UNIVERSITE PARIS DESCARTES

ECOLE DOCTORALE BIOLOGIE SORBONNE PARIS CITE

Année 2016 N°

THESE DE DOCTORAT

Pour obtenir le grade de

DOCTEUR DE L'UNIVERSITE PARIS DESCARTES

Discipline : Génétique

Présentée et soutenue publiquement par

Soto Romuald KIANDO

BASES GENETIQUES DE LA DYSPLASIE FIBROMUSCULAIRE : UNE APPROCHE D'ETUDE D'EXOME ET DE GENETIQUE EPIDEMIOLOGIQUE

Le 8 Juillet 2016

Devant un jury composé de :

Pr	Stéphanie DEBETTE	Rapporteur
Dr	Emmanuelle GENIN	Rapporteur
Pr	Xavier JEUNEMAITRE	Président
Pr	Alexandre PERSU	Examinateur
Dr	Hervé PERDRY	Examinateur
Dr	Nabila BOUATIA-NAJI	Directrice de thèse

A ma famille
Mon Colonel, parti trop tôt
Ma Mounette, toujours là pour moi
Nihad et Ilan, le présent et l'avenir

REMERCIEMENT

Je tiens à remercier chaleureusement l'ensemble des membres de mon jury, et tout particulièrement mes rapporteurs Stéphanie Debette et Emmanuelle Genin. Je suis particulièrement heureux d'avoir pu constituer un jury de votre qualité et c'est un véritable honneur de présenter devant vous mon travail de thèse.

Je remercie aussi ma directrice de thèse, Nabila Bouatia-Naji. Je peux enfin vous tutoyer pour *te dire tout simplement merci*. Merci pour m'avoir confié ce sujet de thèse passionnant et donné les moyens de le conduire sereinement. J'ai beaucoup appris à tes côtés, aussi bien scientifiquement que humainement. Merci pour ta confiance, ta rigueur et ton exigence scientifique indispensables pour la bonne conduite de notre projet. Merci aussi pour ton soutien et tes encouragements, mais aussi tes conseils avisés pendant mes moments de doutes. J'espère avoir été à la hauteur de tes attentes.

Ma reconnaissance va aussi à l'endroit du Pr Xavier Jeunemaitre, pour m'avoir chaleureusement accueilli au sein de son équipe et largement conseillé tout au long de ces années. C'est aussi l'occasion de remercier tous les membres actuels ou anciens de l'équipe 3 gènes et maladies artérielle rares avec qui j'ai toujours eu du plaisir à travailler.

Merci également à l'ensemble des équipes du premier étage du PARCC, aux personnels de la plate-forme administrative, des départements de génétique, d'hypertension et du centre de référence des maladies vasculaires rares de l'HEGP.

Merci aussi à tous mes ami (e)s, sur qui j'ai toujours pu compter.

J'aimerais aussi remercier l'ensemble de ma famille aussi bien au Bénin qu'en France. Merci pour votre soutien et vos encouragements durant toutes ces longues années d'études. Je n'y serais pas arrivé sans vous. Encore une fois Merci.

Merci aussi à tous ceux qui de près ou de loin ont participé d'une manière ou d'une autre à la réussite de ce projet.

Enfin, un grand merci à l'ensemble des malades de la DFM sans qui ce travail n'aurait pu être possible.

Ce travail de thèse a été financé par l'Agence Nationale de la Recherche et dirigé par le

Dr Nabila BOUATIA-NAJI ;

Au sein de l'équipe 3 « gènes et maladies artérielles rares » du

Pr Xavier JEUNEMAITRE ;

Au sein de l'unité INSERM U970 (PARCC-HEGP) dirigée par le

Dr Alain TEDGUI.

RESUME

La dysplasie fibromusculaire artérielle (DFM) est un groupe de pathologies vasculaires non inflammatoires, et non athéromateuses de la paroi artérielle. Elle est caractérisée par la sténose, l'occlusion, l'anévrisme ou la dissection des artères de petit et moyen calibres, en particulier les artères rénales et le tronc supra-aortique. La DFM est un facteur de risque de l'hypertension et de l'accident vasculaire cérébral. Elle touche essentiellement les femmes (80% des cas) de moins de 50 ans. La prévalence en population générale est inconnue et les estimations varient de 0.4% pour les formes cliniques à 4% dans une cohorte de donneurs de reins. Une agrégation familiale a été démontrée et une composante génétique suggérée. L'objectif de mon travail de thèse était de caractériser les bases génétiques la DFM.

Dans la première partie, nous avons analysé des variants génétiques rares générés par séquençage d'exomes chez 16 cas apparentés de DFM issus de 7 fratries. Aucun gène majeur n'était muté pour l'ensemble des fratries ou pour au moins 3 fratries sur 7. Cependant, nous avons pu mettre en évidence puis validé un enrichissement en variants rares à fort potentiel fonctionnel de quatre gènes candidats pour la DFM (*MYLK*, *OBSCN*, *DYNC2H1*, *RNF213*) en combinant l'approche de séquençage d'exomes et l'étude d'association gène entier de 62767 variants rares (MAF < 5%) générés par génotypage avec la puce *Exome-chip* chez 249 cas non apparentés de DFM et 689 témoins. Cependant, l'implication de ces gènes dans la DFM doit être confirmée dans d'autres familles, et par des études de validations fonctionnelles.

Dans la seconde partie, nous avons étudié l'association avec la DFM de 25606 variants fréquents (MAF ≥ 5%) de l'*Exome-chip*. Les résultats majeurs obtenus ont été répliqués dans une première étude (402 cas de DFM et 2537 témoins) puis dans 3 autres études incluant 512 cas de DFM et 669 témoins. La méta-analyse de l'ensemble a permis d'associer à la DFM le polymorphisme rs9349379-A situé dans l'intron du gène *PHACTR1* (OR=1,39 [1,39-1,54] ; P=7,36 ×10^{-10}). Ce variant est aussi un facteur de risque pour la maladie coronaire, la migraine et la dissection de l'artère cervicale. Des études complémentaires conduites chez 2458 volontaires non malades ont permis de montrer que l'allèle à risque pour la DFM, rs9349379-A est associé avec une augmentation de l'épaisseur intima média (P=1,97×10^{-4}) et du rapport de la paroi sur la lumière artérielle (P=0,002), deux paramètres décrits comme augmentés chez les cas de DFM dans des études antérieures. Ensuite, PHACTR1 a été détecté par immunohistochimie dans l'endothélium et les cellules musculaires lisses de carotides dysplasiques et non dysplasiques avec une expression augmentée de PHACTR1 pour les porteurs de l'allèle à risque de DFM dans des cultures primaires de fibroblastes humains (N=86, P=0,003). Enfin, l'invalidation de *Phactr1* chez le poisson zèbre conduit à une dilatation des vaisseaux indiquant un défaut du développement vasculaire.

Ce travail confirme le caractère multifactoriel et hétérogène de la DFM et ouvre de nouvelles perspectives pour évaluer l'ensemble de la variabilité génomique des patients de DFM par des approches massives de génétique épidémiologique.

Mots clés : Dysplasie fibromusculaire artérielle ; séquençage d'exome ; association de variants rares et fréquents ; *Exome-chip* ; *GWAS* ; *SKAT* ; *PHACTR1*.

ABSTRACT

Fibromuscular dysplasia (FMD) is a group of nonatherosclerotic and noninflammatory vascular diseases leading to stenosis, aneurysm, dissection and/or occlusion of medium-sized arteries, in particular the renal and extracranial cervical arteries. Clinical manifestations of FMD are hypertension, dizziness, pulsatile tinnitus, transient ischemic attack or stroke, according to the involved arterial beds. FMD occurs predominantly (80% of cases) in females under 50 years with a variable prevalence estimation from 0.4% for asymptomatic clinical relevant forms to 4% in potential renal donors. The pathogenesis of FMD is unknown and a genetic origin is suspected given its demonstrated familial aggregation. The aim of my thesis work was to characterize genetic basis of FMD.

In the first part of this thesis, we analyzed whole exome sequencing data in 16 related FMD cases from seven families. No gene harbors variants that were shared by all affected members in at least three out seven families. Using combined strategy of whole exome sequencing and gene based association study of 62,767 rare variants (MAF < 5%) generated by Exome-chip arrays in 249 unrelated FMD cases and 689 controls, we have identified and validated an enrichment of rare and putatively functional variants in four candidates genes (*MYLK*, *OBSCN*, *DYNC2H1* and *RNF213*). This results need to be validated in other FMD families and by functional analysis.

In the second part, we analyzed 25,606 common variants (MAF ≥ 5%) generated by Exome-chip array. Top loci were replicated in first replication study (402 cases and 2,537 controls) and in 3 others studies (512 cases and 669 controls). Meta-analysis of all including 1,154 unrelated FMD cases and 3,895 controls allowed identification of association between FMD and rs9349379-A (OR=1.39 [1.39-1.54]; $P=7.4\times10^{-10}$). rs9349379 is intronic to *PHACTR1*, a risk locus for coronary artery disease, migraine, and cervical artery dissection. The analyses of geometrical parameters of carotids from 2,458 healthy volunteers indicated higher intima media thickness ($P = 1.97\times10^{-4}$) and wall to lumen ratio ($P = 0.002$) in rs9349379-A carriers, suggesting indices of carotid hypertrophy as previously described in carotids of FMD patients. Immunohistochemistry detected PHACTR1 in endothelium and smooth muscle cells of FMD and normal human carotids. The expression of PHACTR1 by genotypes in primary human fibroblasts showed higher expression in rs9349379-A carriers (N=86, $P=0.003$). Phactr1 knockdown in zebrafish resulted in dilated vessels indicating subtle impairment of vascular development.

This work confirms the multifactorial and heterogeneous genetic architecture of the FMD and opens new opportunities to evaluate all of genomic variability of FMD patients with massive genetic epidemiology approaches.

Keywords: Fibromuscular dysplasia; FMD; exome sequencing; association of rare and common variants, Exome-chip, GWAS, SKAT, *PHACTR1*.

TABLE DES MATIERES

REMERCIEMENT 3
RESUME 5
ABSTRACT 6
TABLE DES MATIERES 7
LISTE DES PUBLICATIONS 10
Publications issues directement du travail de thèse 10
Publications issues de collaboration au sein du laboratoire de thèse 11
Communications Orales 12
INDEX DES ILLUSTRATIONS 13
INDEX DES TABLES 14
LISTE DES PRINCIPALES ABBREVIATIONS 15
PREMIERE PARTIE : INTRODUCTION GENERALE 17
CHAPITRE I : METHODOLOGIE DE L'ETUDE GENETIQUE DES MALADIES HUMAINES 18
I.1. Structure du génome humain 18
I.2. Mécanismes de régulation de l'expression du gène 19
I.3. Variabilité du génome humain 21
I.3.1. Les séquences répétées en tandem 22
I.3.2. Les variations de nombre de copies ou *CNVs* 23
I.3.3. Les polymorphismes mono-nucléotidiques ou *SNPs* 24
I.4. Génétique des maladies humaines 25
I.4.1. Pénétrance et expressivité variable 26
I.4.2. Les maladies mendéliennes 28
I.4.2.1. Mode de transmission des maladies monogéniques 29
I.4.3. Les maladies polygéniques dites à génétique complexes 31
I.5. Identification de la composante génétique des maladies humaines 32
I.5.1. Analyse de liaison génétique 32
I.5.2. Succès limité des analyses de liaison pour les maladies complexes 34
I.5.3. Etude d'association génétique 35
I.5.3.1. Mesure du risque du variant associé pour les études d'association 36
I.5.3.2. Le Déséquilibre de liaison 38
I.5.3.3. Etude d'association de gènes candidats 40
I.5.3.4. Etudes d'association pangénomiques (*GWAS*) 40
I.5.3.5. Méthodologie des *GWAS* 43

I.5.3.6. *GWAS* : du polymorphisme de prédisposition à la fonction 45
I.5.3.7. Limite des *GWAS* : héritabilité manquante 46
I.5.4. Séquençage nouvelle génération (*Next Generation Sequencing*) 49
I.5.4.1. Technologie du NGS 49
I.5.4.2. Pipeline bio-informatique 51
I.5.5. Etude génétique des maladies humaines par NGS 52
I.5.5.1. Le séquençage d'exome 52
I.5.5.2. Promesses du WES 54
I.5.5.3. Les bases de données de variants 56
I.5.6. Analyse des données de séquençage d'exome 58
I.5.6.1. Analyse du WES dans le cadre de maladies monogénique 59
I.5.6.2. WES et maladies complexes 64
I.5.6.3. Confirmation et validation de la mutation causale 67
CHAPITRE 2 : ETAT DES CONNAISSANCES SUR LA DYSPLASIE FIBROMUSCULAIRE ARTERIELLE (DFM) 69
II.1. DFM : Définition 69
II.2. DFM : Perspectives historiques 70
II.3. DFM : Classification 72
II.3.1. Classification histo-pathologique 73
II.3.2. Corrélation histo-angiographique 75
II.3.3. Classification angiographique binaire 76
II.4. DFM : Diagnostic par imagerie 78
II.4.1. Diagnostic de la DFM des artères rénales 78
II.4.2. Diagnostic de la DFM des artères cervico-encéphaliques 79
II.4.3. L'angiographie classique par cathétérisation pour le diagnostic de la DFM 80
II.5. DFM : Manifestations cliniques 81
II.5.1. Tableau clinique de la DFM rénale 82
II.5.2. Tableau clinique de la DFM cervico-cérébrale 83
II.5.3. Complications de la DFM 84
II.5.3.1. Anévrismes artériels associés à la DFM 84
II.5.3.2. Dissections artérielles associés à la DFM 85
II.5.4. DFM des artères coronaires 86
II.6. DFM : Pathologies associées à la DFM et diagnostic différentiel 89
II.6.1. DFM et athérosclérose 89
II.6.2. DFM et vascularite 89
II.6.3. DFM et autres pathologies associées 90
II.7. DFM : Aspects épidémiologiques 91

II.7.1. Prévalence de la DFM des artères rénales 91
II.7.2. Prévalence de la DFM des artères cervico-cérébrales. 92
II.7.3. Facteurs de risques connus pour la DFM 93
II.8. DFM : Physiopathologie et étiologie de la DFM 94
II.8.1. Facteurs hormonaux et mécaniques pour la DFM 94
II.8.2. La DFM est-elle une maladie génétique? 95
II.9. DFM : Génétique de la DFM 96
II.9.1. La DFM : une maladie génétique difficile à étudier 97
II.9.2. Etude gènes-candidats de la DFM 98
OBJECTIFS DE LA THESE 100
DEUXIEME PARTIE : TRAVAUX ORIGINAUX 102
CHAPITRE III : L'ETUDE DE VARIANTS GENETIQUES RARES SUGGERE L'ABSENCE D'UN GENE MAJEUR POUR LA DFM 103
Article 1 103
III.1. Position du problème 103
III.2. Résultats majeurs 104
CHAPITRE IV : IDENTIFICATION D'UN PREMIER GENE DE PREDISPOSITION A LA DYSPLASIE FIBROMUSCULAIRE 107
Article 2 107
IV.1. Position du problème 107
IV.2. Résultats majeurs 108
TROISIEME PARTIE : DISCUSSION ET PERSPECTIVES 111
CHAPITRE V : DISCUSSION 112
V.1. WES : un outil pour l'identification des gènes de prédisposition à la DFM 112
V.2. Etude d'association gène-entiers de variants rares 114
V.3. Etude d'association des variants fréquents de l'*Exome-chip* 116
V.3.1. *PHACTR1* un locus majeur pour les pathologies vasculaires 116
V.3.2. Du polymorphisme à l'expression du gène 119
V.3.3. Du locus au modèle animal 120
CHAPITRE VI : PERSPECTIVES 123
VI.1. Stratégies de reséquençage ciblé de loci candidats 123
VI.2. Analyse de voies biologiques candidates 124
VI.3. Etude d'association pangénomique de de la DFM 125
VI.4. Autres approches pour l'étude génétique de la DFM 126
CONCLUSION GENERALE 129
QUATRIEME PARTIE : BIBLIOGRAPHIE 132

LISTE DES PUBLICATIONS

Publications issues directement du travail de thèse

Exome sequencing in seven families and gene-based association studies indicate genetic heterogeneity and suggest possible candidates for fibromuscular dysplasia. **Kiando SR**, Barlassina C, Cusi D, Galan P, Lathrop M, Plouin PF, Jeunemaitre X, Bouatia-Naji N. **J Hypertens**. 2015 Sep;33(9):1802-10; discussion 1810. doi: 0.1097/HJH.0000000000000625.

PHACTR1 Is a Genetic Susceptibility Locus f 1 or Fibromuscular Dysplasia. **Soto Romuald Kiando SR**, Nathan R. Tucker, Luis-Jaime Castro-Vega, Alexander Katz, Valentina D'Escamard, Cyrielle Tréard, Daniel Fraher, Juliette Albuisson, Daniella Kadian-Dodov, Zi Ye, Erin Austin, Min-Lee Yang, Kristina Hunker, Cristina Barlassina,Daniele Cusi, Pilar Galan, Jean-Philippe Empana, Xavier Jouven, Anne-Paule Gimenez-Roqueplo, Patrick Bruneval, Esther Soo Hyun Kim, Jeffrey W. Olin, Heather L. Gornik, Michel Azizi1, Pierre-François Plouin, Patrick T. Ellinor, Iftikhar J. Kullo, David J. Milan, Santhi K. Ganesh, Pierre Boutouyrie, Jason C. Kovacic, Xavier Jeunemaitre, Nabila Bouatia-Naji. Article en révision **Plos Genetics** Mai 2016

Publications issues de collaboration au sein du laboratoire de thèse

The *MITF*, p.E318K variant as a risk factor for pheochromocytomas and paragangliomas. Luis Jaime Castro-Vega, **Soto Romuald Kiando**, Nelly Burnichon, Alexandre Buffet, Laurence Amar, Christophe Simian, Amandine Berdelou, Pilar Galan, Martin Schlumberger, Nabila Bouatia-Naji, Judith Favier, Brigitte Bressac-de Paillerets, and Anne-Paule Gimenez-Roqueplo. Article soumis à **J Clin Endocrinol Metab**. Mai 2016

Familial hypocalciuric hypercalcemia types 1 and 3 and primary hyperparathyroidism: similarities and differences.Vargas-Poussou R, Mansour-Hendili L, Baron S, Bertocchio JP, Travers C, Simian C, Treard C, Baudouin V, Beltran S, Broux F, Camard O, Cloarec S, Cormier C, Debussche X, Dubosclard E, Eid C, Haymann JP, **Soto Romuald Kiando**, Kuhn JM, Lefort G, Linglart A, Lucas-Pouliquen B, Macher MA, Maruani G, Ouzounian S, Polak M, Requeda E, Robier D, Silve C, Souberbielle JC, Tack I, Vezzosi D, Jeunemaitre X, Houillier P. **J Clin Endocrinol Metab**. 2016

Investigation of the Matrix Metalloproteinase-2 Gene in Patients with Non- Syndromic Mitral Valve Prolapse. Maëlle Perrocheau, **Soto Romuald Kiando**, Déwi Vernerey, Christian Dina, Pilar Galan, Albert Hagege, Xavier Jeunemaitre and Nabila Bouatia-Naj. **J. Cardiovasc. Dev. Dis**. 2015.

Communications Orales

Common genetic variant in PHACTR1 associates with fibromuscular dysplasia. 25th **International Meeting of the French Society of Hypertension. Paris, Dec 17-18, 2015**

Exome Sequencing in Seven Families and Gene-Based Association Studies Support Genetic Heterogeneity and Suggest Possible Candidates for Fibromuscular Dysplasia. **25th European Meeting on Hypertension and Cardiovascular Protection. Milan, JUNE 12-15, 2015**

Exome Sequencing in Seven Families and Gene-Based Association Studies Support Genetic Heterogeneity and Suggest Possible Candidates for Fibromuscular Dysplasia. **8th International Meeting of the French Society of Hypertension. Paris, Dec 18-19, 2014**

Association pangénomique de variants génétiques codants et fréquents avec la dysplasie fibromusculaire. **7èmes Assises de Génétique Humaine et Médicale, Bordeaux, 29-31 janvier 2014**

INDEX DES ILLUSTRATIONS

Figure I-1. Structure de base du gène. 19

Figure I-2. Mécanismes de cis-régulation de l'expression des gènes. 21

Figure I-3. Différents types de variants génétiques du génome humain. 25

Figure I-4. Facteurs modificateurs de la pénétrance dans les maladies humaines.. 27

Figure I-5. Pénétrance et fréquence allélique dans les maladies humaines.. 28

Figure I-6. SNPs, haplotypes et TagSNPs. 39

Figure I-7. Polymorphismes identifiés par GWAS.. 41

Figure I-8. Principes de l'assignation de génotypes par puce à ADN. 42

Figure I-9. Étapes du séquençage NGS. 50

Figure I-10. Publications portant sur l'étude de traits ou pathologies humaines utilisant les approches d'études de liaison, de GWAS et de WES/WGS. 55

Figure II-11. Lésions dysplasiques de l'artère rénale 70

Figure II-12. Nombre de publications référencées dans PubMed pour la recherche du terme "Fibromuscular dysplasia/hyperplasia" dans le titre et/ou le résumé. 72

Figure II-13. Classification histo-pathologique de la DFM. 74

Figure II-14. Diagramme illustrant la stratégie diagnostique de la DFM des artères rénales et cervico encéphaliques par des techniques d'imagerie médicale. 81

Figure II-15. Aspects angiographiques et cliniques de la DFM. 88

Figure VI-16. Application du NGS pour l'identification de la composante génétique des maladies humaines. 128

INDEX DES TABLES

Tableau III-1. Etude d'association gène-entier de variants rares avec la DFM. 106

Tableau VI-2. Association du polymorphisme rs9349379 avec la DFM dans cinq études indépendantes. 109

LISTE DES PRINCIPALES ABBREVIATIONS

ADN	Acide Désoxyribonucléique
angio-TDM	Angiotomodensitométrie
ARM	Angiographie par résonance magnétique
BP	Paire de bases
CAD-MI	la maladie coronaire et l'infarctus du myocarde (CAD-MI)
CeAD	Dissection de l'artère cervicale
CNVs	Variations du nombre de copies
DFM	Dysplasie fibromusculaire
EVS	Exome variants server
EWAS	Etude d'association de variants codants
GWAS	Etude d'association pangénomique
HAS	Haute Autorité de Santé en France
HEGP	Hôpital Européen Georges-Pompidou
HTA	Hypertension artérielle
KO	Modèle mutants knock-out
LD	Déséquilibre de liaison
LOD	Logarithm Of the Odds
MAF	Fréquence de l'allèle mineur
NCBI	*National Center for Biotechnology Information*
NGS	Séquençage nouvelle génération
NPL	Non parametric linkage
OMIM	Base de données de référence pour les maladies mendéliennes

OR	Odds ratio
PAD	Pression artérielle diastolique
PAM	Pression artérielle moyenne
PAS	Pression artérielle systolique
PCR	Réaction de polymérisation en chaîne
PROFILE	PROgression of FIbromuscular LEsions
QC	Contrôle qualité
SED	Syndrome d'Ehlers-Danlos
SNP	Polymorphisme mono-nucléotidique
TDT	Test de déséquilibre de transmission
TGF-β	Facteur de croissance transformant
TSA	Troncs supra-aortiques
VEGF	Voie du facteur de croissance de l'endothélium vasculaire
VNTR	Séquences répétées en tandem
WES	Séquençage d'exome nouvelle génération
WGS	Séquençage du génome nouvelle génération

PREMIERE PARTIE : INTRODUCTION GENERALE

CHAPITRE I : METHODOLOGIE DE L'ETUDE GENETIQUE DES MALADIES HUMAINES

Dans ce chapitre, nous présenterons quelques concepts généraux relatifs à la génétique humaine et à l'étude génétique des pathologies. Après un bref rappel des méthodes historiques restreintes à l'étude des maladies monogéniques, un accent particulier sera mis sur les différentes stratégies d'identification des variants génétiques de prédisposition aux maladies humaines à l'ère du génotypage et du séquençage haut-débit.

I.1. Structure du génome humain

Le génome est l'ensemble du matériel génétique d'un individu codé dans son acide désoxyribonucléique (ADN). L'ADN nucléaire humain est diploïde et divisé en 23 paires de chromosomes homologues. Les 22 premières paires sont appelées autosomes et numérotées de 1 à 22. La dernière paire consiste en deux chromosomes sexuels : deux chromosomes X chez la femme et un chromosome X et un chromosome Y chez l'homme. La fonction principale du chromosome est de porter le gène qui est l'unité fonctionnelle de base du génome. On estime que le génome humain contient environ 20805 gènes (Ensembl 2016). La distance entre chaque locus (ou position physique précise et invariable du génome) peut être mesurée en paires de base (bp, également kb = 1000 bp et Mb = 10^6 bp). L'ADN nucléaire est composé d'environ trois milliards de paires de bases (bp) et constitue avec l'ADN mitochondrial le génome humain.

Les séquences codantes de gènes, appelées exons (entre 1 et 1.5% du génome humain), sont interrompues par des séquences intermédiaires non codantes, les introns (Lander, Linton et al. 2001). En plus des exons et des introns, les gènes sont constitués d'un certain nombre d'éléments essentiels pour leur expression. Il s'agit d'une séquence promotrice qui se lie à une

enzyme ARN polymérase ; d'un site d'initiation de la transcription ; d'un site d'initiation de traduction ; d'une région non traduite en 5 ' ; d'un codon de terminaison de traduction ; et enfin d'une région non traduite en 3' [cf. Figure I-1].

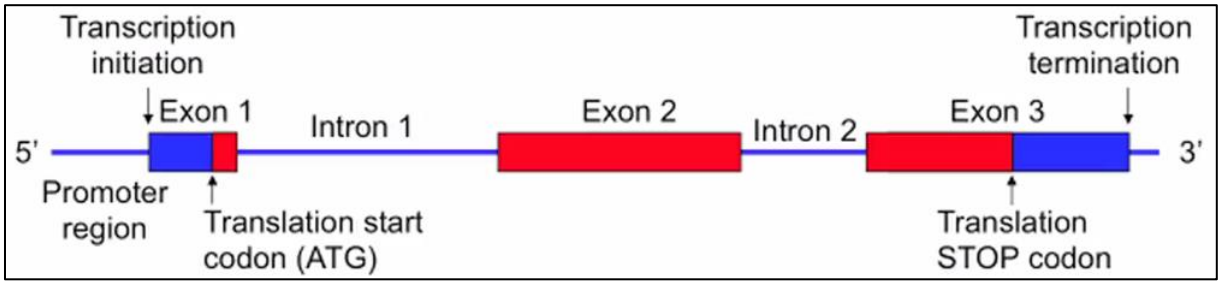


Figure I-1. Structure de base du gène. La structure de base du gène est constituée de séquences codantes appelées exons (entre 1 et 1.5% du génome humain), interrompues par des séquences intermédiaires non codantes, les introns.

I.2. Mécanismes de régulation de l'expression du gène

Le dogme de la biologie moléculaire définissant le gène comme l'unité fonctionnelle centrale de l'hérédité est sans arrêt mis en cause. Il est de plus en plus hasardeux de restreindre la fonction codante du génome uniquement à l'exome compte tenu de la meilleure caractérisation des séquences non codantes du génome par des projets comme ENCODE qui attribue un rôle fonctionnel et/ou de régulation de l'expression des gènes à près de 80% du génome entier (Dunham, Kundaje et al. 2012). Aussi, l'expression du gène est un processus remarquablement complexe régulé à plusieurs niveaux par des séquences *cis* et *trans*-régulatrices (Lemon and Tjian 2000).

Les séquences cis-régulatrices sont principalement des séquences promotrices situées dans la même molécule d'ADN que leurs gènes cibles (Noonan and McCallion 2010). Il s'agit des amplificateurs (*enhancers*), des répresseurs (*repressors*) et des isolateurs (*insulators*) pouvant être situés en amont, en aval ou dans les introns des gènes qu'ils régulent ; ou encore jusqu'à 1 Mb de leurs gènes cibles dépendant ainsi d'une structure appropriée de la chromatine pour leur fonction normale (Kleinjan and van Heyningen 2005). Les amplificateurs et les répresseurs régulent respectivement l'expression du gène de façon positive et négative. Les isolateurs quant à eux, établissent des barrières entre deux gènes ou un groupe de gènes les protégeant ainsi des effets produits par les autres séquences régulatrices et des éventuelles interactions entre elles [cf. Figure I-2].

Les éléments *trans*-régulateurs ont la particularité de réguler l'expression du gène cible à distance. Il s'agit principalement des microARNs (*miARNs*) qui sont des facteurs de régulation post-transcriptionnels capables d'extinction de l'expression d'un gène (Kusenda, Mraz et al. 2006, Bartel 2009). Il en a déjà été identifié près de 700 (Kozomara and Griffiths-Jones 2014) a chacun desquels les outils de prédiction attribuent à chacun un rôle de régulation sur l'expression d'environ 100 à 200 gènes (Krek, Grun et al. 2005).

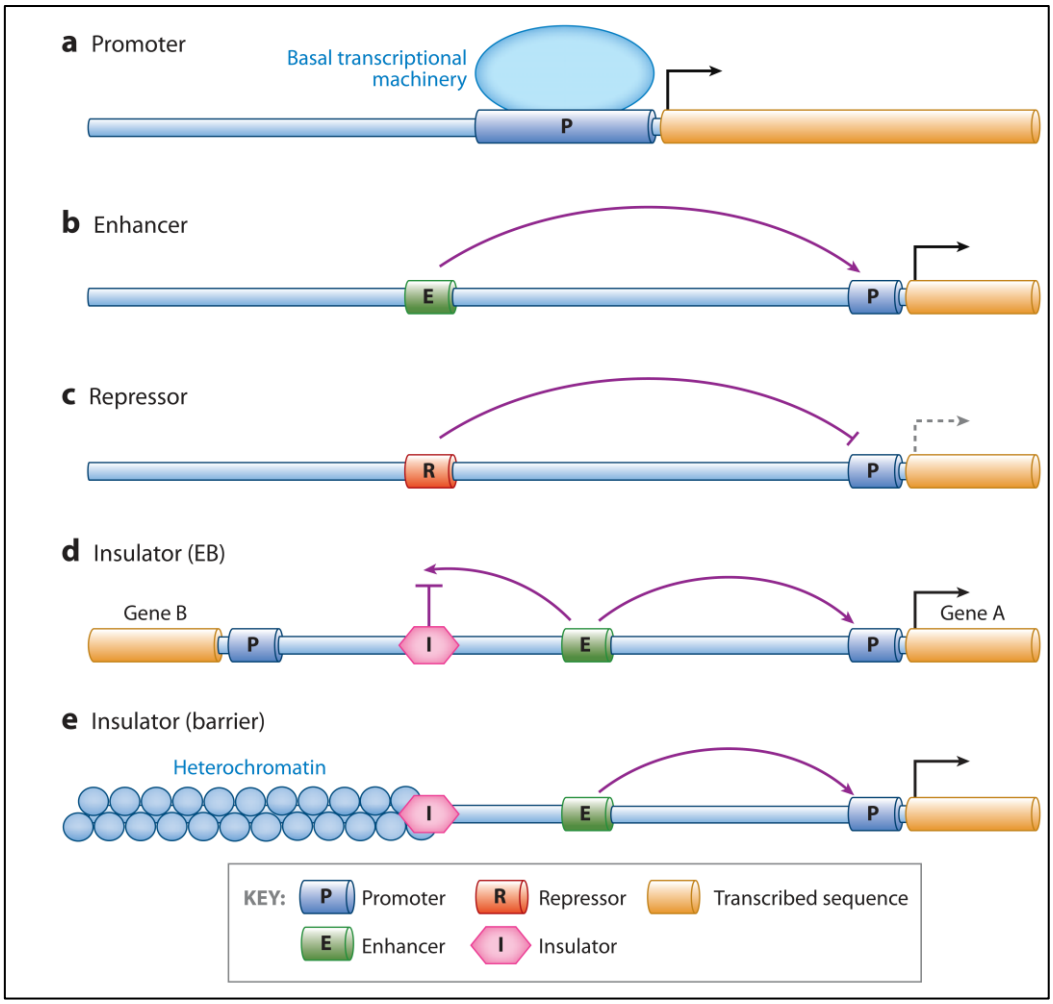


Figure I-2. Mécanismes de cis-régulation de l'expression des gènes. L'expression du gène est un processus remarquablement complexe régulé à plusieurs niveaux par des séquences cis et trans-régulatrices. Illustration extraite de (Noonan and McCallion 2010).

I.3. Variabilité du génome humain

Les séquences nucléotidiques du génome diffèrent de 3,5 millions de bases entre deux individus non apparentés pris au hasard (soit 0.1% du génome) (Genomes Project, Abecasis et al. 2012). Ces variations interindividuelles de la séquence d'ADN qui vont de la modification d'une à plusieurs bases nucléotidiques occupent une localisation précise dans le génome (locus),

sont relativement stables et transmissibles d'une génération à l'autre. Toutes ces propriétés font donc de bons marqueurs pour cartographier le génome et étudier les maladies humaines.

Les différentes structures de la séquence d'ADN pour un variant génétique donné sont appelées allèles et on peut retrouver en un locus donné, une combinaison de 2 allèles (génotype) identiques (homozygote) ou différents (hétérozygote). Les variants génétiques peuvent être localisés aussi bien dans des régions codantes que non codantes du génome. L'effet de chaque variant sur le phénotype est très variable. La majorité des variants génétiques ont cependant un effet en apparence neutre sur le phénotype étudié (Wright 2005).

Il existe plusieurs types de variants génétiques pour lesquels de nombreuses classifications ont été proposées. Nous retiendrons celle qui discrimine les variants en fonction de leur composition nucléotidique [cf. Figure I-3] (Frazer, Murray et al. 2009).

I.3.1. Les séquences répétées en tandem

Les séquences répétées en tandem (VNTR) sont constituées de répétitions en tandem et en nombre variable *n* d'un motif unitaire de taille allant d'une paire de bases à plusieurs milliers de paires de bases. Chaque valeur de *n* définit un allèle. Selon la taille du motif et la répétition on distingue les satellites, les minisatellites et les microsatellites.

Les microsatellites correspondent le plus souvent à des motifs de moins de 9 nucléotides répétés 2 à 50 fois. Ils présentent plusieurs caractéristiques qui en font des bons marqueurs génétiques. Ils sont en effet très polymorphes, facilement détectables par la technique de réaction de polymérisation en chaîne (PCR), avec une distribution relativement homogène sur l'ensemble du génome. Les microsatellites quand ils sont situés dans les régions codantes du

génome peuvent altérer la fonction de la protéine en causant une mutation de type frameshift ou une altération de la longueur de la séquence d'acide aminé (Ellegren 2004). Ils ont été largement utilisés depuis leur identification pour la cartographie du génome (Weissenbach, Gyapay et al. 1992) et pour l'identification des gènes impliqués dans les maladies mendéliennes par les études de liaison génétique (Weissenbach 1993).

I.3.2. Les variations de nombre de copies ou *CNVs*

Les variations de nombre de copies (*CNVs*, *copy number variants*) sont définies comme des segments d'ADN d'une longueur supérieure ou égale à 1 kb qui présentent dans le génome un nombre de copies variables en comparaison avec un génome de référence. Les *CNVs* peuvent être classifiés en délétions, insertions ou duplications (Feuk, Carson et al. 2006). La première carte des *CNVs* du génome humain a été publiée en 2006 et a permis d'identifier 1 447 *CNVs* couvrant 360 Mb d'ADN, soit plus de 12% du génome humain (Redon, Ishikawa et al. 2006). La distinction entre les indels et les *CNVs* est arbitraire et basée sur la taille du segment. Toutefois, avec l'amélioration des techniques de génotypage et surtout de séquençage nouvelle génération (NGS), les *CNVs* sont de plus en plus identifiés et étudiés, y compris les plus petits (Duan, Zhang et al. 2013). Les *CNVs* peuvent modifier le nombre de copies du gène (Wain, Armour et al. 2009) et ont été impliqués dans de nombreuses pathologies (Beckmann, Estivill et al. 2007), aussi bien mendéliennes (Lupski, de Oca-Luna et al. 1991) que complexes (Gonzalez, Kulkarni et al. 2005, Wellcome Trust Case Control 2007).

I.3.3. Les polymorphismes mono-nucléotidiques ou *SNPs*

Les polymorphismes mono-nucléotidiques ou *SNPs* (*Single Nucleotide Polymorphism*) sont des variations ponctuelles d'un seul nucléotide dans la séquence d'ADN. On estime leur nombre à 38 millions (Sudmant, Rausch et al. 2015). Il s'agit du type de variant génétique le plus simple et le plus abondant du génome dont il représente 90% de la variabilité (Genomes Project, Abecasis et al. 2012). Les *SNPs* peuvent être localisés aussi bien dans des régions régulatrices ou codantes que non codantes du génome et peuvent donc avoir des conséquences fonctionnelles directes en modifiant par exemple la structure d'une protéine pour les *SNPs* situés dans une région codante ou son niveau d'expression pour ceux situés dans les régions régulatrices (Chorley, Wang et al. 2008).

Les régions codantes du génome sont caractérisées par une abondance de *SNPs* très rares (MAF < 0.5%) (Nelson, Wegmann et al. 2012). Les *SNPs* peuvent être non-synonymes, c'est à dire entrainant un changement d'acide aminé dans la séquence du gène, ou synonymes n'entrainant aucun changement dans la séquence protéique. Parmi les *SNPs* non-synonymes, on distingue aussi d'une part, les faux-sens, caractérisés par le changement d'un acide aminé par un autre et les non-sens qui sont caractérisés l'introduction prématurée d'un codon stop entrainant une production incomplète de la protéine.

Le terme utilisé pour décrire les *SNPs* dépend aussi de la fréquence du variant en population générale. De façon arbitraire, on évoque souvent le terme polymorphisme pour faire référence aux variants mono-nucléotidiques fréquents dans la population (MAF ≥1%) (Wright 2005). Les variants mono-nucléotidiques rares (MAF < 1%) quant à eux, évoquent généralement la notion de mutation, même si ce terme désigne un variant rare pour lesquels un lien de causalité a été établi avec une maladie. Du fait de leurs propriétés, et grâce aux progrès des techniques de biologie moléculaire permettant le génotypage simultané rapide et à faible

coût, les *SNPs* sont devenus les polymorphismes de l'ADN de prédilection pour l'étude du génome humain.

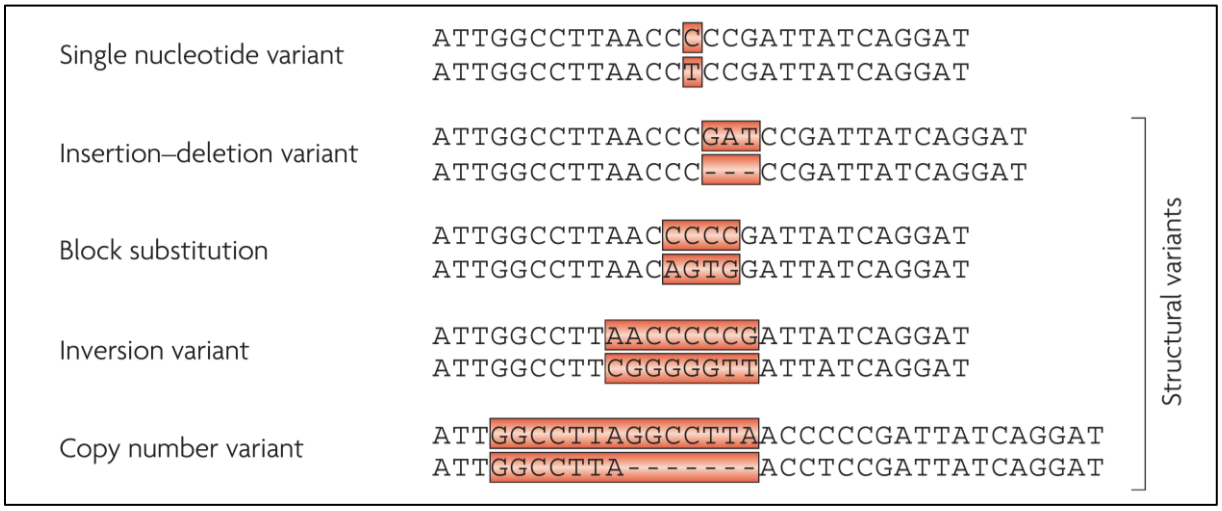


Figure I-3. Différents types de variants génétiques du génome humain. En fonction de la composition nucléotidique des variants, on peut distinguer les polymorphismes mono-nucléotidiques et les variations structurales. D'après Frazer et collaborateurs (Frazer, Murray et al. 2009).

I.4. Génétique des maladies humaines

La majorité des maladies humaines ont une composante génétique sous-jacente (Lander and Schork 1994). Cela signifie qu'il existe à un locus (maladie monogénique) ou à plusieurs loci (maladies oligogéniques et polygéniques) du génome, des variations de la séquence d'ADN pouvant directement prédisposer, augmenter ou diminuer la probabilité d'un individu de développer une maladie par rapport à un individu sans ces variations. De plus, il a été montré que les variations génétiques peuvent influencer le phénotype clinique et l'évolution des

maladies (Zlotogora 2003), la prédisposition aux agents infectieux (Frodsham and Hill 2004) et même la réponse aux traitements (Weinshilboum 2003). L'implication de la composante génétique dans les pathologies humaines varie donc considérablement.

I.4.1. Pénétrance et expressivité variable

La pénétrance peut être définie comme la proportion d'individus portant le variant à risque qui va exprimer le phénotype de la maladie. Dans la plupart des maladies monogéniques, l'hypothèse de base suppose que la présence de la mutation est à elle seule suffisante pour causer la maladie (Strachan 1999). Cela suppose donc une pénétrance complète. Mais bien souvent, la pénétrance est incomplète. Par conséquent, les individus qui portent la mutation peuvent ne pas développer la maladie. Plusieurs concepts sont connexes à la pénétrance avec des différences parfois subtiles. Il s'agit dans un premier temps de la phénocopie qui est la probabilité d'exprimer le phénotype sans porter le variant à risque, puis dans un second de l'expressivité qui mesure la variabilité phénotypique.

La pénétrance et l'expressivité variable dépendent de plusieurs facteurs [cf. Figure I-4]. Il a été suggéré l'implication de gènes modificateurs ainsi que des facteurs environnementaux et épigénétiques (Zlotogora 2003, Cooper, Krawczak et al. 2013) qui peuvent modifier, augmenter ou diminuer l'expression phénotypique du trait. La pénétrance peut aussi être âge-dépendant avec une prédisposition différente en fonction de l'âge (Cooper, Krawczak et al. 2013).

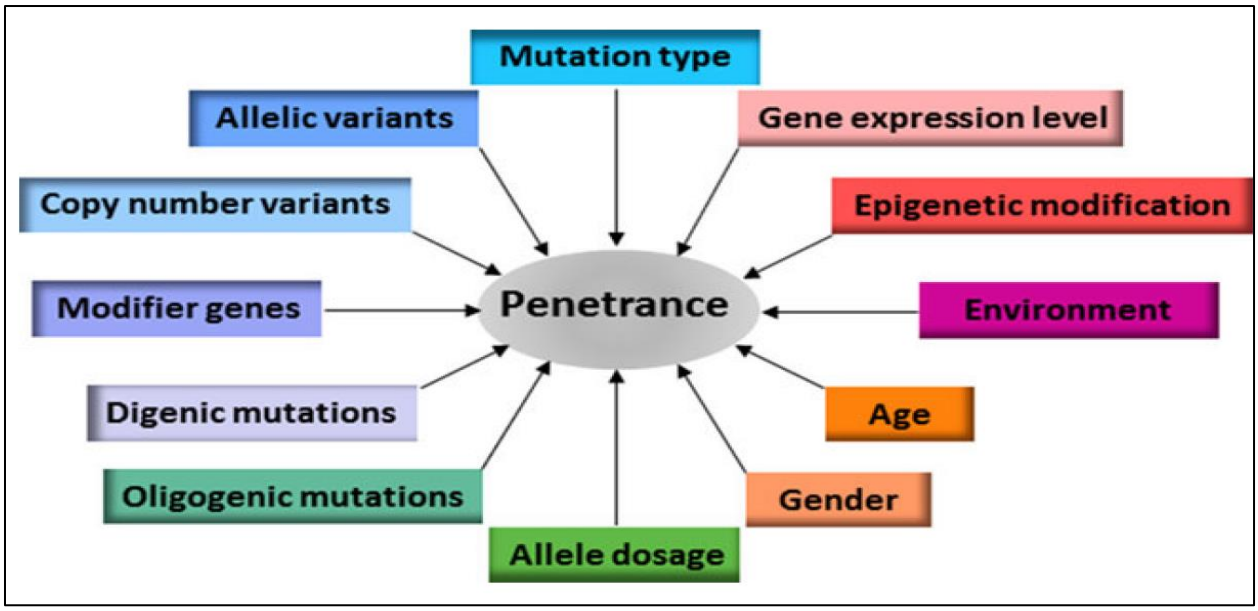


Figure I-4. Facteurs modificateurs de la pénétrance dans les maladies humaines. D'après Cooper et collaborateurs (Cooper, Krawczak et al. 2013).

Dans l'étude des maladies humaines, le spectre de la pénétrance des variants de prédisposition est important pour le modèle de l'étude (McCarthy, Abecasis et al. 2008) [cf. Figure I-5]. Ainsi, on distingue les variants à forte pénétrance, à pénétrance modérée et à faible pénétrance. Les variants à forte pénétrance ont tendance à être rares dans la population générale à cause de la forte pression de sélection qu'ils subissent. Inversement, les variants à faible pénétrance ne subissent pas de pression de sélection parce qu'ils ne contribuent pas seuls au phénotype de la maladie et ont tendance à être plus fréquents.

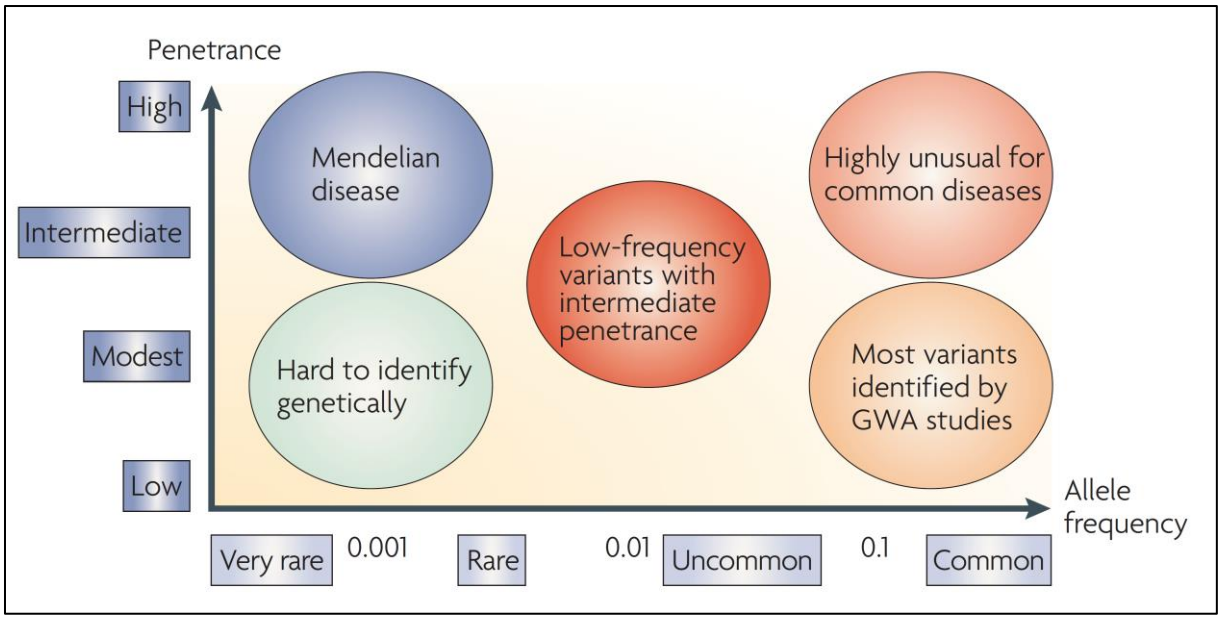


Figure I-5. Pénétrance et fréquence allélique dans les maladies humaines. Les maladies mendéliennes sont caractérisées par l'implication de mutations rares à forte pénétrance, contrairement aux maladies complexes où plusieurs variants fréquents de prédisposition peuvent être identifiés avec des effets modestes. D'après McCarthy et collaborateurs (McCarthy, Abecasis et al. 2008).

I.4.2. Les maladies mendéliennes

Une maladie mendélienne correspond au cas « simple » où la genèse de la maladie est causée par la présence d'une mutation dans un gène qui détermine à elle seule le phénotype. Pour éviter toute confusion avec d'autres maladies pour lesquelles les variants impliqués sont soumis à une hérédité mendélienne, on préfère le terme maladie monofactorielle ou maladie monogénique.

Les maladies monogéniques sont des maladies rares prises individuellement. Elles représentent cependant dans leur ensemble, des millions d'individus de par le monde (Carter

1977, Baird, Anderson et al. 1988) avec un coût important pour la société et les systèmes de santé (Yoon, Olney et al. 1997, McCandless, Brunger et al. 2004, Dye, Brameld et al. 2011). Le nombre de maladies monogéniques connues est estimé à environ 7.000 dont la moitié a une base génétique clairement identifiée (Boycott, Vanstone et al. 2013). Ce nombre est certainement sous-évalué et il y a probablement beaucoup plus de maladies monogéniques non encore identifiées et caractérisées parce qu'elles sont très rares ou surviennent dans des populations peu étudiées et n'ayant pas accès à un système de santé de qualité (Antonarakis and Beckmann 2006). La base de données de référence pour les maladies mendéliennes est sans conteste OMIM (Online Mendelian Inheritance in Man) née dans les années 1960 grâce au travail de Victor McKusick (McKusick 2007). Elle dresse un catalogue de toutes les maladies mendéliennes connues et les relie si possible aux gènes associés.

I.4.2.1. Mode de transmission des maladies monogéniques

Les maladies monogéniques peuvent résulter de mutations sporadiques (de novo) ou de variations dans la séquence d'ADN héritée d'un des parents ou des deux (Boycott, Vanstone et al. 2013). Le génotype au locus impliqué joue un rôle important. De plus, si une seule copie de l'allèle muté est suffisante pour développer la maladie, on parle d'effet dominant. Par contre, si deux copies de l'allèle muté sont nécessaires pour développer le phénotype, on parle d'effet récessif. Ainsi, selon l'effet de l'allèle muté et de sa localisation chromosomique (autosomique ou chromosomes sexuels), cinq différents modèles de transmission de maladies monogéniques sont reconnus. Il s'agit des transmissions autosomique (dominante et/ou récessive), gonosomique liée au chromosome X (dominante et récessive) et gonosomique liée au chromosome Y (Strachan 1999).

Dans le cas de maladies autosomiques dominantes, les individus présentent une histoire familiale de la maladie avec une hérédité parents enfants clairement établie. En cas de pénétrance incomplète, cette transmission héréditaire peut ne pas être effective et le mode autosomique dominant est suggéré par une agrégation familiale. Les mutations sous-jacentes peuvent être des mutations « gain de fonction » conférant un effet délétère dû à l'acquisition d'une nouvelle fonction du gène, ou des mutations de perte de fonction si le gène est haplo insuffisant (la synthèse de la protéine normale par l'allèle non muté ne suffit pas pour maintenir la fonction normale du gène).

Par contre, dans le cas de transmissions autosomiques récessives, peu d'individus sont atteints au sein d'une même famille car seuls ceux portant des mutations homozygotes ou hétérozygotes composites présentent la maladie. Les parents non malades sont porteurs chacun d'une copie de l'allèle muté pour la maladie avec la probabilité de ¼ d'avoir une descendance homozygote malade. Les mutations sont souvent de type perte de fonction et ont tendance à être rares dans la population générale (Gilissen, Hoischen et al. 2011), sauf dans certaines populations isolées où suite à un effet fondateur, la probabilité des individus de porter le variant rare causal peut augmenter par rapport à la population générale (Botstein and Risch 2003). Ainsi, plusieurs maladies autosomiques récessives surviennent dans des populations isolées ou dans des familles consanguines où les deux parents sont porteurs du même variant rare causal hérité du même ancêtre commun (Botstein and Risch 2003).

Les maladies monogéniques liées aux chromosomes sexuels présentent plusieurs particularités compte tenu de la variabilité chromosomique entre les sujets de sexe masculin qui n'ont qu'un seul chromosome X et de sexe féminin qui ont deux chromosomes X dont l'un est inactivé. Dans le cas de maladie monogénique liée au chromosome X, on peut avoir un mode de transmission dominant, ou récessif. L'hérédité liée au chromosome Y est très rare et les seuls

cas connus sont relatifs à la spermatogenèse au déterminisme du sexe (Osborne, Lynch et al. 2007).

L'identification de la mutation (gène) causale dans les maladies monogéniques se fait traditionnellement par le biais de familles, au sein desquelles on cherche des régions du génome co-ségrégant avec la maladie (Lander and Schork 1994).

I.4.3. Les maladies polygéniques dites à génétique complexe

On parle aussi de maladies fréquentes ou multifactorielles. Il s'agit de maladies pour lesquelles on observe l'implication de multiples variants génétiques à faible pénétrance et des facteurs environnementaux sans un mode de transmission héréditaire établi (Zwick, Cutler et al. 2000). Les maladies complexes sont plus fréquentes que les maladies monogéniques parce que chacune des variations sous-jacente prise individuellement n'a qu'un faible effet sur le phénotype et n'est donc pas soumise à la même pression de sélection négative que la mutation causale dans les maladies monogéniques (Blekhman, Man et al. 2008).

L'étude génétique des maladies complexes repose sur deux hypothèses : 1) l'hypothèse maladie fréquente-variants fréquents (CD-CV) qui stipule qu'il y aurait au sein d'un même individu de nombreux variants fréquents (MAF > 5%) qui participent de façon additive ou multiplicative à l'expression du phénotype (Reich and Lander 2001, Pritchard and Cox 2002) et 2) l'hypothèse maladie fréquente-variants rares (CD-RV) qui soutient pour sa part l'implication de multiples variants rares en population générale (MAF < 5%) comme facteur de risque aux maladies complexes (Pritchard 2001). Cependant, pour l'une comme l'autre de ces hypothèses, chacun de ces variants pris individuellement a un effet marginal sur le phénotype.

L'identification des variants génétiques de prédisposition aux maladies complexes est donc un véritable défi à cause du nombre important de variants impliqués attendu, ceux-ci pouvant interagir entre eux (épistasie) ou avec des facteurs environnementaux dont la participation au déterminisme de la maladie est très variable.

I.5. Identification de la composante génétique des maladies humaines

Il existe plusieurs stratégies pour caractériser la composante génétique des maladies humaines. Les analyses de liaison ont été performantes dans le cadre de maladies monogéniques. La disponibilité de la séquence complète du génome humain depuis 2004 (International Human Genome Sequencing 2004) et la constitution de base de données de variants génétique comme HapMap (International HapMap 2003) et 1000 Genomes (Genomes Project, Abecasis et al. 2010) ont révolutionné l'identification des variants génétiques de prédisposition aux maladies humaines grâce aux approches d'études d'association génétique puis de séquençage nouvelle génération (NGS).

I.5.1. Analyse de liaison génétique

L'analyse de liaison génétique consiste à étudier au sein de familles, la transmission d'une série de marqueurs polymorphes conjointement avec le locus de la maladie. On considère que deux loci sont liés s'ils sont transmis ensemble d'un parent à son enfant plus souvent qu'on ne l'attendrait en cas d'indépendance. Concrètement, si durant la méiose, une recombinaison entre deux loci survient à une fréquence inférieure à 50%, alors ces deux loci sont dits liés

(Dawn Teare and Barrett 2005). Une fois la région génomique d'intérêt détectée, l'identification de la mutation causale se fait par séquençage du gène.

Afin de tester la vraisemblance de la liaison entre deux ou plusieurs loci, le paramètre statistique LOD (Logarithm Of the Odds) score « Z-score » a été établi par Morton (Morton 1955). Il s'agit du logarithme décimal du rapport de vraisemblance entre l'hypothèse de liaison de deux loci (liaison à une fréquence de recombinaison « Θ » donnée), et celle de non liaison (correspondant à Θ = 0,5) (Morton 1955).

$$Z(\theta) = Log\ \frac{Probabilité\ de\ liaison\ (\theta = x)}{Probabilité\ de\ non\ liaison\ \ (\theta = 0.5)}$$

Traditionnellement, on conclut à une liaison quand le LOD score est supérieur à 3, c'est à dire quand la probabilité d'être lié est 1 000 fois plus grande que celle d'être non lié, et on rejette la liaison quand il est inférieur à -2 (la probabilité d'être non lié est 100 fois plus grande que celle d'être lié). Entre ces deux seuils, on ne peut tirer aucune conclusion et le recrutement de nouvelles familles est nécessaire pour améliorer la puissance statistique de l'analyse. Il s'agit là d'une des principales limites de cette approche modèle dépendante pour laquelle il faut définir au préalable la fréquence de l'allèle muté et les pénétrances des génotypes au locus de la maladie.

Aussi, plusieurs extensions du LOD score ont été proposées. Il s'agit entre autres du HLOD score proposé par Smith en 1963, permettant de prendre en compte l'hétérogénéité génétique au locus (Smith 1963), par l'introduction d'un paramètre supplémentaire pris en compte lors de l'analyse : le coefficient α, correspondant à la proportion de familles qui présentent la forme génétique de la maladie liée au marqueur d'intérêt. Des algorithmes multipoints ont été aussi proposés (Elston and Stewart 1971), permettant de prendre en compte la dépendance et les distances génétiques de quelques marqueurs adjacents afin de mieux

localiser les recombinaisons. Ils ont été aussi étendus à l'analyse de chromosomes entiers (Morton, MacLean et al. 1986), voire du génome entier et permettent désormais de réaliser l'analyse de liaison avec des *SNPs* contrairement aux première analyses de liaison conduites à partir de marqueurs microsatellites. Bien évidemment, avec l'avènement de ces études de liaison génome-entier et l'augmentation du nombre de marqueurs, il est apparu indispensable de prendre en compte la multiplicité des tests avec un seuil de LOD score désormais fixé à 3,3 (Lander and Kruglyak 1995).

Malgré la nécessité de spécifier un modèle génétique à priori et la nécessité de constituer des familles étendues de cas, les études d'analyses de liaison génétique ont connu des succès notables. Il s'agit entre autres de l'identification des gènes *BRCA1* et *BRCA2* mutés pour le cancer du sein (Miki, Swensen et al. 1994, Wooster, Neuhausen et al. 1994) et les mutations du gène *PARK2* chez les patients présentant une apparition précoce de la maladie de Parkinson (Matsumine, Saito et al. 1997). Plus récemment, grâce aux stratégies combinées avec les nouvelles techniques de séquençage (Bamshad, Ng et al. 2011), les approches d'analyse de liaison ont connu un véritable renouveau permettant à partir de l'étude de quelques familles, l'identification des gènes mutés pour les maladies mendéliennes monogéniques d'une part et les formes monogéniques des maladies complexes d'autre part (Ostergaard, Simpson et al. 2011, Simpson, Irving et al. 2011, Louis-Dit-Picard, Barc et al. 2012).

I.5.2. Succès limité des analyses de liaison pour les maladies complexes

Les analyses de liaison génétique sont particulièrement puissantes dans le cas de maladies monogéniques pour l'identification de gènes présentant des mutations (rares) à forte pénétrance qui coségrègent avec le phénotype de la maladie (Lander and Schork 1994).

Cependant, dans le cas de maladies complexes pour lesquelles on observe l'implication simultanée de de plusieurs polymorphismes génétiques (fréquents) à faible pénétrance et de facteurs environnementaux sans possibilité d'identifier un mode de transmission, les méthodes de LOD score classique (Morton 1955) sont clairement limitées. Des approches de liaison non paramétrique (NPL) ont été proposées reposant sur l'identification des régions génomiques qui sont partagées entre les individus atteints au sein de familles. Cependant, ces analyses de liaison non paramétrique ont une puissance statistique limitée de sorte que les résultats obtenus sont peu reproductibles (Ott, Kamatani et al. 2011).

En outre, compte tenu de la faible agrégation familiale qui caractérise la plupart des maladies complexes, la caractérisation de leurs bases génétiques repose sur des approches d'études d'association sur des jeux de données cas-témoins (Risch and Merikangas 1996).

I.5.3. Etude d'association génétique

L'approche la plus courante consiste à comparer la fréquence d'un polymorphisme particulier entre une population de cas atteints et des témoins non atteints, les deux populations d'études étant le plus comparable possible. Des méthodes statistiques parmi lesquelles les plus simples sont basées sur le test du χ^2 et le test exact de Fisher peuvent donc être utilisées pour mettre en évidence pour un marqueur donné, une différence de fréquences alléliques ou génotypiques statistiquement significative entre les cas et les témoins.

I.5.3.1. Mesure du risque du variant associé pour les études d'association

Le rapport de cotes (*OR*, *odds ratio*) est préféré aux pénétrances des génotypes (par définition toutes faibles) pour quantifier le risque associé au variant génétique de prédisposition identifié par les études d'association (Reich and Lander 2001). L'*OR* est ainsi défini comme le rapport de la cote de l'événement présence du facteur de risque dans le groupe des cas divisé par la cote de l'événement absence du facteur de risque dans le groupe des témoins. Plus concrètement, pour un polymorphisme présentant les allèles A et B, avec A (allèle rare, ou allèle à risque), l'*OR* associé à l'allèle A est :

Allèles	Cas	Témoins
A	cas^+	$témoins^+$
B	cas^-	$témoins^-$

$$OR\,[A] = \frac{cas^+ / \,témoins^+}{cas^- / \,témoins^-} = \frac{cas^+ \times \,témoins^-}{cas^- \times \,témoins^+}\ ;$$

L'*OR* d'un facteur de risque peut être mesuré en prenant en compte l'effet de plusieurs facteurs environnementaux comme le sexe et la consommation tabagique par exemple, on utilise dans ce cas une régression logistique. Une transformation *logit* classique est utilisée pour exprimer la probabilité d'être atteint comme une fonction linéaire des variables $X = (X_1, X_2,\ldots, X_k)$:

$$logit\left(P(atteint|X)\right) = ln\left(\frac{P(atteint|X)}{1 - P(atteint|X)}\right) = \beta_0 + \sum_{i=1}^{k} \beta_i X_i\ ;$$

Avec β_i les paramètres explicatifs du modèle, que l'on peut estimer par maximum de vraisemblance. L'*OR* de chaque variable peut alors s'obtenir par les formules :

$$\beta = ln(OR) \text{ et } OR = e^{\beta}.$$

Les limites de l'intervalle de confiance à 95 % de l'OR (95% CI) correspondent à :

$$e^{[\beta \pm 1{,}96 \times \sqrt{variance(\beta)}]}$$

Pour tester l'association de la variable X_1 avec la maladie, ce qui revient à tester si $\beta_1 \neq 0$ correspondant à $OR \neq 1$, on peut utiliser un test du maximum de vraisemblance comparant les vraisemblances L_1 et L_2 des deux modèles suivants :

Modèle 1 : $logit\ (P) = \beta_0 + \sum_{i=2}^{k} \beta_i X_i$;

Modèle 2 : $logit\ (P) = \beta_0 + \beta_1 X_1 + \sum_{i=2}^{k} \beta_i X_i$;

Avec $-2ln\left(\frac{L_2}{L_1}\right) \sim \chi^2$ à un degré de liberté

Le polymorphisme à risque peut aussi être codé en fonction du modèle génétique à tester. Ainsi, en se référant toujours à notre exemple précédent pour un polymorphisme présentant les allèles A et B, avec A (allèle rare, ou allèle à risque), nous avons les modèles suivants:

Additif	AA (2)	AB (1)	BB (0)
Dominant	AA, AB (1)	*vs*	BB (0)
Récessif	AA (1)	*vs*	AB, BB (0)

Notons que les études d'association peuvent aussi être conduites dans le cadre de données familiales permettant éventuellement de s'affranchir d'un biais de stratification de populations. Cette approche est possible par le test de déséquilibre de transmission (TDT) introduit par Spielman en 1993 qui compare la fréquence de transmission allélique entre un parent hétérozygote et son enfant atteint à la fréquence attendue selon les lois de transmission

mendéliennes (Spielman, McGinnis et al. 1993).

I.5.3.2. Le Déséquilibre de liaison

Dans le cadre des études d'association, le marqueur associé avec le phénotype étudié peut soit être le variant de prédisposition (association directe), soit être en fort déséquilibre de liaison (LD) avec le variant de prédisposition (association indirecte) (Hirschhorn and Daly 2005). Le LD est donc une notion importante dans les études d'association. Il peut se définir comme l'association non aléatoire d'allèles à des loci différents. Ceci correspond à la situation où la probabilité d'observer un couple d'allèles sur un chromosome n'est pas égale au produit des probabilités d'observer ces allèles individuellement (Wall and Pritchard 2003).

Les origines du LD sont multiples. Il survient le plus souvent du fait de l'apparition d'un nouvel allèle par mutation qui se transmettra aux générations suivantes avec les autres allèles situés autour. Le LD diminuera en fonction du nombre de recombinaisons, l'échelle du temps et la survenue de potentielles nouvelles autres mutations. Ainsi, le LD sera plus élevé chez les populations d'origine asiatique que chez celles d'origine européenne, et plus basse chez celles d'origine africaine (Reich, Cargill et al. 2001). D'autres facteurs peuvent aussi influencer le LD. Il s'agit de la sélection naturelle, la réduction de la taille de la population (goulot d'étranglement ou *bottleneck*) et la dérive génétique. Plusieurs méthodes ont été proposées pour quantifier le LD entre deux loci biallélіques. Les plus fréquemment utilisées sont le coefficient *D'* de Lewontin (Lewontin 1964) qui varie entre *-1* et *+1*, avec la valeur *+1* correspondant au déséquilibre maximal ; et le coefficient de corrélation r^2 qui varie entre *0* et *1*, avec la valeur *1* correspondant au déséquilibre complet. Le déséquilibre de liaison d'une région s'observe couramment par un graphique montrant les valeurs de *D'* où r^2 de toutes les paires de *SNPs*

d'une région (Barrett, Fry et al. 2005).

Les études du LD à l'échelle du génome ont permis de montrer que le génome était divisé en régions de fort LD ou blocs de LD avec une faible diversité haplotypique, séparées entre elles par des points chauds de recombinaison (Daly, Rioux et al. 2001, Jeffreys, Kauppi et al. 2001, Myers, Bottolo et al. 2005). Cette répartition est particulièrement intéressante dans le cadre des études d'association génétique car elle permet de restreindre le nombre de *SNPs* à génotyper pour couvrir une région génomique (Bush and Moore 2012). En effet, il existe au sein de chaque bloc de LD un nombre minimal de *SNPs* (Tag*SNPs)* permettant de reconstruire les haplotypes les plus fréquents (Johnson, Esposito et al. 2001). En utilisant cette approche, on estime pouvoir couvrir près de 80% de la variation génétique en génotypant entre 500.000 et 1.000.000 de *SNPs* (Li, Li et al. 2008, International HapMap, Altshuler et al. 2010).

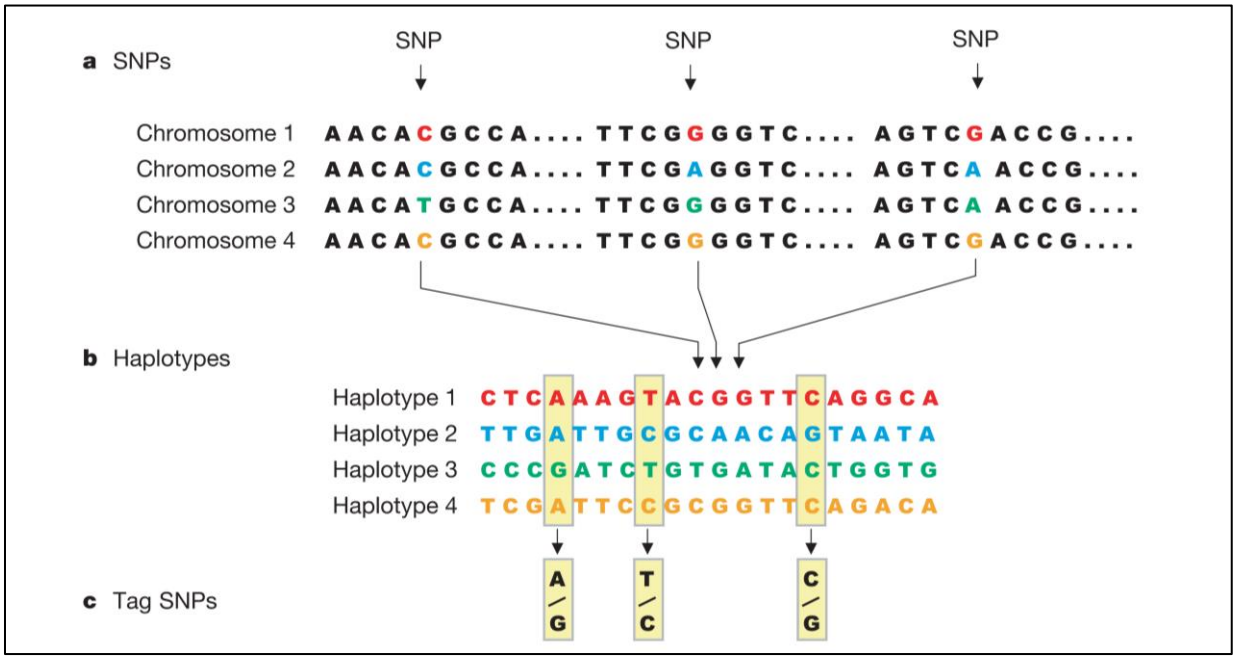


Figure I-6. SNPs, haplotypes et TagSNPs. Il est possible de reconstituer les 4 haplotypes en génotypant juste trois TagSNPs sur les vingt SNPs au total. Illustration tirée de (Li, Li et al.

2008, International HapMap, Altshuler et al. 2010).

I.5.3.3. Etude d'association de gènes candidats

Cette stratégie consiste à rechercher et analyser l'association avec le phénotype de la maladie pour quelques variants génétiques situés dans des gènes particuliers. Le choix de ces régions génomiques se fait sur la base de connaissances ou d'hypothèses physiopathologiques préalables sur la biologie de la maladie, de modèles animaux, ou encore de rapprochement avec d'autres pathologies apparentées dont la composante génétique a été identifiée. Il peut aussi s'agir de régions génomiques identifiées préalablement par des études de liaisons (Jorgensen, Ruczinski et al. 2009).

Cependant, quelques-uns seulement des signaux d'associations résultant de cette approche gène candidats ont été répliqués, la majorité des études s'avérant ultérieurement sous puissantes statistiquement pour conclure à une association (Tabor, Risch et al. 2002, Altshuler, Daly et al. 2008). Ainsi, pour remédier à ces limites, des approches génome-entier sans hypothèses à priori ont été proposées.

I.5.3.4. Etudes d'association pangénomiques (*GWAS*)

Les études d'association pangénomiques (*GWAS, Genome-wide Association Studies*) reposent sur l'exploration de milliers de polymorphismes répartis uniformément sur l'ensemble du génome. Cette approche a été développée en tandem avec l'hypothèse maladie fréquente-variants fréquents (Reich and Lander 2001, Pritchard and Cox 2002) et rendue possible grâce aux progrès de la génomique permettant le génotypage haut débit sur des puces à ADN

(Bumgarner 2013) de milliers d'individus à un coût compétitif et les résultats de projets de recherche collaboratifs (Genomes Project, Abecasis et al. 2010, International HapMap, Altshuler et al. 2010) qui ont permis de constituer un catalogue complet des variants génétiques.

Le premier *GWAS* a été publié en 2005 et portait sur la dégénérescence maculaire liée à l'âge (Klein, Zeiss et al. 2005).Depuis, plus de 15000 loci associés majoritairement avec des effets modestes (OR < 1,5) à des maladies ou des traits phénotypiques ont été identifiés (Burdett T (EBI) , Welter, MacArthur et al. 2014) [cf. Figure I-7].

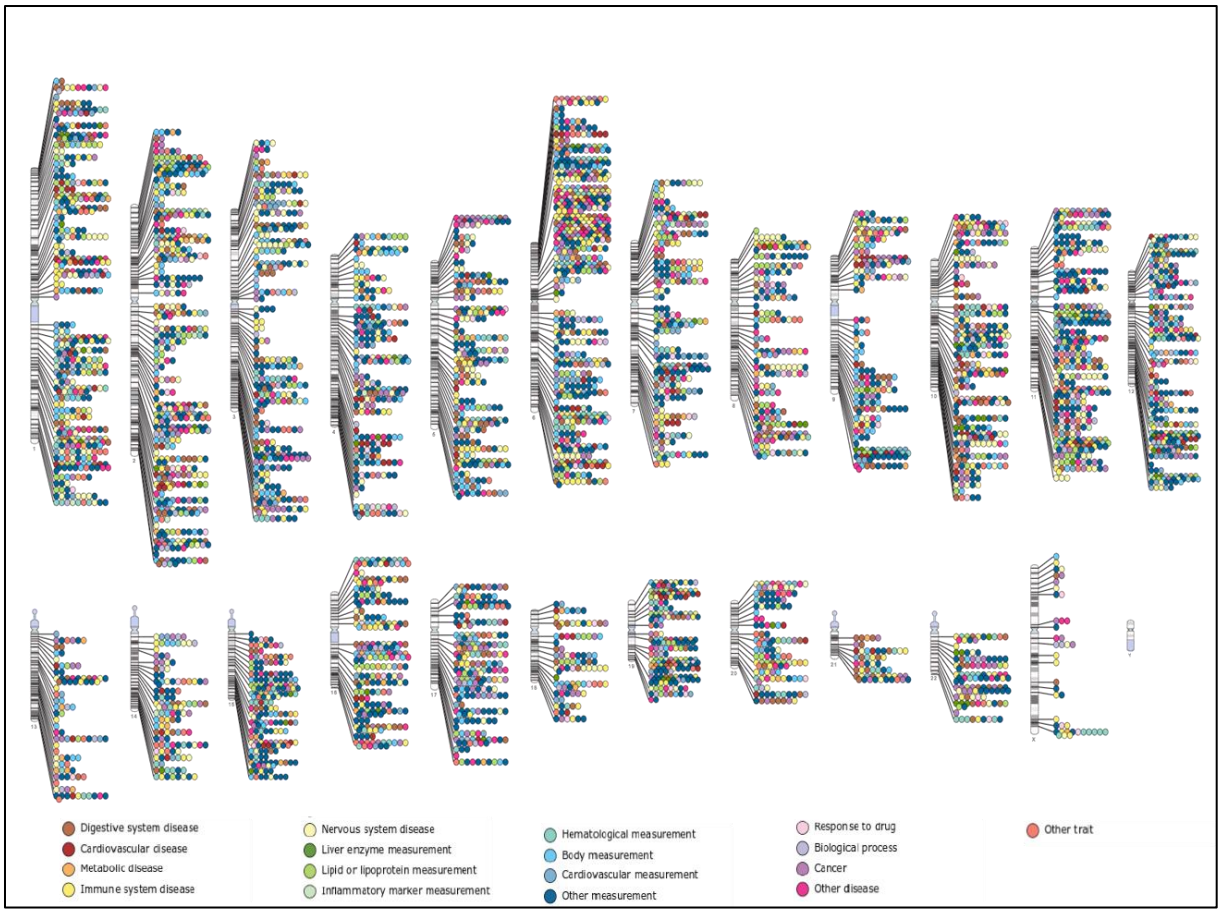


Figure I-7. Polymorphismes identifiés par GWAS. Cette figure montre l'ensemble des loci associés à des maladies ou traits phénotypiques par les GWAS. Cette figure a été générée à partir de données de The NHGRI-EBI Catalog of published GWAS (Burdett T (EBI)).

Compte tenu de leur large utilisation durant ces dernières décennies, plusieurs recommandations ont été émises pour aider à la conception, la réalisation et l'interprétation des études d'association pangénomiques (McCarthy, Abecasis et al. 2008, Bush and Moore 2012, Evangelou and Ioannidis 2013).

I.5.3.4.1. Puces à ADN

Les puces à ADN ont révolutionné l'étude du génome humain. Elles reposent sur les principes d'hybridation entre brins complémentaires, qui permettent de mesurer pour chaque marqueur biallélique une intensité de fluorescence pour ses allèles a et A. Les différents génotypes possibles sont ensuite inférés par un algorithme mathématique à partir des intensités de fluorescence de chaque individu de l'échantillon à étudier.

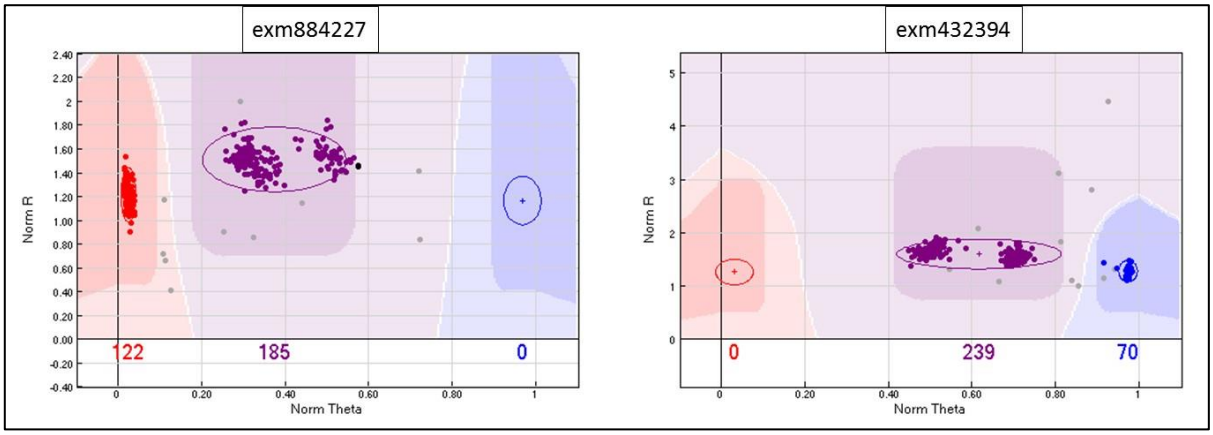


Figure I-8. Principes de l'assignation de génotypes par puce à ADN. Assignation des génotypes par fluorescence pour les marqueurs exm884227 et exm432394 génotypés pour la puce Illumina HumanExome-12v1.

La technologie des puces de génotypage offre de nombreux avantages. Il s'agit notamment de la quantité d'ADN nécessaire pour inférer les génotypes qui est très faible (environ 750 ng) pour la quantité de SNPs génotypés. De plus, l'ensemble du protocole est entièrement standardisé et automatisé (utilisation de kits et de robots), ce qui offre une grande robustesse et une reproductibilité des résultats. Cependant, en cas de délétion, le génotype sera considéré à tort comme homozygote. Ainsi, de nombreux génotypes homozygotes ne reflètent pas une vraie homozygotie, mais la présence de délétions.

Différentes puces de génotypages sont proposées par les deux principales compagnies du secteur (Illumina et Affymetrix). Le choix de la puce de génotypage dans le cadre de projets de recherche prend en compte plusieurs facteurs à savoir la couverture du génome souhaitée, l'ascendance ethnique de la population étudiée, la pathologie étudiée ainsi que le financement disponible. Dans le cadre de mon travail de thèse, nous avons utilisé la puce Illumina HumanExome-12v1, une puce de génotypage qui offre une couverture ciblée des variants exoniques potentiellement fonctionnels. Ces variants ont été observés dans des projets de séquençage d'exomes à grande échelle chez plus de 12000 individus, au minimum 3 fois dans au moins 2 cohortes indépendantes (Exome Chip Design 2014).

I.5.3.5. Méthodologie des *GWAS*

Les *GWAS* sont classiquement divisées en deux grandes phases (Bush and Moore 2012). Dans la première appelée phase de découverte, les données de génotypage par puce de la population d'étude subissent un contrôle qualité approfondi (QC) (Anderson, Pettersson et al. 2010) puis l'analyse statistique visant à tester l'association de chacun des polymorphismes

passant le QC avec le trait étudié est réalisée. L'étape du QC avant la mise en œuvre de différents tests d'association est très importante car elle permet d'éviter les faux positifs en excluant les échantillons d'ADN de mauvaise qualité, les marqueurs mal génotypés, ainsi qu'un éventuel biais de stratification de population. De plus, de nouveaux marqueurs peuvent être imputés à partir de génotype existant dans la phase de découverte (Marchini and Howie 2010), afin de permettre une couverture encore plus large du génome des individus étudiés.

La seconde phase, consiste à confirmer les polymorphismes les plus associés dans la phase de découverte au sein de cohortes indépendantes de réplication (Studies, Chanock et al. 2007). Il s'agit du « gold standard » pour la validation statistique de l'association des polymorphismes identifiés dans la phase de découverte et éventuellement s'affranchir d'un potentiel « winner's curse », synonyme d'effet pour le variant identifié beaucoup plus fort dans la population du *GWAS* que dans la population générale (Zollner and Pritchard 2007). Notons aussi que la cohorte de réplication doit provenir d'une population avec la même ascendance ethnique que celle de la phase de découverte et ayant une caractérisation phénotypique identique. En résumé, la réplication consiste à conduire une nouvelle étude d'association, la plus proche possible de celle de la phase de découverte, en n'examinant cette fois que l'implication spécifique de variants associés au trait sous un certain seuil de significativité. Ainsi, peuvent alors être considérés comme répliqués les variants associés à nouveau dans la cohorte de réplication et présentant un effet sur le trait étudié allant dans la même direction pour les deux études (Bush and Moore 2012).

Notons que compte tenu du nombre important de variants testés dans la phase de découverte, l'interprétation de la p-value d'association tient compte de la problématique des tests multiples. Ainsi, on considère généralement un seuil de significativité à $P = 5 \times 10^{-8}$ après correction de Bonferroni correspondant à environ un million de loci testés. Pour avoir la

puissance statistique nécessaire pour atteindre ce seuil, l'une des stratégies dans le cadre des *GWAS* est d'augmenter la taille de l'échantillon étudié. Cela peut se faire directement à la phase de découverte ou en collaboration entre différentes équipes dans le cadre de méta-analyses de résultats de *GWAS* (Willer, Li et al. 2010, Anttila, Winsvold et al. 2013, Evangelou and Ioannidis 2013).

Le challenge de la méthodologie des *GWAS* est donc d'apporter des arguments statistiques exhaustifs à partir d'études rigoureusement conduites permettant de mettre en évidence des polymorphismes associés au trait étudié. L'objectif ultime étant d'identifier à partir de ces signaux d'associations, des variants de prédisposition qui feront l'objet d'une caractérisation fonctionnelle dans le but de lier le locus à la pathologie/trait étudiée.

I.5.3.6. *GWAS* : du polymorphisme de prédisposition à la fonction

L'évaluation de l'implication causale des variants de prédisposition identifiés par les *GWAS* passe par la mise en œuvre d'expérimentations fonctionnelles. Ces études peuvent être conduites initialement *in-vitro*, par la surexpression de l'allèle à risque dans une lignée de cellules immortalisées, biologiquement pertinentes pour le trait étudié. Le résultat de ces expériences peut ensuite être confirmé dans des cellules de donneurs sains ou malades portant différents génotypes au marqueur d'intérêt. Le but étant de montrer une expression différentielle du gène entre les cas et les témoins d'une part puis en fonction du génotype porté d'autre part.

Ces expérimentations initiales peuvent être complétées par l'introduction de modifications génomiques au locus étudié dans des modèles expérimentaux. De même, le développement de souris mutantes *knock-out* (*KO*) peut fournir des informations importantes

sur les mécanismes moléculaires de médiation des effets pathogènes des loci de prédisposition au trait étudié (Edwards, Beesley et al. 2013).

I.5.3.7. Limite des *GWAS* : héritabilité manquante

Les études d'association pangénomiques ont ouvert la voie à l'identification de milliers de variants génétiques (Burdett T (EBI) , Welter, MacArthur et al. 2014) permettant une avancée significative dans la détermination de la composante génétique des maladies complexes (Manolio 2010). Cependant, malgré toutes ces avancées, ces études présentent quelques limites. Pour la grande majorité des polymorphismes associés identifiés par les *GWAS*, aucun lien de causalité n'a encore été mis en évidence. Cela s'explique par le fait que ces variants sont situés le plus souvent dans des régions non codantes ou intergéniques du génome. Pour ceux-ci, il est difficile d'identifier le variant causal en LD avec le variant associé ou même de leur associer un rôle fonctionnel. De plus, l'ensemble des variants fréquents associés aux maladies complexes pris conjointement ne permettent d'expliquer qu'une part limitée de la variabilité phénotypique (Manolio, Collins et al. 2009). Il en résulte donc une héritabilité manquante pour la majeure partie des maladies complexes que plusieurs facteurs peuvent expliquer.

I.5.3.7.1. Le rôle des variants rares

Les *GWAS* sont focalisés sur l'étude des marqueurs relativement fréquents dans la population générale (MAF ≥ 5%) faisant totalement l'impasse sur les variants rares et très rares (Zeggini, Rayner et al. 2005). Or, on s'attend à ce que ces variants rares aient un effet plus

important sur le trait étudié, participant ainsi de façon importante à l'héritabilité des maladies complexes (Frazer, Murray et al. 2009). Ceci dit, l'étude d'association des variants rares est beaucoup plus ardue. La majeure partie des puces de génotypage offre une couverture limitée de ces variants et les méthodes statistiques utilisées dans les *GWAS* ne sont pas assez puissantes à cause de la faible fréquence allélique et de l'extrême hétérogénéité allélique. Ainsi, des variants génétiques rares de prédisposition aux maladies complexes ont déjà été identifiés (Rivas, Beaudoin et al. 2011, Jonsson, Atwal et al. 2012). Ces variants peuvent donc avoir une implication non négligeable dans les maladies complexes (Iyengar and Elston 2007, Liu and Leal 2010, Ionita-Laza, Buxbaum et al. 2011, Saint Pierre and Genin 2014) et pourraient donc expliquer en partie cette héritabilité manquante.

Notons aussi que le terme variant rare est largement utilisé de nos jours sans consensus clair sur sa définition exacte qui peut être évoquée dans de multiples circonstances soit en synonyme au terme mutation, soit en opposition à polymorphisme ou tout simplement à variant fréquent. Nous retiendrons donc cette dernière définition du terme variant rare en opposition à variant fréquent dans la population avec un seuil de discrimination de la MAF qui varie entre 0,1% (Gorlov, Gorlova et al. 2011), 1% (Frazer, Murray et al. 2009) voire 5% (Bodmer and Bonilla 2008).

I.5.3.7.2. Puissance statistique des *GWAS* et seuil génomique

Le manque de puissance statistique imputable à la taille de la population étudiée est également une source de limitations des *GWAS*. La maximisation des chances de détecter des variants associés nécessiterait de recruter des échantillons de dizaines de milliers d'individus (Park, Wacholder et al. 2010) inenvisageable en pratique pour de nombreuses maladies. En

outre, le seuil de significativité génomique de $P = 5 \times 10^{-8}$, bien que faisant l'objet d'un consensus semble toutefois très conservateur de même que la notion de seuil qui fait penser que l'ensemble des polymorphismes en dessous peuvent n'avoir aucune implication fonctionnelle pour le trait étudié, ce qui n'est pas toujours le cas (Germain, Eyries et al. 2013).

I.5.3.7.3. Autre limites pour les *GWAS* : modèle génétique, interaction, *CNVs*

Le modèle génétique des *GWAS*, qui teste indépendamment l'implication de chaque marqueur dans la maladie peut paraitre assez réducteur. Il est évident que l'architecture génétique des maladies complexes est bien plus compliquée avec l'implication et l'interaction de plusieurs variants appartenant aussi bien au même gène qu'à des gènes différents ou à la même voie métabolique (pathway). D'autres facteurs peuvent aussi être source d'héritabilité manquante. Il s'agit de l'implication de variations structurales comme les *CNVs* (Kidd, Cooper et al. 2008, Sudmant, Rausch et al. 2015), de polymorphismes de novo, de l'hétérogénéité génétique ou même d'une héritabilité surestimée (Clarke and Cooper 2010).

De ce fait, de nouvelles approches ont été pensées et développées pour permettre d'éviter l'héritabilité manquante. Il s'agit notamment des techniques de séquençage nouvelle-génération (NGS), qui ont permis l'accès aux variants rares ouvrant ainsi la voie à de nouvelles opportunités dans l'étude génétique des maladies humaines (Kiezun, Garimella et al. 2012).

I.5.4. Séquençage nouvelle génération (*Next Generation Sequencing*)

Le séquençage de l'ADN est nécessaire pour caractériser les variants génétiques du génome humain. Pendant longtemps, le séquençage Sanger (Sanger, Nicklen et al. 1977) a été la méthode de routine. Toutefois, depuis la fin du projet de séquençage du génome humain (HGP) (Lander, Linton et al. 2001), les méthodes de séquençage nouvelle génération ou séquençage à très haut débit (Metzker 2010) ont connu un essor fulgurant et sont de nos jours très utilisées pour identifier l'implication potentielle des variants génétiques dans les maladies humaines.

I.5.4.1. Technologie du NGS

Le NGS est un ensemble de méthodes de séquençage apparues dans la seconde moitié des années 2000. Elles se caractérisent par l'utilisation d'approches massives en parallèle, permettant de séquencer des centaines de milliers de fragments simultanément (Majewski, Schwartzentruber et al. 2011). Malgré que les détails techniques puissent varier d'une méthode à l'autre, la technologie est globalement la même et peut se résumer en quelques étapes (Voelkerding, Dames et al. 2010)[cf. Figure I-9]. L'ADN génomique subit d'abord une fragmentation enzymatique en petits morceaux aux extrémités desquelles sont liées des séquences adaptatrices. Les librairies ainsi préparées sont ensuite fixées sur un support solide et chaque fragment est amplifié environ 1000 fois pour former des « *clusters* » clonaux qui seront enfin séquencés en parallèle. À chaque cycle, l'ajout d'un nucléotide se traduit par un signal fluorescent associé à chacun des quatre nucléotides de l'ADN.

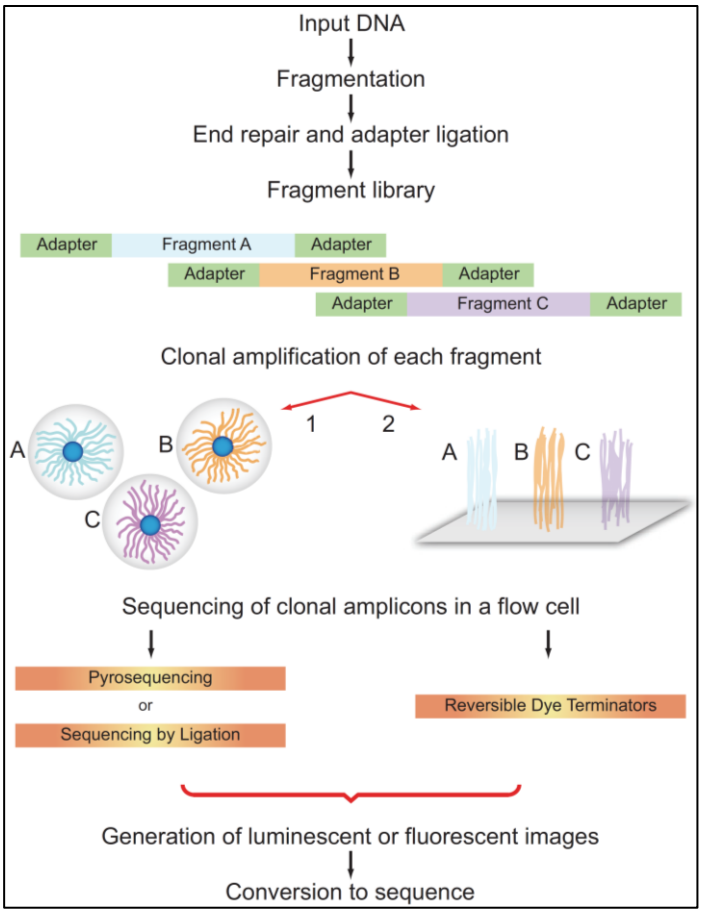


Figure I-9. Étapes du séquençage NGS. Il s'agit des différentes étapes du séquençage NGS pour les plates-formes nécessitant une amplification clonale (Roche 454, Illumina et Life Technologies). Figure extraite de (Voelkerding, Dames et al. 2010).

Les technologies du NGS sont en constantes évolution et la concurrence entre les différents fournisseurs a permis une baisse importante du coût. Il est désormais possible de séquencer un génome entier pour 1000$ (Illumina HiSeq-x (Illumina)). Ainsi les équipes de recherches peuvent désormais séquencer en routine un panel de gènes candidats, les régions codantes du génome (WES) ou même des génomes entiers (WGS).

Aussi, les données brutes de séquenceurs NGS nécessitent-elles la mise en place de véritables pipelines bio-informatiques permettant l'alignement de l'ensemble des reads sur le

génome de référence, mais aussi des étapes de contrôle de qualité basées sur la profondeur de lecture ainsi que des outils pour l'identification et l'annotation des variants (Shendure and Ji 2008, Metzker 2010).

I.5.4.2. Pipeline bio-informatique

Les sorties brutes des séquenceurs ou reads sont de courts fragments d'ADN d'au plus quelques centaines de paires de bases (Metzker 2010, Liu, Li et al. 2012) qu'il va falloir aligner sur le génome humain de référence qui en compte 3 milliards (Lander, Linton et al. 2001). Il s'agit là d'un véritable challenge rendu encore plus laborieux par le fait que chacun de ces reads peut avoir un taux d'erreur pouvant atteindre 2% (Liu, Li et al. 2012), et le génome de chaque individu porte des variations mono-nucléotidiques et structurales par rapport au génome de référence. Les premières méthodes d'alignement étaient basées sur des algorithmes de table de hachage (Flicek and Birney 2009). Toutefois, les méthodes plus récentes, utilisant la transformée de Burrows-Wheeler permettent un alignement plus rapide en utilisant moins de capacité de mémoire (Flicek and Birney 2009, Li and Homer 2010). De plus, la précision du séquençage nécessite que chaque paire de bases soit lue plusieurs fois par différents reads. La profondeur de lecture acceptée varie donc en fonction de la région séquencée, du design de l'étude et de l'objectif de l'étude. Pour un séquençage génome entier par exemple, il est recommandé une profondeur de lecture de 30X et les différents algorithmes d'alignement de reads peuvent prendre en compte la profondeur de lecture. Notons aussi que l'alignement est particulièrement difficile pour les séquences répétées ou présentant des variations structurales par rapport au génome de référence. Dans ce cas, les performances varient en fonction de l'algorithme utilisé, les paramètres choisis et le protocole de séquençage utilisé. Généralement au moins 10% voire plus des reads issus du séquençage ne peuvent pas être alignés sur le

génome de référence (Hatem, Bozdag et al. 2013) et ne peuvent donc être annotés.

L'annotation des variants est le processus consistant à identifier les positions qui diffèrent entre le génome séquencé et le génome de référence. Elle permet aussi d'identifier les gènes et autres éléments fonctionnels. Il s'agit concrètement donc de déterminer pour chaque paire de bases le génotype pour les variations mono-nucléotidiques ou structurales s'il y a lieu en se basant sur la qualité du reads et la probabilité de l'observer à cette position.

En pratique, plusieurs programmes informatiques comme SAMtools (Li, Handsaker et al. 2009) et the Genome Analysis Toolkit (GATK) (McKenna, Hanna et al. 2010) proposent des outils accessibles permettant la manipulation de fichiers issus de séquenceurs NGS, l'alignement des reads, l'annotation et l'assemblage des génomes.

I.5.5. Etude génétique des maladies humaines par NGS

L'avènement du séquençage NG a constitué ces dernières années une véritable révolution pour l'étude génétique des maladies humaines en permettant le développement de nouvelles approches méthodologiques. Parmi celles-ci, le séquençage d'exome est sans doute la plus utilisée et celle qui apporte le plus de succès pour les maladies rares.

I.5.5.1. Le séquençage d'exome

Le séquençage d'exome (WES) permet de séquencer la partie codante du génome soit environ 180.000 exons qui représentent 1 à 1,5% du génome total (Ng, Turner et al. 2009). La technologie est celle du NGS sauf que la préparation des librairies requiert une étape

additionnelle visant à capturer spécifiquement les régions exoniques par hybridation. Cette étape est donc la plus importante dans le WES car c'est elle qui détermine les différentes séquences.

Les premiers kits de capture exoniques reposaient sur une définition assez conservative de l'exome qui se limite aux séquences codantes confirmées décrites dans la base de données du the Consensus Coding Sequence (CCDS) project (Pruitt, Harrow et al. 2009). Toutefois, les kits les plus récents ciblent au minimum toute la collection de séquences codantes du génome des bases de données comme RefSeq (Pruitt, Brown et al. 2014) ou Ensembl (Flicek, Amode et al. 2014) ainsi que de nombreux pseudogènes. Aussi, les nouveaux kits peuvent capturer jusqu'à 62 Mb contrairement aux anciens kits qui étaient limités à 30 Mb. Notons en outre qu'en plus des variants exoniques, le séquençage d'exome a la capacité d'inclure aussi les sites accepteurs et donneurs d'épissage, les régions 5' et 3' UTR ainsi que quelques 500 microARNs permettant une vision plus globale des mécanismes de régulation des gènes.

Malgré les avancées du WES et la performance des kits, la capture ciblée des régions exoniques n'est toujours pas complète. Certains exons peuvent ne pas être couverts ou seront séquencés avec une profondeur de lecture faible conduisant à la possibilité de perdre des variants importants (Majewski, Schwartzentruber et al. 2011). C'est le cas des régions riches en GC par exemple qui présentent une couverture de lecture systématiquement faible (Clark, Chen et al. 2011). En règle générale, une profondeur de lecture minimale de 8-10X est requise pour l'annotation des variants identifiés par WES. En outre, le séquençage de l'exome n'est pas particulièrement performant pour identifier les larges variations structurales du génome telles que les indels, et autres variations de structure comme les larges délétions, substitutions, les réarrangements chromosomiques et les *CNVs* (Majewski et al., 2011) qui peuvent représenter jusqu'à 20 % de la variabilité par individu (Frazer, Murray et al. 2009). Cela est

particulièrement important surtout pour les indels qui peuvent induire des mutations de type frameshift.

En dépit de ces quelques limitations, la grande disponibilité du WES et surtout son coût relativement abordable en ont fait un outil extrêmement populaire pour la recherche en génétique.

I.5.5.2. Promesses du WES

Les séquences codantes du génome abriteraient environ 85% des mutations causales pour les maladies humaines (Majewski, Schwartzentruber et al. 2011) et les variants rares ou extrêmement rares situés dans les régions codantes des génomes sont plus susceptibles d'avoir un effet pathogène (Bamshad, Ng et al. 2011). Sur la base de ces hypothèses, le WES représente donc une alternative intéressante au WGS car il permet d'avoir un accès spécifique à l'ensemble de variations situées dans les régions codantes du génome à un coût relativement abordable avec une meilleure profondeur de lecture pour ces séquences.

Le WES a été conduit pour la première fois en 2008 chez un individu sans maladie génétique connue (Ng, Levy et al. 2008). Très rapidement, une première application de cette technique a été effectuée dans le syndrome de Freeman-Sheldon (FSS), une maladie dominante de diagnostic connu (Ng, Turner et al. 2009). Cela a apporté la preuve de l'utilité du séquençage d'exome pour l'identification des mutations causales à partir d'un petit nombre d'individus atteints non apparentés même en cas d'hétérogénéité génétique. Par la suite, le WES a été utilisé avec succès pour la première fois en 2010 pour l'identification du gène responsable du syndrome de Miller (Ng, Buckingham et al. 2010). Depuis, le WES a relativement tenu ses promesses (Rabbani, Tekin et al. 2014) et prouvé son efficacité par rapport aux études de liaison

et aux *GWAS* pour l'identification de mutations causales pour de nombreuses pathologies humaines [cf. Figure I-10] (Wang, Lu et al. 2015).

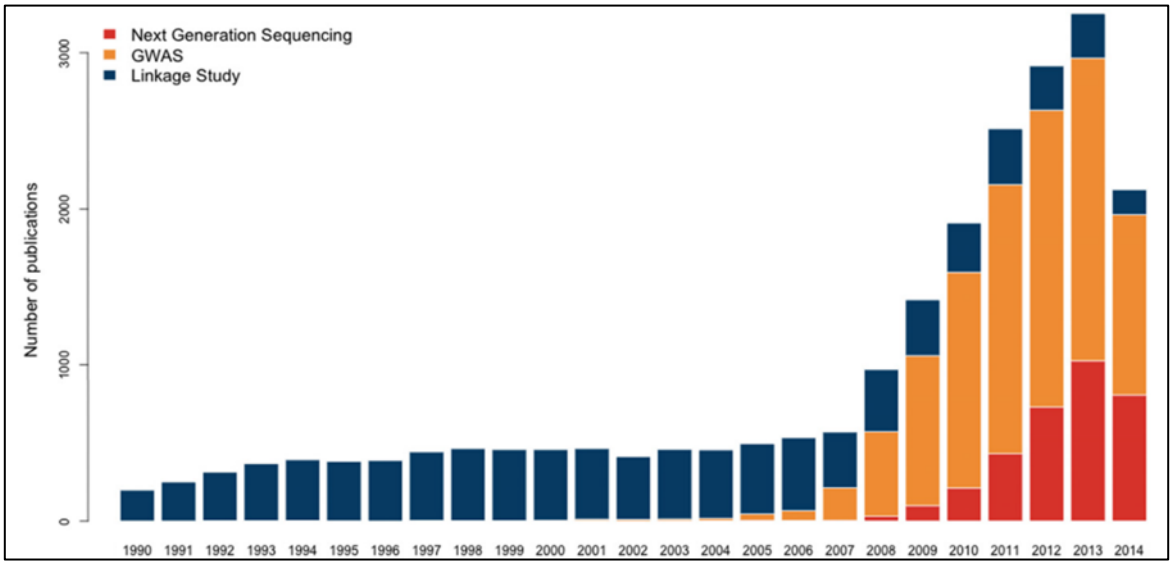


Figure I-10. Publications portant sur l'étude de traits ou pathologies humaines utilisant les approches d'études de liaison, de GWAS et de WES/WGS. Seules les publications répertoriées dans MEDLINE sont incluses. Figure adaptée de (Wang, Lu et al. 2015).

Ainsi les études de WES ont été mises en œuvre pour identifier la composante génétique de nombreuses maladies humaines (Austin, Ma et al. 2012, Albrechtsen, Grarup et al. 2013). Dans les faits, la démarche mise en œuvre dans le cadre de ces études ne fait pas totalement l'impasse sur les études de liaison et d'association génétique et différentes stratégies complémentaires sont souvent mises en œuvre (Walsh, Shahin et al. 2010, Momozawa, Mni et al. 2011, Louis-Dit-Picard, Barc et al. 2012, Albrechtsen, Grarup et al. 2013).

D'autre part, étant donné que le WES permet d'avoir accès à la variabilité exonique individuelle, la quantité d'information obtenue peut être utilisée même en cas d'échec du projet de recherche pour la mise en place de bases de données de variants génétiques.

I.5.5.3. Les bases de données de variants

Les nouvelles techniques de séquençage WGS et WES ont permis un développement fulgurant des bases de données de variants génétiques, permettant à l'ensemble de la communauté scientifique un accès rapide, à une information relativement de bonne qualité à travers des plateformes gratuites et faciles d'accès. Il existe un grand nombre de bases de données en génomique (Rigden, Fernandez-Suarez et al. 2016).

L'une des plus anciennes est la base dbSNP (Sherry, Ward et al. 2001) du *National Center for Biotechnology Information (NCBI*). Elle répertorie uniquement les variations mono-nucléotidiques. Chaque nouveau polymorphisme identifié peut être soumis par la communauté des utilisateurs. Un numéro *rs* lui est alors associé (par exemple rs9349379) et sont répertoriées les informations correspondant à sa position physique sur un assemblage du génome de référence, sa fréquence dans la population et sa pertinence pour un trait ou une maladie humaine selon le cas. Plus récemment le *NCBI* a lancé la base de données *ClinVar* (Landrum, Lee et al. 2014), associée à dbSNP et permettant un accès au phénotype correspondant au variant annoté (Landrum et al. 2014). De façon similaire la base de données Human Gene Mutation Database (HGMD) répertorie aussi les mutations causales ou associées aux maladies humaines directement à partir de données publiées dans la littérature (Stenson, Mort et al. 2014).

Nous avons aussi la base de données 1000 Genomes (Genomes Project, Abecasis et al. 2010) (Genomes Project, Auton et al. 2015), l'une des plus utilisée qui est issue d'un projet

collaboratif international visant à générer un catalogue complet des variations génétiques humaines. Cet objectif est poursuivi à travers l'analyse d'environ 2504 individus provenant de 26 populations humaines. Afin de minimiser le coût du projet, une approche innovante a été mise en place combinant WES avec une faible profondeur de lecture, imputation des données manquantes puis WES avec une profondeur de lecture élevée ≥ 20 X permettant d'augmenter la probabilité de détecter les variants avec une implication fonctionnelle potentielle. L'ensemble des données générées sont en accès libre pour toute la communauté scientifique.

En terme de grand projet de séquençage, nous avons aussi la base de données Exome Variant Server (EVS) (NHLBI Exome Sequencing Project 2016) qui fournit aussi des informations pour des variants identifiées dans plus de 6.500 exomes ainsi que le projet UK10K (UK10K Consortium 2011) qui a séquencé 4000 génomes et 6000 exomes d'individus avec un phénotype clinique vérifié, rendant ainsi disponibles les données pour les études d'association génotype-phénotype.

Une autre initiative en terme de bases de données est le consortium EXAC (*Exome Aggregation Consortium*) (Exome Aggregation Consortium 2016), qui est un projet collaboratifs international d'équipes de recherche avec pour but de regrouper et harmoniser les données issus de divers études génétiques et de grands projets de séquençage d'exome. Ce consortium a déjà collecté les données d'exomes de 60706 individus non apparentés ce qui en fait la plus grande base de données d'exomes.

Les bases de données sont donc des outils indispensables pour la recherche en génomique. Elles permettent à l'ensemble de communauté d'avoir accès à des ressources complètes permettant de nouveaux développements méthodologiques pour l'identification des variants génétiques en cause dans les maladies humaines. D'ailleurs, de nombreuses revues scientifiques encouragent les auteurs à déposer directement dans les bases de données publiques

les séquences nouvellement identifiées lors de la publication.

Cependant, il est à noter que le contenu des bases de données en génomique n'est pas toujours la référence, peut être hétérogène et contenir de nombreuses erreurs. De plus, les bases de données dépendent des thématiques de recherche des équipes participant au contenu, ainsi que de l'expérience des différentes plateformes. Ainsi, l'utilisation des bases de données dans les différentes stratégies de filtration des données de séquençage doit donc prendre en compte tous ces facteurs. Pour la base EVS (NHLBI Exome Sequencing Project 2016) par exemple, les individus inclus sont en majorité hypertendus et présentent des pathologies cardiovasculaires diverses et ne sont donc pas représentatifs de la population générale. En outre, il a été montré pour la base de donnés 1000 génomes qui fait pourtant référence a un niveau élevé de consanguinité inattendu au sein des différentes populations (Gazal, Sahbatou et al. 2015). Malgré l'abondance des ressources, il est donc indispensable de disposer de bases de données les plus épurées possibles et représentatives des différentes populations d'étude. Cela est d'ailleurs indispensable dans le cadre d'études conduites à partir d'analyses d'exome issus de la population française (The French Exome (FREX) Project , GENIN 2016).

I.5.6. Analyse des données de séquençage d'exome

Le séquençage d'exome permet d'identifier en moyenne 20000 variants dans les populations d'ascendance européenne (Bamshad, Ng et al. 2011). Plus de 95% de ces variants sont déjà répertorié dans les bases de données. Environ 10.000 sont des variants non-synonymes, et près de 11.000 sont des variants synonymes. De plus, un certain nombre de ces variants sont susceptibles d'avoir un plus grand impact fonctionnel. Il s'agit de 80 à 100 variants non-sens entrainant un codon stop prématuré, 40 à 50 variants de sites d'épissage et 20 indels

(Bamshad, Ng et al. 2011). Le défi majeur de l'analyse des données de séquençage d'exome est donc de pouvoir identifier des mutations causales parmi ce nombre important de polymorphismes en prenant en compte les artéfacts de séquençage NGS.

I.5.6.1. Analyse du WES dans le cadre de maladies monogénique

L'identification des mutations causales par WES dans le cas de maladie monogénique repose sur les différentes stratégies de filtration des variants rares à fort potentiel fonctionnel. Cependant, des approches complémentaires peuvent être mises en place sur la base de plusieurs facteurs comme des connaissances préalables sur le trait étudié, le mode de transmission du trait, la structure du pedigree, le type de mutation (hérité ou de novo) envisagé pour le trait étudié.

I.5.6.1.1. Stratégie de filtration des variants rares à fort potentiel fonctionnel

C'est la stratégie de filtration la plus commune, dont la pertinence a été démontrée par les travaux de Ng et al pour le syndrome de Freeman-Sheldon (FSS) (Ng, Buckingham et al. 2010). Cette approche consiste à identifier des variants rares à fort potentiel fonctionnel partagé par les individus portant le trait étudié. Cette stratégie a montré son efficacité dans le cas de maladies monogéniques avec de petits échantillons des cas aussi bien sporadiques que familiaux (Gilissen, Hoischen et al. 2011, Ku, Naidoo et al. 2011, Rabbani, Mahdieh et al. 2012).

I.5.6.1.1.1. Filtration des variants rares

Il s'agit d'exclure l'ensemble des variants répertoriés dans les bases de données comme dbSNP (Sherry, Ward et al. 2001) et/ou 1000 Genomes (Genomes Project, Abecasis et al. 2010) et/ou EVS (NHLBI Exome Sequencing Project 2016). Sachant que la grande majorité des polymorphismes identifiés par WES sont déjà répertoriés dans des bases de données (Bamshad, Ng et al. 2011), cette approche est donc relativement puissante et permet de réduire drastiquement le nombre de variants. Cependant, comme nous l'avons développé dans la section précédente, ces bases de données doivent être utilisées avec précaution surtout la base dbSNP qui peut être "polluée" par un certain nombre de variants pathogènes (Zimprich, Biskup et al. 2004, Walsh, Shahin et al. 2010). Outre cet aspect, il semble que le fond génétique des individus constituant la base de données n'ait pas une réelle importance dans la stratégie de filtration. Etant donné que les gènes ont des fonctions similaires au sein des différentes populations humaines, une mutation avec un effet délétère sévère devrait donc avoir le même phénotype pour toutes les populations. Ainsi, l'utilisation de bases de données multi-ethniques uniquement et exclusivement dans le but de filtrer le variant pathogène peut être envisagée. C'est d'ailleurs le cas de la base de données HapMap (International HapMap, Altshuler et al. 2010) utilisée dans sa totalité par Ng et collaborateurs (Ng, Turner et al. 2009, Ng, Buckingham et al. 2010). Cependant, quelques aspects sont à prendre en considération. Les mutations pour les gènes modificateurs peuvent ne pas être les mêmes au sein de différentes populations et toute base de données issue d'un fond génétique différent de celui des individus porteurs du trait peut être moins efficace pour filtrer la mutation causale car la variabilité génomique entre les différentes populations n'est pas forcément chevauchante.

D'autre part, filtrer les variants uniquement sur leur présence ou non dans la base de données, indépendamment de la fréquence de l'allèle mineur (MAF), consiste à prendre le

risque d'exclure des variants pathogènes avec une fréquence extrêmement rare en population générale, particulièrement dans le cas d'étude de maladies récessives. Ainsi, une approche alternative largement utilisée consiste à n'exclure que les variants fréquents avec une MAF maximale fixée en 1% (Bamshad, Ng et al. 2011).

I.5.6.1.1.2. Filtration des variants à fort potentiel fonctionnel

Cette étape est complémentaire à l'exclusion des variants fréquents répertoriés dans les bases de données. Elle permet de prédire la pathogénicité des variants rares filtrés. A cette étape, on peut distinguer deux types de variants. Nous avons d'une part les variants susceptibles d'avoir un impact fonctionnel certain ou probable. Il s'agit de mutations conduisant à un codon stop prématuré, à un décalage du cadre de lecture (*frameshift*) ou de de sites d'épissage (Bamshad, Ng et al. 2011). D'autre part, nous avons les variants faux-sens dont la pathogénicité est prédite à partir d'outils bio-informatiques.

Les outils bio-informatiques de prédiction de la pathogénicité des variants prennent en compte de multiples paramètres comme le degré de conservation de la séquence au cours de l'évolution par rapport aux autres espèces, les propriétés physico-chimiques et biochimiques de l'acide aminé sous-jacent ainsi que la structure de la protéine. Plusieurs outils de prédiction de la pathogénicité des variants sont disponibles (Cooper and Shendure 2011).

L'outil de prédiction *Sorting Tolerant From Intolerant* (SIFT) (Ng and Henikoff 2003, Kumar, Henikoff et al. 2009) est largement utilisé. Il attribue un score au variant entrainant une altération de l'acide aminé en utilisant des séquences homologues. Les variations les plus conservées chez différentes espèces obtiennent un score plus élevé et sont supposées être plus délétères. SIFT est l'exemple typique de l'outil dont les estimations sont directement basées sur

des hypothèses biologiques. D'autres outils sont basés sur des algorithmes d'apprentissage automatique (*machine learning*) combinant les propriétés des mutations existantes dans les bases de données comme HGMD (Stenson, Mort et al. 2014) pour la phase d'apprentissage. Comme exemple, nous pouvons citer l'outil de prédiction PolyPhen-2 (Adzhubei, Schmidt et al. 2010, Adzhubei, Jordan et al. 2013) et MutationTaster(Schwarz, Rodelsperger et al. 2010).

En 2011, Hicks et collaborateurs ont étudié la performance des outils de prédiction PolyPhen-2 and SIFT (Hicks, Wheeler et al. 2011). Les deux avaient respectivement des sensibilités de 0,90 et 0,85, avec des spécificités respectives de 0,40 et 0,52. Etant donné que ces outils sont basés sur différentes hypothèses sous-jacentes, il est habituel de combiner leurs performances au sein d'outils intégratifs de prédiction de la pathogénicité des variants comme CAROL (Lopes, Joyce et al. 2012).

I.5.6.1.2. Approches complémentaires pour les maladies monogéniques

La stratégie de filtration des données n'est pas la seule approche pour l'identification des gènes mutés dans le cas de maladies monogéniques. Plusieurs approches complémentaires sont disponibles et tirent parti des principes de génétique pour identifier efficacement les mutations causales.

Le mode de transmission peut permettre d'optimiser le design expérimental de l'étude et l'analyse des données de WES en permettant la sélection des individus les plus informatifs possibles. Ainsi, pour les maladies récessives, on peut conduire une étude relativement puissante avec peu de cas séquencés. Ceci explique d'ailleurs le succès du WES pour les maladies récessives (Rabbani, Mahdieh et al. 2012). Pour les maladies qui semblent avoir un mode de transmission dominant, la stratégie de filtration peut être optimisée pour exclure les

variants homozygotes ou hétérozygotes composites. En outre, pour les maladies liées au chromosome X, le séquençage du chromosome X uniquement puis la filtration des données peuvent être suffisants pour identifier les gènes mutés (Johnston, Teer et al. 2010). Cependant, dans la pratique, le mode de transmission n'est pas connu de façon certaine et conduire une étude sur la base d'un mode de transmission suspecté peut conduire à l'exclusion par erreur de mutations causales (Simpson, Deshpande et al. 2012).

Lorsque les données d'exomes sont disponibles pour des individus apparentés porteurs du trait étudié, les variants partagés uniquement entre les individus peuvent être considérés pour la filtration. De plus les variants des autres membres de la fratrie non porteurs du trait étudié peuvent également être séquencés pour filtrer les mutations non causales (Wang, Yang et al. 2010).

D'autre part, le WES peut être conduite pour identifier la mutation causale en complément à une élude de liaison génétique qui aurait permis de délimiter au préalable avec succès une région candidate d'intérêt (Depienne, Bouteiller et al. 2012, Rademakers, Baker et al. 2012). Cette approche combinant WES et étude de liaison a aussi montré son intérêt pour l'étude des variants à risque dans les formes monogéniques des maladies complexes (Louis-Dit-Picard, Barc et al. 2012, Cruchaga, Karch et al. 2014).

On peut aussi combiner génotypage par puce et WES ceci surtout au sein de familles consanguines pour lesquelles une maladie autosomique récessive est suspectée. Ainsi, on peut identifier un long segment ou une région homozygote par puce de génotypage que l'on va séquencer par la suite à la recherche de la mutation causale (Walsh, Shahin et al. 2010).

La recherche de mutation de novo est possible par séquençage de trios parents-enfant. La proportion de mutations dans l'exome d'un individu non porté par l'un de ses parents est

très faible (O'Roak, Deriziotis et al. 2011). Cette approche est donc très puissante à condition que l'hypothèse de départ soit bonne, c'est à dire cas sporadique sans histoire familiale de la maladie (O'Roak, Deriziotis et al. 2011, Fromer, Pocklington et al. 2014).

I.5.6.2. WES et maladies complexes

Les techniques de séquençage d'exome peuvent aussi être considérées dans le cadre l'étude de maladies complexes (Kiezun, Garimella et al. 2012), offrant des alternatives à l'hypothèse maladie fréquente variants fréquent largement exploité par les *GWAS* avec un succès relatif. Plusieurs stratégies d'analyse ont été développées pour tester l'hypothèse de l'implication de variants rares et extrêmement rares dans les maladies complexes. Il s'agit entre autres d'études d'association de variants rares (Kozlitina, Smagris et al. 2014, Jackson, Ntalla et al. 2016), de stratégies de WES pour les individus porteurs du phénotype sévère du trait étudié (Cohen, Kiss et al. 2004), au sein d'isolats génétiques (Holm, Gudbjartsson et al. 2011), ou encore de stratégies de reséquençage de loci identifiés au préalable par *GWAS* (Johansen, Wang et al. 2012).

Dans cette section nous aborderons uniquement les études d'association de variants rares.

I.5.6.2.1. Etude d'association de variants rares

L'étude d'association des variants rares est une approche intéressante dont la démarche consiste à comparer l'ensemble des variants identifiés par WES chez des cas porteurs du trait étudié et des témoins non porteurs. Cependant, cette approche à priori séduisante, se heurte à

une difficulté majeure. Il s'agit de la puissance des tests d'association simples –marqueurs qui est largement compromise par la fréquence des variants rares et leur nombre bien plus important. Pour surmonter cette difficulté, plusieurs développements méthodologiques de tests d'association de variants rares proposent de tester l'effet combiné de n variants regroupés au sein d'un gène, une région chromosomique ou un pathway d'intérêt. Cette approche peut être conduite sur des données de séquençage dont le coût est de plus en plus accessible mais aussi et surtout sur des données de puce exoniques développées à partir de projets de séquençage en population générale (Exome Chip Design 2014).

I.5.6.2.1.1. Les "*burden*" tests

L'une des premières méthodes proposées est le *Combined Multivariate and Collapsing* (CMC) test de Li et collaborateurs (Li and Leal 2008). Cette approche propose de combiner les signaux de plusieurs variants rares d'une unité fonctionnelle ou d'un gène candidat au sein d'un signal commun ou score génétique dont on va tester l'association avec le phénotype. Le score génétique d'un individu est de 1 s'il a au moins un variant rare dans le gène et 0 sinon (Li and Leal 2008).

Des évolutions du CMC ont été proposées et sont basées sur des approches différentes pour définir le score génétique par individu (Lin, Zhang et al. 2011). Ainsi, Madsen et Browning proposent d'utiliser un système de pondération qui privilégie les variants rares (*Weighted-Sum test*) (Madsen and Browning 2009). Prince et collaborateurs proposent quant à eux d'utiliser des seuils variables de fréquences pour discriminer les variants rares, et Morris et Zeggini proposent d'utiliser des seuils fixes (Morris and Zeggini 2010). Toutes ces différentes approches constituent ce qu'on appelle les tests de type "*burden*" car elles supposent que

l'ensemble des variants rares regroupés au sein du gène candidat ont un effet sur le phénotype dans la même direction avec un degré similaire (Lee, Wu et al. 2012). Les "*burden*" tests sont donc plus puissants lorsque tous les variants rares ont un effet soit délétère soit protecteur (Basu and Pan 2011). Ainsi, une nouvelle classe de tests d'association de variants rares a été développée pour s'affranchir de ces hypothèses fortes pour une puissance optimale des "*burden*" tests ou "tests de fardeau"

I.5.6.2.1.2. Les "non-*burden*" tests : SKAT

Il s'agit de tests comme le C-alpha (Neale, Rivas et al. 2011) et le SKAT (*Sequence Kernel Association Test*) (Wu, Lee et al. 2011). Ces tests sont optimisés pour prendre en compte des variants avec des effets variés au sein du même gène. Le SKAT est une généralisation du C-alpha et il est représentatif des tests de type "non-*burden*". Il utilise une régression basée sur l'utilisation d'un algorithme d'apprentissage automatique (*kernel machine regression framework*).

Il permet donc d'inclure des covariables dans le modèle de régression et il intègre également un système de pondération qui peut être modifié en fonction des connaissances préalables.

Bien souvent, l'implication de l'effet des différents variants est inconnue et choisir au préalable la bonne approche entre "*burden*" tests et SKAT est impossible. Une approche optimale combinant ces deux tests a été proposée par Lee et collaborateurs. Il s'agit du SKAT-O dont la performance a été démontrée par rapport au "*burden*" test et SKAT pris individuellement dans différents scénarios (Lee, Wu et al. 2012).

Ainsi, le SKAT est largement utilisé dans le cadre des études d'association de variants rares (Huyghe, Jackson et al. 2013, Kozlitina, Smagris et al. 2014) pour lesquelles il représente la méthode de référence. Le SKAT a en outre aussi connu de nombreux autres développements. Il est désormais adapté pour les échantillons de petite taille (Lee, Emond et al. 2012). Il permet aussi de tester l'association combinée de variants rares et fréquents au sein du même gène (Ionita-Laza, Lee et al. 2013), de conduire des méta-analyses d'études d'association de variants rares (Lee, Teslovich et al. 2013) et plus récemment des études d'interactions de variants rares avec les facteurs environnementaux iSKAT (Lin, Lee et al. 2016).

Notons que ces tests d'association de variants rares comme dans le cadre des *GWAS* permettent uniquement de mettre en évidence une association statistique pour des régions chromosomiques ou des gènes susceptibles de contenir les facteurs causatifs. Ils ne permettent pas de présumer de l'implication fonctionnelle des variants constituant la région génomique. Ainsi un gène présentant une association significative avec le trait étudié, des explorations complémentaires doivent être conduites individuellement pour chacun des variants le constituant pour démontrer des mécanismes fonctionnels permettant la preuve de causalité.

I.5.6.3. Confirmation et validation de la mutation causale

La confirmation de la mutation causale est indispensable dans l'étude génétique des maladies humaines par séquençage d'exome. Comme nous l'avons montré précédemment, les données issues de séquenceurs NGS, bien que résultant d'un protocole rigoureux, sont susceptibles de contenir un certain nombre d'erreurs (Liu, Li et al. 2012). Ainsi, la mutation identifiée doit être confirmée par séquençage Sanger (Sanger, Nicklen et al. 1977).

Après confirmation d'un variant rare, d'autres arguments peuvent permettre de justifier

son rôle causal pour le trait étudié. Ainsi, on peut vérifier la ségrégation chez les individus porteurs du trait au sein de plusieurs familles (Rabbani, Mahdieh et al. 2012). Cela peut être rendu périlleux par l'absence de familles étendues, une mauvaise caractérisation du trait étudié pour l'ensemble des membres de la famille et aussi par l'hétérogénéité génétique, qui est un concept récurrent dans l'étude des maladies mendéliennes.

Des arguments fonctionnels peuvent être obtenus expérimentalement pour le variant/gène muté comme décrit précédemment pour la validation fonctionnelle des polymorphismes identifiés par *GWAS* (Edwards, Beesley et al. 2013). Il s'agit de l'invalidation du gène, ou de l'introduction de la mutation dans un modèle animal pertinent pour le trait étudié. Une étude de l'expression du gène peut être conduite par PCR quantitative pour déterminer s'il est exprimé dans un tissu d'intérêt et/ou de façon différentielle en fonction du génotype du variant muté.

CHAPITRE 2 : ETAT DES CONNAISSANCES SUR LA DYSPLASIE FIBROMUSCULAIRE ARTERIELLE (DFM)

Dans le second chapitre, nous ferons une revue de la littérature relative à la dysplasie fibromusculaire en nous attardant sur les aspects justifiant une approche d'étude génétique pour expliquer sa physiopathologie.

II.1. DFM : Définition

La dysplasie fibromusculaire artérielle (DFM), anciennement appelée hyperplasie fibromusculaire artérielle, ou fibroplasie fibromusculaire est un groupe de pathologies vasculaires idiopathiques, segmentaires, non inflammatoires et non athéromateuses de la paroi artérielle. Elle est caractérisée par la sténose, l'occlusion, l'anévrisme ou la dissection des artères de petit et moyen calibre (Plouin, Perdu et al. 2007, Olin and Sealove 2011, Persu, Touze et al. 2012). La DFM se dévoile classiquement par l'aspect angiographique des artères atteintes [cf. `Figure II-11`], qui peuvent présenter une sténose unique (DFM unifocale) ou une succession de sténoses concentriques leur conférant un aspect caractéristique en « collier de perles » (DFM multifocale).

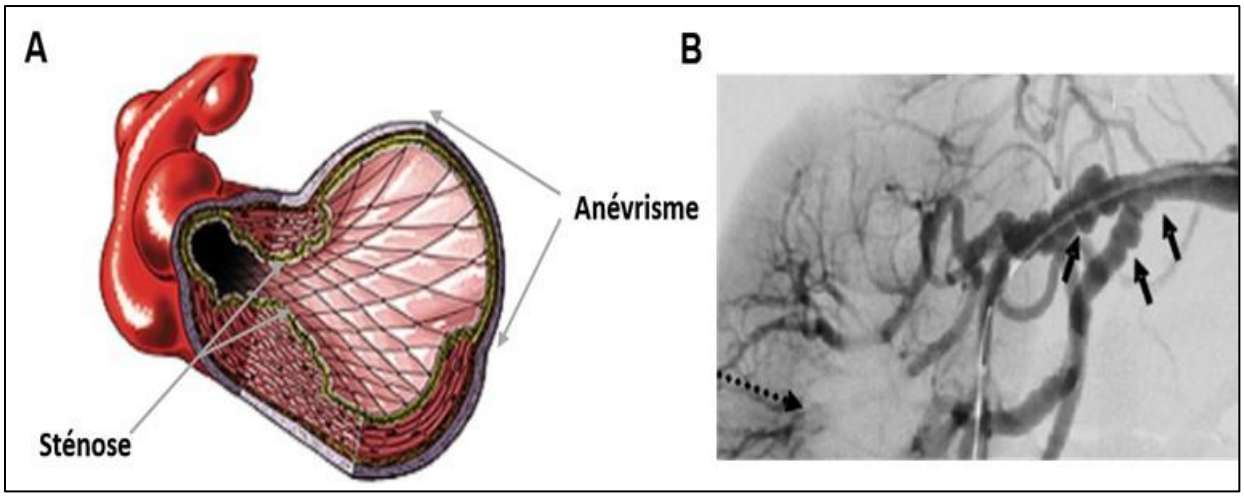


Figure II-11. Lésions dysplasiques de l'artère rénale. (A) Illustration d'une sténose et d'un anévrisme dysplasique, © Inky Mousse Studios 2015. (B) Angiographie de l'artère rénale montrant une DFM Multifocale avec l'aspect typique en « collier de perles » (flèches pleines) et occlusion complète d'une branche de l'artère rénale suite à une sténose dysplasique (flèche en pointillés), © PF Plouin HEGP.

II.2. DFM : Perspectives historiques

La DFM a été décrite pour la première fois en 1938 par Leadbetter et Burkland (Leadbetter and Burkland 1938). Ils évoquaient alors : « une anomalie particulière du développement affectant une artère rénale » chez un garçon de 5 ans présentant une hypertension artérielle sévère associée à une occlusion partielle de l'artère rénale, guéri suite à une néphrectomie unilatérale. Il faudra attendre les travaux de McCormack et collaborateurs (McCormack, Hazard et al. 1958) 20 ans plus tard, pour obtenir une description pathologique précise et l'introduction du terme hyperplasie fibromusculaire, à partir cette fois, de l'observation de 3 patients avec hypertension rénovasculaire. Plus tard, les travaux de Palubinskas et Wylie (Palubinskas and Wylie 1961), Hunt (Hunt 1961), et Kincaid et Davis

(Kincaid and Davis 1961) en 1961 décrivent l'aspect artériographique et les manifestations cliniques de ce qui était alors appelé hyperplasie fibromusculaire.

Dans les années qui ont suivi, la DFM a fait l'objet d'un intérêt croissant de la part de la communauté médicale et scientifique [cf. Figure II-12]. Plusieurs publications ont rapporté des cas pour plusieurs sites artériels autres que rénaux (Palubinskas and Ripley 1964, Connett and Lansche 1965, Ehrenfeld, Stoney et al. 1967, Houser, Baker et al. 1971). Plusieurs systèmes de classification de la DFM sont proposés (McCormack, Poutasse et al. 1966, Kincaid, Davis et al. 1968, Harrison and McCormack 1971, Stanley, Gewertz et al. 1975), des recommandations émises pour le diagnostic et la prise en charge des patients(Olin and Sealove 2011, Persu, Touze et al. 2012, Persu, Giavarini et al. 2014), ainsi que des hypothèses physiopathologiques et génétiques sont envisagées (Rushton 1980, Sang, Whelton et al. 1989, Perdu, Boutouyrie et al. 2007, Savard, Azarine et al. 2013). Plus récemment, la mise en place des cohortes de cas de DFM en France (Savard, Steichen et al. 2012) et aux USA (Olin, Froehlich et al. 2012) a insufflé un nouveau dynamisme aux différents projets de recherche et remobilisé les associations de patients (Kuck, Heidt et al. 2016).

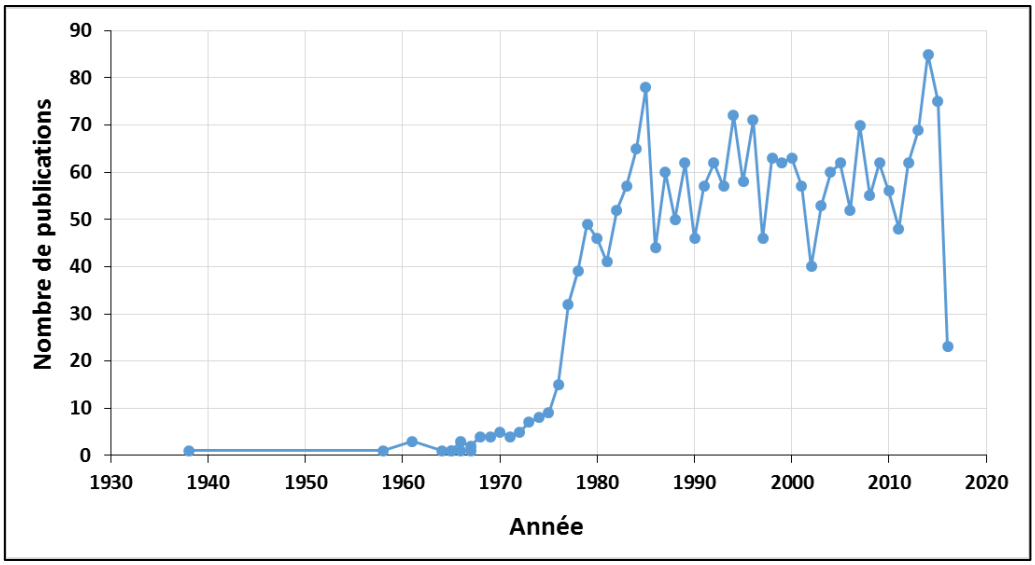


Figure II-12. Nombre de publications référencées dans PubMed pour la recherche du terme "Fibromuscular dysplasia/hyperplasia" dans le titre et/ou le résumé. Au total, 2337 publications sont répertoriées, parmi celles-ci très peu de projets de recherche fondamentale. Il s'agit surtout d'études descriptives, de cases-reports et de recommandations cliniques.

II.3. DFM : Classification

Une classification histologique de la DFM a été proposée par Harrison et McCormack en 1971 (McCormack, Poutasse et al. 1966, Harrison and McCormack 1971) et elle a été améliorée par Stanley et collaborateurs en 1975 (Stanley, Gewertz et al. 1975), puis en 1996 (Stanley 1996). Cette classification est basée sur la couche histologique de la paroi artérielle touchée à savoir l'intima, la média ou l'adventice.

II.3.1. Classification histo-pathologique

La classification histologique de la DFM discrimine trois principaux types de DFM de la paroi vers la lumière de l'artère atteinte. Il s'agit de la DFM intimale, de la DFM médiale et de la DFM périmédiale (Stanley, Gewertz et al. 1975) [cf. Figure II-13].

- La DFM intimale : Elle est caractérisée par un épaississement intimal avec une prolifération lâche et faiblement cellulaire du tissu conjonctif sous-endothélial. La membrane élastique est conservée mais parfois fragmentée avec la média et l'adventice qui restent normales. La DFM intimale est le second sous-type de DFM le plus fréquent et représente environ 5% des cas (Stanley, Gewertz et al. 1975).
- La DFM médiale : c'est la forme histologique la plus fréquente de DFM qui représente 60 à 90% des cas selon les séries (Harrison and McCormack 1971, Stanley and Fry 1975, Olin, Froehlich et al. 2012). Elle est caractérisée par une fibrose dégénérative avec perte de collagène. Elle se présente sous forme d'une succession de zones de sténoses en regard de zones fibreuses alternant avec des zones de cellules musculaires lisses en regard de dilatations. La limitante élastique interne est altérée au niveau des zones fibreuses.
- La DFM périmédiale ou sous-adventicielle : Elle représente 10 à 25% des cas de DFM dans les anciennes séries de cas (Harrison and McCormack 1971, Stanley and Fry 1975) et actuellement moins de 1% des formes chez l'adulte (Olin, Froehlich et al. 2012). Elle résulte de dépôts de collagène dans la partie externe de la média, à la jonction avec l'adventice. L'aspect est très similaire à celle de la DFM médiale, avec des zones de dilatations moins nombreuses et ne dépassant pas le diamètre de l'artère. La limitante élastique interne est détruite.

Deux formes très rares (moins de 1 % des lésions de DFM) ont également été décrites. Il s'agit de l'hyperplasie médiale, qui consiste en une hyperplasie de cellules musculaires lisses, sans fibrose associée et la DFM adventicielle, caractérisée par une fibrose adventicielle.

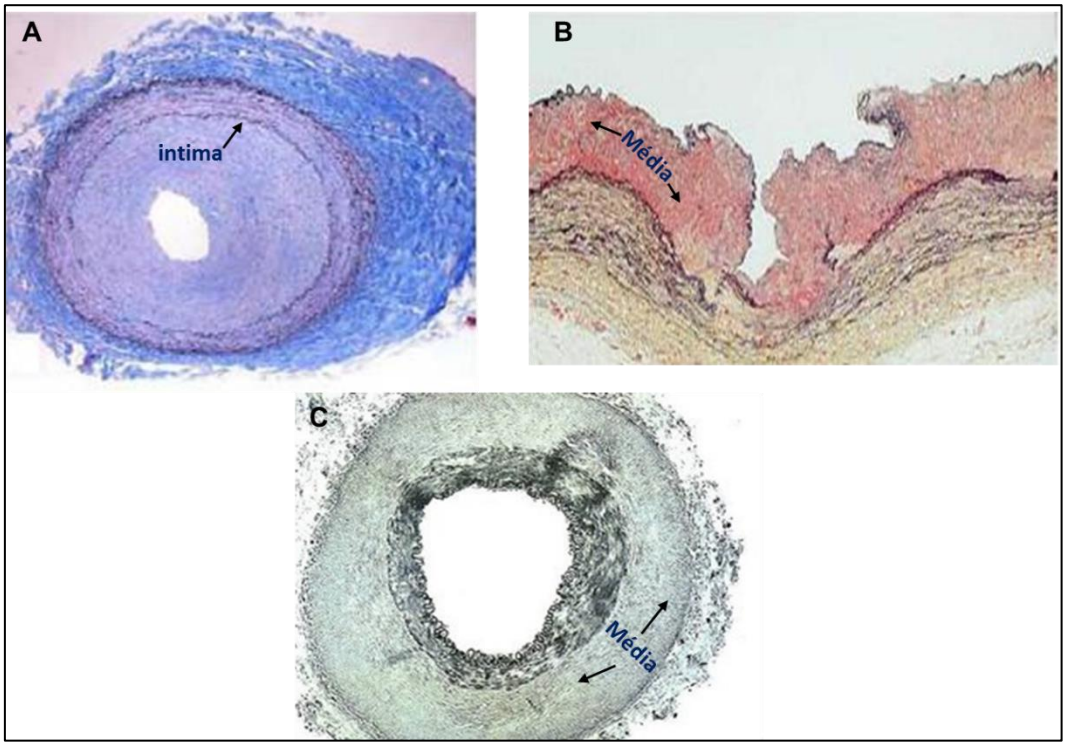


Figure II-13. Classification histo-pathologique de la DFM. (A) DFM intimale. Epaississement concentrique de l'intima avec la media et l'adventice relativement normales ©. (B) DFM Médiale. Altération de la média qui apparait amincie et fibrose du muscle artériel avec perte du collagène. (C) DFM périmédiale. Dépôt de collagène dans les 2/3 extérieurs de la média avec hyperplasie circulaire du 1/3 intérieur. (A) et (B) © (Virmani 2013), (C) (American Urological Association 2016).

Ces différents types histologiques ne sont pas mutuellement exclusifs et il n'est pas rare d'observer l'implication de plusieurs couches histologiques de la paroi au sein du même segment artériel (Alimi, Mercier et al. 1992). Aussi, décrite initialement à partir d'observations faites chez des cas avec une atteinte rénale, cette classification histologiques de la DFM est tout aussi valide pour les artères carotides (Touze, Oppenheim et al. 2010) et a permis d'avoir un cadre cohérent, consensuel et reproductif permettant de discriminer les différents types de la DFM. Cependant, cette classification histologique de la DFM a un intérêt limité de nos jours. En effet, avec l'introduction des techniques de revascularisation percutanée il y a une vingtaine d'années, la chirurgie est rarement actuellement le traitement de première intention pour la DFM, limitant ainsi l'accès et surtout l'analyse de pièces histologiques à des circonstances exceptionnelles. Une nouvelle classification basée sur des corrélations angiographiques avec la classification histologique s'est progressivement imposée dans le cadre de l'étude de la DFM.

II.3.2. Corrélation histo-angiographique

La classification angiographique de la DFM a été proposée en 1968 par Kincaid et collaborateurs (Kincaid, Davis et al. 1968) à partir de la description des caractéristiques angiographiques de l'artère rénale chez 125 cas de DFM. Parmi ceux-ci, les pièces histologiques d'une soixantaine étaient disponibles suite à un traitement par chirurgie permettant d'étudier la corrélation histo-angiographique en plus de la caractérisation angiographique. Ainsi, trois principaux types angiographiques de DFM ont été identifiés.

Il s'agit tout d'abord de la DFM multifocale ou « collier de perles » caractérisée par une succession de multiples sténoses et dilations observée chez 62% des cas étudiés, la DFM

tubulaire avec une sténose longue et concentrique observée chez 14% de cas et enfin la DFM focale avec une sténose solitaire de longueur inférieure à 1 cm, observée chez 7 % des cas. Notons que 17% des cas présentaient un mélange des différents types de sténoses De plus, les corrélations histo-angiographiques ont permis d'associer la DFM médiale à l'aspect angiographique en collier de perles, les sténoses tubulaires et focales n'étant pas spécifiques d'un type histologique particulier (Kincaid, Davis et al. 1968).

De plus, des résultats similaires ont été obtenus à partir d'une autre étude de corrélation histo-angiographique chez 67 cas de DFM (McCormack, Poutasse et al. 1966) permettant ainsi la validation et la généralisation de la classification angiographique qui associe la DFM à un aspect caractéristique en collier de perles avec des lésions dans la média du segment artériel atteint (McCormack, Poutasse et al. 1966, Kincaid, Davis et al. 1968, Lassiter 1998, Plouin, Perdu et al. 2007).

II.3.3. Classification angiographique binaire

La classification de la DFM proposée par Kincaid et collaborateurs (Kincaid, Davis et al. 1968) a permis de décrire trois principaux types angiographiques de la DFM. Compte tenu du fait que les sténoses focales et tubulaires se différencient uniquement par la longueur du segment touché, il parait judicieux de les regrouper en une seule catégorie permettant ainsi de passer à une classification binaire simplifiée (Plouin, Perdu et al. 2007). Cette option a été formalisée par le consensus européen pour le diagnostic et le management de la DFM (Persu, Touze et al. 2012) et sa pertinence démontrée par une étude épidémiologique conduite au sein de notre équipe par Savard et collaborateurs (Savard, Steichen et al. 2012).

Dans cette étude de référence, Savard et collaborateurs ont sélectionné 700 potentiels

cas de DFM de l'artère rénale au sein des différentes bases de données de l'HEGP. Le diagnostic formel a été confirmé par une équipe multidisciplinaire d'experts sur la base des recommandations du consensus européen (Persu, Touze et al. 2012). Il s'agit entre autres de l'absence de sténose athéromateuse, de dosage de marqueurs biochimiques de l'inflammation positifs et toute autre maladie artérielle sténosante associée comme les syndromes d'Ehlers-Danlos vasculaire, de Williams, ou d'Alagille ou encore la neurofibromatose de type 1 (Persu, Touze et al. 2012).

Au total, 337 cas âgés de plus de 18 ans ont été retenus pour l'analyse. Parmi ceux-ci, les patients avec une atteinte unifocale (précédemment focale et tubulaire) comparés aux multifocaux représentaient 18% de la cohorte avec des caractéristiques cliniques spécifiques. Ils étaient le plus souvent des fumeurs (50 vs 26%), de sexe masculin (sexe ratio féminin:masculin, 2:1 vs 5:1). Ils étaient diagnostiqués beaucoup plus jeunes (âge médian 30 vs 49 ans) pour la DFM avec une asymétrie rénale prépondérante, un début de l'HTA plus précoce (âge médian 26 vs 40 ans) et plus sévère (157/97 vs 146/88 mm Hg). Cela a nécessité le plus souvent une revascularisation rénale (90 vs 35%), avec un taux de guérison de l'HTA plus élevé (54 vs 26%). De plus une étude comparative conduite au sein des patients avec une atteinte multifocale avec ou sans aspect caractéristique en collier de perles n'a révélé de différences significatives pour aucune des variables décrites ci-dessus.

Il apparait donc clairement que cette classification angiographique binaire permet de discriminer deux formes cliniques distinctes de la DFM. La forme unifocale étant plus sévère que la forme multifocale plus homogène touchant en majorité les femmes âgées de 30 à 50 ans. Les conclusions de cette étude en termes de prévalence ne peuvent cependant pas être extrapolées à d'autres sites artériels compte tenu du design rétrospectif de cette étude avec de nombreuses données manquantes pour l'exploration angiographique des sites extrarénaux.

II.4. DFM : Diagnostic par imagerie

La démarche diagnostique de la DFM repose sur l'exploration des sites artériels atteints par des techniques d'imagerie médicale non invasifs [cf. Figure II-14] comme l'écho-Doppler, semi invasives comme l'angiotomodensitométrie (angio-TDM) ou l'angiographie par résonance magnétique (ARM) ou encore invasives comme l'angiographie classique par cathétérisation (Vasbinder, Nelemans et al. 2001, Varennes, Tahon et al. 2015). Tout diagnostic confirmé doit être suivi par une exploration complémentaire des sites rénaux et cervico-encéphaliques (Varennes, Tahon et al. 2015).

II.4.1. Diagnostic de la DFM des artères rénales

L'écho-Doopler des artères rénales permet à un coût relativement abordable de mettre en évidence une accélération du flux sanguin dans l'artère associée à des sténoses non spécifiques de la DFM ou une asymétrie rénale. C'est donc un examen de première intention qui ne permet pas de confirmer le diagnostic de la DFM. De plus, il s'agit d'une exploration hautement examinateur-dépendant dont l'acquisition des images peut être perturbée par plusieurs facteurs comme la mobilité rénale à chaque cycle respiratoire ou l'obésité du patient. En outre, il ne permet pas de visualiser l'artère dans son intégrité ni d'explorer les vaisseaux annexes. Cet examen est donc considéré comme moins performant que l'ARM ou l'angio-TDM pour le diagnostic de la DFM (Slovut and Olin 2004, Plouin, Perdu et al. 2007). Cependant, il peut être utile pour le suivi des lésions dysplasiques et de la taille du rein suite à un diagnostic de DFM confirmé par d'autres explorations (La Batide Alanore, Perdu et al. 2007).

D'autre part, l'angio-TDM et l'ARM présentent une spécificité élevée pour détecter les sténoses dysplasiques de l'artère rénale (92 et 84% respectivement), mais ont relativement une

faible sensibilité (64 et 62% respectivement) (Vasbinder, Nelemans et al. 2004). L'ARM a une résolution spatiale plus faible (1-2 mm) mais présente un avantage certain par rapport à l'angio-TDM dont la technologie est ionisante et nécessite l'injection de produits de contraste à base d'iode plus toxique que les agents de contraste à base de gadolinium utilisés pour l'ARM. Ce facteur est important surtout en cas d'atteinte de la fonction rénale. Pour ces patients spécifiquement, une angio-IRM artérielle en temps de vol sans injection de produit de contraste est possible mais nécessite un équipement d'IRM à très haut champ et aucune étude n'a encore validé son utilisation pour le diagnostic de la DFM ni sa performance par rapport à l'angiographie ou aux autres techniques non-invasives (Vasbinder, Nelemans et al. 2004). Cependant, l'angio-TDM permet la visualisation d'un plus grand volume par unité de temps, ce qui réduit considérablement les artefacts. De plus les images sont générées en 3D et peuvent être examinées ou réexaminées en utilisant plusieurs algorithmes de reconstruction, ce qui permet d'améliorer la sensibilité et la spécificité de cette modalité d'imagerie (Sabharwal, Vladica et al. 2007).

Le consensus de la Haute Autorité de Santé en France (HAS) recommande la pratique d'un angio-TDM ou d'une ARM pour établir le diagnostic de DFM des artères rénales. Il considère ces examens comme spécifiques, particulièrement pour le diagnostic de DFM multifocale sur l'aspect en collier de perles, mais peu spécifiques et surtout peu performants pour estimer le degré de sténose de l'artère.

II.4.2. Diagnostic de la DFM des artères cervico-encéphaliques

Le diagnostic de la DFM des artères cervico-encéphaliques est assez similaire à celui des artères rénales. Il peut être suspecté à l'échographie Doppler avec des signes évocateurs plus

spécifiques comme un lambeau intimal décollé, des boucles vasculaires ou encore une ectasie. L'imagerie en coupe, particulièrement l'ARM des troncs supra-aortiques (TSA) ou l'angio-TDM permet de confirmer le diagnostic et d'identifier les potentiels anévrismes et dissections intracrâniens associés (Persu, Touze et al. 2012). De plus, les images peuvent être reconstruites avec des projections en trois dimensions, permettant une évaluation détaillée de la vascularisation cérébrale extracrânienne et intracrânienne de même que l'exclusion des lésions athéromateuses.

Malgré les améliorations apportées aux techniques d'imagerie non invasives ou semi invasives, leur performance pour le diagnostic de la DFM n'est toujours pas optimale. Aussi le gold standard pour le diagnostic de la DFM reste toujours l'angiographie classique par cathétérisation qui est un examen très invasif avec cependant une résolution spatiale inégalée (<0,1 mm) (Plouin, Perdu et al. 2007, Touze, Oppenheim et al. 2010).

II.4.3. L'angiographie classique par cathétérisation pour le diagnostic de la DFM

C'est une technique invasive, ionisante qui expose à des complications artérielles comme l'hématome ou la dissection au point de ponction et à l'artère explorée. Il existe également un risque infectieux dû à l'introduction du cathéter et des complications liées au produit de contraste. C'est donc une exploration dont les risques ne sont pas négligeables et qui est réservée pour des patients chez qui il persiste une incertitude diagnostique après les tests non invasifs et pour lesquels une revascularisation dans le même temps est justifiée.

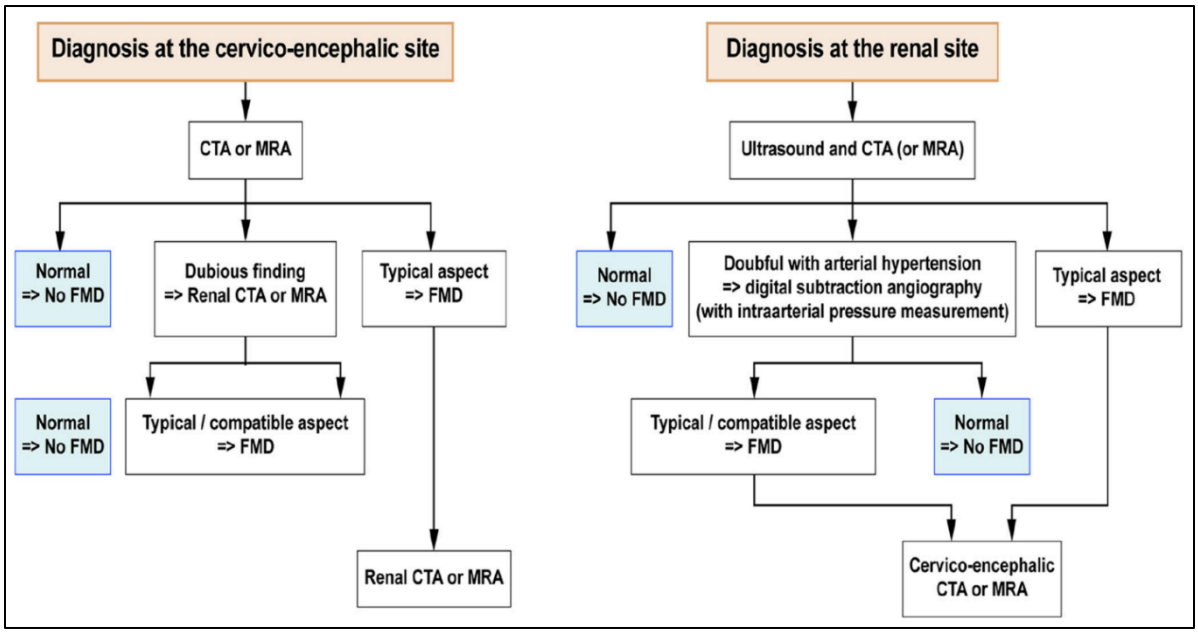


Figure II-14. Diagramme illustrant la stratégie diagnostique de la DFM des artères rénales et cervico encéphaliques par des techniques d'imagerie médicale. D'après Varennes et collaborateurs (Varennes, Tahon et al. 2015)

II.5. DFM : Manifestations cliniques

Les manifestations cliniques de la DFM peuvent être très hétérogènes et relèvent principalement du segment artériel touché, sa localisation et de la nature et de la sévérité des atteintes. Tous les sites artériels de moyen et petit calibre peuvent être atteints par des lésions dysplasiques avec une prédisposition particulière (> 80% des cas) pour les artères rénales, la carotide extracrânienne et les artères vertébrales. Des lésions ont été aussi observées au niveau des artères coronaires, à destinée viscérales (artères mésentériques) et des membres (artères iliaques et brachiales). Les facteurs de risque de progression de la DFM ne sont pas connus avec précision et font l'objet d'une évaluation prospective en cours de réalisation à l'HEGP dans le cadre du protocole PROFILE (PROgression of FIbromuscular LEsions).

II.5.1. Tableau clinique de la DFM rénale

La DFM des artères rénales est la plus fréquente et représente 58% des cas (Mettinger and Ericson 1982, Plouin, Perdu et al. 2007), voire 75% dans certaines séries de cas (Cragg, Smith et al. 1989) (cf. *Figure II-15* (A) (B) (C)]. Elle touche en majorité (80% des cas) des individus de sexe féminin avec une fonction rénale préservée (Pannier-Moreau, Grimbert et al. 1997). Les lésions sont bilatérales dans 60 % des cas (Plouin, Perdu et al. 2007) et surviennent généralement à la moitié ou au 1/3 distal de l'artère rénale avec des extensions dans la partie proximale des branches. L'artère rénale droite est plus souvent atteinte que la gauche (Kincaid, Davis et al. 1968, Luscher, Lie et al. 1987) avec des lésions multifocales en collier de perles retrouvées chez près de 80% à 90 % des cas (Lassiter 1998, Slovut and Olin 2004, Plouin, Perdu et al. 2007, Olin, Froehlich et al. 2012). Les lésions unifocales touchent en majorité les individus de sexe masculin, jeunes, avec des sténoses plus serrées et des lésions rénales d'aval plus fréquentes (Pannier-Moreau, Grimbert et al. 1997).

La manifestation clinique courante de la DFM de l'artère rénale est une HTA rénovasculaire sévère chez la femme âgée entre 15 et 50 ans (Slovut and Olin 2004) avec une fonction rénale préservée le plus souvent (Pannier-Moreau, Grimbert et al. 1997). L'existence d'un souffle abdominal est aussi fréquente (Stanley 1996, Pannier-Moreau, Grimbert et al. 1997). La DFM peut aussi être asymptomatique et découverte de façon fortuite lors d'une consultation pour un autre motif. La progression vers l'insuffisance rénale est rare, même dans les cas de lésions bilatérales (Cragg, Smith et al. 1989, Plouin, Perdu et al. 2007).

II.5.2. Tableau clinique de la DFM cervico-cérébrale

La DFM des artères cervico-cérébrales représente environ 32% des cas et affecte principalement la carotide et les artères vertébrales (Mettinger and Ericson 1982) [cf. *Figure II-15* (D)]. Cette prévalence remonte à l'étude de séries de patients datant des années 1980 et elle est certainement sous-évaluée (74% des cas dans le registre américain de la DFM (Olin, Froehlich et al. 2012)). L'âge moyen de diagnostic est autour de 50 ans (Mettinger and Ericson 1982, Stewart, Moritz et al. 1986, Slovut and Olin 2004. La localisation la plus habituelle est extracrânienne en dehors des sites habituels de l'athérosclérose, au tiers moyen et distal des artères carotides internes et des segments V2-V3 des artères vertébrales, au niveau vertébral C1-C2. L'artère carotide interne est touchée dans la majorité des cas (95 %), souvent de façon bilatérale (60–85 %). L'atteinte vertébrale est moins fréquente et rarement isolée {Touze, 2010 #29)

. Il existe aussi comme dans l'atteinte rénale une prédominance féminine et une majorité de formes histologiques médiales diffuses donnant l'aspect angiographique en collier de perles dans 80 % des cas (Osborn and Anderson 1977, Mettinger and Ericson 1982).

Les manifestations cliniques peuvent être asymptomatiques ou associées à une large variété de symptômes non spécifiques incluant des céphalées parfois de type migraineux, des acouphènes pulsatiles, des étourdissements et le vertige (Mettinger and Ericson 1982, Touze, Oppenheim et al. 2010, Olin, Froehlich et al. 2012). La DFM des artères cervico-cérébrales devient symptomatique lorsqu'il se produit une hypo-perfusion, une embolisation ou une thrombose ou encore lorsqu'une dissection spontanée ou une rupture d'anévrisme associée survient. Dans ce cas, les manifestions cliniques sont par contre beaucoup plus sérieuses compte tenu de leur localisation. Il s'agit de l'accident ischémique transitoire (8-53%), de l'accident vasculaire cérébral (8-35%), de l'hémorragie cérébrale (5-13%) et sous-arachnoïdienne (3-49%)

et de la dissection artérielle (Touze, Oppenheim et al. 2010). La DFM affecte principalement l'artère carotide interne dans sa partie extracrânienne, mais tous les territoires peuvent être atteints (La Batide Alanore, Perdu et al. 2007, Plouin, Perdu et al. 2007).

II.5.3. Complications de la DFM

L'anévrisme et la dissection artérielle représentent les principales complications de la DFM avec laquelle ils ont été décrits conjointement dans des séries de cas étudiées. Notons tout de même que les différentes estimations de prévalences sont très variables d'une étude à l'autre et parfois limitées à l'étude d'anciennes séries de cas pour lesquels le diagnostic formel de la DFM peut être discuté.

II.5.3.1. Anévrismes artériels associés à la DFM

Les lésions anévrismales sont relativement fréquentes dans la DFM et leur prévalence atteint 17% dans le registre américain de la DFM (Olin, Froehlich et al. 2012). Ils sont majoritairement situés au niveau de l'artère rénale (33%) où de grands anévrismes peuvent se développer au sein de la média sans évolution vers la rupture dans la majorité des cas. Les autres localisations les plus fréquentes sont la carotide (21%), l'aorte (19,7%) et l'artère cœliaque (15,8%). Il existe également une association entre DFM et anévrisme intracrânien, qui compte pour 11,8% de l'ensemble des anévrismes rapportés par le registre américain (Olin, Froehlich et al. 2012). Plusieurs autres auteurs rapportent une prévalence entre 22% et 51% (Stanley, Gewertz et al. 1975, Osborn and Anderson 1977, Mettinger and Soderstrom 1978, Mettinger and Ericson 1982). En 1998, dans une méta-analyse de 18 études incluant 615 patients de la

DFM, Cloft et collaborateurs ont estimé la prévalence des anévrismes intracrâniens à 24% (Cloft, Kallmes et al. 1998). Même lorsqu'on se limite à l'échantillon de cas sans hémorragie sous-arachnoïdien, soit 212 individus, la prévalence était toujours de 7,3%, ce qui reste bien supérieur à la prévalence attendue dans la population générale, environ 2,3% (Rinkel, Djibuti et al. 1998) et conforte l'hypothèse de la surreprésentation des anévrismes intracrâniens au sein des populations de cas de DFM.

II.5.3.2. Dissections artérielles associés à la DFM

La dissection artérielle est reconnue comme une complication relativement fréquente de la DFM depuis de nombreuses années (Mokri, Sundt et al. 1986, Stewart, Moritz et al. 1986, Lacombe 2001, Schievink 2001, Debette and Leys 2009). Dans le cadre du registre américain de la DFM, 19,7% des individus inclus présentent une dissection artérielle, avec des atteintes multisites pour 20%. Ces dissections sont localisées dans la grande majorité au niveau des artères carotides (12,1%), rénales (3,6%) et vertébrales (3,4%) (Olin, Froehlich et al. 2012).

La dissection de l'artère rénale peut survenir chez 5 à 10% des cas de DFM, notamment en présence de sténoses tubulaires de l'artère rénale (Lacombe 2001). Elles peuvent être responsables de l'infarctus rénal par dissection occlusive totale de l'artère ou par embolies distales, conduisant à une douleur soudaine dans le flanc, une hématurie avec ou sans une hypertension rapidement progressive (Paris, Bobrie et al. 2006, Plouin, Perdu et al. 2007). Dans les anciennes séries de cas rapportés, cette prévalence de la dissection de l'artère rénale associée à la DFM était plus importante, de l'ordre de 91,7% avec de façon intéressante une prédisposition particulière pour les sujets de sexe masculin (Edwards, Stanson et al. 1982).

La dissection spontanée des artères cervicales (CeAD) correspond à un hématome dans la paroi de la carotide interne ou de l'artère vertébrale. C'est l'une des causes majeures d'accident vasculaire cérébral ischémique chez le jeune adulte (Leys, Bandu et al. 2002, Debette and Leys 2009, Putaala, Metso et al. 2009), et on estime qu'elle est associée avec la DFM (Schievink 2001, Debette and Leys 2009). De façon générale, la fréquence de la DFM dans les séries de cas de CeAD varie considérablement, allant de 5,6% dans une étude multicentrique récente totalisant 983 cas de CeAD (Bejot, Aboa-Eboule et al. 2014) à près de 21% dans l'une des toutes premières cohortes de CeAD (62 patients) (d'Anglejan Chatillon, Ribeiro et al. 1990). Inversement, 12,1 % des cas de DFM présentent une dissection de l'artère cervicale (Olin, Froehlich et al. 2012) qui est désormais considérée comme l'une des complications probables de la DFM. Même si les mécanismes sous-jacents de l'association entre CeAD et la DFM ne sont pas clairement établis, il est désormais habituel d'évoquer une DFM avec une exploration des sites rénaux chez les patients présentant une dissection des artères cervico-encéphaliques, en particulier si elle est spontanée, multifocale ou dans une localisation atypique (Varennes, Tahon et al. 2015).

II.5.4. DFM des artères coronaires

La DFM des artères coronaires représente depuis quelques années un axe de recherche majeur à l'instar des atteintes rénales et cervico-encéphaliques (Michelis, Olin et al. 2014). Les manifestations coronaires de la DFM peuvent être à l'origine d'une dissection spontanée d'un tronc artériel coronaire ou plus rarement d'une branche de division (Saw, Poulter et al. 2012, Tweet, Gulati et al. 2015). Elle se présente cliniquement comme un syndrome coronarien aigu, un infarctus du myocarde, une mort subite, ou une dysfonction ventriculaire gauche chez des patients à faible risque de maladie coronaire avec une atteinte de la DFM confirmée à un autre

site artériel (Pate, Lowe et al. 2005, Michelis, Olin et al. 2014). La prévalence de la DFM des artères coronaires est inconnue. Dans le registre américain, seulement 6,5% des cas présentent une maladie coronaire (DFM et maladie athérome confondus), avec 3,1% des patients signalant un infarctus du myocarde et de 1,3% une procédure de revascularisation coronaire. De façon intrigante, les évènements de mort subite ont été rapportés chez 19,8% des apparentés au premier et au second degré des cas de DFM du registre américain (Olin, Froehlich et al. 2012), largement au-dessus de la prévalence en population générale qui varie entre 50 et 100 morts subites pour 100000 individus (Fishman, Chugh et al. 2010, Deo and Albert 2012). De nouvelles données publiées récemment viennent d'ailleurs corroborer cette observation.

En effet, dans une étude récente, Prasad et collaborateurs ont rapporté une prévalence de 45 % pour la DFM au sein d'un échantillon de 115 patients suivis pour une dissection spontanée de l'artère coronaire tous types confondus (Prasad, Tweet et al. 2015). Dans une autre étude excluant les formes athéromateuses de la dissection spontanée de l'artère coronaire, Saw et collaborateurs ont conclu à une prévalence de près de 72 % de la DFM des artères coronaires chez 168 cas avec un diagnostic de la DFM confirmé pour deux autres sites artériels (Saw, Aymong et al. 2014). Ces données laissent penser que la prévalence de la DFM des artères coronaires est totalement sous-évaluée. Ceci n'est d'ailleurs pas très surprenant compte tenu du fait qu'au diagnostic, l'aspect angiographique typique en collier de perles survient rarement pour les artères coronaires (Camuglia, Manins et al. 2009). On observe plutôt en lieu et place une tortuosité avec un aspect en tire-bouchon (Saw, Bezerra et al. 2016).

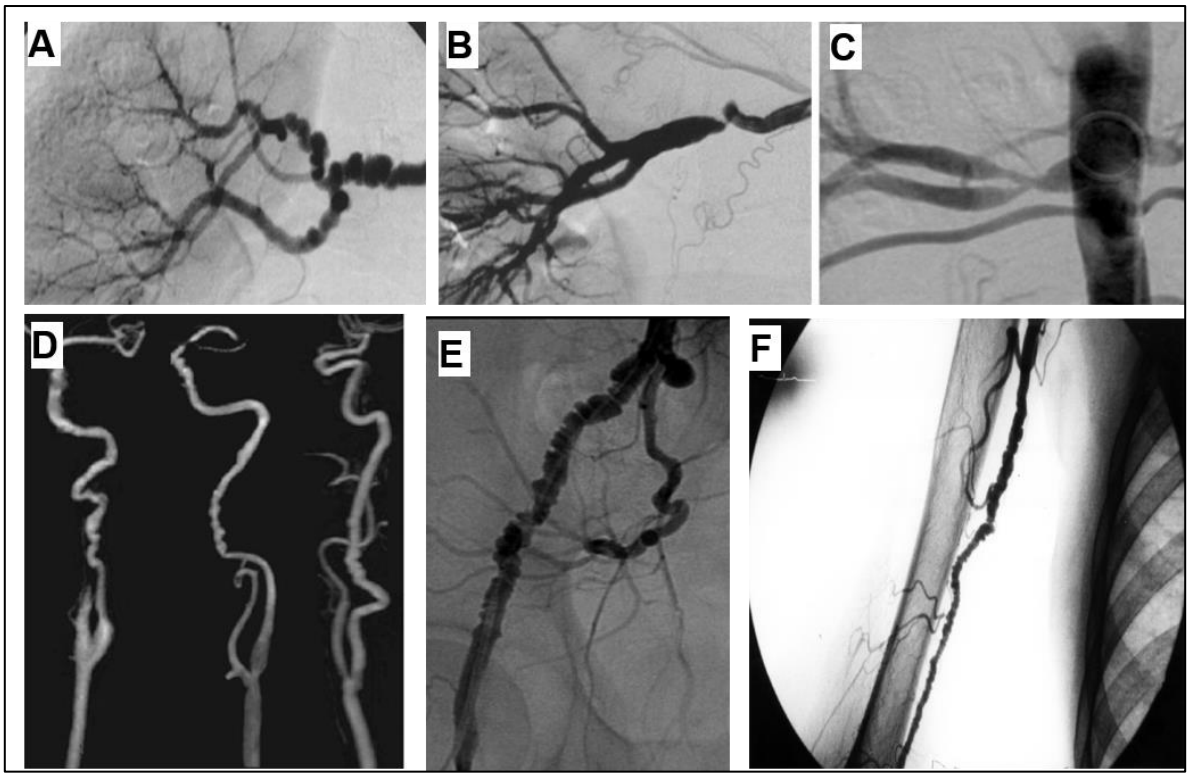


Figure II-15. Aspects angiographiques et cliniques de la DFM. (A) Angiographie de l'artère rénale avec DFM multifocale adaptée de (Plouin, Perdu et al. 2007). (B) Angiographie de l'artère rénale avec DFM focale adaptée de (Plouin, Perdu et al. 2007). (C) Angiographie de l'artère rénale avec DFM tubulaire adaptée de (Plouin, Perdu et al. 2007). (D) DFM multifocale des artères carotides (angiographie par résonnance magnétique) adapté de (Touze, Oppenheim et al. 2010). (E) Angiographie de l'artère iliaque externe avec DFM multifocales adaptée de (Plouin, Perdu et al. 2007, Olin and Sealove 2011, Persu, Touze et al. 2012). (F) Angiographie de l'artère brachiale avec DFM multifocale adaptée de (Suzuki, Daida et al. 1999, Slovut and Olin 2004, Plouin, Perdu et al. 2007, Touze, Oppenheim et al. 2010, Persu, Touze et al. 2012)

II.6. DFM : Pathologies associées à la DFM et diagnostic différentiel

Plusieurs pathologies sont associées à la DFM avec laquelle elles partagent la même symptomatologie clinique non spécifique. Aussi, pour établir le diagnostic formel de la DFM, il est souvent nécessaire de recourir au diagnostic différentiel.

II.6.1. DFM et athérosclérose

Il est plus ou moins facile de différentier les lésions de DFM et d'athérome bien plus fréquentes dans la population générale et qui surviennent généralement à l'origine ou à la portion proximale de l'artère chez des sujets plus âgés cumulant un ou plusieurs facteurs de risque cardiovasculaire. Cependant, il est maintenant reconnu que la DFM peut survenir ou être diagnostiquée à n'importe quel âge, et les patients peuvent présenter simultanément DFM et athérosclérose (Pascual, Bush et al. 2005, Olin, Froehlich et al. 2012).

II.6.2. DFM et vascularite

La DFM est par définition un processus non inflammatoire, alors que la vascularite est caractérisée par une inflammation marquée des vaisseaux sanguins. Dans les vascularites des artères comme la maladie de Takayasu et l'artérite à cellules géantes, l'aspect angiographique des segments artériels montre de longues zones de rétrécissement lisse qui sont classiques mais pas pathognomoniques. Des anévrismes peuvent être rapportés avec des sténoses qui sont le plus souvent tubulaires ou focales (Hata, Noda et al. 1996). Les dosages de protéine C-Réactive et de la vitesse de sédimentation sont habituellement normaux dans les cas de DFM à moins d'infarctus du rein ou de l'intestin. Ainsi, parce que la DFM peut survenir à plusieurs sites

artériels, provoquer une hypertension artérielle, une insuffisance rénale, un AIT, un AVC, et des anomalies comme la sténose solitaire, l'anévrisme ou la dissection, elle peut donc être confondue avec une vascularite. En cas de doute persistant, une échographie intravasculaire et/ou l'absence de sténose aortique peut être nécessaire pour discriminer les lésions dysplasiques (Gowda, Loeb et al. 2003).

II.6.3. DFM et autres pathologies associées

D'autres pathologies peuvent présenter des sténoses au niveau des artères rénales et viscérales et mimer une DFM. Il s'agit du syndrome d'Ehlers-Danlos (SED) vasculaire, le syndrome de Williams ou la neurofibromatose de type 1. Dans l'un ou l'autre des cas, le diagnostic repose sur des traits phénotypiques associés à des explorations génétiques. Il s'agit pour le SED vasculaire d'une hyperextensibilité cutanée, de nombreuses cicatrices cutanées témoins de la fragilité tissulaire et une hyperlaxité articulaire confirmé par la détection des mutations du gène COL3A1) (Germain and Herrera-Guzman 2004); pour le syndrome de Williams d'une dysmorphie faciale, une malformation cardiaque (sténose aortique supra valvulaire), un retard psychomoteur, une dysmorphie du visage et un profil cognitif et comportemental spécifique confirmé par la mise en évidence par FISH de la micro délétion au locus q11.23 du chromosome 7 (Lacolley, Boutouyrie et al. 2002); et pour la neurofibromatose de type I, des lésions cutanées spécifiques en forme de taches café au lait (Gutmann, Aylsworth et al. 1997).

II.7. DFM : Aspects épidémiologiques

La prévalence de la DFM en population générale est inconnue. Cela s'explique par la non-spécificité des symptômes, la démarche diagnostique complexe, avec comme gold standard l'angiographie, une technique d'imagerie médicale invasive, rendant de facto impossible pour des raisons éthiques tout dépistage en population générale. A cela, il faut ajouter que la DFM est globalement encore peu connue par les professionnels de santé. Cependant, plusieurs estimations de la prévalence ont été proposées. Elles varient considérablement en fonction des séries étudiées indiquant toutes une sous-estimation de la prévalence de la DFM.

II.7.1. Prévalence de la DFM des artères rénales

La prévalence clinique (symptomatique) de la DFM rénale a été estimée à 0.4% par Plouin et collaborateurs (Plouin, Perdu et al. 2007) en se basant sur plusieurs postulats à savoir : la proportion de sténoses de l'artère rénale consécutives à la DFM : 10-20% ; la proportion d'HTA-rénovasculaire dans la population d'hypertendus : <1-2% ; prévalence d'HTA chez le sujet d'âge moyen en population générale : 20% (Plouin, Perdu et al. 2007).

A l'opposé, il semble que la DFM rénale asymptomatique soit beaucoup plus fréquente. Plusieurs études, se basant sur des angiographies de potentiels donneurs de reins ont estimé la prévalence de la DFM asymptomatique entre 3,8 (1862 potentiels donneurs) (Cragg, Smith et al. 1989) et 6,6% (716 potentiels donneurs) (Neymark, LaBerge et al. 2000). Ces estimations ont été confirmées par une méta-analyse de quatre études angiographiques totalisant 3181 potentiels donneurs de reins (Spring, Salvatierra et al. 1979, Cragg, Smith et al. 1989, Neymark, LaBerge et al. 2000, Andreoni, Weeks et al. 2002) pour lesquels 139 (4,4%) présentaient une DFM rénale (Plouin, Perdu et al. 2007). Deux autres études beaucoup plus récentes décrivent

des estimations de la prévalence de la DFM asymptomatiques similaires, de l'ordre de 2,6% et 5,8% respectivement chez 2680 potentiels donneurs de rein (McKenzie, Oderich et al. 2013) et 997 patients participant à l'essai clinique CORAL (Hendricks, Matsumoto et al. 2014). Cependant, ces estimations présentent quelques limites et ne peuvent pas être extrapolées à une population générale. En effet, les critères radiologiques de diagnostic ne sont pas disponibles pour toutes les études et n'ont donc pas été standardisées pour homogénéiser la méta-analyse. D'autre part, les donneurs de reins potentiels ne sont pas représentatifs de la population générale car, ils sont le plus souvent apparentés avec des patients atteints de pathologies rénales chroniques avec une possible dimension génétique.

II.7.2. Prévalence de la DFM des artères cervico-cérébrales.

Il existe peu d'études sur la prévalence de la DFM des artères carotides et vertébrales en population générale. Comme pour le DFM rénale, elles reposent sur l'étude de séries angiographiques, et l'estimation de la prévalence varie de 0.3 à 3.2% (Touze, Oppenheim et al. 2010). Toutefois, ces explorations angiographiques sont réalisées pour des indications cliniques spécifiques, il est probable que ces prévalences soient surévaluées. D'ailleurs, Schievink et collaborateurs reportent seulement 4 cas de DFM des artères cervico-cérébrales dans une série de 20244 autopsies consécutives (Schievink and Bjornsson 1996). Cela pose donc la question de la sensibilité de l'examen histo-pathologique en comparaison avec l'angiographie.

II.7.3. Facteurs de risques connus pour la DFM

Deux principaux facteurs de risque à la DFM ont été identifiés. Dans les cohortes françaises (Savard, Steichen et al. 2012) de la DFM et le registre américain (Olin, Froehlich et al. 2012), la quasi majorité des participants (respectivement 85 et 91%) sont de sexe féminin avec en majorité une atteinte multifocale de DFM. Cela fait du sexe féminin le principal facteur de risque à la DFM. Il semble toutefois que le sex-ratio soit légèrement inversé chez les sujets unifocaux qui sont en majorité de sexe masculin. Cependant la forme unifocale étant moins fréquente, les séries étudiées sont limitées à de petits échantillons (Pannier-Moreau, Grimbert et al. 1997).

Le tabagisme a été évoqué comme facteur de risque secondaire à la DFM par plusieurs études (Nicholson, Teichman et al. 1983, Luscher, Lie et al. 1987, Sang, Whelton et al. 1989). Même si la relation de dose-effet (cigarette-risque de DFM) rapportée par Sang et collaborateurs n'a pas été observée dans les études plus récentes (Bofinger, Hawley et al. 1999, Olin, Froehlich et al. 2012, Savard, Azarine et al. 2013), il a été démontré que les patients de DFM qui fument ont une maladie artérielle plus sévère que ceux qui ne fument pas (Bofinger, Hawley et al. 1999). Aussi, dans le registre américain de la DFM (Olin, Froehlich et al. 2012), 37% des participants ont rapporté avoir déjà fumé. Enfin, les travaux de Savard et collaborateurs conduits au sein de notre équipe (Savard, Azarine et al. 2013) ont démontré une différence significative de la proportion de fumeurs entre 337 cas de DFM appariés à des témoins pour l'âge, le sexe, la pression artérielle systolique (PAS) et le nombre de traitements antihypertenseurs utilisés (30% vs18% ; $P < 0.001$; OR = 2.5 [1.6-3.9]). Ils ont aussi montré chez les patients avec une atteinte multifocale de la DFM, une association entre le statut fumeur actuel et un diagnostic plus précoce de l'hypertension (36 vs 42 ans, $P < 0.001$) et de la DFM (43 vs 51 ans; $P < 0.001$), avec une fréquence plus élevée de l'asymétrie rénale et un recours

plus fréquent à la chirurgie rénale. Cette étude va dans le sens de l'hypothèse d'une forme plus sévère de la DFM chez les sujets fumeurs mais pourrait aussi refléter une reconnaissance plus précoce de la DFM chez les sujets hypertendus du fait de leur statut tabagique les prédisposant à un plus grand risque de maladies cardiovasculaires (Powell 1998). Toutefois, les mécanismes par lesquels le tabagisme contribue à l'étiologie de la DFM n'ont cependant pas encore été déterminés et plusieurs autres hypothèses sont avancées pour la physiopathologie de la DFM.

II.8. DFM : Physiopathologie et étiologie de la DFM

De nombreuses hypothèses concernant la responsabilité de facteurs hormonaux, mécaniques ou génétiques ont été proposées pour expliquer la physiopathologie de la DFM. Cependant, aucune étude publiée n'a permis de confirmer clairement et de façon robuste ces hypothèses et la physiopathologie de la DFM demeure toujours inconnue.

II.8.1. Facteurs hormonaux et mécaniques pour la DFM

Dans toutes les séries de cas rapportés pour la DFM, il y a une constante prédominance féminine (McCormack, Poutasse et al. 1966, Kincaid, Davis et al. 1968, Plouin, Perdu et al. 2007, Olin, Froehlich et al. 2012) de la DFM qui laisse supposer que des facteurs hormonaux, notamment une imprégnation œstrogénique jouerait un rôle important dans la physiopathologie de la DFM. De plus, Luscher et collaborateurs ont suggéré que l'étirement répété de l'artère rénale droite plus longue que la gauche peut provoquer des microtraumatismes qui la prédispose beaucoup plus à la DFM (Luscher, Lie et al. 1987). Cette hypothèse a été corroborée par Savard et collaborateurs au sein des cohortes françaises avec une prédisposition pour l'atteinte rénale

droite dans les formes unifocales de la DFM (48% unilatéral droit vs 31% unilatéral gauche ; 21% bilatéral) et multifocales (31% unilatéral droit vs 7% unilatéral gauche ; 61 % bilatéral) (Savard, Steichen et al. 2012). Cependant, toutes ces hypothèses ont été rejetées par les travaux de Sang et collaborateurs (Sang, Whelton et al. 1989) montrant par ailleurs une association avec le tabagisme.

Récemment, Silhol et collaborateurs (Silhol, Sarlon-Bartoli et al. 2015) ont mis en évidence par immunohistochimie, l'existence de récepteurs à la progestérone dans le tissu artériel rénal dysplasique (6 cas de DFM) en comparaison à des artères saines (3 témoins). Ces résultats, bien que préliminaires, relancent ainsi l'hypothèse hormonale, qui reste une piste intéressante et jusque que là sous-exploitée pour expliquer la physiopathologie de la DFM.

II.8.2. La DFM est-elle une maladie génétique?

La composante héréditaire de la DFM a été évoquée dès 1965 par plusieurs auteurs à partir de cas rapportés au sein de fratries et de patients apparentés (Halpern, Sanford et al. 1965, Hansen, Holten et al. 1965, Major, Genest et al. 1977, Stavenow, Henriques et al. 1978, Ouchi, Tagawa et al. 1989) et chez des jumeaux monozygotes (Bigazzi, Bianchi et al. 1998).

En 1980, Rushton (Rushton 1980), Gladstien et collaborateurs (Gladstien, Rushton et al. 1980) ont mené une étude génétique de la DFM auprès de 20 familles avec au moins un cas de DFM. Au sein de 8 familles (40%) seul le cas index semblait atteint et dans les 12 autres (80%), entre 1 et 11 apparentés du cas index présentaient une DFM. Les auteurs ont alors conclu à une transmission héréditaire de la DFM sur un mode autosomique dominant avec une pénétrance variable (Rushton 1980). Cependant, dans cette étude, le diagnostic était en partie basé sur des entretiens, considérant les apparentés comme atteints en présence de lésions cardiovasculaires

plus volontiers d'origine athéroscléreuse. Malgré cette limite fondamentale, cette publication a le mérite d'être la première étude permettant d'émettre une hypothèse claire sur les formes familiales de la DFM et envisageant un mode de transmission s'y afférant.

De même, une récurrence familiale de la DFM chez 11% des cas a été démontrée par Pannier-Moreau et collègues au sein de notre équipe à partir de 104 cas index avec des lésions dysplasiques prouvées par angiographie. Dans cette étude, la forme familiale était cette fois définie par la présence d'au moins deux apparentés au premier degré, diagnostiqués pour la DFM dans les mêmes conditions que le cas index (Pannier-Moreau, Grimbert et al. 1997). Une proportion similaire a été observée de façon descriptive dans le registre américain de la DFM où 7,3% des cas déclarent avoir un apparenté avec un diagnostic confirmé de DFM (Olin, Froehlich et al. 2012).

Compte tenu de tous ces facteurs, nous pouvons affirmer que la DFM est certainement une maladie avec une composante génétique avérée dont l'évaluation et l'identification du modèle sous-jacent ont fait l'objet de quelques études conduites en majorité au sein de notre équipe avec les cohortes françaises de la DFM.

II.9. DFM : Génétique de la DFM

Très peu d'études dans la littérature scientifique traitent de la génétique de la DFM. Lorsqu'elles existent, elles sont limitées à de petits échantillons de cas, et aucune n'a permis d'identifier des variants génétiques associés à la DFM à cause d'un manque récurent de puissance statistique.

II.9.1. La DFM : une maladie génétique difficile à étudier

La principale difficulté pour l'étude génétique de la DFM est donc l'absence de familles étendues ou de grandes cohortes de cas. Cela s'explique par la prévalence de la maladie qui est inconnue, du phénotype qui est variable et hétérogène, du diagnostic qui nécessite des examens d'imagerie invasifs, et l'exclusion des autres causes de sténoses artérielles comme l'athérosclérose bien plus fréquente. De plus, les lésions dysplasiques sont le plus souvent asymptomatiques et l'exploration angiographique systématique d'apparentés "sains" de cas index avec un diagnostic confirmé soulève de nombreuses interrogations éthiques ; ce qui a largement limité l'étude de familles étendues. Cependant, les travaux de Boutouyrie et collaborateurs pourraient constituer une alternative intéressante à l'exploration angiographique (Boutouyrie, Gimenez-Roqueplo et al. 2003).

Boutouyrie et collaborateurs ont montré la présence de lésions spécifiques de la paroi des artères carotides communes et radiales chez des cas de DFM rénale sans atteinte carotidienne à partir d'un système non invasif d'écho-tracking haute résolution (Boutouyrie, Gimenez-Roqueplo et al. 2003). Ces travaux ont aussi permis la détermination d'un score semi-quantitatif combinant les informations fournies par la lecture d'images bidimensionnelles et de signaux de radiofréquence. Ce score phénotypique varie d'une valeur normale de 2 à une valeur très anormale de 7. A partir d'une étude de 70 cas de DFM appariés à des témoins, ils ont déterminé qu'un score > 3 (un aspect échographique en triple signal) conférait un OR de 12,9 (IC95 %, 5,7-29,3) d'être atteint de DFM avec une sensibilité de 73% et une spécificité de 81%. En outre, l'utilisation de cet endophénotype au sein d'études familiales a montré une ségrégation compatible avec un mode de transmission autosomique dominant (Perdu, Boutouyrie et al. 2007). Même si cet endophénotype se montre relativement satisfaisant à première vue, il ne permet pas d'envisager la réalisation d'analyses de liaison dans un plus grand

nombre de familles au sein desquelles le score de chaque individu serait systématiquement établi. En effet, la spécificité est limitée par l'âge de l'individu et l'existence d'un athérome sous-jacent et la sensibilité semble imparfaite. En plus de la question de la fiabilité, il se pose aussi la question de la reproductibilité de ce score dont la détermination nécessite les compétences de praticiens chevronnés et expérimentés sachant toute la subjectivité qui peut résulter de lecture d'image et de la notation d'une image (Swensson 1996). Aussi, l'examen angiographique demeure toujours le gold standard pour le diagnostic de la DFM en attendant la caractérisation d'un endophénotype plus simple et moins subjectif.

Malgré ces difficultés pour la constitution de cohortes étendues de cas de DFM, plusieurs études gènes candidats ont été conduites sur la base d'hypothèses à priori sur la physiopathologie de la DFM avec les séries de cas existants.

II.9.2. Etude gènes-candidats de la DFM

Le diagnostic de la DFM a été rapporté dans quelques cas de déficit en protéine α1-antitrypsine (α1-AT) (Schievink, Bjornsson et al. 1994, Schievink, Puumala et al. 1996, Solder, Streif et al. 1997), qui est un inhibiteur de la sérine protéase, le plus abondant dans le plasma humain. Son principal substrat est l'élastase neutrophile qui est susceptible de dégrader la matrice extracellulaire de la paroi artérielle comme constaté histologiquement au cours de la dysplasie fibromusculaire. Dans une publication parue en 2006, notre équipe a testé l'association de trois polymorphismes au locus AAT avec la DFM dans une étude cas témoins regroupant 161 cas de DFM diagnostiquées pas angiographie rénale et 1085 témoins issus de trois cohortes de témoins constituées respectivement de 353 sujets hypertendus de la base de données HYPERGENE (Charru, Jeunemaitre et al. 1994), 444 témoins de sexe féminin issus

de la base de donnée IPC (Fontana, Gaussem et al. 2003)et 288 témoins issus de la base de donnée TENOR sans maladie cardiovasculaire connue (Jeunemaitre, Inoue et al. 1997). Aucune association significative n'a pu être identifiée entre les polymorphismes testé et la DFM, pas plus qu'avec les phénotypes obtenus par écho-tracking (Perdu, Gimenez-Roqueplo et al. 2006).

Les polymorphismes du système rénine-angiotensine (RA) ont été impliqués dans la pathogenèse de l'hypertension et des maladies vasculaires athéroscléreuses (Rigat, Hubert et al. 1990, Jeunemaitre, Soubrier et al. 1992, Levy 1998). Il a été donc émis l'hypothèse qu'ils peuvent jouer un rôle dans la physiopathologie de la DFM. Bofinger et collègues ont démontré l'association d'un polymorphisme I/D de l'enzyme de conversion de l'angiotensine (P = 0.026) chez 43 cas de DFM rénale diagnostiqués par angiographie comparés 89 témoins normo-tendus (Bofinger, Hawley et al. 2001). Toutefois, cette étude n'a jamais été répliquée.

Sur la base de la même stratégie, l'implication de plusieurs autres gènes a été explorée. Il s'agit notamment du gène de l'élastine *ELN* qui porte des mutations causatives pour la sténose aortique supra-valvulaire dans le syndrome de Williams (Pober, Lacro et al. 1993), le gène *ACTA2* codant pour l'actine des muscles lisses, déjà impliqué dans de rares cas d'anévrismes de l'aorte (Marks, Gullett et al. 2011), ou encore plusieurs gènes impliqués dans des syndromes de perte de la fonction du tissu conjonctif, notamment les gènes *COL31A, TGF-β1* et *TGF-β2* (Poloskey, Kim et al. 2012). Aucune de ces études n'a pu mettre en évidence et/ou répliqué une association significative entre des variants génétiques et la physiopathologie de la DFM à cause d'un manque accru de puissance statistique. La DFM est de nos jours considérée comme une maladie avec une composante génétique sous-jacente (Plouin, Perdu et al. 2007, Olin, Gornik et al. 2014, Persu, Giavarini et al. 2014) pour laquelle la mise en œuvre d'études génétiques en phase avec les nouvelles techniques d'acquisition massives de données à partir de cohortes de cas bien caractérisées phénotypiquement permettra d'expliquer la physiopathologie.

OBJECTIFS DE LA THESE

L'objectif de mon projet de thèse était de caractériser les bases génétiques de la dysplasie fibromusculaire par l'étude des cohortes françaises cas. Pour atteindre cet objectif, j'ai mené deux projets de recherche complémentaires utilisant des approches de séquençage d'exome de cas familiaux de la DFM et de génétique épidémiologique.

Dans la première partie, nous avons émis une première hypothèse qui stipule que des variants génétiques rares à forte pénétrance cumulés dans un certain nombre de gènes ou de voies biologiques, ou agrégés au sein de familles avec plusieurs membres atteints seraient à l'origine de l'étiologie de la DFM, une maladie relativement rare. Ainsi, nous avons mené une étude de séquençage d'exome chez 16 cas familiaux de la DFM, dans le but d'identifier un ou plusieurs gènes enrichis en variants rares codants à forte pénétrance pour lesquels une implication causale dans l'étiologie de la DFM serait envisagée (Article 1).

D'autre part, étant donné les estimations grandissantes de la prévalence de la DFM en population générale, et les facteurs environnementaux comme le tabagisme, l'imprégnation œstrogénique et le stress mécanique suggérés pour la DFM, l'on ne peut exclure l'hypothèse d'une base génétique à caractère polygénique et multifactorielle. Aussi, avons-nous donc conduit dans la seconde partie de ce travail de thèse une étude d'association génétique incluant en phase de découverte 249 cas issus des cohortes françaises de la DFM et 689 témoins, tous génotypés par la puce *Exome-chip*. Le but a été tout d'abord de confirmer/valider par une étude d'association de variants rares (*SKAT* et *Burden tests*), les gènes candidats identifiés par séquençage d'exome. Ensuite, il s'est agi par une étude d'association de variants fréquents d'identifier des polymorphismes de prédisposition à la DFM pour lesquels une réplication dans des cohortes indépendantes et des arguments fonctionnels validera l'implication dans la physiopathologie de la DFM (Article 2).

DEUXIEME PARTIE : TRAVAUX ORIGINAUX

CHAPITRE III : L'ETUDE DE VARIANTS GENETIQUES RARES SUGGERE L'ABSENCE D'UN GENE MAJEUR POUR LA DFM

Article 1

Exome sequencing in seven families and gene-based association studies indicate genetic heterogeneity and suggest possible candidates for fibromuscular dysplasia.

Kiando SR, Barlassina C, Cusi D, Galan P, Lathrop M, Plouin PF, Jeunemaitre X, Bouatia-Naji N. **Journal of Hypertension. 33(9):1802-1810, September 2015.** PMID : 26147384

III.1. Position du problème

La dysplasie fibromusculaire artérielle (DFM) est un groupe de pathologies vasculaires rares, non inflammatoires, et non athéromateuses caractérisées par la sténose, l'occlusion, l'anévrisme ou la dissection des artères de petit et moyen calibre, en particulier les artères rénales et la carotide extracrânienne (Plouin, Perdu et al. 2007, Olin and Sealove 2011, Persu, Touze et al. 2012). Elle prédispose à l'hypertension, et dans certains cas à l'accident vasculaire cérébral. La physiopathologie de la maladie est totalement inconnue. Une agrégation familiale a été démontrée et une composante génétique clairement suggérée (Gladstien, Rushton et al. 1980, Rushton 1980, Pannier-Moreau, Grimbert et al. 1997, Olin, Froehlich et al. 2012). Plusieurs études gènes candidats ont été conduites, aucune d'entre elles n'a permis de mettre en évidence et/ou de répliquer l'association de variants génétiques avec la DFM à cause notamment d'un manque accru de puissance statistique (Pober, Lacro et al. 1993, Bofinger, Hawley et al. 2001, Perdu, Gimenez-Roqueplo et al. 2006, Marks, Gullett et al. 2011, Poloskey,

Kim et al. 2012). Nous avons conduit une étude de séquençage d'exome chez 16 cas apparentés de DFM puis une étude d'association de variants rares génotypés par la puce Illumina HumanExome chez 249 cas non apparentés de DFM et 689 témoins. Ces différentes approches avaient pour but d'identifier un ou plusieurs gènes enrichis en variants rares codants à forte pénétrance impliqués dans l'étiologie de la DFM.

III.2. Résultats majeurs

Le séquençage d'exome réalisé chez 16 cas apparentés de DFM issus de 7 familles a permis d'identifier en moyenne 63.127 variants répartis sur environ 22.000 gènes. A cette étape, nous avons mis en place une stratégie de filtration des données de séquençage d'exome. Il s'agissait de sélectionner les variants codants avec une profondeur de lecture ≥ 20X, provoquant un changement d'acide aminé (faux sens non-sens, et de sites d'épissage), rares avec une MAF ≤ 1% dans les bases de données EVS et/ou 1000Genomes et prédits comme étant potentiellement fonctionnels par Polyphen 2. Cette stratégie de filtration de données a conduit à l'analyse de 3981 variants répartis dans 3971 gènes en vue de l'identification de mutations partagées par toutes les familles ou par des cas au sein de la même fratrie.

Aucun des 3971 gènes ne présentait des mutations partagées par l'ensemble des familles ou au moins 3 sur 7 familles étudiées. Il en était de même pour l'analyse intra-famille ou aucun gène ne présentait des mutations partagées chez tous les cas au sein de la même famille. Nous avons aussi exclu l'implication de 16 gènes connus comme causatifs d'artériopathies en confirmant l'absence de mutations pour ces gènes au sein des familles étudiées. Au terme des analyses inter et intra familles, en l'absence de gènes majeurs mutés pour la DFM, nous avons sélectionné une liste de 22 gènes candidats enrichis en mutations rares à fort potentiel fonctionnel pour lesquels nous avons mené une étude d'association de variants rares génotypés

par la puce Illumina HumanExome en vue de valider leur implication dans la DFM.

Nous avons utilisé la méthode SKAT-O pour tester l'association de 851 variants rares polymorphes pour 20 gènes candidats sélectionnés sur 22 ; et 85 variants rares polymorphes pour 12 gènes sur 16 connus comme étant associés à des pathologies vasculaires. Les analyses ont été conduites pour deux seuils de MAF < 1% et MAF < 5% dans l'échantillon global (249 cas de DFM vs 689 témoins), chez les sujets avec le phénotype multifocal de la DFM (164 cas de DFM vs 689 témoins) et chez les sujets de sexe féminin uniquement (193 cas de DFM vs 344 témoins). Aucun des gènes étudiés ne présentait une association significative avec la DFM dans l'échantillon global. Une association nominale avec la DFM ($P < 0,05$) a par contre été observée pour les gènes *OBSCN*, *DYNC2H1 RNF213* et *MYLK* chez les individus avec le phénotype multifocal de la DFM, et pour les gènes *OBSCN*, *DYNC2H1* et *FSIP2* chez les femmes uniquement [cf. Tableau III-1]. Malheureusement, aucune de ces associations ne passe le seuil de significativité global ($P = 0.001$) après correction de Bonferroni pour tests multiples pour 32 gènes étudiés.

Tableau III-1. Etude d'association gène-entier de variants rares avec la DFM. Seuls les gènes présentant une association nominale avec la DFM chez les multifocaux et chez les femmes sont reportés.

		DFM Multifocaux				DFM Femmes			
		MAF < 0,05		MAF < 0,01		MAF < 0,05		MAF < 0,01	
Chr	Gènes	Variants	*P*	Variants	*P*	Variants	*P*	Variants	*P*
1	*OBSCN*	61	0,030	44	0,003	56	0,024	39	0,411
11	*DYNC2H1*	24	0,072	21	0,020	23	0,045	20	0,154
17	*RNF213*	38	0,069	35	0,010	34	0,739	31	0,368
2	*FSIP2*	26	0,060	24	0,408	26	0,038	24	0,357
3	*MYLK*	12	0,032	11	0,016	11	0,229	10	0,337

Exome Sequencing in Seven Families and Gene-Based Association Studies Indicate Genetic Heterogeneity and Suggest Possible Candidates for Fibromuscular Dysplasia

Short title: Exome sequencing in fibromuscular dysplasia

Authors: Soto Romuald KIANDO[a,b], Cristina BARLASSINA[c,d], Daniele CUSI[c,d], Pilar GALAN[e], Mark LATHROP[f], Pierre-François PLOUIN[b,g], Xavier JEUNEMAITRE[a,b,h], Nabila BOUATIA-NAJI[a,b]

Affiliations

[a]INSERM, UMR970 Paris Cardiovascular Research Center (PARCC), F-75015 Paris, FRANCE
[b]Université Paris-Descartes, PRES Sorbonne Paris Cité, Paris FRANCE
[c]Dept. of Health Sciences, Genomic and Bioinformatics Unit, Viale Ortles 22/4, Milano, ITALY
[d]Chair and Graduate School of Nephrology, University of Milano, Division of Nephrology, San Paolo Hospital, Milano, ITALY
[e]Université Paris 13, Equipe de Recherche en Epidémiologie Nutritionnelle (EREN), Centre d'Epidémiologie et Statistiques Sorbonne Paris Cité, Inserm (U1153), Inra (U1125), Cnam, COMUE Sorbonne Paris Cité, F-93017, Bobigny, FRANCE
[f]Centre National de Génotypage, Evry, FRANCE
[g]AP-HP, Department of Hypertension, Hôpital Européen Georges Pompidou, Paris, FRANCE
[h]AP-HP, Refferal Center for Rare Vascular Diseases, Hôpital Européen Georges Pompidou, Paris, France

Conflicts of interest

None

Corresponding Author

Nabila BOUATIA-NAJI, PhD

Paris Cardiovascular Research Center, INSERM UMR970 - Team 3

56 Rue Leblanc, F-75015, Paris, FRANCE

Email. nabila.bouatia-naji@inserm.fr

Phone. +330153987995

Fax. +33153987952

Abstract

Background: Fibromuscular dysplasia (FMD) is a nonatherosclerotic vascular disease leading to stenosis, aneurysm and dissection, mainly of renal arteries and carotids. FMD occurs predominantly in females with ~4/1000 prevalence and cause hypertension, renal ischemia or stroke. The pathogenesis of FMD is unknown and a genetic origin is suspected given its demonstrated familial aggregation.

Methods: We performed whole exome sequencing (WES) in 16 cases (seven families). Coding variants in 3,971 genes were prioritized on frequency (minor allele frequency <0.01) and *in silico* predicted functionality.

Results: No gene harbors variants that are shared among all affected members of at least three families. Variants from 16 genes of vascular and connective tissue diseases are excluded as causative in these families. Genes with at least four variants in the 16 patients and vascular genes were followed-up using genotypes from 249 unrelated cases and 689 controls. Gene-based association analyses using SKAT-O shows nominal significant association with multifocal FMD (N=164) for myosin light chain kinase (*MYLK, P*=0.01) previously involved in thoracic aortic aneurysm, obscurin *(OBSCN)*, a sarcomeric protein (*P*=0.003), dynein cytoplasmic heavy chain 1 (*DYNC2H1, P*=0.02) and *RNF213* previously associated with Moyamoya disease (*P*=0.01).

Conclusion: Our study indicates genetic heterogeneity and the unlikely existence of a major gene for FMD and excludes the role of several vascular genes in familial FMD. We also suggest four possible candidate genes for multifocal FMD, though these findings need further genetic and functional confirmation. More powerful WES and association studies (e.g. GWAS) will better decipher the genetic basis of FMD.

Keywords: Fibromuscular dysplasia; whole exome sequencing; gene-based association test.

INTRODUCTION

Fibromuscular dysplasia (FMD) is a group of nonatherosclerotic and noninflammatory vascular diseases leading to stenosis, aneurysm, dissection, and/or occlusion of medium-sized arteries, in particular the renal and extracranial cervical arteries [1-3]. Clinical manifestations of FMD are hypertension, dizziness, pulsatile tinnitus, transient ischemic attack or stroke, according to the involved arterial beds [2, 3].

Pathological classifications of FMD identify three main types of FMD affecting the intima, the media or the adventitia [4, 5]. Medial fibroplasia, which represents 80% of FMD cases, is characterized by alternating areas of thinned media and thickened medial collagen ridges creating the typical repeating pattern of arterial stenosis and poststenotic dilation that gives an aspect of “strings-of-beads” appearance on angiography. In the recent years, most patients with renal artery FMD are treated by percutaneous angioplasty rather than surgery [1]. Consequently, the pathological classification has progressively been replaced by a simple binary angiographic classification, with multifocal FMD characterized by the "string-of-beads” aspect and focal FMD characterized by a solitary stenosis on a given arterial segment [6]. Recent investigation from our group demonstrated that two distinct clinical phenotypes are discriminated using this angiographic classification. Patients with focal and multifocal lesions differed significantly in median age at diagnosis of FMD and hypertension, sex distribution, initial blood pressure, need for renal revascularization procedures and efficacy [7]. Intriguingly, a sex ratio is strongly deviated for multifocal FMD since 75% to 90% of patients are females [3, 8].
FMD prevalence in the general population is not known because lesions can be asymptomatic and the diagnosis requires invasive or semi-invasive (conventional, computed tomographic or magnetic resonance) angiography. The prevalence of renal artery stenosis associated with diagnosed FMD in patients with hypertension is estimated to ~4/1000 [2]. On the other hand,

the prevalence of asymptomatic FMD lesions was estimated to ~3% in a large sample of potential kidney donors undergoing computed tomographic angiography (N=2,640) [9] and as high as 5.8% in the CORAL trial supporting under-diagnosis bias for FMD [10].

The pathogenesis of FMD is unknown. Based on pathological studies and occurrence of FMD in several vascular and connective tissue diseases and syndromes, the loss of integrity of the extracellular matrix components (e.g. collagens, fibrillins) in the media and an alteration of TGFβ signaling are highly suspected. In support of this, a recent study reported features of connective tissue disease and the elevation of TGFβ-1 and TGFβ-2 in plasma from FMD patients [11].

We have previously reported in a retrospective analysis of 104 patients with renal FMD seven to 11% prevalence for familial FMD based on angiographic evidence performed in patients and their relatives [12]. Recent data from the US FMD registry are consistent with the French study where ~7% of FMD patients reported a confirmed diagnosis of FMD among a family member [3]. This data motivated the initiation of a genetic investigation for FMD for which only small case series or few candidate gene studies have been reported so far [13-15].

Here we report the whole exome sequencing data from 16 familial cases of FMD that show the lack of evidence for the existence of a major causative gene of FMD and exclude the role of 16 known vascular diseases and connective tissue disorders genes in familial FMD etiology. Using genotypes generated by the exome chip [16] in 249 FMD cases and 689 controls, we provide some evidence of association for *MYLK*, *OBSCN*, *DYNC2H1* and *RNF213*, which suggest these genes as putative candidates for multifocal FMD.

MATERIALS AND METHODS

Familial cases

FMD cases were recruited from the French database of the Rare Vascular Diseases Referral Centre of the European Hospital Georges Pompidou (HEGP), Paris, France. An expert panel using clinical information from the medical history, the examination and the interview of the patient, the interpretation of angiography and / or computed tomography scan defined the FMD diagnosis after the exclusion of other causes of arterial stenosis (mainly atherosclerosis, Takaysu disease and Elhers Danlos syndrome). Familial FMD cases were ascertained as patients with at least one first-degree relative with confirmed FMD as previously reported [17]. We identified seven pedigrees (Fig. 1) from which we selected at least two patients per pedigree with DNA quality and quantity compatible with exome sequencing experiments. Selected patients are sibs in all families, except in Family VII where they are first cousins (Fig. 1). Intra-family controls could not be identified, as angiography explorations in healthy members of these families were not available. The study was approved by local institutional ethics committee (Approval CCPPRB Paris-Cochin # RBM 00–028) and all subjects gave written informed consent.

Case-control association study

Patients were 309 unrelated FMD cases (80 % females and 95% had hypertension) ascertained also from the Rare Vascular Diseases Referral Centre, HEGP. The mean age was 44.9±15.8 years for the diagnosis of FMD and 36.9±14.4 years for hypertension. Controls were 699 individuals ascertained from the SU.VI.MAX study, a national sample of healthy volunteers living in France and enrolled between 1996 and 2001 in a randomized, placebo-controlled trial testing the benefit of anti-oxydants nutrients on the incidence of cancers and cardiovascular diseases, as detailed previously [18].

Exome sequencing, assembly, variant calling and bioinformatics analyses

Exome sequencing, assembly, variant calling and bioinformatics treatment were performed at the IntegraGen Company (Evry, France). Sequence capture, enrichment and elution were performed according to manufacturer's instruction and protocols (SureSelect, Agilent). Briefly, 3 µg of each genomic DNA were fragmented by sonication and purified to yield fragments of 150-200 bp. Paired-end adaptor oligonucleotides from Illumina were ligated on repaired, A-tailed fragments then purified and enriched by six PCR cycles. Five hundred nanograms of the purified libraries were then hybridized to the SureSelect oligo probe capture library for 24 hours. After hybridization, washing, and elution, the eluted fraction was PCR-amplified with 10 to 12 cycles, purified and quantified by quantitative PCR to obtain enough DNA template for subsequent applications. Each eluted-enriched DNA sample was then sequenced on an Illumina HiSeq 2000 as paired-end 75b reads. Image analysis and base calling was performed using Illumina Real Time Analysis (RTA) Pipeline version 1.14 with default parameters.

Bioinformatics analyses of sequencing data were based on the Illumina pipeline (CASAVA1.8), which performs alignment (algorithm ELANDv2) of a sequencing run to a reference genome (hg19), calls the SNPs based on the allele calls and read depth, and detects variants. Only the positions included in the bait coordinates were conserved. Genetic variation annotation was performed by IntegraGen in-house pipeline by gene annotation to the reference genome (RefSeq), identification of known polymorphisms (dbSNP 132 and 1000 Genomes Project [19]) followed by a mutation characterization (exonic / intronic, synonymous/ missense / silent / nonsense). For each position, the exomic frequencies (homozygote and heterozygote) were determined from all existing exomes sequenced by the company IntegraGen and the data available in HapMap [20] to filter out recurrent spurious variants.

Genotyping and quality controls

We used pre-existing genotypes of rare and common variants that were generated for the purposes of an ongoing association study using Illumina HumanExome-12v1_A Beadchip. We analyzed 249 unrelated FMD cases and 689 controls after exclusion of individuals with non-European ancestry (principal components analysis (PCA) with HapMap phase 3 samples using SMARTPCA [21]) and call rate per individual < 99%. We also excluded SNPs with call rates < 95% and a significant deviation from the Hardy-Weinberg Equilibrium ($P < 10^{-5}$) in cases and/or controls.

Statistical analyses

We used SKAT-O [22] to test the gene-based association with FMD, which encompasses burden tests and SKAT [23] as special cases. SKAT-O uses the kernel machine methods to test for association between a set of rare (and common) variants and continuous or binary phenotype [23]. As the underlying biological mechanisms are unknown for FMD and may vary from one gene to another across the genome, SKAT-O has been shown to perform well under a range of scenarios, including variants present in the same gene are null, protective and deleterious, or a large number of variants are causal and associated in the same direction [22]. Default weights were used to perform the analyses using two MAF thresholds (MAF < 0.01 and MAF < 0.05) for variants in the genes covered by the exome chip with at least two successfully genotyped variants. We also performed single SNP association test of common variants (MAF ≥ 0.05) in the same set of genes by logistic regression under the additive model using PLINK [24].

RESULTS

Exome sequencing and variant prioritization

FMD familial cases were predominantly females, affected in the renal arterial bed with the multifocal phenotype (Fig. 1, Table1). All patients except one had hypertension, with age of onset as early as 13 years (Patient VII.2) and ten are reported to be current or former smokers (Table1). Descriptive of the sequencing quality results is displayed in Supp. Table 1. We generated an average of 5.27 gigabases of sequence per patient. The mean depth of coverage of the targeted regions was 71.7X across all samples, with on average 93% of all targeted regions being covered by ≥10X reads and 83% by 20X reads (Supp. Table 1). We identified a mean of total variants of 63,127 per patient, including on average 20,521 SNPs and 395 Indels (Supp. Table 1). We decided to focus on SNPs because more confidently called in WES than indels and applied a prioritization strategy to identify candidate genes for causality in familial FMD (Fig. 2). With a minimal depth threshold of 20X, we identified 16,568 variants in 8,595 genes from the 16 patients. We followed up 3,981 variants in 3,971 different genes, which were coding (missense, nonsense or splice), rare (MAF ≤ 0.01 according to public databases, mainly 1000 Genomes Project [19] and/or Exome Variant Server (EVS) [25]) and *in silico* predicted to be damaging by PolyPhen-2 [26].

Inter-families analyses

Affected sibs share 50% of their genetic variants by chance. However, if several genetic variant are shared by sibs from several independent families and are located in the same gene, one can strongly argue for genetic causality for this gene. We fixed an arbitrary threshold of three families to share variants at the same gene to declare a gene as putative candidate for FMD. Unfortunately, none of 3,971 genes full filled this condition (Fig. 2), which suggests that it is unlikely that a major and unique gene could explain familial FMD, at least in our set of families.

This result indicates that FMD presents important genetic heterogeneity and/or a polygenic mode of inheritance.

Intra-families analyses

FMD shares many clinical, histological and potentially some biological features with several vascular and connective tissue diseases.[11] Mutated molecules of collagen and fibrillin, cause stenosis and dissection, as does impaired TGFβ signaling in Elhers-Danlos and Loeys-Dietz syndrome, for instance. We specifically analyzed the sequencing data for 16 candidate genes that were previously and firmly identified as causative in genetically mediated vascular and connective tissue disorders as reported previously [11, 27]. In the intra-families analyses (Fig. 2), we hypothesized rare and predicted functional variants among these genes to be causative in some of the FMD families. Thus, we looked for shared heterozygous or homozygous variants by all sibs from the same family. This analysis was also unrevealing and allowed us to exclude 16 vascular and connective tissue genes as causative for familial FMD in our patients (Fig. 2).

Gene-based and single common variants association analyses

Here we hypothesize the existence of genes with recurrent rare coding variants among FMD cases, despite a lack of segregation among sibs. If this hypothesis is true, these genes should harbor more variants if we analyze a larger sample of FMD cases, compared to only 16 familial patients. To verify this hypothesis, we tested the association with FMD of genes that harbor multiple variants using a case control study of 249 FMD cases and 689 controls.

We identified 22 genes with at least four rare variants according to the WES data (Fig. 2), as recommended in [28] to attend 80% power from 16 exomes. Supp. Table2 includes the number of coding variants per amino-acid change and MAF categories, those present in the exome chip and those polymorphic in our samples. We focused the analysis on missense and nonsense

variants that are likely to be functional and analyzed 851 variants that mapped in genes identified by WES and 85 variants in vascular and connective tissue disorders causative genes (Supp. Table 2). Most of the WES genes were highly polymorphic in the cases and the controls, except two genes (*BDAG1* and *LINC00955*) that had less than two variants and could not be analyzed by SKAT-O.

MUC16 on chromosome 19 presents the largest number of variants (98 with MAF < 0.05 and 81 with MAF < 0.01) and *LRRC30* the lowest number of variants (N=4). In addition to the association of the global sample of patients, we also tested the association of sequencing genes with multifocal FMD, as the major angiographic phenotype of FMD in our sample (80% of patients), and performed an analysis restricted to females (193 cases *vs.* 344 controls) to explore a putative sex specific association of some genes (Table2).

No gene associates with FMD in the global sample, for both MAF thresholds (Table 2). In the multifocal sample, the obscurin, cytoskeletal calmodulin and titin-interacting RhoGEF encoding gene (*OBSCN*) shows some evidence of association with FMD both for variants with MAF < 0.05 (number of variants=61; *P*=0.03) and MAF < 0.01 (number of variants=44; *P*=0.003). We also obtained significant association with multifocal FMD at the MAF < 0.01 for *DYNC2H1* encoding the dynein, cytoplasmic 2, heavy chain 1 (number of variants=21; *P*=0.02) and *RNF213* encoding ring finger protein 213 (number of variants=35; *P*=0.01). In the female-only analyses, in addition to *OBSCN* (number of variants=56; *P*=0.024) and *DYNC2H1* (number of variants=23; *P*=0.045), we also observed significant association with FMD at MAF < 0.05 for *FSIP2* encoding fibrous sheath interacting protein 2 (number of variants=26; *P*=0.038), but not under MAF < 0.01 (number of variants=24; *P*=0.357). The analyses of 12 vascular and connective tissue disorders genes, for which genotypes were available for at least two variants per gene, showed significant association with multifocal FMD for myosin light chain kinase (*MYLK*) at MAF < 0.05 (number of variants=12; *P*=0.032) and MAF < 0.01 (number of variants=11; P=0.016), though non significant

after Bonferroni correction (Table 3). We note that none of the above associations are significant after a Bonferroni corrected level of significance (P=0.001) that takes into account multiple testing of 32 genes.

The analyses of the association of 203 common (MAF > 0.05) variants with genotypes available from the exome chip in 20 WES genes and 12 vascular genes were non significant overall (Supp. Table 3).

DISCUSSION

Here we describe the first whole exome sequencing (WES) study in fibromuscular dysplasia (FMD), a nonatherosclerotic arterial disease suspected to have a genetic origin. The key finding of our study is the indication of the lack of a major gene for familial FMD based on a hypothesis-free genetic approach conducted in 16 patients belonging to seven different families and presenting predominantly renal multifocal FMD associated with hypertension. We also suggest gene-based nominal association ($P < 0.05$) with multifocal FMD for rare coding variants in *MYLK*, *OBSCN*, *DYNC2H1* and *RNF213* and using the largest case control study described so far for FMD, which included 249 FMD patients and 689 population-based controls.

In the WES data, we mined a large number of high quality, rare coding and likely to be functional variants that belongs to nearly 4000 genes. Our WES analyses identified no gene with rare genetic variations shared among all sibs of the same family and involving at least three different families. Given the small number of families we studied, we believed that an arbitrary threshold of at least three families sharing the variants at the same gene was a good compromise to identify high profile candidate genes for causality. This relatively stringent criterion aimed to compensate the lack of intra-families controls or linkage information, and the close relatedness among the patients who were predominantly sibs ascertained from small pedigrees. We note that the less stringent strategy not to filter out for known variants (e.g. dbSNP and EVS) was also unsuccessful. This recommendation is especially valid for diseases with relatively high prevalence (4/1000 for clinical FMD, ~3% for asymptomatic FMD) because of the constantly increasing number of pathogenic variants reported in these public databases [29]. Our unrevealing results could also be explained by incomplete penetrance of the causative genes in FMD that we are likely to miss in this study design [30]. In this case, our small study (seven families) is lacking statistical power to identify a causative gene, especially if FMD is suspected to be a highly heterogeneous disease both at the clinical and the genetic levels.

We also excluded the role of 16 genes involved in several connective tissue syndromes and rare vascular diseases as rare coding and putatively functional variants in these genes did not segregate with FMD in any of the seven families studied (intra-families analyses). Our WES data are consistent with previous data using direct sequencing reported for an overlapping list of these genes in 47 FMD patients [11]. Interestingly, this study described deep clinical exploration that showed that FMD patients presented in high proportion connective tissue features such as early onset degenerative spine disease, dural ectasia and mild connective tissue dysplasia [11]. In addition, this study reported elevated circulating TGF-β1 and TGF-β2 and in media secreted by dermal fibroblasts from FMD patients, which suggests that the expression alteration of this pathway may be at the origin of these clinical features in FMD patients, similarly to clinical characteristics of patients with Loeys-Dietz and Marfan syndromes [11]. Therefore, causative genes for FMD could be yet to be identified modifiers of the TGF-β signaling and connective tissue components regulators.

In the absence of segregating rare variants in our families, we did not possess valid arguments to carry out direct sequencing for any of the genes identified by WES, despite a large number of variants observed in some genes (e.g. *TTN* with 10 variants). In addition, many of these genes are relatively long and are more likely to be polymorphic by chance (e.g. *MUC16, MUC6, CFTR* and *TTN*) as confirmed by data from public databases of WES data (e.g. EVS). On the other hand, no clear high profile candidate gene based on existing literature about artery biology or development was noticed to encourage gene prioritization. Therefore, as a follow-up investigation of the WES experiments, we studied genotypes generated in a relatively large sample of 249 FMD cases and 689 controls. It is true that genotyping presents a strong limitation as it only covers genetic information of already known variants, compared to sequencing that generates genetic variation specific to the sample studied. Nevertheless, the genotyping information generated at this stage provided critical information that is the presence of rare

variants in a consequent sample of controls for all genes tested, as it is the case from the EVS database. Consequently, this result discards all WES candidates as putative causative genes for FMD, under a full penetrant dominant model for FMD being a monogenic disease as previously suspected [17, 31].

Under the hypothesis of strong genetic heterogeneity and a more polygenic and complex pattern of inheritance for FMD, we complemented our genetic investigation by the study of the association with FMD of 20 genes identified by WES and presenting multiple rare variants and 12 vascular and connective tissue genes. Despite a lack of segregation in the families studied, these genes could be more enriched for rare and coding variants compared to controls and thus increase the risk of FMD. The gene-based association analyses indicate for four different genes (*MYLK, OBSCN*, *DYNC2H1* and *RNF213*) some evidence of association with FMD, especially with the multifocal form, though not significant after adjustment for multiple testing.

MYLK is one of the vascular genes previously involved in familial thoracic aortic dissection [32]. *MYLK* encodes myosin light chain kinase (MLCK) implicated in cell contractile activity and is expressed in smooth muscle cells (SMCs), in which a specific knockdown produced medial degeneration of the aorta in mice [32]. This invalidation cause also important loss of tonic airway smooth muscle contraction [33] and gastrointestinal motility [34]. This data supports important implication of MLCK in the contraction of different arterial beds. In combination with our results of the association of genetic variants in *MYLK* with multifocal FMD, a putative role of MCLK in the contraction of SMCs in the media of renal arteries could be hypothesized. Further specific exploration of renal arteries in SMCs specific knock out mice is needed to test this hypothesis.

OBSCN codes obscurin, which includes several signaling domains (>67 Immunoglobulin domains, 2 fibronectin domains, 1 calcium / calmodulin-biding domain, 1 titin-interacting RhoGEF and 2 serine-threonine domains) [35]. Obscurin is a giant protein and is member of the family of sarcomeric proteins that includes titin and nebulin [35]. *OBSCN* is specifically expressed in

cardiac and skeletal muscle [35] giving minor credit to its putative implication in the maintenance of artery wall, especially at the media, which undergoes important smooth muscle cells and extracellular matrix disorganization.

We also observed significant association with multifocal FMD for *DYNC2H1* encoding dynein cytoplasmic heavy chain 1 (Table 2). Dynein is a large protein involved in intra-flagella transport, a process required for cilia and flagella assembly. To our knowledge, FMD is not reported in the clinical presentation of this rare syndrome and a mechanistic link between cilia and flagella assembly and the establishment of medial fibroplasia is rather speculative at this stage.

Among the associated genes with multifocal FMD, we highlight *RNF213* a strong susceptibility gene for Moyamoya disease. This rare vascular disease is characterized by the progressive occlusion of intracranial carotid arteries, which presents similarities with FMD, particularly, by the absence of inflammation and the increased fibrosis with thickening of arterial walls and high prevalence in females [36]. Following linkage and GWAS, rs112735431, a rare coding variant in *RNF213* (also reported as p.R4859K and c.14576G>A) was strongly associated with Moyamoya disease in Japanese and Korean populations where this disease is ~10 times more prevalent than in Europeans [37, 38]. rs112735431 was not present in the exome array that generated genotypes in our FMD case control sample and thus was not part of the variants that associated with FMD in our study. Using direct genotyping, we did not observe the variant allele of rs112735431 in 249 FMD patients (data not shown), which was consistent with the EVS public database where this variant is also monomorphic in more than 6500 Europeans. *RNF213* codes a protein containing a RING domain with a special Zn-Finger domain that binds two zinc molecules theoretically involved in protein-protein interactions. *RNF213* function is not fully elucidated and clear implication of *RNF213* in cerebrovascular pathology is missing [37, 38].

Additional genetic and functional studies are needed to better elucidate this putative common genetic basis between FMD and Moyamoya disease.

In summary, our study reports data from the first WES investigation conducted for familial forms of FMD and supports the lack of major genetic determinant for this rare disease under a classical Mendelian mode of inheritance. We also excluded the role of rare variants from several known vascular and connective tissue genes in our families. The gene-based association data provide some evidence of association with multifocal FMD for four genes *MYLK, OBSCN*, *DYNC2H1* and *RNF213*. Despite the lack of robust statistical significance, these findings deserve highlight, especially for *MYLK* and RNF213 for which either functional or genetic links are plausible. Further genetic replication and functional investigation will be needed to connect these genes and their protein products with renal artery stenosis that leads to FMD.

ACKNOWLEDGMENTS

We thank all patients who participate in this study. We also thank Laurence Amar, Juliette Albuisson, Michel Azizi, Guillaume Bobrie, Antoine Chédid, Béatrice Fiquet-Kempf, Michael Frank and Anne-Paule Gimenez-Roqueplo for the recruitment of patients and families and discussion on clinical aspects. We thank Christian Dina for fruitful statistical discussion, Emmanuel Martin and Mélanie Letexier from the IntegraGen Company for assistance with the exome sequencing data preparation. We also acknowledge Patrick Bruneval for discussion on the pathology of FMD.

FUNDING SOURCES

This study was supported in part by a European program *FP7 HYPERGENES (FP7 - HEALTH-F4-2007-201550) an Italian joint agency grant (INTEROMICS PB05 - MIUR - CNR Italian Flagship Project),* two French National research agency grants *(ANR GDPM2 and ANR-13-ISV1-0006-0),* a National French clinical grant *(PHRC ARCADIA- PROFILE)* and a *Fondation pour la recherche Médicale support (FRM MEDYA).*

REFERENCES

1. Olin, J.W. and B.A. Sealove, *Diagnosis, management, and future developments of fibromuscular dysplasia.* J Vasc Surg, 2011. **53**(3): p. 826-36 e1.
2. Plouin, P.F., J. Perdu, A. La Batide-Alanore, P. Boutouyrie, A.P. Gimenez-Roqueplo, and X. Jeunemaitre, *Fibromuscular dysplasia.* Orphanet J Rare Dis, 2007. **2**: p. 28.
3. Olin, J.W., J. Froehlich, X. Gu, J.M. Bacharach, K. Eagle, B.H. Gray, et al., *The United States Registry for Fibromuscular Dysplasia: results in the first 447 patients.* Circulation, 2012. **125**(25): p. 3182-90.
4. Harrison, E.G., Jr. and L.J. McCormack, *Pathologic classification of renal arterial disease in renovascular hypertension.* Mayo Clin Proc, 1971. **46**(3): p. 161-7.
5. Stanley, J.C., B.L. Gewertz, E.L. Bove, V. Sottiurai, and W.J. Fry, *Arterial fibrodysplasia. Histopathologic character and current etiologic concepts.* Arch Surg, 1975. **110**(5): p. 561-6.
6. Persu, A., A. Giavarini, E. Touze, A. Januszewicz, M. Sapoval, M. Azizi, et al., *European consensus on the diagnosis and management of fibromuscular dysplasia.* J Hypertens, 2014. **32**(7): p. 1367-78.
7. Savard, S., O. Steichen, A. Azarine, M. Azizi, X. Jeunemaitre, and P.F. Plouin, *Association between 2 angiographic subtypes of renal artery fibromuscular dysplasia and clinical characteristics.* Circulation, 2012. **126**(25): p. 3062-9.
8. Cragg, A.H., T.P. Smith, B.H. Thompson, T.P. Maroney, A.W. Stanson, G.T. Shaw, et al., *Incidental fibromuscular dysplasia in potential renal donors: long-term clinical follow-up.* Radiology, 1989. **172**(1): p. 145-7.
9. McKenzie, G.A., G.S. Oderich, A. Kawashima, and S. Misra, *Renal artery fibromuscular dysplasia in 2,640 renal donor subjects: a CT angiography analysis.* J Vasc Interv Radiol, 2013. **24**(10): p. 1477-80.
10. Hendricks, N.J., A.H. Matsumoto, J.F. Angle, A. Baheti, S.S. Sabri, A.W. Park, et al., *Is fibromuscular dysplasia underdiagnosed? A comparison of the prevalence of FMD seen in CORAL trial participants versus a single institution population of renal donor candidates.* Vasc Med, 2014.
11. Ganesh, S.K., R. Morissette, Z. Xu, F. Schoenhoff, B.F. Griswold, J. Yang, et al., *Clinical and biochemical profiles suggest fibromuscular dysplasia is a systemic disease with altered TGF-beta expression and connective tissue features.* FASEB J, 2014.
12. Pannier-Moreau, I., P. Grimbert, B. Fiquet-Kempf, A. Vuagnat, X. Jeunemaitre, P. Corvol, et al., *Possible familial origin of multifocal renal artery fibromuscular dysplasia.* J Hypertens, 1997. **15**(12 Pt 2): p. 1797-801.
13. Marks, S.D., A.M. Gullett, E. Brennan, K. Tullus, G. Jaureguiberry, E. Klootwijk, et al., *Renal FMD may not confer a familial hypertensive risk nor is it caused by ACTA2 mutations.* Pediatr Nephrol, 2011. **26**(10): p. 1857-61.
14. Perdu, J., A.P. Gimenez-Roqueplo, P. Boutouyrie, S. Beaujour, B. Laloux, V. Nau, et al., *Alpha1-antitrypsin gene polymorphisms are not associated with renal arterial fibromuscular dysplasia.* J Hypertens, 2006. **24**(4): p. 705-10.
15. Poloskey, S.L., E. Kim, R. Sanghani, A.H. Al-Quthami, P. Arscott, R. Moran, et al., *Low yield of genetic testing for known vascular connective tissue disorders in patients with fibromuscular dysplasia.* Vasc Med, 2012. **17**(6): p. 371-8.
16. Exome Chip Design. *Exome Chip Design [cited September 10, 2014]. http://genome.sph.umich.edu/wiki/Exome_Chip_Design.* 2014.

17. Perdu, J., P. Boutouyrie, C. Bourgain, N. Stern, B. Laloux, E. Bozec, et al., *Inheritance of arterial lesions in renal fibromuscular dysplasia.* J Hum Hypertens, 2007. **21**(5): p. 393-400.
18. Hercberg, S., P. Galan, P. Preziosi, S. Bertrais, L. Mennen, D. Malvy, et al., *The SU.VI.MAX Study: a randomized, placebo-controlled trial of the health effects of antioxidant vitamins and minerals.* Arch Intern Med, 2004. **164**(21): p. 2335-42.
19. Genomes Project, C., G.R. Abecasis, D. Altshuler, A. Auton, L.D. Brooks, R.M. Durbin, et al., *A map of human genome variation from population-scale sequencing.* Nature, 2010. **467**(7319): p. 1061-73.
20. International HapMap, C., D.M. Altshuler, R.A. Gibbs, L. Peltonen, D.M. Altshuler, R.A. Gibbs, et al., *Integrating common and rare genetic variation in diverse human populations.* Nature, 2010. **467**(7311): p. 52-8.
21. Patterson, N., A.L. Price, and D. Reich, *Population structure and eigenanalysis.* PLoS Genet, 2006. **2**(12): p. e190.
22. Lee, S., M.C. Wu, and X. Lin, *Optimal tests for rare variant effects in sequencing association studies.* Biostatistics, 2012. **13**(4): p. 762-75.
23. Wu, M.C., S. Lee, T. Cai, Y. Li, M. Boehnke, and X. Lin, *Rare-variant association testing for sequencing data with the sequence kernel association test.* Am J Hum Genet, 2011. **89**(1): p. 82-93.
24. Purcell, S., B. Neale, K. Todd-Brown, L. Thomas, M.A. Ferreira, D. Bender, et al., *PLINK: a tool set for whole-genome association and population-based linkage analyses.* Am J Hum Genet, 2007. **81**(3): p. 559-75.
25. NHLBI Exome Sequencing Project. *Exome Variant Server, NHLBI GO Exome Sequencing Project (ESP), Seattle, WA [April, 2016 accessed].* 2016; Available from: http://evs.gs.washington.edu/EVS/.
26. Adzhubei, I., D.M. Jordan, and S.R. Sunyaev, *Predicting functional effect of human missense mutations using PolyPhen-2.* Curr Protoc Hum Genet, 2013. **Chapter 7**: p. Unit7 20.
27. B. Loeys, D.P., G. Vandeweyer, S. Salemink, M. Kempers, G. Oswald, H. Dietz, G. Mortier, L. Van Laer, *Surprising clinical lessons from targeted next generation sequencing of thoracic aortic aneurysmal genes [Abstract].* ASHG 2013 Meeting.
28. Boyden, L.M., M. Choi, K.A. Choate, C.J. Nelson-Williams, A. Farhi, H.R. Toka, et al., *Mutations in kelch-like 3 and cullin 3 cause hypertension and electrolyte abnormalities.* Nature, 2012. **482**(7383): p. 98-102.
29. Bamshad, M.J., S.B. Ng, A.W. Bigham, H.K. Tabor, M.J. Emond, D.A. Nickerson, et al., *Exome sequencing as a tool for Mendelian disease gene discovery.* Nat Rev Genet, 2011. **12**(11): p. 745-55.
30. Wang, Z., X. Liu, B.Z. Yang, and J. Gelernter, *The role and challenges of exome sequencing in studies of human diseases.* Front Genet, 2013. **4**: p. 160.
31. Rushton, A.R., *The genetics of fibromuscular dysplasia.* Arch Intern Med, 1980. **140**(2): p. 233-6.
32. Wang, L., D.C. Guo, J. Cao, L. Gong, K.E. Kamm, E. Regalado, et al., *Mutations in myosin light chain kinase cause familial aortic dissections.* Am J Hum Genet, 2010. **87**(5): p. 701-7.
33. Zhang, W.C., Y.J. Peng, G.S. Zhang, W.Q. He, Y.N. Qiao, Y.Y. Dong, et al., *Myosin light chain kinase is necessary for tonic airway smooth muscle contraction.* J Biol Chem, 2010. **285**(8): p. 5522-31.
34. He, W.Q., Y.J. Peng, W.C. Zhang, N. Lv, J. Tang, C. Chen, et al., *Myosin light chain kinase is central to smooth muscle contraction and required for gastrointestinal motility in mice.* Gastroenterology, 2008. **135**(2): p. 610-20.

35. Young, P., E. Ehler, and M. Gautel, *Obscurin, a giant sarcomeric Rho guanine nucleotide exchange factor protein involved in sarcomere assembly.* J Cell Biol, 2001. **154**(1): p. 123-36.
36. Fujimura, M., S. Sonobe, Y. Nishijima, K. Niizuma, H. Sakata, S. Kure, et al., *Genetics and Biomarkers of Moyamoya Disease: Significance of RNF213 as a Susceptibility Gene.* J Stroke, 2014. **16**(2): p. 65-72.
37. Liu, W., D. Morito, S. Takashima, Y. Mineharu, H. Kobayashi, T. Hitomi, et al., *Identification of RNF213 as a susceptibility gene for moyamoya disease and its possible role in vascular development.* PLoS One, 2011. **6**(7): p. e22542.
38. Kamada, F., Y. Aoki, A. Narisawa, Y. Abe, S. Komatsuzaki, A. Kikuchi, et al., *A genome-wide association study identifies RNF213 as the first Moyamoya disease gene.* J Hum Genet, 2011. **56**(1): p. 34-40.

Table 1. Clinical features of the 16 familial cases with fibromuscular dysplasia ascertained for exome sequencing.

			FMD		Hypertension		Smoking
Patient ID	**Sex**	**Arterial bed**	**Angiographic phenotype**	**Age at diagnosis[a]**	**Status[b]**	**Age at diagnosis[a]**	**Status[b]**
I.1[c]	F	Carotid	Multifocal	63	1	29	0
I.2	M	Vertebral	NA	56	0	-	1
I.3	F	Renal	NA	66	1	49	0
II.1[c]	F	Renal	Multifocal	61	1	51	1
II.2	F	Vertebral	NA	53	1	47	1
III.1*	F	Carotid	Multifocal	51	1	45	0
III.2	M	Renal	NA	63	1	45	1
IV.1[c]	F	Carotid	Multifocal	54	1	50	1
IV.2	F	Renal	NA	60	1	NA	1
V.1[c]	F	Renal	Multifocal	32	1	13	1
V.2	F	Renal	Unifocal	21	1	21	1
V.3	F	Renal	Unifocal	41	1	33	1
VI.1[c]	F	Iliac and carotid	Multifocal	51	1	36	0
VI.2	F	Carotid	Multifocal	55	1	40	0
VII.1[c]	F	Carotid	NA	41	1	42	1
VII.2	M	Renal	NA	30	1	13	0

FMD, fibromuscular dysplasia; NA, not available.

[a]Age at diagnosis in years;

[b]1, yes; 0, no.

[c]Index cases.

Table 2. Gene-based association results with FMD, multifocal phenotype and in females only for 20 exome sequencing candidate genes

		All Patients (249 cases *vs.* 689 controls)				Multifocal (164 cases vs. 689 controls)				Females (193 cases *vs.* 344 controls)			
		MAF < 0.05		MAF < 0.01		MAF < 0.05		MAF < 0.01		MAF < 0.05		MAF < 0.01	
Chr	**Gene**	**Number of variants**	***P***	**Number of variants**	***P***	**Number of variants**	***P***	**Number of variants**	***P***	**Number of variants**	***P***	**Number of variants**	***P***
1	*OBSCN*	64	0.184	47	0.118	61	0.030	44	0.003	56	0.024	39	0.411
2	*COL6A3*	28	0.876	25	0.995	27	0.870	24	0.671	22	0.317	19	0.240
2	*FSIP2*	26	0.055	24	0.331	26	0.060	24	0.408	26	0.038	24	0.357
2	*TANC1*	11	0.507	10	0.565	11	0.995	10	0.997	11	0.721	10	0.608
2	*TTN*	97	0.871	77	0.156	92	0.487	72	0.264	87	0.857	67	0.318
3	*GBE1*	7	0.281	6	0.163	7	0.337	6	0.675	5	0.180	4	0.256
4	*FAT1*	40	0.487	34	0.335	38	0.348	32	0.324	38	0.732	32	0.459
5	*CMYA5*	34	0.731	32	0.492	33	0.462	31	0.340	31	0.799	29	0.579
6	*BCLAF1*	9	0.780	7	0.792	9	0.529	7	0.285	8	0.333	6	0.784
7	*CFTR*	25	0.838	22	0.474	23	0.607	20	0.419	21	0.470	18	0.461
7	*SSPO*	44	0.674	39	0.492	43	0.482	38	0.262	42	0.787	37	0.565
8	*PLEC*	29	0.985	23	0.215	26	0.853	20	0.321	23	0.835	17	0.386
8	*RECQL4*	10	0.878	9	0.632	9	0.179	8	0.136	9	0.084	8	0.083
10	*TTC40*	7	0.666	5	0.479	7	0.412	5	0.076	7	0.668	5	0.576
11	*DYNC2H1*	25	0.067	22	0.117	24	0.072	21	0.020	23	0.045	20	0.154
11	*MUC6*	15	0.371	12	0.121	15	0.510	12	0.225	13	0.532	10	0.185
16	*HYDIN*	41	0.465	33	0.796	40	0.295	32	0.610	40	0.725	32	0.793
17	*RNF213*	40	0.681	37	0.295	38	0.069	35	0.010	34	0.739	31	0.368
18	*LRRC30*	4	0.262	4	0.233	4	0.220	4	0.225	4	0.172	4	0.187
19	*MUC16*	98	0.620	81	0.232	96	0.685	79	0.222	84	0.300	67	0.127

We note that none of these associations in significant according to an adjusted P value for multiple testing (*P*= 0.001). MAF, minor allele frequency.

Table 3. Gene-based association results with FMD, multifocal phenotype and in females only for 12 vascular diseases and connective tissue disorders genes.

		All Patients (249 cases *vs.* 689 controls)				Multifocal (164 cases vs. 689 controls)				Females (193 cases *vs.* 344 controls)			
		MAF < 0.05		MAF < 0.01		MAF < 0.05		MAF < 0.01		MAF < 0.05		MAF < 0.01	
Chr	**Gene**	**Number of variants**	***P***	**Number of variants**	***P***	**Number of variants**	***P***	**Number of variants**	***P***	**Number of variants**	***P***	**Number of variants**	***P***
1	*PLOD1*	6	0.419	5	0.716	6	0.310	5	0.645	5	0.884	4	0.996
1	*SKI*	2	0.105	2	0.112	2	0.082	2	0.083	2	0.107	2	0.104
1	*TGFB2*	2	0.170	2	0.171	2	0.152	2	0.162	2	0.174	2	0.178
2	*COL3A1*	6	0.208	6	0.219	6	0.721	6	0.709	6	0.642	6	0.594
3	*MYLK*	13	0.125	12	0.122	12	**0.032**	11	**0.016**	11	0.229	10	0.337
3	*TGFBR2*	4	0.064	4	0.064	3	0.076	3	0.076	4	0.111	4	0.111
9	*COL5A1*	13	0.837	11	0.756	13	0.206	11	0.272	12	0.776	10	0.295
9	*NOTCH1*	11	0.429	10	0.099	10	0.628	9	0.180	11	0.097	10	0.062
15	*FBN1*	5	0.361	5	0.341	5	0.115	5	0.124	4	0.261	4	0.245
16	*MYH11*	4	0.405	4	0.389	4	0.678	4	0.681	4	0.799	4	0.799
20	*SLC2A10*	6	0.282	5	0.430	6	0.111	5	0.191	6	0.456	5	0.376
X	*FLNA*	7	0.576	6	0.448	7	0.395	6	0.421	7	0.350	6	0.355

We note that none of these associations in significant according to an adjusted P value for multiple testing (*P*= 0.001). MAF, minor allele frequency.

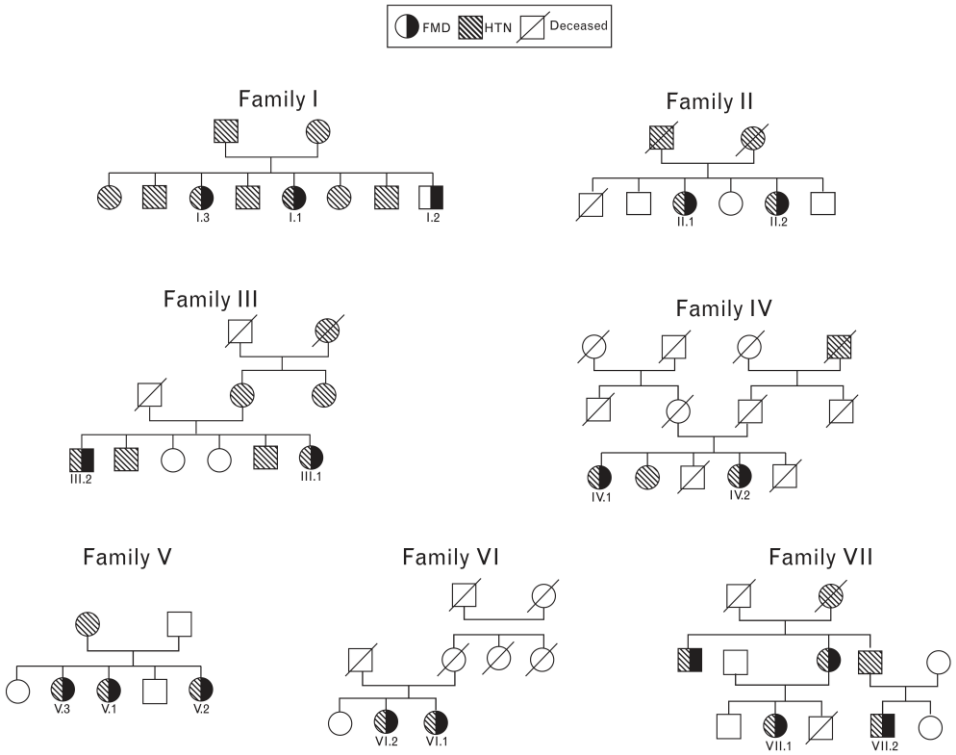
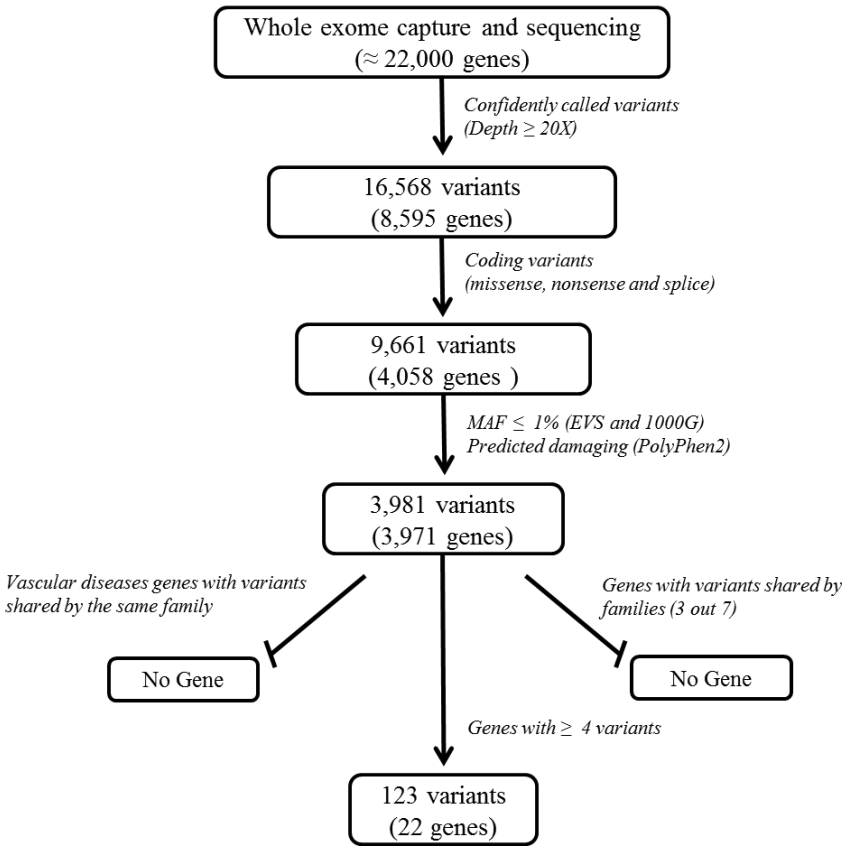


FIGURE 1. Pedigrees of the seven families from which fibromuscular dysplasia patients (numbers indicated) were ascertained for exome sequencing.



(*) ACTA2, COL3A1, COL5A1, EFEMP2, FBN1, FLNA, MYH11, MYLK, NOTCH1, SKI, SLC2A10, SMAD3, TGFB2, TGFBR1, TGFBR2, PLOD1.

FIGURE 2. Filtration strategy to identify candidate genes for causality in familial fibromuscular dysplasia.

Supp. Table 1. Descriptive of the sequencing quality results in the 16 familial FDM cases.

Patient ID	SampleYield (Gigabases)	1X Coverage (%)	4X Coverage (%)	10X Coverage (%)	20X Coverage (%)	Mean depth (X)	Total variants (SNPs / Indels)	Total Coding SNPs	Total Coding Indels
I.1	5.64	99	96	91	77	51	62795	20457	376
I.2	5.50	99	97	94	85	73	62590	20383	397
I.3	4.04	99	96	92	79	55	61801	20122	389
II.1	5.63	99	96	91	77	51	61917	20017	383
II.2	5.07	99	97	93	81	60	62422	20366	379
III.1	5.87	99	97	94	86	90	63560	20664	408
III.2	5.77	99	97	94	85	84	63703	20753	394
IV.1	4.89	99	97	94	84	73	63451	20766	404
IV.2	5.11	99	97	94	85	80	63548	20811	419
V.1	5.31	99	97	94	85	76	63676	20698	401
V.2	5.50	99	97	94	86	79	63669	20580	410
V.3	4.97	99	97	93	82	64	63526	20657	399
VI.1	4.96	99	97	94	84	73	63241	20469	397
VI.2	5.48	99	97	94	86	78	63691	20608	380
VII.1	5.02	99	97	93	84	76	63532	20634	395
VII.2	5.50	99	97	94	85	84	62906	20347	392
Mean	**5.27**	**99**	**97**	**93**	**83**	**71.7**	**63127**	**20521**	**395**

Supp. Table 2. Descriptive of the variants genotyped by the exome chip and tested for association with FMD.

Exome sequencing candidate genes					
	Present in Exome chip	Polymorphic variants after QC			
		MAF < 0.01	MAF < 0.05	MAF ≥ 0.05	Tested
Missense	2307	541	646	196	842
Nonsense	33	8	8	1	9
Silent	200	61	68	15	-
Synonymous	151	40	47	14	-
Total	2691	650	769	226	851
Vascular and connective tissue disorders genes					
	Present in Exome chip	Polymorphic variants after QC			
		MAF < 0.01	MAF < 0.05	MAF ≥ 0.05	Tested
Missense	275	72	79	6	85
Nonsense	1	0	0	0	0
Silent	15	1	1	13	-
Synonymous	14	3	4	3	-
Total	305	76	84	22	85

Supp. Table 3. Results of single-variant association tests of common variants with FMD for exome sequencing candidates and vascular and connective tissue disorders genes. SNPs with MAF≥0.05 in controls were tested by logistic regression under an additive model.

Vascular and connective tissues disorders genes					**All Patients (249 cases vs. 689 controls)**		**Multifocal FMD (164 cases vs. 689 controls)**		**Females (193 cases vs. 344 controls)**	
Chr	**Gene**	**SNP Chip ID**	**Position (bp)**	**Risk allele**	**OR**	***P***	**OR**	***P***	**OR**	***P***
1	*PLOD1*	exm15706	12009956	A	1.02	0.879	1.01	0.943	0.9206	0.6537
1	*PLOD1*	exm15710	12010469	T	1.10	0.633	1.09	0.712	1.046	0.8445
2	*COL3A1*	exm252043	189864080	A	0.90	0.412	0.78	0.095	0.8741	0.3674
3	*MYLK*	exm344374	123512627	T	0.95	0.804	1.20	0.408	1.102	0.6703
9	*COL5A1*	exm794931	137642654	A	0.98	0.916	0.96	0.892	1.17	0.5774
16	*MYH11*	exm1220028	15820863	T	0.98	0.877	1.06	0.663	1.008	0.9522
Exome sequencing genes					**All Patients (249 cases vs. 689 controls)**		**Multifocal FMD (164 cases vs. 689 controls)**		**Females (193 cases vs. 344 controls)**	
Chr	**Gene**	**SNP Chip ID**	**Position (bp)**	**Risk allele**	**OR**	***P***	**OR**	***P***	**OR**	***P***
1	*OBSCN*	exm156282	228525008	A	0.76	0.088	0.70	0.067	0.74	0.126
1	*OBSCN*	exm156120	228505204	A	0.80	0.119	0.76	0.112	0.80	0.192
1	*OBSCN*	exm156780	228564884	A	1.16	0.380	1.24	0.276	0.99	0.968
1	*OBSCN*	exm156457	228548197	A	1.15	0.428	1.27	0.230	0.99	0.954
1	*OBSCN*	exm156606	228558892	T	1.15	0.490	1.15	0.548	0.88	0.578
1	*OBSCN*	exm156096	228504472	T	1.07	0.554	1.18	0.193	1.18	0.202
1	*OBSCN*	exm155617	228464248	T	1.07	0.557	1.19	0.184	1.19	0.185
1	*OBSCN*	exm155739	228468458	A	0.95	0.688	0.89	0.411	0.83	0.179
1	*OBSCN*	exm156022	228494790	A	1.04	0.714	0.97	0.829	0.95	0.640
1	*OBSCN*	exm156346	228528563	G	0.99	0.943	1.02	0.877	0.98	0.895
2	*COL6A3*	exm278759	238243464	C	0.89	0.398	0.81	0.185	0.82	0.220
2	*COL6A3*	exm278779	238244963	A	0.91	0.416	0.86	0.279	0.88	0.365
2	*COL6A3*	exm278798	238247734	G	0.92	0.681	1.06	0.805	0.95	0.825

Exome sequencing genes					All Patients (249 cases vs. 689 controls)		Multifocal FMD (164 cases vs. 689 controls)		Females (193 cases vs. 344 controls)	
2	*COL6A3*	exm278742	238243292	A	0.99	0.895	1.02	0.899	1.04	0.782
2	*FSIP2*	exm251217	186661567	A	0.77	0.176	0.72	0.158	0.74	0.188
2	*FSIP2*	exm251238	186668935	G	1.10	0.401	1.04	0.761	1.01	0.933
2	*FSIP2*	exm251184	186654592	A	0.93	0.483	0.88	0.322	0.87	0.261
2	*FSIP2*	exm251229	186665824	A	0.94	0.552	0.89	0.368	0.87	0.279
2	*FSIP2*	exm251208	186659359	G	0.94	0.552	0.89	0.368	0.87	0.279
2	*FSIP2*	exm251233	186667121	T	0.94	0.552	0.89	0.368	0.87	0.279
2	*FSIP2*	exm251203	186658565	G	0.94	0.552	0.89	0.368	0.87	0.279
2	*FSIP2*	exm251173	186627943	T	0.94	0.555	0.89	0.370	0.87	0.282
2	*FSIP2*	exm251236	186668714	T	1.11	0.580	1.24	0.305	1.14	0.545
2	*FSIP2*	exm251200	186658056	G	0.94	0.588	0.87	0.262	0.87	0.259
2	*FSIP2*	exm251224	186664963	C	0.95	0.606	0.89	0.368	0.88	0.320
2	*FSIP2*	exm251295	186671357	G	0.95	0.606	0.89	0.368	0.88	0.320
2	*FSIP2*	exm251311	186671912	G	0.95	0.606	0.89	0.368	0.88	0.320
2	*FSIP2*	exm251197	186656956	T	0.95	0.606	0.89	0.368	0.88	0.320
2	*FSIP2*	exm251202	186658438	C	0.95	0.606	0.89	0.368	0.88	0.320
2	*FSIP2*	exm251226	186665432	G	0.95	0.607	0.89	0.369	0.88	0.321
2	*FSIP2*	exm251375	186678633	T	1.05	0.786	1.05	0.803	1.10	0.623
2	*FSIP2*	exm251325	186672527	A	1.02	0.846	0.95	0.743	0.99	0.946
2	*TANC1*	exm234920	159954175	T	1.10	0.380	1.06	0.672	1.06	0.681
2	*TTN*	exm249518	179659912	A	0.61	0.067	0.59	0.106	0.49	0.023
2	*TTN*	exm248110	179575511	T	0.77	0.270	0.80	0.408	0.71	0.201
2	*TTN*	exm249099	179623758	C	0.80	0.306	0.73	0.231	0.87	0.584
2	*TTN*	exm249127	179629461	C	0.80	0.306	0.73	0.231	0.87	0.584
2	*TTN*	exm247903	179545859	T	1.13	0.310	1.21	0.171	1.24	0.126
2	*TTN*	exm247987	179558366	C	0.91	0.469	0.80	0.139	0.84	0.261

Exome sequencing genes					All Patients (249 cases vs. 689 controls)		Multifocal FMD (164 cases vs. 689 controls)		Females (193 cases vs. 344 controls)	
2	*TTN*	exm247970	179554305	T	1.08	0.525	1.07	0.606	1.11	0.414
2	*TTN*	exm248245	179582537	T	0.94	0.614	0.86	0.342	0.87	0.377
2	*TTN*	exm248179	179579093	C	0.99	0.935	0.98	0.905	0.94	0.741
2	*TTN*	exm248277	179583496	G	1.01	0.963	1.00	0.991	0.97	0.879
2	*TTN*	exm248353	179587130	G	1.01	0.963	1.00	0.991	0.97	0.879
3	*GBE1*	exm331298	81698130	C	1.12	0.328	1.13	0.376	1.13	0.405
4	*FAT1*	exm438900	187629538	C	1.10	0.355	1.20	0.141	1.06	0.624
4	*FAT1*	exm438950	187630590	A	1.12	0.424	1.29	0.118	1.08	0.656
4	*FAT1*	exm438474	187531022	G	1.14	0.428	1.01	0.947	1.27	0.222
4	*FAT1*	exm438726	187542755	C	1.14	0.433	1.03	0.886	1.26	0.239
4	*FAT1*	exm438781	187557893	T	0.93	0.496	0.94	0.622	0.96	0.759
4	*FAT1*	exm438816	187627792	T	0.94	0.607	0.87	0.381	0.92	0.577
4	*FAT1*	exm438425	187525020	C	1.05	0.646	0.99	0.911	1.07	0.632
4	*FAT1*	exm438564	187539588	C	1.07	0.749	0.97	0.891	0.85	0.530
4	*FAT1*	exm438849	187628398	C	0.97	0.803	0.91	0.517	0.95	0.718
4	*FAT1*	exm438878	187629140	G	0.97	0.803	0.91	0.517	0.95	0.718
4	*FAT1*	exm438534	187538942	T	0.98	0.822	0.98	0.883	1.04	0.766
4	*FAT1*	exm438912	187629770	A	1.02	0.848	1.07	0.590	0.98	0.856
5	*CMYA5*	exm463647	79029594	C	1.27	0.103	1.31	0.114	1.21	0.272
5	*CMYA5*	exm463616	79028726	G	1.27	0.103	1.31	0.114	1.21	0.272
5	*CMYA5*	exm463529	79026539	C	0.73	0.138	0.70	0.168	0.82	0.466
5	*CMYA5*	exm463497	79025634	A	0.73	0.147	0.71	0.175	0.82	0.466
5	*CMYA5*	exm463454	79024779	G	0.75	0.173	0.74	0.227	0.86	0.564
5	*CMYA5*	exm463571	79027605	C	0.75	0.173	0.74	0.227	0.86	0.564
5	*CMYA5*	exm463655	79029749	G	0.75	0.173	0.74	0.227	0.86	0.564
5	*CMYA5*	exm463680	79030346	G	0.75	0.173	0.74	0.227	0.86	0.564

Exome sequencing genes					All Patients (249 cases vs. 689 controls)		Multifocal FMD (164 cases vs. 689 controls)		Females (193 cases vs. 344 controls)	
5	*CMYA5*	exm463639	79029383	G	0.75	0.173	0.74	0.227	0.86	0.564
5	*CMYA5*	exm463797	79033306	C	0.75	0.173	0.74	0.227	0.86	0.564
5	*CMYA5*	exm463633	79029288	A	0.75	0.173	0.74	0.227	0.86	0.564
5	*CMYA5*	exm463902	79086883	A	1.16	0.288	1.19	0.273	1.31	0.093
5	*CMYA5*	exm463612	79028586	T	0.80	0.357	0.75	0.324	0.84	0.563
5	*CMYA5*	exm463611	79028513	G	0.80	0.357	0.75	0.324	0.84	0.563
5	*CMYA5*	exm463651	79029726	A	0.80	0.357	0.75	0.324	0.84	0.563
5	*CMYA5*	exm463771	79032666	T	0.80	0.357	0.75	0.324	0.84	0.563
5	*CMYA5*	exm463774	79032711	A	0.80	0.357	0.75	0.324	0.84	0.563
5	*CMYA5*	exm463840	79034662	G	0.80	0.357	0.75	0.324	0.84	0.563
5	*CMYA5*	exm463675	79030212	T	0.81	0.379	0.76	0.337	0.84	0.563
5	*CMYA5*	exm463679	79030338	G	0.81	0.379	0.76	0.337	0.84	0.563
5	*CMYA5*	exm463574	79027700	C	0.84	0.456	0.81	0.445	0.89	0.687
5	*CMYA5*	exm463919	79095417	T	0.91	0.539	0.91	0.584	0.91	0.616
5	*CMYA5*	exm463480	79025157	G	0.94	0.698	0.91	0.612	0.78	0.168
5	*CMYA5*	exm463724	79031558	A	0.94	0.703	0.92	0.649	0.77	0.157
5	*CMYA5*	exm463609	79028472	T	1.02	0.832	1.02	0.857	0.94	0.616
5	*CMYA5*	exm463862	79041057	G	0.97	0.876	0.98	0.902	0.79	0.257
6	*BCLAF1*	exm580659	136589425	T	0.89	0.395	0.90	0.520	0.80	0.184
6	*BCLAF1*	exm580884	136599885	A	1.02	0.940	1.04	0.872	1.13	0.589
7	*CFTR*	exm652311	117199533	A	0.94	0.598	1.01	0.963	0.93	0.546
7	*SSPO*	exm670206	149528262	T	1.28	0.074	1.14	0.448	1.15	0.415
7	*SSPO*	exm670137	149522162	T	1.24	0.113	1.08	0.632	1.11	0.545
7	*SSPO*	exm669428	149483237	A	0.82	0.246	0.86	0.436	0.81	0.267
7	*SSPO*	exm669471	149485059	C	0.84	0.286	0.88	0.515	0.82	0.319
7	*SSPO*	exm670028	149517978	T	0.89	0.432	1.04	0.817	0.96	0.797

Exome sequencing genes					All Patients (249 cases vs. 689 controls)		Multifocal FMD (164 cases vs. 689 controls)		Females (193 cases vs. 344 controls)	
7	*SSPO*	exm669775	149503843	G	1.10	0.462	0.94	0.714	1.02	0.904
7	*SSPO*	exm669237	149474393	G	1.02	0.887	1.14	0.397	0.90	0.514
7	*SSPO*	exm670154	149523195	G	0.98	0.889	0.94	0.729	0.93	0.705
7	*SSPO*	exm669677	149498893	G	0.98	0.922	0.95	0.821	0.77	0.277
8	*PLEC*	exm729371	144994888	A	0.81	0.277	0.59	0.040	0.82	0.419
8	*PLEC*	exm729805	145001031	C	1.01	0.912	0.89	0.346	1.08	0.556
8	*PLEC*	exm729995	145007187	A	1.00	0.971	0.94	0.600	1.09	0.519
8	*RECQL4*	exm733289	145741702	G	1.00	0.999	1.02	0.896	0.87	0.273
10	*TTC40*	exm866876	134648279	T	1.40	0.110	1.64	0.036	1.32	0.261
10	*TTC40*	exm866909	134659626	G	1.09	0.573	1.08	0.676	1.25	0.230
10	*TTC40*	exm867088	134748331	T	1.01	0.907	0.92	0.543	0.90	0.451
10	*TTC40*	exm866823	134626293	A	0.99	0.949	0.93	0.778	0.96	0.865
11	*DYNC2H1*	exm952045	103158278	T	1.22	0.075	1.22	0.130	1.22	0.132
11	*DYNC2H1*	exm951960	103082590	G	0.91	0.494	0.97	0.876	0.88	0.444
11	*DYNC2H1*	exm951684	102988504	T	1.14	0.509	1.35	0.156	1.21	0.413
11	*DYNC2H1*	exm951796	103029516	A	1.07	0.587	1.18	0.253	1.04	0.770
11	*DYNC2H1*	exm951690	102991197	T	1.03	0.869	1.20	0.390	1.06	0.792
11	*DYNC2H1*	exm951784	103027234	G	1.03	0.870	1.21	0.361	1.07	0.768
11	*MUC6*	exm873765	1017596	G	1.31	0.107	1.07	0.749	1.21	0.338
11	*MUC6*	exm874164	1029320	T	1.14	0.434	1.19	0.360	1.28	0.223
16	*HYDIN*	exm1254853	70883694	A	0.76	0.165	0.83	0.402	0.85	0.457
16	*HYDIN*	exm1255459	71103393	T	0.81	0.376	0.97	0.910	0.71	0.233
16	*HYDIN*	exm1255260	70993639	T	0.87	0.540	0.80	0.441	0.62	0.115
16	*HYDIN*	exm1255254	70993553	T	0.90	0.581	0.98	0.924	0.95	0.800
16	*HYDIN*	exm1255338	71015329	T	1.00	0.966	1.06	0.644	1.07	0.595
17	*RNF213*	exm1364064	78263486	C	0.77	0.075	0.77	0.125	0.75	0.092
17	*RNF213*	exm1364209	78298926	A	0.64	0.089	0.76	0.361	0.57	0.068

Exome sequencing genes					All Patients (249 cases vs. 689 controls)		Multifocal FMD (164 cases vs. 689 controls)		Females (193 cases vs. 344 controls)	
17	*RNF213*	exm1364056	78262161	C	0.79	0.128	0.74	0.112	0.74	0.100
17	*RNF213*	exm1364570	78337058	C	0.81	0.222	0.97	0.876	0.86	0.435
17	*RNF213*	exm1364433	78321157	C	0.86	0.542	0.64	0.187	0.82	0.526
17	*RNF213*	exm1364715	78355494	G	1.10	0.584	0.98	0.920	1.11	0.641
17	*RNF213*	exm1364218	78305871	G	0.95	0.633	0.84	0.186	0.84	0.179
17	*RNF213*	exm1364774	78363054	A	1.08	0.652	0.96	0.855	1.08	0.745
17	*RNF213*	exm1364327	78319136	G	1.05	0.684	0.92	0.548	0.93	0.602
17	*RNF213*	exm1364214	78302157	C	0.97	0.763	0.90	0.437	0.86	0.264
17	*RNF213*	exm1364229	78306280	A	0.99	0.891	0.90	0.439	0.87	0.314
19	*MUC16*	exm1419998	8962389	A	0.63	0.108	0.66	0.218	0.70	0.293
19	*MUC16*	exm1421903	9083143	A	1.18	0.163	1.16	0.287	1.12	0.398
19	*MUC16*	exm1421905	9083174	A	1.18	0.163	1.16	0.287	1.12	0.398
19	*MUC16*	exm1422040	9086123	C	1.18	0.163	1.16	0.287	1.12	0.398
19	*MUC16*	exm1422124	9088017	T	0.85	0.174	0.78	0.074	0.93	0.617
19	*MUC16*	exm1421836	9077581	T	0.84	0.278	0.60	0.019	0.78	0.232
19	*MUC16*	exm1420079	8993018	A	0.84	0.305	0.77	0.212	0.89	0.548
19	*MUC16*	exm1421653	9074073	A	0.85	0.322	0.62	0.027	0.81	0.286
19	*MUC16*	exm1420999	9060059	C	1.11	0.354	1.07	0.602	1.06	0.672
19	*MUC16*	exm1420550	9028373	A	0.90	0.363	0.91	0.488	1.07	0.633
19	*MUC16*	exm1419995	8961981	C	1.18	0.364	1.32	0.192	1.24	0.312
19	*MUC16*	exm1421873	9082514	A	1.19	0.376	1.45	0.089	1.29	0.260
19	*MUC16*	exm1420745	9049149	A	1.11	0.386	1.01	0.928	1.07	0.606
19	*MUC16*	exm1420697	9048035	C	1.11	0.386	1.01	0.928	1.07	0.606
19	*MUC16*	exm1420660	9047267	T	1.11	0.386	1.01	0.928	1.07	0.606
19	*MUC16*	exm1420881	9057750	A	1.10	0.401	1.06	0.682	1.05	0.740
19	*MUC16*	exm1420887	9057896	C	1.10	0.401	1.06	0.682	1.05	0.740

Exome sequencing genes					All Patients (249 cases vs. 689 controls)		Multifocal FMD (164 cases vs. 689 controls)		Females (193 cases vs. 344 controls)	
19	*MUC16*	exm1420958	9059232	C	1.10	0.402	1.06	0.682	1.04	0.786
19	*MUC16*	exm1420832	9056941	G	1.10	0.434	1.01	0.928	1.07	0.606
19	*MUC16*	exm1420234	9003645	A	0.93	0.492	0.93	0.548	0.87	0.258
19	*MUC16*	exm1421763	9076278	A	0.93	0.507	0.93	0.559	1.05	0.700
19	*MUC16*	exm1422163	9088772	C	0.91	0.518	0.87	0.388	0.80	0.198
19	*MUC16*	exm1422143	9088330	A	0.91	0.518	0.87	0.388	0.80	0.198
19	*MUC16*	exm1421920	9083457	T	0.91	0.518	0.87	0.388	0.80	0.198
19	*MUC16*	exm1422018	9085643	G	0.91	0.518	0.87	0.388	0.80	0.198
19	*MUC16*	exm1422057	9086318	A	0.91	0.518	0.87	0.388	0.80	0.198
19	*MUC16*	exm1422110	9087758	A	0.91	0.518	0.87	0.388	0.80	0.198
19	*MUC16*	exm1420858	9057408	A	1.12	0.521	1.23	0.328	1.08	0.740
19	*MUC16*	exm1421293	9066259	C	1.07	0.541	1.08	0.576	1.04	0.775
19	*MUC16*	exm1421996	9085004	C	0.92	0.543	0.87	0.388	0.81	0.212
19	*MUC16*	exm1421121	9062415	T	0.95	0.597	1.01	0.949	1.01	0.944
19	*MUC16*	exm1420874	9057687	A	0.92	0.635	1.07	0.716	0.86	0.452
19	*MUC16*	exm1420283	9006749	T	1.07	0.648	1.09	0.580	1.03	0.864
19	*MUC16*	exm1421845	9077803	C	0.95	0.651	0.94	0.663	1.06	0.658
19	*MUC16*	exm1421830	9077436	A	0.95	0.651	0.94	0.663	1.06	0.658
19	*MUC16*	exm1421745	9075969	T	0.95	0.651	0.94	0.663	1.06	0.658
19	*MUC16*	exm1421752	9076083	A	0.95	0.651	0.94	0.663	1.06	0.658
19	*MUC16*	exm1421590	9072742	C	0.95	0.651	0.94	0.663	1.06	0.658
19	*MUC16*	exm1421795	9076728	T	0.95	0.651	0.94	0.663	1.06	0.658
19	*MUC16*	exm1421806	9076950	A	0.95	0.651	0.94	0.663	1.06	0.658
19	*MUC16*	exm1420878	9057721	T	0.96	0.662	0.89	0.372	1.00	0.995
19	*MUC16*	exm1420924	9058624	A	0.94	0.673	0.83	0.266	0.85	0.343
19	*MUC16*	exm1420787	9049910	T	0.94	0.675	0.83	0.269	0.86	0.372

Exome sequencing genes					All Patients (249 cases vs. 689 controls)		Multifocal FMD (164 cases vs. 689 controls)		Females (193 cases vs. 344 controls)	
19	*MUC16*	exm1420837	9056989	G	0.94	0.675	0.83	0.269	0.86	0.372
19	*MUC16*	exm1421170	9063678	G	0.96	0.696	0.95	0.706	1.07	0.633
19	*MUC16*	exm1421260	9065632	C	0.96	0.696	0.95	0.706	1.07	0.633
19	*MUC16*	exm1420639	9046876	T	0.95	0.699	0.83	0.280	0.86	0.372
19	*MUC16*	exm1421462	9069892	C	0.95	0.702	0.83	0.284	0.86	0.372
19	*MUC16*	exm1421455	9069792	A	0.95	0.702	0.83	0.284	0.86	0.372
19	*MUC16*	exm1421327	9066874	T	0.95	0.702	0.83	0.284	0.86	0.372
19	*MUC16*	exm1422222	9090182	C	1.06	0.722	1.24	0.267	1.01	0.976
19	*MUC16*	exm1421020	9060541	T	0.96	0.734	0.91	0.531	0.93	0.622
19	*MUC16*	exm1421481	9070194	A	0.92	0.740	0.92	0.764	0.96	0.869
19	*MUC16*	exm1420513	9024994	T	0.96	0.741	1.01	0.946	1.12	0.385
19	*MUC16*	exm1421938	9083791	A	1.06	0.745	1.24	0.267	1.00	0.999
19	*MUC16*	exm1422034	9085958	G	1.06	0.745	1.24	0.267	1.00	0.999
19	*MUC16*	exm1420707	9048320	T	0.96	0.763	0.84	0.317	0.86	0.372
19	*MUC16*	exm1422250	9090784	T	1.05	0.773	1.23	0.280	0.99	0.964
19	*MUC16*	exm1421124	9062544	T	0.96	0.826	1.14	0.500	0.90	0.592
19	*MUC16*	exm1420989	9059808	A	0.96	0.826	1.14	0.500	0.90	0.592
19	*MUC16*	exm1420192	9001833	A	1.02	0.865	1.07	0.639	1.09	0.550
19	*MUC16*	exm1420961	9059307	T	0.97	0.889	1.09	0.672	1.07	0.772
19	*MUC16*	exm1421510	9070837	G	1.02	0.902	1.15	0.390	0.99	0.969
19	*MUC16*	exm1420710	9048342	T	1.01	0.923	1.09	0.501	1.12	0.367
19	*MUC16*	exm1421243	9065203	C	0.98	0.927	1.10	0.644	1.08	0.727
19	*MUC16*	exm1421403	9068530	T	0.98	0.927	1.10	0.644	1.08	0.727
19	*MUC16*	exm1420736	9048895	C	0.99	0.938	1.11	0.627	1.07	0.765
19	*MUC16*	exm1421820	9077196	T	1.01	0.967	1.09	0.725	1.21	0.454
19	*MUC16*	exm1420836	9056982	A	1.00	0.968	1.04	0.794	1.08	0.584
19	*MUC16*	exm1420074	8987218	T	1.01	0.970	0.97	0.806	0.97	0.813

CHAPITRE IV : IDENTIFICATION D'UN PREMIER GENE DE PREDISPOSITION A LA DYSPLASIE FIBROMUSCULAIRE

Article 2

PHACTR1 is a genetic susceptibility locus for Fibromuscular Dysplasia.

Soto Romuald Kiando, Nathan R. Tucker, Luis-Jaime Castro-Vega, Alexander Katz, Valentina D'Escamard, Cyrielle Tréard, Daniel Fraher, Juliette Albuisson, Daniella Kadian-Dodov, Zi Ye, Erin Austin, Min-Lee Yang, Kristina Hunker, Cristina Barlassina,Daniele Cusi, Pilar Galan, Jean-Philippe Empana, Xavier Jouven, Anne-Paule Gimenez-Roqueplo, Patrick Bruneval, Esther Soo Hyun Kim, Jeffrey W. Olin, Heather L. Gornik, Michel Azizi1, Pierre-François Plouin, Patrick T. Ellinor, Iftikhar J. Kullo, David J. Milan, Santhi K. Ganesh, Pierre Boutouyrie, Jason C. Kovacic, Xavier Jeunemaitre, Nabila Bouatia-Naji.

Article en révision **Plos Genetics** Mai 2016

IV.1. Position du problème

Dans notre première étude (Kiando, Barlassina et al. 2015), nous avons exclu l'implication d'un gène majeur muté pour la DFM et conclu à une hétérogénéité génétique ; tout ceci évoquant une base génétique polygénique et multifactorielle. Ces résultats sont corroborés par les données récentes de la littérature présentant la DFM asymptomatique comme une pathologie certainement sous-évaluée (Neymark, LaBerge et al. 2000, Plouin, Perdu et al. 2007), pour laquelle plusieurs facteurs environnementaux comme le tabagisme, l'imprégnation

œstrogénique et le stress mécanique sont suggérés (Sang, Whelton et al. 1989, Bofinger, Hawley et al. 1999, Savard, Steichen et al. 2012).

L'ensemble de ces éléments pris conjointement justifient donc une approche d'étude d'association de variants fréquents dans le but d'identifier des polymorphismes génétiques à risque pour la DFM.

IV.2. Résultats majeurs

Le design de notre étude était constitué d'une phase de découverte puis quatre phases de réplications dans des cohortes indépendantes de la DFM. Dans la phase de découverte, nous avons analysé les variants génétiques fréquents (MAF ≥ 5%) de la puce *Exome-chip* précédemment génotypés chez 249 cas de DFM issus des cohortes françaises et 689 témoins. Il s'agissait au total de 25606 variants ayant passé toutes les étapes de contrôle qualité pour lesquels nous avons testé l'association avec la DFM par régression logistique sous le modèle génétique additif ajusté pour les cinq premiers axes de l'analyse en composante principale. Les phases de réplications ont consisté à tester l'association spécifique de marqueurs sélectionnés dans la phase de découverte au sein de cohortes indépendantes, notamment une française et trois autres issus de différents centres participant au registre américain de la DFM. L'ensemble des résultats issus des différentes phases de notre étude ont fait l'objet d'une méta-analyse globale incluant 1154 cas de DFM et 3895 témoins tous d'ascendance européenne.

Cette étude a permis d'identifier puis de confirmer dans trois cohortes de réplication sur quatre, l'association avec la DFM du polymorphisme rs9349379 situé dans l'intron du gène *PHACTR1* (Méta-analyse : OR [A] = 1, 39 (1,25-1,54) ; $P < 7.36 \times 10^{-10}$) [cf. Tableau VI-2].

Tableau VI-2. Association du polymorphisme rs9349379 avec la DFM dans cinq études indépendantes. Pas d'hétérogénéité entre les différentes cohortes de cas et de témoins (Cochran's Q statistical test P = 0,574).

Etude	Cohortes	*n*	EAF (A)	OR (95% CI)	*P -value*
Découverte	Cas RVDRC	249	0,7	1,65	1,47 × 10−5
	Témoins SU.VI.MAX	689	0,59	(1,32-2,07)	
Réplication 1	Cas ARCADIA	393	0,68	1,32	7,21 × 10−4
	Témoins PPS3	2537	0,62	(1,12-1,54)	
Réplication 2	Cas Mayo Clinic	143	0,66	1,37	0,034
	Témoins Mayo Clinic	333	0,59	(1,02-1,83)	
Réplication 3	Cas UM-Cleveland	304	0,69	1,31	0,026
	Témoins UM-Cleveland	289	0,63	(1,03-166)	
Réplication 4	Cas DEFINE-FMD	65	0,77	1,39	0,276
	Témoins DEFINE-FMD	47	0,7	(0,77-2,49)	
Méta-analyse	Total Cas de DFM	1154	0,69	1,39	7,36 × 10−10
	Total Témoins	3895	0,61	(1,25-1,54)	

Des études d'association complémentaires par régression linéaire ont été conduites auprès de 2574 sujets normo-tendus (975 femmes et 1483 hommes) issus de la population générale afin d'évaluer l'implication du polymorphisme rs9349379 au niveau des artères carotides pour lesquels plusieurs paramètres ont été mesurés par écho-tracking haute résolution. Il est apparu que le signal d'association identifié pour la DFM est associé entre autres à une

augmentation de l'épaisseur intima-média, ($P = 1.97\times10^{-4}$) et du ratio de la paroi artérielle sur la lumière ($P = 0.002$), allant dans le sens d'un épaississement de la paroi artérielle.

Des explorations fonctionnelles réalisées en collaboration ont permis de détecter *PHACTR1* par immunohistochimie dans l'endothélium et les cellules musculaires lisses de carotides dysplasiques et non dysplasiques avec une expression augmentée de *PHACTR1* pour les porteurs de l'allèle à risque de DFM dans des cultures primaires de fibroblastes humains (N=86, *P*=0,003). Enfin, l'invalidation de *Phactr1* chez le poisson zèbre conduit à une dilatation des vaisseaux indiquant un défaut du développement vasculaire.

PHACTR1 Is a Genetic Susceptibility Locus for Fibromuscular Dysplasia

Short title: Genetic investigation of FMD

Soto Romuald Kiando,[1,2] Nathan R. Tucker,[3] Luis-Jaime Castro-Vega,[1,2], Alexander Katz,[4] Valentina D'Escamard,[5] Cyrielle Tréard,[1,2] Daniel Fraher[3], Juliette Albuisson,[1,2,6,7] Daniella Kadian-Dodov,[5] Zi Ye,[8] Erin Austin,[8] Min-Lee Yang,[4] Kristina Hunker,[4] Cristina Barlassina,[9] Daniele Cusi,[10] Pilar Galan,[11] Jean-Philippe Empana,[1,2] Xavier Jouven,[1,2,12] Anne-Paule Gimenez-Roqueplo,[1,2,7] Patrick Bruneval,[1,2] Esther Soo Hyun Kim,[13] Jeffrey W. Olin,[5] Heather L. Gornik,[13] Michel Azizi[14,15], Pierre-François Plouin,[1,2,14] Patrick T. Ellinor,[3] Iftikhar J. Kullo,[8] David J. Milan,[3] Santhi K. Ganesh,[4] Pierre Boutouyrie,[1,2,16] Jason C. Kovacic,[5] Xavier Jeunemaitre,[1,2,6,7] Nabila Bouatia-Naji,[1,2*]

1. INSERM, UMR970 Paris Cardiovascular Research Center (PARCC), Paris F-75015, FRANCE
2. Paris-Descartes University, Sorbonne Paris Cité, Paris 75006, FRANCE
3. Cardiovascular research Center, Massachusetts General Hospital, Charlestown, MA 02114, USA, Program in Medical and Population Genetics, The Broad Institute of Harvard and MIT, Cambridge, MA 02142
4. Department of Internal Medicine and Department of Human Genetics, University of Michigan, Ann Arbor, MI 48109, USA
5. The Zena and Michael A. Wiener Cardiovascular Institute, Icahn School of Medicine, Marie-Josée and Henry R. Kravis Cardiovascular Health Center at Mount Sinai, One Gustave L. Levy Place, Box 1030 New York, NY 10029, New York, NY, USA
6. AP-HP, Referral Center for Rare Vascular Diseases, Hôpital Européen Georges Pompidou, Paris, F-75015, FRANCE.
7. AP-HP, Department of Genetics, Hôpital Européen Georges Pompidou, Paris, F-75015, FRANCE
8. Department of Medicine, Division of Cardiovascular Diseases, Mayo Clinic, Rochester, MN 55905, USA
9. Dept. of Health Sciences, Genomic and Bioinformatics Unit, Viale Ortles 22/4, Milano, Chair and Graduate School of Nephrology, University of Milano, Division of Nephrology, San Paolo Hospital, Milano, 20142,ITALY
10. Institute of Biomedical Technologies, Italian National Centre of Research, Via F.lli Cervi 93, 20090 Segrate - Milano
11. Nutritional Epidemiology Research Group, Sorbonne-Paris-Cité, UMR University of Paris 13/Inserm U-557/INRA U-1125/CNAM, Bobigny, France F-93017, Bobigny, FRANCE
12. AP-HP, Department of Cardiology, Hôpital Européen Georges Pompidou, Paris, F-75015, FRANCE
13. Department of Cardiovascular Medicine, Cleveland Clinic Heart and Vascular Institute, Cleveland, OH 44195, USA.
14. AP-HP, Department of Hypertension, Hôpital Européen Georges Pompidou, Paris, F-75015, FRANCE
15. INSERM, Clinical Investigation Center CIC1418, Hôpital Européen Georges Pompidou, Paris, F-75015, FRANCE
16. AP-HP, Department of Pharmacology, Hôpital Européen Georges Pompidou, Paris, F-75015, FRANCE
*Correspondence: nabila.bouatia-naji@inserm.fr

Abstract

Fibromuscular dysplasia (FMD) is a nonatherosclerotic vascular disease leading to stenosis, dissection and aneurysm affecting mainly the renal and cerebrovascular arteries. FMD is often an underdiagnosed cause of hypertension and stroke, has higher prevalence in females (~80%) but its pathophysiology is unclear.

We analyzed ~26K common variants (MAF>0.05) generated by exome-chip arrays in 249 FMD patients and 689 controls. We replicated 13 loci ($P<10^{-4}$) in 402 cases and 2,537 controls and confirmed an association between FMD and a variant in the phosphatase and actin regulator 1 gene (*PHACTR1*). Three additional case control cohorts including 512 cases and 669 replicated this result and overall reached the genomic level of significance (OR=1.39, $P=7.4\times10^{-10}$, 1,154 cases and 3,895 controls). The top variant, rs9349379, is intronic to *PHACTR1*, a risk locus for coronary artery disease, migraine, and cervical artery dissection. The analyses of geometrical parameters of carotids from ~2,500 healthy volunteers indicate higher intima media thickness ($P=1.97\times10^{-4}$) and wall to lumen ratio ($P=0.002$) in rs9349379-A carriers, suggesting indices of carotid hypertrophy previously described in carotids of FMD patients. Immunohistochemistry detected PHACTR1 in endothelium and smooth muscle cells of FMD and normal human carotids. The expression of *PHACTR1* by genotypes in primary human fibroblasts showed higher expression in rs9349379-A carriers (N=86, $P=0.003$). *Phactr1* knockdown in zebrafish resulted in dilated vessels indicating subtle impaired vascular development.

We report the first susceptibility locus for FMD and provide evidence for a complex genetic pattern of inheritance and indices of shared pathophysiology between FMD and other cardiovascular and neurovascular diseases.

Author Summary

Fibromuscular Dysplasia (FMD) is a vascular disease characterized by a succession of occlusions and dilatation of medium-sized arteries (e.g renal, carotid or coronary arteries) with important health consequences, mainly resistant hypertension and stroke. FMD is an atypical vascular disease because it is not associated with overweight or dyslipidemia and 80% of patients are early middle aged women. Our genetic study conducted in >1100 patients and >3800 controls demonstrate that a common variant rs9349379 located on chromosome 6 in the phosphatase and actin regulator 1 gene (*PHACTR1*) increases by ~40% the risk of FMD. This is the first time a genetic risk factor is reported for FMD because it has been longtime considered rare and potentially under a Mendelian mode of inheritance. We also show that rs9349379 correlates with the expression of *PHACTR1* in fibroblasts from FMD patients and controls. Interestingly, the same allele that increases the risk of FMD is at risk for cervical artery dissection and migraine, often reported in FMD patients but protective from myocardial infarction and coronary disease, where atherosclerosis is more common. The clear role of PHACTR1 in maintaining vascular well integrity is not fully elucidated. Using a specific antibody we detected PHACTR1 both on endothelial and smooth muscle cells of human FMD and control carotids, which suggests that PHACTR1 may have multiple functions depending on the cell type and the degree of atherosclerosis of the arteries.

Introduction

Fibromuscular dysplasia (FMD) is a non-atherosclerotic and non-inflammatory vascular disease leading to stenosis, aneurysm, dissection, and/or occlusion of medium-sized arteries, in particular the renal and extracranial cerebrovascular arteries [1-3]. FMD predisposes to hypertension, transient ischemic attack and stroke [2,3]. Intriguingly, 75% to 90% of FMD patients are women[3,4] and FMD is increasingly considered to be a silent and under-diagnosed condition [5].

The angiography-based classification of renal FMD distinguishes between patients with multifocal stenoses, including the "string-of-beads" FMD pattern, and unifocal (or focal) FMD with corresponding differences in sex ratio, median age of diagnosis and smoking status [6,7].

The pathogenesis of FMD is unknown and there are strong arguments in favor of a genetic origin. We have previously reported familiality of FMD in ~10% of patients [8]. The US FMD registry has also described a family history in first or second-degree relatives for FMD (7%) and aneurysm (23%) [3]. Nonetheless, the unknown status of most of family members compromised heritability estimates and made the assessment of the genetic mode of inheritance difficult. Recently, we investigated the coding genomes from 16 affected siblings and excluded the existence of a major mutated gene in familial FMD [9]. The complexity of the diagnosis based on computed tomography angiography (CTA) and/or magnetic resonance angiography (MRA) and a general lack of awareness among affected patients and clinicians result in an under-diagnosis bias.

To advance our understanding of the etiology of FMD, we performed a genetic association study that identifies a first genetic susceptibility locus for FMD. In addition to the high prevalence of asymptomatic FMD (~3-6%) [3,10] and the existence of environmental modifiers (e.g female hormones, lifetime mechanical stress) our study provides genetic and functional evidence supporting for the first time a complex genetic basis for FMD.

Results

Association of rs9349379 in *PHACTR1* with FMD

We carried-out a multi-stage genetic association study, including a discovery and four validation case control cohorts, to identify genetic determinants of FMD. All FMD patients and controls were of European ancestry and have similar overall clinical features (Table 1). First, we analyzed 25,606 common (minor allele frequency ≥ 0.05) genetic variants in 249 FMD cases and 689 controls. Despite the small sample size, we also performed stratified analyses including only females (193 patients and 416 controls). Associations across chromosomes are summarized in Manhattan Plots (Fig S2). In the global analysis, no SNP achieved the adjusted significance threshold for multiple testing (P=1.95×10^{-6} Fig S2). Nevertheless, the strongest association signal located on chromosome 6 (rs9349379, effect allele frequency (EAF) in cases=0.70, odds ratio (OR)=1.65, P=1.47×10^{-5}) surpasses the adjusted threshold in the female only analysis (OR=1.99, P=8.16 × 10^{-7} S2 Fig, S1 Table).

Table 1. Clinical features of the study populations.

Cohorts	N	Females (%)	Age at exam (yrs)	Multifocals (%)	HTN (%)	HTN diagnosis (yrs)	Study Design
RVDRC Cases	249	193 (78%)	44.12 ± 14.31	164 (66%)	228 (92%)	36.6 ± 13.65	Clinical recruitment
SU.VI.MAX controls	689	416 (60%)	49.80 ± 6.20	NR	NA	NA	Population based
ARCADIA cases	402	319 (79%)	42.46 ± 15.68	273 (68%)	341 (85%)	34.9 ± 15.15	Clinical recruitment
PPS3 controls	2,537	1012 (40%)	58.73 ± 5.94	NR	0	NR	Population based
Mayo Cases	143	119 (83)%	61.48 ± 13.43	NA	108 (76%)	58.75 ± 10.31	Clinical recruitment
Mayo Controls	333	286 (86%)	65.78 ± 10.50	NR	169 (51%)	62.52 ± 10.25	Clinical recruitment
UM/Cleveland Cases	304	295 (97%)	53.93 ± 10.45	254 (84%)	193 (63%)	NA	Clinical recruitment
UM/Cleveland Controls	289	280 (97%)	55.13 ± 9.97	NR	NA	NA	Clinical recruitment
DEFINE-FMD Cases	65	65 (100%)	57.03 ± 10.13	65 (100%)	41 (63%)	42.68 ± 13.05	Clinical recruitment
DEFINE-FMD Controls	47	47 (100%)	50.70 ± 9.38	NR	4 (9%)	55.25 ± 13.05	Clinical recruitment

HTN, Hypertension; NR, Not Relevant; NA, Not Available.

In addition to rs9349379, we selected for follow-up loci that showed suggestive association with FMD ($P<10^{-3}$) and were located in or near either biological candidate genes (e.g extracellular matrix degradation) or previous cardiovascular genome-wide association (GWAS) signals. We also prioritized four SNPs located within one megabase interval around rs9349379, the top associated variant. The first follow-up study included 402 FMD patients from the ARCADIA registry, who had similar clinical characteristics to the patients of the discovery stage and were compared to 2,537 controls from PPS3 (Table 1). Of the 16 SNPs selected, 13 passed genotyping QC criteria in cases and controls. Three SNPs, all located in the *PHACTR1* locus, showed replicated association with FMD (rs9349379-A, $P=7.21 \times 10^{-4}$, rs9369640-C, $P=8.45 \times 10^{-4}$ and rs1332844-C, $P=1.72 \times 10^{-3}$, S1 Table). Conditioned regression analyses for rs9349379 in the case control cohort including the discovery and first follow-up study samples indicated association signal redundancy for rs9369640 ($P=0.21$) and rs1332844 (P=0.24) with rs9349379 being the most statistically significant. Next, we investigated the association of rs9349379 with FMD in three additional and independent case-control cohorts from the USA. Overall, of the four follow-up studies, three showed a significant effect of rs9349379 on the risk of FMD; all studies showed consistent direction of effect of the FMD risk allele rs9349379-A being more prevalent in FMD cases (Table 2). The association meta-analysis included 1,154 FMD patients and 3,895 controls and indicated an overall OR of 1.39 for rs9349379 (EAF=0.69 in global cases sample), with a high level of significance ($P=7.36\times10^{-10}$) that is below the genomic threshold and no indices for heterogeneity between studies ($P=0.574$, Table 2).

Table 2. Association of rs9349379 in *PHACTR1* with FMD in five independent case-control cohorts.

Study	Cohorts	N	AA	AG	GG	EAF	OR[a] (95% CI)	P value[a]
Discovery	RVDRC Cases	249	124	103	22	0.70	1.65 (1.32-2.07)	1.47×10^{-5}
	SU.VI.MAX controls	689	237	341	111	0.59		
Follow-up	ARCADIA Cases	393	183	170	40	0.68	1.32 (1.12-1.54)	7.21×10^{-4}
	PPS3 controls	2,537	982	1,174	381	0.62		
	Mayo Cases	143	62	65	16	0.66	1.37 (1.02-1.83)	0.034
	Mayo Controls	333	116	159	58	0.59		
	UM/Cleveland Cases	304	145	130	29	0.69	1.31 (1.03-166)	0.026
	UM/Cleveland Controls	289	116	135	38	0.63		
	DEFINE-FMD Cases	65	41	18	6	0.77	1.39 (0.77-2.49)	0.276
	DEFINE-FMD Controls	47	22	22	3	0.70		
Meta-analysis[b]	All Cases	1,154	555	486	113	0.69	1.39 (1.25-1.54)	7.36×10^{-10}
	All Controls	3,895	1,473	1,831	591	0.61		

[a]Odds Ratio (OR) and P values were computed by logistic regression under the additive genetic model. [b]Meta-analysis was performed using inverse variance-weighted method. Heterogeneity between cohorts was tested using Cochran's Q statistics and was not significant (P=0.574).

Association of rs9349379 with carotid parameters in PPS3

Here we aimed to characterize the association of rs9349379 with artery thickness and stiffness using a non-invasive method in 2,458 healthy volunteers. Genetic association analyses indicate significant association between the FMD increasing risk allele rs9349379-A and greater intima-media thickness (IMT) (β_{add}=11.54 µm, P_{add}=1.97×10^{-4}, Fig 1) and wall to lumen ratio (WLR) (β=0.004, *P*=0.002, Fig 1), and decreased circumferential wall stress at DBP (β_{add}=-0.71, P_{add}=0.004, S2 Table). Interestingly, despite their smaller number (N=975) compared to males (N=1,483), females present more accentuated effect of the rs9349379-A allele on IMT (β=15.15 µm, *P*=8.38×10^{-4}) and WLR (β=0.006, *P*=0.002). However, we detected no significant interaction with sex. rs9349379-A also associated with increased wall cross-sectional area (β_{add}=0.24 mm^2, P_{add}=0.001, S2 Table), peripheral SBP (β_{add}=0.70 mmHg, P_{add}=0.009, S2 Table) and central pulse pressure (β_{add}=0.62 mmHg, P_{add}=0.002, Fig 1). Importantly, rs9349379 associated with none of the stiffness traits (S2 Table).

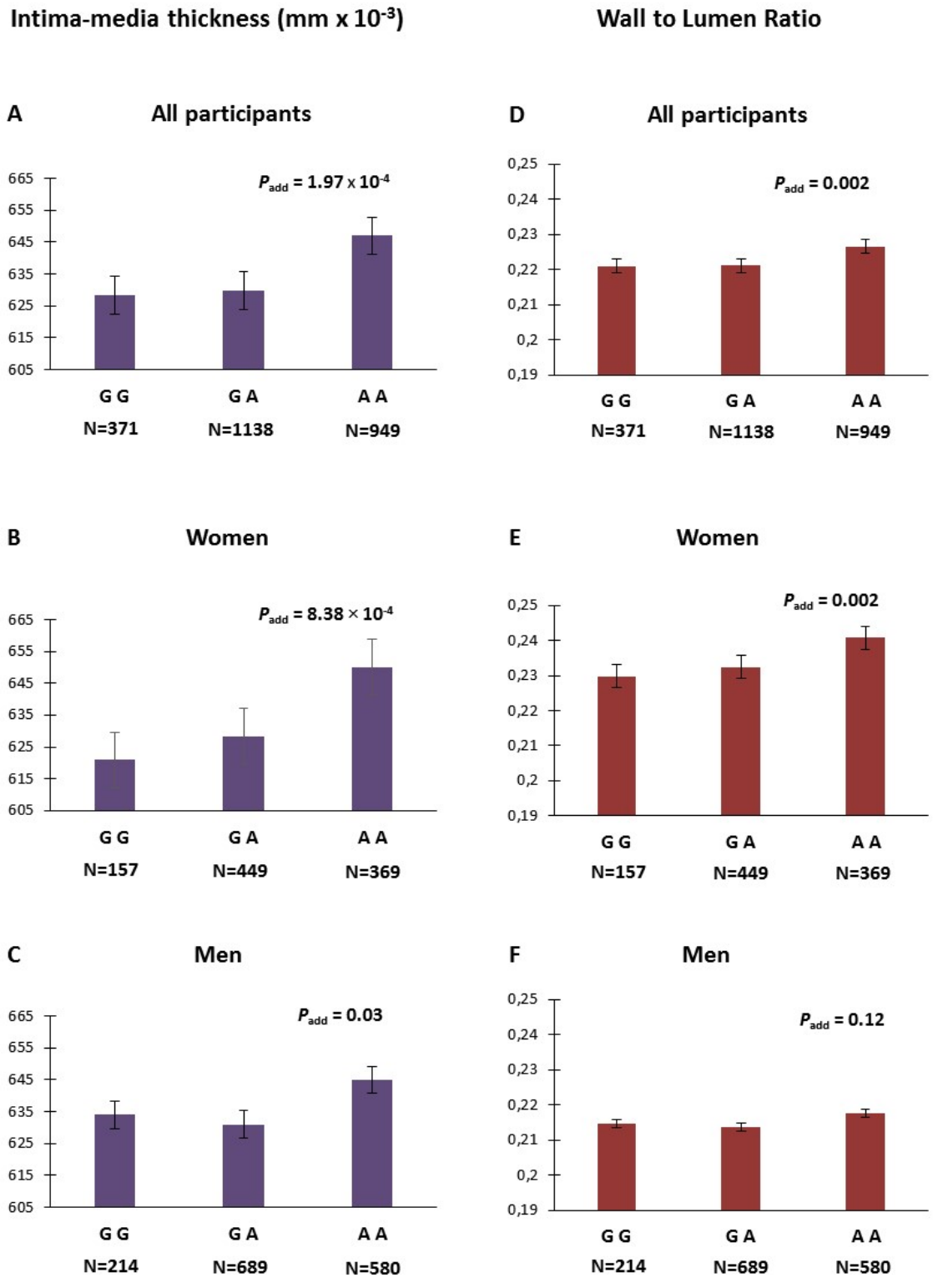


Figure 1. Association between rs9349379 and artery thickness in healthy controls. Intima-media thickness and wall to lumen ratio are presented by genotype in all (A,D) women (B,E) and men (C,F) participants. We indicated p-values for the linear regression analyses under the additive model adjusted for age, sex and body surface area.

***PHACTR1* expression studies**

Variant rs9349379 resides in the fourth intron of the phosphatase actin regulator 1 gene (*PHACTR1*) in a putative regulatory sequence. We assessed *PHACTR1* expression using quantitative real time PCR in mRNAs from primary cultured human fibroblasts. *PHACTR1* expression did not differ in fibroblasts from FMD patients compared to age and sex matched controls (Fig 2). However, stratifying by rs9349379 genotype indicated increased expression in individuals carrying the FMD risk allele rs9349379-A (P_{add}=0.003, Fig 2). At the protein level, using a specific antibody, we found that PHACTR1 is expressed in endothelial and medial smooth muscle cells of carotid arteries from both normal and FMD patients (Fig 2).

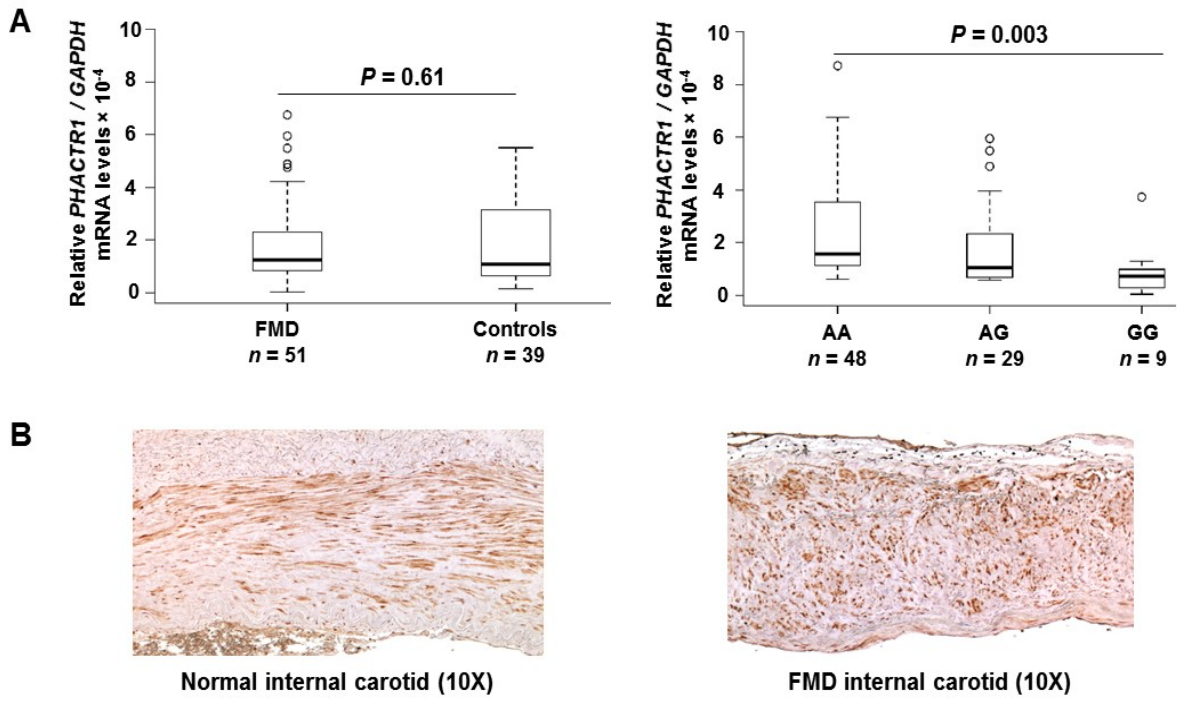


Figure 2. *PHACTR1* mRNA expression and immunostaining.
(A) *PHACTR1* mRNA expression in human fibroblasts. *PHACTR1* mRNA levels were determined by RT-qPCR in cultured fibroblasts from controls (N=39) and FMD cases (N=51). *GAPDH* expression was used as control for normalization and data are expressed as mean fold change of *PHACTR1* relative to *GAPDH*. There is a non-significant trend toward overexpression of *PHACTR1* in FMD cases compared to controls (*P*=0.61, Mann-Whitney test) whereas significant differences are uncovered after stratification by genotypes (*P*=0.003, Kruskal-Wallis test).
(B) Immunostaining of normal and FMD internal carotid using anti-PHACTR1 antibody. PHACTR1 was detected in endothelium and smooth muscle cells in the media. Staining is mostly cytoplasmic with regular alignment in normal carotids and typical disorganized cellular structure in media of FMD carotid.

Functional role of *Phactr1* in vascular development

Given the unclear role for *PHACTR1* in vascular development and maintenance, we assessed the effect of *PHACTR1* perturbation in the zebrafish. Morpholino injected embryos had undetectable *phactr1* transcript levels at 72 hours post fertilization while overall morphology was unchanged (Fig S3). Compared to control injected zebrafish, *phactr1* knockdown resulted in a marked disorganization of the developing hepatic portal vein and the segmental vessels in the developing trunk (Fig 3, Fig S3). Further analysis of the diameter of three major peripheral vessels, the dorsal aorta, caudal artery and posterior cardinal vein, demonstrated a nearly 8% dilatation (P<0.05) of the posterior cardinal vein in *phactr1* suppressed embryos.

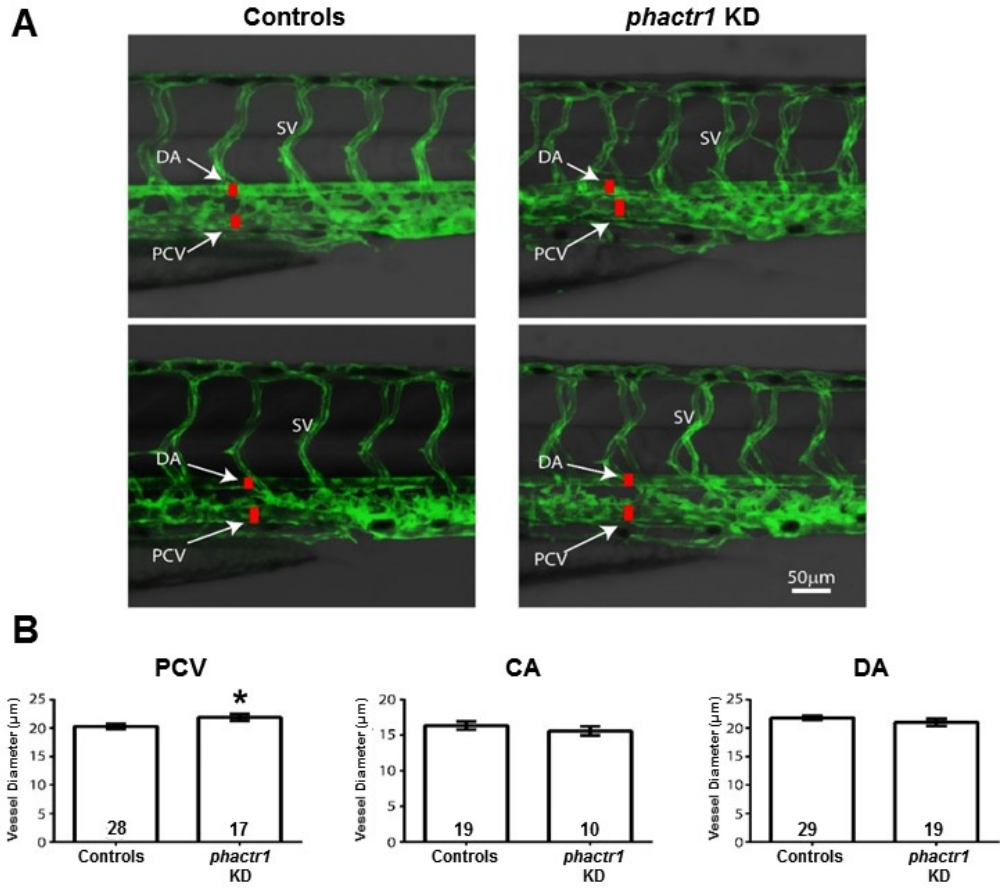


Figure 3. *phactr1* modulation in zebrafish affects vascular dimensions and patterning.
Two-dimensional projections obtained from z-series confocal images in the trunk of control and phactr1 knockdown (KD) zebrafish embryos (two representative images per condition). Green represents the vascular endothelium as marked by EGFP. Greyscale represents the corresponding DIC bright field image of the fish trunk. DA: Dorsal aorta, SV: Segmental vessel, PCV: Posterior cardinal vein. Quantification of inner vascular diameter for the dorsal aorta (DA), posterior cardinal vein (PCV) and caudal artery (CA). (*) represents P<0.05.

Discussion

Our study describes for the first time the genetic association of rs9349379, a common variant in *PHACTR1*, with arterial fibromuscular dysplasia (FMD). We demonstrate that this common variant increases by ~40% the risk of FMD in five independent case-control studies. These findings are based on genetic data from 1,154 FMD patients and 3,895 controls, the largest investigation conducted so far to elucidate the genetic basis of this intriguing vascular disease with unknown etiological origin and challenging clinical features.

To date, the genetic investigation on FMD was limited to the screening of candidate genes involved in rare vascular and arterial syndromes, and/or the study of underpowered series of patients.[9,11,12] For long considered a rare disease, FMD has also been hypothesized to be under the genetic control of highly penetrant genetic defects. Our study provides the first evidence for a complex genetic pattern of inheritance for FMD, involving a common genetic allele (frequency=0.60 in the general population). The identification of a genetic susceptibility locus for FMD supports the concept that this disease is controlled by a large number of genetic determinants in strong interaction with various environmental factors, including female sex and mechanical stresses.

We provide genetic evidence for increased IMT, narrowed artery lumen and no change in arterial stiffness, even after adjusting for blood pressure, in healthy volunteers carrying the rs9349379-A, the FMD risk allele. These genetic associations support the observation of carotid concentric hypertrophy in FMD risk carriers, especially females. This finding is consistent with previous arterial features reported in FMD patients when compared to age, sex and SBP matched controls.[13] Circumferential wall stress was decreased in rs9349379-A carriers, showing that hypertrophy may overcome the moderate increase in blood pressure. Of note, the effect size of the FMD risk allele on IMT corresponds to approximately five years of aging for the IMT, when reported to sex and age reference values for arterial geometry,[14] suggesting potentially an accelerated arterial aging in carriers of the

rs9349379-A FMD risk allele.

PHACTR1 was previously identified by several genome-wide association studies (GWAS) as a risk locus for cardiovascular and neurovascular diseases. *PHACTR1* is a confirmed susceptibility locus for coronary artery disease (CAD) and myocardial infarction(MI),[15,16] migraine[17,18] and more recently cervical artery dissection (CeAD), a rare condition defined as a mural hematoma in a carotid or vertebral artery and a cause of stroke.[19] Of note, the association of rs9349379 is in the opposite direction for FMD, CeAD and migraine, with rs9349379-A at risk, when compared to CAD and MI with rs9349379[G] allele at risk. Migraine and CeAD share several clinical features with FMD. Migraine is also more prevalent in females than in males and it is reported by a third of FMD patients.[3] CeAD is an important risk factor for subarachnoid hemorrhage stroke, as is cerebrovascular FMD.[3] The investigation of a large series of CeAD patients indicated that cervical FMD is reported in 5.6% of the patients[20] and carotid dissection is the presenting manifestation in 12.1% of patients of the US FMD registry.[3] In contrast to migraine and CeAD, the implication of the same common variant in *PHACTR1* in CAD/ MI and FMD is rather unexpected, though it involved a different allele at the same genetic variant. The clinical link between FMD and CAD/MI is less obvious, except for the high proportion of FMD in rare forms of MI involving spontaneous coronary artery dissection, that also present in young females without atherosclerotic risk factors.[21-23] By definition FMD does not involve atherosclerotic stenosis that is concomitant with CAD and MI through pathogenesis processes (e.g dyslipidemia and inflammation). Future comprehensive genetic investigation by full GWAS for FMD will allow the assessment of the putative consistency and opposition of effects between risk loci for these cardiovascular and neurovascular diseases.

The rs9349379 is intronic to *PHACTR1* in a putative noncoding regulatory sequence. *PHACTR1* encodes a phosphatase and actin regulator protein and its function is not fully elucidated. Here we describe significant correlation between rs9349379 genotypes and *PHACTR1* expression in human

fibroblasts from FMD patients and controls. This association is consistent with the expression quantitative trait loci (eQTL) data from several artery beds of the Genotype-Tissue Expression (GTEx) project and a recent study where increased expression is reported in human coronary arteries from donors carrying the rs9349379-A.[16] Using a genome-editing technique with CRISPR-Cas9 applied to human umbilical vein endothelial cells, the deletion of the sequence containing rs9349379 caused a 35% decrease in *PHACTR1* expression and impaired the fixation of the myocyte enhancer factor-2 (MEF2), supporting a functional regulatory effect of this genetic variant *in vitro.*[16] Molecular studies have also linked PHACTR1 to cell adhesion and migration in angiogenesis via vascular endothelial growth factor (VEGF) stimulation.[24,25] In contrast to other studies,[25,26] Beaudoin et al reported a lack of evidence for the induction of PHACTR1 expression in endothelial cells with pro-angiogenic (VEGF), pro-inflammatory stimulations or shear stress.[16] Our immunohistochemistry staining of PHACTR1 clearly indicates the presence of PHACTR1 both in endothelial and smooth muscle cells of normal and FMD carotid arteries. The genetic implication of *PHACTR1* in FMD, the subtle impairment in the development of vasculature in zebrafish, in addition to the evidence from recent GWAS describing an increasing number of loci with genes involved in vessel wall biology in CAD and MI[27] all support PHACTR1 plays a key etiological role in vascular structure. Thus, the manifestation of one specific vascular disease depends on multiple and particular environmental triggers (e.g. female specific hormonal context, mechanical movements of medium size arteries in the case of FMD) that is still to be specifically determined for each of these cardiovascular and neurovascular diseases.

In summary, our study provides genetic and functional evidence to support *PHACTR1* as a first susceptibility locus for FMD. Further functional exploration of this locus and more comprehensive genetic investigation through genome-wide association will provide additional predisposing loci to FMD and help understanding the etiological mechanisms of non-atherosclerotic arterial stenosis.

Methods

Patients and control populations

We used a three-stage association design. First, we performed an exome-chip based genetic association in 249 French FMD patients (RVDRC cohort) and 689 controls from SU.VI.MAX [28]. Second, we followed-up 13 loci in an independent set of 402 French patients (ARCADIA registry) and 2,537 controls from PPS3 [29]. Three additional studies from the USA totaling 512 patients and 669 controls were used for further replication: Mayo Clinic cohort [30], University of Michigan (UM)/Cleveland Clinic cohort and the DEFINE-FMD study. All participants are of European ancestry and presented similar clinical characteristics (Table 1) and homogeneous diagnosis, exclusion and inclusion criteria. We obtained individual informed consent from all participants in France and the USA and study approvals from respective local ethics committees.

RVDRC and ARCADIA cases

We analyzed unrelated FMD followed-up at the Rare Vascular Diseases Reference Center (RVDRC) of the European Hospital Georges Pompidou (HEGP), Paris, France. We ascertained patients from the ARCADIA (Assessment of Renal and Cervical Artery DysplasIA) register, an ongoing national FMD registry at the HEGP, Paris. The diagnosis of FMD in RVDRC and ARCADIA patients was established using clinical information from the medical history, the interpretation of angiography and/or computed tomography scan of arterial beds after the exclusion of other causes of arterial stenosis such as atherosclerosis, Takayasu disease and Elhers Danlos syndrome and neurofibromatosis type 1. Given the complexity of the interpretation of imaging of vascular diseases, a local panel of experts including clinicians from the departments of hypertension, radiology, vascular medicine and medical genetics validated the diagnosis of FMD.

SU.VI.MAX and PPS3 controls

Controls in discovery stage are French Europeans from the SU.VI.MAX study as previously described [28]. PPS3 controls were 2,537 participants ascertained from the Paris Prospective Study 3 (PPS3), an ongoing observational French prospective study evaluating the possible implication of numerous vascular health parameters in cardiovascular disease in healthy subjects aged from 50 to 75 years undergoing a standard health examination in a large preventive medical center that is subsidized by the French national insurance system for salaried workers [29].

Mayo Case-Control study

Potential FMD cases were identified from the Mayo vascular disease bio-repository. Electronic health records of 501 patients were manually reviewed by clinicians and the diagnosis was confirmed for 191 patients according to criteria in [1], of whom 159 had DNA available and 143 (83 % females) had high-density genotyping data. Patients were matched for sex and age (+/- 2 years) to 333 controls without known FMD or other atherosclerotic vascular diseases.

UM/Cleveland Clinic case control study

Patients with FMD were recruited at the University of Michigan and Cleveland Clinic according to inclusion and exclusion criteria previously described [3]. Controls were obtained from the Cleveland Clinic Genebank according to a local IRB protocol.

The DEFINE-FMD Study.

Eligible cases were females with an imaging-confirmed diagnosis of multifocal FMD and who fulfilled other accepted diagnostic criteria [1]. Healthy controls were matched to FMD cases according to age and sex, required to be receiving ≤ 2 blood pressure medications, have a body mass index < 35kg/m2 and to be non-smokers. Healthy controls underwent physical exam and those with bruits; unexplained hypertension or other cardiovascular findings were excluded. Exclusion criteria for all

subjects included male gender, unifocal FMD, use of immunosuppressive agents, major comorbidities, diseases that may confound genetic/genomic analyses (i.e. Crohn's disease, multiple sclerosis etc.) or any other form of heritable vascular disease (i.e. Ehlers-Danlos, Marfan, Loeys-Dietz).

Genotyping

Genotypes were generated using the Illumina-HumanExome-12v1 array in RVDRC, SU.VI.MAX and PPS3,[31] Illumina-Human-Omni-Express-Exome in the DEFINE-FMD Study participants, Illumina-Infinium-Human-CoreExome in Mayo Clinic cohorts and by individual genotyping in ARCADIA (KASP technology®) and UM/Cleveland Clinic cohorts (Taqman®). We applied quality control (QC) filters to the discovery cases and controls as recommended[32] (individual's call rate < 97%, extreme heterozygosity, sex-discordant, duplicate or relatedness unsing PLINK (version 1.07).[33] Individuals with non-European ancestry were detected and excluded (54 cases and 9 controls) using the EIGENSTRAT program[34] and visualized by principal components analysis (PCA) including HapMap phase 3 samples. Cases and controls displayed a comparable distribution after ancestry QC (Figure S1). From 240,748 successfully genotyped SNPs, we excluded monomorphic variants (n=129,890), call rate < 99%, deviation from Hardy-Weinberg Equilibrium in cases and/or controls ($P < 10^{-5}$) and minor allele frequency (MAF)<0.05 in controls. The final analysis included 25,606 variants (25,138 autosomal and 468 X-linked) in a sample of 249 FMD cases and 689 controls. Comparable individuals and markers QC was applied to follow-up cohorts using arrays (PPS3, Mayo Clinic case control cohorts, the DEFINE-FMD study) or individual genotyping (ARCADIA and UM/Cleveland cases control cohort).

Carotid geometry and stiffness measurements in PPS3 cohort

Carotid parameters were measured and calculated as previously reported [31]. Briefly, a 10 MHz 128 transducer linear array probe was positioned on the carotid area. Measurements were performed on a 4 cm segment of the right common carotid artery, 1 cm proximal to the bifurcation/sinus

throughout the cardiac cycle for 6 seconds. A longitudinal section showing clear interfaces for blood/intima and media/adventitia was obtained. The system allows real-time radiofrequency signal analysis with operator-independent determination of external diameter (Dext), internal diameter (Dint), and intima-media thickness (IMT) on 128 lines throughout the cardiac cycle. Distension was measured on 14 lines at high-pulsed radiofrequency (600 Hz). The axial resolution was 34 µm for diameter, 17 µm for IMT, and 1.7 µm for distension.[35] Aortic blood pressure was estimated from the distension waveform according to van Bortel et al.[36] The distensibility coefficient, representing the elastic properties of the artery as a hollow structure, was calculated as dLCSA/(LCSA×central PP), where LCSA is the lumen cross-sectional area and central PP is the aortic pulse pressure. Carotid stiffness (Cstif) was calculated as $DC^{-0.5}$ and circumferential wall stress as diastolic blood pressure×Dint/2×IMT. WCSA is wall cross-sectional area. Young's elastic modulus is calculated as DC^{-1}x3(1+LCSA/WCSA). Here we examined the association of one genetic variant (rs9349379) with 14 interdependent traits in a sub-sample of 2458 participants (975 females and 1483 males). Hypertensive participants (BP over 140 and/or 90 mmHg, and/or use of antihypertensive treatments) were excluded to avoid confounding with genetic effects.

Statistical analyses

We tested the association with FMD using logistic regression under the additive genetic model as implemented in PLINK[33] (version 1.07) in discovery and follow-up studies. In the discovery analysis, we included the first five principal components axes as co-variates to control for hidden population stratification. The Bonferroni adjusted threshold for significance was set to $P=1.95\times10^{-6}$ to account for multiple testing of 25,606 common variants. We used the inverse variance-weighted method for meta-analysis implemented in Metal [37]. Heterogeneity was assessed with Cochran's Q statistics. We tested the association of rs9349379 with carotid parameters in 2,458 normotensive subjects (975 females and 1483 males) from PPS3 using a linear regression on an additive genetic model including age, sex, body surface area (BSA), and mean blood pressure (MBP) as covariates when relevant.

Before performing the analyses, all parameters (all quantitative traits) were quantile-transformed to a standard normal distribution. The 14 traits fall into three main carotid parameters categories: geometry (intima-media-thickness, external and internal carotid diameters, wall to lumen ratio and circumferential wall stress), arterial stiffness (stiffness, cross-sectional distensibility, Young's elastic modulus, cross-sectional compliance, and wall cross-sectional area) and central blood pressure (systolic, diastolic and aortic blood pressure). Considering that carotid parameters are interdependent, we applied a Bonferroni correction (P_{adj} = 0.017) for the three main categories as recommended [31].

***PHACTR1* expression in fibroblasts**

We analyzed the expression of PHACTR1 from fibroblasts of skin biopsy samples of 104 individuals (51 FMD, 39 controls, 12 undetermined). Genotypes and detectable expression levels were available in 86 individuals.

Fibroblasts of the DEFINE-FMD Study were derived from skin biopsy samples using standard explant techniques from 103 individuals (46 FMD and 57 controls). Briefly, skin biopsy samples were dissected into small pieces and cultured under cover slips in DMEM/F12 media containing 20% fetal bovine serum, 1% antibiotic-antimycotic solution, 1% 200 mM L-glutamine, 1% 100mM Sodium pyruvate, and 1% MEM Non-Essential Amino Acids (all from Life Technologies, Grand Island, NY, USA) at 37°C in 5% CO_2. Cell culture medium was replaced every 48 hours and confluent cells from passages 2-3 were used for RNA extractions. RNA was then converted to cDNA using iscript cDNA synthesis kit (Bio Rad, Hercules, CA, USA) and quantitative real time PCR (qRT-PCR) was performed using SYBR green fastmix low rox (Quanta Biosciences, Gaithersburg, MD, USA). Fibroblasts from the University of Michigan were obtained through the Coriell Biorespository (Camden, NJ, USA) for 20 subjects with FMD and apparently 20 healthy control subjects and matched for passage number. DNAs and genotypes were not available. Confluent cells from passages 6-8 were used to extract RNA and generate cDNAs to assess *PHACTR1* expression according to culture and extraction protocol described for the DEFINE-

FMD study. Fibroblasts from the Department of genetic, HGEP study were derived from skin biopsy samples from patients followed up for unknown arterial diseases where FMD and Elhers Danlos syndrome were discarded. FMD status is not available. SNP rs9349379 was genotyped by direct forward and reverse sequencing (BigDye Terminator kit v3.1 cycle sequencing kit) and run on an ABI Prism 3730XL DNA Analyzer Sequencer (Perkin Elmer Applied Biosystems, Foster City, CA). Cell culture and RNA extraction conditions and reagent were identical to the DEFINE-FMD study. Informed consent was obtained from all patients of all centers.

Assessment of the expression of PHACTR1. In total, we analyzed the expression of *PHACTR1* from fibroblasts of skin biopsy samples of 104 individuals (51 FMD, 39 controls, 12 undetermined), and 86 individuals had genotypes and relevant expression levels. The expression of PHACTR1 transcript variant 1 (NCBI Reference Sequence: NM_030948.2) and *GAPDH*, which was used as normalization control, was determined by RT-qPCR following the MIQE guidelines.[38] All expression studies from the three laboratories used the same following primers: PHACTR1 fwd ATGACCGCAGGGCAGATAAG, rev TTCGGATGGCAGCTTTGTCT; GAPDH fwd GGGTGTGAACCATGAGAAGTATGA, rev GGTGCAGGAGGCATTGCT. The efficacy and efficiency of amplifications for PHACTR1 and GAPDH were equivalent as determined by the linearity tests: (*PHACTR1* efficacy 1.97 and efficiency 98.7%); (GAPDH efficacy 2.04 and efficiency 102%). The specificity of amplification products was verified by the presence of a single peak in the melting temperature curve analysis. Outliers displaying *GAPDH* Cq values different from the mean or a ΔCq value >14 (Cq value PHACTR1 – Cq value *GAPDH*) were filtered out. Good correlations for RT and technical qPCR replicates were obtained. Given that data obtained from three independent laboratories was equivalent: mean Cq for *GAPDH* 17,41 +/- 0,65 and mean Cq for *PHACTR1* was 30.46 +/- 1.19, we pooled them as indicated to gain power, especially for the correlation by genotypes where few samples per study were GG carriers. Data is presented as the mean fold change in the expression of PHACTR1 relative to GAPDH, calculated using the formula $2^{-\Delta Cq}$, as recommended.[39] Statistical analyses were performed using a non-parametric Mann-Whitney test for single comparison between FMD cases and controls and Kruskal Wallis test for

comparisons according to rs9349379 genotypes.

PHACTR1 Immunostaining

Protein staining for PHACTR1 was performed on paraffin embedded carotid artery samples using primary anti-PHACTR1 antibody (Sigma-Aldrich, St Louis, MI, USA) and revealed using an ABC peroxidase kit with diaminobenzidine (Vector laboratories, Burlingame, CA, USA). The paraffin blocks of arterial tissues were obtained from surgical pathology archives of the HEGP as remnants of the regular diagnostic procedure. FMD arteries belonged to patients who had surgical reparation for aneurysm resection. Control internal carotid artery was embedded in the surgical resection of a large carotid body paraganglioma necessitating carotid artery resection. The arterial tissues were fixed in formalin and embedded in paraffin. Antigen retrieval were performed by incubating tissue sections in alkaline solution (Dako, Trappes, France) for 40 minutes at 94°C in a hot water bath. The sections were then incubated for 60 minutes with the primary anti-PHACTR1 rabbit polyclonal antibody diluted at 1/50. For revelation we used ABC peroxidase kit with diaminobenzidine (Vector laboratories, Burlingame, CA, USA).

Zebrafish experiments

Anti-*phactr1* morpholino was designed to target the exon7/intron7 boundary of zebrafish *phactr1* transcript. A non-targeting morpholino of equivalent length but differing nucleotide composition was injected at equivalent concentrations as a control. For injections, 0.8nL of control of anti-*phactr1* morpholino was injected into single cell embryos of the Tubingen/AB strain. For confirmation of morpholino efficacy, primers spanning the targeted boundary were used to amplify cDNA constructed from the isolated RNA of injected 72 hour post fertilization embryos. For imaging studies, fish were reared under standard conditions with addition of phenylthiourea to media in order to inhibit pigmentation. At 72hpf, high-speed videography was performed as described[40] used to determine the diameter of major vessels. Measurement of vessels was manually performed using ImageJ. For

confocal reconstructions of the embryonic vasculature, embryos obtained from lines expressing EGFP driven by the *flk* promoter were used. Confocal images were obtained on a Nikon A1SiR confocal. Two-dimensional projections of maximal intensity (head and torso) or averaged intensity (trunk) were used to visualize the overall vascular architecture using ImageJ.

Acknowledgments

We thank all patients who participate in this study in all the participating centers. We also thank clinicians who participated in the recruitment of patients. From the HEGP: Laurence Amar, Arshid Azarine, Guillaume Bobrie, Antoine Chédid, Béatrice Fiquet, Michael Frank, Agnes Labatide-Alanore, Olivier Steinchen. We also acknowledge the contribution to clinical data acquisition by Alessandra Giavarini, University of Milano and Sebastien Savard, Hôtel-Dieu de Québec Hospital. From the ARCADIA register: Jean-Philippe Baguet, Louis Boyer, Béatrice Duly-Bouhanick, Philippe Gosse, Claire Mounier-Vehier, Alexandre Persu, François Sihol, Frédéric Thony and Emmanuel Touzé. We acknowledge Susan Blackburn and James C. Stanley, University of Michigan and Stanley Hazen, Cleveland Clinic and Carin Smith, Mayo Clinic, in the recruitment of USA patients and controls. We thank Louise Marais, Inserm UMR970, Paris, for the acquisition of echo-tracking data, Patrick Lacolley, Inserm UMR_S 1116, Nancy and François Cambien, Inserm UMR_S 1166, Paris who contributed to the genetic investigation in PPS3 and Mark Lathrop, Genome Quebec who contributed to the generation of genotypes in SU.VI.MAX. We thank Siham Bennis, Lama Cheikh-Hussein and Giulia Hardouin for assistance performing *in silico* analyses and expression experiments.

REFERENCES

1. Olin JW, Sealove BA (2011) Diagnosis, management, and future developments of fibromuscular dysplasia. J Vasc Surg 53: 826-836 e821.
2. Plouin PF, Perdu J, La Batide-Alanore A, Boutouyrie P, Gimenez-Roqueplo AP, et al. (2007) Fibromuscular dysplasia. Orphanet J Rare Dis 2: 28.
3. Olin JW, Froehlich J, Gu X, Bacharach JM, Eagle K, et al. (2012) The United States Registry for Fibromuscular Dysplasia: results in the first 447 patients. Circulation 125: 3182-3190.
4. Cragg AH, Smith TP, Thompson BH, Maroney TP, Stanson AW, et al. (1989) Incidental fibromuscular dysplasia in potential renal donors: long-term clinical follow-up. Radiology 172: 145-147.
5. Hendricks NJ, Matsumoto AH, Angle JF, Baheti A, Sabri SS, et al. (2014) Is fibromuscular dysplasia underdiagnosed? A comparison of the prevalence of FMD seen in CORAL trial participants versus a single institution population of renal donor candidates. Vasc Med 19: 363-367.
6. Persu A, Giavarini A, Touze E, Januszewicz A, Sapoval M, et al. (2014) European consensus on the diagnosis and management of fibromuscular dysplasia. J Hypertens 32: 1367-1378.
7. Savard S, Steichen O, Azarine A, Azizi M, Jeunemaitre X, et al. (2012) Association between 2 angiographic subtypes of renal artery fibromuscular dysplasia and clinical characteristics. Circulation 126: 3062-3069.
8. Pannier-Moreau I, Grimbert P, Fiquet-Kempf B, Vuagnat A, Jeunemaitre X, et al. (1997) Possible familial origin of multifocal renal artery fibromuscular dysplasia. J Hypertens 15: 1797-1801.
9. Kiando SR, Barlassina C, Cusi D, Galan P, Lathrop M, et al. (2015) Exome sequencing in seven families and gene-based association studies indicate genetic heterogeneity and suggest possible candidates for fibromuscular dysplasia. J Hypertens 33: 1802-1810.
10. McKenzie GA, Oderich GS, Kawashima A, Misra S (2013) Renal artery fibromuscular dysplasia in 2,640 renal donor subjects: a CT angiography analysis. J Vasc Interv Radiol 24: 1477-1480.
11. Perdu J, Gimenez-Roqueplo AP, Boutouyrie P, Beaujour S, Laloux B, et al. (2006) Alpha1-antitrypsin gene polymorphisms are not associated with renal arterial fibromuscular dysplasia. J Hypertens 24: 705-710.
12. Poloskey SL, Kim E, Sanghani R, Al-Quthami AH, Arscott P, et al. (2012) Low yield of genetic testing for known vascular connective tissue disorders in patients with fibromuscular dysplasia. Vasc Med 17: 371-378.
13. Boutouyrie P, Gimenez-Roqueplo AP, Fine E, Laloux B, Fiquet-Kempf B, et al. (2003) Evidence for carotid and radial artery wall subclinical lesions in renal fibromuscular dysplasia. J Hypertens 21: 2287-2295.
14. Engelen L, Ferreira I, Stehouwer CD, Boutouyrie P, Laurent S, et al. (2013) Reference intervals for common carotid intima-media thickness measured with echotracking: relation with risk factors. Eur Heart J 34: 2368-2380.
15. Coronary Artery Disease Genetics C (2011) A genome-wide association study in Europeans and South Asians identifies five new loci for coronary artery disease. Nat Genet 43: 339-344.
16. Beaudoin M, Gupta RM, Won HH, Lo KS, Do R, et al. (2015) Myocardial Infarction-

Associated SNP at 6p24 Interferes With MEF2 Binding and Associates With PHACTR1 Expression Levels in Human Coronary Arteries. Arterioscler Thromb Vasc Biol 35: 1472-1479.
17. Freilinger T, Anttila V, de Vries B, Malik R, Kallela M, et al. (2012) Genome-wide association analysis identifies susceptibility loci for migraine without aura. Nat Genet 44: 777-782.
18. Anttila V, Winsvold BS, Gormley P, Kurth T, Bettella F, et al. (2013) Genome-wide meta-analysis identifies new susceptibility loci for migraine. Nat Genet 45: 912-917.
19. Debette S, Kamatani Y, Metso TM, Kloss M, Chauhan G, et al. (2015) Common variation in PHACTR1 is associated with susceptibility to cervical artery dissection. Nat Genet 47: 78-83.
20. Bejot Y, Aboa-Eboule C, Debette S, Pezzini A, Tatlisumak T, et al. (2014) Characteristics and outcomes of patients with multiple cervical artery dissection. Stroke 45: 37-41.
21. Saw J, Aymong E, Sedlak T, Buller CE, Starovoytov A, et al. (2014) Spontaneous coronary artery dissection: association with predisposing arteriopathies and precipitating stressors and cardiovascular outcomes. Circ Cardiovasc Interv 7: 645-655.
22. Prasad M, Tweet MS, Hayes SN, Leng S, Liang JJ, et al. (2015) Prevalence of extracoronary vascular abnormalities and fibromuscular dysplasia in patients with spontaneous coronary artery dissection. Am J Cardiol 115: 1672-1677.
23. Michelis KC, Olin JW, Kadian-Dodov D, d'Escamard V, Kovacic JC (2014) Coronary artery manifestations of fibromuscular dysplasia. J Am Coll Cardiol 64: 1033-1046.
24. Allain B, Jarray R, Borriello L, Leforban B, Dufour S, et al. (2012) Neuropilin-1 regulates a new VEGF-induced gene, Phactr-1, which controls tubulogenesis and modulates lamellipodial dynamics in human endothelial cells. Cell Signal 24: 214-223.
25. Jarray R, Allain B, Borriello L, Biard D, Loukaci A, et al. (2011) Depletion of the novel protein PHACTR-1 from human endothelial cells abolishes tube formation and induces cell death receptor apoptosis. Biochimie 93: 1668-1675.
26. Jarray R, Pavoni S, Borriello L, Allain B, Lopez N, et al. (2015) Disruption of phactr-1 pathway triggers pro-inflammatory and pro-atherogenic factors: New insights in atherosclerosis development. Biochimie.
27. Consortium CAD (2015) A comprehensive 1000 Genomes-based genome-wide association meta-analysis of coronary artery disease. Nat Genet 47: 1121-1130.
28. Hercberg S, Galan P, Preziosi P, Bertrais S, Mennen L, et al. (2004) The SU.VI.MAX Study: a randomized, placebo-controlled trial of the health effects of antioxidant vitamins and minerals. Arch Intern Med 164: 2335-2342.
29. Empana JP, Bean K, Guibout C, Thomas F, Bingham A, et al. (2011) Paris Prospective Study III: a study of novel heart rate parameters, baroreflex sensitivity and risk of sudden death. Eur J Epidemiol 26: 887-892.
30. Ye Z, Kalloo FS, Dalenberg AK, Kullo IJ (2013) An electronic medical record-linked biorepository to identify novel biomarkers for atherosclerotic cardiovascular disease. Glob Cardiol Sci Pract 2013: 82-90.
31. Proust C, Empana JP, Boutouyrie P, Alivon M, Challande P, et al. (2015) Contribution of Rare and Common Genetic Variants to Plasma Lipid Levels and Carotid Stiffness and Geometry: A Substudy of the Paris Prospective Study 3. Circ Cardiovasc Genet 8: 628-636.
32. Guo Y, He J, Zhao S, Wu H, Zhong X, et al. (2014) Illumina human exome genotyping array clustering and quality control. Nat Protoc 9: 2643-2662.

33. Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MA, et al. (2007) PLINK: a tool set for whole-genome association and population-based linkage analyses. Am J Hum Genet 81: 559-575.
34. Patterson N, Price AL, Reich D (2006) Population structure and eigenanalysis. PLoS Genet 2: e190.
35. Meinders JM, Brands PJ, Willigers JM, Kornet L, Hoeks AP (2001) Assessment of the spatial homogeneity of artery dimension parameters with high frame rate 2-D B-mode. Ultrasound Med Biol 27: 785-794.
36. Van Bortel LM, Balkestein EJ, van der Heijden-Spek JJ, Vanmolkot FH, Staessen JA, et al. (2001) Non-invasive assessment of local arterial pulse pressure: comparison of applanation tonometry and echo-tracking. J Hypertens 19: 1037-1044.
37. Willer CJ, Li Y, Abecasis GR (2010) METAL: fast and efficient meta-analysis of genomewide association scans. Bioinformatics 26: 2190-2191.
38. Bustin SA, Benes V, Garson JA, Hellemans J, Huggett J, et al. (2009) The MIQE guidelines: minimum information for publication of quantitative real-time PCR experiments. Clin Chem 55: 611-622.
39. Schmittgen TD, Livak KJ (2008) Analyzing real-time PCR data by the comparative C(T) method. Nat Protoc 3: 1101-1108.
40. Dina C, Bouatia-Naji N, Tucker N, Delling FN, Toomer K, et al. (2015) Genetic association analyses highlight biological pathways underlying mitral valve prolapse. Nat Genet 47: 1206-1211.

SUPPLEMENTAL MATERIAL

PHACTR1 is a genetic susceptibility locus for Fibromuscular Dysplasia

Soto Romuald Kiando, MSc,[1,2]; Nathan R. Tucker, PhD,[3]; Luis-Jaime Castro-Vega, MD, PhD,[1,2], ; Alexander Katz, MD,[4]; Valentina D'Escamard, PhD,[5]; Cyrielle Tréard, MSc[1,2]; Daniel Fraher, PhD[3]; Juliette Albuisson, MD, PhD[1,2,6,7]; Daniella Kadian-Dodov, MD[5]; Zi Ye, MD, PhD[8]; Erin Austin, PhD[8]; Min-Lee Yang, PhD,[4]; Kristina Hunker, PhD,[4]; Cristina Barlassina, PhD,[9]; Daniele Cusi, MD, PhD,[10]; Pilar Galan, PhD,[11]; Jean-Philippe Empana, MD, PhD,[1,2]; Xavier Jouven, MD, PhD,[1,2,12]; Anne-Paule Gimenez-Roqueplo, MD, PhD[1,2,7]; Patrick Bruneval, MD, PhD,[1,2]; Esther Soo Hyun Kim, MD[13]; Jeffrey W. Olin, MD[5]; Heather L. Gornik, MD[13]; Michel Azizi, MD, PhD[14,15]; Pierre-François Plouin, MD, PhD[1,2,14]; Patrick T. Ellinor, MD, PhD[3]; Iftikhar J. Kullo, MD, PhD,[8]; David J. Milan, MD[3]; Santhi K. Ganesh, MD,[4]; Pierre Boutouyrie, MD, PhD[1,2,16]; Jason C. Kovacic, MD, PhD,[5]; Xavier Jeunemaitre, MD, PhD[1,2,6,7]; Nabila Bouatia-Naji, PhD[1,2*]

1. INSERM, UMR970 Paris Cardiovascular Research Center (PARCC), Paris F-75015, FRANCE
2. Paris-Descartes University, Sorbonne Paris Cité, Paris 75006, FRANCE
3. Cardiovascular research Center, Massachusetts General Hospital, Charlestown, MA 02114, USA, Program in Medical and Population Genetics, The Broad Institute of Harvard and MIT, Cambridge, MA 02142
4. Department of Internal Medicine and Department of Human Genetics, University of Michigan, Ann Arbor, MI 48109, USA
5. The Zena and Michael A. Wiener Cardiovascular Institute, Icahn School of Medicine, Marie-Josée and Henry R. Kravis Cardiovascular Health Center at Mount Sinai, One Gustave L. Levy Place, Box 1030 New York, NY 10029, New York, NY, USA
6. AP-HP, Referral Center for Rare Vascular Diseases, Hôpital Européen Georges Pompidou, Paris, F-75015, FRANCE.
7. AP-HP, Department of Genetics, Hôpital Européen Georges Pompidou, Paris, F-75015, FRANCE
8. Department of Medicine, Division of Cardiovascular Diseases, Mayo Clinic, Rochester, MN 55905, USA
9. Dept. of Health Sciences, Genomic and Bioinformatics Unit, Viale Ortles 22/4, Milano, Chair and Graduate School of Nephrology, University of Milano, Division of Nephrology, San Paolo Hospital, Milano, 20142, ITALY
10. Institute of Biomedical Technologies, Italian National Centre of Research, Via F.lli Cervi 93, 20090 Segrate – Milano, ITALY
11. Nutritional Epidemiology Research Group, Sorbonne-Paris-Cité, UMR University of Paris 13/Inserm U-557/INRA U-1125/CNAM, Bobigny, France F-93017, Bobigny, FRANCE
12. AP-HP, Department of Cardiology, Hôpital Européen Georges Pompidou, Paris, F-75015, FRANCE
13. Department of Cardiovascular Medicine, Cleveland Clinic Heart and Vascular Institute, Cleveland, OH 44195, USA.
14. AP-HP, Department of Hypertension, Hôpital Européen Georges Pompidou, Paris, F-75015, FRANCE
15. INSERM, Clinical Investigation Center CIC1418, Hôpital Européen Georges Pompidou, Paris, F-75015, FRANCE
16. AP-HP, Department of Pharmacology, Hôpital Européen Georges Pompidou, Paris, F-75015, FRANCE

Supp Table 1. Association of SNPs selected for follow-up in ARCADIA and PPS3 case control analyses.

SNP	Locus[a]	Chr	EA	EA Frequency	Sample	Discovery		Follow-up I	
						OR (95% CI)	P Value	OR (95% CI)	P Value
rs6703791	*IGSF3*	1	C	0.17	All	1.55 (1.21-1.99)	4.98×10^{-4}	0.92 (0.77-1.11)	0.39
				0.16	Females	1.72 (1.28-2.30)	3.29×10^{-4}	0.93 (0.75-1.16)	0.51
				0.17	Multifocals	1.65 (1.24-2.20)	6.23×10^{-4}	0.88 (0.70-1.10)	0.25
rs1332844[b]	*PHACTR1*	6	C	0.37	All	1.24 (1.00-1.53)	0.05	1.27 (1.10-1.48)	1.72×10^{-3}
				0.36	Females	1.38 (1.07-1.76)	0.01	1.55 (1.29-1.86)	2.02×10^{-6}
				0.37	Multifocals	1.30 (1.01-1.66)	0.04	1.36 (1.13-1.62)	9.34×10^{-4}
rs9369640[b]	*PHACTR1*	6	C	0.37	All	1.26 (1.02-1.56)	0.03	1.29 (1.11-1.49)	8.45×10^{-4}
				0.36	Females	1.40 (1.10-1.80)	7.23×10^{-3}	1.49 (1.25-1.78)	1.03×10^{-5}
				0.37	Multifocals	1.32 (1.03-1.69)	0.03	1.35 (1.14-1.61)	7.50×10^{-4}
rs9349379	*PHACTR1*	6	A	0.59	All	1.65 (1.32-2.07)	1.47×10^{-5}	1.32 (1.12-1.54)	7.21×10^{-4}
				0.57	Females	1.99 (1.51-2.62)	8.16×10^{-7}	1.46 (1.20-1.77)	1.23×10^{-4}
				0.59	Multifocals	1.75 (1.34-2.29)	4.41×10^{-5}	1.37 (1.13-1.65)	1.39×10^{-3}
rs4357136[b]	*PHACTR1*	6	T	0.53	All	1.26 (1.02-1.55)	0.03	1.04 (0.90-1.21)	0.58
				0.51	Females	1.29 (1.01-1.65)	0.04	1.06 (0.88-1.27)	0.53
				0.53	Multifocals	1.17 (0.92-1.50)	0.21	1.12 (0.94-1.34)	0.21
rs436268[b]	*PHACTR1*	6	C	0.93	All	1.37 (0.87-2.16)	0.18	0.92 (0.67-1.26)	0.61
				0.93	Females	1.28 (0.76-2.16)	0.35	1.02 (0.69-1.53)	0.91
				0.93	Multifocals	1.40 (0.81-2.42)	0.23	0.91 (0.63-1.31)	0.6
rs2399553	*LOC101928322 / CELF2*	10	T	0.21	All	1.54 (1.22-1.96)	3.18×10^{-4}	0.86 (0.72-1.02)	0.08
				0.21	Females	1.69 (1.27-2.26)	3.67×10^{-4}	0.84 (0.69-1.04)	0.11
				0.21	Multifocals	1.56 (1.19-2.06)	1.44×10^{-3}	0.83 (0.68-1.03)	0.09
rs10837607	*LRRC4C*	11	A	0.12	All	1.52 (1.14-2.02)	3.95×10^{-3}	1.08 (0.87-1.33)	0.49
				0.12	Females	1.52 (1.08-2.15)	0.02	1.06 (0.82-1.35)	0.67
				0.12	Multifocals	1.76 (1.27-2.43)	6.29×10^{-4}	1.17 (0.91-1.48)	0.22
rs1303	*SERPINA1*	14	G	0.23	All	1.57 (1.24-1.98)	2.13×10^{-4}	1.01 (0.85-1.19)	0.92
				0.25	Females	1.39 (1.04-1.86)	0.03	1.00 (0.82-1.23)	0.97
				0.23	Multifocals	1.65 (1.25-2.18)	4.21×10^{-4}	1.02 (0.83-1.25)	0.85
rs11636613	*TSPAN3*	15	A	0.27	All	1.44 (1.15-1.80)	1.30×10^{-3}	1.03 (0.88-1.21)	0.71
				0.27	Females	1.44 (1.09-1.89)	9.64×10^{-3}	0.97 (0.80-1.18)	0.77
				0.27	Multifocals	1.68 (1.29-2.18)	1.03×10^{-4}	0.96 (0.79-1.16)	0.66
rs7188856	*NAGPA*	16	A	0.3	All	1.42 (1.14-1.75)	1.49×10^{-3}	0.93 (0.80-1.09)	0.38
				0.33	Females	1.26 (0.98-1.61)	0.07	0.92 (0.77-1.12)	0.42
				0.3	Multifocals	1.54 (1.21-1.98)	5.88×10^{-4}	0.99 (0.82-1.19)	0.89
rs944895	*LAMA5*	20	A	0.61	All	1.45 (1.17-1.81)	8.70×10^{-4}	1.07 (0.92-1.25)	0.39
				0.61	Females	1.40 (1.08-1.82)	0.01	1.04 (0.87-1.26)	0.65
				0.61	Multifocals	1.30 (1.01-1.68)	0.04	1.13 (0.93-1.36)	0.21
rs228104	*SLC37A1*	21	G	0.09	All	1.55 (1.12-2.14)	8.27×10^{-3}	1.02 (0.81-1.28)	0.88
				0.08	Females	1.96 (1.32-2.90)	8.12×10^{-4}	1.05 (0.80-1.38)	0.73
				0.09	Multifocals	2.25 (1.59-3.20)	5.82×10^{-6}	0.98 (0.75-1.30)	0.91

SNP, Single Nucleotide polymorphism; Chr, Chromosome; EA, Effect allele; OR, Odds-Ratio; CI, Confidence Intervalle.

[a]For genic SNPs, the relevant gene is listed; for intergenic SNPs nearest upstream and downstream genes are listed. [b]SNPs added despite a P value in the discovery stage were above the threshold given their location in or near *PHACTR1* , the most associated locus.

Supp Table 2. Association of rs9349379 in PHACTR1 with carotid artery parameters in PPS3 controls.

		Additive model			Recessive model		
Carotid geometry	**Sample[a]**	**$\beta_{q.t.}$ (95% CI)**	**β**	**P value**	**$\beta_{q.t.}$ (95% CI)**	**β**	**P value**
Intima–media thickness (IMT), μm	All	0.069 (0.033 to 0.106)	11.54	1.97×10^{-4}	0.077 (0.041 to 0.114)	18.73	3.19×10^{-5}
	Females	0.097 (0.040 to 0.154)	15.15	8.38×10^{-4}	0.096 (0.04 to 0.153)	22.3	9.07×10^{-4}
	Males	0.054 (0.006 to 0.101)	9.22	0.03	0.067 (0.02 to 0.114)	16.57	0.006
External diameter (Dext), mm	All	0.028 (-0.006 to 0.062)	0.03	0.1	0.037 (0.003 to 0.071)	0.05	0.03
	Females	0.021 (-0.037 to 0.079)	0.02	0.48	0.026 (-0.032 to 0.084)	0.04	0.38
	Males	0.037 (-0.01 to 0.085)	0.04	0.13	0.049 (0.001 to 0.096)	0.07	0.05
Internal diameter (Dint), mm	All	0.003 (-0.031 to 0.038)	0.006	0.85	0.009 (-0.025 to 0.044)	0.02	0.59
	Females	-0.019 (-0.08 to 0.041)	-0.01	0.53	-0.014 (-0.074 to 0.046)	-0.009	0.65
	Males	0.019 (-0.03 to 0.068)	0.02	0.46	0.025 (-0.024 to 0.074)	0.03	0.31
Wall to lumen ratio (WLR)	All	0.059 (0.022 to 0.096)	0.004	0.002	0.063 (0.025 to 0.10)	0.006	9.80×10^{-4}
	Females	0.095 (0.035 to 0.156)	0.006	0.002	0.092 (0.032 to 0.152)	0.009	0.003
	Males	0.04 (-0.01 to 0.089)	0.003	0.12	0.048 (-0.001 to 0.098)	0.005	0.05
Circumferential wall stress at DBP (CWS-DBP), kPa	All	-0.054 (-0.091 to -0.017)	-0.71	0.004	-0.063 (-0.10 to -0.027)	-1.17	0.001
	Females	-0.065 (-0.125 to -0.005)	-0.74	0.03	-0.067 (-0.127 to -0.007)	-1.04	0.03
	Males	-0.051 (-0.100 to -0.002)	-0.71	0.04	-0.065 (-0.114 to -0.016)	-1.27	0.01
Arterial Stiffness							
Stiffness $m \times s^{-1}$	All	-0.028 (-0.066 to 0.01)	0.05	0.15	-0.025 (-0.063 to 0.013)	0.07	0.2
	Females	-0.007 (-0.066 to 0.053)	0.009	0.82	-0.02 (-0.08 to 0.039)	0.07	0.5
	Males	-0.043 (-0.093 to 0.006)	0.07	0.09	-0.026 (-0.076 to 0.023)	0.06	0.3
Cross-sectional distensibility (Dist), $kPa^{-1} \times 10^{-3}$	All	-0.03 (-0.067 to 0.008)	-0.34	0.13	-0.028 (-0.066 to 0.01)	-0.41	0.14
	Females	-0.005 (-0.065 to 0.054)	-0.07	0.86	-0.016 (-0.075 to 0.044)	-0.19	0.61
	Males	-0.047 (-0.096 to 0.002)	-0.51	0.06	-0.036 (-0.085 to 0.014)	-0.51	0.16
Cross-sectional compliance (Comp), $mm^2 \times kPa^{-1} \times 10^{-3}$	All	-0.026 (-0.064 to 0.012)	-0.07	0.18	-0.023 (-0.061 to 0.015)	-0.08	0.23
	Females	-0.02 (-0.081 to 0.041)	-0.06	0.53	-0.033 (-0.094 to 0.028)	-0.13	0.29
	Males	-0.031 (-0.081 to 0.019)	-0.08	0.23	-0.017 (-0.067 to 0.033)	-0.04	0.51
Young's elastic modulus kPa	All	0.012 (-0.027 to 0.051)	3.08	0.55	0.007 (-0.032 to 0.047)	4.25	0.72
	Females	-0.012 (-0.074 to 0.051)	-4.39	0.72	0.004 (-0.059 to 0.066)	4.23	0.91
	Males	0.029 (-0.022 to 0.08)	8.38	0.26	0.009 (-0.042 to 0.06)	4.04	0.73
Wall cross-sectional area (WCSA), mm^2	All	0.059 (0.024 to 0.094)	0.24	0.001	0.069 (0.034 to 0.104)	0.41	1.26×10^{-4}
	Females	0.079 (0.024 to 0.134)	0.3	0.005	0.081 (0.026 to 0.136)	0.47	0.004
	Males	0.048 (0.002 to 0.095)	0.2	0.04	0.063 (0.016 to 0.109)	0.38	0.008
Blood pressure							
Systolic blood pressure (SBP), mm Hg	All	0.052 (0.013 to 0.09)	0.7	0.009	0.041 (0.002 to 0.079)	0.78	0.04
	Females	0.066 (0.004 to 0.129)	0.96	0.038	0.055 (-0.007 to 0.117)	1.16	0.08
	Males	0.041 (-0.009 to 0.092)	0.51	0.11	0.030 (-0.021 to 0.081)	0.52	0.24
Diastolic blood pressure (DBP), mm Hg	All	0.0008 (-0.038 to 0.04)	0.008	0.97	-0.014 (-0.054 to 0.025)	-0.21	0.48
	Females	0.037 (-0.026 to 0.10)	0.36	0.26	0.023 (-0.040 to 0.086)	0.33	0.47
	Males	-0.024 (-0.075 to 0.027)	-0.24	0.35	-0.039 (-0.090 to 0.012)	-0.56	0.13
Mean blood pressure (MBP), mm Hg	All	0.021 (-0.018 to 0.06)	0.21	0.29	0.008 (-0.031 to 0.046)	0.1	0.7
	Females	0.048 (-0.015 to 0.11)	0.49	0.14	0.037 (-0.025 to 0.10)	0.56	0.24
	Males	0.0009 (-0.05 to 0.052)	0.0002	0.97	-0.015 (-0.066 to 0.036)	-0.22	0.57
Central pulse pressure (CPP), mm Hg	All	0.061 (0.023 to 0.099)	0.62	0.002	0.065 (0.026 to 0.103)	0.94	9.09×10^{-4}
	Females	0.036 (-0.024 to 0.096)	0.39	0.24	0.043 (-0.017 to 0.103)	0.67	0.16
	Males	0.081 (0.031 to 0.13)	0.79	0.001	0.08 (0.03 to 0.129)	1.1	0.002

[a]All (N=2,458), Female (N=975) and Male (N=1,483) participants of the PPS3.

βq.t. is the effect size estimated using quantile-transformed values and β is the effect size estimated using untransformed values. P values indicated are for the analyses of quantile-transformed values. Age, sex, body surface area were included as covariates in all regression models. MBP was an additional covariate for the following carotid geometry (IMT, Dext, Dint and WLR) and arterial stiffness (Stiffness, Dist, Comp and WCSA) parameters.

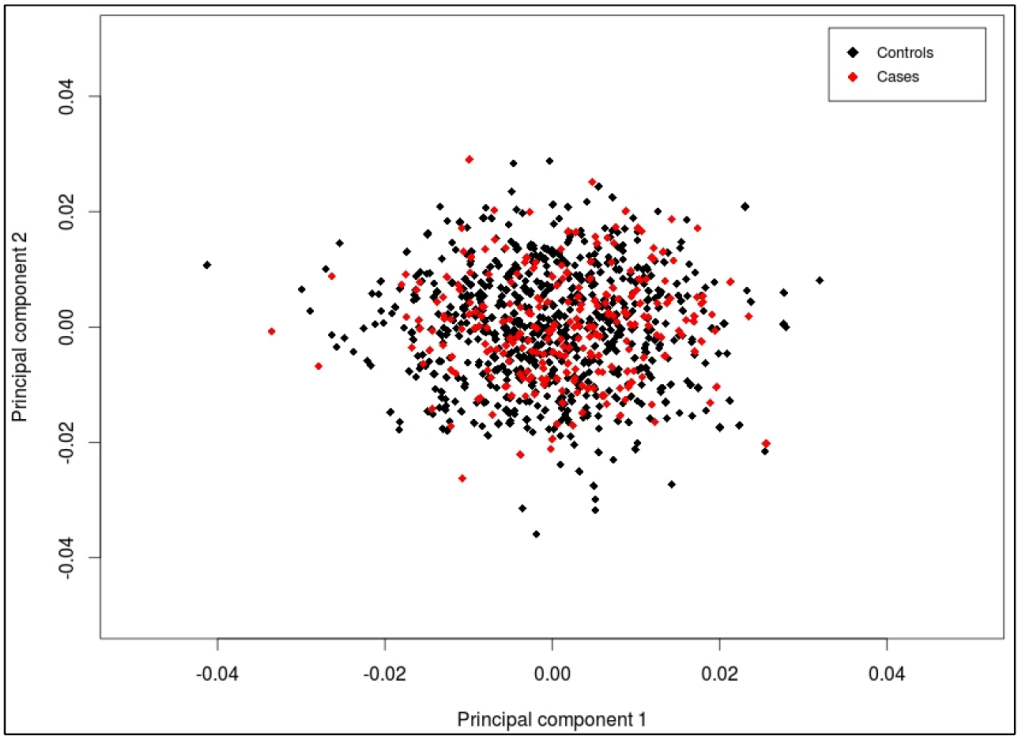


Figure S1. Principal component analysis using genotypes from FMD cases and controls of the discovery stage

Principal component analysis using genotypes from FMD cases and controls of the discovery stage. Data used is post quality control filtration for non-European ancestry origin and relatedness

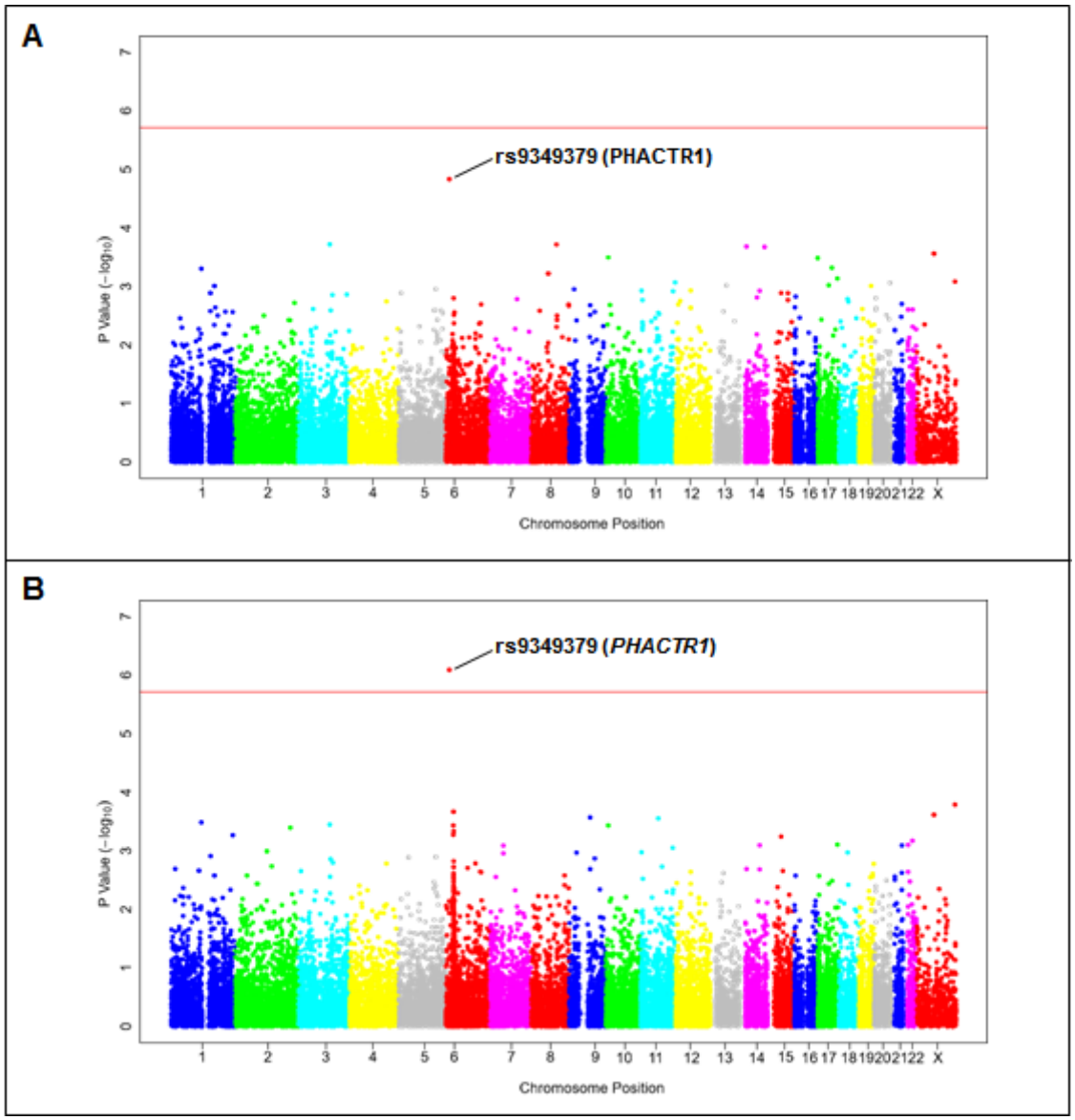


Figure S2. Association between Single-Nucleotide Polymorphisms (SNPs) and Fibromuscular Dysplasia (FMD).

(A) Association between SNPs and FMD in global analysis. (B) Association between SNPs and FMD in females only. The red line indicates the Bonferroni adjusted significance threshold for 25,606 common variants tested set to $P=1.95 \times 10^{-6}$.

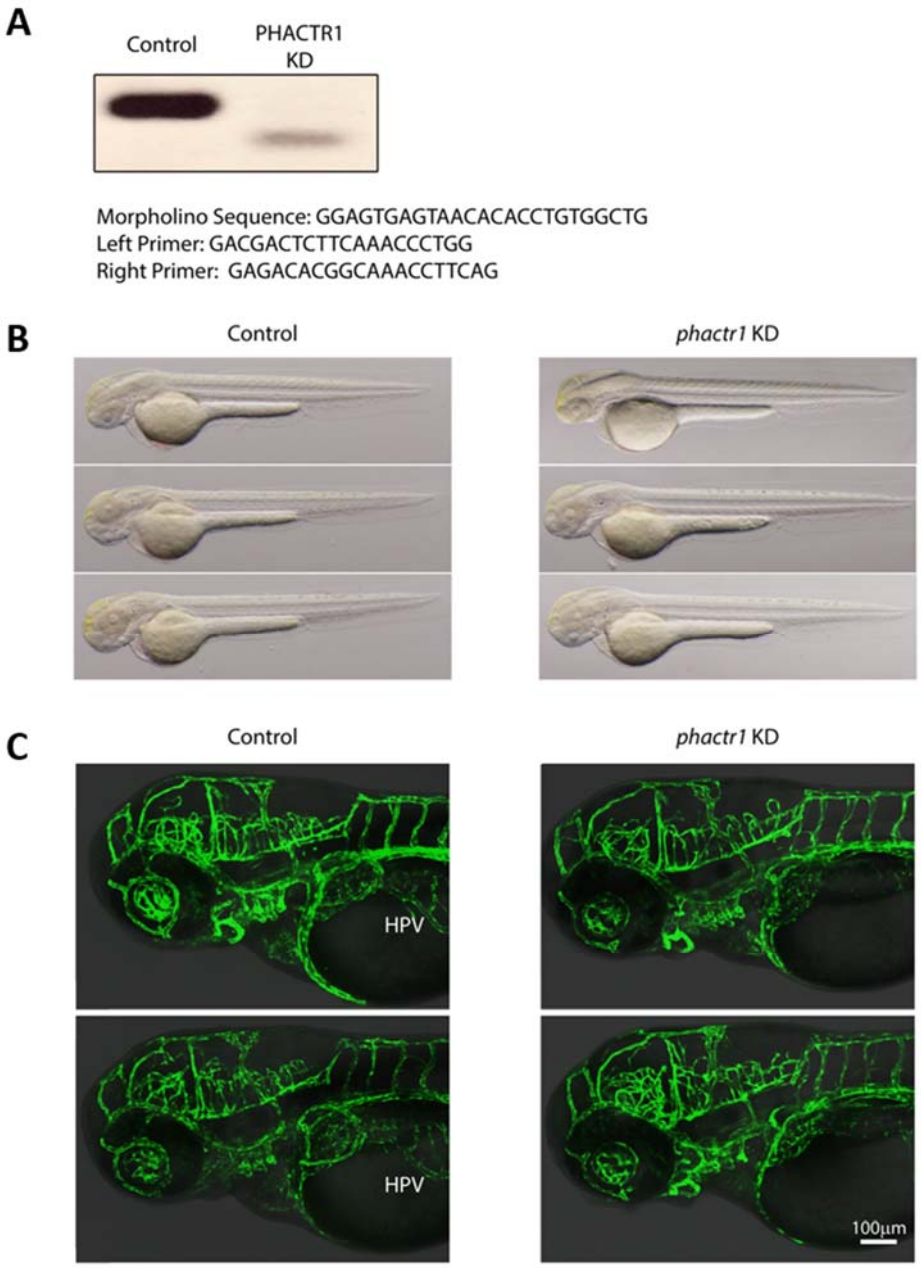


Figure S3. Phenotypic evaluation of *phactr1* knockdown zebrafish

(A) RT-PCR evaluation of splice alteration observed following microinjection of *phactr1* morpholino (PHACTR1 KD). (B) Brightfield micrographs of overt morphology at 60 hours post fertilization. (C) Two-dimensional projections obtained from z-series confocal images in the head and trunk of control and *phactr1* knockdown zebrafish embryos. Green represents the vascular endothelium as marked by EGFP. Greyscale represents the corresponding DIC brightfield image of the fish head and trunk region. HPV indicates the relative position of the developing hepatic portal vein

TROISIEME PARTIE : DISCUSSION ET PERSPECTIVES

CHAPITRE V : DISCUSSION

Dans ce chapitre, nous articulerons la discussion des résultats obtenus autour des deux axes de notre travail de thèse à savoir l'étude d'exome et l'étude d'association de variants rares et fréquents avec la puce *Exome-chip*. Nous présenterons ensuite les perspectives de ce projet et ferons une conclusion générale.

V.1. WES : un outil pour l'identification des gènes de prédisposition à la DFM

Les approches de séquençage d'exome ont montré leur performance pour l'identification de la composante génétique de nombreuses pathologies humaines (Rabbani, Tekin et al. 2014). Toutefois, le succès de ce type d'approche dépend de plusieurs facteurs relatifs aussi bien au design de l'étude qu'à l'approche technologique en soi.

Le design de notre étude s'est heurté à la limite éthique de l'exploration angiographique des apparentés non malades de la DFM au sein des différentes familles. En effet, la possibilité de caractériser phénotypiquement l'ensemble des membres d'une famille permet d'améliorer le design de l'étude d'exome et d'optimiser la puissance pour l'identification des variants génétique de prédisposition. On peut ainsi filtrer les données de séquençage du cas index avec celles d'un apparenté non malade permettant de réduire les potentielles mutations causales partagées, suivre la ségrégation de variants d'intérêt au sein de la famille étudiée ou encore explorer de nouvelles hypothèses physiopathologiques comme l'implication de mutations *de-novo* (Bamshad, Ng et al. 2011). D'ailleurs, dans la suite de la caractérisation des bases génétiques de la DFM par une approche d'étude d'exome conduite au sein de notre équipe

récemment (données non publiées), l'étude de deux trios constitués chacun d'un cas sporadique de DFM avec ses parents non atteints de DFM a permis d'exclure l'implication de ce type de mutations pour ces 2 familles.

En outre, les études de séquençage d'exome en général et celle conduite dans le cadre de mon projet de thèse particulièrement, ne permettent pas en réalité une exploration complète de l'ensemble de la variabilité potentiellement codante du génome ou ayant un rôle dans la régulation de l'expression des gènes (Majewski et al., 2011). Il s'agit là d'une des insuffisances technique du WES qui n'offre qu'une couverture limitée des variations structurales qui peuvent représenter jusqu'à 20 % de la variabilité par individu (Frazer, Murray et al. 2009) et pour lesquels l'implication dans les maladies humaines a déjà été rapportée (Sebat, Lakshmi et al. 2007, Walsh, McClellan et al. 2008). Plus spécifiquement, les *CNVs* sont susceptibles d'altérer le dosage du gène de façon directe en modifiant le nombre de copies d'un gène, ou indirectement en modifiant le nombre d'éléments régulateurs qui augmentent ou répriment son expression (Wain, Armour et al. 2009). De plus, les microARNs ne sont que très partiellement couverts par WES.

Par ailleurs, l'hypothèse même de l'implication spécifique de variants exoniques, excluant *de facto* les régions non codantes pouvant abriter de nombreux mécanismes de régulation de l'expression des gènes (Dunham, Kundaje et al. 2012, Haraksingh and Snyder 2013) pose aussi quelques limitations sur la performance du WES. Aussi, depuis les papiers fondateurs de Ng et collaborateurs (Ng, Levy et al. 2008, Ng, Turner et al. 2009, Ng, Buckingham et al. 2010), le taux de succès des approches par WES dans l'identification des gènes responsables de maladies humaines est difficile à déterminer parce que les études négatives ont tendance à ne pas être publiés (Boycott, Vanstone et al. 2013). L'abondance d'exomes constituant les bases de données (Exome Aggregation Consortium 2016) pourrait

donc être le signe que la ruée vers l'exome n'a pas toujours rapporté que des pépites. Ainsi, même s'il apparait que le WGS devrait représenter à première vue une alternative au WES (Belkadi, Bolze et al. 2015), il n'en demeure pas moins que l'analyse de l'implication de l'ensemble de la variabilité génomique au déterminisme génétique des maladies humaines est un véritable défi (Precone, Del Monaco et al. 2015). Ceci, d'une part à cause de l'augmentation considérable de la quantité de données par rapport au WES (environ 200 fois plus de données), nécessitant une puissance informatique considérable pour la génération, le stockage et l'analyse (Stranneheim and Wedell 2016) ; d'autre part parce que l'implication fonctionnelle des régions introniques, non traduites, des séquences promotrices, amplificateurs, répresseurs du génome est toujours en cours dans le cadre de projets collaboratifs internationaux (Dunham, Kundaje et al. 2012).

V.2. Etude d'association gène-entiers de variants rares

Nous avons rapporté l'association de variants rares génotypés par la puce *Exome-chip* chez 249 cas de DFM et 689 témoins pour plusieurs gènes candidats identifiés par l'analyse des données d'exomes chez 16 cas non apparentés de DFM, et ceux identifiés comme causatifs d'artériopathies proches de la DFM (Kiando, Barlassina et al. 2015). Nous avons également étendu l'analyse à l'ensemble des variants rares disponibles sur la puce *Exome-chip*, qui ont passé les étapes de contrôle qualité de données de génotypage, et qui étaient polymorphes dans notre échantillon d'étude. Il s'agissait au total de 62767 variants rares répartis sur 9967 gènes dont l'association avec la DFM a été testée par la méthode SKAT-O (Lee, Wu et al. 2012). L'analyse de l'ensemble de ces variants rares n'a pas permis d'identifier des gènes significativement associés à la DFM après correction de Bonferroni pour tests multiples, ni des profils candidats pour la physiopathologie de DFM après exploration des gènes présentant les

signaux d'association les plus significatifs. Des approches complémentaires ont été menées incluant des *Burden* tests (Li and Leal 2008); le SKAT-CR (Ionita-Laza, Lee et al. 2013) évaluant l'implication simultanée de variants rares et fréquents au sein du même gène ; la re-caractérisation de l'entité génique a été aussi explorée ainsi que la redéfinition du type de variants pris en compte. Cependant, aucune de ces différentes approches n'a permis d'identifier des variants génétiques d'intérêt pour la DFM entrainant au contraire une inflation significative de l'erreur de type I. A cette étape, une évaluation de l'intérêt des puces exoniques dans le cadre d'étude d'association gène-entiers des variants rares (*Exome-Wide Association Studies*) s'impose.

Les puce exoniques, en particulier l'*Exome-chip* font partie de cette nouvelle génération de puces conçues avec l'ambition d'offrir une alternative intéressante au WES en permettant une couverture exhaustive par génotypage des variants exoniques rares (Exome Chip Design 2014). Cependant, même si elles sont constituées de près de 84% de variants rares, elles n'offrent pas la performance de la couverture du WES. Au-delà des limites en termes de couverture intrinsèque au WES, l'approche du génotypage des variants rares par l'*Exome-chip* présente aussi d'autres insuffisances. Il s'agit notamment du concept même de puce, avec des positions génomiques prédéfinies conduisant à l'exclusion de l'implication des variants exoniques non couverts sans réelles possibilités d'imputation (Martin, Tse et al. 2014) (Tantoso, Wong et al. 2014, Kim, Lee et al. 2015). De plus cette approche fait aussi l'impasse sur les variations structurales, les variants extrêmement rares ou privés hérités au sein de familles et les variations *de-novo* au sein d'un individu.

Ainsi, les études d'association de variants rares génotypés avec l'*Exome-chip* ont surtout été performantes dans les cas de stratégies de reséquençage (Huyghe, Jackson et al. 2013) ou au sein de très grandes cohortes pour lesquelles le coût du WES aurait été prohibitif (Wessel,

Chu et al. 2015, Jackson, Ntalla et al. 2016). Dans une étude récente évaluant la puissance empirique des puces exoniques, Page et collaborateurs estiment qu'il est nécessaire de génotyper 20000 à 30000 individus pour un effet modeste de risque attribuable en population entre 0,5% et 1% avec une puissance de 80% (Page, Baranzini et al. 2015), ce qui est tout à fait inenvisageable pour le cas de la DFM. En conséquence, il serait donc tout à fait hasardeux de conclure à une implication causale des gènes candidats identifiés par notre étude sans de solides arguments fonctionnels.

V.3. Etude d'association des variants fréquents de l'*Exome-chip*

L'étude d'association des variants fréquents a été conduite en complément à l'analyse des variants rares dans le but d'exploiter la totalité du potentiel de la puce *Exome-chip*. Cette étude a permis la mise en évidence puis la réplication dans quatre cohortes indépendantes du polymorphisme rs9349379, premier variant génétique de prédisposition identifié pour la DFM

V.3.1. *PHACTR1* un locus majeur pour les pathologies vasculaires

Le polymorphisme rs9349379 est situé dans l'intron du gène *PHACTR1* (*phosphatase and actin regulator 1*), situé dans une région hautement conservée du génome suggérant qu'il a une implication majeure dans les processus biologiques notamment la régulation de l'activité synaptique et la morphologie dendritique (Allen, Greenfield et al. 2004). Toutefois, il existe très peu de publications étudiant sa fonction. Il interagit avec les filaments d'actine et est impliqué dans la réorganisation du cytosquelette. Des études expérimentales ont identifié un rôle central pour la protéine PHACTR1 dans la formation du tube vasculaire et la polymérisation de l'actine, ce qui suggère qu'il a un rôle potentiel dans les processus d'adhésion et de migration cellulaire au cours de l'angiogenèse induite par la voie du facteur de croissance

de l'endothélium vasculaire (VEGF) (Jarray, Allain et al. 2011, Allain, Jarray et al. 2012). Cependant, ces résultats n'ont pu être répliqués par Beaudoin et collègues (Beaudoin, Gupta et al. 2015). De plus, la régulation positive de *PHACTR1* par le facteur de croissance transformant (TGF-β) a été décrite dans un modèle cellulaire de cancer du sein (Fils-Aime, Dai et al. 2013).

PHACTR1 a déjà été précédemment identifié comme associé à des traits d'intérêt dans les pathologies cardiovasculaires et neurovasculaires (Burdett T (EBI) , Welter, MacArthur et al. 2014). *PHACTR1* a été décrit comme facteur de risque pour la maladie coronaire et l'infarctus du myocarde (CAD-MI) (Coronary Artery Disease Genetics 2011, Beaudoin, Gupta et al. 2015), la migraine (Freilinger, Anttila et al. 2012, Anttila, Winsvold et al. 2013) et plus récemment avec la dissection de l'artère cervicale (CeAD) (Debette, Kamatani et al. 2015). De façon intéressante, l'allèle à risque pour la DFM du polymorphisme rs9349379[A] est le même pour la migraine et la dissection de l'artère cervicale tandis qu'il est protecteur pour la maladie coronaire et l'infarctus du myocarde.

La migraine et la dissection de l'artère cervicale présentent plusieurs caractéristiques communes entre elles d'une part, et avec la DFM d'autre part. En effet, la prévalence de la migraine est plus importante chez les sujets présentant une dissection de l'artère cervicale associée à un AVC ischémique (Metso, Tatlisumak et al. 2012). Elle est aussi plus importante chez les sujets de sexe féminin (Stewart, Shechter et al. 1994) qui représentent la majorité de majorité des atteintes de la DFM (Plouin, Perdu et al. 2007, Olin, Froehlich et al. 2012). De plus, respectivement 60% et 32,2% de l'ensemble des cas participant au registre américain de la DFM indiquent avoir des maux de tête en général et des maux de tête de type migraineux (Olin, Froehlich et al. 2012). Cependant, compte tenu des différences dans la définition du terme migraine d'une étude à l'autre, il n'est pas possible de conclure à un potentiel déterminisme génétique commun entre la DFM et la migraine. La dissection des artères cervicales (CeAD)

est l'une des causes majeures d'accident vasculaire cérébral ischémique chez le jeune adulte (Leys, Bandu et al. 2002, Debette and Leys 2009, Putaala, Metso et al. 2009), et on estime qu'elle est associée avec la DFM (Schievink 2001, Debette and Leys 2009). De façon générale, la fréquence de la DFM dans les séries de cas de CeAD varie considérablement, allant de 5,6% (Bejot, Aboa-Eboule et al. 2014) à près de 21% dans l'une des toutes premières cohortes de CeAD (62 patients) (d'Anglejan Chatillon, Ribeiro et al. 1990). Inversement, 12,1 % des cas de DFM présentent une dissection de l'artère cervicale (Olin, Froehlich et al. 2012) qui est désormais considérée comme l'une des complications probables de la DFM avec laquelle elle semble partager la même architecture multifactorielle avec toutefois une histoire familiale de la maladie moins fréquente pour le CeAD (Debette, Goeggel Simonetti et al. 2014), comparée à la DFM (Pannier-Moreau, Grimbert et al. 1997, Olin, Froehlich et al. 2012). Il existe cependant une différence notable avec la DFM, notamment l'absence de prédominance du sexe féminin chez les patients de CeAD.

Par contre, l'identification par notre étude du même locus génétique entre DFM et la maladie coronaire/infarctus du myocarde (CAD-MI) est beaucoup moins évidente. En effet, la DFM a été décrite uniquement pour les formes non athéromateuses (rares) de la maladie coronaire en majorité chez des sujets de sexe féminin sans facteur de risque cardiovasculaires (Michelis, Olin et al. 2014, Saw, Aymong et al. 2014, Prasad, Tweet et al. 2015). En outre, la présence de la DFM multifocale chez des sujets présentant déjà une dissection spontanée de l'artère coronaire n'a pas été associée avec un risque de récidive, ni d'insuffisance cardiaque congestive ou d'infarctus du myocarde (Prasad, Tweet et al. 2015).

Dans le cadre spécifique de notre étude, la pléiotropie (Sivakumaran, Agakov et al. 2011, Wagner and Zhang 2011) observée pour *PHACTR1* donne une vision globale du niveau de connectivité génétique entre la DFM et les différentes pathologies/traits étudiés. Elle ne permet

cependant pas d'anticiper une classification hiérarchique dans leur survenue. Toutes ces possibilités feront donc l'objet d'un approfondissement dans les perspectives de ce travail de thèse dans le but de discriminer les loci de prédisposition à la DFM des ceux des pathologies associées, ainsi que les potentielles interactions entre elles permettant à terme une meilleure compréhension de la fonction du gène *PHACTR1*.

V.3.2. Du polymorphisme à l'expression du gène

Les *GWAS* ont permis l'identification de milliers de polymorphismes associés aux maladies humaines (Welter, MacArthur et al. 2014). Pour la plupart d'entre eux, situés dans des régions introniques ou intergéniques du génome, il est difficile d'établir leurs implications fonctionnelles qui peuvent mobiliser des mécanismes complexes (Maurano, Humbert et al. 2012, Albert and Kruglyak 2015).

L'un des mécanismes possibles est notamment la modification de l'expression d'un gène par l'action directe ou indirecte d'un SNP sur des éléments régulateurs : on parle alors d'*eQTLs* (Schadt, Monks et al. 2003, Morley, Molony et al. 2004). Plusieurs études ont révélé un enrichissement en *eQTLs* des polymorphismes identifiés par les *GWAS* (Dimas, Deutsch et al. 2009, Nicolae, Gamazon et al. 2010), témoignant de leur utilité dans la compréhension des mécanismes sous-jacent la physiopathologie des traits étudiés. Aussi, dans le cadre de notre étude, nous avons montré que le polymorphisme rs9349379, facteur de prédisposition identifié pour la DFM module le niveau de l'expression de *PHACTR1* dans les cultures primaires de fibroblastes humains issus de biopsies de peau chez 86 individus (44 cas de DFM, 24 témoins et 12 de statut indéterminé). Il s'agit plus précisément d'une expression plus importante de *PHACTR1* chez les individus porteurs de l'allèle rs9349379-A. Notons que des observations

similaires sont rapportées pour l'expression de *PHACTR1* dans les artères coronaires humaines (Beaudoin, Gupta et al. 2015) et pour plusieurs sites artériels dans la base Génotype-Tissue projet (GTEX) qui décrit l'expression de gène dans différents tissus humains et catalogue des milliers de corrélations entre *eQTLs* et tissus d'intérêts (Consortium 2013, Consortium 2015).

Il convient cependant de noter que l'identification d'un eQTL ne fournit qu'une preuve indirecte de la liaison entre le génotype et la transcription du gène. En outre, la situation idéale voudrait qu'on mesure l'effet de l'eQTL dans des cellules normales, représentatives des cellules d'origines pour le trait étudié (Edwards, Beesley et al. 2013), ce qui n'a pas été possible dans le cas de la DFM compte tenu de la rareté des pièces histologiques du fait que les patients sont désormais rarement traités par chirurgie (Persu, Giavarini et al. 2014). Ainsi, pour les tissus difficiles à obtenir, à partir des sites primaires, des cellules souches pluripotentes induites (iPSCs) (Robinton and Daley 2012), qui peuvent être différenciées en différents types de cellules, fournissent une alternative prometteuse. Le type cellulaire est particulièrement important car on estime que près de 50% à 90% des *eQTLs* sont tissus dépendants (Dimas, Deutsch et al. 2009, Nica, Parts et al. 2011, Curtis, Shah et al. 2012, Li, Seo et al. 2013). Aussi des explorations fonctionnelles complémentaires sont-elles nécessaires pour confirmer la pertinence du mécanisme sous-jacent.

V.3.3. Du locus au modèle animal

L'un des plus grands défis dans la validation fonctionnelle des loci identifiés formellement par les études d'association est la capacité à évaluer avec précision le rôle des polymorphismes et des gènes identifiés dans un organisme entier, ouvrant ainsi la voie à la compréhension globale de l'architecture génétique ainsi que des mécanismes cellulaires et

moléculaires. La souris (*Mus musculus*) est généralement le modèle expérimental historique de mammifère par excellence (Petersen, Amundadottir et al. 2010, Flandez, Cendrowski et al. 2014, von Figura, Morris et al. 2014) en raison du fort degré de similarité avec le génome humain (> 80%) (Mouse Genome Sequencing, Waterston et al. 2002), de sa capacité à mimer de nombreux phénotypes de pathologies humaines (Ermann and Glimcher 2012) et à se reproduire relativement rapidement. Cependant, malgré les nombreuses techniques d'édition du génome comme les nucléases à doigts de zinc (ZNFs) (Urnov, Rebar et al. 2010), les nucléases effectrices de type activateur de transcriptions (TALENs) (Miller, Tan et al. 2011) et plus récemment la révolution CRISPR/Cas9 (Cho, Kim et al. 2013, Cong, Ran et al. 2013), la constitution de modèle expérimental de souris reste un processus avec de nombreuses contraintes éthiques, long et onéreux sans possibilité de pouvoir présumer du succès du processus surtout pour des pathologies hétérogènes d'un point de vue clinique et génétique. Ainsi des alternatives au modèle murin sont de plus en plus utilisées.

Il s'agit notamment d'autres modèles animaux parmi lesquels le poisson zèbre (*Danio rerio*) est un des spécimens les plus attractifs (Seth, Stemple et al. 2013, Dina, Bouatia-Naji et al. 2015). En effet le poisson zèbre partage une quantité considérable de son identité génétique codante avec les humains, avec lesquels de nombreuses similitudes existent pour les différents systèmes et organes. Depuis le séquençage de son génome en 2013 (Howe, Clark et al. 2013), il a été montré que 70% des gènes identifiés chez le poisson zèbre ont un orthologue chez l'humain. C'est donc un modèle animal valide pour lequel les nouvelles techniques d'édition de génome par CRICPR/Cas9 sont particulièrement efficientes (Varshney, Pei et al. 2015). En outre, le corps translucide des embryons de poisson zèbre facilite la visualisation microscopique non intrusive des organes et des processus biologiques in-vivo. Enfin, les poissons zèbres sont peu couteux, faciles à entretenir, avec la possibilité de produire un nombre considérable de spécimens, avec un développement vasculaire que l'on peut explorer tout juste 72 heures après

fécondation. Tout ceci a motivé l'utilisation de ce modèle animal dans le cadre de la validation fonctionnelle du locus identifié par notre étude d'association. Ainsi, l'invalidation de *Phactr1* par l'injection de morpholino anti-*phactr1* au stade du développement précoce chez le poisson zèbre a entrainé une dilatation des vaisseaux sanguins au niveau de la veine cave postérieure, indiquant un défaut vasculaire. Cependant, l'invalidation de l'expression du gène par la méthode des morpholinos (Nasevicius and Ekker 2000) est malheureusement un processus transitoire qui permet une étude rapide de la fonction du gène limité aux cinq premiers jours de développement de l'embryon. Aussi, une surexpression de *Phactr1* est toujours en cours chez le même modèle animal et il existe de toute évidence de nombreuses autres possibilités intéressantes de l'utilisation du poisson zèbre comme modèle animal expérimental dans la validation fonctionnelle des loci identifiés pour les maladies humaines (Chu and Sadler 2009, Stoletov, Fang et al. 2009, Baek, Fang et al. 2012).

CHAPITRE VI : PERSPECTIVES

A l'issue de ces trois années de recherches doctorales, nous pouvons envisager de nombreuses perspectives à court, moyen et long termes s'inscrivant dans la confirmation des variants génétiques identifiés pour la DFM et la mise en œuvre de nouvelles approches pour caractérisation de nouveaux déterminants génétiques de la DFM.

VI.1. Stratégies de reséquençage ciblé de loci candidats

Les stratégies de reséquençage ciblé de gènes au sein desquels des polymorphismes avaient été associés ont déjà permis d'identifier de nouveaux variants associés et/ou de nouveaux mécanismes pour les traits d'intérêt (Rivas, Beaudoin et al. 2011). Des approches encore plus puissantes peuvent être mises en œuvre par reséquençage ciblé avec une forte profondeur de lecture permettant d'optimiser d'avantage l'identification de nouveaux variants (Shea, Agarwala et al. 2011). Dans le cadre de notre projet doctoral, nous avons identifié formellement un locus de prédisposition pour la DFM et envisagé l'implication de plusieurs autres gènes candidats. A court terme, nous ferons un reséquençage ciblé de l'ensemble de ces loci pour tous les cas de DFM des différentes cohortes françaises de la DFM (> 700 cas), ceci dans le but de confirmer les variants identifiés et d'identifier de potentielles nouvelles mutations avec un impact fonctionnel pour la DFM. La liste des gènes pour le reséquençage ciblé reste encore à définir. Toutefois elle devra inclure le locus *PHACTR1* et l'ensemble des 22 candidats enrichis en mutations rares à fort potentiel fonctionnel identifié au terme de l'étude de séquençage d'exome. Aussi le reséquençage permettra d'étudier l'implication dans l'étiologie

de la DFM de séquences régulatrices situées dans les régions non codantes du génome des gènes candidats sélectionnés.

VI.2. Analyse de voies biologiques candidates

Les outils de priorisation de gènes et les études d'association de variants rares et fréquents dans une certaine mesure, malgré leur performance, ne permettent d'étudier que l'implication d'un locus ou d'une région génomique unique sans prise en compte des effets cumulés ou des nombreuses interactions dans le cadre d'une approche globale. Aussi, est-il sans doute nécessaire de réanalyser nos données de séquençage d'exome avec des filtres plus larges à la recherche de voies biologiques d'intérêt pour la DFM en rapport avec la matrice extracellulaire, la vasoconstriction et les fonction d'adhésion, de migration et de prolifération cellulaire.

De nombreuses outils comme DAVID (Huang da, Sherman et al. 2009), STRING (Szklarczyk, Franceschini et al. 2015), DEPICT (Pers, Karjalainen et al. 2015) ou encore ENDEAVOUR (Tranchevent, Barriot et al. 2008) permettent de mettre en œuvre ce type d'analyses à partir des bases de données de voies biologiques comme the Kyoto Encyclopedia of Genes and Genomes (KEGG) (Kanehisa and Goto 2000), Reactome (Croft, Mundo et al. 2014) et the Gene Ontology (GO) (Ashburner, Ball et al. 2000). En outre, ces outils étant très dynamiques et subissant des mises à jour régulières, il est sans doute utile de réanalyser les données avec les dernières versions à jour. Plusieurs analyses de ce type sont d'ores et déjà en cours à partir des données générées par le séquençage d'exome chez 16 cas apparentés de DFM et les données de séquençage de 10 nouveaux cas générés ultérieurement (données non publiées).

VI.3. Etude d'association pangénomique de de la DFM

L'*Exome-chip* a constitué une alternative intéressante pour étudier les variants génétiques de prédisposition à la DFM malgré sa couverture limitée par rapport à une puce de génotypage classique. A cet effet, le génotypage de 695 cas de DFM issus des cohortes françaises de l'HEGP avec la puce Illumina OmniExpressExome, une puce de génotypage haut débit incluant, en plus du contenu de l'*Exome-chip*, un nombre important de *SNPs* issus des trois différentes phases de HapMap (International HapMap, Altshuler et al. 2010) soit au total plus de 960K marqueurs permettant une couverture importante de variabilité fréquente du génome. Aussi, avons-nous déjà commencé le contrôle qualité des données de génotypage et envisageons la mise en œuvre des études d'associations à court terme avec une méthodologie similaire à celle utilisée pour l'identification du locus *PHACTR1* et une réplication au sein des cohortes américaines de la DFM (Olin, Froehlich et al. 2012). De plus, grâce à une initiative de cliniciens de notre centre de recherche en collaboration avec des équipes belges, un réseau européen de la DFM est en cours de constitution dont l'objectif est de permettre le recrutement de nouveaux cas de DFM. Cet effort a déjà permis d'initier le projet BEL-FMD qui est une cohorte multicentrique belge de cas de la DFM avec l'objectif de recruter environ 500 cas de DFM et 500 témoins dans les deux prochaines années ; ceci permettant de nouvelles possibilités de réplication/confirmation des variants identifiés pour la DFM.

Globalement, une meilleure caractérisation clinique des différentes cohortes et l'augmentation substantielle du nombre de cas permettront sans doute d'identifier d'autres loci de prédisposition à la DFM ; de stratifier les analyses sur le sexe et les différents phénotypes angiographiques ; et d'étudier l'implication des loci de prédisposition identifiés au déterminisme génétique de pathologies associées à la DFM. En outre, une partie des cohortes française faisant l'objet d'un suivi prospectif dans le cadre du protocole PROFILE (PROgression of FIbromuscular dyslasia LEsions), il serait intéressant de mettre en œuvre une

étude d'association de variants génétique identifiés à la survenue d'évènement de complications à long terme pour la DFM comme la dissection ou l'anévrisme (Kang, Jeon et al. 2015, Sullivan, Kopp et al. 2015).

Au-delà des aspects classiques des *GWAS* à savoir l'identification des polymorphismes de prédisposition, réplication des résultats, méta-analyse et cartographie fine des loci identifiés, il sera sans doute nécessaire de mettre en œuvre des nouvelles approches systémiques pour évaluer l'implication des régions non codantes du génome au déterminisme de la DFM (McPherson and Tybjaerg-Hansen 2016). Ces approches combineront des analyses in silico à partir de données de bases publiques comme RegulomeDB (Boyle, Hong et al. 2012)pour l'annotation de *SNPs* situés dans les régions non codantes du génome connus ou prédits comme étant de potentiels éléments régulateurs de l'expression des gènes, et analyses fonctionnelles incluant études d'expression dans des tissus d'intérêt et modèles animaux pour caractériser les bases génétiques de la DFM.

VI.4. Autres approches pour l'étude génétique de la DFM

Les études d'association génome entier (*GWAS*) et le séquençage d'exome (WES) basés sur des hypothèses de travail différentes permettent une approche complémentaire pour l'identification de la composante génétique des maladies humaines. Malgré la performance de ces deux techniques, il demeure toujours la question récurrente de l'héritabilité manquante persistante pour laquelle les réponses pourraient se trouver dans les parties non codantes du génome, au sein de variations structurales notamment (DeJesus-Hernandez, Mackenzie et al. 2011, Renton, Majounie et al. 2011).

L'étude du rôle fonctionnel des séquences non codantes du génome a connu ces dernières années des avancées significatives grâce notamment à des projets comme ENCODE.

Cela inclut les régions transcrites, les sites de liaison de facteurs de transcription, les régions avec des structures de chromatine distinctes ou encore des modifications d'histone spécifiques (ENCODE Project Consortium 2012). Les résultats obtenus indiquent clairement qu'une proportion importante des séquences non codantes du génome est susceptible d'avoir une implication fonctionnelle, de telle sorte que des mutations/mécanismes en cause pour les maladies humaines peuvent y être identifiés (Dunham, Kundaje et al. 2012). Avec la baisse continue du coût des techniques de séquençage nouvelle génération, il sera alors possible d'explorer l'ensemble de la variabilité génomique des individus à la recherche de mutations causales. En attendant que cette possibilité devienne une réalité, le NGS permet comme dans le cas de l'exome, de tester une hypothèse fonctionnelle particulière par la capture sélective de données de séquences à haut débit. Il s'agit du concept des "*omics*".

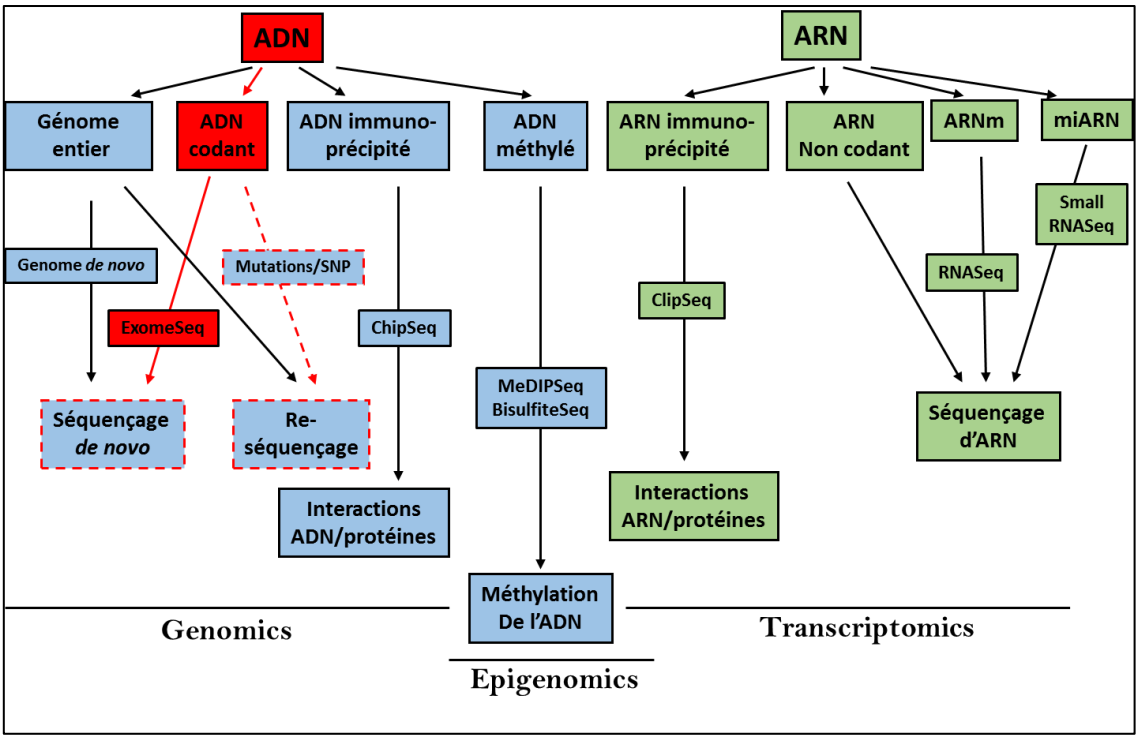


Figure VI-16. Application du NGS pour l'identification de la composante génétique des maladies humaines. Les champs en rouge sont ceux couverts intégralement ou partiellement dans le cadre de cette thèse. Tous les autres restent de potentielles pistes à explorer

Ainsi, le séquençage de l'ARN (RNA-seq) est une alternative efficace aux puces d'expression pour l'étude de l'expression des gènes, permettant la quantification de l'ensemble du transcriptome ou *transcriptomics* d'une cellule ou d'un tissu sous différentes conditions (Wang, Gerstein et al. 2009, Liu, Morley et al. 2015). Aussi, notre équipe est en train d'établir un projet d'étude du transcriptome à partir de tissu artériel rénal provenant de donneurs ayant subi une néphrectomie. Les particularités de ce tissu en comparaison à celui des artères coronaires (Beaudoin, Gupta et al. 2015) pourraient-être étudiées. De plus, les données de transcriptomes pourraient être associées aux données de génotypages permettant ainsi une exploration systématique de corrélation génotypes/expression à la recherche de nouveaux *eQTLs* pour des gènes d'intérêt de la DFM.

CONCLUSION GENERALE

Mon travail de thèse visait à réaliser une caractérisation des bases génétiques de la dysplasie fibromusculaire (DFM). Il s'agit d'une pathologie vasculaire non athéromateuse qui prédispose à l'hypertension et à l'accident vasculaire cérébral pour laquelle l'hypothèse génétique est évoquée, même si aucun modèle génétique ni aucun facteur de risque génétique n'ont encore été identifiés.

L'étude génétique de la DFM confronte à plusieurs difficultés. Il s'agit notamment de l'absence de familles étendues et de grandes cohortes de cas avec une caractérisation phénotypique précise. Cela s'explique par le diagnostic de la maladie qui nécessite des examens d'imagerie invasifs, mais aussi par la nécessité d'exclure les nombreuses pathologies et lésions associées. Ceci avec pour conséquence de restreindre les possibilités d'études familiales et les cohortes de réplications. De plus, les récentes évolutions dans l'estimation de la prévalence de la DFM, qui apparait désormais totalement sous-évaluée, conduisent à reconsidérer le modèle génétique sous-jacent et adapter l'approche d'étude.

Malgré ces quelques difficultés, j'ai pu mettre en œuvre deux projets de recherche complémentaires incluant des approches de séquençage d'exomes de cas familiaux de la DFM et d'étude d'association de variants rares et fréquents génotypés par la puce exome chip. Ces deux études ont permis conjointement d'exclure l'implication d'un gène majeur pour la DFM, de démontrer les caractères polygénique et multifactoriel de cette maladie et d'identifier le premier variant génétique de prédisposition à la DFM pour lequel plusieurs arguments fonctionnels obtenus dans le cadre de collaborations permettent d'étayer l'implication dans la physiopathologie de la DFM. Ces résultats ont fait l'objet de deux articles scientifiques pour lesquels je suis le premier auteur ; le premier est publié dans la revue *Journal of Hypertension* et le second est en cours de révision dans la revue *Plos Genetics*.

Au "terme" de ces trois années d’études doctorales, nous pouvons dire que les résultats obtenus permettent une avancée certaine dans la compréhension des déterminants génétiques impliqués dans la DFM. Ce travail confirme aussi les caractères multifactoriel et hétérogène de la DFM ouvrant ainsi de nouvelles perspectives à des projets de recherche translationnelle pour la génétique et la génomique fonctionnelle de la DFM.

QUATRIEME PARTIE : BIBLIOGRAPHIE

Adzhubei, I., D. M. Jordan and S. R. Sunyaev (2013). "Predicting functional effect of human missense mutations using PolyPhen-2." Curr Protoc Hum Genet **Chapter 7**: Unit7 20.

Adzhubei, I. A., S. Schmidt, L. Peshkin, V. E. Ramensky, A. Gerasimova, P. Bork, et al. (2010). "A method and server for predicting damaging missense mutations." Nat Methods **7**(4): 248-249.

Albert, F. W. and L. Kruglyak (2015). "The role of regulatory variation in complex traits and disease." Nat Rev Genet **16**(4): 197-212.

Albrechtsen, A., N. Grarup, Y. Li, T. Sparso, G. Tian, H. Cao, et al. (2013). "Exome sequencing-driven discovery of coding polymorphisms associated with common metabolic phenotypes." Diabetologia **56**(2): 298-310.

Alimi, Y., C. Mercier, J. F. Pellissier, P. Piquet and P. Tournigand (1992). "Fibromuscular disease of the renal artery: a new histopathologic classification." Ann Vasc Surg **6**(3): 220-224.

Allain, B., R. Jarray, L. Borriello, B. Leforban, S. Dufour, W. Q. Liu, et al. (2012). "Neuropilin-1 regulates a new VEGF-induced gene, Phactr-1, which controls tubulogenesis and modulates lamellipodial dynamics in human endothelial cells." Cell Signal **24**(1): 214-223.

Allen, P. B., A. T. Greenfield, P. Svenningsson, D. C. Haspeslagh and P. Greengard (2004). "Phactrs 1-4: A family of protein phosphatase 1 and actin regulatory proteins." Proc Natl Acad Sci U S A **101**(18): 7187-7192.

Altshuler, D., M. J. Daly and E. S. Lander (2008). "Genetic mapping in human disease." Science **322**(5903): 881-888.

American Urological Association (2016). "American Urological Association."

Anderson, C. A., F. H. Pettersson, G. M. Clarke, L. R. Cardon, A. P. Morris and K. T. Zondervan (2010). "Data quality control in genetic case-control association studies." Nat Protoc **5**(9): 1564-1573.

Andreoni, K. A., S. M. Weeks, D. A. Gerber, J. H. Fair, M. A. Mauro, L. McCoy, et al. (2002). "Incidence of donor renal fibromuscular dysplasia: does it justify routine angiography?" Transplantation **73**(7): 1112-1116.

Antonarakis, S. E. and J. S. Beckmann (2006). "Mendelian disorders deserve more attention." Nat Rev Genet **7**(4): 277-282.

Anttila, V., B. S. Winsvold, P. Gormley, T. Kurth, F. Bettella, G. McMahon, et al. (2013). "Genome-wide meta-analysis identifies new susceptibility loci for migraine." Nat Genet **45**(8): 912-917.

Ashburner, M., C. A. Ball, J. A. Blake, D. Botstein, H. Butler, J. M. Cherry, et al. (2000). "Gene ontology: tool for the unification of biology. The Gene Ontology Consortium." Nat Genet **25**(1): 25-29.

Austin, E. D., L. Ma, C. LeDuc, E. Berman Rosenzweig, A. Borczuk, J. A. Phillips, 3rd, et al. (2012). "Whole exome sequencing to identify a novel gene (caveolin-1) associated with human pulmonary arterial hypertension." Circ Cardiovasc Genet **5**(3): 336-343.

Baek, J. S., L. Fang, A. C. Li and Y. I. Miller (2012). "Ezetimibe and simvastatin reduce cholesterol levels in zebrafish larvae fed a high-cholesterol diet." Cholesterol **2012**: 564705.

Baird, P. A., T. W. Anderson, H. B. Newcombe and R. B. Lowry (1988). "Genetic disorders in children and young adults: a population study." Am J Hum Genet **42**(5): 677-693.

Bamshad, M. J., S. B. Ng, A. W. Bigham, H. K. Tabor, M. J. Emond, D. A. Nickerson, et al. (2011). "Exome sequencing as a tool for Mendelian disease gene discovery." Nat Rev Genet **12**(11): 745-755.

Barrett, J. C., B. Fry, J. Maller and M. J. Daly (2005). "Haploview: analysis and visualization of LD and haplotype maps." Bioinformatics **21**(2): 263-265.

Bartel, D. P. (2009). "MicroRNAs: target recognition and regulatory functions." Cell **136**(2): 215-233.

Basu, S. and W. Pan (2011). "Comparison of statistical tests for disease association with rare variants." Genet Epidemiol **35**(7): 606-619.

Beaudoin, M., R. M. Gupta, H. H. Won, K. S. Lo, R. Do, C. A. Henderson, et al. (2015). "Myocardial Infarction-Associated SNP at 6p24 Interferes With MEF2 Binding and Associates With PHACTR1 Expression Levels in Human Coronary Arteries." Arterioscler Thromb Vasc Biol **35**(6): 1472-1479.

Beckmann, J. S., X. Estivill and S. E. Antonarakis (2007). "Copy number variants and genetic traits: closer to the resolution of phenotypic to genotypic variability." Nat Rev Genet **8**(8): 639-646.

Bejot, Y., C. Aboa-Eboule, S. Debette, A. Pezzini, T. Tatlisumak, S. Engelter, et al. (2014). "Characteristics and outcomes of patients with multiple cervical artery dissection." Stroke **45**(1): 37-41.

Belkadi, A., A. Bolze, Y. Itan, A. Cobat, Q. B. Vincent, A. Antipenko, et al. (2015). "Whole-genome sequencing is more powerful than whole-exome sequencing for detecting exome variants." Proc Natl Acad Sci U S A **112**(17): 5473-5478.

Bigazzi, R., S. Bianchi, N. Quilici, R. Salvadori and G. Baldari (1998). "Bilateral fibromuscular dysplasia in identical twins." Am J Kidney Dis **32**(6): E4.

Blekhman, R., O. Man, L. Herrmann, A. R. Boyko, A. Indap, C. Kosiol, et al. (2008). "Natural selection on genes that underlie human disease susceptibility." Curr Biol **18**(12): 883-889.

Bodmer, W. and C. Bonilla (2008). "Common and rare variants in multifactorial susceptibility to common diseases." Nat Genet **40**(6): 695-701.

Bofinger, A., C. Hawley, P. Fisher, N. Daunt, M. Stowasser and R. Gordon (1999). "Increased severity of multifocal renal arterial fibromuscular dysplasia in smokers." J Hum Hypertens **13**(8): 517-520.

Bofinger, A., C. Hawley, P. Fisher, N. Daunt, M. Stowasser and R. Gordon (2001). "Polymorphisms of the renin-angiotensin system in patients with multifocal renal arterial fibromuscular dysplasia." J Hum Hypertens **15**(3): 185-190.

Botstein, D. and N. Risch (2003). "Discovering genotypes underlying human phenotypes: past successes for mendelian disease, future approaches for complex disease." Nat Genet **33 Suppl**: 228-237.

Boutouyrie, P., A. P. Gimenez-Roqueplo, E. Fine, B. Laloux, B. Fiquet-Kempf, P. F. Plouin, et al. (2003). "Evidence for carotid and radial artery wall subclinical lesions in renal fibromuscular dysplasia." J Hypertens **21**(12): 2287-2295.

Boycott, K. M., M. R. Vanstone, D. E. Bulman and A. E. MacKenzie (2013). "Rare-disease genetics in the era of next-generation sequencing: discovery to translation." Nat Rev Genet **14**(10): 681-691.

Boyle, A. P., E. L. Hong, M. Hariharan, Y. Cheng, M. A. Schaub, M. Kasowski, et al. (2012). "Annotation of functional variation in personal genomes using RegulomeDB." Genome Res **22**(9): 1790-1797.

Bumgarner, R. (2013). "Overview of DNA microarrays: types, applications, and their future." Curr Protoc Mol Biol **Chapter 22**: Unit 22 21.

Burdett T (EBI), H. P. N., Hastings E (EBI), Hindorff LA (NHGRI), Junkins HA (NHGRI), Klemm AK (NHGRI), MacArthur J (EBI), Manolio TA (NHGRI), Morales J (EBI), Parkinson H (EBI) and Welter D (EBI). The NHGRI-EBI Catalog of published genome-wide association studies.Available at: www.ebi.ac.uk/gwas. Accessed [20/04/2016], version [v1.0].

Bush, W. S. and J. H. Moore (2012). "Chapter 11: Genome-wide association studies." PLoS Comput Biol **8**(12): e1002822.

Camuglia, A., V. Manins, A. Taylor and C. Hengel (2009). "Case report and review: epicardial coronary artery fibromuscular dysplasia." Heart Lung Circ **18**(2): 151-154.

Carter, C. O. (1977). "Monogenic disorders." J Med Genet **14**(5): 316-320.

Charru, A., X. Jeunemaitre, F. Soubrier, P. Corvol and G. Chatellier (1994). "Short report: HYPERGENE: a clinical and genetic database for genetic analysis of human hypertension." J Hypertens **12**(8): 981-985.

Cho, S. W., S. Kim, J. M. Kim and J. S. Kim (2013). "Targeted genome engineering in human cells with the Cas9 RNA-guided endonuclease." Nat Biotechnol **31**(3): 230-232.

Chorley, B. N., X. Wang, M. R. Campbell, G. S. Pittman, M. A. Noureddine and D. A. Bell (2008). "Discovery and verification of functional single nucleotide polymorphisms in regulatory genomic regions: current and developing technologies." Mutat Res **659**(1-2): 147-157.

Chu, J. and K. C. Sadler (2009). "New school in liver development: lessons from zebrafish." Hepatology **50**(5): 1656-1663.

Clark, M. J., R. Chen, H. Y. Lam, K. J. Karczewski, R. Chen, G. Euskirchen, et al. (2011). "Performance comparison of exome DNA sequencing technologies." Nat Biotechnol **29**(10): 908-914.

Clarke, A. J. and D. N. Cooper (2010). "GWAS: heritability missing in action?" Eur J Hum Genet **18**(8): 859-861.

Cloft, H. J., D. F. Kallmes, M. H. Kallmes, J. H. Goldstein, M. E. Jensen and J. E. Dion (1998). "Prevalence of cerebral aneurysms in patients with fibromuscular dysplasia: a reassessment." J Neurosurg **88**(3): 436-440.

Cohen, J. C., R. S. Kiss, A. Pertsemlidis, Y. L. Marcel, R. McPherson and H. H. Hobbs (2004). "Multiple rare alleles contribute to low plasma levels of HDL cholesterol." Science **305**(5685): 869-872.

Cong, L., F. A. Ran, D. Cox, S. Lin, R. Barretto, N. Habib, et al. (2013). "Multiplex genome engineering using CRISPR/Cas systems." Science **339**(6121): 819-823.

Connett, M. C. and J. M. Lansche (1965). "Fibromuscular Hyperplasia of the Internal Carotid Artery: Report of a Case." Ann Surg **162**: 59-62.

Consortium, G. T. (2013). "The Genotype-Tissue Expression (GTEx) project." Nat Genet **45**(6): 580-585.

Consortium, G. T. (2015). "Human genomics. The Genotype-Tissue Expression (GTEx) pilot analysis: multitissue gene regulation in humans." Science **348**(6235): 648-660.

Cooper, D. N., M. Krawczak, C. Polychronakos, C. Tyler-Smith and H. Kehrer-Sawatzki (2013). "Where genotype is not predictive of phenotype: towards an understanding of the molecular basis of reduced penetrance in human inherited disease." Hum Genet **132**(10): 1077-1130.

Cooper, G. M. and J. Shendure (2011). "Needles in stacks of needles: finding disease-causal variants in a wealth of genomic data." Nat Rev Genet **12**(9): 628-640.

Coronary Artery Disease Genetics, C. (2011). "A genome-wide association study in Europeans and South Asians identifies five new loci for coronary artery disease." Nat Genet **43**(4): 339-344.

Cragg, A. H., T. P. Smith, B. H. Thompson, T. P. Maroney, A. W. Stanson, G. T. Shaw, et al. (1989). "Incidental fibromuscular dysplasia in potential renal donors: long-term clinical follow-up." Radiology **172**(1): 145-147.

Croft, D., A. F. Mundo, R. Haw, M. Milacic, J. Weiser, G. Wu, et al. (2014). "The Reactome pathway knowledgebase." Nucleic Acids Res **42**(Database issue): D472-477.

Cruchaga, C., C. M. Karch, S. C. Jin, B. A. Benitez, Y. Cai, R. Guerreiro, et al. (2014). "Rare coding variants in the phospholipase D3 gene confer risk for Alzheimer's disease." Nature **505**(7484): 550-554.

Curtis, C., S. P. Shah, S. F. Chin, G. Turashvili, O. M. Rueda, M. J. Dunning, et al. (2012). "The genomic and transcriptomic architecture of 2,000 breast tumours reveals novel subgroups." Nature **486**(7403): 346-352.

d'Anglejan Chatillon, J., V. Ribeiro, J. L. Mas, M. G. Bousser and D. Laplane (1990). "[Dissection of the extracranial internal carotid artery. 62 cases]." Presse Med **19**(14): 661-667.

Daly, M. J., J. D. Rioux, S. F. Schaffner, T. J. Hudson and E. S. Lander (2001). "High-resolution haplotype structure in the human genome." Nat Genet **29**(2): 229-232.

Dawn Teare, M. and J. H. Barrett (2005). "Genetic linkage studies." Lancet **366**(9490): 1036-1044.

Debette, S., B. Goeggel Simonetti, S. Schilling, J. J. Martin, M. Kloss, H. Sarikaya, et al. (2014). "Familial occurrence and heritable connective tissue disorders in cervical artery dissection." Neurology **83**(22): 2023-2031.

Debette, S., Y. Kamatani, T. M. Metso, M. Kloss, G. Chauhan, S. T. Engelter, et al. (2015). "Common variation in PHACTR1 is associated with susceptibility to cervical artery dissection." Nat Genet **47**(1): 78-83.

Debette, S. and D. Leys (2009). "Cervical-artery dissections: predisposing factors, diagnosis, and outcome." Lancet Neurol **8**(7): 668-678.

DeJesus-Hernandez, M., I. R. Mackenzie, B. F. Boeve, A. L. Boxer, M. Baker, N. J. Rutherford, et al. (2011). "Expanded GGGGCC hexanucleotide repeat in noncoding region of C9ORF72 causes chromosome 9p-linked FTD and ALS." Neuron **72**(2): 245-256.

Deo, R. and C. M. Albert (2012). "Epidemiology and genetics of sudden cardiac death." Circulation **125**(4): 620-637.

Depienne, C., D. Bouteiller, A. Meneret, S. Billot, S. Groppa, S. Klebe, et al. (2012). "RAD51 haploinsufficiency causes congenital mirror movements in humans." Am J Hum Genet **90**(2): 301-307.

Dimas, A. S., S. Deutsch, B. E. Stranger, S. B. Montgomery, C. Borel, H. Attar-Cohen, et al. (2009). "Common regulatory variation impacts gene expression in a cell type-dependent manner." Science **325**(5945): 1246-1250.

Dina, C., N. Bouatia-Naji, N. Tucker, F. N. Delling, K. Toomer, R. Durst, et al. (2015). "Genetic association analyses highlight biological pathways underlying mitral valve prolapse." Nat Genet **47**(10): 1206-1211.

Duan, J., J. G. Zhang, H. W. Deng and Y. P. Wang (2013). "Comparative studies of copy number variation detection methods for next-generation sequencing technologies." PLoS One **8**(3): e59128.

Dunham, I., A. Kundaje, S. F. Aldred, P. J. Collins, C. A. Davis, F. Doyle, et al. (2012). "An integrated encyclopedia of DNA elements in the human genome." Nature **489**(7414): 57-74.

Dye, D. E., K. J. Brameld, S. Maxwell, J. Goldblatt, C. Bower, H. Leonard, et al. (2011). "The impact of single gene and chromosomal disorders on hospital admissions of children and adolescents: a population-based study." Public Health Genomics **14**(3): 153-161.

Edwards, B. S., A. W. Stanson, K. E. Holley and S. G. Sheps (1982). "Isolated renal artery dissection, presentation, evaluation, management, and pathology." Mayo Clin Proc **57**(9): 564-571.

Edwards, S. L., J. Beesley, J. D. French and A. M. Dunning (2013). "Beyond GWASs: illuminating the dark road from association to function." Am J Hum Genet **93**(5): 779-797.

Ehrenfeld, W. K., R. J. Stoney and E. J. Wylie (1967). "Fibromuscular hyperplasia of the internal carotid artery." Arch Surg **95**(2): 284-287.

Ellegren, H. (2004). "Microsatellites: simple sequences with complex evolution." Nat Rev Genet **5**(6): 435-445.

Elston, R. C. and J. Stewart (1971). "A general model for the genetic analysis of pedigree data." Hum Hered **21**(6): 523-542.

Ensembl (2016). "http://grch37.ensembl.org/Homo_sapiens/Location/Genome." 04/04/2016.

Ermann, J. and L. H. Glimcher (2012). "After GWAS: mice to the rescue?" Curr Opin Immunol **24**(5): 564-570.

Evangelou, E. and J. P. Ioannidis (2013). "Meta-analysis methods for genome-wide association studies and beyond." Nat Rev Genet **14**(6): 379-389.

Exome Aggregation Consortium (2016). "Exome Aggregation Consortium (ExAC), Cambridge, MA (URL: http://exac.broadinstitute.org) [May, 2016]."

Exome Chip Design (2014). "Exome Chip Design [cited September 10, 2014]. http://genome.sph.umich.edu/wiki/Exome_Chip_Design.".

Feuk, L., A. R. Carson and S. W. Scherer (2006). "Structural variation in the human genome." Nat Rev Genet **7**(2): 85-97.

Fils-Aime, N., M. Dai, J. Guo, M. El-Mousawi, B. Kahramangil, J. C. Neel, et al. (2013). "MicroRNA-584 and the protein phosphatase and actin regulator 1 (PHACTR1), a new signaling route through which transforming growth factor-beta Mediates the migration and actin dynamics of breast cancer cells." J Biol Chem **288**(17): 11807-11823.

Fishman, G. I., S. S. Chugh, J. P. Dimarco, C. M. Albert, M. E. Anderson, R. O. Bonow, et al. (2010). "Sudden cardiac death prediction and prevention: report from a National Heart, Lung, and Blood Institute and Heart Rhythm Society Workshop." Circulation **122**(22): 2335-2348.

Flandez, M., J. Cendrowski, M. Canamero, A. Salas, N. del Pozo, K. Schoonjans, et al. (2014). "Nr5a2 heterozygosity sensitises to, and cooperates with, inflammation in KRas(G12V)-driven pancreatic tumourigenesis." Gut **63**(4): 647-655.

Flicek, P., M. R. Amode, D. Barrell, K. Beal, K. Billis, S. Brent, et al. (2014). "Ensembl 2014." Nucleic Acids Res **42**(Database issue): D749-755.

Flicek, P. and E. Birney (2009). "Sense from sequence reads: methods for alignment and assembly." Nat Methods **6**(11 Suppl): S6-S12.

Fontana, P., P. Gaussem, M. Aiach, J. N. Fiessinger, J. Emmerich and J. L. Reny (2003). "P2Y12 H2 haplotype is associated with peripheral arterial disease: a case-control study." Circulation **108**(24): 2971-2973.

Frazer, K. A., S. S. Murray, N. J. Schork and E. J. Topol (2009). "Human genetic variation and its contribution to complex traits." Nat Rev Genet **10**(4): 241-251.

Freilinger, T., V. Anttila, B. de Vries, R. Malik, M. Kallela, G. M. Terwindt, et al. (2012). "Genome-wide association analysis identifies susceptibility loci for migraine without aura." Nat Genet **44**(7): 777-782.

Frodsham, A. J. and A. V. Hill (2004). "Genetics of infectious diseases." Hum Mol Genet **13 Spec No 2**: R187-194.

Fromer, M., A. J. Pocklington, D. H. Kavanagh, H. J. Williams, S. Dwyer, P. Gormley, et al. (2014). "De novo mutations in schizophrenia implicate synaptic networks." Nature **506**(7487): 179-184.

Gazal, S., M. Sahbatou, M. C. Babron, E. Genin and A. L. Leutenegger (2015). "High level of inbreeding in final phase of 1000 Genomes Project." Sci Rep **5**: 17453.

GENIN, E. D., C.; LUDWIG,T.; QUENEZ, O.; LETORT, S.; LINDENBAUM, P.; GIEMZA,J.; GRENIER-BOLEY,B.; CHARBONNIER,C.; BACQ,D.; LECHNER,D.; BOLAND,A.; MEYER,V.; ROUAULT,K.; BELLENGUEZ,C.; FEREC,C.;LE MAREC,HY.; LETENNEUR,L.; DARTIGUES,J.C.; LAMBERT,J.C.; CAMPION,D.; REDON,R.; DELEUZE,J.F. (2016). "Le projet FREX : une base de données d'exomes des régions françaises." Abstract in 8èmes Assises de Génétique Hmaine et Médicale, Lyon, Février 2016.

Genomes Project, C., G. R. Abecasis, D. Altshuler, A. Auton, L. D. Brooks, R. M. Durbin, et al. (2010). "A map of human genome variation from population-scale sequencing." Nature **467**(7319): 1061-1073.

Genomes Project, C., G. R. Abecasis, A. Auton, L. D. Brooks, M. A. DePristo, R. M. Durbin, et al. (2012). "An integrated map of genetic variation from 1,092 human genomes." Nature **491**(7422): 56-65.

Genomes Project, C., A. Auton, L. D. Brooks, R. M. Durbin, E. P. Garrison, H. M. Kang, et al. (2015). "A global reference for human genetic variation." Nature **526**(7571): 68-74.

Germain, D. P. and Y. Herrera-Guzman (2004). "Vascular Ehlers-Danlos syndrome." Ann Genet **47**(1): 1-9.

Germain, M., M. Eyries, D. Montani, O. Poirier, B. Girerd, P. Dorfmuller, et al. (2013). "Genome-wide association analysis identifies a susceptibility locus for pulmonary arterial hypertension." Nat Genet **45**(5): 518-521.

Gilissen, C., A. Hoischen, H. G. Brunner and J. A. Veltman (2011). "Unlocking Mendelian disease using exome sequencing." Genome Biol **12**(9): 228.

Gladstien, K., A. R. Rushton and K. K. Kidd (1980). "Penetrance estimates and recurrence risks for fibromuscular dysplasia." Clin Genet **17**(2): 115-116.

Gonzalez, E., H. Kulkarni, H. Bolivar, A. Mangano, R. Sanchez, G. Catano, et al. (2005). "The influence of CCL3L1 gene-containing segmental duplications on HIV-1/AIDS susceptibility." Science **307**(5714): 1434-1440.

Gorlov, I. P., O. Y. Gorlova, M. L. Frazier, M. R. Spitz and C. I. Amos (2011). "Evolutionary evidence of the effect of rare variants on disease etiology." Clin Genet **79**(3): 199-206.

Gowda, M. S., A. L. Loeb, L. J. Crouse and P. H. Kramer (2003). "Complementary roles of color-flow duplex imaging and intravascular ultrasound in the diagnosis of renal artery fibromuscular dysplasia: should renal arteriography serve as the "gold standard"?" J Am Coll Cardiol **41**(8): 1305-1311.

Gutmann, D. H., A. Aylsworth, J. C. Carey, B. Korf, J. Marks, R. E. Pyeritz, et al. (1997). "The diagnostic evaluation and multidisciplinary management of neurofibromatosis 1 and neurofibromatosis 2." JAMA **278**(1): 51-57.

Halpern, M. M., H. S. Sanford and M. Viamonte, Jr. (1965). "Renal-artery abnormalities in three hypertensive sisters. Probable familial fibromuscular hyperplasia." JAMA **194**(5): 512-513.

Hansen, J., C. Holten and J. V. Thorborg (1965). "Hypertension in two sisters caused by so-called fibromuscular hyperplasia of the renal arteries." Acta Med Scand **178**(4): 461-474.

Haraksingh, R. R. and M. P. Snyder (2013). "Impacts of variation in the human genome on gene regulation." J Mol Biol **425**(21): 3970-3977.

Harrison, E. G., Jr. and L. J. McCormack (1971). "Pathologic classification of renal arterial disease in renovascular hypertension." Mayo Clin Proc **46**(3): 161-167.

Hata, A., M. Noda, R. Moriwaki and F. Numano (1996). "Angiographic findings of Takayasu arteritis: new classification." Int J Cardiol **54 Suppl**: S155-163.

Hatem, A., D. Bozdag, A. E. Toland and U. V. Catalyurek (2013). "Benchmarking short sequence mapping tools." BMC Bioinformatics **14**: 184.

Hendricks, N. J., A. H. Matsumoto, J. F. Angle, A. Baheti, S. S. Sabri, A. W. Park, et al. (2014). "Is fibromuscular dysplasia underdiagnosed? A comparison of the prevalence of FMD seen in CORAL trial participants versus a single institution population of renal donor candidates." Vasc Med.

Hicks, S., D. A. Wheeler, S. E. Plon and M. Kimmel (2011). "Prediction of missense mutation functionality depends on both the algorithm and sequence alignment employed." Hum Mutat **32**(6): 661-668.

Hirschhorn, J. N. and M. J. Daly (2005). "Genome-wide association studies for common diseases and complex traits." Nat Rev Genet **6**(2): 95-108.

Holm, H., D. F. Gudbjartsson, P. Sulem, G. Masson, H. T. Helgadottir, C. Zanon, et al. (2011). "A rare variant in MYH6 is associated with high risk of sick sinus syndrome." Nat Genet **43**(4): 316-320.

Houser, O. W., H. L. Baker, Jr., B. A. Sandok and K. E. Holley (1971). "Cephalic arterial fibromuscular dysplasia." Radiology **101**(3): 605-611.

Howe, K., M. D. Clark, C. F. Torroja, J. Torrance, C. Berthelot, M. Muffato, et al. (2013). "The zebrafish reference genome sequence and its relationship to the human genome." Nature **496**(7446): 498-503.

Huang da, W., B. T. Sherman and R. A. Lempicki (2009). "Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources." Nat Protoc **4**(1): 44-57.

Hunt, J. C. (1961). "Symposium on hypertension associated with renal artery disease. Clinical aspects." Proc Staff Meet Mayo Clin **36**: 707-712.

Huyghe, J. R., A. U. Jackson, M. P. Fogarty, M. L. Buchkovich, A. Stancakova, H. M. Stringham, et al. (2013). "Exome array analysis identifies new loci and low-frequency variants influencing insulin processing and secretion." Nat Genet **45**(2): 197-201.

Illumina. "Illumina HiSeq-X." 22/04/2016, from http://www.illumina.com/systems/hiseq-x-sequencing-system.html.

International HapMap, C. (2003). "The International HapMap Project." Nature **426**(6968): 789-796.

International HapMap, C., D. M. Altshuler, R. A. Gibbs, L. Peltonen, D. M. Altshuler, R. A. Gibbs, et al. (2010). "Integrating common and rare genetic variation in diverse human populations." Nature **467**(7311): 52-58.

International Human Genome Sequencing, C. (2004). "Finishing the euchromatic sequence of the human genome." Nature **431**(7011): 931-945.

Ionita-Laza, I., J. D. Buxbaum, N. M. Laird and C. Lange (2011). "A new testing strategy to identify rare variants with either risk or protective effect on disease." PLoS Genet **7**(2): e1001289.

Ionita-Laza, I., S. Lee, V. Makarov, J. D. Buxbaum and X. Lin (2013). "Sequence kernel association tests for the combined effect of rare and common variants." Am J Hum Genet **92**(6): 841-853.

Iyengar, S. K. and R. C. Elston (2007). "The genetic basis of complex traits: rare variants or "common gene, common disease"?" Methods Mol Biol **376**: 71-84.

Jackson, V. E., I. Ntalla, I. Sayers, R. Morris, P. Whincup, J. P. Casas, et al. (2016). "Exome-wide analysis of rare coding variation identifies novel associations with COPD and airflow limitation in MOCS3, IFIT3 and SERPINA12." Thorax **71**(6): 501-509.

Jarray, R., B. Allain, L. Borriello, D. Biard, A. Loukaci, J. Larghero, et al. (2011). "Depletion of the novel protein PHACTR-1 from human endothelial cells abolishes tube formation and induces cell death receptor apoptosis." Biochimie **93**(10): 1668-1675.

Jeffreys, A. J., L. Kauppi and R. Neumann (2001). "Intensely punctate meiotic recombination in the class II region of the major histocompatibility complex." Nat Genet **29**(2): 217-222.

Jeunemaitre, X., I. Inoue, C. Williams, A. Charru, J. Tichet, M. Powers, et al. (1997). "Haplotypes of angiotensinogen in essential hypertension." Am J Hum Genet **60**(6): 1448-1460.

Jeunemaitre, X., F. Soubrier, Y. V. Kotelevtsev, R. P. Lifton, C. S. Williams, A. Charru, et al. (1992). "Molecular basis of human hypertension: role of angiotensinogen." Cell **71**(1): 169-180.

Johansen, C. T., J. Wang, A. D. McIntyre, R. A. Martins, M. R. Ban, M. B. Lanktree, et al. (2012). "Excess of rare variants in non-genome-wide association study candidate genes in patients with hypertriglyceridemia." Circ Cardiovasc Genet **5**(1): 66-72.

Johnson, G. C., L. Esposito, B. J. Barratt, A. N. Smith, J. Heward, G. Di Genova, et al. (2001). "Haplotype tagging for the identification of common disease genes." Nat Genet **29**(2): 233-237.

Johnston, J. J., J. K. Teer, P. F. Cherukuri, N. F. Hansen, S. K. Loftus, N. I. H. I. S. Center, et al. (2010). "Massively parallel sequencing of exons on the X chromosome identifies RBM10 as the gene that causes a syndromic form of cleft palate." Am J Hum Genet **86**(5): 743-748.

Jonsson, T., J. K. Atwal, S. Steinberg, J. Snaedal, P. V. Jonsson, S. Bjornsson, et al. (2012). "A mutation in APP protects against Alzheimer's disease and age-related cognitive decline." Nature **488**(7409): 96-99.

Jorgensen, T. J., I. Ruczinski, B. Kessing, M. W. Smith, Y. Y. Shugart and A. J. Alberg (2009). "Hypothesis-driven candidate gene association studies: practical design and analytical considerations." Am J Epidemiol **170**(8): 986-993.

Kanehisa, M. and S. Goto (2000). "KEGG: kyoto encyclopedia of genes and genomes." Nucleic Acids Res **28**(1): 27-30.

Kang, B. W., H. S. Jeon, Y. S. Chae, S. J. Lee, J. Y. Park, J. E. Choi, et al. (2015). "Association between GWAS-identified genetic variations and disease prognosis for patients with colorectal cancer." PLoS One **10**(3): e0119649.

Kiando, S. R., C. Barlassina, D. Cusi, P. Galan, M. Lathrop, P. F. Plouin, et al. (2015). "Exome sequencing in seven families and gene-based association studies indicate genetic heterogeneity and suggest possible candidates for fibromuscular dysplasia." J Hypertens **33**(9): 1802-1810.

Kidd, J. M., G. M. Cooper, W. F. Donahue, H. S. Hayden, N. Sampas, T. Graves, et al. (2008). "Mapping and sequencing of structural variation from eight human genomes." Nature **453**(7191): 56-64.

Kiezun, A., K. Garimella, R. Do, N. O. Stitziel, B. M. Neale, P. J. McLaren, et al. (2012). "Exome sequencing and the genetic basis of complex traits." Nat Genet **44**(6): 623-630.

Kim, Y. J., J. Lee, B. J. Kim, T. D.-G. Consortium and T. Park (2015). "A new strategy for enhancing imputation quality of rare variants from next-generation sequencing data via combining SNP and exome chip data." BMC Genomics **16**: 1109.

Kincaid, O. W. and G. D. Davis (1961). "Renal arteriography in hypertension." Proc Staff Meet Mayo Clin **36**: 689-701.

Kincaid, O. W., G. D. Davis, F. J. Hallermann and J. C. Hunt (1968). "Fibromuscular dysplasia of the renal arteries. Arteriographic features, classification, and observations on natural history of the disease." Am J Roentgenol Radium Ther Nucl Med **104**(2): 271-282.

Klein, R. J., C. Zeiss, E. Y. Chew, J. Y. Tsai, R. S. Sackler, C. Haynes, et al. (2005). "Complement factor H polymorphism in age-related macular degeneration." Science **308**(5720): 385-389.

Kleinjan, D. A. and V. van Heyningen (2005). "Long-range control of gene expression: emerging mechanisms and disruption in disease." Am J Hum Genet **76**(1): 8-32.

Kozlitina, J., E. Smagris, S. Stender, B. G. Nordestgaard, H. H. Zhou, A. Tybjaerg-Hansen, et al. (2014). "Exome-wide association study identifies a TM6SF2 variant that confers susceptibility to nonalcoholic fatty liver disease." Nat Genet **46**(4): 352-356.

Kozomara, A. and S. Griffiths-Jones (2014). "miRBase: annotating high confidence microRNAs using deep sequencing data." Nucleic Acids Res **42**(Database issue): D68-73.

Krek, A., D. Grun, M. N. Poy, R. Wolf, L. Rosenberg, E. J. Epstein, et al. (2005). "Combinatorial microRNA target predictions." Nat Genet **37**(5): 495-500.

Ku, C. S., N. Naidoo and Y. Pawitan (2011). "Revisiting Mendelian disorders through exome sequencing." Hum Genet **129**(4): 351-370.

Kuck, C. E., S. T. Heidt and E. Kline-Rogers (2016). ""Rare" diseases: Motivated patients make the difference." Int J Cardiol **208**: 95-96.

Kumar, P., S. Henikoff and P. C. Ng (2009). "Predicting the effects of coding non-synonymous variants on protein function using the SIFT algorithm." Nat Protoc **4**(7): 1073-1081.

Kusenda, B., M. Mraz, J. Mayer and S. Pospisilova (2006). "MicroRNA biogenesis, functionality and cancer relevance." Biomed Pap Med Fac Univ Palacky Olomouc Czech Repub **150**(2): 205-215.

La Batide Alanore, A., J. Perdu and P. F. Plouin (2007). "[Fibromuscular dysplasia]." Presse Med **36**(6 Pt 2): 1016-1023.

Lacolley, P., P. Boutouyrie, M. Glukhova, J. M. Daniel Lamaziere, P. F. Plouin, P. Bruneval, et al. (2002). "Disruption of the elastin gene in adult Williams syndrome is accompanied by a paradoxical reduction in arterial stiffness." Clin Sci (Lond) **103**(1): 21-29.

Lacombe, M. (2001). "Isolated spontaneous dissection of the renal artery." J Vasc Surg **33**(2): 385-391.

Lander, E. and L. Kruglyak (1995). "Genetic dissection of complex traits: guidelines for interpreting and reporting linkage results." Nat Genet **11**(3): 241-247.

Lander, E. S., L. M. Linton, B. Birren, C. Nusbaum, M. C. Zody, J. Baldwin, et al. (2001). "Initial sequencing and analysis of the human genome." Nature **409**(6822): 860-921.

Lander, E. S. and N. J. Schork (1994). "Genetic dissection of complex traits." Science **265**(5181): 2037-2048.

Landrum, M. J., J. M. Lee, G. R. Riley, W. Jang, W. S. Rubinstein, D. M. Church, et al. (2014). "ClinVar: public archive of relationships among sequence variation and human phenotype." Nucleic Acids Res **42**(Database issue): D980-985.

Lassiter, F. D. (1998). "The string-of-beads sign." Radiology **206**(2): 437-438.

Leadbetter, W. and L. Burkland (1938). "Hypertension in unilateral renal disease." J Urol **39**: 611–626.

Lee, S., M. J. Emond, M. J. Bamshad, K. C. Barnes, M. J. Rieder, D. A. Nickerson, et al. (2012). "Optimal unified approach for rare-variant association testing with application to small-sample case-control whole-exome sequencing studies." Am J Hum Genet **91**(2): 224-237.

Lee, S., T. M. Teslovich, M. Boehnke and X. Lin (2013). "General framework for meta-analysis of rare variants in sequencing association studies." Am J Hum Genet **93**(1): 42-53.

Lee, S., M. C. Wu and X. Lin (2012). "Optimal tests for rare variant effects in sequencing association studies." Biostatistics **13**(4): 762-775.

Lemon, B. and R. Tjian (2000). "Orchestrated response: a symphony of transcription factors for gene control." Genes Dev **14**(20): 2551-2569.

Levy, B. I. (1998). "The potential role of angiotensin II in the vasculature." J Hum Hypertens **12**(5): 283-287.

Lewontin, R. C. (1964). "The Interaction of Selection and Linkage. I. General Considerations; Heterotic Models." Genetics **49**(1): 49-67.

Leys, D., L. Bandu, H. Henon, C. Lucas, F. Mounier-Vehier, P. Rondepierre, et al. (2002). "Clinical outcome in 287 consecutive young adults (15 to 45 years) with ischemic stroke." Neurology **59**(1): 26-33.

Li, B. and S. M. Leal (2008). "Methods for detecting associations with rare variants for common diseases: application to analysis of sequence data." Am J Hum Genet **83**(3): 311-321.

Li, H., B. Handsaker, A. Wysoker, T. Fennell, J. Ruan, N. Homer, et al. (2009). "The Sequence Alignment/Map format and SAMtools." Bioinformatics **25**(16): 2078-2079.

Li, H. and N. Homer (2010). "A survey of sequence alignment algorithms for next-generation sequencing." Brief Bioinform **11**(5): 473-483.

Li, M., C. Li and W. Guan (2008). "Evaluation of coverage variation of SNP chips for genome-wide association studies." Eur J Hum Genet **16**(5): 635-643.

Li, Q., J. H. Seo, B. Stranger, A. McKenna, I. Pe'er, T. Laframboise, et al. (2013). "Integrative eQTL-based analyses reveal the biology of breast cancer risk loci." Cell **152**(3): 633-641.

Lin, W. Y., B. Zhang, N. Yi, G. Gao and N. Liu (2011). "Evaluation of pooled association tests for rare variant identification." BMC Proc **5 Suppl 9**: S118.

Lin, X., S. Lee, M. C. Wu, C. Wang, H. Chen, Z. Li, et al. (2016). "Test for rare variants by environment interactions in sequencing association studies." Biometrics **72**(1): 156-164.

Liu, D. J. and S. M. Leal (2010). "Replication strategies for rare variant complex trait association studies via next-generation sequencing." Am J Hum Genet **87**(6): 790-801.

Liu, L., Y. Li, S. Li, N. Hu, Y. He, R. Pong, et al. (2012). "Comparison of next-generation sequencing systems." J Biomed Biotechnol **2012**: 251364.

Liu, Y., M. Morley, J. Brandimarto, S. Hannenhalli, Y. Hu, E. A. Ashley, et al. (2015). "RNA-Seq identifies novel myocardial gene expression signatures of heart failure." Genomics **105**(2): 83-89.

Lopes, M. C., C. Joyce, G. R. Ritchie, S. L. John, F. Cunningham, J. Asimit, et al. (2012). "A combined functional annotation score for non-synonymous variants." Hum Hered **73**(1): 47-51.

Louis-Dit-Picard, H., J. Barc, D. Trujillano, S. Miserey-Lenkei, N. Bouatia-Naji, O. Pylypenko, et al. (2012). "KLHL3 mutations cause familial hyperkalemic hypertension by impairing ion transport in the distal nephron." Nat Genet **44**(4): 456-460, S451-453.

Lupski, J. R., R. M. de Oca-Luna, S. Slaugenhaupt, L. Pentao, V. Guzzetta, B. J. Trask, et al. (1991). "DNA duplication associated with Charcot-Marie-Tooth disease type 1A." Cell **66**(2): 219-232.

Luscher, T. F., J. T. Lie, A. W. Stanson, O. W. Houser, L. H. Hollier and S. G. Sheps (1987). "Arterial fibromuscular dysplasia." Mayo Clin Proc **62**(10): 931-952.

Madsen, B. E. and S. R. Browning (2009). "A groupwise association test for rare mutations using a weighted sum statistic." PLoS Genet **5**(2): e1000384.

Majewski, J., J. Schwartzentruber, E. Lalonde, A. Montpetit and N. Jabado (2011). "What can exome sequencing do for you?" J Med Genet **48**(9): 580-589.

Major, P., J. Genest, P. Cartier and O. Kuchel (1977). "Hereditary fibromuscular dysplasia with renovascular hypertension." Ann Intern Med **86**(5): 583.

Manolio, T. A. (2010). "Genomewide association studies and assessment of the risk of disease." N Engl J Med **363**(2): 166-176.

Manolio, T. A., F. S. Collins, N. J. Cox, D. B. Goldstein, L. A. Hindorff, D. J. Hunter, et al. (2009). "Finding the missing heritability of complex diseases." Nature **461**(7265): 747-753.

Marchini, J. and B. Howie (2010). "Genotype imputation for genome-wide association studies." Nat Rev Genet **11**(7): 499-511.

Marks, S. D., A. M. Gullett, E. Brennan, K. Tullus, G. Jaureguiberry, E. Klootwijk, et al. (2011). "Renal FMD may not confer a familial hypertensive risk nor is it caused by ACTA2 mutations." Pediatr Nephrol **26**(10): 1857-1861.

Martin, A. R., G. Tse, C. D. Bustamante and E. E. Kenny (2014). "Imputation-based assessment of next generation rare exome variant arrays." Pac Symp Biocomput: 241-252.

Matsumine, H., M. Saito, S. Shimoda-Matsubayashi, H. Tanaka, A. Ishikawa, Y. Nakagawa-Hattori, et al. (1997). "Localization of a gene for an autosomal recessive form of juvenile Parkinsonism to chromosome 6q25.2-27." Am J Hum Genet **60**(3): 588-596.

Maurano, M. T., R. Humbert, E. Rynes, R. E. Thurman, E. Haugen, H. Wang, et al. (2012). "Systematic localization of common disease-associated variation in regulatory DNA." Science **337**(6099): 1190-1195.

McCandless, S. E., J. W. Brunger and S. B. Cassidy (2004). "The burden of genetic disease on inpatient care in a children's hospital." Am J Hum Genet **74**(1): 121-127.

McCarthy, M. I., G. R. Abecasis, L. R. Cardon, D. B. Goldstein, J. Little, J. P. Ioannidis, et al. (2008). "Genome-wide association studies for complex traits: consensus, uncertainty and challenges." Nat Rev Genet **9**(5): 356-369.

McCormack, L., J. Hazard and E. Poutasse (1958). "Obstructive lesions of the renal artery associated with remediable hypertension." Am J Pathol **34**: 582.

McCormack, L. J., E. F. Poutasse, T. F. Meaney, T. J. Noto, Jr. and H. P. Dustan (1966). "A pathologic-arteriographic correlation of renal arterial disease." Am Heart J **72**(2): 188-198.

McKenna, A., M. Hanna, E. Banks, A. Sivachenko, K. Cibulskis, A. Kernytsky, et al. (2010). "The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data." Genome Res **20**(9): 1297-1303.

McKenzie, G. A., G. S. Oderich, A. Kawashima and S. Misra (2013). "Renal artery fibromuscular dysplasia in 2,640 renal donor subjects: a CT angiography analysis." J Vasc Interv Radiol **24**(10): 1477-1480.

McKusick, V. A. (2007). "Mendelian Inheritance in Man and its online version, OMIM." Am J Hum Genet **80**(4): 588-604.

McPherson, R. and A. Tybjaerg-Hansen (2016). "Genetics of Coronary Artery Disease." Circ Res **118**(4): 564-578.

Metso, T. M., T. Tatlisumak, S. Debette, J. Dallongeville, S. T. Engelter, P. A. Lyrer, et al. (2012). "Migraine in cervical artery dissection and ischemic stroke patients." Neurology **78**(16): 1221-1228.

Mettinger, K. L. and K. Ericson (1982). "Fibromuscular dysplasia and the brain. I. Observations on angiographic, clinical and genetic characteristics." Stroke **13**(1): 46-52.

Mettinger, K. L. and C. E. Soderstrom (1978). "Pathogenetic profile of TIA before 55. A three-year investigation." J Neurol Sci **36**(3): 341-348.

Metzker, M. L. (2010). "Sequencing technologies - the next generation." Nat Rev Genet **11**(1): 31-46.

Michelis, K. C., J. W. Olin, D. Kadian-Dodov, V. d'Escamard and J. C. Kovacic (2014). "Coronary artery manifestations of fibromuscular dysplasia." J Am Coll Cardiol **64**(10): 1033-1046.

Miki, Y., J. Swensen, D. Shattuck-Eidens, P. A. Futreal, K. Harshman, S. Tavtigian, et al. (1994). "A strong candidate for the breast and ovarian cancer susceptibility gene BRCA1." Science **266**(5182): 66-71.

Miller, J. C., S. Tan, G. Qiao, K. A. Barlow, J. Wang, D. F. Xia, et al. (2011). "A TALE nuclease architecture for efficient genome editing." Nat Biotechnol **29**(2): 143-148.

Mokri, B., T. M. Sundt, Jr., O. W. Houser and D. G. Piepgras (1986). "Spontaneous dissection of the cervical internal carotid artery." Ann Neurol **19**(2): 126-138.

Momozawa, Y., M. Mni, K. Nakamura, W. Coppieters, S. Almer, L. Amininejad, et al. (2011). "Resequencing of positional candidates identifies low frequency IL23R coding variants protecting against inflammatory bowel disease." Nat Genet **43**(1): 43-47.

Morley, M., C. M. Molony, T. M. Weber, J. L. Devlin, K. G. Ewens, R. S. Spielman, et al. (2004). "Genetic analysis of genome-wide variation in human gene expression." Nature **430**(7001): 743-747.

Morris, A. P. and E. Zeggini (2010). "An evaluation of statistical approaches to rare variant analysis in genetic association studies." Genet Epidemiol **34**(2): 188-193.

Morton, N. E. (1955). "Sequential tests for the detection of linkage." Am J Hum Genet **7**(3): 277-318.

Morton, N. E., C. J. MacLean, R. Lew and S. Yee (1986). "Multipoint linkage analysis." Am J Hum Genet **38**(6): 868-883.

Mouse Genome Sequencing, C., R. H. Waterston, K. Lindblad-Toh, E. Birney, J. Rogers, J. F. Abril, et al. (2002). "Initial sequencing and comparative analysis of the mouse genome." Nature **420**(6915): 520-562.

Myers, S., L. Bottolo, C. Freeman, G. McVean and P. Donnelly (2005). "A fine-scale map of recombination rates and hotspots across the human genome." Science **310**(5746): 321-324.

Nasevicius, A. and S. C. Ekker (2000). "Effective targeted gene 'knockdown' in zebrafish." Nat Genet **26**(2): 216-220.

Neale, B. M., M. A. Rivas, B. F. Voight, D. Altshuler, B. Devlin, M. Orho-Melander, et al. (2011). "Testing for an unusual distribution of rare variants." PLoS Genet **7**(3): e1001322.

Nelson, M. R., D. Wegmann, M. G. Ehm, D. Kessner, P. St Jean, C. Verzilli, et al. (2012). "An abundance of rare functional variants in 202 drug target genes sequenced in 14,002 people." Science **337**(6090): 100-104.

Neymark, E., J. M. LaBerge, R. Hirose, J. S. Melzer, R. K. Kerlan, Jr., M. W. Wilson, et al. (2000). "Arteriographic detection of renovascular disease in potential renal donors: incidence and effect on donor surgery." Radiology **214**(3): 755-760.

Ng, P. C. and S. Henikoff (2003). "SIFT: Predicting amino acid changes that affect protein function." Nucleic Acids Res **31**(13): 3812-3814.

Ng, P. C., S. Levy, J. Huang, T. B. Stockwell, B. P. Walenz, K. Li, et al. (2008). "Genetic variation in an individual human exome." PLoS Genet **4**(8): e1000160.

Ng, S. B., K. J. Buckingham, C. Lee, A. W. Bigham, H. K. Tabor, K. M. Dent, et al. (2010). "Exome sequencing identifies the cause of a mendelian disorder." Nat Genet **42**(1): 30-35.

Ng, S. B., E. H. Turner, P. D. Robertson, S. D. Flygare, A. W. Bigham, C. Lee, et al. (2009). "Targeted capture and massively parallel sequencing of 12 human exomes." Nature **461**(7261): 272-276.

NHLBI Exome Sequencing Project (2016). "Exome Variant Server, NHLBI GO Exome Sequencing Project (ESP), Seattle, WA [April, 2016 accessed].". from http://evs.gs.washington.edu/EVS/.

Nica, A. C., L. Parts, D. Glass, J. Nisbet, A. Barrett, M. Sekowska, et al. (2011). "The architecture of gene regulatory variation across multiple human tissues: the MuTHER study." PLoS Genet **7**(2): e1002003.

Nicholson, J. P., S. L. Teichman, M. H. Alderman, T. A. Sos, T. G. Pickering and J. H. Laragh (1983). "Cigarette smoking and renovascular hypertension." Lancet **2**(8353): 765-766.

Nicolae, D. L., E. Gamazon, W. Zhang, S. Duan, M. E. Dolan and N. J. Cox (2010). "Trait-associated SNPs are more likely to be eQTLs: annotation to enhance discovery from GWAS." PLoS Genet **6**(4): e1000888.

Noonan, J. P. and A. S. McCallion (2010). "Genomics of long-range regulatory elements." Annu Rev Genomics Hum Genet **11**: 1-23.

O'Roak, B. J., P. Deriziotis, C. Lee, L. Vives, J. J. Schwartz, S. Girirajan, et al. (2011). "Exome sequencing in sporadic autism spectrum disorders identifies severe de novo mutations." Nat Genet **43**(6): 585-589.

Olin, J. W., J. Froehlich, X. Gu, J. M. Bacharach, K. Eagle, B. H. Gray, et al. (2012). "The United States Registry for Fibromuscular Dysplasia: results in the first 447 patients." Circulation **125**(25): 3182-3190.

Olin, J. W., H. L. Gornik, J. M. Bacharach, J. Biller, L. J. Fine, B. H. Gray, et al. (2014). "Fibromuscular dysplasia: state of the science and critical unanswered questions: a scientific statement from the American Heart Association." Circulation **129**(9): 1048-1078.

Olin, J. W. and B. A. Sealove (2011). "Diagnosis, management, and future developments of fibromuscular dysplasia." J Vasc Surg **53**(3): 826-836 e821.

Osborn, A. G. and R. E. Anderson (1977). "Angiographic spectrum of cervical and intracranial fibromuscular dysplasia." Stroke **8**(5): 617-626.

Osborne, E. C., M. Lynch, R. McLachlan, A. O. Trounson and D. S. Cram (2007). "Microarray detection of Y chromosome deletions associated with male infertility." Reprod Biomed Online **15**(6): 673-680.

Ostergaard, P., M. A. Simpson, F. C. Connell, C. G. Steward, G. Brice, W. J. Woollard, et al. (2011). "Mutations in GATA2 cause primary lymphedema associated with a predisposition to acute myeloid leukemia (Emberger syndrome)." Nat Genet **43**(10): 929-931.

Ott, J., Y. Kamatani and M. Lathrop (2011). "Family-based designs for genome-wide association studies." Nat Rev Genet **12**(7): 465-474.

Ouchi, Y., H. Tagawa, M. Yamakado, R. Takanashi and S. Tanaka (1989). "Clinical significance of cerebral aneurysm in renovascular hypertension due to fibromuscular dysplasia: two cases in siblings." Angiology **40**(6): 581-588.

Page, C. M., S. E. Baranzini, B. H. Mevik, S. D. Bos, H. F. Harbo and B. K. Andreassen (2015). "Assessing the Power of Exome Chips." PLoS One **10**(10): e0139642.

Palubinskas, A. J. and H. R. Ripley (1964). "Fibromuscular Hyperplasia in Extrarenal Arteries." Radiology **82**: 451-455.

Palubinskas, A. J. and E. J. Wylie (1961). "Roentgen diagnosis of fibromuscular hyperplasia of the renal arteries." Radiology **76**: 634-639.

Pannier-Moreau, I., P. Grimbert, B. Fiquet-Kempf, A. Vuagnat, X. Jeunemaitre, P. Corvol, et al. (1997). "Possible familial origin of multifocal renal artery fibromuscular dysplasia." J Hypertens **15**(12 Pt 2): 1797-1801.

Paris, B., G. Bobrie, P. Rossignol, S. Le Coz, A. Chedid and P. F. Plouin (2006). "Blood pressure and renal outcomes in patients with kidney infarction and hypertension." J Hypertens **24**(8): 1649-1654.

Park, J. H., S. Wacholder, M. H. Gail, U. Peters, K. B. Jacobs, S. J. Chanock, et al. (2010). "Estimation of effect size distribution from genome-wide association studies and implications for future discoveries." Nat Genet **42**(7): 570-575.

Pascual, A., H. S. Bush and J. B. Copley (2005). "Renal fibromuscular dysplasia in elderly persons." Am J Kidney Dis **45**(4): e63-66.

Pate, G. E., R. Lowe and C. E. Buller (2005). "Fibromuscular dysplasia of the coronary and renal arteries?" Catheter Cardiovasc Interv **64**(2): 138-145.

Perdu, J., P. Boutouyrie, C. Bourgain, N. Stern, B. Laloux, E. Bozec, et al. (2007). "Inheritance of arterial lesions in renal fibromuscular dysplasia." J Hum Hypertens **21**(5): 393-400.

Perdu, J., A. P. Gimenez-Roqueplo, P. Boutouyrie, S. Beaujour, B. Laloux, V. Nau, et al. (2006). "Alpha1-antitrypsin gene polymorphisms are not associated with renal arterial fibromuscular dysplasia." J Hypertens **24**(4): 705-710.

Pers, T. H., J. M. Karjalainen, Y. Chan, H. J. Westra, A. R. Wood, J. Yang, et al. (2015). "Biological interpretation of genome-wide association studies using predicted gene functions." Nat Commun **6**: 5890.

Persu, A., A. Giavarini, E. Touze, A. Januszewicz, M. Sapoval, M. Azizi, et al. (2014). "European consensus on the diagnosis and management of fibromuscular dysplasia." J Hypertens **32**(7): 1367-1378.

Persu, A., E. Touze, E. Mousseaux, X. Barral, F. Joffre and P. F. Plouin (2012). "Diagnosis and management of fibromuscular dysplasia: an expert consensus." Eur J Clin Invest **42**(3): 338-347.

Petersen, G. M., L. Amundadottir, C. S. Fuchs, P. Kraft, R. Z. Stolzenberg-Solomon, K. B. Jacobs, et al. (2010). "A genome-wide association study identifies pancreatic cancer susceptibility loci on chromosomes 13q22.1, 1q32.1 and 5p15.33." Nat Genet **42**(3): 224-228.

Plouin, P. F., J. Perdu, A. La Batide-Alanore, P. Boutouyrie, A. P. Gimenez-Roqueplo and X. Jeunemaitre (2007). "Fibromuscular dysplasia." Orphanet J Rare Dis **2**: 28.

Pober, B. R., R. V. Lacro, C. Rice, V. Mandell and R. L. Teele (1993). "Renal findings in 40 individuals with Williams syndrome." Am J Med Genet **46**(3): 271-274.

Poloskey, S. L., E. Kim, R. Sanghani, A. H. Al-Quthami, P. Arscott, R. Moran, et al. (2012). "Low yield of genetic testing for known vascular connective tissue disorders in patients with fibromuscular dysplasia." Vasc Med **17**(6): 371-378.

Powell, J. T. (1998). "Vascular damage from smoking: disease mechanisms at the arterial wall." Vasc Med **3**(1): 21-28.

Prasad, M., M. S. Tweet, S. N. Hayes, S. Leng, J. J. Liang, M. F. Eleid, et al. (2015). "Prevalence of extracoronary vascular abnormalities and fibromuscular dysplasia in patients with spontaneous coronary artery dissection." Am J Cardiol **115**(12): 1672-1677.

Precone, V., V. Del Monaco, M. V. Esposito, F. D. De Palma, A. Ruocco, F. Salvatore, et al. (2015). "Cracking the Code of Human Diseases Using Next-Generation Sequencing: Applications, Challenges, and Perspectives." Biomed Res Int **2015**: 161648.

Pritchard, J. K. (2001). "Are rare variants responsible for susceptibility to complex diseases?" Am J Hum Genet **69**(1): 124-137.

Pritchard, J. K. and N. J. Cox (2002). "The allelic architecture of human disease genes: common disease-common variant...or not?" Hum Mol Genet **11**(20): 2417-2423.

Pruitt, K. D., G. R. Brown, S. M. Hiatt, F. Thibaud-Nissen, A. Astashyn, O. Ermolaeva, et al. (2014). "RefSeq: an update on mammalian reference sequences." Nucleic Acids Res **42**(Database issue): D756-763.

Pruitt, K. D., J. Harrow, R. A. Harte, C. Wallin, M. Diekhans, D. R. Maglott, et al. (2009). "The consensus coding sequence (CCDS) project: Identifying a common protein-coding gene set for the human and mouse genomes." Genome Res **19**(7): 1316-1323.

Putaala, J., A. J. Metso, T. M. Metso, N. Konkola, Y. Kraemer, E. Haapaniemi, et al. (2009). "Analysis of 1008 consecutive patients aged 15 to 49 with first-ever ischemic stroke: the Helsinki young stroke registry." Stroke **40**(4): 1195-1203.

Rabbani, B., N. Mahdieh, K. Hosomichi, H. Nakaoka and I. Inoue (2012). "Next-generation sequencing: impact of exome sequencing in characterizing Mendelian disorders." J Hum Genet **57**(10): 621-632.

Rabbani, B., M. Tekin and N. Mahdieh (2014). "The promise of whole-exome sequencing in medical genetics." J Hum Genet **59**(1): 5-15.

Rademakers, R., M. Baker, A. M. Nicholson, N. J. Rutherford, N. Finch, A. Soto-Ortolaza, et al. (2012). "Mutations in the colony stimulating factor 1 receptor (CSF1R) gene cause hereditary diffuse leukoencephalopathy with spheroids." Nat Genet **44**(2): 200-205.

Redon, R., S. Ishikawa, K. R. Fitch, L. Feuk, G. H. Perry, T. D. Andrews, et al. (2006). "Global variation in copy number in the human genome." Nature **444**(7118): 444-454.

Reich, D. E., M. Cargill, S. Bolk, J. Ireland, P. C. Sabeti, D. J. Richter, et al. (2001). "Linkage disequilibrium in the human genome." Nature **411**(6834): 199-204.

Reich, D. E. and E. S. Lander (2001). "On the allelic spectrum of human disease." Trends Genet **17**(9): 502-510.

Renton, A. E., E. Majounie, A. Waite, J. Simon-Sanchez, S. Rollinson, J. R. Gibbs, et al. (2011). "A hexanucleotide repeat expansion in C9ORF72 is the cause of chromosome 9p21-linked ALS-FTD." Neuron **72**(2): 257-268.

Rigat, B., C. Hubert, F. Alhenc-Gelas, F. Cambien, P. Corvol and F. Soubrier (1990). "An insertion/deletion polymorphism in the angiotensin I-converting enzyme gene accounting for half the variance of serum enzyme levels." J Clin Invest **86**(4): 1343-1346.

Rigden, D. J., X. M. Fernandez-Suarez and M. Y. Galperin (2016). "The 2016 database issue of Nucleic Acids Research and an updated molecular biology database collection." Nucleic Acids Res **44**(D1): D1-6.

Rinkel, G. J., M. Djibuti, A. Algra and J. van Gijn (1998). "Prevalence and risk of rupture of intracranial aneurysms: a systematic review." Stroke **29**(1): 251-256.

Risch, N. and K. Merikangas (1996). "The future of genetic studies of complex human diseases." Science **273**(5281): 1516-1517.

Rivas, M. A., M. Beaudoin, A. Gardet, C. Stevens, Y. Sharma, C. K. Zhang, et al. (2011). "Deep resequencing of GWAS loci identifies independent rare variants associated with inflammatory bowel disease." Nat Genet **43**(11): 1066-1073.

Robinton, D. A. and G. Q. Daley (2012). "The promise of induced pluripotent stem cells in research and therapy." Nature **481**(7381): 295-305.

Rushton, A. R. (1980). "The genetics of fibromuscular dysplasia." Arch Intern Med **140**(2): 233-236.

Sabharwal, R., P. Vladica and P. Coleman (2007). "Multidetector spiral CT renal angiography in the diagnosis of renal artery fibromuscular dysplasia." Eur J Radiol **61**(3): 520-527.

Saint Pierre, A. and E. Genin (2014). "How important are rare variants in common disease?" Brief Funct Genomics **13**(5): 353-361.

Sang, C. N., P. K. Whelton, U. M. Hamper, M. Connolly, S. Kadir, R. I. White, et al. (1989). "Etiologic factors in renovascular fibromuscular dysplasia. A case-control study." Hypertension **14**(5): 472-479.

Sanger, F., S. Nicklen and A. R. Coulson (1977). "DNA sequencing with chain-terminating inhibitors." Proc Natl Acad Sci U S A **74**(12): 5463-5467.

Savard, S., A. Azarine, X. Jeunemaitre, M. Azizi, P. F. Plouin and O. Steichen (2013). "Association of smoking with phenotype at diagnosis and vascular interventions in patients with renal artery fibromuscular dysplasia." Hypertension **61**(6): 1227-1232.

Savard, S., O. Steichen, A. Azarine, M. Azizi, X. Jeunemaitre and P. F. Plouin (2012). "Association between 2 angiographic subtypes of renal artery fibromuscular dysplasia and clinical characteristics." Circulation **126**(25): 3062-3069.

Saw, J., E. Aymong, T. Sedlak, C. E. Buller, A. Starovoytov, D. Ricci, et al. (2014). "Spontaneous coronary artery dissection: association with predisposing arteriopathies and precipitating stressors and cardiovascular outcomes." Circ Cardiovasc Interv **7**(5): 645-655.

Saw, J., H. Bezerra, H. L. Gornik, L. Machan and G. B. Mancini (2016). "Angiographic and Intracoronary Manifestations of Coronary Fibromuscular Dysplasia." Circulation **133**(16): 1548-1559.

Saw, J., R. Poulter, A. Fung, D. Wood, J. Hamburger and C. E. Buller (2012). "Spontaneous coronary artery dissection in patients with fibromuscular dysplasia: a case series." Circ Cardiovasc Interv **5**(1): 134-137.

Schadt, E. E., S. A. Monks, T. A. Drake, A. J. Lusis, N. Che, V. Colinayo, et al. (2003). "Genetics of gene expression surveyed in maize, mouse and man." Nature **422**(6929): 297-302.

Schievink, W. I. (2001). "Spontaneous dissection of the carotid and vertebral arteries." N Engl J Med **344**(12): 898-906.

Schievink, W. I. and J. Bjornsson (1996). "Fibromuscular dysplasia of the internal carotid artery:

a clinicopathological study." Clin Neuropathol **15**(1): 2-6.

Schievink, W. I., J. Bjornsson, J. E. Parisi and U. B. Prakash (1994). "Arterial fibromuscular dysplasia associated with severe alpha 1-antitrypsin deficiency." Mayo Clin Proc **69**(11): 1040-1043.

Schievink, W. I., M. R. Puumala, F. B. Meyer, C. Raffel, J. A. Katzmann and J. E. Parisi (1996). "Giant intracranial aneurysm and fibromuscular dysplasia in an adolescent with alpha 1-antitrypsin deficiency." J Neurosurg **85**(3): 503-506.

Schwarz, J. M., C. Rodelsperger, M. Schuelke and D. Seelow (2010). "MutationTaster evaluates disease-causing potential of sequence alterations." Nat Methods **7**(8): 575-576.

Sebat, J., B. Lakshmi, D. Malhotra, J. Troge, C. Lese-Martin, T. Walsh, et al. (2007). "Strong association of de novo copy number mutations with autism." Science **316**(5823): 445-449.

Seth, A., D. L. Stemple and I. Barroso (2013). "The emerging use of zebrafish to model metabolic disease." Dis Model Mech **6**(5): 1080-1088.

Shea, J., V. Agarwala, A. A. Philippakis, J. Maguire, E. Banks, M. Depristo, et al. (2011). "Comparing strategies to fine-map the association of common SNPs at chromosome 9p21 with type 2 diabetes and myocardial infarction." Nat Genet **43**(8): 801-805.

Shendure, J. and H. Ji (2008). "Next-generation DNA sequencing." Nat Biotechnol **26**(10): 1135-1145.

Sherry, S. T., M. H. Ward, M. Kholodov, J. Baker, L. Phan, E. M. Smigielski, et al. (2001). "dbSNP: the NCBI database of genetic variation." Nucleic Acids Res **29**(1): 308-311.

Silhol, F., G. Sarlon-Bartoli, L. Daniel, J. M. Bartoli, S. Cohen, H. Lepidi, et al. (2015). "Intranuclear expression of progesterone receptors in smooth muscle cells of renovascular fibromuscular dysplasia: a pilot study." Ann Vasc Surg **29**(4): 830-835.

Simpson, M. A., C. Deshpande, D. Dafou, L. E. Vissers, W. J. Woollard, S. E. Holder, et al. (2012). "De novo mutations of the gene encoding the histone acetyltransferase KAT6B cause Genitopatellar syndrome." Am J Hum Genet **90**(2): 290-294.

Simpson, M. A., M. D. Irving, E. Asilmaz, M. J. Gray, D. Dafou, F. V. Elmslie, et al. (2011). "Mutations in NOTCH2 cause Hajdu-Cheney syndrome, a disorder of severe and progressive bone loss." Nat Genet **43**(4): 303-305.

Sivakumaran, S., F. Agakov, E. Theodoratou, J. G. Prendergast, L. Zgaga, T. Manolio, et al. (2011). "Abundant pleiotropy in human complex diseases and traits." Am J Hum Genet **89**(5): 607-618.

Slovut, D. P. and J. W. Olin (2004). "Fibromuscular dysplasia." N Engl J Med **350**(18): 1862-1871.

Smith, C. A. (1963). "Testing for Heterogeneity of Recombination Fraction Values in Human Genetics." Ann Hum Genet **27**: 175-182.

Solder, B., W. Streif, H. Ellemunter, U. Mayr and W. Jaschke (1997). "Fibromuscular dysplasia of the internal carotid artery in a child with alpha-1-antitrypsin deficiency." Dev Med Child Neurol **39**(12): 827-829.

Spielman, R. S., R. E. McGinnis and W. J. Ewens (1993). "Transmission test for linkage disequilibrium: the insulin gene region and insulin-dependent diabetes mellitus (IDDM)." Am J Hum Genet **52**(3): 506-516.

Spring, D. B., O. Salvatierra, Jr., A. J. Palubinskas, W. J. Amend, Jr., F. G. Vincenti and N. J. Feduska (1979). "Results and significance of angiography in potential kidney donors." Radiology **133**(1): 45-47.

Stanley, J. C. (1996). Renal Artery Fibrodysplasia and Renovascular Hypertension. London, UK, WB Saunders.

Stanley, J. C. and W. J. Fry (1975). "Renovascular hypertension secondary to arterial fibrodysplasia in adults: criteria for operation and results of surgical therapy." Arch Surg **110**(8): 922-928.

Stanley, J. C., B. L. Gewertz, E. L. Bove, V. Sottiurai and W. J. Fry (1975). "Arterial fibrodysplasia. Histopathologic character and current etiologic concepts." Arch Surg **110**(5): 561-566.

Stavenow, L., B. Henriques, M. Ekberg, S. E. Bergentz and B. Hood (1978). "Combination of fibromuscular hyperplasia, renal aplasia, hypoplasia or dysplasia and otosclerosis occurring in the same individual or the same family." Acta Med Scand **203**(5): 357-362.

Stenson, P. D., M. Mort, E. V. Ball, K. Shaw, A. Phillips and D. N. Cooper (2014). "The Human Gene Mutation Database: building a comprehensive mutation repository for clinical and molecular genetics, diagnostic testing and personalized genomic medicine." Hum Genet **133**(1): 1-9.

Stewart, M. T., M. W. Moritz, R. B. Smith, 3rd, J. T. Fulenwider and G. D. Perdue (1986). "The natural history of carotid fibromuscular dysplasia." J Vasc Surg **3**(2): 305-310.

Stewart, W. F., A. Shechter and B. K. Rasmussen (1994). "Migraine prevalence. A review of population-based studies." Neurology **44**(6 Suppl 4): S17-23.

Stoletov, K., L. Fang, S. H. Choi, K. Hartvigsen, L. F. Hansen, C. Hall, et al. (2009). "Vascular lipid accumulation, lipoprotein oxidation, and macrophage lipid uptake in hypercholesterolemic zebrafish." Circ Res **104**(8): 952-960.

Strachan, T. R., A.P. (1999). Genes in pedigrees. Human Molecular Genetics. 2nd ed.

. New York: Wiley-Liss.

Stranneheim, H. and A. Wedell (2016). "Exome and genome sequencing: a revolution for the discovery and diagnosis of monogenic disorders." J Intern Med **279**(1): 3-15.

Studies, N.-N. W. G. o. R. i. A., S. J. Chanock, T. Manolio, M. Boehnke, E. Boerwinkle, D. J. Hunter, et al. (2007). "Replicating genotype-phenotype associations." Nature **447**(7145): 655-660.

Sudmant, P. H., T. Rausch, E. J. Gardner, R. E. Handsaker, A. Abyzov, J. Huddleston, et al. (2015). "An integrated map of structural variation in 2,504 human genomes." Nature **526**(7571): 75-81.

Sullivan, J., R. Kopp, K. Stratton, C. Manschreck, M. Corines, R. Rau-Murthy, et al. (2015). "An analysis of the association between prostate cancer risk loci, PSA levels, disease aggressiveness and disease-specific mortality." Br J Cancer **113**(1): 166-172.

Suzuki, H., H. Daida, H. Sakurai and H. Yamaguchi (1999). "Familial fibromuscular dysplasia of bilateral brachial arteries." Heart **82**(2): 251-252.

Swensson, R. G. (1996). "Unified measurement of observer performance in detecting and localizing target objects on images." Med Phys **23**(10): 1709-1725.

Szklarczyk, D., A. Franceschini, S. Wyder, K. Forslund, D. Heller, J. Huerta-Cepas, et al.

(2015). "STRING v10: protein-protein interaction networks, integrated over the tree of life." Nucleic Acids Res **43**(Database issue): D447-452.

Tabor, H. K., N. J. Risch and R. M. Myers (2002). "Candidate-gene approaches for studying complex genetic traits: practical considerations." Nat Rev Genet **3**(5): 391-397.

Tantoso, E., L. P. Wong, B. Li, W. Y. Saw, W. Xu, P. Little, et al. (2014). "Evaluating the coverage and potential of imputing the exome microarray with next-generation imputation using the 1000 Genomes Project." PLoS One **9**(9): e106681.

The French Exome (FREX) Project. "The French Exome (FREX) Project : a reference panel of exomes from French regions [consulté le 25/04/2016]." from https://www.france-genomique.org/spip/spip.php?article158.

Touze, E., C. Oppenheim, D. Trystram, G. Nokam, M. Pasquini, S. Alamowitch, et al. (2010). "Fibromuscular dysplasia of cervical and intracranial arteries." Int J Stroke **5**(4): 296-305.

Tranchevent, L. C., R. Barriot, S. Yu, S. Van Vooren, P. Van Loo, B. Coessens, et al. (2008). "ENDEAVOUR update: a web resource for gene prioritization in multiple species." Nucleic Acids Res **36**(Web Server issue): W377-384.

Tweet, M. S., R. Gulati and S. N. Hayes (2015). "What Clinicians Should Know Alphabout Spontaneous Coronary Artery Dissection." Mayo Clin Proc **90**(8): 1125-1130.

UK10K Consortium (2011). "UK10K - UK10K Goals." from http://www.uk10k.org/goals.html.

Urnov, F. D., E. J. Rebar, M. C. Holmes, H. S. Zhang and P. D. Gregory (2010). "Genome editing with engineered zinc finger nucleases." Nat Rev Genet **11**(9): 636-646.

Varennes, L., F. Tahon, A. Kastler, S. Grand, F. Thony, J. P. Baguet, et al. (2015). "Fibromuscular dysplasia: what the radiologist should know: a pictorial review." Insights Imaging **6**(3): 295-307.

Varshney, G. K., W. Pei, M. C. LaFave, J. Idol, L. Xu, V. Gallardo, et al. (2015). "High-throughput gene targeting and phenotyping in zebrafish using CRISPR/Cas9." Genome Res **25**(7): 1030-1042.

Vasbinder, G. B., P. J. Nelemans, A. G. Kessels, A. A. Kroon, P. W. de Leeuw and J. M. van Engelshoven (2001). "Diagnostic tests for renal artery stenosis in patients suspected of having renovascular hypertension: a meta-analysis." Ann Intern Med **135**(6): 401-411.

Vasbinder, G. B., P. J. Nelemans, A. G. Kessels, A. A. Kroon, J. H. Maki, T. Leiner, et al. (2004). "Accuracy of computed tomographic angiography and magnetic resonance angiography for diagnosing renal artery stenosis." Ann Intern Med **141**(9): 674-682; discussion 682.

Virmani, R. C.-M., N.; Taylor, A.J.; (2013). Congenital anomalies and malformations of the vasculature. In: Creager MA, Beckman JA, Loscalzo J Vascular Medicine: A Companion to Barunwald's Heart Disease. 2nd ed. Philadelphia,, PA: Elsevier Saunders;2013.

Voelkerding, K. V., S. Dames and J. D. Durtschi (2010). "Next generation sequencing for clinical diagnostics-principles and application to targeted resequencing for hypertrophic cardiomyopathy: a paper from the 2009 William Beaumont Hospital Symposium on Molecular Pathology." J Mol Diagn **12**(5): 539-551.

von Figura, G., J. P. t. Morris, C. V. Wright and M. Hebrok (2014). "Nr5a2 maintains acinar cell differentiation and constrains oncogenic Kras-mediated pancreatic neoplastic initiation." Gut **63**(4): 656-664.

Wagner, G. P. and J. Zhang (2011). "The pleiotropic structure of the genotype-phenotype map: the evolvability of complex organisms." Nat Rev Genet **12**(3): 204-213.

Wain, L. V., J. A. Armour and M. D. Tobin (2009). "Genomic copy number variation, human health, and disease." Lancet **374**(9686): 340-350.

Wall, J. D. and J. K. Pritchard (2003). "Haplotype blocks and linkage disequilibrium in the human genome." Nat Rev Genet **4**(8): 587-597.

Walsh, T., J. M. McClellan, S. E. McCarthy, A. M. Addington, S. B. Pierce, G. M. Cooper, et al. (2008). "Rare structural variants disrupt multiple genes in neurodevelopmental pathways in schizophrenia." Science **320**(5875): 539-543.

Walsh, T., H. Shahin, T. Elkan-Miller, M. K. Lee, A. M. Thornton, W. Roeb, et al. (2010). "Whole exome sequencing and homozygosity mapping identify mutation in the cell polarity protein GPSM2 as the cause of nonsyndromic hearing loss DFNB82." Am J Hum Genet **87**(1): 90-94.

Wang, J. L., X. Yang, K. Xia, Z. M. Hu, L. Weng, X. Jin, et al. (2010). "TGM6 identified as a novel causative gene of spinocerebellar ataxias using exome sequencing." Brain **133**(Pt 12): 3510-3518.

Wang, Q., Q. Lu and H. Zhao (2015). "A review of study designs and statistical methods for genomic epidemiology studies using next generation sequencing." Front Genet **6**: 149.

Wang, Z., M. Gerstein and M. Snyder (2009). "RNA-Seq: a revolutionary tool for transcriptomics." Nat Rev Genet **10**(1): 57-63.

Weinshilboum, R. (2003). "Inheritance and drug response." N Engl J Med **348**(6): 529-537.

Weissenbach, J. (1993). "Microsatellite polymorphisms and the genetic linkage map of the human genome." Curr Opin Genet Dev **3**(3): 414-417.

Weissenbach, J., G. Gyapay, C. Dib, A. Vignal, J. Morissette, P. Millasseau, et al. (1992). "A second-generation linkage map of the human genome." Nature **359**(6398): 794-801.

Wellcome Trust Case Control, C. (2007). "Genome-wide association study of 14,000 cases of seven common diseases and 3,000 shared controls." Nature **447**(7145): 661-678.

Welter, D., J. MacArthur, J. Morales, T. Burdett, P. Hall, H. Junkins, et al. (2014). "The NHGRI GWAS Catalog, a curated resource of SNP-trait associations." Nucleic Acids Res **42**(Database issue): D1001-1006.

Wessel, J., A. Y. Chu, S. M. Willems, S. Wang, H. Yaghootkar, J. A. Brody, et al. (2015). "Low-frequency and rare exome chip variants associate with fasting glucose and type 2 diabetes susceptibility." Nat Commun **6**: 5897.

Willer, C. J., Y. Li and G. R. Abecasis (2010). "METAL: fast and efficient meta-analysis of genomewide association scans." Bioinformatics **26**(17): 2190-2191.

Wooster, R., S. L. Neuhausen, J. Mangion, Y. Quirk, D. Ford, N. Collins, et al. (1994). "Localization of a breast cancer susceptibility gene, BRCA2, to chromosome 13q12-13." Science **265**(5181): 2088-2090.

Wright, A. F. (2005). Genetic Variation: Polymorphisms and Mutations. Encyclopedia of Life Sciences.

Wu, M. C., S. Lee, T. Cai, Y. Li, M. Boehnke and X. Lin (2011). "Rare-variant association testing for sequencing data with the sequence kernel association test." Am J Hum Genet **89**(1): 82-93.

Yoon, P. W., R. S. Olney, M. J. Khoury, W. M. Sappenfield, G. F. Chavez and D. Taylor (1997). "Contribution of birth defects and genetic diseases to pediatric hospitalizations. A population-based study." Arch Pediatr Adolesc Med **151**(11): 1096-1103.

Zeggini, E., W. Rayner, A. P. Morris, A. T. Hattersley, M. Walker, G. A. Hitman, et al. (2005). "An evaluation of HapMap sample size and tagging SNP performance in large-scale empirical and simulated data sets." Nat Genet **37**(12): 1320-1322.

Zimprich, A., S. Biskup, P. Leitner, P. Lichtner, M. Farrer, S. Lincoln, et al. (2004). "Mutations in LRRK2 cause autosomal-dominant parkinsonism with pleomorphic pathology." Neuron **44**(4): 601-607.

Zlotogora, J. (2003). "Penetrance and expressivity in the molecular age." Genet Med **5**(5): 347-352.

Zollner, S. and J. K. Pritchard (2007). "Overcoming the winner's curse: estimating penetrance parameters from case-control data." Am J Hum Genet **80**(4): 605-615.

Zwick, M. E., D. J. Cutler and A. Chakravarti (2000). "Patterns of genetic variation in Mendelian and complex traits." Annu Rev Genomics Hum Genet **1**: 387-407.

RESUME

La dysplasie fibromusculaire artérielle (DFM) est un groupe de pathologies vasculaires non inflammatoires, et non athéromateuses de la paroi artérielle. Elle est caractérisée par la sténose, l'occlusion, l'anévrisme ou la dissection des artères de petit et moyen calibre, en particulier les artères rénales et le tronc supra-aortique. La DFM est un facteur de risque de l'hypertension et de l'accident vasculaire cérébral. Elle touche essentiellement les femmes (80% des cas) de moins de 50 ans. La prévalence en population générale est inconnue et les estimations varient de 0.4% pour les formes cliniques à 4% dans une cohorte de donneurs de reins. Une agrégation familiale a été démontrée et une composante génétique suggérée. L'objectif de mon travail de thèse était de caractériser les bases génétiques la DFM.

Dans la première partie, nous avons analysé des variants génétiques rares générés par séquençage d'exome chez 16 cas apparentés de DFM issus de 7 fratries. Aucun gène majeur n'était muté pour l'ensemble des fratries ou pour au moins 3 fratries sur 7. Cependant, nous avons pu mettre en évidence puis validé un enrichissement en variants rares à fort potentiel fonctionnel de quatre gènes candidats pour la DFM (*MYLK*, *OBSCN*, *DYNC2H1*, *RNF213*) en combinant l'approche de séquençage d'exomes et l'étude d'association gène entier de 62767 variants rares (MAF < 5%) générés par génotypage avec la puce *Exome-chip* chez 249 cas non apparentés de DFM et 689 témoins. Cependant, l'implication de ces gènes dans la DFM doit être confirmée dans d'autres familles, et par des études de validations fonctionnelles.

Dans la seconde partie, nous avons étudié l'association avec la DFM de 25606 variants fréquents (MAF ≥ 5%) de l'*Exome-chip*. Les résultats majeurs obtenus ont été répliqués dans une première étude (402 cas de DFM et 2537 témoins) puis dans 3 autres études incluant 512 cas de DFM et 669 témoins. La méta-analyse de l'ensemble ayant permis l'identification de l'association avec la DFM du polymorphisme rs9349379[A] situé dans l'intron du gène *PHACTR1* (OR=1,39 [1,39-1,54] ; P=7,36 $\times 10^{-10}$). Ce variant est aussi un facteur de risque pour la maladie coronaire, la migraine et la dissection de l'artère cervicale. Des études complémentaires conduites chez 2458 volontaires non malades ont permis de montrer que l'allèle à risque pour la DFM est associé avec une augmentation de l'épaisseur intima média (P=1,97$\times 10^{-4}$) et du rapport de la paroi sur la lumière artérielle (P=0,002), deux paramètres précédemment décrits comme augmentés chez les cas de DFM. Ensuite PHACTR1 a été détecté par immunohistochimie dans l'endothélium et les cellules musculaires lisses de carotides dysplasiques et non dysplasiques avec une expression augmentée de PHACTR1 pour les porteurs de l'allèle à risque de DFM dans des cultures primaires de fibroblastes humains (N=86, P=0,003). Enfin, l'invalidation de *Phactr1* chez le poisson zèbre conduit à une dilatation des vaisseaux indiquant un défaut du développement vasculaire.

Ce travail confirme le caractère multifactoriel et hétérogène de la DFM et ouvre de nouvelles perspectives pour évaluer l'ensemble de la variabilité génomique des patients de DFM par des approches massives de génétique épidémiologique.