Identification de nouveaux variants rares associés à la spondyloarthrite par séquençage haut-débit

Hendrick Mambu Mambueni

▶ **To cite this version:**

Hendrick Mambu Mambueni. Identification de nouveaux variants rares associés à la spondyloarthrite par séquençage haut-débit. Génétique. Université Paris-Saclay, 2022. Français. NNT : 2022UPASL064 . tel-03941362

HAL Id: tel-03941362
https://theses.hal.science/tel-03941362

Submitted on 16 Jan 2023

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.

THESE DE DOCTORAT

NNT : 2022UPASL064

Identification de nouveaux variants rares associés à la spondyloarthrite par séquençage haut-débit

Identification of new rare variants associated with spondyloarthritis by high-throughput sequencing

Thèse de doctorat de l'université Paris-Saclay

École doctorale n° 577, Structure et dynamique des systèmes vivants (SDSV)
Spécialité de doctorat : Sciences de la vie et de la santé
Graduate School: Life Sciences and Health.
Référent : Université de Versailles-Saint-Quentin-en-Yvelines

Thèse préparée dans l'unité de recherche 2I (Université Paris-Saclay, UVSQ, Inserm), sous la direction de **Henri-Jean GARCHON**, PU-PH et le co-encadrement de **Félicie COSTANTINO**, PU-PH

Thèse soutenue à Paris-Saclay, le 9 décembre 2022, par

Hendrick MAMBU MAMBUENI

Composition du Jury

Membres du jury avec voix délibérative

Lucienne CHATENOUD PU, Université Paris Cité	Présidente
Nathalie BALANDRAUD PH, Université d'Aix Marseille	Rapporteur & Examinatrice
Frédéric RIEUX-LAUCAT DR, Université Paris Cité	Rapporteur & Examinateur
Elisabeth PETIT-TEIXEIRA PU, UEVE	Examinatrice

université PARIS-SACLAY
ÉCOLE DOCTORALE
Structure et dynamique des systèmes vivants (SDSV)

Titre : Identification de nouveaux variants rares associés à la spondyloarthrite par séquençage haut-débit
Mots clés : spondyloarthrite, étude familiale, variants, séquençage haut-débit, génétique, maladie complexe

Résumé : La spondyloarthrite (SpA) est une maladie multifactorielle avec une héritabilité estimée à plus de 90%, principalement en lien avec le HLA-B27. L'ensemble des facteurs de susceptibilité identifiés, incluant HLA-B27, expliquent moins du tiers de l'héritabilité. L'implication de variants rares pourrait expliquer une partie de cette héritabilité manquante. L'objectif de ce travail était d'identifier des variants rares associés à la SpA via une approche combinant analyses familiales et séquençage haut-débit. D'abord, nous avons séquencé une région de 1,4 Mb significativement liée à la SpA en 13q13 chez 71 patients et 21 témoins sains appartenant à des familles ayant un score de liaison élevée dans cette région. Nous avons identifié un variant rare dans le gène FREM2 présent chez 9 malades d'une famille fortement liée à la région et non retrouvé dans d'autres familles ou chez les cas isolés de SpA. Nous avons ensuite séquencé l'exome de 48 malades venant de 20 familles multiplex. Malheureusement, nous n'avons pas observé de variant récurrent entre les familles. Puis, nous nous sommes concentrés sur un pic de liaison génétique, déjà connu, sur le chromosome 9. L'étude de la famille la plus liée à cette région, qui comprend 12 patients, a conduit à l'identification de plusieurs variants rares codants ségrégeant avec la maladie. Cependant, les études ultérieures ont montré des fréquences alléliques de ces variants équivalentes entre les cas et les témoins. Enfin, le séquençage du génome entier de 413 patients issus de 76 familles multiplex avec 4 malades ou plus a été réalisé. Nous avons identifié 1203 variants rares, codants et non synonymes et partagés par au moins tous les membres atteints d'une famille. Les analyses de validation génétique et fonctionnelle de ces variants sont en cours, tout comme l'analyse des variants non codants. En conclusion, ces différentes approches suggèrent une importante hétérogénéité génétique de la SpA et soulignent également la difficulté de confirmer l'implication de variants rares dans les maladies complexes.

Title: Identification of new rare variants associated with spondyloarthritis by high-throughput sequencing
Keywords: spondyloarthritis, family study, variants, next generation sequencing, genetic, complex disease

Abstract: Spondyloarthritis (SpA) is a multifactorial disease with an estimated heritability of over 90%, mainly related to HLA-B27. All identified susceptibility factors, including HLA-B27, explain less than one third of the heritability. The involvement of rare variants could explain part of this missing heritability. The aim of this work was to identify rare variants associated with SpA via a combined family analysis and high-throughput sequencing approach. First, we sequenced a 1.4 Mb region significantly linked to SpA at 13q13 in 71 patients and 21 healthy controls from families with a high linkage score in this region. We identified a rare variant in the FREM2 gene present in 9 patients from a family with high linkage to the region and not found in other families or isolated cases of SpA. We then sequenced the exome of 48 patients from 20 multiplex families. Unfortunately, we did not observe any recurrent variants between families. We then focused on a second, previously known genetic linkage peak on chromosome 9. The study of the family most linked to this region, which includes 12 patients, led to the identification of several rare coding variants segregating with the disease. However, subsequent studies have shown equivalent allelic frequencies of these variants between cases and controls. Finally, whole genome sequencing of 413 patients from 76 multiplex families with 4 or more patients was performed. We identified 1203 rare, coding, non-synonymous variants shared by at least all affected family members. Genetic and functional validation analyses of these variants are underway, as is the analysis of non-coding variants. In conclusion, these different approaches suggest significant genetic heterogeneity in SpA and also highlight the difficulty of confirming the involvement of rare variants in complex diseases.

REMERCIEMENTS

Tout d'abord, je remercie tous les membres du jury de ma thèse, Dr Frédéric Rieux-Laucat, Pr Lucienne Chatenoud, Dr Nathalie Balandraud et Pr Elisabeth Petit-Teixeira, pour leur lecture de mon manuscrit de thèse et les remarques qu'ils m'adresseront pendant la soutenance de ma thèse. Je les remercie d'avoir accepté d'évaluer ce travail de thèse.

Je remercie les membres de mon comité de suivi de thèse, Monsieur Hervé Perdry et Madame Valérie Chaudru, qui m'ont suivi ces trois dernières années. Vos idées et conseils ont été très précieux pour ma thèse.

Je remercie chaleureusement mon directeur de thèse, Pr Henri-Jean Garchon et ma co-encadrante de thèse, Pr Félicie Costantino. D'abord pour leur accueil pour mon stage de master 2 et ensuite pour ma thèse. Votre confiance m'a permis de remplir mes missions. Vous n'avez pas compté votre temps pour répondre à mes interrogations ainsi qu'à mes nombreux doutes.

Je remercie le directeur de l'équipe, Pr Maxime BREBAN, pour ses conseils. Ils furent d'une aide précieuse tout au long de ma thèse.

Un grand merci à tous les membres de l'équipe IRIS d'hier et d'aujourd'hui, j'ai été ravi de vous connaître et de travailler avec vous.
Luiza et Aude, à vrai dire, je ne sais pas vous dire la joie que je ressentais à vos côtés tout le jour au labo. En une phrase, vous avez rendu mon travail agréable ! Je sais que la suite sera difficile sans vous. Alors, n'oubliez pas de répondre à mes appels lorsque j'aurai besoin de parler. Un grand merci à vous.
Christophe Hue, « Mon président », je sais ce que je te dois. Du début de mon stage de master 2 à ma thèse, en passant par la plateforme, je te suis très reconnaissant.
Simon « l'Américain », dès ton arrivé au labo j'ai su qu'on allait bien s'entendre. You are amazing. Je me souviendrai longtemps de nos soirées ligue des champions autour d'un poulet rôti. Continue de t'entraîner, peut-être que tu me battras à la course un jour.
Ketia et Nadège, merci pour nos échanges. Vous ne m'avez pas oublié même après votre départ. Je suis vos pas comme promis.

Amel « Meli Melo », j'aurai voulu rester un peu plus longtemps au labo pour nos moments fous rires. J'attends toujours mon couscous avec du lait caillé que tu m'avais promis. Merci pour tout mon amie.
Marie, Manon, Bilade, Marc, Thomas, Alaa et Gerlando, je suis navré de ne pas faire une phrase pour chacun de vous, mais je suis sûr que vous comprendrez. J'ai aimé travailler avec vous.
Je n'oublie pas ma stagiaire Clara, j'ai été ravi de te montrer un peu la génomique pendant les deux mois de ton stage. Bon courage pour la suite.

Je remercie également les collègues des autres équipes de l'U1173 d'hier et d'aujourd'hui. Le directeur de l'Unité, Pr Jean-Louis Herrmann, merci pout ton aide capital pour la suite de ma carrière. Je n'oublierai jamais.
Gina, Marion et Pauline, je suis le prochain à partir comme prévu. Je crois que ça va être difficile pour que nos copains de la team BDT nous remplacent. Hamadoun, Laila et Elodie, n'essayer pas de nous remplacer, car votre tour arrive. Mes compagnons de thèse, ça été un plaisir de faire la thèse en même temps que vous.

Je remercie mes amis de longues dates qui sont devenus ma famille, Néhémie, Gaëtan, Dyna, Emmanuel, Thierry, Vailton, Sarah et Aïssata. Merci pour vos conseils. Désolé, pour celles et ceux que j'ai oublié.

A ma famille, Loïc, Aurélien, Adeline, Brenda, Mirnan, Rivaldo, Sophia, Kylian, Kévin, Dylan, Darchi, Messieurs Robert, Gaspard et Alfred Mambu. Merci pour votre soutien indéfectible.

A toi, Laure Madline, merci d'avoir supporté mon travail de nuit, même les week-ends. Merci pour tes conseils, même lorsque tu ne comprenais pas « la génomique ».

A Maman et Papa, je suis arrivé à l'endroit où je vous avais promis. Vous m'avez donné le stylo quand j'étais petit et je l'ai gardé jusqu'au bout. Cela a été long et un peu difficile, mais j'ai tenu bon grâce à votre soutien. Un merci n'est peut-être pas suffisant, mais J'espère que je vous ai rendu fiers.

A mes grands-parents,

TABLE DES MATIERES

REMERCIEMENTS ..... - 1 -
TABLE DES MATIERES ..... - 4 -
ABREVIATIONS ..... - 8 -
INDEX DES FIGURES ..... - 11 -
INDEX DES TABLEAUX ..... - 16 -
INTRODUCTION GENERALE ..... - 17 -
I. La spondyloarthrite : aspects épidémiologiques, cliniques et thérapeutiques. ..... - 17 -
I.1. Epidémiologie ..... - 17 -
I.1.1. Prévalence ..... - 17 -
I.1.2. Population touchée ..... - 18 -
I.2. Diagnostic ..... - 18 -
I.2.1. Manifestations cliniques ..... - 18 -
I.2.2. Examens complémentaires ..... - 21 -
I.3. Les différentes formes de la SpA ..... - 23 -
I.3.1. La spondylarthrite ankylosante (SA) ..... - 23 -
I.3.2. Le rhumatisme psoriasique ..... - 23 -
I.3.3. Les arthrites réactionnelles ..... - 23 -
I.3.4. La SpA associée aux MICI ..... - 23 -
I.3.5. Les spondyloarthrites indifférenciées ..... - 24 -
I.4. Concept unifié de SpA ..... - 24 -
I.5. Critères de classification de la SpA ..... - 25 -
I.5.1. Critères d'Amor ..... - 25 -
I.5.2. Critères de l'ESSG ..... - 26 -
I.5.3. Critères ASAS (Assessment of SpondyloArthritis international Society). ..... - 26 -
I.5.4. Critères de New York modifiés ..... - 27 -
I.6. Traitement ..... - 28 -
I.6.1. Les traitements médicamenteux ..... - 28 -
I.6.2. Les traitements non pharmacologiques ..... - 29 -
I.6.3. La chirurgie ..... - 30 -
II. La génétique des maladies complexes ..... - 31 -
II.1. Les maladies complexes ..... - 31 -

II.1.1. Maladies complexes vs maladies mendéliennes ..... - 31 -
II.1.2. Estimation de la contribution des facteurs génétiques à la survenue d'une maladie complexe ..... - 31 -
II.2. Variabilité génétique humaine ..... - 33 -
II.2.1. Types de variants ..... - 33 -
II.2.2. Détection des variants ..... - 35 -
II.2.3. Choix de la technique ..... - 42 -
II.3. Identification des facteurs génétiques de susceptibilité des maladies complexes ..... - 43 -
II.3.1. Les études de liaison génétique ..... - 43 -
II.3.2. Les analyses d'association génétique ..... - 46 -
II.3.3. Concept d'héritabilité manquante ..... - 52 -
II.4. Les variants rares ..... - 54 -
II.4.1. Implication des variants rares dans les maladies complexes ..... - 55 -
II.4.2. Design d'études pour les variants rares ..... - 55 -
II.4.3. Priorisation des variants issus du NGS ..... - 58 -
II.4.4. Méthodes d'analyses spécifiques de gènes ..... - 61 -
II.4.5. Validation génétique et fonctionnelle des variants ..... - 64 -
III. La génétique de la spondyloarthrite ..... - 66 -
III.1. La spondyloarthrite : une maladie complexe ..... - 66 -
III.1.1. Agrégation familiale ..... - 66 -
III.1.2. Études de jumeaux ..... - 67 -
III.1.3. Études de ségrégation ..... - 68 -
III.1.4. Facteurs environnementaux ..... - 68 -
III.2. Facteurs de susceptibilités génétiques identifiées dans la spondyloarthrite ..... - 68 -
III.2.1. Complexe majeur d'histocompatibilité ..... - 68 -
III.2.2. Les gènes en dehors du complexe majeur d'histocompatibilité ..... - 70 -
III.3. Que nous ont appris les GWAS dans la SpA ? ..... - 75 -
III.4. Où trouver l'héritabilité manquante dans la SpA ? ..... - 75 -
III.4.1. Les interactions gène – gène et gène – environnement ..... - 75 -
III.4.2. Variants structuraux ..... - 76 -
III.4.3. Les variants rares ..... - 76 -
RESULTATS ..... - 79 -
I. Méthodologie commune à toutes les études ..... - 79 -
I.1. Patients ..... - 79 -

I.2. Séquençage et génotypage ..... - 80 -
I.2.1. Séquençage haut-débit ..... - 80 -
I.2.2. Séquençage Sanger ..... - 80 -
I.2.3. Génotypage par sondes TaqMan ..... - 80 -
I.3. Traitement des données de séquençage haut débit ..... - 80 -
I.4. Analyses génétiques ..... - 81 -
I.4.1. Vérification de degré de parenté entre les individus ..... - 81 -
I.4.2. Analyse de ségrégation ..... - 81 -
I.4.3. Analyse d'association ..... - 81 -
II. Partie I : Le reséquençage du locus 13q13 ..... - 82 -
II.1. Introduction ..... - 82 -
II.2. Patients et méthodes ..... - 84 -
II.2.1. Echantillons biologiques ..... - 84 -
II.2.2. Reséquençage du locus 13q13 ..... - 86 -
II.3. Filtrage des variants ..... - 87 -
II.4. Analyses de ségrégation ..... - 90 -
II.5. Analyses d'association cas – témoins ..... - 92 -
II.6. Immunomarquage de FREM2 ..... - 93 -
II.7. Discussion ..... - 96 -
III. Partie II : Le séquençage d'exome ..... - 98 -
III.1. Introduction ..... - 98 -
III.2. Patients et méthodes ..... - 98 -
III.2.1. Echantillons biologiques ..... - 98 -
III.2.2. Séquençage d'exome ..... - 100 -
III.3. Filtrage des variants ..... - 101 -
III.4. Etude des variants d'intérêt ..... - 101 -
III.4.1. Famille 33 et le variant rs144093498 (STIM2*S2195F) ..... - 102 -
III.4.2. Famille 110 et le variant rs200885264 (ENDOG*K66N) ..... - 103 -
III.5. Reconstruction des haplotypes des patients de la famille 110 ..... - 105 -
III.6. Filtrage des variants du chromosome 9 ..... - 107 -
III.7. Analyses de ségrégation des variants proches du variant ENDOG*K66N ..... - 108 -
III.8. Analyses cas – témoins des variants proches du variant ENDOG*K66N ..... - 109 -
III.9. Discussion ..... - 110 -
IV. Partie III : Le séquençage génome entier ..... - 111 -

IV.1. Introduction ..... - 111 -
IV.2. Patients et méthodes ..... - 111 -
IV.2.1. Echantillons biologiques ..... - 111 -
IV.2.2. Séquençage génome entier ..... - 112 -
IV.3. Contrôle qualité de données ..... - 113 -
IV.4. Filtrage des variants ..... - 114 -
IV.4. Caractéristiques des 6083 variants filtrés ..... - 115 -
IV.4.1. Ségrégation ..... - 115 -
IV.4.2. Evaluation du potentiel délétère ..... - 115 -
IV.6. Annotation fonctionnelle ..... - 116 -
DISCUSSION ET PERPECTIVES ..... - 120 -
I. Le reséquençage du locus 13q13 ..... - 120 -
II. Le séquençage d'exome ..... - 123 -
III. Le séquençage du génome entier ..... - 125 -
IV. Le rôle des variants rares dans d'autres maladies auto-immunes ..... - 126 -
CONCLUSION ..... - 128 -
BIBLIOGRAPHIE 130

ABREVIATIONS

ADN : Acide désoxyribonucléique

ADNdb : ADN double brin

AFR : Africain

AINS : Anti-inflammatoires non stéroïdiens

AMR: Américain

ANNOVAR: ANNOtate VARiation

ARN: Acide ribonucléique

ASAS: Assessment of SpondyloArthritis international Society

BASDAI : indice d'activité de la SpA

BASFI : indice de sévérité fonctionnelle de la SpA

BASRI : indice de sévérité radiologique de la SpA

CADD: Combined Annotation Dependent Depletion

CDCV: Common Disease - Common Variants

CDRV: Common Disease - Rare Variants

cM: centimorgan

CMH : Complexe Majeur d'Histocompatibilité

CNRGH : Centre National de la Recherche en Génétique Humaine

CNV: Copy Number Variation

ddNTPs: didésoxyribonucléotides

DICE : Database of Immune Cell Expression, Expression quantitative trait loci and Epigenomics

DL : déséquilibre de liaison

DMARD: Disease Modifying Anti-Rheumatic Drug

dNTPs : désoxyribonucléotides

DZ : jumeaux dizygotes

EAS : Asiatique de l'Est

EC : test de combinaison exponentielle

ESSG : European Spondylarthropathy Study Group

EUR : Européen

FREM2: FRAS1-Related Extracellular Matrix 2

FrEx: French exome

GFEGS : Groupe Français d'Etude Génétique des Spondylarthropathies

gnomAD: Genome Aggregation Database

GO: Gene Ontology

GTEx : Genotype-Tissue Expression

GWAS : études d'association génome entier

H&E : Hématoxyline et éosine

HLA : Human Leukocyte Antigen

HRM : Courbe de fusion à haute résolution

IBD: Identiques par descendance

IGAS: International Genetics of Ankylosing Spondylitis Consortium

Indel : Insertion - délétion

IRM : Imagerie par résonance magnétique

LED : lupus érythémateux disséminé

LOD score : score de logarithme des chances

MAF : fréquence de l'allèle mineur

Mb : mégabases

miARN : microARN

MICI : Maladie Inflammatoire Chronique de l'Intestin

MZ : jumeaux monozygotes

NEG : négatif

NFE : Européen non finlandais

NGS : Nouvelle génération de séquençage

NOD : Non Obese Diabetic

PCR : réaction en chaîne par polymérase

POS : positif

PR : polyarthrite rhumatoïde

PROVEAN : Protein Variation Effect Analyzer

RFLP : Polymorphisme de longueur des fragments de restriction

SA : spondylarthrite ankylosante

SAS : Asiatiques du Sud

SEP: Sclérose en plaques

SIFT: Sorting tolerant from intolerant

SpA: spondyloarthrite

SMRT: Single Molecule Real Time

SNP: Single nucleotide polymorphism

SNV: Single nucleotide variation

SV : Structural variants

λs : indice de récurrence entre frère ou sœur

TADs : domaines topologiquement associés

TASC : Australo-Anglo-Americain Spondyloarthritis Consortium

TDT : Test de déséquilibre de transmission

UKBB: United Kingdom Biobank

WGS: séquençage génome entier

WTCCC2: Wellcome Trust Control Consortium 2

INDEX DES FIGURES

FIGURE 1. LA PREVALENCE MONDIALE DE LA SPA (ADAPTE DE STOLWIJK C ET AL. ARTHRITIS CARE RES 2016) [5]. ..... - 18 -
FIGURE 2. REPRESENTATIONS DES ATTEINTES ARTICULAIRES AXIALES DANS LA SPA. (D'APRES "SPONDYLARTHRITE EN 100 QUESTIONS" DOUGADOS M, EDITION NHA COMMUNICATION, 2005). A – ARTICULATIONS SACRO-ILIAQUES ; B – RACHIS LOMBAIRE ; C – PAROI THORACIQUE ANTERIEURE ..... - 19 -
FIGURE 3. REPRESENTATIONS DES ATTEINTES DES ENTHESES DANS LA SPA. ("SPONDYLARTHRITE EN 100 QUESTIONS" DOUGADOS M, EDITION NHA COMMUNICATION, 2005 ET MAKSYMOWYCH ET AL. [12]). A – ENTHESE TROCHANTERIENNE ; B – SITES ENTHESITIQUES LES PLUS FREQUEMMENT TOUCHES. ..... - 20 -
FIGURE 4. LES DIFFERENTES FORMES DE LA SPA [37] ..... - 25 -
FIGURE 5. EXEMPLE DES MALADIES MENDELIENNES ET MALADIES COMPLEXES. ..... - 31 -
FIGURE 6. EVALUATION DU POIDS GENETIQUE DANS LES MALADIES COMPLEXES [50]. ..... - 33 -
FIGURE 7. LES TYPES DES VARIANTS GENETIQUES. ..... - 35 -
FIGURE 8. PRINCIPE DE LA RFLP (ADAPTE DE GNIS-PEDAGOGIE.ORG). ..... - 36 -
FIGURE 9. PRINCIPE DE LA TECHNIQUE TAQMAN. ..... - 37 -
FIGURE 10. PRINCIPE DE L'ANALYSE DE LA FUSION DE L'ADN A HAUTE RESOLUTION (HRM) [56]. ..... - 38 -
FIGURE 11. PRINCIPE DES MICROARRAYS (ADAPTE DE FRANCE GENOMIQUE 2022). ..... - 39 -
FIGURE 12. PRINCIPE DU SEQUENÇAGE SANGER. ..... - 40 -
FIGURE 13. PRINCIPE DU SEQUENÇAGE DE DEUXIEME GENERATION (ADAPTE DE PLANET-VIE, 2022). ..... - 41 -
FIGURE 14. PRINCIPE DES ANALYSES DE LIAISON GENETIQUE (ADAPTE DE PLANET-VIE, 2022). DEUX PARENTS A (SAINT EN VERT) ET B (MALADE EN ROUGE), ET 5 LOCUS BIALLELIQUES (V, W, X, Y ET Z). CHEZ LES ENFANTS MALADES (C ET D), ON OBSERVE UNE RECOMBINAISON ENTRE LES LOCUS W ET X, ET ENTRE LES LOCUS X ET Y. AU NIVEAU DU LOCUS X, LES ALLELES 4 ET 3 SONT HERITES ENSEMBLE CHEZ LES DEUX ENFANTS MALADES. ..... - 44 -
FIGURE 15. LIAISON GENETIQUE ET FRACTION DE RECOMBINAISON. POUR QU'IL Y AIT LIAISON, IL FAUT QUE LA FRACTION DE RECOMBINAISON SOIT INFERIEURE A 50 %. ..... - 45 -
FIGURE 16. PRINCIPE DES ETUDES D'ASSOCIATION GENETIQUE (ADAPTE DE PLANET-VIE, 2022). ..... - 47 -
FIGURE 17. CALCUL DES MESURES DL, |D'| ET DU COEFFICIENT R^2. ..... - 48 -
FIGURE 18. PRINCIPE DU TDT. ..... - 50 -
FIGURE 19. CONTRIBUTION DE VARIANTS A LA SUSCEPTIBILITE DES MALADIES. ..... - 55 -
FIGURE 20. RELATION ENTRE LA FREQUENCE D'UN VARIANT ET LA TAILLE DE SON EFFET SUR LE PHENOTYPE. ... - 57 -
FIGURE 21. LA PART DE L'HERITABILITE EXPLIQUEE ET INEXPLIQUEE. L'HERITABILITE EXPLIQUEE PAR LE HLA-B27 (ROUGE) ET PAR LES AUTRES LOCUS IDENTIFIES (EN VERT) 7,38 % REPRESENTE 20,44 % ET 7,38 % RESPECTIVEMENT L'HERITABILITE INEXPLIQUEE REPRESENTE UN PEU MOINS DE 75 % (BLEU). ..... - 75 -
FIGURE 22. RESULTATS DE L'ANALYSE DE LIAISON NON PARAMETRIQUE A L'ECHELLE DU GENOME. L'AXE DES X INDIQUE LES POSITIONS DES MARQUEURS SUR LE GENOME, CHAQUE CHROMOSOME ETANT REPRESENTE

PAR UNE COULEUR DISTINCTE. L'AXE DES Y MONTRE LE LOD SCORE. LES SEUILS DE LOD SCORE SIGNIFICATIF ET SUGGESTIF SONT REPRESENTES PAR DES LIGNES PLEINES ET POINTILLEES, RESPECTIVEMENT. ..... - 82 -
FIGURE 23. RESULTATS DE L'ANALYSE DE LIAISON NON PARAMETRIQUE DU CHROMOSOME 13, PRENANT EN COMPTE (LIGNE POINTILLEE) OU NON (LIGNE PLEINE) LE DL. L'AXE DES X REPRESENTE LA DISTANCE PHYSIQUE A PARTIR DU P-TELOMERE ET L'AXE DES Y REPRESENTE LE LOD SCORE. ..... - 83 -
FIGURE 24. INTERVALLE DE LIAISON (ENCADRE EN ROUGE) DELIMITE PAR DES EVENEMENTS DE RECOMBINAISONS INFORMATIVES CHEZ LES PATIENTS. POUR CHAQUE RECOMBINAISON INFORMATIVE, LA BARRE GRISE REPRESENTE LA PORTION DE L'HAPLOTYPE RECOMBINANT QUI SEGREGE AVEC LA MALADIE CHEZ LES AUTRES MEMBRES DE LA FAMILLE. ..... - 83 -
FIGURE 25. SELECTION D'INDIVIDUS POUR LE RESEQUENÇAGE CIBLE DE LA REGION 13Q13. LES BARRES COLOREES REPRESENTENT LES HAPLOTYPES DEDUITS DANS LA REGION 13Q13. LES FLECHES ROUGES INDIQUENT LES INDIVIDUS DONT L'ADN A ETE SEQUENCE. LE PANEL A MONTRE UNE FAMILLE DANS LAQUELLE UN HAPLOTYPE (LE BLEU) COSEGREGE AVEC LA MALADIE : DANS CETTE SITUATION, NOUS AVONS CHOISI POUR LE SEQUENÇAGE DEUX MEMBRES AFFECTES PORTANT L'HAPLOTYPE COSEGREGEANT AVEC LA SPA ET UN MEMBRE NON AFFECTE AVEC DES HAPLOTYPES DIFFERENTS. LE PANEL B MONTRE UNE FAMILLE DANS LAQUELLE TOUS LES MEMBRES AFFECTES SONT HAPLO-IDENTIQUES. DANS CETTE SITUATION, NOUS AVONS SELECTIONNE UN SEUL MEMBRE AFFECTE POUR LE SEQUENÇAGE. ..... - 84 -
FIGURE 26. SCHEMA DE FILTRAGE (A) DES VARIANTS CODANTS ET (B) DES VARIANTS NON CODANTS IDENTIFIES DANS LA REGION CIBLE (GRCH38P12, CHR13 :38125001-39527000)..... - 88 -
FIGURE 27. PEDIGREE DE LA FAMILLE #75 ET SEGREGATION DU VARIANT R727H. TOUS LES MEMBRES SONT IDENTIFIES PAR UN NUMERO D'IDENTIFICATION. LES CARRES ET LES CERCLES CORRESPONDENT AUX HOMMES ET AUX FEMMES, RESPECTIVEMENT. LES SYMBOLES PLEINS ET VIDES CORRESPONDENT AUX PATIENTS ET AUX SUJETS SAINS, RESPECTIVEMENT. LES SYMBOLES GRIS CORRESPONDENT AUX INDIVIDUS POUR LESQUELS L'ADN N'ETAIT PAS DISPONIBLE. LE STATUT HLA-B27 ("POS" OU "NEG") EST EGALEMENT INDIQUE ET LA PRESENCE DU VARIANT R727H EST SPECIFIEE PAR * ET ** POUR LES HETEROZYGOTES ET LES HOMOZYGOTES, RESPECTIVEMENT. LES DEUX BARRES PARALLELES ENTRE 6682 ET 1335 INDIQUENT LA CONSANGUINITE. ..... - 91 -
FIGURE 28. EXPRESSION DE FREM2 DANS LES BIOPSIES DE RAT. COLORATION H&E (A, B, E, F, I, J, M, N) ET IMMUNOFLUORESCENCE AVEC DES ANTICORPS FREM2 ANTIHUMAINS DE LAPIN ET DES ANTICORPS IGG ANTI-LAPIN COUPLES A L'ALEXA FLUOR594 (C, G, K, O), OU ANTICORPS SECONDAIRES SEULS (D, H, L, P) DE REIN (A - D), DE COLON (E - H), DE LA MEMBRANE SYNOVIAL (I - L) ET DE CARTILAGE (M - P) DE RAT. LES FIGURES B, F, J ET N CORRESPONDENT A UN AGRANDISSEMENT PLUS IMPORTANT DE LA ZONE QUADRILLEE DANS LES FIGURES A, E, I, ET M, RESPECTIVEMENT. LES NOYAUX SONT COLORES AVEC DU DAPI DANS LES FIGURES C, D, G, H, K, L, O ET P. LES SECTIONS ONT ETE MONTEES DANS DU PROLONGTM DIAMOND ET EXAMINEES AU MOYEN DE LA MICROSCOPIE CONFOCALE. BARRES D'ECHELLE : 50 M POUR A ; 20 M POUR B, C, D, J, K, L, N, O, P ; 100 M POUR E, I, M ; 10 M POUR F, G, H..... - 95 -

FIGURE 29. EXPRESSION DE FREM2 DANS LES BIOPSIES HUMAINES. COLORATION H&E (A, B, E, F, I, J), IMMUNOFLUORESCENCE (C, D, G, H) ET IMMUNOHISTOCHIMIE (K, L). POUR LE COLON (C, D), DOUBLE MARQUAGE AVEC UN ANTICORPS MONOCLONAL CHROMOGRANINE A ANTI-SOURIS SUIVI D'UN ANTICORPS IGG1 ANTI-SOURIS COUPLE A UN ALEXA FLUOR 594 ET AVEC UN ANTICORPS FREM2 ANTI-LAPIN HUMAIN SUIVI D'UN ANTICORPS IGG ANTI-LAPIN COUPLE A UN ALEXA FLUOR 488. POUR LA SYNOVIE (G, H), MARQUAGE AVEC UN ANTICORPS FREM2 ANTIHUMAIN DE LAPIN SUIVI D'UN ANTICORPS IGG ANTI-LAPIN COUPLE A L'ALEXA FLUOR 594. POUR LE CARTILAGE (K, I), MARQUAGE AVEC L'ANTICORPS PRIMAIRE DE LAPIN ANTIHUMAIN FREM2 SUIVI D'UN ANTICORPS DE CHEVRE ANTI-LAPIN IGG COUPLE A LA PEROXYDASE. LES FLECHES DESIGNENT LES CHONDROCYTES COLOREES POSITIVEMENT (K). LES NOYAUX ONT ETE COLORES AVEC DU DAPI (C, D, G, H). LES SECTIONS ONT ETE MONTEES DANS PROLONGTM DIAMOND ET EXAMINEES PAR MICROSCOPIE CONFOCALE. BARRES D'ECHELLE : 50 M POUR A, J, K, L ; 10 M POUR B, C, D ; 200 M POUR E, I ; 40 M POUR F, G, H. ..... - 96 -
FIGURE 30. EXEMPLE DE CHOIX DES PATIENTS A SEQUENCER DANS UNE FAMILLE MULTIPLEX. LES DEUX PATIENTS SEQUENCES SONT ENCERCLES EN ROUGE. LES CARRES ET LES CERCLES CORRESPONDENT AUX HOMMES ET AUX FEMMES, RESPECTIVEMENT. LES SYMBOLES PLEINS ET VIDES CORRESPONDENT AUX PATIENTS ET AUX SUJETS SAINS, RESPECTIVEMENT. LES SYMBOLES GRIS CORRESPONDENT AUX INDIVIDUS POUR LESQUELS L'ADN N'ETAIT PAS DISPONIBLE. ..... - 99 -
FIGURE 31. EXEMPLE DE CHOIX DES PATIENTS A SEQUENCER DANS UNE FAMILLE MULTIPLEX AVEC DEUX ENTREES DE LA MALADIE. LES QUATRE PATIENTS SEQUENCES SONT ENCERCLES EN ROUGE. LES CARRES ET LES CERCLES CORRESPONDENT AUX HOMMES ET AUX FEMMES, RESPECTIVEMENT. LES SYMBOLES PLEINS ET VIDES CORRESPONDENT AUX PATIENTS ET AUX SUJETS SAINS, RESPECTIVEMENT. LES SYMBOLES GRIS CORRESPONDENT AUX INDIVIDUS POUR LESQUELS L'ADN N'ETAIT PAS DISPONIBLE.. - 99 -
FIGURE 32. REPARTITION DU NOMBRE DE PATIENTS SEQUENCES PAR FAMILLE. ..... - 99 -
FIGURE 33. SCHEMA DE FILTRAGE DES VARIANTS IDENTIFIES DANS LE SEQUENÇAGE D’EXOME. ..... - 101 -
FIGURE 34. PEDIGREE DE LA FAMILLE #33 ET SEGREGATION DU VARIANT S2195F. LES CARRES ET LES CERCLES CORRESPONDENT AUX HOMMES ET AUX FEMMES, RESPECTIVEMENT. LES SYMBOLES PLEINS ET VIDES CORRESPONDENT AUX PATIENTS ET AUX SUJETS SAINS, RESPECTIVEMENT. LES SYMBOLES GRIS CORRESPONDENT AUX INDIVIDUS POUR LESQUELS L'ADN N'ETAIT PAS DISPONIBLE. LE STATUT HLA-B27 ("+" OU "-") EST EGALEMENT INDIQUE ET LA PRESENCE DU VARIANT S2195F EST SPECIFIEE PAR 1/2 ET 2/2 POUR LES HETEROZYGOTES ET LES HOMOZYGOTES, RESPECTIVEMENT TANDIS QUE L’ABSENCE DU VARIANT EST INDIQUE PAR 1/1. LES PATIENTS NON PORTEURS DU VARIANT SONT ENCERCLES EN VERT ET LES PORTEURS SAINS SONT ENCERCLES EN ROUGE. ..... - 103 -
FIGURE 35. PEDIGREE DE LA FAMILLE #110 ET SEGREGATION DU VARIANT K66N. LES CARRES ET LES CERCLES CORRESPONDENT AUX HOMMES ET AUX FEMMES, RESPECTIVEMENT. LES SYMBOLES PLEINS ET VIDES CORRESPONDENT AUX PATIENTS ET AUX SUJETS SAINS, RESPECTIVEMENT. LES SYMBOLES GRIS CORRESPONDENT AUX INDIVIDUS POUR LESQUELS L'ADN N'ETAIT PAS DISPONIBLE. LE STATUT HLA-B27 ("+" OU "-") EST EGALEMENT INDIQUE ET LA PRESENCE DU VARIANT K66N EST SPECIFIEE PAR 1/2 ET 2/2 POUR LES HETEROZYGOTES ET LES HOMOZYGOTES, RESPECTIVEMENT TANDIS QUE L’ABSENCE DU

VARIANT EST INDIQUE PAR 1/1. LE PATIENT NON PORTEUR DU VARIANT EST CERCLE EN VERT ET LES PORTEURS SAINS SONT CERCLES EN ROUGE. ..... - 104 -
FIGURE 36. RECONSTRUCTION DES HAPLOTYPES DANS LA FAMILLE 110. TOUS LES MEMBRES SONT IDENTIFIES PAR UN NUMERO. LES CARRES ET LES CERCLES CORRESPONDENT AUX HOMMES ET AUX FEMMES, RESPECTIVEMENT. LES SYMBOLES PLEINS ET VIDES CORRESPONDENT AUX PATIENTS ET AUX SUJETS SAINS, RESPECTIVEMENT. LES SYMBOLES GRIS CORRESPONDENT AUX INDIVIDUS POUR LESQUELS L'ADN N'ETAIT PAS DISPONIBLE. LES HAPLOTYPES SONT REPRESENTES EN BARRE DE COULEUR ET L'HAPLOTYPE DELETERE EST REPRESENTE EN BARRE DE COULEUR NOIR. LES PATIENTS AYANT LES RECOMBINAISONS INFORMATIVES SONT ENCERCLES EN ORANGE, ROUGE, VERT, BLEU CLAIR ET BLEU FONCE SELON LES EVENEMENTS DE RECOMBINAISON. ..... - 106 -
FIGURE 37. LES HAPLOTYPES DES PATIENTS DE LA FAMILLE 110 AYANT DES EVENEMENTS DE RECOMBINAISON. LES PATIENTS SONT IDENTIFIES PAR UN NUMERO. LA COULEUR ORANGE REPRESENTE LA PARTIE DE L'HAPLOTYPE SEGREGEANT AVEC LA MALADIE. LES DEUX CARRES TRANSPARENTS INDIQUENT LES DEUX HAPLOTYPES PROXIMAL ET DISTAL NON CHEVAUCHANTS. ..... - 106 -
FIGURE 38. SCHEMA DE FILTRAGE DES VARIANTS IDENTIFIES DANS LE SEQUENÇAGE D'EXOMES AU NIVEAU DU CHROMOSOME 9. ..... - 107 -
FIGURE 39. SCHEMA DU CHROMOSOME 9 AVEC LES POSITIONS DES VARIANTS PROCHES DU VARIANT ENDOG*K66N. LES RECOMBINAISONS SONT SITUEES ENTRE LES DEUX CARRES TRANSPARENTS ENTRE LES GENES DENND1A ET PTGES2 (125 MB ET 128 MB). LA POSITION DE CHAQUE VARIANT EST INDIQUEE PAR UNE FLECHE ET EN MEGABASES (MB). LA ZONE OU LES VARIANTS D'INTERET SONT LOCALISES SUR LE CHROMOSOME 9 EST ENCADREE EN ROUGE. ..... - 108 -
FIGURE 40. PEDIGREE DE LA FAMILLE #110 ET SEGREGATION DES VARIANTS DENND1A*A784D ET CRB2*R93H. LES CARRES ET LES CERCLES CORRESPONDENT AUX HOMMES ET AUX FEMMES, RESPECTIVEMENT. LES SYMBOLES PLEINS ET VIDES CORRESPONDENT AUX PATIENTS ET AUX SUJETS SAINS, RESPECTIVEMENT. LES SYMBOLES GRIS CORRESPONDENT AUX INDIVIDUS POUR LESQUELS L'ADN N'ETAIT PAS DISPONIBLE. LE STATUT HLA-B27 ("+" OU "-") EST EGALEMENT INDIQUE ET LA PRESENCE DES VARIANTS A784D ET R93H EST SPECIFIEE PAR 1/2 ET 2/2 POUR LES HETEROZYGOTES ET LES HOMOZYGOTES, RESPECTIVEMENT TANDIS QUE L'ABSENCE DE CES VARIANTS EST INDIQUE PAR 1/1. LE PATIENT NON PORTEUR DES VARIANTS EST ENCERCLE EN BLEU ET LES PORTEURS SAINS SONT ENCERCLES EN ROUGE. ..... - 109 -
FIGURE 41. EXEMPLE D'UNE FAMILLE MULTIPLEX DE 11 PATIENTS. LES CARRES ET LES CERCLES CORRESPONDENT AUX HOMMES ET AUX FEMMES, RESPECTIVEMENT. LES SYMBOLES PLEINS ET VIDES CORRESPONDENT AUX PATIENTS ET AUX SUJETS SAINS, RESPECTIVEMENT. LES SYMBOLES GRIS CORRESPONDENT AUX INDIVIDUS POUR LESQUELS L'ADN N'ETAIT PAS DISPONIBLE. TOUS LES PATIENTS ONT ETE SEQUENCES. ..... - 111 -
FIGURE 42. ANALYSE EN COMPOSANTE PRINCIPALE DES DONNEES DU SEQUENÇAGE GENOME ENTIER. LES POPULATIONS SONT REPRESENTEES AVEC UN CODE COULEUR. LES AFRICAINS (AFR) EN ROUGE, LES AMERICAINS (AMR) EN JAUNE, LES ASIATIQUES DE L'EST EN VERT (EAS), LES EUROPEENS EN BLEU CIEL (EUR), LES ASIATIQUES DU SUD EN BLEU FONCE (SAS) ET NOS ECHANTILLONS EN FUSHIA (WGS). ..... - 113 -

FIGURE 43. ANALYSE EN COMPOSANTE PRINCIPALE. LES DONNEES QUI SE CHEVAUCHENT AVEC D'AUTRES POPULATIONS SONT CERCLEES AVEC LES CODES COULEUR DE CES POPULATIONS. EUR-FIN : EUROPEENS – FINLANDAIS, EUR-AMR : EUROPEENS – AMERICAINS ET EUR-AFR : EUROPEENS – AFRICAINS. ..... - 114 -
FIGURE 44. STRATEGIE DE FILTRAGE DES SNPS. ..... - 115 -
FIGURE 45. REPARTITION DES VARIANTS FILTRES EN FONCTION DU NOMBRE DE FAMILLES OU LE VARIANT COSEGREGE AVEC LA MALADIE (C'EST-A-DIRE EST PRESENT CHEZ AU MOINS N-1 MALADES, N REPRESENTANT LE NOMBRE DE MALADES SEQUENCES DANS LA FAMILLE). ..... - 115 -
FIGURE 46. REPARTITION DES VARIANTS FILTRES EN FONCTION DE LEUR SCORE CADD ..... - 116 -
FIGURE 47. DOT PLOT DES RESULTATS DES ANALYSES D'ENRICHISSEMENT GO. ..... - 118 -
FIGURE 48. EMAP PLOT DES RESULTATS DES ANALYSES D'ENRICHISSEMENT GO. ..... - 119 -

INDEX DES TABLEAUX

TABLEAU 1. LES CRITERES DE CLASSIFICATION D'AMOR [38]. - 25 -
TABLEAU 2. LES CRITERES DE CLASSIFICATION DE L'ESSG [39]. - 26 -
TABLEAU 3. CRITERES DE CLASSIFICATION ASAS [23,40] - 27 -
TABLEAU 4. LES CRITERES DE NEW YORK MODIFIES [41] - 28 -
TABLEAU 5. LES DIFFERENTES TECHNIQUES DE GENOTYPAGE ET DE SEQUENÇAGE. - 42 -
TABLEAU 6. RESUME DES METHODES DE PREDICTION DE L'IMPACT FONCTIONNEL DES VARIANTS. - 61 -
TABLEAU 7. RESUME DES METHODES STATISTIQUES POUR LES TESTS D'ASSOCIATION DES VARIANTS RARES [122]. - 63 -
TABLEAU 8. RESUME DES ETUDES D'AGREGATION FAMILIALE DANS LA SPA. - 66 -
TABLEAU 9. LES AUTRES ALLELES HLA-B ET GENES HLA ASSOCIES AU DEVELOPPEMENT DE LA SPA. - 70 -
TABLEAU 10. RESUME DES ETUDES DE LIAISON CONDUITES DANS LA SPA. - 72 -
TABLEAU 11. RESULTATS DES ETUDES D'ASSOCIATION DANS LA SPA. - 74 -
TABLEAU 12. LES VARIANTS RARES IDENTIFIES DANS LA SPA. - 78 -
TABLEAU 13. CARACTERISTIQUES DES PATIENTS INCLUS DANS LES ETUDES FAMILIALES ET CAS-CONTROLES. - 85 -
TABLEAU 14. TECHNIQUES DE GENOTYPAGE UTILISEES EN FONCTION DES VARIANTS - 86 -
TABLEAU 15. CARACTERISTIQUES DES 24 VARIANTS NON SYNONYMES IDENTIFIES PAR LE RESEQUENÇAGE DE LA REGION CIBLE 13Q13. - 88 -
TABLEAU 16. CARACTERISTIQUES DES 33 VARIANTS NON CODANTS AVEC UN SCORE REGULOMEDB ≤3 IDENTIFIES PAR LE RESEQUENÇAGE DE LA REGION CIBLEE 13Q13. - 89 -
TABLEAU 17. SYNTHESE DES ANALYSES DE SEGREGATION POUR LES VARIANTS CODANTS ET NON CODANTS. - 92 -
TABLEAU 18. ANALYSE D'ASSOCIATION CAS-TEMOINS. - 93 -
TABLEAU 19. FREQUENCE ALLELIQUE DU VARIANT R1840W EN FONCTION DE LA MALADIE ET DU STATUT HLA-B27 - 93 -
TABLEAU 20. CARACTERISTIQUES DES PATIENTS INCLUS DANS LES ETUDES FAMILIALES ET CAS-CONTROLES. - 100 -
TABLEAU 21. TECHNIQUES DE GENOTYPAGE UTILISEES EN FONCTION DES VARIANTS - 101 -
TABLEAU 22. ANALYSE D'ASSOCIATION CAS-TEMOINS DU VARIANT ENDOG*K66N. - 104 -
TABLEAU 23. ANALYSE D'ASSOCIATION CAS-TEMOINS DES VARIANTS PROCHES DU VARIANT K66N. - 110 -
TABLEAU 24. CARACTERISTIQUES DES PATIENTS SEQUENCES. - 112 -

INTRODUCTION GENERALE

I. La spondyloarthrite : aspects épidémiologiques, cliniques et thérapeutiques.

La spondyloarthrite (SpA) est un rhumatisme inflammatoire chronique qui affecte la colonne vertébrale et les articulations périphériques. Chez certains patients, on observe aussi des atteintes extra-articulaires, notamment l'inflammation de la peau, de l'intestin et des yeux. L'atteinte des enthèses, qui sont les sites d'insertion des tendons et des ligaments aux os, est la principale caractéristique de la SpA qui la distingue des autres rhumatismes inflammatoires.

Dans cette première partie, nous aborderons d'abord l'épidémiologie de cette maladie avant d'en détailler les aspects cliniques et thérapeutiques.

I.1. Epidémiologie

I.1.1. Prévalence

La SpA est inégalement répartie sur le plan géographique avec une prévalence plus élevée dans les pays du Nord par rapport aux pays du Sud (Figure 1), en lien avec la fréquence du HLA-B27, le principal facteur génétique de susceptibilité à la maladie. La fréquence la plus élevée apparaît dans le Grand Nord, dans des populations telles que les Esquimaux d'Alaska et de Sibérie et les Lapons scandinaves [1,2]. La prévalence de la SpA varie de 0,3 % à 2,5 % [3] et a été évaluée en 2013 à 0,43 % de la population adulte en France [4].

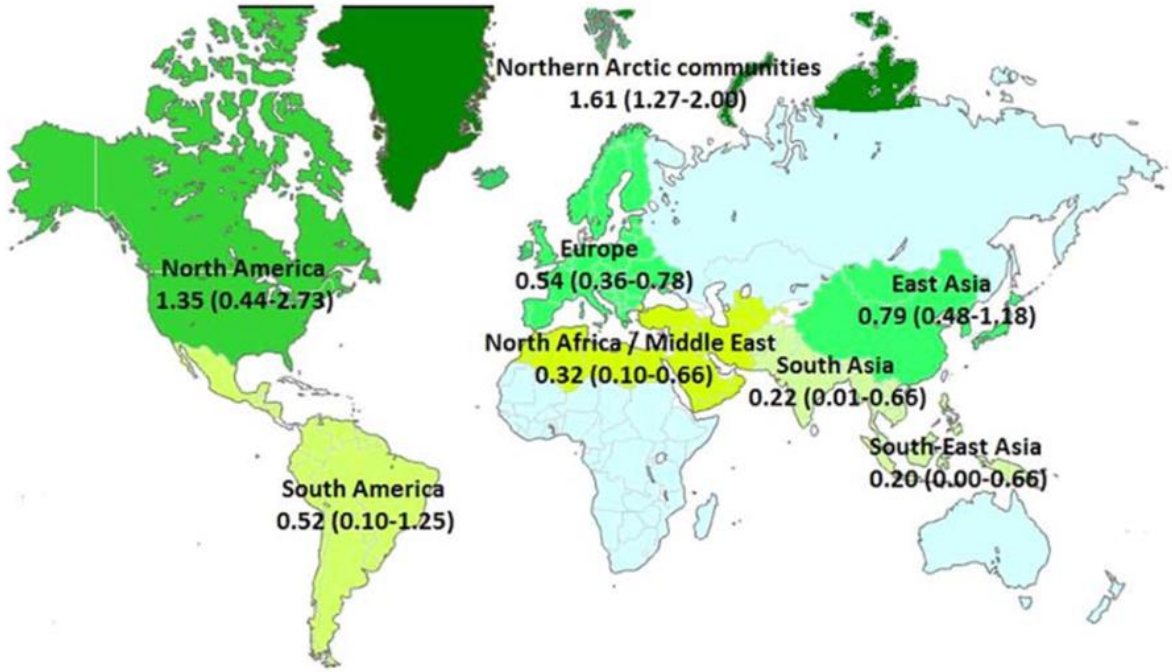

Figure 1. La prévalence mondiale de la SpA (adapté de Stolwijk C et al. Arthritis Care Res 2016) [5].

I.1.2. Population touchée

La SpA débute généralement chez l'adulte jeune (le plus souvent avant 30 ans) sans prédominance de sexe [6].

Le diagnostic de la SpA peut mettre de nombreuses années à être établi (7 ans en moyenne entre les premiers symptômes et le diagnostic définitif) à cause de sa grande hétérogénéité clinique et de l'absence de spécificité de la plupart de ses manifestations cliniques [7,8].

I.2. Diagnostic

Le diagnostic se fait principalement sur la base des manifestations cliniques. Les examens complémentaires biologiques et surtout d'imagerie peuvent aider, mais aucun d'entre eux ne permet d'établir un diagnostic de façon certaine.

I.2.1. Manifestations cliniques

Dans la SpA, on distingue plusieurs types de manifestations cliniques, les principales étant articulaires.

I.2.1.1. Manifestations articulaires

En fonction des sites anatomiques touchés, on distingue trois catégories de manifestations articulaires : les manifestations axiales, les enthésites et les manifestations articulaires

périphériques.

Les manifestations axiales

L'atteinte axiale comprend toutes les manifestations cliniques en rapport avec les articulations sacro-iliaques, les ligaments ou les articulations des vertèbres et de la paroi thoracique antérieure (Figure 2).

L'inflammation de ces trois sites articulaires se traduit par la douleur et la raideur. Dans l'atteinte axiale, le risque principal est la survenue d'une ankylose, conséquence d'une ossification ligamentaire progressive, qui peut entraîner une diminution de la souplesse rachidienne.

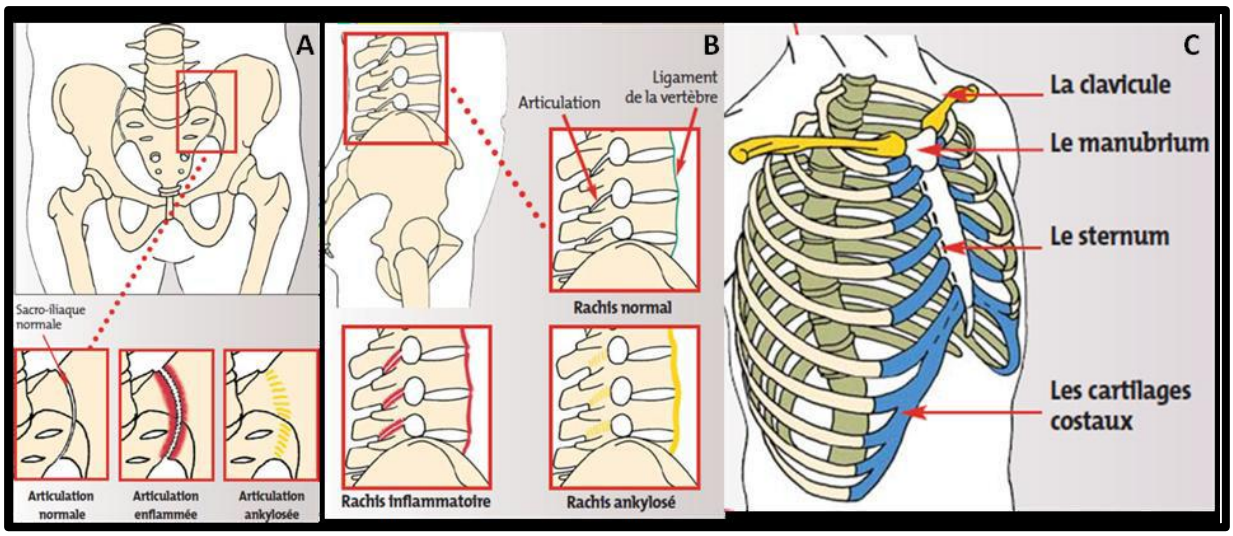

Figure 2. Représentations des atteintes articulaires axiales dans la SpA. (D'après "Spondylarthrite en 100 questions" Dougados M, Edition NHA Communication, 2005). A – articulations sacro-iliaques ; B – rachis lombaire ; C – paroi thoracique antérieure

Les manifestations enthésitiques périphériques

L'atteinte inflammatoire des enthèses (enthésite), sites d'insertion des tendons dans les os, est l'une des atteintes les plus caractéristiques de la SpA (Figure 3). Elle évolue en trois phases successives : d'abord une phase inflammatoire, puis une phase érosive et enfin une phase de cicatrisation qui entraîne une fibrose puis une ossification à l'origine de la formation d'enthésophytes [9].

Toutes les enthèses fibrocartilagineuses peuvent être touchées. L'atteinte la plus fréquente se situe au niveau du tendon d'Achille ou de l'aponévrose plantaire et se traduit par des talalgies.

Les atteintes articulaires périphériques

La fréquence des arthrites dans la SpA varie selon les études et est évaluée entre 40 et 55% [10,11]. Le point de départ de l'inflammation est la membrane synoviale qui peut secondairement retentir sur le cartilage et l'os sous-jacent, entraînant leur destruction : on parle alors d'arthrite érosive.

La dactylite ou « doigt en saucisse » est une manifestation clinique très évocatrice du diagnostic de SpA. Il s'agit d'une tuméfaction globale d'un ou plusieurs doigts ou orteils, douloureuse, associée à une coloration rouge foncé.

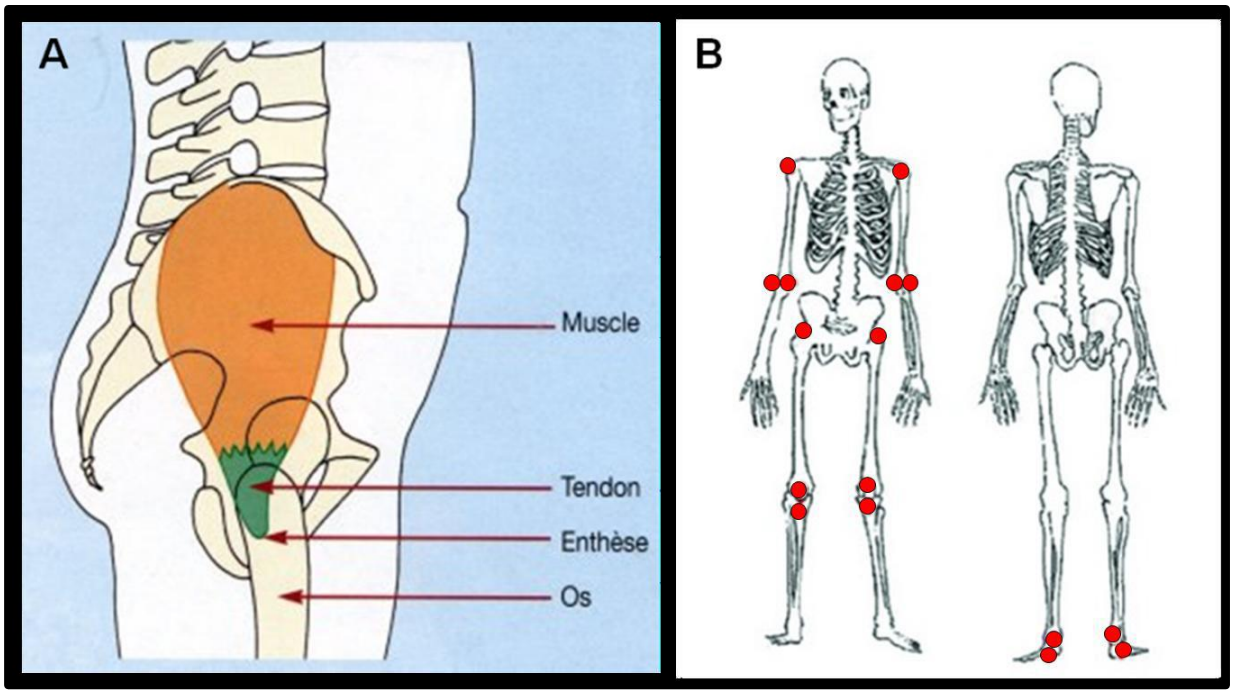

Figure 3. Représentations des atteintes des enthèses dans la SpA. ("Spondylarthrite en 100 questions" Dougados M, Edition NHA Communication, 2005 et Maksymowych et al. [12]). A – enthèse trochantérienne ; B – sites enthésitiques les plus fréquemment touchés.

La SpA peut toucher d'autres sites en dehors des articulations.

I.2.1.2. Manifestations extra-articulaires

Atteinte ophtalmique

L'atteinte ophtalmique la plus typique est l'uvéite antérieure aiguë non granulomateuse, c'est-à-dire l'inflammation de la chambre antérieure de l'œil (iritis) et du corps ciliaire (cyclite), formant le segment antérieur de l'œil. Elle entraîne une douleur, une rougeur oculaire avec larmoiements, photophobie et baisse d'acuité visuelle d'installation brutale.

L'atteinte ophtalmique est l'atteinte extra-articulaire la plus fréquente [13] concernant environ un tiers des patients souffrant de SpA [14].

Atteinte cutanée

L'atteinte cutanée la plus fréquente est le psoriasis dont la prévalence varie entre 8 et 20 % [10,13] dans la SpA alors qu'elle est évaluée entre 1,5 et 3 % dans la population générale [15,16]. Le psoriasis précède dans plus des deux tiers des cas les manifestations articulaires, mais apparaît parfois plusieurs années après le développement de la SpA [17].

Atteinte digestive

La SpA s'accompagne fréquemment d'une inflammation digestive (principalement du colon) qui peut être symptomatique (maladie inflammatoire chronique de l'intestin (MICI)) ou non [18].

Les MICI (maladie de Crohn et rectocolite hémorragique) touchent environ 8% des patients de SpA [13]. Elles se traduisent par une diarrhée glairo-sanglante, chronique ou évoluant en poussées aiguës, associée à de violentes douleurs abdominales. Le diagnostic repose sur l'endoscopie digestive avec étude histologique de biopsies.

Autres atteintes

Plusieurs autres organes peuvent être touchés comme l'appareil uro-génital, le cœur ou les poumons [19,20].

I.2.2. Examens complémentaires

Certains examens complémentaires peuvent être utiles au diagnostic de la SpA car il est difficile d'établir un diagnostic uniquement sur des critères cliniques. Il faut cependant noter qu'aucun examen complémentaire ne permet à lui seul de faire le diagnostic.

I.2.2.1. Examens biologiques

Typage HLA-B27

La détection de l'allèle HLA-B27 constitue une aide au diagnostic de SpA, surtout en cas de contexte clinique évocateur. La fréquence de l'allèle HLA-B27 est plus élevée chez les patients SpA que dans la population générale. Il a été montré que le risque d'être atteint d'une SpA est multiplié par 40 chez les patients porteurs de l'allèle HLA-B27 [4].

La présence de l'allèle HLA-B27 ne doit toutefois pas être considérée comme un test

diagnostique, car elle n'est ni suffisante, ni nécessaire au développement de la SpA [20,21].

Autres examens biologiques

La vitesse de sédimentation et la protéine C-réactive peuvent être utiles dans l'évaluation de l'activité de la maladie, mais ne peuvent être considérées comme des marqueurs diagnostiques, car elles ne sont supérieures à la normale que chez environ la moitié des patients atteints de SpA [20,22]. De plus, elles ne sont pas spécifiques.

I.2.2.2. Imagerie

Radiologie conventionnelle

Les radiographies standard peuvent être utiles au diagnostic, mais la plupart des anomalies constatées apparaissent tardivement. Par conséquent, ce n'est pas un examen de choix pour le diagnostic précoce de SpA [23].

Les atteintes radiologiques typiques sont la présence d'érosions osseuses ou d'ossification, que ce soit au niveau sacro-iliaque, rachidien ou enthésitique.

Imagerie par résonance magnétique

L'IRM est intéressante dans le diagnostic précoce de la maladie. En effet, de nombreuses études ont montré que, grâce à cette technique, il était possible d'observer des anomalies des années avant que celles-ci ne soient détectables sur des radiographies standards [24]. Les anomalies observées sont de deux types : inflammatoires (œdème osseux, synovite, enthésite ou capsulite) ou structurales (sclérose sous-chondrale, érosions, dégénérescence graisseuse ou syndesmophytes) [25,26].

Scanner

Le scanner des articulations sacro-iliaques peut être utile dans la détection d'anomalies structurales trop fines pour être encore observées sur les radiographies standards [27]. L'utilisation est toutefois limitée en raison de l'irradiation importante de cet examen.

Echographie

L'échographie est capable d'analyser finement la morphologie des enthèses et de visualiser des lésions de petite taille. Le couplage aux techniques de Doppler puissance permet le repérage d'une hypervascularisation locale reflétant l'inflammation [28]. L'échographie est

plus sensible que l'examen clinique dans la détection des enthésites périphériques [29,30].

I.3. Les différentes formes de la SpA

En fonction des manifestations cliniques prédominantes, plusieurs sous-types de SpA ont été décrits (Figure 4) :

I.3.1. La spondylarthrite ankylosante (SA)

La spondylarthrite ankylosante est la forme la plus typique de la SpA, caractérisée par une atteinte du squelette axial (rachis et sacro-iliaques) pouvant conduire à une ankylose (Figure 2). Elle se manifeste par la présence de changements structuraux sur les radiographies des articulations sacro-iliaques.

Près de 90 % des patients atteints de la SA sont porteurs de l'allèle HLA-B27.

I.3.2. Le rhumatisme psoriasique

Le rhumatisme psoriasique est un rhumatisme inflammatoire chronique qui fait partie des SpA et se caractérise par un début plus tardif que dans les autres formes de SpA et par une fréquence plus importante des manifestations articulaires périphériques.

L'association avec HLA-B27 est moins forte que pour les autres sous-types.

I.3.3. Les arthrites réactionnelles

Les arthrites réactionnelles sont la conséquence d'une inflammation consécutive à une infection génitale ou digestive, d'origine bactérienne [31]. Il ne s'agit pas d'arthrites septiques, mais d'arthrites post-infectieuses qui se développent en général deux à trois semaines après l'infection initiale. Elles sont caractérisées par une fréquence élevée de manifestations articulaires périphériques (oligoarthrite). L'évolution peut se faire vers une guérison sans séquelles, par la survenue de récidives ou par un passage à la chronicisation.

I.3.4. La SpA associée aux MICI

Les entérocolopathies inflammatoires chroniques regroupent deux maladies inflammatoires du tube digestif : la maladie de Crohn et la rectocolite hémorragique. Dans 10 à 20 % des maladies de Crohn et des rectocolites hémorragiques surviennent des manifestations articulaires inflammatoires axiales ou périphériques.

I.3.5. Les spondyloarthrites indifférenciées

Les SpA indifférenciées sont des SpA répondant aux critères de classification des SpA mais ne correspondant à aucune des catégories précédemment décrites. Elles peuvent évoluer secondairement vers une forme différenciée.

I.4. Concept unifié de SpA

De nombreux arguments laissent penser que ces différentes formes de SpA ne représentent que des manifestations cliniques différentes d'une même pathologie [32]. Il a été observé la coexistence de plusieurs formes de SpA chez les patients d'une même famille, ce qui suggère un terrain génétique commun [33].

Les rats transgéniques pour le HLA-B27 et la β2-microglobuline humaine développent simultanément plusieurs symptômes (arthrite, colite, psoriasis, spondylite) qui correspondent à différents sous-types de SpA humaine [34].

Selon les manifestations cliniques articulaires prédominantes, on distingue désormais la SpA axiale (qui est caractérisée par une atteinte prédominante du rachis et/ou des sacro-iliaques) et la SpA périphérique (arthrite périphérique, enthésite et/ou dactylite)[35,36].

Figure 4. Les différentes formes de la SpA [37]

I.5. Critères de classification de la SpA

Les différentes caractéristiques cliniques et paracliniques de la SpA ont été utilisées pour établir des critères de classification. Il ne s'agit pas de critères diagnostiques utiles à l'échelon individuel, mais de critères ayant comme but de constituer des groupes homogènes de patients pour les études thérapeutiques. Il existe plusieurs critères de classification avec des pourcentages de sensibilité et spécificité variables.

I.5.1. Critères d'Amor

Ils comprennent des critères cliniques et radiologiques avec des scores affectés allant de 1 à 3. Un score supérieur ou égal à 6 est nécessaire pour classer le patient comme atteint d'une SpA.

Tableau 1. Les critères de classification d'Amor [38].

Catégorie	**Critère**	**Points**
Signes cliniques ou histoire clinique	Douleurs nocturnes lombaires ou dorsales et/ou raideur matinale lombaire ou dorsale	1
	Douleurs fessières uni- ou bilatérales	1
	Douleurs fessières à bascule	2
	Oligoarthrite asymétrique	2
	Doigt ou orteil « en saucisse »	2
	Talalgie ou autre enthésopathie	2
	Iritis ou uvéite antérieure aiguë	2
	Urétrite non gonococcique ou cervicite moins d'un mois avant le début d'une arthrite	1
	Diarrhée moins d'un mois avant le début d'une arthrite	1
	Présence ou antécédent de psoriasis et/ou de balanite et/ou d'entérocolopathie chronique	2
Signes radiologiques	Sacro-iliite radiologique ≥ stade 2 si bilatérale ou stade 3 si unilatérale	3
Terrain génétique	Présence de l'antigène HLA-B27 ou antécédents familiaux de pelvispondylite, d'arthrite réactionnelle (ex-syndrome de Fiessinger-Leroy-Reiter), de syndrome de Reiter, de psoriasis, d'entérocolopathies chroniques	2
Sensibilité au traitement	Amélioration en 48 heures des douleurs par AINS et/ou rechute rapide (48 heures) des douleurs à leur arrêt	2

I.5.2. Critères de l'ESSG

Ces critères ont été élaborés afin de prendre en compte les patients ayant des formes indifférenciées de SpA. Le diagnostic est déclaré avec la présence d'un critère majeur et d'au moins un critère mineur.

Tableau 2. Les critères de classification de l'ESSG [39].

Catégorie	**Critère**
Critères majeurs	- Douleurs rachidiennes inflammatoires (début avant 45 ans, amélioration par l'exercice, avec raideur matinale et durée > 3 mois) - Synovites asymétriques ou prédominantes aux membres inférieurs
Critères mineurs	- Antécédents familiaux de spondyloarthrite ou d'uvéite ou d'entérocolopathie - Psoriasis - Maladie inflammatoire intestinale (entérocolopathie) - Urétrite, cervicite ou diarrhée aiguë dans le mois précédant l'arthrite - Diarrhée aiguë - Douleurs fessières à bascule - Enthésopathie - Sacro-iliite radiologique (bilatérale si grade ≥ 2, unilatérale si grade > 3)

I.5.3. Critères ASAS (Assessment of SpondyloArthritis international Society).

Les critères ASAS sont le système de classification de spondyloarthrite le plus récent. Ces critères permettent de séparer les formes périphériques des formes axiales de SpA.

Tableau 3. Critères de classification ASAS [23,40]

Critères de Spondylarthrite axiale	
Sacro-iliite à l'imagerie* et au moins un signe de SpA** Ou HLA-B27 positif et au moins deux signes de SpA**	* Sacro-iliite à l'imagerie : - Inflammation active (aiguë) à l'IRM fortement suggestive de sacro-iliite associée à une SpA - Sacro-iliite radiologique selon les critères modifiés de New York
	** Signes de SpA : - Rachialgie inflammatoire - Arthrite - Enthésite (talon) - Uvéite - Dactylite - Psoriasis - Maladie de Crohn/rectocolite hémorragique - Bonne réponse aux AINS - Antécédent familial de SpA - HLA-B27 - CRP élevée

Critères de Spondylarthrite périphérique	
Arthrite OU Enthésite OU Dactylite Avec :	
**** Au moins deux signes de SpA :** - Arthrite - Enthésite - Dactylite - Lombalgies inflammatoires - Antécédent familial de SpA	*** Au moins un signe de SpA :** - Uvéite - Psoriasis - MICI - Infection précédant les symptômes - HLA-B27 - Sacro-iliite à l'imagerie

I.5.4. Critères de New York modifiés

Les critères de New York ne prennent en compte que les patients ayant la spondyloarthrite ankylosante. Le diagnostic est déclaré en cas de présence d'un critère radiologique et d'au moins un critère clinique.

Tableau 4. Les critères de New York modifiés [41]

Critères cliniques	- Lombalgies avec raideur de plus de trois mois, améliorées à l'effort, mais ne cédant pas au repos - Limitation des mouvements du rachis lombaire à la fois dans le plan frontal et sagittal. - Limitation de l'ampliation thoracique par rapport aux valeurs normales corrigées pour l'âge et le sexe
Critères radiologiques	Sacro-iliite bilatérale de grade ≥ 2, ou sacro-iliite unilatérale de grade ≥ 3

I.6. Traitement

Le traitement de la spondyloarthrite s'appuie à la fois sur des traitements médicamenteux et des mesures non médicamenteuses.

I.6.1. Les traitements médicamenteux

L'arsenal thérapeutique disponible dans la SpA augmente progressivement. Toutefois, les traitements restent symptomatiques, car ils visent à réduire la douleur et l'ankylose mais ne permettent actuellement pas de guérir la maladie.

I.6.1.1. Les anti-inflammatoires non stéroïdiens (AINS)

Les AINS sont le traitement de première ligne de la SpA. Ils sont habituellement rapidement efficaces et si l'AINS est arrêté, une rechute douloureuse est observée en moins de quarante-huit heures.

En absence de bonne évolution après deux semaines de traitement, il est conseillé de tester un autre AINS. Il faut essayer au moins deux AINS successivement sur quatre semaines au total avant de conclure à l'échec du traitement.

I.6.1.2. Les traitements dits de fond (Disease Modifying Anti-Rheumatic Drug ou DMARD)

Pour les patients chez lesquels les AINS sont inefficaces ou insuffisamment efficaces, des traitements de fond peuvent être proposés.

Les traitements de fond conventionnels synthétiques (conventional synthetic DMARD)

Il s'agit notamment de la sulfasalazine, du méthotrexate et du léflunomide. Leur place est actuellement très limitée dans la prise en charge de la SpA car ils n'ont pas démontré d'efficacité notable en dehors de quelques études montrant une efficacité modeste sur les

symptômes articulaires périphériques.

Les traitements de fond par biothérapie (biological DMARD)

a. Les biomédicaments anti-TNF alpha

L'étanercept, l'adalimumab, le certolizumab et le golimumab peuvent être utilisés pour le traitement des spondyloarthrites axiales et du rhumatisme psoriasique.

L'infliximab est indiqué dans le traitement de la SA et du rhumatisme psoriasique.

b. Les biomédicaments anti-IL17

Le sécukinumab et l'ixekizumab sont indiqués dans la SpA axiale et le rhumatisme psoriasique en cas d'échec des traitements conventionnels.

c. Les biomédicaments anti-IL23

L'ustékinumab est indiqué dans le rhumatisme psoriasique en cas d'échec des traitements conventionnels. Il n'a en revanche pas démontré d'efficacité dans les manifestations axiales de la SpA.

Le guselkumab est indiqué dans le rhumatisme psoriasique périphérique en cas d'échec des traitements conventionnels.

Les traitements de fond synthétiques dirigés sur une cible (targeted synthetic DMARD)

Il s'agit de molécules synthétisées chimiquement et dirigées contre des cibles spécifiques, la plupart enzymatiques. Plusieurs d'entre eux ont été développés ces dernières années, visant la phosphodiestérase 4 (apremilast) ou les Janus kinases (tofacitinib, upadacitinib).

Pour le moment, l'upadacitinib a obtenu une autorisation de mise sur le marché dans la SpA en France, et l'apremilast et le tofacitinib ont obtenu une autorisation de mise sur le marché en France chez les patients atteints de rhumatisme psoriasique, mais des études de phase 3 viennent d'être publiées dans la SpA axiale pour le tofacitinib et devraient amener rapidement à leur commercialisation dans cette indication [42–44].

Leur principal avantage par rapport aux traitements biologiques est leur voie orale d'administration (moins contraignante que les voies intraveineuses ou sous-cutanées).

I.6.2. Les traitements non pharmacologiques

Le tabac est un facteur de mauvais pronostic. L'arrêt du tabagisme est une mesure non

pharmacologique indispensable.

L'activité physique régulière et la kinésithérapie accompagnée sont indispensables, car elles permettent de réduire l'enraidissement.

La physiothérapie, l'ergothérapie et les techniques d'appareillage peuvent également être utilisées en fonction des atteintes et de leur évolution sous l'effet du traitement AINS.

I.6.3. La chirurgie

Dans certains cas, la chirurgie peut être indiquée, principalement en cas de remplacement prothétique articulaire.

II. La génétique des maladies complexes

II.1. Les maladies complexes

II.1.1. Maladies complexes vs maladies mendéliennes

La majorité des maladies implique, dans des proportions variables, la participation de facteurs génétiques et environnementaux pouvant interagir entre eux. On distingue schématiquement les maladies monogéniques ou mendéliennes des maladies complexes ou multifactorielles.

Dans le premier cas, l'apparition de la maladie chez un individu est due exclusivement à une ou plusieurs mutations délétères d'un seul gène essentiel à une fonction majeure, telle la mucoviscidose [45]. De l'autre côté, les maladies complexes, comme la SpA, sont dues à une combinaison de plusieurs facteurs génétiques et de facteurs environnementaux pouvant interagir entre eux [46].

Toutefois, il ne faut pas trop opposer l'influence des gènes et de l'environnement dans les maladies monogéniques versus les maladies complexes (Figure 5).

Les mutations impliquées dans les maladies mendéliennes ont des effets plus forts que celles impliquées dans les maladies complexes. Par conséquent, elles sont plus rares.

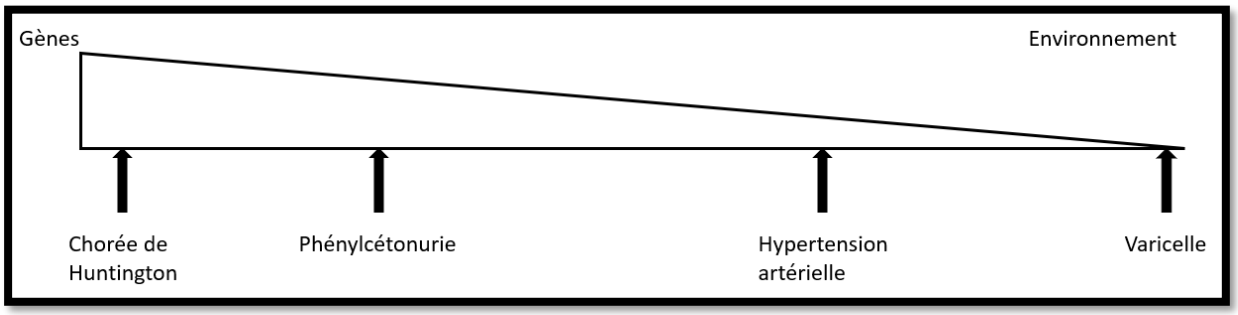

Figure 5. Exemple des maladies mendéliennes et maladies complexes.

II.1.2. Estimation de la contribution des facteurs génétiques à la survenue d'une maladie complexe

Dans les maladies complexes, plusieurs arguments permettent d'évaluer le poids des facteurs génétiques dans la susceptibilité à la maladie.

Le premier argument est celui de l'agrégation familiale. Elle est définie par l'indice de récurrence λ, qui est le rapport de la prévalence de la maladie chez des apparentés d'un sujet

malade et de la prévalence observée dans la population générale (Figure 6A) [47].

L'indice le plus communément utilisé est λs (s pour siblings c'est-à-dire frère ou sœur). Il y a agrégation familiale lorsque λs est nettement supérieur à 1. Dans les maladies mendéliennes, λs est le plus souvent supérieur à 100 alors qu'il est en général inférieur à 50 dans les maladies complexes [48].

Dans un second temps, les études de jumeaux sont utiles pour distinguer la contribution des facteurs génétiques ou environnementaux conduisant à l'agrégation familiale. Elles consistent à comparer le taux de concordance d'une maladie chez les jumeaux monozygotes (MZ) et dizygotes (DZ) (Figure 6B).

Les MZ, issus du même œuf, ont le même patrimoine génétique et donc toute discordance (un jumeau malade, l'autre sain) est par conséquent d'origine environnementale. Les DZ, issus de deux œufs différents, ont la moitié de leur patrimoine génétique en commun, les discordances sont donc d'origine génétique ou environnementale.

Lorsque le taux de concordance des MZ est supérieur à celui des DZ, on peut dire qu'il y a un poids important de la composante génétique dans la survenue de la maladie par rapport aux facteurs environnementaux. Le taux de concordance des jumeaux permet aussi de calculer l'héritabilité, qui est la part attribuée à la génétique dans la variabilité d'un trait ou d'une maladie donnée et dans une population donnée [49].

Enfin, une fois la composante génétique d'une maladie démontrée, il est important d'essayer de déterminer le mode de transmission le plus probable. Pour ce faire, les analyses de ségrégation permettent de déterminer si la répartition familiale de la maladie est compatible avec la ségrégation d'un gène majeur, d'estimer ses caractéristiques (mode de transmission, pénétrance des différents génotypes), de montrer la présence ou non d'une composante polygénique (Figure 6C).

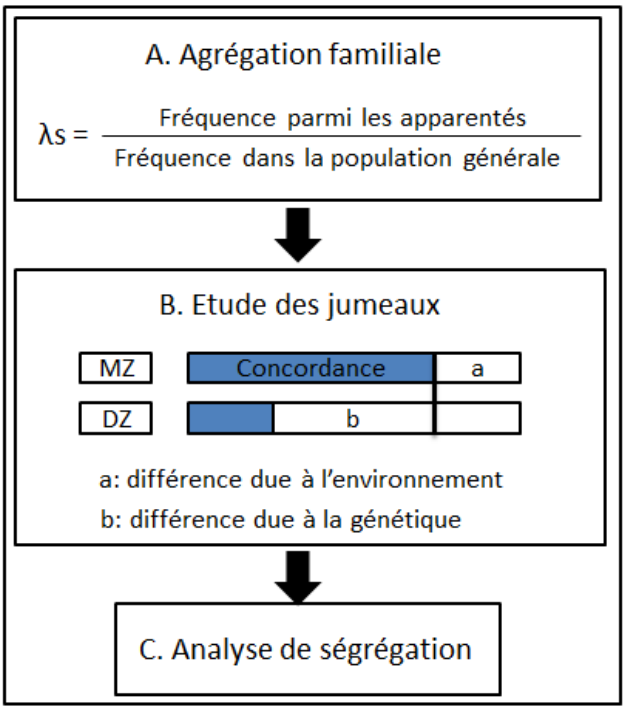

Figure 6. Evaluation du poids génétique dans les maladies complexes [50].

Une fois la prédisposition génétique à une maladie établie, l'étape suivante est d'essayer d'identifier les facteurs génétiques en cause. Cette recherche se base sur l'étude des variations du génome humain qui peuvent être détectées grâce à un nombre croissant de méthodes. Elle s'appuie sur une méthodologie statistique développée pour les maladies complexes. Ces dernières années, un intérêt croissant a été porté aux variants rares, dont l'étude nécessite des outils spécifiques.

II.2. Variabilité génétique humaine

Il existe plusieurs types de variants génétiques dans le génome humain (Figure 7). Les méthodes utilisées afin de les détecter sont différentes en fonction de leur type et de leur taille.

II.2.1. Types de variants

II.2.1.1. Les polymorphismes mono-nucléotidiques (single nucleotide polymorphism ou SNP)

Les SNPs sont aussi appelés SNV pour Single Nucleotide Variation ; il s'agit d'un polymorphisme de substitution au niveau d'un nucléotide (Figure 7A).

Lorsque cette substitution se fait dans une séquence codante, cela peut entraîner :

- Un changement synonyme c'est-à-dire une substitution de nucléotide qui n'entraîne

pas de changement d'acide aminé et donc pas d'impact sur la protéine.

- Un changement non synonyme faux sens. C'est lorsqu'il y a une substitution d'acide aminé qui peut entraîner ou non un variant pathogène en fonction de l'effet de la substitution d'acide aminé sur la fonction et la structure de la protéine.
- Un changement non synonyme non-sens. C'est lorsqu'il y a un changement d'acide aminé qui peut entraîner l'apparition d'un codon stop et par conséquent une protéine tronquée. Généralement, cette protéine est non fonctionnelle.

Les SNPs sont très fréquents dans le génome humain, car il y en a de l'ordre de 5×10^6 par individus. Plus de 1 milliard de SNPs ont déjà été recensés [51].

II.2.1.2. Les Indels

Le terme Indel veut dire Insertion – délétion. Il y a les petits et les grands Indels. Les petits Indels sont fréquents tandis que les grands Indels sont rares. Ici, nous parlons des petits Indels, car les grands Indels sont classés dans les variants structuraux. On parle de petit Indel lorsqu'il y a une insertion ou délétion dans le génome de 3 nucléotides maximum (Figure 7B). Ce phénomène peut entraîner un décalage du cadre de lecture, ce qui peut conduire à une traduction incorrecte.

Il y a en moyenne moins de 1×10^6 d'Indels dans le génome humain. Par exemple, 823 396 Indels ont été découverts dans le génome de Craig Venter, 222 718 Indels ont été découverts dans le génome de James Watson [52]. Au total, plus de 100 millions d'indels ont été recensés [51].

II.2.1.3. Les variants structuraux (Structural variants ou SV)

Les variants structuraux (SV) comprennent plusieurs sous-types de variants, à savoir les variants du nombre de copies ou CNV, les duplications, les grandes délétions et insertions, les inversions et translocations inter et intrachromosomiques (figure 7C). Plus de 7×10^6 de variants structuraux ont été recensés.

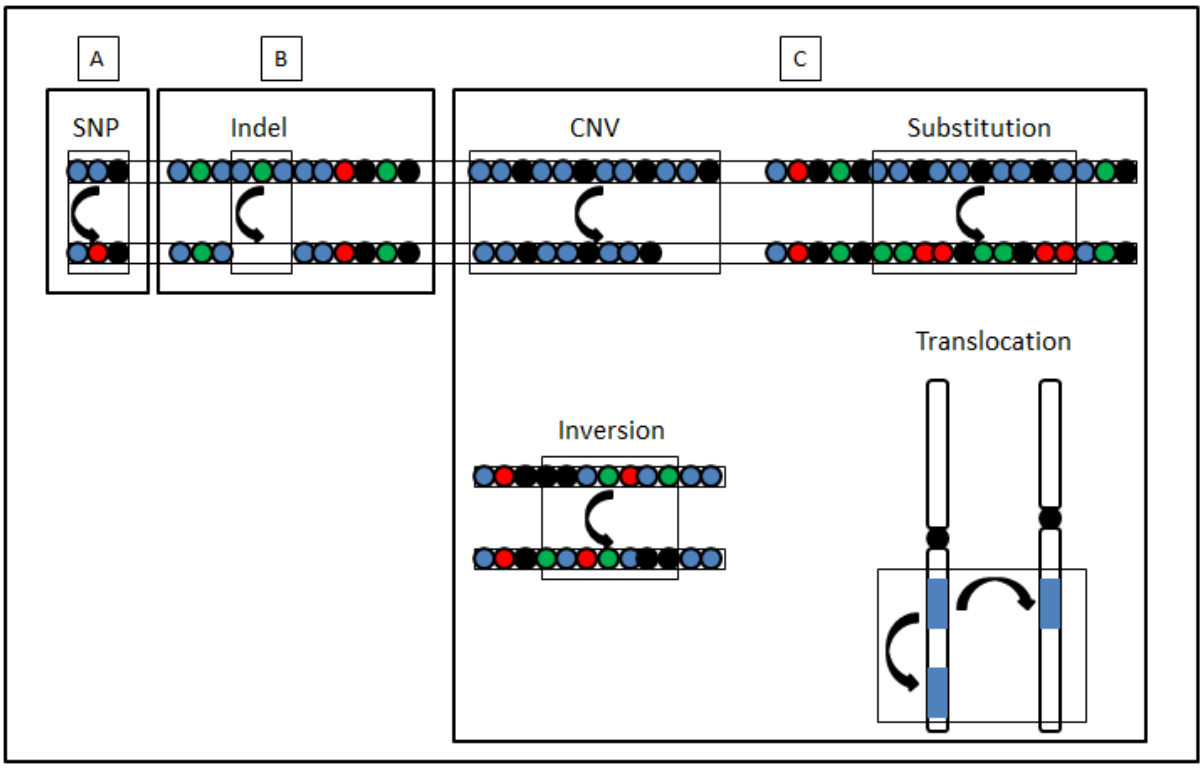

Figure 7. Les types des variants génétiques.

II.2.2. Détection des variants

Plusieurs techniques de génotypage ou de séquençage peuvent être utilisées pour détecter les variants (Tableau 5).

II.2.2.1. Génotypage simple marqueur

Polymorphisme de longueur des fragments de restriction ou RFLP

Cette technique a été utilisée pour la première fois par Botstein et al., pour la publication de la carte génétique du génome humain en 1980 [53]. Cette technique repose sur la digestion par une enzyme de restriction du produit de PCR d'un fragment comprenant le SNP à génotyper.

La présence ou l'absence du variant entraîne la présence ou l'absence de site de restriction, ce qui entraîne ensuite un polymorphisme de longueur de fragments (Figure 8).

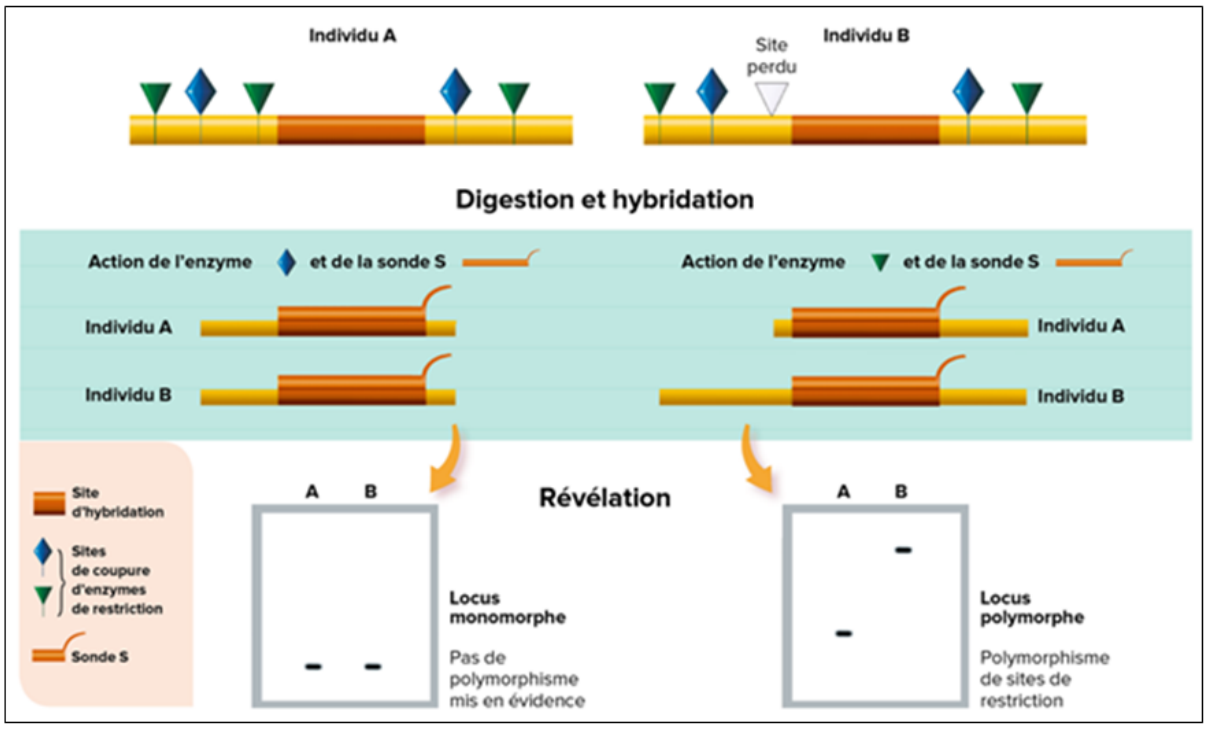

Figure 8. Principe de la RFLP (Adapté de GNIS-PEDAGOGIE.ORG).

TaqMan

La technologie TaqMan est basée sur l'activité exonucléase 5' -> 3' de la Taq polymérase, qui permet de dégrader tout oligonucléotide se trouvant sur son passage au cours de l'amplification de l'ADN (Figure 9) [54].

Cette technologie utilise une sonde spécifique d'un fragment d'ADN d'intérêt, dite sonde TaqMan, bi-marquée par deux fluorophores à chaque extrémité (un rapporteur en 5' et un absorbeur en 3'). L'absorbeur bloque la fluorescence du rapporteur aussi longtemps que les deux fluorophores sont liés à la sonde, car le spectre d'émission du rapporteur chevauche celui de l'excitation de l'absorbeur.

La dégradation de la sonde TaqMan par la Taq polymérase permet une augmentation de la fluorescence du rapporteur puisque les deux fluorophores ne sont plus liés. Cette fluorescence du rapporteur est proportionnelle au produit de l'amplification.

Pour le génotypage des variants, il suffit alors d'utiliser deux sondes spécifiques, l'une doit être complémentaire de l'allèle de référence et l'autre de l'allèle alternatif, mais aussi deux molécules chimiques qui fluorescent en deux couleurs différentes, une couleur spécifique pour chaque allèle.

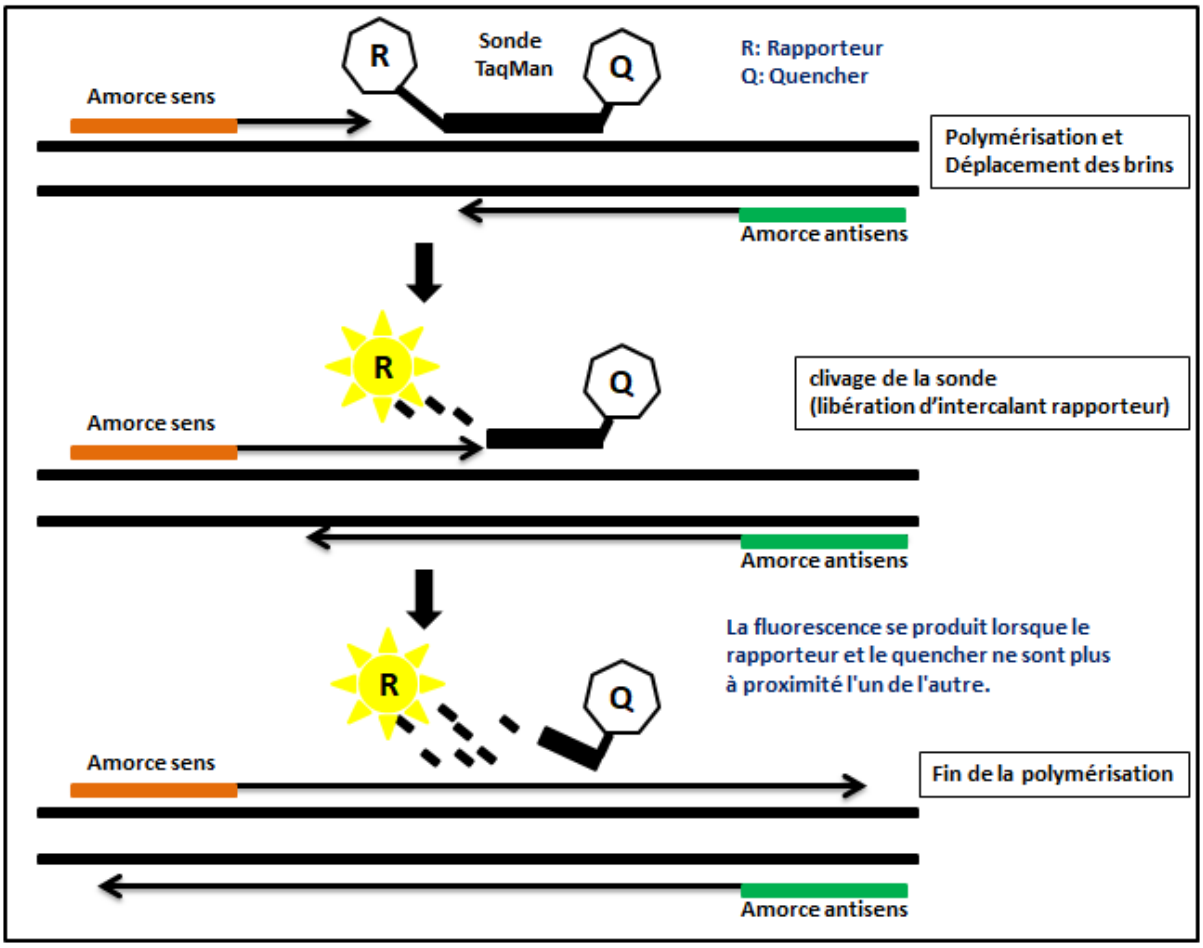

Figure 9. Principe de la technique TaqMan.

Courbe de fusion à haute résolution ou High Resolution Melt (HRM)

La courbe de fusion à haute résolution est une méthode d'analyse post-PCR utilisée pour identifier les SNPs [55]. Cette méthode est basée sur les techniques de courbe de fusion (dissociation) de la PCR et est rendue possible par la disponibilité des intercalants d'ADN double brin (ADNdb) (Figure 10). Ces intercalants ont une forte fluorescence lorsqu'ils sont liés à l'ADN et une faible fluorescence dans le cas inverse.

Pendant une PCR, lorsque l'ADNdb se dissocie en brins simples, l'intercalant est alors libéré, ce qui entraîne un changement de fluorescence. Le résultat est un profil de courbe de fusion caractéristique de l'amplicon.

Cette technique permet de distinguer les séquences d'ADN de chaque amplicon, car il y a variation de la courbe de fusion en fonction de leur séquence, de leur longueur, de leur teneur en GC ou de la complémentarité des brins.

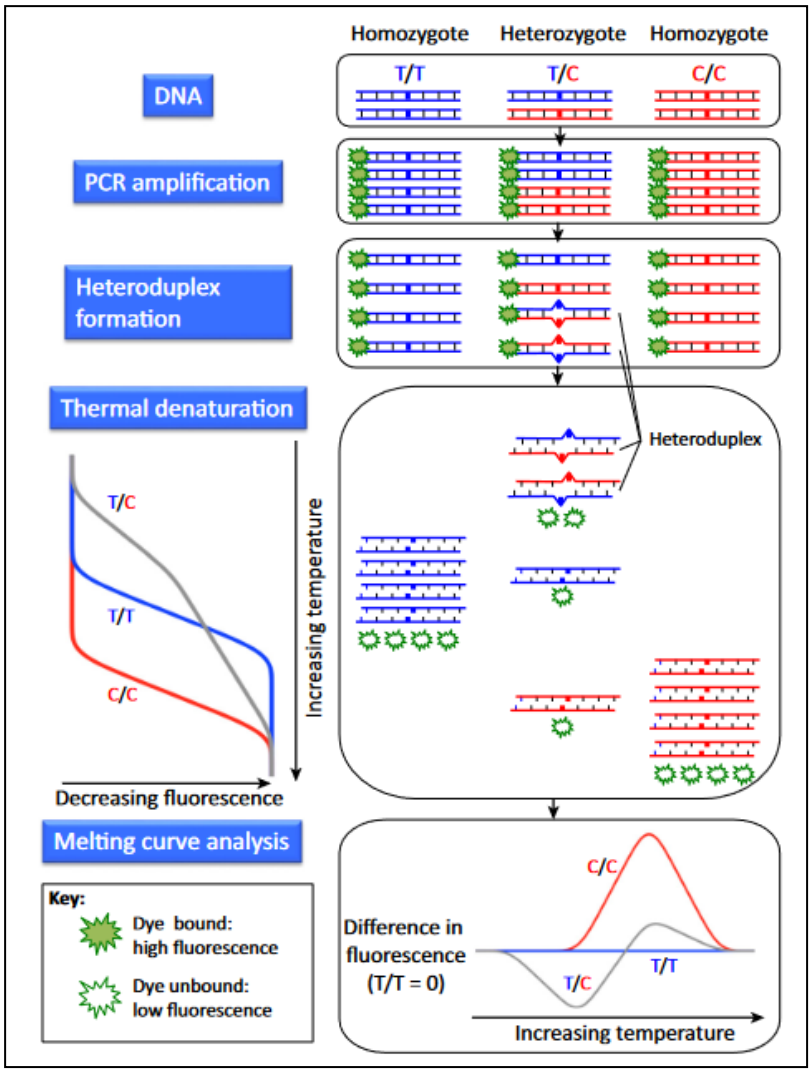

Figure 10. Principe de l'analyse de la fusion de l'ADN à haute résolution (HRM) [56].

II.2.2.2. Génotypage multi-marqueurs : microarrays

Les puces (microarrays)

La technologie des puces a pour principe l'hybridation de l'ADN à des oligonucléotides spécifiques des variations génétiques sur le génome entier (Figure 11).

Sur la surface de chaque puce, des centaines de milliers, voire des millions de génotypes d'un seul individu peuvent être analysés en même temps. Les variants génétiques analysés sur ces puces sont les SNPs et les Indels.

Dans cette technologie, l'ADN est amplifié et fragmenté, et ensuite hybridé de façon complémentaire à des sondes correspondantes aux variations interrogées. La spécificité allélique est déterminée grâce à une étape d'extension d'un nucléotide marqué. Un logiciel permet de déterminer le génotype d'un variant donné en fonction du ratio des couleurs détectées.

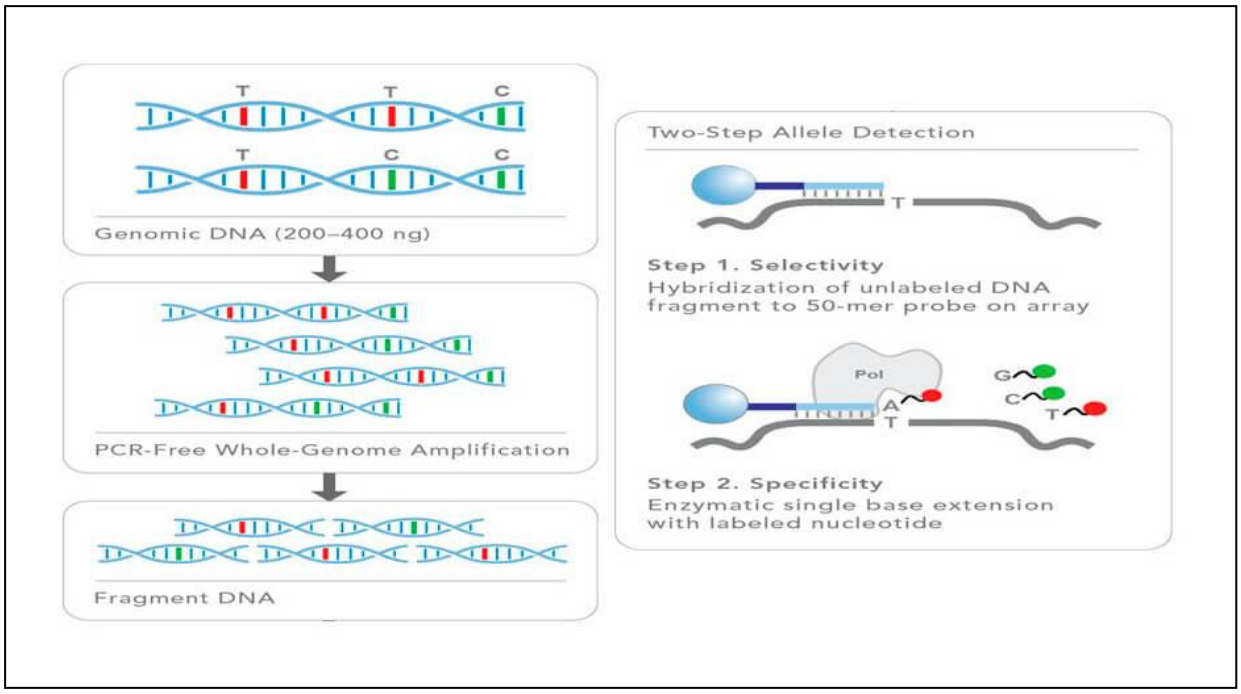

Figure 11. Principe des microarrays (Adapté de France génomique 2022).

II.2.2.3. Séquençage

Le séquençage de l'ADN consiste à déterminer l'ordre d'enchaînement des nucléotides pour un fragment d'ADN donné. L'évolution de la technologie de séquençage, de la première génération avec la technologie de Sanger à la troisième génération a permis de séquencer des génomes de plus en plus grands et en réduisant drastiquement le prix et le temps.

Première génération

En 1977, Sanger d'un côté, Maxam et Gilbert de l'autre, proposent deux approches différentes de séquençage de l'ADN. Celle développée par Sanger est basée sur la synthèse enzymatique tandis que celle de Maxam et Gilbert est basée sur la dégradation chimique [57,58]. La technologie de Sanger est la plus utilisée aujourd'hui [59].

Dans la méthode de Sanger, le principe est l'utilisation de la polymérisation de l'ADN à l'aide de l'ADN polymérase dépourvue d'activité exonucléase 5' – 3' et d'un petit oligonucléotide complémentaire à une partie du fragment d'ADN à séquencer.

Il faut ensuite ajouter les quatre désoxyribonucléotides (dNTPs) et une faible concentration de l'un des quatre didésoxyribonucléotides [60]. Ces quatre didésoxyribonucléotides (ddNTPs) sont liés à quatre différents fluorochromes, chacun de couleur différente (Figure 12).

Une fois qu'un de ces quatre didésoxyribonucléotides est incorporé dans le brin synthétisé, la

réaction s'arrête, car ces molécules sont dépourvues du radical hydroxyle sur le carbone 3' du ribose contrairement aux désoxyribonucléotides. Les différents fragments sont séparés par électrophorèse capillaire.

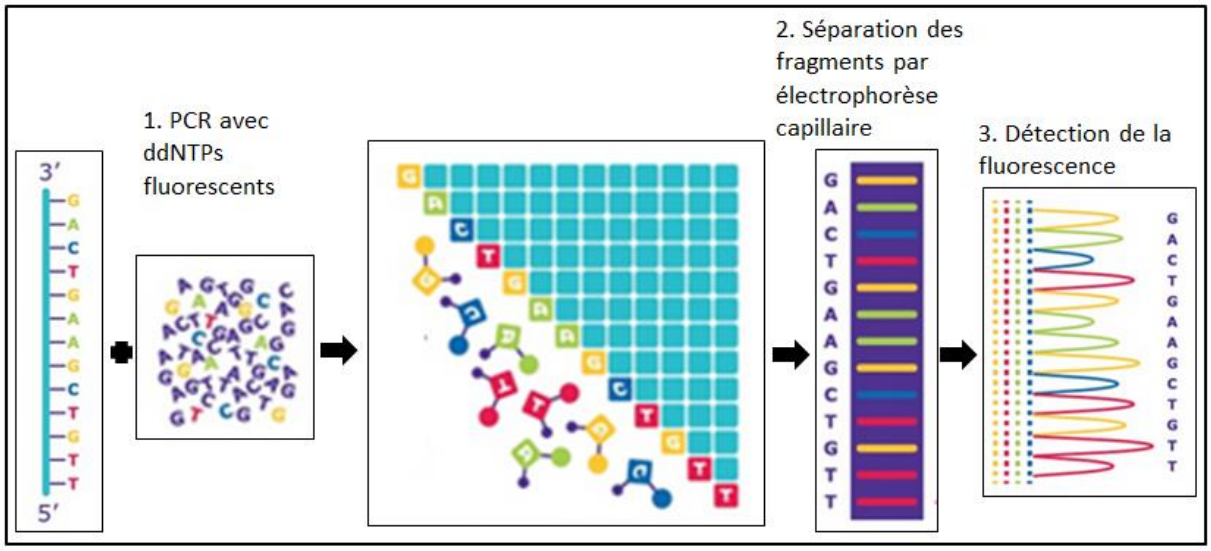

Figure 12. Principe du séquençage Sanger.

Deuxième génération

La deuxième génération de séquençage aussi appelée séquençage haut-débit ou séquençage de nouvelle génération est apparue au début des années 2000. C'est une révolution, car on séquence des centaines des millions voire des milliards des molécules en parallèle dans une chambre de séquençage.

Trois entreprises ont développé des technologies de seconde génération du séquençage : Roche, Life Technology et Illumina. Le séquenceur Ion Torrent développé par Life Technology et le 454 de Roche utilisent la réaction en chaîne par polymérase (PCR) en émulsion alors que les modèles d'Illumina utilisent l'amplification en phase solide [61]. Illumina domine actuellement ses concurrents avec 90 % de part de marché.

Dans la technologie d'Illumina, deux séquences adaptatrices (oligonucléotides) sont liées aux extrémités du fragment d'ADN génomique à séquencer. Ces deux séquences adaptatrices ont pour rôle l'accrochage et l'amplification sur la flow cell (cellule de flux) et l'hybridation d'une amorce qui sert de point de départ du séquençage (Figure 13).

Les brins complémentaires de ces deux séquences adaptatrices sont fixés sur la flow cell. Ensuite, Il y a hybridation d'un de deux oligonucléotides à son complémentaire fixé sur la flow cell et une amplification du fragment d'ADN par un processus d'amplification par pont.

Il y a formation de clusters, car on obtient plusieurs clones provenant d'un même fragment d'ADN (amplification clonale). Après la formation des clusters sur la flow cell, la réaction de séquençage est initiée par l'amorce de séquençage spécifique à la séquence en utilisant la technologie de séquençage par synthèse [62].

La méthode actuelle de la deuxième génération de séquençage s'adresse à des petits fragments.

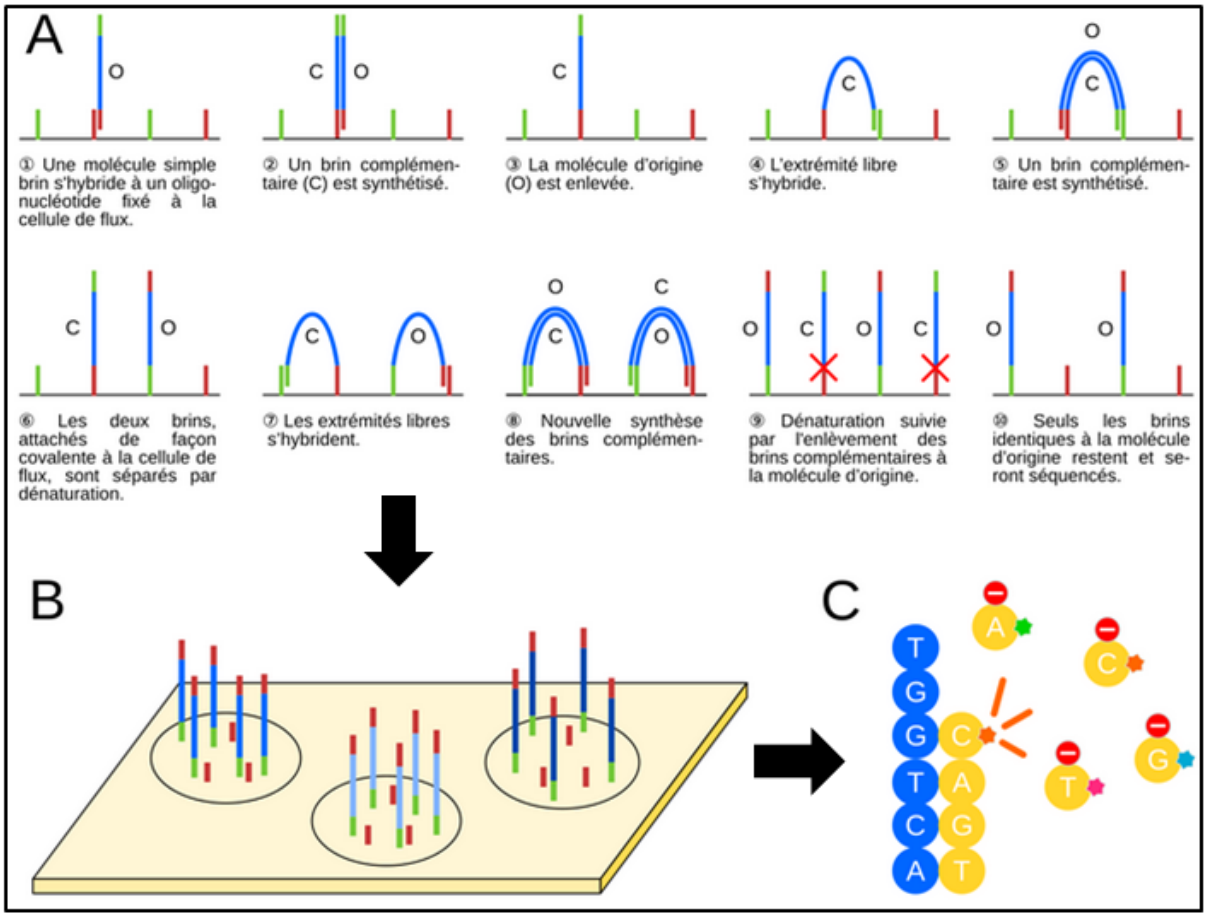

Figure 13. Principe du séquençage de deuxième génération (Adapté de planet-vie, 2022).

Troisième génération

La troisième génération de séquençage est en plein développement pour permettre la lecture des séquences d'ADN plus longues contrairement au NGS. Il y a la technologie SMRT (Single Molecule Real Time) développée par PacBio et celle d'Oxford Nanopore® basée sur les techniques de nanopores [63]. La nouveauté de cette génération de séquençage est la lecture de la molécule unique et sans amplification. Cette technologie permet aussi de séquencer des bases méthylées, en particulier des cytosines méthylées.

Dans la technologie SRMT de PacBio, une molécule d'ADN polymérase est fixée sur une puce

de séquençage pour synthétiser l'ADN et la séquence de nucléotides est déterminée en détectant le changement d'impulsion de fluorescence produit par chaque base en temps réel [64]. Cette fluorescence provient d'un fluorochrome qui est fixé au groupement phosphate de nucléotides.

La technologie ONT d'oxford utilise un tube artificiel appelé « nanopore » qui constitue un chemin de courant. Le flux de courant est interrompu lorsqu'un nucléotide passe dans un nanopore. Dans cette technologie, il n'y a pas d'utilisation de la fluorescence, mais le signal de chaque base est détecté par courant électrique, car chaque base produit un courant différent [65].

Cette technologie n'est pas encore très précise, car il y a encore un taux d'erreur élevé, mais elle reste très importante parce qu'elle permet l'identification des variations des grandes tailles.

II.2.3. Choix de la technique

Le choix de la technique dépend de plusieurs paramètres parmi lesquels le type de variant, les marqueurs détectés ou encore le coût. Le tableau 5 résume les principales caractéristiques de chaque technique de détection de variants.

Tableau 5. Les différentes techniques de génotypage et de séquençage.

	Technique	**Variants**	**Marqueurs**	**Coût**	**Bio-informatique**
Génotypage	RFLP	SNPs Indels	Simple	+	+
	HRM	SNPs Indels	Simple	+	+
	TaqMan	SNPs Indels	Simple	++	+
Microarrays		SNPs Indels CNV	Plusieurs	+++	++
Séquençage	Sanger	SNPs Indels	Plusieurs	+++	+
	Haut débit	SNPs Indels SVs	Plusieurs	++	+++++

Il faut noter qu'on peut identifier les nouveaux variants avec le séquençage alors que ce n'était pas le cas avec les techniques précédentes.

II.3. Identification des facteurs génétiques de susceptibilité des maladies complexes

Historiquement, la recherche des facteurs génétiques de susceptibilité des maladies complexes s'est basée sur les analyses de liaison, développées initialement dans les maladies mendéliennes. Dans un second temps, les études d'association ont été développées pour les maladies complexes.

II.3.1. Les études de liaison génétique

Le terme de liaison fait référence à la tendance des gènes situés à proximité sur le même chromosome à être hérités ensemble. Plus précisément, on dira que deux locus sont génétiquement liés lorsque la fraction de recombinaison méiotique entre eux est inférieure à 50 %.

Les études de liaison génétique ont été d'abord utilisées pour positionner les gènes sur les chromosomes les uns par rapport aux autres et établir ainsi des cartes de liaison génétique. Secondairement, elles ont été utilisées afin de déterminer l'emplacement des gènes prédisposant à une maladie par rapport à des marqueurs génétiques déjà connus (Figure 14). L'analyse de liaison était historiquement la première étape de l'étude génétique d'un trait, car elle permet de cerner des régions génomiques susceptibles de contenir un gène contrôlant ce trait [66]. On pouvait alors rechercher le gène de susceptibilité dans la région, en particulier par des méthodes de clonage positionnel.

Il faut noter que si l'analyse de liaison est un outil puissant pour trouver les gènes impliqués dans les maladies mendéliennes, elle produit souvent des signaux faibles et parfois incohérents dans les études de maladies complexes. Les études de liaison donnent de meilleurs résultats lorsque la pénétrance des allèles de risque est forte [67].

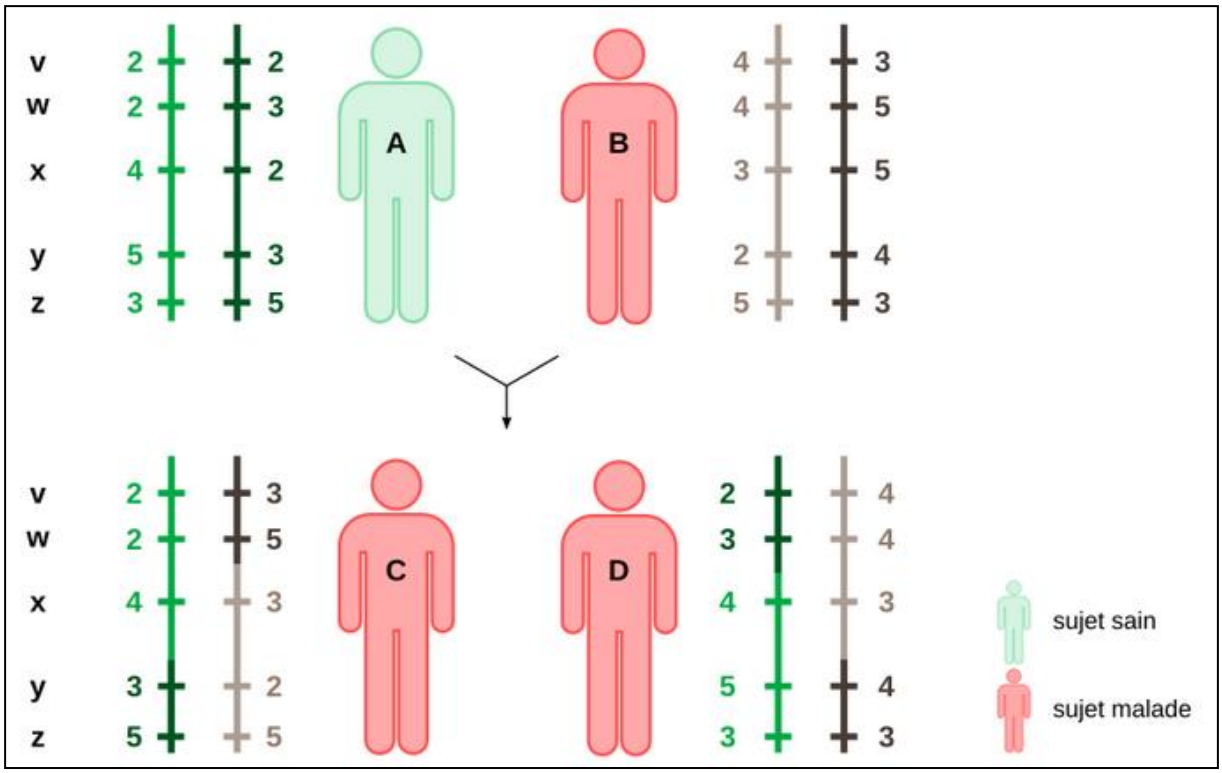

Figure 14. Principe des analyses de liaison génétique (Adapté de planet-vie, 2022). Deux parents A (saint en vert) et B (malade en rouge), et 5 locus bialléliques (v, w, x, y et z). Chez les enfants malades (C et D), on observe une recombinaison entre les locus w et x, et entre les locus x et y. Au niveau du locus x, les allèles 4 et 3 sont hérités ensemble chez les deux enfants malades.

Selon le besoin de spécifier un modèle de transmission de la maladie, il existe deux méthodes d'analyses de liaison génétique : les analyses de liaison paramétriques et non paramétriques.

II.3.1.1. Les analyses de liaison paramétrique

Dans les analyses de liaison paramétrique, la liaison est généralement indiquée sous la forme d'un score de logarithme des chances dit LOD score.

Le LOD score

En 1955 Morton et al. ont proposé ce score [66] basé sur la fraction de recombinaison (ɵ) ou la position chromosomique mesurée en centimorgan ou cM. Il faut spécifier un modèle de transmission de la maladie pour calculer le LOD score.

Le modèle prend en compte la prévalence de la maladie et son mode de transmission (dominant ou récessif) avec la pénétrance de chacun des génotypes.

L'objectif ultime de l'analyse est d'estimer la fraction de recombinaison entre chaque marqueur et le locus de la maladie (deux points) ou la position du locus de la maladie par

rapport à une carte connue de marqueurs (multipoint) (Figure 15).

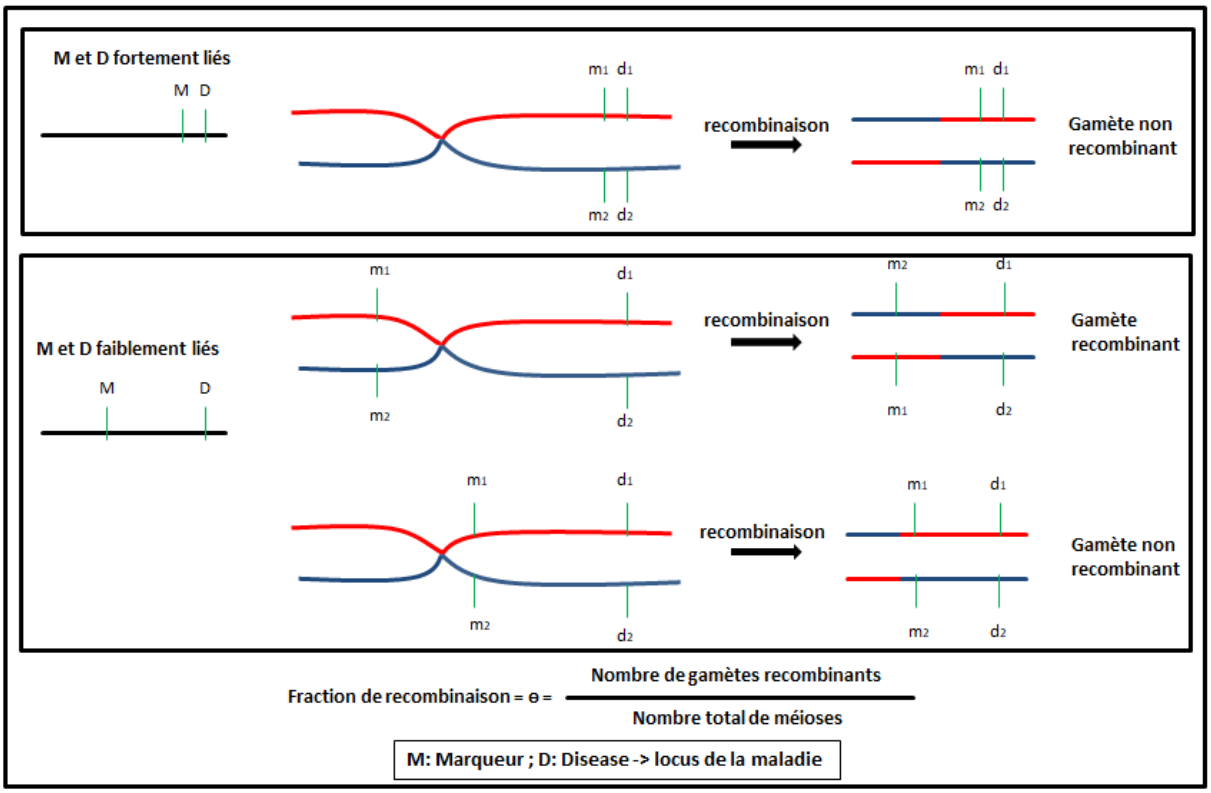

Figure 15. Liaison génétique et fraction de recombinaison. Pour qu'il y ait liaison, il faut que la fraction de recombinaison soit inférieure à 50 %.

Plus le score LOD est élevé, plus importante est la preuve de la liaison.

Traditionnellement, les seuils suivants sont utilisés :

LOD score ≥ 3 : liaison entre le marqueur testé et le locus maladie
LOD score < -2 : absence de liaison sous le modèle de la maladie testé
LOD score entre -2 et 3 : impossible de conclure à la présence ou non d'une liaison

II.3.1.2. Les analyses de liaison non paramétrique

Comme précisé ci-dessus, les analyses de liaison paramétrique nécessitent la spécification d'un modèle de transmission de la maladie. Or ce modèle de transmission est rarement connu dans les maladies complexes du fait que plusieurs gènes peuvent contribuer à la susceptibilité à la maladie, chacun avec des pénétrances faibles.

C'est pourquoi ont été développées des méthodes non paramétriques ou sans modèle. Le raisonnement est le suivant : on s'attend à ce qu'il y ait un partage excessif d'haplotypes identiques par descendance (IBD) dans la région d'un gène responsable de la maladie entre les parents et les malades, quel que soit le mode d'héritage. Diverses méthodes permettent de tester si le partage de l'IBD à un locus est plus important que prévu sous l'hypothèse nulle

d'absence de liaison.

Méthode des paires de germains atteints

L'analyse des paires de frères et sœurs affectés est devenue courante depuis qu'il a été démontré que, sous des hypothèses très spécifiques, ces méthodes constituent un modèle puissant pour l'analyse de liaison [68].

Cette méthode des paires de germains atteints consiste à comparer, chez deux germains atteints de la même maladie, le nombre d'allèles du marqueur génétique transmis en même temps par le père et la mère. Si le marqueur étudié est proche du locus maladie, les paires de germains atteints présenteront un excès d'allèles du marqueur en commun.

En l'absence de liaison, les probabilités que les deux germains atteints partagent 0, 1 ou 2 allèles identiques parentaux sont respectivement de 0,25, 0,5 et 0,25.

Les analyses de liaison ont permis d'identifier les gènes impliqués dans beaucoup de maladies mendéliennes, mais aussi impliqués dans les maladies complexes comme dans les cancers (BRCA1 et 2, PAF, HNPCC), la maladie d'Alzheimer (APP, présélinine 1 et 2) et le diabète (MODY 1, 2 et 3). Tous ces gènes impliqués dans les maladies complexes et identifiés par des analyses de liaison sont ceux qui présentent une faible fréquence d'allèles et une hérédité mendélienne ou quasi-mendélienne [69]. La limite des études de liaison est qu'elles sont inopérantes quand la pénétrance des allèles de risque est faible.

II.3.2. Les analyses d'association génétique

Les études d'association génétique ont pour but de détecter une association entre un ou plusieurs polymorphismes génétiques et une maladie ou un trait (Figure 16).

L'analyse d'association a une plus grande puissance que les études de liaison pour détecter les petits effets, mais elle nécessite l'examen de beaucoup plus de marqueurs [70].

Il est possible de rechercher les gènes de susceptibilité aux maladies en faisant un criblage d'un grand nombre de SNP sur l'ensemble du génome, et c'est ce qu'on appelle les *GWAS* ou les études d'associations génome entier [71].

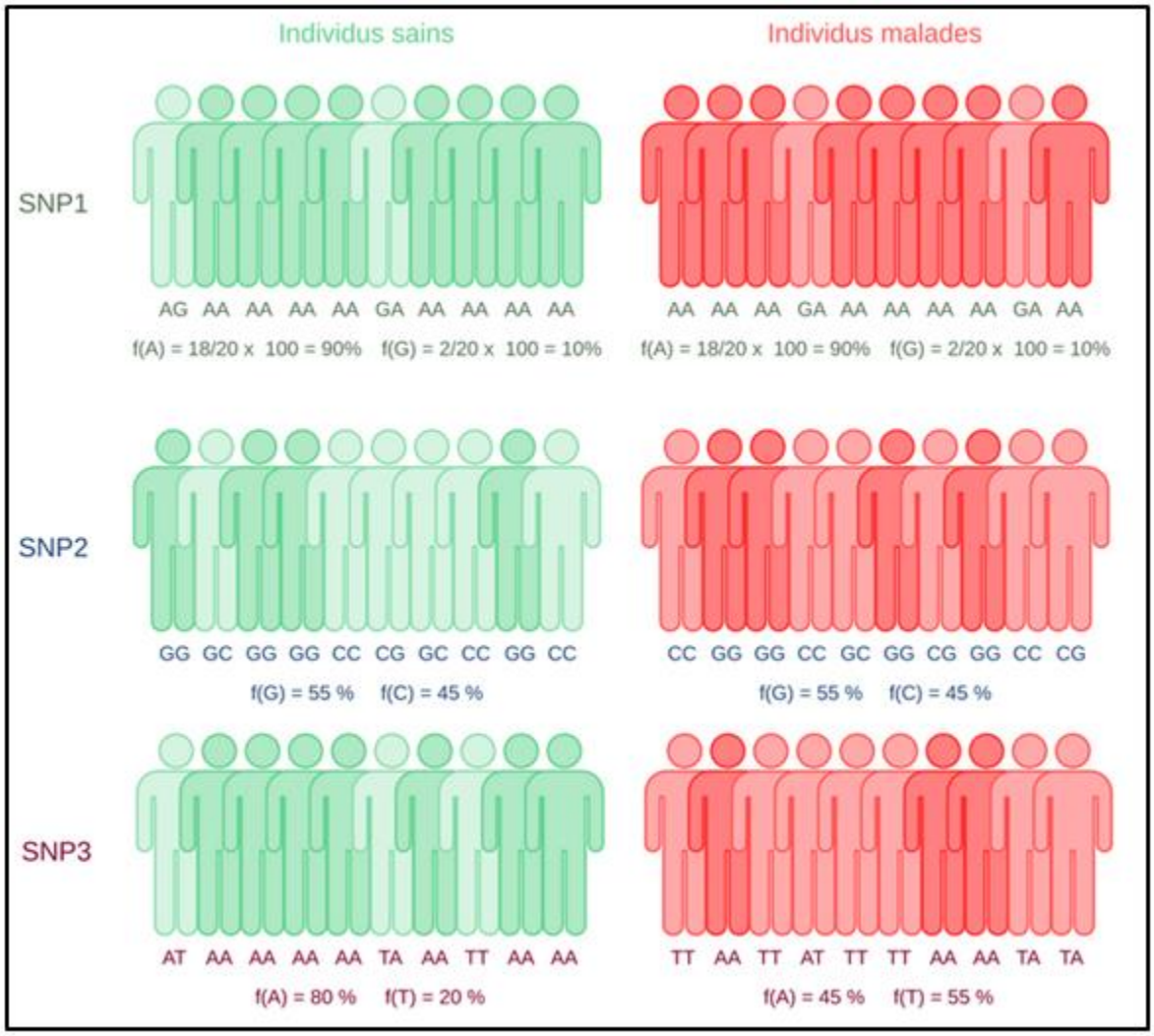

Figure 16. Principe des études d'association génétique (Adapté de planet-vie, 2022).

II.3.2.1. Le déséquilibre de liaison

Le déséquilibre de liaison (DL) est l'association non aléatoire d'allèles à différents locus génétiquement liés et donc deux allèles sont dits associés s'ils sont présents ensemble sur le même chromosome (ou sur le même haplotype) plus souvent que ne le voudrait le hasard. Autrement dit, la fréquence de leur haplotype est différente du produit des fréquences de chacun des allèles.

Ce phénomène est influencé par plusieurs facteurs, notamment le taux de recombinaison, la sélection naturelle, la dérive génétique, le taux de mutation ou encore les facteurs démographiques.

Il existe plusieurs méthodes qui permettent de tester le DL et d'en évaluer la force. La plupart d'entre elles mesurent le DL entre des paires de marqueurs bialléliques. On peut citer les coefficients |D'| et r^2 qui varient de 0, correspondant à l'équilibre de liaison, à 1,

correspondant à un DL complet (Figure 17) [72,73].

Soit une paire de locus A et B présents sur un même chromosome présentant chacun deux allèles 1 et 2.
Les mesures DL, |D'| et le coefficient r^2 se calculent de la manière suivante :

$$DL = f(A_1B_1) - f(A_1) * f(B_1) \qquad |D'| = \frac{DL = f(A_1B_1) - f(A_1) * f(B_1)}{Dmax = f(A_1) - (f(A_1) * f(B_1))} \qquad r^2 = \frac{DL^2 = [f(A_1B_1) - f(A_1) * f(B_1)]^2}{f(A_1) * f(A_2) * f(B_1) * f(B_2)}$$

Où f = fréquence
Dmax = valeur maximale possible du DL

Figure 17. Calcul des mesures DL, |D'| et du coefficient r^2.

Les mesures DL et |D'| sont plutôt utilisées dans des études de populations pour estimer l'évolution des haplotypes alors que le r^2 est beaucoup utilisé en génétique des maladies pour estimer la corrélation entre les variants.

II.3.2.2. Les tags SNPs

On peut capter le signal d'association des SNPs causaux associés aux maladies sans passer par le génotypage de chaque SNP individuellement dans une région chromosomique. Cela permet de réduire les coûts en termes de temps et surtout de dépenses nécessaires à la cartographie des régions du génome associées aux maladies.

Un tag SNP est un SNP représentatif d'une région du génome présentant un déséquilibre de liaison élevé et représentant un groupe de SNP appelé haplotype. Ainsi, le génotypage d'un seul SNP permet de tester l'association entre un trait et tous les SNPs en fort déséquilibre de liaison à celui-ci. L'avantage de cette stratégie est la diminution du nombre de marqueurs à génotyper pour obtenir une couverture pangénomique. Sa limite est que, très souvent, une fois le signal d'association mis en évidence, il est difficile de déterminer avec certitude quel est le variant causal du fait du DL parfois important.

II.3.2.3. Les tests d'association

Deux designs sont possibles pour les études d'association avec des tests statistiques différents en fonction de ce design : les études cas-contrôles et les études intra-familiales.

Les études cas-contrôles

Les études cas-contrôles sont des études dans lesquelles les individus malades ou ayant un trait particulier (les cas) sont sélectionnés pour être comparés à d'autres individus pour qui ce

trait particulier est absent (les contrôles) [74]. Point essentiel, tous les participants à l'étude doivent être non apparentés et provenir de la même population.

Dans ces études, le but est de comparer la fréquence des variants alléliques entre les groupes d'individus.

Il est essentiel de recourir à des tests statistiques afin de déterminer s'il existe une différence significative des fréquences alléliques chez les malades et chez les contrôles.

Les tests d'association génétique sont généralement réalisés séparément pour chaque SNP individuel. Les données pour chaque SNP avec l'allèle mineur a et l'allèle majeur A peuvent être représentées sous la forme d'un tableau de contingence du nombre de génotypes (par exemple, a/a, A/a et A/A) ou du nombre d'allèles (par exemple, a et A) chez les cas et les contrôles.

On s'attend à ce que les fréquences relatives des allèles ou des génotypes soient les mêmes chez les cas et les contrôles sous l'hypothèse nulle (absence d'association). Un test d'association est donc donné par un simple test Chi2 de Pearson ou de Fisher pour l'indépendance des lignes et des colonnes du tableau de contingence [75].

Si le test d'association est positif, on rejette l'hypothèse nulle de non-association et on peut alors calculer l'odds ratio et l'intervalle de confiance. Si l'odds ratio est supérieur à 1, on dit que l'allèle est prédisposant et si l'odds ratio est inférieur à 1, on dit que l'allèle est protecteur.

Des modèles d'association par régression logistique plus complexes sont utilisés pour les traits binaires lorsqu'il est nécessaire d'inclure des covariables supplémentaires pour traiter les maladies complexes. Il s'agit par exemple de situations où l'on s'attend à ce que le risque de maladie soit modifié par des effets environnementaux tels que les facteurs de risque épidémiologiques (le tabagisme et le sexe,…), les variables cliniques (la gravité de la maladie et l'âge d'apparition,…) et la stratification de la population ou par les effets interactifs et conjoints d'autres loci marqueurs [75].

Les études d'association intra-familiale

En 1993 Spielman et al. ont conçu le test de déséquilibre de transmission ou TDT. Ce test est basé sur la transmission des allèles d'un marqueur des parents à leur enfant malade, lorsque le locus marqueur et le locus de la maladie sont liés et se trouvent en déséquilibre de liaison [76]. Sous l'hypothèse nulle de non-association, la transmission des allèles est mendélienne

et équilibrée. L'intérêt de cette méthode est qu'elle est robuste aux stratifications de populations, car dans chaque famille, l'allèle non transmis constitue un contrôle de l'allèle transmis.

Le TDT est une méthode standard pour détecter l'association à l'aide d'un plan de trio familial et consiste à comparer les fréquences attendues aux fréquences réellement observées (Figure 18). Des extensions du TDT ont été proposées pour prendre en compte les informations de transmission des parents hétérozygotes à leur descendance non malade, pour les traits quantitatifs et les marqueurs multialléliques par exemple [77–79].

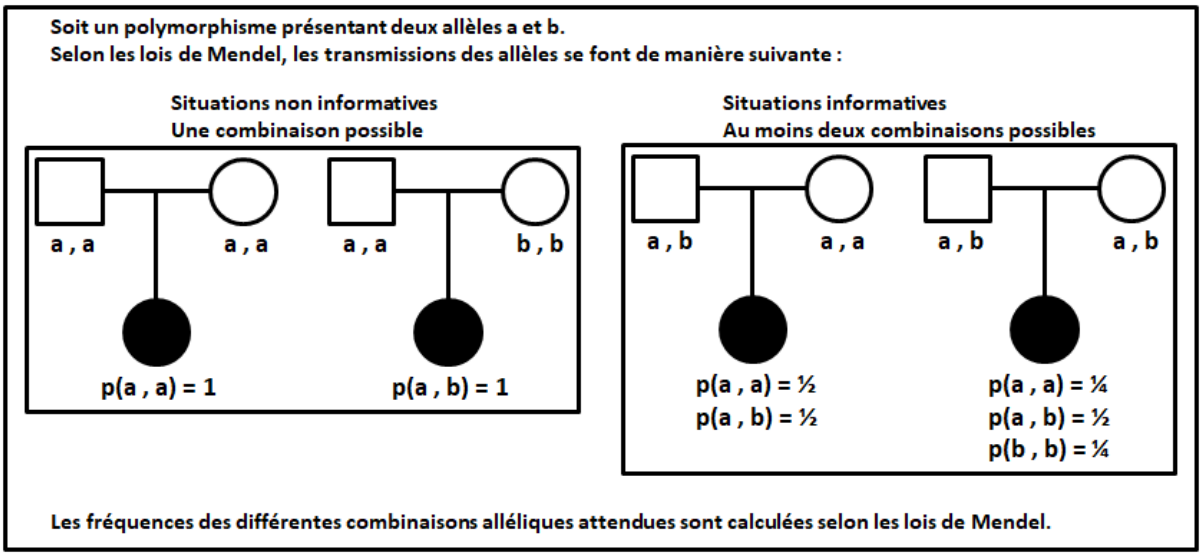

Figure 18. Principe du TDT.

II.3.2.4. Deux grands types d'approche

Gène candidat

Dans cette approche, on choisit les gènes soupçonnés d'être impliqués dans la maladie en se basant sur des connaissances préalables, par exemple physiopathologiques ou cartographiques. On sélectionne les polymorphismes à tester pour leur appartenance à un gène choisi en fonction de son rôle présumé dans la maladie étudiée. On peut aussi se baser sur les études de liaison génétique préalables suggérant la présence d'un facteur de susceptibilité dans une région génétique donnée [80].

La limite de cette approche est l'exclusion des gènes dont le rôle est inconnu ou dont l'implication dans la maladie est insoupçonnable. Pour remédier à ces limites ou à l'absence d'hypothèses préalables, des approches génome-entier ont été développées.

Les études d'association génome entier ou GWAS

Les études d'association à génome entier (GWAS) ont été une révolution dans le domaine de la génétique des maladies complexes, car elles permettent de tester des millions de *SNP* dans le génome de nombreux individus afin d'identifier les associations génotype-phénotype.

Dans ces études, les SNP utilisés sont généralement fréquents, mais sont filtrés pour éviter la redondance de l'information en sélectionnant les tags SNP ou SNP marqueurs dans les zones de déséquilibre de liaison élevé [81].

Ces études ont beaucoup de succès, car elles ont permis l'identification de nombreux gènes de susceptibilité dans plusieurs maladies et traits comme la SpA, les troubles psychiatriques, les maladies auto-immunes, le diabète, … [82,83].

Bien que les GWAS aient été une révolution, les variants mis en évidence par ces études sont rarement les variants causaux, car leur association est le plus souvent indirecte. Surtout ils n'expliquent qu'une petite partie de l'héritabilité [84,85].

II.3.2.5. Limites des GWAS

Il y a plusieurs limites des GWAS [85].

Les GWAS dépendent de l'exhaustivité des panels de référence préexistants.

Les GWAS dépendent de l'exhaustivité des études de séquençage et des panels de référence qui en résultent, car ils sont utilisés pour la conception des puces de génotypage [86]. Les premières puces ont été conçues en sélectionnant des tags SNP à partir de panels de référence de populations essentiellement européennes [87]. Ces réseaux offraient souvent une couverture médiocre dans les populations non européennes, car les modèles de DL varient d'une population à l'autre [87].

Les GWAS sont pénalisés par la nécessité de correction pour tests multiples.

Dans les GWAS, il est nécessaire d'utiliser une correction de Bonferroni pour maintenir à 5 % le taux de faux positifs à l'échelle du génome. Pour cela, on se base sur l'hypothèse d'un million de tests indépendants pour une variation génétique commune. Pour être considérés comme significatifs, les signaux d'association doivent donc atteindre un seuil de $P < 5 \times 10^{-8}$ [88]. Afin d'obtenir la puissance nécessaire, un nombre très important de patients et de témoins doit être recruté et ce d'autant plus que l'influence du variant sur la susceptibilité à la maladie est faible.

Les GWAS n'expliquent qu'une modeste fraction de l'héritabilité.

Les GWAS ont permis d'identifier de variants génétiques associées aux maladies et traits communs, mais ces variants ne représentent qu'une proportion modeste de l'héritabilité estimée de la plupart des traits complexes [88].

Dans les maladies inflammatoires chroniques de l'intestin (MICI), les GWAS ont permis d'identifier plus 215 loci de risque, dévoilant ainsi de multiples voies impliquées dans la pathogenèse des MICI, mais seulement 20 – 25 % de l'héritabilité des MICI a été expliquée par les GWAS [89–91].

Les GWAS ne permettent pas nécessairement d'identifier les variants causaux

La corrélation au niveau local de multiples variants génétiques due au déséquilibre de liaison facilite l'identification initiale d'un locus, mais rend difficile l'identification du ou des variants causaux [92]. Le principe des tags SNPs, propre à la plupart des GWAS fait qu'il est probable que le variant causal n'ait pas été génotypé.

Afin d'identifier les variants causaux, les méthodes dites de "cartographie fine" ont été développées. Elles tentent de hiérarchiser les variants causaux putatifs pour des études de suivi fonctionnel [93].

II.3.3. Concept d'héritabilité manquante

L'héritabilité est la quantification de la variation phénotypique globale qui est attribuable aux facteurs génétiques. Les études des jumeaux MZ et DZ, les études d'adoption et les études familiales permettent de calculer l'héritabilité [94]. Les facteurs génétiques identifiés par les GWAS ne représentent qu'une fraction relativement faible de l'héritabilité des maladies complexes. Par conséquent, une part importante de l'héritabilité des traits complexes est inexpliquée ("héritabilité manquante"). Plusieurs hypothèses ont été avancées pour permettre d'expliquer où se situe cette héritabilité manquante.

II.3.3.1. Surestimation de l'héritabilité des traits complexes

L'héritabilité, définie comme la proportion de la variance phénotypique dans une population attribuable à des facteurs génétiques additifs, est généralement estimée à partir d'études familiales, et on peut s'attendre à ce qu'elle varie selon les environnements.

Les estimations de l'héritabilité chez l'homme peuvent être surestimées si la ressemblance familiale est influencée par des effets génétiques non additifs (épistasie ou interaction gène-

gène), par des environnements familiaux partagés et par des corrélations ou des interactions entre les génotypes et l'environnement.

II.3.3.2. Sous-estimation des effets des allèles identifiés par les GWAS

Les contributions des variants déjà identifiés sont peut-être sous-estimées, car les risques relatifs observés pour les SNPs marqueurs peuvent sous-estimer les risques réels associés aux véritables variants causaux. Par exemple, 11 des 30 gènes porteurs de variants communs associés aux taux de lipides sont également porteurs d'allèles rares connus à effet important identifiés dans les dyslipidémies mendéliennes, notamment ABCA1, PCSK9 et LDLR, ce qui suggère que les gènes contenant des variants communs aux effets modestes sur des traits complexes peuvent également contenir des variants rares aux effets plus importants [95].

II.3.3.3. Interaction gène-gène ou gène-environnement

Les études des interactions gène-gène et gène-environnement peuvent aider à découvrir les mécanismes causant une maladie, en démontrant quels facteurs fonctionnent ensemble pour augmenter ou diminuer le risque. Ces interactions peuvent expliquer une partie de l'héritabilité manquante dans les maladies complexes.

II.3.3.4. Epigénétique

Les processus épigénétiques et les ARNs non codants, dont principalement les microARN (miARN), sont des acteurs importants et largement inexplorés de la régulation des gènes qui pourraient expliquer une partie de l'héritabilité manquante.

II.3.3.5. Variants rares

Les variants rares peuvent avoir des effets considérables (multipliant par deux ou trois le risque de maladie) et peuvent contribuer de manière substantielle à l'héritabilité manquante [96]. En raison de la taille insuffisante des échantillons conduisant à un manque de puissance, il existe actuellement relativement peu d'exemples de tels variants contribuant à des traits complexes.

II.3.3.6. Variants structuraux

Les SVs peuvent aussi expliquer une partie de l'héritabilité inexpliquée. Bien que ce type de variation n'ait pas été explicitement examiné dans la plupart des études d'association pangénomiques jusqu'à présent, les CNV en particulier ont attiré l'attention à mesure que les

méthodes pour les détecter se sont améliorées [94]. D'autres SVs tels que les inversions, les translocations, les répétitions microsatellites et les réarrangements ont été impliqués dans des maladies mendéliennes rares, mais ces variations ont été largement inexplorées en relation avec des traits complexes [95].

Les recherches réalisées durant ma thèse se sont principalement focalisées sur l'étude de l'implication de variants rares dans les maladies complexes, raison pour laquelle cette implication et les méthodes pour la détecter sont détaillées dans les paragraphes suivants.

II.4. Les variants rares

La prédisposition génétique aux maladies complexes a initialement été supposée principalement en lien avec des variants fréquents. Ainsi l'hypothèse Maladie fréquente - variants fréquents ("Common Disease - Common Variants") (CDCV) a été proposée, en opposition à la situation dans les maladies rares associées aux variants rares (« Rare Disease - Rare Variant) [97].

Cette hypothèse a trouvé ses limites dans la part de l'héritabilité des maladies complexes restant inexpliquée après l'identification des milliers de variants communs associés aux maladies complexes.

Pour combler cette héritabilité manquante dans les maladies complexes, l'hypothèse "Common Disease - Rare Variants" (CDRV) a été proposée postulant que les maladies communes peuvent être dues à de multiples variants rares avec une pénétrance relativement élevée [98] (Figure 19).

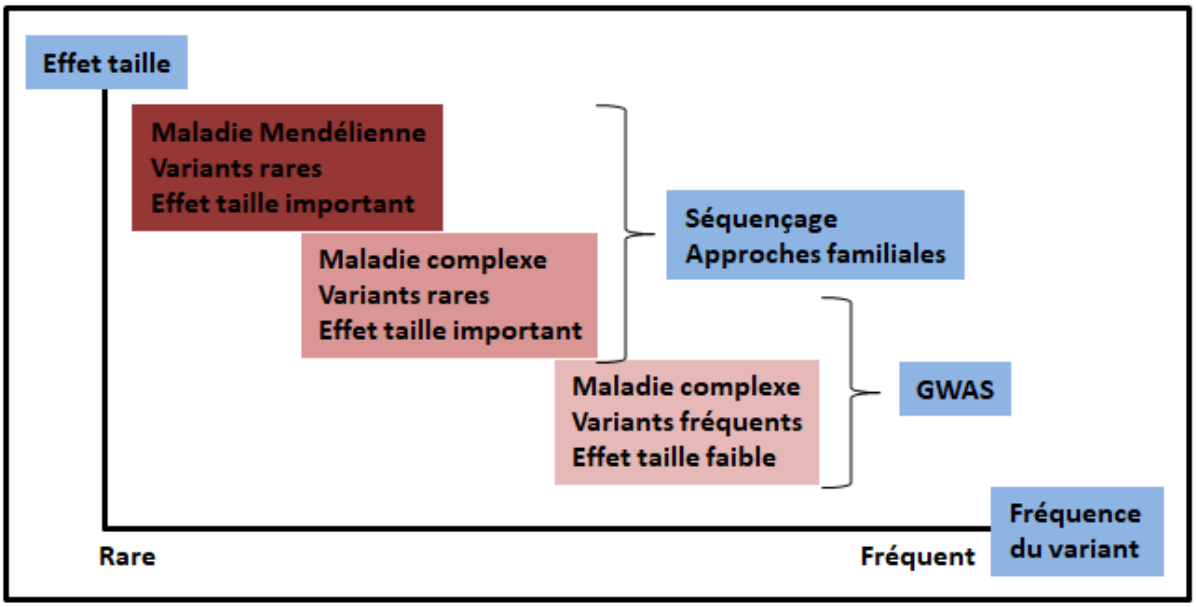

Figure 19. Contribution de variants à la susceptibilité des maladies.

La réduction significative du coût du séquençage de l'exome entier et du génome entier au cours des dernières années a permis de réaliser des expériences de séquençage à grande échelle et a facilité la détection des variants rares.

II.4.1. Implication des variants rares dans les maladies complexes

Les variants rares sont connus pour jouer un rôle important dans les maladies, car c'est parmi les variants rares qu'on trouve majoritairement des variants délétères (qui affectent, décalent ou interrompent le cadre de lecture) associés à une pathologie mendélienne.

La théorie de l'évolution prédit que les allèles délétères restent rares en raison de la sélection purifiante [99,100] [101,102].

Les maladies mendéliennes ne sont pas les seules à être dues à des mutations rares et à forte pénétrance. En effet, de nombreuses maladies chroniques ont des formes familiales en lien avec des variants rares : on peut citer les variants rares responsables du diabète de type 2 (MODY) et les mutations des gènes BRCA1 et BRCA2 à l'origine d'une susceptibilité au cancer du sein [103,104]. Des variants rares ont également été impliqués dans des formes sporadiques de maladie complexe, tels que des variants de NOD2/CARD15 impliqués dans la maladie de Crohn [105].

II.4.2. Design d'études pour les variants rares

Afin de découvrir les gènes impliqués dans les maladies complexes, d'autres plans d'étude et

méthodes statistiques sont nécessaires que ceux utilisés dans les GWAS, car une importante hétérogénéité génétique est attendue dans l'hypothèse Common Disease - Rare Variant [88]. De plus, la rareté des variants recherchés nécessite de très larges effectifs pour obtenir une puissance satisfaisante lorsque l'on applique un design cas-témoins classique. Certains designs pourraient cependant permettre d'obtenir la puissance requise tout en limitant la taille de l'échantillon.

II.4.2.1. Isolats de population

Le pouvoir de détection d'une association peut être élevé si les variants responsables des maladies sont présents à une fréquence élevée dans un isolat d'une population. Les isolats de populations peuvent donc être très utiles pour découvrir des associations de variants rares aux maladies complexes, car le pouvoir de détecter une association est en partie déterminé par la fréquence des allèles [106].

En raison de diverses forces démographiques (par exemple, les épidémies, la famine, la guerre, les migrations), de nombreuses sous-populations dans le monde sont devenues isolées et le sont restées pendant de nombreuses générations, conduisant à une faible variabilité génotypique. A cause de cette diversité génétique réduite et de la dérive génétique accrue (due à l'isolement), les populations isolées présentent souvent des fréquences alléliques très différentes de celles de populations non isolées. De plus, les isolats de population présentent souvent une homogénéité environnementale et culturelle qui peut être avantageuse pour une étude d'association.

Un exemple typique est celui de la Finlande, où des isolats de population sont apparus au sein d'une même région géographique en raison des migrations internes [107].

Un autre exemple est les Indiens Pima d'Arizona qui ont une prévalence très élevée de diabète de type 2 (~20%) [108] et une quasi-absence de diabète de type 1 [109].

II.4.2.2. Échantillonnage de phénotypes extrêmes

Un phénotype est un ensemble de signes, de symptômes et de caractéristiques comportementales observables associés à une maladie. Il est décrit par une variable qui peut être catégorique (sexe masculin ou féminin) ou quantitative (taux de glucose ou seuils d'audition).

La variation phénotypique des caractères quantitatifs peut être représentée par un graphique en forme de cloche où les phénotypes moins graves et ceux plus graves sont situés aux

extrémités de la distribution (Figure 20). Cependant, la majorité des sujets présente un phénotype intermédiaire [110].

L'hypothèse sous-jacente est que les phénotypes extrêmes se produisent dans des cas extrêmes avec un excès de variants rares. Cette stratégie permet d'identifier des variants rares avec un échantillon relativement petit puisque les variants rares qui contribuent à une maladie sont enrichis dans les cas extrêmes de la distribution de la variable représentant le phénotype.

Un exemple remarquable est la détection des mutations dans le gène BRCA1, qui a été identifié dans des familles présentant une incidence élevée de carcinome mammaire à apparition précoce [111].

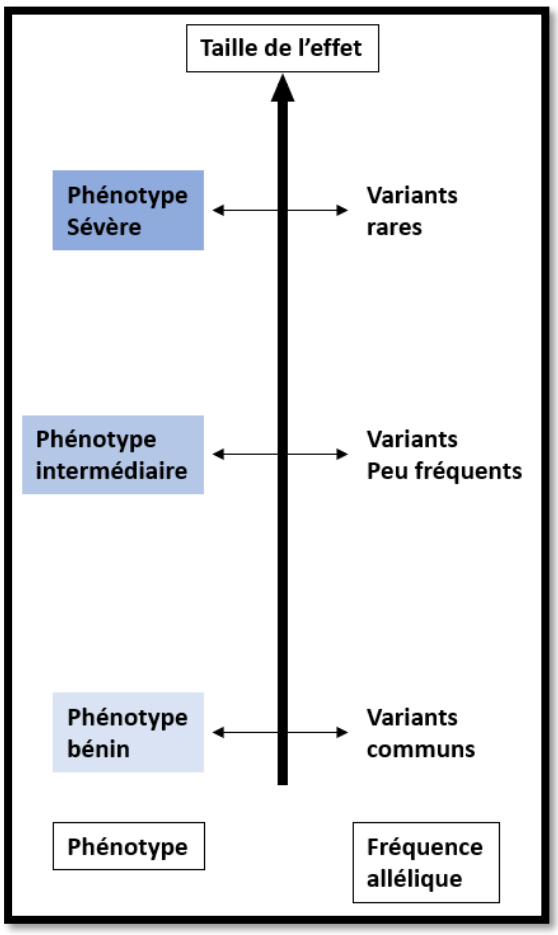

Figure 20. Relation entre la fréquence d'un variant et la taille de son effet sur le phénotype.

II.4.2.3. Les approches familiales

Les approches familiales sont particulièrement bien adaptées à l'analyse des variants rares et présentent plusieurs avantages par rapport aux approches populationnelles [112].

Premièrement, la taille des échantillons requise pour la détection des variants rares peut être beaucoup plus petite que celle requise pour les approches populationnelles. En effet, la ségrégation des variants rares dans un pedigree fournit des copies multiples de ces variants rares, ce qui facilite la détection de leurs effets [112].

Dans une étude sur la maladie d'Alzheimer, Carlos Crucchaga et al ont fourni un exemple de la façon dont les approches familiales peuvent aider à identifier des variants rares ayant des effets importants sur le risque d'une maladie complexe [113]. Ils ont effectué un séquençage d'exome dans 14 grandes familles de patients qui a permis l'identification d'un variant rare du gène PLD3 (phospholipase D3) ségrégeant avec la maladie dans deux familles indépendantes. Les analyses d'association de ce variant candidat dans plusieurs grandes cohortes cas-témoins indépendantes comprenant au total plus de 11 000 cas et témoins d'origine européenne ont montré qu'il doublait le risque d'Alzheimer [113].

II.4.3. Priorisation des variants issus du NGS

Les expériences basées sur le séquençage haut débit génèrent un grand nombre des données de variants génétiques. Leur filtrage reste une tâche très complexe et repose sur leur annotation qui prend en compte la fréquence allélique dans la population concernée, le site de ces variants (exons, introns des gènes, sites d'épissage des messagers, régions régulatrices, régions intergéniques), leur implication préalable dans des maladies, ainsi qu'à des algorithmes visant à estimer leur caractère potentiellement délétère.

Pour cette étape, plusieurs outils informatiques d'annotation des variants, qui produisent des rapports intégrés pouvant être analysés en détail, ont été développés. ANNOVAR est sans doute l'outil le plus utilisé [114]. Une fois les variants annotés, plusieurs critères incluant ceux évoqués ci-dessus peuvent être pris en compte pour sélectionner les variants avec le plus fort potentiel de pathogénicité.

II.4.3.1. Les classes des variants

On peut séparer les variants en deux groupes : les variants codants et les variants non codants. Ces deux groupes de variants ne sont pas évalués de la même façon, car les conséquences sur la protéine ne sont pas les mêmes. Il est plus facile d'étudier les effets d'un variant codant sur la protéine que d'étudier celui d'un variant non codant.

Les variants codants peuvent être divisés en deux groupes : les variants codants synonymes et non synonymes. Les variants codants sont dits non synonymes lorsque le changement du

nucléotide entraîne un changement d'acide aminé tandis que les variants synonymes n'ont pas d'impact sur le codon.

Les variants codants non synonymes de type faux-sens, non-sens, perte de codon stop, décalage du cadre de lecture ainsi que les variants situés sur un site d'épissage sont tous susceptibles d'affecter la structure et donc la fonction des protéines.

II.4.3.2. La MAF

La rareté est l'un des principaux critères permettant de prédire si un variant est susceptible d'avoir un effet fonctionnel sur la protéine. Il est logique de s'attendre à ce qu'un variant rare ait un effet fonctionnel plus important qu'un variant fréquent [115].

Actuellement, la plus grande base de données disponible fournissant des estimations des fréquences des variants est gnomAD (Genome Aggregation Database) [116].

Dans les études s'intéressant aux variants rares, un point de départ couramment utilisé pour filtrer les variants est une MAF inférieure à 1%.

II.4.3.3. Prédiction de l'impact fonctionnel

On peut sélectionner les variants potentiellement délétères en utilisant les algorithmes de prédiction fonctionnelle. Ces outils bio-informatiques ont été développés afin de permettre les analyses fonctionnelles de ces variants (Tableau 6). Ils permettent de savoir si un changement d'acide aminé affecte la fonction d'une protéine afin de classer ces changements par ordre de priorité pour les études plus approfondies.

Parmi ces outils, on peut citer :

- SIFT (Sorting tolerant from intolerant) est conçue pour classer par ordre de priorité les polymorphismes nucléotidiques simples non synonymes survenant dans la région codante du génome et pouvant entraîner une substitution d'acides aminés correspondants, et ce changement peut affecter la fonction de la protéine hôte et le phénotype de l'organisme hôte [117]. SIFT calcule un score de prédiction dérivé de la distribution des résidus d'acides aminés observés à la position donnée dans l'alignement de la séquence.

- PROVEAN (Protein Variation Effect Analyzer) est une méthode de prédiction basée sur le score d'alignement delta d'un alignement de séquences par paires [118]. Le score

d'alignement delta représente qu'il peut interpréter un changement dans le score d'alignement causé par une variation d'acide aminé comme l'impact fonctionnel de la protéine hôte du variant. PROVEAN peut également être utilisé pour prédire l'impact fonctionnel de toutes les classes de variations de séquences protéiques telles que les insertions, les délétions et les substitutions multiples.

- REVEL est une méthode d'ensemble, qui intègre de multiples scores de prédiction fonctionnelle et de conservation de séquence tels que SIFT, PROVEAN et d'autres [119].

- Polyphen-2 est un outil qui prédit l'impact possible d'une substitution d'acide aminé sur la structure et la fonction d'une protéine humaine en utilisant des considérations physiques et comparatives simples [120]. Il effectue l'annotation fonctionnelle des SNPs, fait correspondre les SNPs codants aux transcriptions des gènes, extrait les annotations de séquence des protéines et les attributs structurels, et établit des profils de conservation. Il estime ensuite la probabilité que la mutation faux-sens soit dommageable en se basant sur une combinaison de toutes ces propriétés.

- CADD (Combined Annotation Dependent Depletion) est un algorithme de prédiction qui intègre diverses annotations du génome et évalue toute variante nucléotidique unique humaine possible ou tout petit événement d'insertion/délétion [121]. CADD est un cadre qui intègre de multiples annotations en une seule mesure en opposant des variants qui ont survécu à la sélection naturelle à des mutations simulées. Il peut hiérarchiser quantitativement les variants fonctionnels, délétères et causaux de maladies dans un large éventail de catégories fonctionnelles, d'effet taille et d'architectures génétiques. La capacité de CADD à hiérarchiser les variants fonctionnels, délétères et pathogènes dans de nombreuses catégories fonctionnelles est inégalée par toute annotation actuelle.

Les variants régulateurs peuvent aussi entraîner plusieurs changements, notamment la modification de sites de fixation de facteurs de transcription, de marques d'histones et d'accessibilité de la chromatine à l'origine d'une modification du niveau d'expression de

gènes à proximité ou à distance du variant (Tableau 6).

Tableau 6. Résumé des méthodes de prédiction de l'impact fonctionnel des variants.

Méthode	**Type de variant**	**Mise à jour**
SIFT	Codant non-synonyme	Oui
PROVEAN	Codant non-synonyme	Oui
REVEL	Faux sens	Non
Polyphen-2	Codant non-synonyme	Oui
CADD	Tous	Oui
FATHMM-MKL	Tous	Oui
FunSeq2	Tous	Oui

II.4.3.4. Partage des variants par les patients d'une ou plusieurs familles

On filtre les variants partagés par les patients d'une ou plusieurs familles pour prioriser les variants potentiellement délétères en prenant en compte la structure des familles, en sélectionnant les variants partagés par le plus grand nombre de membres d'une famille porteur du trait étudié.

II.4.4. Méthodes d'analyses spécifiques de gènes

A cause de la faible fréquence des variants rares, les analyses classiques d'association sont peu efficaces. Cependant, plusieurs méthodes statistiques ont été développées pour tester les variants rares associés aux maladies complexes. Ces méthodes testent l'association de plusieurs candidats variants rares d'un gène ou d'une région donnée.

II.4.4.1. Test d'un seul variant

Les tests d'association simple marqueur décrits précédemment (II.3.2.3. Les tests d'association) sont utilisables dans le contexte des variants rares. Toutefois, ils ne permettent de mettre en évidence une association que si la taille des échantillons est suffisamment grande, si les effets sont très importants ou si les variants ne sont pas trop rares [122]. Dans la majorité des cas, ces conditions ne sont pas remplies, ce qui a conduit au développement de nouveaux tests.

II.4.4.2. Les tests d'association par agrégation

Les tests d'agrégation évaluent les effets cumulatifs de plusieurs variants génétiques dans un gène ou une région, ce qui augmente la puissance lorsque plusieurs variants du groupe sont associés à une maladie [122].

De nombreuses méthodes ont été développées, et nous classons ces méthodes en cinq catégories : les tests de charge, les tests de charge adaptatifs, les tests des composantes de la variance, les tests combinés de charge et de composante de variance, et le test de combinaison exponentielle (EC) (Tableau 7).

Tableau 7. Résumé des méthodes statistiques pour les tests d'association des variants rares [122].

Test	Description	Méthodes	Avantage	Inconvénient
Charge	Regrouper les variants rares en scores génétiques	ARIEL test, CAST, CMC method, MZ test, WSS	Sont puissants lorsqu'une grande proportion de variants sont causaux et que les effets vont dans le même sens.	Perdent de la puissance en présence de variants d'effets opposés ou d'une petite fraction de variants causaux.
Charge adaptatifs	Utilise des pondérations ou des seuils adaptés aux données	aSum, Step-up, EREC test, VT, KBAC method, RBT	Sont plus robustes que les tests de charge utilisant des pondérations ou des seuils fixes ; certains tests peuvent améliorer l'interprétation des résultats.	Sont souvent intensifs en termes de calcul ; le VT nécessite les mêmes hypothèses que les tests de charge.
Des composantes de la variance	Teste la variance des effets génétiques.	SKAT, SSU test, C-alpha test	Sont puissants en présence de variants augmentant ou diminuant le trait ou d'une petite fraction de variants causaux.	Sont moins puissants que les tests de charge lorsque la plupart des variants sont causaux et que les effets sont dans la même direction.
Combinés	Combine les tests de charge et de composante de variance.	SKAT-O, Fisher method, MiST	Sont plus robustes en ce qui concerne le pourcentage de variants causaux et la présence de variants augmentant et diminuant le trait.	Peuvent être légèrement moins puissants que les tests de charge ou de composante de variance si leurs hypothèses sont largement tenues.
De combinaison exponentielle	Combine exponentiellement les statistiques de score.	EC test	Est puissant lorsqu'une très faible proportion de variants est causale.	Est moins puissant lorsqu'une proportion modérée ou importante de variants est causale.

Les méthodes citées ci-dessus ont été développées pour les variants codants. Les extensions de ces tests aux variants non codants du génome sont un défi, car les variants non codants sont difficiles à sélectionner puisque leur impact fonctionnel sur l'expression des gènes est difficile à prévoir.

On peut citer par exemple FastSKAT, une extension de SKAT adapté pour les tests d'associations de très grands jeux de données. Cet outil peut être utilisé pour les variants non codants [123]. En plus de permettre une mise en œuvre plus rapide de SKAT, FastSKAT permet également des applications entièrement nouvelles de SKAT, qui n'étaient pas possibles auparavant ; par exemple le regroupement de variants par domaines topologiquement associés (TADs) et de comparaison d'association à l'échelle du chromosome par classe de marqueurs d'histone.

Il faut aussi noter qu'il existe les extensions de ces tests d'associations pour les approches familiales. Par exemple, Ionita-Laza et al. ont proposé FBAT, une extension du test d'association familiale qui inclut comme cas particuliers le test de fardeau et le test des composantes de la variance (SKAT) [124].

Chen et al. ont aussi proposé famSKAT, qui est une extension du test SKAT aux données familiales [125].

II.4.5. Validation génétique et fonctionnelle des variants

Dans les études des variants impliqués dans les maladies, il est essentiel de valider les résultats génétiquement et fonctionnellement.

II.4.5.1. Validation génétique

Il s'agit de confirmer les résultats de l'étude cas-témoins dans une autre cohorte dite de réplication, au minimum dans la même population, ou mieux dans une population différente.

II.4.5.2. Validation fonctionnelle

La validation fonctionnelle permet de tester l'impact du variant ayant un ou des effets inconnus dans les cellules ou les organismes. Plusieurs approches sont possibles selon le type de variant. On peut citer par exemple :

- Tester si la version sauvage du gène dans lequel le variant a été identifié permet de restaurer le phénotype normal.
- Tester les variants dans des systèmes modèles. Il peut s'agir de modèles dérivés de

patients, de modèles de culture cellulaire ou de modèles animaux.

- Analyser l'expression des protéines par Western blot afin de démontrer que les variants génétiques entraînent une perte d'expression ou d’activité de la protéine correspondante.

III. La génétique de la spondyloarthrite

III.1. La spondyloarthrite : une maladie complexe

La SpA est une maladie complexe faisant intervenir des facteurs génétiques et environnementaux.

III.1.1. Agrégation familiale

Comme mentionné précédemment (Figure 6), l'analyse familiale constitue la première étape pour évaluer la prédisposition génétique à une maladie. C'est autour des années 1950 que les premières observations de familles avec plusieurs membres atteints de SpA ont été publiées, suggérant l'implication de facteurs génétiques [126,127].

Le λ d'un apparenté au premier degré d'un patient a été estimé à 80 dans la SA et à 40 dans la SpA [128,129]. D'autres études ont calculé ce risque dans la SA et obtenu des valeurs similaires. Les études de population en Islande ont montré que ce risque au premier degré avait des valeurs allant de 75 à 94 [130,131].

Il faut noter que la prévalence de la SA dans la population générale utilisée dans ces études pour estimer les indices de récurrence était plus faible que dans la plupart des études de prévalence (0,1%), ce qui a pu conduire à une surestimation de l'indice de récurrence.

En revanche, deux études cas-témoins portant sur des patients atteints de SA en Suède ont fait état d'un risque nettement inférieur avec un λ_s compris entre 15 et 20 [132,133].

Récemment, une étude familiale prospective Suisse sur la SA a montré que le taux de récurrence à vie chez les apparentés au premier degré est estimé à 27,1 % [134].

Tableau 8. Résumé des études d'agrégation familiale dans la SpA.

Etude	**Phénotype**	**Taux de récurrence**
Brown MA et al [129]	SA	80
Dernis et al [128]	SpA	40
Thjodleifsson et al [130]	SA	94
Geirsson et al [131]	SA	75
Sundquist et al [132]	SA	17
Morin et al [133]	SA	19

III.1.2. Études de jumeaux

Une fois l'agrégation familiale mise en évidence, il est ensuite nécessaire d'évaluer la part respective des facteurs environnementaux et génétiques à l'origine de cette agrégation.

Dans la SpA, plusieurs études de jumeaux ont montré que le taux de concordance chez les jumeaux MZ est plus élevé que chez les jumeaux DZ :

Une étude a compilé quatre études incluant 83 paires de jumeaux et a démontré que 63 % de jumeaux MZ étaient concordants tandis que seulement 12,5 % étaient concordants chez les jumeaux DZ. L'héritabilité a été estimée à 97 % [49].

Une autre étude a compilé une étude norvégienne et deux études danoises (28 paires de jumeaux) et a montré que le taux de concordance était de 40 % chez les jumeaux MZ et de 4,3 % chez les jumeaux DZ. L'héritabilité a été estimée à 61 % [135].

D'autres études ont calculé l'héritabilité à partir de données des SNPs sur la totalité du génome avec un atout majeur qui est la possibilité d'utiliser des sujets non apparentés, sans risque de confusion due aux facteurs environnementaux partagés, et d'utiliser de très grands ensembles de données provenant des GWAS.

L'héritabilité de la SA basée sur les SNP a été estimée à 60,8 % en utilisant tous les SNPs de l'Immunochip et à 27,8 % en utilisant uniquement 244 signaux d'association indépendants significatifs au niveau du génome entier [136].

L'héritabilité de la SA a été estimée à 39,9 % dans la cohorte UK Biobank (UKBB) à l'aide d'une puce à ADN couvrant l'ensemble du génome avec une meilleure couverture que la puce Immunochip (https://nealelab.github.io/UKBB_ldsc/index.html).

Cependant, l'interprétation de cette estimation doit être faite avec prudence, car la taille de l'échantillon était très faible (584 patients), l'identification des cas dans cette cohorte manquait de précision (système d'information médicalisée) et il a été montré que cette cohorte n'était pas représentative de la population du Royaume-Uni (biais de sélection des volontaires sains) [137].

Toutefois, dans l'ensemble, ces études d'héritabilité suggèrent une forte contribution génétique à la SpA. La différence observée entre les évaluations de l'héritabilité provenant des études jumeaux et celles provenant d'études cas-témoins non apparentées suggère une contribution potentielle de variants rares ou structuraux non capturés par les puces à ADN, car le calcul de l'héritabilité basée sur les SNPs ne capte que les effets additifs des variants causaux marqués par des SNPs communs présents dans la puce utilisée [138].

III.1.3. Études de ségrégation

Une étude comparant les modèles génétiques de la SA afin d'évaluer le mode de transmission le plus probable, en utilisant les rapports de risque de récurrence chez les apparentés des sujets atteints a montré que le modèle de transmission correspondant le mieux était un modèle oligogénique avec une interaction multiplicative prédominante entre les loci [129].

III.1.4. Facteurs environnementaux

Il existe plusieurs éléments qui laissent penser qu'il y a une implication de facteurs environnementaux dans la SpA [139].

D'abord la survenue des arthrites réactionnelles, l'un des sous-types de SpA, déclenchées par des infections uro-génitales ou digestives [140] est en faveur d'une implication bactérienne dans le déclenchement de la maladie. Cette implication bactérienne est renforcée par le fait que l'un des modèles animaux de la SpA, le rat transgénique pour le HLA-B27 et la β2-microglobuline humaine, ne développe pas la maladie s'il est élevé dans les conditions stériles. La reconstitution de la flore intestinale suffit à déclencher la maladie [141–143].

Dans les suites de ces observations suggérant une implication du microbiote intestinal, plusieurs études métagénomiques ont montré une dysbiose dans la SpA, consistant principalement en une réduction de la biodiversité microbienne [144–149]. Cette composition perturbée du microbiote intestinal pourrait contribuer à la pathogenèse de la maladie [150].

D'autres facteurs environnementaux ont été impliqués soit dans l'activité de la maladie (tabac, stress), soit dans sa sévérité radiographique (tabac, sollicitations mécaniques répétées)[151–155].

III.2. Facteurs de susceptibilités génétiques identifiées dans la spondyloarthrite

Plusieurs études ont tenté d'identifier les facteurs de susceptibilité génétique dans la SpA en utilisant différentes méthodes et approches. Les études les plus anciennes se sont concentrées sur la région du complexe majeur d'histocompatibilité (CMH). Le développement progressif des méthodes d'étude du génome a permis secondairement l'étude d'autres régions génétiques, puis de l'ensemble du génome.

III.2.1. Complexe majeur d'histocompatibilité

Le complexe majeur d'histocompatibilité (CMH) humain est une région d'environ 3,6

mégabases (Mb) située sur le bras court du chromosome 6. Ce complexe génomique se compose de plus de 250 gènes parmi lesquels les gènes du système HLA (Human Leucocyte Antigen) sont les plus emblématiques. Les gènes du CMH sont classiquement répartis en trois régions distinctes appelées classes I, II et III [156].

Le CMH est la région la plus dense en gène du génome humain, avec de nombreux gènes impliqués dans la présentation d'antigènes cellulaires et extracellulaires aux cellules T, les réponses inflammatoires et immunitaires, la cascade du complément, la signalisation des cytokines et la régulation de divers aspects du développement, de la différenciation et de l'apoptose cellulaires [157]. Probablement du fait de son importance biologique, le CMH est aussi la région du génome humain où l'on cartographie le plus nombre de signaux d'association à l'aide des GWAS.

Dans la SpA, plusieurs facteurs de susceptibilité génétique ont été identifiés dans cette région dès les années 70.

III.2.1.1. L'allèle HLA-B*27

L'association entre la SpA et l'antigène HLA-B27 a été démontrée en 1973 par deux études distinctes qui rapportaient 96 % de porteurs de HLA-B27 parmi les patients [158,159]. D'autres études ont largement confirmé cette l'association [160].

Il existe plusieurs sous-types de l'antigène HLA-B*27 dont l'évolution à partir de l'allèle parent HLA-B*27:05 reflète les migrations humaines. Certains de ces sous-types sont associés à la maladie (B*27:05, majoritaire chez les caucasiens, B*27:02, B*27:04, B*27:07) alors que d'autres ne le sont pas [160] (B*27:06, fréquent en Asie du Sud Est [161] et B*27:09, décrit en Sardaigne et en Italie du Sud [162]).

En dépit de la force et l'ancienneté de l'association de l'antigène HLA-B*27 avec la SpA, les mécanismes sous-jacents restent méconnus. Plusieurs hypothèses ont été proposées sans qu'elles soient confirmées, notamment l'hypothèse des peptides arthritogènes, celle de la mauvaise conformation des protéines et celle des homodimères de la chaîne lourde HLA-B*27 [163–166].

III.2.1.2. Les autres gènes du complexe majeur d'histocompatibilité

Plusieurs études ont montré l'implication d'autres allèles du gène HLA-B ainsi que d'autres gènes HLA dans le CMH (Tableau 9). Malheureusement, les résultats sont parfois discordants du fait de la difficulté de mettre en évidence une association indépendante de HLA-B*27.

Tableau 9. Les autres allèles HLA-B et gènes HLA associés au développement de la SpA.

CMH	Population	Association
Classe I		
HLA-B*40	Néerlandaise [167], Américaine blanche, noire, chinoise Han [168] Européenne [169]	Associé
HLA-B*13:02	Européenne [169]	Associé
HLA-B*47:01	Européenne [169]	Associé
HLA-B*51:01	Européenne [169]	Associé
HLA-B*57:01	Européenne [169]	Protecteur
HLA-B*07	Blanche, noire, Chinoise Han [168,169]	Protecteur
HLA-B*15	Colombienne [170]	Négativement associé
HLA-C*12:02:02	Taiwanaise [171]	Associé
HLA-A*02:01	Européenne et Américaine blanche [169,172]	Associé
MICA*007:01	Chinoise et Américaine [173,174]	Associé
MICA*019	Chinoise et Taiwanaise [173,175]	Associé
Classe II		
HLA-DPA1	Américaine blanche et européenne [168,169,172,176,177]	Associé
HLA-DPB1	Américaine blanche et Européenne [168,169,177]	Associé
HLA-DQB1*06:02	Européenne et Américaine blanche [168,169,176,178]	Négativement associé
HLA-DRB1*15:01	Européenne et Américaine blanche [168,169]	Négativement associé

III.2.2. Les gènes en dehors du complexe majeur d'histocompatibilité

II.2.2.1 Arguments en faveur de l'implication d'autres facteurs génétiques dans la SpA.

Plusieurs arguments sont en faveur de l'implication d'autres facteurs génétiques dans la SpA en dehors du CMH.

Le premier des arguments est le fait que 5 à 25% des malades ne sont pas porteurs de HLA-B*27 et 3 – 4 % seulement des porteurs de HLA-B*27 dans la population générale développent la maladie [4,179–181].

Le deuxième argument est que dans les études des jumeaux, la différence du taux de concordance entre les jumeaux MZ et DZ persiste même si on ne compare que les paires de jumeaux porteurs de HLA-B27 [49,135,182].

Le troisième argument est que le BASDAI, le BASFI et le BASRI (respectivement indices d'activité, de sévérité fonctionnelle et de sévérité radiologique de la maladie), sont influencés par d'autres facteurs génétiques que le HLA-B27, car ces indices ont été associés à une forte héritabilité dans certaines études faites uniquement chez les patients porteurs de l'antigène HLA-B27 [49,182].

Bien que la plus grande contribution des loci expliquant l'héritabilité de la SpA provienne du CMH, elle n'explique que 20,4% de l'héritabilité totale [136].

Les études de liaison et d'association génétique ont été utilisées afin d'identifier d'autres facteurs génétiques impliqués dans la SpA.

III.2.2.2. Etudes de liaison

Plusieurs études de liaison pangénomique ont été publiées. Dans toutes ces études, la région du CMH était fortement liée à la SpA mais seulement trois autres locus en dehors du CMH ont atteint un score de liaison statistiquement significatif, et ce dans une seule étude pour chaque locus : l'un en 16q dans la SA, le deuxième en 9q31-34 dans la SpA et le dernier en 13q13 dans la SpA (Tableau 10) [183–186]. Deux méta-analyses basées sur les résultats de trois premières études de liaison ont mis en évidence une liaison significative au niveau de la région du CMH [187,188].

En dehors de l'étude de liaison la plus récente [186], toutes sont basées sur l'utilisation d'un nombre limité de microsatellites à l'origine d'une résolution spatiale assez faible. Les études britanniques et nord-américaines [183,184] ont inclus une majorité de paires de germains alors que les deux études françaises [185,186] ont inclus des familles multiplex étendues. Ces deux dernières études ne sont pas totalement indépendantes, car il y a les familles communes aux deux études [185,186].

Tableau 10. Résumé des études de liaison conduites dans la SpA.

Etude	**Phénotype**	**Nombre de familles**	**Marqueur**	**Locus significatifs**
Oxford [183]	SA	99+86	505+367 microsatellites	CMH, 16q23.3
NASC [184]	SA	180	400+16 microsatellites	CMH
GFEGS [185]	SpA	120	369+21 microsatellites	CMH, 9q31 – 34
Costantino et al. [186]	SpA	143	189368 SNPs	CMH, 13q13

Dans une analyse d'association, Pointon et al. ont identifié deux variants associés de manière indépendante à la SA sur le chromosome 16q dans une cohorte anglaise de 730 patients et 2879 témoins [189].

Les locus significatifs 9q31-34 et 13q13 ont aussi fait l'objet d'un suivi afin d'identifier plus précisément les gènes impliqués dans ces régions significativement liées. Cependant, les locus de susceptibilité identifiés dans ces régions n'expliquent pas totalement la liaison [190–192].

Les études de liaison dans la SpA ont donc été une déception, car aucun gène de susceptibilité à la SpA n'a pu être identifié de manière reproductible en dehors du CMH.

III.2.2.3. Etudes d'association

Les études d'association génome entier et gènes candidats à grande échelle ont aussi été menées dans différentes populations afin d'identifier d'autres facteurs génétiques impliqués dans la SpA.

Les études d'association gène – candidat

Plusieurs gènes candidats ont fait l'objet des nombreuses études d'association dans la SpA et dans différentes populations. Malheureusement, la majorité des signaux d'association identifiée n'ont pas pu être répliqués dans d'autres études indépendantes, ce qui remet en question ces résultats.

Les études d'associations génome entier

Plusieurs études génome entier ont été publiées :

- Une première étude d'association publiée en 2007 a rapporté une association de deux loci (en dehors du CMH) avec la maladie, ERAP1 et IL23R [193].
- Une deuxième étude d'association publiée en 2010 a montré une association dans trois loci de plus avec la maladie dans une cohorte européenne et a confirmé l'association d'ERAP1 et IL23R [194].
- Une autre étude d'association publiée en 2011 a montré l'association de 7 autres loci avec la maladie [195].
- Deux nouveaux loci de susceptibilité ont été identifiés dans une étude d'association chinoise de 2011, mais aucun n'a été reproduit, même dans de grandes études [196].

Deux autres grandes études ont consisté en l'analyse simultanée d'un grand nombre de locus potentiellement impliqués dans les maladies auto-immunes ou inflammatoires, à l'aide d'une puce dédiée (Immunochip) :

En 2013, le Consortium international sur la génétique de la spondylarthrite ankylosante (IGAS) a identifié 13 nouveaux loci associés à la maladie dans une large étude cas-contrôles en utilisant la puce Immunochip [172].

En 2016, une analyse combinée des données de la puce Immunochip de 5 maladies inflammatoires chroniques a permis d'identifier 17 nouveaux loci associés à la maladie [136].

L'ensemble de ces études ont conduit à l'identification d'un peu moins de 50 de facteurs génétiques associés à la maladie (Tableau 11). Il faut toutefois souligner que la plupart de ces facteurs de susceptibilité modifient faiblement le risque de développer la maladie.

Tableau 11. Résultats des études d'association dans la SpA.

Étude	**Nombre de cas/témoins (SNP génotypés)**	**Facteurs génétiques identifiés en dehors du CMH**
Burton *et al.* [193]	Découverte : 922/1466 (15333) Réplication : 471/625 (15)	*IL23R, ERAP1*
Reveille *et al.* [194]	Découverte : 2053/5140 (288662) Réplication : 898/1518 (163)	IL23R, 2p15, IL1R2, ANTXR2, ERAP1, 21q22
Evans *et al.* [195]	Découverte : 3023/8779 (2223620) Réplication : 2111/4483 (49)	*RUNX3, IL12B, LTBR-TNFRSF1A, ANTXR2, PTGER4, CARD9, TBKBP, IL23R, ERAP1, 2p15, 21q22, KIF21B*
Lin *et al.* [197]	Découverte : 1837/4231 (1356350) Réplication : 2100/3496 (30)	*EDIL3-HAPLN1, ANO6, 2p15*
IGAS *et al.* [172]	Découverte : 10619/15145 (128935)	*RUNX3, IL23R, GPR25-KIF21B, 2p15, PTGER4, ERAP1, IL12B, CARD9, LTBR-TNFRSF1A, NPEPPS-TBKBP1-TBX21, 21q22, IL6R, FCGR2A, UBE2E3, GPR35, BACH2, ZMIZ1, NKX2-3, SH2B3, GPR65, IL27-SULT1A1, NOS2, TYK2, ICOSLG*
Ellinghaus et *al.* [136]	Découverte : 6463/6069 (9113515)	*ITLN1, ASAP2, CMC1, RNA5SP142-SOX14, NPM1P17, NFKB1, RP11-1C1.5, CDKAL1, FGFR1OP, AGR3-Metazoa_SRP, ACTA2, LINC02098-ETS1, PPP2R3C, ERN1, MAPK3-CORO1A, PTPN2, FAM118A*

III.3. Que nous ont appris les GWAS dans la SpA ?

Les approches GWAS ont permis de mieux comprendre le rôle des facteurs génétiques dans la susceptibilité à la SpA et d'identifier les voies impliquées dans la SpA, telles que les aminopeptidases ou la voie de signalisation IL23/IL17.

La comparaison des résultats de GWAS entre les maladies impliquant le système immunitaire a également conduit au concept de fond génétique commun, car plusieurs facteurs génétiques de susceptibilité à ces maladies sont partagés [136].

Néanmoins, les GWAS ont été aussi décevants, car elles n'ont pas permis d'expliquer complètement l'héritabilité de la maladie.

Ainsi, l'ensemble des facteurs identifiés dans l'étude de l'IGAS (International Genetics of Ankylosing Spondylitis Consortium), incluant la région du CMH, expliquent moins du 1/4 de la prédisposition totale à la maladie (Figure 21) [136,172].

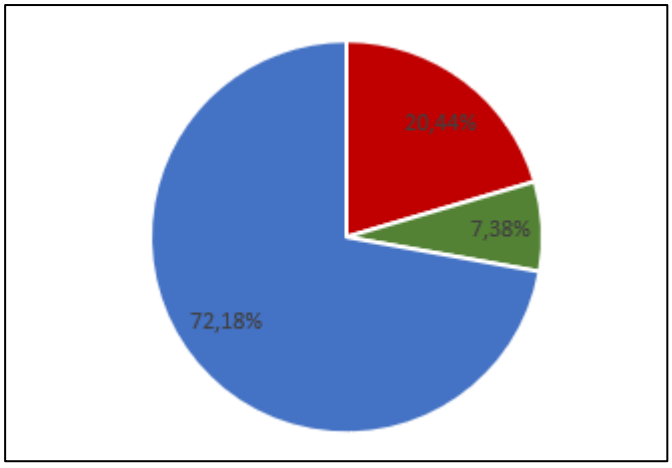

Figure 21. La part de l'héritabilité expliquée et inexpliquée. L'héritabilité expliquée par le HLA-B27 (rouge) et par les autres locus identifiés (en vert) 7,38 % représente 20,44 % et 7,38 % respectivement L'héritabilité inexpliquée représente un peu moins de 75 % (bleu).

III.4. Où trouver l'héritabilité manquante dans la SpA ?

Comme abordé dans le chapitre 2, différentes hypothèses ont été avancées pour expliquer l'héritabilité manquante dans les maladies complexes. Quelques études y ont été consacrées dans la SpA.

III.4.1. Les interactions gène – gène et gène – environnement

Le grand défi que représente la détection des interactions gène-gène et gène-environnement fait qu'il est difficile d'expliquer pleinement l'héritabilité des traits complexes à l'heure

actuelle.

Dans la SpA, seule l'interaction entre ERAP1 et le HLA-B27 a été démontrée. Evans et al, ont testé l'interaction entre le HLA-B27 marqué par le SNP rs4349859 et le SNP rs30187 du gène ERAP1 dans l'ensemble de données WTCCC2 (Wellcome Trust Control Consortium 2), TASC (Australo-Anglo-Americain Spondyloarthritis Consortium) et de réplication, et ont démontré que chez les cas et les témoins HLA-B27 positifs, ERAP1 était associé à la spondylarthrite ankylosante, alors qu'il n'y avait aucune association avec ERAP1 chez les cas HLA-B27 négatifs [195]. Ces résultats ont été confirmés par l'étude de Cortes et al dans une population européenne [169].

Dans la SpA, aucune interaction gène-environnement n'a été clairement démontrée jusqu'à présent, même si le tabac est l'un des principaux facteurs environnementaux impliqués dans la pathogenèse de diverses maladies inflammatoires, notamment la SpA.

Dans une étude mexicaine récente, il a été montré que les SNPs rs30187 et rs27044 du gène ERAP1 semblent potentialiser l'effet du tabagisme chez les patients atteints de SA HLA-B27 positif [198]. Il faudra d'autres études dans d'autres populations et avec des échantillons de plus grande taille pour confirmer ces résultats.

III.4.2. Variants structuraux

Les variants structuraux tels que les CNV, les délétions et les inversions, pourraient aussi jouer un rôle important dans la susceptibilité de la SpA mais ils ont été peu étudiés.

Seuls les CNV ont été étudiés à l'échelle du génome dans une étude coréenne concernant 618 individus (309 patients atteints de SA et 309 contrôles). Dans cette étude, 227 CNV ont été associés à la maladie, mais l'association de seulement 4 CNV a pu être répliquée [199]. Les CNV associées à la SA ont été répliqués dans deux cohortes indépendantes (625 patients et 891 sujets témoins).

Aucune étude ne s'est intéressée aux autres variants structuraux de grande taille.

III.4.3. Les variants rares

Comme suggéré dans la partie 2 (II.4. Les variants rares), les variants rares pourraient expliquer une importante part de l'héritabilité manquante de la SpA. Les GWAS sont incapables de détecter les variants rares associés aux maladies, car la majorité des SNPs inclus dans les GWAS sont des variants fréquents. En outre, peu d'études identifiant les variants rares liées à la SpA ont été publiées.

Les études cas-témoins manquent très souvent de puissance statistique pour détecter les variants rares de manière précise.

Robinson et al, ont mené une étude d'association dans une population caucasienne concernant 5040 patients SA et 21133 témoins sains à l'aide de la puce Illumina Exomechip. Une association significative à l'échelle du génome a été observée au niveau du gène CDKAL1 [200].

Dans une autre étude d'association utilisant une cohorte turque (1001 patients atteints de SA et 1011 témoins) et iranienne (479 patients atteints de SA et 830 témoins), une association d'un variant rare du gène MEFV et la SA a été identifiée [201]. La validation génétique et fonctionnelle de ces variants dans d'autres populations est nécessaire afin de confirmer ces résultats.

Plusieurs études combinant le séquençage haut débit et les approches familiales ont identifié les variants rares dans la SpA mais la plupart du temps ces variants n'ont pas été trouvés dans d'autres familles (Tableau 12). De façon non surprenante, ces variants n'étaient pas été couverts par les GWAS.

Dans une étude combinant séquençage d'exome et approche familiale, O'Rielly et al., ont identifié des délétions rares dans les gènes SEC16A et MAMDC4 et dans une autre étude Liu et al., ont identifié un variant rare dans le gène RFN123. Ces trois variants rares ont été étudiés fonctionnellement.

En outre, d'autres variants rares ont été identifiés dans les gènes UGTB17, IRS1, ANKDD1B et TREML2, RELN et FREM2. Cependant, ces variants n'ont pas été étudiés d'un point de vue fonctionnel.

L'ensemble de ces résultats suggèrent que les variants rares contribuent probablement à la pathogenèse de la maladie, mais soulignent la nécessité d'études complémentaires et de la recherche d'autres variants rares.

Tableau 12. Les variants rares identifiés dans la SpA.

Etude	**Echantillon**	**Ethnie**	**Design**	**Gène**	**Etude fonctionnelle**
Uddin et al. (2013) [202]	1 famille	Caucasien	Génome	UGTB17	Non
Rong et al. (2015) [203]	1 famille	Chinois	Séquençage ciblé	IRS1	Non
O'Rielly et al. (2016)[204]	1 famille	Caucasien	Exome	SEC16A MAMDC4	Oui
Tan et al. (2018) [205]	1 famille	Chinois	Liaison + exome	ANKDD1B	Non
Feng et al. (2018)[206]	1 famille	Chinois	Liaison + exome	TREML2	Non
Garshasbi et al. (2020) [207]	1 famille	Iranien	Exome	RELN	Non
Liu et al. (2020) [208]	1 famille	Chinois	Exome	RNF123	Oui
Mambu Mambueni H et al. (2022) [192]	1 famille	Caucasien	Séquençage ciblé	FREM2	Non

RESULTATS

L'objectif de cette thèse était d'identifier des variants rares associés à la SpA permettant d'expliquer une partie de l'héritabilité manquante dans cette maladie au travers d'une approche combinant séquençage haut-débit et études familiales. Ce projet se divise en trois parties. Nous avons d'abord tenté d'identifier les variants rares d'une région d'intérêt par une approche de séquençage ciblé. Grâce au développement des techniques de séquençage, nous avons ensuite élargi notre approche à l'ensemble du génome, en se focalisant d'abord sur les régions codantes (séquençage d'exomes) puis aux régions codantes et non codantes (séquençage du génome entier).

Par souci de clarté et afin d'éviter les répétitions, nous allons dans un premier temps exposer la méthodologie commune à toutes les études avant d'aborder pour chacune leurs spécificités méthodologiques et d'exposer leurs résultats.

I. Méthodologie commune à toutes les études

I.1. Patients

Les patients ont été recrutés par l'intermédiaire du "Groupe Français d'Etude Génétique des Spondylarthropathies" (GFEGS), comme décrit précédemment [209]. Tous les patients inclus dans les études familiales et cas-contrôles répondaient aux critères de classification d'Amor et/ou de l'ESSG [210,211]. Les critères de classification de l'ASAS pour les SpA axiales et périphériques n'étaient pas élaborés au moment du recrutement, mais ils ont été appliqués a posteriori [212,213].

Au moment de l'inclusion, un échantillon de sang a été prélevé permettant l'extraction d'ADN. Un typage HLA-B*27 a été réalisé chez l'ensemble des sujets inclus et les caractéristiques démographiques, cliniques et radiographiques (radiographies du bassin) ont également été recueillies.

Ces études ont été approuvées par les comités d'éthique de l'hôpital Cochin (Paris, France) et de l'hôpital Ambroise Paré (Boulogne Billancourt, France). Un consentement éclairé écrit a été obtenu de chaque participant.

I.2. Séquençage et génotypage

I.2.1. Séquençage haut-débit

L'ADN génomique (100 ng) a été fragmenté par voie enzymatique et traité pour générer des librairies de séquençage comme recommandé par Illumina. Les librairies ont ensuite été hybridées avec des sondes de capture biotinylées et enrichies avec des billes magnétiques de streptavidine. Les fragments d'ADN ont été élués et séquencés par paires à l'aide d'un séquenceur (Illumina).

I.2.2. Séquençage Sanger

Pour chaque variant testé en Sanger, un amplicon a été amplifié à partir de 200 ng d'ADN génomique. Les amorces utilisées ont été dessinées sur Primer 3 et commandées chez Eurofins MWG. Environ 10-30 ng du produit purifié ont été séquencés sur un séquenceur capillaire ABI 3730xl (Applied Biosystems), et les séquences résultantes ont été analysées à l'aide du logiciel SeqScape version 2.5 (Applied Biosystems) et Chromas version 2.6.6 (Technelysium).

I.2.3. Génotypage par sondes TaqMan

Le génotypage par la méthode de sonde TaqMan a été réalisé selon les recommandations du producteur Thermofisher à partir de 20 ng d'ADN.

A la fin de la réaction, la fluorescence est lue et la combinaison allélique de chaque individu testé est obtenue grâce à une analyse informatique par clustering. L'assignation des génotypes s'est faite à l'aide du logiciel CFX Manager (Biorad).

I.2.4. Courbe de fusion à haute résolution

Lorsque c'était techniquement possible, les SNPs ont été génotypés par la méthode de courbe de fusion à haute résolution, selon les recommandations du producteur Biorad.

Les amorces utilisées ont été dessinées sur Primer 3 et commandées chez Eurofins MWG.

L'assignation des génotypes s'est faite grâce aux logiciels CFX Manager (Biorad) et Precision Melt Analysis (Biorad).

I.3. Traitement des données de séquençage haut débit

Les fichiers Fastq ont été débarrassés des bases de mauvaise qualité et des adaptateurs de lecture en utilisant FASTP [214]. Les lectures ont ensuite été alignées sur la séquence du

génome humain de référence (GRCh38) en utilisant l'algorithme bwa-mem. Les doublons optiques et PCR ont été supprimés avec l'outil MarkDuplicates Picard. La couverture a été déterminée avec BEDTools [215] et Qualimap [216].

Les variants ont été appelés avec l'outil haplotypeCaller de GATK v4. Les variants ont ensuite été recalibrés avec l'outil GATK VariantRecalibrator en utilisant des bases de données de variants nucléotidiques simples et d'indels, conformément aux recommandations de bonnes pratiques du Broad Institute [217].

Le fichier vcf des variants appelés ayant passé le contrôle de qualité a finalement été annoté avec ANNOVAR en utilisant les principales bases de données disponibles pour les variants germinaux (1KG, GnomAD, dbSNP, ClinVar). La pathogénicité des variants codants a été estimée avec les principaux algorithmes de prédiction (SIFT, POLYPHEN2, M-CAP, REVEL, CADD) [114]. La pathogénicité des variants non codants potentiellement régulateurs a été évaluée avec RegulomeDB [218].

I.4. Analyses génétiques

I.4.1. Vérification de degré de parenté entre les individus

Les liens de parenté entre les individus ont été vérifiés en utilisant le logiciel King avec l'option Kinship [219].

I.4.2. Analyse de ségrégation

La première étape permettant d'évaluer l'intérêt d'un variant dans une étude familiale est de vérifier sa ségrégation avec la maladie au sein des familles où le variant est présent, c'est-à-dire sa présence chez les malades et son absence chez les témoins sains.

La visualisation des pedigrees a été réalisée grâce au logiciel Haplopainter [220].

I.4.3. Analyse d'association

L'étude d'association cas-contrôles consiste à comparer la fréquence allélique ou génotypique d'un variant dans une même population entre un groupe de malades et un groupe de sujets sains non apparentés.

Nous avons réalisé les analyses d'association cas-contrôles grâce au logiciel PLINK [221].

La significativité et l'homogénéité des tableaux de contingence de la partie I ont été évaluées avec le progiciel R epiDisplay.

II. Partie I : Le reséquençage du locus 13q13

Cette première partie des résultats a fait l'objet d'un article original en premier auteur publié en 2022 dans la revue Joint Bone Spine [192].

II.1. Introduction

Une liaison significative de la région 13q13 avec la SpA a été rapportée par notre équipe en 2016 grâce aux analyses de liaison non paramétriques à partir de puces Affymetrix 250K chez 914 sujets dont 462 patients de SpA venant de 143 familles multiplex (Figure 22) [222]. Ce pic de liaison s'étendait sur 6 Mb (autour de 40 Mb du p-télomère) (Figure 23).

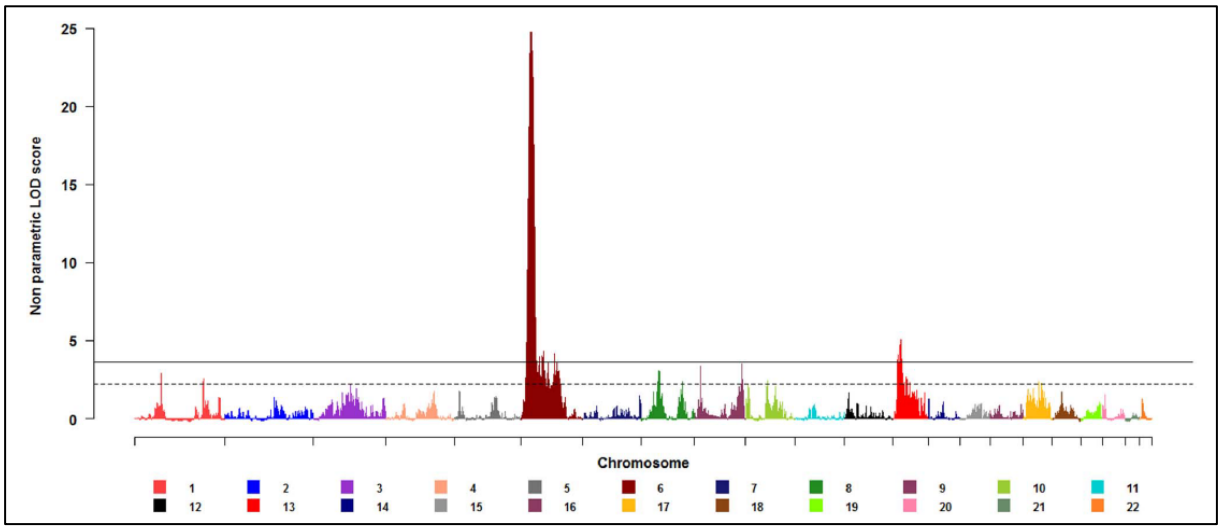

Figure 22. Résultats de l'analyse de liaison non paramétrique à l'échelle du génome. L'axe des x indique les positions des marqueurs sur le génome, chaque chromosome étant représenté par une couleur distincte. L'axe des y montre le LOD score. Les seuils de LOD score significatif et suggestif sont représentés par des lignes pleines et pointillées, respectivement.

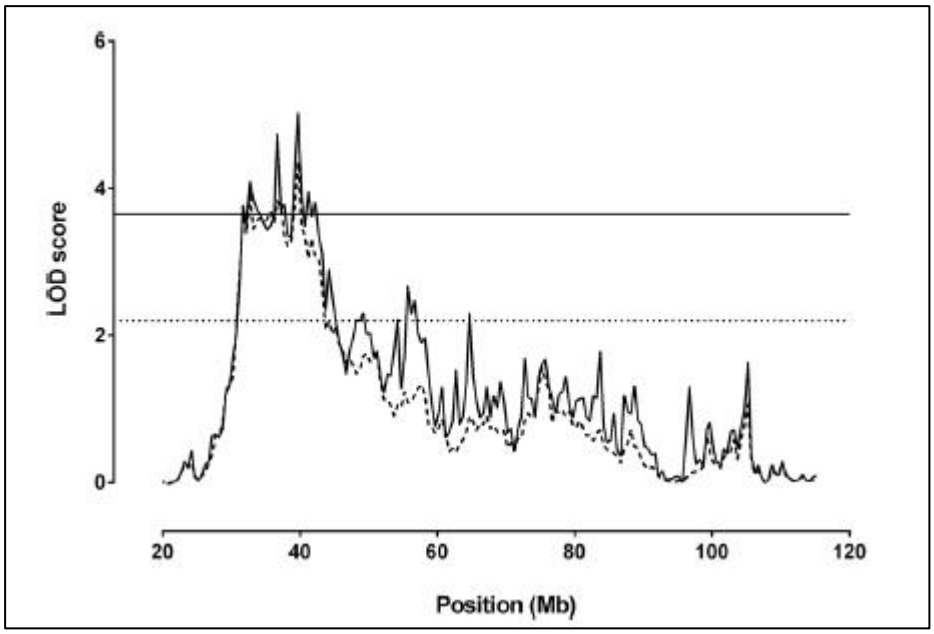

Figure 23. Résultats de l'analyse de liaison non paramétrique du chromosome 13, prenant en compte (ligne pointillée) ou non (ligne pleine) le DL. L'axe des x représente la distance physique à partir du p-télomère et l'axe des y représente le LOD score.

Nous avons reconstruit les haplotypes autour du pic de liaison dans les 43 familles les mieux liées afin de mieux circonscrire la région de liaison avant de l'étudier, nous permettant ainsi d'identifier les haplotypes coségrégeant avec la maladie ainsi que les événements de recombinaison méiotique. Grâce à ces événements de recombinaisons, l'intervalle d'intérêt a pu être restreint à une région de 1,4 Mb (chr13 : 38125001 – 39527000) (Figure 24).

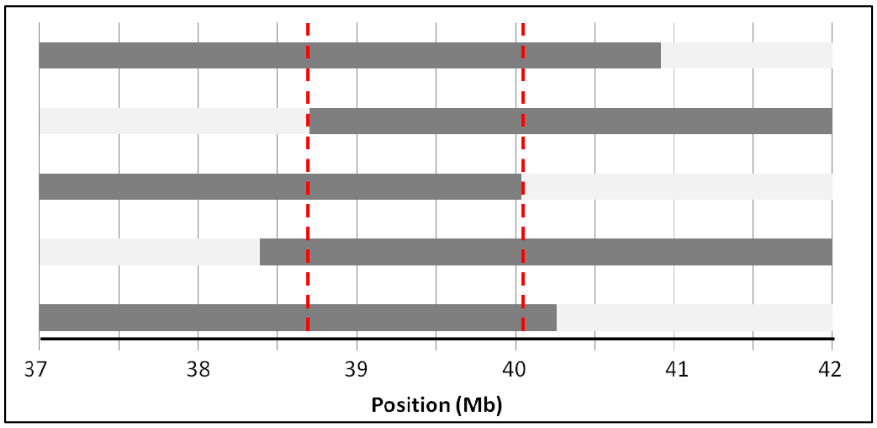

Figure 24. Intervalle de liaison (encadré en rouge) délimité par des événements de recombinaisons informatives chez les patients. Pour chaque recombinaison informative, la barre grise représente la portion de l'haplotype recombinant qui ségrège avec la maladie chez les autres membres de la famille.

L'objectif de cette étude était d'identifier les variants de susceptibilité à la SpA pouvant expliquer le signal de liaison dans cette région. Pour ce faire, nous avons réalisé un

reséquençage ciblé de l'intervalle de la maladie (Figure 24) chez 92 sujets (71 patients atteints de SpA et 21 sujets non affectés) issus des 43 familles les plus liées à cette région.

II.2. Patients et méthodes

II.2.1. Echantillons biologiques

II.2.1.1. ADN

Les patients ont été recrutés comme décrit précédemment (I.1. Patients).

Les individus séquencés appartenaient aux familles ayant obtenu un score de liaison élevé dans la région d'intérêt. Le choix des individus à séquencer a été fait selon deux critères :

- Dans les familles dans lesquelles un haplotype coségrégeant avec la maladie a pu être identifié, nous avons sélectionné deux membres atteints de SpA portant cet haplotype et un membre non malade ne le portant pas (Figure 25A).
- Dans les familles où tous les membres affectés partageaient un seul haplotype, un seul membre atteint a été séquencé (Figure 25B).

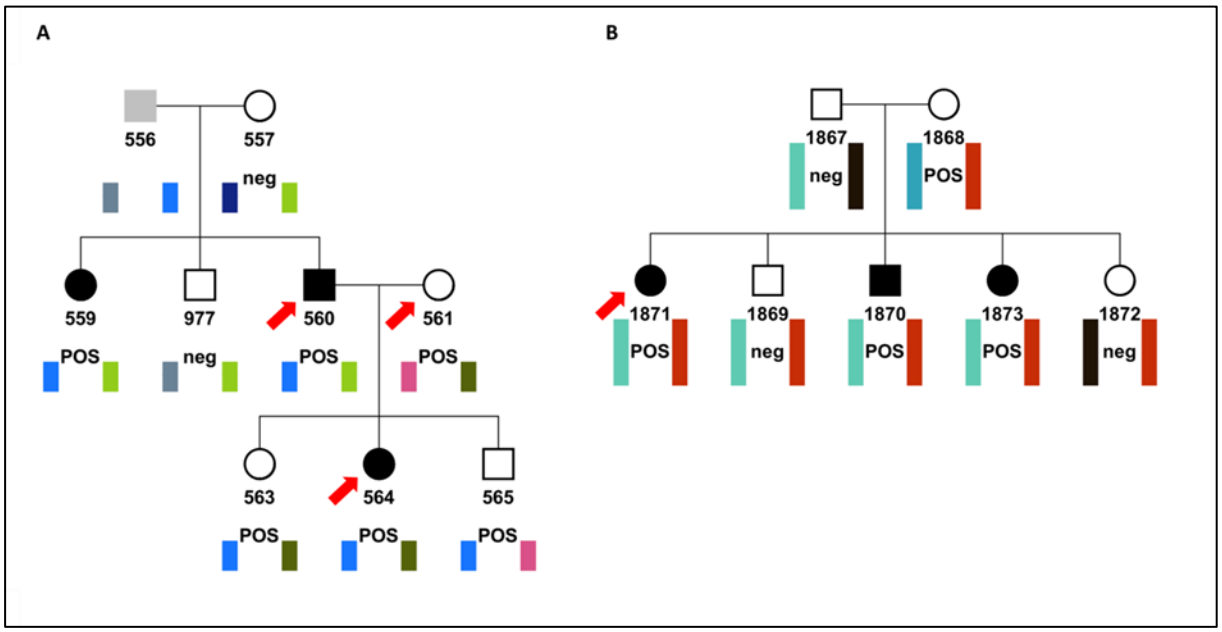

Figure 25. Sélection d'individus pour le reséquençage ciblé de la région 13q13. Les barres colorées représentent les haplotypes déduits dans la région 13q13. Les flèches rouges indiquent les individus dont l'ADN a été séquencé. Le panel A montre une famille dans laquelle un haplotype (le bleu) coségrège avec la maladie : dans cette situation, nous avons choisi pour le séquençage deux membres affectés portant l'haplotype coségrégeant avec la SpA et un membre non affecté avec des haplotypes différents. Le panel B montre une famille dans laquelle tous les membres affectés sont haplo-identiques. Dans cette situation, nous avons sélectionné un seul membre affecté pour le séquençage.

Les caractéristiques cliniques des patients inclus dans les études familiales et cas-contrôles de ce projet sont résumées dans le Tableau 13.

Tableau 13. Caractéristiques des patients inclus dans les études familiales et cas-contrôles.

Caractéristiques	**Séquencés (N = 71)**	**Analyse de ségrégation (N = 134)**	**Etude d'association (N = 1091)**
Age début symptôme : moyenne ± écart-type (année)	49,7± 2,8	42,5±14,3	46,4±10,6
Durée de la maladie : moyenne ± écart-type (année)	25,6±11,7	20,5±18	18,9±10
Sexe ratio (% hommes)	60,5	54,5	48,2
HLA-B*27 positif (%)	92,9	93,3	70
Manifestations axiales (%)			
Lombalgie inflammatoire	90,1	98,5	87,5
Sacro-iliite radiographique	67,6	59,7	40,4
Manifestations périphériques (%)			
Arthrites périphériques	49,2	42,5	36
Enthésites périphériques	78,8	70,1	72,7
Manifestations extra-articulaires (%)			
Uvéites	32,3	25,4	22,2
Psoriasis	29,5	25,4	25,5
MICI	12,6	6,7	7,3

II.2.1.2. Tissus pour immunomarquage

Pour les études d'immunomarquage, des biopsies de côlon et de synoviale provenant de sujets non atteints de SpA ont été fournies par le service d'anatomopathologie de l'hôpital Ambroise Paré (Boulogne-Billancourt, France). Les échantillons de cartilage ont été obtenus chez un patient atteint de SpA lors d'une arthroplastie totale de la hanche.

Des rats Fisher ont été élevés et maintenus dans des conditions conventionnelles. Après le

sacrifice, des échantillons de tissu synovial, cartilage, rein et colon ont été prélevés.

II.2.2. Reséquençage du locus 13q13

Pour le reséquençage profond de la région cible 13q13 (GRCh38p12, chr13:38125001-39527000), nous avons utilisé un protocole de capture par hybridation basé sur la technologie Nextera® d'Illumina. Nous avons conçu un ensemble de 6396 sondes, offrant une couverture d'une densité moyenne de 1 sonde/200 bases. Compte tenu de la taille des sondes (80 nucléotides), l'intervalle inter-sondes moyen était de 140 bases.

Cela a permis de s'assurer que la majeure partie de la séquence cible était couverte par deux sondes, puisque les fragments d'ADN capturés avaient une longueur moyenne de 350 bases. Seules 8 lacunes avaient une longueur supérieure à 250 bases. Au total, ils représentaient ~ 30 kilobases de séquence non couverte et correspondaient à des séquences hautement répétitives.

II.2.2.1. Génotypage

Les différentes méthodes de génotypage utilisées pour les études de ségrégation et cas-témoins sont résumées ci-dessous (Tableau 14).

Tableau 14. Techniques de génotypage utilisées en fonction des variants

Technique	**Variant**
TaqMan	rs201714905
HRM	rs9603422
	rs11841861
	rs17058288
	rs80103953
	rs8002488
	rs79240489
	rs41306662
	rs77958976

II.2.2.2. Immunomarquage sur des échantillons humains et de rats

Tous les tissus prélevés ont été fixés et inclus en paraffine. Des sections de 6 mm ont été déparaffinées.

Coloration à l'hématoxyline et à l'éosine (H&E)

Les coupes de tissus ont été colorées au H&E avec le Multistainer automatisé (Leica Biosystems) et ont été imagées avec le scanner de diapositives Aperio AT2 (Leica Biosystems) avec un objectif 20×.

Immunofluorescence

Les sections de tissu ont été perméabilisées et bloquées avec du BSA/PBS Tween 20 0,1 % avant d'être incubées pendant une nuit avec l'anticorps primaire et 30 minutes avec l'anticorps secondaire. L'imagerie a été réalisée à l'aide d'un microscope confocal à balayage laser (TCS Sp8, Leica Microsystems) avec un objectif 10×, 40× ou 63×. Les images ont été analysées à l'aide du logiciel FIJI (NIH, Bethesda, Maryland, USA).

Immunohistochimie

Les coupes de tissus ont été traitées comme suit : récupération de l'antigène avec du tampon citrate, blocage avec du BSA/PBS 10% Tween 20 0,1%, blocage de l'activité peroxydase endogène par incubation dans du H_2O_2 3%.

Après incubation avec l'anticorps primaire, un anticorps secondaire a ensuite été appliqué, suivi d'une révélation avec le kit Dako EnVision®+, Peroxydase. Les tissus ont ensuite été contre-colorés avec de l'hématoxyline. Les images ont été acquises sur un scanner de diapositives Leica Aperio équipé d'un objectif 40×.

II.3. Filtrage des variants

Après l'étape de contrôle de qualité des variants, 7 563 variants ont été identifiés et annotés. Parmi ceux-ci, 681 étaient nouveaux selon la version 150 de dbSNP.

Nous nous sommes d'abord intéressés aux 51 variants codants et avons conservé les variants non synonymes avec une fréquence allélique <20% et coségrégeant avec la maladie dans au moins une famille (c'est-à-dire présent chez les deux sujets atteints séquencés et absent chez le sujet sain séquencé) (Figure 26A et Tableau 15). Au terme de ce filtrage, deux variants situés dans le gène FREM2 (FRAS1-Related Extracellular Matrix 2) ont été conservés pour des analyses ultérieures : un variant fréquent, rs9603422 (R1840W) et un variant rare, rs201714905 (R727H).

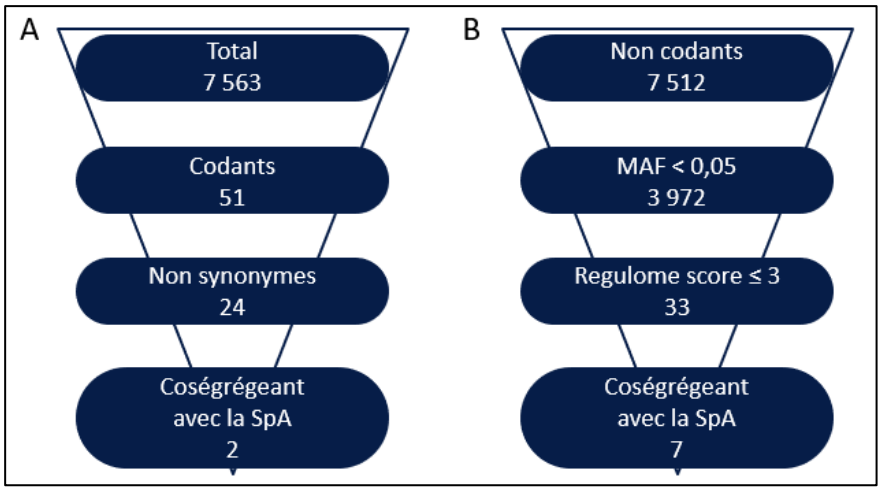

Figure 26. Schéma de filtrage (A) des variants codants et (B) des variants non codants identifiés dans la région cible (GRCh38p12, chr13 :38125001-39527000).

Tableau 15. Caractéristiques des 24 variants non synonymes identifiés par le reséquençage de la région ciblé 13q13.

CHR	Position	Gène	AA	SNP	MAF NFE	CADD
13	38688594	FREM2	L417P	rs571694324	9.85E-05	11.8
13	38689472	FREM2	R710C	rs41292753	0.0041	8.01
13	**38689524**	**FREM2**	**R727H**	**rs201714905**	**0.0003**	**20.2**
13	38690553	FREM2	F1070S	rs2496425	0.2671	12.9
13	38691426	FREM2	F1361S	rs199715291	0.0002	22.5
13	38692335	FREM2	A1664V	rs149779771	0.001	23.1
13	38692347	FREM2	R1668H	rs1868463	0.0502	24.2
13	38692415	FREM2	S1691R	rs966489514	0	27.3
13	**38769685**	**FREM2**	**R1840W**	**rs9603422**	**0.1173**	**24.3**
13	38848487	FREM2	R2066C	rs9548505	0.1048	27.2
13	38850116	FREM2	T2153S	rs9548506	0.1046	3.34
13	38856177	FREM2	T2326I	rs9548509	0.6971	15.9
13	38859500	FREM2	R2477W	rs143571375	0.0014	22.9
13	38878873	FREM2	V2968I	rs116099212	0.0062	15.5
13	38880315	FREM2	T3013M	rs114400765	0.0033	23.1
13	38966959	STOML3	R239C	rs369846460	0.0001	27.9
13	38970370	STOML3	V102I	rs34402894	0.0008	23.3
13	39012842	PROSER1	P804T	rs41286143	0.0255	17.1

13	39013540	PROSER1	V571A	rs3751379	0.1534	0.65
13	39013676	PROSER1	T526A	rs148669353	0.0199	0.032
13	39044165	NHLRC3	L221R	rs9603498	0.0197	28.9
13	39047796	NHLRC3	A305V	rs41286951	0.0045	24.2
13	39047828	NHLRC3	D316Y	rs149175958	0.0035	23.3

CHR : chromosome ; AA : changement d'acide aminé ; NFE : Européens non finlandais ; MAF : fréquence de l'allèle mineur ; SNP : polymorphisme mono nucléotidique

Les caractères gras indiquent les variants co-ségrégeant avec la SpA dans les familles.

Nous nous sommes ensuite intéressés aux 7 512 variants non codants et régulateurs. Ils ont été filtrés en utilisant une MAF ≤ 0,05 et un score RegulomeDB ≤ 3, laissant 33 variants [218]. Nous avons ensuite sélectionné les variants qui étaient présents dans au moins une famille chez les deux patients séquencés et il restait 16 variants après avoir supprimé ceux présents chez les sujets sains ou absents chez les patients (Figure 26B et Tableau 16).

Tableau 16. Caractéristiques des 33 variants non codants avec un score RegulomeDB ≤3 identifiés par le reséquençage de la région ciblée 13q13.

Gène	**Catégorie**	**SNP**	**Score RegulomeDB**	**MAF NFE**	**CADD**
LINC00571 ; UFM1	Intergénique	rs76924815	3a	0.0058	1.2
LINC00571 ; UFM1	**Intergénique**	**rs73460206**	**3a**	**0.0244**	**0.359**
UFM1	**Intronique**	**rs74868151**	**3a**	**0.0003**	**2.87**
UFM1	**Intronique**	**rs10507470**	**3a**	**0.0027**	**2.59**
UFM1 ; LINC00437	**Intergénique**	**rs75764747**	**2a**	**0.0004**	**0.738**
UFM1 ; LINC00437	**Intergénique**	**rs75885484**	**2b**	**0.0002**	**0.992**
UFM1 ; LINC00437	**Intergénique**	**rs17057980**	**3a**	**0.0002**	**0.172**
LINC00437	**ncRNA_intronic**	**rs11841861**	**2b**	**0.0068**	**4.09**
LINC00437	**ncRNA_intronic**	**rs17058288**	**2b**	**0.0076**	**0.118**
LINC00437	**ncRNA_intronic**	**rs79817706**	**2b**	**0.0006**	**1.35**
LINC00437	**ncRNA_intronic**	**rs80103953**	**2b**	**0.0009**	**1.98**
FREM2	**Exonique**	**rs8002488**	**2b**	**0.0009**	**8.65**
FREM2	Intronique	rs115764867	2b	0.0209	5.11

FREM2	Intronique	rs141802208	3a	0.0051	0.568
FREM2	**Intronique**	**rs79240489**	**3a**	**0.0218**	**1.59**
FREM2	**Exonique**	**rs41306662**	**3a**	**0.0216**	**2.06**
FREM2 ; STOML3	Intergénique	rs79506629	3a	0.0327	3.83
FREM2 ; STOML3	**Intergénique**	**rs80121384**	**3a**	**0.0228**	**2.48**
STOML3	Intronique	rs77744003	2b	0.0339	0.190
STOML3 ; PROSER1	Intergénique	rs78684046	3a	0.0407	0.955
PROSER1	Intronique	rs12428338	2a	0.0409	10.3
NHLRC3 ; LHFPL6	Intergénique	rs77435895	3a	0.0346	6.15
NHLRC3 ; LHFPL6	Intergénique	rs74048418	3a	0.0294	1.34
NHLRC3 ; LHFPL6	Intergénique	rs75611908	3a	0.0293	0.05
NHLRC3 ; LHFPL6	**Intergénique**	**rs77958976**	**3a**	**0.0209**	**11.8**
NHLRC3 ; LHFPL6	Intergénique	rs79466993	2b	0.0413	3.13
LHFPL6	Intronique	rs73467356	3a	0.0268	2.01
LHFPL6	Intronique	rs60526930	3a	0.0268	0.123
LHFPL6	Intronique	rs2875298	2b	0.0417	2.04
LHFPL6	Intronique	rs144931287	3a	0.0035	0.095
LHFPL6	Intronique	rs79452404	3a	0.0385	10.1
LHFPL6	Intronique	rs9576813	2c	0.0095	0.01
LHFPL6	**Intronique**	**rs12385888**	**2b**	**0.0033**	**3.90**

Les caractères gras indiquent les variants présents chez les patients et absents chez les sujets sains.

Le score de classement de regulomeDB se réfère aux types de données disponibles pour une seule coordonnée :

- 2a : Liaison TF + motif TF apparié + empreinte DNAse appariée + pic de DNAse
- 2b : Liaison TF + tout motif + empreinte DNAse + pic de DNAse
- 2c : Liaison TF + motif TF apparié + pic de DNAse
- 3a : liaison TF + tout motif + pic de DNAse

II.4. Analyses de ségrégation

Nous avons ensuite génotypé les variants d'intérêt chez l'ensemble des membres des familles dans lesquelles ils avaient été identifiés afin de vérifier si ces variants coségrégeaient avec la

maladie. Les résultats de ces analyses sont résumés dans le tableau 17.

Concernant les 16 variants non codants et régulateurs restants, l'analyse de ségrégation a montré que 7 d'entre eux coségrégeaient avec la maladie dans au moins une famille.

En ce qui concerne le codant, le variant R1840W ségrégeait avec la SpA dans 8 familles sur 12 tandis que le variant R727H n'a été trouvé que dans une seule famille, qui était la famille la plus liée au locus 13q13 dans notre précédente analyse de liaison. Tous les patients de cette famille sauf un étaient porteurs du variant R727H (Figure 27). Deux des porteurs de ce variant étaient homozygotes. À l'inverse, un homme non malade HLA-B27 négatif âgé de 53 ans (id=1334) au moment de l'inclusion était aussi porteur. À noter que deux patients porteurs du variant, qui étaient des frères nés d'une mère homozygote, étaient HLA-B27 négatifs (id=1339 et 1337).

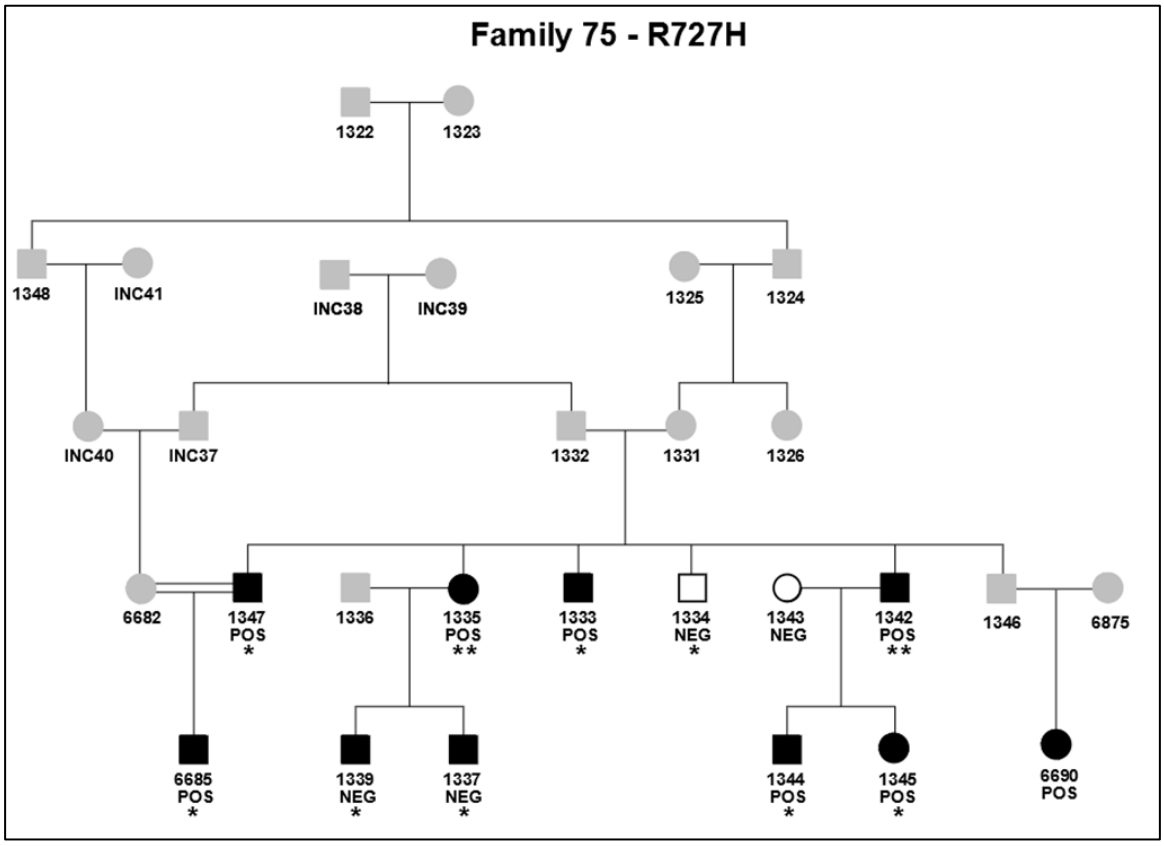

Figure 27. Pedigree de la famille #75 et ségrégation du variant R727H. Tous les membres sont identifiés par un numéro d'identification. Les carrés et les cercles correspondent aux hommes et aux femmes, respectivement. Les symboles pleins et vides correspondent aux patients et aux sujets sains, respectivement. Les symboles gris correspondent aux individus pour lesquels l'ADN n'était pas disponible. Le statut HLA-B27 ("POS" ou "NEG") est également indiqué et la présence du variant R727H est spécifiée par * et ** pour les

hétérozygotes et les homozygotes, respectivement. Les deux barres parallèles entre 6682 et 1335 indiquent la consanguinité.

Tableau 17. Synthèse des analyses de ségrégation pour les variants codants et non codants.

Variant	**MAF**	**Nombre de familles où le variant est présent**	**Nombre de familles où le variant coségrège avec la SpA**
Codants			
rs9603422	0,119	12 familles	8 familles
rs201714905	0,0001	1 famille	1 famille
Non codants et régulateurs			
rs11841861	0,006	5 familles	3 familles
rs17058288	0,007	6 familles	3 familles
rs80103953	0,0007	2 familles	1 famille
rs8002488	0,0007	2 familles	1 famille
rs79240489	0,02	4 familles	1 famille
rs41306662	0,02	4 familles	1 famille
rs77958976	0,02	3 familles	1 famille

II.5. Analyses d'association cas – témoins

Les résultats de l'analyse cas-témoins des variants sont résumés dans le Tableau 18.

La fréquence de l'allèle mineur du variant R1840W n'était pas significativement différente entre les cas et les témoins génotypés (0,109 et 0,101, respectivement ; valeur P = 0,21). Il n'y avait pas non plus d'association préférentielle de ce variant avec HLA-B27 (p=0.37) (Tableau 19).

Le variant codant R727H n'a été retrouvé chez aucun des 969 patients atteints de SpA ni chez les 366 sujets non affectés génotypés. Ce variant n'a pas non plus été détecté dans le projet Exome Français (FrEx), incluant 643 individus sains de 6 régions françaises [223].

Concernant les variants non codants, les MAF de deux premiers variants du tableau étaient significativement différentes entre les 1000 cas génotypés de notre cohorte et les 34000 témoins séquencés de GnomAD, mais similaires dans la même cohorte française (FrEx) de 643 individus. Les fréquences des allèles mineurs des 5 autres variants non codants et régulateurs n'étaient pas significativement différentes entre les cas et les témoins (Tableau 18).

Tableau 18. Analyse d'association cas-témoins.

Variant	Nbre d'individus génotypés (cas/témoins)	MAF (Cas/témoins)	Valeur P	MAF FrEx
Codant				
rs9603422	1091/399	0,109/0,101	0,21	0,114
rs201714905	969/366	0/0	-	0
Non codants et régulateurs				
rs11841861	1010/34013	0,01/0,0062	0,005	0,01
rs17058288	1024/34012	0,01/0,007	0,002	0,01
rs80103953	1010/34014	0,0005/0,0007	0,52	0
rs8002488	1015/34019	0,0005/0,0007	0,52	0
rs79240489	1015/34006	0,02/0,02	0,8	0,02
rs41306662	1015/34018	0,02/0,02	0,82	0,02
rs77958976	1018/34012	0,0176/0,02	0,55	0,02

* Les valeurs p correspondent aux MAF des cas et des témoins de notre cohorte pour les variants codants et de GnomAD pour les variants non codants et régulateurs.

Tableau 19. Fréquence allélique du variant R1840W en fonction de la maladie et du statut HLA-B27

HLA-B27	Groupe	MAF	Nbre d'individus
Positif	SpA	0,27	595
	Témoins	0,26	116
Négatif	SpA	0,25	264
	Témoins	0,2	210

II.6. Immunomarquage de FREM2

Comme mentionné précédemment (II.4. Analyses de ségrégation), le variant R727H, fortement prévalent chez les malades de la famille la plus liée au locus 13q13 appartient au

gène FREM2. Ce gène code pour une protéine qui fait partie d'un groupe de protéines appelé FRAS/FREM et qui joue un rôle important dans la stabilisation et l'intégrité des épithéliums [224]. Plusieurs mutations de ce gène sont associées au syndrome de Fraser, un syndrome malformatif caractérisé par une cryptophtalmie, une syndactylie et des anomalies des voies respiratoires et urogénitales [225].

Nous avons interrogé la base de données GTEx (Genotype-Tissue Expression) et constaté que FREM2 est fortement exprimé dans la thyroïde, les nerfs et les reins et à un niveau moindre dans le pancréas, l'estomac, les muscles, les poumons et le côlon [226]. Le gène n'est pas exprimé dans les 13 types de cellules immunitaires étudiés dans la base de données DICE (Database of Immune Cell Expression, Expression quantitative trait loci (eQTLs) and Epigenomics) [227].

Nous avons décidé de réaliser des expériences d'immunomarquage de FREM2 chez le rat et l'homme afin de préciser l'expression de FREM2 dans le côlon et étudier sa présence dans d'autres tissus pertinents pour la SpA.

Chez le rat, l'expression de FREM2 a été détectée dans le rein avec un fort marquage glomérulaire, comme attendu, mais aussi dans le colon, la membrane synoviale et le cartilage (Figure 28).

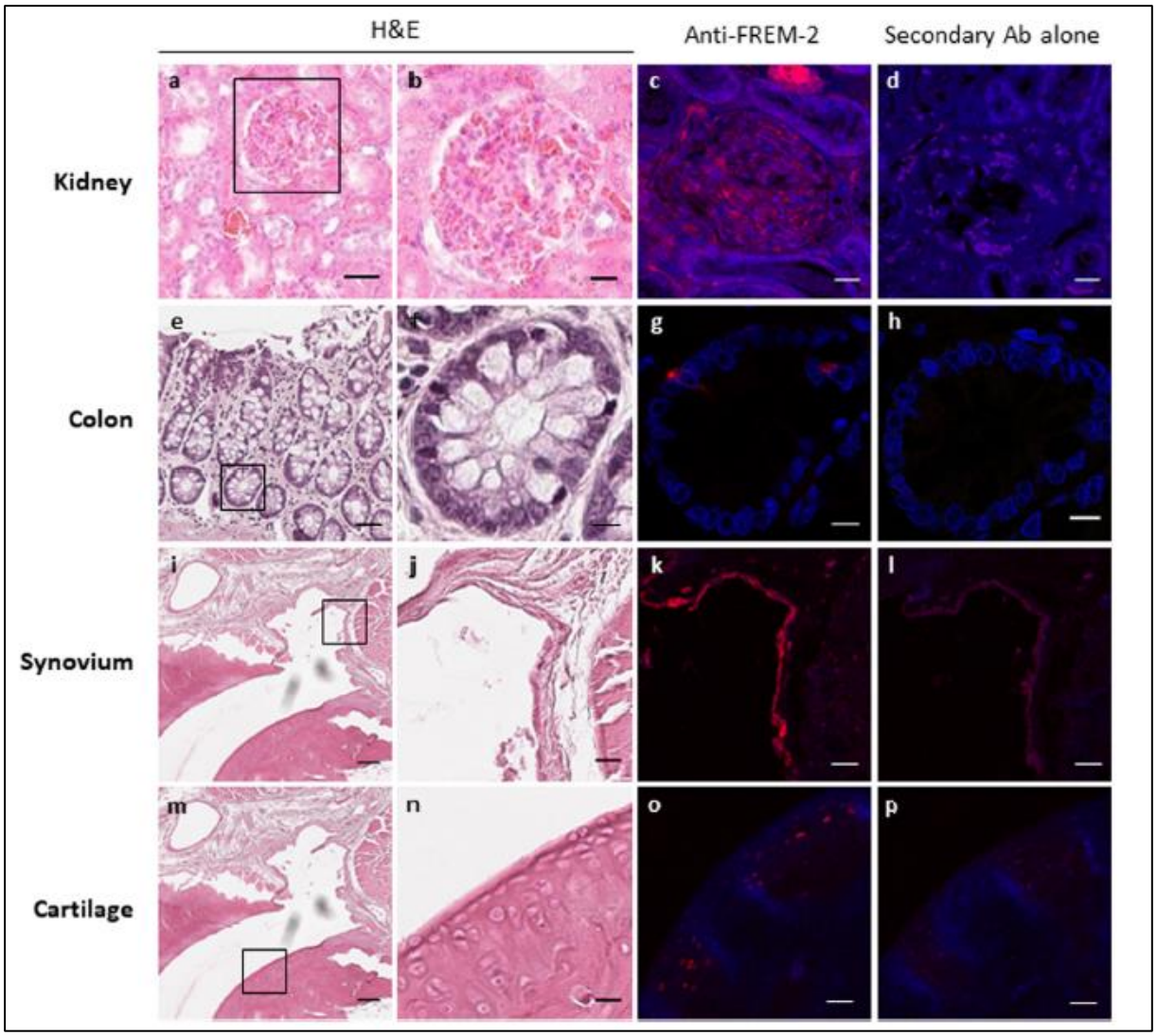

Figure 28. Expression de FREM2 dans les biopsies de rat. Coloration H&E (a, b, e, f, i, j, m, n) et immunofluorescence avec des anticorps FREM2 antihumains de lapin et des anticorps IgG anti-lapin couplés à l'Alexa Fluor594 (c, g, k, o), ou anticorps secondaires seuls (d, h, l, p) de rein (a - d), de colon (e - h), de la membrane synovial (i - l) et de cartilage (m - p) de rat. Les figures b, f, j et n correspondent à un agrandissement plus important de la zone quadrillée dans les figures a, e, i, et m, respectivement. Les noyaux sont colorés avec du DAPI dans les figures c, d, g, h, k, l, o et p. Les sections ont été montées dans du ProlongTM Diamond et examinées au moyen de la microscopie confocale. Barres d'échelle : 50 m pour a ; 20 m pour b, c, d, j, k, l, n, o, p ; 100 m pour e, i, m ; 10 m pour f, g, h.

Chez l'homme, nous avons également trouvé l'expression de FREM2 dans le colon, la membrane synoviale et le cartilage (Figure 29).

Dans le côlon, FREM2 était exprimé dans l'épithélium, dans de rares cellules qui exprimaient également la chromogranine A, un marqueur des cellules entéroendocrines. Dans l'articulation d'un patient atteint de SpA, FREM2 a été détecté dans les chondrocytes et fortement sur la synoviale.

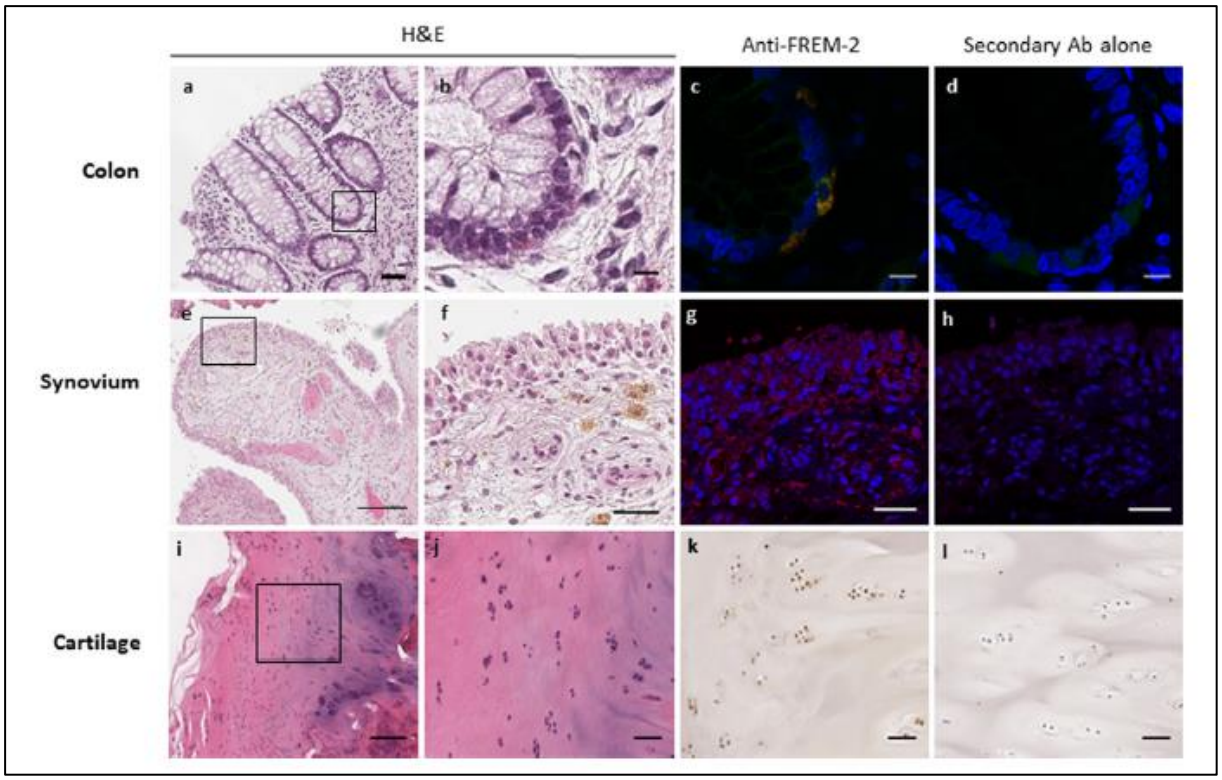

Figure 29. Expression de FREM2 dans les biopsies humaines. Coloration H&E (a, b, e, f, i, j), immunofluorescence (c, d, g, h) et immunohistochimie (k, l). Pour le côlon (c, d), double marquage avec un anticorps monoclonal chromogranine A anti-souris suivi d'un anticorps IgG1 anti-souris couplé à un Alexa Fluor 594 et avec un anticorps FREM2 anti-lapin humain suivi d'un anticorps IgG anti-lapin couplé à un Alexa Fluor 488. Pour la synovie (g, h), marquage avec un anticorps FREM2 antihumain de lapin suivi d'un anticorps IgG anti-lapin couplé à l'Alexa Fluor 594. Pour le cartilage (k, i), marquage avec l'anticorps primaire de lapin antihumain FREM2 suivi d'un anticorps de chèvre anti-lapin IgG couplé à la peroxydase. Les flèches désignent les chondrocytes colorées positivement (k). Les noyaux ont été colorés avec du DAPI (c, d, g, h). Les sections ont été montées dans ProlongTM Diamond et examinées par microscopie confocale. Barres d'échelle : 50 m pour a, j, k, l ; 10 m pour b, c, d ; 200 m pour e, l ; 40 m pour f, g, h.

II.7. Discussion

Malgré les résultats négatifs, nous ne pouvons écarter l'implication des variants non codants sur cette étude, car nous n'avons regardé que les SNPs or il est possible que les autres variants non codants et non couverts par la méthode utilisée dans cette étude, comme les CNV ou les autres types de SVs, soient impliqués.

Le partage du variant R1840W dans les familles peut s'expliquer par le hasard, car le variant est très fréquent dans la population.

Nous n'avons pas pu trouver le variant R727H en dehors de la famille dans laquelle il a été découvert. Par ailleurs, peu d'études publiées ont identifié une association entre la SpA et des

variants rares, la plupart d'entre elles étant basées sur des familles. Toutes les analyses cas-témoins non apparentées réalisées par la suite n'ont pas réussi non plus à valider l'association de ces variants rares avec la SpA.

De plus, dans les autres familles séquencées, nous n'avons pas été en mesure d'identifier des variants rares non synonymes coségrégeant avec la SpA pour expliquer le signal de liaison restant.

III. Partie II : Le séquençage d'exome

III.1. Introduction

Le développement et la réduction des coûts de séquençage nous ont permis de passer d'une approche ciblée à une approche pangénomique. Dans cette deuxième partie du projet, nous nous sommes intéressés à l'ensemble des régions codantes du génome (exome). L'exome ne représente que trois pourcents du génome humain, mais son étude est appropriée dans le cadre des maladies, car les variants situés dans la partie codante du génome ont un impact délétère potentiel plus facile à interpréter. Comme mentionné précédemment (II.4.3. Priorisation des variants issus du NGS), une des difficultés majeures dans la recherche de variants rares est la difficulté du filtrage des nombreux variants identifiés au travers du séquençage. Afin d'aider à ce filtrage, nous avons à nouveau choisi une approche familiale. Afin de limiter les coûts de séquençage, nous n'avons séquencé qu'un nombre limité de patients par famille en sélectionnant les membres avec le plus faible degré d2'apparentement pour minimiser le partage d'allèle par hasard.

Nous avons donc séquencé l'exome entier de 48 malades venant de 20 familles multiplex afin d'identifier d'autres variants rares potentiellement délétères associés à la SpA.

III.2. Patients et méthodes

III.2.1. Echantillons biologiques

III.2.1.1. ADN

Les patients ont été recrutés comme décrit précédemment (I.1. Patients).

Pour le choix de patients à séquencer, nous avons privilégié les patients les plus éloignés possible dans les familles multiplex afin de minimiser la probabilité qu'un variant soit partagé par hasard (Figures 30 et 31). Pour cela, nous avons calculé le coefficient d'apparentement entre chaque paire de malades au sein d'une même famille à l'aide du logiciel KING.

Nous avons donc séquencé deux patients par famille pour la plupart des familles et entre trois et cinq patients pour les familles ayant les structures les plus complexes (par exemple une double entrée de la maladie) (Figure 32).

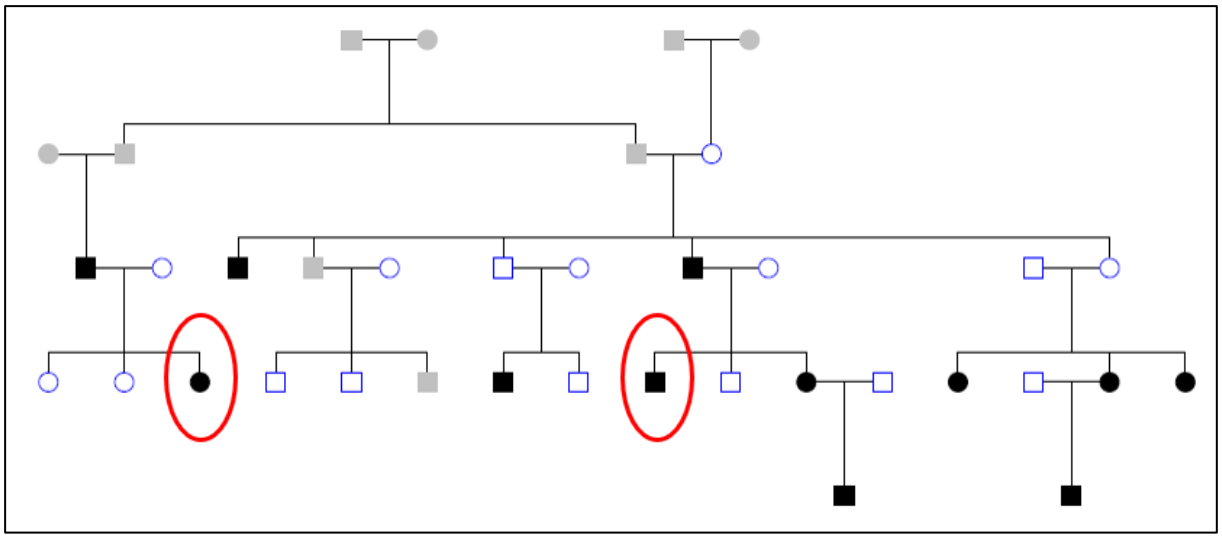

Figure 30. Exemple de choix des patients à séquencer dans une famille multiplex. Les deux patients séquencés sont encerclés en rouge. Les carrés et les cercles correspondent aux hommes et aux femmes, respectivement. Les symboles pleins et vides correspondent aux patients et aux sujets sains, respectivement. Les symboles gris correspondent aux individus pour lesquels l'ADN n'était pas disponible.

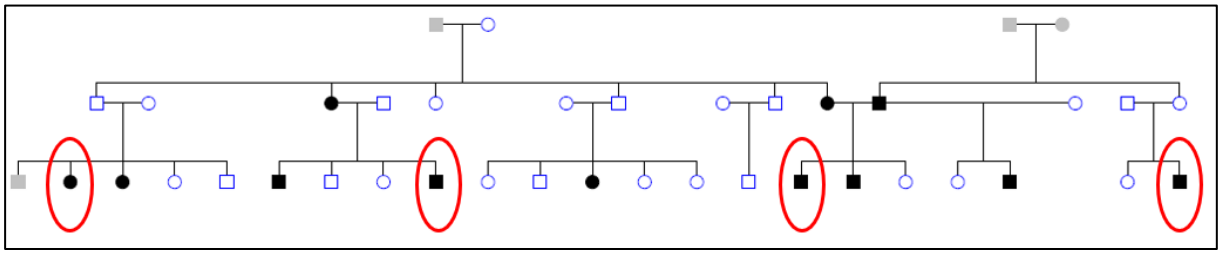

Figure 31. Exemple de choix des patients à séquencer dans une famille multiplex avec deux entrées de la maladie. Les quatre patients séquencés sont encerclés en rouge. Les carrés et les cercles correspondent aux hommes et aux femmes, respectivement. Les symboles pleins et vides correspondent aux patients et aux sujets sains, respectivement. Les symboles gris correspondent aux individus pour lesquels l'ADN n'était pas disponible.

Figure 32. Répartition du nombre de patients séquencés par famille.

Les caractéristiques cliniques des patients inclus dans les études familiales et cas-contrôles sont résumées dans le Tableau 20.

Tableau 20. Caractéristiques des patients inclus dans les études familiales et cas-contrôles.

Caractéristiques	**Séquencés (N = 48)**	**Analyse de ségrégation (N = 134)**	**Etude d'association (N = 1091)**
Age début symptôme : moyenne ± écart-type (année)	49,7±12,8	42,5±14,3	46,4±10,6
Durée de la maladie : moyenne ± écart-type (année)	25,6±11,7	20,5±18	18,9±10
Sexe ratio (% hommes)	60,5	54,5	48,2
HLA-B*27 positif (%)	92,9	93,3	70
Manifestations axiales (%)			
Lombalgie inflammatoire	90,1	98,5	87,5
Sacro-iliite radiographique	67,6	59,7	40,4
Manifestations périphériques (%)			
Arthrites périphériques	49,2	42,5	36
Enthésites périphériques	78,8	70,1	72,7
Manifestations extra-articulaires (%)			
Uvéites	32,3	25,4	22,2
Psoriasis	29,5	25,4	25,5
MICI	12,6	6,7	7,3

III.2.2. Séquençage d'exome

L'ADN génomique a été extrait à partir du sang total des 48 patients venant de 20 familles multiplex. Nous avons utilisé le kit TruSeq DNA Exome d'Illumina. L'ADN (100 ng) a été fragmenté par voie enzymatique et traité pour générer des librairies de séquençage comme recommandé par Illumina. Les fragments d'ADN ont été élués et séquencés par paires à l'aide d'un séquenceur Nextseq500 (Illumina).

III.2.2.1. Génotypage

Les différentes méthodes de génotypage utilisées pour les études de ségrégation et cas-

témoins sont résumées ci-dessous (Tableau 21).

Tableau 21. Techniques de génotypage utilisées en fonction des variants

Technique	Variant
Sanger	rs200885264
HRM	rs144093498
	rs149102151
	rs189947178
	rs138381817

III.3. Filtrage des variants

Nous avons identifié au total 425 593 variants. Parmi ceux-ci, 91 897 étaient des variants codants et 39 817 avaient une MAF inférieure à 0,05 dont 6 192 variants non synonymes. Parmi ces variants non synonymes, 3 786 étaient prédits comme étant délétères par au moins un des algorithmes de prédiction fonctionnelle suivants : SIFT, Polyphen-2 et LRT (Figure 33). Parmi ces variants, nous avons sélectionné ceux présents dans au moins deux familles et partagés par les individus séquencés d'une même famille dans au moins une famille.

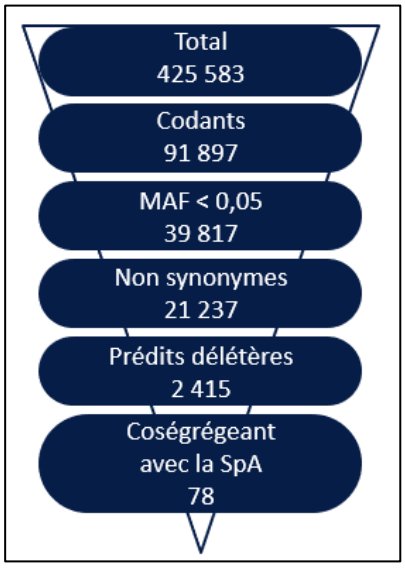

Figure 33. Schéma de filtrage des variants identifiés dans le séquençage d'exome.

III.4. Etude des variants d'intérêt

Afin de vérifier la ségrégation de ces 78 variants avec la maladie chez l'ensemble des malades des familles dans lesquelles chacun d'entre eux avaient été identifiés, nous nous sommes

intéressés sur l'étude des plus grandes familles en nous appuyant sur les résultats de l'analyse de liaison précédemment publiée pour nous aider à sélectionner des variants candidats.

Nous nous sommes focalisés sur deux variants non synonymes les plus intéressants, car ils ont été découverts à proximité de région de liaison dans deux familles avec un grand nombre de patients.

- Le variant rs144093498, localisé dans le gène STIM2 sur le chromosome 4
- Le variant rs200885264, localisé dans le gène ENDOG sur le chromosome 9

III.4.1. Famille 33 et le variant rs144093498 (STIM2*S2195F)

III.4.1.1. Caractéristiques du variant et résultats antérieurs

Le variant rs144093498 est un variant codant, non synonyme et délétère correspondant à une mutation d'une cytosine en une thymine, ce qui se traduit par la mutation de l'acide aminé sérine en phénylalanine (TCC/TTC) à la position 2195 de la protéine STIM2 (S2195F). C'est un variant rare retrouvé à une fréquence de 2/10000 dans la population NFE. La protéine STIM2 joue un rôle important dans le flux d'ions calcium entrant dans la cellule.

Ce variant a été mis en évidence chez les quatre malades d'une même famille inclus dans l'étude d'exome (famille 33). De façon intéressante, les données de liaison dont nous disposions pour cette famille indiquaient un pic de liaison suggestif au niveau du chromosome 4 à proximité de ce variant.

III.4.1.2. Analyse de ségrégation

Le génotypage de ce variant a été réalisé par la méthode HRM et ensuite une analyse de ségrégation entre le variant et la SpA a été réalisée (Figure 34). Nous avons observé la présence du variant chez 8 individus malades sur 12 et chez 9 individus sains sur 13 dans la famille 33.

Les résultats observés dans cette analyse de ségrégation représentée dans la Figure 34 indiquent ainsi qu'il n'y a pas de coségrégation entre le variant STIM2*S2195F et la SpA.

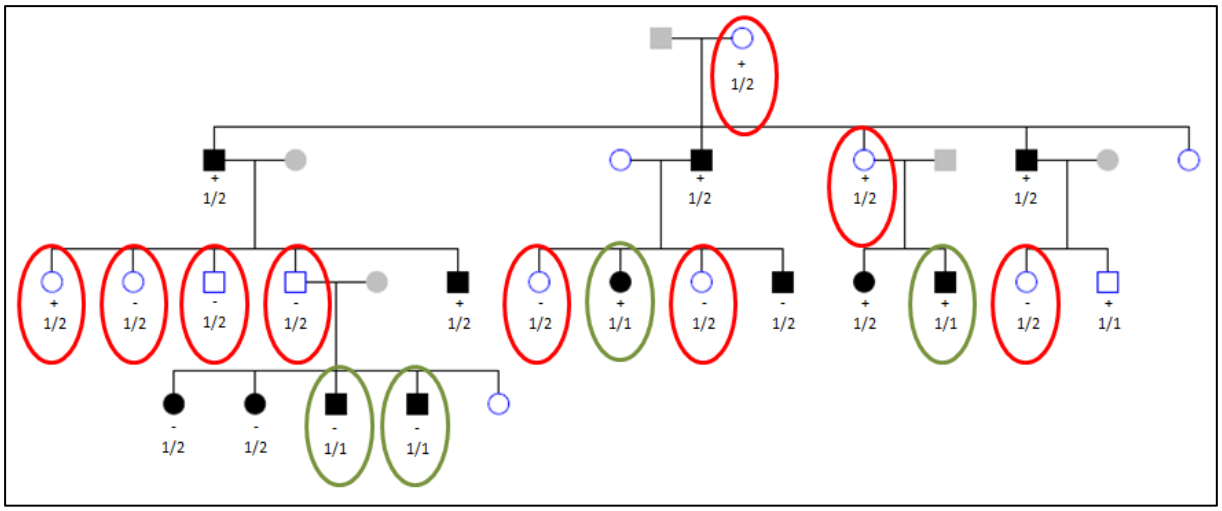

Figure 34. Pedigree de la famille #33 et ségrégation du variant S2195F. Les carrés et les cercles correspondent aux hommes et aux femmes, respectivement. Les symboles pleins et vides correspondent aux patients et aux sujets sains, respectivement. Les symboles gris correspondent aux individus pour lesquels l'ADN n'était pas disponible. Le statut HLA-B27 ("+" ou "-") est également indiqué et la présence du variant S2195F est spécifiée par 1/2 et 2/2 pour les hétérozygotes et les homozygotes, respectivement tandis que l'absence du variant est indiqué par 1/1. Les patients non porteurs du variant sont encerclés en vert et les porteurs sains sont encerclés en rouge.

III.4.2. Famille 110 et le variant rs200885264 (ENDOG*K66N)

III.4.2.1. Caractéristiques du variant et résultats antérieurs

Le variant rs200885264 est un variant non synonyme et délétère correspondant à une mutation d'une guanine en une cytosine (AAG/AAC), ce qui se traduit par la mutation de l'acide aminé lysine en asparagine à la position 66 de la protéine ENDOG (K66N). C'est un variant rare retrouvé à une fréquence de 1 % dans la population NFE. La protéine ENDOG est une endonucléase mitochondriale, mais le gène est nucléaire, qui joue un rôle important dans la réplication de l'ADN mitochondrial et l'apoptose.

Ce variant a été retrouvé chez les deux individus malades d'une même famille inclus dans l'étude d'exome (famille 110). Cette famille est celle qui a obtenu le score de liaison le plus élevé dans la région SPA2, région dont une liaison significative à la SpA a été antérieurement rapportée. Ce variant est localisé dans ce pic de liaison du chromosome 9.

III.4.2.2. Analyse de ségrégation

Tous les patients de la famille 110 étaient porteurs hétérozygotes du variant ENDOG*K66N sauf un encadré en vert (Figure 35). A l'inverse, quatre témoins sains étaient aussi porteurs de

ce variant (encadrés en rouge).

Les résultats de ces analyses de ségrégation représentée dans la Figure 35 montrent que le variant K66N coségrège avec la maladie dans la famille 110 avec une pénétrance incomplète.

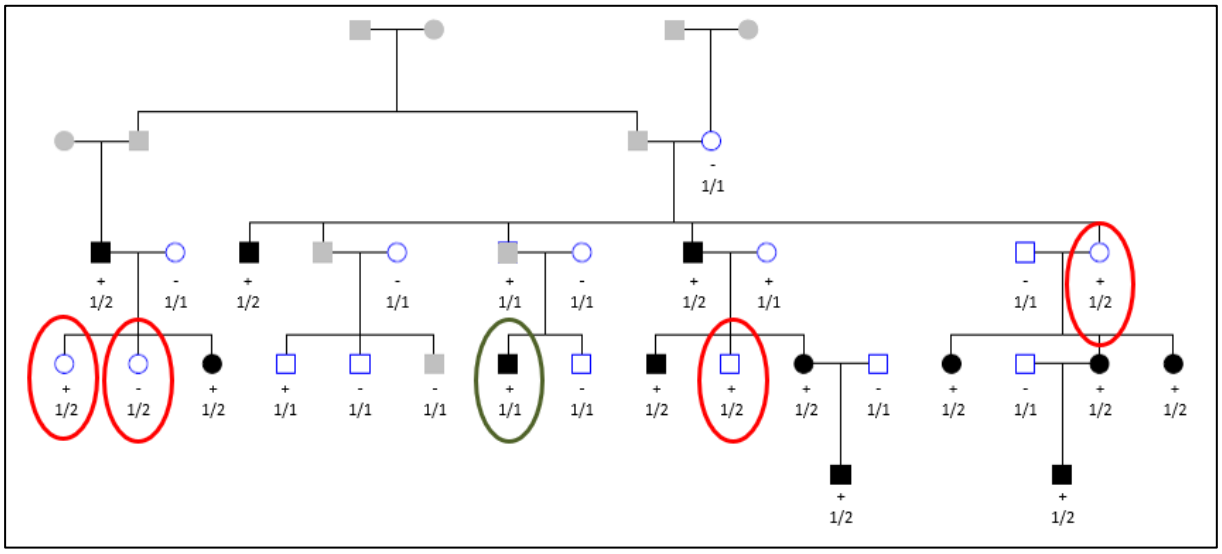

Figure 35. Pedigree de la famille #110 et ségrégation du variant K66N. Les carrés et les cercles correspondent aux hommes et aux femmes, respectivement. Les symboles pleins et vides correspondent aux patients et aux sujets sains, respectivement. Les symboles gris correspondent aux individus pour lesquels l'ADN n'était pas disponible. Le statut HLA-B27 ("+" ou "-") est également indiqué et la présence du variant K66N est spécifiée par 1/2 et 2/2 pour les hétérozygotes et les homozygotes, respectivement tandis que l'absence du variant est indiqué par 1/1. Le patient non porteur du variant est cerclé en vert et les porteurs sains sont cerclés en rouge.

Après avoir obtenu ces résultats encourageants sur le variant ENDOG*K66N, nous avons décidé de poursuivre les analyses en comparant la MAF chez les patients et les témoins.

III.4.2.3. Analyses d'association cas – témoins du variant ENDOG*K66N

Nous avons génotypé le variant ENDOG*K66N chez 460 patients et les résultats sont représentés dans le Tableau 22. La fréquence de l'allèle mineur de ce variant n'était pas significativement différente entre les cas génotypés et les témoins de la base de données GnomAD et du projet FrEx.

Tableau 22. Analyse d'association cas-témoins du variant ENDOG*K66N.

Variant	**Gène**	**Cas Génotypés**	**Cas porteurs**	**MAF (cas)**	**MAF (GnomAD_NFE)**	**MAF FrEx**
K66N	ENDOG	460	11	0,01	0,01	0,01

Ces résultats indiquent que le variant ENDOG*K66N n'est pas le variant causal même si on le trouve dans le locus SPA2. Nous avons décidé d'étudier d'autres variants proches du variant ENDOG*K66N afin d'identifier le variant causal.

Pour cela, nous avons d'abord souhaité reconstruire les haplotypes des individus de cette famille afin d'avoir toutes les informations sur les partages des haplotypes et donc des variants.

III.5. Reconstruction des haplotypes des patients de la famille 110

Afin d'aller plus loin et comprendre le partage de variants dans la famille 110, nous avons décidé de reconstruire les haplotypes autour du variant ENDOG*K66N grâce à des données de puce de génotypage Affymetrix, déjà disponibles au laboratoire. Les résultats de cette reconstruction des haplotypes montrent 5 événements de recombinaison entre les positions 125 Mb et 128 Mb (Figures 36 et 37). Ces recombinaisons entraînent un non-chevauchement des haplotypes des 5 patients.

D'un côté, les patients 1450, 1487 et 1441 avaient trois informations distinctes de recombinaisons conservant la partie proximale de l'haplotype morbide (groupe 1) et de l'autre côté, les patients 1498 et 7026 avaient deux informations distinctes de recombinaisons conservant la partie distale (groupe 2) (Figures 36 et 37). On constate que l'haplotype du patient 1450 est incompatible avec ceux des patients 1498 et 7026 sous l'hypothèse d'un gène unique à l'origine de la maladie dans cette famille. Cela soulève la possibilité d'une phénocopie due à un second gène situé dans la même région.

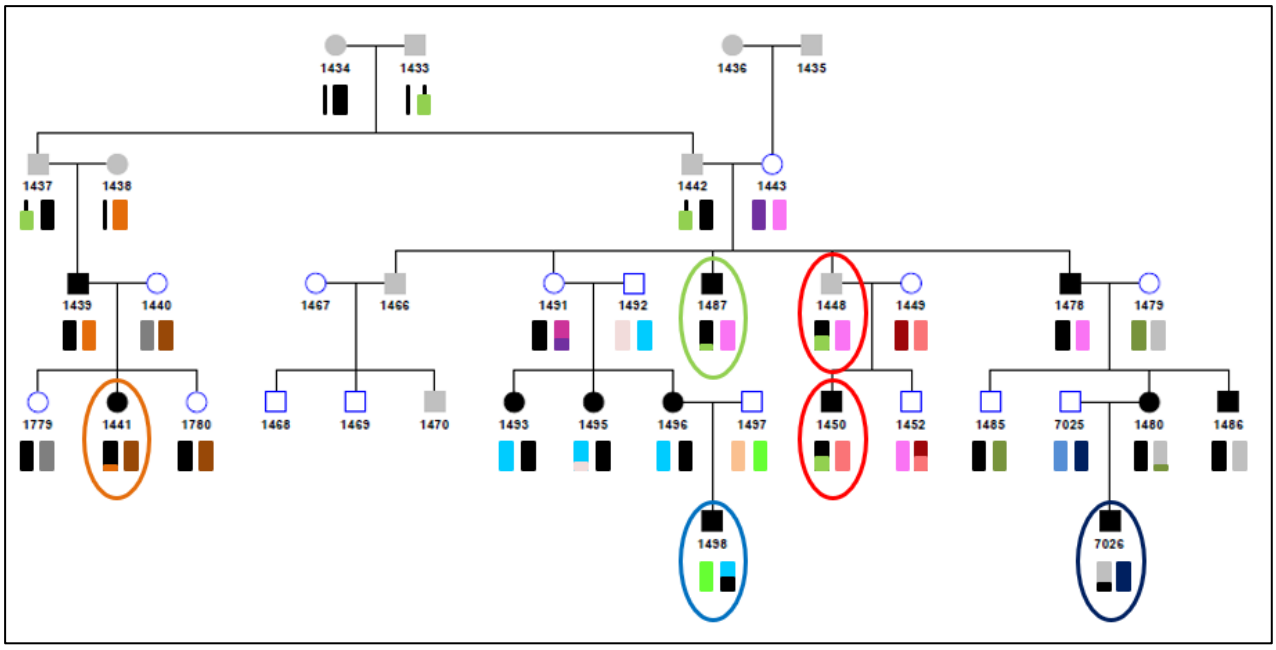

Figure 36. Reconstruction des haplotypes dans la famille 110. Tous les membres sont identifiés par un numéro. Les carrés et les cercles correspondent aux hommes et aux femmes, respectivement. Les symboles pleins et vides correspondent aux patients et aux sujets sains, respectivement. Les symboles gris correspondent aux individus pour lesquels l'ADN n'était pas disponible. Les haplotypes sont représentés en barre de couleur et l'haplotype délétère est représenté en barre de couleur noir. Les patients ayant les recombinaisons informatives sont encerclés en orange, rouge, vert, bleu clair et bleu foncé selon les événements de recombinaison.

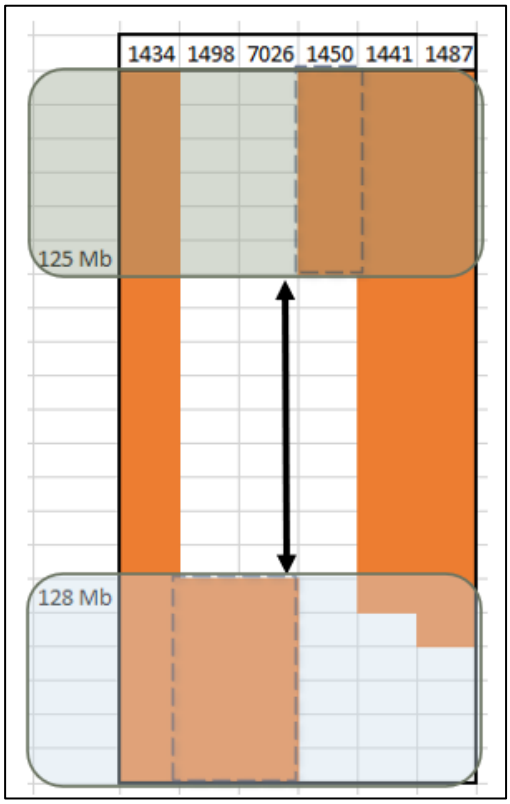

Figure 37. Les haplotypes des patients de la famille 110 ayant des événements de recombinaison. Les patients sont identifiés par un numéro. La couleur orange représente la partie de l'haplotype ségrégeant avec la maladie. Les deux carrés transparents indiquent les deux haplotypes proximal et distal non chevauchants.

Nous avons donc repris l'analyse des variants au niveau du chromosome 9.

III.6. Filtrage des variants du chromosome 9

Sachant que le gène ENDOG est situé sur la partie distale du bras long du chromosome 9, nous avons décidé d'effectuer le filtrage des variants du chromosome 9 et de nous focaliser pour commencer sur ceux partagés par les deux patients séquencés de la famille 110.

Nous avons identifié 17301 variants au total sur le chromosome 9. Parmi ces variants, 3299 étaient partagés par les deux patients séquencés de la famille 110 (Figure 38). Ensuite, nous avons filtré les 576 variants codants et 262 variants non synonymes. Enfin, nous avons gardé 8 variants non synonymes (7 variants + ENDOG*K66N) qui avaient une MAF < 0,05 pour les analyses ultérieures (Figure 39).

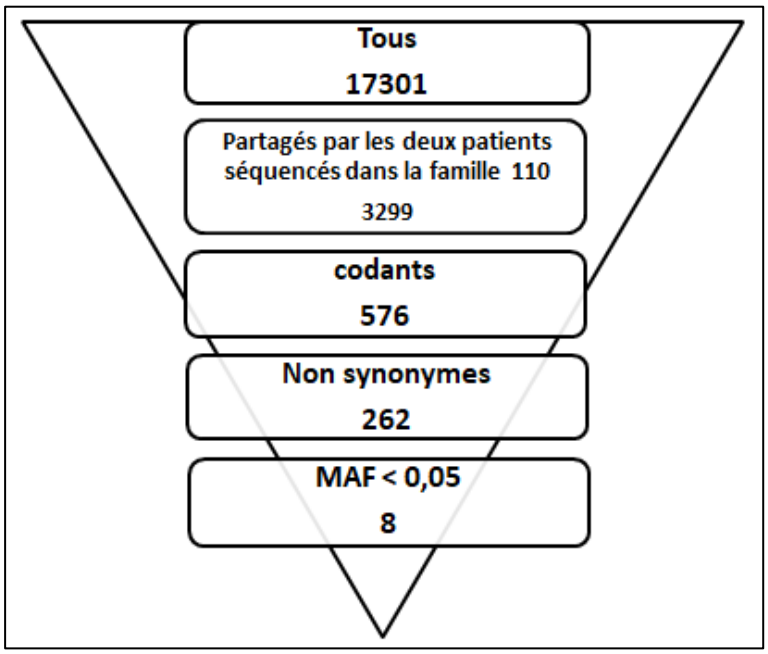

Figure 38. Schéma de filtrage des variants identifiés dans le séquençage d'exomes au niveau du chromosome 9.

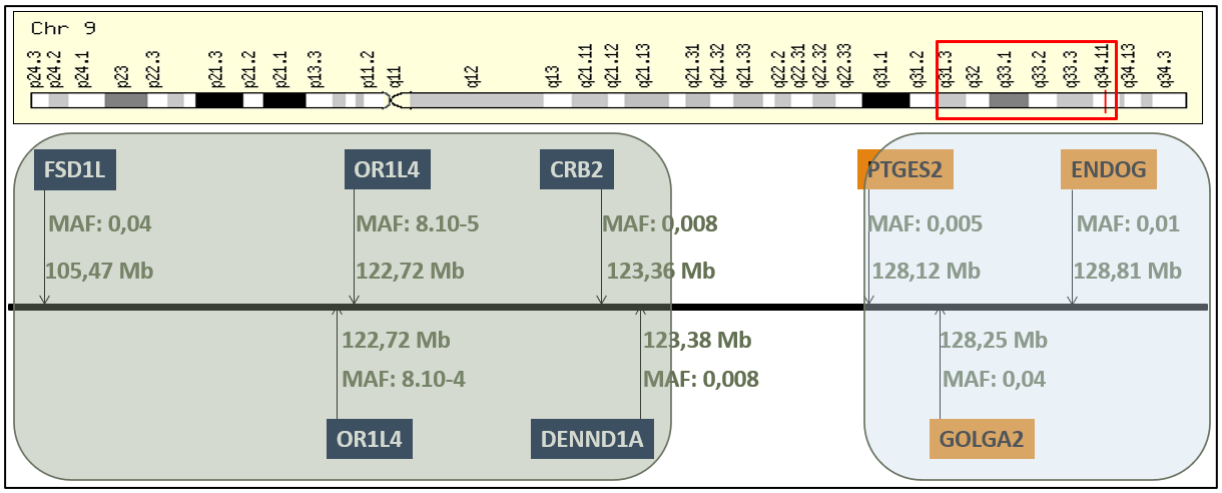

Figure 39. Schéma du chromosome 9 avec les positions des variants proches du variant ENDOG*K66N. Les recombinaisons sont situées entre les deux carrés transparents entre les gènes DENND1A et PTGES2 (125 Mb et 128 Mb). La position de chaque variant est indiquée par une flèche et en mégabases (Mb). La zone où les variants d'intérêt sont localisés sur le chromosome 9 est encadrée en rouge.

III.7. Analyses de ségrégation des variants proches du variant ENDOG*K66N

Le variant rs28503319 du gène FSD1L et les deux variants rs748975619 et rs773750205 du gène OR1L4 n'ont pas été confirmés par le séquençage Sanger. Il s'agissait en fait des faux positifs.

Les variants rs189947178 du gène DENND1A, rs138381817 du gène CRB2 et rs149102151 du gène PTGES2, en revanche, ont été bien confirmés par séquençage Sanger et ont pu faire l'objet d'analyses de ségrégation dans la famille 110. Le variant rs189947178 entraîne un changement de l'acide aminé Alanine à l'acide aspartique en position 784 de la protéine DENND1A (A784D), le variant rs138381817 entraîne un changement de l'acide aminé Arginine en Histidine en position 93 de la protéine CRB2 (R93H) et le variant rs149102151 entraîne un changement de l'acide aminé Glycine en Alanine en position 206 de la protéine PTGES2 (G206A).

Nous avons obtenu les mêmes résultats pour les variants DENND1A*A784D et CRB2*R93H. Tous les patients de la famille 110 étaient porteurs hétérozygotes de ces deux variants sauf deux patients encadrés en bleu (Figure 40). A l'inverse, quatre témoins sains étaient aussi porteurs de ces variants (encadrés en rouge).

Concernant le variant PTGES2*G206A, nous avons obtenu les mêmes résultats que ceux du variant ENDOG*K66N. Tous les patients de la famille 110 étaient porteurs hétérozygotes des deux variants sauf un patient encadré en vert (Figure 35). A l'inverse, quatre témoins sains

étaient aussi porteurs de ces variants (encadrés en rouge).

Les analyses de ségrégation montrent que ces variants ségrégent avec la maladie dans la famille 110 avec une pénétrance incomplète comme le variant ENDOG*K66N.

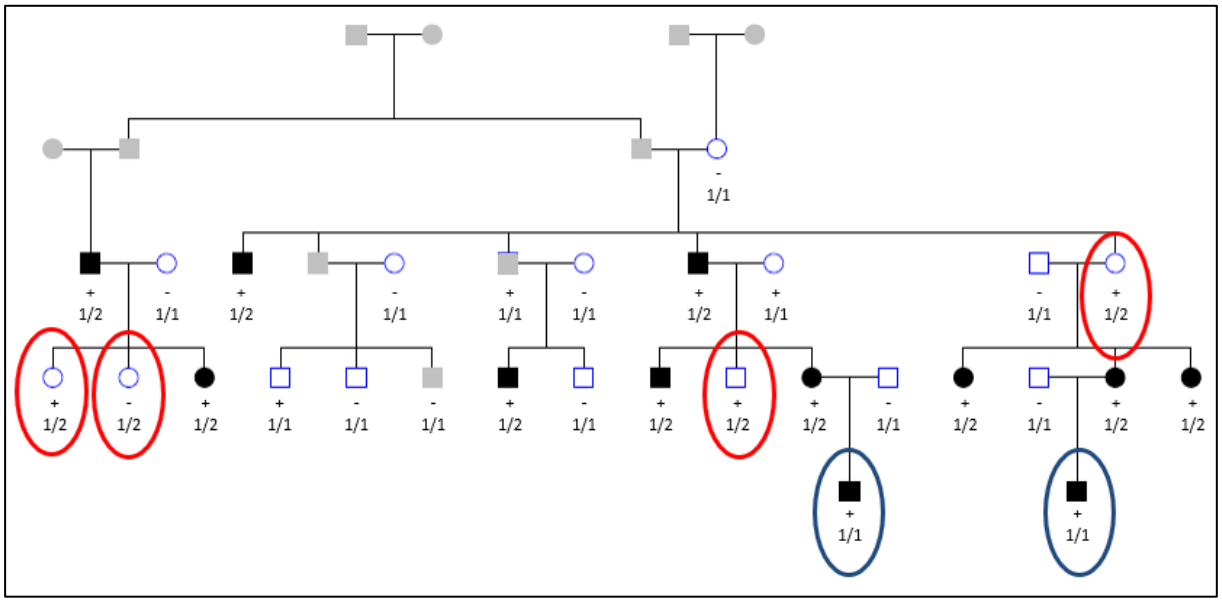

Figure 40. Pedigree de la famille #110 et ségrégation des variants DENND1A*A784D et CRB2*R93H. Les carrés et les cercles correspondent aux hommes et aux femmes, respectivement. Les symboles pleins et vides correspondent aux patients et aux sujets sains, respectivement. Les symboles gris correspondent aux individus pour lesquels l'ADN n'était pas disponible. Le statut HLA-B27 ("+" ou "-") est également indiqué et la présence des variants A784D et R93H est spécifiée par 1/2 et 2/2 pour les hétérozygotes et les homozygotes, respectivement tandis que l'absence de ces variants est indiqué par 1/1. Le patient non porteur des variants est encerclé en bleu et les porteurs sains sont encerclés en rouge.

III.8. Analyses cas – témoins des variants proches du variant ENDOG*K66N

Nous avons génotypé les trois variants proches du variant ENDOG*K66N. Le variant GOLGA2*R762S n'a pas été analysé, car entre temps nous avons obtenu les séquences du génome entier.

Les résultats sont représentés dans le Tableau 23. La fréquence de l'allèle mineur de ces variants n'était pas significativement différente entre les cas génotypés et les témoins de la base de données GnomAD.

Tableau 23. Analyse d'association cas-témoins des variants proches du variant K66N.

Variant	AA	Gène	Cas Génotypés	Cas porteurs	MAF (cas)	MAF (gnomAD_NFE)
rs189947178	A784D	DENND1A	644	11	0,008	0,008
rs138381817	R93H	CRB2	552	11	0,009	0,009
rs149102151	G206A	PTGES2	920	16	0,008	0,006

III.9. Discussion

Aucun des variants étudiés dans cette partie du projet n'étaient impliqué dans la susceptibilité de la SpA car ils n'étaient pas associés à la maladie soit dans l'étude de ségrégation soit dans les analyses cas-témoins. Deux raisons peuvent expliquer cela : soit un autre variant codant présent dans la région, mais non détecté dans ce projet faute de couverture, soit un variant non codant.

Nous n'avons pas étudié d'autres variants d'intérêt, car nous avons obtenu entre-temps les résultats de la troisième partie du projet, le génome entier. Les analyses des régions dans lesquelles ces variants ont été découverts seront donc poursuivies dans la partie génome entier avec un nombre de patients plus important.

IV. Partie III : Le séquençage génome entier

Nous présentons ici les résultats préliminaires de cette troisième partie, car les analyses sont en cours.

IV.1. Introduction

Le séquençage d'exomes a été décevant dans la mesure où les analyses d'association ont montré que les fréquences alléliques de ces variants étaient équivalentes entre les cas et les témoins.

Sachant que dans la stratégie du séquençage d'exomes, nous n'avions pas accès à tous les variants codants ni aux variants non codants, nous avons décidé d'aller plus loin en séquençant le génome entier de plus de patients afin d'identifier les variants rares et potentiellement délétères associés à la SpA.

IV.2. Patients et méthodes

IV.2.1. Echantillons biologiques

IV.2.1.1. ADN

Le choix était donc de séquencer tous les patients de nos grandes familles multiplex. Nous avons séquencé le génome entier de 413 patients issus de 76 familles multiplex avec 4 malades ou plus par famille (Figure 41).

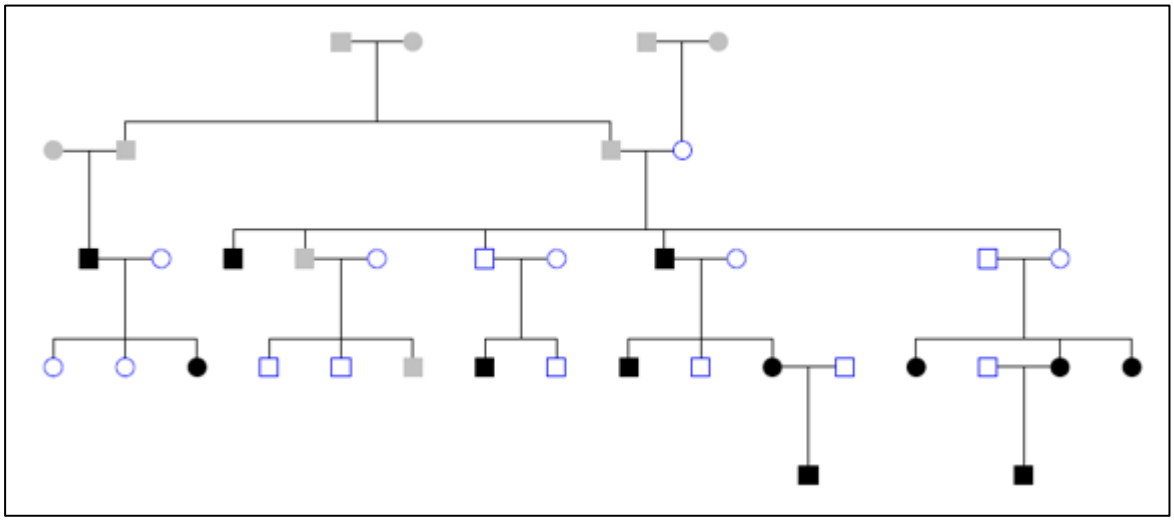

Figure 41. Exemple d'une famille multiplex de 11 patients. Les carrés et les cercles correspondent aux hommes et aux femmes, respectivement. Les symboles pleins et vides correspondent aux patients et aux sujets sains, respectivement. Les symboles gris

correspondent aux individus pour lesquels l'ADN n'était pas disponible. Tous les patients ont été séquencés.

Les caractéristiques cliniques des patients inclus dans les études familiales sont résumées dans le Tableau 24.

Tableau 24. Caractéristiques des patients séquencés.

Caractéristiques	**Séquencés (N = 413)**
Age début symptôme : moyenne +/- écart-type (année)	49,7+/- 12,8
Durée de la maladie : moyenne +/- écart-type (année)	25,6+/-11,7
Sexe ratio (% hommes)	52,8
HLA-B*27 positif (%)	82,5
Manifestations axiales (%)	
Lombalgie inflammatoire	89,5
Sacro-iliite radiographique	48,4
Manifestations périphériques (%)	
Arthrites périphériques	39,9
Enthésites périphériques	74,6
Manifestations extra-articulaires (%)	
Uvéites	39,9
Psoriasis	26,6
MICI	6,2

IV.2.2. Séquençage génome entier

L'ADN génomique a été extrait à partir du sang total des 413 patients venant de 76 familles multiplex. Le séquençage du génome entier a été réalisé au centre national de la recherche en génétique humaine (CNRGH). L'ADN (100 ng) a été fragmenté par voie enzymatique et traité pour générer des librairies de séquençage. Les fragments d'ADN ont été élués et séquencés par paires à l'aide d'un séquenceur Novaseq6000 (Illumina).

IV.3. Contrôle qualité de données

Nous avons réalisé un contrôle qualité supplémentaire des données afin d'éliminer les échantillons présentant des discordances de sexe ou de parenté, ou une contamination inter échantillons et des duplicats. Nous avons retiré ainsi 10 échantillons problématiques et nous avons donc gardé 403 patients venant de 76 familles. Du fait de ce nettoyage, quelques familles présentaient moins de 4 patients. Nous les avons donc retirées. A la fin de cette étape, nous avons donc gardé les données des échantillons de 386 patients venant de 68 familles avec 4 patients ou plus.

Nous avons ensuite projeté nos données avec les données du projet 1000 génomes sur une analyse en composante principale afin de confirmer l'origine européenne des familles étudiées (Figure 42).

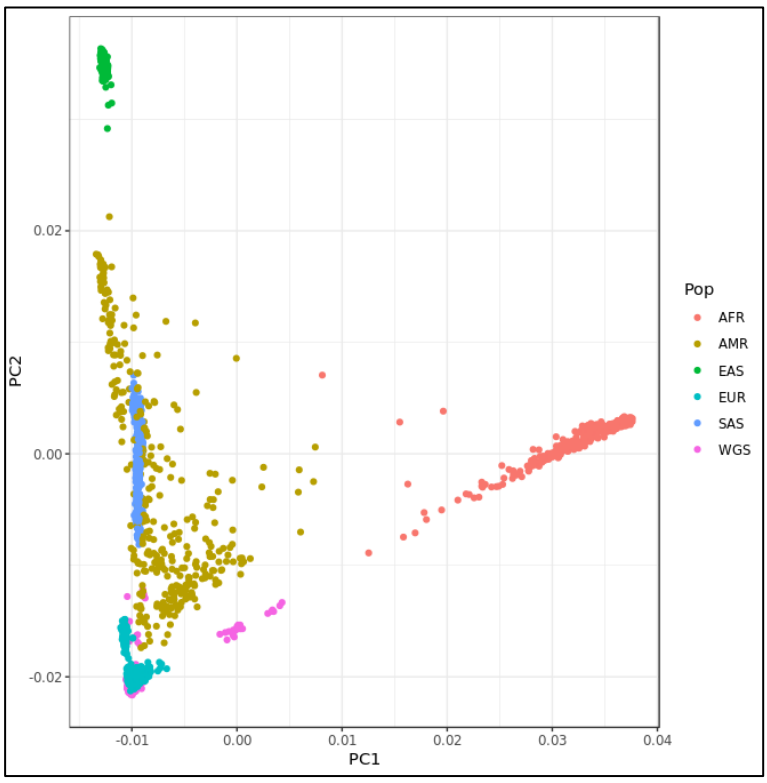

Figure 42. Analyse en composante principale des données du séquençage génome entier. Les populations sont représentées avec un code couleur. Les Africains (AFR) en rouge, les Américains (AMR) en jaune, les Asiatiques de l'Est en vert (EAS), les Européens en bleu ciel (EUR), les Asiatiques du Sud en bleu foncé (SAS) et nos échantillons en fushia (WGS).

La grande majorité des échantillons chevauche les échantillons européens caucasiens. En zoomant cependant, une partie des échantillons chevauche d'autres populations, finlandaise, africaine, latino-américaine (Figure 43). Pour ces patients, nous prendrons en compte les

fréquences alléliques des variants, qui seront identifiés dans ces familles, dans ces populations au moment des analyses cas-témoins.

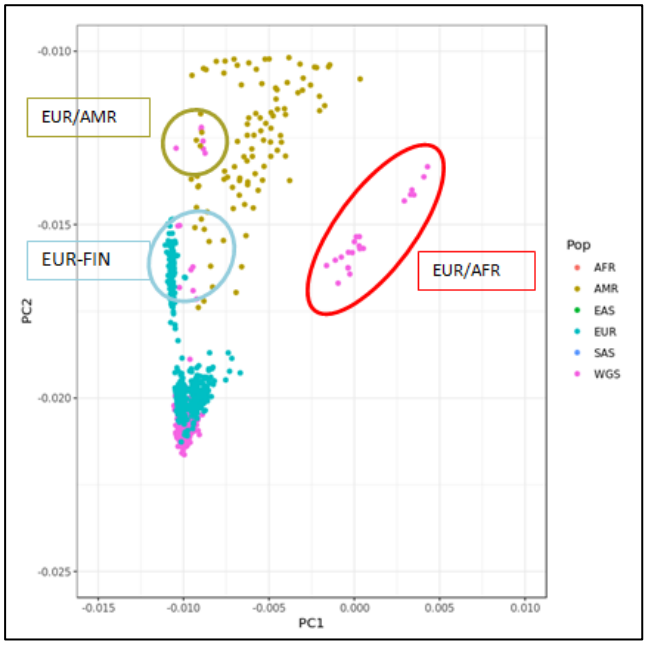

Figure 43. Analyse en composante principale. Les données qui se chevauchent avec d'autres populations sont cerclées avec les codes couleur de ces populations. EUR-FIN : Européens – Finlandais, EUR-AMR : Européens – Américains et EUR-AFR : Européens – Africains.

IV.4. Filtrage des variants

Dans cette partie du projet, notre stratégie de filtrage des variants est différente des deux précédentes, car nous avons déjà les résultats concernant les analyses de ségrégation de tous les variants dans toutes les familles.

Dans un premier temps, nous avons filtré les variants codants d'un côté et les variants non codants de l'autre afin de nous concentrer d'abord sur les variants non synonymes.

Dans un deuxième temps, nous avons utilisé le partage des variants chez tous les membres affectés d'une même famille, la MAF et le caractère délétère des variants pour filtrer les variants non synonymes.

Nous avons identifié au total 15 860 039 SNPs. Parmi ceux-ci, 7 739 620 avaient une MAF inférieure à 0,01 et 471 597 coségrégeaient avec la maladie chez n-1 (n étant le nombre de patients dans la famille) individus dans au moins une famille (Figure 44). Il nous restait 469 968 SNPs après exclusion des variants de la région du CMH.

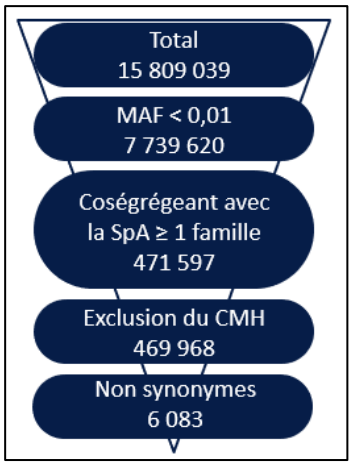

Figure 44. Stratégie de filtrage des SNPs.

IV.4. Caractéristiques des 6083 variants filtrés

IV.4.1. Ségrégation

La majorité de ces variants coségrégeait dans une seule famille comme le montre la Figure 45. Cependant, 332 variants étaient retrouvés dans plus d'une famille.

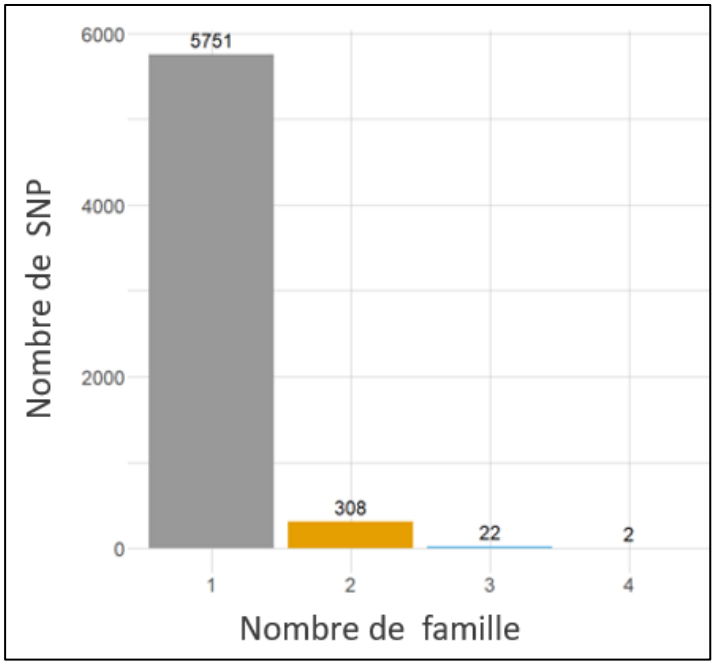

Figure 45. Répartition des variants filtrés en fonction du nombre de familles où le variant coségrège avec la maladie (c'est-à-dire est présent chez au moins n-1 malades, n représentant le nombre de malades séquencés dans la famille).

IV.4.2. Evaluation du potentiel délétère

Pour évaluer la pathogénicité de ces variants d'intérêt, nous avons utilisé le score CADD car

comme décrit dans le chapitre II de l'introduction (II.4.3.3. Prédiction de l'impact fonctionnel), la capacité de CADD à hiérarchiser les variants fonctionnels, délétères et pathogènes dans de nombreuses catégories fonctionnelles est inégalée par toute annotation actuelle.

La majorité des variants avaient des scores CADD inférieurs à 30, 1086 variants avaient un score CADD compris entre 0 et 10, 1124 variants avaient un score CADD compris entre 10 et 20, 2236 variants avaient un score CADD compris entre 20 et 30, 149 variants avaient un score CADD compris entre 30 et 40, 15 variants avaient un score CADD compris entre 40 et 50 et un seul variant avait un score CADD supérieur à 50 (Figure 46).

Habituellement, un score CADD de 10 signifie qu'un variant fait partie des 10 % de variants potentiellement délétères les plus importants du génome humain. Un score CADD de 20 signifie que le variant fait partie des 1 % de variants potentiellement délétères. Un score CADD de 30 signifie que le variant se situe dans le top 0,1 % et ainsi de suite. Toutefois, beaucoup de chercheurs utilisent le score CADD avec un seuil 10.

Le score CADD nous permet essentiellement de trier les mutations causales des mutations non causales. Autrement dit, le score CADD nous permet d'enrichir en variants potentiellement délétères.

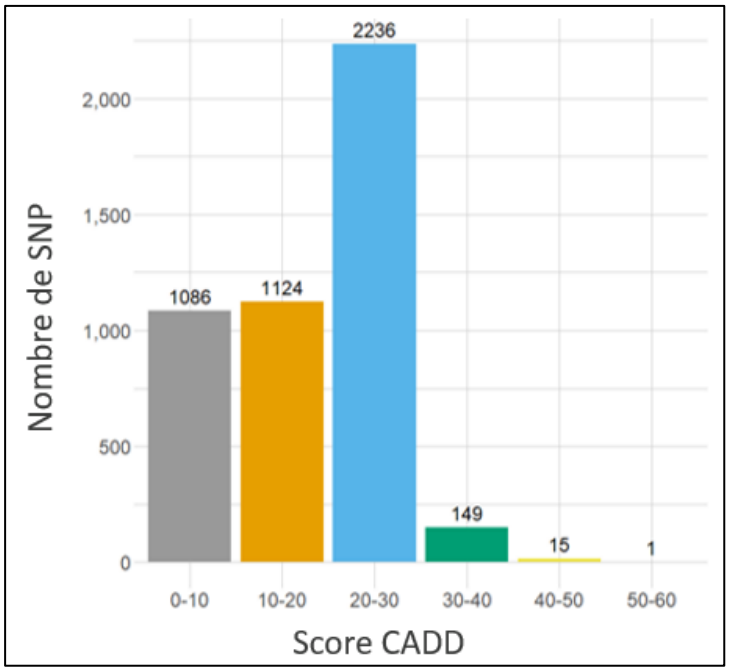

Figure 46. Répartition des variants filtrés en fonction de leur score CADD

IV.6. Annotation fonctionnelle

Nous avons réalisé une étude d'enrichissement de termes de Gene Ontology (GO) à partir de ces variants non synonymes d'intérêt et partagés dans au moins une famille pour rechercher

des voies pertinentes pouvant être impliquées dans la maladie. Les termes GO des processus biologiques sont représentés dans la Figure 47.

Les résultats de l'analyse d'enrichissement GO ont mis en évidence que les voies de la matrice extracellulaire et des cils primaires seraient potentiellement impliquées (Figure 48). Les variants impliqués dans ces voies feront l'objet des analyses plus poussées, car ces voies sont très intéressantes dans le cadre de la SpA puisque les cils primaires sont présents dans presque tous les types de cellules, y compris les chondrocytes. Plusieurs études ont montré que les cils primaires sont nécessaires à la mécanotransduction des chondrocytes et à la synthèse de la matrice extracellulaire associée [228–230].

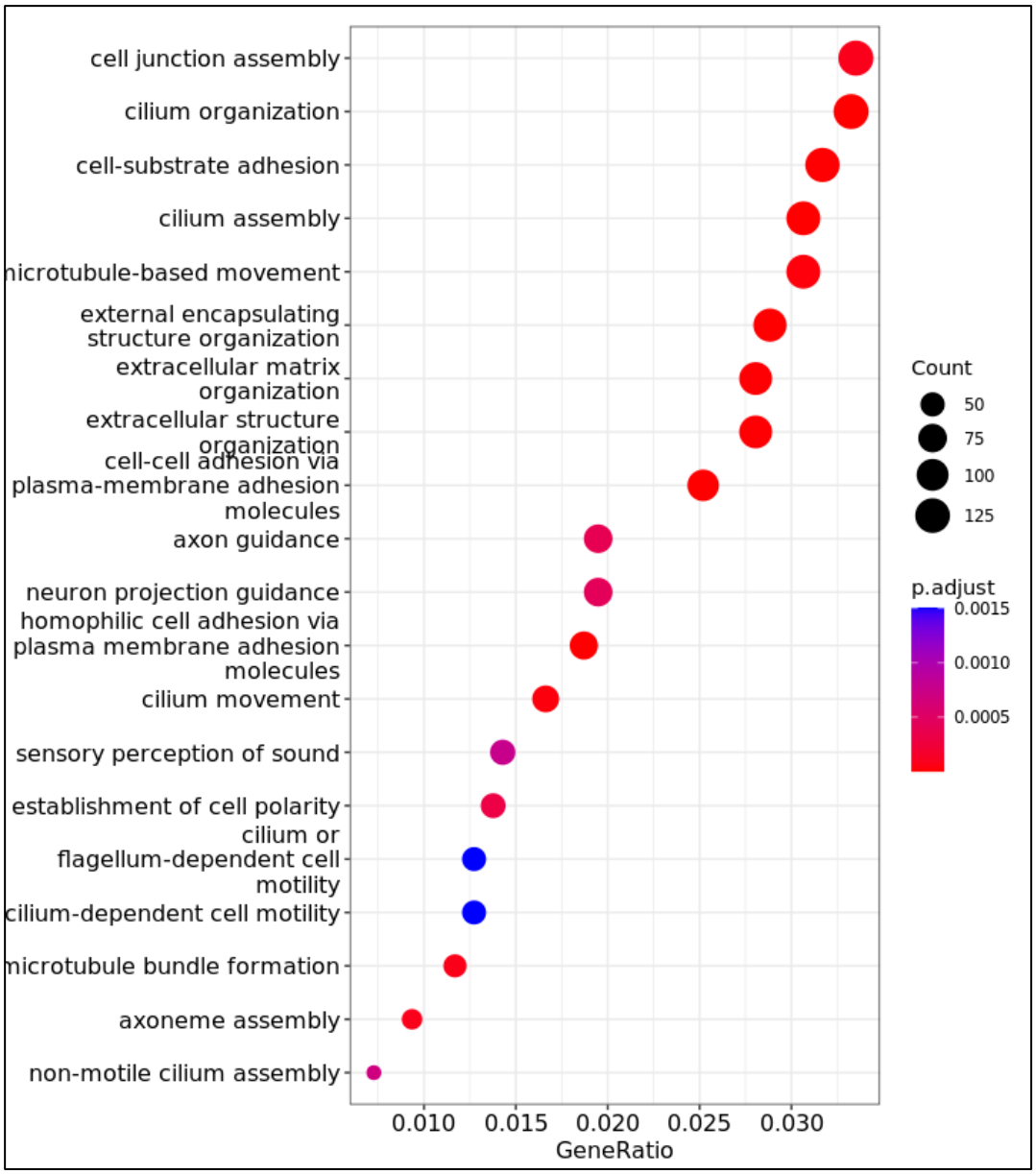

Figure 47. Dot plot des résultats des analyses d'enrichissement GO.

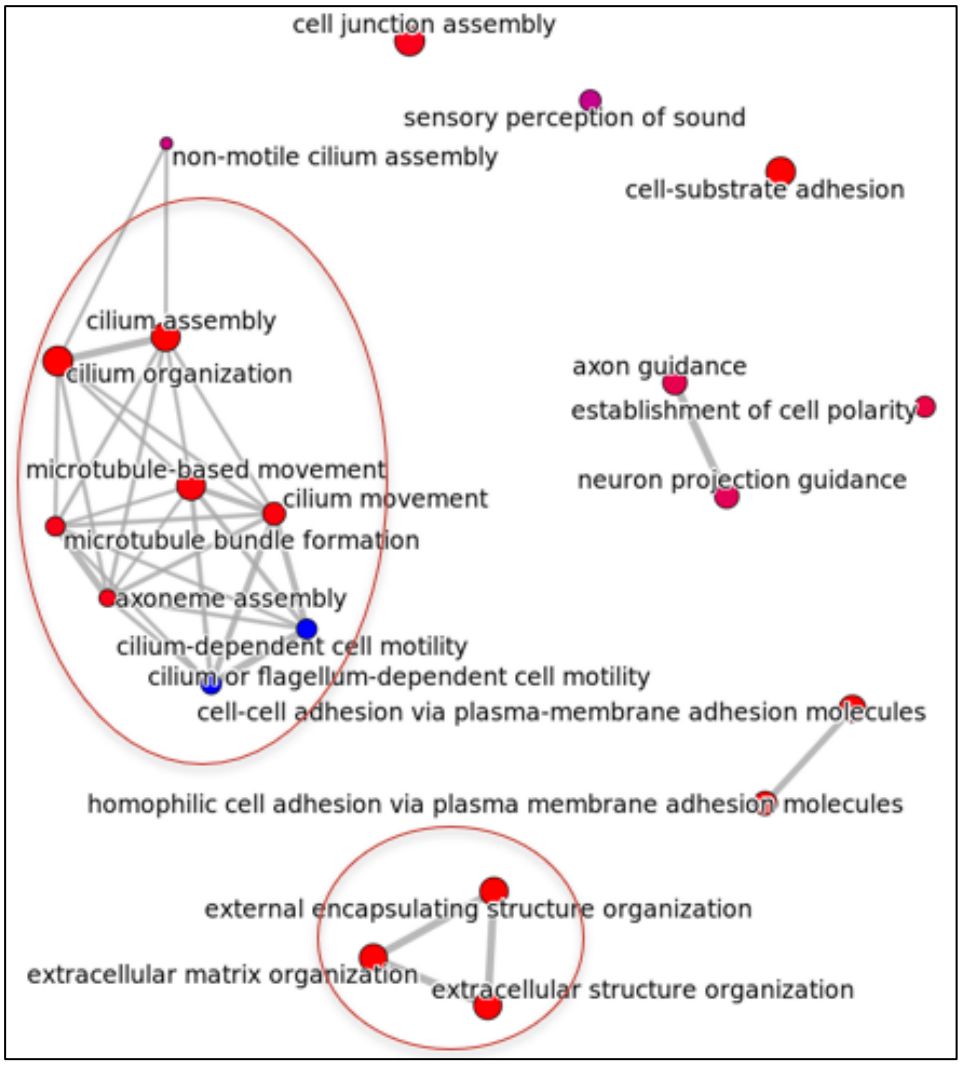

Figure 48. Emap plot des résultats des analyses d'enrichissement GO.

DISCUSSION ET PERPECTIVES

L'objectif de ma thèse était d'identifier de nouveaux variants rares de susceptibilité à la SpA grâce à une approche familiale utilisant le séquençage haut-débit. Le projet a été mis en œuvre en trois étapes successives dans le temps, caractérisées par une proportion croissante du génome séquencé, allant d'une région ciblée de 1,4 Mb jusqu'au génome entier (2800 Mb), en passant par l'exome (35 Mb). Cette extension a été rendue possible par le progrès technologique et le coût décroissant du séquençage haut débit, mais aussi par l'amélioration des méthodes d'analyses des données. La discussion de ces résultats obtenus s'articulera donc autour de ces trois étapes.

I. Le reséquençage du locus 13q13

Dans cette étude, nous avons reséquencé une région de 1,4 Mb sur le chromosome 13, dans un ensemble de familles SpA multiplex génétiquement liées à cette région [222]. Nous nous sommes d'abord concentrés sur les variants non synonymes et avons identifié deux variants dans le gène FREM2 avec des preuves de coségrégation avec la maladie.

Le premier, R1840W, était un variant codant commun avec une MAF de 0,119. Bien que ce variant coségrégeait avec la maladie dans 8 des 12 familles où il était retrouvé, l'absence de son association avec la maladie dans l'analyse cas-témoins, y compris après stratification avec l'allèle B27, n'a cependant pas permis de retenir son implication dans la susceptibilité à la SpA. Sa coségrégation avec la SpA dans ces familles peut s'expliquer par le hasard, car R1840W est un variant fréquent dans la population.

Le second était un variant également non synonyme, mais rare, R727H. Il n'a été trouvé que dans une seule famille qui était la mieux liée au locus 13q13. Ce variant était hautement pénétrant dans cette famille, car tous les membres affectés sauf un étaient porteurs dont deux à l'état homozygote. En revanche, dans l'étude d'association, il n'a été trouvé chez des 1305 autres individus testés (patients SpA ou témoins), ni chez les 600 témoins sains de la base de données FrEx (French Exome). Cependant, étant donné sa MAF (0,0001), notre étude manquait de puissance, avec une probabilité de le détecter de seulement 39%.

Le gène FREM2 code pour une grande protéine transmembranaire de 3169 acides aminés contenant de nombreuses répétitions de l'élément protéoglycane sulfate de chondroïtine

(CSPG) et des domaines Calx-beta. La protéine FREM2 fait partie d'un groupe de protéines appelé FRAS/FREM qui joue un rôle important dans la stabilisation et l'intégrité des épithéliums [224].

La protéine FREM2 est localisée au niveau de la membrane basale de la peau et des épithéliums rénaux. Au niveau de la peau, elle participe à un complexe ternaire avec FRAS1 et FREM1 qui joue un rôle dans les interactions épiderme-derme.

Des mutations de FREM2 ont été associées au syndrome de Fraser, un trouble malformatif caractérisé par une cryptophtalmie, une syndactylie et des anomalies des voies respiratoires et urogénitales. Les phénotypes sont très variables, allant de symptômes mineurs à des malformations létales [225]. FREM2 a également été récemment impliqué dans la progression maligne du glioblastome, avec une expression élevée du transcrit et de la protéine associée à un bon pronostic [231].

R727H est un variant faux-sens substituant une arginine par une histidine dans l'une des douze répétitions de l'élément protéoglycan sulfate de chondroïtine de la protéine (CSPG). Ce variant a un score CADD égal à 20,2, ce qui implique que ce changement d'acide aminé est potentiellement délétère. Il a été récemment démontré que les CSPG jouaient un rôle essentiel de blocage de la transition TLR4 dépendante des cellules immunitaires d'un phénotype inflammatoire à un phénotype « réparateur » dans un modèle de traumatisme de la moelle épinière [232].

Dans le contexte de maladies complexes, on attend des mutations légèrement délétères de préférences à des mutations fortement délétères. De plus, la pathogénicité d'un variant pourrait dépendre de la cooccurrence d'autres variants chez le même individu, comme l'allèle HLA-B27. Il est toutefois intéressant de noter que l'allèle HLA-B27 ne semble pas être indispensable puisque deux patients porteurs de R727H étaient négatifs pour HLA-B27.

Dans ce projet, nous avons aussi montré pour la première fois que FREM2 est exprimé dans des tissus pertinents pour la SpA tels que la membrane synoviale, le cartilage et le colon. L'os et les articulations n'ont pas été inclus dans les grands projets internationaux (tels que GTEx) d'analyse de l'expression des gènes. Dans le côlon, nous avons trouvé une expression de FREM2 dans les cellules entéroendocrines. Ces cellules sont rares, ce qui explique que l'expression de FREM2 dans le côlon ait pu passer inaperçue, et modulent la signalisation des neurones via la sécrétion de neurotransmetteurs, dont la sérotonine [233]. La sérotonine a un large éventail d'effets physiologiques, notamment la régulation de la mobilité intestinale, la

régulation de l'ostéoblastogenèse [234] et le développement et la fonction du système immunitaire [235].

En pathologie, il est intéressant de noter qu'une diminution des taux de sérotonine a été observée dans la SA et qu'elle pourrait être impliquée dans la formation osseuse [236]. Il a également été démontré que les cellules entéroendocrines produisent de l'IL-17C dans le côlon enflammé des patients atteints de la maladie de Crohn [237].

À ce jour, peu d'études ont identifié une association entre la SpA et des variants rares, la plupart d'entre elles étant basées sur des familles (discutées dans [238]). Toutes les analyses cas-témoins non apparentés menées par la suite, y compris la nôtre, n'ont pas réussi à valider l'association de ces variants rares avec la SpA.

Dans notre cas, nous n'avons pas pu trouver le variant R727H en dehors de la famille dans laquelle il a été découvert. Il n'y avait pas non plus de variants associés à la SA dans cette région dans les GWAS publiées précédemment. Cependant, la région n'était pas bien couverte par ces GWAS (seulement 9 SNP dans l'étude IGAS Immunochip avec une valeur p comprise entre 0,22 et 0,94). Il est néanmoins important de signaler ces variants rares, car il est possible qu'ils contribuent à façonner une architecture génétique de la SpA qui serait très hétérogène.

D'autres études génétiques et fonctionnelles sont toutefois nécessaires pour évaluer l'effet pathogène potentiel de ce nouveau variant. En particulier, son génotypage dans des cohortes supplémentaires de patients atteints de SpA serait utile dans la mesure où, comme cela a été évoqué, notre analyse manquait de puissance pour détecter un variant rare comme R727H.

Dans les autres familles séquencées, nous n'avons pas été en mesure d'identifier des variants rares non synonymes coségrégeant avec la SpA pour expliquer le signal de liaison restant. Comme dans la plupart des maladies complexes, les variants régulateurs pourraient jouer un rôle important, mais l'analyse des variants rares non codants du locus 13q13 avec des effets régulateurs prédits en utilisant les outils actuels n'a pas révélé leur implication.

Cependant, nous ne pouvons pas exclure l'implication des variants non codants sur cette étude, car il est également possible que d'autres types de variants génétiques qui n'ont pas été détectés par le séquençage de cette région, en occurrence les variations du nombre de copies ou les grands variants structuraux, jouent un rôle dans la susceptibilité à la SpA.

II. Le séquençage d'exome

Dans cette deuxième partie du projet, nous avons séquencé l'exome complet de 48 patients venant de 20 familles de malades. Nous avons ensuite choisi de poursuivre l'analyse des variants rares par séquençage d'amplicons afin de valider leur existence et de les rechercher chez d'autres patients. Les variants ainsi sélectionnés étaient situés soit dans un gène dont la fonction pouvait avoir un lien avec la maladie soit dans un des pics de liaison observés lors des analyses de liaison effectuées dans le laboratoire avant mon arrivée. Le choix de cibler les fragments à reséquencer autour de ces variants d'intérêt était adapté, car cela permettait aussi d'étudier d'autres variants dans un panel de gènes potentiellement associé à la SpA.

Malheureusement, nous nous sommes heurtés à un taux important de variants faux positifs qui ne nous ont pas permis de continuer avec cette stratégie. L'hybridation/capture était un autre choix possible de séquençage ciblé, mais cette technique a un coût beaucoup plus élevé que celle du séquençage d'amplicons.

Nous nous sommes donc tournés ensuite vers l'étude des plus grandes familles des malades en nous appuyant sur les résultats de l'analyse de liaison précédemment publiée pour nous aider à sélectionner des variants candidats. Nous avons ainsi identifié deux variants non synonymes rares proches de pics de liaison sur les chromosomes 4 et 9.

Le premier est le variant STIM2*S2195F découvert dans la famille 33, au niveau d'un pic de liaison dans le chromosome 4. Ce variant était situé dans le gène STIM2 qui code une protéine transmembranaire jouant un rôle dans l'homéostasie calcique. Il est exprimé dans tous les tissus chez l'homme et il a été récemment montré qu'il participe à la différenciation des cellules T naïves en Th17 [239]. STIM2 a également été proposé comme un acteur pertinent dans les pathologies chroniques liées au vieillissement, maladies d'Alzheimer et de Parkinson, ainsi que dans les maladies auto-immunes et les cancers [240]. L'analyse de ségrégation a cependant montré l'absence de coségrégation de ce variant avec la SpA dans cette grande famille, ce qui rendait peu probable un rôle significatif dans la survenue de la maladie.

Le deuxième variant, ENDOG*K66N, a été découvert dans le gène ENDOG, au niveau d'un pic de liaison du chromosome 9 déjà connu (locus SpA2). Ce variant a été détecté dans la famille la plus liée au locus SPA2.

Le gène ENDOG code une endonucléase localisée dans l'espace intermembranaire de la

mitochondrie et participe à la réplication de l'ADN mitochondrial et au processus d'apoptose indépendant des caspases. Il a été montré que le gène ENDOG est impliqué dans les maladies dégénératives comme la maladie de Parkinson et dans un modèle murin de cardiomyopathie [241].

Les analyses de ségrégation ont démontré que ce variant coségrégeait avec la SpA dans cette famille puisque tous les membres affectés étaient porteurs de ce variant sauf un, qui constituait donc une phénocopie. Cependant, quatre individus étaient des porteurs sains de ce variant dans la famille. Deux d'entre eux avaient peu de risque de développer la maladie, car ils avaient plus de 60 ans alors que la pénétrance de la SpA est de 95% à l'âge de 50 ans. Les deux autres avaient moins de 30 ans et on ne peut pas exclure que la maladie se déclare ultérieurement.

Les études d'associations n'ont cependant pas montré d'association de ce variant avec la SpA dont la MAF était équivalente chez les patients de notre cohorte, mais aussi chez les individus sains des bases de données gnomAD et FrEx. Ces résultats indiquaient que le variant ENDOG*K66N n'était probablement pas causal. On ne pouvait cependant pas exclure qu'un variant d'un gène lié et ségrégeant avec lui dans la famille 110 soit le variant causal.

La reconstruction des haplotypes a mis en évidence cinq événements de recombinaison qui auraient dû permettre de mieux circonscrire la région d'intérêt. Les explications triviales (erreur d'échantillons, ou de phénotypage ou de génotypage) ayant été éliminées, cette discordance suggérait la possibilité de phénocopie mais due dans ce cas à des gènes situés dans la même région chromosomique.

Les analyses de coségrégation ont montré que les variants de gènes proches de ENDOG étaient, comme attendu, aussi bien partagés dans la famille, mais un ou deux patients étaient toujours non-porteur. Les patients non porteurs étaient différents selon que le variant était du côté centromérique ou télomérique de la zone de recombinaison. Toutefois, les analyses d'association cas-témoins de ces variants étaient aussi décevantes que celle du variant K66N du gène ENDOG.

Cette région télomérique du bras long du chromosome 9 est considérée comme paralogue de la région du CMH [242]. Par ailleurs, plusieurs gènes de cette région ont été associés à la prédisposition à la SpA.

Une explication de ces résultats décevants vient peut-être de la nature même des variants dévoilés par le séquençage de l'exome et qui sont essentiellement codants alors que le rôle

de variants non codants est considéré comme prépondérant dans les maladies à hérédité complexe. La possibilité de variants codants, mais situés dans la petite proportion d'exons non couverts par le protocole d'hybridation-capture n'est pas non plus à exclure. Le séquençage du génome entier résout ces deux difficultés.

III. Le séquençage du génome entier

Les analyses des génomes entiers n'en sont qu'à leur début, mais révèlent d'ores et déjà des pistes intéressantes.

Rappelons que nous avons filtré les variants rares, potentiellement délétères, partagés par au moins n-1 patients (n étant le nombre de patients dans la famille) d'au moins une famille et retrouvés dans au moins une autre famille, quel que soit le mode de ségrégation. Comme le montrent les résultats (Figure 45), seuls 332 variants répondaient à ces critères.

De façon intéressante, les études d'enrichissement de termes de Gene Ontology ont révélé l'implication de deux voies pertinentes pour la pathogenèse de la SpA, les cils primaires et la matrice extracellulaire.

Les cils primaires sont présents dans presque tous les types de cellules, et notamment dans les chondrocytes où ils sont nécessaires à la mécanotransduction et à la synthèse de la matrice extracellulaire associée [228–230]. Ils détectent et transmettent des signaux extracellulaires pour réguler divers processus cellulaires au cours du développement et pour maintenir l'homéostasie des tissus.

Nous savons que l'os est composé des plusieurs types des cellules et aussi d'une matrice extracellulaire. Il se renouvelle grâce aux ostéoclastes qui dégradent la matrice osseuse et aux ostéoblastes qui la reconstruisent. L'activité de ces deux types cellulaires doit être coordonnée pour assurer le maintien harmonieux du squelette osseux, et une anomalie dans le fonctionnement de l'un d'eux peut entraîner des conséquences pathologiques.

Il a été ainsi montré que le raccourcissement des cils primaires par la signalisation TGFβ est associé à une transition épithéliale-mésenchymateuse ainsi qu'à une altération de la mécanosensation et de la maturation des ostéoblastes humains [243].

IV. Le rôle des variants rares dans d'autres maladies auto-immunes

Dans la plupart des maladies à hérédité complexe et notamment les maladies auto-immunes, de nombreux variants communs ont été identifiés par des GWAS, mais ils n'expliquent qu'une partie de la composante génétique. La SpA ne fait donc pas exception. Une recherche de variants rares a été entreprise dans la plupart de ces maladies avec des résultats variables, car il est le plus souvent difficile d'apporter une preuve fonctionnelle du rôle délétère de ces variants rares. Compte tenu de l'abondance de la littérature, nous avons choisi de discuter les résultats obtenus dans quatre maladies auto-immunes : la polyarthrite rhumatoïde (PR), le diabète de type 1 (DT1), la sclérose en plaques (SEP) et le lupus érythémateux disséminé (LED).

Dans la PR, plus d'une centaine de variants génétiques ont été associés à la maladie grâce aux GWAS [244,245]. Toutefois, ces variants n'expliquent pas toute l'héritabilité. En effet, l'héritabilité attribuée aux allèles de HLA-DRB1 présentant l'épitope partagé a été estimée entre 11% [246] et 37 % [247] alors que les loci identifiés par les GWAS en dehors du CMH n'expliquent que 5 % supplémentaires [248]. Plusieurs études ont utilisé le séquençage ciblé soit de quelques gènes candidats soit de l'exome entier afin d'identifier les variants rares et d'évaluer leur contribution à l'héritabilité de la PR. Une étude, basée sur une population européenne, a montré une agrégation de variants rares dans les gènes IL2RA et IL2RB [249]. Une autre étude, cette fois-ci basée sur une population coréenne, a mis en évidence une association de variants rares dans 17 gènes différents [250]. D'autres études coréennes ont mis en évidence des variants rares associés à la PR dans plusieurs gènes impliqués dans la production d'espèces réactives de l'oxygène [251] et dans le gène PLB1 [252]. Maëva Veyssiere et al., ont récemment identifié un nouveau variant non-sens associé à la PR dans le gène SUPT20H. Cette étude a été réalisée en couplant l'approche familiale au séquençage de l'exome entier dans 9 familles multiplex européennes [253].

Dans le diabète de type I, plus de 60 loci associés à la maladie ont été identifiés à ce jour grâce aux GWAS [254–256]. La plupart de ces variants sont fréquents dans la population et ont donc des effets modestes [256]. 80 % de la composante génétique du diabète de type I est expliquée [257], les variants HLA représentant la plus grande partie [258]. Les variants rares pourraient expliquer l'héritabilité manquante [258]. Dans le modèle murin de référence

du diabète de type 1, la souche de souris NOD (Non Obese Diabetic), Anne-Perrine Foray et al., ont récemment identifié un variant rare ayant un impact significatif sur le risque de maladie dans le gène DUSP10 [259]. La pathologie de ces souris récapitule les principales caractéristiques de la maladie humaine. Ces résultats suggèrent l'implication de variants rares dans le diabète de type I humain.

Dans la sclérose en plaques, où 230 variants communs ont été associés, une étude portant sur 32 367 patients et autant de contrôles a été réalisée par le consortium international pour la génétique de la SEP grâce à une puce dite « exome-chip » interrogeant 121 000 variants codants dont 88% de faible fréquence (de 0.0001 à 0.05). Elle a permis d'identifier des variants rares dans des gènes déjà associés avec des variants communs. Ces variants rares rendaient compte de 5% de la composante génétique de la SEP, contre 20% pour les variants communs, laissant cependant 75% de l'hérédité inexpliquée [260].

Enfin dans le lupus, où 180 loci ont été détectés par les GWAS, de nombreuses études récentes ont rapporté le rôle de variants rares [261–277]. De plus, dans certaines d'entre elles, une analyse fonctionnelle approfondie permettait d'identifier la voie biologique perturbée par le variant et rendaient vraisemblable son rôle causal.

CONCLUSION

Notre travail ainsi que les résultats obtenus dans d'autres maladies inflammatoires chroniques ou auto-immunes montrent la difficulté d'identifier des variants rares expliquant l'hérédité manquante. Leurs effets sont en général faibles, le lupus érythémateux disséminé semblant être une exception à cet égard. La difficulté de leur détection pourrait cependant aussi être due à une hétérogénéité génétique importante.

BIBLIOGRAPHIE

1. Reveille JD, Weisman MH. The Epidemiology of Back Pain, Axial Spondyloarthritis and HLA-B27 in the United States. The American Journal of the Medical Sciences 2013;345:431–6.

2. Khan MA. HLA-B27 and its subtypes in world populations Editorial review: Current Opinion in Rheumatology 1995;7:263–9.

3. Stolwijk C, Boonen A, van Tubergen A, Reveille JD. Epidemiology of spondyloarthritis. Rheum Dis Clin North Am 2012;38:441–76.

4. Costantino F, Talpin A, Said-Nahal R, Goldberg M, Henny J, Chiocchia G, et al. Prevalence of spondyloarthritis in reference to HLA-B27 in the French population: results of the GAZEL cohort. Ann Rheum Dis 2015;74:689–93.

5. Stolwijk C, van Onna M, Boonen A, van Tubergen A. Global Prevalence of Spondyloarthritis: A Systematic Review and Meta-Regression Analysis: Analysis of SpA Prevalence. Arthritis Care & Research 2016;68:1320–31.

6. Feldtkeller E, Bruckel J, Khan MA. Scientific contributions of ankylosing spondylitis patient advocacy groups: Current Opinion in Rheumatology 2000;12:239–47.

7. Redeker I, Callhoff J, Hoffmann F, Haibel H, Sieper J, Zink A, et al. Determinants of diagnostic delay in axial spondyloarthritis: an analysis based on linked claims and patient-reported survey data. Rheumatology 2019;58:1634–8.

8. Sykes MP, Doll H, Sengupta R, Gaffney K. Delay to diagnosis in axial spondyloarthritis: are we improving in the UK?: Fig. 1. Rheumatology 2015;kev288.

9. Appel H, Sieper J. Spondyloarthritis at the crossroads of imaging, pathology, and structural damage in the era of biologics. Curr Rheumatol Rep 2008;10:356–63.

10. Rudwaleit M, Haibel H, Baraliakos X, Listing J, Märker-Hermann E, Zeidler H, et al. The early disease stage in axial spondylarthritis: Results from the german spondyloarthritis inception cohort. Arthritis Rheum 2009;60:717–27.

11. Dougados M, d'Agostino MA, Benessiano J, Berenbaum F, Breban M, Claudepierre P, et al. The DESIR cohort: A 10-year follow-up of early inflammatory back pain in France: Study design and baseline characteristics of the 708 recruited patients. Joint Bone Spine 2011;78:598–603.

12. Maksymowych WP, Mallon C, Morrow S, Shojania K, Olszynski WP, Wong RL, et al. Development and validation of the Spondyloarthritis Research Consortium of Canada (SPARCC) Enthesitis Index. Annals of the Rheumatic Diseases 2009;68:948–53.

13. Canouï-Poitrine F, Lekpa FK, Farrenq V, Boissinot V, Hacquard-Bouder C, Comet D, et al. Prevalence and factors associated with uveitis in spondylarthritis patients in France: Results from an observational survey. Arthritis Care Res 2012;64:919–24.

14. Zeboulon N, Dougados M, Gossec L. Prevalence and characteristics of uveitis in the spondyloarthropathies: a systematic literature review. Annals of the Rheumatic Diseases 2008;67:955–9.

15. Griffiths CE, Barker JN. Pathogenesis and clinical features of psoriasis. The Lancet 2007;370:263–71.

16. Gelfand JM, Weinstein R, Porter SB, Neimann AL, Berlin JA, Margolis DJ. Prevalence and Treatment of Psoriasis in the United Kingdom: A Population-Based Study. Arch Dermatol [Internet] 2005 [cited 2021 Apr 24];141. Available from: http://archderm.jamanetwork.com/article.aspx?doi=10.1001/archderm.141.12.1537

17. Gladman DD. PSORIATIC ARTHRITIS. Rheumatic Disease Clinics of North America 1998;24:829–44.

18. Mielants H, Veys EM, Cuvelier C, De Vos M, Goemaere S, De Clercq L, et al. The evolution of spondyloarthropathies in relation to gut histology. II. Histological aspects. J Rheumatol 1995;22:2273–8.

19. El Maghraoui A. Extra-articular manifestations of ankylosing spondylitis: Prevalence, characteristics and therapeutic implications. European Journal of Internal Medicine 2011;22:554–60.

20. Rudwaleit M. How to diagnose axial spondyloarthritis early. Annals of the Rheumatic Diseases 2004;63:535–43.

21. Gran JT, Husby G. HLA-B27 and spondyloarthropathy: value for early diagnosis? Journal of Medical Genetics 1995;32:497–501.

22. Dougados M, Gueguen A, Nakache JP, Velicitat P, Zeidler H, Veys E, et al. Clinical relevance of C-reactive protein in axial involvement of ankylosing spondylitis. J Rheumatol 1999;26:971–4.

23. Sieper J, Rudwaleit M, Baraliakos X, Brandt J, Braun J, Burgos-Vargas R, et al. The Assessment of SpondyloArthritis international Society (ASAS) handbook: a guide to assess spondyloarthritis. Annals of the Rheumatic Diseases 2009;68:ii1–44.

24. Rudwaleit M, Khan MA, Sieper J. The challenge of diagnosis and classification in early ankylosing spondylitis: Do we need new criteria? Arthritis Rheum 2005;52:1000–8.

25. Rudwaleit M, Jurik AG, Hermann KGA, Landewe R, van der Heijde D, Baraliakos X, et al. Defining active sacroiliitis on magnetic resonance imaging (MRI) for classification of axial spondyloarthritis: a consensual approach by the ASAS/OMERACT MRI group. Annals of the Rheumatic Diseases 2009;68:1520–7.

26. Hermann KGA, Baraliakos X, van der Heijde DM, Jurik AG, Landewé R, Marzo-Ortega H, et al. Descriptions of spinal MRI lesions and definition of a positive MRI of the spine in axial spondyloarthritis: a consensual approach by the ASAS/OMERACT MRI study group. Ann Rheum Dis 2012;71:1278–88.

27. Devauchelle-Pensec V, D'Agostino MA, Marion J, Lapierre M, Jousse-Joulin S, Colin D, et al. Computed tomography scanning facilitates the diagnosis of sacroiliitis in patients with suspected spondylarthritis: Results of a prospective multicenter French cohort study. Arthritis & Rheumatism 2012;64:1412–9.

28. D'Agostino MA, Said-Nahal R, Hacquard-Bouder C, Brasseur JL, Dougados M, Breban M. Assessment of peripheral enthesitis in the spondylarthropathies by ultrasonography combined with power Doppler: A cross-sectional study. Arthritis & Rheumatism 2003;48:523–33.

29. Lehtinen A, Taavitsainen M, Leirisalo-Repo M. Sonographic analysis of enthesopathy in the lower extremities of patients with spondylarthropathy. Clin Exp Rheumatol 1994;12:143–8.

30. Balint PV. Ultrasonography of entheseal insertions in the lower limb in spondyloarthropathy. Annals of the Rheumatic Diseases 2002;61:905–10.

31. Leirisalo-Repo M. Reactive arthritis. Scandinavian Journal of Rheumatology 2005;34:251–9.

32. Baeten D, Breban M, Lories R, Schett G, Sieper J. Are spondylarthritides related but distinct conditions or a single disease with a heterogeneous phenotype? Arthritis & Rheumatism 2013;65:12–20.

33. Wright V. Seronegative polyarthritis. a unified concept. Arthritis & Rheumatism 1978;21:619–33.

34. Hammer RE, M2aika SD, Richardson JA, Tang JP, Taurog JD. Spontaneous inflammatory disease in transgenic rats expressing HLA-B27 and human β2m: An animal model of HLA-B27-associated human disorders. Cell 1990;63:1099–112.

35. Rudwaleit M, Landewe R, van der Heijde D, Listing J, Brandt J, Braun J, et al. The development of Assessment of SpondyloArthritis international Society classification criteria for axial spondyloarthritis (part I): classification of paper patients by expert opinion including uncertainty appraisal. Annals of the Rheumatic Diseases 2009;68:770–6.

36. Rudwaleit M, van der Heijde D, Landewe R, Listing J, Akkoc N, Brandt J, et al. The development of Assessment of SpondyloArthritis international Society classification criteria for axial spondyloarthritis (part II): validation and final selection. Annals of the Rheumatic Diseases 2009;68:777–83.

37. Costantino F, Breban M. [Inflammatory spondylitis]. Rev Prat 2019;69:e3–10.

38. Amor B, Dougados M, Mijiyawa M. [Criteria of the classification of spondylarthropathies]. Rev Rhum Mal Osteoartic 1990;57:85–9.

39. Dougados M, van der Linden S, Juhlin R, Huitfeldt B, Amor B, Calin A, et al. The European Spondylarthropathy Study Group preliminary criteria for the classification of spondylarthropathy. Arthritis Rheum. 1991;34:1218–27.

40. Rudwaleit M, van der Heijde D, Landewe R, Akkoc N, Brandt J, Chou CT, et al. The Assessment of SpondyloArthritis international Society classification criteria for peripheral spondyloarthritis and for spondyloarthritis in general. Annals of the Rheumatic Diseases 2011;70:25–31.

41. Linden SVD, Valkenburg HA, Cats A. Evaluation of Diagnostic Criteria for Ankylosing Spondylitis. Arthritis & Rheu2matism 1984;27:361–8.

42. Nash P, Coates LC, Fleischmann R, Papp KA, Gomez-Reino JJ, Kanik KS, et al. Efficacy of Tofacitinib for the Treatment of Psoriatic Arthritis: Pooled Analysis of Two Phase 3 Studies. Rheumatol Ther 2018;5:567–82.

43. Keeling S, Maksymowych WP. JAK inhibitors, psoriatic arthritis, and axial spondyloarthritis: a critical review of clinical trials. Expert Review of Clinical Immunology 2021;1–15.

44. Merola JF, Papp KA, Nash P, Gratacós J, Boehncke WH, Thaçi D, et al. Tofacitinib in psoriatic arthritis patients: skin signs and symptoms and health-related quality of life from two randomized phase 3 studies. J Eur Acad Dermatol Venereol 2020;34:2809–20.

45. Elborn JS. Cystic fibrosis. The Lancet 2016;388:2519–31.

46. Spataro N, Rodríguez JA, Navarro A, Bosch E. Properties of human disease genes and the role of genes linked to Mendelian disorders in complex disease aetiology. Hum. Mol. Genet. 2017;ddw405.

47. Risch N. Linkage strategies for genetically complex traits. I. Multilocus models. Am J Hum Genet 1990;46:222–8.

48. Bijanzadeh M. The recurrence risk of genetic complex diseases. J Res Med Sci 2017;22:32.

49. Brown MA, Kennedy LG, MacGregor AJ, Darke C, Duncan E, Shatford JL, et al. Susceptibility to ankylosing spondylitis in twins: the role of genes, HLA, and the environment. Arthritis Rheum. 1997;40:1823–8.

50. Stein C, Elston R. Finding genes underlying human disease. Clinical Genetics 2009;75:101–6.

51. Sherry ST. dbSNP: the NCBI database of genetic variation. Nucleic Acids Research 2001;29:308–11.

52. Mullaney JM, Mills RE, Pittard WS, Devine SE. Small insertions and deletions (INDELs) in human genomes. Human2 Molecular Genetics 2010;19:R131–6.

53. Botstein D, White RL, Skolnick M, Davis RW. Construction of a genetic linkage map in man using restriction fragment length polymorphisms. Am J Hum Genet 1980;32:314–31.

54. Holland PM, Abramson RD, Watson R, Gelfand DH. Detection of specific polymerase chain reaction product by utilizing the 5'----3' exonuclease activity of Thermus aquaticus DNA polymerase. Proceedings of the National Academy of Sciences 1991;88:7276–80.

55. Montgomery J, Wittwer CT, Palais R, Zhou L. Simultaneous mutation scanning and genotyping by high-resolution DNA melting analysis. Nat Protoc 2007;2:59–66.

56. Simko I. High-Resolution DNA Melting Analysis in Plant Research. Trends in Plant Science 2016;21:528–37.

57. Sanger F, Nicklen S, Coulson AR. DNA sequencing with chain-terminating inhibitors. Proceedings of the National Academy of Sciences 1977;74:5463–7.

58. Maxam AM, Gilbert W. A new method for sequencing DNA. Proceedings of the National Academy of Sciences 1977;74:560–4.

59. França LTC, Carrilho E, Kist TBL. A review of DNA sequencing techniques. Quart. Rev. Biophys. 2002;35:169–200.

60. Sanger F, Nicklen S, Coulson AR. DNA sequencing with chain-terminating inhibitors. Proceedings of the National Academy of Sciences 1977;74:5463–7.

61. Metzker ML. Sequencing technologies — the next generation. Nat Rev Genet 2010;11:31–46.

62. Lim ET, Würtz P, Havulinna AS, Palta P, Tukiainen T, Rehnström K, et al. Distribution and medical impact of loss-of-f2unction variants in the Finnish founder population. PLoS Genet. 2014;10:e1004494.

63. Jain M, Koren S, Miga KH, Quick J, Rand AC, Sasani TA, et al. Nanopore sequencing and assembly of a human genome with ultra-long reads. Nat Biotechnol 2018;36:338–45.

64. Rhoads A, Au KF. PacBio Sequencing and Its Applications. Genomics, Proteomics & Bioinformatics 2015;13:278–89.

65. Branton D, Deamer DW, Marziali A, Bayley H, Benner SA, Butler T, et al. The potential and challenges of nanopore sequencing. Nat Biotechnol 2008;26:1146–53.

66. Morton NE. Sequential tests for the detection of linkage. Am J Hum Genet 1955;7:277–318.

67. Mathew C. Science medicine and the future: Postgenomic technologies: hunting the genes for common disorders. BMJ 2001;322:1031–4.

68. Elston RC, Song D, Iyengar SK. Mathematical Assumptions versus Biological Reality: Myths in Affected Sib Pair Linkage Analysis. The American Journal of Human Genetics 2005;76:152–6.

69. Risch NJ. Searching for genetic determinants in the new millennium. Nature 2000;405:847–56.

70. Cordell HJ, Clayton DG. Genetic association studies. The Lancet 2005;366:1121–31.

71. Risch N, Merikangas K. The Future of Genetic Studies of Complex Human Diseases. Science 1996;273:1516–7.

72. Lewontin RC. THE INTERACTION OF SELECTION AND LINKAGE. I. GENERAL CONSIDERATIONS; HETEROTIC MODELS. Genetics 1964;49:49–67.

73. 2 Hill WG, Robertson A. The effect of linkage on limits to artificial selection. Genet Res 1966;8:269–94.

74. Mayo N, Goldberg M. When is a case-control study not a case-control study? J Rehabil Med 2009;41:209–16.

75. Clarke GM, Anderson CA, Pettersson FH, Cardon LR, Morris AP, Zondervan KT. Basic statistical analysis in genetic case-control studies. Nat Protoc 2011;6:121–33.

76. Spielman RS, McGinnis RE, Ewens WJ. Transmission test for linkage disequilibrium: the insulin gene region and insulin-dependent diabetes mellitus (IDDM). Am J Hum Genet 1993;52:506–16.

77. Sham PC, Curtis D. An extended transmission/disequilibrium test (TDT) for multi-allele marker loci. Ann Human Genet 1995;59:323–36.

78. Wei Q, Chen Y, Zeng Z, Shu C, Long L, Lu J, et al. Transmission/Disequilibrium Tests Incorporating Unaffected Offspring. PLoS ONE 2014;9:e114892.

79. Yuan M, Tian X, Zheng G, Yang Y. Adaptive Transmission Disequilibrium Test for Family Trio Design. Statistical Applications in Genetics and Molecular Biology 2009;8:1–20.

80. Jorgensen TJ, Ruczinski I, Kessing B, Smith MW, Shugart YY, Alberg AJ. Hypothesis-Driven Candidate Gene Association Studies: Practical Design and Analytical Considerations. American Journal of Epidemiology 2009;170:986–93.

81. Nsengimana J, Bishop DT. Design Considerations for Genetic Linkage and Association Studies [Internet]. In: Elston RC, editor. Statistical Human Genetics. New York, NY: Springer New York; 2017 [cited 2022 Apr 2]. page 257–81.Available from: http://link.springer.com/10.1007/978-1-4939-7274-6_13

82. Visscher PM, Brown MA, McCarthy MI, Yang J. Five Years of GWAS Discovery. The American Journal of Human Genetics 2012;90:7–24.

83. Klein RJ, Xu X, Mukherjee S, Willis J, Hayes J. Successes of Genome-wide Association Studies. Cell 2010;142:350–1.

84. Goldstein DB. Common Genetic Variation and Human Traits. N Engl J Med 2009;360:1696–8.

85. Tam V, Patel N, Turcotte M, Bossé Y, Paré G, Meyre D. Benefits and limitations of genome-wide association studies. Nat Rev Genet 2019;20:467–84.

86. The 1000 Genomes Project Consortium. A map of human genome variation from population-scale sequencing. Nature 2010;467:1061–73.

87. Rosenberg NA, Huang L, Jewett EM, Szpiech ZA, Jankovic I, Boehnke M. Genome-wide association studies in diverse populations. Nat Rev Genet 2010;11:356–66.

88. Manolio TA, Collins FS, Cox NJ, Goldstein DB, Hindorff LA, Hunter DJ, et al. Finding the missing heritability of complex diseases. Nature 2009;461:747–53.

89. Chen GB, Lee SH, Brion MJA, Montgomery GW, Wray NR, Radford-Smith GL, et al. Estimation and partitioning of (co)heritability of inflammatory bowel disease from GWAS and immunochip data. Human Molecular Genetics 2014;23:4710–20.

90. de Lange KM, Moutsianas L, Lee JC, Lamb CA, Luo Y, Kennedy NA, et al. Genome-wide association study implicates immune activation of multiple integrin genes in inflammatory bowel disease. Nat Genet 2017;49:256–61.

91. Ramos GP, Papadakis KA. Mechanisms of Disease: Inflammatory Bowel Diseases. Mayo Clinic Proceedings 2019;94:155–65.

92. Altshuler D, Daly MJ, Lander ES. Genetic Mapping in Human Disease. Science 2008;322:881–8.

93. LaPierre N, Taraszka K, Huang H, He R, Hormozdiari F, Eskin E. Identifying causal variants by fine mapping across multiple studies. PLoS Genet 2021;17:e1009733.

94. Mayhew AJ, Meyre D. Assessing the Heritability of Complex Traits in Humans: Methodological Challenges and Opportunities. CG [Internet] 2017 [cited 2022 Jul 16];18. Available from: http://www.eurekaselect.com/150709/article

95. Lusis AJ, Pajukanta P. A treasure trove for lipoprotein biology. Nat Genet 2008;40:129–30.

96. Mahajan A, Wessel J, Willems SM, Zhao W, Robertson NR, Chu AY, et al. Refining the accuracy of validated target identification through coding variant fine-mapping in type 2 diabetes. Nat Genet 2018;50:559–71.

97. Reich DE, Lander ES. On the allelic spectrum of human disease. Trends in Genetics 2001;17:502–10.

98. Schork NJ, Murray SS, Frazer KA, Topol EJ. Common vs. rare allele hypotheses for complex diseases. Current Opinion in Genetics & Development 2009;19:212–9.

99. Gibson G. Rare and common variants: twenty arguments. Nat Rev Genet 2012;13:135–45.

100. Kryukov GV, Shpunt A, Stamatoyannopoulos JA, Sunyaev SR. Power of deep, all-exon resequencing for discovery of human trait genes. Proc. Natl. Acad. Sci. U.S.A. 2009;106:3871–6.

101. MacArthur DG, Balasubramanian S, Frankish A, Huang N, Morris J, Walter K, et al. A Systematic Survey of Loss-of-Function Variants in Human Protein-Coding Genes. Science 2012;335:823–8.

102. The 1000 Genomes Project Consortium. An integrated map of genetic variation from 1,092 human genomes. Nature 2012;491:56–65.

103. Goldstein JL, Brown MS. THE LDL RECEPTOR LOCUS AND THE GENETICS OF FAMILIAL HYPERCHOLESTEROLEMIA. Annu. Rev. Genet. 1979;13:259–89.

104. Weedon MN, Frayling TM. Insights on pathogenesis of type 2 diabetes from MODY genetics. Curr Diab Rep 2007;7:131–8.

105. Ogura Y, Bonen DK, Inohara N, Nicolae DL, Chen FF, Ramos R, et al. A frameshift mutation in NOD2 associated with susceptibility to Crohn's disease. Nature 2001;411:603–6.

106. Auer PL, Lettre G. Rare variant association studies: considerations, challenges and opportunities. Genome Med 2015;7:16.

107. Peltonen L, Palotie A, Lange K. Use of population isolates for mapping complex traits. Nat Rev Genet 2000;1:182–90.

108. Baier LJ, Hanson RL. Genetic Studies of the Etiology of Type 2 Diabetes in Pima Indians. Diabetes 2004;53:1181–6.

109. Dabelea D, Hanson RL, Bennett PH, Roumain J, Knowler WC, Pettitt DJ. Increasing prevalence of Type II diabetes in American Indian children. Diabetologia 1998;41:904–10.

110. Amanat S, Requena T, Lopez-Escamez JA. A Systematic Review of Extreme Phenotype Strategies to Search for Rare Variants in Genetic Studies of Complex Disorders. Genes 2020;11:987.

111. Hall JM, Lee MK, Newman B, Morrow JE, Anderson LA, Huey B, et al. Linkage of Early-Onset Familial Breast Cancer to Chromosome 17q21. Science 1990;250:1684–9.

112. Wijsman EM. Family-based approaches: design, imputation, analysis, and beyond. BMC Genet 2016;17:S9.

113. Cruchaga C, Karch CM, Jin SC, Benitez BA, Cai Y, Guerreiro R, et al. Rare coding variants in the phospholipase D3 gene confer risk for Alzheimer's disease. Nature 2014;505:550–4.

114. Wang K, Li M, Hakonarson H. ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. Nucleic Acids Research 2010;38:e164–e164.

115. Gorlov I, Gorlova O, Frazier M, Spitz M, Amos C. Evolutionary evidence of the effect of rare variants on disease etiology. Clinical Genetics 2011;79:199–206.

116. Karczewski KJ, Francioli LC, Tiao G, Cummings BB, Alföldi J, Wang Q, et al. The mutational constraint spectrum quantified from variation in 141,456 humans. Nature 2020;581:434–43.

117. Ng PC. SIFT: predicting amino acid changes that affect protein function. Nucleic Acids Research 2003;31:3812–4.

118. Choi Y, Chan AP. PROVEAN web server: a tool to predict the functional effect of amino acid substitutions and indels. Bioinformatics 2015;31:2745–7.

119. Ioannidis NM, Rothstein JH, Pejaver V, Middha S, McDonnell SK, Baheti S, et al. REVEL: An Ensemble Method for Predicting the Pathogenicity of Rare Missense Variants. The American Journal of Human Genetics 2016;99:877–85.

120. Adzhubei I, Jordan DM, Sunyaev SR. Predicting functional effect of human missense mutations using PolyPhen-2. Curr Protoc Hum Genet 2013;Chapter 7:Unit7.20.

121. Kircher M, Witten DM, Jain P, O'Roak BJ, Cooper GM, Shendure J. A general framework for estimating the relative pathogenicity of human genetic variants. Nat Genet 2014;46:310–5.

122. Lee S, Abecasis GR, Boehnke M, Lin X. Rare-Variant Association Analysis: Study Designs and Statistical Tests. The American Journal of Human Genetics 2014;95:5–23.

123. Lumley T, Brody J, Peloso G, Morrison A, Rice K. FastSKAT: Sequence kernel association tests for very large sets of markers. Genet. Epidemiol. 2018;42:516–27.

124. Ionita-Laza I, Lee S, Makarov V, Buxbaum JD, Lin X. Family-based association tests for sequence data, and comparisons with population-based association tests. Eur J Hum Genet 2013;21:1158–62.

125. Chen H, Meigs JB, Dupuis J. Sequence Kernel Association Test for Quantitative Traits in Family Samples. Genet. Epidemiol. 2013;37:196–204.

126. Jacobs JH, Rose FC. The Familial Occurrence of Ankylosing Spondylitis. BMJ 1954;2:1139–40.

127. Tegner W, Lloyd K. ANKYLOSING SPONDYLITIS IN ALL THREE CHILDREN OF ONE FAMILY. The Lancet 1949;254:196–7.

128. Dernis E, Said-Nahal R, D'Agostino MA, Aegerter P, Dougados M, Breban M. Recurrence of spondylarthropathy among first-degree relatives of patients: a systematic cross-sectional study. Ann. Rheum. Dis. 2009;68:502–7.

129. Brown MA. Recurrence risk modelling of the genetic susceptibility to ankylosing spondylitis. Annals of the Rheumatic Diseases 2000;59:883–6.

130. Thjodleifsson B, Geirsson árni J, Björnsson S, Bjarnason I. A common genetic background for inflammatory bowel disease and ankylosing spondylitis: A genealogic study in Iceland. Arthritis Rheum 2007;56:2633–9.

131. Geirsson AJ, Kristjansson K, Gudbjornsson B. A strong familiality of ankylosing spondylitis through several generations. Annals of the Rheumatic Diseases 2010;69:1346–8.

132. Sundquist K, Martineus JC, Li X, Hemminki K, Sundquist J. Concordant and discordant associations between rheumatoid arthritis, systemic lupus erythematosus and ankylosing spondylitis based on all hospitalizations in Sweden between 1973 and 2004. Rheumatology 2008;47:1199–202.

133. Morin M, Hellgren K, Frisell T. Familial aggregation and heritability of ankylosing spondylitis – a Swedish nested case–control study. Rheumatology 2020;59:1695–702.

134. van der Linden SM, Khan MA, Li Z, Baumberger H, Zandwijk H van, Khan MK, et al. Recurrence of axial spondyloarthritis among first-degree relatives in a prospective 35-year-follow-up family study. RMD Open 2022;8:e002208.

135. Pedersen OB, Svendsen AJ, Ejstrup L, Skytthe A, Harris JR, Junker P. Ankylosing spondylitis in Danish and Norwegian twins: occurrence and the relative importance of genetic vs. environmental effectors in disease causation. Scandinavian Journal of Rheumatology 2008;37:120–6.

136. Ellinghaus D, Jostins L, Spain SL, Cortes A, Bethune J, Han B, et al. Analysis of five chronic inflammatory diseases identifies 27 new associations and highlights disease-specific patterns at shared loci. Nat. Genet. 2016;48:510–8.

137. Fry A, Littlejohns TJ, Sudlow C, Doherty N, Adamska L, Sprosen T, et al. Comparison of Sociodemographic and Health-Related Characteristics of UK Biobank Participants With Those of the General Population. American Journal of Epidemiology 2017;186:1026–34.

138. Yang J, Benyamin B, McEvoy BP, Gordon S, Henders AK, Nyholt DR, et al. Common SNPs explain a large proportion of the heritability for human height. Nat Genet 2010;42:565–9.

139. Hwang MC, Ridley L, Reveille JD. Ankylosing spondylitis risk factors: a systematic literature review. Clin Rheumatol 2021;40:3079–93.

140. Aho K, Ahvonen P, Lassus A, Sievers K, Tiilikainen A. HL-A ANTIGEN 27 AND REACTIVE ARTHRITIS. The Lancet 1973;302:157.

141. Breban M, Hacquard-Bouder C, Falgarone G. Animal Models of HLA-B27-Associated Diseases. CMM 2004;4:31–40.

142. Taurog JD, Richardson JA, Croft JT, Simmons WA, Zhou M, Fernández-Sueiro JL, et al. The germfree state prevents development of gut and joint inflammatory disease in HLA-B27 transgenic rats. Journal of Experimental Medicine 1994;180:2359–64.

143. Rath HC, Herfarth HH, Ikeda JS, Grenther WB, Hamm TE, Balish E, et al. Normal luminal bacteria, especially Bacteroides species, mediate chronic colitis, gastritis, and arthritis in HLA-B27/human beta2 microglobulin transgenic rats. J. Clin. Invest. 1996;98:945–53.

144. Scher JU, Sczesnak A, Longman RS, Segata N, Ubeda C, Bielski C, et al. Expansion of intestinal Prevotella copri correlates with enhanced susceptibility to arthritis. eLife 2013;2:e01202.

145. Zhang X, Zhang D, Jia H, Feng Q, Wang D, Liang D, et al. The oral and gut microbiomes are perturbed in rheumatoid arthritis and partly normalized after treatment. Nat Med 2015;21:895–905.

146. Scher JU, Ubeda C, Artacho A, Attur M, Isaac S, Reddy SM, et al. Decreased Bacterial Diversity Characterizes the Altered Gut Microbiota in Patients With Psoriatic Arthritis, Resembling Dysbiosis in Inflammatory Bowel Disease: Gut Microbiota in PsA. Arthritis & Rheumatology 2015;67:128–39.

147. Stoll ML, Kumar R, Morrow CD, Lefkowitz EJ, Cui X, Genin A, et al. Altered microbiota associated with abnormal humoral immune responses to commensal organisms in enthesitis-related arthritis. Arthritis Res Ther 2014;16:486.

148. Breban M, Tap J, Leboime A, Said-Nahal R, Langella P, Chiocchia G, et al. Faecal microbiota study reveals specific dysbiosis in spondyloarthritis. Ann Rheum Dis 2017;76:1614–22.

149. Guggino G, Mauro D, Rizzo A, Alessandro R, Raimondo S, Bergot A, et al. Inflammasome Activation in Ankylosing Spondylitis Is Associated With Gut Dysbiosis. Arthritis Rheumatol 2021;73:1189–99.

150. Breban M, Beaufrère M, Glatigny S. Intestinal dysbiosis in spondyloarthritis – chicken or egg? Current Opinion in Rheumatology 2021;33:341–7.

151. Chung HY, Machado P, van der Heijde D, D'Agostino MA, Dougados M. Smokers in early axial spondyloarthritis have earlier disease onset, more disease activity, inflammation and damage, and poorer function and health-related quality of life: results from the DESIR cohort. Ann Rheum Dis 2012;71:809–16.

152. Poddubnyy D, Haibel H, Listing J, Märker-Hermann E, Zeidler H, Braun J, et al. Baseline radiographic damage, elevated acute-phase reactant levels, and cigarette smoking status predict spinal radiographic progression in early axial spondylarthritis. Arthritis & Rheumatism 2012;64:1388–98.

153. Zeboulon-Ktorza N, Boelle PY, Nahal RS, D'agostino MA, Vibert JF, Turbelin C, et al. Influence of Environmental Factors on Disease Activity in Spondyloarthritis: A Prospective Cohort Study. J Rheumatol 2013;40:469–75.

154. Ramiro S, Landewé R, van Tubergen A, Boonen A, Stolwijk C, Dougados M, et al. Lifestyle factors may modify the effect of disease activity on radiographic progression in patients with ankylosing spondylitis: a longitudinal analysis. RMD Open 2015;1:e000153.

155. Zhang H, Wan W, Liu J, Dai S, Zou Y, Qian Q, et al. Smoking quantity determines disease activity and function in Chinese patients with ankylosing spondylitis. Clin Rheumatol 2018;37:1605–16.

156. Fernando MMA, Stevens CR, Walsh EC, De Jager PL, Goyette P, Plenge RM, et al. Defining the Role of the MHC in Autoimmunity: A Review and Pooled Analysis. PLoS Genet 2008;4:e1000024.

157. Kulski, Shiina, Dijkstra. Genomic Diversity of the Major Histocompatibility Complex in Health and Disease. Cells 2019;8:1270.

158. Caffrey MFP, James DCO. Human Lymphocyte Antigen Association in Ankylosing Spondylitis. Nature 1973;242:121–121.

159. Brewerton DA, Hart FD, Nicholls A, Caffrey M, James DCO, Sturrock RD. ANKYLOSING SPONDYLITIS AND HL-A 27. The Lancet 1973;301:904–7.

160. Reveille JD. The genetic basis of spondyloarthritis. Annals of the Rheumatic Diseases 2011;70:i44–50.

161. López-Larrea C, Sujirachato K, Mehra NK, Chiewsilp P, Isarangkura D, Kanga U, et al. HLA-B27 subtypes in Asian patients with ankylosing spondylitis Evidence for new associations. Tissue Antigens 1995;45:169–76.

162. Paladini F, Taccari E, Fiorillo MT, Cauli A, Passiu G, Mathieu A, et al. Distribution of HLA-B27 subtypes in Sardinia and continental Italy and their association with spondylarthropathies. Arthritis Rheum 2005;52:3319–21.

163. Benjamin R, Parham P. Guilt by association: HLA-B27 and ankylosing spondylitis. Immunology Today 1990;11:137–42.

164. Colbert R. The Immunobiology of HLA-B27: Variations on a Theme. CMM 2004;4:21–30.

165. Chan AT, Kollnberger SD, Wedderburn LR, Bowness P. Expansion and enhanced survival of natural killer cells expressing the killer immunoglobulin-like receptor KIR3DL2 in spondylarthritis. Arthritis Rheum 2005;52:3586–95.

166. Bowness P, Ridley A, Shaw J, Chan AT, Wong-Baeza I, Fleming M, et al. Th17 Cells Expressing KIR3DL2 $^{+}$ and Responsive to HLA-B27 Homodimers Are Increased in Ankylosing Spondylitis. J.I. 2011;186:2672–80.

167. Boyle LH, Goodall JC, Opat SS, Gaston JSH. The Recognition of HLA-B27 by Human CD4 $^{+}$ T Lymphocytes. J Immunol 2001;167:2619–24.

168. Reveille JD, Zhou X, Lee M, Weisman MH, Yi L, Gensler LS, et al. HLA class I and II alleles in susceptibility to ankylosing spondylitis. Ann Rheum Dis 2019;78:66–73.

169. Cortes A, Pulit SL, Leo PJ, Pointon JJ, Robinson PC, Weisman MH, et al. Major histocompatibility complex associations of ankylosing spondylitis are complex and involve further epistasis with ERAP1. Nat Commun 2015;6:7146.

170. Londono J, Santos AM, Pena P, Calvo E, Espinosa LR, Reveille JD, et al. Analysis of HLA-B15 and HLA-B27 in spondyloarthritis with peripheral and axial clinical patterns. BMJ Open 2015;5:e009092–e009092.

171. Wang CM, Wang SH, Jan Wu YJ, Lin JC, Wu J, Chen JY. Human Leukocyte Antigen C*12:02:02 and Killer Immunoglobulin-Like Receptor 2DL5 are Distinctly Associated with Ankylosing Spondylitis in the Taiwanese. IJMS 2017;18:1775.

172. International Genetics of Ankylosing Spondylitis Consortium (IGAS). Identification of multiple risk variants for ankylosing spondylitis through high-density genotyping of immune-related loci. Nat Genet 2013;45:730–8.

173. Zhou X, Wang J, Zou H, Ward MM, Weisman MH, Espitia MG, et al. MICA, a gene contributing strong susceptibility to ankylosing spondylitis. Ann Rheum Dis 2014;73:1552–7.

174. Cortes A, Gladman D, Raychaudhuri S, Cui J, Wheeler L, Brown MA. Imputation-based analysis of *MICA* alleles in the susceptibility to ankylosing spondylitis. Ann Rheum Dis 2018;77:1691–2.

175. Wang CM, Tan KP, Jan Wu YJ, Lin JC, Zheng JW, Yu AL, et al. MICA*019 Allele and Soluble MICA as Biomarkers for Ankylosing Spondylitis in Taiwanese. JPM 2021;11:564.

176. Huang XF, Li Z, De Guzman E, Robinson P, Gensler L, Ward MM, et al. Genomewide Association Study of Acute Anterior Uveitis Identifies New Susceptibility Loci. Invest. Ophthalmol. Vis. Sci. 2020;61:3.

177. Díaz-Peña R, Castro-Santos P, Aransay AM, Brüges-Armas J, Pimentel-Santos FM, López-Larrea C. Genetic study confirms association of HLA-DPA1∗01:03 subtype with ankylosing spondylitis in HLA-B27-positive populations. Human Immunology 2013;74:764–7.

178. Robinson PC, Claushuis TAM, Cortes A, Martin TM, Evans DM, Leo P, et al. Genetic Dissection of Acute Anterior Uveitis Reveals Similarities and Differences in Associations Observed With Ankylosing Spondylitis: Genetics of Anterior Uveitis. Arthritis & Rheumatology 2015;67:140–51.

179. Reveille J. Major histocompatibility genes and ankylosing spondylitis. Best Practice & Research Clinical Rheumatology 2006;20:601–9.

180. Van Linden SMD, Valkenburg HA, Jongh BMD, Cats A. The risk of developing ankylosing spondylitis in HLA-B27 positive individuals. Arthritis & Rheumatism 1984;27:241–9.

181. Braun J, Bollow M, Remlinger G, Eggens U, Rudwaleit M, Distler A, et al. Prevalence of spondylarthropathies in HLA-B27 positive and negative blood donors. Arthritis & Rheumatism 1998;41:58–67.

182. JÄRvinen P. Occurrence of ankylosing spondylitis in a nationwide series of twins. Arthritis & Rheumatism 1995;38:381–3.

183. Laval SH, Timms A, Edwards S, Bradbury L, Brophy S, Milicic A, et al. Whole-Genome Screening in Ankylosing Spondylitis: Evidence of Non-MHC Genetic-Susceptibility Loci. The American Journal of Human Genetics 2001;68:918–26.

184. Zhang G, Luo J, Bruckel J, Weisman MA, Schumacher HR, Khan MA, et al. Genetic studies in familial ankylosing spondylitis susceptibility. Arthritis & Rheumatism 2004;50:2246–54.

185. Miceli-Richard C. Significant linkage to spondyloarthropathy on 9q31-34. Human Molecular Genetics 2004;13:1641–8.

186. Costantino F, Chaplais E, Leturcq T, Said-Nahal R, Leboime A, Zinovieva E, et al. Whole-genome single nucleotide polymorphism-based linkage analysis in spondyloarthritis multiplex families reveals a new susceptibility locus in 13q13. Ann. Rheum. Dis. 2016;75:1380–5.

187. Lee YH, Rho YH, Choi SJ, Ji JD, Song GG. Ankylosing spondylitis susceptibility loci defined by genome-search meta-analysis. J Hum Genet 2005;50:453–9.

188. Carter KW, Pluzhnikov A, Timms AE, Miceli-Richard C, Bourgain C, Wordsworth BP, et al. Combined analysis of three whole genome linkage scans for Ankylosing Spondylitis. Rheumatology 2007;46:763–71.

189. Pointon JJ, Harvey D, Karaderi T, Appleton LH, Farrar C, Stone MA, et al. The chromosome 16q region associated with ankylosing spondylitis includes the candidate gene tumour necrosis factor receptor type 1-associated death domain (*TRADD*). Ann Rheum Dis 2010;69:1243–6.

190. Zinovieva E, Lebrun N, Letourneur F, Laurent FX, Said-Nahal R, Chiocchia G, et al. Lack of association between Tenascin-C gene and spondyloarthritis. Rheumatology 2008;47:1655–8.

191. Zinovieva E, Kadi A, Letourneur F, Cagnard N, Izac B, Vigier A, et al. Systematic candidate gene investigations in the SPA2 locus (9q32) show an association between TNFSF8 and susceptibility to spondylarthritis. Arthritis & Rheumatism 2011;63:1853–9.

192. Mambueni HM, Hue C, Jobart-Malfait A, Said-Nahal R, Hafci H el, Petite H, et al. Targeted resequencing of the 13q13 spondyloarthritis-linked locus identifies a rare variant in FREM2 possibly associated with familial spondyloarthritis. Joint Bone Spine 2022;105419.

193. Wellcome Trust Case Control Consortium, Australo-Anglo-American Spondylitis Consortium (TASC), Burton PR, Clayton DG, Cardon LR, Craddock N, et al. Association scan of 14,500 nonsynonymous SNPs in four diseases identifies autoimmunity variants. Nat. Genet. 2007;39:1329–37.

194. The Australo-Anglo-American Spondyloarthritis Consortium (TASC). Genome-wide association study of ankylosing spondylitis identifies non-MHC susceptibility loci. Nat Genet 2010;42:123–7.

195. Evans DM, Spencer CCA, Pointon JJ, Su Z, Harvey D, Kochan G, et al. Interaction between ERAP1 and HLA-B27 in ankylosing spondylitis implicates peptide handling in the mechanism for HLA-B27 in disease susceptibility. Nat. Genet. 2011;43:761–7.

196. Lin Z, Bei JX, Shen M, Li Q, Liao Z, Zhang Y, et al. A genome-wide association study in Han Chinese identifies new susceptibility loci for ankylosing spondylitis. Nat. Genet. 2011;44:73–7.

197. Lin Z, Bei JX, Shen M, Li Q, Liao Z, Zhang Y, et al. A genome-wide association study in Han Chinese identifies new susceptibility loci for ankylosing spondylitis. Nat Genet 2012;44:73–7.

198. Fernández-Torres J, Zamudio-Cuevas Y, Montaño-Armendariz N, Luján-Juárez IA, Sánchez-Sánchez R, Martínez-Flores K. HLA-B27 may modulate the interaction between ERAP1 polymorphisms and smoking in ankylosing spondylitis patients. Mol Biol Rep 2022;49:6423–31.

199. Jung SH, Yim SH, Hu HJ, Lee KH, Lee JH, Sheen DH, et al. Genome-Wide Copy Number Variation Analysis Identifies Deletion Variants Associated With Ankylosing Spondylitis: Deletion Variants Associated With AS. Arthritis & Rheumatology 2014;66:2103–12.

200. Australian Osteoporosis Genetics Consortium, Wellcome Trust Case Control Consortium, Management Committee, Data and Analysis Group, DNA, Genotyping, Data QC and Informatics Group, Publications Committee, et al. Exome-wide study of ankylosing spondylitis

demonstrates additional shared genetic background with inflammatory bowel disease. npj Genomic Med 2016;1:16008.

201. Li Z, Akar S, Yarkan H, Lee SK, Çetin P, Can G, et al. Genome-wide association study in Turkish and Iranian populations identify rare familial Mediterranean fever gene (MEFV) polymorphisms associated with ankylosing spondylitis. PLoS Genet 2019;15:e1008038.

202. Uddin M, Maksymowych WP, Inman R, Gladman D, Munn A, Yazdani R, et al. UGT2B17 copy number gain in a large ankylosing spondylitis multiplex family. BMC Genet. 2013;14:67.

203. Rong J, Li Q, Zhang P, Wu X, Huang J, Li C, et al. A rare co-segregation-mutation in the insulin receptor substrate 1 gene in one Chinese family with ankylosing spondylitis. PLoS ONE 2015;10:e0126348.

204. O'Rielly DD, Uddin M, Codner D, Hayley M, Zhou J, Pena-Castillo L, et al. Private rare deletions in *SEC16A* and *MAMDC4* may represent novel pathogenic variants in familial axial spondyloarthritis. Ann Rheum Dis 2016;75:772–9.

205. Tan Z, Zeng H, Xu Z, Tian Q, Gao X, Zhou C, et al. Identification of ANKDD1B variants in an ankylosing spondylitis pedigree and a sporadic patient. BMC Med. Genet. 2018;19:111.

206. Feng Y, Hong Y, Zhang X, Cao C, Yang X, Lai S, et al. Genetic variants of TREML2 are associated with HLA-B27-positive ankylosing spondylitis. Gene 2018;668:121–8.

207. Garshasbi M, Mahmoudi M, Razmara E, Vojdanian M, Aslani S, Farhadi E, et al. Identification of RELN variant p.(Ser2486Gly) in an Iranian family with ankylosing spondylitis; the first association of RELN and AS. Eur J Hum Genet 2020;28:754–62.

208. Liu J, Wan W, Qiao Y, Li Y, Liu Q, Zhu Q, et al. A rare variant of RNF123 in familial ankylosing spondylitis causes dysfunction of osteoclast formation. Rheumatology (Oxford) 2020;59:1174–7.

209. Said-Nahal R, Miceli-Richard C, Berthelot JM, Duché A, Dernis-Labous E, Le Blévec G, et al. The familial form of spondylarthropathy: a clinical study of 115 multiplex families. Groupe Français d'Etude Génétique des Spondylarthropathies. Arthritis Rheum. 2000;43:1356–65.

210. Amor B, Dougados M, Mijiyawa M. [Criteria of the classification of spondylarthropathies]. Rev Rhum Mal Osteoartic 1990;57:85–9.

211. Dougados M, Linden SVD, Juhlin R, Huitfeldt B, Amor B, Calin A, et al. The European Spondylarthropathy Study Group Preliminary Criteria for the Classification of Spondylarthropathy. Arthritis & Rheumatism 1991;34:1218–27.

212. Rudwaleit M, van der Heijde D, Landewe R, Listing J, Akkoc N, Brandt J, et al. The development of Assessment of SpondyloArthritis international Society classification criteria for axial spondyloarthritis (part II): validation and final selection. Annals of the Rheumatic Diseases 2009;68:777–83.

213. Rudwaleit M, van der Heijde D, Landewe R, Akkoc N, Brandt J, Chou CT, et al. The Assessment of SpondyloArthritis international Society classification criteria for peripheral spondyloarthritis and for spondyloarthritis in general. Annals of the Rheumatic Diseases 2011;70:25–31.

214. Chen S, Zhou Y, Chen Y, Gu J. fastp: an ultra-fast all-in-one FASTQ preprocessor. Bioinformatics 2018;34:i884–90.

215. Quinlan AR, Hall IM. BEDTools: a flexible suite of utilities for comparing genomic features. Bioinformatics 2010;26:841–2.

216. Okonechnikov K, Conesa A, García-Alcalde F. Qualimap 2: advanced multi-sample quality control for high-throughput sequencing data. Bioinformatics 2015;btv566.

217. Auwera GA, Carneiro MO, Hartl C, Poplin R, del Angel G, Levy-Moonshine A, et al. From FastQ Data to High-Confidence Variant Calls: The Genome Analysis Toolkit Best Practices Pipeline. Current Protocols in Bioinformatics [Internet] 2013 [cited 2022 Jul 23];43. Available from: https://onlinelibrary.wiley.com/doi/10.1002/0471250953.bi1110s43

218. Boyle AP, Hong EL, Hariharan M, Cheng Y, Schaub MA, Kasowski M, et al. Annotation of functional variation in personal genomes using RegulomeDB. Genome Res. 2012;22:1790–7.

219. Manichaikul A, Mychaleckyj JC, Rich SS, Daly K, Sale M, Chen WM. Robust relationship inference in genome-wide association studies. Bioinformatics 2010;26:2867–73.

220. Thiele H, Nürnberg P. HaploPainter: a tool for drawing pedigrees with complex haplotypes. Bioinformatics 2005;21:1730–2.

221. Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MAR, Bender D, et al. PLINK: A Tool Set for Whole-Genome Association and Population-Based Linkage Analyses. The American Journal of Human Genetics 2007;81:559–75.

222. Costantino F, Chaplais E, Leturcq T, Said-Nahal R, Leboime A, Zinovieva E, et al. Whole-genome single nucleotide polymorphism-based linkage analysis in spondyloarthritis multiplex families reveals a new susceptibility locus in 13q13. Ann. Rheum. Dis. 2016;75:1380–5.

223. Explore FrEx Database [Internet]. [cited 2020 Sep 1];Available from: http://lysine.univ-brest.fr/FrExAC/

224. Pavlakis E, Chiotaki R, Chalepakis G. The role of Fras1/Frem proteins in the structure and function of basement membrane. Int. J. Biochem. Cell Biol. 2011;43:487–95.

225. Bouaoud J, Olivetto M, Testelin S, Dakpe S, Bettoni J, Devauchelle B. Fraser syndrome: review of the literature illustrated by a historical adult case. International Journal of Oral and Maxillofacial Surgery 2020;49:1245–53.

226. Melé M, Ferreira PG, Reverter F, DeLuca DS, Monlong J, Sammeth M, et al. The human transcriptome across tissues and individuals. Science 2015;348:660–5.

227. Schmiedel BJ, Singh D, Madrigal A, Valdovino-Gonzalez AG, White BM, Zapardiel-Gonzalo J, et al. Impact of Genetic Polymorphisms on Human Immune Cell Gene Expression. Cell 2018;175:1701-1715.e16.

228. Poole C. CONFOCAL ANALYSIS OF PRIMARY CILIA STRUCTURE AND COLOCALIZATION WITH THE GOLGI APPARATUS IN CHONDROCYTES AND AORTIC SMOOTH MUSCLE CELLS. Cell Biology International 1997;21:483–94.

229. Jensen CG, Poole CA, McGlashan SR, Marko M, Issa ZI, Vujcich KV, et al. Ultrastructural, tomographic and confocal imaging of the chondrocyte primary cilium in situ. Cell Biology International 2004;28:101–10.

230. McGlashan SR, Cluett EC, Jensen CG, Poole CA. Primary cilia in osteoarthritic chondrocytes: From chondrons to clusters. Dev. Dyn. 2008;237:2013–20.

231. Jovčevska I, Zottel A, Šamec N, Mlakar J, Sorokin M, Nikitin D, et al. High FREM2 Gene and Protein Expression Are Associated with Favorable Prognosis of IDH-WT Glioblastomas. Cancers (Basel) 2019;11.

232. Francos-Quijorna I, Sánchez-Petidier M, Burnside ER, Badea SR, Torres-Espin A, Marshall L, et al. Chondroitin sulfate proteoglycans prevent immune cell phenotypic conversion and inflammation resolution via TLR4 in rodent models of spinal cord injury. Nat Commun [Internet] 2022 [cited 2022 Oct 26];13:2933. Available from: https://www.nature.com/articles/s41467-022-30467-5

233. Braun T, Voland P, Kunz L, Prinz C, Gratzl M. Enterochromaffin cells of the human gut: sensors for spices and odorants. Gastroenterology 2007;132:1890–901.

234. Yadav VK, Ryu JH, Suda N, Tanaka KF, Gingrich JA, Schütz G, et al. Lrp5 Controls Bone Formation by Inhibiting Serotonin Synthesis in the Duodenum. Cell 2008;135:825–37.

235. O'Mahony SM, Clarke G, Borre YE, Dinan TG, Cryan JF. Serotonin, tryptophan metabolism and the brain-gut-microbiome axis. Behavioural Brain Research 2015;277:32–48.

236. Klavdianou K, Liossis SN, Papachristou DJ, Theocharis G, Sirinian C, Kottorou A, et al. Decreased Serotonin Levels and Serotonin-Mediated Osteoblastic Inhibitory Signaling in Patients With Ankylosing Spondylitis: SEROTONIN SIGNALING IN AS. J Bone Miner Res 2016;31:630–9.

237. Friedrich M, Diegelmann J, Schauber J, Auernhammer CJ, Brand S. Intestinal neuroendocrine cells and goblet cells are mediators of IL-17A-amplified epithelial IL-17C production in human inflammatory bowel disease. Mucosal Immunol 2015;8:943–58.

238. Costantino F, Mambu Mambueni H, Said-Nahal R, Garchon HJ, Breban M. What Have We Learned From Family-Based Studies About Spondyloarthritis? Front. Genet. 2021;12:671306.

239. Schuhmann MK, Stegner D, Berna-Erro A, Bittner S, Braun A, Kleinschnitz C, et al. Stromal Interaction Molecules 1 and 2 Are Key Regulators of Autoreactive T Cell Activation in Murine Autoimmune Central Nervous System Inflammation. J.I. 2010;184:1536–42.

240. Berna-Erro A, Jardin I, Salido GM, Rosado JA. Role of STIM2 in cell function and physiopathology: STIM2 in cell function and physiopathology. J Physiol 2017;595:3111–28.

241. Büttner S, Habernig L, Broeskamp F, Ruli D, Vögtle FN, Vlachos M, et al. Endonuclease G mediates α-synuclein cytotoxicity during Parkinson's disease: Endonuclease G executes α-synuclein cytotoxicity. The EMBO Journal 2013;32:3041–54.

242. Humphray SJ, Oliver K, Hunt AR, Plumb RW, Loveland JE, Howe KL, et al. DNA sequence and analysis of human chromosome 9. Nature [Internet] 2004 [cited 2022 Oct 26];429:369–74. Available from: http://www.nature.com/articles/nature02465

243. Ehnert S, Sreekumar V, Aspera-Werz RH, Sajadian SO, Wintermeyer E, Sandmann GH, et al. TGF-β1 impairs mechanosensation of human osteoblasts via HDAC6-mediated shortening and distortion of primary cilia. J Mol Med 2017;95:653–63.

244. Perricone C, Ceccarelli F, Valesini G. An overview on the genetic of rheumatoid arthritis: A never-ending story. Autoimmunity Reviews [Internet] 2011 [cited 2022 Oct 26];10:599–608. Available from: https://linkinghub.elsevier.com/retrieve/pii/S1568997211000899

245. Kurkó J, Besenyei T, Laki J, Glant TT, Mikecz K, Szekanecz Z. Genetics of Rheumatoid Arthritis — A Comprehensive Review. Clinic Rev Allerg Immunol [Internet] 2013 [cited 2022 Oct 26];45:170–9. Available from: http://link.springer.com/10.1007/s12016-012-8346-7

246. van der Woude D, Houwing-Duistermaat JJ, Toes REM, Huizinga TWJ, Thomson W, Worthington J, et al. Quantitative heritability of anti-citrullinated protein antibody-positive and anti-citrullinated protein antibody-negative rheumatoid arthritis. Arthritis Rheum [Internet] 2009 [cited 2022 Oct 26];60:916–23. Available from: https://onlinelibrary.wiley.com/doi/10.1002/art.24385

247. Deighton CM, Walker DJ, Griffiths ID, Roberts DF. The contribution of H LA to rheumatoid arthritis. Clinical Genetics [Internet] 2008 [cited 2022 Oct 26];36:178–82. Available from: https://onlinelibrary.wiley.com/doi/10.1111/j.1399-0004.1989.tb03185.x

248. de Vries RRP, van der Woude D, Houwing JJ, Toes REM. Genetics of ACPA-positive rheumatoid arthritis: the beginning of the end? Annals of the Rheumatic Diseases [Internet] 2011 [cited 2022 Oct 26];70:i51–4. Available from: https://ard.bmj.com/lookup/doi/10.1136/ard.2010.138040

249. Diogo D, Kurreeman F, Stahl EA, Liao KP, Gupta N, Greenberg JD, et al. Rare, Low-Frequency, and Common Variants in the Protein-Coding Sequence of Biological Candidate Genes from GWASs Contribute to Risk of Rheumatoid Arthritis. The American Journal of Human Genetics [Internet] 2013 [cited 2022 Oct 26];92:15–27. Available from: https://linkinghub.elsevier.com/retrieve/pii/S0002929712006258

250. Bang SY, Na YJ, Kim K, Joo YB, Park Y, Lee J, et al. Targeted exon sequencing fails to identify rare coding variants with large effect in rheumatoid arthritis. Arthritis Res Ther [Internet] 2014 [cited 2022 Oct 26];16:447. Available from: http://arthritis-research.biomedcentral.com/articles/10.1186/s13075-014-0447-7

251. Mitsunaga S, Hosomichi K, Okudaira Y, Nakaoka H, Suzuki Y, Kuwana M, et al. Aggregation of rare/low-frequency variants of the mitochondria respiratory chain-related proteins in rheumatoid arthritis patients. J Hum Genet [Internet] 2015 [cited 2022 Oct 26];60:449–54. Available from: http://www.nature.com/articles/jhg201550

252. Okada Y, Diogo D, Greenberg JD, Mouassess F, Achkar WAL, Fulton RS, et al. Integration of Sequence Data from a Consanguineous Family with Genetic Data from an Outbred Population Identifies PLB1 as a Candidate Rheumatoid Arthritis Risk Gene. PLoS ONE [Internet] 2014 [cited 2022 Oct 26];9:e87645. Available from: https://dx.plos.org/10.1371/journal.pone.0087645

253. Veyssiere M, Perea J, Michou L, Boland A, Caloustian C, Olaso R, et al. A novel nonsense variant in SUPT20H gene associated with Rheumatoid Arthritis identified by Whole Exome Sequencing of multiplex families. PLoS ONE [Internet] 2019 [cited 2022 Oct 26];14:e0213387. Available from: https://dx.plos.org/10.1371/journal.pone.0213387

254. Bradfield JP, Qu HQ, Wang K, Zhang H, Sleiman PM, Kim CE, et al. A Genome-Wide Meta-Analysis of Six Type 1 Diabetes Cohorts Identifies Multiple Associated Loci. PLoS Genet [Internet] 2011 [cited 2022 Oct 26];7:e1002293. Available from: https://dx.plos.org/10.1371/journal.pgen.1002293

255. The Type 1 Diabetes Genetics Consortium, Barrett JC, Clayton DG, Concannon P, Akolkar B, Cooper JD, et al. Genome-wide association study and meta-analysis find that over 40 loci affect risk of type 1 diabetes. Nat Genet [Internet] 2009 [cited 2022 Oct 26];41:703–7. Available from: http://www.nature.com/articles/ng.381

256. Pociot F. Type 1 diabetes genome-wide association studies: not to be lost in translation. Clin Trans Immunol [Internet] 2017 [cited 2022 Oct 26];6:e162. Available from: http://doi.wiley.com/10.1038/cti.2017.51

257. Wray NR, Yang J, Goddard ME, Visscher PM. The Genetic Interpretation of Area under the ROC Curve in Genomic Profiling. PLoS Genet [Internet] 2010 [cited 2022 Oct 26];6:e1000864. Available from: https://dx.plos.org/10.1371/journal.pgen.1000864

258. Genetics of Type 1 Diabetes in Finland, the Wellcome Trust Case Control Consortium, Todd JA, Walker NM, Cooper JD, Smyth DJ, et al. Robust associations of four new chromosome regions from genome-wide analyses of type 1 diabetes. Nat Genet [Internet] 2007 [cited 2022 Oct 26];39:857–64. Available from: http://www.nature.com/articles/ng2068

259. Foray AP, Candon S, Hildebrand S, Marquet C, Valette F, Pecquet C, et al. De novo germline mutation in the dual specificity phosphatase 10 gene accelerates autoimmune diabetes. Proc. Natl. Acad. Sci. U.S.A. [Internet] 2021 [cited 2022 Oct 26];118:e2112032118. Available from: https://pnas.org/doi/full/10.1073/pnas.2112032118

260. Mitrovič M, Patsopoulos NA, Beecham AH, Dankowski T, Goris A, Dubois B, et al. Low-Frequency and Rare-Coding Variation Contributes to Multiple Sclerosis Risk. Cell [Internet] 2018 [cited 2022 Oct 26];175:1679-1687.e7. Available from: https://linkinghub.elsevier.com/retrieve/pii/S0092867418312613

261. Duan R, Liu Q, Li J, Bian X, Yuan Q, Li Y, et al. A De Novo Frameshift Mutation in TNFAIP3 Impairs A20 Deubiquitination Function to Cause Neuropsychiatric Systemic Lupus Erythematosus. J Clin Immunol [Internet] 2019 [cited 2022 Oct 28];39:795–804. Available from: http://link.springer.com/10.1007/s10875-019-00695-4

262. Farias FHG, Dahlqvist J, Kozyrev SV, Leonard D, Wilbe M, Abramov SN, et al. A rare regulatory variant in the MEF2D gene affects gene regulation and splicing and is associated with a SLE sub-phenotype in Swedish cohorts. Eur J Hum Genet [Internet] 2019 [cited 2022 Oct 28];27:432–41. Available from: http://www.nature.com/articles/s41431-018-0297-x

263. Qi Y, Zhou X, Nath SK, Sun C, Wang Y, Hou P, et al. A Rare Variant (rs933717) at *FBXO 31- MAP 1 LC 3B* in Chinese Is Associated With Systemic Lupus Erythematosus. Arthritis Rheumatol [Internet] 2018 [cited 2022 Oct 28];70:287–97. Available from: https://onlinelibrary.wiley.com/doi/10.1002/art.40353

264. Ellyard JI, Jerjen R, Martin JL, Lee AYS, Field MA, Jiang SH, et al. Brief Report: Identification of a Pathogenic Variant in TREX1 in Early-Onset Cerebral Systemic Lupus Erythematosus by Whole-Exome Sequencing: Pathogenic TREX1 Deficiency in Early-Onset Cerebral SLE. Arthritis & Rheumatology [Internet] 2014 [cited 2022 Oct 28];66:3382–6. Available from: https://onlinelibrary.wiley.com/doi/10.1002/art.38824

265. Pullabhatla V, Roberts AL, Lewis MJ, Mauro D, Morris DL, Odhams CA, et al. De novo mutations implicate novel genes in systemic lupus erythematosus. Human Molecular Genetics [Internet] 2018 [cited 2022 Oct 28];27:421–9. Available from: https://academic.oup.com/hmg/article/27/3/421/4644524

266. Zoghi S, Ziaee V, Hirschmugl T, Jimenez-Heredia R, Krolo A, Boztug K, et al. Exome sequencing revealed C1Q homozygous mutation in Pediatric Systemic Lupus Erythematosus. Allergologia et Immunopathologia [Internet] 2018 [cited 2022 Oct 28];46:594–8. Available from: https://linkinghub.elsevier.com/retrieve/pii/S0301054618300478

267. Jiang SH, Athanasopoulos V, Ellyard JI, Chuah A, Cappello J, Cook A, et al. Functional rare and low frequency variants in BLK and BANK1 contribute to human lupus. Nat Commun [Internet] 2019 [cited 2022 Oct 28];10:2201. Available from: http://www.nature.com/articles/s41467-019-10242-9

268. Shahin T, Mayr D, Shoeb MR, Kuehn HS, Hoeger B, Giuliani S, et al. Identification of germline monoallelic mutations in *IKZF2* in patients with immune dysregulation. Blood Advances [Internet] 20222 [cited 2022 Oct 28];6:2444–51. Available from: https://ashpublications.org/bloodadvances/article/6/7/2444/483186/Identification-of-germline-monoallelic-mutations

269. Lea-Henry TN, Chuah A, Stanley M, Athanasopoulos V, Starkey MR, Christiadi D, et al. Increased burden of rare variants in genes of the endosomal Toll-like receptor pathway in patients with systemic lupus erythematosus. Lupus [Internet] 2021 [cited 2022 Oct 28];30:1756–63. Available from: http://journals.sagepub.com/doi/10.1177/09612033211033979

270. Al-Mayouf SM, Sunker A, Abdwani R, Abrawi SA, Almurshedi F, Alhashmi N, et al. Loss-of-function variant in DNASE1L3 causes a familial form of systemic lupus erythematosus. Nat Genet [Internet] 2011 [cited 2022 Oct 28];43:1186–8. Available from: http://www.nature.com/articles/ng.975

271. Xu L, Zhao J, Sun Q, Xu X, Wang L, Liu T, et al. Loss-of-function variants in *SAT1* cause X-linked childhood-onset systemic lupus erythematosus. Ann Rheum Dis [Internet] 2022 [cited 2022 Oct 28];annrheumdis-2022-222795. Available from: https://ard.bmj.com/lookup/doi/10.1136/ard-2022-222795

272. He Y, Gallman AE, Xie C, Shen Q, Ma J, Wolfreys FD, et al. P2RY8 variants in lupus patients uncover a role for the receptor in immunological tolerance. Journal of Experimental Medicine [Internet] 2022 [cited 2022 Oct 28];219:e20211004. Available from: https://rupress.org/jem/article/219/1/e20211004/212900/P2RY8-variants-in-lupus-patients-uncover-a-role

273. Roberts AL, Thomas ER, Bhosle S, Game L, Obraztsova O, Aitman TJ, et al. Resequencing the susceptibility gene, ITGAM, identifies two functionally deleterious rare variants in systemic lupus erythematosus cases. Arthritis Res Ther [Internet] 2014 [cited 2022 Oct 28];16:R114. Available from: http://arthritis-research.biomedcentral.com/articles/10.1186/ar4566

274. Postal M, Vivaldo JF, Fernandez-Ruiz R, Paredes JL, Appenzeller S, Niewold TB. Type I interferon in the pathogenesis of systemic lupus erythematosus. Current Opinion in Immunology [Internet] 2020 [cited 2022 Oct 28];67:87–94. Available from: https://linkinghub.elsevier.com/retrieve/pii/S095279152030114X

275. Delgado-Vega AM, Martínez-Bueno M, Oparina NY, López Herráez D, Kristjansdottir H, Steinsson K, et al. Whole Exome Sequencing of Patients from Multicase Families with Systemic Lupus Erythematosus Identifies Multiple Rare Variants. Sci Rep [Internet] 2018 [cited 2022 Oct 28];8:8775. Available from: http://www.nature.com/articles/s41598-018-26274-y

276. Almlöf JC, Nystedt S, Leonard D, Eloranta ML, Grosso G, Sjöwall C, et al. Whole-genome sequencing identifies complex contributions to genetic risk by variants in genes causing monogenic systemic lupus erythematosus. Hum Genet [Internet] 2019 [cited 2022 Oct 28];138:141–50. Available from: http://link.springer.com/10.1007/s00439-018-01966-7

277. Chen J, Zhang P, Chen H, Wang X, He X, Zhong J, et al. Whole-genome sequencing identifies rare missense variants of WNT16 and ERVW-1 causing the systemic lupus erythematosus. Genomics [Internet] 2022 [cited 2022 Oct 28];114:110332. Available from: https://linkinghub.elsevier.com/retrieve/pii/S0888754322000775