Étude des interactions entre la fonction rénale, la thrombose artérielle et la thrombose veineuse par une approche génétique

Romain Charmet

▶ **To cite this version:**

Romain Charmet. Étude des interactions entre la fonction rénale, la thrombose artérielle et la thrombose veineuse par une approche génétique. Médecine humaine et pathologie. Sorbonne Université, 2018. Français. NNT : 2018SORUS392 . tel-02613797

HAL Id: tel-02613797

https://tel.archives-ouvertes.fr/tel-02613797

Submitted on 20 May 2020

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

THÈSE DE DOCTORAT DE L'UNIVERSITÉ PIERRE ET MARIE CURIE

Spécialité

Épidémiologie génétique

Supervisée par l'école doctorale Pierre Louis de santé publique :

Épidémiologie et sciences de l'information biomédicale

Sujet de la thèse :

Étude des interactions entre la fonction rénale, la thrombose artérielle et la thrombose veineuse par une approche génétique

---

Soutenue par

Romain Charmet

Pour obtenir le grade de

Docteur de l'université Pierre et Marie Curie

Préparée au laboratoire **UMRS1166 - Maladies cardiovasculaires, du métabolisme et de la nutrition**

dirigée par **Dr David-Alexandre Trégouët**

23 Mai 2018

Jury :

M. David-Alexandre Trégouët	Directeur de thèse
Mme. Florence Demenais-Diao	Rapportrice et présidente du jury
M. Joseph Emmerich	Rapporteur
M. Frédéric Fumeron	Examinateur
Mme. Genevieve Plu-Bureau	Examinatrice

Why, sometimes I've believed as many as six impossible things before breakfast.
Quelquefois, j'ai cru jusqu'à six choses impossibles avant le petit déjeuner.

Lewis Carroll - Alice de l'autre côté du miroir

Remerciements

Il n'est pas de travail bien achevé sans rendre hommage à ceux qui y ont contribué par leurs actions ou par leur soutien. Cette thèse ne saurait déroger à la règle et je commencerai donc par remercier toutes les personnes qui m'ont apporté leur assistance durant cette période. Bien que la liste des personnes que je dois remercier soit grande je vais tâcher d'être exhaustif.

En premier lieu, je tiens à remercier mon directeur de thèse David Trégouët, qui m'a offert l'opportunité d'effectuer cette thèse, m'a accompagné durant toute la durée de mon travail et qui m'a permis de découvrir le monde de la recherche académique. Je tiens également à remercier les membres de mon jury de thèse qui ont pris le temps de s'y intéresser. Mes rapporteurs, Monsieur Joseph Emmerich ainsi que Madame Florence Demenais, qui ont accepté de relire le manuscrit et m'ont permis de le perfectionner. Mes examinateurs, Monsieur Frédéric Fumeron et Madame Genevieve Plu-Bureau, qui ont eu la gentillesse de bien vouloir participer à l'évaluation de mon travail. Enfin, je remercie Laurence Tiret pour m'avoir accepté en tant que stagiaire dans le laboratoire et qui m'a mis le pied à l'étrier, ainsi que Sophie Garnier qui m'a présenté l'équipe.

Je remercie également tous mes collègues et collaborateurs, mes camarades de bureau François et Nadjim qui m'ont accueillis à mon arrivée mais aussi tous les autres, Dylan, Ulrike, Veronica, Têtê, Badreddine, Waqasuddin, Florian, Béata, Marine, Varma, Anne-Sophie, Claire, Arès, Nathalie grâce à qui j'ai pu me diriger dans le labyrinthe administratif de l'université et Carole pour son bò bún et sans qui je n'aurais toujours pas de dentiste dans cette ville. Impossible d'oublier également mes collègues et amis, Henri et Gwladys, avec qui j'ai pu découvrir toutes les spécialités culinaires que la rive gauche et le centre avaient à offrir et avec qui j'ai pu partager d'agréables vacances en Normandie. Merci à Henri pour avoir relu mon manuscrit et à qui je dois d'ailleurs rendre une visite à Hambourg car on m'a dit le plus grand bien de leurs harengs pomme à l'huile. Enfin merci à Leonardo que j'ai rencontré lors d'un de mes déplacements et qui m'a fait la bonne surprise de me rendre visite à Paris. Merci à mes camarades doctorants, bien que certains d'entre eux auront déjà soutenu leur thèse

au moment où je soutiendrai la mienne, Adeline, Pierre, Héloïse, Sophie, Dhoha et Veronica. Partager mes aventures avec vous a été enrichissant et très amusant.

Naturellement, je remercie ma famille qui m'a bien évidemment soutenue durant cette période. Ma mère qui m'a rendu visite et avec qui j'ai pu profiter de la ville, mon père qui m'a toujours aidé dès que j'en ai eu le besoin. Mes grands-parents, dont je n'ai pas pu prendre des nouvelles régulièrement mais à qui j'ai souvent pensé. Mathys mon cher frère que je n'ai malheureusement pas pu voir autant que je l'aurais voulu durant cette période mais à qui j'ai beaucoup pensé, bien qu'il reste un emboucaneur de service. Mes sœurs Calypso et Pauline dont les visites ont toujours été bienvenues et enfin mon cousin Sacha avec qui j'ai pu passer de très bon week-ends à Lyon qui m'ont permis de m'aérer, la thèse m'a empêché de rester à niveau avec lui mais qu'il ne s'inquiète pas, je vais renfiler le kimono dès que possible.

Il y a aussi mes amis. Hien qui a été l'ami le plus présent pour moi à Paris. Cédric que je n'ai pas vu depuis un moment, mais comme on dit chez nous, c'est pas grave d'être à dache tant qu'on s'engatse pas ! Aurélien, les deux Alex et Guillaume, mes compatriotes du sud avec qui j'ai fait partie de la "Massilia connection" à Paris. Kevin, Renaud et Moncef, mes camarades de lycée qui font aujourd'hui toujours partie de ma vie après tout ce temps. Ophélie avec qui je partage ma malédiction de fatigue perpétuelle et de cycle du sommeil complètement à l'envers, nous nous en remettrons un jour... Astrid, min dansk lærer og min human vækkeur, tak for være dig.

Un dernier merci à tous mes camarades du judo et du jiu-jitsu, c'est toujours un plaisir de pouvoir essuyer le tatami avec votre figure et inversement. Noël et tous les boxeurs du club de Charenton, j'essaie de vous rendre visite dès que je le pourrai. Arthur qui a eu la gentillesse de m'inviter à venir à son club, pratiquer avec toi est toujours un plaisir.

Pour finir, je remercie tous ceux qui m'ont soutenu durant ma thèse sans même me connaître, les musiciens du collectif New Retro Wave, je pense pas être capable de rester concentré sans votre son. Toute l'équipe de Creative Assembly et CDPR, continuez ce que vous faites. Les gens de Air France qui m'ont évité de louper des vols à chaque fois que je suis arrivé à la bourre. Mes camarades du forum libre qui m'ont raconté leur vie quotidiennement. Enfin, les cuisiniers du Café des anges, merci d'avoir eu l'idée de mélanger des frites avec du cantal fondu et du lard.

Table des matières

1 Introduction **1**

1 Les maladies cardiovasculaires . . . 2

2 Les coronaropathies . . . 4

2.1 Physiopathologie . . . 4

2.2 Épidémiologie . . . 7

2.3 Génétique . . . 9

3 La thrombose veineuse . . . 11

3.1 Physiopathologie . . . 12

3.2 Épidémiologie . . . 14

3.3 Génétique . . . 16

4 Voies physiopathologiques partagées entre les coronaropathies et la thrombose veineuse . . . 18

5 Le diabète sucré . . . 19

5.1 Physiopathologie . . . 20

5.2 Génétique . . . 22

5.3 Épidémiologie . . . 23

5.4 Complications associées . . . 24

6 Le rein . . . 30

6.1 Anatomie du rein . . . 30

6.2 La fonction rénale . . . 34

6.3 Rôle dans les maladies cardiovasculaires . . . 38

7 Objectifs . . . 39

7.1 Les coronaropathies chez les diabétiques . . . 40

7.2 Relation entre la fonction rénale et la thrombose veineuse . . . 41

2 Matériel & méthodes **43**

1 L'approche épidémiologique . . . 43

1.1 Les grandes études d'association . . . 44

1.2 Les groupes d'étude . . . 46

1.3 Mesurer les variables biologiques 54
2 Acquisition des données génétiques 58
2.1 Les puces de génotypage 59
3 L'analyse 61
3.1 Le contrôle qualité 61
3.2 L'imputation 65
3.3 Association pan-génomique 66
3.4 Autres analyses statistiques 72
3.5 Exploiter les résultats 75

3 Résultats **77**
1 Étude pan-génomique sur les coronaropathies chez les diabétiques de type I . 77
1.1 Sélection des individus et des marqueurs 79
1.2 Association avec les données imputées et méta-analyse 81
1.3 Réplication des résultats 83
1.4 Analyse des associations déjà identifiées 83
1.5 Discussion autour des gènes candidats mis en évidence 84
2 Impact des marqueurs de la fonction rénale sur la thrombose veineuse 88
2.1 Définition du score polygénique 88
2.2 Analyse dans l'échantillon Français 90
2.3 Analyse dans l'échantillon Néerlandais 91
2.4 Confirmation de l'impact du score sur la fonction rénale 93
2.5 Impact du score sur les facteurs de l'hémostase 93
2.6 Discussion sur l'association entre le score génétique et le risque de MTEV et de sa pertinence vis-à-vis de la relation entre la fonction rénale et les MTEV 95

4 Discussion **99**
1 Intégration des résultats au regard du lien entre la fonction rénale et les maladies cardiovasculaires 99
2 Perspectives 100
3 Conclusion 101

Bibliographie **132**

Annexes **133**

Liste des tableaux

1 Introduction **1**

Tableau 1. Principaux facteurs de risque génétique de la thrombose veineuse . . . 17

2 Matériel & méthodes **43**

Tableau 2A. Caractéristiques des études Genesis/Genediab, Steno et UK-ROI . . 48

Tableau 2B. Caractéristiques des études DCCT/EDIC, WESDR, CACTI, EDC et GoKinD 50

Tableau 2C. Caractéristiques de l'étude MARTHA 52

Tableau 2D. Caractéristiques de l'étude EOVT 52

Tableau 2E. Caractéristiques de l'étude MEGA 53

Tableau 3. Liste des puces de génotypage 61

3 Résultats **77**

Tableau 4. Liste des *loci* candidats retenus lors de la phase exploratoire du GWAS 82

Tableau 5. Réplication des 17 signaux candidats issus la méta-analyse 83

Tableau 6. Réplication des signaux publiés par l'étude C4D-Cardiogram 84

Tableau 7. Effet individuel des SNPs sélectionnés pour le score polygénique sur le risque de thrombose veineuse 90

Tableau 8. Effet individuel des SNPs sélectionnés pour le score polygénique sur l'eGFR 93

Tableau 9. Effet du score polygénique sur les différents paramètres de l'hémostase 94

Table des figures

1 Introduction **1**

Figure 1. Anatomie des artères coronaires 5
Figure 2. Développement de l'athérosclérose 7
Figure 3. Régions génomiques associées aux coronaropathies 11
Figure 4. Cascade de la coagulation 14
Figure 5. Réponse métabolique à l'insuline 22
Figure 6. Principaux gènes associés au diabète de type I 22
Figure 7. Voies pathophysiologiques des complications du diabète 26
Figure 8. Anatomie d'un néphron 33
Figure 9. Synthèse des fonctions du néphron 37

2 Matériel & méthodes **43**

Figure 10. Principe de fonctionnement des puces de génotypage 60
Figure 11. Principe de fonctionnement de l'imputation 66

3 Résultats **77**

Figure 12. Protocole de l'étude sur le diabète est les coronaropathies 78
Figure 13. Sélection des individus 80
Figure 14. Manhattan plot et QQ-plot de la méta-analyse 81
Figure 15. *Plot* d'association régional centré sur *CDK18* 85
Figure 16. Protocole de l'étude de l'impact de la fonction rénale sur le risque de thrombose veineuse 89
Figure 17. Distribution et impact du score dans le groupe Français 91
Figure 18. Distribution et impact du score dans le groupe Néerlandais 92
Figure 19. Taux d'antithrombine moyen par quintile dans MARTHA 95

CHAPITRE 1

Introduction

Depuis le début des années 2000, avec le développement de technologies de biopuces à ADN plus avancées et plus accessibles, de plus en plus d'attention a été portée sur l'identification des facteurs de risque génétiques des maladies communes. Des données de génomique issues de nombreuses sources différentes ont permis depuis de considérablement faire avancer l'état des connaissances scientifiques sur de nombreuses pathologies en mettant en évidence de nouveaux mécanismes causaux, en identifiant des cibles thérapeutiques ou en identifiant des profils génétiques à risque. Malgré toutes ces avancées, les connaissances sur la composante génétique des maladies complexes restent incomplètes et restent donc un axe de recherche important. Le travail présenté dans ce manuscrit s'inscrit dans ce thème et tente d'apporter des éléments de réponse concernant la caractérisation des facteurs de risque génétiques des maladies cardiovasculaires et l'effet des marqueurs génétiques influençant la fonction rénale sur le risque de maladie cardiovasculaire. La démarche présentée ici est cependant pertinente pour d'autres maladies complexes.

Le manuscrit présente mon travail de thèse au cours duquel j'ai été amené à étudier la génomique des maladies cardiovasculaire mais aussi à me documenter sur la physiopathologie et l'épidémiologie de ces maladies mais aussi la physiologie et la biologie humaine. Le manuscrit est divisé en quatre parties qui sont les suivantes :

La première partie est consacrée à la présentation des pathologies traitées dans mes travaux, c'est à dire les maladies cardiovasculaires, en particulier les coronaropathies et la thrombose veineuse, puis le diabète de type I et ce qui relie ces pathologies. Dans cette partie je présente également le rein, son anatomie et sa fonction ainsi que le lien avec les maladies mentionnées plus haut. Enfin je présente les objectifs initiaux pour lesquels ces travaux ont été menés et l'intérêt de l'approche que j'ai employée. Les détails apportés dans cette partie sur la physiologie humaine peuvent sembler superflus mais je pense qu'il est important de les connaitre afin de comprendre la relation complexe qui relie les maladies cardiovasculaires, le diabète de type I et la fonction rénale.

La deuxième partie présente la démarche méthodologique ainsi que son intérêt, en détaillant les principes généraux sur lesquels se base une étude d'épidémiologie génétique. Je parlerai également dans cette section des groupes d'étude à partir desquels ces travaux ont été menés en parlant d'abord des sujets et de leurs caractéristiques puis des paramètres physiologiques mesurés et des méthodes de mesure de ces paramètres. Enfin je présenterai les outils employés lors de ces travaux, allant des puces à ADN permettant de mesurer les variations du génomes aux outils informatiques permettant d'effectuer les analyses statistiques et d'exploiter les résultats. Les méthodes mathématiques sur lesquelles reposent le fonctionnement de ces outils seront également détaillées.

La troisième partie sera consacrée à la présentation des résultats des travaux. Le but est de présenter tous les résultats obtenus dans les deux projets que j'ai menés lors de cette thèse. Le premier projet porte sur l'identification de nouveaux facteurs de risque génétiques des coronaropathies chez les patients diabétiques de type I avec ou sans complications rénales. Dans ce projet nous examinons également l'effet sur cette population particulière de facteurs de risque génétiques connus dans la population générale. Le second projet porte sur l'étude du lien entre la fonction rénale et le risque de thrombose veineuse en utilisant des marqueurs génétiques de la fonction rénale. Nous tentons également d'identifier des paramètres physiologiques influencés par ces marqueurs génétiques qui pourraient relier la fonction rénale au risque de thrombose veineuse.

Pour terminer, la quatrième et dernière partie porte sur l'interprétation des résultats des projets présentés plus haut. Je tenterai d'intégrer les données obtenues grâce à mes travaux dans un cadre plus global de la relation entre les maladies cardiovasculaires et la fonction rénale. Je discuterai également des limites de ces travaux et des questions auxquelles ils n'ont pas pu apporter de réponse mais aussi des hypothèses que nous pouvons formuler à partir des résultats afin d'explorer les nouvelles pistes de recherche qui ont été ouvertes. Enfin, je conclurai sur les éléments apportés par mes travaux dans la cartographie complexe des relations entre la fonction rénale et les maladies cardiovasculaires.

1 Les maladies cardiovasculaires

Les maladies cardiovasculaires regroupent les maladies affectant le cœur et les vaisseaux, qui incluent de manière non exhaustive les coronaropathies, les accidents vasculaires-cérébraux, les cardiomyopathies, les vasculopathies cardiaques, les troubles du rythme cardiaque, les cardiopathies congénitales, les myocardites et la thrombose veineuse. Parmi toutes ces maladies, on peut distinguer les maladies cardiovasculaires causées par l'athérosclérose qui reposent sur un mécanisme commun, la formation de la plaque d'athérome.[1]

Les mécanismes responsables du développement des maladies cardiovasculaires sont extrêmement variables et le seul point commun entre certaines de ces maladies reste le tissu touché. Cependant on peut identifier un certain nombre de fonctions physiologiques impliquées dans plusieurs maladies cardiovasculaire. C'est notamment le cas des radicaux libres qui fragilisent la paroi vasculaire[2], le rôle de la réponse inflammatoire et des lipoprotéines de basse densité dans la formation de la plaque d'athérome[2, 3] ou encore le rôle des facteurs de l'hémostase dans la formation du thrombus[4]. Tous ces éléments sont impliqués dans plusieurs maladies cardiovasculaires, ce qui rend pertinente l'idée de les regrouper dans une seule catégorie.

D'après l'organisation mondiale de la santé, les maladies cardiovasculaires sont la principale cause de mortalité dans le monde, avec un taux de mortalité prématurée allant de 4% dans les pays développés à 42% dans les pays pauvres ou en voie de développement. Près de la moitié de ces décès sont dus à des maladies cardiovasculaires avec athérosclérose.[1] Malgré le fait que les maladies cardiovasculaires, notamment celles reposant sur l'athérosclérose, soient responsables de plusieurs millions de décès chaque année, il a été estimé que 90% d'entre elles seraient évitables en agissant sur les facteurs de risque que ce soit en modifiant les habitudes de vie ou par des traitements médicaux.[5]

Les maladies cardiovasculaires sont des maladies complexes impliquant une multitude de facteurs de risques. Ces facteurs de risques peuvent être regroupés en quatre catégories[1], les facteurs comportementaux (tabagisme, inactivité physique, régime alimentaire trop riche, consommation excessive d'alcool), métaboliques (hypertension, hyperglycémie, hyperlipidémie, obésité), génétiques et environnementaux (sexe, âge, statut socio-économique, stress, dépression, pollution atmosphérique). Cependant ces facteurs interagissent entre eux et influent les uns sur les autres. Certains facteurs de risques sont immuables mais il est cependant possible de prévenir l'apparition de maladies cardiovasculaires en contrôlant d'autres facteurs de risques. L'essentiel de la prévention résiderait donc dans une alimentation saine, la pratique régulière de l'exercice physique et en évitant la consommation de tabac. D'après l'organisation mondiale de la santé, 13% des décès dus aux maladies cardiovasculaires peuvent être attribués à l'hypertension, 9% au tabagisme, 6% à une glycémie trop élevée, 6% au manque d'activité physique et 5% au surpoids et à l'obésité.[1] Il est cependant important de noter que bien que la modification du comportement joue une part importante dans la réduction du risque de développer une maladie cardiovasculaire ce n'est tout de même pas suffisant pour l'éliminer[5]. Il est donc également utile d'étudier les facteurs de risques moins bien compris. C'est le cas de la génétique, car bien qu'un important faisceau de preuves s'accumule sur le rôle des gènes dans la prédisposition à ces maladies[6, 7], le rôle des gènes identifiés, leurs interactions entre eux et avec leur environnement restent à caractériser. Une meilleure compréhension des mécanismes régulés par ces gènes pourrait être un élément clé de la compréhension des

mécanismes derrière le développement de ces maladies.

2 Les coronaropathies

Les coronaropathies regroupent l'angine de poitrine, l'angor instable, l'infarctus du myocarde et l'arrêt cardio-circulatoire. D'après l'institut national de la santé des États-Unis, des douleurs de poitrine pouvant s'étendre aux bras, aux épaules, au dos, au cou et dans la mâchoire sont une manifestation commune de ces maladies. Ces douleurs peuvent apparaître durant une activité physique ou en période de stress.[8] Comme leur nom l'indique, ces maladies affectent les artères coronaires, les artères approvisionnant le muscle cardiaque en sang oxygéné. Ces artères partent de la base de l'aorte, juste à la sortie du ventricule gauche. On en dénombre deux, l'artère coronaire gauche et l'artère coronaire droite, chacune ayant ensuite des embranchements, comme exposé dans la figure 1. Ces artères représentent la seule source d'approvisionnement pour le muscle cardiaque et sont donc essentielles à sa survie.

2.1 Physiopathologie

Les coronaropathies apparaissent le plus souvent lorsque les artères coronaires sont partiellement ou totalement obstruées, principalement en raison de l'athérosclérose. La plaque d'athérome se forme lorsque la paroi endothéliale est exposée à des lipoprotéines de basse densité oxydées. Les lipoprotéines de basse densité (LDL) sont des complexes chargés du transport des lipides. Ils sont composés d'alipoprotéine B-100 et de protéines annexes ainsi que d'acide linoléique, de cholestérol, de triglycérides divers et de vitamines hydrophobes, le tout entouré d'une couche de phospholipides[9]. Ces LDLs sont absorbés par les cellules nécessitant des acides gras et du cholestérol pour leur métabolisme. C'est lorsque les LDLs sont exposés à des radicaux libres qu'ils peuvent s'oxyder.[10] Lorsque l'endothélium vasculaire est exposé aux LDLs oxydés, il devient perméable aux monocytes et aux lymphocytes, des cellules du système immunitaire. Les monocytes peuvent alors adhérer à l'endothélium et migrer dans l'espace sous-endothélial pour se différencier en macrophages, ces macrophages vont endocyter le LDL oxydé sans pouvoir le recycler et vont évoluer en cellules spumeuses, nommées en raison des nombreuses vésicules riches en lipides qu'elles contiennent.[11] Les cellules spumeuses finissent par mourir et laissent derrière elles un noyau nécrotique contenant des résidus lipidiques, la formation de ce noyau entraine le recrutement d'autres monocytes via la sécrétion de cytokines par les cellules endothéliales voisines.

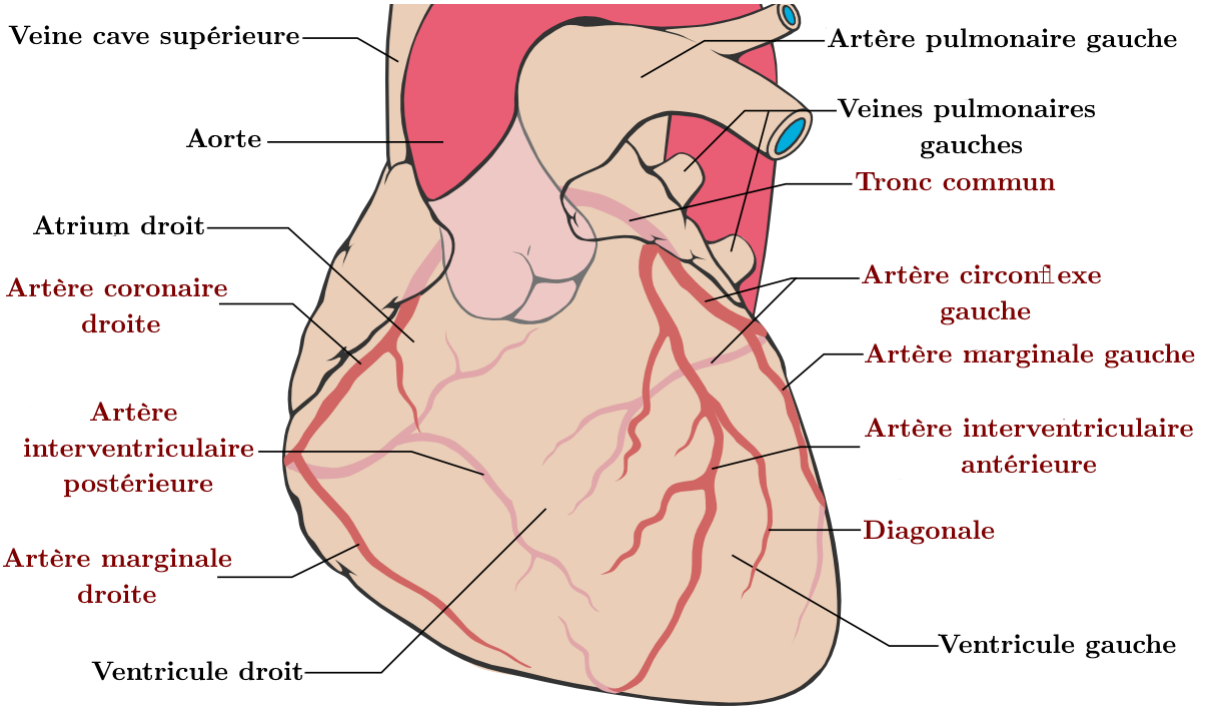


Figure 1 Ici on peut voir les deux artères coronaires principales en rouge, l'artère coronaire droite et ses embranchements, l'artère marginale droite et l'artère interventriculaire postérieure qui alimentent respectivement les faces antérieure et postérieure du ventricule droit. Le tronc commun se divise en deux branches. L'artère circonflexe gauche qui se déploie dans le sillon atrio-ventriculaire et donne l'artère marginale gauche qui alimente la face postérieure du ventricule gauche. L'artère interventriculaire antérieure qui se divise en plusieurs diagonales alimentant le ventricule gauche.

Illustration adaptée du travail de Patrick J. Lynch

Ces monocytes vont à leur tour se différencier et continuer le processus, permettant ainsi à la plaque de grossir. Les plaquettes vont également être recrutées sur le site et le noyau nécrotique va se retrouver entouré par une chape fibreuse qui contribue également à l'évolution de la plaque. Les cellules musculaires lisses en contact avec la plaque accumulent du calcium, ce qui crée des dépôts de calcium extra-cellulaire lorsqu'elles meurent. Ces dépôts cristallisent et entrainent la solidification de l'artère autour de la plaque, la rendant moins tolérante à la déformation entrainée par l'évolution de la plaque.[12] Lorsque la plaque est suffisamment volumineuse, elle va alors obstruer partiellement l'afflux sanguin et entrainer une ischémie, c'est à dire une réduction de l'approvisionnement en oxygène des cellules du muscle cardiaque. Les premiers symptômes peuvent apparaitre sous la forme d'une angine de poitrine stable, qui se manifeste lorsque l'alimentation en oxygène des cellules du muscle cardiaque devient insuffisante par rapport à leurs besoins, typiquement lors d'activités physiques ou durant des périodes de stress.[13] Dans certains cas (plaque lipidique et très inflammatoire), il peut se produire une rupture de la plaque ou une érosion de la chape fibreuse, les éléments fortement pro-thrombotiques du sous-endothélium tels que le collagène et le facteur tissulaire sont alors

exposés à la lumière du vaisseau. Au contact de ces éléments les plaquettes vont alors s'activer puis activer la cascade de la coagulation[14], ce qui va conduire à la formation d'un thrombus ce qui peut entrainer l'occlusion de l'artère.[15] Le développement de l'athérosclérose, synthétisé dans la figure 2, est très long et commence souvent durant l'enfance[16], l'évolution des plaques est souvent asymptomatique mais la rupture de la chape et la formation du thrombus peuvent aboutir à des conséquences très sérieuses.

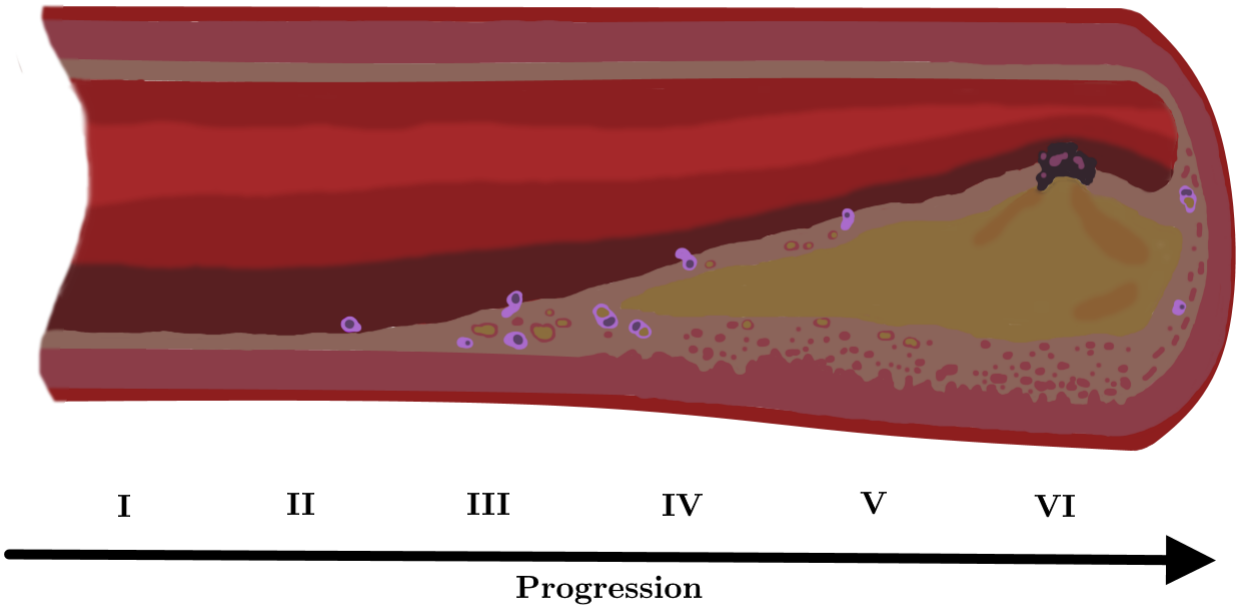


Figure 2 Développement de la plaque d'athérome. Le stade I correspond à un endothélium sain, imperméable aux leucocytes. Le stade II survient durant la première décennie, durant ce stade l'exposition de l'endothélium aux LDLs oxydés permet l'adhésion de l'infiltration des leucocytes. Durant les deux décennies suivantes la plaque d'athérome progresse vers les stades III et IV. Durant le stade III les leucocytes infiltrés dans l'intima vont accumuler des lipides et former des cellules spumeuses. Durant le stade IV les cellules spumeuses meurent et forment le noyau de lipides formant le cœur de la plaque d'athérome. Durant le stade VI les cellules musculaires lisses entourant la plaque meurent et libèrent du calcium, une chape fibreuse se développe autour de la plaque. Au stade V la chape fibreuse s'érode et le contenu de la plaque entre en contact avec la lumière endothéliale, la cascade de la coagulation est activée par la présence de collagène, de facteur tissulaire et d'autres molécules pro-thrombotiques. C'est lorsque le stade VI est atteint qu'on peut observer des syndromes coronaires aigus.

La rupture de la plaque peut entrainer une occlusion complète du vaisseau qui se manifeste sous la forme d'un syndrome coronaire aigu, un événement aigu dont les symptômes sont similaires à ceux de l'angine stable mais qui peut se produire au repos.[17] L'oxygène étant un élément vital pour les cellules, cette privation peut entrainer une nécrose rapide des tissus. On peut constater la nécrose des tissus entre 40 et 60 minutes après l'occlusion du vaisseau.[18] C'est lorsque les cellules du muscle cardiaque meurent que survient l'infarctus du myocarde, dont les symptômes peuvent se manifester sur une période de 3 à 4 heures et qui est un événement potentiellement fatal. Le tissu cardiaque mort ne pouvant pas régénérer, il va alors cicatriser et faire de la fibrose, ce qui est susceptible de perturber le fonctionnement du cœur et entrainer des complications telles qu'une péricardite ou des troubles du rythme cardiaque responsables d'un arrêt cardiaque ou d'une insuffisance cardiaque, en fonction de la taille de la zone touchée et de sa position.[19]

Le diagnostic des coronaropathies peut se faire à l'aide d'examens fonctionnels dont le but est de démontrer la présence d'une ischémie myocardique tels que les électrocardiogrammes, les électrocardiogrammes d'effort, la tomoscintigraphie myocardique de perfusion d'effort, l'échocardiographie de stress et l'IRM de stress. Des examens morphologiques, qui permettent d'examiner directement la morphologie des artères coronaires sont également possibles, tels que la coronarographie ou le scanner coronaire (*CT-scan*).[20, 15] Les événements aigus nécessitent des traitements immédiats, de la nitroglycérine peut être prescrite afin de diminuer la pression artérielle et soulager les douleurs de poitrine.[21] Il est possible de réduire le thrombus avec une angioplastie coronaire. On utilise alors une sonde cathéter gonflable afin de réduire l'occlusion. Enfin la troisième méthode est une méthode chirurgicale, un pontage aorto-coronarien. Il s'agit dans ce cas de contourner l'occlusion en implantant une autre artère en aval[22]. Les deux dernières méthodes permettent d'allonger l'espérance de vie après événement et réduisent le risque d'un arrêt cardiaque.[23]

2.2 Épidémiologie

Malgré le fait que les coronaropathies ont globalement reculé dans les pays développés depuis la fin du 20ème siècle,[24, 25] elles restent les maladies cardiovasculaires les plus communément rencontrées.[26] En 2015, les coronaropathies ont affecté 110 millions de personnes causant 8,9 millions de décès, ce qui en fait la cause de décès la plus commune dans le monde.[26, 27] Les coronaropathies peuvent survenir à tout âge, cependant le risque augmente d'un facteur 2,3-2,7 par tranche de dix ans.[28] Toutes ces donnés suggèrent que la mortalité due aux coronaropathies recule mais reste un phénomène majeur qui mérite amplement le nom d'épidémie.

Afin d'établir des stratégies de prévention et de guider la recherche thérapeutique, il est important d'identifier les causes et les facteurs de risque associés aux maladies cardiovasculaires, ce que la recherche en épidémiologie s'est attelée à faire jusqu'à présent. Comme pour la plupart des autres maladies, on distingue des facteurs de risque communs, dont certains sont immuables et d'autres qui peuvent être modifiés. L'âge reste le plus important, puis vient le sexe avec un risque plus élevé pour les hommes.[28] L'étude INTERHEART a permis de mettre évidence que l'hypercholestérolémie, le tabagisme, le diabète, l'hypertension artérielle, l'obésité abdominale et les facteurs psychosociaux sont des facteurs de risque importants tandis qu'une alimentation riche en fruits et légumes et la pratique de l'exercice physique régulière sont des facteurs atténuants.[29] Les facteurs de risques les plus importants[30, 31, 1] sont le tabagisme avec une augmentation du risque estimée à 3,3 fois pour la femme et 1,9 fois pour l'homme,[32] l'hypertension avec une augmentation du risque estimée à un facteur 2 pour chaque tranche de 20mmHg de pression en systole et 10mmHg en diastole,[33] l'obésité avec une augmentation de 40% du risque de maladie vasculaire pour chaque tranche de 5 d'indice de masse corporelle supplémentaire,[34] le diabète qui double le risque de développer une maladie coronarienne[35, 36] et le manque d'activité physique qui serait à l'origine de 6% des cas de maladie coronarienne.[37] Enfin la dyslipidémie joue également un rôle important. le cas des lipides étant cependant plus complexe car bien que les taux de LDLs et de triglycérides augmentent de manière importante le risque de maladie coronarienne,[38, 39] il semblerait que dans le cas des LDLs la taille des particules[40] ainsi que la durée de l'exposition[41, 42] jouent un rôle important. D'après l'étude de Framingham le HDL-cholestérol joue un rôle protecteur et pourrait donc mitiger l'effet des autres particules lipidiques.[43] On note également l'existence de facteurs de risque plus modestes, comme un environnement de travail stressant qui serait responsable de 3% des cas de maladie coronarienne[44, 45] et la dépression avec 2,9% des cas de maladie coronarienne qui lui seraient attribuables.[46] Les taux de certains facteurs de l'hémostase comme le fibrinogène, le facteur VII ou le facteur XII,[47] le d-dimère, le complexe thrombine-antithrombine et le complexe plasmine-antiplasmine sont également associés à un taux de mortalité plus élevé dans les coronaropathies,[48] ce qui pourrait être en lien avec leur rôle dans la formation du thrombus.

Étant donné l'impact des facteurs de risque comportementaux, la plus grande part du travail de prévention des coronaropathies passe par la modification des comportements à risque. Ainsi, les premiers conseils de prévention sont de pratiquer une activité physique régulière, de maintenir un régime alimentaire sain et équilibré et d'éviter la consommation de tabac.[5] La pratique de l'exercice réduit notamment le risque de développer la maladie, l'exercice physique très fréquent peut réduire le risque de près de 25%.[49] Les bienfaits d'une alimentation équilibrée ont maintes fois été démontrés et il est conseillé de privilégier les fruits et légumes tout en évitant les acides gras trans.[50, 51] Une alimentation peu calorique a également des ef-

fets positifs sur des pathologies associées comme le diabète de type 2.[52] L'arrêt de la consommation du tabac permet de réduire le risque de maladie coronarienne, ce qui montre que même après plusieurs années de consommation il n'est pas trop tard pour contrôler le risque.[53] Pour les individus ayant un profil à risque des traitements pharmacologiques peuvent être envisagés en guise de méthode de prévention. Les statines, des médicaments hypolipémiants, peuvent être utilisées chez les personnes faisant de l'hypercholestérolémie,[54] l'aspirine peut être appropriée pour les profils thrombophiles,[55] et les béta-bloquants peuvent aider les personnes souffrant d'hypertension.[56] Lorsque la prévention ne permet pas d'empêcher l'apparition des coronaropathies, il existe tout de même des moyens de les détecter et de les traiter.

2.3 Génétique

Un autre facteur de risque est bien évidemment la génétique, qui mérite d'être mentionnée à part car l'effet de la génétique n'est pas entièrement caractérisé et les mécanismes physiologiques par lesquels le risque génétique se manifeste ne sont pas tous compris. L'héritabilité des maladies coronarienne a été estimée entre 40% et 60% et le risque relatif pour des jumeaux monozygotes augmente d'un facteur 8,1 par comparaison avec la population générale.[57, 58] Plusieurs types de mutations rares sont associées au risque de coronaropathie, ces mutations ont souvent une forte pénétrance. C'est le cas par exemple des différentes formes d'hypercholestérolémie familiale, qui touchent en général des gènes impliqués dans le métabolisme du cholestérol tels que *APOB*, *LDLR* ou *PCSK9*.[59] Les formes homozygotes des hypercholestérolémies familiales sont généralement associées à un taux élevé de LDL-cholestérol et un risque très élevé d'athérosclérose. Mis à part ces mutations rares, une centaine *loci* ont été associés au risque de coronaropathie.[7, 59, 60, 61] Un résumé de ces *loci* est présenté dans la figure 3. De nombreux *loci* identifés semblent avoir un effet individuel modeste sur le risque de développer une maladie coronarienne mais leurs interactions entre eux et avec l'environnement ainsi que les mécanismes qu'ils affectent restent à déterminer. On sait cependant que certains d'entre eux affectent des mécanismes physiologiques intéressants. Par exemple, on retrouve des mutations de *GUCY1A3* et *CCT7* associées à l'infarctus du myocarde.[62] Ces gènes sont impliqués dans la réponse des plaquettes à l'oxyde nitrique (NO), une molécule de signalisation favorisant la vasodilatation et réduisant la capacité d'activation des plaquettes et la capacité d'adhésion à l'endothélium des leucocytes.[63] De la même manière, des variations du gène *NOS3* ont été associées aux coronaropathies.[64] *NOS3* code pour eNOS, l'enzyme chargée de la production du NO dans les cellules endothéliales. On retrouve des associations pour des *loci* tels que *APOA5-A4-C3-A1* ou *LPA* qui composent les complexes lipoprotéiques (LDL et HDL), *PCSK9* qui est impliqué dans l'homéostasie du LDL, *LDLR* le récepteur du LDL, *COL4A1-A2* qui codent pour des éléments du collagène de type 4, *IL6R* le récepteur à l'interleukine 6, une molécule pro-inflammatoire notamment sécrétée par les

cellules spumeuses, ou encore *SH2B3* et *CXCL12* qui sont impliqués dans la chimiotaxie et l'activation des lymphocytes T.[65, 66, 67] Le gène *PROCR* est lui aussi potentiellement impliqué dans le risque de développer une maladie coronarienne,[68] ce dernier code pour un récepteur à la protéine C, une protéine chargée de réguler l'activité de la thrombine.

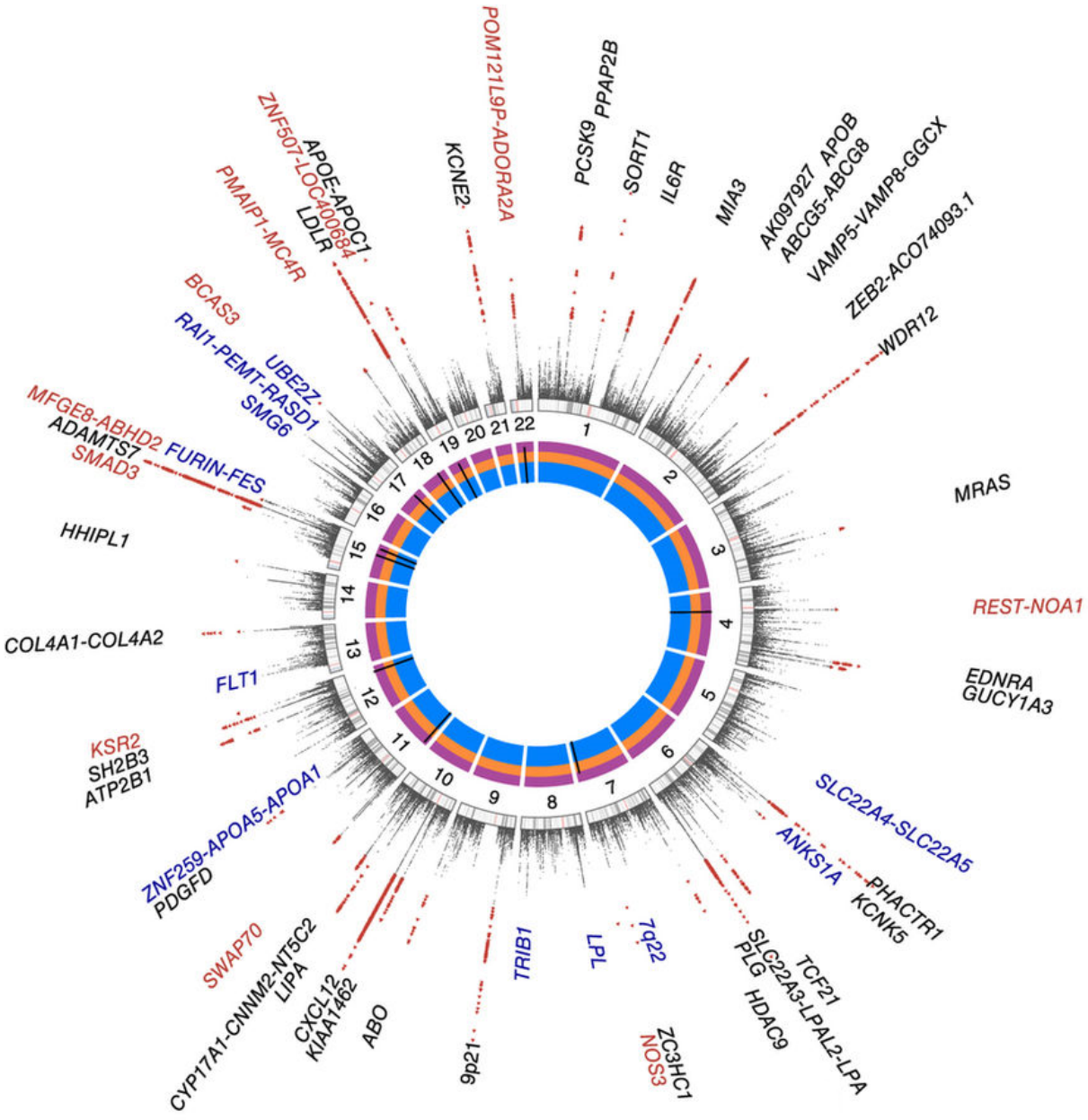


Figure 3 Ce *Manhattan-plot* circulaire offre un bon aperçu des *loci* associés aux coronaropathies dans la population générale. On remarque la présence de gènes du métabolisme des lipides comme *PCSK9*, *LPA* ou encore *LDLR* mais aussi des gènes de la réponse immunitaire comme *IL6R*.

Illustration issue de l'article de Nikpay et al, Nat Genet 2015[7]

On retrouve également des associations avec *ANGPTL4* et *LPL*,[69] des gènes impliqués dans la régulation du taux de triglycérides présents dans la circulation. Les effets de ces variations sont modestes mais il existe un grand nombre de *loci* associés aux coronaropathies. En revanche l'effet de toutes ces variations n'est pas nécessairement additif. L'effet de l'accumulation des variations ne correspond pas à la somme des effets individuels,[70] ce phénomène suggère des interactions plus complexes. En comparaison aux facteurs de risque comportementaux, il peut sembler difficile d'intervenir sur les facteurs génétiques. Leur étude a cependant un intérêt non négligeable car l'identification de gènes associés à la maladie permet de découvrir les mécanismes impliqués, et permet parfois de découvrir des cibles thérapeutiques potentielles, ce qui est par exemple le cas de *PCSK9*. Ce gène a été identifié en 2003 et associé à l'hypercholestérolémie,[71] d'autres mutations ont ensuite été découvertes,[72] ajoutant plus d'importance au gène. La protéine codée par *PCSK9* est impliquée dans la dégradation des récepteurs au LDL, limitant ainsi le nombre de récepteur à la surface des cellules et la réduction de taux de LDLs présents dans la circulation.[73] Il a été montré qu'une réduction de l'activité de *PCSK9* est associée à une réduction du taux de LDLs[74], la protéine a donc naturellement été envisagée comme une cible thérapeutique.[75] Les études de génétique sont une bonne solution lorsqu'il faut clarifier les liens de cause à effet entre une variable physiologique et la maladie, ce qui n'est pas toujours bien établi dans les études d'épidémiologie classique. Les travaux effectués sur les variations de *LDLR* affectant le taux de LDLs dans le sang et leur impact sur le risque de maladie coronarienne en sont un bon exemple.[76] En effet, ces travaux font intervenir des méthodes de randomisation Mendelienne, dans lesquelles sont utilisées des variations génétiques affectant un facteur d'exposition (ici le taux de LDL-cholestérol) afin de les utiliser comme marqueur de l'effet de ce facteur d'exposition sur la pathologie d'intérêt.[77] Ces méthodes sont puissantes car elles permettent de s'affranchir d'un certain nombre de facteurs de confusion et permettent d'établir la direction du lien causal.

3 La thrombose veineuse

La thrombose veineuse est une maladie vasculaire causée par la formation d'un thrombus dans une veine. Une forme commune de thrombose est la thrombose veineuse profonde qui touche une veine profonde généralement dans les jambes, les veines profondes sont responsable de la majeure partie du transport du sang vers le cœur.[78] Le thrombus est susceptible d'obstruer complètement la circulation dans la veine touchée conduisant à une rougeur et un œdème au niveau du membre concerné. Il est possible que le thrombus se lyse et qu'un des fragments soit emporté vers le cœur puis les poumons par la circulation. De là il risque d'obstruer une artère pulmonaire provoquant alors une embolie pulmonaire. L'embolie pulmonaire est une complication sérieuse de la thrombose veineuse.[79] La thrombose veineuse est une maladie complexe impliquant plusieurs facteurs de risque immuables tels que l'âge

et environnementaux tels que les traumatismes ou les immobilisations complètes ou d'un membre sur une relativement longue durée.[80] La génétique est également impliquée et plusieurs mutations connues expliquent une part des cas de thrombose veineuse. C'est le cas des déficits en antithrombine III, en protéine C et en protéine S ainsi que des mutations FV Leiden et 20210A de la prothrombine.[81] Les mécanismes pathophysiologiques de la thrombose veineuse ont été catalogués par le docteur Prusse Rudolf Virchow en 1856, la communauté scientifique les a plus tard regroupés sous le terme de triade de Virchow.[82] Cette triade est composée de trois éléments :

Hypercoagulabilité Changement de la composition du sang favorisant l'activation des facteurs de la coagulation, ces changements peuvent être dus à des mutations génétiques, un état pro-inflammatoire ou à une hyperviscosité sanguine.

Changement hémodynamique Un ralentissement du flux sanguin qui peut être causé par une immobilisation prolongée du membre concerné.

Irritation de la paroi endothélialie Des dégâts subits par la paroi endothéliale, qui peuvent être causés par une blessure ou un traumatisme, une inflammation ou du stress oxydatif.

La présence de ces trois éléments est un élément clé dans le développement de la thrombose veineuse et les facteurs de risque identifiés peuvent généralement être reliés à un de ces composants.

3.1 Physiopathologie

Pour comprendre les mécanismes pathologiques aboutissant à la formation du thrombus il faut connaitre et comprendre le fonctionnement du processus de coagulation. On distingue deux phases dans la coagulation. La première est l'activation des plaquettes et la seconde est la cascade de coagulation qui est présentée en détail dans la figure 4. Le recrutement des plaquettes a lieu immédiatement après la lésion de la paroi endothéliale. Elle est due à l'exposition du collagène à la lumière vasculaire.[83] Les plaquettes disposent de plusieurs récepteurs permettant de se lier au collagène, notamment le complexe glycoprotéique Ia/IIa (GPIa/IIa). La présence de facteur de von Willebrand (vWF) produit par les cellules endothéliales et les plaquettes jouent un rôle dans l'établissement des liaisons entre les fibres de collagène et les glycoprotéines Ib/IX/V des plaquettes.[84] L'activation des plaquettes accumulées au site de la blessure peut se faire de plusieurs manières : par le biais de la liaison entre le collagène et vWF avec la glycoprotéine VI, qui active une cascade métabolique impliquant la PKC ou par la thrombine qui peut activer les plaquettes en se liant au récepteur protéase-activé 4 (PAR4) à la surface des plaquettes. Ce mécanisme fait également appel au facteur tissulaire (TF) et

au facteur VII.[85] Une fois les plaquettes activées, elles libèrent des granules contenant notamment du facteur d'activation des plaquettes (PAF) et du thromboxane A2 (TBXA2). Le PAF va favoriser l'activation d'autres plaquettes tandis que le TBXA2 va modifier la glycoprotéine IIb/IIIa et augmenter son affinité pour le fibrinogène. Ceci va favoriser l'agrégation des plaquettes sur le lieu de la lésion.[84] La cascade de coagulation est concomitante au recrutement et à l'activation des plaquettes et possède deux voies d'amorce différentes, la voie extrinsèque et la voie intrinsèque. Ces deux voies convergent vers une cascade commune qui mène à la génération de fibrine.[86]

La voie extrinsèque est dépendante du TF, déjà mentionné plus haut, qui est exposé à la circulation par la blessure et peut donc rencontrer le FVII avec lequel il forme un complexe TF-FVIIa. Le TF-FVIIa va alors activer le facteur X (FX) qui en s'unissant au facteur V activé (FVa) va catalyser la transformation de la prothrombine en thrombine. La thrombine participe avec le facteur XIII (FXIII) à la transformation du fibrinogène en fibrine en plus d'amplifier la cascade par un mécanisme de rétro-contrôle et d'activer les plaquettes.[87] La voie extrinsèque est régulée par l'inhibiteur de la voie du facteur tissulaire (TFPI) qui inhibe l'activation de FX par le complexe TF-FVIIa. La voie intrinsèque est déclenchée par l'exposition du collagène au kininogène, ce qui permet l'activation du facteur XII, qui permet à son tour d'activer le facteur XI (FXI) puis le facteur IX (FIX), le FIXa se lie à FVIIIa et active FX. La voie intrinsèque rejoint alors la voie extrinsèque dans la génération de thrombine puis de fibrinogène.[86] La fibrine sert d'ancrage aux plaquettes et permet la stabilisation du thrombus.[88] La cascade de coagulation est régulée par plusieurs facteurs en plus du TFPI. Il y a la protéine C, qui désactive FVa et empêche la formation du complexe FVa-FXa. L'antithrombine qui dégrade le FXa et la thrombine, la plasmine qui dégrade la fibrine et la prostacycline qui bloque la libération des granules par les plaquettes activées.[89] C'est lorsque plusieurs éléments se réunissent pour favoriser l'activation du processus de coagulation que le risque de thrombose veineuse émerge. Plusieurs éléments sont susceptibles de favoriser la formation de thrombus. D'abord, l'endothélium sain sécrète normalement des facteurs anti-thrombotiques comme le NO, la prostacycline et l'ADPase endothéliale qui inhibent l'activation des plaquettes.[90] Lorsqu'il y a une dysfonction des cellules endothéliales, cette barrière de protection est susceptible de disparaitre. La dysfonction des cellules endothéliales peut avoir plusieurs origines mais certaines pathologies peuvent en être la cause, notamment le diabète.[91] Un environnement hypoxique est également un élément important, car les cellules endothéliales vont produire des molécules comme HIF-1 et EGR-1 qui favorisent l'activation des monocytes qui vont alors adhérer à l'endothélium et produire du TF.[92, 93] C'est notamment pour cette raison que les valves veineuses sont souvent des sites de formation de thrombi, la zone abritée de la valve est propice à l'hypoxie.[94] Une dysfonction d'un des facteurs de la coagulation peut également provoquer un état pro-thrombotique. C'est le

cas de mutations comme FV Leiden qui entraine une résistance à la protéine C activée[67] ou le déficit en protéine S qui est un cofacteur important de la protéine C activée.[95]

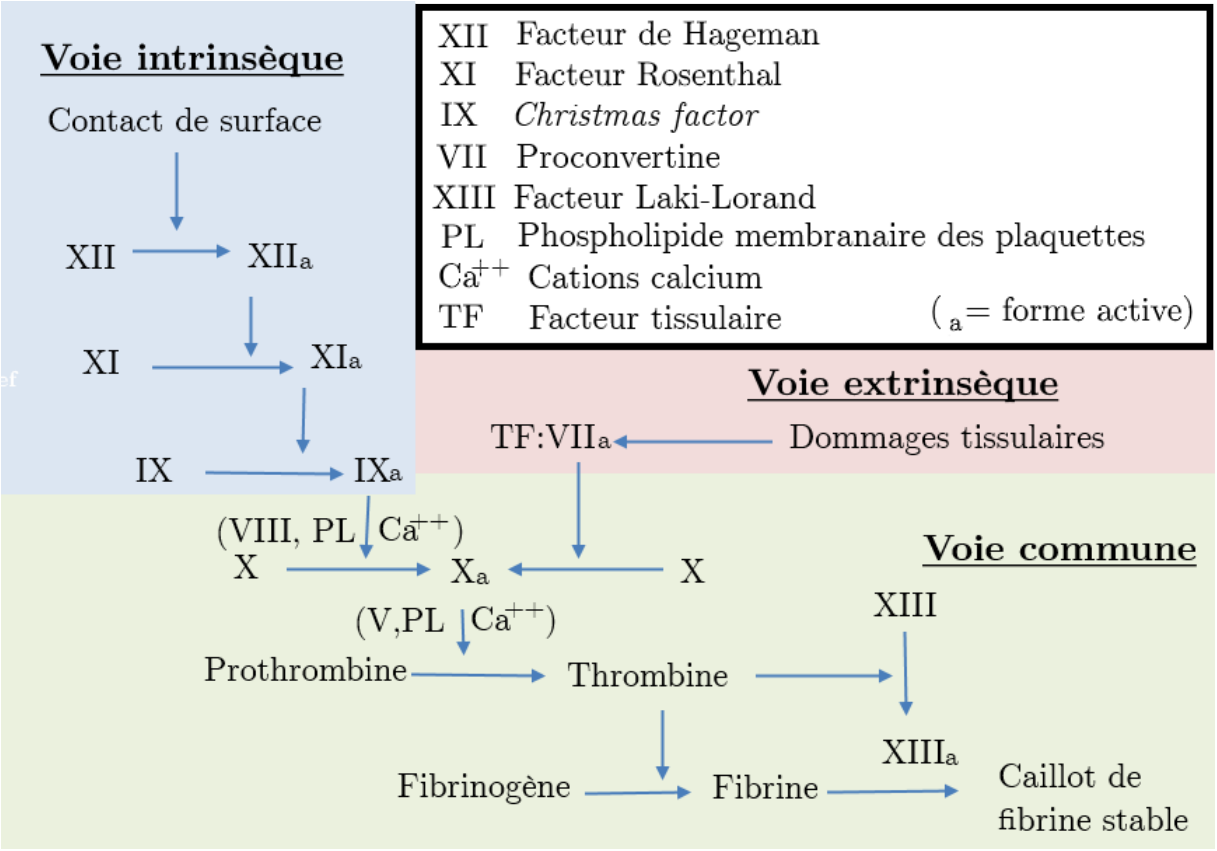


Figure 4 Représentation de la cascade de la coagulation et facteurs de coagulation impliqués. On remarque l'implication du facteur anti-hémophilique A (facteur VIII) dans l'activation du facteur X par le facteur IXa dans la voie intrinsèque. La proaccélérine (facteur V) intervient dans l'activation de la thrombine.

Illustration adaptée du travail du Dr Graham Beards

3.2 Épidémiologie

La thrombose veineuse touche environ 200 personnes par 100 000 par an en Europe, avec un taux de récidive relativement élevé puisque près de 30% des événements sont suivis d'un nouvel événement dans les 10 ans qui suivent.[96] C'est une maladie liée à l'âge, avec un âge moyen à la déclaration de la maladie proche de 60 ans.[97] La thrombose veineuse profonde peut être à l'origine d'une embolie pulmonaire, 10% des embolies pulmonaires sont fatales dans l'heure suivant leur occurrence. Le taux de mortalité sous les 3 mois de l'embolie pulmonaire étant de 15%.[98] Bien que la thrombose veineuse ne soit pas aussi fréquente que les coronaropathies, elle n'en reste pas moins un fardeau majeur dans le monde et représente une

dépense santé entre 1,5 millards et 13,2 milliards d'euros en Europe.[99] Le premier facteur de risque de thrombose veineuse est l'âge, avec un risque passant de 1 sur 10 000 avant 40 ans pour rapidement augmenter et passer à 5 à 6 pour 1 000 à partir de 80 ans. L'âge est également associé à un taux d'embolie pulmonaire et une mortalité plus élevés.[97] D'autres facteurs de risque modifiables ou temporaires connus sont l'obésité. Chez l'homme et la femme un IMC supérieur à 30 est associé à un risque 2 à 3 fois plus élevé.[100] L'immobilisation prolongée et les traumatismes sont également des facteurs de risque temporaires. L'augmentation du risque durant une hospitalisation démontre bien ce phénomène, 20% des patients admis dans un service médical développent une thrombose et c'est également le cas de 40% des patients admis en chirurgie.[100] Certains facteurs de risques sont spécifiques aux femmes, l'incidence de la maladie thromboembolique veineuse est notamment accrue lors de la grossesse et durant la période du post-partum. La prise de contraceptifs oraux ou d'une thérapie hormonale à base d'œstrogène est également un facteur de risque.[96] On observe d'ailleurs une un effet joint important avec d'autres facteurs de risques tels que l'obésité, qui couplée à la prise de contraceptifs oraux est associée à un risque 24 fois plus élevé.[101]

Un traitement prophylactique contre la thrombose est parfois recommandé pour les patients hospitalisés.[102] Les voyages avec de long trajets, par exemple en avion, sont également un facteur de risque associé à une augmentation de deux fois du risque normal observée durant les semaines qui suivent le voyage.[100] Les traumatismes majeurs et mineurs sont également un facteur de risque, notamment les traumatismes au niveau des jambes qui sont associés à un risque cinq fois plus élevé de développer une thrombose veineuse dans les mois suivant la blessure.[103] Certaines maladies sont également associées au risque de thrombose veineuse. C'est le cas des cancers qui sont associés à une incidence de 12,4 cas pour 1 000 cancers six mois après le diagnostic.[104] On retrouve également une association avec d'autres maladies avec atteinte neurologique qui sont associés avec un risque de thrombose veineuse accru d'un facteur 2,4 et 4,9 respectivement[105] et les maladies inflammatoires digestives qui sont associées à un risque 3 fois plus élevé.[106] Le diabète est également associé à un risque plus élevé de thrombose. L'incidence chez les diabétiques est de environ 400 par 100 000 individus par an, ce qui correspond environ au double de l'incidence observée dans la population générale.[107] Les mécanismes et les liens de causalités reliant ces facteurs de risques à la thrombose veineuse ne sont cependant pas élucidés et constituent un axe de recherche important. Le diagnostic d'une thrombose veineuse profonde peut se faire avec plusieurs méthodes. La première est un score clinique appelé score de Wells. Il s'agit d'un barème allant de -2 à 9 dépendant de plusieurs critères physiques et d'informations sur les antécédents médicaux du patient.[108] Cependant, le score de Wells ne permet pas à lui seul de diagnostiquer une thrombose veineuse mais sert à appuyer les résultats des tests paracliniques. La mesure des D-dimères dans le sang est également un outil employé. Les D-dimères sont des produits de

dégradation de la fibrine et sont témoins d'une activité de génération de fibrine. Cependant le taux D-dimères est un marqueur peu spécifique car il est également associé à d'autres états tels que l'inflammation, les cancers, les traumatismes, etc... En définitive, le test des D-dimère est seulement utilsé pour éliminer un diagnostic de thrombose veineuse. L'imagerie peut également être employée, on a alors recours à des méthodes telles que l'écho doppler des membres inférieurs, l'angioscanner pulmonaire, la scintigraphie pulmonaire de ventilation-perfusion et l'échocardiographie transthoracique.[109, 15] Le traitement et la prévention chez les profils à risque se font à l'aide de médicaments anti-coagulants comme l'héparine.[110]

3.3 Génétique

Déjà mentionnée plus haut, la génétique joue un rôle important dans le risque de thrombose veineuse. Des mutations ayant un effet très important ont déjà été identifiées. Des mutations provoquant des déficits en protéine C, protéine S et en antithrombine ont un effet très fort sur le risque de thrombose mais sont très rares, ces déficiences étant en général rencontrées dans moins de 1% de la population.[81] Les mutations à l'origine de ces déficiences sont des mutations de perte de fonction conduisant à une production diminuée de la protéine ou à la production d'une protéine mutée incapable de remplir son rôle.[111, 112, 113] On retrouve également des mutations procurant un risque plus modeste mais qui sont plus fréquentes. C'est le cas de FV Leiden R506Q, un polymorphisme d'un seul nucléotide également désigné par l'identifiant rs6025. Il substitue une arginine par une glutamine au niveau du site de clivage de la protéine C activée. Cette variation engendre un retard de l'inactivation de FVa par la protéine C activée, ce qui conduit à un phénotype de résistance à la protéine C activée qui est à l'origine d'une augmentation de la génération de thrombine.[67] Cette mutation a une fréquence d'environ 5% chez les Européens et est retrouvée chez 2% des patients d'origine Européenne avec une thrombose veineuse.[81] Une mutation dans la région 3'UTR du gène de la prothrombine, la mutation 20210A ou rs1799963, est associée à une quantité légèrement plus élevée de prothrombine dans la circulation et à un risque accru d'un facteur 2 à 3 de thrombose veineuse. Cette mutation est quasiment exclusive aux Européens et est rencontrée dans 3% à 6% des cas de thrombose veineuse en fonction des études.[81, 114] Les deux mutations étant relativement communes, il est possible de rencontrer des porteurs de 20210A et FV Leiden, ces individus doubles hétérozygotes présentent un risque de thrombose veineuse augmenté de 20 fois. Ces mutations ont également un effet combiné avec des facteurs de risque dus à l'environnement, notamment l'obésité, les individus avec un IMC supérieur à 30 et porteurs de FV Leiden ou 20210A ont un risque accru d'un facteur 8 et 7 respectivement.[115, 101] Le même phénomène se produit lorsque la présence d'un groupe sanguin non-O et la présence de la variation FV Leiden sont observées, dans ce cas le risque de maladie thromboembolique veineuse est augmenté d'un facteur 23.[116]

Locus	Polymorphisme	Alleles[(1)]	RAF[(2)]	OR
ABO	rs579459	C/T	0,22	1,50
	rs8176749	A/G	0,10	1,50
F2	rs1799963	A/G	0,008	2,29
	rs3136516	A/G	0,40	1,12
F5	rs6025	A/G	0,03	3,25
	rs4524	A/G	0,75	1,20
F8	rs114209171	T/C	0,80	1,11
F11	rs2289252	T/C	0,41	1,35
	rs2036914	C/T	0,60	1,35
FGG	rs2066865	T/C	0,25	1,25
	rs2227421	C/A	0,34	1,13
KNG1	rs7100446	Thr/Ile	0,40	1,20
PROCR	rs6088735	T/C	0,23	1,11
	rs867186	Gly/Ser	0,10	1,22
PROS1	rs121918472	Pro/Ser	0,002	6,57
	rs121918474	Glu/Lys	0,009	5,00
SERPINC1	rs2227624	Glu/Val	0,004	2,30
	rs2227589	T/C	0,10	1,05
SLC44A2	rs2288904	Arg/Gln	0,76	1,28
STXBP5	rs1039084	A/G	0,55	1,11
THBD	rs16984852	T/G	0,005	2,80
TSPAN15	rs78707713	T/C	0,90	1,42
VWF	rs1063856	Ala/Thr	0,40	1,20
MTHFR	rs1801133	Ala/Val	0,25	1,20
CYP4V2	rs13146272	Lys/Gln	0,56	1,24

Tableau 1. Principales variations associées au risque de thrombose veineuse.
(1) Allèle à risque/allèle protecteur. Pour les polymorphismes non-synonymes, les acides aminés sont indiqués à la place des nucléotides.
(2) Fréquence de l'allèle à risque.

Tableau issu des données des articles de Trégouët et Morange, Br J Haematol 2018[117] et Rosendaal FR, J Thromb Haemost 2009[81] ;

La variation 10034T ou rs2066865 du fibrinogène est également associée à un risque de thrombose accru. Il ságit d'un polymorphisme fréquent puisque la fréquence de l'allèle à risque se situe entre 25% et 30% en fonction des études.[81, 118] Le fibrinogène est un élément clé de la formation du thrombus. Enfin, les différents allèles du gène ABO sont associés au risque de thrombose, les individus avec un groupe non-O ayant un risque accru d'un facteur 2 à 4,[119] les individus du groupe sanguin O présentant des taux de vWF et de FVIII réduits.[120] D'autres variations aux effets moindres ont été identifiées, la plupart grâce à des études d'association pan-génomiques. La plupart sont des variations plutôt communes associées avec

un risque relatif compris entre 1 et 2. On note par exemple la variation rs867186 localisée dans *PROCR*, le gène codant pour le récepteur à la protéine C.[121] Des travaux ont montré que des variations de ce gène sont associées au taux de récepteur soluble à la protéine C qui joue un rôle d'inhibiteur de l'activation de la protéine C.[122] Des variations affectant FXI,[123] et des variations sur des gènes aux fonctions moins connues comme *SLC44A2* et *TSPAN15*.[124]Les analyses génétiques ont permis d'identifier un grand nombre de gènes associés à la thrombose veineuse, les principaux sont résumés dans le tableau 1. Ces études ont permis de révéler des mécanismes importants dans le développement du thrombus, cependant la fonction de nombreux gènes identifiés est encore inconnue et la caractérisation de leur rôle représente un travail colossal.

4 Voies physiopathologiques partagées entre les coronaropathies et la thrombose veineuse

Les deux pathologies cardiovasculaires auxquelles je m'intéresse en particulier dans ce manuscrit sont les coronaropathies et la thrombose veineuse. Les deux maladies ont été présentées avant et on peut remarquer des éléments qui permettent de les distinguer, en particulier la dyslipidémie qui est un facteur de risque majeur des les coronaropathies.[38] Plusieurs travaux suggèrent que la dyslipidémie joue un rôle dans le risque de thrombose veineuse, cependant ces résultats sont controversés.[125, 126] On remarque cependant que les coronaropathies sont favorisées par une hypertension artérielle[33] tandis que la thrombose veineuse se développe plutôt dans un contexte de stase qui fait partie des trois éléments de la triade de Virchow.[82] On ne peut cependant pas s'empêcher de remarquer que des facteurs de risque commun existent, tels que l'obésité qui est un facteur de risque très significatif dans les deux maladies[34, 100, 125] et le diabète qui double grossièrement le risque de développer les deux pathologies.[35, 107, 125] Ces points communs entre les deux pathologies, en plus du fait que dans les deux cas on remarque une augmentation de l'activation des plaquettes et des facteurs de l'hémostase[127, 128] suggèrent un lien entre l'athérosclérose et la thrombose veineuse. Ceci laisse penser que des mécanismes communs aux deux pathologies existent et c'est en effet le cas lorsqu'on s'intéresse au processus d'obstruction des artères coronaires après rupture de la plaque d'athérome puisque cette étape fait intervenir les facteurs de l'hémostase, de la même façon que la formation du thrombus dans la thrombose veineuse, bien que la composition du thrombus soit différente dans les deux cas.[129] Ces éléments ont amené de nombreux chercheurs à s'intéresser aux liens entre les deux pathologies. Il a en premier lieu été suggéré que la présence de plaques d'athérome était un facteur de risque de thrombose veineuse. Des travaux ont montré que la présence d'athérosclérose asymptomatique était un facteur de risque indépendant pour la thrombose veineuse profonde.[130] De la même manière,

plusieurs travaux ont montré un risque plus élevé de développer des maladies cardiovasculaires dues à l'athérosclérose, notamment des infarctus du myocarde, après un événement de thrombose veineuse.[131] On note également que les statines, un médicament utilisé pour le traitement des coronaroptahies en raison de leur effet hypolipémiant et leur effet sur la fonction endothéliale, permettent également de réduire significativement le risque de thrombose veineuse, ce qui pourrait indiquer un mécanisme pathophysiologique partagé entre les deux pathologies.[132] Le fait que ces maladies partagent un certain nombre de facteurs de risque et que chacune soit un facteur de risque indépendant de l'autre remet en question le fait qu'il s'agit de deux pathologies aux mécanismes distincts.[133]

Les données génétiques concernant les deux pathologies semblent également supporter l'existence d'un mécanisme commun à la thrombose veineuse et à la thrombose artérielle. Parmi les *loci* associés aux deux pathologies on retrouve notamment *PROCR*,[68, 121] *ABO*[7, 119] et le fibrinogène.[118, 134] De plus des variations communément associées au risque de thrombose veineuse ont été associés aux coronaropathies ou à des maladies ischémiques similaires comme la mutation G20210A de la prothrombine et la mutation FV Leiden.[134] Des variations associées avec les coronaropathies mais pas avec la thrombose veineuse apportent également des informations intéressantes. C'est le cas de variations de la glycoprotéine IIIa,[129, 134] impliquée dans l'agrégation des plaquettes. On remarque également que la variation C677T du gène *MTHFR* est associée au risque de maladie coronarienne.[134] Or ce gène est impliqué dans le métabolisme de l'homocystéine.[135] L'hyperhomocystéinémie est un facteur de risque de coronaropathies et de thrombose veineuse et l'excès d'homocystéine pourrait endommager la paroi vasculaire.[136] Ces données indiquent que les facteurs de l'hémostase, qui jouent un rôle important dans la thrombose veineuse pourraient également jouer un rôle dans le risque de coronaropathies indépendamment de la dyslipidémie, en particulier pour les événements aigus lors desquels il y a une rupture de la plaque d'athérome qui provoque l'activation de la cascade de coagulation.

5 Le diabète sucré

Le diabète sucré regroupe un ensemble de maladies métaboliques provoquant une augmentation de la glycémie entrainant des complications à plus ou moins long terme. On distingue trois grande catégories regroupant la plupart des cas de diabète.

- Le diabète de type I, qui résulte d'un déficit de production d'insuline.
- Le diabète de type II qui résulte d'une résistance à l'activité de l'insuline des cellules puis à une déficit de production d'insuline.
- Le diabète gestationnel qui correspond à une rapide augmentation de la glycémie chez une femme enceinte sans antécédents de diabète.

Les premiers symptômes sont un besoin fréquent d'uriner, une sensation de soif et de faim accrue, une perte de poids inexplicable et une sensation de fatigue chronique. Ces symptômes apparaissent rapidement après le début de la maladie. C'est une maladie chronique qui ne peut pas être guérie bien que des traitements existent. Il n'y a pas de moyen de prévention à ce jour et la non prise d'un traitement adéquat est fatale.[137]

5.1 Physiopathologie

Pour comprendre comment se développe le diabète, il est important de comprendre le rôle de l'insuline et son importance pour l'organisme. L'insuline est une hormone produite par les cellules β du pancréas. La fonction principale de cet organe est la régulation endocrine de la glycémie. Il dispose de quatre types cellulaires différents, chacun produisant une hormone spécifique dans ce but.[138] La fonction endocrine du pancréas est remplie par des régions appelées îlots de Langerhans, du nom du médecin les ayant identifiés,[139] c'est dans ces îlots que se trouvent les cellules produisant les hormones, notamment les cellules β qui sont responsables de la production, du stockage et de la sécrétion d'insuline. La cellule β est sensible à la quantité de glucose du fait qu'elle présente des transporteurs de glucose GLUT2 à sa surface, ce qui permet au glucose de pénétrer dans la cellule en fonction du gradient de concentration entre le milieu extracellulaire et intracellulaire. Lorsque le glucose rentre dans la cellule, il rentre dans la voie de glycolyse et va modifier le rapport ATP/ADP de la cellule, engendrant un série d'évènements qui vont amener à la libération des vésicules d'insuline.[140] L'insuline est une hormone peptidique composée de 51 acides aminés produite par le gène *INS*.[141] Lorsqu'elle est libérée, elle se lie à des récepteurs spécifiques situés sur les cellules cibles du foie, du tissu adipeux et des muscles squelettiques. Une fois liée l'insuline va activer une cascade d'événements qui va aboutir à la présentation de transporteurs du glucose GLUT4 à la surface des cellules cibles pour leur permettre d'intégrer le glucose en excès dans la circulation et promouvoir la conversion du glucose en glycogène dans les cellules musculaires et les hépatocytes ainsi qu'en triglycérides dans les hépatocytes et les adipocytes afin de le stocker.[142] Les effets physiologiques de l'insuline sont présentés dans la figure 5.

Dans le cas du diabète de type I, pour des raisons encore mal comprises, les cellules β du pancréas sont détruites ce qui engendre une production insuffisante d'insuline. Sans insuline, l'organisme n'est plus capable de réguler la glycémie et le glucose en excès ne peut plus être intégré par les cellules.[137] Ceci peut mener à des valeurs d'hyperglycémie extrêmes qui peuvent avoir des conséquences fatales. Les deux manifestations aiguës de l'hyperglycémie sont l'acidocétose diabétique et le coma hyperosmolaire non-cétonique. Dans la première, le manque d'insuline entraine la libération d'acide gras par le foie, sous la forme de corps cétoniques qui vont diminuer le pH sanguin, la chute de pH peut entraîner une perte de

conscience et la formation d'œdèmes puis la mort.[143] Dans la seconde, c'est la perte du glucose en excès par l'urine qui entraîne la perte d'eau et d'autres solutés par osmose qui débouche sur la perte de conscience.[144] Dans le cas du diabète de type II, la résistance à l'insuline et le déficit de production d'insuline seraient provoqués par l'obésité, des adipokynes et l'effet délétère de certains lipides tels que les acides gras non-estérifiés.[145] Ces mécanismes sont très complexes et encore mal compris.

Une hyperglycémie chronique a également des conséquences à long terme qui se manifestent sous la forme de complications. Elles sont regroupées dans deux grandes classes, les complications microvasculaires telles que la rétinopathie, la néphropathie et les neuropathies, puis les complications macrovasculaires qui se manifestent sous la forme de maladies cardiovasculaires. La dépression est également une complication du diabète mais elle n'est pas classée dans les deux catégories présentées plus haut.[146]

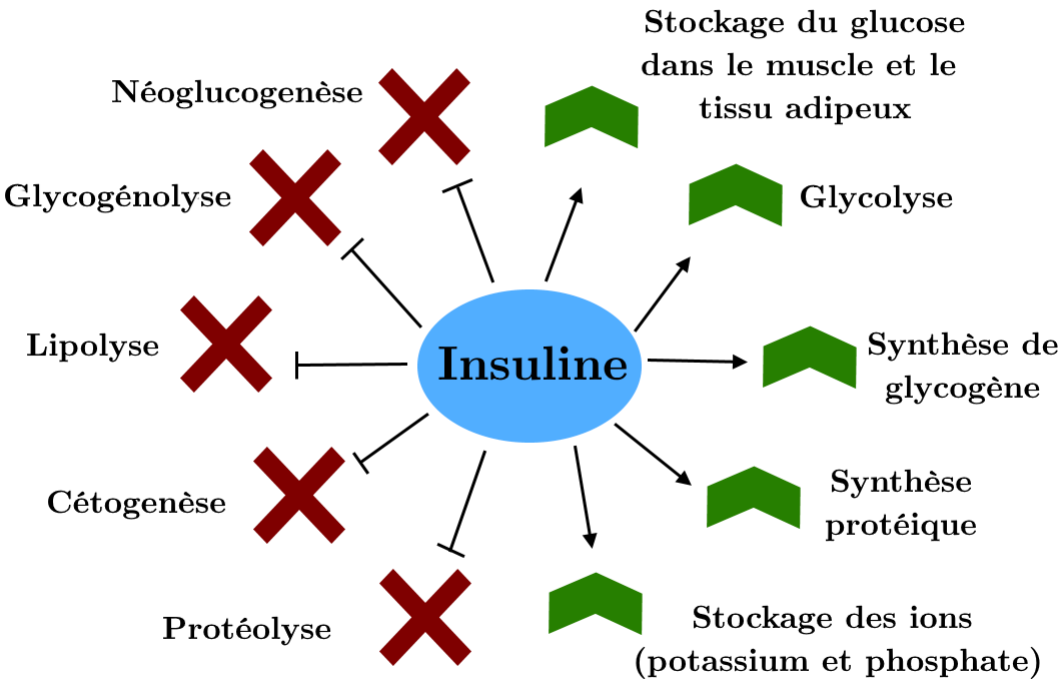


Figure 5 Effets métaboliques de l'insuline. L'insuline favorise le stockage et la conversion en glycogène du glucose circulant mais favorise également l'utilisation du glucose comme source d'énergie prioritaire des cellules. L'insuline est également un hormone anabolisante qui stimule la synthèse protéique et la croissance cellulaire. L'insuline joue également un rôle anti-catabolique en freinant la dégradation des protéines et bloque également le métabolisme des lipides en empêchant leur libération sous la forme de corps cétoniques ainsi qu'en bloquant la lipolyse.

Le contrôle de la glycémie semble être un élément clé dans le développement des complications. La présence de glucose en excès crée un environnement oxydant et favorise la formation de produits de glycation avancée (AGEs) suivant une réaction analogue à la réaction de Maillard. Ces AGEs peuvent se fixer aux récepteurs des produits de glycation avancée (RAGEs) qui vont stimuler l'activation de la réponse inflammatoire et la production de réactifs intermédiaires de l'oxygène. Ces événements vont contribuer à endommager l'endothélium vasculaire et déboucher sur différentes complications en fonction de l'organe touché.[147, 91]

5.2 Génétique

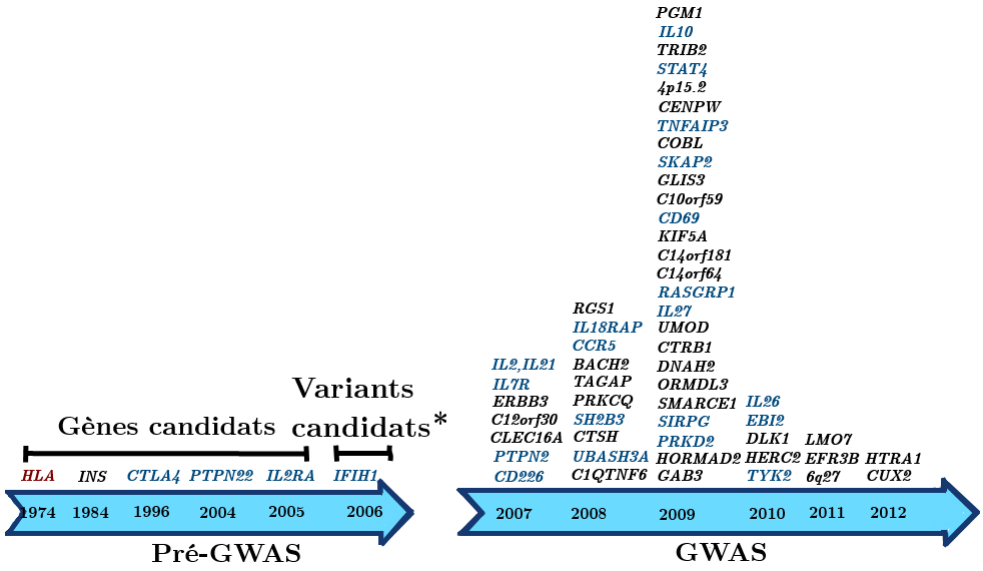


Figure 6 Principaux *loci* associés au diabète de type I. Le *locus HLA* en rouge a été découvert avant les premières analyses pan-génomique et inclue un grand nombre de variations et de gènes dont *HLA-DQB1* mentionné dans le texte. Les *loci* en bleu sont impliqués dans la réponse immunitaire. **Locus* identifié à partir d'une analyse basée sur une puce portant environ 20 000 variations non-synonymes réparties sur tout le génome.

Illustration issue de l'article de Bakay et al, Genes (Basel) 2013[148]

Une composante génétique joue un rôle important dans le développement de la maladie qui a été démontrée comme étant héritable.[149] Une cinquantaine de *loci* ont été associés au diabète de type I.[150] Ceux-ci sont présentés dans la figure 6. Cependant le gène le plus convaincant est *IDDM1* ou *HLA-DQB1*,[151] un gène du *locus* du complexe majeur d'histocompatibilité qui code un récepteur impliqué dans la présentation des antigènes par le CMH2 aux lymphocytes T $CD4^+$.[152] Près de la moitié des signaux associés au diabète de type I sont localisés sur ce *locus*.[153] Le *locus HLA* est également associé à une large variété de maladies inflammatoires

et autoimmunes, ce qui suggère une origine autoimmune pour le diabète de type I.[154, 155, 156] Dans le cas du diabète de type 2, on retrouve environ 80 *loci* associé à la maladie et la plupart des variations identifiées sont des variations fréquentes. Cependant ces variations n'expliquent qu'un faible part de l'héritabilité de la maladie.[157]

Une fois le diabète diagnostiqué, il est nécessaire de suivre une thérapie à l'insuline à vie, une diète stricte, un suivi de la consommation de glucides et la pratique d'activité physique sont également essentiels.[158, 159]

5.3 Épidémiologie

Le diabète, dont une écrasante majorité correspond à des cas de diabète de type 2, est une maladie très courante dans les pays développés et dont l'incidence ne fait que progresser. En France la prévalence est passée de 2,6% en 2000 à 4,4% en 2009, soit 2,9 millions de malades. De la même manière aux États-Unis, on observait une prévalence de 9,8% en 1994 qui est passée à 12,4% en 2012.[160, 161] Parmi tous les cas de diabète, la part estimée de diabètes de type I est de 5% à 10%.[162] L'incidence du diabète de type I varie cependant en fonction des populations. On retrouve notamment une incidence élevée chez les populations homogènes comme c'est le cas en Finlande ou en Sardaigne,[163] d'après la *Diabetesliitto*, 15% des diabètes diagnostiqués en Finlande sont des diabètes de type I.[164]

Avant la découverte de l'insuline par le docteur Frederik Banting et Charles Best en 1920, la première cause de mortalité était le coma diabétique dû à l'hyperglycémie. Cependant après cette date, il y a eu une augmentation du taux de mortalité dû aux complications du diabète,[165] notamment la néphropathie diabétique pour les morts avant 30 ans et les complications cardiovasculaires, à qui sont attribués deux tiers des décès après 30 ans.[166, 167]

Les causes du diabète de type I ne sont pas connues mais différents facteurs environnementaux sont suspectés d'intervenir, peut être même en jouant le rôle déclencheur. Plusieurs hypothèses ont été avancées, certaines impliquant différentes infections virales ou la flore intestinale.[168] Plus de la moitié des cas de diabète de type I sont diagnostiqués avant 30 ans,[162] en général après que le patient ne présente des symptômes aigus d'hyperglycémie. Dans le cas du diabète de type II, les principaux facteurs de risque sont l'âge, la dyslipidémie, l'hypertension artérielle, l'obésité, en particulier l'obésité abdominale, le manque d'activité physique et dans certains cas l'appartenance à un groupe ethnique particulier.[168] Le diabète est la première cause d'insuffisance rénale chronique terminale dans le monde et un près d'un quart des diabétiques développent une maladie rénale.[169] Le diabète de type I serait également associé à une réduction de 12 ans de l'espérance de vie[170] et à un risque doublé de développer une maladie cardiovasculaire.[171] Un autre fait intéressant est le lien entre la néphropathie diabétique et le risque de maladie cardiovasculaire. En effet le risque de

développer une maladie cardiovasculaire est dix fois plus élevé chez les diabétiques de type I avec une néphropathie diabétique que chez les diabétiques de type I non-néphropathes. Le temps écoulé entre le diagnostic du diabète et le développement de la maladie cardiovasculaire est plus court chez les patients avec une néphropathie diabétique, avec en moyenne 12-13 ans contre 16-17 ans. En revanche, les mécanismes reliant la néphropathie diabétique et les maladies cardiovasculaires ne sont pas connus.[172] Les mêmes résultats sont observés dans le cas du diabète de type II, avec cependant une réduction de l'espérance de vie moins importante mais un impact similaire de le néphropathie diabétique.[173, 174]

5.4 Complications associées

Les complications du diabète ont été présentées plus haut. Il est cependant nécessaire de présenter plus en détails les complications vasculaires dans ce manuscrit. On distingue deux catégories de complications vasculaires. La première catégorie comprend les complications microvasculaires qui sont dues aux dommages sur les petits vaisseaux et qui ont des conséquences plus ou moins sévères en fonction de l'organe touché. Ces complications sont la néphropathie, la neuropathie et la rétinopathie. La seconde catégorie est celle des complications macrovasculaires, qui sont dues aux dommages sur les artères, les manifestations les plus sévères étant les infarctus du myocarde, les accidents vasculaire-cérébraux et l'artériopathie oblitérante des membres inférieurs. Les complications neurologiques du diabète ne seront en revanche pas discutées ici. Comme expliqué plus haut, le contrôle de la glycémie est un élément clé dans le développement des complications du diabète et il a été montré qu'un bon contrôle de la glycémie réduisait le risque de complications. Les effets d'une hyperglycémie chronique sont synthétisés dans la figure 7.[175] L'hyperglycémie affecte principalement les tissus par le biais de quatre voies métaboliques différentes qui sont cependant étroitement liées. Ces voies métaboliques favorisent d'autres substrats dans les conditions physiologiques normales mais l'excès de glucose favorise son utilisation en tant que substrat.[176]

La première voie est celle des polyols ou glycols, dont l'enzyme clé est l'aldose réductase qui catalyse la conversion du glucose en sorbitol qui est ensuite oxydé en fructose par la sorbitol déshydrogénase. La première conséquence de la suractivité de cette voie métabolique est un manque de NAD(P)H qui est consommé par l'aldose réductase, ce qui empêche la conversion du disulfide de glutathion (GSSG) en glutathion (GSH) par la glutathion réductase. Le GSH est une molécule clé dans la protection contre le stress oxydatif. La seconde conséquence vient du fait que la sorbitol déshydrogénase utilise du NAD^+ pour oxyder le sorbitol, ce qui résulte en un changement dans le rapport entre NADH et NAD^+ qui favorise l'inhibition de la glyceraldéhyde-3-phosphate déshydrogénase (GAPDH) ce qui à son tour favorise l'accumulation de triose phosphate, un élément précurseur du méthylglyoxal et du diacylglycérol

(DAG) tous deux favorisant la formation d'AGEs.[177]

La seconde voie est celle des AGEs qui a déjà été présentée plus haut. Elle est favorisée par des produits dérivés du glucose comme les DAGs qui sont en partie issus de l'activité de la voie des polyols mais aussi issus d'autres processus. Les AGEs sont formés par la modification post-traductionnelle non-enzymatiques des protéines par ces produits et sont présents dans tous les compartiments tissulaires.[178] Ces AGEs ont des fonctions et des propriétés altérées et il a été montré que la formation d'AGEs altère les interactions entre les protéines de la lame basale, ce qui entraîne un épaississement de cette dernière.[179] Les AGEs peuvent se retrouver dans le milieu extracellulaire et activer les RAGEs des cellules endothéliales et des monocytes ce qui favorise notamment l'activation de la protéine kinase C (PKC) et donc la production de facteur de croissance vasculaire endothélial (VEGF), l'activation de la voie NF-kB et la production de réactifs intermédiaires de l'oxygène (ROS).[180] Tous ces éléments ont pour effet de favoriser à leur tour la différentiation des monocytes en macrophages, la production de TNF-β et de cytokines, favorisant ainsi la création d'un environnement pro-inflammatoire.[176]

La troisième voie est la voie PKC, qui est dépendante des AGEs mais peut être directement activée par les DAGs. La PKC stimule la production de VEGF ce qui favorise la perméabilisation des vaisseaux, active la voie NF-kB pro-inflammatoire, augmente la production de ROS et inhibe la production d'oxyde nitrique synthase endothéliale (eNOS), réduisant ainsi la production d'oxyde nitrique et entrainant la suppression du message pro-vasodilatation.[176] La quatrième voie est celle des hexosamines. Elle est dépendante de l'excès en glucose comme les autres mais serait également favorisée par la production de ROS. Le glucose en excès serait converti en fructose-6-phosphate qui sert de substrat à la glutamine-fructose-6-phosphate aminotransférase (GFAT) pour former du glucosamine-6-phosphate, lui même lié aux radicaux des tyrosines et des sérines par l'enzyme O-GlcNAc transferase (OGT). Ces modifications post-traductionnelles peuvent agir de la même façon que les AGEs sur la fonction des protéines mais peuvent également entrer en compétition avec les groupements phosphates qui utilisent les mêmes substrats.[181]

Une autre hypothèse avancée est celle de la perte du contrôle de la production d'énergie dans la cellule, qui pourrait être en partie une conséquence des quatre voies mentionnées au-dessus. En effet les cellules des tissus cibles ne peuvent pas contrôler le transport du glucose et se retrouvent en excès de substrat,[182] ce qui favorise grandement la glycolyse et peut mener à des dysfonctionnement de la mitochondrie qui favoriserait la production de ROS, notamment de superoxyde (O_2^-). Ces dysfonctionnements pourraient également déboucher sur un découplage de la chaîne respiratoire et de la production d'ATP.[91]

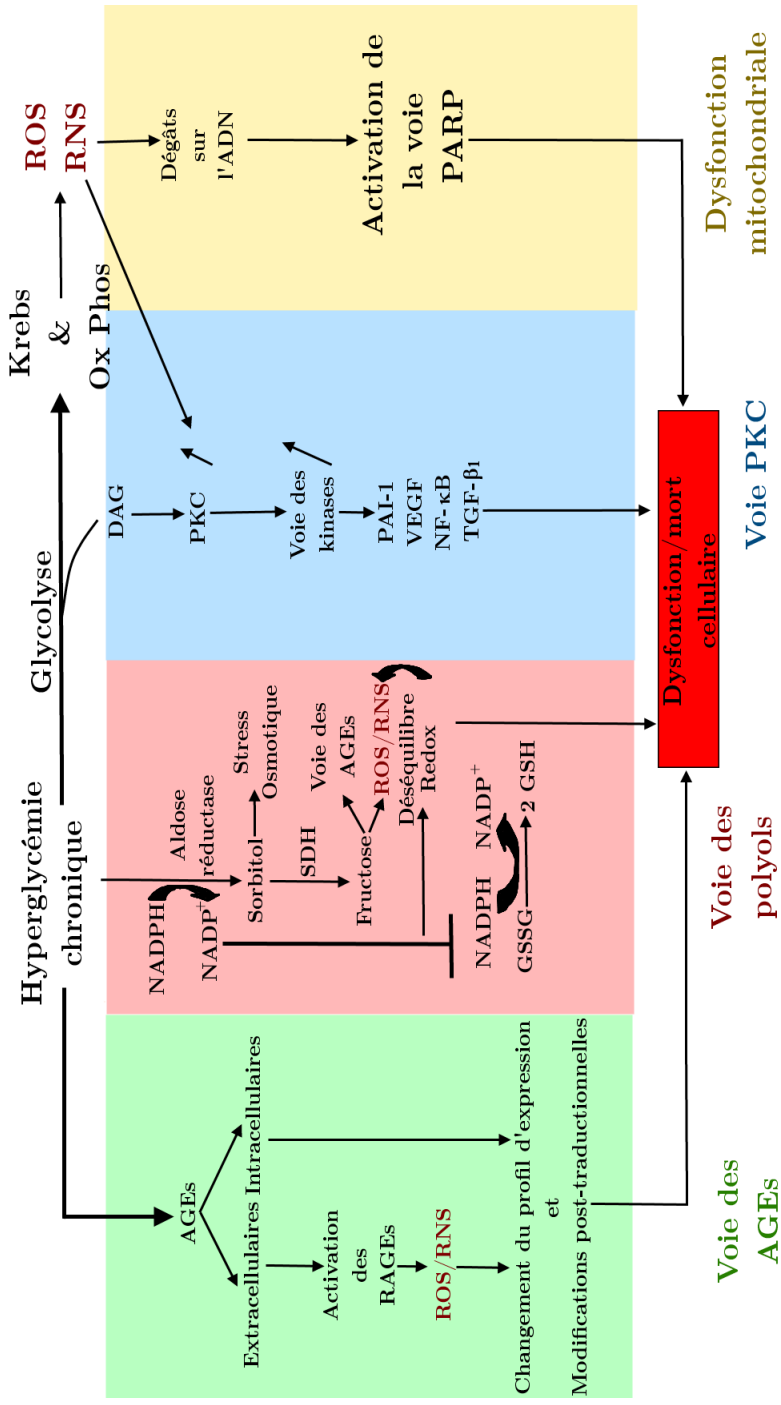


Figure 7 Synthèse des voies physiologiques affectées par l'hyperglycémie chronique. Cette figure montre les effets indépendants de chacune de ces voies métaboliques. Il est cependant important de rappeler que ces processus interagissent fortement entre eux puisque la voie des polyols influence la formation de DAGs et que les DAGs favorisent la formation d'AGEs, l'activation des RAGEs favorise également l'activation de la PKC. Une grande partie des effets nocifs de l'hyperglycémie chronique est due à la formation de radicaux libres et de ROS/RNS qui endommagent l'ADN des cellules cibles des complications.
Figure composée à partir des données présentées dans les articles cités dans le manuscrit [91, 178, 181, 179, 180, 177, 176]

Bien que le contrôle de la glycémie soit un élément clé il n'est pas le seul facteur déterminant et des complications peuvent tout de même apparaitre malgré un contrôle strict de la part du malade.[183] Un processus auto-immunitaire similaire à celui impliqué dans le développement de la maladie pourrait également contribuer au développement des complications.[184] Il a notamment été montré que les cellules endothéliales de la rétine présentent une expression anormale du complexe majeur d'histocompatibilité en particulier les allèles *HLA-DR* et *HLA-DQ* déjà mentionnées, ce qui entraine l'activation de lymphocytes T pouvant induire des réponses auto-immunitaires.[185] De la même manière il a été observé qu'il y avait une accumulation juxtaglomérulaire de lymphocytes T et une expression accrue de molécules d'adhésion par les cellules endothéliales, épithéliales et mésengiales du rein coïncidant avec le développement de la néphropathie.[186] De plus les troubles du système nerveux autonome, un symptôme de la neuropathie, ont également été associés à la présence d'auto-anticorps dirigés contre les structures nerveuses autonomes.[187] L'existence ou non d'une relation entre le contrôle de la glycémie et le développement de l'auto-immunité n'est pas encore claire mais l'auto-immunité pourrait être un facteur de risque indépendant à étudier. La génétique joue également un rôle important et il a été montré que les complications du diabète sont hautement héritables. Par exemple le risque de développer une néphropathie diabétique est doublé lorsque la pathologie est présente chez un frère ou une sœur.[188] La conclusion est similaire dans le cas de la rétinopathie progressive.[189] Les éléments manquent dans le cas de la neuropathie mais les approches "gène candidat" dans le cas des neuropathies non-diabétiques donnent des pistes intéressantes, car les gènes identifiés pourraient être testés dans le contexte de la néphropathie diabétique.[190] L'architecture génétique des complications macrovasculaires est en revanche plus évasive, bien que l'impact de la génétique sur les maladies cardiovasculaire telles que les coronaropathies a déjà été démontré.[7] Les complications macrovasculaires du diabète de type I sont souvent précédées par une néphropathie diabétique.[172] De plus des travaux ont montré qu'une tendance familiale à la néphropathie est également accompagnée d'une tendance à développer des complications macrovasculaires, même chez les membres de la famille épargnés pas la néphropathie.[191] Ces éléments suggèrent un profil génétique particulier commun à la néphropathie et aux complications macrovasculaires, potentiellement différent de celui observé dans le cas des coronaropathies classiques. Les différentes complications du diabète partagent toutes des mécanismes effecteurs clés mais on distingue des pathologies distinctes en fonction du tissu touché, elles sont présentées ci-dessous :

Néphropathie La néphropathie diabétique est une complication majeure touchant entre 30% et 40% des diabétiques de type I et 20% et 30% des diabétiques de type II.[192, 193] Cette complication se développe lentement et apparait généralement 10 à 20 ans après la déclaration du diabète avec une protéinurie suivie d'une diminution progressive du débit de filtration glomérulaire.[91] L'environnement cytotoxique créé par les

mécanismes mentionnés plus haut affecte les cellules endothéliales des petits vaisseaux du glomérule mais aussi les podocytes et les cellules mésengiales qui sont en contact direct avec ces vaisseaux. Les dommages soutenus par l'endothélium vasculaire, la lame basale, qui est la membrane protéique servant de support aux cellules endothéliales,[194] et les podocytes conduisent à la fragilisation de la barrière glomérulaire ce qui conduit à la protéinurie.[195] La réponse initiale de l'organisme est une expansion du rein par hypertrophie et une hyperplasie du tissu mésengial, ce qui résulte en une hyperfiltration au stade précoce de la complication. Alors que la pathologie avance, la lame basale séparant les cellules endothéliales et les podocytes s'épaissit et la croissance mésengiale envahi puis ferme le glomérule, lorsque le processus s'étend à tous les glomérules il conduit à une insuffisance rénale chronique terminale.[91] Lorsque la protéinurie devient trop importante elle peut entrainer une hypoprotéinémie qui va mener à un flux d'eau des vaisseaux vers les tissus interstitiels en raison de la baisse de la pression oncotique, c'est à dire la pression exercée sur l'eau par la différence de concentration en protéines entre le plasma et les tissus environnants.[196] Ce phénomène contribue à former des épanchements et des œdèmes, en particulier dans les chevilles. La perte du débit de filtration peut mener à une urémie en absence de traitement adapté, c'est une condition fatale causée par l'excès de déchets azotés dans la circulation.[197] La néphropathie diabétique représente une part importante des mises en dialyse et est un facteur de risque pour les complications macrovasculaires du diabète.[169, 172]

Neuropathie La neuropathie est une complication affectant plus de la moitié des individus diabétiques. Elle est due à la dégénérescence des fibres nerveuses périphériques qui fait suite à celle des petits vaissaux irrigant les nerfs.[198] Les dégâts causés par les mécanismes liés au contrôle de la glycémie provoquent l'épaississement de la lame basale et une hyperplasie des cellules endothéliales des vaisseaux, ce qui débouche sur une perte de capacité de diffusion de l'oxygène. Les nerfs les plus touchés sont les nerfs périphériques sensoriels et moteurs ainsi que le système nerveux autonome bien que plus rarement. Ce phénomène touche plus particulièrement les fibres nerveuses longues et affecte donc en premier les extrémités, soit les pieds et les mains. La perte de conductivité des terminaisons nerveuses peut se manifester par une perte de la sensibilité dans les membres touchés ainsi que par des sensations douloureuses, des paresthésies et de l'allodynie.[91] Ces phénomènes conduisent à une perte de précision motrice débouchant sur un risque de chute et de blessure accru.[199] L'altération de la fonction sensorielle peut engendrer des douleurs chroniques diminuant la qualité de vie. De plus le risque accru de blessure couplé à une immunité et capacité de régénération diminuée augmente fortement le risque de gangrène des membres concernés, le diabète étant d'ailleurs associé à un risque accru d'amputation des membres inférieurs.[200] La dégénérescence du système nerveux autonome est associé à des dysfonctionnements

du tube digestif.[201]

Rétinopathie Cette complication est fréquente chez les diabétiques et est une cause majeure de cécité chez les moins de 70 ans.[202, 203] C'est une complication qui évolue sur une longue durée et qui se manifeste en deux étapes, la première étant appelée "non-proliférative" et la seconde "proliférative". Les dégâts subits par les petites artères de la rétine, en particulier ceux causés par l'hyperglycémie, sont à l'origine de cette complication.[91] Au stade non-prolifératif, l'épaississement de la lame basale des vaisseaux et la dysfonction des cellules endothéliales peut avoir pour conséquence de causer la mort des péricytes, des cellules formant la barrière hémato-rétinienne.[204] L'œil étant un organe immuno-privilégié, la barrière hémato-rétinienne permet d'empêcher l'infiltration de cellules immunitaires pouvant avoir un effet toxique sur la rétine, la mort des péricyte permet donc l'infiltration de la rétine par les cellules immunitaires et autres éléments normalement absents.[205] À ce stade de la maladie, il n'y a pas de perte de vision remarquable. Les dégâts sur les petits vaisseaux peuvent amener à leur occlusion et sont fortement associés à une évolution vers le stade supérieur. Au stade proliférative, l'état hypoxique provoque la libération de facteurs entrainant une néo-vascularisation tels que le VEGF,[202] ce qui engendre la formation de nouveaux vaisseaux vont être formés sur la face antérieure de la rétine. Les vaisseaux fragilisés et les nouveaux vaisseaux également fragiles peuvent céder et provoquer un œdème rétinien qui va endommager la vision. Cette néovascularisation engendre également des changements de l'architecture de la rétine et peut mener à la formation d'une membrane fibrovasculaire,[91] en suit un risque de décollement de la rétine. La conséquence à long terme de la rétinopathie est une dégénérescence maculaire aboutissant à la perte de la vue.

Complications macrovasculaires Les complications macrovasculaires sont dues au dysfonctionnement des cellules endothéliales des artères. Ces dysfonctionnement affectent la production de NO et d'angiotensine II (ANG II) et favorisent l'adhésion des macrophages et des lymphocytes.[91] Dans le cas du diabète de type I, les complications macrovasculaires sont souvent accompagnées par une néphropathie qui contribue à ces dysfonctionnements en augmentant le risque d'hypertension.[172] La perte de la fonction régulatrice des cellules endothéliales, l'augmentation de la perméabilité, l'activité pro-inflammatoire et l'augmentation de la pression artérielle accélèrent la formation de la plaque d'athérome. Il en résulte que les diabétiques présentent un risque élevé de développer des maladies cardiovasculaires. Près de la moitié des décès chez les diabétiques sont dus à des maladies cardiovasculaires, notamment des coronaropathies.[206] Il est important de mentionner que les complications macrovasculaires touchent en particulier les diabétiques de type II, qui présentent souvent d'autres facteurs de risque

des maladies cardiovasculaires en plus du diabète, tels que l'obésité, l'hypertension artérielle ou la dyslipidémie.

Bien qu'il soit clairement établi que ces complications sont en partie héréditaires et qu'il y ait des interactions entre celles-ci, la nature des interactions entre ces complications n'est pas clairement établie et les facteurs génétiques favorisant leur développement encore inconnus.

6 Le rein

La fonction rénale est un élément régulièrement mentionné dans le développement des maladies cardiovasculaires. Il a déjà été montré que la néphropathie est une complication courante du diabète et que celle-ci mène à un risque accru de complication macrovasculaire.[171] Cependant la perte de fonction rénale est également associée au risque de maladie cardiovasculaire dans un contexte non-diabétique.[207] Les mécanismes impliqués ne sont pas encore complètement connus mais plusieurs hypothèses ont été mises en avant. Il est cependant nécessaire de présenter l'anatomie et la fonction rénale avant d'entrer dans la relation entre le rein et les maladies cardiovasculaires.

6.1 Anatomie du rein

Les reins sont des organes jumeaux localisés dans la cavité abdominale. Chaque rein est irrigué par une artère rénale et est connecté à un uretère servant à conduire l'urine produite vers la vessie.[196] Le rein est également parcouru par le plexus rénal qui le relie au système nerveux sympathique.[208] L'artère rénale pénètre dans le rein et se sépare en plusieurs embranchements avant de former les artérioles afférentes qui alimentent chacune un néphron en passant dans le corpuscule rénal. Le sang quitte le corpuscule rénal par l'artériole efférente qui se divise en capillaires péritubulaires qui parcourent tout le néphron avant de se regrouper et de rejoindre la veine rénale pour quitter le rein.[208] Le rein est composé d'un parenchyme entouré par une capsule fibreuse, le parenchyme est constitué d'un cortex en périphérie et d'une médullaire vers l'intérieur.[196] La médullaire est organisée en structures coniques appelées pyramides de Malpighi. Entre ces pyramides, le cortex forme des projections appelées colonnes rénales.[208]

Le néphron est l'unité fonctionnelle du rein et un rein en contient près d'un million.[196] Leur anatomie générale est présentée dans la figure 8. Le néphron s'étend du cortex à la médullaire en formant une boucle qui remonte ensuite vers le cortex. Il existe chez l'homme deux types de néphrons, les néphrons corticaux dont la boucle s'enfonce peu profondément dans la médullaire et les néphron juxtamédullaires qui s'enfoncent plus profondément dans

la médullaire.[196] Le néphron est constitué de deux parties, le corpuscule rénal situé dans le cortex et le tubule rénal formant le reste de la boucle.[196] Les tubules rénaux de plusieurs néphrons convergent vers un canal collecteur qui traverse la pyramide de Malpighi jusqu'à son sommet. Les sommets des pyramides débouchent sur les calices qui récoltent l'urine et l'amènent vers une structure en entonnoir appelée le bassinet. Le bassinet devient ensuite l'uretère. Le corpuscule rénal est divisé en deux pôles, le pôle vasculaire qui inclue le glomérule et le pôle urinaire qui inclue la capsule de Bowman. Le corpuscule rénal reçoit le sang de l'artériole afférente qui le quitte ensuite par l'artériole efférente. L'artériole afférente a cependant un diamètre plus large que l'artériole efférente ce qui génère une différence de débit.[208]

Le glomérule : Le glomérule est composé d'un nœud de petits capillaires possédant de larges fenestrations (50 à 100 nanomètres) sans diaphragmes,[209] les cellules épithéliales reposent sur une lame basale composé de collagène IV, de laminine et de sulfate d'héparane, une molécule chargée négativement.[210] L'espace entre les capillaires est occupé par des cellules mésengiales. Ce sont des péricytes spécialisés dérivés des cellules musculaires lisses, leurs filaments d'actine et de myosine leur confèrent une capacité contractile. Elles produisent également des protéines de la matrice extracellulaire comme le collagène IV et la laminine et ont une activité phagocytaire.[211]

Capsule de Bowman : La capsule entoure entièrement le glomérule. Sur la face interne elle est tapissée de cellules spécialisées appelées podocytes. Les podocytes forment des projections appelées pédicelles qui recouvrent les capillaires du glomérule et forment des diaphragmes. Les pédicelles ont une capacité contractile limitée qui leur permet de contrôler la taille des ouvertures. La surface des pédicelles exprime des intégrines qui leur permettent de se fixer à la lame basale des capillaires grâce à la laminine.[209] La face extérieure de la capsule de Bowman est couverte de cellules épithéliales squameuses. L'espace à l'intérieur de la capsule de Bowman est appelé espace de Bowman, il est accessible par les diaphragmes formés par les pédicelles et possède une ouverture sur le tubule rénal.[208]

Les fenestrations des capillaires, la lame basale et les pédicelles forment la barrière glomérulaire qui sépare le pôle vasculaire du pôle urinaire. La seconde partie du néphron, le tubule rénal forme la structure en boucle qui descend vers la médullaire avant de remonter et de mener au canal collecteur. Il est divisé en trois parties, le tube contourné proximal, l'anse de Henle et le tube contourné distal.[196] Le tubule rénal est parcouru par des capillaires péritubulaires et le sens de circulation dans le tubule se fait à contre-sens de la circulation dans les capillaires.

Le tube contourné proximal : Le tubule proximal est couvert de cellules épithéliales présentant des microvillosités organisées en bordure en brosse, ce qui augmente la surface de contact avec la lumière du tube. Ces cellules sont également riches en mitochondries, qui sont concentrées dans le pôle basal. Les cellules du tube contourné proximal expriment des cotransporteurs sodium/glucose, sodium/acides aminés mais aussi des échangeurs sodium/potassium.[196]

L'anse de Henle : L'anse de Henle est une structure en épingle qui forme l'essentiel de la boucle du néphron. C'est d'ailleurs cette région qui est allongée dans les néphrons juxtamédullaires. Elle est divisée en deux parties, l'anse descendante et l'anse ascendante. L'anse descendante est riche en aquaporines ce qui lui confère une importante perméabilité à l'eau mais n'exprime pas de transporteurs pour les électrolytes et les petites molécules. La partie inférieure de l'anse ascendante est couverte de cellules épithéliales fines qui n'expriment pas d'aquaporines et ne sont donc que très peu perméables à l'eau tandis que la partie supérieure de l'anse est couverte de cellules épithéliales plus épaisses qui n'expriment également pas d'aquaporines mais présentent des transporteurs pour différents électrolytes.[196]

Le tube contourné distal : Comme pour le tube contourné proximal, les cellules épithéliales du tube contourné distal sont enrichies en mitochondries. Les cellules du tube contourné distal expriment plusieurs transporteurs dont des cotransporteurs sodium/chlorure, des échangeurs sodium/potassium et des échangeurs sodium/calcium. Les cellules présentent également des récepteurs à la vasopressine.

La sortie du néphron débouche sur le canal collecteur. Les cellules du canal collecteur expriment des canaux sodium et potassium sur leur membrane apical et possèdent elles aussi des récepteurs à la vasopressine.

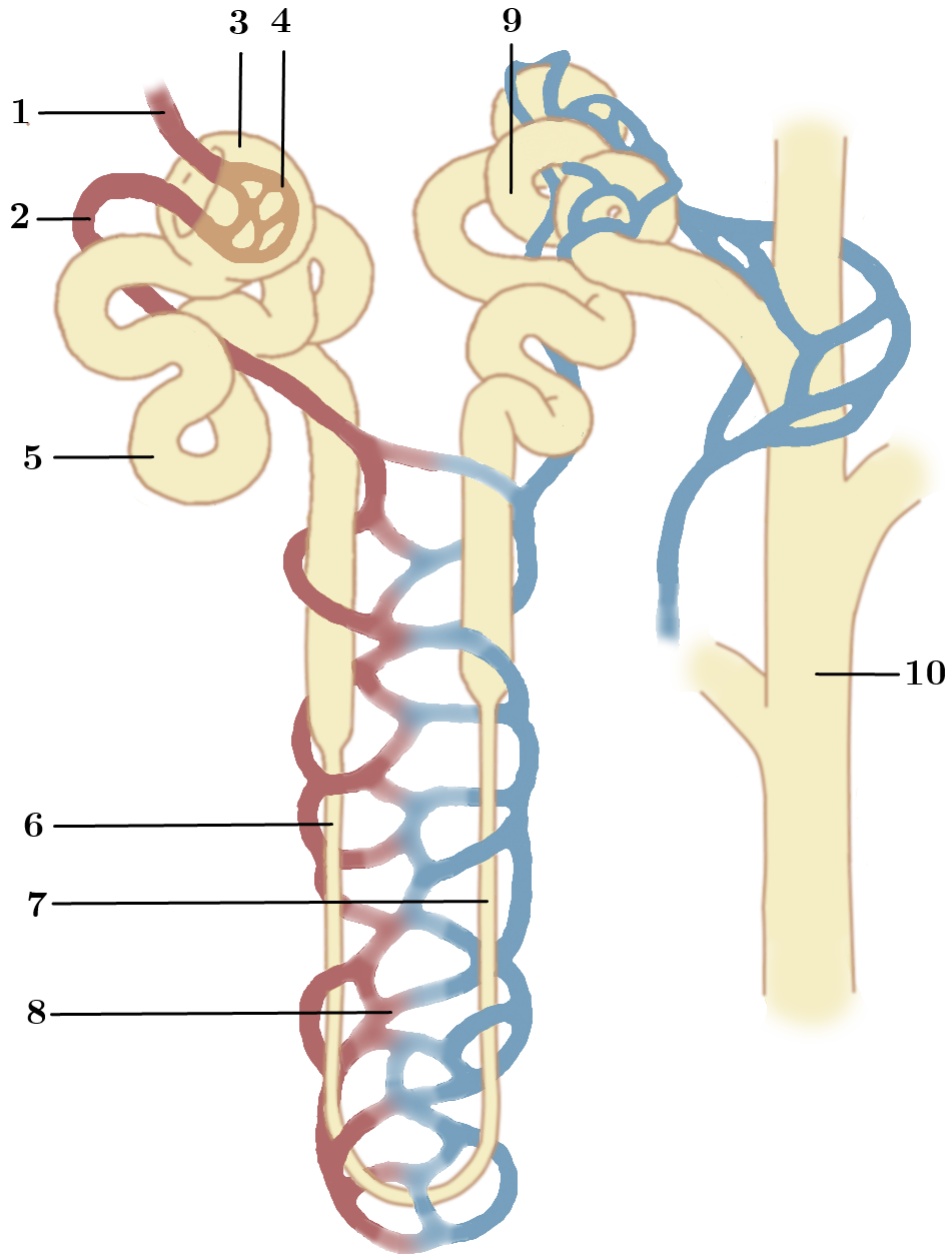


Figure 8 Anatomie d'un néphron. L'artériole afférente (1) pénètre dans la capsule de Bowman (3) et se divise en un nœud de capillaires formant le glomérule (4). Le glomérule et la capsule de Bowman forment le corpuscule rénal. Une partie du plasma est filtrée dans le corpuscule rénal tandis que le reste est évacué par l'artériole efférente (2). Le filtrat pénètre ensuite le tubule rénal en passant par le tube contourné proximal (5) et continue vers l'anse de Henle qui est divisée en une partie descendante (6) et une partie ascendante (7). L'anse de Henle est entourée par les capillaires péritubulaires (8) qui jouent un rôle capital dans la réabsorption des solutés depuis le filtrat. Le filtrat rejoint ensuite le tube contourné distal (9) et est récolté par le canal collecteur (10) qui ne fait plus tout à fait partie du néphron mais joue un rôle dans la filtration.

6.2 La fonction rénale

La fonction principale du rein est l'épuration du sang et un rein est capable de filtrer jusqu'à 180 litres de plasma par jour.[196] Cette fonction est assurée par les néphrons et consiste en quatre mécanismes qui sont la filtration du plasma, la réabsorption des solutés et de l'eau, la sécrétion des éléments indésirables et l'excrétion du produit final. La fonction du néphron est importante pour le maintient de l'osmolalité, le maintient de l'équilibre acido-basique, la régulation de la pression artérielle et la clairance des toxines.[196] Chaque compartiment du néphron présenté ci-dessus joue un rôle dans un des processus de l'épuration sanguine. Le corpuscule rénal est chargé de séparer le plasma des cellules et des protéines. Le tubule rénal est chargé de la réabsorption des molécules importantes et de l'eau et de la sécrétion des toxines et des déchets métaboliques. Le canal collecteur est chargé d'acheminer le produit final de la filtration vers l'uretère afin qu'il soit excrété.[196]

Le filtration : La filtration a lieu dans le corpuscule rénal. Durant cette étape le plasma passe du compartiment vasculaire à l'espace de Bowman grâce à la pression des capillaires du glomérule qui force le plasma à passer par les fenestrations des capillaires. Cette pression est générée par la différence de débit entre les artérioles afférentes et efférentes mais est également modulée par la contraction des cellules mésengiales.[196, 211] Environ un cinquième du plasma traversant le glomérule est filtré, le reste continuant dans les artérioles efférentes puis les capillaires péritubulaires. La barrière glomérulaire permet de séparer le plasma des cellules, plaquettes et protéines grâce à la charge négative de la lame basale et les diaphragmes formés par les podocytes.[209] Le filtrat brut recueilli dans l'espace de Bowman est appelé ultrafiltrat et est ensuite transféré au tubule rénal.

La réabsorption : Durant la réabsorption, le néphron réabsorbe l'eau et les petites molécules de l'ultrafiltrat telles que le glucose, les acides aminés et les électrolytes, les molécules réabsorbées regagnent ensuite la circulation via les capillaires péritubulaires.[196] La quasi totalité des molécules de glucose et des acides aminés sont réabsorbés dans le tube contourné proximal via des cotransporteurs dépendant du sodium ou des transporteurs nécessitant de l'ATP. Ce transport peut être saturé par une présence excessive de ces molécules dans le plasma. C'est notamment le cas du glucose chez les diabétiques puisque la réabsorption du glucose sature autour de 180mg/dL de filtrat.[196] Les deux tiers du volume du filtrat sont réabsorbés dans le tube contourné proximal par diffusion. Cette diffusion est permise grâce à la création d'un gradient de concentration en sodium entre le filtrat et le plasma des capillaires péritubulaires. Ce gradient est créé par le transport actif du sodium depuis le pôle basal des cellules du tube vers le milieu interne. Les ions chlorures suivent les ions sodium par

attraction électrique et l'eau les rejoint par osmose. Cette réabsorption passive n'est cependant pas suffisante puisque qu'il reste encore 60 Litres de filtrat journalier non réabsorbé, la réabsorption du reste du filtrat se faisant dans la région médullaire du néphron où un milieu hypertonique est créé grâce à l'anse de Henle.[196] La circulation à contre courant au niveau de l'anse est importante pour ce phénomène puisque les capillaires péritubulaires récupèrent le sel transporté activement depuis l'anse ascendante imperméable à l'eau. Ceci permet de créer un plasma hypertonique qui circule à contre courant du filtrat et remonte alors vers l'anse descendante perméable, ce qui permet à l'eau de diffuser vers les capillaires. Dans ce système, plus l'anse est longue plus le gradient créé est important.[196] Ce phénomène est amplifié par l'urée qui peut également diffuser depuis l'anse descendante.[212] Le filtrat qui quitte l'anse est hypotonique ce qui facilite la récupération de l'eau restante dans le tube contourné distal et le canal collecteur. La récupération de l'eau dans ces parties du néphron est cependant dépendante de la vasopressine qui stimule l'expression des aquaporines.[196] La réabsorption du sodium est également partiellement modulée par l'aldostérone et le néphron réabsorbe d'autres électrolytes comme le bicarbonate, le potassium et le calcium.[196] Ces mécanismes peuvent également être partiellement régulés par des hormones comme la parathormone ou le calcitriol.[196]

La sécrétion : La sécrétion est un processus similaire à la réabsorption mais allant dans une direction opposée puisque les molécules sont transportées du milieu interne vers le filtrat. Ce processus prend place sur toute la longueur du tubule rénal. La sécrétion concerne des toxines et des déchets métaboliques produits par l'organisme mais aussi des molécules exogènes et peut se faire par diffusion ou par transport actif. Les molécules sécrétées sont notamment l'urée issue du métabolisme des protéines qui est sécrétée par diffusion et aussi par transport actif pour compenser la réabsorption dans l'anse de Henle, l'acide urique issu du métabolisme des acides nucléiques, la créatinine mais aussi des ions hydronium et ammonium, transportés dans le filtrat depuis le tube contourné distal.[196, 213, 214]

L'excrétion : Une fois le filtrat final obtenu, il est amené vers les uretères puis la vessie pour être enfin excrété. L'urine est nettement plus concentrée que le plasma et ceci est principalement dû aux néphron juxtamédullaires. L'urine contient principalement de l'urée, de l'acide nucléique et d'autres acides, ce qui en fait une solution nettement plus acide que le plasma mais aussi des molécules exogènes aux origines diverses.[196]

Les mécanismes cités ci-dessus sont des mécanismes clés du maintient de l'homéostase. La réabsorption de l'eau et des électrolytes et sa modulation par différentes hormones permet de maintenir l'osmolalité de l'organisme, c'est à dire la concentration en solutés de l'organisme. Elle est principalement régulée via le rein par des hormones telles que la vaso-

pressine et l'aldostérone. Ce mécanisme de régulation a pour principal effet de limiter la déshydratation.[196] Le rein participe également à la régulation de la pression artérielle car il contrôle le volume de plasma circulant, notamment via le système rénine-angiotensine qui régule la réabsorption du sel et donc de l'eau par osmose.[196] Enfin, le rein participe également au maintient de l'équilibre acido-basique de l'organisme d'un part en sécrétant des ions hydroniums et d'autres acides, mais aussi en recyclant les réserves de bicarbonate du corps grâce à la réabsorption. La sécrétion de bicarbonate dans le plasma est un des principaux moyens qu'a le corps de compenser la diminution du pH extracellulaire.[196] Le rein possède également des fonctions n'ayant pas de rapport avec sa capacité d'épuration du sang, comme la production d'érythropoïétine mais produit également des molécules et des hormones régulant sa propre fonction, comme la rénine.[196] L'ensemble des fonctions physiologiques du néphron est présenté dans la figure 10. La fonction du néphron est représentée par deux paramètres souvent utilisés dans les études de néphrologie, le débit de filtration glomérulaire ou taux de clairance, qui représente la quantité de néphrons capables d'exercer leur fonction de filtration. Lorsque ce débit ou ce taux diminue, cela signifie qu'une partie ou la totalité des néphrons ne filtrent pas de plasma. Cette donnée peut être mesurée à l'aide de marqueurs physiologiques ou estimée à partir de relevés sanguins.[215] Le second paramètre est la protéinurie, qui représente la quantité de protéines présentes dans l'urine. On utilise également l'albuminurie à la place, l'albumine étant la protéine la plus présente dans le plasma. La protéinurie représente l'état des néphrons encore capables de filtrer le plasma, ou plus précisément l'état de leur barrière glomérulaire. C'est un marqueur d'une maladie appelée syndrome néphrotique.[216] Ces deux paramètres sont couramment utilisés ensemble mais sont des marqueurs de maladies rénales distincts et ne sont parfois pas associés.

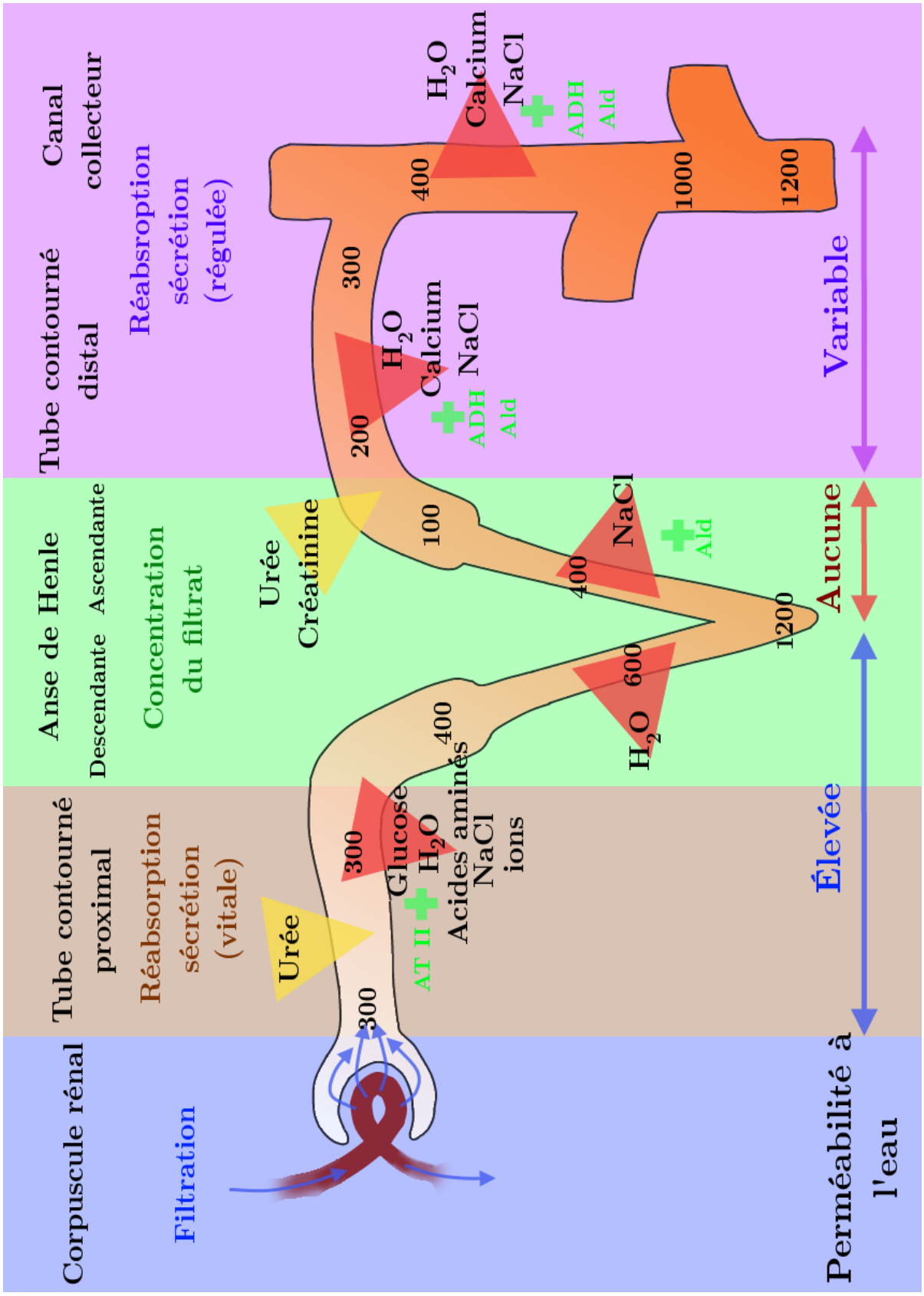


Figure 9 Synthèse des fonctions physiologiques du néphron. La filtration est effectuée par le corpuscule rénal tandis que les autres fonctions sont remplies par le tubule rénal. La réabsorption a lieu sur toute la longueur du tubule mais la majeure partie des solutés (glucose, acides aminés, sel, calcium, potassium) sont réabsorbés par le tube contourné proximal. La sécrétion active des molécules à évacuer a lieu principalement dans les tubes contournés. La réabsorption peut être partiellement ou entièrement contrôlée par différentes hormones comme l'angiotensine II (AT II), l'aldostérone (Ald) ou la vasopressine (ADH). Le chiffre apparaissant dans le tubule rénal représente la concentration en urée exprimée en milliosmole par litre (mOsm/L).
Illustration adaptée du travail de Michal Komorniczak

6.3 Rôle dans les maladies cardiovasculaires

Comme nous l'avons vu précédemment, la fonction rénale joue un rôle important dans le maintien de l'hémodynamique via plusieurs mécanismes. Le rein entretient une relation interdépendante avec le cœur et la dysfonction d'un des deux organes peut mener à la dysfonction de l'autre. Ces situations sont regroupées sous le terme de syndrome cardio-rénal, cependant il s'agit d'un terme simpliste car il regroupe des situations avec des mécanismes pathologiques très variés.[217] Il a d'abord été constaté grâce à des études épidémiologiques que les maladies rénales chroniques augmentaient le risque de maladie cardiovasculaire. C'est notamment le cas de la thrombose veineuse pour laquelle il a été montré une association entre le débit de filtration glomérulaire et le risque de thrombose veineuse ainsi que le taux circulant de certains facteurs de coagulation comme FVIII ou le facteur de von Willebrand.[218, 120] De même, plusieurs études ont relié le syndrome néphrotique au risque de thrombose veineuse,[219] les patients touchés par le syndrome néphrotique présentant aussi des taux diminués d'antithrombine.[220] Le débit de filtration glomérulaire a également été associé à un risque accru de coronaropathie ainsi qu'à des événements plus sévères.[221] Plusieurs hypothèses ont été avancées pour expliquer le lien entre la fonction rénale et le risque de maladie cardiovasculaire. Tout d'abord le rein participe à la régulation de la pression artérielle en contrôlant le volume de plasma. Cependant dans le cas des maladies rénales chroniques, le déficit de filtration débouche sur une pression artérielle élevée et la pression artérielle est notoirement associée au risque de maladie cardiovasculaire.[222, 33]

Le rein participe également à la régulation du taux d'électrolytes dans la circulation, notamment du calcium. Or en cas de maladie rénale chronique, le calcium en excès ne peut être sécrété et pourrait participer à la calcification des tissus vasculaires et des plaques d'athérome, ce qui les fragilise.[12, 223] Ce phénomène pourrait expliquer la sévérité accrue des événements coronariens chez les patients atteints de maladie rénale chronique. Le rein participe au métabolisme de l'homocystéine en sécrétant l'homocystéine en excès,[135] présente dans la circulation en excès, cette molécule dérivée d'un acide aminé possède des propriétés oxydantes et pourrait endommager les cellules endothéliales lorsqu'elle est présente en trop grande quantité.[136] Ceci pourrait avoir un effet pro-inflammatoire et fragiliser la plaque d'athérome. L'hyperhomocystéinémie est associée au risque de thrombose veineuse mais aussi au risque de maladie coronarienne et d'accident vasculaire-cérébral.[136] Dans le cas du syndrome néphrotique, la protéinurie pourrait causer la perte de protéines antithrombotiques comme l'antithrombine,[220] ce qui favorise la réaction de coagulation, le syndrome néphrotique a déjà été associé à la thrombose veineuse[219] mais pourrait également avoir un impact sur la sévérité des événements coronariens en favorisant la formation du thrombus lors de la rupture de la plaque.

Pour finir, le rein est également responsable de la production d'érythropoïétine. En cas de maladie rénale chronique, le rein n'est plus capable de produire l'hormone en quantité suffisante ce qui entraine une anémie, fréquemment observée chez les patients atteints de maladie rénale chronique.[224] Les patients en anémie présentent un risque d'événement aigu accru en cas de maladie coronarienne.[225] Malgré toutes ces hypothèses pertinentes, de nombreuses interrogations persistent sur le lien entre la fonction rénale et les maladies cardiovasculaires. Tout d'abord, bien que des études montrent un risque accru de maladie cardiovasculaire en cas de maladie rénale chronique, le lien de causalité n'est pas clair puisqu'on observe également un déclin de la fonction rénale chez les patients atteints de maladies cardiovasculaires.[226] De même, le rein est associé à un risque d'hypertension mais une pression artérielle élevée est également un facteur de risque de déclin de la fonction rénale.[227]

Dans le cadre plus particulier de la thrombose veineuse, les éléments permettant d'expliquer le lien de cause à effet ne sont pas suffisamment nombreux. En effet bien que la protéinurie soit associée à une déficience en antithrombine, on ne sait pas pour quelle raison l'antithrombine est affectée mais pas les autres facteurs de l'hémostase. De même le débit de filtration glomérulaire est associé au facteur VIII et au facteur de von Willebrand mais on ne sait pas pourquoi un déficit de filtration augmenterait la quantité de facteur VIII circulant. De même la plupart des études épidémiologiques concernant la fonction rénale et la thrombose veineuse sont des études transversales et elles ne peuvent donc pas explorer le lien de cause à effet. Le dernier point d'interrogation nous concernant est celui des facteurs de confusion. En effet il existe d'autres facteurs de risques associés aux maladies cardiovasculaires et aux maladies rénales et les études d'épidémiologies ne peuvent pas toujours tous les prendre en compte. Il est donc difficile d'exclure la responsabilité de facteurs de confusion comme le diabète, le surpoids ou même la pression artérielle. Il est donc difficile avec les donnés en notre possession d'établir des relations de cause à effet claires entre les différents événements mais nous verrons par la suite comment apporter des éléments de réponse à ces questions grâce à l'approche génétique.

7 Objectifs

J'ai présenté plus haut tous les acteurs de l'étude et il est temps de s'intéresser au sujet de mon travail. Nous avons vu que tous les acteurs présentés plus haut, à savoir, les maladies cardiovasculaires, en particulier les coronaropathies et la thrombose veineuse, le diabète de type I et la fonction rénale sont liés par des mécanismes complexes et encore confus. Mon objectif est d'essayer d'apporter un éclaircissement sur ce réseau de liens. Ce travail peut être séparé en deux parties distinctes, d'abord celle concernant les coronaropathies dans un contexte diabétique puis une partie concernant la thrombose veineuse. On y découvrira

d'abord une approche génétique classique avec pour objectif l'identification de marqueurs génétiques du risque de coronaropathie et une approche différente où je tente d'utiliser des marqueurs de la fonction rénale pour estimer le risque de thrombose veineuse en fonction de l'état de la fonction rénale.

7.1 Les coronaropathies chez les diabétiques

Le premier sujet présenté porte sur l'identification de marqueurs de risque génétique de maladie coronarienne chez les diabétiques de type I. Comme nous l'avons vu avant, les coronaropathies sont les maladies cardiovasculaires les plus fréquentes et sont responsables de plusieurs millions de décès chaque année. Le risque de développer une coronaropathie est influencé par une multitude de facteurs environnementaux et génétiques qui on déjà été présentés. Le diabète fait partie de ces facteurs de risque, à plus forte raison lorsque ce diabète est combiné à une néphropathie diabétique. Le diabète double approximativement le risque de développer une maladie coronarienne.[36, 172] Cependant, le diabète de type I est un cas particulier en ce qui concerne le développement des coronaropathies. Chez le diabétique de type I, les facteurs de risque principaux sont le manque de contrôle de la glycémie et la néphropathie diabétique.[91, 172] Les coronaropathies dans le contexte du diabète de type I semblent dépendre de mécanismes physiopathologiques distincts et on pourrait être amené à penser que des marqueurs de risque génétique spécifiques pourraient exister pour ce cas. Cependant, si des marqueurs génétiques spécifiques aux complications coronariennes du diabète de type I existent, il est peu vraisemblable qu'ils puissent être mis en évidence par une étude sur la population générale lorsqu'on considère la prévalence du diabète de type I.

Au commencement de ma thèse, aucune étude d'association pan-génomique sur les complications coronariennes du diabète de type I n'avait été menée mais des éléments suggèraient tout de même une composante génétique affectant le risque de complication macrovasculaire chez les patients diabétiques de type I.[228] De la même façon pour la néphropathie diabétique, bien qu'il existe des études montrant une ségrégation familiale de la complication suggérant une composante génétique,[186] les études d'association n'ont pu trouver que très peu de candidats. Le mécanisme derrière le lien entre la néphropathie diabétique et les complication macrovasculaires du diabète n'est pas encore bien compris, et il pourrait s'agir d'une relation purement physiologique ou de marqueurs génétiques partagés. Cependant si des marqueurs de la néphropathie diabétique existent, considérant le fort lien avec les complication macrovasculaires ces marqueurs pourraient aussi bien être associés au coronaropathies chez les diabétiques de type I.

Le but de ce premier travail est d'identifier des marqueurs génétiques du risque de maladie coronarienne chez les diabétiques de type I. J'en profite également pour tester l'association

entre les marqueurs génétiques des coronaropathies déjà identifiés dans la population générale dans le contexte particulier du diabète de type I. Pour cela, j'ai mené une étude d'association pan-génomique sur un groupe composé entièrement de patients diabétiques de type I, en séparant les individus avec et sans maladie coronarienne et en tenant compte de la présence ou de l'absence de néphropathie diabétique. Le but est de mettre en évidence des *loci* spécifiques au contexte diabétique et des *loci* permettant d'apporter une information fonctionnelle sur le lien entre la néphropathie et les complications macrovasculaires du diabète.

7.2 Relation entre la fonction rénale et la thrombose veineuse

Le second sujet présenté est une étude de l'impact de la fonction rénale sur le risque de développer une thrombose veineuse. La thrombose veineuse a un impact sociétal plus anecdotique comparée aux coronaropathies mais les données épidémiologiques présentées plus tôt montrent tout de même qu'il s'agit d'une maladie fréquente. Le risque de développer la maladie est dépendant de plusieurs facteurs environnementaux déjà présentés mais aussi de facteurs génétiques ayant des effets plus ou moins importants entrainant souvent un dysfonctionnement dans la cascade de coagulation.[100, 81]

Dans cette étude, je m'intéresse en particulier à un facteur de risque qui est la fonction rénale qu'un faisceau de preuves relie au risque de thrombose veineuse.[219, 218] Pourtant, le lien entre la fonction rénale et le risque de thrombose veineuse n'est pas clairement élucidé. En effet, comme nous l'avons vu plusieurs études épidémiologiques font état d'une association entre le débit de filtration glomérulaire et le risque de thrombose veineuse.[218] Cependant le lien de causalité n'a pas été examiné à ce jour et ces études ne peuvent pas prendre en compte le rôle de certains facteurs de confusion comme le statut diabétique, l'obésité ou la pression artérielle. Afin de mieux aborder la question du lien de causalité et le problème posé par les facteurs de confusion j'ai choisi d'étudier la relation entre la fonction rénale et la thrombose veineuse par le biais de marqueurs génétiques de la fonction rénale. Ces marqueurs ont été sélectionnés à partir des résultats de plusieurs études d'association pan-génomiques sur le débit de filtration glomérulaire réalisées dans différentes cohortes.[229, 230, 231, 232] Les marqueurs sélectionnés ont ensuite été rassemblés en un score de risque polygénique qui a été testé en association avec la thrombose veineuse dans deux cohortes indépendantes. Le but de cette approche est avant tout d'utiliser les marqueurs génétiques comme point de référence du lien de causalité car ils sont attribués à la naissance et permettre de fixer la direction du lien de causalité. De plus l'utilisation de marqueurs génétiques permet également d'atténuer l'effet de certains facteurs de confusion car si on considère que les marqueurs utilisés n'ont d'effet que sur le débit de filtration glomérulaire alors leur distribution est indépendante des facteurs de confusion affectant ce phénotype ou le risque de thrombose veineuse et donc que l'étude de

l'association entre ces marqueurs et la thrombose veineuse est libre de toute influence par ces facteurs de confusion. Pour renforcer ces résultats j'ai également confirmé l'association entre les marqueurs et le débit de filtration glomérulaire dans deux autres cohortes indépendantes et enfin, j'ai utilisé le score afin de tester plusieurs hypothèses reliant la fonction rénale à la quantité de certains facteurs de l'hémostase dans la circulation, notamment l'antithrombine, le facteur VIII et le facteur de von Willebrand, qui avaient déjà été mis en avant par d'autres études.

CHAPITRE 2

Matériel & méthodes

Les stratégies de recherche présentées plus haut nécessitent des méthodes appropriées. Ces méthodes sont décrites dans la partie qui va suivre. Premièrement, on cherche à déterminer s'il existe un profil génétique spécifique au contexte diabétique du risque de développer une maladie cardiovasculaire. Ici, le choix de l'approche génétique prend tout son sens, la génétique étant la science des gènes, de leurs variations et de l'hérédité des caractères. Deuxièmement, nous cherchons à exposer le rôle des complications rénales dans le développement des maladies cardiovasculaires, et la génétique est également pertinente dans ce cas là puisque le but est d'identifier des marqueurs de la fonction rénale en association avec les maladies cardiovasculaires. Pour ma part, j'estime que la génétique est une discipline essentielle pour l'étude des pathologies complexes qui est très complémentaire avec les approches d'épidémiologie classique et de biologie moléculaire et cellulaire. La génétique s'enrichit des résultats et des méthodes de l'épidémiologie classique et permet d'apporter des éléments de réponse sur l'héritabilité et la répartition des pathologies au sein de la population.[1] La génétique s'intègre également très bien avec la biologie cellulaire et moléculaire pour élucider le mécanisme d'action derrière la pathologie. L'approche employée dans ce manuscrit est une approche centrée sur l'étude de la génétique dans les populations, plus particulièrement l'étude de la distribution des variations génétiques au sein de la population et leur effet sur le développement de maladies, c'est le domaine de l'épidémiologie génétique. Dans cette partie, l'objectif est de présenter la discipline, expliquer son intérêt et présenter les outils et les méthodes mises en œuvre dans les travaux qui seront présentés par la suite.

1 L'approche épidémiologique

La génétique est une science jeune, qui a fait son apparition au début du 20ème siècle avec la redécouverte des travaux de Gregor Mendel, le terme de "génétique" est cependant apparu au 19ème siècle et a été attribué à Imre Festetics.[233] Cependant la démarche initiale de la génétique classique était une approche fonctionnelle, centrée sur l'étude des mécanismes de l'hérédité. Ce n'est que vers les années 70, avec le développement du séquençage Sanger et à la découverte des polymorphismes génétiques que la génétique va faire sa ren-

contre avec l'épidémiologie. Le développement des puces à ADN et des technologies haut-débit entre les années 90 et 2000 ont permis l'étude de l'ensemble du génome dans de grands échantillons.[234, 235] L'épidémiologie génétique a dans un premier temps traité de l'héritabilité des caractères, avec des études de liaison génétique ainsi que des études de ségrégation qui ont précédé les grandes études d'association. L'épidémiologie est la science de la distribution des pathologies au sein des populations et de l'évaluation des facteurs de risque. C'est une discipline clé dans le domaine de la santé publique qui emploie principalement des outils statistiques. L'épidémiologie génétique est l'étude de la distribution des caractères génétiques au sein des grandes populations et de l'estimation du risque associé à ces caractères, les méthodes d'analyse restent cependant assez proches de celles de l'épidémiologie classique. Les études d'épidémiologie génétique peuvent être organisées de nombreuses manières différentes, cependant dans ce manuscrit nous ne nous intéresseront qu'aux études d'association. L'épidémiologie génétique s'attache à répondre à trois types de questions :

1 Établir l'existence d'une composante génétique contribuant à une maladie.

2 Estimer l'effet de cette composante génétique sur le risque de développer la maladie.

3 Identifier les gènes impliqués dans cette composante génétique.

Les études d'association apportent des éléments de réponse aux trois questions posées mais sont plus pertinentes pour répondre aux deux dernières questions. Depuis qu'il est possible d'observer des marqueurs génétiques répartis sur tout le génome dans de grandes populations, les études d'association génétiques les plus fréquemment rencontrées sont les études d'association pan-génomiques (GWAS) dans lesquelles des milliers de marqueurs génétiques, en général des polymorphismes d'un seul nucléotide (SNPs pour *Single Nucleotide Polymorphism*), sont testés en association avec un trait biologique ou une maladie. L'approche la plus souvent utilisée pour ces études est l'approche cas-témoin. Nous verrons cependant qu'il est nécessaire de suivre plusieurs étapes de contrôle de qualité à la fois dans la composition de l'échantillon, la sélection des marqueurs génétiques et l'analyse statistique, ces étapes seront décrites plus loin dans cette partie.

1.1 Les grandes études d'association

Les études d'association pan-génomiques sont des études d'association réalisées sur des échantillons de grande taille et qui étudient un grand nombre de variations génétiques réparties sur tout le génome. Ce type d'analyse a été rendu possible grâce au développement des biopuces et d'ordinateurs capables d'effectuer de plus en plus d'opérations rapidement, l'outil informatique est d'ailleurs absolument nécessaire pour réaliser une GWAS. La première

GWAS publiée date de 2005[236] mais depuis des milliers de GWASes ont été réalisées. Ces études sont en général des études de type cas-témoins ou des études de cohortes portant sur une maladie ou un caractère biologique, elles peuvent être transversales ou longitudinales. Il s'agit d'un approche naïve opposée aux études basées sur des gènes candidats puisqu'il n'y a pas de sélection des variations basée sur la fonction des gènes proches, cette approche permet d'être particulièrement exhaustif dans l'identification des variations associées à la condition étudiée.[237] De plus, la grande taille des échantillon permet d'avoir une forte puissance statistique et d'identifier un grand nombre de variations même lorsqu'elles ont des effets modestes.[238] Tout ceci rend les GWASes particulièrement utiles pour l'étude de la génétique des maladies complexes où plusieurs variations avec des effets faibles contribuent au risque. Les SNPs utilisés dans les GWASes sont généralement obtenus par un génotypage à l'aide d'une biopuce à ADN suivi par une imputation des SNPs absents sur la biopuce. Les SNPs choisis pour l'étude sont issus de biobanques telles que HapMap[239] et sont choisis de manière à couvrir l'ensemble du génome. Avant l'analyse, on choisi pour chaque SNP quel sera l'allèle testé et on détermine sous quel modèle l'analyse sera effectuée. On choisi alors entre modèle additif, modèle dominant ou modèle récessif. Dans le modèle additif, on compte la somme des allèles à tester tandis que dans un modèle dominant on compare les individus ayant au moins un allèle à tester aux autres, à l'inverse dans un modèle récessif on compare les individus ayant deux allèles à tester aux autres.

Dans les GWASes cas-témoins on recherche les SNPs dont les fréquences alléliques sont différentes entre les cas et les témoins. L'ampleur de l'effet est souvent rapportée sous la forme d'un *odds ratio* tandis que la force du signal est indiquée par une *p-value*.[237] La GWAS possède cependant quelques limites. D'abord la GWAS ne permet pas d'identifier directement les SNPs causaux mais des SNPs situés à proximité de la vraie variation causale.[238] Ensuite, l'analyse individuelle de chaque SNP ne permet pas d'identifier les phénomènes d'épistasie,[240] c'est à dire les interactions entre plusieurs variations qui modifient l'effet individuel de chacune d'entre elles. La grande taille des échantillons augmente le risque d'avoir des échantillons peu homogènes et il est alors nécessaire de contrôler la présence de sous-structures dans la population d'étude.[241] L'hétérogénéité de l'échantillon est susceptible de créer des associations fausses entre les SNPs testés et les phénotypes étudiés.[237] Enfin, le grand nombre de SNPs testés fait que la proportion de faux-positifs est élevée. Il est donc nécessaire de prendre en compte l'effet des tests multiples.[237] Malgré tout cela la GWAS est un outil puissant pour identifier des gènes impliqués dans la composante génétique d'une pathologie et constitue donc une approche adaptée à nos objectifs.

1.2 Les groupes d'étude

Il existe plusieurs méthodes pour assembler un groupe d'étude. Cependant dans ces travaux je me suis limité au design le plus fréquent, l'étude cas-témoin. Il s'agit ici de comparer un groupe de cas, des individus étant atteints de la maladie ou du trait étudié à un groupe témoin qui peut soit être composé d'individus tirés aléatoirement dans la population étudiée ou d'individus non affectés par la maladie ou le trait. Lorsque le trait ou la maladie est commune, il est plus pertinent d'utiliser un groupe d'individus sans la maladie en tant que groupe témoin afin de gagner en puissance statistique.[242] Il est important que le groupe témoin et le groupe cas soient issus de la même population et qu'ils aient si possible des caractéristiques similaires en dehors de la maladie étudiée afin de limiter au maximum la variabilité intergroupe. Il n'y a pas nécessairement de rapport numérique témoin/cas à respecter mais si le nombre de cas est faible il peut être utile d'augmenter le nombre de témoins afin d'augmenter la puissance de l'étude. Cependant certains travaux indiquent que dépasser un rapport de 4 témoins par cas serait contre-productif.[243]

Les groupes d'études présentés ici sont les cohortes qui ont été utilisées pour composer les groupes cas-témoins utilisés pour mes travaux. Tous ces groupes sont issus de populations d'origine Européenne, venant soit du continent Européen (France, Pays-bas, Danemark, îles Britanniques, Allemagne) ou d'Amérique du Nord (Canada, États-Unis). Les individus pour lesquels le sexe n'était pas renseigné ou pour lesquels le sexe renseigné ne correspondait pas au sexe déduit des marqueurs génétiques ont été exclus. Pour les groupes d'individus avec un score de parenté trop élevé, un seul membre a été retenu pour les analyses, les cohortes sont regroupées en fonction du projet pour lequel elles ont été employées. Tous les patients recrutés ont émis leur consentement informé par écrit et les différents protocoles de recrutement ont été approuvés par les comités d'éthique des institutions participantes.

Travail sur les marqueurs de risque génétique des coronaropathies chez les diabétiques de type I

Les cohortes suivantes ont été employées dans le projet d'étude d'association pan-génomique sur les coronaropathies chez les diabétiques de type I avec ou sans néphropathie diabétique. Les trois premières cohortes, Genesis/Genediab, Steno et UK-ROI ont été employées dans la phase de découverte tandis que les cohortes DCCT/EDIC, WESDR, CACTI, EDC et Go-KinD ont été employées durant la phase de réplication.

Genesis/Genediab

Cette cohorte comporte 1370 patients français ayant un diabète de type 1 depuis au moins 15

ans au moment de l'inclusion. Le diabète de type 1 est défini par un diagnostic avant 35 ans et une prise d'insuline définitive dans l'année suivant le premier diagnostic. La néphropathie diabétique a été définie par un taux d'excrétion d'albumine supérieur à 300mg par 24 heures sur deux mesures parmi trois mesures consécutives d'un échantillon d'urine stérile. Le groupe inclut également 85 cas de coronaropathie, définis par un ou plusieurs événements d'infarctus du myocarde renseignés ou une revascularisation avec ou sans néphropathie. Les 1285 individus sans antécédents de coronaropathie sont définis comme témoins. Les caractéristiques du groupe sont présentées dans le tableau 2A.[244]

Steno

Steno est une cohorte de 868 patients danois diabétiques de type 1 recrutés au *Steno Diabetes center* à Copenhague au Danemark. Le diabète a été défini comme de type 1 par un diagnostic avant 35 ans et le démarrage d'une thérapie à l'insuline définitive moins d'un an après le premier diagnostic. La néphropathie diabétique a été définie par une albuminurie persistante (supérieure à 300mg par 24 heures) sur deux mesures parmi trois mesures consécutives, la présence d'une rétinopathie et l'absence d'autres maladies rénales ou du canal urinaire. L'absence de néphropathie diabétique a été définie par une normoalbuminurie persistante (inférieure à 300mg par 24 heures) après 15 ans de diabète, sans traitement avec des inhibiteurs de l'ECA ou d'antagonistes des récepteurs de l'angiotensine II. Les individus avec des antécédents de coronaropathies confirmés cliniquement au recrutement ont été classés en cas tandis que les individus sans antécédents de maladies cardiovasculaire en général ont été classés comme témoins. Les individus avec des antécédents de maladies cardiovasculaires autres que des coronaropathies ont été exclus. Les caractéristiques de la cohortes sont présentées dans le tableau 2A.[245]

Ireland, Warren 3 and Genetics of kidneys in Diabetes UK collection

Ireland, Warren 3 and Genetics of kidneys in Diabetes UK collection (UK-ROI) est une cohorte de 1318 patients européens diabétiques de type 1 recrutés dans différents centres des îles Britanniques. Tous les patients ont été diagnostiqués avec un diabète de type 1 avant leurs 31 ans et 112 d'entre eux ont également une maladie coronarienne caractérisée par la présence d'une angine de poitrine, un événement d'infarctus du myocarde ou un pontage aorto-coronarien, les 1206 individus restants ont un diabète de type 1 sans antécédents de coronaropathie. Les caractéristiques des patients d'UK-ROI sont présentées dans le tableau 2A.[245]

	Genesis/Genediab			Steno			UK-ROI		
	Cas	Témoins	*P-value*	Cas	Témoins	*P-value*	Cas	Témoins	*P-value*
	N=85	N=1285		N=237	N=632		N=112	N=1206	
Sexe (% femmes)	41,2	47,8	0,28	40,1	44,1	0,28	46,4	49,4	0,50
Âge moyen en années (σ)	51 (10)	40 (12)	$3{,}04\text{x}10^{-16}$	48,1 (10,9)	42,3 (10,7)	$5{,}98\text{x}10^{-12}$	44,09 (11,50)	44,12 (11,12)	0,98
Durée moyenne du diabète en années (σ)	32 (11)	23 (11)	$2{,}55\text{x}10^{-12}$	32,1 (9,8)	26,6 (8,8)	$9{,}42\text{x}10^{-15}$	30,28 (10,08)	30,45 (9,69)	0,90
IMC moyen kg/m^2 (σ)	24,5 (3,6)	24,1 (3,5)	0,29	24,1 (3,3)	24,2 (3,1)	0,51	25,86 (11,50)	26,19 (4,44)	0,50
Fumeurs (%)	43	44	0,96	48,6	40,9	0,049	18,8	17,1	0,60
% HbA1c moyen (σ)	8,4 (1,3)	8,4 (1,7)	0,87	9,3 (1,4)	8,6 (1,3)	$1{,}7\text{x}10^{-10}$	9,10 (1,96)	8,60 (1,60)	0,01
eGFR moyen mg/mL (σ)	66,4 (29,1)	88,3 (29,1)	$2{,}7\text{x}10^{-9}$	62,0 (27,3)	82,2 (21,4)	$<2\text{x}10^{-16}$	43,11 (28,32)	60,79 (27,64)	<0,01
Néphropathie diabétique (%)	47	25	$1{,}01\text{x}10^{-4}$	75	40	$<2\text{x}10^{-16}$	87	43	$<2\text{x}10^{-16}$
AER* moyen mg/24h (σ)	234,67 (433,26)	133,71 (349,59)	0,046	NA	NA	NA	NA	NA	NA

Tableau 2A Caractéristiques générales des études *Genesis/Genediab*, *Steno* et *UK-ROI*. Les *p-values* correspondent aux *p-values* d'un test de Student ou d'un test de χ^2 de comparaison entre les cas et les témoins en fonction de la caractéristique examinée.
IMC = indice de masse corporelle.
HbA1c = pourcentage d'hémoglobine glyquée, voir plus bas.
eGFR = taux de filtration glomérulaire estimé, voir plus bas.
*Taux d'albumine excrétée, voir plus bas.

Diabetes Control and Complications Trial/Epidemiology of Diabetes Interventions and Complications

Diabetes Control and Complications Trial/Epidemiology of Diabetes Interventions and Complications (DCCT/EDIC) est une cohorte qui regroupe 1304 participants âgés de 13 à 39 ans recrutés entre 1983 et 1989 dans 21 cliniques réparties dans plusieurs états des États-Unis (Ohio, Pennsylvanie, New York, Michigan, Massachusetts, South Carolina, Minnesota, Iowa, Tennessee, Texas, Washington, Missouri et Connecticut) ainsi qu'en Ontario, au Canada. Le recrutement des patients a été effectué en deux phases. Les patients recrutés lors de la première phase avaient un diabète de type 1 déclaré depuis au moins 1 an et au maximum depuis 5 ans et ne présentaient pas de signe de rétinopathie. Les patients recrutés dans la deuxième phase avaient un diabète de type 1 déclaré depuis au moins 1 an et jusqu'à 15 ans au maximum et présentaient des signes de rétinopathie non-proliférative légère. Les détails de cette cohorte sont présentés dans le tableau 2B.[246]

Wisconsin Epidemiology Study of Diabetic Retinopathy

Wisconsin Epidemiology Study of Diabetic Retinopathy (WESDR) regroupe 594 patients atteints de diabète de type 1 recrutés dans le Wisconsin entre 1979 et 1980. Tous les patients ont été diagnostiqués avant leurs 30 ans et suivent une thérapie à l'insuline. Les détails de WESDR sont présentés dans le tableau 2B.[247]

Coronary Artery Calcificaltion in Type 1 Diabetes

Coronary Artery Calcificaltion in Type 1 Diabetes (CACTI) regroupe 520 patients diabétiques de type 1 recrutés entre 2000 et 2002 dans le Colorado. Tous les patients étaient âgés de 19 à 56 ans et n'avaient pas d'antécédents de maladies cardiovasculaires au moment du recrutement. Les détails de CACTI sont présentés dans le tableau 2B.[248]

Pittsburgh Epidemiology of Diabetic Complications Study

Pittsburgh Epidemiology of Diabetic Complications Study (EDC) regroupe 428 patients ayant été diagnostiqués avec un diabète de type 1 avant 17 ans, recrutés entre 1986 et 1988 à l'hôpital pédiatrique de Pittsburgh, en Pennsylvanie. Tous les patients ont été suivis bi-annuellement pendant 25 ans, les détails sont disponibles dans le tableau 2B.[249]

Genetics of Kidneys in Diabetes US Study

La cohorte *Genetics of Kidneys in Diabetes US Study* (GoKinD US) est composée de 351 patients diabétiques de type 1 atteints ou non de néphropathie diabétique, ayant été diagnos-

tiqués pour un diabète de type 1 avant 31 ans et ayant commencé une thérapie à l'insuline un an après le diagnostic. La néphropathie diabétique est définie comme la présence d'une insuffisance rénale chronique terminale, la nécessité de dialyse ou d'une transplantation rénale et une albuminurie persistante sur deux mesures parmi trois mesures consécutive(rapport albumine/créatinine d'au moins 300mg/g ou une mesure de $\geq$ 1+ par bande urinaire). Les participants sont tous d'origine caucasienne et ont été recrutés entre 2000 et 2003 dans le centre de l'université George Washington et le *Joslin Diabetes Center* aux États-Unis. Les caractéristiques de GoKinD sont disponibles dans le tableau 2B.[245]

Bien que Genesis/Genediab, UK-ROI, Steno, DCCT/EDIC, WESDR, CACTI, EDC et GoKinD US soient des cohortes de patients diabétiques de type 1, elles ont été utilisées dans ces travaux comme des groupes cas/témoins pour les coronaropathies, le groupe cas étant les individus diabétiques avec des antécédents de coronaropathie (infarctus du myocarde, angioplastie des artères coronaires, angine de poitrine...) tandis que le groupe témoin était composé de patients diabétiques sans antécédents de coronaropathies.

Mesures à l'inclusion	DCCT/EDIC N=1304	WESDR N=594	CACTI N=520	EDC N=428	GoKinD N=351
Dates de recrutement (années)	1983-1989	1979-1980	2000-2002	1986-1988	2000-2003
Sexe (% femmes)	47	50	44	48	66
Âge au recrutement* (années)	27 (7)†	26 (11)†	36 (9)†	27 (8)†	37 (9)†
Durée du diabète* (années)	6 (4)†	12 (9)	23 (9)†	19 (8)†	24 (7)
Durée du suivi (années)	25 (5)	37 (9)	11 (5)	39 (8)	NA
CVD**	166 (13)	209 (35)	49 (10)	158 (37)	50 (14)
CAD**	152 (12)	202 (34)	42 (8)	139 (32)	50 (14)

Tableau 2B Caractéristiques générales des études DCCT/EDIC, WESDR, CACTI, EDC et GoKinD.
CVD = *cardiovascular disease.*
CAD = *coronary artery disease.*
*moyenne (écart-type).
** Nombre d'évènements (%).
†*P-value* d'association ≤ 0.05 avec CVD dans un modèle logistique.

Travail sur la fonction rénale et la thrombose veineuse

Les cohortes suivantes ont été employées dans l'étude d'estimation de l'impact de la fonction rénale sur le risque de thrombose veineuse à l'aide de marqueurs génétiques.

Marseille Thrombosis Association

La cohorte *MARseille THRombosis Association* (MARTHA) est une cohorte de 1592 patients non apparentés ayant été recrutés au centre d'hématologie du centre hospitalier de la Timone à Marseille. Ces patients ont tous été recrutés entre janvier 1994 et octobre 2005 après un événement de thrombose veineuse vérifié par veinographie, ultrason Doppler, angiographie et/ou scintigraphie pulmonaire. Tous ces patients étaient dépourvus de toute condition chronique et sans aucun facteur de risque génétique de thrombose connu (mutation antithrombine, déficience en protéine S ou protéine C, porteur homozygote de la mutation FV Leiden ou de la mutation FII 20210A et lupus anticoagulant). Les patients ont été séparés en deux groupes MARTHA08 et MARTHA10, qui ont été génotypés sur deux puces différentes. Les caractéristiques des patients de MARTHA sont présentées dans le tableau 2C.[250]

	MARTHA N=1542
Âge moyen (σ)	46,99 (15,27)
Sexe (% femmes)	66,1
IMC moyen, kg/m^2 (σ)	25,14 (4,61)
Fumeurs (%)	24,1
Buveurs d'alcool (%)	0,58
Porteurs hétérozygotes de FV (%)	21,98
Porteurs hétérozygotes de FII (%)	13,94
Groupe ABO (%)	
A1	55
A1B	4
A2	6
B	12
O	23
aPTT* (σ)	31,40 (3,33)
Intervale	21,2-43,3
FVIIIc (σ)	134,44 (51,18)
Intervale	43,0-434,0
vWF (σ)	152,50 (66,56)
Intervale	41,0-506,0

Tableau 2C Caractéristiques générales de l'étude MARTHA.

*Temps de céphaline activée, exprimé en secondes, voir plus bas.

Early Onset Venous Thrombosis

Early Onset Venous Thrombosis (EOVT) est une cohorte de 453 individus ayant eu un premier événement de thrombose veineuse profonde ou d'embolie pulmonaire avant 50 ans, ces patients ont été recrutés dans quatre centres différents (Grenoble, Marseille, Montpellier et Paris) et leur statut a été vérifié par veinographie, ultrasons Doppler, angiographie et/ou scintigraphie pulmonaire. Les patients présentant des risques connus environnementaux (hospitalisation avec immobilisation, traumatisme, chirurgie, grossesse, post-partum, prise de contraceptifs oraux, cancer, maladie autoimmune) ou génétiques (anti-thrombine, déficiences en protéine S ou en protéine C, homozygote FV Leiden ou FII 20210A) n'ont pas été inclus. Les caractéristiques des individus de EOVT sont détaillées dans le tableau 2D.[251]

	EOVT N=453
Âge moyen en années(σ)	36 (9)
Sexe (hommes/femmes)	250/203
Porteurs hétérozygotes de FV (%)	18,3
Porteurs hétérozygotes de FII (%)	7,1

Tableau 2D Caractéristiques générales de l'étude EOVT.

Multiple Environmental and Genetic Assessment of risk factors for venous thrombosis

Multiple Environmental and Genetic Assessment of risk factors for venous thrombosis (MEGA) est une cohorte composée de 3220 patients Néerlandais entre 18 et 70 ans ayant eu un événement de thrombose veineuse profonde ou une embolie pulmonaire. Les participants ont été recrutés dans 6 centres hospitaliers différents des Pays-Bas et les patients présentant des troubles psychiatriques ou ne parlant pas le Néerlandais n'ont pas été inclus. Les participants ont répondu à un questionnaire détaillant le diagnostic d'un cancer, et le cas échéant, la date, le type de traitement administré et la présence ou non de métastases connues. Les participants ont également été testés pour la présence de mutations FII 20210A et FV Leiden puis les individus homozygotes pour au moins une de ces deux variations ont été exclus. Les caractéristiques de MEGA au recrutement sont détaillées dans le tableau 2E.[252, 105]

	MEGA N=3220
Âge médian en années (5ème - 95ème percentiles)	49,7 (25,9-67,8)
Sexe (% femmes)	54
IMC médian, kg/m^2 (5ème - 95ème percentiles)	26,2 (20,1-35,5)
Immobilisation* (%)	38,6
FV Leiden (%)	15,8
groupe sanguin non-O (%)	71,2
FVIII médian en IU/dL (5ème - 95ème percentiles)	154 (83-280)
Facteur IX médian en IU/dL (5ème - 95ème percentiles)	108 (80-144)
vWF** médian en IU/dL (5ème - 95ème percentiles)	138 (75-255)
Thrombose veineuse profonde seule (%)	57,5
Embolie pulmonaire (%)	42,5

Tableau 2E Caractéristiques générales de l'étude MEGA.
*Alité pendant au moins 4 jours, chirurgie ou hospitalisation dans l'année précédant l'inclusion.
**Facteur de von Willebrand.

Trois Cités

Les Trois Cités (3C) est une cohorte initialement conçue pour étudier le lien entre les maladies vasculaires et la démence. Les participants sont Français et ont été recrutés dans les villes de Bordeaux, Montpellier et Dijon.[253] Un sous-échantillon de 2159 individus dépourvus de toute condition chronique ont été sélectionnés et ont été répartis en deux groupes de 1110 et 1049 individus afin de servir de groupes témoins pour MARTHA et MEGA. Dans le cas de MEGA, il n'y avait pas d'échantillon de sujets sains génotypés disponibles pour servir de groupe témoin, le choix d'utiliser 3C a donc été fait car il s'agit d'un groupe d'individus sains homogène d'origine Européenne, donc suffisamment proches des individus de la cohorte MEGA pour notre étude.

Suvimax

La cohorte SU.VI.MAX (Supplementation en Vitamines et Mineraux Antioxydants) est une cohorte de 1327 individus en bonne santé et d'origine Européenne, initialement employée pour des essais cliniques randomisés sur les effets des vitamines antioxydantes. Les patients ont été recrutés en 1994 et n'ont pas d'antécédents de maladies cardiovasculaires.[254] Un sous-échantillon de 1228 individus a été utilisé comme groupe témoin pour l'étude EOVT.

Autres cohortes

La cohorte présentée ci-dessous a été employée pour effectuer des analyses complémentaires pour les deux projets présentés plus tôt mais aussi pour des analyses secondaires sur des projets indirectement liés à ma thèse.

AtheroGene

AtheroGene est une cohorte Allemande composée de 1762 participants recrutés entre 1996 et 2004 dans le département de cardiologie du centre médical universitaire de Johannes Gutenberg-University à Mainz et le département de médecine interne des forces armées fédérales de Koblenz, en Allemagne. Les participants inclus présentaient tous des signes indicateurs d'une maladie coronarienne comme de la dyspnée, des douleurs dans la poitrine et des altérations de l'électrocardiogramme. La maladie coronarienne a été établie pour les patients présentant une sténose avec une réduction de plus de 30% de l'espace vasculaire dans au moins une des artères coronaires majeures.[255]

1.3 Mesurer les variables biologiques

Les grandes études d'association génétiques nécessitent la mesure de variables biologiques qui vont être testées ou être utilisées en tant que variables explicatives. Ces variables peuvent être mesurées à partir de prélèvements sanguins ou d'échantillons urinaires ou bien être obtenues par des mesures de paramètres physiologiques ou morphologiques. Tous les paramètres biologiques utilisés dans les travaux présentés sont listés dans cette section avec les méthodes de mesures détaillées.

L'indice de masse corporelle

L'indice de masse corporelle (IMC) a été calculé conformément à la formule établie par

Adolphe Quetelet[256] qui est la suivante :

$$IMC = \frac{masse}{taille^2}$$

Avec la masse exprimée en kilogrammes et la taille en mètres. L'IMC est utilisé pour établir l'obésité de la manière suggérée par l'organisation mondiale de la santé[257] :

IMC<18.5 Maigreur.

18.5≤IMC<25 Corpulence normale.

25≤IMC<30 Surpoids.

30≤IMC<35 Obésité modérée.

35≤IMC<40 Obésité sévère.

40≤IMC Obésité morbide.

Bien que l'IMC ne soit pas le marqueur d'obésité le plus performant[258], il reste un marqueur efficace de mortalité et des morbidités associées à l'obésité.[259, 260, 261]

Plusieurs métabolites, protéines et marqueurs physiologiques ont été mesurés à partir de prélèvements de sang périphérique et d'urine pour les différentes études présentées dans ce manuscrit. Les différentes méthodes de dosage sont présentées ci-dessous.

Albuminurie

L'albumine est une protéine produite par le foie et qui est la protéine la plus présente dans la circulation sanguine.[262] La présence d'albumine dans l'urine est un marqueur de maladie rénale chronique. L'albuminurie s'exprime sous la forme d'un taux d'albumine excrétée par jour (AER en mg par 24h) ou en rapport albumine excrétée par créatinine excrétée (ACR en mg albumine par g créatinine). La présence d'albumine dans l'urine est dosée par colorimétrie à l'aide de bleu de bromophénol, la gravité de l'albuminurie est divisée en trois catégories suivant la quantité d'albumine présente dans l'urine[263].

A1 (albuminurie normale ou légère) ACR<30mg/g.

A2 (albuminurie modérée) 30mg/g≤ACR<300mg/g.

A3 (albuminurie sévère) ACR≥300mg/g.

Créatinine

La créatinine est un métabolite libéré par les muscles dans la circulation et excrété dans l'urine. La créatinine sérique est mesurée par colorimétrie. Le taux de créatinine sérique entre en jeu dans l'estimation taux de filtration glomérulaire (eGFR), utilisé comme marqueur de la fonction rénale.[264]

Débit de filtration glomérulaire estimé et maladie rénale chronique

Le débit de filtration glomérulaire estimé (eGFR) est un marqueur de la fonction rénale. Il est calculé à partir des taux sanguins de cystatine C ou de créatinine. Il existe plusieurs équations permettant d'obtenir l'eGFR et celles-ci prennent en compte plusieurs paramètres tels que le sexe, l'âge ou l'ethnie des individus. Pour les travaux présentés, deux équations ont été employées.

La première est la formule MDRD (*Modification of Diet in Renal Disease*)[265] qui nécessite la créatinine sérique en mg/dL, l'âge, le sexe et l'ethnie et dont la formule est la suivante :

$$eGFR = 186 * SCr^{-1,154} * \hat{\text{A}}\text{ge}^{-0,203} * [1,210 \text{ si origine africaine}] * [0,742 \text{ si femme}]$$

La seconde formule est l'équation CKD-EPI (*Chronic Kidney Disease Epidemiology Collaboration*)[266], elle nécessite le sexe, l'âge et la créatinine sérique. La formule est la suivante :

$$eGFR = 141 * \max(\frac{SCr}{k}, 1)^{a} * \max(\frac{SCr}{k}, 1)^{-1,209} * [1,159 \text{ si origine afircaine}] * [1,018 \text{ si femme}]$$

$$k \begin{cases} 0,7 \text{ si femme} \\ 0,9 \text{ si homme} \end{cases} \quad a \begin{cases} -0,329 \text{ si femme} \\ -0,411 \text{ si homme} \end{cases}$$

Le terme SCr représente la quantité de créatinine dans le plasma (exprimée mg par dL). L'eGFR est exprimé en mL/min/1,73m^2. L'eGFR a été utilisé afin d'établir l'existence ou non de maladie rénale chronique. La maladie rénale chronique a été définie comme ayant plusieurs paliers qui ont été définis de la manière suivante :

Niveau de la maladie rénale chronique	eGFR (mL/min/1.73m^2)
niveau 0	$\geq$90
niveau 1	60-89
niveau 2	30-59
niveau 3	15-29
niveau 4	<15

Le niveau 0 correspond à une absence de maladie rénale chronique, le niveau 1 correspond à une maladie rénale chronique légère, le niveau 2 à une maladie rénale chronique modérée, le niveau 3 à une maladie rénale chronique sévère et le niveau 4 à une insuffisance rénale chronique terminale (IRCT).

Hémoglobine glyquée

L'hémoglobine glyquée (HbA1c) correspond à la forme glyquée de l'hémoglobine et est utilisée comme marqueur de la glycémie. L'HbA1c est mesurée par chromatographie (HPLC), cette mesure permet de déduire la concentration plasmatique de glucose.[267]

Tests de coagulation

Il existe de nombreux tests visant à détecter des anomalies de la voie de coagulation, pour les différents travaux présentés ici deux d'entre eux ont été employés. Le test de temps de céphaline activée, également appelée aPTT. Il s'agit d'un test réalisé à partir d'un prélèvement sanguin servant a détecter des anomalies dans la voie intrinsèque de la coagulation. L'échantillon de sang est recalcifié puis une molécule activatrice est ajoutée afin de démarrer la réaction de coagulation. Le temps de formation du thrombus est mesuré et exprimé en secondes. L'aPTT peut également être exprimée sous la forme d'un rapport entre le temps d'activation du sujet et un temps d'activation étalon servant de témoin. Le temps de génération de thrombine est un test visant à observer la cinétique de la formation de thrombine après activation de la voie de coagulation à l'aide d'un mélange de TF, de phospholipides et de chlorure de calcium. Le temps de génération de thrombine est mesuré à l'aide d'un *CAT assay* qui utilise une molécule fluorogène réactive à la thrombine. Le *CAT assay* permet d'obtenir des informations comme le temps de latence, qui correspond au temps écoulé avant que la thrombine ne commence à être générée, le temps de pic qui correspond au temps écoulé avant d'atteindre la quantité maximum de thrombine activée, la hauteur de pic qui est la quantité de thrombine activée et enfin le potentiel de thrombine endogène (ETP) qui correspond à la quantité thrombine

pouvant être générée.[268]

Énumération cellulaire

Le comptage cellulaire est utilisé pour estimer la quantité de globules blancs et d'érythrocytes présents dans le sang. Des méthodes similaires sont également utilisées pour quantifier les plaquettes. Le comptage de cellules peut se faire manuellement à l'aide d'un hémocytomètre, dans lequel du sang est placé dans une chambre à la surface quadrillée. On peut alors compter le nombre de cellules dans plusieurs segments et en déduire la concentration en cellules. Le comptage peut être réalisé de manière automatique à l'aide d'un cytomètre en flux qui fait passer les cellules individuellement dans un canal et les compte. Le nombre de cellules est généralement exprimé en nombre de cellules par mm^3 de sang.[269]

Mesure des protéines plasmatiques

Dans les deux études présentées, en particulier celle portant sur la thrombose veineuse, plusieurs protéines présentes dans le sang ont été étudiées. Cette liste inclut le facteur VIII, le facteur IX, le facteur XI, le vWF, l'antithrombine, la PC, la PS, le fibrinogène, le TFPI, le récepteur soluble à la protéine C (sEPCR), l'inhibiteur de la fibrinolyse activable par la thrombine (TAFI) et les D-dimères. La méthode la plus utilisée pour la mesure des quantités de protéines dans le sang est un test ELISA, un test basé sur la détection à l'aide d'anticorps spécifiques. Le sérum est mis en contact avec une plaque solide sur laquelle sont fixés des anticorps spécifiques de la protéine étudiée, après lavage du sérum la protéine reste sur la plaque, fixée par les anticorps. Elle est ensuite mise en contact avec un anticorps primaire spécifique puis un anticorps secondaire portant une enzyme catalysant une réaction produisant de la fluorescence ou de la couleur. Le substrat de l'enzyme est ensuite ajouté et la quantité de la protéine est estimée à partir de l'absorbance ou de l'intensité de la lumière dégagée.[270]

2 Acquisition des données génétiques

Les variations génétiques et l'expression des gènes peuvent être mesurées à l'aide de biopuces. Les biopuces regroupent un ensemble de technologies permettant d'analyser les séquences ADN avec un très haut débit et d'obtenir des informations sur les SNPs, l'expression des gènes ou encore les marques épigénétiques. Le fonctionnement des biopuces s'appuie sur les propriétés d'hybridation du brin d'ADN. Pour rappel, l'ADN est organisé en double hélice et les deux brins sont composés de nucléotides qui s'associent en paires complémentaires, l'adénine et la thymine (A-T) ainsi que la cytosine et la guanine (C-G).

Lorsque deux brins d'ADN sont complémentaires, ils se lient de manière non-covalente pour former la double hélice, la force de la liaison dépend du degré de complémentarité entre les deux brins. La biopuce consiste généralement en une surface solide sur laquelle sont fixées des sondes conçues spécifiquement pour être complémentaires avec les séquences d'intérêt qui vont être analysées. L'ADN de l'échantillon biologique est préalablement fragmenté puis marqué à l'aide de fluorochromes, des molécules émettant de la lumière après excitation à l'aide d'un laser. L'ADN de l'échantillon est mis en contact avec la biopuce et les séquences complémentaires aux sondes vont s'y fixer. Après un lavage pour éliminer les fragments d'ADN non fixés un laser est utilisé pour exciter les fluorochromes et la lecture de la réponse lumineuse va permettre d'obtenir les informations recherchées. Le fonctionnement des biopuces et les méthodes de lecture peuvent varier en fonction du type de biopuce utilisé et du constructeur mes le principe sur lequel s'appuient ces technologies reste le même.[271]

2.1 Les puces de génotypage

Les puces de génotypages sont spécifiquement conçues pour lire les SNPs et les courtes insertions/délétions du génome. Le panel de SNPs analysés est déterminé à l'aide d'un génome de référence. Une biopuce à ADN est capable de génotyper plus d'un million de marqueurs à la fois. On retrouve fréquemment deux types de puces, les puces solides où les sondes sont fixées à une lame unique, c'est la représentation traditionnelle des puces à ADN, mais aussi des *bead arrays* dans lesquelles chaque type de sonde est fixé à une micro-bille. Plusieurs technologies de génotypage ont été développées, cependant, afin de rester simple je vais me limiter à seulement présenter la technologie du constructeurs Illumina, présentée dans la Figure 10.[272]

Illumina utilise des *bead arrays* avec des sondes spécifiques de chaque SNP portées par différentes billes. Chaque sonde est complémentaire à une séquence flanquante du SNP mais ne comporte pas le SNP, à ces sondes sont ajoutés des oligonucléotides marqués avec différents fluorochromes. Lorsque les fragments d'ADN de l'échantillon sont mis en contact avec les sondes ils vont s'apparier aux sondes complémentaires sans que l'allèle du SNP n'influence l'appariement sonde-fragment. Une fois la sonde et le fragment appariés, l'oligonucléotide complémentaire à l'allèle du SNP va venir rallonger la sonde. Après lavage il ne reste que la sonde avec le nucléotide marqué complémentaire de l'allèle du SNP.

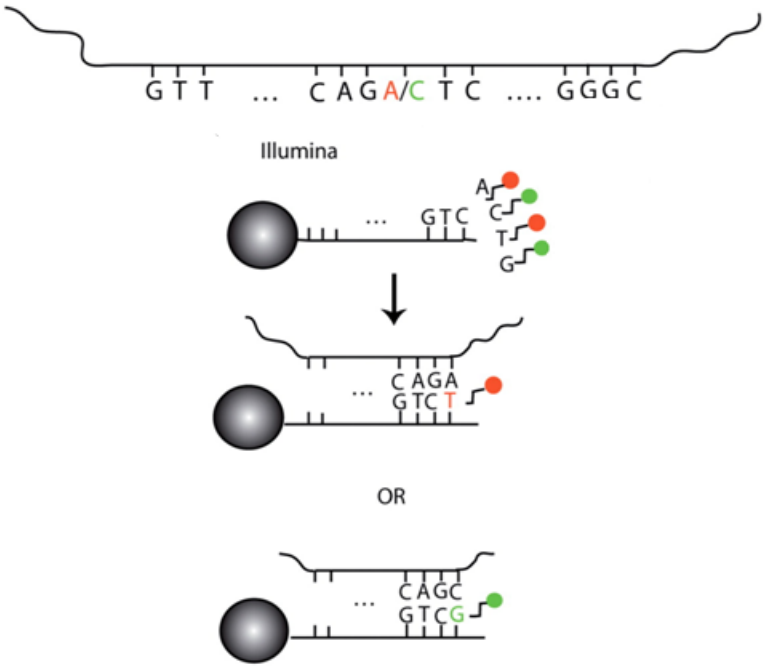


Figure 10 Principe de fonctionnement de base des puces Illumina. Des sondes complémentaires à une des séquences flaquantes du SNP sont fixées à des billes et des nucléotides portant différents fluorochromes sont présents dans la solution, l'ADN testé se fixe à la sonde sauf au niveau de l'allèle du SNP qui est libre, un nucléotide complémentaire vient alors se fixer avec son fluorochrome.

Illustration issue de l'article de LaFramboise, Nucleic Acids Res, 2009[272]

Le signal lumineux de chaque sonde est ensuite lu et les signaux lumineux pour chaque SNP sont regroupés en deux paramètres X_i et Y_i qui représentent l'intensité lumineuse des sondes de chaque SNP pour l'allèle A et B respectivement. Il n'y a pas de normalisation pour chaque puce mais une normalisation interne pour chaque échantillon. Pour chaque SNP i on calcule le rapport σ des intensités lumineuses normalisées des allèles A et B :

$$\sigma_i = (\frac{2}{\pi}) \times \arctan\frac{Y_i}{X_i}$$

Chaque génotype AA, AB ou BB est ensuite représenté par un groupe de valeurs de σ. Pour un SNP i, la proximité de σ_i à un des groupes permet d'interpréter le génotype. La distance entre σ_i et le centre du groupe est utilisée pour attribuer une note de qualité pour le SNP. Différentes puces de génotypage ont été utilisées dans le cadre des études présentées dans ce manuscrit, elles sont listées dans le tableau 3.

Constructeur	Modèle	SNPs génotypés	cohortes testées
Illumina	HumanCore Exome 12v1	542 585	Genesis/Genediab, Steno, UK-ROI
Illumina	Human 1M	1 199 187	DCCT/EDIC
Illumina	HumanOmni1-Quad	1 140 419	WESDR, EDC
Illumina	HumanCore Exome v3	∼500 000	CACTI, GoKinD
Illumina	Human 660W-Quad beadchip	657 366	MARTHA, MEGA
Illumina	Human 610-Quad	∼610 000	MARTHA, 3C
Illumina	Sentrix HumanHap300 beadchip	>317 000	EOVT, SUVIMAX

Tableau 3 Liste des puces de génotypage utilisées dans les deux projets présentés.

3 L'analyse

Maintenant que les outils et méthodes permettant l'acquisition des données physiologiques et génétiques ont été présentés, il est temps d'aborder les méthodes d'analyse. Ci-dessous sont présentées les différentes méthodes permettant de réaliser le tri et le contrôle qualité des données mais aussi les outils utilisés. L'environnement GNU R, un logiciel sous licence GNU, a été utilisé pour un certains nombres d'analyses statistiques, du traitement de données et la génération de la plupart des graphiques présentés.[273]

3.1 Le contrôle qualité

Afin de réduire au maximum le biais dans les analyses statistiques qui vont suivre, il est nécessaire de trier les données afin d'exclure les sujets qui pourraient être apparentés ou dont les caractères phénotypiques ont été mal renseignés et les SNPs pour lesquels l'attribution du génotype pourrait être incorrecte. Les individus peuvent être exclus en raison d'un manque d'informations phénotypiques mais aussi lorsqu'ils ont trop de SNPs non-génotypés (plus de 5% de SNPs manquants). Les individus avec un taux d'hétorozygotie trop élevé par rapport au taux attendu sont exclus car c'est un indicateur de génotypage de mauvaise qualité. Une mauvaise attribution du sexe, des liens de parenté et un profil d'*outlier* sont également des motifs d'exclusion. Pour les SNPs, on exclut les SNPs mal génotypés chez trop d'individus (plus de 1% des individus de la cohorte non-génotypés) et les SNPs qui ne sont pas en équilibre de Hardy-Weinberg chez les témoins.

Pour cette étape, je me suis essentiellement appuyé sur deux logiciels, le logiciel PLINK (v1.07) sous licence GNU et développé par Shaun Purcell[274, 275] et GCTA un logiciel sous licence GNU développé par Jian Yang disposant de l'outil GRM, très utilisé pour les analyses de parenté entre les individus.[276] Ces deux logiciels offrent un large choix d'outils décrits plus bas, permettant de réaliser toutes les opérations de contrôle qualité de mes travaux.

Vérification du sexe

Le sexe est un paramètre crucial dans les études d'épidémiologie puisque, comme nous l'avons vu avec les maladies cardiovasculaires, c'est un facteur de risque important susceptible d'interagir avec les facteurs génétiques étudiés. Il est donc important que le sexe soit correctement renseigné afin de ne pas introduire d'erreur dans l'analyse. En utilisant PLINK, on estime le coefficient de consanguinité sur le chromosome X (hors région pseudo-autosomale) de chaque individu. PLINK calcule le coefficient de consanguinité à partir des SNPs de la manière suivante :

$$F_{SNP} = \frac{O - E}{N - E}$$

Avec O le nombre de SNPs homozygotes observé, E le nombre de marqueurs homozygotes attendus sous l'hypothèse de l'équilibre de Hardy-Weinberg et N le nombre total de SNPs utilisés.[275] Les individus ayant un coefficient de consanguinité inférieur à 0,20 sont considérés comme des femmes tandis que ceux ayant un coefficient supérieur à 0,80 sont considérés comme des hommes. Les individus ayant un coefficient de consanguinité incompatible avec leur sexe déclaré et les individus ayant un coefficient de consanguinité situé entre 0,20 et 0,80 sont exclus de l'étude.

Stratifications de la population et parenté

Comme nous l'avons vu plus haut, il est important pour une étude d'association que la population d'étude soit homogène, de plus les méthodes statistiques employées sont valides à la condition que les individus soient indépendants. Si l'échantillon est composé de plusieurs sous-populations, on dit donc qu'il y a des stratifications de la population. Cette situation est problématique pour les études d'association car la fréquence allélique des SNPs varie en fonction de la population et une analyse sur plusieurs populations distinctes peut conduire à de fausses associations. La parenté est problématique car elle entre en conflit avec la nécessité d'indépendance entre les individus, si un certain nombre d'individus sont apparentés alors on ne peut plus s'assurer que les SNPs soient indépendamment et identiquement distribués. Les outils utilisés pour contrôler ces deux paramètres sont basés sur le même principe, la mesure de l'identité par état (IBS). L'IBS est définie comme la moitié du nombre d'allèles partagés. En conséquence si deux individus partagent 2, 1 ou aucun allèle sur un SNP donné, alors l'IBS pour ce SNP sera de 1, 1/2 ou 0 respectivement. De là, il est possible d'obtenir un IBS moyen pour tous les marqueurs disponibles dont la valeur sera comprise entre 0 et 1. Il est possible d'estimer l'IBS à partir des fréquences alléliques p et q du SNP de la façon suivante :

p(IBS=0)	$2p^2q^2$
p(IBS=0,5)	$1-(p^4+q^4+6p^2q^2)$
p(IBS=1)	$p^4+q^4+4p^2q^2$

À partir de ces information on peut en déduire que l'espérance de l'IBS pour des fréquences alléliques données p et q s'exprime telle que :

$$\frac{1+(p^2+q^2)^2}{2}$$

En supposant que les fréquences alléliques des SNPs testés soient uniformément réparties entre 0 et 1, on en déduit que l'espérance de l'IBS entre deux individus indépendants se situe vers 0,75. En utilisant PLINK, il est possible d'obtenir une matrice des IBS de tous les couples d'individus du groupe d'étude. Pour étudier la stratification de notre population d'étude on peut effectuer une analyse en composantes principales (ACP) sur la matrice des IBS. L'ACP est une transformation orthogonale qui permet de convertir un ensemble d'observations en un jeu de vecteurs appelés composantes principales. La transformation est définie de façon à ce que la première composante explique le plus possible de la variance des observations et que les composantes suivantes en expliquent une portion moindre de manière décroissante.[277]

Cette méthode permet de représenter la variation des observations sur un nombre de dimensions réduit. Il est alors possible de visualiser la stratification de la population en projetant les premières composantes principales calculées sur les données d'IBS. En projetant les deux premières composantes principales on obtient un nuage de points représentant les individus. On peut alors identifier les sous-groupes de l'échantillon et les *outliers*. On peut alors exclure les individus trop éloignés du centre du nuage de points (situés à plus de 4 écarts-types par exemple), en itérant le processus une deuxième fois, pour homogénéiser l'échantillon. Pour l'analyse de la parenté entre les individus, j'ai eu recours à l'outil GRM de GCTA qui permet d'obtenir la matrice des coefficients de parenté de tous les couples d'individus. Le coefficient de parenté est une mesure dérivée de l'IBS qui est définie de la manière suivante :

$$A_{jk}=\frac{1}{N}\sum_{i=1}^{N}\frac{(x_{ij}-2p_i)(x_{ik}-2p_i)}{2p_i(1-p_i)}$$

On définit ensuite un seuil de 0,025 à partir duquel on considère les individus apparentés et on exclut donc un des individus du binôme.

l'équilibre d'Hardy-Weinberg

L'équilibre d'Hardy-Weinberg est utilisé pour décrire la distribution des génotypes d'un SNP donné dans la population en admettant les conditions suivantes :

— Les individus sont diploïdes.
— La population est de taille infinie.
— La reproduction est uniquement sexuée.
— Pas de chevauchement entre les générations.
— Panmixie.
— Pas de différence sexuelle des fréquences alléliques.
— Pas de mutations, de migrations et de sélection.

Dans l'étude des populations humaines, les conditions de diploïdie et de reproduction exclusivement sexuée sont automatiquement admises et la taille des populations étudiées fait qu'on peut approcher la condition de taille infinie des populations. La panmixie signifie que les individus sont répartis de manière homogène et que les individus se reproduisant sont choisis de manière aléatoire. Le raison de l'hypothèse d'absence de différences de fréquences alléliques dues au sexe, le principe ne peut être appliqué aux chromosomes sexuels. L'équilibre d'Hardy-Weinberg ne s'applique pas non plus au chromosome mitochondrial car ce dernier est polyploïde. Toutes les conditions listées ne peuvent pas s'appliquer à une population réelle mais le test d'équilibre d'Hardy-Weinberg a tout de même sa place dans les études d'associations. Lorsque ces conditions sont remplies les fréquences des génotypes p_{AA}, p_{AB} et p_{BB} sont uniquement fonctions des fréquences alléliques p et q telles que :

$$\begin{cases} p_{AA} = p^2 \\ p_{AB} = 2pq \\ p_{BB} = q^2 \end{cases}$$

Avec p et q les fréquences des allèles A et B et $q = 1 - p$. Lorsque la distribution des génotypes d'un SNP est conforme à cette distribution on dit que le SNP est à l'équilibre d'Hardy-Weinberg.[278, 279] Un écart significatif de la distribution des génotypes à l'équilibre d'Hardy-Weinberg peut être expliqué de plusieurs façons :

— Une stratification de la population.
— Une pression de sélection exercée sur le *locus* testé.
— Un écart à la panmixie.
— Un problème de génotypage.

Dans tous les cas, il est préférable de retirer les SNPs non-conformes à l'équilibre d'Hardy-Weinberg afin d'éviter les fausses associations. Pour tester l'équilibre d'Hardy-Weinberg, on peut appliquer un test de χ^2 d'adéquation au modèle dans lequel on compare la distribution observée à la distribution attendue sous l'équilibre d'Hardy-Weinberg.[280] On récupère la distribution des génotypes AA, AB et BB qui est telle que $n_{AA} + n_{AB} + n_{BB} = N$ et on la compare aux effectifs attendus qui sont donc p^2N, $2pqN$ et q^2N. On peut alors calculer le score de χ^2 de la manière suivante :

$$\chi^2 = \frac{(|n_{AA} - p^2N|)^2}{p^2N} + \frac{(|n_{AB} - 2pqN|)^2}{2pqN} + \frac{(|n_{BB} - q^2N|)^2}{q^2N}$$

Le score de χ^2 est ensuite comparé à une distribution de χ^2 à un degré de liberté et les SNPs pour lesquels la distribution diffère significativement de la distribution attendue à l'équilibre avec une *p-value* inférieure à 10^{-5} sont exclus de l'analyse. PLINK intègre des outils permettant d'effectuer le test d'équilibre d'Hardy-Weinberg.

3.2 L'imputation

Une fois les marqueurs génotypés et le contrôle qualité effectué, il est encore possible d'améliorer la couverture de l'ensemble du génome en procédant à l'imputation des marqueurs non-observés. L'imputation des marqueurs manquants repose sur l'existence de segments haplotypiques partagés entre les individus hérités d'ancêtres communs. Il suffit donc d'identifier les haplotypes existants dans la population et de comparer les marqueurs génotypés de chaque individu à ces haplotypes. Les différents haplotypes ont été identifiés dans des populations de référence qui servent de base de comparaison pour l'imputation. Les populations de référence utilisées sont les populations du projet HapMap[239] ou celles du projet 1000 génomes,[281] plus récent. Le principe de fonctionnement de l'imputation est présenté dans la figure 11.

Les marqueurs de chaque individu génotypé sont comparés à l'ensemble des haplotypes possibles issus de la population de référence. Pour chaque marqueur, les haplotypes susceptibles de posséder l'allèle porté par l'individu sont sélectionnés puis les marqueurs manquants chez l'individu sont inférés des marqueurs trouvés sur les haplotypes compatibles.[282] Pour chaque SNP, le génotype imputé est accompagné d'une note permettant d'évaluer le degré de certitude de l'imputation.

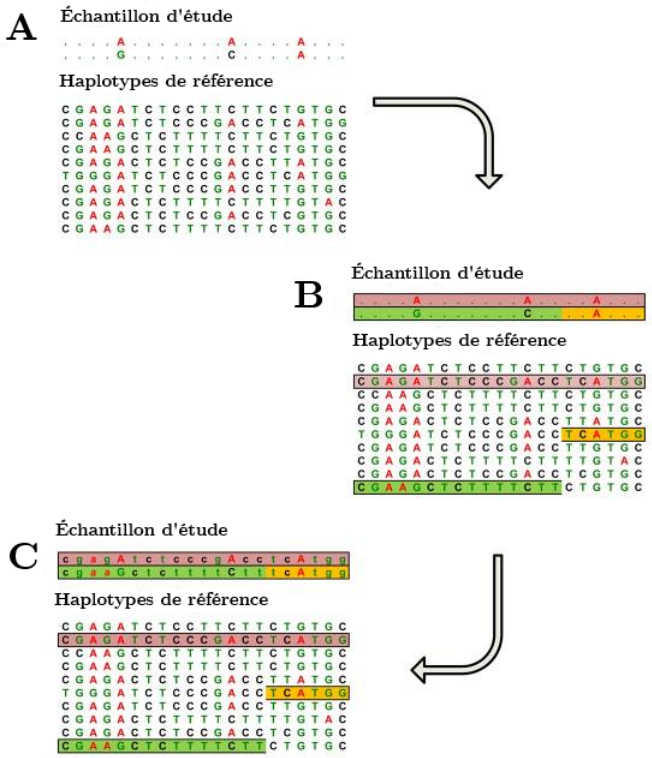


Figure 11 Principe de fonctionnement de l'imputation. **A** L'échantillon à imputer est comparé aux haplotypes de référence disponibles dans les bases de données. **B** En utilisant les polymorphismes disponibles dans l'échantillon à imputer, les sections haplotypiques partagées avec les haplotypes de référence sont identifiées. **C** Les polymorphismes à imputer sont inférés à partir des poymorphismes des haplotypes de référence correspondants.

Illustration issue de l'article de Li, Annu Rev Genomics Hum Genet, 2009[282]

Dans les travaux présentés ici, l'imputation a été réalisée avec deux outils différents, IMPUTE2[283] et MaCH.[284] IMPUTE2 fonctionne avec une méthode de Monte-Carlo par chaînes de Markov tandis que MaCH utilise un modèle de Markov caché mais les deux méthodes fonctionnent sur le même principe, avec une première étape dans laquelle l'haplotype de chaque individu est estimé à partir des SNPs génotypés et de l'ensemble des haplotypes de la population de référence puis une seconde étape d'imputation des marqueurs manquants à partir de l'haplotype reconstitué.

3.3 Association pan-génomique

Une fois que les données sont préparées, il reste à passer à l'analyse statistique dans laquelle on va tester l'association entre les paramètres phénotypique d'intérêt et le génotype des individus. Les études d'associations pan-génomiques peuvent être réalisées avec différents tests, Mach2dat[284] et PLINK incluent notamment des tests comme le test exact de Fischer et le test de tendance de Cochran-Armitage mais j'ai utilisé les modèles de régression dans

mes analyses car ils permettent de prendre en compte plusieurs variables. Les analyses d'association ont été effectuées avec les outils disponibles dans PLINK et l'environnement R. Les méta-analyses ont été réalisées avec l'extension "rmeta" de R[285] et le logiciel de méta-analyse METAL.[286] Les calculs de puissance statistique ont été réalisés à l'aide du logiciel CatS.[287]

Les modèles de régression

Les différentes études d'associations présentées dans ce manuscrit ont été réalisées à l'aide de différents modèles de régression linéaire ou logistique. Une analyse de régression sert à estimer la relation entre une variable expliquée et une ou plusieurs variables indépendantes, ou prédicteurs. L'analyse de régression permet de modéliser la relation entre la variable explicative et chacune des variables indépendantes sans faire varier les autres, la relation entre la variable expliquée et les variables indépendantes est estimée par une fonction de régression. Dans le cas de mon travail, le but des analyses de régression est de mettre en évidence l'existence d'une association entre la variable expliquée et une variable indépendante d'intérêt.

Dans le cas des analyses génétiques d'association les variables indépendantes sont évidemment le génotype du SNP testé mais aussi tous les facteurs de confusions susecptibles d'avoir un effet sur la variable expliquée. Elles dépendent donc du caractère testé mais on retrouve souvent l'âge et le sexe. Les composantes principales décrivant la variabilité de l'échantillon qui ont été présentées un peu avant sont également utilisées car elle permettent de contrôler efficacement la stratification au sein de l'échantillon.[288] Le choix des variables à intégrer dans le modèle peut être déterminé par des réalités biologiques mais il existe également des outils mathématiques permettant de choisir les variables afin d'obtenir un modèle le plus informatif possible que nous verrons par la suite. Un modèle de régression implique les trois paramètres suivant :

— La variable expliquée Y.
— Les variables indépendantes X.
— Un vecteur de paramètres inconnus β.

Y est estimé par une fonction de X et β tel que $Y \approx f(X, \beta)$, cette approximation est notée $E(Y|X) = f(X, \beta)$, Y et X étant disponibles dans l'échantillon, il faut utiliser les données pour estimer le vecteur de paramètres β. Il existe plusieurs critères nécessaires pour que le modèle soit robuste :

— L'échantillon utilisé doit être représentatif de la population sur laquelle est appliquée

le modèle de prédiction.

— Les variables indépendantes doivent être linéairement indépendantes.

Lorsque la variable expliquée est exprimée comme une fonction affine des variables indépendantes, on parle alors de régression linéaire, ce type de régression est utilisée pour les variables expliquées prenant des valeurs continues, ce qui est généralement le cas des variables biologiques comme le taux de LDL, la quantité de créatinine dans le sang etc... Un modèle de régression linéaire prend la forme suivante :

$$y_i = \beta_0 + \beta_1 x_{i1} + ... + \beta_p x_{ip} + \epsilon_i = X_i^T \beta + \epsilon \quad i \in 1, ..., n$$

$X_i^T \beta$ étant le produit scalaire des vecteurs $X_i = (x_{i1}, ..., x_{ip})$ et $\beta = (\beta_0, ..., \beta_p)$, l'ensemble des n équations est regroupé sous le terme $y = X\beta + \epsilon$. Les paramètres β peuvent être estimés par différentes méthodes mais la plus fréquente est la méthode des moindres carrés[289]. La méthode des moindres carrés s'appuie sur la somme des carrés des résidus, qui sont les distances verticales entre les points (x_i, y_i) et l'hyperplan $x^T \beta$, soit la distance verticale entre les points de données et la droite de régression. La somme des carrés des résidus s'exprime de la manière suivante :

$$S(b) = \sum_{i=1}^{n} (y_i - x_i^T b)^2 = (y - Xb)^T (y - Xb)$$

La valeur de b permettant de minimiser l'expression S est alors désignée comme un estimateur de β.

Dans le cas d'un événement à issue binaire, comme une maladie, dont le statut peut être défini par deux états, présence ou absence, on utilise un modèle de régression logistique. Ce modèle est utilisé pour prédire une issue binaire en fonction d'une ou plusieurs variables indépendantes. Le modèle repose sur l'utilisation de la fonction logistique, qui prend un argument réel pour donner une valeur comprise entre 0 et 1, cette fonction est définie de la manière suivante :

$$\sigma(t) = \frac{e^t}{e^t + 1} = \frac{1}{1 - e^{-t}} \quad \text{avec} \quad t \in \mathbb{R}$$

Si t est une fonction linéaire de $x \in \mathbb{R}$ telle que $t(x) = \beta_0 + \beta_1 x$, alors on peut écrire la fonction de répartition $F(x)$ telle que :

$$F(x) = \frac{1}{1 + e^{-(\beta_0 + \beta_1 x)}}$$

$F(x)$ correspond à la probabilité de "succès"(ou de présence dans le cas d'une pathologie). On peut ensuite prendre l'inverse de la fonction $F(x)$, le *logit* (ou logarithme des *odds*) :

$$g(F(x)) = ln(\frac{F(x)}{1-F(x)}) = \beta_0 + \beta_1 x$$

On peut alors en déduire les *odds*, qui sont le rapport entre la probabilité de succès et la probabilité d'échec :

$$\frac{F(x)}{1-F(x)} = e^{(\beta_0+\beta_1 x)}$$

On peut ensuite en déduire l'*odds ratio* :

$$OR = \frac{odds(x+1)}{odds(x)} = \frac{\frac{F(x+1)}{1-F(x+1)}}{\frac{F(x)}{1-F(x)}} = \frac{e^{\beta_0+\beta_1(x+1)}}{e^{\beta_0+\beta_1 x}} = e^{\beta_1}$$

Soit le facteur par lequel est multiplié le risque à chaque fois que x augmente de 1. De la même manière que la régression linéaire, la méthode couramment utilisée pour l'estimation des paramètres de la régression logistique est la méthode des moindres carrés.

Méta-analyse

Pour certains travaux présentés dans ce manuscrit, les résultats provenant de plusieurs études ont été combinés pour réaliser une méta-analyse. Une méta-analyse est une analyse statistique réalisée à partir des résultats de plusieurs études similaires effectuées sur des échantillons indépendants. L'objectif est d'obtenir une estimation du paramètre étudié dans l'ensemble des échantillons mais la méta-analyse permet également de mesurer l'hétérogénéité des résultats des différentes études utilisées. Le principal bénéfice d'une méta-analyse est d'augmenter la taille de l'échantillon et donc la puissance statistique de l'étude. Pour effectuer une méta-analyse il convient de choisir un modèle qui va dépendre de l'hypothèse de départ :

Effet fixe : On suppose ici que le paramètre qui est estimé prend la même valeur dans toutes les populations représentées par les échantillons inclus dans la méta-analyse et donc que les différents estimateurs propres à chaque échantillon sont des estimateurs du même paramètres.

Effet aléatoire : Dans ce cas, le paramètre estimé n'est pas identique pour toutes les populations représentées par les échantillons utilisés, il y a donc une certaine hétérogénéité entre les estimateurs des différents échantillons inhérente au fait qu'ils n'estiment pas exactement la même valeur du paramètre.

Le choix du modèle va influencer la méthodologie utilisée. Dans le cas d'un modèle à effet fixe, l'estimateur du paramètre est une moyenne pondérée par l'inverse de la variance. Soit, une méta-analyse de n études pour estimer le paramètre y, on a alors l'estimateur suivant :

$$\hat{y} = \frac{\sum_{i=1}^{n} \frac{y_i}{\sigma_i^2}}{\sum_{i=1}^{n} \frac{1}{\sigma_i^2}}$$

Avec sa variance

$$D^2(\hat{y}) = \frac{1}{\sum_{i=1}^{n} \frac{1}{\sigma_i^2}}$$

Avec σ_i^2 l'estimateur de la variance du paramètre y dans l'étude i.
En revanche, dans le cas d'un modèle avec effet aléatoire bien que la première étape est également de calculer la moyenne pondérée par l'inverse de la variance, il faut ensuite atténuer cette pondération en y ajoutant un élément de variance aléatoire dépendant de l'hétérogénéité entre les études.

Pour tester l'hétérogénéité entre les études d'une méta-analyse, il existe deux tests couramment utilisés. Le premier est le Q-test de Cochran[290], la statistique correspond à la somme pondérée des carrés des différences entre l'effet global et les effets des k études utilisées, elle est définie de la manière suivante :

$$Q = \sum_{i=1}^{k} w_i(\bar{y} - y_i)^2$$

Avec $\bar{y}$ l'effet global, et w_i le poids de chaque étude, qui est défini comme $\frac{1}{\sigma^2}$ dans le cas d'une méta-analyse avec un modèle à effet fixe. La statistique Q suit une distribution de χ^2 à $df = k - 1$ degrés de liberté, il est donc possible d'estimer le seuil de significativité de Q pour un α grâce à une table de χ^2 à $k - 1$ degrés de liberté. C'est la statistique Q qui est ensuite utilisée pour définir l'élément de variance aléatoire τ^2 pour la méta-analyse à effet aléatoire, tel que

$$\tau^2 = \frac{Q - k}{c}$$

Avec k le nombre d'études et c la variance en excès définie comme

$$c = \sum w_i - \frac{\sum w_i^2}{\sum w_i}$$

avec $w_i = \frac{1}{\sigma^2}$ la variance inversée. Au final, dans une méta-analyse avec effet aléatoire, la pondération par l'inverse de la variance prend en compte l'élément de variance aléatoire tel

que $w_i = \frac{1}{\sigma_i^2+\tau^2}$, il en résulte que l'effet estimé est donc défini tel que :

$$\hat{y} = \frac{\sum_{i=1}^{n} \frac{y_i}{\sigma_i^2} + \tau^2}{\sum_{i=1}^{n} \frac{1}{\sigma_i^2+\tau^2}}$$

Avec sa variance

$$D^2(\hat{y}) = \frac{1}{\sum_{i=1}^{n} \frac{1}{\sigma_i^2+\tau^2}}$$

La deuxième statistique est le I^2 qui est définie de la manière suivante :

$$I^2 = 100 \times \frac{Q - df}{Q} \quad \text{avec} \quad df = k - 1$$

Le I^2 correspond au pourcentage de variation inter-études dû à l'hétérogénéité plutôt qu'au hasard, dans le cas ou $Q < df$ la statistique est rabotée à zéro.[291]

Facteur d'inflation génétique

Une fois l'analyse d'association effectuée, on peut utiliser les résultats pour vérifier l'absence de stratification de la population d'étude. L'estimation du facteur d'inflation génétique permet de tester la présence de stratification de la population ou la conformité du modèle utilisé pour l'analyse. Pour cela il faut calculer le score du test de tendance de Cochran-Armitage avec une hypothèse d'effet additif pour chaque SNP i bi-allélique A/B, qui est défini de cette façon :

$$Y_i^2 = \frac{N(N(r_1 + 2r_2) - R(n_1 + 2n_2)^2}{R(N - R)(2N(n_1 + 2n_2) - (n_1 + 2n_2)^2)}$$

Avec N le nombre total d'individus, R le nombre de cas, r_1 et r_2 les nombres d'individus avec des génoypes AB et BB chez les cas respectivement et enfin n_1 et n_2 les nombres d'individus avec les génotypes AB et BB chez les témoins. Dans le cas d'une analyse avec un modèle de régression linéaire ou logistique, $Y_i^2 = \frac{\beta^2}{\sigma^2}$ est équivalent.[241] En supposant qu'il n'y a pas de stratification de la population, alors les Y_i^2 suivent une distribution χ^2 à 1 degré de liberté telle que $Y^2 \sim \lambda\chi_1^2$ avec λ qui est le facteur d'inflation génétique. Ce λ peut être estimé de la façon suivante[292] :

$$L = \frac{\text{médiane}(Y_1^2, ..., Y_n^2)}{0,456}$$

Calcul de puissance statistique

La puissance statistique d'une hypothèse correspond à la probabilité de rejeter une hypothèse nulle sachant qu'elle est fausse. La puissance statistique correspond à $1 - \beta$ avec β la probabilité d'erreur de type II, ou de non-rejet d'une hypothèse nulle incorrecte. La puissance statistique d'une hypothèse peut être choisie *a priori* afin de déterminer la taille nécessaire de l'échantillon afin de détecter un effet donné, ou d'estimer *a posteriori* les effets qu'il est possible de détecter avec un seuil de significativité satisfaisant avec la taille de l'échantillon disponible. Il existe plusieurs méthodes pour estimer la puissance statistique mais elles nécessitent toutes de connaître la taille de l'échantillon, l'intensité de l'effet recherché, la prévalence de la pathologie testée et le seuil α pour la probabilité d'erreur de type I choisi. Pour un modèle de régression logistique multivarié à p covariables, il est possible d'obtenir un estimateur de β à l'aide de la relation suivante[293] :

$$\frac{n}{1 - \rho^2_{1,2...p}} = \frac{(Z_{1-\frac{\alpha}{2}} + Z_{1-\beta})^2}{P1(1-P1)B^2}$$

Avec n l'effectif de l'échantillon, $\rho_{1,2...p}$ le coefficient de corrélation multiple reliant la covariable X_1 aux covariables $X_2, ..., X_p$, $Z_{1-\frac{\alpha}{2}}$ et Z_u le *Z-score* de la loi normale centrée-réduite au $u^{\text{ème}}$ percentile, sachant que α correspond au seuil d'erreur de type I fixé et β au seuil d'erreur de type II, choisi au préalable en cas de test de puissance statistique *a priori* et inconnu sinon, $P1$ correspond à la probabilité d'évênement ou prévalence et B correspond à l'intensité de l'effet estimé. Dans mes travaux j'ai utilisé l'outil CaTS qui permet de calculer la puissance statistique *post-hoc* des hypothèses testées.[287]

3.4 Autres analyses statistiques

Sélection des variables à intégrer au modèle

Comme nous l'avons vu avant, les modèles de régressions peuvent intégrer plusieurs variables indépendantes pour modéliser le comportement de la variable expliquée. Le choix des variables peut être motivé par les réalités biologiques et les données épidémiologiques disponible. Mais lorsque qu'il y a un très grand choix de variables possibles il est possible d'avoir recours à des outils mathématiques permettant de sélectionner les variables à intégrer au modèle. C'est le cas du critère d'information d'Akaike (AIC),[294] qui propose une estimation de la perte d'information entre le modèle utilisé et le processus étudié. Le calcul de l'AIC s'inspire de la distance de Kullback-Leibler,[295] qui est définie comme suit :

$$D(\theta, \theta_0) = 2 \int_{-\infty}^{+\infty} f(x|\theta_0) log \frac{f(x|\theta_0)}{f(x|\theta) dx}$$

où $f(x|\theta)$ est le modèle testé est $f(x|\theta_0)$ est le "vrai" modèle. Le modèle le plus adéquat est le modèle réduisant au maximum la distance de Kullback-Leibler. En pratique, le "vrai" modèle est inconnu et la distance de Kullback-Leibler est remplacée par le terme $-2ln(\hat{L})$ avec $\hat{L}$ la valeur maximum de la fonction de vraisemblance pour le modèle testé. Ce terme est ensuite pénalisé par le terme $2k$, k étant le nombre de paramètres estimés par le modèle. La formule permettant d'obtenir l'AIC du modèle testé est donc la suivante :

$$AIC = 2k - 2ln(\hat{L})$$

La valeur d'AIC correspond à un compromis entre adéquation du modèle aux données et simplicité du modèle. Pour sélectionner le modèle le plus approprié, il faut tester les différents modèles possibles et sélectionner celui avec l'AIC le plus faible. R dispose de fonctions permettant de sélectionner la combinaison de variables aléatoires afin de minimiser la valeur de l'AIC, la fonction *step* permet notamment de tester différentes séries de variables jusqu'à obtenir le modèle avec l'AIC le plus faible.

Score de risque polygénique

Depuis l'arrivée des GWAS, de nombreuses variations ont été associées aux maladies complexes. Ces variations ont souvent des effets modestes mais les maladies auxquelles elles sont associées sont hautement héritables. Cette héritabilité pourrait être expliquée par l'effet joint d'une multitude de variations dont les effets sont additifs. Afin de tester cette hypothèse, on peut calculer un score de risque polygénique à partir des différentes variations associées à la pathologie d'intérêt. La manière la plus simple de définir le score de risque polygénique est d'additionner le nombre d'allèles associés au risque des différentes variations inclues dans le score. Il faut cependant prendre soin à retirer les variations en déséquilibre de liaison fort avec d'autres variations inclues dans le score afin d'éviter de compter plusieurs fois des effets redondants. Ce score allélique peut ensuite être testé avec d'autres variables pertinentes dans un modèle de régression avec la pathologie d'intérêt.

Il est cependant possible que les différentes variations ne contribuent pas également au score et il existe plusieurs méthodes de pondération apportant des solutions à ce problème. Si les variations sont issues d'un GWAS il est possible de pondérer la somme du nombre des allèles de chaque variation avec l'effet de l'allèle dans une analyse de régression univariée (l'effet associé à l'allèle dans le GWAS en question). Il existe également des moyens de pondérer les allèles en fonction d'un potentiel effet délétère sur la séquence d'ADN, les outils SIFT et CADD sont des outils attribuant un score à un SNP en fonction de l'effet prédit sur la séquence. SIFT[296] est un outil utilisé pour estimer l'effet d'une variation dans la séquence

codante d'un gène sur la protéine produite, SIFT attribue un score situé entre 0 et 1 à chaque SNP codant non-synonyme, les SNPs avec un score proche de 0 sont prédit comme délétères pour la fonction de la protéine. CADD[297] est un outil de prédiction similaire mais qui a l'avantage de ne pas être limité aux variations codantes, le calcul du score CADD est basé sur le fait que les allèles délétères sont négativement affectés par la sélection naturelle. Le potentiel délétère d'une variation est déduit de sa fréquence dans la population. Les scores de risque polygénique sont souvent utilisés comme des outils de prédiction[298] mais permettent de mettre en évidence un risque génétique commun pour des groupes de maladies et permettent d'estimer l'héritabilité d'une pathologie.[299]

Analyse de randomisation Mendelienne

Les analyses de randomisation Mendelienne sont utilisées pour étudier l'effet d'un phénotype ou d'un facteur de risque environnemental sur le risque de développer une maladie d'intérêt en utilisant des marqueurs génétiques du phénotype ou du facteur de risque. Les marqueurs génétiques sont choisis de façon à reproduire l'effet causé par des causes environnementales, il peut s'agir par exemple de marqueurs génétiques affectant le taux circulant de LDLs. Il s'agit d'une analyse d'association entre la maladie et les marqueurs génétiques choisis pour simuler le phénotype testé. Le principal avantage de l'utilisation de marqueurs génétiques est le fait qu'ils permettent d'éviter un certain nombre de facteurs de confusion. Le fait que les allèles soient attribués au hasards aux individus fait que les marqueurs génétiques sont des variables indépendantes de l'âge, du sexe ou de tout autre facteur de confusion autrement susceptibles d'influencer le facteur de risque qui est reproduit. De plus, l'utilisation de marqueurs génétiques permet de fixer la relation de cause à effet puisque les marqueurs sont attribués à la naissance et précèdent donc la maladie.[77] Pour préparer une analyse de randomisation Mendelienne il faut sélectionner des marqueurs génétiques pour reproduire l'effet du facteur de risque d'intérêt sur la la maladie étudiée, ces marqueurs doivent remplir trois conditions.

- Il doit s'agir de marqueurs établis du facteur de risque.
- Les marqueurs doivent être indépendants des facteurs de confusion affectant l'association entre le facteur de risque et la maladie.
- Les marqueurs ne doivent pas être associés à la maladie indépendamment de leur effet sur le facteur de risque.

De manière générale, il faut choisir des marqueurs dont l'association avec le facteur de risque est bien documentée et qui n'ont pas d'effet pléiotropique. Une fois les marqueurs sélectionnés on peut estimer l'effet du facteur de risque sur la maladie comme le rapport de l'effet des

marqueurs sur la maladie et de l'effet des marqueurs sur le facteur de risque. Pour un facteur de risque X et une maladie Y on sélectionne les marqueurs génétiques $k \in \{1, ..., K\}$ pour l'analyse. On note X_k l'effet du marqueur k sur X avec une écart-type σ_{Xk}, de même on note Y_k l'effet du marqueur k sur Y avec une écart-type σ_{Yk} (si Y est binaire on utilise le *log-odds*). L'estimateur $\hat{\beta}$ de l'effet de X sur Y peut être obtenu en suivant la méthode proposée par *Stephen Burgess*[300] qui est la suivante :

$$\hat{\beta} = \frac{\sum_k X_k Y_k \sigma_{Yk}^{-2}}{\sum_k X_k^2 \sigma_{Yk}^{-2}}$$

Avec son écart-type $se(\hat{\beta})$ qui est obtenu de la manière suivante :

$$se(\hat{\beta}) = \sqrt{\frac{1}{\sum_k X_k^2 \sigma_{Yk}^{-2}}}$$

L'utilisation de plusieurs marqueurs génétiques implique que ces marqueurs sont indépendants, il est donc nécessaire de retirer les marqueurs en déséquilibre de liaison.

3.5 Exploiter les résultats

Une fois les résultats des analyses disponibles, il faut en tirer le plus d'informations possible afin de rendre l'interprétation de ces résultats plus pertinente, c'est là que les outils d'exploration des données sont utiles. D'abord, lorsque des marqueurs de risque génétiques de la pathologie sont identifiés il est nécessaire de savoir à quel *locus* ils sont associés grâce à des outils d'annotation. Annovar[301] est un outil d'annotation de variations permettant d'identifier le gène associé à chaque marqueur, la position du marqueur dans la séquence génomique et l'effet de ce marqueur sur la séquence. Un fois un gène identifié, il faut déterminer son potentiel en tant que gène candidat, il y a alors plusieurs possibilités.

En fonction des données disponibles, il est possible de poursuivre les analyses avec des analyses de type *eQTL* (*expression Quantitative Trait Loci*),[302] dans lesquelles on teste l'effet des différents génotypes du marqueur d'intérêt sur l'expression d'un ou plusieurs gènes d'intérêt. L'analyse peut se porter sur les gènes directement adjacents au marqueurs (*cis-eQTL*) ou sur des gènes distants (*trans-eQTL*). En fonction du nombre de marqueurs à analyser il est possible de pousser l'analyse à l'ensemble des gènes exprimés. Une analyse *eQTL* emploie un modèle de régression linéaire avec l'expression du gène d'intérêt en variable à expliquer (après avoir appliqué la normalisation appropriée) et le génotype et d'autres paramètres pertinents (sexe, âge, composantes principales, taux plasmatiques de cellules, etc...) en tant que variables indépendantes. Les bases de données telles que celles du *ROADMAP epigenomic project*[303] et ENCODE (**Enc**yclopedia **Of D**NA **E**lements)[304] mettent également à dis-

position des données d'épigénomique qui permettent de relier les variations étudiées à des marques épigénétiques et donc d'étudier l'effet potentiel de ces variations sur la régulation de l'expression des gènes. Les bases de données telles que Genecards[305] et OMIM[306] donnent des informations sur la fonction du gène et les recherches menées sur ce gène, des outils tels que GeneMANIA[307] et STRING[308] permettent de connaitre les gènes associés au gène d'intérêt, que ce soit par co-expression ou par interactions physiques entre les produits des gènes. GRASP[309] est un outil permettant de chercher les données d'association sur un gène ou un marqueur d'intérêt et permet donc de savoir si les marqueurs identifiés ont déjà été associés à la pathologie étudiée ou à des conditions proches. GTeX portal[310] est une base de données d'expression donnant des informations sur l'expression des gènes dans les différents tissus humain, une donnée très pertinente en fonction de la pathologie étudiée. Les marqueurs identifiés par les études d'association sont rarement les marqueurs causaux, il est donc nécessaire de rechercher des candidats potentiels en explorant les donnés de déséquilibre de liaison. Haploreg (v3)[311] et LDlink[312] sont des bases de données de déséquilibre de liaison qui permettent de trouver des marqueurs en déséquilibre de liaison avec le marqueur étudié. Locuszoom[313] est un outil permettant de générer des graphiques utiles pour observer le comportement des SNPs en déséquilibre de liaison avec le marqueur identifié. Grâce à ces outils, il est possible de réunir une grande quantité d'informations sur les gènes candidats identifiés afin de formuler des hypothèses sur leur lien avec la pathologie étudiée.

CHAPITRE 3

Résultats

Dans cette partie je présente les résultats obtenus dans les différents travaux que j'ai effectués pendant la thèse et je discute de leurs implications. Il s'agit simplement d'exposer l'ensemble des données obtenues afin de pouvoir en tirer des conclusions globales plus tard.

1 Étude pan-génomique sur les coronaropathies chez les diabétiques de type I

La première étude réalisée est une étude d'association pan-génomique portant sur les coronaropathies dans un contexte de diabète de type I. Pour cela j'ai réuni les groupes Genesis/Genediab, Steno et UK-ROI et j'ai récupéré les polymorphismes génotypés sur tout le génome des patients afin de les tester en association avec le risque de maladie coronarienne. Les données génotypées ont été enrichies avec une imputation des marqueurs manquants afin d'augmenter le nombre de marqueurs disponibles, les résultats des trois groupes ont ensuite fait l'objet d'une méta-analyse.

Les résultats obtenus dans la méta-analyse ont été répliqués dans une seconde méta-analyse regroupant DCCT/EDIC, WESDR, CACTI, EDC et GoKinD, le plan d'analyse est présenté dans la figure 12. Plusieurs analyses complémentaires portant sur le transcriptome ont été réalisées en parallèle et sont présentées plus bas. Les résultats de ces travaux ont fait l'objet d'une publication qui a été rajouté en annexe à la fin de ce manuscrit de thèse.[314]

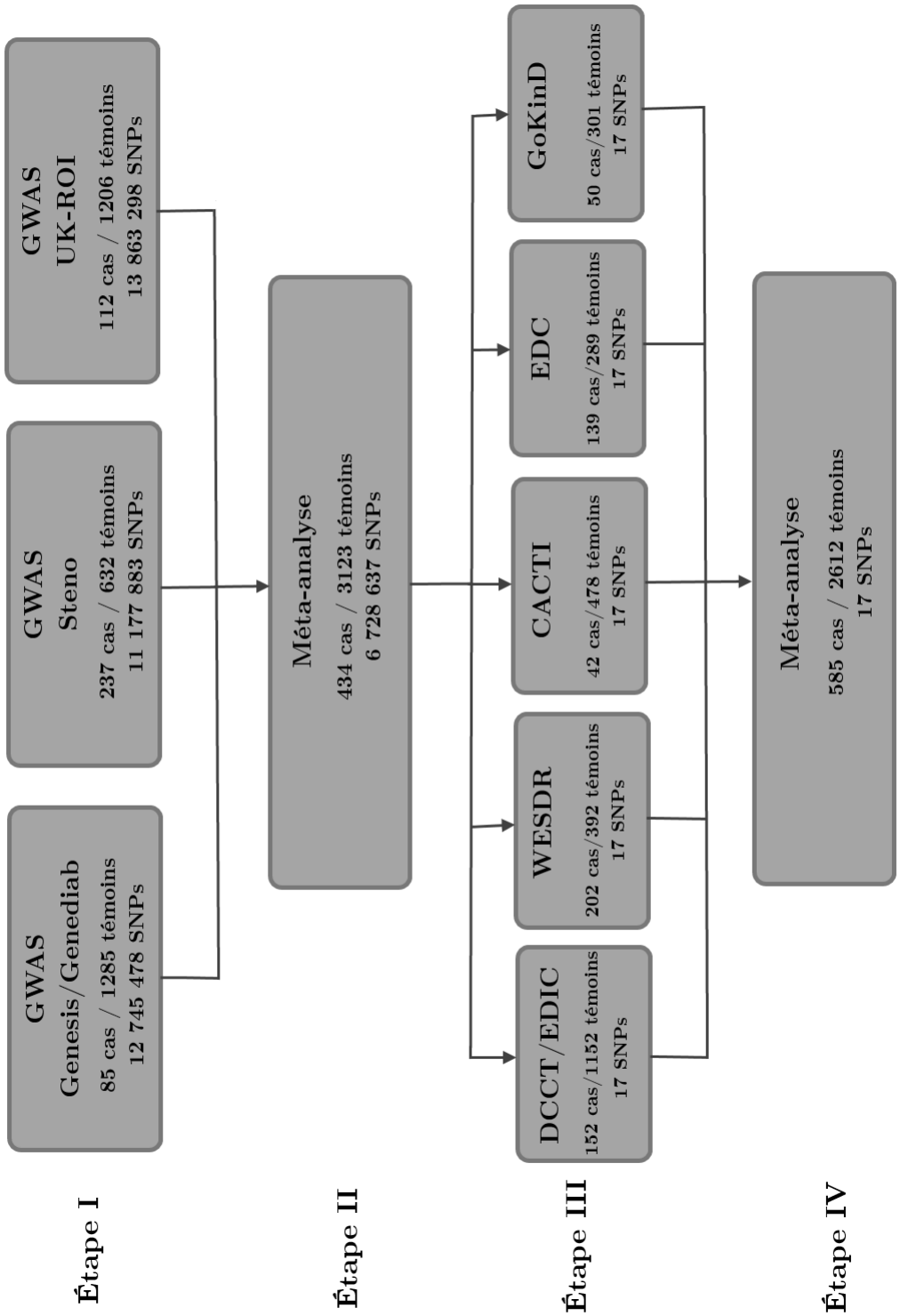


Figure 12 Protocole de l'étude. Durant l'étape I des GWASes indépendantes ont été réalisées sur les trois cohortes Genesis/Geneidab, Steno et UK-ROI. Ensuite durant l'étape II les résultats de ces trois groupes ont été regroupés dans une méta-analyse. Les 17 signaux qui ont obtenu un signal inférieur à *p-value* = 10^{-5} et qui étaient disponibles dans les groupes de réplication ont été répliqués dans cinq groupes indépendants, DCCT/EDIC, WESDR, CACTI, EDC et GoKinD, durant l'étape III. Les résultats de DCCT/EDIC, WESDR, CACTI, EDC et GokinD on été regroupés dans une méta-analyse à l'étape IV.

1.1 Sélection des individus et des marqueurs

Avant l'analyse des données du groupe Genesis/Genediab, il a été nécessaire de trier les données avec un certain nombre de filtres appliqués aux individus puis aux marqueurs. Le groupe était initialement composé de 1586 individus mais 69 individus ont du être retirés en raison soit d'un sexe incorrectement renseigné, d'un phénotype cardiaque manquant, d'un trop grand nombre de marqueurs non génotypés (plus de 5% du total) ou car il s'agissait de doublons. Pour les marqueurs, sur les 538 448 SNPs génotypés par la puce, 271 039 SNPs ont été retirés car ils appartenaient aux chromosomes sexuels ou à la mitochondrie, soit car ils s'agissait de doublons, soit car leur fréquence allélique mineure était inférieure à 1%, car il étaient génotypés chez moins de 99% des individus ou soit car la distribution des allèles ne respectait pas l'équilibre Hardy-Weinberg (p-value$<10^{-5}$) chez le groupe contrôle.

Ensuite, j'ai voulu examiner la proximité entre les individus via l'estimation de leur IBS. J'ai calculé la matrice d'IBS et l'ai utilisée pour estimer la relation génétique avec l'outil GRM de CGTA, cet outil permet d'obtenir une estimation de la proximité des individus grâce à une coefficient de similitude. J'ai pu regrouper les individus en branches d'individus apparentés, représentées dans la figure 13A, grâce à un *clustering* basé sur la distance entre les individus. Parmi les individus avec un score GRM supérieur à 0,025 je n'ai conservé qu'un seul individu, celui avec le plus de SNPs correctement génotypés. Au total 147 individus ont été retirés de l'étude. La matrice d'IBS a ensuite été utilisée pour mettre en évidence la présence d'une stratification de l'échantillon. Pour cela j'ai réalisé une analyse en composante principales sur la matrice des IBS et j'ai calculé le centroïde du nuage de points à l'aide des deux premières composantes principales. J'ai défini comme *outlier* tous les individus situés à plus de quatre écarts-types du centroïde mais aucun individu ne s'éloignait autant du centre du nuage de points mis à part un individu qui était cependant proche de la limite et a été conservé, comme le montre la figure 13B.

Enfin, j'ai voulu comparer les individus du groupe aux populations du projet HapMap dont les données de génotypages sont disponibles sur le site du projet. J'ai calculé l'IBS de tous les individus et avons réalisé une nouvelle analyse en composantes principales, la représentation des deux premières composantes principales de la figure 13C montre encore une fois un groupe plutôt homogène. Après le contrôle qualité nous avons obtenu un échantillon final de 1370 individus avec 267 409 marqueurs à analyser.

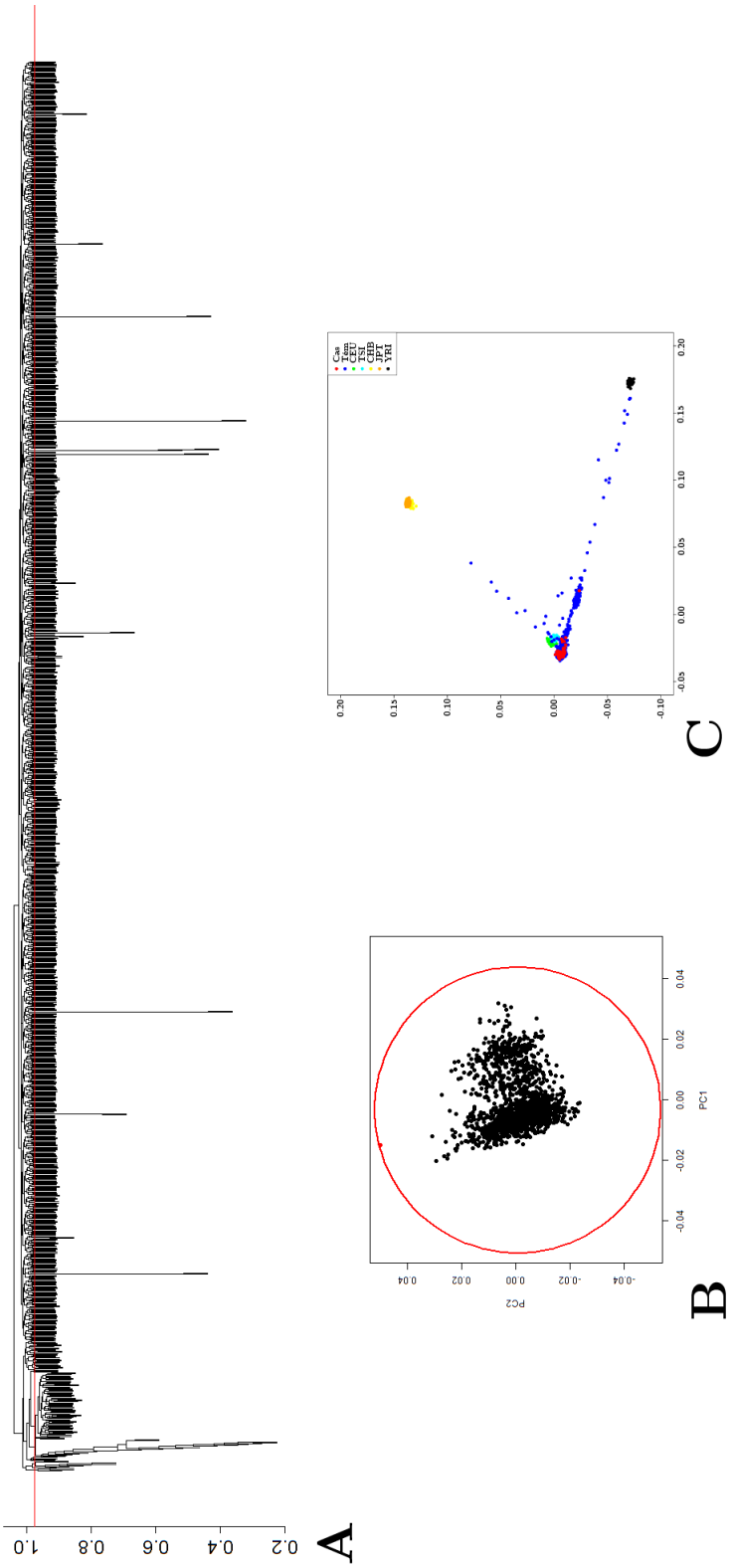


Figure 13 A Dendrogramme des individus après le premier tri (n=1517) réalisé à partir d'une matrice de distance basée sur le score GRM des individus. Mis à part quelques paires d'individus proches on observe deux sous-groupes plus importants (branches de gauche). **B** Détection des *outliers* après exclusion des individus en fonction du score GRM (n=1370), on observe qu'aucun individu ne se trouve à plus de quatre écart-types du centroïde du nuage de points, mis à part un individu en rouge qui a été conservé malgré tout. **C** ACP sur les IBS de l'ensemble de la cohorte Genesis/Genediab (n=1586) avec les groupes du projet HapMap. Les cas sont représentés en rouge et les témoins en bleu, les populations de HapMap sont en vert pour les Américains d'origine Européenne (CEU), turquoise pour les Toscans (TSI), jaune pour les Han (CHB), orange pour les Japonais (JPT) et noir pour les Nigérians (YRI). Mis à part un groupe d'*outliers* l'ensemble du groupe Genesis/Genediab colocalise avec les groupes HapMap d'origine Européenne.

1.2 Association avec les données imputées et méta-analyse

Une fois l'imputation et le tri final effectués, les 12 745 477 marqueurs disponibles dans Genesis/Genediab avec un critère d'imputation r^2>0,3 ont été testés en association avec la coronaropathie dans un modèle de régression logistique avec l'âge, le sexe, la néphropathie diabétique et les quatre premières composantes principales en utilisant un modèle additif pour les allèles. Ces résultats ont ensuite été regroupés avec les résultats d'analyses suivant le même protocole dans Steno et UK-ROI. Les résultats des 6 728 637 SNPs communs à ces trois analyses ont ensuite été regroupés dans une méta-analyse avec un modèle en effet fixe regroupant 3123 contrôles et 434 cas. Après la méta-analyse le facteur d'inflation est de 1,04 une synthèse de l'ensemble des signaux est présentée dans la figure 14. Un *locus* a dépassé le seuil significatif de $5x10^{-8}$. Il s'agit de rs115829748, proche du gène *MAP1B* sur le chromosome 5, l'allèle mineur T, avec une fréquence allélique de 0,04, étant associé à un risque accru de coronaropathie avec un *odds ratio* de 3,16 [2,18-4,59] et une *p-value* de $1,36x10^{-9}$. En plus de ce résultat, 20 autres *loci* ont été sélectionnés comme candidats, 19 *loci* avaient un signal dépassant le seuil de significativité de 10^{-5} et un autre *locus* avec une *p-value* de $1,72x10^{-5}$ a été retenu car il s'agissait d'une variation non-synonyme, ces *loci* sont présentés dans le tableau 4.

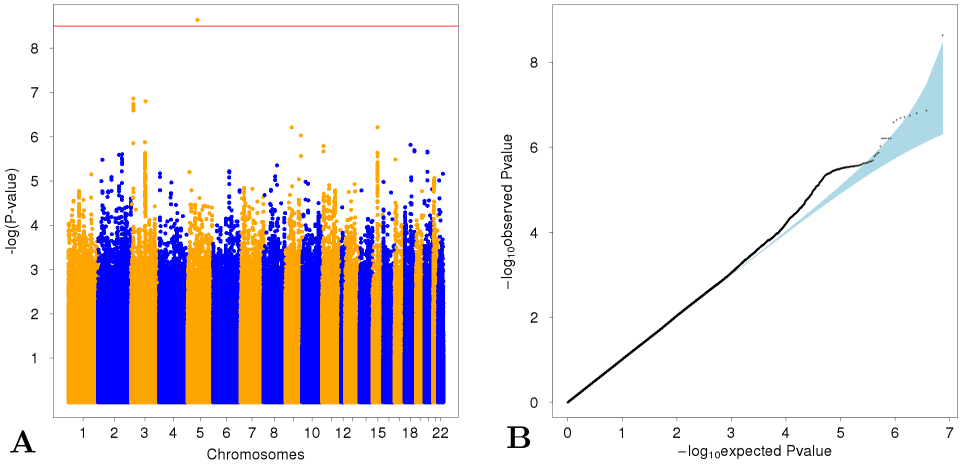


Figure 14 A Manhattan plot de l'ensemble des signaux, on constate qu'un seul signal dépasse le seuil de significativité pan-génomique de $5x10^{-8}$ sur le chromosome 5. **B** QQ-plot de l'ensemble des signaux, on n'observe pas de déviation évidente par rapport à la distribution attendue des *p-values*, mis à part pour le signal le plus fort.

Chr	Bp	*Locus*	SNP	Variation	EA/NEA[(1)]	EAF[(2)]	OR[(3)]	p[(4)]	Direction[(5)]	$I^{2(6)}$	$P^{Het\,(7)}$
1	205484373	*CDK18*	rs138760780	Intronic	T/C	0,018	3,48	$9{,}20\text{x}10^{-6}$	+++	0	0,943
2	97455276	*CNNM4*	rs116656846	Intronic	A/G	0,023	2,57	$9{,}88\text{x}10^{-6}$	+++	0	0,661
2	155225182	*GALNT13*	rs17206992	Intronic	G/A	0,057	2,32	$6{,}33\text{x}10^{-6}$	+++	0	0,817
2	177645590	*AC092162,1*	rs113517532	Intergenic	AGAT/A	0,106	1,75	$1{,}90\text{x}10^{-6}$	+++	0	0,912
3	13370674	*NUP210*	rs73018809	Intronic	T/A	0,024	3,50	$1{,}89\text{x}10^{-7}$	+++	59,1%	0,087
3	103975418	*MIR548A3*	rs28641753	Intergenic	T/C	0,071	2,10	$1{,}59\text{x}10^{-7}$	+++	0	0,863
4	6171230	*JAKMIP1*	rs78031527	Intronic	T/C	0,201	1,70	$5{,}41\text{x}10^{-6}$	+++	0	0,493
5	10500646	*ROPN1L*	rs143537377	Intronic	C/A	0,096	1,88	$7{,}89\text{x}10^{-6}$	+++	1,1%	0,577
5	71394387	*MAP1B*	rs115829748	Intergenic	T/C	0,040	3,16	$1{,}36\text{x}10^{-9}$	+++	71,9%	0,028
6	95557471	*MANEA-AS1*	rs9354144	Intergenic	A/T	0,105	1,71	$5{,}78\text{x}10^{-6}$	+++	0	0,785
8	73842523	*KCNB2*	rs571622299	Intronic	A/G	0,015	4,12	$6{,}47\text{x}10^{-6}$	+++	47,2%	0,15
9	37034095	*PAX5*	rs143723948	UTR5	T/C	0,499	1,65	$6{,}01\text{x}10^{-7}$	+++	0	0,985
9	71955717	*FAM189A2*	rs12344245	Intronic	G/A	0,035	2,52	$9{,}23\text{x}10^{-6}$	+++	0	0,950
10	19457387	*ARL5B*	rs117826205	Intronic	C/T	0,026	2,72	$6{,}30\text{x}10^{-6}$	+++	0	0,482
11	8080425	*TUB*	rs61879614	Intronic	C/T	0,048	2,94	$2{,}60\text{x}10^{-6}$	+++	9,4%	0,332
16	2160973	*PKD1*	rs116092985	Missense (W1399R)	G/A	0,097	1,85	$1{,}72\text{x}10^{-5}$	+++	29,5%	0,242
17	4328164	*SPNS3*	rs34319244	Intergenic	C/CT	0,440	1,51	$2{,}59\text{x}10^{-6}$	+++	48,8%	0,142
18	45399356	*SMAD2*	rs113114656	Intronic	T/C	0,040	2,67	$2{,}30\text{x}10^{-6}$	+++	2,4%	0,359
21	21347156	*NCRNA00320*	rs67213764	Intergenic	G/A	0,261	1,48	$9{,}28\text{x}10^{-6}$	+++	0	0,536
21	24929109	*AP000459,7*	rs12482425	Intergenic	A/G	0,314	0,66	$9{,}33\text{x}10^{-6}$	−−−	0	0,509
22	25988780	*ADRBK2*	rs373009901	Intronic	C/G	0,019	3,94	$9{,}42\text{x}10^{-6}$	+++	0	0,599

Tableau 4 Signaux retenus avec une *p-value* inférieure à 10^{-5} lors de la première étape de méta-analyses.
(1) *Estimated allele/Non-estimated allele*, désigne l'allèle qui a été utilisé pour le test d'association.
(2) *Estimated allele frequency*, fréquence moyenne de l'allèle estimé dans les populations de l'étude.
(3) *Odds ratio* pour le risque de coronaropathie.
(4) *P-value* associée à l'*odds ratio* pour la méta-analyse.
(5) Direction des effets observés dans chacune des études de la méta-analyse.
(6) Score I^2 d'hétérogénéité des résultats de chaque étude.
(7) *P-value* d'hétérogénéité unidirectionnelle.

1.3 Réplication des résultats

Parmi les 21 *loci* candidats, 17 (les SNPs rs571622299, rs143723948, rs34319244 et rs373009901 n'étaient pas disponibles dans toutes les études de la réplication) ont pu ensuite être répliqués dans une méta-analyse regroupant les groupes DCCT/EDIC, CACDI, WESDR, EDC et GoKinD. Les *loci* ont été analysés en association avec les coronaropathies dans un modèle de régression logistique avec l'âge, le sexe, la durée du diabète et les composantes principales de la population en variables indépendantes et avec un modèle allélique additif. La méta-analyse a été réalisée avec un effet fixe. Les résultats de la réplications sont présentés dans le tableau 5. Aucun signal n'atteint le seuil de Bonferroni de $3x10^{-3}$ mais trois *loci* étaient associés au risque de coronaropathie avec une *p-value* inférieure à 0,05. Ces trois signaux sont rs138760780 sur le gène *CDK18*, rs116092985 sur le gène *PKD1* et rs12344245 sur le gène *FAM189A2*. Le seul signal ayant dépassé le seuil de $5x10^{-8}$ dans la découverte, rs115829748 n'était pas associé au risque de coronaropathie dans la réplication (p = 0,279).

SNP	EAF	OR	$P^{(1)}$	Direction	I^2	P^{Het}
rs138760780	0,020	1,88	0,014	+ + + + −	0	0,509
rs116656846	0,028	0,89	0,696	− − − + −	0	0,876
rs17206992	0,052	0,97	0,580	+ + − + −	0	0,934
rs113517532	0,105	0,86	0,887	+ − + − −	0	0,658
rs73018809	0,022	0,60	0,955	− − − − +	0	0,722
rs28641753	0,067	1,02	0,440	− + + − −	62,7%	0,030
rs78031527	0,209	1,13	0,114	+ + − + −	0	0,622
rs143537377	0,100	0,85	0,885	− − + − +	19,2%	0,292
rs115829748	0,036	1,13	0,283	− + + + −	0	0,655
rs9354144	0,097	1,05	0,330	− + + − −	23,9%	0,262
rs12344245	0,038	1,45	0,022	+ − + + +	0	0,560
rs117826205	0,029	1,11	0,315	− − − + +	5,7%	0,374
rs61879614	0,046	1,11	0,300	+ − − − −	0	0,775
rs116092985	0,096	1,29	0,026	+ + + + −	0	0,488
rs113114656	0,038	1,04	0,437	− + + − +	43,4%	0,132
rs67213764	0,260	1,03	0,363	+ + − − −	32,2%	0,207
rs12482425	0,308	1,09	0,871	+ + + + +	0	0,931

Tableau 5 Réplication des 17 signaux candidats de la méta-analyse.
(1) *P-value* d'association unidirectionnelle.

1.4 Analyse des associations déjà identifiées

Le risque de coronaropathie dans la population générale est un sujet bien étudié et a déjà fait l'objet de plusieurs études d'association pan-génomiques. Au moment où ce travail

a été initié 95 *loci* avaient déjà été identifiés et j'ai tenté de répliquer ces résultats dans notre groupe d'étude. 95 SNPs ont été recueillis à partir des études réalisées par Cardiogram et C4D[7, 66, 59, 60, 61] et ont été testés dans un modèle de régression logistique avec l'âge, le sexe, la néphropathie diabétique et les quatre premières composantes principales, les résultats de la réplication sont présentés dans le tableau 6.

			Résultats publiés			Réplication		
SNP	Chr	Locus	EA[1]	EAF[(2)]	OR[(3)]	EAF[(4)]	OR[(5)]	P-value[(6)]
rs17609940	6	*ANKS1A*	G	0,82	1,03	0,80	1,32	0,004
rs11838776	13	*COL4A2*	A	0,26	1,07	0,28	1,33	0,001
rs2075650	19	*TOMM40*	A	0,86	0,93	0,87	0,74	0,003

Tableau 6 Réplication des signaux publiés par l'étude C4D-Cardiogram. Par soucis de place, ce tableau ne présente que les signaux les plus fort dans mon étude (*p-value*<5x10^{-3}), la totalité des 95 SNPs testés est présentée en annexe avec le manuscrit.
(1) Allèle de référence utilisé dans la méta-analyse C4D-Cardiogram.[7]
(2) Fréquence de l'allèle de référence rapportée dans la méta-analyse C4D-Cardiogram.
(3) Effet de l'allèle de référence estimé dans la ;éta-analyse C4D-Cardiogram.
(4) Fréquence de l'allèle de référence dans la méta-analyse Genesis/Genediab, Steno, UK-ROI.
(5) Effet estimé dans la méta-analyse Genesis/Genediab, Steno, UK-ROI.
(6) *P-value* unidirectionnelle associée à l'effet estimé dans la méta-analyse Genesis/Genediab, Steno, UK-ROI.

1.5 Discussion autour des gènes candidats mis en évidence

Dans cette étude je cherchais à mettre en évidence des marqueurs génétiques du risque de coronaropathie chez les patients diabétiques de type I. Dans l'étape de découverte, le SNP rs115829748 proche du gène *MAP1B* a été identifié avec un signal dépassant le seuil de significativité pour une étude pan-génomique (5x10^{-8}), cependant ce résultat n'a pas pu être répliqué. Nous avons cependant pu observer des associations intéressantes à la fois dans la méta-analyse et dans la réplication pour trois *loci*, *CDK18*, *PKD1* et *FAM189A2*. L'allèle T de rs138760780 situé sur *CDK18* est associé à un risque 2,5 fois plus élevé de coronaropathie, il s'agit d'un allèle rare n'étant pas en déséquilibre de liaison fort (r^2>0,8) avec un autre SNP (figure 18). La position du SNP ne nous permet pas d'affirmer un rôle fonctionnel évident. De plus, *CDK18* code pour une kinase cycline-dépendante impliquée dans la transition G2/M du cycle cellulaire.[315] Cette fonction vitale pour l'ensemble des cellules eucaryotes n'indique pas de rôle évident dans le diabète de type I ou les complications vasculaires. Un SNP sur le même *locus* a été associé au diabète de type I mais il n'a pas été répliqué.[316] À ce stade,

il n'y a pas suffisamment de données disponibles sur ce gène pour estimer sa pertinence en tant que gène candidat.

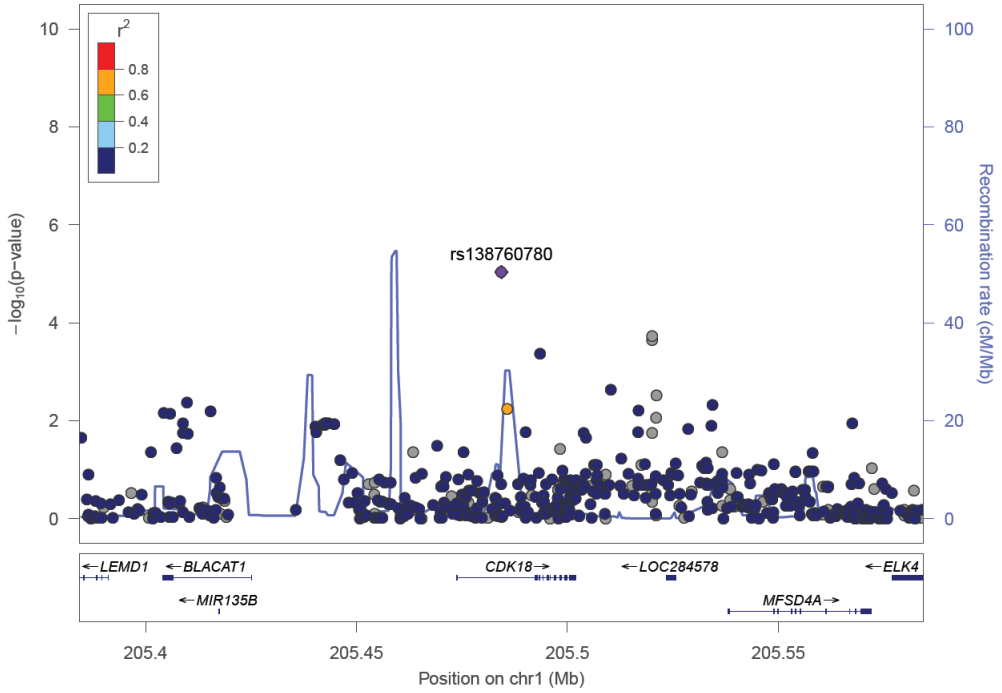


Figure 15 *Plot* d'association régional centré sur *CDK18*, on n'observe aucun SNP en déséquilibre de liaison fort avec rs138760780 dans la région.

rs116092985 situé sur *PKD1* est plus intéressant car il s'agit d'une mutation non-synonyme (Trp1399Arg), la présence de l'allèle G (Arg1399) est associé avec un risque accru de coronaropathie d'un facteur 1,5. *PKD1* code pour la polycystine 1, une protéine transmembranaire qui est exprimée à la surface des cellules de l'épithélium rénal lors de la morphogénèse des tubules rénaux.[317] Des mutations dans le gène *PKD1* sont la cause principale de polykylose rénale dominante, une maladie qui se manifeste durant l'enfance avec des dysfonctionnements de la fonction rénale.[318] La substitution d'un acide aminé aromatique hydrophobe comme le tryptophane par un acide aminé chargé comme l'arginine pourrait avoir des conséquences sur la structure de la protéine, surtout dans le cas d'une protéine transmembranaire où la substitution d'un acide aminé hydrophobe par un acide aminé hydrophile peut compromettre la structure de la région membranaire de la protéine. *PKD1* pourrait être un bon gène candidat pour étudier le rôle de la néphropathie diabétique dans le risque de complication cardiovasculaire.

L'allèle G de rs12344245 est associé à un risque accru de coronaropathie d'un facteur 1,8. Ce SNP est localisé dans un intron de *FAM189A2*, un gène dont la fonction n'est pas caractérisée. Le SNP rs10120442 situé dans le même gène a été associé au rapport albumine/créatinine urinaire avec une *p-value* de $2{,}6\text{x}10^{-4}$,[319] tandis que le rs10780297 a été associé au risque de coronaropathie dans la population générale, avec cependant un signal modeste (*p-value* de $9{,}3\text{x}10^{-4}$).[66] Malgré ces éléments on ne peut pas affirmer que ce gène constitue un candidat particulièrement pertinent.

Cette étude était également l'occasion de tester l'effet de variations identifiées dans la population générale sur cette population de patients diabétiques de type I. Sur les 95 SNPs testé, trois SNPs étaient associé au risque de coronaropathie avec une *p-value* suggestive, rs17609940, rs11838776 et rs2075650 qui se trouvent sur les *loci* *ANKS1A*, *COL4A2* et *TOMM40-APOE* respectivement. Ces résultats suggèrent en plus que ces SNPs pourraient avoir un effet plus fort chez les individus diabétiques que dans la population générale mais cette hypothèse doit tout de même être validée avec un échantillon de taille plus importante et avec un groupe de réplication. *ANKS1A* code pour une protéine qui est impliquée dans la régulation de l'activité des facteurs de croissance,[320] plusieurs SNPs localisés dans le même *locus* sont d'ailleurs associés à la taille.[321, 322] Il est possible que ce gène soit impliqué dans d'autres voies physiologiques mais les données manquent pour le relier aux complications du diabète. *COL4A2* est le gène de la sous-unité *alpha* 2 du collagène de type IV. Le collagène de type IV est un des principaux éléments composants la lame basale. Or l'endommagement de la lame basale est un événement commun à toutes les complications microvascualaires et macrovasculaires du diabète. Les dégâts sur la plaque pourraient contribuer à l'accélération de la formation de la plaque d'athérome et augmenter le risque de formation d'un thrombus. L'endommagement de la lame basale contribue également à la néphropathie diabétique car la lame basale est un élément clé de la barrière glomérulaire. *COL4A2* pourrait donc être un candidat intéressant pour le risque de complications dans le diabète de type I étant donné que la lame basale est particulièrement sensible à l'hyperglycémie chronique. Le *locus* *TOMM40-APOE* est principalement impliqué dans le métabolisme des acides gras, *APOE* code pour une alipoprotéine et plusieurs marqueurs sur ce *locus* sont associés aux LDL et au cholestérol total.[323] Le cholestérol est un facteur de risque des coronaropathies mais il est difficile de trouver une explication permettant de justifier un effet plus fort chez les individus diabétiques de type I.

Malgré ces découvertes, l'étude présente une limitation majeure qui est la taille de l'échantillon. Avec 1019 cas et 5735 témoins, nous sommes assez loin des échantillons généralement utilisés pour les études pan-génomiques, ce qui limite la capacité de notre étude à surmonter le seuil de significativité normalement imposé dans ce type d'études (*p-value* de $5\text{x}10^{-8}$). En effet

seuls des *odds ratio* supérieurs à 1,6 sur des variations fréquentes (MAF>0,28) auraient pu être détectés à ce seuil de significativité avec une puissance statistique acceptable (>80%). De même, je n'avais pas suffisamment d'effectifs dans le groupe de réplication pour détecter les marqueurs de *CDK18*, *PKD1* et *FAM189A2* avec un seuil de Bonferroni de $3x10^{-3}$, la puissance de détection d'un signal pour rs138760780 avec ce seuil était de 53%, elle était de 38% pour rs116092985 et de 26% pour rs12344245.

Cette limitation pourrait être surmontée en agrandissant l'échantillon. Cependant la relative rareté des patients diabétiques de type I (5% à 10% des patients diabétiques) pourrait être un obstacle dans la résolution de ce problème. Il serait également possible d'améliorer la puissance statistique en ayant plutôt recours à des analyses *Gene-based* ou *Pathway-based* qui permettraient de réduire le nombre de tests à effectuer en plus de permettre de prendre en compte des allèles rares. Deuxièmement, j'avais prévu d'effectuer une analyse prenant en compte l'interaction entre la présence de coronaropathie et le statut de néphropathie diabétique afin d'étudier l'interaction entre la néphropathie et les marqueurs génétiques. Malheureusement une division supplémentaire de l'échantillon en sous-groupe aurait réduit davantage la puissance de l'étude. Ce serait cependant une prochaine étape logique afin d'apporter des réponses concernant l'hypothèse d'une relation entre les complications cardio-vasculaires et la néphropathie diabétique.

2 Impact des marqueurs de la fonction rénale sur la thrombose veineuse

La deuxième étude présentée porte sur la relation entre la fonction rénale et la thrombose veineuse et utilise un score de risque polygénique. J'avais sélectionné plusieurs marqueurs génétiques associés à l'eGFR, un marqueur de la fonction rénale, et je les ai testés en association avec la thrombose veineuse dans un modèle de régression logistique. Par la suite, les marqueurs ont été rassemblés dans un score de risque polygénique qui a été à son tour testé en association avec la thrombose veineuse dans un modèle de régression logistique. Cette association a ensuite été répliquée dans une seconde cohorte indépendante. Pour finir, l'étude a été complétée par des tests d'association du score avec différents facteurs de l'hémostase et des analyses *in silico* ont été réalisées sur les différents *loci* proches des SNPs sélectionnés. Pour plus de détails, cette étude à fait l'objet d'un article qui a été publié dans une revue spécialisée. L'article a été ajouté en annexe à la fin du manuscrit.[324] Le plan d'analyse de l'étude est présenté dans la figure 16.

2.1 Définition du score polygénique

69 SNPs précédemment associés à l'eGFR ont été étudiés. Un SNP a été retiré pour chacune des paires de SNPs en déséquilibre de liaison et les 3 SNPs avec une qualité d'imputation faible (r^2<0,3) ont ensuite été retirés. Les 53 SNPs restants ont été testés individuellement dans un modèle de régression logistique ajusté sur le sexe, les quatre premières composantes principales de l'ACP basée sur l'IBS des individus et un paramètre de distinction entre MARTHA et EOVT. L'hypothèse départ étant qu'une fonction rénale diminuée est associée à un risque accru de thrombose veineuse, j'ai donc conservé les 28 SNPs pour lesquels l'allèle associé à une diminution de l'eGFR était également associé à un risque accru de thrombose veineuse. Ces SNPs ont été testés dans une modèle de régression multivarié ajusté sur le sexe, les quatre premières composantes principales et un paramètre de distinction entre MARTHA et EOVT et la combinaison de SNPs avec l'AIC le plus faible a été sélectionnée pour composer le score de risque polygénique. Le score final correspond à la somme des allèles associées à la diminution de l'eGFR des 9 SNPs sélectionnés sans pondération. L'effet individuel des 9 SNPs sélectionnés pour le score polygénique est présenté dans le tableau 7. Les 9 SNPs choisis constituent le groupe de SNPs le plus parcimonieux permettant de séparer les cas et les témoins de MARTHA/EOVT. Le SNP avec le signal le plus fort est rs807601 (p=0,011) dans le *locus DDX1* et le SNP ayant le signal le moins fort est rs6420094 (p=0,255) situé dans le *locus SLC34A1*.

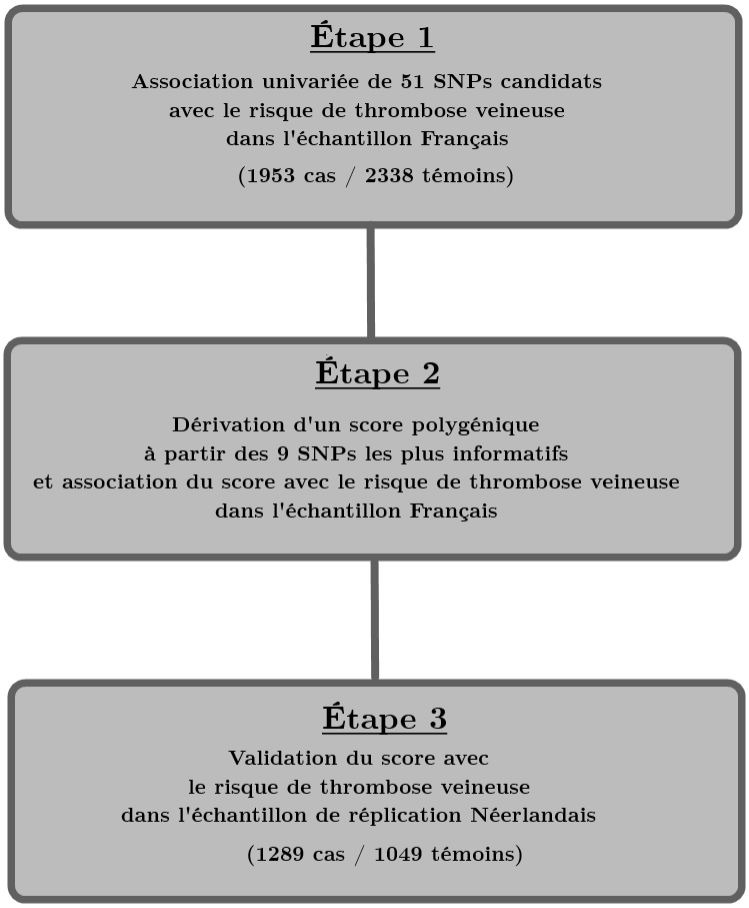


Figure 16 Protocole de l'étude. Les SNPs sont d'abord recueillis dans la bibliographie puis testés individuellement dans le groupe MARTHA/EOVT. Le score est ensuite composé en utilisant le groupe de SNPs permettant de minimiser l'AIC du modèle. Le score est ensuite testé dans le groupe MARTHA/EOVT puis répliqué dans le groupe MEGA.

				MARTHA/EOVT			MEGA		
SNP	Chr	Locus	EA[1]	EAF[2]	OR	P-value	EAF[2]	OR	P-value
rs807601	2	*DDX1*	G	0,65	1,14	0,011	0,66	1,17	0,288
rs1047891	2	*CPS1*	A	0,32	1,14	0,016	0,31	1,03	0,846
rs9682041	3	*SKIL*	T	0,86	1,15	0,051	0,86	0,96	0,821
rs6420094	5	*SLC34A1*	G	0,32	1,06	0,255	0,31	1,12	0,445
rs7759001	6	*ZNF204P*	A	0,77	1,10	0,100	0,77	1,37	0,068
rs3758086	8	*STC1,ADAM28*	A	0,41	1,06	0,243	0,39	1,19	0,204
rs9916302	17	*FBXL20*	T	0,70	1,09	0,116	0,71	1,15	0,373
rs12460876	19	*SLC7A9*	T	0,57	1,09	0,084	0,57	1,05	0,699
rs6088580	20	*PIGU,TP53INP2*	C	0,48	1,07	0,144	0,47	1,23	0,127

Tableau 7 Effet individuel des SNPs sélectionnés pour le score polygénique sur le risque de thrombose veineuse dans MARTHA/EOVT et MEGA. L'effet et la p-value ont été rapportés après ajustement sur le sexe, les quatre premières composantes principales et un paramètre de distinction entre MARTHA et EOVT.
(1) Allèle de référence associé à une diminution de l'eGFR dans l'étude publiée en 2016.[229]
(2) Fréquence de l'allèle de référence chez les groupes témoins de MARTHA/EOVT et MEGA.

2.2 Analyse dans l'échantillon Français

Le score polygénique des 9 SNPs a été calculé dans le groupe MARTHA/EOVT et la distribution du score au sein de ce groupe est présentée dans la figure 17A. Le score est significativement différent entre les cas et les témoins, avec des valeurs moyennes de 10,4±0,04 et 10,1±0,047 respectivement ($p=1{,}03 \times 10^{-7}$) Le score a ensuite été testé en association avec la thrombose veineuse dans un modèle de régression logistique ajusté sur le sexe, les quatre premières composantes principales et un paramètre de distinction entre MARTHA et EOVT. Une augmentation d'une unité du score polygénique est associée à une augmentation du risque de thrombose veineuse d'un facteur 1,09 [1,06-1,15] ($p=1{,}44 \times 10^{-7}$). La population a ensuite été divisée en quintiles en fonction de la distribution du score et le risque de thrombose veineuse en fonction du quintile a été estimé en prenant les 20% de la distribution avec le score le plus faible (1er quintile) comme référence. Les quintiles ont été testés dans un modèle de régression logistique ajusté sur le sexe, les quatre premières composantes principales et un paramètre de distinction entre MARTHA et EOVT. Le 2ème quintile n'est pas significativement associé à un risque accru de thrombose veineuse mais les 3ème, 4ème et 5ème quintiles sont associés à un risque de thrombose veineuses avec des OR de 1,39 [1,12-1,72], 1,32 [1,06-1,64] et 1,54 [1,24-1,92] respectivement. Les résultats de l'analyse en quintiles sont présentés dans la figure 17B.

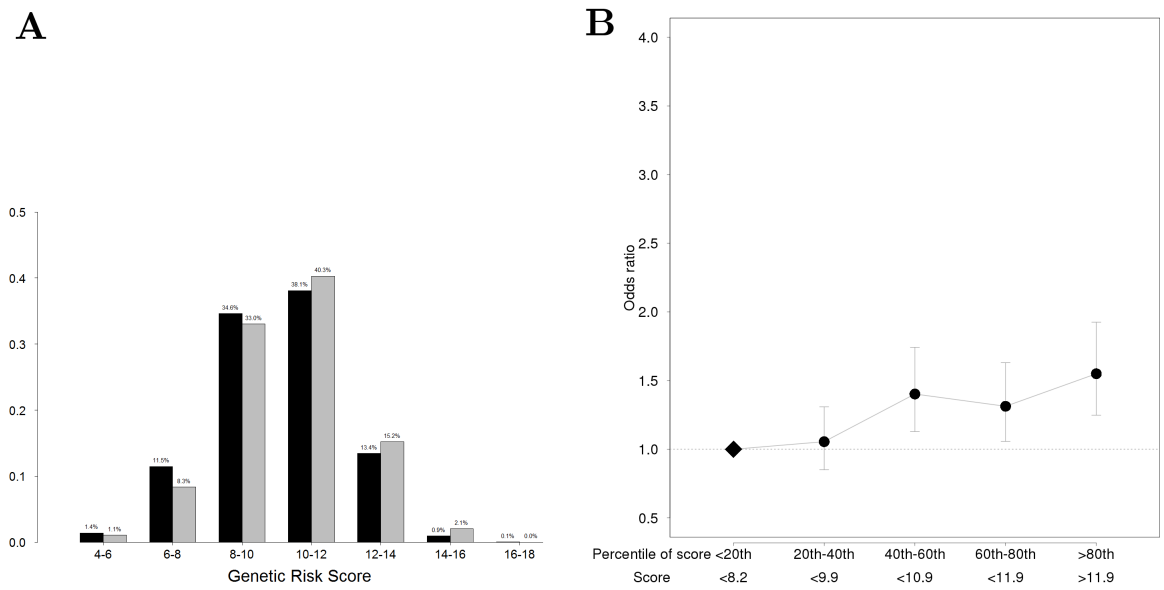


Figure 17 A Distribution du score dans la cohorte MARTHA/EOVT, les barres noires représentent la distribution des témoins tandis que les barres grises représentent la distribution des cas. **B** *Odds ratio* du risque de thrombose veineuse par quintile de score dans le groupe MARTHA/EOVT, le premier quintile est utilisé comme référence pour l'estimation du risque.

2.3 Analyse dans l'échantillon Néerlandais

Le score polygénique a été calculé à partir des 9 SNPs dans MEGA, une cohorte indépendante, en suivant la même méthode. La distribution du score dans MEGA est présentée dans la figure 18A. Les résultats trouvés dans l'analyse de MARTHA/EOVT ont ensuite été répliqués dans MEGA, le score a été testé avec le risque de thrombose veineuse dans un modèle de régression logistique ajusté sur le sexe et les quatre premières composantes principales. Le score polygénique est associé à un risque accru de thrombose veineuse d'un facteur 1,18 [1,11-1,26] ($p=8{,}86 \times 10^{-8}$). En utilisant les quintiles définis dans l'analyses de MARTHA/EOVT, MEGA a été divisée en groupes et ces groupes ont été testés en association avec le risque de thrombose veineuse en utilisant le groupe avec le score le plus faible comme référence. L'analyse a été réalisée avec le même modèle de régression que celui utilisé dans l'analyse avec la cohorte Française. Les 2ème et 3ème groupes ne sont pas significativement associés à un risque accru de thrombose veineuse mais les 4ème et 5ème le sont avec des facteurs de risque de 2,20 [1,54-3,14] et 2,23 [1,56-3,19] respectivement, les résultats de l'analyse en groupes sont présentés dans la figure 18B.

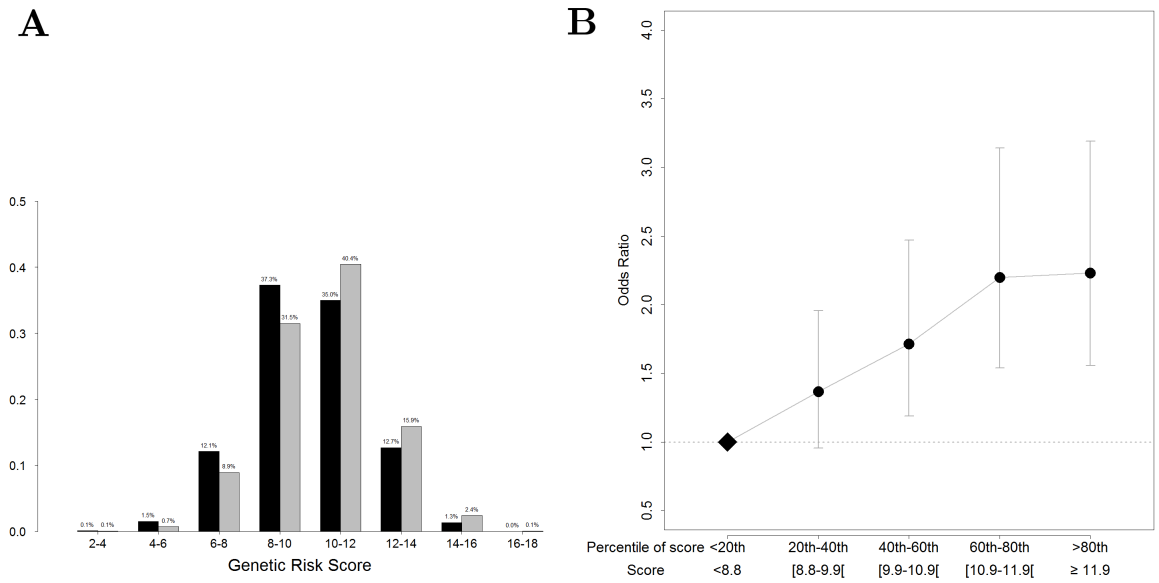


Figure 18 A Distribution du score dans la cohorte MEGA, les barres noires représentent la distribution des témoins tandis que les barres grises représentent la distribution des cas. **B** *Odds ratio* du risque de thrombose veineuse par quintile de score dans le groupe MEGA, le premier quintile est utilisé comme référence pour l'estimation du risque.

Les résultats de l'association du score avec la thrombose veineuse ne sont pas significativement différents entre MARTHA/EOVT et MEGA (p=0,06) et en combinant les résultats de MARTHA/EOVT et MEGA dans une méta-analyse, on obtient une augmentation du risque de thrombose veineuse d'un facteur 1,12 [1,09-1,16] ($p=3{,}77 \times 10^{-13}$). L'analyse par quintile des deux études combinées indique un risque accru de thrombose veineuse pour les 3ème, 4ème et 5ème groupes avec des facteurs de risque de 1,47 [1,22-1,77], 1,52 [1,26-1,82] et 1,70 [1,41-2,05] respectivement. Il est intéressant de noter qu'un des SNPs inclus dans le score polygénique, rs6088580, est en déséquilibre de liaison ($r^2=0{,}15$; D'=1) avec rs867186, un SNP du locus PROCR connu pour être associé au risque de thrombose veineuse.[325] Les analyses de l'effet du score sur le risque de thrombose veineuse après ajustement donne un score significativement associé au risque de thrombose veineuse avec un facteur de risque de 1,10 ($p=6{,}24 \times 10^{-7}$) dans MARTHA/EOVT et 1,17 ($p=1{,}20 \times 10^{-6}$) dans MEGA. En composant un score génétique sans rs6088580 et en répétant les analyses effectuées haut dessus, on obtient toujours un score associé au risque de thrombose veineuse, avec un facteur de risque de 1,10 ($p=7.08 \times 10^{-7}$) dans MARTHA/EOVT, 1,18 ($p=5.48 \times 10^{-7}$) dans MEGA et un facteur de risque de 1,13 [1,09-1,16] dans l'analyse combinée de MARTHA/EOVT et MEGA.

2.4 Confirmation de l'impact du score sur la fonction rénale

Les 9 SNPs du score polygénique avaient déjà été individuellement associés à l'eGFR dans des études antérieures. Cependant il fallait confirmer l'effet du score polygénique composé à partir de ces 9 SNPs sur l'eGFR. J'ai donc ensuite testés les 9 SNPs du score en association avec l'eGFR dans une autre cohortes indépendantes, Genesis/Genediab. L'analyse a été réalisée avec un modèle de régression linéaire ajusté sur l'âge, le sexe et les quatre premières composantes principales. Les résultats sont présentés dans le tableau 8. Un seul des 9 SNPs est associé à l'eGFR avec une p-value inférieure à 0,05 dans Genesis/Genediab, il s'agit de rs3758086 du locus *STC1* (p=0,035). Le score polygénique a ensuite été testé en association avec l'eGFR en utilisant le même modèle de régression linéaire et il est associé à une diminution de l'eGFR (β=-1,19, p=3,54x10^{-3}). L'association entre le score et l'eGFR a ensuite été testée dans la cohorte AtheroGene mais n'a pas montré de résultat significatif (β=-0,453, p=0,076).

SNP	EA[(1)]	Effect	Se	P-value
rs807601	G	-1,330	1,186	0,262
rs1047891	A	-1,015	1,302	0,435
rs9682041	T	-2,340	1,670	0,161
rs6420094	G	-0,569	1,178	0,629
rs7759001	A	-1,475	1,352	0,275
rs3758086	A	-2,345	1,109	0,035
rs9916302	T	-0,725	1,216	0,551
rs12460876	T	-0,575	1,085	0,596
rs6088580	C	-0,819	1,109	0,460
GRS	NA	-1,187	0,406	3,545x10^{-3}

Tableau 8 Effet individuel des SNPs sélectionnés pour le score polygénique sur l'eGFR dans Genesis/Genediab. L'effet et la *p-value* ont été rapportés après ajustement sur l'âge, le sexe et les quatre premières composantes principales.
(1) Allèle de référence associé à une diminution de l'eGFR dans l'étude publiée en 2016.[229]

2.5 Impact du score sur les facteurs de l'hémostase

Le score polygénique a ensuite été testé avec plusieurs paramètres biologiques de l'hémostase disponibles dans MARTHA et AtheroGene. Les seuls paramètres associés au score sont l'antithrombine dans MARTHA (p=4,01x10^{-3}) qui reste associée au score après le retrait des individus suivant un traitement anticoagulant (p=2x10^{-3}) et le taux de récepteurs endothéliaux soluble à la protéine C (sEPCR) dans AtheroGene (p=8,89x10^{-4}), mais cette association ne persiste pas après ajustement sur rs867186. Le score est associé au taux de fibrinogène dans

MARTHA (p=5x10^{-3}) même après retrait des individus suivant un traitement anticoagulant (p=7,74x10^{-4}) et aussi dans AtheroGene après exclusion des individus suivant un traitement anticoagulant (p=0,045). L'ensemble des analyses menées sur les paramètres de l'hémostase est présenté dans le tableau 9. L'association entre le score et le taux d'antithrombine est également observable en comparant le taux d'antithrombine moyen en fonction du quintile de score dans MARTHA (figure 19).

Phénotype	Étude	Effectifs	Effet	P-value
FVIIIc	MARTHA	1092	-1,03	0,20
vWF	MARTHA	1375	-0,3	0,75
Antithrombine	MARTHA	1403	-0,49	4,01x10^{-3}
Antithrombine**	MARTHA	897	-0,63	2x10^{-3}
PC	MARTHA	1445	-1,02	0,064
PS	MARTHA	1430	-0,21	0,60
Fibrinogène	MARTHA	1542	-0,03	5x10^{-3}
Fibrinogène**	MARTHA	985	-0,04	7,74x10^{-4}
TFPI	MARTHA	1170	0,18	0,23
FXI	MARTHA	1170	-0,005	0,32
Génération de thrombine	MARTHA	705	1,71	0,80
APTT	MARTHA	1537	0,042	0,67
Globules blancs	MARTHA	1540	0,015	0,59
Plaquettes	MARTHA	1540	-0,50	0,58
Plaquettes**	MARTHA	984	-1,58	0,13
Érythrocytes	MARTHA	1539	-0,004	0,46
HDL	MARTHA	1002	0,004	0,52
TFPI	AtheroGene	892	0,65	0,37
sEPCR	AtheroGene	892	4,62	8,89x10^{-4}
sEPCR*	AtheroGene	892	0,66	0,43
TAFIa/ai	AtheroGene	891	0,28	0,12
D Dimère	AtheroGene	790	-3,47	0,42
Fibrinogène	AtheroGene	1503	-0,02	0,27
Fibrinogène**	AtheroGene	704	-0,046	0,045
HDL	AtheroGene	1761	-0,32	0,059

Tableau 9 Effet du score polygénique sur les différents paramètres de l'hémostase et autres paramètres biologiques disponibles dans MARTHA et AtheroGene.
*Résultat après ajustement sur rs867186.
**Résultat après le retrait des patients traités avec des anticoagulants.

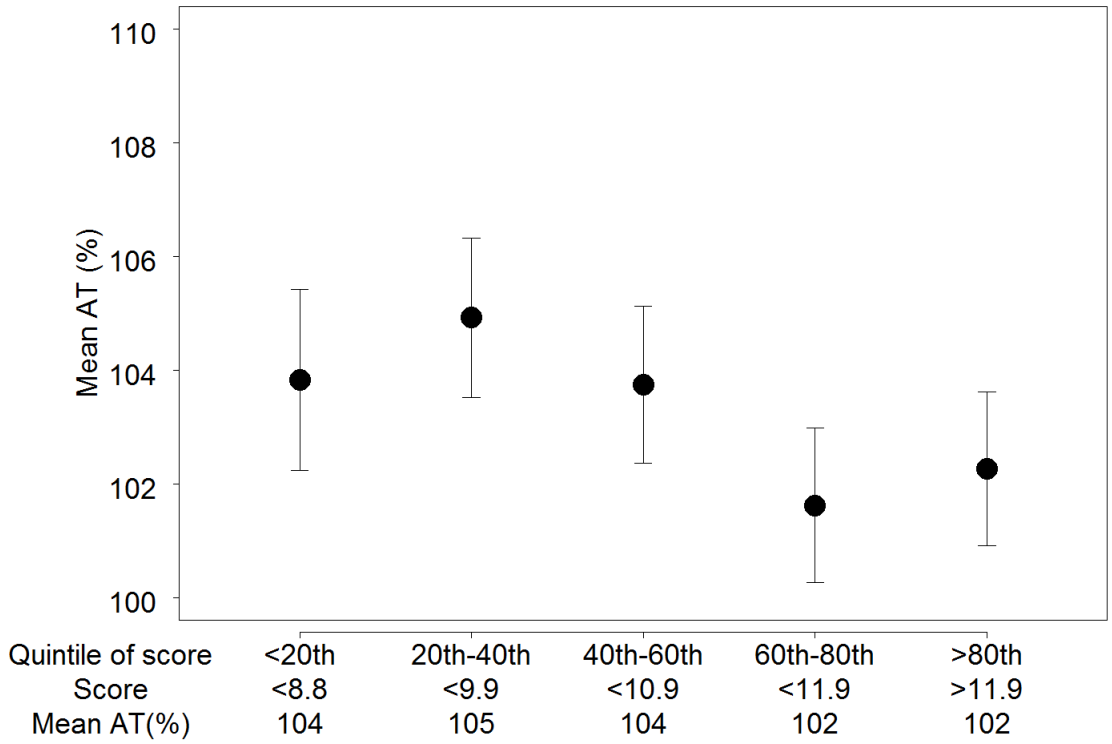


Figure 19 Taux d'antithrombine moyen par quintile de score polygénique dans la population MARTHA.

2.6 Discussion sur l'association entre le score génétique et le risque de MTEV et de sa pertinence vis-à-vis de la relation entre la fonction rénale et les MTEV

Les données obtenues à partir de ce travail sur la thrombose veineuse et les marqueurs de la fonction rénale ont permis de montrer une association entre un score allélique de neuf SNPs associés à l'eGFR et le risque de thrombose veineuse dans deux groupes cas/témoins indépendants. Malgré le fait qu'aucun de ces SNPs ne soit individuellement associé significativement au risque de thrombose veineuse, on constate qu'une accumulation de onze allèles à risque ou plus (score>11) est associée à un *odds ratio* de 1,42 pour le risque de thrombose veineuse. Afin de chercher des pistes expliquant l'association entre le score allélique et le risque de thrombose veineuse il est pertinent de regarder l'état des connaissances sur les neuf *loci* auxquels appartiennent ces SNPs.

rs807601 est un SNP intergénique proche du gène *DDX1*. Ce gène code pour une hélicase à motif DEAD-Box, qui aurait une activité hélicase spécifique au doubles-brins d'ARN et aux

hybrides ADN-ARN. Il a été montré que *DDX1* est nécessaire à l'expression des gènes *CCND2* et *CD9* dans des cellules de lignée TGTC (cellules de tumeur germinale testiculaire), ces gènes sont régulièrement sur-exprimés dans différents types de carcinomes.[326] Bien que ces informations ne semblent pas pertinentes dans le cadre de la fonction rénale il est intéressant de noter qu'un autre SNP sur le même *locus*, rs3755132, a été associé à la tumeur de Wilms.[327] La tumeur de Wilms est un cancer touchant les reins et qui survient généralement durant l'enfance. Il est accompagné de troubles de la fonction rénale comme de la protéinurie. Dans les populations HapMap et 1000 Génomes, les SNPs rs807601 et rs3755132 sont en léger déséquilibre de liaison (r^2=0,36 et D'=0,98), cependant l'allèle à risque rs807601(G) n'est quasiment jamais retrouvé sur le même haplotype que l'allèle à risque rs3755132(G) pour la tumeur de Wilms.

CPS1 est un gène codant pour la carbamoyl-phosphate synthase, une enzyme impliquée dans le métabolisme de l'azote, notamment la conversion de l'ammoniac en urée. Il n'y a pas d'élément pertinent permettant de lier ce gène à la fonction rénale dans les bases de données.

SKIL code pour une protéine *SKI-like*, cette protéine semble être un régulateur de la voie SMAD et donc de l'activation de la voie TGFβ.[328] Cependant les fonctions de la voie TGFβ sont si étendues qu'il serait difficile d'énoncer une hypothèse sur le rôle que pourrait jouer ce gène dans la fonction rénale.

SLC34A1 est vraisemblablement un transporteur de solutés qui serait impliqué dans la réabsorption du phosphate et d'autres solutés dans les tubules rénaux. Des mutations dans ce gènes ont été associées à des déficiences en phosphate dus à un manque de réabsorption dans les tubules rénaux mais aussi au syndrome de Fanconi, une maladie dans laquelle on observe un défaut de la réabsorption des solutés dans les tubules rénaux.[329, 330] Il est intéressant de constater que l'allèle A de rs6420094 a été associé à diminution d'antigènes du facteur VIII[331] et que le SNP rs7708314 du même *locus* est associé à l'aPTT,[332] ces éléments suggèrent que *SLC34A1* joue probablement un rôle dans le risque de thrombose veineuse, il est cependant plus difficile de lier ces élément à la déficience en phosphates.

ZNF204P est un pseudogène sans fonction ou activité connue. D'après les données disponibles sur GTeX portal il est cependant exprimé dans plusieurs tissus, principalement des organes du système nerveux. *STC1* code pour la stanniocalcine 1, une protéine que l'on sait capable de stimuler la réabsorption des phosphates chez les petits mammifères,[333] il n'y a cependant aucun élément permettant de confirmer ce rôle chez l'homme. Ces données permettraient de regrouper *STC1* et *SLC34A1* dans la même voie métabolique. Le produit d'*ADAM28* serait reconnu par les intégrines ITGA4/ITGB1,[334] ce qui suggère un rôle dans la migration des lymphocytes, il est difficile d'envisager un rôle dans la fonction rénale à partir de ces données.

FBXL20 aurait une activité régulatrice sur le largage des vésicules synaptiques.[335] Il n'y a aucun lien évident avec la fonction rénale ou la thrombose veineuse. *SLC7A9* est un transporteur responsable de la réabsorption des acides aminés basiques depuis le tubule rénal, plusieurs mutations de ce gène ont été reliées à un déficit de réabsorption de la cystine,[336] il est possible que ce SNP exerce une influence sur le risque de thrombose veineuse via un mécanisme impliquant l'homocystéine.

PIGU code pour une sous-unité du complexe GPI-transamidase qui est chargée de fixer l'ancre GPI sur les protéines cibles. *TP53INP2* est un gène inductible par p53 qui serait impliqué dans la régulation de l'autophagie. Pour le *loci PIGU-TP53INP2*, il est cependant probable que la contribution au risque de thrombose veineuse soit due au fait que l'allèle C de rs6088580 soit en phase avec l'allèle G de rs867186 (D'=1) et l'allèle G de rs764598 (D'=0,96) qui sont associés au taux circulant de protéine C et au risque de thrombose veineuse.[325, 337] L'effet de rs6088580 est donc sans doute représentatif d'un marqueur affectant le gène *PROCR*, codant pour le récepteur à la protéine C, qui est très proche de *PIGU*.

Certains *loci* présentés ici pourraient fournir des pistes intéressantes pour explorer les mécanismes reliant la fonction rénale au risque de thrombose veineuse. Il faut cependant rappeler que les SNPs sélectionnés ne sont sans doute pas les SNPs affectant l'eGFR. Ils peuvent en revanche être déséquilibre de liaison avec des SNPs situés sur des *loci* voisins ou affectant des gènes distants en modifiant des séquences régulatrices distantes. Les *loci* à proximité des SNPs utilisés pour le score n'ont à présent pas été associés au risque de thrombose veineuse dans des études pan-génomiques, à l'exception de *PIGU-TP53INP2* qui est voisin de *PROCR*. Nous avons cependant montré que le score reste significativement associé au risque de thrombose veineuse sans la contribution du SNP lié à ce *locus*. S'il est possible de confirmer que les SNPs liés à *SLC34A1*, *STC1* et *SLC7A9* ont bien un effet sur ces gènes, alors ils offrent des hypothèses intéressantes sur les mécanismes d'action par lesquels la fonction rénale affecte le risque de thrombose veineuse.

Les analyses d'associations entre le score et les paramètres de l'hémostase n'ont pas permis d'apporter plus d'éléments en faveur de l'hypothèse du lien entre l'eGFR et le taux de FVIII suggérée par d'autres résultats.[218] Ces analyses n'ont pas apporté d'élément solide en faveur d'un impact du score sur les paramètres de l'hémostase en général. Nous avons pu confirmer une association entre le score et l'eGFR dans une cohorte indépendante. En revanche il est nécessaire de souligner que le score a un impact très faible sur la variation de l'eGFR puisqu'il n'en explique que 0,67%. Cependant on peut tout de même observer une forte association. Ceci pourrait indiquer l'existence d'effets pléiotropiques venant du score et qu'il pourrait

donc être un marqueur imparfait de l'eGFR.

La stratégie employée présente cependant un avantage car le choix d'un score génétique comme marqueur d'un phénotype permet de garantir une certaine indépendance vis-à-vis des facteurs de confusions et permettre de fixer la relation de causalité car les marqueurs sont attribués à la naissance. Dans notre cas, nous avons montré que le score est indépendant des facteurs susceptibles d'affecter la relation entre l'eGFR et la risque de thrombose veineuse qui sont le sexe, l'âge et l'IMC. Malgré tout cette étude possède quelques limitations qui doivent être mentionnées. En premier lieu, l'eGFR des individus n'était pas disponible dans les groupes utilisés pour analyser l'association entre le score et le risque de thrombose veineuse, une analyse ajustée sur l'eGFR aurait permis de mettre en évidence des effets pléiotropiques du score. De plus, l'absence de données rénales en général ne nous a pas permis de vérifier l'association entre le score et la présence de maladie rénale chronique. Bien que les SNPs utilisés pour composer le score soient issus d'études pan-génomiques robustes, ils ne sont sans doute pas fonctionnels mais sont plutôt associés à des marqueurs fonctionnels qui n'ont pas été génotypés. Identifier des marqueurs fonctionnels directement liés à la fonction rénale pourrait permettre de composer une score plus représentatif de ce phénotype.

Pour finir, le manque de résultats significatifs concernant l'association entre le score et les facteurs de l'hémostase nous laisse peu d'hypothèses pour expliquer l'effet de ces SNPs sur le risque de thrombose veineuse. Sous réserve que ces résultats soient répliqués dans une autre étude et que des solutions soient apportées pour remédier aux limitations mentionnées plus haut, il serait intéressant d'analyser l'impact de ce score sur la récidive des événements de thrombose veineuse. En cas de réplication, il serait également intéressant de s'attarder sur l'impact de ce score sur le taux d'homocystéine et de phosphates dans le plasma ainsi que sur l'effet qu'aurait un ajustement sur ces paramètres sur l'association en le score et le risque de thrombose veineuse. Un dernier détail concerne la pondération du score, traditionnellement, chaque variation inclue dans le score est pondérée par son effet sur phénotype d'intérêt ce qui permet de prendre en compte la contribution au risque de chaque allèle. Dans le cas de ce score polygénique, étant donné que les SNPs sont issus de plusieurs études indépendantes sur des populations distinctes, le choix de ne pas appliquer de pondération a été fait afin d'éviter de mélanger des effets qui pourraient être différents dans les différentes populations testées.

CHAPITRE 4

Discussion

Ces travaux avaient chacun un objectif, identifier des marqueurs génétiques du risque de coronaropathie spécifique au contexte du diabète de type I et clarifier le lien de causalité entre la fonction rénale et le risque de thrombose veineuse. Ces résultats dans leur ensemble devaient ensuite permettre d'apporter de nouvelles informations concernant le rôle de la fonction rénale dans le risque de développer des maladies cardiovasculaires. Après avoir discuté des résultats de ces travaux indépendamment j'aborde dans cette partie l'interprétation de l'ensemble des résultats au regard du lien entre la fonction rénale et les maladies cardiovasculaires ainsi que les perspectives envisageables après ces travaux.

1 Intégration des résultats au regard du lien entre la fonction rénale et les maladies cardiovasculaires

Il est maintenant temps de regrouper les données des deux études afin d'énumérer les éléments nouveaux qu'elles apportent à la relation entre la fonction rénale et le risque de maladie cardiovasculaire. La première étude apporte finalement peu d'éléments sur la relation entre la fonction rénale et le risque de coronaropathie en raison du fait que l'analyse d'interaction des marqueurs avec le statut de néphropathie diabétique n'a pas pu être menée à bien. *PKD1* offre cependant un candidat intéressant étant donné son rôle dans le développement des tubules rénaux et les conséquences d'une perte de fonction du gène. Il est cependant impossible d'avancer des hypothèses sur son rôle dans le risque de maladie cardiovasculaire avec si peu d'informations. Dans ce contexte, il serait intéressant d'estimer l'impact de rs116092985 sur le risque de néphropathie diabétique, l'eGFR ou la protéinurie puis d'effectuer une analyse de randomisation Mendelienne avec les phénotypes rénaux et le risque de coronaropathie. Ce type d'analyse permettrait de déterminer si le rôle de ce *locus* est porté par son action sur la fonction rénale ou non.

Dans la deuxième étude, nous avons montré qu'un score allélique composé de marqueurs génétiques de l'eGFR était associé au risque de thrombose veineuse. Cette association est un

élément en faveur de l'hypothèse déjà avancée par d'autres études épidémiologiques que la dégradation de la fonction rénale serait un facteur de risque de thrombose veineuse. Malgré le fait que les analyses de l'effet du score sur les paramètres de l'hémostase ne permettent pas de mettre en évidence de mécanisme explicatif reliant la thrombose à la fonction rénale et bien qu'il n'y ait pas d'élément permettant d'exclure un effet pléiotropique du score, nous avons tout de même des données intéressantes. Plusieurs *loci* attachés à ce score portent des gènes exprimés dans les tubules rénaux et impliqués dans la réabsorption des molécules depuis le filtrat vers le plasma, c'est le cas de *SLC34A1* et *STC1* qui sont impliqués dans la réabsorption des phosphates et *SLC7A9* qui est impliqué dans la réabsorption de la cystine et des acides aminés basiques. Les gènes listés ci-dessus semblent tous jouer un rôle au niveau des tubules rénaux qui sont le lieu où sont réabsorbées les molécules filtrées nécessaires comme le glucose, les solutés ioniques (sel, phosphates, calcium) et les acides aminés mais aussi où sont excrétés les métabolites devant être éliminés.

Ces données nous permettent de penser qu'une dysfonction des tubules rénaux pourrait entrainer un déficit de réabsorption ou de sécrétion d'un soluté qui pourrait à son tour augmenter le risque de maladie cardiovasculaire. C'est notamment au niveau du tubule rénal que le rein exerce son contrôle sur le métabolisme de l'homocystéine, qui est associée au risque d'événement cardiovasculaire. C'est également dans les tubules rénaux que le phosphate est réabsorbé, permettant à l'organisme de maintenir son taux de phosphates, or chez l'animal on observe une diminution du nombre d'événements de thromboses après injection d'une solution concentrée en phosphates.[338] Ces résultats suggèrent que l'activité des tubules rénaux pourrait influencer le risque de coronaropathie ou de thrombose veineuse, possiblement par le biais d'un soluté sécrété ou réabsorbé.

2 Perspectives

Le protocole définit pour mes travaux sur le risque de coronaropathie chez les patients diabétiques de type I envisageait une méta-analyse exploratoire avec les études Genesis/Genediab, Steno et UK-ROI puis une étape de réplication avec les études DCCT/EDIC, WESDR, CACTI, EDC et GokinD. En raison des contraintes temporelles de la thèse je n'ai pas dévié de ce plan de départ. En revanche il serait pertinent de combiner les résultats de ces huit cohortes dans une méta-analyse afin d'augmenter la puissance statistique de l'étude. Cette méta-analyse pourrait être la prochaine étape de ces travaux d'autant plus que les données génétiques des huit cohortes sont déjà disponibles. Il faudrait cependant trouver des cohortes adéquates pour une étape de réplication pour cette nouvelle méta-analyse, les cohortes de FinnDiane[339] (Finlande) et SDRNT1BIO[340] (Écosse) pourraient être de bons candidats car il s'agit de cohortes de patients diabétiques de type I d'origine Européenne qui ont déjà été

employées dans des travaux de génétique. Une méta-analyse combinant plus d'études permettrait d'obtenir des résultats plus robustes.

Sous réserve qu'il soit possible de répliquer les résultats de ces travaux, il y aurait alors plusieurs voies intéressantes à explorer par la suite. D'abord, il serait intéressant de poursuivre ces études avec les analyses déjà suggérées plus haut, à savoir une analyse stratifiée sur la néphropathie diabétique pour le risque de coronaropathie et des analyses de l'effet de rs116092985 (*PKD1*) sur des paramètres de la fonction rénale et une analyse de randomisation Mendelienne. Il serait intéressant d'effectuer ses analyses sur l'infarctus du myocarde plutôt que sur les coronaropathies en général, au risque de perdre de la puissance, car l'infarctus du myocarde correspond au moment où les facteurs de coagulation interviennent dans la coronaropathie.

Pour le score, les analyses pourraient être complétées avec des analyses d'haplotypes et afin d'identifier des haplotypes plus à risque. Rechercher les marqueurs causaux associés aux marqueurs utilisés pour le score peut s'avérer fastidieux mais pourrait permettre d'obtenir un score ayant un meilleur potentiel prédictif. Le score peut également être testé en association avec la récidive de thrombose veineuse et avec d'autres maladies cardiovasculaires. Afin de mieux caractériser la fonction des marqueurs identifiés, il y a ensuite la possibilité d'effectuer diverses analyses bioinformatiques, des études de type eQTL (*expression quantitative trait loci*) et 3C (*chromatin configuration capture*) permettraient de mettre en évidence un impact des marqueurs identifiés sur l'expression d'un ou plusieurs gènes et de mettre en évidence l'existence d'interactions à longue distance. Des analyses sur les marques de méthylation et éventuellement des études de Chip-Seq (*chromatin immunoprecipitation sequencing*) pourraient permettre d'observer une impact épigénétique de ces marqueurs.

Toutes ces analyses pourraient ensuite être étendues à un ensemble plus large de maladies cardiovasculaires afin de généraliser le lien entre la fonction rénale et ces pathologies. Il est a noter les analyses de score génétique et de randomisation Mendelienne offrent un bon moyen d'étudier les relations entre divers phénotypes et le risque de pathologie, il serait donc possible de répéter ces analyses avec des phénotypes tels que la pression artérielle et la fonction hépatique afin d'estimer leur impact sur les maladies cardiovasculaires.

3 Conclusion

Pour conclure, ces analyses nous ont permis de mettre en évidence des marqueurs génétiques associés au risque de coronaropathie chez les individus diabétiques de type I et au risque de thrombose veineuse. Une partie de ces marqueurs sont situés dans des *loci* impliqués dans la

fonction rénale et/ou sont associés à des phénotypes rénaux, ce qui permet d'apporter des éléments suggérant que la perte de fonction rénale augmente le risque de maladie cardiovasculaire. Les mécanismes par lesquels la fonction rénale a un impact sur le risque de maladie cardiovasculaire n'ont pas pu être élucidés mais les données suggèrent que la capacité des reins à réguler l'homéostasie de certains solutés pourrait être une voie d'étude importante.

Bibliographie

[1] Shanthi Mendis ; Pekka Puska ; Bo Norrving. Global Atlas on cardiovascular disease prevention and control. World Health Organization ; World Health Federation ; World Stroke Organization, Geneva ; Switzerland, 2011.

[2] Raja B Singh ; Sushma A Mengi ; Yan-Jun Xu ; Amarjit S Arneja ; Naranjan S Dhalla. Pathogenesis of atherosclerosis : A multifactorial process. Exp Clin Cardiol, 7(1) :40–53, 2002.

[3] Russell Ross. The pathogenesis of atherosclerosis : a perspective for the 1990s. Nature, 362 :801–809, April 1993.

[4] Bruce Furie ; Barbara C. Furie. Mechanisms of thrombus formation. N Engl J Med, 359(9) :938–949, August 2008.

[5] McGill HC Jr ; McMahan CA ; Gidding SS. Preventing heart disease in the 21st century : implications of the pathobiological determinants of atherosclerosis in youth (pday) study. Circulation, 117(9) :1216–1227, March 2008.

[6] Sekar Kathiresan ; Deepak Srivastava. Genetics of human cardiovascular disease. Cell, 148(6) :1242–1257, March 2012.

[7] Nikpay M ; Goel A ; Won HH ; Hall LM ; Willenborg C ; Kanoni S ; Saleheen D ; Kyriakou T ; Nelson CP ; Hopewell JC ; Webb TR ; Zeng L ; Dehghan A ; Alver M ; Armasu SM ; Auro K ; Bjonnes A ; Chasman DI ; Chen S ; Ford I ; Franceschini N ; Gieger C ; Grace C ; Gustafsson S ; Huang J ; Hwang SJ ; Kim YK ; Kleber ME ; Lau KW ; Lu X ; Lu Y ; Lyytikäinen LP ; Mihailov E ; Morrison AC ; Pervjakova N ; Qu L ; Rose LM ; Salfati E ; Saxena R ; Scholz M ; Smith AV ; Tikkanen E ; Uitterlinden A ; Yang X ; Zhang W ; Zhao W ; de Andrade M ; de Vries PS ; van Zuydam NR ; Anand SS ; Bertram L ; Beutner F ; Dedoussis G ; Frossard P ; Gauguier D ; Goodall AH ; Gottesman O ; Haber M ; Han BG ; Huang J ; Jalilzadeh S ; Kessler T ; König IR ; Lannfelt L ; Lieb W ; Lind L ; Lindgren CM ; Lokki ML ; Magnusson PK ; Mallick NH ; Mehra N ; Meitinger T ; Memon FU ; Morris AP ; Nieminen MS ; Pedersen NL ; Peters A ; Rallidis LS ; Rasheed A ; Samuel M ; Shah SH ; Sinisalo J ; Stirrups KE ; Trompet S ; Wang L ; Zaman KS ; Ardissino D ; Boerwinkle E ; ; Borecki IB ; Bottinger EP ; Buring JE ; Chambers JC ; Collins R ; Cupples LA ; Danesh J ; Demuth I ; Elosua R ; Epstein SE ; Esko T ; Feitosa MF ; Franco OH ; Franzosi MG ; Granger CB ; Gu D ; Gudnason V ; Hall AS ; Hamsten A ; Harris TB ; Hazen SL ; Hengstenberg C ; Hofman A ; Ingelsson E ; Iribarren C ; Jukema JW ; Karhunen PJ ; Kim BJ ; Kooner JS ; Kullo IJ ; Lehtimäki T ; Loos RJF ; Melander O ; Metspalu A ; März W ; Palmer CN ; Perola M ; Quertermous T ; Rader DJ ; Ridker PM ; Ripatti S ; Roberts R ; Salomaa V ; Sanghera DK ; Schwartz SM ; Seedorf U ; Stewart AF ; Stott DJ ; Thiery J ; Zalloua PA ; O'Donnell CJ ; Reilly MP ; Assimes TL ; Thompson JR ; Erdmann J ; Clarke R ; Watkins H ; Kathiresan S ; McPherson R ; Deloukas P ; Schunkert H ; Samani NJ ; Farrall M. A comprehensive 1,000 genomes-based genomde associate-wiion meta-analysis of coronary artery disease. Nat Genet, 47(10) :1121–1130, October 2015.

[8] https ://www.nhlbi.nih.gov/health/health-topics/topics/cad/signs.

[9] Kumar V ; Butcher SJ ; Öörni K ; Engelhardt P ; Heikkonen J ; Kaski K ; Ala-Korpela M ; Kovanen PT. Three-dimensional cryoem reconstruction of native ldl particles to 16Å resolution at physiological body temperature. PLoS One, 6(5) :e18841, May 2011.

[10] Sparrow CP ; Olszewski J. Cellular oxidation of low density lipoprotein is caused by thiol production in media containing transition metal ions. J Lipid Res, 34(7) :1219–1228, July 1993.

[11] Schwartz CJ ; Valente AJ ; Sprague EA ; Kelley JL ; Cayatte AJ ; Mowery J. Atherosclerosis. potential targets for stabilization and regression. Circulation, 86(6 Suppl) :III117–123, December 1992.

[12] Miller JD. Cardiovascular calcification : Orbicular origins. Nat Mater, 12(6) :476–478, June 2013.

[13] Tobin KJ. Stable angina pectoris : what does the current clinical evidence tell us ? J Am Osteopath Assoc, 110(7) :364–370, July 2010.

[14] Virmani R ; Burke AP ; Farb A ; Kolodgie FD. Pathology of the vulnerable plaque. J Am Coll Cardiol, 47(8 Suppl) :C13–18, April 2006.

[15] Collège National des Enseignants de Cardiologie ; Société française de cardiologie. Cardiologie. ELSEVIER / MASSON, Paris ; France, 2015.

[16] Shanthi Mendis ; P. Nordet ; J.E. Fernandez-Britto ; N. Sternby ;Pathobiological Determinants of Atherosclerosis in Youth(PBDAY) Research Group. Atherosclerosis in children and young adults : An overview of the world health organization and international society and federation of cardiology study on pathobiological determinants of atherosclerosis in youth study (1985–1995). Prevention and Control, 1(1) :3–15, March 2005.

[17] Yeghiazarians Y ; Braunstein JB ; Askari A ; Stone PH. Unstable angina pectoris. N Engl J Med, 342(2) :101–114, January 2000.

[18] Buja LM. Myocardial ischemia and reperfusion injury. Cardiovasc Pathol, 14(4) :170–175, July 2005.

[19] Kutty RS ; Jones N ; Moorjani N. Mechanical complications of acute myocardial infarction. Cardiol Clin, 31(4) :519–531, November 2013.

[20] https ://www.nhlbi.nih.gov/health/health-topics/topics/cad/diagnosis.

[21] Yasue H ; Nakagawa H ; Itoh T ; Harada E ; Mizuno Y. Coronary artery spasm–clinical features, diagnosis, pathogenesis, and treatment. J Cardiol, 51(1) :2–17, February 2008.

[22] Deb S ; Wijeysundera HC ; Ko DT ; Tsubota H ; Hill S ; Fremes SE. Coronary artery bypass graft surgery vs percutaneous interventions in coronary revascularization : a systematic review. JAMA, 310(19) :2086–2095, November 2013.

[23] Zhao DF ; Edelman JJ ; Seco M ; Bannon PG ; Wilson MK ; Byrom MJ ; Thourani V ; Lamy A ; Taggart DP ; Puskas JD ; Vallely MP. Coronary artery bypass grafting with and without manipulation of the ascending aorta : A network meta-analysis. J Am Coll Cardiol, 69(8) :924–936, February 2017.

[24] Moran AE ; Forouzanfar MH ; Roth GA ; Mensah GA ; Ezzati M ; Murray CJ ; Naghavi M. Temporal trends in ischemic heart disease mortality in 21 world regions, 1980 to 2010 : the global burden of disease 2010 study. Circulation, 129(14) :1483–1492, April 2014.

[25] Moran AE ; Forouzanfar MH ; Roth GA ; Mensah GA ; Ezzati M ; Murray CJ ; Naghavi M. The global burden of ischemic heart disease in 1990 and 2010 : the global burden of disease 2010 study. Circulation, 129(14) :1493–1501, April 2014.

[26] GBD 2015 Mortality and Causes of Death Collaborators. Global, regional, and national life expectancy, all-cause mortality, and cause-specific mortality for 249 causes of death, 1980-2015 : a systematic analysis for the global burden of disease study 2015. Lancet, 388(10053) :1459–1544, October 2016.

[27] GBD 2015 Disease, Injury Incidence, and Prevalence Collaborators. Global, regional, and national incidence, prevalence, and years lived with disability for 310 diseases and injuries, 1990-2015 : a systematic analysis for the global burden of disease study 2015. Lancet, 388(10053) :1545–1602, October 2016.

[28] Finegold JA ; Asaria P ; Francis DP. Mortality from ischaemic heart disease by country, region, and age : statistics from world health organisation and united nations. Int J Cardiol, 168(2) :934–945, September 2013.

[29] Yusuf S ; Hawken S ; Ounpuu S ; Dans T ; Avezum A ; Lanas F ; McQueen M ; Budaj A ; Pais P ; Varigos J ; Lisheng L ; INTERHEART Study Investigators. Effect of potentially modifiable risk factors associated with myocardial infarction in 52 countries (the interheart study) : case-control study. Lancet, 364(9438) :937–952, September 2004.

[30] Calderon R Jr ; Schneider RH ; Alexander CN ; Myers HF ; Nidich SI ; Haney. Stress, stress reduction and hypercholesterolemia in african americans : a review. Ethn Dis, 9(3) :451–462, 1999.

[31] World Health Organization. Global status report on noncommunicable diseases 2010. World Health Organization, Geneva ; Switzerland, 2010.

[32] Njølstad I ; Arnesen E ; Lund-Larsen PG. Smoking, serum lipids, blood pressure, and sex differences in myocardial infarction. a 12-year follow-up of the finnmark study. Circulation, 93(3) :450–456, February 1996.

[33] Lewington S ; Clarke R ; Qizilbash N ; Peto R ; Collins R ; Prospective Studies Collaboration. Age-specific relevance of usual blood pressure to vascular mortality : a meta-analysis of individual data for one million adults in 61 prospective studies. Lancet, 360(9349) :1903–1913, December 2002.

[34] Prospective Studies Collaboration ; Whitlock G ; Lewington S ; Sherliker P ; Clarke R ; Emberson J ; Halsey J ; Qizilbash N ; Collins R ; Peto R. Body-mass index and cause-specific mortality in 900 000 adults : collaborative analyses of 57 prospective studies. Lancet, 373(9669) :1083–1096, March 2009.

[35] Emerging Risk Factors Collaboration ; Sarwar N ; Gao P ; Seshasai SR ; Gobin R ; Kaptoge S ; Di Angelantonio E ; Ingelsson E ; Lawlor DA ; Selvin E ; Stampfer M ; Stehouwer CD ; Lewington S ; Pennells L ; Thompson A ; Sattar N ; White IR ; Ray KK ; Danesh J. Diabetes mellitus, fasting blood glucose concentration, and risk of vascular disease : a collaborative meta-analysis of 102 prospective studies. Lancet, 375(9733) :2215–2222, June 2010.

[36] Rao Kondapally Seshasai S ; Kaptoge S ; Thompson A ; Di Angelantonio E ; Gao P ; Sarwar N ; Whincup PH ; Mukamal KJ ; Gillum RF ; Holme I ; Njølstad I ; Fletcher A ; Nilsson P ; Lewington S ; Collins R ; Gudnason V ; Thompson SG ; Sattar N ; Selvin E ; Hu FB ; Danesh J ; Emerging Risk Factors Collaboration. Diabetes mellitus, fasting glucose, and risk of cause-specific death. N Engl J Med, 364(9) :829–841, March 2011.

[37] Lee IM ; Shiroma EJ ; Lobelo F ; Puska P ; Blair SN ; Katzmarzyk PT ; Lancet Physical Activity Series Working Group. Effect of physical inactivity on major non-communicable diseases worldwide : an analysis of burden of disease and life expectancy. Lancet, 380(9838) :219–229, July 2012.

[38] RSKannel WB ; Vasan RS. Triglycerides as vascular risk factors : new epidemiologic insights. Curr Opin Cardiol, 24(5) :345–350, July 2009.

[39] Cholesterol Treatment Trialists' (CTT) Collaboration ; Baigent C ; Blackwell L ; Emberson J ; Holland LE ; Reith C ; Bhala N ; Peto R ; Barnes EH ; Keech A ; Simes J ; Collins R. Efficacy and safety of more intensive lowering of ldl cholesterol : a meta-analysis of data from 170,000 participants in 26 randomised trials. Lancet, 376(9753) :1670–1681, November 2010.

[40] Stampfer MJ ; Krauss RM ; Ma J ; Blanche PJ ; Holl LG ; Sacks FM ; Hennekens CH. A prospective study of triglyceride level, low-density lipoprotein particle diameter, and risk of myocardial infarction. JAMA, 276(11) :882–888, September 1996.

[41] Brown MS ; Goldstein JL. Biomedicine. lowering ldl–not only how low, but how long ? Science, 311(5768) :1721–1723, March 2006.

[42] Ference BA ; Yoo W ; Alesh I ; Mahajan N ; Mirowska KK ; Mewada A ; Kahn J ; Afonso L ; Williams KA Sr ; Flack JM. Effect of long-term exposure to lower low-density lipoprotein cholesterol beginning early in life on the risk of coronary heart disease : a mendelian randomization analysis. J Am Coll Cardiol, 60(25) :2631–2639, December 2012.

[43] Mahmood SS ; Levy D ; Vasan RS ; Wang TJ. The framingham heart study and the epidemiology of cardiovascular disease : a historical perspective. Lancet, 383(9921) :999–1008, March 2014.

[44] Kivimäki M ; Nyberg ST ; Batty GD ; Fransson EI ; Heikkilä K ; Alfredsson L ; Bjorner JB ; Borritz M ; Burr H ; Casini A ; Clays E ; De Bacquer D ; Dragano N ; Ferrie JE ; Geuskens GA ; Goldberg M ; Hamer M ; Hooftman WE ; Houtman IL ; Joensuu M ; Jokela M ; Kittel F ; Knutsson A ; Koskenvuo M ; Koskinen A ; Kouvonen A ; Kumari M ; Madsen IE ; Marmot MG ; Nielsen ML ; Nordin M ; Oksanen T ; Pentti J ; Rugulies R ; Salo P ; Siegrist J ; Singh-Manoux A ; Suominen SB ; Väänänen A ; Vahtera J ; Virtanen M ; Westerholm PJ ; Westerlund H ; Zins M ; Steptoe A ; Theorell T ; IPD-Work Consortium. Job strain as a risk factor for coronary heart disease : a collaborative meta-analysis of individual participant data. Lancet, 380(9852) :1491–1497, October 2012.

[45] imäki M ; Jokela M ; Nyberg ST ; Singh-Manoux A ; Fransson EI ; Alfredsson L ; Bjorner JB ; Borritz M ; Burr H ; Casini A ; Clays E ; De Bacquer D ; Dragano N ; Erbel R ; Geuskens GA ; Hamer M ; Hooftman WE ; Houtman IL ; Jöckel KH ; Kittel F ; Knutsson A ; Koskenvuo M ; Lunau T ; Madsen IE ; Nielsen ML ; Nordin M ; Oksanen T ; Pejtersen JH ; Pentti J ; Rugulies R ; Salo P ; Shipley MJ ; Siegrist J ; Steptoe A ; Suominen SB ; Theorell T ; Vahtera J ; Westerholm PJ ; Westerlund H ; O'Reilly D ; Kumari M ; Batty GD ; Ferrie JE ; Virtanen M ; IPD-Work Consortium. Long working hours and risk of coronary heart disease and stroke : a systematic review and meta-analysis of published and unpublished data for 603,838 individuals. Lancet, 386(10005) :1739–1746, October 2015.

[46] Charlson FJ ; Moran AE ; Freedman G ; Norman RE ; Stapelberg NJ ; Baxter AJ ; Vos T ; Whiteford HA. The contribution of major depression to the global burden of ischemic heart disease : a comparative risk assessment. BMC Med, 11 :250, November 2013.

[47] Grant PJ. The genetics of atherothrombotic disorders : a clinician's view. J Thromb Haemost, 1(7) :1381–1390, July 2003.

[48] Morange PE ; Bickel C ; Nicaud V ; Schnabel R ; Rupprecht HJ ; Peetz D ; Lackner KJ ; Cambien F ; Blankenberg S ; Tiret L ; AtheroGene Investigators. Arterioscler thromb vasc biol. Arterioscler Thromb Vasc Biol, 26(12) :2793–2799, December 2006.

[49] Kyu HH ; Bachman VF2 ; Alexander LT ; Mumford JE ; Afshin A ; Estep K ; Veerman JL3 ; Delwiche K4 ; Iannarone ML ; Moyer ML ; Cercy K ; Vos T ; Murray CJ ; Forouzanfar MH. Physical activity and risk of breast cancer, colon cancer, diabetes, ischemic heart disease, and ischemic stroke events : systematic review and dose-response meta-analysis for the global burden of disease study 2013. BMJ, 354 :i3857, August 2016.

[50] Wang X ; Ouyang Y ; Liu J ; Zhu M ; Zhao G ; Bao W ; Hu FB. Fruit and vegetable consumption and mortality from all causes, cardiovascular disease, and cancer : systematic review and dose-response meta-analysis of prospective cohort studies. BMJ, 349 :g4490, July 2014.

[51] Lopez-Garcia E ; Schulze MB ; Meigs JB ; Manson JE ; Rifai N ; Stampfer MJ ; Willett WC ; Hu FB. Consumption of trans fatty acids is related to plasma biomarkers of inflammation and endothelial dysfunction. J Nutr, 135(3) :562–566, March 2005.

[52] Pedersen LR ; Olsen RH ; Jürs A ; Anholm C ; Fenger M ; Haugaard SB ; Prescott E. A randomized trial comparing the effect of weight loss and exercise training on insulin sensitivity and glucose metabolism in coronary artery disease. Metabolism, 64(10) :1298–1307, October 2015.

[53] Hermanson B ; Omenn GS ; Kronmal RA ; Gersh BJ. Beneficial six-year outcome of smoking cessation in older men and women with coronary artery disease. results from the cass registry. N Engl J Med, 319(21) :1365–1369, November 1988.

[54] Taylor F ; Huffman MD ; Macedo AF ; Moore TH ; Burke M ; Davey Smith G ; Ward K ; Ebrahim S. Statins for the primary prevention of cardiovascular disease. Cochrane Database Syst Rev, page CD004816, January 2013.

[55] Smith SC Jr ; Benjamin EJ ; Bonow RO ; Braun LT ; Creager MA ; Franklin BA ; Gibbons RJ ; Grundy SM ; Hiratzka LF ; Jones DW ; Lloyd-Jones DM ; Minissian M ; Mosca L ; Peterson ED ; Sacco RL ; Spertus J ; Stein JH ; Taubert KA. Aha/accf secondary prevention and risk reduction therapy for patients with coronary and other atherosclerotic vascular disease : 2011 update : a guideline from the american heart association and american college of cardiology foundation endorsed by the world heart federation and the preventive cardiovascular nurses association. J Am Coll Cardiol, 58(23) :2432–2446, November 2011.

[56] Freemantle N ; Cleland J ; Young P ; Mason J ; Harrison J. beta blockade after myocardial infarction : systematic review and meta regression analysis. BMJ, 318(7200) :1730–1737, June 1999.

[57] McPherson R ; Tybjaerg-Hansen A. Genetics of coronary artery disease. Circ Res, 118(4) :564–578, February 2016.

[58] Dai X ; Wiernek S ; Evans JP ; Runge MS. Genetics of coronary artery disease and myocardial infarction. World J Cardiol, 8(1) :1–23, January 2016.

[59] Webb TR ; Erdmann J ; Stirrups KE ; Stitziel NO ; Masca NG ; Jansen H ; Kanoni S ; Nelson CP ; Ferrario PG ; König IR ; Eicher JD ; Johnson AD ; Hamby SE ; Betsholtz C ; Ruusalepp A ; Franzén O ; Schadt EE ; Björkegren JL ; Weeke PE ; Auer PL ; Schick UM ; Lu Y ; Zhang H ; Dube MP ; Goel A ; Farrall M ; Peloso GM ; Won HH ; Do R ; van Iperen E ; Kruppa J ; Mahajan A ; Scott RA ; Willenborg C ; Braund PS ; van Capelleveen JC ; Doney AS ; Donnelly LA ; Asselta R ; Merlini PA ; Duga S ; Marziliano N ; Denny JC ; Shaffer C ; El-Mokhtari NE ; Franke A ; Heilmann S ; Hengstenberg C ; Hoffmann P ; Holmen OL ; Hveem K ; Jansson JH ; Jöckel KH ; Kessler T ; Kriebel J ; Laugwitz KL ; Marouli E ; Martinelli N ; McCarthy MI ; Van Zuydam NR ; Meisinger C ; Esko T ; Mihailov E ; Escher SA ; Alver M ; Moebus S ; Morris AD ; Virtamo J ; Nikpay M ; Olivieri O ; Provost S ; AlQarawi A ; Robertson NR ; Akinsansya KO ; Reilly DF ; Vogt TF ; Yin W ; Asselbergs FW ; Kooperberg C ; Jackson RD ; Stahl E ; Müller-Nurasyid M ; Strauch K ; Varga TV ; Waldenberger M ; Wellcome Trust Case Control Consortium ; Zeng L ; Chowdhury R ; Salomaa V ; Ford I ; Jukema JW ; Amouyel P ; Kontto J ; MORGAM Investigators ; Nordestgaard BG ; Ferrières J ; Saleheen D ; Sattar N ; Surendran P ; Wagner A ; Young R ; Howson JM ; Butterworth AS ; Danesh J ; Ardissino D ; Bottinger EP ; Erbel R ; Franks PW ; Girelli D ; Hall AS ; Hovingh GK ; Kastrati A ; Lieb W ; Meitinger T ; Kraus WE ; Shah SH ; McPherson R ; Orho-Melander M ; Melander O ; Metspalu A ; Palmer CN ; Peters A ; Rader DJ ; Reilly MP ; Loos RJ ; Reiner AP ; Roden DM ; Tardif JC ; Thompson JR ; Wareham NJ ; Watkins H ; Willer CJ ; Samani NJ ; Schunkert H ; Deloukas P ; Kathiresan S ; Myocardial Infarction Genetics and CARDIoGRAM Exome Consortia Investigators. Systematic evaluation of pleiotropy identifies 6 further loci associated with coronary artery disease. J Am Coll Cardiol, 69(7) :823–836, February 2017.

[60] Howson JMM ; Zhao W ; Barnes DR ; Ho WK ; Young R ; ; Paul DS ; Waite LL ; Freitag DF ; Fauman EB ; Salfati EL ; Sun BB ; Eicher JD ; Johnson AD ; Sheu WHH ; Nielsen SF ; Lin WY ; Surendran P ; Malarstig A ; Wilk JB ; Tybjærg-Hansen A ; Rasmussen KL ; Kamstrup PR ; Deloukas P ; Erdmann J ; Kathiresan S ; Samani NJ ; Schunkert H ; Watkins H ; CARDIoGRAMplusCD ; Do R ; Rader DJ ; Johnson JA ; Hazen SL ; Quyyumi AA ; Spertus JA ; ; Pepine CJ ; Franceschini N ; Justice A ; Reiner AP ; Buyske S ; Hindorff LA ; Carty CL ; North KE ; ; Kooperberg C ; Boerwinkle E ; Young K ; Graff M ; Peters U ; Absher D ; Hsiung CA ; Lee WJ ; Taylor KD ; Chen YH ; Lee IT ; Guo X ; Chung RH ; Hung YJ ; Rotter JI ; Juang JJ ; Quertermous T ; Wang TD ; Rasheed A ; Frossard P ; Alam DS ; Majumder AAS ; Di Angelantonio E ; ; Chowdhury R ; EPIC-CVD ; Chen YI ; Nordestgaard BG ; Assimes TL ; ; Danesh J ; Butterworth AS ; Saleheen D. Fifteen new risk loci for coronary artery disease highlight arterial-wall-specific mechanisms. Nat Genet, 49(7) :1113–1119, July 2017.

[61] Nelson CP ; Goel A ; Butterworth AS ; Kanoni S ; Webb TR ; Marouli E ; Zeng L ; Ntalla I ; Lai FY ; Hopewell JC ; Giannakopoulou O ; Jiang T ; Hamby SE ; Di Angelantonio E ; Assimes TL ; Bottinger EP ; Chambers JC ; Clarke R ; Palmer CNA ; Cubbon RM ; Ellinor P ; Ermel R ; Evangelou E ; Franks PW ; Grace C ; Gu D ; Hingorani AD ; Howson JMM ; Ingelsson E ; Kastrati A ; Kessler T ; Kyriakou T ; Lehtimäki T ; Lu X ; Lu Y ; ; März W ; McPherson R ; Metspalu A ; Pujades-Rodriguez M ; Ruusalepp A ; Schadt EE ; Schmidt AF ; Sweeting MJ ; Zalloua PA ; AlGhalayini K ; Keavney BD ; Kooner JS ; Loos RJF ; Patel RS ; Rutter MK ; Tomaszewski M ; Tzoulaki I ; Zeggini E ; Erdmann J ; Dedoussis G ; Björkegren JLM ; EPIC-CVD Consortium ; CARDIoGRAMplusCD ; UK Biobank CardioMetabolic Consortium CHD working group ; Schunkert H ; Farrall M ; Danesh J ; Samani NJ ; Watkins H ; Deloukas P. Association analyses based on false discovery rate implicate new loci for coronary artery disease. Nat Genet, 49(9) :1385–1391, September 2017.

[62] Erdmann J ; Stark K ; Esslinger UB ; Rumpf PM ; Koesling D ; de Wit C ; Kaiser FJ ; Braunholz D ; Medack A ; Fischer M ; Zimmermann ME ; Tennstedt S ; Graf E ; Eck S ; Aherrahrou Z ; Nahrstaedt J ; Willenborg C ; Bruse P ; Brænne I ; Nöthen MM ; Hofmann P ; Braund PS ; Mergia E ; Reinhard W ; Burgdorf C ; Schreiber S ; Balmforth AJ ; Hall AS ; Bertram L ; Steinhagen-Thiessen E ; Li SC ; März W ; Reilly M ; Kathiresan S ; McPherson R ; Walter U ; CARDIoGRAM ; Ott J ; Samani NJ ; Strom TM ; Meitinger T ; Hengstenberg C ; Schunkert HErdmann J ; Stark K ; Esslinger UB ; Rumpf PM ; Koesling D ; de Wit C ; Kaiser FJ ; Braunholz D ; Medack A ; Fischer M ; Zimmermann ME ; Tennstedt S ; Graf E ; Eck S ; Aherrahrou Z ; Nahrstaedt J ; Willenborg C ; Bruse P ; Brænne I ; Nöthen MM ; Hofmann P ; Braund PS ; Mergia E ; Reinhard W ; Burgdorf C ; Schreiber S ; Balmforth AJ ; Hall AS ; Bertram L ; Steinhagen-Thiessen E ; Li SC ; März W ; Reilly M ; Kathiresan S ; McPherson R ; Walter U ; CARDIoGRAM ; Ott J ; Samani NJ ; Strom TM ; Meitinger T ; Hengstenberg C ; Schunkert H. Dysfunctional nitric oxide signalling increases risk of myocardial infarction. Nature, 504(7480) :432–436, December 2013.

[63] Culotta E ; Koshland DE Jr. No news is good news. Science, 258(5090) :1862–1865, December 1992.

[64] Vecoli C. Endothelial nitric oxide synthase gene polymorphisms in cardiovascular disease. Vitam Horm, 94 :387–406, 2014.

[65] Schunkert H ; König IR ; Kathiresan S ; Reilly MP ; Assimes TL ; Holm H ; Preuss M ; Stewart AF ; Barbalic M ; Gieger C ; Absher D ; Aherrahrou Z ; Allayee H ; Altshuler D ; Anand SS ; Andersen K ; Anderson JL ; Ardissino D ; Ball SG ; Balmforth AJ ; Barnes TA ; Becker DM ; Becker LC ; Berger K ; Bis JC ; Boekholdt SM ; Boerwinkle E ; Braund PS ; Brown MJ ; Burnett MS ; Buysschaert I ; Cardiogenics ; Carlquist JF ; Chen L ; Cichon S ; Codd V ; Davies RW ; Dedoussis G ; Dehghan A ; Demissie S ; Devaney JM ; Diemert P ; Do R ; Doering A ; Eifert S ; Mokhtari NE ; Ellis SG ; Elosua R ; Engert JC ; Epstein SE ; de Faire U ; Fischer M ; Folsom AR ; Freyer J ; Gigante B ; Girelli D ;

Gretarsdottir S ; Gudnason V ; Gulcher JR ; Halperin E ; Hammond N ; Hazen SL ; Hofman A ; Horne BD ; Illig T ; Iribarren C ; Jones GT ; Jukema JW ; Kaiser MA ; Kaplan LM ; Kastelein JJ ; Khaw KT ; Knowles JW ; Kolovou G ; Kong A ; Laaksonen R ; Lambrechts D ; Leander K ; Lettre G ; Li M ; Lieb W ; Loley C ; Lotery AJ ; Mannucci PM ; Maouche S ; Martinelli N ; McKeown PP ; Meisinger C ; Meitinger T ; Melander O ; Merlini PA ; Mooser V ; Morgan T ; Mühleisen TW ; Muhlestein JB ; Münzel T ; Musunuru K ; Nahrstaedt J ; Nelson CP ; Nöthen MM ; Olivieri O ; Patel RS ; Patterson CC ; Peters A ; Peyvandi F ; Qu L ; Quyyumi AA ; Rader DJ ; Rallidis LS ; Rice C ; Rosendaal FR ; Rubin D ; Salomaa V ; Sampietro ML ; Sandhu MS ; Schadt E ; Schäfer A ; Schillert A ; Schreiber S ; Schrezenmeir J ; Schwartz SM ; Siscovick DS ; Sivananthan M ; Sivapalaratnam S ; Smith A ; Smith TB ; Snoep JD ; Soranzo N ; Spertus JA ; Stark K ; Stirrups K ; Stoll M ; Tang WH ; Tennstedt S ; Thorgeirsson G ; Thorleifsson G ; Tomaszewski M ; Uitterlinden AG ; van Rij AM ; Voight BF ; Wareham NJ ; Wells GA ; Wichmann HE ; Wild PS ; Willenborg C ; Witteman JC ; Wright BJ ; Ye S ; Zeller T ; Ziegler A ; Cambien F ; Goodall AH ; Cupples LA ; Quertermous T ; März W ; Hengstenberg C ; Blankenberg S ; Ouwehand WH ; Hall AS ; Deloukas P ; Thompson JR ; Stefansson K ; Roberts R ; Thorsteinsdottir U ; O'Donnell CJ ; McPherson R ; Erdmann J ; CARDIoGRAM Consortium ; Samani NJ. Large-scale association analysis identifies 13 new susceptibility loci for coronary artery disease. Nat Genet, 43(4) :333–338, March 2011.

[66] CARDIoGRAMplusC4D Consortium ; Deloukas P ; Kanoni S ; Willenborg C ; Farrall M ; Assimes TL ; Thompson JR ; Ingelsson E ; Saleheen D ; Erdmann J ; Goldstein BA ; Stirrups K ; König IR ; Cazier JB ; Johansson A ; Hall AS ; Lee JY ; Willer CJ ; Chambers JC ; Esko T ; Folkersen L ; Goel A ; Grundberg E ; Havulinna AS ; Ho WK ; Hopewell JC ; Eriksson N ; Kleber ME ; Kristiansson K ; Lundmark P ; Lyytikäinen LP ; Rafelt S ; Shungin D ; Strawbridge RJ ; Thorleifsson G ; Tikkanen E ; Van Zuydam N ; Voight BF ; Waite LL ; Zhang W ; Ziegler A ; Absher D ; Altshuler D ; Balmforth AJ ; Barroso I ; Braund PS ; Burgdorf C ; Claudi-Boehm S ; Cox D ; Dimitriou M ; Do R ; DIAGRAM Consortium ; CARDIOGENICS Consortium ; Doney AS ; El Mokhtari N ; Eriksson P ; Fischer K ; Fontanillas P ; Franco-Cereceda A ; Gigante B ; Groop L ; Gustafsson S ; Hager J ; Hallmans G ; Han BG ; Hunt SE ; Kang HM ; Illig T ; Kessler T ; Knowles JW ; Kolovou G ; Kuusisto J ; Langenberg C ; Langford C ; Leander K ; Lokki ML ; Lundmark A ; McCarthy MI ; Meisinger C ; Melander O ; Mihailov E ; Maouche S ; Morris AD ; Müller-Nurasyid M ; MuTHER Consortium ; Nikus K ; Peden JF ; Rayner NW ; Rasheed A ; Rosinger S ; Rubin D ; Rumpf MP ; Schäfer A ; Sivananthan M ; Song C ; Stewart AF ; Tan ST ; Thorgeirsson G ; van der Schoot CE ; Wagner PJ ; Wellcome Trust Case Control Consortium ; Wells GA ; Wild PS ; Yang TP ; Amouyel P ; Arveiler D ; Basart H ; Boehnke M ; Boerwinkle E ; Brambilla P ; Cambien F ; Cupples AL ; de Faire U ; Dehghan A ; Diemert P ; Epstein SE ; Evans A ; Ferrario MM ; Ferrières J ; Gauguier D ; Go AS ; Goodall AH ; Gudnason V ; Hazen SL ; Holm H ; Iribarren C ; Jang Y ; Kähönen M ; Kee F ; Kim HS ; Klopp N ; Koenig W ; Kratzer W ; Kuulasmaa K ; Laakso M ; Laaksonen R ; Lee JY ; Lind L ; Ouwehand WH ; Parish S ; Park JE ; Pedersen NL ; Peters A ; Quertermous T ; Rader DJ ; Salomaa V ; Schadt E ; Shah SH ; Sinisalo J ; Stark K ; Stefansson K ; Trégouët DA ; Virtamo J ; Wallentin L ; Wareham N ; Zimmermann ME ; Nieminen MS ; Hengstenberg C ; Sandhu MS ; Pastinen T ; Syvänen AC ; Hovingh GK ; Dedoussis G ; Franks PW ; Lehtimäki T ; Metspalu A ; Zalloua PA ; Siegbahn A ; Schreiber S ; Ripatti S ; Blankenberg SS ; Perola M ; Clarke R ; Boehm BO ; O'Donnell C ; Reilly MP ; März W ; Collins R ; Kathiresan S ; Hamsten A ; Kooner JS ; Thorsteinsdottir U ; Danesh J ; Palmer CN ; Roberts R ; Watkins H ; Schunkert H ; Samani NJ. Large-scale association analysis identifies new susceptibility loci for coronary artery disease. Nat Genet, 45(1) :25–33, January 2013.

[67] Bertina RM ; Koeleman BP ; Koster T ; Rosendaal FR ; Dirven RJ ; de Ronde H ; van der Velden PA ; Reitsma PH. Mutation in blood coagulation factor v associated with resistance to activated protein c.

Nature, 369(6475) :64–67, May 1994.

[68] Tang W ; Schwienbacher C ; Lopez LM ; Ben-Shlomo Y ; Oudot-Mellakh T ; Johnson AD ; Samani NJ ; Basu S ; Gögele M ; Davies G ; Lowe GD ; Tregouet DA ; Tan A ; Pankow JS ; Tenesa A ; Levy D ; Volpato CB ; Rumley A ; Gow AJ ; Minelli C ; Yarnell JW ; Porteous DJ ; Starr JM ; Gallacher J ; Boerwinkle E ; Visscher PM ; Pramstaller PP ; Cushman M ; Emilsson V ; Plump AS ; Matijevic N ; Morange PE ; Deary IJ ; Hicks AA ; Folsom A. Genetic associations for activated partial thromboplastin time and prothrombin time, their gene expression profiles, and risk of coronary artery disease. Am J Hum Genet, 91(1) :152–162, July 2012.

[69] Myocardial Infarction Genetics and CARDIoGRAM Exome Consortia Investigators ; Stitziel NO ; Stirrups KE ; Masca NG ; Erdmann J ; Ferrario PG ; König IR ; Weeke PE ; Webb TR ; Auer PL ; Schick UM ; Lu Y ; Zhang H ; Dube MP ; Goel A ; Farrall M ; Peloso GM ; Won HH ; Do R ; van Iperen E ; Kanoni S ; Kruppa J ; Mahajan A ; Scott RA ; Willenberg C ; Braund PS ; van Capelleveen JC ; Doney AS ; Donnelly LA ; Asselta R ; Merlini PA ; Duga S ; Marziliano N ; Denny JC ; Shaffer CM ; El-Mokhtari NE ; Franke A ; Gottesman O ; Heilmann S ; Hengstenberg C ; Hoffman P ; Holmen OL ; Hveem K ; Jansson JH ; Jöckel KH ; Kessler T ; Kriebel J ; Laugwitz KL ; Marouli E ; Martinelli N ; McCarthy MI ; Van Zuydam NR ; Meisinger C ; Esko T ; Mihailov E ; Escher SA ; Alver M ; Moebus S ; Morris AD ; Müller-Nurasyid M ; Nikpay M ; Olivieri O ; Lemieux Perreault LP ; AlQarawi A ; Robertson NR ; Akinsanya KO ; Reilly DF ; Vogt TF ; Yin W ; Asselbergs FW ; Kooperberg C ; Jackson RD ; Stahl E ; Strauch K ; Varga TV ; Waldenberger M ; Zeng L ; Kraja AT ; Liu C ; Ehret GB ; Newton-Cheh C ; Chasman DI ; Chowdhury R ; Ferrario M ; Ford I ; Jukema JW ; Kee F ; Kuulasmaa K ; Nordestgaard BG ; Perola M ; Saleheen D ; Sattar N ; Surendran P ; Tregouet D ; Young R ; Howson JM ; Butterworth AS ; Danesh J ; Ardissino D ; Bottinger EP ; Erbel R ; Franks PW ; Girelli D ; Hall AS ; Hovingh GK ; Kastrati A ; Lieb W ; Meitinger T ; Kraus WE ; Shah SH ; McPherson R ; Orho-Melander M ; Melander O ; Metspalu A ; Palmer CN ; Peters A ; Rader D ; Reilly MP ; Loos RJ ; Reiner AP ; Roden DM ; Tardif JC ; Thompson JR ; Wareham NJ ; Watkins H ; Willer CJ ; Kathiresan S ; Deloukas P ; Samani NJ ; Schunkert H. Coding variation in angptl4, lpl, and svep1 and the risk of coronary disease. N Engl J Med, 374(12) :1134–1144, March 2016.

[70] Krarup NT ; Borglykke A ; Allin KH ; Sandholt CH ; Justesen JM ; Andersson EA ; Grarup N ; Jørgensen T ; Pedersen O ; Hansen T. A genetic risk score of 45 coronary artery disease risk variants associates with increased risk of myocardial infarction in 6041 danish individuals. Atherosclerosis, 240(2) :305–310, June 2015.

[71] Abifadel M ; Varret M ; Rabès JP ; Allard D ; Ouguerram K ; Devillers M ; Cruaud C ; Benjannet S ; Wickham L ; Erlich D ; Derré A ; Villéger L ; Farnier M ; Beucler I ; Bruckert E ; Chambaz J ; Chanu B ; Lecerf JM ; Luc G ; Moulin P ; Weissenbach J ; Prat A ; Krempf M ; Junien C ; Seidah NG ; Boileau C. Mutations in pcsk9 cause autosomal dominant hypercholesterolemia. Nat Genet, 34(2) :154–156, June 2003.

[72] Abifadel M ; Elbitar S ; El Khoury P ; Ghaleb Y ; Chémaly M ; Moussalli ML ; Rabès JP ; Varret M ; Boileau C. Living the pcsk9 adventure : from the identification of a new gene in familial hypercholesterolemia towards a potential new class of anticholesterol drugs. Curr Atheroscler Rep, 16(9) :439, September 2014.

[73] Lingling Zhang ; Kangping Song ; Mengting Zhu ; Jinling Shi ; Huijuan Zhang ; Liang Xu ; Yingzhu Chen. Proprotein convertase subtilisin/kexin type 9 (pcsk9) in lipid metabolism, atherosclerosis and ischemic stroke. Int J Neurosci, 126(8) :675–680, August 2016.

[74] Cohen JC ; Boerwinkle E ; Mosley TH Jr ; Hobbs HH. Sequence variations in pcsk9, low ldl, and protection against coronary heart disease. N Engl J Med, 354(12) :1264–1272, March 2006.

[75] Norata GD ; Tibolla G ; Catapano AL. Pcsk9 inhibition for the treatment of hypercholesterolemia : promises and emerging challenges. Vascul Pharmacol, 62(2) :103–111, August 2014.

[76] Linsel-Nitschke P ; Götz A ; Erdmann J ; Braenne I ; Braund P ; Hengstenberg C ; Stark K ; Fischer M ; Schreiber S ; El Mokhtari NE ; Schaefer A ; Schrezenmeir J ; Rubin D ; Hinney A ; Reinehr T ; Roth C ; Ortlepp J ; Hanrath P ; Hall AS ; Mangino M ; Lieb W ; Lamina C ; Heid IM ; Doering A ; Gieger C ; Peters A ; Meitinger T ; Wichmann HE ; König IR ; Ziegler A ; Kronenberg F ; Samani NJ ; Schunkert H ; Wellcome Trust Case Control Consortium (WTCCC) ; Cardiogenics Consortium. Lifelong reduction of ldl-cholesterol related to a common variant in the ldl-receptor gene decreases the risk of coronary artery disease–a mendelian randomisation study. PLoS One, 3(8) :e2986, August 2008.

[77] Davey Smith G ; Ebrahim S. What can mendelian randomisation tell us about modifiable behavioural and environmental exposures ? BMJ, 330(7499) :1076–1079, May 2005.

[78] Tretbar LL. Deep veins. Dermatol Surg, 21(1) :47–51, January 1995.

[79] National Clinical Guideline Centre – Acute and Chronic Conditions (UK). Venous thromboembolism : Reducing the risk of venous thromboembolism (deep vein thrombosis and pulmonary embolism) in patients admitted to hospital. London : Royal College of Physicians (UK), 2010.

[80] Cushman M. Epidemiology and risk factors for venous thrombosis. Semin Hematol, 44(2) :62–69, April 2007.

[81] Rosendaal FR ; Reitsma PH. Genetics of venous thrombosis. J Thromb Haemost, Suppl 1 :301–304, July 2009.

[82] Bagot CN ; Arya R. Virchow and his triad : a question of attribution. Br J Haematol, 143(2) :180–190, October 2008.

[83] Furie B ; Furie BC. Thrombus formation in vivo. J Clin Invest, 115(12) :3355–3362, December 2005.

[84] Li Z ; Delaney MK ; O'Brien KA ; Du X. Signaling during platelet adhesion and activation. Arterioscler Thromb Vasc Biol, 30(12) :2341–2349, December 2010.

[85] De Candia E. Mechanisms of platelet activation by thrombin : a short history. Thromb Res, 129(3) :250–256, March 2012.

[86] Johari V ; Loke C. Brief overview of the coagulation cascade. Dis Mon, 58(8) :421–423, August 2012.

[87] Owens AP 3rd ; Mackman N. Tissue factor and thrombosis : The clot starts here. Thromb Haemost, 104(3) :432–439, September 2010.

[88] Chapin JC ; Hajjar KA. Fibrinolysis and the control of blood coagulation. Blood Rev, 29(1) :17–24, January 2015.

[89] Palta S ; Saroa R ; Palta A. Overview of the coagulation system. Indian J Anaesth, 58(5) :515–523, September 2014.

[90] Botting R ; Vane JR. Mediators and the anti-thrombotic properties of the vascular endothelium. Ann Med, 21(1) :31–38, February 1989.

[91] Forbes JM ; Cooper ME. Mechanisms of diabetic complications. Physiol Rev, 93(1) :137–188, January 2013.

[92] Tai TC ; Wong-Faull DC ; Claycomb R ; Wong DL. Hypoxia and adrenergic function : molecular mechanisms related to egr-1 and sp1 activation. Brain Res, 1353 :14–27, September 2010.

[93] Benizri E ; Ginouvès A ; Berra E. The magic of the hypoxia-signaling cascade. Cell Mol Life Sci, 65(7-8) :1133–1149, April 2008.

[94] Bovill EG ; van der Vliet A. Venous valvular stasis-associated hypoxia and thrombosis : what is the link ? Annu Rev Physiol, 73 :527–545, 2011.

[95] García de Frutos P ; Fuentes-Prior P ; Hurtado B ; Sala N. Molecular basis of protein s deficiency. Thromb Haemost, 98(3) :543–556, September 2007.

[96] Heit JA ; Spencer FA ; White RH. The epidemiology of venous thromboembolism. J Thromb Thrombolysis, 41(1) :3–14, Janauary 2016.

[97] Silverstein MD ; Heit JA ; Mohr DN ; Petterson TM ; O'Fallon WM ; Melton LJ 3rd. Trends in the incidence of deep vein thrombosis and pulmonary embolism : a 25-year population-based study. Arch Intern Med, 158(6) :585–593, March 1998.

[98] Rahimtoola A ; Bergin JD. Acute pulmonary embolism : an update on diagnosis and management. Curr Probl Cardiol, 30(2) :61–114, February 2005.

[99] Kröger K ; Küpper-Nybelen J ; Moerchel C ; Moysidis T ; Kienitz C ; Schubert I. Prevalence and economic burden of pulmonary embolism in germany. Vasc Med, 17(5) :303–309, October 2012.

[100] Cushman M. Epidemiology and risk factors for venous thrombosis. Semin Hematol, 44(2) :62–69, April 2007.

[101] Pomp ER ; le Cessie S ; Rosendaal FR ; Doggen CJ. Risk of venous thrombosis : obesity and its joint effect with oral contraceptive use and prothrombotic mutations. Br J Haematol, 139(2) :289–296, October 2007.

[102] Geerts WH ; Pineo GF ; Heit JA ; Bergqvist D ; Lassen MR ; Colwell CW ; Ray JG. Prevention of venous thromboembolism : the seventh accp conference on antithrombotic and thrombolytic therapy. Chest, 126 (3 Suppl) :338S–400S, September 2004.

[103] van Stralen KJ ; Rosendaal FR ; Doggen CJ. Minor injuries as a risk factor for venous thrombosis. Arch Intern Med, 168(1) :21–26, January 2008.

[104] Blom JW ; Vanderschoot JP ; Oostindiër MJ ; Osanto S ; van der Meer FJ ; Rosendaal FR. Incidence of venous thrombosis in a large cohort of 66,329 cancer patients : results of a record linkage study. J Thromb Haemost, 4(3) :529–535, March 2006.

[105] Ocak G ; Vossen CY ; Verduijn M ; Dekker FW ; Rosendaal FR ; Cannegieter SC ; Lijfering WM. Risk of venous thrombosis in patients with major illnesses : results from the mega study. J Thromb Haemost, 11(1) :116–123, January 2013.

[106] Giannotta M ; Tapete G ; Emmi G ; Silvestri E ; Milla M. Thrombosis in inflammatory bowel diseases : what's the link ? Thromb J, 13 :14, April 2015.

[107] Petrauskiene V ; Falk M ; Waernbaum I ; Norberg M ; Eriksson JW. The risk of venous thromboembolism is markedly elevated in patients with diabetes. Diabetologia, 48(5) :1017–1021, May 2005.

[108] Wells PS ; Hirsh J ; Anderson DR ; Lensing AW ; Foster G ; Kearon C ; Weitz J ; D'Ovidio R ; Cogo A ; Prandoni P. Accuracy of clinical assessment of deep-vein thrombosis. Lancet, 345(8961) :1326–1330, May 1995.

[109] Bates SM ; Jaeschke R ; Stevens SM ; Goodacre S ; Wells PS ; Stevenson MD ; Kearon C ; Schunemann HJ ; Crowther M ; Pauker SG ; Makdissi R ; Guyatt GH. Diagnosis of dvt : Antithrombotic therapy and prevention of thrombosis, 9th ed : American college of chest physicians evidence-based clinical practice guidelines. Chest, 141(2 Suppl) :e351S–e418S, February 2012.

[110] Kearon C ; Akl EA ; Comerota AJ ; Prandoni P ; Bounameaux H ; Goldhaber SZ ; Nelson ME ; Wells PS ; Gould MK ; Dentali F ; Crowther M ; Kahn SR. Antithrombotic therapy for vte disease : Antithrombotic

therapy and prevention of thrombosis, 9th ed : American college of chest physicians evidence-based clinical practice guidelines. Chest, 141(2 Suppl) :e419S–e496S, February 2012.

[111] Egeberg O. Inherited antithrombin deficiency causing thrombophilia. Thromb Diath Haemorrh, 13 :516–530, June 1965.

[112] Griffin JH ; Evatt B ; Zimmerman TS ; Kleiss AJ ; Wideman C. Deficiency of protein c in congenital thrombotic disease. J Clin Invest, 68(5) :1370–1373, November 1981.

[113] Schwarz HP ; Fischer M ; Hopmeier P ; Batard MA ; Griffin JH. Plasma protein s deficiency in familial thrombotic disease. Blood, 64(6) :1297–1300, December 1984.

[114] Poort SR ; Rosendaal FR ; Reitsma PH ; Bertina RM. A common genetic variation in the 3'-untranslated region of the prothrombin gene is associated with elevated plasma prothrombin levels and an increase in venous thrombosis. Blood, 88(10) :3698–3703, November 1996.

[115] Emmerich J ; Rosendaal FR ; Cattaneo M ; Margaglione M ; De Stefano V ; Cumming T ; Arruda V ; Hillarp A ; Reny JL. Combined effect of factor v leiden and prothrombin 20210a on the risk of venous thromboembolism–pooled analysis of 8 case-control studies including 2310 cases and 3204 controls. study group for pooled-analysis in venous thromboembolism. Thromb Haemost, 86(3) :809–816, September 2001.

[116] Morelli VM ; De Visser MC ; Vos HL ; Bertina RM ; Rosendaal FR. Abo blood group genotypes and the risk of venous thrombosis : effect of factor v leiden. J Thromb Haemost, 3(1) :183–185, January 2005.

[117] Trégouët DA ; Morange PE. What is currently known about the genetics of venous thromboembolism at the dawn of next generation sequencing technologies. Br J Haematol, 180(3) :335–345, February 2018.

[118] Uitte de Willige S ; de Visser MC ; Houwing-Duistermaat JJ ; Rosendaal FR ; Vos HL ; Bertina RM. Genetic variation in the fibrinogen gamma gene increases the risk for deep venous thrombosis by reducing plasma fibrinogen gamma' levels. Blood, 106(13) :4176–4183, December 2005.

[119] Jick H ; Slone D ; Westerholm B ; Inman WH ; Vessey MP ; Shapiro S ; Lewis GP ; Worcester J. Venous thromboembolic disease and abo blood type. a cooperative study. Lancet, 1(7594) :539–542, March 1969.

[120] Koster T ; Blann AD ; Briët E ; Vandenbroucke JP ; Rosendaal FR. Role of clotting factor viii in effect of von willebrand factor on occurrence of deep-vein thrombosis. Lancet, 345(8943) :152–155, January 1995.

[121] Athanasiadis G ; Buil A ; Souto JC ; Borrell M ; López S ; Martinez-Perez A ; Lathrop M ; Fontcuberta J ; Almasy L ; Soria JM. A genome-wide association study of the protein c anticoagulant pathway. PLoS One, 6(12) :e29168, 2011.

[122] Saposnik B ; Reny JL ; Gaussem P ; Emmerich J ; Aiach M ; Gandrille S. A haplotype of the epcr gene is associated with increased plasma levels of sepcr and is a candidate risk factor for thrombosis. Blood, 103(4) :1311–1318, February 2004.

[123] Sennblad B ; Basu S ; Mazur J ; Suchon P ; Martinez-Perez A ; van Hylckama Vlieg A ; Truong V ; Li Y ; Gådin JR ; Tang W ; Grossman V ; de Haan HG ; Handin N ; Silveira A ; Souto JC ; Franco-Cereceda A ; Morange PE ; Gagnon F ; Soria JM ; Eriksson P ; Hamsten A ; Maegdefessel L ; Rosendaal FR ; Wild P ; Folsom AR ; Trégouët DA ; Sabater-Lleal M. Genome-wide association study with additional genetic and post-transcriptional analyses reveals novel regulators of plasma factor xi levels. Hum Mol Genet, 26(3) :637–649, February 2017.

[124] Germain M ; Chasman DI ; de Haan H ; Tang W ; Lindström S ; Weng LC ; de Andrade M ; de Visser MC ; Wiggins KL ; Suchon P ; Saut N ; Smadja DM ; Le Gal G ; van Hylckama Vlieg A ; Di Narzo A ; Hao K ; Nelson CP ; Rocanin-Arjo A ; Folkersen L ; Monajemi R ; Rose LM ; Brody JA ; Slagboom E ; Aïssi D ; Gagnon F ; Deleuze JF ; Deloukas P ; Tzourio C ; Dartigues JF ; Berr C ; Taylor KD ; Civelek M ; Eriksson P ; Cardiogenics Consortium ; Psaty BM ; Houwing-Duitermaat J ; Goodall AH ; Cambien F ; Kraft P ; Amouyel P ; Samani NJ ; Basu S ; Ridker PM ; Rosendaal FR ; Kabrhel C ; Folsom AR ; Heit J ; Reitsma PH ; Trégouët DA ; Smith NL ; Morange PE. Meta-analysis of 65,734 individuals identifies tspan15 and slc44a2 as two susceptibility loci for venous thromboembolism. Am J Hum Genet, 96(4) :532–542, April 2015.

[125] Ageno W ; Becattini C ; Brighton T ; Selby R ; Kamphuisen PW. Cardiovascular risk factors and venous thromboembolism : a meta-analysis. Circulation, 117(1) :93–102, January 2008.

[126] Morelli VM ; Lijfering WM ; Bos MHA ; Rosendaal FR ; Cannegieter SC. Lipid levels and risk of venous thrombosis : results from the mega-study. Eur J Epidemiol, 32(8) :669–681, August 2017.

[127] FitzGerald GA ; Tigges J ; Barry P ; Lawson JA. Markers of platelet activation and oxidant stress in atherothrombotic disease. Thromb Haemost, 78(1) :280–284, July 1997.

[128] Libby P. The interface of atherosclerosis and thrombosis : basic mechanisms. Vasc Med, 3(3) :225–229, 1998.

[129] Lane DA ; Grant PJ. Role of hemostatic gene polymorphisms in venous and arterial thrombotic disease. Blood, 95(5) :1517–1532, March 2000.

[130] Prandoni P ; Bilora F ; Marchiori A ; Bernardi E ; Petrobelli F ; Lensing AW ; Prins MH ; Girolami A. An association between atherosclerosis and venous thrombosis. N Engl J Med, 348(15) :1435–1441, April 2003.

[131] Prandoni P. Venous and arterial thrombosis : Two aspects of the same disease ? Clin Epidemiol, 1 :1–6, August 2009.

[132] Lijfering WM ; Flinterman LE ; Vandenbroucke JP ; Rosendaal FR ; Cannegieter SC. Relationship between venous and arterial thrombosis : a review of the literature from a causal perspective. Semin Thromb Hemost, 37(8) :885–896, November 2011.

[133] Prandoni P. Venous and arterial thrombosis : Is there a link ? Adv Exp Med Biol, 906 :273–283, 2017.

[134] Arnett DK ; Baird AE ; Barkley RA ; Basson CT ; Boerwinkle E ; Ganesh SK ; Herrington DM ; Hong Y ; Jaquish C ; McDermott DA ; O'Donnell CJ ; American Heart Association Council on Epidemiology, Prevention ; American Heart Association Stroke Council ; Functional Genomics, and Translational Biology Interdisciplinary Working Group. Relevance of genetics and genomics for prevention and treatment of cardiovascular disease : a scientific statement from the american heart association council on epidemiology and prevention, the stroke council, and the functional genomics and translational biology interdisciplinary working group. Circulation, 115(22) :2878–2901, June 2007.

[135] Friedman AN ; Bostom AG ; Selhub J ; Levey AS ; Rosenberg IH. The kidney and homocysteine metabolism. J Am Soc Nephrol, 12(10) :2181–2189, October 2001.

[136] Cattaneo M. Hyperhomocysteinemia, atherosclerosis and thrombosis. Thromb Haemost, 81(2) :165–176, February 1999.

[137] World Health Organisation. Diabetes fact sheet n312. July 2017.

[138] Ionescu-Tirgoviste C ; Gagniuc PA ; Gubceac E ; Mardare L ; Popescu I ; Dima S ; Militaru M. A 3d map of the islet routes throughout the healthy human pancreas. Sci Rep, 5 :14634, September 2015.

[139] Paul Langerhans. Beiträge zur mikroskopischen Anatomie der Bauchspeicheldrüse. Buchdruckerei von Gustav Lange, Berlin ; Deutschland, 1869.

[140] Fu Z ; Gilbert ER ; Liu D. Regulation of insulin synthesis and secretion and pancreatic beta-cell dysfunction in diabetes. Curr Diabetes Rev, 9(1) :25–53, January 2013.

[141] Bell GI ; Pictet RL ; Rutter WJ ; Cordell B ; Tischer E ; Goodman HM. Sequence of the human insulin gene. Nature, 284(5751) :26–32, March 1980.

[142] Dimitriadis G ; Mitrou P ; Lambadiari V ; Maratou E ; Raptis SA. Insulin effects in muscle and adipose tissue. Diabetes Res Clin Pract, 93(Suppl 1) :S52–59, August 2011.

[143] Misra S ; Oliver NS. Diabetic ketoacidosis in adults. BMJ, 351 :h5660, October 2015.

[144] Stoner GD. Hyperosmolar hyperglycemic state. Am Fam Physician, 71(9) :1723–1730, May 2005.

[145] Scheen AJ. Pathophysiology of type 2 diabetes. Acta Clin Belg, 58(6) :335–341, November 2003.

[146] Roy T ; Lloyd CE. Epidemiology of depression and diabetes : a systematic review. J Affect Disord, 142 Suppl :S8–21, October 2012.

[147] Giacco F ; Brownlee M. Oxidative stress and diabetic complications. Circ Res, 107(9) :1058–1070, October 2010.

[148] Hakonarson H Bakay M1, Pandey R2. Genes involved in type 1 diabetes : an update. Genes (Basel), 4(3) :499–521, September 2013.

[149] Redondo MJ ; Yu L ; Hawa M ; Mackenzie T ; Pyke DA ; Eisenbarth GS ; Leslie RD. Heterogeneity of type i diabetes : analysis of monozygotic twins in great britain and the united states. Diabetologia, 44(3) :354–362, March 2001.

[150] Ionescu-Tîrgoviste C ; Gagniuc PA ; Guja C. Structural properties of gene promoters highlight more than two phenotypes of diabetes. PLoS One, 10(9) :e0137950, September 2015.

[151] Todd JA. Genetic control of autoimmunity in type 1 diabetes. Immunol Today, 11(4) :122–129, April 1990.

[152] Dorman JS ; Bunker CH. Hla-dq locus of the human leukocyte antigen complex and type 1 diabetes mellitus : a huge review. Epidemiol Rev, 22(2) :218–227, 2000.

[153] Pociot F ; Lernmark Å. Genetic risk factors for type 1 diabetes. Lancet, 387(10035) :2331–2339, June 2016.

[154] Dyment DA ; Sadovnick AD ; Ebers GC. Genetics of multiple sclerosis. Hum Mol Genet, 6(10) :1693–1698, 1997.

[155] Wu J ; Li J ; Li S ; Zhang TP ; Li LJ ; Lv TT ; Pan HF ; Ye DQ. Association of hla-dqb1 polymorphisms with rheumatoid arthritis : a meta-analysis. Postgrad Med J, 93(1104) :618–625, October 2017.

[156] Murray JA ; Moore SB ; Van Dyke CT ; Lahr BD ; Dierkhising RA ; Zinsmeister AR ; Melton LJ 3rd ; Kroning CM ; El-Yousseff M ; Czaja AJ. Hla dq gene dosage and risk and severity of celiac disease. Clin Gastroenterol Hepatol, 5(12) :1406–1412, December 2007.

[157] Fuchsberger C ; Flannick J ; Teslovich TM ; Mahajan A ; Agarwala V ; Gaulton KJ ; Ma C ; Fontanillas P ; Moutsianas L ; McCarthy DJ ; Rivas MA ; Perry JRB ; Sim X ; Blackwell TW ; Robertson NR ; Rayner NW ; Cingolani P ; Locke AE ; Tajes JF ; Highland HM ; Dupuis J ; Chines PS ; Lindgren CM ; Hartl C ; Jackson AU ; Chen H ; Huyghe JR ; van de Bunt M ; Pearson RD ; Kumar A ; Müller-Nurasyid M ; Grarup N ; Stringham HM ; Gamazon ER ; Lee J ; Chen Y ; Scott RA ; Below JE ; Chen P ; Huang J ; Go MJ ; Stitzel ML ; Pasko D ; Parker SCJ ; Varga TV ; Green T ; Beer NL ; Day-Williams AG ; Ferreira T ; Fingerlin T ; Horikoshi M ; Hu C ; Huh I ; Ikram MK ; Kim BJ ; Kim Y ; Kim YJ ; Kwon

MS ; Lee J ; Lee S ; Lin KH ; Maxwell TJ ; Nagai Y ; Wang X ; Welch RP ; Yoon J ; Zhang W ; Barzilai N ; Voight BF ; Han BG ; Jenkinson CP ; Kuulasmaa T ; Kuusisto J ; Manning A ; Ng MCY ; Palmer ND ; Balkau B ; Stančáková A ; Abboud HE ; Boeing H ; Giedraitis V ; Prabhakaran D ; Gottesman O ; Scott J ; Carey J ; Kwan P ; Grant G ; Smith JD ; Neale BM ; Purcell S ; Butterworth AS ; Howson JMM ; Lee HM ; Lu Y ; Kwak SH ; Zhao W ; Danesh J ; Lam VKL ; Park KS ; ; Saleheen D ; So WY ; Tam CHT ; Afzal U ; Aguilar D ; Arya R ; Aung T ; Chan E ; Navarro C ; Cheng CY ; Palli D ; Correa A ; Curran JE ; Rybin D ; Farook VS ; Fowler SP ; Freedman BI ; Griswold M ; Hale DE ; Hicks PJ ; Khor CC ; Kumar S ; Lehne B ; Thuillier D ; Lim WY ; Liu J ; van der Schouw YT ; Loh M ; Musani SK ; Puppala S ; Scott WR ; Yengo L ; Tan ST ; Taylor HA Jr ; Thameem F ; Wilson G Sr ; Wong TY ; Njølstad PR ; Levy JC ; Mangino M ; Bonnycastle LL ; Schwarzmayr T ; Fadista J ; Surdulescu GL ; Herder C ; Groves CJ ; Wieland T ; Bork-Jensen J ; Brandslund I ; Christensen C ; Koistinen HA ; Doney ASF ; Kinnunen L ; Esko T ; Farmer AJ ; Hakaste L ; Hodgkiss D ; Kravic J ; Lyssenko V ; Hollensted M ; Jørgensen ME ; Jørgensen T ; Ladenvall C ; Justesen JM ; Käräjämäki A ; Kriebel J ; Rathmann W ; Lannfelt L ; Lauritzen T ; Narisu N ; Linneberg A ; Melander O ; Milani L ; Neville M ; Orho-Melander M ; Qi L ; Qi Q ; Roden M ; Rolandsson O ; Swift A ; Rosengren AH ; Stirrups K ; Wood AR ; Mihailov E ; Blancher C ; Carneiro MO ; Maguire J ; Poplin R ; Shakir K ; Fennell T ; DePristo M ; de Angelis MH ; Deloukas P ; Gjesing AP ; Jun G ; Nilsson P ; Murphy J ; Onofrio R ; Thorand B ; Hansen T ; Meisinger C ; Hu FB ; Isomaa B ; Karpe F ; Liang L ; Peters A ; Huth C ; O'Rahilly SP ; Palmer CNA ; Pedersen O ; Rauramaa R ; Tuomilehto J ; Salomaa V ; Watanabe RM ; Syvänen AC ; Bergman RN ; Bharadwaj D ; Bottinger EP ; Cho YS ; Chandak GR ; Chan JCN ; Chia KS ; Daly MJ ; Ebrahim SB ; Langenberg C ; Elliott P ; Jablonski KA ; Lehman DM ; Jia W ; Ma RCW ; Pollin TI ; Sandhu M ; Tandon N ; Froguel P ; Barroso I ; Teo YY ; Zeggini E ; Loos RJF ; Small KS ; Ried JS ; DeFronzo RA ; Grallert H ; Glaser B ; Metspalu A ; Wareham NJ ; Walker M ; Banks E ; Gieger C ; Ingelsson E ; Im HK ; Illig T ; Franks PW ; Buck G ; Trakalo J ; Buck D ; Prokopenko I ; Mägi R ; Lind L ; Farjoun Y ; Owen KR ; Gloyn AL ; Strauch K ; Tuomi T ; Kooner JS ; Lee JY ; Park T ; Donnelly P ; Morris AD ; Hattersley AT ; Bowden DW ; Collins FS ; Atzmon G ; Chambers JC ; Spector TD ; Laakso M ; Strom TM ; Bell GI ; Blangero J ; Duggirala R ; Tai ES ; McVean G ; Hanis CL ; Wilson JG ; Seielstad M ; Frayling TM ; Meigs JB ; Cox NJ ; Sladek R ; Lander ES ; Gabriel S ; Burtt NP ; Mohlke KL ; Meitinger T ; Groop L ; Abecasis G ; Florez JC ; Scott LJ ; Morris AP ; Kang HM ; Boehnke M ; Altshuler D ; McCarthy MI. The genetic architecture of type 2 diabetes. Nature, 536(7614) :41–47, August 2016.

[158] Feinman RD ; Pogozelski WK ; Astrup A ; Bernstein RK ; Fine EJ ; Westman EC ; Accurso A ; Frassetto L ; Gower BA ; McFarlane SI ; Nielsen JV ; Krarup T ; Saslow L ; Roth KS ; Vernon MC ; Volek JS ; Wilshire GB ; Dahlqvist A ; Sundberg R ; Childers A ; Morrison K ; Manninen AH ; Dashti HM ; Wood RJ ; Wortman J ; Worm N. Dietary carbohydrate restriction as the first approach in diabetes management : critical review and evidence base. Nutrition, 31(1) :1–13, January 2015.

[159] Colberg SR ; Sigal RJ ; Yardley JE ; Riddell MC ; Dunstan DW ; Dempsey PC ; Horton ES ; Castorino K ; Tate DF. Physical activity/exercise and diabetes : A position statement of the american diabetes association. Diabetes Care, 39(11) :2065–2079, November 2016.

[160] Pauline Ricci ; Pierre-Olivier Blotière ; Alain Weill ; Dominique Simon ; Philippe Tuppin ; Philippe Ricordeau ; Hubert Allemand. Diabète traité : quelles évolutions entre 2000 et 2009 en france ? BEH, 42-43 :425–431, November 2010.

[161] Menke A Casagrande S ; Geiss L ; Cowie CC. Prevalence of and trends in diabetes among adults in the united states, 1988-2012. JAMA, 314(10) :1021–1029, September 2015.

[162] Daneman D. Type 1 diabetes. Lancet, 367(9513) :847–858, March 2006.

[163] Songini M ; Lombardo C. The sardinian way to type 1 diabetes. J Diabetes Sci Technol, 4(5) :1248–1255, September 2010.

[164] https ://www.diabetes.fi/en/finnish_diabetes_association/diabetes_in_finland.

[165] Entmacher PS ; Root HF ; Marks HH. Longevity of diabetic patients in recent years. Diabetes, 13 :373–377, July-August 1964.

[166] Borch-Johnsen K ; Nissen H ; Henriksen E ; Kreiner S ; Salling N ; Deckert T ; Nerup J. The natural history of insulin-dependent diabetes mellitus in denmark : 1. long-term survival with and without late diabetic complications. Diabet Med, 4(3) :201–210, May-June 1987.

[167] Krolewski AS ; Warram JH ; Christlieb AR ; Busick EJ ; Kahn CR. The changing natural history of nephropathy in type i diabetes. Am J Med, 78(5) :785–794, May 1985.

[168] Rewers M ; Ludvigsson J. Environmental risk factors for type 1 diabetes. Lancet, 387(10035) :2340–2348, June 2016.

[169] Afkarian M ; Zelnick LR ; Hall YN ; Heagerty PJ ; Tuttle K ; Weiss NS ; de Boer IH. Clinical manifestations of kidney disease among us adults with diabetes, 1988-2014. JAMA, 316(6) :602–610, August 2016.

[170] Livingstone SJ ; Levin D ; Looker HC ; Lindsay RS ; Wild SH ; Joss N ; Leese G ; Leslie P ; McCrimmon RJ ; Metcalfe W ; McKnight JA ; Morris AD ; Pearson DW ; Petrie JR ; Philip S ; Sattar NA ; Traynor JP ; Colhoun HM ; Scottish Diabetes Research Network epidemiology group ; Scottish Renal Registry. Estimated life expectancy in a scottish cohort with type 1 diabetes, 2008-2010. JAMA, 313(1) :37–44, January 2015.

[171] Lind M ; Svensson AM ; Kosiborod M ; Gudbjörnsdottir S ; Pivodic A ; Wedel H ; Dahlqvist S ; Clements M ; Rosengren A. Glycemic control and excess mortality in type 1 diabetes. N Engl J Med, 371(21) :1972–1982, November 2014.

[172] Tuomilehto J ; Borch-Johnsen K ; Molarius A ; Forsén T ; Rastenyte D ; Sarti C ; Reunanen A. Incidence of cardiovascular disease in type 1 (insulin-dependent) diabetic subjects with and without diabetic nephropathy in finland. Diabetologia, 41(7) :784–790, July 1998.

[173] Walker J ; Colhoun H ; Livingstone S ; McCrimmon R ; Petrie J ; Sattar N ; Wild S ; Scottish Diabetes Research Network Epidemiology Group. Type 2 diabetes, socioeconomic status and life expectancy in scotland (2012-2014) : a population-based observational study. Diabetologia, 61(1) :108–116, January 2018.

[174] Kelly PJ ; Clarke PM ; Hayes AJ ; Gerdtham UG ; Cederholm J ; Nilsson P ; Eliasson B ; Gudbjornsdottir S. Predicting mortality in people with type 2 diabetes mellitus after major complications : a study using swedish national diabetes register data. Diabet Med, 31(8) :954–962, August 2014.

[175] Diabetes Control and Complications Trial Research Group ; Nathan DM ; Genuth S ; Lachin J ; Cleary P ; Crofford O ; Davis M ; Rand L ; Siebert C. The effect of intensive treatment of diabetes on the development and progression of long-term complications in insulin-dependent diabetes mellitus. N Engl J Med, 329(14) :977–986, September 1993.

[176] Brownlee M. Biochemistry and molecular cell biology of diabetic complications. Nature, 414(6865) :813–820, December 2001.

[177] Lee AY ; Chung SK ; Chung SS. Demonstration that polyol accumulation is responsible for diabetic cataract by the use of transgenic mice expressing the aldose reductase gene in the lens. Proc Natl Acad Sci U S A, 92(7) :2780–2784, March 1995.

[178] Fu MX ; Wells-Knecht KJ ; Blackledge JA ; Lyons TJ ; Thorpe SR ; Baynes JW. Glycation, glycoxidation, and cross-linking of collagen by glucose. kinetics, mechanisms, and inhibition of late stages of the maillard reaction. Diabetes, 43(5) :676–683, May 1994.

[179] Goh SY ; Cooper ME. Clinical review : The role of advanced glycation end products in progression and complications of diabetes. J Clin Endocrinol Metab, 93(4) :1143–1152, April 2008.

[180] Coughlan MT ; Thorburn DR ; Penfold SA ; Laskowski A ; Harcourt BE ; Sourris KC ; Tan AL ; Fukami K ; Thallas-Bonke V ; Nawroth PP ; Brownlee M ; Bierhaus A ; Cooper ME ; Forbes JM. Rage-induced cytosolic ros promote mitochondrial superoxide generation in diabetes. J Am Soc Nephrol, 20(4) :742–752, April 2009.

[181] Du XL ; Edelstein D ; Rossetti L ; Fantus IG ; Goldberg H ; Ziyadeh F ; Wu J ; Brownlee M. Hyperglycemia-induced mitochondrial superoxide overproduction activates the hexosamine pathway and induces plasminogen activator inhibitor-1 expression by increasing sp1 glycosylation. Proc Natl Acad Sci U S A, 97(22) :12222–12226, October 2000.

[182] Heilig CW ; Concepcion LA ; Riser BL ; Freytag SO ; Zhu M ; Cortes P. Overexpression of glucose transporters in rat mesangial cells cultured in a normal glucose milieu mimics the diabetic phenotype. J Clin Invest, 96(4) :1802–1814, October 1995.

[183] Pittenger GL ; Liu D ; Vinik AI. The toxic effects of serum from patients with type 1 diabetes mellitus on mouse neuroblastoma cells : a new mechanism for development of diabetic autonomic neuropathy. Diabet Med, 10(10) :925–932, December 1993.

[184] Kaštelan S ; Zjačić-Rotkvić V ; Kaštelan Z. Could diabetic retinopathy be an autoimmune disease ? Med Hypotheses, 68(5) :1016–1018, 2007.

[185] Baudouin C ; Fredj-Reygrobellet D ; Brignole F ; Lapalus P ; Gastaud P. Mhc class ii antigen expression by ocular cells in proliferative diabetic retinopathy. Fundam Clin Pharmacol, 7(9) :523–530, 1993.

[186] Ichinose K ; Kawasaki E ; Eguchi K. Recent advancement of understanding pathogenesis of type 1 diabetes and potential relevance to diabetic nephropathy. Am J Nephrol, 27(6) :554–564, 2007.

[187] Granberg V ; Ejskjaer N ; Peakman M ; Sundkvist G. Autoantibodies to autonomic nerves associated with cardiac and peripheral autonomic neuropathy. Diabetes Care, 28(8) :1959–1964, August 2005.

[188] Harjutsalo V ; Katoh S ; Sarti C ; Tajima N ; Tuomilehto J. Population-based assessment of familial clustering of diabetic nephropathy in type 1 diabetes. Diabetes, 53(9) :2449–2454, September 2004.

[189] Hietala K ; Forsblom C ; Summanen P ; Groop PH ; FinnDiane Study Group. Heritability of proliferative diabetic retinopathy. Diabetes, 57(8) :2176–2180, August 2008.

[190] Bennett DL ; Woods CG. Painful and painless channelopathies. Lancet Neurol, 13(6) :587–599, June 2014.

[191] Earle K ; Walker J ; Hill C ; Viberti G. Familial clustering of cardiovascular disease in patients with insulin-dependent diabetes and nephropathy. N Engl J Med, 326(10) :673–677, March 1992.

[192] Krolewski AS ; Warram JH ; Christlieb AR ; Busick EJ ; Kahn CR. The changing natural history of nephropathy in type i diabetes. Am J Med, 78(5) :758–794, May 1985.

[193] American Diabetes Association. Nephropathy in diabetes. Diabetes Care, 27(Suppl 1) :s79–s83, January 2004.

[194] Paulsson M. Basement membrane proteins : structure, assembly, and cellular interactions. Crit Rev Biochem Mol Biol, 27(1-2) :93–127, 1992.

[195] Schlöndorff D ; Banas B. The mesangial cell revisited : no cell is an island. J Am Soc Nephrol, 20(6) :1179–1187, June 2009.

[196] Peter Hamilton Raven ; George Brooks Johnson ; Jonathan B Losos ; Susan R Singer ; Jules Bouharmont ; Pierre L Masson ; Charles Van Hove ; Charles-Marie Évrard ; Annie Mamecier ; et al. Sciences du vivant. De Boeck, Bruxelles, Belgique, 2009.

[197] Meyer TW ; Hostetter TH. Uremia. N Engl J Med, 357(13) :1316–1325, September 2007.

[198] Abbott CA ; Malik RA ; van Ross ER ; Kulkarni J ; Boulton AJ. Prevalence and characteristics of painful diabetic neuropathy in a large community-based diabetic population in the u.k. Diabetes Care, 34(10) :2220–2224, October 2011.

[199] Yang Y ; Hu X ; Zhang Q ; Zou R. Diabetes mellitus and risk of falls in older adults : a systematic review and meta-analysis. Age Ageing, 45(6) :761–767, November 2016.

[200] Warren Clayton Jr ; Tom A. Elasy. A review of the pathophysiology, classification, and treatment of foot ulcers in diabetic patients. Clinical Diabetes, 27(2) :52–58, April 2009.

[201] Vinik AI ; Erbas T. Diabetic autonomic neuropathy. Handb Clin Neurol, 117 :279–294, 2013.

[202] Frank RN. Diabetic retinopathy. N Engl J Med, 350(1) :48–58, January 2004.

[203] Klein R ; Klein BE ; Moss SE ; Davis MD ; DeMets DL. The wisconsin epidemiologic study of diabetic retinopathy. ii. prevalence and risk of diabetic retinopathy when age at diagnosis is less than 30 years. Arch Ophthalmol, 102(4) :520–526, April 1984.

[204] Stitt AW ; Lois N ; Medina RJ ; Adamson P ; Curtis TM. Advances in our understanding of diabetic retinopathy. Clin Sci (Lond), 125(1) :1–17, March 2013.

[205] Zhang C ; Wang H ; Nie J ; Wang F. Protective factors in diabetic retinopathy : focus on blood-retinal barrier. Discov Med, 18(98) :105–112, September 2014.

[206] Burden AC ; Morris A ; Waugh NR ; Gatling W ; Bingley PJ ; Patterson CC. Mortality from heart disease in a cohort of 23,000 patients with insulin-treated diabetes. Diabetologia, 46(6) :760–765, June 2003.

[207] Sarnak MJ ; Levey AS ; Schoolwerth AC ; Coresh J ; Culleton B ; Hamm LL ; McCullough PA ; Kasiske BL ; Kelepouris E ; Klag MJ ; Parfrey P ; Pfeffer M ; Raij L ; Spinosa DJ ; Wilson PW ; American Heart Association Councils on Kidney in Cardiovascular Disease ; High Blood Pressure Research ; Clinical Cardiology ;, Epidemiology, PreventionSarnak MJ ; Levey AS ; Schoolwerth AC ; Coresh J ; Culleton B ; Hamm LL ; McCullough PA ; Kasiske BL ; Kelepouris E ; Klag MJ ; Parfrey P ; Pfeffer M ; Raij L ; Spinosa DJ ; Wilson PW ; American Heart Association Councils on Kidney in Cardiovascular Disease ; High Blood Pressure Research ; Clinical Cardiology ;, Epidemiology, and Prevention. Kidney disease as a risk factor for development of cardiovascular disease : a statement from the american heart association councils on kidney in cardiovascular disease, high blood pressure research, clinical cardiology, and epidemiology and prevention. Circulation, 108(17) :2154–2169, October 2003.

[208] Henry Gray. Anatomy of the Human Body. Lea and Febiger, Philadelphia, United States of America, 1918.

[209] Tryggvason K. Unraveling the mechanisms of glomerular ultrafiltration : nephrin, a key component of the slit diaphragm. J Am Soc Nephrol, 10(11) :2440–2445, November 1999.

[210] Suh JH ; Miner JH. The glomerular basement membrane as a barrier to albumin. Nat Rev Nephrol, 9(8) :470–477, August 2013.

[211] Schlondorff D. The glomerular mesangial cell : an expanding role for a specialized pericyte. FASEB J, 1(4) :272–281, October 1987.

[212] Pannabecker TL. Structure and function of the thin limbs of the loop of henle. Compr Physiol, 2(3) :2063–2086, July 2012.

[213] Volek JS ; Ballard KD ; Forsythe CE. Essentials of Creatine in Sports and Health. Humana Press, New York City, United States of America, 2008.

[214] Kim HY. Renal handling of ammonium and acid base regulation. Electrolyte Blood Press, 7(1) :9–13, June 2009.

[215] Stevens LA ; Coresh J ; Greene T ; Levey AS. Assessing kidney function–measured and estimated glomerular filtration rate. N Engl J Med, 354(23) :2473–2483, June 2006.

[216] Simerville JA ; Maxted WC ; Pahira JJ. Urinalysis : a comprehensive review. Am Fam Physician, 71(6) :1153–1162, March 2005.

[217] Ronco C ; McCullough P ; Anker SD ; Anand I ; Aspromonte N ; Bagshaw SM ; Bellomo R ; Berl T ; Bobek I ; Cruz DN ; Daliento L ; Davenport A ; Haapio M ; Hillege H ; House AA ; Katz N ; Maisel A ; Mankad S ; Zanco P ; Mebazaa A ; Palazzuoli A ; Ronco F ; Shaw A ; Sheinfeld G ; Soni S ; Vescovo G ; Zamperetti N ; Ponikowski P ; Acute Dialysis Quality Initiative (ADQI) consensus group. Cardio-renal syndromes : report from the consensus conference of the acute dialysis quality initiative. Eur Heart J, 31(6) :703–711, March 2010.

[218] Ocak G ; Vossen CY ; Lijfering WM ; Verduijn M ; Dekker FW ; Rosendaal FR ; Cannegieter SC. Role of hemostatic factors on the risk of venous thrombosis in people with impaired kidney function. Circulation, 129(6) :683–691, February 2014.

[219] Loscalzo J. Venous thrombosis in the nephrotic syndrome. N Engl J Med, 368(10) :956–958, March 2013.

[220] Kauffmann RH ; Veltkamp JJ ; Van Tilburg NH ; Van Es LA. Acquired antithrombin iii deficiency and thrombosis in the nephrotic syndrome. Am J Med, 65(4) :607–613, October 1978.

[221] Damman K ; Valente MA ; Voors AA ; O'Connor CM ; van Veldhuisen DJ ; Hillege HL. Renal impairment, worsening renal function, and outcome in patients with heart failure : an updated meta-analysis. Eur Heart J, 35(7) :455–469, February 2014.

[222] Romagnani P ; Remuzzi G ; Glassock R ; Levin A ; Jager KJ ; Tonelli M ; Massy Z ; Wanner C ; Anders HJ. Chronic kidney disease. Nat Rev Dis Primers, 3 :17088, November 2017.

[223] O'Neill WC. Understanding the pathogenesis of vascular calcification : timing is everything. Kidney Int, 92(6) :1316–1318, December 2017.

[224] Wu CJ ; Chen CY ; Lai TS ; Wu PC ; Chuang CK ; Sun FJ ; Liu HL ; Chen HH ; Yeh HI ; Lin CS ; Lin CJ. The role of indoxyl sulfate in renal anemia in patients with chronic kidney disease. Oncotarget, 8(47) :83030–83037, June 2017.

[225] Lee WC ; Fang HY ; Chen HC ; Chen CJ ; Yang CH ; Hang CL ; Wu CJ ; Fang CY. Anemia : A significant cardiovascular mortality risk after st-segment elevation myocardial infarction complicated by the comorbidities of hypertension and kidney disease. PLoS One, 12(7) :e0180165, July 2017.

[226] Viswanathan G ; Gilbert S. The cardiorenal syndrome : making the connection. Int J Nephrol, 2011 :283137, October 2010.

[227] Kazancioğlu R. Risk factors for chronic kidney disease : an update. Kidney Int Suppl (2011), 3(4) :368–371, December 2013.

[228] de Ferranti SD ; de Boer IH ; Fonseca V ; Fox CS ; Golden SH ; Lavie CJ ; Magge SN ; Marx N ; McGuire DK ; Orchard TJ ; Zinman B ; Eckel RH. Type 1 diabetes mellitus and cardiovascular disease : a scientific statement from the american heart association and american diabetes association. Diabetes Care, 37(10) :2843–2863, October 2014.

[229] Pattaro C ; Teumer A ; Gorski M ; Chu AY ; Li M ; Mijatovic V ; Garnaas M ; Tin A ; Sorice R ; Li Y ; Taliun D ; Olden M ; Foster M ; Yang Q ; Chen MH ; Pers TH ; Johnson AD ; Ko YA ; Fuchsberger C ; Tayo B ; Nalls M ; Feitosa MF ; Isaacs A ; Dehghan A ; d'Adamo P ; Adeyemo A ; Dieffenbach AK ; Zonderman AB ; Nolte IM ; van der Most PJ ; Wright AF ; Shuldiner AR ; Morrison AC ; Hofman A ; Smith AV ; Dreisbach AW ; Franke A ; Uitterlinden AG ; Metspalu A ; Tonjes A ; Lupo A ; Robino A ; Johansson Å ; Demirkan A ; Kollerits B ; Freedman BI ; Ponte B ; Oostra BA ; Paulweber B ; Krämer BK ; Mitchell BD ; Buckley BM ; Peralta CA ; Hayward C ; Helmer C ; Rotimi CN ; Shaffer CM ; Müller C ; Sala C ; van Duijn CM ; Saint-Pierre A ; Ackermann D ; Shriner D ; Ruggiero D ; Toniolo D ; Lu Y ; Cusi D ; Czamara D ; Ellinghaus D ; Siscovick DS ; Ruderfer D ; Gieger C ; Grallert H ; Rochtchina E ; Atkinson EJ ; Holliday EG ; Boerwinkle E ; Salvi E ; Bottinger EP ; Murgia F ; Rivadeneira F ; Ernst F ; Kronenberg F ; Hu FB ; Navis GJ ; Curhan GC ; Ehret GB ; Homuth G ; Coassin S ; Thun GA ; Pistis G ; Gambaro G ; Malerba G ; Montgomery GW ; Eiriksdottir G ; Jacobs G ; Li G ; Wichmann HE ; Campbell H ; Schmidt H ; Wallaschofski H ; Völzke H ; Brenner H ; Kroemer HK ; Kramer H ; Lin H ; Leach IM ; Ford I ; Guessous I ; Rudan I ; Prokopenko I ; Borecki I ; Heid IM ; Kolcic I ; Persico I ; Jukema JW ; Wilson JF ; Felix JF ; Divers J ; Lambert JC ; Stafford JM ; Gaspoz JM ; Smith JA ; Faul JD ; Wang JJ ; Ding J ; Hirschhorn JN ; Attia J ; Whitfield JB ; Chalmers J ; Viikari J ; Coresh J ; Denny JC ; Karjalainen J ; Fernandes JK ; Endlich K ; Butterbach K ; Keene KL ; Lohman K ; Portas L ; Launer LJ ; Lyytikäinen LP ; Yengo L ; Franke L ; Ferrucci L ; Rose LM ; Kedenko L ; Rao M ; Struchalin M ; Kleber ME ; Cavalieri M ; Haun M ; Cornelis MC ; Ciullo M ; Pirastu M ; de Andrade M ; McEvoy MA ; Woodward M ; Adam M ; Cocca M ; Nauck M ; Imboden M ; Waldenberger M ; Pruijm M ; Metzger M ; Stumvoll M ; Evans MK ; Sale MM ; Kähönen M ; Boban M ; Bochud M ; Rheinberger M ; Verweij N ; Bouatia-Naji N ; Martin NG ; Hastie N ; Probst-Hensch N ; Soranzo N ; Devuyst O ; Raitakari O ; Gottesman O ; Franco OH ; Polasek O ; Gasparini P ; Munroe PB ; Ridker PM ; Mitchell P ; Muntner P ; Meisinger C ; Smit JH ; ICBP Consortium ; AGEN Consortium ; CARDIOGRAM ; CHARGe-Heart Failure Group ; ECHOGen Consortium ; Kovacs P ; Wild PS ; Froguel P ; Rettig R ; Mägi R ; Biffar R ; Schmidt R ; Middelberg RP ; Carroll RJ ; Penninx BW ; Scott RJ ; Katz R ; Sedaghat S ; Wild SH ; Kardia SL ; Ulivi S ; Hwang SJ ; Enroth S ; Kloiber S ; Trompet S ; Stengel B ; Hancock SJ ; Turner ST ; Rosas SE ; Stracke S ; Harris TB ; Zeller T ; Zemunik T ; Lehtimäki T ; Illig T ; Aspelund T ; Nikopensius T ; Esko T ; Tanaka T ; Gyllensten U ; Völker U ; Emilsson V ; Vitart V ; Aalto V ; Gudnason V ; Chouraki V ; Chen WM ; Igl W ; März W ; Koenig W ; Lieb W ; Loos RJ ; Liu Y ; Snieder H ; Pramstaller PP ; Parsa A ; O'Connell JR ; Susztak K ; Hamet P ; Tremblay J ; de Boer IH ; Böger CA ; Goessling W ; Chasman DI ; Köttgen A ; Kao WH ; Fox CS. Genetic associations at 53 loci highlight cell types and biological pathways relevant for kidney function. Nat Commun, 7 :10023, January 2016.

[230] Pattaro C ; Köttgen A ; Teumer A ; Garnaas M ; Böger CA ; Fuchsberger C ; Olden M ; Chen MH ; Tin A ; Taliun D ; Li M ; Gao X ; Gorski M ; Yang Q ; Hundertmark C ; Foster MC ; O'Seaghdha CM ; Glazer N ; Isaacs A ; Liu CT ; Smith AV ; O'Connell JR ; Struchalin M ; Tanaka T ; Li G ; Johnson AD ; Gierman HJ ; Feitosa M ; Hwang SJ ; Atkinson EJ ; Lohman K ; Cornelis MC ; Johansson Å ; Tönjes A ; Dehghan A ; Chouraki V ; Holliday EG ; Sorice R ; Kutalik Z ; Lehtimäki T ; Esko T ; Deshmukh H ; Ulivi S ; Chu AY ; Murgia F ; Trompet S ; Imboden M ; Kollerits B ; Pistis G ; CARDIoGRAM Consortium ; ICBP Consortium ; CARe Consortium ; Wellcome Trust Case Control Consortium 2 (WTCCC2) ; Harris TB ; Launer LJ ; Aspelund T ; Eiriksdottir G ; Mitchell BD ; Boerwinkle E ; Schmidt H ; Cavalieri M ; Rao M ; Hu FB ; Demirkan A ; Oostra BA ; de Andrade M ; Turner ST ; Ding J ; Andrews JS ; Freedman BI ; Koenig W ; Illig T ; Döring A ; Wichmann HE ; Kolcic I ; Zemunik T ; Boban M ; Minelli C ; Wheeler HE ; Igl W ; Zaboli G ; Wild SH ; Wright AF ; Campbell H ; Ellinghaus D ; Nöthlings U ; Jacobs G ; Biffar R ; Endlich K ; Ernst F ; Homuth G ; Kroemer HK ; Nauck M ; Stracke S ; Völker U ; Völzke

H ; Kovacs P ; Stumvoll M ; Mägi R ; Hofman A ; Uitterlinden AG ; Rivadeneira F ; Aulchenko YS ; Polasek O ; Hastie N ; Vitart V ; Helmer C ; Wang JJ ; Ruggiero D ; Bergmann S ; Kähönen M ; Viikari J ; Nikopensius T ; Province M ; Ketkar S ; Colhoun H ; Doney A ; Robino A ; Giulianini F ; Krämer BK ; Portas L ; Ford I ; Buckley BM ; Adam M ; Thun GA ; Paulweber B ; Haun M ; Sala C ; Metzger M ; Mitchell P ; Ciullo M ; Kim SK ; Vollenweider P ; Raitakari O ; Metspalu A ; Palmer C ; Gasparini P ; Pirastu M ; Jukema JW ; Probst-Hensch NM ; Kronenberg F ; Toniolo D ; Gudnason V ; Shuldiner AR ; Coresh J ; Schmidt R ; Ferrucci L ; Siscovick DS ; van Duijn CM ; Borecki I ; Kardia SL ; Liu Y ; Curhan GC ; Rudan I ; Gyllensten U ; Wilson JF ; Franke A ; Pramstaller PP ; Rettig R ; Prokopenko I ; Witteman JC ; Hayward C ; Ridker P ; Parsa A ; Bochud M ; Heid IM ; Goessling W ; Chasman DI ; Kao WH ; Fox CS. Genome-wide association and functional follow-up reveals new loci for kidney function. PLoS Genet, 7(3) :e1002584, 2012.

[231] Okada Y ; Sim X ; Go MJ ; Wu JY ; Gu D ; Takeuchi F ; Takahashi A ; Maeda S ; Tsunoda T ; Chen P ; Lim SC ; Wong TY ; Liu J ; Young TL ; Aung T ; Seielstad M ; Teo YY ; Kim YJ ; Lee JY ; Han BG ; Kang D ; Chen CH ; Tsai FJ ; Chang LC ; Fann SJ ; Mei H ; Rao DC ; Hixson JE ; Chen S ; Katsuya T ; Isono M ; Ogihara T ; Chambers JC ; Zhang W ; Kooner JS ; KidneyGen Consortium ; CKDGen Consortium ; Albrecht E ; GUGC consortium ; Yamamoto K ; Kubo M ; Nakamura Y ; Kamatani N ; Kato N ; He J ; Chen YT ; Cho YS ; Tai ES ; Tanaka T. Meta-analysis identifies multiple loci associated with kidney function-related traits in east asian populations. Nat Genet, 44(8) :904–909, July 2012.

[232] Köttgen A ; Pattaro C ; Böger CA ; Fuchsberger C ; Olden M ; Glazer NL ; Parsa A ; Gao X ; Yang Q ; Smith AV ; O'Connell JR ; Li M ; Schmidt H ; Tanaka T ; Isaacs A ; Ketkar S ; Hwang SJ ; Johnson AD ; Dehghan A ; Teumer A ; Paré G ; Atkinson EJ ; Zeller T ; Lohman K ; Cornelis MC ; Probst-Hensch NM ; Kronenberg F ; Tönjes A ; Hayward C ; Aspelund T ; Eiriksdottir G ; Launer LJ ; Harris TB ; Rampersaud E ; Mitchell BD ; Arking DE ; Boerwinkle E ; Struchalin M ; Cavalieri M ; Singleton A ; Giallauria F ; Metter J ; de Boer IH ; Haritunians T ; Lumley T ; Siscovick D ; Psaty BM ; Zillikens MC ; Oostra BA ; Feitosa M ; Province M ; de Andrade M ; Turner ST ; Schillert A ; Ziegler A ; Wild PS ; Schnabel RB ; Wilde S ; Munzel TF ; Leak TS ; Illig T ; Klopp N ; Meisinger C ; Wichmann HE ; Koenig W ; Zgaga L ; Zemunik T ; Kolcic I ; Minelli C ; Hu FB ; Johansson A ; Igl W ; Zaboli G ; Wild SH ; Wright AF ; Campbell H ; Ellinghaus D ; Schreiber S ; Aulchenko YS ; Felix JF ; Rivadeneira F ; Uitterlinden AG ; Hofman A ; Imboden M ; Nitsch D ; Brandstätter A ; Kollerits B ; Kedenko L ; Mägi R ; Stumvoll M ; Kovacs P ; Boban M ; Campbell S ; Endlich K ; Völzke H ; Kroemer HK ; Nauck M ; Völker U ; Polasek O ; Vitart V ; Badola S ; Parker AN ; Ridker PM ; Kardia SL ; Blankenberg S ; Liu Y ; Curhan GC ; Franke A ; Rochat T ; Paulweber B ; Prokopenko I ; Wang W ; Gudnason V ; Shuldiner AR ; Coresh J ; Schmidt R ; Ferrucci L ; Shlipak MG ; van Duijn CM ; Borecki I ; Krämer BK ; Rudan I ; Gyllensten U ; Wilson JF ; Witteman JC ; Pramstaller PP ; Rettig R ; Hastie N ; Chasman DI ; Kao WH ; Heid IM ; Fox CS. New loci associated with kidney function and chronic kidney disease. Nat Genet, 42(5) :376–384, May 2010.

[233] Irme Festetics. Die Genetische Gesätze der Natur. 1819.

[234] Bumgarner R. Overview of dna microarrays : types, applications, and their future. Curr Protoc Mol Biol, January 2013.

[235] Ronaghi M ; Karamohamed S ; Pettersson B ; Uhlén M ; Nyrén P. Real-time dna sequencing using detection of pyrophosphate release. Anal Biochem, 242(1) :84–89, November 1996.

[236] Klein RJ ; Zeiss C ; Chew EY ; Tsai JY ; Sackler RS ; Haynes C ; Henning AK ; SanGiovanni JP ; Mane SM ; Mayne ST ; Bracken MB ; Ferris FL ; Ott J ; Barnstable C ; Hoh J. Complement factor h polymorphism in age-related macular degeneration. Science, 308(5720) :385–389, April 2005.

[237] Clarke GM ; Anderson CA ; Pettersson FH ; Cardon LR ; Morris AP ; Zondervan KT. Basic statistical analysis in genetic case-control studies. Nat Protoc, 6(2) :121–133, February 2011.

[238] Manolio TA. Genomewide association studies and assessment of the risk of disease. N Engl J Med, 363(2) :166–176, July 2010.

[239] International HapMap Consortium. The international hapmap project. Nature, 426(6968) :789–796, December 2003.

[240] Cordell HJ. Epistasis : what it means, what it doesn't mean, and statistical methods to detect it in humans. Hum Mol Genet, 11(20) :2463–2468, October 2002.

[241] Devlin B ; Roeder K. Genomic control for association studies. Biometrics, 55(4) :997–1004, December 1999.

[242] Lewis CM ; Knight J. Introduction to genetic association studies. Cold Spring Harb Protoc, 2012(3) :297–306, March 2012.

[243] Grimes DA ; Schulz KF. Compared to what ? finding controls for case-control studies. Lancet, 365(9468) :1429–1433, April 2005.

[244] Samy Hadjadj ; Bertrand Cariou ; Frederic Fumeron ; Elise Gand ; Guillaume Charpentier ; Ronan Roussel ; Ahmed-Amine Kasmi ; Jean-François Gautier ; Kammel Mohammedi ; Pierre Gourdy ; Pierre-Jean Saulnier ; Eva Feigerlova ; Michel Marreon ; French JDRF Diabetic Nephropathy Collaborative Research Initiative (search for genes determining time to onset of ESRD in T1D patients with proteinuria) ; the SURDIAGENE and DIABHYCAR study groups. Death, end-stage renal disease and renal function decline in patients with diabetic nephropathy in french cohorts of type 1 and type 2 diabetes. Diabetologia, 59(1) :208–216, January 2016.

[245] Sandholm N ; Salem RM ; McKnight AJ ; Brennan EP ; Forsblom C ; Isakova T ; McKay GJ ; Williams WW ; Sadlier DM ; Mäkinen VP ; Swan EJ ; Palmer C ; Boright AP ; Ahlqvist E ; Deshmukh HA ; Keller BJ ; Huang H ; Ahola AJ ; Fagerholm E ; Gordin D ; Harjutsalo V ; He B ; Heikkilä O ; Hietala K ; Kytö J ; Lahermo P ; Lehto M ; Lithovius R ; Osterholm AM ; Parkkonen M ; Pitkäniemi J ; Rosengård-Bärlund M ; Saraheimo M ; Sarti C ; Söderlund J ; Soro-Paavonen A ; Syreeni A ; Thorn LM ; Tikkanen H ; Tolonen N ; Tryggvason K ; Tuomilehto J ; Wadén J ; Gill GV ; Prior S ; Guiducci C ; Mirel DB ; Taylor A ; Hosseini SM ; DCCT/EDIC Research Group ; Parving HH ; Rossing P ; Tarnow L ; Ladenvall C ; Alhenc-Gelas F ; Lefebvre P ; Rigalleau V ; Roussel R ; Tregouet DA ; Maestroni A ; Maestroni S ; Falhammar H ; Gu T ; Möllsten A ; Cimponeriu D ; Ioana M ; Mota M ; Mota E ; Serafinceanu C ; Stavarachi M ; Hanson RL ; Nelson RG ; Kretzler M ; Colhoun HM ; Panduru NM ; Gu HF ; Brismar K ; Zerbini G ; Hadjadj S ; Marre M ; Groop L ; Lajer M ; Bull SB ; Waggott D ; Paterson AD ; Savage DA ; Bain SC ; Martin F ; Hirschhorn JN ; Godson C ; Florez JC ; Groop PH ; Maxwell AP. New susceptibility loci associated with kidney disease in type 1 diabetes. PLoS Genet, 8(9) :e1002921, September 2012.

[246] DCCT/EDIC Research Group ; Nathan DM ; Bebu I ; Hainsworth D ; Klein R ; Tamborlane W ; Lorenzi G ; Gubitosi-Klug R ; Lachin JM. Frequency of evidence-based screening for retinopathy in type 1 diabetes. N Engl J Med, 376(16) :1507–1516, April 2017.

[247] Klein R ; Knudtson MD ; Lee KE ; Gangnon R Klein BE. The wisconsin epidemiologic study of diabetic retinopathy : Xxii the twenty-five-year progression of retinopathy in persons with type 1 diabetes. Ophthalmology, 115(11) :1859–1868, November 2008.

[248] Bjornstad P ; Maahs DM ; Rivard CJ ; Pyle L ; Rewers M ; Johnson RJ ; Snell-Bergeon JK. Serum uric acid predicts vascular complications in adults with type 1 diabetes : the coronary artery calcification in type 1 diabetes study. Acta Diabetol, 51(5) :783–791, October 2014.

[249] Orchard TJ ; Chang YF ; Ferrell RE ; Petro N ; Ellis DE. Nephropathy in type 1 diabetes : a manifestation of insulin resistance and multiple genetic susceptibilities ? further evidence from the pittsburgh epidemiology of diabetes complication study. Kidney Int, 62(3) :963–970, September 2002.

[250] Morange PE ; Oudot-Mellakh T ; Cohen W ; Germain M ; Saut N ; Guillemette A ; Alessi MC ; Bertrand M ; Dupuy AM ; Letenneur C ; Lathrop M ; Lopez LM ; Lambert JC ; Emmerich J ; Amouyel P ; Trégouët DA. Kng1 ile581thr and susceptibility to venous thrombosis. Blood, 117 :3692–3694, March 2011.

[251] David-Alexandre Trégouët ; Simon Heath ; Noémie Saut ; Christine Biron-Andreani ; Jean-François Schved ; Gilles Pernod ; Pilar Galan ; Ludovic Drouet ; Diana Zelenika ; Irène Juhan-Vague ; Marie-Christine Alessi ; Laurence Tiret ; Mark Lathrop ; Joseph Emmerich ; Pierre-Emmanuel Morange. Common susceptibility alleles are unlikely to contribute as strongly as the fv and abo loci to vte risk : results from a gwas approach. Blood, 113 :5298–5303, May 2009.

[252] Jeanet W. Blom ; Carine J. M. Doggen ; Susanne Osanto ; Frits R. Rosendaal. Malignancies, prothrombotic mutations, and the risk of venous thrombosis. JAMA, 293(6) :715–722, February 2005.

[253] Marine Germain ; Noémie Saut ; Nicolas Greliche ; Christian Dina ; Jean-Charles Lambert ; Claire Perret ; William Cohen ; Tiphaine Oudot-Mellakh ; Guillemette Antoni ; Marie-Christine Alessi ; Diana Zelenika ; François Cambien ; Laurence Tiret ; Marion Bertrand ; Anne-Marie Dupuy ; Luc Letenneur ; Mark Lathrop ; Joseph Emmerich ; Philippe Amouyel ; David-Alexandre Trégouët ; Pierre-Emmanuel Morange. Genetics of venous thrombosis : insights from a new genome wide association study. PLoS One, 6(9) :e25581, 2011.

[254] Hercberg S ; Galan P ; Preziosi P ; Bertrais S ; Mennen L ; Malvy D ; Roussel AM ; Favier A ; Briançon S. The su.vi.max study : a randomized, placebo-controlled trial of the health effects of antioxidant vitamins and minerals. Arch Intern Med, 164(21) :2335–2342, November 2004.

[255] Waldeyer C ; Karakas M ; Scheurle C ; Ojeda F ; Schnabel RB ; Zeller T ; Zengin E ; Westermann D ; Schrage B ; Bickel C ; Rupprecht HJ ; Lackner KJ ; Blankenberg S ; Seiffert M ; Sinning C. The predictive value of different equations for estimation of glomerular filtration rate in patients with coronary artery disease - results from the atherogene study. Int J Cardiol, 221 :908–913, October 2016.

[256] Quetelet A. Recherches sur le poids de l'homme aux différents âges. Nouveau mémoire de l'Académie Royale des Sciences et des Belles-lettres de Bruxelles, 1832.

[257] World Health Organization. Physical status : the use and interpretation of anthropometry. WHO Technical Report Series, 854, 1995.

[258] Romero-Corral A ; Somers VK ; Sierra-Johnson J ; Thomas RJ ; Collazo-Clavell ML ; Korinek J ; Allison TG ; Batsis JA ; Sert-Kuniyoshi FH ; Lopez-Jimenez F. Accuracy of body mass index in diagnosing obesity in the adult general population. Int J Obes (Lond), 36(6) :959–966, June 2008.

[259] Prospective Studies Collaboration ; Whitlock G ; Lewington S ; Sherliker P ; Clarke R ; Emberson J ; Halsey J ; Qizilbash N ; Collins R ; Peto R. Body-mass index and cause-specific mortality in 900 000 adults : collaborative analyses of 57 prospective studies. Lancet, 373(9669) :1083–1096, March 2009.

[260] Bhaskaran K ; Douglas I ; Forbes H ; dos Santos-Silva I ; Leon DA ; Smeeth L. Body-mass index and risk of 22 specific cancers : a population-based cohort study of 5·24 million uk adults. Lancet, 384(9945) :755–765, August 2014.

[261] Sanada H ; Yokokawa H ; Yoneda M ; Yatabe J ; Sasaki Yatabe M ; Williams SM ; Felder RA ; Jose PA. High body mass index is an important risk factor for the development of type 2 diabetes. Intern Med, 51(14) :1821–1826, 2012.

[262] Ha CE ; Bhagavan NV. Novel insights into the pleiotropic effects of human serum albumin in health and disease. Biochim Biophys Acta, 1830(12) :5486–8493, December 2013.

[263] Kidney Disease Improving Global Outcomes. Kdigo 2012 clinical practice guideline for the evaluation and management of chronic kidney disease. Kidney International Supplements, 3(1), January 2013.

[264] Stevens LA ; Coresh J ; Greene T ; Levey AS. Assessing kidney function–measured and estimated glomerular filtration rate. N Engl J Med, 354(23) :2473–2483, June 2006.

[265] Levey AS1 ; Bosch JP ; Lewis JB ; Greene T ; Rogers N ; Roth D. A more accurate method to estimate glomerular filtration rate from serum creatinine : a new prediction equation. modification of diet in renal disease study group. Ann Intern Med, 130(6) :461–470, March 1999.

[266] Levey AS ; Stevens LA ; Schmid CH ; Zhang YL ; Castro AF 3rd ; Feldman HI ; Kusek JW ; Eggers P ; Van Lente ; Greene T ; Coresh J ; CKD-EPI (Chronic Kidney Disease Epidemiology Collaboration). A new equation to estimate glomerular filtration rate. Ann Intern Med, 150(9) :604–612, May 2009.

[267] Rohlfing CL ; Wiedmeyer HM ; Little RR ; England JD ; Tennill A ; Goldstein DE. Defining the relationship between plasma glucose and hba(1c) : analysis of glucose profiles and hba(1c) in the diabetes control and complications trial. Diabetes Care, 25(2) :275–278, February 2002.

[268] Castoldi E ; Rosing J. Thrombin generation tests. Thromb Res, 127 Suppl 3 :S21–25, February 2011.

[269] Dzik S ; Moroff G ; Dumont L. A multicenter study evaluating three methods for counting residual wbcs in wbc-reduced blood components : Nageotte hemocytometry, flow cytometry, and microfluorometry. Transfusion, 40(5) :513–520, May 2000.

[270] Bounameaux H ; Schneider PA ; Reber G ; de Moerloose P ; Krahenbuhl B. Measurement of plasma d-dimer for diagnosis of deep venous thrombosis. Am J Clin Pathol, 91(1) :82–85, January 1989.

[271] Taub FE ; DeLeo JM ; Thompson EB. Sequential comparative hybridizations analyzed by computerized image processing can identify and quantitate regulated rnas. DNA, 2(4) :309–327, 1983.

[272] LaFramboise T. Single nucleotide polymorphism arrays : a decade of biological, computational and technological advances. Nucleic Acids Res, 37(13) :4181–4193, July 2009.

[273] https ://www.r-project.org/.

[274] http ://zzz.bwh.harvard.edu/plink/index.shtml.

[275] Purcell S ; Neale B ; Todd-Brown K ; Thomas L ; Ferreira MA ; Bender D ; Maller J ; Sklar P ; de Bakker PI ; Daly MJ ; Sham PC. Plink : a tool set for whole-genome association and population-based linkage analyses. Am J Hum Genet, 81(3) :559–575, September 2007.

[276] Yang J ; Lee SH ; Goddard ME ; Visscher PM. Gcta : a tool for genome-wide complex trait analysis. Am J Hum Genet, 88(1) :76–82, January 2011.

[277] Pearson K. On lines and planes of closest fit to systems of points in space. Philosophical Magazine, 2(11) :559–572, 1901.

[278] Hardy GH. Mendelian proportions in a mixed population. Science, 28(706) :49–50, July 1908.

[279] Weinberg W. Über den nachweis der vererbung beim menschen. Jahreshefte des Vereins für vaterländische Naturkunde in Württemberg, 64 :368–382, 1908.

[280] Pearson K. On the criterion that a given system of deviations from the probable in the case of a correlated system of variables is such that it can be reasonably supposed to have arisen from random sampling. Philosophical Magazine, 50(302) :157–175, 1900.

[281] 1000 Genomes Project Consortium ; Abecasis GR ; Altshuler D ; Auton A ; Brooks LD ; Durbin RM ; Gibbs RA ; Hurles ME ; McVean GA. A map of human genome variation from population-scale sequencing. Nature, 467(7319) :1061–1073, October 2010.

[282] Li Y ; Willer C ; Sanna S ; Abecasis G. Genotype imputation. Annu Rev Genomics Hum Genet, 10 :387–406, 2009.

[283] Howie BN ; Donnelly P ; Marchini J. A flexible and accurate genotype imputation method for the next generation of genome-wide association studies. PLoS Genet, 5(6) :e1000529, June 2009.

[284] Li Y ; Willer CJ ; Ding J ; Scheet P ; Abecasis GR. Mach : using sequence and genotype data to estimate haplotypes and unobserved genotypes. Genet Epidemiol, 34(8) :816–834, December 2010.

[285] https ://cran.r-project.org/web/packages/rmeta/index.html.

[286] Willer CJ ; Li Y ; Abecasis GR. Metal : fast and efficient meta-analysis of genomewide association scans. Bioinformatics, 26(17) :2190–2191, September 2010.

[287] Skol AD ; Scott LJ ; Abecasis GR ; Boehnke M. Joint analysis is more efficient than replication-based analysis for two-stage genome-wide association studies. Nat Genet, 38(2) :209–213, February 2006.

[288] 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, 38(8) :904–909, August 2006.

[289] Lengendre AM. Nouvelles méthodes pour la détermination des orbites des comètes. F. Didot, Rue de Thionville, Paris, France, 1805.

[290] Cochran WG. The comparison of percentages in matched samples. Biometrika, 37(3) :256–266, December 1950.

[291] Higgins JP ; Thompson SG ; Deeks JJ ; Altman DG. Measuring inconsistency in meta-analyses. BMJ, 327(7414) :557–560, September 2003.

[292] Devlin B ; Roeder K ; Wasserman L. Genomic control, a new approach to genetic-based association studies. Theor Popul Biol, 60(3) :155–166, November 2001.

[293] Hsieh FY ; Bloch DA ; Larsen MD. A simple method of sample size calculation for linear and logistic regression. Stat Med, 17(14) :1623–1634, July 1998.

[294] Akaike H. A new look at the statistical model identification. IEEE Trans Autom Control, 19 :716–723, 1974.

[295] Kullback S ; Leibler RA. On information and sufficiency. Ann Math Statist, 1 :79–86, 1951.

[296] Ng PC ; Henikoff S. Sift : Predicting amino acid changes that affect protein function. Nucleic Acids Res, 31(13) :3812–3814, July 2003.

[297] Kircher M ; Witten DM ; Jain P ; O'Roak BJ ; Cooper GM ; Shendure J. A general framework for estimating the relative pathogenicity of human genetic variants. Nat Genet, 46(3) :310–315, March 2014.

[298] Dudbridge F. Power and predictive accuracy of polygenic risk scores. PLoS Genet, 9(3) :e1003348, March 2013.

[299] Guo W ; Samuels JF ; Wang Y ; Cao H ; Ritter M ; Nestadt PS ; Krasnow J ; Greenberg BD ; Fyer AJ ; McCracken JT ; Geller DA ; Murphy DL ; Knowles JA ; Grados MA ; Riddle MA ; Rasmussen SA ; McLaughlin NC ; Nurmi EL ; Askland KD ; Cullen BA ; Piacentini J ; Pauls DL ; Bienvenu OJ ; Stewart SE ; Goes FS ; Maher B ; Pulver AE ; Valle D ; Mattheisen M ; Qian J ; Nestadt G ; Shugart YY. Polygenic risk score and heritability estimates reveals a genetic relationship between asd and ocd. Eur Neuropsychopharmacol, 27(7) :657–666, July 2017.

[300] Burgess S ; Butterworth A ; Thompson SG. Mendelian randomization analysis with multiple genetic variants using summarized data. Genet Epidemiol, 37(7) :658–665, November 2013.

[301] Wang K ; Li M ; Hakonarson H. Annovar : functional annotation of genetic variants from high-throughput sequencing data. Nucleic Acids Res, 38(16) :e164, September 2010.

[302] Rockman MV ; Kruglyak. Genetics of global gene expression. Nat Rev Genet, 7(11) :862–872, November 2006.

[303] Bernstein BE ; Stamatoyannopoulos JA ; Costello JF ; Ren B ; Milosavljevic A ; Meissner A ; Kellis M ; Marra MA ; Beaudet AL ; Ecker JR ; Farnham PJ ; Hirst M ; Lander ES ; Mikkelsen TS ; Thomson JA. The nih roadmap epigenomics mapping consortium. Nat Biotechnol, 28(10) :1045–1048, October 2010.

[304] ENCODE Project Consortium. An integrated encyclopedia of dna elements in the human genome. Nature, 489(7414) :57–74, September 2012.

[305] Rebhan M ; Chalifa-Caspi V ; Prilusky J ; Lancet D. Genecards : integrating information about genes, proteins and diseases. Trends Genet, 13(4) :163, April 1997.

[306] Hamosh A ; Scott AF ; Amberger JS ; Bocchini CA ; McKusick VA. Online mendelian inheritance in man (omim), a knowledgebase of human genes and genetic disorders. Nucleic Acids Res, 33 :D514–517, January 2005.

[307] Warde-Farley D ; Donaldson SL ; Comes O ; Zuberi K ; Badrawi R ; Chao P ; Franz M ; Grouios C ; Kazi F ; Lopes CT ; Maitland A ; Mostafavi S ; Montojo J ; Shao Q ; Wright G ; Bader GD ; Morris Q. The genemania prediction server : biological network integration for gene prioritization and predicting gene function. Nucleic Acids Res, 38 :W214–220, July 2010.

[308] Szklarczyk D ; Morris JH ; Cook H ; Kuhn M ; Wyder S ; Simonovic M ; Santos A ; Doncheva NT ; Roth A ; Bork P ; Jensen LJ ; von Mering C. The string database in 2017 : quality-controlled protein-protein association networks, made broadly accessible. Nucleic Acids Res, 45(D1) :D362–D368, January 2017.

[309] Leslie R ; O'Donnell CJ ; Johnson AD. Grasp : analysis of genotype-phenotype results from 1390 genome-wide association studies and corresponding open access database. Bioinformatics, 30(12) :i185–194, June 2014.

[310] Carithers LJ ; Ardlie K ; Barcus M ; Branton PA ; Britton A ; Buia SA ; Compton CC ; DeLuca DS ; Peter-Demchok J ; Gelfand ET ; Guan P ; Korzeniewski GE ; Lockhart NC ; Rabiner CA ; Rao AK ; Robinson KL ; Roche NV ; Sawyer SJ ; Segrè AV ; Shive CE ; Smith AM ; Sobin LH ; Undale AH ; Valentino KM ; Vaught J ; Young TR ; Moore HM ; GTEx Consortium. A novel approach to high-quality postmortem tissue procurement : The gtex project. Biopreserv Biobank, 13(5) :311–319, October 2015.

[311] Ward LD ; Kellis M. Haploreg : a resource for exploring chromatin states, conservation, and regulatory motif alterations within sets of genetically linked variants. Nucleic Acids Res, 40 :D930–934, January 2012.

[312] Machiela MJ ; Chanock SJ. Ldlink : a web-based application for exploring population-specific haplotype structure and linking correlated alleles of possible functional variants. Bioinformatics, 31(21) :3555–3557, November 2015.

[313] Pruim RJ ; Welch RP ; Sanna S ; Teslovich TM ; Chines PS ; Gliedt TP ; Boehnke M ; Abecasis GR ; Willer CJ. Locuszoom : regional visualization of genome-wide association scan results. Bioinformatics, 26(18) :2336–2337, September 2010.

[314] Charmet R ; Duffy S ; Keshavarzi S ; Gyorgy B ; Marre M ; Rossing P ; McKnight AJ ; Maxwell AP ; Ahluwalia TVS ; Paterson AD ; Trégouët DA ; Hadjadj S. Novel risk genes identified in a genome-wide association study for coronary artery disease in patients with type 1 diabetes. Cardiovasc Diabetol, 17(1) :61, April 2018.

[315] Okuda T ; Cleveland JL ; Downing JR. Pctaire-1 and pctaire-3, two members of a novel cdc2/cdc28-related protein kinase gene family. Oncogene, 7(11) :2249–2258, November 1992.

[316] Bradfield JP ; Qu HQ ; Wang K ; Zhang H ; Sleiman PM ; Kim CE ; Mentch FD ; Qiu H ; Glessner JT ; Bradfield JP1 ; Qu HQ ; Wang K ; Zhang H ; Sleiman PM ; Kim CE ; Mentch FD ; Qiu H ; Glessner JT ; Thomas KA ; Frackelton EC ; Chiavacci RM ; Imielinski M ; Monos DS ; Pandey R ; Bakay M ; Grant SF ; Polychronakos C ; Hakonarson H. A genome-wide meta-analysis of six type 1 diabetes cohorts identifies multiple associated loci. PLoS Genet, 7(9) :e1002293, September 2011.

[317] Qian F ; Boletta A ; Bhunia AK ; Xu H ; Liu L ; Ahrabi AK ; Watnick TJ ; Zhou F ; Germino GG. Cleavage of polycystin-1 requires the receptor for egg jelly domain and is disrupted by human autosomal-dominant polycystic kidney disease 1-associated mutations. Proc Natl Acad Sci U S A, 99(26) :16981–16986, December 2002.

[318] Polycystic kidney disease : the complete structure of the pkd1 gene and its protein. the international polycystic kidney disease consortium. Cell, 81(2) :289–298, April 1995.

[319] Böger CA ; Chen MH ; Tin A ; Olden M ; Köttgen A ; de Boer IH ; Fuchsberger C ; O'Seaghdha CM ; Pattaro C ; Teumer A ; Liu CT ; Glazer NL ; Li M ; O'Connell JR ; Tanaka T ; Peralta CA ; Kutalik Z ; Luan J ; Zhao JH ; Hwang SJ ; Akylbekova E ; Kramer H ; van der Harst P ; Smith AV ; Lohman K ; de Andrade M ; Hayward C ; Kollerits B ; Tönjes A ; Aspelund T ; Ingelsson E ; Eiriksdottir G ; Launer LJ ; Harris TB ; Shuldiner AR ; Mitchell BD ; Arking DE ; Franceschini N ; Boerwinkle E ; Egan J ; Hernandez D ; Reilly M ; Townsend RR ; Lumley T ; Siscovick DS ; Psaty BM ; Kestenbaum B ; Haritunians T ; Bergmann S ; Vollenweider P ; Waeber G ; Mooser V ; Waterworth D ; Johnson AD ; Florez JC ; Meigs JB ; Lu X ; Turner ST ; Atkinson EJ ; Leak TS ; Aasarød K ; Skorpen F ; Syvänen AC ; Illig T ; Baumert J ; Koenig W ; Krämer BK ; Devuyst O ; Mychaleckyj JC ; Minelli C ; Bakker SJ ; Kedenko L ; Paulweber B ; Coassin S ; Endlich K ; Kroemer HK ; Biffar R ; Stracke S ; Völzke H ; Stumvoll M ; Mägi R ; Campbell H ; Vitart V ; Hastie ND ; Gudnason V ; Kardia SL ; Liu Y ; Polasek O ; Curhan G ; Kronenberg F ; Prokopenko I ; Rudan I ; Arnlöv J ; Hallan S ; Navis G ; CKDGen Consortium ; Parsa A ; Ferrucci L ; Coresh J ; Shlipak MG ; Bull SB ; Paterson NJ ; Wichmann HE ; Wareham NJ ; Loos RJ ; Rotter JI ; Pramstaller PP ; Cupples LA ; Beckmann JS ; Yang Q ; Heid IM ; Rettig R ; Dreisbach AW ; Bochud M ; Fox CS ; Kao WH. Cubn is a gene locus for albuminuria. J Am Soc Nephrol, 22(3) :555–570, March 2011.

[320] Pandey A ; Blagoev B ; Kratchmarova I ; Fernandez M ; Nielsen M ; Kristiansen TZ ; Ohara O ; Podtelejnikov AV ; Roche S ; Lodish HF ; Mann M. Cloning of a novel phosphotyrosine binding domain containing molecule, odin, involved in signaling by receptor tyrosine kinases. Oncogene, 21(52) :8029–8036, November 2002.

[321] Berndt SI ; Gustafsson S ; Mägi R ; Ganna A ; Wheeler E ; Feitosa MF ; Justice AE ; Monda KL ; Croteau-Chonka DC ; Day FR ; Esko T ; Fall T ; Ferreira T ; Gentilini D ; Jackson AU ; Luan J ; Randall JC ; Vedantam S ; Willer CJ ; Winkler TW ; Wood AR ; Workalemahu T ; Hu YJ ; Lee SH ; Liang L ; Lin DY ; Min JL ; Neale BM ; Thorleifsson G ; Yang J ; Albrecht E ; Amin N ; Bragg-Gresham JL ; Cadby G ; den Heijer M ; Eklund N ; Fischer K ; Goel A ; Hottenga JJ ; Huffman JE ; Jarick I ; Johansson Å ; Johnson T ; Kanoni S ; Kleber ME ; König IR ; Kristiansson K ; Kutalik Z ; Lamina C ; Lecoeur C ; Li G ; Mangino M ; McArdle WL ; Medina-Gomez C ; Müller-Nurasyid M ; Ngwa JS ; Nolte IM ; Paternoster L ; Pechlivanis S ; Perola M ; Peters MJ ; Preuss M ; Rose LM ; Shi J ; Shungin D ; Smith AV ; Strawbridge RJ ; Surakka I ; Teumer A ; Trip MD ; Tyrer J ; Van Vliet-Ostaptchouk JV ; Vandenput L ; Waite LL ; Zhao JH ; Absher D ; Asselbergs FW ; Atalay M ; Attwood AP ; Balmforth AJ ; Basart H ; Beilby J ; Bonnycastle LL ; Brambilla P ; Bruinenberg M ; Campbell H ; Chasman DI ; Chines PS ; Collins FS ; Connell JM ; Cookson WO ; de Faire U ; de Vegt F ; Dei M ; Dimitriou M ;

Edkins S ; Estrada K ; Evans DM ; Farrall M ; Ferrario MM ; Ferrières J ; Franke L ; Frau F ; Gejman PV ; Grallert H ; Grönberg H ; Gudnason V ; Hall AS ; Hall P ; Hartikainen AL ; Hayward C ; Heard-Costa NL ; Heath AC ; Hebebrand J ; Homuth G ; Hu FB ; Hunt SE ; Hyppönen E ; Iribarren C ; Jacobs KB ; Jansson JO ; Jula A ; Kähönen M ; Kathiresan S ; Kee F ; Khaw KT ; Kivimäki M ; Koenig W ; Kraja AT ; Kumari M ; Kuulasmaa K ; Kuusisto J ; Laitinen JH ; Lakka TA ; Langenberg C ; Launer LJ ; Lind L ; Lindström J ; Liu J ; Liuzzi A ; Lokki ML ; Lorentzon M ; Madden PA ; Magnusson PK ; Manunta P ; Marek D ; März W ; Mateo Leach I ; McKnight B ; Medland SE ; Mihailov E ; Milani L ; Montgomery GW ; Mooser V ; Mühleisen TW ; Munroe PB ; Musk AW ; Narisu N ; Navis G ; Nicholson G ; Nohr EA ; Ong KK ; Oostra BA ; Palmer CN ; Palotie A ; Peden JF ; Pedersen N ; Peters A ; Polasek O ; Pouta A ; Pramstaller PP ; Prokopenko I ; Pütter C ; Radhakrishnan A ; Raitakari O ; Rendon A ; Rivadeneira F ; Rudan I ; Saaristo TE ; Sambrook JG ; Sanders AR ; Sanna S ; Saramies J ; Schipf S ; Schreiber S ; Schunkert H ; Shin SY ; Signorini S ; Sinisalo J ; Skrobek B ; Soranzo N ; Stančáková A ; Stark K ; Stephens JC ; Stirrups K ; Stolk RP ; Stumvoll M ; Swift AJ ; Theodoraki EV ; Thorand B ; Tregouet DA ; Tremoli E ; Van der Klauw MM ; van Meurs JB ; Vermeulen SH ; Viikari J ; Virtamo J ; Vitart V ; Waeber G ; Wang Z ; Widén E ; Wild SH ; Willemsen G ; Winkelmann BR ; Witteman JC ; Wolffenbuttel BH ; Wong A ; Wright AF ; Zillikens MC ; Amouyel P ; Boehm BO ; Boerwinkle E ; Boomsma DI ; Caulfield MJ ; Chanock SJ ; Cupples LA ; Cusi D ; Dedoussis GV ; Erdmann J ; Eriksson JG ; Franks PW ; Froguel P ; Gieger C ; Gyllensten U ; Hamsten A ; Harris TB ; Hengstenberg C ; Hicks AA ; Hingorani A ; Hinney A ; Hofman A ; Hovingh KG ; Hveem K ; Illig T ; Jarvelin MR ; Jöckel KH ; Keinanen-Kiukaanniemi SM ; Kiemeney LA ; Kuh D ; Laakso M ; Lehtimäki T ; Levinson DF ; Martin NG ; Metspalu A ; Morris AD ; Nieminen MS ; Njølstad I ; Ohlsson C ; Oldehinkel AJ ; Ouwehand WH ; Palmer LJ ; Penninx B ; Power C ; Province MA ; Psaty BM ; Qi L ; Rauramaa R ; Ridker PM ; Ripatti S ; Salomaa V ; Samani NJ ; Snieder H ; Sørensen TI ; Spector TD ; Stefansson K ; Tönjes A ; Tuomilehto J ; Uitterlinden AG ; Uusitupa M ; van der Harst P ; Vollenweider P ; Wallaschofski H ; Wareham NJ ; Watkins H ; Wichmann HE ; Wilson JF ; Abecasis GR ; Assimes TL ; Barroso I ; Boehnke M ; Borecki IB ; Deloukas P ; Fox CS ; Frayling T ; Groop LC ; Haritunian T ; Heid IM ; Hunter D ; Kaplan RC ; Karpe F ; Moffatt MF ; Mohlke KL ; O'Connell JR ; Pawitan Y ; Schadt EE ; Schlessinger D ; Steinthorsdottir V ; Strachan DP ; Thorsteinsdottir U ; van Duijn CM ; Visscher PM ; Di Blasio AM ; Hirschhorn JN ; Lindgren CM ; Morris AP ; Meyre D ; Scherag A ; McCarthy MI ; Speliotes EK ; North KE ; Loos RJ ; Ingelsson E. Genome-wide meta-analysis identifies 11 new loci for anthropometric traits and provides insights into genetic architecture. Nat Genet, 45(5) :501–512, May 2013.

[322] Lango Allen H ; Estrada K ; Lettre G ; Berndt SI ; Weedon MN ; Rivadeneira F ; Willer CJ ; Jackson AU ; Vedantam S ; Raychaudhuri S ; Ferreira T ; Wood AR ; Weyant RJ ; Segrè AV ; Speliotes EK ; Wheeler E ; Soranzo N ; Park JH ; Yang J ; Gudbjartsson D ; Heard-Costa NL ; Randall JC ; Qi L ; Vernon Smith A ; Mägi R ; Pastinen T ; Liang L ; Heid IM ; Luan J ; Thorleifsson G ; Winkler TW ; Goddard ME ; Sin Lo K ; Palmer C ; Workalemahu T ; Aulchenko YS ; Johansson A ; Zillikens MC ; Feitosa MF ; Esko T ; Johnson T ; Ketkar S ; Kraft P ; Mangino M ; Prokopenko I ; Absher D ; Albrecht E ; Ernst F ; Glazer NL ; Hayward C ; Hottenga JJ ; Jacobs KB ; Knowles JW ; Kutalik Z ; Monda KL ; Polasek O ; Preuss M ; Rayner NW ; Robertson NR ; Steinthorsdottir V ; Tyrer JP ; Voight BF ; Wiklund F ; Xu J ; Zhao JH ; Nyholt DR ; Pellikka N ; Perola M ; Perry JR ; Surakka I ; Tammesoo ML ; Altmaier EL ; Amin N ; Aspelund T ; Bhangale T ; Boucher G ; Chasman DI ; Chen C ; Coin L ; Cooper MN ; Dixon AL ; Gibson Q ; Grundberg E ; Hao K ; Juhani Junttila M ; Kaplan LM ; Kettunen J ; König IR ; Kwan T ; Lawrence RW ; Levinson DF ; Lorentzon M ; McKnight B ; Morris AP ; Müller M ; Suh Ngwa J ; Purcell S ; Rafelt S ; Salem RM ; Salvi E ; Sanna S ; Shi J ; Sovio U ; Thompson JR ; Turchin MC ; Vandenput L ; Verlaan DJ ; Vitart V ; White CC ; Ziegler A ; Almgren P ; Balmforth AJ ; Campbell H ; Citterio

L ; De Grandi A ; Dominiczak A ; Duan J ; Elliott P ; Elosua R ; Eriksson JG ; Freimer NB ; Geus EJ ; Glorioso N ; Haiqing S ; Hartikainen AL ; Havulinna AS ; Hicks AA ; Hui J ; Igl W ; Illig T ; Jula A ; Kajantie E ; Kilpeläinen TO ; Koiranen M ; Kolcic I ; Koskinen S ; Kovacs P ; Laitinen J ; Liu J ; Lokki ML ; Marusic A ; Maschio A ; Meitinger T ; Mulas A ; Paré G ; Parker AN ; Peden JF ; Petersmann A ; Pichler I ; Pietiläinen KH ; Pouta A ; Ridderstråle M ; Rotter JI ; Sambrook JG ; Sanders AR ; Schmidt CO ; Sinisalo J ; Smit JH ; Stringham HM ; Bragi Walters G ; Widen E ; Wild SH ; Willemsen G ; Zagato L ; Zgaga L ; Zitting P ; Alavere H ; Farrall M ; McArdle WL ; Nelis M ; Peters MJ ; Ripatti S ; van Meurs JB ; Aben KK ; Ardlie KG ; Beckmann JS ; Beilby JP ; Bergman RN ; Bergmann S ; Collins FS ; Cusi D ; den Heijer M ; Eiriksdottir G ; Gejman PV ; Hall AS ; Hamsten A ; Huikuri HV ; Iribarren C ; Kähönen M ; Kaprio J ; Kathiresan S ; Kiemeney L ; Kocher T ; Launer LJ ; Lehtimäki T ; Melander O ; Mosley TH Jr ; Musk AW ; Nieminen MS ; O'Donnell CJ ; Ohlsson C ; Oostra B ; Palmer LJ ; Raitakari O ; Ridker PM ; Rioux JD ; Rissanen A ; Rivolta C ; Schunkert H ; Shuldiner AR ; Siscovick DS ; Stumvoll M ; Tönjes A ; Tuomilehto J ; van Ommen GJ ; Viikari J ; Heath AC ; Martin NG ; Montgomery GW ; Province MA ; Kayser M ; Arnold AM ; Atwood LD ; Boerwinkle E ; Chanock SJ ; Deloukas P ; Gieger C ; Grönberg H ; Hall P ; Hattersley AT ; Hengstenberg C ; Hoffman W ; Lathrop GM ; Salomaa V ; Schreiber S ; Uda M ; Waterworth D ; Wright AF ; Assimes TL ; Barroso I ; Hofman A ; Mohlke KL ; Boomsma DI ; Caulfield MJ ; Cupples LA ; Erdmann J ; Fox CS ; Gudnason V ; Gyllensten U ; Harris TB ; Hayes RB ; Jarvelin MR ; Mooser V ; Munroe PB ; Ouwehand WH ; Penninx BW ; Pramstaller PP ; Quertermous T ; Rudan I ; Samani NJ ; Spector TD ; Völzke H ; Watkins H ; Wilson JF ; Groop LC ; Haritunians T ; Hu FB ; Kaplan RC ; Metspalu A ; North KE ; Schlessinger D ; Wareham NJ ; Hunter DJ ; O'Connell JR ; Strachan DP ; Wichmann HE ; Borecki IB ; van Duijn CM ; Schadt EE ; Thorsteinsdottir U ; Peltonen L ; Uitterlinden AG ; Visscher PM ; Chatterjee N ; Loos RJ ; Boehnke M ; McCarthy MI ; Ingelsson E ; Lindgren CM ; Abecasis GR ; Stefansson K ; Frayling TM ; Hirschhorn JN. Hundreds of variants clustered in genomic loci and biological pathways affect human height. Nature, 467(7317) :832–838, October 2010.

[323] Hopewell JC ; Parish S ; Offer A ; Link E ; Clarke R ; Lathrop M ; Armitage J ; Collins R ; MRC/BHF Heart Protection Study Collaborative Group. Impact of common genetic variation on response to simvastatin therapy among 18 705 participants in the heart protection study. Eur Heart J, 34(13) :982–992, April 2013.

[324] Charmet R ; van Hylckama Vlieg A ; Germain M ; Roussel R ; Marre M ; Debette S ; Amouyel P ; Deleuze JF ; Hadjadj S ; Rosendaal FR ; Morange PE ; Trégouët DA. Association of impaired renal function with venous thrombosis : A genetic risk score approach. Thromb Res, 158 :102–107, October 2017.

[325] Dennis J ; Johnson CY ; Adediran AS ; de Andrade M ; Heit JA ; Morange PE ; Trégouët DA ; Gagnon F. The endothelial protein c receptor (procr) ser219gly variant and risk of common thrombotic disorders : a huge review and meta-analysis of evidence from observational studies. Blood, 119(10) :2392–2400, March 2012.

[326] Tanaka K ; Okamoto S ; Ishikawa Y ; Tamura H ; Hara T. Ddx1 is required for testicular tumorigenesis, partially through the transcriptional activation of 12p stem cell genes. Oncogene, 28(21) :2142–2151, May 2009.

[327] Turnbull C ; Perdeaux ER ; Pernet D ; Naranjo A ; Renwick A ; Seal S ; Munoz-Xicola RM ; Hanks S ; Slade I ; Zachariou A ; Warren-Perry M ; Ruark E ; Gerrard M ; Hale J ; Hewitt M ; Kohler J ; Lane S ; Levitt G ; Madi M ; Morland B ; Neefjes V ; Nicholson J ; Picton S ; Pizer B ; Ronghe M ; Stevens M ; Traunecker H ; Stiller CA ; Pritchard-Jones K ; Dome J ; Grundy P ; Rahman N. A genome-wide association study identifies susceptibility loci for wilms tumor. Nat Genet, 44(6) :681–684, April 2012.

[328] Stroschein SL ; Wang W ; Zhou S ; Zhou Q ; Luo K. Negative feedback regulation of tgf-beta signaling by the snon oncoprotein. Science, 286(5440) :771–774, October 1999.

[329] Prié D ; Huart V ; Bakouh N ; Planelles G ; Dellis O ; Gérard B ; Hulin P ; Benqué-Blanchet F ; Silve C ; Grandchamp B ; Friedlander G. Nephrolithiasis and osteoporosis associated with hypophosphatemia caused by mutations in the type 2a sodium-phosphate cotransporter. N Engl J Med, 347(13) :983–991, September 2002.

[330] Magen D ; Berger L ; Coady MJ ; Ilivitzki A ; Militianu D ; Tieder M ; Selig S ; Lapointe JY ; Zelikovic I ; Skorecki K. A loss-of-function mutation in napi-iia and renal fanconi's syndrome. N Engl J Med, 362(12) :1102–1109, March 2010.

[331] Williams FM ; Carter AM ; Hysi PG ; Surdulescu G ; Hodgkiss D ; Soranzo N ; Traylor M ; Bevan S ; Dichgans M ; Rothwell PM ; Sudlow C ; Farrall M ; Silander K ; Kaunisto M ; Wagner P ; Saarela O ; Kuulasmaa K ; Virtamo J ; Salomaa V ; Amouyel P ; Arveiler D ; Ferrieres J ; Wiklund PG ; Ikram MA ; Hofman A ; Boncoraglio GB ; Parati EA ; Helgadottir A ; Gretarsdottir S ; Thorsteinsdottir U ; Thorleifsson G ; Stefansson K ; Seshadri S ; DeStefano A ; Gschwendtner A ; Psaty B ; Longstreth W ; Mitchell BD ; Cheng YC ; Clarke R ; Ferrario M ; Bis JC ; Levi C ; Attia J ; Holliday EG ; Scott RJ ; Fornage M ; Sharma P ; Furie KL ; Rosand J ; Nalls M ; Meschia J ; Mosely TH ; Evans A ; Palotie A ; Markus HS ; Grant PJ ; Spector TD ; EuroCLOT Investigators ; Wellcome Trust Case Control Consortium 2 ; MOnica Risk ; Genetics ; Archiving and Monograph ; MetaStroke ; International Stroke Genetics Consortium. Ischemic stroke is associated with the abo locus : the euroclot study. Ann Neurol, 73(1) :16–31, January 2013.

[332] Tang W ; Schwienbacher C ; Lopez LM ; Ben-Shlomo Y ; Oudot-Mellakh T ; Johnson AD ; Samani NJ ; Basu S ; Gögele M ; Davies G ; Lowe GD ; Tregouet DA ; Tan A ; Pankow JS ; Tenesa A ; Levy D ; Volpato CB ; Rumley A ; Gow AJ ; Minelli C ; Yarnell JW ; Porteous DJ ; Starr JM ; Gallacher J ; Boerwinkle E ; Visscher PM ; Pramstaller PP ; Cushman M ; Emilsson V ; Plump AS ; Matijevic N ; Morange PE ; Deary IJ ; Hicks AA ; Folsom AR. Genetic associations for activated partial thromboplastin time and prothrombin time, their gene expression profiles, and risk of coronary artery disease. Am J Hum Genet, 91(1) :152–162, July 2012.

[333] Wagner GF ; Vozzolo BL ; Jaworski E ; Haddad M ; Kline RL ; Olsen HS ; Rosen CA ; Davidson MB ; Renfro JL. Human stanniocalcin inhibits renal phosphate excretion in the rat. J Bone Miner Res, 12(2) :165–171, February 1997.

[334] Bridges LC ; Tani PH ; Hanson KR ; Roberts CM ; Judkins MB ; Bowditch RD. The lymphocyte metalloprotease mdc-l (adam 28) is a ligand for the integrin alpha4beta1. J Biol Chem, 277(5) :3784–3792, February 2002.

[335] Yao I ; Takagi H ; Ageta H ; Kahyo T ; Sato S ; Hatanaka K ; Fukuda Y ; Chiba T ; Morone N ; Yuasa S ; Inokuchi K ; Ohtsuka T ; Macgregor GR ; Tanaka K ; Setou M. Scrapper-dependent ubiquitination of active zone protein rim1 regulates synaptic vesicle release. Cell, 130(5) :943–957, September 2007.

[336] Feliubadaló L ; Font M ; Purroy J ; Rousaud F ; Estivill X ; Nunes V ; Golomb E ; Centola M ; Aksentijevich I ; Kreiss Y ; Goldman B ; Pras M ; Kastner DL ; Pras E ; Gasparini P ; Bisceglia L ; Beccia E ; Gallucci M ; de Sanctis L ; Ponzone A ; Rizzoni GF ; Zelante L ; Bassi MT ; George AL Jr ; Manzoni M ; De Grandi A ; Riboni M ; Endsley JK ; Ballabio A ; Borsani G ; Reig N ; Fernández E ; Estévez R ; Pineda M ; Torrents D ; Camps M ; Lloberas J ; Zorzano A ; Palacín M ; International Cystinuria Consortium. Non-type i cystinuria caused by mutations in slc7a9, encoding a subunit (bo,+at) of rbat. Nat Genet, 23(1) :52–57, September 1999.

[337] Tang W ; Basu S ; Kong X ; Pankow JS ; Aleksic N ; Tan A ; Cushman M ; Boerwinkle E ; Folsom AR. Genome-wide association study identifies novel loci for plasma levels of protein c : the aric study. Blood, 116(23) :5032–5036, December 2010.

[338] Brown DM ; Kania N ; Lantieri LA ; Pasia EN ; Khouri RK. Effect of phosphate concentration in buffered irrigating solutions on arterial thrombosis. Microsurgery, 16(11) :757–759, 1995.

[339] Thorn LM ; Forsblom C ; Fagerudd J ; Thomas MC ; Pettersson-Fernholm K ; Saraheimo M ; Wadén J ; Rönnback M ; Rosengård-Bärlund M ; Björkesten CG ; Taskinen MR ; Groop PH ; FinnDiane Study Group. Metabolic syndrome in type 1 diabetes : association with diabetic nephropathy and glycemic control (the finndiane study). Diabetes Care, 28(8) :2019–2024, August 2005.

[340] Akbar T ; McGurnaghan S ; Palmer CNA ; Livingstone SJ ; Petrie J ; Chalmers J ; Lindsay RS ; McKnight JA ; Pearson DWM ; Patrick AW ; Walker J ; Looker HC ; Colhoun HM. Cohort profile : Scottish diabetes research network type 1 bioresource study (sdrnt1bio). Int J Epidemiol, 46(3) :796–796i, June 2017.

Annexes

Travaux publiés ou soumis

Article publié Charmet R, van Hylckama Vlieg A, Germain M, Roussel R, Marre M, Debette S, Amouyel P, Deleuze JF, Hadjadj S, Rosendaal FR, Morange PE, Trégouët DA. *Association of impaired renal function with venous thrombosis : A genetic risk score approach.* Thromb Res. 2017 Oct ;158 :102-107.

Article publié Charmet R, Duffy S, Keshavarzi S, Gyorgy B, Marre M, Rossing P, McKnight AJ, Maxwell AP, Ahluwalia TVS, Paterson AD, Trégouët DA, Hadjadj S. *Novel risk genes identified in a genome-wide association study for coronary artery disease in patients with type 1 diabetes.* Cardiovasc Diabetol. 2018 Apr 25 ;17(1) :61.

Association of impaired renal function with venous thrombosis : A genetic risk score approach

Thrombosis Research 158 (2017) 102–107

Contents lists available at ScienceDirect

Thrombosis Research

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

Full Length Article

Association of impaired renal function with venous thrombosis: A genetic risk score approach

Romain Charmet[a,b], Astrid van Hylckama Vlieg[c], Marine Germain[a,b], Ronan Roussel[d,e], Michel Marre[d,e], Stéphanie Debette[f,g], Philippe Amouyel[h], Jean-François Deleuze[i,j], Samy Hadjadj[k,l,m], Frits R. Rosendaal[c], Pierre-Emmanuel Morange[n,o], David-Alexandre Trégouët[a,b,*]

[a] Sorbonne Universités, UPMC Univ. Paris 06, Institut National pour la Santé et la Recherche Médicale (INSERM), Unité Mixte de Recherche en Santé (UMR_S) 1166, Team Genomics & Pathophysiology of Cardiovascular Diseases, Paris, France
[b] ICAN Institute for Cardiometabolism and Nutrition, Paris, France
[c] Department of Clinical Epidemiology, Leiden University Medical Center, Leiden, The Netherlands
[d] Assistance Publique Hôpitaux de Paris, Hôpital Bichat, DHU FIRE, Département de Diabétologie, Endocrinologie et Nutrition, Paris, France.
[e] Université Paris Diderot, Sorbonne Paris Cité, UFR de Médecine, Paris, France.
[f] INSERM UMR_S 1219, Bordeaux Population Health Research Center, University of Bordeaux, France
[g] Department of Neurology, Bordeaux University Hospital, Bordeaux, France.
[h] Univ. Lille, Inserm, CHU Lille, Institut Pasteur de Lille, U1167 - RID-AGE - Risk Factors and Molecular Determinants of Aging-related Diseases, F-59000 Lille, France
[i] Centre National Génotypage, Institut de Génomique, CEA, 91057 Evry, France
[j] CEPH, Fondation Jean Dausset, Paris, France
[k] Université de Poitiers, UFR de Médecine et Pharmacie, Poitiers, France
[l] INSERM, CIC 1402 & U1082, Poitiers, France
[m] CHU de Poitiers, Service d'Endocrinologie & CIC 1402, Poitiers, France
[n] Laboratory of Haematology, La Timone Hospital, Marseille, France
[o] INSERM UMR_S 1062, Nutrition Obesity and Risk of Thrombosis, Aix-Marseille University, Marseille, France

ARTICLE INFO

Keywords:
Genetics
Estimated glomerular filtration rate
Renal function
Venous thrombosis
Genetic risk score
Association study

ABSTRACT

Objective: The association between impaired kidney function and venous thrombosis has been previously reported but supportive data are still sparse. We here wish to strengthen this association by investigating, by use of a genetic risk score approach, whether single nucleotide polymorphisms (SNPs) known to decrease the estimated glomerular filtration rate (eGFR), a surrogate marker for renal dysfunction, are associated with increased risk of venous thrombosis.

Approach and results: Fifty-one polymorphisms selected from the literature to robustly associate with eGFR were first tested for association with venous thrombosis in a French case-control collection of 1953 patients and 2338 healthy individuals. This led to the identification of a genetic risk score based on 9 polymorphisms that strongly associated with increased risk (odds ratio (OR) = 1.09 [1.06–1.15], $p = 1.44 \cdot 10^{-7}$). This genetic score association replicated (OR = 1.18 [1.11–1.26], $p = 8.86 \cdot 10^{-8}$) in an independent sample of 1289 patients and 1049 healthy controls part of the Dutch MEGA study. We then categorized the genetic score distribution observed in the combined samples into quintiles. Compared with the lowest quintile, the OR for increased risk of disease associated with the second, third, fourth and fifth quintiles were 1.13 [0.94–1.16], 1.47 [1.22–1.77], 1.52 [1.26–1.82] and 1.70 [1.41–2.05], respectively.

Conclusions: Using a genetic risk score analysis, our study provides new elements supporting the association between impaired renal function and the risk of venous thrombosis.

1. Introduction

Venous thrombosis (VT), that encompasses both deep vein thrombosis and pulmonary embolism, is the second most common cardiovascular disease and the third in terms of mortality. Its incidence is estimated to be ~200 per 100,000 person-years in Europe [1] and its

Abbreviations: CKD, chronic kidney disease; eGFR, estimated glomerular filtration rate; GRS, genetic risk score; GWAS, genome wide association study; SNP, single nucleotide polymorphisms; VT, venous thrombosis

* Corresponding author at: INSERM UMR_S 1166, 91 Boulevard de l'Hopital, 75013 Paris, France.
E-mail address: david.tregouet@upmc.fr (D.-A. Trégouët).

http://dx.doi.org/10.1016/j.thromres.2017.08.015
Received 29 May 2017; Received in revised form 14 August 2017; Accepted 22 August 2017
Available online 25 August 2017
0049-3848/ © 2017 Elsevier Ltd. All rights reserved.

recurrence rate within 10 years is close to 30%. The disease also represents a significant spending in health care [2], it is for example associated with a total annual cost ranging from €1.5 to 13.2€ billion for the EU [3]. The nature of the complex pathophysiological mechanisms that lead to VT is not fully characterized. Disturbances of the coagulation and fibrinolysis cascade leading to hypercoagulable states are clearly the most important contributor to VT etiology. Obesity-related mechanisms, inflammation [4], platelets [5] and complement cascade [6] are additional biological systems that contribute to the development of the disease. However, despite these results of intensive research efforts, there are so far limited prognosis tools that help clinicians to predict which individuals are at risk for a VT event. More investigations are thus needed to better disentangle the etiological architecture underlying VT.

A meta-analysis of 5 community-based cohorts has shown that decreased estimated glomerular filtration rate (eGFR) was associated with increased VT risk [7]. It was further hypothesized that this association was partially mediated by elevated levels of coagulation Factor VIII and von Willebrand factor [8]. Inflammation and coagulation biomarkers are associated with decreasing kidney function in ambulatory adults without established cardiovascular disease or chronic kidney disease. Investigations in the ARIC study [9] demonstrated that increased levels of factor VIIIc, fibrinogen, and von Willebrand Factor were significantly associated with increased risk of incident chronic kidney disease, of which decreased eGFR is a surrogate marker, in European Americans. The pathophysiology underlying the association of hemostatic factors to kidney function decline is unclear, although triggers of hemostatic activation, including vascular injury, endothelial dysfunction, and inflammation, have been proposed as potential mechanisms.

To provide additional support to this hypothesis, we here propose the first genetic risk score approach addressing where genetic polymorphisms that have been robustly associated with eGFR, a marker for renal function, associate with VT risk.

Main outlines of the workflow are summarized in Fig. 1. In a first step, we sought the literature for single nucleotide polymorphisms (SNPs) robustly associated with eGFR. These SNPs were then tested for association with VT risk in 1953 VT patients and 2338 healthy individuals from a French case control study (MARTHA/EOVT) [10]. We derived a genetic risk score (GRS) based on the VT-associated SNPs and tested it for replication in an independent sample of 1289 VT patients and 1049 controls from a Dutch case control study (MEGA study) [11]. We finally confirmed the association of the derived GRS with eGFR in a third independent study.

2. Material & method

2.1. Studied populations

Informed consent was obtained from all participants in accordance with the Declaration of Helsinki, and the study met all institutional ethics requirements.

2.1.1. Discovery VT case-control samples

The discovery cohort was composed of 1953 VT patients and 2338 healthy individuals from two French VTE case-control studies, MARTHA and EOVT that have been extensively described before [10,12]. Patients were individuals with documented personal VTE history and lacking strong genetic risk factors (antithrombin, protein C or protein S deficiencies, FV Leiden homozygosity, FII G20210A homozygosity). Controls were apparently healthy individuals free of any chronic conditions and of a personal VT history.

2.1.2. Replication VT case-control samples

The replication stage was based on the same data from the Multiple Environmental and Genetic Assessment of risk factor for venous thrombosis (MEGA) study [11] as those that were used in a recent meta-analysis of Genome-Wide Association Study (GWAS) data for VT [12].1289 Dutch VT patients with no prior event of VT and no cancer were compared with 1049 controls genetically matched for geographical ancestry [[12].].

2.1.3. Additional validation cohorts

The Genesis/Genediab study group was composed of 1370 individuals with Type 1 diabetes from the French population [13].

Clinical characteristics of this population are provided in Supplementary Table A.

2.2. Genotyping

All individuals used in the work have previously been typed for genome-wide genotype SNPs using dedicated DNA microarrays and imputed with 1000 Genomes reference. Genotyping, quality controls and imputation analyses of the genotype data have been previously described in [12] for the MARTHA, EOVT and MEGA studies and in [14] for Genesis/Genediab.

2.3. SNP selection

Sixty-nine SNPs were identified to robustly associate with eGFR in different GWAS studies [15–18]. These were selected as they have demonstrated genome-wide significance ($p < 5 \cdot 10^{-8}$) association with eGFR and replicated in independent studies. From this list of candidate SNPs, we discarded SNPs from any pair of SNPs in linkage disequilibrium and excluded SNPs with bad imputation quality ($r^2 < 0.3$) in both French GWAS cohorts. This resulted in 51 candidate SNPs (Supplementary Table B).

2.4. Statistical methods

2.4.1. Derivation of the genetic risk score (GRS)

Each of the 51 selected candidate SNPs was tested for association with VT in the French case-control samples using logistic regression analysis where the expected number of alleles (often referred to as imputed dose) at each SNP was used as a covariate in a model adjusting for age, sex and principal components derived from genotype data. Of note all SNPs had imputation quality $r^2 > 0.30$. From these analyses, we selected only SNPs for which the risk-allele was the one that was reported in the literature to associate with eGFR decrease, as compatible with the hypothesized relationship between renal dysfunction and VT risk. Only SNPs (Supplementary Table B) satisfying this condition were kept for the subsequent GRS analysis. In order to identify the most parsimonious and information SNP combination with respect to VT, an Akaike Information Criterion (AIC) [19] strategy was adopted using a backward logistic regression procedure, while adjusting for sex, principal components and French sub-study group. SNPs selected by this AIC procedure were then entered into a GRS defined, for each individual, as the sum of the imputed risk-allele dose at each selected SNP. We confirmed that selected SNPs were also well imputed in the replication MEGA study.

2.4.2. Association of the GRS with VT risk

Association of the derived GRS with VT was then tested using logistic regression analysis using the same covariates as for the AIC strategy, both in the discovery (MARTHA/EOVT) and replication (MEGA) studies. Results observed in the discovery and replication studies were then pooled using fixed effect model (as implemented in the rmeta R package).

2.4.3. Association analysis with eGFR

The association of the proposed GRS was investigated in relation to eGFR in the Genesis/Genediab study where eGFR was estimated using plasma creatinin and the CKD-EPI eq. [20] Association was tested using

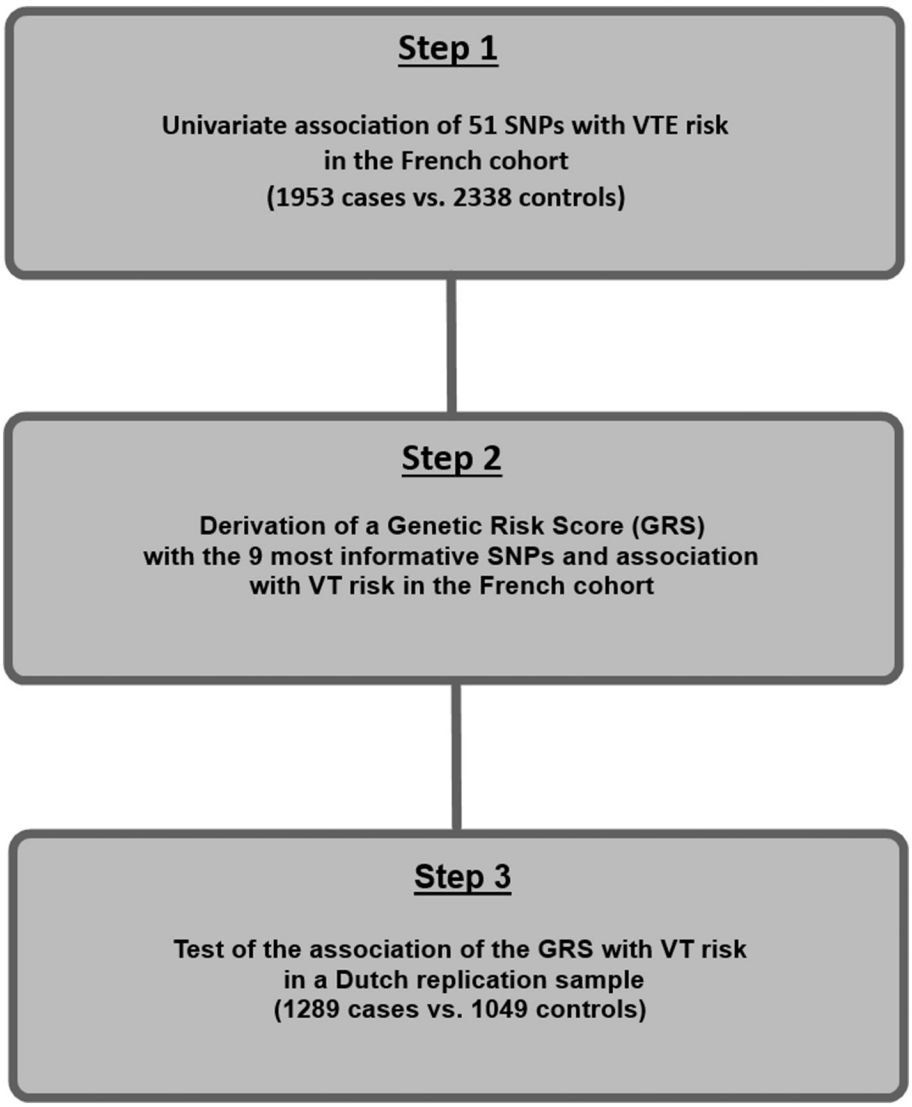


Fig. 1. Main outlines of the adopted research strategy.

linear regression analyses adjusted for age, sex and principal components derived from genotype data.

3. Results

From the 51 independent SNPs reported to robustly associate with eGFR in GWAS studies, only 27 presented with an eGFR decreasing allele that was also associated with an increased risk of VT (Supplementary Table B). The AIC based strategy applied to these 27 SNPs further selected a subset of 9 SNPs (Table 1) as the most parsimonious and informative group of SNPs to discriminate between patients and controls in the French populations. The minimal p-value of association of these 9 SNPs with VT in the French study was $p = 0.011$ for the *DDX1* rs807601 and the highest p-value was $p = 0.255$ for the *SLC34A1* rs6420094 (Table 1).

A GRS variable was then derived from these 9 SNPs for each individual as the expected number of alleles associated with decreased eGFR levels carried by this individual. The distribution of the GRS in the French populations was shown in Supplementary Fig. 1A. The GRS was significantly higher in VT patients than in controls (10.4 ± 0.04 vs 10.1 ± 0.047, $p = 1.03 \cdot 10^{-7}$). One unit increase of the GRS was associated with an increased risk for VT of 1.09 [1.06–1.15] ($p = 1.44 \cdot 10^{-7}$). In an independent collection of 1289 Dutch patients and 1049 controls, the proposed GRS exhibited a distribution similar to that observed in the French population (Supplementary Fig. 1B). One unit increase of the GRS was found associated with an increase in VT risk of 1.18 [1.11–1.26] ($p = 8.86 \cdot 10^{-8}$). The GRS association did not differ between the two studies ($p = 0.06$) and the combined odds ratio (OR) for VT associated with one unit increase of the GRS was 1.12 [1.09–1.16] ($p = 3.77 \cdot 10^{-13}$). When individual data from French and Dutch populations were combined together, the corresponding OR was 1.10 [1.06–1.14].

Fig. 2 displayed the association of the GRS with VTE according to the quintiles of its distribution. In the French samples, the association was not completely linear as individuals in the 3rd, 4th and 5th GRS quintiles were at similar risk of VT, OR = 1.39 [1.12–1.73], 1.32 [1.06–1.64] and 1.54 [1.24–1.92] respectively, compared to individuals in the lowest quintile. This association was not completely linear in the Dutch population either. When the GRS distribution observed in the Dutch population was divided according to the same GRS quintiles as those used in the French studies, individuals in the 4th and 5th highest quintiles were at increased odds of 2.20 [1.54–3.14] and 2.23 [1.56–3.19], respectively, compared with individuals in the lowest quintiles.

In the combined population, compared with lowest quintile, the OR for increased risk of VT associated with the second, third, fourth and fifth quintiles were 1.13 [0.94–1.36], 1.47 [1.22–1.77], 1.52 [1.26–1.82] and 1.70 [1.41–2.05], respectively. Based on these

Table 1
Association with VT risk of the nine SNPs defining the identified GRS.

SNP	Chromosome	Position	Locus	eGFR associated allele	MARTHA + EOVT			MEGA		
					Allele frequency[a]	Odds ratio[b]	p value	Allele frequency	Odds ratio	p value
rs807601	2	15793014	DDX1	G	0.65	1.14	0.011	0.66	1.17	0.288
rs1047891	2	211540507	CPS1	A	0.32	1.14	0.016	0.31	1.03	0.846
rs9682041	3	170091902	SKIL	T	0.86	1.15	0.051	0.86	0.96	0.821
rs6420094	5	176817636	SLC34A1	G	0.32	1.06	0.255	0.31	1.12	0.445
rs7759001	6	27341409	ZNF204P	A	0.77	1.10	0.100	0.77	1.37	0.068
rs3758086	8	23714992	STC1,ADAM28	A	0.41	1.06	0.243	0.39	1.19	0.204
rs9916302	17	37499949	FBXL20	T	0.70	1.09	0.116	0.71	1.15	0.373
rs12460876	19	33356891	SLC7A9	T	0.57	1.09	0.084	0.57	1.05	0.699
rs6088580	20	33285053	PIGU,TP53INP2	C	0.48	1.07	0.144	0.47	1.23	0.127

[a] Allele frequency in the control population.
[b] Odds ratio and p-value were adjusted for sex and principal components.

observations and the GRS quintiles distribution, these results can be translated into a 1.42 [1.25–1.61] increased risk of VTE associated with a GRS > 11 in the combined studied populations. Corresponding ORs were 1.31 [1.13–1.51] and 1.78 [1.39–2.27] in the French and Dutch samples, respectively.

Note that one of the SNPs, rs6088580, included in the definition of the GRS, is in moderate linkage disequilibrium ($r^2 = 0.15$; $D' = 1$) with the rs867186 known to associate with VT risk [21]. Nevertheless, the GRS still remained significantly associated with VT risk in the discovery (OR = 1.10 for one unit increase, $p = 6.24 \cdot 10^{-7}$) and replication case-control studies (OR = 1.17, $p = 1.20 \cdot 10^{-6}$) after adjustment for the rs867186. The resulting meta-analyzed OR was 1.12 [1.08–1.15]. In addition, when the rs6088580 was discarded from the GRS definition, the restricted GRS was still associated with VT in the French (OR = 1.10 for one unit increase, $p = 7.08 \cdot 10^{-7}$) and the Dutch (OR = 1.18, $p = 5.48 \cdot 10^{-7}$) populations. The meta-analyzed OR for VT associated with such restricted GRS was 1.13 [1.09–1.16].

SNPs defining the GRS had been previously reported to individually associate with eGFR but their joint effect have never been assessed. Therefore, we investigated the impact on eGFR levels of the proposed GRS in a collection of 1325 patients with Type 1 diabetes. In this study, the distribution of the GRS was similar to those observed in the discovery and replication VT studies (Supplementary Fig. 1C). In this study, only one of the 9 assessed SNPs, the *STC1* rs3758086, showed nominal association with eGFR (p = 0.035), all other SNPs having association p-value > 0.05 (Supplementary Table C). Nevertheless, one unit increase of the GRS was associated with a decrease of −1.19 [−1.92 to −0.39] ($p = 3.54 \cdot 10^{-3}$) in eGFR levels (Supplementary Table C).

4. Discussion

Employing a two-phase discovery-validation strategy on individual genetic data of two independent studies, we identified a GRS derived from 9 SNPs that, collectively, strongly associated with the risk of VT. These 9 SNPs combined altogether define a GRS that associates with the risk of the disease, individuals with an elevated GRS > 11 being at 1.42 increased risk of VT, whereas individually these SNPs showed very modest associations with the disease. These SNPs map to loci (DDX1, CPS1, SKIL, SLC34A1, ZNF204P, STC1, FBXL20, SLC7A9, TP53INP2) that have not been clearly associated with VT in recent large scale association studies [12,22] and whose associated biology is poorly documented with respect to VT. Nevertheless, it could be mentioned that genetic variations at the *SLC34A1* locus have been found associated with Factor XII antigen levels [23]. In addition, the *CPS1* and *STC1* genes both play role in the phosphate mechanisms of action and are involved in vascular angiogenesis [24–26]. These observations could be of particular relevance and open novel research avenues in light of recent works emphasizing the emerging role of the polyphosphate / factor XII pathway in thrombosis [27]. Finally, rare mutations

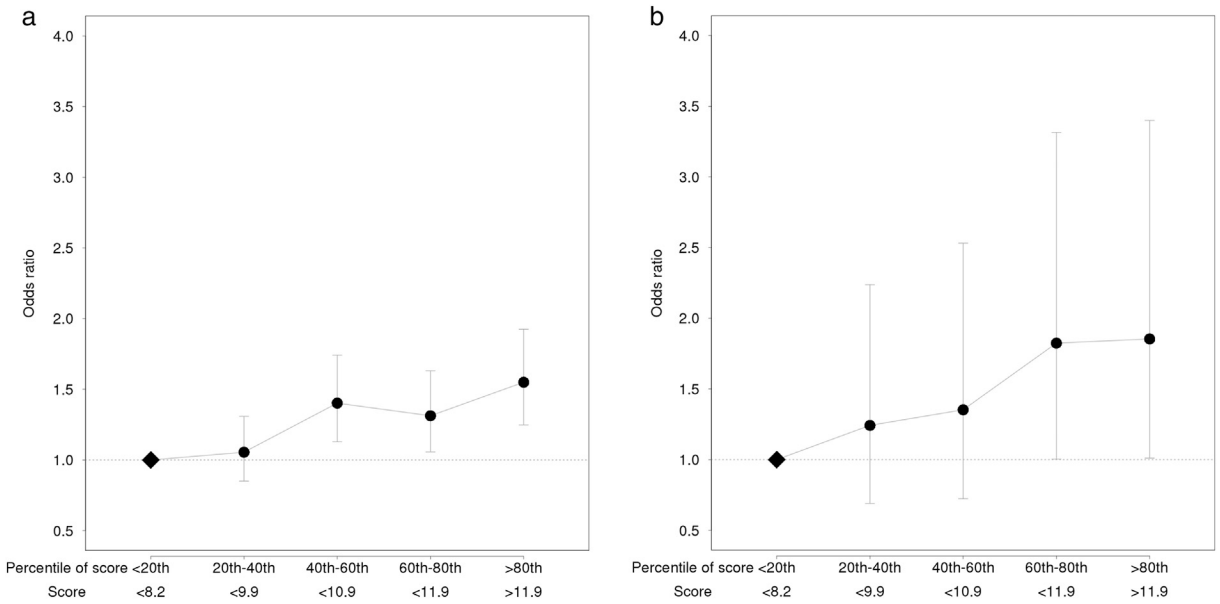


Fig. 2. Association between quintiles of the eGFR based GRS and the risk of venous thrombosis in the French (left) and Dutch (right) populations.

in SLC7A9 are known to cause cystinuria, which could link this gene to VT via some homocysteine associated mechanisms [28].

The SNPs defining this GRS had been individually associated with eGFR in previous large meta-analysis of GWAS studies and we here showed that, while these SNPs were mildly associated with eGFR in a moderate sample size study, the derived GRS was much more significantly associated with eGFR. However, the GRS impact on eGFR variability was rather modest, R^2 = 0.67%, a value that must be compared to the 3.22% of eGFR variability explained by the 53 SNPs found to significantly associate with eGFR in the recent GWAS studies [15]. This observation indicates that the influence of the GRS on VTE risk could not be solely mediated by eGFR that is an imperfect surrogate marker of impaired renal function.

The use of genetic score approach to evidence a causative association between impaired renal function and VT is appealing as the genetic risk score is fixed at birth for each individual and should thus be independent from confounders normally affecting eGFR such as age or body mass index. We did not observe any association between the proposed GRS and BMI, age or sex (data not shown).

However, several limitations need to be acknowledged. The main one relates to the fact that we do not have eGFR measurements in the studied case-control samples preventing us from conducting mediation analysis to assess whether the impact of the GRS on VT risk vanishes after adjusting for eGFR. Similarly, proteinuria or cardiac biomarkers were not available in our samples to assess how the proposed GRS associates with CKD. While we obtained strong evidence that the GRS is associated with VT risk and eGFR, we have so far no evidence of inflammatory, vascular or hemostatic markers that could be influenced by this GRS. We investigated its association with several biological traits available in the MARTHA participants (including D-dimers, endogenous thrombin generation, plasma antigen or activity levels of fibrinogen, coagulation factors II, VIII, and XI, von Willebrand factor, antithrombin, protein C, protein S, activated partial thromboplastin time, haemoglobin, and white blood cell and platelet counts) but without any robust findings (data not shown). Unfortunately, factor XII antigen had not been measured and, according to the point discussed above, its association with the proposed GRS deserves further investigations. It is important to stress that the GRS was derived from SNPs that have been found robustly associated with eGFR in several studies but without any evidence they could be functional. They likely serve as markers for the underlying functional variants that need to be identified. One could speculate that once these functional variants are identified, the GRS they define would likely be a more accurate marker of the risk of VT. All these points deserve further investigations as well as further validation of the proposed GRS in large prospective cohorts for incidence, but also for recurrence of the disease.

5. Conclusion

This work provided additional elements supporting the association of impaired renal function with increased risk of venous thrombosis. However, the mechanisms underlying this association risk are far to be understood and require further deep investigations.

Supplementary data to this article can be found online at http://dx.doi.org/10.1016/j.thromres.2017.08.015.

Acknowledgments

We are grateful to Elise GAND (Pole DUNE – CHU Poitiers) for data-management, Ahmed Amine KASMI (INSERM CIC1402, Poitiers) for clinical management and CRB BB 0033-00068 for biological handling and shipment of DNA to genomic platform.

Sources of funding

R.C. was supported by a PhD grant from the Region Ile de France (CORDDIM 2014 program). MARTHA genetics research program was supported by the GENMED Laboratory of Excellence on Medical Genomics (ANR-10-LABX-0013), the French Clinical Research Infrastructure Network on Venous Thrombo-Embolism (F-CRIN INNOVTE) and the ICAN Institute for Cardiometabolism and Nutrition (ANR-10-IAHU-05), three research programs managed by the National Research Agency (ANR) part of the French Investment for the Future initiative. The Juvenile Diabetes Research foundation (JDRF-I) funded the genotyping of the Genesis/Genediab population (Grant - Diabetic Nephropathy Initiative).

Disclosures

None.

References

[1] J.A. Heit, F.A. Spencer, R.H. White, The epidemiology of venous thromboembolism, J. Thromb. Thrombolysis 41 (2016) 3–14.
[2] S.D. Grosse, R.E. Nelson, K.A. Nyarko, L.C. Richardson, G.E. Raskob, The economic burden of incident venous thromboembolism in the United States: a review of estimated attributable healthcare costs, Thromb. Res. 137 (2016) 3–10.
[3] K. Kröger, J. Küpper-Nybelen, C. Moerchel, T. Moysidis, C. Kienitz, I. Schubert, Prevalence and economic burden of pulmonary embolism in Germany, Vasc. Med. 17 (2012) 303–309.
[4] B.A. Lwaleed, R.S. Kazmi, A.J. Cooper, Inflammation, endothelial dysfunction, and thromboembolism, Semin. Thromb. Hemost. 41 (2015) 547–548.
[5] S. Montoro-García, M. Schindewolf, S. Stanford, O.H. Larsen, T. Thiele, The role of platelets in venous thromboembolism, Semin. Thromb. Hemost. 42 (2016) 242–251.
[6] J.H. Foley, Examining coagulation-complement crosstalk: complement activation and thrombosis, Thromb. Res. 141 (2016) S50–4.
[7] B.K. Mahmoodi, R.T. Gansevoort, I.A. Næss, P.L. Lutsey, S.K. Brækkan, N.J. Veeger, E.E. Brodin, K. Meijer, Y. Sang, K. Matsushita, S.I. Hallan, J. Hammerstrøm, S.C. Cannegieter, B.C. Astor, J. Coresh, A.R. Folsom, J.B. Hansen, M. Cushman, Association of mild to moderate chronic kidney disease with venous thromboembolism: pooled analysis of five prospective general population cohorts, Circulation 126 (16) (2012) 1964–1971.
[8] G. Ocak, C.Y. Vossen, W.M. Lijfering, M. Verduijn, F.W. Dekker, F.R. Rosendaal, S.C. Cannegieter, Role of hemostatic factors on the risk of venous thrombosis in people with impaired kidney function, Circulation 129 (2014) 683–691.
[9] L.D. Bash, T.P. Erlinger, J. Coresh, J. Marsh-Manzi, A.R. Folsom, B.C. Astor, Inflammation, hemostasis, and the risk of kidney function decline in the atherosclerosis risk in communities (ARIC) study, Am. J. Kidney Dis. 53 (2009) 596–605.
[10] M. Germain, N. Saut, N. Greliche, et al., Genetics of venous thrombosis: insights from a new genome wide association study, PLoS One 6 (2011) e25581.
[11] J.W. Blom, C.J. Doggen, S. Osanto, F.R. Rosendaal, Malignancies, prothrombotic mutations, and the risk of venous thrombosis, JAMA 293 (2005) 715–722.
[12] M. Germain, D.I. Chasman, H. de Haan, et al., Meta-analysis of 65,734 individuals identifies TSPAN15 and SLC44A2 as two susceptibility loci for venous thromboembolism, Am. J. Hum. Genet. 96 (2015) 532–542.
[13] S. Hadjadj, B. Cariou, F. Fumeron, et al., Death, end-stage renal disease and renal function decline in patients with diabetic nephropathy in French cohorts of type 1 and type 2 diabetes, Diabetologia 59 (2016) 208–216.
[14] M. Germain, M.G. Pezzolesi, N. Sandholm, et al., SORBS1 gene, a new candidate for diabetic nephropathy: results from a multi-stage genome-wide association study in patients with type 1 diabetes, Diabetologia 58 (2015) 543–548.
[15] C. Pattaro, A. Teumer, M. Gorski, et al., Genetic associations at 53 loci highlight cell types and biological pathways relevant for kidney function, Nat. Commun. 7 (2016) 10023.
[16] Y. Okada, X. Sim, M.J. Go, et al., Meta-analysis identifies multiple loci associated with kidney function-related traits in east Asian populations, Nat. Genet. 44 (2012) 904–909.
[17] C. Pattaro, A. Köttgen, A. Teumer, et al., Genome-wide association and functional follow-up reveals new loci for kidney function, PLoS Genet. 8 (2012) e1002584.
[18] A. Köttgen, C. Pattaro, C.A. Böger, et al., New loci associated with kidney function and chronic kidney disease, Nat. Genet. 42 (2010) 376–384.
[19] H. Akaike, A new look at the statistical model identification, IEEE Trans. Autom. Control 19 (1974) 716–723.
[20] A.S. Levey, L.A. Stevens, C.H. Schmid, Y.L. Zhang, A.F. Castro III, H.I. Feldman, J.W. Kusek, P. Eggers, F. Van Lente, T. Greene, J. CoreshCKD-EPI (Chronic Kidney Disease Epidemiology Collaboration), A new equation to estimate glomerular filtration rate, Ann. Intern. Med. 150 (2009) 604–612.
[21] J. Dennis, C.Y. Johnson, A.S. Adediran, M. de Andrade, J.A. Heit, P.E. Morange, D.A. Trégouët, F. Gagnon, The endothelial protein C receptor (PROCR) Ser219Gly variant and risk of common thrombotic disorders: a HuGE review and meta-analysis of evidence from observational studies, Blood 119 (2012) 2392–2400.
[22] D.A. Hinds, A. Buil, D. Ziemek, et al., Genome-wide association analysis of self-reported events in 6135 individuals and 252 827 controls identifies 8 loci associated with thrombosis, Hum. Mol. Genet. 25 (2016) 1867–1874.

[23] F.M. Williams, A.M. Carter, Hysi, et al., Ischemic stroke is associated with the ABO locus: the EuroCLOT study, Ann. Neurol. 73 (2013) 16–31.
[24] C. Zlot, G. Ingle, J. Hongo, et al., Stanniocalcin 1 is an autocrine modulator of endothelial angiogenic responses to hepatocyte growth factor, J. Biol. Chem. 278 (2003) 47654–47659.
[25] Y. Yoshiko, J.E. Aubin, Stanniocalcin 1 as a pleiotropic factor in mammals, Peptides 25 (2004) 1663–1669.
[26] M.L. Summar, J.V. Gainer, M. Pretorius, et al., Relationship between carbamoyl-phosphate synthetase genotype and systemic vascular function, Hypertension 43 (2004) 186–191.
[27] K.F. Nickel, L. Labberton, A.T. Long, et al., The polyphosphate/factor XII pathway in cancer-associated thrombosis: novel perspectives for safe anticoagulation in patients with malignancies, Thromb. Res. 141 (Suppl. 2) (2016) S4–7.
[28] R. Ragone, Homocystine solubility and vascular disease, FASEB J. 16 (2002) 401–404.

	Genesis/Genediab
N	1370
Age (mean±se)	41±0.3
Sex (% women)	47
BMI (mean±se)	24.1±0.1*
eGFR (mean±se) [25th-75th percentiles]	86.9±0.8 [71.1-108.4]**
GRS (mean±se)	10.24±0.05
Diabetes (%)	100
CAD (%)	6

Supplementary table A Clinical characteristics of the Genesis/Genediab population.
* measured on a subsample of 1305 individuals
**measured on a subsample of 1325 individuals

		MARTHA			EOVT					
		Allele Frequency			Allele Frequency					
eGFR increasing allele	Frequency in 1000G	Controls	Cases	Imputation quality (r^2)	Controls	Cases	Imputation quality (r^2)	Combined Odds ratio	Selected SNP	Reference (PMID)
T	0.30	0.30	0.27	1.00	0.29	0.32	1.00	0.99	NO	22479191, 26831199
G	0.30	0.27	0.27	0.99	0.29	0.29	0.99	1.02	YES	26831199
T	0.79	0.80	0.81	0.97	0.80	0.81	0.96	0.98	NO	20383146, 26831199
G	0.88	0.89	0.89	1.00	0.88	0.90	1.00	0.99	NO	26831199
A	0.43	0.46	0.47	0.90	0.47	0.47	0.70	1.01	YES	26831199
G	0.66	0.65	0.67	0.93	0.65	0.68	0.94	1.14	YES	22479191, 26831199
C	0.58	0.53	0.52	1.00	0.56	0.57	0.94	1.07	YES	20383146, 26831199
A	0.76	0.75	0.77	1.00	0.76	0.76	1.00	1.06	YES	20383146, 26831199
A	0.53	0.53	0.53	0.98	0.54	0.55	0.98	1.04	YES	26831199
A	0.32	0.32	0.33	0.77	0.33	0.36	0.70	1.14	YES	20383146, 26831199
A	0.58	0.59	0.58	1.00	0.58	0.58	1.00	0.97	NO	26831199
G	0.85	0.83	0.83	1.00	0.84	0.84	0.91	1.08	YES	26831199
C	0.73	0.73	0.75	0.98	0.74	0.76	0.97	1.03	YES	20383146, 26831199
T	0.87	0.86	0.86	1.00	0.86	0.88	1.00	1.15	YES	26831199
G	0.13	0.13	0.13	0.97	0.14	0.15	0.98	1.06	YES	26831199
A	0.43	0.43	0.41	0.98	0.43	0.42	0.93	0.95	NO	20383146, 26831199
A	0.44	0.46	0.46	0.97	0.45	0.48	0.95	0.99	NO	20383146, 26831199
G	0.34	0.31	0.34	0.98	0.33	0.33	0.98	1.06	YES	20383146, 26831199
A	0.76	0.77	0.80	0.98	0.77	0.80	0.98	1.11	YES	26831199
T	0.71	0.72	0.70	0.89	0.70	0.69	0.95	0.94	NO	20383146, 26831199
C	0.90	0.92	0.91	1.00	0.91	0.90	1.00	0.99	NO	20383146, 26831199
T	0.67	0.67	0.66	0.92	0.66	0.67	0.87	0.94	NO	26831199
G	0.27	0.28	0.27	0.96	0.28	0.26	0.89	0.97	NO	20383146, 26831199
A	0.25	0.29	0.28	1.00	0.29	0.29	0.96	1.01	YES	26831199
T	0.74	0.75	0.72	0.95	0.73	0.73	0.91	0.94	NO	26831199
A	0.42	0.41	0.41	0.97	0.42	0.41	0.97	1.06	YES	20383146, 26831199
A	0.40	0.41	0.40	1.00	0.41	0.41	1.00	1.03	YES	20383146, 26831199
T	0.08	0.08	0.07	1.00	0.07	0.09	1.00	1.01	YES	20383146, 26831199
C	0.81	0.83	0.82	0.90	0.82	0.81	0.86	0.98	NO	26831199
G	0.18	0.19	0.18	0.99	0.18	0.17	0.98	0.94	NO	26831199
T	0.54	0.51	0.51	1.00	0.54	0.52	0.99	0.99	NO	22479191, 26831199
G	0.36	0.34	0.35	0.98	0.36	0.31	0.96	0.94	NO	20383146, 26831199
T	0.65	0.66	0.64	1.00	0.65	0.64	1.00	0.99	NO	20383146, 26831199
A	0.10	0.10	0.11	1.00	0.11	0.10	0.99	0.89	NO	26831199
T	0.81	0.82	0.82	1.00	0.81	0.82	1.00	1.03	YES	26831199
C	0.78	0.81	0.81	0.97	0.80	0.80	0.97	0.91	NO	26831199
G	0.60	0.60	0.58	0.99	0.59	0.59	0.99	0.94	NO	20383146, 26831199
G	0.43	0.44	0.46	0.97	0.45	0.46	0.96	1.03	YES	22479191, 26831199
G	0.38	0.38	0.40	0.96	0.39	0.42	0.91	1.02	YES	20383146, 26831199
A	0.78	0.82	0.82	0.99	0.81	0.80	0.98	0.90	NO	20383146, 26831199
A	0.35	0.33	0.33	0.65	0.35	0.39	0.98	0.97	NO	20383146, 26831199
T	0.81	0.81	0.79	0.98	0.80	0.80	0.97	0.96	NO	20383146, 26831199
G	0.47	0.53	0.55	0.86	0.52	0.55	0.83	1.01	YES	26831199
C	0.41	0.41	0.42	0.60	0.40	0.44	0.97	1.04	YES	22479191, 26831199
T	0.74	0.72	0.73	0.86	0.69	0.72	0.66	1.09	YES	22479191, 26831199
T	0.19	0.19	0.22	0.68	0.21	0.24	0.64	1.06	YES	20383146, 26831199
A	0.56	0.55	0.58	1.00	0.59	0.57	0.75	1.05	YES	26831199
T	0.60	0.56	0.58	0.99	0.58	0.55	0.97	1.09	YES	20383146, 26831199
T	0.18	0.19	0.18	0.70	0.18	0.19	0.57	0.95	NO	26831199
C	0.47	0.47	0.51	0.95	0.49	0.50	0.96	1.07	YES	26831199
T	0.79	0.81	0.80	0.88	0.82	0.81	0.80	0.96	NO	26831199

Supplementary Table B Association analysis of 51 eGFR associated SNPs with the risk of venous thrombosis in two French case-control populations, MARTHA and EOVT.

		Genesis/Genediab		
SNP	A1	Beta	SE	P
rs807601	G	-1.330	1.186	0.262
rs1047891	A	-1.015	1.302	0.435
rs9682041	T	-2.340	1.670	0.161
rs6420094	G	-0.569	1.178	0.629
rs7759001	A	-1.475	1.352	0.275
rs3758086	A	-2.345	1.109	0.035
rs9916302	T	-0.725	1.216	0.551
rs12460876	T	-0.575	1.085	0.596
rs6088580	C	-0.819	1.109	0.460
GRS	NA	-1.187	0.406	$3.545\ 10^{-3}$

Supplementary Table C Association of the eGFR based GRS and its individual SNP component with eGFR levels in the Genesis/Genediab study.

Novel risk genes identified in a genome-wide association study for coronary artery disease in patients with type 1 diabetes

Charmet *et al. Cardiovasc Diabetol (2018) 17:61*
https://doi.org/10.1186/s12933-018-0705-0

Cardiovascular Diabetology

ORIGINAL INVESTIGATION

Open Access

Novel risk genes identified in a genome-wide association study for coronary artery disease in patients with type 1 diabetes

Romain Charmet[1,2], Seamus Duffy[3], Sareh Keshavarzi[4], Beata Gyorgy[1,2], Michel Marre[5,6], Peter Rossing[7,8], Amy Jayne McKnight[3], Alexander P. Maxwell[3], Tarun veer Singh Ahluwalia[7], Andrew D. Paterson[4,9], David-Alexandre Trégouët[1,2*†] and Samy Hadjadj[10,11,12*†]

Abstract

Background: Patients with type 1 diabetes are more at risk of coronary artery disease than the general population. Although evidence points to a genetic risk there have been no study investigating genetic risk factors of coronary artery disease specific to individuals with type 1 diabetes. To identify low frequency and common genetic variations associated with coronary artery disease in populations of individuals with type 1 diabetes.

Methods: A two-stage genome wide association study was conducted. The discovery phase involved the meta-analysis of three genome-wide association cohorts totaling 434 patients with type 1 diabetes and coronary artery disease (cases) and 3123 T1D individuals with no evidence of coronary artery disease (controls). Replication of the top association signals ($p < 10^{-5}$) was performed in five additional independent cohorts totaling 585 cases and 2612 controls.

Results: One locus (rs115829748, located upstream of the MAP1B gene) reached the statistical threshold of 5×10^{-8} for genome-wide significance but did not replicate. Nevertheless, three single nucleotide polymorphisms provided suggestive evidence for association with coronary artery disease in the combined studies: *CDK18* rs138760780 (OR = 2.60 95% confidence interval [1.75–3.85], $p = 2.02 \times 10^{-6}$), *FAM189A2* rs12344245 (OR = 1.85 [1.41–2.43], $p = 8.52 \times 10^{-6}$) and *PKD1* rs116092985 (OR = 1.53 [1.27–1.85], $p = 1.01 \times 10^{-5}$). In addition, our analyses suggested that genetic variations at the *ANKS1A*, *COL4A2* and *APOE* loci previously found associated with coronary artery disease in the general population could have stronger effects in patients with type 1 diabetes.

Conclusions: This study suggests three novel candidate genes for coronary artery disease in the subgroup of patients affected with type 1 diabetes. The detected associations deserve to be definitively validated in additional epidemiological studies.

Keywords: Type 1 diabetes, Coronary artery disease, Diabetic nephropathy, Genome-wide association study, Case control study, Epidemiology, Genetic association studies, Meta-analysis

*Correspondence: david.tregouet@upmc.fr; samy.hadjadj@gmail.com
†David-Alexandre Trégouët and Samy Hadjadj contributed equally to the work
[1] Institut National pour la Santé et la Recherche Médicale (INSERM), Unité Mixte de Recherche en Santé (UMR_S) 1166, Team Genomics & Pathophysiology of Cardiovascular Diseases, Sorbonne Universités, UPMC Univ. Paris 06, Paris, France
[12] Service d'Endocrinologie-Diabétologie and Centre d'Investigation Clinique, CHU de Poitiers, BP 577, 86021 Poitiers Cedex, France
Full list of author information is available at the end of the article

© The Author(s) 2018. This article is distributed under the terms of the Creative Commons Attribution 4.0 International License (http://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The Creative Commons Public Domain Dedication waiver (http://creativecommons.org/publicdomain/zero/1.0/) applies to the data made available in this article, unless otherwise stated.

Background

Type 1 diabetes (T1D) is a chronic disease characterized by an increase in blood glucose due to a lack of insulin production. Diabetes is a major health concern globally with a prevalence ranging between 4 and 7.8% in industrialized countries [1, 2]; among persons with diabetes, it is estimated that 5–10% are affected with T1D [3]. Recent large-scale epidemiological studies suggest that T1D is associated with a reduced lifespan of approximately 12 years [4] due to at least two-fold increased risk for death due to cardiovascular (CV) events.

Of note, in a nation-based registry of type 1 diabetes, coronary heart disease was 4 times more prevalent in patients with type 1 diabetes compared to sex- and age-matched controls [5].

However, studies exploring cardiovascular disease (CVD) risk variants in T1D patients are lacking. Traditional risk factors for CVD including age, circulating lipid levels, and smoking are well established in the general population [6] and in patients with type 1 diabetes [7]. Coronary artery disease (CAD) could be strongly influenced by genetic factors [8]. However, to the best of our knowledge, the relationship between genetic factors and CAD in type 1 diabetes was largely examined with a candidate gene approach [9–12]. Another important question is whether genetic markers of CAD established in the general population, known to roughly explain 10% of the heritability [13, 14], also play a role in individuals with T1D.

In an attempt to unravel the genetic determinants of CAD among T1D patients, we examined the association of genome-wide genotype array data with CAD in multiple T1D cohorts of European descent (The British Isles, Denmark and France).

Materials and methods

General workflow

The present work reports the results of a two-stage research strategy for common genetic variations associated with CAD risk in T1D patients. The first (discovery) stage was based on the meta-analysis of three GWAS cohorts totaling 434 T1D patients with CAD (cases) and 3123 T1D patients with no evidence of CAD (controls). The second stage consisted of a replication of the top discovery signals with association $p < 10^{-5}$ in five additional T1D studies totaling 585 cases and 2612 controls.

Participating cohorts for the discovery and replication stages

All participants were patients with T1D diagnosed using ADA criteria [15].

Controls were patients with T1D without history of CAD while cases were patients with T1D and a personal history of myocardial infarction or coronary artery revascularization (coronary artery angioplasty or by-pass grafting).

The discovery phase was composed of European-ancestry adults from (1) France (pooled cohorts of T1D from Corbeil Essonnes, Poitiers, Nantes, Paris, Toulouse [16] and two large scale multicenter cohorts i.e. GENESIS/GENEDIAB [16]), (2) Denmark (After-EU cohort [17]) and (3) British Isles (UK-ROI study [18]). The replication phase included five independent T1D cohorts recruited in North America (Additional file 1: Table S1).

All participating studies were approved by their respective institutional review board/ethics committee and an informed consent was obtained from all participating individuals.

Genotype determinations and imputation

For each participating study, DNA samples were genotyped with high-density SNP arrays and further imputed for SNPs available in the 1000 Genomes reference dataset. Summary descriptions of genotyping technologies, quality control procedures, and used imputation methods (MACH/Impute2) are shown in Additional file 1: Table S1.

Discovery phase: meta-analysis of discovery GWAS

Association analyses of imputed SNPs with CAD risk were performed separately in each study. Analyses were performed using either of the MACH [19], Quicktest (http://toby.freeshell.org/software/quicktest.shtml), or Plink [20] analyses tools implementing a logistic regression model where the allele dosage representing the expected number of a given reference allele at the imputed SNP was used as covariate to estimate SNP effect. Analyses were adjusted for sex, age, DN status and potential population sub-structure as defined by SNPs derived principal components.

Only SNPs with acceptable imputation quality ($r^2 > 0.3$) in the three discovery cohorts and with estimated minor allele frequency (MAF) $\geq 1\%$ were kept for meta-analysis. This was performed by use of a fixed-effects model based on the inverse-variance weighting method as implemented in the METAL software [21]. The statistical threshold ($p < 5 \times 10^{-8}$) was used for declaring genome-wide statistical significance while controlling for the number of independent tests across the genome. The Cochran's Q statistic was used to assess heterogeneity of the SNP associations across studies whose magnitude was expressed by the I^2 index [22]. Power calculations were performed using the CaTS power calculator (http://csg.sph.umich.edu/abecasis/cats/) [23].

Replication phase

Similar logistic regression models as those employed in the discovery were used for assessing the association of tested SNPs with CAD risk (Additional file 1: Table S1). Results obtained in the independent replication cohorts were then meta-analyzed using the same methodology as in the discovery step. The Bonferroni threshold corresponding to 0.05 divided by the number of tested SNPs was used to declare statistical replication. Unilateral hypothesis testing was adopted at the replication stage. For SNPs that replicated, a meta-analysis of the combined discovery and replications cohorts was performed to produce a more robust estimate of the effect size.

Results

A total of 6,728,637 imputed SNPs were tested for association with CAD in 3557 T1D patients made of 434 with CAD cases and 3123 controls in the discovery dataset. The meta-analysis results of the discovery GWAS have been summarized in the Manhattan and QQ plots shown in Additional file 2: Figure S1, Additional file 3: Figure S2, respectively.

One locus at chromosome 5q13.2 reached genome-wide significance ($p < 5 \times 10^{-8}$) with the lead SNP, rs115829748, located upstream of the *MAP1B* gene. The T allele of this low frequency SNP (MAF ~ 0.04), was associated with an Odds Ratio (OR) of 3.16 [95% confidence interval (CI) 2.18–4.59] ($p = 1.36 \times 10^{-9}$). No other SNP demonstrated suggestive association with CAD at this locus (Additional file 4: Figure S3).

At the $p < \sim 1.0 \times 10^{-5}$ threshold, 20 additional loci demonstrated evidence for suggestive association with CAD with little heterogeneity across cohorts (Table 1). Imputation metrics of the top SNPs are provided in the Additional file 5: Table S2. Therefore, we sought to replicate the top 21 signals in five independent T1D cohorts totaling 585 CAD cases and 2612 controls. Replication was feasible for 17 SNPs while four SNPs (rs34319244, rs373009901, rs143723948, rs571622299) were not properly imputed in the replication stage (Table 1). While none of the 17 SNPs reached the pre-specified Bonferroni threshold of 3.0×10^{-3} for positive statistical replication, three were however nominally ($p < 0.05$) associated with CAD in the replication stage, with genetic effects consistent between the discovery and replication studies (Table 1). Of note, no trend for association was observed ($p = 0.279$) with the *MAP1B* rs115829748 that came out first in the discovery GWAS and that showed similar allele frequencies in the discovery and replication studies.

The strongest association was observed at the *CDK18* locus where, in the replication stage, the rs138760780-T allele, with frequency 0.02, was associated with an increased odds ratio (OR) for disease of 1.88 [1.07–3.31] ($p = 0.014$). This value has to compare with 3.48 [2.00–6.04] observed in the discovery cohorts. In the combined discovery and replication cohorts, the meta-analyzed OR associated with the rs138760780 T allele was 2.60 [1.75–3.85] ($p = 2.02 \times 10^{-6}$) with no heterogeneity across the discovery and replication stage ($p = 0.545$).

The second suggestive association holds at the *FAM189A2* locus. The rs12344245-G allele found associated with an increased OR of 2.52 [1.68–3.81] in the discovery cohorts also demonstrated a trend for association with the disease in the replication stage, OR = 1.45 [1.01–2.08] ($p = 0.022$). Combining the discovery and replication study led to a meta-analyzed OR for disease of 1.85 [1.41–2.43] ($p = 8.52 \times 10^{-6}$) with no significant evidence for heterogeneity across stage ($p = 0.426$).

The third suggestive association was observed for the *PKD1* locus. The rs116092985 minor G associated with a 1.85 [1.408–2.44] and a 1.29 [1.00–1.67] increased risk of CAD in the discovery and replication cohorts, respectively. Altogether, the combined statistical evidence for association of the rs116092985-G allele reached $p = 1.01 \times 10^{-5}$ (OR = 1.53 [1.27–1.85]) ($p = 0.220$ for heterogeneity across studies).

Candidate CAD SNPs

About 90 loci have been found, through GWAS studies, to harbor common susceptibility alleles associated with CAD in the general population. We sought to investigate how these loci associate with CAD in T1D patients. Results of this investigation are summarized in Table 2. From CAD SNPs identified in previous GWAS [13, 14, 24–26], 95 were well imputed in our discovery phase and showed genetic effects with directionality in our T1D populations that were consistent with those previously reported (Table 2). Imputation quality for these 95 SNPs is given in Additional file 6: Table S3. For three SNPs *ANKS1A*_rs17609940, *COL4A2*_rs11838776 and *TOMM40*_rs2075650 (near the *APOE* locus), the statistical evidence for association with the disease was rather strong with $p < 5 \times 10^{-3}$ even though it did not achieve multiple testing correction for the number of tested SNPs ($\sim 5 \times 10^{-4} = 0.05/95$). For these three SNPs, the amplitude of the genetic association even tended to be stronger in our T1D patients than that previously reported (Table 2). As an illustration, in our discovery T1D population, the *COL4A1* rs11838776-A allele was associated with an OR of 1.33 [1.11–1.61] while the OR reported in the literature was slightly lower (OR = 1.07). Conversely, the association of the polymorphism at the non-coding ANRIL loci on 9p21, that is known to associate the most with CAD among common polymorphisms, showed a very similar association in our T1D patients (OR = 1.16 [0.993–1.362], $p = 0.03$) compared to that previously reported (OR ~ 1.21).

Table 1 Lead SNPs in discovery and replication

						Discovery						Replication					
CHR	BP	Locus	SNP	Type of variation	EA/NEA[a]	EAF[b]	OR[c]	P[d]	Direction[e]	I^{2} [f]	P_{het} [g]	EAF	OR	P[h]	Direction	I^{2}	P_{het}
1	205484373	CDK18	rs138760780	Intronic	T/C	0.018	3.48	$9.20\ 10^{-6}$	+++	0	0.943	0.020	1.88	0.014	++++−	0	0.509
2	97455276	CNNM4	rs116656846	Intronic	A/G	0.023	2.57	$9.88\ 10^{-6}$	+++	0	0.661	0.028	0.89	0.696	−+−	0	0.876
2	155225182	GALNT13	rs17206992	Intronic	G/A	0.057	2.32	$6.33\ 10^{-6}$	+++	0	0.817	0.052	0.97	0.580	++−+−	0	0.934
2	177645590	AC092162.1	rs113517532	Intergenic	AGAT/A	0.106	1.75	$1.90\ 10^{-6}$	+++	0	0.912	0.105	0.86	0.887	+−+−	0	0.658
3	13370674	NUP210	rs73018809	Intronic	T/A	0.024	3.50	$1.89\ 10^{-7}$	+++	59.1%	0.087	0.022	0.60	0.955	−+	0	0.722
3	103975418	MIR548A3	rs28641753	Intergenic	T/C	0.071	2.10	$1.59\ 10^{-7}$	+++	0	0.863	0.067	1.02	0.440	−++−	62.70%	0.030
4	6171230	JAKMIP1	rs78031527	Intronic	T/C	0.201	1.70	$5.41\ 10^{-6}$	+++	0	0.493	0.209	1.13	0.114	++−+−	0	0.622
5	10500646	ROPN1L	rs143537377	Intronic	C/A	0.096	1.88	$7.89\ 10^{-6}$	+++	1.1%	0.577	0.100	0.85	0.885	−+−+	19.20%	0.292
5	71394387	MAP1B	rs115829748	Intergenic	T/C	0.040	3.16	$1.36\ 10^{-9}$	+++	71.9%	0.028	0.036	1.13	0.283	−+++−	0	0.655
6	95557471	MANEA-AS1	rs9354144	Intergenic	A/T	0.105	1.71	$5.78\ 10^{-6}$	+++	0	0.785	0.097	1.05	0.330	−++−	23.90%	0.262
8	73842523	KCNB2	rs571622299	Intronic	A/G	0.015	4.12	$6.47\ 10^{-6}$	+++	47.2%	0.150	NA	NA	NA	NA	NA	NA
9	37034095	PAX5	rs143723948	UTR5	T/C	0.499	1.65	$6.01\ 10^{-7}$	+++	0	0.985	NA	NA	NA	NA	NA	NA
9	71955717	FAM189A2	rs12344245	Intronic	G/A	0.035	2.52	$9.23\ 10^{-6}$	+++	0	0.950	0.038	1.45	0.022	+−+++	0	0.560
10	19457387	ARL5B	rs117826205	Intronic	C/T	0.026	2.72	$6.30\ 10^{-6}$	+++	0	0.482	0.029	1.11	0.315	−++	5.70%	0.374
11	8080425	TUB	rs61879614	Intronic	C/T	0.048	2.94	$2.60\ 10^{-6}$	+++	9.4%	0.332	0.046	1.11	0.300	+−	0	0.775
16	2160973	PKD1	rs116092985	Missense (W1399R)	G/A	0.097	1.85	$1.72\ 10^{-5}$	+++	29.5%	0.242	0.096	1.29	0.026	++++−	0	0.488
17	4328164	SPNS3	rs34319244	Intergenic	C/CT	0.440	1.51	$2.59\ 10^{-6}$	+++	48.8%	0.142	NA	NA	NA	NA	NA	NA
18	45399356	SMAD2	rs113114656	Intronic	T/C	0.040	2.67	$2.30\ 10^{-6}$	+++	2.4%	0.359	0.038	1.04	0.437	−++−+	43.40%	0.132
21	21347156	NCRNA00320	rs67213764	Intergenic	G/A	0.261	1.48	$9.28\ 10^{-6}$	+++	0	0.536	0.260	1.03	0.363	++−	32.20%	0.207
21	24929109	AP000459.7	rs12482425	Intergenic	A/G	0.314	0.66	$9.33\ 10^{-6}$	−	0	0.509	0.308	1.09	0.871	+++++	0	0.931
22	25988780	ADRBK2	rs373009901	Intronic	C/G	0.019	3.94	$9.42\ 10^{-6}$	+++	0	0.599	NA	NA	NA	NA	NA	NA

[a] Estimated allele/non estimated allele

[b] Allele frequency of the estimate allele

[c] Odds ratio for disease

[d] Association p value derived from the meta-analysis of the three discovery cohorts

[e] Directionality of the effects across the contributing cohorts

[f] I^{2} statistics for heterogeneity across the contributing cohorts

[g] p value for homogeneity across the contributing cohorts

[h] One sided test p value of association

Table 2 Replication of CAD SNPs previously identified in the general population

SNP	CHR	BP	Locus	Published GWAS results			Discovery GWAS in T1D patients					
				EA[a]	EAF[b]	OR[c]	EAF[d]	OR[e]	[95% CI]	P[f]	Direction[g]	Power[h]
rs11206510	1	55496039	PCSK9	T/C	0.848	1.08	0.815	1.08	[0.883–1.324]	0.223	−++	0.150
rs9970807	1	56965664	PPAP2B	C/T	0.915	1.13	0.906	1.21	[0.904–1.625]	0.098	+++	0.380
rs7528419	1	109817192	SORT1	A/G	0.786	1.12	0.784	1.11	[0.918–1.348]	0.138	+++	0.260
rs11810571	1	151762308	TDRKH	G/C	0.849	1.06	0.854	1.05	[0.836–1.316]	0.339	−+−	0.080
rs1892094	1	169094459	ATP1B1	T/C	0.500	0.96	0.529	1.08	[0.919–1.266]	0.822	−++	0.200
rs6700559	1	200646073	DDX59	T/C	0.470	0.96	0.460	1.15	[0.982–1.338]	0.958	+−+	0.520
rs2820315	1	201872264	LMOD1	T/C	0.300	1.05	0.298	0.97	[0.820–0.156]	0.618	−++	0.060
rs17464857	1	222762709	MIA3	T/G	0.861	1.06	0.842	1.12	[0.901–1.405]	0.148	+++	0.250
rs16986953	2	19942473	AK097927	A/G	0.105	1.09	0.075	1.07	[0.799–1.450]	0.314	+++	0.080
rs7567217	2	21303470	APOB	T/C	0.860	1.07	0.858	1.02	[0.811–1.297]	0.415	−+	0.060
rs7568458	2	85788175	GGCX	A/T	0.449	1.06	0.460	1.06	[0.904–1.247]	0.232	+−+	0.140
rs17678683	2	145286559	ZEB2	G/T	0.088	1.10	0.079	1.23	[0.900–1.692]	0.095	+−	0.390
rs2252641	2	145801461	ZEB2	C/T	0.475	1.03	0.450	1.01	[0.867–1.186]	0.430	++−	0.050
rs1250229	2	216304384	FN1	T/C	0.256	1.07	0.260	0.88	[0.730–1.060]	0.911	−	0.390
rs2571445	2	218683154	TNS1	A/G	0.390	1.04	0.388	1.05	[0.894–1.234]	0.275	+−+	0.110
rs1801251	2	233633460	KCNJ13	A/G	0.350	1.05	0.344	1.15	[0.975–1.352]	0.049	++−	0.490
rs7623687	3	49448566	RHOA	A/C	0.855	1.08	0.864	0.96	[0.756–1.212]	0.626	+++	0.060
rs142695226	3	124475201	ITGB5	G/T	0.138	1.07	0.152	0.95	[0.754–1.195]	0.672	−+−	0.090
rs201477372	3	138099161	MRAS	TTTC/T	0.163	1.08	0.155	1.02	[0.829–1.264]	0.411	−+−	0.060
rs12493885	3	153839866	ARHGEF26	C/G	0.886	1.07	0.866	0.87	[0.691–1.089]	0.890	+−	0.260
rs17087335	4	57838583	NOA1	T/G	0.210	1.06	0.177	0.99	[0.812–1.215]	0.523	−++	0.050
rs10857147	4	81181072	FGF5	T/A	0.275	1.05	0.282	1.02	[0.855–1.215]	0.417	++−	0.060
rs7678555	4	120909501	MAD2L1	C/A	0.301	1.05	0.287	0.92	[0.770–1.100]	0.818	−+	0.200
rs4593108	4	148281001	MIR548G	C/G	0.795	1.07	0.831	0.93	[0.761–1.145]	0.743	+−	0.110
rs1878406	4	148393664	EDNRA	C/T	0.844	0.94	0.894	0.94	[0.740–1.208]	0.328	−+	0.070
rs72689147	4	156639888	GUCY1A3	G/T	0.817	1.07	0.812	1.13	[0.919–1.390]	0.122	+++	0.300
rs273909	5	131667353	SLC22A4	G/A	0.117	1.06	0.104	0.94	[0.726–1.217]	0.680	−	0.090
rs246600	5	142516897	ARHGAP26	T/C	0.480	1.05	0.486	0.98	[0.839–1.155]	0.575	−+−	0.060
rs9349379	6	12903957	PHACTR1	G/A	0.432	1.14	0.392	1.07	[0.919–1.267]	0.174	+++	0.200
rs7454157	6	12909874	PHACTR1	G/A	0.651	1.10	0.622	1.15	[0.978–1.371]	0.044	+++	0.550
rs6909752	6	22612629	HDGFL1	A/G	0.351	1.05	0.374	1.01	[0.851–1.186]	0.476	−+−	0.050
rs3130683	6	31888367	C2	T/C	0.860	1.09	0.960	0.69	[0.455–1.034]	0.964	+−	0.580
rs17609940	6	35034800	ANKS1A	G/C	0.824	1.03	0.796	1.32	[1.077–1.635]	0.004	+++	0.890
rs56336142	6	39134099	KCNK5	T/C	0.807	1.07	0.786	1.04	[0.860–1.265]	0.333	+−	0.080
rs10947789	6	39174922	KCNK5	T/C	0.775	1.05	0.753	1.07	[0.890–1.293]	0.229	+−+	0.150
rs12202017	6	134173151	TCF21	A/G	0.700	1.07	0.713	1.02	[0.856–1.215]	0.411	+−	0.060
rs12190287	6	134214525	TCF21	C/G	0.617	1.06	0.633	1.06	[0.896–1.270]	0.232	+−	0.150
rs2048327	6	160863532	SLC22A3	T/C	0.646	0.94	0.633	1.00	[0.854–1.188]	0.538	−+−	0.050
rs3798220	6	160961137	LPA	T/C	0.975	0.70	0.986	0.74	[0.397–1.393]	0.178	+−	0.140
rs55730499	6	161005610	LPA	T/C	0.056	1.37	0.079	1.22	[0.925–1.609]	0.078	+0+	0.350
rs4252185	6	161123451	PLG	C/T	0.060	1.34	0.087	1.15	[0.862–1.541]	0.168	+++	0.210
rs4252120	6	161143608	PLG	T/C	0.740	1.03	0.709	0.91	[0.766–1.084]	0.852	−+	0.240
rs2023938	7	19036775	HDAC9	T/C	0.897	0.94	0.899	0.87	[0.679–1.135]	0.161	+−	0.200
rs2107595	7	19049388	HDAC9	A/G	0.200	1.08	0.174	1.14	[0.926–1.406]	0.106	+++	0.310
rs12539895	7	107091849	COG5	C/A	0.807	1.04	0.788	1.14	[0.931–1.394]	0.101	+−+	0.350
rs10953541	7	107244545	BCAP29	C/T	0.783	1.05	0.750	1.16	[0.967–1.401]	0.054	++−	0.490
rs11556924	7	129663496	ZC3HC1	C/T	0.687	1.08	0.593	0.97	[0.832–1.147]	0.610	−+	0.060
rs10237377	7	139757136	PARP12	T/G	0.350	0.95	0.362	0.91	[0.766–1.084]	0.148	−+−	0.260

Table 2 continued

SNP	CHR	BP	Locus	Published GWAS results			Discovery GWAS in T1D patients					
				EA[a]	EAF[b]	OR[c]	EAF[d]	OR[e]	[95% CI]	P[f]	Direction[g]	Power[h]
rs3918226	7	150690176	NOS3	T/C	0.060	1.14	0.088	0.90	[0.672–1.212]	0.751	−++	0.140
rs264	8	19813180	LPL	G/A	0.853	1.06	0.846	0.94	[0.756–1.168]	0.711	−+	0.100
rs2954029	8	126490972	TRIB1	A/T	0.551	1.04	0.529	1.05	[0.896–1.232]	0.269	++−	0.110
rs3217992	9	22003223	CDKN2BAS1	C/T	0.607	0.88	0.614	0.92	[0.789–1.084]	0.169	−	0.200
rs2891168	9	22098619	CDKN2BAS1	G/A	0.489	1.21	0.505	1.16	[0.993–1.362]	0.030	+++	0.6
rs2519093	9	136141870	ABO	T/C	0.191	1.08	0.186	0.98	[0.800–1.203]	0.571	−++	0.060
rs2487928	10	30323892	KIAA1462	A/G	0.418	1.06	0.458	1.08	[0.924–1.271]	0.161	−++	0.220
rs2624695	10	44549767	CXCL12	C/T	0.534	0.94	0.503	1.08	[0.926–1.270]	0.843	+++	0.220
rs501120	10	44753867	CXCL12	T/C	0.813	1.08	0.863	1.18	[0.927–1.508]	0.088	+++	0.400
rs11203043	10	90989279	LIPA	G/A	0.576	1.04	0.551	0.89	[0.764–1.052]	0.909	−	0.360
rs1412444	10	91002927	LIPA	T/C	0.369	1.07	0.326	0.94	[0.798–1.116]	0.749	−+−	0.130
rs11191416	10	104604916	CYP17A1	T/G	0.873	1.08	0.915	0.90	[0.686–1.186]	0.768	+−	0.120
rs11042937	11	10745394	MRVI1-CTR9	T/G	0.490	1.04	0.499	1.06	[0.908–1.248]	0.220	++−	0.150
rs3993105	11	13303071	ARNTL	T/C	0.704	1.05	0.693	1.03	[0.866–1.228]	0.364	−+−	0.070
rs12801636	11	65391317	PCNX3	A/G	0.230	0.95	0.225	0.98	[0.810–1.189]	0.425	+−	0.060
rs590121	11	75274150	SERPINH1	T/G	0.300	1.05	0.295	0.80	[0.669–0.961]	0.992	−	0.830
rs9319428	13	28973621	FLT1	A/G	0.314	1.04	0.289	1.03	[0.874–1.227]	0.339	−++	0.070
rs4773144	13	110960712	COL4A2	A/G	0.572	0.95	0.555	0.99	[0.850–1.172]	0.493	−++	0.050
rs11838776	13	111040681	COL4A2	A/G	0.263	1.07	0.284	1.33	[1.113–1.606]	0.001	++−	0.980
rs9515203	13	111049623	COL4A2	T/C	0.761	1.07	0.736	1.20	[0.983–1.468]	0.036	+++	0.650
rs10139550	14	100145710	HHIPL1	G/C	0.423	1.06	0.413	0.96	[0.819–1.145]	0.645	−+	0.070
rs6494488	15	65024204	RBPMS2	G/A	0.180	0.95	0.151	1.01	[0.803–1.280]	0.545	+−	0.050
rs56062135	15	67455630	SMAD3	C/T	0.790	1.07	0.758	1.07	[0.893–1.302]	0.214	++−	0.160
rs7173743	15	79141784	ADAMTS7	T/C	0.564	1.08	0.515	1.00	[0.855–1.171]	0.495	−++	0.050
rs8042271	15	89574218	ABHD2	G/A	0.900	1.10	0.947	0.93	[0.602–1.457]	0.613	−+	0.070
rs17514846	15	91416550	FURIN-FES	A/C	0.440	1.05	0.461	1.08	[0.924–1.267]	0.163	++−	0.210
rs1800775	16	56995236	CETP	C/A	0.510	1.04	0.524	0.99	[0.844–1.153]	0.568	−+−	0.050
rs1050362	16	72130815	DHX38	A/C	0.380	1.04	0.351	1.11	[0.939–1.313]	0.110	+−+	0.310
rs7500448	16	83045790	CDH13	A/G	0.752	1.06	0.755	1.09	[0.894–1.321]	0.202	+−+	0.190
rs216172	17	2126504	SMG6	C/G	0.350	1.05	0.368	1.13	[0.963–1.329]	0.067	+++	0.420
rs12936587	17	17543722	RAI1	G/A	0.611	1.03	0.542	1.00	[0.857–1.174]	0.480	−++	0.050
rs17608766	17	45013271	GOSR2	C/T	0.140	1.07	0.133	0.89	[0.705–1.121]	0.840	−	0.220
rs999474	17	46987665	UBE2Z	G/A	0.600	1.04	0.572	0.92	[0.789–1.084]	0.832	−	0.130
rs7212798	17	59013488	BCAS3	C/T	0.150	1.08	0.155	0.98	[0.786–1.221]	0.570	−++	0.060
rs1867624	17	62387091	PECAM1	C/T	0.390	0.96	0.377	1.06	[0.898–1.241]	0.744	+++	0.130
rs663129	18	57838401	U4/MC4R	A/G	0.260	1.06	0.237	1.01	[0.847–1.227]	0.418	++−	0.060
rs1122608	19	11163601	LDLR	G/T	0.770	1.07	0.756	0.87	[0.731–1.048]	0.926	+−	0.380
rs56289821	19	11188247	LDLR	G/A	0.900	1.14	0.880	1.10	[0.855–1.427]	0.223	+−+	0.170
rs12976411	19	32882020	ZNF507	A/T	0.910	1.61	0.958	0.96	[0.639–1.460]	0.564	−+	0.050
rs8108632	19	41854534	TGFB1	T/A	0.488	1.05	0.445	1.11	[0.938–1.307]	0.113	++−	0.330
rs2075650	19	45395619	TOMM40	A/G	0.865	0.93	0.869	0.74	[0.596–0.919]	0.003	−	0.870
rs445925	19	45415640	APOE/APOC1	G/A	0.902	1.09	0.893	0.87	[0.681–1.129]	0.844	++−	0.210
rs4420638	19	45422946	APOE/APOC1	G/A	0.166	1.10	0.171	1.20	[0.985–1.461]	0.034	+−+	0.530
rs1964272	19	46190268	SNRPD2	G/A	0.510	1.05	0.509	0.95	[0.805–1.113]	0.747	−+	0.110
rs867186	20	33764554	PROCR	G/A	0.110	0.93	0.094	1.26	[0.974–1.623]	0.961	+++	0.490

Table 2 continued

SNP	CHR	BP	Locus	Published GWAS results			Discovery GWAS in T1D patients					
				EA[a]	EAF[b]	OR[c]	EAF[d]	OR[e]	[95% CI]	P[f]	Direction[g]	Power[h]
rs28451064	21	35593827	KCNE2	A/G	0.121	1.14	0.126	1.36	[1.066–1.751]	0.006	+−+	0.890
rs180803	22	24658858	POM121L9P	G/T	0.970	1.20	0.979	1.18	[0.510–2.743]	0.348	−++	0.120

[a] Estimated allele/non estimated allele

[b] Allele frequency of the estimate allele reported in [14, 24–26]

[c] Odds ratio for CAD reported in [14, 24–26]

[d] Allele frequency of the estimated allele in the discovery GWAS of T1D patients

[e] Odds ratio for CAD [95% confidence interval] observed in the discovery GWAS of T1D patients

[f] One sided test p value of association

[g] Directionality of the effects across the contributing cohorts

[h] Power estimates were provided by the CaTS program [23] and correspond to the power of our discovery GWAS to achieve 0.05 statistical significance at the observed associations based on EAF[d] and OR[e] under the assumption of a multiplicative model (on log-scale)

Discussion

The present work was aimed at identifying susceptibility alleles for CAD risk in patient population of T1D using a GWAS approach with a two-step framework (discovery + replication). Albeit we identified one locus (*MAP1B*) reaching genome-wide significance in the discovery stage, it did not replicate with similar effects. Nevertheless, in the combined T1D dataset of 1019 cases and 5735 controls, we observed strong statistical evidence for association with CAD at three biological candidate genes, *CDK18, PKD1* and *FAM189A2*.

We consider our study as very original as it is the first one, to the best of our knowledge, to use a GWAS approach for CAD, in patients with type 1 diabetes. We have thus pooled the largest collection of type 1 diabetes patients with available genetic data.

The low frequency *CDK18* rs138760780-T allele (frequency ~ 0.02) was found associated with ~ 2.5 fold increased risk of CAD. According to public database (e.g. Haploreg [27]), this SNP does not show strong linkage disequilibrium (LD) (pairwise $r^2 > 0.80$) with other SNPs at this locus, consistent with the regional association plot that does not suggest any evidence of disease associated SNPs (Additional file 7: Figure S4). Interrogating the functional status of this SNP through HaploReg tool [27] suggested that this SNP may be involved in some epigenetic regulatory mechanisms. *CDK18* encodes for a cyclin-dependent kinase, suggesting a role in cell cycle. This predicted protein is also related to *CDK1*, which is involved in the G2/M transition in eukaryotic cells [28]. Although cell cycle is a very broad pathway, *CDK1* has also been associated with T1D [29], but at this point little is known about a potential involvement of *CDK18* in the pathophysiology of T1D or its complications.

We also observed some evidence that the low frequency *FAM189A2* rs12344245 G allele, (frequency ~ 0.04), associated with a ~ 1.8 fold-increased risk of CAD. We did not find any evidence suggesting that this intronic SNP, or any other SNPs in LD (Additional file 8: Figure S5: regional association plot) with it, could be functional. Nevertheless, even though not much is known about the role of the encoded protein, this locus is a good candidate. Indeed, genetic variations at this locus have been found associated with albumin to creatinine ratio [30]. More interestingly, two *FAM189A2* SNPs (rs10780297 and rs10120442) have been reported to moderately associate ($p = 9.3 \times 10^{-4}$) in a large GWAS for CAD in ~ 63,000 non-diabetic populations [14], suggesting that this locus could be a CAD locus in some specific at-risk groups of diabetic patients. The latter two SNPs are in moderate LD ($D' = 1$ but $r^2 = 0.05$) with our lead rs12344245 SNP, indicating that a fine mapping analysis of this locus would warrant further investigations. Of interest, it was not identified as a common gene in both type 2 diabetes and CAD. It can thus be speculated this gene is an important gene in high-glucose environment rather than a gene leading to high-glucose.

Finally, we observed an association of the non-synonymous *PKD1* rs116092985 (Trp1399Arg) with CAD among T1D patients where the Arg1399 minor allele (frequency ~ 0.10), was associated with an increased CAD risk (OR ~ 1.5). The regional plot (Additional file 9: Figure S6) shows that there are several SNPs in LD with this *PKD1* top SNP that associate with CAD. PKD1 encodes for the Polycystin 1, Transient Receptor Potential Channel Interacting protein, a member of the polycystin protein family. Recent reports have suggested a role of *PKD1* not only in renal tubular function and structure [31]

but rare mutations in this gene as the main also cause underlying polycystic kidney disease [32], highlighting its importance in kidney complications. One important question is whether *PKD1* risk allele is involved in a common genetic background linking DN and CAD. This question was not duly analyzed due to power issue. However, no clear association was established with DN in previous GWAS focusing on this question [18].

Our study also enabled us to assess in patients with T1D the impact of common SNPs that have been found associated with CAD in large GWAS performed in unselected individuals. Beyond the observation that most of the previously reported SNPs showed consistent association with CAD in our T1D population, this look-up identified a few CAD loci (*ANKS1A*, *COL4A2*, *TOMM40/APOE*) where the reported CAD associated SNP could have a stronger effect in T1D patients. However, this hypothesis would require further investigation.

Some limitations must be acknowledged. We did not consider differently T1D patients with and without DN, and all of the analyses were not stratified on DN status in order to keep all CAD patients in the analysis. However, this should have limited impact on our main results as none of the SNPs identified here were positive considering previously reported GWAS with regard to DN as primary endpoint [18]. Another limitation pertains to limited power of our sample size required, particular to overcome the harsh genome-wide statistical significance threshold. Indeed, our discovery GWAS was not well powered to identify common SNPs associated with moderate genetic effects as those frequently encountered in a GWAS context. For instance, our discovery study had no power to detect at the genome-wide statistical threshold the genetic effect of a variant with an associated allelic OR less than 1.40. It was only well powered (> 80%) to detect OR greater than 1.6 as soon as the allele frequency of the disease allele is greater than 0.28 and well powered to detect OR greater than ~ 2 for allele frequency greater than 0.05. In particular, we had no power to detect the well-established association of the 9p21 locus at the 5×10^{-8} threshold while we had a chance of 60% to detect it would the liberal threshold of 0.05 had been used. Similarly, we acknowledge the low power of our replication studies where none of the tested associations achieved the Bonferroni threshold of 3×10^{-3}. We only had a power of 53, 38 and 26% to detect at this threshold a significant association at the *CDK18* rs13876070, *PKD1* rs116092985 and the *FAM189A2* loci, respectively.

Despite these limitations, we have assembled the largest cohort available and conducted novel analyses to discover novel candidate loci for CAD in T1D patients that need to be further studied with additional epidemiological data and functional work to confirm our findings. Of interest, our negative study could suggest a role of epigenetics beyond genomics. In this regard, insulin promotes the expression of DNA methyltransferases leading to methylation resulting in atherosclerosis [33], broadening the research field of CAD in type 1 diabetes.

Conclusions

We identified three new candidate loci for CAD in individuals with T1D, these loci weren't previously found associated with CAD in the general population. Three other loci previously reported in the general population were found associated with CAD in our setting, namely *ANKS1A*, *COL4A2* and *TOMM40/APOE*. Although this work needs further investigation, studying the function of these loci could lead to a better understanding of the physiological pathways involved in the development of CAD as a complication of T1D.

Additional files

Additional file 1: Table S1. Main design and sample characteristics of the discovery and replication studies.

Additional file 2: Figure S1. Quantile-Quantile plot representation of the discovery meta-GWAS results.

Additional file 3: Figure S2. Manhattan plot representation of the discovery meta-GWAS results.

Additional file 4: Figure S3. Regional association plot at the *MAP1B* locus.

Additional file 5: Table S2. Imputation quality of SNPs with association p-values < 1.0×10^{-5} in the discovery cohorts.

Additional file 6: Table S3. Imputation quality of the established CAD associated SNPs in the discovery cohorts.

Additional file 7: Figure S4. Regional association plot at the *CKD18* locus.

Additional file 8: Figure S5. Regional association plot at the *FAM189A2* locus.

Additional file 9: Figure S6. Regional association plot at the *PKD1* locus.

Abbreviations

CAD: coronary artery disease; T1D: type 1 diabetes; GWAS: genome wide association study; CVD: cardiovascular disease; DN: diabetic nephropathy; SNP: single nucleotide polymorphism; MAF: minor allele frequency; OR: odds ratio; LD: linkage disequilibrium; CI: confidence interval.

Authors' contributions

RC performed the GWAS for the French cohort as well as the meta-analyses for the three discovery cohorts and the five replication cohorts. SD performed the GWAS for the Ireland-UK cohort. SK performed the GWASes for the five replication cohorts. TSA performed the GWAS for the Danish cohort. DAT and SH interpreted the results of the analysis and contributed in writing the manuscript. All authors read and approved the final manuscript.

Author details

[1] Institut National pour la Santé et la Recherche Médicale (INSERM), Unité Mixte de Recherche en Santé (UMR_S) 1166, Team Genomics & Pathophysiology of Cardiovascular Diseases, Sorbonne Universités, UPMC Univ. Paris 06, Paris, France. [2] ICAN Institute for Cardiometabolism and Nutrition, Paris, France. [3] Centre for Public Health, Queen's University of Belfast, Belfast, Northern Ireland, UK. [4] Genetics & Genome Biology Program, Hospital for Sick Children,

Toronto, Canada. [5] Département de Diabétologie, Endocrinologie et Nutrition, Assistance Publique Hôpitaux de Paris, Hôpital Bichat, DHU FIRE, Paris, France. [6] UFR de Médecine, Université Paris Diderot, Sorbonne Paris Cité, Paris, France. [7] Steno Diabetes Center Copenhagen, Gentofte, Denmark. [8] University of Copenhagen, Copenhagen, Denmark. [9] Dalla Lana School of Public Health, University of Toronto, Toronto, Canada. [10] UFR de Médecine et Pharmacie, Université de Poitiers, Poitiers, France. [11] INSERM, CIC 1402 & U1082, Poitiers, France. [12] Service d'Endocrinologie-Diabétologie and Centre d'Investigation Clinique, CHU de Poitiers, BP 577, 86021 Poitiers Cedex, France.

Competing interests
The authors declare that they have no competing interests.

Availability of data and materials
Summary statistics of the discovery GWAS are available in the GRASP resources database (https://grasp.nhlbi.nih.gov/FullResults.aspx). The French cohorts details are available in additional files.

Ethics approval and consent to participate
All participating studies were approved by their respective institutional review board/ethics committee and an informed consent was obtained from all participating individuals.

Funding
R.C was supported by a Ph.D. grant from the Region Ile de France (CORDDIM program).

The UK-ROI component was funded by Diabetes UK, the JDRF, the Health Research Board, Science Foundation Ireland, and Department for the Economy NI 15/IA/3152; this includes the Warren 3/UK GoKinD Study Group, which includes, Belfast: A. P. Maxwell, A. J. McKnight, D. A. Savage; Edinburgh: J. Walker; London: S. Thomas, G. C. Viberti; Manchester: A. J. M. Boulton; Newcastle: S. Marshall; Plymouth: A. G. Demaine, B. A. Millward; Swansea: S. C. Bain. SD was supported by a Ph.D. studentship from the Northern Ireland Department of Education and Learning.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Received: 8 January 2018 Accepted: 16 April 2018
Published online: 25 April 2018

References

1. Menke A, Casagrande S, Geiss L, Cowie CC. Prevalence of and trends in diabetes among adults in the United States, 1988–2012. JAMA. 2015;314:1021–9.
2. Kusnik-Joinville O, Weill A, Ricordeau P, Allemand H. Diabète traité en France en 2007: un taux de prévalence proche de 4% et des disparités géographiques croissantes. BEH. 2008;43:409–13.
3. Daneman D. Type 1 diabetes. Lancet. 2006;367:847–58.
4. Livingstone SJ, Levin D, Looker HC, Lindsay RS, Wild SH, Joss N, et al. Estimated life expectancy in a Scottish cohort with type 1 diabetes, 2008–2010. JAMA. 2015;313:37–44.
5. Lind M, Svensson AM, Kosiborod M, Gudbjörnsdottir S, Pivodic A, Wedel H, et al. Glycemic control and excess mortality in type 1 diabetes. N Engl J Med. 2014;371:1972–82.
6. Yusuf S, Hawken S, Ounpuu S, Dans T, Avezum A, Lanas F, et al. Effect of potentially modifiable risk factors associated with myocardial infarction in 52 countries (the INTERHEART study): case-control study. Lancet. 2004;364:937–52.
7. Nathan DM, Cleary PA, Backlund J-YC, Genuth SM, Lachin JM, Orchard TJ, et al. Intensive diabetes treatment and cardiovascular disease in patients with type 1 diabetes. N Engl J Med. 2005;353:2643–53.
8. Earle K, Walker J, Hill C, Viberti G. Familial clustering of cardiovascular disease in patients with insulin-dependent diabetes and nephropathy. N Engl J Med. 1992;326:673–7.
9. Hokanson JE, Cheng S, Snell-Bergeon JK, Fijal BA, Grow MA, Hung C, et al. A common promoter polymorphism in the hepatic lipase gene (LIPC-480C>T) is associated with an increase in coronary calcification in type 1 diabetes. Diabetes. 2002;51:1208–13.
10. Pettersson-Fernholm K, Karvonen MK, Kallio J, Forsblom CM, Koulu M, Pesonen U, et al. Leucine 7 to proline 7 polymorphism in the preproneuropeptide Y is associated with proteinuria, coronary heart disease, and glycemic control in type 1 diabetic patients. Diabetes Care. 2004;27:503–9.
11. Kretowski A, Hokanson JE, McFann K, Kinney GL, Snell-Bergeon JK, Maahs DM, et al. The apolipoprotein A-IV Gln360His polymorphism predicts progression of coronary artery calcification in patients with type 1 diabetes. Diabetologia. 2006;49:1946–54.
12. Pettersson-Fernholm K, Forsblom C, Hudson BI, Perola M, Grant PJ, Groop P-H, et al. The functional -374 T/A RAGE gene polymorphism is associated with proteinuria and cardiovascular disease in type 1 diabetic patients. Diabetes. 2003;52:891–4.
13. CARDIoGRAMplusC4D Consortium, Deloukas P, Kanoni S, Willenborg C, Farrall M, Assimes TL, et al. Large-scale association analysis identifies new risk loci for coronary artery disease. Nat Genet. 2013;45:25–33.
14. Nikpay M, Goel A, Won H-H, Hall LM, Willenborg C, Kanoni S, et al. A comprehensive 1,000 Genomes-based genome-wide association meta-analysis of coronary artery disease. Nat Genet. 2015;47:1121–30.
15. Association AD. Diagnosis and classification of diabetes mellitus. Diabetes Care. 2010;33:S62–9.
16. Hadjadj S, Cariou B, Fumeron F, Gand E, Charpentier G, Roussel R, et al. Death, end-stage renal disease and renal function decline in patients with diabetic nephropathy in French cohorts of type 1 and type 2 diabetes. Diabetologia. 2016;59:208–16.
17. Lajer M, Jorsal A, Tarnow L, Parving H-H, Rossing P. Plasma growth differentiation factor-15 independently predicts all-cause and cardiovascular mortality as well as deterioration of kidney function in type 1 diabetic patients with nephropathy. Diabetes Care. 2010;33:1567–72.
18. Sandholm N, Salem RM, McKnight AJ, Brennan EP, Forsblom C, Isakova T, et al. New susceptibility loci associated with kidney disease in type 1 diabetes. PLoS Genet. 2012;8:e1002921.
19. Li Y, Willer CJ, Ding J, Scheet P, Abecasis GR. MaCH: using sequence and genotype data to estimate haplotypes and unobserved genotypes. Genet Epidemiol. 2010;34:816–34.
20. Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MAR, Bender D, et al. PLINK: a tool set for whole-genome association and population-based linkage analyses. Am J Hum Genet. 2007;81:559–75.
21. Willer CJ, Li Y, Abecasis GR. METAL: fast and efficient meta-analysis of genomewide association scans. Bioinformatics. 2010;26:2190–1.
22. Higgins JPT, Thompson SG. Quantifying heterogeneity in a meta-analysis. Stat Med. 2002;21:1539–58.
23. Hsieh FY, Bloch DA, Larsen MD. A simple method of sample size calculation for linear and logistic regression. Stat Med. 1998;17:1623–34.
24. Nelson CP, Goel A, Butterworth AS, Kanoni S, Webb TR, Marouli E, et al. Association analyses based on false discovery rate implicate new loci for coronary artery disease. Nat Genet. 2017;49:1385–91.
25. Howson JMM, Zhao W, Barnes DR, Ho W-K, Young R, Paul DS, et al. Fifteen new risk loci for coronary artery disease highlight arterial-wall-specific mechanisms. Nat Genet. 2017;49:1113–9.
26. Webb TR, Erdmann J, Stirrups KE, Stitziel NO, Masca NGD, Jansen H, et al. Systematic evaluation of pleiotropy identifies 6 further loci associated with coronary artery disease. J Am Coll Cardiol. 2017;69:823–36.
27. Ward LD, Kellis M. HaploReg: a resource for exploring chromatin states, conservation, and regulatory motif alterations within sets of genetically linked variants. Nucleic Acids Res. 2012;40:D930–4.
28. Okuda T, Cleveland JL, Downing JR. PCTAIRE-1 and PCTAIRE-3, two members of a novel cdc2/CDC28-related protein kinase gene family. Oncogene. 1992;7:2249–58.
29. Bradfield JP, Qu H-Q, Wang K, Zhang H, Sleiman PM, Kim CE, et al. A genome-wide meta-analysis of six type 1 diabetes cohorts identifies multiple associated loci. PLoS Genet. 2011;7:e1002293.
30. Böger CA, Chen M-H, Tin A, Olden M, Köttgen A, de Boer IH, et al. CUBN is a gene locus for albuminuria. J Am Soc Nephrol. 2011;22:555–70.
31. Song X, Di Giovanni V, He N, Wang K, Ingram A, Rosenblum ND, et al. Systems biology of autosomal dominant polycystic kidney disease

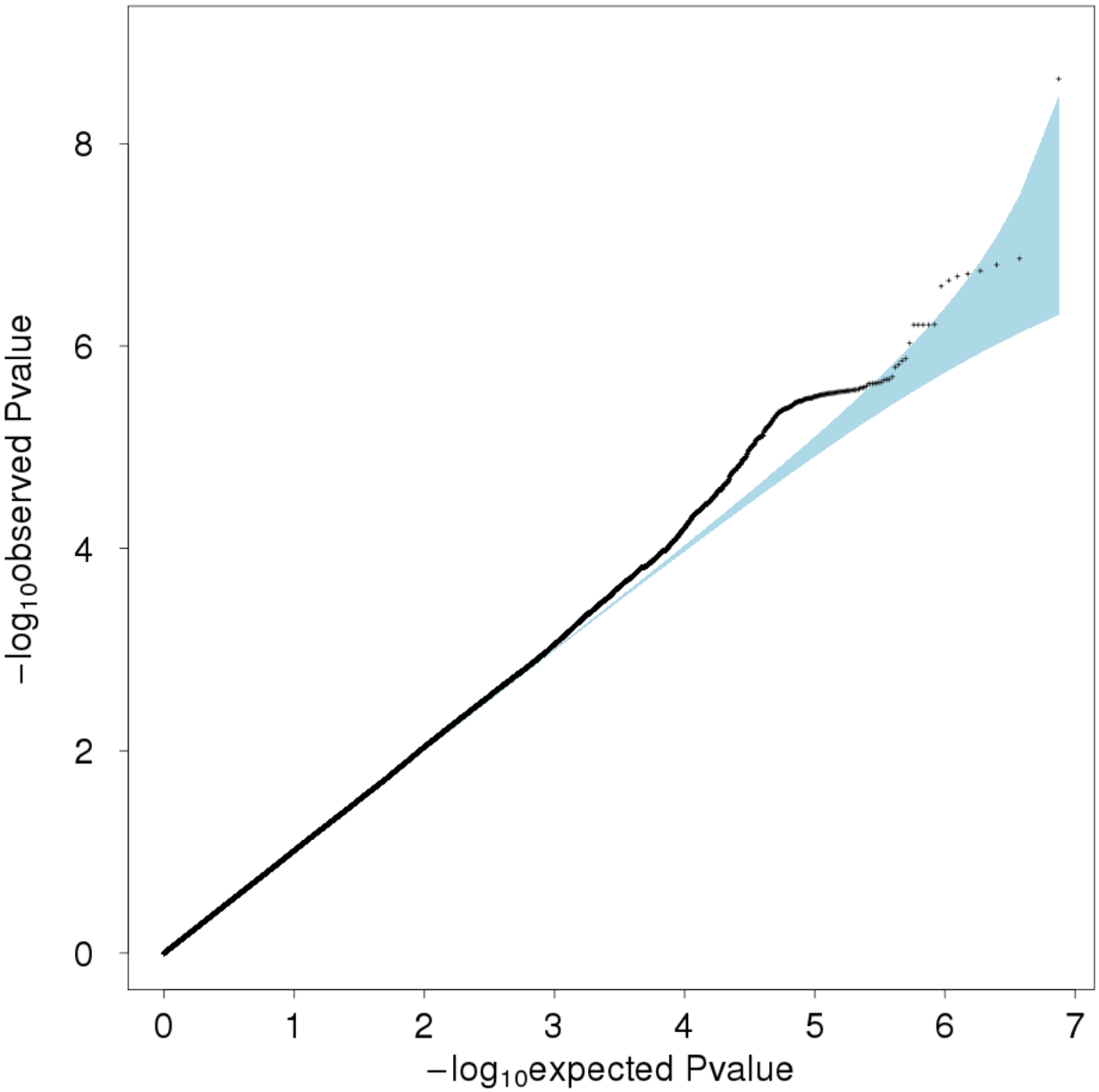
(ADPKD): computational identification of gene expression pathways and integrated regulatory networks. Hum Mol Genet. 2009;18:2328–43.
32. The International Polycystic Kidney Disease Consortium. Polycystic kidney disease: the complete structure of the PKD1 gene and its protein. Cell. 1995;81:289–98.
33. Min J, Weitian Z, Peng C, Yan P, Bo Z, Yan W, Yun B, Xukai W. Correlation between insulin-induced estrogen receptor methylation and atherosclerosis. Cardiovasc Diabetol. 2016;15:156.

Ready to submit your research? Choose BMC and benefit from:

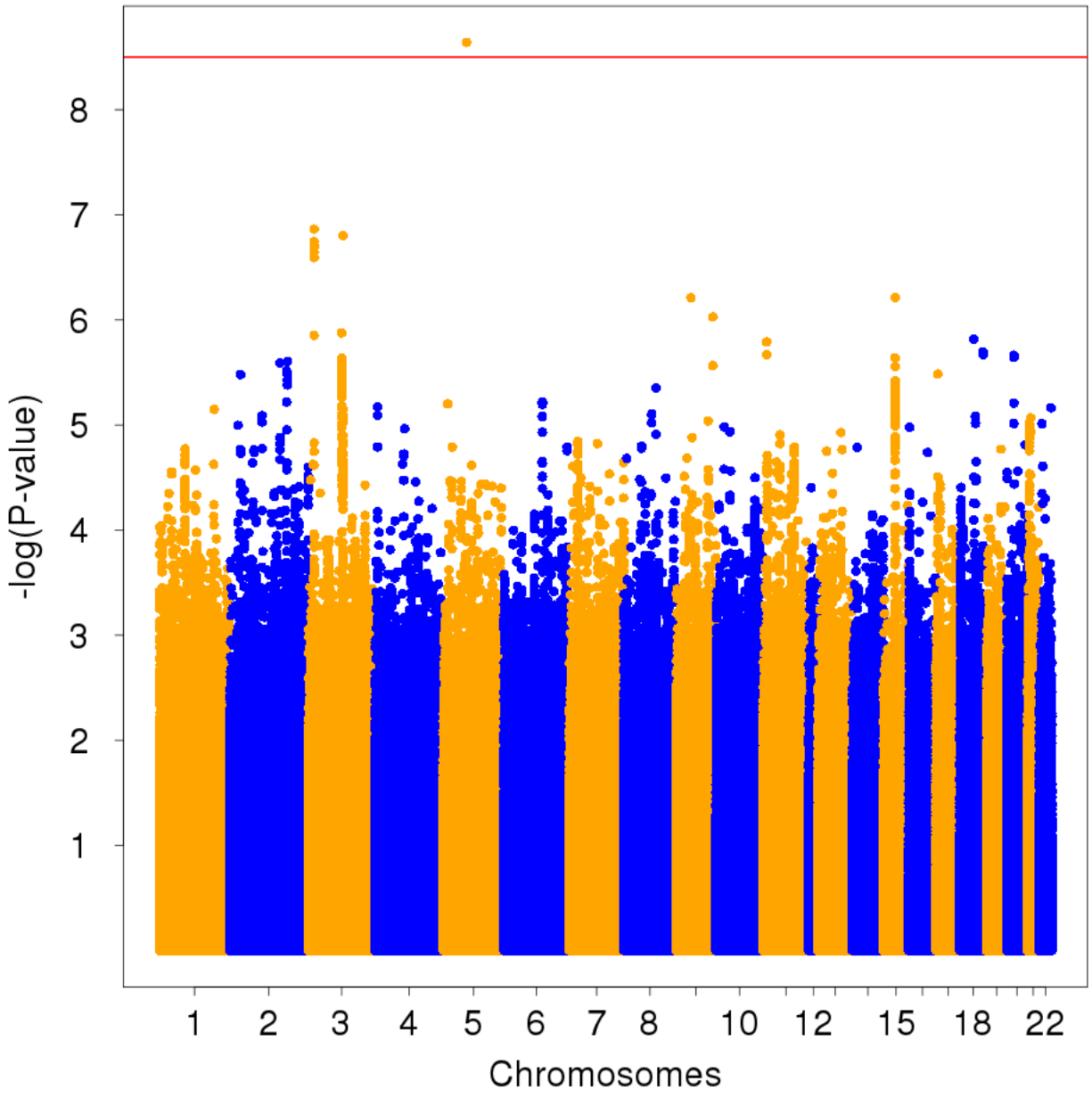
- fast, convenient online submission
- thorough peer review by experienced researchers in your field
- rapid publication on acceptance
- support for research data, including large and complex data types
- gold Open Access which fosters wider collaboration and increased citations
- maximum visibility for your research: over 100M website views per year

At BMC, research is always in progress.

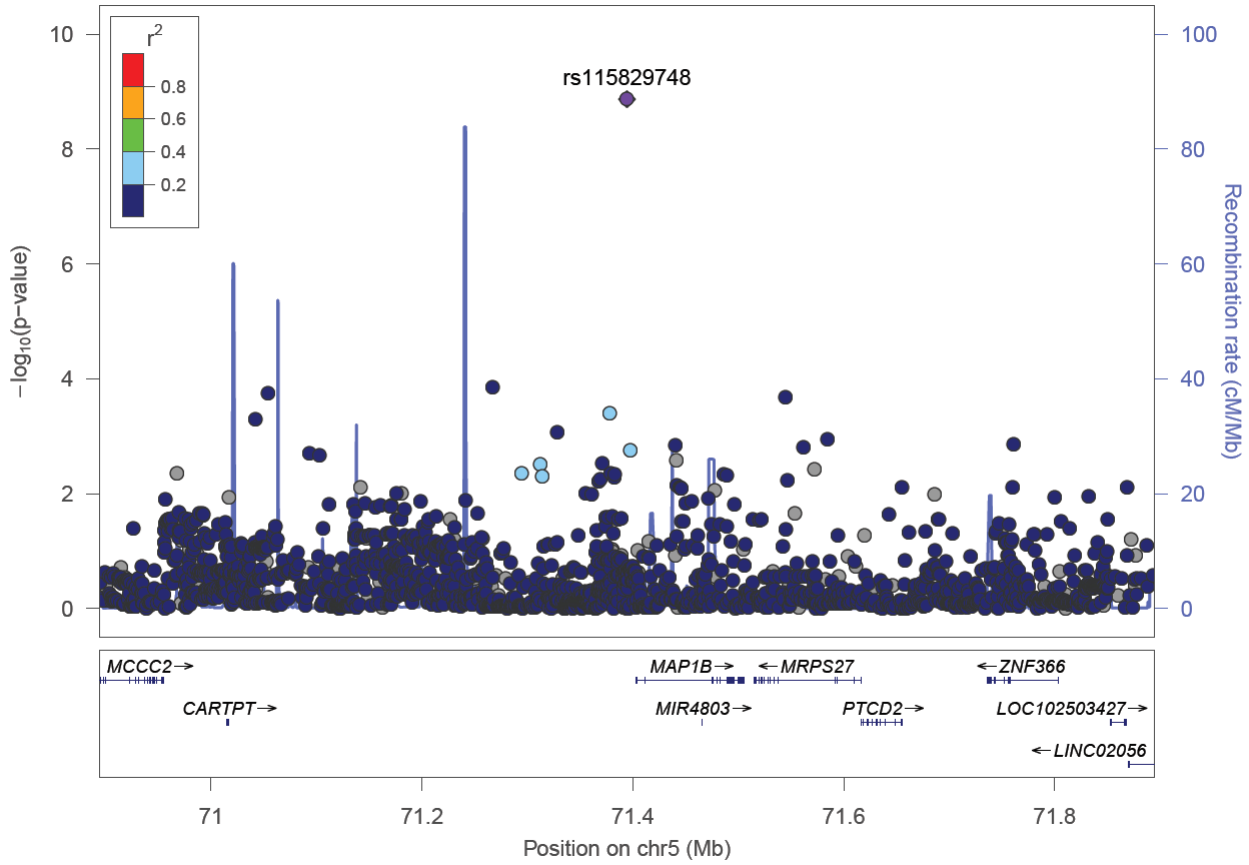
Learn more biomedcentral.com/submissions



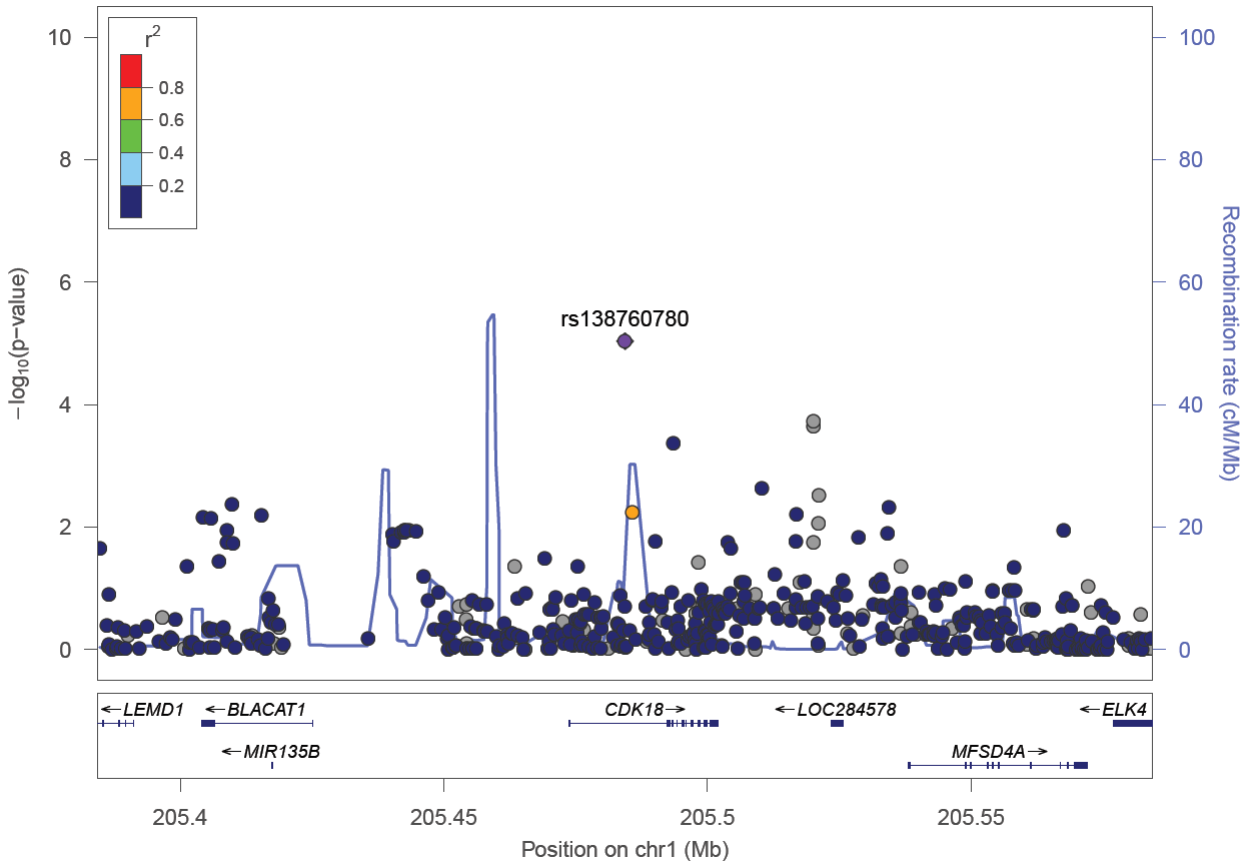
Supplementary figure 1



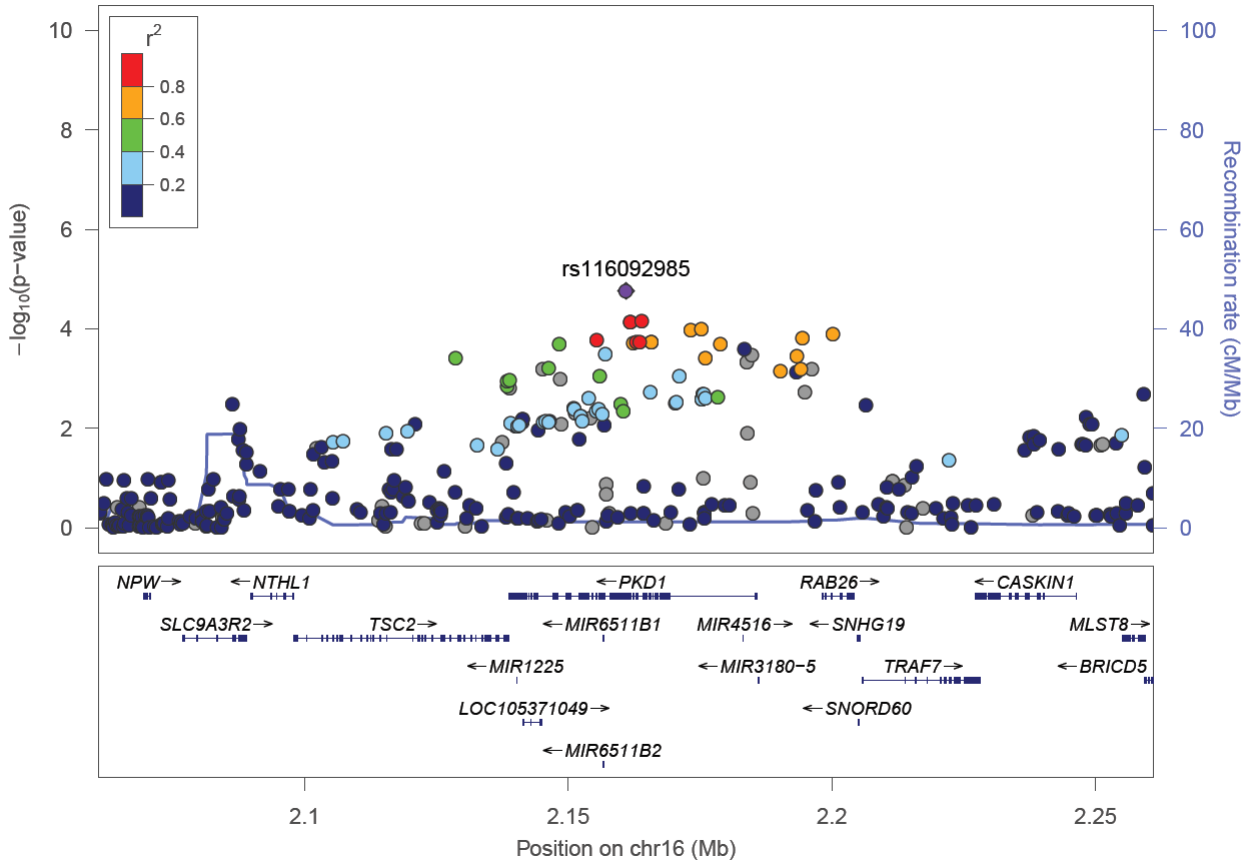
Supplementary figure 2



Supplementary figure 3



Supplementary figure 4



Supplementary figure 5

Study name	Category	Number of cases	Number of controls	Years	Population origin origin	Inclusion	Number of diabetic (Cases/Controls) nephropathy (Cases/Controls)	Case ascertainment	Mean age (SD)	Female (%)
Genesis/Genediab	Discovery	85	1285	1994-2000	France	T1D diagnosed and before 35 years of age and with diabetes duration $\geq$ 15 years	40/325	History of myocardial infarction or revascularization	40,7 (12,2)	47,4
Steno	Discovery	237	632	1993-2000	Denmark	T1D before 35 years old	177/251	History of CVD except stroke at baseline or during follow up	43,9 (11,1)	43
UK-ROI	Discovery	112	1206	1996-2000	British isles	T1D before 31 years old	97/514	Angina, MI or coronary artery bypass graft bypass graft	44,1 (11,1)	49,1
DCCT-EDIC	Replication	152	1152	1983-1989	United States Canada	T1D subjects aged between 13 and 39 years and with diabetes duration $\geq$ 1 year	28/165	Any CVD event except stroke	27 (7)	47
WESDR	Replication	202	392	1979-1980	United States (Wisconsin)	T1D individuals diagnosed <32 years of age and treated with insulin	NA	Any CVD event except stroke	26 (11)	50
CACTI	Replication	42	478	2000-2002	United States (Colorado)	T1D subjects with diabetes duration >10 years without CVD at baseline	NA	Any CVD event except stroke	39 (9)	44
EDC	Replication	139	289	1968-1988	United States (Pennsylvania)	childhood-onset (<17 years of age) T1D cases diagnosed or seen within 1 year of diagnosis	NA	Any CVD event except stroke	27 (8)	48
GoKIND	Replication	50	301	2000-2003	United States	subjects with long-term (>10 years) T1D diagnosed before 31 years old	NA	Any CVD event except stroke	37 (9)	66

Supplementary table 1 Part 1.

Study name	Genotyping chip and version	Subject exclusion	SNP exclusion	SNP count prior to imputation	Imputation panel and software	SNP count post imputation	Statistical analyses
Genesis/Genediab	Illumina HumanCore Exome 12v1	Unmatched sex missing call rate >5% outliers based on PCAs high kinship	Missing call rate >1% mitochondria sexual chromosomes HWE p-value<1E-05 MAF<0,1 r-square<0,3 for imputed SNPs	267 409	1000 Genomes phase 3v5 MaCH mach2dat	12 745 478	Logistic model with additive effect Age, sex, 4 first PCs and diabetic nephropathy as covariates
Steno	Illumina HumanCore Exome 12v1	Unmatched sex, missing call rate >5% genetic duplicates outliers based on PCAs extreme heterozygocity	Missing call rate >5% monomorphic SNPs HWE p-value<1E-06 INFO value<0,3 for imputed SNPs	286 626	1000 Genomes phase 3v5 MaCH Quicktest	11 177 883	Logistic model with additive effect Age, sex, 2 first PCs and diabetic nephropathy as covariates
UK-ROI	Illumina HumanCore Exome 12v1	Unmatched sex missing call rate >5% genetic duplicates outliers based on PCAs extreme heterozygosity	Mitochondria sex chromosomes MAF <0.01 missing call rate >5% monomorphic SNPs HWE p-value<1E-08 r-square < 0.29 for imputed SNPs	253 271	1000 Genomes phase 3v5 MaCH Plink 1.9	13 863 298	Logistic model with additive effect Age, sex, PCs and diabetic nephropathy as covariates
DCCT-EDIC	Illumina Human 1M	Unmatched sex missing call rate >2% outliers based on PCAs	Missing call rate >2% mitochondria sex chromosomes MAF<0.01 info > 0.8 for imputed SNPs	841 366	1000 Genomes phase 3 , IMPUTE2* SNPTEST v2.5†	19 210 835	Logistic model with additive effect Age, sex duration of diabetes as covariates
WESDR	Illumina HumanOmni1-Quad	Unmatched sex missing call rate >2% outliers based on PCAs	Missing call rate >5% mitochondria sex chromosomes MAF<0.01 info > 0.8 for imputed SNPs	861 503	1000 Genomes phase 3 IMPUTE2 SNPTEST v2.5 SNPTEST v2.5	12 969 212	Logistic model with additive effect Age, sex duration of diabetes as covariates
CACTI	Illumina HumanCore Exome v3	Unmatched sex missing call rate >2% outliers based on PCAs	Missing call rate >5% mitochondria sex chromosomes MAF<0.01 info > 0.8 for imputed SNPs	522 731	1000 Genomes phase 3 IMPUTE2 SNPTEST v2.5	14 036 939	Logistic model with additive effect Age, sex duration of diabetes as covariates
EDC	Illumina HumanOmni1-Quad	Unmatched sex missing call rate >2% outliers based on PCAs	Missing call rate >5% mitochondria sex chromosomes MAF<0.01 info > 0.8 for imputed SNPs	311 840	1000 Genomes phase 3 MaCH SNPTEST v2.5	12 091 401	Logistic model with additive effect Age, sex duration of diabetes as covariates
GoKIND	Illumina HumanCore Exome v3	Unmatched sex missing call rate >2% outliers based on PCAs	Missing call rate >5% mitochondria sex chromosomes MAF<0.01 info > 0.8 for imputed SNPs	330 567	1000 Genomes phase 3 MaCH SNPTEST v2.5	11 876 576	Logistic model with additive effect Age, sex duration of diabetes as covariates

Supplementary table 1 Part 2.

*J. Marchini and B. Howie (2010) Genotype imputation for genome-wide association studies. Nature Reviews Genetics

†J. Marchini, B. Howie, S. Myers, G. McVean, and P. Donnelly (2007) A new multipoint method for genome-wide association studies via imputation of genotypes. Nature Genetics 39 : 906-913T

Titre : Étude des interactions entre la fonction rénale, la thrombose artérielle et la thrombose veineuse par une approche génétique

Résumé : L'objet de cette thèse est l'étude de la génétique des coronaropathies dans le contexte du diabète de type I et de la génétique de la thrombose veineuse ainsi que de l'interaction entre ces facteurs génétiques et la fonction rénale. Le but de ces travaux et d'apporter de nouveaux éléments permettant de caractériser le rôle de la fonction rénale dans le développement des maladies cardiovasculaires. À l'aide d'une approche GWAS je cherche à identifier des marqueurs génétiques associés au risque de coronaropathie dans un groupe cas-témoin composé de patients atteints de diabète de type I et d'étudier l'effet de l'interaction entre ces marqueurs génétiques et la néphropathie diabétique sur le risque de coronaropathie. Ensuite, j'étudie le rôle de la fonction rénale dans le risque de thrombose veineuse en rassemblant des marqueurs génétiques de la fonction rénale dans un score polygénique. Ces travaux ont permis de mettre en évidence des gènes candidats associés au risque de coronaropathie chez les patients diabétiques de type I qui sont impliqués dans le développement du rein et la fonction rénale et l'association entre le score polygénique et le risque de thrombose veineuse a permis de renforcer l'hypothèse de la fonction rénale en tant que facteur de risque de la thrombose veineuse. Les résultats obtenus dans ces travaux doivent être confirmés mais dans l'ensemble ils montrent que la fonction rénale joue un rôle dans le développement des maladies cardiovasculaires.

Mot-clés : Maladies cardiovasculaires, diabète de type I, fonction rénale, génomique, études d'association pan-génomiques, score polygénétique.

Title : Study of interactions between renal function, arterial thrombosis and venous thrombosis using a genetic approach

Abstract : This thesis deals with the study of genetic factors of coronary artery diseases as complications of type I diabetes, genetics factors of venous thrombosis and the interactions between these genetic factors and renal function. This work aims to add new pieces of knowledge concerning the role of renal function in the development of cardiovascular diseases. From a GWAS approach I aim to identify genetic markers associated with coronary artery disease within a group of type I diabetic patients a study the effect of the interaction between these markers and diabetic nephropathy. Then, I will study the link between renal function and venous thrombosis using a polygenetic risk score composed from genetic markers of renal function. This work highlighted new candidate genes associated with coronary artery disease among type I diabetic patients that are involved in kidney development and renal function. The association between the polygenetic risk score and renal function strengthened the hypothesis of renal function as a risk factor for venous thrombosis. The obtained results must be confirmed but they globally demonstrate that renal function plays a role in the development of cardiovascular diseases.

Key-words : cardiovascular diseases, type I diabetes, renal function, genomics, genome-wide association studies, polygenetic score.