



**HAL**  
open science

# Étude du rôle de facteurs génétiques liés à la physiologie des femmes dans la prédisposition à la dysplasie fibromusculaire (DFM) et à la dissection spontanée de l'artère coronaire (SCAD), artériopathies à forte prédominance féminine

Inès Sadeg-Sayoud

## ► To cite this version:

Inès Sadeg-Sayoud. Étude du rôle de facteurs génétiques liés à la physiologie des femmes dans la prédisposition à la dysplasie fibromusculaire (DFM) et à la dissection spontanée de l'artère coronaire (SCAD), artériopathies à forte prédominance féminine. Génétique. Université Paris Cité, 2022. Français. NNT : 2022UNIP5196 . tel-04659970

**HAL Id: tel-04659970**

**<https://theses.hal.science/tel-04659970v1>**

Submitted on 23 Jul 2024

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

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

# Université Paris Cité

Bio Sorbonne Paris Cité ED 562

Paris centre de recherche cardio-vasculaire (PARCC)

Equipe : « Approches génétiques pour comprendre les maladies artérielles »

**Etude du rôle de facteurs génétiques liés à la physiologie des femmes dans la prédisposition à la dysplasie fibromusculaire (DFM) et à la dissection spontanée de l'artère coronaire (SCAD), artériopathies à forte prédominance féminine.**

Par Inès SADEG-SAYOUD

Doctorat en Génétique

Dirigée par Nabila BOUATIA-NAJI

Présentée et soutenue publiquement le 16 Novembre 2022

Devant le jury composé de :

Dr. Marie-Christine BOUTRON-RUAULT	DR1 INSERM	Université Paris-Saclay	Rapporteur
Dr. Hervé PERDRY	MCU-HDR	Université Paris-Saclay	Rapporteur
Pr. Elisabeth TOURNIER-LASSERVE	PU-PH	Université Paris cité	Examineur
Pr. Alexandre PERSU	Professeur des universités	Université catholique de Louvain	Examineur
Dr. Nabila BOUATIA-NAJI	DR2 INSERM	Université Paris cité	Directrice de thèse

*A mes chers parents*

*A mon cher mari*

*A mes chers enfants Maryam-Zina et Adam*

*A mes chers frère et sœurs*

## Remerciements

Je remercie ma directrice de thèse Mme Nabila BOUATIA-NAJI, d'avoir dirigé cette thèse du début jusqu'à la fin, et de m'avoir offert cette chance d'appliquer mes connaissances statistiques dans la génétique, un domaine fascinant que j'ai découvert à travers ce projet de thèse.

Je remercie Marie-Christine BOUTRON-RUAULT et Hervé PERDRY d'avoir accepté de juger mon travail de thèse ainsi que Elisabeth TOURNIER-LASSERVE et Alexandre PERSU d'avoir consacré du temps pour l'examiner.

Je remercie tous les membres de l'équipe 3A, Joéphine, Lu, Yingwei, Adrien, ainsi que tous ceux que j'ai rencontrés pendant la thèse. Je remercie également Muriel DELACROIX pour sa présence et sa réactivité concernant l'aspect administratif.

Je remercie Takiy de m'avoir appris beaucoup de choses, d'avoir répondu présent à chaque fois que je le sollicite ainsi que sa contribution dans chaque étape de mon travail.

Je remercie Chaïma pour sa contribution dans la correction de ma thèse, ses encouragements et ses conseils, sa réactivité ainsi que ses petites phrases qui me font rire.

Je remercie mes chers parents d'avoir toujours cru en moi, mon mari pour sa présence et son aide précieuse dans ma première et ma seconde thèse, mes sœurs, mon frère, mes amies pour leurs conseils et leurs encouragements, et toute ma famille et belle famille.

Je remercie particulièrement mes deux trésors Maryam-Zina et Adam pour tout le bonheur qu'ils m'offrent au quotidien.

Je remercie toutes les personnes qui ont contribué de près ou de loin à la réalisation de mon projet de thèse.

## Table des matières

<b>Remerciments</b> .....	<b>2</b>
<b>Table des matières</b> .....	<b>3</b>
<b>Liste des Publications</b> .....	<b>12</b>
<b>Liste des figures et tableaux</b> .....	<b>14</b>
<b>Liste des tableaux</b> .....	<b>16</b>
<b>Liste des abreviations</b> .....	<b>17</b>
<b>Résumé</b> .....	<b>19</b>
<b>Abstract</b> .....	<b>20</b>
<b>Partie I. Introduction générale</b> .....	<b>21</b>
<b>Chapitre 1.</b> .....	<b>21</b>
<b>Maladies cardiovasculaires</b> .....	<b>21</b>
1. Généralités sur les maladies cardiovasculaires .....	22
a. Introduction aux maladies cardiovasculaires .....	22
b. Les principaux groupes de maladies cardiovasculaires.....	25
i. Cardiopathie coronarienne .....	25

ii.	Accident vasculaire cérébral .....	26
iii.	Maladie artérielle périphérique .....	27
i.	Facteurs de risque majeurs .....	30
▪	Hypertension artérielle .....	30
▪	Dyslipidémie .....	30
▪	Diabète.....	31
▪	Tabagisme .....	31
▪	Âge et sexe .....	31
ii.	Facteurs globaux.....	32
c.	Risque cardiovasculaire selon le sexe .....	32
2.	Facteurs de risques de maladies cardiovasculaires liés à la physiologie de la femme	34
a.	Les étapes physiologiques critiques dans la vie de la femme .....	34
i.	Puberté.....	34
ii.	Ménopause.....	34
iii.	Grossesse et allaitement .....	34
b.	Fluctuations hormonales et risque de maladies cardiovasculaires .....	35
i.	Définition des hormones féminines.....	35
ii.	Risque cardiovasculaires lié aux étapes physiologiques chez la femme..	37
3.	La dysplasie fibromusculaire artérielle .....	39

a.	Définition clinique.....	39
b.	Historique .....	41
c.	Diagnostique différentiel de la DFM.....	41
d.	Facteurs de risque de la DFM.....	44
e.	Prédominance féminine dans la DFM .....	44
4.	La dissection spontanée de l'artère coronaire .....	45
a.	Définition clinique.....	45
b.	Diagnostique différentiel de la SCAD.....	45
c.	Facteurs de risque de la SCAD.....	46
	<b>Chapitre 2. ....</b>	<b>47</b>
	<b>Méthodes d'épidémiogénétique de détection des facteurs de risque génétiques aux maladies à génétique complexe .....</b>	<b>47</b>
1.	Introduction .....	48
2.	La base génétique des maladies cardiovasculaires.....	50
a.	Héritabilité des maladies et traits liés à la santé cardiovasculaire.....	50
b.	Etudes de liaisons génétiques et candidats biologiques .....	51
i.	Définition et principe de l'étude de liaisons génétiques.....	51
ii.	Définition et principe de l'étude gène candidat.....	52
c.	Apport des études de séquençage d'exomes .....	53
i.	Définition et principe.....	53

ii.	Quelques succès .....	54
3.	Contributions des études d'associations pangénomiques à la génétique des maladies	55
a.	Principes généraux des études d'association pan-génomique (GWAS) .....	55
b.	Apports à la base génétique des maladies cardiovasculaires .....	58
c.	Rôle des variations génétiques fréquentes dans la DFM et dans la SCAD .....	60
4.	Rôles des variations génétiques liés aux chromosomes sexuels dans le risque génétique cardiovasculaire .....	61
5.	Apports de l'approche des scores polygéniques à partir des données GWAS .....	62
a.	Définition et principe.....	63
b.	Applications des scores polygéniques dans le domaine cardiovasculaire....	69
6.	Apport de la corrélation génétique et de la randomisation Mendélienne.....	69
a.	La corrélation génétique .....	70
i.	Principales méthodes .....	70
b.	La randomisation Mendélienne .....	71
i.	Définition et principe.....	71
c.	Contribution de la corrélation génétique et de la randomisation mendélienne à l'étude de la génétique des maladies cardiovasculaires .....	75
	<b>Partie II. Résultats .....</b>	<b>77</b>
	<b>Chapitre 3. ....</b>	<b>77</b>

**L'étude du rôle des variations génétiques fréquentes du chromosome X dans le risque de la DFM et de la SCAD ..... 77**

1. Méthodologies appliquées ..... 78

a. Populations étudiées ..... 78

i. Études DFM ..... 78

▪ ARCADIA ..... 78

▪ 3 Cités ..... 78

▪ ARCADIA-POL ..... 79

▪ WOBASZII ..... 79

▪ FMD European/International Registry and Initiative (FEIRI) ..... 79

▪ ASKLEPIOS. .... 80

▪ Étude cas-témoin Mayo VDB. .... 80

▪ Étude DEFINE-DFM. .... 81

▪ Étude cas-témoin de l'Université du Michigan-Michigan Genomics Initiative/Cleveland Clinic. .... 81

ii. Études SCAD ..... 84

▪ Étude cas-témoins DISCO-3C..... 84

▪ Études cas-témoins SCAD-UK. .... 84

▪ Étude cas-témoins de la Mayo Clinic ..... 85

▪ Étude cas-témoins CanSCAD / MGI (UBC/MGI)..... 86

▪	Études cas-témoins VCCRI.....	86
▪	Étude cas-témoïn DEFINE (MS) .....	87
b.	Étapes majeures de la réalisation des GWAS Françaises.....	88
i.	Génotypage.....	88
ii.	Contrôle qualité .....	90
▪	Etape 1 : QC sur les individus .....	90
▪	Etape 2 : QC sur les SNPs. ....	93
▪	Etape 3 : Imputation .....	94
▪	Etape 4 : QC post imputation .....	94
c.	Méthodologies appliquées pour l’analyse des variants du Chromosome X.....	96
i.	Principe et contexte de l’étude des chromosomes sexuels.....	96
ii.	Rappel du précepte et mécanisme de l’inactivation du chromosome X... 96	
▪	Définition de l’inactivation de X.....	99
▪	Quelques aspects du mécanisme de l’inactivation du Chr X (ICX).....	100
▪	Mise en place de ICX .....	100
iii.	Méthodologie X-WAS .....	102
▪	Définition.....	102
▪	Contrôle qualité .....	102
▪	Tests d'association avec XWAS .....	103
▪	Test d'association à un seul marqueur sur le chromosome X.....	103

▪	Analyse stratifiée par sexe à marqueur unique sur le chromosome X ...	104
2.	Résultats .....	104
a.	Étude de DFM .....	104
i.	Modèle 1 : sans l'inactivation de X.....	104
▪	QC standard des données de géotypage .....	105
▪	Imputation et QC post-imputation.....	107
▪	Test d'association .....	107
ii.	Modèle 2 : avec Inactivation de X.....	115
▪	Analyse d'association des trois cohortes ensemble.....	120
iii.	XWAS .....	122
iv.	Test d'hétérogénéité .....	125
b.	Étude SCAD .....	125
i.	Modèle : avec Inactivation de X.....	125
▪	QC standard, imputation et QC-post imputation.....	125
▪	Test d'association .....	128
ii.	Méta-analyse .....	128
iii.	Test d'hétérogénéité .....	130
3.	Conclusion.....	131
	<b>Chapitre 4. ....</b>	<b>132</b>

**Exploitations de données résumées de GWAS de traits liés à la physiologie des femmes pour étudier leurs liens avec le risque génétique de la DFM ..... 132**

1. Méthodologies appliquées .....	133
a. Exploitations de données résumées GWAS .....	133
b. Descriptive des données exploitées .....	133
i. Données individuelles DFM.....	133
ii. Données d'âge de puberté, d'âge de ménopause et de la durée de la vie reproductive.....	135
c. Construction des scores polygéniques.....	135
i. Contrôle qualité .....	135
ii. Calcul du score polygénique .....	137
iii. Test d'association avec la DFM .....	137
d. Corrélations génétiques .....	138
i. Le principe de la méthode du LD score .....	138
ii. Contrôle qualité des données et analyse de la corrélation.....	138
▪ Etape 1. L'interception de régression du score LD pour un GWAS.....	138
▪ Etape 2. Le calcul de l'héritabilité des SNPs pour la DFM .....	139
▪ Etape 3. La corrélation génétique entre la DFM et le trait lié à la physiologie de la femme.....	139
e. Randomisation Mendélienne .....	139
i. Présentation des différentes approches MR appliquées .....	139

▪	Approche de pondération inverse de la variance (IVW).....	140
▪	L'approche RM-Egger .....	142
▪	Approche de la médiane pondérée. ....	143
▪	L'approche du mode pondéré.....	143
ii.	Prérequis de l'analyse RM et sélection des variables instruments par trait 144	
2.	Résultats .....	144
a.	Scores Polygénique .....	144
b.	Corrélation génétique .....	146
c.	Randomisation mendélienne .....	147
3.	Conclusion.....	150
	<b>Partie III .....</b>	<b>151</b>
	<b>Discussion générale et perspectives .....</b>	<b>151</b>
	<b>Références bibliographiques .....</b>	<b>155</b>

## Liste des Publications

### ➤ Papier en cours de publication

**Genetic determinants of reproductive age milestones do not associate with fibromuscular dysplasia.** Ines Sadeg-Sayoud, Joséphine Henry, Aleksander Prejbisz, Piotr Dobrowolski, Andrzej Januszewicz, Miika Viikkula, Alexandre Persu, Takiy-Eddine Berrandou, Nabila Bouatia-Naji. **Hypertension (en preparation).**

### ➤ Publications issues de collaboration au sein du laboratoire de thèse

**Genome-wide association meta-analysis of spontaneous coronary artery dissection reveals common variants and genes related to artery integrity and tissue-mediated coagulation.** David Adlam, Takiy-Eddine Berrandou, Adrien Georges, Christopher P. Nelson, Eleni Giannoulatou, Joséphine Henry, Lijiang Ma, Montgomery Blencowe, Tamiel N. Turley, Min-Lee Yang, Peter S. Braund, **Ines Sadeg-Sayoud**, Siiri E. Iismaa, Matthew L. Kosel, Xiang Zhou, Stephen E. Hamby, Jenny Cheng, Lu Liu, Ingrid Tarr, David W.M. Muller, Valentina d'Escamard, Annette King, Liam R. Brunham, Ania A. Baranowska-Clarke, Stéphanie Debette, Philippe Amouyel, Jeffrey W. Olin, Snehal Patil, Stephanie E. Hesselton, Keerat Junday, Stavroula Kanoni, Krishna Aragam, Adam S. Butterworth, CARDIoGRAMPlusC4D, MEGASTROKE, International stroke genetics consortium (ISGC) intracranial aneurysm working group, Marysia S. Tweet, Rajiv Gulati, Nicolas Combaret, DISCO register investigators, Daniella Kadian-Dodov, Jon Kalman, Diane Fatkin, Jacqueline Saw, Tom R. Webb, Sharonne N. Hayes, Xia Yang, Santhi K. Ganesh, Timothy M. Olson, Jason C. Kovacic, Robert M. Graham, Nilesh J. Samani, Nabila Bouatia-Naji doi: <https://doi.org/10.1101/2022.07.05.22277238>. 2021.

**Genetic investigation of fibromuscular dysplasia identifies risk loci and shared genetics with common cardiovascular diseases.** Georges A, Yang ML, Berrandou TE, Bakker MK, Dikilitas O, Kiando SR, Ma L, Satterfield BA, Sengupta S, Yu M, Deleuze JF, Dupré D, Hunker KL, Kyryachenko S, Liu L, **Sayoud-Sadeg I**, Amar L, Brummett CM, Coleman DM, d'Escamard V, de Leeuw P, Fendrikova-Mahlay N, Kadian-Dodov D, Li JZ, Lorthioir A, Pappaccogli M, Prejbisz A, Smigielski W, Stanley JC, Zawistowski M, Zhou X, Zöllner S; FEIRI investigators; International Stroke Genetics Consortium (ISGC) Intracranial Aneurysm

Working Group; MEGASTROKE, Amouyel P, De Buyzere ML, Debette S, Dobrowolski P, Drygas W, Gornik HL, Olin JW, Piwonski J, Rietzschel ER, Ruigrok YM, Vikkula M, Warchol Celinska E, Januszewicz A, Kullo IJ, Azizi M; ARCADIA Investigators, Jeunemaitre X, Persu A, Kovacic JC, Ganesh SK, Bouatia-Naji N. **Nature Communication**. <https://doi.org/10.1038/s41467-021-26174-2>.

➤ **Poster**

**Etude d'association entre les traits liés à la physiologie des femmes et la dysplasie fibromusculaire.** Ines Sayoud-Sadeg, Takiy Berrandou, Nabila Bouatia-Naji. 11èmes Assises de Génétiques Humaines et Médicale. Février 2022.

## Liste des figures et tableaux

<b>Figure 1.</b> Nombre de décès annuels mondiaux par principales maladies non transmissible en 2018. .	23
<b>Figure 2.</b> Variation en pourcentage du taux de mortalité par MCV standardisé selon l'âge de 2010 à 2019.....	24
<b>Figure 3.</b> Incidence des maladies cardiovasculaires chez les hommes et les femmes en fonction de l'âge. ....	33
<b>Figure 4.</b> Niveaux des hormones féminines dans les différentes étapes de la vie de de la femme.....	36
<b>Figure 5.</b> Schéma de principe de l'influence des hormones sexuelles sur les maladies cardiovasculaires (MCV) .....	38
<b>Figure 6.</b> Angiographie par cathéter des artères rénales montrant. ....	40
<b>Figure 7.</b> Différents outils d'imagerie pour visualiser la dysplasie fibromusculaire multi-focale (DFM). ....	43
<b>Figure 8.</b> Caractéristiques des maladies monogéniques (mendéliennes) et des maladies polygéniques (complexes).....	49
<b>Figure 9.</b> L'étude d'association à l'échelle du génome est généralement basée sur une conception cas-témoin dans laquelle les SNPs du génome humain sont génotypés. ....	57
<b>Figure 10.</b> Jalons de la recherche en génomique cardiovasculaire de 2007 à 2017 et au-delà.....	59
<b>Figure 11.</b> Étapes d'analyse du score de risque polygénique (PRS) montrant les données de base, les données cibles et différentes méthodes en regroupant différentes approches. ....	65
<b>Figure 12.</b> Graphe acyclique dirigé (en anglais, <i>directed acyclic graph</i> ou DAG) .....	72
<b>Figure 13.</b> DAG partiel représentant une violation de la condition (A2). ....	73
<b>Figure 14.</b> Exemple d'une violation de la condition (A3). ....	74
<b>Figure 15.</b> Autre violation de la c (A3).....	74
<b>Figure 16.</b> Exemple de graphe d'intensité de signal d'allèle.....	89
<b>Figure 17.</b> Imputation d'un génotype d'échantillon manquant à l'aide d'un panel d'haplotypes de référence .....	95
<b>Figure 18.</b> Structure des chromosomes sexuels masculins humains et localisation des deux régions RPA1 (PAR1 en anglais) et RPA2 (PAR2 en anglais).....	98

<b>Figure 19.</b> Processus de l'inactivation du chromosome X et la formation du crépuscule de Barr .....	101
<b>Figure 20.</b> ACP du GWAS de la DFM avec les populations de 1000 génomes Phases 3.....	106
<b>Figure 21.</b> QQ plot et Manhattan plot de l'étude d'association de la DFM pour la cohorte ARCADIA-3C en utilisant le <b>modèle 1</b> .....	110
<b>Figure 22.</b> QQ plot et Manhattan plot de l'étude d'association de la DFM pour la cohorte ARCADIA-POL en utilisant le <b>modèle 1</b> .....	112
<b>Figure 23.</b> QQ plot et Manhattan plot de l'étude d'association de la DFM pour la cohorte FEIRI en utilisant le <b>modèle 1</b> .....	114
<b>Figure 24.</b> QQ plot et Manhattan plot de l'étude d'association de la DFM pour la cohorte ARCADIA-3C en utilisant le <b>modèle 2</b> .....	116
<b>Figure 25.</b> QQ plot et Manhattan plot de l'étude d'association de la DFM pour la cohorte ARCADIA-POL en utilisant le <b>modèle 2</b> .....	118
<b>Figure 26.</b> QQ plot et Manhattan plot de l'étude d'association de la DFM pour la cohorte FEIRI en utilisant le <b>modèle 2</b> .....	119
<b>Figure 27.</b> QQ plot et Manhattan plot de l'approche « pooled-analysis » de la DFM avec le <b>modèle 2</b> .....	121
<b>Figure 28.</b> QQ plot et Manhattan plot de la méta-analyse du Chr X des six études de la DFM avec le <b>modèle 2</b> .....	123
<b>Figure 29.</b> QQ plot et Manhattan plot de la méta-analyse du Chr X des six études de la DFM stratifiée uniquement sur les femmes avec le <b>modèle 2</b> .....	124
<b>Figure 30.</b> Manhattan plot de la méta-analyse des huit études de la SCAD avec le <b>modèle 2</b> .....	129
<b>Figure 31.</b> Schéma des différents types de pléiotropie.....	142
<b>Figure 32.</b> Forest plot de la corrélation génétique entre la DFM et les traits liés à la physiologie de la femme.....	146
<b>Figure 33.</b> Forest plot de la randomisation mendélienne entre la DFM et les traits liés à la physiologie de la femme.....	147

## Liste des tableaux

<b>Tableau 1.</b> Les différents facteurs de risque cardiovasculaire identifiés [13]. .....	29
<b>Tableau 2.</b> Description des différentes approches utilisées pour le calcul du score de risque polygénique (PRS) .....	67
<b>Tableau 3.</b> Comparaison des différentes approches pour effectuer des analyses PRS .....	68
<b>Tableau 4.</b> Caractéristiques cliniques des populations étudiées [60]. .....	83
<b>Tableau 5.</b> Informations détaillées pour les différentes étapes, génotypages, QC, QC-post imputation pour chaque étude DFM.....	92
<b>Tableau 6.</b> Informations détaillées pour les différentes étapes, génotypages, QC, QC-post imputation pour chaque étude de SCAD. ....	127
<b>Tableau 7.</b> Données descriptives des cohortes individuelles de la DFM. ....	134
<b>Tableau 8.</b> Nombre d'individus et le nombre de SNP pour chaque trait inclus dans la construction des scores polygéniques. ....	136
<b>Tableau 9.</b> Résultats d'association entre la DFM et les différents traits en utilisant l'approche du score polygénique (PRS).....	145
<b>Tableau 10.</b> Résultats de la corrélation génétique entre la DFM femmes uniquement et les trois traits : âge de puberté, âge de ménopause et la durée d'exposition aux hormones.....	146
<b>Tableau 11.</b> Résultats de l'analyse de la randomisation mendélienne (RM) entre la DFM, considérée comme phénotype d'issue et les différents traits, considérés comme des facteurs de risque.....	148

## Liste des abreviations

**ACP** : Analyse en Composantes Principales

**ADN** : Acide DésoxyriboseNucléique

**ARN** : Acide RiboNucléique

**AVC** : Accident Vasculaire Cérébral

**CHD** : Maladies Coronariennes

**CPI** : Cardiopathie Ischémique

**CR** : Callrate

**DFM** : Dysplasie Fibromusculaire

**FHS**: Framingham Heart Study

**GBD**: Global Burden of Disease

**GWAS**: genome-wide association studies

**HDL**: High Density Lipoprotein

**HDL-C**: High-Density Lipoprotein Cholesterol

**HTA** : Hypertension Artérielle

**IC** : Intervalle de confiance

**Chr X** : Chromosome X

**DZ** : jumeaux dizygotes

**HWE** : Équilibre de Hardy-Weinberg

**ICX** : Inactivation du chromosome X

**IM** : Infarctus du myocarde

**IMC** : Indice de Masse Corporelle

**LDL-C** : Low-Density Lipoprotein Cholesterol

**LH** : hormone lutéinisante

**MCV** : Maladie Cardiovasculaire

**MZ** : jumeaux monozygotes

**OMS** : Organisation Mondiale de la Santé

**OR** : Odds Ratio

**PAD** : Pression Artérielle Diastolique

**PSA** : Pression Artérielle Systolique

**$p_T$**  : p-value thresholding

**QC** : contrôle qualité

**RM** : Randomisation Mendélienne

**R<sub>g</sub>** : corrélation génétique

**SCAD** : Dissection Spontanée de l'artère coronaire

**SNP** : Single-nucleotide polymorphism

**TA** : Tension artérielle

**T<sub>six</sub>** : transcrit anti-sens

**X<sub>ic</sub>** : centre d'inactivation de l'X

**X<sub>m</sub>** : X maternel

**X<sub>p</sub>** : X paternel

## Résumé

### **Etude du rôle de facteurs génétiques liés à la physiologie des femmes dans la prédisposition à la dysplasie fibromusculaire (DFM) et à la dissection spontanée de l'artère coronaire (SCAD), artériopathies à forte prédominance féminine.**

La dysplasie fibromusculaire (DFM) et la dissection spontanée de l'artère coronaire (SCAD) sont deux artériopathies associées à l'hypertension artérielle sévère, aux accidents vasculaires cérébraux et à l'infarctus du myocarde. Elles présentent plusieurs caractéristiques cliniques similaires, comme l'absence de plaque d'athérome et d'obésité. Elles touchent majoritairement des femmes d'âge moyen (>80% des cas). La complexité du diagnostic différentiel, nécessitant une imagerie semi-invasive, limite l'accès à de grandes populations de malades et la réalisation d'études épidémiologique prospectives pour comprendre la forte proportion de femmes parmi les patients de ces deux artériopathies à modèle génétique polygénique et complexe.

L'objectif de cette thèse était d'étudier le rôle des facteurs génétiques liés à la physiologie des femmes dans la prédisposition à la DFM et à la SCAD. Pour cela, j'ai étudié le rôle des variations génétiques fréquentes (SNPs) du Chr X, et exploré les liens génétiques potentiels avec les déterminants génétiques de traits physiologiques liés aux caractéristiques de la reproduction féminine.

J'ai d'abord réalisé une étude d'association pangénomique spécifiquement sur le Chr X. J'ai utilisé les données génotypiques individuelles de 3 études cas-témoin européennes (1115 patients DFM et 2397 témoins) et d'une étude française (313 patients SCAD et 1478 témoins). J'ai appliqué deux modèles analytiques qui tiennent en compte l'inactivation du X chez les femmes : l'approche XWAS qui permet de tester toutes les variations génétiques simultanément avec un génotype masculin codé 0..1 (équivalent d'un hétérozygote féminin) et 0..2 (équivalent d'un homozygote féminin) ; et le modèle implémenté dans PLINK v2.0, sous l'hypothèse d'inactivation de X (les hommes et les femmes codés en 0..2). J'ai obtenu des résultats comparables par les deux méthodes. J'ai réalisé deux méta-analyses des données du Chr X incluant les résumés statistiques pour six études GWAS de la DFM (1962 cas et 7100 témoins) et huit pour la SCAD (1917 cas et 9292 témoins). Cette dernière a révélé une association significative pour le locus SRSF2P1/TMEM47 (OR=1.36,  $p=3 \times 10^{-8}$ ).

Dans la deuxième partie, j'ai étudié les liens génétiques potentiels entre le risque de DFM et les déterminants génétiques de l'âge de puberté, de l'âge de ménopause, ainsi que ceux de la période reproductive de la vie des femmes définie comme la différence entre les deux âges. J'ai exploité des résultats GWAS issus de l'étude UK Biobank (N=109 758). La méthode du LD score n'a montré aucune corrélation génétique significative entre la DFM et les trois traits. L'utilisation de 42, 28 et 25 SNPs associés au seuil génomique respectivement avec l'âge de puberté, l'âge de ménopause et la période reproductive n'ont montré aucune association significative avec la DFM par l'approche de randomisation mendélienne. Afin de compenser le faible nombre de variants dans cette analyse, j'ai calculé des scores génétiques à partir de 9 seuils de significativité obtenus dans l'étude UK Biobank (de  $P < 5 \times 10^{-8}$  : entre 151 et 196 SNPs à  $P < 0.50$  : entre 293 799 et 298 857 SNPs selon, les traits). J'ai testé les associations des scores construits dans les trois études européennes (1115 cas et 2397 témoins). Cette approche a confirmé l'absence d'association entre les déterminants génétiques des étapes majeures de la vie reproductive des femmes et la DFM.

En conclusion, mon travail de thèse a permis de décrire pour la première fois un nouveau locus de prédisposition au risque de la SCAD sur le Chr X. L'analyse d'échantillons plus grands sera nécessaire pour évaluer l'apport de ce chromosome sexuel dans le risque de DFM et de SCAD. La collecte des données sur les traits liés à la reproduction directement chez les patientes pourraient apporter des réponses plus précises sur le lien entre la génétique de ces traits et le risque d'artériopathies prédominantes chez les femmes comme la DFM et la SCAD.

**Mots clés :** DFM, SCAD, XWAS, scores génétiques, corrélation génétique, randomisation mendélienne.

## Abstract

### **Study of the role of genetic factors related to female physiology in the predisposition to fibromuscular dysplasia (FMD) and spontaneous coronary artery dissection (SCAD), arteriopathies with high female predominance.**

Fibromuscular dysplasia (FMD) and spontaneous coronary artery dissection (SCAD) are two arterial diseases associated with hypertension, stroke and myocardial infarction. Both diseases present several similar clinical features, mainly the absence of atherosclerosis and obesity. Patients are predominantly (>80%) middle-aged women. The complexity of the differential diagnosis requiring semi-invasive imaging limits access to large populations of patients and prospective epidemiological studies to understand the high proportion of women among the patients for these arteriopathies with a complex genetic pattern of inheritance.

The aim of this thesis was to investigate the role of genetic factors related to female physiology in the genetic risk for FMD and SCAD. To achieve this goal, I investigated the role of common genetic variations (SNPs) in the X chromosome (ChrX), and explored potential links between genetic determinants of physiological traits related to female reproductive characteristics and FMD risk.

I first performed a genome-wide association study specifically for SNPs on ChrX. I used individual genotyping data from 3 European case-control studies (1115 FMD patients and 2397 controls) and one French study (313 SCAD patients and 1478 controls). I applied 2 analytical models that consider X inactivation in females: the XWAS approach that allows testing all SNPs simultaneously with a male genotype coded 0.1 (equivalent to a female heterozygote) and 0.2 (equivalent to a female homozygote); and the model implemented in PLINK v2.0, under the assumption of X inactivation (males and females coded as 0.2) I obtained comparable results by both methods. I then performed 2 meta-analyses including summary statistics for 6 ChrX GWAS for FMD (1962 cases and 7100 controls) and 8 for SCAD (1917 cases and 9292 controls). The latter revealed a significant association for the *SRSF2P1/TMEM47* locus (OR=1.36, P=3.0×10<sup>-8</sup>).

In the second part of my thesis, I investigated the potential genetic links with the risk of FMD and the genetic determinants of age at menarche, age at menopause, as well as the reproductive period defined as the difference between the 2 ages using GWAS results from the UK Biobank study (N=109,758). The LD score method showed no significant genetic correlation between FMD and the 3 traits. The use of 42, 28 and 25 SNPs as instrumental variables for age at puberty, age at menopause and reproductive period respectively showed no significant association with FMD using the Mendelian randomization approach. To compensate for the low number of variants in this analysis, I calculated genetic scores based on 9 significance thresholds obtained in the *UK Biobank* GWAS (from P<5×10<sup>-8</sup>: 151 - 196 SNPs; to P<0.50: 293,799 - 298,857 SNPs according to traits). I tested the associations of the scores constructed in the 3 European studies (1115 cases and 2397 controls). This approach confirmed the lack of association between genetic determinants of major reproductive life stages of women and FMD.

In conclusion, my thesis work has described for the first time a new risk susceptibility locus on ChrX for SCAD. Analysis of larger samples will be required to assess the contribution of this major sex chromosome to the risk of FMD and SCAD. Collection of data on reproductive traits directly from patients could provide more precise answers to the link between the genetics of these traits and the risk of female-predominant arterial diseases such as FMD and SCAD.

**Keywords:** FMD, SCAD, XWAS, genetic scores, genetic correlation, Mendelian randomization.

# **Partie I. Introduction générale**

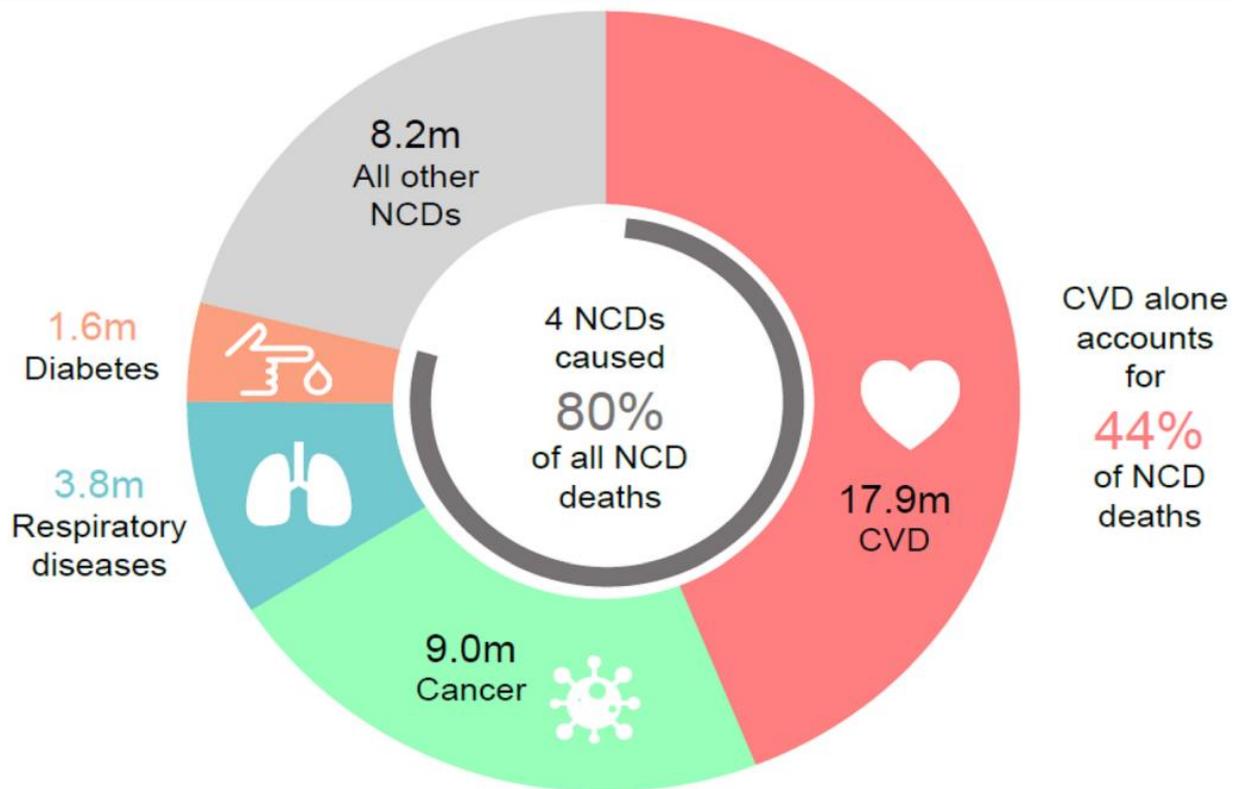
## **Chapitre 1.**

### **Maladies cardiovasculaires**

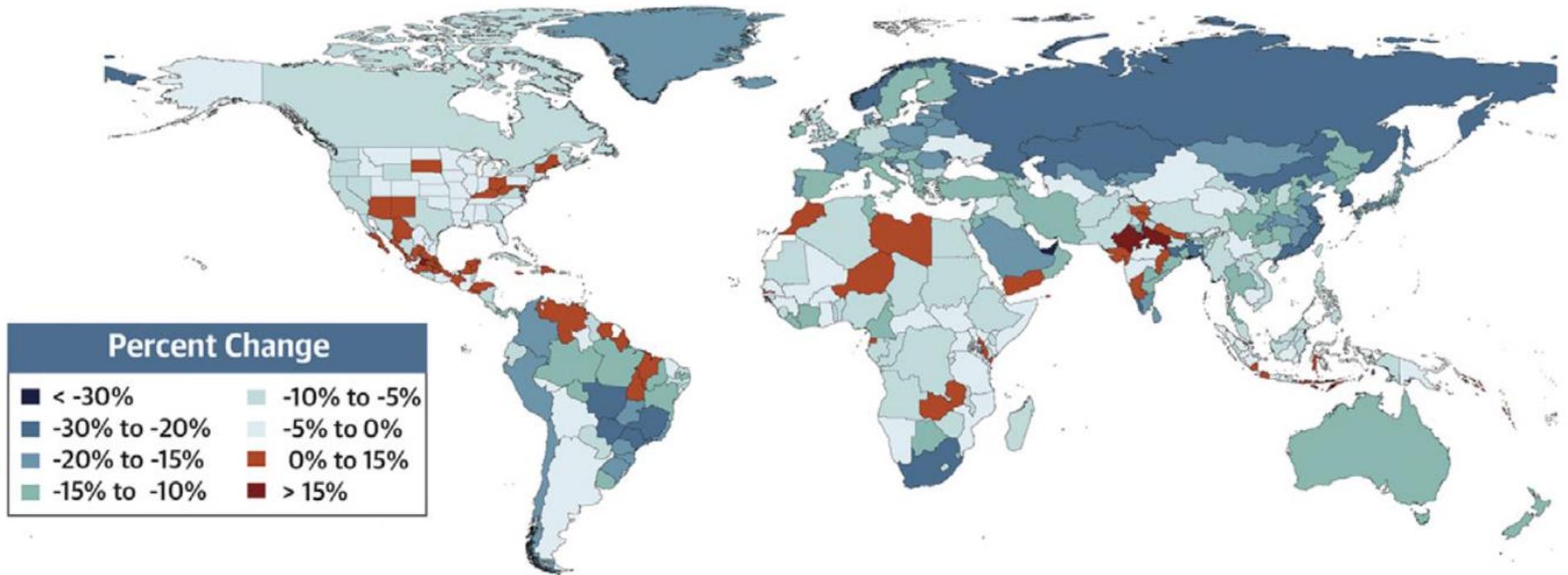
# 1. Généralités sur les maladies cardiovasculaires

## a. Introduction aux maladies cardiovasculaires

Avec plus de 17,9 millions de décès en 2018, les maladies cardiovasculaires (MCV) sont la principale cause de mortalité dans le monde selon les estimations de *Global Burden of Disease* (GBD), une étude épidémiologique observationnelle mondiale (Figure 1) [1], [2]. La Figure 2 montre les variations des taux de morbidité et de mortalité par MCV dans le monde : le nombre total de cas prévalents de MCV a presque doublé, passant de 271 millions en 1990 à 523 millions en 2019, soit +94% en 29 ans [1], et le nombre de décès par MCV a régulièrement augmenté de 12,1 millions en 1990 à 18,6 millions en 2019, soit +54% en 29 ans [1]. Selon les estimations données par l'Organisation Mondiale de la Santé (OMS), les MCV représentent 32% de tous les décès dans le monde [3]. Parmi ces décès, 85 % étaient dus à un événement cardiaque ou à un accident vasculaire cérébral [3]. Sur les 17 millions de décès prématurés (moins de 70 ans) par maladies non transmissibles en 2019, 38 % ont été causés par des MCV [3]. De même, en Europe, les MCV restent la cause de décès la plus fréquente [4]–[6]. Des études antérieures ont rapporté que les maladies cardiovasculaires tuent près de 4 millions de personnes en Europe chaque année, soit environ 44 % de tous les décès, les cardiopathies ischémiques (CPI) et les accidents vasculaires cérébraux (AVC) représentant respectivement 44 % et 25 % de décès par MCV [6].



**Figure 1.** Nombre de décès annuels mondiaux par principales maladies non transmissibles en 2018 selon l'OMS [7].



**Figure 2.** Variation en pourcentage du taux de mortalité par MCV standardisé selon l'âge de 2010 à 2019.

## **b. Les principaux groupes de maladies cardiovasculaires**

Les MCV sont un groupe de maladies affectant le cœur et l'ensemble des vaisseaux sanguins. Elles regroupent notamment les maladies coronariennes, les maladies cérébro-vasculaires, les artériopathies périphériques, et l'insuffisance cardiaque. Dans cette section, nous aborderons en détails les différents types de MCV, ainsi que leur diagnostic et le traitement approprié.

### **i. Cardiopathie coronarienne**

La cardiopathie ischémique appelée aussi maladie coronarienne est un type de maladie cardiaque où les artères du cœur ne peuvent pas fournir suffisamment de sang riche en oxygène au muscle cardiaque [8]. Parmi ses symptômes, citons : difficultés à respirer, gêne et douleurs dans la poitrine (angor), fatigue extrême à l'effort. Elle peut être fugace ou vive et ressentie dans l'abdomen, dans le dos ou dans le bras. La cardiopathie ischémique recouvre différentes entités cliniques : l'infarctus du myocarde, l'angor instable, l'angor stable et l'ischémie myocardique silencieuse. Elle représente la deuxième cause de mortalité en France après les cancers.

Diagnostic : Le diagnostic de maladie coronarienne se fonde sur deux types d'examens.

*Les examens anatomiques* visent à observer le rétrécissement artériel. Le coro-scanner (scanner des artères) est aujourd'hui utilisé en routine pour le dépistage de la maladie coronaire chez les patients symptomatiques présentant peu ou pas de facteurs de risque cardiovasculaire. L'examen dure une quinzaine de minutes, installation comprise, et permet de visualiser de façon exhaustive le réseau artériel coronarien. La coronarographie est une radiographie avec opacification des artères coronaires, plus invasive que le coro-scanner, car elle nécessite de ponctionner une artère pour monter un cathéter permettant de parfaitement visualiser l'ensemble des artères coronaires : placées en couronne autour du cœur, celles-ci lui apportent le sang nécessaire à son fonctionnement. Néanmoins, elle reste l'examen le plus fiable pour apprécier le degré de rétrécissement des artères coronaires [9].

*Les examens fonctionnels* permettent d'estimer l'impact du rétrécissement artériel sur le fonctionnement du cœur : l'échographie d'effort, celle-ci consiste à faire pédaler le patient tout en réalisant une échographie cardiaque et un électrocardiogramme, l'imagerie par résonance magnétique (IRM) de stress ou encore une scintigraphie myocardique pourront être proposés.

Traitement : La prise en charge médicale actuelle de l'ischémie comprend plusieurs traitements, tels que :

*Les bêta-bloquants* (ils ralentissent le rythme cardiaque et diminuent la pression artérielle, réduisant ainsi les besoins du cœur en oxygène).

La trinitrine (qui contrôle l'angor par dilatation des vaisseaux sanguins, et améliore ainsi le flux sanguin).

Les inhibiteurs calciques (ils relâchent les muscles de la paroi des artères coronaires ouvrant les vaisseaux et augmentant ainsi le flux sanguin autour du cœur).

## **ii. Accident vasculaire cérébral**

L'Accident Vasculaire Cérébral (AVC) est un déficit neurologique soudain, d'origine vasculaire, causé par une interruption brutale du flux sanguin cérébral par vaisseau sanguin bouché (AVC ischémique) ou rompu (AVC hémorragique). L'accident ischémique est la cause la plus fréquente des AVC avec 80% des cas. Il touche chaque année environ 150 000 nouveaux patients en France, soit 1 toutes les 4 minutes. L'AVC est l'une des premières causes de décès en France, dont le taux est plus élevé chez les femmes que chez les hommes [10].

Diagnostic : Un AVC est évoqué devant des symptômes neurologiques d'apparition soudaine (paralysie, perte visuelle, difficulté à s'exprimer, etc.). Le diagnostic d'AVC peut être confirmé par l'imagerie. Deux techniques sont utilisées : le scanner permet de définir le type d'AVC en cause (ischémique ou hémorragique). Il peut être complété d'un examen par imagerie par résonance magnétique (IRM) en cas d'AVC de petite taille. Lors d'un accident ischémique, d'autres examens (angiographie, échographie) sont utilisés pour déterminer l'origine du caillot ou de la plaque d'athérome responsable de l'obstruction du vaisseau

Traitement : Le traitement de l'AVC diffère selon le type. Dans le cas d'un AVC ischémique, le but est de déboucher le vaisseau cérébral le plus rapidement possible afin de limiter les lésions, et donc les séquelles. Des médicaments ou un geste invasif peuvent être proposés pour dissoudre ou retirer le caillot qui obstrue le vaisseau sanguin si le patient est vu dans les heures suivant l'AVC (< 4h30 pour une fibrinolyse, > 6h pour une thrombectomie). Après un AVC hémorragique, le traitement peut être chirurgical, pour vider la poche de sang qui comprime une partie du cerveau. Des traitements à long terme peuvent également être prescrits pour prévenir des rechutes, par exemple, des anticoagulants. La rééducation est un élément essentiel de la prise en charge de l'AVC qui a prouvé son efficacité, même dans le cas d'AVC sévère.

### **iii. Maladie artérielle périphérique**

La maladie artérielle périphérique dite aussi maladie vasculaire périphérique, est généralement une manifestation de l'athérosclérose ; elle se présente sous forme de sténose ou d'occlusion des artères des membres inférieurs. Elle provoque une diminution du flux sanguin vers les artères du tronc, des bras et des jambes. Les maladies artérielles périphériques peuvent être : 1) *l'artériopathie périphérique oblitérante* : c'est une maladie artérielle occlusive chronique des membres inférieurs causée par l'athérosclérose [11]. La dysplasie fibromusculaire artérielle [12], est un autre exemple d'artériopathie périphérique oblitérante ; 2) *l'artériopathie périphérique fonctionnelle* est une diminution du flux sanguin parce que les artères ne fonctionnent pas correctement. En général, la dysfonction entraîne une contraction soudaine et anormale des muscles (spasmes) dans les parois des vaisseaux sanguins. Le spasme provoque un rétrécissement temporaire qui réduit le flux sanguin. Plus rarement, la maladie est due à une relaxation anormale des muscles dans les parois des vaisseaux sanguins, provoquant un élargissement des artères.

Diagnostic : La claudication intermittente est la manifestation typique de la maladie artérielle périphérique. La claudication intermittente est une sensation douloureuse, à type de crampes, d'inconfort ou de fatigue des jambes qui se produit pendant la marche et est soulagée par le repos. La douleur de claudication se produit habituellement dans les mollets, mais peut apparaître également dans les pieds, les cuisses, les hanches, les fesses ou, beaucoup plus rarement, les bras. Le diagnostic de cette maladie est confirmé par des examens complémentaires non invasifs: échographie-doppler des membres inférieurs avec mesure de

l'index de pression systolique, incluant l'exploration de l'aorte abdominale. De plus, un test de marche standardisé peut être justifié chez certains patients pour quantifier le degré de gêne fonctionnelle.

Traitement : Tous les patients doivent bénéficier d'une modification « agressive » des facteurs de risque pour soulager les symptômes de la maladie artérielle périphérique et pour prévenir son aggravation. L'exercice physique, 35 à 50 mn de tapis roulant ou de parcours de marche, en alternant des périodes de marche et de repos, 3 à 4 fois/semaine, est un aspect important du traitement mais sous utilisé. Les antiagrégants plaquettaires peuvent légèrement diminuer les symptômes et améliorer la distance de marche chez les patients présentant une maladie artérielle périphérique.

L'identification des facteurs de risque (FDR) est essentielle car elle conduit à une meilleure prévention des MCV. Éliminer ou réduire les facteurs de risque permet de réduire la probabilité de survenue des événements cardiovasculaires ou de leurs complications. Les recommandations en Europe et aux États-Unis [13], [14] ont identifié 3 groupes de facteurs de risque cardiovasculaire en fonction de la causalité (Tableau 1).

Lien de causalité	Facteur de risque identifié	Impact du facteur de risque
Facteurs de risque <b>majeurs</b>	<ul style="list-style-type: none"> <li>▪ Tabagisme</li> <li>▪ Hypertension artérielle</li> <li>▪ Élévation du cholestérol total</li> <li>▪ Élévation du LDL cholestérol</li> <li>▪ Diminution du HDL cholestérol</li> <li>▪ Diabète de type 2</li> <li>▪ Âge</li> </ul>	<ul style="list-style-type: none"> <li>▪ Effet multiplicateur du risque cardiovasculaire indépendamment des autres Facteurs de risque</li> </ul>
Facteurs de risque <b>prédisposants</b>	<ul style="list-style-type: none"> <li>▪ Obésité androïde</li> <li>▪ Sédentarité</li> <li>▪ Antécédents familiaux de maladie coronarienne précoce (H &lt; 55 ans, F &lt; 65 ans)</li> <li>▪ Origine géographique</li> <li>▪ Précarité</li> <li>▪ Ménopause</li> </ul>	<ul style="list-style-type: none"> <li>▪ Effet potentialisateur lorsqu'ils sont associés aux facteurs de risque majeurs</li> </ul>
Facteurs de risque <b>discutés</b>	<ul style="list-style-type: none"> <li>▪ Élévation des triglycérides</li> <li>▪ Lipoprotéines LDL petites et denses</li> <li>▪ Élévation de l'homocystéine</li> <li>▪ Élévation de la lipoprotéine A</li> <li>▪ Facteurs prothrombotiques (fibrinogène, inhibiteur de l'activateur du plasminogène)</li> <li>▪ Marqueurs de l'inflammation (CRP, IL6)</li> <li>▪ Ménopause</li> <li>▪ Facteurs infectieux (Chlamydia pneumoniae, Helicobacter pylori, cytomégalovirus)</li> </ul>	<ul style="list-style-type: none"> <li>▪ Associé à un risque augmenté de maladie cardiovasculaire (coronarienne et/ou vasculaire cérébrale) mais le degré d'imputabilité est méconnu</li> </ul>

**Tableau 1.** Les différents facteurs de risque cardiovasculaire identifiés [13].  
(H) = homme; F = femme; CRP = C reactive protein; IL-6 = interleukine 6; LDL = low density lipoproteins cholesterol; HDL= high density lipoproteins cholesterol

### **i. Facteurs de risque majeurs**

Ce sont l'hypertension artérielle, les dyslipidémies, le diabète et le tabagisme pour les facteurs modifiables ; l'âge et le sexe en matière de facteurs de risque non modifiables. On parle de facteur de risque majeur lorsqu'il lui est rattaché un effet multiplicateur du risque cardiovasculaire indépendamment des autres facteurs de risque.

#### **▪ Hypertension artérielle**

L'hypertension artérielle (HTA) touche à peu près un adulte sur trois dans le monde [15], [16]. C'est le facteur de risque le plus commun à toutes les MCV, la relation entre tension artérielle et risque accru de MCV est linéaire [16].

L'HTA est cliniquement définie par la pression artérielle c'est à dire la force exercée par le sang contre les parois des vaisseaux sanguins où il circule. Elle est exprimée par deux chiffres. La Pression Artérielle Systolique (PAS) est le chiffre le plus élevé, sa valeur représente la pression dans l'artère au moment où le cœur se contracte, tandis que la Pression Artérielle Diastolique, est le chiffre le plus faible (PAD) et la valeur de la pression dans l'artère lorsque le cœur est au repos entre deux contractions. Une augmentation de la PAS de 20 mmHg et de la PAD de 10 mmHg multiplie par deux le risque de décès par AVC ou par cardiopathie ischémiques [17]. Il a été mis en évidence que les événements cardiovasculaires se manifestent généralement 5 ans plus tôt chez les patients hypertendus que chez ceux qui ont une TA normale [18].

#### **▪ Dyslipidémie**

Parmi les anomalies des lipides circulants, le principal facteur de risque des MCV est l'élévation du LDL-cholestérol (Low Density Lipoprotein) et la diminution du HDL-cholestérol (High Density Lipoprotein).

En France, environ 25% des sujets de plus de 65 ans présentent une dyslipidémie. Les données actuelles semblent démontrer que le rapport cholestérol total/HDL-cholestérol est une meilleure mesure du risque que le taux de cholestérol total (tableau 1).

- **Diabète**

Les études épidémiologiques révèlent que le diabète accroît aussi bien le risque de survenue de MCV que le risque de décès par MCV [19]–[22]. Le diabète de type 2 est un état pathologique caractérisé par une glycémie élevée [23]. Le diabète de type 2, qui est souvent caractérisé par une résistance à l'insuline entraîne une hyperglycémie qui favorise le développement de l'athérosclérose (dépôt de plaque de lipides dans les artères du cœur) [28], [29].

- **Tabagisme**

Du point de vue épidémiologique, il est admis que le tabagisme qu'il soit passif ou actif accroît significativement le risque de survenue de MCV [24]. Le tabagisme passif augmente de 30% le risque de survenue de MCV, tandis que le tabagisme actif augmente de 80% le risque de survenue de MCV [24]. Sur le plan physiopathologique, c'est l'action de la nicotine, du monoxyde de carbone (CO) et des gaz oxydants de la cigarette qui peut expliquer la toxicité cardiovasculaire de la cigarette [24], [25].

- **Âge et sexe**

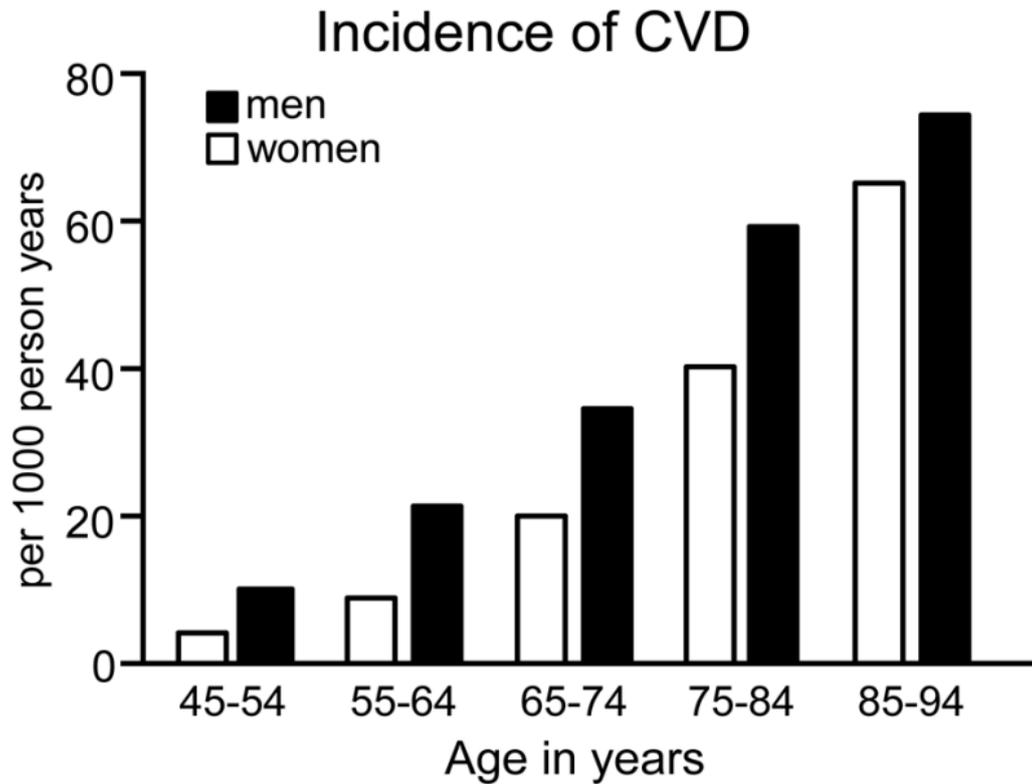
L'âge joue un rôle essentiel dans la détérioration de la fonctionnalité cardiovasculaire, entraînant un risque accru de MCV chez les personnes âgées [26], [27]. Il a également été démontré que la prévalence des maladies cardiovasculaires augmente avec l'âge, tant chez les hommes que chez les femmes, y compris la prévalence de l'athérosclérose, des AVC et de l'infarctus du myocarde (IM) [28]. L'*American Heart Association* (AHA) rapporte que cette prévalence des MCV chez les hommes et les femmes aux États-Unis est d'environ 40 % entre 40 et 59 ans, d'environ 75 % entre 60 et 79 ans et d'environ 86 % chez les personnes de plus de 80 ans [28]. De même, l'incidence des MCV, y compris les coronaropathies, les accidents vasculaires cérébraux ou hémorragie intracérébrale, passe de 4 à 10 pour 1 000 années-personnes chez les adultes âgés de 45 à 54 ans à 65-75 pour 1 000 années-personnes chez les adultes âgés de 85 à 94 ans [28].

## **ii. Facteurs globaux**

L'obésité, la sédentarité, l'alcool et les facteurs psychosociaux comme la précarité sont considérés comme des FDR prédisposants. Il a été démontré que ces FDR entraînent notamment une élévation de la TA, c'est le cas pour l'obésité et le surpoids [29], [30], les régimes riches en sodium, les régimes pauvres en potassium, la consommation excessive d'alcool [31] et la sédentarité [16].

## **c. Risque cardiovasculaire selon le sexe**

Les hommes et les femmes ont un risque cardiovasculaire similaire, mais les hommes développent une maladie cardiovasculaire plus tôt que les femmes [32], [33] (Figure 3). À 55 ans, les hommes ont un risque plus élevé de première maladie coronarienne que les femmes, mais les femmes ont un risque plus élevé de première maladie cérébrovasculaire ou d'insuffisance cardiaque que les hommes [32], [34]. C'est pour cette raison que les MCV surviennent et évoluent différemment chez les hommes et les femmes [34], [35]. En effet, les facteurs biologiques qui diffèrent entre les sexes ont des effets protecteurs et délétères spécifiques au sexe [34]. Les hormones gonadiques, en particulier les œstrogènes et les androgènes, expliquent depuis longtemps certaines différences entre les sexes dans les MCV [36], [37]. Les hormones sexuelles peuvent affecter le risque de MCV à la fois directement et indirectement, c'est-à-dire par le biais de divers facteurs de risque métaboliques [38]. L'estradiol par exemple contribue à l'accumulation de dépôts graisseux sous-cutanés chez les femmes [39], et la testostérone réduit la graisse corporelle en favorisant la prolifération des cellules musculaires chez les hommes [40]. Cependant, les hommes ont également tendance à accumuler de la graisse autour des organes internes sous forme de tissu adipeux viscéral [38]. Cette différence de distribution de masse grasse contribue au syndrome métabolique, un facteur de risque distinct majeur de maladie cardiovasculaire caractérisé par une obésité centrale, une résistance à l'insuline, une dyslipidémie et une hypertension artérielle.



**Figure 3.** Incidence des maladies cardiovasculaires chez les hommes et les femmes en fonction de l'âge. Données sur l'incidence des maladies cardiovasculaires définies comme les maladies coronariennes, l'insuffisance cardiaque, les accidents vasculaires cérébraux et la claudication intermittente chez les femmes (barres blanches) et les hommes (barres noires) de 45 à 94 ans [34].

## **2. Facteurs de risques de maladies cardiovasculaires liés à la physiologie de la femme**

### **a. Les étapes physiologiques critiques dans la vie de la femme**

Cette section est consacrée à définir les différentes étapes de la vie de la femme, ainsi que les différentes hormones.

#### **i. Puberté**

La puberté est définie comme la première période menstruelle chez une adolescente. L'âge de la puberté survient généralement entre 10 et 16 ans, l'âge moyen d'apparition étant de 12,4 ans [41]. Pendant cette période, le niveau des hormones féminines (estrogène et progestérone) commence à croître progressivement.

#### **ii. Ménopause**

La ménopause, définie comme la fin des menstruations due à la perte de l'activité folliculaire ovarienne, dont on sait qu'elle diminue à la fin de la trentaine avec une perte complète chez la plupart des femmes vers 50 ans [42], [43]. Cependant, il n'est pas toujours aisé de déterminer cet instant puisqu'une ou plusieurs menstruations irrégulières peuvent encore survenir après une période d'aménorrhée de quelques mois. Le diagnostic de ménopause est donc rétrospectif devant la constatation d'une aménorrhée d'une durée supérieure ou égale à 12 mois. Aucune intervention médicale ou d'ordre hygiéno-diététique ne semble susceptible de modifier l'âge de sa survenue. On parle de ménopause précoce avant 40 ans et tardive après 55 ans. Au cours de la pré-ménopause, la plupart des femmes commencent à ressentir des changements au niveau du cycle menstruel, premières bouffées de chaleur, irritabilité, fatigue, troubles de sommeil, ostéoporose.

#### **iii. Grossesse et allaitement**

L'organisme de la femme produit constamment des hormones pour réguler son fonctionnement. Pendant la grossesse, il augmente la production de certaines hormones spécifiques qu'il sécrétait déjà durant le cycle, comme la progestérone ou les œstrogènes. Ce sont ces modifications hormonales qui vont permettre la gestation, le développement du fœtus, la

préparation du corps à l'accouchement, le déclenchement du travail et l'allaitement..., mais aussi la protection contre certains risques comme un renforcement de la coagulation pour se protéger des hémorragies.

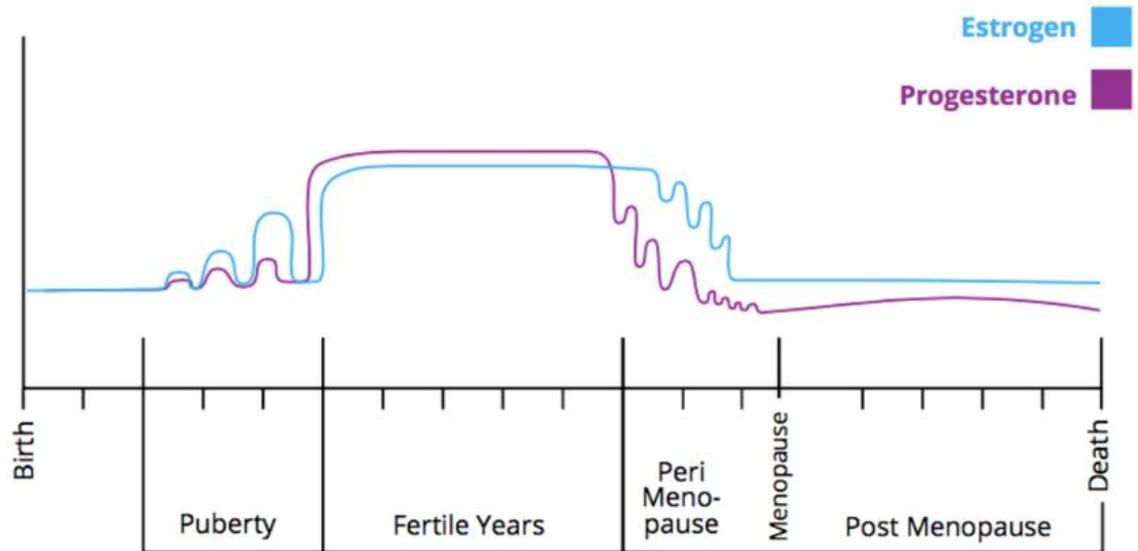
## **b. Fluctuations hormonales et risque de maladies cardiovasculaires**

### **i. Définition des hormones féminines**

Il est important de noter que les hormones féminines telles que l'estrogène et la progestérone contribuent dans le développement de chacune des étapes définies précédemment. L'**estrogène** garantit le bon fonctionnement du corps féminin pour l'ovulation et la grossesse. La **progestérone** quant à elle, est l'hormone qui prépare la nidation et l'implantation de l'ovule dans l'utérus lorsqu'il y a fécondation. Bien que considérée comme une hormone masculine, la **testostérone** est présente chez les femmes en quantité plus faible que chez les hommes. Elle reste néanmoins essentielle, car elle intervient dans de nombreux domaines tels que le développement musculaire, la masse osseuse et la mémoire.

La Figure 4 présente les niveaux d'estrogène et de progestérone au cours de chaque période de la vie de la femme, ainsi que leur rôle.

De nombreuses études ont montré qu'un niveau élevé d'estrogène protège le cœur et empêche la prolifération des cellules musculaires lisses [45], [46].

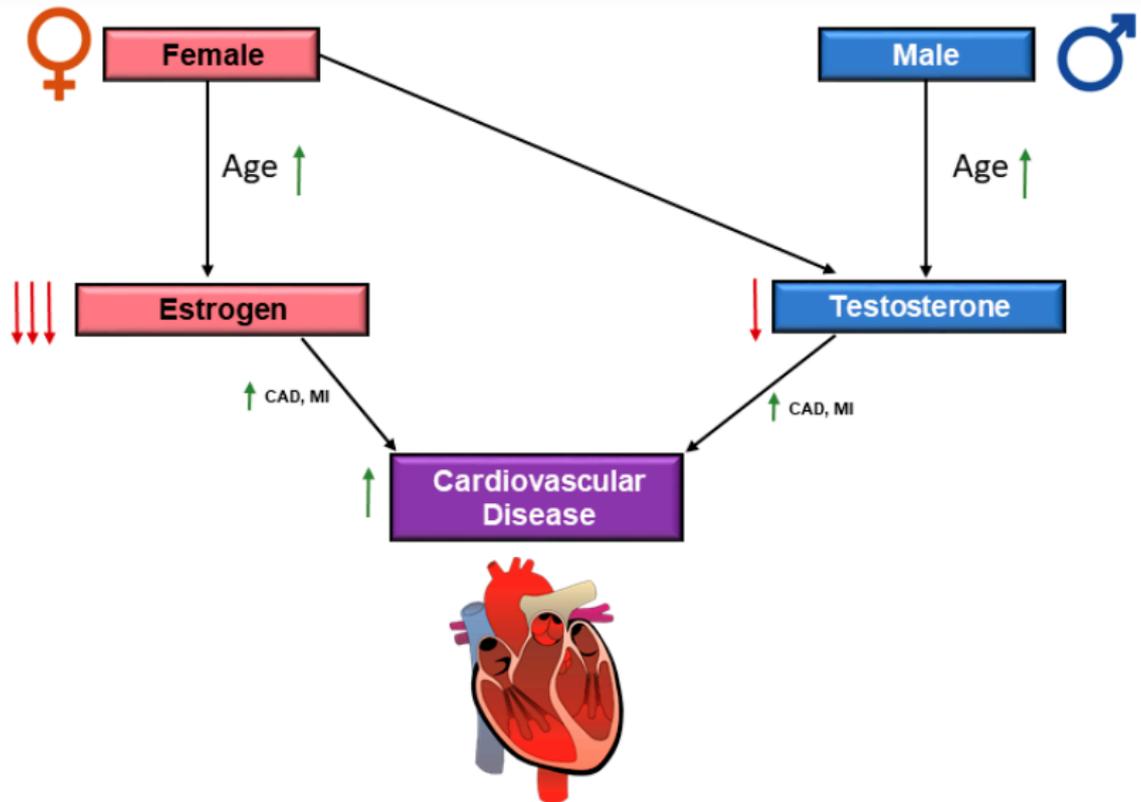


**Figure 4.** Niveaux des hormones féminines dans les différentes étapes de la vie de de la femme. Puberty : la période de la puberté de la femme. Fertile years : les années de de reproduction maximale d'une femme qui se situe entre la fin de l'adolescence et la fin de la vingtaine. Peri-menopause : est la période qui suit les toutes dernières règles. Menopause : l'arrêt définitif des règles. Post-menopause : la période d'absence des règles pendant plus d'un an.

## **ii. Risque cardiovasculaires lié aux étapes physiologiques chez la femme**

Comme nous avons vu précédemment dans la partie présentant les différents FDR liés au MCV, l'âge est un facteur de risque indépendant de MCV chez les hommes et chez les femmes. Les femmes âgées sont plus sujettes à certaines complications liées aux MCV [47]. Généralement, avant la ménopause, les femmes sont relativement protégées des maladies cardiovasculaires, puis, après la ménopause, le risque de maladie cardiaque augmente fortement. Il a été démontré que la diminution des hormones sexuelles joue un rôle important dans le développement des maladies cardiovasculaires avec l'apparition de l'âge avancé (Figure 5), tant chez les hommes que chez les femmes [48]. Les oestrogènes sont souvent reconnus pour leur rôle cardioprotecteur et seraient directement associés avec l'incidence globale plus faible des maladies cardiovasculaires chez les femmes pré-ménopausées, par rapport aux femmes appariées selon l'âge hommes [49], [50]. Une étude a rapporté que les hommes sont susceptibles de développer une maladie cardiaque 10 à 15 ans plus tôt que les femmes en raison de la baisse progressive des taux d'oestrogènes après la puberté [51]. À l'inverse, les hommes de 70 ans ont un risque cardiovasculaire global plus faible que les femmes de 50 ans, l'âge moyen de la ménopause chez les femmes [52], ce qui suggère que le déclin des estrogènes aurait un impact plus important sur le risque de MCV chez les femmes que chez les hommes.

D'autre part, de nombreuses études ont montré une association entre l'âge à la ménarche ou âge de puberté et le risque de MCV. La ménarche est une étape physiologique complexe de la reproduction qui influence la maturation ultérieure du corps féminin [53]. La ménarche a également été associée à des facteurs de risque de MCV [54]–[56]. La ménarche précoce a été associée à une adiposité plus importante, au syndrome métabolique et à un risque accru de cancer du sein, ce qui semble être dû, au moins en partie, à l'augmentation de l'exposition vie entière aux estrogènes [54]. L'âge précoce et tardif de la ménarche a été associé à des risques accrus de maladie coronarienne [57]. D'autres ont rapporté qu'un âge précoce à la ménarche est associé à un risque accru d'événements cardiovasculaires et de mortalité toutes causes confondues [58, p. 1976–88].



**Figure 5.** Schéma de principe de l'influence des hormones sexuelles sur les maladies cardiovasculaires (MCV) [59]. Homme et les femmes présentent toutes deux une baisse des hormones sexuelles, tandis que la baisse hormonale est plus drastique chez les femmes avec le début de la ménopause. Le déclin des hormones sexuelles est associé à une augmentation de risque cardiaque chez les hommes et les femmes, avec la baisse de la testostérone et des œstrogènes, respectivement. La flèche en vert indique régulation à la hausse, et la flèche en rouge indique une régulation à la baisse. Maladie de l'artère coronaire (CAD) et l'infarctus du myocarde (IM).

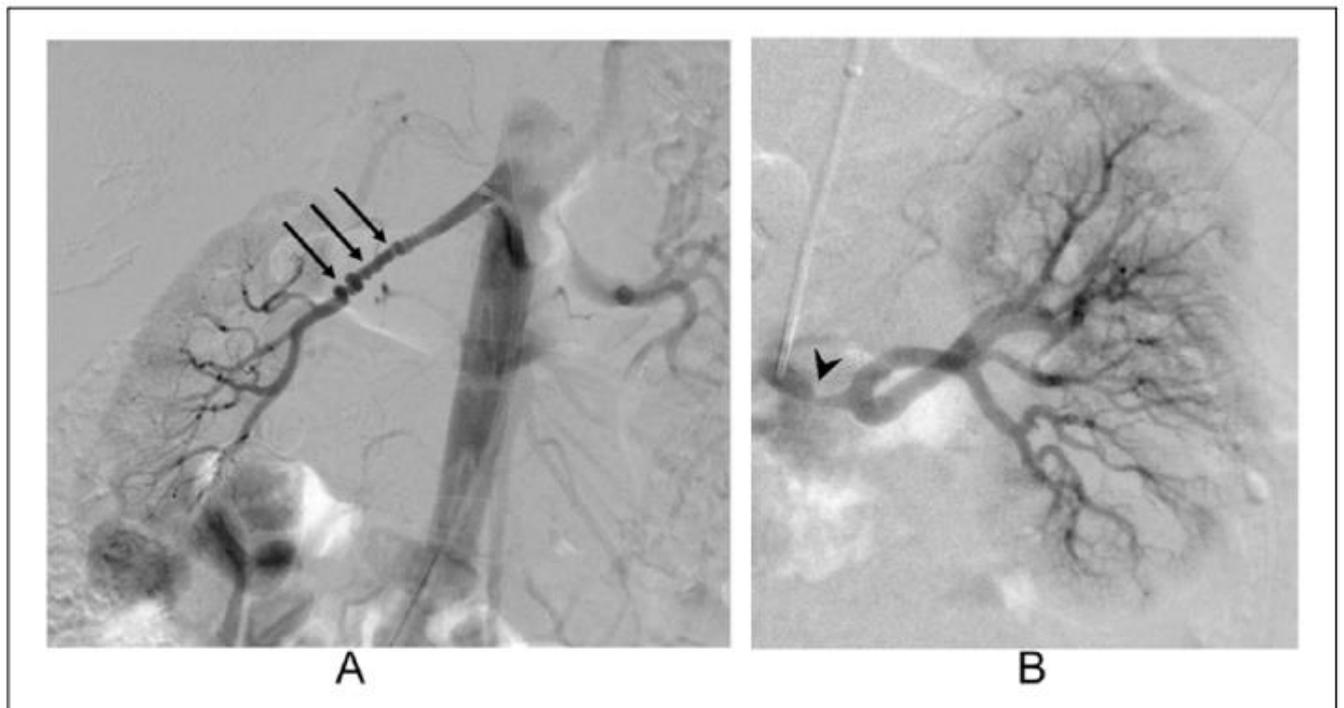
Malgré les connaissances acquises sur les risques de MCV chez les femmes, certaines d'entre elles échappent à ce raisonnement. Elles ne respectent pas forcément les FDR classiques des MCV, et touchent majoritairement des femmes jeunes (< 45 ans). C'est le cas de la dysplasie fibromusculaire artérielle et la dissection spontanée de l'artère coronaire, qui vont faire l'objet de mon travail de thèse, et qui seront définies dans les sections suivantes.

### **3. La dysplasie fibromusculaire artérielle**

#### **a. Définition clinique**

La Dysplasie Fibromusculaire Artérielle (DFM) est une maladie vasculaire non athéroscléreuse et non inflammatoire. Elle provoque un rétrécissement (sténose) des artères de moyen calibre, ce qui réduit le flux sanguin et affecte le fonctionnement des reins et du cœur [60]. La majorité des patients atteints de DFM sont des femmes (environ quatre femmes pour un homme) âgées de 40 à 55 ans. Cependant, le diagnostic de DFM peut être aussi posé chez les hommes [61]–[64].

Selon les experts de cette pathologie [65], il existe deux types de DFM : multi-focale et uni-focale. La forme multi-focale est la plus fréquente, et elle affecte 90 % ou plus des patients atteints de DFM. Elle se caractérise par l'apparence de l'artère en « collier de perles » (Figure 6A), et elle est formée par la succession de multiples sténoses et dilatations post-sténotiques, alors que la DFM uni-focale est beaucoup moins courante. Elle est caractérisée par un seul rétrécissement plus lisse dans une artère (Figure 6B), et elle est plus fréquente chez les enfants et les adolescents diagnostiqués d'une DFM [61].



**Figure 6.** Angiographie par cathéter des artères rénales montrant. Image A : un « collier de perles » (trois flèches) dans la dysplasie fibromusculaire multifocale (DFM). Image B un seul rétrécissement de la DFM focale (tête d'une seule flèche) [61].

## **b. Historique**

La première description remonte à 1938 par Leadbetter et Burkland [66]. Le terme de « dysplasie fibromusculaire » est employé pour la première fois en 1958. Les premières publications datent des années 1970-1980 [67]–[69], puis les publications ultérieures ont presque toutes été des cas uniques ou des rapports de petites séries de cas. Il y a eu très peu de nouveautés publiées sur la DFM au cours des 30 dernières années.

## **c. Diagnostique différentiel de la DFM**

La DFM ne provoque généralement aucun symptôme (forme asymptomatique) quelle que soit la localisation. Les symptômes, quand ils se produisent, varient en fonction de la localisation [70]. Les manifestations dépendent alors des organes irrigués par la (ou les) artère(s) atteinte(s), à savoir : le rein (environ deux cas sur trois), le cerveau (environ un cas sur trois), très rarement, l'intestin, les membres, et le cœur. La manifestation la plus fréquente de la DFM des artères rénales est l'HTA. Ce phénomène est très fréquent chez les adultes, chez des personnes qui n'ont pas de DFM, mais c'est aussi la principale conséquence des sténoses d'une artère rénale liées à la DFM. Les différentes manifestations de la DFM sont :

- L'infarctus rénal qui se manifeste par une douleur très violente et brutale dans la région lombaire qui ressemble à celle du blocage d'un calcul rénal (colique néphrétique) ; il y a souvent un peu de sang dans les urines (hématurie). La dissection d'une artère rénale dysplasique survient presque toujours sans raison particulière, très rarement elle peut être provoquée par un traumatisme.

- La DFM des artères du cou est souvent découverte par hasard, par exemple sur un examen radiologique demandé pour une autre raison. Elle peut aussi être découverte à l'occasion d'une complication neurologique (AVC) ou lors du bilan réalisé chez une personne ayant une DFM des artères rénales.

- La DFM peut se compliquer d'un AVC et être diagnostiquée à la suite de cet événement. Il s'agit d'une obstruction brutale par un caillot ou la déchirure (dissection) d'une zone artérielle touchée par la DFM qui fait qu'une partie du cerveau ne reçoit plus de sang (ischémie), provoquant un accident ischémique cérébral, aussi appelé AVC ischémique ou infarctus cérébral.

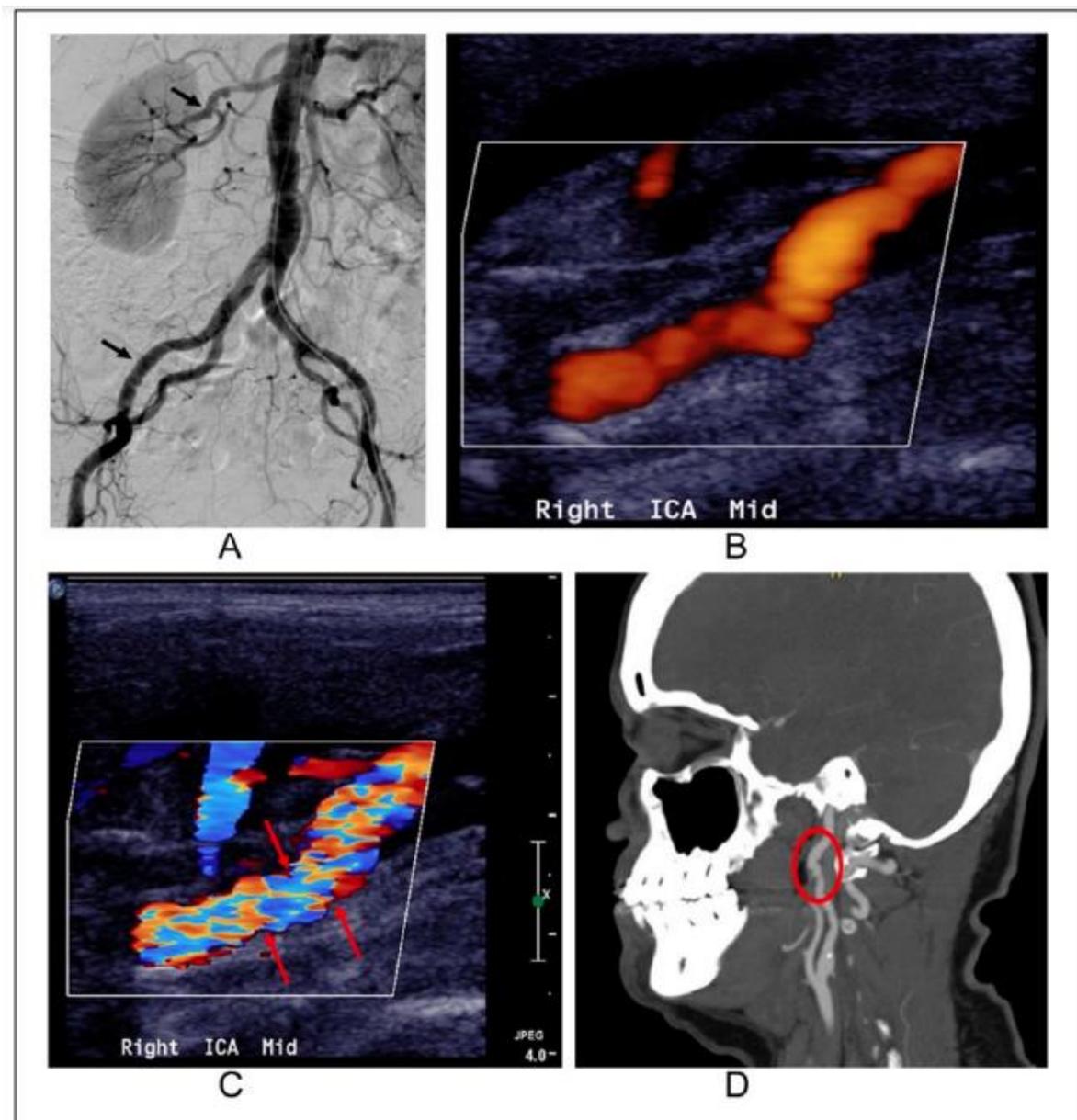
L'atteinte d'autres artères que celles des reins et du cou se manifeste par des douleurs après les repas (atteinte des artères digestives) ou des crampes musculaires à l'effort (artères iliaques). L'atteinte des artères du cœur (les coronaires) est exceptionnelle.

Le diagnostic de DFM peut parfois s'avérer difficile car d'autres affections provoquant des anomalies des artères peuvent ressembler à la DFM (comme l'athérosclérose ou la dissection d'une artère). Différentes techniques d'imagerie sont utilisées pour diagnostiquer la DFM (Figure 7) :

Échographie-Doppler ou écho-Doppler : C'est un examen indolore qui utilise les ultrasons. Le médecin passe une sonde sur la peau tout le long du trajet d'un vaisseau. La sonde émet des ultrasons et enregistre leur écho. L'appareil traduit ces échos en images et en sons. En cas de sténose, le flux sanguin est accéléré dans la zone rétrécie, ce qui se traduit par des sons et des images caractéristiques. Mais l'écho-Doppler ne permet pas d'affirmer que la sténose est due à une DFM.

Angio-scanner, angio-IRM : Ces examens permettent de reconnaître une sténose et en général de l'attribuer à la DFM. L'angio-scanner est un examen basé sur les rayons X, il nécessite l'injection d'un produit de contraste iodé et aussi d'un produit de contraste à base de gadolinium, sans iode ; elle est moins précise que l'angio-scanner pour examiner les artères rénales.

L'angiographie (angio-scanner) par cathéter (invasive) : Il reste l'examen principal qui permet de reconnaître une sténose et en général de l'attribuer à la DFM. Ce test consiste à créer des images radiographiques de haute qualité basées sur les rayons X, il nécessite l'injection d'un produit de contraste iodé des artères à travers des cathéters. Cette méthode n'est pas utilisée pour diagnostiquer DFM dans la pratique clinique de routine compte tenu de sa nature invasive, de son coût et de l'utilisation des rayonnements et colorant de contraste à base d'iode. Au lieu de cela, l'angiographie est utilisée pour diagnostiquer la DFM uniquement si d'autres études d'imagerie ont été pas clair ou si une intervention est nécessaire [61].



**Figure 7.** Différents outils d'imagerie pour visualiser la dysplasie fibromusculaire multi-focale (DFM). (A) Angiographie par cathéter montrant une "chaîne de perles" des artères rénales (flèche supérieure) et iliaques externes (flèche inférieure). (B) L'échographie (écho-Doppler) montre un bourrelet. (C) un flux turbulent (indiqué par des flèches rouges pointant vers une mosaïque de couleurs) dans l'artère carotide interne droite. (D) Angiographie par cathéter montre un bourrelet dans l'artère carotide interne (ovale rouge).

#### **d. Facteurs de risque de la DFM**

La physio-pathologie de la DFM n'est toujours pas connue avec précision, et la cause principale de développer une DFM n'est pas encore identifiée. Plusieurs hypothèses ont été émises sur certains facteurs qui peuvent être impliqués dans le risque de DFM, notamment les facteurs hormonaux, environnementaux et génétiques [71]–[73].

Parmi les facteurs environnementaux, le tabagisme a été identifié comme un facteur pathogène potentiel associé à la DFM [72]. Des études cas-témoins ont mis en évidence une association entre le tabagisme actuel (OR [2.5-4.05]) et passé (OR [1.8-4.1]) et la DFM rénale [71], [74], [75]. Dans le registre américain de la DFM, il a été rapporté que les patients atteints de DFM ayant des antécédents de tabagisme avaient un taux significativement plus élevé d'anévrismes que ceux qui n'avaient jamais fumé avec une tendance à l'augmentation de la prévalence des événements vasculaires majeurs chez les fumeurs [71], [76]. Malgré ces données, le tabagisme ne peut être considéré comme un facteur indispensable au développement d'une DFM [65], [76].

#### **e. Prédominance féminine dans la DFM**

La prédominance des cas chez les femmes a été considérée comme le principal facteur de risque de DFM. Il y a plus de 85 % de femmes avec une DFM multi-focale dans la cohorte française de la DFM [77], et 91 % dans le registre américain [78]. A l'inverse, il apparaît que les sujets présentant une DFM uni-focale sont majoritairement de sexe masculin. Cependant, la forme uni-focale est moins fréquente, les séries étudiées sont limitées à de petits effectifs [79]. L'exposition à des hormones féminines endogènes ou exogènes a également été associée à la DFM, mais les données disponibles ne permettent pas d'affirmer ces associations. En effet, bien que la maladie soit beaucoup plus répandue chez les femmes que chez les hommes, aucun lien de causalité clair n'a été identifié chez celles qui ont utilisé des contraceptifs oraux ou d'autres hormones féminines exogènes [75], [80], [81].

Une étude histologique cas-témoin a suggéré un équilibre anormal entre les récepteurs aux estrogènes et à la progestérone dans les échantillons d'artère rénale des patients ayant subi une intervention chirurgicale pour DFM rénale, caractérisé par une expression intense des récepteurs à la progestérone dans les noyaux des cellules musculaires lisses qui n'a pas été trouvé dans les échantillons des patients témoins [82]. Ces résultats préliminaires suggèrent que la progestérone

peut également jouer un rôle dans la pathogenèse de la DFM, mais cela doit être reproduit dans d'autres études.

## **4. La dissection spontanée de l'artère coronaire**

### **a. Définition clinique**

La dissection spontanée de l'artère coronaire ou SCAD (pour Spontaneous Coronary Artery Dissection) est un type de crise cardiaque causé par la déchirure (dissection) d'une ou de plusieurs artères du cœur [83]. La dissection sera soit entre les couches de la paroi artérielle, ce qui provoque une accumulation de sang (hématome), soit dans la couche interne (intima). Ces deux types de dissection peuvent bloquer la circulation sanguine et provoquer une ischémie cardiaque, et parfois la mort. La SCAD risque également d'entraîner des lésions cardiaques ou des anomalies du rythme cardiaque. Depuis la première description de SCAD par Pretty en 1931 lors d'une autopsie, la compréhension de celle-ci a énormément évolué au cours des dernières années. Sur la base de rapports de cas isolés et de petites séries, la SCAD a été initialement décrite comme une cause rare et presque universellement mortelle de syndrome coronarien aigu, d'IM et de mort cardiaque subite chez les femmes en péri-partum [84]. En effet, les progrès dans notre compréhension de l'épidémiologie de la SCAD, la disponibilité des techniques d'imagerie intravasculaire, le développement d'une classification angiographique spécifique à la SCAD, la sensibilisation accrue des prestataires et les efforts parallèles des patients utilisant les médias sociaux pour diffuser largement l'information suggèrent que la SCAD est beaucoup plus fréquent qu'on ne le pensait auparavant, en particulier chez les jeunes femmes.

### **b. Diagnostique différentiel de la SCAD**

Depuis 2010, le diagnostic de SCAD s'est rapidement élargi en raison de la reconnaissance accrue de l'utilisation de l'imagerie intra-coronaire à haute résolution, en particulier la tomographie par cohérence optique, pour identifier la SCAD [70]. La SCAD est maintenant de plus en plus acceptée comme une cause importante d'infarctus de myocarde chez les femmes jeunes d'âge moyen [70], [85]. Cependant, l'incidence et la prévalence réelles de la SCAD restent inconnues, en raison du sous-diagnostic et de l'erreur de diagnostic de cette maladie. Parmi les patients présentant un syndrome coronarien aigu, la SCAD a été signalée

comme étant la cause dans 1 % à 4 % des coronarographies et représentait 0,5 % des morts cardiaques subites dans les cas d'autopsie. Dans des séries récentes, la SCAD a été signalée comme étant la cause d'IM chez 25 % à 35 % des femmes âgées de moins de 50 ans et jusqu'à 25 % des femmes âgées de moins de 60 ans. C'est également la cause la plus fréquente d'infarctus du myocarde associé à la grossesse (jusqu'à 43 %), survenant principalement au cours du troisième trimestre ou de la phase post-partum [86]–[89].

### **c. Facteurs de risque de la SCAD**

La SCAD affecte principalement les femmes dans plus de 90 % des cas, dont 60 % étaient ménopausées [70], [89], et l'âge moyen des patientes atteintes de SCAD varie de 45 à 53 ans. La plupart des patientes présentent peu ou pas de facteurs de risques cardiovasculaires conventionnels ; cependant, la migraine ( $\approx 40\%$ ), la dépression ( $\approx 25\%$ ) et l'anxiété ( $\approx 15\%$ ) sont fréquemment observées chez des femmes atteintes de SCAD [90], [91].

Les patients atteints de SCAD présentent des taux moindres de facteurs de risque cardiovasculaire habituels [92]. Les facteurs de risque de SCAD incluent notamment le stress émotionnel et physique, la grossesse, et DFM. Bien que les patients atteints de SCAD aient des taux inférieurs de FDR cardiovasculaires traditionnels tels que l'HTA, l'hyperlipidémie et le tabagisme par rapport aux patients atteints d'athérosclérose [93]–[96], la prévalence de l'hypertension (32 % à 37 %) et de l'hyperlipidémie (20 % à 35 %) chez les patients SCAD est similaire à celle de la population générale de même âge et sexe [93], [96], [97].

Les recherches récentes ont montré que les patients ayant souffert d'une SCAD présentaient plus souvent des lésions de DFM (30 à 80% des cas) [86], en particulier au niveau des artères rénales, cervicales et iliaques. Les lésions atteignent souvent plusieurs lits vasculaires et sont quasi exclusivement de type multifocal. Cependant, bien que la prévalence de la DFM soit plus élevée chez les patients ayant présenté une SCAD que dans la population générale, la SCAD reste une complication rare chez les patients avec DFM.

## **Chapitre 2.**

### **Méthodes d'épidémiogénétique de détection des facteurs de risque génétiques aux maladies à génétique complexe**

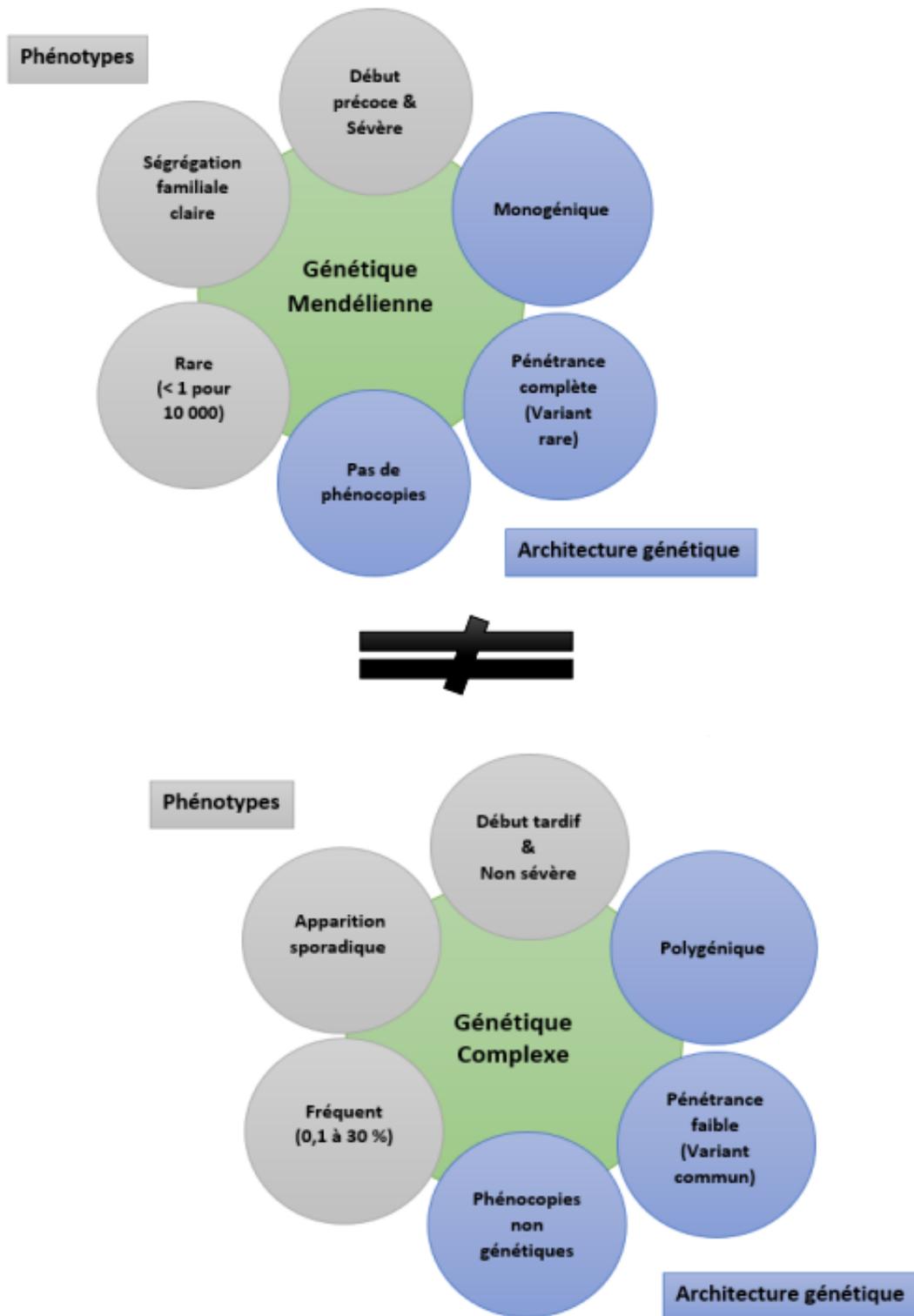
## 1. Introduction

Les maladies cardiovasculaires sont des maladies fréquentes, car leur incidence dans la population générale est élevée contrairement aux maladies rares. Les MCV sont considérées comme des maladies complexes appelées aussi maladies multifactorielles. Elles résultent de l'interaction de facteurs génétiques et environnementaux qui augmentent chacun d'eux, peu à peu, le risque de développer la maladie. Ces troubles génétiques complexes impliquent un grand nombre de maladies courantes, telles que le diabète de type 1 et de type 2, les maladies cardiovasculaires, certains cancers courants, les troubles neurologiques et psychiatriques, l'asthme et les maladies auto-immunes [98].

La définition des maladies complexes est à l'opposé de celle des maladies mendéliennes qui sont dues à une anomalie génétique n'impliquant qu'un seul gène, également appelées maladies monogéniques (voir Figure 8).

L'héritabilité de ce type de pathologie est liée à plusieurs facteurs génétiques, chacun d'entre eux n'est ni indispensable, ni suffisant, pour causer à lui seul la maladie. On parle donc plutôt de gènes de prédisposition qui résulte de l'action combinée (effets additifs, effets multiplicatifs) d'un grand nombre de gènes. L'identification de ces facteurs génétiques est d'une grande importance et conduit à une meilleure compréhension de la pathogenèse de la maladie, à une meilleure prédiction du risque, du diagnostic, du pronostic et du traitement de la maladie.

La recherche et la caractérisation des déterminants génétiques impliqués dans les maladies génétiques "complexes" se repose sur les études d'épidémiologies, en se basant sur les trois étapes suivantes : (i) montrer que la maladie est familiale, (ii) montrer que cette tendance familiale est due à des facteurs génétiques et (iii) identifier les gènes impliqués.



**Figure 8.** Caractéristiques des maladies monogéniques (mendéliennes) et des maladies polygéniques (complexes).

## **2. La base génétique des maladies cardiovasculaires**

Les maladies cardiovasculaires englobent un spectre de phénotypes cliniques très divers où la composante héréditaire a été mise en évidence par différentes méthodes de l'épidémiologie génétique. Ainsi, dans cette section nous aborderons les différentes approches qui ont permis à identifier les gènes associés aux MCV.

### **a. Héritabilité des maladies et traits liés à la santé cardiovasculaire**

L'héritabilité exprime la variance génotypique et la variance phénotypique dans une population donnée. L'héritabilité est une estimation statistique du degré auquel les gènes contribuent à un trait multifactoriel [99]. Dans ce cas, le phénotype par rapport au génotype correspond à l'héritabilité par rapport à l'hérédité. Dans une étude épidémiologique [100], 50% des personnes hypertendus avaient des antécédents de maladies cardiovasculaires. Dans la même étude, l'héritabilité était de 27 % pour la pression artérielle diastolique et de 36 % pour la pression artérielle systolique [100]. la contribution génétique globale au développement de la coronaropathie est estimée entre 20 % et 60 % [99], les facteurs génétiques ayant une plus grande contribution chez ceux qui développent une coronaropathie à un plus jeune âge [101].

Dans le même concept, les études de jumeaux permettent de prédire l'existence ou non d'une prédisposition génétique à une maladie en question. Cette approche permet d'évaluer le taux de concordance de la maladie chez les jumeaux monozygotes (MZ) (génome identique à 100 %), et les jumeaux dizygotes (DZ) (ils partagent en moyenne 50% de leurs gènes). Les taux de concordance se réfèrent au pourcentage de cas dans lesquels, si un jumeau a un phénotype particulier, l'autre jumeau aussi présente ou développe le même phénotype.

Le principe de cette approche est de supposer que les facteurs environnementaux interviennent de la même manière sur les deux types de jumeaux. Ce qui veut dire que la différence du taux de concordance observée entre les types de jumeaux est d'origine génétique. Si le taux de concordance entre les jumeaux MZ est égal au taux de concordance entre les jumeaux DZ, donc il n'y a pas de facteur causal génétique. En revanche, plus la différence entre les deux taux de concordance est élevée, plus la part génétique est importante dans la maladie.

Cette étude a été largement utilisée pour étudier l'impact génétique sur la survenue des MCV. Selon certaines études, le taux de concordance pour la survenue d'un infarctus est ainsi de 39 à 65 % chez les jumeaux MZ, alors que la concordance est seulement de 22 à 26 % chez les jumeaux DZ [99].

## **b. Etudes de liaisons génétiques et candidats biologiques**

### **i. Définition et principe de l'étude de liaisons génétiques**

L'analyse des liaisons génétiques utilise des marqueurs génétiques distribués sur les chromosomes (dont l'emplacement sur le génome est connu) pour localiser les gènes (que nous essayons de localiser) qui influencent l'expression phénotypique à travers des observations d'individus apparentés. Il s'agit donc d'une analyse statistique effectuée à partir de grandes familles dont plusieurs sont atteints d'une maladie d'intérêt. Cela consiste à parcourir le génome afin de détecter la liaison telle que mesurée par la fraction de recombinaison. La recombinaison est utilisée pour déterminer la distance génétique entre le gène et le marqueur. Cette distance est exprimée en unité de recombinaison ou centiMorgan: 1 cM correspond à une fréquence de recombinaison de 1% entre le gène et le marqueur.

Cette approche repose donc sur la fraction de recombinaison. Deux gènes sont liés si, au cours de la méiose, une recombinaison se produit entre eux avec une probabilité inférieure à 50% [102]. Par conséquent, il existe un lien entre ces deux gènes, de sorte que le gène à cartographier est situé à proximité du gène à l'emplacement connu. Sinon, si la probabilité de recombinaison est supérieure ou égale à 50 %, dans ce cas on peut conclure que le gène de la maladie en question n'est pas à proximité du marqueur déjà cartographié.

Lorsque la fraction de recombinaison est importante, c'est-à-dire lorsqu'elle s'approche de 1/2, nous concluons que le gène de la maladie n'est pas à proximité du marqueur déjà cartographié. Lorsque la fraction de recombinaison était faible, les deux gènes semblaient être transmis ensemble plus souvent que séparément. Dans ce cas, on peut conclure qu'il existe un lien entre ces deux gènes, de sorte que le gène à cartographier est situé à proximité du gène à l'emplacement connu. 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 co-ségrègent avec le phénotype de la maladie [103].

Le succès limité de l'analyse de liaison des maladies complexes à ce jour est en partie dû au fait que les études étaient trop petites pour détecter des gènes avec un effet individuel faible. L'explication des résultats apparemment négatifs dépend en grande partie de la puissance. La taille de l'échantillon nécessaire pour détecter un lien avec des gènes présentant un risque relatif de génotype inférieur à 2 pourrait être irréalisable. Les erreurs de génotypage et les mutations non détectées peuvent affecter la puissance en ayant un impact important sur les informations de liaison [104]. Pour les grands pedigrees, ces erreurs de génotypage sont faciles à détecter car elles conduisent souvent à une incompatibilité mendélienne claire, mais peuvent ne pas être possibles lorsque seuls les frères et sœurs affectés sont génotypés sans aucune autre donnée familiale, donc le type de gène incorrect peut ne pas être déterminé.

## **ii. Définition et principe de l'étude gène candidat**

Cette approche consiste à rechercher et analyser les variants génétiques dans des gènes dont la fonction pourrait jouer un rôle dans la pathologie en question. 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 [105].

La DFM touche environ 3 à 4 % de la population générale [60], [106]. Néanmoins, la majorité des familles diagnostiquées de la DFM sont de petites tailles (par exemple, frères et sœurs, paires mère-fille), ce qui rend l'utilisation des études de liaison génétique moins puissantes pour identifier les causes génétiques. Afin de contourner cette limitation, les chercheurs ont appliqué l'approche gène candidat sur de petits échantillons de patients DFM. Les gènes ciblés tels que: FBN1, ACTA2, SMAD3, TGFBR1/2 et COL3A1 inclus dans cette approche n'ont montré aucune association significative avec le risque de la DFM, à cause de manque de puissance statistique [107], [108].

Cette approche a également contribué dans la recherche de gènes candidats pour la SCAD. Poloskey *et al* [107] ont identifié 11 gènes candidats significatifs, notamment PAM, GLI3, SEC24B, COL18A1, NFATC4, ARNTL, TBX2, HDAC9, SOX9, SORBS2 et COL4A2. Chacun de ces gènes représente un candidat crédible pour la SCAD, basé sur la fonction, l'expression et le phénotype de la souris, mais nécessite une validation dans de futures études.

Carss et ses collaborateurs [109] ont confirmé par l'approche de séquençage de l'exome, les rapports précédents indiquant une association significative entre le gène PDK1 et la SCAD [110]. Le répertoire de gènes CTD signalés comme étant impliqués dans la SCAD, peut être aussi intéressant par l'identification des patients présentant des variants pathogènes dans TGFB2, LOX, MYLK et YY1AP. On constate que parmi ces variants impliqués dans la DFM et la SCAD, il y a ceux qui sont communs pour ces deux phénotypes, c'est le résultat trouvé dans l'étude de Kim *et al* [70].

Cependant, l'approche gène candidat a été critiquée en raison de manque de puissance statistique pour conclure à une association [111], [112]. En effet, les tailles d'effets observées dans les études par gène candidat sont de faible ampleur, et la probabilité de survenue d'effets faux positifs est élevée. Heureusement, le développement du séquençage de nouvelle génération ainsi que les études d'association pan-génomiques, ont permis de mener des expériences sans d'hypothèses physiopathologiques, préalables sur la biologie de la maladie.

### **c. Apport des études de séquençage d'exomes**

#### **i. Définition et principe**

Le séquençage de l'exome entier (WES : Whole Exom Sequencing) peut identifier les variants codants pour les protéines dans tout le génome (appelé exome) [113]. Cette approche offre une méthode efficace pour étudier les maladies, car les régions exoniques représentent 2% du génome entier, mais peuvent contenir jusqu'à 85% de variants fonctionnels responsables des maladies. Par conséquent, cette technique se base sur deux étapes : La première étape consiste à sélectionner uniquement un sous-ensemble d'ADN qui code la protéine. La deuxième étape consiste à séquencer l'ADN exonique en utilisant n'importe quel système à haut débit. Technologie de séquençage de l'ADN [114]. L'approche WES offre plusieurs avantages, elle permet d'avoir toutes les mutations présentes au niveau des exons entre des sujets malades et non malades. Elle peut identifier les variants rares qui ne sont pas capturés par les GWAS, et enfin, elle est également plus rapide et moins coûteuse. À ce jour, l'utilisation du séquençage de l'exome dans de nombreuses maladies a donné de bons résultats, avec un succès relatif pour la DFM et la SCAD.

## ii. Quelques succès

Dans l'étude de Kiando et ses collaborateurs [115], 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. Après contrôle qualité, cela 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. Aucun gène n'héberge des variants partagés par tous les membres affectés d'au moins trois familles. Les gènes avec au moins quatre variants chez les 16 patients et les gènes vasculaires ont été suivis à l'aide des génotypes de 249 cas non apparentés et de 689 témoins.

Les analyses d'association basées sur les gènes à l'aide de SKAT-O montrent une association significative nominale avec la DFM multi-focale (N = 164) pour la kinase de la chaîne légère de la myosine (MYLK,  $P = 0,01$ ) précédemment impliquée dans l'anévrisme de l'aorte thoracique, l'obscurine (OBSCN), une protéine sarcomère ( $P = 0,003$ ), la chaîne lourde cytoplasmique de dynéine 1 (DYNC2H1,  $P = 0,02$ ) et RNF213 précédemment associée à la maladie de Moyamoya ( $P = 0,01$ ).

Une autre étude de séquençage de l'exome de 85 cas de SCAD et de 296 témoins d'origine chinoise a impliqué le gène du facteur de maturation du ribosome TSR1 [116, p. 1]. Cette étude familiale a également identifié un variant faux-sens rare dans TLN1, codant pour la protéine cytosquelettique talin1, et d'autres variants faux-sens rares dans des cohortes sporadiques, suggérant que TLN1 est un nouveau gène de susceptibilité. TLN1 est également fortement exprimé dans les tissus vasculaires et les cellules endothéliales [117]. Cela suggère que des altérations de l'expression de TLN1 (talin1) entraînent une réduction de l'intégrité de la paroi artérielle, un mécanisme pathogène putatif associé à SCAD.

Georges *et al* [118] ont identifié des mutations rares avec perte de fonction dans la DFM et la SCAD, le gène du récepteur de la prostacycline (PTGIR), un récepteur couplé aux protéines G à fonctions cardioprotectrices, impliqué dans la thrombose athérosclérotique [119]. Plus précisément, 4 patients atteints de la DFM et 2 patients atteints de SCAD portaient des allèles non synonymes (L67P, rs1397542892) et de perte de fonction (Q163X, rs199560500), qui retardaient significativement ou altéraient totalement la signalisation de l'AMPc,

respectivement, après stimulation par l'iloprost, un agoniste du récepteur de la prostacycline [118]. Les patients SCAD porteurs d'une perte de fonction PTGIR ne présentaient pas de DFM dans d'autres lits artériels. Cette observation peut suggérer que l'impact pathologique de ces mutations pourrait être différent dans la SCAD et la DFM. La plus grande cohorte de séquençage de patients signalée à ce jour comprenait 384 cas de SCAD qui ont été analysés par rapport à 13 722 témoins de la biobanque britannique

Fait intéressant, 2 variants hétérozygotes ont été identifiés dans *YY1AP1*, dont des mutations homozygotes ou hétérozygotes ont été rapportées dans le syndrome de Grange, un trouble avec des caractéristiques vasculaires de type DFM. Cette étude a utilisé une stratégie intéressante pour la hiérarchisation des gènes basée sur une approche d'apprentissage automatique qui a identifié 10 gènes liés à SCAD avec des variants ultra-rares, y compris tous les gènes définis manuellement.

Enfin, à l'aide d'une analyse d'enrichissement de gènes, les auteurs rapportent un enrichissement intéressant de variants dans la boucle génétique du développement de Henle, principalement dû à cinq individus porteurs de variants dans *PKD1*. En plus des rapports précédents [110], ce résultat suggère l'existence d'une association génétique potentielle entre la SCAD et les troubles génétiques rénaux qui mérite une enquête plus approfondie pour clarifier les mécanismes moléculaires et cellulaires partagés par ces 2 maladies.

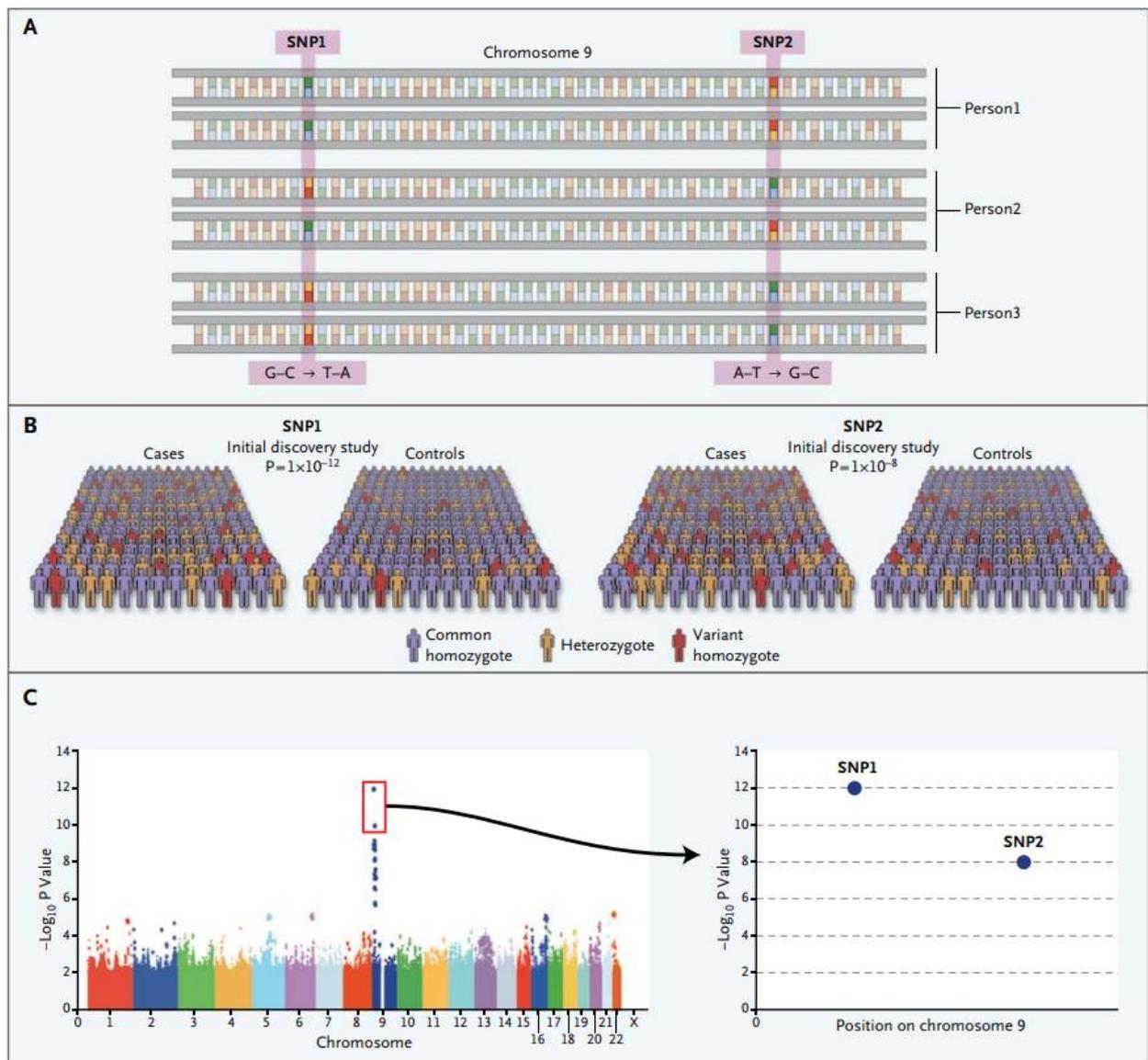
### **3. Contributions des études d'associations pangénomiques à la génétique des maladies**

Au cours de la période où les études de gènes candidats sur de petits échantillons cas-témoins produisaient souvent des résultats controversés, de nouvelles conceptions de puces et des coûts en baisse ont facilité le développement du génotypage à l'échelle du génome [120]. En parallèle, les méthodes statistiques s'amélioraient et des consortiums internationaux ont été créés pour mener une études d'association pan-génomique sur la maladie coronarienne [120].

#### **a. Principes généraux des études d'association pan-génomique (GWAS)**

Les études d'association pan-génomique ou *genome-wide association studies* (GWAS) utilisent des cartes denses de centaines de milliers de polymorphismes di-nucléotidiques (SNP)

répartis dans le génome humain (à l'exception du chromosome Y) pour rechercher des associations entre ces différents variants génétiques et les phénotypes étudiés [121] (Figure 9). Une percée technologique impérative qui a contribué au succès rapide de GWAS a été l'achèvement du projet international HapMap (abréviation de carte d'haplotype) en 2007, et la mise à disposition publique des données résultantes [122]–[124], qui ont permis la cartographie des paysages d'haplotypes de SNPs dans trois populations continentales. Huit ans plus tard, en 2015, le projet 1000 Genomes a profité de technologies de séquençage plus abordables et a publié gratuitement des données de variation génétique humaine, basées sur un séquençage du génome entier à faible couverture, atteignant un sommet avec le panel de référence 1000GP3 [125] . Récemment, le Haplotype Reference Consortium (HRC) [126], [127] a combiné tous les ensembles de données de séquençage du génome entier en un seul panel de référence d'haplotype pour faciliter l'imputation du génotype.

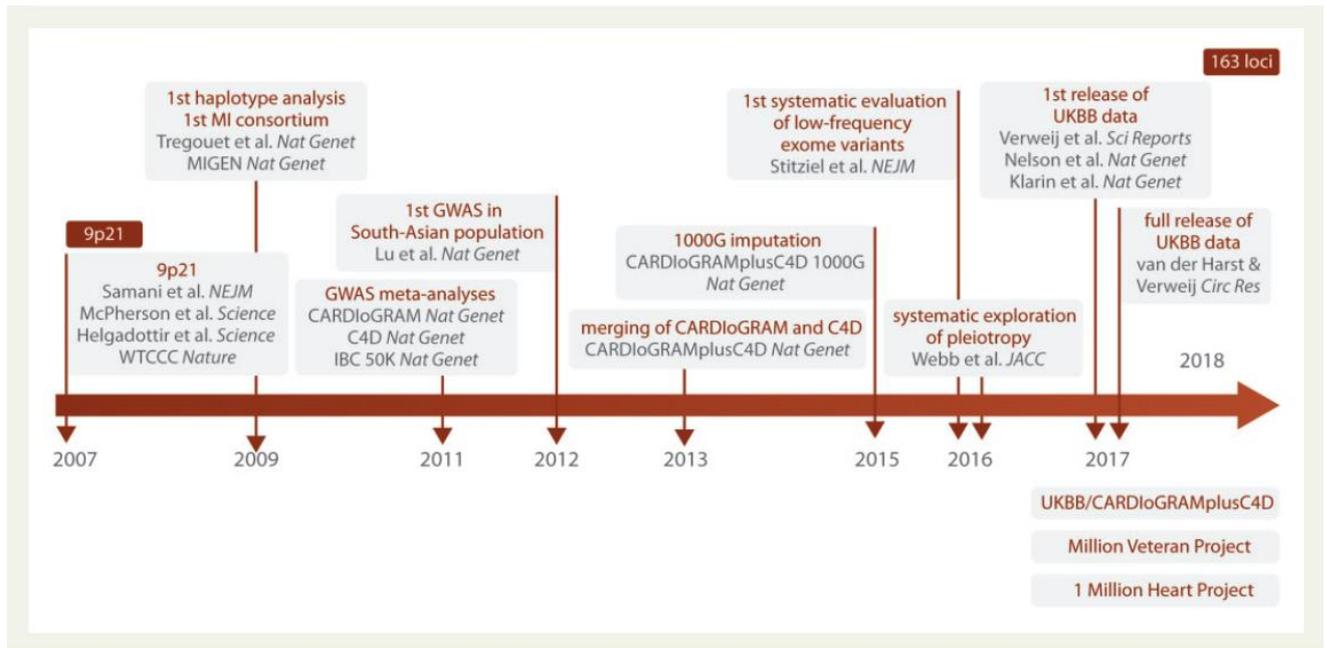


**Figure 9.** L'étude d'association à l'échelle du génome est généralement basée sur une conception cas-témoin dans laquelle les SNPs du génome humain sont génotypés. Le panneau **A** représente un petit locus sur le chromosome 9, et donc un très petit fragment du génome. Dans le panneau **B**, la force de l'association entre chaque SNP et la maladie est calculée sur la base de la prévalence de chaque SNP chez les cas et les témoins. Dans cet exemple, les SNP 1 et 2 sur le chromosome 9 sont associés à la maladie, avec des valeurs P de  $10^{-12}$  et  $10^{-8}$ , respectivement. Le graphique du panneau **C** montre les valeurs P pour tous les SNPs génotypés retenus après le contrôle de qualité, chaque chromosome étant affiché dans une couleur différente. Les résultats impliquent un locus sur le chromosome 9, marqué par les SNP 1 et 2, qui sont adjacents l'un à l'autre (graphique à droite), et d'autres SNP voisins. Adaptée de Manolio et collaborateurs [122].

## **b. Apports à la base génétique des maladies cardiovasculaires**

Les GWAS ont généré de nombreux résultats d'association significatifs pour les coronaropathies et les IM (Figure 10). En 2007, quatre groupes de recherche indépendants [128]–[131] ont découvert pour la première fois le locus de risque chromosome 9p21, fortement associé à la maladie coronarienne. Ce même locus a été observé dans la réplication dans l'étude familiale allemande, et a été aussi associé au risque d'IM. Deux autres études [132], [133] ont découvert 12 variants associés au risque de coronaropathies. Par la suite, de grands consortiums GWAS ont été formés et finalement fusionnés, pour analyser des centaines de milliers d'individus, à savoir, le consortium *Coronary ARtery DIsease Genome-wide Replication and Meta-analysis* (CARDIoGRAM) [134], le *MIGen Consortium* [132], le *Coronary Artery Disease (C4D) Genetics Consortium* [135], et, plus récemment, la *UK Biobank* (UKBB) [136]. En 2011 et 2013, les consortiums *CARDIoGRAM* [134] et *C4D* [168], ont identifié respectivement 25 et 46 nouveaux loci associés à la coronaropathie.

Trois autres études [139]–[141] ont trouvé chacune 13, 14 et 15 nouveaux loci associés à la coronaropathie, dont 7 se chevauchent entre les trois études. Très récemment, en analysant l'ensemble de données UKBB (34 541 cas de maladie coronarienne et 261 984 témoins) et en utilisant les données *CARDIoGRAMplusC4D* 1000 G pour la réplication, 64 nouveaux loci de risque de coronaropathie ont été identifiés [142].



**Figure 10.** Jalons de la recherche en génomique cardiovasculaire de 2007 à 2017 et au-delà. D'après Erdmann et al [120].

### c. Rôle des variations génétiques fréquentes dans la DFM et dans la SCAD

L'approche GWAS a également contribué dans les recherches des variants génétiques de la DFM et de la SCAD, cela a permis de trouver des SNPs non identifiés auparavant par l'approche exome entier chez les patients atteints DFM [115], [143].

Kiando *et al* [144] ont analysé environ 26 000 variants communs (la fréquence de l'allèle mineur (MAF)  $\geq 5\%$ ) générés par des puces d'exome chez 249 cas de DFM issus des cohortes françaises et 689 témoins. Ce travail a permis d'identifier puis de confirmer dans trois cohortes de réplification sur quatre, l'association de la DFM avec le polymorphisme rs9349379. Ce dernier est intronique à *PHACTR1*, un locus de risque de maladie coronarienne, de migraine et de dissection de l'artère cervicale. Les analyses des paramètres géométriques des carotides d'environ 2 500 volontaires sains indiquent une épaisseur plus élevée de l'intima média ( $P = 1,97 \times 10^{-4}$ ) et du ratio de la paroi artérielle sur la lumière ( $P = 0.002$ ), allant dans le sens d'un épaissement de la paroi artérielle. Des explorations fonctionnelles ont permis de détecter la protéine *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,  $P=0,003$ ). Dès 2015, l'étude *Carotid Artery Dissection and Ischemic Stroke Patient* (CADISP) dédiée à la génétique de la dissection de l'artère carotide a identifié *PHACTR1* comme site à risque d'AVC chez le jeune adulte [145]. En particulier, l'allèle rs9349379[A] était associé à un risque plus élevé de la dissection de l'artère carotide (OR = 1,33,  $P = 4,46 \times 10^{-10}$ ) [146], et de la migraine selon une méta-analyse de 29 études de GWAS incluant un total de 23 285 patients migraineux et 95 425 témoins appariés dans la population [147] ; la migraine étant une manifestation clinique commune à la DFM et à la CeAD. Ces résultats sont intéressants comptes tenus du chevauchement entre la DFM, la dissection carotidienne et la migraine.

L'approche GWAS a également permis l'identification de nouveaux variants génétiques associés à la SCAD. Turley *et al* [148] ont mené la première étude GWAS chez des femmes européenne atteintes de SCAD (180 cas et 340 témoins). Ils ont identifié un total de cinq loci à risque répliqués, y compris *PHACTR1*, initialement décrit par notre équipe par une approche candidate [149]. Le signal d'association le plus fort a été détecté dans le locus du chromosome 1q21.3, avec au moins 25 gènes cartographiés dans la région pertinente de 400 kb définie par

les auteurs, qui ont rapporté que le gène de la protéine ribosomique mitochondriale S21 (*MRPS21*), gène de la protéine de matrice extracellulaire 1 (*ECM1*) et le gène ADAMTS like 4 (*ADAMTSL4*) comme candidats les plus probables.

Un autre GWAS SCAD indépendant a rapporté des résultats d'association similaires sur le chromosome 1q21.3 sur la base de 2 études cas-témoins incluant 433 patients et 8470 témoins [150]. Les auteurs ont effectué une analyse de colocalisation à l'aide des données d'expression du portail d'expression du génotype et de GTEx et ont signalé *ADAMTSL4* comme un gène candidat potentiel pour ce locus. Il s'agissait d'un variant intronique, rs11172113, significativement associé à l'expression de ce gène dans plusieurs tissus artériels [148], [150]. Un autre locus a été signalé au locus Chr 21q22.11, qui peut impliquer plusieurs gènes montrant des signes d'expression dans les artères. Turley *et al* [148] ont suggéré que *LINC00310* pourrait être le candidat le plus probable pour ce locus, étant donné la compréhension croissante du rôle biologique des longs ARN non codants. Cependant, Saw *et al* [150] ont effectué une analyse de colocalisation en utilisant l'expression génique dans les artères et l'association avec SCAD à ce locus et a indiqué 2 candidats : *MRPS6* codant pour une protéine ribosomique mitochondriale S6 et *KCNE2* codant pour la sous-unité régulatrice E de la sous-famille E des canaux potassiques voltage-dépendants 2. Des études plus importantes impliquant idéalement d'autres origines ethniques seront certainement nécessaires pour affiner les variants causaux et les gènes à ce locus. Un locus SCAD supplémentaire a été identifié par Turley *et al* [148] dans le gène de la fibrilline (*FBNI*) au locus 15q21.1. Ce signal d'association est piloté par plusieurs variants communs couvrant une grande région de déséquilibre de liaison qui comprend *FBNI*, un gène qui abrite des mutations rares responsables du syndrome de Marfan. Comme souligné par les auteurs, la protéine codée par *ADAMTSL4*, le meilleur gène candidat au locus 1q21.3, se lie à *FBNI* (fibrilline 1) et favorise l'assemblage des microfibrilles.

#### **4. Rôles des variations génétiques liés aux chromosomes sexuels dans le risque génétique cardiovasculaire**

Comme il a été mentionné dans le chapitre 1, le sexe est un des facteurs de risque des MCV. Donc il est important de comprendre les mécanismes sous-jacents aux différences sexuelles pour concevoir des approches préventives et thérapeutiques optimales pour tous les individus. Notons que les maladies neurodégénératives, la sclérose en plaques, la maladie

d'Alzheimer [151], la DFM et la SCAD ont une incidence plus élevée chez les femmes [61], [65], [70].

De nombreuses études ont étudié les différences d'héritabilité entre les sexes [152] et l'écart des corrélations génétiques [153]–[155] et ont réalisé des études d'association pan-génomique stratifiées par sexe (GWAS) pour évaluer directement les différences dans les effets des variants génétiques entre les sexes [156]–[159].

Parmi elles, une étude à grande échelle sur une population britannique a analysé l'hétérogénéité génétique entre les sexes en considérant 19 traits complexes. En particulier, il a été constaté que les génotypes spécifiques au sexe contribuent avec une taille d'effet modérée au phénotype d'une proportion de traits, notamment la pression artérielle, la taille, la taille et le tour de hanches. Une méta-analyse de 114 GWAS incluant plus de 320,485 individus d'origine européenne, a été réalisée sur l'adiposité viscérale [160], un des facteurs de risque de MCV. Le résultat obtenu est l'identification de 44 variants génétiques autosomiques, dont 28 loci ont montré des effets plus importants chez les femmes que chez les hommes, 5 loci ont montré des effets plus importants chez les hommes que chez les femmes et 11 ont montré des effets opposés entre les sexes.

## **5. Apports de l'approche des scores polygéniques à partir des données GWAS**

Plusieurs méthodes post-GWAS ont été introduites pour surmonter les limitations de puissance lors des GWAS [161]. En raison de problèmes de confidentialité [162], [163], tels que l'accès aux ensembles de données GWAS, la plupart des méthodes post-GWAS ne nécessitent que des résumés statistiques (résultats du GWAS). Cette approche unique de l'analyse post-GWAS est appelée analyse du score de risque polygénique (PRS), également connue sous le nom de score de risque génétique. L'approche PRS a récemment suscité beaucoup d'intérêt pour évaluer le pouvoir explicatif d'un ensemble de marqueurs [164]. Les détails de cette méthode sont présentés dans la section suivante.

## a. Définition et principe

La prédisposition génétique d'un individu à un trait donné peut être résumée dans un score polygénique [165]. La méthode PRS cartographie les données de génotype des résumés statistiques GWAS en une seule variable qui est utilisée pour estimer les scores de risque au niveau individuel pour les traits phénotypiques. L'analyse PRS a été utilisée pour prédire l'héritabilité individuelle en combinant tous les SNP sélectionnés [164]. En d'autres termes, la PRS est une méthode puissante pour estimer la prédisposition phénotypique d'un individu [166]. Son calcul PRS se fait en additionnant les allèles à risque que possède un individu, pondérés par la taille d'effet des allèles à risque estimés par GWAS pour le phénotype étudié [166]. Pour chaque individu, le nombre d'allèles à risque portés à chaque variant (0, 1 ou 2) est additionné, pondéré par sa taille d'effet (c'est-à-dire log (OR) pour les traits binaires ou coefficient bêta pour les traits continus). Le résultat est un score unique de la charge génétique de chaque individu pour une maladie ou pour un trait continu [167].

Une série de méthodes ont été développées pour calculer le PRS [168]. Il existe trois approches principales [167]: (i) une approche pondérée, prenant en compte les tailles d'effet dérivées des résultats de GWAS ; (ii) une approche non pondérée prenant en compte les analyses à marqueur unique ; (iii) une approche de retrait prenant en compte les analyses. Ces différentes méthodes d'analyse PRS sont présentées dans la Figure 10 et le Tableau 2. Elles sont également divisées en deux catégories : les méthodes bayésiennes et non bayésiennes. Les méthodes PRS peuvent également être classées selon l'utilisation du déséquilibre de liaison (LD) : les méthodes PRS qui incorporent les méthodes LD, et d'autres méthodes qui appliquent le *clumping* LD. Afin de faciliter la compréhension de son algorithme sous-jacent, les méthodes d'analyse PRS sont divisées en quatre catégories (voir Tableau 3) [161].

Parmi les différentes méthodes mentionnées ci-dessus, le "*clumping + thresholding*" (C+T, également connu sous le nom de "*pruning + thresholding*") est la méthode la plus couramment utilisée pour construire PRS. Cette méthode est basée sur des résumés statistiques univariés de GWAS. Ces résumés représentent les résultats de tests d'association entre les SNPs et les phénotypes étudiés. Ces tests d'associations sont basés sur les modèles mathématiques donnés dans les équations (1) et (2), pour les traits continus et binaires, respectivement.

$$y = \alpha_j + \beta_j G_j + \gamma_j^{(1)} COV^{(1)} + \dots + \gamma_j^{(K)} COV^{(K)} + \epsilon, \quad (1)$$

$y$  est le phénotype continu,  $\alpha_j$  est l'ordonnée à l'origine,  $G_j$  est le  $j^{ième}$  SNP avec son effet estimé  $\beta_j$ .  $COV^{(1)}, \dots, COV^{(K)}$  sont les  $K$  covariables d'effet  $\gamma_j^{(1)}, \dots, \gamma_j^{(K)}$  respectivement,  $y$  compris les composantes principales et d'autres covariables telles que l'âge et le sexe.

$$\log\left(\frac{p}{1-p}\right) = \alpha_j + \beta_j G_j + \gamma_j^{(1)} COV^{(1)} + \dots + \gamma_j^{(K)} COV^{(K)}, \quad (2)$$

$p = P(Y = 1)$ , où  $Y$  représente le phénotype binaire (par exemple, l'état de la maladie), une régression logistique est utilisée et un test Z-test est effectué sur  $\beta_j$  pour chaque SNP.

L'approche « *Clumping+Thresholding* » se compose de deux étapes importantes comme son nom l'indique. La première étape nommée « **Clumping** » consiste à regrouper Les SNPs afin qu'il ne reste que des SNP faiblement corrélés entre eux, en se basant sur le déséquilibre de liaison (en anglais *Linkage Disequilibrium* LD). Le LD est une mesure de l'association non aléatoire entre des allèles à différents loci sur le même chromosome dans une population donnée [166]. L'association aléatoire se produit lorsque la probabilité d'observer une paire d'allèles sur un chromosome n'est pas égale au produit des probabilités d'observer ces allèles individuellement [169]. Un SNP est en LD lorsque la fréquence associée de ses allèles est plus élevée que prévu dans la classification aléatoire. Le LD implique des modèles de corrélation entre les SNPs dans tout le génome. Il est donc extrêmement difficile d'identifier les contributions des variants génétiques causaux indépendants. Une façon de capturer approximativement le bon niveau de signal causal consiste à effectuer un **clumping**, qui supprime les SNPs de manière à ne conserver que les SNPs faiblement corrélés. Cette étape consiste donc à examiner d'abord le SNP le plus significatif, calculer la corrélation entre ce SNP et les SNP proches (à une distance génétique de, par exemple, 250 kb) et supprimer tous les SNP proches qui sont corrélés avec ce SNP au-delà d'un seuil prédéfini. Cette étape vise à améliorer la précision de l'analyse PRS.

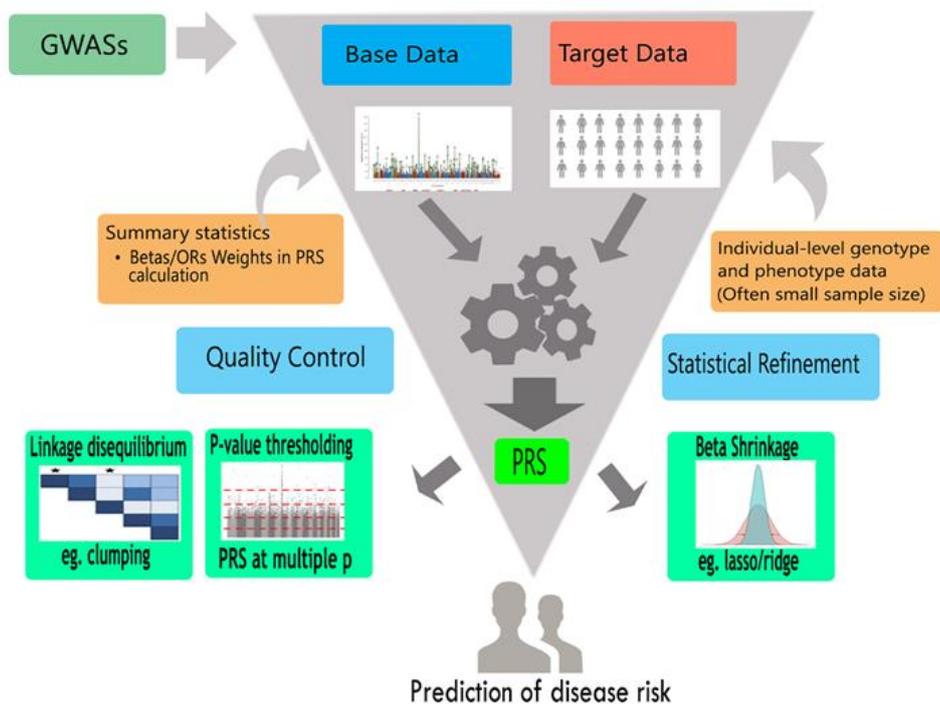
La seconde étape « **Thresholding** » consiste à supprimer les SNPs avec une  $P$ -value supérieure à un seuil de  $P$ -value ( $p_T$ ) ( $p_T = 5 \times 10^{-8}, 1 \times 10^{-5}, 1 \times 10^{-3}, 0.01, 0.05, 0.1, 0.2, 0.3, 0.4, 0.5$ . [166]) afin de réduire le bruit dans le score. Généralement, plusieurs seuils de valeur  $p$  sont testés pour maximiser la prédiction. Un PRS est finalement défini comme la somme des nombres d'allèles des SNP restants (après C+T) pondérée par les tailles d'effet

GWAS correspondantes [164], [170], [171]. Son équation mathématique est donnée ci-dessous :

$$PRS_i = \sum_{\substack{j \in \text{clumping} \\ p_j < p_T}} \beta_j \cdot G_{i,j}, \quad (3)$$

Où  $\beta_j$  est la taille d'effet estimé par GWAS, et  $G_{i,j}$  est le nombre d'allèles observés pour le  $i^{\text{ème}}$  individu et le  $j^{\text{ème}}$  SNP.

## Workflow in Polygenic Risk Score Calculation and Analysis



**Figure 11.** Étapes d'analyse du score de risque polygénique (PRS) montrant les données de base, les données cibles et différentes méthodes en regroupant différentes approches. À l'aide de données de génotype et de phénotype au niveau individuel ou résumés statistiques, des approches telles que la régression lasso/ridge, « *clumping & thresholding* » peuvent être utilisées pour augmenter la précision prédictive de l'analyse PRS. De plus, les résultats peuvent être utilisés pour prédire le risque de santé ou de maladie et donner un aperçu des traitements appropriés [161].

Tool	Approach	Computational platform	User friendly	Functionality
LDpred <sup>16</sup>	Bayesian Shrinkage Prior	Python	Difficult	Uses a prior on effect sizes and LD information from an external reference panel
PRS-CS <sup>24</sup>	Bayesian regression framework	Python	Difficult	Utilizes a high- dimensional Bayesian regression framework, by placing a continuous shrinkage (CS) prior on SNP effect sizes
EB-PRS <sup>19</sup>	Empirical Bayes approach	R	Difficult	A novel method that leverages information for effect sizes across all the markers
AnnoPred <sup>20</sup>	Bayesian Shrinkage Prior	Python	Difficult	A framework that leverages diverse types of genomic and epigenomic functional annotations
PRSice <sup>36</sup>	Clumping + thresholding (C+T)	R	Difficult	For calculating, applying, evaluating and plotting the results of PRS analysis
PRSice2 <sup>37</sup>	Clumping +thresholding (C+T)	C++, R	Easy	An efficient and scalable software program for automating and simplifying PRS analyses on large-scale data
LDpred2 <sup>38</sup>	Bayesian Shrinkage	R	Difficult	A faster and more robust implementation of LDpred in R package bigsnpr
BSLMM <sup>39</sup>	Bayesian sparse linear mixed model	R	Difficult	Prior specification for the hyper-parameters and a novel Markov chain Monte Carlo algorithm for posterior inference
BayesR <sup>23</sup>	Hierarchical Bayesian Mixture Model	Fortran	Difficult	Bayesian mixture model that simultaneously allows variant discovery, estimation of genetic variance explained by all variants.
DPR software <sup>40</sup>	Latent Dirichlet process regression model	C++	Easy	Dirichlet process regression to flexibly and adaptively model the effect size distribution.
SMTpred <sup>41</sup>		Python	Difficult	Combines SNP effects or individual scores from multiple traits according to their sample size, SNP-heritability ( $h^2$ ) and genetic correlation ( $r_G$ ).
Lassosum <sup>21</sup>	Penalised Regression	R	Difficult	A method for constructing PGS using summary statistics and a reference panel in a penalized regression framework.
Plink <sup>42</sup>	$p$ -value thresholding approach	C/C++	Easy	Open-source C/C++ toolset for GWAS analysis and research in population genetics.

**Tableau 2.** Description des différentes approches utilisées pour le calcul du score de risque polygénique (PRS) [161] .

Key factors	Approaches			
	<i>p</i> -value thresholding with clumping	Penalised regression	Clumping + thresholding (C+T)	Bayesian shrinkage prior
Controlling for Linkage Disequilibrium	N/A	LD matrix is integral to algorithm	Clumping	Shrink effect sizes with respect to LD
Shrinkage of GWAS effect size estimates	<i>P</i> -value threshold	LASSO, Elastic Net, penalty parameters Bayesian	<i>P</i> -value threshold standard	Prior distribution, e.g. fraction of causal SNPs

**Tableau 3.** Comparaison des différentes approches pour effectuer des analyses PRS [161].

## **b. Applications des scores polygéniques dans le domaine cardiovasculaire**

Comme nous l'avons déjà vu dans les sections précédentes, les MCV sont des maladies complexes avec des bases génétiques importantes, elles représentent l'une des principales causes de morbidité et de mortalité dans le monde. L'approche PRS a été utilisée dans plusieurs études comme un outil clinique pour prédire les MCV [172]–[177]

Dans l'étude de Agbaedeng *et al* [178], on trouve une prédiction de risque des coronaropathies, allant de 8 % à 49 %. Les études de Arnold *et al* [172], Hadley *et al* [173] ont démontré que le risque capturé par le PRS semble largement indépendant des facteurs de risque traditionnels. Même si la prédisposition génétique ne peut pas être modifiée directement, le risque héréditaire peut être atténué avec succès par un profil de mode de vie favorable et, chez certains sujets, même en combinaison avec des interventions pharmacologiques précoces, par ex. dans le métabolisme des lipides. L'approche PRS permet d'améliorer la stratification du risque des maladies cardiovasculaires athéroscléroses et d'identifier les patients susceptibles de tirer un bénéfice accru de la prévention primaire et secondaire. Le PRS est aussi significativement associé au risque d'IM (1.06 (1.02-1.11),  $p = 0.01$ ) [179]. La stratification sur le sexe a montré un PRS associé au risque d'IM chez les hommes (1,06 (1,01-1,12),  $p = 0,02$ ) et chez les individus au-dessus de la médiane de 45,11 ans (1,06 (1,00-1.12),  $p = 0.03$ ). Ce résultat a été aussi démontré dans l'étude de Manikpurage *et al* [180].

## **6. Apport de la corrélation génétique et de la randomisation Mendélienne**

Comme nous l'avons vu dans les sections précédentes, les GWAS ont énormément contribué dans la compréhension des relations complexes entre les traits humains et les maladies. Afin de mieux comprendre la relation génétique entre deux traits, la corrélation génétique peut estimer la corrélation de l'effet des allèles entre deux traits sans donner l'information sur le sens de la corrélation. Alors que la randomisation mendélienne peut évaluer l'effet causal d'une exposition en se concentrant sur les variants génétiques significativement associés à l'échelle du génome en tant que variables instrumentales pour l'exposition d'intérêt. Cela signifie que les deux méthodes sont complémentaires. Dans la suite de cette section, nous allons voir de plus près ces deux approches complémentaires ainsi que leur contribution dans la recherche des variants génétiques responsables de plusieurs maladies.

## a. La corrélation génétique

### i. Principales méthodes

Les études de corrélation génétique estiment la corrélation des effets alléliques entre deux traits via les SNPs responsables dans le génome. La corrélation génétique  $r_g$  peut être définie comme [181] :

$$\frac{\sum_k b_k c_k}{\sqrt{\sum_k b_k^2} \sqrt{\sum_k c_k^2}} \quad (4)$$

où  $b_k$  est l'effet par allèle du SNP  $k$  à partir d'une régression multivariée du trait  $Y_1$  sur tous les SNPs dans un échantillon infini d'une population donnée, et  $c_k$  est le même pour le trait  $Y_2$ . En pratique, lorsque les tailles d'échantillon pour les études de  $Y_1$  et  $Y_2$  sont généralement inférieures au nombre de SNP à analyser, la régression multivariée ne peut pas être utilisée pour estimer  $b_k$  ou  $c_k$  ; des hypothèses supplémentaires sont nécessaires [182]. Il existe deux méthodes pour calculer la corrélation génétique :

#### ***Méthode 1 : Accès à des données individuelles*** [181]

Lorsque des données au niveau individuel sont disponibles, la corrélation génétique à l'échelle du génome entre deux traits peut être estimée à l'aide d'une analyse bivariée du maximum de vraisemblance restreinte basée sur le génome (GREML) [183] sur des données GWAS d'individus indépendants, telles qu'implémentées dans le logiciel GCTA largement utilisé [184].

#### ***Méthode 2 : Absence des données individuelles*** [181]

En l'absence de données au niveau individuel, les corrélations génétiques entre deux traits peuvent être estimées par la régression du score LD inter-traits (LDSC) de plus en plus utilisée [183]. Le LDSC s'appuie sur le fait que les statistiques d'association spécifiques aux SNPs incluent les effets de tous les SNPs en déséquilibre de liaison (LD) avec ce SNP. Ainsi, pour un trait polygénique, les SNPs dans les régions à LD élevée auront en moyenne des statistiques de  $\chi^2$  plus élevées que les SNPs dans les régions à faible LD.

Nous pouvons estimer la corrélation génétique entre deux traits en utilisant l'équation suivante :

$$E[\beta_j \gamma_j] = \frac{\rho_g}{M} l_j + \frac{N_s \rho}{N_{Y_1} N_{Y_2}} \quad (5)$$

Où  $\beta_j$  and  $\gamma_j$  est la taille d'effet pour le SNP  $j$  sur le trait  $Y_1$  et le trait  $Y_2$  respectivement,  $\rho_g$  est la covariance génétique,  $M$  est le nombre de SNP,  $N_1$  et  $N_2$  sont les tailles d'échantillon pour le trait  $Y_1$  et  $Y_2$  respectivement,  $N_s$  est le nombre d'échantillons qui se chevauchent,  $\rho$  est la corrélation phénotypique entre les échantillons qui se chevauchent et  $l(j, c)$  est le score LD du  $j^{ieme}$  SNP, défini comme  $l(j, c) = \sum_{k \in c} r^2(j, k)$  [185]. Pour obtenir la corrélation génétique  $r_g$ , on peut normaliser la covariance génétique  $\rho_g$  par les héritabilités SNP estimées pour les deux traits :  $r_g = \frac{\rho_g}{h_{g1}^2 h_{g2}^2}$ ,

où  $h_{g1}^2$  et  $h_{g2}^2$  sont les héritabilités SNP pour les traits  $Y_1$  et  $Y_2$  respectivement [185].

La méthode LDSC permet le chevauchement d'individus entre les études et elle est rapide en termes de calcul. En fait, même dans les situations où des données au niveau individuel sont disponibles, LDSC pourrait être préférable en raison de la faisabilité informatique. Cependant, cette méthode est moins précise que l'analyse GREML et nécessite que le panel de référence externe LD corresponde à la population étudiée [186]. Les analyses de corrélation génétique supposent que chacun des traits à l'étude a une composante génétique, de sorte que les traits avec  $h_g^2 > 2$  ne doivent pas être inclus dans ces analyses.

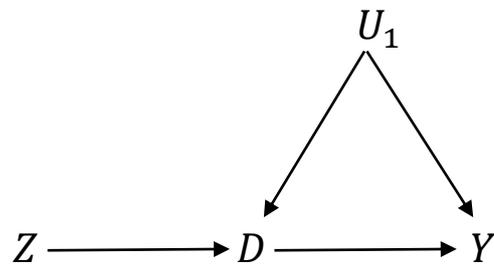
## **b. La randomisation Mendélienne**

L'un des développements récents les plus importants en génétique épidémiologique se trouve dans un autre modèle de recherche qui croise la génétique et l'environnement : la randomisation mendélienne (RM). Elle permet d'utiliser la génétique pour étudier les liens entre phénotypes intermédiaires et phénotypes d'intérêt afin de déterminer si ces liens sont causaux ou simplement corrélés [187].

### **i. Définition et principe**

La randomisation mendélienne (RM) est une technique d'utilisation de variants génétiques pour estimer l'effet causal d'un facteur de risque modifiable à partir de données

d'observation [188]–[190]. Ce modèle d'étude est identique à la méthode des variables instrumentales très utilisée en économétrie. L'instrumentation est une technique d'estimation d'un effet causal entre l'exposition  $D$  et une issue  $Y$  lorsque cet effet est influencé par une (ou plusieurs) variable de confusion non mesurée  $U_1$ , tel qu'illustré à la Figure 12.



**Figure 12.** Graphe acyclique dirigé (en anglais, *directed acyclic graph* ou DAG), un outil permettant de représenter la structure de cause à effet entre les variables. Il illustre toutes les suppositions faites à propos des liens de causalité entre les variables d'exposition  $D$  et le phénotype d'issue  $Y$ . Il contient également l'instrument génétique  $Z$  et la variable de confusion  $U_1$ .

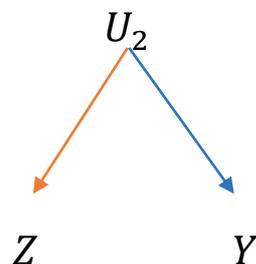
L'interprétation d'un DAG se fait comme suit. L'instrument  $Z$  est corrélé avec  $D$  mais indépendant des variables de confusion  $U_1$ , afin d'estimer l'effet causal de  $D$  sur  $Y$ . La flèche  $Z \rightarrow D$  implique le lien entre  $Z$  et  $D$ . Ce n'est pas nécessairement une relation causale, mais c'est un abus de notation et n'apparaît que dans la connexion entre les représentations. Ces outils sont définis comme des marqueurs génétiques dans les méthodes RM [189], [191], [192]. Bien que plusieurs marqueurs génétiques puissent être utilisés comme outils dans la randomisation mendélienne, par souci de simplicité nous nous limiterons aux cas les plus courants où un ou plusieurs SNPs sont utilisés comme outils. Dans un problème donné, un SNP peut être utilisé comme instrument s'il satisfait les conditions suivantes [193] :

- (A1) : il est corrélé avec le phénotype exposition  $D$  ;
- (A2) : il n'existe pas de cause commune non mesurée entre  $Z$  et le phénotype issue  $Y$  ;
- (A3) : tout chemin dirigé (séquence de flèches) de  $Z$  allant vers  $Y$  passe uniquement par  $D$  ;

La condition (A1) signifie que l'instrument  $Z$  prédit  $D$ . Cela signifie que (A1) requiert que  $Z$  prédise  $D$ , et non pas que  $Z$  cause  $D$ , et tout SNP avec une forte association avec  $D$  peut être un instrument valide au niveau de (A1). Notez que plus la corrélation entre  $Z$  et  $D$  est forte,

plus la fraction de D prédite par Z est grande et meilleure est la force de l'instrument Z. Notons que la condition (A1) peut être validée dans les données à l'aide d'un modèle d'association statistique approprié : régression logistique si D est binaire ; régression linéaire simple si D est continue.

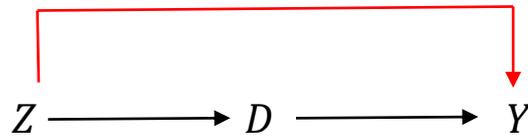
La condition (A2) implique qu'il n'existe pas de variable non ou mal mesurée précédant l'assignation de Z dans le temps qui influence à la fois Z et Y. Remarquons d'abord qu'une variable pouvant influencer à la fois Z et Y doit être préalable à Z et Y chacun. Dans la vaste majorité des cas, la seule variable qui peut être préalable à Z est l'origine ethnique puisque la composition génétique d'un individu est déterminée avant sa naissance. L'origine ethnique est une cause de Z parce qu'elle cause l'apparition plus ou moins fréquente de mutations génétiques dans une population donnée. De plus, puisque l'origine a une influence sur les gènes d'individus issus d'une population donnée, elle influence également l'apparition de certains traits physiques, donc de phénotypes, à l'intérieur d'une population donnée. Ainsi, l'origine ethnique est une variable qui pourrait être une cause commune à l'instrument et au phénotype issue. Cet exemple est illustré ci-dessous :



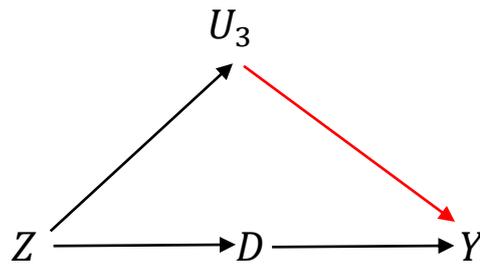
**Figure 13.** DAG partiel représentant une violation de la condition (A2).

La flèche de gauche (orange) de la Figure 13 implique que  $U_2$  précède l'assignation des allèles de risque à l'individu lors de sa conception, tandis que la flèche de droite (bleue) implique que  $U_2$  influence le phénotype issu indépendamment de la fréquence des allèles de risque. C'est l'existence des deux flèches dans un même DAG qui implique une violation de (A2). Dans le cas où  $U_2$  représente l'origine ethnique, il y aurait un problème si la fréquence des allèles de risque ainsi que la prévalence de l'issue différaient selon les origines ethniques et que l'étude contenait des individus de plusieurs ethnies différentes. Toutefois, en pratique, il est habituellement possible de déterminer et d'ajuster pour l'origine ethnique dans le modèle de randomisation mendélienne, minimisant ainsi la possibilité d'une violation de (A2) [194].

La supposition (A3) implique que  $Z$  influence  $Y$  à travers  $D$  uniquement. Supposons que  $Z$  influence  $Y$  par un chemin autre que  $Z \rightarrow D \rightarrow Y$  tel qu'illustré dans les Figures 14 et 15.



**Figure 14.** Exemple d'une violation de la condition (A3).



**Figure 15.** Autre violation de la c (A3),  $U_3$  est un phénotype inconnu.

À la Figure 14, l'existence de la flèche allant de  $Z$  vers  $Y$  directement implique que le SNP utilisé comme instrument influence non seulement l'expression du phénotype exposition, mais aussi celle du phénotype issu. De façon similaire, à la Figure 15, si  $U_3$  est un autre phénotype inconnu, l'existence du chemin  $Z \rightarrow U_3 \rightarrow Y$  implique que le SNP utilisé comme instrument influence l'expression d'un autre phénotype qui a également une influence sur le phénotype issu. Notons que, au contraire de (A2) qui ne pose pas de problèmes importants qui ne peuvent pas être résolus en pratique, la préoccupation avec (A3) n'est pas que théorique, car les gènes peuvent agir sur plusieurs phénotypes ce qu'on appelle la pléiotropie [195]. La plupart de ces mécanismes d'action sont inconnus ou mal compris.

Par conséquent, l'approche RM consiste à utiliser des variables instrumentales génétiques pour obtenir des inférences causales à partir de données d'observation. Il existe différentes méthodes d'estimation des variables instrumentales [191]. Notons qu'une approche de données au niveau individuel nécessite des données individuelles sur la variation génétique à utiliser comme variables instrumentales, facteurs de risque et résultats. Les méthodes basées sur des résumés statistiques ne nécessitent que les résultats de l'association de variables instrumentales avec des facteurs de risque. Si la disponibilité des données au niveau individuel est limitée. Par exemple, des données individuelles existent pour la variable instrumentale  $D$  en tant que facteur

de risque, mais pas pour la variable instrumentale Y en tant que phénotype d'issue (Figure 11), un résumé statistique de l'association peut être obtenu à partir des données à un niveau personnel,

### **c. Contribution de la corrélation génétique et de la randomisation mendélienne à l'étude de la génétique des maladies cardiovasculaires**

La corrélation génétique et la randomisation mendélienne, deux approches qui ont apporté une grande contribution dans l'étude génétique des MCV. Cela a été démontré dans de nombreuses études. Dans l'étude apportée sur la relation entre l'asthme et les MCV, Zhou *et al* [196] se sont servis des résumés statistiques de GWAS récemment publiées, afin d'examiner la corrélation génétique, les variants génétiques partagés et la relation causale entre l'asthme (N = 127 669) [197] et les maladies cardiovasculaires (N = 86 995 – 521 612) [137], [198], [199]. En utilisant la méthode statistique HDL [200] similaire à la méthode LDSC décrite précédemment, cela a permis d'observer une corrélation génétique significative entre l'asthme et l'insuffisance cardiaque (IC) ( $R_g = 0,278$ ,  $P = 5 \times 10^{-4}$ ). En utilisant l'analyse RM basée sur la méthode pondérée par l'inverse de la variance, cela a montré que l'asthme était significativement associé à l'insuffisance cardiaque, avec un Odd Ratio (OR) de 1,07 [IC à 95 % : 1,03 – 1,12 ;  $P = 1,31 \times 10^{-3}$ ].

La corrélation génétique et de la randomisation mendélienne bidirectionnelle ont également été utilisés pour l'exploration de la relation entre la schizophrénie et les maladies cardiovasculaires : Rada *et al* [201] ont utilisé des résultats de GWAS sur la schizophrénie (N = 130 644), l'insuffisance cardiaque (N = 977 323), la maladie coronarienne (N = 332 477), la tension artérielle systolique et diastolique (N = 757 601), la variabilité de la fréquence cardiaque (N = 46 952), intervalle QT (N = 103 331), schémas ECG de repolarisation précoce et de cardiomyopathie dilatée (N = 63 700). Avec la RM multivariée, les auteurs de cette étude ont recherché si les effets de causalité étaient médiés par le tabagisme, l'indice de masse corporelle, l'activité physique, les taux de lipides ou le diabète de type 2. Les résultats de la corrélation génétique entre la schizophrénie et les maladies cardiovasculaires étaient proches de zéro ( $-0,02$  ;  $-0,04$ ). Il y avait des preuves que la responsabilité de la schizophrénie augmente causalement le risque d'insuffisance cardiaque. Cet effet est resté compatible avec l'approche RM multivariée. Il y avait peu de preuves qu'une pression artérielle systolique plus élevée augmente le risque de schizophrénie. Une autre étude basée sur la méthode RM à 2 échantillons

pour étudier les associations causales de 18 facteurs de risque cardiovasculaire et des comportements liés au mode de vie avec l'hypertension [202].

## **Partie II. Résultats**

### **Chapitre 3.**

#### **L'étude du rôle des variations génétiques fréquentes du chromosome X dans le risque de la DFM et de la SCAD**

# 1. Méthodologies appliquées

## a. Populations étudiées

### i. Études DFM

#### ▪ **ARCADIA**

Les patients atteints de la DFM ont été inclus à partir du registre ARCADIA (*Assessment of Renal and Cervical Artery DysplasIA*), un registre national de la DFM à l'Hôpital Européen Georges Pompidou, APHP, Paris8 [203]. Le diagnostic des patients DFM a été établi sur la base des informations cliniques issues des antécédents médicaux, de l'interprétation de l'angiographie et/ou de la tomodensitométrie des lits artériels après exclusion d'autres causes de sténose artérielle telles que l'athérosclérose, la maladie de Takayasu, le syndrome d'Ehlers Danlos et la neurofibromatose de type 1. Etant donné la complexité de l'interprétation de l'imagerie des maladies vasculaires, un panel local d'experts comprenant des cliniciens des services d'hypertension, de radiologie, de médecine vasculaire et de génétique médicale a validé le diagnostic. Le protocole ARCADIA/PROFILE a été validé par le Comité de Protection des Personnes d'Île-de-France (CPP d'île de France) le 04/03/2009 (ID : 2009-A00288-49).

#### ▪ **3 Cités**

L'étude de trois cités (étude 3C) est une étude longitudinale basée sur la population de la relation entre les maladies vasculaires et la démence chez les personnes âgées de 65 ans et plus [204]. Les participants ont été recrutés dans trois villes Françaises : Bordeaux (Sud-Ouest), Dijon (Nord-Est) et Montpellier (Sud-Est). L'étude 3C s'est étendue de 1999 à 2012. Les participants ont subi des examens approfondis réguliers. L'examen comprenait des mesures des facteurs de risque vasculaires traditionnels (tension artérielle, glycémie, lipides, etc.), des fonctions cognitives et des maladies vasculaires subcliniques par échographie carotidienne et imagerie par résonance magnétique (IRM) cérébrale. Le protocole d'étude a été validé par le comité consultatif de protection des personnes dans la recherche biomédicale Bicêtre Hôpital Bicêtre n°99-28 CCPPRB agréé les 10/06/99, 11/03/2003 et 17/03/2006.

- **ARCADIA-POL**

Les patients ont été recrutés dans le cadre de l'étude ARCADIA-POL (*Assessment of Renal and Cervical Artery Dysplasia - POLAND*), un registre national polonais de la DFM [205]. De janvier 2015 à décembre 2018, 343 patients, femmes et hommes âgés de plus de 18 ans ont été référés depuis 32 centres en Pologne et ont été évalués pour suspicion de la DFM à l'Institut de cardiologie de Varsovie, Pologne. Des lésions de DFM dans au moins un lit vasculaire ont été confirmées chez 232 patients par tomodensitométrie corps entier. Nous avons analysé 129 patients pour lesquels nous avons exclu la sténose athéroscléreuse et les syndromes où des lésions de type DFM sont souvent observées (par exemple, Ehlers Danlos, Loeys-Dietz). L'étude ARCADIA-Pol a été approuvée par le comité d'éthique local, Institut de cardiologie, IK-NPIA-0021017/1482/17.

- **WOBASZII**

Nous avons utilisé 298 témoins d'un échantillon déterminé au hasard de l'étude WOBASZ II (Enquête nationale multicentrique sur la santé de la population), une cohorte polonaise basée sur la population [206]. L'étude WOBASZ II a été conçue comme une enquête transversale auprès d'un échantillon aléatoire de résidents polonais âgés de plus de 20 ans. La sélection, à l'aide du Registre national des cartes d'identité du ministère de l'Intérieur, a été réalisée sous la forme d'un échantillonnage à trois degrés, stratifié selon les unités administratives (voïvodies), le type d'urbanisation (commune) et le sexe. Le protocole de l'étude consistait en un questionnaire utilisé lors d'entretiens en face à face, d'un examen physique et d'échantillons sanguins. WOBASZ II a été coordonnée par le Département d'épidémiologie, de prévention des maladies cardiovasculaires et de promotion de la santé de l'Institut de cardiologie de Varsovie en coopération avec les universités de médecine de Gdansk, Katowice, Cracovie, Lodz et Poznan. L'étude a été acceptée par le comité de bioéthique de terrain de l'Institut de cardiologie de Varsovie.

- **FMD European/International Registry and Initiative (FEIRI)**

Les patients ont été recrutés par le biais de l'initiative et registre européen/international sur la DFM (FEIRI) [207]. La DFM a été définie comme la présence d'une sténose idiopathique, segmentaire, non athérosclérotique et non inflammatoire (avec un aspect focal ou en chapelet)

d'une artère de petite ou moyenne taille dans au moins 1 lit vasculaire, documentée par tomographie, angiographie, angiographie par résonance magnétique ou imagerie angiographique par cathéter. Les patients suspects de DFM uniquement sur la base de l'échodoppler ont été exclus. Les patients dont le diagnostic principal était une dissection coronarienne spontanée n'étaient pas éligibles, même en présence de lésions extracoronaires de DFM. Les données sur les caractéristiques démographiques et la DFM ont été recueillies par le biais du registre de la DFM. Tous les centres inclus dans FEIRI ont reçu l'approbation des comités d'éthique locaux/nationaux respectifs.

- **ASKLEPIOS.**

L'étude Asklepios est une étude de population longitudinale axée sur l'interaction entre le vieillissement, l'hémodynamique cardiovasculaire et l'inflammation dans les maladies cardiovasculaires précliniques [208]. Les participants âgés de 35 à 55 ans ont été échantillonnés dans les communes belges jumelées d'Erpe-Mere et de Nieuwerkerken. Les critères d'exclusion étaient la présence de lésions athérosclérotiques et la concomitance avec une maladie grave, le diabète ou des conditions empêchant une évaluation hémodynamique précise. Tous les participants ont subi un examen physique systématique pendant une période continue de 2 ans, entre octobre 2002 et septembre 2004 sur un seul site d'étude à Erpe-Mere impliquant : mesure des données cliniques de base, prélèvement d'échantillons sanguins, examen échocardiographique et échographie vasculaire et mesures tonométriques. Le comité d'éthique de l'hôpital universitaire de Gand a approuvé le protocole d'étude.

- **Étude cas-témoin Mayo VDB.**

Les patients DFM et des témoins ont été identifiés à partir des maladies vasculaires Mayo Biodépôt (VDB) [209]. Les dossiers de santé électroniques de 209 patients ont été examinés manuellement par des cliniciens et le diagnostic a été confirmé pour 175 patients selon des rapports cités précédemment [72], dont 169 (83,4 % de femmes) avaient des données de génotypage à haute densité après retenus après l'étape de contrôle qualité (QC). Les témoins (n = 1141) étaient des participants sans DFM connue ni maladie vasculaire athéroscléreuse. L'étude a été approuvée par les comités d'examen institutionnels de la Mayo Clinic (IRB # 08–008355).

- **Étude DEFINE-DFM.**

Les cas éligibles étaient des femmes avec un diagnostic confirmé par imagerie de DFM multifocale et qui remplissaient d'autres critères diagnostiques acceptés [72], [210]. L'objectif de DEFINE-DFM était d'établir une bibliothèque de fibroblastes, d'ADN, de plasma et de sérum de patients atteints de DFM et de sujets témoins non affectés. Les cas de DFM devaient avoir un diagnostic clinique de DFM multifocale confirmé par imagerie [angiographie par tomodensitométrie (CTA), angiographie par résonance magnétique ou angiographie par cathéter]. Les témoins ont été appariés aux cas de DFM en fonction de l'âge et du sexe, devaient recevoir moins de 2 médicaments contre l'hypertension, avoir un indice de masse corporelle < 35 kg/m<sup>2</sup> et être non-fumeurs. Les critères d'exclusion pour tous les sujets comprenaient le sexe masculin, la DFM unifocale, l'utilisation d'agents immunosuppresseurs, les comorbidités majeures, la maladie de Crohn, la sclérose en plaques, ou toute autre forme de maladie vasculaire héréditaire (i.e, Ehlers-Danlos, Marfan, Loeys-Dietz) [210]. L'étude DEFINE-DFM a été approuvée par le comité d'éthique de la recherche humaine de l'école de médecine Icahn du mont Sinaï (ID d'étude : HS#13-00575/GCO#13-1118) et est enregistrée auprès de l'identifiant ClinicalTrials.gov : NCT01967511. Le consentement du patient n'impliquait pas la possibilité de redistribuer du matériel biologique et toute demande devait être adressée aux responsables de DEFINE-DFM.

- **Étude cas-témoin de l'Université du Michigan-Michigan Genomics Initiative/Cleveland Clinic.**

Les cas de DFM ont été recrutés à l'UM et/ou à la Cleveland Clinic. Les cas de DFM UM/CCF ont été recrutés dans une étude approuvée par l'IRB par l'intermédiaire d'une clinique de référence à l'UM et par auto-référence à l'étude. Les cas de la Cleveland Clinic ont été inscrits parmi des patients consécutifs vus dans une clinique de référence dédiée à la DFM. Le diagnostic clinique de la DFM a été établi par un spécialiste en médecine vasculaire après examen de l'imagerie diagnostique (tomodensitométrie, résonance magnétique ou angiographie par cathéter) et avant le prélèvement d'un échantillon de sang. L'ADN génomique a été isolé à partir d'un échantillon de sang périphérique ou de salive. La *Michigan Genomics Initiative* (MGI) est un programme qui a recruté des participants en attendant des procédures diagnostiques, interventionnelles et chirurgicales. Les participants ont fourni un échantillon de sang pour analyse génétique et ont accepté de lier leur échantillon à leur dossier de santé

électronique et à d'autres sources d'informations sur la santé [211]. Plusieurs codes CIM correspondant à des diagnostics d'anévrisme artériel, de dissection, de dysplasie et de sténose non athéroscléreuses ont été exclus, ainsi que les troubles du tissu conjonctif, comme décrit précédemment [150]. La variable hypertension a été définie comme tout code CIM contenant le terme « hypertension » pour les témoins MGI. L'étude de cas-témoins UM/Cleve Land a été approuvée par les protocoles IRB de chaque établissement : le numéro IRB de l'Université du Michigan est HUM00044507 et le numéro IRB de la Cleveland Clinic est le 10-318.

Country	Study	Type [status]	Total	Women	Age at inclusion (y)	Age of FMD diagnosis (y)	BMI (kg/m <sup>2</sup> )	HTN	FMD subtype MultiF/ Focal/ N/A	FMD location MultiS/ Unisite/ N/A
			n	n (%)	Mean ± SD [min -max]	Mean ± SD [min -max]	Mean ± SD	n (%)	n	n
France	ARCADIA	Clinical based [Cases]	643	533 (82.9)	54.05 ± 13.33 [19 - 89]	50.11 ± 14.27 [11 - 89]	23.81 ± 4.27	515 (80.1)	431/52/160	183/424/36
France	3 Cities study	Population based [Controls]	1487	876 (58.9)	74.36 ± 5.5 [65 - 94]	NR	N/A	1171 (78.7)	NR	NR
Poland	ARCADIA-POL	Clinical based [Cases]	125	100 (80.0)	45.73 ± 14.41 [18 - 75]	42.91 ± 14.95 [15 - 75]	22.13 ± 8.60	109	107/18/0	41/84/0
Poland	WOBASZ II	Population based [Controls]	295	227 (76.9)	40.97 ± 11.05 [21 - 86]	NR	26.37 ± 5.43	59	NR	NR
Europe	FEIRI	Clinical based [Cases]	344	281 (81.7)	N/A	48.27 ± 13.65 [5 - 85]	N/A	N/A	243/77/23	138/202/4
Europe	ASKLEPIOS	Population based [Controls]	615	611 (99.3)	45.67 ± 6.23 [15 - 58]	NR	N/A	N/A	NR	NR
USA	MayoVDB	Clinical based [Cases]	169	141 (83.4)	61.68 ± 13.56 [27 - 88]	N/A	26.84 ± 5.82	128 (75.7)	116	N/A
USA	MayoVDB	Clinical based [Controls]	1141	572 (50.1)	64.98 ± 11.30 [18 - 94]	NR	29.33 ± 6.02	680 (59.6)	NR	NR
USA	DEFINE-FMD	Clinical based [Cases]	108	108 (100.0)	56.13 ± 10.00 [32 - 80]	52.35 ± 10.49 [24 - 73]	24.13 ± 4.07	62 (51.4)	108/0/0	73/35/0
USA	DEFINE-FMD	Clinical based [Controls]	126	126 (100.0)	49.49 ± 11.53 [23 - 74]	NR	23.99 ± 3.86	N/A	NR	NR
USA	UM-FMD	Clinical based [Cases]	573	552(96.3)	53.39 ± 12.05 [11 - 84]	49.79 ± 12.19 [1 - 78]	25.13 ± 5.03	300 (52.4)	551/0/22	274/272/27
USA	UM-MGI	Population based [Controls]	3436	3312(96.4)	54.42 ± 12.46 [18 - 89]	NR	30.57 ± 9.23	1512 (44.0)	NR	NR

**Tableau 4.** Caractéristiques cliniques des populations étudiées [60].

BMI: Body-Mass Index, HTN: Hypertension, n: count, min: minimum range, max: maximum range, SD: Standard Deviation, y: years, MultiF: Multifocal, MultiS: Multisite, N/A: not available, NR: not relevant.

## **ii. Études SCAD**

### **▪ Étude cas-témoins DISCO-3C.**

L'étude DISCO a été enregistrée sous l'identifiant d'essai clinique : NCT02799186, et approuvée par le comité régional CPP (comité de protection des personnes) Sud-Est 6 2016 AU-1258. Protocole d'étude 3C approuvé par le comité consultatif de protection des personnes dans la recherche biomédicale Bicêtre Hôpital Bicêtre n°99-28 CCPPRB approuvé les 10/06/99, 11/03/2003 et 17/03/2006.

Les données d'un registre national Français des cas de SCAD (51 centres de cardiologie) ont été analysées de manière prospective et rétrospective. Tous les sujets âgés de 18 ans et plus ont donné leur consentement écrit et éclairé pour participer à l'étude. Critères d'exclusion inclus dans la garde ou la tutelle, les patients ayant une dissection iatrogène ou traumatique ou une dissection athéroscléreuse. La SCAD a été identifiée comme un syndrome coronarien aigu ou chronique défini selon les recommandations universelles avec des signes angiographiques évocateurs de SCAD [212] sur la coronarographie initiale. En contexte de MINOCA, une résonance magnétique cardiaque (RMC) pathologique pourrait conduire rétrospectivement au diagnostic de SCAD par la relecture des angiogrammes. En cas d'ambiguïté, une échographie intravasculaire (IVUS), une tomographie par cohérence optique (OCT) ou une angiographie coronarienne répétée a été réalisée à la discrétion de l'opérateur pour confirmer le diagnostic de SCAD. Tous les cas de SCAD ont ensuite été confirmés par un laboratoire central de 3 opérateurs (NC, PM, GS) expérimentés dans le domaine de la SCAD et de l'imagerie intra-coronaire. La segmentation coronaire a été définie selon la classification AHA [14]. Le dépistage de la DFM se faisait au choix des différents centres soit par scanner, résonance magnétique, angiographie ou Doppler dans les artères cervico-céphaliques, viscérales et iliaques. Le diagnostic de DFM était considéré comme confirmé si au moins une localisation extra-cardiaque était vérifiée par un radiologue expérimenté (LC) selon les critères établis [213] un dépistage complet était défini du cerveau au bassin.

### **▪ Études cas-témoins SCAD-UK.**

Etude UK SCAD (ISRCTN42661582) a été approuvée par *UK National Research Ethics Service (14/EM/0056)* et *UK Health Research Authority*. Tous les patients ont donné leur

consentement éclairé écrit pour participer à l'étude. Les patients atteints de SCAD ont été identifiés soit par référence du clinicien au service clinique national britannique de SCAD, soit par auto-référence à un portail en ligne. Seuls les patients atteints de SCAD confirmés par angiographie invasive ont été inclus. Les patients ayant des dissections coronaires athéroscléreuses, iatrogènes ou traumatiques ont été exclus. Les données démographiques ont été obtenues à partir du dossier médical des patients et du patient par le biais d'un sondage en ligne et d'un contact direct au besoin. L'hypertension a été définie comme étant sous traitement avant le premier événement clinique de SCAD. L'hypertension gestationnelle ou le diabète n'étaient pas inclus. Tous les angiogrammes ont été rapportés conjointement par au moins 2 cardiologues interventionnels expérimentés en SCAD. L'imagerie en coupe transversale par ARM ou CTA, réalisée à des fins de recherche ou cliniques, a été utilisée pour déterminer la présence ou l'absence d'artériopathies extra-coronaires associées à la SCAD. Ceux qui avaient la DFM, des anévrismes ou des dissections étaient considérés comme anormaux. Ceux qui n'en avaient pas étaient considérés comme normaux. La tortuosité ou la sténose focale n'étaient pas considérées comme anormales en l'absence de ces autres caractéristiques.

- **Étude cas-témoins de la Mayo Clinic**

Les participants à l'étude ont fourni un consentement éclairé écrit dans le cadre des protocoles de recherche clinique et génétique approuvés par le *Mayo Clinic Institutional Review Board* (NCT01429727 ; NCT01427179). Les sujets ont été recrutés parmi la population de patients de la Mayo Clinic, y compris les résidents locaux, les patients auto-référés et recommandés par un médecin, et les personnes qui ont contacté les investigateurs via le site Web de l'étude ([www.mayo.edu/research/SCAD](http://www.mayo.edu/research/SCAD)) et les médias sociaux. Les sujets de l'étude ont été inscrits consécutivement dans le registre SCAD de la Mayo Clinic après confirmation du diagnostic de SCAD par l'examen des angiographies coronariennes par un cardiologue interventionnel expérimenté. Les données démographiques et cliniques ont été extraites des questionnaires et des dossiers médicaux. La majorité des individus (95 %) vivaient aux États-Unis. Le dépistage de la DFM a été effectué chez un sous-ensemble de sujets par tomodensitométrie angiographie imagerie d'au moins deux lits artériels du cerveau au bassin.

- **Étude cas-témoins CanSCAD / MGI (UBC/MGI)**

Les approbations du comité d'éthique de la recherche ont été obtenues à chaque site d'inclusion de patients SCAD, et tous les patients ont fourni un consentement éclairé pour la participation. L'étude GenSCAD a été approuvée par l'IRB : HUM00113268, SCAD Registry. Les approbations du comité d'éthique de la recherche ont été obtenues (approbation de l'IRB : HUM00112101), l'analyse génétique de la dysplasie et du remodelage artériels (AMGI/AOS) et tous les patients ont reçu consentement éclairé à participer. La sous-étude génétique CanSCAD (N = 502) a inclus des patients atteints de SCAD de l'étude prospective de cohorte SCAD canadienne et de l'étude NACAD (*Nonatherosclerotic Coronary Artery Disease*). Des patients atteints de SCAD aigu ont été recrutés de manière prospective dans 22 sites en Amérique du Nord (20 sites au Canada et 2 aux États-Unis). Le diagnostic de SCAD a été confirmé par coronarographie par l'équipe de recherche du Laboratoire central de l'UBC. Des données démographiques de base détaillées, des antécédents ciblés pour les conditions pré-disposantes et les facteurs de stress précipitants, et un dépistage en laboratoire des conditions pré-disposantes ont été effectués. Le dépistage de la DFM a été recommandé pour tous les patients atteints de SCAD, et la DFM multifocale a été définie selon les directives consensuelles [213]. Les patients ont été suivis de manière prospective après la sortie à 1, 6 et 12 mois, puis annuellement pendant 3 ans pour les événements cardiovasculaires (CV).

- **Études cas-témoins VCCRI**

L'étude a été approuvée par le comité d'éthique de la recherche humaine de l'hôpital Saint-Vincent (HREC/16/SVH/338, numéro de protocole SVH 16/245) et menée conformément à la déclaration nationale sur la conduite éthique chez l'homme de l'*Australian National Health and Medical Research Council. Research and the CPMP/ICH Note for Guidance on Good Clinical Practice*. L'étude sur l'échantillon témoin du bras II a été approuvée par le comité d'éthique de la recherche humaine de l'hôpital St. Vincent (HREC/17/SVH/315) et menée conformément à la déclaration nationale sur la conduite éthique dans la recherche humaine de l'*Australian National Health and Medical Research Council* et au CPMP /Note d'orientation de l'ICH sur les bonnes pratiques cliniques. Numéro d'approbation éthique 2015.028 pour les échantillons du *Royal Melbourne Hospital*, Melbourne, Victoria, Australie. Les cas de SCAD ont été recrutés via une plateforme de médias sociaux ou par recommandation directe de cardiologues. Le diagnostic de SCAD a été confirmé par l'examen des images d'angiographie coronarienne

par un cardiologue interventionnel expert (DM) aveuglé par les résultats de l'analyse génétique. Le diagnostic de la DFM ou l'analyse négative a été auto-rapporté par les sujets. Les sujets étaient considérés comme souffrant d'hypertension s'ils recevaient des médicaments antihypertenseurs avant la survenue de la SCAD. Tous les sujets ont été interrogés sur les critères de diagnostic clinique des migraines. Les échantillons témoins du bras I de l'étude VCCRI proviennent de la cohorte MRGB [214]. Pour le bras II de l'étude VCCRI, les échantillons témoins étaient des cas de fibrillation auriculaire non apparentés et des patients atteints de cardiomyopathie dilatée non apparentés et des membres de la famille non affectés non apparentés.

- **Étude cas-témoin DEFINE (MS)**

Le protocole de l'étude DEFINE (NCT01967511) est approuvé par le comité d'éthique de la recherche humaine de l'*Icahn School of Medicine de Mount Sinai*. Les patients suspects d'artériopathies ont été vus et évalués à la clinique de médecine vasculaire du mont Sinaï. Les critères d'inclusion pour l'entrée dans DEFINE comprenaient des individus avec un âge supérieur ou égale à 18 ans, la volonté de participer librement et la maîtrise de l'anglais. Les cas de DFM devaient avoir un diagnostic clinique de DFM multifocale confirmé par imagerie [angiographie par tomodensitométrie, angiographie par résonance magnétique ou angiographie par cathéter]. Les patients SCAD devaient avoir un diagnostic clinique de SCAD confirmé par angiographie coronarienne invasive. Au départ, seules les femmes étaient autorisées à participer à l'étude, mais à partir du début de 2017, les hommes et les femmes étaient inscrits. Pour les témoins, les critères d'inclusion comprenaient l'absence de caractéristiques cliniques de DFM, de dissection de l'artère cervicale ou de SCAD (y compris l'absence de bruits cervicaux ou abdominaux, l'absence d'antécédents familiaux de mort subite ou d'anévrisme) et l'absence de toute maladie systémique majeure en cours, y compris toute condition nécessitant une hospitalisation, une immunosuppression, des médicaments intraveineux ou injectés ou qui entraînent une déficience fonctionnelle dans l'exécution des activités de la vie quotidienne. Des témoins ont été recrutés dans la population générale et ont été présélectionnés par la même équipe clinique et appariés aux cas d'étude en fonction de l'âge, du sexe, de la race/ethnie et de l'indice de masse corporelle (IMC). Des témoins ont été recrutés pour correspondre à l'âge, au sexe et à la distribution de l'IMC des cas. Les critères d'exclusion (pour les cas et les témoins) comprenaient : les comorbidités qui réduisent l'espérance de vie à 1 an ; toute transplantation

d'organe solide ou hématologique, ou ceux chez qui une transplantation est envisagée ; maladie auto-immune active ; consommation de drogues illicites ; séropositif ; malignité antérieure. Chez les témoins, un critère d'exclusion supplémentaire était l'apparition précoce d'antécédents familiaux de toute forme de maladie vasculaire. Les témoins ont également subi une évaluation clinique de dépistage, avec une attention particulière portée à tout antécédent ou résultat d'examen physique suggérant la DFM ou une autre maladie vasculaire, par deux experts cliniques de la DFM (J.W.O. et D.K.-D.).

## **b. Étapes majeures de la réalisation des GWAS Françaises**

Nous avons réalisé deux études d'association cas-témoin sur des populations Françaises, dans le but d'identifier de nouveaux SNPs sur le chromosome X, contribuant au risque de DFM et de SCAD. La collecte de ces données Françaises est considérée comme la première étape d'une étude pan-génomique. La procédure utilisée pour la collecte de ces données DFM et SCAD a été décrite en détails dans le chapitre 2. Dans cette partie, nous décrivons toutes les étapes suivantes pour la réalisation de ces deux études.

### **i. Génotypage**

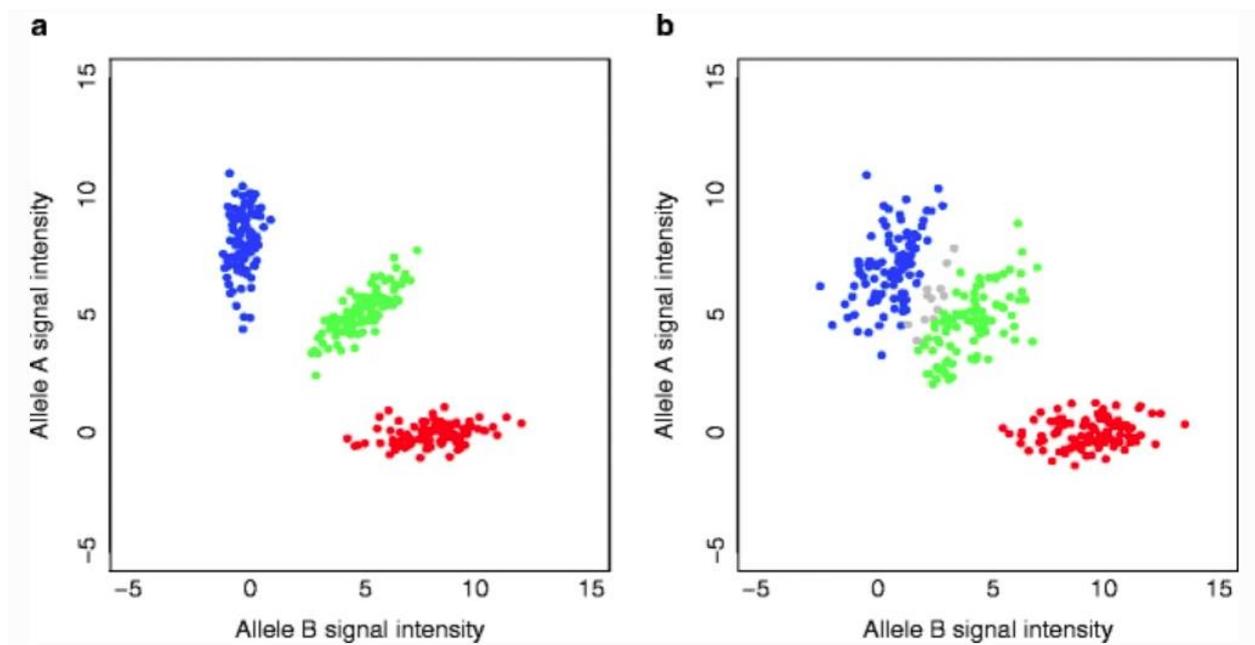
Une fois que les échantillons sanguins ont été collectés, l'étape suivante est de les génotyper en utilisant des puces à ADN. Le but de cette technique est de déterminer l'allèle d'un marqueur donné d'une variation génétique, à une position spécifique dans le génome pour un groupe d'individus donné.

Le principe des puces à ADN est basé sur la complémentarité des deux brins de l'ADN et donc de façon plus générale sur la technique d'hybridation, elle permet de mesurer pour chaque marqueur bi-allélique une intensité de fluorescence pour ses allèles a et A. Les différents génotypes possibles sont ensuite inférés par un algorithme mathématique, permettant d'attribuer un génotype aux groupes définis par l'intensité de chaque individu de l'échantillon à étudier. Si on représente graphiquement les deux intensités pour tous les individus, on voit apparaître, dans le meilleur des cas, trois groupes correspondant aux génotypes (deux homozygotes, un hétérozygote) (Figure 16).

Dans ces études Françaises de DFM et SCAD, les cas sont issus des études ARCADIA [203] et DISCO pour la DFM et la SCAD respectivement. Les données ont été génotypées avec

les puces Infinium OmniExpressExome-8v1.3, incluant 786 cas de DFM (24342 SNPs chromosome X), et 412 cas de SCAD (19363 SNPs pour le chromosome X). Pour les témoins, les données sont issues de l'étude 3C [204]. Elles ont été génotypées avec la puce Illumina-Human660W-Quad v1.0, incluant 1487 témoins (14962 SNPs pour le chromosome X).

Le génotypage à haut débit est susceptible de contenir des erreurs s'il n'a pas été fait sous de bonnes conditions, notamment la qualité de l'ADN, un mauvais transfert du tube d'ADN. Cela peut conduire à l'obtention de données génotypées de mauvaise qualité, entraînant une augmentation du nombre des résultats de faux positifs. Une analyse de contrôle qualité (QC) peut être effectuée pour éliminer les échantillons ou les marqueurs de faible qualité et minimiser l'impact de ces erreurs lors des analyses ultérieures [215].



**Figure 16.** Exemple de graphe d'intensité de signal d'allèle. Les groupes colorés représentent les individus avec les génotypes AA (bleu), AB (vert) et BB (rouge). (a) Les trois génotypes apparaissent distinctement et l'algorithme attribue correctement les génotypes. (b) En raison de chevauchement d'une partie des groupes AA et AB est classée en données manquantes par l'algorithme (en gris) [216].

## ii. Contrôle qualité

Le QC est une phase cruciale dans la réalisation des analyses d'association. Le QC consiste à nettoyer les données de génotypage en utilisant des critères classiques utilisés [216], [217]. Dans notre travail, nous avons réalisé un QC à l'aide de PLINK v2.0 [218]. Nous distinguons deux parties : le filtrage au niveau des individus et au niveau des SNPs.

### ▪ Etape 1 : QC sur les individus

#### • *Vérification du taux de génotypage*

Une mauvaise qualité de l'ADN ou un faible signal fluorescent empêche de déterminer le génotype, ce qui en résulte donc un nombre important de données manquantes (génotypes manquants). Dans notre étude, nous avons exclus les individus dont le taux de génotypage (CR pour call rate) est faible, c'est-à-dire  $CR < 99\%$ .

#### • *Identification des individus avec des informations sexuelles discordantes*

Ceci se fait en calculant le taux d'homozygotie sur le chromosome X. Ce taux est supérieur à 0,8 pour un homme et inférieur à 0,2 pour une femme. Si le taux d'homozygotie d'un individu s'écarte du taux d'homozygotie attendu selon les informations sur le sexe vérifiées, cela indique que l'information sexuelle est discordante, dans ce cas il faudra vérifier l'information et réassigner le sexe correspondant.

#### • *Identification des doublons.*

Une matrice « Identity By State » (IBS) est utilisée pour identifier les individus en double et apparentés. Ceci se fait en calculant pour chaque paire d'individus en fonction de la proportion moyenne d'allèles partagés en commun au niveau des SNPs génotypés. Pour notre étude, nous avons appliqué l'approche implémentée dans PLINK v2.0 [218], qui consiste à calculer la proportion d'allèles partagés (IBD) (allèles provenant d'un même ancêtre), en utilisant la méthode statistique de « chaîne de Markov cachée » pour estimer les probabilités suivantes :  $Z_0 = P(IBD = 0)$ ,  $Z_1 = P(IBD = 1)$ ,  $Z_2 = P(IBD = 2)$ . La proportion des allèles partagés par descendance entre deux individus est:

$$\alpha = \frac{2 \times Z_2 + 1 \times Z_1 + 0 \times Z_0}{2} = Z_2 + 0.5 \times Z_1 \quad (6)$$

Pour  $\alpha = 1$  signifie que l'individu est dupliqué (ADN envoyé deux fois) ou que les deux sujets sont des vrais jumeaux. Pour  $\alpha = \frac{1}{2}, \frac{1}{4}$  et  $\frac{1}{8}$  représentent un lien de parenté du premier, deuxième et troisième degré, respectivement.

- ***Stratification de population.***

Les personnes ayant une ethnie différente par rapport à la cohorte étudiée ont des différences systématiques de fréquence allélique. Cela entraînera la détection de grandes différences génotypiques dues à des sous-groupes de population. Par conséquent, ces individus sont considérés comme des outliers dont la fréquence allélique est aberrante par rapport à la fréquence de la population étudiée. Ces outliers doivent être identifiés et supprimés pour réduire la présence des faux positifs. C'est pour cette raison que le contrôle de cette stratification de population est considéré comme essentiel dans les étapes de QC. La stratification de population se fait en utilisant la méthode l'analyse en composantes principales (ACP). Elle est souvent appliquée pour identifier ses valeurs aberrantes. L'ACP est une méthode statistique descriptive multivariée, elle est utilisée pour transformer un ensemble de variables corrélées en un ensemble sensiblement plus petit de variables orthogonales et donc non corrélées appelées composantes principales. [215], [219]. Ces composantes principales sont présentées graphiquement sur des axes appelées vecteurs propres, de telle sorte que la première composante principale contient la plus grande variation possible des données en une seule composante. De la même manière, la deuxième composante principale maximise aussi la variance sous la contrainte d'être orthogonale à la composante principale précédente. L'ACP est devenu un outil extrêmement utilisé, car elle permet de structurer les populations à l'aide d'un faible nombre d'axes principaux, ainsi de déterminer les valeurs aberrantes. On définit un individu comme une valeur aberrante si sa structure génétique était à au moins 6 écarts-types de la moyenne des meilleurs composantes principales [220]. Dans notre travail, afin de réaliser nos analyses statistiques sur la population Européenne, nous avons appliqué la méthode de l'ACP sur nos cohortes à l'aide de PLINK v2.0 [221], en utilisant aussi 1000 Génome (incluant 28 ethnies) comme population de référence (<https://www.internationalgenome.org/category/phase-3/>), qui nous a permis d'obtenir un aperçu des grandes structures de population (européenne, asiatique, africaine). Un tableau récapitulatif de cette partie du QC sur les individus est donné ci-dessous (Tableau 5).

Country	cohort	status	genotyping array	QC steps				N included in GWAS
				N (genotyped)	Heterosigosity / Call Rate	Relatdness/ Duplication (IBD>0.185)	Non European excluded (PCA analysis)	
France	ARCADIA 3 Cities study	Cases	Infinium OmniExpressExome-8v1.3	786	27	59	54	646
		controls	Illumina-Human660W-Quad v1.0	1487	0	0	0	1487
POL	ARCADIA-POL WOBASZ II	Cases	Illumina-Infinium Global Screening Array-24 v2.0 +MD	129	0	4	0	125
		controls		298	3	0	0	295
Europe	FEIRI ASKLEPIOS	Cases	Illumina-Infinium Global Screening Array-24 v2.0 +MD	374	7	4	19	344
		controls		658	9	32	2	615
USA	MayoVDB MayoVDB	Cases	Illumina-Infinium-Human-Core-Exome	175	3	3	0	169
		controls		1177	16	20	0	1141
USA	DEFINE-FMD DEFINE-FMD	Cases	Illumina-Human-Omni-Express-Exome	134	1	7	18	108
		controls		164	5	8	25	126
USA	UM-FMD UM-MGI	Cases	illumina-Infinium CoreExome-24v1.1 BeadArray with 607,778 SNP markers (UM_HUNT_Biobank_v1-1_20006200_A)	584	0	11	0*	573
		controls		3438	0	2	0*	3436

**Tableau 5.** Informations détaillées pour les différentes étapes, génotypages, QC, QC-post imputation pour chaque étude DFM [60].

- **Etape 2 : QC sur les SNPs.**

- ***Taux de données manquantes.***

Un SNP mal géotypé implique une proportion élevée de données manquantes. Cela entraînera un taux élevé d'erreurs de géotypage. Nous n'avons gardé que les SNPs ayant un CR supérieur ou égale à 99%.

- ***Fréquence d'allèle mineur (MAF).***

La MAF est la fréquence de l'allèle mineur d'un SNP donné dans une population donnée. Elle est largement utilisée dans les études de génétique des populations, car elle fournit des informations permettant de différencier les variants communs et rares dans la population à partir d'un seuil donné. La MAF indique également la qualité de géotypage. En effet, les SNPs avec une MAF très faible sont également plus facilement affectés par une erreur de géotypage car l'allèle plus rare ne se produit que chez quelques individus. Comme les tests d'associations de l'approche GWAS sont sensibles aux erreurs de géotypages, nous avons appliqué un filtre sur nos deux études DFM et SCAD, et je n'ai gardé que les SNPs présentant une  $MAF > 1\%$ .

- ***Équilibre de Hardy-Weinberg (HWE).***

L'équilibre d'Hardy-Weinberg (HWE) est aussi utilisé dans le QC pour garantir une bonne qualité du géotypage. La théorie de HWE se repose sur l'hypothèse suivante : dans une population nombreuse et homogène avec un accouplement aléatoire, les fréquences alléliques et géotypiques demeurent constantes de génération en génération en l'absence d'autres influences évolutives. Par exemple, soit  $p$  la fréquence d'un allèle (A) et  $q$  la fréquence d'un allèle alternatif (a), alors les fréquences géotypiques attendues pour AA, Aa et aa sont :  $p^2$ ,  $2pq$  et  $q^2$ , respectivement. L'écart des fréquences observées par rapport à HWE peut être calculé à l'aide du test d'ajustement de  $\chi^2$ . Quand un SNP présente un écart important par rapport à HWE, cela est généralement dû aux erreurs techniques de géotypage. Dans la partie QC de notre étude, j'ai supprimé les SNPs de la cohorte témoin ayant une  $p$ -value du test de HWE inférieur au seuil de significativité  $10^{-5}$ .

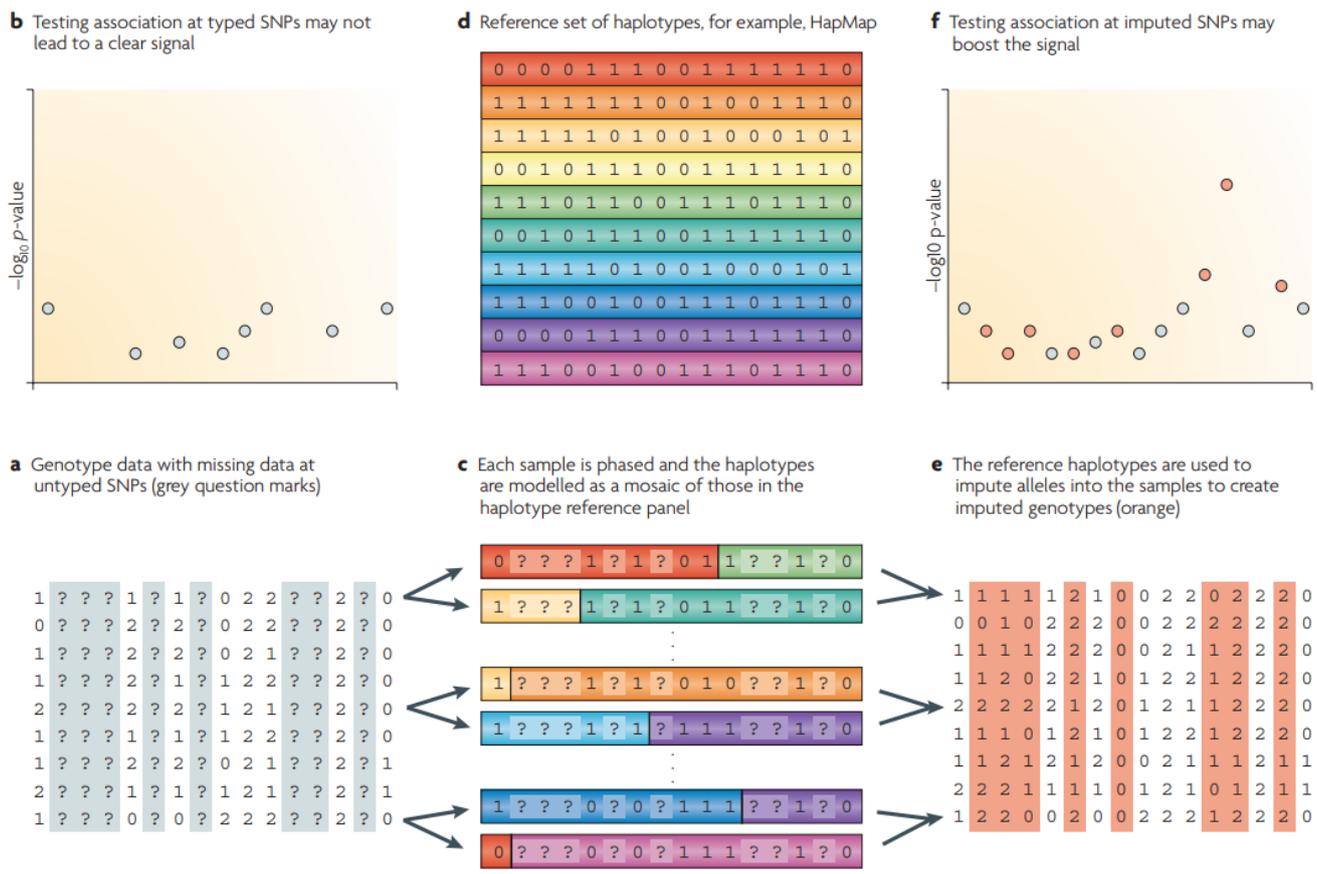
- **Etape 3 : Imputation**

Après avoir appliqué le QC sur nos données QC, l'étape suivante est l'imputation des données obtenues. Cette dernière est d'une grande importance dans les études d'associations. Elle permet d'augmenter considérablement la taille de l'étude et la puissance de l'analyse d'association. Le principe de l'imputation consiste à prédire le génotype de marqueurs non génotypés chez tous nos individus après QC [222], [223] (Figure 17).

Pour les cohortes Françaises, j'ai imputé des génotypes au panel de référence HRC v1.1 le plus récent [127] sur le Michigan Imputation Server [224]. Un des plus grands avantages des serveurs de l'imputation est la rapidité d'obtenir des données imputées en un temps très court. Un autre avantage de ces serveurs est l'utilisation de procédures de QC prédéfinies avant le phasage, telles que la vérification et la conversion de la construction du génome entre hg19 et hg38, l'inversion de brin, la conversion allélique, le filtrage des SNPs à faible taux de génotypage, la suppression des multi-alléliques. Il est également possible de sélectionner un autre panel de référence. Après l'imputation, les données imputées et les fichiers d'informations détaillées peuvent être téléchargés. De nos jours, les serveurs d'imputation les plus utilisés sont : le Michigan Imputation Server [224] et Sanger Imputation Server [225].

- **Etape 4 : QC post imputation**

Après l'imputation et la récupération des ensembles de données imputées, une autre étape de QC est mise en œuvre pour créer un ensemble de données de haute qualité et prêt pour l'analyse pour les études d'association à l'échelle du génome. Pour ce faire, j'ai procédé de la même façon que le QC présenté précédemment sur les SNPs uniquement, en plus du filtre spécifique à la qualité d'imputation. J'ai appliqué ce QC post-imputation sur l'étude Française de DFM et SCAD. J'ai supprimé les SNPs avec un taux élevé de données manquantes (supérieur à 0.01), une fréquence d'allèle mineurs ( $MAF < 0.01$ ), un écart par rapport à HWE ( $P < 10^{-5}$ ) et une faible qualité d'imputation ( $info < 0.8$  pour les cas et les témoins).



**Figure 17.** Imputation d'un génotype d'échantillon manquant à l'aide d'un panel d'haplotypes de référence [222]. La figure ci-dessus illustre l'imputation pour un échantillon d'individus non apparentés. Les données brutes consistent en un ensemble de SNPs génotypés qui comporte un grand nombre de SNPs sans aucune donnée de génotype (partie a). Le test d'association pour ces seuls SNPs peut ne pas conduire à une association significative (partie b). L'imputation tente de prédire ces génotypes manquants. Les algorithmes diffèrent dans leurs détails mais tous impliquent essentiellement le phasage de chaque individu de l'étude au niveau des SNPs typés. La figure met en évidence trois individus phasés (partie c). Ces haplotypes sont comparés aux haplotypes denses du panel de référence (partie d). L'alignement des brins entre les ensembles de données doit être effectué avant que cette comparaison n'ait lieu. Les haplotypes de l'étude en phase ont été colorés en fonction des haplotypes de référence auxquels ils correspondent. Cela met en évidence l'idée implicite dans la plupart des modèles de phasage et d'imputation que les haplotypes d'un individu donné sont modélisés comme une mosaïque d'haplotypes d'autres individus. Les génotypes manquants dans l'échantillon étudié sont ensuite imputés à l'aide des haplotypes correspondants dans l'ensemble de référence (partie e). Dans les exemples réels, les génotypes sont imputés avec incertitude et une distribution de probabilité sur les trois génotypes possibles est produite. Il est nécessaire de tenir compte de cette incertitude dans toute analyse en aval des données imputées. Tester ces SNPs imputés peut conduire à des associations plus significatives (partie f) et à une vue plus détaillée des régions associées.

## **c. Méthodologies appliquées pour l'analyse des variants du Chromosome X**

### **i. Principe et contexte de l'étude des chromosomes sexuels**

Comme nous l'avons vu dans le chapitre 2, les MCV se manifestent différemment chez les hommes et les femmes [34], [38]. Le sexe est donc un facteur de risque pour de nombreux types de MCV, et son rôle génétique (chromosomes XX versus XY) dans les MCV a été difficile à évaluer chez l'homme [226], en partie à cause de l'incapacité de démêler les génotypes XX et XY de la présence du type de gonade correspondant [226]. Une introduction à l'inactivation du chromosome X est nécessaire afin de mieux comprendre ce processus.

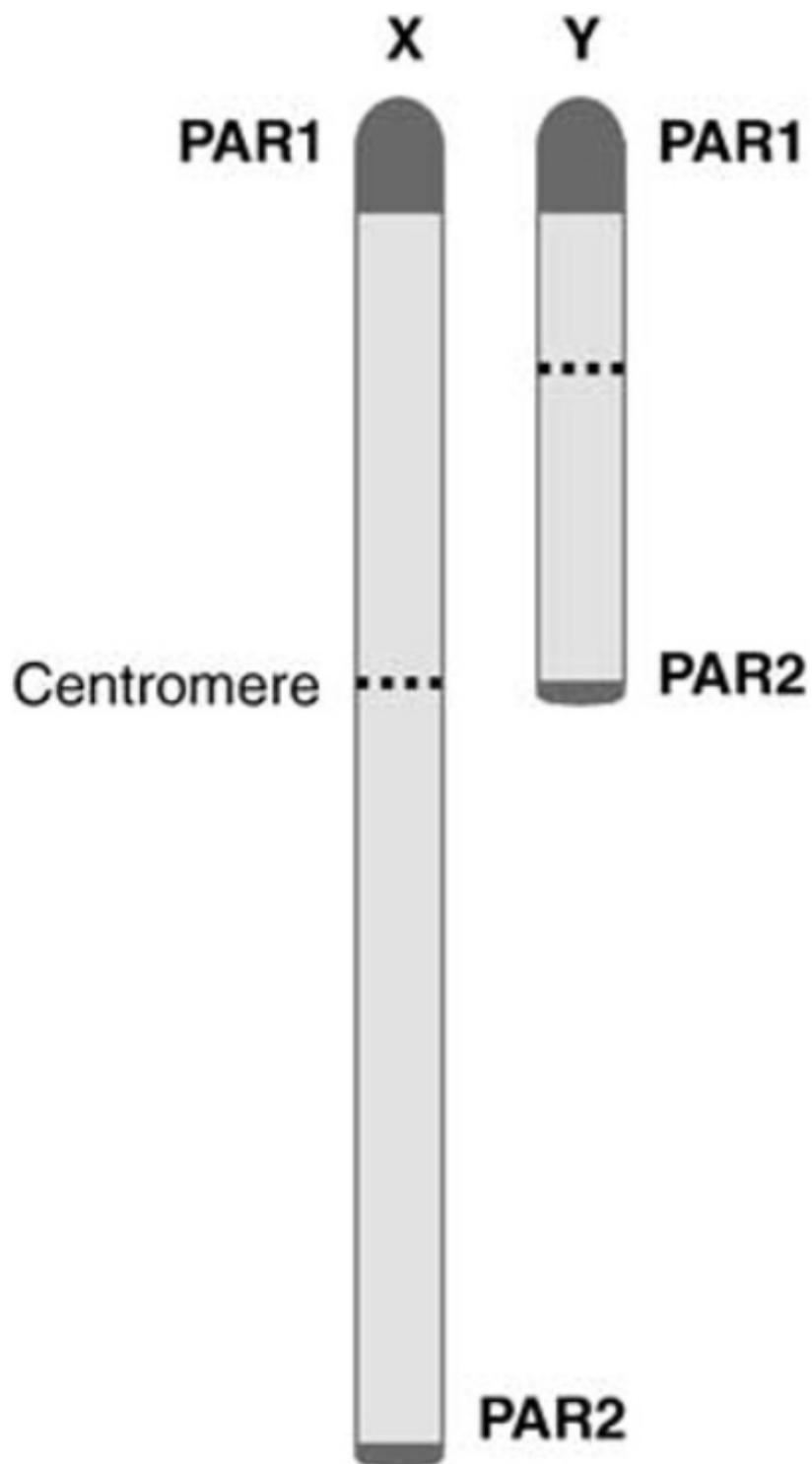
### **ii. Rappel du précepte et mécanisme de l'inactivation du chromosome X**

Toutes les particularités des chromosomes sexuels des mammifères, y compris leur remarquable enrichissement en gènes impliqués dans les gonades et les gamètes, résultent de leur représentation différente chez les mâles et les femelles [227]. Ces deux chromosomes seraient issus d'une paire d'autosomes et auraient divergé il y a environ 300 millions d'années [228]. Au cours de cette évolution, le chromosome Y aurait subi divers remaniements, dont la perte de la majorité de ses gènes fonctionnels. De plus, les gènes conservés auraient muté, ce qui aurait conduit à l'émergence du chromosome déterminant le sexe masculin. Même si les chromosomes X et Y sont très différents, ils peuvent encore subir des recombinaisons, spécifiquement dans les régions situées aux extrémités, nommées régions pseudo-autosomales (RPA) [229]. Les RPA sont les sections des chromosomes sexuels (X et Y) qui se comportent comme des autosomes, c'est-à-dire qu'elles y partagent des gènes (voir la Figure 17).

Chez les femelles, la présence de deux copies du Chr X (XX) par rapport au mâle (XY), celui-ci contient environ 70 gènes, alors qu'on trouve environ 900 à 1500 gènes dans le Chr X [230]. Il en résulte donc un déséquilibre entre les deux sexes quant à la quantité de gènes codés par les gènes du Chr X [231], donc de produits géniques (ARN et protéines), que leurs cellules contiennent.

L'absence d'une copie du Chr X chez les hommes est associée avec des maladies génétiques récessives reconnues, par exemple l'hémophilie et le daltonisme. Cependant, en dehors de ce contexte de maladies très rares, il n'est pas tout à fait clair comment le dosage génétique représenté par 2 chromosomes chez les femmes et 1 seul chez les hommes pourrait

influencer le risque génétique pour des maladies polygéniques et multifactorielles, notamment en tenant en compte du phénomène de l'inactivation somatique du ChrX. Il s'agit d'un mécanisme qui permet de s'assurer d'établir des niveaux équivalents d'expression pour les gènes liés au Chr X entre les hommes et chez les femmes dans de nombreux tissus somatiques [230], [232]–[234].



**Figure 18.** Structure des chromosomes sexuels masculins humains et localisation des deux régions RPA1 (PAR1 en anglais) et RPA2 (PAR2 en anglais) [235].

## ▪ Définition de l'inactivation de X

L'Inactivation du Chr X (ICX) est un processus d'une importance capitale chez les mammifères femelles, il aboutit à la compensation de doses des gènes liés à l'X entre les deux sexes. Ce dernier se produit au début du développement embryonnaire féminin [236]–[238]. Un des deux chromosomes X passe, durant le développement embryonnaire précoce, d'un état euchromatique actif (structure observable non condensée de la chromatine) à un état hétérochromatique inactif (structure condensée de la chromatine). L'ICX commence peu de temps après l'implantation de l'embryon [239].

C'est en 1961 que la généticienne anglaise Mary Lyon a pour la première fois décrit ce phénomène biologique fondamental de l'inactivation du Chr X [240]. Elle avait observé chez les souris femelles des phénotypes inattendus « en mosaïque » de la couleur de leur pelage – contrairement aux mâles, ainsi que dans les cellules diploïdes, un des 2 chromosomes X se trouvait sous forme d'hétérochromatine nommé « corpuscule de Barr » [241]. Plus précisément, elle a émis l'hypothèse « Il est ici suggéré que ce phénotype de mosaïque est dû à l'inactivation de l'un ou l'autre Chr X au début du développement embryonnaire. » [236], [242]. Ce phénotype mosaïque est composé d'un mélange de cellules qui expriment le Chr X paternel (Xp) et de cellules qui expriment le Chr X maternel (Xm). Normalement, chez une femelle, la moitié des cellules expriment Xp et l'autre moitié expriment Xm. Cependant, de nombreuses femelles présentent un ratio d'inactivation Xp/Xm favorisant l'un des chromosomes (par exemple 75/25 ou 25/75). On évoquera un peu plus tard le rôle du ratio d'inactivation.

Lyon a ainsi suggéré (1) que ce corpuscule de Barr était un Chr X génétiquement inactif, (2) que, dans différentes cellules d'un même animal, le chromosome inactivé pouvait être soit d'origine paternelle ou maternelle et (3) que cette inactivation se produisait tôt dans le développement embryonnaire [243]. Ceci explique qu'au début du développement embryonnaire, soit le Chr Xp, soit le Chr Xm est inactivé dans chaque cellule et que cet état inactif est alors propagé à travers les divisions cellulaires conduisant à la mosaïque cellulaire chez les femelles pour l'expression des gènes liés à l'X.

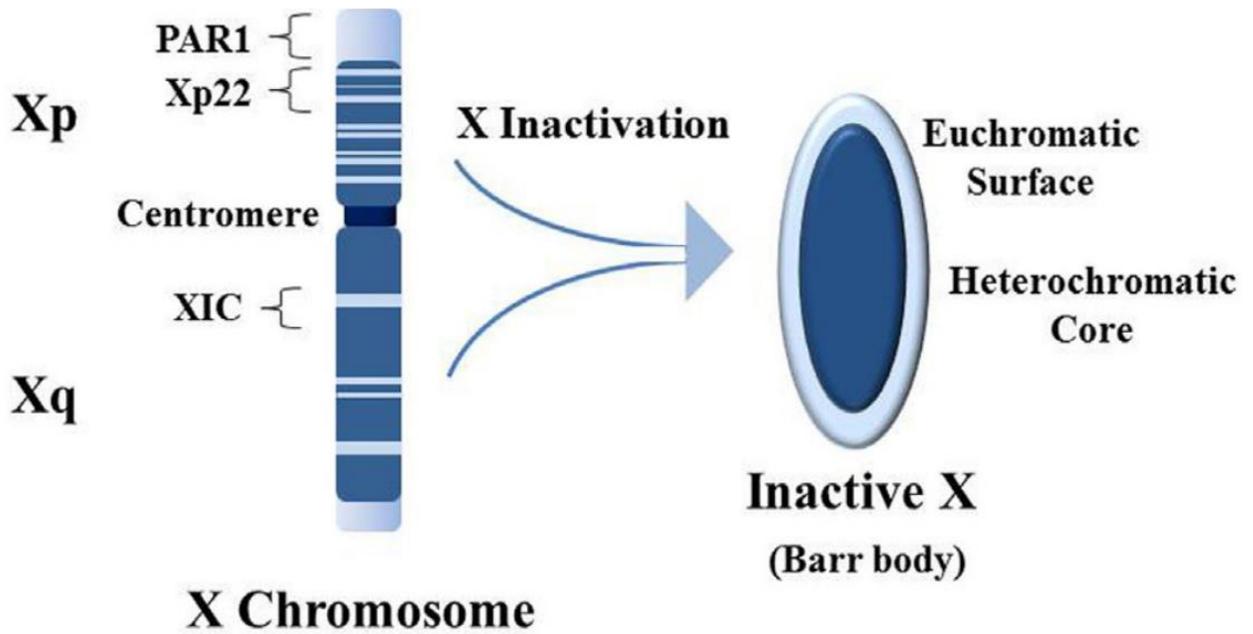
- **Quelques aspects du mécanisme de l'inactivation du Chr X (ICX)**

L'aspect aléatoire du mécanisme de l'ICX implique plusieurs contraintes pour sa réalisation. Pour se faire, quelques aspects de sa mise en place et de sa maintenance peuvent être indiqués [244]–[246].

- **Mise en place de ICX**

Dans les années 1970, Heard *et al* (2001) [246] ont mené des travaux de recherche sur des souris pour mieux comprendre le mécanisme de l'ICX. L'inactivation d'un des deux chromosomes X dans toutes les cellules d'une femme se produit durant le développement, à un moment encore précisément indéterminé. Ces chercheurs ont découvert un locus sur le Chr X long de plusieurs millions de bases, connu sous le nom de centre d'inactivation de l'X (Xic). En supprimant ce centre, cela empêchait la mise en place de l'ICX dans des embryons femelles de souris. Puis, en 1991, il y a eu la découverte du gène nommé Xist (*X-inactive specific transcript*), situé dans le locus Xic [247], [248].

Xist est considéré comme un gène majeur de l'ICX, et s'exprime de façon aléatoire soit sur le chromosome Xp, soit sur le Xm. Il gouverne la synthèse d'un ARN non codant (c'est-à-dire ne donnant pas une protéine) qui recouvre l'ensemble du Chr X dont il est issu voir [246], [249]. Cet ARN Xist empêche ainsi l'expression des gènes de ce chromosome en le rendant entièrement inactif, et produisant donc le corpuscule de Barr (Figure 19). Cette inactivation est exercée dans toutes les cellules, donnant lieu à un individu mosaïque quant à l'origine du Chr X inactivé et elle est maintenue tout au long de la vie de l'individu. Toutefois, un autre ARN non codant, Tsix (le transcrit anti-sens de Xist) peut aussi bloquer la production de Xist ; un phénomène d'opposition complexe entre ces deux ARN non codants gouverne l'extinction du X. Heard et ses collègues [245] ont observé que chez les souris, le chromosome Xp est toujours éteint au début du développement, puis réactivé ; c'est seulement dans un deuxième temps que se produit l'inactivation aléatoire du X aboutissant à une mosaïque de cellules, certaines exprimant le Xp, les autres le Xm. Les biologistes ont en outre montré que le choix du Chr X à éteindre varie d'une espèce de mammifères à l'autre. Chez le lapin et l'homme, le Xp n'est pas inactivé pour être réactivé (avant que l'inactivation aléatoire d'un des chromosomes X ait lieu), et le gène Tsix ne semble pas exister.



**Figure 19.** Processus de l'inactivation du Chr X et la formation du crépuscule de Barr [250]. Pendant le processus de l'inactivation de X, Environ 95 % des gènes du bras long (Xm) et 65 % des gènes du bras court (Xp) sont réduits au silence. Les gènes silencieux sont représentés en bleu foncé, tandis que les gènes qui échappent à l'inactivation sont représentés en bleu clair. Le résultat est le corps de Barr qui apparaît comme une structure hétérochromatique dense près de la membrane nucléaire.

De nombreuses maladies humaines couramment étudiées dans les études d'association pan-génomiques présentent un dimorphisme sexuel (MCV [251], maladies auto-immunes [252]), ce qui suggère une contribution potentielle du chromosome X. Une attention particulière doit être accordée à son analyse en raison de son schéma d'héritage unique, ce qui entraîne des complications analytiques qui font que la majorité des études d'association pan-génomique (GWAS) ne prennent pas en compte le Chr X ou le manipulent mal avec des outils conçus pour les chromosomes non sexuels. Gao et ses collaborateurs [253] ont développé l'approche X-WAS pour répondre au besoin d'outils spécialement conçus pour l'analyse de l'X. Cette approche sera abordée avec plus de détails dans la section ci-dessous.

### **iii. Méthodologie X-WAS**

#### **▪ Définition**

XWAS est un logiciel spécialement conçu pour l'analyse du Chr X dans les études d'association [253]. Il a été implémenté en tant que version étendue de PLINK [218] appelée PLINK/XWAS. Au-delà de ces extensions de PLINK, XWAS inclut plusieurs scripts shell et Perl qui rassemblent les différentes fonctionnalités dans un pipeline qui effectue une analyse intégrée.

L'approche XWAS peut effectuer divers tests statistiques, notamment le test d'association entre un SNP et le risque de maladie, un test pour un effet différentiel d'un SNP sur la maladie entre les hommes et les femmes, motivé par l'inactivation du X. XWAS comprend également des procédures de contrôle qualité pour les données liées au chromosome X.

#### **▪ Contrôle qualité**

L'approche XWAS implémente un processus complet pour effectuer un QC sur des données liées au Chr X [253]. Ce processus se compose de deux parties de QC.

La première partie, consiste à appliquer les étapes QC standard de GWAS présenté dans la section précédente. Ces étapes sont effectuées séparément sur l'ensemble des hommes uniquement et des femmes. Elles incluent l'élimination des individus et des SNPs selon plusieurs critères [253].

La deuxième partie du processus applique des étapes de QC spécifiques au chromosome X, qui sont exclusivement intégrées dans XWAS V3.0 [254]. Elles sont effectuées sur les hommes et les femmes ensemble après l'étape du QC général, et incluent :

- La suppression des SNPs avec des MAF significativement différentes entre les échantillons masculins et féminins du groupe témoin (le seuil de *p-value* recommandé est 0,05) ; cela revient à filtrer les variants qui ne respectent pas à l'équilibre de HWE chez les femelles.
- La suppression des SNPs avec des taux de données manquantes significativement différents entre les contrôles masculins et féminins (le seuil de *p-value* recommandé est 0,05).
- La suppression des SNPs dans les régions pseudo-autosomiques (PAR).
- Les 2 premiers points de cette étape de ce QC capturent les problèmes de génotypage lorsque les plaques incluent à la fois des mâles et des femelles [255].

#### ▪ **Tests d'association avec XWAS**

Comme je l'ai mentionné précédemment, l'approche XWAS inclut plusieurs tests d'association liés à l'X. Dans mon travail, j'ai utilisé le test d'association d'un SNP sur le chromosome X, ainsi le test stratifié par sexe à marqueur unique sur le chromosome X. Ces deux tests sont définis comme suit :

#### ▪ **Test d'association à un seul marqueur sur le chromosome X**

Pour un SNP lié à l'X, les femelles ont 0, 1 ou 2 copies d'un allèle, alors que les mâles ne portent que 0 ou 1 copie. Si les loci subissent une inactivation complète de X, le Chr X produit une expression mono-allélique des gènes codant pour les protéines liées à l'X chez les femelles. Par conséquent, on considère que les mâles sont codés en 0/2 correspondant aux homo-zygotes femelles. Le test d'association adéquat pour un seul SNP codé en 0/2 est effectué en utilisant l'option `--xchr-model 2` dans PLINK [218].

Pour d'autres scénarios, par exemple lorsque différents gènes échappent à l'inactivation de X, il est plus approprié de coder les mâles comme ayant des allèles 0/1. Dans ce cas, le test d'association entre un seul SNP et le risque de maladie est effectué en utilisant les options

suivantes dans PLINK [218]: *--logistic* et *--linear* pour les traits binaires et quantitatifs, respectivement [253].

- **Analyse stratifiée par sexe à marqueur unique sur le chromosome X**

Le logiciel étendu, PLINK/XWAS, fournit de plus un nouveau test stratifié par sexe qui n'existe pas dans PLINK. Ce test est particulièrement pertinent pour les analyses liées à l'X, car les SNPs et les locus présents sur les chromosomes sexuels sont potentiellement plus susceptibles de présenter des effets différents sur le risque de maladie entre les hommes et les femmes [253]. Cette fonctionnalité est accessible par l'option *--stratsex*. Il effectue dans un premier temps un test d'association chez chacun des hommes et des femmes séparément, puis il combine les résultats des 2 tests en utilisant la méthode de Fisher (*--fishers*) [256] pour obtenir un niveau de signification final stratifié par sexe. Étant donné que le test chez les hommes est séparé, le test stratifié selon le sexe est insensible au fait que les mâles soient codés comme 0/2 ou comme 0/1, ne faisant ainsi aucune hypothèse concernant le statut d'inactivation de X. De plus, ce test peut être plus puissant dans les situations où un allèle a des effets opposés entre les hommes et les femmes en association avec le trait testé, ainsi que lorsque l'effet n'est observé que chez un sexe [253].

## 2. Résultats

Dans le but d'identifier de nouveaux variants associés à la DFM et à la SCAD, j'ai utilisé les méthodes spécifiques aux analyses d'associations sur les variants génétiques du chromosome X. Ces analyses ont été effectuées sur la cohorte Française (ARCADIA-3C pour DFM et DISCO-3C pour SCAD), la cohorte Polonaise (ARCADIA-POL) et la cohorte Belge. De plus, deux méta-analyses ont été réalisées en ajoutant d'autres études GWAS issus des collaborations internationales Européennes et Américaines pour les phénotypes DFM et SCAD. Les détails de ces analyses ainsi que leurs résultats sont présentés dans la suite de cette section.

### a. Étude de DFM

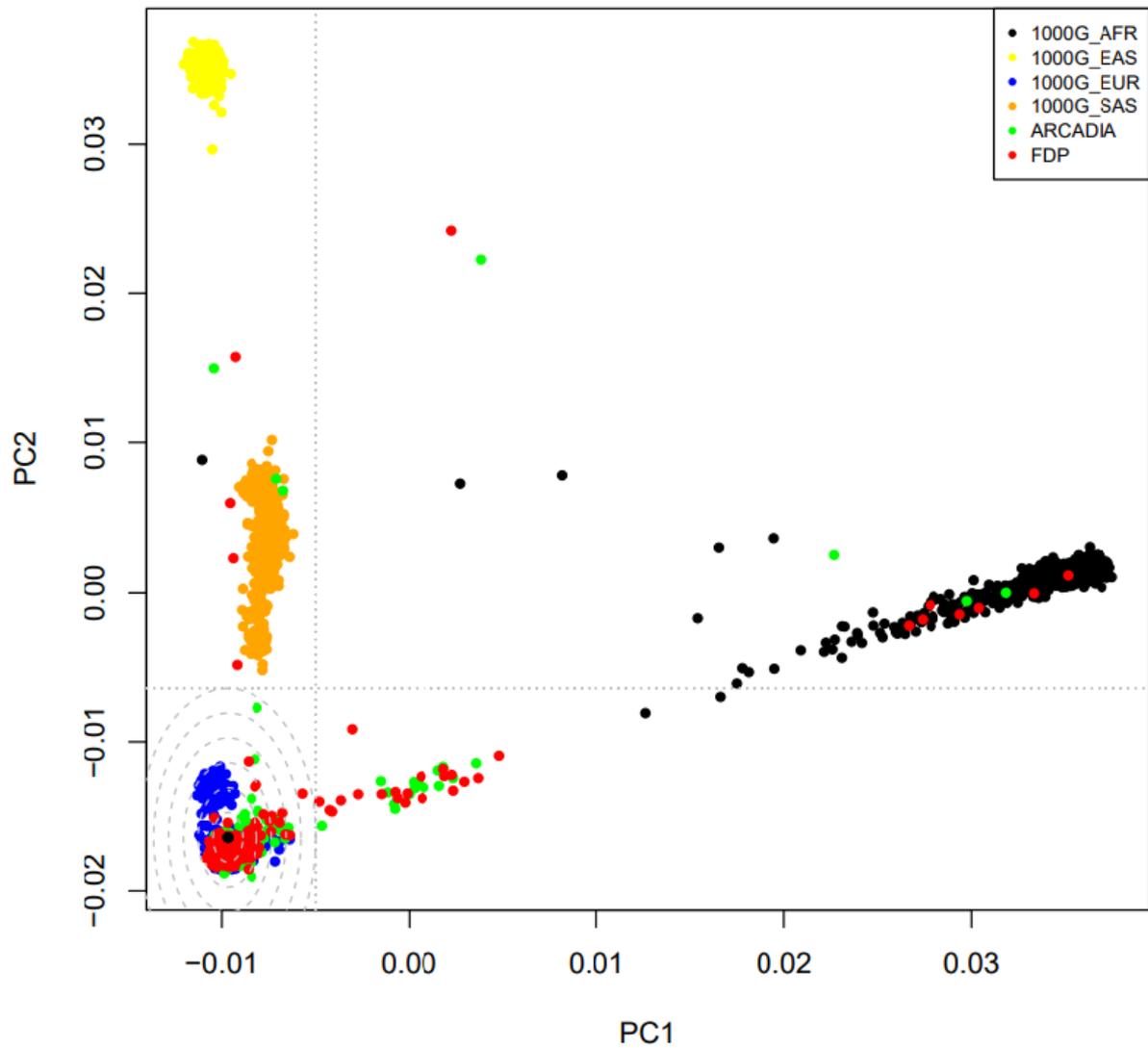
#### i. Modèle 1 : sans l'inactivation de X

Dans cette première analyse, j'ai réalisé des études d'associations sur nos trois cohortes cas-témoins en utilisant le modèle avec le codage classique sans la particularité de l'inactivation

de X, que je note **modèle 1**. Pour ce faire, j'ai utilisé la méthode GWAS avec PLINK v2.0 [221] qui permet de lire les fichiers binaires de très grande capacité. Les résultats obtenus sont présentés par cohorte comme suit.

- **QC standard des données de génotypage**

Les différentes étapes du QC des données de génotypage que j'ai effectué sur les trois cohortes cas-témoins sont données en détails dans le tableau 2. J'ai appliqué les mêmes critères de QC vu précédemment sur l'ensemble de données contenant tous les chromosomes en utilisant PLINK v2.0 [221]. J'ai d'abord exclu les individus et les variants ayant un taux des données manquantes  $> 1\%$  (*--geno 0.01*). J'ai vérifié s'il existe information discordante sur le sexe par rapport au sexe déclaré, ainsi que le taux d'hétérozygotie ne respectant pas les valeurs normales pour les hommes (taux entre 0.2 et 1) et les femmes (taux  $< 0.2$ ). J'ai également exclu les doublons et les individus avec un lien de parenté au premier degré, et enfin, en utilisant l'ensemble des individus retenus pour l'analyse après les différents filtres appliqués, j'ai réalisé une ACP (*--pca*), avec les populations de 1000 génomes (1KG) phase3 (<https://www.internationalgenome.org/category/phase-3/>) dans le but de garder la population Européenne uniquement (Figure 20). Ensuite j'ai supprimé les variants avec une MAF inférieure à 1% (*--maf 0.01*), ceux qui présentent des écarts par rapport à l'équilibre de HWE ( $P < 10^{-5}$ ). Il est important de mentionner que pour l'étude Française, le QC a été fait sur les cas et les témoins séparément, car les données n'ont pas été génotypées avec la même puce.



**Figure 20.** ACP du GWAS de la DFM avec les populations de 1000 génomes Phases 3. 1000G\_AFR = Africain; 1000G\_EAS = Asie de l'Est; 1000G\_EUR = Européen; 1000G\_SAS = Asie du sud; ARCADIA= les temoins de notre échantillon ; FDP = notre échantillon DFM

- **Imputation et QC post-imputation**

Une fois que l'étape de QC est faite, j'ai filtré l'ensemble des données des trois cohortes afin de ne garder que celles du chromosome X, à l'aide de la fonction `--chr 23` de PLINK v2.0 [221]. Les fichiers résultants contenaient 18 157 SNPs et 14 513 SNPs pour l'étude ARCADIA et 3C respectivement. J'ai aussi obtenu 15 651 SNPs pour l'étude Polonaise et 21680 SNPs pour l'étude Belge. L'étape suivante consistait à imputer les SNPs manquants afin d'augmenter la couverture génétique dans les échantillons étudiés. Pour cela j'ai utilisé l'algorithme minimac4 intégré dans Michigan Imputation Server [224], en utilisant le panel de référence HRC v1.1 le plus récent [127]. Par la suite, j'ai utilisé le bcftools v1.9 (<https://samtools.github.io/bcftools/bcftools.html>) et le vcftools [257], d'une part pour le contrôle qualité après l'imputation pour supprimer les variants avec une mauvaise qualité d'imputation, c'est-à-dire  $\text{info} < 0.8$ . D'autre part pour rassembler les données imputées d'ARCADIA et 3C dans le même fichier, afin de pouvoir appliquer le test d'associations.

Après avoir appliqué le QC sur les données imputées, les fichiers obtenus ont inclus les nombres de SNPs suivant : 410397 SNPs, 488809 SNPs et 461973 SNPs pour la cohorte Française, Polonaise et Belge respectivement. La dernière étape de cette analyse consiste à tester l'association entre chaque variant du Chr X et la DFM.

- **Test d'association**

Le but de mener des études cas-témoins est d'estimer l'association entre chaque SNP et le phénotype étudié. A cet effet, j'utilise la régression logistique avec l'hypothèse que les données sont indépendantes et identiquement distribuées. Cette méthode teste si la fréquence de l'allèle est significativement modifiée entre le groupe cas et le groupe témoin. En terme statistique, cela revient à tester si la distribution observée n'a pas d'influence sur le phénotype étudié. Le principe du test d'adéquation utilisé est de calculer l'odds ratio, qui est le rapport de deux cotes. L'odds ratio est la probabilité qu'un individu porteur d'un allèle particulier développe la maladie versus la probabilité qu'un individu ne possède pas le même allèle. Lorsque la fréquence de l'allèle chez les cas est beaucoup plus élevée que chez les témoins, le rapport de cotes est supérieur à 1. De plus, une valeur de P pour la significativité de l'OR est typiquement calculée en utilisant un test du  $\chi^2$  basé sur l'hypothèse nulle ( $H_0$ ) d'absence

d'association du SNP avec le phénotype étudié, contre l'hypothèse alternative présentant la présence de l'effet significatif du SNP sur le phénotype.

$$H_0: \beta = 0 \text{ (Effet nul du SNP sur le phénotype étudié)}$$

$$H_1: \beta \neq 0$$

Dans notre étude d'association, j'ai utilisé le modèle logistique mixte qui permet d'intégrer des facteurs de confusion potentiels (covariable) tels que l'ethnie, le sexe, l'âge, etc. La prise en compte d'une ou plusieurs covariables va permettre d'ajuster l'association et de contrôler une partie de la variation de la variable indépendante d'intérêt.

Si l'on considère  $Y$  un phénotype binaire (0 pour non malade et 1 pour malade) mesuré sur  $n$  individus,  $X$  la matrice de  $n \times p$  covariables sur lesquelles on souhaite ajuster le modèle.  $G_i$  est le vecteur de génotypes (habituellement codé 0, 1 ou 2), le modèle linéaire s'écrit comme suit [258] :

$$\text{logit}P(Y_i = 1) = \alpha_i + \beta_i G_i + \gamma_i X_i \quad (7)$$

Avec :

$\alpha_i$  est un vecteur représentant l'effet aléatoire de l'individu  $i$ , sa distribution suit une loi  $N(\mathbf{0}, \sigma)$ , avec  $\sigma$  la matrice de corrélation génétique des individus.

$\beta_i$  et  $\gamma_i$  correspondants à des effets fixes.  $\beta_i$  est l'effet du  $i$ -ème SNP que l'on cherche à estimer et  $\gamma_i$  est l'effet associé aux covariables.

Comme mentionné au début de cette section, le premier modèle d'association (modèle 1) que j'ai utilisé ne tient pas compte de l'inactivation de X, c'est-à-dire que pour la covariable sexe, les doses masculines sont sur une échelle de 0..1 sur le chromosome X, tandis que les femmes sont de 0..2, c'était la valeur par défaut de PLINK 1.9 [218]. Pour l'étude d'association des SNPs du Chr X et la DFM, j'ai appliqué le **modèle 1** en utilisant la fonction `--glm`, en choisissant `--xchr-model 1`, qui est aussi implémenté dans PLINK v2.0 [221]. Tous les tests d'association ont été ajustés sur le sexe et les 20 premières composantes principales. Les résultats obtenus sont donnés dans un grand tableau sous forme d'un résumé statistique,

contenant toutes les informations statistiques calculées par le test pour chaque SNP. L'interprétation de ces tests d'associations se base sur la valeur de la *p-value* obtenue pour chaque SNP. Cette *p-value* est comparée au seuil de significativité à l'échelle du génome qui est fixé  $5.10^{-8}$  (soit 1 million de tests indépendants au seuil de 5%) [259], permet d'accepter ou de rejeter l'hypothèse  $H_0$ .

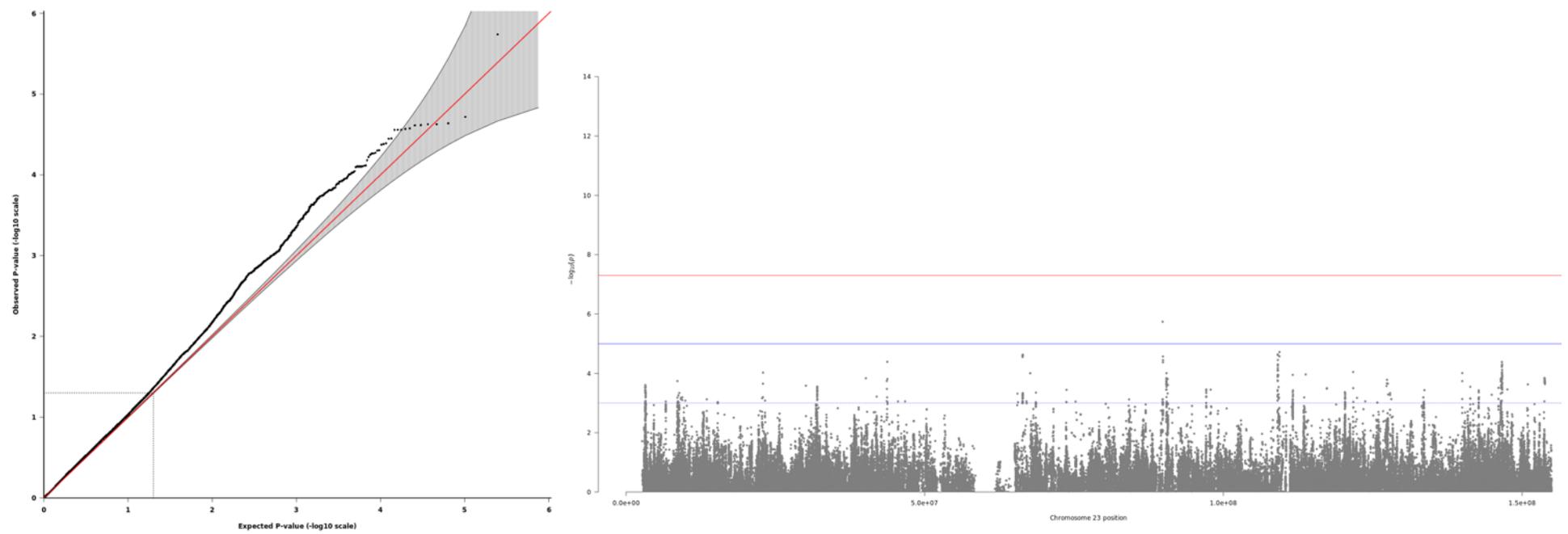
Les résultats des tests sont présentés graphiquement sous forme d'un QQ plot et d'un Manhattan plot, un petit rappel de l'utilité de ces deux graphiques est nécessaire.

Le **QQ plot** est un diagramme de Quantiles-Quantiles qui permet de comparer la distribution des *p-values* observées (axe des ordonnées) à la distribution des valeurs de *p-value* attendues (axe des abscisses) (sous l'hypothèse  $H_0$ , les valeurs de *p-value* suivent une distribution de loi uniforme sur  $[0,1]$ ). Soit  $y = x$ , la droite de regression représentant la distribution de *p-value* sous l'hypothèse d'absence d'association  $H_0$ . On considère que le modèle est bien ajusté si les valeurs de *p-values* observées sont bien alignées représentant la distribution théorique. Cependant, s'il existe une déviation de certains points de cette droite, cela signifie la présence des signaux d'association.

Le **Manhattan plot** quant à lui, permet de visualiser les résultats d'association en représentant les valeurs de  $-\log_{10}(p\text{-values observées})$  en ordonnée (axe des y), en fonction de la position chromosomique en abscisse (axe des x). Cette présentation graphique est très utile pour identifier les SNPs les plus significatifs, de sorte qu'ils apparaissent en hauteur traversant la droite horizontale présente sur le graphe selon le seuil de significativité prédéfini (la droite rouge pour les valeurs de *p-value*  $< 5.10^{-8}$ , la droite bleue pour les *p-values*  $< 10^{-5}$ ). Les résultats d'association de chacune de nos trois études sont présentés ci-dessous.

- **Cohorte ARCADIA-3C**

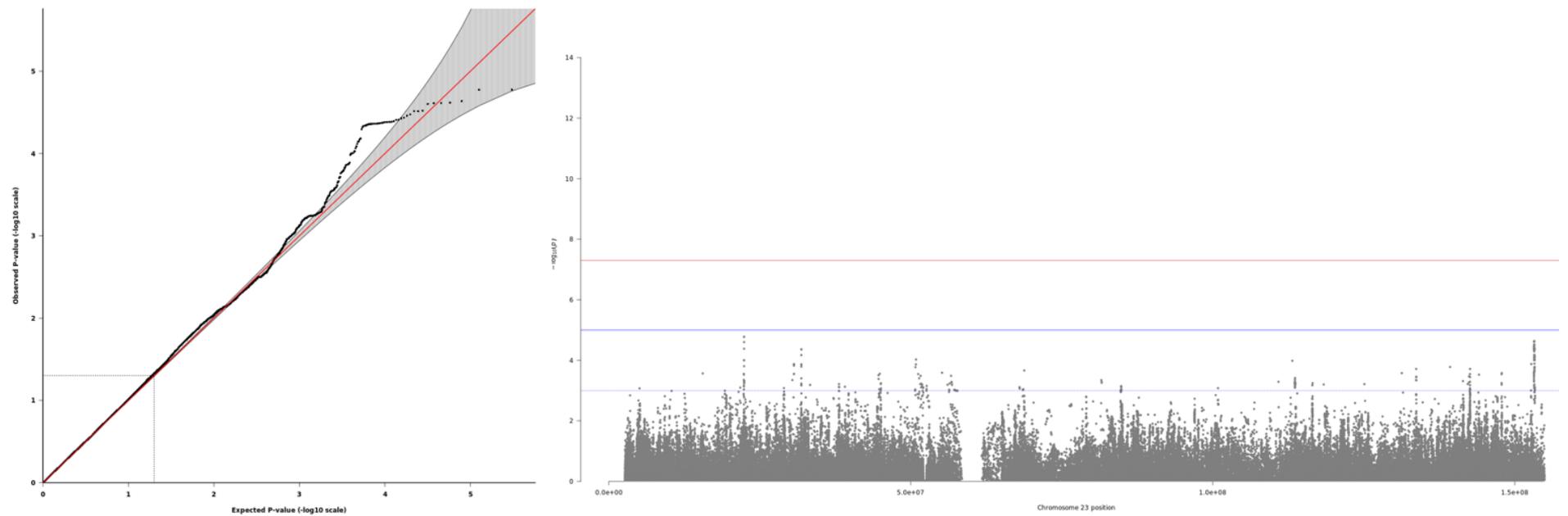
Dans l'étude ARCADIA-3C, j'ai effectué le test d'association sur 150 839 SNPs, le résultat obtenu n'a montré aucune association significative au seuil de  $5.10^{-8}$  entre ces SNPs et le risque de la DFM au seuil de  $5.10^{-8}$ . Le QQ plot et le Manhattan plot sont donnés dans la Figure 21.



**Figure 21.** QQ plot et Manhattan plot de l'étude d'association de la DFMA pour la cohorte ARCADIA-3C en utilisant le **modèle 1**.

- *Cohorte ARCADIA-POL*

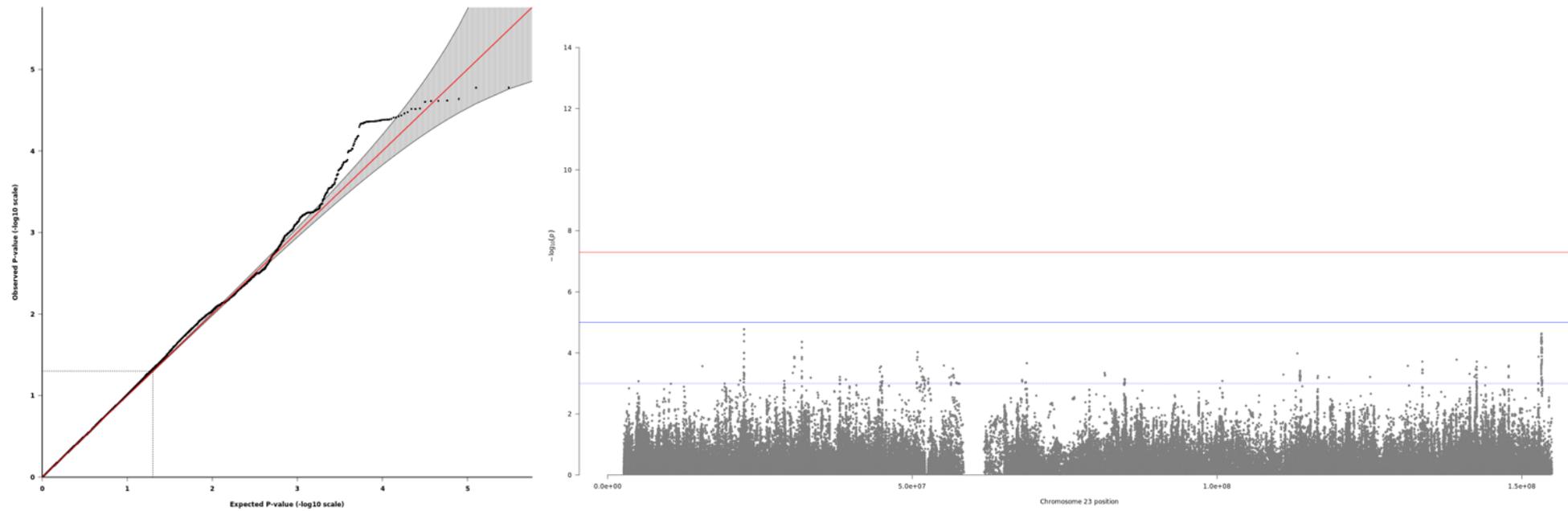
Concernant l'étude ARCADIA-POL, le test d'association été effectué sur 163 559 SNPs, les résultats présentés graphiquement (voir la Figure 22) montrent l'absence d'association significative entre ces SNPs et la DFM.



**Figure 22.** QQ plot et Manhattan plot de l'étude d'association de la DFM pour la cohorte ARCADIA-POL en utilisant le **modèle 1**.

- ***Cohorte FEIRI***

Enfin, pour la cohorte FEIRI, j'ai testé l'association de 170 814 SNPs et le risque de DFM. L'association significative est également absente comme pour les deux études précédentes. Les résultats ont été illustrés dans la Figure 23.



**Figure 23.** QQ plot et Manhattan plot de l'étude d'association de la DFM pour la cohorte FEIRI en utilisant le **modèle 1**.

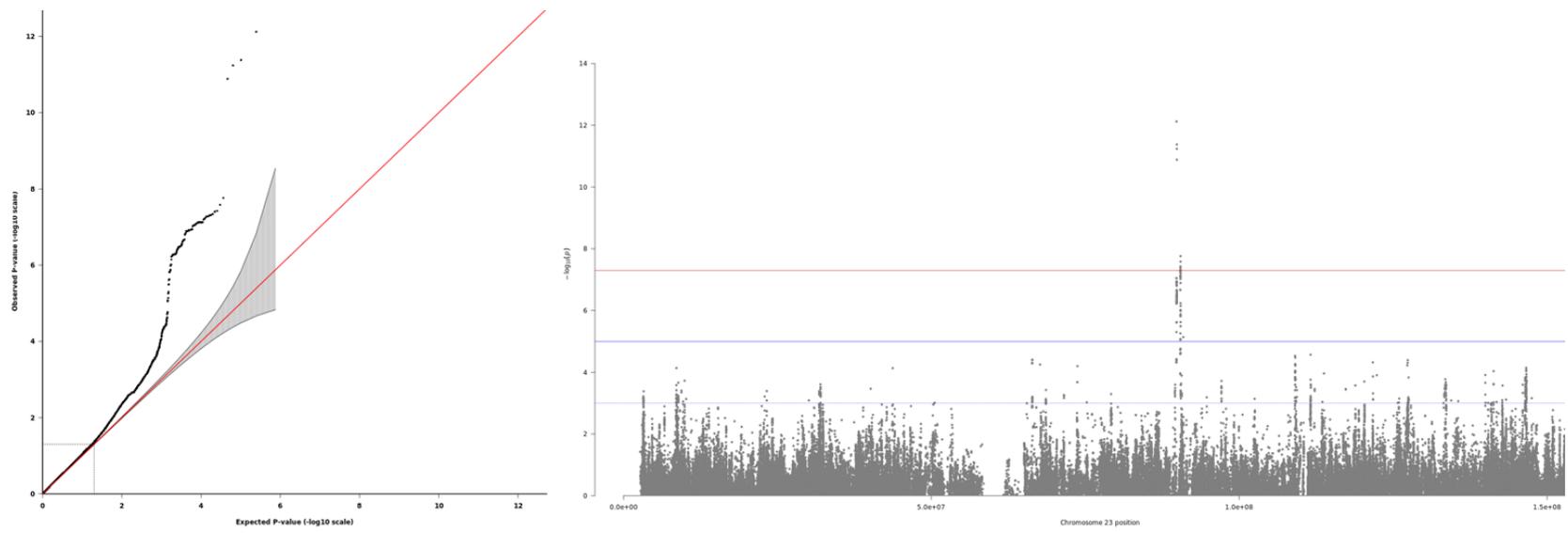
## ii. Modèle 2 : avec Inactivation de X

Comme nous l'avons vu dans les résultats du **modèle 1**, nous avons constaté l'absence d'association entre les SNPs du Chr X et le risque de la DFM. Je rappelle que dans ce modèle le processus de l'inactivation de X n'est pas pris en considération, ce qui est une limite de cette méthode car il existe un risque d'interprétation erronée des résultats obtenus si cette spécificité n'est pas prise en compte. Afin de pallier ce problème, j'ai utilisé un autre modèle (**modèle 2**) avec l'approche XWAS v3.0 [253] implémenté sur PLINK v 2.0 [221], en tenant compte de l'inactivation de X.

L'analyse a été faite en suivant les étapes présentées dans la section c. J'ai d'abord effectué un contrôle qualité sur l'ensemble des données des trois cohortes, en utilisant les mêmes critères que ceux du **modèle 1**. Ensuite j'ai appliqué un autre contrôle qualité sur les hommes et les femmes séparément (étape supplémentaire incluse dans XWAS). Les SNPs avec un taux de données manquantes supérieur à 1%, une fréquence allélique mineure supérieur à 1%, et les SNPs présents dans les régions pseudo-autosomiques (PAR) ont été exclus dans chacune des études. Après avoir rassemblé les données hommes et femmes, j'ai utilisé la même procédure d'imputation que dans le **modèle 1**. Et enfin, l'analyse d'association a été réalisée avec le modèle de régression logistique, en ajustant le modèle sur le sexe. Pour cette Covariable le codage est identique chez les mâles et les femmes, ils sont tous les deux sur une échelle de 0..2. Cette option est implémenté par défaut dans PLINK v2.0 [221], en utilisant la commande `--glm` avec l'option `--xchr-model`. Les résultats d'association des SNPs avec la DFM pour le **modèle 2** sont donnés ci-dessous.

- **Cohorte ARCADIA-3C**

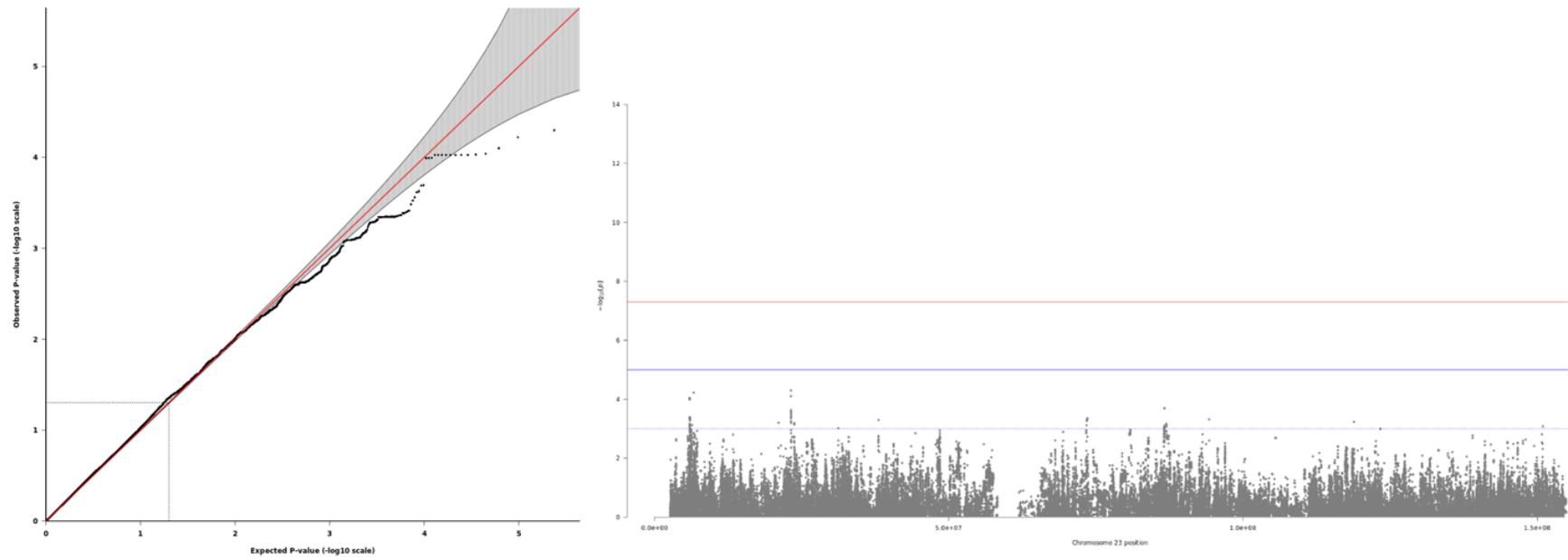
Le test d'association a été fait sur 170 926 SNPs. Le résultat obtenu dans le Manhattan plot montre un signal d'association significatif pour la DFM. Ce qui signifie l'identification d'un SNP significatif au seuil de GWAS. Le QQ plot et le Manhattan plot sont illustrés dans la Figure 24.



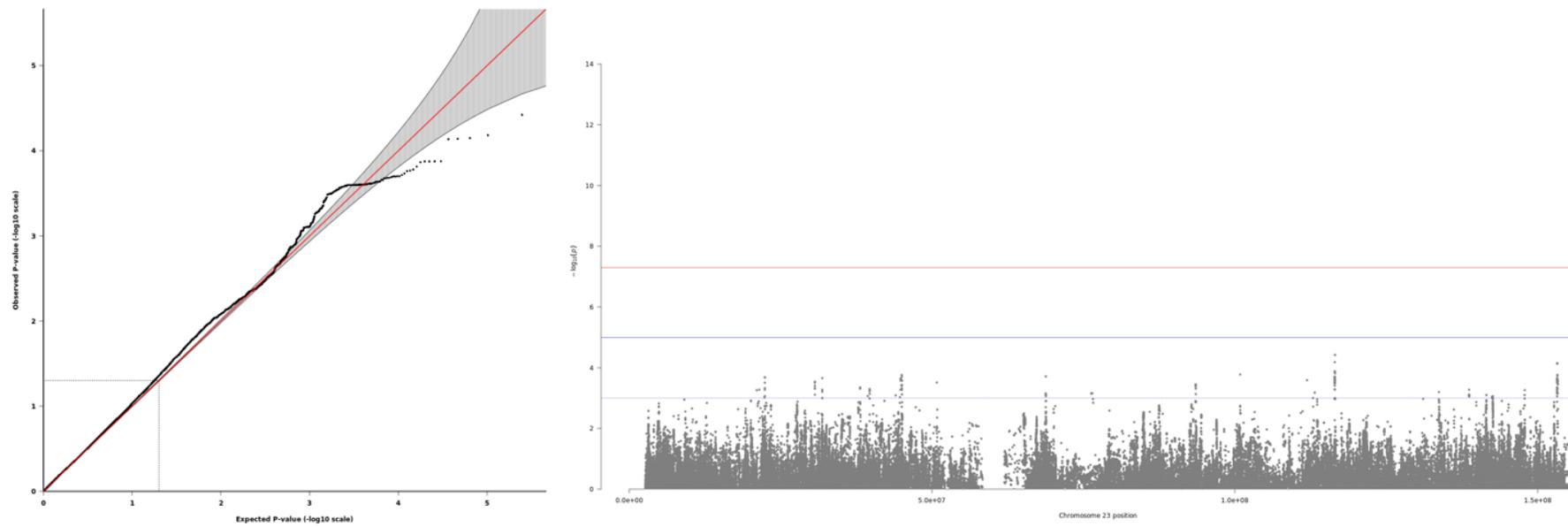
**Figure 24.** QQ plot et Manhattan plot de l'étude d'association de la DFM pour la cohorte ARCADIA-3C en utilisant le **modèle 2**.

- *Les cohortes ARCADIA-POL et FEIRI*

Pour ces deux cohortes, les résultats obtenus n'ont montré aucune association significative des SNPs inclus dans le test (163 904 SNPs pour ARCADIA-POL, 171 648 SNPs pour FEIRI). Les QQ plot et les Manhattan plot sont présentés dans les Figures 25 et 26 pour ARCADIA-POL et FEIRI respectivement.



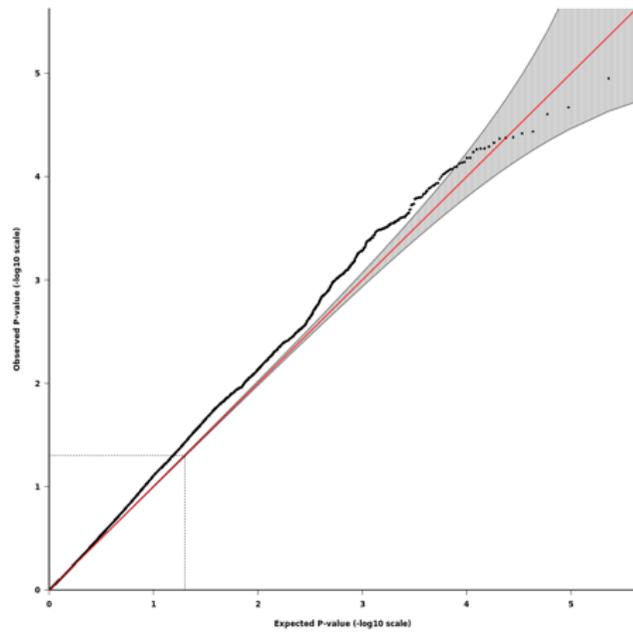
**Figure 25.** QQ plot et Manhattan plot de l'étude d'association de la DFM pour la cohorte ARCADIA-POL en utilisant le **modèle 2**.



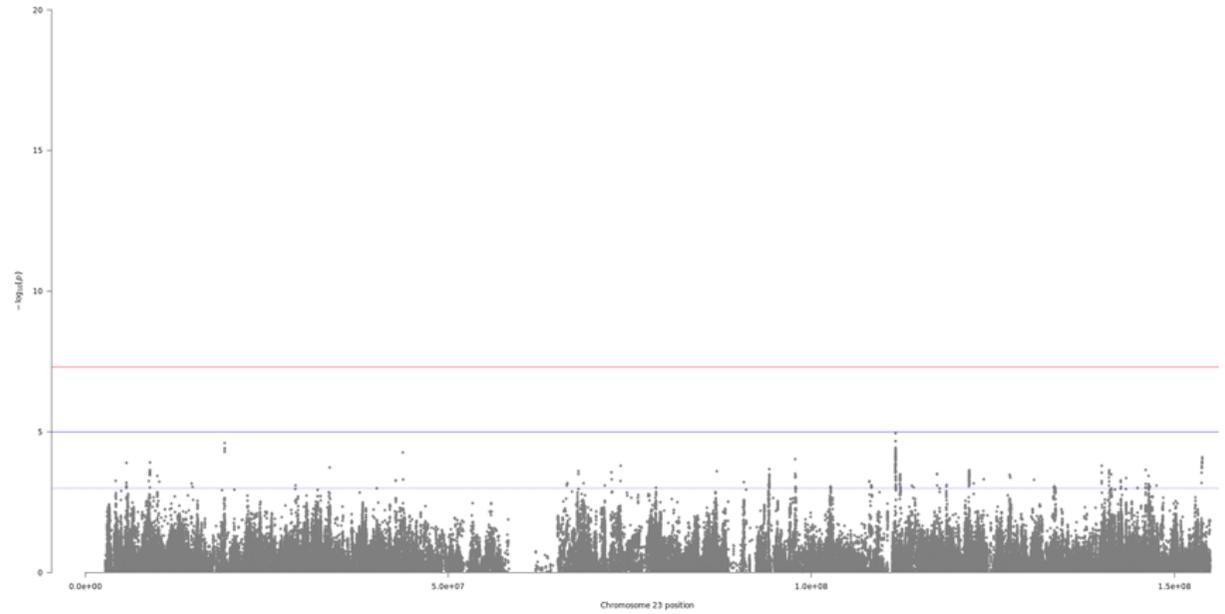
**Figure 26.** QQ plot et Manhattan plot de l'étude d'association de la DFM pour la cohorte FEIRI en utilisant le **modèle 2**.

- **Analyse d'association des trois cohortes ensemble**

L'approche XWAS permet aussi de faire l'analyse stratifiée par sexe. J'ai appliqué ce test sur chacune de nos études individuelles. Cependant, j'ai rencontré des difficultés à le faire à cause de l'hétérogénéité du nombre des hommes et des femmes dans nos données. Pour rappel, les cohortes de la DFM et de la SCAD sont composées de 80 à 100% femmes. Pour contourner ce problème, j'ai utilisé l'approche « *pooled\_analysis* », ceci a consisté à regrouper la fois les 3 cohortes ARCADIA-3C, ARCADIA-POL et FEIRI, pour appliquer le test sur la cohorte finale. Ensuite, j'ai effectué une analyse d'association, le résultat obtenu reste globalement le même que les précédentes approches. Le QQ plot et le Manhattan plot donnés ci-dessous.



QQ plot



Manhattan plot

**Figure 27.** QQ plot et Manhattan plot de l'approche « pooled-analysis » de la DFM avec le **modèle 2**.

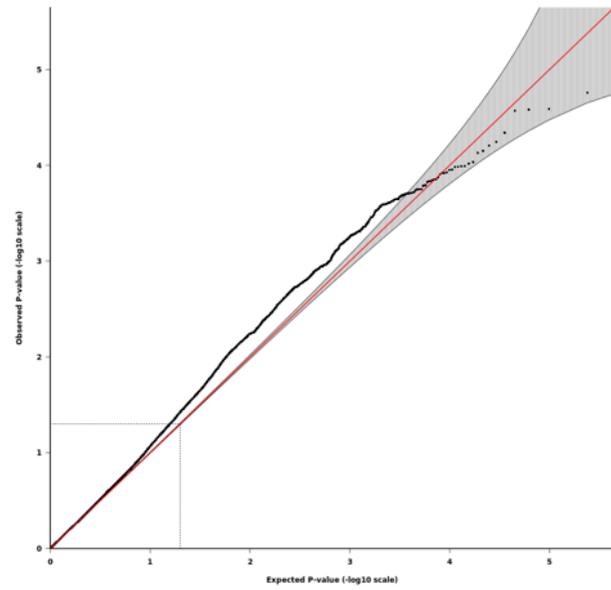
### iii. XWAS

Les résultats GWAS obtenus pour nos cohortes individuelles n'ont montré aucun locus significatif au seuil pan-génomique ( $P\text{-value} < 5.10^{-8}$ ) pour la DFM, cela est sûrement dû à la petite taille de nos échantillons. Afin de pallier ce problème, j'ai réalisé une méta-analyse incluant 6 études de GWAS indépendantes pour le même phénotype. Le principe de cette approche est de quantifier l'effet global pour l'ensemble des études considérées. Parmi les avantages de la méta-analyse est qu'elle permet de combiner les informations de ces différentes études en utilisant leurs résumés statistiques. Un autre avantage aussi est d'augmenter la puissance statistique de l'étude en considération d'un nombre de sujets plus important qu'une, ce qui pourra permettre d'identifier des SNPs significatifs sans qu'ils soient nécessairement dans chacune des études individuelles.

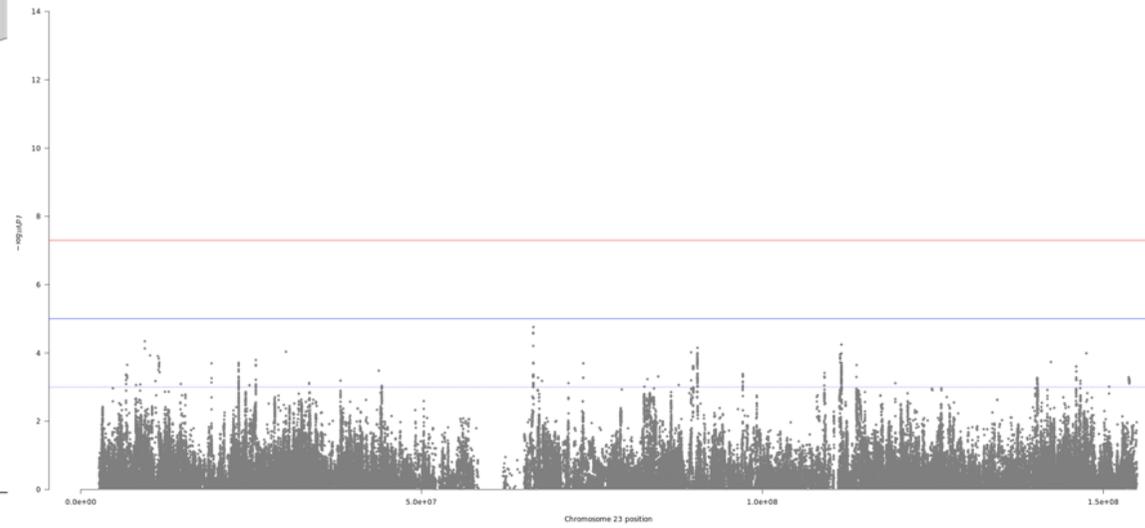
Pour cette méta-analyse, j'ai testé 165525 SNPs en fusionnant six études indépendantes (trois résumés statistiques de mes analyses d'associations précédentes, et trois autres issus des collaborations internationales Européennes et Américaines). Le nombre total d'individus inclus dans cette méta-analyse est 1556 cas et 7100 témoins. J'ai exclu les SNPs une  $MAF < 0.01$ , une mauvaise qualité d'imputation ( $info < 0.8$ ), et avec un écart rapport à l'équilibre de HWE ( $P < 10^{-5}$ ). J'ai également ajusté le modèle sur les 20 premières composantes et le sexe en considérant l'inactivation de X (**modèle 2**). Ensuite, J'ai utilisé la méthode à effet fixes en pondérant l'effet observé dans chaque cohorte par l'inverse de la variance telle qu'implémenté dans le logiciel METAL [260].

J'ai également réalisé une autre méta-analyse, j'ai combiné les six études précédentes sur les femmes seulement. J'ai fait le même QC avec les mêmes critères appliqués dans la méta-analyse précédente. J'ai réalisé une étude d'association sur 165013 SNPs.

Les résultats des deux méta-analyses ne montrent aucun SNP associé significativement avec le risque de la DFM. Ce qui a été confirmé dans les figures ci-dessous.

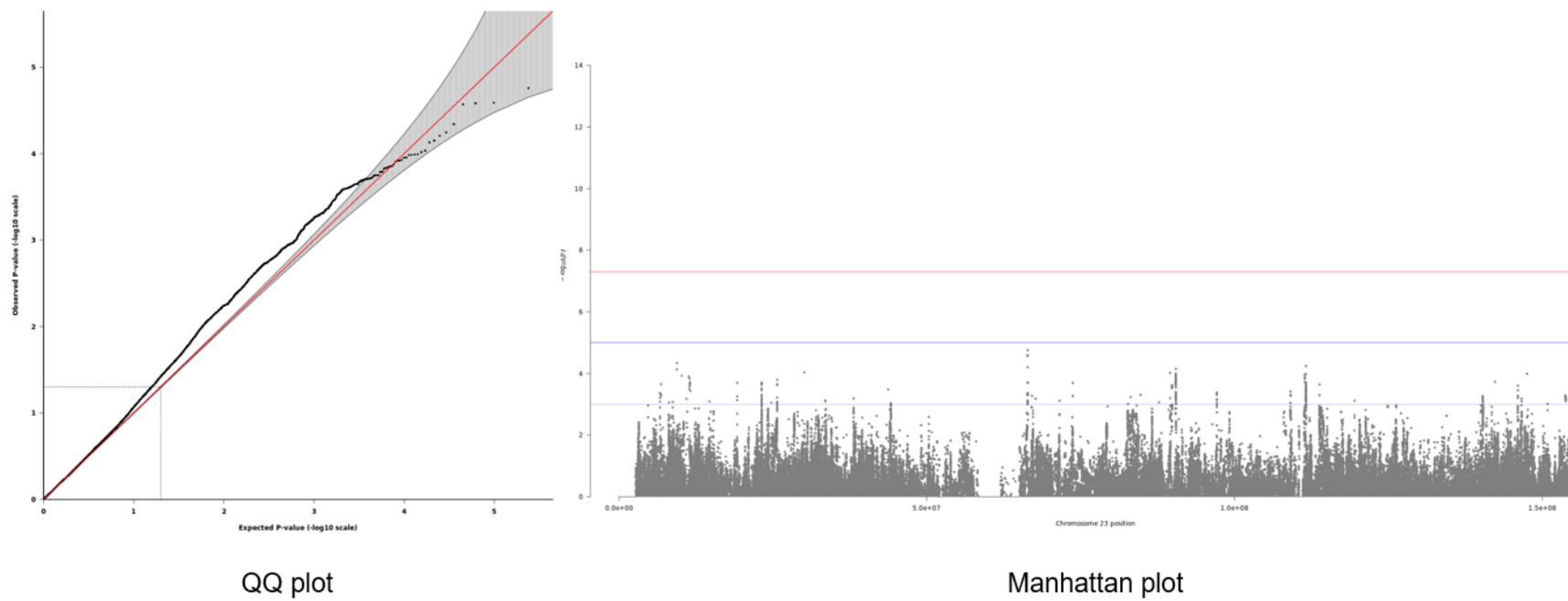


QQ plot



Manhattan plot

**Figure 28.** QQ plot et Manhattan plot de la méta-analyse du Chr X des six études de la DFM avec le **modèle 2**.



**Figure 29.** QQ plot et Manhattan plot de la méta-analyse du Chr X des six études de la DFM stratifiée uniquement sur les femmes avec le **modèle 2**.

#### **iv. Test d'hétérogénéité**

Après une méta-analyse, il convient de tester l'hétérogénéité des associations entre les six études. Ce qui revient à tester les effets tester l'hétérogénéité des effets  $\beta_i$  entre études à l'aide du Q-test de Cochran [261]. Ce test correspond à la somme pondérée des carrés des différences entre l'effet combiné estimé et les effets estimés dans chacune des N études. Pour notre analyse, le test a montré une homogénéité entre les études incluses de la méta-analyse.

#### **b. Étude SCAD**

Dans la deuxième partie de cette section, je vais présenter les résultats de l'étude GWAS pour le phénotype SCAD de la cohorte Française DISCO-3C, incluant 313 cas et 1487 témoins. L'analyse d'association a été effectuée suivant la même procédure pour l'étude de la DFM (QC, imputation, QC post imputation, test d'association) en utilisant PLINK v2.0 [221]. Les détails de cette analyse sont donnés comme suit.

##### **i. Modèle : avec Inactivation de X**

- **QC standard, imputation et QC-post imputation**

Pour bien mener l'analyse d'association de la SCAD en utilisant la cohorte DISCO-3C2. J'ai effectué le contrôle qualité en respectant les mêmes critères présentés d'étude DFM.

Dans un premier temps, les individus et les SNPs mal génotypés (taux de génotypage >1%), les doublons, et les individus avec un lien de parenté ont été supprimés. J'ai vérifié l'information du sexe des individus, en comparant le taux d'homozygotie attribué à chaque individu avec le taux d'homozygotie attendu. J'ai restreint mon étude sur la population Européenne en appliquant l'ACP avec 1000 génomes phase3 comme population de référence (<https://www.internationalgenome.org/category/phase-3/>). Enfin, j'ai exclu les SNPs ne respectant pas une MAF supérieure à 1% et une déviation par rapport à loi de Hardy Weinberg (HWE ( $P < 10^{-5}$ )).

Dans la deuxième partie, j'ai préparé le fichier de mes données contenant que le Chr X pour l'imputation (algorithme minima 4, panel de référence HRC v1, serveur : *Michigan Imputation*), dans le but d'augmenter le nombre de SNPs non génotypés.

Enfin, la dernière partie consiste à nettoyer les données imputées. J'ai sélectionné les variants avec une MAF > 0.01 et une qualité d'imputation info > 0.8. Le nombre final de SNPs à analyser pour cette cohorte DISCO-3C est de 154725 SNPs. Le Tableau 6 ci-dessous résume toutes les informations de l'étape du génotypage, le QC standard et le QC post-imputation.

Country	Cohort	Status	Genotyping array	N (genotyped)	QC steps							N included in GWAS	
					sub-Cohort selection filter	Heterozygosity / Call Rate	Relatdness (IBD>0.185)	Non European excluded. (PCA analysis)	Non SCAD or unknown	genetic syndrome cases	Sex check		
FRA	DISCO	Cases	Infinium OmniExpressExome-8v1.3	412	-	12	1	41	45	-	-	313	
	3 Cities study	Controls	Illumina-Human660W-Quad v1.0	1487	-	0	0	0	NR	-	-	1487	
UK	SCAD-UK Study I	Cases	Sequenced	383				0				383	
		Controls*	Affymetrix Axiom UK Biobank array*	1487		0	0	0	NR			1919	
UK	SCAD-UK Study II	Cases	Illumina-Infinium Global Screening Array-24 v2.0 +MD	163		20	0	0	0			143	
		Controls*	Affymetrix Axiom UK Biobank array*	815		0	0	0	NR			815	
USA	MAYO	Cases	Infinium OmniExpress-24 v1.2	506		0	0	0	0			506	
		Controls	Infinium HumanHap550, 610, 660; OmniExpress	1549		0	0	0	NR			1549	
USA	CanSCAD	Case	illumina-Infinium CoreExome-24v1.1 BeadArray with 607,778 SNP markers	502	49	1	1	88	2	2	2	357**	
	MGI	Controls	(UM_HUNT_Biobank_v1-1_20006200_A)	13756		0	0	11631	NR			2125***	
USA	DEFINE-SCAD Study	Cases	Infinium OmniExpressExome-8v1.6	47		1	0	0	30			16	42
		Cases	Illumina-Human-Omni-Express-Exome	172		3	0	0	143			26	
		Controls	Infinium OmniExpressExome-8v1.6	1		0	0	0	0			1	153
		Controls	Illumina-Human-Omni-Express-Exome	164		2	6	0	1		3	152	
AU	VCCRI	Cases	Whole Genome Sequencing	88		0	0	0	0			88	
		Controls		1127		0	0	0	NR			1127	
AU	VCCRI2	Cases	Axiom UK Biobank array-cases Axiom	91		0	2	4	0			85	
		Controls	PMDA-controls	111		0	0	0	NR			111	

**Tableau 6.** Informations détaillées pour les différentes étapes, génotypages, QC, QC-post imputation pour chaque étude de SCAD.

- **Test d'association**

Pour examiner l'association entre le SNPS du Chr X et le risque de SCAD, j'ai utilisé une régression logistique mixte, en utilisant le modèle 2 en tenant compte de l'inactivation X et en ajustant sur les 20 premières composantes principales. La SCAD étant cliniquement associée à la DFM [86], [262]–[264], celle-ci est signalée chez 40 à 60 % des patients atteints de SCAD [96], [263]. Cela m'a incité à considérer les cas DFM dans mon étude d'association. Cette analyse d'association de la cohorte Française a été réalisée en considérant les scénarios suivants :

**Premier scénario** Consistait à tester l'association de 154 725 SNPs avec le phénotype SCAD en ajustant sur le sexe. Ensuite, j'ai testé l'association de 156158 SNPs avec la SCAD sur les femmes uniquement.

**Deuxième scénario** Je me suis intéressée aux patients diagnostiqués avec la SCAD et la DFM au même temps, dans le but de savoir s'il existe un éventuel lien entre les SNPs du Chr X et le risque de ces deux phénotypes au même temps. J'ai réalisé ce test d'association en incluant 154 072 SNPs en ajustant sur le sexe.

**Troisième scénario** J'ai testé l'association de 153 358 SNPs avec la SCAD en supprimant les individus avec une DFM signalée.

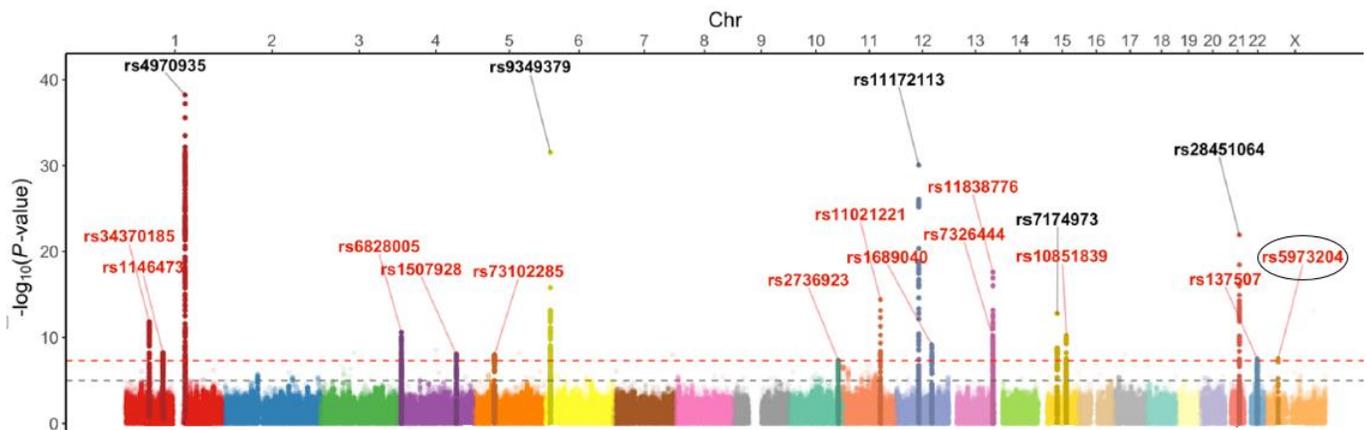
Les différents résultats d'association obtenus n'ont révélé aucun SNP sur le Chr X associé à la SCAD. Afin d'apporter plus de puissance statistique à cette étude d'association, j'ai réalisé une méta-analyse en incluant d'autres étude indépendante pour le phénotype SCAD.

## **ii. Méta-analyse**

La méta-analyse de l'étude SCAD a été réalisée avec le logiciel METAL [260] en choisissant la méthode à effets fixe pondérée par l'inverse de la variance. J'ai ajouté à mon étude Française sept autres études Européennes et Américaine, ce qui a donné un total de 1917 cas et 9292 témoins. Afin de garantir des résultats cohérents, j'ai appliqué un contrôle qualité sur la totalité des SNPs en utilisant les mêmes critères que dans l'étude d'association de la DFM. J'ai supprimé les SNPs avec une fréquence d'allèle mineur ( $MAF < 0.01$ ), une mauvaise qualité d'imputation ( $info < 0.8$ ), et

ceux ne respectant pas l'équilibre de HWE ( $P < 10^{-5}$ ). J'ai réalisé une méta-analyse pour chaque scénario décrit dans la partie précédente en ajustant à chaque fois sur les 20 premières composantes principales, ce qui donne un total de quatre méta-analyses les détails sont donnés comme suit :

**Premier scénario** La première méta-analyse a inclus huit études GWAS de la SCAD, avec 1917 cas et 9292 témoins. En utilisant une régression logistique mixte ajustée par sexe. Le résultat obtenu a révélé une association significative pour le locus *SRSF2P1/TMEM47* ( $OR=1.36$ ,  $P=3.0 \times 10^{-8}$ ). Le Manhattan plot est donné dans la Figure 30. Concernant la deuxième méta-analyse sur les femmes seulement contenant, le résultat obtenu a montré l'absence d'une association significative pour la SCAD.



**Figure 30.** Manhattan plot de la méta-analyse des huit études de la SCAD avec le **modèle 2**.

**Deuxième scénario** Dans la troisième méta-analyse a été faite, j'ai combiné quatre études GWAS en tenant compte des patients DFM. Je n'ai obtenu aucun SNP sur le Chr X associé au risque de la SCAD.

**Troisième scénario** Pour la quatrième et dernière méta-analyse des quatre GWAS SCAD sans les patients DFM. Le test d'association n'a identifié aucun SNP associé à la SCAD.

### **iii. Test d'hétérogénéité**

Les tests d'hétérogénéité réalisés dans ces quatre méta-analyses ont validé l'hypothèse d'homogénéité des effets des SNPs dans toutes les études utilisées, ce qui signifie que toutes les études dans les méta-analyses précédentes évoluent le même effet.

### 3. Conclusion

Dans cette première partie de ma de thèse, je me suis intéressée à l'étude de l'impact des variations génétiques fréquentes du Chr X dans la prédisposition à la DFM et à la SCAD, deux maladies à prédominance féminine. Mon travail a consisté à mener une étude d'association spécifiquement sur le chromosome X.

Dans un premier temps, j'ai utilisé deux approches différentes dans l'étude d'association de la DFM. Cette différence est dans le choix du modèle de régression utilisée. Le premier modèle ne tient pas compte de l'inactivation de X. Pour le deuxième modèle, j'ai utilisé l'approche XWAS qui intègre une procédure d'analyse bien spécifique au chromosome X, notamment l'étape du contrôle qualité. Cette étude d'association a été réalisée sur chacune des cohortes suivantes : ARCADIA-3C, ARCADIA-POL et FEIRI. En combinant ces trois études, j'ai réalisé la première méta-analyse pour les deux modèles décrits précédemment. Les résultats obtenus n'ont montré aucun locus sur le Chr X associé significativement à la DFM. Quant à la deuxième méta-analyse, je l'ai effectuée en utilisant les résumés statistiques issus de six études GWAS indépendants d'origine Européenne et Américaine, en se basant sur le modèle implémenté dans PLINK v2.0 [221]. Les résultats obtenus n'ont également révélé aucune association significative.

Dans un second temps, j'ai testé l'association entre les SNPs sur le Chr X et la SCAD de la cohorte DISCO-3C. J'ai appliqué les mêmes critères de contrôle qualité de l'analyse DFM, ainsi que le modèle de régression implémenté dans PLINK v2.0 [221]. Le résultat montre l'absence d'une association significative au risque de la SCAD. Afin d'augmenter la puissance statistique, j'ai fait une méta-analyse de huit études GWAS indépendantes. Cette méta-analyse a permis d'identifier pour la première fois un nouveau locus situé sur Chr X associé à la SCAD : *SRSF2P1/TMEM47* (rs5973204, OR= 1.36, P =  $3.0 \times 10^{-8}$ ). Pour les autres méta-analyses effectuées en stratifiant par la DFM et en tenant compte que des femmes, les résultats obtenus n'étaient pas significatifs.

## **Chapitre 4.**

**Exploitations de données résumées de GWAS de traits liés à la physiologie des femmes pour étudier leurs liens avec le risque génétique de la DFM**

## **1. Méthodologies appliquées**

### **a. Exploitations de données résumées GWAS**

Dans cette deuxième partie de mon travail de thèse, j'ai utilisé différentes approches statistiques pour étudier les liens génétiques potentiels entre les traits physiologiques liés aux caractéristiques de la reproduction féminine (âge de puberté, âge de ménopause, le temps d'exposition aux hormones de reproduction et les taux circulants d'hormones sexuelles) et le risque génétique de la DFM.

Cette étude analytique a été réalisée sur la base des résumés statistiques du GWAS disponibles dans l'étude *UK Biobank* (<http://nealelab.is/uk-biobank>). *UK Biobank* est une très grande cohorte prospective britannique contenant des informations génétiques et médicales de plus d'un demi-million de participants bénévoles. Elle vise à faire de nouvelles découvertes scientifiques pour améliorer la compréhension de la prévention, du diagnostic et du traitement d'un large éventail des maladies les plus courantes et les plus mortelles [136]. Le recrutement de cette étude a été effectué entre 2006 et 2010 et a inclus plus de 500 000 participants hommes et femmes, âgés de 40 à 69 ans, dont la santé a été suivie à long terme. La base de données a été régulièrement complétée par des données supplémentaires. En 2012, la ressource *UK Biobank* a été lancée en tant que ressource en libre accès à l'usage de tous les chercheurs du monde entier souhaitant mener des recherches liées à la santé dans l'intérêt public. Les informations des données que j'ai utilisées pour mener cette analyse d'association sont données sont la section suivante.

### **b. Descriptive des données exploitées**

#### **i. Données individuelles DFM**

Pour toutes les analyses d'associations, j'ai utilisé les mêmes études cas-témoins de la DFM que dans les analyses présentées dans le chapitre 3, en incluant que les femmes. Je précise que ces cohortes ont déjà subi un contrôle qualité sur les individus et les variants. Les informations de chaque cohorte sont décrites dans le Tableau 7.

<b>Cohorte</b>	<b>Nombre Femme</b>	<b>Nombre total inclus dans GWAS</b>
ARCADIA-3C	534 cas 876 témoins	1410
FEIRI	281 cas 611 témoins	892
ARCADIA-POL	100 cas 227 témoins	327

**Tableau 7.** Données descriptives des cohortes individuelles de la DFM.

## **ii. Données d'âge de puberté, d'âge de ménopause et de la durée de la vie reproductive**

Plusieurs études de recherche ont démontré que les facteurs reproductifs des femmes sont associés au risque de maladie cardiovasculaire [265], [266].

Le trait de l'âge de puberté chez les femmes a été mentionné dans des questionnaires remplis par les participantes, en écrivant en année entière l'âge de l'apparition des règles pour la première fois. Les personnes d'origine non européenne, celles n'ayant pas déclaré l'âge de puberté, ou celles qui avaient déclaré un âge très extrême (inférieur à 8 ans ou supérieur à 19 ans) ont été exclus de l'étude [56]. Concernant l'âge de ménopause, les participantes ont mentionné dans le questionnaire l'âge de la dernière période de règles. Les femmes ayant un âge précoce de ménopause (< 47 ans) [265], ou un âge tardif de ménopause (> 55 ans) ne sont pas inclus dans l'étude. Enfin, le dernier trait de cette catégorie représente la durée de la vie reproductive. Il s'agit d'une variable quantitative que nous avons générée, et qui est égale à la différence entre l'âge de ménopause et l'âge de puberté pour chaque participante.

## **c. Construction des scores polygéniques**

### **i. Contrôle qualité**

Après avoir récupéré les résumés statistiques des GWAS de *UK Biobank* pour les traits cités précédemment, il est nécessaire d'appliquer un contrôle qualité avant de calculer les scores polygéniques. J'ai écarté tous les SNPs ayant une MAF < 1%, une mauvaise qualité d'imputation (info < 0.08) et un écart par rapport à HWE ( $P < 10^{-6}$ ). J'ai également exclu les variants d'insertion et de délétion (indels), car un indel peut incorporer ou supprimer un ou plusieurs nucléotides. Le nombre de SNPs avant et après le QC pour chaque trait est donné dans le tableau suivant.

Trait	Nombre de femme	Nombre de SNPs près-QC	Nombre de SNPs post-QC
Age de puberté	109758	13791467	8506555
Age de ménopause	109758	13791467	8506555
Durée d'exposition aux hormones	109758	13791467	8506555

**Tableau 8.** Nombre d'individus et le nombre de SNP pour chaque trait inclus dans la construction des scores polygéniques.

## ii. Calcul du score polygénique

Après le contrôle qualité, l'étape suivante consiste à appliquer la méthode du « *clumping* + *thresholding* » qui permettent de réduire la redondance génétique.

Dans un premier temps, j'ai supprimé les SNPs corrélés entre eux en utilisant les fonctions, implémentées dans PLINK v2.0 [221], décrites comme suit :

--*clump-p1* 1 : inclure les SNPs au seuil de la valeur de p1 (p1 =1 revient à inclure tous les SNPs pour le clumping)

--*clump-r2* 0.1 : sélectionner les SNPs ayant une valeur de corrélation ( $r^2 > 0.1$ )

--*clump-kb* 250 : garder les SNPs qui ont une distance de moins de 250 kb

Dans un second, j'ai sélectionné à partir des SNPs obtenus du *clumping* ceux ayant une valeur de *p-value* significative aux 9 seuils que j'ai choisis, tels que :  $p < 0.5$ ,  $p < 0.4$ ,  $p < 0.3$ ,  $p < 0.2$ ,  $p < 0.01$ ,  $p < 0.05$ ,  $p < 10^{-3}$ ,  $p < 10^{-5}$ ,  $p < 5 \times 10^{-8}$ . J'ai obtenu au total 9 fichiers contenant la liste des SNPs significatifs au seuil correspondant.

Enfin, la dernière étape est dédiée au calcul des scores génétiques de chaque trait. Les PRS sont générés en sommant les allèles de risques d'un individu pondérés par leur beta ( $PRS_i = \sum_{i=1}^N G_i \times \beta_i$ ) pour les 9 seuils de significativité ( $p < 0.5$ ,  $p < 0.4$ ,  $p < 0.3$ ,  $p < 0.2$ ,  $p < 0.01$ ,  $p < 0.05$ ,  $p < 10^{-3}$ ,  $p < 10^{-5}$ ,  $p < 5 \times 10^{-8}$ ). Une fois que les scores génétiques sont calculés, la prochaine étape consiste à faire le test d'association avec la DFM, ce qui est décrit dans la section ci-dessous.

## iii. Test d'association avec la DFM

Comme je l'ai mentionné précédemment, les scores pour les différents seuils de significativités ont été générés pour les 5 traits liés à la physiologie de la femme. Le but de ce calcul est de tester l'association entre chaque score de chaque trait et le risque de la DFM, dans les trois études GWAS incluant les femmes uniquement : ARCADIA-3C (534 cas et 876 témoins), FEIRI (281 cas et 611 témoins) et ARCADIA-POL (100 cas et 227 témoins). J'ai fait le teste d'association

de chaque score avec le risque de la DFM, en utilisant la régression logistique et en ajustant sur les 5 premières composantes dans les 3 études séparément pour tenir compte de la stratification de population. Ensuite, j'ai réalisé une méta-analyse totalisant 918 cas et 1714 témoins.

#### **d. Corrélations génétiques**

##### **i. Le principe de la méthode du LD score**

Dans cette deuxième partie de mon travail, pour évaluer la corrélation génétique de chaque trait lié à la physiologie étudié dans la section précédente et la DFM. Cela revient à calculer la corrélation entre les influences génétiques du trait 1 (âge de puberté, âge de ménopause et la durée d'exposition aux hormones) et les influences génétiques du trait 2 qui est la DFM. Pour se faire, j'ai utilisé la méthode LDSC [267]. En bref, cette méthode estime l'héritabilité en régressant les résumés statistiques GWAS sur les scores LD, le détail de ce calcul est donné dans la section ci-dessous.

##### **ii. Contrôle qualité des données et analyse de la corrélation**

Avant de procéder au calcul de la corrélation génétique, j'ai fait un contrôle qualité sur les résumés statistiques utilisés dans cette méthode. J'ai supprimé les variants avec une MAF < 1%, et les variants indels. Le nombre total de SNPs inclus dans cette étude est 917040 pour les trois traits. Après avoir effectué le contrôle qualité, j'ai effectué l'analyse de corrélation génétique de chaque trait avec la DFM par la méthode LDSC en passant par les étapes suivantes :

- **Etape 1. L'interception de régression du score LD pour un GWAS**

Cette étape consiste dans un premier temps à préparer le résumé statistique de GWAS de la DFM sous le format recommandé, en incluant que les paramètres nécessaires pour calculer la corrélation entre deux traits. Ce nouveau format contient les colonnes suivantes : (1) un identifiant unique de chaque SNP, (2) allèle 1 (allèle d'effet), (3) allèle 2 (allèle sans effet), (4) taille de l'échantillon, (5) *p-value*, (6) effet estimé de chaque SNP ( $\beta$ , OR, *z-score*, etc). Ensuite, j'ai normalisé mon nouveau résumé statistique de GWAS de la DFM en utilisant le package *MungeSumstats v 1.4.5* [268], permettant la normalisation des données, le contrôle de la

qualité des résumés statistiques du GWAS, par exemple, garder les SNPs ayant une bonne qualité d'imputation, supprimer SNPs multi-alléliques ou ambigus, vérifier le retournement d'allèle. Il permet également d'ajuster des informations manquantes telles que la taille de l'échantillon, la qualité d'imputation (info) en se basant sur un panel de référence (1000 génomes dans notre étude).

- **Etape 2. Le calcul de l'héritabilité des SNPs pour la DFM**

La deuxième étape de cette analyse de corrélation, consiste d'abord à sélectionner les SNPs significatifs selon le seuil de GWAS ( $p = 5 \times 10^{-8}$ ) pour l'âge de puberté, l'âge de ménopause et la durée d'exposition aux hormones. J'ai supprimé tous les SNPs qui sont en déséquilibre de liaisons avec le SNPs significatif (sélectionné auparavant), et ne garder que les SNPs indépendants pour former des variables instrumentales. Ensuite, j'ai estimé l'héritabilité des SNPs indépendants sur la DFM en utilisant la commande `--h2` implémentée dans le package `ldsc` (<https://github.com/bulik/ldsc/tree/master/ldscore>).

- **Etape 3. La corrélation génétique entre la DFM et le trait lié à la physiologie de la femme.**

Dans cette dernière étape, j'ai calculé la corrélation génétique entre la DFM et chacun des traits étudiés en utilisant la commande `--rg` avec le package `ldsc`. Cette commande utilise l'intersection des SNPs de la DFM et les SNPs du trait étudié pour la régression du score LD.

## e. Randomisation Mendélienne

### i. Présentation des différentes approches MR appliquées

La randomisation mendélienne est une méthode permettant grâce à la génétique de déterminer si le lien entre la variable d'exposition, et phénotype d'issue (D et Y dans la Figure 13 respectivement) est causal ou simplement corrélé. Dans notre analyse RM, on a appliqué la méthode à deux échantillons d'études différentes, en utilisant deux résumés statistiques (un pour le facteur de risque qui représente les traits reproductifs dans notre étude, et l'autre pour le phénotype d'issue la DFM sur les femmes uniquement). Le but principal de cette analyse est d'estimer les effets causaux du trait liés à la physiologie de la femme sur la DFM. On suppose que toutes les

relations entre ces variables (en particulier, les associations génétiques avec le trait étudié et la DFM, et l'effet causal de ce trait sur la DFM) sont linéaires sans modification d'effet. On suppose également que toutes les variants génétiques ne sont pas corrélées (c'est-à-dire qu'elles ne sont pas en déséquilibre de liaison).

Soit  $\beta_{Yj}$  l'association entre le variant génétique du trait  $G_j$  ( $j = 1, 2, \dots, J$ ). L'association génétique avec le résultat peut être décomposée en la somme d'un effet direct (pléiotrorandomizpique) et d'un effet indirect (causal) :

$$\beta_{Yj} = \alpha_j + \theta\beta_{Xj} \quad (8)$$

Où:  $\alpha_j$  est l'effet du variants génétiques sur la DFM qui n'est pas médié par le facteur de risque (trait étudié), et  $\theta$  est l'effet causal du facteur de risque sur la DFM [269], voir Figure X.

un variant génétique ayant des associations avec plus d'un facteur de risque sur différentes voies causale, est appelé variant génétique est pléiotropie [269]. Tout effet de ce type est inclus dans le paramètre  $\alpha_j$  ; un variant génétique est pléiotrope si  $\alpha_j \neq 0$ . Donc il en résulte que c'est une variable instrumentale non valide (violation une des hypothèses données dans le chapitre 2.

Dans la section suivante, je présenterai plusieurs approches inclues la méthode à deux échantillons d'études différentes décrites précédemment, permettant d'étudier le lien de causalité entre le trait lié à la physiologie de la femme dans notre étude et la DFM, en tenant compte des différents cas d'analyse.

- **Approche de pondération inverse de la variance (IVW)**

Le principe de cette méthode est la régression pondérée en variance inversée. IVW fournit la première indication pour savoir si les traits reproductifs indépendamment causent la DFM. L'estimateur d'IVW est le rapport obtenu en divisant l'effet d'un SNP sur la DFM par l'effet qu'il a sur le trait reproductif utilisé (l'âge de puberté, l'âge de ménopause ou la durée d'exposition aux hormones), en supposant que toutes les hypothèses sont satisfaites, c'est-à-dire que l'intercepte est

nulle (pléiotropie équilibrée) [270]. Le calcul de l'estimateur d'IVW se fait en utilisant l'équation suivante :

$$\hat{\theta}_{IVW} = \frac{\sum_j \beta_{Yj} \beta_{Xj} se(\beta_{Yj})^{-2}}{\sum_j \beta_{Xj}^2 se(\beta_{Yj})^{-2}} \quad (9)$$

Avec ;

$\beta_{Yj}$  : L'effet causal de chaque variant  $j$  sur la DFM.

$\beta_{Xj}$  : L'effet causal de chaque variant  $j$  sur le trait étudié.

$se(\beta_{Yj})^2$ : la variance de l'estimation de l'association du gène de la DFM pour le SNP  $j$ .

$\beta_{Xj}^2$  : la variance de l'estimation de l'association du gène du trait étudié pour le SNP  $j$ .

La même estimation peut également être obtenue à partir d'une régression linéaire pondérée des associations génétiques avec la DFM  $\hat{\beta}_{Yj}$ , sur les associations génétiques avec le trait étudié  $\hat{\beta}_{Xj}$ , en utilisant des poids de variance inverse  $se(\beta_{Yj})^{-2}$  lorsque l'intercepte est nulle :

$$\hat{\beta}_{Yj} = \hat{\theta}_{IVW} \hat{\beta}_{Xj} + \epsilon_{Ij} \quad \epsilon_{Ij} \sim N(0, \delta^2 se(\hat{\beta}_{Yj}^2)) \quad (10)$$

Avec:

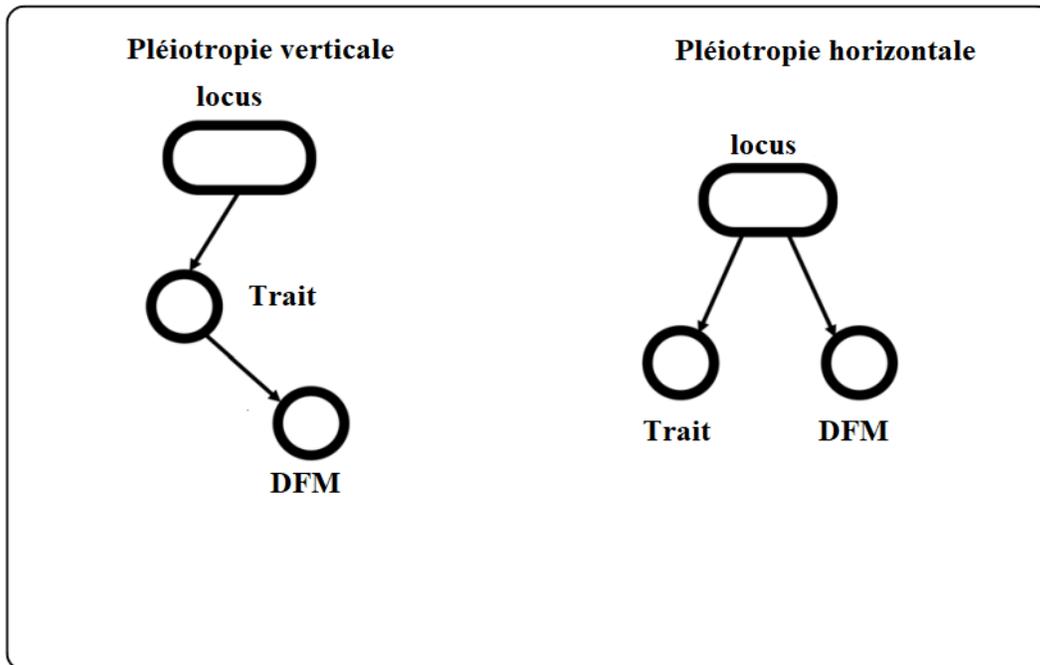
$\hat{\theta}_{IVW}$  : est l'estimateur d'IVW obtenu de l'équation (9)

$\epsilon_{Ij}$  : est le terme résiduel.

Notons que l'estimation IVW est une méthode d'analyse efficace lorsque tous les variants génétiques sont des instruments valides. Cependant, si l'intercepte est différente de zéro, cela indique une pléiotropie horizontale directionnelle [270], donc l'estimateur IVW est biaisé. Pour remédier à ce problème, l'approche RM-Egger est proposée dans ce cas.

- **L'approche RM-Egger**

Contrairement à l'approche IVW, *RM-Egger* permet la pléiotropie horizontale. Cette dernière fait référence aux SNPs ayant des effets indépendants sur la DFM et le trait en question, comme le montre la Figure 30 ci-dessous.



**Figure 31.** Schéma des différents types de pléiotropie. Pour la pléiotropie verticale, les effets sur un trait sont médiés par des effets sur un autre trait, et la pléiotropie horizontale, les effets sur plusieurs traits sont indépendants.

L'estimateur de MR-Egger est réalisé par une simple modification de la régression linéaire pondérée décrite dans l'approche d'IVW. Plutôt qu'en fixant le terme d'interception à zéro, le terme est estimé comme suit [271].

$$\hat{\beta}_{Yj} = \hat{\theta}_{0E} + \hat{\theta}_{1E} \hat{\beta}_{Xj} + \epsilon_{1j} \quad \epsilon_{1j} \sim N(0, \delta'^2 se(\hat{\beta}_{Yj}^2)) \quad (11)$$

Avec ;

$\hat{\theta}_{0E}$  : est l'intercepte du modèle (l'ordonnée à l'origine).

$\hat{\theta}_{1E}$  : est l'estimateur de RM-Egger (la pente).

- **Approche de la médiane pondérée.**

La régression médiane pondérée peut fournir une estimation causale cohérente en supposant que la majorité des variants génétiques sont valides (c'est-à-dire ne présentant aucune pléiotropie horizontale, aucune association avec des facteurs de confusion, une association robuste avec le trait étudié), pour que l'estimation de l'effet causal soit impartiale. L'estimation médiane pondérée permet aux SNPs les plus forts de contribuer davantage à l'estimation et peut être obtenue en pondérant la contribution de chaque SNP par la variance inverse de son association avec le résultat [270]. Plus précisément, soit  $\hat{\beta}_j$  la  $j$  ième estimation du rapport ordonné (classée de la plus petite à la plus grande). Si le nombre total de variants génétiques est impair  $j = (2k + 1)$ , l'estimateur médian simple est l'estimation du rapport médian  $\hat{\beta}_{k+1}$ . S'il est pair  $= 2k$ , la médiane est interpolée entre les deux estimations médianes  $\frac{1}{2} (\hat{\beta}_k + \hat{\beta}_{k+1})$ .

- **L'approche du mode pondéré**

Cette dernière approche utilisée dans mes analyses RM est basée sur un mode pondéré. L'approche du mode pondéré regroupe les SNPs en groupes déterminés par leur similarité d'effets de causalité, et renvoie l'estimation de l'effet de causalité en fonction du cluster qui a le plus grand nombre de SNP. Cette approche basée sur le mode renvoie un effet causal non biaisé si les SNPs du plus grand groupe sont des instruments valides [272]. Le mode pondéré introduit un élément supplémentaire similaire à IVW et à la médiane pondérée, pondérant la contribution de chaque SNP au regroupement par la variance inverse de son effet de résultat.

## ii. Prérequis de l'analyse RM et sélection des variables instruments par trait

Avant de procéder à l'analyse RM, il est recommandé de faire un contrôle qualité sur les résumés statistiques de chaque trait étudié (âge de puberté, âge de ménopause et la durée d'exposition aux hormones), y compris le phénotype d'issue. Le QC a été fait en se basant sur les critères standards définis dans les analyses précédentes. On a exclu les indels, les SNPs avec une MAF inférieure à 1% et les SNPs ayant un des écarts par rapport à l'équilibre de HWE ( $P < 10^{-6}$ ). Une fois le QC réalisé, l'étape suivante est de sélectionner les instruments pour faire l'analyse RM.

Cette sélection commence par conserver tous les SNPs significativement associés au seuil GWAS ( $5 \times 10^{-8}$ ) de la variable d'exposition (le trait en question). Ensuite, j'ai effectué un *clumping*, en utilisant pour critères une corrélation entre les SNPs de 10% ( $r^2 = 0.001$ ) et une fenêtre de 1000 kb. Cette étape m'a permis de ne garder, parmi ces SNPs associés, que des instruments indépendants. Avant d'entamer l'analyse d'association, j'ai utilisé la fonction incluse dans le package *TwoSampleMR* permettant d'harmoniser les SNPs par brins d'ADN, et de vérifier que les instruments choisis ne sont pas associés avec le phénotype d'issue.

La dernière étape de cette analyse, a consisté à faire un test d'association pour déterminer la causalité du trait étudié sur la DFM. Ceci a été fait en utilisant la régression logistique avec la fonction (*mr*) implémentée dans le package *TwoSampleMR*.

## 2. Résultats

### a. Scores Polygénique

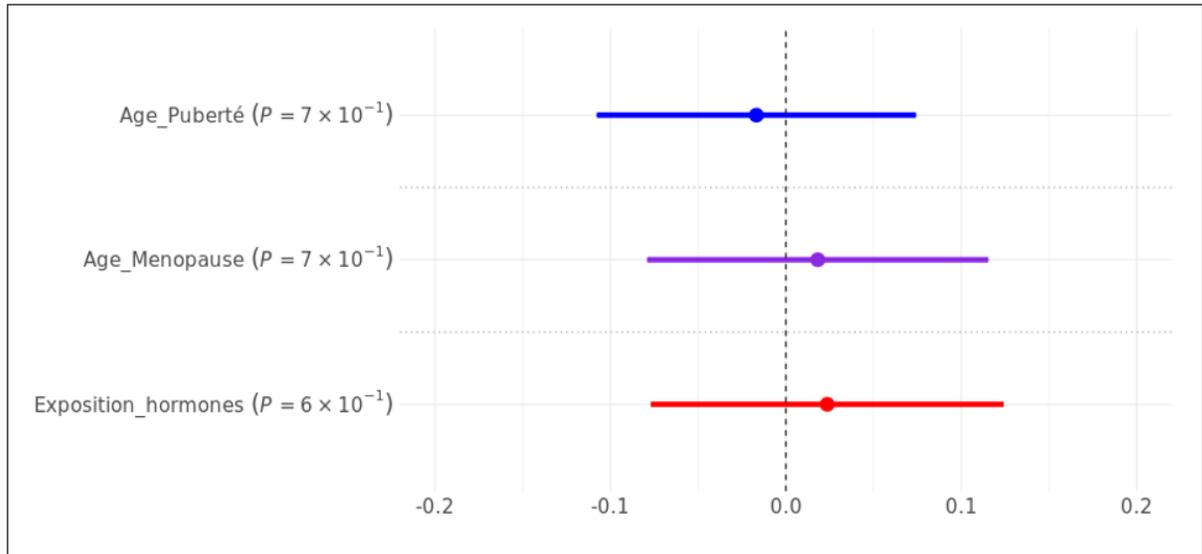
L'analyse d'association entre les scores génétiques des traits liés aux étapes majeurs de la vie reproductive des femmes et la DFM n'a montré aucun résultat significatif. Le test d'hétérogénéité est significativement positif faut au seuil de  $10^{-3}$  ( $P_{\text{Hétérogénéité}} = 0.24$ ) (Tableau 9).

Scores polygéniques (PRS)								
	Thresholds	N.SNP FR	N.SNP BEL	N.SNP POL	OR	IC 95%	P <sub>Méta-analyse</sub>	P <sub>Hétérogénéité</sub>
Age de puberté	8×10 <sup>-5</sup>	154	152	153	1.00	[0.87,1.15]	0.99	0.01
	10 <sup>-5</sup>	501	503	500	1.00	[0.86,1.15]	0.95	0.08
	10 <sup>-3</sup>	4390	4364	4380	1.01	[0.88,1.16]	0.89	0.24
	0.05	62474	62408	62676	0.98	[0.85,1.13]	0.81	0.08
	0.1	101988	101725	101982	0.94	[0.82,1.09]	0.44	0.02
	0.2	165504	164479	165048	0.92	[0.80,1.06]	0.25	0.02
	0.3	217301	215392	216306	0.92	[0.80,1.06]	0.27	0.03
	0.4	261176	258486	259841	0.93	[0.80,1.07]	0.30	0.03
	0.5	298857	295675	297522	0.92	[0.80,1.06]	0.25	0.03
Age de ménopause	8×10 <sup>-5</sup>	195	195	196	1.04	[0.90,1.20]	0.63	0.80
	10 <sup>-5</sup>	486	495	492	1.01	[0.87,1.17]	0.91	0.94
	10 <sup>-3</sup>	3459	3448	3464	0.97	[0.84,1.13]	0.72	0.90
	0.05	57793	57943	57972	0.93	[0.81,1.07]	0.30	0.53
	0.1	97183	97027	97297	0.95	[0.82,1.09]	0.44	0.75
	0.2	161274	160320	161126	0.94	[0.81,1.08]	0.37	0.92
	0.3	213703	211975	213068	0.93	[0.81,1.07]	0.34	1.00
	0.4	258656	256004	257548	0.92	[0.80,1.06]	0.27	0.98
	0.5	297387	293799	295592	0.92	[0.80,1.06]	0.23	1.00
Durée de la vie reproductive	8×10 <sup>-5</sup>	153	153	151	1.05	[0.91,1.22]	0.48	0.67
	10 <sup>-5</sup>	425	427	428	1.04	[0.90,1.21]	0.60	0.99
	10 <sup>-3</sup>	3503	3519	3535	0.99	[0.86,1.15]	0.92	0.84
	0.05	58114	58169	58287	0.99	[0.86,1.14]	0.86	0.51
	0.1	97711	97594	97853	0.97	[0.84,1.12]	0.67	0.31
	0.2	161889	160947	161587	0.95	[0.83,1.10]	0.51	0.77
	0.3	214611	212863	213930	0.94	[0.82,1.08]	0.39	0.80
	0.4	259554	256813	258153	0.95	[0.83,1.10]	0.50	0.74
	0.5	297893	294400	296132	0.95	[0.83,1.09]	0.47	0.63

**Tableau 9.** Résultats d'association entre la DFM et les différents traits en utilisant l'approche du score polygénique (PRS). N.SNP : nombre de polymorphismes mono-nucléotidiques (SNP) pour différents seuils significatifs pour chaque cohorte (FR : Française, BEL : Belge, POL : Polonaise). OR : rapport des cotes, IC à 95 % : intervalle de confiance à 95 %. P<sub>Méta-analyse</sub> : valeur P de la méta-analyse. P<sub>Hétérogénéité</sub> : p-value du test d'hétérogénéité.

## b. Corrélation génétique

Les résultats de la corrélation génétique entre les traits liés à la physiologie de la femme et la DFM sont donnés dans le Tableau 10, et sont représentés graphiquement dans la Figure 32.



**Figure 32.** Forest plot de la corrélation génétique entre la DFM et les traits liés à la physiologie de la femme. Age\_Puberté (bleu) : le trait âge de puberté. Age\_Menopause (violet) : le trait de l'âge de la ménopause. Exposition\_hormones (rouge) : la durée d'exposition aux hormones. P : la probabilité d'association de chaque trait étudié avec la DFM.

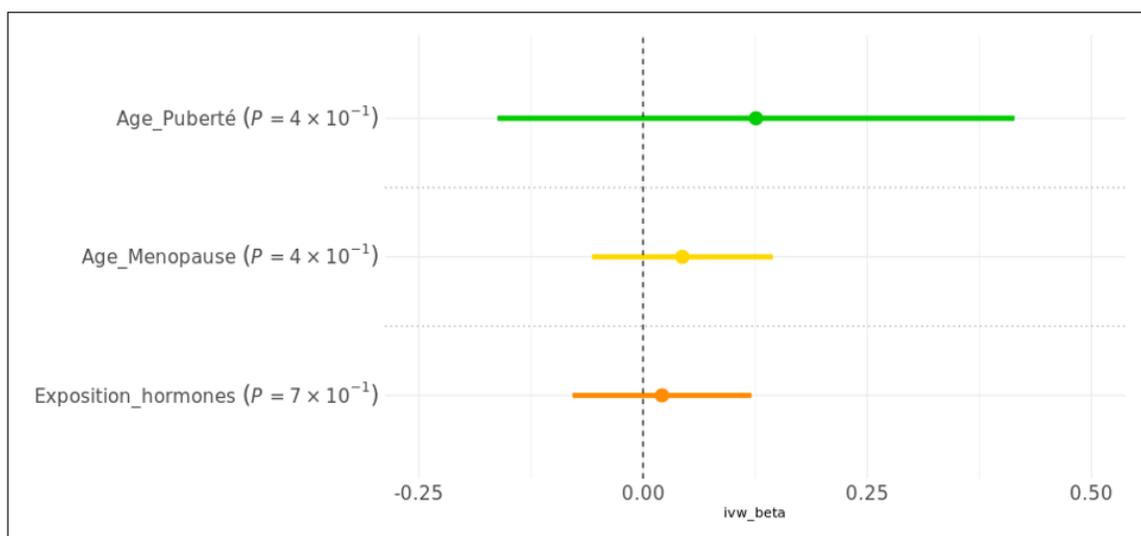
Trait 1	Trait 2	N.SNP	R <sub>g</sub>	SE	L95_R <sub>g</sub>	U95_R <sub>g</sub>	Z	P
DFM (femme)	Age de puberté	917559	-0.02	0.05	-0.11	0.07	-0.47	0.64
	Age de ménopause	917559	0.01	0.05	-0.09	0.10	0.12	0.91
	Durée d'exposition aux hormones	917559	0.01	0.05	-0.09	0.11	0.28	0.78

**Tableau 10.** Résultats de la corrélation génétique entre la DFM femmes uniquement et les trois traits : âge de puberté, âge de ménopause et la durée d'exposition aux hormones. N.SNP : nombre de SNP utilisés pour l'analyse. R<sub>g</sub> : coefficient rho de la corrélation génétique, SE : erreur standard de la taille de l'effet, L95\_R<sub>g</sub> : limite inférieure de 95% IC, U95\_R<sub>g</sub> : limite supérieure de 95% IC, Z : score de la corrélation, P: p-value de la corrélation.

D'après la Figure 32, je constate que la base génétique responsable de l'âge de la puberté aurait un effet protecteur (réduirait le risque) sur la DFM. Cependant, l'absence de significativité de la *p-value* ( $P=0.64$ ) ne permet pas de conclure sur la corrélation génétique entre l'âge de la puberté et la DFM. De plus, je constate que la base génétique responsable, d'une part de l'âge de la ménopause, et d'autre part de la durée d'exposition aux hormones aurait un effet délétère (augmenterait le risque) sur la DFM. Cependant, la significativité des *p-values* étant inférieures au seuil de 5% ( $P=0.91$  et  $P=0.78$  respectivement), il n'est pas possible de rendre une conclusion concernant la corrélation génétique entre ces deux traits et la DFM.

### c. Randomisation mendélienne

L'analyse de la randomisation mendélienne que j'ai effectuée pour l'identification d'un éventuel lien de causalité entre les trois traits étudiés précédemment et la DFM. Les résultats de cette analyse sont présentés graphiquement dans la Figure 33, et numériquement dans le Tableau 11.



**Figure 33.** Forest plot de la randomisation mendélienne entre la DFM et les traits liés à la physiologie de la femme. Age\_Puberté (vert): le trait âge de puberté. Age\_Menopause (jaune): le trait de l'âge de la ménopause. Exposition\_hormones (orange): la durée d'exposition aux hormones. P: la probabilité d'association de chaque trait étudié avec la DFM.

Phénotype d'issue	Approches	Facteurs de risque	N SNP	BETA	SE	P	P <sub>Hétérogénéité</sub>	Intercepte	P <sub>Pléiotropie</sub>
DFM	IVW	Age de la puberté	42	0.15	0.15	0.31	0.42	-0.02 [-0.05 , 0.01]	0.42
		Age de la ménopause	28	0.05	0.05	0.39	0.76	0.02 [0.00 , 0.05]	0.33
		Durée d'exposition aux hormones	25	0.02	0.05	0.68	0.58	0.01 [-0.02 , 0.03]	0.80
	RM-Egger	Age de la puberté	42	0.57	0.54	0.30	0.41		
		Age de la ménopause	28	-0.07	0.13	0.59	0.77		
		Durée d'exposition aux hormones	25	-0.01	0.13	0.95	0.52		
	Mediane pondérée	Age de la puberté	42	0.24	0.21	0.25			
		Age de la ménopause	28	-0.03	0.08	0.75			
		Durée d'exposition aux hormones	25	-0.02	0.08	0.76			
	Mode pondérée	Age de la puberté	42	0.25	0.43	0.57			
		Age de la ménopause	28	-0.05	0.10	0.63			
		Durée d'exposition aux hormones	25	-0.05	0.08	0.62			

**Tableau 11.** Résultats de l'analyse de la randomisation mendélienne (RM) entre la DFM, considérée comme phénotype d'issue et les différents traits, considérés comme des facteurs de risque. N.SNP : nombre de SNP utilisé comme variable instrumentale dans l'analyse RM. BETA : la taille de l'effet obtenue à partir de l'analyse RM avec les quatre approches. SE : erreur standard de la taille de l'effet, P : valeur P de l'association, P<sub>Hétérogénéité</sub> : *p-value* du test d'hétérogénéité, Intercepte : valeur de l'ordonnée à l'origine et son IC à 95 %, Pléiotropie : *p-value* du test de pléiotropie.

La Figure 33 représente le Forest plot de la randomisation mendélienne entre la DFM et les trois traits étudiés (l'âge de la puberté, l'âge de la ménopause et la durée d'expositions aux hormones), en utilisant les résultats de l'approche IVW, car en absence de la pléiotropie ( $P_{\text{pléiotropie}} = 0.42, 0.33$  et  $0.80$  respectivement), les estimations d'IVW restent les plus robustes. Ces résultats montrent que la base génétique responsable de l'âge de la puberté, l'âge de la ménopause et la durés d'expositions aux hormones augmenteraient le risque de la DFM (les  $BETA > 0$ , voir Tableau 11). Cependant, l'absence de significativité de chaque *p-value* ( $P=0.31, 0.39$  et  $0.68$  des trois traits respectifs) ne permet pas de conclure sur la causalité entre les traits liés à la physiologie de la femme et la DFM.

### 3. Conclusion

Dans cette deuxième partie de mon travail, je me suis intéressée à comprendre pourquoi la DFM est majoritairement présente plus j'ai les femmes que chez les hommes. Autrement dit, j'ai voulu savoir si les traits liés à la physiologie de la femme tels que : l'âge de la puberté, l'âge de la ménopause et la durée d'exposition aux hormones ont un impact sur de la DFM. Pour bien mener cette étude, j'ai utilisé plusieurs méthodes permettant de définir l'existence de ce lien, et de quelle nature.

La première partie de mon travail a consisté à analyser l'association entre les traits cités précédemment et le risque de la DFM. Pour se faire, j'ai utilisé la méthode du score polygénique (PRS), en se basant sur les résumés statistiques de ces traits disponibles sur *UK Biobank*. Cette méthode m'a permis de tester l'association entre les scores obtenus pour chaque trait et la DFM, en utilisant les 3 cohortes Européennes dont nous disposons. Les résultats de la méta-analyse que j'ai obtenu n'ont montré aucune association entre des trois traits étudiés et le risque de la DFM.

Dans la deuxième partie de mon travail, je me suis intéressée à connaître le lien entre ces traits étudiés précédemment et la DFM. Pour cela, j'ai utilisé la méthode de la corrélation génétique. Les résultats obtenus ne sont pas significatifs, cependant, il n'est pas possible de conclure sur la corrélation entre les traits étudiés et la DFM. En ajoutant à ça, j'ai utilisé la méthode de la randomisation mendélienne pour savoir s'il existe un lien de causalité entre les traits liés à la physiologie de la femme et la DFM. Les résultats de cette analyse ont également montré l'absence d'association significative entre ces traits et la DFM, mais cela ne permet pas encore une fois de tirer une conclusion sur le lien de causalité.

## Partie III

### Discussion générale et perspectives

Les différences de sexe dans les maladies cardiovasculaires ont été le sujet de milliers de publications pendant des années. A savoir, la dysplasie fibromusculaire artérielle (DFM) et la dissection spontanée de l'Artère Coronaire (SCAD), deux pathologies affectant principalement les femmes d'âge moyen, faisant le sujet principal de ma thèse.

Notre équipe de recherche a pour objectif d'étudier et de comprendre l'origine génétique de la DFM et de la SCAD, en menant des études d'association à l'échelle du génome dans les plus grandes cohortes existantes de patients atteints de ces deux pathologies. Des résultats intéressants pour la DFM et la SCAD ont été rapportés dans les travaux de George *et al* [60] Adlam et al [273], respectivement. Ma contribution dans cette équipe de recherche est de comprendre et d'expliquer cette prédominance féminine dans la DFM et la SCAD. Pour cela j'ai été amenée à réaliser des analyses d'associations génétiques spécifiquement sur les variants génétiques du chromosome X, ainsi que l'étude des liens génétiques entre les traits liés à la physiologie de la femme et le risque de la DFM et la SCAD.

Dans cette première partie de ma de thèse, je me suis intéressée à l'étude de l'impact des variations génétiques fréquentes du Chr X dans la prédisposition à la DFM et à la SCAD. Mon travail a consisté à réaliser des analyses d'associations génétiques spécifiquement sur les variants génétiques du chromosome X. Dans un premier temps, j'ai utilisé deux approches différentes dans l'étude d'association de la DFM pour les trois études cas-témoins d'origines Européennes : ARCADIA-3C, ARCADIA-POL et FEIRI. Cette différence est dans le choix du modèle de régression utilisée. Le premier modèle (modèle 1) ne tient pas compte de l'inactivation de X. Quant au deuxième modèle (modèle 2), j'ai utilisé l'approche XWAS qui intègre une procédure d'analyse bien spécifique au chromosome X, notamment l'étape du contrôle qualité. Les résultats d'association obtenus de ces deux modèles analytiques n'ont identifié aucun nouveau locus associé significativement avec le risque de la DFM, sauf pour la cohorte ARCADIA-3C avec le modèle 2 (Figure 24), où j'ai obtenu un nouveau locus associé à la DFM. Néanmoins ce SNP significatif a disparu après la méta-analyse. Pour expliquer cette disparition, on a émis l'hypothèse que ce SNP pourrait être un outlier parce que les cas et les

témoins n'étaient pas génotypés sur le même puce, ou que les cas pourraient avoir un profil de recrutement différent des autres cohortes incluses dans l'analyse. L'absence d'un résultat significatif dans les analyses d'association précédente pour la DFM peut être expliquée par les faibles tailles des cohortes utilisées. Afin de contourner ce problème, j'ai réalisé une méta-analyse en regroupant les trois études précédentes. Ensuite, j'ai testé l'association entre les SNPs sur le Chr X et la DFM, en utilisant les deux modèles de régression à effet fixe sur l'échantillon final. Les résultats obtenus ne montrent aucune association significative à la DFM, même celui identifié précédemment dans l'étude ARCADIA-3C, ce qui explique probablement que ce signal était un faux positif qui a été supprimé dans l'étape du QC de la méta-analyse. Cependant, j'ai constaté que le modèle sans l'inactivation de X inclus plus de fluctuations que le modèle avec l'inactivation de X, parce que le codage entre les hommes et les femmes n'est pas adéquat à l'inactivation de X chez les femmes. Donc l'approche XWAS est de loin plus adéquate pour étudier l'association entre les variants du Chr X et la DFM.

En voulant creuser un peu plus sur le choix du modèle, j'ai utilisé dans un second temps, le modèle d'association intégré dans PLINK v2.0 [221], qui prend en considération l'inactivation de X par défaut (le même codage 0.2 pour les hommes et les femmes). Les résultats obtenus pour ces deux modèles sont similaires, à part une différence dans le nombre de SNPs après le contrôle qualité, où le modèle XWAS a tendance à exclure plus de SNPs, car il applique trois fois le QC dans l'analyse (le QC standard, chez les hommes et chez les femmes séparément). Donc c'est pour cette raison que j'ai choisi d'utiliser le modèle implémenté dans PLINK v2.0 [221] pour le reste des analyses d'associations.

J'ai réalisé par la suite une méta-analyse en combinant six études GWAS de la DFM (1556 cas et 7100 témoins), afin d'apporter plus de puissance statistique à mon étude d'association. Le résultat reste le même que les analyses précédentes, l'absence d'un ou plusieurs signaux significatifs avec la DFM à cause de taille de l'échantillon qui reste relativement faible pour cette étude d'association.

En ce qui concerne l'étude d'association de la SCAD, la méta-analyse de 8 études GWAS indépendantes a identifié un nouveau locus situé sur Chr X associé à la SCAD : *SRSF2P1/TMEM47* (rs5973204, OR= 1.36, P =  $3.0 \times 10^{-8}$ ). Pour les autres méta-analyses effectuées en stratifiant par la DFM et en tenant compte que des femmes, les résultats obtenus

n'étaient pas significatifs. La raison principale de cette absence d'association dans la majorité des analyses est le nombre faible des données dans les cohortes.

Dans la deuxième partie de ma thèse, je me suis intéressée à comprendre le fait que la DFM est majoritairement présente chez les femmes que chez les hommes. Autrement dit, j'ai voulu savoir si les traits liés à la physiologie de la femme tels que : l'âge de la puberté, l'âge de la ménopause et la durée d'exposition aux hormones ont un impact sur de la DFM. Pour bien mener cette étude, j'ai utilisé plusieurs méthodes permettant de définir l'existence de ce lien et sa nature.

La première méthode a consisté à analyser l'association entre les traits en question et le risque de la DFM. Pour se faire, j'ai utilisé la méthode du score polygénique (PRS), en se basant sur les résumés statistiques de ces traits disponibles sur *UK Biobank*. Cette méthode m'a permis de tester l'association entre les scores obtenus pour chaque trait et la DFM, en utilisant les 3 cohortes Européennes dont nous disposons. Les résultats de la méta-analyse que j'ai obtenus n'ont montré aucune association entre les trois traits étudiés et le risque de la DFM. Ceci peut s'expliquer par les deux hypothèses suivantes : la première est le manque de puissance statistique, c'est-à-dire que la taille de nos trois études cas-témoins est petite pour pouvoir permettre l'identification des SNPs associés avec la DFM. La deuxième hypothèse est que l'âge de la puberté, l'âge de la ménopause et la durée d'exposition aux hormones n'ont aucun rôle dans le risque de la DFM.

Après cette analyse, je me suis intéressée à connaître le lien entre ces traits étudiés précédemment (Age de puberté, Age de ménopause et là la durée d'exposition aux hormones) et la DFM. Pour cela, j'ai utilisé la méthode de la corrélation génétique. Les résultats obtenus ne sont pas significatifs, cependant, il n'est pas possible de conclure sur la corrélation entre les traits étudiés et la DFM. En ajoutant à ça, j'ai utilisé la méthode de la randomisation mendélienne pour savoir s'il existe un lien de causalité entre les traits liés à la physiologie de la femme et la DFM. Les résultats de cette analyse ont également montré l'absence d'association significative entre ces traits et la DFM, mais cela ne permet pas encore une fois de tirer une conclusion sur le lien de causalité.

Pour conclure, le travail de recherche que j'ai effectué durant cette thèse m'a permis d'identifier pour la première fois un nouveau locus sur le Chr X associé à la SCAD. Ce résultat permet d'ouvrir la voie à de nouvelles recherches potentielles :

- Effectuer des analyses complémentaires pour connaître le gène cible et le mécanisme biologique impliqué.
- Analyser de plus grands échantillons de patients SCAD, pour évaluer l'apport de ce chromosome sexuel dans le risque de la DFM et de la SCAD.
- Etudier l'association des traits liés à la reproduction directement chez les patients en stratifiant par groupe d'âge, afin apporter des réponses plus précises sur le lien entre la génétique de ces traits et le risque de la DFM et la SCAD.
- Etudier l'association des scores génétiques sur différentes phases d'âge, telle que la période pré ménopause et post ménopause.
- Utiliser l'approche de la corrélation génétique et la randomisation mendélienne pour identifier le lien entre les traits liés à la reproduction chez les femmes le risque de la SCAD.

## Références bibliographiques

- [1] G. A. Roth *et al.*, « Global Burden of Cardiovascular Diseases and Risk Factors, 1990–2019: Update From the GBD 2019 Study », *Journal of the American College of Cardiology*, vol. 76, n° 25, p. 2982-3021, déc. 2020, doi: 10.1016/j.jacc.2020.11.010.
- [2] T. Vos *et al.*, « Global burden of 369 diseases and injuries in 204 countries and territories, 1990–2019: a systematic analysis for the Global Burden of Disease Study 2019 », *The Lancet*, vol. 396, n° 10258, p. 1204-1222, oct. 2020, doi: 10.1016/S0140-6736(20)30925-9.
- [3] « Cardiovascular diseases (CVDs) ». [https://www.who.int/fr/news-room/fact-sheets/detail/cardiovascular-diseases-\(cvds\)](https://www.who.int/fr/news-room/fact-sheets/detail/cardiovascular-diseases-(cvds)) (consulté le 1 juin 2022).
- [4] A. Timmis *et al.*, « European Society of Cardiology: Cardiovascular Disease Statistics 2017 », *European Heart Journal*, vol. 39, n° 7, p. 508-579, févr. 2018, doi: 10.1093/eurheartj/ehx628.
- [5] N. Townsend, L. Wilson, P. Bhatnagar, K. Wickramasinghe, M. Rayner, et M. Nichols, « Cardiovascular disease in Europe: epidemiological update 2016 », *Eur Heart J*, vol. 37, n° 42, p. 3232-3245, nov. 2016, doi: 10.1093/eurheartj/ehw334.
- [6] N. Townsend *et al.*, « Epidemiology of cardiovascular disease in Europe », *Nat Rev Cardiol*, vol. 19, n° 2, p. 133-143, févr. 2022, doi: 10.1038/s41569-021-00607-3.
- [7] K. H. Zou, J. Z. Li, L. A. Salem, J. Imperato, J. Edwards, et A. Ray, « Harnessing real-world evidence to reduce the burden of noncommunicable disease: health information technology and innovation to generate insights », *Health Serv Outcomes Res Method*, vol. 21, n° 1, p. 8-20, mars 2021, doi: 10.1007/s10742-020-00223-7.
- [8] « Coronary Heart Disease - What Is Coronary Heart Disease? | NHLBI, NIH ». <https://www.nhlbi.nih.gov/health/coronary-heart-disease> (consulté le 26 septembre 2022).
- [9] H. Id, « Épidémiologie des facteurs de risque cardiovasculaire en population tropicale - cas du Bénin », p. 236, 2015.
- [10] E. A. Khramtsova, L. K. Davis, et B. E. Stranger, « The role of sex in the genomics of human complex traits », *Nat Rev Genet*, vol. 20, n° 3, p. 173-190, mars 2019, doi: 10.1038/s41576-018-0083-1.
- [11] W. S. Aronow, « Peripheral arterial disease of the lower extremities », *Arch Med Sci*, vol. 8, n° 2, p. 375-388, mai 2012, doi: 10.5114/aoms.2012.28568.
- [12] K. Cassar, « Intermittent claudication », *BMJ*, vol. 333, n° 7576, p. 1002-1005, nov. 2006, doi: 10.1136/bmj.39001.562813.DE.

- [13] S. M. Grundy, R. Pasternak, P. Greenland, S. Smith, et V. Fuster, « Assessment of cardiovascular risk by use of multiple-risk-factor assessment equations: a statement for healthcare professionals from the American Heart Association and the American College of Cardiology », *Circulation*, vol. 100, n° 13, p. 1481-1492, sept. 1999, doi: 10.1161/01.cir.100.13.1481.
- [14] G. De Backer *et al.*, « European guidelines on cardiovascular disease prevention in clinical practice. Third Joint Task Force of European and Other Societies on Cardiovascular Disease Prevention in Clinical Practice », *Eur Heart J*, vol. 24, n° 17, p. 1601-1610, sept. 2003, doi: 10.1016/s0195-668x(03)00347-6.
- [15] M. H. Forouzanfar *et al.*, « Global Burden of Hypertension and Systolic Blood Pressure of at Least 110 to 115 mm Hg, 1990-2015 », *JAMA*, vol. 317, n° 2, p. 165-182, janv. 2017, doi: 10.1001/jama.2016.19043.
- [16] S. Oparil *et al.*, « Hypertension », *Nat Rev Dis Primers*, vol. 4, p. 18014, mars 2018, doi: 10.1038/nrdp.2018.14.
- [17] « Age-specific relevance of usual blood pressure to vascular mortality: a meta-analysis of individual data for one million adults in 61 prospective studies », *The Lancet*, vol. 360, n° 9349, p. 1903-1913, déc. 2002, doi: 10.1016/S0140-6736(02)11911-8.
- [18] E. Rapsomaniki *et al.*, « Blood pressure and incidence of twelve cardiovascular diseases: lifetime risks, healthy life-years lost, and age-specific associations in 1·25 million people », *Lancet*, vol. 383, n° 9932, p. 1899-1911, mai 2014, doi: 10.1016/S0140-6736(14)60685-1.
- [19] A. D. Shah *et al.*, « Type 2 diabetes and incidence of cardiovascular diseases: a cohort study in 1·9 million people », *Lancet Diabetes Endocrinol*, vol. 3, n° 2, p. 105-113, févr. 2015, doi: 10.1016/S2213-8587(14)70219-0.
- [20] L. Liu, B. Simon, J. Shi, A. K. Mallhi, et H. J. Eisen, « Impact of diabetes mellitus on risk of cardiovascular disease and all-cause mortality: Evidence on health outcomes and antidiabetic treatment in United States adults », *World J Diabetes*, vol. 7, n° 18, p. 449-461, oct. 2016, doi: 10.4239/wjd.v7.i18.449.
- [21] A. Z. Chase-Vilchez, I. H. Y. Chan, S. A. E. Peters, et M. Woodward, « Diabetes as a risk factor for incident peripheral arterial disease in women compared to men: a systematic review and meta-analysis », *Cardiovasc Diabetol*, vol. 19, n° 1, p. 151, sept. 2020, doi: 10.1186/s12933-020-01130-4.
- [22] S. Kodama *et al.*, « Diabetes mellitus and risk of new-onset and recurrent heart failure: a systematic review and meta-analysis », *ESC Heart Fail*, vol. 7, n° 5, p. 2146-2174, juill. 2020, doi: 10.1002/ehf2.12782.
- [23] H. E. Bays *et al.*, « Ten things to know about ten cardiovascular disease risk factors – 2022 », *American Journal of Preventive Cardiology*, vol. 10, p. 100342, juin 2022, doi: 10.1016/j.ajpc.2022.100342.

- [24] G. Gallucci, A. Tartarone, R. Lerose, A. V. Lalinga, et A. M. Capobianco, « Cardiovascular risk of smoking and benefits of smoking cessation », *J Thorac Dis*, vol. 12, n° 7, p. 3866-3876, juill. 2020, doi: 10.21037/jtd.2020.02.47.
- [25] S. Dalkou et C. Clair, « Tabagisme, vapotage et risque cardiovasculaire : mise au point », *Revue Médicale Suisse*, vol. 13, n° 566, p. 1186-1190, 2017, doi: 10.53738/REVMED.2017.13.566.1186.
- [26] B. J. North et D. A. Sinclair, « The Intersection Between Aging and Cardiovascular Disease », *Circulation Research*, vol. 110, n° 8, p. 1097-1108, avr. 2012, doi: 10.1161/CIRCRESAHA.111.246876.
- [27] A. B. Curtis, R. Karki, A. Hattoum, et U. C. Sharma, « Arrhythmias in Patients  $\geq 80$  Years of Age: Pathophysiology, Management, and Outcomes », *Journal of the American College of Cardiology*, vol. 71, n° 18, p. 2041-2057, mai 2018, doi: 10.1016/j.jacc.2018.03.019.
- [28] A. Yazdanyar et A. B. Newman, « The Burden of Cardiovascular Disease in the Elderly: Morbidity, Mortality, and Costs », *Clin Geriatr Med*, vol. 25, n° 4, p. 563-vii, nov. 2009, doi: 10.1016/j.cger.2009.07.007.
- [29] I. Csige *et al.*, « The Impact of Obesity on the Cardiovascular System », *J Diabetes Res*, vol. 2018, p. 3407306, 2018, doi: 10.1155/2018/3407306.
- [30] T. Mandviwala, U. Khalid, et A. Deswal, « Obesity and Cardiovascular Disease: a Risk Factor or a Risk Marker? », *Curr Atheroscler Rep*, vol. 18, n° 5, p. 21, mai 2016, doi: 10.1007/s11883-016-0575-4.
- [31] A. Briasoulis, V. Agarwal, et F. H. Messerli, « Alcohol consumption and the risk of hypertension in men and women: a systematic review and meta-analysis », *J Clin Hypertens (Greenwich)*, vol. 14, n° 11, p. 792-798, nov. 2012, doi: 10.1111/jch.12008.
- [32] M. J. G. Leening *et al.*, « Sex differences in lifetime risk and first manifestation of cardiovascular disease: prospective population based cohort study », *BMJ*, vol. 349, p. g5992, nov. 2014, doi: 10.1136/bmj.g5992.
- [33] A. Shi, Z. Tao, P. Wei, et J. Zhao, « Epidemiological aspects of heart diseases », *Exp Ther Med*, vol. 12, n° 3, p. 1645-1650, sept. 2016, doi: 10.3892/etm.2016.3541.
- [34] A. P. Arnold, L. A. Cassis, M. Eghbali, K. Reue, et K. Sandberg, « Sex Hormones and Sex Chromosomes Cause Sex Differences in the Development of Cardiovascular Diseases », *Arterioscler Thromb Vasc Biol*, vol. 37, n° 5, p. 746-756, mai 2017, doi: 10.1161/ATVBAHA.116.307301.
- [35] C. Maric-Bilkan *et al.*, « Report of the National Heart, Lung, and Blood Institute Working Group on Sex Differences Research in Cardiovascular Disease: Scientific Questions and Challenges », *Hypertension*, vol. 67, n° 5, p. 802-807, mai 2016, doi: 10.1161/HYPERTENSIONAHA.115.06967.

- [36] A. P. Arnold, « The organizational-activational hypothesis as the foundation for a unified theory of sexual differentiation of all mammalian tissues », *Horm Behav*, vol. 55, n° 5, p. 570-578, mai 2009, doi: 10.1016/j.yhbeh.2009.03.011.
- [37] S. Eggers, T. Ohnesorg, et A. Sinclair, « Genetic regulation of mammalian gonad development », *Nat Rev Endocrinol*, vol. 10, n° 11, p. 673-683, nov. 2014, doi: 10.1038/nrendo.2014.163.
- [38] F. Vaura, J. Palmu, J. Aittokallio, A. Kauko, et T. Niiranen, « Genetic, Molecular, and Cellular Determinants of Sex-Specific Cardiovascular Traits », *Circulation Research*, vol. 130, n° 4, p. 611-631, févr. 2022, doi: 10.1161/CIRCRESAHA.121.319891.
- [39] A. Wawrzekiewicz-Jałowicka, A. Lalik, et G. Soveral, « Recent Update on the Molecular Mechanisms of Gonadal Steroids Action in Adipose Tissue », *Int J Mol Sci*, vol. 22, n° 10, p. 5226, mai 2021, doi: 10.3390/ijms22105226.
- [40] B. T. Palmisano, L. Zhu, R. H. Eckel, et J. M. Stafford, « Sex differences in lipid and lipoprotein metabolism », *Mol Metab*, vol. 15, p. 45-55, sept. 2018, doi: 10.1016/j.molmet.2018.05.008.
- [41] P. Marques, T. Madeira, et A. Gama, « Menstrual cycle among adolescents: girls' awareness and influence of age at menarche and overweight », *Rev Paul Pediatr*, vol. 40, p. e2020494, 2022, doi: 10.1590/1984-0462/2022/40/2020494.
- [42] I. M. P. S. Ilankoon, K. Samarasinghe, et C. Elgán, « Menopause is a natural stage of aging: a qualitative study », *BMC Women's Health*, vol. 21, n° 1, p. 47, févr. 2021, doi: 10.1186/s12905-020-01164-6.
- [43] N. F. Goodman *et al.*, « American Association of Clinical Endocrinologists Medical Guidelines for Clinical Practice for the Diagnosis and Treatment of Menopause », *Endocrine Practice*, vol. 17, p. 1-25, nov. 2011, doi: 10.4158/EP.17.S6.1.
- [44] « Préménopause et péri-ménopause : Effets sur l'organisme et gestion de l'activité physique », *valdemarne.fr*. <https://www.valdemarne.fr/newsletters/sport-sante-et-preparation-physique/premenopause-et-perimenopause-effets-sur-lorganisme-et-gestion-de-lactivite-physique> (consulté le 1 juin 2022).
- [45] D. Xiang, Y. Liu, S. Zhou, E. Zhou, et Y. Wang, « Protective Effects of Estrogen on Cardiovascular Disease Mediated by Oxidative Stress », *Oxidative Medicine and Cellular Longevity*, vol. 2021, p. 1-15, juin 2021, doi: 10.1155/2021/5523516.
- [46] D. Zhao *et al.*, « Endogenous Sex Hormones and Incident Cardiovascular Disease in Post-Menopausal Women », *J Am Coll Cardiol*, vol. 71, n° 22, p. 2555-2566, juin 2018, doi: 10.1016/j.jacc.2018.01.083.
- [47] E. J. Benjamin *et al.*, « Heart Disease and Stroke Statistics—2019 Update: A Report From the American Heart Association », *Circulation*, vol. 139, n° 10, p. e56-e528, mars 2019, doi: 10.1161/CIR.0000000000000659.

- [48] A. Villa, N. Rizzi, E. Vegeto, P. Ciana, et A. Maggi, « Estrogen accelerates the resolution of inflammation in macrophagic cells », *Sci Rep*, vol. 5, p. 15224, oct. 2015, doi: 10.1038/srep15224.
- [49] A. Iorga, C. M. Cunningham, S. Moazeni, G. Ruffenach, S. Umar, et M. Eghbali, « The protective role of estrogen and estrogen receptors in cardiovascular disease and the controversial use of estrogen therapy », *Biol Sex Differ*, vol. 8, n° 1, p. 33, oct. 2017, doi: 10.1186/s13293-017-0152-8.
- [50] J. R. Bell *et al.*, « Aromatase transgenic upregulation modulates basal cardiac performance and the response to ischemic stress in male mice », *Am J Physiol Heart Circ Physiol*, vol. 306, n° 9, p. H1265-1274, mai 2014, doi: 10.1152/ajpheart.00012.2014.
- [51] W. H. Parker, V. Jacoby, D. Shoupe, et W. Rocca, « Effect of bilateral oophorectomy on women's long-term health », *Womens Health (Lond)*, vol. 5, n° 5, p. 565-576, sept. 2009, doi: 10.2217/whe.09.42.
- [52] D. M. Lloyd-Jones *et al.*, « Prediction of Lifetime Risk for Cardiovascular Disease by Risk Factor Burden at 50 Years of Age », *Circulation*, vol. 113, n° 6, p. 791-798, févr. 2006, doi: 10.1161/CIRCULATIONAHA.105.548206.
- [53] J. J. Lee *et al.*, « Age at Menarche and Risk of Cardiovascular Disease Outcomes: Findings From the National Heart Lung and Blood Institute-Sponsored Women's Ischemia Syndrome Evaluation », *Journal of the American Heart Association*, vol. 8, n° 12, p. e012406, juin 2019, doi: 10.1161/JAHA.119.012406.
- [54] K. E. Remsberg, E. W. Demerath, C. M. Schubert, Wm. C. Chumlea, S. S. Sun, et R. M. Siervogel, « Early Menarche and the Development of Cardiovascular Disease Risk Factors in Adolescent Girls: The Fels Longitudinal Study », *The Journal of Clinical Endocrinology & Metabolism*, vol. 90, n° 5, p. 2718-2724, mai 2005, doi: 10.1210/jc.2004-1991.
- [55] R. Lakshman *et al.*, « Early Age at Menarche Associated with Cardiovascular Disease and Mortality », *The Journal of Clinical Endocrinology & Metabolism*, vol. 94, n° 12, p. 4953-4960, déc. 2009, doi: 10.1210/jc.2009-1789.
- [56] F. R. Day, C. E. Elks, A. Murray, K. K. Ong, et J. R. B. Perry, « Puberty timing associated with diabetes, cardiovascular disease and also diverse health outcomes in men and women: the UK Biobank study », *Sci Rep*, vol. 5, n° 1, Art. n° 1, juin 2015, doi: 10.1038/srep11208.
- [57] D. Canoy *et al.*, « Age at Menarche and Risks of Coronary Heart and Other Vascular Diseases in a Large UK Cohort », *Circulation*, vol. 131, n° 3, p. 237-244, janv. 2015, doi: 10.1161/CIRCULATIONAHA.114.010070.
- [58] B. K. Jacobsen, K. Oda, S. F. Knutsen, et G. E. Fraser, « Age at menarche, total mortality and mortality from ischaemic heart disease and stroke: the Adventist Health Study, 1976-88 », *International Journal of Epidemiology*, vol. 38, n° 1, p. 245-252, févr. 2009, doi: 10.1093/ije/dyn251.

- [59] J. L. Rodgers *et al.*, « Cardiovascular Risks Associated with Gender and Aging », *J Cardiovasc Dev Dis*, vol. 6, n° 2, p. 19, avr. 2019, doi: 10.3390/jcdd6020019.
- [60] A. Georges *et al.*, « Genetic investigation of fibromuscular dysplasia identifies risk loci and shared genetics with common cardiovascular diseases », *Nat Commun*, vol. 12, n° 1, p. 6031, oct. 2021, doi: 10.1038/s41467-021-26174-2.
- [61] M. H. Khoury et H. L. Gornik, « Fibromuscular dysplasia (FMD) », *Vasc Med*, vol. 22, n° 3, p. 248-252, juin 2017, doi: 10.1177/1358863X17700716.
- [62] P.-F. Plouin, J. Perdu, A. La Batide-Alanore, P. Boutouyrie, A.-P. Gimenez-Roqueplo, et X. Jeunemaitre, « Fibromuscular dysplasia », *Orphanet J Rare Dis*, vol. 2, p. 28, juin 2007, doi: 10.1186/1750-1172-2-28.
- [63] K. C. Michelis, J. W. Olin, D. Kadian-Dodov, V. d'Escamard, et J. C. Kovacic, « Coronary Artery Manifestations of Fibromuscular Dysplasia », *J Am Coll Cardiol*, vol. 64, n° 10, p. 1033-1046, sept. 2014, doi: 10.1016/j.jacc.2014.07.014.
- [64] E. K. Brinza et H. L. Gornik, « Fibromuscular dysplasia: Advances in understanding and management », *Cleve Clin J Med*, vol. 83, n° 11 Suppl 2, p. S45-S51, nov. 2016, doi: 10.3949/ccjm.83.s2.06.
- [65] J. W. Olin *et al.*, « Fibromuscular dysplasia: state of the science and critical unanswered questions: a scientific statement from the American Heart Association », *Circulation*, vol. 129, n° 9, p. 1048-1078, mars 2014, doi: 10.1161/01.cir.0000442577.96802.8c.
- [66] W. F. Leadbetter et C. E. Burkland, « Hypertension in Unilateral Renal Disease », *Journal of Urology*, vol. 39, n° 5, p. 611-626, mai 1938, doi: 10.1016/S0022-5347(17)71895-2.
- [67] E. G. Harrison et L. J. McCormack, « Pathologic classification of renal arterial disease in renovascular hypertension », *Mayo Clin Proc*, vol. 46, n° 3, p. 161-167, mars 1971.
- [68] « Progression of renal artery fibromuscular dysplasia in 42 patients as seen on angiography - PubMed ». <https://pubmed.ncbi.nlm.nih.gov/7208940/> (consulté le 2 juin 2022).
- [69] J. C. Stanley et W. J. Fry, « Renovascular hypertension secondary to arterial fibrodysplasia in adults: criteria for operation and results of surgical therapy », *Arch Surg*, vol. 110, n° 8, p. 922-928, août 1975, doi: 10.1001/archsurg.1975.01360140066014.
- [70] E. S. H. Kim, J. Saw, D. Kadian-Dodov, M. Wood, et S. K. Ganesh, « FMD and SCAD: Sex-Biased Arterial Diseases With Clinical and Genetic Pleiotropy », *Circ Res*, vol. 128, n° 12, p. 1958-1972, juin 2021, doi: 10.1161/CIRCRESAHA.121.318300.
- [71] S. Savard, A. Azarine, X. Jeunemaitre, M. Azizi, P.-F. Plouin, et O. Steichen, « Association of smoking with phenotype at diagnosis and vascular interventions in

- patients with renal artery fibromuscular dysplasia », *Hypertension*, vol. 61, n° 6, p. 1227-1232, juin 2013, doi: 10.1161/HYPERTENSIONAHA.111.00838.
- [72] H. L. Gornik *et al.*, « First International Consensus on the diagnosis and management of fibromuscular dysplasia », *Vasc Med*, vol. 24, n° 2, p. 164-189, avr. 2019, doi: 10.1177/1358863X18821816.
- [73] J. Perdu *et al.*, « Inheritance of arterial lesions in renal fibromuscular dysplasia », *J Hum Hypertens*, vol. 21, n° 5, p. 393-400, mai 2007, doi: 10.1038/sj.jhh.1002156.
- [74] A. Mackay, J. J. Brown, A. M. Cumming, C. Isles, A. F. Lever, et J. I. Robertson, « Smoking and renal artery stenosis. », *Br Med J*, vol. 2, n° 6193, p. 770, sept. 1979.
- [75] C. N. Sang *et al.*, « Etiologic factors in renovascular fibromuscular dysplasia. A case-control study », *Hypertension*, vol. 14, n° 5, p. 472-479, nov. 1989, doi: 10.1161/01.hyp.14.5.472.
- [76] S. O'Connor *et al.*, « Smoking and Adverse Outcomes in Fibromuscular Dysplasia: U.S. Registry Report », *Journal of the American College of Cardiology*, vol. 67, n° 14, p. 1750-1751, avr. 2016, doi: 10.1016/j.jacc.2016.01.058.
- [77] S. Savard, O. Steichen, A. Azarine, M. Azizi, X. Jeunemaitre, et P.-F. Plouin, « Association Between 2 Angiographic Subtypes of Renal Artery Fibromuscular Dysplasia and Clinical Characteristics », *Circulation*, vol. 126, n° 25, p. 3062-3069, déc. 2012, doi: 10.1161/CIRCULATIONAHA.112.117499.
- [78] J. W. Olin *et al.*, « The United States Registry for Fibromuscular Dysplasia: results in the first 447 patients », *Circulation*, vol. 125, n° 25, p. 3182-3190, juin 2012, doi: 10.1161/CIRCULATIONAHA.112.091223.
- [79] I. Pannier-Moreau *et al.*, « Possible familial origin of multifocal renal artery fibromuscular dysplasia », *J Hypertens*, vol. 15, n° 12 Pt 2, p. 1797-1801, déc. 1997, doi: 10.1097/00004872-199715120-00092.
- [80] T. F. Lüscher, J. T. Lie, A. W. Stanson, O. W. Houser, L. H. Hollier, et S. G. Sheps, « Arterial Fibromuscular Dysplasia », *Mayo Clinic Proceedings*, vol. 62, n° 10, p. 931-952, oct. 1987, doi: 10.1016/S0025-6196(12)65051-4.
- [81] E. S. H. Kim *et al.*, « Clinical Manifestations of Fibromuscular Dysplasia Vary by Patient Sex: A Report of the United States Registry for Fibromuscular Dysplasia », *Journal of the American College of Cardiology*, vol. 62, n° 21, p. 2026-2028, nov. 2013, doi: 10.1016/j.jacc.2013.07.038.
- [82] F. Silhol *et al.*, « Intranuclear Expression of Progesterone Receptors in Smooth Muscle Cells of Renovascular Fibromuscular Dysplasia: A Pilot Study », *Annals of Vascular Surgery*, vol. 29, n° 4, p. 830-835, mai 2015, doi: 10.1016/j.avsg.2014.10.025.
- [83] D. Adlam *et al.*, « European Society of Cardiology, acute cardiovascular care association, SCAD study group: a position paper on spontaneous coronary artery

- dissection », *European Heart Journal*, vol. 39, n° 36, p. 3353-3368, sept. 2018, doi: 10.1093/eurheartj/ehy080.
- [84] M. S. Tweet, S. N. Hayes, E. Codsí, R. Gulati, C. H. Rose, et P. J. M. Best, « Spontaneous Coronary Artery Dissection Associated With Pregnancy », *Journal of the American College of Cardiology*, vol. 70, n° 4, p. 426-435, juill. 2017, doi: 10.1016/j.jacc.2017.05.055.
- [85] K. Poon, B. Bell, O. C. Raffel, D. L. Walters, et I.-K. Jang, « Spontaneous Coronary Artery Dissection: Utility of Intravascular Ultrasound and Optical Coherence Tomography During Percutaneous Coronary Intervention », *Circ: Cardiovascular Interventions*, vol. 4, n° 2, avr. 2011, doi: 10.1161/CIRCINTERVENTIONS.110.959593.
- [86] J. Saw, D. Ricci, A. Starovoytov, R. Fox, et C. E. Buller, « Spontaneous coronary artery dissection: prevalence of predisposing conditions including fibromuscular dysplasia in a tertiary center cohort », *JACC Cardiovasc Interv*, vol. 6, n° 1, p. 44-52, janv. 2013, doi: 10.1016/j.jcin.2012.08.017.
- [87] M. S. Tweet *et al.*, « Spontaneous Coronary Artery Dissection: Revascularization Versus Conservative Therapy », *Circ: Cardiovascular Interventions*, vol. 7, n° 6, p. 777-786, déc. 2014, doi: 10.1161/CIRCINTERVENTIONS.114.001659.
- [88] S. Rogowski *et al.*, « Spontaneous Coronary Artery Dissection: Angiographic Follow-Up and Long-Term Clinical Outcome in a Predominantly Medically Treated Population », *Cathet. Cardiovasc. Intervent.*, vol. 89, n° 1, p. 59-68, janv. 2017, doi: 10.1002/ccd.26383.
- [89] P. Díez-Villanueva *et al.*, « Spontaneous Coronary Artery Dissection and Menopause », *The American Journal of Cardiology*, vol. 148, p. 53-59, juin 2021, doi: 10.1016/j.amjcard.2021.02.007.
- [90] J. Saw *et al.*, « Spontaneous Coronary Artery Dissection », *Journal of the American College of Cardiology*, vol. 70, n° 9, p. 1148-1158, août 2017, doi: 10.1016/j.jacc.2017.06.053.
- [91] L. McGrath-Cadell, P. McKenzie, S. Emmanuel, D. W. M. Muller, R. M. Graham, et C. J. Holloway, « Outcomes of patients with spontaneous coronary artery dissection », *Open Heart*, vol. 3, n° 2, p. e000491, 2016, doi: 10.1136/openhrt-2016-000491.
- [92] D. Adlam, M. García-Guimaraes, et A. H. E. M. Maas, « Spontaneous coronary artery dissection: no longer a rare disease », *Eur Heart J*, vol. 40, n° 15, p. 1198-1201, avr. 2019, doi: 10.1093/eurheartj/ehz048.
- [93] S. N. Kok *et al.*, « Prevalence and Clinical Factors of Migraine in Patients With Spontaneous Coronary Artery Dissection », *JAHA*, vol. 7, n° 24, p. e010140, déc. 2018, doi: 10.1161/JAHA.118.010140.

- [94] R. Clare *et al.*, « Characteristics and Clinical Outcomes of Patients With Spontaneous Coronary Artery Dissection », *JAHA*, vol. 8, n° 10, p. e012570, mai 2019, doi: 10.1161/JAHA.119.012570.
- [95] S. Sharma *et al.*, « Risk Factors, Imaging Findings, and Sex Differences in Spontaneous Coronary Artery Dissection », *Am J Cardiol*, vol. 123, n° 11, p. 1783-1787, juin 2019, doi: 10.1016/j.amjcard.2019.02.040.
- [96] J. Saw *et al.*, « Canadian spontaneous coronary artery dissection cohort study: in-hospital and 30-day outcomes », *European Heart Journal*, vol. 40, n° 15, p. 1188-1197, avr. 2019, doi: 10.1093/eurheartj/ehz007.
- [97] P. W. McNair, A. Parker, A. Taylor, R. Battle, P. Norton, et A. M. Sharma, « Spontaneous Coronary Artery Dissection and Its Association With Fibromuscular Dysplasia and Other Vascular Abnormalities », *American Journal of Cardiology*, vol. 125, n° 1, p. 34-39, janv. 2020, doi: 10.1016/j.amjcard.2019.09.043.
- [98] D. J. Hunter, « Gene–environment interactions in human diseases », *Nat Rev Genet*, vol. 6, n° 4, p. 287-298, avr. 2005, doi: 10.1038/nrg1578.
- [99] M. A. Nordlie, L. E. Wold, et R. A. Kloner, « Genetic contributors toward increased risk for ischemic heart disease », *Journal of Molecular and Cellular Cardiology*, vol. 39, n° 4, p. 667-679, oct. 2005, doi: 10.1016/j.yjmcc.2005.06.006.
- [100] G. Biino *et al.*, « Environmental and Genetic Contribution to Hypertension Prevalence: Data from an Epidemiological Survey on Sardinian Genetic Isolates », *PLoS One*, vol. 8, n° 3, p. e59612, mars 2013, doi: 10.1371/journal.pone.0059612.
- [101] R. A. Chaer, R. Billeh, et M. G. Massad, « Genetics and Gene Manipulation Therapy of Premature Coronary Artery Disease », *CRD*, vol. 101, n° 1-3, p. 122-130, 2004, doi: 10.1159/000075993.
- [102] P. R. Burton, M. D. Tobin, et J. L. Hopper, « Genetic Epidemiology 1 Key concepts in genetic epidemiology », vol. 366, p. 11, 2005.
- [103] E. S. Lander et N. J. Schork, « Genetic Dissection of Complex Traits », vol. 265, p. 12, 1994.
- [104] J. A. Douglas, M. Boehnke, et K. Lange, « A Multipoint Method for Detecting Genotyping Errors and Mutations in Sibling-Pair Linkage Data », *The American Journal of Human Genetics*, vol. 66, n° 4, p. 1287-1297, avr. 2000, doi: 10.1086/302861.
- [105] T. J. Jorgensen, I. Ruczinski, B. Kessing, M. W. Smith, Y. Y. Shugart, et A. J. Alberg, « Hypothesis-Driven Candidate Gene Association Studies: Practical Design and Analytical Considerations », *American Journal of Epidemiology*, vol. 170, n° 8, p. 986-993, oct. 2009, doi: 10.1093/aje/kwp242.

- [106] D. M. Shivapour, P. Erwin, et E. S. Kim, « Epidemiology of fibromuscular dysplasia: A review of the literature », *Vasc Med*, vol. 21, n° 4, p. 376-381, août 2016, doi: 10.1177/1358863X16637913.
- [107] S. L. Poloskey *et al.*, « Low yield of genetic testing for known vascular connective tissue disorders in patients with fibromuscular dysplasia », *Vasc Med*, vol. 17, n° 6, p. 371-378, déc. 2012, doi: 10.1177/1358863X12459650.
- [108] S. K. Ganesh *et al.*, « Clinical and biochemical profiles suggest fibromuscular dysplasia is a systemic disease with altered TGF- $\beta$  expression and connective tissue features », *The FASEB Journal*, vol. 28, n° 8, p. 3313-3324, 2014, doi: 10.1096/fj.14-251207.
- [109] K. J. Carss *et al.*, « Spontaneous Coronary Artery Dissection », *Circulation: Genomic and Precision Medicine*, vol. 13, n° 6, p. e003030, déc. 2020, doi: 10.1161/CIRCGEN.120.003030.
- [110] M. I. Kaadan *et al.*, « Prospective Cardiovascular Genetics Evaluation in Spontaneous Coronary Artery Dissection », *Circulation: Genomic and Precision Medicine*, vol. 11, n° 4, p. e001933, avr. 2018, doi: 10.1161/CIRCGENETICS.117.001933.
- [111] H. K. Tabor, N. J. Risch, et R. M. Myers, « Candidate-gene approaches for studying complex genetic traits: practical considerations », *Nat Rev Genet*, vol. 3, n° 5, p. 391-397, mai 2002, doi: 10.1038/nrg796.
- [112] D. Altshuler, M. J. Daly, et E. S. Lander, « Genetic Mapping in Human Disease », *Science*, vol. 322, n° 5903, p. 881-888, nov. 2008, doi: 10.1126/science.1156409.
- [113] A. B. Singleton, « Exome sequencing: a transformative technology », *The Lancet Neurology*, vol. 10, n° 10, p. 942-946, oct. 2011, doi: 10.1016/S1474-4422(11)70196-X.
- [114] S. B. Ng *et al.*, « Targeted Capture and Massively Parallel Sequencing of Twelve Human Exomes », *Nature*, vol. 461, n° 7261, p. 272-276, sept. 2009, doi: 10.1038/nature08250.
- [115] S. R. Kiando *et al.*, « Exome sequencing in seven families and gene-based association studies indicate genetic heterogeneity and suggest possible candidates for fibromuscular dysplasia », *Journal of Hypertension*, vol. 33, n° 9, p. 1802-1810, sept. 2015, doi: 10.1097/HJH.0000000000000625.
- [116] Y. Sun *et al.*, « Association of TSR1 Variants and Spontaneous Coronary Artery Dissection », *Journal of the American College of Cardiology*, vol. 74, n° 2, p. 167-176, juill. 2019, doi: 10.1016/j.jacc.2019.04.062.
- [117] S. J. Monkley *et al.*, « Endothelial cell talin1 is essential for embryonic angiogenesis », *Dev Biol*, vol. 349, n° 2-3, p. 494-502, janv. 2011, doi: 10.1016/j.ydbio.2010.11.010.
- [118] A. Georges *et al.*, « Rare loss-of-function mutations of PTGIR are enriched in fibromuscular dysplasia », *Cardiovasc Res*, vol. 117, n° 4, p. 1154-1165, juin 2020, doi: 10.1093/cvr/cvaa161.

- [119] B. H. Majed et R. A. Khalil, « Molecular mechanisms regulating the vascular prostacyclin pathways and their adaptation during pregnancy and in the newborn », *Pharmacol Rev*, vol. 64, n° 3, p. 540-582, juill. 2012, doi: 10.1124/pr.111.004770.
- [120] J. Erdmann, T. Kessler, L. Munoz Venegas, et H. Schunkert, « A decade of genome-wide association studies for coronary artery disease: the challenges ahead », *Cardiovasc Res*, vol. 114, n° 9, p. 1241-1257, juill. 2018, doi: 10.1093/cvr/cvy084.
- [121] G. B. Ehret, « Genome-Wide Association Studies: Contribution of Genomics to Understanding Blood Pressure and Essential Hypertension », *Curr Hypertens Rep*, vol. 12, n° 1, p. 17-25, févr. 2010, doi: 10.1007/s11906-009-0086-6.
- [122] T. A. Manolio, « Genomewide Association Studies and Assessment of the Risk of Disease », *N Engl J Med*, vol. 363, n° 2, p. 166-176, juill. 2010, doi: 10.1056/NEJMra0905980.
- [123] Y. Nannya, K. Taura, M. Kurokawa, S. Chiba, et S. Ogawa, « Evaluation of genome-wide power of genetic association studies based on empirical data from the HapMap project », *Human Molecular Genetics*, vol. 16, n° 20, p. 2494-2505, oct. 2007, doi: 10.1093/hmg/ddm205.
- [124] International HapMap Consortium, « The International HapMap Project », *Nature*, vol. 426, n° 6968, p. 789-796, déc. 2003, doi: 10.1038/nature02168.
- [125] « A global reference for human genetic variation », *Nature*, vol. 526, n° 7571, p. 68-74, 2015, doi: 10.1038/nature15393.
- [126] A. I. Iglesias *et al.*, « Haplotype reference consortium panel: Practical implications of imputations with large reference panels », *Human Mutation*, vol. 38, n° 8, p. 1025-1032, 2017, doi: 10.1002/humu.23247.
- [127] P.-R. Loh *et al.*, « Reference-based phasing using the Haplotype Reference Consortium panel », *Nat Genet*, vol. 48, n° 11, p. 1443-1448, nov. 2016, doi: 10.1038/ng.3679.
- [128] N. J. Samani *et al.*, « Genomewide Association Analysis of Coronary Artery Disease », *New England Journal of Medicine*, vol. 357, n° 5, p. 443-453, août 2007, doi: 10.1056/NEJMoa072366.
- [129] R. McPherson *et al.*, « A Common Allele on Chromosome 9 Associated with Coronary Heart Disease », *Science*, vol. 316, n° 5830, p. 1488-1491, juin 2007, doi: 10.1126/science.1142447.
- [130] A. Helgadottir *et al.*, « A Common Variant on Chromosome 9p21 Affects the Risk of Myocardial Infarction », *Science*, vol. 316, n° 5830, p. 1491-1493, juin 2007, doi: 10.1126/science.1142842.
- [131] « Genome-wide association study of 14,000 cases of seven common diseases and 3,000 shared controls », *Nature*, vol. 447, n° 7145, p. 661-678, juin 2007, doi: 10.1038/nature05911.

- [132] « Genome-wide association of early-onset myocardial infarction with common single nucleotide polymorphisms, common copy number variants, and rare copy number variants », *Nat Genet*, vol. 41, n° 3, p. 334-341, mars 2009, doi: 10.1038/ng.327.
- [133] J. Erdmann *et al.*, « New susceptibility locus for coronary artery disease on chromosome 3q22.3 », *Nat Genet*, vol. 41, n° 3, p. 280-282, mars 2009, doi: 10.1038/ng.307.
- [134] M. Preuss *et al.*, « Design of the Coronary ARtery Disease Genome-Wide Replication And Meta-Analysis (CARDIoGRAM) Study », *Circ Cardiovasc Genet*, vol. 3, n° 5, p. 475-483, oct. 2010, doi: 10.1161/CIRCGENETICS.109.899443.
- [135] J. F. Peden *et al.*, « A genome-wide association study in Europeans and South Asians identifies five new loci for coronary artery disease », *Nat Genet*, vol. 43, n° 4, Art. n° 4, avr. 2011, doi: 10.1038/ng.782.
- [136] C. Sudlow *et al.*, « UK Biobank: An Open Access Resource for Identifying the Causes of a Wide Range of Complex Diseases of Middle and Old Age », *PLoS Med*, vol. 12, n° 3, p. e1001779, mars 2015, doi: 10.1371/journal.pmed.1001779.
- [137] H. Schunkert *et al.*, « Large-scale association analyses identifies 13 new susceptibility loci for coronary artery disease », *Nat Genet*, vol. 43, n° 4, p. 333-338, mars 2011, doi: 10.1038/ng.784.
- [138] P. Deloukas *et al.*, « Large-scale association analysis identifies new risk loci for coronary artery disease », *Nat Genet*, vol. 45, n° 1, Art. n° 1, janv. 2013, doi: 10.1038/ng.2480.
- [139] D. Klarin *et al.*, « Genetic Analysis in UK Biobank Links Insulin Resistance and Transendothelial Migration Pathways to Coronary Artery Disease », *Nat Genet*, vol. 49, n° 9, p. 1392-1397, sept. 2017, doi: 10.1038/ng.3914.
- [140] N. Verweij, R. N. Eppinga, Y. Hagemmeijer, et P. van der Harst, « Identification of 15 novel risk loci for coronary artery disease and genetic risk of recurrent events, atrial fibrillation and heart failure », *Sci Rep*, vol. 7, n° 1, p. 2761, juin 2017, doi: 10.1038/s41598-017-03062-8.
- [141] C. P. Nelson *et al.*, « Association analyses based on false discovery rate implicate new loci for coronary artery disease », *Nat Genet*, vol. 49, n° 9, p. 1385-1391, sept. 2017, doi: 10.1038/ng.3913.
- [142] P. van der Harst et N. Verweij, « Identification of 64 Novel Genetic Loci Provides an Expanded View on the Genetic Architecture of Coronary Artery Disease », *Circ Res*, vol. 122, n° 3, p. 433-443, févr. 2018, doi: 10.1161/CIRCRESAHA.117.312086.
- [143] D. Guo *et al.*, « Loss-of-Function Mutations in YY1AP1 Lead to Grange Syndrome and a Fibromuscular Dysplasia-Like Vascular Disease », *The American Journal of Human Genetics*, vol. 100, n° 1, p. 21-30, janv. 2017, doi: 10.1016/j.ajhg.2016.11.008.

- [144] S. R. Kiando *et al.*, « PHACTR1 Is a Genetic Susceptibility Locus for Fibromuscular Dysplasia Supporting Its Complex Genetic Pattern of Inheritance », *PLoS Genet*, vol. 12, n° 10, p. e1006367, oct. 2016, doi: 10.1371/journal.pgen.1006367.
- [145] S. Di Monaco, A. Georges, J.-P. Lengelé, M. Vikkula, et A. Persu, « Genomics of Fibromuscular Dysplasia », *Int J Mol Sci*, vol. 19, n° 5, p. 1526, mai 2018, doi: 10.3390/ijms19051526.
- [146] S. O'Connor *et al.*, « Systemic connective tissue features in women with fibromuscular dysplasia », *Vasc Med*, vol. 20, n° 5, p. 454-462, oct. 2015, doi: 10.1177/1358863X15592192.
- [147] V. Anttila *et al.*, « Genome-wide meta-analysis identifies new susceptibility loci for migraine », *Nat Genet*, vol. 45, n° 8, p. 912-917, août 2013, doi: 10.1038/ng.2676.
- [148] T. N. Turley *et al.*, « Identification of Susceptibility Loci for Spontaneous Coronary Artery Dissection », *JAMA Cardiol*, vol. 5, n° 8, p. 1-10, août 2020, doi: 10.1001/jamacardio.2020.0872.
- [149] D. Adlam *et al.*, « Association of the PHACTR1/EDN1 Genetic Locus With Spontaneous Coronary Artery Dissection », *J Am Coll Cardiol*, vol. 73, n° 1, p. 58-66, janv. 2019, doi: 10.1016/j.jacc.2018.09.085.
- [150] J. Saw *et al.*, « Chromosome 1q21.2 and additional loci influence risk of spontaneous coronary artery dissection and myocardial infarction », *Nat Commun*, vol. 11, p. 4432, sept. 2020, doi: 10.1038/s41467-020-17558-x.
- [151] Y. Deming *et al.*, « Sex-specific genetic predictors of Alzheimer's disease biomarkers », *Acta Neuropathol*, vol. 136, n° 6, p. 857-872, déc. 2018, doi: 10.1007/s00401-018-1881-4.
- [152] E. L. I. Salfati, « Genetic determinants of cardiovascular disease : heritability and genetic risk score », phdthesis, Université René Descartes - Paris V, 2014. Consulté le: 1 juin 2022. [En ligne]. Disponible sur: <https://tel.archives-ouvertes.fr/tel-01127451>
- [153] E. Bernabeu, O. Canela-Xandri, K. Rawlik, A. Talenti, J. Prendergast, et A. Tenesa, « Sex differences in genetic architecture in the UK Biobank », *Nat Genet*, vol. 53, n° 9, p. 1283-1289, sept. 2021, doi: 10.1038/s41588-021-00912-0.
- [154] Y. Deming *et al.*, « Genome-wide association study identifies four novel loci associated with Alzheimer's endophenotypes and disease modifiers », *Acta Neuropathol*, vol. 133, n° 5, p. 839-856, mai 2017, doi: 10.1007/s00401-017-1685-y.
- [155] M. Traglia *et al.*, « Genetic Mechanisms Leading to Sex Differences Across Common Diseases and Anthropometric Traits », *Genetics*, vol. 205, n° 2, p. 979-992, févr. 2017, doi: 10.1534/genetics.116.193623.
- [156] T. Barnay et É. Defebvre, « L'effet des contraintes physiques du travail sur les maladies cardiovasculaires chez les femmes. Enquête Santé et itinéraire professionnel, vagues

2006 et 2010 », *Bulletin d'épidémiologie hebdomadaire*, 2016, Consulté le: 1 juin 2022. [En ligne]. Disponible sur: <https://hal-upec-upem.archives-ouvertes.fr/hal-01297582>

- [157] J. Yang *et al.*, « Genome-wide genetic homogeneity between sexes and populations for human height and body mass index », *Human Molecular Genetics*, vol. 24, n° 25, p. 7445-7449, déc. 2015, doi: 10.1093/hmg/ddv443.
- [158] S. L. Pulit *et al.*, « Meta-analysis of genome-wide association studies for body fat distribution in 694 649 individuals of European ancestry », *Hum Mol Genet*, vol. 28, n° 1, p. 166-174, janv. 2019, doi: 10.1093/hmg/ddy327.
- [159] A. Köttgen *et al.*, « Genome-wide association analyses identify 18 new loci associated with serum urate concentrations », *Nat Genet*, vol. 45, n° 2, p. 145-154, févr. 2013, doi: 10.1038/ng.2500.
- [160] T. W. Winkler *et al.*, « The Influence of Age and Sex on Genetic Associations with Adult Body Size and Shape: A Large-Scale Genome-Wide Interaction Study », *PLoS Genet*, vol. 11, n° 10, p. e1005378, oct. 2015, doi: 10.1371/journal.pgen.1005378.
- [161] Y. Adam *et al.*, « Polygenic Risk Score in African populations: progress and challenges ». F1000Research, 14 février 2022. doi: 10.12688/f1000research.76218.1.
- [162] E. R. Chimusa, S. Dalvie, C. Dandara, A. Wonkam, et G. K. Mazandu, « Post genome-wide association analysis: dissecting computational pathway/network-based approaches », *Briefings in Bioinformatics*, vol. 20, n° 2, p. 690-700, mars 2019, doi: 10.1093/bib/bby035.
- [163] B. Pasaniuc et A. L. Price, « Dissecting the genetics of complex traits using summary association statistics », *Nat Rev Genet*, vol. 18, n° 2, Art. n° 2, févr. 2017, doi: 10.1038/nrg.2016.142.
- [164] F. Dudbridge, « Power and Predictive Accuracy of Polygenic Risk Scores », *PLOS Genetics*, vol. 9, n° 3, p. e1003348, mars 2013, doi: 10.1371/journal.pgen.1003348.
- [165] J. A. Collister, X. Liu, et L. Clifton, « Calculating Polygenic Risk Scores (PRS) in UK Biobank: A Practical Guide for Epidemiologists », *Frontiers in Genetics*, vol. 13, 2022, Consulté le: 13 juillet 2022. [En ligne]. Disponible sur: <https://www.frontiersin.org/articles/10.3389/fgene.2022.818574>
- [166] S. W. Choi, T. S.-H. Mak, et P. F. O'Reilly, « Tutorial: a guide to performing polygenic risk score analyses », *Nat Protoc*, vol. 15, n° 9, Art. n° 9, sept. 2020, doi: 10.1038/s41596-020-0353-1.
- [167] C. M. Lewis et E. Vassos, « Polygenic risk scores: from research tools to clinical instruments », *Genome Medicine*, vol. 12, n° 1, p. 44, mai 2020, doi: 10.1186/s13073-020-00742-5.

- [168] O. Pain *et al.*, « Evaluation of polygenic prediction methodology within a reference-standardized framework », *PLoS Genetics*, vol. 17, n° 5, p. e1009021, mai 2021, doi: 10.1371/journal.pgen.1009021.
- [169] J. D. Wall et J. K. Pritchard, « Haplotype blocks and linkage disequilibrium in the human genome », *Nat Rev Genet*, vol. 4, n° 8, p. 587-597, août 2003, doi: 10.1038/nrg1123.
- [170] J. Euesden, C. M. Lewis, et P. F. O'Reilly, « PRSice: Polygenic Risk Score software », *Bioinformatics*, vol. 31, n° 9, p. 1466-1468, mai 2015, doi: 10.1093/bioinformatics/btu848.
- [171] N. R. Wray, S. H. Lee, D. Mehta, A. A. E. Vinkhuyzen, F. Dudbridge, et C. M. Middeldorp, « Research Review: Polygenic methods and their application to psychiatric traits », *Journal of Child Psychology and Psychiatry*, vol. 55, n° 10, p. 1068-1087, 2014, doi: 10.1111/jcpp.12295.
- [172] N. Arnold et W. Koenig, « Polygenic Risk Score: Clinically Useful Tool for Prediction of Cardiovascular Disease and Benefit from Lipid-Lowering Therapy? », *Cardiovasc Drugs Ther*, vol. 35, n° 3, p. 627-635, 2021, doi: 10.1007/s10557-020-07105-7.
- [173] T. D. Hadley, A. M. Agha, et C. M. Ballantyne, « How Do We Incorporate Polygenic Risk Scores in Cardiovascular Disease Risk Assessment and Management? », *Curr Atheroscler Rep*, vol. 23, n° 6, p. 28, avr. 2021, doi: 10.1007/s11883-021-00915-6.
- [174] F. Wünnemann *et al.*, « Validation of Genome-Wide Polygenic Risk Scores for Coronary Artery Disease in French Canadians », *Circ Genom Precis Med*, vol. 12, n° 6, p. e002481, juin 2019, doi: 10.1161/CIRCGEN.119.002481.
- [175] P. Natarajan *et al.*, « Polygenic Risk Score Identifies Subgroup With Higher Burden of Atherosclerosis and Greater Relative Benefit From Statin Therapy in the Primary Prevention Setting », *Circulation*, vol. 135, n° 22, p. 2091-2101, mai 2017, doi: 10.1161/CIRCULATIONAHA.116.024436.
- [176] N. Kurniansyah *et al.*, « A multi-ethnic polygenic risk score is associated with hypertension prevalence and progression throughout adulthood », *Nat Commun*, vol. 13, p. 3549, juin 2022, doi: 10.1038/s41467-022-31080-2.
- [177] J. D. Newman *et al.*, « Associations of a polygenic risk score with coronary artery disease phenotypes in the Prospective Multicenter Imaging Study for Evaluation of Chest Pain (PROMISE) trial », *American Heart Journal*, vol. 252, p. 12-15, oct. 2022, doi: 10.1016/j.ahj.2022.05.012.
- [178] T. A. Agbaedeng *et al.*, « Polygenic risk score and coronary artery disease: A meta-analysis of 979,286 participant data », *Atherosclerosis*, vol. 333, p. 48-55, sept. 2021, doi: 10.1016/j.atherosclerosis.2021.08.020.
- [179] N. T. Krarup *et al.*, « A genetic risk score of 45 coronary artery disease risk variants associates with increased risk of myocardial infarction in 6041 Danish individuals »,

*Atherosclerosis*, vol. 240, n° 2, p. 305-310, juin 2015, doi: 10.1016/j.atherosclerosis.2015.03.022.

- [180] « Polygenic Risk Score for Coronary Artery Disease Improves the Prediction of Early-Onset Myocardial Infarction and Mortality in Men | *Circulation: Genomic and Precision Medicine* ». [https://www-ahajournals-org.proxy.insermbiblio.inist.fr/doi/10.1161/CIRCGEN.121.003452?url\\_ver=Z39.88-2003&rfr\\_id=ori:rid:crossref.org&rfr\\_dat=cr\\_pub%20%20pubmed](https://www-ahajournals-org.proxy.insermbiblio.inist.fr/doi/10.1161/CIRCGEN.121.003452?url_ver=Z39.88-2003&rfr_id=ori:rid:crossref.org&rfr_dat=cr_pub%20%20pubmed) (consulté le 18 juillet 2022).
- [181] P. Kraft, H. Chen, et S. Lindström, « The Use Of Genetic Correlation And Mendelian Randomization Studies To Increase Our Understanding of Relationships Between Complex Traits. », *Curr Epidemiol Rep*, vol. 7, n° 2, p. 104-112, juin 2020, doi: 10.1007/s40471-020-00233-6.
- [182] K. Hou *et al.*, « Accurate estimation of SNP-heritability from biobank-scale data irrespective of genetic architecture », *Nat Genet*, vol. 51, n° 8, Art. n° 8, août 2019, doi: 10.1038/s41588-019-0465-0.
- [183] S. H. Lee, J. Yang, M. E. Goddard, P. M. Visscher, et N. R. Wray, « Estimation of pleiotropy between complex diseases using single-nucleotide polymorphism-derived genomic relationships and restricted maximum likelihood », *Bioinformatics*, vol. 28, n° 19, p. 2540-2542, oct. 2012, doi: 10.1093/bioinformatics/bts474.
- [184] J. Yang, S. H. Lee, M. E. Goddard, et P. M. Visscher, « GCTA: A Tool for Genome-wide Complex Trait Analysis », *The American Journal of Human Genetics*, vol. 88, n° 1, p. 76-82, janv. 2011, doi: 10.1016/j.ajhg.2010.11.011.
- [185] B. K. Bulik-Sullivan *et al.*, « LD Score regression distinguishes confounding from polygenicity in genome-wide association studies », *Nat Genet*, vol. 47, n° 3, p. 291-295, mars 2015, doi: 10.1038/ng.3211.
- [186] G. Ni, G. Moser, Schizophrenia Working Group of the Psychiatric Genomics Consortium, N. R. Wray, et S. H. Lee, « Estimation of Genetic Correlation via Linkage Disequilibrium Score Regression and Genomic Restricted Maximum Likelihood », *Am J Hum Genet*, vol. 102, n° 6, p. 1185-1194, juin 2018, doi: 10.1016/j.ajhg.2018.03.021.
- [187] G. Davey Smith et G. Hemani, « Mendelian randomization: genetic anchors for causal inference in epidemiological studies », *Hum Mol Genet*, vol. 23, n° R1, p. R89-98, sept. 2014, doi: 10.1093/hmg/ddu328.
- [188] S. Burgess, A. Butterworth, et S. G. Thompson, « Mendelian Randomization Analysis With Multiple Genetic Variants Using Summarized Data », *Genet Epidemiol*, vol. 37, n° 7, p. 658-665, nov. 2013, doi: 10.1002/gepi.21758.
- [189] G. D. Smith et S. Ebrahim, « “Mendelian randomization”: can genetic epidemiology contribute to understanding environmental determinants of disease? », *Int J Epidemiol*, vol. 32, n° 1, p. 1-22, févr. 2003, doi: 10.1093/ije/dyg070.

- [190] S. Burgess, R. M. Daniel, A. S. Butterworth, et S. G. Thompson, « Network Mendelian randomization: using genetic variants as instrumental variables to investigate mediation in causal pathways », *Int J Epidemiol*, vol. 44, n° 2, p. 484-495, avr. 2015, doi: 10.1093/ije/dyu176.
- [191] S. Burgess, F. Dudbridge, et S. G. Thompson, « Combining information on multiple instrumental variables in Mendelian randomization: comparison of allele score and summarized data methods », *Stat Med*, vol. 35, n° 11, p. 1880-1906, mai 2016, doi: 10.1002/sim.6835.
- [192] S. Burgess, A. Butterworth, A. Malarstig, et S. G. Thompson, « Use of Mendelian randomisation to assess potential benefit of clinical intervention », *BMJ*, vol. 345, p. e7325, nov. 2012, doi: 10.1136/bmj.e7325.
- [193] M. M. Glymour, E. J. Tchetgen Tchetgen, et J. M. Robins, « Credible Mendelian Randomization Studies: Approaches for Evaluating the Instrumental Variable Assumptions », *American Journal of Epidemiology*, vol. 175, n° 4, p. 332-339, févr. 2012, doi: 10.1093/aje/kwr323.
- [194] T. M. Palmer *et al.*, « Using multiple genetic variants as instrumental variables for modifiable risk factors », *Stat Methods Med Res*, vol. 21, n° 3, p. 223-242, juin 2012, doi: 10.1177/0962280210394459.
- [195] F. W. Stearns, « One Hundred Years of Pleiotropy: A Retrospective », *Genetics*, vol. 186, n° 3, p. 767-773, nov. 2010, doi: 10.1534/genetics.110.122549.
- [196] Y. Zhou *et al.*, « Shared Genetic Architecture and Causal Relationship Between Asthma and Cardiovascular Diseases: A Large-Scale Cross-Trait Analysis », *Frontiers in Genetics*, vol. 12, 2022, Consulté le: 24 juillet 2022. [En ligne]. Disponible sur: <https://www.frontiersin.org/articles/10.3389/fgene.2021.775591>
- [197] F. Demenais *et al.*, « Multiancestry association study identifies new asthma risk loci that colocalize with immune-cell enhancer marks », *Nat Genet*, vol. 50, n° 1, p. 42-53, janv. 2018, doi: 10.1038/s41588-017-0014-7.
- [198] K. G. Aragam *et al.*, « Phenotypic Refinement of Heart Failure in a National Biobank Facilitates Genetic Discovery », *Circulation*, nov. 2018, doi: 10.1161/CIRCULATIONAHA.118.035774.
- [199] R. Malik *et al.*, « Multiancestry genome-wide association study of 520,000 subjects identifies 32 loci associated with stroke and stroke subtypes », *Nat Genet*, vol. 50, n° 4, p. 524-537, avr. 2018, doi: 10.1038/s41588-018-0058-3.
- [200] Z. Ning, Y. Pawitan, et X. Shen, « High-definition likelihood inference of genetic correlations across human complex traits », *Nat Genet*, vol. 52, n° 8, p. 859-864, août 2020, doi: 10.1038/s41588-020-0653-y.
- [201] R. R. Veeneman *et al.*, « Exploring the Relationship Between Schizophrenia and Cardiovascular Disease: A Genetic Correlation and Multivariable Mendelian

- Randomization Study », *Schizophr Bull*, vol. 48, n° 2, p. 463-473, mars 2022, doi: 10.1093/schbul/sbab132.
- [202] S. van Oort, J. W. J. Beulens, A. J. van Ballegooijen, D. E. Grobbee, et S. C. Larsson, « Association of Cardiovascular Risk Factors and Lifestyle Behaviors With Hypertension: A Mendelian Randomization Study », *Hypertension*, vol. 76, n° 6, p. 1971-1979, déc. 2020, doi: 10.1161/HYPERTENSIONAHA.120.15761.
- [203] P.-F. Plouin *et al.*, « High Prevalence of Multiple Arterial Bed Lesions in Patients With Fibromuscular Dysplasia: The ARCADIA Registry (Assessment of Renal and Cervical Artery Dysplasia) », *Hypertension*, vol. 70, n° 3, p. 652-658, sept. 2017, doi: 10.1161/HYPERTENSIONAHA.117.09539.
- [204] « Vascular Factors and Risk of Dementia: Design of the Three-City Study and Baseline Characteristics of the Study Population », *Neuroepidemiology*, vol. 22, n° 6, p. 316-325, 2003, doi: 10.1159/000072920.
- [205] P. Dobrowolski *et al.*, « Echocardiographic assessment of left ventricular morphology and function in patients with fibromuscular dysplasia: the ARCADIA-POL study », *Journal of Hypertension*, vol. 36, n° 6, p. 1318-1325, juin 2018, doi: 10.1097/HJH.0000000000001706.
- [206] W. Drygas *et al.*, « Multi-centre National Population Health Examination Survey (WOBASZ II study): assumptions, methods, and implementation », *Kardiol Pol*, vol. 74, n° 7, p. 681-690, 2016, doi: 10.5603/KP.a2015.0235.
- [207] M. Pappaccogli *et al.*, « The European/International Fibromuscular Dysplasia Registry and Initiative (FEIRI)—clinical phenotypes and their predictors based on a cohort of 1000 patients », *Cardiovascular Research*, vol. 117, n° 3, p. 950-959, mars 2021, doi: 10.1093/cvr/cvaa102.
- [208] E.-R. Rietzschel *et al.*, « Rationale, design, methods and baseline characteristics of the Asklepios Study », *Eur J Cardiovasc Prev Rehabil*, vol. 14, n° 2, p. 179-191, avr. 2007, doi: 10.1097/HJR.0b013e328012c380.
- [209] Z. Ye, F. S. Kalloo, A. K. Dalenberg, et I. J. Kullo, « An electronic medical record-linked biorepository to identify novel biomarkers for atherosclerotic cardiovascular disease », *Glob Cardiol Sci Pract*, vol. 2013, n° 1, p. 82-90, nov. 2013, doi: 10.5339/gcsp.2013.10.
- [210] J. W. Olin *et al.*, « A plasma proteogenomic signature for fibromuscular dysplasia », *Cardiovasc Res*, vol. 116, n° 1, p. 63-77, janv. 2020, doi: 10.1093/cvr/cvz219.
- [211] L. G. Fritsche *et al.*, « Association of Polygenic Risk Scores for Multiple Cancers in a Phenome-wide Study: Results from The Michigan Genomics Initiative », *Am J Hum Genet*, vol. 102, n° 6, p. 1048-1061, juin 2018, doi: 10.1016/j.ajhg.2018.04.001.
- [212] B. Vogel *et al.*, « The Lancet women and cardiovascular disease Commission: reducing the global burden by 2030 », *Lancet*, vol. 397, n° 10292, p. 2385-2438, juin 2021, doi: 10.1016/S0140-6736(21)00684-X.

- [213] J.-P. Collet *et al.*, « 2020 ESC Guidelines for the management of acute coronary syndromes in patients presenting without persistent ST-segment elevation », *Eur Heart J*, vol. 42, n° 14, p. 1289-1367, avr. 2021, doi: 10.1093/eurheartj/ehaa575.
- [214] M. Prasad *et al.*, « Prevalence of Extracoronary Vascular Abnormalities and Fibromuscular Dysplasia in Patients With Spontaneous Coronary Artery Dissection », *American Journal of Cardiology*, vol. 115, n° 12, p. 1672-1677, juin 2015, doi: 10.1016/j.amjcard.2015.03.011.
- [215] M. H. Wang, H. J. Cordell, et K. Van Steen, « Statistical methods for genome-wide association studies », *Semin Cancer Biol*, vol. 55, p. 53-60, avr. 2019, doi: 10.1016/j.semcancer.2018.04.008.
- [216] M. E. Weale, « Quality Control for Genome-Wide Association Studies », in *Genetic Variation: Methods and Protocols*, M. R. Barnes et G. Breen, Éd. Totowa, NJ: Humana Press, 2010, p. 341-372. doi: 10.1007/978-1-60327-367-1\_19.
- [217] C. A. Anderson, F. H. Pettersson, G. M. Clarke, L. R. Cardon, A. P. Morris, et K. T. Zondervan, « Data quality control in genetic case-control association studies », *Nat Protoc*, vol. 5, n° 9, p. 1564-1573, sept. 2010, doi: 10.1038/nprot.2010.116.
- [218] S. Purcell *et al.*, « PLINK: A Tool Set for Whole-Genome Association and Population-Based Linkage Analyses », *Am J Hum Genet*, vol. 81, n° 3, p. 559-575, sept. 2007.
- [219] M. Ringnér, « What is principal component analysis? », *Nat Biotechnol*, vol. 26, n° 3, Art. n° 3, mars 2008, doi: 10.1038/nbt0308-303.
- [220] A. L. Price, N. J. Patterson, R. M. Plenge, M. E. Weinblatt, N. A. Shadick, et D. Reich, « Principal components analysis corrects for stratification in genome-wide association studies », *Nat Genet*, vol. 38, n° 8, Art. n° 8, août 2006, doi: 10.1038/ng1847.
- [221] C. C. Chang, C. C. Chow, L. C. Tellier, S. Vattikuti, S. M. Purcell, et J. J. Lee, « Second-generation PLINK: rising to the challenge of larger and richer datasets », *GigaScience*, vol. 4, n° 1, p. s13742-015-0047-8, déc. 2015, doi: 10.1186/s13742-015-0047-8.
- [222] J. Marchini et B. Howie, « Genotype imputation for genome-wide association studies », *Nat Rev Genet*, vol. 11, n° 7, p. 499-511, juill. 2010, doi: 10.1038/nrg2796.
- [223] B. N. Howie, P. Donnelly, et J. Marchini, « A Flexible and Accurate Genotype Imputation Method for the Next Generation of Genome-Wide Association Studies », *PLOS Genetics*, vol. 5, n° 6, p. e1000529, juin 2009, doi: 10.1371/journal.pgen.1000529.
- [224] S. Das *et al.*, « Next-generation genotype imputation service and methods », *Nat Genet*, vol. 48, n° 10, p. 1284-1287, oct. 2016, doi: 10.1038/ng.3656.
- [225] S. McCarthy *et al.*, « A reference panel of 64,976 haplotypes for genotype imputation », *Nat Genet*, vol. 48, n° 10, p. 1279-1283, oct. 2016, doi: 10.1038/ng.3643.

- [226] K. Reue et C. B. Wiese, « Illuminating the Mechanisms Underlying Sex Differences in Cardiovascular Disease », *Circulation Research*, vol. 130, n° 12, p. 1747-1762, juin 2022, doi: 10.1161/CIRCRESAHA.122.320259.
- [227] M. Delbridge et J. Graves, « Origin and evolution of spermatogenesis genes on the human sex chromosomes », *Society of Reproduction and Fertility supplement*, vol. 65, p. 1-17, févr. 2007.
- [228] N. Takagi et K. Abe, « Detrimental effects of two active X chromosomes on early mouse development », *Development*, vol. 109, n° 1, p. 189-201, mai 1990, doi: 10.1242/dev.109.1.189.
- [229] P. S. Burgoyne, « Genetic homology and crossing over in the X and Y chromosomes of mammals », *Hum Genet*, vol. 61, n° 2, p. 85-90, sept. 1982, doi: 10.1007/BF00274192.
- [230] G. Pereira et S. Dória, « X-chromosome inactivation: implications in human disease », *J Genet*, vol. 100, n° 2, p. 63, oct. 2021, doi: 10.1007/s12041-021-01314-1.
- [231] C. M. Disteché, « Dosage compensation of the sex chromosomes and autosomes », *Semin Cell Dev Biol*, vol. 56, p. 9-18, août 2016, doi: 10.1016/j.semcdb.2016.04.013.
- [232] E. V. Dementyeva, A. I. Shevchenko, et S. M. Zakian, « X-chromosome upregulation and inactivation: two sides of the dosage compensation mechanism in mammals », *BioEssays*, vol. 31, n° 1, p. 21-28, 2009, doi: 10.1002/bies.080149.
- [233] X. Deng, J. B. Berletch, D. K. Nguyen, et C. M. Disteché, « X chromosome regulation: diverse patterns in development, tissues and disease », *Nat Rev Genet*, vol. 15, n° 6, p. 367-378, juin 2014, doi: 10.1038/nrg3687.
- [234] I. Talon *et al.*, « Enhanced chromatin accessibility contributes to X chromosome dosage compensation in mammals », *Genome Biol*, vol. 22, n° 1, p. 302, déc. 2021, doi: 10.1186/s13059-021-02518-5.
- [235] A. Flaquer, G. A. Rappold, T. F. Wienker, et C. Fischer, « The human pseudoautosomal regions: a review for genetic epidemiologists », *Eur J Hum Genet*, vol. 16, n° 7, Art. n° 7, juill. 2008, doi: 10.1038/ejhg.2008.63.
- [236] C. M. DISTECHE et J. B. BERLETCH, « X-chromosome inactivation and escape », *J Genet*, vol. 94, n° 4, p. 591-599, déc. 2015.
- [237] M. Monk, M. Boubelik, et S. Lehnert, « Temporal and regional changes in DNA methylation in the embryonic, extraembryonic and germ cell lineages during mouse embryo development », *Development*, vol. 99, n° 3, p. 371-382, mars 1987, doi: 10.1242/dev.99.3.371.
- [238] M. Taherian, H. Maghsoudi, K. Bidaki, et R. Taherian, « The Relationship between Skewed X-chromosome Inactivation and Neurological Disorders Development », vol. 3, n° 2, p. 12, 2016.

- [239] H. Fang, C. M. Distèche, et J. B. Berletch, « X Inactivation and Escape: Epigenetic and Structural Features », *Frontiers in Cell and Developmental Biology*, vol. 7, 2019, Consulté le: 9 juin 2022. [En ligne]. Disponible sur: <https://www.frontiersin.org/article/10.3389/fcell.2019.00219>
- [240] M. F. Lyon, « Gene Action in the X-chromosome of the Mouse (*Mus musculus* L.) », *Nature*, vol. 190, n° 4773, Art. n° 4773, avr. 1961, doi: 10.1038/190372a0.
- [241] M. L. Barr et E. G. Bertram, « A Morphological Distinction between Neurones of the Male and Female, and the Behaviour of the Nucleolar Satellite during Accelerated Nucleoprotein Synthesis », *Nature*, vol. 163, n° 4148, Art. n° 4148, avr. 1949, doi: 10.1038/163676a0.
- [242] M. F. Lyon, « Gene Action in the X-chromosome of the Mouse (*Mus musculus* L.) », *Nature*, vol. 190, n° 4773, Art. n° 4773, avr. 1961, doi: 10.1038/190372a0.
- [243] M. F. Lyon, « Sex Chromatin and Gene Action in the Mammalian X-Chromosome », *Am J Hum Genet*, vol. 14, n° 2, p. 135-148, juin 1962.
- [244] E. Heard et C. M. Distèche, « Dosage compensation in mammals: fine-tuning the expression of the X chromosome », *Genes Dev.*, vol. 20, n° 14, p. 1848-1867, juill. 2006, doi: 10.1101/gad.1422906.
- [245] J. Chaumeil, P. Le Baccon, A. Wutz, et E. Heard, « A novel role for Xist RNA in the formation of a repressive nuclear compartment into which genes are recruited when silenced », *Genes Dev.*, vol. 20, n° 16, p. 2223-2237, août 2006, doi: 10.1101/gad.380906.
- [246] P. Avner et E. Heard, « X-chromosome inactivation: counting, choice and initiation », *Nat Rev Genet*, vol. 2, n° 1, p. 59-67, janv. 2001, doi: 10.1038/35047580.
- [247] C. J. Brown *et al.*, « A gene from the region of the human X inactivation centre is expressed exclusively from the inactive X chromosome », *Nature*, vol. 349, n° 6304, Art. n° 6304, janv. 1991, doi: 10.1038/349038a0.
- [248] G. Borsani *et al.*, « Characterization of a murine gene expressed from the inactive X chromosome », *Nature*, vol. 351, n° 6324, p. 325-329, mai 1991, doi: 10.1038/351325a0.
- [249] K. Plath, S. Mlynarczyk-Evans, D. A. Nusinow, et B. Panning, « Xist RNA and the Mechanism of X Chromosome Inactivation », *Annu. Rev. Genet.*, vol. 36, n° 1, p. 233-278, déc. 2002, doi: 10.1146/annurev.genet.36.042902.092433.
- [250] W. H. Brooks, « Viral Impact in Autoimmune Diseases: Expanding the “X Chromosome–Nucleolus Nexus” Hypothesis », *Front Immunol*, vol. 8, p. 1657, nov. 2017, doi: 10.3389/fimmu.2017.01657.
- [251] D. J. Lerner et W. B. Kannel, « Patterns of coronary heart disease morbidity and mortality in the sexes: a 26-year follow-up of the Framingham population », *Am Heart J*, vol. 111, n° 2, p. 383-390, févr. 1986, doi: 10.1016/0002-8703(86)90155-9.

- [252] R. Voskuhl, « Sex differences in autoimmune diseases », *Biology of Sex Differences*, vol. 2, n° 1, p. 1, janv. 2011, doi: 10.1186/2042-6410-2-1.
- [253] F. Gao *et al.*, « XWAS: A Software Toolset for Genetic Data Analysis and Association Studies of the X Chromosome », *J Hered*, vol. 106, n° 5, p. 666-671, oct. 2015, doi: 10.1093/jhered/esv059.
- [254] « XWAS ». KeinanLab, 11 septembre 2020. Consulté le: 6 septembre 2022. [En ligne]. Disponible sur: <https://github.com/KeinanLab/xwas-3.0>
- [255] J. M. Korn *et al.*, « Integrated genotype calling and association analysis of SNPs, common copy number polymorphisms and rare CNVs », *Nat Genet*, vol. 40, n° 10, p. 1253-1260, oct. 2008, doi: 10.1038/ng.237.
- [256] L. T. Hogben et E. Ponder, « BIOLOGICAL MONOGRAPHS AND MANUALS », p. 336.
- [257] P. Danecek *et al.*, « The variant call format and VCFtools », *Bioinformatics*, vol. 27, n° 15, p. 2156-2158, août 2011, doi: 10.1093/bioinformatics/btr330.
- [258] J. Milet, D. Courtin, A. Garcia, et H. Perdry, « Mixed logistic regression in genome-wide association studies », *BMC Bioinformatics*, vol. 21, n° 1, p. 536, nov. 2020, doi: 10.1186/s12859-020-03862-2.
- [259] I. Pe'er, R. Yelensky, D. Altshuler, et M. J. Daly, « Estimation of the multiple testing burden for genomewide association studies of nearly all common variants », *Genetic Epidemiology*, vol. 32, n° 4, p. 381-385, 2008, doi: 10.1002/gepi.20303.
- [260] C. J. Willer, Y. Li, et G. R. Abecasis, « METAL: fast and efficient meta-analysis of genomewide association scans », *Bioinformatics*, vol. 26, n° 17, p. 2190-2191, sept. 2010, doi: 10.1093/bioinformatics/btq340.
- [261] J. P. T. Higgins, S. G. Thompson, J. J. Deeks, et D. G. Altman, « Measuring inconsistency in meta-analyses », *BMJ*, vol. 327, n° 7414, p. 557-560, sept. 2003.
- [262] N. Combaret *et al.*, « National French registry of spontaneous coronary artery dissections: prevalence of fibromuscular dysplasia and genetic analyses », *EuroIntervention*, vol. 17, n° 6, p. 508-515, août 2021, doi: 10.4244/EIJ-D-20-01046.
- [263] A. Persu *et al.*, « Prevalence and Disease Spectrum of Extracoronary Arterial Abnormalities in Spontaneous Coronary Artery Dissection », *JAMA Cardiol*, vol. 7, n° 2, p. 159-166, févr. 2022, doi: 10.1001/jamacardio.2021.4690.
- [264] M. Prasad *et al.*, « Prevalence of Extracoronary Vascular Abnormalities and Fibromuscular Dysplasia in Patients With Spontaneous Coronary Artery Dissection », *The American Journal of Cardiology*, vol. 115, n° 12, p. 1672-1677, juin 2015, doi: 10.1016/j.amjcard.2015.03.011.

- [265] S. A. Peters et M. Woodward, « Women's reproductive factors and incident cardiovascular disease in the UK Biobank », *Heart*, vol. 104, n° 13, p. 1069-1075, juill. 2018, doi: 10.1136/heartjnl-2017-312289.
- [266] C. M. Otto, « Heartbeat: cardiovascular disease risk and reproductive factors in women », *Heart*, vol. 104, n° 13, p. 1045-1047, juill. 2018, doi: 10.1136/heartjnl-2018-313503.
- [267] B. Bulik-Sullivan *et al.*, « An atlas of genetic correlations across human diseases and traits », *Nat Genet*, vol. 47, n° 11, Art. n° 11, nov. 2015, doi: 10.1038/ng.3406.
- [268] A. E. Murphy, B. M. Schilder, et N. G. Skene, « MungeSumstats: a Bioconductor package for the standardization and quality control of many GWAS summary statistics », *Bioinformatics*, vol. 37, n° 23, p. 4593-4596, déc. 2021, doi: 10.1093/bioinformatics/btab665.
- [269] S. Burgess et S. G. Thompson, « Interpreting findings from Mendelian randomization using the MR-Egger method », *Eur J Epidemiol*, vol. 32, n° 5, p. 377-389, mai 2017, doi: 10.1007/s10654-017-0255-x.
- [270] J. Bowden, F. Del Greco M, C. Minelli, G. Davey Smith, N. A. Sheehan, et J. R. Thompson, « Assessing the suitability of summary data for two-sample Mendelian randomization analyses using MR-Egger regression: the role of the I2 statistic », *Int J Epidemiol*, vol. 45, n° 6, p. 1961-1974, déc. 2016, doi: 10.1093/ije/dyw220.
- [271] J. Bowden, G. Davey Smith, et S. Burgess, « Mendelian randomization with invalid instruments: effect estimation and bias detection through Egger regression », *International Journal of Epidemiology*, vol. 44, n° 2, p. 512-525, avr. 2015, doi: 10.1093/ije/dyv080.
- [272] G. Hemani *et al.*, « The MR-Base platform supports systematic causal inference across the human phenome », *eLife*, vol. 7, p. e34408, doi: 10.7554/eLife.34408.
- [273] D. Adlam *et al.*, « Genome-wide association meta-analysis of spontaneous coronary artery dissection reveals common variants and genes related to artery integrity and tissue-mediated coagulation ». medRxiv, p. 2022.07.05.22277238, 6 juillet 2022. doi: 10.1101/2022.07.05.22277238.