.

Further details of the study samples are available in the **Supplemental Methods**, section 1.

Laboratory measurements of VEGF levels

VEGF levels were measured in serum for the FHS and PIVUS and plasma for the SFS (**Supplemental Methods**, section 2).

Genotyping

FHS

Genome-wide genotyping in the FHS was performed on the Affymetrix GeneChip Human Mapping 500K Array Set® and 50K Human Gene Focused Panel®. Genotyping, quality control and imputation methods are detailed in the **Supplemental Methods**, section 3.

SFS

The SNPs were genotyped by Genoscreen using a Sequenom® iPLEX Gold assay – Medium Throughput Genotyping Technology.[30]

PIVUS study

The SNPs were genotyped as part of a 96-plex assay at the SNP technology platform in Uppsala University (http://www.genotyping.se/) using the Illumina BeadXpress system from Illumina Inc.[31] Genotyping calls were done with Illumina BeadStudio software.

Statistics

VEGF levels were natural log-transformed to normalize their distribution.

Genome-wide association analysis in the FHS

A linear mixed effects model accounting for familial relatedness was used to evaluate the association of each SNP with VEGF levels.[32] An additive genetic model with one degree of freedom was used. In a first step (model A), analyses were adjusted for age, sex, familial structure and the ninth principal component (**Supplemental Methods**, section 4). In a second step designed to explore potential mechanisms, we additionally adjusted our most significant associations for covariates previously found to be associated with serum VEGF levels:[25] compared to model A, model B was additionally adjusted

for hypertension; model C for smoking; model D for central obesity and model E for the presence of a metabolic syndrome (**Supplemental Methods**, section 5).

Genetic association study in the SFS and the PIVUS study

In order to confirm our findings in the FHS, we genotyped 25 SNPs in two independent samples. To select a parsimonious number of SNPs for replication we used criteria of strength of association (p-value), whether the SNP was genotyped or imputed, linkage disequilibrium (LD) between SNPs and functionality (**Supplemental Methods**, section 6). A linear regression model using the same covariates and analytic strategy as in the FHS was implemented.

Joint analysis of the FHS, SFS and PIVUS study

For SNPs that were successfully genotyped in the SFS and the PIVUS study we performed a meta-analysis of the SNP-phenotype associations, using a fixed effects inverse-variance meta-analysis technique (**Supplemental Methods**, section 7).

Genetic liability score

The methods used for computing a genetic liability score are detailed in the **Supplemental Methods**, section 8, and **Supplemental Table 1**. The phenotypic variance explained by this genetic liability score was separately calculated in the FHS, the SFS and the PIVUS study, using regression models that included age and sex as covariates.

VEGF gene expression analysis in peripheral blood mononuclear cells (PBMCs)

Sample preparation and quantification of the PBMC messenger RNA (mRNA) of *VEGF* spliced forms and statistical analyses of these data are described in the **Supplemental Methods** (section 10).

Biological pathway analysis

Methods for the biological pathway analysis are provided in the **Supplemental Methods**, section 9.

Results

Characteristics of the 5,273 study participants are presented in **Table 1**.

GWAS of VEGF levels in the FHS

The quantile-quantile plot showed an excess of extreme p-values but no evidence of systematic inflation of the genomic control inflation factor (λ=1.02) (**Supplemental Figure 1**). The genome-wide plot of p-values for the individual SNPs against their genomic position is shown in **Supplemental Figure 2**. A total of 140 SNPs cleared the threshold for genome-wide significance at $5x10^{-8}$ (**Supplemental Table 2**). These were located in three chromosomal regions: 6p21.1, 8q23.1, 9p24.2 (**Table 2**). The most significant association was found with rs6921438 on chromosome 6p21.1 ($p=6.11x10^{-506}$), at 171 kb downstream of the *VEGF* gene, and close to the *mitochondrial ribosomal protein L14* gene (*MRPL14*) and the *MCG45491* gene (*C6orf223*), encoding an uncharacterized protein. Sixty-seven other SNPs on chromosome 6p21.1 were also associated with VEGF levels at $p<5x10^{-8}$ (**Figure 1a**). When running a conditional GWAS adjusting for rs6921438, one other SNP in 6p21.1 (rs4416670) still yielded a genome-wide significant association, suggesting that two variants in this region independently modulate VEGF levels. In the 8q23.1 region the SNP yielding the most significant association with VEGF levels (rs6993770, $p=2.50x10^{-16}$) is located in the *zinc finger protein, multitype 2 (ZFPM2)* gene and 980.4 kb away from the *low-density lipoprotein receptor-related protein gene (LRP12*). Forty-three SNPs in LD with rs6993770 were also associated with VEGF levels at $p<5x10^{-8}$ (**Figure 1b**). A conditional GWAS adjusting for rs6993770, rs6921438 and rs4416670 did not yield any other genome-wide significant association in chromosome 8q23.1. The most significant association on 9p24.2 was observed with rs10738760 ($p=1.96x10^{-34}$), located close to the *very low density lipoprotein receptor (VLDLR)* and *potassium voltage-gated channel subfamily V, member 2 (KCNV2)* genes. Twenty-nine SNPs in LD with rs10738760 were also associated with VEGF levels at $p<5x10^{-8}$ (**Figure 1c**). None reached genome-

wide significance in a conditional GWAS adjusted for rs6921438, rs4416670, rs6993770 and rs10738760.

We computed a genetic liability score including the four SNPs yielding genome-wide significant associations with VEGF levels in the conditional GWAS (**Supplemental Table 1**). This score explained 47.6% of serum VEGF variability ($p=2.19x10^{-644}$).

Replication studies

We sought to replicate our most significant results in two independent cohorts. Of the 25 SNPs selected for replication, 24 were successfully genotyped in the SFS and 20 in the PIVUS study (**Table 2**). Of these, 17 and 20 respectively reached nominal significance in association with VEGF levels (**Table 2**). When meta-analyzing the results of both replication cohorts for the 19 SNPs genotyped in both studies, all 19 SNPs were associated with VEGF levels at $p<0.05$ (**Table 2**). The joint meta-analysis of results from the discovery and replication cohorts is displayed in **Table 2**. The genetic liability score explained 16.6% ($p=1.75x10^{-36}$) of observed plasma VEGF variability in the SFS and 48.4% ($p=3.31x10^{-180}$) of observed serum VEGF variability in the PIVUS study.

The observed associations remained unchanged in each of the three cohorts after adjusting for hypertension, current smoking, central obesity and metabolic syndrome (**Supplemental Table 3**).

VEGF gene expression analysis

In order to better characterize the functional role of the SNPs identified in the GWAS we quantified mRNA expression of the two splice variants corresponding to the diffusible isoforms of VEGF, $VEGF_{121}$ and $VEGF_{165}$, in PBMCs of 220 SFS participants. The association of VEGF mRNA levels with the 24 SNPs successfully genotyped in the SFS was assessed.

At the nominal significance level, 1 SNP on chromosome 6p21.1, 4 SNPs on chromosome 8q23.1 and 1 SNP on chromosome 9p24.2 were associated with $VEGF_{121}$ mRNA levels (**Table 3**).

Biological pathway analysis

Using the Ingenuity Pathway Analysis software (IPA, Ingenuity Systems, www.ingenuity.com) we explored functional relationships between *VEGF* and the genes closest to the SNPs on chromosome 8q23.1 and 9p24.2 that were significantly associated with circulating VEGF levels. In each case we selected the genes closest to the identified SNPs, as in **Table 2**, to identify plausible biological pathways. We selected five focus genes: *VEGF, ZFPM2, LRP12, VLDLR, KCNV2*. The IPA network analysis identified relationships among three of these five focus genes (*VEGF*, *ZFPM2* and *VLDLR*) as part of a larger network. The presence of three of our five focus genes in this network was unlikely to occur by chance ($p=10^{-8}$). In **Figure 2** we present a subset of this network, including only interactions between *VEGF* and the two other focus genes in the network, with 2 or fewer intermediate nodes.

Discussion

Principal findings

In this first GWAS of circulating VEGF levels undertaken in 3,546 community individuals of European descent, we identified novel genetic associations: 140 SNPs reached genome-wide significance. Of these, 4 SNPs were independently associated with VEGF levels (rs6921438 and rs4416670 on chromosome 6p21.1, rs6993770 on chromosome 8q23.1 and rs10738760 on chromosome 9p24.2). We found evidence of replication for selected SNPs in 1,727 individuals of European descent from two independent community-based samples. The SNPs are located close to the *VEGF* and *MRPL14* genes *(*chromosome 6p21.1), within the *ZFPM2* gene (chromosome 8q23.1), and between the *VLDLR* and *KCNV2* genes (chromosome 9p24.2). In a subset of participants we found that 6 of 25 selected SNPs yielding genome-wide significant associations with circulating VEGF levels were also associated with *VEGF* mRNA levels ($VEGF_{121}$ splice variant) in PBMCs.

In the context of the current literature

Genetic association study

Candidate gene studies exploring associations between *VEGF* polymorphisms and circulating VEGF levels have yielded controversial results (**Supplemental Methods**, section 11; **Supplemental Table 4**). Eight studies have found significant associations with candidate polymorphisms (rs699947, rs1570360, rs833061, rs2010963, rs3025039 and -2549 18bp I/D) in the promoter, 5' and 3' untranslated region of the *VEGF* gene.[33-40] However, several other studies did not identify any association with these and other *VEGF* SNPs (**Supplemental Table 4**). Using a hypothesis-free genome-wide approach, the present study revealed novel associations with 140 SNPs. Of these, 68 SNPs are located on chromosome 6 approximately 150 kb downstream from the 3' end of the *VEGF* gene, far from previously tested candidate SNPs. Although we do replicate previously described associations of 2 *VEGF* promoter

polymorphisms (rs699947 and rs833061) with serum VEGF levels at $p<5x10^{-7}$ (**Supplemental Methods**, section 11; **Supplemental Table 4**; **Supplemental Figure 3**), none of the SNPs that reached genome-wide significance in our analysis, on chromosome 6p21.1, 8q23.1 and 9p24.2, had been examined previously in relation with circulating VEGF levels.

Transcriptomic analysis

While several studies have examined the association of candidate genetic variants with *VEGF* gene expression in pathological tissues,[41, 42] little is known about the genetic variants influencing *VEGF* expression in normal cells. Our data suggest that six of the SNPs associated with circulating VEGF levels in our GWAS also modulate the expression of the $VEGF_{121}$ splice variant in PBMCs of community-based persons. The diffusible VEGF isoforms, $VEGF_{165}$ and $VEGF_{121}$, are released by a variety of tumor and normal cells, including PBMCs.[43] $VEGF_{121}$ lacks a heparin-binding domain and has a higher migration but lower mitogenic potency than $VEGF_{165}$.[44]

Potential mechanisms mediating observed genetic associations

Our data suggest that almost half the inherited component of circulating VEGF levels is explained by genetic variants located downstream from the *VEGF* gene on chromosome 6p21.1. The conditional GWAS suggests that this region could harbor at least two distinct loci that are independently associated with circulating VEGF levels. Although located relatively far from the *VEGF* gene, results from our transcriptomic analysis indicate that this region could indeed contain functional variants modulating *VEGF* gene expression.

Genome-wide significant associations with circulating VEGF levels were also identified for SNPs located on chromosome 8q23.1 and 9p24.2. Although these *trans* effects explain a much smaller proportion of the heritability of VEGF levels, they provide important clues about the pathways involved in the regulation of *VEGF* expression. The SNPs on chromosome 8q23.1 are located in introns 4 and 5 of the *ZFPM2* gene. This gene encodes a widely expressed member of the Friend of GATA family of

transcription factors that modulate the activity of the GATA family proteins, which are important regulators of embryogenesis and also seem to play a significant role in endothelial cell biology.[45-47] The second closest gene to the SNPs identified on chromosome 8q23.1 is *LRP12*, encoding a low-density lipoprotein receptor-related protein that interacts with proteins related to signal transduction pathways and is differentially expressed in many cancer cells. The SNPs on chromosome 9p24.2 are located between the *VLDLR* and *KCNV2* genes. *VLDLR* encodes a lipoprotein receptor involved in the metabolism of apolipoprotein-E-containing triacylglycerol-rich lipoproteins. Like *VEGF, VLDLR* appears to modify the risk of developing age-related macular degeneration,[48] and recent data suggest that *VLDLR* could play a central role in a network of interacting angiogenic genes activated in response to hypoxia.[49] *KCNV2* encodes a member of the potassium voltage-gated channel subfamily V involved in regulation of neurotransmitter release, neuronal excitability and heart rate. While the present data does not permit us to formally determine which of these genes underlie the observed SNP associations with circulating VEGF levels, our *in silico* biological pathway analysis suggests that *ZFPM2* and *VLDLR* are the most likely candidates. Further research is needed to explore the mechanisms underlying the associations of *cis* and *trans* acting genetic variants with circulating VEGF levels, such as modulation of gene expression, differential splicing or mRNA degradation.

Strengths and limitations

The findings from this first GWAS of circulating VEGF levels emphasize the importance of screening for genetic variation modulating biomarker levels not only within and in close proximity to the gene encoding the protein under investigation, but also in more distant potentially regulatory regions, including on other chromosomes. The strength of the observed associations and the fact that we were able to replicate our findings in two independent cohorts suggest that these associations are real. This is further supported by the association of several of these genetic variants with *VEGF* gene expression in PBMCs. Our study also had several limitations. Whereas focusing on white populations of European

descent has the advantage of minimizing potential population stratification issues, our findings cannot be generalized to other ethnic groups. We may not have identified the true causal variants but merely SNPs in LD with the latter. Plasma levels of VEGF were measured in one of the replication cohorts (SFS), while serum VEGF levels had been measured in the discovery cohort. Serum VEGF concentrations are higher than plasma concentrations due to the release of VEGF from platelets during the clotting process and may have partly distinct genetic susceptibility factors.[50] The vast majority of associations found in the FHS did however replicate in the SFS and explained a substantial proportion of plasma VEGF variability, suggesting an important overlap between genetic susceptibility factors of serum and plasma VEGF levels. Finally, although our transcriptomic analysis does provide some support for a functional role of SNPs associated with mRNA levels of $VEGF_{121}$, these results are exploratory and were not corrected for multiple testing. We may have been underpowered for this analysis due to limited sample size and also because VEGF expression was measured on PBMCs only, which are not the sole contributors to circulating VEGF levels. Further studies looking at the association of genetic markers with VEGF expression in other cell types, including endothelial cells, would be of great value.

Clinical Implications

VEGF plays a key role in various diseases including atherosclerosis, inflammatory and neurodegenerative disorders, and cancer.[1] Anti-VEGF and proangiogenic VEGF based treatments have recently been developed for several therapeutic indications.[19-23] The identification of polymorphisms linked to VEGF levels could help in identifying patients who are more likely to respond favorably to such treatments. These therapies can have major side-effects,[24] and optimizing the risk-benefit ratio of their administration could lead to substantial improvements in patient care. The discovery of *trans*-acting genetic variants influencing VEGF levels could also spur the discovery of new molecular targets for pro- or anti-angiogenic therapies.

Conclusions

In a large population-based sample of European ancestry we identified novel genetic variants associated with circulating VEGF levels, on chromosome 6p21.1, 8q23.1 and 9p24.2, which explain almost half of the observed phenotypic variation.

Acknowledgements

Framingham Heart Study: This work was supported by the National Heart, Lung and Blood Institute's Framingham Heart Study (Contract No. N01-HC-25195) and its contract with Affymetrix, Inc for genotyping services (Contract No. N02-HL-6-4278). A portion of this research utilized the Linux Cluster for Genetic Analysis funded by the Robert Dawson Evans Endowment of the Department of Medicine at Boston University School of Medicine and Boston Medical Center. Analyses reflect intellectual input and resource development from the Framingham Heart Study investigators participating in the SNP Health Association Resource (SHARe) project. This study was also supported by grants from the National Heart Lung and Blood Institute (HL-077477, HL093029, HL-K24-04334), the National Institute of Neurological Disorders and Stroke (NS17950) and the National Institute of Aging (AG08122, AG033193, AG031287, AG033040, P30AG013846). Dr. Debette received an award from the Bettencourt-Schueller Foundation.

STANISLAS Family Study: The STANISLAS Family study samples and data used here are part of the Biological Resources Bank (BRC) "Interactions Gène-Environnement en Physiopathologie CardioVasculaire" (IGE-PCV) in Nancy, France. The STANISLAS Family Study, as part of the BRC, was supported by the "Caisse Nationale d'Assurance Maladies des Travailleurs Salariés" (CNAM), the "Institut National de la Santé et de la Recherche Médicale" (INSERM), the "Région Lorraine", the "Communauté Urbaine du Grand Nancy," and the "Henri Poincaré" University of Nancy I. We are deeply grateful to the cooperation of the families participating in the STANISLAS Cohort. We thank the staff of the "Centre de Médecine Préventive" of Vandoeuvre-lès-Nancy (France) for their involvement in the recruitment of the STANISLAS cohort.

Prospective Investigation of the Vasculature in Uppsala Seniors Study: Genotyping was performed by the SNP Technology Platform in Uppsala (www.genotyping.se). We thank Tomas Axelsson, Ann-Christine Wiman and Caisa Pöntinen for their excellent assistance with genotyping. The SNP Technology Platform is supported by Uppsala University and the Knut and Alice Wallenberg Foundation.

E.I. is supported by grants from the Swedish Research Council, the Swedish Foundation for Strategic Research, and the Royal Swedish Academy of Science.

Figure Legends

Figure 1: Regional plot for associations in region centered on rs6921438 (**1a**), rs6993770 (**1b**) and rs10738760 (**1c**).

All SNPs (triangles) are plotted with their meta-analysis p-values against their genomic position. The color of the triangles represents the linkage disequilibrium between each of the SNPs in the region with rs6921438: purple: $r2 \leq 0.05$, light blue: $0.05 < r2 \leq 0.10$, green: $0.10 < r2 \leq 0.30$, yellow: $0.30 < r2 \leq 0.60$, orange: $0.60 < r2 \leq 0.80$, red: $r2 > 0.80$. Light blue line represents estimated recombination rates. Genes are shown as dark green arrows. LD and recombination rates were drawn from Hapmap (release #22).

Figure 2: Putative Gene Network based on Ingenuity Path Analysis. Edges are displayed with labels describing the nature of the relationship between the nodes. The lines between genes represent known interactions and the nodes are displayed using various shapes which represent the functional class of the gene product (legend).

Figure 1a

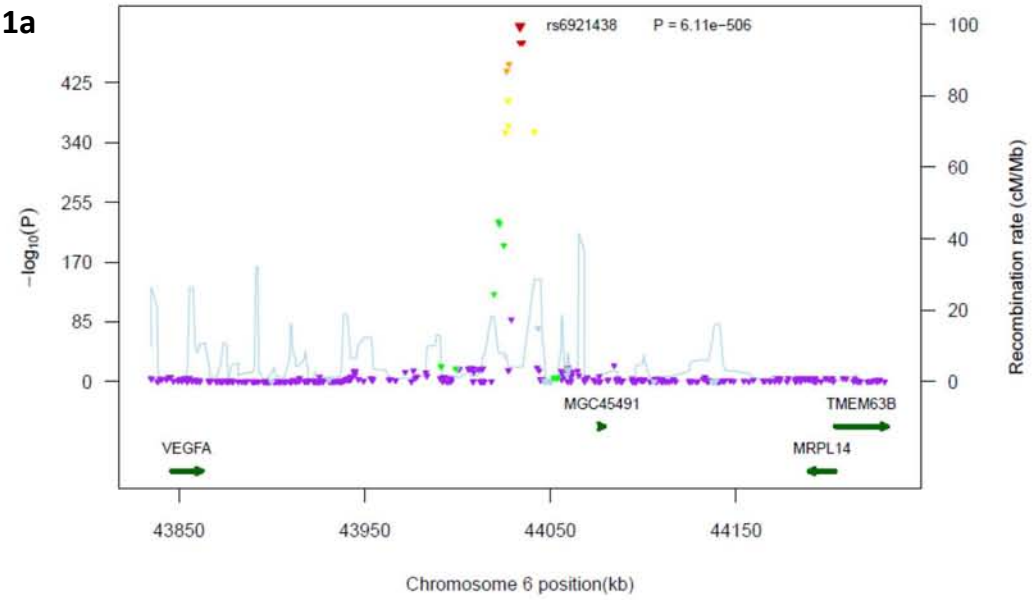


Figure 1b

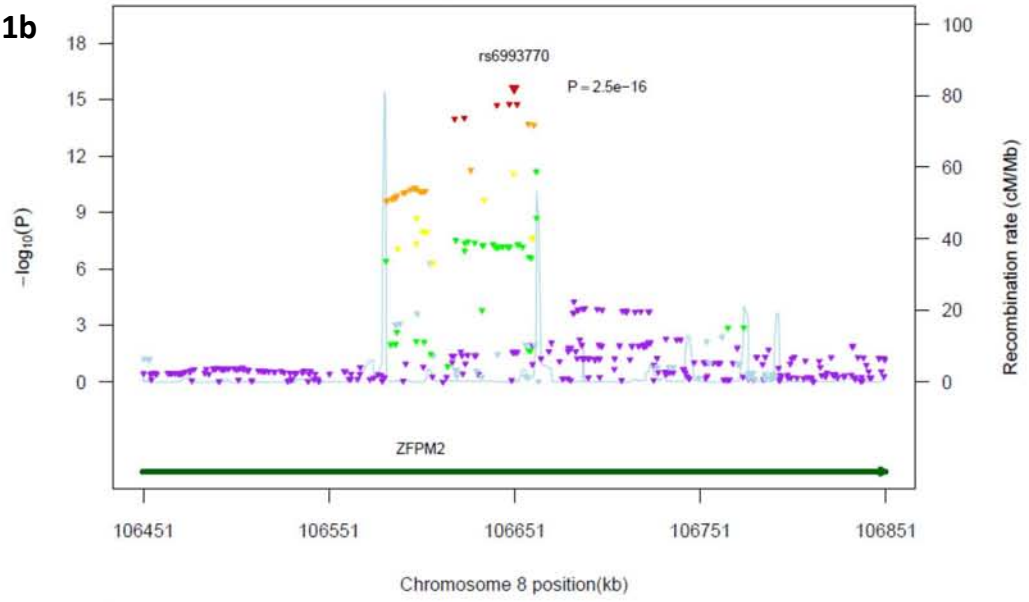


Figure 1c

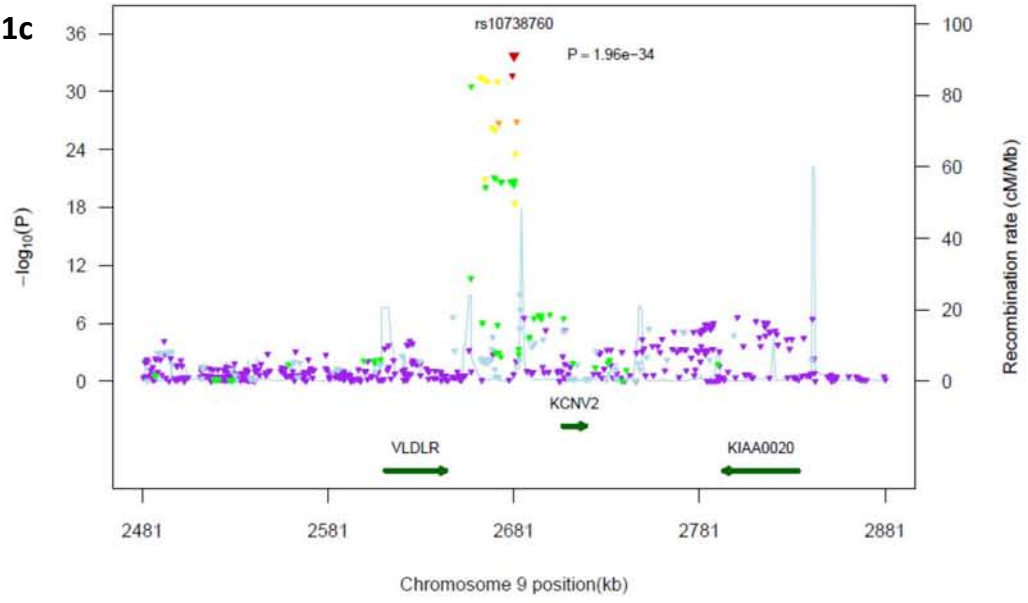


Figure 2

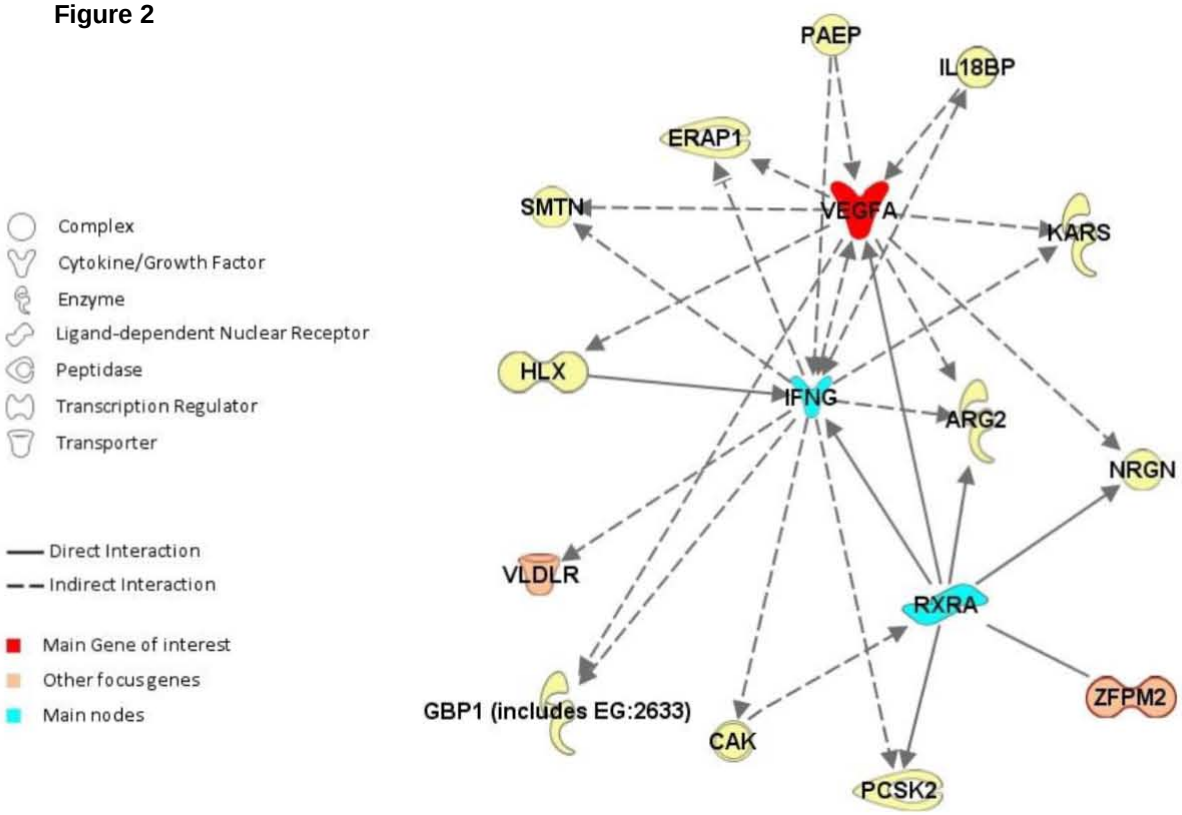

Complex
Cytokine/Growth Factor
Enzyme
Ligand-dependent Nuclear Receptor
Peptidase
Transcription Regulator
Transporter
Direct Interaction
Indirect Interaction
Main Gene of interest
Other focus genes
Main nodes
PAEP
IL18BP
ERAP1
SMTN
VEGFA
KARS
HLX
IFNG
ARG2
NRGN
VLDLR
RXRA
ZFPM2
GBP1 (includes EG:2633)
CAK
PCSK2

Table 1: Characteristics of Study Participants

Characteristics	FHS	SFS†	PIVUS Study
Number of participants	3546	859	868
Mean circulating VEGF level (ng/L), mean (SD) *	341.7 (254.7)	37.4 (35.7)	239.3 (185.4)
Mean age (SD) at VEGF measurement	40.0 (8.7)	29.83 (14.5)	70.2 (0.2)
Women (%)	1890 (53.2)	428 (49.8)	454 (52.3)
Cardiovascular risk factor at VEGF measurement			
Systolic blood pressure, mean (SD)	116.7 (14.0)	120.3 (12.8)	149.6 (22.7)
Hypertension (%)	561 (15.9)	23 (2.7)	606 (69.8)
Diabetes mellitus (%)	89 (2.5)	0	68 (7.8)
Current smoker (%)	544 (15.3)	188 (21.9)	93 (10.7)
Central obesity (%)	1315 (37.2)	37 (44.3)	266 (30.6)
Metabolic syndrome (%)	693 (19.6)	19 (2.2)	198 (22.8)

* Serum levels for the FHS and the PIVUS study, and plasma levels for the SFS; † by design, SFS participants were free of chronic disorders (cardiovascular or cancer) and had no personal history of cardiovascular disease at the time of inclusion (VEGF levels and covariates for the present analysis were measured during the second examination cycle in 1998–2000); all individuals with CVD (cardiovascular disease), defined in the FHS as presence of stroke, congestive heart failure, coronary heart disease or intermittent claudication, were excluded before analyses in FHS and PIVUS

Table 2: Genome-wide significant single nucleotide polymorphism (SNP)-phenotype associations in genome-wide association analysis of circulating VEGF levels ($p<5x10^{-8}$)

SNP	Chr	Position*	Function	MAF	MA	strand	beta† (FHS)	SE (FHS)	p (FHS)	p (SFS)	p (PIVUS)	Meta-p (replication)	Meta-p (all)	Gene1	Distance (kb)	Gene2	Distance (kb)
rs6921438	6	44033585	intergenic	0.51	G	+	0.72	0.01	$6.11x10^{-506}$	$2.93x10^{-39}$	NA	$2.93x10^{-39}$	$1.38x10^{-527}$	MGC45491	42.7	MRPL14	155.8
rs4513773	6	44033504	intergenic	0.47	G	+	-0.70	0.01	$2.08x10^{-482}$	NA	$6.23x10^{-139}$	$6.23x10^{-139}$	$4.45x10^{-619}$	MGC45491	42.8	MRPL14	155.8
rs9472159	6	44027673	intergenic	0.50	C	+	0.76	0.02	$4.30x10^{-452}$	$4.11x10^{-35}$	$3.27x10^{-109}$	$3.99x10^{-134}$	$6.54x10^{-570}$	MGC45491	48.6	MRPL14	161.7
rs9369434	6	44026385	intergenic	0.53	C	+	0.84	0.02	$2.15x10^{-442}$	$5.31x10^{-28}$	$1.43x10^{-63}$	$2.71x10^{-86}$	$3.91x10^{-490}$	MGC45491	49.9	MRPL14	163.0
rs1776717	6	44059314	intergenic	0.21	A	+	-0.23	0.02	$8.10x10^{-20}$	$9.75x10^{-6}$	$3.74x10^{-4}$	$1.44x10^{-8}$	$2.29x10^{-26}$	MGC45491	17.0	MRPL14	130.0
rs1776721	6	43998961	intronic	0.31	T	+	-0.18	0.02	$1.52x10^{-19}$	0.02	$3.43x10^{-8}$	$5.17x10^{-8}$	$6.70x10^{-26}$	MGC45491	77.3	VEGF	136.8
rs1886979	6	44012879	3'UTR	0.41	G	+	0.17	0.02	$3.71x10^{-19}$	0.01	$3.23x10^{-6}$	$6.73x10^{-7}$	$3.77x10^{-24}$	MGC45491	63.4	VEGF	150.7
rs9472155	6	44005705	intronic	0.22	T	+	-0.20	0.02	$4.45x10^{-19}$	0.01	$3.93x10^{-9}$	$7.52x10^{-9}$	$2.58x10^{-26}$	MGC45491	70.6	VEGF	143.5
rs844294	6	44008685	intronic	0.52	C	+	-0.15	0.02	$1.19x10^{-14}$	0.09	$2.25x10^{-5}$	$4.25x10^{-5}$	$5.92x10^{-18}$	MGC45491	67.6	VEGF	146.5
rs4416670	6	44058431	intergenic	0.55	T	+	0.13	0.02	$1.47x10^{-12}$	$2.87x10^{-4}$	0.10	$1.41x10^{-4}$	$1.80x10^{-15}$	MGC45491	17.9	MRPL14	130.9
rs910611	6	44058829	intergenic	0.08	C	+	-0.26	0.04	$2.61x10^{-10}$	0.11	$6.36x10^{-6}$	$2.83x10^{-5}$	$6.58x10^{-14}$	MGC45491	17.5	MRPL14	130.5
rs6993770	8	106650704	intronic	0.32	T	+	-0.17	0.02	$2.50x10^{-16}$	0.02	$3.99x10^{-8}$	$9.96x10^{-8}$	$1.51x10^{-22}$	ZFPM2	0	LRP12	980.4
rs16873402	8	106658423	intronic	0.33	T	+	-0.15	0.02	$1.97x10^{-14}$	0.16	$9.49x10^{-9}$	$1.86x10^{-6}$	$2.14x10^{-19}$	ZFPM2	0	LRP12	988.1
rs16873365	8	106627411	intronic	0.22	T	+	-0.16	0.02	$5.65x10^{-12}$	0.37	$2.09x10^{-6}$	$1.57x10^{-4}$	$4.06x10^{-15}$	ZFPM2	0	LRP12	957.1
rs7013321	8	106662734	intronic	0.49	A	+	-0.14	0.02	$6.75x10^{-12}$	0.01	NA	0.0144	$9.51x10^{-13}$	ZFPM2	0	LRP12	992.4
rs6993696	8	106650460	intronic	0.46	A	+	-0.13	0.02	$8.54x10^{-12}$	0.05	$1.49x10^{-4}$	$6.18x10^{-5}$	$3.33x10^{-15}$	ZFPM2	0	LRP12	980.1
rs16873291	8	106597206	intronic	0.31	T	+	-0.13	0.02	$5.30x10^{-11}$	0.07	$7.65x10^{-7}$	$4.08x10^{-6}$	$1.07x10^{-15}$	ZFPM2	0	LRP12	926.9
rs1349319	8	106625810	intronic	0.39	A	+	0.11	0.02	$3.59x10^{-8}$	0.05	$1.32x10^{-3}$	$3.03x10^{-4}$	$4.86x10^{-11}$	ZFPM2	0	LRP12	955.5
rs10738760	9	2681186	intergenic	0.49	A	+	0.28	0.02	$1.96x10^{-34}$	0.03	$1.12x10^{-8}$	$5.69x10^{-8}$	$3.91x10^{-37}$	KCNV2	26.3	VLDLR	36.7
rs6475920	9	2663933	intergenic	0.36	A	+	-0.24	0.02	$3.76x10^{-32}$	0.02	$2.40x10^{-8}$	$5.94x10^{-8}$	$6.55x10^{-37}$	VLDLR	19.4	KCNV2	43.6
rs4741756	9	2658187	intergenic	0.28	C	+	-0.25	0.02	$2.95x10^{-31}$	0.09	$8.64x10^{-5}$	$1.24x10^{-4}$	$4.08x10^{-31}$	VLDLR	13.7	KCNV2	49.3
rs2375980	9	2682622	intergenic	0.42	G	+	-0.25	0.02	$1.30x10^{-27}$	0.02	$2.25x10^{-8}$	$4.67x10^{-8}$	$5.50x10^{-6}$	KCNV2	24.9	VLDLR	38.1
rs10122587	9	2681951	intergenic	0.28	T	+	-0.22	0.02	$3.02x10^{-24}$	0.02	NA	0.0243	$1.81x10^{-23}$	KCNV2	25.6	VLDLR	37.5
rs10967492	9	2671175	intergenic	0.21	A	+	-0.22	0.02	$1.02x10^{-21}$	0.10	NA	0.104	$6.07x10^{-20}$	VLDLR	26.7	KCNV2	36.3
rs10967470	9	2665698	intergenic	0.24	G	+	-0.22	0.02	$1.17x10^{-21}$	0.04	NA	0.0442	$1.56x10^{-20}$	VLDLR	21.2	KCNV2	41.8

SNP: Single Nucleotide Polymorphism; Chr: chromosome; MA: Minor Allele; MAF : Minor Allele Frequency; SE: Standard Error; Meta-p: meta-analysis p-value; Gene1: closest referenced gene; Gene2: second closest referenced gene; *genome build 36.3; †effect estimate for the minor allele

Table 3: Significant correlations between SNPs and VEGF transcripts

phenotype	SNP	Chr	position	MA	MAF	beta †	SE	p	h2q (%)
mRNA_121	rs16873365	8	106627411	T	0.22	22.71	7.22	0.002	4.73
mRNA_121	rs16873402	8	106658423	T	0.33	12.15	5.10	0.017	2.84
mRNA_121	rs6993770	8	106650704	T	0.32	12.06	5.23	0.021	2.82
mRNA_121	rs16873291	8	106597206	T	0.31	11.95	5.37	0.026	2.47
mRNA_121	rs2375980	9	2682622	G	0.42	10.15	4.75	0.032	2.03
mRNA_121	rs910611	6	44058829	C	0.08	-19.47	9.49	0.040	2.13

*log-transformed; †effect estimate for the minor allele; Chr: chromosome; MA: Minor Allele; MAF: Minor Allele Frequency; SE: standard error; h2q: variance explained

Reference

1. Ferrara N, Gerber HP, LeCouter J. The biology of VEGF and its receptors. *Nat Med.* 2003;9:669-676.

2. Ferrara N, Davis-Smyth T. The biology of vascular endothelial growth factor. *Endocr Rev.* 1997;18:4-25.

3. Wei W, Jin H, Chen ZW, Zioncheck TF, Yim AP, He GW. Vascular endothelial growth factor-induced nitric oxide- and PGI2-dependent relaxation in human internal mammary arteries: a comparative study with KDR and Flt-1 selective mutants. *J Cardiovasc Pharmacol.* 2004;44:615-621.

4. Zachary I. Neuroprotective role of vascular endothelial growth factor: signalling mechanisms, biological function, and therapeutic potential. *Neurosignals.* 2005;14:207-221.

5. Hojo Y, Ikeda U, Zhu Y, Okada M, Ueno S, Arakawa H, Fujikawa H, Katsuki T, Shimada K. Expression of vascular endothelial growth factor in patients with acute myocardial infarction. *J Am Coll Cardiol.* 2000;35:968-973.

6. Blann AD, Belgore FM, McCollum CN, Silverman S, Lip PL, Lip GY. Vascular endothelial growth factor and its receptor, Flt-1, in the plasma of patients with coronary or peripheral atherosclerosis, or Type II diabetes. *Clin Sci (Lond).* 2002;102:187-194.

7. Chin BS, Chung NA, Gibbs CR, Blann AD, Lip GY. Vascular endothelial growth factor and soluble P-selectin in acute and chronic congestive heart failure. *Am J Cardiol.* 2002;90:1258-1260.

8. Slevin M, Krupinski J, Slowik A, Kumar P, Szczudlik A, Gaffney J. Serial measurement of vascular endothelial growth factor and transforming growth factor-beta1 in serum of patients with acute ischemic stroke. *Stroke.* 2000;31:1863-1870.

9. Wirostko B, Wong TY, Simo R. Vascular endothelial growth factor and diabetic complications. *Prog Retin Eye Res.* 2008;27:608-621.

10. Cao L, Jiao X, Zuzga DS, Liu Y, Fong DM, Young D, During MJ. VEGF links hippocampal activity with neurogenesis, learning and memory. *Nat Genet.* 2004;36:827-835.

11. Del Bo R, Scarlato M, Ghezzi S, Martinelli Boneschi F, Fenoglio C, Galbiati S, Virgilio R, Galimberti D, Galimberti G, Crimi M, Ferrarese C, Scarpini E, Bresolin N, Comi GP. Vascular endothelial growth factor gene variability is associated with increased risk for AD. *Ann Neurol.* 2005;57:373-380.

12. Ferrara N, Frantz G, LeCouter J, Dillard-Telm L, Pham T, Draksharapu A, Giordano T, Peale F. Differential expression of the angiogenic factor genes vascular endothelial growth factor (VEGF) and endocrine gland-derived VEGF in normal and polycystic human ovaries. *Am J Pathol.* 2003;162:1881-1893.

13. Peitsidis P, Agrawal R. Role of vascular endothelial growth factor in women with PCO and PCOS: a systematic review. *Reprod Biomed Online.*20:444-452.

14. Smith GC, Wear H. The perinatal implications of angiogenic factors. *Curr Opin Obstet Gynecol.* 2009;21:111-116.

15. Distler JH, Gay S, Distler O. Angiogenesis and vasculogenesis in systemic sclerosis. *Rheumatology (Oxford).* 2006;45 Suppl 3:iii26-27.

16. Paleolog EM. The vasculature in rheumatoid arthritis: cause or consequence? *Int J Exp Pathol.* 2009;90:249-261.

17. Nishimura R, Nagao K, Miyayama H, Matsuda M, Baba K, Yamashita H, Fukuda M. Higher plasma vascular endothelial growth factor levels correlate with menopause, overexpression of p53, and recurrence of breast cancer. *Breast Cancer.* 2003;10:120-128.

18. Moon HS, Kim SC, Ahn JJ, Woo BH. Concentration of vascular endothelial growth factor (VEGF) and transforming growth factor-beta1 (TGF-beta1) in the serum of patients with cervical cancer: prediction of response. *Int J Gynecol Cancer.* 2000;10:151-156.

19. Isner JM, Pieczek A, Schainfeld R, Blair R, Haley L, Asahara T, Rosenfield K, Razvi S, Walsh K, Symes JF. Clinical evidence of angiogenesis after arterial gene transfer of phVEGF165 in patient with ischaemic limb. *Lancet.* 1996;348:370-374.
20. Yeung YA, Wu X, Reyes AE, 2nd, Vernes JM, Lien S, Lowe J, Maia M, Forrest WF, Meng YG, Damico LA, Ferrara N, Lowman HB. A Therapeutic Anti-VEGF Antibody with Increased Potency Independent of Pharmacokinetic Half-life. *Cancer Res.*
21. Hasskarl J. Sorafenib. *Recent Results Cancer Res.*184:61-70.
22. Dixon JA, Oliver SC, Olson JL, Mandava N. VEGF Trap-Eye for the treatment of neovascular age-related macular degeneration. *Expert Opin Investig Drugs.* 2009;18:1573-1580.
23. Schoettler N, Brahn E. Angiogenesis inhibitors for the treatment of chronic autoimmune inflammatory arthritis. *Curr Opin Investig Drugs.* 2009;10:425-433.
24. Wagner AD, Arnold D, Grothey AA, Haerting J, Unverzagt S. Anti-angiogenic therapies for metastatic colorectal cancer. *Cochrane Database Syst Rev.* 2009:CD005392.
25. Lieb W, Safa R, Benjamin EJ, Xanthakis V, Yin X, Sullivan LM, Larson MG, Smith HM, Vita JA, Mitchell GF, Sawyer DB, Vasan RS. Vascular endothelial growth factor, its soluble receptor, and hepatocyte growth factor: clinical and genetic correlates and association with vascular function. *Eur Heart J.* 2009;30:1121-1127.
26. Pantsulaia I, Trofimov S, Kobyliansky E, Livshits G. Heritability of circulating growth factors involved in the angiogenesis in healthy human population. *Cytokine.* 2004;27:152-158.
27. Berrahmoune H, Herbeth B, Lamont JV, Masson C, Fitzgerald PS, Visvikis-Siest S. Heritability for plasma VEGF concentration in the Stanislas family study. *Ann Hum Genet.* 2007;71:54-63.
28. Visvikis-Siest S, Siest G. The STANISLAS Cohort: a 10-year follow-up of supposed healthy families. Gene-environment interactions, reference values and evaluation of biomarkers in prevention of cardiovascular diseases. *Clin Chem Lab Med.* 2008;46:733-747.
29. Lind L, Fors N, Hall J, Marttala K, Stenborg A. A comparison of three different methods to evaluate endothelium-dependent vasodilation in the elderly: the Prospective Investigation of the Vasculature in Uppsala Seniors (PIVUS) study. *Arterioscler Thromb Vasc Biol.* 2005;25:2368-2375.
30. Ehrich M, Bocker S, van den Boom D. Multiplexed discovery of sequence polymorphisms using base-specific cleavage and MALDI-TOF MS. *Nucleic Acids Res.* 2005;33:e38.
31. Fan JB, Oliphant A, Shen R, Kermani BG, Garcia F, Gunderson KL, Hansen M, Steemers F, Butler SL, Deloukas P, Galver L, Hunt S, McBride C, Bibikova M, Rubano T, Chen J, Wickham E, Doucet D, Chang W, Campbell D, Zhang B, Kruglyak S, Bentley D, Haas J, Rigault P, Zhou L, Stuelpnagel J, Chee MS. Highly parallel SNP genotyping. *Cold Spring Harb Symp Quant Biol.* 2003;68:69-78.
32. Chen MH, Yang Q. GWAF: an R package for genome-wide association analyses with family data. *Bioinformatics.*26:580-581.
33. Steffensen KD, Waldstrom M, Brandslund I, Jakobsen A. The relationship of VEGF polymorphisms with serum VEGF levels and progression-free survival in patients with epithelial ovarian cancer. *Gynecol Oncol.*117:109-116.
34. Kamoun M, Houman MH, Hamzaoui A, Hamzaoui K. Vascular endothelial growth factor gene polymorphisms and serum levels in Behcet's disease. *Tissue Antigens.* 2008;72:581-587.
35. Zhai R, Gong MN, Zhou W, Thompson TB, Kraft P, Su L, Christiani DC. Genotypes and haplotypes of the VEGF gene are associated with higher mortality and lower VEGF plasma levels in patients with ARDS. *Thorax.* 2007;62:718-722.
36. Ferrante M, Pierik M, Henckaerts L, Joossens M, Claes K, Van Schuerbeek N, Vlietinck R, Rutgeerts P, Van Assche G, Vermeire S. The role of vascular endothelial growth factor (VEGF) in inflammatory bowel disease. *Inflamm Bowel Dis.* 2006;12:870-878.
37. Krippl P, Langsenlehner U, Renner W, Yazdani-Biuki B, Wolf G, Wascher TC, Paulweber B, Haas J, Samonigg H. A common 936 C/T gene polymorphism of vascular endothelial growth factor is associated with decreased breast cancer risk. *Int J Cancer.* 2003;106:468-471.

38. Awata T, Inoue K, Kurihara S, Ohkubo T, Watanabe M, Inukai K, Inoue I, Katayama S. A common polymorphism in the 5'-untranslated region of the VEGF gene is associated with diabetic retinopathy in type 2 diabetes. *Diabetes.* 2002;51:1635-1639.

39. Renner W, Kotschan S, Hoffmann C, Obermayer-Pietsch B, Pilger E. A common 936 C/T mutation in the gene for vascular endothelial growth factor is associated with vascular endothelial growth factor plasma levels. *J Vasc Res.* 2000;37:443-448.

40. Petrovic MG, Korosec P, Kosnik M, Osredkar J, Hawlina M, Peterlin B, Petrovic D. Local and genetic determinants of vascular endothelial growth factor expression in advanced proliferative diabetic retinopathy. *Mol Vis.* 2008;14:1382-1387.

41. Cacev T, Loncar B, Seiwerth S, Spaventi S, Kapitanovic S. Vascular endothelial growth factor polymorphisms -1154 G/A and -460 C/T are not associated with VEGF mRNA expression and susceptibility to sporadic colon cancer. *DNA Cell Biol.* 2008;27:569-574.

42. Cosin R, Gilabert-Estelles J, Ramon LA, Espana F, Gilabert J, Romeu A, Estelles A. Vascular endothelial growth factor polymorphisms (-460C/T, +405G/C, and 936C/T) and endometriosis: their influence on vascular endothelial growth factor expression. *Fertil Steril.* 2009;92:1214-1220.

43. Visvikis-Siest S, Marteau JB, Samara A, Berrahmoune H, Marie B, Pfister M. Peripheral blood mononuclear cells (PBMCs): a possible model for studying cardiovascular biology systems. *Clin Chem Lab Med.* 2007;45:1154-1168.

44. Keyt BA, Berleau LT, Nguyen HV, Chen H, Heinsohn H, Vandlen R, Ferrara N. The carboxyl-terminal domain (111-165) of vascular endothelial growth factor is critical for its mitogenic potency. *J Biol Chem.* 1996;271:7788-7795.

45. Tevosian SG, Deconinck AE, Tanaka M, Schinke M, Litovsky SH, Izumo S, Fujiwara Y, Orkin SH. FOG-2, a cofactor for GATA transcription factors, is essential for heart morphogenesis and development of coronary vessels from epicardium. *Cell.* 2000;101:729-739.

46. Song H, Suehiro J, Kanki Y, Kawai Y, Inoue K, Daida H, Yano K, Ohhashi T, Oettgen P, Aird WC, Kodama T, Minami T. Critical role for GATA3 in mediating Tie2 expression and function in large vessel endothelial cells. *J Biol Chem.* 2009;284:29109-29124.

47. von Kodolitsch Y, Ito WD, Franzen O, Lund GK, Koschyk DH, Meinertz T. Coronary artery anomalies. Part I: Recent insights from molecular embryology. *Z Kardiol.* 2004;93:929-937.

48. Haines JL, Schnetz-Boutaud N, Schmidt S, Scott WK, Agarwal A, Postel EA, Olson L, Kenealy SJ, Hauser M, Gilbert JR, Pericak-Vance MA. Functional candidate genes in age-related macular degeneration: significant association with VEGF, VLDLR, and LRP6. *Invest Ophthalmol Vis Sci.* 2006;47:329-335.

49. Loewen N, Chen J, Dudley VJ, Sarthy VP, Mathura JR, Jr. Genomic response of hypoxic Muller cells involves the very low density lipoprotein receptor as part of an angiogenic network. *Exp Eye Res.* 2009;88:928-937.

50. Verheul HM, Hoekman K, Luykx-de Bakker S, Eekman CA, Folman CC, Broxterman HJ, Pinedo HM. Platelet: transporter of vascular endothelial growth factor. *Clin Cancer Res.* 1997;3:2187-2190.

Supplemental Material

Identification of Cis- and Trans-acting Genetic Variants Explaining up to Half the Heritability of Circulating VEGF levels

A. Supplemental Methods

1. Cohorts

Framingham Heart Study (FHS)
The FHS is a three-generation, single-site, community-based, prospective cohort study that was initiated in 1948 to investigate risk factors for cardiovascular disease including stroke in Framingham, MA, USA. It now comprises 3 generations of participants: the original cohort followed since 1948 (Original, Gen 1)[1]; their offspring and spouses of the offspring, followed since 1971 (Offspring, Gen 2);[2] and grandchildren of the Original cohort in 2002 (Gen 3).[3] Vascular endothelial growth factor (VEGF) levels have at this time only been measured in Gen 3 participants. Starting in 2002, 4,095 participants with at least one parent in the Offspring cohort were enrolled in the Gen 3 cohort.[3] At their first examination (2002-2005) Gen 3 participants underwent a targeted medical history, physical examination, anthropometry, and laboratory assessment of traditional cardiovascular risk factors. VEGF levels were available in 3,946 participants, among whom we excluded persons with prevalent cardiovascular disease (n=66), serum creatinine >2 mg/dl (n=1) and those in whom data for more covariates (n=125) were not available. Of the remaining 3,754 participants, 3,553 also had DNA extracted and genome-wide genotyping data available. Genome-wide genotyping was performed at Affymetrix (Santa Clara, CA) through an NHLBI funded SNP-Health Association Resource (SHARe) project and was successful in 3,546 persons. These 3,546 persons constitute the FHS sample for this study.
All study participants provided informed consent and the study was approved by the Institutional Review Board at the Boston University Medical Center.

STANISLAS Family Study (SFS)
Initiated in 1993, the STANISLAS Family Study is a 10-year longitudinal survey set up to seek out gene-gene and gene-environment interactions in the field of cardiovascular diseases.[4] 1,006 volunteer families (2 parents and at least two siblings) were initially recruited between 1993 and 1995 during a free health check-up at the Centre for Preventive Medicine in Vandoeuvre-lès-Nancy, France. Individuals with chronic disorders (cardiovascular or cancer) or having a personal history of cardiovascular disease were not included, in order to assess the effects of genetic susceptibility factors on the variability of intermediate phenotypes in physiological conditions without the influence of any long term medication and disease. Biological and clinical measurements, health and lifestyle information were thus collected from 4,488 individuals belonging to 2 generations using appropriate, validated questionnaires and procedures as described previously.[4, 5] DNA was extracted from all participants and serum/plasma/DNA biobanks were constructed. In 1998–2000, 756 of the original families attended a second check-up and the same examination and biobanking protocols undertaken at the first examination were repeated. Using the plasma collected at this examination, specific measurements like VEGF plasma concentrations were measured on a random subsample of 1,000 subjects. A third check-up took place between 2003 and 2005 and at this visit mRNA and lymphocytes extract biobanks were constructed in addition to serum/plasma/DNA biobanks. Individuals with both VEGF plasma concentrations and DNA available were selected for the present study (919 individuals from 263 families). Some parents or children were secondarily excluded because DNA amplification was not successful (n=58), or results of genotyping were ambiguous (n=2). Finally, 859 individuals from 217 families met our selection criteria. Each subject gave written informed consent for participating in this study, which was approved by the Local Ethics Committee of Nancy.

Prospective Investigation of the Vasculature in Uppsala Seniors (PIVUS) study
All 70-year old individuals living in the community of Uppsala, Sweden, between April 2001 and June 2004 were eligible for the Prospective Investigation of the Vasculature in Uppsala Seniors (PIVUS) study, which

has been described in detail previously.[6] The individuals were selected randomly, and were examined within one month of their 70th birthday in order to standardize for age. Of 2,025 individuals invited, 1,016 were investigated giving a participation rate of 50%, and 999 of these individuals provided DNA for genetic studies. For the purpose of the present investigation, we excluded participants with sample call rate <90% (n=11), prevalent myocardial infarction, stroke or congestive heart failure (n=116), serum creatinine >2 mg/dl (n=1), or missing data on VEGF (n=3). Thus, 868 participants (52% women) were eligible for the present study. The PIVUS study has been approved by the Ethics Committee of Uppsala University and the participants gave written informed consent.

2. Laboratory measurements of vascular endothelial growth factor

VEGF levels were measured in serum for the FHS and PIVUS and plasma for the SFS, at the first examination cycle (2002-2005) in the FHS, the second examination in the SFS (1998-2000) and the baseline examination (2001-2004) in the PIVUS study.
In all 3 studies venous blood samples were drawn after an overnight fast, immediately centrifuged and stored appropriately (at -80°C in FHS and PIVUS and at -196°C in liquid nitrogen in SFS) until VEGF measurements were undertaken. At FHS, serum VEGF was measured using a commercial ELISA assay (R&D Inc.). In SFS and PIVUS plasma VEGF and serum VEGF quantification respectively was performed by Randox Ltd (Crumlin, UK), using a biochip array analyzer (Evidence®).[7, 8] In all studies both diffusible VEGF isoforms ($VEGF_{121}$ and $VEGF_{165}$) were detected.
The average inter-assay coefficients of variation were 2.1% for serum VEGF in the FHS, less than 9% in the SFS and less than 15% in the PIVUS study.
Since serum VEGF had been measured in the FHS and PIVUS and plasma for VEGF in the SFS, we checked the correlation between the 2 types of specimens. VEGF was measured in a subset (n=18) of matched plasma and serum samples. Plasma VEGF was lower (42±28 ng/L, mean±SD) than serum VEGF (361±223 ng/L); There was a strong correlation between plasma and serum VEGF (r=0.76, p=0.0002) which strengthens our study.

3. Genotyping, quality controls, and imputation

Genome-wide genotyping in the FHS study was performed on the Affymetrix GeneChip Human Mapping 500K Array Set® and 50K Human Gene Focused Panel®.
The set of genotyped input SNPs used for imputation was selected based on their highest quality GWA data. From a total of 534,982 genotyped autosomal SNPs in FHS, we used 378,163 SNPs in the imputation after filtering out 15,586 SNPs for Hardy-Weinberg disequilibrium ($p<1x10^{-6}$), 64,511 SNPs for missingness >0.03, 45,361 SNPs for a test of differential missingness yielding a $p<1x10^{-9}$ (mishap test in PLINK, http://pngu.mgh.harvard.edu/purcell/plink/), 4,857 SNPs for >100 Mendel errors, 67,269 SNPs for a minor allele frequency < 0.01, 2 SNPs due to strandedness issues upon merging data with HapMap, and a further 13,394 SNPs because they were not present on HapMap. We used the Markov Chain Haplotyping (MaCH) package (http://www.sph.umich.edu/csg/abecasis/MACH, version 1.0.15 software) and imputed to the plus strand of NCBI build 36, HapMap release #22. For each imputed SNP, imputation quality was estimated as the ratio of the empirically observed dosage variance to the expected binomial dosage variance. After quality control and filtering, FHS had either genotyped or imputed data for 2,540,223 autosomal SNPs. From a total of 10,886 genotyped SNPs on the X chromosome, we used 7,795 SNPs in the imputation after filtering out 3,091 SNPs for Hardy-Weinberg $p<1x10^{-6}$ (n=159), missingness >0.03 (n=450), minor allele frequency < 0.01 (n=1851), male heterozygote count > 45 (n=12), and a further 619 SNPs because they were not present on HapMap. We used the IMPUTE package (https://mathgen.stats.ox.ac.uk/impute/impute.html, version 0.5.0) and imputed to the plus strand of NCBI Build 35, Hapmap release #21.

4. Screening for latent population substructure

FHS was screened for latent population substructure, including cryptic relatedness, using EIGENSTRAT.[9, 10] The ninth principal component was significantly associated with serum VEGF levels and was therefore included as a covariate in the GWAS.

5. Covariate definition

In all 3 studies, hypertension was defined as systolic blood pressure ≥140, diastolic blood pressure ≥90 or use of anti-hypertensive treatment and smokers were identified based on current smoking status. Standing waist circumference (WC) was obtained at the level of the umbilicus and central obesity was defined as a WC ≥102 cm in men and ≥89 cm in women. Metabolic syndrome was defined as having any 3 of the following traits: (i) central obesity, (ii) hypertension, (iii) hyperglycemia defined as fasting glucose ≥100 mg/dL or use of anti-diabetic medications or insulin, (iv) hypertriglyceridemia defined as serum or plasma triglycerides ≥150 mg/dL or being on any lipid-lowering medication, (v) low HDL cholesterol (<40 mg/d in men and <50 mg/d in women).[11] These covariates were measured at the same time as VEGF levels, at the first Gen 3 examination cycle (2002-2005) in the FHS, at the second examination in the SFS (1998-2000) and at the baseline examination (2001-2004) in the PIVUS study.

6. Selection of SNPs for replication

From all SNPs associated with VEGF levels at a p-value < $5x10^{-8}$ in the discovery cohort, we excluded SNPs with minor allele frequencies (MAF) <0.05 as well as those imputed SNPs with low imputation quality (ratio of the empirically observed dosage variance to the expected binomial dosage variance <0.6). The remaining SNPs were grouped by "bins", each bin comprising SNPs that are in very strong linkage disequilibrium (LD) with each other, i.e. with an r^2>0.8 with the most significant SNP in the bin. Within each bin we selected one SNP for replication (except for the bin with the most significant associations where 2 SNPs were selected). Typically, we selected the SNP with the lowest p-value within each bin. In 6 bins a SNP with a slightly higher p-value was chosen either because this SNP had been directly genotyped in the discovery cohort, whereas the SNP with the lowest p-value in the bin had been imputed (rs1776721 and rs1886979) or because there were stronger arguments for functionality for the SNP with the slightly higher p-value (rs16873291, rs1349319, rs6475920 and rs10967492).

7. Meta-analysis

For the meta-analyses we used a fixed effects inverse-variance meta-analysis technique. Beta estimates were weighted by their inverse variance and a combined estimate was obtained by summing the weighted betas and dividing by the summed weights. We undertook the meta-analyses at FHS using METAL software (http://www.sph.umich.edu/csg/abecasis/Metal/index.html).

8. Genetic liability score calculation

Step 1:
We selected SNPs with independent effects by running a conditional GWAS. This was done in a forward stepwise fashion. First we ran a GWAS adjusting for age, gender, the ninth principal component and the most significantly associated SNP (rs6921438). We then ran a GWAS additionally adjusting for the most significantly associated SNP in the aforementioned conditional GWAS. This process was repeated, by adding the most significantly associated SNP as a new covariate to the regression model, until all SNPs independently associated with VEGF level at a p-value < $5x10^{-8}$ were selected. All selected independent SNPs with a minor allele frequency > 5% were used to compute the genetic liability score (**Supplementary Table 1**). In the PIVUS study, as rs6921438 genotypes were not available, rs4513773 was used instead to compute the genetic liability score (r^2 with rs6921438 = 0.90)

Step 2:
Genotypes for the SNPs selected in step 1 were coded as 0, 1 or 2 for genotyped SNPs (according to the number of minor alleles) and the imputed allele dosage was used for imputed SNPs.
To compute the genetic liability score the genotype value was weighted by the effect size estimate from the GWAS (**Supplementary Table 1**):
Risk Score = SNP1_estimate * SNP1_genotype + SNP2_estimate * SNP2_genotype + SNP3_estimate * SNP3_genotype + SNP4_estimate * SNP4_genotype

Step 3:

We estimated the proportion of phenotype variance explained by each SNP incorporated in the risk score (h2q, see **Supplementary Table 1**).[12]

9. Biological pathway analysis

Genes located close to associated SNPs were investigated for relevant networks by the Ingenuity Pathway Analysis (IPA) software (Ingenuity Systems, www.ingenuity.com). To build networks, IPA queries the Ingenuity Pathways Knowledge Base for biological interactions between identified "focus genes", in this case genes close to SNPs significantly associated with circulating VEGF levels in the GWAS, and all other gene objects stored in the knowledge base. It then generates a set of networks with a maximum network size of 35 genes. An underlying assumption is that highly-interconnected networks are likely to represent significant biological function,[13] thus IPA optimizes for triangular relationships between genes, favoring denser networks over more sparsely connected ones. Networks are displayed graphically as "nodes" (corresponding to genes or gene products) and the biological relationships between the nodes, referred to as "edges". IPA also computes a score, representing the –log (p-value), where the p-value is the probability of finding f or more focus genes in a set of n genes randomly selected.

10. VEGF gene expression analysis in peripheral blood mononuclear cells (PBMCs)

Sample preparation and quantification of the messenger RNA (mRNA) VEGF spliced forms
Two-hundred and twenty samples from the SFS were randomly selected for inclusion in the PBMCs transcriptomic study. Fresh whole blood (10mL) was collected by standardized venipuncture in EDTA tubes (Vacutainer^{TM}; Becton Dickinson, NJ, USA). PBMCs were isolated by centrifugation on a density gradient of Ficoll (Ficoll-Paque^{TM} PLUS; Amersham BioSciences) and stored at -80°c until RNA extraction according to a well-validated protocol,[14] with high recovery of lymphocyte(97.5%). Total RNA was subsequently extracted with the MagNaPure automate, using the MagNA Pure LC RNA HP isolation kit and RNA HP Blood External lysis protocol (Roche Diagnostics, France). Reverse transcription of total RNA was performed using 200 units of M-MuLV Reverse Transcriptase with 0.25 µg of oligos(dt) (Promega, France) according to a previous described protocol.[14] Quantification of the transcripts coding for the $VEGF_{121}$ and $VEGF_{165}$ isoforms, and the beta 2 microglobulin (B2M) control gene, was performed using TaqMan® and LightCycler technologies (LC TaqMan Master kitRoche Diagnostics, France). All experiments were performed in duplicate. The detection level for each transcript was between 1 and 10 copies for both transcripts and for B2M. RT-PCR optimization and specificity of RT-PCR products were examined using SYBR® Green technology (LC FastStart DNA Master^{PLUS} SYBR Green I kit, Roche Diagnostic, France), melting curves analysis and agarose gel electrophoresis of the PCR amplicons. Primers and probes were designed to specifically amplify the spliced forms of VEGF based on their splicing sites[15] with specific reverse primers or hydrolyzation probes spanning the variant specific exon boundaries, which also avoids amplification of contaminating genomic DNA. Hydrolyzation probes were labeled with the reporter dye FAM (6-carboxy-fluorescein phosphoramidite) at the 5' end and the quencher dye TAMRA (5 -carboxyl- tetramethyl-rhodamine) at the 3' end. For all assays, intra- and inter-run variability were 11% and 5% respectively. PCR products for each VEGF mRNA spliced variant amplification were purified with a PCR purification kit (QiaQuick, Qiagen, France). The product concentrations were measured in a spectrophotometer, the molecule concentrations were calculated, and a standard curve was generated for each transcript using serial dilutions of products ranging from 1 or 10 to 10^7 molecules/µL. The copy number of unknown samples was calculated by setting their PCR cycle number (Crossing Point: CP) to the standard curve and normalized to the housekeeping B2M gene. Results are presented as copies of the target gene product per 10^6 copies of B2M. Primer efficiencies were calculated according to the equation $E = 10[-1/slope]$. All investigated transcripts had real-time PCR efficiency rates above 1.9.
PBMCs VEGF protein measurements
PBMC VEGF (121 and 165) protein quantification was performed by Randox Ltd (Crumlin, UK), with a biochip array analyzer (Evidence®) using a high sensitivity kit as previously described.[7] PBMC VEGF concentrations were log10-transformed in all analyses in order to improve normality, were adjusted for the effect of between-run variation and regressed on mean values of all samples measured in each run. Grubbs' test was applied for detection of the extreme values in the data (log transformed VEGF) and there were no outliers at the 5% level. The average inter-assay coefficient of variation was 5.7%.

Statistical analysis
A linear mixed effects model that accounts for within family correlation was used to evaluate the association of each of the SNPs successfully genotyped in the SFS with each of the two transcript levels ($VEGF_{121}$ and $VEGF_{165}$) and with natural log-transformed PBMC VEGF concentration, assuming an additive genetic model. These analyses were adjusted for age and sex.

11. Candidate gene association studies

We searched Pubmed for original articles published before May 14 2010, which reported statistically significant associations between specific genotypes and circulating VEGF levels. We used the search terms VEGF, gene and polymorphism. We identified 14 candidate gene association studies,[16-29] who examined 11 different polymorphisms in the VEGF gene in association with plasma or serum VEGF levels (**Supplementary Table 4**). Of these, 8 studies reported a significant association of at least one polymorphism (rs699947, rs1570360, rs833061, rs2010963, rs3025039, -2549 18bp I/D) with circulating VEGF levels.[22-29] In the FHS sample we replicated the association of rs699947 and rs833061 with serum VEGF levels at $p < 5x10^{-7}$ (**Supplementary Table 4**). The other 4 polymorphisms (rs1570360, rs2010963, rs3025039, -2549 18bp I/D) had neither been genotyped nor imputed in the FHS GWAS. As these SNPs are not in the Hapmap database, we could not identify any proxy either to test their association with serum VEGF levels in the FHS dataset.

B. Supplemental Tables

Supplemental Table 1: Results of the main and conditional GWAS within the Framingham Heart Study sample for the SNPs retained in the genetic liability score calculation

SNP	chr	position	Minor allele	MAF	Estimate in main GWAS	SE in main GWAS	p in main GWAS	p in conditional GWAS	h2q (%)
rs6921438	6	44033585	G	0.51	0.7199	0.0149	$<5x10^{-324}$	$<5x10^{-324}$	41.19
rs10738760	9	2681186	A	0.49	0.2812	0.0230	$1.96x10^{-34}$	$3.78x10^{-47}$	4.97
rs6993770	8	106650704	T	0.32	-0.1667	0.0203	$2.50x10^{-16}$	$5.45x10^{-30}$	2.03
rs4416670	6	44058431	T	0.55	0.1342	0.0190	$1.47x10^{-12}$	$4.79x10^{-9}$	1.46

Chr: chromosome; GWAS: genome-wide association study; h2q: percentage of phenotypic variance explained; MAF: minor allele frequency; SE: standard error

Supplemental Table2: SNPs associated with serum VEGF levels at a p-value < $5x10^{-8}$ in the discovery GWAS

SNP	Chr	Position	MA	MAF	beta	SE	p	Gene 1	Distance	Gene 2	Distance	Function	O/E ratio	Imputed
rs6921438	6	44033585	G	0,52	0,72	0,01	6.11E-506	MGC45491	42731	MRPL14	155764	INTERGENIC	0,99	0
rs7767396	6	44035028	G	0,47	-0,71	0,02	1.71E-482	MGC45491	41288	MRPL14	154321	INTERGENIC	0,99	1
rs4513773	6	44033504	G	0,46	-0,71	0,02	2.08E-482	MGC45491	42812	MRPL14	155845	INTERGENIC	1,00	0
rs9472159	6	44027673	C	0,51	0,76	0,02	4.30E-452	MGC45491	48643	MRPL14	161676	INTERGENIC	0,82	1
rs9369434	6	44026385	C	0,53	0,85	0,02	2.15E-442	MGC45491	49931	MRPL14	162964	INTERGENIC	0,66	1
rs6916314	6	44027140	G	0,38	0,92	0,02	3.13E-400	MGC45491	49176	MRPL14	162209	INTERGENIC	0,54	1
rs9472158	6	44026875	G	0,38	0,92	0,02	6.22E-400	MGC45491	49441	MRPL14	162474	INTERGENIC	0,54	1
rs6916540	6	44027394	C	0,40	0,87	0,02	3.60E-364	MGC45491	48922	MRPL14	161955	INTERGENIC	0,56	1
rs9472173	6	44041234	C	0,47	0,93	0,02	1.16E-355	MGC45491	35082	MRPL14	148115	INTERGENIC	0,47	1
rs729391	6	44025870	C	0,39	1,05	0,03	7.17E-355	MGC45491	50446	MRPL14	163479	INTERGENIC	0,38	1
rs7764227	6	44021966	G	0,26	1,32	0,04	1,41E-228	MGC45491	54350	VEGFA	159767	INTERGENIC	0,22	1
rs9369433	6	44022528	C	0,23	1,28	0,04	6,87E-225	MGC45491	53788	VEGFA	160329	INTERGENIC	0,25	1
rs1359617	6	44024870	T	0,11	1,87	0,06	1,56E-194	MGC45491	51446	VEGFA	162671	INTERGENIC	0,18	1
rs865585	6	44019529	C	0,29	0,83	0,04	3,16E-125	MGC45491	56787	VEGFA	157330	INTERGENIC	0,30	1
rs6458360	6	44028870	T	0,11	1,10	0,05	3,14E-89	MGC45491	47446	MRPL14	160479	INTERGENIC	0,28	1
rs17209449	6	44043245	C	0,09	-1,20	0,06	1,30E-76	MGC45491	33071	MRPL14	146104	INTERGENIC	0,24	1
rs10738760	9	2681186	A	0,49	0,28	0,02	1,96E-34	KCNV2	26339	VLDLR	36701	INTERGENIC	0,68	1
rs10757631	9	2680295	G	0,47	-0,26	0,02	2,30E-32	KCNV2	27230	VLDLR	35810	INTERGENIC	0,73	1
rs6475919	9	2662809	T	0,37	-0,24	0,02	3,63E-32	VLDLR	18324	KCNV2	44716	INTERGENIC	0,95	1
rs6475920	9	2663933	A	0,37	-0,24	0,02	3,76E-32	VLDLR	19448	KCNV2	43592	INTERGENIC	0,96	1
rs1020651	9	2666360	G	0,37	-0,23	0,02	8,60E-32	VLDLR	21875	KCNV2	41165	INTERGENIC	1,01	1
rs1020652	9	2667012	T	0,37	-0,23	0,02	8,69E-32	VLDLR	22527	KCNV2	40513	INTERGENIC	1,01	0
rs10738758	9	2672355	G	0,37	-0,23	0,02	9,11E-32	VLDLR	27870	KCNV2	35170	INTERGENIC	0,98	1
rs4741756	9	2658187	C	0,28	-0,25	0,02	2,95E-31	VLDLR	13702	KCNV2	49338	INTERGENIC	0,97	0
rs2375980	9	2682622	G	0,42	-0,25	0,02	1,30E-27	KCNV2	24903	VLDLR	38137	INTERGENIC	0,70	1
rs7856084	9	2672943	C	0,39	-0,22	0,02	1,75E-27	VLDLR	28458	KCNV2	34582	INTERGENIC	0,97	1
rs10125071	9	2669579	C	0,40	-0,21	0,02	5,88E-27	VLDLR	25094	KCNV2	37946	INTERGENIC	1,00	1
rs4317630	9	2671025	C	0,39	-0,21	0,02	9,65E-27	VLDLR	26540	KCNV2	36500	INTERGENIC	0,98	1

rs10122587	9	2681951	T	0,28	-0,22	0,02	3,02E-24	KCNV2	25574	VLDLR	37466	INTERGENIC	0,98	1
rs9381273	6	44084246	G	0,01	5,92	0,59	5,37E-24	MGC45491	3381	MRPL14	105103	DOWNSTREAM	0,03	1
rs7747448	6	43990902	A	0,34	-0,25	0,03	6,88E-23	MGC45491	85414	VEGFA	128703	INTRONIC	0,58	1
rs4276495	6	43991237	T	0,32	-0,23	0,02	6,46E-22	MGC45491	85079	VEGFA	129038	INTRONIC	0,69	1
rs10117473	9	2670425	G	0,21	-0,22	0,02	7,67E-22	VLDLR	25940	KCNV2	37100	INTERGENIC	1,01	1
rs10967492	9	2671175	A	0,21	-0,22	0,02	1,02E-21	VLDLR	26690	KCNV2	36350	INTERGENIC	1,01	1
rs10967470	9	2665698	G	0,24	-0,22	0,02	1,17E-21	VLDLR	21213	KCNV2	41827	INTERGENIC	0,97	1
rs10812475	9	2681546	C	0,21	-0,22	0,02	1,84E-21	KCNV2	25979	VLDLR	37061	INTERGENIC	0,98	1
rs10812474	9	2681446	G	0,21	-0,22	0,02	1,86E-21	KCNV2	26079	VLDLR	36961	INTERGENIC	0,98	1
rs10812473	9	2681329	A	0,21	-0,22	0,02	1,86E-21	KCNV2	26196	VLDLR	36844	INTERGENIC	0,98	1
rs10812471	9	2680980	T	0,21	-0,22	0,02	1,89E-21	KCNV2	26545	VLDLR	36495	INTERGENIC	0,99	1
rs10125245	9	2679244	T	0,21	-0,22	0,02	2,09E-21	KCNV2	28281	VLDLR	34759	INTERGENIC	1,00	1
rs10967512	9	2673939	C	0,21	-0,22	0,02	2,18E-21	VLDLR	29454	KCNV2	33586	INTERGENIC	1,02	0
rs10757628	9	2674656	T	0,21	-0,22	0,02	2,22E-21	VLDLR	30171	KCNV2	32869	INTERGENIC	1,01	0
rs10812472	9	2681095	G	0,20	-0,23	0,02	4,89E-21	KCNV2	26430	VLDLR	36610	INTERGENIC	0,94	1
rs7867894	9	2665867	C	0,25	-0,21	0,02	8,62E-21	VLDLR	21382	KCNV2	41658	INTERGENIC	0,98	1
rs9472175	6	44042760	T	0,15	-0,56	0,06	9,73E-21	MGC45491	33556	MRPL14	146589	INTERGENIC	0,20	1
rs1776717	6	44059314	A	0,21	-0,23	0,02	8,10E-20	MGC45491	17002	MRPL14	130035	INTERGENIC	0,86	1
rs910609	6	44059634	A	0,21	-0,23	0,02	9,14E-20	MGC45491	16682	MRPL14	129715	INTERGENIC	0,86	1
rs833623	6	44006755	C	0,41	0,18	0,02	1,07E-19	MGC45491	69561	VEGFA	144556	INTRONIC	0,92	1
rs833622	6	44008382	T	0,40	0,17	0,02	1,21E-19	MGC45491	67934	VEGFA	146183	INTRONIC	0,98	0
rs1776721	6	43998961	T	0,30	-0,18	0,02	1,52E-19	MGC45491	77355	VEGFA	136762	INTRONIC	0,99	0
rs9381267	6	44013484	G	0,41	0,17	0,02	3,15E-19	MGC45491	62832	VEGFA	151285	3PRIME_UTR	0,97	1
rs1886979	6	44012879	G	0,41	0,17	0,02	3,72E-19	MGC45491	63437	VEGFA	150680	3PRIME_UTR	0,97	0
rs10122524	9	2681667	T	0,30	-0,19	0,02	3,73E-19	KCNV2	25858	VLDLR	37182	INTERGENIC	0,99	1
rs9472155	6	44005705	T	0,22	-0,20	0,02	4,45E-19	MGC45491	70611	VEGFA	143506	INTRONIC	0,96	1
rs7767550	6	44007230	A	0,22	-0,20	0,02	6,36E-19	MGC45491	69086	VEGFA	145031	INTRONIC	0,99	0
rs3888006	6	44000916	A	0,22	-0,20	0,02	7,60E-19	MGC45491	75400	VEGFA	138717	INTRONIC	0,95	1
rs1750570	6	44000213	T	0,22	-0,20	0,02	8,79E-19	MGC45491	76103	VEGFA	138014	INTRONIC	0,95	1
rs1326141	6	44011255	A	0,22	-0,20	0,02	2,43E-18	MGC45491	65061	VEGFA	149056	INTRONIC	0,98	0
rs910608	6	44059804	C	0,22	-0,21	0,02	3,39E-18	MGC45491	16512	MRPL14	129545	INTERGENIC	0,87	1
rs7356919	6	43999343	A	0,22	-0,20	0,02	5,37E-18	MGC45491	76973	VEGFA	137144	INTRONIC	0,96	1

rs910610	6	44059418	A	0,23	-0,21	0,03	2,50E-17	MGC45491	16898	MRPL14	129931	INTERGENIC	0,79	1
rs1418898	6	44027285	A	0,01	-2,60	0,31	7,27E-17	MGC45491	49031	MRPL14	162064	INTERGENIC	0,14	1
rs1418897	6	44027365	G	0,01	-2,60	0,31	7,37E-17	MGC45491	48951	MRPL14	161984	INTERGENIC	0,14	1
rs9296424	6	44043864	C	0,18	-0,49	0,06	9,98E-17	MGC45491	32452	MRPL14	145485	INTERGENIC	0,18	1
rs1631938	6	44060024	A	0,23	-0,20	0,02	1,74E-16	MGC45491	16292	MRPL14	129325	INTERGENIC	0,83	1
rs7761865	6	43976194	G	0,36	-0,51	0,06	2,21E-16	MGC45491	100122	VEGFA	113995	INTRONIC	0,10	1
rs6993770	8	106650704	T	0,32	-0,17	0,02	2,50E-16	ZFPM2	0	LRP12	980360	INTRONIC	0,97	1
rs9367181	6	44064593	G	0,20	-0,19	0,02	3,61E-16	MGC45491	11723	MRPL14	124756	INTERGENIC	0,98	0
rs1740074	6	44055768	G	0,06	-0,46	0,06	8,17E-16	MGC45491	20548	MRPL14	133581	INTERGENIC	0,44	1
rs11968152	6	43944908	A	0,04	1,16	0,14	1,18E-15	VEGFA	82709	MGC45491	131408	INTERGENIC	0,11	1
rs3932536	6	43944278	A	0,04	1,15	0,14	1,43E-15	VEGFA	82079	MGC45491	132038	INTERGENIC	0,11	1
rs6936668	6	43944048	C	0,04	1,15	0,14	1,45E-15	VEGFA	81849	MGC45491	132268	INTERGENIC	0,11	1
rs7832219	8	106648153	C	0,30	-0,16	0,02	1,72E-15	ZFPM2	0	LRP12	977809	INTRONIC	0,98	1
rs4734879	8	106652300	G	0,30	-0,16	0,02	1,78E-15	ZFPM2	0	LRP12	981956	INTRONIC	0,98	1
rs2343592	8	106641446	G	0,30	-0,16	0,02	1,98E-15	ZFPM2	0	LRP12	971102	INTRONIC	0,97	1
rs4734875	8	106623913	C	0,26	-0,16	0,02	9,46E-15	ZFPM2	0	LRP12	953569	INTRONIC	1,00	0
rs16873346	8	106618657	C	0,26	-0,16	0,02	1,07E-14	ZFPM2	0	LRP12	948313	INTRONIC	0,99	1
rs844294	6	44008685	C	0,51	-0,15	0,02	1,19E-14	MGC45491	67631	VEGFA	146486	INTRONIC	0,97	1
rs4236085	6	43971742	C	0,38	-0,43	0,06	1,27E-14	MGC45491	104574	VEGFA	109543	INTRONIC	0,12	1
rs16873402	8	106658423	T	0,33	-0,15	0,02	1,97E-14	ZFPM2	0	LRP12	988079	INTRONIC	0,99	1
rs910604	6	44061649	A	0,20	-0,18	0,02	2,15E-14	MGC45491	14667	MRPL14	127700	INTERGENIC	0,98	0
rs16873418	8	106661321	G	0,33	-0,15	0,02	2,28E-14	ZFPM2	0	LRP12	990977	INTRONIC	0,99	1
rs6931378	6	43982797	A	0,50	0,37	0,05	7,27E-14	MGC45491	93519	VEGFA	120598	INTRONIC	0,14	1
rs4416670	6	44058431	T	0,55	0,13	0,02	1,47E-12	MGC45491	17885	MRPL14	130918	INTERGENIC	0,99	0
rs1740077	6	44057885	C	0,56	0,13	0,02	5,21E-12	MGC45491	18431	MRPL14	131464	INTERGENIC	0,96	1
rs1776704	6	44057919	G	0,56	0,13	0,02	5,38E-12	MGC45491	18397	MRPL14	131430	INTERGENIC	0,96	1
rs16873365	8	106627411	T	0,21	-0,16	0,02	5,65E-12	ZFPM2	0	LRP12	957067	INTRONIC	0,92	1
rs1776706	6	44057965	C	0,56	0,13	0,02	5,79E-12	MGC45491	18351	MRPL14	131384	INTERGENIC	0,97	1
rs4320369	6	44058040	C	0,56	0,13	0,02	5,98E-12	MGC45491	18276	MRPL14	131309	INTERGENIC	0,97	1
rs1740079	6	44059217	T	0,56	0,13	0,02	6,02E-12	MGC45491	17099	MRPL14	130132	INTERGENIC	0,96	1
rs910612	6	44058684	T	0,44	-0,13	0,02	6,07E-12	MGC45491	17632	MRPL14	130665	INTERGENIC	0,97	1
rs910613	6	44058614	T	0,56	0,13	0,02	6,10E-12	MGC45491	17702	MRPL14	130735	INTERGENIC	0,97	1

rs910614	6	44058596	T	0,44	-0,13	0,02	6,12E-12	MGC45491	17720	MRPL14	130753	INTERGENIC	0,97	1
rs7013321	8	106662734	A	0,48	-0,14	0,02	6,75E-12	ZFPM2	0	LRP12	992390	INTRONIC	0,83	1
rs6993696	8	106650460	A	0,46	-0,13	0,02	8,54E-12	ZFPM2	0	LRP12	980116	INTRONIC	0,99	1
rs4741755	9	2657929	C	0,30	-0,23	0,03	2,26E-11	VLDLR	13444	KCNV2	49596	INTERGENIC	0,37	1
rs16873287	8	106597151	G	0,30	-0,13	0,02	5,13E-11	ZFPM2	0	LRP12	926807	INTRONIC	1,02	0
rs16873291	8	106597206	T	0,30	-0,13	0,02	5,30E-11	ZFPM2	0	LRP12	926862	INTRONIC	1,02	0
rs12676726	8	106597884	C	0,30	-0,13	0,02	5,44E-11	ZFPM2	0	LRP12	927540	INTRONIC	1,01	1
rs9472147	6	43944737	A	0,17	-0,49	0,07	5,78E-11	VEGFA	82538	MGC45491	131579	INTERGENIC	0,12	1
rs10105733	8	106595037	G	0,29	-0,14	0,02	6,09E-11	ZFPM2	0	LRP12	924693	INTRONIC	1,01	1
rs4734119	8	106599699	G	0,30	-0,13	0,02	7,33E-11	ZFPM2	0	LRP12	929355	INTRONIC	1,01	0
rs1157141	8	106600486	G	0,30	-0,13	0,02	7,33E-11	ZFPM2	0	LRP12	930142	INTRONIC	1,01	1
rs1157142	8	106600850	A	0,30	-0,13	0,02	7,34E-11	ZFPM2	0	LRP12	930506	INTRONIC	1,01	1
rs6996138	8	106601653	G	0,30	-0,13	0,02	7,35E-11	ZFPM2	0	LRP12	931309	INTRONIC	1,01	1
rs4734873	8	106602506	A	0,30	-0,13	0,02	7,36E-11	ZFPM2	0	LRP12	932162	INTRONIC	1,01	0
rs7001868	8	106602795	G	0,30	-0,13	0,02	7,38E-11	ZFPM2	0	LRP12	932451	INTRONIC	1,01	1
rs10094510	8	106592361	A	0,29	-0,13	0,02	8,33E-11	ZFPM2	0	LRP12	922017	INTRONIC	1,02	0
rs7007968	8	106591021	G	0,29	-0,13	0,02	9,27E-11	ZFPM2	0	LRP12	920677	INTRONIC	1,02	1
rs16873231	8	106587411	G	0,29	-0,13	0,02	1,36E-10	ZFPM2	0	LRP12	917067	INTRONIC	1,01	1
rs6997293	8	106587104	C	0,29	-0,13	0,02	1,47E-10	ZFPM2	0	LRP12	916760	INTRONIC	1,01	1
rs1868650	8	106587088	C	0,29	-0,13	0,02	1,52E-10	ZFPM2	0	LRP12	916744	INTRONIC	1,01	1
rs1868649	8	106587041	A	0,29	-0,13	0,02	1,57E-10	ZFPM2	0	LRP12	916697	INTRONIC	1,01	1
rs2291192	8	106586248	A	0,29	-0,13	0,02	1,77E-10	ZFPM2	0	LRP12	915904	INTRONIC	1,01	1
rs12678719	8	106585230	G	0,29	-0,13	0,02	1,95E-10	ZFPM2	0	LRP12	914886	INTRONIC	1,01	1
rs10093110	8	106634590	A	0,43	-0,12	0,02	2,18E-10	ZFPM2	0	LRP12	964246	INTRONIC	0,99	0
rs6988664	8	106582477	G	0,29	-0,13	0,02	2,33E-10	ZFPM2	0	LRP12	912133	INTRONIC	1,01	1
rs4734869	8	106582004	A	0,29	-0,13	0,02	2,60E-10	ZFPM2	0	LRP12	911660	INTRONIC	1,01	0
rs910611	6	44058829	C	0,08	-0,26	0,04	2,61E-10	MGC45491	17487	MRPL14	130520	INTERGENIC	0,77	1
rs748785	9	2684347	C	0,13	0,24	0,04	1,01E-09	KCNV2	23178	VLDLR	39862	INTERGENIC	0,52	1
rs7767854	6	44065304	T	0,11	-0,22	0,04	1,02E-09	MGC45491	11012	MRPL14	124045	INTERGENIC	0,74	1
rs6995272	8	106662838	T	0,46	0,14	0,02	1,98E-09	ZFPM2	0	LRP12	992494	INTRONIC	0,66	1
rs12675041	8	106598221	A	0,34	-0,12	0,02	2,10E-09	ZFPM2	0	LRP12	927877	INTRONIC	0,96	1
rs1740080	6	44060390	A	0,12	-0,18	0,03	2,76E-09	MGC45491	15926	MRPL14	128959	INTERGENIC	0,93	1

rs910606	6	44061307	A	0,12	-0,18	0,03	2,90E-09	MGC45491	15009	MRPL14	128042	INTERGENIC	0,94	0
rs1631662	6	44061336	T	0,12	-0,18	0,03	2,93E-09	MGC45491	14980	MRPL14	128013	INTERGENIC	0,94	1
rs2051074	6	44056431	T	0,04	-0,41	0,07	5,80E-09	MGC45491	19885	MRPL14	132918	INTERGENIC	0,46	1
rs9381262	6	43983427	A	0,45	-0,29	0,05	6,78E-09	MGC45491	92889	VEGFA	121228	INTRONIC	0,14	1
rs1450163	8	106601298	C	0,37	-0,11	0,02	1,11E-08	ZFPM2	0	LRP12	930954	INTRONIC	1,00	1
rs7836542	8	106603619	T	0,37	-0,11	0,02	1,13E-08	ZFPM2	0	LRP12	933275	INTRONIC	1,01	0
rs16897113	6	44076194	T	0,03	0,62	0,11	1,34E-08	MGC45491	122	MRPL14	113155	UPSTREAM	0,29	1
rs2343595	8	106660383	G	0,45	-0,11	0,02	2,26E-08	ZFPM2	0	LRP12	990039	INTRONIC	1,01	1
rs16873415	8	106660978	G	0,45	-0,11	0,02	2,39E-08	ZFPM2	0	LRP12	990634	INTRONIC	1,01	0
rs1901061	8	106619426	A	0,39	0,11	0,02	3,08E-08	ZFPM2	0	LRP12	949082	INTRONIC	0,94	1
rs1349319	8	106625810	A	0,39	0,11	0,02	3,59E-08	ZFPM2	0	LRP12	955466	INTRONIC	0,94	1
rs9369430	6	43942229	G	0,16	0,24	0,04	3,68E-08	VEGFA	80030	MGC45491	134087	INTERGENIC	0,33	1
rs748786	9	2684436	T	0,12	0,23	0,04	3,91E-08	KCNV2	23089	VLDLR	39951	INTERGENIC	0,52	1
rs4734122	8	106629546	A	0,41	0,11	0,02	4,15E-08	ZFPM2	0	LRP12	959202	INTRONIC	0,99	1
rs1375955	8	106624215	T	0,39	0,11	0,02	4,30E-08	ZFPM2	0	LRP12	953871	INTRONIC	0,94	1
rs12679049	8	106598022	G	0,22	-0,12	0,02	4,75E-08	ZFPM2	0	LRP12	927678	INTRONIC	0,99	0

SNP: single nucleotide polymorphism; Chr: chromosome; MA: minor allele; MAF: minor allele frequency; SE: standard error; p: p-value; O/E ratio: observed over expected ratio (measuring imputation accuracy); KCNV2: potassium voltage-gated channel subfamily V, member 2; LRP12: low-density lipoprotein receptor-related protein gene; MCG45491: uncharacterized protein, also known as C6orf223; MRPL14: mitochondrial ribosomal protein L14; VEGF: vascular endothelial growth factor; VLDLR: very low density lipoprotein receptor; ZFPM2: zinc finger protein, multitype 2

Supplemental Table 3: Secondary genetic association analysis adjusting for clinical covariates previously found to be associated with VEGF levels

SNPID	Chr	Position	p (FHS)	p (SFS)	p (PIVUS)	meta-p (replication)	meta-p (all)
Model B							
rs6921438	6	44033585	$1.72x10^{-506}$	$1.84x10^{-39}$	NA	$1.84x10^{-39}$	$3.05x10^{-528}$
rs4513773	6	44033504	$1.58x10^{-482}$	NA	$7.98x10^{-139}$	$7.98x10^{-139}$	$4.41x10^{-619}$
rs9472159	6	44027673	$2.89x10^{-452}$	$3.58x10^{-35}$	$2.90x10^{-109}$	$4.14x10^{-134}$	$5.27x10^{-570}$
rs9369434	6	44026385	$8.70x10^{-443}$	$5.81x10^{-28}$	$3.02x10^{-63}$	$6.46x10^{-86}$	$8.61x10^{-490}$
rs1776717	6	44059314	$1.23x10^{-19}$	$8.96x10^{-6}$	$2.27x10^{-4}$	$7.97x10^{-9}$	$1.68x10^{-26}$
rs1776721	6	43998961	$1.17x10^{-19}$	0.018	$4.34x10^{-8}$	$4.88x10^{-8}$	$4.88x10^{-26}$
rs1886979	6	44012879	$3.23x10^{-19}$	0.013	$2.71x10^{-6}$	$5.99x10^{-7}$	$2.91x10^{-24}$
rs9472155	6	44005705	$4.98x10^{-19}$	0.014	$4.83x10^{-9}$	$7.64x10^{-9}$	$2.91x10^{-26}$
rs844294	6	44008685	$1.34x10^{-14}$	0.087	$1.95x10^{-5}$	$3.93x10^{-5}$	$6.09x10^{-18}$
rs4416670	6	44058431	$2.04x10^{-12}$	$2.79x10^{-4}$	0,1	$1.39x10^{-4}$	$2.38x10^{-15}$
rs910611	6	44058829	$4.77x10^{-10}$	0.11	$4.72x10^{-6}$	$2.42x10^{-5}$	$9.49x10^{-14}$
rs6993770	8	106650704	$2.03x10^{-16}$	0.016	$3.61x10^{-8}$	$7.60x10^{-8}$	$9.31x10^{-23}$
rs16873402	8	106658423	$1.45x10^{-14}$	0.14	$1.16x10^{-8}$	$1.77x10^{-6}$	$1.52x10^{-19}$
rs16873365	8	106627411	$6.93x10^{-12}$	0.37	$2.10x10^{-6}$	$1.64x10^{-4}$	$5.21x10^{-15}$
rs7013321	8	106662734	$4.66x10^{-12}$	0.013	NA	0.013	$6.03x10^{-13}$
rs6993696	8	106650460	$8.83x10^{-12}$	0.040	$1.57x10^{-4}$	$5.11x10^{-5}$	$2.79x10^{-15}$
rs16873291	8	106597206	$5.43x10^{-11}$	0.061	$1.42x10^{-6}$	$5.51x10^{-6}$	$1.47x10^{-15}$
rs1349319	8	106625810	$2.93x10^{-8}$	0.040	$1.73x10^{-3}$	$3.11x10^{-4}$	$4.11x10^{-11}$
rs10738760	9	2681186	$1.17x10^{-34}$	0.035	$3.24x10^{-9}$	$2.70x10^{-8}$	$9.12x10^{-38}$
rs6475920	9	2663933	$2.23x10^{-32}$	0.022	$1.46x10^{-8}$	$4.39x10^{-8}$	$2.88x10^{-37}$
rs4741756	9	2658187	$2.29x10^{-31}$	0.097	$7.05x10^{-5}$	$1.20x10^{-4}$	$3.33x10^{-31}$
rs2375980	9	2682622	$1.53x10^{-27}$	0.022	$8.92x10^{-9}$	$2.59x10^{-8}$	$8.85x10^{-6}$
rs10122587	9	2681951	$3.89x10^{-24}$	0.026	NA	0.026	$2.67x10^{-23}$
rs10967492	9	2671175	$1.86x10^{-21}$	0.11	NA	0.11	$1.14x10^{-19}$
rs10967470	9	2665698	$2.19x10^{-21}$	0.044	NA	0.044	$2.79x10^{-20}$
Model C							
rs6921438	6	44033585	$3.73x10^{-507}$	$3.54x10^{-32}$	NA	$3.54x10^{-32}$	$4.16x10^{-523}$
rs4513773	6	44033504	$4.54x10^{-484}$	NA	$1.13x10^{-143}$	$1.13x10^{-143}$	$1.93x10^{-625}$
rs9472159	6	44027673	$5.87x10^{-454}$	$3.49x10^{-30}$	$6.67x10^{-113}$	$3.65x10^{-133}$	$8.42x10^{-573}$
rs9369434	6	44026385	$4.64x10^{-445}$	$1.76x10^{-26}$	$3.27x10^{-65}$	$4.00x10^{-87}$	$1.28x10^{-497}$
rs1776717	6	44059314	$8.01x10^{-20}$	$2.77x10^{-5}$	$5.20x10^{-4}$	$5.79x10^{-8}$	$8.46x10^{-26}$
rs1776721	6	43998961	$1.05x10^{-19}$	0.025	$5.89x10^{-9}$	$1.15x10^{-8}$	$8.29x10^{-27}$
rs1886979	6	44012879	$2.65x10^{-19}$	$2.84x10^{-3}$	$2.12x10^{-6}$	$4.80x10^{-8}$	$1.08x10^{-25}$
rs9472155	6	44005705	$5.96x10^{-19}$	0.010	$1.33x10^{-9}$	$1.07x10^{-9}$	$4.19x10^{-27}$
rs844294	6	44008685	$1.63x10^{-14}$	0.041	$1.95x10^{-5}$	$9.33x10^{-6}$	$1.13x10^{-18}$
rs4416670	6	44058431	$1.16x10^{-12}$	$1.49x10^{-4}$	0,11	$1.19x10^{-4}$	$1.03x10^{-15}$
rs910611	6	44058829	$3.37x10^{-10}$	0.076	$7.04x10^{-6}$	$1.16x10^{-5}$	$2.59x10^{-14}$
rs6993770	8	106650704	$2.11x10^{-16}$	$2.67x10^{-3}$	$6.87x10^{-8}$	$5.07x10^{-9}$	$7.00x10^{-24}$
rs16873402	8	106658423	$1.26x10^{-14}$	0.043	$1.75x10^{-8}$	$1.26x10^{-7}$	$9.13x10^{-21}$
rs16873365	8	106627411	$7.36x10^{-12}$	0.26	$2.02x10^{-6}$	$4.49x10^{-5}$	$1.56x10^{-15}$
rs7013321	8	106662734	$3.41x10^{-12}$	$2.21x10^{-3}$	NA	$2.21x10^{-3}$	$3.85x10^{-14}$
rs6993696	8	106650460	$6.23x10^{-12}$	$2.57x10^{-3}$	$2.09x10^{-4}$	$2.10x10^{-6}$	$6.73x10^{-17}$

rs16873291	8	106597206	$4.25x10^{-11}$	0.031	$9.77x10^{-7}$	$9.07x10^{-7}$	$2.19x10^{-16}$
rs1349319	8	106625810	$2.33x10^{-8}$	$7.71x10^{-3}$	$1.61x10^{-3}$	$4.08x10^{-5}$	$4.21x10^{-12}$
rs10738760	9	2681186	$1.91x10^{-34}$	$9.24x10^{-3}$	$5.53x10^{-8}$	$1.39x10^{-8}$	$1.69x10^{-38}$
rs6475920	9	2663933	$3.43x10^{-32}$	$6.55x10^{-3}$	$9.54x10^{-8}$	$1.53x10^{-8}$	$4.95x10^{-38}$
rs4741756	9	2658187	$5.22x10^{-31}$	0.049	$2.27x10^{-4}$	$7.77x10^{-5}$	$1.09x10^{-31}$
rs2375980	9	2682622	$1.15x10^{-27}$	0.021	$8.63x10^{-8}$	$6.45x10^{-8}$	$1.37x10^{-6}$
rs10122587	9	2681951	$4.10x10^{-24}$	0.027	NA	0.027	$1.07x10^{-23}$
rs10967492	9	2671175	$2.39x10^{-21}$	0.19	NA	0.19	$1.59x10^{-19}$
rs10967470	9	2665698	$2.66x10^{-21}$	0.096	NA	0.096	$5.93x10^{-20}$
Model D							
rs6921438	6	44033585	$1.53x10^{-508}$	$1.40x10^{-40}$	NA	$1.40x10^{-40}$	$2.18x10^{-532}$
rs4513773	6	44033504	$2.35x10^{-485}$	NA	$6.36x10^{-136}$	$6.36x10^{-136}$	$4.72x10^{-619}$
rs9472159	6	44027673	$1.61x10^{-454}$	$1.52x10^{-36}$	$4.22x10^{-108}$	$3.75x10^{-135}$	$7.36x10^{-574}$
rs9369434	6	44026385	$6.46x10^{-445}$	$2.70x10^{-29}$	$7.63x10^{-63}$	$2.57x10^{-87}$	$1.97x10^{-494}$
rs1776717	6	44059314	$2.43x10^{-19}$	$7.08x10^{-6}$	$2.96x10^{-4}$	$8.35x10^{-9}$	$3.06x10^{-26}$
rs1776721	6	43998961	$7.12x10^{-20}$	0.013	$4.46x10^{-8}$	$3.03x10^{-8}$	$1.74x10^{-26}$
rs1886979	6	44012879	$1.54x10^{-19}$	$7.09x10^{-3}$	$2.15x10^{-6}$	$2.20x10^{-7}$	$4.29x10^{-25}$
rs9472155	6	44005705	$1.37x10^{-19}$	$7.43x10^{-3}$	$4.11x10^{-9}$	$2.40x10^{-9}$	$2.42x10^{-27}$
rs844294	6	44008685	$2.77x10^{-15}$	0.048	$1.48x10^{-5}$	$1.32x10^{-5}$	$3.74x10^{-19}$
rs4416670	6	44058431	$1.95x10^{-12}$	$2.05x10^{-4}$	0,088	$9.72x10^{-5}$	$1.44x10^{-15}$
rs910611	6	44058829	$3.19x10^{-10}$	0.11	$3.45x10^{-6}$	$2.17x10^{-5}$	$5.64x10^{-14}$
rs6993770	8	106650704	$5.19x10^{-16}$	0.016	$5.30x10^{-8}$	$1.00x10^{-7}$	$3.07x10^{-22}$
rs16873402	8	106658423	$3.72x10^{-14}$	0.15	$1.34x10^{-8}$	$2.06x10^{-6}$	$4.34x10^{-19}$
rs16873365	8	106627411	$2.38x10^{-11}$	0.36	$3.87x10^{-6}$	$2.02x10^{-4}$	$2.13x10^{-14}$
rs7013321	8	106662734	$8.07x10^{-12}$	0.013	NA	0.013	$8.97x10^{-13}$
rs6993696	8	106650460	$1.36x10^{-11}$	0.040	$2.41x10^{-4}$	$6.88x10^{-5}$	$5.59x10^{-15}$
rs16873291	8	106597206	$9.39x10^{-11}$	0.071	$1.37x10^{-6}$	$6.34x10^{-6}$	$2.91x10^{-15}$
rs1349319	8	106625810	$3.87x10^{-8}$	0.057	$2.46x10^{-3}$	$6.05x10^{-4}$	$1.06x10^{-10}$
rs10738760	9	2681186	$1.36x10^{-35}$	0.046	$1.06x10^{-8}$	$8.27x10^{-8}$	$6.96x10^{-38}$
rs6475920	9	2663933	$9.18x10^{-33}$	0.018	$2.67x10^{-8}$	$4.29x10^{-8}$	$1.06x10^{-37}$
rs4741756	9	2658187	$9.04x10^{-32}$	0.071	$1.28x10^{-4}$	$1.07x10^{-4}$	$9.64x10^{-32}$
rs2375980	9	2682622	$3.07x10^{-28}$	0.015	$3.55x10^{-8}$	$3.36x10^{-8}$	$3.16x10^{-6}$
rs10122587	9	2681951	$5.73x10^{-25}$	0.020	NA	0.020	$2.70x10^{-24}$
rs10967492	9	2671175	$3.72x10^{-22}$	0.091	NA	0.091	$1.58x10^{-20}$
rs10967470	9	2665698	$5.48x10^{-22}$	0.045	NA	0.045	$6.85x10^{-21}$
Model E							
rs6921438	6	44033585	$6.58x10^{-506}$	$2.91x10^{-39}$	NA	$2.91x10^{-39}$	$1.60x10^{-527}$
rs4513773	6	44033504	$6.46x10^{-482}$	NA	$3.04x10^{-138}$	$3.04x10^{-138}$	$6.76x10^{-618}$
rs9472159	6	44027673	$9.84x10^{-452}$	$4.30x10^{-35}$	$4.86x10^{-110}$	$9.54x10^{-135}$	$2.47x10^{-570}$
rs9369434	6	44026385	$2.44x10^{-442}$	$5.33x10^{-28}$	$3.05x10^{-65}$	$1.06x10^{-87}$	$5.57x10^{-492}$
rs1776717	6	44059314	$9.92x10^{-20}$	$1.00x10^{-5}$	$2.53x10^{-4}$	$9.88x10^{-9}$	$1.75x10^{-26}$
rs1776721	6	43998961	$5.41x10^{-20}$	0.020	$2.57x10^{-8}$	$4.32x10^{-8}$	$2.05x10^{-26}$
rs1886979	6	44012879	$2.55x10^{-19}$	0.013	$1.66x10^{-6}$	$4.30x10^{-7}$	$1.57x10^{-24}$
rs9472155	6	44005705	$1.74x10^{-19}$	0.015	$3.04x10^{-9}$	$6.16x10^{-9}$	$8.38x10^{-27}$
rs844294	6	44008685	$8.15x10^{-15}$	0.086	$1.39x10^{-5}$	$3.13x10^{-5}$	$2.92x10^{-18}$
rs4416670	6	44058431	$1.86x10^{-12}$	$2.89x10^{-4}$	0,11	$1.68x10^{-4}$	$2.73x10^{-15}$
rs910611	6	44058829	$2.35x10^{-10}$	0.11	$1.99x10^{-6}$	$1.42x10^{-5}$	$2.75x10^{-14}$

rs6993770	8	106650704	$1.34x10^{-16}$	0.018	$3.63x10^{-8}$	$9.82x10^{-8}$	$8.06x10^{-23}$
rs16873402	8	106658423	$1.24x10^{-14}$	0.16	$8.58x10^{-9}$	$1.83x10^{-6}$	$1.35x10^{-19}$
rs16873365	8	106627411	$8.10x10^{-12}$	0.37	$4.01x10^{-6}$	$2.35x10^{-4}$	$8.66x10^{-15}$
rs7013321	8	106662734	$3.64x10^{-12}$	0.014	NA	0.014	$5.40x10^{-13}$
rs6993696	8	106650460	$5.82x10^{-12}$	0.047	$1.73x10^{-4}$	$6.89x10^{-5}$	$2.62x10^{-15}$
rs16873291	8	106597206	$3.14x10^{-11}$	0.066	$7.92x10^{-7}$	$4.31x10^{-6}$	$6.74x10^{-16}$
rs1349319	8	106625810	$3.53x10^{-8}$	0.046	$1.84x10^{-3}$	$3.93x10^{-4}$	$6.26x10^{-11}$
rs10738760	9	2681186	$6.77x10^{-35}$	0.034	$3.90x10^{-9}$	$2.98x10^{-8}$	$6.55x10^{-38}$
rs6475920	9	2663933	$1.85x10^{-32}$	0.022	$9.61x10^{-9}$	$3.41x10^{-8}$	$1.66x10^{-37}$
rs4741756	9	2658187	$1.32x10^{-31}$	0.092	$6.02x10^{-5}$	$1.02x10^{-4}$	$1.51x10^{-31}$
rs2375980	9	2682622	$7.94x10^{-28}$	0.022	$1.45x10^{-8}$	$3.67x10^{-8}$	$5.38x10^{-6}$
rs10122587	9	2681951	$1.47x10^{-24}$	0.024	NA	0.024	$1.01x10^{-23}$
rs10967492	9	2671175	$6.75x10^{-22}$	0.10	NA	0.10	$4.51x10^{-20}$
rs10967470	9	2665698	$9.41x10^{-22}$	0.045	NA	0.045	$1.37x10^{-20}$

All analyses were adjusted for age, sex, familial structure and principal components. Model B was additionally adjusted for hypertension; model C for smoking; model D for central obesity; model E for the presence of a metabolic syndrome

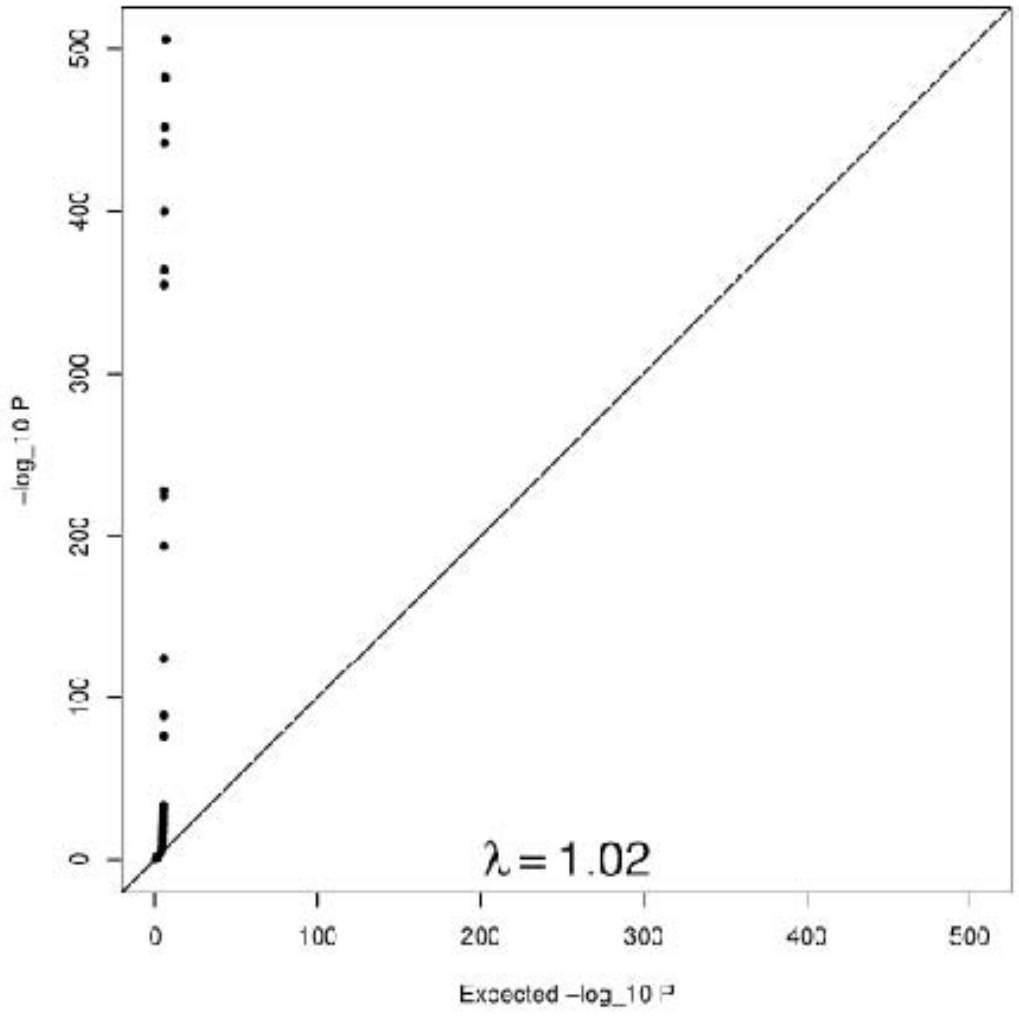
Supplemental Table 4: Candidate gene association studies of VEGF variants and circulating VEGF levels

Author	Population	N	Sample	SNP										
				rs699947	rs1570360	rs833061	rs2010963	rs3025039	rs1005230	rs25648	rs10434	-1190G>A	702C>T	-2549 18bp I/D
Steffensen [22]	ovarian cancer patients	143	serum	C ↑ VEGF	NS	T ↑ VEGF	C ↑ VEGF	NS	-	-	-	-	-	-
Costa [16]	post-menopausal women	252	serum	-	-	-	NS	NS	-	-	-	-	-	-
Medford [17]	patients with (and at risk of) ARDS	70	plasma	-	-	-	-	NS	-	-	-	-	-	-
Kamoun [23]	Behcet patients and healthy controls	135	serum	-	-	-	NS	NS	-	-	-	-	-	D ↑ VEGF
Langsenlehner [18]	healthy persons	81	plasma	NS	-	NS	NS	NS	NS	NS	NS	-	-	-
Petrovic [29]	PDR patients	104	serum	-	-	-	CC ↑ VEGF	-	-	-	-	-	-	-
Mateo [19]	AD patients and healthy controls	117	serum	NS	-	-	NS	-	-	-	-	-	-	-
Balasubramanian, [20]	post-menopausal women	62	serum and plasma	-	-	NS	NS	NS	-	NS	-	-	-	-
Zhai [24]	ARDS patients	71	plasma	-	-	NS	NS	T ↓ VEGF	-	-	-	-	-	-
Berrahmoune [21]	community sample	647	plasma	-	-	NS	NS	NS	-	-	-	-	-	-
Ferrante [25]	IBD patients and healthy controls	1142	serum	NS	GG ↑ VEGF	-	NS	NS	-	-	-	-	-	-
Krippl [26]	post-menopausal women	21	plasma	-	-	-	-	T ↓ VEGF	-	-	-	-	-	-
Awata [27]	healthy persons	64	serum	-	NS	NS	CC ↑ VEGF	NS	-	NS	NS	NS	-	-
Renner [28]	healthy persons	23	plasma	-	-	-	-	T ↓ VEGF	-	-	NS	-	NS	-
Present study	**community sample**	**3546**	**serum**	**C ↑ VEGF $p=2.33x10^{-7}$ (beta±SE= 0.13±0.02)**	**NA**	**T ↑ VEGF $p=3.05x10^{-7}$ (beta±SE= 0.12±0.02)**	**NA**	**NA**						**NA**

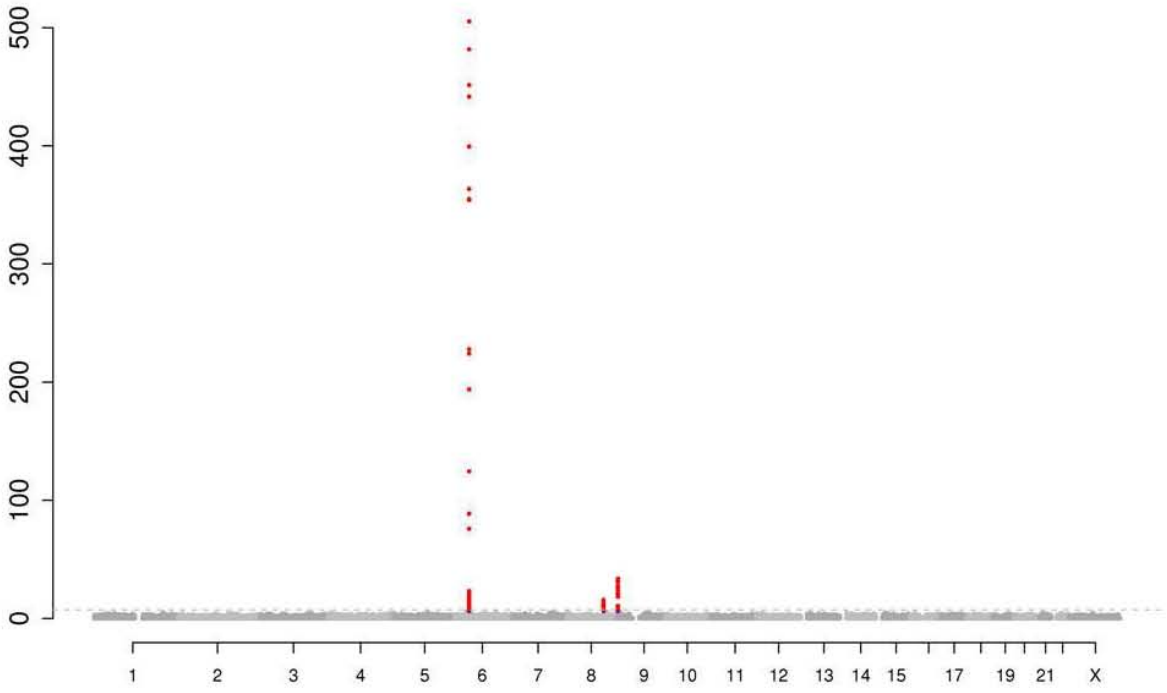
AD: Alzheimer Disease; ARDS: Acute Respiratory Distress Syndrome; IBD: Inflammatory Bowel Disease; HAM/TSP: HTLV-I- associated myelopathy/tropical spastic paraparesis; NS: non significant (no association); NA: not available; PDR = Proliferative Diabetic Retinopathy

C. Supplemental Figures

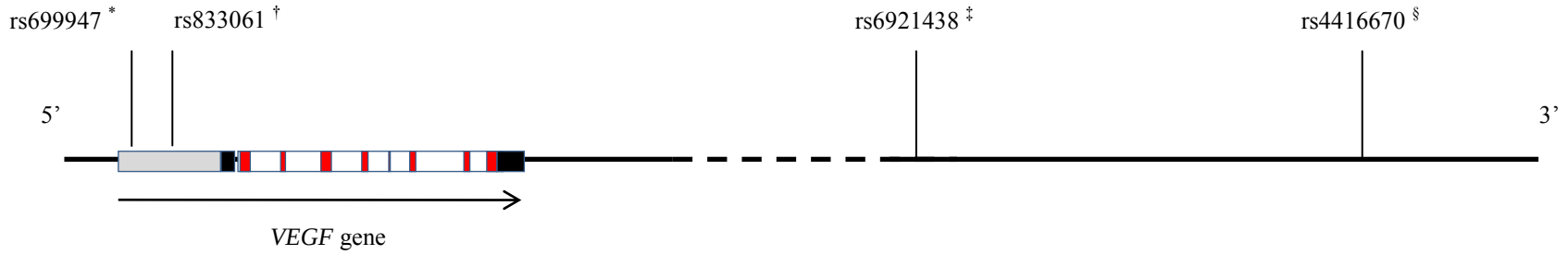
Supplemental Figure 1: Quantile-quantile QQ-plot showing the observed versus the expected p-values after meta-analysis for serum VEGF levels (the solid line shows the distribution under the null-hypothesis)



Supplemental Figure 2: Manhattan plot showing individual p-values against their genomic position for GWAS of serum VEGF levels. Within each chromosome (x-axis), results are plotted left to right from p-terminal end. Dashed line indicates preset threshold for genome-wide significance, p=5.0x10-8; solid line threshold for suggestive associations, $p=4.0x10^{-7}$



Supplemental Figure 3: Genomic organization of the VEGF gene and localization of the SNPs identified in the genetic association and transcriptomic studies



*chr6:43844367, p=2.33x10^{-7} replicates findings from previous association studies (Supplementary Table 4)
†chr6:43845464, p=3.05x10^{-7}, replicates findings from previous association studies (Supplementary Table 4)
‡ chr6:44033585, p=6.11x10-^{506}, top SNP, explains 41.2% of the phenotypic variance in the FHS
§ chr6:44058431, p=1.47x10^{-12}, associated with serum VEGF levels independently of rs6921438 in conditional GWAS, explains 1.5% of the phenotypic variance in the FHS

References

1. Dawber TR, Kannel WB. The Framingham study. An epidemiological approach to coronary heart disease. *Circulation.* 1966;34:553-555.

2. Feinleib M, Kannel WB, Garrison RJ, McNamara PM, Castelli WP. The Framingham Offspring Study. Design and preliminary data. *Prev Med.* 1975;4:518-525.

3. Splansky GL, Corey D, Yang Q, Atwood LD, Cupples LA, Benjamin EJ, D'Agostino RB, Sr., Fox CS, Larson MG, Murabito JM, O'Donnell CJ, Vasan RS, Wolf PA, Levy D. The Third Generation Cohort of the National Heart, Lung, and Blood Institute's Framingham Heart Study: design, recruitment, and initial examination. *Am J Epidemiol.* 2007;165:1328-1335.

4. Visvikis-Siest S, Siest G. The STANISLAS Cohort: a 10-year follow-up of supposed healthy families. Gene-environment interactions, reference values and evaluation of biomarkers in prevention of cardiovascular diseases. *Clin Chem Lab Med.* 2008;46:733-747.

5. Siest G, Visvikis S, Herbeth B, Gueguen R, Vincent-Viry M, Sass C, Beaud B, Lecomte E, Steinmetz J, Locuty J, Chevrier P. Objectives, design and recruitment of a familial and longitudinal cohort for studying gene-environment interactions in the field of cardiovascular risk: the Stanislas cohort. *Clin Chem Lab Med.* 1998;36:35-42.

6. Lind L, Fors N, Hall J, Marttala K, Stenborg A. A comparison of three different methods to evaluate endothelium-dependent vasodilation in the elderly: the Prospective Investigation of the Vasculature in Uppsala Seniors (PIVUS) study. *Arterioscler Thromb Vasc Biol.* 2005;25:2368-2375.

7. Fitzgerald SP, Lamont JV, McConnell RI, Benchikh el O. Development of a high-throughput automated analyzer using biochip array technology. *Clin Chem.* 2005;51:1165-1176.

8. Fan JB, Oliphant A, Shen R, Kermani BG, Garcia F, Gunderson KL, Hansen M, Steemers F, Butler SL, Deloukas P, Galver L, Hunt S, McBride C, Bibikova M, Rubano T, Chen J, Wickham E, Doucet D, Chang W, Campbell D, Zhang B, Kruglyak S, Bentley D, Haas J, Rigault P, Zhou L, Stuelpnagel J, Chee MS. Highly parallel SNP genotyping. *Cold Spring Harb Symp Quant Biol.* 2003;68:69-78.

9. Price AL, Patterson NJ, Plenge RM, Weinblatt ME, Shadick NA, Reich D. Principal components analysis corrects for stratification in genome-wide association studies. *Nat Genet.* 2006;38:904-909.

10. Patterson N, Price AL, Reich D. Population structure and eigenanalysis. *PLoS Genet.* 2006;2:e190.

11. Grundy SM, Cleeman JI, Daniels SR, Donato KA, Eckel RH, Franklin BA, Gordon DJ, Krauss RM, Savage PJ, Smith SC, Jr., Spertus JA, Costa F. Diagnosis and management of the metabolic syndrome: an American Heart Association/National Heart, Lung, and Blood Institute Scientific Statement. *Circulation.* 2005;112:2735-2752.

12. Chen MH, Yang Q. GWAF: an R package for genome-wide association analyses with family data. *Bioinformatics.*26:580-581.

13. Ma HW, Zhao XM, Yuan YJ, Zeng AP. Decomposition of metabolic network into functional modules based on the global connectivity structure of reaction graph. *Bioinformatics.* 2004;20:1870-1876.

14. Marteau JB, Mohr S, Pfister M, Visvikis-Siest S. Collection and storage of human blood cells for mRNA expression profiling: a 15-month stability study. *Clin Chem.* 2005;51:1250-1252.

15. Mourah S, Porcher R, Lescaille G, Rousselot P, Podgorniak MP, Labarchede G, Naimi B, Medioni J, Dombret H, Calvo F. Quantification of VEGF isoforms and VEGFR transcripts by qRT-PCR and their significance in acute myeloid leukemia. *Int J Biol Markers.* 2009;24:22-31.

16. Costa N, Paramanathan S, Mac Donald D, Wierzbicki AS, Hampson G. Factors regulating circulating vascular endothelial growth factor (VEGF): association with bone mineral density (BMD) in post-menopausal osteoporosis. *Cytokine.* 2009;46:376-381.

17. Medford AR, Godinho SI, Keen LJ, Bidwell JL, Millar AB. Relationship between vascular endothelial growth factor + 936 genotype and plasma/epithelial lining fluid vascular endothelial growth factor protein levels in patients with and at risk for ARDS. *Chest.* 2009;136:457-464.

18. Langsenlehner U, Wolf G, Langsenlehner T, Gerger A, Hofmann G, Clar H, Wascher TC, Paulweber B, Samonigg H, Krippl P, Renner W. Genetic polymorphisms in the vascular endothelial growth factor gene and breast cancer risk. The Austrian "tumor of breast tissue: incidence, genetics, and environmental risk factors" study. *Breast Cancer Res Treat.* 2008;109:297-304.

19. Mateo I, Llorca J, Infante J, Rodriguez-Rodriguez E, Fernandez-Viadero C, Pena N, Berciano J, Combarros O. Low serum VEGF levels are associated with Alzheimer's disease. *Acta Neurol Scand.* 2007;116:56-58.

20. Balasubramanian SP, Cox A, Cross SS, Higham SE, Brown NJ, Reed MW. Influence of VEGF-A gene variation and protein levels in breast cancer susceptibility and severity. *Int J Cancer.* 2007;121:1009-1016.

21. Berrahmoune H, Herbeth B, Lamont JV, Masson C, Fitzgerald PS, Visvikis-Siest S. Heritability for plasma VEGF concentration in the Stanislas family study. *Ann Hum Genet.* 2007;71:54-63.

22. Steffensen KD, Waldstrom M, Brandslund I, Jakobsen A. The relationship of VEGF polymorphisms with serum VEGF levels and progression-free survival in patients with epithelial ovarian cancer. *Gynecol Oncol.*117:109-116.

23. Kamoun M, Houman MH, Hamzaoui A, Hamzaoui K. Vascular endothelial growth factor gene polymorphisms and serum levels in Behcet's disease. *Tissue Antigens.* 2008;72:581-587.

24. Zhai R, Gong MN, Zhou W, Thompson TB, Kraft P, Su L, Christiani DC. Genotypes and haplotypes of the VEGF gene are associated with higher mortality and lower VEGF plasma levels in patients with ARDS. *Thorax.* 2007;62:718-722.

25. Ferrante M, Pierik M, Henckaerts L, Joossens M, Claes K, Van Schuerbeek N, Vlietinck R, Rutgeerts P, Van Assche G, Vermeire S. The role of vascular endothelial growth factor (VEGF) in inflammatory bowel disease. *Inflamm Bowel Dis.* 2006;12:870-878.

26. Krippl P, Langsenlehner U, Renner W, Yazdani-Biuki B, Wolf G, Wascher TC, Paulweber B, Haas J, Samonigg H. A common 936 C/T gene polymorphism of vascular endothelial growth factor is associated with decreased breast cancer risk. *Int J Cancer.* 2003;106:468-471.

27. Awata T, Inoue K, Kurihara S, Ohkubo T, Watanabe M, Inukai K, Inoue I, Katayama S. A common polymorphism in the 5'-untranslated region of the VEGF gene is associated with diabetic retinopathy in type 2 diabetes. *Diabetes.* 2002;51:1635-1639.

28. Renner W, Kotschan S, Hoffmann C, Obermayer-Pietsch B, Pilger E. A common 936 C/T mutation in the gene for vascular endothelial growth factor is associated with vascular endothelial growth factor plasma levels. *J Vasc Res.* 2000;37:443-448.

29. Petrovic MG, Korosec P, Kosnik M, Osredkar J, Hawlina M, Peterlin B, Petrovic D. Local and genetic determinants of vascular endothelial growth factor expression in advanced proliferative diabetic retinopathy. *Mol Vis.* 2008;14:1382-1387.