



HAL
open science

Génétique humaine de la lèpre au Vietnam : une histoire de familles

Chaïma Gzara

► **To cite this version:**

Chaïma Gzara. Génétique humaine de la lèpre au Vietnam : une histoire de familles. Santé publique et épidémiologie. Université Paris Cité, 2021. Français. NNT : 2021UNIP5234 . tel-04186899

HAL Id: tel-04186899

<https://theses.hal.science/tel-04186899v1>

Submitted on 24 Aug 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.

UNIVERSITÉ DE PARIS

*Thèse De Doctorat
d'Épidémiologie Génétique
Grade De Docteur*

École Doctorale 393 : Pierre Louis de santé publique à Paris
épidémiologie et sciences de l'information biomédicale

Laboratoire de Génétique Humaine des Maladies Infectieuses

*Génétique humaine de la lèpre au Vietnam :
une histoire de familles*

Présentée et soutenue publiquement le 31 Mars 2021

Par Chaïma Gzara

Et dirigée par Dr Alexandre Alcaïs

Devant un jury composé de :

- Pr Antoine Mahé, prof conventionné & HDR, Université de Strasbourg (rapporteur)
- Pr Bertram Müller-Myhsok, Max Planck Institute of Psychiatry, Munich (rapporteur)
- Dr Laure Gineau, CR, Université de Paris (Examinatrice)
- Dr Nabila Bouatia-Naji, DR-HDR, Université de Paris (Examinatrice)
- Pr Roch-Christian Johnson, MCU-HDR, Université d'Abomey-Calavi, Bénin (Examineur)
- Dr Alexandre Alcaïs, DR, Université de Paris (Directeur de thèse)

*À Dieu,
À ma mère,
À ma famille,*

Table des matières titres et sous-titres

TABLE DES MATIÈRES TITRES ET SOUS-TITRES	1
TABLE DES MATIÈRES FIGURES ET TABLEAUX	7
LES FIGURES.....	7
LES TABLEAUX.....	9
CARTES MENTALES	11
MES REMERCIEMENTS.....	17
LISTE DES ABRÉVIATIONS	19
RÉSUMÉ.....	23
ABSTRACT	25
LISTE DES PUBLICATIONS	27
INTRODUCTION ET MOTIVATIONS DE L'ÉTUDE.....	29
PARADIGME DES MALADIES INFECTIEUSES	33
CHAPITRE 1 : LA LÈPRE.....	41
1.1 Un peu d'histoire.....	41
1.1.1 Hypothèse historique : origine et diffusion de la lèpre.....	41
1.1.2 Hypothèse « génomique » : origine et diffusion de la lèpre.....	43
1.1.3 Stigmatisation des patients atteints de la lèpre.....	44
1.1.4 La paléomicrobiologie	46
1.2 Histoire de <i>Mycobacterium leprae</i>	49
1.2.1 Taxonomie	49
1.2.2 Morphologie et structure	49
➤ Microscopie optique.....	49
➤ Microscopie électronique.....	50
1.2.3 Constitution chimique et antigénique de la paroi.....	50
1.2.4 Identification de <i>M. leprae</i>	50
1.2.5 Réservoirs naturels	53
1.2.6 Croissance de <i>M. leprae</i>	55
1.2.7 Modèles d'étude.....	55

1.2.8	Génome de <i>M. leprae</i>	57
➤	Description du génome de <i>M. leprae</i>	57
➤	<i>M. leprae</i> versus <i>M. lepromatosis</i>	58
1.3	Spectre clinique et classification.....	62
1.3.1	Forme tuberculoïde	62
1.3.2	Forme lépromateuse	63
1.3.3	Classification de Ridley-Jopling et formes « borderline »	64
1.3.4	Classifications de l'OMS.....	65
1.4	Épidémiologie de la lèpre.....	68
1.4.1	Campagne éradication de l'Organisation Mondiale de la Santé.....	68
1.4.2	Situation actuelle.....	70
1.4.3	Facteurs influençant le développement de la lèpre.....	74
➤	Facteurs d'exposition : contact avec un sujet atteint	74
➤	Géographie	74
➤	Dépistage tardif, niveau socio-économique et nutrition	75
➤	Âge	76
➤	Sexe.....	77
➤	BCG	77
➤	VIH	78
1.4.4	La pléiotropie :	80
➤	Maladies inflammatoires de l'intestin : la maladie de Crohn	80
➤	Lèpre et COVID-19 : un lourd impact pour les patients	81
•	Un accès aux soins très limité	81
•	Interaction entre <i>M. leprae</i> et le SARS-Cov2	82
1.5	Diagnostic clinique et paraclinique de la lèpre.....	83
1.5.1	Diagnostic clinique.....	83
➤	Interrogatoire du patient.....	83
➤	Examen de la peau.....	84
➤	Test de sensibilité des taches cutanées	84
➤	Palpation des nerfs	84
➤	Examen des mains et des pieds.....	84

1.5.2	Diagnostic paraclinique	85
➤	Histopathologique : le Gold Standard	85
➤	Bactériologique : L'index bacillaire (IB) ou Morphologique (IM)	86
➤	Immunologique : le test de Mitsuda	86
➤	Immunologique : la sérologie	87
➤	L'apport de la biologie moléculaire : la PCR	87
1.5.3	Infection infraclinique.....	88
1.6	Antibiothérapie, séquelles, prophylaxie	89
1.6.1	Antibiothérapie : historique et recommandations de l'OMS.....	89
➤	Historique	89
➤	Recommandations de l'OMS	90
➤	Rechutes	91
1.6.2	Antibiothérapie : réactions lépreuses	94
➤	Les réactions de type 1 ou réaction de réversion	94
➤	Les réactions de type 2 ou Erythème Nouveau Lépreux.....	95
➤	Prise en charge des réactions lépreuses	95
1.6.3	Séquelles.....	96
1.6.4	Prophylaxie	96
➤	« Éducation » des populations et prévention du handicap	96
➤	Un traitement prophylactique : une dose unique de rifampicine	97
➤	La vaccination	98
1.7	Physiopathologie de l'infection à <i>M. leprae</i>	99
1.7.1	Transmission	99
1.7.2	Immunologie : la région HLA	101
1.7.3	Interaction de <i>M. leprae</i> avec le macrophage	104
1.7.4	Mécanisme de l'atteinte nerveuse.....	104
1.7.5	Immunologie : polarisation de la réponse immunitaire de l'hôte	107
1.8	Génétique humaine de la lèpre	111
1.8.1	La lèpre : une maladie « héréditaire » ?.....	111
1.8.2	Modèles expérimentaux chez l'animal.....	118

1.8.3	Épidémiologie génétique	120
➤	Analyses sans marqueurs génétiques	120
•	Analyses d'agrégation familiale	120
•	Analyses de jumeau	122
•	Analyses de ségrégation	123
➤	Analyses avec marqueurs génétiques	129
•	Analyse de liaison.....	130
•	Analyse d'association.....	136
1.9	Motivations et objectifs	148
CHAPITRE 2 : ANALYSE D'ASSOCIATION PANGÉNOMIQUE FAMILIALE AU VIETNAM.....		149
2.1	Données phénotypiques	149
2.1.1	Région d'étude : le Vietnam, milieu physique et humain	149
➤	Le Vietnam.....	149
➤	La lèpre au Vietnam depuis le recrutement jusqu'à aujourd'hui	150
2.1.2	Recueil et description de l'échantillon	153
➤	Stratégie de recueil des données	153
➤	Description de l'échantillon initial.....	157
➤	Choix du phénotype étudié	159
2.2	Données génétiques	163
2.2.1	Génotypage	163
➤	Principe général de génotypages par puces à ADN.....	163
➤	Génotypage des individus de notre échantillon.....	166
2.2.2	Le contrôle qualité.....	168
➤	Marqueurs	168
➤	Individus.....	170
2.2.3	Imputation de génotypes additionnels	174
2.3	Méthodologie	177
2.3.1	Analyse en composante principale.....	177
2.3.2	Déséquilibre de liaison	179

2.3.3	Analyses d'association pangénomique	181
➤	Méthodes d'association	181
➤	Modèles génétiques : codage du phénotype	182
➤	Principe du Test de Déséquilibre de transmission	186
➤	Extension du TDT	188
➤	FBAT : Family-Based Association Test	189
➤	FBAT Dosage : cas des SNPs imputés	193
➤	P-values minimales	194
➤	Lecture des résultats	195
2.3.4	Réplication des signaux de la cohorte primaire	199
➤	Cohorte de réplication.....	199
➤	Sélection des SNPs pour l'étude de réplication.....	200
➤	Méthode d'analyse : la régression logistique.....	202
2.3.5	Analyse combinée : les « pseudo-contrôles »	203
➤	Définition des « pseudo-germains ».....	203
➤	Modèle statistique.....	204
➤	Stratification sur les covariables non génétiques.....	206
2.3.6	Analyse multivariée	207
➤	Analyses haplotypiques	207
➤	Analyse par régression logistique conditionnelle.....	208
2.4	Résultat de l'analyse d'association pangénomique	211
2.4.1	Résultat de l'analyse familiale primaire par FBATdos	211
➤	Robustesse de l'analyse : QQ-plot.....	211
➤	Résultats d'association : Manhattan-plot	212
2.4.2	Résultat de la réplication en cas-contrôles	214
➤	Résultats des associations univariées.....	214
➤	Résultats des associations multivariées dans la région HLA	218
2.5	Conclusion sur l'analyse d'association familiale pangénomique	224

CHAPITRE 3 : VALIDITÉ DE LA RÉPLICATION PARENTALE DANS LES GWAS FAMILIAUX.....	229
3.1 Méthodes et simulations utilisées	229
3.1.1 Paramètres utilisés pour la simulation	229
3.1.2 Hypothèses de simulation	231
3.2 Résultats.....	232
3.2.1 Indépendance du TDT et du CCP	232
3.2.2 Puissance du CCP sous H1	233
3.3 Exemples	233
3.4 Conclusion sur la réplication parentale dans le GWAS familial	238
 CHAPITRE 4 : CONCLUSION GÉNÉRALE ET PERSPECTIVES	 241
 BIBLIOGRAPHIE.....	 253
 ANNEXES	 293
ANNEXE 1 : Résultats d'association des 33 signaux hors-HLA associés à la lèpre, et précédemment publiés, lors de l'étude primaire familiale.....	293
ANNEXE 2 : Résultats d'association des 358 SNPs dans la région HLA dépassant le seuil évocateur lors de l'étude primaire familiale	297
ANNEXE 3 : Article 1	313
ANNEXE 4 : Article 2	331
ANNEXE 5 : Article 3	335

Table des matières figures et tableaux

LES FIGURES

Figure 0 : De la théorie microbienne à la théorie génétique des maladies infectieuses.....	37
Figure 0bis : Caractéristiques de la génétique mendélienne et épidémiologique	39
Figure 1 : Dissémination de la lèpre dans le monde en fonction du génome du bacille.....	47
Figure 2 : Exclusion des lépreux	48
Figure 3 : Lèpre et Église	48
Figure 4 : Observations de <i>M. leprae</i>	51
Figure 5 : Représentation schématique de l'enveloppe de <i>M. leprae</i>	52
Figure 6 : Hôtes de <i>M. leprae</i> , le tatou à 9 bandes et l'écureuil roux.....	56
Figure 7 : Modèles d'étude de <i>M. leprae</i>	56
Figure 8 : Carte du génome de <i>M. leprae</i> avec la position et l'orientation des gènes connus, des pseudogènes et des séquences répétées	60
Figure 9 : Comparaison des génomes de <i>M. leprae</i> et <i>M. tuberculosis</i>	60
Figure 10 : Pourcentage d'échantillons contenant <i>M. leprae</i> et/ou <i>M. lepromatosis</i>	61
Figure 11 : Phylogénie des souches de <i>M. lepromatosis</i> et <i>M. leprae</i>	61
Figure 12 : Images cliniques de la lèpre tuberculoïde et lépromateuse.....	66
Figure 13 : Répartition géographique des nouveaux cas de lèpre dans le monde en 2018....	73
Figure 14 : Test de la sensibilité des paumes de la main et des plantes des pieds lors du diagnostic clinique	85
Figure 15 : Plaquette de PCT des formes multibacillaires et paucibacillaires	93
Figure 16 : Caractéristiques de la région HLA chez l'Homme	103
Figure 17 : Les macrophages infectés par <i>M. leprae</i> sont à l'origine de la destruction des cellules de Schwann	106
Figure 18 : <i>M. leprae</i> est capable de reprogrammer les cellules de Schwann	106
Figure 19 : Spectre immuno-clinique de la lèpre	110

Figure 20 : Immunopathologie du spectre de la lèpre.....	110
Figure 21 : Du contact avec <i>M. leprae</i> au développement de signes cliniques.....	116
Figure 22 : Représentation schématique des facteurs de risque génétiques établis, conférant une susceptibilité accrue à la lèpre per se ou à certains sous-types de lèpre	117
Figure 23 : Prédiposition génétiques aux maladies infectieuses en fonction du nombre de gènes impliqués, de la pénétrance allélique et du nombre de pathogènes	126
Figure 24 : Stratégies d'analyse communément utilisées en épidémiologie génétique pour l'étude des maladies à hérédité complexe	144
Figure 25 : Carte des zones de liaison et des gènes au sein desquels des variants ont été identifiés comme associés à la lèpre	145
Figure 26 : Le Vietnam.....	152
Figure 27 : Statistiques aux Vietnam.....	152
Figure 28 : Fiche individuelle utilisée pour le recrutement des individus	156
Figure 29 : ACP des individus fondateurs par origine ethnique (populations HapMap)	161
Figure 30 : ACP du GWAS avec les populations de 1000 génomes Phases 3.....	162
Figure 31 : Schéma général du principe des puces à ADN de génotypage	165
Figure 32 : Principe de l'assignation de génotypes par puce à ADN.....	165
Figure 33 : Biopuce <i>Illumina Human 660W-Quad</i> et sa couverture à $r^2 > 0,8$ dans différentes populations de la base de données HapMap	167
Figure 34 : Principe de l'imputation de génotypes	175
Figure 35 : Association indirecte, le variant génotypé est en LD avec le variant causal	198
Figure 36 : Création de pseudo-germains pour l'analyse par régression logistique conditionnelle d'un marqueur bi-allélique (Aa).....	204
Figure 37 : QQ-plot de l'analyse d'association pangénomique familiale de la lèpre	211
Figure 38 : Manhattan-plot, résultats d'association pangénomique familiale avec la lèpre per se dans l'échantillon primaire vietnamien.....	213
Figure 39 : Association des SNPs HLA de classe I et II dans la réplication cas-contrôles.....	220
Figure 40 : Indépendance entre le TDT et le cas-contrôle parental	234
Figure 41 : Puissance du TDT et de l'analyse cas-témoins parentale en fonction de la prévalence parentale	236

LES TABLEAUX

Tableau 0 : Caractéristiques de la génétique mendélienne et épidémiologique	38
Tableau 1 : Caractéristiques du spectre clinique de la lèpre selon la classification de Ridley-Jopling	67
Tableau 2 : Antibiotiques actifs sur <i>M. leprae</i>	92
Tableau 3 : Recommandations de l'OMS pour le traitement de la lèpre	93
Tableau 4 : Principales analyses de ségrégation sur la lèpre	127
Tableau 5 : Marqueurs génétiques utilisés en épidémiologie génétique	129
Tableau 6 : Principales méthodes utilisées en épidémiologie génétique	143
Tableau 7 : Variants, hors-HLA, associés à la lèpre per se et précédemment publiés	146
Tableau 8 : Périodes de recrutement du laboratoire au Vietnam entre 1991 et 2008	155
Tableau 9 : Répartition des 2 235 individus génotypés selon le type de famille	160
Tableau 10 : Caractéristiques phénotypiques de l'échantillon initial	160
Tableau 11 : Caractéristiques de la puce Illumina Human 660W-Quad	167
Tableau 12 : Caractéristiques cliniques de l'échantillon familiale primaire après QC	173
Tableau 13 : Risques relatifs génotypiques de chaque modèle génétique	183
Tableau 14 : Table de contingence du TDT	186
Tableau 15 : Probabilités de cellules du tableau de contingence du TDT	187
Tableau 16 : Codage des génotypes selon les différents modèles génétiques	190
Tableau 17 : Espérance sous H_0 du génotype X d'un enfant, notée $E(X)$, selon les génotypes parentaux et le modèle génétique considéré	191
Tableau 18 : Description de l'échantillon de réplication	200
Tableau 19 : Résultats d'associations suite à la procédure de « clumping » pour les 8 signaux évocateurs, hors-HLA, de l'échantillon primaire	216
Tableau 20 : Résultats d'association dans l'échantillon primaire et la réplication cas-contrôles pour les 6 signaux significatifs de la région HLA et les 5 signaux, hors-HLA, précédemment publiés	217

Tableau 21 : Résultats de l'analyse multivariée dans la région HLA (échantillon de réplication)	221
Tableau 22 : Résultats d'association dans l'échantillon primaire, la réplication et l'analyse combinée pour les 5 signaux indépendants identifiés par l'analyse multivariée.....	222
Tableau 23 : Caractéristiques des modèles génétiques simulés.....	230
Tableau 24 : Statistique du test génotypique	231
Tableau 25 : Distribution des faux positifs du cas-contrôle parental selon l'erreur de type I et en fonction des résultats du TDT	235
Tableau 26 : Résultats d'association dans l'échantillon primaire familial des 33 SNPs hors-HLA et précédemment publiés.....	295
Tableau 27 : Résultats d'association de l'étude primaire familiale et de la réplication pour les 358 SNPs HLA dépassant le seuil évocateur	312

CARTES MENTALES

Carte mentale 1 : La lèpre	147
Carte mentale 2 : Contrôle qualité de l'échantillon primaire (marqueurs et individus)	176
Carte mentale 3 : Méthodes du GWAS	210
Carte mentale 4 : Principaux résultats du GWAS.....	223
Carte mentale 5 : Réplication cas-contrôle parentale	237

La lépreuse abandonnée de Taenga

Non, cela ne s'est pas passé au Moyen-Âge, mais en 1930 au Tuamotu.

Lorsque le médecin qui effectuait sa tournée annuelle est arrivé à Taenga un petit atoll des Tuamotu, il a examiné la jeune femme qui avait quelques tâches suspectes. « Cela pourrait être la lèpre » a-t-il dit à mi-voix. Il l'avait dit trop fort cependant. À peine était-il remonté à bord que le chef du village faisait saisir la femme « suspecte ». On l'arracha à son mari, à ses cinq enfants. On la mena de force dans une pirogue et on la jeta sur un motu (îlot) situé à 2 heures de bateau du village. Elle et son chien. On n'entendit ni ses pleurs, ni ses cris, ni l'animal épouvanté qui hurlait à la mort. Cette femme avait vingt-cinq ans...



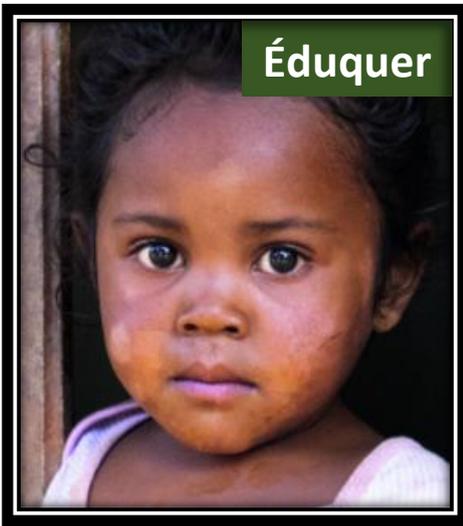
Pendant six ans, elle vécut seule. Seule avec son chien. Dressé pour la pêche, il allait lui chercher sa misérable pitance. Chaque semaine, une pirogue s'approchait du rivage maudit. Sans descendre, on lui jetait quelques vivres et un bidon d'eau. « La dernière fois, elle était entièrement rongée... Ses pieds ne la supportaient plus... Je lui ai laissé un peu de farine, mais qu'en fera-t-elle ? Ses doigts sont tellement pourris qu'en pétrissant la pâte, elle y laissera des morceaux de sa propre chair... » raconta, les yeux pleins de larmes, l'homme qui la ravitaillait. « Un jour elle ne pourra plus se lever. Le chien aura faim, très faim, très faim... Il flairera l'agonisante ».

Vers la fin de ces six années, un médecin s'approcha en bateau du motu. « J'ai eu le temps de compter les huit cocotiers qui en sont l'ornement ! Quelques arbustes rabougris les entourent : c'est tout. Une mesure faite avec rien, pire, avec n'importe quoi. Elle est sortie et m'a crié « N'approche pas ! J'ai la lèpre... ». Je lui ai répondu « Qu'est-ce que tu veux que ça me foute : je suis toubib ». Alors elle m'a souri. Son premier sourire depuis tant d'années. Je lui ai remis ce que j'avais : quelques pansements, un peu d'aspirine. Les sulfones n'étaient pas encore arrivés en Océanie. Elle m'a supplié de l'emmener. Je ne pouvais pas. L'équipage refusait. L'équipage avait déclaré qu'il quitterait le bateau si elle y montait. Elle n'a pas eu de colère, pas même d'amertume. Elle m'a dit : Je comprends. Alors j'ai eu envie de lui demander pardon. Et quand je suis parti, il s'est passé quelque chose, quelque chose d'incroyable. Elle m'a crié : la Orana – la Orana ! (Au revoir, au revoir) ... Puis elle a chanté la « Marseillaise » ! Je serais les poings pour ne pas pleurer... tandis que j'entendais, dans le silence bleu du Pacifique, la voix rauque qui s'élevait : Amour sacré de la Patrie... »

Cinq jours plus tard, le navire Tamara avec, à son bord l'administrateur des Tuamotu M. Ahnne et un infirmier partait la récupérer. Et deux jours après, la lépreuse de Taenga débarquait sur un brancard à la Pointe Vénus en chantant la « Marseillaise ». Et pendant tout le temps qu'elle chanta, le Colonel demeura au garde à vous, c'était un émouvant tableau. Après la Marseillaise, elle cria : « Merci la France, je suis sauvée. Merci, merci ! ». Enfin, elle arriva à la léproserie d'Orofara à Mahina et tous les malades lui souhaitèrent la bienvenue.

Raoul Follereau, Tour du monde chez les lépreux, 1953, Flammarion Paris p 134

En 2020...



1 malade de la lèpre
sur **10** est un enfant



+150 pays touchés

+200 000 nouveau cas
dépistés chaque année

1 nouveau cas toutes les **2**
minutes

3 millions de malades avec
des invalidités/mutilations



+16 millions de
malades guéris sur les 20
dernières années



« La lèpre est une maladie de la pauvreté qui nécessite des soins coûteux et longs en cas de complications chez des patients très démunis. Les malades infectés par la lèpre sont stigmatisés du fait de l'absence de sensibilisation de la population. Enfin, les malades affectés par la lèpre et leur famille sont peu suivis et accompagnés au niveau psychosocial. »

Dr Lahla Fall, dermatologue spécialiste de la lèpre au centre hospitalier de l'ordre de malte (Dakar, 2020)

« Il est des plaies qui, pareilles à la lèpre, rongent l'âme, lentement, dans la solitude. »

Sadegh Hedayat, La chouette aveugle, 1953

« Il a eu peur à cause des conséquences, l'infirmité, le rejet social, tout ce qui est associé à la lèpre. Les gens te fuient, c'est comme si Dieu t'avait abandonné. »

Dr Fomba, léproserie à Bamako (Mali, 2018)

« Les lépreux qu'est-ce que cela peut leur faire qu'on leur donne quelque chose, si on ne leur donne pas la main. »

Raoul Follereau, Trente fois le tour du monde, 1961

« La maladie s'est déclarée il y a plus de 5 ans par des douleurs au niveau des mains et des gonflements aux pieds. (...) Au début, quand j'ai su que j'avais la lèpre, j'en ai parlé à ma femme et à ma famille. J'ai été stigmatisé au début, mais ma femme m'a toujours soutenu et accompagné. »

Homme de 52 ans (Sénégal, 2020)

« Le lépreux, atteint de la plaie, portera ses vêtements déchirés, et aura la tête nue; il se couvrira la barbe, et criera: Impur! Impur! Aussi longtemps qu'il aura la plaie, il sera impur: il est impur. Il habitera seul; sa demeure sera hors du camp. »

Protocole observé à l'égard des lépreux au Moyen Âge (inspiré de la Bible)

Mes remerciements

À mon jury de thèse

Je tiens à remercier très sincèrement Antoine Mahé et Bertram Müller-Myhsok de m'avoir fait l'honneur d'être rapporteur de cette thèse et d'avoir toléré quelques retards calendaires. Mes chaleureux remerciements vont également à Laure Gineau, Nabila Bouatia-Naji et Roch-Christian Johnson qui ont accepté d'être examinateurs. C'est un grand honneur pour moi d'avoir pu réunir un tel jury, dont les champs de compétences complémentaires permettent de couvrir de nombreux aspects de ce travail. J'adresse un grand merci à tous les membres du jury pour le temps passé et l'attention portée à ce manuscrit et espère qu'ils auront trouvé un intérêt à sa lecture...

À Laurent Abel

Co-directeur du Laboratoire de Génétique Humaine des Maladies Infectieuses, je le remercie très sincèrement pour son accueil, ses conseils précieux, sa gentillesse, sa patience et sa grande rigueur scientifique.

À Jean Laurent Casanova

Je tiens également à remercier Jean-Laurent Casanova de m'avoir accueillie dans l'équipe qu'il dirige avec Laurent Abel. C'est une grande chance d'avoir pu réaliser ma thèse dans un laboratoire d'un tel niveau scientifique.

À mes amis du laboratoire, passés et présents

Je remercie chaque membre de l'équipe, pour leur partage enthousiaste d'idées scientifiques et... moins scientifiques, leur présence et... leur folie. Une pensée particulière pour les épidémiologistes : Fabienne pour sa générosité, elle a guidé mes premiers pas dans l'univers de l'épidémiologie génétique ; Aurélie pour son efficacité et son aide toujours précise ; Frédégonde pour sa joie de vivre et sa sincérité ; Jérémy pour ses discussions et sa disponibilité ; Jocelyn pour sa précision et son amour du PSG ; Yohann pour sa simplicité. Gaspard, Vimel et Rosa, le « trio sans parents », je les remercie pour leur amitié sans faille et pour réussir à me supporter tous les jours en vrai ou sur WhatsApp, particulièrement, dans cette dernière ligne de cette aventure exigeante.

À Alexandre Alcaïs

Après m'avoir accompagnée et guidée lors de mon Master Recherche, il a été mon directeur de thèse. Je voudrais remercier très sincèrement, AA, pour son enthousiasme, sa patience, sa sincérité infaillible, sa confiance absolue, son aide constante, sa présence... tout simplement pour ce qu'il est et tout ce qu'il m'a apporté et m'apportera (« jamais je ne vous laisserai tranquille AAAAA »). C'est un immense privilège et toujours un grand plaisir de travailler et discuter avec lui, j'admire la richesse de ses idées, son écoute, sa capacité à se remettre en question et bien d'autre de ses qualités, au-delà, du scientifique. Il a été pour moi un repère, un protecteur contre tous, un révélateur de qui je suis (les rayures ça rapprochent 😊), il fait partie de ces relations que l'on croise que très rarement et qui restent gravées à jamais. Merci pour toutes ces nuits blanches à travailler ensemble et à refaire le monde. Merci à Alexandre Alcaïs, le directeur, le père, l'ami...

À ma famille et mes amis

Merci à ma famille et à mes amis qui me suivent dans toutes les circonstances. Merci à « Mama », sans elle, je ne serais pas là (outre la définition parentale), merci pour ses sacrifices et pour avoir été, tous les jours depuis bientôt 30 ans, présente à mes côtés. Merci à Papa pour son amour de la science (ça se transmet génétiquement apparemment) ; Heger pour sa rigueur et ses ordonnances ; Sarra pour son réalisme et sa présence (elle fait partie des personnes sur qui l'on peut toujours compter, et ce, à tout moment) ; Heythem pour notre enfance, Omeyma pour sa douceur, sa bienveillance et ses trolls permanents, Hamoudh pour son humour et son bricolage dans toutes circonstances. Enfin, Houcem, le dernier de la fratrie, je le remercie pour sa confiance inaltérée, sa profondeur, sa protection, sa loyauté et... tout simplement pour notre « vérité » qui, moi-même, me dépasse. Merci aux enfants, en particulier à mes nièces et mon neveu, Lémys, Nora, Lilia et Leyth d'exprimer si librement leur énergie de vie et de la partager. Merci à Cimoun'z d'être constamment là malgré mon caractère explosif. Merci à Mika d'avoir relu et corrigé (orthographiquement) ce manuscrit. Merci à Julie S, Amani, Sun, Julie D, Noémie, Hugo et tant d'autres, d'être toujours présents à mes côtés.

À tous ceux qui m'ont aidé et permis d'en être là aujourd'hui 😊

Liste des abréviations

- 1KG : 1000 génomes
- ACP : Analyse en Composantes Principales
- ADN : Acide DésoxyriboseNucléique
- ARNm : Acide Ribo-Nucléique messenger
- B : Borderline
- BAAR : Bacille Acido-Alcool-Résistant
- BB : Borderline Borderline
- BCG : Bacille de Calmette-Guérin
- BH : Benjamini et Hochberg
- BL : Borderline Lépromateuse
- BT : Borderline Tuberculoïde
- CC : Cas-Contrôles
- CCP : Cas-Contrôle Parental
- CHR : Chromosome
- CMH : Complexe Majeur d'Histocompatibilité
- CNV : Copy Number Variants
- CR : Callrate
- CRAN : Comprehensive R Archive Network
- ddl : Degré de liberté
- DM : Differential Missingness
- DV : Dermato-Vénérologie
- DZ : Dizygote
- ELISA : Enzyme-Linked Immunosorbent Assay
- EM : Erreur Mendélienne
- ENL : Erythème Nouveux Lépreux
- FBAT : Family-Based Association Tests
- FBATdos : Family-Based Association Tests Dosage
- FDR : False Discovery Rate

- FP : Faux Positif
- GATK : Genome Analysis Toolkit
- gnomAD = Genome Aggregation Database
- GO : Gene Ontology
- GRM : Genetic Relatedness Matrix
- gVCF : Genomic Variant Call Format
- GWAS : Genome-Wide Association Study
- HAART : High Antiretroviral Therapy
- HLA : Human Leukocyte Antigen
- HSE : Herpes Simplex Encephalitis (Encéphalite Herpétique)
- HW : Hardy-Weinberg
- HZ : Taux d'Hétérozygotie
- I : Indéterminée
- IB : Index Bacillaire
- IBD : Identity By Descent
- IBS : Identity By State
- IC : Intervalle de Confiance
- IFG : Facteur d'Inflation Génomique
- IM : Index Morphologique
- INDEL : Insertion-Délétion
- iNOS : inducible Nitric Oxyde Synthase
- LL : Lépromateuse Lépromateuse
- LLs : Lépromateuse Lépromateuse subpolaire
- m/M = allèles mineur/Majeur
- MAF : Minor Allele Frequency
- MB : Multibacillaire
- MC : Maladie de Crohn
- MICI : Maladies Inflammatoires Chroniques de l'Intestin
- MV : Multivarié
- MZ : Monozygote
- NA : Non Attribué

- NS : Non Significatif
- OMS : Organisation Mondiale de la Santé
- OR : Odds Ratio
- p : p value
- p-val : p value
- PB : Paucibacillaire
- PC : Principal Component
- PC : Principal Component Analysis
- PCR : Polymerase Chain Reaction
- PCT : PolyChimioThérapie
- PGL-1 : Phenolic Glycolipid I
- PIDs : primary immunodeficiencies
- REML : Maximum de Vraisemblance Restreint
- RR : Réaction de Réversion
- S : Significatif
- SCID : Severe Combined ImmunoDeficient
- SIDA : Syndrome d'ImmunoDéficiency Acquis
- SIDCS : Syndrome d'ImmunoDéficiency Combinée Sévère
- SNP : Single Nucleotide Polymorphism
- T : Tuberculoïde
- T1R : Type 1 Reaction
- TDT : Transmission Disequilibrium Test
- TNF : Tumor Necrosis Factor
- TT : Tuberculoïde Tuberculoïde
- QC : Quality Control
- VIH : Virus de l'ImmunoDéficiency Humaine
- WES : Whole Exome Sequencing
- WHO : World Health Organization (WHO)

Résumé

La lèpre, maladie infectieuse chronique causée par *Mycobacterium leprae*, affecte principalement la peau, les nerfs et les yeux avec des séquelles majeures en l'absence de traitement. Avec 200 000 nouveaux cas diagnostiqués chaque année (1 toutes les 2 minutes), il s'agit de la mycobactériose la plus commune après la tuberculose et requalifiée « maladie tropicale négligée » en 2017 par l'Organisation Mondiale de la Santé (OMS). Si la contribution génétique de l'hôte dans l'histoire naturelle de la maladie est maintenant bien établie, son architecture reste lacunaire. Dans cette continuité et afin de la préciser, nous avons, pour la première fois, utilisé une approche de génétique épidémiologique familiale.

Plus précisément, nous avons réalisé la première étude d'association pangénomique (Genome-Wide Association Study, GWAS) familiale sur la lèpre. Ainsi, au cours des 20 dernières années, un échantillon de 481 familles nucléaires, parents et enfants, sélectionnées à partir d'un enfant atteint, a été constitué au Vietnam. Sur cet échantillon primaire de 1 749 individus incluant 622 enfants atteints, nous avons testé l'association de près de 6 millions de variants bi-alléliques (Single Nucleotide Polymorphism, SNP), génotypés ou imputés, avec la lèpre. Dans un second temps, nous avons testé les signaux les plus prometteurs dans un échantillon de réplication, c'est-à-dire, indépendant et issu de la même population, constitué de 1 181 cas et 668 contrôles.

Les résultats les plus significatifs ont été observés au sein de la région HLA et l'analyse multivariée a permis d'identifier trois signaux indépendants. Deux dans la région HLA classe I : rs1265048 [Odds-ratio (OR) = 0,69 ; p-val = $5,5 \cdot 10^{-11}$] et rs114598080 [OR = 1,47 ; p-val = $8,8 \cdot 10^{-13}$] ; Et un dans la région HLA classe II : rs3187964 [OR = 1,67 ; p-val = $8,4 \cdot 10^{-16}$]. Nous avons également identifié deux signaux hors-HLA : un variant faux-sens dans le gène *LACC1* (rs3764147 : OR = 1,52 ; p-val = $5,1 \cdot 10^{-14}$), et un variant à proximité du gène *IL12B* (rs6871626 : OR = 0,73 ; p-val = $6,4 \cdot 10^{-8}$).

Les contraintes de coûts des études pangénomiques imposent une réduction majeure du nombre de SNPs à tester dans d'autres échantillons. Dans les études familiales, les parents sont de fait génotypés et pourraient permettre une réplification immédiate sans coûts ajoutés. Au moyen d'une large étude de simulation, nous avons montré que cette approche était pertinente. Une étude cas-contrôle chez les parents de l'échantillon primaire est une réplification valide, statistiquement indépendante de l'étude d'association familiale. C'est un argument fort en faveur des approches familiales pour l'exploration pangénomique de la contribution génétique de l'hôte dans les phénotypes complexes.

La compréhension de la physiopathologie de l'infection à *M. leprae* est cruciale pour optimiser les approches préventives selon les profils génétiques à plus haut risque, et ouvrir de nouvelles pistes thérapeutiques en précisant les cascades fonctionnelles pertinentes. En ce sens, la dissection du contrôle génétique de l'infection par l'hôte est indispensable. Enfin, remettre la famille au cœur de la quête génétique, c'est remettre la génétique dans son milieu naturel.

Mots-clés :

Maladie infectieuse, Épidémiologie, Biostatistiques, Génétique humaine, Mycobactériose, Lèpre, Mycobacterium leprae, Vietnam, Familles, Analyse d'association pangénomique (GWAS), Approche familiale (Family-Based), Réplification (Population-Based), Études cas-contrôles parentales.

Abstract

Leprosy is a chronic infectious disease caused by *Mycobacterium leprae*. It primarily affects the skin and peripheral nerves, and can cause an irreversible impairment of nerve function, often leading to severe disabilities and social stigma if left untreated. The disease, re-qualified by WHO (World Health Organization) as a “Neglected Tropical Disease” in 2017, remains a major public health problem in regions of endemic countries, with over 200,000 new cases per year (one every two minutes). It is ranked second as the most common mycobacterial infectious disease, right after tuberculosis. While it has been well established that there is a genetic contribution to this disease, the underlying genetic causes remains unknown. In our study, we sought to reveal the host’s genetic architecture of leprosy by taking of a familial epidemiological approach.

We conducted the first Family-Based Genome-Wide Association Study (GWAS) of leprosy in 481 Vietnamese nuclear families (parents and children) selected based on one affected child and collected over the past 20 years. Using this sample of 1,749 individuals, including 622 affected offspring, we performed association tests between six million biallelic genetic variants (Single-Nucleotide Polymorphism, genotyped or imputed) and the binary phenotype of disease status. Following this first analysis, we conducted a replication analysis of the most promising results in an independent sample of the same ethnic origin, accounting for 1,181 cases and 668 controls.

The most significant results were observed within the HLA (Human Leukocyte Antigen) region, in which 3 independent SNPs displayed genome-wide significant associations. Among these, two were for the HLA class I region and one for the HLA class II (rs1265048 [OR = 0.69; p-value = 5.5×10^{-11}], rs114598080 [OR = 1.47; p-value = 8.8×10^{-13}] and rs3187964 [OR = 1.67; p-value = 8.4×10^{-16}] respectively). We also identified a missense variant in the *LACC1* gene (rs3764147: OR = 1.52; p-value = 5.1×10^{-14}) and an intergenic variant located close to the *IL12B* gene (rs6871626: OR = 0.73; p-value = 6.4×10^{-8}). *LACC1* encodes a central regulator of the metabolic function and bioenergetic state of macrophages and *IL12B* encodes IL-12p40, which is common to two interleukins, *IL-12* and *IL-23*.

Large GWAS are expensive, strongly limiting the number of variants to test in a replication set. Here, we took advantage of the available parental phenotypic and genotypic information to perform a classical case-control study among the parents of the family-based sample. Indeed, using of extensive computer simulations, we demonstrated that this population-based parental study is a valid, powerful and costless replication strategy to confirm family-based associations. Overall, our observations add to the attractiveness of family-based designs and should provide valuable help for investigators planning to perform GWA studies.

Understanding leprosy pathophysiology infection is crucial to optimize preventive approaches based on genetic profiles. Dissection of the genetic control of the infection by *M. leprae* by its human host, therefore, constitutes an indispensable step. Finally, repositioning the family at the heart of the genetic quest means repositioning genetics into its natural environment.

Keywords :

Infectious disease, Epidemiology, Biostatistics, Human genetics, Mycobacteriosis, Leprosy, Mycobacterium leprae, Vietnam, Families, Genome-Wide Association Study (GWAS), Family-Based Approach, Replication (Population-Based), Parental case-control studies.

Liste des publications

Publications issues directement du travail de thèse

- **Gzara C**, Dallmann-Sauer M, Orlova M...//...Abel L, Alcaïs A, Schurr E, Cobat A.
Family-based genome-wide association study of leprosy in Vietnam.
PLoS Pathog. **2020 May** 18;16(5):e1008565.doi: 10.1371/journal.ppat.1008565.
- **Gzara C**, Mallat L...//...Abel L, Alcaïs A.
Family-based associations: case-control replication for free.
Open Journal of Statistics (Minor Revision)
- **Gzara C**, Gaschignard J...//...Abel L, Alcaïs A.
A Family-based genome-wide association study of leprosy polarization in Vietnam (À soumettre très prochainement)

Publications issues de collaborations au sein du laboratoire au cours de la thèse

- Dallmann-Sauer M, Fava VM, **Gzara C**...//...Alcaïs A, Abel L, Cobat A, Schurr E.
The complex pattern of genetic associations of leprosy with HLA class I and class II alleles can be reduced to four amino acid positions.
PLoS Pathog. **2020 Aug** 10;16(8):e1008818.doi: 10.1371/journal.ppat.1008818.
- Fava VM, Cobat A, **Gzara C**, Alcaïs A, Abel L, Schurr E.
The differential role of LRRK2 variants in nested leprosy phenotypes.
Proc Natl Acad Sci U S A. **2020 May** 12;117(19):10124-10125. doi: 10.1073/pnas.2002654117.

Introduction et motivations de l'étude

Dans ce manuscrit, vous allez, je l'espère, vous immerger dans la démarche scientifique que j'ai entreprise au cours de mes 4 années de thèse au Laboratoire de Génétique Humaine des Maladies Infectieuses. Plus précisément, dans l'équipe de Génétique épidémiologique dirigée par Laurent Abel et dont le principal objectif est de « développer des modèles mathématiques et statistiques originaux et de les appliquer en particulier à l'étude du déterminisme génétique des maladies infectieuses communes ». Ce manuscrit va vous retracer l'essentiel de mes travaux de recherche de la manière la plus claire, la plus précise et, je l'espère, la plus sincère. J'ai essayé de respecter le plus possible la chronologie exacte des événements. Parfois, j'ai eu l'impression qu'une légère interversion pouvait aider à la compréhension du document et je l'ai donc effectuée. Il y a donc ici et là quelques anachronismes volontaires.

Notre principale hypothèse de travail est qu'il existe des déterminants génétiques de l'hôte qui prédisposent (ou confère une résistance) au développement de la lèpre per se (lèpre indépendamment de ses formes cliniques). Ainsi, afin de préciser la signature génétique à la prédisposition de la maladie et donc d'identifier cette composante génétique de l'hôte, notre stratégie d'analyse a associé les outils classiques mais également plus récents de l'épidémiologie génétique. *« Les activités méthodologiques et appliquées me paraissent tout à fait indissociables, car complémentaires : l'analyse de données soulève souvent des problèmes méthodologiques, et de nouvelles méthodes doivent être validées dans le contexte de données réelles »*. Cette phrase tirée de la thèse d'Alexandre Alcaïs en 2002 [1] prend ici tout son sens et est toujours d'actualité. C'est pourquoi j'ai choisi de raconter dans ce mémoire de thèse comment un travail appliqué, l'étude du déterminisme génétique de la lèpre via une étude d'association pangénomique, m'a conduit à entrevoir une nouvelle approche méthodologique dans le cadre de l'analyse de répliation dans les études d'association pangénomique au design familial.

Plus précisément, mon travail et sa réalisation s'articulent en 4 parties :

- Tout d'abord, la description de la lèpre, une maladie qui de la Bible à Emmanuel Macron a toujours résonné comme une métaphore de la condition humaine et dont, pour parodier Mark Twain, « les rumeurs sur sa disparition ont été grandement exagérées ». Principalement bibliographique, ce premier chapitre propose une lecture détaillée de cette maladie. Depuis une perspective historique jusqu'aux mécanismes physiopathologiques de l'infection à *Mycobacterium leprae*, nous passerons en revue l'épidémiologie (facteurs de l'hôte non-génétiques), le phénotype, le mode de transmission, les moyens diagnostiques et thérapeutiques de la lèpre. Un état des connaissances détaillé sur la génétique humaine de la lèpre sera présenté. Enfin, cette première partie expose de manière plus détaillée que dans cette partie, les hypothèses de travail et les motivations de notre étude.
- Le deuxième chapitre constitue le cœur de cette thèse et correspond à l'étude d'association pangénomique (GWAS) de la lèpre per se. Le recueil de notre base de données sera également décrit. La population d'étude est composée de familles vietnamiennes avec des enfants atteints et l'information génétique à notre disposition est générée par une puce à ADN pangénomique.
- Le troisième chapitre, un peu plus méthodologique, a pour but de montrer que lors d'un GWAS au design familial, il est pertinent d'utiliser les parents pour l'étude de réplique. De ce fait, un gain de temps et financier n'est pas à négliger. Cette alternative sera illustrée au travers de différentes applications dont l'analyse de nos données sur la lèpre.
- Enfin, la discussion met en perspective les principaux résultats de notre travail, aborde les analyses complémentaires qui vont le prolonger et dont certaines sont déjà en cours d'exploration, en particulier l'étude d'association pangénomique de la polarisation de la lèpre toujours au Vietnam. Elle présente aussi des approches plus originales, non restreintes à la génétique humaine, dont nous pensons qu'elles permettront de mieux disséquer la contribution de la génétique de l'hôte.

Pour conclure, il est important d'insister sur le fait que l'objectif de cette thèse est double : il s'agit en premier lieu de disséquer l'architecture génétique de l'infection à *Mycobacterium leprae* et donc de déterminer les variants génétiques associés à la lèpre per se chez l'Homme. Pour ce faire, nous avons réalisé le premier GWAS sur données familiales sur la lèpre au Vietnam. Le second objectif est de proposer une alternative concrète et efficace à la réplication dans les études d'association au design familial : la réplication parentale.

Paradigme des maladies infectieuses

2 217 579 personnes dans le monde, 436 780 aux États-Unis, 223 971 au Brésil, 154 147 en Inde, 91 822 en Angleterre, 75 862 en France... Ces chiffres représentent le nombre de décès liés à la COVID-19 dans le monde et par pays au 31 Janvier 2021. Les maladies infectieuses sont, pour cette dernière année, la première cause de décès dans « *les pays à faible revenu* » [2] et peuvent mettre « à genoux » (économiquement, financièrement, humainement, socialement) même les pays les plus riches et les plus développés. En effet, depuis maintenant un an, nous vivons la crise sanitaire la plus importante de ce dernier siècle.

« Ainsi l'humanité a-t-elle pris conscience qu'elle était vulnérable face aux maladies infectieuses, et ce début de sagesse devrait pouvoir s'accompagner de mesures actives pour surveiller et détecter l'apparition de nouvelles maladies, évaluer l'évolution des épidémies et mettre en place les moyens de prévention et de traitement permettant d'éradiquer les épidémies identifiées » Pr Raoult [3].

Depuis ses origines, l'homme est en interaction dynamique avec son environnement. D'importance particulière est sa relation avec les microbes puisque l'infection représente la plus grande cause de mortalité de toute l'histoire de l'humanité. Cette nature dynamique de l'interaction hôte-pathogène explique qu'aujourd'hui encore les maladies infectieuses sont à l'origine de plus de 13 millions de décès chaque année, et que dans les pays à faible revenu une personne sur deux meurt prématurément d'une maladie infectieuse. Les catastrophes infectieuses sont donc bien réelles et sont liées à l'évolution permanente du monde, aux déplacements humains, aux guerres, aux échanges commerciaux, aux conquêtes et, bien sûr, à l'interaction hôte-pathogène. Nous sommes, ainsi, vulnérables à une multitude de maladies infectieuses endémiques et épidémiques [4], l'espérance de vie à la naissance ne dépassait pas 25 ans jusqu'aux progrès récents de l'hygiène, de la vaccination et des traitements antimicrobiens construits sur la théorie microbienne de la maladie développée par Pasteur. En effet, l'espérance de vie était en moyenne de 25 ans dans le monde jusqu'à la révolution industrielle [5], même en Europe, seulement 35% des personnes atteignaient l'âge de 40 ans à la fin du XIX^e siècle [6]. En Angleterre, durant la même période, 60% des décès étaient dus à des maladies infectieuses, en particulier lors de graves épidémies.

L'augmentation de l'espérance de vie, qui atteint aujourd'hui 80 ans dans ces régions, est récente et contribue largement à l'explosion démographique actuelle. Elle reflète principalement les progrès accomplis dans la lutte contre les maladies infectieuses sur la base de trois facteurs : le développement de l'hygiène à partir du milieu du XIX^e siècle (prévention de la transmission de l'infection); la mise au point de vaccins à partir du début du XX^{ème} siècle (prévention de la maladie chez les individus infectés); et la mise au point de médicaments anti-infectieux à partir de la première moitié du XX^{ème} siècle (prévention de la mort chez les patients atteints d'une maladie clinique) [6].

Confrontée à la formidable diversité de microbes coévoluants, l'immunité naturelle a assuré la survie globale des espèces en dépit de la mort prématurée de la plupart des individus. En identifiant la cause des maladies infectieuses, la théorie microbienne de la maladie a résolu la plus importante question de la médecine humaine mais n'a pas expliqué les causes de la grande hétérogénéité clinique intrafamiliale au sein des familles exposées au même environnement, probablement la deuxième question la plus importante en médecine. C'est la finalité de la génétique humaine des maladies infectieuses de disséquer cette variabilité *in Natura* : c'est-à-dire là où sélection naturelle agit, dans les conditions naturelles. De telles études, de par leur insistance sur les conditions naturelles, ont le potentiel unique de définir les fonctions des gènes de l'hôte favorisées par l'évolution qui sont écologiquement pertinente pour la survie passée et présente [7].

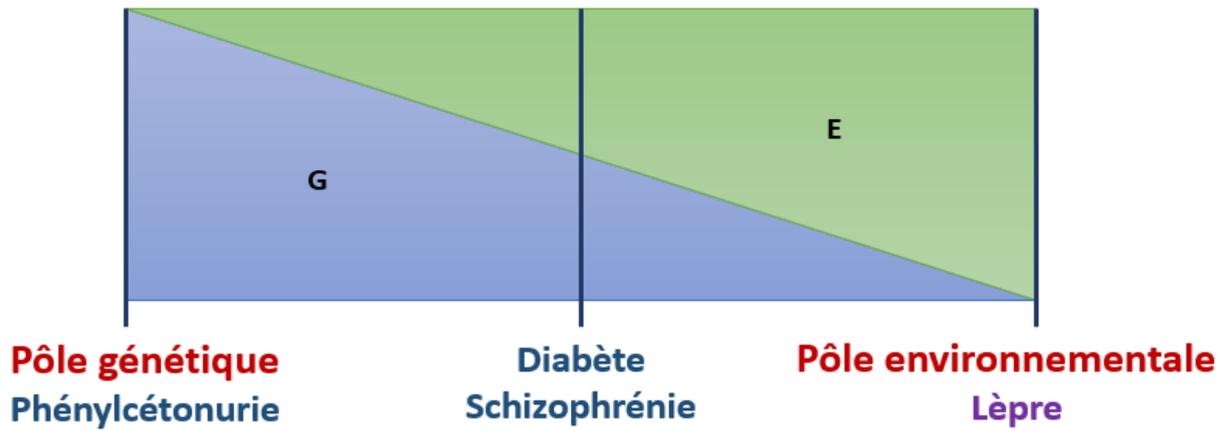
Plus précisément, depuis les travaux de Pasteur, il est admis que les agents infectieux (virus, bactéries, parasites et champignons) sont responsables des maladies infectieuses. Cependant, dès le début du XX^{ème} siècle, il a été constaté que, dans les régions d'endémie, là où les pathogènes persistent très longtemps, il existe toujours une grande variabilité de résistance aux infections selon les individus, le même pathogène pouvant être mortel chez les uns, bénin ou asymptomatique chez les autres. Ainsi, la découverte d'un grand nombre d'infections asymptomatiques a soulevé une question majeure dans le monde scientifique, a été de comprendre la variabilité interindividuelle observée au cours de l'infection. Charles Nicolle a été le premier à démontrer formellement, entre 1911 et 1917, la coexistence d'infections symptomatiques et asymptomatiques dans les populations humaines [8].

Cette variabilité a, souvent, été attribuée au rôle de l'immunité acquise, comme le démontrent les premières expériences de vaccination où les réponses immunitaires sont variables d'un individu à l'autre, cependant, elle ne peut expliquer la variabilité interindividuelle observée dans les infections primaires chez l'enfant par exemple. En effet, la description des premières immunodéficiences primaires (primary immunodeficiencies, PIDs) dans les années 1950 dues à des mutations génétiques instaura l'idée, parmi de nombreux généticiens, que la prédisposition au développement des maladies infectieuses pouvait être transmise de génération en génération. Néanmoins, l'hypothèse génétique a eu du mal à trouver des adeptes : le paradigme établi par Pasteur et Koch avec leur théorie microbienne des maladies infectieuses, a pris du temps pour être remis en cause, répétant le célèbre aphorisme de Pasteur : « *une maladie, un microbe, un vaccin* ». De plus, l'impossibilité à tester concrètement l'hypothèse génétique en l'absence de marqueurs génétiques jusque dans les années 90, a été un obstacle majeur dans le soutien de cette théorie génétique.

Dans la deuxième moitié du XX^{ème} siècle, des données d'épidémiologie génétique ont clairement établi que la génétique humaine joue un rôle central dans la susceptibilité-résistance aux maladies infectieuses. Plus précisément, certaines études agissent comme pierres fondatrices d'une théorie génétique des maladies infectieuses. Tout d'abord, l'étude des formes sévères de malaria causée par *Plasmodium falciparum* par Anthony Allison est l'une des plus remarquables. Il montra que le développement de ces formes avait lieu plus fréquemment dans les zones de forte prévalence d'une variation génétique, en l'occurrence, la variation de l'Hémoglobine S du gène Hb β (HBB), causant le trait drépanocytaire, et que les porteurs hétérozygotes développaient souvent des formes moins sévères [9,10]. Durant la même période, les premières descriptions de PIDs associés à un seul agent infectieux tel que l'épidermodysplasie verruciforme (prédisposant au papillomavirus humain) et la maladie lymphoproliférative liée au chromosome X (prédisposant au virus Epstein-Barr) ont été rapportées. Ces observations ont contribué indirectement à la découverte, dans les années 90, aux premières mutations prédisposant à un seul agent infectieux, telles les infections par mycobactérie ou l'encéphalite herpétique [11]. Bien évidemment et cela représente le cœur de cette thèse, la lèpre n'échappe pas à ce constat. Cette prédisposition humaine aux maladies infectieuses recouvre apparemment un spectre continu depuis l'hérédité monogénique (mendélienne) jusqu'à l'hérédité complexe [12].

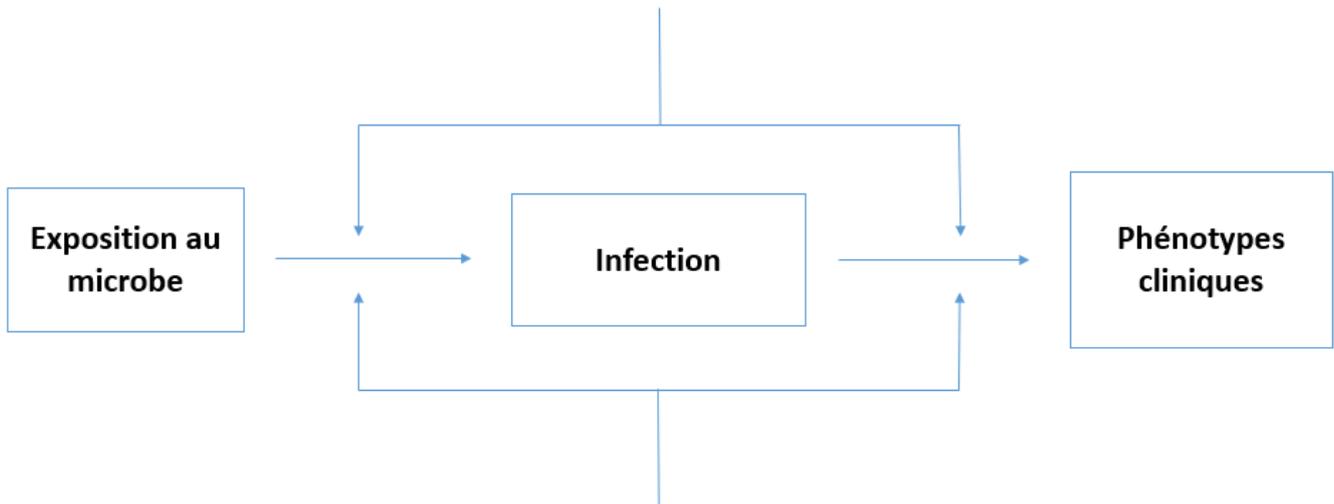
De nombreux déficits immunitaires primaires ont été décrits cliniquement et expliqués génétiquement et dans la grande majorité des cas ils prédisposent les individus atteints à des infections récurrentes par de multiples pathogènes. D'autres traits monogéniques prédisposant spécifiquement à un pathogène chez des individus par ailleurs sains ont été récemment identifiés [13]. Au niveau populationnel, des gènes d'importance particulière (dits gènes majeurs) ont été identifiés dans un nombre limité mais croissant d'infections communes comme la lèpre ou la tuberculose [14–16].

Pour résumé, on peut dire que l'existence d'une composante génétique influençant la réponse de l'hôte aux agents infectieux est maintenant admise sur la base d'observations épidémiologiques et de résultats expérimentaux chez l'animal. Chez l'homme, l'identification de ces facteurs génétiques fait appel à l'épidémiologie génétique. La théorie microbienne des maladies infectieuses a démontré que les microbes sont nécessaires au développement des maladies infectieuses (L. Pasteur, R. Koch). Cependant, ils ne sont pas suffisants, car seule une minorité des individus infectés développent une maladie clinique (C. Nicolle). Un « défaut » génétique est nécessaire pour le développement d'une infection (J-L. Casanova) **(figure 0)**. Cette théorie génétique des maladies infectieuses est, aujourd'hui, bien ancrée dans la pensée scientifique et est en constante évolution grâce à un domaine de recherche : la génétique humaine des maladies infectieuses. Ces implications cliniques sont considérables dans la mesure où la compréhension moléculaire de la pathogénèse d'une maladie infectieuse ouvre la voie au diagnostic génétique et aux interventions immunologiques visant à restaurer physiologiquement une réponse immunitaire partiellement déficiente, à l'image de l'insuline du diabétique. L'étude génétique des maladies infectieuses, vues comme des « expériences de la Nature » fournit également une opportunité unique de définir la fonction des gènes *in Natura*, dans le contexte d'un écosystème naturel par opposition à l'expérimental [7].



Théorie microbienne des maladies infectieuses

« pas d'infection sans microbe »



« pas d'infection sans défaut génétique »

Théorie génétique des maladies infectieuses



Alliance entre environnement et génétique

Lèpre

Figure 0 : De la théorie microbienne à la théorie génétique des maladies infectieuses

« Ce n'est pas le microbe qui est transmis des parents aux enfants, mais la prédisposition à la maladie » Pasteur [17].

Les maladies infectieuses (comme ici la lèpre) peuvent être considérées comme maladies « complexes », pour lesquelles l'objectif est de rechercher des variants génétiques qui ne sont pas individuellement responsables du phénotype, mais qui augmentent chacun d'eux, peu à peu, le risque de développer la maladie ou une forme clinique particulière. Ceci par opposition aux maladies que l'on peut qualifier de « mendéliennes » pour lesquelles un variant spécifique confère la maladie, selon un mode d'hérédité et de pénétrance définis (**tableau 0 et figure 0bis**). Il existe un continuum entre ces deux formes d'hérédité, puisqu'un variant d'une maladie complexe avec un odds-ratio élevé va se rapprocher d'un variant « d'hérédité mendélienne » à pénétrance faible [18].

Historiquement, les approches méthodologiques pour ces deux types d'hérédité étaient bien distinctes : l'épidémiologie génétique pour la première et la génétique moléculaire pour la seconde. L'interaction entre ces deux disciplines pouvait être féconde, sous un angle séquentiel : des régions de liaison génétique étaient identifiées au sein de familles avant que les gènes de cette région ne soient étudiés à la recherche de la mutation causale. L'avènement des techniques de séquençage dites de « nouvelle génération » permet une synchronisation de ces deux disciplines. En effet, leur matériel génétique, les séquences d'exome et du génome entier, est maintenant commun.

<i>Caractéristiques</i>	<i>Génétique Mendélienne</i>	<i>Génétique Épidémiologique</i>
<i>Phénotype</i>	Rare (Formes très sévères)	Commun (Atteint/non atteint)
<i>Causalité</i>	Monogénique	Complexe
<i>Type de mutation</i>	Mutation rare	Polymorphismes communs
<i>Effet de la mutation</i>	Fort	Modeste

Tableau 0 : Caractéristiques de la génétique mendélienne et épidémiologique

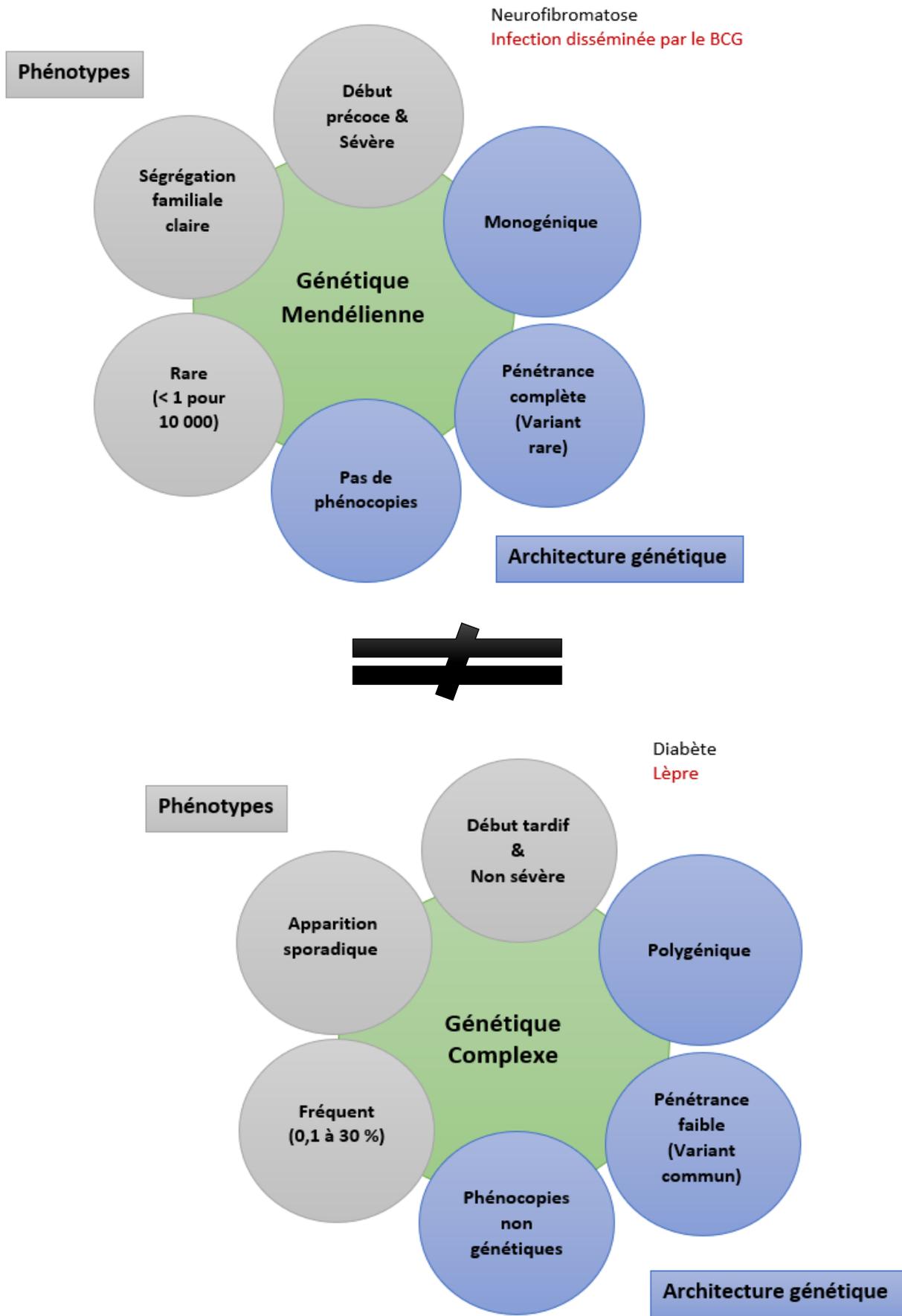


Figure Obis : Caractéristiques de la génétique mendélienne et épidémiologique

Chapitre 1 : La Lèpre

1.1 Un peu d'histoire...

Un des fléaux de tout temps le plus redouté de l'humanité, attribué à une malédiction divine, la lèpre, longtemps incurable et très mutilante, a trop souvent suscité des mesures inhumaines envers les malades. La seule évocation de la lèpre suscite invariablement des réactions de répulsion, de dégoût et/ou de pitié. De nos jours, malgré des progrès incontestables réalisés sur ce plan, la lèpre véhicule encore, dans de nombreux pays, une image péjorative.

C'est très récemment (au cours de la Seconde Guerre mondiale) que des traitements efficaces ont transformé la lèpre en maladie curable. Par ses effets physiques, comme par les drames psychologiques qu'elle provoque, la lèpre crée un important problème social, justifiant la poursuite des efforts qui, au cours des dernières décennies, ont permis de réduire la prévalence de la maladie dans de nombreux pays [19].

1.1.1 Hypothèse historique : origine et diffusion de la lèpre

L'origine de la lèpre se perd dans la nuit des temps. À la lecture des textes anciens, indiens et chinois, et des livres sacrés, telle la Bible, certaines maladies indiquées comme redoutables et infamantes semblent correspondre à la lèpre, mais l'imprécision des signes évoqués ne permet pas d'en être certain. Ainsi, l'origine géographique de la lèpre reste incertaine mais plusieurs éléments laissent penser que l'Inde fut, sinon le lieu d'origine, du moins un des premiers foyers du fléau. Documentée depuis l'antiquité dans plusieurs civilisations, la lèpre est mentionnée dès le quatorzième siècle avant Jésus-Christ dans les écrits sacrés de Vedas sous le nom de « Kushtha », mot sanscrit signifiant « manger autre part » probablement en raison de l'exclusion dont étaient déjà victimes les malades. Le fameux médecin indien Sushruta (IV^{ème} siècle apr. J.-C.), dans son livre « Sushruta Samhita » rapportant les pratiques médicales des anciens hindous, mentionne des écrits du VII^{ème} siècle av. J.-C. décrivant les différentes formes de lésions cutanées ainsi que les atteintes

neurologiques périphériques et allant jusqu'à rassembler et structurer ces manifestations au sein d'une même entité médicale.

Ainsi, l'Inde a longtemps été considérée comme le berceau originel de la lèpre et aujourd'hui encore, la lèpre se dit « Kushtha Roj » en indien [1,20]. Depuis l'Inde, la lèpre se serait répandue à l'est vers l'Indochine, l'Insulinde, la Chine (ou des traités médicaux décrivent une maladie, lei-fon, ayant de nombreuses ressemblances avec la lèpre dès le V^{ème} siècle av. J.-C.) et le Japon. À l'Ouest, aucune description clinique de lèpre n'est rapportée avant le III^{ème} siècle av. J.-C., en particulier Hippocrate (460-377 av. J.-C.) n'en fait aucune mention dans ses écrits, et il faudra attendre le « Canon de la Médecine » d'Avicenne (980-1037) pour une présentation détaillée de la maladie. Transportée par les guerriers de Darius et d'Alexandre, elle gagne la Perse, le Proche Orient et l'Égypte (l'examen de momies égyptiennes indique que la lèpre existait dans ce pays dès le II^{ème} siècle av. J.-C.). Les Phéniciens contribuent à son extension à tout le littoral méditerranéen, et au début de l'ère chrétienne les légions romaines la font pénétrer au cœur de l'Occident [1,20].

Les navigateurs, les invasions barbares et sarrasines favorisent sa dissémination à toute l'Europe et jusqu'en Islande. A la suite des Croisades, enfin, elle atteint son apogée en Europe au XII^{ème} et XIII^{ème} siècles, elle inspire à cette époque une terreur telle que les lépreux sont frappés de mort civile et rejetés de la communauté humaine [21]. En Europe de l'Ouest, la lèpre ne fut endémique que durant la période médiévale. Les raisons de son déclin, antérieur à l'utilisation du traitement antibiotique, ne sont pas clairement établies. L'efficacité des principes de prévention ou en d'autres termes l'efficacité de cette impitoyable ségrégation, les épidémies concomitantes d'infections rapidement létales (peste, variole, choléra), et l'amélioration des conditions d'habitation (logements plus salubres, eau courante, meilleure nutrition) sont probablement conjointement responsables de cette évolution [22]. Une autre hypothèse serait l'acquisition d'une immunité croisée au contact du bacille tuberculeux [23]. Néanmoins, quelques foyers ont persisté en plusieurs régions jusqu'au début du XX^{ème} siècle, notamment la Scandinavie et le pourtour méditerranéen, ce qui explique probablement qu'un Norvégien, Hansen, ait pu, au XIX^{ème} siècle, découvrir le bacille, agent pathogène de l'infection, classant celle-ci parmi les maladies infectieuses et contagieuses [1,20].

Depuis l'Europe, la lèpre a vraisemblablement été exportée vers le continent américain par les explorateurs portugais et espagnols d'abord, puis par la traite des esclaves d'Afrique ensuite. C'est finalement au cours du XIX^{ème} siècle que la lèpre fut identifiée comme une « maladie tropicale contagieuse ». Tropicale en raison de sa répartition prépondérante dans les territoires coloniaux européens, et contagieuse grâce aux travaux du médecin danois Daniel Danielsson sur la lèpre lépromateuse qui est, de fait, la forme contagieuse de la maladie. L'identification de l'organisme causal, *M. leprae*, sur des lépromes de patients Norvégiens par Armauer Hansen en 1873 est sans doute la plus grande découverte à nos jours quant à l'étiologie d'une pathologie puisque pour la première fois (avec l'anthrax) un microbe était identifié comme l'agent causal d'une maladie (quelques années avant que Koch ne décrive le bacille de la tuberculose) [24].

1.1.2 Hypothèse « génomique » : origine et diffusion de la lèpre

Jusqu'en 2005, la connaissance historique de la lèpre était, comme décrit dans le paragraphe précédent, purement bibliographique et archéologique. Cette vision a été remise en question par l'étude de polymorphismes du génome de *M. leprae* (technique de génomique comparative) qui, en comparant différentes souches existantes dans le monde, a permis de reconstruire l'histoire naturelle du pathogène et de considérer deux scénarios plausibles sur l'origine géographique de la lèpre dans le monde : elle pourrait effectivement être originaire d'Asie, mais il semble cependant plus probable que son point de départ se situe en Afrique de l'Est [25]. Ainsi, ils ont comparé les séquences de 400 échantillons de *M. leprae* prélevés sur des malades et des restes anciens, provenant de 28 pays. Les polymorphismes identifiés (des SNPs ou polymorphisme d'un seul nucléotide) leur ont permis de classer les échantillons en 4 types et 16 sous-types avant de déterminer les génotypes ancestraux les plus probables. Chaque génotype présentant une forte corrélation avec l'origine géographique de l'échantillon, ils ont donc pu proposer une hypothèse phylogéographique concernant le patron de dissémination du bacille.

« En tout état de cause, il est maintenant clairement établi que la pandémie résulte de la dissémination d'une souche unique qui ne s'est guère modifiée au cours des siècles » Pr. Cole

Tout d'abord, et de façon inattendue, la maladie aurait plutôt été introduite en Afrique de l'Ouest par des explorateurs, des commerçants ou des colons porteurs de l'infection provenant d'Afrique du Nord ou d'Europe, plutôt que par des migrants originaires d'Afrique de l'Est. En effet, le fait que le génome de *M. leprae* présent en Afrique du Nord soit plus proche de celui de *M. leprae* européen que du génome de *M. leprae* de l'Afrique de l'Est tend à montrer cette hypothèse. D'Afrique de l'Ouest, la lèpre aurait ensuite été introduite par le commerce des esclaves au XVIII^{ème} siècle dans les îles Caraïbes, au Brésil et vraisemblablement dans les autres régions d'Amérique du Sud ainsi qu'en Amérique du Nord par le colonialisme, des isolats au même profil que ceux d'Afrique de l'Ouest ayant été découverts dans ces régions. Ainsi, le génome de la souche de *M. leprae* majoritairement responsable de la maladie sur le continent américain est plus proche du type présent en Europe et en Afrique du Nord, ce qui indique que le colonialisme et l'émigration depuis le vieux continent ont très certainement contribué à l'introduction de la lèpre dans le Nouveau Monde [25,26] (**figure 1**).

1.1.3 Stigmatisation des patients atteints de la lèpre

Les stigmates endurés par les patients atteints de lèpre sont terribles. En effet, la lèpre a donné lieu à des mesures de ségrégation et d'exclusion sociale, quelquefois héréditaires, comme dans le cas des Cagots du sud-ouest de la France. Au Moyen-Âge, les malades sont identifiés par les prêtres ou d'autres autorités spirituelles, et séparés de la communauté, le plus souvent de façon ritualiste. Considérés comme « mort pour la société », des funérailles factices étaient prononcées en présence des malades, ceux-ci se tenant symboliquement dans une tombe. Ils étaient ensuite conduits sur un terrain vague et leur accès à la ville ou au village sévèrement limité. Les lois de l'époque illustrent la peur irrationnelle que générait la maladie. Ils doivent porter une cape grise ou noire ornée d'une croix jaune, « le manteau du lépreux », ainsi qu'un signe distinctif mentionnant leur « misérable présence » aux autres individus (une cloche, une crécelle, une corne) ; ils ne peuvent approcher les enfants ; ils ne peuvent ni boire ni manger en présence d'individus non atteints ; ils ne peuvent parler à un adulte que face au vent ; ils ne peuvent se laver les mains à une source ou un ruisseau ; ils doivent pointer avec une perche ce qu'ils désirent acheter et mendier en mettant un sac au bout de cette perche ; les églises, les tavernes, les maisons leur sont interdites...

Il existe alors, pour recueillir les lépreux, environ dix-neuf mille léproseries dans toute la chrétienté dont près de deux mille en France [1,21] (**figure 2**). La plupart des prêtres et des savants de cette époque pensaient que la lèpre était la maladie *tsara'ath* de l'Ancien Testament [qui devient *lepra* (du grec *lepros*, « qui pèle ») lorsque l'Ancien Testament fut traduit en grec environ un siècle avant J.-C., terme que Jérôme conserva en 405 dans le Latin Vulgate]. C'est ce malentendu qui fut, au moins partiellement, la cause d'une telle stigmatisation dans les cultures occidentales [27]. En effet, le *tsara'ath* de l'Ancien Testament est associé à l'impureté morale, une punition de Dieu pour avoir péché dont les victimes doivent être isolées des autres (**figure 3**). Cependant, la description de la maladie dans le Lévitique 13 et 14, qui associe une tache blanche sur la peau et un blanchissement des cheveux, ne mentionne aucun des traits caractéristiques de la maladie. Autre élément troublant, si l'atteinte cutanée recouvre l'ensemble du corps, le sujet est alors considéré comme « propre », c'est-à-dire non contagieux, ce qui n'est guère en faveur d'une atteinte lépreuse. Enfin, on peut même y lire que la maladie s'attaque aux objets, aux vêtements, aux maisons... Il s'agit donc probablement d'un ensemble de maladies contagieuses de la peau, mais aussi de moisissures ou de lichens. Toutefois, l'association de la lèpre et du péché n'est pas spécifique à la culture occidentale. L'attitude chinoise et hindoue était très semblable, bien que pour des raisons évidemment différentes. Les musulmans ont été prétendus plus tolérants face aux patients atteints de lèpre, mais c'est oublier le massacre par les Sarrasins, en 1253, de tous les malades de l'hospice de l'Ordre Hospitalier Saint Lazare de Jérusalem [1,20,21].

En raison de cette terrible stigmatisation dont ont été victimes les malades de la lèpre, il n'est pas surprenant que nombreux soient ceux qui préfèrent l'utilisation du terme éponyme « Maladie de Hansen » à la place de « lèpre ». Déjà en 1948, lors du Cinquième Congrès International sur la Lèpre, une résolution avait été adoptée bannissant l'usage du mot lépreux pour désigner un malade de la lèpre.

1.1.4 La paléomicrobiologie

« Une telle étude aurait été impossible il y a encore quelques années, la paléomicrobiologie doit son essor aux développements technologiques réalisés ces dernières années en biologie moléculaire. » Pr Cole [28].

En 2013, une étude a permis de séquencer entièrement le génome de bactéries retrouvées dans des ossements humains datant du Moyen-Âge. Les sources bibliographiques ont conduit à explorer plusieurs cimetières d'Europe du Nord, au Royaume-Uni, au Danemark et en Suède sur des sites connus pour avoir hébergé des communautés importantes de lépreux lorsque la maladie était encore endémique en Europe, ils ont pu exhumer différents restes humains (dents, crânes, os) qui se sont avérés porteurs de *M. leprae*.

Le premier défi technique a été d'extraire l'ADN contenu dans les ossements, datant du XI^{ème} siècle après J.-C. pour les plus vieux. Il a ensuite fallu séparer l'ADN des bactéries de celui de leur hôte. Une gageure quand on sait que certains échantillons ne contenaient pas plus de 0,1% d'ADN bactérien. Néanmoins, le séquençage des fragments d'ADN ainsi isolés a permis de mettre en évidence cinq souches distinctes de bactéries. Ils ont ainsi pu comparer le génome de ces bactéries anciennes à celui des souches de la lèpre contemporaine. De manière inattendue, le génome de *M. leprae* n'a quasiment pas évolué en plus de neuf siècles. Ceci alors que la plupart des agents pathogènes sont connus pour muter très fréquemment, ce qui leur permet de faire face à des modifications de leur environnement. Ces résultats indiqueraient donc que la lèpre n'a pas disparu d'Europe à cause d'une mutation génétique qui aurait rendu la bactérie moins virulente (hypothèse jusque-là souvent avancée). La maladie aurait été éradiquée grâce à une meilleure résistance de ses hôtes [29].

« Nos résultats plaident en faveur d'une certaine sélection naturelle qui se serait opérée au fil des siècles et aurait favorisé ceux dont le système immunitaire permettait de résister à la maladie. Ceci est appuyé par des études antérieures qui ont révélé qu'aujourd'hui, les Européens sont plus résistants à la lèpre que les autres peuples. » Pr Cole [28].

Cette résistance et où prédisposition serait donc liée à la présence de certaines mutations dans des zones spécifiques de notre génome, c'est ce que nous allons étayer tout au long de ce manuscrit.

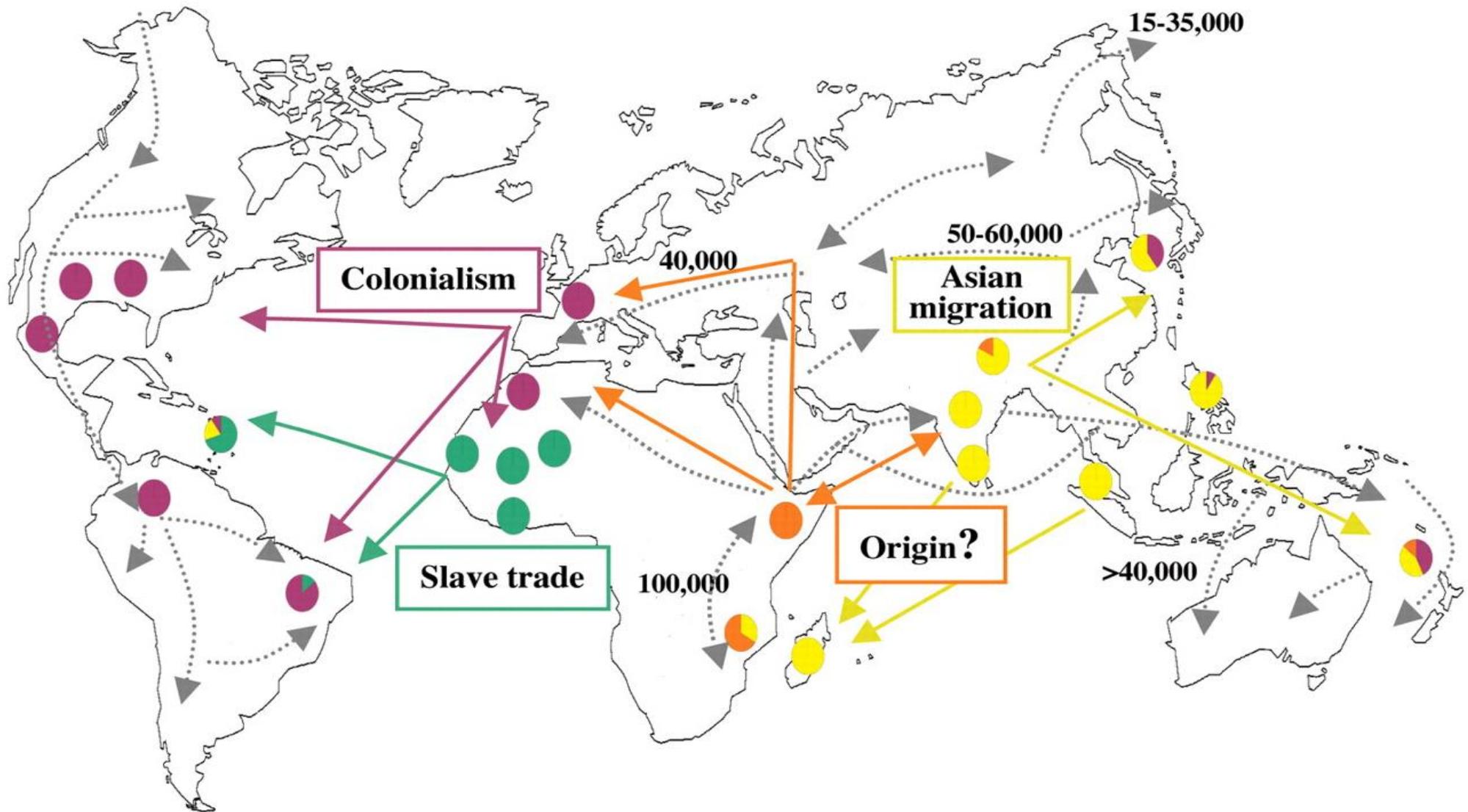
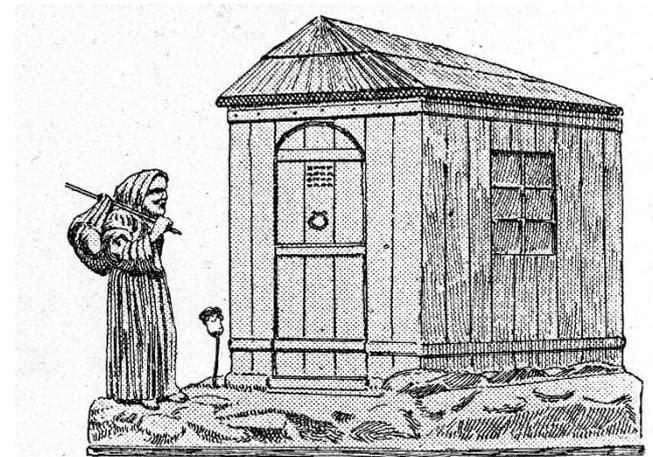


Figure 1 : Dissémination de la lèpre dans le monde en fonction du génome du bacille [25]

Les cercles indiquent les foyers endémiques au cours de l'histoire, les pays d'origine des souches examinées et leurs distributions à travers les 4 SNPs-types. Chaque couleur représente un type de bacille (le SNP type 1 apparaît en jaune, le 2 en orange, le 3 en violet et le 4 en vert). Les flèches de couleur indiquent la direction des migrations humaines prédite ou déduite par l'analyse de ces 4 SNPs. Les flèches grises montrent les migrations humaines dérivées d'études génétiques, archéologiques et anthropologiques (temps de migration estimé en année).



Peintures représentant des « Lépreux » sonnant leur cloche (ou crécelles ou cliquettes) pour avertir les passants pour ne pas être approchés, (manuscrit latin du XIV^{ème} siècle).



Groupe de « lépreux » arabes durant la domination ottomane, près de la tombe de Myriam, située hors des murs de la vieille ville de Jérusalem, 1910.

Retraite des « lépreux » en fonte, XV^{ème} siècle.

Figure 2 : Exclusion des lépreux [30]



Jésus guérit un « lépreux », mosaïque médiévale, cathédrale de Monreale (Sicile), XII-XIII^{ème} siècles.



Victimes de la lèpre recevant la parole d'un évêque. Omne Bonum, de James le Palmer, Londres, 1360-1375.

Figure 3 : Lèpre et Église [30]

1.2 Histoire de *Mycobacterium leprae*

La lèpre est, avec l'anthrax, l'une des toutes premières maladies infectieuses pour laquelle l'agent infectieux causal a été identifié, par Hansen, en 1873 en Norvège, quelques années avant que Koch ne décrive le bacille de la tuberculose [24].

1.2.1 Taxonomie

Le bacille de Hansen est un schisomycète [31] :

- Classe : Actinomycetales
- Ordre : Mycobacteriales
- Famille : Mycobateries
- Genre : Mycobacterium

1.2.2 Morphologie et structure

➤ **Microscopie optique**

M. leprae est un bacille intracellulaire, micro-aérophile et se colore uniformément en rouge vif avec la coloration de Ziehl-Neelsen. Cette méthode consiste à colorer le bacille avec de la fuschine dans un premier temps puis à le soumettre à l'action de l'acide et de l'alcool ; si le bacille résiste et qu'il conserve la coloration, on dit qu'il est alcool-acido-résistant (BAAR). C'est le cas pour *M. leprae* qui apparaît sous la forme d'un bâtonnet rouge, non mobile, rectiligne, aux extrémités arrondies, long de 1 à 8 µm sur 0,3 à 0,5 µm de large (**figure 4.A**).

Une des particularités de *M. leprae* est qu'il s'assemble « en paquets de cigares » dans les lésions lépromateuses : ils se regroupent en amas arrondis de 50 à 100 µm de diamètre que l'on appelle globi. Intracellulaire, le bacille pénètre dans un macrophage, il s'y multiplie puis envahit la cellule en repoussant le noyau et le cytoplasme sur la membrane cellulaire et provoque, ainsi, la dégénérescence de la cellule [31] (**figure 4.B**).

Lorsque le bacille est en dégénérescence, c'est le cas chez les patients en cours de traitement, il ne se colore pas uniformément par la méthode de Ziehl-Neelsen et présente un aspect granuleux [32].

➤ Microscopie électronique

M. leprae a un aspect mousseux car il est entouré d'une zone constituée de mycosides, transparente aux électrons et qui ressemble à une capsule. La paroi est composée de deux couches principales : une interne (opaque aux électrons correspondant au peptidoglycane), une externe (transparente aux électrons correspondant au polysaccharide). La membrane cytoplasmique est constituée de deux couches de 3 µm d'épaisseur séparées par un espace de faible densité, elle est donc symétrique. Enfin, le cytoplasme de *M. leprae* peut parfois renfermer des inclusions paracrystallines (particules de 8 nm de diamètre), particularité propre à son espèce [32] (**figure 4.C**).

1.2.3 Constitution chimique et antigénique de la paroi

La nature chimique de la paroi de *M. leprae* est quelques peu différente de celle des autres mycobactéries. De ce fait, son mur cellulaire, est composé depuis la membrane cellulaire vers l'extérieur (**figure 5**) :

- De peptidoglycane contenant bien le N-acétylglucosamine mais dans les chaînes tétra-peptidiques latérales, le 1^{er} acide aminé n'est pas le L-alanine mais la glycine,
- Autour se trouve un réseau d'arabino-galactane auquel sont rattachés les acides mycoliques, mycosérosoïques à longue chaîne de carbone liés de façon covalente,
- Dans la zone appelée « capsule » (couche externe), il existe de nombreux lipides et en particulier des glycolipides phénoliques (PGL).

Le lipide qui donne à *M. leprae* sa spécificité immunologique est le PGL-1, contre lequel se développe une réponse humorale détectable par sérologie. D'autre part, les gènes codants pour la formation de cette paroi cellulaire ont été conservés [33,32].

1.2.4 Identification de *M. leprae*

L'identification de *M. leprae* comprend plusieurs critères [32] que nous détaillerons dans la suite de ce manuscrit : néanmoins, il s'agit principalement de l'étude des acides mycoliques, la présence de PGL-1 (spécifique), le séquençage spécifique de l'ARN ribosomal, l'essai de multiplication sur le coussinet plantaire de la souris.

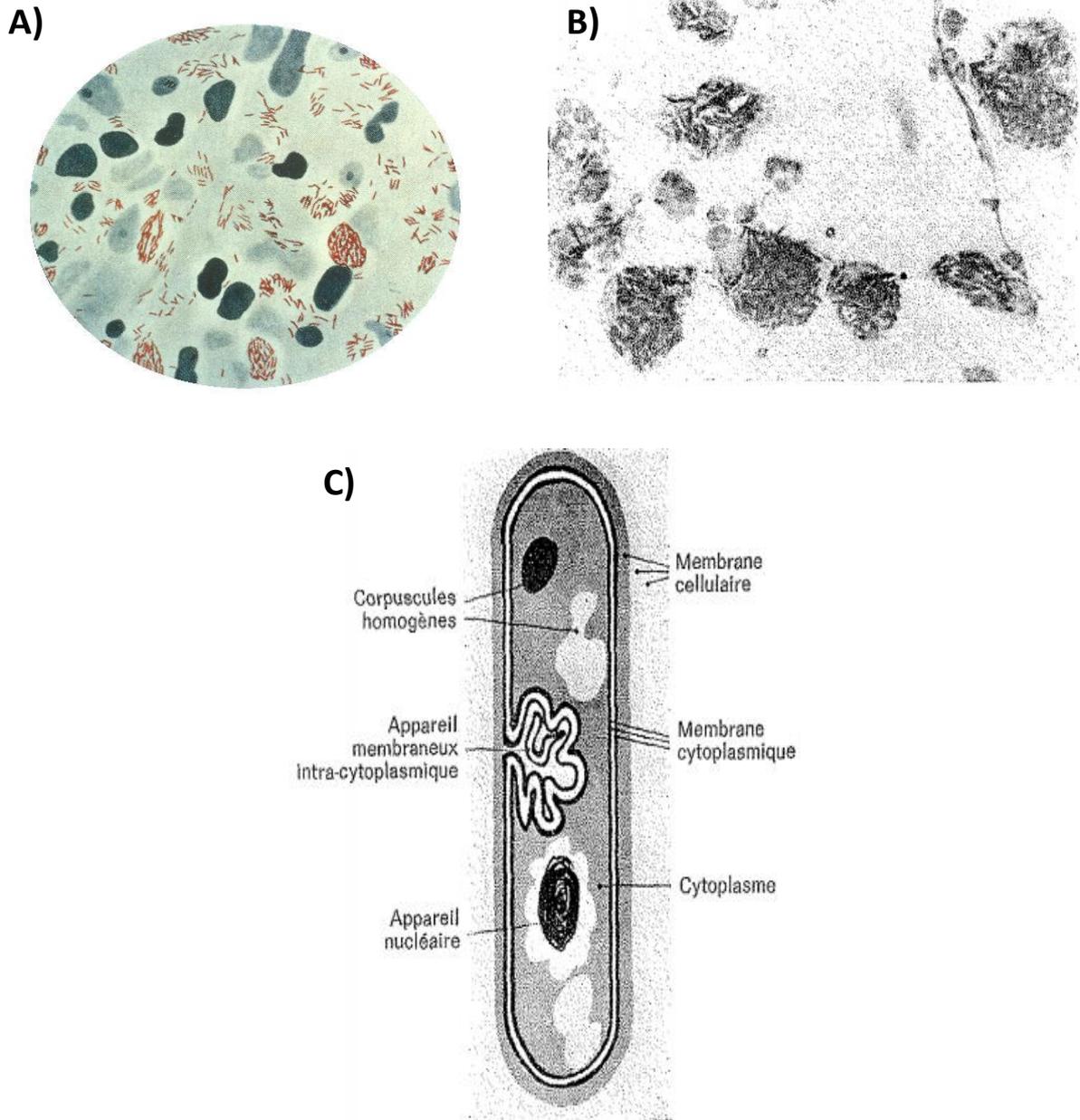


Figure 4 : Observations de *M. leprae*

- A) Microscopie optique montrant les bacilles en forme de bâtonnets rouges après la coloration de Ziehl-Neelsen [34].
- B) Amas de bacilles : « des globis » (prélèvement issu de la muqueuse nasale) [31].
- C) Schéma de *M. leprae* au microscope électronique [32].

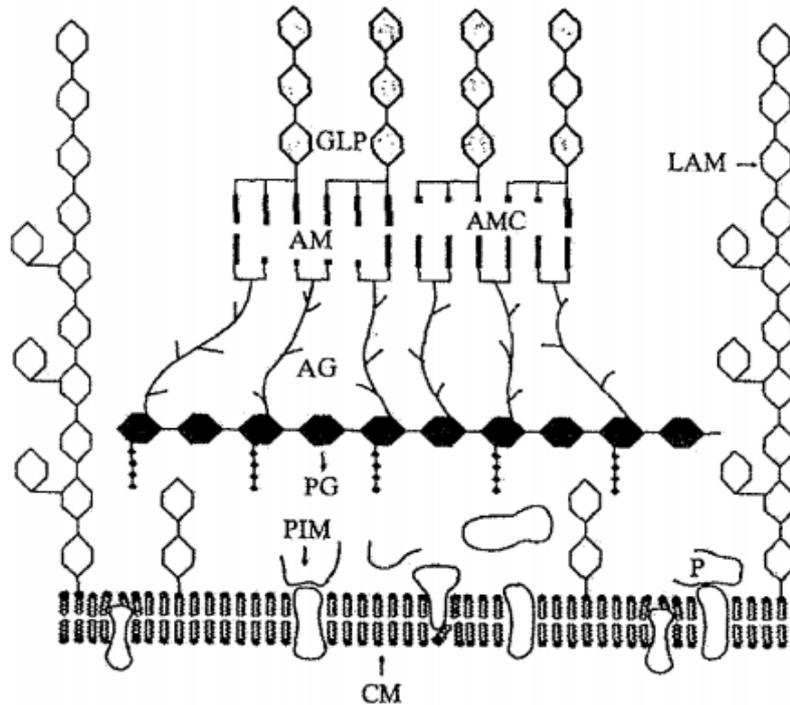


Figure 5 : Représentation schématique de l'enveloppe de *M. leprae* [33]

AG : arabino-galactane ; AM : acides mycoliques ; AMC : acides mycosérosiques ; CM : membrane cytoplasmique ; PGL : glycolipides phénolés ; LAM : lipoarabinomannane ; P : protéine ; PG : peptidoglycane ; PIM : phosphatidyl-inositole mannoside.

1.2.5 Réservoirs naturels

L'homme est son principal réservoir et le tatou à 9 bandes a d'abord été identifié comme un hôte naturel du bacille sur le continent américain [35]. D'autre part, en 2016 en Grande-Bretagne et plus précisément sur l'île de Brownsea, il a été montré que les écureuils roux représentent également un réservoir naturel de *M. leprae* [36] (figure 6).

De manière plus anecdotique, des infections spontanées ont été observées également chez un chimpanzé et deux singes mangabeys (*Cercocebus atys*), l'un d'entre eux en situation d'immunodépression acquise, importés d'Afrique aux USA dans les années soixante-dix et chez lesquels la maladie se développa plusieurs années après leur arrivée [37–39]. De même, fin des années 90 a été décrit un singe macaque (*Macaca fascicularis*) importé également vers les USA en provenance des Philippines, ayant développé une lèpre diagnostiquée par la technique de PCR (Polymerase Chain Reaction) et une sérologie ELISA contre PGL-I et LAM (protéines membranaires du bacille, PGL, glycolipides phénolés ; LAM, lipoarabinomannane). Il est en effet important d'affirmer que *M. leprae* est bien l'agent infectieux animal par des techniques plus fiables que les colorations histologiques dites à tort spécifiques, d'autant plus que *M. leprae* n'est toujours pas cultivable *in vitro* [40]. Il faut signaler ainsi que la lèpre murine naturelle (très proche de l'infection expérimentale par *M. leprae*) est due à *M. leprae murium* [41]. La présence de ces réservoirs pourrait expliquer pourquoi, malgré les traitements efficaces actuels censés diminuer les risques de transmission interhumaines, on dénombre encore plus de 220 000 cas touchés chaque année dans le monde.

« La lèpre n'est plus un problème de santé publique au Royaume-Uni depuis le XVII^{ème} siècle, mais le fait de retrouver cette bactérie des centaines d'années plus tard chez l'écureuil roux nous montre que même lorsque l'on élimine le réservoir humain, il peut en subsister d'autres dans l'environnement. Cela pourrait être suffisant pour expliquer un certain pourcentage de nouveaux cas et compliquer la maîtrise de la maladie. » Pr Cole [42]

L'hypothèse d'un réservoir inerte a également été proposée dès 1910 [43] et reste encore débattue aujourd'hui [44]. Si les premières études cherchant à identifier *M. leprae* dans le milieu naturel étaient basées sur des techniques de mise en évidence pour le moins indirectes (identification de bacilles non cultivables, coloration de Ziehl-Neelsen), le développement de la PCR a donné lieu à plusieurs campagnes d'analyses du sol et de l'eau

dans des régions endémiques en Inde et Indonésie. Ainsi, des résultats de PCR positifs concernant 25% des 90 puits échantillonnés à Java ont été obtenus [45]. Sur cette même île, il a été dénombré 21 points d'eau sur 44 qui contiendraient l'ADN du bacille et une différence significative de prévalence de la maladie entre les groupes d'individus utilisant les puits « PCR positifs » et ceux utilisant les puits « PCR négatifs » a été observée [46]. Par ailleurs, des résultats similaires avec des échantillons de terre prélevés en Inde ont été constatés. Ainsi, 50% des 40 échantillons collectés dans des zones où habitaient des malades présentaient une amplification de l'ARN du bacille contre un pourcentage significativement inférieur dans les échantillons (15% de 40) provenant de zones sans malades recensés [47]. Enfin, une équipe américaine a réussi à infecter, en laboratoire, des amibes présentes dans l'eau : la bactérie ne s'y multiplie pas, mais elle ne meurt pas restant de ce fait potentiellement infectieuse. Cette observation nous laisse penser que cela pourrait également se produire dans la nature [48].

Cependant, la seule utilisation de la PCR, en particulier lorsque la cible de l'amplification est un unique segment du génome, est discutable pour rechercher *M. leprae* [44]. Outre les problèmes de contamination, l'amplification d'une seule portion d'ADN bactérien ne permet pas de conclure à l'identification d'une espèce sans un risque important de faux positifs, dans la mesure où une grande partie de la diversité, notamment génétique, du monde bactérien nous est encore inconnue. Même en s'affranchissant intellectuellement de cette limite méthodologique, ces études soulèvent d'autres questions.

Ainsi, il faut souligner que la présence dans le milieu de l'ADN du bacille ne signifie pas que ce dernier soit vivant, infectieux ou encore capable de s'y multiplier. De plus, les différences significatives observées entre les échantillons supposés « en contact » et « sans contact » avec les malades peuvent refléter une contamination humaine de l'environnement plutôt que l'inverse. Concernant la survie du bacille dans l'environnement, il a été rapporté que *M. leprae* reste capable de se multiplier dans des coussinets de souris après avoir passé jusqu'à 46 jours dans de la terre humide [49].

Finalement, à notre connaissance, aucune étude ne fait état d'une multiplication de *M. leprae* dans le milieu extérieur, élément indispensable à l'établissement d'un réservoir pérenne et véritable du point de vue épidémiologique. L'idée d'une multiplication du bacille dans un milieu extracellulaire semble d'ailleurs de plus en plus improbable au vu de l'analyse du génome de *M. leprae*.

1.2.6 Croissance de *M. leprae*

En Afrique, on dit souvent que la lèpre « lèche la peau et mord le nerf ». En effet, le bacille de Hansen agit selon un double tropisme : cutané et neurologique. Ce bacille a un tropisme particulier pour la peau au niveau de laquelle il infecte les macrophages dermiques et les nerfs où il se multiplie dans les cellules de Schwann (gaine protectrice entourant les neurones) [50]. Sa première manifestation se traduit ainsi par des lésions dermatologiques, sous forme de dépigmentation légère, indolore.

Sa croissance est la plus lente du monde bactérien avec un temps de duplication de 12 à 20 jours et optimale pour des températures plus faibles que les autres bactéries pathogènes chez l'homme, entre 32 et 35°C et un pH acide de 5,6 [51].

1.2.7 Modèles d'étude

La culture de *M. leprae* n'a jamais pu être réalisée *in vitro* (en milieu artificiel), le bacille ne peut croître que sur des cellules vivantes. Néanmoins, *M. leprae* peut être maintenu à 33°C dans un état métabolique stable durant quelques semaines [52]. De ce fait, sa croissance n'est possible qu'*in vivo*, dans son réservoir naturel (l'homme, et dans une moindre mesure le tatou) et chez quelques modèles expérimentaux d'animaux (**figure 7**) :

- Le coussinet plantaire de souris : modèle privilégié depuis 1960 [53],
- Le tatou à neuf bandes qui est un hôte naturel [54,55],
- Les souris athymiques nude [56] ou SCID (severe combined immunodeficient) [57] et le rat Lewis [58],
- Dans une moindre mesure le singe mangabey (vu précédemment) [59].



Figure 6 : Hôtes de *M. leprae*, le tatou à 9 bandes et l'écureuil roux
[42,60]

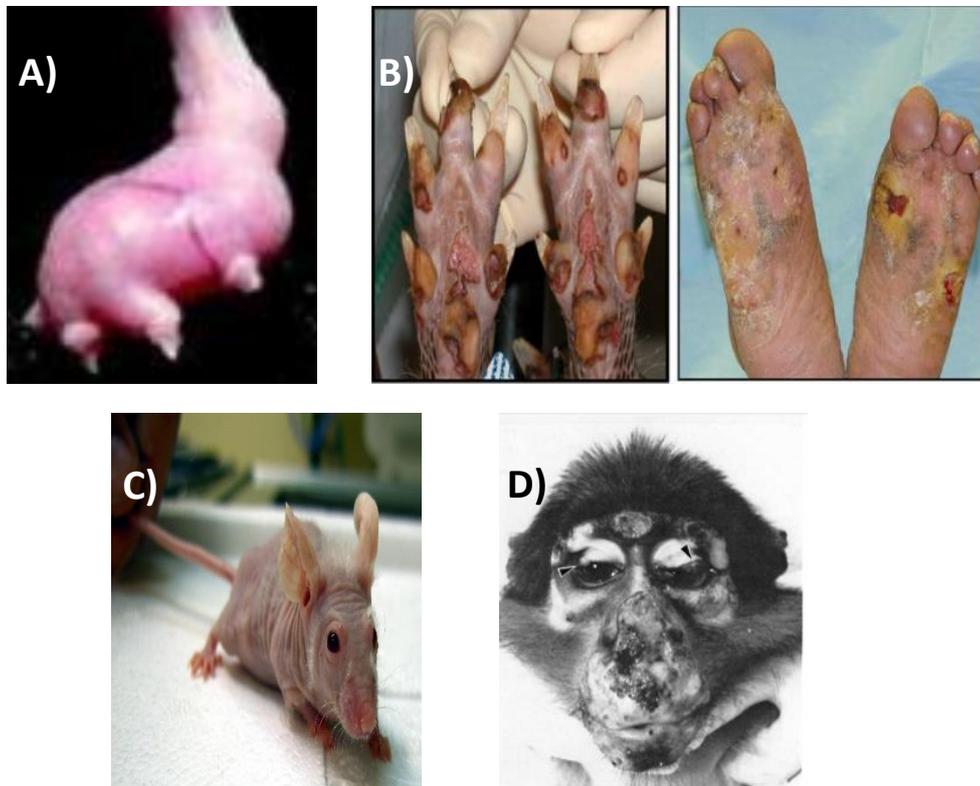


Figure 7 : Modèles d'étude de *M. leprae*

- A) Photo d'un coussinet plantaire de souris inoculé par *M. leprae* (modèle privilégié depuis 1960) [61].
- B) Photos des pieds de tatou à 9 bandes (à gauche) et d'humain (à droite) montrant des lésions à un stade avancé de la maladie [55].
- C) Photo d'une souris athymique nude immunodéficente [62].
- D) Photo de visage d'un singe mangabey naturellement atteint de lèpre [39].

1.2.8 Génome de *M. leprae*

➤ Description du génome de *M. leprae*

Le génome de *M. leprae* a entièrement été séquencé en 2001 (**figure 8**), 3 ans après celui de *M. tuberculosis* [63]. L'étude génomique comparative de *M. leprae* avec *M. tuberculosis* a permis de montrer que ce dernier représentait un cas extrême d'évolution réductive. En effet, le génome de *M. leprae* est nettement plus petit que celui de *M. tuberculosis* (3,3 vs. 4,4 mégabases) avec une réduction majeure du nombre de gènes potentiellement codants (1 614 vs. 3 993) et un nombre très élevé de pseudogènes (1 133 vs. 6). Au total moins de 50% de son génome code pour des gènes fonctionnels (vs 90% chez *M. tuberculosis*) (**figure 9**) [64]. En conséquence, plusieurs voies métaboliques ont disparu au cours de l'évolution : ainsi, *M. leprae* possède des enzymes des cycles de la glycolyse, du pentose phosphate et de l'acide tricarboxilique, de la synthèse de pyrimidines mais dépend de son hôte pour les purines [32]. Cette réduction évolutive par la perte de voies métaboliques entières (catabolisme des acides gras et des polykétides, synthèse de l'enveloppe, transport de certains métabolites et défense contre les radicaux libres) explique son caractère strictement intracellulaire, son parasitisme quasi exclusivement humain, sa niche cellulaire très limitée, son incapacité à être cultivé en milieu artificiel et sa croissance lente. Historiquement, on a constaté le terrible isolement auquel étaient réduits les malades de la lèpre. De façon assez ironique, il semble que le bacille causal soit lui-même destiné à subir un tel sort.

La variabilité génomique de *M. leprae* est négligeable à l'échelle du globe : le bacille infectant est quasi identique quelle que soit la région du globe où il a été caractérisé. En effet, le séquençage d'échantillons provenant d'Inde, de Thaïlande, d'Amérique du Nord et du Brésil a permis de déterminer que la diversité génétique du bacille semble très faible puisque les échantillons partagent une identité de séquence de 99,995% [65]. Cette quasi-clonalité du bacille a permis de montrer que *M. leprae* est une espèce distincte, et est en particulier éloignée de *M. lepraemurium* qui infecte naturellement les souris [66]. Elle a également contribué à identifier, par contraste, une nouvelle espèce de mycobactérie qui pourrait être impliquée dans certaines formes de la maladie.

➤ *M. leprae* versus *M. lepromatosis*

Dans ce sens, en 2008, il a été décrit au Mexique une nouvelle espèce chez deux patients atteints d'une lèpre diffuse (ou de Lucio), nommée *M. lepromatosis* : le bacille diffère de *M. leprae* de 2% pour le gène codant l'ARN ribosomal 16S, et de 6,9% pour 14 autres gènes codants [67,68]. Une équipe indépendante a confirmé en 2012 l'incrimination probable de cette mycobactérie chez une femme atteinte de lèpre diffuse également au Mexique : l'histologie révélait des BAAR, la PCR était négative pour *M. leprae* mais positive pour *M. lepromatosis* [69]. De même et toujours au Mexique, deux cas de lèpre chez des frères et sœurs causés par *M. lepromatosis* a été confirmé. C'est le premier cas familial mais cette fois-ci les lésions suggéraient une forme borderline lépromateuse (et non de Lucio), la PCR était négative pour *M. leprae* mais positive pour *M. lepromatosis* [70]. D'ailleurs, concernant la pathogénie de cette nouvelle espèce, il a tout d'abord été soutenu qu'elle était spécifiquement responsable de la lèpre de Lucio, une forme sévère de lèpre retrouvée majoritairement au Mexique et aux Caraïbes [71]. Cependant et comme vu précédemment, les premières descriptions de *M. lepromatosis* étaient effectivement retrouvées exclusivement chez des patients de cette zone géographique et présentant cette forme particulière de la maladie. Or, comme nous l'avons vu dans le cas familial, *M. lepromatosis* a été également identifiée chez des patients présentant une lèpre tuberculoïde ou encore lépromateuse au Canada [72], à Singapour [73], en Birmanie et au Brésil [74], aux États-Unis mais chez des cas ayant voyagé dans des zones endémiques [75], laissant penser que son implication pourrait être plus large que ce qui était initialement envisagé. Un fait plutôt intrigant avait été mis en lumière : *M. lepromatosis* a été souvent retrouvé conjointement avec *M. leprae* sur les prélèvements des patients (16% de 87 biopsies considérées) [76]. Enfin, une récente étude, publiée fin 2019, a montré que des réactions de Lucio se sont produites chez des patients uniquement infectés par *M. leprae* [77].

Comme cela avait été suggéré par Gillis en 2011 [78], l'inoculation expérimentale de *M. lepromatosis*, dans les coussinets plantaires de souris ou chez le tatou à 9 bandes, semble déterminante afin de montrer que *M. lepromatosis* provoque, tout comme celle de *M. leprae*, la maladie. Cela procurerait des arguments supplémentaires pour attribuer à cette nouvelle bactérie une pathogénie similaire ou différente de celle de *M. leprae* et d'étayer le lien de cause à effet entre la présence de *M. lepromatosis* chez les patients et le développement de la maladie. Dans ce sens, cette même étude a permis d'isoler *M. lepromatosis* et de l'inoculer dans des coussinets plantaires de souris et permis sa multiplication sans reproduction de la maladie. De même que *M. leprae*, il a également été montré que *M. lepromatosis* est incultivable en milieu artificiel. Ils ont également développé un diagnostic moléculaire afin de distinguer spécifiquement *M. leprae* et *M. lepromatosis* et testé plusieurs échantillons. De plus, compte tenu des similitudes entre *M. leprae* et *M. lepromatosis*, il est intuitif de penser que les tatous peuvent être infectés par *M. lepromatosis*. Ainsi, ils ont également examiné des échantillons de tatous provenant de zones endémiques aux États-Unis où ils sont connus pour être porteurs de la lèpre, aucun échantillon n'a été positif pour *M. lepromatosis* (**figure 10**) [77].

De ce fait, les arguments ne sont pas encore suffisants pour attribuer à cette nouvelle bactérie une pathogénie différente et spécifique de celle de *M. leprae*.

Néanmoins, le séquençage complet du génome de *M. lepromatosis* suivi d'une comparaison avec celui de *M. leprae* a confirmé que les deux entités bactériennes peuvent être qualifiées d'espèces distinctes du point de vue génétique [79,80]. En effet, les gènes codant des protéines partagent en moyenne 93% d'identité de séquence et les pseudogènes 82%. Les auteurs ont également pu estimer que la divergence entre les deux espèces remonterait à 13,9 millions d'années et aurait été suivie d'événements de pseudogénéisation spécifiques à chacune des bactéries (**figure 11**).

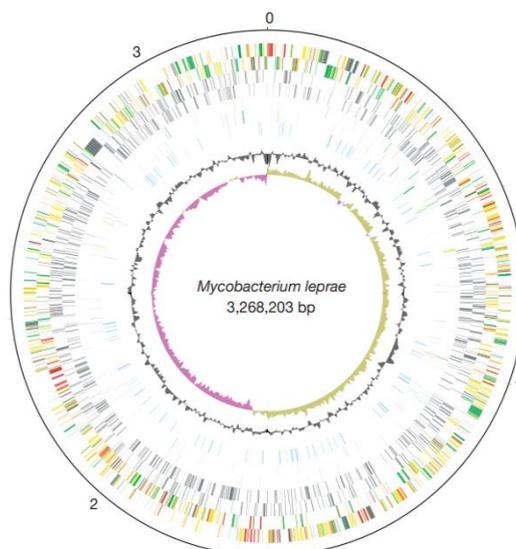


Figure 8 : Carte du génome de *M. leprae* avec la position et l'orientation des gènes connus, des pseudogènes et des séquences répétées [64]

Depuis l'extérieur, les cercles :

1 et 2 : gènes présents sur les brins moins et plus ; 3 et 4 : pseudogènes ; 5 et 6 : gènes spécifiques à *M. leprae* ; 7 : séquences répétées ; 8 : contenu en G+C (57,8%) 9 : rapport (G+C)/(G-C).

Couleur différente suivant la fonction du gène :

noir : métabolisme des lipides ; jaune : respiration et métabolisme intermédiaire ; rouge : machinerie cellulaire ; bleu clair : protéines régulatrices ; orange : protéines hypothétiques conservées ; vert clair : protéines ayant une fonction inconnue ; rose : séquences d'insertion et relatives aux phages ; vert : paroi cellulaire ; magenta : les familles de protéines PE et PPE ; marron : adaptation, virulence et détoxification.

Les nombres à l'extérieur indiquent l'échelle en Mégabases.

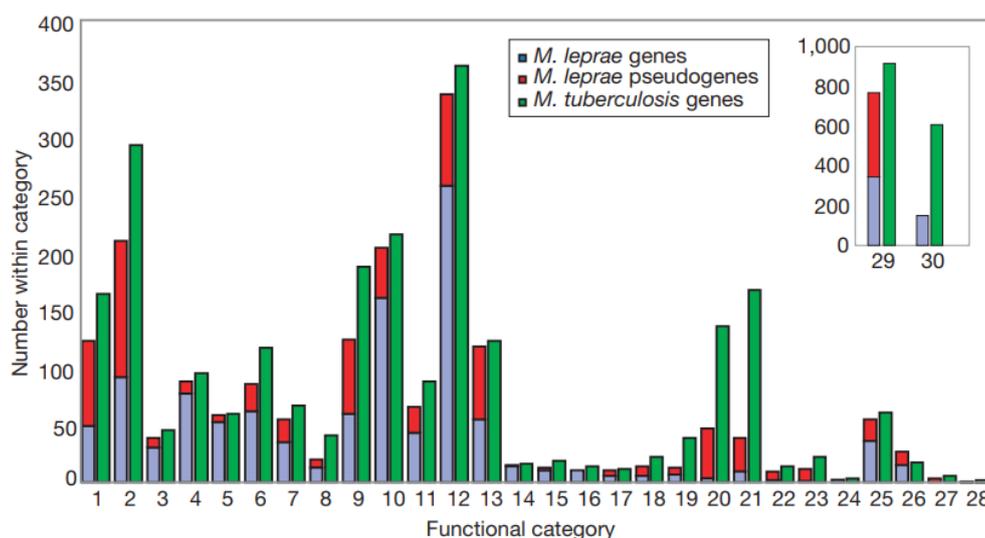


Figure 9 : Comparaison des génomes de *M. leprae* et *M. tuberculosis* [64]

Répartition par catégorie fonctionnelle des gènes (bleu) et pseudogènes (rouge) de *M. leprae* et des gènes de *M. tuberculosis* (vert). Parmi les plus différentes : 1 : catabolisme des petites molécules ; 2 : métabolisme énergétique ; 9 : protéines régulatrices ; 12 : constituants de l'enveloppe cellulaire ; 18 : détoxification ; 20 : éléments IS et protéines dérivées des phages ; 22 : production d'antibiotiques et de résistances ; 23 : enzymes du cytochrome P450 ; 29 : protéines hypothétiques conservées.

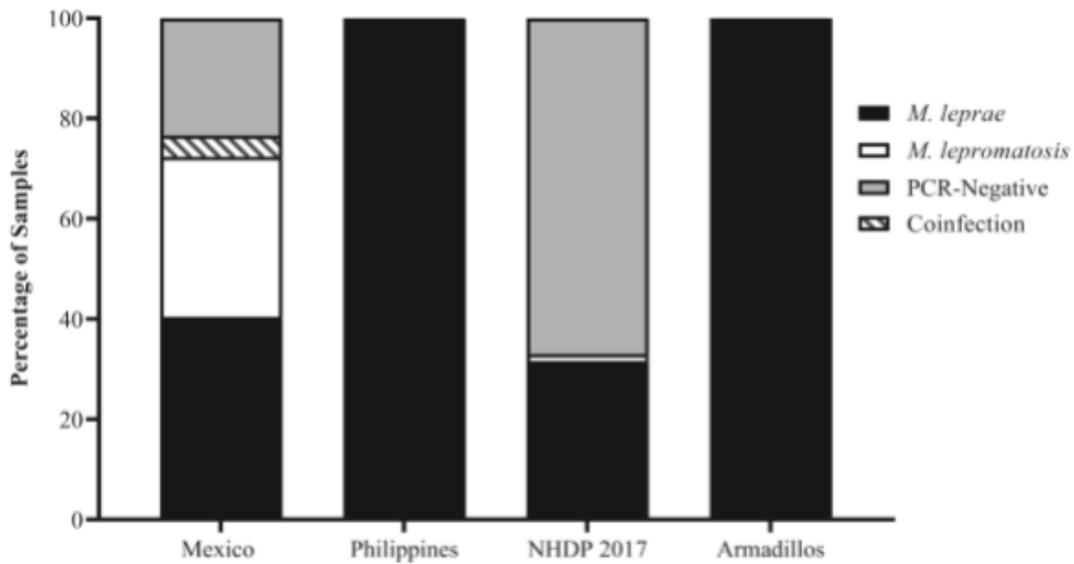


Figure 10 : Pourcentage d'échantillons contenant *M. leprae* et/ou *M. lepromatosis* [77]

Des échantillons cliniques du Mexique ($n = 47$), des Philippines ($n = 180$) des États-Unis ($n = 218$ testés en 2017) et des échantillons de tatous ($n = 106$) ont été testés en utilisant les tests RLEP et RLPM qPCR pour *M. leprae* et *M. lepromatosis*.

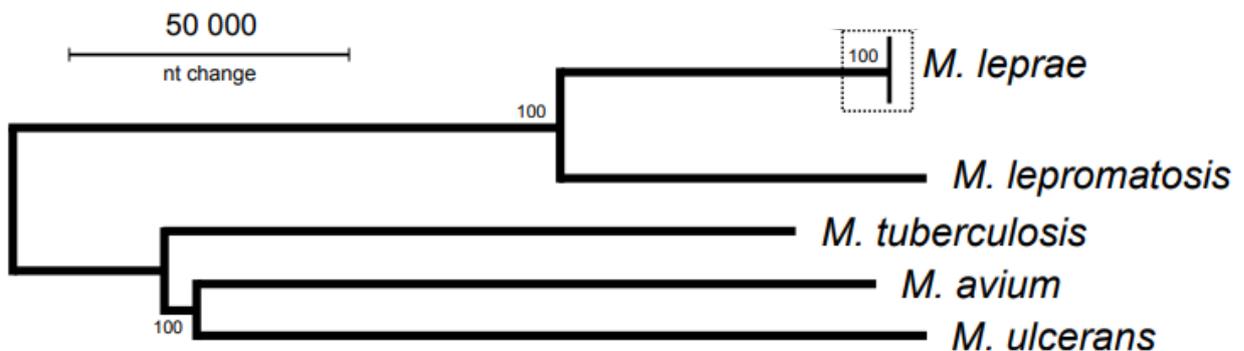


Figure 11 : Phylogénie des souches de *M. lepromatosis* et *M. leprae* [79]

Relation phylogénétique du génome de *M. lepromatosis* en utilisant un arbre de parcimonie maximale, incluant *M. tuberculosis*, *M. avium* et *M. ulcerans* comme sous-groupes.

1.3 Spectre clinique et classification

Le tropisme particulier et limité de *M. leprae* aux cellules de Schwann du système nerveux périphérique et aux macrophages dermiques, fait que la traduction clinique de la maladie est dans 90% des cas dermatologique (avec ou sans signes neurologiques) et dans 10% des cas purement neurologique [50]. La dénomination « lèpre » ou « lèpre per se » regroupe une variété de manifestations cliniques suivant un spectre clinique continu entre deux pôles extrêmes nommés « lèpre tuberculoïde » et « lèpre lépromateuse » : c'est-à-dire la lèpre considérée indépendamment de la forme clinique développée. Quelle que soit la forme de lèpre, la maladie se déclare après une période d'incubation longue, allant de 2 à 12 ans [81].

1.3.1 Forme tuberculoïde

La forme tuberculoïde est caractérisée par une réponse immunitaire cellulaire spécifique adaptée conduisant à la formation de lésions cutanées peu nombreuses (inférieures à 10), des grandes tailles (supérieures à 3 cm de diamètre), bien délimitées par une bordure papuleuse en relief, érythémateuses ou hypo-pigmentées, maculeuses (planes) ou infiltrées. Elles sont toujours hypo ou anesthésiques, caractéristique pathognomonique de la lèpre (**figure 12.A.B**). La charge bacillaire au sein des lésions étant faible, cette forme est donc communément appelée « paucibacillaire » (PB) [50]. Histologiquement, on met en évidence la formation de granulomes giganto-épithéliocellulaires sans nécrose caséuse, bien différenciés et la recherche de bacilles est souvent négative.

Des lésions nerveuses périphériques sont rapidement présentes, importantes dans l'évolution de la maladie et souvent douloureuse. En effet, l'atteinte neurologique, précoce, concerne non seulement les petites fibres nerveuses des lésions cutanées mais également les gros troncs nerveux périphériques (par ordre de fréquence décroissante : tibial postérieur, cubital, médian, sciatique poplitée externe et facial) [82]. L'infiltrat inflammatoire touche la peau et les nerfs, avec une hypertrophie nerveuse souvent palpable cliniquement. Il peut, ainsi, s'agir d'une simple hypertrophie d'un nerf périphérique, mais le déficit moteur peut

entraîner au niveau des mains et des pieds des déformations décrites comme des « déformations en griffe » [83] (**figure 12.E**).

Des ulcérations, des perforations et des destructions osseuses distales peuvent compléter le tableau de cette atteinte clinique.

1.3.2 Forme lépromateuse

A l'opposé, la forme lépromateuse correspond à l'absence de réponse immunitaire cellulaire spécifique résultant en une maladie disséminée. Elle est caractérisée par de nombreuses lésions cutanées (supérieures à 20), maculeuses ou infiltrées, de petites tailles (inférieures à 2 cm de diamètre). Les lésions infiltrées peuvent prendre l'allure de nodules appelés « lépromes » ou, par confluence, de placards annulaires mal délimités (**figure 12.C.D**). Les lésions cutanées ne sont pas ou très peu hypoesthésiques [50]. L'examen histologique retrouve une infiltration lympho-histiocytaire désorganisée et la présence de nombreux bacilles dans les macrophages de la peau et des nerfs, d'où le nom de lèpre « multibacillaire » (MB).

Les manifestations extra-cutanées de la lèpre lépromateuse sont nombreuses, fréquentes et une dissémination systémique est possible, touchant principalement les yeux, la muqueuse nasale, les os, les testicules et les reins. Plus précisément, la rhinite lépreuse aboutit à l'ulcération de la cloison nasale et à l'effondrement du nez. Les yeux sont fréquemment atteints (kératite, uvéite) et peuvent conduire à la cécité [84] (**figure 12.F**). L'atteinte testiculaire entraîne une azoospermie et une gynécomastie. Les autres organes internes (foie, ganglions lymphatiques, reins, rate, poumons, glandes endocrines...) peuvent également être atteints. Contrairement à ce qui est observé dans la forme paucibacillaire, l'atteinte nerveuse clinique est classiquement plus tardive, moins constante et progresse moins rapidement lorsqu'elle existe. Toutefois une atteinte sensitive infraclinique survient précocement, et peut être mise en évidence en pratique courante par l'utilisation de monofilaments de Semmes-Weins [85–87].

1.3.3 Classification de Ridley-Jopling et formes « borderline »

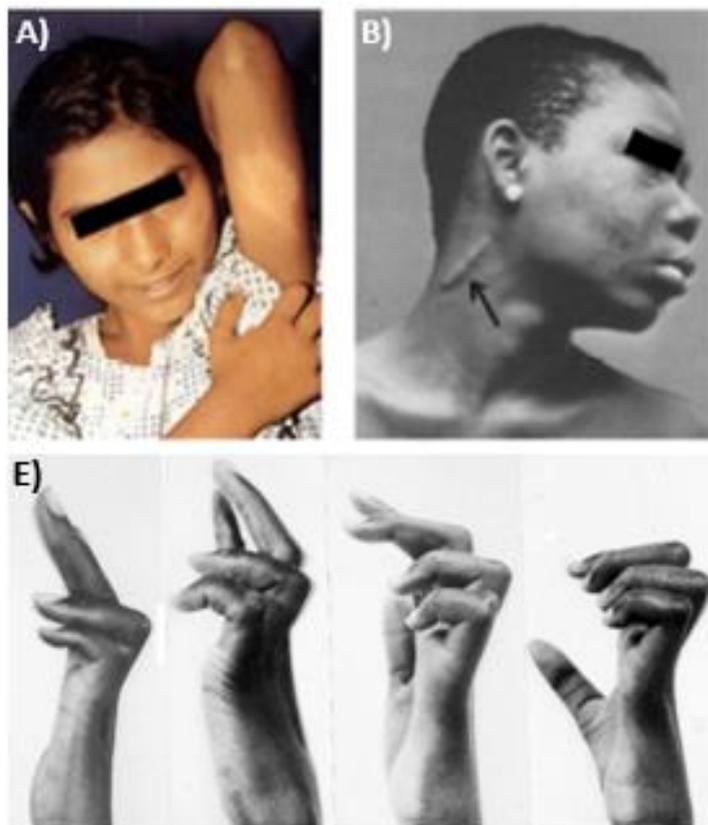
Les formes cliniques comprises entre ces deux extrêmes sont diverses et qualifiées de « borderline ». Ridley et Jopling ont proposé en 1963 une classification, encore très utilisée, des formes de lèpre selon leurs caractéristiques cliniques, bactériologiques, histologiques et immunologiques [88]. Cette classification retient au total 5 formes et une forme indéterminée (I) : les formes à l'extrémité tuberculoïde du spectre clinique sont qualifiées de Tuberculoïde Tuberculoïde (TT), celles à l'extrémité lépromateuse de Lépromateuse Lépromateuse (LL) et les formes intermédiaires de « borderline » sont au nombre de trois et sont nommées Borderline Tuberculoïde (BT), Borderline Borderline (BB) et Borderline Lépromateuse (BL) en fonction de leur degré de similitude avec les formes polaires. Bien qu'initialement dérivée d'arguments histopathologiques, immunologiques, bactériologiques et cliniques, l'attribution d'une classe est fondée quasi exclusivement sur les caractéristiques des lésions cutanées et des atteintes nerveuses (**tableau 1**).

Il existe par ailleurs des entités cliniques bien définies, comme la forme neurale pure, la lèpre histoïde de Wade ou la lèpre de Lucio (précédemment mentionnée) [89]. Cette dernière s'exprime par une infiltration diffuse du derme, sans nodules, avec nécrose et ulcération des lésions. Quelle que soit sa forme, l'évolution de la maladie est lente et l'état général reste longtemps conservé malgré la dissémination du bacille. L'évolution peut cependant être émaillée de deux réactions inflammatoires aiguës.

1.3.4 Classifications de l'OMS

Plus récemment, en 1982, l'OMS a recommandé d'utiliser une classification binaire [90], qui a l'avantage de séparer deux groupes : les formes paucibacillaire (PB) et multibacillaire (MB). L'examen clinique est si possible complété par un frottis des lésions cutanées qui permet la quantification de la charge bacillaire (index bacillaire coté de 1 à 6 qui correspond à la moyenne sur plusieurs sites comme les lobes d'oreilles, une lésion cutanée, le mucus nasal du nombre de bacilles par champ sur le frottis observé au microscope (échelle logarithmique), qui est nulle au pôle tuberculoïde et croissante jusqu'au pôle lépromateux, le plus contagieux. La classe PB incluait initialement les formes TT et BT avec indice bacillaire (IB) négatif et un nombre de lésions cutanées inférieur à 5, toutes les autres situations étant classées MB. En 1988, l'OMS a modifié ses critères bactériologiques : tout patient avec un IB positif était désormais classé comme MB [91]. Depuis 1996, cette classification a été encore simplifiée pour être plus opérationnelle, la lèpre étant classée MB s'il existe au moins cinq lésions, et PB sinon [92]. Cette nouvelle classification, très utile sur le terrain, conduit toutefois à un surtraitement des patients puisque plus 60% des formes MB ainsi diagnostiquées ont en réalité un index bacillaire négatif [93]. Depuis 2018, les patients PB et MB ont le même traitement mais sur des temps différents (voir plus bas).

FORME TUBERCULOÏDE



FORME LÉPROMATEUSE

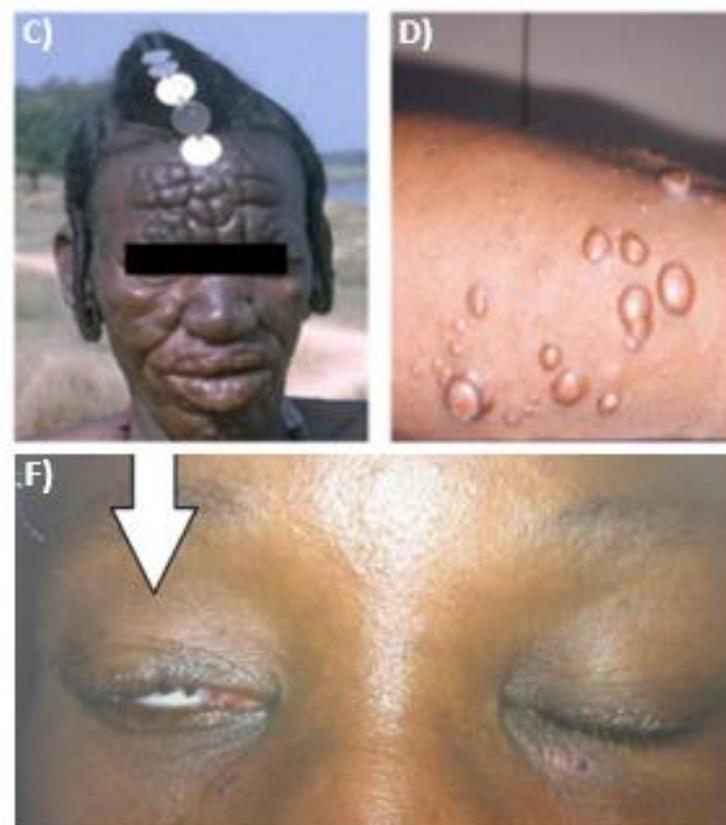


Figure 12 : Images cliniques de la lèpre tuberculoïde et lépromateuse [94]

- A) Lésion cutanée hypopigmentée [95].
- B) Hypertrophie du nerf auriculaire cervicale [96].
- C) Faciès dit « léonin » encore rencontré en cas d'absence de traitement [97].
- D) Lépromes cutanés [98].
- E) Mains décrites comme des « déformations en griffe » [83].
- F) Lagophtalmie droite [84].

Forme de lèpre	TT	BT	BB	BL	LL	
CLINIQUE						
➤ Signes cutanés						
➤ Signes généraux	-	-	-	±	+	
➤ Signes muqueux	-	-	+	++	+++	
➤ Signes nerveux	+++	++	+	+	+	
➤ Signes viscéraux	-	-	+	++	++	
BACILLE DE HANSEN						
➤ Présence	-	±	+	++	+++	
➤ Globi	-	-	-	-	+	
IMMUNOLOGIE						
➤ IDR de Mitsuda	+	±	-	-	-	
BIOPSIE (PEAU)					LLs	LLp
➤ Histologie (infiltrat)	TB	TB	Mixte	MB	MB	MB + VIRCHOW
➤ Bacille de Hansen						
○ Présence	-	±	+	++	+++	+++
○ Golbi	-	-	-	-	+	+++

Tableau 1 : Caractéristiques du spectre clinique de la lèpre selon la classification de Ridley-Jopling [99]

1.4 Épidémiologie de la lèpre

1.4.1 Campagne éradication de l'Organisation Mondiale de la Santé

La lèpre est, parmi les maladies transmissibles, la principale cause de handicap physique dans le monde, et les difformités terribles dues à la forme avancée de la maladie ont depuis toujours contribué aux préjugés profonds ayant conduit au rejet des malades et de leurs proches par la société. Dans un souci de clarté, rappelons que l'OMS a défini en 1988 un cas de lèpre comme « *tout malade présentant des signes cliniques de la lèpre avec ou sans confirmation bactériologique et ayant besoin de recevoir un traitement antibactérien spécifique* ».

En 1941, le premier traitement spécifique et efficace contre la lèpre, la dapsoné, est mis à la disposition des patients et représente un tournant majeur pour les malades [100]. Malheureusement, trente ans plus tard les premiers cas de résistance sont décrits, et à la fin des années 70 leur fréquence est telle que de nombreux échecs thérapeutiques sont enregistrés. En conséquence, l'OMS préconise dès 1981 l'utilisation d'une polychimiothérapie (PCT) associant dapsoné, rifampicine et clofazimine qui reste encore aujourd'hui le traitement de choix. Ce traitement, très efficace et peu coûteux, était cependant inaccessible à de nombreux malades vivant en situation de grande précarité et exclus des circuits diagnostiques et thérapeutiques habituels. Ainsi, au milieu des années 80, la prévalence (nombre de cas sous traitement en fin d'année au 31 Décembre) mondiale de la lèpre atteignait 10 millions de cas, soit un taux de prévalence (prévalence rapportée à la population) d'environ 12/10 000 habitants), l'Inde en comptabilisant à elle seule plus de 8 millions.

En 1991, la prévalence mondiale était encore voisine de 6 millions de cas (plus de 1/1000 habitants), ce qui conduisit l'OMS à faire de l'élimination de la lèpre une de ses priorités pour la fin de l'année 2000. [101]. Pour ce faire, dès 1995, l'OMS a mis la PCT gratuitement à la disposition de tous les sujets atteints dans le monde. Le traitement était initialement financé par la Nippon Foundation et, depuis 2000, il est donné dans le cadre d'un accord avec Novartis qui s'est récemment engagé à prolonger cette action (il faut compter 150 euros pour dépister et soigner un malade de la lèpre).

Tous ces efforts importants et concertés des pays membres et de nombreuses associations caritatives ont permis de ramener la prévalence à 750 000 cas (1,25/10 000 habitants) en 2000, tandis que parmi les 122 pays considérés comme endémiques en 1985, 107 ont atteint le but de l'élimination [102,103]. En mai 2001, l'OMS déclare la lèpre « *éliminée comme problème de santé publique au niveau global* ». Au vu des données épidémiologiques récentes, il est important de rappeler que cette élimination, définie en termes de prévalence mondiale globale, est loin de correspondre à une éradication de la maladie dans les pays endémiques [101]. En effet, le terme « élimination » peut porter à confusion : il ne désigne pas l'absence de survenue de nouveau cas (comme cela a été le cas pour la variole, déclarée éradiquée en 1980), mais un taux de prévalence inférieur à 1/10 000. Cependant, l'incidence (nombre de cas déclaré sur une période et une population donnée) de la lèpre était restée stable, voire s'était légèrement majorée, dépassant les 700 000 cas par an de 1999 à 2001. Certains éléments pouvaient en partie expliquer ce phénomène comme l'élargissement de la couverture géographique ou l'intensification du programme de dépistage. Au vu du long délai pouvant s'écouler entre l'infection et les manifestations cliniques (parfois plus de vingt ans), il est également possible que les nouveaux cas dépistés soient le reflet de transmissions ayant eu lieu avant la mise en place du programme de l'OMS. Dans ce contexte, un programme final d'élimination « *final push strategy* » au niveau national a été mis en place pour les pays n'ayant pas atteint les critères d'élimination. La particularité de celui-ci est d'intégrer les services de soins « anti lépreux » dans les établissements généraux de santé aisément accessibles aux malades [104].

Une dernière campagne débutée en 2016 et intitulée « *une stratégie mondiale de lutte contre la lèpre 2016-2020 - parvenir plus rapidement à un monde exempt de lèpre* » s'articule autour des trois piliers : 1) renforcer l'appropriation par les autorités, la coordination et le partenariat, 2) mettre fin à la lèpre et à ses complications, 3) mettre fin à la discrimination et promouvoir l'inclusion [105,106]. De plus, en août 2016, l'OMS a publié un manuel opérationnel pour faciliter l'adaptation et la mise en œuvre cette stratégie : le but étant de donner des orientations aux administrateurs des programmes nationaux de lutte contre la lèpre [107].

1.4.2 Situation actuelle

« La lèpre est une maladie tropicale négligée, ne négligeons pas ceux qui en souffrent »

« Elle ne fait pas beaucoup de bruit et fait son lit sur la pauvreté, la précarité, l'insuffisance d'accès à la santé. Malheureusement, dans les pays où la Fondation Raoul Follereau agit, tous ces éléments se dégradent et cela dans le silence. » 67^{ème} journée de lèpre en Janvier 2020 [103].

Malgré de nombreuses rumeurs sur sa disparition, la lèpre représente toujours la seconde mycobactériose humaine la plus fréquente après la tuberculose [108]. À l'occasion de la 67^{ème} journée mondiale des lépreux, les 24, 25 et 26 Janvier 2020, le point sur cette maladie, trop souvent passée sous silence a été mis en lumière [109].

Si, en près de 10 ans, les nouveaux cas de lèpre détectés diminuent légèrement dans le monde, il reste encore de nombreux foyers fortement endémiques (244 796 en 2009 à 208 619 en 2018 [110]) (**figure 13**). Cette année encore, plus de 200 000 nouveaux cas sont répertoriés dont près de 5% d'enfants de moins de 15 ans engendrant des dizaines de milliers d'infirmes. Même si la tendance indique donc un déclin progressif du nombre de nouveaux cas détectés dans le monde, en réalité, depuis que l'OMS a déclaré l'infection éliminée comme problème de santé publique globale, le nombre de nouveaux cas dépistés chaque année n'a pas diminué et reste d'ailleurs très probablement sous-estimé [103,109].

Dans le monde, on compte encore près de 3 millions de « lépreux » avec des infirmités ou des mutilations, ce chiffre étant très certainement inférieur à la réalité car ces personnes sont considérées comme guéries et ne font donc plus l'objet d'un suivi des services médicaux. Une personne est touchée toutes les 2 minutes par l'infection. Ces chiffres étant globaux, la pathologie est surtout surreprésentée dans certaines régions du globe. En effet, si ces 20 dernières années, plus de 16 millions de « lépreux » ont été guéris, et la maladie éliminée dans 108 des 122 pays où elle était considérée comme un problème de santé publique, elle demeure toutefois un enjeu majeur dans 14 pays d'Afrique, d'Asie et d'Amérique Latine. Plus précisément pour six d'entre eux qui regroupent 83% de la prévalence mondiale et 88% des nouveaux cas annuels mondiaux : Inde, Brésil, Indonésie, Népal, République Démocratique du Congo et Mozambique.

D'autre part, la lèpre reste endémique dans de nombreux autres pays (Angola, Madagascar, République Centrafricaine, Tanzanie...) ainsi qu'à Mayotte ! De nos jours, la grande majorité des cas de lèpre diagnostiqués dans les pays développés concerne des migrants en provenance des pays d'endémie de la ceinture tropicale où les ressources économiques et sociales sont insuffisantes pour obtenir les standards de vie nécessaire à la contention de la maladie. Ainsi, même la France n'est pas épargnée, néanmoins, le chiffre exact de nouveaux cas reste discutable car la déclaration de la maladie n'est pas obligatoire sur notre territoire. Ainsi, selon les sources (si l'on considère les cas diagnostiqués seulement en France ou avec les territoires d'outre-mer, si l'on considère que les cas de nationalité française ou non), entre une trentaine à une soixantaine de nouveaux cas serait répertoriés cette année [103,109].

De manière inquiétante, depuis 2005, la prévalence et l'incidence de la lèpre sont restées stables et identiques aux alentours de 200 000 cas annuels. Par ailleurs, une certaine prudence doit accompagner l'interprétation de ces statistiques, qui reposent sur une déclaration active des cas de lèpre des États à l'OMS. Il est maintenant admis que l'ensemble des cas ne sont pas déclarés. En effet, la pression sur les pays les plus touchés était importante, en particulier sur le taux de prévalence, pour que la déclaration d'élimination puisse être actée et médiatisée. Cette « pression » a pris plusieurs formes mais dans tous les cas a eu pour effet secondaire une sous-déclaration de certains cas, des changements de politique de Santé publique et même des changements de méthode de comptage afin de corroborer, du moins sur papier, la déclaration d'élimination. Mais surtout, les patients dont le traitement de 6 mois est débuté avant juin terminent leur traitement avant décembre, et sortent des registres à la date utilisée pour l'estimation de la prévalence : ils ne sont donc pas comptabilisés. On peut enfin s'étonner sur la précision (au patient près) des données sur la lèpre où, dans ces mêmes pays, les chiffres pour la tuberculose ou le syndrome d'immunodéficience acquise (SIDA) sont arrondis à la dizaine de milliers. Enfin, certains pays, comme l'Inde, ont mis fin aux activités de dépistage actif des cas et à l'examen des contacts. En 2014, il a été estimé qu'en Inde, environ la moitié des cas de lèpre reste non déclarée [111].

Le nombre de cas recensés ne représente que la partie émergée de l'iceberg du problème de santé publique que pose la lèpre. S'il est aujourd'hui, en théorie, techniquement simple de traiter la lèpre sur le plan infectieux, la prévalence et l'incidence ne traduisent

cependant pas la comorbidité importante associée à la lèpre, en particulier les séquelles invalidantes que peut entraîner la maladie. En effet, à l'issue du traitement, ces malades sont considérés comme guéris, même s'ils sont porteurs d'infirmités et ne sont donc plus comptabilisés comme « lépreux ».

Cette définition est intéressante dans le fait qu'elle permet d'évaluer réellement le potentiel de contagiosité de la maladie. Il est ainsi également très important de les comptabiliser selon la gravité, en degrés 1 et 2 (classification de l'OMS). Ces porteurs d'infirmités sont à l'origine de l'image péjorative de la lèpre véhiculée depuis des millénaires. Ainsi, malgré l'existence et la disponibilité d'un traitement permettant de guérir de l'infection, le nombre de nouveaux patients présentant une incapacité de grade 2 (ID2 définie par une déformation visible, ou une acuité visuelle $< 1/10$) reste toujours élevé avec, dans le monde, 11 323 nouveaux cas ayant une ID2 au moment du diagnostic en 2018 contre 14 322 en 2009 [110] et cette proportion augmente à la fin du traitement.

Bien que la lèpre soit considérée comme éliminée à l'échelle mondiale depuis 20 ans, l'éradication de la maladie est toujours loin d'être une réalité. Le ralentissement des résultats obtenus en termes de prévalence et d'incidence, l'existence d'une dizaine de pays encore endémiques ainsi que le pourcentage de malades souffrant de séquelles invalidantes font encore aujourd'hui de la lèpre un enjeu de santé publique. De ce fait, dans ces pays, les moyens de lutte contre la lèpre : diagnostic, traitement et prophylaxie doivent continuer à être intensifiés. De même, la recherche expérimentale d'un nouveau traitement (de durée plus courte) devrait être poursuivie afin d'augmenter l'observance et donc de favoriser l'élimination de la maladie.

Le microbiologiste Pr Cole, qui travaille sur le sujet depuis plus de dix ans, déplore que la lèpre soit devenue une « *maladie délaissée* » et que beaucoup de ses collègues aient réorienté leurs travaux sur *M. tuberculosis*, très proche de celui de la lèpre mais qui, lui, concerne bien plus les pays occidentaux [28].

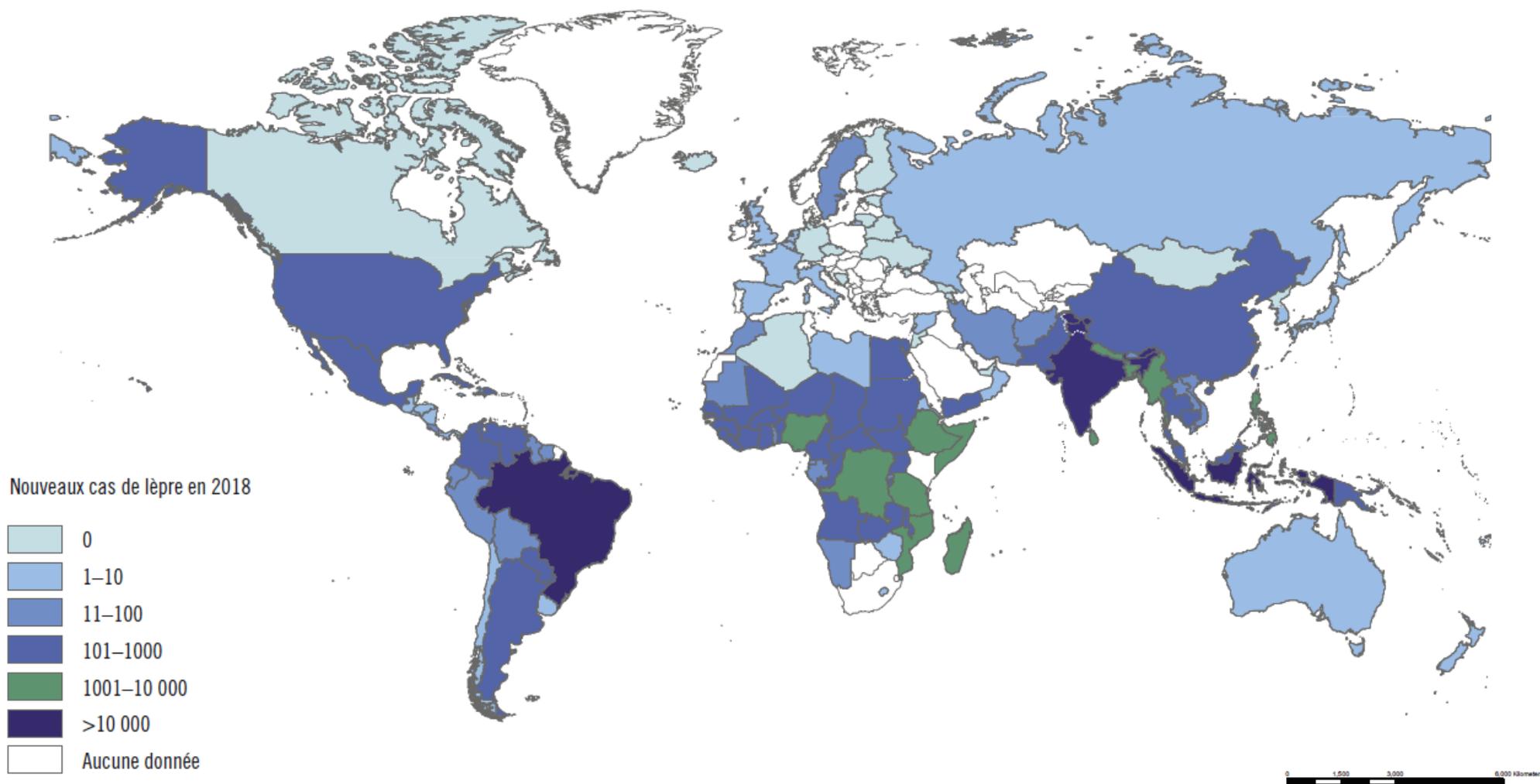


Figure 13 : Répartition géographique des nouveaux cas de lèpre dans le monde en 2018 [110]

1.4.3 Facteurs influençant le développement de la lèpre

Les études épidémiologiques ont permis de mettre à jour plusieurs hypothèses et facteurs environnementaux pouvant influencer le développement de la lèpre.

➤ **Facteurs d'exposition : contact avec un sujet atteint**

Le contact domiciliaire avec un patient multibacillaire est le facteur de risque le plus important (risque relatif de 5 à 10), tandis que l'augmentation du risque lors de contacts avec un patient paucibacillaire est encore controversée [112–115]. Cela s'explique sûrement par le fait que ces patients sont porteurs de lésions riches en bacilles. Néanmoins, la majorité des patients lépreux ne décrivent aucun contact domiciliaire avec un individu lépreux, en particulier dans les régions de haute prévalence [116].

➤ **Géographie**

Des facteurs géographiques et ethniques entrent également en jeu. Ainsi, la proportion de patients multibacillaires diffère nettement en fonction de l'origine ethnique mais aussi au sein de chaque continent, y compris entre pays limitrophes. Prenons un exemple, parmi les pays qui ont déclaré plus de 1 000 nouveaux cas en 2018, la proportion des cas MB (classification OMS) variait de 40% (Bangladesh) à 90% (Philippines) [110]. Il est peu probable que les données démographiques (âge et sexe) expliquent la variabilité observée, étant donné que les pyramides des âges sont relativement homogènes au sein d'un continent et que, sauf exception, le sex-ratio au sein d'une population est toujours proche de 1.

En 1966, K. Newell, un épidémiologiste qui se définit comme « béotien » au sujet de la lèpre, a publié un article passé bien trop inaperçu intitulé « *Regard d'un épidémiologiste sur la lèpre* ». Il s'est appuyé essentiellement sur l'observation suivante : dans les pays de forte endémicité, la prévalence des cas lépromateux (qui s'apparentent aux cas MB) était relativement stable (entre 5 et 10/1000), quelle que soit la prévalence totale de la lèpre (entre 10 et 40/1000).

Cette observation l'a conduit à rejeter certaines hypothèses jusqu'alors avancées : 1) la forme lépromateuse est causée par une souche plus virulente de *M. leprae*, 2) les cas lépromateux sont associés à un plus grand *inoculum* et/ou une plus grande exposition, 3) la polarisation est liée à un facteur environnemental commun à toutes les populations humaines. Sur la base de cette même observation, il propose comme la plus plausible, l'hypothèse selon laquelle une proportion fixe de la population infectée présente une ou plusieurs caractéristiques qui la rendent plus susceptible de développer une forme lépromateuse de la lèpre. L'observation de ce phénomène dans des populations aux modes de vie très différents rend raisonnable l'hypothèse de la nature génétique de cette ou ces caractéristiques qui représente le cœur de ce manuscrit [117].

➤ Dépistage tardif, niveau socio-économique et nutrition

La lèpre demeure une maladie de la pauvreté, de l'exclusion, du manque d'hygiène et la promiscuité et s'attaque généralement à des personnes éloignées du système de soins. Il est aujourd'hui admis que l'amélioration des conditions de vie a contribué à la disparition de la lèpre en Europe [22]. Ainsi, le niveau socio-économique joue un rôle majeur dans la dissémination de la maladie, notamment les mauvaises conditions d'habitat et un bas niveau d'éducation qui diminue les chances d'accès aux soins. Tout d'abord, comme les malades ne souffrent pas des premiers signes cliniques, ils ne perçoivent donc pas l'urgence qu'il y a à consulter un médecin. Au début, la lèpre est souvent considérée comme une simple dermatose, courante dans les zones rurales des pays concernés et qui est donc rarement traitée. Ajoutons à cela que dans beaucoup de pays d'Afrique, la médecine traditionnelle a encore une grande valeur : ainsi, les malades se tournent donc en priorité vers des « guérisseurs », qui ne connaissent pas la maladie et sont incapables de la guérir.

Autre contrainte importante du dépistage passif et qui limite son efficacité : l'éloignement des centres sanitaires. Pour se rendre dans un hôpital ou au dispensaire le plus proche, les malades doivent parfois marcher durant des jours. Le fait de s'absenter de leur travail sur une longue période peut peser lourdement sur l'économie familiale. Sans compter que, si la prise en charge de la lèpre est gratuite pour le patient, la consultation générale nécessaire au dépistage est payante : un frein supplémentaire pour ces populations pauvres.

Enfin, cette maladie reste encore largement stigmatisée. Restant « tabou » dans de nombreux pays, la honte qui y est associée demeure un obstacle à la consultation spontanée. Les patients préfèrent notamment se cacher ou se rassembler dans des lieux confinés plutôt que de se rendre aux centres de santé pour y recevoir des soins adaptés. Un dépistage précoce permet, pourtant, d'éviter la propagation de l'infection, de traiter le malade et de le guérir définitivement et sans séquelle. Or, la pratique la plus courante pour dépister la pathologie consiste « à attendre » que les cas suspects viennent de leur propre initiative pour consulter dans les services de santé. Cette méthode ayant montré ses limites, un dépistage actif semble une solution prometteuse [113,118,119].

Enfin, le statut nutritif pourrait également influencer le risque de lèpre, en particulier durant l'enfance [120]. Ainsi, l'épidémie de lèpre qui commença en 1912 dans l'île de Nauru (Océan Pacifique) explosa en 1920 après qu'une terrible épidémie de grippe (30% de mortalité) eut laissé une population affaiblie avec de nettes carences alimentaires [121]. De façon similaire, de nettes aggravations cliniques ont été décrites dans les léproseries chinoises au cours de la deuxième guerre mondiale [122]. Il est cependant difficile, dans ces observations, de dissocier la malnutrition d'autres facteurs pouvant influencer l'état immunitaire des individus.

➤ Âge

Bien qu'une lèpre puisse se développer à n'importe quelle période de la vie, il existe une répartition bimodale de l'âge de survenue, avec un premier pic d'incidence entre 15 et 20 ans et un second autour de 40 ans [112]. D'autre part, l'âge moyen des cas MB est généralement plus élevé que celui des cas PB : l'explication commune est une plus longue période d'incubation pour la forme MB [123]. Une étude réalisée en Inde en 1962 a montré que la proportion des cas MB varie de moins de 5% chez les patients âgés de 5 à 9 ans à plus de 25% chez ceux âgés de 25 à 64 ans [124]. Deux études réalisées au cours de la même période, l'une dans la région indienne du Tamil Nadu, l'autre en Birmanie retrouvaient cette même augmentation avec l'âge [125,126]. Une étude plus récente menée entre 2001 et 2010 et toujours en Inde retrouve la même tendance avec une proportion des cas MB de 54% chez les enfants (seuil d'âge non précisé) contre 77% chez les adultes ($p < 0,001$) [127].

➤ Sexe

De façon encore inexpliquée mais constante dans les études, le nombre de cas de lèpre rapportés chez les adultes est significativement plus élevé chez les hommes que chez les femmes, avec un sex-ratio homme/femme compris entre 1,5 et 3 [112]. Les biais potentiels liés à l'exposition, à la déclaration de la maladie ou à l'accès aux soins, souvent avancés comme explications d'un sex-ratio supérieur à 1, semblent ne pas pouvoir expliquer entièrement une telle différence. Dans certains pays africains comme le Burkina Faso, le Kenya ou l'Ouganda, un sex-ratio inférieur à 1 a été décrit [128].

Peu d'études se sont intéressées spécifiquement à l'estimation du sex-ratio chez les patients PB et MB. Une vaste étude brésilienne, qui a utilisé les données du recensement du ministère de la Santé entre 2006 et 2010 (40 544 cas PB et 29 764 cas MB), a estimé un sex-ratio de plus de 2 pour les patients MB, mais légèrement inférieur à 1 pour les cas PB [129]. Ceci expliquerait qu'en fonction de la proportion relative des formes PB et MB dans une population donnée, le sex-ratio global puisse varier considérablement, et que dans certains pays africains, où la forme PB est prédominante, les femmes soient surreprésentées parmi les cas de lèpre. Une autre façon d'étudier l'influence du sexe sur la polarisation est de regarder séparément la proportion de cas PB ou MB parmi les malades, selon qu'ils soient des hommes ou des femmes.

➤ BCG

Deux méta-analyses comprenant respectivement 26 et 28 études ont montré que la vaccination par le BCG réduisait significativement le risque de présenter des signes cliniques de lèpre [130,131]. En effet, la diminution du nombre de cas de lèpre grâce au BCG variait de 26% à 68%, selon que les méta-analyses se concentraient sur les études observationnelles, expérimentales ou l'ensemble. Plus précisément, les études expérimentales [132–137] montrent un effet protecteur global de 26% sur l'incidence de la lèpre (intervalle de confiance (IC) : 14-37%), effet que les études observationnelles avaient tendance à surestimer, avec une efficacité globale estimée à 60% [138]. D'autre part, l'âge de vaccination par le BCG ne semble pas avoir d'incidence sur cette protection. Par ailleurs, deux études sur trois ont montré que la re-vaccination par le BCG majorait cet effet protecteur contre le lèpre [135,136,139].

Concernant le statut MB ou PB, la première méta-analyse, montre une réduction plus importante dans le groupe MB que dans le groupe PB, bien que seulement marginalement significative au risque 5%, (78% versus 59% avec $p = 0,04$). Dans la seconde, un ajustement a été fait sur le nombre de doses de BCG reçues et sur le type des études incluses : l'impact du BCG variait alors de 41 à 68% mais n'était plus significativement différent dans les groupes PB ou MB. Enfin, une étude a suggéré que les patients LL étaient insensibles à la vaccination car ces patients ne sont pas en mesure de développer une réponse granulomateuse [140].

Enfin, l'adjonction de *M. leprae* tués par la chaleur ou de *Mycobacterium w* (une mycobactérie environnementale à croissance rapide) dans le produit vaccinal pourrait augmenter modestement l'effet protecteur du BCG, et la combinaison avec le bacille ICRC (une mycobactérie à croissance lente cultivée à partir d'un lépreux humain depuis 1958) semblerait encore plus efficace [136]. Cependant, si les effets observés sont significatifs, ils sont insuffisants pour justifier un programme de couverture systématique dans les pays endémiques.

➤ VIH

Depuis qu'il est admis qu'il existe un effet aggravant du VIH sur la tuberculose, il a été légitime de se demander si par mimétisme, le VIH a un effet aggravant sur la lèpre. En effet, le rôle de l'immunosuppression relative constituée par le SIDA dans le déclenchement de la tuberculose a motivé plus d'une dizaine d'études [141] alors que, à l'inverse, les études cherchant à établir les interactions entre l'infection par le VIH et la lèpre sont moindres [142,143]. De manière générale, il semblerait que le VIH n'a pas d'effet sur le développement de la lèpre clinique, ce qui est surprenant. D'une part, la prévalence de la lèpre n'a pas augmenté avec l'explosion de la prévalence de l'infection par le VIH [144–147]. La majorité des études cas-témoins montrent que l'infection par le VIH n'est pas plus fréquente chez les patients lépreux que dans le reste de la population [142,143]. L'infection par le VIH ne semble pas non plus modifier le type de lèpre [142]. En effet, la présentation clinique et le rapport des formes multibacillaire/paucibacillaire sont identiques à ceux observés dans la population non infectée par le VIH [148–150]. Néanmoins, deux études d'Afrique de l'Est montrent une augmentation des cas MB [151,152].

Histologiquement, la réponse immunitaire cellulaire T-médiée est conservée à l'extrémité tuberculoïde du spectre [153–155], même chez des patients ayant un taux très faible de CD4 circulants, contrairement à ce qui est observé dans la tuberculose. Le seul lien entre les deux pathologies semble être le développement de réactions lépreuses de réversion chez des patients lépreux infectés par le VIH dans les semaines suivant l'instauration d'une thérapie antirétrovirale hautement active (HAART) [156–162]. Ces réactions correspondent probablement à des syndromes de reconstitution immunitaire, à l'instar des réactions paradoxales observées au décours de la mise sous HAART chez les patients non lépreux infectés par le VIH, compliquant des maladies infectieuses concomitantes à l'infection par le VIH (principalement la tuberculose) [163]. Néanmoins, ces réactions peuvent être gérées avec des corticostéroïdes ou de la thalidomide selon les cas [143,164,165].

Malheureusement, comme l'expliquent les auteurs, la longueur de la période d'incubation et la faible incidence de la lèpre n'a pas permis de mener des études prospectives sur l'incidence de la maladie chez des patients VIH-positifs versus des individus VIH-négatifs, ni même, des études de prévalence [142,165]. Les études publiées s'intéressent plutôt à la lèpre comme facteur de risque de développement du SIDA que l'inverse. D'autres études devront confirmer ou infirmer ces hypothèses.

1.4.4 La pléiotropie :

➤ Maladies inflammatoires de l'intestin : la maladie de Crohn

En l'absence d'études fonctionnelles spécifiques à la lèpre, certains gènes ont été explorés dans d'autres maladies qui partagent un fond génétique commun avec la lèpre, comme les maladies inflammatoires chroniques de l'intestin (MICI). En effet, certaines études ont révélé des voies de signalisation communes entre la lèpre et la maladie de Crohn (MC) [166–168] dont l'origine bactérienne par une potentielle mycobactérie est toujours débattue [169–172]. Prenons un exemple, dans le premier GWAS en population sur la lèpre, quatre gènes (*NOD2*, *TNFSF15*, *RIPK2* et *HLA DR-DQ*) appartiennent à la voie de signalisation impliquant *NOD2* [166]. *NOD2* est également le premier gène de susceptibilité découvert par une approche pangénomique dans la maladie de Crohn. Les odds-ratio des variants de *NOD2* étaient supérieur à 3. C'est ainsi l'un des gènes avec l'effet le plus fort pour une maladie humaine complexe [173].

Plus précisément, la protéine *NOD2* (gène *NOD2/CARD15* - Nucleotide-binding Oligomerization Domain 2) appartient à la famille des récepteurs du soi appelé « Pattern Recognition Receptor » (PRR). Il est exprimé dans le cytosol des monocytes, des macrophages, des cellules dendritiques et des cellules intestinales de Paneth [174]. Il se lie à un motif bactérien ubiquitaire, le muramyl-dipeptide et stimule une cascade moléculaire aboutissant à l'activation de NF- κ B. Les récepteurs de la famille NOD sont largement étudiés en particulier au moyen de souris *knock-out* pour ces gènes. Ces découvertes ont relancé l'hypothèse d'une contribution infectieuse à la maladie de Crohn [169], avec un intérêt particulier pour *M. avium paratuberculosis*. Cette bactérie est l'agent étiologique de la maladie de Jones, pathologie digestive semblable à la maladie de Crohn qui sévit chez les bovins. Il est par ailleurs intéressant de souligner que *M. avium paratuberculosis*, dont la contribution est énergiquement discutée, valide les 4 postulats de causalité définis par Koch vis-à-vis de la maladie de Crohn, à la différence de *M. leprae* vis-à-vis de la lèpre [167].

Le gène LTA est également associé à la maladie de Crohn : la lymphotoxine α soluble est sécrétée par les lymphocytes B et T CD4+ et les cellules Natural Killer (NK). Les études murines ont montré que la LTA participe au recrutement des macrophages et des lymphocytes aux sites d'infection par des bactéries intracellulaires [175]. Chez l'homme, la production de lymphotoxine α est diminuée chez les patients qui présentent l'haplotype LTA+80A-LTA+368C. Ces patients sont ceux qui présentaient un risque accru de lèpre per se dans l'enfance dans deux études vietnamienne (données familiales) et indienne (cas-contrôles) [14,176]. Les analogies entre les MICI et la lèpre per se peuvent donc inviter les scientifiques travaillant sur la lèpre à utiliser les progrès de la génétique des MICI pour comprendre l'architecture des déterminants génétiques de la lèpre.

➤ Lèpre et COVID-19 : un lourd impact pour les patients

- **Un accès aux soins très limité**

Comme pour la majorité des maladies, la crise sanitaire actuelle de la COVID-19 paralyse la chaîne de soins pour la lèpre. Avec les différents confinements et restrictions à l'échelle mondiale, les défis sanitaires sont encore plus grands qu'auparavant. Les services humanitaires manquent de médicaments et même de pansements pour soulager et soigner les malades de la lèpre. D'autant que la plupart des services médicaux ont été détournés vers la lutte contre la propagation de la COVID-19. Sans compter les pénuries, qui renforcent la malnutrition, favorisant ainsi le développement de la maladie. Pour soigner la lèpre, le traitement est long et demande une certaine logistique. Il nécessite l'administration d'un médicament pendant 12 mois au total. Le président de l'International Leprosy Association (ILA), Roch Christian Johnson explique que « *le cordon sanitaire mis en place en Afrique a empêché les malades de venir les chercher* ». Les interventions non urgentes telles que la chirurgie reconstructive ont été reportées [177,178].

Le dépistage actif de nouveaux patients est aussi mis à l'arrêt. C'est ainsi que les risques de développer des handicaps à vie à cause de la lèpre augmentent chaque jour avec la crise. Les hôpitaux continuent les consultations et les hospitalisations, mais la situation est grave. Les déplacements sont compliqués et les gens ont du mal à y arriver, ce qui a réduit les consultations de près de deux tiers [179].

« Lorsque l'on regarde les chiffres de l'année dernière, nous constatons une baisse du nombre de nouveaux cas recensés. Ces données ne reflètent pas la réalité, car beaucoup de malades n'ont pu être dépistés ... alors que la maladie est présente », souligne encore le Dr Johnson [178].

- **Interaction entre *M. leprae* et le SARS-Cov2**

De manière générale, aucune interaction particulière n'a été mise en évidence entre *M. leprae* et le virus de la COVID-19. Chez les patients atteints de lèpre, il a été vérifié que les porteurs de COVID-19 peuvent être soit asymptomatiques, soit présentent des degrés variables d'insuffisance respiratoire jusqu'au stade « grave » et en association avec la tempête cytokinique pouvant aller jusqu'au décès des patients. Cela représente le tableau d'atteinte « standard » de la COVID-19 et avoir la lèpre ne semble donc pas être un facteur de risque ni de protection [180].

Néanmoins, la tempête cytokinique caractéristique de la COVID-19 semble grandement ressembler aux réactions de réversion (voir plus bas) dans le cas de la lèpre. Une étude montre que chez les patients ENL (voir plus bas) et chez les patients COVID+, l'expression des neutrophiles semble être similaires [180]. D'autre part et sans vraiment de preuves scientifiques, un communiqué de chercheurs de la « Société Brésilienne de la maladie de Hansen » pensent que « à mesure que le nombre de nouveaux cas d'infection par la COVID-19 augmente, l'incidence des réactions lépreuses pourrait également augmenter de manière considérable » et que « il convient de rappeler que les patients ayant des RR sont traités avec des médicaments qui affectent la stabilité du système immunitaire, ce qui peut contribuer à la manifestation du SRAS, en particulier chez les personnes âgées présentant des comorbidités » [181].

1.5 Diagnostic clinique et paraclinique de la lèpre

On estime qu'environ 5% des individus exposés à *M. leprae* développent une lèpre, avec une durée d'incubation de quelques mois à plus de trente ans, en moyenne quatre ans pour les formes tuberculoïdes et dix ans pour les formes lépromateuses [182]. De manière générale le diagnostic de la lèpre est le plus souvent clinique. À l'heure actuelle, il n'existe aucun test paraclinique simple qui permette le diagnostic de lèpre, notamment dans sa forme tuberculoïde (ou paucibacillaire). L'OMS insiste donc sur le diagnostic de la lèpre selon des critères purement cliniques car dans la majorité des pays où la lèpre est endémique, les équipements permettant un diagnostic bactériologique, immunitaire ou histopathologique, ne sont pas suffisamment accessibles au personnel de santé pour être utilisés. En effet, le recours à l'histologie n'est possible que dans les grands centres de référence, ainsi, sur le terrain, la plupart des patients sont donc diagnostiqués et classés sur la seule clinique.

1.5.1 Diagnostic clinique

Il est aisé lorsque des signes pathognomoniques sont présents principalement une hypertrophie d'un nerf périphérique et une ou des lésions cutanées anesthésiques.

➤ Interrogatoire du patient

L'interrogatoire du patient doit commencer par une reconstitution aussi exacte que possible de ses antécédents médicaux. Les questions fréquemment posées sont [95] :

- *Depuis quand est apparue la tache cutanée ? Comment a-t-elle commencé ? A-t-elle changé ? En général, les taches de lèpre apparaissent lentement.*
- *Est-ce que les taches démangent ? Y a-t-il une douleur ? Les lésions cutanées de lèpre ne démangent pas et sont indolores en général.*
- *La personne sent-elle que ses mains ou ses pieds sont devenus plus faibles ? A-t-elle des problèmes à tenir ou soulever des objets et à bouger les mains et les pieds ? La perte de force dans les mains ou les pieds peut être un signe de lèpre.*
- *La personne vit-elle des problèmes sociaux ? Ceci serait plus plausible si le sujet a déjà des infirmités dues à la lèpre.*

➤ **Examen de la peau**

La peau du patient doit être examinée dans un lieu discret, isolé, à la lumière du jour ou dans une pièce bien éclairée, elle doit être examinée des pieds à la tête, aussi bien de face que de dos. Les taches de la lèpre sont habituellement plus claires que la peau environnante ; elles peuvent être rougeâtres avec un bord surélevé mais elles peuvent également prendre d'autres formes d'aspect très différentes (brillantes et sèches au touché). Parfois un épaissement de la peau sans taches cutanées peut être présent. La lèpre cause parfois des nodules ou des nodosités sur la peau, en général, il s'agit d'infection grave [95]. Avant toute conclusion hâtive, il faut faire un test de la sensibilité des taches cutanées.

➤ **Test de sensibilité des taches cutanées**

Ce test permet de vérifier la sensibilité des lésions cutanées. Le praticien demande de fermer les yeux au patient, puis il va toucher les lésions cutanées avec une mèche de coton ou la mine d'un stylo et demander alors au patient d'indiquer le point concerné. Le test doit être réalisé à la fois sur les taches et sur la peau normale. Si les lésions cutanées sont hypoesthésiques, il pourrait s'agir d'une forme plutôt tuberculoïde [95].

➤ **Palpation des nerfs**

L'hypertrophie d'un nerf peut être un signe évocateur de la lèpre. Le nerf cubital, qui passe derrière le coude, est le nerf le plus fréquemment hypertrophié ainsi que le nerf sciatique poplité externe. La palpation de ces deux nerfs cubitaux représente donc une étape au diagnostic clinique de la maladie [95].

➤ **Examen des mains et des pieds**

L'atteinte des nerfs peut provoquer une perte de sensibilité des mains et des pieds. Une personne avec une insensibilité, peut se blesser sans s'en rendre compte, expliquant la fréquence des plaies et des ulcères chez les sujets ayant la lèpre. La perte de sensibilité est rare dans les autres maladies, aussi ceci peut aider à confirmer le diagnostic.

De la même manière que pour le test de la sensibilité des taches cutanées, le praticien de santé va cette fois-ci tester la sensibilité à quatre endroits spécifiques de la paume de la main et des pieds et demander au sujet qui a fermé les yeux de montrer l'endroit concerné par la piqûre [95] (**figure 14**).

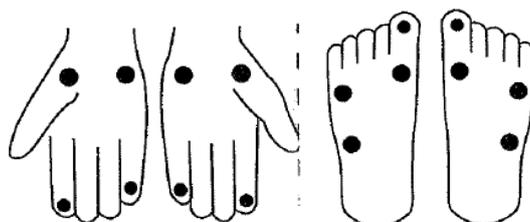


Figure 14 : Test de la sensibilité des paumes de la main et des plantes des pieds lors du diagnostic clinique [95]

1.5.2 Diagnostic paraclinique

➤ Histopathologique : le Gold Standard

L'histologie d'une lésion cutanée est ensuite l'examen le plus contributif, à la fois pour le diagnostic (identification de bacilles) et pour la classification (nature de l'infiltrat). En effet, l'examen de référence, rarement réalisable sur le terrain, reste la biopsie de peau profonde à la marge d'une lésion cutanée active, fixée dans la paraffine et examinée par un anatomopathologiste expérimenté. Les principaux critères à évaluer sont : 1) le type de réponse immunitaire de l'hôte (granulome giganto-épithéliocellulaire en cas de forme tuberculoïde, infiltrat mononucléé avec histiocytes spumeux (ou cellules de Virchow) en cas de forme lépromateuse), 2) l'identification de bacilles acido-alcool-résistants (BAAR) après coloration de Ziehl-Neelsen, et 3) l'atteinte nerveuse (la localisation de BAAR dans les nerfs est pathognomonique). En cas de lèpre tuberculoïde, les bacilles peuvent être difficiles à mettre en évidence et le diagnostic différentiel est alors celui d'un granulome giganto-épithéliocellulaire (tuberculose, sarcoïdose, granulome annulaire) [88,183,61].

➤ Bactériologique : L'index bacillaire (IB) ou Morphologique (IM)

L'examen au microscope optique après coloration de Ziehl-Neelsen, du suc dermique d'une lésion cutanée ou du lobe de l'oreille, ou des sécrétions nasales permet une évaluation semi-quantitative de la charge bacillaire par champ microscopique selon l'échelle logarithmique de Ridley-Jopling [88]. Le résultat formulé en 7 classes (de 0 à 6), est appelé index bacillaire. Très utile pour le classement de la maladie (tout IB positif signifiant un classement dans la catégorie multibacillaire), il participe de ce fait au choix thérapeutique et facilite également le suivi des patients pendant ou après traitement. Il n'est cependant fiable que s'il est pratiqué par des techniciens expérimentés. D'autre part, l'Index Morphologique (IM) exprime le pourcentage de bacilles uniformément colorés et morphologiquement intacts (de détermination délicate) ; il n'est pas recommandé en pratique de routine.

➤ Immunologique : le test de Mitsuda

Il ne s'agit pas d'un test diagnostique mais d'un test permettant d'évaluer la capacité à développer une réaction immunitaire cellulaire adaptée à *M. leprae*. De nos jours, il n'est néanmoins plus utilisé. Il consiste en l'injection intradermique de 0,1 ml de lépromine (suspension stérile de bacilles lépreux tués par la chaleur), dont la lecture se fait 21 à 28 jours plus tard en mesurant le diamètre d'induration de la réaction cutanée. Une infiltration strictement inférieure à 3 mm est considérée comme négative. Néanmoins, le test de Mitsuda ne peut être utilisé pour le diagnostic car il peut être positif chez des individus non infectés par *M. leprae*. Ainsi, en zone endémique, mais aussi non endémique, 70 à 93% des individus sains ont un test positif [184,185]. En effet, d'autres mycobactéries que *M. leprae* sont capables de positiver la réaction, en particulier *M. tuberculosis* et *M. bovis*. Sur le plan immunologique, la réaction de Mitsuda correspond à la formation d'un granulome giganto-épithéliocellulaire en réponse à l'injection de *M. leprae*, elle-même fortement corrélée à la clairance bactérienne [186,187]. Chez un patient atteint, une réponse négative au test de Mitsuda est très fortement prédictive du développement d'une forme lépromateuse (toujours négatif chez les patients LL). De ce fait, il n'est aujourd'hui plus utilisé.

➤ Immunologique : la sérologie

La réponse immunitaire humorale au cours de l'infection par *M. leprae* est globalement plus développée chez les patients du pôle lépromateux, avec une prédominance des anticorps d'isotype IgG2. Différents constituants protéiques ou polysaccharidiques de la paroi mycobactérienne sont capables d'induire cette réponse humorale, la plupart étant communs à plusieurs mycobactéries [188]. Parmi eux, le glycophénolique PGL-1 a été particulièrement étudié en raison de sa spécificité pour *M. leprae*. Malheureusement, sa sérologie s'est avérée peu utile pour le diagnostic de lèpre en raison de son manque de sensibilité chez les patients PB. En effet, le dosage d'anticorps anti-PGL-1 est positif chez 90% des patients MB mais chez seulement 40 à 50% des patients PB. De plus, 1 à 5% des individus sains ont également un dosage positif en région d'endémie [189].

➤ L'apport de la biologie moléculaire : la PCR

Contournant l'obstacle de l'impossibilité de culture in vitro de *M. leprae*, la PCR permet désormais la détection de gènes, d'ARN 16S ou de polymorphismes spécifiques de *M. leprae* [61], avec une spécificité proche de 100%. La sensibilité de la PCR sur biopsie ou frottis cutané est élevée (de l'ordre de 95%) chez les patients MB mais nettement moins (autour de 55%) chez les patients PB [190]. La PCR pourrait donc permettre un diagnostic aussi sensible et plus spécifique que l'examen histologique, sur des types de prélèvements variés (biopsie cutanée, nerveuse ou ganglionnaire, frottis cutané, sécrétions nasales, sang). Néanmoins, les formes indéterminées sont presque constamment négatives. La PCR à visée diagnostique n'a donc pas actuellement sa place dans le diagnostic de la lèpre et cette technique n'est cependant pas recommandée en pratique clinique, en dehors peut-être du diagnostic différentiel avec une autre infection mycobactérienne, lorsque des BAAR sont visibles à l'examen histologique mais qu'il existe des atypies cliniques. L'intérêt des techniques de PCR est également en cours d'investigation dans le suivi des patients [191], sachant que seule l'amplification des ARN messagers (et non des ADN) de *M. leprae* est corrélée à la viabilité des bacilles.

Enfin, cette technique a révolutionné le domaine de la détection de la résistance aux antibiotiques. En effet, *M. leprae* ne pouvant être cultivé in vitro, le seul moyen de détecter ces résistances était auparavant l'inoculation des coussinets plantaires de 20 à 40 souris par

des organismes viables issus des lésions du patient et les résultats n'étaient disponibles que 6 à 12 mois plus tard [192].

Grâce à la PCR, il est désormais possible d'identifier directement les mutations génétiques responsables de résistances aux antibiotiques en réduisant considérablement le nombre de bacilles nécessaires et surtout le délai d'obtention des résultats (quelques heures au lieu de quelques mois). Plusieurs variantes de la technique sont utilisées en routine [61], et une puce à oligonucléotides permettant l'investigation simultanée de toutes les substitutions nucléotidiques connues impliquées dans la résistance de *M. leprae* a été mise au point [193].

1.5.3 Infection infraclinique

Un signe médical infraclinique est un signe que l'on peut observer chez le malade (pour une maladie donnée) avant qu'il n'en présente les symptômes. Un des objectifs majeurs des recherches en immunologie a été la mise au point d'un test permettant de déceler cette infection infraclinique qui permettrait d'établir un diagnostic précoce et donc une amélioration de la prise en charge et du traitement. À nos jours, l'absence de marqueur fiable d'infection infraclinique pour la lèpre (à la différence de l'intra-dermo réaction à la tuberculine et les tests de production d'interféron *in vitro* pour la tuberculose par exemple), ne permet pas de comparer des individus exposés à des individus infectés, ni des individus infectés à des malades.

Dans le cas de la lèpre, seul 5% des personnes exposées à de fortes sources d'infection contractent la lèpre, l'infection infraclinique doit donc être fréquente [136,139,194]. L'hypothèse d'une contamination à partir de ces sujets porteurs d'une forme infraclinique de lèpre, ou même à partir de porteurs sains, reste d'actualité. Ces observations suggèrent donc l'existence de facteurs génétiques de l'hôte qui prédisposeraient certains individus à développer une lèpre après l'infection par *M. leprae*, tandis que d'autres seraient porteurs sains.

1.6 Antibiothérapie, séquelles, prophylaxie

1.6.1 Antibiothérapie : historique et recommandations de l'OMS

➤ Historique

En 1941, le premier traitement spécifique et efficace contre la lèpre, la Dapsone (famille des sulfones), est mis à la disposition des patients. Cet antibiotique révolutionna le pronostic des patients atteints de lèpre. Malheureusement, la monothérapie à long terme associée à une compliance imparfaite des patients entraîna, 30 ans plus tard, l'apparition de résistance. Ainsi, à la fin des années 70 leur fréquence est telle qu'elle rend inefficace la Dapsone en monothérapie avec de nombreux échecs thérapeutiques (jusqu'à 40% dans certaines régions du monde) [195]. La rifampicine, introduite par la suite, s'est avérée extrêmement efficace contre *M. leprae* et est la plus bactéricide (une dose de 600 mg est capable de tuer 99.9% des organismes viables) mais des résistances à cet antibiotique apparurent également rapidement. Le troisième antibiotique disponible, la clofazimine, ne pouvait être utilisé en monothérapie en raison de sa faible bactéricidie. En conséquence, l'OMS recommanda en 1981 l'utilisation d'une PolyChimioThérapie (PCT) qui reste encore aujourd'hui le traitement de choix de la maladie et institua la classification clinique binaire de la forme de lèpre que nous avons définie plus haut, où les patients PB et MB bénéficiaient d'un traitement différent [196]. Si les cas de résistances aux différents antibiotiques augmentent (en 2018, des résistances sont identifiées dans 8% des cas dont : dapsone 4,5%, rifampicine 3,8%, ofloxacine 1% et dapsone + rifampicine 1%) [197], aucune résistance à la PCT n'a jusqu'à présent été rapportée [198]. L'efficacité de cette PCT ne fait aucun doute en pratique, et des millions de patients ont ainsi été traités et guéris depuis 25 ans.

Trois autres antibiotiques ont également couramment été utilisés pour traiter des cas de lèpre : l'ofloxacine, la minocycline et la clarythromycine [89]. Les principales caractéristiques pharmacologiques ainsi que les mécanismes de résistance, les modalités de prescription et les effets secondaires de ces molécules sont bien détaillés (**tableau 2**). De manière plus anecdotique, d'autres quinolones, la moxifloxacine, la péfloxacine, la levofloxacine et la sparfloxacine sont également actifs [199,200].

➤ Recommandations de l'OMS

Plus particulièrement, en 2018, l'OMS a émis de nouvelles recommandations afin de faciliter la mise en place du traitement, d'éviter les erreurs de diagnostic PB vs. MB. En effet, un inconvénient de la classification clinique PB vs. MB est le risque de sur-classement de cas PB en tant que MB, mais aussi, de sous-classement de cas MB faussement étiquetés PB. Ainsi, une même trithérapie est recommandée pour les PB et les MB mais avec des temps différents : six mois et 12 mois respectivement. Il s'agit d'une administration journalière de dapsons (100mg/j) associée à une administration mensuelle supervisée de rifampicine (600mg/mois) et d'une administration de clofamizine journalière (50mg/j) et mensuelle (300mg/mois). Pour les cas PB, la clairance bactérienne est en effet obtenue en six mois mais les bacilles lépreux tués persistent longtemps dans les tissus infectés et les lésions cutanées peuvent demeurer actives pendant un à deux ans. Ainsi, 2 à 44% des patients ont encore des lésions cutanées actives à la fin du traitement, sans que cela constitue un facteur de risque de rechute [201–203]. Pour les cas MB, le schéma initialement recommandé par l'OMS était l'administration de cette trithérapie pendant au minimum deux ans mais cette durée a été réduite à 12 mois en 1997 [92], bien qu'aucune étude contrôlée randomisée ne permette d'étayer ces dernières recommandations [138], et reflète plutôt une réduction des ressources allouées à la lèpre dans les pays en voie de développement. Une année supplémentaire de PCT est donc envisageable, et particulièrement recommandée si les lésions se détériorent après l'arrêt du traitement [204]. Ainsi, le traitement est poursuivi jusqu'à guérison clinique chez les patients tuberculoïdes et jusqu'à deux ans de traitement (mais pas plus) pour les patients avec IB élevé (4+ à 6+) chez les patients lépromateux.

Pour les formes à lésion unique, une monoprise de rifampicine, ofloxacine et minocycline est recommandée par l'OMS depuis 1997, mais là encore le schéma proposé est très controversé. En effet, une étude comparative randomisée a montré que ce traitement était moins efficace que le traitement standard de type PB en termes de guérison [205] et le taux de rechute semble très élevé après traitement en monodose [206]. Ces individus bénéficient donc actuellement du traitement du groupe PB (**figure 15 et tableau 3**).

Nous aborderons plus tard le traitement prophylactique, dont l'efficacité à long terme est discutée [207,208]. La PCT est mis gratuitement à disposition par Novartis qui s'est récemment engagé à prolonger cette action jusqu'à 2020 et est distribué par l'OMS au niveau de chaque pays. Les controverses concernent actuellement la durée de traitement, sa voie d'administration (supervisée ou non), le taux d'échec, plus que le choix des molécules [209]. Certains médicaments en développement pour la tuberculose pourraient bénéficier aux patients lépreux. Ainsi la bédaquiline (ou R207910), qui a une activité bactéricide sur *M. leprae* chez la souris, attend d'être utilisé chez l'homme. Les effets secondaires des traitements actuels devraient stimuler la recherche de nouveaux traitements. Une étude génétique a identifié un polymorphisme du CMH (le HLA-B*13 :01) qui prédit le risque de développer une réaction d'hypersensibilité à la dapsonne pouvant conduire à des décès [210]. Ce polymorphisme serait présent chez 1 à 12% des Indiens, pays qui déclare le plus de cas de lèpre.

➤ Rechutes

Malgré l'efficacité indiscutable de la PCT à court terme, des rechutes peuvent survenir dans les dix ans qui suivent la fin du traitement. Chez les patients PB, le taux de rechute varie d'une étude à l'autre, mais reste globalement très faible (< 2,5%) [201–203,211–217], et non corrélé à la persistance de lésions à l'arrêt du traitement [216]. Chez les patients MB, le risque de rechute est un peu plus élevé, avec là encore de grandes disparités inter-études (de 0,7 à 8%) [203,211–215]. Comme nous l'avons évoqué plus haut, ce taux de rechute est beaucoup plus important chez les patients MB avec fort IB (16 à 39%) [211,218,219] mais semble bien contrôlé par la prolongation du traitement [211].

Antibiotique	Famille	Cible	Action sur <i>M. leprae</i>	Mécanisme de résistance	Posologie	Effets secondaires	Indication
Rifampicine	Rifamycine	Sous-unités β de l'ARN polymérase, codée par le gène rpoB	Excellent bactéricide	Mutations faux sens dans rpoB (dont Ser425Leu)	Dose mensuelle 1200mg	Rares en dose mensuelle (élévation des transaminases)	1 ^{ère} intention
Dapsone	Sulfone synthétique	Dihydroptéroate synthase (enzyme clé dans la voie de synthèse des folates chez les bactéries = inhibiteur compétitif de l'acide p-aminobenzoïque)	Bactériostatique, Faible bactéricide	Mutations faux sens de folP1 (codant pour la dihydroptéroate synthase)	100mg/j	Rares : hypersensibilité, éruptions cutanées, anémie hémolytiques méthémoglobinémique, (sujets G6PD déficients)	1 ^{ère} intention
Clofazimine	Imino-phenazine	Partiellement inconnue... se lie préférentiellement à l'ADN mycobactérien sur les séquences contenant des guanines	Bactéricide, Faible anti-inflammatoire	Rarissime et relevant probablement de différents mécanismes	50mg/j + dose mensuelle de 300mg	Coloration brune de la peau et xérodémie à partir du 3 ^{ème} mois (réversible en 1 an à l'arrêt du traitement)	1 ^{ère} intention + traitement des ENL
Ofloxacine	Fluoroquinolone	Inhibe l'ADN gyrase bactérienne	Bactéricide	Mutation faux sens d'une région du gène gyrA	400mg/j	Rares : photosensibilité, troubles neurosensoriels, tendinopathie	Résistance ou intolérance à rifampicine ou clofazimine
Minocycline	Tétracyclines	Inhibe la synthèse des protéines en se liant à la sous-unité ribosomale 30S	Bactériostatique	3 mécanismes possibles*	100mg/j	Rares : troubles digestifs, vestibulaires, photosensibilité	Résistance ou intolérance à la rifampicine
Clarithromycine	Macrolide semi-synthétique	Inhibe la synthèse des protéines par liaison au ribosome	Bactériostatique, Faible bactéricide	Inconnu (pas de mutation du gène ribosomal RNA23S)	100mg/j	Rares : élévation des transaminases	Résistance ou intolérance à la rifampicine

Tableau 2 : Antibiotiques actifs sur *M. leprae* [220]

* Efflux énergie-dépendant (par une protéine de membrane), inactivation enzymatique de la tétracycline, ou protection du ribosome par une protéine soluble

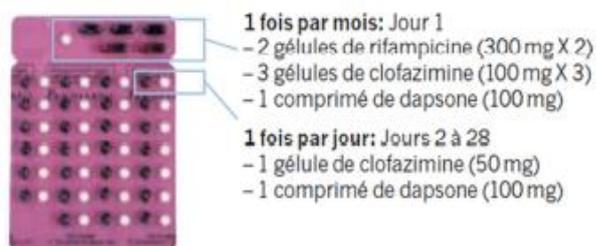


Figure 15 : Plaquette de PCT des formes multibacillaires et paucibacillaires

<i>Type de lèpre</i>	<i>Antibiotiques</i>	<i>Doses</i>	<i>Périodicité</i>	<i>Délivrance supervisée</i>	<i>Durée</i>
Multibacillaire	Rifampicine	600mg	Mensuel	Oui	12 mois
	Clofazimine	300mg	Mensuel	Oui	12 mois
	Clofazimine	50mg	Quotidien	Non	12 mois
	Dapsone	100mg	Quotidien	Non	12 mois
Paucibacillaire	Rifampicine	600mg	Mensuel	Oui	6 mois
	Clofazimine	300mg	Mensuel	Oui	6 mois
	Clofazimine	50mg	Quotidien	Non	6 mois
	Dapsone	100mg	Quotidien	Non	6 mois

Tableau 3 : Recommandations de l'OMS pour le traitement de la lèpre

1.6.2 Antibiothérapie : réactions lépreuses

Depuis l'utilisation systématique de la PCT, la majorité des atteintes neurologiques résulte du développement de réactions immunitaires aiguës : les réactions lépreuses [221].

➤ **Les réactions de type 1 ou réaction de réversion**

La réaction de type 1 (T1R), ou réaction de réversion (RR), est une poussée inflammatoire brutale au niveau des lésions cutanées (œdème, érythème et desquamation) et au niveau des nerfs périphériques (névrite hypertrophique douloureuse). Le risque est un déficit sensitivomoteur total et définitif en l'absence de traitement avec une atteinte articulaire (polyarthrite et ténosynovite). Il s'agit d'une réaction d'hypersensibilité retardée, liée à une amélioration de la réponse cellulaire vis-à-vis de *M. leprae* chez les patients borderline, en relation fréquente avec le traitement. Sur le plan bactériologique, les bacilles sont rares dans les lésions [99]. Plus particulièrement, elles correspondent à une variation d'intensité de la réponse immunitaire cellulaire de l'hôte avec en théorie deux sous-types de réactions possibles : 1) les réactions de dégradation immunitaire (down-grading) qui sont une perte progressive de l'immunité cellulaire, se manifestant par un lent glissement de la forme borderline vers la forme lépromateuse, chez des patients non ou mal traités, et qui semblent de nos jours devenues extrêmement rares ; et 2) les réactions de réversion (RR) (reverse ou up-grading) qui sont des épisodes immunitaires aigus de type hypersensibilité cellulaire (type IV dans la classification de Gell et Coombs) dans la peau ou les nerfs, correspondant donc à une augmentation de l'immunité cellulaire de l'hôte. La prévalence des RR au cours de la lèpre voisine les 30%, avec de grandes variations inter-études [222,223]. Elles s'observent très majoritairement chez des patients borderline et peuvent survenir à n'importe quel moment de l'évolution de la maladie (avant, pendant ou même plusieurs années après la PCT), mais la majorité est observée en cours des six premiers mois de traitement [213,224–226]. Les autres facteurs de risque de RR sont l'extension de la maladie (nombre de lésions cutanées, de nerfs ou d'aires corporelles atteints), et de l'index bacillaire [222,226]. Les manifestations cliniques sont une inflammation aiguë des lésions cutanées préexistantes ou la survenue de nouvelles lésions cutanées inflammatoires, et une multinévrite douloureuse, souvent déficitaire. Des signes généraux (fièvre, malaise, dème des extrémités) peuvent coexister [227]. Les réactions de réversion sont de nos jours la première cause de séquelles neurologiques de la lèpre.

➤ Les réactions de type 2 ou Erythème Noveux Lépreux

La réaction de type 2, ou Erythème Noveux Lépreux (ENL), surviennent uniquement dans les formes BL (environ 10% d'entre elles) et les formes LLs (sous-groupe « subpolaire » de LL) (20%). Les facteurs déclenchants communément retrouvés sont l'initiation du traitement, une infection ou une grossesse. Le début est brutal avec une altération de l'état général et cliniquement caractérisé par l'apparition de nodules inflammatoires dermo-hypodermiques (érythème noveux) fréquemment associés à des signes généraux marqués (fièvre, fatigue, amaigrissement, polyarthralgie, œdème des extrémités). À distance des lésions cutanées existantes, apparaissent des nodules dermo-hypodermiques rouges, chauds, douloureux, de 1 à 3 cm de diamètre, disséminés (nouures). L'évolution des nodules est une desquamation en 48-72 heures, associée le plus souvent à une névrite aiguë de mauvais pronostic, hypertrophique et douloureuse. Des atteintes systémiques (articulaire, ophtalmologique, testiculaire, rénale) sont également observées. Sur le plan immunologique, il s'agit des réponses inflammatoires aiguës systémiques liées à la déposition de complexes immuns dans différents organes (hypersensibilité de type III dans la classification de Gell et Coombs). L'évolution est le plus souvent marquée par plusieurs poussées avant la guérison. Elles sont relativement peu fréquentes, avec une prévalence inférieure à 5% chez l'ensemble des patients lépreux [213,222,224,226,228].

➤ Prise en charge des réactions lépreuses

Les RR nécessitent une prise en charge urgente, de préférence en hospitalisation, et sont traitées par une corticothérapie orale (40-60 mg/j de prednisolone) à décroissance progressive, classiquement pendant 3 à 4 mois [227]. Malgré cette prise en charge, des séquelles neurologiques persistent dans 30 à 50% des cas [229]. Il a été montré que des traitements prolongés (6 mois) limitaient le risque de rechute et de séquelles [230]. L'administration préventive de prednisolone a été évaluée (20 mg/j durant les trois premiers mois de PCT, et arrêt progressif durant le 4^{ème} mois), mais l'effet protecteur initialement observé s'est avéré nul à long terme [231]. Les ENL de gravité modérée répondent aux anti-inflammatoires non stéroïdiens, à la pentoxifilline ou aux antipaludéens de synthèse.

En revanche, les formes graves nécessitent le recours à des cures de corticoïdes (0,5 à 1mg/kg/j) souvent répétées en raison de fréquentes corticorésistances. L'administration de thalidomide (400 mg/j), très efficace, en est le traitement de choix, mais elle n'est disponible que dans les centres de référence. Son utilisation chez la jeune femme doit être exceptionnelle et associée à une contraception efficace en raison des risques de phocomélie durant la grossesse. La clofazimine (300 mg/j) est souvent utilisée en appoint pour son action anti-inflammatoire.

La correction chirurgicale des paralysies (transferts ligamentaires et musculaires) ou des déformations (chirurgie plastique) peut être envisagée lorsque la maladie est contrôlée.

1.6.3 Séquelles

Le fardeau de la maladie ne saurait être évalué par les seuls chiffres d'incidence ou de prévalence : en effet, ils ne sont que la partie émergée de l'iceberg, dont le socle est formé d'individus, certes guéris sur le plan microbiologique, mais qui présentent des séquelles physiques et/ou sociales majeures dues à la maladie. Les réactions de réversion sont la première cause de séquelles neurologiques au cours de la lèpre. On estime à 3 millions le nombre de personnes vivant actuellement avec une séquelle de la maladie [99].

1.6.4 Prophylaxie

➤ « Éducation » des populations et prévention du handicap

Dans beaucoup de pays d'Afrique, la médecine traditionnelle a encore une grande valeur : ainsi, les malades se tournent donc en priorité vers des « guérisseurs », qui ne connaissent pas la maladie et sont incapables de la guérir. Il est donc fondamental pour garantir une bonne observance thérapeutique de fournir aux patients des explications claires et détaillées sur l'origine bactérienne de la maladie, l'efficacité du traitement, l'importance des prises antibiotiques quotidiennes, et les effets secondaires possibles.

Ces informations seront idéalement réitérées lors des visites mensuelles. Par ailleurs, les patients doivent également être informés de la possibilité de rechutes et de réactions et

être incités à consulter rapidement en présence de symptômes évocateurs. Enfin, la prévention des ulcères cutanés est extrêmement importante en cas d'atteinte neurologique, grâce à l'auto-examen à la recherche de traumatisme des zones anesthésiées, particulièrement les mains et les pieds, et la protection de ces zones (notamment port de chaussures orthopédiques en cas de déformations) [232]. En cas d'échec des mesures préventives, les ulcères plantaires débutants requièrent une décharge complète [233] et des soins locaux quotidiens.

➤ Un traitement prophylactique : une dose unique de rifampicine

La stagnation du nombre de nouveaux cas, la nécessité de raccourcir le traitement de la lèpre et de réduire le temps de contagiosité a poussé les chercheurs à essayer de mettre en place un traitement prophylactique. Pour ce faire, il a été proposé d'administrer une dose unique de rifampicine aux personnes ayant été en contact avec des patients atteints de lèpre et une large étude a été menée au Bangladesh. En effet, plus de 28 000 contacts de plus de 1000 patients repartis en deux groupes (placebo contre une dose unique de rifampicine) ont été étudiés. Une réduction de l'incidence de nouveaux cas de lèpre de 57% a été déclarée [234].

Néanmoins, bien qu'elle semble efficace, cette approche montre plusieurs limites : tout d'abord, son utilisation n'a jamais été ajoutée dans les recommandations officielles notamment de l'OMS. L'administration préventive de rifampicine en dose unique ne peut se concevoir que dans un contexte politique sanitaire réfléchi et globalisé. Par ailleurs, son efficacité a été uniquement démontrée chez les personnes n'ayant qu'un contact occasionnel avec un individu atteint de lèpre (« voisins de voisins ») et cela qu'avec des formes PB, donc les formes connues pour être les moins contagieuses. Cette protection semble limitée dans le temps et ne durerait pas plus de deux ans. De plus, la logistique et notamment les rapports coût efficacité reste discutables et soulève plusieurs questions éthiques dans un contexte où la lèpre reste une maladie stigmatisante. Enfin, et afin d'éviter les résistances de *M. leprae* contre la rifampicine, d'autres protocoles sont en cours d'étude comme des prises répétées de rifampicine associées à d'autres antibiotiques [235].

➤ La vaccination

Malgré les efforts de la communauté scientifique, aucun vaccin anti-lépreux efficace n'a encore pu voir le jour. Néanmoins, un vaccin en cours de développement appelé LepVax semble très prometteur. En effet, il s'agit d'une combinaison antigénique (ML2055, ML2380 et ML2028) associée à un adjuvant spécifique sans aluminium (Glucopyranosyl Lipid Adjuvant in stable emulsion ou GLA-SE) stimulant la réponse anti-TH1 et permettant, ainsi, de diminuer les niveaux d'infection de *M. leprae* chez la souris. De plus, il a été observé, chez des tatous infectés par *M. leprae* puis immunisés par LepVax, une diminution et un retard de l'atteinte nerveuse. De ce fait, l'immunisation par LepVax n'a pas exacerbé l'atteinte nerveuse due à l'infection par *M. leprae* et son utilisation semble sans risque. D'autre part, il n'y a pas eu d'inflammation intra-neuronale mais une réduction des lésions axonales, suggérant ainsi que le traitement par LepVax pourrait restaurer une certaine fonction sensorielle. Enfin, ces données indiquent que la prophylaxie post-exposition à *M. leprae* avec LepVax semble non seulement sans danger et, contrairement au BCG, atténue et retarde les atteintes neurologiques causées par l'infection à *M. leprae* [236]. Ce vaccin très prometteur a été très bien tolérés et sans danger chez des volontaires sains aux États-Unis et est aujourd'hui en cours d'essai de phase 1 randomisé contrôle contre placebo au Brésil [237,238].

1.7 Physiopathologie de l'infection à *M. leprae*

Depuis la découverte de *M. leprae* par Hansen en 1873, la recherche scientifique sur la lèpre fut considérablement ralentie par l'impossibilité de cultiver le bacille en milieu artificiel. La découverte d'une méthode de culture dans les coussinets plantaires de souris en 1960, puis de l'existence d'une infection naturelle chez le tatou [54], facilitèrent grandement son étude. Ce n'est qu'à la suite du séquençage de son génome en 2001 [239], que ses caractéristiques structurales et métaboliques furent éclaircies, comme nous l'avons vu plus haut.

Nous aborderons, ici, les mécanismes d'interaction du bacille avec ses deux principales cellules hôtes, le macrophage et la cellule de Schwann, puis les deux polarisations possibles du système immunitaire de l'hôte en réponse à cette infection [61].

1.7.1 Transmission

Un des obstacles majeurs à la prévention de la maladie est l'absence de certitude quant au mode de contamination des patients. Comme vu précédemment, les seuls réservoirs naturels connus de *M. leprae* sont l'homme et, de façon plus anecdotique, le tatou et l'écureuil roux. Néanmoins, et malgré un tropisme limité à deux types cellulaires, les modes de transmission supposés du bacille semblent plus divers. Ainsi, ont été considérées comme portes d'entrée possibles les lésions cutanées et les muqueuses buccales et nasales [240,241]. Plus précisément, le contact cutané fut longtemps considéré comme le principal mode de transmission mais cette hypothèse est contestée, d'une part, lorsqu'elle est intacte, la peau d'individus fortement infectés ne libère qu'un très petit nombre de bacilles, et seules les lésions ulcérées sont donc susceptibles de décharger une quantité conséquente de *M. leprae*. D'autre part, la première lésion cutanée est le plus souvent localisée à des endroits où la peau est couverte par des vêtements [242].

Si la transmission cutanée paraît limitée, la voie aérienne reste considérée comme le mode de transmission principal : la transmission est directe par les sécrétions nasales (gouttelettes nasales, buccales et aérosol) et indirecte par les objets souillés [43].

Malgré l'absence d'une démonstration de la primauté de ce mode de transmission par rapport à d'autres [243], diverses études ont démontré *in vitro* que l'entrée et la survie de *M. leprae* sont possibles dans les cellules épithéliales nasales et alvéolaires pendant plus d'une semaine [244,245] et les souris immunodéficientes développent une lèpre disséminée après l'inhalation d'aérosols contenant des bacilles [246]. De plus, des modifications de la muqueuse nasale chez 39 patients dans des phases précoces de la maladie, suggérant que les voies respiratoires supérieures pourraient être le site premier de l'infection, ont été observées [247]. Cela principalement à partir des patients atteints de forme lépromateuse ou multibacillaires : en effet, Hansen avait déjà établi que la muqueuse nasale de patients lépromateux contenait des quantités massives de bacilles et des études récentes ont montré que le bacille est retrouvé par PCR au niveau des muqueuses nasales chez 14 à 92% des patients multibacillaires et 13% des patients paucibacillaires [248]. Cependant, l'incidence de lèpre ne semble pas forcément plus élevée dans les régions où la proportion de cas multibacillaires est importante.

D'autres modes de transmission ont été décrits comme la transmission materno-fetale [249], l'allaitement maternel ou, des transmissions par voie sanguine [250], via l'allaitement maternel, *in utero*, par piqûres d'insectes ou encore le contact avec des tatous [251–253]. Cependant leur impact en termes de Santé Publique semble limité.

Enfin, l'hypothèse d'une contamination à partir de sujets porteurs d'une forme infraclinique de lèpre, ou même à partir de porteurs sains, reste d'actualité. En effet, la recherche par PCR de *M. leprae* dans les sécrétions et biopsies nasales d'individus vivant en régions endémiques montre que certains individus pourraient héberger le bacille dans leurs cavités nasales : 2 à 8% des individus sains sont porteurs de *M. leprae*, qu'ils soient ou non en contact avec un individu lépreux [248,254,255]. Ces résultats, qui suggèrent l'existence de porteurs sains ou, tout du moins, d'infections infracliniques, doivent cependant être nuancés. En effet, l'utilisation de la PCR reste sujette à des faux positifs, et cela plus particulièrement dans le domaine de l'identification des espèces bactériennes [44]. Par ailleurs, ces observations suggèrent l'existence de facteurs génétiques de l'hôte qui prédisposeraient certains individus à développer une lèpre après l'infection par *M. leprae*, tandis que d'autres seraient porteurs sains. Dans l'état actuel des connaissances on admet que la voie de pénétration est principalement respiratoire.

1.7.2 Immunologie : la région HLA

Le système de l'Human Leucocyte Antigen (HLA), (Complexe majeur d'histocompatibilité ou CMH chez l'homme) est une partie importante du système immunitaire et, est contrôlé par des gènes situés sur le bras court chromosome 6 dans la région 6p21.31. Il code pour les molécules de surface cellulaires spécialisées dans la présentation des peptides antigéniques au récepteur des lymphocytes T (TCR) sur les lymphocytes T [256]. Ce système immunogène est caractérisé par son polygénisme et son polymorphisme qui sont à l'origine d'une grande variabilité interindividuelle. L'ensemble des gènes HLA sont subdivisés en trois régions du chromosome 6 qui contiennent chacune de nombreux gènes avec ou sans fonction immunologique. On distingue ainsi la région CMH de classe I, de classe II, et de classe III [257].

Les molécules du CMH présentant les antigènes (Ag) sont classées en 2 groupes [256] :

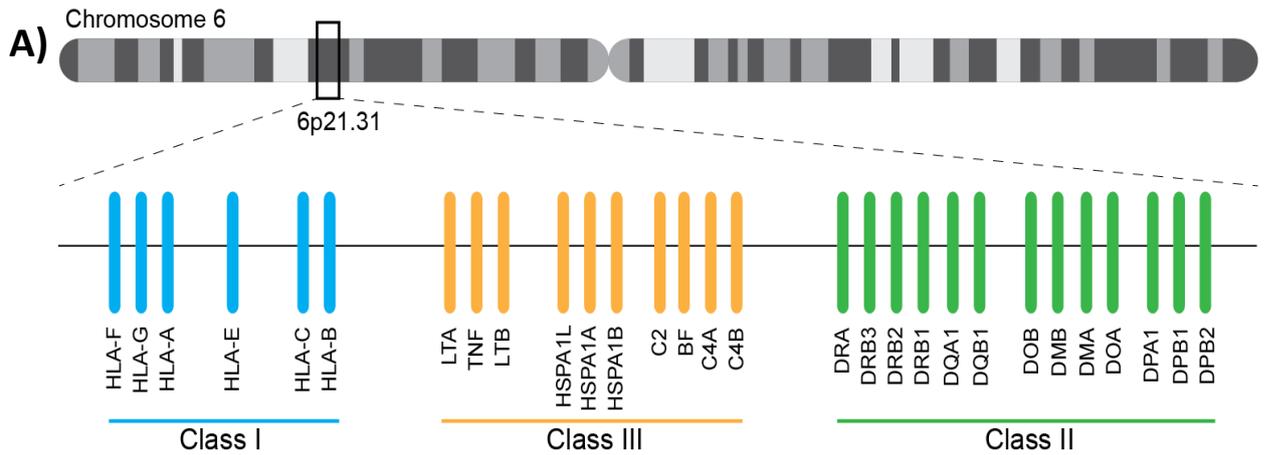
- **Les molécules HLA de classe I** sont présentes en tant que glycoprotéines transmembranaires à la surface de toutes les cellules nucléées. Les molécules intactes de classe I consistent en une chaîne lourde alpha liée à une molécule de bêta-2-microglobuline (β 2m). La chaîne lourde se compose de 2 domaines sur lesquels se fixent les peptides (α 1 et α 2), un domaine de type immunoglobuline (Ig) (α 3) avec une région transmembranaire et une terminaison cytoplasmique (**figure 16.A.B**). La chaîne lourde des molécules de classe I est codée par des gènes des loci *HLA-A*, *HLA-B* ou *HLA-C*. Les lymphocytes T qui expriment les molécules CD8 réagissent avec les molécules du CMH de classe I. Ces lymphocytes ont souvent une fonction cytotoxique, les obligeant à être capables de reconnaître n'importe quelle cellule infectée. Comme chaque cellule nucléée exprime les molécules du CMH de classe I, toutes les cellules infectées peuvent être des cellules présentatrices d'Ag aux lymphocytes T CD8 (les CD8 se lient à la partie non polymorphique de la chaîne lourde de classe I). Certains gènes de classe I non classiques codent les molécules du CMH, telles que HLA-G (qui peuvent jouer un rôle dans la protection du fœtus à partir de la réponse immunitaire maternelle) et HLA-E (qui présente des peptides à certains récepteurs des cellules NK (Natural Killer)).

- **Les molécules HLA de classe II** sont généralement présentes sur les cellules présentatrices de l'Ag (lymphocytes B, macrophages, cellules dendritiques, cellules de Langerhans), l'épithélium thymique et les lymphocytes T activés (mais pas sur les lymphocytes quiescents) ; la plupart des cellules nucléées peuvent être amenées à exprimer les molécules du CMH de classe II par l'interféron (IFN)- γ . Les molécules de classe II se composent de 2 chaînes polypeptidiques (α et β) ; chaque chaîne a un domaine sur lequel se lient les peptides (α 1 et β 1), un domaine de type Ig (α 2 et β 2) et une région transmembranaire avec une terminaison cytoplasmique (**figure 16.A.B**). Les deux chaînes polypeptidiques sont codées par des gènes situés dans la région HLA-DP, -DQ ou -DR du chromosome 6. Les lymphocytes T stimulés par les molécules de classe II expriment des molécules CD4 et sont souvent des lymphocytes helper.

Les Ag définis sérologiquement et codés par des loci de gènes de classe I et II du système HLA reçoivent des désignations normalisées (*ex.*, *HLA-A1, -B5, -C1, -DR1*). Les allèles définis par séquençage de l'ADN sont associés à un nom pour identifier le gène, suivi d'un astérisque, les nombres représentant les groupes d'allèles (souvent l'équivalent sérologique de l'Ag codé par cet allèle), deux points et des nombres représentant l'allèle spécifique (*ex.*, *A*02:01, DRB1*01:03, DQA1*01:02*). Parfois des numéros supplémentaires sont ajoutés après deux points pour identifier des variants alléliques codant pour des protéines identiques, et après un autre deux points, d'autres nombres sont ajoutés pour désigner des polymorphismes des introns ou des régions non traduites en 5' ou 3' (*ex.*, *A*02:101:01:02, DRB1*03:01:01:02*) [256].

La région du génome **du CMH de classe III** code pour plusieurs molécules de l'inflammation importantes ; elles comprennent les fractions du complément C2, C4 et le facteur B ; le tumor necrosis factor (TNF)- α , la lymphotoxine- α , et 3 protéines de choc thermique.

Ces notions seront utilisées et parfois détaillées tout au long de ce manuscrit.



Expression des molécules HLA

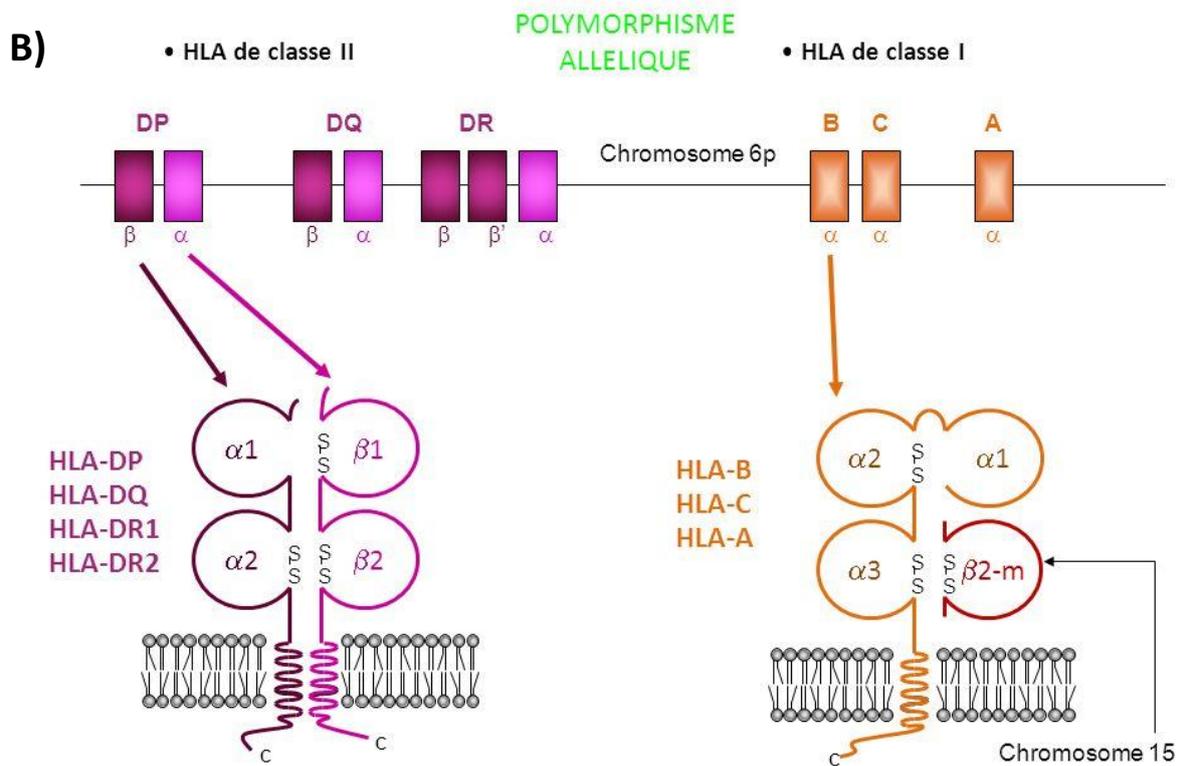


Figure 16 : Caractéristiques de la région HLA chez l'Homme

A) Régions chromosomiques des différentes classes du système HLA [257].

B) Structure peptidique des molécules du CMH de classe I et II [258].

1.7.3 Interaction de *M. leprae* avec le macrophage

Le macrophage est la cellule hôte primaire pour *M. leprae* : en l'absence d'immunité de type cellulaire, ces bacilles peuvent se multiplier dans les macrophages jusqu'à 1100 organismes par cellule. Cette croissance est rendue possible par différents mécanismes. Les macrophages infectés sont réfractaires à la stimulation par l'interféron (IFN)- γ à la fois pour l'induction des molécules CMH classe II de surface et la bactéricidie [282]. La présence de la mycobactérie vivante entraîne le blocage de la maturation des phagosomes qui les contiennent et inhibent leur fusion avec les endosomes, qui est nécessaire à la destruction de la mycobactérie [259]. De plus, *M. leprae* vivante inhibe l'autophagie et l'apoptose des macrophages infectés et empêche ainsi la destruction de la cellule hôte par les phagocytes [260]. Les macrophages activés sont normalement capables d'inactiver les pathogènes invasifs par la génération de réactifs oxydés ou de réactifs nitrogénés. Le premier mécanisme est inefficace contre *M. leprae* car la présence de PGL-1 entraîne une hyporégulation de la génération de superoxyde et le bacille possède de surcroît une bonne résistance aux réactifs oxygénés. Les réactifs nitrogénés semblent en revanche capables d'inhiber l'activité métabolique du bacille. De fait, plusieurs études ont observé un taux élevé d'iNOS (inducible nitric oxide synthase, qui produit l'oxyde nitrique (NO) à partir d'arginine chez les macrophages activés) dans les lésions TB et plus encore au cours des RR [261].

1.7.4 Mécanisme de l'atteinte nerveuse

La cellule de Schwann est la deuxième cellule hôte de prédilection de *M. leprae*, caractéristique unique dans le monde bactérien. De grands progrès ont été accomplis dans la compréhension des mécanismes de l'infection nerveuse par *M. leprae*. Des études chez le tatou ont montré que cette infection survenait par voie sanguine ou lymphatique périneurale (et non de l'intérieur par voie ascendante à partir de la peau comme on le croyait auparavant) [61]. La bactérie atteint ainsi l'endoneurium puis adhère à la surface des cellules de Schwann grâce à une liaison spécifique au domaine G de la chaîne α -2 de la laminine 2, protéine de la lame basale au niveau des cellules de Schwann dont la présence se restreint aux nerfs périphériques [262].

Le bacille est ensuite ingéré par les cellules de Schwann qui, incapables de le détruire, lui fournissent un environnement protégé pour sa multiplication. En retour, *M. leprae* favorise leur prolifération en activant Erk1/2 (extracellular activated kinase), qui enclenche leur cycle cellulaire via la cycline D1 [263]. *M. leprae* induirait également une démyélinisation de ces cellules, même en l'absence de cellule immunitaire, par le biais d'un mécanisme dépendant du PGL-1 [264]. Une étude sur des larves de poisson-zèbre montre que la démyélinisation et les lésions axonales ne sont pas directement provoquées par *M. leprae*, mais par des macrophages infectés qui se fixent sur les axones ; la démyélinisation se produit dans ses zones de contact et c'est PGL-1 qui confère cette capacité neurotoxique aux macrophages pour détruire les axones. Plus précisément, PGL-1 active une enzyme, l'iNOS dans les macrophages infectés qui va induire l'augmentation de réactifs nitrogénés qui vont détruire les mitochondries des cellules de Schwann et donc provoquer la démyélinisation et la destruction des axones [265] (**figure 17**). La neuropathie résulte non seulement de l'infection directe des nerfs périphériques mais aussi de la réponse inflammatoire et immunitaire cytotoxique de l'hôte, secondaire à la présentation restreinte par le CMH de classe II de peptides mycobactériens par les cellules de Schwann [266]. Ce phénomène explique donc le tropisme particulier de *M. leprae* pour les nerfs périphériques. Les déficits neurologiques les plus sévères surviennent dans les formes à forte réponse immunitaire spécifique (TB ou RR).

Une autre très belle étude et plutôt surprenante chez les souris montre que *M. leprae* est capable de détourner le processus cellulaire pour reprogrammer les cellules de Schwann infectées pour qu'elles se différencient en un type de cellule souche mésenchymateuse, (« stem-like cells) en régulant à la baisse les gènes associés à la différenciation et à la hausse ceux qui régulent le développement du mésoderme. La reprogrammation s'accompagne de changements épigénétiques augmentant la plasticité cellulaire et rendant les cellules infectées migratoires et immunomodulatrices. En effet, ces propriétés favorisent la propagation de *M. leprae* par deux mécanismes : la différenciation directe vers les tissus mésenchymateux (muscles squelettiques et lisses), et la formation de structures semblables à des granulomes libérant des macrophages chargés de bactéries (**figure 18**). Ces résultats soutiennent un modèle de reprogrammation de la cellule hôte dans lequel un pathogène bactérien utilise la plasticité de sa niche cellulaire pour favoriser sa dissémination et fournit un lien inattendu entre la reprogrammation cellulaire et l'interaction hôte-pathogène [267].

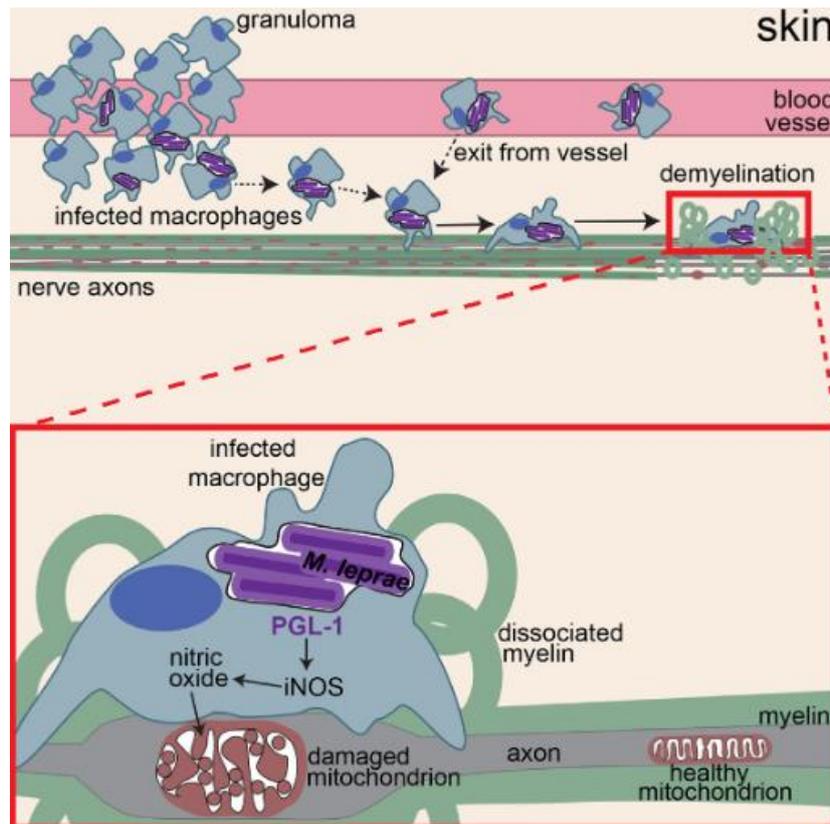


Figure 17 : Les macrophages infectés par *M. leprae* sont à l'origine de la destruction des cellules de Schwann [265]

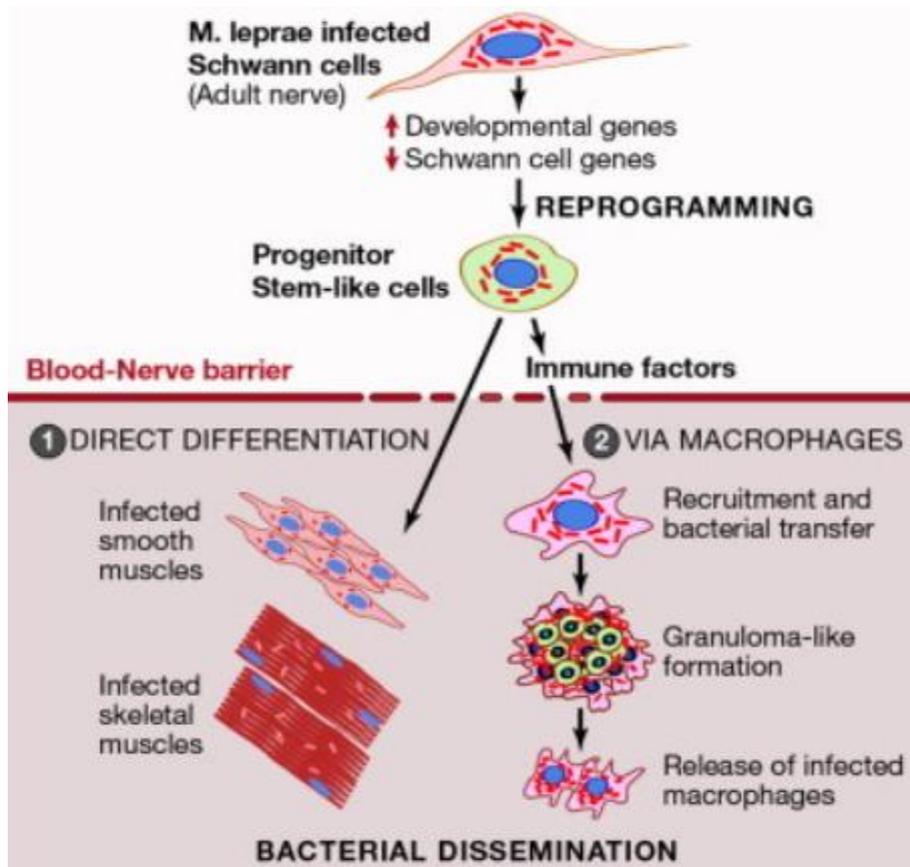


Figure 18 : *M. leprae* est capable de reprogrammer les cellules de Schwann [267]

1.7.5 Immunologie : polarisation de la réponse immunitaire de l'hôte

M. leprae est présente dans les tissus de tous les individus qui développent une lèpre clinique. Toutefois, les différentes formes cliniques de lèpre sont déterminées par la réaction immunologique sous-jacente de l'hôte à *M. leprae* (**figure 19**). La lèpre est ainsi un exemple remarquable de l'impact de l'équilibre entre immunité cellulaire et humorale sur les aspects cliniques de la maladie. Alors que l'immunité cellulaire des patients présentant une lèpre de type tuberculoïde est capable de contrôler la multiplication des micro-organismes en développant une forte réponse immunitaire cellulaire contre la mycobactérie (ce qui limite la maladie à quelques tâches cutanées et aux troncs nerveux), les patients lépromateux présentent une anergie qui semble spécifique de *M. leprae*, car ils ne sont prédisposés ni aux autres infections ni aux cancers. De nombreuses études ont mesuré la réponse immune locale dans les lésions cutanées lépreuses, avec différentes méthodes [183,268].

Chez les TT, il se forme des granulomes gigantoépithélio-cellulaires bien délimités autour de nerfs dermiques. Les lésions sont vierges, ou presque, de *M. leprae*. La forte immunité cellulaire est confirmée par la prolifération des cellules T après stimulation par des antigènes de *M. leprae in vitro* [209]. *In vivo*, le test à la lépromine (ou test de Mitsuda) est positif. En revanche, la réponse humorale des patients tuberculoïdes, évaluée par la production d'anticorps dirigés contre les antigènes de *M. leprae* (dont PGL-1), est faible ou nulle. A l'opposé, la lèpre lépromateuse (LL) est caractérisée par l'absence d'immunité cellulaire spécifique (malgré une immunité intacte contre *M. tuberculosis* chez les patients). *M. leprae* prolifère au sein des lésions. Le derme contient des macrophages spumeux remplis de bactéries mais peu de lymphocytes T CD4+ ou CD8+. Ceux-ci ne s'organisent d'ailleurs pas en granulome. On détecte des titres élevés d'anticorps anti-PGL-1 dans le sang ou les urines. Le test de Mitsuda est négatif chez les individus lépromateux (diamètre < 3 mm). Les formes cliniques intermédiaires ou « borderline » sont caractérisées par une réduction progressive de la réponse cellulaire depuis le pôle tuberculoïde vers le pôle lépromateux. Ceci est associé à une augmentation de la charge bacillaire dans la peau et les nerfs (**figure 20**). Ces formes borderline sont cliniquement instables, et peuvent évoluer soit lentement vers le pôle lépromateux, soit brutalement par une réaction de réversion.

D'autre part, l'analyse différentielle des profils d'expression génique de lésions cutanées tuberculoïdes et lépromateuses a confirmé l'inversion de la balance cytokinique TH1/TH2 entre les deux types de lésions cutanées [269]. En effet, la polarisation immunitaire est attestée : les lésions cutanées et nerveuses de la lèpre tuberculoïde sont infiltrées par les cellules T dites « Th1 » (pour « T Helper type 1 »), qui produisent l'IFN- α , Tumor necrosis factor α (TNF- α), ainsi qu'*IL-2* et *IL-15*, qui sont des facteurs de croissance des cellules T. Les transcrits des interleukines 12 et 18, requises pour le développement des cellules Th1, sont abondants dans ces lésions. À l'opposé, les lésions de patients lépromateux contiennent des ARNm des cytokines « Th2 » (*IL-4* et *IL-10*). Certains patients lépromateux ne produisent pas de cellules T qui réagissent à la stimulation par *M. leprae*, ce qui suggère soit une inhibition de ce répertoire, soit la présence de lymphocytes T suppresseurs [270]. Or l'association d'*IL-2* et d'*IL-12* recombinantes restaure la réponse Th1 des lymphocytes chez certains individus lépromateux [271]. Ceci suggère la réversibilité d'un effet « inhibiteur ». Chez ces patients, la présentation par des cellules dendritiques de *M. leprae* permet de changer la production cytokinique des lymphocytes T d'un profil Th2 vers un profil Th1. De façon intéressante, l'injection intracutanée de IFN- γ ou d'*IL-2* améliore la réponse immune cellulaire locale mais ne modifie pas l'immunité systémique chez les patients lépromateux [272,273]. Cette absence de réponse des lymphocytes T est au cœur des recherches et parmi les hypothèses figurent : 1) la polarisation de la réponse des cellules T CD4+, 2) la suppression des lymphocytes T réactifs à *M. leprae*, 3) la présence de lymphocytes T régulateurs ou suppresseurs.

Des populations de lymphocytes T plus récemment décrites, les cellules T $\gamma\delta$ et T $\alpha\beta$ « doubles négatifs » (CD4-/CD8-) participent aussi à la réponse immunitaire. Ces cellules reconnaissent des antigènes non peptidiques, comme le lipoarabinomannane et l'acide mycolique, qui sont présentés via le récepteur CD1 par des cellules présentatrices d'antigène, indépendamment du Complexe Majeur d'Histocompatibilité (CMH) de classe I ou II. Le CD1 est fortement exprimé à la surface des cellules dendritiques au sein des granulomes cutanés des individus tuberculoïdes, mais il n'est pas retrouvé dans les lésions de patients lépromateux. Ceci suggère que la population T CD1+ contribue au contrôle de la bactérie [270]. De même, on retrouve dans les lésions cutanées des patients tuberculoïdes, mais aussi borderline, des lymphocytes « Th17 ». Ces lymphocytes T ne sont ni Th1 ni Th2, mais produisent entre autres l'*IL-17* [274,275]. Leur rôle dans la lèpre reste à préciser.

L'intérêt pour la réponse immunitaire adaptative ne doit pas faire oublier le rôle important de la réponse innée pour le contrôle de la bactérie. En effet, contrairement à ce qui était admis, l'immunité innée a certainement une responsabilité importante dans la polarisation de la réponse immune à *M. leprae*. Les cellules dendritiques, qui sont des cellules présentatrices d'antigène, participent aussi à l'activation des cellules T. Or, *M. leprae*, à la différence des autres mycobactéries, peut inhiber les cellules dendritiques. Le nombre de ces cellules est d'ailleurs diminué dans les lésions cutanées lépromateuses [276,277]. Le Toll-like receptor (TLR)-2 des monocytes et le système CD1 des cellules dendritiques, semblent y jouer des rôles intriqués. Les TLR sont une famille de récepteurs reconnaissant différents motifs microbiens invariants, qui provoquent l'activation de facteurs de transcription, tels que NF- κ B, induisant à leur tour la production de cytokines et de molécules de costimulation pour la réponse immune adaptative. Les molécules CD1, quant à elles, sont un système de présentation d'antigènes lipidiques (par exemple le lipoarabinomannan [278] dans le cas de *M. leprae*), et permettent l'activation de lymphocytes T CD1a restreints. Les monocytes et cellules dendritiques des lésions cutanées tuberculoïdes expriment plus fortement TLR-1 et -2, que ceux des lésions lépromateuses. L'activation de ces TLR par *M. leprae* induirait d'une part la production de cytokines de type TH1 [279] et d'autre part la différenciation des monocytes en cellules dendritiques CD1+ chez les patients TT. Chez les patients LL, au contraire, les cellules dendritiques CD1+, dont les cellules de Langerhans, sont particulièrement rares [278]. En revanche dans les lésions lépromateuses, un autre type de récepteur est abondamment exprimé par les macrophages : DC-SIGN, une lectine de type C associée aux réponses de type TH2, qui pourrait servir de récepteur d'entrée pour *M. leprae* via sa liaison au lipoarabinomannan [280].

Si ces observations éclairent en partie l'absence de réponse adaptative locale chez les patients lépromateux, l'origine des disparités de réponse immune entre patients est encore mal connue, et probablement le fruit d'une susceptibilité génétique individuelle.

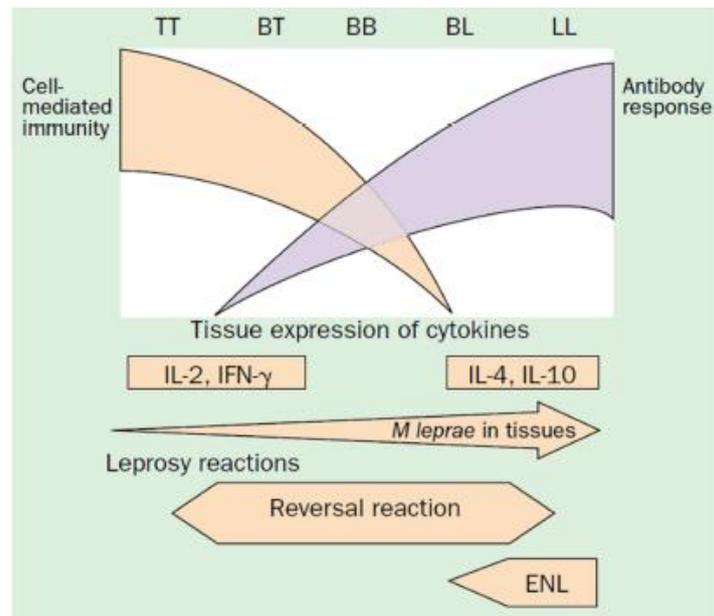


Figure 19 : Spectre immuno-clinique de la lèpre [209]

IL = Interleukine, IFN = interféron, ENL = érythème noueux lépreux.

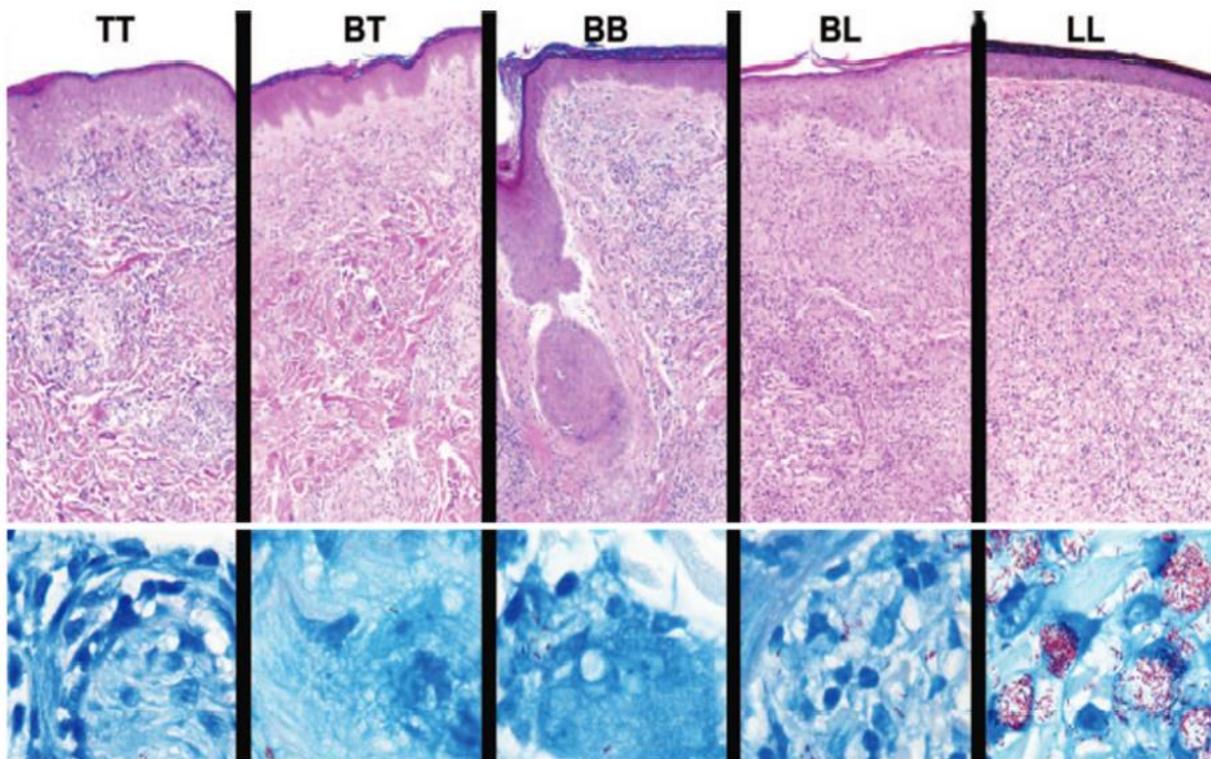


Figure 20 : Immunopathologie du spectre de la lèpre [61]

Chaque champ représente l'histopathologie d'un type de lèpre selon la classification de Ridley-Jopling :
 Panneau supérieur : coloration à l'hématoxyline et à l'éosine. Les granulomes épithélioïdes bien formés (TT) se désorganisent dans les tranches successives jusqu'à des agrégats d'histiocytes spumeux, avec seulement quelques lymphocytes (LL).

Panneau inférieur : coloration Fite-Faraco. 50 champs ont été nécessaires pour trouver les deux organismes visualisés dans un nerf cutané dans le champ TT. Le nombre d'organismes augmente à mesure qu'on s'approche du champ LL.

1.8 Génétique humaine de la lèpre

1.8.1 La lèpre : une maladie « héréditaire » ?

« I want to say that, if I formerly may have doubted the theory about the heredity of the disease, I have now no longer any doubts about it. » [281]

[« Si j'ai pu, par le passé, douter de la théorie selon laquelle la maladie est héréditaire, je n'ai maintenant plus aucun doute sur le sujet. »] Carl Wilhelm Boeck (1808–1875)

« ...your opinions about leprosy are completely wrong. You believe that the disease is hereditary but not infectious. The truth is that it is infectious but not hereditary. » [282]

[« ...vos convictions à propos de la lèpre sont totalement erronées. Vous croyez que la maladie est héréditaire et non infectieuse. La vérité est qu'elle est infectieuse et non pas héréditaire. »] Gerhard Henrik Armauer Hansen (1841–1912)

Ces deux citations pourraient résumer le débat qui a animé la communauté des « léprologistes » sur l'origine de la maladie au cours du XIXe siècle. En effet, jusqu'à la deuxième moitié de ce siècle, l'idée que la lèpre est une maladie héréditaire est fortement ancrée dans les esprits. Loin d'être seulement une croyance populaire associée au stigma social qu'entraîne la maladie, cette opinion est partagée et défendue par la communauté médicale. Ainsi, Danielssen et Boeck publient en 1847 le résultat d'une étude portant sur 213 patients et leurs liens de parenté, sous le nom Om Spedalskhed [De la lèpre] [283]. Le texte, qui affirme que la lèpre est une affection sanguine héréditaire, est alors considéré comme un ouvrage de référence par la communauté scientifique [281]. Dans ce contexte, la mise en évidence du bacille de la lèpre par Hansen en 1873 ravive un ancien débat entre partisans de l'hérédité et ceux d'une contagiosité de la maladie [282]. Nourri par les échecs répétés des expériences de Hansen pour mettre en évidence la contagiosité de la maladie, le débat perdure au cours des deux décennies suivantes [284,285] et ne prend fin qu'avec le premier congrès international de léprologie qui eut lieu à Berlin en 1897. Malgré l'absence de preuves expérimentales à cette époque et, en particulier, le fait que *M. leprae* ne satisfaisaient pas les postulats de Koch (l'impossibilité de cultiver in vitro le bacille de Hansen fait qu'il ne répond pas, même à l'heure actuelle, au deuxième postulat de Koch).

En outre, les 3^{ème} et 4^{ème} postulats n'ont pu être satisfaits expérimentalement avant 1960 et la découverte de la multiplication du bacille dans les coussinets de souris après inoculation [192]), les congressistes concluent :

« Leprosy Bacillus is the real cause of the disease. Leprosy is infectious but not hereditary »
[« Le bacille de la lèpre est la cause réelle de la maladie. La lèpre est de nature infectieuse et non héréditaire »] [282].

L'étiologie microbienne de la maladie étant reconnue, la lèpre est considérée pendant la première moitié du XX^{ème} siècle comme un exemple de maladie purement environnementale, la théorie, pourtant ancienne, d'une transmission héréditaire de la maladie fut totalement abandonnée, et il faudra attendre le début des années soixante pour qu'elle soit à nouveau considérée. Dès le milieu du siècle dernier, l'idée d'une prédisposition génétique au développement de la maladie ressurgit peu à peu. L'élément le plus ancien, et probablement le plus important, en faveur de cette participation génétique de l'hôte dans les pathologies infectieuses est la grande variabilité de la réponse observée chez des individus exposés au même agent pathogène : une certaine proportion ne sera jamais infectée ; parmi les sujets infectés, seule une partie présentera une expression clinique ; parmi les sujets développant la maladie, il y a une grande diversité dans les manifestations cliniques selon le type de réponse immunitaire spécifique de *M. leprae* développée par l'hôte. Cette diversité n'a à ce jour encore trouvé aucune explication épidémiologique ou microbiologique. On admet ainsi que moins de 5% des êtres humains vont développer la maladie après un contact prolongé avec *M. leprae* : ainsi, les études de cohortes vaccinales montrent que seule une minorité des personnes exposées à *M. leprae* développent une forme clinique de lèpre (probablement moins de 5%), même si un chiffre précis ne peut être avancé en raison de l'ignorance de la voie de transmission de la maladie [136,139,194]. À la fin du XIX^{ème} siècle, tous les essais (éthiquement discutables puisque le jour de leur exécution on proposait à des condamnés de se prêter à des expériences sur la lèpre) d'infection par le bacille lépreux se sont soldés par un échec. De façon instructive, la seule exception est un individu issu d'une famille incluant de nombreux malades.

Finalement, face à cette grande diversité des formes cliniques, l'idée d'une prédisposition génétique est maintenant bien ancrée. La variabilité de réponse suite à l'exposition à un agent pathogène résulte donc de la variabilité de quatre types de facteurs : les facteurs non génétiques (état nutritionnel, comorbidités, immunosuppression par exemple) et génétiques inhérents à l'hôte, les facteurs d'exposition et les facteurs microbiens (**figure 21**).

- La variabilité des facteurs microbiens est considérée comme négligeable au vu de la quasi-clonalité des souches de *M. leprae* à l'échelle du globe et, de fait, aucune différence de virulence n'a jamais été mise en évidence entre différentes souches de la mycobactérie [65]. Cela explique que les facteurs génétiques de l'hôte de susceptibilité à la lèpre qui sont identifiés dans une population donnée soient souvent retrouvés dans d'autres populations. Ceci est moins observé pour la tuberculose, dont l'agent causal, *M. tuberculosis* présente une grande diversité génomique à travers le globe, diversité qui influence le phénotype de la maladie [286,287]. Néanmoins, la découverte de *M. lepromatosis* pourrait nuancer cette position à condition non seulement que des arguments supplémentaires viennent étayer son implication dans la maladie, mais plus particulièrement que cette espèce soit effectivement associée à une forme particulière de lèpre et soit représentée de manière plus significative car elle n'a pas été identifiée dans l'ensemble des pays endémiques, dont l'Inde.
- Concernant les facteurs d'exposition, il a été observé que le risque de transmission après contact avec un patient atteint est plus élevé lorsque le patient index présente une forme MB, cela s'explique par le fait que ces patients sont porteurs de lésions riches en bacilles.
- De façon surprenante, les facteurs non génétiques relatifs à l'hôte semblent n'avoir été que peu explorés. Ainsi, une méta-analyse concernant les acteurs de risque de développement de la lèpre per se a été réalisée. Pour ce faire, les auteurs ont rassemblé tous les articles publiés entre 1940 et 2003 traitants des facteurs de risque de la lèpre. Or sur les 253 publications trouvées, seules deux s'intéressent aux facteurs socio-économiques et à l'état nutritionnel et, de plus, présentent des résultats contradictoires. De même, la recherche de facteurs de risque qui pourraient être constitués par des comorbidités n'est pas relevée par les auteurs [288]. À l'inverse, le rôle de

l'immunosuppression relative constituée par le SIDA dans le déclenchement de la lèpre a motivé de multiples études qui n'ont malheureusement pas permis d'apporter de réelles conclusions [142]. Ainsi, l'implication de facteurs environnementaux n'est pas réellement établie malgré les très nombreuses études ayant tenté de les identifier. Il s'agit d'un argument fort en faveur d'une composante génétique importante de l'hôte dans la variabilité interindividuelle de la réponse à *M. leprae*.

- Cela est confirmé par l'observation d'un nombre croissant de publications concernant la génétique de la lèpre depuis les années 1980, et dans le même temps d'une diminution significative des publications concernant son immunologie [289]. En effet, si pour les scientifiques du XIXe siècle les idées de maladie infectieuse et d'hérédité semblaient incompatibles, il est aujourd'hui reconnu qu'un certain nombre de facteurs de l'hôte, en particulier des facteurs génétiques, sont impliqués à plusieurs stades [290]. Un modèle génétique à deux niveaux a été proposé, différenciant deux groupes de gènes. Plus précisément, plusieurs niveaux de contrôle génétique sont envisageables, correspondant chacun à une étape physiopathologique de la lèpre. Certains facteurs génétiques pourraient influencer le passage de l'exposition à l'infection par *M. leprae* (hypothèse difficile à valider en l'absence de mesure fiable de l'exposition ni de l'infection), certains influenceraient la susceptibilité à la lèpre per se (c'est-à-dire au fait de développer une maladie lépreuse indépendamment de sa forme clinique), et d'autres la polarisation vers une forme clinique spécifique. Il est également envisageable que certains facteurs génétiques puissent directement influencer la survenue d'une forme ou l'autre de la maladie, sans nécessiter le stade pré symptomatique d'une lèpre per se (**figure 22**). Dans la suite de ce manuscrit, nous nous intéresserons exclusivement aux gènes du groupe 1, à savoir ceux impliqués dans la sensibilité (prédisposition ou protection) à la lèpre per se.

L'idée que la diversité de réponse de l'hôte humain à l'exposition à *M. leprae* ait une origine génétique est désormais étayée par de nombreuses études d'épidémiologie génétique et est considérée comme le facteur principal. Certains auteurs n'hésitent d'ailleurs pas à écrire que « la lèpre est peut-être la maladie infectieuse commune où la contribution génétique de l'hôte est la plus forte » [167,291]. Un biais possible est que la quasi-clonalité de *M. leprae* rend la reproductibilité des résultats plus facile que pour d'autres pathogènes. De manière générale, la démonstration d'une participation génétique de l'hôte dans la susceptibilité aux maladies infectieuses est maintenant bien établie, mais certains ouvrages présentent encore les maladies infectieuses comme un exemple du pôle « purement environnemental » sur le spectre d'intensité de la contribution génétique à une pathologie. La lèpre ne fait pas exception. Les modèles expérimentaux chez l'animal et les études d'épidémiologie génétique ayant participé à la dissection de ce déterminisme génétique complexe seront exposés ici.

Temps de contact avec un patient
lépromateux

Génome *quasi* clonal, pas de
différence de virulence entre
les différentes souches

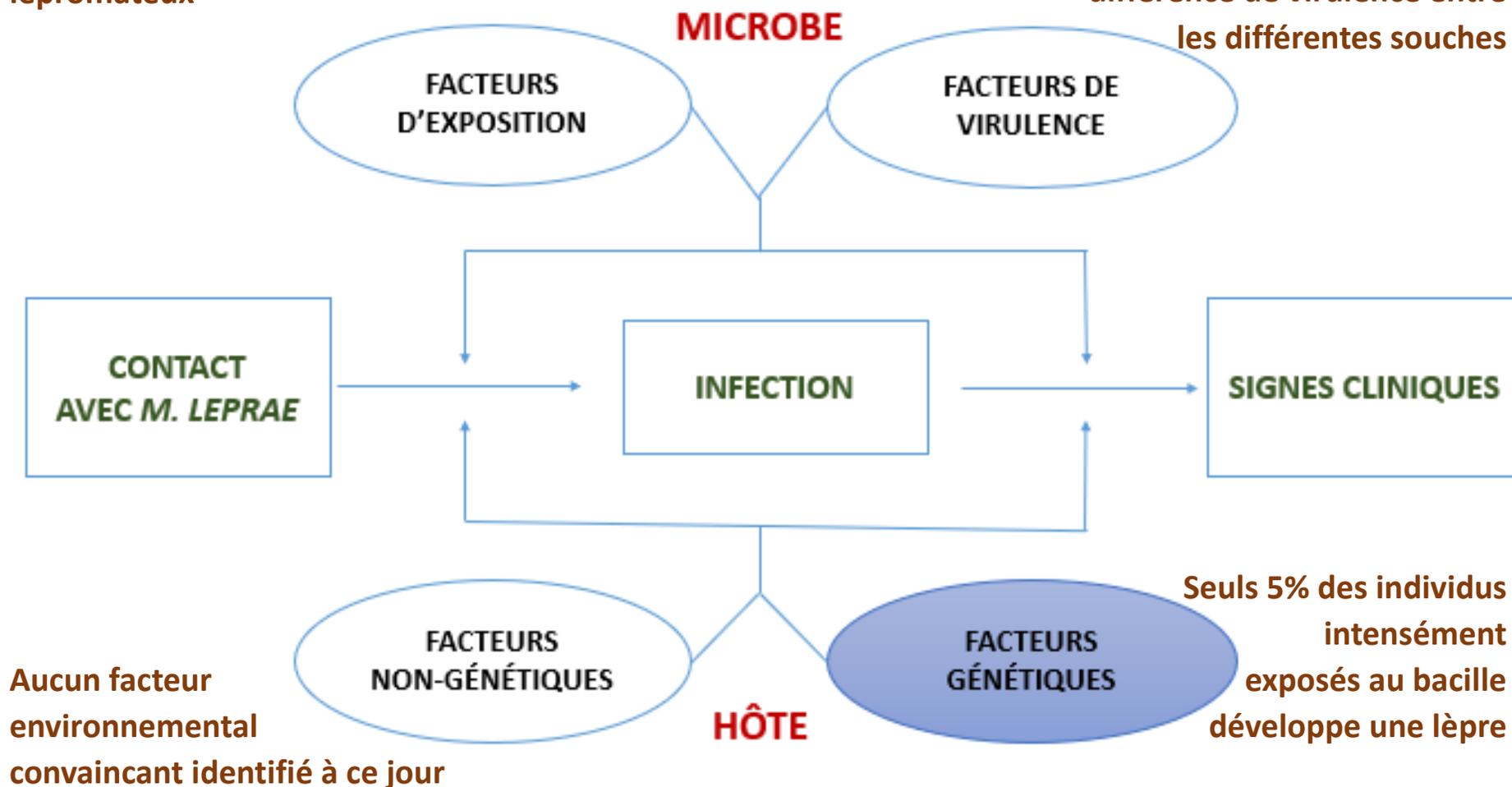


Figure 21 : Du contact avec *M. leprae* au développement de signes cliniques [292]

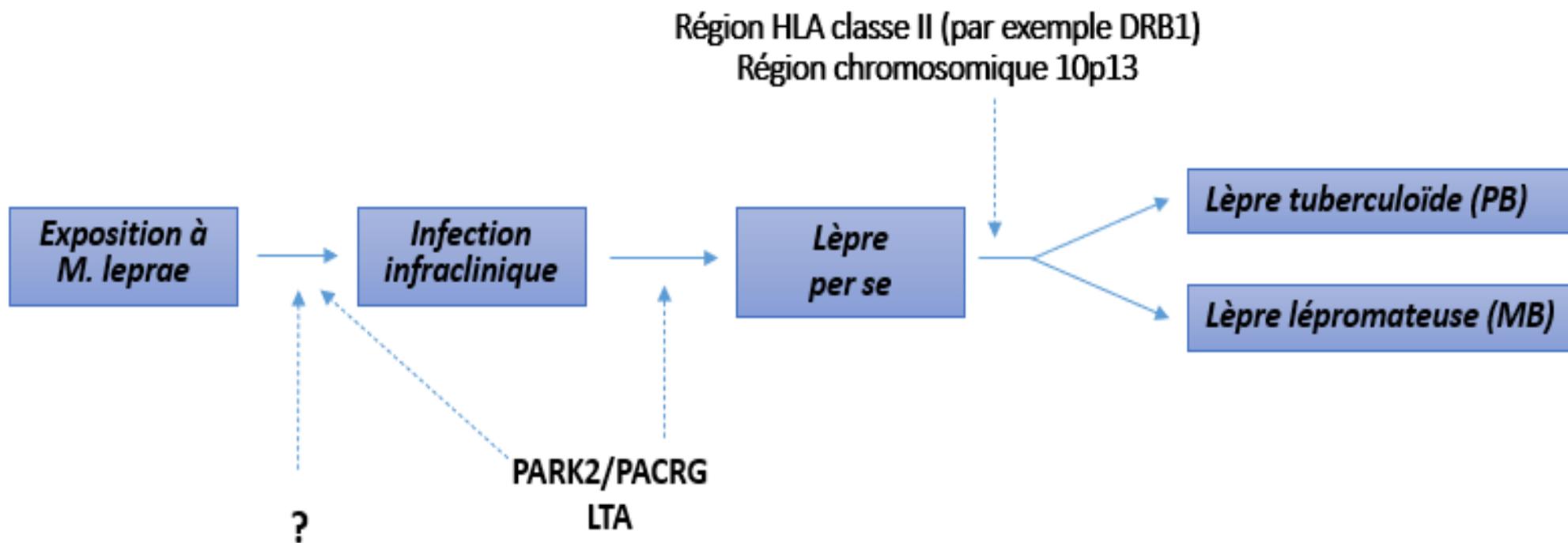


Figure 22 : Représentation schématique des facteurs de risque génétiques établis, conférant une susceptibilité accrue à la lèpre per se ou à certains sous-types de lèpre [293]

1.8.2 Modèles expérimentaux chez l'animal

Il y a plus de 45 ans, il fut établi que parmi certaines lignées de souris congéniques, un facteur génétique était impliqué dans la réponse précoce aux infections par différents agents pathogènes intracellulaires incluant *Salmonella typhimurium* [294,295], *Leishmania donovani* [296] et *Mycobacterium bovis* [297]. Les gènes responsables de la susceptibilité/résistance à ces agents pathogènes furent respectivement dénommés *Ity*, *Lsh*, et *Bcg*. Ces trois groupes réalisèrent rapidement qu'un gène commun était probablement en cause puisqu'ils furent indépendamment localisés sur la même région du chromosome 1 [295,298]. En 1982, deux papiers publiés simultanément confirment qu'il s'agit bien du même gène surnommé *Ity/Lsh/bcg*. En plus du résultat majeur qu'ils rapportent, ces papiers sont remarquables par la pertinence des prédictions émises. Ils présument 1) que ce gène intervient probablement dans le milieu intracellulaire du macrophage en affectant un composant biochimique ou nutritionnel nécessaire à la croissance du pathogène, 2) qu'il s'agit d'une mutation récente et non d'un variant d'un gène polymorphe, et 3) qu'un homologue humain du gène doit certainement réguler la susceptibilité à la tuberculose et à la lèpre. Vingt ans plus tard, l'ensemble de ces prédictions se sont révélées exactes. Rapidement ce gène est mis en cause dans de nombreuses autres infections murines et en particulier *M. lepraemurium*, l'équivalent murin de *M. leprae*. In vivo, il fut démontré que les macrophages des souris susceptibles (*Bcg*^S) n'étaient pas capable d'empêcher la multiplication de ces différents agents pathogènes [299–301].

Les auteurs employèrent ensuite le clonage positionnel (connu à l'époque sous le nom de « reverse genetics ») pour identifier le gène *Ity/Lsh/bcg* [302]. Il s'agit d'une procédure visant à identifier des gènes dont la fonction et le produit sont inconnus en se fondant sur leur localisation chromosomique et en effectuant une analyse moléculaire de la région ainsi identifiée. Une analyse de liaison génétique révéla la présence d'un marqueur étroitement lié au gène putatif et une carte génétique haute résolution de la région fut développée [303,304]. Une carte physique de la région identifia un contig de 400 kb contenant le gène en question et la méthode d'amplification d'exons décrite l'année précédente [305] identifia sept candidats potentiels. Seul l'un d'entre eux était exprimé exclusivement dans les cellules du système réticuloendothélial et, plus important, dans les macrophages dérivés. *Nramp1* venait d'être identifié [306]. Il est constitué de 15 exons et code pour une protéine de membrane de

100 kD dont une unique substitution non conservatrice d'une glycine vers un acide aspartique dans le 169^{ème} acide aminé (G169D) fut identifiée dans les lignées Bcg^S mais pas Bcg^R suggérant que *Nramp1* était bien *Bcg*. L'utilisation de marqueurs microsatellites adjacents établit que la région partagée par les souris Bcg^S s'étendait sur plus de 2 mégabases évocateur d'une mutation récente. Un an plus tard l'orthologue humain de *Nramp1*, noté par convention *NRAMP1*, était identifié sur le chromosome 2q35 [307].

L'identité entre *Nramp1* et *Ity/Lsh/Bcg* fut démontrée par la production d'une souris invalidée pour le gène *Nramp1*. Cette souris était susceptible aux infections par *BCG*, *L. donovani*, et *S. typhimurium*. De façon instructive ces souris, si elles ne peuvent pas limiter la prolifération bactérienne au stade initial de l'infection, sont capables de s'en débarrasser à un stade tardif. Finalement, de nombreux travaux ont essayé de déterminer la fonction exacte de *Nramp1* sans parvenir à des conclusions définitives. Il est probable que cette protéine agisse comme un transporteur de certains cations divalents comme le fer, le manganèse, le magnésium, le cuivre, le zinc ou encore le calcium qui sont métaboliquement essentiels pour la survie et la réplication microbienne dans le macrophage. En plus de réduire la survie du pathogène, les modifications dans la concentration intramacrophagique de ces cations entraînent une diminution des capacités de détoxification et de modulation de la fusion endosomale codés par le micro-organisme [308]. Cette hypothèse est séduisante puisqu'à la fois les microbes et l'hôte ont développé des systèmes sophistiqués et souvent concurrents pour importer ou éliminer ces cations et métaux du milieu intra macrophagique. En particulier, les 40% d'homologie entre les protéines *NRAMP1* de l'homme et de *M. leprae* [309] suggère que l'hôte et la mycobactérie pourraient être en compétition pour une substance commune essentielle pour la survie du micro-organisme.

1.8.3 Épidémiologie génétique

Les études d'épidémiologie génétique reposent sur l'utilisation conjointe d'informations de nature épidémiologique, telles que la mesure de facteurs de risque connus pour influencer l'expression de la maladie étudiée, et de nature génétique, comme les liens familiaux entre individus de l'échantillon ou encore le génotypage de marqueurs génétiques. Les deux sections ci-dessous rapportent les différentes observations ayant contribué à objectiver la présence d'un déterminisme génétique dans l'infection par *M. leprae*. On distingue les analyses n'incluant pas d'étude de marqueurs génétiques (aucune nécessité de génotypage de marqueurs génétiques) de celles qui l'incluent.

➤ Analyses sans marqueurs génétiques

Ce sont ces observations épidémiologiques qui attirent l'attention sur la possibilité d'un déterminisme génétique pour une maladie donnée.

- **Analyses d'agrégation familiale**

La nature familiale de la lèpre a été remarquée par de nombreux auteurs et même par Hansen qui cite cette observation, souvent utilisée par les partisans de la lèpre héréditaire, comme, au contraire, un argument de la contagiosité de l'infection [310]. Il considère, en effet et à juste titre, que l'exposition au bacille est plus importante dans l'entourage familial d'un malade excréteur. Cependant, l'observation d'une agrégation familiale ne correspond pas seulement à l'observation d'une prévalence de la maladie significativement plus élevée chez des apparentés à un cas que chez le reste de la population mais elle suppose aussi que le reste de la population est tout aussi exposée au pathogène que les membres de la famille du patient index. Toutefois, même en cas d'exposition homogène, une agrégation familiale des cas n'implique pas que la maladie est génétiquement déterminée, elle peut aussi résulter d'un environnement et/ou d'un mode de vie communs à risque.

Ainsi, une première approche consiste à comparer les taux de prévalence de la lèpre au sein de populations plus ou moins proches génétiquement. La codistribution de certaines données socio-économiques et du taux de prévalence de la lèpre dans 183 villages de la province de Bogia en Papouasie Nouvelle Guinée a été étudiée. Les auteurs ont trouvé que les taux de prévalence étaient corrélés avec la similarité linguistique entre les villages, interprétée comme une mesure des liens génétiques ancestraux [311]. Bien entendu, cette interprétation n'est valide que si l'on fait l'hypothèse d'une exposition homogène des villages à l'agent pathogène.

D'autre part, de nombreuses études d'agrégation familiales dans différentes populations ont rapporté une concentration familiale des cas de lèpre (une prévalence accrue de lèpre chez les individus apparentés à des patients lépreux) [312], mais là encore l'interprétation en termes de prédisposition génétique n'est envisageable qu'en admettant une intensité d'exposition comparable. Afin de surmonter ce problème, une étude réalisée en Papouasie-Nouvelle Guinée dans la province de Karimui où le mode de vie communautaire permet d'éliminer le biais potentiel d'une exposition familiale spécifique a été réalisée. Plus précisément, cette province est endémique pour la lèpre, l'organisation sociale particulière de la tribu est basée non pas sur une unité familiale mais sur une unité communautaire et la polygamie y est admise et commune si bien que les maris ne résident pas avec leurs femmes ni leurs enfants permettant, ainsi, pour des individus apparentés, de découpler les risques liés à l'hérédité et ceux liés à l'exposition. Il a été montré que les mères atteintes n'avaient pas plus de risque d'avoir des enfants atteints que les pères atteints en dépit du fait que la majeure partie de l'enfance soit passée avec la mère. Cette constatation est évocatrice d'une exposition homogène de la communauté. Dans ces conditions, le risque pour un enfant de développer la maladie était associé avec le risque génétique parental, défini entre autres par le nombre d'atteints dans la famille des parents. Le patron de répartition de la maladie observé par les auteurs reste donc familial. En outre, l'étude des 269 familles a montré que le risque de développer la maladie diminue de façon significative avec le nombre d'ascendants malades [313].

- **Analyses de jumeau**

Une agrégation familiale des cas n'implique pas que la maladie est génétiquement déterminée. C'est l'objectif des études de jumeaux de distinguer la contribution relative de facteurs génétiques d'autres facteurs familiaux pouvant participer à cette agrégation. Les jumeaux monozygotes (MZ) partagent le même patrimoine génétique. Comme des frères et sœurs classiques, les jumeaux dizygotes (DZ) partagent en moyenne la moitié de leur patrimoine génétique. On admet ordinairement que toute discordance phénotypique est d'origine environnementale chez des jumeaux MZ (car même information génétique mises à part les mutations somatiques et modifications épigénétiques considérées comme négligeables) et d'origine mixte (génétique et environnementale) chez les DZ. La comparaison des taux de concordances (taux de paires de jumeaux où les deux individus présentent le même phénotype) entre ces deux types de jumeaux doit donc permettre d'estimer la composante génétique. Les MZ doivent présenter des taux de concordance significativement plus élevés que des jumeaux dizygotes pour des phénotypes ayant une composante génétique forte. Au contraire, les taux de concordances de phénotypes environnementaux sont censés être équivalents. De fait, le taux de concordance de la lèpre chez les jumeaux MZ s'étend de 60 à 85% dans les trois études publiées tandis que chez les DZ ce taux est constamment bas autour de 15-20% [314–316]. Plus précisément, une première analyse en 1966 a étudié les phénotypes constitués, d'une part, par la lèpre *per se* et, d'autre part, par la forme clinique de lèpre chez 35 paires de jumeaux. Dix-neuf des 23 paires de MZ étaient atteintes de lèpre (soit un taux de concordance de 82,6%) contre 2 des 12 paires de DZ (correspondant à un taux de concordance de 16,7%).

De même, concernant la forme clinique développée, 17 des 19 paires de MZ malades souffraient du même type de lèpre (taux de concordance de 89,5%) contre 0 des 2 paires de DZ [315]. Ces résultats ont été partiellement confirmés en 1973 avec des effectifs plus importants. Ainsi, en comparant 62 paires de MZ et 40 paires de DZ, les auteurs ont observé respectivement 37 et 8 paires concordantes pour la lèpre *per se* soit des taux de concordance de 59,7%. Concernant le type de lèpre MB développé 32 des 37 paires monozygotes (86,5%) et 6 des 8 paires dizygotes (75%) étaient concordantes [316]. À l'inverse, pour la forme PB, les taux de concordance entre jumeaux MZ et DZ étaient très proches. Comme précédemment évoqué, il existe une plus grande proportion de MB chez les garçons [317].

Ce biais contribue sans doute marginalement à l'augmentation du taux de concordance chez les jumeaux MZ vs. DZ (qui peuvent être de sexe opposé). Ainsi, ces études apportent des arguments en faveur de l'existence d'une prédisposition génétique, non seulement, dans le développement de la lèpre *per se*, mais également, dans la polarisation de maladie.

Néanmoins, ces études, toutes anciennes, présentent une sérieuse limitation : l'établissement de la zygote n'était pas simple et en cas de doute la paire était écartée, ce qui conduit naturellement à un excès de paires DZ composées d'un garçon et d'une fille. Dans la mesure où l'on a observé un excès de lèpre chez les garçons, il est clair que cet écueil va artificiellement diminuer le taux de concordance chez les DZ. Enfin, ce type d'approche fait l'hypothèse d'un environnement identique pré et post natal chez les deux types de jumeaux et ignore donc *de facto* l'influence de l'environnement fœtal. Quant à la similarité de l'environnement post natal on peut citer cet exemple (librement) traduit de Penrose [318] : « Imaginons qu'on demande à un étudiant en hérédité humaine fraîchement débarqué d'une autre planète d'utiliser la méthode des jumeaux afin d'étudier si la tenue vestimentaire est génétiquement déterminée. Il observerait que les jumeaux MZ sont habillés de la même façon, souvent jusque dans les moindres détails, tandis que c'est beaucoup plus inhabituel chez les DZ. Il conclurait en toute confiance que le choix des vêtements est quasi exclusivement sous l'influence de facteurs génétiques ». Néanmoins, utilisée avec précaution, l'étude des jumeaux demeure un outil appréciable en génétique humaine.

- **Analyses de ségrégation**

L'analyse de ségrégation est une méthode statistique qui cherche à déterminer l'origine des corrélations familiales observées et particulièrement à isoler la présence d'un gène unique (gène majeur), parmi l'ensemble des facteurs génétiques et environnementaux intervenant dans le déterminisme du phénotype considéré ; le terme gène majeur ne signifie pas qu'il s'agit du seul gène intervenant dans le phénotype étudié mais que, parmi l'ensemble des gènes impliqués il en existe un dont l'effet est suffisamment important pour être distingué des autres.

Lorsque le phénotype d'intérêt est binaire, cette méthode repose sur une modélisation mathématique de la probabilité d'observer le phénotype d'un individu en fonction des différents facteurs génétiques et/ou environnementaux, des corrélations familiales pouvant influencer ce phénotype en testant plusieurs modèles de transmission.

Différentes analyses de ségrégation ont été effectuées et ont mis en évidence une forte composante familiale influençant la lèpre *per se*, mais les modèles génétiques obtenus n'étaient pas toujours reproductibles d'une étude à l'autre (**tableau 4**). Tout d'abord, une étude a analysé 91 familles philippines incluant au moins un enfant lépromateux. Leurs résultats sont en accord avec l'hypothèse d'un gène autosomique récessif conférant une susceptibilité à la lèpre lépromateuse, mais les auteurs précisent qu'un modèle multigénique semble plus vraisemblable rejoignant les conclusions auxquelles était parvenue une seconde équipe en Papouasie Nouvelle Guinée [311,319]. Si la ségrégation d'un gène majeur récessif avec la forme TB a été observée dans différentes populations et avec différentes méthodes d'analyses [320–323], les études sur la lèpre *per se* ont rapporté des résultats beaucoup plus contrastés. Un mode de transmission récessif était le plus vraisemblable dans une population indienne [321] tandis qu'un modèle additif quasi dominant était suggéré par l'analyse d'un échantillon de 72 familles thaïlandaises [323]. Une analyse de ségrégation complexe sur 27 familles multigénérationnelles originaires de l'île antillaise de la Désirade a mis en évidence un gène majeur récessif contrôlant la susceptibilité à la lèpre *per se* [322]. La fréquence de l'allèle de susceptibilité était estimée à 30% soit 9% de sujets homozygotes prédisposés à la maladie. Sous cette hypothèse, la probabilité pour un sujet susceptible de développer la maladie avant l'âge de 60 ans était de l'ordre de 60%. Enfin, la proportion de patients lépreux non porteurs du génotype de prédisposition ne représentait que 23% de l'ensemble des cas de lèpre dans cette population. Dans une vaste étude incluant 10 886 individus distribués dans 1 568 familles brésiliennes, les données étaient en faveur d'un gène majeur récessif avec une fréquence de l'allèle de susceptibilité aux alentours de 5%, proche de l'étude précédente. Cependant, une nette déviation par rapport aux taux attendus de transmission mendélienne des allèles au locus majeur ne permet pas de conclure clairement [324].

Enfin, à l'aide d'un modèle de ségrégation dérivé des méthodes d'analyse de la survie et permettant de prendre en compte l'âge de survenue de la maladie, un gène majeur codominant prédisposant à la lèpre per se dans un échantillon de familles vietnamiennes a été mis en évidence [325]. À l'inverse, aucun gène n'était retrouvé dans les familles chinoises qui faisaient partie de l'échantillon initial de cette étude.

En définitive, si aucune des études précédentes n'est individuellement suffisante pour établir de manière univoque le rôle de facteurs génétiques dans la susceptibilité à la lèpre ou à sa polarisation et leur mode de transmission, il faut admettre que leur combinaison est clairement en faveur de cette hypothèse et les auteurs concluent tous à l'existence d'une composante génétique portée par un ou plusieurs gènes à effet majeur.

Par rapport au modèle mendélien strict, les gènes à effet majeur présentent une pénétrance plus faible, le phénotype étant alors le résultat de l'expression d'autres gènes, de facteurs environnementaux et de leurs interactions [290]. Ainsi, dans le spectre des prédispositions aux maladies infectieuses, la sensibilité à la lèpre constitue un exemple de modèle complexe avec un nombre réduit de gènes à effet majeur, à mi-chemin entre les modèles mendéliens monogéniques à pénétrance complète (résistance à la malaria causée par *Plasmodium vivax*) et les modèles complexes polygéniques (infections associées à des polymorphismes au sein du complexe majeur d'histocompatibilité (CMH)) (**figure 23**).

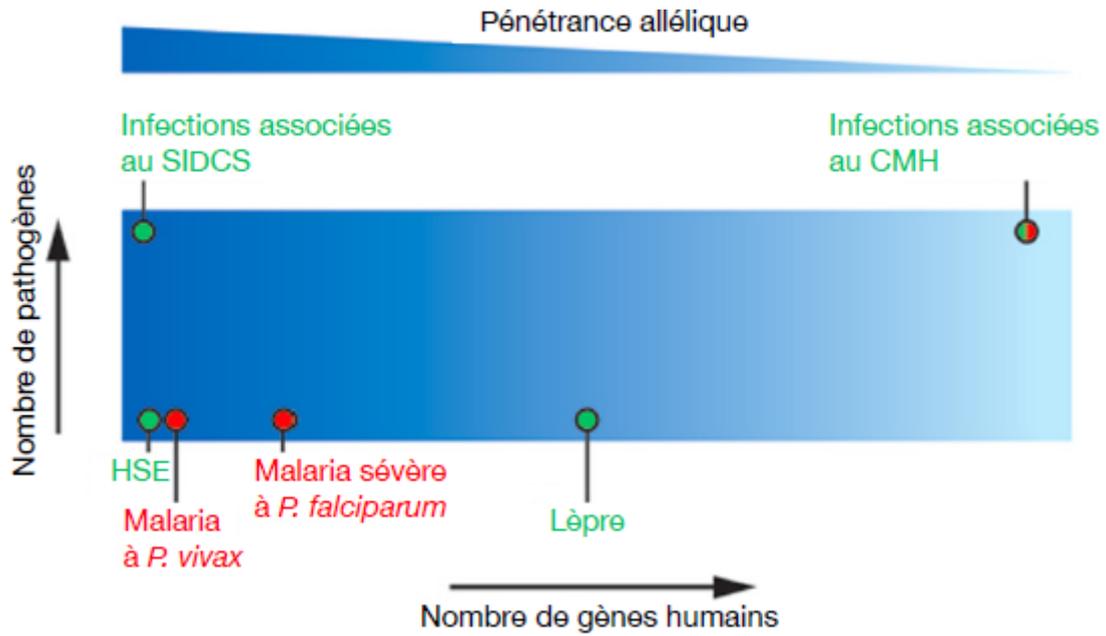


Figure 23 : Prédiposition génétiques aux maladies infectieuses en fonction du nombre de gènes impliqués, de la pénétrance allélique et du nombre de pathogènes [290]

En rouge : résistance,

En vert : sensibilité,

SIDCS : Syndrome d'Immunodéficience Combinée Sévère,

HSE : Herpes Simplex Encephalitis (Encéphalite Herpétique).

Référence	Population (année)	Échantillon	Conclusion
[319]	Philippines (1979)	91 familles nucléaires	Lèpre lépromateuse : modèle récessif Lèpre per se : modèle multigénique
[311]	Papouasie Nouvelle Guinée (1979)	340 proposants et apparentés 1 ^{er} degré	Lèpre per se : modèle multigénique
[321]	Inde (1985)	72 familles nucléaires	Lèpre tuberculoïde : modèle récessif
[320]	Ile de la Désirade (1985)	16 familles étendues	Lèpre per se : modèle multigénique Lèpre tuberculoïde : modèle récessif
[323]	Thaïlande (1988)	63 familles nucléaires	Lèpre per se : modèle additif Lèpre tuberculoïde : modèle récessif
[322]	Ile de la Désirade (1988)	27 familles étendues	Lèpre per se : modèle récessif Lèpre tuberculoïde : modèle récessif
[325]	Vietnam (1995)	402 familles nucléaires	Lèpre per se : modèle codominant
[324]	Brésil (1995)	1 568 familles	Lèpre per se : modèle récessif

Tableau 4 : Principales analyses de ségrégation sur la lèpre

Depuis, la majorité des études a consisté en l'analyse de marqueurs génétiques, selon différentes stratégies pour identifier les gènes de susceptibilité (**figure 24**). Plusieurs approches sont disponibles pour clarifier les bases moléculaires d'une prédisposition mendélienne aux maladies rares. L'analyse de liaison génétique est habituellement la première étape du clonage positionnel, bien que l'identification d'une anomalie cytogénétique puisse être très utile. La stratégie gène candidat nécessite de sélectionner a priori un certain nombre de gènes (du fait de leur fonction, par homologie avec les modèles animaux...) dont la fonction sera évaluée à l'aide d'essais adaptés et qui pourront être séquencés à la recherche d'une mutation. Pour évaluer la base moléculaire d'une prédisposition complexe aux maladies plus communes comme la lèpre, il est possible d'utiliser la stratégie gène candidat décrite plus haut ou de générer de nouvelles hypothèses de travail par des criblages du génome. Une autre approche potentiellement fructueuse consiste à comparer l'expression de différents gènes dans les tissus d'individus atteints et sains. Le rôle de polymorphismes au sein des gènes candidats définis a priori ou identifiés « par expérience » (c'est-à-dire localisés dans une région identifiée par un criblage du génome) est testé par des études d'association qui peuvent être de type cas/témoins ou familiale. Enfin, les évidences statistiques de l'épidémiologie génétique, aussi fortes soient elles, doivent être validées par des études fonctionnelles.

Dès lors que la prédisposition génétique à la lèpre est complexe et non mendélienne, il ne s'agit plus de montrer un lien de causalité à l'échelle cellulaire entre un polymorphisme et un phénotype mais bien de prendre en compte plusieurs polymorphismes, leurs interactions avec l'agent pathogène et les effets environnementaux qui participent tous à l'expression de la maladie. Cette évaluation se réalise à l'échelle d'une population. Les outils nécessaires, ceux de l'épidémiologie génétique, sont donc statistiques.

Dans la lèpre, de nombreuses études d'épidémiologie génétique ont objectivé l'implication de facteurs génétiques de l'hôte dans la susceptibilité à la maladie. Cette prédisposition génétique à la lèpre a été principalement analysée par le biais de deux types d'études (analyses de liaison ou d'association) et à l'aide de deux stratégies de sélection des marqueurs génétiques (gène candidat ou pangénomique).

➤ Analyses avec marqueurs génétiques

Ce type d'analyse permettent de spécifier les régions chromosomiques, les gènes voire les variants alléliques impliqués dans le déterminisme de la maladie. Au sein de ces approches nécessitant le typage de marqueurs génétiques, on distingue plusieurs types de marqueurs génétiques qui sont des polymorphismes génétiques : courtes régions d'ADN qui peuvent varier d'un individu à l'autre dans une population saine. Les plus utilisés en épidémiologie génétique sont les microsatellites ou les polymorphismes mono-nucléotidiques (SNP pour « single nucleotide polymorphism ») (**tableau 5**). Dans ce manuscrit, le terme variant réfère aux SNPs, polymorphismes bi-alléliques de l'ADN qui diffère par une seule paire de bases et représentent la variation la plus abondante dans le génome humain. De plus, ils sont stables et uniformément distribués le long du génome. Ces variations sont associées à la diversité entre populations ou individus, comme une différence de sensibilité à des maladies ou encore la réponse aux médicaments. Dans les différentes stratégies d'analyse possibles dans la cadre de la lèpre comme de toute autre maladie commune, dite « à hérédité complexe », et nécessitant le typage de marqueurs génétiques, on distingue très schématiquement les analyses de liaison génétique et les études d'association (**tableau 6**).

Type	Nature du changement nucléotidique	Nombre d'allèles	Localisation	Nombre sur le génome	Fonction	Utilisation
Micro-satellite	N répétitions de (groupes de) bases → ex : (CAA) _n	4 à 20	Généralement extra-génique	Quelques milliers	Inconnue	Analyses de liaison
SNP	Remplacement d'une base par une autre → ex : C devient T	2	Extra ou intra-génique	5-10 millions	Neutre ou fonctionnel	Analyses de liaison ou d'association

Tableau 5 : Marqueurs génétiques utilisés en épidémiologie génétique

Après avoir détaillé les principaux gènes candidats qui ont été investigués par des analyses de liaison ou des études d'association, nous verrons les études par criblage du génome réalisées dans la lèpre. Les résultats les plus convaincants sont résumés dans le **tableau 7 et la figure 25**. Ces parties n'ont pas pour objectif de proposer une revue exhaustive des gènes associés à la sensibilité à la lèpre mais seulement de fournir des exemples, d'une part, des types de gènes impliqués dans cette sensibilité et, d'autre part, de souligner que les types de gènes identifiés par ces deux approches peuvent différer.

- **Analyse de liaison**

L'analyse de liaison vise à identifier et localiser sur le génome des régions chromosomiques contenant un ou plusieurs gène(s) d'intérêt (englobant les notions de prédisposition mais, également, de protection vis-à-vis du développement de la lèpre.) Elle étudie la co-ségrégation d'un locus génétique à risque pour le phénotype étudié (que l'on veut localiser) et de marqueurs génétiques (dont on connaît la localisation sur le génome). La co-ségrégation est la tendance, pour deux loci situés sur le même chromosome et proches physiquement, à ne pas être séparés lors de la méiose (faible taux de recombinaison) et donc à être hérités ensemble d'une génération à l'autre. L'analyse nécessite donc obligatoirement un échantillon familial, avec au moins deux générations disponibles. En pratique, il est testé pour chaque marqueur si deux frères/sœurs malades partagent plus de 50% d'allèles hérités en commun de leurs parents. Elle peut être réalisée grâce à des méthodes paramétriques (qui nécessitent de spécifier un modèle génétique pour pouvoir inférer le génotype au locus causal à partir des phénotypes observés) ou non paramétriques. L'analyse de liaison permet soit d'étudier un nombre restreint de régions candidates, soit de cribler le génome entier. Le principal attrait de la stratégie de criblage entier est de s'assurer que tous les loci majeurs impliqués dans la genèse d'un phénotype soient identifiés, et de fournir l'opportunité de détecter de nouveaux gènes majeurs (et donc de nouvelles voies physiopathologiques) dont le rôle n'était pas soupçonné a priori [326]. Malheureusement, dans le contexte d'un phénotype infectieux complexe comme la lèpre, on ne peut pas espérer une localisation fine des gènes d'intérêt mais plutôt une région dont la taille restera importante de l'ordre de 10 à 15 centimorgans (10 à 15 millions de bases).

Au départ, le plus souvent les études focalisaient sur des gènes candidats (candidats par hypothèse ou par expérience), mais il est également possible de réaliser un criblage complet du génome (génération d'hypothèses).

Approche gène candidat

Les études de liaison génétique portant sur une région de taille limitée, structurée autour d'un gène d'intérêt particulier du fait de sa fonction (gène candidat par hypothèse) ont été les premières à utiliser l'information génétique issue de l'ADN des individus échantillonnés. En raison de leur importance dans le déclenchement de la réponse immunitaire spécifique, les antigènes de la région HLA, discriminables par des techniques de sérologie bien avant l'avènement des techniques de génotypage, ont été les premiers polymorphismes à être étudiés.

- **Le système HLA :** Dans la mesure où les molécules du complexe majeur d'histocompatibilité de classe (CMH) I et II sont hautement polymorphiques et situés dans un intervalle réduit (3,6 mégabases) du chromosome 6. Ils jouent un rôle important en présentant les peptides antigéniques respectivement aux cellules T auxiliaires et cytotoxiques au cours de la réponse immune cellulaire, leurs gènes ont été extensivement étudiés dans la lèpre mais les résultats sont controversés. Les études de liaison génétique modèle-indépendantes ont apporté des arguments supplémentaires en faveur d'une implication de HLA dans les formes polaires de lèpre. Une ségrégation non aléatoire des haplotypes parentaux a été observée chez des enfants présentant la forme tuberculoïde originaires du Surinam [327], du Sud et du centre de l'Inde [328,329], du Venezuela [330], et d'Égypte [331]. Des résultats similaires furent obtenus chez des enfants lépromateux au Venezuela et en Chine. En revanche, les haplotypes HLA ségrégaient aléatoirement chez des enfants atteints [328] et sains [330] dans des familles multiplex du sud de l'Inde suggérant un rôle modeste du système HLA sur la lèpre per se [332]. Cette observation est consistante avec l'hypothèse d'un complexe majeur d'histocompatibilité intervenant à un stade relativement tardif de l'infection, l'immunité cellulaire antigène-spécifique régulant la polarisation de la lèpre vers une forme tuberculoïde ou lépromateuse.

- **SCL11A1 (anciennement NRAMP1)** : Ce gène humain est un excellent gène candidat dans la mesure où il s'agit de l'orthologue humain [307] du gène murin NRAMP1 (Natural Resistance Associated Macrophage Protein 1) et qui fut le premier gène de susceptibilité aux infections mycobactériennes décrit chez la souris [306]. Plus précisément, chez la souris les protéines Nramp influencent la viabilité des agents pathogènes à l'intérieur des macrophages en transportant du fer et autres cations divalents à travers la membrane des phagosomes, et ont également des effets sur l'activation des macrophages et la régulation de la balance TH1/TH2. Une étude de paires de germains atteints au Vietnam a trouvé une liaison significative entre la lèpre *per se* et la région de ce gène [333]. Ce résultat est en accord avec l'analyse de ségrégation réalisée dans la même population, néanmoins, cette étude suggère également une possible hétérogénéité génétique selon l'origine ethnique des familles (vietnamienne vs. chinoise). Cette hétérogénéité génétique pourrait expliquer au moins en partie l'absence de liaison génétique entre la lèpre et SCL11A1 observée dans des échantillons de familles brésiliennes et pakistanaises [334], polynésienne [335]. Aucune association entre trois polymorphismes de SCL11A1 et la lèpre *per se* n'a été retrouvée en Thaïlande [336].
- Les études sur un échantillon Brésilien ont rapporté une liaison de la lèpre *per se* avec les régions 17q11-q21 [337], 17q22 et 20p13 [338]. Ces études ont défini ultérieurement deux groupes T* et L* : le groupe T* rassemble les malades classés TT, BT et BB selon la classification de Ridley-Jopling, et le groupe L* les malades BL et LL. En utilisant cette classification inhabituelle (les individus BB sont généralement classés avec les BL et LL dans le groupe MB), la liaison observée avec les régions 17q11 et 20p13 semblait portée par le groupe L*, car on ne la retrouvait pas dans le groupe T*, pourtant de taille équivalente (109 et 93 enfants, respectivement).
- Si la grande majorité de ces études a porté sur la lèpre *per se*, une étude de liaison s'est intéressée à la polarisation de la lèpre telle que définie par Ridley et Jopling. Elle comprenait 118 germains (c'est-à-dire frères et/ou sœurs) distribués dans 20 familles vietnamiennes et a analysé six régions d'intérêt : 1p36 autour du gène Tumor Necrosis Factor Receptor 2 (*TNFR2*), 5q31-33 autour d'un groupe de cytokines, 6p21 autour de *TNF*, 10p13 rapportée dans une étude préalable [339], 12p13 autour de *TNFR1* et 12q13 autour du récepteur de la vitamine D (*VDR*). La polarisation a été codée comme un phénotype

ordinal, donnant un poids croissant selon la classe d'appartenance (TT < BT < BB < BL < LL). Elle a mis en évidence une liaison significative avec la région 6p21 autour du gène *TNF* [340]. La liaison était moins forte pour le phénotype « lèpre *per se* », suggérant un contrôle génétique spécifique de la polarisation.

- Enfin, une étude de liaison au Malawi a identifié un signal significatif entre la polarisation et des marqueurs génétiques de la région 21q22 [341]. Toutefois, comme pour la plupart des régions identifiées par les études de liaison génétique, les explorations fonctionnelles nécessaires pour déterminer le mécanisme d'implication du ou des gènes situés dans ces régions, si elles ont été réalisées, n'ont pas été publiées.

Approche génome entier

Alors que les analyses de gènes candidats par hypothèse se sont avérées peu fructueuses dans le cadre de la lèpre *per se*, de nouveaux axes de recherche ont pu être générés grâce à l'exploration du génome complet par analyse de liaison.

La première étude de liaison pangénomique et modèle-indépendante de la lèpre, a été réalisée sur un échantillon familial indien et publié en 2001. L'échantillon rassemblait 224 familles incluant 245 paires de germains atteints par la lèpre et originaires du sud de l'Inde [339]. Les patients ont été diagnostiqués selon les directives de l'OMS. À l'exception de quatre patients, l'ensemble des enfants présentait la forme PB de la maladie. Une liaison significative ($p < 2.10^{-5}$) a été retrouvée sur le chromosome 10p13 suggérant qu'un gène localisé dans cette région pourrait influencer de façon substantielle la survenue d'une lèpre PB. En effet, en raison de la très forte proportion de cas de PB (98%), il n'a pas été possible de distinguer si cette région influençait la lèpre *per se* ou spécifiquement la forme PB, ces deux phénotypes étant indissociables chez les individus inclus dans l'étude. Une étude ultérieure réalisée au Vietnam supporte plutôt l'hypothèse d'un impact spécifique sur la forme PB. Elle incluait un nombre équilibré de PB et de MB et le signal de liaison en 10p13 n'a été retrouvé que dans le sous-groupe PB [342]. Le test formel d'une hétérogénéité entre PB et MB quant à l'impact de la région 10p13 était significatif ($p = 0,03$).

Une seconde étude indienne, incluant très sensiblement les mêmes familles que la première (243 familles au total, dont 233 avec des enfants PB exclusivement) a identifié un second signal de liaison dans la région 20p12 [343]. Dans l'ensemble, les résultats des études de liaison de la lèpre chez les patients indiens et vietnamiens semblent donc indiquer que les régions 10p13 et 20p12 sont impliquées dans le phénomène de polarisation. De façon surprenante, aucune liaison génétique ne fut retrouvée avec la région HLA alors que plusieurs études préalables gènes candidats, y compris dans une population du sud de l'Inde, avaient indiqué que des gènes de cette région jouaient très probablement un rôle dans les formes lépreuses tuberculoïdes paucibacillaires [328].

Comme déjà indiqué, un second criblage du génome a été entrepris sur un échantillon de familles vietnamiennes incluant plusieurs enfants porteurs de lèpre PB ou MB [342] :

- **PARK2/PACRG** : Ce criblage a mis en évidence une liaison génétique très significative ($p < 10^{-6}$) entre la lèpre per se et la région chromosomique 6q25. La région a été saturée en SNPs et une association a été retrouvée avec des SNPs localisés dans la région régulatrice commune au gène PARK2, dont des mutations sont responsables de certaines formes juvéniles de maladie de Parkinson, et au gène corégulé PACRG [16]. En particulier, la combinaison de deux allèles à risque sur deux SNPs significatifs était associée à un risque cinq fois plus élevé de développer la maladie. Ce résultat a été confirmé dans une population brésilienne [16], en revanche aucun des SNPs à risque dans les populations brésilienne et vietnamienne n'était significativement associé à la lèpre dans une étude cas-témoins en Inde [344,345]. Ce résultat négatif peut s'expliquer soit par une hétérogénéité ethnique dans le déterminisme génétique de la lèpre, comme cela a déjà été suggéré pour d'autres régions chromosomiques [333,343], soit par un profil de déséquilibre de liaison différent entre les populations indienne et vietnamienne ou brésilienne, qui nécessiterait de génotyper la région de façon plus extensive dans la population indienne avant de pouvoir conclure à l'absence d'implication de la région PARK2/PACRG (comme cela a été nécessaire pour la réplique de l'effet du gène LTA [14], décrit plus bas). Sur le plan fonctionnel, PARK2 code pour PARKIN, une protéine impliquée dans l'une des cascades de dégradation des protéines cellulaires, la cascade d'ubiquitination protéolyse, qui est un mécanisme général de régulation des protéines, dont celles impliquées dans les réponses

immunitaires. Des études récentes ont ainsi montré que plusieurs E3 ubiquitine-ligases interagissaient avec des protéines impliquées dans l'induction de l'anergie par les cellules T ou dans la régulation de la signalisation par les TLRs [346].

- **Lymphotoxine- α** : Le même criblage du génome mettait également en évidence un signal de liaison ($p < 10^{-5}$) au niveau de la région chromosomique 6p21 (région HLA). Après densification de la région par des SNPs supplémentaires, une association a été retrouvée avec des marqueurs situés dans le gène de la LTA (lymphotoxine- α , situé dans la région HLA de classe III). L'association avec l'un de ces SNPs (LTA+80) a été répliquée dans des familles indépendantes au Vietnam, ainsi que dans des populations non apparentées au Brésil et en Inde. Précisons que ce marqueur n'était pas en déséquilibre de liaison avec les régions HLA de classe I et II. De façon intéressante, cette association n'était significative que chez les individus de moins de 15 ans, avec un risque cinq fois plus élevé de développer la lèpre lorsque le sujet est porteur d'au moins un allèle A (effet dominant) [14]. Dans la littérature, l'allèle A de LTA+80 est associé à de bas niveaux de production de LTA [347] et les souris invalidées pour la LTA présentent une susceptibilité accrue aux infections mycobactériennes [348]. Ces données renforcent donc la plausibilité du rôle direct de LTA+80 dans la susceptibilité à la lèpre. Ce marqueur sera utilisé comme exemple dans notre population lors de la réplique cas-contrôles parentale.

En résumé, les analyses de liaison gènes candidats ont ainsi permis d'identifier huit régions d'intérêt pour la lèpre et/ou sa polarisation, situées sur les chromosomes 2, 6, 10, 17, 20 et 21 (**figure 25**).

- **Analyse d'association**

Il s'agit de l'étape suivante qui permet alors d'évaluer le rôle potentiel de polymorphismes au sein des gènes candidats localisés dans la région en question. Leurs buts sont 1) de localiser plus précisément la région contenant les gènes à risque pour le phénotype en étudiant des marqueurs supplémentaires dont on espère que l'un d'eux sera en déséquilibre de liaison avec le gène à identifier, 2) au sein d'un gène particulier, de mettre en évidence un variant génétique responsable d'un risque accru de maladie ou en déséquilibre de liaison avec le variant génétique réellement fonctionnel, et d'estimer l'importance de son effet. Il existe un déséquilibre de liaison entre deux locus lorsqu'il existe à la fois une liaison génétique forte et une association, c'est-à-dire que non seulement les loci co-ségrègent d'une génération à l'autre au sein de la famille en raison de leur grande proximité physique mais aussi leurs allèles s'associent préférentiellement (alors que lors d'une liaison génétique les allèles liés au locus maladie peuvent différer selon les familles). Ainsi, les études d'association ont, pour but principal d'identifier, non plus des portions du génome, mais des variants causaux ou en LD avec. La méthode classique est de type cas-témoin et consiste à comparer la distribution des allèles (ou génotypes) d'un polymorphisme donné chez des cas et des témoins non apparentés. Afin de surmonter le délicat problème du choix des témoins et de limiter les associations factices causées par des structures particulières de population, des méthodes dites intrafamiliales « family-based » se sont développées comme le Transmission Disequilibrium Test (TDT) qui teste si les enfants atteints ont reçu de leurs parents hétérozygotes pour un allèle donné, une fréquence significativement différente de ce que voudrait le hasard (une sur/sous-transmission) [349]. Cette méthode sera détaillée dans le chapitre 3.

Au départ, le plus souvent les études d'associations se focalisaient sur des gènes candidats (candidats par hypothèse ou par expérience, c'est-à-dire mis en évidence par l'analyse de liaison et seuls les variants correspondants sont alors testés), mais il est maintenant très courant de réaliser un criblage complet du génome en utilisant ces méthodes et c'est le cœur même de cette thèse [350].

Contrairement à l'approche gène candidat, les études d'association pangénomique (ou GWAS pour Genome Wide Association Study) permettent de se libérer de l'étape de la génération d'hypothèses et d'explorer des gènes qui n'auraient pas été sélectionnés sur la base d'*a priori* biologiques. En effet, ce ne sont plus quelques variants qui sont testés mais l'ensemble des variants du génome (plusieurs centaines de milliers de SNPs), le plus souvent correspondant à une puce de génotypage. Ces *tagging-SNPs* sont choisis pour couvrir la totalité du génome, c'est-à-dire pour capter l'information portée par les SNPs non génotypés par la puce, du fait du déséquilibre de liaison. Le génotype de certains SNPs peut être inféré de la connaissance du génotype d'un seul SNP ciblant la région considérée, c'est ce que l'on appelle l'imputation. Ces méthodes seront également détaillées dans le chapitre 3.

- **Approche gène candidat**

Comme cela a été précédemment indiqué, les études d'association de type gène candidat ont essentiellement porté sur des gènes susceptibles *a priori* de jouer un rôle dans la sensibilité à la lèpre. Ainsi, de nombreuses études ont testé une association entre la lèpre per se et des variants génétiques localisés sur les gènes du CMH, de la famille des « *Toll like receptors* » (*TLR*) et de la famille des cytokines [291,351]. Comme pour la liaison, seul un sous-ensemble de ces études s'est intéressé aux formes cliniques de la maladie.

Le système HLA : Au vu des éléments de physiopathologie, il est aisément compréhensible que la majorité des études a été faite au niveau des gènes ayant un rôle dans le système immunitaire ou l'inflammation. Les polymorphismes ayant été le plus souvent associés à la sensibilité à la lèpre sont :

- Pour le CMH de classe II, les gènes *HLA-DR2* et *HLA-DR3* [352], dont les associations ont été retrouvées dans différentes populations et par des équipes différentes, ce qui souligne la solidité de ces résultats,
- Concernant le CMH de classe I, des allèles sont associés aux cas de lèpre aussi bien pour les gènes *HLA-A*, *HLA-B* que *HLA-C* dans des échantillons turcs, indiens et vietnamiens. Cependant, aucun des allèles trouvés dans l'échantillon turc n'est commun avec les polymorphismes indiens ou vietnamiens [167]. Bien qu'il soit tout à fait possible que certains polymorphismes soient restreints à une population donnée,

l'implication de tous ces allèles en tant que polymorphismes causaux reste discutable. Enfin, la séquence A du polypeptide lié au CMH classe I (*MICA*) en Chine et au Brésil et *HLA-G* au Brésil [353,354] ont été testés.

- Parmi les gènes du CMH de classe III, *TNF- α* et *LTA*, qui codent des cytokines pro-inflammatoires, sont identifiés par plusieurs études. Concernant *TNF- α* , le même SNP est retrouvé dans plusieurs études portant sur différentes populations en Inde et en Thaïlande [336,355]. Cependant, ce polymorphisme est associé à des phénotypes différents en fonction de la population considérée par l'étude. Ainsi, tandis que ce SNP est associé au risque de développer la lèpre *per se* dans certains échantillons, il est surreprésenté seulement chez les patients lépromateux dans un échantillon indien et, au contraire, associé aux formes tuberculoïdes dans une étude portant sur des Brésiliens [351]. La lymphotoxine- α (*LTA*) au Malawi [356] a également été testée, la fraction 4 du complément (*C4*) en Thaïlande [357], *IFN γ* au Brésil [348].

Au-delà des contradictions entre les publications, les résultats d'association concernant les polymorphismes de gènes du CMH sont d'autant plus délicats à interpréter que la région physique du CMH, du fait de la très faible distance physique entre les loci considérés, les allèles étudiés sont donc sujets à un fort déséquilibre de liaison.

Récepteurs de l'immunité innée : les gènes codant les récepteurs impliqués dans l'immunité innée ont été les plus étudiés. Ainsi, six gènes ayant fait l'objet d'investigations ont été recensés [351] : *TLR1*, *TLR2*, *TLR4*, *NOD2*, *MRC1* et *VDR*. Nous verrons que si pour certains les arguments concernant leur implication dans la prédisposition à la lèpre semblent solides, d'autres sont l'objet de résultats contradictoires.

- **VDR :** L'intérêt pour le récepteur à la vitamine D (*VDR*) se justifie par le large spectre de modulations que présente la vitamine D sur le système immunitaire : elle peut présenter des actions inhibitrices ou au contraire stimulatrices en fonction du type de cellule impliquée. La forme active de la vitamine D, module la différenciation, la croissance et la fonction d'un grand nombre de cellules y compris les cellules du système immunitaire avec des effets immunosuppresseurs [359]. Au moins trois études indépendantes ont recherché des associations entre les polymorphismes de

VDR et le fait, d'une part, de développer une lèpre *per se* puis, d'autre part, de présenter une forme clinique particulière de la maladie. Plusieurs variants alléliques du gène *VDR* ont été décrits et l'influence d'un polymorphisme du codon 352 avec les deux allèles désignés T et t a été évaluée dans une étude cas-témoins dans une population indienne [360].

- **NOD2** : La famille des récepteurs *Nod-like* regroupe des récepteurs cytosoliques qui se lient à certains éléments de la paroi bactérienne. Dans cette famille, le récepteur *NOD2* semble particulièrement intéressant à explorer puisqu'il a été impliqué dans la sensibilité à la tuberculose, dont l'agent pathogène *Mycobacterium tuberculosis* partage des caractéristiques avec *M. leprae*. Ainsi, la délétion du gène *NOD2* a été associée dans les modèles murins, *in vivo* et *in vitro*, à une altération globale de la réponse immunitaire à *M. tuberculosis*. De plus, des mutations du gène humain *NOD2* ont été associées à une sensibilité accrue ou à une protection vis-à-vis de la tuberculose chez des Afro-Américains. Une étude « gène candida » a identifié 8 SNPs dans les régions flanquantes de ce gène, chacun de ces polymorphismes est associé à une sensibilité augmentée à la lèpre *per se* [361].
- **MRC1** : codant pour le récepteur au mannose de type C1, nous fournit l'exemple rare d'un cas où des résultats d'épidémiologie génétique ont fait l'objet d'une tentative de validation fonctionnelle. Le choix de *MRC1* a été motivé par son implication dans la phagocytose, en effet le récepteur au mannose a la propriété de se lier à certaines lipoprotéines bactériennes et, ainsi, d'induire la phagocytose. Une étude sur un échantillon vietnamien a permis d'identifier un unique SNP, nommé G396S, associé à la résistance à la lèpre. Ce résultat a été répliqué par la même équipe dans un échantillon brésilien. La phagocytose a été évaluée dans des cellules HEK portant le SNP G396S mais aucune différence n'a été observée entre les phénotypes associés aux allèles sauvages et G396S. Cet exemple souligne que la validation fonctionnelle des résultats d'épidémiologie génétique est délicate et constitue le plus souvent une étape limitante [167].

- **Famille des récepteurs Toll-like :**
 - **TLR4** : code un récepteur présent à la surface des macrophages, monocytes et cellules dendritiques qui se lie au LPS des bactéries Gram-négative. Concernant la sensibilité à la lèpre per se, *TLR4* a été l'objet de deux études, l'une réalisée sur un échantillon constitué d'individus éthiopiens et l'autre venant du Malawi. La première a mis en évidence une association entre deux SNPs, déjà identifiés dans la sensibilité à d'autres infections à bactéries Gram-négative telle que la légionellose, et une protection vis-à-vis de la lèpre. Cependant, ces résultats n'ont pas été répliqués [362].
 - **TLR1 et TLR2** : sont d'un intérêt particulier dans l'étude de la sensibilité à la lèpre puisqu'ils forment un hétérodimère qui semble être impliqué dans la reconnaissance de *M. leprae*. Des études portant sur *TLR2* en Corée du Sud [363], en Ethiopie [364] et au Malawi [356] ont testé la sensibilité à la lèpre per se et l'implication du gène *TLR1* est considérée comme l'une des plus solide [351]. Ainsi, le SNP 1805G (guanosine en position 1805) a été identifié comme protecteur vis-à-vis de la lèpre per se dans une étude en Turquie. Le fait que cette association n'ait pas été retrouvée au Népal et au Bangladesh dans deux études ultérieures n'est pas, dans ce cas particulier, interprété comme une remise en question des résultats précédents. En effet, la fréquence de l'allèle 1805G est variable selon la population considérée. De plus, il a été montré que ce polymorphisme est sous sélection positive dans certaines populations seulement. Ces résultats suggèrent que *TLR1* pourrait avoir des impacts différents sur la sensibilité à la lèpre per se selon la population considérée. Enfin, plusieurs études fonctionnelles ont mis en évidence l'importance de ce polymorphisme dans l'expression du récepteur *TLR1* et dans la voie de signalisation de NF-κB.

Outre les gènes des deux catégories précédentes, les études gènes candidats ont également porté sur des gènes codant des cytokines dont la plupart des associations n'ont été rapportées que dans une seule étude :

- *L'IL-10* a été associé à la polarisation dans deux échantillons brésiliens indépendants [365,366], *l'IL-12B* au Mexique, *l'IL-12RB2* au Japon et *IL-23R* en Inde [367,368] ;
- La Mannose Binding Lectin 2 (*MBL2*) au Brésil et au Népal [369,370] ;
- La protéine de résistance naturelle associée aux macrophages 1 (*NRAMP1*) au Mali [371] et *KIR2DS3* [372] ;
- La laminine 2 (*LAMA2*) en Indonésie [373] ;
- La défensine β 1 (*DEFB1*) au Mexique [374].

Approche génome entier

La première étude d'association pangénomique sur la sensibilité à la lèpre *per se* a été publiée en 2009 [166]. Elle comprenait 706 cas et 1 225 contrôles originaires de l'Est de la Chine. Parmi les 491 883 *SNP* testés couvrant l'ensemble du génome, 93 *SNPs* étaient associés au seuil évocateur avec le phénotype lèpre *per se* de façon ($p < 10^{-5}$). Ces signaux, les plus prometteur, ont ensuite été testés dans trois autres échantillons indépendants dits de réplication totalisant 3 254 cas et 5 955 contrôles, également chinois, principalement de la Chine de l'Est mais aussi issus de minorités ethniques. Une seconde étude a utilisé 4 367 contrôles supplémentaires dans l'échantillon initial et des échantillons de réplication similaires [375].

Ces deux études ont permis d'identifier un total de 18 polymorphismes significativement associés à la lèpre et localisés dans huit gènes : *HLA-DR-DQ*, *RIPK2* (codant pour le « Receptor-Interacting Protein Kinase 2 »), *TNFSF15* (Tumor Necrosis Factor SuperFamily member 15), *CCDC122* (Coiled- Coil Domain Containing 122), *LACC1* (LACCase Containing domain 1), *NOD2* (Nucleotidebinding Oligomerization Domain containing 2), *RAB32* (protéine Rab32) et *IL-23R* (récepteur de l'interleukine 23).

Une hétérogénéité entre les groupes PB et MB a été observée pour quatre polymorphismes dans quatre gènes (*LACC1*, *LRRK2*, *NOD2*, et *RIPK2*). Deux études menées en Inde et au Vietnam ont cherché à valider ces associations en utilisant les mêmes SNPs. Au Vietnam dans un échantillon familial incluant 188 cas PB et 286 cas MB, les signaux localisés dans les gènes *HLADR-DQ*, *RIPK2*, *CCDC122* et *LACC1* ont été validés ($p < 0.05$). L'effet était plus fort chez les patients MB pour tous les marqueurs à l'exception de *CCDC122* [168]. Dans l'étude cas-témoin indienne à l'effectif relativement faible (211 cas dont 74 PB et 137 MB, seuls les signaux dans les gènes *LRRK2* et *RIPK2* ont été retrouvés ; l'effet observé était systématiquement plus significatif chez les patients PB [376]. Enfin, notre GWAS familial va permettre de répliquer le signal du gène *LACC1*.

À l'heure actuelle, plusieurs autres études pangénomiques, à partir de celles-ci, ont été publiées et ont montré l'implication de nouveaux gènes tels *IL23R*, *RAB23*, *BATF3*, *IL18RAP/IL18R1* et *IL12B* que nous avons également répliqués [377–379].

Un second « GWAS » (48 000 marqueurs couvrant 2 092 gènes sur les 20-25 000 du génome) a été réalisé en 2010 dans une population indienne. L'échantillon primaire comportait seulement 258 cas et 300 contrôles de la région de New Delhi. Des marqueurs dans les gènes *TLR1* et *HLA-DRB1/DQA1* étaient significativement associés à la lèpre *per se*. Le test d'hétérogénéité entre les groupes PB et MB n'était pas significatif [380]. Néanmoins, dans les deux GWAS, le phénotype d'intérêt principal était la lèpre *per se*, et seuls les SNPs significatifs pour la lèpre *per se* ont ensuite été testés pour l'hétérogénéité entre les groupes PB et MB. Ceci ne constitue pas la stratégie optimale pour identifier les SNPs associés à la polarisation car certains SNPs ayant un impact spécifique sur la polarisation peuvent apparaître comme de faux négatifs lors de l'étude de la lèpre *per se*.

Les approches gènes candidats et pangénomiques permettent ainsi d'atteindre des catégories différentes de gènes et pourraient donc être pensées comme complémentaires. Cependant, l'étude de leurs limites met en évidence qu'elles ne peuvent explorer que certains groupes de gènes particuliers. D'autre part, les évidences statistiques de l'épidémiologie génétique, aussi fortes soient elles, doivent être validées par des études fonctionnelles ce qui n'est pas le cas. Le **tableau 7** répertorie tous les SNPs, hors-HLA associés à lèpre *per se* et déjà publiés.

ADN ?	Analyse	Échantillon	Marqueurs	Objectif	Principe	Limitations
Analyses ADN-indépendantes (sans marqueurs génétique)	Agrégation familiale	Familles	Pas de marqueurs génétiques	Mettre en évidence un excès de risque en cas d'antécédent familial	Comparer le risque de développer la maladie chez les apparentés de sujets atteints et de sujets sains	Distinction impossible entre un risque lié à des facteurs génétiques et celui lié à un environnement commun
	Jumeaux	Jumeaux		Distinguer la contribution relative de facteurs génétiques de celle d'autres facteurs familiaux partagés pouvant participer à l'agrégation familiale	Comparer les taux de concordance phénotypique chez les jumeaux dizygotes et monozygotes	Rareté des cas, Distinction mono/dizygote incertaine (temps plus anciens)
	Adoptés	Adoptés et leurs parents biologiques et adoptifs		Comparer la concordance phénotypique entre les enfants et leurs parents naturels et adoptifs	Rareté des cas, Âge à l'adoption	
	Ségrégation	Familles ≥ 2 générations		Déterminer l'origine des corrélations familiales, en particulier rechercher l'effet d'un gène majeur parmi l'ensemble des facteurs de risque génétiques et environnementaux	Modéliser le risque d'être atteint en fonction de l'effet d'un gène majeur, de corrélations familiales et de facteurs environnementaux	Grands échantillons, Généalogie fiable, Modèles génétiques simples
Analyses ADN-dépendantes (avec marqueurs génétique)	Liaison	Familles ≥ 2 enfants atteints	Région candidate (2-10) Pangénomique (400-10,000)	Localiser sur le génome des régions chromosomiques contenant un ou plusieurs gène(s) de susceptibilité	Étudier le lien entre la ressemblance phénotypique et génotypique de deux apparentés. Dans la méthode des paires de germains atteints, il est testé, pour chaque marqueur, si 2 frères/sœurs malades partagent plus de 50% d'allèles hérités en commun de leurs parents	Maladies peu fréquentes, Localisation génomique de l'ordre de 5 à 20 mégabases
	Association	En population : individus sains et atteints non apparentés En familles : Majoritairement des trios (parents et 1 enfant atteint)	Région candidate (1-10) GWAS (500,00-5M) Séquençage (30M-3G)	Identifier au sein d'une région particulière un variant génétique associé à une modification du risque de développer la maladie	En population : Comparer la fréquence des allèles ou génotypes d'un marqueur donné chez des cas et des témoins non apparentés En familles : Tester s'il existe chez les parents hétérozygotes une sur/sous-transmission d'un de leurs allèles aux enfants atteints	En population : Choix des témoins délicat, Associations indirectes en raison du déséquilibre de liaison, ou factices dues aux structures génétiques des populations En familles : Génotypage nécessaire des parents (donc vivants), Pas de variants de la survie foetale

Tableau 6 : Principales méthodes utilisées en épidémiologie génétique [94]

Phénotypes complexes (exemple la lèpre)

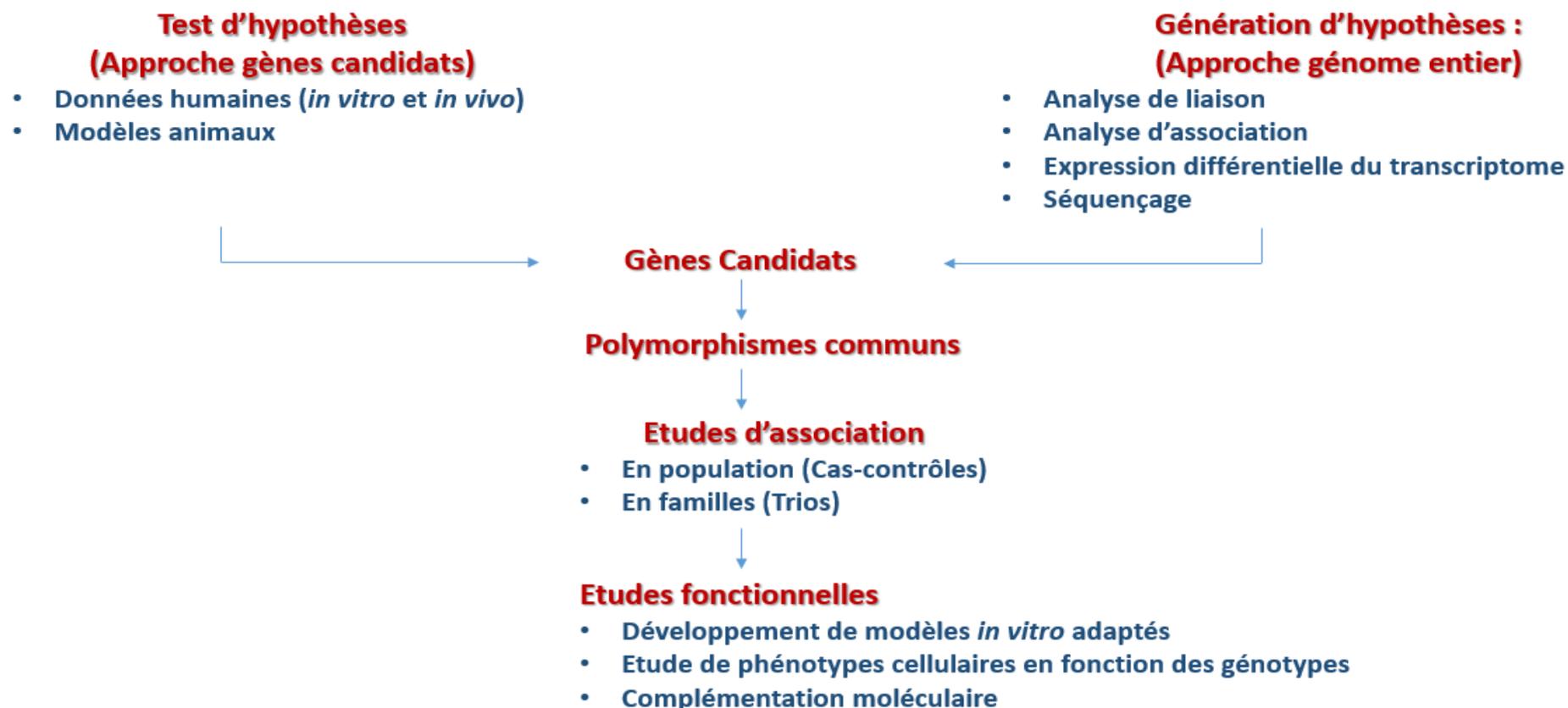


Figure 24 : Stratégies d'analyse communément utilisées en épidémiologie génétique pour l'étude des maladies à hérédité complexe

L'approche « test d'hypothèse » consiste à sélectionner des gènes candidats a priori en se fondant sur des données issues d'études humaines *in vitro* or *in vivo* ou de modèles animaux (peu contributifs dans le cas de la lèpre), de connaissances théoriques concernant la maladie ou des maladies proches (ex : tuberculose). L'approche « génération d'hypothèse » définit les gènes candidats sur la base des résultats d'une analyse de liaison génétique, d'association, transcriptomique explorant le génome complet. Le rôle des gènes candidats identifiés est ensuite étudié en testant l'association entre le phénotype étudié et les polymorphismes de ces gènes (cas-témoins ou familiale). Il est difficile de valider fonctionnellement les polymorphismes associés car l'effet attendu est subtil (maladie à déterminisme complexe), et le développement d'un modèle *in vitro* approprié est difficile (*M. leprae* non cultivable *in vitro*).

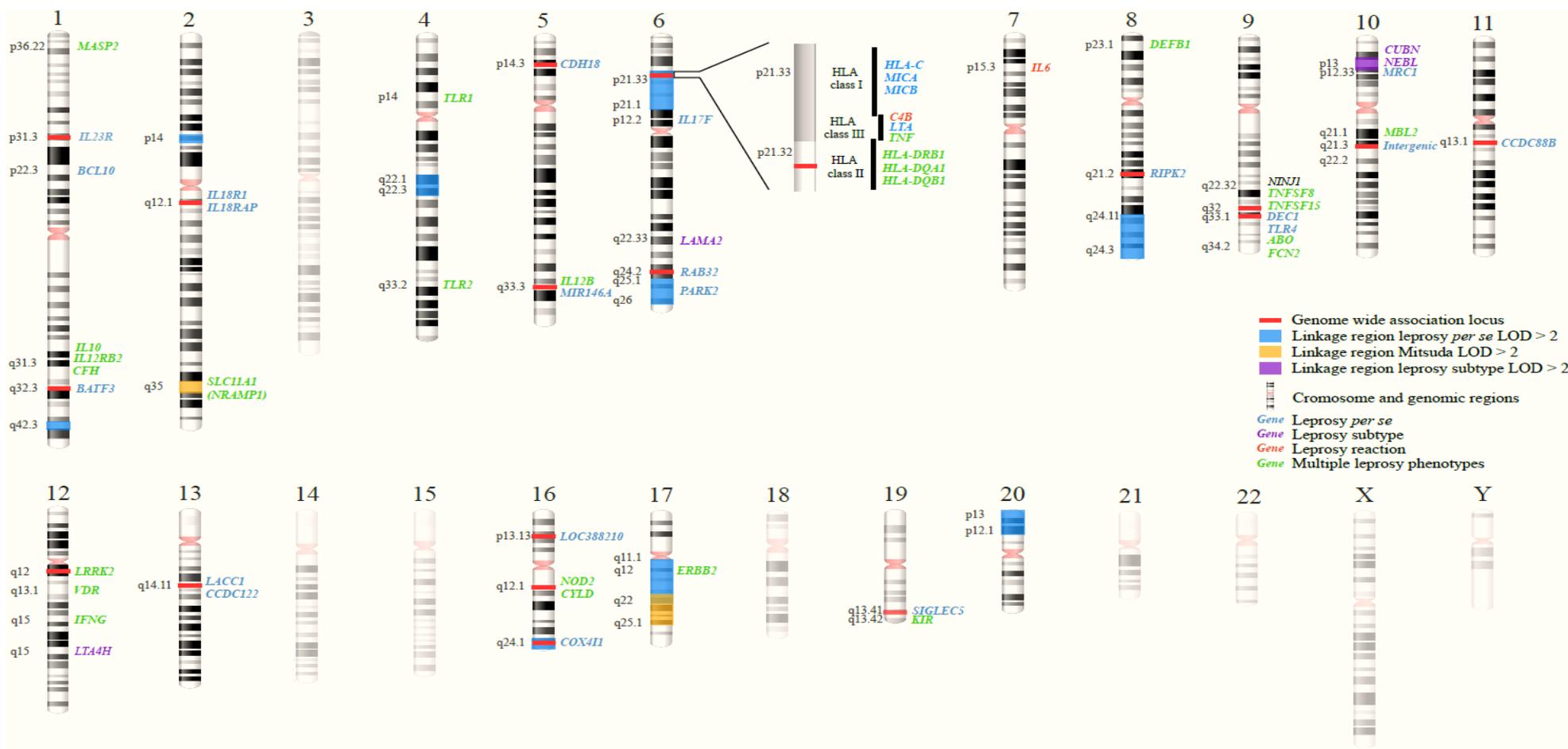


Figure 25 : Carte des zones de liaison et des gènes au sein desquels des variants ont été identifiés comme associés à la lèpre [381]

À gauche de chaque chromosome figurent les gènes pour lesquels un marqueur a été associé à la lèpre, et à droite les régions de liaison. Les gènes associés, soit dans un GWAS, soit dans plusieurs études d'association, sont en gras. La région du CMH située en 6p21 a été élargie pour détailler les gènes de classe I, II et III. Aucune association n'a été rapportée sur les chromosomes sexuels. Le positionnement des gènes et des régions est ordonné sans être à l'échelle.

SNP	CHR	Position	Gène	m/M	RA	P-val	Référence
<i>rs3762318</i>	1	67597119	<i>IL23R</i>	G/A	A	3,27.10 ⁻¹¹	[375]
<i>rs76418789</i>	1	67648596	<i>IL23R</i>	A/G	A	1,03.10 ⁻¹⁰	[378]
<i>rs2735591</i>	1	85744472	<i>BCL10</i>	G/A	A	1,03.10 ⁻⁰⁹	[382]
<i>rs146466242</i>	1	152275298	<i>FLG</i>	A/T	A	3,39.10 ⁻¹²	[378]
<i>rs2221593</i>	1	212873431	<i>BATF3</i>	T/C	T	3,09.10 ⁻⁰⁸	[377]
<i>rs2058660</i>	2	103054449	<i>IL18RAP/IL18R1</i>	A/G	A	4,57.10 ⁻¹⁹	[379]
<i>rs6807915</i>	3	12313846	<i>SYN2</i>	C/T	T	1,94.10 ⁻⁰⁸	[383]
<i>rs145562243</i>	3	48719549	<i>NCKIPSD</i>	T/C	T	1,71.10 ⁻⁰⁹	[378]
<i>rs5743618</i>	4	38798648	<i>TLR1</i>	C/A	A	1,70.10 ⁻⁰⁹	[380]
<i>rs73058713</i>	5	20374213	Intergénique	A/C	A	9,54.10 ⁻⁰⁹	[377]
<i>rs6871626</i>	5	158826792	<i>IL12B</i>	A/C	C	3,95.10 ⁻¹⁸	[379]
<i>rs2275606</i>	6	146918950	<i>RAB32</i>	A/G	A	3,94.10 ⁻¹⁴	[375]
<i>rs4720118</i>	7	33469241	<i>BBS9</i>	T/C	T	3,85.10 ⁻¹⁰	[383]
<i>rs55894533</i>	8	11749242	<i>CTSB</i>	C/A	C	5,07.10 ⁻¹¹	[383]
<i>rs13259978</i>	8	17396415	<i>SLC7A2</i>	C/G	C	1,74.10 ⁻⁰⁸	[384]
<i>rs160451</i>	8	90662941	<i>RIPK2</i>	T/C	C	7,45.10 ⁻¹²	[377]
<i>rs42490</i>	8	90778513	<i>RIPK2</i>	G/A	G	1,38.10 ⁻¹⁶	[166]
<i>rs10100465</i>	8	118626279	<i>M.10D30</i>	A/G	G	2,85.10 ⁻¹¹	[383]
<i>rs6478108</i>	9	117558703	<i>TNFSF15</i>	G/A	A	3,39.10 ⁻²¹	[166]
<i>rs10817758</i>	9	118143933	<i>DEC1</i>	T/C	T	1,15.10 ⁻⁰⁸	[377]
<i>rs149308743</i>	9	139258965	<i>CARD9</i>	T/C	T	2,09.10 ⁻⁰⁸	[378]
<i>rs58600253</i>	10	64507904	Intergénique	T/C	T	3,02.10 ⁻¹²	[377]
<i>rs780668</i>	10	73111408	<i>SLC29A3</i>	T/C	T	2,17.10 ⁻⁰⁹	[378]
<i>rs663743</i>	11	64107735	<i>CCDC88B</i>	A/G	A	8,84.10 ⁻¹⁴	[377]
<i>rs925368</i>	12	110390979	<i>GIT2</i>	C/T	C	9,18.10 ⁻¹⁷	[384]
<i>rs671</i>	12	112241766	<i>ALDH2</i>	A/G	A	2,00.10 ⁻²⁰	[384]
<i>rs3764147</i>	13	44457925	<i>LACC1</i>	A/G	G	3,72.10 ⁻⁵⁴	[166]
<i>rs8002861</i>	13	44474517	<i>LACC1</i>	G/A	A	1,07.10 ⁻¹²	[377]
<i>rs142179458</i>	14	62203623	<i>HIF1A</i>	G/A	A	4,95.10 ⁻⁰⁹	[385]
<i>rs77061563</i>	16	11452542	<i>LOC388210</i>	T/C	C	6,23.10 ⁻¹⁵	[377]
<i>rs9302752</i>	16	50719103	<i>NOD2</i>	A/G	G	3,77.10 ⁻⁴⁰	[166]
<i>rs16948876</i>	16	50854418	<i>CYLD</i>	A/G	A	1,64.10 ⁻¹⁰	[375]
<i>rs75680863</i>	22	31007023	<i>TCN2</i>	T/A	A	8,37.10 ⁻²¹	[384]

Tableau 7 : Variants, hors-HLA, associés à la lèpre per se et précédemment publiés

RA : Allèle à Risque,

m/M = allèles mineur/majeur,

Les positions sont données en fonction de hg19,

rs145562243 ne fait pas partie du panel de 1000 génomes (Phase 1 et 3),

Les SNPs en vert sont significatifs dans l'échantillon de réplication.

Carte mentale 1 :

La lèpre

Situation actuelle :

- 200 000 nouveaux cas/an
- > 150 pays touchés
- 3 millions de malades avec des invalidités
- 1 malade sur 10 est un enfant

Mycobacterium leprae :

- 1873 par Hansen : *Bacille intra-cellulaire (BAAR)*
- Non cultivable *in vitro*
- Réservoirs naturels : *Homme, tatou, écureuil roux*
- Tropicisme : *Macrophages dermiques, cellule de Schwann*
- Croissance très lente : 12 à 20 jours pour doubler
- Modèles d'étude : *Coussinets de souris, Tatou*
- Génome quasi-clonal et 50% de pseudogènes

Génétique Humaine de la lèpre :

- **Facteurs de variabilité du phénotype**
 - Facteurs microbiens : exposition, virulence,
 - Facteurs de l'hôte : non génétique et génétique
- **Analyse sous marqueur génétique :**
 - Agrégation familiale, jumeaux et ségrégation
- **Analyse avec marqueur génétique**
 - Liaison et Association
 - Gène candidat ou pangénomique
HLA-DR/HLA-DQ/TLR2/NOD2/LTA/TNF- α /LACC1...

La lèpre

Transmission :

- Voie aérienne la plus probable : *Sécrétions nasales, aérosols, contact objets souillés...*

Phénotype :

- **Forme tuberculoïde**
 - Réponse immunitaire cellulaire
 - Lésions cutanées peu nombreuses, délimitées et anesthésiques
→ Charge bacillaire faible (peu contagieuse)
 - Lésions nerveuses rapides et importantes
- **Forme lépromateuse**
 - Pas de réponse immunitaire cellulaire
 - Lésions cutanées nombreuses (lépromes), peu hypoesthésiques
→ Charge bacillaire importante (très contagieuse)
 - Dissémination systémique (yeux, os, testicules...)
 - Lésions nerveuses tardives et faibles

Diagnostic/Antibiothérapie :

- **Clinique :** *Interrogatoire, examen cutané (sensibilité), palpation des nerfs*
- **Paraclinique :** *Histopathologie, bactériologie (index bacillaire), immunologie, PCR*
- **Polychimiothérapie :** *Rifampicine, Dapsone, Clofazimine*
- **Réactions immunitaires aiguës après antibiothérapie :**
Réactions lépreuses (RR/ENL)
- **Prophylaxie :** *prévention et « éducation » des population*

1.9 Motivations et objectifs

Mon travail de thèse porte principalement sur la dissection de l'architecture génétique de la lèpre mais, pour rappel, il est important d'insister sur le fait que l'objectif de cette thèse est double :

- Il s'agit en premier lieu de disséquer l'architecture génétique de l'infection à *Mycobacterium leprae* et donc de déterminer les variants génétiques de susceptibilités/résistances associées à la lèpre per se chez l'Homme. Pour ce faire, nous avons réalisé le premier GWAS sur données familiales de la lèpre au Vietnam. Nous allons, dans le prochain chapitre suivre les différentes étapes que nous avons effectuées pour réaliser ce GWAS qui a donné lieu à un article principal en première auteure (**annexe 3**) et à un deuxième article en collaboration avec le laboratoire de Médecine et de Génétique Humaine à l'Université de McGill au Canada (**annexe 5**).
- Le second objectif, et ce sera le thème du chapitre 3, est de proposer une alternative efficace à la réplique dans un échantillon indépendant dans les études d'association pangénomiques au design familial. Ayant les parents déjà génotypés, il s'agit donc de réaliser une réplique cas-contrôles parentale et nous avons montré que cette approche était pertinente, valide et, statistiquement indépendante de l'étude d'association familiale. Ce travail a donné lieu à un deuxième article en première auteure (en review).
- Nous terminerons, et ce sera l'objet du chapitre 4, par une conclusion détaillant les limites et les perspectives de notre étude. Enfin, nous discuterons de la méthode même des GWAS et de leur impact dans la génétique épidémiologique.

Pour le confort de lecture, et afin que chaque partie puisse être lue indépendamment, certaines notions font l'objet de répétitions.

Chapitre 2 : Analyse d'association pangénomique familiale au Vietnam

Au cours de ce chapitre, nous allons détailler l'échantillon familial sur lequel a été réalisée l'étude, la génération des données génétiques par puce pangénomique, les critères de qualité que nous avons retenus et les outils méthodologiques que nous avons utilisés pour analyser nos données afin de relier information phénotypique et génétique dans un contexte d'étude familiale. Enfin, nous rapporterons les principaux résultats obtenus de l'analyse familiale primaire, de l'analyse de réplication sur un échantillon indépendant, et de l'analyse multivariée.

2.1 Données phénotypiques

Les données ont été recueillies au Vietnam à l'Hôpital Dermatovénérologie (DV), au début des années 90 dans la ville d'Ho Chi Minh (anciennement Saigon) située au sud du pays non loin du fameux delta du Mékong. Cette zone était particulièrement adaptée à cette étude sur la génétique humaine de la lèpre pour plusieurs raisons : 1) la prévalence y était à l'époque parmi les plus élevées du pays, 2) il existe un hôpital largement dédié au diagnostic, au traitement et au suivi des patients atteints de la lèpre, et 3) la population est relativement homogène en termes d'ethnie par rapport au reste du pays.

2.1.1 Région d'étude : le Vietnam, milieu physique et humain

➤ Le Vietnam

Longue bande de terre montagneuse en forme de S bordant le Laos au Nord-Ouest puis le Cambodge au Sud-Ouest en surplomb de la mer de Chine, le Vietnam est traversé de bassins hydrographiques importants qui font de ses deux grandes plaines des greniers abondants : le delta du fleuve Rouge au nord, qui irrigue Hanoi, et celui du Mékong au Sud du pays, où se trouve Hô Chi Minh-Ville. Sur une superficie de totale de 330 967 Km², vivent ainsi près de 98 millions d'individus (**figure 26**).

ANALYSE D'ASSOCIATION PANGÉNOMIQUE FAMILIALE

Le Viêt-Nam reste l'un des microcosmes ethniques les plus complexes du monde puisque pas moins de 54 ethnies cohabitent. Originaires de la Chine méridionale, les Viet (ou Kinh) représentent près de 85% de la population, les chinois 5% tandis que les 10% restants sont composés d'ethnies minoritaires vivant principalement dans les montagnes. Il s'agit d'un pays jeune puisque le quart de la population a moins de 15 ans, et pauvre avec un revenu moyen par habitant de 2 540 \$ par an (France, 46 481 [386]) et un taux de mortalité infantile de 16 pour mille [387,388].

➤ La lèpre au Vietnam depuis le recrutement jusqu'à aujourd'hui

Avant de décrire notre population d'étude, nous allons situer le contexte de la lèpre au Vietnam, en nous appuyant sur le rapport détaillé de l'OMS publié en 2008 concernant le Vietnam entre 1983 et 2006 [389] qui correspond à la période de recrutement des données (1991-2008). En 1983, année d'introduction de la polychimiothérapie de l'OMS au Vietnam, le pays a déclaré 2 021 nouveaux cas. La prévalence était alors de 6,78 pour 10 000 habitants. Le niveau de couverture par la polychimiothérapie a atteint 50% en 1985, et pratiquement 100% en 1991. Le Vietnam a atteint l'objectif de l'élimination (telle que définie par l'OMS par une prévalence < 1/10 000) au niveau national en 1995. L'épidémiologie de la lèpre dans cette région a évolué de façon comparable à celle observée dans les autres pays endémiques : une nette diminution de la prévalence a été observée entre 1993 et 2001 (respectivement 43 et 3 pour 100 000), tandis que l'incidence est restée relativement constante (3.5 contre 2 pour 100 000). La prévalence a continué de diminuer jusqu'à 0,1/10 000 en 2006. La diminution rapide peut s'expliquer par : 1) la réduction de la durée du traitement grâce à l'efficacité de la polychimiothérapie, 2) l'utilisation d'une définition standard des cas pour le traitement, 3) les changements des pratiques d'enregistrement – comme retirer des registres les sujets qui ont terminé leur traitement, sont décédés, ont quitté la zone ou ont interrompu le traitement.

Par opposition à la prévalence, le nombre annuel de nouveaux cas dépistés a dans un premier temps augmenté de 1991 à 1997, principalement en raison d'un dépistage actif renforcé. Depuis 1997, parallèlement à la baisse de la prévalence, ce nombre a diminué chaque année. En 2006, seuls 666 cas ont été notifiés, soit un taux de 0,75/100 000. Ce faible nombre de nouveaux cas annuels explique pourquoi nous n'avons recruté que peu de nouveaux cas après 2000. La proportion des femmes parmi les nouveaux cas dépistés a fluctué

ANALYSE D'ASSOCIATION PANGÉNOMIQUE FAMILIALE

entre 1983 et 1995, avant de rester stable autour de 37%. La proportion d'enfants de moins de 15 ans parmi les nouveaux aurait diminué de 7,3% en 1996 à 5,3% en 2006 (tel qu'écrit dans le rapport, mais avec les données fournies, cette diminution est non significative). Si l'on suppose que la politique de détection active de cas parmi les écoliers est restée stable, alors la réduction de la proportion observée d'enfants parmi les nouveaux cas suggère que la transmission de la maladie est en régression. Cette réduction de l'incidence de la lèpre chez les enfants a également été observée dans plusieurs autres pays d'Asie du Sud-Est, tels le Bhoutan, la Birmanie et la Thaïlande.

Enfin, si la proportion de cas présentant une incapacité de grade 2 au diagnostic a diminué depuis 1996, elle était toujours de 1 cas sur 6 en 2006. Ceci témoigne d'un retard de dépistage de certains cas, probable reflet d'une forte stigmatisation dans la communauté et d'une faible sensibilisation à la maladie dans le grand public. Les données concernant la lèpre au Vietnam pour l'année 2012 confirment les différentes tendances observées jusqu'en 2006 : 296 nouveaux cas déclarés, dont 105 femmes (35%), 10 enfants (3%), 191 MB (65%) et 44 qui présentaient une incapacité de grade 2 (15%) [390].

En 2018, seuls 77 cas ont été détectés incluant 24 femmes et 0 enfant correspondant à un taux de 0,1 pour 100 000 habitants et avec une proportion de cas présentant une incapacité de grade 2 de 18% (**figure 27**) [391]. Comme dans les autres pays touchés par la lèpre, le fardeau de la maladie ne saurait être évalué par les seuls chiffres d'incidence ou de prévalence. Ils ne sont que la partie émergée de l'iceberg, dont le socle est formé d'individus, certes guéris sur le plan microbiologique, mais qui présentent des séquelles physiques et/ou sociales majeures dues à la maladie. En dépit de soins médicaux de qualité et d'aides sociales importantes offerts par l'État aux individus atteints de lèpre, des populations rurales pauvres restent volontairement à l'écart du dépistage et des traitements. Ceci s'explique par la crainte d'une stigmatisation importante mais également par la nécessité de quitter le village en cas de diagnostic de lèpre, pour être pris en charge dans une léproserie. Ces populations restent donc très probablement ignorées par les statistiques officielles.

ANALYSE D'ASSOCIATION PANGÉNOMIQUE FAMILIALE



Figure 26 : Le Vietnam

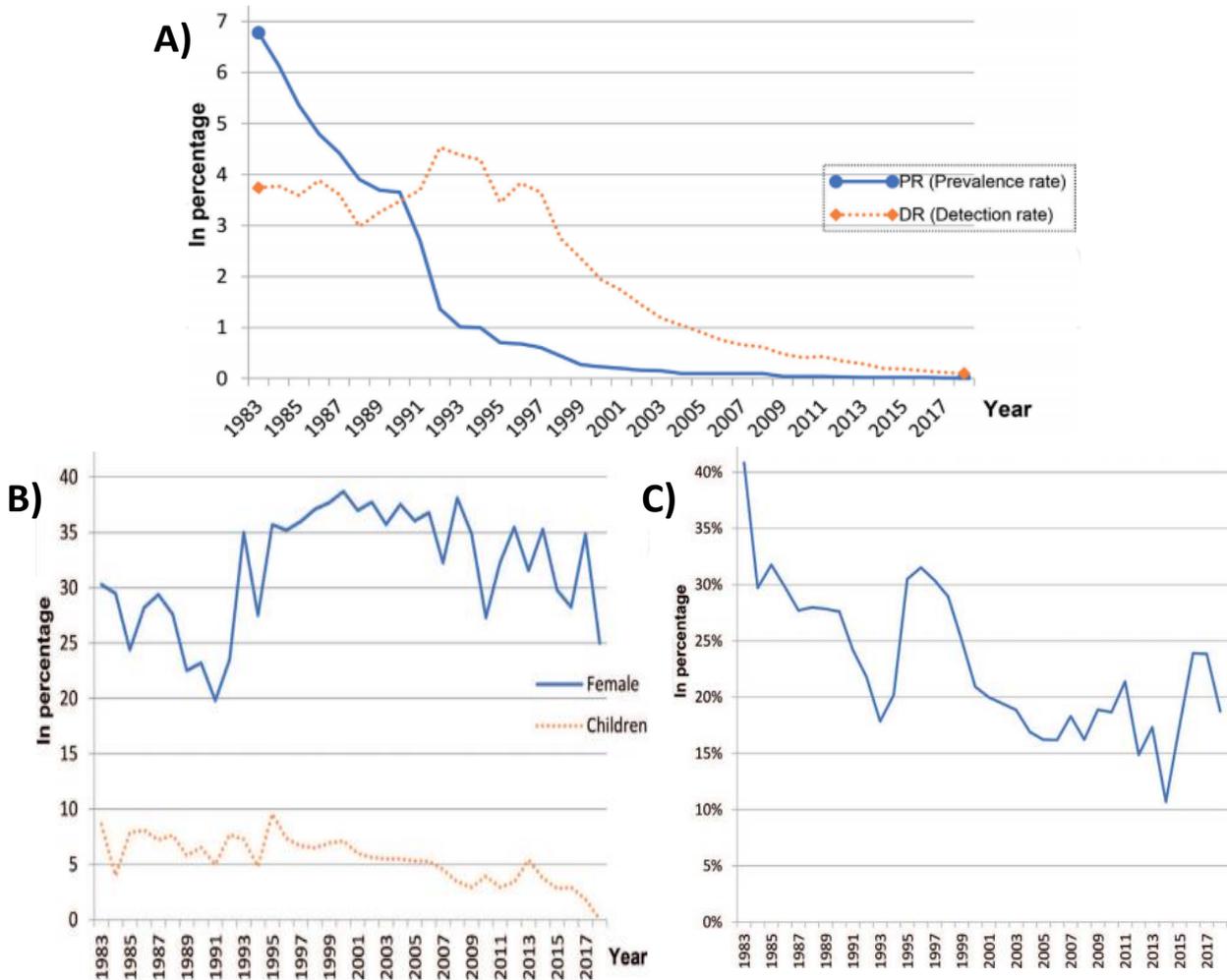


Figure 27 : Statistiques aux Vietnam [391]

- A) Taux de prévalence (PR = proportion de nouveaux cas pour 100 000 habitants) et taux de détection (DR = proportion de cas pour 10 000 habitants) de la lèpre entre 1983 et 2018.
- B) Proportion de femmes et d'enfants parmi les nouveaux cas.
- C) Proportion d'incapacités de grade 2 parmi les nouveaux cas.

2.1.2 Recueil et description de l'échantillon

➤ Stratégie de recueil des données

Les familles ont été identifiées à partir des registres de l'Hôpital de DV d'Ho Chi Minh-Ville, centre de référence pour le traitement des patients lépreux dans cette zone géographique qui couvre plus de 5 millions d'individus. Depuis la fin des années 80, le laboratoire de Génétique Humaine des Maladies Infectieuses a noué une collaboration étroite avec cet hôpital. Les données présentées et étudiées dans ce manuscrit ont été recueillies lors de plusieurs campagnes indépendantes effectuées de 1991 à 2008 (**tableau 8**). Différentes études ont été menées, essentiellement de génétique épidémiologique (ségrégation, liaison et association), mais aussi d'immunologie fonctionnelle. Pour les projets s'intéressant aux familles, le mode de recrutement s'est toujours effectué rétrospectivement à partir d'un enfant atteint : c'est l'identification d'un enfant atteint qui menait au recrutement des parents et de la fratrie. Deux enfants atteints étaient nécessaires pour les études de liaison. Les projets « Mitsuda » et « Réaction de réversion » sont les seuls projets pour lesquels la lèpre per se n'était pas le phénotype principal d'intérêt. Le premier a conduit à recruter des individus (et non des familles) pour identifier les facteurs génétiques (ou non) influençant le résultat du test de Mitsuda [392,393]. Le second a ciblé des individus qui présentaient une RR ou qui étaient susceptibles d'en présenter une, essentiellement des individus classés « borderline » (BT, BB et BL selon Ridley et Jopling) [394]. Au total, 2 784 individus ont été recrutés en 18 ans, dont 1 182 cas de lèpre et l'échantillon qui a servi pour ce travail est extrait de l'ensemble de ces projets avec comme phénotype binaire : individus atteints versus non atteints.

L'ensemble des individus a été examiné par des léprologues vietnamiens expérimentés et le diagnostic de lèpre a été porté par au moins deux léprologues indépendants sur un faisceau d'arguments cliniques, microbiologique et éventuellement histopathologiques standards après examen des lésions cutanées. La forme clinique était établie suivant la classification de Ridley et Jopling. Finalement, la lecture du test de Mitsuda fut réalisée par deux léprologues différents, 28 jours après l'injection intradermique de lépromine chez tous les individus de l'échantillon (qu'ils soient sains ou atteints).

ANALYSE D'ASSOCIATION PANGÉNOMIQUE FAMILIALE

Deux codages furent considérés : un codage quantitatif correspondant à la valeur moyenne des lectures réalisées par les deux léprologues (de 0 à 14 mm) et un codage ordinal en quatre classes (<3, [3 ; 5],]5 ; 10[et ≥ 10 mm) selon les recommandations de l'OMS.

Un formulaire sous forme de questionnaire a été rempli pour chaque individu permettant de collecter les informations suivantes : patronyme, ethnie, lieu et année de naissance, sexe, nom des parents, statut vaccination au BCG, statut lèpre, et si atteint : année de survenue de la maladie, classification Ridley Jopling (TT, BT, BB, BL ou LL), classification par le clinicien (PB ou MB), nombre de lésions cutanées et nerveuses, indice bacillaire, traitement, présence ou non de réaction au test de Mitsuda et le développement ou non d'une réaction de type 1 (RR) ou de type 2 (ENL) (**figure 28**).

Enfin, il faut noter qu'un consentement, dont la formulation a bien entendu évolué au cours des années, fut obtenu auprès de l'ensemble des participants de l'étude en accord conjoint avec l'International Review Board de l'Hôpital DV d'Ho Chi Minh Ville, le Ministère de la Santé Vietnamien et les directives du Medical Research Council du Canada.

ANALYSE D'ASSOCIATION PANGÉNOMIQUE FAMILIALE

Projet	Période	Inclusion	Type	Total (N)	CNG (N)	Génotypes validés (N)	Cas de lèpre avec génotype validé					
							N	%♂	Âge médian	%MB	%RR	%ENL
OVN	1991	≥ 2 enfants atteints	Familles multiplex	187	160	140	63	51	22	56	0	0
NVN	1998-2000	≥ 1 enfant atteint	Trios/Familles	1 056	921	833	465	59	23	55	22	7
RR	2002-2004	RR	Trios	297	284	251	97	57	21	78	96	4
zRR	2005	RR	Trios	60	57	40	15	53	24	86	100	0
BL	2006-2007	Borderline	Trios/Individus	78	72	48	23	56	19	42	0	0
MIT	2006-2008	Sains et atteints	Familles/Individus	500	500	434	112	57	33	54	22	4
PB	2006-2007	PB	Trios	30	30	26	11	62	23	18	9	0
mBL	2007-2008	BL	Trios/Familles	216	204	178	74	53	18	47	2	0
mPB	2007	PB	Trios	123	114	101	43	62	16	0	3	0
mRR	2007-2008	RR	Trios	102	90	71	34	49	17	74	96	0
RS	2007	Borderline	Individus	22	21	17	14	41	18	70	45	0
yRR	2007	RR	Trios	51	51	41	19	45	20	43	93	8
PS	2008	Borderline	Trios	63	60	55	54	76	26	79	21	0
Total				2 785	2 564	2 235	1 024	57	24	55	31	4

Tableau 8 : Périodes de recrutement du laboratoire au Vietnam entre 1991 et 2008

CNG : nombre d'échantillons envoyés au Centre National de Génotypage.

Les génotypes validés correspondent, au nombre de génotypes de bonne qualité qui ont été conservés, une fois les duplicats et triplicats retirés.

La proportion de MB, de RR et d'ENL varie selon le critère de recrutement du projet.

ANALYSE D'ASSOCIATION PANGÉNOMIQUE FAMILIALE

CẨM ỨNG DI TRUYỀN ĐÒI VỚI BỆNH PHONG TẠI VIỆT NAM

Trắc nghiệm cá nhân
BỘ * CQ



Code famille	Số gia đình : HCM 111
Nom et prénom	Họ tên người được nghiên cứu : [Redacted]
Ethnie	1. Cho tất cả mọi người được lấy máu (Cha, mẹ và các con) :
Lieu de naissance	1.1 Họ (chữ thường), tên (chữ hoa) : [Redacted] (con)
Adresse	1.2 Dân tộc : Việt Nam <input type="checkbox"/> Hoa <input checked="" type="checkbox"/> Khác <input type="checkbox"/> Ghi rõ :
Sexe	1.3 Nơi sanh (thành phố, vùng) : [Redacted]
Année de naissance	1.4 Thường trú (thành phố, vùng) : [Redacted]
	1.5 Phái : Nam <input checked="" type="checkbox"/> Nữ <input type="checkbox"/>
	1.6 Ngày, tháng, năm sinh (hay tuổi) : 1981
Identité père	2. Chỉ dành cho bệnh nhân :
Identité mère	2.1 Họ (chữ thường), tên (chữ hoa) của mẹ : [Redacted]
Statut BCG	2.2 Họ (chữ thường), tên (chữ hoa) của cha : [Redacted]
Date du diagnostic	2.3 Sẹo BCG : Có <input type="checkbox"/> Không <input checked="" type="checkbox"/> Không rõ <input type="checkbox"/>
Classification clinicien	2.4 Năm chẩn đoán bệnh : [Redacted] 1964
Nombre de lésions	2.5 Thể bệnh :
Classification RJ	Ít khuẩn (PB) <input type="checkbox"/> Nhiều khuẩn (MB) <input checked="" type="checkbox"/>
Indice bacillaire	Số lượng thương tổn da : 720
Réaction de type 1	I <input type="checkbox"/> T <input type="checkbox"/> B <input type="checkbox"/> L <input checked="" type="checkbox"/>
Réaction de type 2	I <input type="checkbox"/> TT <input type="checkbox"/> BT <input type="checkbox"/> BB <input type="checkbox"/> BL <input type="checkbox"/> LL <input checked="" type="checkbox"/>
Traitement	2.6 Chỉ số vi trùng (BI) lúc chẩn đoán : 2+
	2.7 Phản ứng :
	Thời gian khởi phát
	Trước Trong Sau (ĐT)
	Loại I : Phản ứng đảo ngược : Không <input checked="" type="checkbox"/> Có <input type="checkbox"/> Nếu có <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/>
	Loại II : Phản ứng HBN : Không <input type="checkbox"/> Có <input checked="" type="checkbox"/> Nếu có <input type="checkbox"/> <input checked="" type="checkbox"/> <input type="checkbox"/>
	2.8 Tình trạng bệnh nhân :
	DHTL <input type="checkbox"/> Giám sát <input type="checkbox"/> Ngừng giám sát <input checked="" type="checkbox"/>

Figure 28 : Fiche individuelle utilisée pour le recrutement des individus

La traduction française des renseignements de la fiche est indiquée à gauche.

➤ Description de l'échantillon initial

L'objectif principal du projet a été de réaliser une analyse d'association pangénomique en utilisant des données familiales. Compte-tenu du projet, il a été pertinent d'inclure la totalité des familles recrutées.

Pour chacun des individus de l'échantillon, nous disposons d'ADN extrait à partir de sang total et, des données phénotypiques précédemment décrites. Parmi les 2 785 individus, répartis en 840 « familles », de la base de données, 1 181 sont des cas de lèpre, 1064 des individus sains et 540, principalement des fondateurs, pour lesquels les données phénotypiques sont inconnues ou insuffisamment documentées (NA – Non Attribué).

L'ADN était disponible (ADN de bonne qualité) pour 2 564 individus. Le Centre National de Génotypage (CNG) d'Évry a pu en analyser 2 506. Les données de génotypage, comme la corrélation des génotypes entre individus, nous ont aidés à structurer la population. Par exemple, il est apparu, avant le début de l'analyse, que de nombreux individus figuraient en deux (N = 137) ou trois (N = 18) exemplaires. La principale raison (avant les erreurs techniques - étiquetage, pipetage...) est que certains individus avaient réellement été génotypés plusieurs fois... car recrutés plusieurs fois lors des différentes études menées. Il n'avait pas toujours été possible de faire ce recoupement en amont avec les données patronymiques, compte tenu des interprétations phonétiques variées des différents recruteurs !

Un long travail de correction des duplicats/triplicats a donc été mené. Le problème était celui de données phénotypiques différentes déclarées par un même individu lors de recrutements successifs : nombre de lésions, présence ou non de réaction de réversion, âge au diagnostic... Ceci pouvait provenir d'un biais de mémoire de l'individu, d'erreurs de saisie, ou simplement refléter l'évolution de la maladie. Il s'agissait alors de « fixer » le phénotype de l'individu. Nos collaborateurs Vietnamiens ont alors effectué un minutieux travail d'analyse rétrospective des fiches individuelles, conservées à l'Hôpital de DV d'Ho Chi Minh. Certaines erreurs de saisie ont pu être facilement élucidées (sexe discordant, patient non atteint avec un indice bacillaire positif, date de naissance incompatible avec l'âge au diagnostic...), d'autres ont nécessité un choix concerté avec nos collaborateurs vietnamiens.

ANALYSE D'ASSOCIATION PANGÉNOMIQUE FAMILIALE

De ce fait, après avoir retiré les duplicats/triplicats, notre échantillon comportait 2 235 individus appartenant à 680 « familles ». Nous distinguons plusieurs types de familles : les individus isolés ou singletons (non apparentés), les familles nucléaires (avec un ou deux parents génotypés), les individus non liés (cousins ou enfants probables d'un même couple non génotypé), les trios qui sont majoritaires (deux parents et un enfant atteint), des duos (un parent ou un enfant manquant), les familles multiplex (deux parents et plusieurs enfants avec au moins un atteint), et les familles complexes (comportant plusieurs couples de parents, éventuellement sur plusieurs générations). Nous distinguons ensuite les « enfants » (N = 997) des « fondateurs » (N = 1 238). Les fondateurs sont les individus dont aucun parent n'a été génotypé et les autres sont désignés « enfants », quel que soit leur âge. Ces enfants peuvent eux-mêmes être parents, dans des familles complexes à trois générations par exemple. Parmi les 680 familles, 575 avaient au moins un parent génotypé et, 365 sont des trios (**tableau 9**).

Parmi les 2 335 individus, 939 sont atteints (692 enfants et 247 fondateurs). Parmi les cas, le sex-ratio hommes-femmes chez les enfants est de presque 2 et reflète la prépondérance masculine de la maladie classiquement rapportée. L'âge moyen est de 56 ans (de 4 à 75 ans), plus bas que ce qui est classiquement reporté mais consistant avec le mode de recrutement qui s'est effectué via un enfant atteint et la nécessité de parents vivants. Enfin, la répartition du type de lèpre est de 63% de lèpre MB et 37% PB ce qui est en accord avec les proportions précédemment rapportées au Vietnam (**tableau 10**). Les caractéristiques des fondateurs différaient radicalement des enfants. Ceci est attendu compte tenu du mode de recrutement des familles, principalement à partir d'un ou plusieurs enfants atteints.

Enfin, concernant l'origine des individus, ils sont tous vietnamiens mais peuvent appartenir à des ethnies différentes. Comme attendu puisque c'est l'une des raisons pour lesquelles nous avons choisi cette zone d'étude, une majorité des individus de l'étude se déclare d'origine *Kinh* (principale ethnie), quelques-uns aux minorités *Cham*, *Chn (ou Han)*, *Raglay*, *Lanh* et *Nung (ou Tay)*, enfin, les familles MIX (parents d'ethnies différentes) sont minoritaires (7 individus). Plus précisément, les 1 205 fondateurs majoritairement *Kinh* (N = 1 158), quelques-uns provenaient des minorités *Cham* (11), *Han* (22), *Raglay* (10) et *Nung/Tay* (4), et deux étaient d'origine mixte. De façon intéressante, ces ethnicités auto-reportées par les sujets de l'étude sont très fiables et en très grande majorité confirmées par l'analyse en composante principale (voir méthode) que nous avons réalisée après la phase de génotypage.

ANALYSE D'ASSOCIATION PANGÉNOMIQUE FAMILIALE

Pour ce faire, nous avons vérifié les origines déclarées en calculant les 10 premières composantes principales d'analyse de nos fondateurs. Nous les avons, pour cela, mélangés aux individus de HapMap issus des populations « *Chinois de Beijing* », « *Chinois de Denver* » et « *Japonais* ». Nous donnons une représentation graphique des 2 premières composantes principales (**figure 29**). Les fondateurs *Cham* et *Raglay* se distinguaient des autres ; aucun individu de ces 2 ethnies ne semblait avoir été mal déclaré. Certains individus déclarés d'origine *Kinh* s'approchaient plus des groupes *Han* ou *Cham*. Il est intéressant de voir que notre population était plus proche de la population des *Chinois de Denver* que des *Chinois de Beijing*. Nous pouvons imaginer que parmi les *Chinois de Denver* figuraient des individus issus du Vietnam ou de mariages mixtes sino-vietnamiens, plus que parmi les *Chinois de Beijing*.

De même, nous avons, réalisé une ACP avec les populations de 1000 génomes (1KG) phase 3 (**figure 30**). Cela a encore confirmé la fiabilité des ethnies auto-rapportés, en effet, notre population forme un cluster avec la population KHV (Kinh d'Ho Chi Minh) de 1KG.

Dans notre étude, la variabilité observée n'est pas très importante et ne concerne qu'un nombre très restreint d'individus (rappelons également que l'approche familiale est insensible à la stratification de population et que l'impact d'un mélange de populations s'observerait en présence d'hétérogénéité génétique entre ces populations, c'est-à-dire, des gènes différents gouvernent le phénotype en fonction des populations). En effet, le fait d'avoir des données familiales d'un même pays permet d'être robuste concernant les problèmes de stratification de population souvent rencontrés dans les GWAS cas-contrôles.

➤ **Choix du phénotype étudié**

Nos analyses d'association ont étudié le phénotype binaire atteint Versus non atteint à la lèpre per se. La définition de cas est purement clinique et ce statut a été défini par au moins deux léprologues indépendants après examen des lésions cutanées. Comme dans toutes les études, nous ne pourrions pas exclure que certains diagnostics puissent être erronés, ce qui peut affecter la puissance de notre étude.

ANALYSE D'ASSOCIATION PANGÉNOMIQUE FAMILIALE

<i>Description</i>	<i>N</i>	<i>N (≥ 1 enfant atteint)</i>	
<i>Singletons</i>	89	0	
<i>Familles nucléaires</i>	1 enfant	365	365
	2 enfants	55	55
	3 enfants	68	64
	4 enfants	11	10
	2 parents 5 enfants	14	14
	6 enfants	1	1
	7 enfants	3	3
	8 enfants	1	1
	11 enfants	1	1
	1 parent 1 enfant	26	26
	2 enfants	11	10
	3 enfants	2	2
	4 enfants	2	2
	12 enfants	1	1
<i>Individus non liés</i>	16	0	
<i>Familles complexes</i>	14	14	
<i>Total</i>	680	569	

Tableau 9 : Répartition des 2 235 individus génotypés selon le type de famille

	<i>Fondateurs N (%)</i>	<i>Enfants N (%)</i>	<i>Total (%)</i>
<i>Total</i>	1238	997	2 235
Hommes	634	651	1 285
Femmes	604 (49)	346 (35)	950 (43)
<i>Cas de lèpre</i>	247	692	939
Âge moyen au diagnostic	37,3	18,3	55,6
Répartition par groupe d'âge			
4-9	3 (1)	57 (8)	60 (6)
10-19	21 (8)	367 (53)	388 (42)
20-59	191 (77)	266 (39)	457 (50)
≥ 60	17 (4)	0 (0)	17 (2)
PB (clinicien)	54 (25)	263 (41)	317 (37)
MB (clinicien)	164 (75)	378 (59)	542 (63)
Réaction	58 (23)	250 (36)	308 (33)
Type 1 (RR)	49 (20)	232 (34)	281 (30)
Type 2 (ENL)	12 (5)	26 (4)	38 (4)

Tableau 10 : Caractéristiques phénotypiques de l'échantillon initial

ANALYSE D'ASSOCIATION PANGÉNOMIQUE FAMILIALE

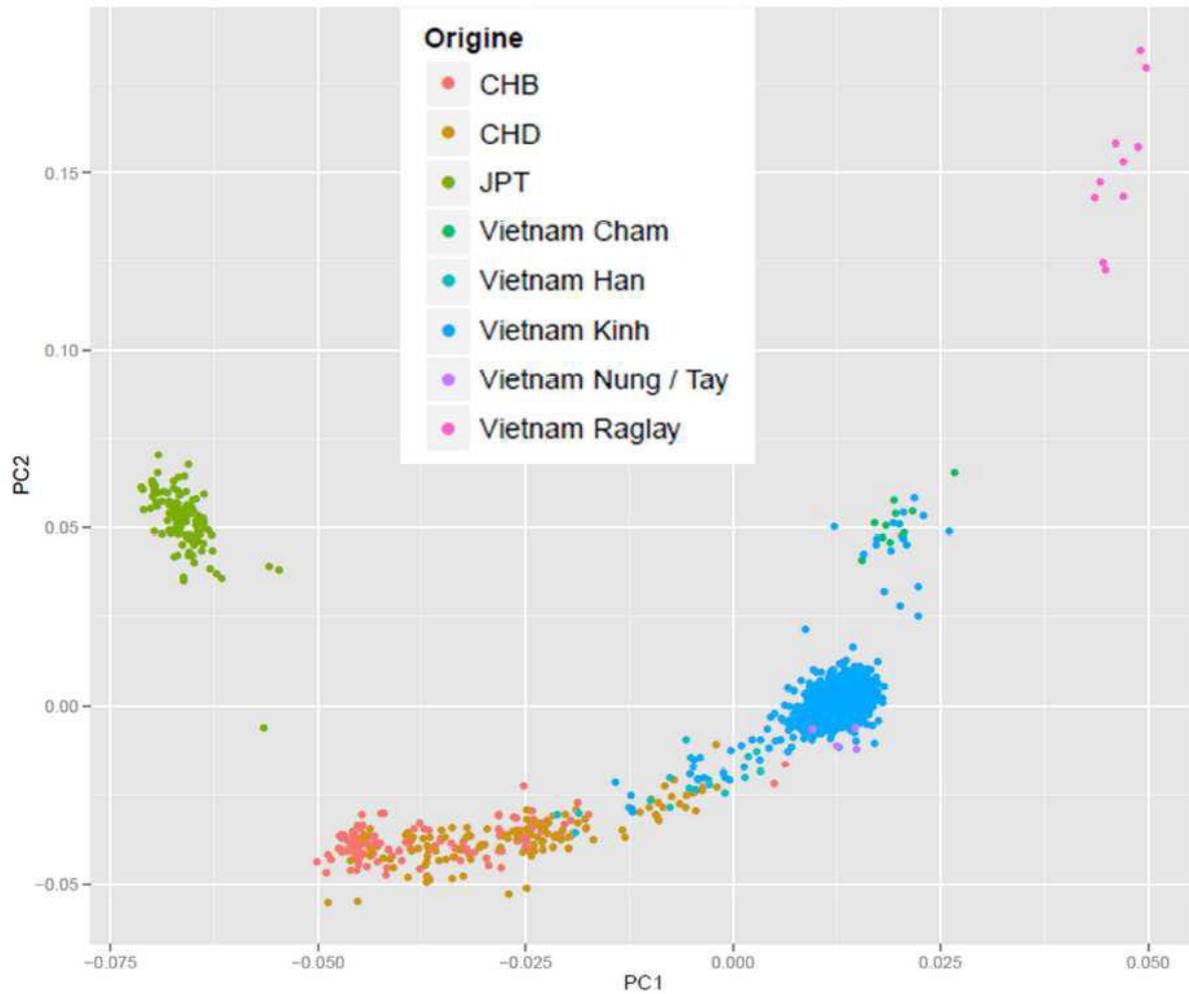


Figure 29 : ACP des individus fondateurs par origine ethnique (populations HapMap)

ACP selon la première (axe des x) et la deuxième (axe des y) composantes principales en incluant les populations de HapMap les plus proches génétiquement (CHB = Chinois Han de Beijing, CHD = Chinois de Denver aux États-Unis, JPT = Japonais).

ANALYSE D'ASSOCIATION PANGÉNOMIQUE FAMILIALE

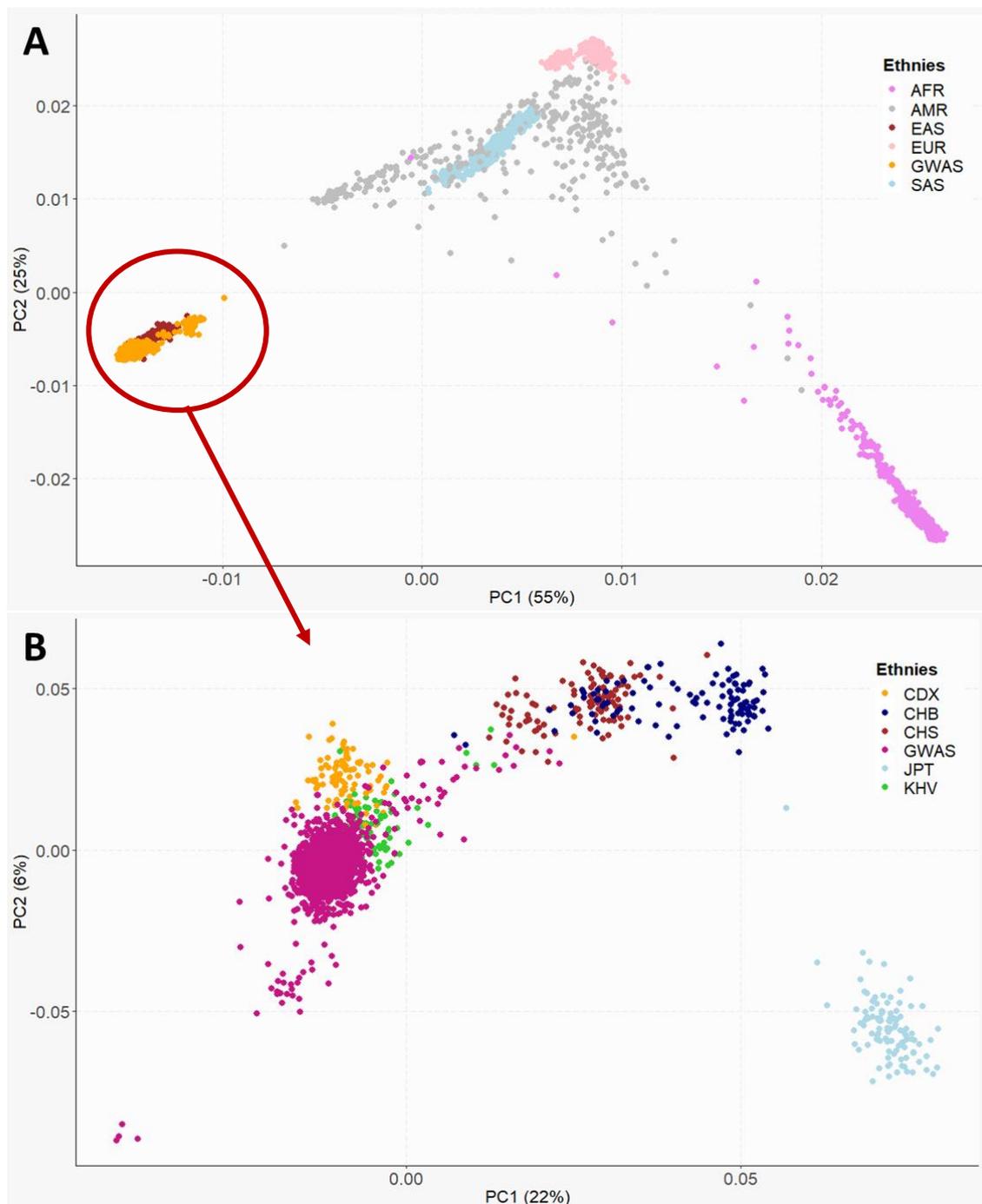


Figure 30 : ACP du GWAS avec les populations de 1000 génomes Phases 3

A) Toutes les populations de 1KG phase 3.

B) Les populations d'Asie de l'Est e 1KG phase 3.

AFR = Africain ; AMR = Américain ; EAS = Asie de l'Est ; EUR = Européen ; GWAS = notre échantillon ; SAS = Asie du Sud ; CDX = Chinois Dai de Xishuangbanna ; CHB = Chinois Han de Beijing ; CHS = Chinois Han du Sud ; JPT = Japonais de Tokyo ; KHV = Kinh de Ho Chi Minh City au Vietnam.

2.2 Données génétiques

2.2.1 Génotypage

➤ Principe général de génotypages par puces à ADN

Bien que sa biochimie exacte varie d'une plateforme à l'autre, le principe du génotypage à haut-débit par puce à ADN reste constant. L'objectif des puces de génotypage pangénomique à ADN est la connaissance de l'identité allélique des SNPs présents sur l'ensemble du génome d'un organisme. Le génotypage des SNPs se fonde sur le principe d'hybridation : deux fragments d'acides nucléiques complémentaires peuvent s'associer et se dissocier de façon réversible sous l'action de la chaleur et de la concentration saline du milieu. Concrètement, une puce à ADN est un support rigide de verre de quelques centimètres carrés, sur lequel de courtes séquences d'ADN spécifiques des SNPs (appelées sondes) ont été déposées. Ainsi, après amplification, fragmentation et dénaturation du génome, les fragments d'ADN (ou cibles) vont pouvoir s'hybrider sur les sondes. S'en suit une étape d'extension à partir de la position du SNP où chaque allèle du SNP sera marqué par des fluorochromes ayant une couleur différente. Après, cette étape expérimentale, les données vont être traitées et analysées avec, tout d'abord, une acquisition des images d'hybridation. De ce fait, l'intensité de fluorescence calculée reflète la séquence du SNP ainsi que son niveau relatif (**figure 31**).

Plus précisément, le nucléotide préalablement marqué par un fluorochrome émet un signal lumineux qui est enregistré. C'est l'analyse de l'intensité des couleurs correspondant aux différents allèles du SNP qui permet la détermination du génotype de l'individu à ce SNP. Néanmoins, les intensités lumineuses émises et mesurées dépendent de nombreux paramètres (plateforme, lots de réactifs...), et l'assignation d'un génotype se fait de manière graphique par la représentation des intensités lumineuses de chaque allèle pour l'ensemble des individus génotypés au même moment. La base des algorithmes de génotypage consiste en l'identification de 3 groupes de points correspondant aux trois génotypes possibles AA, AB et BB. Les points intermédiaires entre les groupes ne sont pas assignés à un groupe. Un échec du génotypage à un SNP peut se manifester de plusieurs façons. La répartition des intensités lumineuses peut être relativement uniforme, aboutissant à une définition peu précise des groupes, et donc un taux de génotypage faible.

ANALYSE D'ASSOCIATION PANGÉNOMIQUE FAMILIALE

Par ailleurs, certaines valeurs extrêmes d'intensité peuvent tromper l'algorithme et aboutir à la fusion du groupe hétérozygote et d'un groupe homozygote. Dans ce cas, il y aura un excès d'hétérozygotes qui provoque un écart artéfactuel à l'équilibre de Hardy-Weinberg (**figure 32**). Enfin, les SNPs ayant une faible fréquence de l'allèle mineur sont caractérisés par l'absence fréquente du groupe des homozygotes pour l'allèle mineur, ce qui rend les algorithmes de génotypage moins fiables. Par ailleurs, la puissance pour détecter un effet de ces allèles dans le cadre de l'étude d'association pangénomique est faible.

Ces considérations amènent à définir et à réfléchir sur trois critères classiques et essentiels quant à la qualité des marqueurs dans le cadre d'une étude d'association pangénomique : le taux de génotypage (seuil généralement entre 95 et 99%), la fréquence de l'allèle mineur (seuil généralement entre 1 et 5%) et un écart majeur à l'équilibre de Hardy-Weinberg (seuil de p-value généralement entre 10^{-4} et 10^{-7}) (voir contrôle qualité).

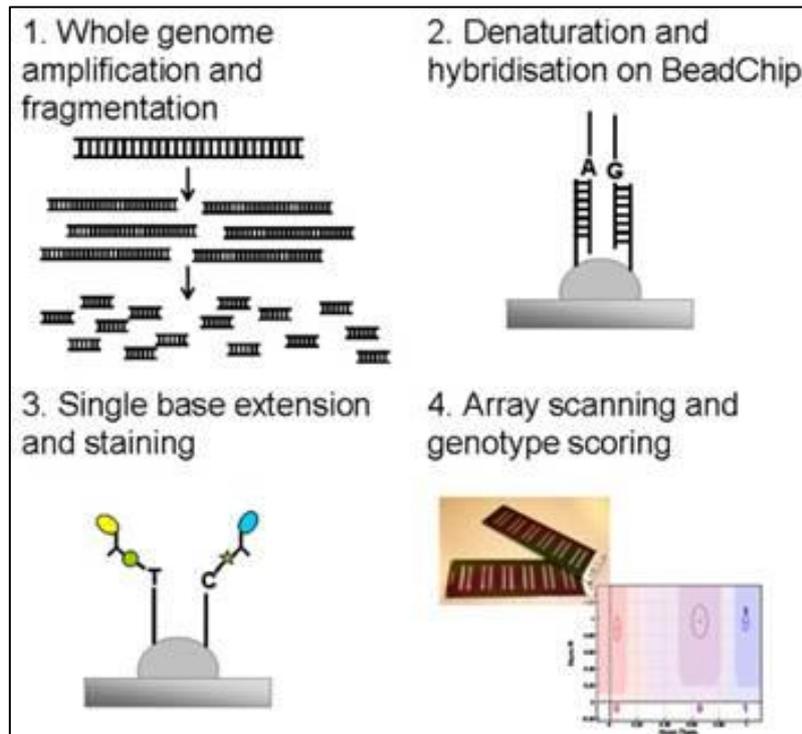


Figure 31 : Schéma général du principe des puces à ADN de génotypage [395]

- 1) Après amplification, fragmentation et dénaturation du génome,
- 2) Les cibles vont pouvoir, après dénaturation, s'hybrider sur les sondes,
- 3) Il s'en suit une étape d'extension à partir de la position du SNP,
- 4) Enfin, les données seront traitées et analysées par bio-informatique.

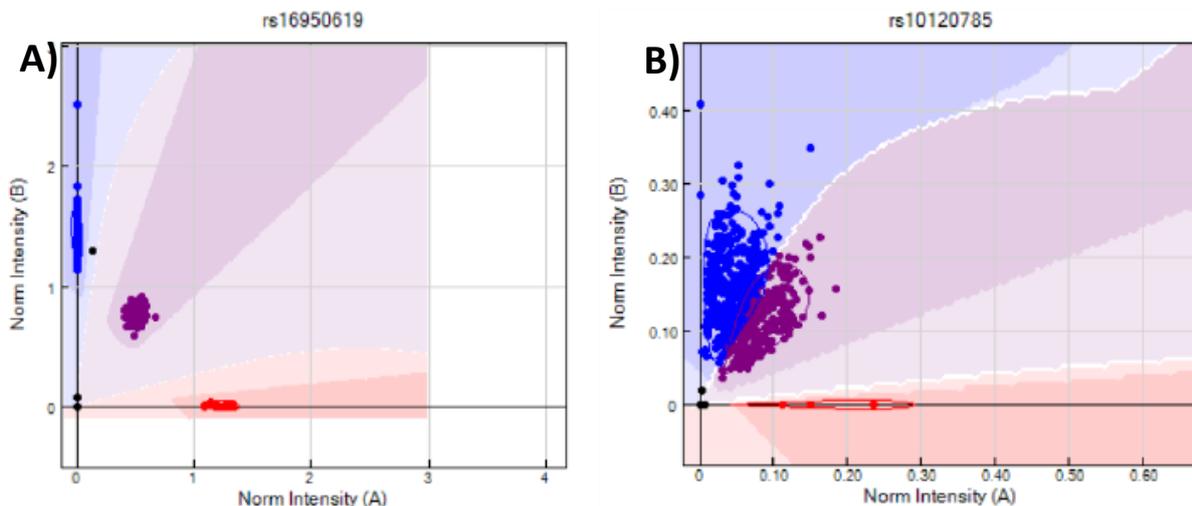


Figure 32 : Principe de l'assignation de génotypes par puce à ADN

- A) Exemple d'un SNP avec assignation correcte. En abscisse, l'intensité d'hybridation normalisée pour l'allèle A et en ordonnée celle de l'allèle B. Trois groupes se distinguent clairement et permettent l'assignation d'un génotype automatisée à la quasi-totalité des individus.
- B) Exemple d'un SNP avec assignation problématique. La difficile séparation des groupes résulte en un écart artificiel à l'équilibre de Hardy-Weinberg.

➤ Génotypage des individus de notre échantillon

Nous avons utilisé la puce de génotypage *Human 660W-Quad* fabriquée par *Illumina* qui comporte 657 366 SNPs, dont 640 663 sur les chromosomes autosomes (n'incluant pas les chromosomes sexuels ni l'ADN mitochondrial) et permet un débit de traitement contenant jusqu'à 12 échantillons simultanément. Cette puce a été conçue pour optimiser la couverture des variations du génome chez les Caucasiens. Cette optimisation est fondée sur deux bases de données, HapMap et 1000 Génomes. HapMap dresse une carte haplotypique du génome humain qui décrit les variations courantes dans la séquence de l'ADN humain [396]. Cette carte repose sur le génotypage d'environ 4 millions de variations connues dans onze populations différentes d'Afrique, d'Europe, d'Asie, d'Inde et du Mexique. Le projet 1000 Génomes a lui pour but d'identifier « toutes » les variations du génome humain, en particulier les plus rares qui n'ont pas été génotypées par le projet HapMap. À ce jour, et issu de la phase 3 de 1000G, environ 30 millions de variations ont été décrites chez 3 202 individus issus de 26 populations différentes [397] La puce couvre presque 90% des variants fréquents ($MAF \geq 5\%$) décrits chez les Caucasiens avec un $r^2 > 0,8$.

Même si nous ne disposons pas de données de couverture de la puce *Human 660W-Quad* pour notre population vietnamienne, elle est a priori élevée. En effet, 85% des variations communes référencées dans les bases de données HapMap au niveau des populations chinoises et japonaises, sont couvertes par la puce avec un $r^2 > 0,80$ (**figure 33**). Ces pays sont proches géographiquement du Vietnam, et un lien entre distance géographique et génétique a été établi, en particulier pour les populations Européennes [398]. Ainsi, et à cette époque, cette puce était la plus adaptée pour notre population. La médiane d'espacement entre 2 SNPs consécutifs sur la puce est de 2,3 kilobases (kb) (**tableau 11**). Compte-tenu du fort taux de recombinaison dans la région du CMH, la densité de SNPs de la puce y est plus élevée : 3 177 SNPs couvrent les 4 Mb de la région, soit une distance moyenne de 1,3 kb entre 2 SNPS. La puce comporte un grand nombre de Copy Number Variants (CNV), polymorphismes non bi-alléliques comme les SNPs. Nous les avons donc exclus de nos analyses. Ainsi, les échantillons d'ADN extraits, à partir de sang total à l'Hôpital de DV, ont été génotypés en 2010 au Centre National de Génotypage (CNG) d'Évry. Au total, après cette étape de génotypage, notre échantillon d'étude se compose de 592 633 SNPs uniformément réparties sur tous les autosomes ainsi que sur le X pour 2 335 individus.

ANALYSE D'ASSOCIATION PANGÉNOMIQUE FAMILIALE

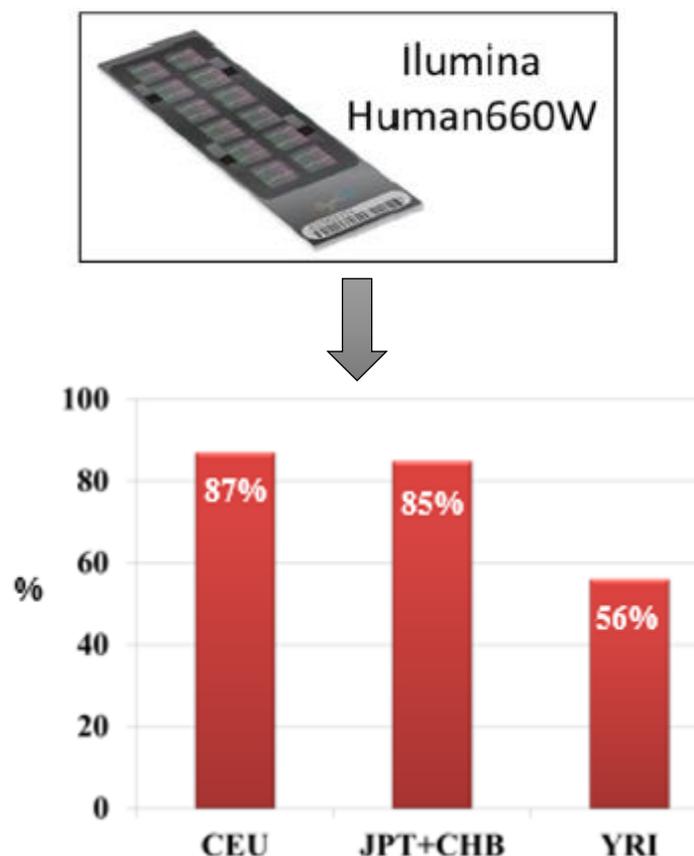


Figure 33 : Biopuce Illumina Human 660W-Quad et sa couverture à $r^2 > 0,8$ dans différentes populations de la base de données HapMap

CEU = ancêtres d'Europe du Nord et de l'Ouest (Utah) ; JPT+CHB = Japonais (Tokyo) + Chinois (population Han à Beijing) et YRI = Nigériens (population Yoruba à Ibadan).

À partir de plus de 2,3 millions de variants communs.

Illumina Human660W-Quad	
SNPs (N)	657 366
Couverture génomique (CHN-JAP)	86% avec $r^2 > 0,8$
Distance médiane entre 2 SNPs	2,3 kb
90^{ème} percentile du plus grand espace entre 2 SNPs	10,6 kb
SNPs non-synonymes (N)	10 051
SNPs du CMH/Insertions-délétions (N)	3 177/0
SNPs sur gonosomes (X/Y/PAR)	16 509/44/15
SNPs mitochondriaux	135

Tableau 11 : Caractéristiques de la puce Illumina Human 660W-Quad [399]

2.2.2 Le contrôle qualité

Le contrôle qualité (ou QC) est une étape essentielle du traitement de l'information génétique, au même titre que la validation des données phénotypiques puisque l'objectif final est de lier les deux. Il est donc essentiel avant de pouvoir réaliser les tests statistiques de rendre les données plus homogènes par individus et par marqueurs et donc comparables entre elles en minimisant les sources de variations causées par l'expérimentation (exemple : perte d'information génotypique ou apparition d'erreurs mendéliennes dues aux scanners). D'autre part, les variants et individus vont avoir des valeurs types à vérifier comme le taux d'hétérozygotie. De fait, le QC va éliminer les effets qui pourraient fausser nos résultats par une inflation d'erreurs de type 1 c'est-à-dire de faux positifs (FP) correspondant à des fausses associations mais aussi un manque de puissance.

Dans notre cas, nous avons réalisé le QC de nos données en 2 étapes : d'abord sur les marqueurs puis sur les individus. Cet ordre n'est pas neutre quant à la sélection finale des marqueurs et des individus qui vont être emmenés à l'étape suivante.

➤ **Marqueurs**

Tout d'abord, nous avons vérifié la qualité des 592 633 SNPs génotypés sur 5 critères :

- **Callrate (CR) : taux de génotypage par marqueur**

Pour un SNP donné, le Callrate (CR) correspond à la proportion d'individus pour lequel un génotype a pu être assigné (on remarquera qu'un assignement n'est pas synonyme d'un assignement correct, c'est d'autres éléments du QC qui permettront de nous aider en ce sens, (e.g. l'équilibre de Hardy Weinberg) : cette valeur mesure la quantité de données manquantes. Dans notre étude, seuls les SNPs avec un CR supérieur ou égal à 95% ont été conservés pour la suite des analyses, ce qui se traduit par le fait que pour chaque marqueur gardé dans l'analyse au moins 95% des individus ont été génotypés. En effet, un taux de génotypes manquants supérieur à 5%, pour un SNP donné, indique un non assignement qui peut être dû par exemple à une erreur de conception de la sonde sur la puce à ADN et rend incertain la validité des génotypes attribués aux autres individus

- **Differential missingness (DM)**

Le DM teste pour chaque SNP s'il existe une différence significative du CR entre les individus atteints et non atteints. C'est un raffinement du QC précédent : en effet, un DM significatif peut entraîner aussi bien une inflation de l'erreur de type I qu'un manque de puissance. Dans notre étude principale qui est, rappelons-le, familiale, ce critère n'est pas pertinent (même si on pourrait imaginer un DM entre l'enfant et ses parents). Il n'a donc été une mesure du QC que pour l'analyse de répllication.

- **Minor allele frequency (MAF)**

La MAF indique la fréquence de l'allèle mineur d'un SNP donné dans une population donnée, ici la population est représentée par les fondateurs de notre échantillon. Du point de vue « théorique », cette valeur permet de classer un variant comme rare ou fréquent à partir d'un seuil donné. Plus important, en pratique la qualité du génotypage pour un SNP donné est influencée par sa MAF, en effet, plus la MAF est petite plus il est difficile de mettre en évidence la présence d'homozygote rare. Cette étape permet donc d'éliminer les SNPs monomorphes donc non informatifs pour l'étude d'association proprement dite. La MAF a été estimée uniquement chez les fondateurs afin de ne pas biaiser les résultats dus au fait que les parents homozygotes vont forcément transmettre un allèle donné et pas l'autre. Pour la suite de notre étude, nous avons arbitrairement sélectionné les SNPs ayant une MAF supérieure à 5%.

- **Équilibre de Hardy-Weinberg (HW)**

L'équilibre de HW est un autre élément important du QC : en effet, il s'agit d'un indicateur de la qualité d'un marqueur dans la population. Le théorème d'HW stipule que dans une population neutre, les fréquences alléliques et génotypiques demeurent constantes de génération en génération en l'absence d'autres influences évolutives. Il existe donc un équilibre entre ces deux fréquences, et tout SNP qui dériverait de cet équilibre est suspect, ce d'autant plus qu'en l'absence d'explications comme le mélange de population au sein de l'échantillon, le motif le plus fréquent d'une déviation à HW est un problème technique lors de la phase de génotypage. Ce test n'a été réalisé que chez les fondateurs non atteints : en effet, l'équilibre de HW est supposé vrai chez les non atteints mais pas forcément chez les atteints ou certains SNPs peuvent avoir été sous pression de sélection. De même, le fait de ne

ANALYSE D'ASSOCIATION PANGÉNOMIQUE FAMILIALE

considérer que les fondateurs permet d'éliminer le biais de fréquence allélique dû à la transmission des parents homozygotes aux enfants. Nous avons conservé pour la suite de l'étude les SNPs ayant une p-value supérieure au seuil de 10^{-5} .

- **Erreurs mendéliennes (EM)**

Le taux d'EM est un critère de QC spécifique aux études familiales de types TDT. Par EM, on entend une incohérence entre le génotype des parents et celui des enfants : prenons un exemple d'un SNP bi-alléliques A/B où les deux parents sont homozygotes AA avec un enfant hétérozygote AB, on considère donc qu'il y a une EM. Ces EM sont quasi exclusivement causées par des erreurs de génotypage (chez l'enfant ou les parents) et dans une très moindre mesure par des mutations de novo. À l'inverse des autres critères de QC, il est inféré uniquement à partir des enfants. Notre étude a confirmé que ce critère est un élément non seulement spécifique mais également majeur quant à son impact sur les résultats, en particulier en augmentant l'erreur de type I. Nous avons décidé d'un seuil de 25 EM maximum pour conserver un SNP dans l'analyse. Ce seuil est arbitraire même s'il dérive d'une étude que nous avons faite mais qui dépasse le cadre de ce manuscrit. Toutefois, si l'on considère que l'on a environ 700 familles de type trios, il est possible d'avoir 1400 EM par SNP (deux par enfant), ainsi 25 EM correspond à 1,8% d'erreurs.

Nos valeurs de filtres sont dans la continuité de celles rapportées dans différentes études [400]. Au total et après QC, nous disposons de 422 546 SNPs dont 9 907 SNPs sur le chromosome X chez 2 235 individus, dont 692 enfants atteints.

➤ **Individus**

Suite au QC sur les marqueurs, le QC a été réalisé sur les 2 335 individus de l'étude. Une première étape de QC individus avait été réalisée avant mon arrivée au laboratoire (par exemple l'identification des doublons et le réajustement des relations familiales si nécessaire), il a été néanmoins intéressant et pédagogique de le refaire en partie. Pour cela, nous avons également pris en compte cinq conditions :

ANALYSE D'ASSOCIATION PANGÉNOMIQUE FAMILIALE

- **Callrate**

De même que le CR sur les marqueurs, le CR sur les individus correspond à la proportion de SNPs génotypés pour un individu donné. Nous avons également choisi un seuil de 95% pour conserver un individu dans l'analyse. Un taux important de génotypes manquants indique généralement une mauvaise qualité de l'ADN et/ou un problème dans la procédure de génotypage et doit amener à douter de la validité des génotypes pour l'ensemble des autres SNPs chez cet individu.

- **Taux d'hétérozygotie (HZ)**

Le taux d'hétérozygotie reflète le nombre de marqueurs à l'état hétérozygote chez un individu donné. Cette valeur est à peu près similaire entre individus dans la population : ainsi, nous regardons pour chaque individu, s'il a un HZ aberrant. Les individus avec un taux excessif (ou trop faible) de génotypes hétérozygotes peuvent être le signe de contamination des échantillons ou de consanguinité. Plus précisément, il faut calculer le taux d'hétérozygotie pour chaque individu et éliminer ceux qui ont un taux d'hétérozygotie supérieure ou inférieure au taux moyen sur l'ensemble des individus plus ou moins 3 fois l'écart-type [401]. Il est également possible de tracer l'histogramme du HZ par individu et de supprimer ceux qui apparaissent visuellement donc subjectivement comme des outliers.

- **Vérification des sexes déclarés**

Cette étape permet de contrôler la cohérence du sexe des individus renseignés avec le sexe des individus déterminé génétiquement. Il s'agit d'un critère essentiel de fiabilité pour l'ensemble de l'échantillon (qualité du recrutement). Une absence de correspondance indique potentiellement l'inversion de plusieurs prélèvements, et remet donc en cause le phénotype enregistré. Le sexe d'un individu peut être inféré par l'analyse des marqueurs du chromosome X et Y. La méthode la plus simple consiste à mesurer le taux d'hétérozygotie du chromosome X, présent en une copie chez l'homme (haploïde, taux d'hétérozygotie attendu nul) et en deux copies chez la femme (diploïde, taux d'hétérozygotie attendu non-nul). Ainsi, le taux moyen d'homozygotie le long du chromosome X est supérieur à 0,8 pour un homme et inférieur à 0,2 pour une femme. Cette étape déjà réalisée auparavant n'a pas identifié de nouveau problème qui serait passé inaperçu.

ANALYSE D'ASSOCIATION PANGÉNOMIQUE FAMILIALE

- **Vérification des doublons/triplets et des liens familiaux**

L'étape de vérification des doublons et triplets a été réalisée en calculant la matrice « Identity By State » (IBS), elle a également été utile pour redéfinir certains liens familiaux. L'IBS se base sur le nombre d'allèles que peuvent partager en commun deux individus, pour un SNP donné. Deux individus peuvent donc : ne pas partager d'allèles, partager un allèle ou partager les deux allèles en commun. Chaque case de la matrice calculée correspond à la proportion d'allèles identiques entre 2 individus, ainsi, une valeur d'IBS > 0,99 entre deux individus signait une identité. Nous avons ensuite calculé la proportion d'allèles partagés « Identical By Descent » (IBD) (allèles provenant d'un même ancêtre). Cette proportion a permis d'identifier des liens de parenté ignorés (parent-enfant ou frère-sœur), afin de reconstruire les familles nucléaires. Certaines filiations déclarées ont ainsi été réfutées, car incompatibles avec les données génotypiques, et d'autres découvertes. Nous avons ainsi pu commencer le travail d'analyse génétique proprement dit sur une base nettoyée et clairement annotée.

- **EM par individu**

Enfin, nous avons contrôlé le nombre d'EM par individu à un seuil maximum de 10000 erreurs correspondant à 1% d'erreur si on considère nos 592 633 SNPs avec pour chacun des marqueurs deux erreurs possibles par individu.

Après contrôle qualité individuel, l'échantillon est donc composé de 1 749 individus comprenant 481 familles informatives dont 622 enfants atteints. L'âge moyen global est de 18 ans chez les enfants et de 40 ans chez les founders (**tableau 12**).

ANALYSE D'ASSOCIATION PANGÉNOMIQUE FAMILIALE

Caractéristiques	Étude primaire	
	Founders	Enfants
N Statut	938	811
N (atteint/non atteint)	120/ 818	622/189
N (MB/PB/NA)	60/29/31	202/327/93
Âge en années : moyenne (SD)		
Atteint	40,8 (13,1)	18,6 (7,8)
Non atteint	-	-
% d'hommes		
Atteint	73,3	69,3
Non atteint	46,2	58,2

Tableau 12 : Caractéristiques cliniques de l'échantillon familiale primaire après QC

Les familles sont réparties en 481 familles informatives : 379 trios ; 80 familles multiplex avec au moins deux enfants atteints et deux parents ; 9 familles multiplex avec au moins deux enfants atteints et un parent ; 5 familles multiplex avec au moins deux enfants atteints ; un parent et au moins un enfant non atteint ; 1 famille multiplex avec quatre enfants atteints et plus de deux enfants non atteints mais pas de parents ; 6 familles composées d'un enfant atteint ; d'un parent et d'au moins un enfant non atteint ; et 1 famille trigénérationnelle composée de quatre enfants atteints et quatre enfants non atteints.

Afin de réaliser ces différentes étapes du QC (**carte mentale 2**), le logiciel et le langage de programmation principalement utilisé est R version 3.1.3, c'est un environnement de développement gratuit pour le traitement de données et l'analyse statistique [402]. Par ailleurs, Bioconductor, plateforme gratuite et open source, a été utile car elle contient les nombreux packages R dédiés à l'analyse statistique de données génomiques ainsi que le CRAN (Comprehensive R Archive Network). Plus précisément, nous avons utilisé package « Gaston » d'une part et d'autre part, nous avons utilisé le logiciel Plink [403,404].

2.2.3 Imputation de génotypes additionnels

L'imputation de génotypes consiste à prédire le génotype de marqueurs non directement génotypés chez un ou plusieurs individus [405]. Un panel d'haplotypes de référence, typiquement le panel 1000 génomes, extrêmement dense en marqueurs, est utilisé pour prédire les génotypes absents d'un groupe d'individus génotypés pour un sous-groupe des marqueurs du panel de référence (**figure 34**) [406]. Les génotypes sont imputés sous la forme d'un vecteur de probabilité estimant la probabilité que l'individu soit AA, AB ou BB au SNP imputé sachant ses génotypes alentour et connaissant les haplotypes de référence. IMPUTE2 semble être la méthode la plus efficace en termes de rapidité d'exécution et de qualité d'imputation [407]. Par ailleurs, un grand gain computationnel sans perte significative de performance est obtenu en fournissant directement à IMPUTE2 les haplotypes des individus à imputer, plutôt que leurs simples génotypes [408]. La méthode actuellement recommandée par les auteurs d'IMPUTE2 est SHAPEIT2 [409]. Cette méthode est également celle sélectionnée par le projet 1000 génomes.

Ainsi, avant la phase d'imputation, nous avons, supprimé les SNPs A/T ou G/C afin d'éviter les discordances entre notre échantillon et les panels de référence utilisés pour l'imputation. Nous avons, ensuite, phasé nos haplotypes avec SHAPEIT2 [410] en utilisant l'option duoHMM (sauf pour le chromosome X) permettant d'intégrer les pedigrees familiaux connus et donc d'améliorer la qualité du phasage [411] (excepté pour le chromosome X pour lequel cette option n'est pas applicable). Enfin, nous avons imputé les génotypes manquants avec IMPUTE2 [412]. L'imputation a été réalisée à partir du panel de référence 1000 génomes phase 3 (version d'Octobre 2014) avec les paramètres par défaut et un nombre d'haplotypes de référence maximal (khap = 2 184, fenêtres de 5 Mb).

Nous avons obtenu les génotypes de plus de 60 millions de SNPs et INDELS (insertion ou délétion). En supprimant les variants monomorphes et en appliquant un seuil d'information d'imputation (Info) > 0,9 et une MAF > 5%, plus de 5 millions de variants supplémentaires et, de grandes qualités, ont été ajoutés à notre étude.

ANALYSE D'ASSOCIATION PANGÉNOMIQUE FAMILIALE

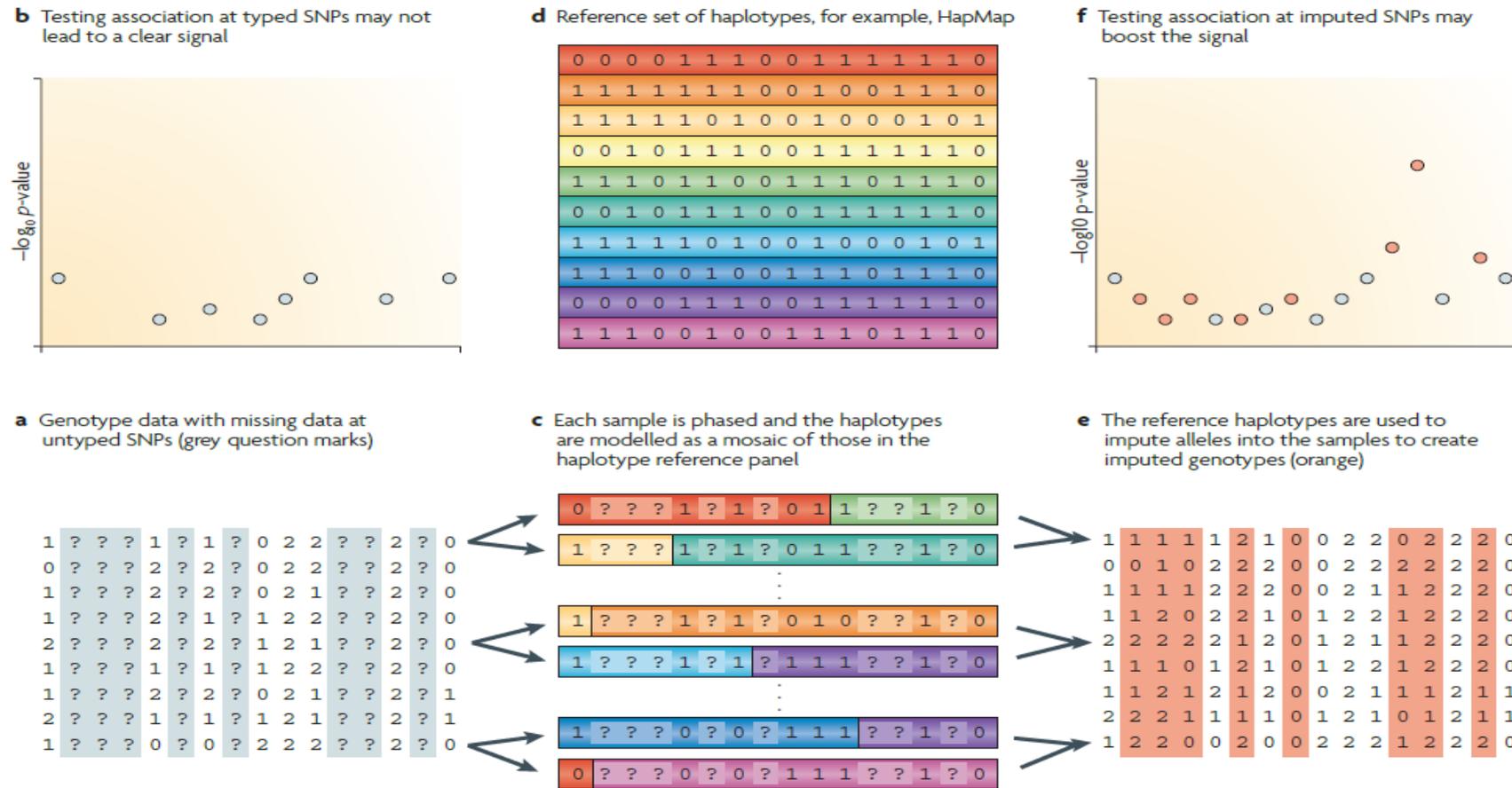
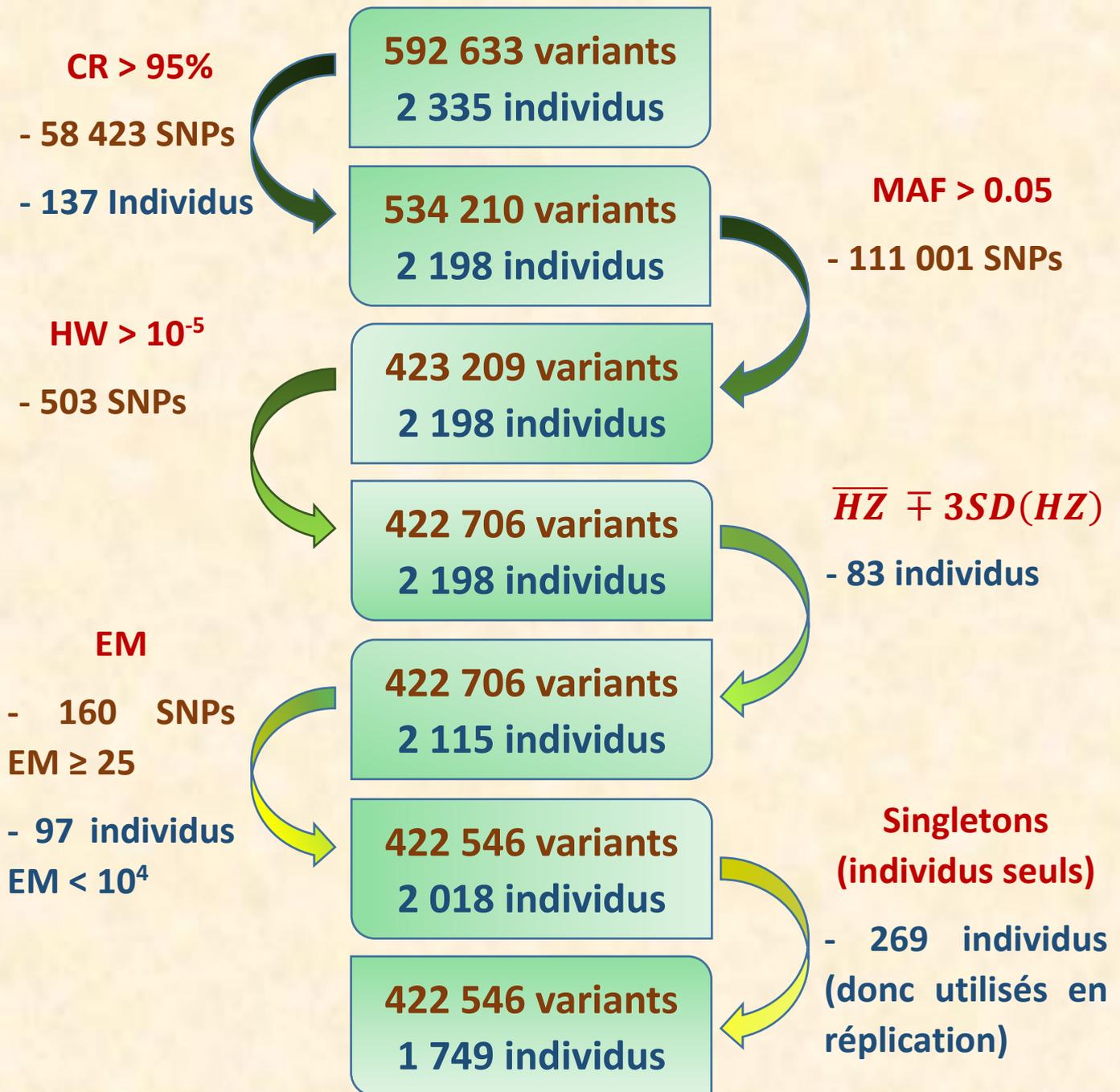


Figure 34 : Principe de l'imputation de génotypes [405]

- a) Un ensemble d'individus est génotypé pour un certain nombre de SNPs
- b) dans le cadre d'une étude d'association.
- c) Pour imputer les génotypes inconnus, ces génotypes sont phasés
- d) puis comparés à des haplotypes de référence
- e) pour inférer les génotypes probables aux positions manquantes.
- f) Ceci peut notablement améliorer le signal d'association.

En réalité, le tableau (e) est plus complexe puisqu'il contient pour chaque individu et chaque SNP le vecteur des probabilités des trois génotypes possibles.



Imputation des génotypes manquants

- ⇒ Panel de référence 1000 génomes phases 3
- ⇒ +60 millions de variants SNPs et INDELS
- QC → INFO > 0,99 & MAF > 0,05**
- ⇒ **5 607 170 variants & 1 749 individus**

Carte mentale 2 : Contrôle qualité de l'échantillon primaire (marqueurs et individus)

2.3 Méthodologie

2.3.1 Analyse en composante principale

Afin de détecter la présence d'une structure de population due aux différentes sous-populations d'un échantillon, l'analyse en composante principale (ACP) d'une matrice mesurant la similarité génétique entre individus, est une méthode très utilisée en épidémiologie génétique. Il s'agit d'une méthode faisant partie des statistiques descriptives multidimensionnelles qui consiste à représenter simultanément un nombre important de variables quantitatives dans un espace de dimension réduit en déformant le moins possible la réalité. Elle permet donc d'obtenir le résumé le plus pertinent possible des données initiales [413]. Les axes recherchés maximisent l'inertie en étant orthogonaux, les facteurs explicatifs, ou composantes principales (PC), correspondent à des combinaisons linéaires des variables et permettent la représentation des individus.

Une matrice d'apparentements basée sur la ressemblance génotypique est calculée, cela nécessite une matrice de génotypes centrés et réduits : le fait de centrer ($\mu=0$) ne modifie pas la forme du nuage, c'est seulement une translation. La réduction ($\sigma^2=1$) permet d'avoir des variables sans unités donc comparables et de mêmes importances. Plus précisément, pour résumer la ressemblance génétique de n individus sur m marqueurs (matrice X de dimension $m*n$ des génotypes centrés et réduits), une réduction rapide et importante de dimension est obtenue en estimant la covariance des génotypes pour chaque paire d'individus (matrice de variance-covariance Σ , proportionnelle à $X^T X$, de dimension $n*n$, symétrique). Deux individus dont les génotypes sont systématiquement proches auront une valeur de covariance fortement positive ; à l'inverse deux individus dont les génotypes sont systématiquement opposés auront une valeur de covariance fortement négative. La matrice XX^T peut ensuite être reformulée comme le produit d'une matrice V de n vecteurs propres indépendants et d'une matrice diagonale S^2 de n valeurs propres ou eigenvalues ($X^T X = VS^2 V^T$). La matrice V contient les coordonnées de chaque individu sur les n CP et la matrice S^2 contient l'importance relative des différentes composantes. La matrice d'origine X peut elle-même être exprimée comme le produit USV^T , avec U la matrice de dimension $m*n$ des coordonnées de chaque marqueur sur les n CP (également appelés poids ou loadings du marqueur sur chaque CP).

ANALYSE D'ASSOCIATION PANGÉNOMIQUE FAMILIALE

Cette ACP, ou décomposition en valeur singulière de la matrice X, est à la base de l'approche EIGENSTRAT [414,415]. Le principe géométrique sous-jacent à cette approche est la recherche de l'axe de projection qui conserve au maximum la dispersion des données (1^{ère} CP), puis d'un second axe de projection orthogonale conservant au maximum la dispersion des données (2nd CP) et ainsi de suite. Il s'agit d'un changement de base dans un espace à n dimensions, tel que les premiers axes soient les plus informatifs. Chaque individu, initialement caractérisé par m génotypes résumés en n covariances, est maintenant caractérisé par n coordonnées sur n CP, combinaisons linéaires des covariances initiales (elles-mêmes combinaisons des génotypes initiaux). Les CP sont ordonnées en fonction de leur valeur propre, une mesure de la proportion de la dispersion initiale qu'elles capturent. En acceptant une perte d'information minimisée, les données peuvent donc être tronquées à p premières CP, ce qui constitue la deuxième étape importante de réduction de dimension de la matrice initiale X.

Autrement dit, l'information génétique relative d'un individu au sein d'un groupe peut être décrite exhaustivement par son génotype à m marqueurs ou résumée par ses coordonnées sur p CP ($p \sim 10$). Chaque CP représente un axe continu de variation génétique, et sépare des groupes génétiquement homogènes : les individus appartenant à la population A auront par exemple des coordonnées négatives sur la CP 1 et ceux appartenant à la population B des coordonnées positives.

La capacité à séparer des individus par groupe sur les premières CP dépend fortement des niveaux de structure présents dans l'échantillon. Ainsi, une ACP réalisée sur un échantillon composé de populations très hétérogènes (Europe et Afrique par exemple) ne séparera pas les sous-populations de chaque groupe de manière optimale (différentes populations africaines par exemple) [416]. À une échelle locale, la présence de quelques individus fortement reliés, comme les membres d'une même famille, peut accaparer les premières CP et dissimuler la structure génétique d'intérêt [414]. Il est donc préférable d'identifier a priori les individus fortement liés entre eux et de réaliser l'ACP sur le sous-échantillon indépendants.

C'est d'ailleurs pour cette raison que nous avons utilisé uniquement les fondateurs afin de réaliser nos ACP. De même, les SNPs du chromosome X, ont été enlevés afin d'éviter les biais dus à l'homozygotie du X pour les hommes. Ainsi, on obtient une matrice symétrique dont la dimension (lignes et colonnes) correspond aux individus et où chaque case peut être interprétée comme un coefficient d'estimation du lien entre eux. Nous avons réalisé une ACP en considérant l'ensemble des variants ou seulement une partie selon la prise en compte ou non du déséquilibre de liaison afin de s'assurer que s'il y a un effet, celui-ci soit bien retrouvé sur l'ensemble des variables.

2.3.2 Déséquilibre de liaison

Le déséquilibre de liaison (en anglais linkage disequilibrium ou LD) est défini comme l'association non aléatoire d'allèles à des loci différents qui perdure en raison d'une liaison génétique. Plus précisément, lorsque deux loci co-ségrègent d'une génération à l'autre du fait de leur proximité physique (distance inter-locus souvent inférieure à 1 Mb) et que leurs allèles s'associent de manière non aléatoire dans une population panmictique (c'est-à-dire où les couples se forment au hasard indépendamment du phénotype), on parle alors de LD.

Considérons deux loci bi-alléliques ayant respectivement pour allèles Aa et Bb tels que $f(A) > f(a), f(B) > f(b)$, et $f(A) \geq f(B)$ où $f(A)$ est la fréquence de l'allèle A dans la population. En théorie, quatre haplotypes peuvent être observés dans la population : AB, Ab, aB et ab . En l'absence de LD, la présence de l'allèle B au deuxième locus ne dépend pas de l'allèle présent au premier et $f(AB) = f(A)f(B)$ où f représente la fréquence allélique ou haplotypique. En revanche, dans une population panmictique, dès lors que $f(AB) \neq f(A)f(B)$, il y a LD.

On parlera de LD « complet » si l'un des allèles du locus 2 (par exemple B) est exclusivement associé à l'un des allèles du locus 1 (par exemple A). Dans ce cas, l'haplotype aB n'existe pas et seulement trois haplotypes seront observés (AB, ab et Ab). Dans le cas particulier de LD « complet » où $f(A) = f(B)$, l'haplotype Ab n'existe pas non plus et seulement deux haplotypes seront observés (AB et ab : on parle alors d'association complète ou de LD « parfait »).

ANALYSE D'ASSOCIATION PANGÉNOMIQUE FAMILIALE

Il existe différentes mesures du déséquilibre de liaison communément utilisées [417,418] :

- $D = f(AB) - f(A)f(B)$. Cependant, cette mesure du LD dépend des fréquences alléliques et ne permet donc pas de comparer des valeurs entre plusieurs paires de loci. De fait d'autres mesures ont été proposées :
- $D' = \frac{|D|}{D_{max}}$, où D_{max} est la valeur théorique que prendrait D si le LD était complet ($D_{max} = f(A)f(B)$ si $D > 0$ et $D_{max} = f(a)f(b)$ si $D < 0$). Ainsi, D et D' varient de 0 (équilibre de liaison ou absence totale de LD) à 1 (lorsqu'au moins un haplotype n'est pas observé = LD complet). D' permet également de définir des blocs de SNPs (SNPs contigus présentant un D' , supérieur à un seuil prédéterminé, classiquement 0,8).
- Enfin, $r^2 = \frac{D^2}{f(A)f(a)f(B)f(b)}$ et représente le coefficient de corrélation. Il varie également de 0 (équilibre) à 1. Contrairement au D' , il ne vaut 1 que lorsque le LD est parfait (uniquement deux haplotypes sont observés).

En pratique, r^2 reflète mieux la redondance de deux polymorphismes et est préférentiellement utilisé lors du choix des SNPs pour une étude d'association. En effet, lorsqu'on cherche à savoir si une maladie est associée à un polymorphisme d'une région chromosomique (par exemple un gène candidat), il serait trop coûteux de génotyper d'emblée tous les polymorphismes de cette région. On cherche donc à sélectionner un sous-groupe de SNPs qui soit représentatif du plus grand nombre possible des SNPs de la région. Pour cela on définit des bins (ensemble de SNPs, contigus ou non, en fort LD), et l'on sélectionne un tag-SNP représentant ce bin et, défini de telle manière qu'il ait un r^2 supérieur à un certain seuil (généralement 0,8) avec chacun des autres SNPs du bin. L'information véhiculée par ce tag-SNP capture alors celle de tous les SNPs du bin correspondant et il suffit de génotyper les tag-SNPs de la région pour couvrir l'ensemble des variations alléliques et donc la variabilité génétique.

ANALYSE D'ASSOCIATION PANGÉNOMIQUE FAMILIALE

Cependant, la mesure du D' peut aussi avoir un intérêt, notamment pour comprendre l'histoire d'un LD. En effet, lorsqu'une mutation *de novo* dans un chromosome « originel » provoque l'apparition d'un nouveau variant allélique, le variant muté sera préférentiellement transmis aux générations suivantes en étant associé aux variants alléliques voisins présents sur le chromosome originel (en l'absence de recombinaison entre le nouveau variant et les variants voisins, et en l'absence de nouvelle mutation chez les variants voisins). Le LD entre le nouveau variant et les variants voisins sera alors complet, c'est à dire $D' = 1$ tandis que la valeur du r^2 dépendra des fréquences alléliques des variants et pourra être très différent de 1 si les allèles voisins du nouveau variant sur le chromosome originel étaient fréquents. Dans notre travail les calculs de r^2 et D' ont été réalisés grâce aux logiciels PLINK et Haploview [403,404,419].

2.3.3 Analyses d'association pangénomique

➤ Méthodes d'association

Dans les études d'épidémiologie classiques, une association est définie comme l'occurrence chez un même individu de deux caractéristiques plus souvent que ne le voudrait le hasard. En épidémiologie génétique, ces deux caractéristiques sont le phénotype (e.g. une maladie ou une mesure quantitative) et le génotype d'un marqueur spécifique. Dans le cas d'un trait binaire, le principe des études d'association est de comparer la fréquence d'un allèle marqueur donné (ou la distribution des génotypes au marqueur), noté m , entre un échantillon de cas et un échantillon de témoins non atteints [420,421]. Pour un trait quantitatif, cela revient à faire une régression linéaire du phénotype sur les génotypes des individus dans une population.

La mise en évidence d'une association statistiquement significative entre un allèle marqueur m et le phénotype d'intérêt peut avoir plusieurs significations : 1) l'association est due au hasard (erreur de type I), 2) le variant testé est le variant fonctionnel, 3) le variant testé n'a pas d'effet sur le phénotype étudié mais est en déséquilibre de liaison avec le variant causal, 4) l'association est due à un mélange de populations.

ANALYSE D'ASSOCIATION PANGÉNOMIQUE FAMILIALE

Par exemple, une étude d'association cas-contrôles réalisée dans une population composée de deux sous-populations conclura à une association positive si la fréquence de la maladie est plus élevée dans une sous-population et que la fréquence de l'allèle m est également plus élevée dans cette sous-population. Une solution proposée pour éviter le biais dû à un mélange de populations est de conditionner sur les génotypes parentaux au locus marqueur. Ce conditionnement est le fondement des méthodes d'associations en famille comme le test de déséquilibre de transmission (TDT) [422] dont le principe est détaillé dans les paragraphes suivants.

➤ Modèles génétiques : codage du phénotype

L'impact d'un allèle délétère sur le risque de développer une maladie (ou symétriquement d'un allèle protecteur sur la résistance à la maladie) peut suivre différents mécanismes aboutissant à différents modèles génétiques. Un allèle récessif ne modifie le risque de maladie que s'il est présent de manière homozygote. Un allèle dominant modifie le risque de maladie dès sa présence à l'état hétérozygote et cette modification du risque est identique à l'état homozygote.

En génétique mendélienne, on trouve des explications physiopathologiques à ces modes de transmission : si la présence d'un seul allèle sauvage suffit à assurer la fonction de la protéine considérée, la maladie ne survient qu'en l'absence d'allèle sauvage et le mode de transmission est récessif. Si la présence d'un seul allèle sauvage provoque un déficit quantitatif incompatible avec la fonction de la protéine, le mode de transmission est dominant (mécanisme d'haploinsuffisance). C'est également le cas si la présence d'un allèle muté empêche l'action normale de l'allèle sauvage, par exemple par la formation de complexes protéiques non-fonctionnels (mécanisme de dominance négative) [423].

L'allèle à risque peut également augmenter le risque de maladie de façon multiplicative en fonction de sa présence à l'état hétérozygote ou homozygote. C'est l'hypothèse du modèle additif. Ce terme n'est généralement pas utilisé en génétique mendélienne, mais il existe des situations s'y rapportant : les porteurs homozygotes de la mutation du facteur V de Leiden ont un risque thrombotique multiplié par 80, tandis que les hétérozygotes ont un risque multiplié par 7 (la transmission est toutefois classiquement

ANALYSE D'ASSOCIATION PANGÉNOMIQUE FAMILIALE

qualifiée de dominante) [424]. Le surrisque de maladie peut également n'exister qu'en cas d'homozygotie, que ce soit pour l'allèle sauvage ou l'allèle à risque (situation dite « d'avantage aux hétérozygotes »). Ce type de mécanisme est documenté en génétique animale et végétale, et également observé chez l'homme, tout particulièrement dans le domaine des maladies infectieuses où il pourrait être associé à de meilleures capacités de reconnaissance des pathogènes [425–431].

La génétique complexe a repris les modèles hérités de la génétique mendélienne et les a formalisés en fonction des risques relatifs génotypiques (**tableau 13**). Néanmoins, le modèle le plus populaire pour les études d'association pangénomique est nettement le modèle additif. Il est difficile de trouver l'origine de cet état de fait. Le choix du modèle additif n'est pas justifié dans un article standard [432]. Le registre des associations phénotypes-génotypes (« GWAS catalog ») maintenu par le National Human Genome Research Institute (NHGRI) ne retient d'ailleurs que les p-values et odds-ratios du modèle allélique, qui sont asymptotiquement égaux aux p-values et odds-ratios du modèle additif en cas d'équilibre de Hardy-Weinberg [433,434].

Risque Relatif Génotypique			
Mode génétique	Génotype AA	Génotype AB	Génotype BB
Récessif	1	1	$\gamma_{BB} \neq 1$
Dominant	1	$\gamma_{AB} = \gamma_{BB} \neq 1$	$\gamma_{BB} = \gamma_{AB} \neq 1$
Additif/ Allélique	1	$\gamma_{AB} = (\gamma_{BB})^{0,5} \neq 1$	$\gamma_{BB} = (\gamma_{AB})^2 \neq 1$
Avantage aux hétérozygotes	$\gamma_{AA} = \gamma_{BB} \neq 1$	1	$\gamma_{BB} = \gamma_{AA} \neq 1$
Génotypique	1	$\gamma_{AB} \neq 1$	$\gamma_{BB} \neq 1$

Tableau 13 : Risques relatifs génotypiques de chaque modèle génétique

Un risque génotypique > 1 implique que le génotype est à risque et un risque relatif génotypique < 1 implique que le génotype est protecteur.

ANALYSE D'ASSOCIATION PANGÉNOMIQUE FAMILIALE

Les arguments apportés pour ne tester (officiellement) qu'un seul modèle génétique sont les suivants [435] : 1) le modèle additif est asymptotiquement équivalent en cas d'équilibre de Hardy-Weinberg au modèle allélique qui compare directement la fréquence de l'allèle considéré chez les cas et les témoins, ce qui simplifie la description de l'allèle comme « facteur de risque » au sens de l'épidémiologie classique, et ce qui simplifie également la conduite de méta-analyses; 2) le modèle additif serait relativement neutre et pourrait capturer d'autres modèles vrais ; 3) si le variant étudié n'est pas le variant causal mais un marqueur en déséquilibre de liaison avec lui, la puissance de détection d'un effet réellement non-additif du variant causal serait minime au variant étudié dès lors que le déséquilibre de liaison n'est pas complet [435] ; 4) les questions des tests multiples et du choix des marqueurs à génotyper dans la cohorte de réplication sont bien assez difficiles à gérer avec un seul modèle génétique.

La génétique mendélienne montre qu'il n'y a pas de justification biologique à une préséance d'un modèle additif. Des études de simulation et des dérivations analytiques rappellent d'ailleurs que le modèle le plus puissant pour la détection de l'effet d'un variant est le modèle « vrai », que nous ne connaissons pas. L'idée d'une « neutralité » du modèle additif vient du fait qu'il modélise un risque à l'état hétérozygote (comme le modèle dominant et le modèle avantage aux hétérozygotes) et à l'état homozygote (comme le modèle dominant et le modèle récessif). Il impose néanmoins une relation multiplicative entre l'état hétérozygote et homozygote. Sa neutralité est donc partielle ; un variant à effet récessif est particulièrement difficile à capturer par ce biais [436,437].

Il existe d'ailleurs un modèle plus neutre que le modèle additif, le modèle génotypique, qui n'impose aucune relation entre l'état hétérozygote et homozygote. Ce modèle requiert néanmoins l'estimation de deux paramètres de risque, alors que tous les autres modèles aboutissent à l'estimation d'un seul paramètre de risque. Encore une fois, ce modèle est le plus puissant s'il correspond au modèle vrai. C'est le modèle le plus universel, puisqu'en cas de modèle vrai récessif, l'estimation du risque pour les hétérozygotes tendra vers 1 et qu'en cas de modèle vrai dominant, les estimations des risques pour les hétérozygotes et les homozygotes tendront à être égales. En cas de modèle vrai additif ou avantage aux hétérozygotes, le modèle génotypique tendra également vers les modèles vrais. Mais les tests

ANALYSE D'ASSOCIATION PANGÉNOMIQUE FAMILIALE

statistiques dérivés auront un degré de liberté supplémentaire, ce qui aboutira à une perte de puissance dans toutes les situations où le modèle vrai est additif, récessif, dominant ou avantage aux hétérozygotes. Le coût statistique moyen de ce paramètre supplémentaire est à l'origine de la rare adoption de ce modèle génétique. La célèbre étude du Wellcome Trust Case Control Consortium a cependant choisi de tester systématiquement deux modèles, le modèle additif et le modèle génotypique. Les modèles dominants et récessifs ne sont considérés que si le modèle génotypique améliore significativement l'adéquation aux données (model fit) [438].

Néanmoins, il y a un certain pragmatisme à ne traiter que le modèle additif. Pour beaucoup de marqueurs, la fréquence des homozygotes pour l'allèle mineur est faible, ce qui rend en pratique l'estimation d'un OR pour un effet dominant ou un avantage aux hétérozygotes proche de l'OR estimé par le modèle additif. Le seul modèle réellement négligé par cette approche est le modèle récessif. Certains considèrent que la puissance pour détecter un effet récessif dans un GWAS est trop faible pour que l'on s'y attarde. À odds-ratio et fréquence allélique constants, la puissance de détection d'un effet récessif est bien plus faible que celle d'un effet dominant, en raison de la rareté globale des cas et témoins exposés dans le cas récessif (homozygotes). Enfin, on pourrait argumenter que la capacité technologique ou même conceptuelle de validation mécanistique d'un signal de GWAS est bien trop faible pour que le mode réel d'action de l'allèle à risque ait une importance à l'heure actuelle. Qui plus est, tester un seul modèle réduit la dimension du problème de la correction pour tests multiples.

Il reste une constatation simple : nous sommes des êtres diploïdes et la biologie mendélienne regorge d'exemples de mode d'action de toute nature, en particulier récessifs et dominants. La question des tests multiples ne nous paraît pas centrale dans une perspective exploratoire de génération d'hypothèses. Nous avons donc choisi d'explorer l'ensemble des modèles génétiques.

➤ Principe du Test de Déséquilibre de transmission

En anglais, le Transmission Disequilibrium Test (ou TDT) a été développé par Spielman en 1993 pour tester la liaison génétique entre un locus marqueur et le locus causal en présence d'association, c'est à dire pour distinguer le déséquilibre de liaison d'une association factice entre deux loci non liés due à une stratification ou un mélange de population [422]. Une des particularités du TDT est qu'il ne peut détecter la liaison génétique qu'en présence d'association. L'unité familiale de base du TDT est le trio, c'est-à-dire un enfant atteint et ses deux parents. Le principe est de tester si pour un SNP donné, les parents hétérozygotes Aa transmettent les allèles A et a à leur enfant atteint de façon équiprobable. Sous l'hypothèse nulle d'absence de liaison ou d'association entre le marqueur et le locus causal, les parents hétérozygotes transmettent l'un ou l'autre de leurs allèles à leur enfant atteint avec la même probabilité 0,5. En revanche, sous l'hypothèse alternative de liaison et d'association (ou déséquilibre de liaison) entre le marqueur et le locus causal, la probabilité de transmission d'un des allèles est supérieure à l'autre. Le principe du test est d'utiliser les allèles parentaux non transmis aux enfants atteints comme témoins appariés des allèles parentaux transmis. Le test est un simple χ^2 sur données appariées suivant sous H_0 une loi de χ^2 à 1 degré de liberté (χ^2 de Mac Nemar).

Détaillons la formulation du test. Les allèles parentaux transmis et non transmis sont comptabilisés dans un tableau de contingence à quatre cases (**tableau 14**) où

- n_{AA} est le nombre de parents homozygotes AA ayant transmis A ,
- n_{Aa} est le nombre de parents hétérozygotes Aa ayant transmis a ,
- naA est le nombre de parents hétérozygotes Aa ayant transmis A et
- naa est le nombre de parents homozygotes aa ayant transmis a .

		Allèles transmis	
		A	a
Allèles non transmis	A	n_{AA}	n_{Aa}
	a	naA	naa

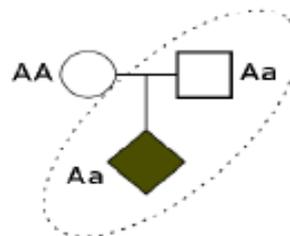


Tableau 14 : Table de contingence du TDT

ANALYSE D'ASSOCIATION PANGÉNOMIQUE FAMILIALE

Soient :

$P(A)$ la probabilité pour un parent hétérozygote Aa de transmettre A à un enfant atteint et,

$P(a)$ la probabilité pour un parent hétérozygote Aa de transmettre a à un enfant atteint, les hypothèses s'écrivent :

$$H_0 = P(A) = P(a), \text{ et}$$

$$H_1 = P(A) \neq P(a).$$

Le test de ces hypothèses est un χ^2 de Mac Nemar : $T = \frac{(naA - nAa)^2}{(naA + nAa)}$ qui suit sous H_0 une loi de χ^2 à 1 degré de liberté. Notons que comme les tests de liaison classiques, seule l'information des parents hétérozygotes est utilisée.

Nous avons représenté les probabilités associées aux cellules du tableau de contingence du TDT, dans le cas d'une maladie récessive, où m désigne la fréquence de l'allèle A dans la population, p désigne la fréquence de l'allèle de susceptibilité D au locus causal, $\delta = freq(AD) - mp$ est le déséquilibre de liaison entre le locus causal et le locus marqueur θ , est le taux de recombinaison entre le locus marqueur et le locus causal. Sous l'hypothèse nulle d'absence de liaison génétique ($\theta = 1/2$), nous constatons que l'on obtient bien $E(nAa) = E(naA)$ quelques soient les valeurs de m, p et δ . De plus, la statistique du TDT dépend de la liaison et de l'association ($\delta > 0$), le test ne peut mettre en évidence la liaison génétique que lorsqu'il y a également association allélique entre les deux loci (**tableau 15**).

		Allèles transmis		
		A	a	Total
Allèles non transmis	A	$m^2 + (m\delta)/p$	$m(1-m) + [(1-\theta-m)\delta/p]$	$m + [(1-\theta)\delta/p]$
	a	$m(1-m) + [(1-\theta-m)\delta/p]$	$(1-m)^2 - [(1-m)\delta/p]$	$1-m - [(1-\theta)\delta/p]$
	Total	$m + (\theta\delta/p)$	$1-m - (\theta\delta/p)$	1

Tableau 15 : Probabilités de cellules du tableau de contingence du TDT

ANALYSE D'ASSOCIATION PANGÉNOMIQUE FAMILIALE

Bien que développé comme un test de liaison, le TDT est également un test valide d'association lorsque l'échantillon est constitué de trios [439]. En effet, sous l'hypothèse d'absence d'association ($\delta = 0$), $E(nAa) = E(naA)$ quelques soient les valeurs de m, p et θ et la variance de $(nAa - naA)$ peut également être estimée par $(nAa + naA)$. Le TDT peut également être appliqué à l'analyse de familles composées de plusieurs enfants atteints en décomposant la famille en trios indépendants. Cependant, dans ce cas, le TDT n'est plus un test valide d'association puisque sous l'hypothèse d'absence d'association les allèles marqueurs transmis à des individus atteints apparentés ne sont plus nécessairement indépendants (e.g. en cas de liaison génétique) et la variance de $(nAa - naA)$ ne peut plus être estimée simplement par $(nAa + naA)$.

➤ Extension du TDT

Tel que nous l'avons formulé dans le paragraphe précédent, le TDT nécessite un échantillon de familles nucléaires composés d'un enfant atteint et de ses deux parents génotypés pour un marqueur bi-allélique. Il peut être appliqué à des familles multiplex pour tester la liaison en présence d'association mais n'est pas un test valide d'association en présence de liaison dans ce cas. De fait, de nombreuses extensions du TDT ont été proposées permettant l'analyse de marqueurs multi-alléliques [440], de tester l'association en présence de liaison dans des familles multiplex [441], ou encore de prendre en compte des données parentales incomplètes. Par exemple, dans le cas d'une maladie à âge de survenue tardif, il est souvent difficile d'obtenir les génotypes parentaux et un autre test a été proposé connu sous le nom de Sib-TDT [442]. Le principe est d'utiliser le génotype des germains non atteints à la place des parents et de comparer la fréquence de l'allèle marqueur chez les cas et leurs germains non atteints.

Pour les traits quantitatifs, des tests d'association en famille basés sur le principe du TDT, ont été introduits [443–445]. Le principe du test proposé par Allison (noté TDT-Q) revient à comparer la moyenne phénotypique des enfants ayant reçu l'allèle A d'un parent hétérozygote à la moyenne phénotypique des enfants ayant reçu l'allèle a d'un parent hétérozygote. Tout comme le TDT original, ces tests ont comme unité de base le trio et supposent que les deux parents soient génotypés.

ANALYSE D'ASSOCIATION PANGÉNOMIQUE FAMILIALE

Basé sur l'approche de Fulker et al. [446], qui partitionne l'effet d'association en une composante interfamiliale (« between family » en anglais, correspondant à la composante des méthodes d'association en population) et une composante intrafamiliale (« within family » en anglais, correspondant à la composante du TDT), Abecasis et al. [447] ont proposé un test d'association en famille applicable à des familles nucléaires de taille variable et incluant des données génotypiques parentales manquantes (noté QTDT). Le modèle est basé sur les méthodes de décomposition de la variance et permet de tester la liaison, l'association et la stratification de population. La statistique de test d'association du QTDT est un rapport de vraisemblance qui teste si l'effet intrafamilial est supérieur à 0, laissant l'effet interfamilial non spécifié. Cette approche fait l'hypothèse de normalité du phénotype.

➤ **FBAT : Family-Based Association Test**

Afin de réaliser nos tests d'association, nous avons, utilisé l'approche FBAT (acronyme pour Family Based Association Tests) qui est une généralisation des tests d'association en famille reposant sur le TDT (précédemment décrit) [448].

- **Avantages de FBAT**

Cette approche permet d'intéressantes extensions comme la possibilité de spécifier différents modèles génétiques et différents types de traits (e.g. binaire, quantitatif, censuré), de réaliser des analyses haplotypiques ou multivariées standards, de prendre en compte les familles de structures variables et à information génotypique parentale incomplète, d'analyser conjointement plusieurs phénotypes, ou encore de tester l'hypothèse nulle d'absence d'association en présence de liaison dans des familles multiplex. Il permet également d'étudier l'empreinte parentale (ou imprinting) c'est-à-dire si un allèle est plutôt transmis par la mère plutôt que le père ou inversement [449–453].

- **Statistique de test du FBAT**

De même que pour le TDT, les hypothèses testées sont :

H0 : Absence de liaison ou d'association entre SNP et locus causal ($U = 0$)

H1 : Liaison et association (ou LD) entre SNP et locus causal ($U \neq 0$)

ANALYSE D'ASSOCIATION PANGÉNOMIQUE FAMILIALE

Avec FBAT, la statistique de test U est une combinaison linéaire des génotypes des enfants et de leur phénotype :

$$U = S - E(S),$$

avec

$$S = \sum_{ij} T_{ij} X_{ij},$$

où

X_{ij} est une fonction du génotype de l'enfant j de la famille i dont le codage dépend du modèle génétique considéré (additif, dominant ou récessif, **tableau 16**). Il s'agit d'une variable aléatoire dont la distribution dépend des lois de Mendel.

T_{ij} est une fonction du phénotype de l'enfant j de la famille i .

$E(S)$ est l'espérance de S .

Génotype	X		
	Additif	Dominant	Récessif
aa	0	0	0
aA	1	1	0
AA	2	1	1

Tableau 16 : Codage des génotypes selon les différents modèles génétiques

A est l'allèle de référence.

- Le codage de T_{ij} dépend du phénotype et il s'agit d'une variable fixée, il n'y a donc pas d'hypothèses sur la distribution du trait :
 - Dans le cas binaire, comme dans notre étude, T_{ij} est codé 1 pour les enfants atteints et 0 pour les enfants non atteints, ainsi seuls les enfants atteints vont contribuer à S et sa formulation se simplifie en $S = \sum_{ij} X_{ij}$. Les enfants sains peuvent permettre de reconstruire des génotypes parentaux manquants.
 - Dans le cas quantitatif, si T_{ij} est codé $(y_{ij} - \mu)$, où μ est la moyenne phénotypique dans la population, la statistique U est équivalente à la statistique de Rabinowitz [444]. Dans la mesure où les phénotypes sont considérés comme fixés, aucune hypothèse n'est faite sur la distribution phénotypique, contrairement au QTDT basé sur les méthodes de décomposition de la variance.

ANALYSE D'ASSOCIATION PANGÉNOMIQUE FAMILIALE

- $E(S)$ est l'espérance de S sous l'hypothèse nulle d'absence de liaison ou d'association conditionnellement aux phénotypes et aux génotypes parentaux :

$$E(S_{ij}) = \sum_{ij} T_{ij} E(X_{ij}) = \sum_{ij} T_{ij} E(X_{ij}),$$

avec

$$E(X_{ij}) = \sum_g P(g) X(g),$$

où

$X(g)$ est le génotype de l'enfant dont le codage dépend du modèle génétique considéré.

$P(g)$ est la probabilité sous H_0 de $X(g)$.

\sum_g est la somme sur tous les génotypes possibles des enfants de la famille i .

Lorsque les génotypes parentaux sont connus, il est possible de déterminer les valeurs de $E(X_{ij})$ selon les combinaisons parentales et le modèle génétique considéré (**tableau 17**).

Génotypes parentaux	E(X)		
	Additif	Dominant	Récessif
AA AA	2	1	1
AA Aa	1,5	1	0,5
AA aa	1	1	0
Aa Aa	1	0,75	0,25
Aa aa	0,5	0,5	0
aa aa	0	0	0

Tableau 17 : Espérance sous H_0 du génotype X d'un enfant, notée $E(X)$, selon les génotypes parentaux et le modèle génétique considéré

A est l'allèle de référence.

$V(S)$ est la variance de S sous l'hypothèse nulle d'absence de liaison ou d'association conditionnellement aux phénotypes et aux génotypes parentaux :

$$V(S_{ij}) = V(T_{ij}X_{ij}) = T_{ij}^2 V(X_{ij}), \text{ avec}$$

$$V(X_{ij}) = E(X_{ij}^2) - E(X_{ij})^2, \text{ et}$$

$$E(X_{ij}^2) = \sum_{\text{Génotypes}} X_{ij}^2 P(\text{Génotype} | \text{Génotypes parentaux})$$

ANALYSE D'ASSOCIATION PANGÉNOMIQUE FAMILIALE

La variance V de U peut être calculée à partir de la distribution attendue des génotypes X_{ij} des enfants sous H_0 , c'est-à-dire en considérant que les X_{ij} sont distribués aléatoirement conditionnellement au génotype des parents et que les phénotypes T_{ij} sont fixés. Dans le cas où les données parentales sont incomplètes, la distribution de la statistique s'obtient en conditionnant également sur les configurations génotypiques observées chez les enfants [453]. Dans le cas général, la variance devient :

$$V = \sum_{ij} T_{ij}^2 V(X_{ij}) + \sum_i \sum_j \sum_{k \neq j} T_{ij} T_{ik} \text{cov}(X_{ij}, X_{ik})$$

Pour un échantillon suffisamment grand (en pratique au moins 10 familles informatives), la statistique de test est :

$$ZFBAT = \frac{U}{\sqrt{V_U}} \text{ et suit une loi normale centrée réduite sous } H_0.$$

$$X^2FBAT = \frac{U^2}{V_U} \text{ et suit une loi du } \chi^2 \text{ à 1 degré de liberté sous } H_0.$$

Jusqu'à présent, nous avons considéré l'hypothèse nulle d'absence de liaison ou d'association. Dans le cas d'une étude d'association de familles multiplex au sein d'un intervalle de confiance d'un pic de liaison, il est judicieux de reconsidérer les hypothèses. En effet, l'hypothèse nulle adéquate devient une absence d'association en présence de liaison et l'hypothèse alternative est une association en présence de liaison.

Plus précisément, l'approche la plus intuitive serait de décomposer ces familles multiplex en plusieurs familles trios. Cette hypothèse serait valide si la transmission des allèles parentaux à chaque enfant est indépendante c'est-à-dire en absence de liaison génétique. En revanche, si la transmission des allèles parentaux à chaque enfant n'est pas indépendante c'est-à-dire en présence de liaison génétique, cette décomposition va induire une inflation d'erreur de type 1 correspondant à une augmentation de FP. En effet, en présence de liaison, au sein d'une même famille multiplex, le même allèle d'un marqueur situé dans la région de liaison ségrégera de manière non aléatoire parmi les enfants même s'il ne s'agit ni de l'allèle causal, ni d'un allèle en déséquilibre de liaison avec l'allèle causal.

ANALYSE D'ASSOCIATION PANGÉNOMIQUE FAMILIALE

La distribution de U peut alors être déterminée en conditionnant sur le nombre d'allèles partagés (IBD) c'est-à-dire provenant d'un même ancêtre. Lorsque les parents sont génotypés, cette distribution peut être construite à l'aide de permutations. Pour chaque parent hétérozygote, la valeur de ses deux allèles est permutée indépendamment tout en fixant l'allèle transmis à chaque enfant. Ainsi, les allèles A et a sont transmis aussi souvent mais la distribution du nombre d'allèles partagés IBD reste fixe [454]. En pratique, afin de pallier ce problème, une variance empirique de U qui corrige la corrélation des transmissions peut être utilisée [451] et est donnée par $Ve = \sum_i [S_i - E(S_i)]^2$. Elle est implémentée dans le logiciel FBAT sous l'option « $-e$ ».

Ce test a été effectué pour chaque SNP indépendamment après le QC avec le logiciel FBAT v2.0.4 [449]. Nous avons testé plusieurs modèles génétiques (additif, dominant et récessif) avec ou sans la variance empirique. Nous avons considéré comme significativement associé avec la lèpre, les SNPs ayant une p -value $< 10^{-5}$ pour au moins l'un des trois modèles génétiques testés.

➤ FBAT Dosage : cas des SNPs imputés

Du fait de leur nature probabiliste, une attention particulière doit être apportée à l'analyse des variants imputés. En effet, la contribution de chaque individu n'est pas un couple {Phénotype ; Génotype} mais un quadruplet {Phénotype ; Probabilité du Génotype AA, P_{AA} ; Probabilité du Génotype AB, P_{AB} ; Probabilité du Génotype BB, P_{BB} }. L'approche la plus simple est de considérer uniquement le génotype le plus probable en assignant un génotype à chaque individu lorsque la probabilité d'un des trois génotypes dépasse un seuil donné, par exemple 90%. On rejoint alors le cas général d'un SNP génotypé, appelé génotype « best guess ». Néanmoins, ne pas prendre en compte l'incertitude génotypique peut générer des faux positifs et des faux négatifs dans les analyses d'association. Une approche répandue dans les études en population (par exemple cas-témoins) est de capturer l'incertitude génotypique par le dosage allélique 59 définit comme deux fois la probabilité d'être homozygote pour l'allèle de référence plus une fois la probabilité d'être hétérozygote [455].

ANALYSE D'ASSOCIATION PANGÉNOMIQUE FAMILIALE

Pour cela, la prise en compte de cette incertitude est essentielle et le logiciel FBAT Dosage (Fbatdos) [456] permet donc d'analyser ce type de données dans les études familiales. Plus précisément, cette approche, généralise l'approche précédente et permet en particulier d'utiliser cette mesure de dosage allélique dans les tests d'association en famille précédemment décrits [449]. Des études de simulations ont montré que la distribution asymptotique de FBATdosage était valide sous les hypothèses nulles « absence d'association et absence de liaison » et « absence d'association en présence de liaison ». De plus, ces simulations ont montré qu'elle apportait un gain de puissance substantiel comparé à la méthode « best guess » [456]. L'application de cet outil aux données de la lèpre a de plus permis d'affiner l'association à deux loci précédemment identifiés dans la population chinoise [375,456].

Nous avons donc utilisé FBATdos pour tester l'association des 5 607 170 SNPs imputés et ayant passés les filtres de QC. Les 3 modèles génétiques (add, dom, réc) ont été étudiés. Pour les SNPs situés sur le chromosome X, les probabilités des génotypes imputées ont été convertis en « best-guess » en utilisant un seuil de probabilité d'imputation de 0,9. Suite à cela, les SNPs ayant un CR inférieur à 99% et un écart à Hardy Weinberg avec $p < 10^{-3}$ ont été éliminés de l'étude. Ainsi, 9 907 SNPs sur le chromosome X ont été analysés avec FBAT v2.0.4 [449] sous le modèle additif.

➤ **P-values minimales**

Dans la section précédente, nous avons vu qu'il était possible de tester plusieurs modèles génétiques et ces différentes approches d'analyse ne donnent pas la même liste des meilleurs signaux. Or il est difficile de donner une préférence à l'une ou l'autre sur des arguments objectifs. Une possibilité pragmatique dès lors est d'assigner à chaque SNP sa p-value minimale pour les 3 modèles testés afin d'identifier les signaux les plus prometteurs qui seront ensuite emmenés en réplification.

➤ Lecture des résultats

A la suite des tests statistiques, nous avons récupéré une liste de p-values par la méthode FBATdos. Ainsi, la première étape a été de déterminer les SNPs potentiellement associés à la lèpre : il s'agit des marqueurs dits significatifs, puis, étudier leur déséquilibre de liaison, enfin leur gène associé.

Tout d'abord, afin de visualiser graphiquement les résultats et de déterminer les variants significatifs, nous avons utilisé deux représentations implémentées dans le package « qqman » sur R [457,458] :

- **Manhattan-plot**

Le Manhattan-plot représente la p-value du résultat d'association (plus précisément, le logarithme négatif de la p-value) de chaque SNP le long des chromosomes. Plus précisément, il permet de représenter sur l'axe des abscisses (X) les marqueurs par chromosome selon leur position génomique et sur l'axe des ordonnées (Y) le logarithme négatif de la p-value d'association avec le trait étudié. Cette figure illustre la masse de données générées par les analyses d'association (ici presque 6 millions de tests) et permet de faire ressortir visuellement les SNPs les plus significatifs (avec les p-values les plus faibles), qui apparaîtront en haut sur le graphique. L'observation d'un « gratte-ciel » (ou ligne verticale de points) évoquera une association significative de SNPs proches sur le génome.

- **Quantile-quantile-plot**

Le QQ-plot permet d'évaluer la pertinence de l'ajustement d'une distribution donnée à un modèle théorique (c'est-à-dire évaluer l'impact des structures génétiques identifiées sur l'erreur de type I). Ces graphes représentent pour chaque SNP la valeur de la statistique observée dans l'échantillon en fonction de la valeur attendue de la statistique sous l'hypothèse nulle : ainsi, sur l'axe des X est représenté le $-\log_{10}(P)$ attendu et sur l'axe des Y l'observée. Par exemple, après classement par ordre croissant de la statistique de test sur l'ensemble des SNPs testés, le 180000^{ème} SNP sur 1800000 SNPs (10^{ème} percentile) a une valeur attendue de χ^2 à 1 degré de liberté de 0,016. Une inflation de l'erreur de type I est probable si le ratio de la valeur observée sur la valeur attendue est supérieur à 1. Autrement dit, pour une statistique contrôlant l'erreur de type I, le diagramme qq est la droite $y = x$ et

représente la distribution des p-values sous l'hypothèse nulle (absence d'association). De manière générale, on s'attend à une déviation de la distribution attendue uniquement pour les valeurs de P extrêmes, ce qui reflète vraisemblablement de réelles associations.

• **Facteur d'Inflation Génomique**

Une façon fréquente de résumer un diagramme quantile-quantile est de donner le ratio de la statistique de test observée sur la statistique attendue sous H_0 à la médiane de la distribution (valeur attendue 0,45 pour un χ^2 à 1 degré de liberté). Dans le cadre des analyses d'association pangénomique, ce ratio est nommé facteur d'inflation génomique (Genomic Inflation Factor, GIF) et est noté λ . Il permet de vérifier qu'il n'y a pas d'inflation systématique des statistiques qui signifierait que le QC n'a pas été réalisé de façon optimale. Dans le cas où λ est compris dans l'intervalle $[0,95 - 1,05]$ donc proche de 1, on peut considérer qu'il n'y a pas de déviation systématique entre la distribution observée et celle attendue (causée, par exemple, par un phénomène de stratification de population dans les études de type cas-témoins ; mais dans notre étude, l'analyse de type FBAT présente justement l'avantage d'être insensible à la stratification de population) et l'interprétation des p-values reste inchangée. Il est à noter que l'ensemble de cette procédure est valable sous l'hypothèse que la majorité des SNPs n'est pas associée à la pathologie d'intérêt et que les statistiques obtenues sont bien distribuées selon l'hypothèse nulle.

Pour ces deux représentations, toutes les p-values sont transformées en $-\log_{10}(P)$, cela a deux principaux avantages : tout d'abord, le $-\log$ permet d'inverser le signe et donc d'avoir les plus petites p-values « en haut » du graphe sur les Y . De plus, le log en base 10 facilite le lien avec la p-value car $-\log_{10}(10^{-3}) = 3$ correspond à une p-value de 10^{-3} .

• **Seuil de significativité et tests multiples**

Dans les études d'association, du fait que des centaines de milliers de tests (voire des millions comme dans notre étude) sont réalisés (un par SNP), la barre de significativité de 0,05 n'est plus valide (car dans le cas où l'on réalise des tests sur 500 000 SNPs satisfaisant l'hypothèse nulle de non-association, cela supposerait qu'environ 25 000 SNPs seraient déclarés significativement associés à la pathologie alors qu'ils ne le sont pas réellement). Une correction de Bonferroni exigerait une valeur nominale de significativité de $\frac{0,05}{500\,000} \sim 10^{-8}$, mais

ANALYSE D'ASSOCIATION PANGÉNOMIQUE FAMILIALE

celle-ci est trop conservatrice. De plus, compte-tenu du déséquilibre de liaison entre les SNPs, les tests ne sont pas tous indépendants. Il existe alors un consensus tacite quant à ce seuil de significativité à l'échelle du génome qui est fixé à $5 \cdot 10^{-8}$ (soit 1 million de tests indépendants au seuil de 5%) et ce quel que soit le nombre de marqueurs étudiés [459]. Ce seuil est figuré par une ligne horizontale rouge sur nos Manhattan plots. Tout SNP ayant une p-value inférieure est considéré comme significatif au sein d'une étude GWAS. Cependant en pratique, les SNPs ayant une p-value d'association de l'ordre de 10^{-5} sont généralement conservés et testés au sein d'un autre jeu de données indépendant afin de valider ou non l'association, il s'agit de l'étude de réplication.

Ainsi, nous avons considéré comme significativement associés au phénotype lèpre, tous les marqueurs ayant une p-value inférieure au seuil 10^{-5} , une ligne bleue représente ce seuil nominal évocateur d'association à l'échelle du génome. Celui-ci peut paraître peu stringent car il ne prend pas réellement en compte les tests multiples liés au nombre de SNPs et aux différents modèles testés, ce qui induit incontestablement une inflation d'erreurs de type 1. Néanmoins, dans ce premier temps d'analyse, nous considérons être plutôt dans un contexte exploratoire visant à générer des hypothèses, et il nous a semblé plus important de minimiser l'erreur de type II, c'est-à-dire le manque de puissance. Sans bien entendu perdre de vue cet écueil des tests multiples dans nos interprétations.

- **Interprétation d'une association entre un SNP et un phénotype**

Il est important de rappeler les limites inhérentes à ces approches et qui peuvent conduire à des faux positifs comme à des faux négatifs. Un problème méthodologique majeur est la multiplicité des tests. En particulier, tester plusieurs gènes candidats (et souvent plusieurs polymorphismes par gène) sur la même maladie et dans la même population nécessite de corriger le degré de signification en fonction du nombre de tests réalisés (ou au moins de discuter le problème) même si les résultats sont publiés dans des papiers successifs. De plus, il faut réaliser que les résultats négatifs (au sens de non significatif statistiquement) sont rarement rapportés rendant plus difficile encore l'évaluation objective d'une association dite « significative ».

ANALYSE D'ASSOCIATION PANGÉNOMIQUE FAMILIALE

Un deuxième problème concerne le déséquilibre de liaison. En présence d'une association réelle entre un allèle et une pathologie, il est possible que l'allèle associé ne soit pas l'allèle fonctionnel mais en déséquilibre de liaison avec lui. Dans ce cas, la difficulté à répliquer l'association avec cet allèle donné dans une autre population peut s'expliquer par une hétérogénéité dans le déséquilibre de liaison entre les populations alors que cette explication ne tient pas si l'on teste le variant fonctionnel proprement dit. À l'inverse, si aucune association n'est retrouvée, une explication peut être que l'allèle testé n'est pas en fort déséquilibre de liaison avec l'allèle fonctionnel même s'il est localisé dans le même gène. Ainsi, l'absence d'association avec un allèle n'écarte pas un effet du gène.

Pour résumé, la mise en évidence d'une association statistiquement significative entre un allèle marqueur m et le phénotype d'intérêt peut avoir plusieurs significations :

- Le variant testé est le variant fonctionnel,
- Le variant testé n'a pas d'effet sur le phénotype étudié mais est en déséquilibre de liaison avec le variant causal (**figure 35**),
- L'association est fallacieuse est due :
 - Au hasard (erreur de type I),
 - À un mélange de populations (étude d'association cas-contrôles)
 - À la problématique des tests multiples
 - À d'autres conditions expérimentales et/ou d'analyses statistiques.

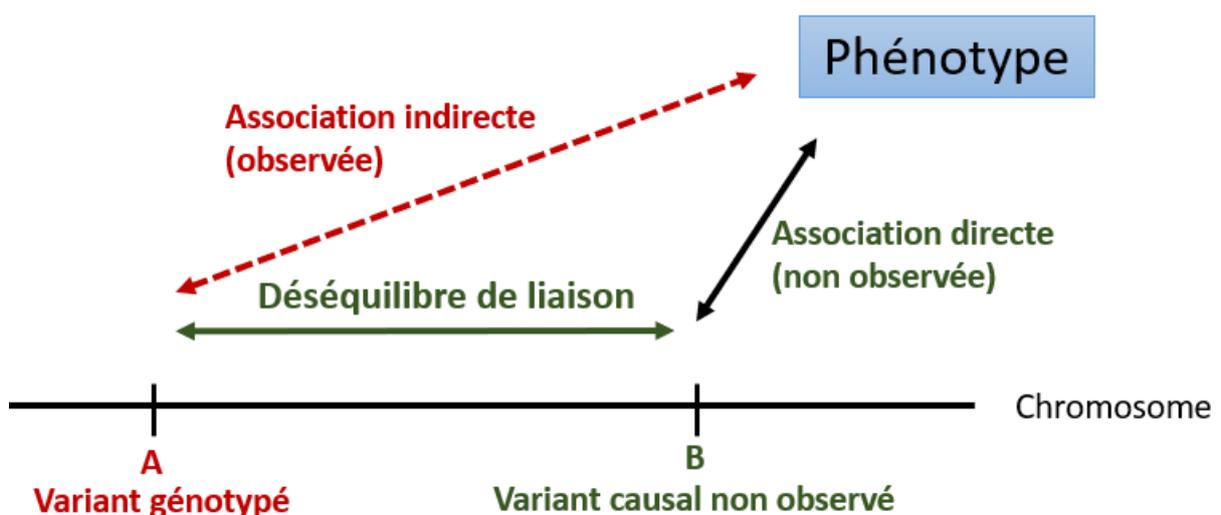


Figure 35 : Association indirecte, le variant génotypé (A) est en LD avec le variant causal (B)

2.3.4 Réplication des signaux de la cohorte primaire

Au cours de ce travail, nous avons réalisé des millions de tests statistiques. Les analyses d'association pangénomique en général, et la nôtre en particulier, interrogent la problématique des tests multiples (comme nous l'avons vu dans le précédent paragraphe) avec une intensité inédite dans l'histoire de la théorie statistique. Les corrections faisant l'hypothèse de tests indépendants (Sidak, Bonferroni) semblent trop sévères dans une perspective de génération d'hypothèses, et aboutissent à l'élimination excessive de faux négatifs. La manière la plus raisonnable de se départir des faux positifs tout en contrôlant en partie l'erreur de type II est le passage des meilleurs signaux au révélateur de l'étude de réplication [460]. La réplication consiste à tester les meilleurs signaux dans un nouvel échantillon indépendant et de ne retenir que les signaux significatifs à deux reprises.

➤ Cohorte de réplication

La cohorte de réplication est un échantillon indépendant de cas et de contrôles recrutés au même endroit que la cohorte primaire. Les mêmes critères de diagnostics de lèpre et de recrutements que dans l'étude primaire ont été utilisés. Les contrôles n'ont pas d'antécédents personnels ou familiaux (parmi les parents au premier, deuxième ou troisième degré) de lèpre ou de tuberculose. Elle est composée de 1 863 individus, dont 1 849 passent le contrôle qualité par individu. Les 14 individus ont été éliminés pour un taux de génotypage (CR) inférieur à 95%. Ces 1 849 individus sont répartis en 1 181 cas et 668 contrôles. La proportion d'hommes parmi les cas et témoins est respectivement de 66 et 43%. L'âge moyen des cas est de 23 ans et celui des contrôles est de 37 ans (**tableau 18**).

ANALYSE D'ASSOCIATION PANGÉNOMIQUE FAMILIALE

<i>Caractéristiques</i>	<i>Cas</i>	<i>Contrôles</i>
<i>N individus</i>	1 181	668
<i>N (atteint/non atteint)</i>	1 181/0	0/668
<i>N (MB/PB/NA)</i>	584/536/61	–
<i>Âge en années : moyenne (SD)</i>		
<i>Atteint</i>	23,0 (8,0)	–
<i>Non atteint</i>	–	37,9 (12,1)
<i>% d'hommes</i>		
<i>Atteint</i>	66,4	–
<i>Non atteint</i>	–	44,3

Tableau 18 : Description de l'échantillon de réplication

Âge au moment du diagnostic pour les cas de lèpre et au moment du recrutement pour les contrôles.

➤ Sélection des SNPs pour l'étude de réplication

Nous avons utilisé trois stratégies différentes pour sélectionner les SNPs pour l'étude de réplication :

- Tout d'abord, nous avons sélectionné les signaux évocateurs à l'échelle du génome ($p < 10^{-5}$) et situés en dehors de la région HLA (hors-HLA). Pour identifier les loci indépendants, nous avons utilisé un LD basé sur la procédure de « Clumping » implémentée dans PLINK 1.9 [403,404]. Cette procédure prend en compte tous les SNPs ayant une p-val $< 10^{-5}$ et forme des clusters de ces SNPs « index », avec tous les autres SNPs en LD ($r^2 > 0,5$) et à proximité physique (à moins de 5 Mb des SNPs « index »). Seuls les SNPs ayant une p-val $< 10^{-4}$ sont gardés. Comme « input », nous avons utilisé les p-val d'association de l'étude primaire et les structures de LD des KHV de la population de 1000 génomes phase 3. Un SNP de chaque cluster (Tag-SNP) a été sélectionné et nous avons génotypé 8 Tag-SNPs.

- Nous avons, ensuite, appliqué une deuxième stratégie, toujours pour les SNPs hors-HLA mais ayant, cette fois ci, une p-value en dessous du seuil de significativité à l'échelle du génome ($p < 5 \cdot 10^{-8}$) précédemment décrite dans d'autres études. Pour ces SNPs, nous avons utilisé un seuil de p de 10^{-2} dans notre étude et les SNPs ayant le même allèle à risque (ou protecteur) dans les deux études. Nous avons donc sélectionné 5 SNPs.

ANALYSE D'ASSOCIATION PANGÉNOMIQUE FAMILIALE

- Enfin, pour les SNPs situés dans la région HLA, nous avons utilisé une troisième stratégie, car un grand nombre de SNPs présentaient des signes d'association, notamment au niveau du seuil de significativité ($p < 5.10^{-8}$) avec des structures complexes de LD. Nous avons sélectionné un ensemble de 88 SNPs situés entre 30 Mb et 33 Mb sur le chromosome 6, sur la base des résultats de l'association, la structure des LD et les considérations techniques pour le génotypage, et nous avons génotypé ces SNPs dans l'échantillon de réplication. Nous avons, ensuite, utilisé ces SNPs pour imputer, dans l'échantillon de réplication tous les variants pour lesquelles il y avait des signaux évocateurs d'association (358 SNPs).

Les génotypes ont été déterminés par séquençage ciblé FLUIDIGM ou avec la technologie SEQUENOM. Pour les données FLUIDIGM, nous avons utilisé Genome Analysis Toolkit (GATK v3.3) [461] pour traiter les fichiers BAM. Les fichiers, par individu, genomic variant call files (gVCF) ont été générés avec GATK et HaplotypeCaller, et le génotypage a été effectué avec GATK et GenotypeGVCFs.

Les SNPs ayant un CR < 95% ou un écart significatif à Hardy-Weinberg ($p < 0,001$) ont été retirés de l'analyse. Ainsi, 10 SNPs sur les 13 choisis hors-HLA, ont été analysés. L'imputation dans la région du CMH a été effectuée avec SHAPEIT v2 [410], IMPUTE2 [412] et le panel de référence 1000 génomes Phase 3.

➤ Méthode d'analyse : la régression logistique

Le design cas-témoin de notre étude de réplication oriente naturellement la méthode d'analyse vers la régression logistique, c'est-à-dire un cas particulier de modèle linéaire généralisé dans lequel la probabilité de développer la maladie est modélisée par une combinaison linéaire de facteurs explicatifs que sont le génotype (codé selon différents modèles génétiques) et des variables d'ajustement. Plus précisément, le logarithme de la cote de la probabilité de développer la maladie est modélisé (fonction de lien logit). Ce modèle linéaire est adapté à l'estimation de probabilités dont les valeurs sont par essence contraintes à l'intervalle [0,1]. Le choix de modéliser la cote rend par ailleurs aisée l'estimation des odds-ratios (rapport de cote) quantifiant le risque conféré par tel ou tel génotype comme l'exponentielle du coefficient de régression linéaire. La formulation d'une fonction de vraisemblance des données en fonction de ce coefficient permet son estimation au maximum de vraisemblance. Le caractère significativement différent de 0 de ce coefficient est évalué par un test reposant sur la statistique du rapport de vraisemblance. La prise en compte d'une structure génétique de l'échantillon de type stratification de population, ou effet batch, peut être réalisée par l'introduction des composantes principales (issues de la décomposition de la matrice de variance-covariance des génotypes entre individus) comme variables d'ajustement dans le modèle de régression logistique. Nous avons réalisé l'analyse sans et avec ajustement sur les premières composantes principales.

L'ensemble de ces procédures a été réalisé avec le logiciel R par la fonction glm [402] et le logiciel SNPTEST en testant les 3 modèles génétiques (additif, dominant et récessif) [412,438]. Un test unilatéral a été effectué avec l'hypothèse alternative que les allèles de susceptibilité trouvés dans l'échantillon familial primaire étaient également des facteurs de susceptibilité dans la réplication. Un seuil de p de 0,01 a été utilisé dans ce test unilatéral pour la détection des signaux de réplication.

2.3.5 Analyse combinée : les « pseudo-contrôles »

➤ **Définition des « pseudo-germains »**

Les données familiales en trios peuvent également s'analyser par le biais d'une régression logistique conditionnelle, comparant les cas (enfants atteints) et des pseudo-témoins appariés qui sont des frères ou sœurs fictifs des cas (ou « pseudo-germains ») dont les génotypes aux différents marqueurs étudiés sont créés à partir des génotypes des cas et de leurs parents, comme décrit ci-après. Connaissant les génotypes parentaux, il existe au maximum quatre combinaisons alléliques possibles chez les enfants d'une même fratrie, ayant chacune une probabilité de 25% sous l'hypothèse nulle.

Pour chaque cas réel (ayant un génotype observé à chacun des marqueurs d'intérêt), on crée trois pseudo-germains (non atteints) ayant chacun pour génotype aux marqueurs d'intérêt un des trois génotypes non transmis au cas qui auraient pu être observés conditionnellement aux génotypes parentaux (**figure 36**). Par exemple, si les deux parents sont hétérozygotes Aa et que l'enfant atteint a le génotype aa , les pseudo-germains auront pour génotypes AA , Aa et aA . Si l'un des parents est homozygote, seul un pseudo-germain sera créé, puisqu'il n'y a dans ce cas que deux génotypes transmis possibles (les génotypes des deux autres pseudo-contrôles seront codés comme des valeurs manquantes). Par exemple, si les parents ont pour génotypes AA et Aa et l'enfant atteint AA , le pseudo-contrôle aura le génotype Aa . Les génotypes des cas et des pseudo-germains sont ensuite recodés selon les modes de transmission que l'on souhaite tester selon la méthode décrite précédemment (voir FBAT).

ANALYSE D'ASSOCIATION PANGÉNOMIQUE FAMILIALE

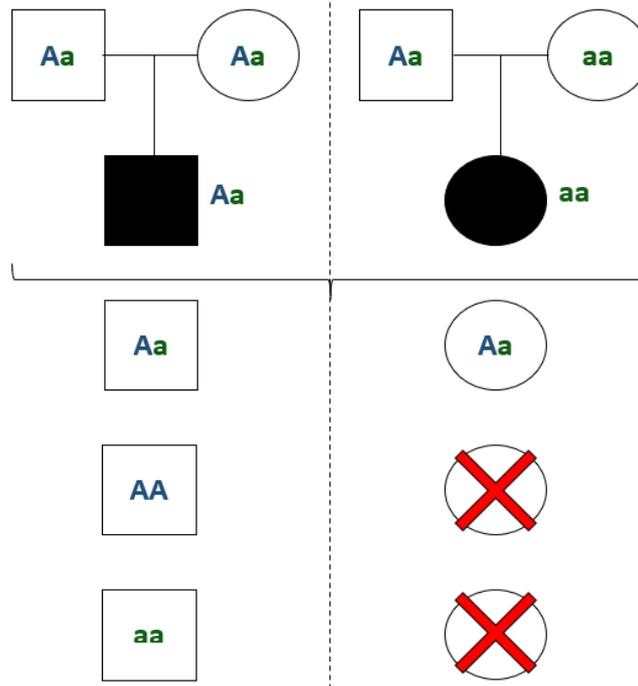


Figure 36 : Création de pseudo-germains pour l'analyse par régression logistique conditionnelle d'un marqueur bi-allélique (Aa)

Cas d'un enfant avec deux parents hétérozygotes (à gauche) et avec un parent homozygote et un parent hétérozygote (à droite).

➤ Modèle statistique

L'analyse consiste alors à comparer les génotypes des cas et des pseudo-témoins par le biais d'une régression logistique conditionnelle. Cette méthode est classiquement utilisée en épidémiologie dans les études cas-témoins avec appariement. Elle dérive de la méthode de régression logistique classique, sa particularité étant de prendre en compte le fait que les individus de chaque groupe d'appariement (ou strate) ont un risque de base commun qui est potentiellement différent de celui des autres groupes (par exemple lié à des variables environnementales communes non identifiées, ou dans notre cas, à un patrimoine génétique parental particulier).

Sans chercher à estimer ce risque, elle analyse l'effet des covariables étudiées de façon stratifiée (séparément dans chaque groupe) et réalise la synthèse des effets observés strate par strate. Dans notre cas, chaque strate est constituée par une pseudo-famille, c'est-à-dire un cas et ses trois pseudo-germains. Dans le cas d'une famille multiplex avec n enfants atteints, la strate familiale contenait n cas et $3n$ pseudo-témoins.

ANALYSE D'ASSOCIATION PANGÉNOMIQUE FAMILIALE

Le principe de cette méthode est en réalité le même que celui du TDT puisque les allèles non transmis (attribués ici aux pseudo-témoins) servent de témoins aux allèles transmis (observés chez les cas). Schaid et Rowland [261] ont montré que lorsqu'on considérait un modèle génétique additif, la statistique du Score de la régression logistique conditionnelle était strictement égale au χ^2 du TDT.

Cette méthode offre une grande souplesse d'analyse et a notamment plusieurs avantages qui explique pourquoi nous l'avons utilisée : elle permet une estimation directe des odds ratios (OR) correspondant aux génotypes, de réaliser des analyses multivariées avec estimation conjointe de l'effet de plusieurs marqueurs génétiques, et de tester l'existence d'une interaction entre marqueurs ou entre un marqueur génétique et des covariables non génétiques. Par convention, pour chacun des marqueurs nous avons choisi comme allèle de référence l'allèle A le plus fréquent. L'OR correspondant à l'effet d'un génotype, qui estime l'excès de risque d'avoir une lèpre associée à ce génotype, est égal à l'exponentielle du paramètre de régression logistique β correspondant à ce génotype. Dans le cas d'un modèle additif, le génotype de référence est AA , et l'OR de l'allèle mineur a est modélisé comme étant multiplicatif ($\beta_{aa} = 2\beta_{aA}$ donc $OR_{aa} = e^{2\beta_{aA}} = OR^2_{aA}$). Dans le cas du modèle dominant, l'OR compare l'effet des génotypes aa et aA à celui du génotype AA et dans le cas du modèle récessif l'OR compare l'effet du génotype aa à celui des génotypes AA et Aa .

D'autre part, et il s'agit de la raison principale de cette analyse, en ajoutant les individus cas-contrôles comme une strate supplémentaire et indépendante, par la même méthode, nous avons pu obtenir une p-value combinée de l'étude primaire et de réplification. En effet, le fait d'avoir des données familiales pour l'étude primaire puis des données cas-contrôles pour la réplification, complique l'obtention d'une p-value combinée qui se calcule habituellement en mergeant les données du primaire et de la réplification.

➤ Stratification sur les covariables non génétiques

Les pseudo-témoins se voyant attribuer les mêmes valeurs que leur cas pour les covariables phénotypiques (par exemple le sexe, l'âge ou la forme de lèpre), il n'est pas possible d'étudier l'effet conjoint des génotypes aux marqueurs et de ces covariables en incluant simplement ces covariables dans le modèle de régression logistique conditionnelle. En revanche, il est possible d'inclure dans le modèle un terme d'interaction entre un génotype et une covariable pour tester un effet différentiel du marqueur selon la valeur de cette covariable.

Pour la création automatique des pseudo-germains, nous avons utilisé un programme écrit en langage C et qui avait été fait en interne au laboratoire. Dans le cas particulier où les deux parents sont homozygotes à un marqueur donné, les génotypes du cas et de ses pseudo-germains à ce marqueur sont recodés selon une valeur particulière (« -1 ») qui ne correspond à aucun génotype réel. Cette pseudo-famille ne contribuera pas à l'estimation de l'effet de ce marqueur puisque tous les génotypes auront la même valeur, mais elle pourra tout de même être prise en compte en analyse multivariée pour l'estimation conjointe de l'effet des autres marqueurs, alors qu'elle aurait été retirée de l'analyse multivariée si tous ses génotypes étaient codés en valeur manquante à ce marqueur. Or la probabilité P pour une pseudo-famille d'avoir deux parents homozygotes à au moins un marqueur augmente très rapidement lorsque le nombre de marqueurs étudiés augmente (d'autant plus que les fréquences alléliques des marqueurs étudiés s'éloignent de 0,5). Par exemple, pour 5 marqueurs de fréquences alléliques 0,2 et 0,8, $P = 0,96$. Cette procédure permet donc d'éviter une réduction drastique du nombre de pseudo-familles prises en compte en analyse multivariée.

La régression logistique conditionnelle a été réalisée avec le logiciel SASv.9.3 [462] et a permis de déterminer une p-val combinée de l'étude primaire familiale avec l'étude de réplification cas-contrôles..

2.3.6 Analyse multivariée

Jusqu'à présent, toutes les analyses décrites étaient des analyses univariées (l'association de chaque SNP avec la lèpre a été testée indépendamment des autres SNPs). En revanche, l'analyse multivariée vise à évaluer l'effet conjoint de plusieurs marqueurs à la fois. Deux types d'approches sont possibles : les analyses multivariées (MV) selon que l'on considère la phase ou non (l'analyse haplotypique et l'analyse par régression logistique conditionnelle). Pour les SNPs ne présentant aucun LD ($D'=0$), il n'y a pas d'association allélique préférentielle, l'analyse haplotypique n'a pas d'intérêt. Pour les SNPs présentant un LD parfait ($r^2=1$), ni l'analyse haplotypique ni la régression multivariée n'ont d'intérêt car la redondance d'information est totale. Les analyses multivariées ont donc été effectuées en conséquence.

➤ **Analyses haplotypiques**

Le test haplotypique consiste à évaluer une probabilité d'association entre un haplotype et un trait phénotypique d'intérêt. Elle étudie l'action conjointe d'allèles en phase, c'est-à-dire situés sur le même brin d'ADN. Elles peuvent permettre de capturer un effet biologique en capturant une interaction en cis entre 2 ou plusieurs variants [463]. Elles peuvent également permettre de capturer des variants rares [463,464]. Dans la mesure où nous disposons d'un échantillon familial, l'inférence des haplotypes est facilitée et les analyses peuvent être réalisées avec le logiciel FBAT grâce à la fonction hbat (haplotype-based association test). En effet, dans certains cas, il est possible de déterminer l'haplotype sans ambiguïté à partir des génotypes à chaque locus. Cependant, à partir du moment où un individu est hétérozygote pour au moins deux des marqueurs considérés, on ne peut déterminer la phase allélique que sous certaines configurations génotypiques parentales. Les individus pour lesquels la phase allélique est incertaine peuvent néanmoins être inclus dans l'analyse après attribution d'un « poids » à chaque phase possible, le poids étant déterminé par la probabilité d'observer chaque haplotype conditionnellement aux génotypes (non phasé) de l'individu et de ses parents. Cette probabilité peut être estimée par la méthode dite EM (pour Expectation Maximization) [465]. La méthode EM est directement implémentée dans la procédure « hbat » du logiciel FBAT.

➤ Analyse par régression logistique conditionnelle

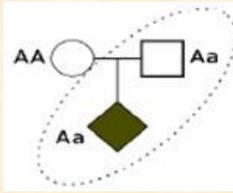
Les analyses multivariées non phasées permettent d'étudier des interactions d'ordre plus général et donc de l'action conjointe de plusieurs variants indépendamment de la phase (considération simultanée des deux allèles présents à chaque marqueur). Cette approche consiste à prendre plusieurs marqueurs en considération et à utiliser une méthode de sélection multivariée :

- **Sélection forward** : elle démarre du modèle nul ($\text{logit}(p) = \alpha$) et les principaux marqueurs sont ajoutés au modèle en fonction de leur pertinence c'est-à-dire l'information qu'ils apportent pour expliquer la maladie. Ils sont ajoutés du plus pertinent au moins pertinent avec un critère d'arrêt du type : plus aucun marqueur ne contribue significativement à améliorer l'explication du phénotype. Pour cela, la p-val de chaque SNP est mesurée et à chaque étape le meilleur SNP, si sa p-valeur est inférieure à 0,05, est intégré dans le modèle. On obtient ainsi un test à plusieurs variables, dont le nombre augmente à chaque étape. À la fin, on obtient un ensemble de SNPs qui contribuent au phénotype de manière coopérative. Un tel test permet de mettre en évidence des interactions éventuelles entre les SNPs associés. Néanmoins, cette approche est coûteuse en temps lorsque le nombre de variables (ici les SNPs) est très grand.
- **Sélection backward** : à l'inverse elle démarre du modèle complet incluant tous les marqueurs et soustraira un à un les marqueurs les moins pertinents avec un critère d'arrêt du type : le retrait d'un prédicteur supplémentaire diminue significativement l'explication du phénotype.
- **Sélection stepwise** : c'est celle que nous avons principalement utilisée, est une combinaison des deux précédentes. Elle démarre avec l'algorithme forward, puis, chaque fois qu'une variable est sélectionnée, l'algorithme backward, pour tester la significativité des variables précédemment sélectionnées, va être utilisé. La procédure s'arrête lorsqu'aucune variable ne peut être introduite et aucune variable ne peut être éliminée.

ANALYSE D'ASSOCIATION PANGÉNOMIQUE FAMILIALE

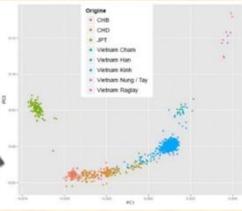
Dans notre cas, il faut noter que les analyses multivariées n'ont pas été réalisées de façon systématique mais uniquement sur les SNPs emmenés à la réplication et donc considérés comme « significatifs » en univarié, ceci en raison d'un problème de temps computationnel. Plus précisément, nous avons effectué, une analyse multivariée des SNPs HLA dans l'échantillon de réplication de cas-contrôles a été réalisée afin de disséquer de manière fine ce signal et de détecter les signaux indépendants. Toutes ces analyses ont été effectuées avec le logiciel SAS 9.3 [462] en utilisant la procédure « LOGISTIC » et la sélection « stepwise ».

Afin de conclure sur cette partie méthodologie (**carte mentale 3**), nous pouvons dire que l'étape de contrôle qualité des données (marqueurs et individus) a été essentielle afin de pouvoir réaliser les tests d'association permettant d'obtenir une liste de variants potentiellement associés au phénotype maladie. Une étape d'imputation des génotypes manquants a également été décisive afin de couvrir le maximum de positions sur l'ensemble du génome. Les analyses d'association univariées (marqueur par marqueur) ont été effectuées par FBATdos [456] (spécifiques aux études d'association familiales). Néanmoins, il a été important de répliquer les résultats au sein d'un échantillon indépendant de la même population afin de confirmer les signaux évocateurs d'associations en particulier en raison du seuil peu stringent de significativité que nous avons choisi. Enfin, les analyses de LD et multivariées permettent de détecter les signaux indépendants. La détermination de SNPs associés donc potentiellement de gènes associés est une première étape dans la compréhension de la génétique humaine de la lèpre et ouvre la possibilité vers d'autres études.



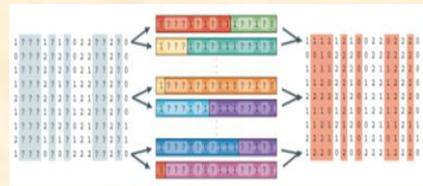
Recueil des données familiales primaires (Vietnam) & Génométypage

2 785 Individus
592 633 SNPs



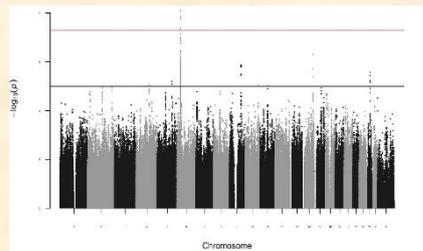
Contrôle qualité des données sur les SNPs et les individus

1 749 Individus
422 546 SNPs
(9 907 sur le X)



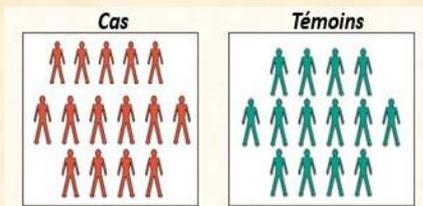
Imputation des génotypes manquants
⇒ Cohorte primaire finale

1 749 Individus
5 607 170 SNPs
(9 907 sur le X)



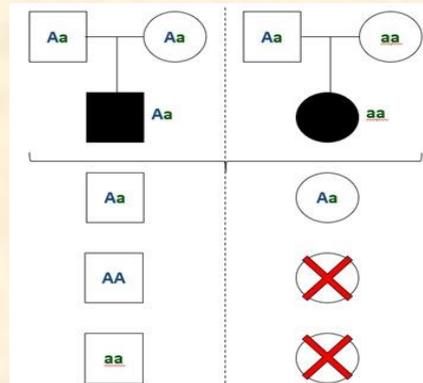
Analyse univariée
→ Association pangénomique familiale

Signaux significatifs et indépendants (LD) testés dans une cohorte indépendante



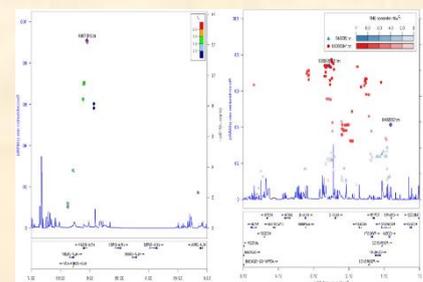
Réplication cas-témoin + QC + imputation HLA
⇒ Cohorte de réplication

1 181 cas et
668 contrôles
98 + 358 SNPs



Analyse combinée (primaire + réplication)
⇒ Pseudo-contrôles

2 162 pseudo-contrôles + 1 849 de la réplication
101 + 358 SNPs



Analyse multivariée HLA sur cohorte de réplication
→ Signaux indépendants dans la région HLA

Signaux indépendants dans région HLA (LD)

1 181 cas et
668 contrôles
18 SNPs HLA

Carte mentale 3 : Méthodes du GWAS

2.4 Résultat de l'analyse d'association pangénomique

Nous allons, ici, présenter les principaux résultats de l'analyse d'association pangénomique familiale à la lèpre per se. Malgré l'étude de tous les modèles génétiques possibles, tous les résultats seront, ici, rapportés sous le modèle additif car suffisant pour expliquer l'ensemble de nos données.

2.4.1 Résultat de l'analyse familiale primaire par FBATdos

➤ Robustesse de l'analyse : QQ-plot

À l'échelle génomique, le tracé du QQ-plot montre une distribution observée très proche de la distribution attendue sous H_0 (droite rouge) avec quelques SNPs qui « décrochent » de la droite attendue pour des p-values inférieures à 10^{-5} pouvant correspondre aux marqueurs réellement associés à la maladie. Plus précisément, sans la région HLA (droite grise) qui est notre principal signal, nous sommes juste dessous de la distribution attendue. Enfin, l'estimation du l'IFG, $\lambda = 1,014$, proche de 1, parachève de démontrer la robustesse de l'analyse par FBATdos (**figure 37**).

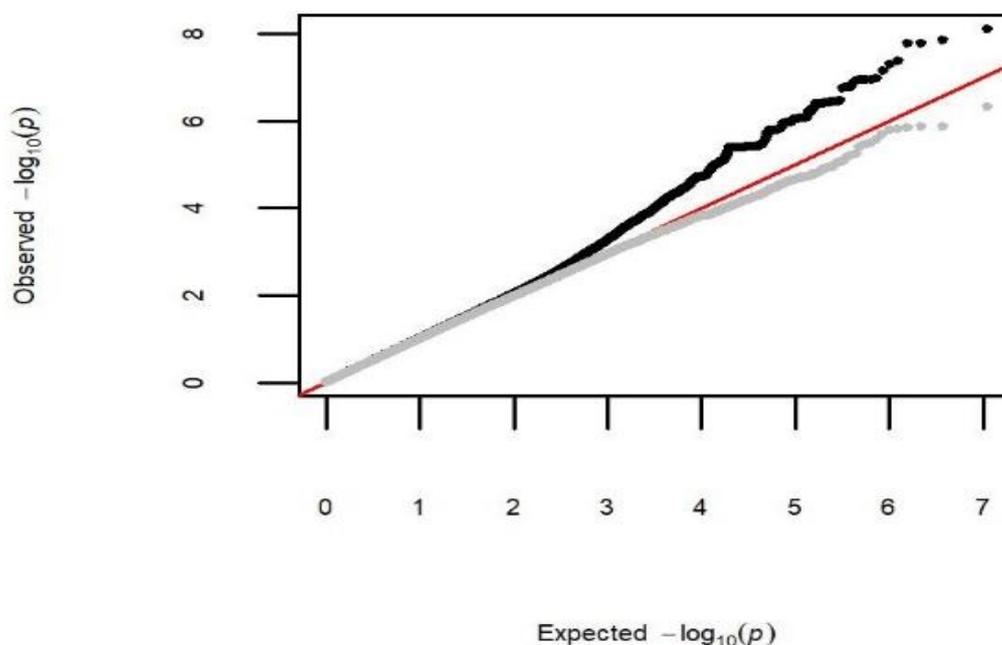


Figure 37 : QQ-plot de l'analyse d'association pangénomique familiale de la lèpre

Modèle additif,

En noir : tous les SNPs,

En gris : tous les SNPs sans la région HLA,

La droite rouge (y=x) représente la distribution attendue sous l'hypothèse nulle de non association.

➤ Résultats d'association : Manhattan-plot

Les résultats d'association sont présentés sous le modèle additif (**figure 38**).

- **Hors-HLA**

- **Signaux évocateurs**

Au total, 21 SNPs formant 8 bins indépendants (**tableau 19**) sur les chromosomes 4, 5, 9, 10, 11, 14 et 21 (2 bins) ont montré une association évocatrice à l'échelle du génome avec la lèpre ($p\text{-val} < 10^{-5}$).

- **Signaux précédemment publiés**

Parmi les 33 résultats significatifs à l'échelle du génome précédemment publiés dans d'autres études, une preuve d'association ($p\text{-val} < 10^{-2}$) avec le même allèle à risque que dans l'étude initiale a été constatée pour 5 de ces SNPs dans notre échantillon primaire vietnamien (**tableau 20**).

- **Région HLA**

Les résultats les plus significatifs ont été obtenus dans la région HLA, dans laquelle une association significative avec la lèpre a été détectée à l'échelle du génome pour 6 SNPs ($p\text{-val} < 5.10^{-8}$) (**tableau 20**). De même, 358 SNPs ayant dépassé le seuil évocateur d'association à l'échelle du génome ont été détectés ($p\text{-val} < 10^{-5}$).

ANALYSE D'ASSOCIATION PANGÉNOMIQUE FAMILIALE

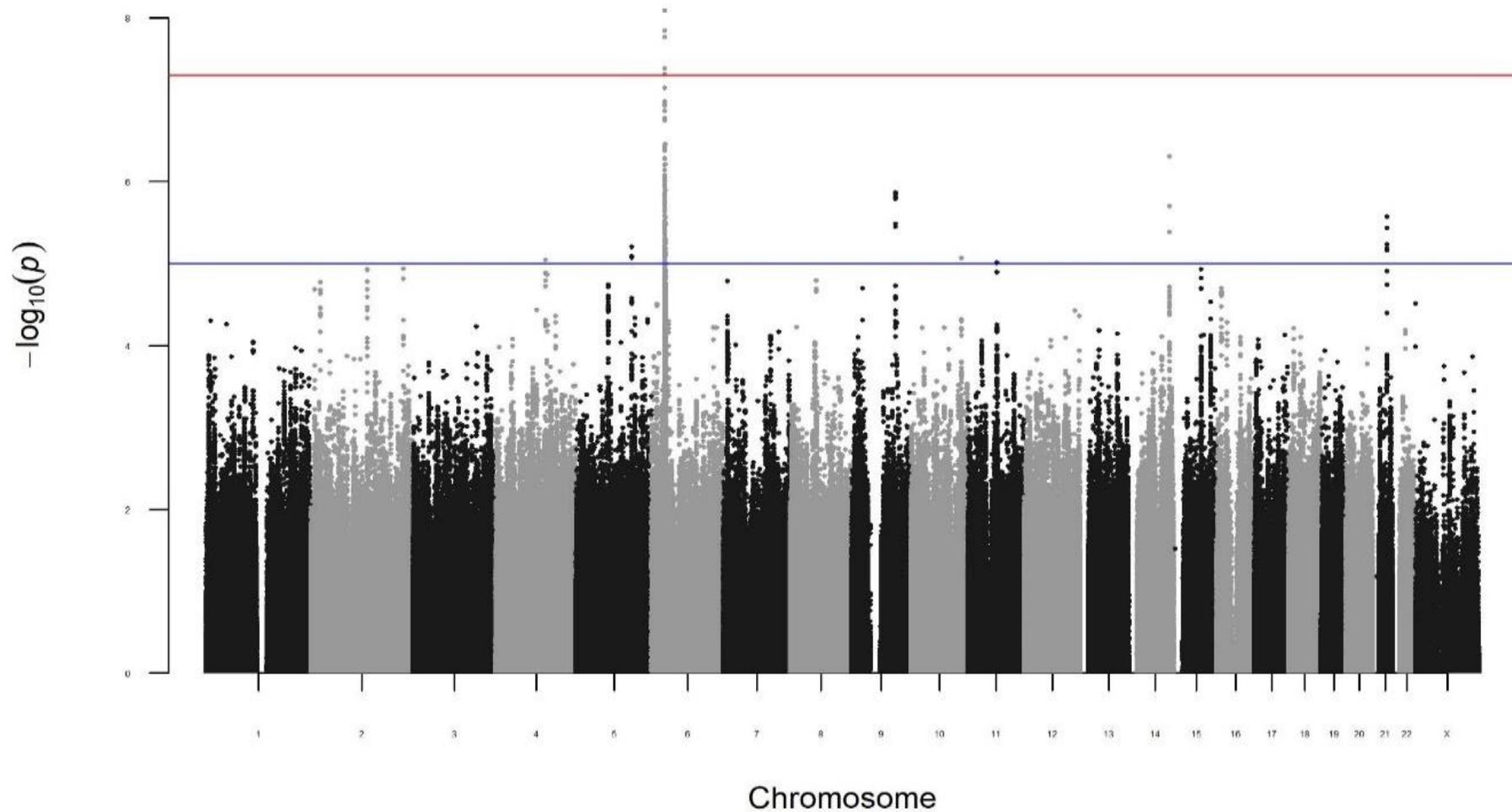


Figure 38 : Manhattan-plot, résultats d'association pangénomique familiale avec la lèpre per se dans l'échantillon primaire vietnamien
*Analyse d'association pangénomique familiale de la lèpre chez 622 enfants atteints avec 5 607 170 SNPs.
Le $-\log_{10}(p\text{-val})$ de chaque variants (axe des y) est tracé en fonction de sa position chromosomique (axe des x, positions à partir de hg19).
Les lignes rouge et bleue indiquent respectivement le seuil de significativité ($5 \cdot 10^{-8}$) et le seuil évocateur (10^{-5}) à l'échelle du génome.*

2.4.2 Résultat de la réplication en cas-contrôles

➤ **Résultats des associations univariées**

- **Hors-HLA**

- **Signaux évocateurs**

Parmi les 8 bins pour lesquels une association suggestive avait été détectée, aucune réplication n'a été validée. Plus précisément, 2 SNPs (régions 4q26 and 9q31.1) sur 8 n'ont pas passé le QC (CR < 95%). Pour les 6 autres, ils présentent une p-val > 0,01 (seuil utilisé pour la détection des signaux de réplication avec un test unilatéral) (**tableau 19**).

- **Signaux précédemment publiés**

Parmi les 5 SNPs précédemment publiés comme étant significatifs à l'échelle du génome et ayant une p-val < 10^{-2} dans notre échantillon familial primaire, 2 signaux ont été répliqués. Plus précisément (**tableau 20**) :

- 1 SNP sur 5 n'a pas passé le QC (CR < 95%).
- 2 SNPs présentent une p-val > 0,01 et ne sont donc pas répliqués.
- 2 SNPs présentent une p-val < 0,01 et sont donc répliqués.
 - Le premier, rs3764147, situé sur le chromosome 13, est un variant faux-sens (p.Ile254Val) du gène *LACC1* ayant une p-val dans l'analyse combinée (échantillon primaire familiale + échantillon de réplication cas-contrôles) de $5,1 \cdot 10^{-14}$. L'OR pour le développement de la lèpre chez les sujets AG par rapport aux sujets GG (ou AA par rapport à AG) est de 1,52 avec un intervalle de confiance à 95% de 1,41-1,63. Ce SNP appartient à un ensemble de 10 SNPs en fort LD ($r^2 > 0,8$), dans la population KHV de 1000 génomes, qui sont situés à l'intérieur ou juste en aval du gène *LACC1*.
 - Le deuxième répliqué, rs6871626, est situé sur le chromosome 5, à une distance de 69 kb en 5' du gène *IL12B*. Une p-val très proche du seuil de significativité de $6,4 \cdot 10^{-8}$ a été obtenue, avec un OR de AC contre AA (ou

ANALYSE D'ASSOCIATION PANGÉNOMIQUE FAMILIALE

CC contre AC) de 0,73 [0,61-0,84]. De plus, ce SNP ne fait pas partie du panel de référence de la phase 3 de 1000 génomes et aucune information sur les LD n'était disponible pour la population KHV. Néanmoins, il a été séquencé avec succès dans la base de données gnomAD (The Genome Aggregation Database) [466].

- **Région HLA**

Nous avons génotypé 88 SNPs dans la région du CMH et nous les avons utilisés pour imputer les 364 autres SNPs ayant une p-val $< 10^{-5}$ dans l'échantillon familial primaire. Plus de 98% de ces SNPs (358/364) ont été imputés avec un score d'information $> 0,8$.

- Les 6 SNPs HLA ayant une p-val significative à l'échelle du génome ont été répliqués, tous avec une p-val unilatérale $< 5 \cdot 10^{-4}$ (**tableau 20**).
- En outre, 319 des 358 SNPs HLA ayant une p-val évocatrice à l'échelle du génome ont tous été répliqués, avec un seuil de valeur p unilatéral de 0,01. Ils se sont regroupés en deux intervalles, le premier de 31 Mo à 31,5 Mo et, le second, de 32,5 Mo à 32,7 Mo (**figure 39 et annexe 2**).

ANALYSE D'ASSOCIATION PANGÉNOMIQUE FAMILIALE

SNP	CHR	SNPs du cluster (en LD)	Génotypé réplication	Étude primaire			Réplication		
				MAF	OR [IC 95%]	P-val	MAF	OR [IC 95%]	P-val
rs72679464	4	-	rs72679464	0,058	0,36 [-0,09-0,81]	9,02.10 ⁻⁰⁶	CR < 95%		
rs329308	5	rs329311, rs329316	rs329308	0,16	1,59 [1,34-1,84]	6,22.10 ⁻⁰⁶	0,16	1,00]0 - 1,16]	0,48
rs72745332	9	rs138270577, rs72745331, rs7873495, rs7846860, rs72745334, rs139715851	rs72745332	0,097	0,46 [0,17-0,75]	1,37.10 ⁻⁰⁶	CR < 95%		
rs10749270	10	rs12245921	rs12245921	0,35	0,68 [0,49-0,86]	8,61.10 ⁻⁰⁶	0,36	-	>0,5
rs10898992	11	-	rs10898992	0,105	0,52 [0,23-0,8]	9,88.10 ⁻⁰⁶	0,11	-	>0,5
rs572428555	14	rs61982614, rs34179372	rs34179372	0,083	2,24 [1,87-2,61]	4,92.10 ⁻⁰⁷	0,07	1,13 [0,91 - Inf[0,175
rs7281793	21	rs9981205, rs7280840	rs7281793	0,461	0,66 [0,48-0,84]	3,69.10 ⁻⁰⁶	0,46	0,98]0 - 1,10]	0,376
rs75025892	21	rs4817512	rs75025892	0,271	0,64 [0,44-0,83]	2,68.10 ⁻⁰⁶	0,26	0,98]0 - 1,11]	0,387

Tableau 19 : Résultats d'associations suite à la procédure de « clumping » pour les 8 signaux évocateurs, hors-HLA, de l'échantillon primaire

Les positions sont données en fonction de hg19,

OR est donné par rapport à l'allèle mineur,

Pour la réplication, il s'agit de tests unilatéraux.

ANALYSE D'ASSOCIATION PANGÉNOMIQUE FAMILIALE

SNP	Chr	Position	m/M	Étude primaire				Réplication		
				Type	MAF	OR [IC 95%]	P-val	MAF	P-val	OR [IC 95%]
Signaux significatifs de la région HLA										
<i>rs3095309</i>	6	31091475	T/C	Imputé	0,476	1,6 [1,43-1,77]	8,2.10 ⁻⁰⁹	0,44	9,46.10 ⁻⁰⁵	1,30 [1,16 - Inf[
<i>rs3094194</i>	6	31136240	G/A	Imputé	0,46	1,71 [1,53-1,88]	1,4.10 ⁻⁰⁸	0,44	5,03.10 ⁻⁰⁶	1,36 [1,21 - Inf[
<i>rs2844633</i>	6	31096189	T/C	Imputé	0,269	0,56 [0,38-0,75]	1,7.10 ⁻⁰⁸	0,30	4,20.10 ⁻⁰⁴	0,79]0 - 0,89]
<i>rs2844634</i>	6	31096184	C/G	Imputé	0,269	0,56 [0,38-0,75]	1,7.10 ⁻⁰⁸	0,30	4,65.10 ⁻⁰⁴	0,79]0 - 0,89]
<i>rs2394945</i>	6	31220971	G/C	Imputé	0,353	1,73 [1,54-1,93]	4,2.10 ⁻⁰⁸	0,35	9,20.10 ⁻⁰⁷	1,42 [1,25 - Inf[
<i>rs3094196</i>	6	31093947	G/A	Imputé	0,269	0,57 [0,38-0,75]	4,9.10 ⁻⁰⁸	0,30	4,70.10 ⁻⁰⁴	0,79]0 - 0,89]
Signaux significatifs hors-HLA précédemment publiés										
<i>rs3762318</i>	1	67597119	G/A	Génotypé	0,05	0,52 [0,12-0,93]	4,41.10 ⁻⁰³	0,06	1,57.10 ⁻⁰¹	0,86]0 - 1,10]
<i>rs2221593</i>	1	212873431	T/C	Génotypé	0,229	1,33 [1,13-1,53]	1,94.10 ⁻⁰⁴	CR < 95%		
<i>rs2058660</i>	2	103054449	G/A	Imputé	0,414	0,75 [0,57-0,92]	5,66.10 ⁻⁰³	0,40	6,46.10 ⁻⁰²	0,90]0 - 1,01]
<i>rs6871626</i>	5	158826792	A/C	Imputé	0,324	0,71 [0,58 - 0,86]	5,12.10 ⁻⁰⁴	0,32	1,37.10 ⁻⁰⁵	0,74]0 - 0,83]
<i>rs3764147</i>	13	44457925	G/A	Génotypé	0,396	1,41 [1,18 - 1,67]	6,56.10 ⁻⁰⁵	0,39	2,36.10 ⁻¹¹	1,61 [1,43 - Inf[

Tableau 20 : Résultats d'association dans l'échantillon primaire et la réplication cas-contrôles pour les 6 signaux significatifs de la région HLA et les 5 signaux, hors-HLA, précédemment publiés

m/M = allèles mineur/majeur,

Les positions sont données en fonction de hg19,

OR est donné par rapport à l'allèle mineur,

Pour la réplication, il s'agit de tests unilatéraux,

rs6871626 a été imputé en utilisant le panel de référence de la phase 1 de 1000 génomes,

Les SNPs en vert sont significatifs dans l'échantillon de réplication.

➤ Résultats des associations multivariées dans la région HLA

Nous avons également étudié le nombre de SNPs dans la région HLA requis pour capturer le signal d'association complet, en effectuant une analyse multivariée. Cette analyse a été menée dans l'échantillon de réplication, qui a fourni l'ensemble le plus fiable de SNPs et a facilité l'analyse en raison de sa nature en cas-contrôles.

Avant d'effectuer cette analyse multivariée, nous avons regroupé les SNPs, tels que décrits dans la section « méthodologie » (procédure de « Clumping »), sur la base de leur le statut génotypique dans l'échantillon de réplication (les SNPs génotypés ont été privilégiés par rapport aux SNPs imputés), leur p-val de réplication unilatérale et un seuil de LD de $r^2 > 0,5$. Ainsi, la procédure de « Clumping » a permis d'identifier 18 SNPs « index » avec une p-val unilatérale $< 0,01$, qui ont ensuite été testés dans le modèle multivariée.

Nous avons constaté que trois SNPs (rs3187964, rs114598080 et rs1265048) sont nécessaires et suffisants pour saisir l'association complète dans la région HLA. L'ajout d'un quatrième SNP n'a pas eu d'effet significatif sur l'amélioration du modèle (p-val = 0,23) (**tableau 21**). Le SNP le plus significatif dans le modèle multivarié est rs3187964 (p-val multivariée = $7,8 \cdot 10^{-8}$). La force de l'association était beaucoup plus faible pour rs114598080 (p-val multivariée = $3,75 \cdot 10^{-5}$), et encore plus faible pour rs1265048, (p-val multivariée = 0,011). Ces trois SNPs sont donc suffisants pour expliquer l'ensemble du signal d'association dans la région HLA trouvé dans notre GWAS.

Nous avons ensuite étudié plus en détail la localisation, le LD dans la population KHV de 1000 génomes et les résultats univariés dans l'échantillon combiné de ces trois SNPs HLA (**tableau 22**).

- Tout d'abord, le SNP rs3187964, appartient à la région HLA classe II (**figure 39.B**). Dans l'analyse combinée, il a une p-val univariée de $8,3 \cdot 10^{-16}$ avec un OR de CG contre GG (ou CC contre CG) de 1,67 [1,55- 1.80]. Ce SNP est situé dans la région 5' (non traduite) du gène *HLA-DQA1* et n'est pas en fort LD ($r^2 > 0,8$) avec d'autres SNPs dans la population KHV de 1000 génomes (analyse LDproxy [467]).
- Les deux autres SNP se trouvent dans la région HLA de classe I avec de nombreux autres gènes (**figure 39.A**).

ANALYSE D'ASSOCIATION PANGÉNOMIQUE FAMILIALE

- Le SNP rs114598080, situé dans un CTCF (ENSR00000262841) 15,6 kb en amont du gène *HLA-C*, a une p-val univariée de $8,8 \cdot 10^{-13}$ dans l'analyse combinée, avec un OR de AG vs. GG (ou AA vs. AG) de 1,47 [1,46-1,48]. Ce SNP appartient à un ensemble de 245 SNPs en fort LD qui s'étend sur près de 200 kb.
- Enfin, le troisième SNP, rs1265048, est situé près des gènes *C6orf15* (1,07 kb), *PSORS1C1* (1,12 kb) et *CDSN* (1,2 kb) et il a donné une p-val univariée de $5,5 \cdot 10^{-11}$ dans l'analyse combinée, avec un OR de CT vs. TT (ou CC vs. CT) de 0.69 [0.58–0.80]. Il n'est pas en fort LD avec d'autres SNPs dans la population KHV de 1000 génomes (analyse LDproxy [467]).

ANALYSE D'ASSOCIATION PANGÉNOMIQUE FAMILIALE

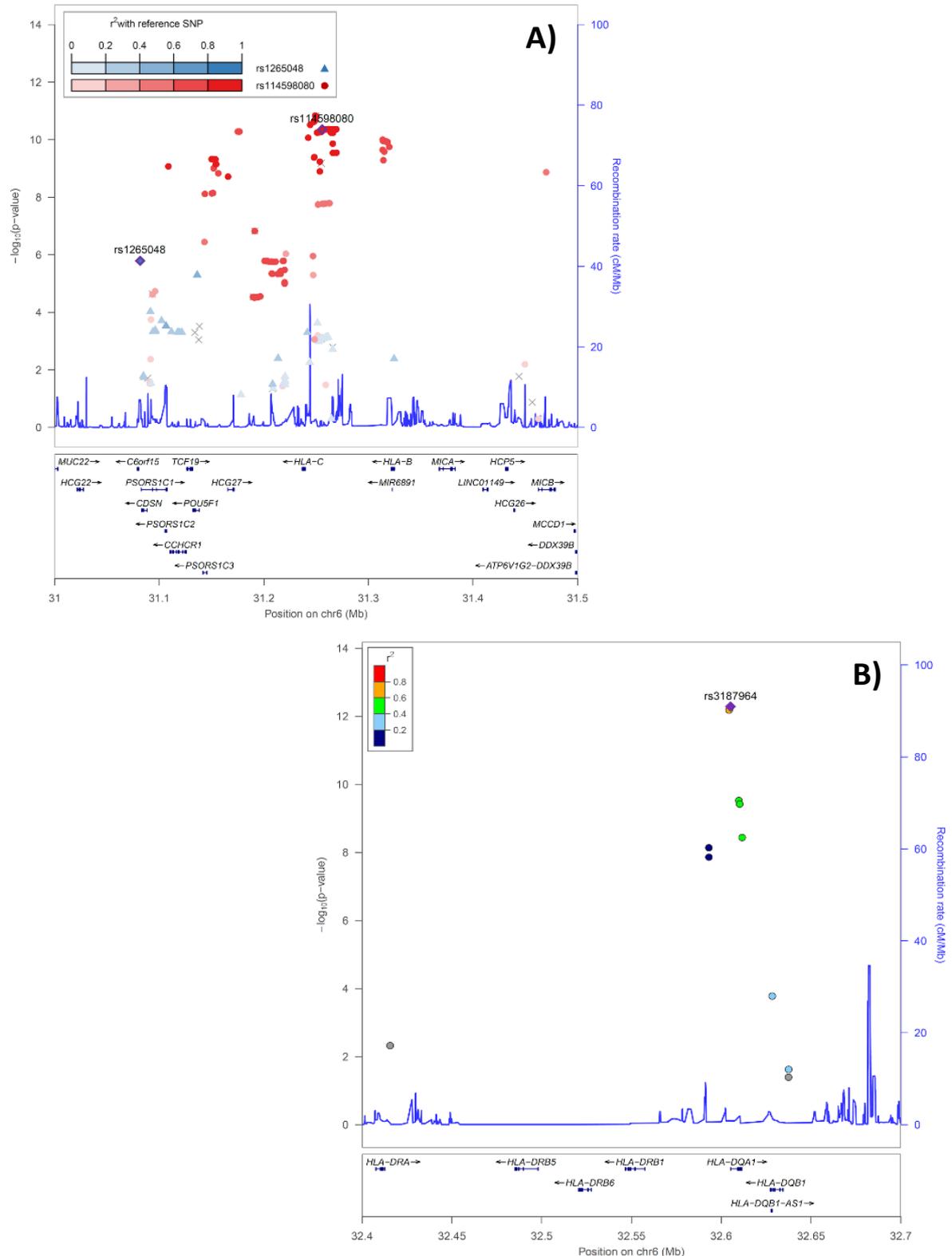


Figure 39 : Association des SNPs HLA de classe I et II dans la réplication cas-contrôles

Graphique tiré de Locus zoom [468] montrant les p -val unilatérales d'association dans la région HLA. Les trois SNPs indépendants identifiés par l'analyse multivariée sont représentés :

- A) *rs1265048* et *rs114598080* par des diamants bleus et rouges respectivement,
- B) *rs3187964* (diamant violet).

Les couleurs indiquent le LD (r^2) avec les 3 SNPs indépendants, tel que calculé dans la population KHV de 1000 génomes. Les SNPs non présents dans la population de référence sont représentés en gris.

ANALYSE D'ASSOCIATION PANGÉNOMIQUE FAMILIALE

SNP (chr:position)	m/M	Gene	MAF	Étude primaire		Réplication			
				Analyse univariée		Analyse univariée		Analyse multivariée	
				P-val	OR [IC 95%]	P-val	OR [IC 95%]	P-val	OR [IC 95%]
rs1265048 (6:31081409)	C/T	<i>C6orf15</i>	0,49	8,72.10 ⁻⁰⁶	0,63 [0,52 -0,63]	1,58.10 ⁻⁰⁶	0,72]0 - 0,81]	1,11.10 ⁻⁰²	0,83 [0,72-0,96]
rs114598080 (6:31255442)	A/G	<i>HLA-C</i>	0,10	3,90.10 ⁻⁰⁶	1,85 [1,39 -2,44]	4,39.10 ⁻¹¹	2,06 [1,7 - Inf[3,75.10 ⁻⁰⁵	1,64 [1,29 -2,09]
rs3187964 (6:32605207)	C/G	<i>HLA-DQA1</i>	0,41	1,28.10 ⁻⁰⁶	1,52 [1,25 - 1,85]	5,28.10 ⁻¹³	1,72 [1,51 - Inf[7,80.10 ⁻⁰⁸	1,53 [1,31 - 1,79]
rs1065386 (6:31324547)	C/G	<i>HLA-B</i>	0,44	3,87.10 ⁻⁰⁷	0,57 [0,37-0,78]	4,13.10 ⁻⁰³	0,82]0 - 0,93]	2,20.10 ⁻⁰¹	

Tableau 21 : Résultats de l'analyse multivariée dans la région HLA (échantillon de réplication)

m/M = allèles mineur/majeur,

Les positions sont données en fonction de hg19,

OR est donné par rapport à l'allèle mineur,

Pour la réplication, il s'agit de tests unilatéraux.

ANALYSE D'ASSOCIATION PANGÉNOMIQUE FAMILIALE

SNP (chr:position)	m/M	Gene	MAF	Étude primaire		Réplication		Analyse combinée (Primaire + réplication)	
				P-val	OR [IC 95%]	P-val	OR [IC 95%]	P-val	OR [IC 95%]
rs1265048 (6:31081409)	C/T	<i>C6orf15</i>	0,49	8,72.10 ⁻⁰⁶	0,63 [0,52 -0,63]	1,58.10 ⁻⁰⁶	0,72]0 - 0,81]	5,53.10 ⁻¹¹	0,69 [0,58-0,80]
rs114598080 (6:31255442)	A/G	<i>HLA-C</i>	0,10	3,90.10 ⁻⁰⁶	1,85 [1,39 -2,44]	4,39.10 ⁻¹¹	2,06 [1,7 - Inf[8,77.10 ⁻¹³	1,47 [1,46-1,48]
rs3187964 (6:32605207)	C/G	<i>HLA-DQA1</i>	0,41	1,28.10 ⁻⁰⁶	1,52 [1,25 - 1,85]	5,28.10 ⁻¹³	1,72 [1,51 - Inf[8,35.10 ⁻¹⁶	1,67 [1,55-1,80]
rs3764147 (13:44457925)	G/A	<i>LACC1</i>	0,35	6,56.10 ⁻⁰⁵	1,41 [1,18 - 1,67]	2,36.10 ⁻¹¹	1,61 [1,49-Inf[5,06.10 ⁻¹⁴	1,52 [1,41-1,63]
rs6871626^f (5:158826792)	A/C	<i>IL12B</i>	0,35	5,12.10 ⁻⁰⁴	0,71 [0,58 - 0,86]	1,37.10 ⁻⁰⁵	0,74]0-0,86]	6,44.10 ⁻⁰⁸	0,73 [0,61-0,84]

Tableau 22 : Résultats d'association dans l'échantillon primaire, la réplication et l'analyse combinée pour les 5 signaux indépendants identifiés par l'analyse multivariée

m/M = allèles mineur/majeur,

Les positions sont données en fonction de hg19,

La MAF a été estimée à partir des 939 parents non atteints de lèpre de l'échantillon primaire et des 668 témoins de l'échantillon de réplication.

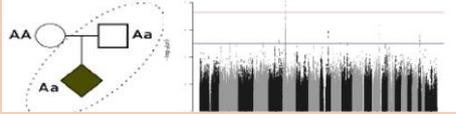
OR est donné par rapport à l'allèle mineur,

Pour la réplication, il s'agit de tests unilatéraux,

rs6871626 a été imputé en utilisant le panel de référence de la phase 1 de 1000 génomes.

Signaux significatifs : hors-HLA

Étude primaire familiale



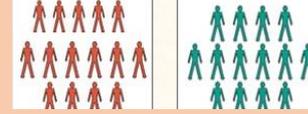
Signaux évocateurs (tableau 19)

- 8 bins indépendants testés en réplification
- 21 SNPs → p-val < 10^{-5}
- Chr : 4, 5, 9, 10, 11, 14 et 21 (2 bins)

Signaux précédemment publiés (tableau 20)

- 5 SNPs testés en réplification
- Significatifs dans d'autres études (p-val < $5 \cdot 10^{-8}$)
- Preuve d'association dans notre étude (même allèle à risque et p-val < 10^{-2})
- Chr : 1 (2 SNPs), 2, 5 et 13

Réplication cas-témoins



Pas de signaux répliqués

- 2 SNPs « failed »
- 6 SNPs non répliqués avec p-val unilatérale > 0,01

2 signaux répliqués

- 1 SNP « failed »
- 2 SNPs non répliqués avec p-val unilatérale > 0,01
- 2 SNPs répliqués
 - rs3764147 (chr13 – *LACC1*, faux sens, p-val_{comb} = $5,1 \cdot 10^{-14}$)
 - rs6871626 (chr5 – intergénique proche *IL12B*, p-val_{comb} = $6,4 \cdot 10^{-8}$)

Signaux significatifs : région HLA

Signaux significatifs (tableau 20)

- 6 SNPs → p-val < $5 \cdot 10^{-8}$

Signaux évocateurs (ANNEXE 2)

- 358 SNPs → p-val < 10^{-5}
- ⇒ 88 signaux choisis pour la réplification (LD clumping) et les 364 signaux sont imputés
- ⇒ 18 SNPs choisis pour l'analyse multivariée (LD clumping)

6 signaux répliqués

- p-val unilatérale < $5 \cdot 10^{-4}$

319 signaux répliqués

- p-val unilatérale < 0,01
- Regroupés en 2 intervalles
 - 31 Mo à 31,5 Mo
 - 32,5 Mo à 32,7 Mo

Analyse multivariée (tableau 21)

- 3 signaux HLA indépendants
- rs3187964 (classe II - *HLA-DQA1*, p-val MV = $7,8 \cdot 10^{-8}$)
- rs114598080 (classe I – proche *HLA-C*, p-val MV = $3,8 \cdot 10^{-5}$)
- rs1265048, (classe I – proche *CDSN*, *PSORS1C1*, p-val MV = 0,01)

Carte mentale 4 : Principaux résultats du GWAS

2.5 Conclusion sur l'analyse d'association familiale pangénomique

Nous rapportons, ici, la première étude d'association pangénomique familiale présentant l'architecture génétique à la susceptibilité de la lèpre au Vietnam. Les principaux signaux identifiés se trouvent dans la région HLA :

- **rs3187964** : le SNP le plus significatif, est situé dans la région HLA de classe II, plus précisément dans la région non traduite en 5' du gène *HLA-DQA1* et ne présente aucun LD ($r^2 > 0,8$) avec d'autres SNPs à proximité dans la population KHV de 1000 génomes. Plusieurs études ont régulièrement fait l'état de l'implication d'allèles HLA de classe II, principalement des allèles du gène *HLA-DRB1*, en tant que facteurs génétiques clés contrôlant la susceptibilité à diverses formes de lèpre [377,469–472]. Une récente méta-analyse basée sur l'imputation des allèles HLA dans une population chinoise a identifié les allèles des gènes *HLA-DRB1* et *HLA-DQA1* et les acides aminés spécifiques de *HLA-DRB1* comme facteurs de protection pour la lèpre [472]. Il est, également, intéressant de noter que dans la base de données GTex V7 [473], le rs3187964 a été identifié comme un fort eQTL (expression Quantitative Trait Loci) pour plusieurs gènes de la région HLA de classe II avec une p inférieure à 10^{-50} pour *HLA-DQB1*, *HLA-DQB2*, *HLA-DQA1* et *HLA-DQA2* dans de multiples tissus, y compris la peau et les nerfs. Il a également été constaté qu'il s'agissait d'un QTL d'épissage pour *HLA-DQA1* ($p = 3,3 \cdot 10^{-7}$). D'autre part, les gènes *HLA-DQ* ont été impliqués dans plusieurs maladies auto-immunes et inflammatoires [474] et, récemment, dans la susceptibilité aux infections streptococciques [475]. D'autre part, on sait que les patients atteints de lèpre sont susceptibles de développer une réaction inflammatoire aiguë, T1R, et une concordance frappante entre les allèles à risques de T1R et les maladies inflammatoires chroniques de l'intestin (MICI), a récemment été montrée [476]. Cependant, les SNPs de la région HLA ne sont pas des loci de risque majeur de T1R, suggérant ainsi, un effet des SNPs HLA de la susceptibilité à la lèpre et non de T1R. Il est, enfin, intéressant de noter qu'il existe plusieurs exemples de loci ayant des effets génétiques miroirs entre des maladies infectieuses et des troubles inflammatoires, où l'allèle protecteur contre l'infection est délétère pour les troubles inflammatoires, ce

ANALYSE D'ASSOCIATION PANGÉNOMIQUE FAMILIALE

qui correspond à l'opinion selon laquelle la fréquence élevée actuelle des troubles inflammatoires/maladies auto-immunes peuvent refléter d'une sélection passée de fortes réponses immunitaires pour combattre l'infection [477,478]. Des études supplémentaires sont nécessaires pour préciser le rôle des gènes *HLA-DQ* dans la lèpre, T1R et aux maladies inflammatoires.

Les deux autres SNPs indépendants et associés à la lèpre dans la région HLA sont situés dans la région de classe I :

- **rs114598080** : le plus significatif appartient à un grand groupe de SNPs en fort LD couvrant les gènes *HLA-C* et *HLA-B* et les gènes hors-HLA *CCHCR1*, *TCF19*, *POU5F1*, *PSORS1C3* et *HCG27*. La structure complexe de LD rend difficile d'identifier le gène responsable de cette association. Néanmoins, un SNP du bin, le rs2394885, a été montré comme tag pour l'allèle *HLA-C*15:05* dans la population vietnamienne [479]. Ce même SNP est également un locus de susceptibilité à la lèpre dans une population indienne [479]. En Chine, la méta-analyse basée sur l'imputation HLA, a montré que l'allèle *HLA-C*08:01* et d'autres acides aminés spécifiques de *HLA-C* sont des facteurs de protection et de risques pour la lèpre [472]. De plus, rs114598080 est également en fort LD avec un variant faux-sens du gène *CCHCR1* potentiellement dommageable. Le gène *CCHCR1* correspond à une protéine dont la fonction est inconnue, mais il a récemment été découvert qu'elle a les propriétés typiques des « P body-resident protéins » [480]. Les corps P sont des composants cytoplasmiques conservés par l'évolution et qui contiennent les protéines impliquées dans la dégradation de l'ARNm, la répression de la traduction et le stockage de l'ARNm [481]. De manière intéressante, il a été démontré que les composants du corps P régulent chez la plante le déclenchement de l'immunité innée [482]. Enfin, le rs114598080 est un fort eQTL, au niveau de la peau et l'artère tibiale, du gène *POU5F1* ($p < 10^{-10}$), également connu sous le nom de *OCT4*, qui code un facteur de transcription principal pour l'auto-renouvellement des cellules pluripotentes [483].

ANALYSE D'ASSOCIATION PANGÉNOMIQUE FAMILIALE

- **rs1265048** : est situé à 1,1 kb en amont de *C6orf15*, 1,1 kb en amont de *PSORS1C1* et 1,2 kb en amont du *CDSN*. Il est intéressant de noter que ce SNP est corrélé avec l'expression de plusieurs gènes, dont *HLA-C* ($p = 1,3 \cdot 10^{-5}$), et *POU5F1* ($p = 2,8 \cdot 10^{-5}$) et de façon sporadique dans certains tissus.

Une étude plus raffinée basée sur le séquençage des gènes HLA, pour caractériser davantage la contribution précise des allèles HLA de classe I et II à la susceptibilité à la lèpre, a été effectuée dans un laboratoire partenaire du nôtre et dont je suis coauteur [484] (**annexe 5**). Un séquençage des génotypes des gènes HLA de classe I et II a été réalisé pour 11 gènes de la région chez 1 155 individus provenant de notre échantillon cas-contrôle précédemment décrit. Les allèles HLA appartenant à un haplotype étendu de *HLA-A* à *HLA-DPB1* ont été associés à un risque de lèpre. Ce signal de susceptibilité pourrait être réduit aux allèles *HLA-DRB1*10:01* et *HLA-DQA1*01:05* qui sont en LD complet. En outre, les haplotypes contenant les allèles *HLA-DRB3~HLA-DRB1*12:02* et *HLA-C*07:06~HLA-B*44:03~HLA-DRB1*07:01* ont été détectés comme deux signaux indépendants et protecteurs pour la lèpre. D'autre part, des signaux précédemment publiés ont été répliqués : *HLA-DRB1*15:01* comme facteur de risque à la lèpre et *HLA-DRB1*04:05~HLA-DQA1*03:03* comme allèles protecteurs. Lorsque les analyses ont été réduites au niveau d'un seul acide aminé, il a été constaté que les associations des allèles HLA ont été largement capturées par quatre acides aminés indépendants à *HLA-DRB1* aux positions 57 (D ou acide aspartique) et 13 (F ou phénylalanine), *HLA-B* à la position 63 (E ou acide glutamique) et *HLA-A* à la position 19 (K ou lysine). Ainsi, les analyses au niveau des acides aminés ont permis de contourner l'ambiguïté causée par les forts LD entre les différents allèles HLA et à identifier quatre facteurs distincts de susceptibilités à la lèpre.

ANALYSE D'ASSOCIATION PANGÉNOMIQUE FAMILIALE

En plus du locus HLA, nous avons pu répliquer les signaux des gènes *LACC1* et *IL12B* précédemment associés à la lèpre dans une population chinoise aussi bien dans notre étude primaire et de répllication :

- **rs3764147** : ce SNP du gène *LACC1* est un variant faux-sens (p.Ile254Val) identifié pour la première fois dans un GWAS population-based de la lèpre dans la population chinoise Han [166], puis répliqué chez les populations Indienne, Africaine [485], Vietnamiennne [168], Yi [486] et Wenshan, Yuxi [385] en Chine. Il a également été démontré que l'allèle de susceptibilité à la lèpre est associé à une sensibilité accrue à la maladie de Crohn [487,488] et aux maladies inflammatoires de l'intestin [489,490]. Ce SNP, rs3764147, a un score CADD (Combined Annotation Dependent Depletion) de 16 [491,492], ce qui le place dans les 2,5% supérieurs des variants délétères par rapport à tous les variants possibles du génome de référence. D'autre part, des études fonctionnelles récentes ont montré que *LACC1* code un régulateur central de la fonction métabolique et de l'état bioénergétique des macrophages [493]. En outre, l'allèle à risque G de rs3764147 est associé, à des niveaux plus faibles d'immunité innée dans les monocytes dérivés de macrophages [494], ce qui pourraient expliquer la plus grande sensibilité aux maladies inflammatoires intestinales et à la lèpre.
- **rs6871626** : est situé dans une région intergénique du chromosome 5. Dans une étude antérieure du génome de la colite ulcéreuse par une méta-analyse, le gène candidat le plus probable attribué à ce SNP par la méthode GRAIL (Gene Relationship Across Implicated Loci) était *IL12B* [495]. Le GRAIL est un outil statistique qui évalue le degré de proximité d'une liste de maladies avec des résumés PubMed (text mining), pour l'annotation de gènes candidats [496]. Le gène *IL12B* code l'IL-12p40, commune à deux interleukines, *IL-12* et *IL-23* qui jouent un rôle essentiel dans l'immunité et, plus précisément, dans la lutte anti-mycobactérienne par l'induction de l'IFN- γ [497,498]. Un déficit autosomique récessif complet en *IL12B* est une étiologie génétique à la susceptibilité mendélienne aux maladies mycobactériennes (MSMD), une maladie rare caractérisée par une prédisposition aux mycobactéries faiblement virulentes, comme les vaccins BCG et les mycobactéries environnementales [499]. Par ailleurs,

ANALYSE D'ASSOCIATION PANGÉNOMIQUE FAMILIALE

l'allèle C du rs6871626 est associé à un risque de lèpre, mais il a été montré comme protecteur pour la colite ulcéreuse [495], la spondylarthrite ankylosante et la maladie de Crohn [500], la maladie inflammatoire de l'intestin [489] et l'artérite de Takayasu [501]. Il est intéressant de noter que chez les patients atteints de l'artérite de Takayasu, l'allèle C, qui confère un risque à la lèpre, a été associée à des niveaux anormalement bas d'*IL-12* (composé d'*IL-12p40* et d'*IL-12p35*) dans le plasma et de faibles niveaux d'*IL-12p40* et d'*IL12* dans les surnageants de culture de monocytes/macrophages de patients [502].

Chapitre 3 : Validité de la réplication parentale dans les GWAS familiaux

Dans les études d'association pangénomique, qu'elles soient familiales ou populationnelles, des réplifications sont nécessaires pour confirmer les signaux d'associations génotype-phénotype. Toutefois, les études de réplication dans des échantillons supplémentaires nécessitent des efforts importants. Ici, nous évaluons la possibilité de tirer parti des informations phénotypiques et génotypiques parentales disponibles dans les études familiales pour réaliser une étude cas-témoins classique avec ces mêmes parents des échantillons familiaux. Nous démontrons que cette réplication parentale est une stratégie de réplication valable, puissante et sans coût pour confirmer les associations familiales. Enfin, nous illustrons l'utilité de la méthode en reproduisant une association décrite entre l'allèle A du gène (LTA)+80 et l'apparition précoce de la lèpre.

3.1 Méthodes et simulations utilisées

3.1.1 Paramètres utilisés pour la simulation

Tout d'abord, nous avons décidé d'évaluer la validité de notre stratégie par des études de simulation. Nous avons considéré une maladie commune avec une prévalence en fonction de l'âge de 0,08 et 0,02 chez les parents et les enfants, respectivement. Nous avons, ensuite, généré des trios (deux parents et un enfant atteint), ainsi qu'un locus causal bi-allélique et un locus de marqueur bi-allélique.

Le locus causal est caractérisé par deux allèles D et d dont les fréquences dans la population sont q et $1 - q$, respectivement. L'effet de l'allèle D dans la population a été modélisé par trois pénétrances f_{dd} , f_{dD} et f_{DD} qui sont les probabilités de présenter la maladie pour un sujet ayant un génotype dd , dD ou DD , respectivement. Les pénétrances ont été paramétrées en termes de risques génotypiques relatifs (RGR), $\gamma_1 = \frac{f_{dD}}{f_{dd}}$, $\gamma_2 = \frac{f_{DD}}{f_{dd}}$. Ces deux paramètres représentent le risque relatif d'être atteint pour une personne portant une (génotype dD) ou deux (génotype DD) copies de cet allèle par rapport à une personne ne

VALIDITÉ DE LA RÉPLICATION PARENTALE DANS LES GWAS FAMILIAUX

portant aucune copie (génotype dd). Plusieurs modèles génétiques ont été caractérisés : un modèle multiplicatif correspond à $\gamma_1 = (\gamma_2)^2$, un modèle dominant est $\gamma_1 = \gamma_2 \geq 1$ et un modèle récessif $\gamma_1 = 1$ et $\gamma_2 \geq 1$. Les modèles $X1, D1$ et $R1$ correspondent à un effet génétique modeste tandis que $X2, D2$ et $R2$ à un effet génétique fort (**tableau 23**).

Pour résumer, nous avons considéré une maladie commune avec une prévalence F dépendant de l'âge, c'est-à-dire que la prévalence a été fixée à 0,08 et 0,02 chez les parents et les enfants, respectivement. Notez qu'une fois que F, q, γ_1 et γ_2 ont été fixés, fdd et les deux autres pénétrances peuvent être dérivées de la relation suivante [503] :

$$F = [q^2\gamma_2 + 2q(1 - q)\gamma_1 + (1 - q)^2]fdd$$

Modèles	γ_1	γ_2	q
Multiplicatif			
X1	1,5	2,5	0,2
X2	2,5	6,25	0,2
Dominant			
D1	2	2	0,05
D2	4	4	0,05
Récessif			
R1	1	2	0,3
R2	1	4	0,3

Tableau 23 : Caractéristiques des modèles génétiques simulés

$\gamma_1 = fdD/fdd$ et $\gamma_2 = fDD/fdd$ où fdd, fdD et fDD sont les pénétrances (probabilités de présenter la maladie pour un sujet ayant un génotype dd, dD ou DD , respectivement).

3.1.2 Hypothèses de simulation

Sous l'hypothèse nulle (H_0) d'absence de LD, nous avons supposé que le locus causal et celui du marqueur ne sont pas liés et que la fréquence de l'allèle du marqueur était de 0,5. Sous l'hypothèse alternative (H_1), nous avons supposé un LD parfait ($r^2 = 1$) entre les allèles du locus causal et le second marqueur.

Les trios ont été générés selon différents modèles de simulation. Chaque réplicat est constituée de 400 trios avec un enfant atteint. Pour chacun d'entre eux, nous avons calculé la statistique du TDT [422], noté χ^2 TDT, et la statistique du test génotypique (χ^2 avec 2 ddl, **tableau 24**) de l'étude cas-témoin parentale (CCP) correspondante (les cas sont les parents atteints et les témoins les parents non atteints), notée χ^2 CCP. De plus, sous H_1 , lorsque des modèles récessifs ou dominants ont été simulés, nous avons également pris en compte les tests spécifiques correspondants pour le TDT (comme décrit dans FBAT [451] et pour les études CCP (χ^2 avec 1 ddl). Enfin, nous avons répété le processus de simulation jusqu'à ce que le nombre de réplicats souhaité soit atteint dans chaque modèle.

Individu Génotype	Cas : parents atteints	Contrôles : parents non atteints
<i>Dd</i>	Effectifs observés Et Calculer les effectifs attendus sous H_0	
<i>dD</i>		
<i>DD</i>		

$$\chi^2_{2ddl} = \sum \frac{(\text{Obs} - \text{Att})^2}{\text{Att}}$$

Tableau 24 : Statistique du test génotypique

H_0 : pas d'association entre le phénotype et le marqueur testé,

H_1 : association entre le phénotype et le marqueur testé,

Sous le modèle dominant (D dominant) : on pool les individus de génotypes dD et DD (χ^2 avec 1 ddl),

Sous le modèle récessif (D récessif) : on pool les individus de génotypes dd et dD (χ^2 avec 1 ddl).

3.2 Résultats

3.2.1 Indépendance du TDT et du CCP

Comme nous l'avons préalablement énoncé, nous avons d'abord évalué la validité de la stratégie proposée au moyen d'études de simulation. Sous H_0 , nous avons cherché à savoir si la distribution du χ^2 CCP dépendait ou non de celle du χ^2 TDT. De notre processus de simulation, nous avons retenu les 100 000 premiers réplicats ayant un χ^2 TDT $\leq 3,84$ (non significatif au seuil 5%), et les 100 000 premiers réplicats ayant un χ^2 TDT $> 3,84$ (faux positif au seuil 5%). Ensuite, nous avons calculé la proportion de χ^2 CCP qui étaient significatifs aux seuils de 5%, 1% et 0,1% dans ces deux échantillons (c'est-à-dire les erreurs de type I observées de 0,05, 0,01 et 0,001 du test CCP) (**tableau 25**). Quel que soit le seuil d'erreur de type I ou le modèle génétique, la proportion observée de faux positifs χ^2 CCP n'était pas différente entre les deux échantillons (significatifs ou non significatifs pour le TDT), et ils appartenait tous à leurs intervalles de prédiction de 95%.

En outre, nous avons calculé le coefficient de corrélation Spearman (non paramétrique) entre le χ^2 TDT et χ^2 CCP sur 100 000 réplicats pour les six modèles génétiques générés sous H_0 (**figure 40**). Les coefficients de corrélation variaient entre -0,06 et 0,03 et aucun n'était significatif au niveau de 5%. Ces résultats étaient identiques lorsque les coefficients de corrélation étaient estimés séparément dans les échantillons répliqués significatifs et non significatifs du χ^2 TDT (données non présentées). L'ensemble de ces résultats indiquent que la distribution du χ^2 CCP est indépendante de la distribution du χ^2 TDT sous H_0 , et que le χ^2 CCP est donc un test de réplification valable.

3.2.2 Puissance du CCP sous H1

Sous l'hypothèse alternative $H1$ de LD parfait, nous avons généré 1 000 réplicats pour les trois modèles à effet génétique modeste ($X1, D1$ et $R1$). Tout d'abord, nous avons comparé la puissance de l'étude CCP à celle de l'étude TDT. Comme attendu, la puissance de l'étude CCP augmente avec la prévalence parentale, c'est-à-dire avec le nombre de parents atteints, et elle atteint finalement celle du TDT pour une prévalence de 0,15 (**figure 41**).

Ensuite, nous avons comparé la puissance du CCP dérivé de l'échantillon TDT (sélectionné par l'intermédiaire d'un enfant atteint) à celle d'un échantillon parental indépendant du phénotype de l'enfant (similaire à un cas-témoins standard basé sur la population). Par souci de comparabilité, chaque réplicat de ce second échantillon a été générée de manière à inclure le même nombre de parents atteints et non atteints que celui dérivé de l'échantillon CCP. La moyenne (écart-type) de la statistique du test selon les modèles $X1, D1$ et $R1$ était respectivement de 7,6 (6,13), 8,06 (6,89) et 6,51 (6,00) pour l'étude CCP dérivée du TDT, contre 7,35 (5,92), 6,47 (6,27) et 6,20 (5,59) pour l'échantillon indépendant du phénotype de l'enfant ($p < 10^{-6}$ pour toutes les comparaisons). Ce gain de puissance reflète probablement l'enrichissement des cas génétiques lors de la sélection par les enfants atteints, comme cela a déjà été noté pour d'autres modèles familiaux [504].

3.3 Exemples

Pour illustrer notre stratégie, nous avons effectué une analyse cas-témoins sur un sous ensemble des données familiales précédemment décrites. Nous avons sélectionné 194 trios, correspondant à 388 parents, pour lesquelles l'information phénotypique et génotypique étaient complètes et disponibles. Ce même sous-groupe avait été utilisé pour montrer l'association dominante de l'allèle A du gène de la lymphotoxine α (LTA+80A) comme facteur de risque majeur de l'apparition précoce de la lèpre [14]. Soixante-treize parents ont été atteints contre 295 non atteints, et nous avons testé si LTA+80A était un allèle de risque dans cet échantillon CCP. Bien que non significatif au niveau de 5% ($p = 0,16$), nous avons observé un enrichissement de l'allèle A (LTA+80A) chez les parents atteints avec un OR ajusté sur le sexe pour AA/AC vs. CC de 1,32 [0,76-2,27].

VALIDITÉ DE LA RÉPLICATION PARENTALE DANS LES GWAS FAMILIAUX

Étant donné que l'effet s'est d'abord avéré beaucoup plus fort chez les jeunes patients [14], nous avons limité notre analyse aux parents atteints dont l'âge était inférieur à 30 ans ($n = 25$) et avons constaté un effet dominant significatif ($p = 0,025$) de l'allèle A (OR = 2,45 [1,00-6,16]). En raison de l'indépendance de l'échantillon parental démontrée dans les paragraphes ci-dessus, ce résultat est une réplique valide [505] de l'association précédemment décrite entre LTA+80A et l'apparition précoce de la lèpre.

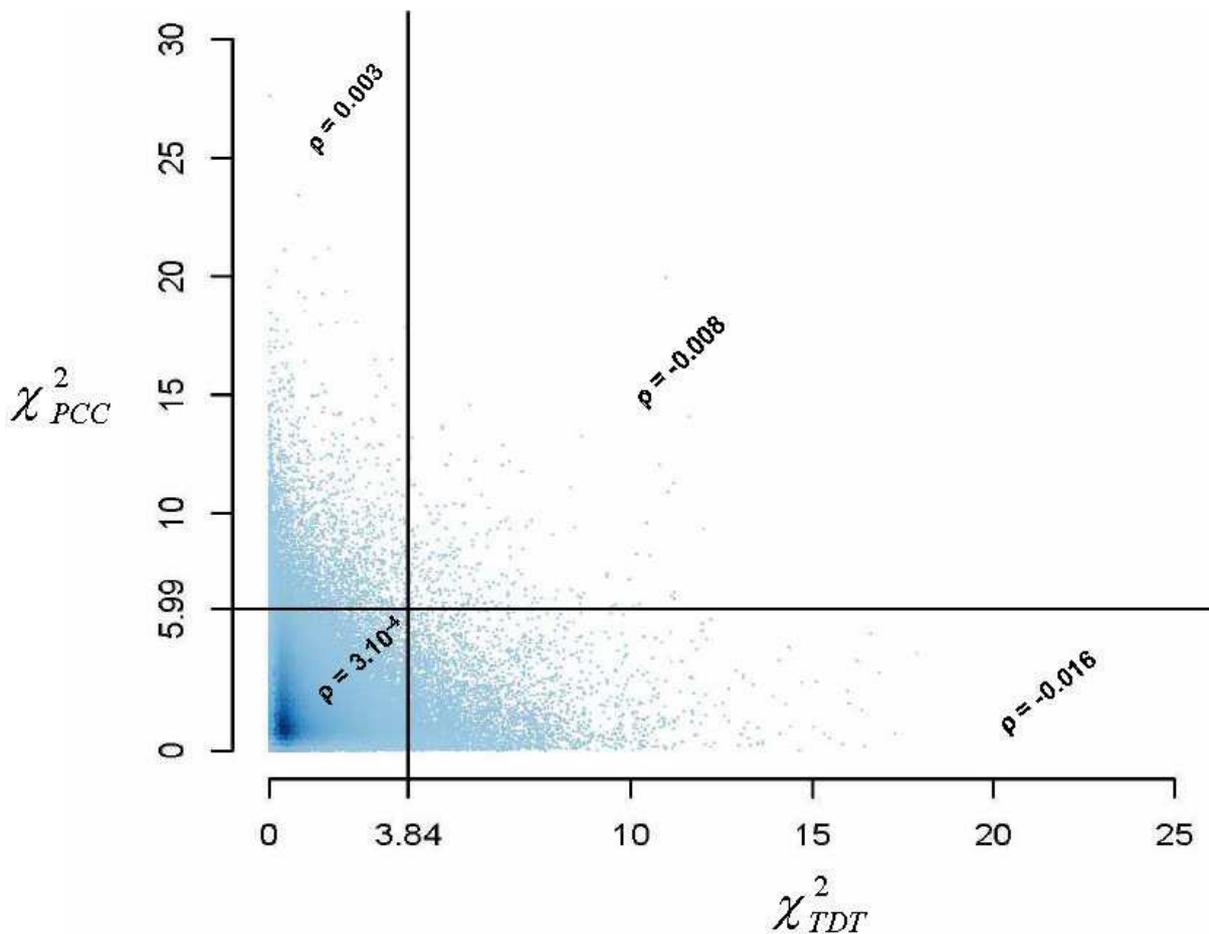


Figure 40 : Indépendance entre le TDT et le cas-contrôle parental

Corrélation entre le χ^2 du TDT et le χ^2 du cas-contrôle parental observée pour 100 000 répliques générés sous R1, correspondant à un effet génétique modéré, ayant pour risque relatif d'être atteint, sachant les génotypes à risques dD et DD, respectivement $\gamma_1 = 1$ et $\gamma_2 = 2$ et une fréquence allélique $d = q = 0,3$. Les lignes noires indiquent au seuil de 5%, les erreurs de type I pour les deux tests. Le coefficient global de corrélation de Spearman ρ est de 0,0014 et ceux de chacun des quatre quadrants sont indiqués sur la figure.

VALIDITÉ DE LA RÉPLICATION PARENTALE DANS LES GWAS FAMILIAUX

α	[IP 95%]	TDT	Modèle génétique					
			X1	X2	D1	D2	R1	R2
0,05	[0,0486-0,0513]	Significatif	0,0507	0,0493	0,0499	0,0498	0,0500	0,0493
		Non significatif	0,0493	0,0498	0,0492	0,0498	0,0498	0,0496
0,01	[0,0094-0,0106]	Significatif	0,0100	0,0098	0,0096	0,0094	0,0100	0,0096
		Non significatif	0,0097	0,0098	0,0098	0,0099	0,0097	0,0099
0,001	[0,0008-0,0012]	Significatif	0,0010	0,0010	0,0011	0,0010	0,0012	0,0009
		Non significatif	0,0010	0,0010	0,0010	0,0010	0,0010	0,0010

Tableau 25 : Distribution des faux positifs du cas-contrôle parental selon l'erreur de type I et en fonction des résultats du TDT

a représente l'erreur de type I pour l'étude cas-témoins parentale.

Les intervalles de prédiction (IP) à 95% représentent la proportion de faux positifs du cas-témoins calculée sur 100 000 simulations.

Le TDT est significatif ou non significatif à un niveau de 5%.

VALIDITÉ DE LA RÉPLICATION PARENTALE DANS LES GWAS FAMILIAUX

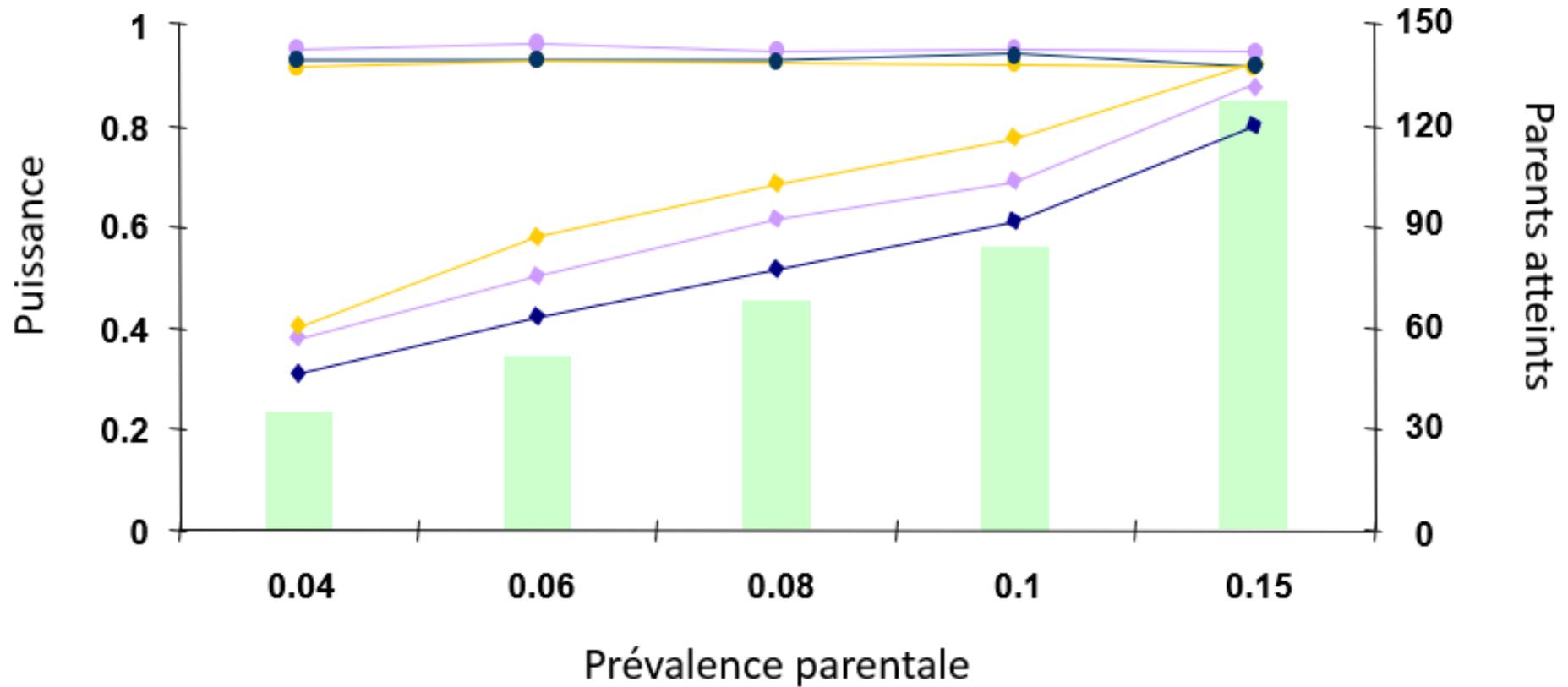


Figure 41 : Puissance du TDT et de l'analyse cas-témoins parentale en fonction de la prévalence parentale

L'axe y (à gauche) représente la puissance au niveau $\alpha = 5\%$.

L'axe Y (à droite) est le nombre moyen de parents atteints par réplicat (le nombre total de parents est de 800 par réplication).

La puissance du TDT (cercles) et du cas-contrôle parental (diamants) est dessinée en bleu, violet et jaune sous les modèles X1, R1 et D1, respectivement.

Le nombre de parents concernés est indiqué par des barres vertes.

Simulations et résultats

Paramètres de simulations :

- **Maladie commune, prévalences F**

- 0,08 chez les parents
- 0,02 chez les enfants

- **Locus marqueur Aa et causal Dd**

- Fréquence de D = q (allèle à risque)
- Fréquence de d = 1 - q
- RGR : $\gamma_1 = \frac{fdD}{fdd}$ et $\gamma_2 = \frac{fDD}{fdd}$

- **Modèles génétiques avec deux effets**

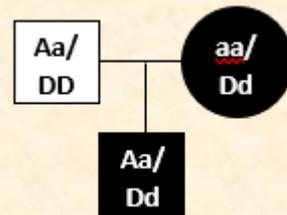
- Multiplicatif $\rightarrow \gamma_1 = (\gamma_2)^2$
- Dominant $\rightarrow \gamma_1 = \gamma_2 \geq 1$
- Récessif $\rightarrow \gamma_1 = 1$ et $\gamma_2 \geq 1$
 - Modeste $\rightarrow X_1, D_1$ et R1
 - Fort $\rightarrow X_2, D_2$ et R2

- **Hypothèses de tests**

- H0 : pas de LD entre les deux loci et q = 0,5
- H1 : LD parfait ($r^2 = 1$) entre les loci

Le TDT et le CCP sont-ils indépendants sous H0 ?

- 400 trios générés par réplicat
- 2 échantillons de 100 000 réplicats
 - $\chi^2_{TDT} \leq 3,84$ (non significatif à 5%)
 - $\chi^2_{TDT} > 3,84$ (faux positifs à 5%)
- ➔ Proportion de χ^2_{CCP} significatifs à 5%, 1% et 0,1% dans ces deux échantillons ?
- ➔ Coefficient de corrélation de Spearman entre le χ^2_{TDT} et χ^2_{CCP} sur 100 000 réplicats pour les six modèles génétiques générés sous H0.



- Quantité de faux positifs χ^2_{CCP} identique entre les 2 échantillons (\forall le modèle et α) et \in [IP 95%].
- Coefficients entre [-0,06;0,03] (non significatif à 5%).

Calcul de la puissance sous H1 ?

- 1 000 réplicats pour X1, D1 et R1
- ➔ Comparaison des puissances de l'étude CCP à celle de l'étude TDT
- 1 000 réplicats pour X1, D1 et R1
- ➔ Comparaison de la puissance du CCP dérivé de l'échantillon TDT (via enfant atteint) à celle d'un échantillon parental indépendant (cas-témoin standard basé sur la population).
 - ⇒ Comparaison des moyennes des statistiques de tests.

- La puissance du CCP augmente avec la prévalence parentale (nombre de parents atteints), elle atteint celle du TDT pour une prévalence de 0,15.
- $p < 10^{-6}$ pour toutes les comparaisons
 - ⇒ Gain de puissance pour l'étude dérivée du CCP

Carte mentale 5 : Réplication cas-contrôle parentale

3.4 Conclusion sur la réplication parentale dans le GWAS familial

Bien que l'utilisation d'un ensemble de données parentales constitue une méthode valable et très rentable pour répliquer et donc confirmer les résultats d'une étude d'association familiale, une limitation évidente est la nécessité d'avoir un nombre suffisant de parents. En effet, comme le TDT est basé sur les génotypes des parents et des enfants atteints uniquement, la mention du phénotype parental lors du recrutement des données n'a pas été systématique dans notre étude. Cependant, trois considérations devraient limiter le risque d'une étude insuffisante.

- Premièrement, dans le contexte d'un GWAS, la taille de l'échantillon doit être assez importante (probablement supérieure aux 400 trios utilisés dans notre étude de simulation), ce qui augmente naturellement la taille de l'échantillon de réplication proposé.
- Deuxièmement, pour une maladie courante, la prévalence est par définition raisonnablement élevée chez les adultes (exemple $> 0,10$ pour l'hypertension ou l'asthme [506]).
- Enfin, on pourrait décider de favoriser l'inclusion des familles qui contiennent un parent atteint, ce qui augmenterait la probabilité de cas génétiques et donc la puissance.

Un deuxième problème associé à l'utilisation d'un ensemble de données parentales comme échantillon de réplication cas-témoins est l'hypothèse d'absence de stratification de population (comme dans toute étude d'association basée sur la population) et d'indépendance entre les parents (spécifique à l'ensemble de données parentales). Dans le contexte des GWAS, il est assez simple de sélectionner un sous-ensemble des SNPs génotypés comme contrôle génomique pour tenir compte d'une éventuelle stratification cryptique de population [507,508]. De même, il est possible d'identifier la parenté entre les parents (par exemple en raison d'un accouplement non aléatoire) en comparant l'IBS observé vs. attendu [509], et de limiter l'analyse aux parents réellement non apparentés.

RÉPLICATION PARENTALE DANS UN GWAS FAMILIAL

La stratégie proposée augmentera généralement l'homogénéité génétique de l'échantillon de réplication et donc la capacité de réplication d'un résultat réel [510]. Ce dernier point ne sera pas toujours le cas, en particulier si la prédisposition génétique est fortement liée à l'âge. Toutefois, il est toujours possible d'effectuer des analyses stratifiées supplémentaires, comme l'illustre notre exemple sur la lèpre. Une autre caractéristique intéressante de notre approche est qu'elle s'étend facilement aux phénotypes catégoriques et quantitatifs, bien que pour ces phénotypes, Van Steen et ses collègues aient déjà proposé une élégante technique de réduction des données qui équivaut à une réplication interne en utilisant un seul ensemble de données familiales [511,512]. Dans l'ensemble, nos observations renforcent l'attrait des modèles familiaux et devraient fournir une aide précieuse aux chercheurs qui prévoient de réaliser des GWAS sur échantillons familiaux.

RÉPLICATION PARENTALE DANS UN GWAS FAMILIAL

Chapitre 4 : Conclusion générale et Perspectives

En résumé, nous avons réalisé le premier GWAS familial de la lèpre chez une population vietnamienne. Les signaux d'association les plus forts ont été observés pour les SNPs dans la région HLA. Nous avons, en outre, décomposé ces signaux d'association de la région HLA en trois signaux indépendants, dont deux correspondent à la région HLA de classe I, et l'autre, le signal de tête, situé dans la région de classe II près du gène *HLA-DQA1*. Nous avons, enfin, validé l'impact de deux variants, l'un dans *LACC1* et l'autre proche de *IL12B*, dont des rôles fonctionnels ont récemment été démontrés. Dans un deuxième temps et au moyen d'une large étude de simulation, nous avons montré que l'approche cas-contrôles parentale pour l'étude de réplication était pertinente : le cas-contrôle chez les parents de l'échantillon primaire est une réplication valide, statistiquement indépendante de l'étude d'association familiale. C'est un argument fort en faveur des approches familiales pour l'exploration pangénomiques de la contribution génétique de l'hôte dans les phénotypes complexes.

Outre ces résultats, il est important de noter que notre étude présente certaines limites : tout d'abord, la taille de notre échantillon peut être considérée comme importante mais reste limitée, avec 622 enfants atteints. En supposant qu'il s'agisse uniquement de trios donc 622 trios, la puissance du TDT est inférieure à 80% pour les SNPs ayant une MAF de 0,05 avec un risque relatif inférieur à 2 ou pour les SNPs ayant une MAF de 0,3 avec un risque relatif inférieur à 1,5. De ce fait, nous avons certainement manqué de nouvelles associations et nous n'avons pas réussi à répliquer certains des loci précédemment identifiés et ayant un effet génétique modeste. Deuxièmement, comme dans toutes les études sur la lèpre, la définition de cas est purement clinique. Bien que la lèpre ait été diagnostiquée par deux médecins expérimentés, nous ne pouvons pas exclure que certains diagnostics soient erronés, ce qui peut affecter la puissance de l'étude. Troisièmement, deux sur huit des signaux évocateurs hors-HLA et un des cinq signaux hors-HLA précédemment publiés n'ont pas pu être testés dans l'échantillon de réplication en raison d'un échec de génotypage. Enfin, l'étude des loci identifiés est uniquement basée sur l'analyse *in silico* en utilisant les ressources publiques disponibles et non pas sur des études fonctionnelles.

CONCLUSION GÉNÉRALE

Dans le but de compléter cette étude, plusieurs approches à suivre pourraient être intéressantes et ce sur cette même cohorte : dans un premier temps, il serait nécessaire d'effectuer des tests d'enrichissement et donc de déterminer si parmi les marqueurs associés des pathways sont « sur-enrichis » avec le phénotype lèpre per se, ce qui permettrait d'avoir une interprétation biologique plus précise des gènes associés et ainsi voir dans quelles voies métaboliques ils sont impliqués. Ainsi, cela pourrait ouvrir la porte à de nouvelles analyses comme le pathway-based, c'est-à-dire tester plusieurs pathways potentiellement associés à la lèpre per se. D'autre part, il serait possible de chercher une association entre marqueurs et un phénotype plus précis que « atteints versus non atteints » comme tester la polarisation, ceci afin de définir des profils génétiques causaux plus précis selon quels phénotypes risquent d'être développés et ainsi de pouvoir, peut-être, prévoir l'apparition de tels ou tels phénotypes et agir avant son développement ou d'adapter les traitements (analyses en cours). Par ailleurs, un des grands avantages des études familiales est de disposer de l'ADN des parents et donc de pouvoir facilement tester le phénomène d'empreinte parentale, tant pour la liaison que pour l'association [503]. Cette analyse pourra donc être réalisée prochainement, sans coût supplémentaire. Enfin, nous n'avons pas testé l'association sur le chromosome sexuel Y dont notre puce d'ADN pouvait en génotyper au maximum 44. Certes, la majorité des GWAS publiés ne rapportent pas l'association sur les chromosomes sexuels, qui posent des problématiques statistiques particulières [513]. Dans FBAT par exemple, il n'est pas possible d'étudier la transmission des allèles sur la portion non commune du chromosome Y, puisqu'il n'y a pas d'hétérozygotie à ces loci. Cependant, des outils robustes sont maintenant disponibles, y compris pour des données familiales [514].

De manière complémentaire, d'autres approches, cherchant à pallier les problèmes liés aux GWAS, existent afin d'identifier de nouveaux gènes de susceptibilité aux maladies complexes et pourraient être utilisées dans le cas de la lèpre. Tout d'abord, les GWAS prennent en compte les SNPs mais pas les autres variants génétiques dits de « grande taille » (comparés aux SNPs) comme les délétions, les duplications (CNVs), les inversions qui sont communes dans les populations humaines et non accessibles par les GWAS [515]. Ces variants représentent une part non négligeable de la variabilité génétique, les prendre en compte dans les analyses génétiques permettrait de couvrir au maximum l'ensemble de cette variabilité. Ainsi, afin d'approfondir les études pangénomiques, des méthodes d'analyse d'exons (qui

représentent qu'une faible partie du génome mais concentrent 85% des mutations potentiellement responsables des maladies) ont été développées : c'est le « Whole Exome Sequencing » (WES) et permet de capturer l'ensemble des exons et d'en déterminer la séquence [516]. Elle permet d'identifier, en théorie, la quasi-totalité des variations de séquences qui existent au niveau des exons entre des sujets malades et non malades. De plus, elle est adaptée à la caractérisation des variants rares qui ne sont pas capturés par les GWAS : en effet, les GWAS considèrent l'hypothèse du « common disease, common variant » s'expliquant par le fait que les maladies communes sont fréquemment retrouvées dans les populations et sont donc dues à des variants fréquents. Or, l'étude de la variabilité génétique des variants rares permettrait peut-être d'avoir suffisamment de puissance statistique pour détecter les variants associés à la maladie. Enfin, d'autres études de types transcriptomiques, ayant pour but d'identifier des gènes différentiellement exprimés (donc au niveau de l'ARNm) entre deux échantillons (exemple : atteints et non atteints ou encore PB ou MB...) pourraient également être envisagées afin de connaître les profils d'expression, ce qui serait une étape supplémentaire et complémentaire à la compréhension du phénotype lèpre per se.

Une approche originale pourrait être d'utiliser le tatou, hôte naturel de *M. leprae*. Son génome du tatou a été séquencé et annoté [517,518]. Le séquençage de plusieurs tatous permettrait d'établir une liste des variants génétiques de cette espèce. Dans un second temps, nous pourrions comparer les variants des mammifères qui développent la lèpre (le tatou et l'homme exclusivement) avec les variants des autres mammifères qui ne la développent pas. Les variants spécifiques à l'homme et au tatou pourraient offrir de précieux renseignements sur la prédisposition génétique à la lèpre per se. Les variants différenciant les tatous PB et MB seraient de bons candidats de gènes influençant la polarisation chez l'homme. Le tatou présente aussi une caractéristique unique qui le rend intéressant pour les études génétiques, la polyembryonie. Les femelles donnent en effet systématiquement naissance à des quadruplés monozygotes. Cette particularité peut nous aider à distinguer les facteurs environnementaux des facteurs génétiques influençant le phénotype. Il a été observé que le nombre de bactéries récoltées dans les biopsies de tatous inoculés expérimentalement par *M. leprae* était semblable parmi les quadruplés élevés ensemble. Ceci est un argument supplémentaire en faveur d'une contribution majeure de la génétique de l'hôte dans la capacité à lutter contre la bactérie [519].

Des limites intrinsèques aux analyses d'association pangénomique sont également à préciser : de manière générale, et contrairement aux études basées sur des gènes candidats, le GWAS permet une étude globale du génome et donc d'éviter tout *à priori* pouvant correspondre à une limite de notre imagination. Pour ce faire, il existe deux grandes méthodes : le family-based et le population-based au design cas-témoins avec des individus non apparentés. Néanmoins, cette deuxième méthode, connaît une importante limitation qui est la stratification de population : en effet, il existe des disparités géographiques et ethniques des fréquences des marqueurs alléliques ou haplotypiques. Ainsi, une association observée peut résulter d'un artéfact dû à la structure de la population étudiée qui est un mélange de sous populations ayant des fréquences alléliques différentes au niveau des SNPs et de la maladie. Il est donc important de sélectionner correctement la population pour éviter de conclure faussement au rôle d'un variant et donc d'un gène candidat. Dans notre étude, la cohorte est assez homogène car tous les individus sont vietnamiens et nous l'avons vérifié par l'ACP réalisée sur les fondateurs afin d'éviter des structures familles sur des familles multiplex avec beaucoup d'enfants par exemple. Des méthodes hybrides alliant les avantages du family-based et du population-based ont été développées afin d'augmenter la puissance de chacune d'elles [520]. De ce fait, l'avantage principal du family-based et la robustesse de ce test vis-à-vis des biais de stratification de population car il ne dépend pas des fréquences alléliques dans la population générale. Néanmoins, cela se paie toutefois par une diminution de la puissance à détecter l'association car il est nécessaire que les parents soient hétérozygotes pour pouvoir réaliser ce test. D'autre part, il requiert le recrutement de familles (préférentiellement trios), ce qui est plus difficile à mettre en œuvre comparé à des individus indépendants et nécessite la connaissance des génotypes parentaux, ce qui n'est pas toujours accessible pour des maladies à début tardif. Des méthodes analogues ont été développées afin de se passer partiellement des données parentales (comme le sib-TDT [442] utilisant les germains du patient) afin de considérer l'ensemble des apparentés pour reconstituer les données parentales, celles-ci sont implémentées dans FBAT et ont donc été utilisées dans notre étude.

D'autre part, quel que soit le test d'association utilisé, la question des faux-positifs (FP) engendrés par les tests multiples demeure : tester n marqueurs avec k allèles revient à réaliser environ $n(k-1)$ tests indépendants, ce qui revient à n tests indépendants. La signification de chaque test doit, en toute rigueur, être divisée par ce facteur (correction de Bonferroni). Par exemple, une valeur de signification nominale de $p = 0,0005$ est requise pour atteindre un simple taux de 5% de faux positifs (l'erreur de type I) lorsque l'on teste indépendamment 100 polymorphismes bi-alléliques sur des gènes différents. Cette correction est cependant très conservatrice, surtout dans le cas des maladies multifactorielles, correspondant à un contexte exploratoire de générations d'hypothèses, où les nombreux tests réalisés ne sont pas souvent indépendants entre eux par l'existence d'un LD. Ainsi, et ce malgré qu'il existe d'autres critères moins stringents de correction comme le FDR (False Discovery Rate) ou taux de fausses découvertes, introduit en 1995 par Benjamini et Hochberg (BH), et défini par l'espérance du taux de FP au niveau α si les hypothèses sont indépendantes : $FDR = E\left(\frac{FP}{P}\right)$ si le nombre de positifs $P > 0$, 0 sinon [521–525]. Nous avons choisi un seuil de significativité de 10^{-5} et d'être moins stringent vis-à-vis du seuil limite de détection, en tolérant un seuil de faux-positifs élevé.

En résumé, le fait de conclure à une association entre un marqueur et le phénotype maladie peut avoir plusieurs interprétations : tout d'abord, il peut s'agir du variant causal ou bien d'un variant en LD avec le réel variant causal. Dans ces deux cas, on obtient un résultat intéressant pour la suite des recherches pour cette maladie. En revanche, il est également possible d'avoir des FP c'est-à-dire une fausse association pouvant être due soit à une stratification de population, soit aux problèmes des tests multiples ou d'autres conditions expérimentales et/ou d'analyses statistiques. De façon générale, lorsqu'il est conclu une association génétique, il n'existe aucun moyen statistique pour trancher s'il s'agit d'une réelle association ou non. Pour le savoir, il faut réaliser des études fonctionnelles sur modèle animal par exemple.

Il est maintenant important d'orienter la suite de cette discussion sur l'approche GWAS en elle-même et les perspectives d'évolution plus générales de la génétique humaine des maladies complexes. Les études d'association pangénomique ont connu une grande popularité à partir de 2005. Ces études ont été un moteur pour la mise en place de collaborations impliquant en général de nombreuses équipes internationales et rassemblant parfois un nombre astronomique de patients et de témoins (plusieurs milliers à dizaines de milliers). Les méta-analyses de GWAS atteignent des sommets étonnants, avec jusqu'à 260 000 individus dans une étude sur la susceptibilité génétique à l'obésité [526]. La dimension de ces études semble justifier parfois à lui seul leur publication dans des revues de très haut niveau. 3 567 publications ont été recensées en 2018 correspondant à 71 673 variants associés à un phénotype donné par le catalogue des GWAS (certains articles comportant plusieurs phénotypes) [527].

Malgré ce succès éditorial, plusieurs voix se sont élevées pour critiquer les GWAS. Ces critiques sont de plusieurs ordres : certains estiment que l'allocation de ressources économiques pour les GWAS a été disproportionnée, d'autres estiment que les résultats scientifiques ont été décevants et n'ont pas contribué à la compréhension physiopathologique des maladies, et certains cliniciens (particulièrement pressés) estiment que le bénéfice des GWAS pour le patient est inexistant [528,529]. Une frange de la communauté scientifique de la génétique épidémiologique même s'est levée contre le GWAS qui, étonnamment, parvient à faire de la génétique sans échantillonner de familles. Une tension est également née du caractère presque automatisé en apparence de la production de GWAS par des consortiums de plus en plus grands. Le GWAS a probablement transposé à la génétique humaine une opposition entre « big science » et « small science », dont les objectifs prioritaires sont parfois divergents, puisqu'il est difficile d'atteindre certaines tailles d'échantillon si l'on s'astreint à une certaine qualité et homogénéité dans la description du phénotype d'intérêt. Quoi qu'il en soit, on peut rester quelque peu dubitatif face à un article rapportant pas moins de 7 GWAS sur des phénotypes indépendants dans un seul et même texte [438]. C'est l'aspect mécanique du GWAS qui prévaut alors, puisqu'il mêle sur quelques pages des résultats concernant des milliers de patients atteints de troubles bipolaires, des milliers de patients atteints de diabète de type II et des milliers de patients ayant fait un infarctus du myocarde.

Un paramètre célèbre de la génétique quantitative animale a pris une place de choix dans le débat sur l'efficacité ou l'inefficacité des GWAS en génétique humaine : l'héritabilité. Les GWAS seraient inefficaces puisque les variants identifiés n'expliquent qu'une proportion modeste (souvent moins de 10%, au mieux 20 à 30%) de l'héritabilité estimée d'une maladie. Le papier ayant introduit le concept d'héritabilité manquante, post-GWAS, a été cité plus de 7 800 fois en 12 ans [18]. La notion d'héritabilité manquante fait consensuellement partie de toute introduction de demande de financement, ou de publication, d'une étude génétique sur une maladie complexe. Les limitations dans l'estimation de l'héritabilité sont en effet tellement manifestes dans le cadre de la génétique humaine que, pour pouvoir poursuivre sa recherche, chacun s'empresserait de les rappeler si cette estimation aboutissait à l'absence d'héritabilité manquante. Reprocher aux GWAS d'être inefficaces car ils n'expliquent qu'une faible part de l'héritabilité, c'est considérer l'estimation de l'héritabilité comme correcte et donc considérer que les maladies complexes humaines sont issues de variants à effet additif, sans interaction gène-gène ni interaction gène-environnement. C'est exactement sous ces hypothèses que sont également construits les modèles génétiques tentant de quantifier le risque de développement d'une maladie complexe à partir des variants associés en GWAS. Il n'est donc pas particulièrement étonnant de constater leur relative inefficacité actuelle [530–532]. Il semble encore actuellement que la prédiction de risque de diabète soit plus fiable par l'interrogatoire simple du patient sur ses antécédents familiaux et personnels que par le génotypage des variants associés par GWAS [533]. La vague de mécontentement anti-GWAS a été suffisamment forte pour générer une série d'articles de revue en défense du GWAS [528,534,535]. Ces articles rappellent que les GWAS ont apporté des preuves solides qu'une susceptibilité génétique complexe existe dans de nombreuses maladies communes. En effet, des *milliers* de signaux répliqués ont été listés en quelques années à peine dans le catalogue des GWAS du NHGRI. Ils rappellent également que l'approche agnostique pangénomique du GWAS est une force, puisqu'elle permet de générer des hypothèses de recherche totalement inattendues, comme cela a été le cas par exemple dans la dégénérescence maculaire liée à l'âge, la maladie de Crohn et l'infection par le virus de l'hépatite C. On peut noter qu'il existe quelques exemples d'élucidation fonctionnelle du mécanisme par lequel l'allèle à risque confère ce risque [536,537]. Ces mécanismes d'élucidation reposent en général sur l'identification d'un effet du génotype à risque sur le niveau d'expression d'un gène d'intérêt (eQTL), ou de méthylation d'une région donnée (metQTL).

Une difficulté importante dans l'acceptation des GWAS comme ayant contribué aux connaissances scientifiques pourrait venir de la nature de leur contribution : la mise au jour de corrélations, qui questionnent la notion de causalité et de preuve scientifique. L'épidémiologie, certes génétique, s'invite dans des revues scientifiques habituées à publier des résultats de biologie moléculaire. Un article de génétique moléculaire contient fréquemment toute la chaîne causale reliant une mutation à une maladie génétique : la mutation est séquencée, la protéine change de taille, sa fonction est annulée, la complémentation par un allèle sauvage restaure cette fonction. Mais voilà qu'à la place de gels et d'imageries confocales apparaissent des Manhattan plots et tableaux d'odds-ratios dont le maximum dépasse exceptionnellement 2. Les seuils de significativité apparaissent eux-mêmes variables. Nous quittons le cadre de la causalité déterministe : ici, les cas comme les témoins portent, ou ne portent pas, le variant déclaré à risque. Par ailleurs, probablement pour des questions de finesse de résolution, la biologie moléculaire actuelle peine à identifier des phénotypes cellulaires covariant avec le génotype à risque dans les maladies complexes. Autrement dit, le signal reste statistique. Aucune cause nécessaire ou suffisante n'est dévoilée : une suite de corrélations ni nécessaires ni suffisantes est listée.

Pour aller un peu plus loin, il semble que la génétique statistique souffre d'un mal très mendélien : la volonté de binariser la découverte. Un variant est répliqué ou ne l'est pas. Mais la difficulté à définir ce qu'est un variant répliqué montre d'emblée l'ampleur du problème. La question du seuil de p-value, par essence arbitraire, est sans cesse reposée. Les variants répliqués publiés sont souvent tout simplement les variants les *mieux* répliqués. Dans cette perspective, il ne s'agit pas tant de valeurs absolues de p-values que de classement des meilleures p-values, comme le montre le processus de sélection des SNPs pour la réplication. On risque alors de tomber dans cette idée irritante qu'un GWAS ne perd jamais : à un certain seuil, certains SNPs seront associés. Mais cette formulation elle-même est binaire, comme l'est l'idée que la recherche est une guerre que l'on perd ou que l'on gagne. En réalité, si les phénotypes et les génotypes sont correctement caractérisés, un GWAS est ce qu'il est : une somme de corrélations, permettant peut-être de prioriser la suite de la recherche. Voilà peut-être la conclusion : les meilleurs signaux de GWAS sont enrichis en découvertes scientifiques. Ce sont des briques pour le futur.

Les GWAS poussent hors de ses limites la théorie statistique en posant des questions insolubles de seuils de significativité. Je pense également que les GWAS poussent l'esprit humain à ses limites. Comment conceptualiser les résultats d'un GWAS ? Comment se les représenter ? Les GWAS communiquent essentiellement leurs résultats par un tableau de quelques signaux principaux associés au phénotype. Des millions de marqueurs ont été testés, et une vingtaine apparaissent dans la publication et sont commentés et seront cités dans l'abstract ou peut-être même dans le titre. L'auteur même d'un GWAS est généralement incapable de résumer les meilleurs signaux de son étude au-delà peut-être de quelques marqueurs. Il ne peut en faire mémoire ni en dessiner un schéma, puisqu'ils ne sont généralement rattachés à aucun sens unifié et restent potentiellement des listes de marqueurs nombreux et anonymes.

Pour finir, il semble important de rappeler que les difficultés statistiques posées par les GWAS sont multipliées de façon littéralement exponentielle dès lors que l'on complexifie légèrement le modèle causal envisagé. Aussi sophistiqués et immenses qu'ils soient, les GWAS ne testent pas moins de millions d'hypothèses simples et ouvertement trop simples : l'association marginale d'un allèle à un phénotype clinique. La génétique complexe, qui repose sur l'hypothèse d'une hérédité polygénique impliquant l'interaction de dizaines ou centaines de variants modifiant le risque de la maladie, est essentiellement cantonnée à tester séquentiellement l'effet marginal de chaque variant dans des modèles linéaires univariés. Il semble pourtant extrêmement vraisemblable que certains variants agissent de concert, par exemple dans une voie métabolique. On pourrait imaginer à l'instar Kenneth Rothman en épidémiologie non génétique que certains complexes causaux, formés d'éléments causaux individuellement ni nécessaires ni suffisants, se révèlent eux-mêmes causes suffisantes [538]. Autrement dit, aucun des dix allèles associés à la maladie dans dix gènes n'est suffisant pour déterminer la survenue de la maladie, mais la présence conjointe chez un individu de, disons, 4 de ces variants suffit à la provoquer. Cette théorie séduisante soulagerait le chercheur attaché à la causalité déterministe. Intuitivement, une autre possibilité est que la génétique humaine ait à faire un saut conceptuel en termes de déterminisme causal analogue à celui que la physique mécaniste a fait au cours du XX^{ème} siècle pour appréhender une réalité quantique.

Sur le plan des phénotypes d'intérêt, une perspective intéressante de la génétique humaine me semble être la génétique de sous-phénotypes quantitatifs : étudier la glycémie plutôt que le diabète, ou la production d'une cytokine en réponse à un agent infectieux plutôt que la maladie infectieuse. L'écart et le nombre de facteurs confondants séparant un sous-phénotype d'un allèle modificateur me paraissent moins abyssaux que celui entre un phénotype clinique et ce même allèle. Ce type d'étude a d'ores et déjà commencé [539]. Dans ce contexte, des phénotypes fortement corrélés sont souvent produits (production de telle cytokine en réponse au stimulus A ou B) et l'analyse combinée de ces phénotypes est également une perspective majeure de développement théorique et appliqué.

Comme précédemment décrit, sur le plan des génotypes d'intérêt, la génétique humaine complexe a progressivement basculé vers le séquençage du génome entier de larges échantillons de cas et de témoins. Ceci va résoudre deux limites du GWAS : l'identification de variants rares (fréquence de l'allèle mineur inférieur à 1%) impliqués dans la susceptibilité aux maladies complexes et l'incomplète couverture du génome pour les variants communs. Une étude montre que les couvertures du génome par les puces à ADN, c'est-à-dire la proportion de variants connus avec une fréquence de l'allèle mineur supérieure à 1% étant en $r^2 > 0.8$ avec un SNP présent sur la puce, était largement surestimée dans les brochures commerciales des sociétés Affymetrix et Illumina. Dans le meilleur des cas (populations d'origine européenne), cette étude révèle que seuls 50% des variants sont effectivement couverts selon cette définition [540]. La meilleure couverture future du génome, associée à l'amélioration progressive de la compréhension du potentiel régulateur des différentes régions du génome humain par l'étude de ses modifications épigénétiques, permettra de faire des pas supplémentaires sur le chemin de ceux qui, en explorant des corrélations, espèrent faire des découvertes.

CONCLUSION GÉNÉRALE

Fort des résultats des trente dernières années, la communauté scientifique peut légitimement fonder des espoirs dans l'étude de la génétique humaine de la lèpre. Notre thèse a illustré que de nombreuses associations sont bien spécifiques à la lèpre et notre modeste contribution a permis d'identifier trois loci HLA et de valider deux loci hors-HLA.

L'absence de culture possible de *M. leprae* in vitro, et l'absence de modèle animal pratique limitent les études biologiques conventionnelles. L'approche de « génétique inverse » soutenue par l'épidémiologie génétique reste donc non seulement attrayante, mais aussi déterminante pour comprendre la physiopathologie de cette maladie tropicale négligée. La mise en évidence de marqueurs associés à un phénotype maladie (ici la lèpre) afin de déceler les causes génétiques de diverses pathologies complexes apparaît comme un enjeu majeur de santé publique.

Toutefois, ces analyses demandent un nombre élevé d'individus de préférence de même origine afin d'augmenter au maximum la puissance, d'éviter les biais de stratification de populations et ainsi être confiant dans les résultats d'analyses statistiques. Ces conditions sont parfois difficiles et longues à satisfaire notamment dans le cas des maladies rares. Les avancées nécessiteront une étroite collaboration entre les médecins, biologistes, épidémiologistes et les généticiens. En effet, l'identification de marqueurs associés à la lèpre permettrait de mieux prévoir l'évolution de la maladie et de proposer des traitements plus ciblés selon le risque génétique individuel. Un petit pas dans la marche pour véritablement éradiquer la maladie.

Cela peut donner suite à des études fonctionnelles, d'autres analyses statistiques et *in fine* à de nouveaux diagnostics et pourquoi pas de nouvelles thérapies plutôt génétiques qu'antibiotiques et ainsi compléter la fonction déficiente pour les maladies infectieuses comme la lèpre. Cette nouvelle approche permettrait d'éviter le phénomène de résistance des bactéries aux antibiotiques (déjà connu pour *M. leprae*) et dont une récente étude britannique a montré que « dix millions de personnes supplémentaires par an pourraient décéder à cause de la résistance aux antibiotiques à l'horizon 2050, soit une toutes les trois secondes » [541,542].

Bibliographie

- [1] ALCAÏS A. DE LA LEPRE AU MAXIMUM DE VRAISEMBLANCE BINOMIALE: HISTOIRE NATURELLE D'UNE NOUVELLE METHODE D'ANALYSE DE LIAISON GENETIQUE. PARIS. 2002;11.
- [2] WHO. LES 10 PRINCIPALES CAUSES DE MORTALITE [INTERNET]. 2020 [CITED 2021 FEB 4]. AVAILABLE FROM: [HTTPS://WWW.WHO.INT/FR/NEWS-ROOM/FACT-SHEETS/DETAIL/THE-TOP-10-CAUSES-OF-DEATH](https://www.who.int/fr/news-room/factsheets/detail/the-top-10-causes-of-death).
- [3] RAOULT D. L'HISTOIRE DES EPIDEMIES ET DES MALADIES INFECTIEUSES [INTERNET]. CAFES THEMA. [CITED 2021 FEB 3]. AVAILABLE FROM: [HTTP://CAFES-THEMA.COM/EVENTS/HISTOIRE-DES-EPIDEMIES-ET-DES-MALADIES-INFECTIEUSES/](http://cafes-thema.com/events/histoire-des-epidemies-et-des-maladies-infectieuses/).
- [4] ANDERSON RM, MAY RM. INFECTIOUS DISEASES OF HUMANS: DYNAMICS AND CONTROL. OXFORD UNIVERSITY PRESS; 1992.
- [5] CAIRNS J, SINGER AL. MATTERS OF LIFE AND DEATH: PERSPECTIVES ON PUBLIC HEALTH, MOLECULAR BIOLOGY, CANCER, AND THE PROSPECTS FOR THE HUMAN RACE. DIANE PUB CO; 1997. 257 P.
- [6] CASANOVA J-L, ABEL L. INBORN ERRORS OF IMMUNITY TO INFECTION: THE RULE RATHER THAN THE EXCEPTION. J EXP MED. 2005;202:197–201.
- [7] QUINTANA-MURCI L, ALCAÏS A, ABEL L, ET AL. IMMUNOLOGY IN NATURA: CLINICAL, EPIDEMIOLOGICAL AND EVOLUTIONARY GENETICS OF INFECTIOUS DISEASES. NAT IMMUNOL. 2007;8:1165–1171.
- [8] NICOLLE C. DESTIN DES MALADIES INFECTIEUSES. 1993.
- [9] ALLISON AC. THE DISTRIBUTION OF THE SICKLE-CELL TRAIT IN EAST AFRICA AND ELSEWHERE, AND ITS APPARENT RELATIONSHIP TO THE INCIDENCE OF SUBTERTIAN MALARIA. TRANS R SOC TROP MED HYG. 1954;48:312–318.
- [10] ALLISON AC. GENETIC CONTROL OF RESISTANCE TO HUMAN MALARIA. CURR OPIN IMMUNOL. 2009;21:499–505.
- [11] CASANOVA J-L, ABEL L. THE GENETIC THEORY OF INFECTIOUS DISEASES: A BRIEF HISTORY AND SELECTED ILLUSTRATIONS. ANNU REV GENOMICS HUM GENET. 2013;14:215–243.
- [12] CASANOVA J-L, ABEL L. HUMAN GENETICS OF INFECTIOUS DISEASES: A UNIFIED THEORY. EMBO J. 2007;26:915–922.
- [13] CASANOVA J-L, ABEL L. PRIMARY IMMUNODEFICIENCIES: A FIELD IN ITS INFANCY. SCIENCE. 2007;317:617–619.

- [14] ALCAIS A, ALTER A, ANTONI G, ET AL. STEPWISE REPLICATION IDENTIFIES A LOW-PRODUCING LYMPHOTOXIN-A ALLELE AS A MAJOR RISK FACTOR FOR EARLY-ONSET LEPROSY. NAT GENET. 2007;39:517–522.
- [15] BAGHDADI JE, ORLOVA M, ALTER A, ET AL. AN AUTOSOMAL DOMINANT MAJOR GENE CONFERS PREDISPOSITION TO PULMONARY TUBERCULOSIS IN ADULTS. J EXP MED. 2006;203:1679–1684.
- [16] MIRA MT, ALCAIS A, VAN THUC N, ET AL. SUSCEPTIBILITY TO LEPROSY IS ASSOCIATED WITH PARK2 AND PACRG. NATURE. 2004;427:636–640.
- [17] PASTEUR L. ETUDES SUR LA MALADIE DES VERS A SOIE: 1: LA PEBRINE ET LA FLACHERIE. GAUTHIER-VILLARS; 1870.
- [18] MANOLIO TA, COLLINS FS, COX NJ, ET AL. FINDING THE MISSING HERITABILITY OF COMPLEX DISEASES. NATURE. 2009;461:747–753.
- [19] MESSALI J-P. LA "LEPRE" DANS LES ECRITS BIBLIQUES ET RABBINIQUES : ASPECTS HISTORIQUES, TEXTUELS ET RITUELS. RELIG UNIV SORBONNE PARIS CITE. 2016;2016USPCA037:01539423.
- [20] TORAL MORENO, LOPEZ DIAZ E, TERESA. LA LEPRE: HISTOIRE D'UNE MALADIE STIGMATISANTE. HIST SCI MEDICALES. 2001;35:13–20.
- [21] [HTTP://LARTDESMETS.E-MONSITE.COM/PAGES/MEDECINE-MEDIEVALE/LA-LEPRE-AU-MOYEN-AGE.HTML](http://lartdesmets.e-monsite.com/pages/medecine-medievale/la-lepre-au-moyen-age.html).
- [22] ALFONSO JL, VICH FA, VILATA JJ, ET AL. FACTORS CONTRIBUTING TO THE DECLINE OF LEPROSY IN SPAIN IN THE SECOND HALF OF THE TWENTIETH CENTURY. INT J LEPR MYCOBACT DIS. 2005;73.
- [23] SREERAMAREDDY C, MENEZES R, KISHORE P. CONCOMITANT AGE OLD INFECTIONS OF MANKIND - TUBERCULOSIS AND LEPROSY : A CASE REPORT. J MED CASE REP. 2007;1:43,.
- [24] HANSEN GA ET LOOFT, C. LEPROSY IN ITS CLINICAL AND PATHOLOGICAL ASPECTS. BRISTOL: J. WRIGHT & CO; 1895.
- [25] MONOT M, HONORÉ N, GARNIER T. ON THE ORIGIN OF LEPROSY. SCIENCE. 2005;308:1040–1042.
- [26] CHASSAGNY LAURE J. LA LEPRE : LE POINT EN 2006, THESE D'EXERCICE [INTERNET]. LIMOGES: UNIVERSITE DE LIMOGES; 2006. AVAILABLE FROM: [HTTP://AURORE.UNILIM.FR/ORI-OAI-SEARCH/NOTICE/VIEW/UNILIM](http://aurore.unilim.fr/ori-oai-search/notice/view/unilim).
- [27] SKINSNESS O. THE PATTERN OF CONCEPT AND REACTION TO LEPROSY IN ORIENTAL ANTIQUITY. LEPR REV. 1964;35:106–122.
- [28] [HTTPS://WWW.LETEMPS.CH/SCIENCES/LADN-PERCER-SECRETS-LEPRE](https://www.letemps.ch/sciences/ladn-percer-secrets-lepre).

- [29] SCHUENEMANN VJ, SINGH P, MENDUM TA. GENOME-WIDE COMPARISON OF MEDIEVAL AND MODERN MYCOBACTERIUM LEPRAE. SCIENCE. 2013;341:179–183.
- [30] [HTTPS://FR.WIKIPEDIA.ORG/WIKI/L%C3%A8PRE](https://fr.wikipedia.org/wiki/L%C3%A8pre).
- [31] LANGUILLON J, CARAYON A, THOIRE D, ET AL. PRECIS DE LEPROLOGIE. ACTA LEPROL. 1988;6.
- [32] BOBIN P. LEPRE. IN : MANIFESTATIONS DERMATOLOGIQUES DES MALADIES INFECTIEUSES, METABOLIQUES ET TOXIQUES. PARIS: SPRINGER; 2008.
- [33] LECLERC H, GAILLARD J-L, SIMONET, ET AL. MICROBIOLOGIE GENERALE: LA BACTERIE ET LE MONDE BACTERIEN. DOIN; 1995.
- [34] SKERMAN VBD, MCGOWAN V, SNEATH PHA. APPROVED LISTS OF BACTERIAL NAMES. INT J SYST EVOL MICROBIOL. 1980;30:225–420, 326.
- [35] WALSH GP, MEYERS WM, BINFORD, ET AL. NATURALLY ACQUIRED LEPROSY IN THE NINE-BANDED ARMADILLO: A DECADE OF EXPERIENCE 1975–1985. J LEUKOC BIOL. 1986;40:645–656.
- [36] AVANZI C, DEL-POZO J, BENJAK A. RED SQUIRRELS IN THE BRITISH ISLES ARE INFECTED WITH LEPROSY BACILLI. SCIENCE. 2016;354:744–747.
- [37] DONHAM KJ, LEININGER J. SPONTANEOUS LEPROSY-LIKE DISEASE IN A CHIMPANZEE. J INFECT DIS. 1977;136:886203.
- [38] MEYERS WM, WALSH GP, BROWN HL. NATURALLY ACQUIRED LEPROSY IN A MANGABEY MONKEY (CERCOCEBUS-SP. INT J LEPR MYCOBACT DIS GWL HANSENS DIS CENT LSU VET SCH RM 11022 SOUTH STAD DR BATON ROUGE 70803-9999. 1980;495–496.
- [39] GORMUS BJ, WOLF RH, BASKIN GB. A SECOND SOOTY MANGABEY MONKEY WITH NATURALLY ACQUIRED LEPROSY: FIRST REPORTED POSSIBLE MONKEY-TO-MONKEY TRANSMISSION. INT J LEPR OTHER MYCOBACT DIS. 1988;56:61–65.
- [40] VALVERDE CR, CANFIELD D, TARARA R, ET AL. SPONTANEOUS LEPROSY IN A WILD-CAUGHT CYNOMOLGUS MACAQUE. INT J LEPR. 1998;66:140–148.
- [41] ROJAS-ESPINOSA O. MYCOBACTERIAL INFECTIONS IN DOMESTIC AND WILD ANIMALS: MYCOBACTERIUM LEPRAE AND MYCOBACTERIUM LEPRÆMURIUM. REV SCI TECH. 2001;20:219–251.
- [42] [HTTPS://WWW.LETEMPS.CH/SCIENCES/ECUREUILS-ROUX-VECTEURS-LEPRE](https://www.letemps.ch/sciences/ecureuils-roux-vecteurs-lepre).
- [43] BLAKE LA, WEST BC, LARY CH, ET AL. ENVIRONMENTAL NONHUMAN SOURCES OF LEPROSY. REV INFECT DIS. 1987;9:562–577.
- [44] TRUMAN R, FINE PE. ENVIRONMENTAL SOURCES OF MYCOBACTERIUM LEPRAE : ISSUES AND EVIDENCE. LEPR REV. 2010;81:89–95.

- [45] WAHYUNI R, ADRIATY D, ISWAHYUDI I, ET AL. MYCOBACTERIUM LEPRAE IN DAILY WATER RESOURCES OF INHABITANTS WHO LIVE IN LEPROSY ENDEMIC AREA OF EAST JAVA. *INDONES J TROP INFECT DIS*. 2010;1:65–68.
- [46] MATSUOKA M, IZUMI S, BUDIAWAN T, ET AL. MYCOBACTERIUM LEPRAE DNA IN DAILY USING WATER AS A POSSIBLE SOURCE OF LEPROSY INFECTION. *INDIAN J LEPR*. 1998;71:61–67.
- [47] LAVANIA M, KATOCH K, KATOCH VM, ET AL. DETECTION OF VIABLE MYCOBACTERIUM LEPRAE IN SOIL SAMPLES : INSIGHTS INTO POSSIBLE SOURCES OF TRANSMISSION OF LEPROSY. *INFECT GENET EVOL*. 2008;8:627–631.
- [48] WHEAT WH, CASALI AL, THOMAS V. LONG-TERM SURVIVAL AND VIRULENCE OF MYCOBACTERIUM LEPRAE IN AMOEBAL CYSTS. *PLOS NEGL TROP DIS*. 2014;8.
- [49] DESIKAN K, SREEVATSA. EXTENDED STUDIES ON THE VIABILITY OF MYCOBACTERIUM LEPRAE OUTSIDE THE HUMAN BODY. *LEPR REV*. 1995;66:287–295.
- [50] FLAGEUL B. LE DIAGNOSTIC DE LA LEPRE. *REV FRANCO PH LAB*. 2011;37–42.
- [51] REIBEL F, CAMBAU E, AUBRY, ET AL. UPDATE ON THE EPIDEMIOLOGY, DIAGNOSIS, AND TREATMENT OF LEPROSY. *MED MAL INFECT*. 2015;45:383–393.
- [52] TRUMAN RW, KRAHENBUHL JL. VIABLE M. LEPRAE AS A RESEARCH REAGENT. *INT J LEPR*.
- [53] SHEPARD CC. THE EXPERIMENTAL DISEASE THAT FOLLOWS THE INJECTION OF HUMAN LEPROSY BACILLI INTO FOOT-PADS OF MICE. *J EXP MED*. 2001;69:445–454.
- [54] KIRCHHEIMER WF ET STORRS, E E. ATTEMPTS TO ESTABLISH THE ARMADILLO AS A MODEL FOR THE STUDY OF LEPROSY. I. REPORT OF LEPROMATOID LEPROSY IN AN EXPERIMENTALLY INFECTED ARMADILLO. *INT J LEPR*. 1971;39:692–701.
- [55] SHARMA R, LAHIRI R, SCOLLARD DM. THE ARMADILLO: A MODEL FOR THE NEUROPATHY OF LEPROSY AND POTENTIALLY OTHER NEURODEGENERATIVE DISEASES. *DIS MODEL MECH*. 2013;6:19–24.
- [56] COLSTON MJ, HILSON GR. GROWTH OF MYCOBACTERIUM LEPRAE AND M. MARINUM IN CONGENITALLY ATHYMIC (NUDE) MICE. *NATURE*. 1976;262:399-401,.
- [57] CONVERSE PJ, HAINES VL, WONDIMU A. INFECTION OF SCID MICE WITH MYCOBACTERIUM LEPRAE AND CONTROL WITH ANTIGEN-ACTIVATED" IMMUNE" HUMAN PERIPHERAL BLOOD MONONUCLEAR CELLS. *INFECT IMMUN*. 1995;63:1047–1054.
- [58] WINTERS MARKA ET HUMPHRES, C ROBERT. INDUCTION OF ANTIGEN-SPECIFIC IMMUNITY AND TOLERANCE TO MYCOBACTERIUM LEPRAE IN LEWIS RATS. *INFECT IMMUN*. 1990;58:495–501.
- [59] AVRIL J-L, DABERNAT H, DENIS F. *BACTERIOLOGIE CLINIQUE*. ELLIPSES EDITION MARKETING SA; 2000.

- [60] [HTTP://WWW.SUR-LES-ANIMAUX.FR/TATOU-A-NEUF-BANDES/](http://www.sur-les-animaux.fr/tatou-a-neuf-bandes/).
- [61] SCOLLARD DM, ADAMS LB, GILLIS TP. THE CONTINUING CHALLENGES OF LEPROSY. *CLIN MICROBIOL REV.* 2006;19:338–381.
- [62] [HTTPS://FR.WIKIPEDIA.ORG/WIKI/SOURIS_NUDE](https://fr.wikipedia.org/wiki/Souris_nude).
- [63] COLE ST, BROSCHE R, PARKHILL J. DECIPHERING THE BIOLOGY OF MYCOBACTERIUM TUBERCULOSIS FROM THE COMPLETE GENOME SEQUENCE. *NATURE.* 1998;393:537–544.
- [64] COLE ST, EIGLMEIER K, PARKHILL J. MASSIVE GENE DECAY IN THE LEPROSY BACILLUS. *NATURE.* 2001;409:1007–1011.
- [65] MONOT M, HONORÉ N, GARNIER T. COMPARATIVE GENOMIC AND PHYLOGEOGRAPHIC ANALYSIS OF MYCOBACTERIUM LEPRAE. *NAT GENET.* 2009;41:1282.
- [66] ALCAÏS A, MIRA M, CASANOVA J-L, ET AL. GENETIC DISSECTION OF IMMUNITY IN LEPROSY. *CURR OPIN IMMUNOL.* 2005;17:44–48.
- [67] HAN XY, SEO Y-H, SIZER KC, ET AL. A NEW MYCOBACTERIUM SPECIES CAUSING DIFFUSE LEPROMATOUS LEPROSY. *AM J CLIN PATHOL.* 2008;130:856–864.
- [68] HAN XY, SIZER KC, THOMPSON EJ, ET AL. COMPARATIVE SEQUENCE ANALYSIS OF MYCOBACTERIUM LEPRAE AND THE NEW LEPROSY-CAUSING MYCOBACTERIUM LEPROMATOSIS. *J BACTERIOL.* 2009;191:6067–6074.
- [69] VERA-CABRERA L, ESCALANTE-FUENTES WG, GOMEZ-FLORES M, ET AL. CASE OF DIFFUSE LEPROMATOUS LEPROSY ASSOCIATED WITH “MYCOBACTERIUM LEPROMATOSIS. *J CLIN MICROBIOL.* 2011;49:4366–4368.
- [70] SOTIRIOU MC, STRYJEWSKA BM, HILL, ET AL. TWO CASES OF LEPROSY IN SIBLINGS CAUSED BY MYCOBACTERIUM LEPROMATOSIS AND REVIEW OF THE LITERATURE. *AM J TROP MED HYG.* 2016;95:522–527.
- [71] VELARDE-FÉLIX JS, ALVARADO-VILLA G, VERA-CABRERA, ET AL. LUCIO’S PHENOMENON” ASSOCIATED WITH MYCOBACTERIUM LEPROMATOSIS. *AM J TROP MED HYG.* 2016;94:483.
- [72] JESSAMINE PG, DESJARDINS M, GILLIS T, ET AL. LEPROSY-LIKE ILLNESS IN A PATIENT WITH MYCOBACTERIUM LEPROMATOSIS FROM ONTARIO, CANADA. *J DRUGS DERMATOL JDD.* 2012;11:229–233.
- [73] HAN XY, SIZER KC, TAN HH. IDENTIFICATION OF THE LEPROSY AGENT MYCOBACTERIUM LEPROMATOSIS IN SINGAPORE. *J DRUGS DERMATOL JDD.* 2012;11:168–172.
- [74] AUNG FM, CHOON SE, WERNER B. ANALYSIS OF THE LEPROSY AGENTS MYCOBACTERIUM LEPRAE AND MYCOBACTERIUM LEPROMATOSIS IN FOUR COUNTRIES. *J DERMATOL.* 2014;51:952–959.

- [75] VIRK A, PRITT B, PATEL R. MYCOBACTERIUM LEPROMATOSIS LEPROMATOUS LEPROSY IN US CITIZEN WHO TRAVELED TO DISEASE-ENDEMIC AREAS. EMERG INFECT DIS. 2017;23:1864.
- [76] HAN XY, SIZER KC, VELARDE-FELIX JS, ET AL. THE LEPROSY AGENTS MYCOBACTERIUM LEPROMATOSIS AND MYCOBACTERIUM LEPRAE IN MEXICO. INT J DERMATOL. 2012;51:952–959.
- [77] SHARMA R, SINGH P, MCCOY RC. ISOLATION OF MYCOBACTERIUM LEPROMATOSIS AND DEVELOPMENT OF MOLECULAR DIAGNOSTIC ASSAYS TO DISTINGUISH M. LEPRAE M LEPROMATOSIS CLIN INFECT DIS. 2019;
- [78] GILLIS TP, SCOLLARD DM, LOCKWOOD DN. WHAT IS THE EVIDENCE THAT THE PUTATIVE MYCOBACTERIUM LEPROMATOSIS SPECIES CAUSES DIFFUSE LEPROMATOUS LEPROSY. LEPR REV. 2011;82:205–209.
- [79] SINGH P, BENJAK A, SCHUENEMANN VJ, ET AL. INSIGHT INTO THE EVOLUTION AND ORIGIN OF LEPROSY BACILLI FROM THE GENOME SEQUENCE OF MYCOBACTERIUM LEPROMATOSIS. PROC NATL ACAD SCI. 2015;112:4459–4464.
- [80] HAN XY, MISTRY NA, THOMPSON EJ. DRAFT GENOME SEQUENCE OF NEW LEPROSY AGENT MYCOBACTERIUM LEPROMATOSIS. GENOME ANNOUNC. 2015;3:00513–00515.
- [81] RODRIGUES LC, LOCKWOOD DN. LEPROSY NOW: EPIDEMIOLOGY, PROGRESS, CHALLENGES, AND RESEARCH GAPS. LANCET INFECT DIS. 2011;11:464–470.
- [82] SAUNDERSON P, GEBRE S, BYASS P. REVERSAL REACTIONS IN THE SKIN LESIONS OF AMFES PATIENTS: INCIDENCE AND RISK FACTORS. LEPR REV. 2000;71:309–317.
- [83] LA LEPRE ET LA MAIN - CHIRURGIE DE LA MAIN LEPREUSE : PLACE DE P. BOURREL DANS SON DEVELOPPEMENT. THESE D'AUDREY DELHOM [INTERNET]. [CITED 2020 MAR 29]. AVAILABLE FROM: [HTTPS://SITES.GOOGLE.COM/SITE/THESEAUDREYDELHOM/LA-LEPRE-ET-LA-MAIN](https://sites.google.com/site/theseaudreydelhom/la-lepre-et-la-main).
- [84] DIOP MM, BERTHE A, DIOUSSE P, ET AL. LA LEPRE: QUATRE CAS DE RETARD TARDIF. REV AFR MALGACHE RECH SCI SANTE. 2014;1.
- [85] MEIMA A, SAUNDERSON PR, GEBRE S, ET AL. DYNAMICS OF IMPAIRMENT DURING AND AFTER TREATMENT: THE AMFES COHORT. LEPR REV. 2001;72:158–170.
- [86] VAN BRAKEL WH, ANDERSON AM, WITHINGTON SG, ET AL. THE PROGNOSTIC IMPORTANCE OF DETECTING MILD SENSORY IMPAIRMENT IN LEPROSY: A RANDOMIZED CONTROLLED TRIAL (TRIPOD 2). LEPR REV. 2003;74:300–310.
- [87] RICHARDUS JH, NICHOLLS PG, CROFT RP, ET AL. INCIDENCE OF ACUTE NERVE FUNCTION IMPAIRMENT AND REACTIONS IN LEPROSY: A PROSPECTIVE COHORT ANALYSIS AFTER 5 YEARS OF FOLLOW-UP. INT J EPIDEMIOL. 2004;33:337–343.
- [88] RIDÉY D, JOPLING WH. CLASSIFICATION OF LEPROSY ACCORDING TO IMMUNITY. A FIVE-GROUP SYSTEM. INT J LEPR. 1966;34:255–273.

- [89] LANGUILLON J. PRECIS DE LEPROLOGIE. J. LANGUILLON; 1999.
- [90] GROUP WS. CHEMOTHERAPY OF LEPROSY FOR CONTROL PROGRAM. WHO TECH REP SER. 1982;
- [91] BLOOM BR, GROSSET JH, LECHAT MF, ET AL. WHO EXPERT COMMITTEE ON LEPROSY. WORLD HEALTH ORGAN-TECH REP SER. 1988;7–50.
- [92] WHO. TECHNICAL REPORT, GENEVA, 1997.
- [93] VAN BRAKEL WH, NICHOLLS PG, DAS L, ET AL. THE INFIR COHORT STUDY: INVESTIGATING PREDICTION, DETECTION AND PATHOGENESIS OF NEUROPATHY AND REACTIONS IN LEPROSY. METHODS AND BASELINE RESULTS OF A COHORT OF MULTIBACILLARY LEPROSY PATIENTS IN NORTH INDIA. LEPR REV. 2005;76:14–34.
- [94] GASCHIGNARD J. GENETIQUE HUMAINE DES FORMES CLINIQUES DE LA LEPRE. SORBONNE PARIS CITE; 2015.
- [95] INTERNATIONAL FEDERATION OF ANTI-LEPROSY ASSOCIATION. COMMENT DIAGNOSTIQUER ET TRAITER LA LEPRE. LONDRES: LA FEDERATION INTERNATIONALE DES ASSOCIATIONS CONTRE LA LEPRE; 2002.
- [96] PATHOLOGY5.PATHOLOGY.JHMI.EDU/MICRO/V20N10.HTM.
- [97] LEPROSYINFORMATIONWEB.BLOGSPOT.COM.
- [98] CHAUFFOUR A, MOUGARI F, LECORCHE E, ET AL. DIAGNOSTIC BACTERIOLOGIE DE LA LEPRE EN FRANCE. :18.
- [99] AUBRY P, GAÛZERE B-A. «LEPRE OU MALADIE DE HANSEN: ACTUALITES 2017. MEDECINE TROP. 2017;1–12.
- [100] PARASCANDOLA J. SULFONES AND THE MIRACLE AT CARVILLE. REV HIST PHARM. 1996;84:409–412.
- [101] WHO. GLOBAL LEPROSY SITUATION, 2009 [INTERNET]. [CITED 2020 FEB 27]. AVAILABLE FROM:
[HTTPS://APPS.WHO.INT/IRIS/BITSTREAM/HANDLE/10665/241399/WER8433_333-340.PDF](https://apps.who.int/iris/bitstream/handle/10665/241399/wer8433_333-340.pdf).
- [102] WHO. LA LEPRE [INTERNET]. [CITED 2020 FEB 27]. AVAILABLE FROM:
[HTTPS://WWW.WHO.INT/FR/NEWS-ROOM/FACT-SHEETS/DETAIL/LEPROSY](https://www.who.int/fr/news-room/fact-sheets/detail/leprosy).
- [103] FAGET M. FONDATION FOLLEREAU. DOSSIER DE PRESSE : 67EME JOURNÉE MONDIALE DES MALADES DE LA LÈPRE. :16.
- [104] WHO | THE FINAL PUSH STRATEGY: QUESTIONS AND ANSWERS [INTERNET]. WHO. WORLD HEALTH ORGANIZATION; [CITED 2020 FEB 27]. AVAILABLE FROM:
[HTTP://WWW.WHO.INT/LEP/STRATEGY/FINAL_PUSH/EN/](http://www.who.int/lep/strategy/final_push/en/).

- [105] WHO. STRATEGIE MONDIALE DE LUTTE CONTRE LA LEPRE 2016–2020 - PARVENIR PLUS RAPIDEMENT A UN MONDE EXEMPT DE LEPRE [INTERNET]. [CITED 2020 FEB 27]. AVAILABLE FROM: HTTPS://APPS.WHO.INT/IRIS/BITSTREAM/HANDLE/10665/208824/9789290225102_FR.PDF?SEQUENCE=15.
- [106] WHO. STRATEGIE MONDIALE DE LUTTE CONTRE LA LEPRE 2016–2020 : GUIDE DE SUIVI ET EVALUATION [INTERNET]. [CITED 2020 MAR 3]. AVAILABLE FROM: <HTTPS://APPS.WHO.INT/IRIS/BITSTREAM/HANDLE/10665/254907/9789290225867-FRE.PDF?SEQUENCE=5>.
- [107] WHO. STRATEGIE MONDIALE DE LUTTE CONTRE LA LEPRE 2016–2020 [INTERNET]. [CITED 2020 FEB 27]. AVAILABLE FROM: <HTTPS://APPS.WHO.INT/IRIS/BITSTREAM/HANDLE/10665/250119/9789290225409-FRE.PDF?SEQUENCE=6>.
- [108] PORTAELS F, DE JONG B. NOUVEAUX DEFIS DANS LA LUTTE CONTRE LA LEPRE. BULL SEANCES ACADEMIE R D'OUTRE-MER. 2014;60.
- [109] ORDRE DE MALTE. DOSSIER-PRESSE-JOURNEE-MONDIALE-DES-LEPREUX_2020.PDF [INTERNET]. [CITED 2020 MAR 3]. AVAILABLE FROM: HTTPS://WWW.ORDREDEMALTEFRANCE.ORG/IMAGES/CAMPAGNES/JML_2020/DOSSIER-PRESSE-JOURNEE-MONDIALE-DES-LEPREUX_2020.PDF.
- [110] WHO. SITUATION DE LA LEPRE DANS LE MONDE, 2018: PARVENIR A UN MONDE EXEMPT DE LEPRE. [INTERNET]. 2019. AVAILABLE FROM: <HTTPS://APPS.WHO.INT/IRIS/BITSTREAM/HANDLE/10665/326775/WER9435-36-EN-FR.PDF>.
- [111] LOCKWOOD DN, SHETTY V, PENNA GO. HAZARDS OF SETTING TARGETS TO ELIMINATE DISEASE: LESSONS FROM THE LEPROSY ELIMINATION CAMPAIGN. BMJ. 2014;348:G1136.
- [112] MOET FJ, PAHAN D, SCHURING RP, ET AL. PHYSICAL DISTANCE, GENETIC RELATIONSHIP, AGE, AND LEPROSY CLASSIFICATION ARE INDEPENDENT RISK FACTORS FOR LEPROSY IN CONTACTS OF PATIENTS WITH LEPROSY. J INFECT DIS. 2006;193:346–353.
- [113] BAKKER MI, HATTA M, KWENANG A, ET AL. RISK FACTORS FOR DEVELOPING LEPROSY-A POPULATION-BASED COHORT STUDY IN INDONESIA. LEPR REV. 2006;77:48–61.
- [114] VAN BEERS SM, HATTA M, KLATSER PR. PATIENT CONTACT IS THE MAJOR DETERMINANT IN INCIDENT LEPROSY: IMPLICATIONS FOR FUTURE CONTROL1. INT J LEPR OTHER MYCOBACT DIS. 1999;67:119.
- [115] FINE PE, STEME JAC, PÖNNIGHAUS JM, ET AL. HOUSEHOLD AND DWELLING CONTACT AS RISK FACTORS FOR LEPROSY IN NORTHERN MALAWI. AM J EPIDEMIOL. 1997;146:91–102.

- [116] RICHARDUS JH, MEIMA A, VAN MARREWIJK CJ, ET AL. CLOSE CONTACTS WITH LEPROSY IN NEWLY DIAGNOSED LEPROSY PATIENTS IN A HIGH AND LOW ENDEMIC AREA: COMPARISON BETWEEN BANGLADESH AND THAILAND. *INT J LEPR OTHER MYCOBACT DIS*. 2005;73:249.
- [117] NEWELL KW. A EPIDEMIOLOGIST'S VIEW OF LEPROSY. *BULL WORLD HEALTH ORGAN*. 1966;34:827.
- [118] PONNIGHAUS JM, FINE PE, STERNE JA, ET AL. EXTENDED SCHOOLING AND GOOD HOUSING CONDITIONS ARE ASSOCIATED WITH REDUCED RISK OF LEPROSY IN RURAL MALAWI. *INT J LEPR MYCOBACT DIS OFF ORGAN INT LEPR ASSOC*. 1994;62:345–352.
- [119] KERR-PONTES LR, BARRETO ML, EVANGELISTA CM, ET AL. SOCIOECONOMIC, ENVIRONMENTAL, AND BEHAVIOURAL RISK FACTORS FOR LEPROSY IN NORTH-EAST BRAZIL: RESULTS OF A CASE–CONTROL STUDY. *INT J EPIDEMIOL*. 2006;35:994–1000.
- [120] SOMMERFELT H, IRGENS LM, CHRISTIAN M. GEOGRAPHICAL VARIATIONS IN THE OCCURRENCE OF LEPROSY: POSSIBLE ROLES PLAYED BY NUTRITION AND SOME OTHER ENVIRONMENTAL FACTORS. *INT J LEPR MYCOBACT DIS OFF ORGAN INT LEPR ASSOC*. 1985;53:524–532.
- [121] GRANT A. LEPROSY AT NAURU SINCE 1928. *INT J LEPR*. 1934;2.
- [122] SKINSNES O.K. EFFECT OF MALNUTRITION ON LEPROSY. *INT J LEPR OTHER MYCOBACT DIS*, 44(3) :374–5, 1976.
- [123] FINE PE. LEPROSY: THE EPIDEMIOLOGY OF A SLOW BACTERIUM. *EPIDEMIOL REV*. 1982;4:161–188.
- [124] DE VRIES JL, PERRY BH. LEPROSY CASE DETECTION RATES BY AGE, SEX, AND POLAR TYPE UNDER LEPROSY CONTROL CONDITIONS. *AM J EPIDEMIOL*. 1985;121:403–413.
- [125] RAO PSS, KARAT ABA, KALIAPERUMAL VG, ET AL. PREVALENCE OF LEPROSY IN GUDIYATHAM TALUK, SOUTH INDIA. PART I. SPECIFIC RATES WITH REFERENCE TO AGE, SEX, AND TYPE. *INT J LEPR*. 1972;40:157–163.
- [126] DOMINGUEZ VM, GARBAJOSA PG, GYI MM, ET AL. EPIDEMIOLOGICAL INFORMATION ON LEPROSY IN THE SINGU AREA OF UPPER BURMA. *BULL WORLD HEALTH ORGAN*. 1980;58:81.
- [127] MUKHERJEE PK, DAS P, RAO PSS. TIME TRENDS IN MB-PB RATIO AMONG UNTREATED LEPROSY PATIENTS ATTENDING A REFERRAL HOSPITAL IN UP, INDIA DURING 2001 TO 2010. *INDIAN J LEPR*. 2013;85:59–64.
- [128] LE GRAND J. KNIGHTS, KNAVES OR PAWNS? HUMAN BEHAVIOUR AND SOCIAL POLICY. *J SOC POLICY*. 1997;26:149–169.
- [129] GUERRA-SILVEIRA F, ABAD-FRANCH F. SEX BIAS IN INFECTIOUS DISEASE EPIDEMIOLOGY: PATTERNS AND PROCESSES. *PLOS ONE*. 2013;8:E62390.

- [130] SETIA MS, STEINMAUS C, HO CS, ET AL. THE ROLE OF BCG IN PREVENTION OF LEPROSY: A META-ANALYSIS. LANCET INFECT DIS. 2006;6:162–170.
- [131] MERLE CS, CUNHA SS, RODRIGUES LC. BCG VACCINATION AND LEPROSY PROTECTION: REVIEW OF CURRENT EVIDENCE AND STATUS OF BCG IN LEPROSY CONTROL. EXPERT REV VACCINES. 2010;9:209–222.
- [132] STANLEY SJ, HOWLAND C, STONE MM, ET AL. BCG VACCINATION OF CHILDREN AGAINST LEPROSY IN UGANDA: FINAL RESULTS. EPIDEMIOL INFECT. 1981;87:233–248.
- [133] LWIN K, SUNDARESAN T, GYI MM, ET AL. BCG VACCINATION OF CHILDREN AGAINST LEPROSY: FOURTEEN-YEAR FINDINGS OF THE TRIAL IN BURMA. BULL WORLD HEALTH ORGAN. 1985;63:1069.
- [134] BAGSHAWE A, SCOTT GC, RUSSELL DA, ET AL. BCG VACCINATION IN LEPROSY: FINAL RESULTS OF THE TRIAL IN KARIMUI, PAPUA NEW GUINEA, 1963-79. BULL WORLD HEALTH ORGAN. 1989;67:389.
- [135] GROUP KPT. RANDOMISED CONTROLLED TRIAL OF SINGLE BCG, REPEATED BCG, OR COMBINED BCG AND KILLED MYCOBACTERIUM LEPRAE VACCINE FOR PREVENTION OF LEPROSY AND TUBERCULOSIS IN MALAWI. THE LANCET. 1996;348:17–24.
- [136] GUPTA MD, VALLISHAYEE RS, ANANTHARAMAN DS, ET AL. COMPARATIVE LEPROSY VACCINE TRIAL IN SOUTH INDIA. INDIAN J LEPR. 1998;70:369–388.
- [137] LY HM, THUY NK, TRACH DD, ET AL. VACCINATION AGAINST LEPROSY AT BEN SAN LEPROSY CENTRE, HO CHI MINH CITY, VIETNAM. VACCINE. 2001;19:3451–3458.
- [138] LOCKWOOD D. LEPROSY. CLIN EVID. 2006;1079–1087.
- [139] CONVIT J, SAMPSON C, ZUNIGA M, ET AL. IMMUNOPROPHYLACTIC TRIAL WITH COMBINED MYCOBACTERIUM LEPRAE/BCG VACCINE AGAINST LEPROSY: PRELIMINARY RESULTS. THE LANCET. 1992;339:446–450.
- [140] VAN THUC N, ABEL L, LAP VD, ET AL. PROTECTIVE EFFECT OF BCG AGAINST LEPROSY AND ITS SUBTYPES: A CASE-CONTROL STUDY IN SOUTHERN VIETNAM. INT J LEPR. 1994;62:532–538.
- [141] KWAN CK, ERNST JD. HIV AND TUBERCULOSIS: A DEADLY HUMAN SYNDROME. CLIN MICROBIOL REV. 2011;24:351–376.
- [142] USTIANOWSKI AP, LAWN SD, LOCKWOOD DN. INTERACTIONS BETWEEN HIV INFECTION AND LEPROSY: A PARADOX. LANCET INFECT DIS. 2006;6:350–360.
- [143] LOCKWOOD DN, LAMBERT SM. HUMAN IMMUNODEFICIENCY VIRUS AND LEPROSY: AN UPDATE. DERMATOL CLIN. 2011;29:125–128.
- [144] PÖNNIGHAUS JM, MWANJASI LJ, FINE PE, ET AL. IS HIV INFECTION A RISK FACTOR FOR LEPROSY? INT J LEPR MYCOBACT DIS OFF ORGAN INT LEPR ASSOC. 1991;59:221–228.

- [145] LUCAS S. HUMAN IMMUNODEFICIENCY VIRUS AND LEPROSY. 1993.
- [146] HUSSAIN T, SINHA S, KULSHRESHTHA KK, ET AL. SEROPREVALENCE OF HIV INFECTION AMONG LEPROSY PATIENTS IN AGRA, INDIA: TRENDS AND PERSPECTIVE¹. INT J LEPR OTHER MYCOBACT DIS. 2005;73:93.
- [147] LEONARD G, SANGARE A, VERDIER M, ET AL. PREVALENCE OF HIV INFECTION AMONG PATIENTS WITH LEPROSY IN AFRICAN COUNTRIES AND YEMEN. J ACQUIR IMMUNE DEFIC SYNDR. 1990;3:1109–1113.
- [148] FROMMEL D, NEGESSE Y, TEKLE-HAIMANOT R, ET AL. HIV INFECTION AND LEPROSY: A FOUR-YEAR SURVEY IN ETHIOPIA. THE LANCET. 1994;344:165–166.
- [149] VAN DEN BROEK J, CHUM HJ, SWAI R, ET AL. ASSOCIATION BETWEEN LEPROSY AND HIV INFECTION IN TANZANIA. INT J LEPR MYCOBACT DIS. 1997;65:203–210.
- [150] GEBRE S, SAUNDERSON P, MESSELE T, ET AL. THE EFFECT OF HIV STATUS ON THE CLINICAL PICTURE OF LEPROSY: A PROSPECTIVE STUDY IN ETHIOPIA. LEPR REV. 2000;71:338–343.
- [151] BORGDORFF MW, VAN DEN BROEK J, CHUM HJ, ET AL. HIV-1 INFECTION AS A RISK FACTOR FOR LEPROSY: A CASE-CONTROL STUDY IN TANZANIA. INT J LEPR OTHER MYCOBACT DIS. 1993;61:556–556.
- [152] OREGÉ PA, FINE PEM, LUCAS SB, ET AL. A CASE CONTROL STUDY ON HUMAN IMMUNODEFICIENCY VIRUS-1 (HIV-1) INFECTION AS A RISK FACTOR FOR TUBERCULOSIS AND LEPROSY IN WESTERN KENYA. TUBER LUNG DIS. 1993;74:377–381.
- [153] KENNEDY C, LIEN RA, STOLZ E, ET AL. LEPROSY AND HUMAN IMMUNODEFICIENCY VIRUS INFECTION. A CLOSER LOOK AT THE LESIONS. INT J DERMATOL. 1990;29:139–140.
- [154] SAMPAIO EP, CANESHI JR, NERY JA, ET AL. CELLULAR IMMUNE RESPONSE TO MYCOBACTERIUM LEPRAE INFECTION IN HUMAN IMMUNODEFICIENCY VIRUS-INFECTED INDIVIDUALS. INFECT IMMUN. 1995;63:1848–1854.
- [155] PEREIRA GA, STEFANI MM, ARAUJO FILHO JA, ET AL. HUMAN IMMUNODEFICIENCY VIRUS TYPE 1 (HIV-1) AND MYCOBACTERIUM LEPRAE CO-INFECTION: HIV-1 SUBTYPES AND CLINICAL, IMMUNOLOGIC, AND HISTOPATHOLOGIC PROFILES IN A BRAZILIAN COHORT. AM J TROP MED HYG. 2004;71:679–684.
- [156] BLUM L, FLAGEUL B, SOW S, ET AL. LEPROSY REVERSAL REACTION IN HIV-POSITIVE PATIENTS. INT J LEPR OTHER MYCOBACT DIS. 1993;61:214–214.
- [157] LAWN SD, WOOD C, LOCKWOOD DN. BORDERLINE TUBERCULOID LEPROSY: AN IMMUNE RECONSTITUTION PHENOMENON IN A HUMAN IMMUNODEFICIENCY VIRUS-INFECTED PERSON. CLIN INFECT DIS. 2003;36:E5–E6.
- [158] COUPPIE P, ABEL S, VOINCHET H, ET AL. IMMUNE RECONSTITUTION INFLAMMATORY SYNDROME ASSOCIATED WITH HIV AND LEPROSY. ARCH DERMATOL. 2004;140:997–1000.

- [159] PIGNATARO P, DA SILVA ROCHA A, NERY JC, ET AL. LEPROSY AND AIDS: TWO CASES OF INCREASING INFLAMMATORY REACTIONS AT THE START OF HIGHLY ACTIVE ANTIRETROVIRAL THERAPY. *EUR J CLIN MICROBIOL INFECT DIS*. 2004;23:408–411.
- [160] PEREIRA GA, STEFANI MM, ARAUJO FILHO JA, ET AL. HUMAN IMMUNODEFICIENCY VIRUS TYPE 1 (HIV-1) AND MYCOBACTERIUM LEPRAE CO-INFECTION: HIV-1 SUBTYPES AND CLINICAL, IMMUNOLOGIC, AND HISTOPATHOLOGIC PROFILES IN A BRAZILIAN COHORT. *AM J TROP MED HYG*. 2004;71:679–684.
- [161] VISCO-COMANDINI U, LONGO B, CUZZI T, ET AL. TUBERCULOID LEPROSY IN A PATIENT WITH AIDS: A MANIFESTATION OF IMMUNE RESTORATION SYNDROME. *SCAND J INFECT DIS*. 2004;36:881–883.
- [162] BWIRE R, KAWUMA HJS. TYPE 1 REACTIONS IN LEPROSY, NEURITIS AND STEROID THERAPY: THE IMPACT OF THE HUMAN IMMUNODEFICIENCY VIRUS. *TRANS R SOC TROP MED HYG*. 1994;88:315–316.
- [163] GOEBEL FD. IMMUNE RECONSTITUTION INFLAMMATORY SYNDROME (IRIS)—ANOTHER NEW DISEASE ENTITY FOLLOWING TREATMENT INITIATION OF HIV INFECTION. *INFECTION*. 2005;33:43–45.
- [164] TALHARI C, MIRA MT, MASSONE C, ET AL. LEPROSY AND HIV COINFECTION: A CLINICAL, PATHOLOGICAL, IMMUNOLOGICAL, AND THERAPEUTIC STUDY OF A COHORT FROM A BRAZILIAN REFERRAL CENTER FOR INFECTIOUS DISEASES. *J INFECT DIS*. 2010;202:345–354.
- [165] LOCKWOOD DNJ, LAMBERT SM. LEPROSY AND HIV, WHERE ARE WE AT? *LEPR REV*. 2010;81:169–175.
- [166] ZHANG F-R, HUANG W, CHEN S-M, ET AL. GENOMEWIDE ASSOCIATION STUDY OF LEPROSY. *N ENGL J MED*. 2009;361:2609–2618.
- [167] ALTER A, GRANT A, ABEL L, ET AL. LEPROSY AS A GENETIC DISEASE. *MAMM GENOME*. 2011;22:19–31.
- [168] GRANT AV, ALTER A, HUONG NT, ET AL. CROHN'S DISEASE SUSCEPTIBILITY GENES ARE ASSOCIATED WITH LEPROSY IN THE VIETNAMESE POPULATION. *J INFECT DIS*. 2012;206:1763–1767.
- [169] GREENSTEIN RJ. IS CROHN'S DISEASE CAUSED BY A MYCOBACTERIUM? COMPARISONS WITH LEPROSY, TUBERCULOSIS, AND JOHNE'S DISEASE. *LANCET INFECT DIS*. 2003;3:507–514.
- [170] SARTOR RB. DOES MYCOBACTERIUM AVIUM SUBSPECIES PARATUBERCULOSIS CAUSE CROHN'S DISEASE? *GUT*. 2005;54:896–898.
- [171] UZOIGWE JC, KHAITSA ML, GIBBS PS. EPIDEMIOLOGICAL EVIDENCE FOR MYCOBACTERIUM AVIUM SUBSPECIES PARATUBERCULOSIS AS A CAUSE OF CROHN'S DISEASE. *EPIDEMIOL INFECT*. 2007;135:1057–1068.

- [172] PIERCE SB, WALSH T, CHISHOLM KM, ET AL. MUTATIONS IN THE DBP-DEFICIENCY PROTEIN HSD17B4 CAUSE OVARIAN DYSGENESIS, HEARING LOSS, AND ATAXIA OF PERRAULT SYNDROME. *AM J HUM GENET.* 2010;87:282–288.
- [173] HUGOT J-P, CHAMAILLARD M, ZOUALI H, ET AL. ASSOCIATION OF NOD2 LEUCINE-RICH REPEAT VARIANTS WITH SUSCEPTIBILITY TO CROHN'S DISEASE. *NATURE.* 2001;411:599–603.
- [174] STROBER W, WATANABE T. NOD2, AN INTRACELLULAR INNATE IMMUNE SENSOR INVOLVED IN HOST DEFENSE AND CROHN'S DISEASE. *MUCOSAL IMMUNOL.* 2011;4:484–495.
- [175] ROACH DR, BRISCOE H, SAUNDERS B, ET AL. SECRETED LYMPHOTOXIN-A IS ESSENTIAL FOR THE CONTROL OF AN INTRACELLULAR BACTERIAL INFECTION. *J EXP MED.* 2001;193:239–246.
- [176] ALTER A, ALCAÏS A, ABEL L, ET AL. LEPROSY AS A GENETIC MODEL FOR SUSCEPTIBILITY TO COMMON INFECTIOUS DISEASES. *HUM GENET.* 2008;123:227–235.
- [177] TUCKER A, CRUZ A, DUCK M, ET AL. PERSONS AFFECTED BY LEPROSY AND THE COVID-19 GLOBAL HEALTH CRISIS: A CONSULTATIVE CALLS REPORT FROM GPZL'S EMERGENCY RESPONSE WORKING GROUP 2. *LEPR REV.* 2020;91:425–430.
- [178] TOPSANTE.COM. LUTTE CONTRE LA LEPRE : LE LOURD IMPACT DU COVID-19 - TOP SANTE [INTERNET]. 2021 [CITED 2021 MAR 5]. AVAILABLE FROM: [HTTPS://WWW.TOPSANTE.COM/MEDECINE/MALADIES-INFECTIEUSES/ZOONOSES/LEPRE-IMPACT-COVID-641429](https://www.topsante.com/medecine/maladies-infectieuses/zoonoses/lepre-impact-covid-641429).
- [179] LEPRE ET COVID-19 : D'UN COMBAT A L'AUTRE AVEC LE DOCTEUR ILBOUDO | VIA @CARENEWS [INTERNET]. [CITED 2021 MAR 5]. AVAILABLE FROM: [HTTPS://WWW.CARENEWS.COM/FR/NEWS/LEPRE-ET-COVID-19-D-UN-COMBAT-A-L-AUTRE-AVEC-LE-DOCTEUR-ILBOUDO](https://www.carenews.com/fr/news/lepre-et-covid-19-d-un-combat-a-l-autre-avec-le-docteur-ilboudo).
- [180] SCHMITZ V, DOS SANTOS JB. COVID-19, LEPROSY, AND NEUTROPHILS. *PLOS NEGL TROP DIS.* 2021;15:E0009019.
- [181] ANTUNES DE, GOULART IMB, GOULART LR. WILL CASES OF LEPROSY REACTION INCREASE WITH COVID-19 INFECTION? *PLOS NEGL TROP DIS.* 2020;14:E0008460.
- [182] WJ B. LOCKWOOD DN. LEPROSY. *LANCET.* 2004;363:1209–1219.
- [183] SANSONETTI P, LAGRANGE PH. THE IMMUNOLOGY OF LEPROSY: SPECULATIONS ON THE LEPROSY SPECTRUM. *REV INFECT DIS.* 1981;3:422–469.
- [184] ILANGUMARAN S, RAMANATHAN S, SHANKERNARAYAN N, ET AL. IMMUNOLOGICAL PROFILES OF LEPROSY PATIENTS AND HEALTHY FAMILY CONTACTS TOWARD M. LEPRAE ANTIGENS. *INT J LEPR MYCOBACT DIS OFF ORGAN INT LEPR ASSOC.* 1996;64:6–14.
- [185] SHAH DH, VARTAK RR, NAIK SS, ET AL. EPIDEMIOLOGICAL STUDIES IN CHILDREN OF A LOW-ENDEMIC REGION, A HIGH-ENDEMIC REGION, AND DWELLERS OF A LEPROSY

- COLONY: EVALUATION OF ANTI-ND-BSA ANTIBODIES AND LEPROMIN RESPONSE. INT J LEPR OTHER MYCOBACT DIS. 1994;62:539–539.
- [186] LASTORIA LC, OPROMOLLA DV, FLEURY RN, ET AL. SERIAL MITSUDA TESTS FOR IDENTIFICATION OF REACTIONAL TUBERCULOID AND REACTIONAL BORDERLINE LEPROSY FORMS. INT J LEPR MYCOBACT DIS OFF ORGAN INT LEPR ASSOC. 1998;66:190–200.
- [187] CHAUDHURI S, HAJRA SK, MUKHERJEE A, ET AL. WHY RELAPSE OCCURS IN PB LEPROSY PATIENTS AFTER ADEQUATE MDT DESPITE THEY ARE MITSUDA REACTIVE: LESSONS FROM CONVIT'S EXPERIMENT ON BACTERIA-CLEARING CAPACITY OF LEPROMIN-INDUCED GRANULOMA. INT J LEPR MYCOBACT DIS OFF ORGAN INT LEPR ASSOC. 1998;66:182–189.
- [188] SOUSA AO, HENRY S, MAROJA FM, ET AL. IGG SUBCLASS DISTRIBUTION OF ANTIBODY RESPONSES TO PROTEIN AND POLYSACCHARIDE MYCOBACTERIAL ANTIGENS IN LEPROSY AND TUBERCULOSIS PATIENTS. CLIN EXP IMMUNOL. 1998;111:48–55.
- [189] CHO SN, CELLONA RV, VILLAHERMOSA LG, ET AL. DETECTION OF PHENOLIC GLYCOLIPID I OF MYCOBACTERIUM LEPRAE IN SERA FROM LEPROSY PATIENTS BEFORE AND AFTER START OF MULTIDRUG THERAPY. CLIN DIAGN LAB IMMUNOL. 2001;8:138–142.
- [190] TORRES P, CAMARENA JJ, GOMEZ JR, ET AL. COMPARISON OF PCR MEDIATED AMPLIFICATION OF DNA AND THE CLASSICAL METHODS FOR DETECTION OF MYCOBACTERIUM LEPRAE IN DIFFERENT TYPES OF CLINICAL SAMPLES IN LEPROSY PATIENTS AND CONTACTS. LEPR REV. 2003;74:18–30.
- [191] KAMAL R, DAYAL R, KATOCH VM, ET AL. ANALYSIS OF GENE PROBES AND GENE AMPLIFICATION TECHNIQUES FOR DIAGNOSIS AND MONITORING OF TREATMENT IN CHILDHOOD LEPROSY. LEPR REV. 2006;77:141–146.
- [192] SHEPARD CC, CHANG YT. EFFECT OF SEVERAL ANTI-LEPROSY DRUGS ON MULTIPLICATION OF HUMAN LEPROSY BACILLI IN FOOTPADS OF MICE. PROC SOC EXP BIOL MED SOC EXP BIOL MED N Y N. 1962;109:636–638.
- [193] SUZUKI Y, MATSUOKA M. [DNA MICROARRAY BASED RAPID DRUG SUSCEPTIBILITY TEST FOR MYCOBACTERIUM LEPRAE]. NIHON HANSEN GAKKAI ZASSHI JPN J LEPR OFF ORGAN JPN LEPR ASSOC. 2006;75:271–277.
- [194] CHAUDHURY S, HAZRA SK, SAHA B, ET AL. AN EIGHT-YEAR FIELD TRIAL ON ANTILEPROSY VACCINES AMONG HIGH-RISK HOUSEHOLD CONTACTS IN THE CALCUTTA METROPOLIS. INT J LEPR OTHER MYCOBACT DIS. 1994;62:389–389.
- [195] WHO. TECHNICAL REPORT, GENEVA, 1977.
- [196] WHO. CHEMOTHERAPY OF LEPROSY FOR CONTROL PROGRAMS. GENEVA, 1988.
- [197] CAMBAU E, SAUNDERSON P, MATSUOKA M, ET AL. ANTIMICROBIAL RESISTANCE IN LEPROSY: RESULTS OF THE FIRST PROSPECTIVE OPEN SURVEY CONDUCTED BY A WHO

- SURVEILLANCE NETWORK FOR THE PERIOD 2009–15. *CLIN MICROBIOL INFECT.* 2018;24:1305–1310.
- [198] WHO (2013). GLOBAL LEPROSY: UPDATE ON THE 2012 SITUATION. *WKLY. EPIDEMIOL. REC.* 365–380.
- [199] FAJARDO TT, VILLAHERMOSA LG, CRUZ ECD, ET AL. A CLINICAL TRIAL OF PEFLOXACIN AND OFLOXACIN IN LEPROMATOUS LEPROSY. *LEPR REV.* 2004;75:389–397.
- [200] PARDILLO FEF, BURGOS J, FAJARDO TT, ET AL. POWERFUL BACTERICIDAL ACTIVITY OF MOXIFLOXACIN IN HUMAN LEPROSY. *ANTIMICROB AGENTS CHEMOTHER.* 2008;52:3113–3117.
- [201] BOERRIGTER G, PONNIGHAUS JM, FINE PE. PRELIMINARY APPRAISAL OF A WHO-RECOMMENDED MULTIPLE DRUG REGIMEN IN PAUCIBACILLARY LEPROSY PATIENTS IN MALAWI. *INT J LEPR MYCOBACT DIS OFF ORGAN INT LEPR ASSOC.* 1988;56:408–417.
- [202] KAR PK, ARORA PN, RAMASASTRY CV, ET AL. A CLINICOPATHOLOGICAL STUDY OF MULTIDRUG THERAPY IN BORDERLINE TUBERCULOID LEPROSY. *J INDIAN MED ASSOC.* 1994;92:336–337.
- [203] DASANANJALI K, SCHREUDER PA, PIRAYAVARAPORN C. A STUDY ON THE EFFECTIVENESS AND SAFETY OF THE WHO/MDT REGIMEN IN THE NORTHEAST OF THAILAND; A PROSPECTIVE STUDY, 1984-1996. *INT J LEPR MYCOBACT DIS OFF ORGAN INT LEPR ASSOC.* 1997;65:28–36.
- [204] ISHII N. RECENT ADVANCES IN THE TREATMENT OF LEPROSY. *DERMATOL ONLINE J.* 2003;9:5.
- [205] EFFICACY OF SINGLE DOSE MULTIDRUG THERAPY FOR THE TREATMENT OF SINGLE-LESION PAUCIBACILLARY LEPROSY. SINGLE-LESION MULTICENTRE TRIAL GROUP. *INDIAN J LEPR.* 1997;69:121–129.
- [206] PATTYN SR, GHYS P, JANSSENS L, ET AL. A RANDOMIZED CLINICAL TRIAL OF TWO SINGLE-DOSE TREATMENTS FOR PAUCIBACILLARY LEPROSY. *LEPR REV.* 1994;65:45–57.
- [207] BAKKER MI, HATTA M, KWENANG A, ET AL. PREVENTION OF LEPROSY USING RIFAMPICIN AS CHEMOPROPHYLAXIS. *AM J TROP MED HYG.* 2005;72:443–448.
- [208] OSKAM L, MI B. REPORT OF THE WORKSHOP ON THE USE OF CHEMOPROPHYLAXIS IN THE CONTROL OF LEPROSY HELD IN AMSTERDAM, THE NETHERLANDS ON 14 DECEMBER 2006. *LEPR REV.* 2007;78:173–185.
- [209] BRITTON WJ, LOCKWOOD DNJ. LEPROSY. *LANCET LOND ENGL.* 2004;363:1209–1219.
- [210] ZHANG F-R, LIU H, IRWANTO A, ET AL. HLA-B*13:01 AND THE DAPSONE HYPERSENSITIVITY SYNDROME. *N ENGL J MED.* 2013;369:1620–1628.
- [211] GIRDHAR BK, GIRDHAR A, KUMAR A. RELAPSES IN MULTIBACILLARY LEPROSY PATIENTS: EFFECT OF LENGTH OF THERAPY. *LEPR REV.* 2000;71:144–153.

- [212] LI HY, HU LF, HUANG WB, ET AL. RISK OF RELAPSE IN LEPROSY AFTER FIXED-DURATION MULTIDRUG THERAPY. *INT J LEPR MYCOBACT DIS OFF ORGAN INT LEPR ASSOC.* 1997;65:238–245.
- [213] SCHREUDER PA. THE OCCURRENCE OF REACTIONS AND IMPAIRMENTS IN LEPROSY: EXPERIENCE IN THE LEPROSY CONTROL PROGRAM OF THREE PROVINCES IN NORTHEASTERN THAILAND, 1987-1995 [CORRECTION OF 1978-1995]. III. NEURAL AND OTHER IMPAIRMENTS. *INT J LEPR MYCOBACT DIS OFF ORGAN INT LEPR ASSOC.* 1998;66:170–181.
- [214] GEBRE S, SAUNDERSON P, BYASS P. RELAPSES AFTER FIXED DURATION MULTIPLE DRUG THERAPY: THE AMFES COHORT. *LEPR REV.* 2000;71:325–331.
- [215] NORMAN G, JOSEPH G, RICHARD J. RELAPSES IN MULTIBACILLARY PATIENTS TREATED WITH MULTI-DRUG THERAPY UNTIL SMEAR NEGATIVITY: FINDINGS AFTER TWENTY YEARS. *INT J LEPR MYCOBACT DIS OFF ORGAN INT LEPR ASSOC.* 2004;72:1–7.
- [216] BOERRIGTER G, PÖNNIGHAUS JM, FINE PE, ET AL. FOUR-YEAR FOLLOW-UP RESULTS OF A WHO-RECOMMENDED MULTIPLE-DRUG REGIMEN IN PAUCIBACILLARY LEPROSY PATIENTS IN MALAWI. *INT J LEPR MYCOBACT DIS OFF ORGAN INT LEPR ASSOC.* 1991;59:255–261.
- [217] CHOPRA NK, AGARAWAL JS, PANDYA PG. A STUDY OF RELAPSE IN PAUCIBACILLARY LEPROSY IN A MULTIDRUG THERAPY PROJECT, BARODA DISTRICT, INDIA. *LEPR REV.* 1990;61:157–162.
- [218] JAMET P, JI B. RELAPSE AFTER LONG-TERM FOLLOW UP OF MULTIBACILLARY PATIENTS TREATED BY WHO MULTIDRUG REGIMEN. MARCHOUX CHEMOTHERAPY STUDY GROUP. *INT J LEPR MYCOBACT DIS OFF ORGAN INT LEPR ASSOC.* 1995;63:195–201.
- [219] GELBER RH, BALAGON VF, CELLONA RV. THE RELAPSE RATE IN MB LEPROSY PATIENTS TREATED WITH 2-YEARS OF WHO-MDT IS NOT LOW. *INT J LEPR MYCOBACT DIS OFF ORGAN INT LEPR ASSOC.* 2004;72:493–500.
- [220] RANQUE-FRANÇOIS, BRIGITTE. RECHERCHE DE FACTEURS GENETIQUES INFLUENÇANT DEUX PHENOTYPES LIES A LA LEPRE: LA REACTION GRANULOMATEUSE DE MITSUDA ET LES REACTIONS DE REVERSION. 2007. THESE DE DOCTORAT. PARIS 11.
- [221] LIENHARDT C, FINE PEM. TYPE 1 REACTION, NEURITIS AND DISABILITY IN LEPROSY. WHAT IS THE CURRENT EPIDEMIOLOGICAL SITUATION? *LEPR REV.* 1994;65:9–33.
- [222] VAN WB, KHAWAS IB, LUCAS SB. REACTIONS IN LEPROSY: AN EPIDEMIOLOGICAL STUDY OF 386 PATIENTS IN WEST NEPAL. *LEPR REV.* 1994;65:190–203.
- [223] SCHREUDER PA. THE OCCURRENCE OF REACTIONS AND IMPAIRMENTS IN LEPROSY : EXPERIENCE IN THE LEPROSY CONTROL PROGRAM OF THREE PROVINCES IN NORTHEASTERN THAILAND, 1987-1995 [CORRECTION OF 1978-1995]. II. REACTIONS. *INT J LEPR MYCOBACT* 662 159–69 1998.

- [224] BECX-BLEUMINK M, BERHE D. OCCURRENCE OF REACTIONS, THEIR DIAGNOSIS AND MANAGEMENT IN LEPROSY PATIENTS TREATED WITH MULTIDRUG THERAPY; EXPERIENCE IN THE LEPROSY CONTROL PROGRAM OF THE ALL AFRICA LEPROSY AND REHABILITATION TRAINING CENTER (ALERT) IN ETHIOPIA. INT J LEPR MYCOBACT DIS OFF ORGAN INT LEPR ASSOC. 1992;60:173–184.
- [225] LOCKWOOD DN, VINAYAKUMAR S, STANLEY JN, ET AL. CLINICAL FEATURES AND OUTCOME OF REVERSAL (TYPE 1) REACTIONS IN HYDERABAD, INDIA. INT J LEPR OTHER MYCOBACT DIS. 1993;61:8–15.
- [226] KUMAR B, DOGRA S, KAUR I. EPIDEMIOLOGICAL CHARACTERISTICS OF LEPROSY REACTIONS: 15 YEARS EXPERIENCE FROM NORTH INDIA. INT J LEPR OTHER MYCOBACT DIS. 2004;72:125.
- [227] ROSE P, WATERS MF. REVERSAL REACTIONS IN LEPROSY AND THEIR MANAGEMENT. LEPR REV. 1991;62:113–121.
- [228] SHARMA N, KORANNE RV, MENDIRATTA V, ET AL. A STUDY OF LEPROSY REACTIONS IN A TERTIARY HOSPITAL IN DELHI. J DERMATOL. 2004;31:898–903.
- [229] BRITTON WJ. THE MANAGEMENT OF LEPROSY REVERSAL REACTIONS. LEPR REV. 1998;69:225–234.
- [230] RAO PSSS, SUGAMARAN DST, RICHARD J, ET AL. MULTI-CENTRE, DOUBLE BLIND, RANDOMIZED TRIAL OF THREE STEROID REGIMENS IN THE TREATMENT OF TYPE-1 REACTIONS IN LEPROSY. LEPR REV. 2006;77:25–33.
- [231] ANDERSSON AK, CHADUVULA M, ATKINSON SE, ET AL. EFFECTS OF PREDNISOLONE TREATMENT ON CYTOKINE EXPRESSION IN PATIENTS WITH LEPROSY TYPE 1 REACTIONS. INFECT IMMUN. 2005;73:3725–3733.
- [232] CROSS H, NEWCOMBE L. AN INTENSIVE SELF-CARE TRAINING PROGRAMME REDUCES ADMISSIONS FOR THE TREATMENT OF PLANTAR ULCERS. LEPR REV. 2001;72:276–284.
- [233] KAZEN RO. MANAGEMENT OF PLANTAR ULCERS IN LEPROSY. LEPR REV. 1999;70:63–69.
- [234] MOET FJ, PAHAN D, OSKAM L, ET AL. EFFECTIVENESS OF SINGLE DOSE RIFAMPICIN IN PREVENTING LEPROSY IN CLOSE CONTACTS OF PATIENTS WITH NEWLY DIAGNOSED LEPROSY: CLUSTER RANDOMISED CONTROLLED TRIAL. BMJ. 2008;336:761–764.
- [235] LOCKWOOD DN, KRISHNAMURTHY P, KUMAR B, ET AL. SINGLE-DOSE RIFAMPICIN CHEMOPROPHYLAXIS PROTECTS THOSE WHO NEED IT LEAST AND IS NOT A COST-EFFECTIVE INTERVENTION. PLOS NEGL TROP DIS. 2018;12:E0006403.
- [236] DUTHIE MS, PENA MT, EBENEZER GJ, ET AL. LEPVAX, A DEFINED SUBUNIT VACCINE THAT PROVIDES EFFECTIVE PRE-EXPOSURE AND POST-EXPOSURE PROPHYLAXIS OF M. LEPRAE INFECTION. NPJ VACCINES. 2018;3:1–9.

- [237] VELDHUIJZEN N. LEPROSY RESEARCH INITIATIVE - LEPVAX: LEPROSY VACCINE TRIAL, PHASE 1B [INTERNET]. [CITED 2021 MAR 15]. AVAILABLE FROM: [HTTPS://LEPROSYRESEARCH.ORG/PROJECTS/TRANSMISSION/LEPVAX](https://leprosyresearch.org/projects/transmission/lepvax).
- [238] DUTHIE MS, FREVOL A, DAY T, ET AL. A PHASE 1 ANTIGEN DOSE ESCALATION TRIAL TO EVALUATE SAFETY, TOLERABILITY AND IMMUNOGENICITY OF THE LEPROSY VACCINE CANDIDATE LEPVAX (LEP-F1+ GLA-SE) IN HEALTHY ADULTS. *VACCINE*. 2020;38:1700–1707.
- [239] ORÉ NAEH, TEKAIA FR, TI AT, ET AL. THE DECAYING GENOME OF MYCOBACTERIUM LEPRAE. *LEPR REV*. 2001;72:387–398.
- [240] JOB CK. NASAL MUCOSA AND ABRADED SKIN ARE THE TWO ROUTES OF ENTRY OF MYCOBACTERIUM LEPRAE. *STAR*. 1990;49:1.
- [241] DE ABREU MM, ROSELINO AM, ENOKIHARA M, ET AL. MYCOBACTERIUM LEPRAE IS IDENTIFIED IN THE ORAL MUCOSA FROM PAUCIBACILLARY AND MULTIBACILLARY LEPROSY PATIENTS. *CLIN MICROBIOL INFECT*. 2014;20:59–64.
- [242] BECHELLI LM, GARBAJOSA PG, GYI MM, ET AL. SITE OF EARLY SKIN LESIONS IN CHILDREN WITH LEPROSY. *BULL WORLD HEALTH ORGAN*. 1973;48:107.
- [243] JOB CK, JAYAKUMAR J, KEARNEY M, ET AL. TRANSMISSION OF LEPROSY: A STUDY OF SKIN AND NASAL SECRETIONS OF HOUSEHOLD CONTACTS OF LEPROSY PATIENTS USING PCR. *AM J TROP MED HYG*. 2008;78:518–521.
- [244] SILVA CA, DANELISHVILI L, MCNAMARA M, ET AL. INTERACTION OF MYCOBACTERIUM LEPRAE WITH HUMAN AIRWAY EPITHELIAL CELLS: ADHERENCE, ENTRY, SURVIVAL, AND IDENTIFICATION OF POTENTIAL ADHESINS BY SURFACE PROTEOME ANALYSIS. *INFECT IMMUN*. 2013;81:2645–2659.
- [245] DAVEY TF, REES RJW. THE NASAL DISCHARGE IN LEPROSY: CLINICAL AND BACTERIOLOGICAL ASPECTS. *LEPR REV*. 1974;45:121–134.
- [246] REES RJW, MCDUGALL AC. AIRBORNE INFECTION WITH MYCOBACTERIUM LEPRAE IN MICE. *J MED MICROBIOL*. 1977;10:63–68.
- [247] JOB A. HISTOLOGICAL STUDIES IN PRIMARY NEURITIC LEPROSY: CHANGES IN THE NASAL MUCOSA. *LEPR REV*. 1998;69:358–366.
- [248] PATROCINIO LG, GOULART IMB, GOULART LR, ET AL. DETECTION OF MYCOBACTERIUM LEPRAE IN NASAL MUCOSA BIOPSIES BY THE POLYMERASE CHAIN REACTION. *FEMS IMMUNOL MED MICROBIOL*. 2005;44:311–316.
- [249] MELSOM R, HARBOE M, DUNCAN ME. IGA, IGM AND IGG ANTI-M. LEPRAE ANTIBODIES IN BABIES OF LEPROSY MOTHERS DURING THE FIRST 2 YEARS OF LIFE. *CLIN EXP IMMUNOL*. 1982;49:532.
- [250] GOULART IMB, ARAUJO S, BOTELHO FILHO A, ET AL. ASYMPTOMATIC LEPROSY INFECTION AMONG BLOOD DONORS MAY PREDICT DISEASE DEVELOPMENT AND

- SUGGESTS A POTENTIAL MODE OF TRANSMISSION. J CLIN MICROBIOL. 2015;53:3345–3348.
- [251] LUMPKIN III LR, COX GF, WOLF JR JE. LEPROSY IN FIVE ARMADILLO HANDLERS. J AM ACAD DERMATOL. 1983;9:899–903.
- [252] PEDLEY JC. THE PRESENCE OF M. LEPRAE IN HUMAN MILK. LEPR REV. 1967;38:239–242.
- [253] LASTORIA JC, ABREU MAMM DE. LEPROSY: REVIEW OF THE EPIDEMIOLOGICAL, CLINICAL, AND ETIOPATHOGENIC ASPECTS-PART 1. AN BRAS DERMATOL. 2014;89:205–218.
- [254] KLATSER PR, VAN BEERS S, MADJID B, ET AL. DETECTION OF MYCOBACTERIUM LEPRAE NASAL CARRIERS IN POPULATIONS FOR WHICH LEPROSY IS ENDEMIC. J CLIN MICROBIOL. 1993;31:2947–2951.
- [255] BEYENE D, ASEFFA A, HARBOE M, ET AL. NASAL CARRIAGE OF MYCOBACTERIUM LEPRAE DNA IN HEALTHY INDIVIDUALS IN LEGA ROBI VILLAGE, ETHIOPIA. EPIDEMIOLOGICAL INFECT. 2003;131:841–848.
- [256] SYSTEME DE L'HUMAN LEUKOCYTE ANTIGEN (HLA) - IMMUNOLOGIE; TROUBLES ALLERGIQUES [INTERNET]. ÉDITION PROF. MAN. MSD. [CITED 2021 MAR 6]. AVAILABLE FROM: [HTTPS://WWW.MSDMANUALS.COM/FR/PROFESSIONAL/IMMUNOLOGIE-TROUBLES-ALLERGIQUES/BIOLOGIE-DU-SYST%C3%A8ME-IMMUNITAIRE/SYST%C3%A8ME-DE-HUMAN-LEUKOCYTE-ANTIGEN](https://www.msmanuals.com/fr/professional/immunologie-troubles-allergiques/biologie-du-syst%C3%A8me-immunitaire/syst%C3%A8me-de-human-leukocyte-antigen).
- [257] HLA | BIORIGAMI [INTERNET]. [CITED 2021 MAR 6]. AVAILABLE FROM: [HTTPS://WWW.BIORIGAMI.COM/?TAG=HLA](https://www.biorigami.com/?tag=HLA).
- [258] SYSTEME HLA, PRESENTATION DE L'ANTIGENE ET REPOSE IMMUNITAIRE - PPT VIDEO ONLINE TELECHARGER [INTERNET]. [CITED 2021 MAR 6]. AVAILABLE FROM: [HTTPS://SLIDEPLAYER.FR/SLIDE/9488571/](https://slideplayer.fr/slide/9488571/).
- [259] FREHEL C, RASTOGI N. MYCOBACTERIUM LEPRAE SURFACE COMPONENTS INTERVENE IN THE EARLY PHAGOSOME-LYSOSOME FUSION INHIBITION EVENT. INFECT IMMUN. 1987;55:2916–2921.
- [260] MA Y, PEI Q, ZHANG L, ET AL. LIVE MYCOBACTERIUM LEPRAE INHIBITS AUTOPHAGY AND APOPTOSIS OF INFECTED MACROPHAGES AND PREVENTS ENGULFMENT OF HOST CELL BY PHAGOCYTES. AM J TRANSL RES. 2018;10:2929–2939.
- [261] LITTLE D, KHANOLKAR-YOUNG S, COULTHART A, ET AL. IMMUNOHISTOCHEMICAL ANALYSIS OF CELLULAR INFILTRATE AND GAMMA INTERFERON, INTERLEUKIN-12, AND INDUCIBLE NITRIC OXIDE SYNTHASE EXPRESSION IN LEPROSY TYPE 1 (REVERSAL) REACTIONS BEFORE AND DURING PREDNISOLONE TREATMENT. INFECT IMMUN. 2001;69:3413–3417.

- [262] RAMBUKKANA A, SALZER JL, YURCHENCO PD, ET AL. NEURAL TARGETING OF MYCOBACTERIUM LEPRAE MEDIATED BY THE G DOMAIN OF THE LAMININ-A2 CHAIN. *CELL*. 1997;88:811–821.
- [263] TAPINOS N, OHNISHI M, RAMBUKKANA A. ERBB2 RECEPTOR TYROSINE KINASE SIGNALING MEDIATES EARLY DEMYELINATION INDUCED BY LEPROSY BACILLI. *NAT MED*. 2006;12:961–966.
- [264] RAMBUKKANA A, ZANAZZI G, TAPINOS N, ET AL. CONTACT-DEPENDENT DEMYELINATION BY MYCOBACTERIUM LEPRAE IN THE ABSENCE OF IMMUNE CELLS. *SCIENCE*. 2002;296:927–931.
- [265] MADIGAN CA, CAMBIER CJ, KELLY-SCUMPIA KM, ET AL. A MACROPHAGE RESPONSE TO MYCOBACTERIUM LEPRAE PHENOLIC GLYCOLIPID INITIATES NERVE DAMAGE IN LEPROSY. *CELL*. 2017;170:973-985.E10.
- [266] SPIERINGS E, DE BOER T, ZULIANELLO L, ET AL. NOVEL MECHANISMS IN THE IMMUNOPATHOGENESIS OF LEPROSY NERVE DAMAGE: THE ROLE OF SCHWANN CELLS, T CELLS AND MYCOBACTERIUM LEPRAE. *IMMUNOL CELL BIOL*. 2000;78:349–355.
- [267] MASAKI T, QU J, CHOLEWA-WACLAW J, ET AL. REPROGRAMMING ADULT SCHWANN CELLS TO STEM CELL-LIKE CELLS BY LEPROSY BACILLI PROMOTES DISSEMINATION OF INFECTION. *CELL*. 2013;152:51–67.
- [268] SANTOS AR, SUFFYS PN, VANDERBORGH PR, ET AL. ROLE OF TUMOR NECROSIS FACTOR–A AND INTERLEUKIN-10 PROMOTER GENE POLYMORPHISMS IN LEPROSY. *J INFECT DIS*. 2002;186:1687–1691.
- [269] BLEHARSKI JR, LI H, MEINKEN C, ET AL. USE OF GENETIC PROFILING IN LEPROSY TO DISCRIMINATE CLINICAL FORMS OF THE DISEASE. *SCIENCE*. 2003;301:1527–1530.
- [270] SIELING PA, MODLIN RL. CYTOKINE PATTERNS AT THE SITE OF MYCOBACTERIAL INFECTION. *IMMUNOBIOLOGY*. 1994;191:378–387.
- [271] DE JONG R, JANSON AA, FABER WR, ET AL. IL-2 AND IL-12 ACT IN SYNERGY TO OVERCOME ANTIGEN-SPECIFIC T CELL UNRESPONSIVENESS IN MYCOBACTERIAL DISEASE. *J IMMUNOL*. 1997;159:786–793.
- [272] NATHAN CF, KAPLAN G, LEVIS WR, ET AL. LOCAL AND SYSTEMIC EFFECTS OF INTRADERMAL RECOMBINANT INTERFERON- γ IN PATIENTS WITH LEPROMATOUS LEPROSY. *N ENGL J MED*. 1986;315:6–15.
- [273] KAPLAN G, KIESSLING R, TEKLEMARIAM S, ET AL. THE RECONSTITUTION OF CELL-MEDIATED IMMUNITY IN THE CUTANEOUS LESIONS OF LEPROMATOUS LEPROSY BY RECOMBINANT INTERLEUKIN 2. *J EXP MED*. 1989;169:893–907.
- [274] QUARESMA JAS, AARÃO TLS, SOUSA JR, ET AL. T-HELPER 17 CYTOKINES EXPRESSION IN LEPROSY SKIN LESIONS. *BR J DERMATOL*. 2015;173:565–567.

- [275] SAINI C, RAMESH V, NATH I. CD4+ TH17 CELLS DISCRIMINATE CLINICAL TYPES AND CONSTITUTE A THIRD SUBSET OF NON TH1, NON TH2 T CELLS IN HUMAN LEPROSY. PLOS NEGL TROP DIS. 2013;7:E2338.
- [276] MURRAY RA, SIDDIQUI MR, MENDILLO M, ET AL. MYCOBACTERIUM LEPRAE INHIBITS DENDRITIC CELL ACTIVATION AND MATURATION. J IMMUNOL. 2007;178:338–344.
- [277] MODLIN RL. THE INNATE IMMUNE RESPONSE IN LEPROSY. CURR OPIN IMMUNOL. 2010;22:48–54.
- [278] SIELING PA, JULLIEN D, DAHLEM M, ET AL. CD1 EXPRESSION BY DENDRITIC CELLS IN HUMAN LEPROSY LESIONS: CORRELATION WITH EFFECTIVE HOST IMMUNITY. J IMMUNOL. 1999;162:1851–1858.
- [279] KRUTZIK SR, OCHOA MT, SIELING PA, ET AL. ACTIVATION AND REGULATION OF TOLL-LIKE RECEPTORS 2 AND 1 IN HUMAN LEPROSY. NAT MED. 2003;9:525–532.
- [280] SOILLEUX EJ, SARNO EN, HERNANDEZ MO, ET AL. DC-SIGN ASSOCIATION WITH THE TH2 ENVIRONMENT OF LEPROMATOUS LESIONS: CAUSE OR EFFECT? J PATHOL J PATHOL SOC G B IREL. 2006;209:182–189.
- [281] MARMOR MF. THE OPHTHALMIC TRIALS OF GHA HANSEN. SURV OPHTHALMOL. 2002;47:275–287.
- [282] PANDYA SS. ANTI-CONTAGIONISM IN LEPROSY, 1844-1897. INT J LEPR OTHER MYCOBACT DIS. 1998;66:374.
- [283] DANIELSSEN DC, BOECK W. OM SPEDALSKHED. C. GRÖNDAHL (PRINTERS); 1847.
- [284] ABRAHAM PS. II. ON THE CONTAGIOUSNESS OF LEPROSY. ANN SURG. 1886;4:324.
- [285] KIRKPATRICK J. ETIOLOGY OF LEPROSY. BR MED J. 1889;1:1491.
- [286] CAWS M, THWAITES G, DUNSTAN S, ET AL. THE INFLUENCE OF HOST AND BACTERIAL GENOTYPE ON THE DEVELOPMENT OF DISSEMINATED DISEASE WITH MYCOBACTERIUM TUBERCULOSIS. PLOS PATHOG. 2008;4.
- [287] SINGH P, COLE ST. MYCOBACTERIUM LEPRAE: GENES, PSEUDOGENES AND GENETIC DIVERSITY. FUTURE MICROBIOL. 2011;6:57–71.
- [288] RICHARDUS JH. RISK FACTORS FOR THE DEVELOPMENT OF CLINICAL LEPROSY AMONG CONTACTS, AND THEIR RELEVANCE FOR TARGETED INTERVENTIONS. LEPR REV. 2004;75:310–326.
- [289] SCOLLARD DM. LEPROSY RESEARCH DECLINES, BUT MOST OF THE BASIC QUESTIONS REMAIN UNANSWERED. INT J LEPR OTHER MYCOBACT DIS. 2005;73:25.
- [290] ALCAÏS A, ABEL L, CASANOVA J-L. HUMAN GENETICS OF INFECTIOUS DISEASES: BETWEEN PROOF OF PRINCIPLE AND PARADIGM. J CLIN INVEST. 2009;119:2506–2514.

- [291] GASCHIGNARD J, SCURR E, ALCAÏS A. LA LEPRE, PILIER DE LA GENETIQUE DES MALADIES INFECTIEUSES. *PATHOL BIOL.* 2013;61:120–128.
- [292] CASANOVA J-L, ABEL L. GENETIC DISSECTION OF IMMUNITY TO MYCOBACTERIA: THE HUMAN MODEL. *ANNU REV IMMUNOL.* 2002;20:581–620.
- [293] RANQUE B, ALTER A, SCHURR E, ET AL. LA LEPRE-UN PARADIGME POUR L'ETUDE DE LA PREDISPOSITION GENETIQUE AUX MALADIES INFECTIEUSES. *MEDECINE/SCIENCES.* 2008;24:491–498.
- [294] PLANT J, GLYNN AA. NATURAL RESISTANCE TO SALMONELLA INFECTION, DELAYED HYPERSENSITIVITY AND IR GENES IN DIFFERENT STRAINS OF MICE. *NATURE.* 1974;248:345–347.
- [295] PLANT J, GLYNN AA. LOCATING SALMONELLA RESISTANCE GENE ON MOUSE CHROMOSOME 1. *CLIN EXP IMMUNOL.* 1979;37:1.
- [296] BRADLEY DJ. REGULATION OF LEISHMANIA POPULATIONS WITHIN THE HOST. II. GENETIC CONTROL OF ACUTE SUSCEPTIBILITY OF MICE TO LEISHMANIA DONOVANI INFECTION. *CLIN EXP IMMUNOL.* 1977;30:130.
- [297] GROS P, SKAMENE E, FORGET A. GENETIC CONTROL OF NATURAL RESISTANCE TO MYCOBACTERIUM BOVIS (BCG) IN MICE. *J IMMUNOL.* 1981;127:2417–2421.
- [298] BRADLEY DJ, TAYLOR BA, BLACKWELL JM, ET AL. REGULATION OF LEISHMANIA POPULATIONS WITHIN THE HOST. III. MAPPING OF THE LOCUS CONTROLLING SUSCEPTIBILITY TO VISCERAL LEISHMANIASIS IN THE MOUSE. *CLIN EXP IMMUNOL.* 1979;37:7.
- [299] LISSNER CR, SWANSON RN, O'BRIEN AD. GENETIC CONTROL OF THE INNATE RESISTANCE OF MICE TO SALMONELLA TYPHIMURIUM: EXPRESSION OF THE ITY GENE IN PERITONEAL AND SPLENIC MACROPHAGES ISOLATED IN VITRO. *J IMMUNOL.* 1983;131:3006–3013.
- [300] CROCKER PR, BLACKWELL JM, BRADLEY DJ. EXPRESSION OF THE NATURAL RESISTANCE GENE LSH IN RESIDENT LIVER MACROPHAGES. *INFECT IMMUN.* 1984;43:1033–1040.
- [301] STACH JL, GROS P, FORGET A, ET AL. PHENOTYPIC EXPRESSION OF GENETICALLY-CONTROLLED NATURAL RESISTANCE TO MYCOBACTERIUM BOVIS (BCG). *J IMMUNOL.* 1984;132:888–892.
- [302] GROS P, MALO D. A REVERSE GENETICS APPROACH TO BCG/ITY/LSH GENE CLONING. *RES IMMUNOL.* 1989;140:774–777.
- [303] SCHURR E, SKAMENE E, FORGET A, ET AL. LINKAGE ANALYSIS OF THE BCG GENE ON MOUSE CHROMOSOME 1. IDENTIFICATION OF A TIGHTLY LINKED MARKER. *J IMMUNOL.* 1989;142:4507–4513.

- [304] MALO D, VIDAL S, LIEMAN JH, ET AL. PHYSICAL DELINEATION OF THE MINIMAL CHROMOSOMAL SEGMENT ENCOMPASSING THE MURINE HOST RESISTANCE LOCUS BCG. GENOMICS. 1993;17:667–675.
- [305] DUYK GM, KIM SW, MYERS RM, ET AL. EXON TRAPPING: A GENETIC SCREEN TO IDENTIFY CANDIDATE TRANSCRIBED SEQUENCES IN CLONED MAMMALIAN GENOMIC DNA. PROC NATL ACAD SCI. 1990;87:8995–8999.
- [306] VIDAL SM, MALO D, VOGAN K, ET AL. NATURAL RESISTANCE TO INFECTION WITH INTRACELLULAR PARASITES: ISOLATION OF A CANDIDATE FOR BCG. CELL. 1993;73:469–485.
- [307] CELLIER M, GOVONI G, VIDAL S, ET AL. HUMAN NATURAL RESISTANCE-ASSOCIATED MACROPHAGE PROTEIN: CDNA CLONING, CHROMOSOMAL MAPPING, GENOMIC ORGANIZATION, AND TISSUE-SPECIFIC EXPRESSION. J EXP MED. 1994;180:1741–1752.
- [308] CANONNE-HERGAUX F, GRUENHEID S, GOVONI G, ET AL. THE NRAMP1 PROTEIN AND ITS ROLE IN RESISTANCE TO INFECTION AND MACROPHAGE FUNCTION. PROC ASSOC AM PHYSICIANS. 1999;111:283–289.
- [309] CELLIER M, BELOUCHI A, GROS P. RESISTANCE TO INTRACELLULAR INFECTIONS: COMPARATIVE GENOMIC ANALYSIS OF NRAMP. TRENDS GENET. 1996;12:201–204.
- [310] HANSEN GA. ON THE ETIOLOGY OF LEPROSY. LONDRES, J. & A. CHURCHILL. 1875;
- [311] SERJEANTSON S, WILSON SR, KEATS BJB. THE GENETICS OF LEPROSY. ANN HUM BIOL. 1979;6:375–393.
- [312] BEIGUELMAN B. LEPROSY AND GENETICS: A REVIEW. REV BRAS GENET. 1983;109–172.
- [313] SHIELDS ED, RUSSELL DA, PERICAK-VANCE MA. GENETIC EPIDEMIOLOGY OF THE SUSCEPTIBILITY TO LEPROSY. J CLIN INVEST. 1987;79:1139–1143.
- [314] SPICKETT SG. GENETICS AND THE EPIDEMIOLOGY OF LEPROSY. LEPR REV. 1962;1:73.
- [315] MOHAMED A, RAMANUJAM K. LEPROSY IN TWINS. INT J LEPR. 1966;34:405–407.
- [316] CHAKRAVARTTI MR. A TWIN STUDY ON LEPROSY. TOP HUM GENET. 1973;1–123.
- [317] BEIGUELMAN B, DA SILVA E, DALL'AGLIO FF. LEPROSY AND SEX. REV PAUL MED. 1968;72:120.
- [318] PENROSE L. OUTLINE OF HUMAN GENETICS. 1963;
- [319] SMITH DG. THE GENETIC HYPOTHESIS FOR SUSCEPTIBILITY TO LEPROMATOUS LEPROSY. HUM GENET. 1979;50:163–177.
- [320] DEMENAIIS F, GOULET V, FEINGOLD N, ET AL. GENETIC STUDY OF LEPROSY IN A CARIBBEAN ISLAND: DESIRADE. GENET CONTROL HOST RESIST INFECT MALIG N Y ALAN R LISS. 1985;319–324.

- [321] HAILE RWC, ISELIUS L, FINE PEM, ET AL. SEGREGATION AND LINKAGE ANALYSES OF 72 LEPROSY PEDIGREES. HUM HERED. 1985;35:43–52.
- [322] ABEL L, DEMENAIIS F. DETECTION OF MAJOR GENES FOR SUSCEPTIBILITY TO LEPROSY AND ITS SUBTYPES IN A CARIBBEAN ISLAND: DESIRADE ISLAND. AM J HUM GENET. 1988;42:256.
- [323] WAGENER DK, SCHAUF V, NELSON KE, ET AL. SEGREGATION ANALYSIS OF LEPROSY IN FAMILIES OF NORTHERN THAILAND. GENET EPIDEMIOL. 1988;5:95–105.
- [324] FEITOSA MF, BORECKI I, KRIEGER H, ET AL. THE GENETIC EPIDEMIOLOGY OF LEPROSY IN A BRAZILIAN POPULATION. AM J HUM GENET. 1995;56:1179.
- [325] ABEL L, LAP VD, OBERTI J, ET AL. COMPLEX SEGREGATION ANALYSIS OF LEPROSY IN SOUTHERN VIETNAM. GENET EPIDEMIOL. 1995;12:63–82.
- [326] TEARE MD, BARRETT JH. GENETIC LINKAGE STUDIES. THE LANCET. 2005;366:1036–1044.
- [327] DE V, DE VRIES R, RFMLA F, ET AL. HLA-LINKED GENETIC CONTROL OF HOST RESPONSE TO MYCOBACTERIUM LEPRAE. 1976;
- [328] FINE PE, WOLF E, PRITCHARD J, ET AL. HLA-LINKED GENES AND LEPROSY: A FAMILY STUDY IN KARIGIRI, SOUTH INDIA. J INFECT DIS. 1979;140:152–161.
- [329] DE VRIES RR, MEHRA NK, VAIDYA MC, ET AL. HLA-LINKED CONTROL OF SUSCEPTIBILITY TO TUBERCULOID LEPROSY AND ASSOCIATION WITH HLA–DR TYPES. TISSUE ANTIGENS. 1980;16:294–304.
- [330] VAN EDEN W, GONZALEZ NM, DE VRIES RR, ET AL. HLA-LINKED CONTROL OF PREDISPOSITION TO LEPROMATOUS LEPROSY. J INFECT DIS. 1985;151:9–14.
- [331] DESSOUKEY MW, EL-SHIEMY S, SALLAM T. HLA AND LEPROSY: SEGREGATION AND LINKAGE STUDY. INT J DERMATOL. 1996;35:257–264.
- [332] OTTENHOFF TM, DE VRIES RP. HLA CLASS II IMMUNE RESPONSE AND SUPPRESSION GENES IN LEPROSY. INT J LEPR OTHER MYCOBACT DIS. 1987;55:521–534.
- [333] ABEL L, SANCHEZ FO, OBERTI J, ET AL. SUSCEPTIBILITY TO LEPROSY IS LINKED TO THE HUMAN NRAMP1 GENE. J INFECT DIS. 1998;177:133–145.
- [334] SHAW M-A, ATKINSON S, DOCKRELL H, ET AL. AN RFLP MAP FOR 2Q33-Q37 FROM MULTICASE MYCOBACTERIAL AND LEISHMANIAL DISEASE FAMILIES: NO EVIDENCE FOR AN LSH/ITY/BCG GENE HOMOLOGUE INFLUENCING SUSCEPTIBILITY TO LEPROSY. ANN HUM GENET. 1993;57:251–271.
- [335] LEVEE G, LIU J, GICQUEL B, ET AL. GENETIC CONTROL OF SUSCEPTIBILITY TO LEPROSY IN FRENCH POLYNESIA; NO EVIDENCE FOR LINKAGE WITH MARKERS ON TELOMERIC HUMAN CHROMOSOME 2. INT J LEPR MYCOBACT DIS OFF ORGAN INT LEPR ASSOC. 1994;62:499–511.

- [336] VEJBAESYA S, MAHAISAVARIYA P, LUANGTRAKOOL P, ET AL. TNF ALPHA AND NRAMP1 POLYMORPHISMS IN LEPROSY. *J-MED ASSOC THAIL.* 2007;90:1188.
- [337] JAMIESON SE, MILLER EN, BLACK GF, ET AL. EVIDENCE FOR A CLUSTER OF GENES ON CHROMOSOME 17Q11–Q21 CONTROLLING SUSCEPTIBILITY TO TUBERCULOSIS AND LEPROSY IN BRAZILIANS. *GENES IMMUN.* 2004;5:46–57.
- [338] MILLER EN, JAMIESON SE, JOBERTY C, ET AL. GENOME-WIDE SCANS FOR LEPROSY AND TUBERCULOSIS SUSCEPTIBILITY GENES IN BRAZILIANS. *GENES IMMUN.* 2004;5:63–67.
- [339] SIDDIQUI MR, MEISNER S, TOSH K, ET AL. A MAJOR SUSCEPTIBILITY LOCUS FOR LEPROSY IN INDIA MAPS TO CHROMOSOME 10P13. *NAT GENET.* 2001;27:439–441.
- [340] MIRA MT, ALCAIS A, DI PIETRANTONIO T, ET AL. SEGREGATION OF HLA/TNF REGION IS LINKED TO LEPROSY CLINICAL SPECTRUM IN FAMILIES DISPLAYING MIXED LEPROSY SUBTYPES. *GENES IMMUN.* 2003;4:67–73.
- [341] WALLACE C, FITNESS J, HENNIG B, ET AL. LINKAGE ANALYSIS OF SUSCEPTIBILITY TO LEPROSY TYPE USING AN IBD REGRESSION METHOD. *GENES IMMUN.* 2004;5:221–225.
- [342] MIRA MT, ALCAÍS A, VAN THUC N, ET AL. CHROMOSOME 6Q25 IS LINKED TO SUSCEPTIBILITY TO LEPROSY IN A VIETNAMESE POPULATION. *NAT GENET.* 2003;33:412–415.
- [343] TOSH K, MEISNER S, SIDDIQUI MR, ET AL. A REGION OF CHROMOSOME 20 IS LINKED TO LEPROSY SUSCEPTIBILITY IN A SOUTH INDIAN POPULATION. *J INFECT DIS.* 2002;186:1190–1193.
- [344] MALHOTRA D, DARVISHI K, LOHRA M, ET AL. ASSOCIATION STUDY OF MAJOR RISK SINGLE NUCLEOTIDE POLYMORPHISMS IN THE COMMON REGULATORY REGION OF PARK2 AND PACRG GENES WITH LEPROSY IN AN INDIAN POPULATION. *EUR J HUM GENET.* 2006;14:438–442.
- [345] MALHOTRA D, DARVISHI K, LOHRA M, ET AL. ERRATUM: ASSOCIATION STUDY OF MAJOR RISK SINGLE NUCLEOTIDE POLYMORPHISMS IN THE COMMON REGULATORY REGION OF PARK2 AND PACRG GENES WITH LEPROSY IN AN INDIAN POPULATION. *EUR J HUM GENET.* 2007;15:908–908.
- [346] CHUANG T-H, ULEVITCH RJ. TRIAD3A, AN E3 UBIQUITIN-PROTEIN LIGASE REGULATING TOLL-LIKE RECEPTORS. *NAT IMMUNOL.* 2004;5:495–502.
- [347] KNIGHT JC, KEATING BJ, KWIATKOWSKI DP. ALLELE-SPECIFIC REPRESSION OF LYMPHOTOXIN-A BY ACTIVATED B CELL FACTOR-1. *NAT GENET.* 2004;36:394–399.
- [348] WARE CF. NETWORK COMMUNICATIONS: LYMPHOTOXINS, LIGHT, AND TNF. *ANNU REV IMMUNOL.* 2005;23:787–819.
- [349] CORDELL HJ, CLAYTON DG. GENETIC ASSOCIATION STUDIES. *THE LANCET.* 2005;366:1121–1131.

- [350] WANG WY, BARRATT BJ, CLAYTON DG, ET AL. GENOME-WIDE ASSOCIATION STUDIES: THEORETICAL AND PRACTICAL CONCERNS. *NAT REV GENET*. 2005;6:109–118.
- [351] MISCH EA, BERRINGTON WR, VARY JC, ET AL. LEPROSY AND THE HUMAN GENOME. *MICROBIOL MOL BIOL REV*. 2010;74:589–620.
- [352] SAUER ME, SALOMÃO H, RAMOS GB, ET AL. GENETICS OF LEPROSY: EXPECTED AND UNEXPECTED DEVELOPMENTS AND PERSPECTIVES. *CLIN DERMATOL*. 2015;33:99–107.
- [353] SERGIO DO SACRAMENTO W, MAZINI PS, FRANCESCHI DAS, ET AL. FREQUENCIES OF MICA ALLELES IN PATIENTS FROM SOUTHERN BRAZIL WITH MULTIBACILLARY AND PAUCIBACILLARY LEPROSY. *INT J IMMUNOGENET*. 2012;39:210–215.
- [354] LUCENA-SILVA N, TEIXEIRA MAG, RAMOS A DE L, ET AL. THE+ 3187A/G HLA-G POLYMORPHIC SITE IS ASSOCIATED WITH POLAR FORMS AND REACTIVE REACTION IN LEPROSY. *MOL GENET GENOMIC MED*. 2013;1:123–130.
- [355] ROY S, MCGUIRE W, MASCIE-TAYLOR CGN, ET AL. TUMOR NECROSIS FACTOR PROMOTER POLYMORPHISM AND SUSCEPTIBILITY TO LEPROMATOUS LEPROSY. *J INFECT DIS*. 1997;176:530–532.
- [356] FITNESS J, FLOYD S, WARNDORFF DK, ET AL. LARGE-SCALE CANDIDATE GENE STUDY OF TUBERCULOSIS SUSCEPTIBILITY IN THE KARONGA DISTRICT OF NORTHERN MALAWI. *AM J TROP MED HYG*. 2004;71:341–349.
- [357] GREINER J, WEBER FJ, MAUFF G, ET AL. GENETIC POLYMORPHISMS OF PROPERDIN FACTOR B (BF), THE SECOND COMPONENT (C2), AND THE FOURTH COMPONENT (C4) OF COMPLEMENT IN LEPROSY PATIENTS AND HEALTHY CONTROLS FROM THAILAND. *IMMUNOBIOLOGY*. 1980;158:134–138.
- [358] REYNARD MP, TURNER D, JUNQUEIRA-KIPNIS AP, ET AL. ALLELE FREQUENCIES FOR AN INTERFERON- γ MICROSATELLITE IN A POPULATION OF BRAZILIAN LEPROSY PATIENTS. *EUR J IMMUNOGENET*. 2003;30:149–151.
- [359] PIEMONTI L, MONTI P, SIRONI M, ET AL. VITAMIN D3 AFFECTS DIFFERENTIATION, MATURATION, AND FUNCTION OF HUMAN MONOCYTE-DERIVED DENDRITIC CELLS. *J IMMUNOL*. 2000;164:4443–4451.
- [360] ROY S, FRODSHAM A, SAHA B, ET AL. ASSOCIATION OF VITAMIN D RECEPTOR GENOTYPE WITH LEPROSY TYPE. *J INFECT DIS*. 1999;179:187–191.
- [361] BERRINGTON WR, MACDONALD M, KHADGE S, ET AL. COMMON POLYMORPHISMS IN THE NOD2 GENE REGION ARE ASSOCIATED WITH LEPROSY AND ITS REACTIVE STATES. *J INFECT DIS*. 2010;201:1422–1435.
- [362] BOCHUD P-Y, SINSIMER D, ADEREM A, ET AL. POLYMORPHISMS IN TOLL-LIKE RECEPTOR 4 (TLR4) ARE ASSOCIATED WITH PROTECTION AGAINST LEPROSY. *EUR J CLIN MICROBIOL INFECT DIS*. 2009;28:1055–1065.

- [363] KANG T-J, CHAE G-T. DETECTION OF TOLL-LIKE RECEPTOR 2 (TLR2) MUTATION IN THE LEPROMATOUS LEPROSY PATIENTS. *FEMS IMMUNOL MED MICROBIOL.* 2001;31:53–58.
- [364] BOCHUD P-Y, HAWN TR, SIDDIQUI MR, ET AL. TOLL-LIKE RECEPTOR 2 (TLR2) POLYMORPHISMS ARE ASSOCIATED WITH REVERSAL REACTION IN LEPROSY. *J INFECT DIS.* 2008;197:253–261.
- [365] MORAES MO, PACHECO AG, SCHONKEREN JJM, ET AL. INTERLEUKIN-10 PROMOTER SINGLE-NUCLEOTIDE POLYMORPHISMS AS MARKERS FOR DISEASE SUSCEPTIBILITY AND DISEASE SEVERITY IN LEPROSY. *GENES IMMUN.* 2004;5:592–595.
- [366] SANTOS AR, SUFFYS PN, VANDERBORGHT PR, ET AL. ROLE OF TUMOR NECROSIS FACTOR- α AND INTERLEUKIN-10 PROMOTER GENE POLYMORPHISMS IN LEPROSY. *J INFECT DIS.* 2002;186:1687–1691.
- [367] RAJALINGAM R, MEHRA NK, SINGAL DP. POLYMORPHISM IN HEAT-SHOCK PROTEIN 70-1 (HSP70-1) GENE PROMOTER REGION AND SUSCEPTIBILITY TO TUBERCULOID LEPROSY AND PULMONARY TUBERCULOSIS IN ASIAN INDIANS. 2000;
- [368] ALI S, SRIVASTAVA AK, CHOPRA R, ET AL. IL12B SNPS AND COPY NUMBER VARIATION IN IL23R GENE ASSOCIATED WITH SUSCEPTIBILITY TO LEPROSY. *J MED GENET.* 2013;50:34–42.
- [369] DE MESSIAS-REASON IJ, BOLDT AB, BRAGA ACM, ET AL. THE ASSOCIATION BETWEEN MANNAN-BINDING LECTIN GENE POLYMORPHISM AND CLINICAL LEPROSY: NEW INSIGHT INTO AN OLD PARADIGM. *J INFECT DIS.* 2007;196:1379–1385.
- [370] SAPKOTA BR, MACDONALD M, BERRINGTON WR, ET AL. ASSOCIATION OF TNF, MBL, AND VDR POLYMORPHISMS WITH LEPROSY PHENOTYPES. *HUM IMMUNOL.* 2010;71:992–998.
- [371] MEISNER SJ, MUCKLOW S, WARNER G, ET AL. ASSOCIATION OF NRAMP1 POLYMORPHISM WITH LEPROSY TYPE BUT NOT SUSCEPTIBILITY TO LEPROSY PER SE IN WEST AFRICANS. *AM J TROP MED HYG.* 2001;65:733–735.
- [372] FRANCESCHI DSA, MAZINI PS, RUDNICK CCC, ET AL. ASSOCIATION BETWEEN KILLER-CELL IMMUNOGLOBULIN-LIKE RECEPTOR GENOTYPES AND LEPROSY IN BRAZIL. *TISSUE ANTIGENS.* 2008;72:478–482.
- [373] WIBAWA T, SOEBONO H, MATSUO M. ASSOCIATION OF A MISSENSE MUTATION OF THE LAMININ A2 GENE WITH TUBERCULOID TYPE OF LEPROSY IN INDONESIAN PATIENTS. *TROP MED INT HEALTH.* 2002;7:631–636.
- [374] DE OCA EP-M, VELARDE-FELIX JS, RIOS-TOSTADO JJ, ET AL. SNP 668C (– 44) ALTERS A NF-KB1 PUTATIVE BINDING SITE IN NON-CODING STRAND OF HUMAN B-DEFENSIN 1 (DEFB1) AND IS ASSOCIATED WITH LEPROMATOUS LEPROSY. *INFECT GENET EVOL.* 2009;9:617–625.

- [375] ZHANG F, LIU H, CHEN S, ET AL. IDENTIFICATION OF TWO NEW LOCI AT IL23R AND RAB32 THAT INFLUENCE SUSCEPTIBILITY TO LEPROSY. *NAT GENET.* 2011;43:1247–1251.
- [376] MARCINEK P, JHA AN, SHINDE V, ET AL. LRRK2 AND RIPK2 VARIANTS IN THE NOD 2-MEDIATED SIGNALING PATHWAY ARE ASSOCIATED WITH SUSCEPTIBILITY TO MYCOBACTERIUM LEPRAE IN INDIAN POPULATIONS. *PLOS ONE.* 2013;8:E73103.
- [377] LIU H, IRWANTO A, YU G, ET AL. DISCOVERY OF SIX NEW SUSCEPTIBILITY LOCI AND ANALYSIS OF PLEIOTROPIC EFFECTS IN LEPROSY. *NAT GENET.* 2015;47:267–271.
- [378] LIU H, WANG Z, LI Y, ET AL. GENOME-WIDE ANALYSIS OF PROTEIN-CODING VARIANTS IN LEPROSY. *J INVEST DERMATOL.* 2017;137:2544–2551.
- [379] LIU H, IRWANTO A, TIAN H, ET AL. IDENTIFICATION OF IL18RAP/IL18R1 AND IL12B AS LEPROSY RISK GENES DEMONSTRATES SHARED PATHOGENESIS BETWEEN INFLAMMATION AND INFECTIOUS DISEASES. *AM J HUM GENET.* 2012;91:935–941.
- [380] WONG SH, GOCHHAIT S, MALHOTRA D, ET AL. LEPROSY AND THE ADAPTATION OF HUMAN TOLL-LIKE RECEPTOR 1. *PLOS PATHOG.* 2010;6:E1000979.
- [381] FAVA VM, SCHURR E. THE COMPLEXITY OF THE HOST GENETIC CONTRIBUTION TO THE HUMAN RESPONSE TO MYCOBACTERIUM LEPRAE. *INT TEXTB LEPR.* 2016;1–33.
- [382] LIU H, BAO F, IRWANTO A, ET AL. AN ASSOCIATION STUDY OF TOLL AND CARD WITH LEPROSY SUSCEPTIBILITY IN CHINESE POPULATION. *HUM MOL GENET.* 2013;22:4430–4437.
- [383] WANG Z, SUN Y, FU X, ET AL. A LARGE-SCALE GENOME-WIDE ASSOCIATION AND META-ANALYSIS IDENTIFIED FOUR NOVEL SUSCEPTIBILITY LOCI FOR LEPROSY. *NAT COMMUN.* 2016;7:1–8.
- [384] WANG Z, MI Z, WANG H, ET AL. DISCOVERY OF 4 EXONIC AND 1 INTERGENIC NOVEL SUSCEPTIBILITY LOCI FOR LEPROSY. *CLIN GENET.* 2018;94:259–263.
- [385] WANG D, FAN Y, MALHI M, ET AL. MISSENSE VARIANTS IN HIF1A AND LACC1 CONTRIBUTE TO LEPROSY RISK IN HAN CHINESE. *AM J HUM GENET.* 2018;102:794–805.
- [386] LE SALAIRE MOYEN EN FRANCE EN 2021 [INTERNET]. *BUS. COOL.* 2021 [CITED 2021 FEB 13]. AVAILABLE FROM: [HTTPS://BUSINESS-COOL.COM/VIE-PRO/SE-FAIRE-RECRUTER/SALAIRE-MOYEN-CADRE-FRANCE/](https://business-cool.com/vie-pro/se-faire-recruter/salaire-moyen-cadre-france/).
- [387] VIET NAM • FICHE PAYS • POPULATIONDATA.NET [INTERNET]. *POPULATIONDATA.NET.* [CITED 2021 FEB 13]. AVAILABLE FROM: [HTTPS://WWW.POPULATIONDATA.NET/PAYS/VIET-NAM/](https://www.populationdata.net/pays/viet-nam/).
- [388] VIET NAM [INTERNET]. *WIKIPEDIA.* 2021 [CITED 2021 FEB 13]. AVAILABLE FROM: [HTTPS://FR.WIKIPEDIA.ORG/W/INDEX.PHP?TITLE=VI%C3%AAT_NAM&OLDID=179693895](https://fr.wikipedia.org/w/index.php?title=VI%C3%AAT_NAM&oldid=179693895).

- [389] ORGANIZATION WH. TRENDS IN THE EPIDEMIOLOGY OF LEPROSY—VIET NAM, 1983-2006. WKLY EPIDEMIOL REC RELEVÉ ÉPIDÉMIOLOGIQUE HEBD. 2008;83:217–224.
- [390] ORGANIZATION WH. GLOBAL LEPROSY UPDATE, 2013; REDUCING DISEASE BURDEN. WKLY EPIDEMIOL REC RELEVÉ ÉPIDÉMIOLOGIQUE HEBD. 2014;89:389–400.
- [391] KHANG TH, THAO NM. EPIDEMIOLOGY OF LEPROSY IN VIETNAM AND THE EFFECTIVENESS OF MULTIDRUG THERAPY (MDT) IN THE MANAGEMENT OF THE DISEASE. CURR TOP NEGLECTED TROP DIS. INTECHOPEN; 2019.
- [392] RANQUE B, ABEL L, ALCAÏS A. LA LEPRE, UNE MALADIE ELIMINÉE... OU NEGLIGÉE? ANTIBIOTIQUES. 2007;9:99–114.
- [393] RANQUE B, ALCAÏS A, THUC NV, ET AL. A RECESSIVE MAJOR GENE CONTROLS THE MITSUDA REACTION IN A REGION ENDEMIC FOR LEPROSY. J INFECT DIS. 2005;192:1475–1482.
- [394] RANQUE B, VAN THUC N, VU HT, ET AL. AGE IS AN IMPORTANT RISK FACTOR FOR ONSET AND SEQUELAE OF REVERSAL REACTIONS IN VIETNAMESE PATIENTS WITH LEPROSY. CLIN INFECT DIS. 2007;44:33–40.
- [395] [HTTP://MOLMED.MEDSCI.UU.SE/SNP+SEQ+TECHNOLOGY+PLATFORM/GENOTYPING/SNP_METHODS_AND_REFERENCES/IL_LUMINA_INFINIUM_ASSAY/](http://molmed.medsci.uu.se/snp+seq+technology+platform/genotyping/snp_methods_and_references/il_lumina_infinium_assay/).
- [396] [HTTP://HAPMAP.NCBI.NLM.NIH.GOV/INDEX.HTML.EN](http://hapmap.ncbi.nlm.nih.gov/index.html.en).
- [397] [WWW.INTERNATIONALGENOME.ORG/](http://www.internationalgenome.org/).
- [398] NOVEMBRE J, JOHNSON T, BRYC K, ET AL. GENES MIRROR GEOGRAPHY WITHIN EUROPE. NATURE. 2008;456:98–101.
- [399] [HTTP://SUPPORT.ILLUMINA.COM/ARRAY/ARRAY_KITS/HUMAN660W-QUAD_DNA_ANALYSIS_KIT.HTML](http://support.illumina.com/array/array_kits/human660w-quad_dna_analysis_kit.html).
- [400] ANDERSON CA, PETERSSON FH, CLARKE GM, ET AL. DATA QUALITY CONTROL IN GENETIC CASE-CONTROL ASSOCIATION STUDIES. NAT PROTOC. 2010;5:1564–1573.
- [401] TURNER S, ARMSTRONG LL, BRADFORD Y, ET AL. QUALITY CONTROL PROCEDURES FOR GENOME-WIDE ASSOCIATION STUDIES. CURR PROTOC HUM GENET. 2011;68:1.19. 1-1.19. 18.
- [402] TEAM RC. R: A LANGUAGE AND ENVIRONMENT FOR STATISTICAL COMPUTING. 2013;
- [403] PURCELL S, NEALE B, TODD-BROWN K, ET AL. PLINK: A TOOL SET FOR WHOLE-GENOME ASSOCIATION AND POPULATION-BASED LINKAGE ANALYSES. AM J HUM GENET. 2007;81:559–575.
- [404] PURCELL S, CHANG C. PLINK 1.9. AVAILABLE FROM: [WWW.COGENOMICS.ORG/PLINK/1.9/](http://www.cog-genomics.org/plink/1.9/).

- [405] MARCHINI J, HOWIE B. GENOTYPE IMPUTATION FOR GENOME-WIDE ASSOCIATION STUDIES. *NAT REV GENET.* 2010;11:499–511.
- [406] CONSORTIUM 1000 GENOMES PROJECT. AN INTEGRATED MAP OF GENETIC VARIATION FROM 1,092 HUMAN GENOMES. *NATURE.* 2012;491:56.
- [407] HOWIE BN, DONNELLY P, MARCHINI J. A FLEXIBLE AND ACCURATE GENOTYPE IMPUTATION METHOD FOR THE NEXT GENERATION OF GENOME-WIDE ASSOCIATION STUDIES. *PLOS GENET.* 2009;5:E1000529.
- [408] HOWIE B, FUCHSBERGER C, STEPHENS M, ET AL. FAST AND ACCURATE GENOTYPE IMPUTATION IN GENOME-WIDE ASSOCIATION STUDIES THROUGH PRE-PHASING. *NAT GENET.* 2012;44:955–959.
- [409] DELANEAU O, ZAGURY J-F, MARCHINI J. IMPROVED WHOLE-CHROMOSOME PHASING FOR DISEASE AND POPULATION GENETIC STUDIES. *NAT METHODS.* 2013;10:5–6.
- [410] DELANEAU O, MARCHINI J, ZAGURY J-F. A LINEAR COMPLEXITY PHASING METHOD FOR THOUSANDS OF GENOMES. *NAT METHODS.* 2012;9:179–181.
- [411] O’CONNELL J, GURDASANI D, DELANEAU O, ET AL. A GENERAL APPROACH FOR HAPLOTYPE PHASING ACROSS THE FULL SPECTRUM OF RELATEDNESS. *PLOS GENET.* 2014;10:E1004234.
- [412] MARCHINI J, HOWIE B, MYERS S, ET AL. A NEW MULTIPOINT METHOD FOR GENOME-WIDE ASSOCIATION STUDIES BY IMPUTATION OF GENOTYPES. *NAT GENET.* 2007;39:906–913.
- [413] DUBY C, ROBIN S. ANALYSE EN COMPOSANTES PRINCIPALES. *INST NATL AGRON PARIS-GRIGNON.* 2006;80.
- [414] PATTERSON N, PRICE AL, REICH D. POPULATION STRUCTURE AND EIGENANALYSIS. *PLOS GENET.* 2006;2:E190.
- [415] PRICE AL, PATTERSON NJ, PLENGE RM, ET AL. PRINCIPAL COMPONENTS ANALYSIS CORRECTS FOR STRATIFICATION IN GENOME-WIDE ASSOCIATION STUDIES. *NAT GENET.* 2006;38:904–909.
- [416] INTARAPANICH A, SHAW PJ, ASSAWAMAKIN A, ET AL. ITERATIVE PRUNING PCA IMPROVES RESOLUTION OF HIGHLY STRUCTURED POPULATIONS. *BMC BIOINFORMATICS.* 2009;10:1–17.
- [417] DEVLIN B, RISCH N. A COMPARISON OF LINKAGE DISEQUILIBRIUM MEASURES FOR FINE-SCALE MAPPING. *GENOMICS.* 1995;29:311–322.
- [418] PRITCHARD JK, PRZEWORSKI M. LINKAGE DISEQUILIBRIUM IN HUMANS: MODELS AND DATA. *AM J HUM GENET.* 2001;69:1–14.
- [419] BARRETT JC, FRY B, MALLER J, ET AL. HAPLOVIEW: ANALYSIS AND VISUALIZATION OF LD AND HAPLOTYPE MAPS. *BIOINFORMATICS.* 2005;21:263–265.

- [420] LANDER ES, SCHORK NJ. GENETIC DISSECTION OF COMPLEX TRAITS. SCIENCE. 1994;265:2037–2048.
- [421] KHOURY MJ, BEATY TH, COHEN BH. FUNDAMENTALS OF GENETIC EPIDEMIOLOGY OXFORD UNIVERSITY PRESS. NEW YORKGOOGLE SCH. 1993;
- [422] 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.
- [423] WILKIE AO. THE MOLECULAR BASIS OF GENETIC DOMINANCE. J MED GENET. 1994;31:89–98.
- [424] ROSENDAAL FR, KOSTER T, VANDENBROUCKE JP, ET AL. HIGH RISK OF THROMBOSIS IN PATIENTS HOMOZYGOUS FOR FACTOR V LEIDEN (ACTIVATED PROTEIN C RESISTANCE)[SEE COMMENTS]. 1995;
- [425] CARRINGTON M, NELSON GW, MARTIN MP, ET AL. HLA AND HIV-1: HETEROZYGOTE ADVANTAGE AND B* 35-CW* 04 DISADVANTAGE. SCIENCE. 1999;283:1748–1752.
- [426] CATANO G, AGAN BK, KULKARNI H, ET AL. INDEPENDENT EFFECTS OF GENETIC VARIATIONS IN MANNOSE-BINDING LECTIN INFLUENCE THE COURSE OF HIV DISEASE: THE ADVANTAGE OF HETEROZYGOSITY FOR CODING MUTATIONS. J INFECT DIS. 2008;198:72–80.
- [427] HRABER P, KUIKEN C, YUSIM K. EVIDENCE FOR HUMAN LEUKOCYTE ANTIGEN HETEROZYGOTE ADVANTAGE AGAINST HEPATITIS C VIRUS INFECTION. HEPATOLOGY. 2007;46:1713–1721.
- [428] THURSZ MR, THOMAS HC, GREENWOOD BM, ET AL. HETEROZYGOTE ADVANTAGE FOR HLA CLASS-II TYPE IN HEPATITIS B VIRUS INFECTION. NAT GENET. 1997;17:11–12.
- [429] HELLEMANN D, LARSSON A, MADSEN HO, ET AL. HETEROZYGOSITY OF MANNOSE-BINDING LECTIN (MBL2) GENOTYPES PREDICTS ADVANTAGE (HETEROSIS) IN RELATION TO FATAL OUTCOME IN INTENSIVE CARE PATIENTS. HUM MOL GENET. 2007;16:3071–3080.
- [430] PENN DJ, DAMJANOVICH K, POTTS WK. MHC HETEROZYGOSITY CONFERS A SELECTIVE ADVANTAGE AGAINST MULTIPLE-STRAIN INFECTIONS. PROC NATL ACAD SCI. 2002;99:11260–11264.
- [431] SELLIS D, CALLAHAN BJ, PETROV DA, ET AL. HETEROZYGOTE ADVANTAGE AS A NATURAL CONSEQUENCE OF ADAPTATION IN DIPLOIDS. PROC NATL ACAD SCI. 2011;108:20666–20671.
- [432] MINELLI C, THOMPSON JR, ABRAMS KR, ET AL. THE QUALITY OF META-ANALYSES OF GENETIC ASSOCIATION STUDIES: A REVIEW WITH RECOMMENDATIONS. AM J EPIDEMIOL. 2009;170:1333–1343.

- [433] WELTER D, MACARTHUR J, MORALES J, ET AL. THE NHGRI GWAS CATALOG, A CURATED RESOURCE OF SNP-TRAIT ASSOCIATIONS. *NUCLEIC ACIDS RES.* 2014;42:D1001–D1006.
- [434] SASIENI PD. FROM GENOTYPES TO GENES: DOUBLING THE SAMPLE SIZE. *BIOMETRICS.* 1997;1253–1261.
- [435] MCCARTHY MI, ABECASIS GR, CARDON LR, ET AL. GENOME-WIDE ASSOCIATION STUDIES FOR COMPLEX TRAITS: CONSENSUS, UNCERTAINTY AND CHALLENGES. *NAT REV GENET.* 2008;9:356–369.
- [436] LETTRE G, LANGE C, HIRSCHHORN JN. GENETIC MODEL TESTING AND STATISTICAL POWER IN POPULATION-BASED ASSOCIATION STUDIES OF QUANTITATIVE TRAITS. *GENET EPIDEMIOL OFF PUBL INT GENET EPIDEMIOL SOC.* 2007;31:358–362.
- [437] FREIDLIN B, ZHENG G, LI Z, ET AL. TREND TESTS FOR CASE-CONTROL STUDIES OF GENETIC MARKERS: POWER, SAMPLE SIZE AND ROBUSTNESS. *HUM HERED.* 2002;53:146–152.
- [438] CONSORTIUM WTCC. GENOME-WIDE ASSOCIATION STUDY OF 14,000 CASES OF SEVEN COMMON DISEASES AND 3,000 SHARED CONTROLS. *NATURE.* 2007;447:661.
- [439] EWENS WJ, SPIELMAN RS. THE TRANSMISSION/DISEQUILIBRIUM TEST: HISTORY, SUBDIVISION, AND ADMIXTURE. *AM J HUM GENET.* 1995;57:455.
- [440] SPIELMAN RS, EWENS WJ. THE TDT AND OTHER FAMILY-BASED TESTS FOR LINKAGE DISEQUILIBRIUM AND ASSOCIATION. *AM J HUM GENET.* 1996;59:983.
- [441] MARTIN ER, KAPLAN NL, WEIR BS. TESTS FOR LINKAGE AND ASSOCIATION IN NUCLEAR FAMILIES. *AM J HUM GENET.* 1997;61:439–448.
- [442] SPIELMAN RS, EWENS WJ. A SIBSHIP TEST FOR LINKAGE IN THE PRESENCE OF ASSOCIATION: THE SIB TRANSMISSION/DISEQUILIBRIUM TEST. *AM J HUM GENET.* 1998;62:450–458.
- [443] ALLISON DB. TRANSMISSION-DISEQUILIBRIUM TESTS FOR QUANTITATIVE TRAITS. *AM J HUM GENET.* 1997;60:676.
- [444] RABINOWITZ D. A TRANSMISSION DISEQUILIBRIUM TEST FOR QUANTITATIVE TRAIT LOCI. *HUM HERED.* 1997;47:342–350.
- [445] SUN FZ, FLANDERS WD, YANG QH, ET AL. TRANSMISSION/DISEQUILIBRIUM TESTS FOR QUANTITATIVE TRAITS. *ANN HUM GENET.* 2000;64:555–565.
- [446] FULKER DW, CHERNY SS, SHAM PC, ET AL. COMBINED LINKAGE AND ASSOCIATION SIB-PAIR ANALYSIS FOR QUANTITATIVE TRAITS. *AM J HUM GENET.* 1999;64:259–267.
- [447] ABECASIS GR, CARDON LR, COOKSON WOC. A GENERAL TEST OF ASSOCIATION FOR QUANTITATIVE TRAITS IN NUCLEAR FAMILIES. *AM J HUM GENET.* 2000;66:279–292.

- [448] HORVATH S, XU X, LAIRD NM. THE FAMILY BASED ASSOCIATION TEST METHOD: STRATEGIES FOR STUDYING GENERAL GENOTYPE–PHENOTYPE ASSOCIATIONS. *EUR J HUM GENET*. 2001;9:301–306.
- [449] LAIRD NM, HORVATH S, XU X. IMPLEMENTING A UNIFIED APPROACH TO FAMILY-BASED TESTS OF ASSOCIATION. *GENET EPIDEMIOL OFF PUBL INT GENET EPIDEMIOL SOC*. 2000;19:S36–S42.
- [450] LAIRD NM, LANGE C. FAMILY-BASED METHODS FOR LINKAGE AND ASSOCIATION ANALYSIS. *ADV GENET*. 2008;60:219–252.
- [451] LAKE SL, BLACKER D, LAIRD NM. FAMILY-BASED TESTS OF ASSOCIATION IN THE PRESENCE OF LINKAGE. *AM J HUM GENET*. 2000;67:1515–1525.
- [452] LANGE C, SILVERMAN EK, XU X, ET AL. A MULTIVARIATE FAMILY-BASED ASSOCIATION TEST USING GENERALIZED ESTIMATING EQUATIONS: FBAT-GEE. *BIostatistics*. 2003;4:195–206.
- [453] RABINOWITZ D, LAIRD N. A UNIFIED APPROACH TO ADJUSTING ASSOCIATION TESTS FOR POPULATION ADMIXTURE WITH ARBITRARY PEDIGREE STRUCTURE AND ARBITRARY MISSING MARKER INFORMATION. *HUM HERED*. 2000;50:211–223.
- [454] LAIRD NM, LANGE C. FAMILY-BASED DESIGNS IN THE AGE OF LARGE-SCALE GENE-ASSOCIATION STUDIES. *NAT REV GENET*. 2006;7:385–394.
- [455] ZHENG J, LI Y, ABECASIS GR, ET AL. A COMPARISON OF APPROACHES TO ACCOUNT FOR UNCERTAINTY IN ANALYSIS OF IMPUTED GENOTYPES. *GENET EPIDEMIOL*. 2011;35:102–110.
- [456] COBAT A, ABEL L, ALCAÏS A, ET AL. A GENERAL EFFICIENT AND FLEXIBLE APPROACH FOR GENOME-WIDE ASSOCIATION ANALYSES OF IMPUTED GENOTYPES IN FAMILY-BASED DESIGNS. *GENET EPIDEMIOL*. 2014;38:560–571.
- [457] DEBETTE S. COMMENT LIRE UNE ETUDE D’ASSOCIATION GENETIQUE PANGENOMIQUE (GWAS)? *SANG THROMB VAISS*. 2012;24:240–247.
- [458] TURNER SD. QQMAN: AN R PACKAGE FOR VISUALIZING GWAS RESULTS USING QQ AND MANHATTAN PLOTS. *BIORXIV*. 2014;005165.
- [459] JANNOT A-S, EHRET G, PERNEGER T. $P < 5 \times 10^{-8}$ HAS EMERGED AS A STANDARD OF STATISTICAL SIGNIFICANCE FOR GENOME-WIDE ASSOCIATION STUDIES. *J CLIN EPIDEMIOL*. 2015;68:460–465.
- [460] HUNTER DJ, KRAFT P. DRINKING FROM THE FIRE HOSE—STATISTICAL ISSUES IN GENOMEWIDE ASSOCIATION STUDIES. *N ENGL J MED*. 2007;357:436–439.
- [461] DEPRISTO MA, BANKS E, POPLIN R, ET AL. A FRAMEWORK FOR VARIATION DISCOVERY AND GENOTYPING USING NEXT-GENERATION DNA SEQUENCING DATA. *NAT GENET*. 2011;43:491.

- [462] SAS INSTITUTE, CARY, NORTH CAROLINA, USA.
- [463] CLARK AG. THE ROLE OF HAPLOTYPES IN CANDIDATE GENE STUDIES. *GENET EPIDEMIOL OFF PUBL INT GENET EPIDEMIOL SOC.* 2004;27:321–333.
- [464] TREGOUËT D-A, KÖNIG IR, ERDMANN J, ET AL. GENOME-WIDE HAPLOTYPE ASSOCIATION STUDY IDENTIFIES THE SLC22A3-LPAL2-LPA GENE CLUSTER AS A RISK LOCUS FOR CORONARY ARTERY DISEASE. *NAT GENET.* 2009;41:283–285.
- [465] NIU T. ALGORITHMS FOR INFERRING HAPLOTYPES. *GENET EPIDEMIOL OFF PUBL INT GENET EPIDEMIOL SOC.* 2004;27:334–347.
- [466] KARCZEWSKI KJ, FRANCIOLI LC, TIAO G, ET AL. VARIATION ACROSS 141,456 HUMAN EXOMES AND GENOMES REVEALS THE SPECTRUM OF LOSS-OF-FUNCTION INTOLERANCE ACROSS HUMAN PROTEIN-CODING GENES. *BIORXIV.* 2019;531210.
- [467] MACHIELA MJ, CHANOCK SJ. LDLINK: A WEB-BASED APPLICATION FOR EXPLORING POPULATION-SPECIFIC HAPLOTYPE STRUCTURE AND LINKING CORRELATED ALLELES OF POSSIBLE FUNCTIONAL VARIANTS. *BIOINFORMATICS.* 2015;31:3555–3557.
- [468] PRUIM RJ, WELCH RP, SANNA S, ET AL. LOCUSZOOM: REGIONAL VISUALIZATION OF GENOME-WIDE ASSOCIATION SCAN RESULTS. *BIOINFORMATICS.* 2010;26:2336–2337.
- [469] CAMBRI G, MIRA MT. GENETIC SUSCEPTIBILITY TO LEPROSY—FROM CLASSIC IMMUNE-RELATED CANDIDATE GENES TO HYPOTHESIS-FREE, WHOLE GENOME APPROACHES. *FRONT IMMUNOL.* 2018;9:1674.
- [470] JARDULI LR, SELL AM, REIS PG, ET AL. ROLE OF HLA, KIR, MICA, AND CYTOKINES GENES IN LEPROSY. *BIOMED RES INT.* 2013;2013.
- [471] KRAUSE-KYORA B, NUTSUA M, BOEHME L, ET AL. ANCIENT DNA STUDY REVEALS HLA SUSCEPTIBILITY LOCUS FOR LEPROSY IN MEDIEVAL EUROPEANS. *NAT COMMUN.* 2018;9:1–11.
- [472] ZHANG X, CHENG Y, ZHANG Q, ET AL. META-ANALYSIS IDENTIFIES MAJOR HISTOCOMPATIBILITY COMPLEX LOCI IN OR NEAR HLA-DRB1, HLA-DQA1, HLA-C AS ASSOCIATED WITH LEPROSY IN CHINESE HAN POPULATION. 2019.
- [473] [HTTPS:// GTEXPORTAL.ORG/HOME/](https://gtexportal.org/home/).
- [474] MATZARAKI V, KUMAR V, WIJMENGA C, ET AL. THE MHC LOCUS AND GENETIC SUSCEPTIBILITY TO AUTOIMMUNE AND INFECTIOUS DISEASES. *GENOME BIOL.* 2017;18:1–21.
- [475] PARKS T, ELLIOTT K, LAMAGNI T, ET AL. ELEVATED RISK OF INVASIVE GROUP A STREPTOCOCCAL DISEASE AND HOST GENETIC VARIATION IN THE HUMAN LEUCOCYTE ANTIGEN LOCUS. *GENES IMMUN.* 2020;21:63–70.

- [476] FAVA VM, MANRY J, COBAT A, ET AL. A GENOME WIDE ASSOCIATION STUDY IDENTIFIES A LNCRNA AS RISK FACTOR FOR PATHOLOGICAL INFLAMMATORY RESPONSES IN LEPROSY. PLOS GENET. 2017;13:E1006637.
- [477] BARREIRO LB, QUINTANA-MURCI L. FROM EVOLUTIONARY GENETICS TO HUMAN IMMUNOLOGY: HOW SELECTION SHAPES HOST DEFENCE GENES. NAT REV GENET. 2010;11:17–30.
- [478] MANRY J, VINCENT QB, JOHNSON C, ET AL. GENOME-WIDE ASSOCIATION STUDY OF BURULI ULCER IN RURAL BENIN HIGHLIGHTS ROLE OF TWO LNCRNAs AND THE AUTOPHAGY PATHWAY. COMMUN BIOL. 2020;3:1–10.
- [479] ALTER A, HUONG NT, SINGH M, ET AL. HUMAN LEUKOCYTE ANTIGEN CLASS I REGION SINGLE-NUCLEOTIDE POLYMORPHISMS ARE ASSOCIATED WITH LEPROSY SUSCEPTIBILITY IN VIETNAM AND INDIA. J INFECT DIS. 2011;203:1274–1281.
- [480] LUO Y, NA Z, SLAVOFF SA. P-BODIES: COMPOSITION, PROPERTIES, AND FUNCTIONS. BIOCHEMISTRY. 2018;57:2424–2431.
- [481] PARKER R, SHETH U. P BODIES AND THE CONTROL OF MRNA TRANSLATION AND DEGRADATION. MOL CELL. 2007;25:635–646.
- [482] YU X, LI B, JANG G-J, ET AL. ORCHESTRATION OF PROCESSING BODY DYNAMICS AND MRNA DECAY IN ARABIDOPSIS IMMUNITY. CELL REP. 2019;28:2194-2205. E6.
- [483] RADZISHEUSKAYA A, CHIA GLB, DOS SANTOS RL, ET AL. A DEFINED OCT4 LEVEL GOVERNS CELL STATE TRANSITIONS OF PLURIPOTENCY ENTRY AND DIFFERENTIATION INTO ALL EMBRYONIC LINEAGES. NAT CELL BIOL. 2013;15:579–590.
- [484] DALLMANN-SAUER M, FAVA VM, GZARA C, ET AL. THE COMPLEX PATTERN OF GENETIC ASSOCIATIONS OF LEPROSY WITH HLA CLASS I AND CLASS II ALLELES CAN BE REDUCED TO FOUR AMINO ACID POSITIONS. PLOS PATHOG. 2020;16:E1008818.
- [485] WONG SH, HILL AV, VANNBERG FO, ET AL. GENOMEWIDE ASSOCIATION STUDY OF LEPROSY. N ENGL J MED. 2010;362:1446.
- [486] XIONG J-H, MAO C, SHA X-W, ET AL. ASSOCIATION BETWEEN GENETIC VARIANTS IN NOD 2, C13ORF31, AND CCDC 122 GENES AND LEPROSY AMONG THE CHINESE YI POPULATION. INT J DERMATOL. 2016;55:65–69.
- [487] BARRETT JC, HANSOUL S, NICOLAE DL, ET AL. GENOME-WIDE ASSOCIATION DEFINES MORE THAN 30 DISTINCT SUSCEPTIBILITY LOCI FOR CROHN'S DISEASE. NAT GENET. 2008;40:955–962.
- [488] FRANKE A, MCGOVERN DP, BARRETT JC, ET AL. GENOME-WIDE META-ANALYSIS INCREASES TO 71 THE NUMBER OF CONFIRMED CROHN'S DISEASE SUSCEPTIBILITY LOCI. NAT GENET. 2010;42:1118–1125.

- [489] JOSTINS L, RIPKE S, WEERSMA RK, ET AL. HOST–MICROBE INTERACTIONS HAVE SHAPED THE GENETIC ARCHITECTURE OF INFLAMMATORY BOWEL DISEASE. *NATURE*. 2012;491:119–124.
- [490] LIU JZ, VAN SOMMEREN S, HUANG H, ET AL. ASSOCIATION ANALYSES IDENTIFY 38 SUSCEPTIBILITY LOCI FOR INFLAMMATORY BOWEL DISEASE AND HIGHLIGHT SHARED GENETIC RISK ACROSS POPULATIONS. *NAT GENET*. 2015;47:979–986.
- [491] KIRCHER M, WITTEN DM, JAIN P, ET AL. A GENERAL FRAMEWORK FOR ESTIMATING THE RELATIVE PATHOGENICITY OF HUMAN GENETIC VARIANTS. *NAT GENET*. 2014;46:310–315.
- [492] RENTZSCH P, WITTEN D, COOPER GM, ET AL. CADD: PREDICTING THE DELETERIOUSNESS OF VARIANTS THROUGHOUT THE HUMAN GENOME. *NUCLEIC ACIDS RES*. 2019;47:D886–D894.
- [493] CADER MZ, BOROVIK K, ZHANG Q, ET AL. C13ORF31 (FAMIN) IS A CENTRAL REGULATOR OF IMMUNOMETABOLIC FUNCTION. *NAT IMMUNOL*. 2016;17:1046.
- [494] LAHIRI A, HEDL M, YAN J, ET AL. HUMAN LACC1 INCREASES INNATE RECEPTOR-INDUCED RESPONSES AND A LACC1 DISEASE-RISK VARIANT MODULATES THESE OUTCOMES. *NAT COMMUN*. 2017;8:1–13.
- [495] ANDERSON CA, BOUCHER G, LEES CW, ET AL. META-ANALYSIS IDENTIFIES 29 ADDITIONAL ULCERATIVE COLITIS RISK LOCI, INCREASING THE NUMBER OF CONFIRMED ASSOCIATIONS TO 47. *NAT GENET*. 2011;43:246–252.
- [496] RAYCHAUDHURI S, PLENGE RM, ROSSIN EJ, ET AL. IDENTIFYING RELATIONSHIPS AMONG GENOMIC DISEASE REGIONS: PREDICTING GENES AT PATHOGENIC SNP ASSOCIATIONS AND RARE DELETIONS. *PLOS GENET*. 2009;5:E1000534.
- [497] BOISSON-DUPOUIS S, RAMIREZ-ALEJO N, LI Z, ET AL. TUBERCULOSIS AND IMPAIRED IL-23-DEPENDENT IFN-GAMMA IMMUNITY IN HUMANS HOMOZYGOUS FOR A COMMON TYK2 MISSENSE VARIANT. *SCI IMMUNOL*. 2018.
- [498] MARTINEZ-BARRICARTE R, MARKLE JG, MA CS, ET AL. HUMAN IFN- γ IMMUNITY TO MYCOBACTERIA IS GOVERNED BY BOTH IL-12 AND IL-23. *SCI IMMUNOL*. 2018;3.
- [499] BUSTAMANTE J, BOISSON-DUPOUIS S, ABEL L, ET AL. MENDELIAN SUSCEPTIBILITY TO MYCOBACTERIAL DISEASE: GENETIC, IMMUNOLOGICAL, AND CLINICAL FEATURES OF INBORN ERRORS OF IFN- γ IMMUNITY. *SEMIN IMMUNOL*. ELSEVIER; 2014. P. 454–470.
- [500] CORTES A, HADLER J, POINTON JP, ET AL. IDENTIFICATION OF MULTIPLE RISK VARIANTS FOR ANKYLOSING SPONDYLITIS THROUGH HIGH-DENSITY GENOTYPING OF IMMUNE-RELATED LOCI. *NAT GENET*. 2013;45:730.
- [501] TERAO C, YOSHIFUJI H, KIMURA A, ET AL. TWO SUSCEPTIBILITY LOCI TO TAKAYASU ARTERITIS REVEAL A SYNERGISTIC ROLE OF THE IL12B AND HLA-B REGIONS IN A JAPANESE POPULATION. *AM J HUM GENET*. 2013;93:289–297.

- [502] NAKAJIMA T, YOSHIFUJI H, SHIMIZU M, ET AL. A NOVEL SUSCEPTIBILITY LOCUS IN THE IL12B REGION IS ASSOCIATED WITH THE PATHOPHYSIOLOGY OF TAKAYASU ARTERITIS THROUGH IL-12P40 AND IL-12P70 PRODUCTION. *ARTHRITIS RES THER.* 2017;19:1–8.
- [503] VINCENT Q, ALCAÏS A, ALTER A, ET AL. QUANTIFYING GENOMIC IMPRINTING IN THE PRESENCE OF LINKAGE. *BIOMETRICS.* 2006;62:1071–1080.
- [504] ANTONIOU AC, EASTON DF. POLYGENIC INHERITANCE OF BREAST CANCER: IMPLICATIONS FOR DESIGN OF ASSOCIATION STUDIES. *GENET EPIDEMIOL OFF PUBL INT GENET EPIDEMIOL SOC.* 2003;25:190–202.
- [505] CLARKE GM, CARTER KW, PALMER LJ, ET AL. FINE MAPPING VERSUS REPLICATION IN WHOLE-GENOME ASSOCIATION STUDIES. *AM J HUM GENET.* 2007;81:995–1005.
- [506] WWW.CDC.GOV/NCHS/ABOUT/MAJOR/NHCS/NHCS_QUICKSTATS.HTM.
- [507] DEVLIN B, ROEDER K. GENOMIC CONTROL FOR ASSOCIATION STUDIES. *BIOMETRICS.* 1999;55:997–1004.
- [508] BACANU S-A, DEVLIN B, ROEDER K. THE POWER OF GENOMIC CONTROL. *AM J HUM GENET.* 2000;66:1933–1944.
- [509] CONSORTIUM IH. A SECOND GENERATION HUMAN HAPLOTYPE MAP OF OVER 3.1 MILLION SNPS. *NATURE.* 2007;449:851.
- [510] ZONDERVAN KT, CARDON LR. DESIGNING CANDIDATE GENE AND GENOME-WIDE CASE–CONTROL ASSOCIATION STUDIES. *NAT PROTOC.* 2007;2:2492.
- [511] HERBERT A, GERRY NP, MCQUEEN MB, ET AL. A COMMON GENETIC VARIANT IS ASSOCIATED WITH ADULT AND CHILDHOOD OBESITY. *SCIENCE.* 2006;312:279–283.
- [512] VAN STEEN K, MCQUEEN MB, HERBERT A, ET AL. GENOMIC SCREENING AND REPLICATION USING THE SAME DATA SET IN FAMILY-BASED ASSOCIATION TESTING. *NAT GENET.* 2005;37:683–691.
- [513] WISE AL, GYI L, MANOLIO TA. EXCLUSION: TOWARD INTEGRATING THE X CHROMOSOME IN GENOME-WIDE ASSOCIATION ANALYSES. *AM J HUM GENET.* 2013;92:643–647.
- [514] KÖNIG IR, LOLEY C, ERDMANN J, ET AL. HOW TO INCLUDE CHROMOSOME X IN YOUR GENOME-WIDE ASSOCIATION STUDY. *GENET EPIDEMIOL.* 2014;38:97–103.
- [515] EICHLER EE, FLINT J, GIBSON G, ET AL. MISSING HERITABILITY AND STRATEGIES FOR FINDING THE UNDERLYING CAUSES OF COMPLEX DISEASE. *NAT REV GENET.* 2010;11:446–450.
- [516] CHOI M, SCHOLL UI, JI W, ET AL. GENETIC DIAGNOSIS BY WHOLE EXOME CAPTURE AND MASSIVELY PARALLEL DNA SEQUENCING. *PROC NATL ACAD SCI.* 2009;106:19096–19101.

- [517] LINDBLAD-TOH K, GARBER M, ZUK O, ET AL. A HIGH-RESOLUTION MAP OF HUMAN EVOLUTIONARY CONSTRAINT USING 29 MAMMALS. *NATURE*. 2011;478:476–482.
- [518] LYNCH VJ, LECLERC RD, MAY G, ET AL. TRANSPOSON-MEDIATED REWIRING OF GENE REGULATORY NETWORKS CONTRIBUTED TO THE EVOLUTION OF PREGNANCY IN MAMMALS. *NAT GENET*. 2011;43:1154–1159.
- [519] ADAMS LB, PENA MT, SHARMA R, ET AL. INSIGHTS FROM ANIMAL MODELS ON THE IMMUNOGENETICS OF LEPROSY: A REVIEW. *MEM INST OSWALDO CRUZ*. 2012;107:197–208.
- [520] INFANTE-RIVARD C, MIREA L, BULL SB. COMBINING CASE-CONTROL AND CASE-TRIO DATA FROM THE SAME POPULATION IN GENETIC ASSOCIATION ANALYSES: OVERVIEW OF APPROACHES AND ILLUSTRATION WITH A CANDIDATE GENE STUDY. *AM J EPIDEMIOL*. 2009;170:657–664.
- [521] DALMASSO C, BROËT P. PROCEDURES D’ESTIMATION DU FALSE DISCOVERY RATE BASEES SUR LA DISTRIBUTION DES DEGRES DE SIGNIFICATION. *J SOCIETE FR STAT*. 2005;146:63–75.
- [522] DALMASSO C, BROËT P, MOREAU T. MULTIPLE COMPARISON PROCEDURES: PRINCIPLES, LIMITS. APPLICATIONS TO MICROARRAY PHENOTYPE-GENOTYPE ANALYSIS. *REV EPIDEMIOL SANTE PUBLIQUE*. 2004;52:523–537.
- [523] BENJAMINI Y, DRAI D, ELMER G, ET AL. CONTROLLING THE FALSE DISCOVERY RATE IN BEHAVIOR GENETICS RESEARCH. *BEHAV BRAIN RES*. 2001;125:279–284.
- [524] BENJAMINI Y, YEKUTIELI D. THE CONTROL OF THE FALSE DISCOVERY RATE IN MULTIPLE TESTING UNDER DEPENDENCY. *ANN STAT*. 2001;1165–1188.
- [525] BENJAMINI Y, HOCHBERG Y. CONTROLLING THE FALSE DISCOVERY RATE: A PRACTICAL AND POWERFUL APPROACH TO MULTIPLE TESTING. *J R STAT SOC SER B METHODOL*. 1995;57:289–300.
- [526] BERNDT SI, GUSTAFSSON S, MÄGI R, ET AL. GENOME-WIDE META-ANALYSIS IDENTIFIES 11 NEW LOCI FOR ANTHROPOMETRIC TRAITS AND PROVIDES INSIGHTS INTO GENETIC ARCHITECTURE. *NAT GENET*. 2013;45:501–512.
- [527] BUNIELLO A, MACARTHUR JAL, CEREZO M, ET AL. THE NHGRI-EBI GWAS CATALOG OF PUBLISHED GENOME-WIDE ASSOCIATION STUDIES, TARGETED ARRAYS AND SUMMARY STATISTICS 2019. *NUCLEIC ACIDS RES*. 2019;47:D1005–D1012.
- [528] VISSCHER PM, BROWN MA, MCCARTHY MI, ET AL. FIVE YEARS OF GWAS DISCOVERY. *AM J HUM GENET*. 2012;90:7–24.
- [529] VISSCHER PM, WRAY NR, ZHANG Q, ET AL. 10 YEARS OF GWAS DISCOVERY: BIOLOGY, FUNCTION, AND TRANSLATION. *AM J HUM GENET*. 2017;101:5–22.
- [530] DO CB, HINDS DA, FRANCKE U, ET AL. COMPARISON OF FAMILY HISTORY AND SNPS FOR PREDICTING RISK OF COMPLEX DISEASE. *PLOS GENET*. 2012;8:E1002973.

- [531] JOSTINS L, BARRETT JC. GENETIC RISK PREDICTION IN COMPLEX DISEASE. HUM MOL GENET. 2011;20:R182–R188.
- [532] EVANS DM, VISSCHER PM, WRAY NR. HARNESSING THE INFORMATION CONTAINED WITHIN GENOME-WIDE ASSOCIATION STUDIES TO IMPROVE INDIVIDUAL PREDICTION OF COMPLEX DISEASE RISK. HUM MOL GENET. 2009;18:3525–3531.
- [533] VASSY JL, MEIGS JB. IS GENETIC TESTING USEFUL TO PREDICT TYPE 2 DIABETES? BEST PRACT RES CLIN ENDOCRINOL METAB. 2012;26:189–201.
- [534] HIRSCHHORN JN, GAJDOS ZK. GENOME-WIDE ASSOCIATION STUDIES: RESULTS FROM THE FIRST FEW YEARS AND POTENTIAL IMPLICATIONS FOR CLINICAL MEDICINE. ANNU REV MED. 2011;62:11–24.
- [535] MANOLIO TA. BRINGING GENOME-WIDE ASSOCIATION FINDINGS INTO CLINICAL USE. NAT REV GENET. 2013;14:549–558.
- [536] MUSUNURU K, STRONG A, FRANK-KAMENETSKY M, ET AL. FROM NONCODING VARIANT TO PHENOTYPE VIA SORT1 AT THE 1P13 CHOLESTEROL LOCUS. NATURE. 2010;466:714–719.
- [537] EDWARDS SL, BEESLEY J, FRENCH JD, ET AL. BEYOND GWASS: ILLUMINATING THE DARK ROAD FROM ASSOCIATION TO FUNCTION. AM J HUM GENET. 2013;93:779–797.
- [538] ROTHMAN KJ. CAUSES. AM J EPIDEMIOL. 1976;104:587–592.
- [539] PACIS A, NEDELEC Y, BARREIRO LB. WHEN GENETICS MEETS EPIGENETICS: DECIPHERING THE MECHANISMS CONTROLLING INTER-INDIVIDUAL VARIATION IN IMMUNE RESPONSES TO INFECTION. CURR OPIN IMMUNOL. 2014;29:119–126.
- [540] HA N-T, FREYTAG S, BICKEBOELLER H. COVERAGE AND EFFICIENCY IN CURRENT SNP CHIPS. EUR J HUM GENET. 2014;22:1124–1130.
- [541] SHALLCROSS LJ, DAVIES DSC. ANTIBIOTIC OVERUSE: A KEY DRIVER OF ANTIMICROBIAL RESISTANCE. BRITISH JOURNAL OF GENERAL PRACTICE; 2014.
- [542] RESISTANCE A. TACKLING A CRISIS FOR THE HEALTH AND WEALTH OF NATIONS. REV ANTIMICROB RESIST. 2014;

Annexes

ANNEXE 1 : Résultats d'association des 33 signaux hors-HLA associés à la lèpre, et précédemment publiés, lors de l'étude primaire familiale

SNP	CHR	Position	Gène	Études précédemment publiées				Notre GWAS au Vietnam		
				m/M	RA	P-val	Référence	RA	Info	P
rs3762318	1	67597119	IL23R	G/A	A	3,27.10 ⁻¹¹	[375]	A	1	4,41.10 ⁻⁰³
rs76418789	1	67648596	IL23R	A/G	A	1,03.10 ⁻¹⁰	[378]	A	0,557	4,60.10 ⁻⁰¹
rs2735591	1	85744472	BCL10	G/A	A	1,03.10 ⁻⁰⁹	[382]	G	1	0,87
rs146466242	1	152275298	FLG	A/T	A	3,39.10 ⁻¹²	[378]	T	0,933	6,51.10 ⁻⁰¹
rs2221593	1	212873431	BATF3	T/C	T	3,09.10 ⁻⁰⁸	[377]	T	1	1,94.10 ⁻⁰⁴
rs2058660	2	103054449	IL18RAP/IL18R1	A/G	A	4,57.10 ⁻¹⁹	[379]	A	0,999	5,66.10 ⁻⁰³
rs6807915	3	12313846	SYN2	C/T	T	1,94.10 ⁻⁰⁸	[383]	C	0,984	3,71.10 ⁻⁰¹
rs145562243	3	48719549	NCKIPSD	T/C	T	1,71.10 ⁻⁰⁹	[378]	-	-	-
rs5743618	4	38798648	TLR1	C/A	A	1,70.10 ⁻⁰⁹	[380]	C	0,67	9,85.10 ⁻⁰¹
rs73058713	5	20374213	Intergénique	A/C	A	9,54.10 ⁻⁰⁹	[377]	C	0,984	2,14.10 ⁻⁰¹
rs6871626	5	158826792	IL12B	A/C	C	3,95.10 ⁻¹⁸	[379]	C	0,778	5,12.10 ⁻⁰⁴
rs2275606	6	146918950	RAB32	A/G	A	3,94.10 ⁻¹⁴	[375]	A	1	1,99.10 ⁻⁰¹
rs4720118	7	33469241	BBS9	T/C	T	3,85.10 ⁻¹⁰	[383]	T	0,998	1,85.10 ⁻⁰¹
rs55894533	8	11749242	CTSB	C/A	C	5,07.10 ⁻¹¹	[383]	A	0,964	3,53.10 ⁻⁰¹
rs13259978	8	17396415	SLC7A2	C/G	C	1,74.10 ⁻⁰⁸	[384]	G	0,907	8,96.10 ⁻⁰¹
rs160451	8	90662941	RIPK2	T/C	C	7,45.10 ⁻¹²	[377]	C	0,999	8,33.10 ⁻⁰¹
rs42490	8	90778513	RIPK2	G/A	G	1,38.10 ⁻¹⁶	[166]	G	1	3,15.10 ⁻⁰²
rs10100465	8	118626279	M.10D30	A/G	G	2,85.10 ⁻¹¹	[383]	G	1	1,39.10 ⁻⁰¹
rs6478108	9	117558703	TNFSF15	G/A	A	3,39.10 ⁻²¹	[166]	C	1	3,88.10 ⁻⁰²
rs10817758	9	118143933	DEC1	T/C	T	1,15.10 ⁻⁰⁸	[377]	T	1	3,54.10 ⁻⁰¹
rs149308743	9	139258965	CARD9	T/C	T	2,09.10 ⁻⁰⁸	[378]	T	0,514	9,79.10 ⁻⁰¹
rs58600253	10	64507904	Intergénique	T/C	T	3,02.10 ⁻¹²	[377]	T	0,987	3,17.10 ⁻⁰¹
rs780668	10	73111408	SLC29A3	T/C	T	2,17.10 ⁻⁰⁹	[378]	T	1	4,46.10 ⁻⁰²

SNP	CHR	Position	Gène	Études précédemment publiées				Notre GWAS au Vietnam		
				m/M	RA	P-val	Référence	RA	Info	P
rs663743	11	64107735	CCDC88B	A/G	A	8,84.10 ⁻¹⁴	[377]	A	0,985	2,14.10 ⁻⁰²
rs925368	12	110390979	GIT2	C/T	C	9,18.10 ⁻¹⁷	[384]	C	0,877	7,80.10 ⁻⁰¹
rs671	12	112241766	ALDH2	A/G	A	2,00.10 ⁻²⁰	[384]	A	1	2,33.10 ⁻⁰¹
rs3764147	13	44457925	LACC1	A/G	G	3,72.10⁻⁵⁴	[166]	G	1	6,56.10⁻⁰⁵
rs8002861	13	44474517	LACC1	G/A	A	1,07.10 ⁻¹²	[377]	A	1	5,70.10 ⁻⁰²
rs142179458	14	62203623	HIF1A	G/A	A	4,95.10 ⁻⁰⁹	[385]	A	0,687	1,52.10 ⁻⁰¹
rs77061563	16	11452542	LOC388210	T/C	C	6,23.10 ⁻¹⁵	[377]	T	0,904	7,08.10 ⁻⁰¹
rs9302752	16	50719103	NOD2	A/G	G	3,77.10 ⁻⁴⁰	[166]	C	1	9,83.10 ⁻⁰²
rs16948876	16	50854418	CYLD	A/G	A	1,64.10 ⁻¹⁰	[375]	A	0,777	2,17.10 ⁻⁰¹
rs75680863	22	31007023	TCN2	T/A	A	8,37.10 ⁻²¹	[384]	T	0,804	2,54.10 ⁻⁰¹

Tableau 26 : Résultats d'association dans l'échantillon primaire familial des 33 SNPs hors-HLA et précédemment publiés

RA = allèle à risque,

Les positions sont données en fonction de hg19,

rs145562243 ne fait pas partie du panel de 1000 génomes (Phase 1 et 3),

Les SNPs en vert sont significatifs dans l'échantillon de réplication.

**ANNEXE 2 : Résultats d'association des 358 SNPs
dans la région HLA dépassant le seuil évocateur
lors de l'étude primaire familiale**

SNP	Position	m/M	Étude primaire					Réplication				
			MAF	info	RA	nfam	P	MAF	info	RA	P	OR [IC 95%]
rs1265048	31081409	C/T	0,462	1	T	310	8,72049.10 ⁻⁰⁶	0,468884	0,942	T	1,58285.10 ⁻⁰⁶	0,721]0 - 0,81]
rs707913	31084787	G/A	0,252	0,997	A	298	3,11437.10 ⁻⁰⁶	0,269246	0,985	A	0,0162387	0,851]0 - 0,96]
rs1265045	31085340	A/G	0,252	0,997	G	299	3,17807.10 ⁻⁰⁶	0,267198	0,979	G	0,01828185	0,853]0 - 0,97]
rs28383830	31088950	TGG/T	0,301	0,995	T	331	3,80686.10 ⁻⁰⁶	0,318851	0,967	T	0,0194322	0,86]0 - 0,97]
rs3094208	31090672	G/A	0,426	0,999	A	320	1,07251.10 ⁻⁰⁶	0,443509	0,997	A	0,02495445	0,874]0 - 0,98]
rs3130998	31090945	T/C	0,431	0,999	C	321	1,13543.10 ⁻⁰⁶	0,447631	0,992	C	0,02938365	0,878]0 - 0,98]
rs3132542	31091163	A/G	0,431	0,999	G	318	1,14391.10 ⁻⁰⁶	0,447556	0,993	G	0,02978155	0,878]0 - 0,98]
rs3095310	31091447	T/C	0,431	0,999	C	318	1,16338.10 ⁻⁰⁶	0,44754	0,993	C	0,03009215	0,879]0 - 0,98]
rs3095309	31091475	T/C	0,476	0,999	T	324	8,16477.10 ⁻⁰⁹	0,444361	0,977	T	9,46.10 ⁻⁰⁵	1,297 [1,16 - Inf[
rs3094206	31091642	G/C	0,416	0,999	C	324	2,11699.10 ⁻⁰⁶	0,437176	0,993	C	0,00424399	0,835]0 - 0,93]
rs3094205	31091862	G/A	0,395	1	A	318	3,91571.10 ⁻⁰⁷	0,415089	1	A	0,0289926	0,877]0 - 0,98]
rs3094204	31091992	A/G	0,391	1	A	324	1,38138.10 ⁻⁰⁷	0,376639	0,983	A	0,000178394	1,29 [1,15 - Inf[
rs3094199	31093326	A/G	0,184	0,998	A	221	5,19623.10 ⁻⁰⁷	0,171306	0,998	A	2,29996.10 ⁻⁰⁵	1,475 [1,26 - Inf[
rs3094197	31093485	A/G	0,184	0,998	A	221	5,252.10 ⁻⁰⁷	0,171714	1	A	2,37825.10 ⁻⁰⁵	1,473 [1,26 - Inf[
rs3094196	31093947	G/A	0,269	0,997	A	305	4,94733.10 ⁻⁰⁸	0,304487	0,974	A	4,70.10 ⁻⁰⁴	0,789]0 - 0,89]
rs3131005	31095864	G/A	0,18	0,997	G	221	1,39111.10 ⁻⁰⁶	0,169985	0,994	G	1,84629.10 ⁻⁰⁵	1,481 [1,26 - Inf[
rs2844634	31096184	C/G	0,269	0,998	G	307	1,73449.10 ⁻⁰⁸	0,302326	1	G	4,65.10 ⁻⁰⁴	0,793]0 - 0,89]
rs2844633	31096189	T/C	0,269	0,998	C	307	1,73308.10 ⁻⁰⁸	0,302055	1	C	4,20.10 ⁻⁰⁴	0,791]0 - 0,89]
rs10947140	31102143	A/G	0,364	0,991	G	337	7,46156.10 ⁻⁰⁶	0,37956	0,938	G	0,000196483	0,778]0 - 0,87]
rs3132557	31106409	G/A	0,384	0,989	A	319	6,16345.10 ⁻⁰⁶	0,377354	0,937	A	0,000297875	0,784]0 - 0,88]
rs114240643	31108628	G/A	0,117	0,994	G	147	9,02682.10 ⁻⁰⁶	0,113536	0,972	G	8,44205.10 ⁻¹⁰	2,042 [1,66 - Inf[
rs1265082	31111606	A/G	0,236	0,997	G	274	7,19662.10 ⁻⁰⁷	0,243239	0,96	G	0,000474965	0,773]0 - 0,88]
rs1265113	31117577	G/C	0,237	0,998	C	274	8,35912.10 ⁻⁰⁷	0,24291	0,962	C	0,000486452	0,773]0 - 0,88]
rs1265110	31119422	T/C	0,237	0,999	C	274	9,12075.10 ⁻⁰⁷	0,242722	0,963	C	0,000494401	0,774]0 - 0,88]

SNP	Position	m/M	Étude primaire					Réplication				
			MAF	info	RA	nfam	P	MAF	info	RA	P	OR [IC 95%]
rs1265107	31121529	T/G	0,237	0,999	G	274	9,33801.10 ⁻⁰⁷	0,242684	0,963	G	0,000496658	0,774]0 - 0,88]
rs1265163	31134087	G/C	0,237	0,998	C	273	1,78593.10 ⁻⁰⁶	0,242372	0,966	C	0,00050538	0,774]0 - 0,88]
rs3094194	31136240	G/A	0,46	1	G	314	1,42614.10 ⁻⁰⁸	0,444076	0,989	G	5,03.10 ⁻⁰⁶	1,359 [1,21 - Inf[
rs9281239	31137715	CCAGA/C	0,275	0,997	C	302	8,95487.10 ⁻⁰⁷	0,279159	0,969	C	0,00090405	0,793]0 - 0,9]
rs1265160	31138371	A/G	0,244	0,989	G	328	1,94129.10 ⁻⁰⁶	0,254178	0,951	G	0,000305988	0,767]0 - 0,87]
rs9461650	31143241	A/G	0,298	1	A	271	9,52788.10 ⁻⁰⁷	0,28186	0,954	A	3,60472.10 ⁻⁰⁷	1,479 [1,3 - Inf[
rs3734869	31143639	A/G	0,134	1	A	159	8,86805.10 ⁻⁰⁶	0,136638	0,956	A	7,63355.10 ⁻⁰⁹	1,852 [1,54 - Inf[
rs56178408	31150039	T/C	0,121	0,999	T	150	9,42962.10 ⁻⁰⁶	0,123997	0,973	T	4,71539.10 ⁻¹⁰	2,015 [1,66 - Inf[
rs138693838	31150159	CTG/C	0,134	0,999	CTG	160	9,02978.10 ⁻⁰⁶	0,136464	0,956	CTG	7,26615.10 ⁻⁰⁹	1,855 [1,54 - Inf[
rs4713440	31150323	A/C	0,121	0,999	A	150	9,39967.10 ⁻⁰⁶	0,12465	0,969	A	4,67131.10 ⁻¹⁰	2,015 [1,66 - Inf[
rs9468884	31151319	C/T	0,121	0,999	C	150	9,45919.10 ⁻⁰⁶	0,124037	0,972	C	4,68199.10 ⁻¹⁰	2,016 [1,66 - Inf[
rs9468885	31151413	C/A	0,134	0,999	C	160	9,06342.10 ⁻⁰⁶	0,136441	0,956	C	7,15915.10 ⁻⁰⁹	1,855 [1,54 - Inf[
rs9468886	31151573	A/G	0,121	0,999	A	150	9,45767.10 ⁻⁰⁶	0,124042	0,972	A	4,75245.10 ⁻¹⁰	2,015 [1,65 - Inf[
rs6913894	31151997	A/G	0,121	0,999	A	150	9,41752.10 ⁻⁰⁶	0,122963	0,972	A	9,85525.10 ⁻¹⁰	1,996 [1,64 - Inf[
rs6913631	31152098	C/A	0,121	0,999	C	150	9,47041.10 ⁻⁰⁶	0,124082	0,973	C	4,76474.10 ⁻¹⁰	2,015 [1,65 - Inf[
rs56332747	31152321	A/G	0,121	0,999	A	150	9,47286.10 ⁻⁰⁶	0,124044	0,972	A	4,75571.10 ⁻¹⁰	2,015 [1,65 - Inf[
rs9501351	31153126	T/C	0,121	0,999	T	150	9,49217.10 ⁻⁰⁶	0,124042	0,972	T	4,75975.10 ⁻¹⁰	2,015 [1,65 - Inf[
rs9501506	31153573	C/T	0,121	0,999	C	150	9,49869.10 ⁻⁰⁶	0,124041	0,972	C	4,76662.10 ⁻¹⁰	2,015 [1,65 - Inf[
rs9501507	31153716	C/T	0,121	0,999	C	150	9,49891.10 ⁻⁰⁶	0,124079	0,973	C	4,7744.10 ⁻¹⁰	2,015 [1,65 - Inf[
rs115737821	31154138	G/C	0,121	0,999	G	150	9,5077.10 ⁻⁰⁶	0,123598	0,972	G	7,067.10 ⁻¹⁰	2,001 [1,64 - Inf[
rs17191040	31154206	G/T	0,121	0,999	G	150	9,5077.10 ⁻⁰⁶	0,123598	0,972	G	7,067.10 ⁻¹⁰	2,001 [1,64 - Inf[
rs17197402	31154384	A/G	0,121	0,999	A	150	9,51401.10 ⁻⁰⁶	0,123598	0,972	A	7,07455.10 ⁻¹⁰	2,001 [1,64 - Inf[
rs199987285	31154473	CTTG/C	0,121	0,999	CTTG	150	9,51424.10 ⁻⁰⁶	0,123598	0,972	CTTG	7,0731.10 ⁻¹⁰	2,001 [1,64 - Inf[
rs149432671	31156350	C/A	0,121	0,999	C	150	9,43943.10 ⁻⁰⁶	0,122507	0,972	C	1,46689.10 ⁻⁰⁹	1,982 [1,63 - Inf[

SNP	Position	m/M	Étude primaire					Réplication				
			MAF	info	RA	nfam	P	MAF	info	RA	P	OR [IC 95%]
rs9500884	31165733	T/C	0,12	0,998	T	154	9,4223.10 ⁻⁰⁶	0,114105	0,977	T	1,92309.10 ⁻⁰⁹	2,005 [1,64 - Inf[
rs9501079	31175097	A/G	0,123	1	A	154	8,68744.10 ⁻⁰⁶	0,12375	0,984	A	5,21245.10 ⁻¹¹	2,093 [1,72 - Inf[
rs9501082	31175251	T/C	0,123	1	T	154	8,68744.10 ⁻⁰⁶	0,12375	0,984	T	5,2169.10 ⁻¹¹	2,093 [1,72 - Inf[
rs4516988	31176602	G/A	0,123	0,999	G	155	9,3182.10 ⁻⁰⁶	0,123715	0,984	G	5,21045.10 ⁻¹¹	2,093 [1,72 - Inf[
rs9501513	31176678	G/A	0,123	0,999	G	154	9,48551.10 ⁻⁰⁶	0,123717	0,984	G	5,2056.10 ⁻¹¹	2,093 [1,72 - Inf[
rs3130513	31178084	G/C	0,364	0,998	C	330	2,30119.10 ⁻⁰⁶	0,385403	0,962	C	0,072923	0,901]0 - 1,01]
rs145021772	31189460	T/C	0,157	1	T	189	8,92082.10 ⁻⁰⁶	0,151984	0,992	T	2,91541.10 ⁻⁰⁵	1,493 [1,26 - Inf[
rs201234764	31190027	G/GC	0,157	1	G	189	8,91635.10 ⁻⁰⁶	0,151987	0,992	G	2,92239.10 ⁻⁰⁵	1,493 [1,26 - Inf[
rs113565451	31190175	T/C	0,157	1	T	189	8,91635.10 ⁻⁰⁶	0,151985	0,992	T	2,92303.10 ⁻⁰⁵	1,493 [1,26 - Inf[
rs35391481	31190736	A/C	0,157	1	A	198	8,77935.10 ⁻⁰⁶	0,152039	0,992	A	2,94642.10 ⁻⁰⁵	1,492 [1,26 - Inf[
rs12662501	31190850	T/C	0,157	1	T	187	8,98763.10 ⁻⁰⁶	0,151986	0,992	T	2,92634.10 ⁻⁰⁵	1,493 [1,26 - Inf[
rs12662527	31190897	A/G	0,157	1	A	188	8,91189.10 ⁻⁰⁶	0,15244	0,989	A	3,10144.10 ⁻⁰⁵	1,491 [1,26 - Inf[
rs35891864	31191099	A/G	0,157	1	A	188	8,91189.10 ⁻⁰⁶	0,151987	0,992	A	2,93037.10 ⁻⁰⁵	1,492 [1,26 - Inf[
rs141822551	31191249	A/AC	0,148	0,968	A	201	3,75455.10 ⁻⁰⁶	0,138323	0,972	A	1,49189.10 ⁻⁰⁷	1,725 [1,44 - Inf[
rs139975409	31191259	A/C	0,148	0,968	A	201	3,75455.10 ⁻⁰⁶	0,138323	0,972	A	1,49227.10 ⁻⁰⁷	1,725 [1,44 - Inf[
rs113831989	31191294	T/C	0,148	0,968	T	201	3,75455.10 ⁻⁰⁶	0,138323	0,972	T	1,49234.10 ⁻⁰⁷	1,725 [1,44 - Inf[
rs112056872	31192675	T/A	0,157	1	T	206	8,9523.10 ⁻⁰⁶	0,152397	0,992	T	3,01669.10 ⁻⁰⁵	1,491 [1,26 - Inf[
rs2394887	31194711	T/G	0,157	1	T	188	8,87668.10 ⁻⁰⁶	0,15226	0,993	T	2,9419.10 ⁻⁰⁵	1,492 [1,26 - Inf[
rs112815193	31196390	A/G	0,157	1	A	189	8,80734.10 ⁻⁰⁶	0,152571	0,993	A	2,81927.10 ⁻⁰⁵	1,493 [1,26 - Inf[
rs56210254	31196640	G/A	0,157	1	G	188	8,86354.10 ⁻⁰⁶	0,152454	0,994	G	2,80283.10 ⁻⁰⁵	1,493 [1,26 - Inf[
rs12664433	31200672	A/G	0,172	0,996	A	205	9,68689.10 ⁻⁰⁷	0,164657	0,994	A	1,61681.10 ⁻⁰⁶	1,568 [1,33 - Inf[
rs112758902	31201534	T/C	0,17	1	T	198	8,82611.10 ⁻⁰⁷	0,164638	0,995	T	1,62742.10 ⁻⁰⁶	1,567 [1,33 - Inf[
rs55783607	31202764	C/T	0,17	1	C	198	8,80925.10 ⁻⁰⁷	0,164638	0,995	C	1,65595.10 ⁻⁰⁶	1,567 [1,33 - Inf[
rs72863879	31203352	A/G	0,17	1	A	198	8,80204.10 ⁻⁰⁷	0,164644	0,995	A	1,6662.10 ⁻⁰⁶	1,567 [1,33 - Inf[

SNP	Position	m/M	Étude primaire					Réplication				
			MAF	info	RA	nfam	P	MAF	info	RA	P	OR [IC 95%]
rs113288104	31204402	T/G	0,17	1	T	198	8,78767.10 ⁻⁰⁷	0,164652	0,995	T	1,68731.10 ⁻⁰⁶	1,566 [1,33 - Inf[
rs377461698	31205030	C/A	0,17	1	C	198	8,7781.10 ⁻⁰⁷	0,164658	0,995	C	1,70092.10 ⁻⁰⁶	1,566 [1,33 - Inf[
rs371249684	31205031	C/A	0,17	1	C	198	8,7781.10 ⁻⁰⁷	0,164658	0,995	C	1,70092.10 ⁻⁰⁶	1,566 [1,33 - Inf[
rs3869118	31206136	C/G	0,17	1	C	198	8,76142.10 ⁻⁰⁷	0,164671	0,995	C	1,73001.10 ⁻⁰⁶	1,565 [1,33 - Inf[
rs113390008	31207351	A/G	0,17	1	A	197	8,79774.10 ⁻⁰⁷	0,164792	0,995	A	1,70099.10 ⁻⁰⁶	1,566 [1,33 - Inf[
rs111411820	31207589	A/C	0,171	1	A	196	1,10717.10 ⁻⁰⁶	0,166734	0,995	A	4,47998.10 ⁻⁰⁶	1,531 [1,3 - Inf[
rs12665166	31208242	T/C	0,171	1	T	196	1,10717.10 ⁻⁰⁶	0,166777	0,995	T	4,56799.10 ⁻⁰⁶	1,53 [1,3 - Inf[
rs35593555	31208289	C/CATG	0,351	1	CATG	308	1,13294.10 ⁻⁰⁶	0,376349	0,994	CATG	0,04523695	0,888]0 - 1]
rs3095250	31208340	T/C	0,394	1	C	314	7,17926.10 ⁻⁰⁸	0,412424	0,991	C	3,13.10 ⁻⁰²	0,879]0 - 0,99]
rs3130534	31209045	A/G	0,351	1	G	308	1,13294.10 ⁻⁰⁶	0,376372	0,994	G	0,04525615	0,888]0 - 1]
rs112030670	31209380	A/G	0,169	1	A	197	8,92517.10 ⁻⁰⁷	0,164788	0,996	A	1,76596.10 ⁻⁰⁶	1,564 [1,33 - Inf[
rs114475699	31211201	C/T	0,169	1	C	197	8,99479.10 ⁻⁰⁷	0,164842	0,996	C	1,71991.10 ⁻⁰⁶	1,565 [1,33 - Inf[
rs4332019	31213413	G/T	0,171	1	G	199	1,10093.10 ⁻⁰⁶	0,166998	0,995	G	4,6198.10 ⁻⁰⁶	1,529 [1,3 - Inf[
rs2394897	31213502	A/G	0,171	1	A	197	1,10776.10 ⁻⁰⁶	0,166952	0,996	A	4,58201.10 ⁻⁰⁶	1,53 [1,3 - Inf[
rs2394898	31213583	C/T	0,377	0,911	T	395	1,84941.10 ⁻⁰⁶	0,377365	0,924	T	0,00396969	0,829]0 - 0,93]
rs79997604	31213587	T/C	0,377	0,911	C	395	1,85077.10 ⁻⁰⁶	0,377365	0,924	C	0,00396964	0,829]0 - 0,93]
rs3899471	31215037	C/T	0,171	1	C	198	1,10836.10 ⁻⁰⁶	0,166989	0,996	C	4,59355.10 ⁻⁰⁶	1,529 [1,3 - Inf[
rs12662550	31215732	G/A	0,171	1	G	198	1,10866.10 ⁻⁰⁶	0,167012	0,996	G	4,59927.10 ⁻⁰⁶	1,529 [1,3 - Inf[
rs12665531	31215916	A/G	0,171	0,999	A	201	1,00621.10 ⁻⁰⁶	0,167492	0,994	A	3,7092.10 ⁻⁰⁶	1,536 [1,31 - Inf[
rs2394904	31216105	A/G	0,171	1	A	198	1,10866.10 ⁻⁰⁶	0,167025	0,996	A	4,6026.10 ⁻⁰⁶	1,529 [1,3 - Inf[
rs2394907	31217943	T/C	0,348	1	C	309	8,43286.10 ⁻⁰⁷	0,370554	0,997	C	0,0370317	0,882]0 - 0,99]
rs55651925	31218131	C/G	0,169	0,999	C	199	9,16257.10 ⁻⁰⁷	0,164949	0,996	C	1,63857.10 ⁻⁰⁶	1,566 [1,33 - Inf[
rs4084262	31218889	T/C	0,169	0,999	T	205	9,23004.10 ⁻⁰⁷	0,165075	0,996	T	1,62957.10 ⁻⁰⁶	1,566 [1,33 - Inf[
rs3130434	31219803	C/G	0,379	0,982	G	332	1,70931.10 ⁻⁰⁷	0,382905	0,995	G	0,0326289	0,882]0 - 0,99]

SNP	Position	m/M	Étude primaire					Réplication				
			MAF	info	RA	nfam	P	MAF	info	RA	P	OR [IC 95%]
rs3130435	31219805	A/G	0,373	0,981	G	334	1,12853.10 ⁻⁰⁷	0,382028	0,996	G	0,0250199	0,875]0 - 0,98]
rs3130540	31219862	C/T	0,373	0,982	T	333	1,14159.10 ⁻⁰⁷	0,382299	0,997	T	0,0246597	0,875]0 - 0,98]
rs9264179	31219869	C/T	0,379	0,982	T	332	1,69298.10 ⁻⁰⁷	0,383698	0,996	T	0,0275646	0,878]0 - 0,98]
rs9264180	31219902	A/C	0,377	0,981	C	335	1,67864.10 ⁻⁰⁷	0,3894	1	C	0,0303032	0,881]0 - 0,98]
rs9264181	31219922	T/C	0,311	0,977	C	334	6,92365.10 ⁻⁰⁶	0,317469	1	C	0,01907375	0,866]0 - 0,97]
rs9264183	31219934	G/A	0,311	0,977	A	334	6,92657.10 ⁻⁰⁶	0,317469	1	A	0,01907375	0,866]0 - 0,97]
rs56375204	31219935	C/G	0,175	0,984	C	232	1,94489.10 ⁻⁰⁶	0,171714	1	C	9,0263.10 ⁻⁰⁶	1,485 [1,27 - Inf[
rs55685143	31219936	T/G	0,175	0,984	T	232	1,94489.10 ⁻⁰⁶	0,171444	1	T	1,00655.10 ⁻⁰⁵	1,483 [1,27 - Inf[
rs9264184	31219945	A/G	0,311	0,977	G	334	6,92495.10 ⁻⁰⁶	0,317469	1	G	0,01907375	0,866]0 - 0,97]
rs55839536	31219958	A/G	0,173	0,984	A	232	1,62703.10 ⁻⁰⁶	0,169551	1	A	3,35542.10 ⁻⁰⁶	1,519 [1,3 - Inf[
rs9264186	31220017	A/G	0,374	0,982	G	337	1,18063.10 ⁻⁰⁷	0,38314	0,998	G	0,0250835	0,876]0 - 0,98]
rs9264187	31220027	C/T	0,379	0,982	T	335	1,77849.10 ⁻⁰⁷	0,384213	0,997	T	0,0307139	0,881]0 - 0,98]
rs9264188	31220109	T/C	0,373	0,981	C	335	1,14434.10 ⁻⁰⁷	0,382047	0,997	C	0,0236131	0,874]0 - 0,98]
rs3095257	31220318	T/G	0,375	0,981	G	354	1,04746.10 ⁻⁰⁷	0,384173	0,997	G	0,0168978	0,866]0 - 0,97]
rs2394945	31220971	G/C	0,353	0,984	G	327	4,19088.10 ⁻⁰⁸	0,348447	0,971	G	9,20.10 ⁻⁰⁷	1,416 [1,25 - Inf[
rs2524084	31241639	G/A	0,369	0,999	A	317	3,32959.10 ⁻⁰⁶	0,385343	1	A	0,00050099	0,795]0 - 0,89]
rs12662096	31242445	C/T	0,131	0,998	C	166	3,52123.10 ⁻⁰⁶	0,125198	0,998	C	8,505.10 ⁻¹¹	2,058 [1,69 - Inf[
rs2524075	31243859	T/C	0,194	0,999	C	239	9,63573.10 ⁻⁰⁶	0,208626	0,994	C	0,0054809	0,808]0 - 0,93]
rs9468913	31244332	T/C	0,132	1	T	165	3,6746.10 ⁻⁰⁶	0,125633	0,999	T	3,0757.10 ⁻¹¹	2,095 [1,72 - Inf[
rs72855954	31246969	G/A	0,159	0,996	G	200	6,16496.10 ⁻⁰⁶	0,153597	1	G	1,10058.10 ⁻⁰⁶	1,608 [1,36 - Inf[
rs2853946	31247203	T/A	0,33	0,998	T	305	1,10848.10 ⁻⁰⁷	0,321525	1	T	5,0475.10 ⁻⁰⁶	1,392 [1,23 - Inf[
rs28457789	31247801	G/T	0,132	0,999	G	165	3,76142.10 ⁻⁰⁶	0,125461	0,998	G	2,38057.10 ⁻¹¹	2,106 [1,73 - Inf[
rs28709322	31247804	A/G	0,132	0,999	A	165	3,76227.10 ⁻⁰⁶	0,125459	0,998	A	2,3711.10 ⁻¹¹	2,107 [1,73 - Inf[
rs17192281	31248075	A/G	0,133	0,998	A	167	4,26831.10 ⁻⁰⁶	0,128141	0,994	A	4,16853.10 ⁻¹⁰	1,983 [1,64 - Inf[

SNP	Position	m/M	Étude primaire					Réplication				
			MAF	info	RA	nfam	P	MAF	info	RA	P	OR [IC 95%]
rs17192288	31248091	A/G	0,131	0,999	A	166	3,79097.10 ⁻⁰⁶	0,125431	0,998	A	2,31154.10 ⁻¹¹	2,108 [1,73 - Inf[
rs17192295	31248114	T/C	0,131	0,999	T	166	3,79358.10 ⁻⁰⁶	0,125429	0,998	T	2,30558.10 ⁻¹¹	2,108 [1,73 - Inf[
rs17192302	31248256	T/C	0,131	0,999	T	166	3,80848.10 ⁻⁰⁶	0,125413	0,998	T	2,27169.10 ⁻¹¹	2,109 [1,73 - Inf[
rs17192323	31248324	T/C	0,131	0,999	T	166	3,81642.10 ⁻⁰⁶	0,125405	0,998	T	2,25474.10 ⁻¹¹	2,109 [1,73 - Inf[
rs17198664	31248382	T/C	0,133	0,998	T	167	4,33799.10 ⁻⁰⁶	0,12811	0,994	T	4,00321.10 ⁻¹⁰	1,985 [1,64 - Inf[
rs2844606	31248432	A/G	0,472	0,996	G	339	6,73388.10 ⁻⁰⁶	0,485447	0,994	G	0,000870425	0,806]0 - 0,9]
rs2844605	31248591	A/C	0,473	0,996	C	339	6,76544.10 ⁻⁰⁶	0,485469	0,994	C	0,000863865	0,805]0 - 0,9]
rs17198706	31248693	G/C	0,131	0,999	G	166	3,86318.10 ⁻⁰⁶	0,125358	0,998	G	2,15749.10 ⁻¹¹	2,111 [1,74 - Inf[
rs58114162	31248825	T/C	0,131	0,999	T	166	3,88432.10 ⁻⁰⁶	0,125337	0,998	T	2,11841.10 ⁻¹¹	2,112 [1,74 - Inf[
rs73728770	31248887	C/G	0,131	0,999	C	166	3,89267.10 ⁻⁰⁶	0,125328	0,998	C	2,1005.10 ⁻¹¹	2,112 [1,74 - Inf[
rs189616187	31248995	T/C	0,131	0,999	T	166	3,91138.10 ⁻⁰⁶	0,125309	0,998	T	2,06818.10 ⁻¹¹	2,113 [1,74 - Inf[
rs2253489	31249195	G/C	0,131	0,999	G	166	3,94824.10 ⁻⁰⁶	0,125276	0,998	G	2,00547.10 ⁻¹¹	2,114 [1,74 - Inf[
rs73392854	31249326	G/A	0,131	0,999	G	167	3,95785.10 ⁻⁰⁶	0,125982	0,996	G	1,45965.10 ⁻¹¹	2,126 [1,75 - Inf[
rs141206831	31249358	A/AAAG	0,131	0,997	A	169	4,35254.10 ⁻⁰⁶	0,125577	0,995	A	1,64104.10 ⁻¹¹	2,123 [1,75 - Inf[
rs28383843	31249893	GT/G	0,133	0,999	GT	168	3,73449.10 ⁻⁰⁶	0,127243	0,996	GT	1,48942.10 ⁻¹¹	2,118 [1,74 - Inf[
rs114612857	31250876	G/C	0,134	0,999	G	169	3,74927.10 ⁻⁰⁶	0,129976	0,999	G	5,6673.10 ⁻¹¹	2,046 [1,69 - Inf[
rs6921855	31251312	C/T	0,164	0,935	T	293	1,12988.10 ⁻⁰⁶	0,180483	0,888	T	0,000234215	0,723]0 - 0,84]
rs2524058	31251569	A/G	0,488	0,998	G	339	1,77248.10 ⁻⁰⁶	0,492191	0,99	G	0,000635135	0,801]0 - 0,9]
rs6899660	31251643	G/A	0,187	1	A	229	1,59688.10 ⁻⁰⁶	0,202864	0,986	A	0,000893955	0,769]0 - 0,88]
rs2524057	31251895	G/A	0,304	1	G	281	4,0282.10 ⁻⁰⁷	0,296222	0,998	G	1,7932.10 ⁻⁰⁸	1,53 [1,34 - Inf[
rs115182712	31251954	G/A	0,133	0,999	G	169	3,77605.10 ⁻⁰⁶	0,129955	0,999	G	5,58695.10 ⁻¹¹	2,047 [1,69 - Inf[
rs114020623	31251962	T/C	0,133	0,999	T	169	3,77605.10 ⁻⁰⁶	0,129955	0,999	T	5,58695.10 ⁻¹¹	2,047 [1,69 - Inf[
rs115699806	31252031	G/A	0,133	0,999	G	169	3,77713.10 ⁻⁰⁶	0,12966	0,996	G	5,43885.10 ⁻¹¹	2,051 [1,69 - Inf[
rs115931507	31252406	C/T	0,133	0,999	C	169	3,78687.10 ⁻⁰⁶	0,129945	0,999	C	5,54655.10 ⁻¹¹	2,047 [1,69 - Inf[

SNP	Position	m/M	Étude primaire					Réplication				
			MAF	info	RA	nfam	P	MAF	info	RA	P	OR [IC 95%]
rs116281893	31252527	C/T	0,133	0,999	C	169	3,79012.10 ⁻⁰⁶	0,129942	0,999	C	5,5382.10 ⁻¹¹	2,047 [1,69 - Inf[
rs9348860	31252667	C/T	0,187	1	T	229	1,58306.10 ⁻⁰⁶	0,203671	0,984	T	0,00105205	0,772]0 - 0,89]
rs9348861	31252722	G/A	0,187	1	A	229	1,58214.10 ⁻⁰⁶	0,202924	0,986	A	0,000880905	0,769]0 - 0,88]
rs9348862	31252747	C/T	0,187	1	T	229	1,58182.10 ⁻⁰⁶	0,202925	0,986	T	0,00088099	0,769]0 - 0,88]
rs9368670	31252882	T/C	0,187	1	C	229	1,58182.10 ⁻⁰⁶	0,202934	0,986	C	0,0008783	0,769]0 - 0,88]
rs116350468	31252951	T/G	0,133	0,999	T	169	3,80534.10 ⁻⁰⁶	0,12993	0,999	T	5,5031.10 ⁻¹¹	2,048 [1,69 - Inf[
rs115727572	31253034	G/T	0,133	0,999	G	169	3,80865.10 ⁻⁰⁶	0,129926	0,999	G	5,48845.10 ⁻¹¹	2,048 [1,69 - Inf[
rs114206206	31253468	T/G	0,133	0,999	T	169	3,82846.10 ⁻⁰⁶	0,12991	0,999	T	5,43405.10 ⁻¹¹	2,048 [1,69 - Inf[
rs114798972	31253514	T/C	0,133	0,999	T	169	3,83177.10 ⁻⁰⁶	0,129909	0,999	T	5,42715.10 ⁻¹¹	2,048 [1,69 - Inf[
rs73728779	31253605	C/T	0,136	0,995	C	192	7,24496.10 ⁻⁰⁶	0,133893	0,984	C	5,7844.10 ⁻¹⁰	1,95 [1,62 - Inf[
rs73728781	31253619	T/C	0,136	0,994	T	187	5,19599.10 ⁻⁰⁶	0,134276	0,984	T	1,2662.10 ⁻⁰⁹	1,917 [1,59 - Inf[
rs59217589	31254280	A/T	0,133	1	A	169	3,85852.10 ⁻⁰⁶	0,129886	0,999	A	5,35205.10 ⁻¹¹	2,049 [1,69 - Inf[
rs58410461	31254299	C/A	0,133	1	C	169	3,86191.10 ⁻⁰⁶	0,129885	0,999	C	5,34815.10 ⁻¹¹	2,049 [1,69 - Inf[
rs58022700	31254334	T/C	0,133	1	T	169	3,86303.10 ⁻⁰⁶	0,129884	0,999	T	5,3448.10 ⁻¹¹	2,049 [1,69 - Inf[
rs373238011	31254364	C/CT	0,304	1	C	281	4,08278.10 ⁻⁰⁷	0,296201	0,999	C	1,71315.10 ⁻⁰⁸	1,53 [1,34 - Inf[
rs58860177	31254485	C/T	0,133	1	C	169	3,8664.10 ⁻⁰⁶	0,12988	0,999	C	5,33195.10 ⁻¹¹	2,049 [1,69 - Inf[
rs56328367	31254594	C/T	0,133	1	C	169	3,87203.10 ⁻⁰⁶	0,129877	0,999	C	5,32345.10 ⁻¹¹	2,049 [1,69 - Inf[
rs9380236	31254664	G/A	0,187	1	A	229	1,58182.10 ⁻⁰⁶	0,20308	0,987	A	0,00084798	0,768]0 - 0,88]
rs55822325	31254691	G/A	0,133	1	G	169	3,87316.10 ⁻⁰⁶	0,129874	0,999	G	5,3097.10 ⁻¹¹	2,049 [1,69 - Inf[
rs9368671	31254704	T/C	0,187	1	C	229	1,58182.10 ⁻⁰⁶	0,203085	0,987	C	0,00084659	0,768]0 - 0,88]
rs369387030	31254836	T/TC	0,187	1	TC	229	1,58182.10 ⁻⁰⁶	0,203094	0,987	TC	0,000845165	0,768]0 - 0,88]
rs201734761	31254964	G/GAAA	0,135	0,999	G	170	4,80941.10 ⁻⁰⁶	0,132857	0,993	G	6,5957.10 ⁻¹⁰	1,943 [1,61 - Inf[
rs199720990	31254972	C/A	0,133	1	C	169	3,88336.10 ⁻⁰⁶	0,129867	0,999	C	5,284.10 ⁻¹¹	2,049 [1,69 - Inf[
rs72861464	31255063	G/A	0,187	1	A	229	1,58182.10 ⁻⁰⁶	0,203109	0,987	A	0,000841955	0,768]0 - 0,88]

SNP	Position	m/M	Étude primaire					Réplication				
			MAF	info	RA	nfam	P	MAF	info	RA	P	OR [IC 95%]
rs57599285	31255183	T/C	0,133	1	T	169	3,89018.10 ⁻⁰⁶	0,129862	0,999	T	5,2683.10 ⁻¹¹	2,049 [1,69 - Inf[
rs16899147	31255201	G/A	0,187	1	A	229	1,58182.10 ⁻⁰⁶	0,20312	0,987	A	0,000839995	0,768]0 - 0,88]
rs57742463	31255210	A/C	0,133	1	A	169	3,89245.10 ⁻⁰⁶	0,129861	0,999	A	5,26505.10 ⁻¹¹	2,049 [1,69 - Inf[
rs72861465	31255344	A/G	0,187	1	G	229	1,58182.10 ⁻⁰⁶	0,203129	0,987	G	0,000837985	0,768]0 - 0,88]
rs114598080	31255442	A/G	0,133	1	A	169	3,9016.10 ⁻⁰⁶	0,13007	1	A	4,39264.10 ⁻¹¹	2,055 [1,7 - Inf[
rs55825863	31255465	A/C	0,133	1	A	169	3,9016.10 ⁻⁰⁶	0,130064	1	A	4,41027.10 ⁻¹¹	2,055 [1,7 - Inf[
rs56184612	31255523	G/A	0,133	1	G	169	3,90389.10 ⁻⁰⁶	0,130064	1	G	4,41026.10 ⁻¹¹	2,055 [1,7 - Inf[
rs56016471	31255555	C/A	0,133	1	C	169	3,90504.10 ⁻⁰⁶	0,130064	1	C	4,41026.10 ⁻¹¹	2,055 [1,7 - Inf[
rs55918251	31255811	G/A	0,133	1	G	169	3,91308.10 ⁻⁰⁶	0,130065	1	G	4,40947.10 ⁻¹¹	2,055 [1,7 - Inf[
rs2524049	31255953	G/C	0,304	1	G	281	4,11721.10 ⁻⁰⁷	0,296192	1	G	1,68619.10 ⁻⁰⁸	1,53 [1,35 - Inf[
rs17198727	31256013	T/C	0,187	1	C	229	1,58182.10 ⁻⁰⁶	0,203174	0,987	C	0,00082846	0,768]0 - 0,88]
rs17192393	31256065	G/C	0,133	1	G	169	3,92231.10 ⁻⁰⁶	0,130064	1	G	4,41029.10 ⁻¹¹	2,055 [1,7 - Inf[
rs9366776	31256630	A/G	0,187	1	G	229	1,58182.10 ⁻⁰⁶	0,203217	0,987	G	0,00081942	0,768]0 - 0,88]
rs17192421	31256693	C/T	0,133	1	C	168	3,94602.10 ⁻⁰⁶	0,130065	1	C	4,40948.10 ⁻¹¹	2,055 [1,7 - Inf[
rs17198776	31256732	T/C	0,133	1	T	167	3,94836.10 ⁻⁰⁶	0,130065	1	T	4,40948.10 ⁻¹¹	2,055 [1,7 - Inf[
rs17198783	31256746	C/T	0,133	1	C	167	3,94836.10 ⁻⁰⁶	0,130113	0,999	C	4,19371.10 ⁻¹¹	2,057 [1,7 - Inf[
rs2524044	31256753	G/T	0,304	1	G	280	4,17908.10 ⁻⁰⁷	0,296217	0,999	G	1,68302.10 ⁻⁰⁸	1,53 [1,35 - Inf[
rs10947169	31256916	C/T	0,187	1	T	229	1,58182.10 ⁻⁰⁶	0,203238	0,987	T	0,0008152	0,768]0 - 0,88]
rs114015300	31256965	A/C	0,133	1	A	167	3,95081.10 ⁻⁰⁶	0,130065	1	A	4,40947.10 ⁻¹¹	2,055 [1,7 - Inf[
rs115007911	31256982	A/C	0,133	1	A	167	3,95163.10 ⁻⁰⁶	0,130065	1	A	4,40947.10 ⁻¹¹	2,055 [1,7 - Inf[
rs146303266	31257093	GTGTCTC/G	0,133	1	GTGTCTC	167	3,95492.10 ⁻⁰⁶	0,130065	1	GTGTCTC	4,40947.10 ⁻¹¹	2,055 [1,7 - Inf[
rs115226524	31257120	C/T	0,133	1	C	167	3,95656.10 ⁻⁰⁶	0,130065	1	C	4,40866.10 ⁻¹¹	2,055 [1,7 - Inf[
rs2524042	31257160	G/C	0,304	1	G	280	4,14596.10 ⁻⁰⁷	0,296216	0,999	G	1,67362.10 ⁻⁰⁸	1,531 [1,35 - Inf[
rs115482903	31257198	A/G	0,133	1	A	167	3,96233.10 ⁻⁰⁶	0,130065	1	A	4,40866.10 ⁻¹¹	2,055 [1,7 - Inf[

SNP	Position	m/M	Étude primaire					Réplication				
			MAF	info	RA	nfam	P	MAF	info	RA	P	OR [IC 95%]
rs116736925	31257229	A/G	0,133	1	A	167	3,96399.10 ⁻⁰⁶	0,130065	1	A	4,40866.10 ⁻¹¹	2,055 [1,7 - Inf[
rs115504545	31257251	G/A	0,133	1	G	167	3,96564.10 ⁻⁰⁶	0,130065	1	G	4,40785.10 ⁻¹¹	2,055 [1,7 - Inf[
rs116337797	31257422	A/G	0,133	1	A	167	3,97645.10 ⁻⁰⁶	0,130065	1	A	4,40705.10 ⁻¹¹	2,055 [1,7 - Inf[
rs115651636	31257533	C/T	0,133	1	C	167	3,98147.10 ⁻⁰⁶	0,130065	1	C	4,40703.10 ⁻¹¹	2,055 [1,7 - Inf[
rs114110521	31257617	A/G	0,133	1	A	167	3,98398.10 ⁻⁰⁶	0,130065	1	A	4,40703.10 ⁻¹¹	2,055 [1,7 - Inf[
rs115545796	31258036	C/A	0,133	1	C	167	3,98987.10 ⁻⁰⁶	0,130065	1	C	4,40705.10 ⁻¹¹	2,055 [1,7 - Inf[
rs182445956	31258046	C/T	0,133	1	C	167	3,98987.10 ⁻⁰⁶	0,130065	1	C	4,40705.10 ⁻¹¹	2,055 [1,7 - Inf[
rs2524039	31258226	G/A	0,304	1	G	280	4,17399.10 ⁻⁰⁷	0,296253	0,999	G	1,6586.10 ⁻⁰⁸	1,531 [1,35 - Inf[
rs114093160	31258371	A/G	0,133	1	A	170	4,00562.10 ⁻⁰⁶	0,130079	1	A	4,37174.10 ⁻¹¹	2,055 [1,7 - Inf[
rs367893790	31258626	C/CATAT	0,133	1	C	167	3,99833.10 ⁻⁰⁶	0,130066	1	C	4,40624.10 ⁻¹¹	2,055 [1,7 - Inf[
rs2256583	31258846	C/T	0,304	1	C	280	4,18126.10 ⁻⁰⁷	0,296261	0,999	C	1,65498.10 ⁻⁰⁸	1,531 [1,35 - Inf[
rs3873377	31258868	C/T	0,187	1	T	229	1,58182.10 ⁻⁰⁶	0,203506	0,988	T	0,0007632	0,767]0 - 0,88]
rs78950083	31258986	C/T	0,133	1	C	167	4,00087.10 ⁻⁰⁶	0,130066	1	C	4,40624.10 ⁻¹¹	2,055 [1,7 - Inf[
rs16899170	31259017	C/T	0,133	1	C	167	4,00087.10 ⁻⁰⁶	0,130066	1	C	4,40624.10 ⁻¹¹	2,055 [1,7 - Inf[
rs4307215	31259188	GTT/G	0,371	0,982	G	367	6,25105.10 ⁻⁰⁷	0,393949	0,948	G	0,0329502	0,877]0 - 0,99]
rs35667233	31259269	C/T	0,133	1	C	167	4,00087.10 ⁻⁰⁶	0,130066	1	C	4,40544.10 ⁻¹¹	2,055 [1,7 - Inf[
rs34115632	31259298	G/A	0,133	1	G	167	4,00087.10 ⁻⁰⁶	0,130066	1	G	4,40544.10 ⁻¹¹	2,055 [1,7 - Inf[
rs36087596	31259299	C/CT	0,133	1	C	167	4,00087.10 ⁻⁰⁶	0,130066	1	C	4,40544.10 ⁻¹¹	2,055 [1,7 - Inf[
rs35401506	31259332	C/T	0,133	1	C	167	4,00087.10 ⁻⁰⁶	0,130066	1	C	4,40544.10 ⁻¹¹	2,055 [1,7 - Inf[
rs36060788	31259338	G/A	0,133	1	G	167	4,00087.10 ⁻⁰⁶	0,130066	1	G	4,40544.10 ⁻¹¹	2,055 [1,7 - Inf[
rs35870025	31259435	T/C	0,133	1	T	167	4,00087.10 ⁻⁰⁶	0,130066	1	T	4,40544.10 ⁻¹¹	2,055 [1,7 - Inf[
rs34743644	31259443	T/C	0,133	1	T	167	4,00087.10 ⁻⁰⁶	0,130066	1	T	4,40544.10 ⁻¹¹	2,055 [1,7 - Inf[
rs116506525	31259458	C/T	0,133	1	C	167	4,00087.10 ⁻⁰⁶	0,130066	1	C	4,40544.10 ⁻¹¹	2,055 [1,7 - Inf[
rs116030808	31259495	G/A	0,133	1	G	167	4,00087.10 ⁻⁰⁶	0,130066	1	G	4,40544.10 ⁻¹¹	2,055 [1,7 - Inf[

SNP	Position	m/M	Étude primaire					Réplication				
			MAF	info	RA	nfam	P	MAF	info	RA	P	OR [IC 95%]
rs115175858	31259543	C/T	0,133	1	C	167	4,00087.10 ⁻⁰⁶	0,130066	1	C	4,40544.10 ⁻¹¹	2,055 [1,7 - Inf[
rs147841590	31259588	G/GA	0,133	1	G	167	4,00087.10 ⁻⁰⁶	0,130076	1	G	4,40336.10 ⁻¹¹	2,055 [1,7 - Inf[
rs114454776	31259605	A/G	0,133	1	A	167	4,00087.10 ⁻⁰⁶	0,130076	1	A	4,40336.10 ⁻¹¹	2,055 [1,7 - Inf[
rs115106318	31259704	G/A	0,133	1	G	167	4,00087.10 ⁻⁰⁶	0,130066	1	G	4,40544.10 ⁻¹¹	2,055 [1,7 - Inf[
rs34332804	31259729	C/G	0,133	1	C	167	4,00087.10 ⁻⁰⁶	0,130066	1	C	4,40544.10 ⁻¹¹	2,055 [1,7 - Inf[
rs34254374	31259834	G/T	0,133	1	G	167	4,00087.10 ⁻⁰⁶	0,130066	1	G	4,40544.10 ⁻¹¹	2,055 [1,7 - Inf[
rs35206735	31259837	A/G	0,133	1	A	167	4,00087.10 ⁻⁰⁶	0,130066	1	A	4,40544.10 ⁻¹¹	2,055 [1,7 - Inf[
rs114500903	31259869	C/T	0,133	1	C	167	4,00087.10 ⁻⁰⁶	0,130066	1	C	4,40544.10 ⁻¹¹	2,055 [1,7 - Inf[
rs116452682	31259870	A/G	0,133	1	A	167	4,00087.10 ⁻⁰⁶	0,130066	1	A	4,40544.10 ⁻¹¹	2,055 [1,7 - Inf[
rs6916525	31259902	G/A	0,187	1	A	229	1,58182.10 ⁻⁰⁶	0,203556	0,989	A	0,00075227	0,766]0 - 0,88]
rs6939681	31259952	G/T	0,187	1	T	229	1,58182.10 ⁻⁰⁶	0,203582	0,988	T	0,00074003	0,766]0 - 0,88]
rs34925694	31259955	G/A	0,133	1	G	167	4,00087.10 ⁻⁰⁶	0,130066	1	G	4,40544.10 ⁻¹¹	2,055 [1,7 - Inf[
rs34505728	31260004	A/C	0,133	1	A	167	4,00087.10 ⁻⁰⁶	0,130066	1	A	4,40542.10 ⁻¹¹	2,055 [1,7 - Inf[
rs35298376	31260040	G/T	0,133	1	G	167	4,00087.10 ⁻⁰⁶	0,130066	1	G	4,40463.10 ⁻¹¹	2,055 [1,7 - Inf[
rs35389342	31260049	C/T	0,133	1	C	167	4,00087.10 ⁻⁰⁶	0,130066	1	C	4,40463.10 ⁻¹¹	2,055 [1,7 - Inf[
rs35235581	31260105	A/G	0,133	1	A	167	4,00087.10 ⁻⁰⁶	0,130066	1	A	4,40463.10 ⁻¹¹	2,055 [1,7 - Inf[
rs35626283	31260168	CT/C	0,133	1	CT	167	4,00087.10 ⁻⁰⁶	0,130066	1	CT	4,40463.10 ⁻¹¹	2,055 [1,7 - Inf[
rs114392121	31260185	C/T	0,133	1	C	167	4,00087.10 ⁻⁰⁶	0,130066	1	C	4,40463.10 ⁻¹¹	2,055 [1,7 - Inf[
rs115042633	31260201	A/T	0,133	1	A	167	4,00087.10 ⁻⁰⁶	0,130066	1	A	4,40463.10 ⁻¹¹	2,055 [1,7 - Inf[
rs9357122	31260345	C/T	0,187	1	T	229	1,58182.10 ⁻⁰⁶	0,20359	0,989	T	0,000745405	0,766]0 - 0,88]
rs117127791	31260398	G/A	0,133	1	G	167	4,00087.10 ⁻⁰⁶	0,130066	1	G	4,40463.10 ⁻¹¹	2,055 [1,7 - Inf[
rs114601215	31260462	G/C	0,133	1	G	167	4,00002.10 ⁻⁰⁶	0,130066	1	G	4,40463.10 ⁻¹¹	2,055 [1,7 - Inf[
rs9394051	31260569	C/G	0,187	1	G	231	1,54218.10 ⁻⁰⁶	0,203612	0,989	G	0,000741785	0,766]0 - 0,88]
rs114489270	31260583	C/T	0,133	1	C	167	4,00002.10 ⁻⁰⁶	0,130066	1	C	4,40462.10 ⁻¹¹	2,055 [1,7 - Inf[

SNP	Position	m/M	Étude primaire					Réplication				
			MAF	info	RA	nfam	P	MAF	info	RA	P	OR [IC 95%]
rs116548532	31260594	G/T	0,133	1	G	167	4,00002.10 ⁻⁰⁶	0,130066	1	G	4,40462.10 ⁻¹¹	2,055 [1,7 - Inf[
rs116537416	31260633	C/T	0,133	1	C	167	4,00002.10 ⁻⁰⁶	0,130067	1	C	4,40383.10 ⁻¹¹	2,055 [1,7 - Inf[
rs9394052	31260708	T/C	0,187	1	C	231	1,51956.10 ⁻⁰⁶	0,203627	0,989	C	0,000739345	0,766]0 - 0,88]
rs55993314	31261056	T/C	0,133	1	T	167	3,99833.10 ⁻⁰⁶	0,130067	1	T	4,40382.10 ⁻¹¹	2,055 [1,7 - Inf[
rs16899178	31261137	T/C	0,133	1	T	167	3,99833.10 ⁻⁰⁶	0,130078	1	T	4,37508.10 ⁻¹¹	2,055 [1,7 - Inf[
rs56409019	31261358	G/C	0,133	1	G	167	3,99833.10 ⁻⁰⁶	0,130067	1	G	4,40383.10 ⁻¹¹	2,055 [1,7 - Inf[
rs55793836	31261393	T/C	0,133	1	T	167	3,99833.10 ⁻⁰⁶	0,130067	1	T	4,40383.10 ⁻¹¹	2,055 [1,7 - Inf[
rs9295969	31261470	A/G	0,188	0,928	G	418	1,46444.10 ⁻⁰⁶	0,203731	0,99	G	0,000719405	0,766]0 - 0,88]
rs34504767	31261492	T/C	0,133	1	T	167	3,99833.10 ⁻⁰⁶	0,130084	1	T	4,32812.10 ⁻¹¹	2,055 [1,7 - Inf[
rs115748458	31261570	G/T	0,133	1	G	167	3,99833.10 ⁻⁰⁶	0,130084	1	G	4,32812.10 ⁻¹¹	2,055 [1,7 - Inf[
rs116134275	31261571	C/T	0,133	1	C	167	3,99833.10 ⁻⁰⁶	0,130084	1	C	4,32812.10 ⁻¹¹	2,055 [1,7 - Inf[
rs35742578	31261645	A/G	0,133	1	A	167	3,99833.10 ⁻⁰⁶	0,130084	1	A	4,32812.10 ⁻¹¹	2,055 [1,7 - Inf[
rs35196124	31261777	T/A	0,133	1	T	167	3,99833.10 ⁻⁰⁶	0,130084	1	T	4,32733.10 ⁻¹¹	2,055 [1,7 - Inf[
rs35933419	31261781	G/A	0,133	1	G	167	3,99833.10 ⁻⁰⁶	0,130084	1	G	4,32733.10 ⁻¹¹	2,055 [1,7 - Inf[
rs114265643	31261851	T/C	0,133	1	T	167	3,99833.10 ⁻⁰⁶	0,130084	1	T	4,32733.10 ⁻¹¹	2,055 [1,7 - Inf[
rs187999352	31261887	T/A	0,133	1	T	167	3,99833.10 ⁻⁰⁶	0,130084	1	T	4,32733.10 ⁻¹¹	2,055 [1,7 - Inf[
rs192233466	31261888	A/T	0,133	1	A	167	3,99833.10 ⁻⁰⁶	0,130084	1	A	4,32733.10 ⁻¹¹	2,055 [1,7 - Inf[
rs2524152	31261929	T/C	0,303	0,993	T	301	3,73601.10 ⁻⁰⁷	0,296316	0,999	T	1,60616.10 ⁻⁰⁸	1,531 [1,35 - Inf[
rs35611056	31261956	G/A	0,133	1	G	167	3,99833.10 ⁻⁰⁶	0,130052	0,999	G	4,28072.10 ⁻¹¹	2,056 [1,7 - Inf[
rs34187356	31261983	C/T	0,133	1	C	167	3,99833.10 ⁻⁰⁶	0,130052	0,999	C	4,28072.10 ⁻¹¹	2,056 [1,7 - Inf[
rs56042166	31262024	G/A	0,133	1	G	167	3,99833.10 ⁻⁰⁶	0,130052	0,999	G	4,28072.10 ⁻¹¹	2,056 [1,7 - Inf[
rs2524150	31262650	A/T	0,301	0,988	A	306	3,7192.10 ⁻⁰⁷	0,296332	0,999	A	1,60888.10 ⁻⁰⁸	1,531 [1,35 - Inf[
rs2524149	31262651	T/C	0,301	0,988	T	306	3,72358.10 ⁻⁰⁷	0,296332	0,999	T	1,60888.10 ⁻⁰⁸	1,531 [1,35 - Inf[
rs189869162	31262677	C/T	0,133	1	C	167	3,99833.10 ⁻⁰⁶	0,130068	1	C	4,39983.10 ⁻¹¹	2,055 [1,7 - Inf[

SNP	Position	m/M	Étude primaire					Réplication				
			MAF	info	RA	nfam	P	MAF	info	RA	P	OR [IC 95%]
rs115667045	31262717	A/T	0,133	1	A	167	3,99833.10 ⁻⁰⁶	0,130068	1	A	4,39982.10 ⁻¹¹	2,055 [1,7 - Inf[
rs114899495	31262761	T/C	0,133	1	T	167	3,99833.10 ⁻⁰⁶	0,130068	1	T	4,39982.10 ⁻¹¹	2,055 [1,7 - Inf[
rs116654277	31262769	T/C	0,133	1	T	167	3,99833.10 ⁻⁰⁶	0,130068	1	T	4,39982.10 ⁻¹¹	2,055 [1,7 - Inf[
rs2524147	31262781	A/G	0,3	0,987	A	306	3,71854.10 ⁻⁰⁷	0,29637	0,999	A	1,61089.10 ⁻⁰⁸	1,531 [1,35 - Inf[
rs2524146	31262783	C/T	0,3	0,987	C	307	3,7245.10 ⁻⁰⁷	0,29637	0,999	C	1,61089.10 ⁻⁰⁸	1,531 [1,35 - Inf[
rs116405890	31262870	T/C	0,133	1	T	167	3,99833.10 ⁻⁰⁶	0,130068	1	T	4,39982.10 ⁻¹¹	2,055 [1,7 - Inf[
rs115219503	31262959	C/T	0,133	1	C	167	3,99833.10 ⁻⁰⁶	0,130085	1	C	4,38813.10 ⁻¹¹	2,055 [1,7 - Inf[
rs115973915	31263349	A/C	0,133	1	A	167	4,00378.10 ⁻⁰⁶	0,130068	1	A	4,39821.10 ⁻¹¹	2,055 [1,7 - Inf[
rs114016160	31263410	A/G	0,133	1	A	167	4,00463.10 ⁻⁰⁶	0,130069	1	A	4,39742.10 ⁻¹¹	2,055 [1,7 - Inf[
rs115101855	31263482	T/A	0,133	1	T	167	4,00447.10 ⁻⁰⁶	0,130069	1	T	4,39742.10 ⁻¹¹	2,055 [1,7 - Inf[
rs114601085	31263548	T/C	0,133	1	T	167	4,00583.10 ⁻⁰⁶	0,130069	1	T	4,39742.10 ⁻¹¹	2,055 [1,7 - Inf[
rs114294380	31263601	G/T	0,133	1	G	167	4,00702.10 ⁻⁰⁶	0,130069	1	G	4,39742.10 ⁻¹¹	2,055 [1,7 - Inf[
rs200307568	31263721	T/G	0,133	1	T	167	4,0079.10 ⁻⁰⁶	0,130069	1	T	4,39743.10 ⁻¹¹	2,055 [1,7 - Inf[
rs114583506	31263801	T/G	0,133	1	T	167	4,00927.10 ⁻⁰⁶	0,130069	1	T	4,39742.10 ⁻¹¹	2,055 [1,7 - Inf[
rs116165038	31263852	C/G	0,136	0,99	C	181	6,72289.10 ⁻⁰⁶	0,130097	1	C	4,39803.10 ⁻¹¹	2,055 [1,7 - Inf[
rs115865502	31264178	T/C	0,133	1	T	167	4,01408.10 ⁻⁰⁶	0,130069	1	T	4,39582.10 ⁻¹¹	2,055 [1,7 - Inf[
rs115228090	31264284	T/C	0,132	0,996	T	170	3,89798.10 ⁻⁰⁶	0,127262	0,982	T	5,62565.10 ⁻¹¹	2,073 [1,71 - Inf[
rs115805679	31264288	C/T	0,132	0,996	C	170	3,89999.10 ⁻⁰⁶	0,127214	0,982	C	5,4538.10 ⁻¹¹	2,075 [1,71 - Inf[
rs115494471	31264408	T/C	0,133	1	T	167	4,01494.10 ⁻⁰⁶	0,130069	1	T	4,39583.10 ⁻¹¹	2,055 [1,7 - Inf[
rs114973736	31264497	C/T	0,133	1	C	167	4,01494.10 ⁻⁰⁶	0,130069	1	C	4,39583.10 ⁻¹¹	2,055 [1,7 - Inf[
rs115964138	31264512	A/G	0,133	1	A	167	4,01494.10 ⁻⁰⁶	0,130069	1	A	4,39583.10 ⁻¹¹	2,055 [1,7 - Inf[
rs116334679	31264517	C/T	0,133	1	C	167	4,01494.10 ⁻⁰⁶	0,130069	1	C	4,39583.10 ⁻¹¹	2,055 [1,7 - Inf[
rs115700453	31264577	C/T	0,133	1	C	167	4,01494.10 ⁻⁰⁶	0,130069	1	C	4,39582.10 ⁻¹¹	2,055 [1,7 - Inf[
rs200436274	31264580	T/TC	0,133	1	T	167	4,01494.10 ⁻⁰⁶	0,130069	1	T	4,39503.10 ⁻¹¹	2,055 [1,7 - Inf[

SNP	Position	m/M	Étude primaire					Réplication				
			MAF	info	RA	nfam	P	MAF	info	RA	P	OR [IC 95%]
rs574228159	31264583	T/G	0,133	1	T	167	4,01494.10 ⁻⁰⁶	0,130069	1	T	4,39503.10 ⁻¹¹	2,055 [1,7 - Inf[
rs9366777	31264715	T/A	0,188	1	A	229	2,65474.10 ⁻⁰⁶	0,205846	0,988	T	0,5	ns
rs114770325	31264845	C/T	0,133	1	C	167	4,01371.10 ⁻⁰⁶	0,13007	1	C	4,39264.10 ⁻¹¹	2,055 [1,7 - Inf[
rs114904826	31264960	C/T	0,133	1	C	167	4,01097.10 ⁻⁰⁶	0,13007	1	C	4,39264.10 ⁻¹¹	2,055 [1,7 - Inf[
rs115753074	31265168	A/G	0,133	1	A	167	4,00754.10 ⁻⁰⁶	0,13007	1	A	4,39324.10 ⁻¹¹	2,055 [1,7 - Inf[
rs115975714	31265242	G/T	0,133	1	G	167	4,00617.10 ⁻⁰⁶	0,130071	1	G	4,38807.10 ⁻¹¹	2,055 [1,7 - Inf[
rs17198811	31265634	T/C	0,133	1	T	167	4,00275.10 ⁻⁰⁶	0,130072	1	T	4,38595.10 ⁻¹¹	2,055 [1,7 - Inf[
rs17198818	31265695	A/G	0,133	1	A	167	4,00139.10 ⁻⁰⁶	0,130073	1	A	4,38258.10 ⁻¹¹	2,055 [1,7 - Inf[
rs3828860	31265736	T/C	0,133	1	T	167	4,00139.10 ⁻⁰⁶	0,130073	1	T	4,3826.10 ⁻¹¹	2,055 [1,7 - Inf[
rs17192470	31265743	T/C	0,133	1	T	167	4,00275.10 ⁻⁰⁶	0,130072	1	T	4,38339.10 ⁻¹¹	2,055 [1,7 - Inf[
rs17192491	31265788	G/C	0,133	1	G	167	4,00002.10 ⁻⁰⁶	0,130073	1	G	4,37889.10 ⁻¹¹	2,055 [1,7 - Inf[
rs17198839	31265807	C/A	0,133	1	C	167	4,00002.10 ⁻⁰⁶	0,130073	1	C	4,37889.10 ⁻¹¹	2,055 [1,7 - Inf[
rs9348864	31265855	C/T	0,17	0,998	T	269	3,00517.10 ⁻⁰⁶	0,190582	0,996	T	0,001920905	0,78]0 - 0,9]
rs17198860	31265875	A/G	0,133	1	A	167	4,00002.10 ⁻⁰⁶	0,130074	1	A	4,37294.10 ⁻¹¹	2,055 [1,7 - Inf[
rs148136265	31265884	A/ACAAT	0,17	0,999	ACAAT	261	3,00195.10 ⁻⁰⁶	0,190855	0,996	ACAAT	0,00164068	0,776]0 - 0,89]
rs17192498	31265887	G/A	0,138	0,983	G	254	6,55644.10 ⁻⁰⁶	0,138722	0,948	G	1,37604.10 ⁻¹⁰	2,01 [1,66 - Inf[
rs17192505	31265910	G/T	0,133	1	G	167	4,00172.10 ⁻⁰⁶	0,130091	1	G	4,31903.10 ⁻¹¹	2,055 [1,7 - Inf[
rs17192512	31265914	G/A	0,133	1	G	167	4,00002.10 ⁻⁰⁶	0,130074	1	G	4,37245.10 ⁻¹¹	2,055 [1,7 - Inf[
rs17192533	31266085	G/C	0,135	1	G	166	5,13922.10 ⁻⁰⁶	0,132485	0,994	G	2,87749.10 ⁻¹⁰	1,976 [1,64 - Inf[
rs17198909	31266189	T/A	0,133	0,997	T	166	5,45889.10 ⁻⁰⁶	0,129273	0,995	T	5,70035.10 ⁻¹¹	2,053 [1,69 - Inf[
rs138758697	31266310	A/AG	0,133	1	A	167	4,00002.10 ⁻⁰⁶	0,130074	1	A	4,37559.10 ⁻¹¹	2,055 [1,7 - Inf[
rs16899203	31266335	C/T	0,133	1	C	167	4,00002.10 ⁻⁰⁶	0,130073	1	C	4,37718.10 ⁻¹¹	2,055 [1,7 - Inf[
rs139795635	31266411	TA/T	0,133	1	TA	167	4,00002.10 ⁻⁰⁶	0,130074	1	TA	4,37304.10 ⁻¹¹	2,055 [1,7 - Inf[
rs115491491	31266465	G/A	0,133	1	G	167	4,00002.10 ⁻⁰⁶	0,130074	1	G	4,37064.10 ⁻¹¹	2,055 [1,7 - Inf[

SNP	Position	m/M	Étude primaire					Réplication				
			MAF	info	RA	nfam	P	MAF	info	RA	P	OR [IC 95%]
rs116364562	31266479	C/G	0,133	1	C	167	4,00002.10 ⁻⁰⁶	0,130074	1	C	4,37064.10 ⁻¹¹	2,055 [1,7 - Inf[
rs201641102	31267776	G/A	0,133	1	G	167	3,99833.10 ⁻⁰⁶	0,13008	0,999	G	4,33859.10 ⁻¹¹	2,055 [1,7 - Inf[
rs116647350	31268833	A/G	0,133	1	A	167	3,99833.10 ⁻⁰⁶	0,130079	0,999	A	4,33426.10 ⁻¹¹	2,055 [1,7 - Inf[
rs12663075	31269256	A/C	0,133	1	A	167	3,99833.10 ⁻⁰⁶	0,130082	0,999	A	4,33811.10 ⁻¹¹	2,055 [1,7 - Inf[
rs56138398	31269299	C/G	0,133	1	C	167	3,99833.10 ⁻⁰⁶	0,130081	0,999	C	4,32352.10 ⁻¹¹	2,056 [1,7 - Inf[
rs3915970	31269460	C/T	0,135	1	C	166	5,15058.10 ⁻⁰⁶	0,132603	0,993	C	2,82343.10 ⁻¹⁰	1,977 [1,64 - Inf[
rs2854007	31313586	G/A	0,134	1	G	167	8,10524.10 ⁻⁰⁶	0,134388	0,989	G	2,24385.10 ⁻¹⁰	1,962 [1,63 - Inf[
rs2854006	31313591	A/G	0,134	1	A	167	8,10524.10 ⁻⁰⁶	0,134387	0,989	A	2,24389.10 ⁻¹⁰	1,962 [1,63 - Inf[
rs2596526	31313724	C/T	0,134	0,999	C	167	8,15884.10 ⁻⁰⁶	0,132264	0,997	C	9,99385.10 ⁻¹¹	1,997 [1,66 - Inf[
rs2596524	31313906	A/G	0,134	1	A	167	8,10206.10 ⁻⁰⁶	0,132368	0,998	A	1,09354.10 ⁻¹⁰	1,993 [1,65 - Inf[
rs7453967	31314243	G/T	0,134	1	G	165	8,26057.10 ⁻⁰⁶	0,132291	0,997	G	1,04372.10 ⁻¹⁰	1,995 [1,65 - Inf[
rs4475359	31314298	T/G	0,136	1	T	164	8,91404.10 ⁻⁰⁶	0,136445	0,987	T	5,2197.10 ⁻¹⁰	1,926 [1,6 - Inf[
rs9265929	31314313	C/T	0,134	1	C	166	8,13151.10 ⁻⁰⁶	0,132359	0,998	C	1,10244.10 ⁻¹⁰	1,992 [1,65 - Inf[
rs9265930	31314347	A/C	0,134	1	A	166	8,13151.10 ⁻⁰⁶	0,132359	0,998	A	1,10298.10 ⁻¹⁰	1,992 [1,65 - Inf[
rs2508000	31314588	T/C	0,134	1	T	166	8,13151.10 ⁻⁰⁶	0,132354	0,998	T	1,10754.10 ⁻¹⁰	1,992 [1,65 - Inf[
rs9265950	31315261	C/A	0,134	0,999	C	178	7,4791.10 ⁻⁰⁶	0,134039	0,989	C	2,58221.10 ⁻¹⁰	1,96 [1,63 - Inf[
rs9265959	31315425	T/C	0,134	1	T	166	8,13151.10 ⁻⁰⁶	0,132336	0,998	T	1,12515.10 ⁻¹⁰	1,992 [1,65 - Inf[
rs9265963	31315540	A/G	0,134	1	A	166	8,13151.10 ⁻⁰⁶	0,132334	0,998	A	1,12709.10 ⁻¹⁰	1,991 [1,65 - Inf[
rs2596520	31316481	T/C	0,134	0,999	T	166	8,31101.10 ⁻⁰⁶	0,132288	0,998	T	1,1321.10 ⁻¹⁰	1,992 [1,65 - Inf[
rs2596519	31316563	T/A	0,134	0,999	T	166	8,31101.10 ⁻⁰⁶	0,132289	0,998	T	1,13392.10 ⁻¹⁰	1,992 [1,65 - Inf[
rs2596518	31316759	G/A	0,134	1	G	166	8,13487.10 ⁻⁰⁶	0,132313	0,998	G	1,14603.10 ⁻¹⁰	1,991 [1,65 - Inf[
rs2523620	31317514	C/A	0,134	1	C	167	8,11875.10 ⁻⁰⁶	0,132311	0,998	C	1,15886.10 ⁻¹⁰	1,991 [1,65 - Inf[
rs2596515	31317619	A/G	0,134	1	A	167	8,11061.10 ⁻⁰⁶	0,13231	0,998	A	1,16083.10 ⁻¹⁰	1,991 [1,65 - Inf[
rs2596513	31317762	G/A	0,134	1	G	167	8,09789.10 ⁻⁰⁶	0,132309	0,997	G	1,16282.10 ⁻¹⁰	1,991 [1,65 - Inf[

SNP	Position	m/M	Étude primaire					Réplication				
			MAF	info	RA	nfam	P	MAF	info	RA	P	OR [IC 95%]
rs2596511	31317855	T/A	0,134	1	T	169	8,08291.10 ⁻⁰⁶	0,13237	0,997	T	1,23237.10 ⁻¹⁰	1,989 [1,65 - Inf[
rs2523618	31318303	C/T	0,134	0,999	C	171	7,82489.10 ⁻⁰⁶	0,132286	0,996	C	1,21028.10 ⁻¹⁰	1,991 [1,65 - Inf[
rs2442724	31319907	C/T	0,134	0,998	C	177	5,98684.10 ⁻⁰⁶	0,132108	0,985	C	1,77509.10 ⁻¹⁰	1,985 [1,64 - Inf[
rs1065386	31324547	C/G	0,442	0,945	G	421	3,87273.10 ⁻⁰⁷	0,442768	0,83	G	0,00413253	0,82]0 - 0,93]
rs1050570	31324549	C/T	0,442	0,945	T	421	3,8725.10 ⁻⁰⁷	0,442768	0,83	T	0,004132705	0,82]0 - 0,93]
rs2523656	31444048	C/G	0,469	0,995	G	330	9,65793.10 ⁻⁰⁶	0,461882	0,949	G	0,0169342	0,863]0 - 0,97]
rs2523647	31449778	G/A	0,457	0,994	A	339	2,61139.10 ⁻⁰⁶	0,452727	0,958	A	0,00641935	0,842]0 - 0,94]
rs78485683	31456691	CTT/CT	0,394	0,987	CT	358	9,15212.10 ⁻⁰⁶	0,401221	0,976	CT	0,134367	0,925]0 - 1,04]
rs10661543	31462605	CAAAA/C	0,188	0,993	C	259	7,77775.10 ⁻⁰⁶	0,206645	0,929	CAAAA	0,5	ns
rs375883533	31470020	A/G	0,083	0,978	A	111	9,74943.10 ⁻⁰⁶	0,0685344	0,926	A	1,35607.10 ⁻⁰⁹	2,568 [1,94 - Inf[
rs3830668	31502882	A/ATGT	0,181	0,992	ATGT	254	5,22218.10 ⁻⁰⁶	0,199162	0,967	ATGT	0,4264245	0,984]0 - 1,13]
rs61523215	32415606	A/AT	0,176	0,981	AT	241	5,09436.10 ⁻⁰⁶	0,16619	0,857	AT	0,00476605	0,778]0 - 0,91]
rs116673414	32593097	G/T	0,151	0,978	G	217	7,9479.10 ⁻⁰⁶	0,145342	0,938	G	7,2913.10 ⁻⁰⁹	1,828 [1,52 - Inf[
rs35675538	32593127	T/C	0,151	0,978	T	217	8,18407.10 ⁻⁰⁶	0,147271	0,945	T	1,38032.10 ⁻⁰⁸	1,796 [1,5 - Inf[
rs9272346	32604372	G/A	0,469	0,904	G	430	5,47794.10 ⁻⁰⁶	0,429895	0,86	G	6,5357.10 ⁻¹³	1,71 [1,51 - Inf[
rs3187964	32605207	C/G	0,457	0,916	C	419	1,28095.10 ⁻⁰⁶	0,426129	0,863	C	5,2794.10 ⁻¹³	1,717 [1,51 - Inf[
rs707951	32609813	T/C	0,313	0,976	T	343	3,37407.10 ⁻⁰⁶	0,274689	0,841	T	2,96954.10 ⁻¹⁰	1,696 [1,47 - Inf[
rs9272776	32610270	T/C	0,349	0,982	T	366	6,55514.10 ⁻⁰⁶	0,309258	0,866	T	3,77476.10 ⁻¹⁰	1,649 [1,44 - Inf[
rs78570891	32611590	G/A	0,365	0,946	G	427	2,73452.10 ⁻⁰⁶	0,338893	0,806	G	3,66137.10 ⁻⁰⁹	1,605 [1,4 - Inf[
rs9273508	32628439	A/G	0,295	0,974	G	378	7,2503.10 ⁻⁰⁶	0,324356	0,897	G	0,000165505	0,763]0 - 0,86]
rs17212014	32637472	T/C	0,206	0,935	C	365	6,13472.10 ⁻⁰⁷	0,222999	0,826	C	0,0401682	0,857]0 - 0,99]
rs28371211	32637528	A/C	0,209	0,941	C	363	3,47303.10 ⁻⁰⁷	0,225738	0,835	C	0,02366855	0,841]0 - 0,97]

Tableau 27 : Résultats d'association de l'étude primaire familiale et de la réplication pour les 358 SNPs HLA dépassant le seuil évocateur

RA = allèle à risque ; Les positions sont données en fonction de hg19 ; OR est donné par rapport à l'allèle mineur ; Pour la réplication, il s'agit de tests unilatéraux.

ANNEXE 3 : Article 1

Family-based genome-wide association study of leprosy in Vietnam

Chaïma Gzara, Monica Dallmann-Sauer, Marianna Orlova,
Nguyen Van Thuc, Vu Hong Thai, Vinicius M Fava,
Marie-Thérèse Bihoreau, Anne Boland, Laurent Abel,
Alexandre Alcaïs, Erwin Schurr, Aurélie Cobat

Publié dans PLoS Pathogen,

2020 May 18;16(5):e1008565.

doi: 10.1371/journal.ppat.1008565.

RESEARCH ARTICLE

Family-based genome-wide association study of leprosy in Vietnam

Chaima Gzara^{1,2}, Monica Dallmann-Sauer^{3,4,5}, Marianna Orlova^{3,4,5}, Nguyen Van Thuc⁶, Vu Hong Thai⁶, Vinicius M. Fava^{3,4}, Marie-Thérèse Bihoreau⁷, Anne Boland⁷, Laurent Abel^{1,2,8}, Alexandre Alcaïs^{1,2‡}, Erwin Schurr^{3,4,5‡}, Aurélie Cobat^{1,2‡*}

1 Laboratory of Human Genetics of Infectious Diseases, Necker Branch, INSERM UMR1163, Paris, France, **2** Université de Paris, Imagine Institute, Paris, France, **3** McGill International TB Centre, Montreal, QC, Canada, **4** Program in Infectious Diseases and Immunity in Global Health, The Research Institute of the McGill University Health Centre, Montreal, QC, Canada, **5** Department of Medicine and Human Genetics, Faculty of Medicine, McGill University, Montreal, QC, Canada, **6** Hospital for Dermato-Venereology, District, Ho Chi Minh City, Vietnam, **7** Université Paris-Saclay, CEA, Centre National de Recherche en Génomique Humaine, Evry, France, **8** St. Giles Laboratory of Human Genetics of Infectious Diseases, Rockefeller Branch, The Rockefeller University, New York, United States of America

☞ These authors contributed equally to this work.

‡ These authors are joint senior authors on this work

* aurelie.cobat@inserm.fr



OPEN ACCESS

Citation: Gzara C, Dallmann-Sauer M, Orlova M, Van Thuc N, Thai VH, Fava VM, et al. (2020) Family-based genome-wide association study of leprosy in Vietnam. *PLoS Pathog* 16(5): e1008565. <https://doi.org/10.1371/journal.ppat.1008565>

Editor: Thomas R. Hawn, University of Washington, UNITED STATES

Received: February 4, 2020

Accepted: April 20, 2020

Published: May 18, 2020

Copyright: © 2020 Gzara et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Data Availability Statement: All relevant data are within the manuscript and its Supporting Information files.

Funding: This work was supported by grants from the Canadian Institutes of Health Research (CIHR; FDN-143332) to ES, MALTALP from l'Ordre de Malte to AA and ES, and the Agence Nationale de la Recherche (ANR) to AA. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing interests: The authors have declared that no competing interests exist.

Abstract

Leprosy is a chronic infectious disease of the skin and peripheral nerves with a strong genetic predisposition. Recent genome-wide approaches have identified numerous common variants associated with leprosy, almost all in the Chinese population. We conducted the first family-based genome-wide association study of leprosy in 622 affected offspring from Vietnam, followed by replication in an independent sample of 1181 leprosy cases and 668 controls of the same ethnic origin. The most significant results were observed within the HLA region, in which six SNPs displayed genome-wide significant associations, all of which were replicated in the independent case/control sample. We investigated the signal in the HLA region in more detail, by conducting a multivariate analysis on the case/control sample of 319 GWAS-suggestive HLA hits for which evidence for replication was obtained. We identified three independently associated SNPs, two located in the HLA class I region (rs1265048: OR = 0.69 [0.58–0.80], combined p -value = 5.53×10^{-11} ; and rs114598080: OR = 1.47 [1.46–1.48], combined p -value = 8.77×10^{-13}), and one located in the HLA class II region (rs3187964 (OR = 1.67 [1.55–1.80], combined p -value = 8.35×10^{-16}). We also validated two previously identified risk factors for leprosy: the missense variant rs3764147 in the *LACC1* gene (OR = 1.52 [1.41–1.63], combined p -value = 5.06×10^{-14}), and the intergenic variant rs6871626 located close to the *IL12B* gene (OR = 0.73 [0.61–0.84], combined p -value = 6.44×10^{-8}). These results shed new light on the genetic control of leprosy, by dissecting the influence of HLA SNPs, and validating the independent role of two additional variants in a large Vietnamese sample.

Author summary

Due to its extreme reductive evolution *Mycobacterium leprae* is a rare example of an essentially clonal bacterial pathogen that affects humans. However, exposed individuals display a wide diversity of symptoms reflecting the interactions between the host immune response and the bacterium. There is now accumulating evidence, in particular from genome-wide association study (GWAS), that common genetic variants play a role explaining this variability. Previous GWAS were based on case control design and almost all of them were conducted in the Chinese population. We conducted the first family-based GWAS of leprosy in the Vietnamese population and identified several genome-wide significant association signals within the HLA region. By performing a multivariate analysis in an independent case-control sample of the same ethnic origin we were able to decipher the HLA association signal and to reduce it to three independent SNPs, two in the class I and one in the class II region. We also validated two previously identified risk factors for leprosy, a missense variant in the *LACC1* gene, and an intergenic variant located close to the *IL12B* gene. These results shed new light on the genetic control of leprosy and, in particular, on the influence of HLA SNPs in a large Vietnamese sample.

Introduction

Leprosy is a chronic infectious disease caused by either *Mycobacterium leprae* or *Mycobacterium lepromatosis*. It primarily affects the skin and peripheral nerves, and can cause an irreversible impairment of nerve function, often leading to severe disabilities and social stigma. Despite a decrease in its prevalence over the last two decades, leprosy remains a major public health problem in regions of endemic countries, with over 200,000 new cases detected in 2018 (https://www.who.int/gho/neglected_diseases/leprosy/en/). However, this number of cases is probably a severe underestimate of the true incidence [1]. The clinical disease develops in a minority of exposed individuals, manifesting as a spectrum of disease symptoms reflecting the interactions between the host immune response and the bacterium [2]. Tuberculoid and lepromatous leprosy are at opposite ends of the clinical spectrum and are associated with a relatively stable host immune status while borderline categories of the disease are associated with an unstable immune response to the bacilli.

Only a subset of individuals develops clinical leprosy after sustained exposure to *M. leprae*. Due to the extreme reductive evolution of the *M. leprae* genome [3], it is highly unlikely that differences in susceptibility or clinical manifestations are governed by the *M. leprae* strain or by intrastrain variation [4]. From the early observations of familial clusters of leprosy cases to recent whole-exome sequencing studies identifying genetic variants associated with leprosy, there is strong evidence to suggest that the development of this disease is under tight human genetic control [2, 5]. Genetic studies, including positional cloning analyses [6–8], genome-wide association studies [9–15], and, more recently, whole-exome sequencing [16] have identified several susceptibility loci for leprosy (reviewed in [17] and [5]), and have demonstrated the involvement of both innate and adaptive immune responses in this disease [18]. Almost all the susceptibility loci identified in genome-wide studies to date were found in the Chinese population. The association with leprosy has been replicated in the Vietnamese population for some of these loci [19, 20], but others were found to be associated with type-1 reactions (T1Rs), a pathological inflammatory response afflicting a subgroup of leprosy patients and resulting in peripheral nerve damage [21–23], rather than with leprosy itself. We previously investigated susceptibility factors for T1Rs in Vietnamese families by GWAS and identified an

eQTL (rs1875147) for a lncRNA gene as a global risk-factor for T1R but not for leprosy [24]. Here, we performed a two-stage family-based genome-wide association study on leprosy *per se*, i.e. leprosy independent of its clinical subtype, in a large collection of Vietnamese families including 622 affected offspring, followed by a replication study in an independent sample of 1181 leprosy cases and 668 controls.

Materials and methods

Ethics statement

The study was conducted in accordance with the Declaration of Helsinki. Written informed consent was obtained from all adult subjects participating in the study. All minors agreed to take part, and a parent or guardian provided informed consent on their behalf. The study was approved by the Research Ethics Board at the Research Institute at McGill University Health Centre in Montreal (REC98-041) and the regulatory authorities and ethics committees of Ho Chi Minh City in Vietnam (So3813/UB-VX and 4933/UBND-VX).

Samples and study design

Over twenty-five years (1990–2015), we have, in close collaboration with the Dermatology and Venereal Diseases Hospital of Ho-Chi-Minh City (Vietnam), recruited a large sample of nuclear families with at least one child diagnosed with leprosy and including either both parents or unaffected sibling(s) if one of the parents was unavailable. Leprosy was diagnosed based on a committee decision involving at least two independent and experienced physicians and was classified into multibacillary (MB) or paucibacillary (PB) subtype according to the operational WHO-96 definition based solely on the number of lesions [2]. All patients responded to therapy which was a strong argument in favor of the leprosy diagnosis. For replication purposes, we collected an independent case/control sample of Vietnamese origin. The same criteria were used for leprosy diagnosis as in the family-based discovery study. The controls had no personal or family (among first-, second- or third-degree relatives) history of leprosy or tuberculosis.

GWAS genotyping and imputation

The genotypes of all children and parents (or siblings) were determined with the Illumina Human 660w Quad v1 bead chip containing 592,633 single-nucleotide polymorphisms (SNPs) during the discovery phase. Genotypes were called with the Illuminus algorithm [25]. Quality control (QC) was performed on genotype data with PLINK software [26] and the “GASTON” package in R (<https://CRAN.R-project.org/package=gaston>). SNPs with a call-rate < 0.95, more than 10 Mendelian errors, a minor allele frequency (MAF) < 0.05 or displaying significant departure from Hardy-Weinberg equilibrium ($p < 10^{-5}$) were removed from the analyses, resulting in a final set of 422,546 high-quality SNPs, including 9,907 SNPs on the X chromosome. Individuals with a call-rate < 0.95, more than 10 Mendelian errors and a heterozygosity rate more than three standard deviations on either side of the mean were excluded. In total, 1850 individuals were genotyped and 1749 fulfilled the quality control criteria, including 622 leprosy-affected offspring, from 481 informative families (S1 Table).

Before imputation in the discovery sample, A/T or G/C SNPs were removed to prevent strand mismatches between the study sample and the reference sample used for imputation. The remaining high-quality SNPs were pre-phased with SHAPEIT [27] v2 software using to the duoHMM method, to incorporate the known pedigree information and improve phasing [28], except for the X chromosome for which duoHMM is not applicable. Genotypes from the

1000 Genomes Project phase 3 haplotype reference panel released in October 2014 were further imputed with IMPUTE2 software [29]. An information imputation threshold (Info) > 0.9 was applied to capture most of the common variants (MAF > 5%) with reasonable confidence, leading to the retention of ~5 million additional variants (SNPs and INDELs).

Replication strategy

We used three different strategies to select SNPs for genotyping in the replication sample, according to the type of variant. First, for genome-wide suggestive (p -value < 10^{-5}) hits located outside the HLA region (non-HLA SNPs), we used a linkage disequilibrium (LD)-based clumping procedure, implemented in PLINK 1.9 [26, 30], to identify independent loci. This procedure considers all SNPs with p -values for association below 10^{-5} , and forms clumps of these 'index' SNPs, together with all other SNPs in LD with them at an r^2 threshold of 0.5 and in close physical proximity (less than 5 Mb away from the index SNPs). Only SNPs with a p -value below a secondary significance threshold of 10^{-4} are clumped. As input for the clumping procedure, we used association p -values from the discovery GWAS and LD patterns estimated from the Kinh in Ho Chi Minh City, Vietnam (KHV) population of the 1000 Genomes phase 3 reference set. We selected one SNP from each clump for genotyping in the replication sample, based on technical considerations. We applied a second strategy specifically to the non-HLA SNPs previously associated with leprosy at a genome-wide level of significance (S2 Table). For these SNPs, we used a 10^{-2} p -value threshold to select the SNPs for the replication study. Finally, for SNPs located within the HLA region, we used a third strategy, because there were large numbers of SNPs displaying suggestive evidence of association, including genome-wide significant (p -value < 5×10^{-8}) results, and complex LD patterns. We selected a set of 88 SNPs located at positions from 30 Mb to 33 Mb on chromosome 6, on the basis of preliminary association results, LD structure and technical considerations for genotyping, and we genotyped these SNPs in the replication sample. We then used these SNPs to impute, in the replication sample, all the variants for which there was genome-wide suggestive evidence of association.

For the replication sample, genotypes were determined by FLUIDIGM targeted sequencing or with SEQUENOM technology. For FLUIDIGM data, we used the Genome Analysis Toolkit [31] (GATK v3.3) to process the bam files. Individual genomic variant call files (gVCF) were generated with GATK HaplotypeCaller, and joint genotyping was performed with GATK GenotypeGVCFs. SNPs with a call-rate < 0.95 or significant departure from Hardy-Weinberg equilibrium ($p < 0.001$) were removed from the analyses. Individuals with a call-rate < 0.95 were excluded. Imputation in the MHC region was performed with SHAPEIT v2 [27] and IMPUTE2 [29] software and the 1000 Genomes Project phase 3 haplotype reference panel.

Statistical methods

In the discovery phase, a family-based association test was used to estimate the non-random transmission of alleles from heterozygous parents to leprosy-affected offspring. For autosomal SNPs the analysis was carried out under the additive model, with FBATdosage v2.6 to account for the post-imputation genotype uncertainty [19]. For SNPs on the X chromosome, the association with leprosy was carried out under the additive model with FBAT v2.0.4 [32]. Prior to association testing with FBAT, the imputed genotype probabilities for the X chromosomal SNPs were converted into best-guess genotypes using a probability threshold of 0.9. Imputed SNPs with call-rate below 99% or showing departure from Hardy Weinberg equilibrium ($p < 10^{-3}$) after best-guess transformation were excluded. Alleles for which some evidence of association was obtained were also analyzed by conditional logistic regression, as previously

described [33–35]. Briefly, for each affected child, up to three matched unaffected pseudosibs were formed by all other possible combinations of parental alleles. The real affected offspring were compared with the pseudosibs in a matched case-control design. This approach generated odds ratio estimates and made it possible to perform multivariate stepwise regression. Conditional logistic regression was also used for a combined analysis of the discovery family and replication case-control samples. In this combined analysis, all cases and controls from the case-control study were grouped together in a stratum for analysis, in combination with the strata consisting of the affected offspring and their pseudosibs, as previously described [34, 35]. Conditional logistic analysis was performed with the LOGISTIC procedure implemented in SAS software v. 9.3 (SAS Institute, Cary, North Carolina, USA).

Case/control association analyses of genotyped and imputed SNPs in the replication sample were performed by logistic regression analysis, under the additive model, with SNPTTEST software [29, 36]. A one-tailed test was performed with the alternative hypothesis that the leprosy susceptibility alleles of the discovery sample were also susceptibility factors in the replication sample, and a p -value threshold of 0.01 was used in this one-tailed test for the detection of significant replication. We also performed a multivariable analysis of the HLA SNPs replicated in the case/control replication sample, using the best subset selection method of the SAS 9.3 logistic regression procedure.

Results

Genome-Wide Analyses in the Primary Cohort

We carried out a family-based GWAS on leprosy *per se* in 481 informative families consisting of 1749 individuals, including 622 offspring with leprosy (S1 Table). For the discovery phase, 5,607,170 high-quality genotyped and imputed autosomal variants were analyzed. Principal component analysis of our samples with data for individuals from the 1000 Genomes database revealed clustering with the 1000 Genomes KHV population (S1 Fig). A Manhattan plot of the association with leprosy *per se* is shown in Fig 1. The most significant results were obtained for the HLA region, in which genome-wide significant association with leprosy was detected for six SNPs (p -value $< 5 \times 10^{-8}$) (Table 1), and genome-wide suggestive association with leprosy was detected for 358 SNPs (p -value $< 10^{-5}$). Outside the HLA region, there was no significant deviation from expectations on the quantile-quantile plot of the GWAS results (S2 Fig). In total, 21 SNPs from eight independent clusters (S2 Table) on chromosomes 4, 5, 9, 10, 11, 14 and 21 (2 clusters) displayed genome-wide suggestive association with leprosy (p -value $< 10^{-5}$). Finally, among the 33 genome-wide significant previously published non-HLA hits, evidence of association ($p < 10^{-2}$) with the same risk allele as in the original study was found for five of these SNPs in the Vietnamese discovery sample (S3 Table).

Replication study

Outside the MHC region. Within each of the 8 loci for which suggestive association with leprosy was detected, we genotyped one SNP in an independent replication cohort of 1189 leprosy cases and 674 controls. We also genotyped the 5 previously published non-HLA hits replicated in the discovery cohort with a p -value threshold of 10^{-2} . In total, 10 of the 13 SNPs satisfying the quality control filters in the replication cohort were tested for association in the 1181 leprosy cases and 668 controls. No evidence of replication was obtained for any of the suggestive GWAS hits that could be tested in the replication sample (S2 Table). For two suggestive GWAS hits, on chromosomal region 4q26 and 9q31.1, the genotyping failed in the replication sample. Replication was observed for two previously published hits fulfilling the genotyping quality criteria (Table 1 and S3 Table). The first, rs3764147, located on

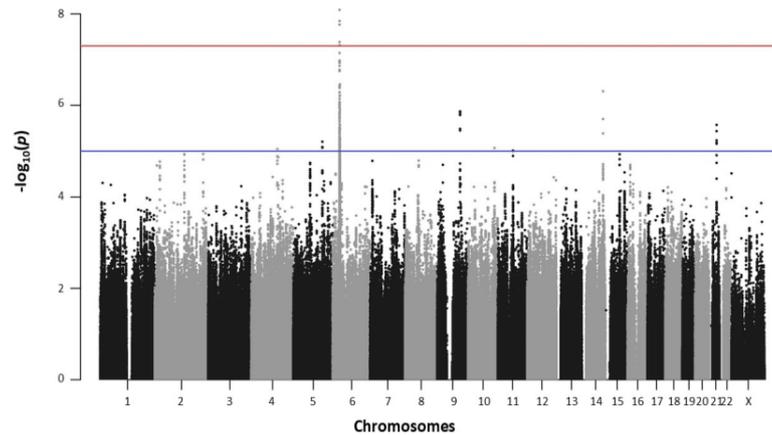


Fig 1. Manhattan plot of the results for association with leprosy *per se* in the family-based GWAS. Manhattan plot showing results of the family-based genome-wide association study of leprosy *per se* in 622 affected offspring, for 5,607,170 variants. The $-\log_{10}(p\text{-value})$ for each variant (y-axis) is plotted according to its chromosomal position (x-axis, build hg19). The red and blue lines indicate the genome-wide significance (5.0×10^{-8}) and suggestive (10^{-5}) thresholds, respectively.

<https://doi.org/10.1371/journal.ppat.1008565.g001>

chromosome 13, is a missense variant (p.Ile254Val) of the *LACCI* gene with a highly significant genome-wide *p*-value in the combined analysis of the discovery family and the replication case/control samples ($p = 5.1 \times 10^{-14}$). The odds ratio (OR) for developing leprosy for AG subjects vs. GG subjects (or AA vs. AG) was 1.52 [95% confidence interval: 1.41–1.63] (Table 2). This SNP belongs to a bin of 10 SNPs in strong LD ($r^2 > 0.8$) in the 1000 Genomes KHV population that are located either within or just downstream of the *LACCI* gene (S4 Table). The

Table 1. Association statistics in the discovery and replication samples for the six genome-wide significant HLA SNPs and the five previously published non-HLA SNPs replicated in the discovery GWAS.

SNP	Chr	Position ^S	m/M	Type	Discovery			Replication		
					MAF	OR* [95% CI]	P	MAF	P [#]	OR* [#] [95% CI]
<i>Significant HLA hits</i>										
rs3095309	6	31091475	T/C	imputed	0.476	1.6 [1.43–1.77]	8.2×10^{-09}	0.44	9.46×10^{-05}	1.30 [1.16–Inf]
rs3094194	6	31136240	G/A	imputed	0.46	1.71 [1.53–1.88]	1.4×10^{-08}	0.44	5.03×10^{-06}	1.36 [1.21–Inf]
rs2844633	6	31096189	T/C	imputed	0.269	0.56 [0.38–0.75]	1.7×10^{-08}	0.30	4.20×10^{-04}	0.79 [0–0.89]
rs2844634	6	31096184	C/G	imputed	0.269	0.56 [0.38–0.75]	1.7×10^{-08}	0.30	4.65×10^{-04}	0.79 [0–0.89]
rs2394945	6	31220971	G/C	imputed	0.353	1.73 [1.54–1.93]	4.2×10^{-08}	0.35	9.20×10^{-07}	1.42 [1.25–Inf]
rs3094196	6	31093947	G/A	imputed	0.269	0.57 [0.38–0.75]	4.9×10^{-08}	0.30	4.70×10^{-04}	0.79 [0–0.89]
<i>Previously published non-HLA hits replicated in the discovery GWAS</i>										
rs3762318	1	67597119	G/A	genotyped	0.05	0.52 [0.12–0.93]	4.41×10^{-03}	0.06	1.57×10^{-01}	0.86 [0–1.10]
rs2221593	1	212873431	T/C	genotyped	0.229	1.33 [1.13–1.53]	1.94×10^{-04}	Failed call rate < 95%		
rs2058660	2	103054449	G/A	imputed	0.414	0.75 [0.57–0.92]	5.66×10^{-03}	0.40	6.46×10^{-02}	0.90 [0–1.01]
rs6871626	5	158826792	A/C	imputed ^ε	0.324	0.71 [0.58–0.86]	5.12×10^{-04}	0.32	1.37×10^{-05}	0.74 [0–0.83]
rs3764147	13	44457925	G/A	genotyped	0.396	1.41 [1.18–1.67]	6.56×10^{-05}	0.39	2.36×10^{-11}	1.61 [1.43–Inf]

m/M = minor/Major alleles; MAF = minor allele frequency; OR = odds ratio; CI = confidence interval

^S positions are given according to the hg19 genome build

* with respect to the minor allele

[#] one-tailed test

^ε imputed using the 1000 Genomes phase 1 reference panel

Bold typeface indicates SNPs significant in the replication sample

<https://doi.org/10.1371/journal.ppat.1008565.t001>

Table 2. Association statistics in the discovery GWAS, the case/control replication and the combined GWAS and replication samples for the five independent SNPs identified by multivariate analysis.

SNP (chr:position [§])	m/M	Gene	MAF [@]	Discovery GWAS		Replication		Combined GWAS and replication sample	
				P value	OR* [95% CI]	P value [#]	OR* [#] [95% CI]	P value	OR* [95% CI]
rs1265048 (6:31081409)	C/T	<i>C6orf15</i>	0.49	8.72x10 ⁻⁰⁶	0.63 [0.52–0.63]	1.58x10 ⁻⁰⁶	0.72 [0–0.81]	5.53x10 ⁻¹¹	0.69 [0.58–0.80]
rs114598080 (6:31255442)	A/G	<i>HLA-C</i>	0.10	3.90x10 ⁻⁰⁶	1.85 [1.39–2.44]	4.39x10 ⁻¹¹	2.06 [1.7–Inf]	8.77x10 ⁻¹³	1.47 [1.46–1.48]
rs3187964 (6:32605207)	C/G	<i>HLA-DQA1</i>	0.41	1.28x10 ⁻⁰⁶	1.52 [1.25–1.85]	5.28x10 ⁻¹³	1.72 [1.51–Inf]	8.35x10 ⁻¹⁶	1.67 [1.55–1.80]
rs3764147 (13:44457925)	G/A	<i>LACC1</i>	0.35	6.56x10 ⁻⁰⁵	1.41 [1.18–1.67]	2.36x10 ⁻¹¹	1.61 [1.49–Inf]	5.06x10 ⁻¹⁴	1.52 [1.41–1.63]
rs6871626 [£] (5:158826792)	A/C	<i>IL12B</i>	0.35	5.12x10 ⁻⁰⁴	0.71 [0.58–0.86]	1.37x10 ⁻⁰⁵	0.74 [0–0.86]	6.44x10 ⁻⁰⁸	0.73 [0.61–0.84]

m/M = minor/Major alleles; MAF = minor allele frequency; OR = odds ratio; CI = confidence interval

[§] positions are given according to the hg19 genome build

[@] MAF was estimated from the 939 leprosy unaffected parents of the discovery GWAS and 668 controls of the replication sample

* with respect to the minor allele

[#] one-tailed test

[£] imputed using the 1000 Genomes phase 1 reference panel

<https://doi.org/10.1371/journal.ppat.1008565.t002>

second replicated hit, rs6871626, is located on chromosome 5, 69 kb 5' to *IL12B*. A borderline non-significant genome-wide *p*-value of 6.4x10⁻⁸ was obtained for this SNP, with an OR of developing leprosy for AC vs. AA (or CC vs. AC) of 0.73 [0.61–0.84]. This SNP is not part of the 1000 Genomes phase 3 reference panel and no LD information was available for the KHV population, but it was successfully sequenced in the genome part of The Genome Aggregation Database (gnomAD) [37].

Within the MHC region. We genotyped 88 SNPs within the MHC region, and used them to impute the 364 SNPs with a discovery *p*-value < 10⁻⁵. More than 98% of these SNPs (358/364) were imputed with an information score > 0.8. The six HLA SNPs with genome-wide significance were replicated, all with a one-tailed *p*-value < 5x10⁻⁴ (Table 1). In addition, 319 of the 358 genome-wide suggestive HLA SNPs were replicated, with a one-tailed *p*-value threshold of 0.01. They clustered in two intervals, from 31 Mb to 31.5 Mb and from 32.5 Mb to 32.7 Mb (Fig 2, S5 Table). We further investigated the number of SNPs in the HLA region required to capture the full association signal, by performing a multivariable analysis. This analysis was conducted in the replication sample, which provided the most reliable set of SNPs and facilitated the analysis because of its case/control nature. Before performing this multivariable analysis, we clumped the SNPs, as described in the methods section, on the basis of their genotyping status in the replication sample (genotyped SNPs were favored over imputed ones), their one-tailed replication *p*-value and a LD *r*² threshold of 0.5.

The clumping procedure identified 18 index SNPs with one-tailed *p*-value < 0.01, which were then tested in the multivariable model with the best subset method (SCORE) implemented in the logistic regression procedure of the SAS software (S6 Table). We found that three SNPs (rs3187964, rs114598080 and rs1265048) were required to capture the full association signal in the HLA region (S7 Table). The addition of a fourth SNP did not significantly improve the multivariable model (*p*-value = 0.23). The most significant SNP in the multivariable model was rs3187964 (multivariable *p*-value = 7.8x10⁻⁸). The strength of association was much lower for rs114598080 (multivariable *p*-value = 3.75x10⁻⁵), and even lower for rs1265048, (multivariable *p*-value = 0.011). These three SNPs were sufficient to account for the HLA association signal found in our GWAS.

We then investigated in more detail the location, LD in the 1000 Genome KHV population and univariate results in the combined sample of these three HLA SNPs (Table 2). The

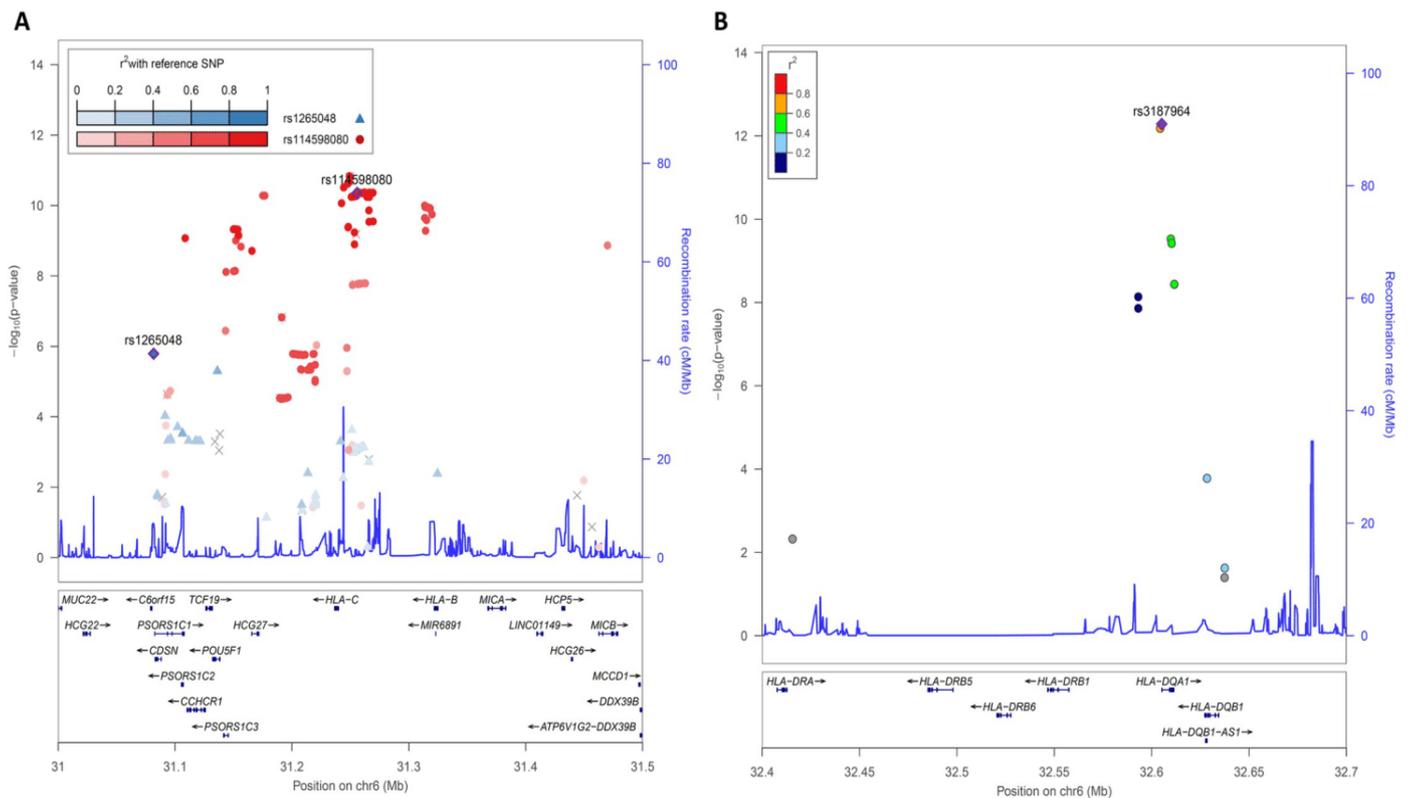


Fig 2. Regional association plot for HLA class I (A) and class II (B) SNPs in the case/control replication sample. Locus zoom plot showing one-tailed p -values for association within the HLA region. The three independent SNPs identified by multivariate analysis, rs1265048 (A), rs114598080 (A) and rs3187964 (B), are represented as blue, red and purple diamonds, respectively. The colors indicate the pairwise linkage disequilibrium (r^2) with the three independent SNPs, as calculated in the 1000 Genomes KHV population. SNPs not found in the reference populations are shown in gray.

<https://doi.org/10.1371/journal.ppat.1008565.g002>

rs3187964 SNP belongs to the HLA interval corresponding to the HLA class II region (Fig 2B). In the combined analysis of the discovery and replication samples, it gave a univariate p -value of 8.3×10^{-16} with an OR for developing leprosy for CG vs. GG (or CC vs. CG) of 1.67 [1.55–1.80]. This SNP is located in the 5' untranslated region of the *HLA-DQA1* gene and was not in strong LD ($r^2 > 0.8$) with other SNPs in the 1000 Genomes KHV population in an LDproxy analysis [38]. The other two SNPs are in the first HLA interval, which includes HLA class I genes, together with many other genes (Fig 2A). The rs114598080 SNP, located in a CTCF binding site (ENSR00000262841) 15.6 kb upstream of the *HLA-C* gene, gave a univariate p -value of 8.8×10^{-13} in the combined analysis, with an OR for developing leprosy for AG vs. GG (or AA vs. AG) of 1.47 [1.46–1.48]. This SNP belongs to a bin of 245 SNPs in strong LD that span almost 200 kb (S8 Table). The third SNP, rs1265048, is located near the *C6orf15* (1.07 kb), *PSORS1C1* (1.12 kb), and *CDSN* (1.2 kb) genes and it gave a univariate p -value of 5.5×10^{-11} in the combined analysis, with an OR for developing leprosy for CT vs. TT (or CC vs. CT) of 0.69 [0.58–0.80]. It was not in strong LD with other SNPs in the 1000 Genomes KHV population in LDproxy analysis.

Discussion

We performed the first family-based GWAS for leprosy in a Vietnamese population. The main susceptibility loci identified were in the HLA region. The most significant SNP, rs3187964, is located in the MHC class II region, in the 5' untranslated region of *HLA-DQA1* and did not display strong LD ($r^2 > 0.8$) with other SNPs in the region in the 1000 Genomes KHV

population. Several studies have consistently reported the involvement of HLA class II alleles, mostly *HLA-DRB1* alleles, as key genetic factors controlling susceptibility to various forms of leprosy [11, 18, 39–41]. A recent HLA imputation-based meta-analysis in the Chinese population identified *HLA-DRB1* and *HLA-DQA1* alleles and specific amino acids of HLA-DR β 1 as independent protective factors for leprosy [39]. Interestingly, in the GTex project V7 (<https://gtexportal.org/home/>), rs3187964 was identified as a strong eQTL for several HLA class II genes with *p*-values below 10^{-50} for *HLA-DQB1*, *HLA-DQB2*, *HLA-DQA1* and *HLA-DQA2* in multiple tissues, including skin and nerves (S9 Table). It was also found to be a splice QTL for *HLA-DQA1* (nominal *p*-value = 3.3×10^{-7} , FDR = 0.000155). HLA-DQ genes have been implicated in several autoimmune and inflammatory diseases [42] and, recently, in susceptibility to streptococcal disease [43]. Leprosy patients are prone to develop acute inflammatory reaction, T1R, and we have recently shown a striking overlap between T1R and inflammatory bowel disease (IBD) risk alleles [24]. However, SNPs of the HLA region were not major risk loci for T1R suggesting an effect of HLA SNPs on leprosy *per se* susceptibility and not on T1R. Interestingly, there are several examples of loci with mirror genetic effects between infectious diseases and inflammatory disorders, where the allele protective against infection is deleterious for inflammatory disorders, consistent with the view that the current high frequency of inflammatory/autoimmune diseases may reflect past selection for strong immune responses to combat infection [44, 45]. Additional studies are needed to precise the role of HLA-DQ genes in leprosy susceptibility with respect to T1R and inflammatory diseases.

The other two independent SNPs associated with leprosy in the HLA region are located in the class I region. The most significant, rs114598080, belongs to a large bin of SNPs in strong LD spanning the HLA class I genes *HLA-C* and *HLA-B* and the non-HLA genes *CCHCR1*, *TCF19*, *POU5F1*, *PSORS1C3*, and *HCG27*. The complex LD structure makes it difficult to identify the gene related to this association hit. Nevertheless, one SNP of the bin, rs2394885, was shown to tag the *HLA-C*15:05* allele in the Vietnamese population [46]. This SNP was also validated as a leprosy susceptibility locus in an Indian population [46]. In the Chinese HLA imputation-based meta-analysis, *HLA-C*08:01* and specific *HLA-C* amino acids were found to be independent protective and risk factors for leprosy [39]. The rs114598080 SNP is also in strong LD with a possibly damaging missense variant of the *CCHCR1* gene. *CCHCR1* is a protein of unknown function that was recently found to have properties typical of P body-resident proteins [47]. P bodies are evolutionary conserved cytoplasmic components that contain proteins involved in mRNA degradation, translation repression, and mRNA storage [48]. Interestingly, P body core components were recently shown to positively regulate plant pattern-triggered innate immunity [49]. In addition, rs114598080 is a strong eQTL in sun-exposed skin and tibial artery for *POU5F1* ($p < 10^{-10}$), also known as *OCT4*, which encodes a master transcription factor for pluripotent cell self-renewal [50] (S10 Table). The second SNP of the HLA class I region, rs1265048, is located 1.1 kb upstream of *C6orf15*, 1.1 kb upstream of *PSORS1C1* and 1.2 kb upstream of *CDSN*. Interestingly, the SNP correlates with the expression of several genes, including *HLA-C* (minimum *p*-value = 1.3×10^{-5}), and *POU5F1* (minimum *p*-value = 2.8×10^{-5}) in particular, sporadically in some tissues (S11 Table).

In addition to the HLA locus, we were able to consistently replicate the *LACC1* and *IL12B* loci previously associated with leprosy in the Chinese population both in the discovery GWAS and the replication sample. The *LACC1* leprosy susceptibility SNP rs3764147 is a missense variant (p.Ile254Val) first identified by a leprosy GWAS in the Han Chinese population [9] and then replicated in Indian, African [51], Vietnamese [20], Yi [52], Wenshan and Yuxi [16] Chinese populations. The leprosy susceptibility allele was also shown to be associated with an increased susceptibility to Crohn's [53, 54] and inflammatory bowel diseases [55, 56]. SNP rs3764147 has a PHRED-scaled CADD score [57, 58] of 16, placing it in the top 2.5% of

deleterious variants relative to all possible reference genome single-nucleotide variants. Interestingly, recent functional studies have shown that *LACCI* encodes a central regulator of the metabolic function and bioenergetic state of macrophages [59]. Furthermore, the rs3764147 G risk allele is associated with lower levels of innate receptor-induced activity in primary human monocyte-derived macrophages [60], potentially accounting for the higher susceptibility to intestinal inflammatory diseases and leprosy.

SNP rs6871626 is located in an intergenic region on chromosome 5. In a previous genome-wide meta-analysis study of ulcerative colitis, the most likely candidate gene assigned to this SNP by Gene Relationship Across Implicated Loci (GRAIL) pathway analysis was *IL12B* [61]. GRAIL is a statistical tool that assesses the degree of relatedness of a list of disease regions by text mining PubMed abstracts, for the annotation of candidate genes for disease risk [62]. *IL12B* encodes IL-12p40, which is common to two interleukins, IL-12 and IL-23. Both IL-12- and IL-23-dependent signaling pathways play critical roles in human antimycobacterial immunity, through the induction of IFN- γ [63, 64]. Autosomal recessive complete *IL12B* deficiency is a genetic etiology of Mendelian susceptibility to mycobacterial diseases (MSMD), a rare condition characterized by predisposition to clinical disease caused by weakly virulent mycobacteria, such as BCG vaccines and environmental mycobacteria, in otherwise healthy individuals [65]. The rs6871626 C allele is associated with a risk of leprosy, but has been shown to be protective against ulcerative colitis [61], ankylosing spondylitis and Crohn's disease [66], inflammatory bowel disease [55] and Takayasu arteritis [67]. Interestingly, in patients with Takayasu arteritis, the rs6871626 C allele, which confers susceptibility to leprosy, was associated with abnormally low levels of IL-12 (composed of IL-12p40 and IL-12p35) in plasma and low levels of both IL-12p40 and IL12 in the culture supernatants of patient monocytes/macrophages stimulated with LPS [68].

Our study had some limitations. First, our sample size was limited with 622 affected offspring. Assuming 622 trios, the power of the transmission disequilibrium test was below 80% for SNPs with a MAF of 0.05 with a relative risk below 2 or for SNPs with a MAF of 0.3 with a relative risk below 1.5, and we may have missed novel associations and failed to replicate some of the previously identified loci with modest genetic effect size. Second, as in all leprosy studies, the case definition was purely clinical. Although leprosy was diagnosed by two independent and experienced physicians, we cannot exclude some misdiagnosis which may affect the power of the study. Third, two out of eight of the non-HLA suggestive hits and one out of five of the previously published non-HLA hits could not be tested for association in the replication sample because of genotyping failure. Finally, the investigation of the identified loci was based solely on in silico analysis using publicly available resources and not on functional studies. In conclusion, we performed the first family-based GWAS of leprosy *per se* in the Vietnamese population. The strongest association signals were observed for SNPs within the HLA region. We also validated the independent impact of two additional variants, one in *LACCI* and the second close to *IL12B*, recently reported to play functional roles. We further broke down the association signal within the HLA region into three independent signals, two mapping to the HLA class I region, and one, the leading signal, located in the class II region close to *HLA-DQA1*. More refined analyses based on direct HLA typing, which is lacking so far, are required to characterize further the precise contribution of HLA class I and class II alleles to leprosy susceptibility.

Supporting information

S1 Table. Clinical characteristics of study populations from the Vietnamese family-based and case-control studies.

(XLSX)

S2 Table. LD-based clumping of genome-wide suggestive non-HLA hits.
(XLSX)

S3 Table. Previously published genome-wide significant SNPs outside the HLA region.
(XLSX)

S4 Table. SNPs in high linkage disequilibrium ($r^2 > 0.8$) with rs3764147 in the 1000 Genomes KHV population.
(XLSX)

S5 Table. Association statistics of genome-wide suggestive HLA SNPs in the discovery cohort and the replication cohort.
(XLSX)

S6 Table. LD-based clumping of replicated SNPs within the HLA region in the case/control replication sample.
(XLSX)

S7 Table. Results of the multivariate analysis within the HLA region in the case/control replication sample.
(XLSX)

S8 Table. SNPs in high linkage disequilibrium ($r^2 > 0.8$) with rs114598080 in the 1000 genomes KHV population.
(XLSX)

S9 Table. Single-Tissue eQTL for rs3187964 in the GTEx project.
(XLSX)

S10 Table. Single-Tissue eQTL for rs114598080 in the GTEx project.
(XLSX)

S11 Table. Single-Tissue eQTL for rs1265048 in the GTEx project.
(XLSX)

S1 Fig. Principal component analysis of the GWAS discovery sample with all 1000 genomes phase 3 samples (A) and with the East-Asian 1000 genomes samples (B) AFR = African; AMR = American; EAS = East Asian; EUR = European; GWAS = Genome-wide association study samples; SAS = South Asian; CDX = Chinese Dai in Xishuangbanna, China; CHB = Han Chinese in Beijing, China; CHS = Southern Han Chinese; JPT = Japanese in Tokyo, Japan; KHV = Kinh in Ho Chi Minh City, Vietnam.
(TIF)

S2 Fig. Quantile-quantile plot of the association results with leprosy *per se* in the family-based GWAS for all SNPs (black) and all SNPs outside the HLA region (grey).
(TIF)

Acknowledgments

We thank all leprosy patients and their family who participated in this study. We thank the staff of leprosy program in Southern Việt Nam and the members of leprosy control program at Hospital of D-V of Ho Chi Minh City, Việt Nam. We thank members of the laboratory of human genetics of infectious diseases and the Erwin Schurr laboratory for helpful discussions. We thank the French National Centre of Human Genomics Research (CNRGH) and Genome Quebec for performing genotyping.

Author Contributions

Conceptualization: Alexandre Alcaïs, Erwin Schurr.

Data curation: Chaima Gzara, Marianna Orlova.

Formal analysis: Chaima Gzara, Monica Dallmann-Sauer.

Funding acquisition: Alexandre Alcaïs, Erwin Schurr.

Methodology: Chaima Gzara, Aurélie Cobat.

Resources: Nguyen Van Thuc, Vu Hong Thai, Vinicius M. Fava, Marie-Thérèse Bihoreau, Anne Boland.

Supervision: Laurent Abel, Alexandre Alcaïs, Aurélie Cobat.

Validation: Aurélie Cobat.

Visualization: Chaima Gzara.

Writing – original draft: Chaima Gzara, Laurent Abel, Aurélie Cobat.

Writing – review & editing: Chaima Gzara, Monica Dallmann-Sauer, Marianna Orlova, Vinicius M. Fava, Alexandre Alcaïs, Erwin Schurr.

References

1. Smith WC, van Brakel W, Gillis T, Saunderson P, Richardus JH. The missing millions: a threat to the elimination of leprosy. *PLoS Negl Trop Dis*. 2015; 9(4):e0003658. Epub 2015/04/24. <https://doi.org/10.1371/journal.pntd.0003658> PMID: 25905706; PubMed Central PMCID: PMC4408099.
2. Gaschignard J, Grant AV, Thuc NV, Orlova M, Cobat A, Huong NT, et al. Pauci- and Multibacillary Leprosy: Two Distinct, Genetically Neglected Diseases. *PLoS Negl Trop Dis*. 2016; 10(5):e0004345. <https://doi.org/10.1371/journal.pntd.0004345> PMID: 27219008; PubMed Central PMCID: PMC4878860.
3. Cole ST, Eiglmeier K, Parkhill J, James KD, Thomson NR, Wheeler PR, et al. Massive gene decay in the leprosy bacillus. *Nature*. 2001; 409(6823):1007–11. Epub 2001/03/10. <https://doi.org/10.1038/35059006> PMID: 11234002.
4. Monot M, Honore N, Garnier T, Araoz R, Coppee JY, Lacroix C, et al. On the origin of leprosy. *Science*. 2005; 308(5724):1040–2. Epub 2005/05/17. <https://doi.org/10.1126/science.1109759> PMID: 15894530.
5. Fava VM, Dallmann-Sauer M, Schurr E. Genetics of leprosy: today and beyond. *Hum Genet*. 2019. Epub 2019/11/13. <https://doi.org/10.1007/s00439-019-02087-5> PMID: 31713021.
6. Mira MT, Alcais A, Van Thuc N, Thai VH, Huong NT, Ba NN, et al. Chromosome 6q25 is linked to susceptibility to leprosy in a Vietnamese population. *Nat Genet*. 2003; 33(3):412–5. Epub 2003/02/11. <https://doi.org/10.1038/ng1096> PMID: 12577057.
7. Mira MT, Alcais A, Nguyen VT, Moraes MO, Di Flumeri C, Vu HT, et al. Susceptibility to leprosy is associated with PARK2 and PACRG. *Nature*. 2004; 427(6975):636–40. Epub 2004/01/23. <https://doi.org/10.1038/nature02326> PMID: 14737177.
8. Alcais A, Alter A, Antoni G, Orlova M, Nguyen VT, Singh M, et al. Stepwise replication identifies a low-producing lymphotoxin-alpha allele as a major risk factor for early-onset leprosy. *Nat Genet*. 2007; 39(4):517–22. Epub 2007/03/14. <https://doi.org/10.1038/ng2000> PMID: 17353895.
9. Zhang FR, Huang W, Chen SM, Sun LD, Liu H, Li Y, et al. Genomewide association study of leprosy. *N Engl J Med*. 2009; 361(27):2609–18. Epub 2009/12/19. <https://doi.org/10.1056/NEJMoa0903753> PMID: 20018961.
10. Zhang F, Liu H, Chen S, Low H, Sun L, Cui Y, et al. Identification of two new loci at IL23R and RAB32 that influence susceptibility to leprosy. *Nat Genet*. 2011; 43(12):1247–51. Epub 2011/10/25. <https://doi.org/10.1038/ng.973> PMID: 22019778.
11. Liu H, Irwanto A, Fu X, Yu G, Yu Y, Sun Y, et al. Discovery of six new susceptibility loci and analysis of pleiotropic effects in leprosy. *Nat Genet*. 2015; 47(3):267–71. Epub 2015/02/03. <https://doi.org/10.1038/ng.3212> PMID: 25642632.

12. Wang Z, Sun Y, Fu X, Yu G, Wang C, Bao F, et al. A large-scale genome-wide association and meta-analysis identified four novel susceptibility loci for leprosy. *Nat Commun.* 2016; 7:13760. Epub 2016/12/16. <https://doi.org/10.1038/ncomms13760> PMID: 27976721; PubMed Central PMCID: PMC5172377.
13. Wong SH, Gochhait S, Malhotra D, Pettersson FH, Teo YY, Khor CC, et al. Leprosy and the adaptation of human toll-like receptor 1. *PLoS Pathog.* 2010; 6:e1000979. Epub 2010/07/10. <https://doi.org/10.1371/journal.ppat.1000979> PMID: 20617178; PubMed Central PMCID: PMC2895660.
14. Liu H, Wang Z, Li Y, Yu G, Fu X, Wang C, et al. Genome-Wide Analysis of Protein-Coding Variants in Leprosy. *J Invest Dermatol.* 2017; 137(12):2544–51. Epub 2017/08/27. <https://doi.org/10.1016/j.jid.2017.08.004> PMID: 28842327.
15. Wang Z, Mi Z, Wang H, Sun L, Yu G, Fu X, et al. Discovery of 4 exonic and 1 intergenic novel susceptibility loci for leprosy. *Clin Genet.* 2018; 94(2):259–63. Epub 2018/05/04. <https://doi.org/10.1111/cge.13376> PMID: 29722023.
16. Wang D, Fan Y, Malhi M, Bi R, Wu Y, Xu M, et al. Missense Variants in HIF1A and LACC1 Contribute to Leprosy Risk in Han Chinese. *Am J Hum Genet.* 2018; 102(5):794–805. Epub 2018/05/01. <https://doi.org/10.1016/j.ajhg.2018.03.006> PMID: 29706348; PubMed Central PMCID: PMC5986702.
17. Dallmann-Sauer M, Correa-Macedo W, Schurr E. Human genetics of mycobacterial disease. *Mamm Genome.* 2018; 29(7–8):523–38. Epub 2018/08/18. <https://doi.org/10.1007/s00335-018-9765-4> PMID: 30116885; PubMed Central PMCID: PMC6132723.
18. Cambri G, Mira MT. Genetic Susceptibility to Leprosy-From Classic Immune-Related Candidate Genes to Hypothesis-Free, Whole Genome Approaches. *Front Immunol.* 2018; 9:1674. Epub 2018/08/07. <https://doi.org/10.3389/fimmu.2018.01674> PMID: 30079069; PubMed Central PMCID: PMC6062607.
19. Cobat A, Abel L, Alcais A, Schurr E. A general efficient and flexible approach for genome-wide association analyses of imputed genotypes in family-based designs. *Genet Epidemiol.* 2014; 38(6):560–71. Epub 2014/07/22. <https://doi.org/10.1002/gepi.21842> PMID: 25044438.
20. Grant AV, Alter A, Huong NT, Orlova M, Van Thuc N, Ba NN, et al. Crohn's disease susceptibility genes are associated with leprosy in the Vietnamese population. *J Infect Dis.* 2012; 206(11):1763–7. Epub 2012/09/18. <https://doi.org/10.1093/infdis/jis588> PMID: 22984114.
21. Fava VM, Cobat A, Van Thuc N, Latini AC, Stefani MM, Belone AF, et al. Association of TNFSF8 regulatory variants with excessive inflammatory responses but not leprosy per se. *J Infect Dis.* 2015; 211(6):968–77. Epub 2014/10/17. <https://doi.org/10.1093/infdis/jiu566> PMID: 25320285.
22. Fava VM, Manry J, Cobat A, Orlova M, Van Thuc N, Ba NN, et al. A Missense LRRK2 Variant Is a Risk Factor for Excessive Inflammatory Responses in Leprosy. *PLoS Negl Trop Dis.* 2016; 10(2):e0004412. <https://doi.org/10.1371/journal.pntd.0004412> PMID: 26844546; PubMed Central PMCID: PMC4742274.
23. Fava VM, Xu YZ, Lettre G, Van Thuc N, Orlova M, Thai VH, et al. Pleiotropic effects for Parkin and LRRK2 in leprosy type-1 reactions and Parkinson's disease. *Proc Natl Acad Sci U S A.* 2019; 116(31):15616–24. Epub 2019/07/17. <https://doi.org/10.1073/pnas.1901805116> PMID: 31308240; PubMed Central PMCID: PMC6681704.
24. Fava VM, Manry J, Cobat A, Orlova M, Van Thuc N, Moraes MO, et al. A genome wide association study identifies a lncRNA as risk factor for pathological inflammatory responses in leprosy. *PLoS Genet.* 2017; 13(2):e1006637. <https://doi.org/10.1371/journal.pgen.1006637> PMID: 28222097; PubMed Central PMCID: PMC5340414.
25. Teo YY, Inouye M, Small KS, Gwilliam R, Deloukas P, Kwiatkowski DP, et al. A genotype calling algorithm for the Illumina BeadArray platform. *Bioinformatics.* 2007; 23(20):2741–6. Epub 2007/09/12. <https://doi.org/10.1093/bioinformatics/btm443> PMID: 17846035; PubMed Central PMCID: PMC2666488.
26. Purcell S, Chang C. PLINK 1.9. Available from: www.cog-genomics.org/plink/1.9/.
27. Delaneau O, Marchini J, Zagury JF. A linear complexity phasing method for thousands of genomes. *Nat Methods.* 2011; 9(2):179–81. Epub 2011/12/06. <https://doi.org/10.1038/nmeth.1785> PMID: 22138821.
28. O'Connell J, Gurdasani D, Delaneau O, Pirastu N, Ulivi S, Cocca M, et al. A general approach for haplotype phasing across the full spectrum of relatedness. *PLoS Genet.* 2014; 10(4):e1004234. Epub 2014/04/20. <https://doi.org/10.1371/journal.pgen.1004234> PMID: 24743097; PubMed Central PMCID: PMC3990520.
29. Marchini J, Howie B, Myers S, McVean G, Donnelly P. A new multipoint method for genome-wide association studies by imputation of genotypes. *Nat Genet.* 2007; 39(7):906–13. Epub 2007/06/19. <https://doi.org/10.1038/ng2088> PMID: 17572673.
30. Chang CC, Chow CC, Tellier LC, Vattikuti S, Purcell SM, Lee JJ. Second-generation PLINK: rising to the challenge of larger and richer datasets. *Gigascience.* 2015; 4:7. Epub 2015/02/28. <https://doi.org/10.1186/s13742-015-0047-8> PMID: 25722852; PubMed Central PMCID: PMC4342193.

31. DePristo MA, Banks E, Poplin R, Garimella KV, Maguire JR, Hartl C, et al. A framework for variation discovery and genotyping using next-generation DNA sequencing data. *Nat Genet.* 2011; 43(5):491–8. Epub 2011/04/12. <https://doi.org/10.1038/ng.806> PMID: 21478889; PubMed Central PMCID: PMC3083463.
32. Laird NM, Horvath S, Xu X. Implementing a unified approach to family-based tests of association. *Genet Epidemiol.* 2000; 19 Suppl 1:S36–42. [https://doi.org/10.1002/1098-2272\(2000\)19:1+<::AID-GEPI6>3.0.CO;2-M](https://doi.org/10.1002/1098-2272(2000)19:1+<::AID-GEPI6>3.0.CO;2-M) PMID: 11055368.
33. Cordell HJ, Barratt BJ, Clayton DG. Case/pseudocontrol analysis in genetic association studies: A unified framework for detection of genotype and haplotype associations, gene-gene and gene-environment interactions, and parent-of-origin effects. *Genet Epidemiol.* 2004; 26(3):167–85. Epub 2004/03/17. <https://doi.org/10.1002/gepi.10307> PMID: 15022205.
34. Grant AV, El Baghdadi J, Sabri A, El Azbaoui S, Alaoui-Tahiri K, Abderrahmani Rhorfi I, et al. Age-dependent association between pulmonary tuberculosis and common TOX variants in the 8q12-13 linkage region. *Am J Hum Genet.* 2013; 92(3):407–14. Epub 2013/02/19. <https://doi.org/10.1016/j.ajhg.2013.01.013> PMID: 23415668; PubMed Central PMCID: PMC3591857.
35. Grant AV, Sabri A, Abid A, Abderrahmani Rhorfi I, Benkirane M, Souhi H, et al. A genome-wide association study of pulmonary tuberculosis in Morocco. *Hum Genet.* 2016; 135(3):299–307. Epub 2016/01/16. <https://doi.org/10.1007/s00439-016-1633-2> PMID: 26767831; PubMed Central PMCID: PMC5042142.
36. Wellcome Trust Case Control C. Genome-wide association study of 14,000 cases of seven common diseases and 3,000 shared controls. *Nature.* 2007; 447(7145):661–78. Epub 2007/06/08. <https://doi.org/10.1038/nature05911> PMID: 17554300; PubMed Central PMCID: PMC2719288.
37. Karczewski Konrad J., Francioli Laurent C., Tiao Grace, Cummings Beryl B., Jessica Alfoldi Qingbo Wang, et al. Variation across 141,456 human exomes and genomes reveals the spectrum of loss-of-function intolerance across human protein-coding genes. *bioRxiv.* 2019. <https://doi.org/10.1101/531210>.
38. Machiela MJ, Chanock SJ. LDlink: a web-based application for exploring population-specific haplotype structure and linking correlated alleles of possible functional variants. *Bioinformatics.* 2015; 31(21):3555–7. Epub 2015/07/04. <https://doi.org/10.1093/bioinformatics/btv402> PMID: 26139635; PubMed Central PMCID: PMC4626747.
39. Zhang X, Cheng Y, Zhang Q, Wang X, Lin Y, Yang C, et al. Meta-Analysis Identifies Major Histocompatibility Complex Loci in or Near HLA-DRB1, HLA-DQA1, HLA-C as Associated with Leprosy in Chinese Han Population. *J Invest Dermatol.* 2019; 139(4):957–60. Epub 2018/11/06. <https://doi.org/10.1016/j.jid.2018.09.029> PMID: 30389493.
40. Jarduli LR, Sell AM, Reis PG, Sippert EA, Ayo CM, Mazini PS, et al. Role of HLA, KIR, MICA, and cytokines genes in leprosy. *Biomed Res Int.* 2013; 2013:989837. Epub 2013/08/13. <https://doi.org/10.1155/2013/989837> PMID: 23936864; PubMed Central PMCID: PMC3722889.
41. Krause-Kyora B, Nutsua M, Boehme L, Pierini F, Pedersen DD, Kornell SC, et al. Ancient DNA study reveals HLA susceptibility locus for leprosy in medieval Europeans. *Nat Commun.* 2018; 9(1):1569. Epub 2018/05/03. <https://doi.org/10.1038/s41467-018-03857-x> PMID: 29717136; PubMed Central PMCID: PMC5931558.
42. Matzaraki V, Kumar V, Wijmenga C, Zhernakova A. The MHC locus and genetic susceptibility to autoimmune and infectious diseases. *Genome Biol.* 2017; 18(1):76. Epub 2017/04/30. <https://doi.org/10.1186/s13059-017-1207-1> PMID: 28449694; PubMed Central PMCID: PMC5406920.
43. Parks T, Elliott K, Lamagni T, Auckland K, Mentzer AJ, Guy R, et al. Elevated risk of invasive group A streptococcal disease and host genetic variation in the human leucocyte antigen locus. *Genes Immun.* 2019. Epub 2019/08/30. <https://doi.org/10.1038/s41435-019-0082-z> PMID: 31462703.
44. Barreiro LB, Quintana-Murci L. From evolutionary genetics to human immunology: how selection shapes host defence genes. *Nat Rev Genet.* 2010; 11(1):17–30. <https://doi.org/10.1038/nrg2698> PMID: 19953080.
45. Manry J, Vincent QB, Johnson C, Chrabieh M, Lorenzo L, Theodorou I, et al. Genome-wide association study of Buruli ulcer in rural Benin highlights role of two LncRNAs and the autophagy pathway. *Commun Biol.* 2020; 3(1):177.
46. Alter A, Huong NT, Singh M, Orlova M, Van Thuc N, Katoch K, et al. Human leukocyte antigen class I region single-nucleotide polymorphisms are associated with leprosy susceptibility in Vietnam and India. *J Infect Dis.* 2011; 203(9):1274–81. Epub 2011/04/05. <https://doi.org/10.1093/infdis/jir024> PMID: 21459816; PubMed Central PMCID: PMC3069725.
47. Luo Y, Na Z, Slavoff SA. P-Bodies: Composition, Properties, and Functions. *Biochemistry.* 2018; 57(17):2424–31. Epub 2018/01/31. <https://doi.org/10.1021/acs.biochem.7b01162> PMID: 29381060; PubMed Central PMCID: PMC6296482.

48. Parker R, Sheth U. P bodies and the control of mRNA translation and degradation. *Mol Cell*. 2007; 25(5):635–46. <https://doi.org/10.1016/j.molcel.2007.02.011> PMID: 17349952.
49. Yu X, Li B, Jang GJ, Jiang S, Jiang D, Jang JC, et al. Orchestration of Processing Body Dynamics and mRNA Decay in Arabidopsis Immunity. *Cell Rep*. 2019; 28(8):2194–205 e6. <https://doi.org/10.1016/j.celrep.2019.07.054> PMID: 31433992; PubMed Central PMCID: PMC6716526.
50. Radzisheuskaya A, Chia Gle B, dos Santos RL, Theunissen TW, Castro LF, Nichols J, et al. A defined Oct4 level governs cell state transitions of pluripotency entry and differentiation into all embryonic lineages. *Nat Cell Biol*. 2013; 15(6):579–90. Epub 2013/05/01. <https://doi.org/10.1038/ncb2742> PMID: 23629142; PubMed Central PMCID: PMC3671976.
51. Wong SH, Hill AV, Vannberg FO, India-Africa-United Kingdom Leprosy Genetics C. Genomewide association study of leprosy. *N Engl J Med*. 2010; 362(15):1446–7; author reply 7–8. Epub 2010/04/16. <https://doi.org/10.1056/NEJMc1001451> PMID: 20393182.
52. Xiong JH, Mao C, Sha XW, Jin Z, Wang H, Liu YY, et al. Association between genetic variants in NOD2, C13orf31, and CCDC122 genes and leprosy among the Chinese Yi population. *Int J Dermatol*. 2016; 55(1):65–9. Epub 2015/08/04. <https://doi.org/10.1111/ijd.12981> PMID: 26235265.
53. Barrett JC, Hansoul S, Nicolae DL, Cho JH, Duerr RH, Rioux JD, et al. Genome-wide association defines more than 30 distinct susceptibility loci for Crohn's disease. *Nat Genet*. 2008; 40(8):955–62. Epub 2008/07/01. <https://doi.org/10.1038/ng.175> PMID: 18587394; PubMed Central PMCID: PMC2574810.
54. Franke A, McGovern DP, Barrett JC, Wang K, Radford-Smith GL, Ahmad T, et al. Genome-wide meta-analysis increases to 71 the number of confirmed Crohn's disease susceptibility loci. *Nat Genet*. 2010; 42(12):1118–25. Epub 2010/11/26. <https://doi.org/10.1038/ng.717> PMID: 21102463; PubMed Central PMCID: PMC3299551.
55. Jostins L, Ripke S, Weersma RK, Duerr RH, McGovern DP, Hui KY, et al. Host-microbe interactions have shaped the genetic architecture of inflammatory bowel disease. *Nature*. 2012; 491(7422):119–24. Epub 2012/11/07. <https://doi.org/10.1038/nature11582> PMID: 23128233; PubMed Central PMCID: PMC3491803.
56. Liu JZ, van Sommeren S, Huang H, Ng SC, Alberts R, Takahashi A, et al. Association analyses identify 38 susceptibility loci for inflammatory bowel disease and highlight shared genetic risk across populations. *Nat Genet*. 2015; 47(9):979–86. Epub 2015/07/21. <https://doi.org/10.1038/ng.3359> PMID: 26192919; PubMed Central PMCID: PMC4881818.
57. 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(3):310–5. Epub 2014/02/04. <https://doi.org/10.1038/ng.2892> PMID: 24487276; PubMed Central PMCID: PMC3992975.
58. Rentzsch P, Witten D, Cooper GM, Shendure J, Kircher M. CADD: predicting the deleteriousness of variants throughout the human genome. *Nucleic Acids Res*. 2019; 47(D1):D886–D94. Epub 2018/10/30. <https://doi.org/10.1093/nar/gky1016> PMID: 30371827; PubMed Central PMCID: PMC6323892.
59. Cader MZ, Boroviak K, Zhang Q, Assadi G, Kempster SL, Sewell GW, et al. C13orf31 (FAMIN) is a central regulator of immunometabolic function. *Nat Immunol*. 2016; 17(9):1046–56. Epub 2016/08/02. <https://doi.org/10.1038/ni.3532> PMID: 27478939; PubMed Central PMCID: PMC6581540.
60. Lahiri A, Hedl M, Yan J, Abraham C. Human LACC1 increases innate receptor-induced responses and a LACC1 disease-risk variant modulates these outcomes. *Nat Commun*. 2017; 8:15614. Epub 2017/06/09. <https://doi.org/10.1038/ncomms15614> PMID: 28593945; PubMed Central PMCID: PMC5472760.
61. Anderson CA, Boucher G, Lees CW, Franke A, D'Amato M, Taylor KD, et al. Meta-analysis identifies 29 additional ulcerative colitis risk loci, increasing the number of confirmed associations to 47. *Nat Genet*. 2011; 43(3):246–52. Epub 2011/02/08. <https://doi.org/10.1038/ng.764> PMID: 21297633; PubMed Central PMCID: PMC3084597.
62. Raychaudhuri S, Plenge RM, Rossin EJ, Ng AC, International Schizophrenia C, Purcell SM, et al. Identifying relationships among genomic disease regions: predicting genes at pathogenic SNP associations and rare deletions. *PLoS Genet*. 2009; 5(6):e1000534. Epub 2009/06/27. <https://doi.org/10.1371/journal.pgen.1000534> PMID: 19557189; PubMed Central PMCID: PMC2694358.
63. Boisson-Dupuis S, Ramirez-Alejo N, Li Z, Patin E, Rao G, Kerner G, et al. Tuberculosis and impaired IL-23-dependent IFN-gamma immunity in humans homozygous for a common TYK2 missense variant. *Sci Immunol*. 2018; 3(30). Epub 2018/12/24. <https://doi.org/10.1126/sciimmunol.aau8714> PMID: 30578352; PubMed Central PMCID: PMC6341984.
64. Martinez-Barricarte R, Markle JG, Ma CS, Deenick EK, Ramirez-Alejo N, Mele F, et al. Human IFN-gamma immunity to mycobacteria is governed by both IL-12 and IL-23. *Sci Immunol*. 2018; 3(30). Epub 2018/12/24. <https://doi.org/10.1126/sciimmunol.aau6759> PMID: 30578351; PubMed Central PMCID: PMC6380365.

65. Bustamante J, Boisson-Dupuis S, Abel L, Casanova JL. Mendelian susceptibility to mycobacterial disease: genetic, immunological, and clinical features of inborn errors of IFN-gamma immunity. *Semin Immunol*. 2014; 26(6):454–70. Epub 2014/12/03. <https://doi.org/10.1016/j.smim.2014.09.008> PMID: 25453225; PubMed Central PMCID: PMC4357480.
66. International Genetics of Ankylosing Spondylitis C, Cortes A, Hadler J, Pointon JP, Robinson PC, Karaderi T, et al. Identification of multiple risk variants for ankylosing spondylitis through high-density genotyping of immune-related loci. *Nat Genet*. 2013; 45(7):730–8. Epub 2013/06/12. <https://doi.org/10.1038/ng.2667> PMID: 23749187; PubMed Central PMCID: PMC3757343.
67. Terao C, Yoshifuji H, Kimura A, Matsumura T, Ohmura K, Takahashi M, et al. Two susceptibility loci to Takayasu arteritis reveal a synergistic role of the IL12B and HLA-B regions in a Japanese population. *Am J Hum Genet*. 2013; 93(2):289–97. Epub 2013/07/09. <https://doi.org/10.1016/j.ajhg.2013.05.024> PMID: 23830516; PubMed Central PMCID: PMC3738822.
68. Nakajima T, Yoshifuji H, Shimizu M, Kitagori K, Murakami K, Nakashima R, et al. A novel susceptibility locus in the IL12B region is associated with the pathophysiology of Takayasu arteritis through IL-12p40 and IL-12p70 production. *Arthritis Res Ther*. 2017; 19(1):197. Epub 2017/09/07. <https://doi.org/10.1186/s13075-017-1408-8> PMID: 28874185; PubMed Central PMCID: PMC5585951.

ANNEXE 4 : Article 2

The differential role of LRRK2 variants in nested leprosy phenotypes

Vinicius M Fava, Aurélie Cobat, Chaïma Gzara,
Alexandre Alcaïs, Laurent Abel, Erwin Schurr

*Publié dans Proceedings of the National Academy of Sciences USA,
2020 May 12;117(19):10124-10125.*

doi: 10.1073/pnas.200265411



REPLY TO ZHANG ET AL.:

The differential role of LRRK2 variants in nested leprosy phenotypes

Vinicius M. Fava^{a,b}, Aurélie Cobat^{c,d}, Chaïma Gzara^{c,d}, Alexandre Alcaïs^{c,d}, Laurent Abel^{c,d,e}, and Erwin Schurr^{a,b,f,1}

We thank Dr. Zhang et al. (1) for their thought-provoking comments and addition of exciting results to our recent study that identifies an overlap in the genetic control of type-1 reaction (T1R) and Parkinson's disease (PD) (2). To put the comments by Zhang et al. into context, the purpose and design of our study is to analyze the contribution of rare variants to T1R by contrasting T1R-affected with T1R-free leprosy patients. Our study is not designed to test for common variants or the impact of genetic variants on leprosy per se susceptibility. Rather, our study deals with the role of specific genetic variants in the development of T1R among leprosy patients, and we never state or imply that the *PRKN* and *LRRK2* genes have an exclusive impact on T1R but not leprosy. Indeed, we have published extensively on the role of common *PRKN* variants in leprosy susceptibility (3, 4)

As for *PRKN*, the two markers listed by Zhang et al. (1) do not show evidence for significant association with T1R in our study (2). We find that the V380L marker, which is not significant on its own, contributes to the rare variant analysis but not to an extent that impacts the overall conclusion about the contribution of rare *PRKN* variants to T1R risk. Given the nonsignificant effect of V380L, we do not attempt to do any functional validation of this particular variant.

Stimulated by the data of Zhang et al. (1) for *LRRK2*, we compared the R1628P distribution between T1R-affected and T1R-free leprosy cases of our sample against 472 ethnically matched healthy controls. For T1R-free

leprosy patients, we observed a trend for a risk effect of R1628P (odds ratios [OR] = 1.64, $P = 0.07$) of similar effect size to the one reported by Zhang et al. when using ethnically matched controls (OR = 1.6; table 1 in ref. 1). Conversely, for T1R-affected cases, a protective effect was detected (OR = 0.44, $P = 0.04$). Hence, R1628P is not a leprosy per se (i.e., leprosy irrespective of the clinical manifestation) variant, but is highly dependent on clinical specifics of leprosy. Since the incidence of T1R in leprosy can be as low as 8 to 9% depending on patient characteristics, it is possible that the sample of Zhang et al. comprises a small number of T1R cases, which would result in the observed impact of the R1628P polymorphism on leprosy risk (5, 6). The opposite effect for R1628P between T1R-affected and T1R-free patients stresses the importance of assessing subgroup heterogeneity and the increased power to detect associations with secondary phenotypes when comparing extremes, as we discussed previously (7–9).

We conclude, from our results, that *LRRK2* is an inhibitor of host inflammation (2). This conclusion is supported by the interesting observation, also quoted by Zhang et al. (1), that *Mycobacterium tuberculosis* (*Mtb*) induced neuroinflammation in *Lrrk2* KO mice (10). Similarly, a *Lrrk2* knockout had been shown to enhance the proinflammatory cytokine response to *Mtb* (11). Given that *LRRK2* R1628P is a gain of function mutation, collectively, the results of multiple studies support a role of *LRRK2* in the down-regulation of host inflammatory responses.

1 D. -F. Zhang, D. Wang, Y. -Y. Li, Y. -G. Yao, Is there an antagonistic pleiotropic effect of a *LRRK2* mutation on leprosy and Parkinson's disease? *Proc. Natl. Acad. Sci. U.S.A.* **117**, 10122–10123 (2020).

2 V. M. Fava et al., Pleiotropic effects for Parkin and *LRRK2* in leprosy type-1 reactions and Parkinson's disease. *Proc. Natl. Acad. Sci. U.S.A.* **116**, 15616–15624 (2019).

^aProgram in Infectious Diseases and Immunity in Global Health, The Research Institute of the McGill University Health Centre, Montreal, QC, Canada H4A3J1; ^bMcGill International TB Centre, Montreal, QC, Canada H4A 3J1; ^cLaboratory of Human Genetics of Infectious Diseases, Necker Branch, Institut National de la Santé et de la Recherche Médicale 1163, 75015 Paris, France; ^dImagine Institute, Paris Descartes-Sorbonne Paris Cité University, 75015 Paris, France; ^eSt. Giles Laboratory of Human Genetics of Infectious Diseases, Rockefeller Branch, The Rockefeller University, New York, NY 10065; and ^fDepartment of Human Genetics, Faculty of Medicine, McGill University, Montreal, QC, Canada H3A 0C7

Author contributions: V.M.F., A.A., L.A., and E.S. designed research; V.M.F., A.C., and C.G. analyzed data; and V.M.F., L.A., and E.S. wrote the paper.

The authors declare no competing interest.

Published under the PNAS license.

¹To whom correspondence should be addressed. Email: erwin.schurr@mcgill.ca.

First published April 28, 2020.

- 3 M. T. Mira *et al.*, Susceptibility to leprosy is associated with PARK2 and PACRG. *Nature* **427**, 636–640 (2004).
- 4 A. Alter *et al.*, Linkage disequilibrium pattern and age-at-diagnosis are critical for replicating genetic associations across ethnic groups in leprosy. *Hum. Genet.* **132**, 107–116 (2013).
- 5 N. Sharma, R. V. Koranne, V. Mendiratta, R. C. Sharma, A study of leprosy reactions in a tertiary hospital in Delhi. *J. Dermatol.* **31**, 898–903 (2004).
- 6 C. Lienhardt, P. E. Fine, Type 1 reaction, neuritis and disability in leprosy. What is the current epidemiological situation? *Lepr. Rev.* **65**, 9–33 (1994).
- 7 V. M. Fava, C. Sales-Marques, A. Alcaïs, M. O. Moraes, E. Schurr, Age-dependent association of *TNFSF15/TNFSF8* variants and leprosy type 1 reaction. *Front. Immunol.* **8**, 155 (2017).
- 8 V. M. Fava *et al.*, Association of *TNFSF8* regulatory variants with excessive inflammatory responses but not leprosy per se. *J. Infect. Dis.* **211**, 968–977 (2015).
- 9 J. Gaschignard *et al.*, Pauci- and multibacillary leprosy: Two distinct, genetically neglected diseases. *PLoS Negl. Trop. Dis.* **10**, e0004345 (2016).
- 10 C. G. Weindel *et al.*, LRRK2 regulates innate immune responses and neuroinflammation during *Mycobacterium tuberculosis* infection. *bioRxiv*:10.1101/699066. (26 July 2019).
- 11 A. Härtlova *et al.*, LRRK2 is a negative regulator of *Mycobacterium tuberculosis* phagosome maturation in macrophages. *EMBO J.* **37**, 37 (2018).

ANNEXE 5 : Article 3

The complex pattern of genetic associations of leprosy with HLA class I and class II alleles can be reduced to four amino acid positions

Monica Dallmann-Sauer, Vinicius M Fava, Chaïma Gzara, Marianna Orlova, Nguyen Van Thuc, Vu Hong Thai, Alexandre Alcaïs, Laurent Abel, Aurélie Cobat, Erwin Schurr

Publié dans PLoS Pathogen,

2020 Aug 10;16(8):e1008818.

doi: 10.1371/journal.ppat.1008818.

RESEARCH ARTICLE

The complex pattern of genetic associations of leprosy with HLA class I and class II alleles can be reduced to four amino acid positions

Monica Dallmann-Sauer^{1,2,3}, Vinicius M. Fava^{1,2}, Chaïma Gzara^{4,5}, Marianna Orlova^{1,2,3}, Nguyen Van Thuc⁶, Vu Hong Thai⁶, Alexandre Alcaïs^{4,5,7}, Laurent Abel^{4,5,7}, Aurélie Cobat^{4,5,7}, Erwin Schurr^{1,2,3*}

1 Program in Infectious Diseases and Immunity in Global Health, The Research Institute of the McGill University Health Centre, Montreal, Quebec, Canada, **2** McGill International TB Centre, McGill University, Montreal, Quebec, Canada, **3** Departments of Human Genetics and Medicine, Faculty of Medicine, McGill University, Montreal, Quebec, Canada, **4** Laboratory of Human Genetics of Infectious Diseases, Necker Branch, INSERM UMR1163, Paris, France, **5** Université de Paris, Imagine Institute, Paris, France, **6** Hospital for Dermato-Venerology, Ho Chi Minh City, Vietnam, **7** St. Giles Laboratory of Human Genetics of Infectious Diseases, Rockefeller Branch, The Rockefeller University, New York, New York, United States of America

* erwin.schurr@mcgill.ca



OPEN ACCESS

Citation: Dallmann-Sauer M, Fava VM, Gzara C, Orlova M, Van Thuc N, Thai VH, et al. (2020) The complex pattern of genetic associations of leprosy with HLA class I and class II alleles can be reduced to four amino acid positions. *PLoS Pathog* 16(8): e1008818. <https://doi.org/10.1371/journal.ppat.1008818>

Editor: Thomas R. Hawn, University of Washington, UNITED STATES

Received: May 6, 2020

Accepted: July 16, 2020

Published: August 10, 2020

Copyright: © 2020 Dallmann-Sauer et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Data Availability Statement: All relevant data are within the manuscript and its Supporting Information files.

Funding: This work was supported by grants from the Canadian Institutes of Health Research (CIHR) to E.S. (FDN-143332) and from l'Ordre de Malte to A.A. and E.S. This research was supported through resource allocation in the Cedar high performance computing cluster by Compute Canada (<https://www.computeCanada.ca/>) and WestGrid (<https://www.westgrid.ca/>)

Abstract

Leprosy is a chronic disease caused by *Mycobacterium leprae*. Worldwide, more than 200,000 new patients are affected by leprosy annually, making it the second most common mycobacterial disease after tuberculosis. The MHC/HLA region has been consistently identified as carrying major leprosy susceptibility variants in different populations at times with inconsistent results. To establish the unambiguous molecular identity of classical HLA class I and class II leprosy susceptibility factors, we applied next-generation sequencing to genotype with high-resolution 11 HLA class I and class II genes in 1,155 individuals from a Vietnamese leprosy case-control sample. HLA alleles belonging to an extended haplotype from *HLA-A* to *HLA-DPB1* were associated with risk to leprosy. This susceptibility signal could be reduced to the *HLA-DRB1*10:01~HLA-DQA1*01:05* alleles which were in complete linkage disequilibrium (LD). In addition, haplotypes containing *HLA-DRB3~HLA-DRB1*12:02* and *HLA-C*07:06~HLA-B*44:03~HLA-DRB1*07:01* alleles were found as two independent protective factors for leprosy. Moreover, we replicated the previously associated *HLA-DRB1*15:01* as leprosy risk factor and *HLA-DRB1*04:05~HLA-DQA1*03:03* as protective alleles. When we narrowed the analysis to the single amino acid level, we found that the associations of the HLA alleles were largely captured by four independent amino acids at HLA-DRβ1 positions 57 (D) and 13 (F), HLA-B position 63 (E) and HLA-A position 19 (K). Hence, analyses at the amino acid level circumvented the ambiguity caused by strong LD of leprosy susceptibility HLA alleles and identified four distinct leprosy susceptibility factors.

www.westgrid.ca/). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing interests: The authors have declared that no competing interests exist.

Author summary

Despite global efforts to eliminate leprosy over the past 25 years, more than 200,000 new cases are reported annually, and leprosy still represents a major public health problem in endemic regions. Leprosy presents a strong link with the host genetic background. The most significant susceptibility factors are located in the MHC region and likely involve classical HLA genes. However, the molecular identity of the HLA class I/II-leprosy risk factor(s) has been a matter of longstanding scientific dispute. By conducting a comprehensive sequenced-based analysis of HLA class I and class II genes, we are able to provide a unifying view of the complex relationship of leprosy susceptibility and HLA alleles. In addition, we show that four amino acid polymorphisms in HLA-DR β 1, HLA-B and HLA-A are sufficient to explain the majority of leprosy-HLA associations which opens the way for select protein-HLA peptide binding studies.

Introduction

Leprosy is a chronic human disease of the skin and peripheral nerves that results from an infection with *Mycobacterium leprae*. Permanent nerve damage and disabilities can occur in leprosy patients, mainly due to delayed diagnoses and exacerbated inflammatory episodes known as type-1 reactions (T1R). Human genetic factors strongly influence susceptibility to leprosy and more than 30 loci throughout the genome have been associated with leprosy phenotypes (rev in [1]). Genome-wide association studies (GWAS) have identified single nucleotide polymorphisms (SNPs) within the Major Histocompatibility Complex (MHC) on chromosome 6p21 as the most significantly leprosy-associated genetic variants [2–7]. Divided into three classes, the MHC region harbors hundreds of genes including the classical Human Leukocyte Antigen (HLA) genes of the MHC class I and class II regions. These genes encode transmembrane receptors that present short antigenic peptides to T-cells, and, in the case of class I molecules, to NK-cells and specialized cells of the monocyte lineage (rev in [8]). The highly polymorphic class I/II genes present several bi-allelic and multi-allelic amino acid variants that impact on the strength and specificity of peptide binding. Immunologically defined HLA alleles are, on the molecular level, multi-variant haplotypes of specific genes. Molecular defined HLA alleles are now generally used for HLA genetics studies since they provide a higher resolution over immunologically-defined alleles.

Historically, a large number of HLA class I and class II alleles have been associated with leprosy or its clinical subtypes in different ethnicities (rev in [9]). Association analyses of HLA genes are challenging due to the difficulty of high-resolution allele genotyping and the complex linkage disequilibrium (LD) pattern among HLA alleles. Since earlier studies were typically conducted employing low-resolution candidate gene approaches that did not allow to adjust on the impact of complex LD pattern across the HLA region, no unifying picture of class I and class II leprosy risk alleles emerged. More recently, using HLA allele imputation, the *HLA-DRB1*15:01* allele was identified as the major driver of the MHC association signal for leprosy *per se* in the Chinese population [3, 4]. In addition to *HLA-DRB1*, variants in *HLA-DQA1* and *HLA-C* were also associated with leprosy in the Chinese population [10]. However, population-specific effects of HLA genes were suggested by two association scans conducted in Vietnamese and Indian populations that identified association signals that were independent of *HLA-DRB1* alleles and implicated a class I *HLA-C* allele [11, 12]. Hence, while the major genetic effects of leprosy susceptibility were associated with class II genes in Chinese

leprosy patients it seemed possible that in Non-Chinese patients, class I alleles had a stronger effect on leprosy susceptibility.

In the present study, we used accurate HLA typing by next-generation sequencing (NGS) of three class I (*HLA-A*, *HLA-C* and *HLA-B*) and eight class II HLA genes (*HLA-DRB1*, *HLA-DRB3*, *HLA-DRB4*, *HLA-DRB5*, *HLA-DQA1*, *HLA-DQB1*, *HLA-DPA1* and *HLA-DPB1*) in a Vietnamese case-control sample. Gene-centric analyses revealed an intricate pattern of associated class I and class II alleles in line with the complex historical results. However, by taking into account allele correlations across the HLA region and by considering individual amino acid variants we showed that the complexity of associations could be reduced to the presence of specific amino acids at only four positions in the *HLA-DRB1*, *HLA-B* and *HLA-A* genes.

Materials and methods

Ethics statement

The study received approval from the regulatory authorities of Ho Chi Minh City, Vietnam (So3813/UB-VX and 4933/UBND-VX) and the Research Ethics Board of the Research Institute at McGill University Health Centre in Montreal, Canada (REC98-041). Written informed consent was obtained from all participants in the study. For children, assent was given by subjects, and written informed consent was obtained from parents or guardians.

Population sample

For the present study, we enrolled 1,164 Vietnamese individuals recruited from Ho Chi Minh City in Vietnam, as previously described [7, 13, 14]. After quality control, 1,155 participants were included in the association study for leprosy *per se*, corresponding to 687 leprosy cases and 468 healthy controls (Table 1). Healthy controls were subjects living in the same districts as leprosy cases with no family history of leprosy up to third-degree relatives [7]. Due to low prevalence of leprosy, it is possible that the inclusion of genetically susceptible subjects into the group of controls will result in an underestimate of the strength of the genetic effect. The present WHO classification recognizes two leprosy subtypes: paucibacillary (PB) and multibacillary (MB) leprosy [15]. In the present dataset, 491 (71.5%) and 196 (28.5%) cases were classified as MB and PB leprosy respectively. All leprosy cases were included in the analysis of leprosy subtype using the WHO clinical classification as phenotype (Table 1). Finally, a subset of 460 borderline leprosy cases was selected for the analysis of T1R and was previously described [16]. This population sample included 230 T1R-free and 230 T1R-affected leprosy patients that were matched for gender, Ridley and Jopling subtypes and age at leprosy diagnosis (Table 1).

HLA typing by next-generation sequencing

In the present study we aimed to accurately type HLA alleles up to the second field resolution (HLA proteins). Hence, we conducted HLA typing by NGS in all 1,164 individuals for 11 HLA genes, including three class I genes—*HLA-A*, *HLA-C* and *HLA-B*—and eight loci from the class II region—*HLA-DRB1/3/4/5*, *HLA-DQA1*, *HLA-DQB1*, *HLA-DPA1* and *HLA-DPB1*. We used HoloType HLA 24/11 and 96/11 kit v2 (Omixon Biocomputing Ltd, Budapest, Hungary) to amplify the genes and generate sequencing libraries [17]. The libraries were sequenced on a MiSeq platform (Illumina, San Diego, CA, USA) in a standard flow cell using MiSeq Reagent Kit v2, 500 cycle (Illumina). The sequencing data composed of 250 bp paired-end reads was analyzed for HLA typing using Omixon HLA Twin software v2.1.4 (Omixon) with default settings and using the IPD-IMGT/HLA release 3.29 reference [18, 19]. For typing of *HLA-DRB3/*

Table 1. Clinical characteristics of the Vietnamese individuals included in leprosy *per se*, leprosy subtype polarization and type-1 reaction association analyses with HLA genes.

		Leprosy <i>per se</i>		Polarization to leprosy subtypes		Type 1 reaction ^b	
		Cases	Controls	MB	PB	T1R-affected	T1R-free
Number of samples	N	687	468	491	196	230	230
Gender	Male	480 (69.9%)	257 (54.9%)	354 (72.1%)	126 (64.3%)	163 (70.9%)	159 (69.1%)
	Female	207 (30.1%)	211 (45.1%)	137 (27.9%)	70 (35.7%)	67 (29.1%)	71 (30.9%)
Age ^a (yr)	Mean (SD)	19.7 (±6.7)	30.1 (±8.3)	20.2 (±6.7)	18.4 (±6.5)	19.8 (±5.5)	20.6 (±4.7)
	Median	20	29	21	18	20	21
Subtype (WHO classification)	MB	491 (71.5%)	-	491 (100%)	-	184 (80%)	180 (78.3%)
	PB	196 (28.5%)	-	-	196 (100%)	46 (20%)	50 (21.7%)
Subtype (Ridley and Jopling classification)	TT	88 (12.8%)	-	1 (0.2%)	87 (44.4%)	-	-
	BT	121 (17.6%)	-	14 (2.9%)	107 (54.6%)	51 (22%)	51 (22%)
	BB	214 (31.2%)	-	212 (43.2%)	2 (1.0%)	99 (43%)	99 (43%)
	BL	174 (25.3%)	-	174 (35.4%)	-	80 (35%)	80 (35%)
	LL	90 (13.1%)	-	90 (18.3%)	-	-	-
Type 1 reaction	Yes	244 (35.5%)	-	196 (39.9%)	48 (24.5%)	230 (100%)	-
	No	436 (63.5%)	-	292 (59.5%)	144 (73.5%)	-	230 (100%)
	Unknown	7 (1.0%)	-	3 (0.6%)	4 (2.0%)	-	-

BB: borderline-borderline; BL: borderline-lepromatous; BT: borderline-tuberculoid; LL: lepromatous-lepromatous; MB: multibacillary; PB: paucibacillary; SD: standard deviation; T1R: Type-1 reaction; TT: tuberculoid-tuberculoid; yr: years.

a. Age at recruitment is presented for the healthy control group in leprosy *per se* analysis, while age at leprosy diagnosis is presented in the remaining groups.

b. A subset of borderline leprosy cases was selected for T1R analysis, where T1R-free and T1R-affected groups were matched for age, gender and leprosy subtype.

<https://doi.org/10.1371/journal.ppat.1008818.t001>

4/5, LD option was turned on for the software to use LD information in publicly available databases for determining if these loci were hemizygous. Precision of the employed method has been reported to range from 95.6% in *HLA-DQB1* to 99.53% in *HLA-A* for four-digit alleles based on data from 424 samples originated from 197 reference cell lines [20]. In 253 samples, the NGS-based HLA genotype calls have been reported to be 97.4% concordant with the high-resolution genotypes derived from a combination of Sanger-based typing and sequence-specific primer technology [17]. In our dataset, HLA typing failed in nine out of the 1,164 sequenced samples (0.77%) and those were excluded from the study. We obtained call-rates higher than 99% for HLA class I genes, while the call-rates for class II loci ranged from 90.6% in *HLA-DQB1* to more than 98% in *HLA-DRB1/3/4/5* genes. In total, we identified 104 and 252 alleles at the first (two-digit) and second field (four-digit) resolution, respectively.

HLA alleles and amino acid markers

For the association analyses, binary (biallelic) markers based on the presence (P) / absence (A) of each specific HLA allele or single amino acid were generated as described elsewhere [21]. These binary markers presented three possible genotypes: homozygous for the presence of the allele/amino acid (PP), heterozygous (AP) and homozygous for the absence of the allele/amino acid (AA). In the analysis of HLA alleles, binary markers were created for each two-digit and four-digit allele from the 11 HLA loci that were sequenced. In addition, we created binary markers for the presence/absence of *HLA-DRB3/4/5* genes regardless of their alleles. Gene and allele markers for *HLA-DRB3/4/5* were created using R software v3.4.4, while markers for the remaining HLA loci were generated using MakeReference command from SNP2HLA software v1.0.3 [21, 22].

HLA amino acid markers corresponded to the minor/major amino acids in biallelic protein positions (one marker per position) and binary markers for the presence/absence of each amino acid residue in multiallelic positions (three or more binary markers for single amino acid residues per multiallelic position). The amino acid markers were generated with MakeReference command from SNP2HLA software using an HLA protein reference panel with protein sequences from IPD-IMGT/HLA release 3.34 [19]. The reference panel was manually curated to include new HLA alleles detected by NGS-based typing and to complete gaps in a few known alleles from regions for which we had obtained sequencing data. The positions that we were unable to complete were coded as missing data in the MakeReference output file. A total of 491 single amino acid markers were created for the polymorphic positions in HLA-A, HLA-C, HLA-B, HLA-DR β 1 and HLA-DQ α 1 proteins.

HLA alleles or amino acids with call-rate lower than 90%, with minor allele frequency (MAF) lower than 1% in the 1,155 individuals or with deviations from Hardy-Weinberg Equilibrium in the control group with $P < 0.01$ were excluded from the analysis. In total, 198 binary markers for HLA alleles were used in the association analyses, including three markers for *HLA-DRB3/4/5* genes, 72 two-digit alleles and 123 four-digit HLA alleles. In the amino acid analysis, 424 markers passed filtering and were included in the association analyses. The tested amino acid markers corresponded to 192 biallelic amino acid changes and 232 binary markers for single residues in 71 multiallelic protein positions. The genotypes of these alleles and amino acid binary markers are available in the [S1 Data](#).

Statistical analyses

To account for the strong correlation caused by linkage disequilibrium (LD) among HLA alleles and among HLA amino acid markers, GEC (Genetic Type I Error Calculator) v0.2 was used to find the effective number of independent tests [23]. For that, the genomic positions of all markers were coded as being consecutives to avoid exclusion of markers that were at the same position. Although efficiency of programs for estimation of the effective number of independent tests has not been demonstrated specifically for HLA alleles/amino acids binary markers, we selected to use GEC as it has been shown to be robust to variable LD patterns across the genome [23]. Using GEC, the 198 HLA alleles markers were determined to be equivalent to 121.41 independent tests in the total population sample. Bonferroni correction for multiple testing was applied using the effective number of allele markers. This yielded a significance P cut-off of 4.12×10^{-4} ($0.05/121.41$), which was used in the univariable association analysis of HLA alleles. The effective number of independent amino acid tests was determined as 178 using GEC software. Hence, $P < 2.81 \times 10^{-4}$ ($0.05/178$) was used as multiple-testing cut-off in the univariable analysis of single amino acids. Univariable association analyses of HLA alleles or amino acids and leprosy susceptibility were done by logistic regression under an additive model as implemented in PLINK v1.9 [24]. For that, the three genotypes of the biallelic/binary markers (e.g. AA, AP and PP of an HLA allele marker) were coded as 0, 1 and 2 to reflect the minor allele dosage. In this model, the direction of the regression coefficient [presented as odds ratio (OR)] represented the effect of each additional minor allele. Forward conditional association analysis and pairwise reciprocal conditional analysis were also done in the additive model using PLINK, by the inclusion of the conditioned marker(s) into the model as covariate (s). In the conditional analyses, the $OR_{\text{conditional}}$ represented the effect of each additional minor allele of the tested marker whilst controlling for the covariate(s). $P_{\text{conditional}} < 0.01$ was used as cut-off for independent association signals in the multivariable analyses. When the tested marker and a covariate presented collinearity due to complete LD, results were reported as not available (NA). Gender distribution between cases and controls was tested using

Pearson's Chi-squared test with Yates' continuity correction. We included gender as a covariate in the univariable and multivariable association analyses for leprosy *per se*, since it is a known leprosy risk factor that was differently distributed between cases and controls ($P_{\text{Chi-square}} = 3.84 \times 10^{-7}$, Table 1). To assess the distribution of HLA alleles among leprosy cases based on two disease endophenotypes, we used logistic regression under an additive model with no covariate included, where MB were compared to PB cases and T1R-affected patients were compared to T1R-free leprosy cases [15].

We used Akaike Information Criterion (AIC), a model selection method, to compare the allele-based and amino acid-based models and select the model that best represented the data [25]. AIC was calculated for a subset of 1,136 samples with no missing genotypes among six tag allele and amino acid markers (98% of the samples), using the LOGISTIC procedure in SAS v 9.4 (SAS Institute, Cary, North Carolina, USA). The model presenting the lowest AIC (AIC_{min}) was selected as presenting the best support among the tested models. Delta (Δ) AIC was calculated by subtracting the AIC of each model by the AIC_{min} to compare the two models. Models with Δ AIC of 0–2, 3–9 and ≥ 10 were considered to present substantial support, considerably less support and no additional support regarding the best model, respectively [26].

Haplotype structure and LD

For LD analysis of associated markers, pairwise correlation coefficient r^2 and D' were calculated in healthy controls using PLINK v1.9. Phased haplotype between HLA amino acids from the same gene was determined by the NGS-based HLA typing method. For markers in different loci, we used Beagle v 3.0.4 to estimate the phased haplotypes [27]. OR and P -values for the haplotypic analyses were calculated using epitools R package v 0.5–10.1 [28]. Finally, Disentangler software was used for the haplotype structure visualization of the HLA genes presenting at least one four-digit allele associated with leprosy *per se* [29, 30].

3D visualization of HLA molecules

To identify the location of the four key leprosy-associated HLA amino acids in the protein structure, the three-dimensional structure of the HLA molecules was analyzed using UCSF Chimera v 1.13.1. [31]. HLA-DR, HLA-B and HLA-A structures were based on Protein Data Bank (PDB): 3PDO, 1SYS and 3UTQ, respectively [32–34]. These entries correspond to *HLA-DR1*, *HLA-B*44:03* and *HLA-A*02:01* alleles respectively.

Results

HLA allele association with leprosy

We conducted HLA genotyping by NGS of 11 HLA genes in 687 leprosy cases and 468 healthy controls from Vietnam (Table 1). In the first part of our study, 198 HLA alleles were tested for association with leprosy *per se* (S1 Table). We found 20 HLA alleles in the *HLA-A*, *HLA-C*, *HLA-B*, *HLA-DRB3*, *HLA-DRB1*, *HLA-DQA1*, *HLA-DQB1* and *HLA-DPB1* genes to be significantly associated with leprosy *per se* under an additive model (Table 2). As previous studies had reported HLA alleles to be associated with leprosy clinical subtypes (rev in [9]), we analyzed the allele distribution between MB and PB leprosy patients and between T1R free and affected cases (Table 1). We observed no significant differences of HLA alleles for MB-PB leprosy clinical subtype or T1R state (S1 Table).

Among the leprosy *per se* associated HLA alleles, *HLA-DQA1*01:05* presented the most significant evidence of association (OR = 3.11, $P = 7.61 \times 10^{-10}$) followed by *HLA-DRB1*10:01*

Table 2. HLA alleles significantly associated with risk or protection from leprosy *per se* in Vietnam.

MHC region	Resolution	HLA allele	HLA allele frequency		Univariable analysis ^a			Signal #
			Leprosy cases	Healthy controls	OR	(95% CI)	P-value ^b	
Class II	Four-digit	<i>HLA-DQA1*01:05</i>	12.7%	4.5%	3.11	(2.16–4.45)	7.61×10^{-10}	1
Class II	Four-digit	<i>HLA-DRB1*10:01</i>	12.6%	4.5%	3.08	(2.15–4.42)	1.02×10^{-9}	1
Class II	Two-digit	<i>HLA-DRB1*10</i>	12.6%	4.5%	3.08	(2.15–4.42)	1.02×10^{-9}	1
Class I	Four-digit	<i>HLA-C*15:05</i>	10.4%	3.6%	3.08	(2.08–4.58)	2.30×10^{-8}	1
Class I	Four-digit	<i>HLA-B*07:05</i>	10.3%	3.6%	3.09	(2.08–4.60)	2.41×10^{-8}	1
Class II	Four-digit	<i>HLA-DQB1*05:01</i>	18.7%	9.3%	2.25	(1.68–3.02)	5.26×10^{-8}	1
Class I	Two-digit	<i>HLA-C*15</i>	14.0%	6.5%	2.29	(1.68–3.12)	1.34×10^{-7}	1
Class I	Two-digit	<i>HLA-B*07</i>	13.5%	6.6%	2.20	(1.61–2.99)	6.52×10^{-7}	1
Class II	Two-digit	<i>HLA-DQA1*01</i>	43.1%	32.8%	1.56	(1.31–1.87)	8.34×10^{-7}	1
Class I	Four-digit	<i>HLA-A*29:01</i>	10.9%	5.6%	2.10	(1.49–2.95)	2.09×10^{-5}	1
Class I	Two-digit	<i>HLA-A*29</i>	10.9%	5.6%	2.10	(1.49–2.95)	2.09×10^{-5}	1
Class II	Two-digit	<i>HLA-DQB1*05</i>	29.9%	21.6%	1.55	(1.25–1.92)	7.11×10^{-5}	1
Class II	Two-digit	<i>HLA-DPBI*104</i>	6.6%	2.5%	2.67	(1.65–4.34)	7.14×10^{-5}	1
Class II	Four-digit	<i>HLA-DPBI*104:01</i>	6.5%	2.5%	2.64	(1.63–4.29)	8.81×10^{-5}	1
Class I	Four-digit	<i>HLA-C*07:06</i>	1.2%	3.9%	0.32	(0.18–0.57)	9.95×10^{-5}	2
Class I	Four-digit	<i>HLA-B*44:03</i>	1.3%	3.7%	0.35	(0.20–0.61)	2.51×10^{-4}	2
Class I	Two-digit	<i>HLA-B*44</i>	1.5%	3.8%	0.37	(0.22–0.64)	3.83×10^{-4}	2
Class II	Two-digit	<i>HLA-DRB1*12</i>	23.1%	29.7%	0.69	(0.57–0.84)	1.80×10^{-4}	3
Class II	Four-digit	<i>HLA-DRB1*12:02</i>	23.0%	29.4%	0.69	(0.57–0.84)	2.59×10^{-4}	3
Class II	Gene	<i>HLA-DRB3</i>	39.9%	47.2%	0.72	(0.60–0.86)	3.78×10^{-4}	3

CI: confidence interval; OR: odds ratio.

a) Results in the additive model adjusted for gender. OR (95% CI) refer to the presence of the allele.

b) P-value threshold for multiple testing was set as 4.12×10^{-4} for the univariable analysis of HLA alleles.

<https://doi.org/10.1371/journal.ppat.1008818.t002>

(OR = 3.08, $P = 1.02 \times 10^{-9}$, Table 2). Of note, the *HLA-DRB1*10:01* and *HLA-DQA1*01:05* alleles were statistically equivalent due to complete LD ($r^2 = 1$). The small difference in their strength of association was due to differences in missing genotypes. To test for leprosy associations independent of *HLA-DRB1*10:01*~*HLA-DQA1*01:05*, we conducted a forward conditional association analysis of the 20 significantly associated alleles until markers presented $P_{\text{conditional}} \geq 0.01$. When *HLA-DQA1*01:05* entered as first independent genetic variable in the model, all HLA leprosy risk alleles lost significance (S2 Table). This indicated that the associations of risk alleles were almost entirely captured by *HLA-DRB1*10:01*~*HLA-DQA1*01:05*. We analyzed the haplotype structure of the four-digit HLA alleles and found that the HLA class I and class II leprosy risk alleles belonged to a long-range haplotype spanning from *HLA-A* to *HLA-DPBI*, with a frequency of 4.37% in cases and 1.5% in controls (haplotype *HLA-A*29:01*~*HLA-C*15:05*~*HLA-B*07:05*~*HLA-DRB1*10:01*~*HLA-DQA1*01:05*~*HLA-DQB1*05:01*~*HLA-DPBI*104:01*, Fig 1 and S1 Fig).

After adjusting on the effect of *HLA-DQA1*01:05* in the forward conditional analysis, significant associations were still observed for the alleles associated with protection from leprosy and the lowest P-value was found for the protective class I allele *HLA-C*07:06* (OR_{conditional} = 0.33, $P_{\text{conditional}} = 2.35 \times 10^{-4}$, S2 Table). When *HLA-C*07:06* entered in the forward conditional analysis, *HLA-B*44:03* and *HLA-B*44* lost significance while a residual effect was still observed for the remaining protective alleles (S2 Table). *HLA-B*44:03* and *HLA-C*07:06* were in strong LD ($r^2 = 0.97$) and statistically equivalent in pairwise multivariate analysis (S3 Table). Hence, the second independent association signal was due to *HLA-C*07:06*~*HLA-B*44:03*

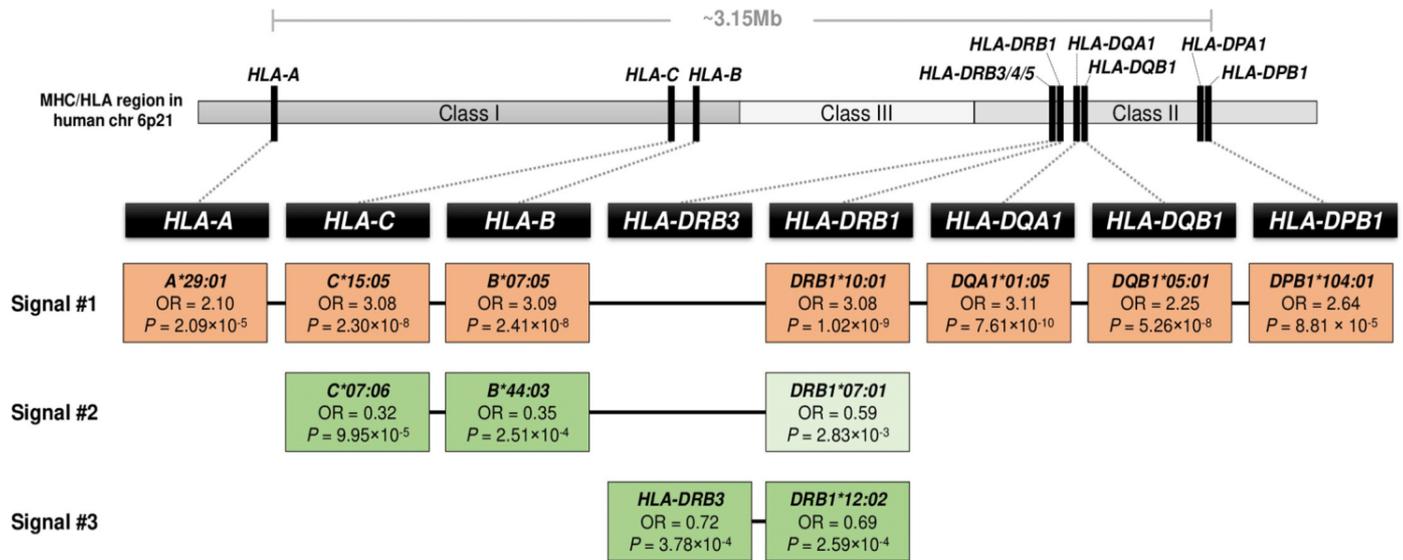


Fig 1. HLA class I and class II alleles with study-wide significance for association with leprosy. A schematic representation of the MHC/HLA region highlighting the genes sequenced in the present study is shown on top. The four-digit HLA alleles associated with risk and protection from leprosy *per se* are shown in orange and green, respectively. The *HLA-DRB1* allele with suggestive significance captured by the alleles in signal #2 is indicated by a box with lighter color. Odds ratios (OR) and *P*-values presented in the boxes refer to the results of univariable association analyses under an additive model (Table 2 and S1 Table). HLA alleles that belong to the same association signal based on the forward and pairwise conditional analyses are connected by lines (S2 and S3 Tables). The signals indicated common haplotypes of the corresponding alleles (S1 Fig).

<https://doi.org/10.1371/journal.ppat.1008818.g001>

(Fig 1). After adjustment for *HLA-DQA1*01:05* and *HLA-C*07:06*, the two-digit class II allele *HLA-DRB1*12* presented the lowest conditional *P*-value (OR_{conditional} = 0.71, *P*_{conditional} = 1.01 × 10⁻³, S2 Table). The association signal of *HLA-DRB1*12* was primarily driven by the common *HLA-DRB1*12:02* allele which had a statistically equivalent effect with the presence of the *HLA-DRB3* gene (S3 Table) making *HLA-DRB3* ~ *HLA-DRB1*12:02* the third independent association signal in our dataset (Fig 1).

HLA amino acid associations with leprosy

To disentangle the HLA association signals and assess the contribution of individual HLA amino acid polymorphisms, we tested if specific amino acid substitutions in HLA proteins, rather than the multi-amino acid HLA alleles, were associated with leprosy. A total of 424 single amino acids belonging to the HLA-A, HLA-C, HLA-B, HLA-DRβ1 and HLA-DQα1 proteins were selected for this analysis (S4 Table). In univariable analyses, 64 amino acids at 55 protein positions were significantly associated with leprosy *per se* (Fig 2A and S4 Table). The presence of aspartic acid at HLA-DRβ1 position 57 (HLA-DRβ1 57D, OR = 1.81, *P* = 1.81 × 10⁻¹⁰) was the most significant association (Fig 2A). In pairwise reciprocal conditional analysis for HLA-DRβ1 57D versus each of the remaining 63 amino acids, a significant residual effect was consistently detected for HLA-DRβ1 57D (*P*_{conditional} ≤ 5.47 × 10⁻³). This observation indicated a unique contribution of this amino acid to leprosy risk.

To test for leprosy associations independent of HLA-DRβ1 57D among the amino acids with study-wide significance, we used forward conditional regression until amino acids reached *P*_{conditional} ≥ 0.01 (Fig 2B–2D). When we adjusted on HLA-DRβ1 57D, the strongest remaining risk signal was the presence of phenylalanine at HLA-DRβ1 position 13 (HLA-DRβ1 13F, OR_{conditional} = 1.69, *P*_{conditional} = 6.49 × 10⁻⁶, Fig 2B). However, HLA-DRβ1 13F and phenylalanine at HLA-DRβ1 position 31 (HLA-DRβ1 31F) were in complete LD (*r*² = 1) since presence of HLA-DRβ1 13F was always in haplotype with absence of HLA-DRβ1 31F

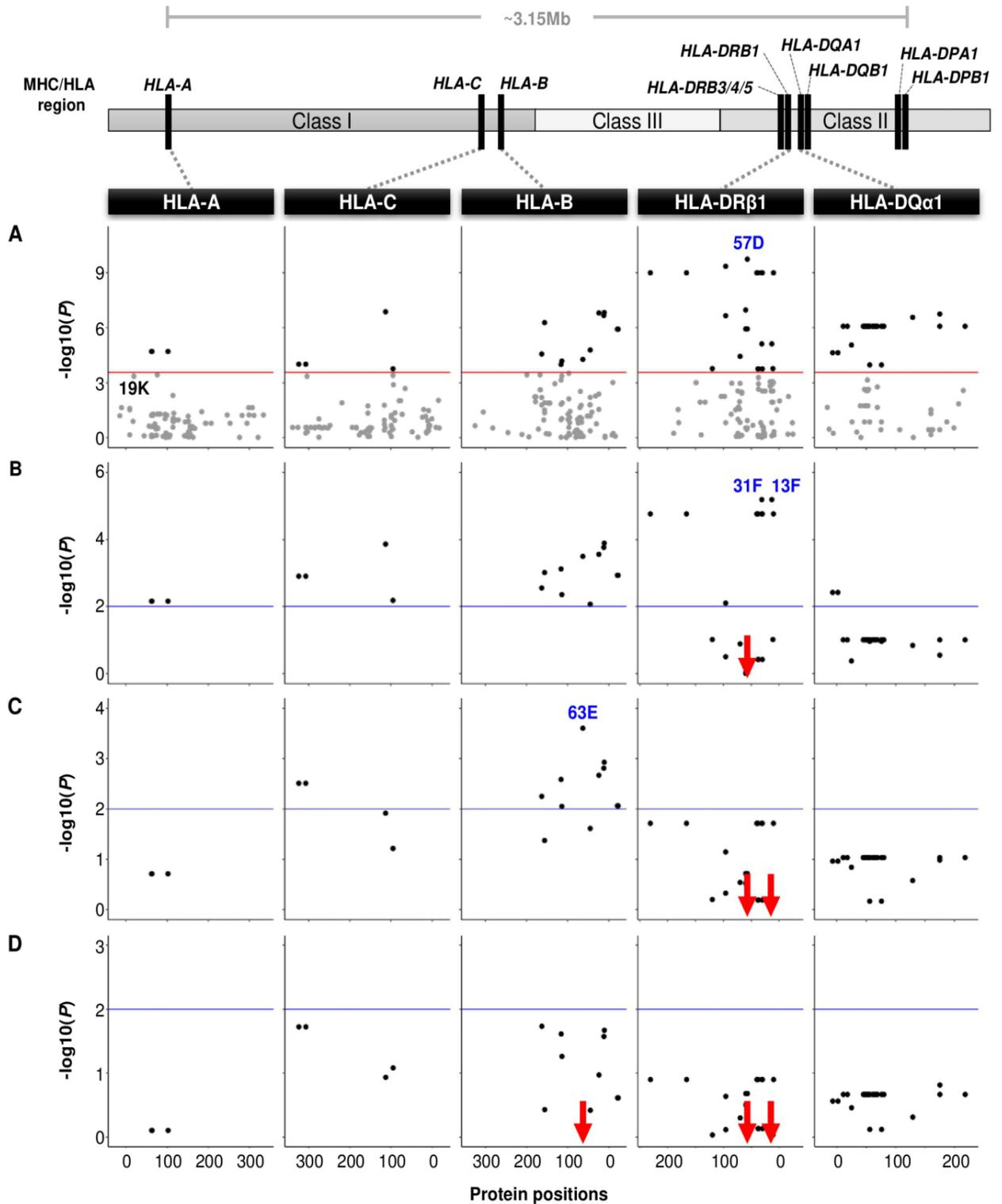


Fig 2. Forward conditional analysis of HLA amino acids with study-wide significance for association with leprosy. The figure presents the results for the genotypic association test under an additive model adjusted for gender and A) unconditioned for any marker (univariable analysis, S4 Table), B) conditioned on HLA-DRβ1 57D (forward step 1), C) conditioned on HLA-DRβ1 57D and HLA-DRβ1 13F (forward step 2) and D) conditioned on HLA-DRβ1 57D, HLA-DRβ1 13F and HLA-B 63E (forward step 3). A schematic representation of the MHC/HLA region highlighting the genes sequenced in the present study is shown on top. In the graphs, the y axes indicate the negative log₁₀ of the P-value for association of amino acids with leprosy *per se* and the x axes correspond to the amino acid position in the mature HLA proteins. The results for each HLA gene are shown by column and the dots represent single amino acids. The HLA loci and amino acid markers are ordered based on their genomic positions. The red line corresponds to the significance P-value threshold for multiple-testing of amino acid markers ($P = 2.81 \times 10^{-4}$) in the univariable analysis (A) and the blue lines represent a $P_{\text{conditional}}$ of 0.01 in the multivariable analyses (B-D). Dots shown in black reached the multiple testing P-value threshold in the univariable step (A) while grey dots were not significant (A). Only amino acids significantly associated in the univariable step (A) were included in the

multivariable analyses (B-D). The most significant association signal at each step is indicated by the name of the amino acid(s) shown in blue. Borderline-associated HLA-A 19K is presented in black in step (A). The locations of the markers used for stepwise conditioning are indicated by red arrows.

<https://doi.org/10.1371/journal.ppat.1008818.g002>

and vice-versa. As a consequence of the inverse relationship between phenylalanines at positions 13 and 31, presence of 31F was associated with protection from leprosy ($OR_{\text{conditional}} = 0.59$, $P_{\text{conditional}} = 6.49 \times 10^{-6}$). When we adjusted for risk HLA-DR β 1 57D and HLA-DR β 1 13F, all remaining class II amino acid markers in HLA-DR β 1 and HLA-DQ α 1 lost evidence of association (Fig 2C). Conversely, significant residual associations were still observed for HLA-C and HLA-B amino acids. At this step, glutamic acid at HLA-B position 63 displayed the lowest P -value (HLA-B 63E, $OR_{\text{conditional}} = 0.72$, $P_{\text{conditional}} = 2.50 \times 10^{-4}$). Inclusion of HLA-DR β 1 57D, HLA-DR β 1 13F and HLA-B 63E in the third step of the forward conditional analysis accounted for the full association signal of the remaining 61 HLA amino acid markers at $P_{\text{conditional}} \geq 0.01$ (Fig 2D). Interestingly, HLA-B position 63 and HLA-DR β 1 positions 13 and 57 are located in the peptide-binding grooves of the HLA-B and HLA-DR β 1 proteins (S2A and S2B Fig).

Additional HLA alleles and amino acids involved in leprosy risk

We further investigated the presence of additional independent signals, that were missed by the stringent study-wide significance cut-offs. For this, we adjusted each of the formally non-significant HLA alleles on the presence of *HLA-DQA1*01:05*, *HLA-C*07:06* and *HLA-DRB1*12* and identified seven HLA four-digit alleles with evidence ($P_{\text{conditional}} < 0.01$) for association with leprosy (S5 Table). Among the seven alleles, *HLA-A*11:02*, *HLA-B*38:02*, *HLA-DRB1*15:01* and *HLA-DRB1*04:05~HLA-DQA1*03:03* (both with $r^2 = 0.87$) had displayed a trend toward association in the univariable analysis with P -values ranging from 2.56×10^{-3} to 4.38×10^{-4} (S1 and S5 Tables). In the amino acid conditional analyses, lysine in HLA-A position 19 (HLA-A 19K) was the only amino acid that remained with a $P_{\text{conditional}}$ lower than 0.01 after removing the effect of HLA-DR β 1 57D, HLA-DR β 1 13F and HLA-B 63E ($OR_{\text{conditional}} = 0.55$, $P_{\text{conditional}} = 8.07 \times 10^{-3}$). HLA-A 19K presented a trend toward association as a protective amino acid in the univariable analysis ($P = 4.49 \times 10^{-4}$, Fig 2A).

Comparison of HLA allele-based and amino acid-based models

Next, we tested to what extent the amino acids captured the leprosy association of the HLA alleles. Except for *HLA-A*11:02*, the association signals of the study-wide significant and borderline-significant HLA alleles were accounted for when the analysis was adjusted on the HLA-DR β 1 57D, HLA-DR β 1 13F and HLA-B 63E amino acids (S5 and S6 Tables). However, to capture the association of *HLA-A*11:02* it was necessary to include HLA-A 19K in the conditional analysis (S5 Table). HLA-A 19K is located outside the protein-binding groove (S2C Fig) and is found exclusively in the *HLA-A*11:02* allele. Taking together, these results suggested that the four amino acids offered a good explanation for the observed HLA allele associations. To further validate this conclusion, we used the AIC to select the best multivariate model from the HLA allele and amino acid analyses. The four-amino acid model with gender, HLA-DR β 1 57D, HLA-DR β 1 13F, HLA-B 63E and HLA-A 19K presented the lowest AIC among the tested models (model A in Table 3). Conversely, the models including the three significant amino acids (gender + HLA-DR β 1 57D + HLA-DR β 1 13F + HLA-B 63E, model C) and the three significant HLA alleles (gender + *HLA-DQA1*01:05* + *HLA-C*07:06* + *HLA-DRB1*12*, model D) presented considerably less support (Δ AIC of 5 and 12, respectively; Table 3). Indeed, even after inclusion of *HLA-A*11:02* in the four-allele model (model B), the

Table 3. Comparison of HLA allele-based and HLA amino acid-based models using the Akaike Information Criterion (AIC).

	Models	df	Chi-Square (SCORE)	AIC	Δ AIC ^a
A	Gender + HLA-DR β 1 57D + HLA-DR β 1 13F + HLA-B 63E + HLA-A 19K	5	105.2	1436	0
B	Gender + <i>HLA-DQA1*01:05</i> + <i>HLA-C*07:06</i> + <i>HLA-DRB1*12</i> + <i>HLA-A*11:02</i>	5	100.6	1440	4
C	Gender + HLA-DR β 1 57D + HLA-DR β 1 13F + HLA-B 63E	4	98.7	1441	5
D	Gender + <i>HLA-DQA1*01:05</i> + <i>HLA-C*07:06</i> + <i>HLA-DRB1*12</i>	4	91.0	1448	12

AIC: Akaike Information Criterion, df: degrees of freedom.

a) Delta (Δ) AIC presents the difference between the AICs of each model with the model with the lowest AIC (in bold). Δ AIC of 0–2 represents substantial support, 3–9 shows considerably less support and models with Δ AIC \geq 10 presents essentially no support.

<https://doi.org/10.1371/journal.ppat.1008818.t003>

four-amino acid model still provided the best explanation of the data (Δ AIC = 4, Table 3). Hence, the four-amino acid model was the model best fitting the data.

Haplotype analysis of HLA amino acids

To better understand how the HLA-DR β 1 amino acid residues contributed to the association of the *HLA-DRB1* alleles, we analyzed the phased haplotype between HLA-DR β 1 13F and 57D based on the protein sequence of the four-digit HLA alleles (Fig 3A). HLA-DR β 1 position 13 and 57 were multi-allelic in the studied population with five and three additional amino acids besides 13F and 57D, respectively. Interestingly, the haplotype of the two risk residues was only present on the risk allele *HLA-DRB1*10:01* and two rare alleles, *HLA-DRB1*01:01* and *HLA-DRB1*01:02* (HLA-DR β 1 13F present ~ 57D present haplotype in Fig 3A). Due to their low frequency, the *HLA-DRB1*01* alleles were not included in the association analyses (see Methods). However, we noted that the *HLA-DRB1*01* group was more frequent in leprosy cases than in healthy controls (frequency of 0.7% and 0.3% respectively). On the other hand, both risk amino acids were absent in 13 *HLA-DRB1* alleles including the protective *HLA-DRB1*12* (HLA-DR β 1 13F absent ~ 57D absent haplotype in Fig 3A). This suggested the haplotypes HLA-DR β 1 13F present ~ 57D present and HLA-DR β 1 13F absent ~ 57D absent as main causes of the risk and protective effects of leprosy associated *HLA-DRB1* alleles.

In the HLA-B protein, position 63 was biallelic in our dataset with the risk asparagine (HLA-B 63N) as the major allele among cases and the protective glutamic acid (HLA-B 63E) as the major allele among controls (HLA-B 63E had frequencies of 46.3% in cases and 52.1% in controls). We analyzed the haplotypic effect between the amino acid at HLA-B position 63 with the major risk HLA-DR β 1 13F present ~ 57D present and the major protective HLA-DR β 1 13F absent ~ 57D absent haplotypes. For that, we estimated the phased haplotype between *HLA-B* and *HLA-DRB1*. We observed that the strength of association of the protective HLA-DR β 1 13F absent ~ 57D absent was increased by the presence of the protective HLA-B 63E on the same haplotype, while HLA-DR β 1 13F present ~ 57D present had a stronger risk effect when on the same haplotype with HLA-B 63N (Fig 3B).

*HLA-B*07:05* and *HLA-DRB1*10:01*, both from signal #1 in the HLA allele analysis (Fig 1), belonged to the risk combination HLA-B 63N ~ HLA-DR β 1 13F present ~ 57D present and were the majority of samples in this group (Fig 3B). The protective *HLA-B*44:03* from signal #2 (Fig 1) was mostly found on the background of *HLA-DRB1*07:01* haplotype in the Vietnamese population (S1 Fig) and *HLA-B*44:03* ~ *HLA-DRB1*07:01* haplotype belonged to the protective HLA-B 63E ~ HLA-DR β 1 13F absent ~ 57D absent group (Fig 3B). Interestingly, *HLA-DRB1*07:01* presented a trend toward association as a protective allele in the univariable analysis (OR = 0.59, $P = 2.83 \times 10^{-3}$). When we analyzed *HLA-B*44:03* and *HLA-DRB1*07:01*

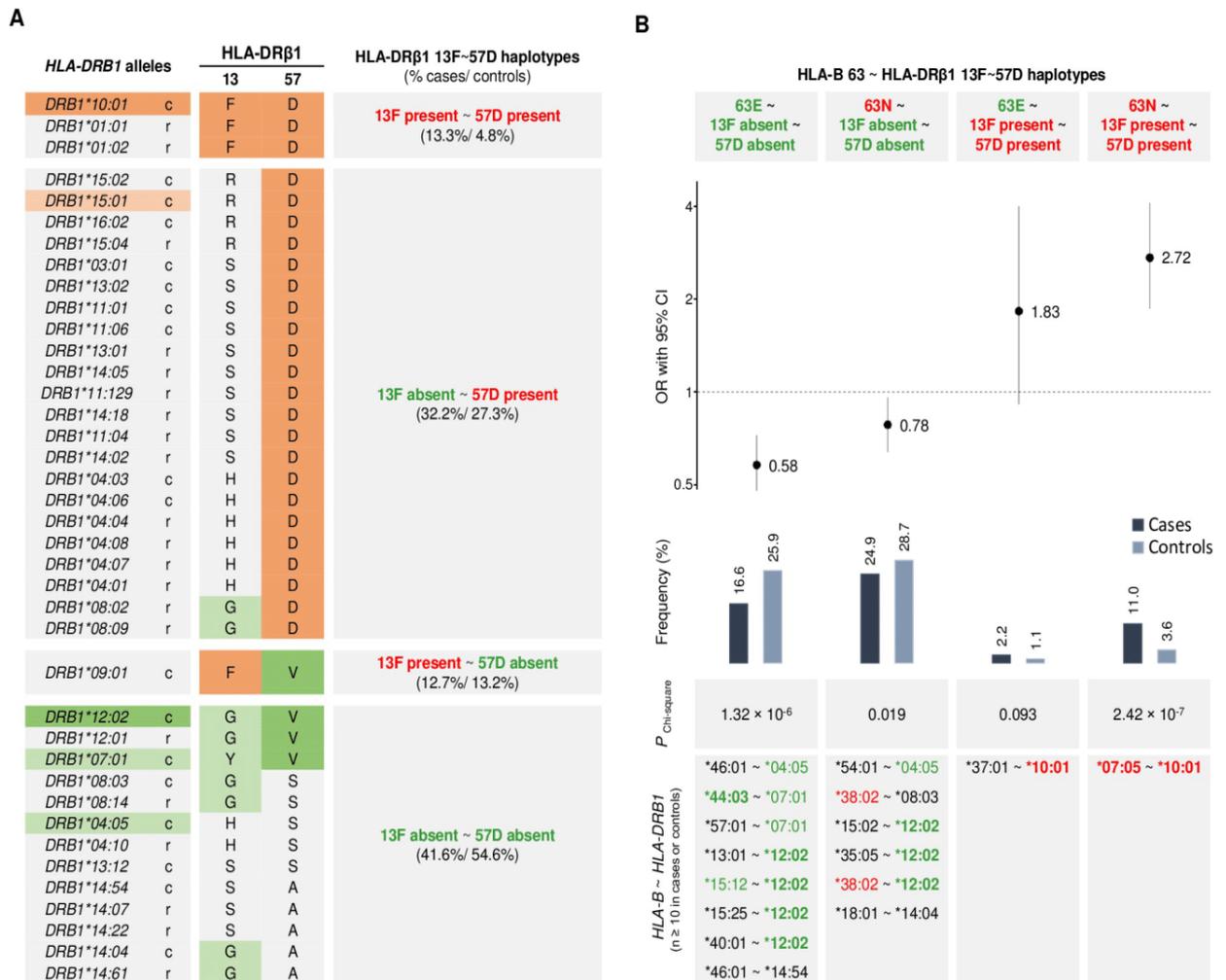


Fig 3. HLA-DRβ1 13F~57D~HLA-B 63N/E haplotypes correlated with the associated HLA-DRB1 and HLA-B alleles. **A)** Distribution of phased haplotypes of the risk amino acids HLA-DRβ1 13F and 57D among *HLA-DRB1* alleles. The left column presents the 39 four-digit *HLA-DRB1* alleles genotyped in the present Vietnamese sample. Rare (< 1%) and common (≥ 1%) alleles are indicated by r and c, respectively. The middle column presents the amino acid residues at HLA-DRβ1 positions 13 and 57 for each HLA allele. In the left and middle columns, the HLA alleles and amino acids with $P < 0.01$ are colored in orange or green if their presence corresponded to risk or protection from leprosy, respectively (S1 and S4 Tables). Markers with multiple-testing significance in the univariable analysis are indicated by a darker orange/green. In the right column, HLA alleles are grouped based on the haplotype of the presence/absence of HLA-DRβ1 13F and 57D. **B)** Contribution of the HLA-B 63N/E biallelic amino acid polymorphism to the association of HLA-DRβ1 13F present ~ 57D present and HLA-DRβ1 13F absent ~ 57D absent with leprosy. The results of the four estimated phased haplotypes are presented by columns. Odds ratios (OR), 95% confidence intervals (95% CI) and $P_{\text{Chi-square}}$ were determined for the four haplotypes using the pool of remaining haplotypes as reference (44.7% in cases and 40.4% in controls), except haplotypes with missing data (0.7% in cases and 0.3% in controls). Four-digit *HLA-B*~*HLA-DRB1* haplotypes belonging to each group and presenting counts higher or equal to ten chromosomes in cases and/or controls are listed at the bottom. HLA alleles and single amino acids that as binary markers presented $P < 0.01$ in the genotypic univariable analysis are colored in red and green when associated with risk or protection from leprosy, respectively. Markers that reached the multiple-testing thresholds are indicated in bold (S1 and S4 Tables).

<https://doi.org/10.1371/journal.ppat.1008818.g003>

by reciprocal conditional analysis we found that the *HLA-DRB1*07:01* suggestive signal was dependent on the significant *HLA-B*44:03* allele (S3 Table) and was included in signal #2 (Fig 1). Finally, haplotypes including the protective *HLA-DRB1*12:02* allele from signal #3 in the protective HLA-B 63E ~ HLA-DRβ1 13F absent~ 57D absent haplotype group consisted of different less common and non-significant *HLA-B* alleles with glutamic acid at position 63 (Fig 3B). These results indicated that the class I HLA-B amino acid change contributed to the association signal of the class II variants and illustrates how protective and susceptibility effects of classical HLA alleles are dependent on the presence/absence of specific amino acids.

Discussion

The MHC region has been consistently identified as chromosomal location of strong genetic leprosy risk factors, including in our recent Vietnamese GWAS where three independent association signals were identified in the MHC region [7]. While genetic studies have implicated non-classical MHC genes like *LTA* and *MICA* in leprosy susceptibility, the main effects have been assigned to classical class I and class II HLA genes [11, 35]. Employing candidate gene approaches, these genes have long been the object of genetic studies of leprosy, mainly trying to identify clinical subtype specific effects (rev in [9]). However, the extreme polymorphic nature of class I and class II HLA genes combined with extensive, population-specific LD have provided a major challenge for the replication of reported allele associations across studies. HLA imputation using GWAS data has been demonstrated to be an effective method for MHC fine mapping. However, this approach also presents some limitations. Imputation accuracy of low frequency HLA alleles is dependent on the similarity of the ethnicities between the pre-built reference panel and the studied population sample since alleles not present in the panel cannot be imputed. HLA imputation is limited to the genes present in the reference panel. Moreover, HLA allele resolution of the imputed data is defined by the resolution of the HLA typing method used to generate the reference panel. Hence, in the present study, we elected to perform high-resolution HLA typing by NGS to obtain accurate HLA genotypes (low ambiguity rate) independently to our GWAS data and pre-build reference panels. The use of NGS to genotype classical HLA genes in the Vietnamese population allowed us to gain a comprehensive understanding of the alleles implicated in leprosy susceptibility. Based on HLA allele analysis, we detected independent signals of association with risk and protection from leprosy of class I and class II alleles (Fig 1).

The strongest leprosy susceptibility factors in our study were the *HLA-DRB1*10:01* and *HLA-DQA1*01:05* alleles which are in complete LD in the Vietnamese population. These two alleles are tagging an extended haplotype which spans from *HLA-A* to *HLA-DPB1*, a distance of approximately 3.15 Mb (signal #1 in Fig 1). These findings are strongly supported by a study of three Indian leprosy cohorts and our family-based leprosy GWAS in Vietnam [5, 7]. In the Indian population samples, the authors reported a strong risk effect for leprosy of MHC class II SNP rs1071630 in two case-control samples from New Delhi and Kolkata [5]. A significant but weaker effect was seen in a family-based population sample from Southern India for the same SNP [5]. In the Vietnamese leprosy GWAS, SNP rs3187964 was identified as the most significant independent signal in the MHC region [7]. Employing a subset of imputed data, we found that both rs1071630 and rs3187964 were in complete LD with *HLA-DQA1*01* in our sample ($r^2 = 1$ and $r^2 = 0.98$ respectively). *HLA-DQA1*01* was significantly associated with leprosy and presented frequencies of 43.1% and 32.8% in cases and controls, respectively (OR = 1.56, $P = 8.34 \times 10^{-7}$, Table 2). *HLA-DQA1*01* is composed of five four-digit alleles that share specific amino acid residues (*01:01 to *01:05), including the major risk allele *HLA-DQA1*01:05* (Table 2). Conditioning on *HLA-DQA1*01* did not abolish the association of *HLA-DQA1*01:05* ($P_{\text{conditional}} = 1.25 \times 10^{-6}$). Conversely, conditioning on *HLA-DQA1*01:05* accounted for most of the signal of *HLA-DQA1*01* ($P_{\text{conditional}} = 0.0103$). Together, these data showed that rs1071630 and rs3187964 are tag SNPs for the major risk signal in our sample.

Similarly, we had shown SNP rs2394885 to be a major leprosy risk factor in two large samples from Vietnam and India [12]. We had also shown that this SNP tags the *HLA-C*15:05* allele. Here, we replicated this previous observation ($r^2 = 0.63$) and found the same SNP also to be in strong LD with *HLA-B*07:05* ($r^2 = 0.63$) and *HLA-A*29:01* ($r^2 = 0.58$). In the Vietnamese leprosy GWAS, we found that rs2394885 belonged to a large bin of SNPs in strong LD that included the most significant independent class I region SNP rs114598080 of the GWAS [7].

All three tagged class I alleles (*HLA-C*15:05*, *HLA-B*07:05*, *HLA-A*29:01*) are part of signal #1 in our study (Fig 1). Taken together the combined results of the present and previous studies strongly support the role of *HLA-DRB1*10:01*~*HLA-DQA1*01:05* as major leprosy risk factor in independent Indian and Vietnamese samples.

Previous studies in large, well powered samples of Chinese leprosy patients had identified *HLA-DRB1*15:01* as the strongest leprosy associated HLA allele (OR = 2.17, $P = 4.21 \times 10^{-44}$) [3, 4]. These studies had not detected a significant risk effect of the *HLA-DRB1*10:01* and *HLA-DQA1*01:05* alleles. In our sample, we detected a risk effect of *HLA-DRB1*15:01* with similar OR to the one reported for the Chinese patients which failed to pass correction for multiple testing (OR = 2.02, $P = 1.11 \times 10^{-3}$). The reduced significance of *HLA-DRB1*15:01* in the Vietnamese patients is most likely a result of the substantially lower allele frequency in our sample (6.5% and 3.2% in cases and controls) as compared to the Chinese sample (11.7%) [3]. Accounting for the signals tagged by *HLA-DQA1*01:05*, *HLA-C*07:06* and *HLA-DRB1*12* did not abolish the borderline effect for the *HLA-DRB1*15:01* allele ($P_{\text{conditional}} = 1.86 \times 10^{-3}$, S5 Table). Hence, *HLA-DRB1*15:01* was indeed an independent leprosy risk factor for the Vietnamese patients which was also supported by the observation that *HLA-DRB1*15:01* contributed significantly to leprosy risk independently of *HLA-DRB1*10:01*~*HLA-DQA1*01:05*. Previously, two small studies reported a leprosy *per se* risk effect of *HLA-DRB1*15:01* and a serotype linked to *HLA-DQA1*01*, and a well powered study of Brazilian leprosy patients detected a risk effect for both *HLA-DRB1*10* and *HLA-DRB1*15* [36–38]. Taken together, these results implicated *HLA-DRB1*10:01*/*HLA-DQA1*01:05* and *HLA-DRB1*15:01* as strong global leprosy risk factors.

The strongest protective signal was associated with the *HLA-C*07:06*/*HLA-B*44:03* alleles. The two alleles were mostly found on a haplotype with *HLA-DRB1*07:01* ($D' = 0.77$), and collectively made up signal #2 in our study (Fig 1 and S1 Fig). The univariable borderline association of *HLA-DRB1*07:01* with leprosy was fully explained by *HLA-C*07:06*/*HLA-B*44:03* by multivariable analysis. Hence, signal #2 is consistent with previous results in a large case-control sample from Brazil that did not test class I alleles but detected *HLA-DRB1*07* as protective factor [38]. The second resistance signal (signal #3) in our study was associated with allele *HLA-DRB1*12:02*. This finding was consistent with the reported protective effect of *HLA-DRB1*12* in Brazilian and Indonesian leprosy patients [38, 39]. In large studies in Chinese patients, the strongest protective effect was assigned to *HLA-DQB1*04:01* [3]. In both the Vietnamese and the Chinese populations, *HLA-DQB1*04:01* is part of a haplotype which includes *HLA-DQA1*03:03* and *HLA-DRB1*04:05* (S1 Fig), and in a meta-analysis of studies in Chinese patients *HLA-DQA1*03:03* was identified as main leprosy resistance factor [10]. While in our study both *HLA-DQA1*03:03* and *HLA-DRB1*04:05* did not pass the significance threshold required by multiple testing, we did observe a strong protective effect for both alleles independently of the three significant HLA allele signals (OR_{conditional} = 0.50, $P_{\text{conditional}} = 2.65 \times 10^{-3}$ and OR_{conditional} = 0.49, $P_{\text{conditional}} = 3.02 \times 10^{-3}$, respectively) that can be considered as replication of the Chinese studies (S5 Table). Hence, while significance of the protective effect differed among Chinese and Vietnamese patients the effect sizes obtained in Vietnamese patients were consistent with those obtained for Chinese patients.

The leprosy protective signal #2 is dominated by alleles that are part of a *HLA-C*07:06* ~ *HLA-B*44:03* ~ *HLA-DRB1*07:01* haplotype (Fig 1). Of note, *HLA-DRB1*07:01* had previously been implicated in two main inflammatory bowel diseases: Crohn's disease (CD) and ulcerative colitis (UC). Interestingly, *HLA-DRB1*07:01* was protective for UC while it was a risk marker for CD [40]. A meta-analysis of *HLA-DRB1* alleles and risk of tuberculosis (TB) in Asian patients found *HLA-DRB1*07:01* associated with protection from TB [41]. The same meta-analysis also identified *HLA-DRB1*12:02* (protective signal #3 in Fig 1) as TB protective marker with borderline significance [41]. In addition, *HLA-DRB1*12* and **12:02* had also been

associated with protection from periodontal oral infections and recurrent typhoid fever [42, 43]. An allele that needs to be considered is *HLA-DRB1*04:05* which was associated with protection from leprosy in the present as well as other studies [38, 44, 45]. Previously, *HLA-DRB1*04:05* had been found protective for enteric fever in Vietnamese patients [46]. In addition, *HLA-DRB1*04:05* had been reported as associated with protection from UC and risk for CD in the Japanese population [47]. Finally, the leprosy risk allele *HLA-DRB1*15:01* was a major risk factor for multiple sclerosis in earlier studies and has been associated with risk for UC and Parkinson's disease [40, 48–50]. The combined results showed that HLA alleles implicated in leprosy susceptibility have a wider impact on infectious and inflammatory diseases. Importantly, these results also further support the genetic overlap between leprosy, inflammatory bowel disease and Parkinson's disease [1, 16, 51].

Our study demonstrated the challenges posed by the very strong LD across the HLA complex. While independent risk and protective signals could be defined, it was in general not possible by genetic means to implicate a single class I/II allele as cause of the observed associations. To disentangle the complex pattern of correlated alleles, we investigated the association of single polymorphic amino acids within class I and class II proteins. We identified four amino acids in HLA-DR β 1 (phenylalanine and aspartic acid in positions 13 and 57), HLA-B (glutamic acid in position 63), and HLA-A (lysine in position 19) that fully explained the associations of class I and class II alleles with leprosy in our sample (S5 and S6 Tables). Model selection metrics, such as AIC, have been useful to identify the combination of independent HLA variants that represent the best-fitting model for the data in previous studies [52, 53]. Hence, we used AIC to compare the allele-based and amino acid-based models in our dataset. We showed that the model based on only four amino acids better explained the HLA effect on leprosy than the model based on HLA alleles that tag a number of highly correlated alleles (Table 3). Three of the four amino acids are located within the peptide-binding groove of the HLA-DR β 1 and HLA-B proteins (S2 Fig). Amino acid changes in the peptide-binding groove can impact on the HLA molecule interaction with the antigen peptide. In the HLA-DR molecule, amino acid changes at the two HLA-DR β 1 polymorphic positions 57 and 13 are predicted to impact on the peptide-binding repertoire size [54]. Moreover, aspartic acid at position 57 of HLA-DR β 1 interacts with a conserved residue in HLA-DR α and this interaction appears lost in the absence of aspartic acid [55]. Interestingly, this HLA-DR interaction is the same as in HLA-DQ for the well-known Type 1 diabetes-protective HLA-DQ β 1 57D [56–58]. In HLA-B, an *in silico* analysis has indicated that a single substitution in position 63 of the protein can have a strong impact on the protein repertoire that bind the molecule, suggesting this position as crucial for peptide binding [59]. Taken together, these findings suggest that the strong impact on leprosy susceptibility found in our study can be traced to changes in the HLA molecule structure as well as to specific HLA molecule-antigen peptide interactions. Additional studies will be required to fully comprehend the exact range of HLA-*M. leprae* antigen interactions as well as to investigate the non-additive and interactive risk effect of HLA variants in leprosy. Functional studies analyzing the impact of HLA-DR β 1 57D and 13F, HLA-B 63E and HLA-A 19K in the context of the specific protein structure of the associated HLA alleles will provide additional insight in the role of HLA molecules in leprosy susceptibility. Given the ongoing efforts of developing a leprosy vaccine, it will be useful to assure that vaccine antigens can overcome this specific genetic restriction of leprosy susceptibility.

Supporting information

S1 Fig. Haplotype structure of the four-digit alleles for class I HLA-A, -C and -B and class II HLA-DRB1, -DQA1, -DQB1 and -DPB1 in the Vietnamese sample. The two panels show

the HLA haplotype structure in **A**) healthy controls (N = 468) and **B**) leprosy cases (N = 687). The columns represent observed alleles of seven HLA class I and class II genes, where each box corresponds to a specific four-digit HLA allele. The height of the box is relative to the observed HLA allele frequency. Alleles associated with risk or protection from leprosy *per se* are shown in black and green boxes, respectively (see Fig 1). Each grey line connects two HLA alleles in consecutive HLA genes, where the thickness of the line is based on the frequency of the haplotype. Haplotypes between associated alleles in consecutive genes are highlighted in darker grey.

(TIF)

S2 Fig. HLA-DR β 1 positions 13 and 57 and HLA-B position 63, but not HLA-A position 19, are located in the peptide-binding grooves of the respective HLA proteins. Three-dimensional ribbon representation of **A**) HLA-DR, **B**) HLA-B and **C**) HLA-A peptide binding grooves. HLA-DR α , HLA-DR β 1, HLA-B, HLA-A and microglobulin are shown in green, blue, pink, purple and grey, respectively. HLA-DR β 1 positions 13 and 57 (**A**), HLA-B position 63 (**B**) and HLA-A position 19 (**C**) are shown as spheres and highlighted in red.

(TIF)

S1 Table. Association analysis of leprosy *per se*, leprosy subtype polarization or type-1 reaction phenotypes against HLA alleles binary markers from 11 classic HLA genes.

(XLSX)

S2 Table. Forward conditional analysis of HLA alleles that reached study-wide significant association with leprosy *per se*.

(XLSX)

S3 Table. Linkage disequilibrium analysis in healthy controls and pairwise reciprocal conditional analysis of leprosy-associated HLA alleles.

(XLSX)

S4 Table. Association analysis of leprosy *per se* and HLA amino acids in HLA-A, HLA-C, HLA-B, HLA-DR β 1 and HLA-DQ α 1.

(XLSX)

S5 Table. HLA alleles with conditional *P*-value lower than 0.01 after adjusting on HLA-DQAI*01:05, HLA-C*07:06 and HLA-DRB1*12.

(XLSX)

S6 Table. Impact on HLA alleles association with leprosy *per se* by conditioning the logistic regression on HLA-DR β 1 57D, HLA-DR β 1 13F and HLA-B 63E in a forward fashion.

(XLSX)

S1 Data. Genotypes of the 622 tested HLA alleles and amino acid binary markers in 468 controls and 687 leprosy cases (coded as 1 and 2, respectively). Genotypes are provided in PLINK binary format (bed/bim/fam files) from PLINK v1.9 (<http://www.cog-genomics.org/plink/1.9/>).

(GZ)

Acknowledgments

We are grateful to all patients and controls who participated in the study. We thank the members of the leprosy control program at the Hospital for Dermato-Venereology in Ho Chi Minh City and the staff of the leprosy control program in Southern Việt Nam. We thank all members

of the Schurr laboratory in Montreal and the members of the Human Genetics of Infectious Diseases laboratory in Paris for helpful discussions and suggestions.

Author Contributions

Conceptualization: Monica Dallmann-Sauer, Vinicius M. Fava, Alexandre Alcaïs, Laurent Abel, Aurélie Cobat, Erwin Schurr.

Data curation: Monica Dallmann-Sauer, Marianna Orlova.

Formal analysis: Monica Dallmann-Sauer, Chaïma Gzara.

Funding acquisition: Alexandre Alcaïs, Erwin Schurr.

Investigation: Monica Dallmann-Sauer, Vinicius M. Fava.

Project administration: Erwin Schurr.

Resources: Chaïma Gzara, Marianna Orlova, Nguyen Van Thuc, Vu Hong Thai, Alexandre Alcaïs, Aurélie Cobat, Erwin Schurr.

Supervision: Erwin Schurr.

Visualization: Monica Dallmann-Sauer.

Writing – original draft: Monica Dallmann-Sauer, Erwin Schurr.

Writing – review & editing: Monica Dallmann-Sauer, Vinicius M. Fava, Laurent Abel, Aurélie Cobat, Erwin Schurr.

References

1. Fava VM, Dallmann-Sauer M, Schurr E. Genetics of leprosy: today and beyond. *Hum Genet.* 2019. Epub 2019/11/13. <https://doi.org/10.1007/s00439-019-02087-5> PMID: 31713021.
2. Zhang FR, Huang W, Chen SM, Sun LD, Liu H, Li Y, et al. Genomewide association study of leprosy. *N Engl J Med.* 2009; 361(27):2609–18. Epub 2009/12/19. <https://doi.org/10.1056/NEJMoa0903753> PMID: 20018961.
3. Liu H, Irwanto A, Fu X, Yu G, Yu Y, Sun Y, et al. Discovery of six new susceptibility loci and analysis of pleiotropic effects in leprosy. *Nat Genet.* 2015; 47(3):267–71. Epub 2015/02/03. <https://doi.org/10.1038/ng.3212> PMID: 25642632.
4. Wang Z, Sun Y, Fu X, Yu G, Wang C, Bao F, et al. A large-scale genome-wide association and meta-analysis identified four novel susceptibility loci for leprosy. *Nat Commun.* 2016; 7:13760. Epub 2016/12/16. <https://doi.org/10.1038/ncomms13760> PMID: 27976721; PubMed Central PMCID: PMC5172377.
5. Wong SH, Gochhait S, Malhotra D, Pettersson FH, Teo YY, Khor CC, et al. Leprosy and the adaptation of human toll-like receptor 1. *PLoS Pathog.* 2010; 6:e1000979. Epub 2010/07/10. <https://doi.org/10.1371/journal.ppat.1000979> PMID: 20617178; PubMed Central PMCID: PMC2895660.
6. Liu H, Wang Z, Li Y, Yu G, Fu X, Wang C, et al. Genome-Wide Analysis of Protein-Coding Variants in Leprosy. *J Invest Dermatol.* 2017; 137(12):2544–51. Epub 2017/08/27. <https://doi.org/10.1016/j.jid.2017.08.004> PMID: 28842327.
7. Gzara C, Dallmann-Sauer M, Orlova M, Van Thuc N, Thai VH, Fava VM, et al. Family-based genome-wide association study of leprosy in Vietnam. *PLoS Pathog.* 2020; 16(5):e1008565. Epub 2020/05/19. <https://doi.org/10.1371/journal.ppat.1008565> PMID: 32421744; PubMed Central PMCID: PMC7259797.
8. Trowsdale J, Knight JC. Major histocompatibility complex genomics and human disease. *Annu Rev Genomics Hum Genet.* 2013; 14:301–23. Epub 2013/07/24. <https://doi.org/10.1146/annurev-genom-091212-153455> PMID: 23875801; PubMed Central PMCID: PMC4426292.
9. Jarduli LR, Sell AM, Reis PG, Sippert EA, Ayo CM, Mazini PS, et al. Role of HLA, KIR, MICA, and cytokines genes in leprosy. *Biomed Res Int.* 2013; 2013:989837. Epub 2013/08/13. <https://doi.org/10.1155/2013/989837> PMID: 23936864; PubMed Central PMCID: PMC3722889.
10. Zhang X, Cheng Y, Zhang Q, Wang X, Lin Y, Yang C, et al. Meta-Analysis Identifies Major Histocompatibility Complex Loci in or Near HLA-DRB1, HLA-DQA1, HLA-C as Associated with Leprosy in Chinese

- Han Population. *J Invest Dermatol*. 2019; 139(4):957–60. Epub 2018/11/06. <https://doi.org/10.1016/j.jid.2018.09.029> PMID: 30389493.
11. Alcais A, Alter A, Antoni G, Orlova M, Nguyen VT, Singh M, et al. Stepwise replication identifies a low-producing lymphotoxin-alpha allele as a major risk factor for early-onset leprosy. *Nat Genet*. 2007; 39(4):517–22. Epub 2007/03/14. <https://doi.org/10.1038/ng2000> PMID: 17353895.
 12. Alter A, Huong NT, Singh M, Orlova M, Van Thuc N, Katoch K, et al. Human leukocyte antigen class I region single-nucleotide polymorphisms are associated with leprosy susceptibility in Vietnam and India. *J Infect Dis*. 2011; 203(9):1274–81. Epub 2011/04/05. <https://doi.org/10.1093/infdis/jir024> PMID: 21459816; PubMed Central PMCID: PMC3069725.
 13. Fava VM, Manry J, Cobat A, Orlova M, Van Thuc N, Moraes MO, et al. A genome wide association study identifies a lncRNA as risk factor for pathological inflammatory responses in leprosy. *PLoS Genet*. 2017; 13(2):e1006637. Epub 2017/02/22. <https://doi.org/10.1371/journal.pgen.1006637> PMID: 28222097; PubMed Central PMCID: PMC5340414.
 14. Grant AV, Alter A, Huong NT, Orlova M, Van Thuc N, Ba NN, et al. Crohn's disease susceptibility genes are associated with leprosy in the Vietnamese population. *J Infect Dis*. 2012; 206(11):1763–7. Epub 2012/09/18. <https://doi.org/10.1093/infdis/jis588> PMID: 22984114.
 15. Gaschignard J, Grant AV, Thuc NV, Orlova M, Cobat A, Huong NT, et al. Pauci- and Multibacillary Leprosy: Two Distinct, Genetically Neglected Diseases. *PLoS Negl Trop Dis*. 2016; 10(5):e0004345. Epub 2016/05/25. <https://doi.org/10.1371/journal.pntd.0004345> PMID: 27219008; PubMed Central PMCID: PMC4878860.
 16. Fava VM, Xu YZ, Lettre G, Van Thuc N, Orlova M, Thai VH, et al. Pleiotropic effects for Parkin and LRRK2 in leprosy type-1 reactions and Parkinson's disease. *Proc Natl Acad Sci U S A*. 2019; 116(31):15616–24. Epub 2019/07/17. <https://doi.org/10.1073/pnas.1901805116> PMID: 31308240; PubMed Central PMCID: PMC6681704.
 17. Duke JL, Lind C, Mackiewicz K, Ferriola D, Papazoglou A, Gasiewski A, et al. Determining performance characteristics of an NGS-based HLA typing method for clinical applications. *HLA*. 2016; 87(3):141–52. Epub 2016/02/18. <https://doi.org/10.1111/tan.12736> PMID: 26880737.
 18. Omixon Biocomputing Ltd. Omixon HLA Twin RUO—2.1.4: Handbook. Budapest, Hungary.
 19. Robinson J, Barker DJ, Georgiou X, Cooper MA, Flicek P, Marsh SGE. IPD-IMGT/HLA Database. *Nucleic Acids Res*. 2020; 48(D1):D948–D55. Epub 2019/11/02. <https://doi.org/10.1093/nar/gkz950> PMID: 31667505; PubMed Central PMCID: PMC7145640.
 20. Omixon Biocomputing Ltd. Omixon HLA Twin RUO—2.5.0: User guide. Budapest, Hungary.
 21. Jia X, Han B, Onengut-Gumuscu S, Chen WM, Concannon PJ, Rich SS, et al. Imputing amino acid polymorphisms in human leukocyte antigens. *PLoS One*. 2013; 8(6):e64683. Epub 2013/06/14. <https://doi.org/10.1371/journal.pone.0064683> PMID: 23762245; PubMed Central PMCID: PMC3675122.
 22. R Core Team. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. 2020. Available from: <https://www.R-project.org/>.
 23. Li MX, Yeung JM, Cherny SS, Sham PC. Evaluating the effective numbers of independent tests and significant p-value thresholds in commercial genotyping arrays and public imputation reference datasets. *Hum Genet*. 2012; 131(5):747–56. Epub 2011/12/07. <https://doi.org/10.1007/s00439-011-1118-2> PMID: 22143225; PubMed Central PMCID: PMC3325408.
 24. Chang CC, Chow CC, Tellier LC, Vattikuti S, Purcell SM, Lee JJ. Second-generation PLINK: rising to the challenge of larger and richer datasets. *Gigascience*. 2015; 4:7. Epub 2015/02/28. <https://doi.org/10.1186/s13742-015-0047-8> PMID: 25722852; PubMed Central PMCID: PMC4342193.
 25. Akaike H. A new look at the statistical model identification. *IEEE Transactions on Automatic Control*. 1974; 19(6):716–23. <https://doi.org/10.1109/tac.1974.1100705>
 26. Burnham K, Anderson D. Model selection and multimodel inference: a practical information-theoretic approach. 2nd ed. New York: Springer-Verlag; 2002.
 27. Browning BL, Browning SR. A unified approach to genotype imputation and haplotype-phase inference for large data sets of trios and unrelated individuals. *Am J Hum Genet*. 2009; 84(2):210–23. Epub 2009/02/10. <https://doi.org/10.1016/j.ajhg.2009.01.005> PMID: 19200528; PubMed Central PMCID: PMC2668004.
 28. Aragon TJ. epitools: Epidemiology Tools. R package version 0.5–10. 2017. Available from: <https://CRAN.R-project.org/package=epitools>.
 29. Kumasaka N, Nakamura Y, Kamatani N. The textile plot: a new linkage disequilibrium display of multiple-single nucleotide polymorphism genotype data. *PLoS One*. 2010; 5(4):e10207. Epub 2010/05/04. <https://doi.org/10.1371/journal.pone.0010207> PMID: 20436909; PubMed Central PMCID: PMC2860502.

30. Okada Y, Momozawa Y, Ashikawa K, Kanai M, Matsuda K, Kamatani Y, et al. Construction of a population-specific HLA imputation reference panel and its application to Graves' disease risk in Japanese. *Nat Genet.* 2015; 47(7):798–802. Epub 2015/06/02. <https://doi.org/10.1038/ng.3310> PMID: 26029868.
31. Pettersen EF, Goddard TD, Huang CC, Couch GS, Greenblatt DM, Meng EC, et al. UCSF Chimera—a visualization system for exploratory research and analysis. *J Comput Chem.* 2004; 25(13):1605–12. Epub 2004/07/21. <https://doi.org/10.1002/jcc.20084> PMID: 15264254.
32. Gunther S, Schlundt A, Sticht J, Roske Y, Heinemann U, Wiesmuller KH, et al. Bidirectional binding of invariant chain peptides to an MHC class II molecule. *Proc Natl Acad Sci U S A.* 2010; 107(51):22219–24. Epub 2010/12/01. <https://doi.org/10.1073/pnas.1014708107> PMID: 21115828; PubMed Central PMCID: PMC3009805.
33. Zernich D, Purcell AW, Macdonald WA, Kjer-Nielsen L, Ely LK, Laham N, et al. Natural HLA class I polymorphism controls the pathway of antigen presentation and susceptibility to viral evasion. *J Exp Med.* 2004; 200(1):13–24. Epub 2004/07/01. <https://doi.org/10.1084/jem.20031680> PMID: 15226359; PubMed Central PMCID: PMC2213310.
34. Bulek AM, Cole DK, Skowera A, Dolton G, Gras S, Madura F, et al. Structural basis for the killing of human beta cells by CD8(+) T cells in type 1 diabetes. *Nat Immunol.* 2012; 13(3):283–9. Epub 2012/01/17. <https://doi.org/10.1038/ni.2206> PMID: 22245737; PubMed Central PMCID: PMC3378510.
35. Tosh K, Ravikumar M, Bell JT, Meisner S, Hill AV, Pitchappan R. Variation in MICA and MICB genes and enhanced susceptibility to paucibacillary leprosy in South India. *Hum Mol Genet.* 2006; 15(19):2880–7. Epub 2006/08/23. <https://doi.org/10.1093/hmg/ddl229> PMID: 16923796.
36. Rani R, Fernandez-Vina MA, Zaheer SA, Beena KR, Stastny P. Study of HLA class II alleles by PCR oligotyping in leprosy patients from north India. *Tissue Antigens.* 1993; 42(3):133–7. Epub 1993/09/01. <https://doi.org/10.1111/j.1399-0039.1993.tb02179.x> PMID: 8284786.
37. Schauf V, Ryan S, Scollard D, Jonasson O, Brown A, Nelson K, et al. Leprosy associated with HLA-DR2 and DQw1 in the population of northern Thailand. *Tissue Antigens.* 1985; 26(4):243–7. Epub 1985/10/01. <https://doi.org/10.1111/j.1399-0039.1985.tb00966.x> PMID: 3878012.
38. Vanderborgh PR, Pacheco AG, Moraes ME, Antoni G, Romero M, Verville A, et al. HLA-DRB1*04 and DRB1*10 are associated with resistance and susceptibility, respectively, in Brazilian and Vietnamese leprosy patients. *Genes Immun.* 2007; 8(4):320–4. Epub 2007/03/31. <https://doi.org/10.1038/sj.gene.6364390> PMID: 17396103.
39. Soebono H, Giphart MJ, Schreuder GM, Klatser PR, de Vries RR. Associations between HLA-DRB1 alleles and leprosy in an Indonesian population. *Int J Lepr Other Mycobact Dis.* 1997; 65(2):190–6. Epub 1997/06/01. PMID: 9251590.
40. Goyette P, Boucher G, Mallon D, Ellinghaus E, Jostins L, Huang H, et al. High-density mapping of the MHC identifies a shared role for HLA-DRB1*01:03 in inflammatory bowel diseases and heterozygous advantage in ulcerative colitis. *Nat Genet.* 2015; 47(2):172–9. Epub 2015/01/07. <https://doi.org/10.1038/ng.3176> PMID: 25559196; PubMed Central PMCID: PMC4310771.
41. Li CP, Zhou Y, Xiang X, Zhou Y, He M. Relationship of HLA-DRB1 gene polymorphism with susceptibility to pulmonary tuberculosis: updated meta-analysis. *Int J Tuberc Lung Dis.* 2015; 19(7):841–9. Epub 2015/06/10. <https://doi.org/10.5588/ijtld.14.0521> PMID: 26056112.
42. Mauramo M, Ramseier AM, Buser A, Tiercy JM, Weiger R, Waltimo T. Associations of HLA-A, -B and -DRB1 types with oral diseases in Swiss adults. *PLoS One.* 2014; 9(7):e103527. Epub 2014/07/30. <https://doi.org/10.1371/journal.pone.0103527> PMID: 25072155; PubMed Central PMCID: PMC4114782.
43. Dharmana E, Joosten I, Tijssen HJ, Gasem MH, Indarwidayati R, Keuter M, et al. HLA-DRB1*12 is associated with protection against complicated typhoid fever, independent of tumour necrosis factor alpha. *Eur J Immunogenet.* 2002; 29(4):297–300. Epub 2002/07/18. <https://doi.org/10.1046/j.1365-2370.2002.00318.x> PMID: 12121274.
44. Joko S, Numaga J, Kawashima H, Namisato M, Maeda H. Human leukocyte antigens in forms of leprosy among Japanese patients. *Int J Lepr Other Mycobact Dis.* 2000; 68(1):49–56. Epub 2000/06/02. PMID: 10834069.
45. Hsieh NK, Chu CC, Lee NS, Lee HL, Lin M. Association of HLA-DRB1*0405 with resistance to multibacillary leprosy in Taiwanese. *Hum Immunol.* 2010; 71(7):712–6. Epub 2010/04/01. <https://doi.org/10.1016/j.humimm.2010.03.007> PMID: 20353806.
46. Dunstan SJ, Hue NT, Han B, Li Z, Tram TT, Sim KS, et al. Variation at HLA-DRB1 is associated with resistance to enteric fever. *Nat Genet.* 2014; 46(12):1333–6. Epub 2014/11/11. <https://doi.org/10.1038/ng.3143> PMID: 25383971; PubMed Central PMCID: PMC5099079.
47. Arimura Y, Isshiki H, Onodera K, Nagaishi K, Yamashita K, Sonoda T, et al. Characteristics of Japanese inflammatory bowel disease susceptibility loci. *Journal of Gastroenterology.* 2013; 49(8):1217–30. Epub 2013/08/15. <https://doi.org/10.1007/s00535-013-0866-2> PMID: 23942620.

48. Patsopoulos NA, Barcellos LF, Hintzen RQ, Schaefer C, van Duijn CM, Noble JA, et al. Fine-mapping the genetic association of the major histocompatibility complex in multiple sclerosis: HLA and non-HLA effects. *PLoS Genet.* 2013; 9(11):e1003926. Epub 2013/11/28. <https://doi.org/10.1371/journal.pgen.1003926> PMID: 24278027; PubMed Central PMCID: PMC3836799.
49. Wissemann WT, Hill-Burns EM, Zabetian CP, Factor SA, Patsopoulos N, Hoglund B, et al. Association of Parkinson disease with structural and regulatory variants in the HLA region. *Am J Hum Genet.* 2013; 93(5):984–93. Epub 2013/11/05. <https://doi.org/10.1016/j.ajhg.2013.10.009> PMID: 24183452; PubMed Central PMCID: PMC3824116.
50. Sulzer D, Alcalay RN, Garretti F, Cote L, Kanter E, Agin-Lieb J, et al. T cells from patients with Parkinson's disease recognize alpha-synuclein peptides. *Nature.* 2017; 546(7660):656–61. Epub 2017/06/22. <https://doi.org/10.1038/nature22815> PMID: 28636593; PubMed Central PMCID: PMC5626019.
51. Hui KY, Fernandez-Hernandez H, Hu J, Schaffner A, Pankratz N, Hsu NY, et al. Functional variants in the LRRK2 gene confer shared effects on risk for Crohn's disease and Parkinson's disease. *Sci Transl Med.* 2018; 10(423):eaai7795. Epub 2018/01/13. <https://doi.org/10.1126/scitranslmed.aai7795> PMID: 29321258; PubMed Central PMCID: PMC6028002.
52. Morris DL, Taylor KE, Fernando MM, Nititham J, Alarcon-Riquelme ME, Barcellos LF, et al. Unraveling multiple MHC gene associations with systemic lupus erythematosus: model choice indicates a role for HLA alleles and non-HLA genes in Europeans. *Am J Hum Genet.* 2012; 91(5):778–93. Epub 2012/10/23. <https://doi.org/10.1016/j.ajhg.2012.08.026> PMID: 23084292; PubMed Central PMCID: PMC3487133.
53. Ferreiro-Iglesias A, Lesseur C, McKay J, Hung RJ, Han Y, Zong X, et al. Fine mapping of MHC region in lung cancer highlights independent susceptibility loci by ethnicity. *Nat Commun.* 2018; 9(1):3927. Epub 2018/09/27. <https://doi.org/10.1038/s41467-018-05890-2> PMID: 30254314; PubMed Central PMCID: PMC6156406.
54. Manczinger M, Boross G, Kemeny L, Muller V, Lenz TL, Papp B, et al. Pathogen diversity drives the evolution of generalist MHC-II alleles in human populations. *PLoS Biol.* 2019; 17(1):e3000131. Epub 2019/02/01. <https://doi.org/10.1371/journal.pbio.3000131> PMID: 30703088; PubMed Central PMCID: PMC6372212.
55. Brown JH, Jardetzky TS, Gorga JC, Stern LJ, Urban RG, Strominger JL, et al. Three-dimensional structure of the human class II histocompatibility antigen HLA-DR1. *Nature.* 1993; 364(6432):33–9. Epub 1993/07/01. <https://doi.org/10.1038/364033a0> PMID: 8316295.
56. Todd JA, Bell JL, McDevitt HO. HLA-DQ beta gene contributes to susceptibility and resistance to insulin-dependent diabetes mellitus. *Nature.* 1987; 329(6140):599–604. Epub 1987/10/15. <https://doi.org/10.1038/329599a0> PMID: 3309680.
57. Hu X, Deutsch AJ, Lenz TL, Onengut-Gumuscu S, Han B, Chen WM, et al. Additive and interaction effects at three amino acid positions in HLA-DQ and HLA-DR molecules drive type 1 diabetes risk. *Nat Genet.* 2015; 47(8):898–905. Epub 2015/07/15. <https://doi.org/10.1038/ng.3353> PMID: 26168013; PubMed Central PMCID: PMC4930791.
58. Gerasimou P, Nicolaidou V, Skordis N, Picolos M, Monos D, Costeas PA. Combined effect of glutamine at position 70 of HLA-DRB1 and alanine at position 57 of HLA-DQB1 in type 1 diabetes: An epitope analysis. *PLoS One.* 2018; 13(3):e0193684. Epub 2018/03/02. <https://doi.org/10.1371/journal.pone.0193684> PMID: 29494662; PubMed Central PMCID: PMC5832312.
59. van Deutekom HW, Kesmir C. Zooming into the binding groove of HLA molecules: which positions and which substitutions change peptide binding most? *Immunogenetics.* 2015; 67(8):425–36. Epub 2015/06/05. <https://doi.org/10.1007/s00251-015-0849-y> PMID: 26040913; PubMed Central PMCID: PMC4498290.

Résumé

La lèpre, maladie infectieuse chronique causée par *Mycobacterium leprae*, affecte principalement la peau, les nerfs et les yeux avec des séquelles majeures en l'absence de traitement. Avec 200 000 nouveaux cas diagnostiqués chaque année (1 toutes les 2 minutes), il s'agit de la mycobactériose la plus commune après la tuberculose et requilifiée « maladie tropicale négligée » en 2017 par l'Organisation Mondiale de la Santé (OMS). Si la contribution génétique de l'hôte dans l'histoire naturelle de la maladie est maintenant bien établie, son architecture reste lacunaire. Dans cette continuité et afin de la préciser, nous avons, pour la première fois, utilisé une approche de génétique épidémiologique familiale.

Plus précisément, nous avons réalisé la première étude d'association pangénomique (Genome-Wide Association Study, GWAS) familiale sur la lèpre. Ainsi, au cours des 20 dernières années, un échantillon de 481 familles nucléaires (parents et enfants), sélectionnées à partir d'un enfant atteint, a été constitué au Vietnam. Sur cet échantillon primaire de 1 749 individus incluant 622 enfants atteints, nous avons testé l'association de près de 6 millions de variants bi-alléliques (Single Nucleotide Polymorphism, SNP), génotypés ou imputés, avec la lèpre. Dans un second temps, nous avons testé les signaux les plus prometteurs dans un échantillon de réplication, c'est-à-dire, indépendant et issu de la même population, constitué de 1 181 cas et 668 contrôles.

Les résultats les plus significatifs ont été observés au sein de la région HLA et l'analyse multivariée a permis d'identifier trois signaux indépendants. Deux dans la région HLA classe I : rs1265048 [OR = 0,69 ; p-val = $5,5 \cdot 10^{-11}$] et rs114598080 [OR = 1,47 ; p-val = $8,8 \cdot 10^{-13}$] ; Et un dans la région HLA classe II : rs3187964 [OR = 1,67 ; p-val = $8,4 \cdot 10^{-16}$]. Nous avons également identifié deux signaux hors HLA : un variant faux-sens dans le gène *LACC1* (rs3764147 : OR = 1,52 ; p-val = $5,1 \cdot 10^{-14}$), et un variant à proximité du gène *IL12B* (rs6871626 : OR = 0,73 ; p-val = $6,4 \cdot 10^{-8}$).

Les contraintes de coûts des études pangénomiques imposent une réduction majeure du nombre de SNPs à tester dans d'autres échantillons. Dans les études familiales, les parents sont de fait génotypés et pourraient permettre une réplication immédiate sans coûts ajoutés. Au moyen d'une large étude de simulation, nous avons montré que cette approche était pertinente. Une étude cas-contrôle chez les parents de l'échantillon primaire est une réplication valide, statistiquement indépendante de l'étude d'association familiale. C'est un argument fort en faveur des approches familiales pour l'exploration pangénomique de la contribution génétique de l'hôte dans les phénotypes complexes.

La compréhension de la physiopathologie de l'infection à *M. leprae* est cruciale pour optimiser les approches préventives selon les profils génétiques à plus haut risque, et ouvrir de nouvelles pistes thérapeutiques en précisant les cascades fonctionnelles pertinentes. En ce sens, la dissection du contrôle génétique de l'infection par l'hôte est indispensable. Enfin, remettre la famille au cœur de la quête génétique, c'est remettre la génétique dans son milieu naturel.

Mots-clés :

Maladie infectieuse, Épidémiologie, Biostatistiques, Génétique humaine, Mycobactériose, Lèpre, *Mycobacterium leprae*, Vietnam, Familles, Analyse d'association pangénomique (GWAS), Approche familiale (Family-Based), Réplication (Population-Based), Études cas-contrôles parentales.