

UNIVERSITE PARIS V – DESCARTES

Année 2015

THESE

Pour obtenir le grade de

DOCTEUR DE L'UNIVERSITE PARIS V – DESCARTES

Spécialité : Génétique Statistique

Ecole Doctorale de rattachement : ED420

Présentée et soutenue publiquement

par

Jean GASCHIGNARD

le 27 mars 2015

Génétique humaine des formes cliniques de la lèpre

Dirigée par M. Alexandre Alcaïs et M. Laurent Abel

JURY

Mme Florence Demenais	Présidente
Mme Geneviève Milon	Rapporteur
M. Bertram Müller-Myhsok	Rapporteur
M. Alain Dessein	Examineur
M. Marcelo Mira	Examineur
M. Laurent Abel	Directeur
M. Alexandre Alcaïs	Directeur

Titre : Génétique humaine des formes cliniques de la lèpre

Résumé : La lèpre est une maladie tropicale négligée qui atteint près de 200 000 personnes chaque année et dont l'agent causal est *Mycobacterium leprae*. La susceptibilité génétique de l'hôte à la maladie est bien établie, et a permis de comprendre certains mécanismes de la physiopathologie de la maladie. Il existe par ailleurs une grande variabilité inter-individuelle des manifestations cliniques de la maladie, qui s'étendent d'un pôle dit tuberculoïde à un pôle dit lépromateux. Nous avons cherché à identifier les facteurs de susceptibilité génétique à cette polarisation de la maladie. Nous avons tout d'abord décrit que le sexe et l'âge sont des facteurs non-génétiques associés à ce phénotype. Notre travail s'est ensuite appuyé sur les outils classiques de l'épidémiologie génétique, c'est à dire les études de liaison et d'association, pour identifier des variants génétiques qui influencent la polarisation de la lèpre. Nous avons utilisé une puce à ADN pangénomique avec plus de 500 000 marqueurs pour génotyper un échantillon de familles vietnamiennes comprenant 939 malades, dont 692 enfants. Nous avons identifié une liaison de la région 19p12 avec la polarisation de la lèpre. L'étude d'association n'a pas permis d'identifier de signal significatif à l'échelle du génome. Nous avons développé un nouveau test d'association pour des données familiales qui a permis d'améliorer les résultats sans atteindre la significativité. Notre travail sera prolongé par des études d'association dans deux populations cas-témoins du Vietnam et du Brésil. Nous chercherons à identifier les marqueurs causaux au sein de la région de liaison 19p12 d'une part, et à découvrir de nouveaux variants d'autre part. L'identification de marqueurs associés à la polarisation de la lèpre permettra de mieux prévoir l'évolution de la maladie et de proposer des traitements plus ciblés selon le risque génétique individuel.

Mots clés : Epidémiologie génétique, génétique humaine, lèpre, *Mycobacterium leprae*, tuberculoïde, paucibacillaire, lépromateux, multibacillaire, polarisation, tatou, transmission familiale, analyse de liaison génétique, analyse d'association, association familiale, trait catégoriel.

Title: Human genetics of clinical forms of leprosy

Abstract: Leprosy is a neglected tropical disease that affects nearly 200,000 people each year and caused by *Mycobacterium leprae*. Genetic host susceptibility to the disease is well established, and helped to understand some of the mechanisms of the disease's physiopathology. There is also a wide inter-individual variability of the clinical manifestations of the disease, which runs from a so-said tuberculoid to a so-said lepromatous pole. We sought to identify genetic susceptibility factors for this polarization of the disease. We initially described gender and age as non-genetic factors associated with this phenotype. We then based our analysis on the classical tools from the field of genetic epidemiology, namely linkage and association studies, to identify genetic variants that influence leprosy polarization. We used a DNA-microarray with 500,000 markers spanning over the whole genome to genotype a sample of Vietnamese families including 939 patients, of which 692 were children. We have found a region linked to leprosy polarization on chromosome 19p12. The association study could not identify a significant signal across the genome. We developed a new test for association designed for familial data that improved the results without reaching significance. Our work will be pursued by association studies in two case-control populations from Vietnam and Brazil. We will try to identify the causal markers within the 19p12 linkage region on the one hand, and to discover new variants on the other hand. The identification of markers associated with leprosy polarization could help us to better predict the evolution of the disease, and to offer more targeted treatments to patients, based on their individual genetic risk.

Keywords: Genetic epidemiology, human genetics, leprosy, *Mycobacterium leprae*, tuberculoid paucibacillary, lepromatous, multibacillary, polarization, armadillo, familial transmission, linkage analysis, association analysis, familial association, categorical trait.

Remerciements

Je remercie Florence Demenais qui me fait l'honneur de présider le jury. Ma thèse est un clin d'œil au travail qu'elle a mené sur la génétique humaine de la lèpre sur l'île de la Désirade au début de sa carrière. Je remercie aussi vivement Geneviève Milon et Bertram Müller-Myhsok d'avoir accepté la charge de rapporteurs. Leur lecture attentive et curieuse ainsi que leurs encouragements m'ont porté dans cette dernière ligne droite. Ma reconnaissance va également à Alain Dessein pour avoir accepté avec enthousiasme le rôle d'examineur malgré des délais raccourcis.

Obrigado à Marcelo Mira, pour son accueil chaleureux à Belem et pour la rencontre qu'il m'a offerte avec les habitants de la communauté de Prata, au Brésil.

Mes pensées vont aux personnes rencontrées au Vietnam, en particulier au Docteur Khoa, chirurgien dont j'ai eu le privilège d'être l'aide-opérateur pendant une semaine auprès des patients atteints de lèpre. Et aussi à sœur Modeste, qui soigne humblement les malades des communautés rurales de la vallée de Kon Tum. Je garde en mémoire leur dévouement.

Je n'oublie pas Erwin Schurr, qui m'a soutenu dans bien des moments, de la rédaction d'un article à la reformulation d'un concept, en passant par l'accueil d'un jeune couple sous la neige de Montréal. Il a été, avec Marianna Orlova, un guide hors pair pour déguster un *phở* dans les ruelles d'Hô Chi Minh.

Merci à Laurent pour sa bienveillance et sa constance à chaque étape du travail. Je me sens tout petit devant son exigence intellectuelle et sa probité qui ont été des phares lors des moments plus rudes de la traversée de ma thèse.

Merci à Alexandre, pour sa confiance en moi. Il regarde toujours son étudiant comme une personne et non simplement comme un anonyme qui fait avancer un projet. Il m'a permis de rencontrer des personnes extraordinaires. Ma culture culinaire lui doit également beaucoup. Il n'y a pas de raison que nous ne nous recroisions pas à l'*Epicuriste*...

Merci à Jean-Laurent de m'avoir accueilli dans le laboratoire et de m'avoir posé de bonnes questions avant que je ne me lance dans cette aventure. Elles trottent encore dans ma tête, et j'y apporte des réponses, à mon rythme.

Je remercie la Fondation pour la Recherche Médicale, qui m'a donné et renouvelé sa confiance pour financer les trois années de ma thèse. Cela m'a permis de me consacrer entièrement à ce travail. Très concrètement, savoir que des personnes avaient fait des dons pour financer mon travail m'a aussi aidé à me lever certains matins.

Je remercie chaque membre de l'équipe, pour leur partage enthousiaste d'idées scientifiques et... moins scientifiques. Une pensée particulière pour les épidémiologistes : Aurélie pour sa disponibilité, son aide toujours délicate et ses bons conseils pour s'expatrier avec une femme kangourou au Québec, Fabienne pour sa générosité à partager ses gâteaux quand je n'avais plus de Prince, Frédégonde pour sa joie de vivre et son précieux rôle de secrétaire à distance, Audrey, pour les grandes discussions même en-dehors des pauses café, Vincent pour sa patience à me coacher pour les couleurs de mes présentations, Aziz, pour son soutien viril

indéfectible, et Quentin qui a guidé mes premiers pas dans l'univers de l'épidémiologie génétique.

Je remercie mes parents, mes frères et sœurs, mes beaux-frères et belles sœurs, dont je mesure chaque jour le soutien précieux, dans les joies et dans les épreuves.

Et mon épouse, Béatrice, qui a partagé la dernière ligne de cette aventure exigeante, avec douceur et patience. J'apprends tellement à ses côtés.

Liste des publications

Publications issues des résultats de la thèse

J. Gaschignard, A.V. Grant, E. Schurr and A. Alcaïs. Génétique épidémiologique de la lèpre : mythe ou réalité ?, *Bulletin de l'association des Léprologistes de Langue Française*, vol.2, 2014.

J. Gaschignard, E. Schurr, and A. Alcaïs. La lèpre, pilier de la génétique des maladies infectieuses, *Pathol Biol*, vol n°6, p 120-8, 2013.

J. Gaschignard, Q.B. Vincent, A. Cobat, J.P. Jaïs, and A. Alcaïs. Implicit hypotheses are hidden power droppers in family-based association studies of secondary outcomes, *Open Journal of Statistics*, vol n°5, p 35-45, 2015 (*Article 3 en annexe*).

J. Gaschignard, A.V. Grant, N.V. Thuc, M. Orlova, N.T. Huong, N.N. Ba, V.H. Thai, L. Abel, E. Schurr and A. Alcaïs. Pauci- and multibacillary leprosy: two distinct genetically neglected diseases, *soumis à Plos Neglected Tropical Diseases* (*Article 1 en annexe*).

J. Gaschignard, A. Cobat, N.V. Thuc, M. Orlova, A. Boland-Auge, N.T. Huong, N.N. Ba, V.H. Thai, E. Schurr and A. Alcaïs, Genomewide linkage analysis of leprosy polarization implicates chromosomal region 19p12-p13, *en cours de soumission à Journal of Infectious Diseases* (*Article 2 en annexe*).

Autres publications au cours de la thèse

AV. Grant, A. Cobat, N Van Thuc, M. Orlova, NT Huong, **J. Gaschignard**, A. Alter, NN Ba, VH Thai, L. Abel, A. Alcaïs and E. Schurr. CUBN and NEBL common variants in the chromosome 10p13 linkage region are associated with multibacillary leprosy in Vietnam, *Hum Genet*, vol. 133, n°7, p. 883-93, 2014.

J. Gaschignard, C. Levy, M. Chrabieh, B. Boisson, C. Bost-Bru, S. Dager, F. Dubos, P. Durand, J. Gaudelus, D. Gendrel, C. Gras Le Guen, E. Grimpel, G. Guyon, C. Jeudy, E. Jeziorski, F. Leclerc, P.L. Léger, F. Lesage, M. Lorrot, I. Pellier, D. Pinquier, L. de Pontual, P. Sachs, C. Thomas, P. Tissières, F.V. Valla, P. Desprez, V. Frémeaux-Bacchi, E. Varon, X. Bossuyt, R. Cohen, L. Abel, J.L. Casanova, A. Puel and C. Picard. Invasive pneumococcal disease can reveal primary immune deficiency, *Clin Infect Dis*, vol. 59, n°2, p. 244-51, 2014.

Communication orale issue de la thèse

J. Gaschignard, A. Cobat, M. Orlova, NV. Thuc, NT. Huong, NN. Ba, VH. Thai, E. Schurr, A. Alcaïs. Human Genetics of leprosy polarization. *XVIII^{ème} Congrès international de la lèpre*, Bruxelles, Septembre 2013. *A reçu le prix de Best oral presentation on the basic science theme.*

Sommaire

TABLE DES ILLUSTRATIONS ET DEMONSTRATIONS.....	15
INTRODUCTION.....	17
DE LA LEPRE A SES FORMES CLINIQUES	19
1.1 BREVE HISTOIRE ET EPIDEMIOLOGIE.....	19
1.2 MYCOBACTERIUM LEPRAE	22
1.3 MANIFESTATIONS CLINIQUES	24
1.4 BREVE HISTOIRE DE LA CLASSIFICATION DES FORMES DE LEPRE	27
1.5 IMMUNOLOGIE.....	31
1.6 DIAGNOSTIC ET TRAITEMENT.....	35
1.7 LA LEPRE, UNE MALADIE HEREDITAIRE ?.....	36
1.8 UN PHENOTYPE ET UNE PROBLEMATIQUE : DES FORMES CLINIQUES A LA POLARISATION	39
ETAT DES CONNAISSANCES DE L'EPIDEMIOLOGIE CLASSIQUE ET GENETIQUE DES FORMES CLINIQUES DE LA LEPRE	43
2.1 ÉPIDEMIOLOGIE DES FORMES CLINIQUES DE LA LEPRE	43
2.1.1 Sexe	44
2.1.2 Age	44
2.1.3 BCG.....	46
2.1.4 Géographie	46
2.2 GENETIQUE DES FORMES CLINIQUES DE LA LEPRE.....	47
2.2.1 Analyses ADN-indépendantes : corrélations familiales, études de jumeaux, études de ségrégation	49
2.2.2 Etudes de liaison région-candidate puis pangénomiques.....	49
2.2.3 Analyses d'association gène-candidat, région-candidate puis pangénomiques	51
2.2.4 Méthodologie.....	56

SUJETS ET MATERIEL GENETIQUE	59
3.1 RECUEIL DES SUJETS.....	59
3.2 LA LEPRE AU VIETNAM	62
3.3 MATERIEL GENETIQUE	64
3.4 CONSTITUTION DE LA BASE DE DONNEES	65
3.5 DESCRIPTION DES SUJETS.....	68
3.6 CHOIX DU PHENOTYPE	73
 ANALYSE DE LIAISON	 75
4.1 METHODE DU MAXIMUM DE VRAISEMBLANCE BINOMIAL.....	75
4.1.1 <i>Cas binaire</i>	75
4.1.2 <i>Extension aux phénotypes quantitatifs : méthode nMLB–QTL</i>	77
4.2 CHOIX DES MARQUEURS GENETIQUES.....	80
4.3 DESCRIPTION DE L’ECHANTILLON DE FAMILLES MULTIPLES	80
4.4 AGREGATION FAMILIALE DES SOUS-TYPES DE LEPRE.....	81
4.5 RESULTATS DE L’ANALYSE DE LIAISON	82
4.6 DISCUSSION	86
 METHODES D’ETUDE D’ASSOCIATION GENETIQUE.....	 89
5.1 FBAT : PRESENTATION GENERALE.....	89
5.2 FBAT POUR UN TRAIT BINAIRE	92
5.2.1 <i>FBAT</i>	92
5.2.2 <i>FBAT_{+1/-1}</i>	94
5.3 NOUVEAUX TESTS D’ASSOCIATION FAMILIALE POUR UN TRAIT BINAIRE	95
5.3.1 <i>FBAT_{het}</i>	95
5.3.2 <i>FBAT_{het free} (modèle génétique additif)</i>	97
5.4 COMPARAISON ENTRE FBAT _{+1/-1} , FBAT _{HET} ET FBAT _{HET FREE}	99
5.4.1 <i>Comparaison entre FBAT_{het} et FBAT_{het free}</i>	99
5.4.2 <i>Autre écriture des tests</i>	102
5.4.3 <i>Comparaison entre FBAT_{+1/-1} et FBAT_{het free}</i>	102
5.5 DISCUSSION	104
	12

RESULTATS DE L'ETUDE D'ASSOCIATION	107
6.1 CHOIX DES MARQUEURS.....	107
6.2 ANALYSE PAN-GENOMIQUE PRIMAIRE.....	108
6.2.1 Analyse de la polarisation (définition OMS-82).....	108
6.2.2 Comparaison des associations de la polarisation et des formes PB et MB.....	110
6.2.3 Analyse de la polarisation en contrastant TT et LL.....	113
6.2.4 Synthèse graphique des analyses de liaison et d'association.....	114
6.2.5 Discussion.....	117
6.3 PROJET DE REPLICATION.....	117
PERSPECTIVES	119
SYNTHESE.....	119
LIMITES ET PERSPECTIVES IMMEDIATES	120
PERSPECTIVES DE RECHERCHE ORIGINALES.....	121
Recherche d'une hérédité mendélienne dans certaines familles.....	121
Le tatou à 9 bandes.....	122
La pléiotropie : lien avec les maladies inflammatoires de l'intestin	123
CONCLUSION	125
BIBLIOGRAPHIE	127
ANNEXES	143
DEMONSTRATIONS.....	143
PROTOCOLE D'IMPLEMENTATION DE FBAT _{HET FREE}	147
ARTICLE 1.....	149
ARTICLE 2.....	181
ARTICLE 3.....	199

Table des Illustrations et Démonstrations

Figures et tableaux

FIGURE 1. FRISE HISTORIQUE DES AVANCEES DANS LA LEPRE	19
FIGURE 2. DECLARATION DE NOUVEAUX CAS DE LEPRE DANS LE MONDE PENDANT L'ANNEE 2012	21
FIGURE 3. COMPARAISON DES GENOMES DE M. LEPRAE ET M. TUBERCULOSIS.	23
FIGURE 4. IMAGES CLINIQUES DE LA LEPRE TUBERCULOÏDE ET LEPROMATEUSE.	25
FIGURE 5. CARACTERISTIQUES DU SPECTRE CLINIQUE DE LA LEPRE.	26
FIGURE 6. LES DIFFERENTES CLASSIFICATIONS DE LA LEPRE.	30
FIGURE 6BIS. SCHEMA SIMPLIFIE DES CORRESPONDANCES ENTRE CLASSIFICATIONS.	30
FIGURE 7. SPECTRE IMMUNO-CLINIQUE DE LA LEPRE.	31
FIGURE 8. IMMUNOPATHOLOGIE DU SPECTRE DE LA LEPRE	34
FIGURE 9. PLAQUETTES DE POLYCHIMIOThERAPIE DES FORMES PAUCI ET MULTIBACILLAIRES	36
FIGURE 10. INFLUENCE DE L'HOTE ET DU PATHOGENE SUR LE COURS DE LA MALADIE	37
FIGURE 11. GENES DONT DES VARIANTS ONT ETE ASSOCIES A LA LEPRE.	38
FIGURE 12. MODELE GENETIQUE DE CONTROLE DE LA LEPRE EN DEUX ETAPES	40
FIGURE 13. LE CONCEPT DE POLARISATION POUR ETUDIER LA GENETIQUE DES FORMES CLINIQUES DE LEPRE	42
FIGURE 14. PROPORTION DES CAS MULTIBACILLAIRES EN FONCTION DE L'AGE ET DU SEXE.....	45
TABLEAU 1. PRINCIPAUX TYPES D'ANALYSES UTILISEES EN EPIDEMIOLOGIE GENETIQUE.	48
TABLEAU 2. LISTE DE GENES ASSOCIES AVEC UNE FORME CLINIQUE DE LEPRE OU AVEC SA POLARISATION.....	53
FIGURE 15. REPARTITION DES ETUDES GENETIQUES SELON LE PHENOTYPE DE LEPRE ASSOCIE	56
FIGURE 16. STRATEGIES D'ETUDE DE LA POLARISATION	58
TABLEAU 3. PROJETS DU LABORATOIRE AU VIETNAM ENTRE 1991 ET 2008	60
FIGURE 17. EXEMPLE DE FICHE INDIVIDUELLE	61
TABLEAU 4. PRINCIPALES CARACTERISTIQUES DE LA PUCE ILLUMINA HUMAN660W-QUAD.....	65
FIGURE 18. ARBRE DECISIONNEL D'ATTRIBUTION DU PHENOTYPE POUR DES DUPLICATS DISCORDANTS.	66
TABLEAU 5. REPARTITION DES 2 235 INDIVIDUS GENOTYPES SELON LE TYPE DE FAMILLE.	68
FIGURE 19. DISTRIBUTION DES FONDATEURS DE L'ECHANTILLON PAR ORIGINE ETHNIQUE	69
TABLEAU 6. CARACTERISTIQUES DE L'ECHANTILLON.	70
FIGURE 20. DIAGRAMME DE VENN DES CAS PB ET MB SELON 3 CLASSIFICATIONS.	71
TABLEAU 7. CARACTERISTIQUES DES 939 CAS DE LEPRE PARMI LES ENFANTS ET LES FONDATEURS.....	72

FIGURE 21. PROPORTION DE CAS MB SELON L'AGE ET LE SEXE DANS NOTRE ECHANTILLON VIETNAMIEN.....	72
TABEAU 8. CARACTERISTIQUES DE 162 ENFANTS DANS LES 72 FAMILLES DE L'ANALYSE DE LIAISON.....	81
FIGURE 22. REPARTITION DES FRATRIES DE L'ANALYSE DE LIAISON SELON LEUR CONCORDANCE CLINIQUE.....	82
FIGURE 23. RESULTAT DE L'ANALYSE DE LIAISON PANGENOMIQUE DE LA POLARISATION DE LA LEPRE	83
FIGURE 25. ANALYSE DE LIAISON PANGENOMIQUE POUR 5 PHENOTYPES.....	85
FIGURE 25. IMPACT DES FAMILLES CONTRIBUANT AU PIC DE LIAISON DU CHROMOSOME 19 SUR LE PIC DE LIAISON DU CHROMOSOME 16 ET INVERSEMENT.....	86
TABEAU 9. CODAGE DES GENOTYPES SELON LE MODELE GENETIQUE.....	90
TABEAU 10. ESPERANCE SOUS H_0 DU GENOTYPE X D'UN ENFANT.	91
FIGURE 26. CONTRIBUTION D'UN TRIO A FBAT, $FBAT_{HET}$ ET $FBAT_{HET FREE}$	98
FIGURE 27. DISTRIBUTION DE P SELON NA/N.	100
FIGURE 28. PUISSANCES COMPAREES DE $FBAT_{HET}$ ET $FBAT_{HET FREE}$ SOUS UN MODELE ADDITIF.	101
FIGURE 29. CRITERES D'EXCLUSION POUR LA SELECTION DES SNPs	107
FIGURE 30. ASSOCIATION POUR LA POLARISATION SELON L'OMS-82 AVEC FBAT ET $FBAT_{HET}$	109
FIGURE 31. ASSOCIATION POUR LA POLARISATION SELON L'OMS-82 AVEC $FBAT_{HET FREE}$ ET COMPARAISON AVEC LES MODELES ADDITIFS DE FBAT ET $FBAT_{HET}$	111
FIGURE 32. COMPARAISON DES ASSOCIATIONS POUR 3 PHENOTYPES AVEC FBAT.....	112
FIGURE 33. ASSOCIATION DE LA POLARISATION TT VS. LL AVEC FBAT ET $FBAT_{HET FREE}$	113
FIGURE 34. SYNTHESE DES RESULTATS DE LIAISON ET D'ASSOCIATION PANGENOMMIQUE.	114
FIGURE 35. SYNTHESE DES RESULTATS DE LIAISON ET D'ASSOCIATION SUR LE CHROMOSOME 19.....	115
TABEAU 11. MEILLEURS SNPs POUR L'ASSOCIATION AVEC LA POLARISATION.....	116
TABEAU 12. CARACTERISTIQUES DES CAS DE LEPRE DANS L'ECHANTILLON DE REPLICATION VIETNAMIEN.	118

Démonstrations

DEMONSTRATION 1. $FBAT_{+1/-1}$ EST UN TEST D'HETEROGENEITE Q DE COCHRAN	143
DEMONSTRATION 2. EQUIVALENCES DE CODAGE DANS FBAT SOUS UN MODELE ADDITIF	143
DEMONSTRATION 3. OFFSET QUI MINIMISE LA VARIANCE SOUS UN MODELE ADDITIF	144
DEMONSTRATION 4. $FBAT_{HET FREE}$ EST UN TEST DU χ^2 DE PEARSON.....	145
DEMONSTRATION 5. $FBAT_{HET} = pFBAT_{HET FREE}$	146

Introduction

Ce mémoire reprend la démarche que j'ai suivie pendant les trois années de ma thèse, dont la chronologie a parfois été modifiée pour rendre la narration plus fluide. Notre hypothèse de travail est qu'il existe des déterminants génétiques de l'hôte spécifiques aux différentes formes cliniques de la lèpre. Afin d'identifier cette composante génétique de l'hôte, notre stratégie d'analyse associe les outils classiques mais également plus récents de l'épidémiologie génétique.

La description de la lèpre fait l'objet du premier chapitre. Nous y précisons le phénotype d'intérêt que sont les formes paucibacillaire (PB) et multibacillaire (MB) de la lèpre. La recherche d'une définition univoque de notre phénotype d'intérêt nous a conduits à questionner les classifications utilisées par les médecins et chercheurs. Ces deux formes cliniques de lèpre peuvent être regroupées au sein d'un terme phénotypique unique, **la polarisation**, qui capture le contraste entre la forme PB et la forme MB.

Le deuxième chapitre étudie les facteurs de l'hôte, génétiques et non-génétiques, connus pour être associés à la polarisation de la lèpre. Principalement bibliographique, cette partie propose une relecture et des analyses originales à partir de la littérature la plus pertinente. Elle essaie en particulier de comprendre les limites des études antérieures qui ont contribué à la relative pauvreté de notre connaissance des facteurs génétiques impliqués dans le phénomène de polarisation de la lèpre.

La constitution de notre base de données est décrite dans le troisième chapitre. Notre population d'étude est composée de familles vietnamiennes avec des enfants PB, MB ou indemnes de lèpre clinique. L'information génétique à notre disposition est générée par une puce à ADN pangénomique.

Le quatrième chapitre est consacré à l'analyse de liaison génétique pangénomique de la polarisation de la lèpre. Nous utilisons un sous-échantillon de notre population, composé de familles incluant au moins deux enfants atteints de lèpre. Nous avons utilisé une méthode d'analyse de liaison développée au sein du laboratoire, le maximum de vraisemblance binomial,

en considérant notre phénotype soit comme un trait binaire soit comme un trait ordonné à cinq classes.

L'étude d'association fait l'objet des cinquième et sixième chapitres. La particularité de notre phénotype, ainsi que la structure particulière de notre échantillon - des familles, nous a conduits à proposer et implémenter une extension de la méthode de référence pour l'analyse d'association en familles aux traits binaires. Nous avons donc initié l'analyse de l'association pangénomique de la polarisation en incluant notre nouvelle approche.

Enfin, la discussion met en perspective les principaux résultats de notre travail, aborde les analyses complémentaires qui vont le prolonger, en particulier l'étude d'association. 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 à ce phénotype fascinant qu'est le phénomène de polarisation de la lèpre.

De la lèpre à ses formes cliniques

1.1 Brève histoire et épidémiologie

La lèpre, ou maladie de Hansen, est une maladie infectieuse chronique due à *Mycobacterium leprae*. Bien qu'elle soit mentionnée depuis l'Antiquité dans l'Histoire de plusieurs civilisations, les études récentes du génome de *M. leprae* semblent situer l'origine de la lèpre en Afrique de l'Est (Monot et al., 2005). C'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 (Hansen and Looft, 1895), quelques années avant que Koch ne décrive le bacille de la tuberculose. Endémique en Europe jusqu'au XVIII^{ème} siècle, et même jusqu'au début du XX^{ème} siècle en Norvège, la lèpre y a été apparemment éradiquée avant l'avènement de son traitement par les antibiotiques. Parmi les explications avancées figurent l'amélioration des conditions sanitaires et l'acquisition d'une immunité croisée avec la tuberculose (Ranque, 2007a).

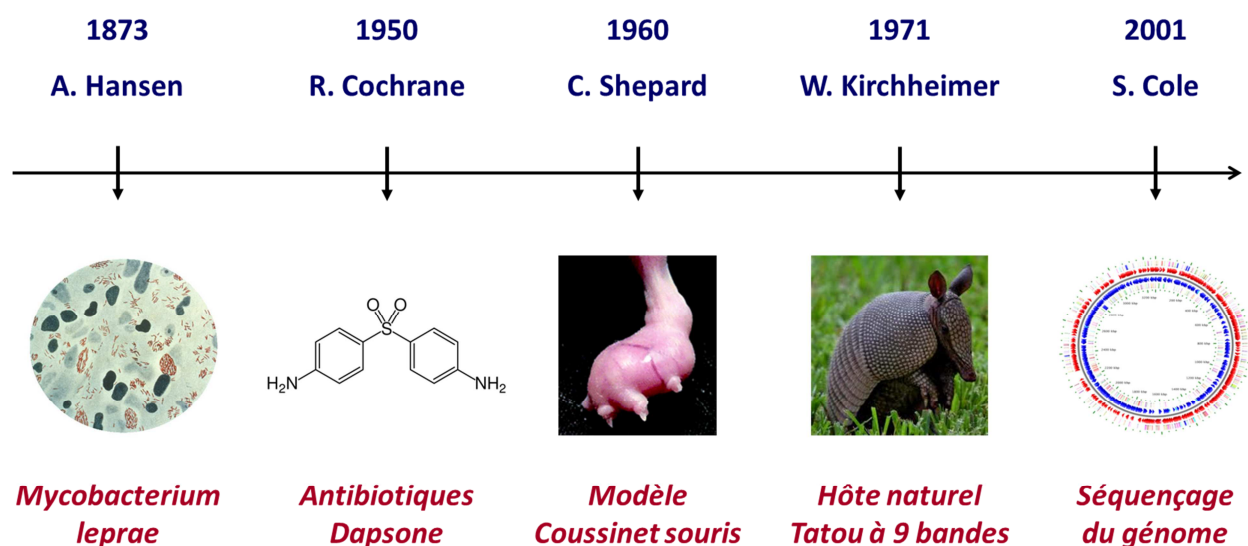


Figure 1. Frise historique des avancées dans la lèpre

Malgré de nombreuses rumeurs sur sa disparition, la lèpre représente toujours en 2015 la seconde mycobactériose humaine la plus fréquente après la tuberculose (causée par *M. tuberculosis*) (Portaels and de Jong, 2014). La découverte de la dapsonne comme traitement efficace de la lèpre par Cochrane en 1950 a été un tournant majeur pour les malades (Parascandola, 1996) (Figure 1). En 1981, afin de contrer la résistance de *M. leprae* à la dapsonne acquise pourtant dès 1964, l'Organisation Mondiale de la Santé (OMS) a promu une polychimiothérapie (PCT), associant dapsonne, rifampicine et clofazimine. L'OMS a ensuite initié, avec le support de fondations privées, une grande campagne visant à l'élimination de la maladie, en mettant gratuitement la PCT à disposition des malades. 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 $< 1 / 10\,000$. Au cours des 20 dernières années, plus de 14 millions de malades de la lèpre ont été guéris (WHO, 2014a). La prévalence (*i.e.* le nombre de malades sous traitement en fin d'année) dans le monde a chuté considérablement, passant de 5,2 millions en 1985 à 805 000 en 1995 et 180 618 en 2013.

Le taux de prévalence mondiale de la lèpre est ainsi passé de 21,1 pour 10 000 en 1985 à moins de 1 pour 10 000 en 2000, qui était le seuil fixé par l'OMS en 1991 pour déclarer la lèpre « éliminée » au niveau mondial (Britton and Lockwood, 2004). En 2013, la lèpre a été éliminée de 119 des 122 pays où elle constituait un problème de santé publique en 1985. Les efforts de l'OMS se concentrent actuellement sur l'élimination de la lèpre au niveau national (dans les pays d'endémie) et dans les poches d'endémie régionale, ainsi que sur la réduction de la stigmatisation et des séquelles neurologiques de la maladie.

Même « éliminée », la lèpre reste présente dans de nombreux pays en voie de développement, principalement en Afrique sub-saharienne, en Amérique du Sud et en Asie. Plus de 80 % des 215 656 nouveaux cas déclarés en 2013 l'ont été dans trois pays : l'Inde (126 913 cas), le Brésil (31 044 cas) et l'Indonésie (16 856 cas) (WHO, 2014b) (Figure 2, (WHO, 2013)). De manière inquiétante, depuis 2005, la prévalence et l'incidence de la lèpre sont restées stables *et identiques* (une anomalie dans le cadre d'une maladie chronique !) aux alentours de 200 000 cas annuels. Une augmentation du nombre de nouveaux cas a même été observée entre 2011 et 2012, avant une nouvelle baisse entre 2012 et 2013 (WHO, 2014b).

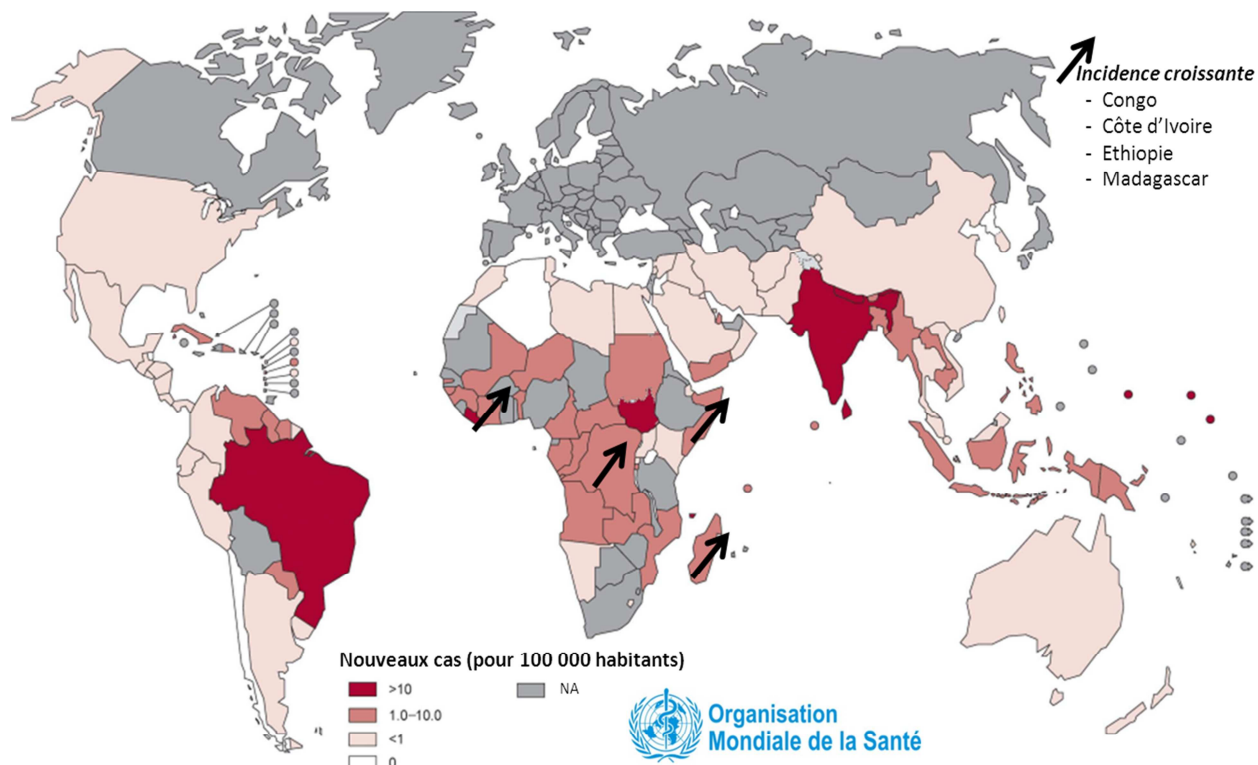


Figure 2. Déclaration de nouveaux cas de lèpre dans le monde pendant l'année 2012

Une certaine prudence doit accompagner l'interprétation de ces statistiques, qui reposent sur une déclaration active des cas de lèpre des Etats 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. Certains pays, comme l'Inde, ont mis fin aux activités de dépistage actif des cas et à l'examen des contacts. On estime qu'actuellement, en Inde, environ la moitié des cas de lèpre reste non déclarée (Lockwood et al., 2014). 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,

dans les mêmes pays où les chiffres pour la tuberculose ou le syndrome d'immunodéficience acquise (SIDA) sont arrondis à la dizaine de milliers.

Enfin, 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 la plan infectieux, un patient sur 15 présente une incapacité de grade 2 (définie par une déformation visible, ou une acuité visuelle $< 1/10$) (14 409 individus parmi les nouveaux cas en 2012) (WHO, 2013), et cette proportion augmente à la fin du traitement. On estime entre 2 et 3 millions le nombre de personnes vivant actuellement avec une séquelle de la maladie (Aubry). L'OMS a d'ailleurs fixé comme objectif de réduire, au niveau mondial, la proportion de nouveaux cas présentant une incapacité de degré 2 de plus de 35 % entre 2010 et 2015. La lèpre demeure ainsi à bien des égards une maladie tropicale négligée.

1.2 *Mycobacterium leprae*

Mycobacterium leprae, ou bacille de Hansen, est un bacille alcool-acido-résistant (BAAR) intracellulaire, non mobile, micro-aérophile, long de 1 à 8 μm et large de 0,3 μm . Le bâtonnet apparaît rouge avec la coloration de Ziehl-Neelsen. Il appartient à la famille des mycobactéries. Sa croissance est la plus lente du monde bactérien, avec un temps de duplication de 12 à 20 jours chez l'homme (Aubry) : cette croissance est optimale pour des températures plus faibles que les autres bactéries pathogènes chez l'homme, entre 32 et 35°C. Le lipide qui donne à *M. leprae* sa spécificité d'espèce est le glycolipide phénolique 1 (PGL-1), contre lequel peut se développer une réponse humorale détectable par sérologie. Le 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 (Flageul, 2011).

L'homme est son principal réservoir et le tatou à 9 bandes a plus récemment été identifié comme un hôte naturel du bacille (Kirchheimer and Storrs, 1971) (Figure 1). La culture de *M. leprae* n'a jamais pu être réalisée *in vitro*, mais elle est possible dans les coussinets plantaires de souris qui ont constitué le modèle expérimental de choix depuis 1960 (Shepard, 1960) (Figure 1). *M. leprae* est éliminé par la muqueuse nasale et par la peau en cas de plaie ou

d'ulcération. La transmission est directe par les sécrétions nasales (mouchage, crachats, aérosol). La transmission indirecte est possible par les objets souillés (Blake et al., 1987). Dans l'état actuel des connaissances on admet que la voie de pénétration est principalement respiratoire.

Le génome de *M. leprae* a été entièrement séquencé en 2001, 3 ans après celui de *M. tuberculosis* (Cole et al., 1998) (Figure 1). Le génome de *M. leprae* est plus petit que celui de *M. tuberculosis* (3,3 vs. 4,4 mégabases) avec une réduction du nombre de gènes potentiellement codants (1 614 vs. 3 993) et un nombre très élevé de pseudogènes (1 133 vs. 6) (Cole et al., 2001). La perte de voies métaboliques entières (catabolisme des acides gras et synthèse de l'enveloppe entre autres) explique sans doute son caractère strictement intracellulaire, son parasitisme quasi exclusivement humain, sa niche cellulaire très limitée (macrophages et cellules de Schwann) et sa croissance lente (Figure 3).

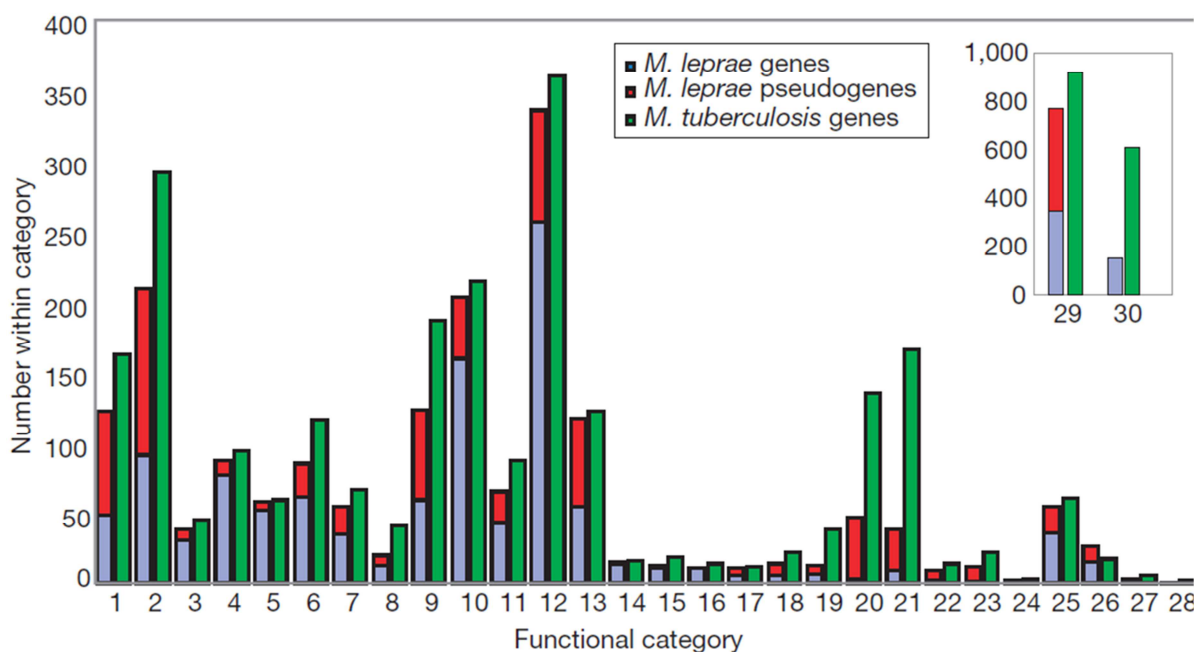


Figure 3. Comparaison des génomes de *M. leprae* et *M. tuberculosis*

Figure empruntée et légende adaptée de Cole (Cole et al., 2001). Répartition par catégorie fonctionnelle des gènes (bleu) et pseudogènes (rouges) de *M. leprae* et des gènes de *M. tuberculosis* (vert), d'après la séquence du génome publiée (Cole et al., 1998). 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.

La variabilité génomique de *M. leprae* est négligeable à l'échelle du globe (Monot et al., 2009), *i.e.* le bacille infectant est quasi identique quelle que soit la région du globe où il a été caractérisé. Sur le plan phylogénique, *M. leprae* est une espèce distincte, et est en particulier éloignée de *M. lepraemurium* qui infecte naturellement les souris (Alcaïs et al., 2005).

En 2008, Han *et al* ont décrit au Mexique une nouvelle espèce chez deux patients atteints d'une lèpre diffuse qu'ils ont nommée *M. lepromatosis* : le bacille différait de *M. leprae* (de 2 % pour le gène codant l'ARN ribosomal 16S, et de 6.9 % pour 14 autres gènes codants) (Han et al., 2008, 2009). Une équipe indépendante a confirmé en 2012 l'incrimination probable de cette mycobactérie chez une femme atteinte de lèpre diffuse au Mexique : l'histologie révélait des BAAR, la PCR était négative pour *M. leprae* mais positive pour *M. lepromatosis* (Vera-Cabrera et al., 2011). *M. lepromatosis* a ensuite été retrouvé au Brésil et en Birmanie, chez des patients présentant une lèpre tuberculoïde. Un fait intrigant est qu'il était souvent retrouvé conjointement avec *M. leprae* sur les prélèvements des patients (Han et al., 2014). Les arguments ne sont pas encore suffisants pour attribuer à cette nouvelle bactérie une pathogénie différente de celle de *M. leprae* (Gillis et al., 2011).

1.3 Manifestations cliniques

Le tropisme particulier de *M. leprae* pour la peau et les nerfs périphériques 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 (Flageul, 2011). Les manifestations cliniques de la lèpre se distribuent suivant un spectre clinique qui va d'un pôle dit « tuberculoïde » à un pôle dit « lépromateux ».

La forme tuberculoïde est caractérisée par un petit nombre de lésions cutanées (< 10), de grande taille (> 3 cm), bien délimitées, érythémateuses ou hypopigmentées, maculeuses (planes) ou infiltrées. Elles sont toujours hypo ou anesthésiques, caractéristique pathognomonique de la lèpre. La charge bacillaire au sein des lésions est faible. Cette forme est donc communément appelée « paucibacillaire » (PB). Des lésions nerveuses périphériques sont rapidement présentes dans l'évolution de la maladie. Il peut s'agir d'une simple hypertrophie

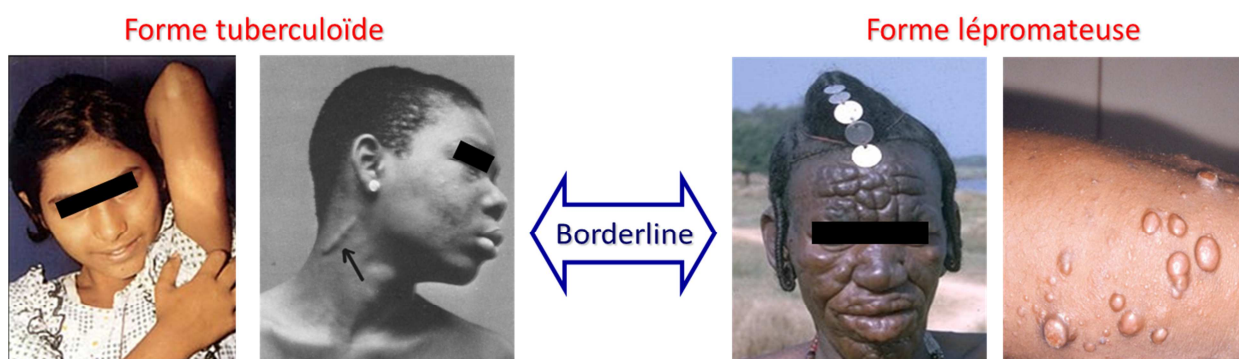


Figure 4. Images cliniques de la lèpre tuberculoïde et lépromateuse

De gauche à droite, lésion hypopigmentée, hypertrophie du nerf auriculaire cervical ([pathology 5.pathology.jhmi.edu/micro/v20n10.htm](http://pathology.5.pathology.jhmi.edu/micro/v20n10.htm)), faciès léonin (leprosyinformationweb.blogspot.com) et lépromes cutanés. (acupuncture.voila.net/epidemiologiedelalepreamayotte/index.html)

d'un nerf périphérique (Figure 4), mais le déficit moteur peut entraîner au niveau des mains et des pieds des déformations décrites comme des « déformations en griffe ». Des ulcérations, des maux perforants et des destructions osseuses distales peuvent compléter le tableau de cette atteinte nerveuse.

La forme lépromateuse est caractérisée par de nombreuses lésions cutanées (>20), maculeuses ou infiltrées, de petite taille (< 2cm). Les lésions infiltrées peuvent prendre l'allure de nodules appelés « lépromes » ou, par confluence, de placards annulaires mal délimités (Figure 4). Les lésions cutanées ne sont pas ou très peu hypoesthésiques (Flageul, 2011). La charge bacillaire est élevée, d'où le nom de lèpre « multibacillaire » (MB). L'atteinte des nerfs périphériques est moins constante que dans la forme tuberculoïde et progresse moins rapidement. Les manifestations extra-cutanées de la lèpre lépromateuse sont fréquentes. La rhinite lépreuse aboutit à l'ulcération de la cloison nasale et à l'effondrement du nez. L'œil est fréquemment atteint, et la maladie conduit à la cécité. L'atteinte testiculaire entraîne une azoospermie et une gynécomastie. Les autres organes internes (foie, ganglions lymphatiques, reins, rate, poumons, glandes endocrines...) peuvent être atteints.

Entre ces deux pôles, les patients sont dits « borderline ». La classification de Ridley et Jopling retient au total 5 groupes : Tuberculoïde tuberculoïde (TT), Borderline tuberculoïde (BT), Borderline borderline (BB), Borderline lépromateuse (BL) et Lépromateuse lépromateuse (LL). 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 (Figure 5) (Ridley and Jopling, 1966). Il existe par ailleurs des entités cliniques bien définies, comme la forme neurale pure, la lèpre histoïde de Wade ou la lèpre de Lucio (Languillon, 1999). Cette dernière s'exprime par une infiltration diffuse du derme, sans nodules, avec nécrose et ulcération des lésions. Nous mentionnons la lèpre de Lucio, principalement présente au Mexique, car c'est chez ces patients qu'a initialement été décrit *M. lepromatosis* (Han et al., 2008). 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.

FORME DE LEPRE	TT	BT	BB	BL	LL						
CLINIQUE											
signes généraux	—	—	—	±	+						
— cutanés											
— muqueux	—	—	+	++	+++						
— nerveux	+++	++	+	+	+						
— viscéraux	—	—	+	++	++						
BACILLE DE HANSEN											
• présence	—	±	±	++	+++						
• Globi	—	—	—	—	+						
IMMUNOLOGIE											
IDR de Mitsuda	+	±	—	—	—						
BIOPSIE (peau)											
— HISTOLOGIE	infiltrat tuberculoïde	infiltrat tuberculoïde ⊕ bande de UNNA	infiltrat mixte	infiltrat lépromateux	<table><tr><th>LL_s</th><th>LL_p</th></tr><tr><td>infiltrat lépromateux</td><td>infiltrat lépromateux ⊕ cellules de VIRCHOW</td></tr><tr><td>+++ +</td><td>+++ +++</td></tr></table>	LL _s	LL _p	infiltrat lépromateux	infiltrat lépromateux ⊕ cellules de VIRCHOW	+++ +	+++ +++
LL _s	LL _p										
infiltrat lépromateux	infiltrat lépromateux ⊕ cellules de VIRCHOW										
+++ +	+++ +++										
— BACILLE DE HANSEN											
• présence	—	±	+	++	+++						
• Globi	—	—	—	—	+++						

Figure 5. Caractéristiques du spectre clinique de la lèpre. Figure empruntée à Aubry (Aubry).

La réaction de type 1 (T1R), ou réaction de réversion, est une poussée inflammatoire brutale au niveau des lésions cutanées (oedème, 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. Une atteinte articulaire (polyarthrite et ténosynovite) peut compléter le tableau. Cette réaction est actuellement la principale cause de séquelle chez les malades de la lèpre (Ranque et al., 2007a). 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 donc rares dans les lésions (Aubry).

La réaction de type 2, ou érythème noueux lépreux (ENL), survient chez les patients LLs (sous groupe « subpolaire » de LL, Figure 5). Un facteur déclenchant, comme l'initiation du traitement, une infection ou une grossesse, est souvent retrouvé. Le début est brutal avec une altération de l'état général : fièvre, asthénie, arthromyalgies et céphalées. A 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 hypertrophique et douloureuse. Des atteintes systémiques (articulaire, ophtalmologique, testiculaire, rénale) sont également observées. Sur le plan immunologique, il s'agit d'une vascularite à complexes immuns circulants. L'évolution est le plus souvent marquée par plusieurs poussées avant la guérison.

1.4 Brève histoire de la classification des formes de lèpre

Dès 600 avant JC, le Sushruta Samhita – le plus ancien livre de médecine indienne - distinguait le *kushta*, une forme lépromateuse de la maladie, du *rakta*, une forme neuronale (Sharma, 1999). En 1881, Danielssen et Boeck différenciaient une lèpre nodulaire d'une lèpre anesthésique (Languillon, 1999). En 1895, Hansen et Looft précisait le terme «anesthésique» en «maculo-anesthésique» (Hansen and Looft, 1895). En 1935, Wade remarquait que certains patients peuvent passer d'une forme tuberculoïde à une forme lépromateuse, et dénommait «borderline» la forme de lèpre chez ces patients (Wade, 1935). De nos jours, une fois le

diagnostic de lèpre établi, l'identification de la forme clinique est essentielle, car elle permet de déterminer la durée du traitement antibiotique (WHO, 1982) et dans une certaine mesure la contagiosité du cas. De même, la détermination clinique du sous-type de lèpre fournit des renseignements physiopathologiques majeurs, qui ont rapidement suscité l'intérêt des immunologistes.

Les Conférences Internationales sur la Lèpre du Caire (1938), de La Havane (1948) et de Madrid (1953) ont ouvert la voie pour le classement de la maladie selon des critères cliniques (aspect de lésions), bactériologiques (identification des bacilles), immunologiques (test à la lépromine) et histologiques (nature de l'infiltrat, regroupement des bacilles). La classification dite de « Madrid » distingue ainsi trois formes de lèpre : Tuberculoïde (T), Borderline (ou Dimorphique) (B) et Lépromateuse (ou Virchowienne) (L) (Anonymous, 1953). Cette classification est encore utilisée dans de nombreux pays hispanophones ou lusophones, comme le Brésil, actuellement le deuxième pays le plus touché par la lèpre.

En 1966, Ridley et Jopling ont apporté des modifications importantes à cette classification. L'objectif était de mieux caractériser le groupe borderline qui comprenait des patients avec des caractéristiques cliniques, immunologiques, bactériologiques ou histologiques différentes. Ils ont alors proposé une nouvelle classification en cinq groupes : Tuberculoïde Tuberculoïde (TT), Borderline Tuberculoïde (BT), Borderline Borderline (BB), Borderline Lépromateuse (BL) et Lépromateuse Lépromateuse (LL) (Ridley and Jopling, 1966). Cette dernière forme peut elle-même être affinée en « sub-polaire » (LLs) et « polaire » (LLp) sur une base clinico-pathologique (Ridley and Waters, 1969) (cf figure 5).

Pour faciliter le diagnostic et le traitement de la lèpre sur le terrain, (OMS) a proposé en 1982 une classification simplifiée. Elle utilise l'indice bacillaire (IB), qui correspond à la moyenne sur plusieurs sites (lobes d'oreilles, lésion cutanée, mucus nasal) du nombre de bacilles par champ sur le frottis observé au microscope (échelle logarithmique) : un IB = 3 correspond ainsi à une moyenne de 1 000 bacilles par champ. Cette classification « OMS-82 » distingue la lèpre PB, qui comprend les patients TT et BT avec un $IB \leq 1$, de la lèpre MB, qui comprend les patients BB, BL et LL, ainsi que ceux ayant un indice bacillaire ≥ 2 (WHO, 1982).. En 1988, l'OMS a modifié ses critères bactériologiques : tout patient avec un IB positif était désormais classé comme MB (classification « OMS-88 » (WHO Expert Committee on Leprosy, 1988). Enfin,

en 1996, l'OMS a opté pour une définition opérationnelle afin d'optimiser le choix du traitement: les patients avec moins de cinq lésions cutanées sont diagnostiqués comme PB, tandis que tous les autres sont diagnostiqués comme MB (classification « OMS-96 »).

Dans le contexte de la diminution de la prévalence de la maladie – et aussi de sa connaissance par les soignants, cette classification présente l'avantage de la simplicité. Elle conduit cependant à un excès de diagnostic MB afin de prolonger le traitement. Ainsi, 2 % des cas MB seraient traités à tort comme PB et 20 % des cas PB traités à tort comme MB (Aubry). Mais dans les pays où la proportion de MB est élevée, comme aux Philippines (78 %), l'inverse semble se produire (Pardillo et al., 2007). Un autre effet secondaire de cette classification opérationnelle est sa décorrélation des fondements physiopathologiques qui sous-tendaient les classifications précédentes.

L'ensemble des classifications est résumé dans la figure 6. Dans la suite du manuscrit, nous nous référerons aux classifications établies par l'OMS en 1982, 1988 et 1996 par les acronymes « OMS-82 », « OMS-88 » et « OMS-96 ». La classification de référence sera celle de Ridley et Jopling en 5 classes. Il n'est pas toujours aisé de comparer des études utilisant des classifications différentes. Cependant, on peut situer schématiquement les frontières entre MB et PB des classifications OMS-82 et OMS-88 en fonction du spectre de Ridley-Jopling. La figure 6bis permettra au lecteur de comparer la plupart des classifications entre elles. Seule la classification OMS-96 n'apparaît pas, car comme nous le verrons dans la troisième partie, la concordance avec les autres classifications n'est pas optimale.

Madrid (1953) : T (Tuberculoïde), B (Borderline) et L (Lépromateuse), selon des critères cliniques (aspect des lésions cutanées et nerveuses), bactériologiques (indice bactériologique), immunologiques (réaction à la lépromine) et histologiques (nature de l'infiltrat).

Ridley-Jopling (1966) : TT (Tuberculoïde Tuberculoïde), BT (Borderline Tuberculoïde), BB (Borderline Borderline), BL (Borderline Lépromateuse) et LL (Lépromateuse Lépromateuse) selon des critères cliniques, bactériologiques, immunologiques et histologiques.

OMS (1982) : PB (Paucibacillaire) et MB (Multibacillaire) selon les classifications de Ridley et Jopling/Madrid et de l'indice bacillaire (IB) : PB = BT + TT (ou T) et IB ≤ 1 ; MB = LL + BL + BB (ou B + L) et/ou IB ≥ 2.

OMS (1988) : PB = BT + TT et IB = 0 ; MB = LL + BL + BB et/ou IB ≥ 1.

OMS (1996) : PB ou MB selon le seul nombre de lésions cutanées (≤/≥5).

Figure 6. Les différentes classifications de la lèpre

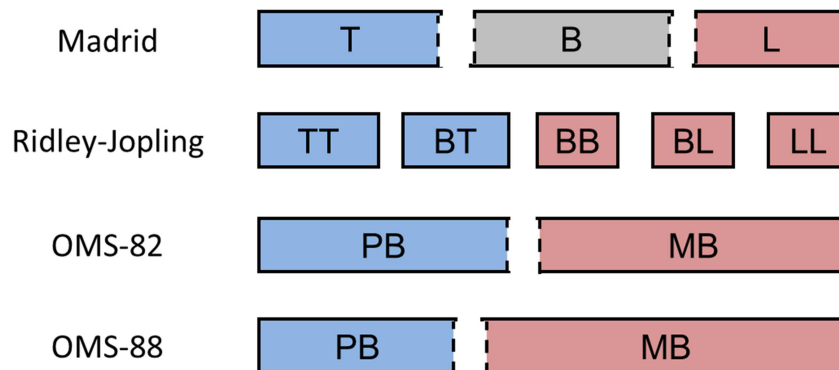


Figure 6bis. Schéma simplifié des correspondances entre classifications

Les cas paucibacillaires (PB) sont en bleu, les multibacillaires (MB) en rouge. Les cas Borderline (B) de la classification de Madrid ne peuvent être dichotomisés en PB ou MB.

1.5 Immunologie

Les différentes formes cliniques de lèpre sont déterminées par la réaction immunologique sous-jacente de l'hôte à *M. leprae* (figure 7). La lèpre est un exemple remarquable de l'impact de l'équilibre entre immunité cellulaire et humorale sur les aspects cliniques de la maladie.

Les patients atteints de la lèpre tuberculoïde (TT) développent une forte réponse immunitaire cellulaire contre la mycobactérie, ce qui limite la maladie à quelques tâches cutanées et troncs nerveux. Les lésions sont infiltrées par des lymphocytes T CD4⁺ qui sécrètent de l'interféron- γ (IFN- γ) et de l'interleukine 2 (IL-2). Ils forment des granulomes giganto-épithélio-cellulaires bien délimités autour de nerfs dermiques. Les lésions sont vierges, ou presque, de mycobactéries. La forte immunité cellulaire est confirmée par la prolifération des cellules T après stimulation par des antigènes de *M. leprae* *in vitro* (Britton and Lockwood, 2004). *In vivo*, le test à la lépromine (ou test de Mitsuda) est positif. Ce test consiste en l'injection intradermique de 0,1 ml de lépromine (suspension stérile de bacilles tués par la

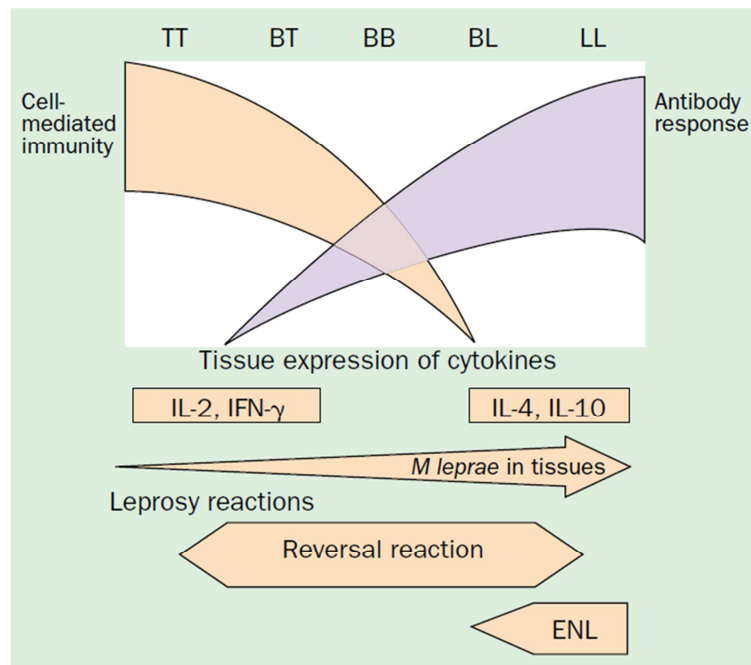


Figure 7. Spectre immuno-clinique de la lèpre. Figure empruntée à Britton (Britton and Lockwood, 2004). IL = Interleukine, IFN = interféron, ENL = érythème noueux lépreux.

chaleur : se forme alors au point d'injection un granulome giganto-épithélio-cellulaire, associées à un infiltrat lymphocytaire. La lecture du test se fait après 28 jours par la mesure de l'induration cutanée. Une infiltration ≥ 3 mm est considérée comme positive (Wade, 1953; Ilangumaran et al., 1996). 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 < 3mm).

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. La figure 8 illustre le parallèle entre ces deux phénomènes. Ces formes borderline sont cliniquement instables, et peuvent évoluer soit lentement vers le pôle lépromateux, soit brutalement par une réaction de réversion.

L'absence de réponse des lymphocytes T dans la lèpre lépromateuse est au cœur des recherches. 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. 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 IFN- γ , 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 (García et al., 1999). A 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 (Sieling and Modlin, 1994). Or l'association d'IL-2 et d'IL-12 recombinantes restaure la réponse Th1 des lymphocytes chez certains individus lépromateux (de Jong et al., 1997). 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.

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 (Sieling and Modlin, 1994). 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 (Quaresma et al., 2014; Saini et al., 2013). 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. 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 diminuée dans les lésions cutanées lépromateuses (Murray et al., 2007; Modlin, 2010). Par ailleurs, Masaki et al montrent dans une très belle étude comment *M. leprae* parvient à reprogrammer les cellules de Schwann qu'il infecte pour qu'elles se différencient en un type de cellule souche mésenchymateuse, phénomène qui contribue à la formation de granulomes (Masaki et al., 2013).

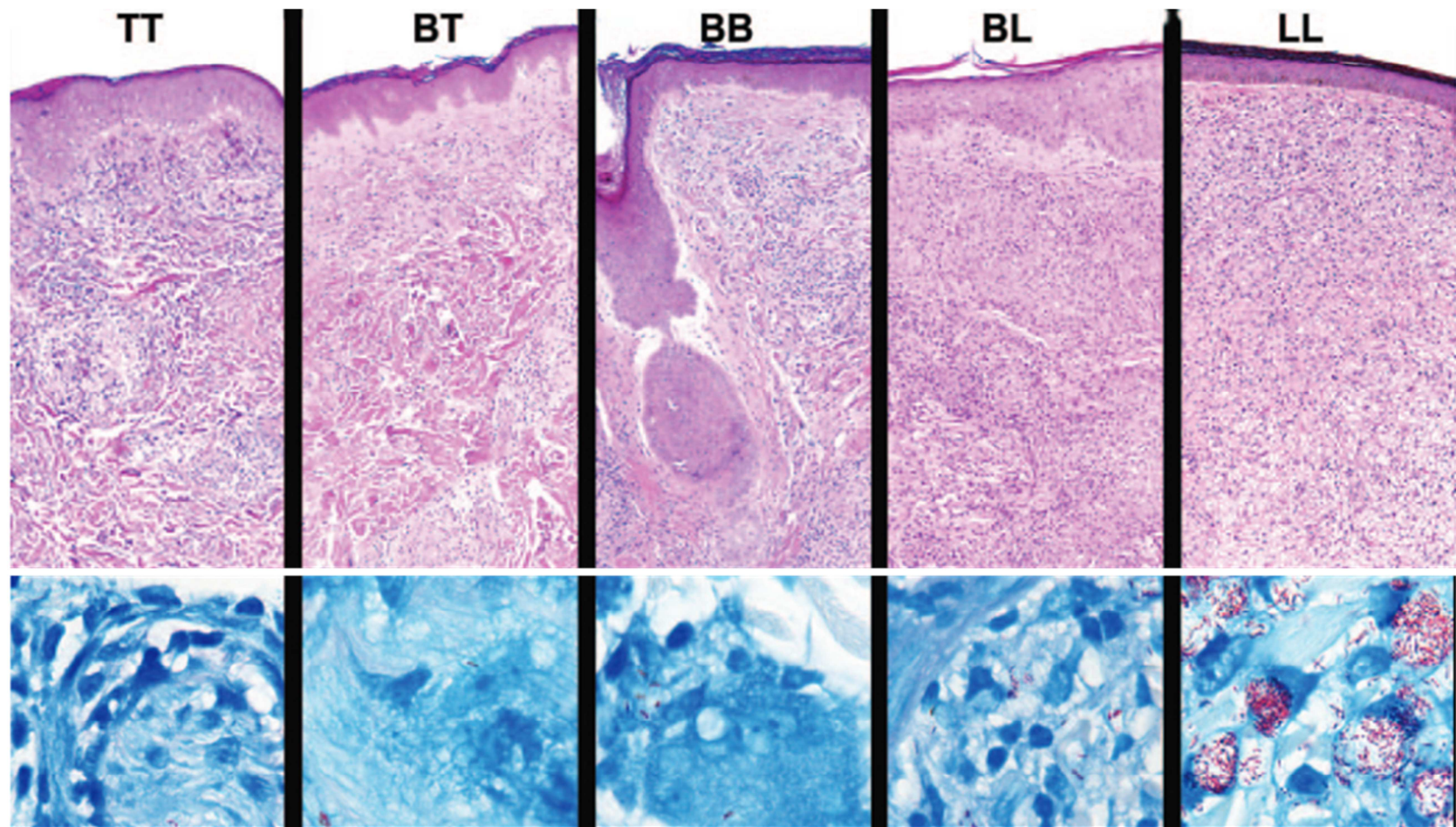


Figure 8. Immunopathologie du spectre de la lèpre. Figure empruntée à Scollard (Scollard et al., 2006).

Chaque champ représente l'histopathologie d'un type de lèpre selon la classification de Ridley et 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.

1.6 Diagnostic et traitement

Le diagnostic de la lèpre est le plus souvent clinique. Il est aisé lorsque des signes pathognomoniques sont présents : hypertrophie d'un nerf périphérique et lésion cutanée anesthésique. 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). Pour la forme tuberculoïde, il n'existe d'ailleurs pas d'autre examen paraclinique à visée diagnostique. En effet, la recherche de bacille et la sérologie anti-PGL-1 sont le plus souvent négatives, et la Polymerase Chain Reaction (PCR) à la recherche d'antigènes de *M. leprae* est peu sensible chez ces patients (Flageul, 2011). Le test de Mitsuda ne peut non plus être utilisé pour le diagnostic. Bien que toujours négatif chez les patients LL, il peut être positif chez des individus non infectés par *M. leprae*. En effet, l'infection par d'autres mycobactéries, voire des bactéries intracellulaires, le positivent. Par ailleurs, le recours à l'histologie n'est possible que dans les centres de référence. La plupart des patients sont donc diagnostiqués et classés sur la seule clinique.

Six antibiotiques sont couramment utilisés pour traiter des cas de lèpre : la dapsonne, la rifampicine, la clofazimine, l'ofloxacine, la minocycline et la clarythromycine (Languillon, 1999). Deux autres quinolones, la moxifloxacine et la péfloxacine, ont également été utilisées avec succès (Fajardo et al., 2004; Pardillo et al., 2008). La rifampicine, introduite après la survenue de résistance à la dapsonne (le 1^{er} antibiotique utilisé) est la plus bactéricide : 1 dose de 600 mg tue 99,9 % des organismes viables. Les cas de résistance à la rifampicine ont conduit à réaliser des essais en 1982 avec des polychimiothérapies (PCT) associant rifampicine, dapsonne et clofazimine. 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 (Britton and Lockwood, 2004). 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 utilisée 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 au décès (Zhang et al., 2013). Ce polymorphisme serait présent chez 1 à 12 % des Indiens, pays qui déclare le plus de cas de lèpre.

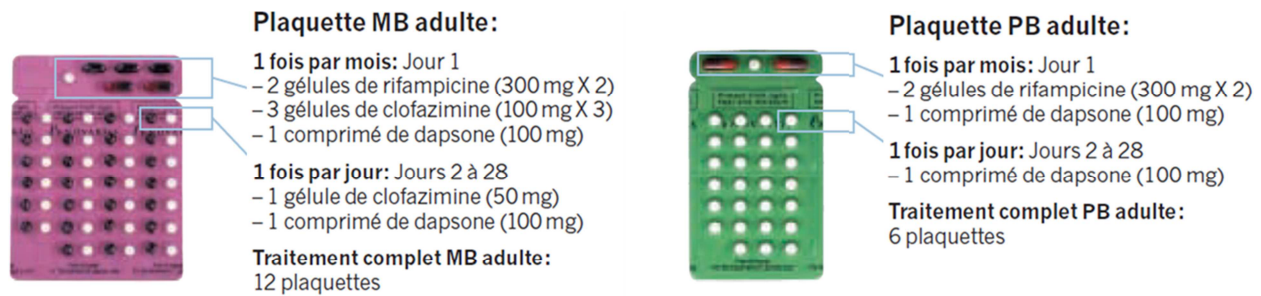


Figure 9. Plaquettes de polychimiothérapie des formes pauci et multibacillaires

L'OMS recommande actuellement une bithérapie de 6 mois pour les cas PB et une trithérapie de 12 mois pour les cas MB (Figure 9). Ce traitement est mis gratuitement à disposition par une fondation privée et est distribué par l'OMS au niveau de chaque pays. Si les cas de résistances à l'un des trois antibiotiques augmente (3 196 cas en 2012), aucune résistance à la PCT n'a jusqu'à présent été rapportée (WHO, 2013). L'ofloxacine et la minocycline ont été utilisées en monodose (en association avec la rifampicine) pour traiter les patients qui ne présentaient qu'une seule lésion : ces individus bénéficient actuellement du traitement du groupe PB. Nous n'abordons pas le traitement prophylactique, dont l'efficacité à long terme est discutée (Bakker et al., 2005; Oskam and Mi, 2007).

1.7 La lèpre, Une maladie héréditaire ?

Jusqu'au XIX^{ème} siècle, le caractère « héréditaire » de la lèpre était bien ancré dans les mentalités. La découverte du bacille de Hansen en 1873 et de l'étiologie microbienne de la lèpre a cependant mis à mal l'idée d'une contribution potentielle de la génétique de l'homme dans le développement de la maladie. L'absence d'un modèle expérimental satisfaisant et l'incapacité à cultiver l'agent pathogène ont également contribué à faire de la lèpre un exemple récurrent de maladie purement environnementale, y compris dans de nombreux ouvrages classiques de génétique humaine. Au milieu du siècle dernier, l'idée d'une prédisposition génétique au développement de la maladie ressurgit peu à peu, suggérée notamment par la grande diversité des réponses cliniques à l'exposition à *M. leprae*. On considère d'ailleurs aujourd'hui que la lèpre est peut-être la maladie infectieuse commune où la contribution génétique de l'hôte est la

plus forte (Alter et al., 2011; Gaschignard et al., 2013). Un biais possible est que la quasi-clonalité de *M. leprae* rend la reproductibilité des résultats plus facile que pour d'autres pathogènes.

La figure 10 présente un schéma général de l'influence des caractéristiques du pathogène et de l'hôte dans le passage de l'exposition à l'infection, puis de l'infection à la maladie. Dans le cas particulier de la lèpre, il est observé que le risque de transmission après contact avec un patient atteint est plus élevé lorsque le patient index est porteur de lésions riches en bacilles (formes MB). Mais les études de cohortes vaccinales montrent que seuls 5 % des individus exposés à *M. leprae* développent une lèpre (Convit et al., 1992; Gupte et al., 1998).

Ensuite, l'agent pathogène ne joue qu'un rôle marginal car *M. leprae* présente une variabilité génomique négligeable à l'échelle du globe (Monot et al., 2009). 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 (Caws et al., 2008; Singh and Cole, 2011). La découverte de *M. lepromatosis* pourrait nous amener à réviser cette position. Cependant, d'une part *M. lepromatosis* n'a pas été identifié dans l'ensemble des pays endémiques – dont l'Inde, d'autre part ce bacille ne semble pas associé comme on le pensait initialement à un type particulier de lèpre (Han et al., 2014).

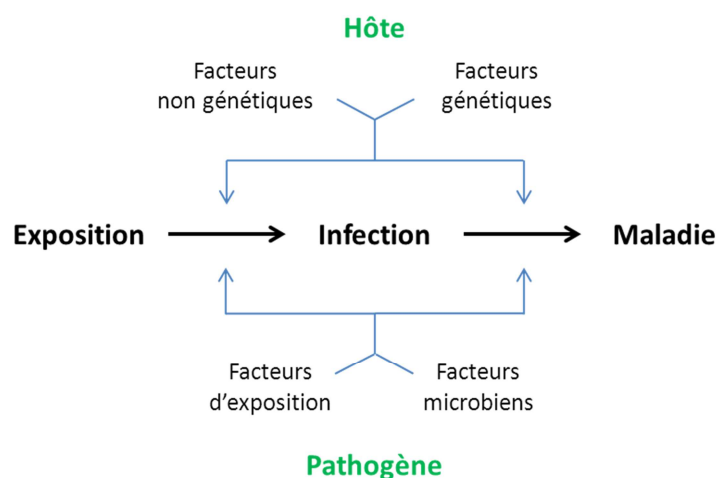


Figure 10. Influence de l'hôte et du pathogène sur le cours de la maladie

Différents éléments non spécifiquement génétiques pourraient contribuer à cette variabilité inter-individuelle (état nutritionnel, comorbidités, immunosuppression,...). Cependant, il n'y a aucun facteur de risque environnemental clairement établi, en dehors de l'intensité du contact avec un malade (le plus souvent MB), malgré le grand nombre d'études menées afin de les identifier (Moet et al., 2004; Ranque, 2007b). 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 (Scollard, 2005).

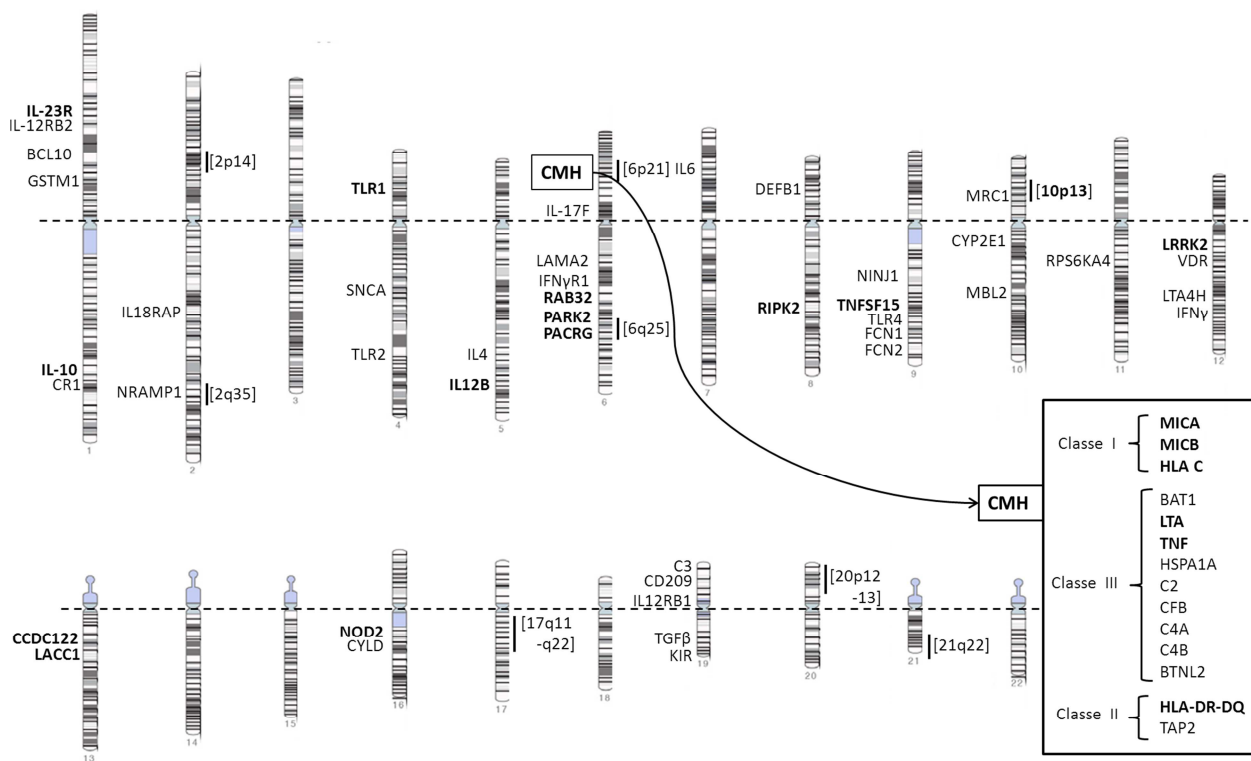


Figure 11. Gènes dont des variants ont été associés à la lèpre. Figure tirée de Gaschignard et al (Gaschignard et al., 2013). A gauche de chaque chromosome figurent les gènes pour lesquels un marqueur a été associé à la lèpre, et à droite les régions de liaison. Les gènes associés, soit dans un GWAS, 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. La description détaillée des différents éléments de nomenclature - liaison, association, GWAS sont détaillées dans le chapitre suivant.

Trois revues (entre autres) détaillent les gènes identifiés dans la lèpre (Misch et al., 2010; Alter et al., 2011; Gaschignard et al., 2013). Ces découvertes sont synthétisées dans la figure 11 où sont représentés les gènes pour lesquels un variant a été associé à la lèpre ainsi que les régions liées à la maladie. La plupart des associations proviennent d'études dites « gène-candidat », pour lesquelles un gène est choisi en raison d'*a priori* biologiques. Aussi, la part belle est faite aux gènes reliés au système immunitaire, comme ceux du CMH et ceux codant pour des cytokines. Des approches sans *a priori* ont permis d'identifier, entre autres, le gène *PARK2/PACRG* codant pour la protéine parkin et sa région régulatrice, également impliquée dans les formes précoces de la maladie de Parkinson (Mira et al., 2004). Une étude fonctionnelle a ensuite mis en évidence le rôle de *PARK2* dans la production d'IL-6 et le chimiotactisme des monocytes (de Léséleuc et al., 2013). Une autre a montré que parkin jouait un rôle dans l'autophagie de *M. tuberculosis* (Manzanillo et al., 2013). L'identification du gène *PARK2* est ainsi une belle illustration du succès de l'approche par « génétique inverse », où l'implication d'un gène dans une pathologie permet de découvrir ou de mieux comprendre les fonctions de ce gène.

1.8 Un phénotype et une problématique : des formes cliniques à la polarisation

Suite aux succès des études de génétique humaine concernant la lèpre, un « modèle en deux étapes » a été proposé. Un premier groupe de gènes (comme *PARK2/PACRG* ou *LTA* codant pour la lymphotoxine α) influencerait le passage de l'exposition à la lèpre *per se*, i.e. indépendamment de sa forme clinique. Un second groupe de gènes (comme le gène codant pour le CMH DR) influencerait le développement de l'une ou l'autre forme de lèpre (Ranque et al., 2008; Alter et al., 2011) (Figure 12). La figure 10 sépare le passage de l'exposition à la maladie en deux étapes : exposition à infection, puis infection à maladie. L'absence de marqueur fiable d'infection 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. Aussi, il n'est

pas possible de savoir si les gènes du 1^{er} groupe jouent un rôle au niveau de l'infection ou de la maladie.

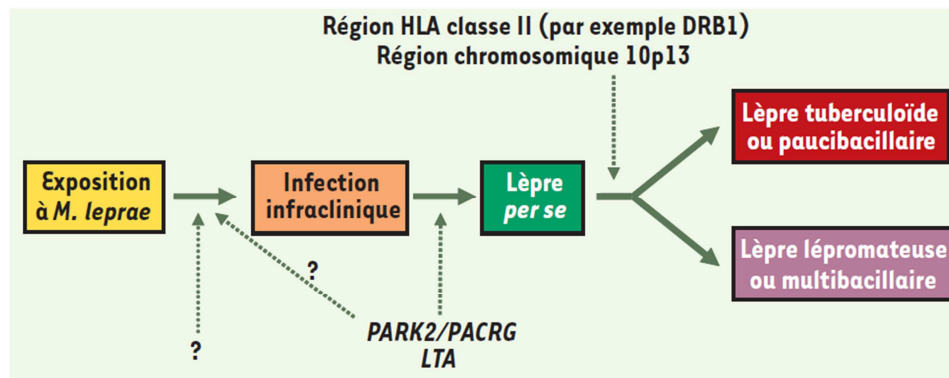


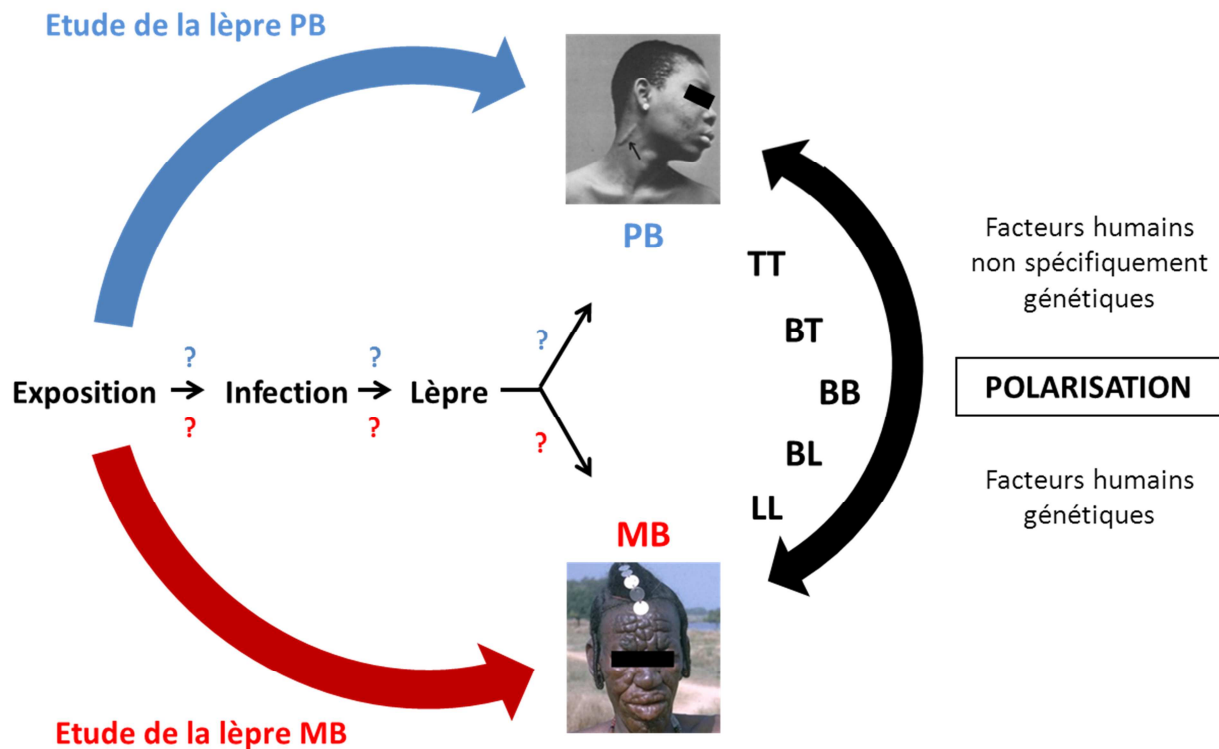
Figure 12. Modèle génétique de contrôle de la lèpre en deux étapes
Figure empruntée à Ranque (Ranque et al., 2008)

La figure 13 montre qu'il n'est pas non plus aisé lorsqu'on s'intéresse aux formes cliniques de la lèpre, de différencier les gènes qui impactent spécifiquement l'étape de la lèpre *per se* à la forme PB, si l'on compare des sujets exposés à des sujets PB (idem pour la forme MB). Notre hypothèse de travail est qu'il existe des gènes qui agissent spécifiquement dans la détermination du sous-type de lèpre. Nous utiliserons le mot « polarisation » pour désigner tout processus qui différencie un individu infecté par *M. leprae* développant une forme PB d'un individu développant une forme MB. Et ce, quelle que soit la classification utilisée pour définir les cas PB et MB (classifications de Madrid, OMS-82, OMS-88 et OMS-96, cf figure 6).

Cette première partie introductive sur les généralités de la lèpre et ses formes cliniques nous a appris, qu'à côté des définitions binaires PB et MB, il est possible de définir des formes cliniques plus précises, comme le proposent Ridley et Jopling. Leur classification en 5 groupes repose sur notre compréhension actuelle de la physiopathologie de la lèpre. Ce raffinement peut être incorporé dans notre concept de polarisation, et le phénotype binaire devenir un phénotype catégoriel ordonné.

Pour mener à bien l'étude de la génétique humaine des formes cliniques de lèpre, envisagée comme une polarisation, il nous faut aussi considérer que des facteurs non génétiques de l'hôte puissent contribuer à cette polarisation. A la différence du schéma général de la figure 10, nous excluons la possibilité d'un rôle majeur du génome de *M. leprae* du fait de sa population quasi clonale (Cole et al., 2001; Monot et al., 2009; Singh and Cole, 2011). Cette situation est assez rare dans le monde microbien et de nombreux exemples du monde bactérien (comme *M. tuberculosis*) ou virale (comme le virus de l'hépatite C), illustrent que le génome du pathogène peuvent modifier non seulement la clinique mais également les cascades physiopathologiques mises en œuvre par l'hôte (Di Pietrantonio et al., 2010; Portevin et al., 2011; Coscolla and Gagneux, 2014).

Le deuxième chapitre de notre manuscrit a donc cherché à identifier des facteurs de risque de l'hôte le prédisposant à une forme de lèpre plutôt qu'à une autre. Cette partie précisera aussi, parmi les gènes connus pour être associés à la lèpre *per se*, ceux qui semblent jouer un rôle dans la polarisation de la maladie.



Polarisation : processus qui conduit un individu infecté par *M. leprae* vers la forme paucibacillaire ou la forme multibacillaire, quelle que soit la classification utilisée pour définir les cas PB et MB (classifications de *Madrid*, *OMS-82*, *OMS-88* et *OMS-96*).

Figure 13. Le concept de polarisation pour étudier la génétique des formes cliniques de lèpre
Les études de la lèpre PB (MB) comparent des sujets exposés à des cas PB (MB). Elles ne peuvent distinguer la contribution au développement de l'infection ou de la lèpre per se de la contribution au sous-type de lèpre (figurées par des points d'interrogation). Les études de polarisation comparent elles directement des cas PB à des cas MB. Elles s'affranchissent donc d'une possible contribution à l'infection ou à la lèpre per se. De plus, dans les études de polarisation, le phénotype binaire peut-être raffiné, comme ici en un phénotype ordonné à 5 catégories reprenant la classification de Ridley et Jopling.

Etat des connaissances de l'épidémiologie classique et génétique des formes cliniques de la lèpre

2.1 Épidémiologie des formes cliniques de la lèpre

A notre connaissance, aucune étude épidémiologique portant spécifiquement sur la forme MB ou PB n'a jamais été publiée. Nous pouvons méconnaître de nombreuses publications confidentielles publiées dans les pays endémique. Par ailleurs, une revue des données épidémiologiques sur l'érythème noueux lépreux, une des complications majeures de la lèpre, a révélé la grande difficulté de synthétiser les données lorsque le phénotype principal d'intérêt des études n'est pas la lèpre *per se* (Voorend and Post, 2013). Par conséquent, afin de décrire l'épidémiologie des formes cliniques de la lèpre, nous avons extrait des données d'études épidémiologiques plus générales portant sur la lèpre *per se* et qui mentionnaient les formes cliniques. Nous mesurons que ce procédé introduit un biais de sélection des données. De plus, utiliser les données les plus récentes n'a pas été possible en raison même de l'épidémiologie de la maladie. En effet, avec la diminution du nombre de nouveaux cas dans les années 1990, les études épidémiologiques récentes se concentrent plus sur les facteurs de risque de survenue de cas secondaires que primaires : avec un taux de détection de nouveaux cas souvent inférieur à 1 % après 4 ans de suivi, plus de 10 000 contacts sont nécessaires pour finalement étudier 100 cas secondaires de lèpre (Antunes et al., 2013; Bakker et al., 2006; Moet et al., 2004, 2006; Dos Santos et al., 2013). Ceci explique les nombreuses références à des publications antérieures à 1990.

Nous avons pu étudier de façon convaincante un nombre relativement restreint de facteurs de risque liés à l'hôte: le sexe, l'âge, l'origine géographique et le statut de la vaccination par le BCG. Sauf précision contraire, nous avons réalisé les analyses avec les fonctions de base du logiciel libre R, version 3.1.2 (R Development Core Team, 2008).

2.1.1 Sexe

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 2 (WHO, 2013). 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 (Le Grand, 1997).

Peu d'études se sont intéressées spécifiquement à l'estimation du sex-ratio chez les patients PB et les patients 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 (définis selon la classification de Madrid) (Guerra-Silveira and Abad-Franch, 2013). 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. Le docteur Abad-Franch, de l'Institut Leonidas et Maria Deane – Fiocruz de Manaus, a gracieusement mis à notre disposition les données du recensement brésilien des cas de lèpre entre 2006 et 2010. Nous avons pu calculer la proportion de cas MB, définie comme le rapport $L/(T + L)$ à partir de la classification de Madrid. Entre 2006 et 2010, cette proportion était au Brésil de 26,3 % chez les femmes contre 54,6 % chez les hommes ($p < 10^{-10}$) (Guerra-Silveira and Abad-Franch, 2013).

2.1.2 Age

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 (Fine, 1982). Une étude réalisée en Andhra Pradesh (Inde) en 1962 a montré que la proportion des cas MB (définition de Madrid) 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 (de Vries and Perry, 1985). Deux études réalisées au cours de la même période, l'une dans la région indienne du Tamil Nadu, l'autre en Birmanie (toujours en utilisant la classification de Madrid) retrouvaient cette même augmentation avec l'âge (Rao et al., 1972; Martínez Domínguez et al., 1980) (figure 14). Une étude plus récente menée en Uttar Pradesh (Inde) entre 2005 et 2010 retrouve la même tendance avec une proportion des cas MB (classification non précisée) de 54 % chez les enfants (seuil d'âge malheureusement non précisé) contre 77 % chez les adultes ($p < 0,001$) (Mukherjee et al., 2013).

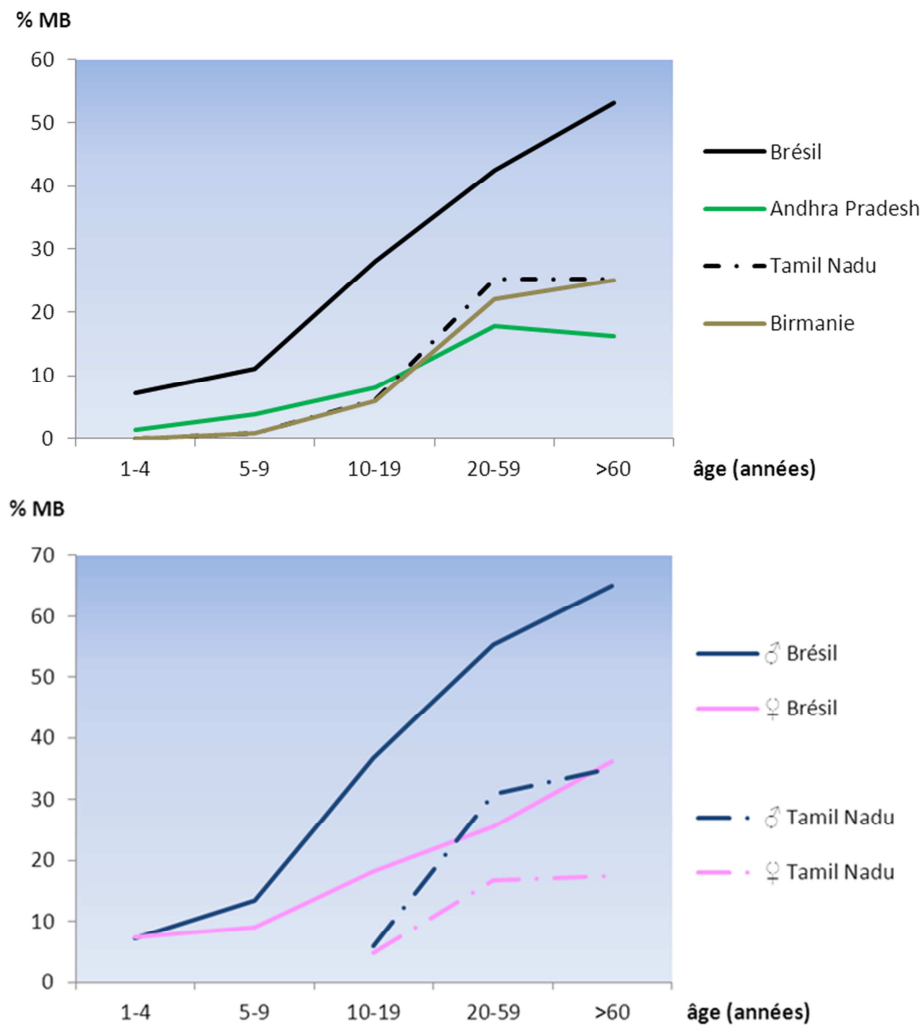


Figure 14. Proportion des cas multibacillaires en fonction de l'âge et du sexe

Brésil : données nationales, période 2006-2010, classification de Madrid (Guerra-Silveira and Abad-Franch, 2013) ; Tamil Nadu : 8 909 cas de la région de Gudiyatham, entre 1962 et 1970, classification de Madrid (Rao et al., 1972) ; Andhra Pradesh : 47 530 cas entre 1962 et 1970, classification non précisée (de Vries and Perry, 1985) ; Birmanie : 1 962 cas dans la région de Singu entre 1964 et 1972, classification de Madrid (Martínez Domínguez et al., 1980).

Enfin, nous avons réanalysé les données de l'étude brésilienne de Guerra-Silveira *et al* et nous avons là encore observé que la proportion de cas MB augmentait avec l'âge, passant de moins de 10 % avant 10 ans à plus de 50 % après 60 ans (figure 14) (Guerra-Silveira and Abad-Franch, 2013).

2.1.3 BCG

Deux méta-analyses ont montré que la vaccination par le BCG réduisait significativement le risque de présenter des signes cliniques de lèpre (Setia et al., 2006; Merle et al., 2010). Les deux méta-analyses comprenaient respectivement 26 et 28 études, qui elles-mêmes utilisaient diverses classifications des sous-types de lèpre (Madrid, Ridley-Jopling, OMS-82, OMS-88 et OMS-96); 22 études étaient communes aux deux méta-analyses. 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, intentionnelles ou l'ensemble des études. Dans la première, la réduction était plus importante dans le groupe MB que dans le groupe PB, bien que seulement marginalement significative au risque 5%, (78 % vs. 59 %, $p = 0,04$). Dans la seconde, un ajustement a été fait sur le nombre de doses de BCG reçues et sur le type des études incluses (observationnelles vs. interventionnelles) : l'impact du BCG variait alors de 41 à 68 % mais n'était plus significativement différent dans les groupes PB ou MB. 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) (Thuc et al., 1994).

2.1.4 Géographie

La proportion des cas MB varie considérablement entre les continents, mais aussi au sein de chaque continent, y compris entre pays limitrophes. Parmi les pays qui ont déclaré plus de 1 000 nouveaux cas en 2012, la proportion des cas MB (classification OMS-96) variait de 45% (Bangladesh) à 92 % (Philippines) (WHO, 2013). 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, Kenneth Newell, un épidémiologiste

qui se définit lui-même comme « béotien » au sujet de la lèpre, a publié un article passé bien trop inaperçu intitulé « Regard d'un épidémiologiste sur la lèpre » (Newell, 1966). 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 selon la classification de Madrid) 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 rend 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 (Newell, 1966).

2.2 Génétique des formes cliniques de la lèpre

Il existe de nombreux outils méthodologiques en épidémiologie génétique selon les données dont on dispose et les hypothèses que l'on souhaite tester. Leurs principales caractéristiques sont résumées dans le tableau 1. On remarquera qu'il est possible d'aborder la question d'une contribution génétique sur la seule base des liens familiaux sans disposer d'échantillons biologiques. Ces méthodes ADN-indépendantes sont en apparence faciles à mettre en œuvre et peu coûteuses. Deux éléments importants dans le contexte du transfert de connaissance vers les pays endémiques. Toutefois, elles supposent la possibilité de reconstruire sans ambiguïté des arbres généalogiques ce qui n'est pas toujours aisé, car la définition du lien familial varie selon les cultures. Et dans certains domaines, le coût des phénotypes étudiés a considérablement augmenté. La section suivante illustre toutefois comment elles peuvent contribuer à l'étude du déterminisme de la polarisation de la lèpre.

Tableau 1. Principaux types d'analyses utilisées en épidémiologie génétique

	Echantillon	Marqueurs	Objectif	Principe	Limitations	
Analyses ADN- indépendantes	Agrégation familiale	Familles	Pas de marqueur génétique	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 le risque lié à des facteurs génétiques et celui lié à un mode de vie et/ou un environnement communs
	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 dans les études les plus anciennes
	Adoptés	Adoptés et leurs parents biologiques et adoptifs		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	Comparer la concordance phénotypique entre les enfants et leurs parents naturels et adoptifs	Rareté des cas Age à l’adoption
	Ségrégation	Familles ≥ 2 générations		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	Liaison	Familles ≥ 2 enfants atteints	Région candidate (2-10)	Localiser sur le génome des régions chromosomiques contenant un ou plusieurs gène(s) de susceptibilité	Etudier le lien entre la ressemblance phénotypique de deux apparentés et leur ressemblance génotypique. Ainsi, dans la méthode des paires de germains atteints, on teste 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	Pangénomique (400–10,000)			
		- En familles : Historiquement trios (enfant atteint et ses parents)	Région candidate (1-10)	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 population : Choix des témoins délicat Associations indirectes en raison du déséquilibre de liaison ou factices dues aux structures génétiques des populations
		GWAS (500,000-5M)	- En familles : Tester s’il existe chez les parents hétérozygotes une sur/sous-transmission d’un de leurs allèles aux enfants atteints.		- En familles : Génotypage nécessaire des parents Variants n’impactant pas la probabilité de survie foetale	
		Séquençage (30M-3G)				

2.2.1 Analyses ADN-indépendantes : corrélations familiales, études de jumeaux, études de ségrégation

De nombreuses études de jumeaux ont étudié le phénomène de polarisation en plus de la lèpre *per se*. Par exemple, une étude portant sur des couples de jumeaux dont au moins un individu était MB (classification de Madrid) a observé un taux de concordance de la forme MB très significativement supérieur chez les monozygotes (MZ) que chez les dizygotes (DZ) (70 % chez les MZ vs. 20 % chez les DZ) (Chakravarti and Vogel, 1973; Smith, 1979) ; A l'inverse, pour la forme PB, les taux de concordance entre jumeaux MZ et DZ étaient très proches. Comme nous l'avons évoqué plus haut, il existe une plus grande proportion de MB chez les garçons (Beiguelman et al., 1968). Ce biais contribue sans doute marginalement à l'augmentation du taux de concordance chez les jumeaux MZ vs. les DZ (qui peuvent être de sexe opposé). Enfin, plusieurs études de ségrégation familiale ont été réalisées en particulier aux Philippines, au Vietnam, sur l'Ile de la Désirade, en Thaïlande et au Brésil : elles sont en faveur d'une contribution génétique significative dans le contrôle par l'hôte de la polarisation de la lèpre, même si les modèles génétiques sont rarement concordants d'une étude à l'autre (Abel and Demenais, 1988; Feitosa et al., 1995; Smith, 1979; Wagener et al., 1988).

2.2.2 Etudes de liaison région-candidate puis pangénomiques

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. 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, *i.e.* en 5 classes. 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 (Siddiqui et al., 2001), 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* (Mira et al., 2003a). La liaison était moins forte pour le phénotype « lèpre *per se* », suggérant un contrôle génétique spécifique de la polarisation.

Les études sur un échantillon Brésilien ont rapporté une liaison de la lèpre *per se* avec les régions 17q11-q21 (Jamieson et al., 2004), 6p21, 17q22 et 20p13 (Miller et al., 2004). 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, cf figure 6bis), 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). Enfin, une étude de liaison au Malawi a identifié un signal significatif entre la polarisation (selon la classification OMS 88) et des marqueurs génétiques de la région 21q22 (Wallace et al., 2004). 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.

La première étude de liaison pangénomique (c'est-à-dire analysant l'ensemble du génome tandis que les études mentionnées précédemment se limitaient à une ou quelques régions particulières) de la lèpre (et non de la polarisation), 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 (Siddiqui et al., 2001). Une liaison significative ($p < 2 \cdot 10^{-5}$) a été observée avec la région 10p13. Toutefois, en raison de la très forte proportion de cas de PB (98 %; définition de Madrid), 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 (OMS-88) (Mira et al., 2003b). 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 (Tosh et al., 2002). Encore une fois, cette région a été retrouvée lors de l'étude du phénotype de polarisation mais pas de la lèpre *per se* dans l'étude vietnamienne précédentes (E. Schurr, données non publiées). 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.

2.2.3 Analyses d'association gène-candidat, région-candidate puis pangénomiques

De nombreuses études ont testé une association entre la lèpre *per se* et des variants génétiques localisés dans des gènes candidats par hypothèse, du fait de leur(s) fonction(s) connue(s). Elles ont principalement porté sur les gènes du complexe majeur d'histocompatibilité, de la famille des 'Toll like receptors' (TLR) et de la famille des cytokines (Misch et al., 2010; Gaschignard et al., 2013). Comme pour la liaison, seul un sous-ensemble de ces études s'est intéressé aux formes cliniques de la maladie (Tableau 2).

Cette stratégie d'études d'association avec un « gène-candidat » se heurte au problème classique des tests multiples (plusieurs gènes testés et plusieurs variants au sein de chacun de ces gènes). Faute d'une prise en compte adéquate, il entraîne une inflation de l'erreur de type 1. Pour s'en prémunir, les études de réplification sont absolument indispensables pour pouvoir envisager une association réellement significative. De fait, seules quelques associations ont été retrouvées dans plus d'un échantillon. Elles concernent des Single Nucleotide Polymorphisms (SNPs, qui sont des marqueurs bi-alléliques) au sein des gènes suivants : *TLR2* en Corée du Sud (Kang and Chae, 2001), en Ethiopie (Bochud et al., 2008) et au Malawi (Fitness et al., 2004) ; *TNF α* en Inde et en Thaïlande (Roy et al., 1997; Vejbaesya et al., 2007) ; Mannose Binding Lectin 2 (*MBL2*) au Brésil et au Népal (de Messias-Reason et al., 2007; Sapkota et al., 2010) ; le récepteur du mannose de type C 1 (*MRC1*) au Brésil et au Vietnam (Alter et al., 2010) ; et la séquence A du polypeptide lié au CMH classe I (*MICA*) en Chine et au Brésil (do Sacramento et al., 2012). Le gène *IL-10* a quant à lui été associé à la polarisation dans deux échantillons brésiliens indépendants (Moraes et al., 2004; Santos et al., 2002).

La plupart des associations n'ont été rapportées que dans une seule étude : *IFN γ* et *HLA-G* au Brésil (Reynard et al., 2003; Lucena-Silva et al., 2013), *TLR4* et la leucotriène A4

hydrolase (*LTA4H*) en Ethiopie (Bochud et al., 2009; Tobin et al., 2010), *VDR* (Roy et al., 1999), *KIR2DS3* (Franceschi et al., 2008) , la protéine de choc thermique 1A (*HSPA1A*) et *IL-23R* en Inde (Rajalingam et al., 2000; Ali et al., 2013), la lymphotoxine α (*LTA*) au Malawi (Fitness et al., 2004), la fraction 4 du complément (*C4*) en Thaïlande (Greiner et al., 1980), l'*IL-12B* au Mexique, l'*IL-12RB2* au Japon, la protéine de résistance naturelle associée aux macrophages 1 (*NRAMP1*) au Mali (Meisner et al., 2001), la laminine 2 (*LAMA2*) en Indonésie (Wibawa et al., 2002) et la défensine $\beta 1$ (*DEFB1*) au Mexique (Prado-Montes de Oca et al., 2009). Enfin une étude brésilienne illustre l'importance de la classification des formes de lèpre. Dans celle-ci, l'association des SNPs du gène codant pour l'IL-17F était significative pour la polarisation avec la classification de l'OMS-88 (88 patients PB et 52 MB, OR 3.2, $p < 0.05$), mais pas avec la classification de l'OMS-96 (53 individus PB et 87 MB) (Chaitanya et al., 2013). Ceci ne s'expliquait pas par un problème de puissance puisqu'elle était a posteriori (*i.e.*, sur la base des effets identifiés dans l'étude de la première classification) de 90 % dans cette seconde analyse, telle qu'estimée avec le logiciel QUANTO°, www.ypdra.usc.edu/gxe/.

Comme mentionné plus haut, la liaison de la région 10p13 avec la forme PB de la lèpre en Inde (Siddiqui et al., 2001) a été confirmée au Vietnam (Mira et al., 2003b). Afin de caractériser l'identité moléculaire du ou des gènes de la région responsable de ce signal, nous avons réalisé une étude d'association fine à l'aide de 1 522 marqueurs couvrant les 9 mégabases (Mb) de cette 'région-candidate' (Grant et al., 2014). L'échantillon primaire comprenait 294 familles nucléaires (parents et enfants) de la région de Hô-Chi-Minh ville incluant 303 enfants malades, dont 63 % présentaient une forme MB selon la définition OMS-96. L'échantillon de réplication issu de la même région regroupait 192 trios dont 55 % des enfants étaient MB. Trois SNPs localisés dans les gènes codant pour les protéines nébulette (*NEBL*) et cubiline (récepteur du facteur de cobalamine intrinsèque ou *CUBN*) et distants de 3 Mb, étaient spécifiquement et indépendamment associés avec la forme MB. Compte tenu du nombre moins élevé de cas PB (200 contre 295 MB), la puissance de détection était certes plus faible que pour le groupe MB, mais supérieure à 65 % pour les 3 SNPs identifiés (Grant et al., 2014).

Tableau 2. Liste de gènes associés avec une forme clinique de lèpre ou avec sa polarisation

Le tableau est divisé en deux parties. La partie supérieure regroupe les études qui ont analysé le phénotype « polarisation », la partie inférieure celles qui ont comparé les patients PB ou MB à des individus non-atteints. Les études sont triées par effectif décroissant. Les gènes identifiés par un GWAS sont en gras. Les études dont l'objectif principal a été d'identifier les gènes spécifiques de la polarisation ou d'un sous-type sont en rose.

Comparaison d'individus PB vs. MB						
Gène	Pays	Phénotype	Définition	N (MB/PB/Contrôles)	p	Marqueur (allèle, modèle génétique)
<i>MBL2</i> (Sapkota et al., 2010)	Népal	MB vs. PB	OMS-88	581/343/101	0.01	polymorphisme G161A (récessif)
<i>LTA4H</i> (Tobin et al., 2010)	Ethiopie	MB vs. PB	OMS-88	443/335/0	0.001	rs1978331 T/C et rs2660898 T/G
<i>LACC1</i> (Zhang et al., 2009a)	Chine	MB vs. PB	OMS-88	305/397/1225	<0.0001	rs3764147 et rs10507522 (allèle ? additif)
<i>LRRK2</i> (Zhang et al., 2009a)	Chine	MB vs. PB	OMS-88	305/397/1225	0.0008	rs1491938 (allèle ? additif)
<i>NOD2</i> (Zhang et al., 2009a)	Chine	MB vs. PB	OMS-88	305/397/1225	0.00004	rs9302752 (allèle ? additif)
<i>RIPK2</i> (Zhang et al., 2009a)	Chine	MB vs. PB	OMS-88	305/397/1225	0.0008	rs42490 (allèle ? additif)
<i>TLR4</i> (Bochud et al., 2009)	Ethiopie	MB vs. PB	OMS-88	298/138/197	0.05	polymorphisme T1196C (allèle C, dominant)
<i>NRAMP1</i> (Meisner et al., 2001)	Mali	MB vs. PB	OMS-88	181/92/201	0.003	insertion/délétion 3'-UTR 4-pb TGTG (hét vs. del/del)
<i>MBL2</i> (de Messias-Reason et al., 2007)	Brésil	MB vs. PB	Madrid	150/36/214	0.012	
<i>IL-10</i> (Moraes et al., 2004)	Brésil	MB vs. PB	OMS-96	166/131/283	0.006	Haplotype -375A/-2849G/-2763C
<i>VDR</i> (Roy et al., 1999)	Inde	MB vs. PB	Madrid	124/107/165	0.001	TaqI T -> C
<i>IL-10</i> (Santos et al., 2002)	Brésil	MB vs. PB	OMS-96	143/59/62	< 0.01	Haplotype -375A/-2849G/-2763C
<i>MICA</i> (do Sacramento et al., 2012)	Brésil	MB vs. PB	Madrid	153/45/201	0.01	MICA*010 et MICA*027
<i>IL-12RB2</i> (Ohyama et al., 2008)	Japon	MB vs. PB	Madrid	130/46/68	< 0.01	A-1035G (allèle G, allélique) A-1023G (allèle G, allélique) -650delG (allèle G, allélique) A-464G (allèle G, allélique) Haplotype -1035A/-1023A/-650G/-464A
<i>TLR2</i> (Fitness et al., 2004)	Malawi	MB vs. PB	OMS-96	26/184/379	0.04	microsatellite 224pb (intron 2) (modèle non spécifié)
<i>IL-17F</i> (Chaitanya et al., 2013)	Inde	MB vs. PB	OMS-88	88/52/84	< 0.05	T7488C (TT vs. TC)
<i>IL-10</i> (Franceschi et al., 2009)	Brésil	MB vs. PB	Madrid	65(LL)/43(TT)/240	< 0.01	haplotype -1082G/-819C/-592C

<i>LAMA2</i> (Wibawa et al., 2002)	Indonésie	MB vs. PB	OMS-88	27/26/58	< 0.005	mutation faux-sens (T7809C) substitution valine pour alanine (V2587A) (TC vs. TT)
<i>KIR2DS3</i> (Franceschi et al., 2008)	Brésil	LL vs. TT	RJ	65(LL)/42(TT)/289	0.04	
Comparaison d'individus PB ou MB vs. des individus non-atteints						
Gène	Pays	Phénotype	Définition	N (MB/PB/Contrôles)	p	Marqueur (allèle, modèle génétique)
<i>IL-23R</i> (Prado-Montes de Oca et al., 2009)	Inde	PB vs. CTL	OMS-88	416/427/1502	0.02	CNV - amplicon de 77 bases - situé dans l'exon 11
<i>MRC1</i> (Alter et al., 2010)	Brésil	MB vs. CTL	OMS-88	373/331/396	0.001	rs1926736 (G/A = G396S, allèle G récessif) rs2437257 (C/G = F407L, allèle C récessif) haplotype G396-F407
<i>CMCC122</i> (Grant et al., 2012)	Vietnam	MB vs. CTL	OMS-88	286/188	0.0064	rs3088362 (C/A, allèle A, additif)
<i>HLA-DR-DQ</i> (Grant et al., 2012)	Vietnam	MB vs. CTL	OMS-88	286/188	0.0003	rs602875 (A/G, allèle A, additif)
<i>LACC1</i> (Grant et al., 2012)	Vietnam	MB vs. CTL	OMS-88	286/188	0.00073	rs3764147 (A/G, allèle G, additif) rs10507522 (A/G, allèle A, additif)
<i>TLR2</i> (Bochud et al., 2008)	Ethiopie	MB vs. CTL	OMS-88	298/138/187	0.02	microsatellite 288pb (dominant)
<i>TNF</i> (Roy et al., 1997)	Inde	MB vs. CTL	Madrid	121/107/160	0.01	G308A (allèle A, allélique)
<i>RIPK2</i> (Marcinek et al., 2013)	Inde	PB vs. CTL	OMS-96	137/74/230	0.02	hapotype rs40457 G - rs42490 A
<i>LRRK2</i> (Marcinek et al., 2013)	Inde	PB vs. CTL	OMS-88	137/74/230	0.00001	rs1873613A/G (allèle A, allélique)
<i>LTA</i> (Fitness et al., 2004)	Malawi	MB vs. CTL	OMS-96	26/184/379	0.003	5' UTR microsatellite (AC/GT) _n à -3.5kb (allèle 101-pb)
<i>HLA-G</i> (Lucena-Silva et al., 2013)	Brésil	MB vs. CTL	OMS-88	76/70/128	0.018	polymorphisme + A3187G (allèle A, dominant)
<i>MICA</i> (Wang et al., 1999)	Chine	MB vs. CTL	X	50/19/112	< 0.05	MICA-A5 (allélique)
<i>IFNγ</i> (Reynard et al., 2003)	Brésil	PB vs. CTL	Madrid	59/10/98	0.01	CA repeat in intron 1 (< 122pb = "court", 122-126pb = "long") à risque = "long"
<i>DEFB1</i> (Prado-Montes de Oca et al., 2009)	Mexico	MB vs. CTL	Madrid	46/18/151	0.02	G668C (allèle C, dominant)
<i>TNFα</i> (Vejbaesya et al., 2007)	Thaïlande	MB vs. CTL	OMS-96	24/13/140	0.04	G308A (allèle A, modèle non spécifié)
<i>C4A</i> (Greiner et al., 1980)	Thaïlande	LL vs. CTL	RJ	71(LL)/27(TT)/201	< 0.01	allèle F1 ("fonctionnellement inactif")
<i>HSPA1A</i> (Rajalingam et al., 2000)	Inde	TT vs. CTL	RJ	0/49(TT)/38	0.03	allèle à risque HSP70-1A (allélique)

La première étude d'association pangénomique (GWAS pour Genome Wide Association Study) a été publiée en 2009 (Zhang et al., 2009a). Elle a inclus 706 cas et 1 225 contrôles chinois ; et environ 500 000 SNPs couvrant l'ensemble du génome ont été analysés. Les signaux les plus prometteurs ont été testés dans un échantillon dit de réplication incluant 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 (Zhang et al., 2011). Ces deux études ont identifié 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* (Nucleotide-binding Oligomerization Domain containing 2), *RAB32* (protéine Rab32) et *IL-23R* (récepteur de l'interleukine 23) (Zhang et al., 2011, 2009a). Une hétérogénéité entre les groupes PB et MB (OMS-88) 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 selon la classification OMS-88, les signaux localisés dans les gènes *HLA-DR-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* (Grant et al., 2012). Dans l'étude cas-témoin indienne à l'effectif relativement faible (211 cas dont 74 PB et 137 MB selon la classification OMS-88), seuls les signaux dans les gènes *LRRK2* et *RIPK2* ont été retrouvés ; l'effet observé était systématiquement plus significatif chez les patients PB (Marcinek et al., 2013).

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 (définition OMS-88) n'était pas significatif (Wong et al., 2010). 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 (voir la section suivante) 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*.

2.2.4 Méthodologie

La figure 15 montre la répartition des 82 études qui ont mis en évidence une association avec un phénotype de lèpre : lèpre *per se*, forme PB, MB ou polarisation (PB vs. MB). Parmi celles-ci figurent les 36 études du Tableau 2 qui ont mis en évidence une association avec un sous-type de lèpre ou avec sa polarisation. Pour 4 études, le résultat primaire était l'association avec la polarisation ou la forme PB seule (quadrants roses) et non la lèpre *per se*. Quatre autres études n'identifiaient pas d'association avec la lèpre *per se* (phénotype primaire), mais avec l'un de ses sous-types (quadrants bleus). Enfin, pour 28 études, une association était mise en évidence soit avec un sous-type, soit avec la polarisation, secondairement à une association avec la lèpre *per se* (quadrants verts). Par ailleurs, 46 études ont révélé une association avec la seule lèpre *per se* (quadrant gris). Au total, il y a donc 74 (46 + 28) études qui ont montré une association significative avec la lèpre *per se*, soit le double du nombre rapporté pour la polarisation ou l'un de ses sous-type. Plusieurs scénari peuvent expliquer cette différence.

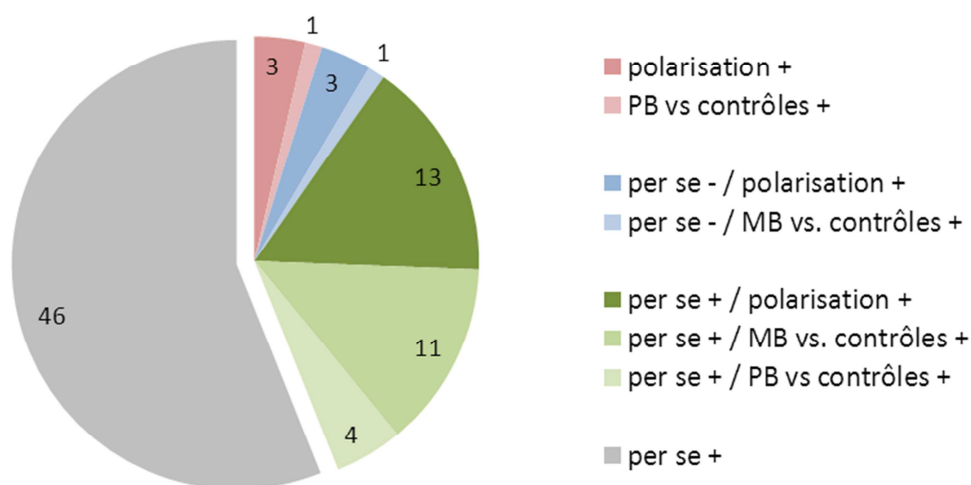


Figure 15. Répartition des études génétiques selon le phénotype de lèpre associé

Les signes + et – désignent la présence ou l'absence d'association avec le phénotype considéré, soit lèpre *per se*, forme MB ou PB, ou polarisation

Premièrement, la lèpre *per se* était le phénotype principal d'intérêt dans la majorité des études : aussi le nombre de patients pour chaque sous-type était plus petit, et la puissance pour détecter les variants génétiques influençant les sous-types était plus faible. Notons qu'une grande partie des associations significatives avec un sous-type concernait la forme MB, forme majoritaire dans la plupart des pays, à l'exception de l'Inde.

Deuxièmement, le schéma de classification des formes PB et MB peut aussi diminuer la puissance, comme le montre l'étude de l'*IL-17F* (Chaitanya et al., 2013). Troisièmement, la plupart des études de génétique de la polarisation de la lèpre ne tiennent pas compte des facteurs de risque « non génétiques ». Des études de la lèpre *per se* ont illustré la contribution importante de ces facteurs de risque aux résultats d'association : par exemple, l'association de la lèpre *per se* avec *PARK2* ou *LTA* était 3-5 fois plus forte chez les jeunes patients (Alcaïs et al., 2007; Alter et al., 2012). Plus généralement cet effet différentiel en fonction de l'âge semble être l'élément le plus informatif quant à la conduite de sous analyses.

Schématiquement, la plupart des études recherchent d'abord une association en comparant des cas de lèpre à des contrôles. Une valeur de p ($p_{\text{per se}}$) est calculée pour un SNP et sous un modèle génétique donné. Ensuite, des tests supplémentaires sont menés dans les sous-groupes en comparant les cas MB puis les cas PB à un même groupe contrôle indemne de lèpre. Des valeurs p_{MB} et p_{PB} sont calculées. Si $p_{\text{MB}} < p_{\text{per se}} < p_{\text{PB}}$, on pourrait conclure que ce SNP est associé à la lèpre et possède un effet majeur chez les patients MB (si $p_{\text{MB}} < p_{\text{per se}} < \text{seuil}$) ou que ce SNP est associé à la lèpre MB (si $p_{\text{MB}} < \text{seuil} < p_{\text{per se}}$). Cette stratégie générale présente plusieurs limites. Tout d'abord, pour conclure à un effet plus fort d'un sous-type de lèpre, l'odds-ratio ou le risque relatif selon le design de l'étude semble plus informatif que la valeur de p . Nous montrons dans la figure 16 une meilleure stratégie lorsque le but est d'identifier les variantes qui distinguent spécifiquement les cas PB des cas MB. Elle consiste, après avoir affiné le phénotype et équilibré le nombre de cas dans chaque sous-groupe, à tester le phénotype polarisation.

Par ailleurs, l'étude directe de la polarisation présente l'avantage d'une interprétation univoque (si l'on met de côté les interprétations relatives à l'erreur de type I, d'une stratification de population, etc...). Comme illustré dans la problématique (Figure 13), l'identification d'une association avec un sous-type de lèpre en comparant des cas et des contrôles (par définition non

malades) ne permet pas de savoir si le variant agit dans le passage de l'exposition à l'infection, de l'infection à la maladie, ou de la maladie à sa forme clinique.

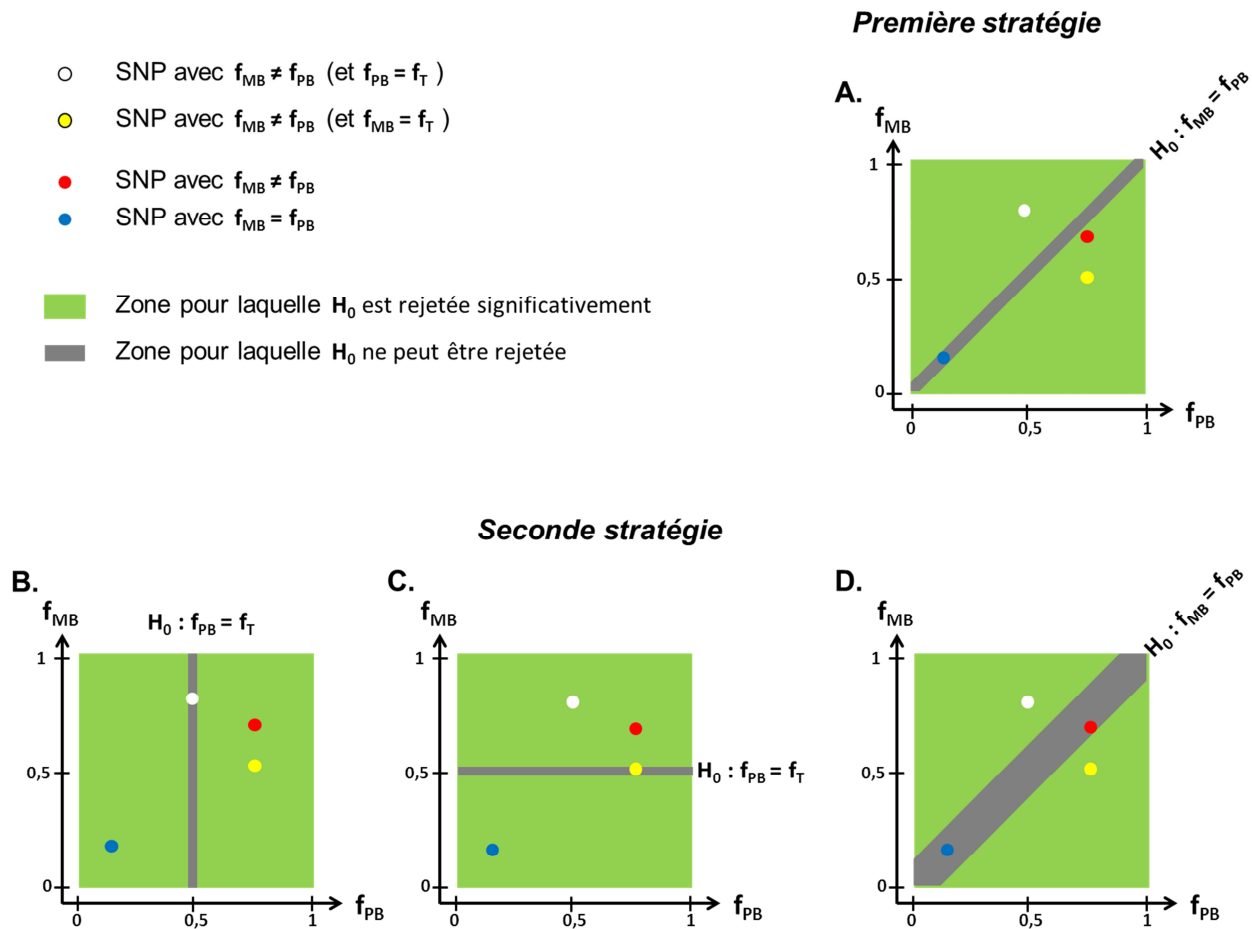


Figure 16. Stratégies d'étude de la polarisation

Dans une étude cas-témoin, un test standard pour l'association entre un variant génétique et le phénotype est de comparer la fréquence des allèles (f) entre un groupe de cas et un groupe de contrôles. Nous considérons ici un échantillon comprenant autant de contrôles que de cas de lèpre, équi-répartis entre PB et MB. La fréquence de l'allèle d'intérêt a été fixée à 0,5 chez les témoins ($f_T = 0,5$). Dans chaque panneau, f_{PB} et f_{MB} (axe des x et des y) sont les fréquences de l'allèle chez les personnes PB et MB.

La première stratégie compare directement les groupes MB et PB (Panneau A).

La seconde stratégie teste l'opposition entre les groupes PB et contrôles (Panneau B), celle entre les groupes MB et contrôles (Panneau C) avant de réaliser un test d'hétérogénéité PB vs. MB (panneau D). Le SNP rouge qui correspond à un variant pour lequel $f_{MB} \neq f_{PB}$. L'association est significative avec la première stratégie, pas avec la seconde (SNP rouge dans la zone grise où H_0 ne peut être rejetée). Cette seconde stratégie est en effet entravée par la correction pour les tests multiples (test MB vs. PB après avoir testé MB vs. contrôles et PB vs. contrôles) et par la puissance souvent faible des tests d'hétérogénéité.

Sujets et matériel génétique

3.1 Recueil des sujets

Depuis la fin des années 80, notre laboratoire a noué une collaboration étroite avec l'Hôpital de Dermato-Vénérologie de Ho Chi Minh. Il y a eu plusieurs phases de recrutement entre 1991 et 2008. Différentes études ont été menées, essentiellement de génétique épidémiologique (ségrégation, liaison et association), mais aussi d'immunologie fonctionnelle. Pour chacune, les consentements éthiques ont été obtenus auprès des autorités de Santé à Ho Chi Minh, et auprès du Research Institute of the McGill University Health Centre de Montréal au Canada.

Les caractéristiques des échantillons sont résumées dans le tableau 3. Pour les projets s'intéressant aux familles, le mode de recrutement a toujours été centré sur l'enfant : 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 (Ranque et al., 2005, 2007b). Le second a ciblé des individus qui présentaient une réaction de réversion ou qui étaient susceptibles d'en présenter une, essentiellement des individus classés « borderline » (BT, BB et BL selon Ridley et Jopling (Ranque et al., 2007a). Au total, 2 784 individus ont été recrutés en 18 ans, dont 1 182 cas de lèpre.

Un formulaire était rempli pour chaque individu. Les informations suivantes étaient renseignées : patronyme, année de naissance, nom des parents, 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, présence d'une réaction de type 1 (réaction de réversion) ou de type 2 (érythème noueux) (Figure 17). L'échantillon qui a servi pour notre thèse est extrait de l'ensemble de ces projets.

Tableau 3. Projets du laboratoire au Vietnam entre 1991 et 2008

CNG : nombre d'échantillons envoyés au Centre National de Génotypage. Les génotypes validés correspondent, pour chaque projet au nombre de génotypes de bonne qualité qui ont été conservés, une fois les duplicats et triplicats retirés. La proportion de MB (selon la classification OMS-82), de réaction de réversion (RR) et d'érythème noueux lépreux (ENL) varie selon le critère de recrutement du projet.

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	1056	921	833	465	59	23	55	22	7
RR	2002-2004	Réaction de réversion	Trios	297	284	251	97	57	21	78	96	4
zRR	2005	Réaction de réversion	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
MITS	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	Réaction de réversion	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	Réaction de réversion	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

CẨM ỨNG DI TRUYỀN ĐÒI VỚI BỆNH PHONG TÂY VIỆT NAM

Trắc nghiệm cá nhân
BỘ * C



Code famille

Số gia đình : HCM 111

Nom et prénom

Họ tên người được nghiên cứu : [Redacted]

Ethnie

1.2 Dân tộc : Việt Nam ☐ Hoa ☒ Khác ☐ Ghi rõ :

Lieu de naissance

1.3 Nơi sinh (thành phố, vùng) : [Redacted]

Adresse

1.4 Thường trú (thành phố, vùng) : [Redacted]

Sexe

1.5 Giới : Nam ☒ Nữ ☐

Année de naissance

1.6 Ngày, tháng, năm sinh (hay tuổi) : 1941

2. Chỉ dành cho bệnh nhân :

Identité père

2.1 Họ (chữ thường), tên (chữ hoa) của mẹ : [Redacted]

Identité mère

2.2 Họ (chữ thường), tên (chữ hoa) của cha : [Redacted]

Statut BCG

2.3 Sẹo BCG : Có ☐ Không ☒ Không rõ ☐

Date du diagnostic

2.4 Năm chẩn đoán bệnh : 1964

Classification clinicien

2.5 Thể bệnh :

Ít khuẩn (PB) ☐

Nhiều khuẩn (MB) ☒

Nombre de lésions

Số lượng thương tổn da : 720

I ☐

T ☐

B ☐

L ☒

Classification RJ

I ☐

TT ☐

BT ☐

BB ☐

BL ☐

LL ☒

Indice bacillaire

2.6 Chỉ số vi trùng (BI) lúc chẩn đoán : 2.4

2.7 Phản ứng :

Thời gian khởi phát

Trước

Trong

Sau (ĐT)

Réaction de type 1

Loại I : Phản ứng đảo nghịch : Không ☒

Có ☐

Nếu có

☐

☐

☐

Réaction de type 2

Loại II : Phản ứng HBN : Không ☐

Có ☒

Nếu có

☐

☒

☐

2.8 Tình trạng bệnh nhân :

Traitement

DHTL ☐

Giám sát ☐

Ngừng giám sát ☒

Figure 17. Exemple de fiche individuelle

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

Avant de le décrire, nous le situons dans 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 (WHO, 2008) qui correspond à la période de recrutement (1991-2008).

3.2 La lèpre au Vietnam

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. Jusqu'en 1998, les cas de lèpre MB étaient traités pendant 24 mois ou jusqu'à ce que l'IB devienne négatif. 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. 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. A partir de 1998, la réduction de la durée de traitement des sujets MB (de 24 mois ou plus à 12 mois, quel que soit l'IB en fin de traitement) n'a pas interrompu la diminution de la prévalence.

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 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.

La proportion des cas MB parmi les nouveaux cas de lèpre a augmenté très significativement entre le début des années 80 et 2006 passant de 30-40 % à ~ 67 %. Cette augmentation peut s'expliquer en partie par la modification des critères de classification. Malgré cette augmentation en terme de proportion, le nombre absolu annuel de nouveaux cas MB a lui diminué depuis 1992 (2 062 en 1992, 1 189 en 1998 et 443 en 2006). Enfin, si la proportion de cas présentant une incapacité de grade 2 au diagnostic (définie par une déformation visible ou une acuité visuelle $< 1/10$) 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 %) (WHO, 2013).

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. Notre visite sur le terrain à Ho Chi Minh et à Kon Tum, ville située dans la région des Hauts-Plateaux au centre du Vietnam, nous l'a confirmé. En 10 jours de consultation à l'hôpital de dermatologie-vénéréologie de Ho Chi Minh, nous n'avons diagnostiqué qu'un seul nouveau cas de lèpre chez un adolescent. En revanche, en moyenne 6 patients étaient hospitalisés pour des complications de la lèpre (e.g. infection de plaie...). Par ailleurs, 4 à 10 personnes consultaient chaque jour pour des soins ambulatoires de plaie, une chirurgie correctrice ambulatoire, une confection de chaussures adaptées... Près de Kon Tum, nous avons visité un « village de lépreux » et suivi une religieuse dans ses visites auprès de populations indigènes. Nous avons pu observer qu'en dépit de soins médicaux de qualité et d'aides sociales importantes offerts par l'Etat aux individus atteints de lèpre, des populations rurales pauvres restaient 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.

3.3 Matériel génétique

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). 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 (<http://hapmap.ncbi.nlm.nih.gov/index.html.en>). 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. A ce jour, environ 30 millions de variations ont été décrites chez 1 092 individus issus de 21 populations différentes (1000 Genomes Project Consortium et al., 2012). La puce couvre 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 une population vietnamienne, elle est *a priori* élevée. En effet, la couverture est, avec un $r^2 > 0,8$, de 86% pour les variants ($MAF \geq 5\%$) des populations japonaise et chinoise référencées dans HapMap et 1000 Génomes. Ces pays sont proches géographiquement du Vietnam, et un lien entre distance géographique et génétique a été établi, en particulier pour les populations européennes (Novembre et al., 2008). La médiane d'espacement entre 2 SNPs consécutifs sur la puce est de 2,3 kilobases (kb) (tableau 4). 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 par ailleurs un grand nombre de Copy Number Variants (CNV), polymorphismes non bi-alléliques comme les SNPs. Nous ne les avons pas inclus dans nos analyses.

Tableau 4. Principales caractéristiques de la Puce Illumina Human660W-Quad d'après les données du fabricant. CMH = Complexe Majeur d'Histocompatibilité

Illumina Human660W-Quad	
SNPs (N)	657 366
Couverture génomique (CHN-JAP)	86% $r^2 > 0.8$
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)	3 177/8 440/0
SNPs sur gonosomes (X/Y/PAR)	16 509/44/15
SNPs mitochondriaux	135

3.4 Constitution de la base de données

L'objectif principal du projet était de réaliser une association génome-entier en utilisant des données familiales. Compte-tenu du budget alloué, il a été possible d'inclure la totalité des familles recrutées entre 1991 et 2008.

Parmi les 2 784 individus de la base de données, l'ADN était disponible pour 2 564. Le Centre National de Génotypage (CNG) d'Evry 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 que de nombreux individus figuraient en deux (N = 137) ou trois (N = 18) exemplaires. La principale raison (avant les erreurs techniques - étiquetage, pipetage...) était que certains individus avaient réellement été génotypés plusieurs fois... car recrutés plusieurs fois. 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 !

Nous avons utilisé la matrice « Identity By State » (IBS) pour identifier les duplicats/triplicats, ainsi que pour redéfinir certains liens familiaux. Chaque case de la matrice correspond à la proportion d'allèles identiques entre 2 individus. Ainsi, une valeur d'IBS > 0,99

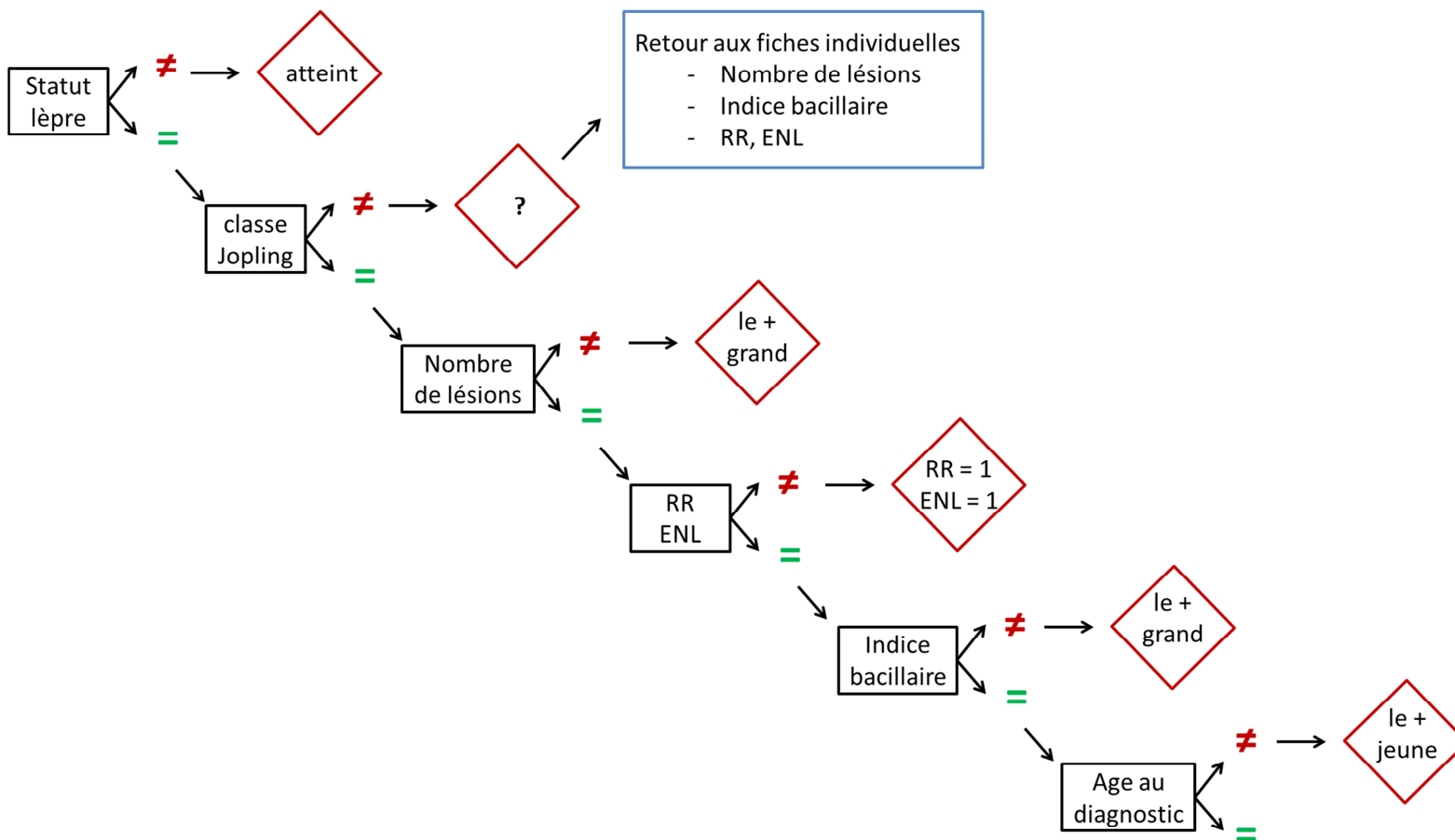


Figure 18. Arbre décisionnel d'attribution du phénotype pour des duplicats discordants

Chaque critère est comparé successivement : en cas de discordance, la décision figure dans le carré rouge. RR : réaction de réversion, ENL = Erythème noueux lépreux

entre deux individus signait une identité. Nous avons ensuite calculé la proportion d'allèles partagés « Identical By Descent » (IBD), *i.e* d'allèles provenant d'un même ancêtre, avec le logiciel PLINK (<http://pngu.mgh.harvard.edu/~purcell/plink/>) (Purcell et al., 2007). 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.

Le second problème posé par les duplicats/triplicats 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 pour nous de « fixer » le phénotype de l'individu pour nos analyses ultérieures. Nos collaborateurs Vietnamiens ont alors effectué un minutieux travail d'analyse rétrospective des fiches individuelles, conservées à l'Hôpital de dermato-vénérologie 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.

Nous avons globalement suivi l'algorithme présenté dans la figure 18. Seules les discordances de classification selon Ridley-Jopling ou selon l'OMS-96 ont posé problème. En effet, ces classifications influencent directement notre phénotype d'intérêt. A titre d'exemple, comment considérer un individu classé BB en 2005 puis BL en 2007? Si la polarisation de la lèpre est un phénomène dynamique, les passages entre certaines classes sont moins fréquents que d'autres, en particulier un passage BB vers BT (expertise de nos confrères vietnamiens). Comme nos études ont d'abord concerné le phénotype binaire ($PB = TT + BT$ vs. $MB = BB + BL + LL$), les cas de discordance autour de cette frontière nous ont particulièrement soucieux : les données du nombre de lésions, de l'IB et de la présence ou non de réactions lépreuses de type 1 ou 2 nous ont permis d'attribuer une classe selon Ridley et Jopling, et à défaut la valeur « Non Attribué » (NA).

Après avoir retiré les duplicats/triplicats, notre échantillon comportait 2 235 individus pour lesquels le génotypage avait été un succès pour plus de 95 % des SNPs de la puce. Les individus avec un succès moindre avaient possiblement une qualité d'ADN moindre et les

données pour les SNPs génotypés risquaient d’être erronés. Ces individus appartenaient à 680 « famille ». Nous distinguons plusieurs types de familles : les individus isolés ou singletons (non apparentés), les familles nucléaires (comprenant un ou deux parents génotypés), les individus non liés (cousins ou enfants probables d’un même couple non génotypé), 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. 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 (Tableau 5).

Tableau 5. Répartition des 2 235 individus génotypés selon le type de famille

<i>Description</i>		N	N (≥ 1enfant PB ou MB)
Singletons		89	0
Familles nucléaires	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 enfant	26	26
	2 enfants	11	10
1 parent	3 enfants	2	2
	4 enfants	2	2
	12 enfants	1	1
Individus non liés		16	0
Familles complexes		14	14
Total		680	569

3.5 Description des sujets

Les 1 205 fondateurs dont l’ethnie était déclarée étaient majoritairement *Kinh* (N = 1 158). Quelques-uns provenaient des minorités *Cham* (11), *Han* (22), *Raglay* (10) et *Nung ou Tay* (4), et deux étaient d’origine mixte. 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 19). 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 mariage mixtes sino-vietnamiens, plus que parmi les Chinois de Beijing.

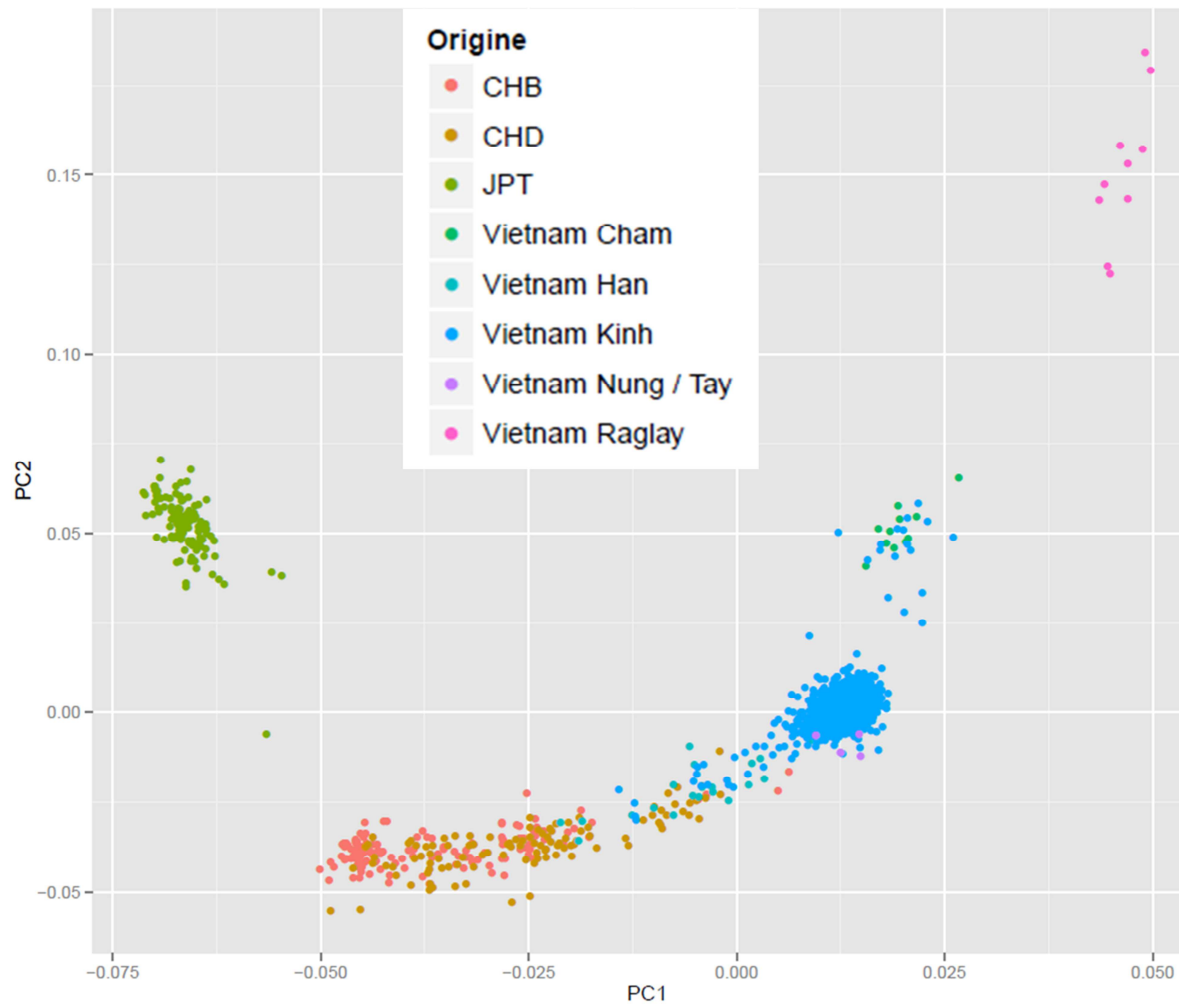


Figure 19. Distribution des fondateurs de l'échantillon par origine ethnique selon la première (axe des x) et la deuxième (axe des y) composantes principales d'analyse en incluant les populations de Hapmap les plus proches génétiquement (CHB = Chinois Han de Beijing, CHD = Chinois de Denver – Etats-Unis, JPT = Japonais)

Tableau 6. Caractéristiques de l'échantillon

	Fondateurs	Enfants	Total	p
	N (%)	N (%)	(%)	PB vs. MB
Total	1238	997	2 235	
Hommes	634	651	1 285	
Femmes	604 (49)	346 (35)	950 (43)	$< 10^{-6}$
Cas de lèpre	247	692	939	
Age moyen au diagnostic	37,3	18,3	55,6	$< 10^{-6}$
Répartition par groupe d'âge				$< 10^{-6}$
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)	$2 \cdot 10^{-5}$
PB (OMS 96)	80 (45)	385 (63)	465 (58)	
MB (OMS 96)	98 (55)	229 (37)	327 (42)	$3 \cdot 10^{-5}$
PB (OMS 82)	64 (30)	315 (46)	379 (41)	
MB (OMS 82)	152 (70)	363 (54)	515 (59)	$2 \cdot 10^{-5}$
Classification de Ridley et Jopling				$< 10^{-6}$
TT	30 (13)	123 (18)	153 (17)	
BT	34 (15)	192 (28)	226 (25)	
BB	70 (30)	193 (29)	263 (29)	
BL	59 (26)	125 (18)	184 (20)	
LL	38 (16)	45 (7)	83 (9)	
Réaction	58 (23)	250 (36)	308 (33)	$4 \cdot 10^{-4}$
Type 1 (RR)	49 (20)	232 (34)	281 (30)	$8 \cdot 10^{-5}$
Type 2 (ENL)	12 (5)	26 (4)	38 (4)	0,57

Les principales caractéristiques de l'échantillon sont résumées dans le tableau 6. Les caractéristiques des fondateurs différaient radicalement des enfants. Ceci était attendu compte-tenu du mode de recrutement des familles, principalement à partir d'un ou plusieurs enfants atteints. La polarisation en tant que trait binaire (PB/MB) pouvait être définie selon différentes classifications de la lèpre (figure 6) : OMS-82, OMS-88, OMS-96, ainsi que la classification retenue par le clinicien, qui apparaissait sur les fiches de recueil de données de chaque individu (figure 17). Les classifications OMS-82 et OMS-88 ne diffèrent que par le niveau d'IB retenu comme seuil. A 2 patients près, elles coïncidaient dans notre échantillon. Nous avons conservé la classification OMS-82 qui se superposait exactement à celle de Ridley et Jopling, en regroupant les patients TT et BT dans le groupe PB, et les patients BB, BL et LL dans le groupe MB (figure 6bis).

La proportion de cas MB, calculée à partir des données du tableau 6, variait considérablement selon la classification utilisée, de 41,2 % pour OMS-96 à 57,6 % pour OMS-82 et 63,1 % pour le clinicien. Un diagramme de Venn permet de visualiser les recouvrements entre classification (Figure 20) : il utilise les 733 malades pour lesquels nous disposons des 3 classifications. Nous avons quantifié la concordance des trois classifications par le calcul du coefficient Kappa, implémenté dans le logiciel R (Revelle, 2014). Il était plus élevé entre les classifications OMS-82 et clinicien (0,86 – intervalle de confiance à 95 % [0,82; 0,89]), qu’entre OMS-82 et OMS-96 (0,6 [0,55; 0,65]), ou entre OMS-96 et clinicien (0,61 [0,56; 0,66]). Une interprétation courante du coefficient Kappa est une corrélation faible [0; 0,4], moyenne [0,4; 0,6], bonne [0,6; 0,8] et excellente [0,8; 1]. Cette minoration de la proportion de cas MB par la classification OMS-96 par rapport à celle de l’OMS-82 confirme une observation de Pardillo *et al* aux Philippines, autre pays de haute endémicité des cas MB (78 % dans leur étude) (Pardillo et al., 2007). Le contraire est attendu dans les pays où la forme MB est moins prévalente (Aubry).

La proportion de patients MB selon la classification la plus récente de l’OMS (1996) était plus faible que celle observée au Vietnam entre 1983 et 2006, période proche de celle de notre recrutement (WHO, 2008). Toutefois, notre recrutement, centré sur le phénotype pour certains projets (tableau 3), ne cherchait pas à refléter la proportion de cas MB observée au Vietnam. Il est difficile de comparer cette proportion avec celles des études brésilienne, indienne ou birmane citées dans le paragraphe « connaissances de l’épidémiologie des formes de lèpre » car elles ont utilisé la classification de Madrid, non rapportée au Vietnam (Guerra-Silveira and Abad-Franch, 2013; Mukherjee et al., 2013; Rao et al., 1972).

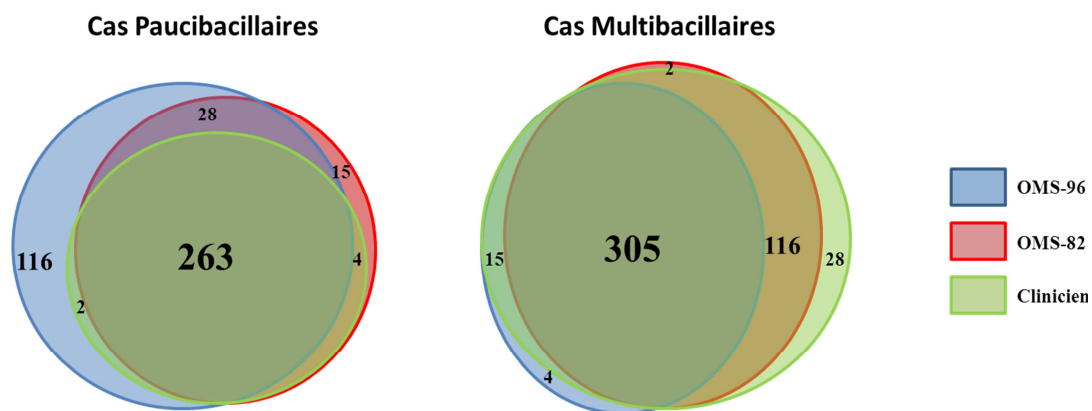


Figure 20. Diagramme de Venn des cas PB et MB selon 3 classifications
Figure réalisée à partir des 733 cas de lèpre classés selon les 3 méthodes.

Tableau 7. Caractéristiques des 939 cas de lèpre parmi les enfants et les fondateurs

Forme clinique de lèpre (classification OMS 82)	Enfants			Fondateurs		
	PB N (%)	MB N (%)	p PB vs. MB	PB N (%)	MB N (%)	p PB vs. MB
Total	315	363		64	152	
Hommes	227	249		47	113	
Femmes	88 (28)	114 (31)	0,37	17 (27)	39 (26)	0,87
Age moyen au diagnostic	18,4	18,5	0,80	39,3	36,8	0,03
Répartition par groupe d'âge			0,91			0,41
4-9	24 (8)	29 (8)		1 (2)	1 (0)	
10-19	169 (54)	190 (52)		8 (12)	12 (8)	
20-59	120 (38)	144 (40)		49 (77)	127 (85)	
> 60	0 (0)	0		6 (9)	10 (7)	
Nombre moyen de lésions	2,3	9,1	$< 10^{-6}$	2,2	11,1	$< 10^{-6}$
> 5 lésions	17 (5)	212 (58)	$< 10^{-6}$	2 (3)	94 (62)	$< 10^{-6}$
IB > 0	5 (2)	135 (38)	$< 10^{-6}$	3 (5)	76 (50)	$< 10^{-6}$
Réaction	58 (18)	181 (50)	$< 10^{-6}$	5 (8)	50 (33)	$6 \cdot 10^{-5}$
Type 1 (RR)	58 (18)	173 (48)	$< 10^{-6}$	5 (8)	41 (27)	$2 \cdot 10^{-3}$
Type 2 (ENL)	0 (0)	26 (7)	$< 10^{-6}$	0 (0)	12 (8)	0,02

Aucun effet significatif du sexe sur la proportion de cas MB n'a été observé, quelle que soit la classification utilisée. Cette proportion était, chez les hommes et chez les femmes, de 42 et 41 % (OMS-96), 58 et 57 % (OMS-82), 63 et 62 % (clinicien). Le mode de recrutement, non orienté par le sexe, ne devrait pas influencer ce résultat. La proportion de cas MB augmentait avec l'âge, et la corrélation était plus forte pour les hommes que pour les femmes. L'influence de l'âge était toutefois moindre qu'au Brésil (Figure 21 comparée à la figure 14). L'évaluation dynamique de l'effet du sexe par l'analyse multivariée, c'est à dire prenant en compte l'âge, ne modifiait pas les conclusions précédentes.

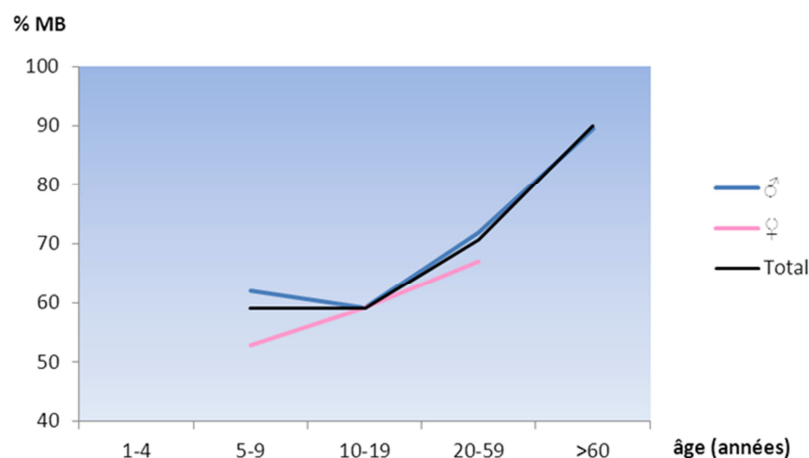


Figure 21. Proportion de cas MB selon l'âge et le sexe dans notre échantillon vietnamien

3.6 Choix du phénotype

Nous avons donc la possibilité de définir la polarisation comme :

1. Un trait binaire (MB vs. PB) selon les classifications dont nous disposions :
 - OMS-82 (équivalente dans notre échantillon à celle de Ridley-Jopling en définissant le groupe PB = TT + BT et le groupe MB = BB + BL + LL, cf figure 6bis) : 678 enfants classés, 315 PB et 363 MB.
 - OMS-88 : 678 enfants classés, 298 PB et 380 MB.
 - OMS-96 : 614 enfants classés, 385 PB et 229 MB.
 - Clinicen : 641 enfants classés, 263 PB et 378 MB.
2. Un trait catégoriel ordonné (TT < BT < BB < BL < LL) : 123 enfants TT, 192 BT, 193 BB, 125 BL et 45 LL.
3. Un trait quantitatif, en attribuant des valeurs aux différentes classes de Ridley-Jopling.

Pour le trait binaire, la classification OMS-82 nous semble le mieux séparer les deux groupes PB et MB sur le plan physiopathologique. Choisir le meilleur classement nous permettait d'espérer à la fois une meilleure puissance, mais aussi un nombre limité de faux positifs. De plus, la classification OMS-82 était celle qui était disponible pour le plus grand nombre d'enfants ($N = 678$), et qui équilibrait le mieux les deux groupes PB et MB. Ceci nous plaçait dans les conditions optimales de puissance statistique.

Nos analyses de liaison et d'association se sont donc appuyées sur cette classification. Nous avons étudié la polarisation comme trait catégoriel ordonné en liaison, mais pas en association. Notre thèse n'aborde pas l'analyse du trait quantitatif.

Analyse de liaison

4.1 Méthode du maximum de vraisemblance binomial

4.1.1 Cas binaire

La méthode MLB (pour *Maximum Likelihood Binomial*) est une méthode d'analyse de liaison modèle indépendante orientée fratrie. Elle permet d'analyser des fratries de grande taille sans recourir à un découpage des fratries de taille > 2 en paires de germains. La méthode repose sur l'utilisation d'une loi binomiale pour modéliser la distribution du nombre d'enfants atteints qui ont reçu un allèle donné, par exemple A , d'un parent hétérozygote Aa (Abel et al., 1998).

Pour une famille nucléaire donnée i et un parent j hétérozygote Aa , notons S_i le nombre d'enfants atteints et n_{Aij} le nombre d'enfants atteints ayant reçu l'allèle A . L'expression de la vraisemblance repose sur l'idée d'une distribution binomiale des allèles parentaux parmi les enfants. Sous l'hypothèse nulle H_0 d'absence de liaison entre le marqueur étudié et le locus influençant le phénotype, les probabilités de transmettre A ou a à un enfant atteint sont égales à 0,5. La vraisemblance de l'échantillon peut s'écrire comme le produit sur i (les N familles) et j (les deux parents) de distributions binomiales $(S_i, 0,5)$. Un test de liaison génétique simple peut donc être construit en testant l'écart par rapport à 0,5 du paramètre de probabilité de ces distributions binomiales. Si on note α ce paramètre, la vraisemblance de l'échantillon, $L(\alpha)$, s'écrit :

$$L(\alpha) = \prod_{i=1}^N \prod_{j=1}^2 L_{ij}(\alpha)$$

où $L_{ij}(\alpha)$ est la vraisemblance d'une fratrie dans laquelle la distribution du nombre d'enfants atteints ayant reçu l'allèle A d'un parent hétérozygote Aa est une binomiale (S_i, α) . Dans le cas habituel où la phase est inconnue, on ne sait pas d'une famille à l'autre quel allèle a la probabilité α et quel autre a la probabilité $1-\alpha$ d'être reçu par un enfant atteint. Chacune des configurations est *a priori* équiprobable. En conséquence, $L_{ij}(\alpha)$ est la somme de la moitié de la vraisemblance de la fratrie sachant qu'un allèle donné, disons A , a été reçu avec une probabilité

α , et de la moitié de la vraisemblance de la fratrie en considérant que A a été reçu avec une probabilité $1-\alpha$, soit :

$$\begin{aligned} L_{ij}(\alpha) &= 0,5 \binom{S_i}{n_{A_{ij}}} \left[\alpha^{n_{A_{ij}}} (1-\alpha)^{S_i - n_{A_{ij}}} \right] + 0,5 \binom{S_i}{n_{A_{ij}}} \left[\alpha^{S_i - n_{A_{ij}}} (1-\alpha)^{n_{A_{ij}}} \right] \\ &= 0,5 \binom{S_i}{n_{A_{ij}}} \left[\alpha^{n_{A_{ij}}} (1-\alpha)^{S_i - n_{A_{ij}}} + \alpha^{S_i - n_{A_{ij}}} (1-\alpha)^{n_{A_{ij}}} \right] = C_{ij} f_{ij}(\alpha) \end{aligned}$$

où $f_{ij}(\alpha)$ est la vraisemblance de travail dépendant de α .

L'expression de la vraisemblance est totalement symétrique en α et l'hypothèse alternative peut être définie comme $\alpha < 0,5$ (et $1-\alpha > 0,5$) ou $\alpha > 0,5$ (et $1-\alpha < 0,5$), en fonction de l'interprétation de α . Dans le premier cas, α correspond à la probabilité pour un enfant atteint d'avoir reçu l'allèle marqueur non transmis avec l'allèle de susceptibilité. Dans le second cas, c'est la probabilité d'avoir reçu l'allèle marqueur transmis avec l'allèle de susceptibilité. Dans les deux cas, le test est unilatéral une fois que l'interprétation de α a été déterminée. Dans la suite de ce chapitre, l'hypothèse alternative sera définie par $\alpha > 0,5$. Soit α_0 l'estimateur du maximum de vraisemblance de α . La statistique du test de liaison génétique est un rapport de vraisemblance λ_{MLB} défini comme :

$$\lambda_{MLB} = 2 \text{Ln} \frac{L(\alpha = \alpha_0)}{L(\alpha = 0,5)} = 2 \text{Ln} \frac{\prod_{ij} f_{ij}(\alpha = \alpha_0)}{\prod_{ij} f_{ij}(\alpha = 0,5)}$$

Sous H_0 , λ_{MLB} est asymptotiquement distribué comme un mélange 50 % : 50 % de χ^2 à 0 et 1 degré de liberté. La raison du point de masse en zéro est que le test est unilatéral, c'est-à-dire que α est borné à 0,5 quand le maximum non restreint est $< 0,5$, et la probabilité que cette situation survienne est de 0,5 sous H_0 . λ_{MLB} peut également être exprimé sous la forme d'un écart réduit standardisé normal unilatéral $Z_{MLB} = \sqrt{\lambda_{MLB}}$. La statistique de test peut aussi s'exprimer comme un LOD-score [égal à $\lambda_{MLB}/2\text{Ln}(10)$] qui suit sous H_0 la même distribution qu'un LOD-score « classique » basé sur l'estimation du taux de recombinaison. Dans le cas de données parentales manquantes, la vraisemblance peut être calculée en sommant sur tous les génotypes parentaux possibles au locus marqueur. Chaque terme de la somme est pondéré par la probabilité

du génotype parental correspondant, obtenu à partir des fréquences alléliques du marqueur dans la population. Enfin, les parents homozygotes ne contribuent pas au calcul de la vraisemblance, i.e. $f_{ij}(\alpha) = 1$ quand le parent j de la famille i est homozygote au locus marqueur.

4.1.2 Extension aux phénotypes quantitatifs : méthode nMLB-QTL

Nous exposons ici la méthode du new MLB Quantitative Trait Loci (nMLB-QTL) proposée par Aurélie Cobat. Cette approche est une variation du MLB-QTL (Alcais et al., 1999), qui, tout en conservant la robustesse du MLB-QTL original quant aux hypothèses relatives à la taille de la fratrie et à la distribution du phénotype observé, se révèle plus puissante pour les modèles génétiques les plus vraisemblables dans le contexte d'une maladie complexe (Cobat, 2010; Cobat et al., 2011a). L'idée générale de la méthode consiste à introduire une variable latente binaire qui va capturer l'information de liaison entre le phénotype quantitatif et le marqueur étudié. La vraisemblance est alors composée d'une partie modélisant le lien entre le phénotype quantitatif et la variable latente, et d'une autre partie modélisant le lien entre la variable binaire et le marqueur. La seconde partie de la vraisemblance repose sur la même notion de distribution binomiale décrite dans le paragraphe précédent.

Considérons un trait quantitatif z . Pour une famille nucléaire de n enfants, notons $Z = (z_1, z_2, \dots, z_n)$ le vecteur des n phénotypes quantitatifs observés et $M = (m_{11}, m_{12}, \dots, m_{n1}, m_{n2})$ le vecteur des allèles marqueurs chez les enfants. Nous sommes intéressés par la vraisemblance des observations au locus marqueur sachant les phénotypes des enfants, $P(M/Z)$. Sous l'hypothèse nulle d'absence de liaison génétique, il y a indépendance de transmission des allèles marqueurs parentaux. Ainsi, pour une fratrie, la vraisemblance $P(M/Z)$ des observations au locus marqueur sachant les phénotypes des enfants s'écrit :

$$P(M/Z) = \prod_{j=1}^2 P(M_j/Z)$$

où $P(M_j/Z)$ est la vraisemblance du vecteur des allèles marqueurs reçus du parent j sachant les phénotypes des enfants.

Une variable latente binaire individuelle y_i , qui absorbe l'information de liaison entre le phénotype quantitatif et le marqueur est introduite ; la valeur de cette variable pour le $i^{\text{ème}}$ enfant

est y_i ($y_i \in \{0;1\}$), et $Y = (y_1, y_2, \dots, y_n)$ est le vecteur de ces variables pour la fratrie. Dans la mesure où les y_i ne sont pas observées, l'introduction de Y dans $P(M_j/Z)$ nécessite de sommer sur les 2^n vecteurs Y_k possibles, avec $Y_k = (y_1^k, y_2^k, \dots, y_n^k)$:

$$P(M_j/Z) = \sum_{k=1}^{2^n} P(Y_k/Z)P(M_j/Y_k, Z) = \sum_{k=1}^{2^n} P(Y_k/Z)P(M_j/Y_k)$$

puisque par définition, conditionnellement à Y_k , M_j est indépendant de Z , et où $P(M_j/Y_k)$ est la vraisemblance de travail d'une fratrie pour le parent j .

$P(Y_k/Z)$. Conditionnellement à z_i , et en posant qu' y_i ne dépend que de z_i (*i.e.* le phénotype d'un enfant donné ne dépend pas du phénotype des autres enfants de la fratrie – une hypothèse intuitivement valide sous H_0), la valeur de la variable latente pour un enfant est indépendante de celles des autres membres de la fratrie. L'expression générale de $P(Y_k/Z)$ s'écrit alors :

$$P(Y_k/Z) = \prod_{i=1}^n P(y_i^k/z_i)$$

Pour définir $P(y_i^k/z_i)$, la cohérence de la méthode nécessite de spécifier une fonction de lien monotone entre y_i^k et z_i . Ainsi, une valeur plus élevée de z_i devra être associée à une valeur plus élevée de $P(y_i^k = 1/z_i)$. Pour des raisons de symétrie, il serait équivalent d'associer une valeur élevée de z_i avec une valeur élevée de $P(y_i^k = 0/z_i)$. Une modélisation naturelle consiste à utiliser une fonction cumulée. Une première approche est d'utiliser une fonction paramétrique comme la fonction de répartition normale :

$$P(y_i^k = 1/z_i) = \Phi(z_i)$$

et $P(y_i^k = 0/z_i) = 1 - \Phi(z_i)$

où $\Phi(z)$ représente la fonction cumulée standardisée normale. Alternativement, on peut utiliser une fonction de répartition empirique basée sur la distribution empirique du trait comme les déciles. Le choix des déciles a été suggéré par certains auteurs (Risch and Zhang, 1995). Ce choix est arbitraire et d'autres fonctions monotones, qui représenteraient mieux les données,

peuvent être utilisées. Si on note $d(z_i)$ le décile correspondant à la valeur z_i , $d(z_i) = \{1 ; 2 ; \dots ; 10\}$, on définit :

$$P(y_i^k = 1/z_i) = d(z_i)/10 - 0.05$$

$$P(y_i^k = 0/z_i) = 1 - P(y_i^k = 1/z_i)$$

On remarquera que l'utilisation d'une fonction de lien empirique permet également l'analyse des phénotypes ordinaux (Alcais et al., 2001).

$P(M_j/Y_k)$. La contribution à la vraisemblance d'un enfant i avec $y_i^k = r$ ($r = 0$ ou 1) est $\alpha^r(1-\alpha)^{1-r}$ ou $\alpha^{1-r}(1-\alpha)^r$ selon qu'il a reçu l'allèle A ou a de son parent Aa . Dans la situation usuelle où la phase est inconnue, et en reprenant le paramètre α introduit dans le cas binaire, la vraisemblance de la fratrie pour le parent j s'écrit :

$$\begin{aligned} P(M_j/Y_k) &= 0,5\alpha^{n_{Ajk}(1)}(1-\alpha)^{T_k-n_{Ajk}(1)}\alpha^{n-T_k-n_{Ajk}(0)}(1-\alpha)^{n_{Ajk}(0)} \\ &+ 0,5(1-\alpha)^{n_{Ajk}(1)}\alpha^{T_k-n_{Ajk}(1)}(1-\alpha)^{n-T_k-n_{Ajk}(0)}\alpha^{n_{Ajk}(0)} = g_j^k(\alpha) \end{aligned}$$

où T_k correspond au nombre d'enfants ayant un $y_i^k = 1$ pour un vecteur Y_k donné, et $n_{Ajk}(r)$ est le nombre d'enfants avec $y_i^k = r$ ayant reçu l'allèle A du parent hétérozygote j .

Finalement, la vraisemblance pour une famille avec deux parents et n enfants, $P(M/Z)$, s'écrit :

$$P(M/Z) = \prod_{j=1}^2 \sum_{k=1}^{2^n} g_j^k(\alpha) \prod_{i=1}^n P(y_i^k/z_i)$$

Comme dans le cas binaire, on peut construire un test de liaison fondé sur un rapport de vraisemblance ou l'hypothèse alternative de liaison génétique est définie par $\alpha > 0,5$, où α s'interprète comme la probabilité pour qu'un enfant avec $y_i^k = 1$ ait reçu l'allèle marqueur transmis avec l'allèle prédisposant à une valeur élevée du phénotype. La statistique de test, $\lambda_{nMLB-QTL}$, a la même distribution que dans le cas binaire, i.e. un mélange 50 % : 50 % de χ^2 à 0 et 1 degré de liberté.

La méthode nMLB-QTL a été implémentée en langage C et liée au programme GENEHUNTER (Kruglyak et al., 1996; Cobat et al., 2011a). Elle permet l'analyse simple point et multipoint de la même façon que la méthode MLB-QTL originale (Abel and Muller-Myhsok, 1998). La version 3.0 du logiciel MLB-QTL qui incorpore la version originale et la nouvelle version de la méthode est disponible sur le site www.hgid.net.

4.2 Choix des marqueurs génétiques

Le contrôle de qualité des génotypes a été réalisé en utilisant le logiciel PLINK (Purcell et al., 2007). Tous les individus retenus pour l'analyse de liaison avaient un call-rate > 95 %. Le choix du sous-ensemble de SNPs pour l'analyse de liaison a été obtenu en appliquant les filtres de sélection suivants : call-rate > 99,5 %, fréquence de l'allèle mineur (MAF) $\geq 0,05$, valeur de p pour le test de Hardy-Weinberg > 0,001 et aucune erreur mendélienne. Ces critères stringents ne sont pas un handicap puisqu'au total, 342 568 SNPs ont passé l'ensemble de ces filtres. Nous avons donc réalisé un second filtrage pour sélectionner 45 947 SNPs « communs » (définis par une MAF $\geq 0,1$) affichant de faibles niveaux de déséquilibre de liaison deux à deux ($r^2 < 0,2$ pour chaque paire de SNPs conservés). Une modélisation théorique montre qu'avec ce nombre de marqueurs, une couverture presque continue du génome est assurée (Kruglyak et al., 1996). De fait, l'information était en moyenne supérieure à 99 % sur les autosomes. Elle variait de 99,2 % sur le chromosome 20 à 99,6 % sur le chromosome 9. Moins de 5 % du génome (159 Mbs) était couvert par des SNPs espacés de plus de 1 Mb. Seuls 8 SNPs se situaient dans des régions chromosomiques dont l'information était inférieure à 95 %.

4.3 Description de l'échantillon de familles multiples

Nous avons extrait de notre échantillon primaire les familles nucléaires de l'ethnie *Kinh* (cf dont les deux parents étaient génotypés et qui incluaient au moins deux enfants atteints pour lesquels la définition phénotypique selon Ridley et Jopling était disponible, soit au total 306 individus distribués dans 72 familles. Le tableau 8 montre la répartition des enfants atteints (74 PB et 88 MB – définition OMS-82) selon le sexe, l'âge au diagnostic, le nombre de lésions et

l'indice bacillaire. Dans notre échantillon, cette classification correspond exactement à regrouper les individus TT et BT dans le groupe PB, et les individus BB, BL et LL dans le groupe MB.

L'âge moyen des enfants au moment du diagnostic était de 17 ans. Le sex-ratio hommes-femmes était de 1,9 (106 hommes et 56 femmes). L'âge et le sexe n'étaient pas significativement associés à la polarisation selon la définition OMS-82 dans notre échantillon, en analyse univariée comme multivariée. Ainsi, nous avons décidé de ne pas ajuster les phénotypes sur ces variables, comme nous l'aurait permis le nMLB-QTL.

Tableau 8. Caractéristiques de 162 enfants dans les 72 familles de l'analyse de liaison

Classification OMS-82	PB (N = 74)	MB (N = 88)	p
Hommes	52	54	
Femmes	22	34	
Sex ratio	2,4	1,6	0,31
Age (moyen [min;max])	17 [4;35]	17 [5;37]	0,43
Classification de Ridley et Jopling			$< 10^{-6}$
TT	40	0	
BT	34	0	
BB	0	42	
BL	0	32	
LL	0	14	
Nombre de lésions (moyen [min;max])	2 [1;10]	9 [1;>20]	$< 10^{-6}$
> 5 lésions	4/69	47/85	$< 10^{-6}$
IB > 0	1/66	22/86	$2 \cdot 10^{-5}$
Cicatrice de BCG présente	19/63	21/82	0,67

4.4 Agrégation familiale des sous-types de lèpre

La proportion de familles dites concordantes (*i.e.* incluant exclusivement des enfants PB ou MB, définition OMS-82) était significativement plus élevée que celle attendue sur la base des seules proportions de PB et de MB (46 et 54 %) ($p = 0,005$; figure 22). Dans cet échantillon, nous disposions également du nombre de lésions pour 147 enfants. Ainsi, 64 familles (275 personnes) étaient disponibles pour analyser la polarisation selon la classification de l'OMS-96 : il y avait 98 enfants PB et 49 enfants MB. L'hypothèse d'une agrégation non-aléatoire des cas (*i.e.* la proportion de familles concordantes) restait significative avec cette classification OMS-96 ($p = 0,02$).

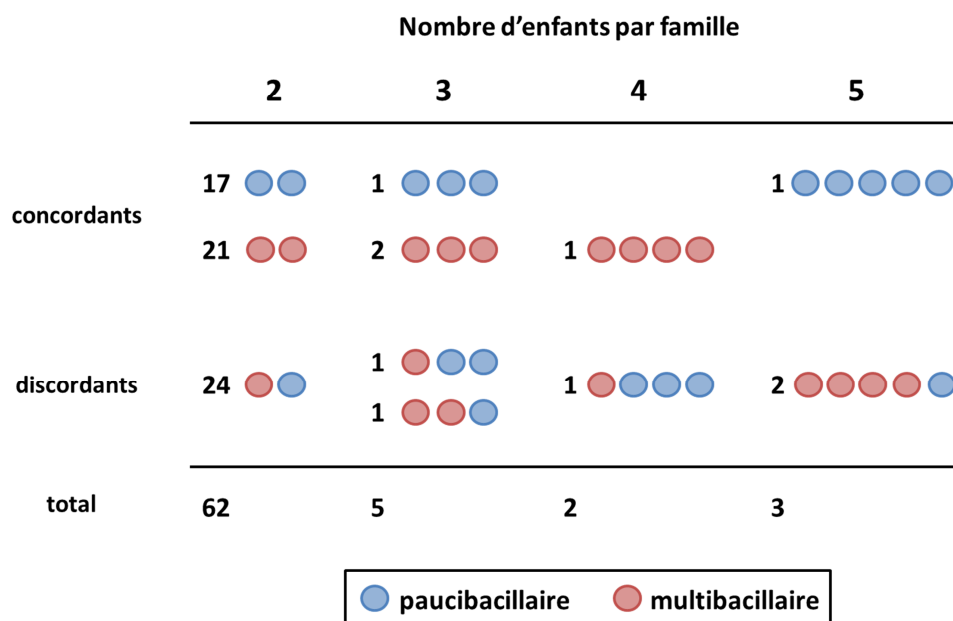


Figure 22. Répartition des fratries de l'analyse de liaison selon leur concordance clinique
 Chaque cercle représente un enfant PB (bleu) ou MB (rouge). Les familles sont ordonnées selon la taille de la fratrie et sa concordance de leurs fratries (partage du même sous-type de lèpre ou non). Le nombre de chaque type de fratrie est indiqué à gauche.

4.5 Résultats de l'analyse de liaison

Les résultats de l'étude de la liaison pangénomique de la polarisation de la lèpre sont représentés sur la Figure 23. Dans la figure 23A, le phénotype d'intérêt est la polarisation codée comme un trait binaire (PB vs. MB, OMS-82). Nous avons identifié une région du chromosome 19p suggestive de liaison, culminant au SNP rs2049051 (position 19:18315831; LOD-score de 3,19, $p = 6 \cdot 10^{-5}$). En effet, le seuil de significativité à l'échelle du génome pour une analyse de liaison multimarqueurs est de 3,6. Le seuil évocateur de liaison est quant à lui de 2,2 (Kruglyak et al., 1996). Les intervalles de confiance à 90 % et 95 % des meilleurs loci ont été estimés selon la méthode de l'intervalle de support (Dupuis and Siegmund, 1999). Ces intervalles peuvent être approchés de façon satisfaisante en délimitant, de part et d'autre du marqueur responsable du meilleur LOD-score (LOD_{max}), la région dans laquelle le LOD-score reste supérieur au $LOD_{max} - 1$ (intervalle à 90 %) et au $LOD_{max} - 1,5$ (intervalle à 95 %).

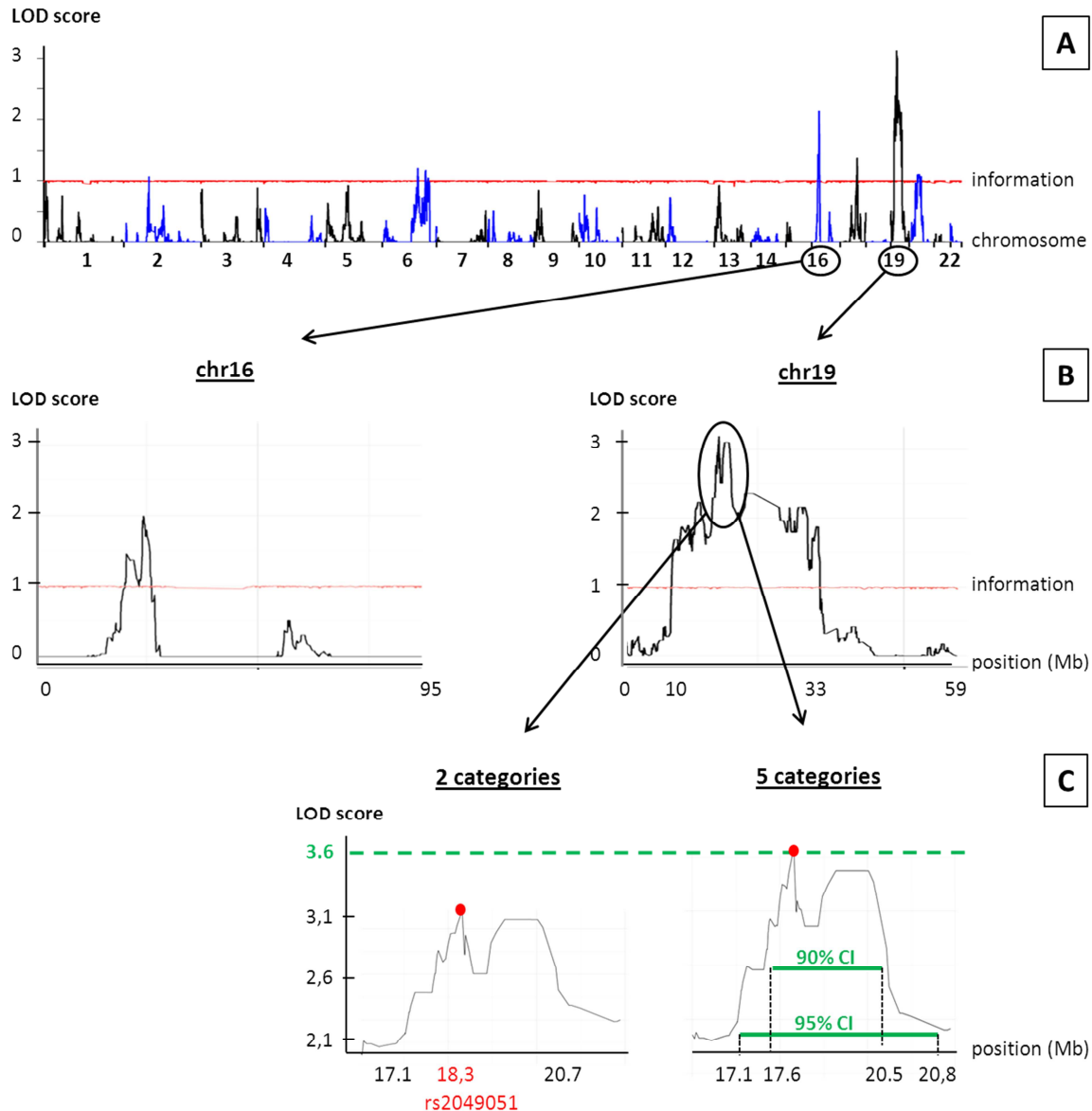


Figure 23. Résultat de l'analyse de liaison pangénomique de la polarisation de la lèpre

A. LOD-score multimarqueurs (axe des y) calculé par la méthode nMLB le long de chaque chromosome (axe des x) pour la liaison de la polarisation de la lèpre en deux catégories (TT + BT vs. BB + BL + LL). La ligne rouge indique la qualité de l'informativité moyenne de l'échantillon en chaque position (0 = nulle, 1 = parfaite).

B. Zoom sur les chromosomes 16 et 19, où le seuil évocateur de liaison (LOD-score > 2.2) est dépassé. La position sur l'axe des x est indiquée en millions de paires de bases (Mb).

C. Zoom sur la zone d'intervalle à 90% du chromosome 19 (LOD-score > 2.1). Le LOD-score est calculé pour la polarisation codée en 2 catégories (gauche) ou en 5 catégories (TT < BT < BB < BL < LL, droite). Le point rouge représente le pic de LOD-score obtenu pour rs2049051: 3,19 avec la classification en deux catégories (gauche) et 3,57 avec la classification en 5 catégories (droite). Les intervalles de confiance à 90% et 95% pour la liaison de la polarisation correspondent à une diminution de 1 et 1,5 du LOD_{max} , soit 2,57 et 2,07. La ligne pointillée en vert (LOD-score = 3,6) indique le seuil de liaison significatif à l'échelle du génome.

Ainsi l'intervalle de confiance à 90 % définissait une région de 3,6 Mbs s'étendant de 17,1 à 20,7 Mbs. la seconde région de liaison la plus significative est localisée sur le bras court du chromosome 16, avec un LOD_{max} de 1,99 au SNP rs198174 (pos 16:24247031, $p = 1 \cdot 10^{-3}$).

Nous avons ensuite affiné le phénotype en 5 catégories pour utiliser le nMLB-QTL avec des quintiles. Les probabilités d'être MB ont respectivement été fixées à 0 ; 0,1 ; 0,8 ; 0,9 et 1 pour les enfants classés TT, BT, BB, BL et LL selon Ridley et Jopling. Ce choix de probabilités non symétriques par rapport à 0,5 a pour objectif de modéliser au mieux les similitudes cliniques entre les individus TT et BT d'une part, et les individus BB, BL et LL d'autre part.

Avec le nMLB-QTL, le LOD-score du pic de liaison sur le chromosome 19 augmentait à 3,57 pour le SNP rs2049051 ($p = 2,5 \cdot 10^{-5}$). L'intervalle de confiance à 90 % se réduisait à une région de 2,9 Mb s'étendant de 17,6 à 20,5 Mbs (figure 23C) tandis que l'intervalle de confiance de 95 % s'étendait de 17,0 à 20,8 Mbs. A l'inverse, avec cette classification en 5 catégories, le pic sur le chromosome 16 diminuait (LOD-score de 1,63 pour le SNP rs1976194, distant de 180 kb du SNP rs198174, dont le LOD-score était de 1,59, $p = 3 \cdot 10^{-4}$). En outre, le raffinement du phénotype n'a pas identifié de nouvelles régions d'intérêt (Figure 24).

Pour tester l'indépendance des deux pics des chromosomes 16 et 19, nous avons réalisé l'analyse de liaison en retirant tour à tour les familles qui contribuaient « le plus » à chacun des pics. Le LOD-score pour chaque famille variait entre - 0,33 et + 0,36 aux deux pics : nous avons retiré les familles dont le LOD-score était $> 0,15$. L'analyse de liaison excluant les familles qui contribuaient au pic sur le chromosome 19 révélait que la valeur du LOD_{max} du chromosome 16 augmentait à plus de 4, tandis que le LOD_{max} du chromosome 19 diminuait à moins de 1. L'inverse s'observait en retirant les familles contribuant au pic du chromosome 16 (toutes différentes de celles contribuant au pic du chromosome 19) (Figure 25). Ces valeurs absolues de LOD-score n'ont aucune significativité, en revanche, elles suggèrent fortement que des groupes différents de familles sont liées à des régions différentes en ce qui concerne la polarisation de la lèpre.

En utilisant la classification OMS-96 (PB vs. MB), aucune liaison même évocatrice n'a été observée: le LOD-score le plus élevé était de 2,07 sur le chromosome 20 (pour le SNP rs10891749 en position 20:58406750) (Figure 25). Avec cette définition, le LOD-score était de 0 pour rs1976194 sur le chromosome 16 et de 0,40 pour rs2049051 sur le chromosome 19.

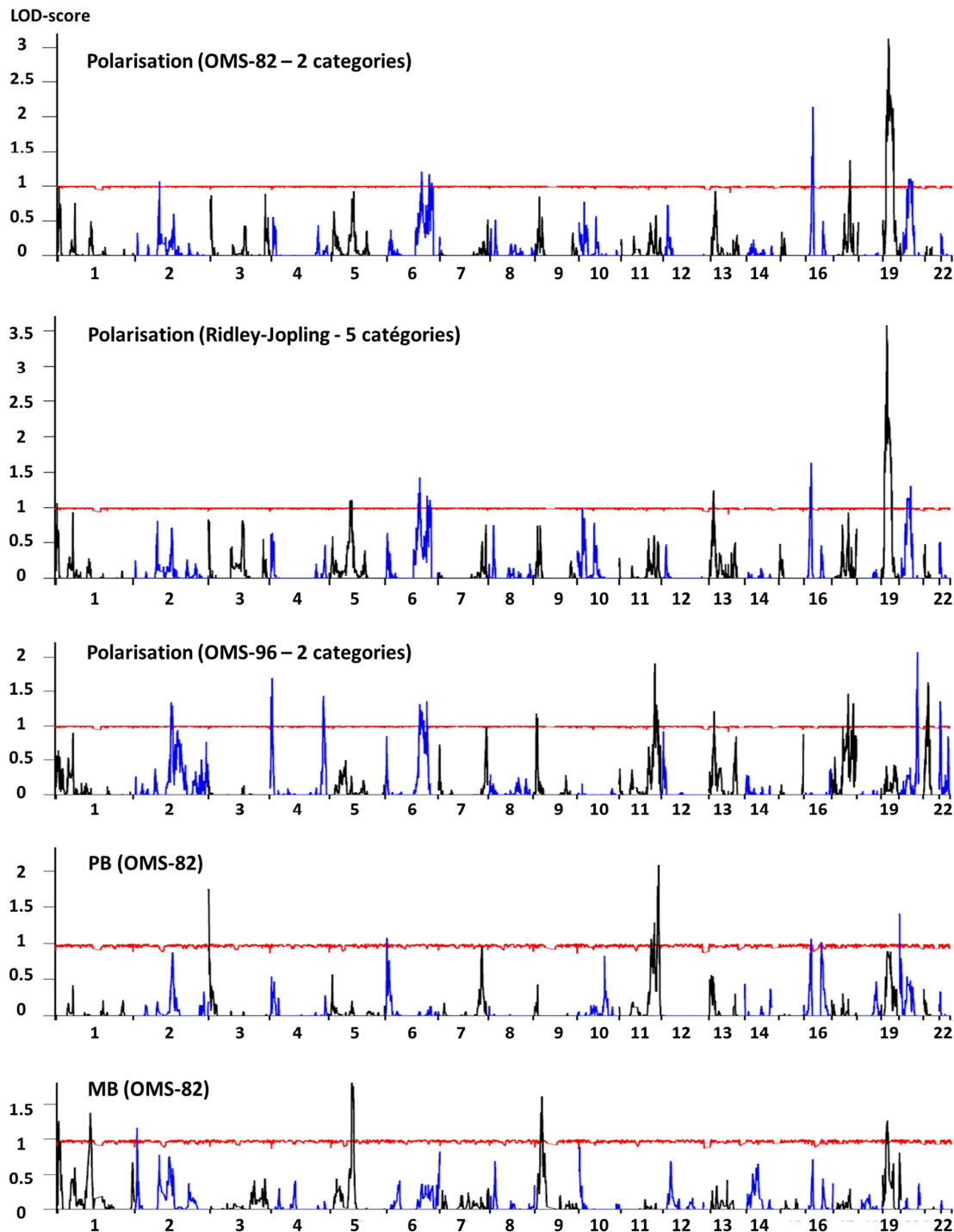


Figure 25. Analyse de liaison pangénomique pour 5 phénotypes

LOD-score multimaqueurs (axe des y) calculé par la méthode nMLB le long de chaque chromosome (axe des x) pour la polarisation de la lèpre (OMS-82); la polarisation codée en 5 catégories (TT < BT < BB < BL < LL, selon Ridley et Jopling); la polarisation codée en 2 catégories (OMS-96); la forme PB (OMS-82, équivalent ici à TT + BT) et la forme MB (OMS-82, équivalent ici à BB + BL + LL). La ligne rouge indique l'informativité moyenne de l'échantillon en chaque position, qui était supérieure à 0,95 sur l'ensemble du génome.

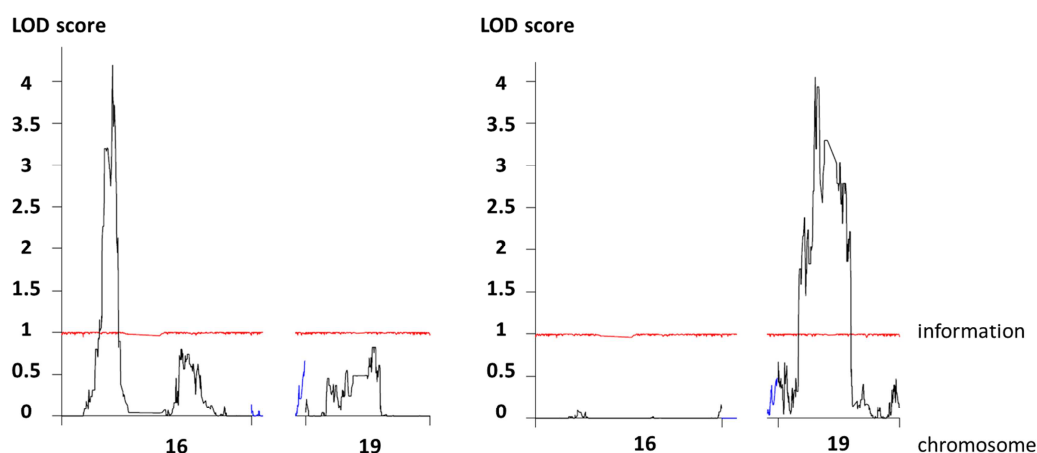


Figure 25. Impact des familles contribuant au pic de liaison du chromosome 19 sur le pic de liaison du chromosome 16 et inversement

LOD-score multimaqueurs (axe des y) calculé par la méthode nMLB le long des chromosomes 16 et 19 (axe des x) sur l'ensemble des familles sauf celles contribuant au pic de liaison du chromosome 19 (gauche) ou au pic de liaison du chromosome 16 (droite).

Par ailleurs, le signal de liaison de la polarisation avec le chromosome 19 n'était pas dû seulement à l'effet de l'un des sous-groupe PB ou MB. En effet, le LOD_{max} sur le chromosome 19 était inférieur à 1,0 pour PB (en utilisant le MLB en mode binaire codant TT = BT = 1). Il était inférieur à 1,2 pour MB (en codant BB = BL = LL = 1, toujours avec le MLB) (Figure 25). La somme des LOD scores pour MB et PB ne dépassait pas 2,0 sur le chromosome 19 (maximum sur le génome). Ce résultat est inférieur au LOD score de 3,19 obtenu en analysant la polarisation en mode binaire.

4.6 Discussion

Notre étude a identifié une région sur le chromosome 19p12-p13 liée à la polarisation de la lèpre, définie selon la classification de Ridley et Jopling. La région définie par un intervalle de confiance à 95 %, large de 3,8 Mbs, comporte 180 gènes. Un premier regard en identifie 6 d'intérêt a priori, sans oublier pas que la beauté des approches pangénomiques, libres d'hypothèses, est justement d'identifier des gènes dont l'intérêt ne saute pas aux yeux : UNC-13 homologue A (*UNC13A*), Janus kinase 3 (*JAK3*), le récepteur $\beta 1$ de l'IL-12 (*IL12R β 1*), l'interféron gamma-inductible 30 (*IFI30*), l'oncogène *RAB3A* de la famille RAS et le facteur de

régulation associé à l’X contenant Ankyrin (*RFXANK*). En particulier, rs2049051 est situé à 230kb de *IL12Rβ1* et à 1 Mb de *RFXANK*. Une mutation homozygote dans le second gène est responsable d’un déficit complexe d’expression du CMH de classe II (Ouederni et al., 2011). *IL12Rβ1* a été associé à des formes mendélienne d’infection par *M. tuberculosis* et *M. tilburgii* (Schepers et al., 2013). En outre, une étude a montré une association entre la polarisation de la lèpre et *IL12Rβ2*, situé sur le chromosome 1 (Ohyama et al., 2005). La liste des 180 gènes figure dans l’annexe 2 (tableau S2 de l’article « Genomewide linkage analysis of leprosy polarization implicates chromosomal region 19p12-p13 »).

Par ailleurs, la région de liaison 19p12 a été rapportée dans des études d’autres phénotypes : la cirrhose biliaire primitive (Liu et al., 2012), la neuropathie de Charcot-Mary-Tooth (Kennerson et al., 2001), et les maladies inflammatoires de l’intestin (maladie de Crohn et rectocolite hémorragique) (van Heel et al., 2003; Shugart et al., 2008). Pour les deux dernières, des mécanismes physiopathologiques communs avec la lèpre ont été mis en évidence.

L’observation d’un signal de liaison plus faible lorsque la polarisation est définie selon la classification OMS-96 est intéressante. En effet, sous l’hypothèse que le signal de liaison observé avec la classification de Ridley-Jopling est réel (*i.e.* un vrai positif), la diminution du signal de liaison observée avec la définition OMS-96 pourrait refléter une pertinence biologique moindre de cette classification opérationnelle. Les résultats de cette étude suggèrent que l’utilisation d’une classification clinique fine, si elle n’est pas nécessaire sur le terrain, peut augmenter substantiellement la puissance des études portant sur la polarisation de la lèpre.

Nous n’avons pas observé de signaux de liaison significatifs avec des régions préalablement rapportées dans des études sur la lèpre *per se* – les régions 2p14, 6p21 et 6q25, 10p13, 17q11-21, 20p12-13 (Misch et al., 2010). Ceci est en faveur d’un modèle en deux étapes, dans lequel il existe des déterminants génétiques spécifiques de la polarisation et différents de la lèpre *per se*. Nous n’avons pas non plus observé de région de liaison commune à d’autres études sur un sous-type de lèpre – la région 10p13 avec la forme PB et les régions 17q11 et 20p12 avec la forme MB (Siddiqui et al., 2001; Miller et al., 2004; Jamieson et al., 2004; Grant et al., 2014).

Une suite naturelle à l’identification d’une région de liaison, est un *fine-mapping* de cette région. Cela consiste en une étude d’association avec un grand nombre de marqueurs au sein de la région de liaison identifiée. Le but est d’identifier un ou plusieurs marqueurs qui sont soit causaux, soit en déséquilibre de liaison avec les variations causales. Le déroulement de notre thèse nous a plutôt conduit à réaliser d’emblée une étude pan-génomique pour notre phénotype.

L'identification d'une région de liaison, si elle représente une avancée (du génome à 30 Mb), ne permet pas l'identification moléculaire du ou des variants ayant généré ce signal. Il est donc nécessaire de poursuivre l'investigation de cette (ou ces) région(s) par des études dites de *fine-mapping*. Le principe de ces études consiste à tester l'association préférentielle d'un allèle donné avec la survenue d'un phénotype particulier en saturant en nombre de marqueurs la région de liaison identifiée. Toutefois, l'interprétation d'une association significative restera toujours incertaine depuis l'identification du variant effectivement causal jusqu'aux faux positifs observés en raison du nombre de variants testés en passant par l'identification d'un variant sans lien causal avec le phénotype étudié mais en déséquilibre de liaison avec un variant biologiquement pertinent. Si le déroulement de notre thèse et l'évolution du coût des efforts de génotypage nous a conduit à réaliser une étude d'association pan-génomique, il faut garder en mémoire que notre analyse de liaison préalable nous invite à considérer différemment des variants également significatifs selon qu'ils appartiennent ou non à une région de liaison préalablement identifiée. Cette stratégie est d'ailleurs préconisée et modélisée par Roeder et al (Roeder et al., 2006).

Le chapitre suivant décrit les différentes méthodes disponibles pour tester une telle association dans le cas particulier d'un échantillon familial. Et comment les limites de ces méthodes pour l'analyse d'un phénotype « secondaire » (il faut développer une lèpre pour pouvoir la polariser – en ce sens l'étude de la polarisation est l'étude d'un phénotype secondaire) nous ont amené à proposer une nouvelle approche pour l'étude de ces phénotypes secondaires.

Méthodes d'étude d'association génétique

5.1 FBAT : présentation générale

L'approche FBAT, acronyme pour « Family Based Association Tests », est une généralisation des tests d'association en famille qui reposent sur le *Transmission Disequilibrium Test* ou TDT (Spielman et al., 1993). FBAT 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, 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 avec plusieurs enfants (Laird et al., 2000; Rabinowitz and Laird, 2000; Lange et al., 2003).

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

$$U = S - E(S)$$
$$\text{avec } S = \sum_{i,j} T_{ij} X_{ij}$$

où X_{ij} et T_{ij} sont fonction du génotype et du phénotype de l'enfant j de la famille i . Le codage de X_{ij} dépend du modèle génétique considéré (Tableau 9). Dans le cas binaire, si T_{ij} est codé 1 pour les enfants atteints et 0 pour les enfants non atteints, seuls les enfants atteints contribuent à S et sa formulation se simplifie en $S = \sum_{ij} X_{ij}$. Dans le cas quantitatif, il est proposé de coder T_{ij} par $(y_{ij} - \mu)$, où μ est la moyenne phénotypique dans la population ; la statistique U est alors équivalente à la statistique de Rabinowit (Rabinowitz, 1997). Dans la mesure où les phénotypes sont considérés comme fixés, aucune hypothèse n'est faite sur la distribution des phénotypes. Ceci constitue un des grands avantages de FBAT par rapport au Quantitative-TDT qui repose sur les méthodes de décomposition de la variance (Waldman et al., 1999).

Génotype	X		
	Additif	Dominant	Récessif
<i>aa</i>	0	0	0
<i>Aa</i>	1	1	0
<i>AA</i>	2	1	1

Tableau 9. Codage des génotypes selon le modèle génétique

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

$$E(S) = \sum_{i,j} T_{ij} E(X_{ij})$$

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 $X(g)$, et $\sum_g$ est la somme sur tous les génotypes possibles des enfants j de la famille i . Le Tableau 10 donne les valeurs de $E(X_{ij})$ lorsque les génotypes parentaux sont connus selon les combinaisons parentales et le modèle génétique considéré.

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 (Rabinowitz and Laird, 2000). Dans le cas général, la variance s'écrit :

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

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 10. Espérance sous H_0 du génotype X d'un enfant

Espérance de X , notée $E(X)$, selon les génotypes parentaux et le modèle génétique.

Pour un échantillon suffisamment grand (en pratique au moins 10 transmissions), la statistique de test $Z = \frac{U}{\sqrt{V}}$ suit une loi normale centrée réduite.

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. L'hypothèse nulle adéquate est une absence d'association en présence de liaison et l'hypothèse alternative est une association en présence de liaison. En présence de liaison, au sein d'une 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. La distribution de U peut alors être déterminée en conditionnant sur le nombre d'allèles partagés « Identical By Descent » (IBD), ie des allèles 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 (Laird and Lange, 2006). En pratique, une variance empirique qui corrige la corrélation des transmissions peut être utilisée (Lake et al., 2000). Cette variance empirique de U est donnée par $V = \sum_i [S_i - E(S_i)]^2$ et elle est implémentée dans le logiciel FBAT (option « -e »).

5.2 FBAT pour un trait binaire

Notre sujet de thèse nous a conduit à étudier la polarisation de la lèpre comme un phénotype à deux issues mutuellement exclusives : PB ou MB, quelle que soit la définition utiliser pour définir ces deux formes. Nous présentons différentes stratégies pour tester l'association d'un phénotype binaire dans un échantillon familial. Nous montrons aussi certaines limites de FBAT pour l'analyse d'un phénotype binaire secondaire, situation qui nous intéresse tout particulièrement dans la mesure où la polarisation est un phénotype binaire secondaire, i.e. il faut avoir développé une lèpre pour la polariser.

Par souci de simplicité, nous considérons l'analyse d'un marqueur **biallélique** dans un échantillon de **trios** (familles formées de deux parents et un enfant atteint) sans données manquantes.

5.2.1 FBAT

En utilisant les mêmes notations que dans la publication originale de FBAT (Laird et al., 2000),

$$\text{soit } S = \sum_i T_i X_i$$

où X_i et T_i représentent le génotype au locus testé et le phénotype de l'enfant du trio i . L'espérance de X_i est calculée conditionnellement aux génotypes parentaux sous l'hypothèse nulle d'absence d'association ou de liaison.

$$\text{Soit } E = E(S) = E\left(\sum_i T_i X_i\right) = \sum_i T_i E(X_i)$$

$$\text{Soit } V = \text{Var}(S) = \text{Var}\left(\sum_i T_i X_i\right) = \sum_i \text{Var}(T_i X_i) = \sum_i T_i^2 \text{Var}(X_i)$$

puisque les génotypes des enfants ne dépendent que de leurs propres parents, et sont indépendants des autres familles.

La statistique FBAT s'écrit :

$$FBAT = \frac{(S - E)^2}{V}$$

$$FBAT \underset{H_0}{\sim} \chi^2_{1df}$$

Sous un modèle additif, X_i représente le nombre de copies de l'allèle étudié (0, 1 ou 2). Un codage naturel du phénotype est $T = 1$ pour les personnes atteintes et $T = 0$ pour les non-atteintes. S s'écrit alors:

$$S = \sum_{i \in \text{atteints}} 1 \times X_i + \sum_{i \in \text{non-atteints}} 0 \times X_i = \sum_{i \in \text{atteints}} X_i$$

Dans un échantillon sans données parentales manquantes, les individus non atteints ne contribuent donc pas à la statistique. Toutefois, lorsque certaines données parentales manquent, ces individus non-atteints auront un impact sur la statistique, car ils peuvent être utilisés pour reconstruire les génotypes parentaux, sous certaines conditions (Knapp, 1999). Une situation implémentée dans FBAT.

L'hypothèse nulle d'absence d'association entre le phénotype et un allèle correspond à une transmission aléatoire de cet allèle des parents hétérozygotes aux enfants atteints. En notant p la probabilité de transmission de cet allèle, les hypothèses H_0 et H_1 peuvent s'écrire :

$$H_0 : p = \frac{1}{2}$$

$$H_1 : p \neq \frac{1}{2}$$

L'allèle sera considéré comme respectivement « à risque » ou « protecteur » vis-à-vis de la maladie, si $p > \frac{1}{2}$ ou $p < \frac{1}{2}$.

Il est courant d'étudier un phénotype « primaire » (dans le cas d'une maladie : atteint vs. non-atteint). Cependant, les chercheurs peuvent aussi s'intéresser à la contribution génétique à un phénotype dit « secondaire ». Dans le cas particulier d'une maladie, ce phénotype peut être par exemple la forme clinique : sévère vs. non-sévère. A première vue, FBAT peut être utilisé pour analyser ces phénotypes secondaires en calculant la statistique originale indépendamment pour les modalités du phénotype secondaire, deux dans le cas binaire. En notant D_1 et D_2 les modalités du phénotype secondaire, p_1 et p_2 les probabilités de transmission de l'allèle testé aux enfants D_1 et D_2 , nous avons :

$$S_1 = \sum_{i \in 1} T_i X_i, S_1 = \sum_{i \in 1} X_i, \text{FBAT}_1 = \frac{(S_1 - E_1)^2}{V_1}$$

$$\mathbf{H}_0 : \mathbf{p}_1 = \frac{1}{2}$$

$$\mathbf{H}_1 : \mathbf{p}_1 \neq \frac{1}{2}$$

$$S_2 = \sum_{i \in 2} T_i X_i, S_2 = \sum_{i \in 2} X_i, \text{FBAT}_2 = \frac{(S_2 - E_2)^2}{V_2}$$

$$\mathbf{H}_0 : \mathbf{p}_2 = \frac{1}{2}$$

$$\mathbf{H}_1 : \mathbf{p}_2 \neq \frac{1}{2}$$

Toutefois, en raison de la nature bidimensionnelle du phénotype étudié (ie atteint/non-atteint et sévère/non-sévère), le rejet de l'hypothèse nulle ne permet pas de distinguer les allèles associés à la maladie *per se* (c'est à dire indépendamment de sa sévérité) des allèles spécifiquement associés à la sévérité de la maladie. FBAT ne propose pas de test ni même de stratégie pour analyser ces phénotypes secondaires.

Une première idée consiste à considérer le phénotype binaire comme un phénotype quantitatif que l'on analysera avec FBAT. Toutefois nous montrons dans le paragraphe suivant que ce procédé n'est pas optimal. Cette constatation nous a amené à proposer deux nouvelles approches, FBAT_{het} et $\text{FBAT}_{\text{het free}}$, qui forment le cœur de ce chapitre et ont donné lieu à une première publication (Gaschignard et al. en Annexe 3).

5.2.2 $\text{FBAT}_{+1/-1}$

En attribuant une valeur numérique à chaque réalisation du phénotype secondaire, nous pouvons analyser le phénotype comme un trait quantitatif. La stratégie proposée consiste alors à tester une transmission différentielle d'un allèle donné entre les enfants D_1 et les enfants D_2 . Ce test aura a priori une puissance optimale si la transmission d'un allèle qui contribue positivement à S pour les enfants D_1 contribue négativement à S pour les enfants D_2 . On modélisera cette situation optimale en attribuant un signe contraire à T_1 et T_2 (le phénotype des enfants D_1 et D_2), par exemple $+1$ et -1 . Compte-tenu de la linéarité de FBAT, choisir un codage de type $+k/-k$, $\forall k \in \mathbb{R}$ donnera des résultats identiques. Nous appelons $\text{FBAT}_{+1/-1}$ ce test.

Notons $S_{+1/-1}$, $E_{+1/-1}$ et $V_{+1/-1}$ les valeurs de S , E et V pour le test $\text{FBAT}_{+1/-1}$. En reprenant les notations du paragraphe précédent,

$$\begin{aligned}
S_{+1/-1} &= \sum_{i \in 1} 1 \times X_i + \sum_{i \in 2} (-1) \times X_i = S_1 - S_2 \\
E_{+1/-1} &= E(S_1 - S_2) = E(S_1) - E(S_2) = E_1 - E_2 \\
V_{+1/-1} &= \text{Var}(S_1 - S_2) = \text{Var}(S_1) + \text{Var}(S_2) = V_1 + V_2 \\
\mathbf{FBAT}_{+1/-1} &= \frac{((S_1 - S_2) - (E_1 - E_2))^2}{V_1 + V_2}
\end{aligned}$$

On montre que $\mathbf{FBAT}_{+1/-1}$ est un test d'hétérogénéité Q de Cochran pour S_1 et S_2 (Démonstration 1). Cependant, les hypothèses de ce test s'écrivent :

$$\mathbf{H}_0 : S_1 = S_2$$

$$\mathbf{H}_1 : S_1 \neq S_2$$

Ces hypothèses diffèrent, ce qui était intuitivement prévisible, de celles que nous souhaitons tester, à savoir : $H_0 : p_1 = p_2$ et $H_1 : p_1 \neq p_2$. Nous proposons donc dans la suite du texte deux nouveaux tests mieux adaptés à la question initialement posée.

5.3 Nouveaux tests d'association familiale pour un trait binaire

5.3.1 $\mathbf{FBAT}_{het}$

Pour étudier un phénotype secondaire, une seconde approche naturelle (la première étant $\mathbf{FBAT}_{+1/-1}$) consiste à effectuer un test d'homogénéité entre les deux statistiques S_1 et S_2 . Nous appelons $\mathbf{FBAT}_{het}$ ce test.

$$\begin{aligned}
\mathbf{FBAT}_{het}(\mathbf{D}_1, \mathbf{D}_2) &= \text{Test d'homogénéité}(S_1, S_2) \\
&= \frac{(S_1 - E_1)^2}{V_1} + \frac{(S_2 - E_2)^2}{V_2} - \frac{(S_1 - E_1 + S_2 - E_2)^2}{V_1 + V_2} \\
&= \frac{(V_1 + V_2)(V_2(S_1 - E_1)^2 + V_1(S_2 - E_2)^2)}{V_1 V_2 (V_1 + V_2)} - \frac{V_1 V_2 (S_1 - E_1 + S_2 - E_2)^2}{V_1 V_2 (V_1 + V_2)} \\
&= \frac{V_2^2 (S_1 - E_1)^2 - 2V_1 V_2 (S_1 - E_1)(S_2 - E_2) + V_1^2 (S_2 - E_2)^2}{V_1 V_2 (V_1 + V_2)}
\end{aligned}$$

$$\begin{aligned}\mathbf{FBAT}_{\text{het}}(\mathbf{D}_1, \mathbf{D}_2) &= \frac{\left(\frac{S_1 - E_1}{V_1}\right)^2 - 2\left(\frac{S_1 - E_1}{V_1}\right)\left(\frac{S_2 - E_2}{V_2}\right) + \left(\frac{S_2 - E_2}{V_2}\right)^2}{\frac{(V_1 + V_2)}{V_1 V_2}} \\ &= \frac{\left(\frac{S_1 - E_1}{V_1} - \frac{S_2 - E_2}{V_2}\right)^2}{\frac{1}{V_1} + \frac{1}{V_2}} = \frac{\left(\frac{S_1 - E_1}{V_1} - \frac{S_2 - E_2}{V_2}\right)^2}{\frac{1}{V_1^2} V_1 + \frac{1}{V_2^2} V_2}\end{aligned}$$

Si on analyse l'échantillon formé de l'ensemble des individus malades ($D_1 + D_2$) en codant le phénotype D_1 par $\frac{1}{V_1}$ et D_2 par $-\frac{1}{V_2}$ (et non $+1$ et -1 comme dans $\mathbf{FBAT}_{+1/-1}$), alors $\mathbf{FBAT}_{\text{het}} \equiv \mathbf{FBAT}$.

En effet,

$$\begin{aligned}S\left(T_1 = \frac{1}{V_1}, T_2 = -\frac{1}{V_2}\right) &= \sum_{i \in 1} \frac{1}{V_1} X_i - \sum_{i \in 2} \frac{1}{V_2} X_i = \frac{1}{V_1} S_1 - \frac{1}{V_2} S_2 \\ E &= \frac{1}{V_1} E_1 - \frac{1}{V_2} E_2 \text{ et } V = \frac{1}{V_1^2} V_1 + \frac{1}{V_2^2} V_2 \\ \text{et } \mathbf{FBAT}\left(T_1 = \frac{1}{V_1}, T_2 = -\frac{1}{V_2}\right) &= \frac{\left(\frac{S_1 - E_1}{V_1} - \frac{S_2 - E_2}{V_2}\right)^2}{\frac{1}{V_1} + \frac{1}{V_2}} = \mathbf{FBAT}_{\text{het}}\end{aligned}$$

Les hypothèses du test s'écrivent alors:

$$\mathbf{H}_0 : \mathbf{p}_1 = \mathbf{p}_2 = \frac{1}{2}$$

$$\mathbf{H}_1 : \mathbf{p}_1 \neq \mathbf{p}_2$$

Notons que, pour un modèle génétique additif et pour un échantillon de trios sans données parentales manquantes, le codage $T_1 = \frac{1}{V_1}$ et $T_2 = -\frac{1}{V_2}$ est équivalent au codage $T_1 = \frac{1}{n_1}$ et $T_2 = -\frac{1}{n_2}$, où n_1 et n_2 désignent le nombre de parents hétérozygotes d'enfants atteints de phénotype D_1 et D_2 (Démonstration 2).

$\mathbf{FBAT}_{\text{het}}$ peut se calculer directement dans le logiciel FBAT en utilisant l'option offset “-o” en codant $T_1 = 1$ et $T_2 = 0$: le logiciel calcule, pour chaque allèle m , un offset μ_m (cf

paragraphe introductif de FBAT) pour transformer les phénotypes : $T_1 = 1 - \mu_m$ et $T_2 = -\mu_m$ qui minimisent la variance de la statistique. Nous montrons que cette option est équivalente au codage $T_1 = \frac{1}{V_1}$ et $T_2 = -\frac{1}{V_2}$, permettant de tester un phénotype secondaire binaire (Démonstration 3). Dans ce cas particulier, il ne faut pas coder les individus non atteints par 0 mais comme manquants pour éviter qu'ils ne contribuent à la statistique.

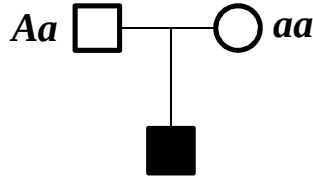
5.3.2 $FBAT_{het\ free}$ (modèle génétique additif)

Une contrainte implicite de $FBAT_{het}$ est que sous l'hypothèse nulle, les probabilités de transmission dans chacun des deux groupes est de 0,5. Bien que valable et efficace dans un certain nombre de situations pratiques, cet implicite peut avoir un impact sur la puissance du test d'association pour un résultat secondaire.

Nous proposons, dans le seul cadre d'un modèle génétique additif, une nouvelle statistique $FBAT_{het\ free}$ qui s'affranchit de cette contrainte d'une transmission de 0,5 sous l'hypothèse nulle. Considérons un locus biallélique (A et a) et notons n_{A1} (n_{A2}) le nombre de transmissions de l'allèle A de parents hétérozygotes Aa vers leurs enfants D_1 (D_2). Alors $\frac{n_{A1} + n_{A2}}{n_1 + n_2} = \frac{n_A}{N}$ représente la transmission moyenne de l'allèle A des parents hétérozygotes Aa aux enfants atteints (qu'ils soient D_1 ou D_2).

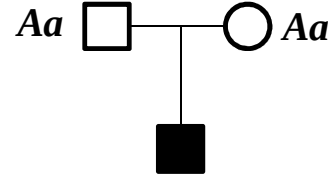
Sous l'hypothèse nulle d'absence association, alors que pour FBAT et $FBAT_{het}$ la transmission attendue de l'allèle d'intérêt est de 0,5, pour $FBAT_{het\ free}$, elle vaut $\frac{n_A}{N}$. Nous pouvons calculer S , E et V pour FBAT, $FBAT_{het}$ et $FBAT_{het\ free}$. La contribution à $(S - E)$ de chaque transmission d'un allèle A à partir d'un parent AA est de $\frac{1}{2}$ pour FBAT et $FBAT_{het}$, et de $\frac{n_A}{N}$ pour $FBAT_{het\ free}$. De même, la contribution à V de chaque transmission est de $\frac{1}{4}$ pour FBAT et $FBAT_{het}$, et de $\left(1 - \frac{n_A}{N}\right)\frac{n_A}{N}$ pour $FBAT_{het\ free}$ (Figure 26). Remarquons que pour les trois statistiques, l'espérance et la variance d'un trio dont les deux parents sont hétérozygotes sont doubles de celles d'un trio avec un seul parent hétérozygote. De manière symétrique, les parents hétérozygotes Aa qui transmettent l'allèle a contribuent chacun pour $\frac{1}{2}$ et $\left(1 - \frac{n_A}{N}\right)$ à $(S - E)$ et pour $\frac{1}{4}$ et $\left(1 - \frac{n_A}{N}\right)\frac{n_A}{N}$ à V pour FBAT ou $FBAT_{het}$ et $FBAT_{het\ free}$.

Trio avec 1 parent hétérozygote



Génotype de l'enfant		aa	Aa	AA
	X_i	0	1	2
FBAT & FBAT _{het}	$p(X_i)$	$\frac{1}{2}$	$\frac{1}{2}$	0
	$X_i - E(X)$	$-\frac{1}{2}$	$\frac{1}{2}$	
FBAT _{het free}	$p(X_i)$	$1 - \frac{n_A}{N}$	$\frac{n_A}{N}$	0
	$X_i - E(X)$	$-\frac{n_A}{N}$	$1 - \frac{n_A}{N}$	

Trio avec 2 parents hétérozygotes



Génotype de l'enfant		aa	Aa	AA
	X_i	0	1	2
FBAT & FBAT _{het}	$p(X_i)$	$\frac{1}{4}$	$\frac{1}{2}$	$\frac{1}{4}$
	$X_i - E(X)$	-1	0	1
FBAT _{het free}	$p(X_i)$	$\left(1 - \frac{n_A}{N}\right)^2$	$\left(1 - \frac{n_A}{N}\right)\frac{2n_A}{N}$	$\left(\frac{n_A}{N}\right)^2$
	$X_i - E(X)$	$-2\frac{n_A}{N}$	$1 - 2\frac{n_A}{N}$	$2\left(1 - \frac{n_A}{N}\right)$

Pour FBAT/FBAT_{hets} $E(X) = \frac{1}{2}$

$$\text{Var}(X) = \frac{1}{4}$$

Pour FBAT_{het free}, $E(X) = \frac{n_A}{N}$

$$\text{Var}(X) = \left(1 - \frac{n_A}{N}\right)\frac{n_A}{N}$$

$E(X) = 1$

$$\text{Var}(X) = \frac{1}{2}$$

$E(X) = 2\frac{n_A}{N}$

$$\text{Var}(X) = 2\left(1 - \frac{n_A}{N}\right)\frac{n_A}{N}$$

Figure 26. Contribution d'un trio à FBAT, FBAT_{het} et FBAT_{het free}

Dans un trio comprenant 1 (partie gauche) ou 2 (partie droite) parent(s) hétérozygote(s), les probabilités des génotypes aa, Aa et AA des enfants dépendent de la statistique considérée. Pour FBAT et FBAT_{het}, la probabilité de transmission d'un allèle A d'un parent hétérozygote est de $\frac{1}{2}$ alors qu'elle est de $\frac{n_A}{N}$ pour FBAT_{het free} (où N et n_A correspondent au nombre total d'allèles et au nombre d'allèles A transmis dans l'échantillon). La situation d'un couple Aa x aa est symétrique à celle d'un couple Aa x AA.

Puis, avec $T_1 = \frac{1}{n_1}$ et $T_2 = -\frac{1}{n_2}$, nous avons :

$$\begin{aligned} \mathbf{FBAT}_{\text{het free}} &= \frac{\left(\frac{n_{A1}}{n_1} \left(1 - \frac{n_A}{N}\right) + \frac{n_{a1}}{n_1} \left(0 - \frac{n_A}{N}\right) - \frac{n_{A2}}{n_2} \left(1 - \frac{n_A}{N}\right) - \frac{n_{a2}}{n_2} \left(0 - \frac{n_A}{N}\right) \right)^2}{\frac{n_1}{n_1^2} \frac{n_A n_a}{N^2} + \frac{n_2}{n_2^2} \frac{n_A n_a}{N^2}} \\ &= \frac{N}{n_1 n_2 n_A n_a} (n_{A1} n_{a2} - n_{a1} n_{A2})^2 \end{aligned}$$

Nous montrons que $\mathbf{FBAT}_{\text{het free}}$ est un test du Chi2 de Pearson (Démonstration 4). Enfin, les hypothèses du test $\mathbf{FBAT}_{\text{het free}}$ s'écrivent :

$$\mathbf{H}_0 : \mathbf{p}_1 = \mathbf{p}_2$$

$$\mathbf{H}_1 : \mathbf{p}_1 \neq \mathbf{p}_2$$

On remarque que contrairement à FBAT et $\mathbf{FBAT}_{\text{het}}$, la contrainte implicite d'une transmission allélique moyenne de 0,5 sous l'hypothèse nulle a disparu.

Une procédure d'implémentation du test $\mathbf{FBAT}_{\text{het free}}$ à l'aide des logiciels PLINK et R (Purcell et al., 2007; R Development Core Team, 2008) est présentée en annexe (Protocole).

5.4 Comparaison entre $\mathbf{FBAT}_{+1/-1}$, $\mathbf{FBAT}_{\text{het}}$ et $\mathbf{FBAT}_{\text{het free}}$

5.4.1 Comparaison entre $\mathbf{FBAT}_{\text{het}}$ et $\mathbf{FBAT}_{\text{het free}}$

La comparaison analytique des tests $\mathbf{FBAT}_{\text{het}}$ et $\mathbf{FBAT}_{\text{het free}}$ présentée dans la démonstration 5 établit que :

$$\mathbf{FBAT}_{\text{het}} = \rho \mathbf{FBAT}_{\text{het free}}$$

$$\text{avec } \rho = \frac{4n_A(N-n_A)}{N^2} \quad \rho \in [0,1]$$

Les propriétés (robustesse et puissance) des deux tests peuvent donc être directement dérivées. La distribution de ρ selon nA/N est illustrée dans la Figure 27. A titre d'exemple, considérons un échantillon de 300 trios (150 de phénotype D_1 et 150 de phénotype D_2), avec dans chaque trio un seul parent hétérozygote, et des transmissions moyennes de l'allèle A respectivement de 0,7 et 0,8 pour les enfants D_1 et D_2 . Il s'ensuit que $p = 0,75$, $\rho = 0,75$, $FBAT_{het} = 3$ ($p = 0,083$) et $FBAT_{het free} = 4$ ($p = 0,046$).

Lorsque la transmission moyenne de l'allèle A entre tous les individus atteints (D_1 ou D_2) est de 0,5, $n_A = n_a = -$ et $\rho = 1$. Dans ce cas particulier, $FBAT_{het} = FBAT_{het free}$. Dans tous les autres cas, $\rho < 1$ et $FBAT_{het} < FBAT_{het free}$, i.e. $FBAT_{het free}$ est le test le plus puissant. Ceci est illustré dans la Figure 28.

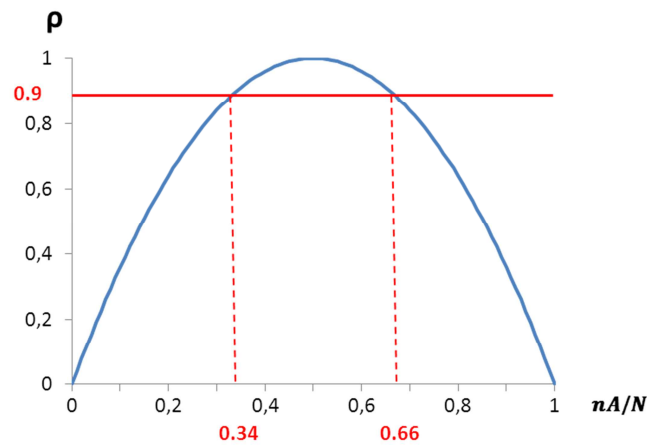


Figure 27. Distribution de ρ selon nA/N

La fonction de lien entre $FBAT_{het}$ et $FBAT_{het free}$ est $\rho = 4n_A(N-n_A)/N^2$. La ligne rouge $y = 0,9$ coupe la courbe de ρ pour les valeurs de $n_A/N = 0,34$ et $0,66$.

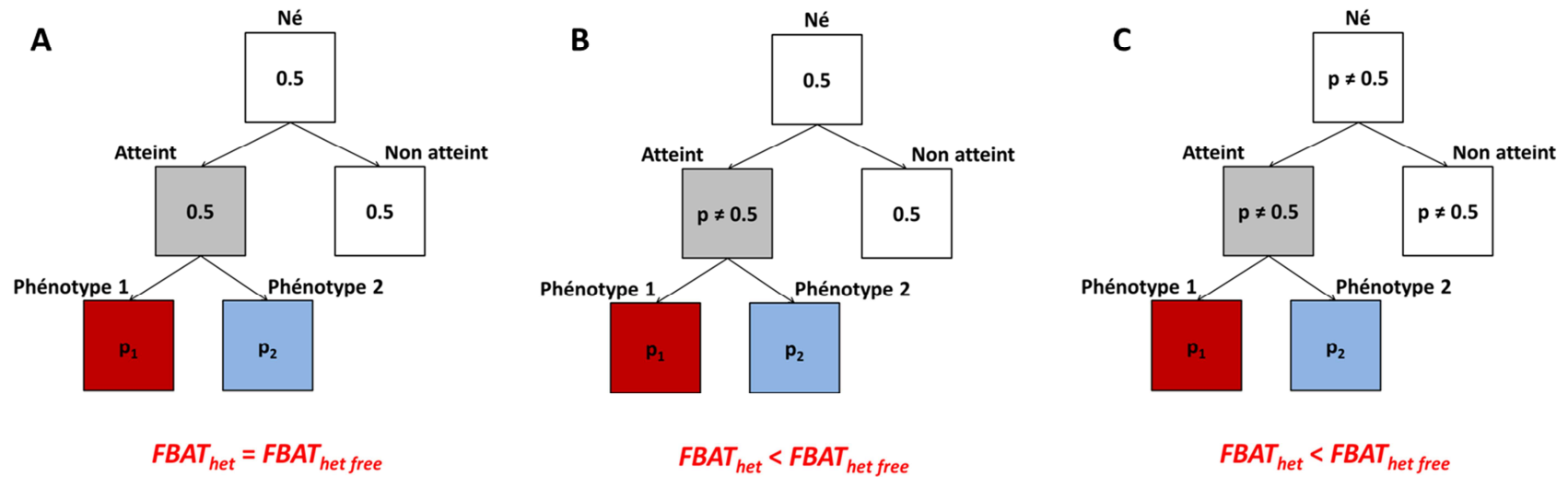


Figure 28. Puissances comparées de $FBAT_{het}$ et $FBAT_{het\ free}$ sous un modèle additif

Les chiffres désignent les transmissions moyennes de l'allèle considéré des parents hétérozygotes vers les enfants nés, atteints ou non-atteints ; p_1 et p_2 les transmissions moyennes vers les enfants de phénotype 1 et 2. Lorsque la transmission moyenne chez les atteints est égale à 0,5, $FBAT_{het}$ et $FBAT_{het\ free}$ sont égaux (panneau A). La transmission aux enfants atteints peut être différente de 0,5 si l'allèle est associé à la maladie (panneau B) ou associé à la survie à la naissance (panneau C). Dans ces deux cas, $FBAT_{het}$ est moins puissant que $FBAT_{het\ free}$.

5.4.2 Autre écriture des tests

Sous un modèle additif, en notant $p_1 = \frac{n_{A1}}{n_1}$, $p_2 = \frac{n_{A2}}{n_2}$ et $p = \frac{n_{A1} + n_{A2}}{n_1 + n_2} = \frac{n_A}{n}$ les transmissions moyennes observées de l'allèle A des parents hétérozygotes aux enfants présentant un phénotype D_1 , D_2 ou (D_1 ou D_2), les différents tests s'écrivent comme suit :

$$FBAT_{+1/-1} = \frac{(n_{A1} - n_{a1} - n_{A2} + n_{a2})^2}{\frac{1}{4}(n_{A1} + n_{a1} + n_{A2} + n_{a2})} = \frac{4}{N} \left(n_1 \left(p_1 - \frac{1}{2} \right) - n_2 \left(p_2 - \frac{1}{2} \right) \right)^2$$

$$\begin{aligned} FBAT_{het\ free} &= \frac{N}{n_1 n_2 n_A n_a} (n_{A1} n_{a2} - n_{a1} n_{A2})^2 \\ &= \frac{N}{n_1 n_2 p N (1-p) N} (n_1 p_1 n_2 (1-p_2) - n_1 (1-p_1) n_2 p_2)^2 \\ &= \frac{(n_1 n_2)^2}{n_1 n_2 p (1-p) N} (p_1 (1-p_2) - p (1-p))^2 = \frac{n_1 n_2}{p (1-p) N} (p_1 - p_2)^2 \end{aligned}$$

$$FBAT_{het} = \frac{4n_A(N - n_A)}{N^2} FBAT_{het\ free} = 4p(1-p) FBAT_{het\ free} = \frac{4n_1 n_2}{N} (p_1 - p_2)^2$$

L'hypothèse implicite de FBAT ($p_1 = p_2 = 0,5$) apparaît clairement dans la formulation de la statistique $FBAT_{+1/-1}$. Elle disparaît dans la formulation de $FBAT_{het}$ et de $FBAT_{het\ free}$. La valeur de la statistique $FBAT_{het}$ est indépendante de la transmission moyenne p , à la différence de $FBAT_{het\ free}$: un même écart de transmission entre les deux groupes D_1 et D_2 sera d'autant plus facilement décelé que p s'écarte de 0,5, puisque la fonction $\frac{1}{p(1-p)}$ prend son minimum en 0,5. Enfin, les deux statistiques $FBAT_{het}$ et $FBAT_{het\ free}$ sont sensibles aux tailles respectives des deux échantillons : les tests seront d'autant plus puissants que $n_1 = n_2$, qui maximise la fonction $\frac{4n_1 n_2}{N}$.

5.4.3 Comparaison entre $FBAT_{+1/-1}$ et $FBAT_{het\ free}$

Nous ne comparerons $FBAT_{+1/-1}$ et $FBAT_{het\ free}$ que dans le cas particulier où ces deux tests partagent les mêmes hypothèses nulle et alternative. Ceci est le cas lorsque les effectifs des deux sous-groupes sont égaux ($n_1 = n_2$). En effet, nous avons alors :

$$n_1 = n_2 = \frac{N}{2} \quad , \quad p = \frac{p_1 + p_2}{2}$$

Et les écritures de S_1 et S_2 se simplifient en :

$$S_1 = \sum_{i \in 1} 1 \times X_i = n_{A1} = p_1 n_1 = \frac{N}{2} p_1$$

$$S_2 = \sum_{i \in 2} 1 \times X_i = n_{A1} = p_2 n_2 = \frac{N}{2} p_2$$

Les hypothèses de $\text{FBAT}_{+1/-1}$ ($H_0 : S_1 = S_2 = 0,5$ et $H_1 : S_1 \neq S_2$) deviennent alors :

$$\mathbf{H_0 : } p_1 = p_2 = 0,5$$

$$\mathbf{H_1 : } p_1 \neq p_2$$

qui sont aussi celles de $\text{FBAT}_{\text{het free}}$.

En utilisant la nouvelle écriture des tests, nous avons :

$$\begin{aligned} \frac{\text{FBAT}_{+1/-1}}{\text{FBAT}_{\text{het free}}} &= \frac{\frac{4}{N} \left(n_1 \left(p_1 - \frac{1}{2} \right) - n_2 \left(p_2 - \frac{1}{2} \right) \right)^2}{\frac{n_1 n_2}{p(1-p)N} (p_1 - p_2)^2} \\ &= \frac{4p(1-p)}{n_1 n_2} \frac{\left(n_1 \left(p_1 - \frac{1}{2} \right) - n_2 \left(p_2 - \frac{1}{2} \right) \right)^2}{(p_1 - p_2)^2} \\ &= \frac{4p(1-p)}{\frac{N}{2} \frac{N}{2}} \frac{\left(\frac{N}{2} \left(p_1 - \frac{1}{2} \right) - \frac{N}{2} \left(p_2 - \frac{1}{2} \right) \right)^2}{(p_1 - p_2)^2} \\ &= \frac{4p(1-p)}{\frac{N}{2} \frac{N}{2}} \frac{\left(\frac{N}{2} \right)^2 (p_1 - p_2)^2}{(p_1 - p_2)^2} \\ &= 4p(1-p) = \rho \\ &\quad \rho < 1 \quad \forall p \neq \frac{1}{2} \end{aligned}$$

Autrement dit, lorsque les deux sous-groupes sont de même taille, $FBAT_{het\ free}$ est systématiquement « plus puissant » que $FBAT_{+1/-1}$. Et dans ce cas, les deux tests sont égaux si la transmission moyenne de l'allèle A aux deux sous-groupes est égale à 0,5.

5.5 Discussion

Les études d'association avec données familiales ont gagné en popularité pour disséquer l'architecture génétique des caractères complexes. FBAT est probablement l'outil le plus populaire pour effectuer de telles études. Nous avons montré qu'à première vue, FBAT pouvait être facilement utilisé pour tester les résultats secondaires, comme l'hétérogénéité génétique entre les formes sévères ou non d'une maladie. A titre d'exemple, dans un échantillon de trios, on peut pondérer chaque sous-phénotype (non sévère et sévère) par l'inverse de la variance de chaque statistique. Nous avons appelé cette statistique $FBAT_{het}$, pour laquelle les hypothèses nulle et alternative sont $H_0 : p_1 = p_2 = 1/2$ et $H_1 : p_1 \neq p_2$.

Cependant, dans le test précédent, les probabilités de transmission sous l'hypothèse nulle sont fixées à 0,5 dans les deux groupes. Cela peut ne pas être optimal lorsque la transmission de l'allèle testé diffère déjà significativement de 0,5 parmi les individus atteints. Nous montrons qu'il est possible d'assouplir cette contrainte en modifiant l'espérance de la statistique $FBAT_{het}$ pour que les hypothèses nulle et alternative deviennent $H_0 : p_1 = p_2$ et $H_1 : p_1 \neq p_2$, qui sont les hypothèses classiques des tests d'homogénéité. Ce nouveau test, $FBAT_{het\ free}$, s'avère être égal à un test du χ^2 d'homogénéité de Pearson. $FBAT_{het\ free}$ est le test le plus puissant lorsque la transmission de moyenne aux enfants touchés ($D_1 + D_2$) diffère de 0,5. Autrement dit, à chaque fois un allèle est déjà associé à la maladie, $FBAT_{het\ free}$ sera plus puissant pour détecter l'hétérogénéité entre les transmissions de cet allèle aux phénotypes secondaires.

Nos développements présentent toutefois des limitations. La première est la restriction aux marqueurs bi-alléliques ($FBAT_{het}$ et $FBAT_{het\ free}$). Elle semble la moins importante compte tenu que les SNPs bi-alléliques constituent actuellement le principal matériel génétique utilisé en épidémiologie génétique. Il sera toujours possible de tester un marqueur à n allèles par $n-1$ tests bi-alléliques, même si cette stratégie sera peu puissante si l'on tient compte de la correction pour les tests multiples.

La deuxième limitation est la restriction aux trios (FBAT_{het} et $\text{FBAT}_{\text{het free}}$). On notera toutefois que le logiciel FBAT ne permet pas d'utiliser simultanément les options « -e » et « -o ». Certaines études mentionnent l'usage de l'option « -o », qui permet d'augmenter la puissance, et n'utilisent donc *a priori* pas l'option « -e », même lorsque l'échantillon est principalement constitué de familles multiplex (≥ 2 enfants) (Yang et al., 2006; Jonas et al., 2008; Thakur et al., 2012). La véritable hypothèse nulle rejetée est alors celle d'une absence de liaison ou d'association, et non pas l'absence d'association en présence de liaison.

Enfin, la troisième limitation est la restriction au seul modèle additif (pour $\text{FBAT}_{\text{het free}}$ seulement). Notre intuition est que la non-linéarité de la variance en fonction du nombre de transmissions alléliques sous les modèles récessif et dominant empêche de généraliser l'approche développée pour le modèle additif sans toutefois parvenir à le démontrer analytiquement. En pratique, un grand nombre de GWAS ne teste que le modèle additif ; la plupart des SNPs pour lesquels le meilleur modèle serait dominant sont aussi significatifs avec le modèle additif ; mais ce n'est pas le cas du modèle récessif ni de celui qui donne l'avantage aux hétérozygotes. Le modèle récessif est souvent limité en puissance par le nombre d'individus homozygotes pour l'allèle rare ; et l'avantage aux hétérozygotes n'est qu'exceptionnellement rapporté dans les études (Vincent, 2014).

Résultats de l'étude d'association

6.1 Choix des marqueurs

Le contrôle de qualité des génotypes a été réalisé en utilisant le logiciel PLINK (Purcell et al., 2007). 2 564 individus ont été génotypés. Après avoir retiré les individus pour lesquels moins de 95 % des SNPs avaient été validés par le logiciel, puis les duplicats et les triplicats, il restait 2 235 individus. Nous avons exclu les SNPs qui avaient les caractéristiques suivantes : taux de génotypage < 99 % individus, valeur de p pour le test de Hardy-Weinberg < 10^{-5} , fréquence de l'allèle mineur (MAF) < 0,01, présence de 10 erreurs mendéliennes ou plus. Ces filtres ont été appliqués successivement selon l'algorithme de la figure 29, qui donne le nombre de SNPs exclus à chaque étape. L'équilibre de Hardy Weinberg a été testé sur les fondateurs non atteints (N = 991) Nous avons ensuite exclu les SNPs mitochondriaux ou des chromosomes sexuels, nous restreignant ainsi aux chromosomes 1 à 22.

La stratégie de sélectionner les individus avant les SNPs est recommandée (Anderson et al., 2010). Cette approche évite que des marqueurs soient exclus du fait de ne pas avoir été appelés chez quelques individus de mauvaise qualité. Chaque SNP retiré constitue un possible faux négatif pour l'association avec le phénotype d'intérêt, tandis qu'un individu retiré ne

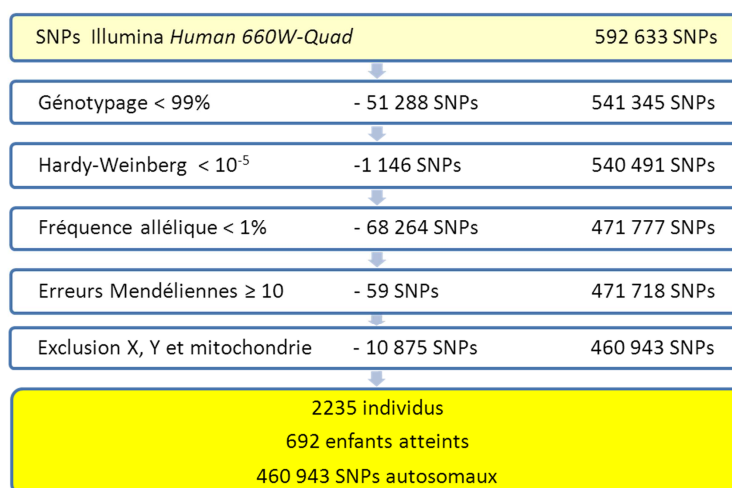


Figure 29. Critères d'exclusion pour la sélection des SNPs

modifie que marginalement la valeur du test d'association. Il n'y a pas de consensus pour les seuils des filtres. Plus stricts, ils diminuent le risque de faux positifs, plus lâches, ils diminuent le risque de faux négatifs. Nos valeurs de filtres (génotypage < 99 %, Hardy Weinberg < 10^{-5} , MAF < 0,01) sont dans la moyenne de celles rapportées (Anderson et al., 2010).

Au total, nous disposons de 460 943 SNPs sur les autosomes chez 2 235 individus, dont 692 enfants atteints. Parmi ceux-ci, 315 étaient PB et 363 MB selon la classification OMS-82.

6.2 Analyse pan-génomique primaire

6.2.1 Analyse de la polarisation (définition OMS-82)

Les analyses que nous présentons ne prennent pas en compte, par construction, le phénotype des parents. L'association de chaque SNP avec le phénotype est testée indépendamment des autres SNPs. Les résultats sont illustrés par deux types de graphiques : le « Manhattan plot » et le « QQ plot », tous deux obtenus avec R (Turner, 2014). Dans le Manhattan plot, l'axe des abscisses représente la position sur le génome (du chromosome 1 à 22) et l'axe des ordonnées $-\log_{10}(p)$, où p désigne le résultat du test d'association. Ce graphique permet de faire ressortir visuellement les SNPs les plus significatifs, qui apparaîtront en haut sur le graphique, parmi l'ensemble des 460 943 SNPs. L'observation d'un « gratte-ciel » (ou ligne verticale de points) évoquera une association significative de SNPs proches sur le génome. Le « Quantile-Quantile plot » ou QQ plot donne un aperçu immédiat de l'erreur de type 1 du test. Il compare les quantiles de la distribution théorique sous H_0 aux quantiles de la distribution observée de la statistique. Nous avons utilisé les statistiques implémentées dans FBAT, qui suivent toutes sous H_0 une loi du χ^2 à 1 degré de liberté.

Compte-tenu du grand nombre de SNPs et donc de tests réalisés, le seuil de significativité pour un SNP n'est pas égal à 0,05. Une correction de Bonferroni exigerait une valeur nominale de significativité de $0,05/460\,943 \sim 10^{-8}$, mais celle-ci est trop conservatrice. Compte-tenu du déséquilibre de liaison entre les SNPs, les tests ne sont pas tous indépendants. Le seuil de significativité à l'échelle du génome généralement retenu est de $5 \cdot 10^{-8}$, et ce quel que soit le nombre de SNPs étudiés (Jannot et al.). Ce seuil est figuré par une ligne horizontale rouge sur nos Manhattan plots. Une ligne bleue figure le seuil nominal évocateur d'association à l'échelle du génome, qui est de 10^{-5} .

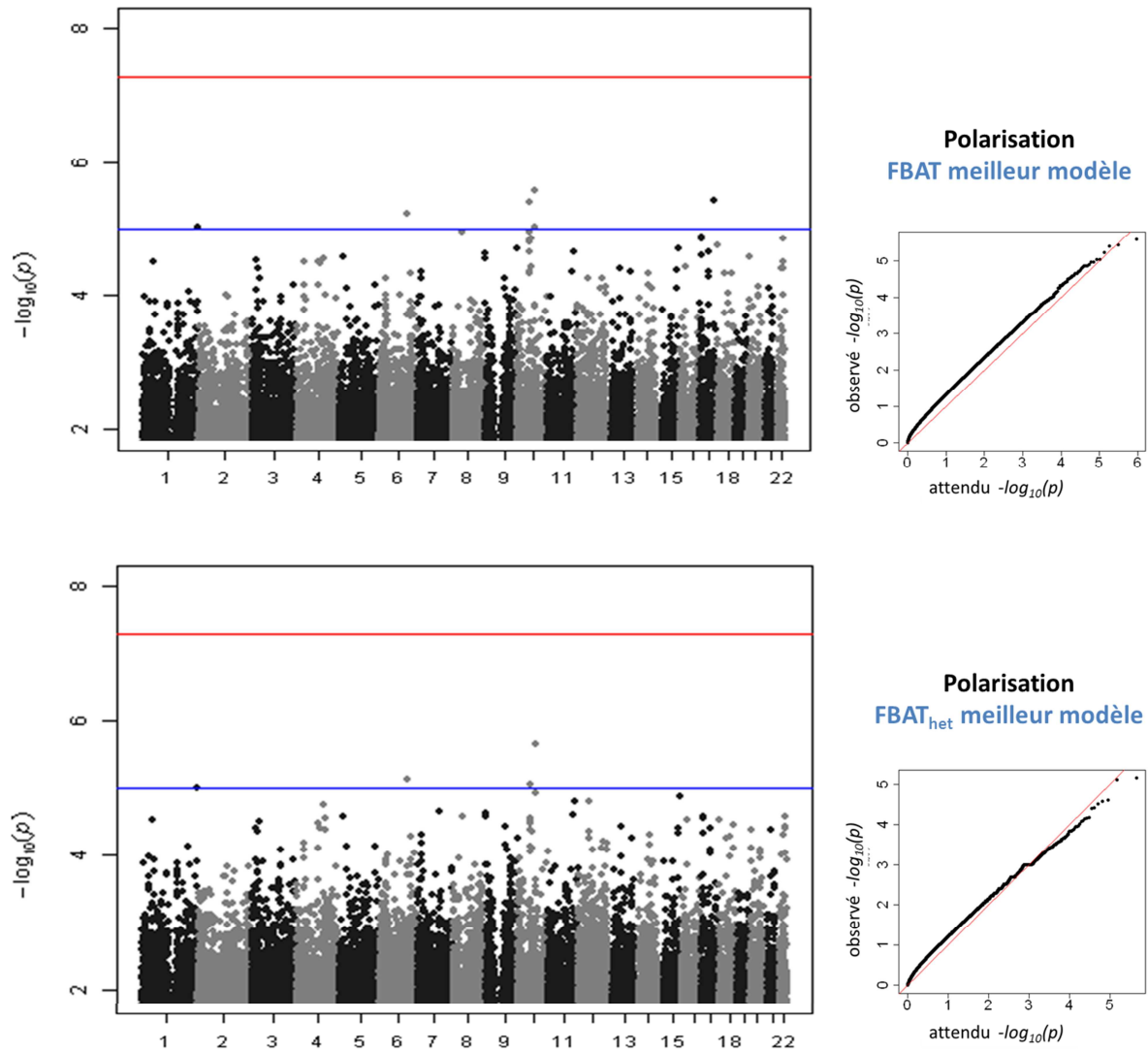


Figure 30. Association pour la polarisation selon l'OMS-82 avec *FBAT* et *FBAT_{het}* Manhattan et QQ-plots pour les modèles additif, dominant et récessif (de haut en bas).

Nous souhaitons analyser la polarisation, et non la forme PB ou la forme MB seule. Nous avons d'abord codé la polarisation comme un trait quantitatif, en attribuant une valeur positive (+1) aux enfants MB, et une valeur négative symétrique (-1) aux cas PB. C'est le test $FBAT_{+1/-1}$ (explicité au chapitre précédent), et que nous appellerons simplement *FBAT* ici. Puis nous avons utilisé le test $FBAT_{het}$, introduit chapitre précédent, qui pondère les poids des groupes PB et MB par l'inverse de leur « taille ». Cette taille ne correspond pas exactement au nombre d'enfants du groupe, mais au nombre de parents hétérozygotes disponibles dans les deux

groupes. Aussi cette pondération varie pour chaque SNP en fonction des répartition des allèles chez les parents. Les trois modèles additif (a), dominant (d) et récessif (r) ont été testés avec ces deux méthodes.

La figure 30 présente les résultats de l'association pour la polarisation (définition OMS-82) avec les méthodes FBAT et FBAT_{het}. Les valeurs de p représentées sont celles obtenues pour le meilleur des trois modèles a, d et r. Aussi, les valeurs du QQ-plot sont au-dessus de la droite d'équation $y = x$. Aucun SNP n'a atteint le seuil significatif d'association. Seuls 6 SNPs ont atteint le seuil suggérant une association avec FBAT : rs3811445 sur le chromosome 1 (modèle r, $p = 9 \cdot 10^{-6}$), rs781484 (a, $p = 6,05 \cdot 10^{-6}$) sur le 6, rs1329202 (a, $p = 2,63 \cdot 10^{-6}$), rs42455 (d, $p = 9,19 \cdot 10^{-6}$) et rs1903942 (a, $p = 3,95 \cdot 10^{-6}$) sur le 10 et rs1034764 (d, $3,67 \cdot 10^{-6}$) sur le 17. Quatre des 6 SNPs ont aussi atteint le seuil de 10^{-5} avec FBAT_{het}, avec le même modèle que pour FBAT, avec à chaque fois des valeurs de p plus grandes. Nous avons tronqué les Manhattan plot aux valeurs supérieures à 10^{-2} pour alléger les figures : cela concernait plus de 400 000 SNPs.

La figure 31 présente les résultats obtenus avec FBAT_{het free} sous ceux obtenus avec FBAT et FBAT_{het} sous le modèle additif, puisque FBAT_{het free} n'existe que sous ce modèle génétique. FBAT_{het free} n'a identifié aucun SNP significatif mais 5 évocateurs d'association avec la polarisation, tous différents de ceux trouvés avec FBAT ou FBAT_{het} : rs10914232 sur le chromosome 1 ($p = 7,20 \cdot 10^{-6}$), rs1021377 sur le 4 ($p = 8,61 \cdot 10^{-6}$), rs1140809 sur le 6 ($p = 2,80 \cdot 10^{-6}$), rs2884739 sur le 14 ($p = 5,06 \cdot 10^{-6}$) et rs4622543 sur le 17 ($p = 7,92 \cdot 10^{-6}$).

6.2.2 Comparaison des associations de la polarisation et des formes PB et MB

Notre phénotype d'intérêt était la polarisation. Afin d'illustrer que les signaux identifiés pour la polarisation peuvent être différents de ceux associés à l'une des formes cliniques, nous présentons, comme pour l'analyse de liaison, les résultats pour les formes PB et MB de la lèpre. Nous les comparons sur la figure 32 aux résultats de la polarisation, pour l'analyse avec FBAT. Si aucun SNP n'a atteint la significativité, une association est suggérée pour 16 et 6 SNPs avec les phénotypes MB (363 enfants) et PB (315 enfants). Aucun de ces SNPs n'est commun avec ceux associés à la polarisation. Pour la forme MB, 4 SNPs rapprochés dans la région du CMH sur le chromosome 6p21 ont des valeurs $< 10^{-5}$, et 15 une valeur $< 10^{-4}$, réalisant l'aspect de « gratte-ciel » évocateur d'association dans les études.

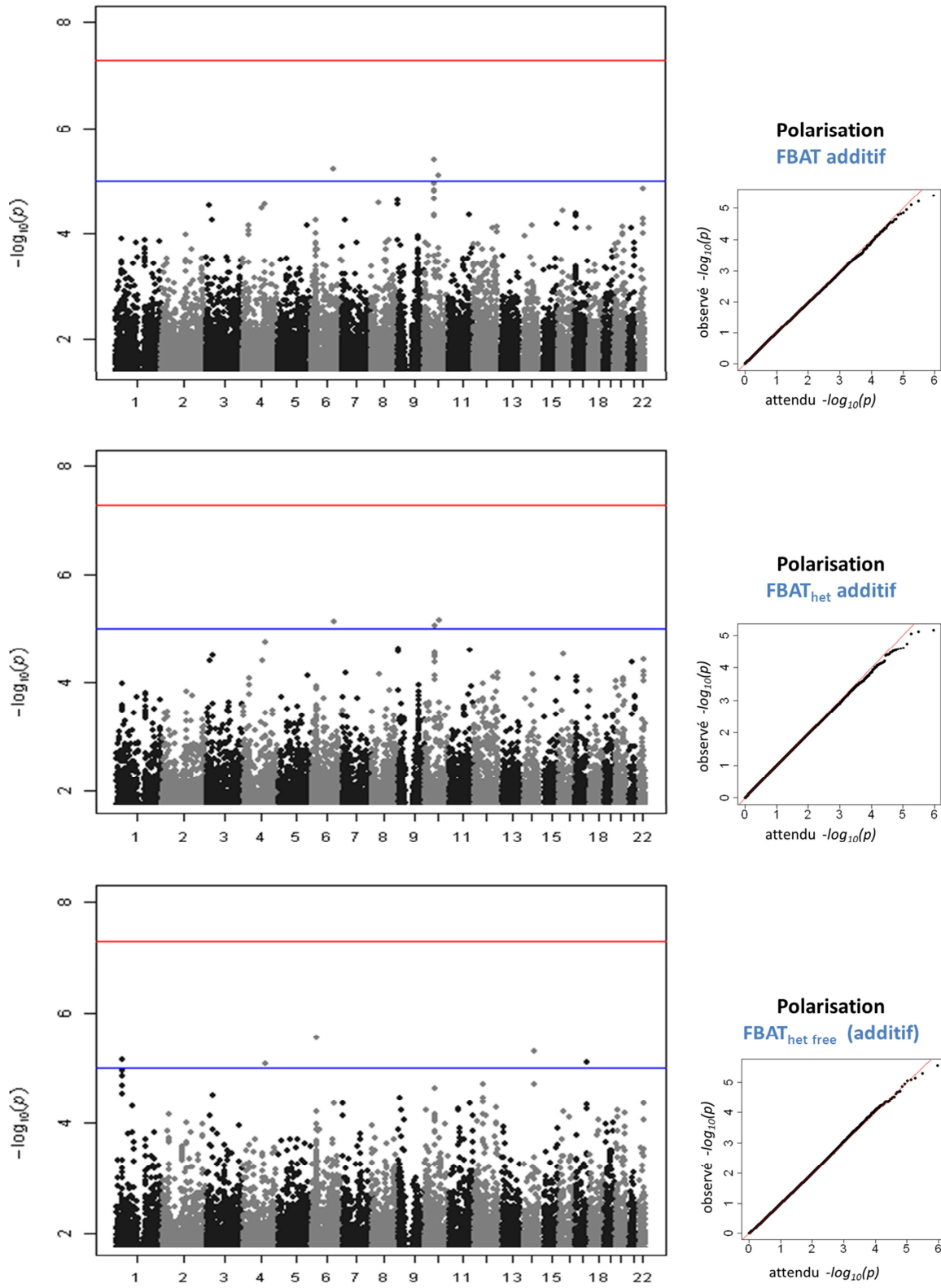


Figure 31. Association pour la polarisation selon l'OMS-82 avec FBAT_{het free} et comparaison avec les modèles additifs de FBAT et FBAT_{het}

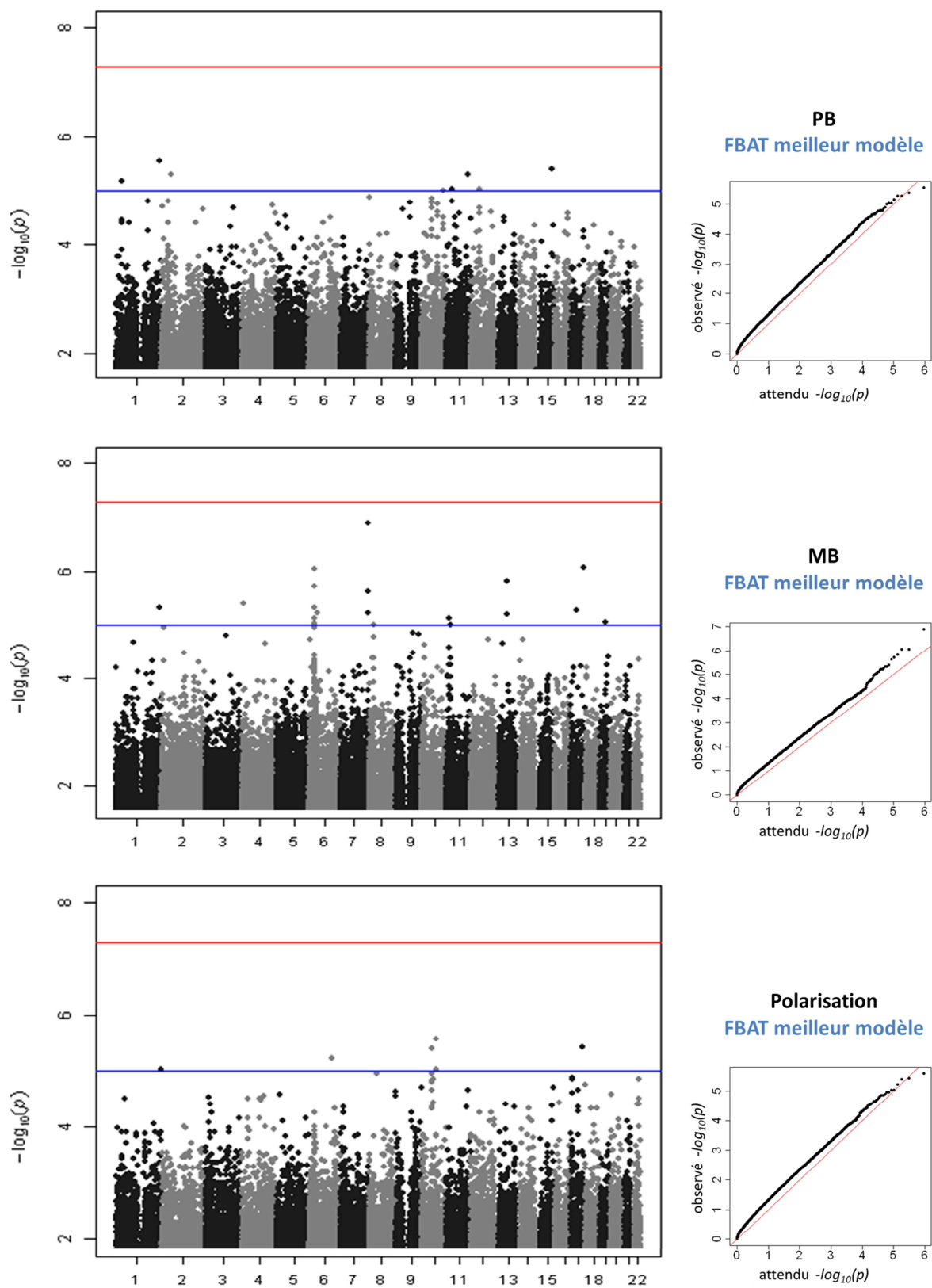


Figure 32. Comparaison des associations pour 3 phénotypes avec FBAT

6.2.3 Analyse de la polarisation en contrastant TT et LL

Nous n'avons pu identifier un marqueur significativement associé à la polarisation en comparant les groupes PB et MB selon la classification OMS-82. Nous avons ensuite augmenté le contraste de notre phénotype et comparé, toujours avec FBAT et FBAT_{het}, les groupes TT et LL (Figure 33). Ils comprenaient respectivement 123 et 45 enfants. Malgré une puissance statistique environ 4 fois moindre que pour la polarisation PB vs. MB (168 vs. 678 équivalents trios), une valeur de $p < 10^{-6}$ était obtenue sur le chromosome 3 avec FBAT_{het} (rs843354 modèle d, $p = 3,12 \cdot 10^{-7}$).

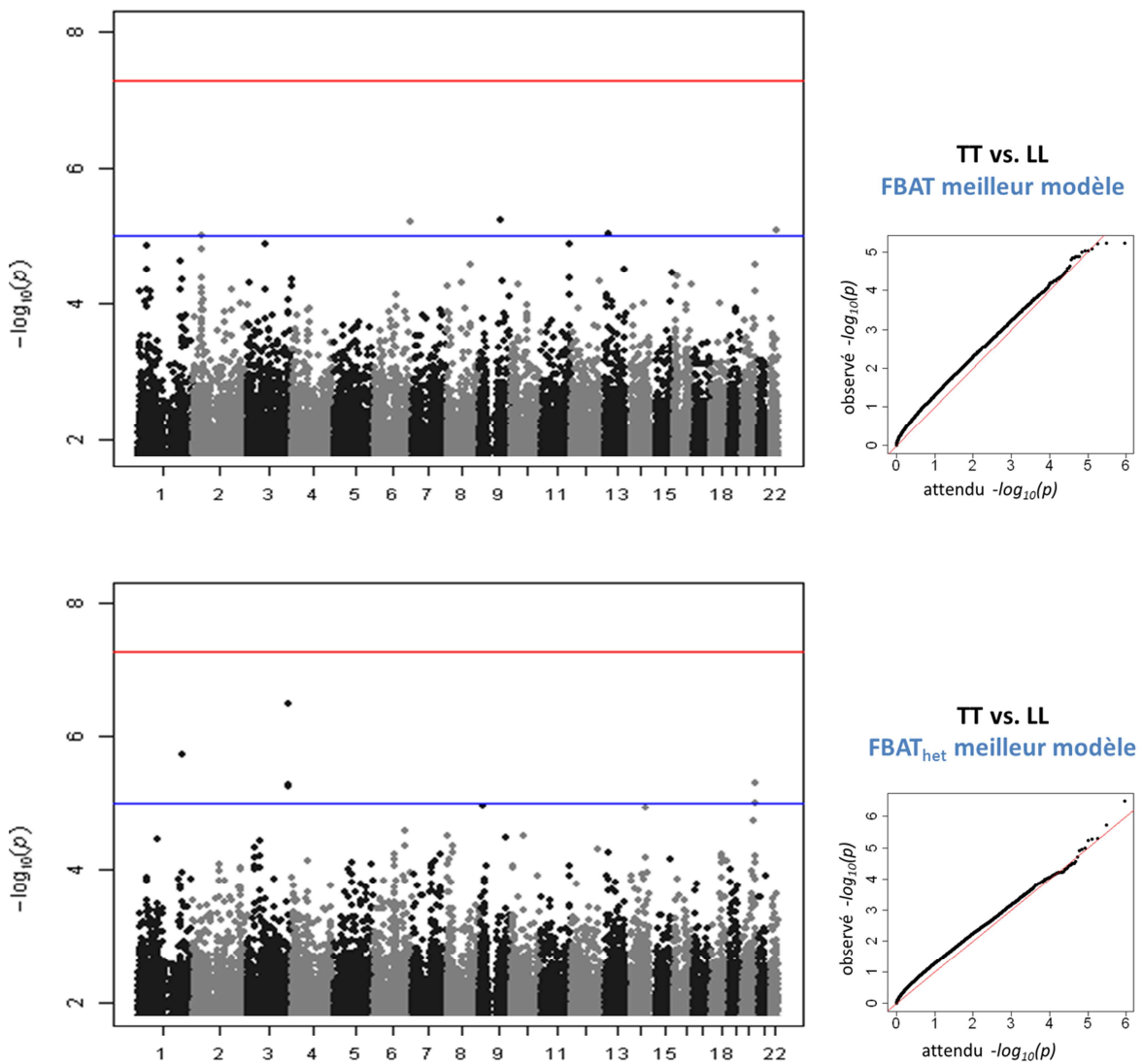


Figure 33. Association de la polarisation TT vs. LL avec FBAT et FBAT_{het} free

Ce SNP était en $LD > 0.5$ avec deux autres SNPs, rs843350 et rs902417, avec des valeurs de $p < 10^{-5}$ avec $FBAT_{het}$, toujours sous un modèle dominant ($p = 5,25 \cdot 10^{-7}$ et $5,69 \cdot 10^{-7}$, respectivement). Les MAF de ces SNPs étaient comprises entre 0,26 et 0,29.

6.2.4 Synthèse graphique des analyses de liaison et d'association

Nous synthétisons les résultats des analyses d'association et de liaison dans la figure 34. L'association a été testée pour 13 configurations : 3 modèles (a, d et r) pour 2 phénotypes (PB vs. MB et TT vs. LL) et pour 2 analyses (FBAT et $FBAT_{het}$), ainsi que le modèle additif pour PB vs. MB) avec $FBAT_{het free}$. Parmi les 243 valeurs de $p < 10^{-4}$, 101 étaient obtenues en additif, 85 en dominant et 57 en récessif. FBAT donnait la meilleure valeur 107 fois, $FBAT_{het}$ 94 fois et 42 $FBAT_{het free}$. Nous avons fait figurer le résultat de la liaison codée en 5 catégories selon Ridley et Jopling. Le tableau 11 donne plus de renseignements sur le 48 SNPs avec un $p < 2 \cdot 10^{-5}$ à l'échelle du génome.

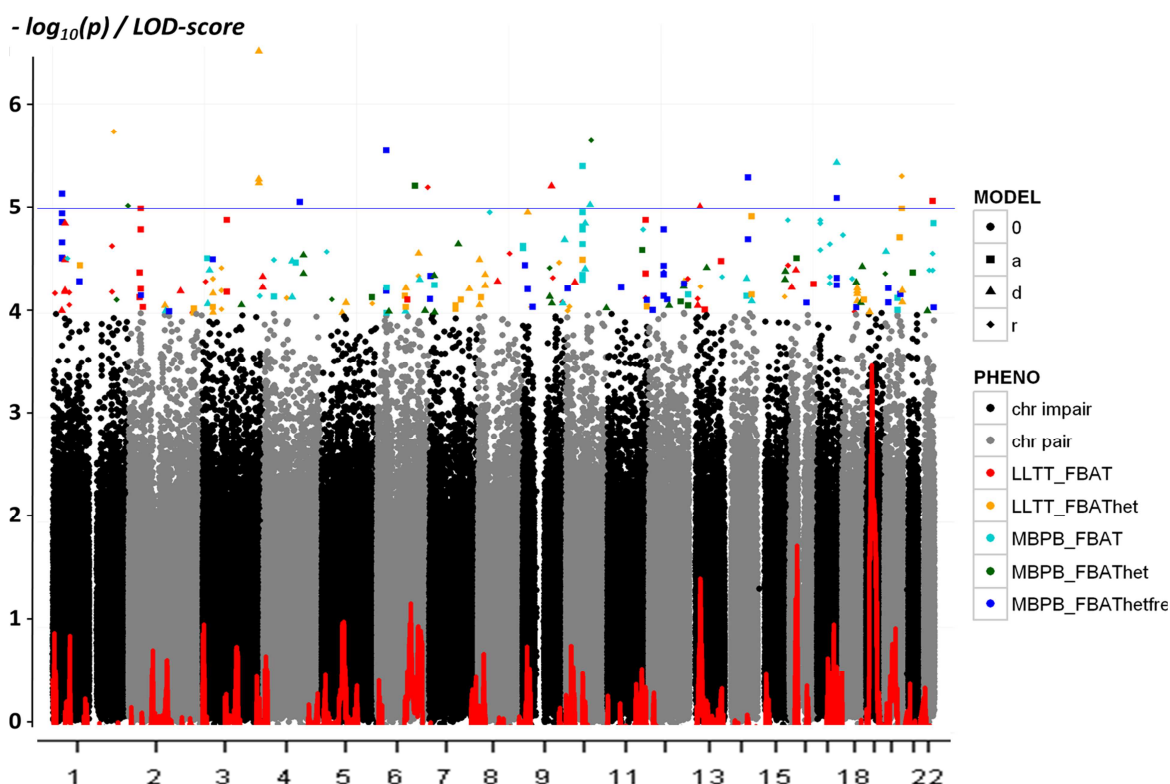


Figure 34. Synthèse des résultats de liaison et d'association pangénomique
LOD-score de la liaison (rouge) et Manhattan plot de l'association de la polarisation. La valeur de p correspond à la meilleure pour les tests FBAT, $FBAT_{het}$ et $FBAT_{het free}$ pour les phénotypes TT vs. LL et PB vs. MB. Le modèle et le phénotype sont précisés lorsque $p < 10^{-4}$.

La figure 35 zoome sur le chromosome 19 où une liaison a été identifiée. Seul 1 SNP (rs7256672, 19:10579474) a obtenu une valeur de $p < 10^{-4}$. ce SNP se situait en-dehors de l'intervalle de confiance à 95% du pic de liaison (entre 17,1 et 20,8 Mbs). Dans cet intervalle, seuls 5 SNPs avaient un $p < 10^{-3}$: rs17654580, rs2238675, rs1035064, rs2075090 et rs8111389. Parmi ceux-ci, rs17654580 (19:18128388) était à une distance de 40 kb de l'extrémité 3' du gène *IL12RB1*.

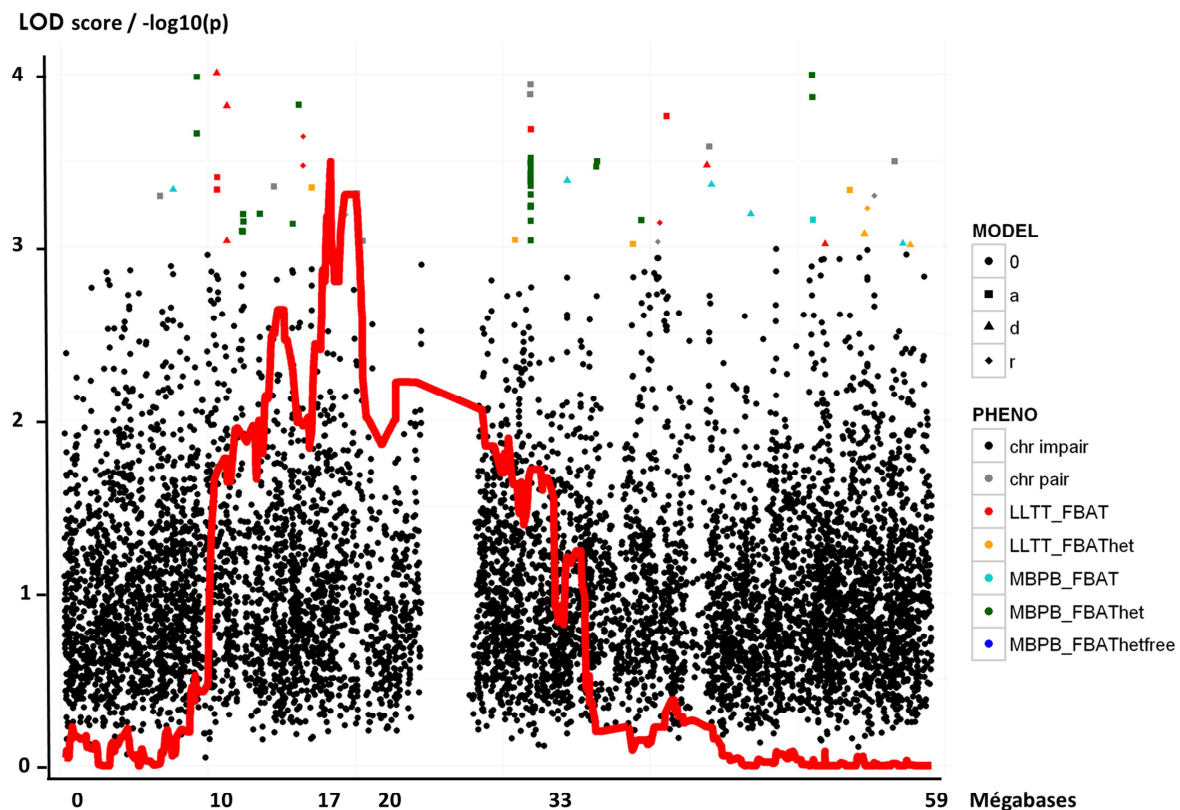


Figure 35. Synthèse des résultats de liaison et d'association sur le chromosome 19

LOD-score de la liaison (rouge) et Manhattan plot de l'association de la polarisation sur le chromosome 19. La valeur de p correspond à la meilleure pour les tests FBAT, $FBAT_{het}$ et $FBAT_{het free}$ pour les phénotypes TT vs. LL et PB vs. MB. Le modèle et le phénotype sont précisés lorsque $p < 0.001$.

Tableau 11. Meilleurs SNPs pour l'association avec la polarisation

Le tableau présente les SNPs dont une valeur de p pour l'association avec un des phénotypes de polarisation a été $< 2 \cdot 10^{-5}$. Les SNPs sont classés par ordre croissant de p . MAF fréquence de l'allèle mineur. La colonne "gène proche" indique le gène le plus proche du SNP (distance max 100 kb). Le phénotype et le modèle pour lesquels le meilleur p a été obtenus sont mentionnés. La colonne "LOD" indique le LOD-score de la liaison pour la polarisation selon Ridley-Jopling.

CHR	POSITION	SNP	A1	A2	MAF	gène proche	distance	PHENOTYPE	MODÈLE	LOD	p
3	184006395	rs843354	A	C	0,26	ECE2	0	LLTT_FBAThet	d	0,25	3,12E-07
1	201345487	rs10753241	T	C	0.3585	TNNT2	0	LLTT_FBAThet	r	0,00	1,84E-06
10	83188768	rs1329202	C	T	0,08			MBPB_FBAThet	r	0,00	2,24E-06
6	30611676	rs1140809	G	T	0,38	ATAT1	0	MBPB_FBAThetfree	a	0,00	2,80E-06
17	63726773	rs1034764	T	C	0,06	CEP112	0	MBPB_FBAT	d	0,56	3,67E-06
10	54924612	rs1903942	A	G	0,35			MBPB_FBAT	a	0,42	3,95E-06
20	56599408	rs2095295	A	G	0,19			LLTT_FBAThet	r	0,00	4,90E-06
14	75831121	rs2884739	G	A	0,49			MBPB_FBAThetfree	a	0,11	5,06E-06
3	183946274	rs843350	C	T	0,29	VWA5B2	[s:2043;e:13843]	LLTT_FBAThet	d	0,17	5,25E-06
3	183951431	rs902417	A	G	0,29	VWA5B2	0	LLTT_FBAThet	d	0,18	5,69E-06
6	124977655	rs781484	T	C	0,43	NKAIN2	0	MBPB_FBAThet	a	0,44	6,05E-06
9	94202927	rs11535536	T	C	0,30			LLTT_FBAT	d	0,00	6,11E-06
9	94207042	rs1106610	C	T	0,29			LLTT_FBAT	d	0,00	6,11E-06
6	167276965	rs9356531	T	C	0,13	RPS6KA2	[s:454111;e:1194]	LLTT_FBAT	r	0,00	6,26E-06
1	31388559	rs10753242	T	C	0.4827	SDC3	[s:46246;e:7079]	MBPB_FBAThetfree	a	0,00	7,20E-06
17	64510410	rs4622543	T	C	0,22	PRKCA	0	MBPB_FBAThetfree	a	0,30	7,92E-06
22	44940363	rs2142832	A	G	0,20			LLTT_FBAT	a	0,00	8,44E-06
4	119867342	rs1021377	T	C	0,29	SYNPO2	0	MBPB_FBAThetfree	a	0,00	8,61E-06
10	79162342	rs42455	C	T	0,13	KCNMA1	0	MBPB_FBAT	d	0,00	9,19E-06
1	248039713	rs10863849	T	C	0.4942	TRIM58	0	MBPB_FBAThet	r	0,00	9,34E-06
13	32886213	rs9534154	T	C	0,16	ZAR1L	[s:8305;e:122]	LLTT_FBAT	d	1,31	9,58E-06
13	32889363	rs3092989	A	G	0,16	ZAR1L	[s:11455;e:3272]	LLTT_FBAT	d	1,31	9,58E-06
2	40373872	rs444999	C	T	0,24	LOC100128590	0	LLTT_FBAT	a	0,01	1,00E-05
20	56594279	rs6099886	A	G	0,19			LLTT_FBAThet	a	0,00	1,00E-05
8	38813046	rs7007530	C	T	0,09	PLEKHA2	0	MBPB_FBAT	r	0,00	1,10E-05
9	16392810	rs7041706	C	T	0,15			LLTT_FBAThet	d	0,08	1,10E-05
10	54862114	rs1903971	A	G	0,36			MBPB_FBAT	a	0,38	1,10E-05
1	31566838	rs10914232	A	G	0.486	SEPW1P	[s:580;e:1647]	MBPB_FBAThetfree	a	0,00	1,13E-05
14	87320278	rs10498606	G	T	0,42			LLTT_FBAThet	a	0,00	1,20E-05
3	79264611	rs2077753	G	A	0,38	ROBO1	0	LLTT_FBAT	a	0,10	1,30E-05
11	125489621	rs540723	T	C	0,18	STT3A	0	LLTT_FBAT	a	0,32	1,30E-05
11	125613572	rs2217034	G	A	0,19	PATE1	[s:2616;e:6171]	LLTT_FBAT	a	0,33	1,30E-05
15	98166554	rs8039196	T	C	0,33			MBPB_FBAT	r	0,00	1,30E-05
17	10295497	rs2270056	T	C	0,14	MYH8	0	MBPB_FBAT	r	0,00	1,30E-05
1	31471702	rs11811369	C	A	0.4747	PUM1	0	MBPB_FBAThetfree	a	0,00	1,37E-05
1	40990560	rs12024378	A	G	0.4269	DEM1	[s:16127;e:8332]	LLTT_FBAT	d	0,20	1,40E-05
1	40993873	rs12042242	T	C	0.4257	ZNF684	[s:3360;e:19968]	LLTT_FBAT	d	0,20	1,40E-05
10	63023314	rs4145320	T	C	0,14			MBPB_FBAT	d	0,26	1,40E-05
17	10317940	rs4791408	A	G	0,14	MYH8	0	MBPB_FBAT	r	0,00	1,40E-05
22	48250736	rs2103957	C	T	0,06			MBPB_FBAT	a	0,00	1,40E-05
10	54883866	rs11003396	A	G	0,35			MBPB_FBAT	a	0,39	1,50E-05
2	40371095	rs7572256	G	A	0,23	LOC100128590	0	LLTT_FBAT	a	0,01	1,60E-05
10	54867617	rs1904056	G	A	0,36			MBPB_FBAT	a	0,38	1,60E-05
11	117595511	rs11216517	A	G	0,28	DSCAML1	0	MBPB_FBAT	r	0,03	1,60E-05
12	49308284	rs10747556	C	T	0,45	CCDC65	0	MBPB_FBAThetfree	a	0,00	1,60E-05
18	2461449	rs4594325	C	T	0,27			MBPB_FBAT	r	0,00	1,80E-05
20	50419395	rs6021451	C	A	0,20	SALL4	[s:18814;e:347]	LLTT_FBAThet	a	0,00	1,90E-05
14	75823156	rs1882830	G	A	0,46			MBPB_FBAThetfree	a	0,11	1,98E-05

6.2.5 Discussion

Nous n'avons pas identifié de variant significativement associé à la polarisation de la lèpre à l'échelle du génome, même en prenant la meilleure valeur des 13 tests d'associations réalisés. Ces résultats seraient encore moins significatifs si l'on corrigeait pour la réalisation des tests multiples, par un facteur 13 selon Bonferroni. Cependant, un facteur 13 serait bien trop conservateur. En effet, nos tests (FBAT, FBAT_{het} et FBAT_{het free}), nos phénotypes (PB vs. MB) et nos modèles (a, d et r) ne sont pas indépendants entre eux. Une manière d'approcher une valeur nominale de p conforme serait de calculer par permutations la distribution empirique. Nous ne l'avons pas fait, car notre stratégie va reposer sur la réplication de nos résultats. Aussi, la valeur exacte de la valeur de p à moins d'importance que son rang. L'objectif est en effet de tenter de répliquer les SNPs plus significatifs « quelle que soit cette significativité » de l'analyse primaire dans un échantillon indépendant. Nous réfléchissons actuellement, compte tenu du budget alloué, au nombre de SNPs que nous pourrions génotyper dans un échantillon vietnamien déjà recruté.

6.3 Projet de réplication

Nous chercherons à comparer des individus MB (que nous appellerons « cas ») à des individus PB (que nous appellerons « contrôles ») dans un échantillon vietnamien indépendant. Il a été constitué entre 2010 et 2013. Il comprend 1 182 cas de lèpre, dont le diagnostic est ancien pour la plupart. Cela n'a été possible que grâce à un archivage rigoureux par nos collaborateurs vietnamiens des dossiers de tous les patients traités. Les cas ne sont pas connus pour être apparentés, et tous sont déclarés être de l'ethnie *Kinh*.

Le grand avantage de notre phénotype pour une étude cas-témoin est qu'il n'a pas été nécessaire de recruter des contrôles « sains ». Il est en effet parfois difficile de recruter des contrôles ressemblant le plus possible aux cas hormis la maladie. Lorsque la maladie est infectieuse, le minimum est de garantir une exposition des contrôles comparable à celle des cas. Cela est difficile aujourd'hui pour la lèpre qui n'est heureusement plus aussi endémique qu'autrefois, et pour laquelle il n'existe pas de marqueur d'infection fiable qui témoignerait de l'exposition. Dans notre étude, nos cas MB et nos contrôles PB ont tous été exposés et infectés par *M. leprae*. Il s'agit alors de nous assurer que les principales variables démographiques (âge et sexe) sont comparables entre les deux groupes.

Le tableau 12 donne les caractéristiques des 1 080 individus qui ont été classés selon l'OMS-82 : 340 sont PB et 740 MB. Si la proportion de femmes n'était pas significativement différente entre les groupes MB et PB, celle-ci avait tendance à être plus grande dans le groupe PB ($p = 0,09$). De même, l'âge médian au diagnostic avait tendance à être plus élevé dans le groupe MB ($p = 0,08$). Ceci était attendu en l'absence de consigne de recrutement selon l'âge. Cela reflète l'âge plus tardif au diagnostic des formes MB, au Vietnam comme dans le reste du monde (Languillon, 1999). Cependant, la répartition par grands groupes d'âge (4-9 ans, 10-19 ans, 20-59 ans, > 60 ans) n'était pas significativement différente entre les deux groupes ($p = 0,25$). Le nombre moyen de lésions, la proportion de cas avec plus de 5 lésions ou un IB > 1, étaient comparables entre les groupes PB de l'échantillon de réplication et des enfants de l'étude initiale. De même pour le groupe MB. En revanche, la proportion de réaction de type 1 ou 2 variait considérablement entre les enfants de l'étude initiale et l'échantillon de réplication. La proportion de réaction de type 2 chez les cas MB de réplication (16 %) paraît même très élevée. Ce sera un point du phénotype à clarifier avec les recruteurs.

Tableau 12. Caractéristiques des cas de lèpre dans l'échantillon de réplication vietnamien

Forme clinique de lèpre (classification OMS 82)	Réplication		
	PB N (%)	MB N (%)	p PB vs. MB
Total	340	740	
Hommes	218	514	
Femmes	122 (36)	226 (31)	0,09
Age médian au diagnostic	23,3	24,4	0,08
Répartition par groupe d'âge			0,25
4-9	1	0	
10-19	99 (29)	186 (25)	
20-59	235 (69)	541 (74)	
> 60	5 (2)	10 (1)	
Nombre moyen de lésions	2,1	8,5	$< 10^{-6}$
> 5 lésions	21	539	$< 10^{-6}$
IB > 0	11	423	$< 10^{-6}$
Réactions			
Type 1 (RR)	40 (12)	159 (21)	$2 \cdot 10^{-4}$
Type 2 (ENL)	3 (1)	120 (16)	$< 10^{-6}$

Perspectives

Synthèse

Notre sujet de thèse nous a conduits à étudier l'épidémiologie classique et génétique des formes cliniques de la lèpre. La revue de la littérature a pu identifier le sexe et l'âge comme deux facteurs de risque indépendants de la polarisation. Les hommes âgés ont plus de risque de développer une forme multibacillaire que les femmes jeunes. Nous n'avons pu statuer sur le rôle du BCG et les données étaient insuffisantes pour étudier d'autres facteurs de risque non génétiques.

Sur le plan génétique, nous avons identifié des études antérieures pour lesquelles des associations de polymorphismes génétiques avec le phénotype polarisation différaient des associations identifiées pour le lèpre *per se*. Nous avons montré certaines limites méthodologiques de la stratégie qui consiste à étudier la lèpre *per se* avant ses formes cliniques, lorsque l'objectif est précisément d'étudier ces formes cliniques particulières.

Nous avons ensuite étudié la liaison génétique avec la polarisation de la lèpre dans un échantillon de 64 familles de l'ethnie *Kinh* du Vietnam. Celle-ci a montré qu'en codant le phénotype polarisation comme un trait ordonné à 5 catégories selon la classification de Ridley-Jopling, la liaison avec la région 19p12 était presque significative à l'échelle du génome. L'intervalle de confiance du pic de liaison autour de rs2049051 définissait une région de 3,7 Mb.

L'analyse d'association de la polarisation menée avec plus de 400 000 marqueurs dans des familles comprenant 692 enfants atteints de lèpre n'a en revanche pas identifié de marqueur proche de la significativité. La méthode d'association sous un modèle additif que nous avons développée n'a pas plus permis d'atteindre le seuil de significativité. Nous disposons cependant d'un échantillon indépendant d'individus PB et MB qui nous permettra d'essayer de répliquer les meilleures associations de notre échantillon primaire.

Limites et perspectives immédiates

Notre étude d'association pangénomique a été limitée à l'étude de la polarisation traitée comme un phénotype binaire, en l'absence de covariables, par un nombre limité de méthodes s'inspirant toutes de FBAT. L'absence de résultat significatif et clair d'association avec un marqueur dans notre analyse doit nous conduire à améliorer notre stratégie, en utilisant les connaissances disponibles, à la fois sur le phénotype et sur les méthodes statistiques.

Nous avons montré, dans la revue de la littérature du deuxième chapitre, que la polarisation de la lèpre était influencée par le sexe et par l'âge (de Vries and Perry, 1985; Richard et al., 2010; Guerra-Silveira and Abad-Franch, 2013). Nous avons pu nous affranchir de ces variables dans notre analyse de liaison car elles n'étaient pas corrélées au phénotype dans le sous-échantillon de familles multiples. Cependant, l'âge (mais pas le sexe) influait sur la polarisation dans l'échantillon complet. Comme nous l'avons rappelé dans notre discussion méthodologique, la lèpre est un bel exemple où l'ajout de l'âge dans le modèle a permis de préciser une association (Alter et al., 2012).

Ensuite, nous n'avons envisagé la polarisation que comme un phénotype binaire. Un des intérêts d'envisager les formes cliniques de lèpre sous l'angle de la polarisation est que ce phénotype peut être ordonné en plus de 2 catégories, et même quantifié. L'utilisation de la classification de Ridley-Jopling (en 5 catégories) dans l'analyse de liaison nous a permis d'augmenter le LOD-score de la région de liaison, et d'approcher le seuil de significativité à l'échelle du génome. Or cette classification est disponible pour la quasi-totalité des cas de lèpre de notre échantillon, et FBAT s'utilise aussi facilement avec un trait binaire que quantitatif. En revanche, l'interprétation biologique d'un résultat d'association avec un trait quantitatif est moins immédiate.

La principale limite de notre travail concerne sans doute l'aspect méthodologique. Si FBAT est une méthode d'analyse robuste, facile à utiliser, et qui a permis de belles découvertes, notamment dans la lèpre, d'autres méthodes ont été développées pour étudier l'association dans une population d'apparentés.

En particulier, nous n'avons pas exploité la donnée du phénotype des parents dans notre analyse d'association. Or, parmi les 939 cas de lèpre de notre échantillon, 247 étaient des parents ou des individus isolés. Trois types de stratégies permettent de valoriser cette donnée dans le cadre des études d'association. La première consiste simplement à se servir des parents comme

échantillon de réplication. Le petit nombre de cas (64 MB avec la définition OMS-82) et de contrôles (152 PB) rendait cependant cette approche peu séduisante. La seconde stratégie combine la valeur de p obtenue par FBAT avec les génotypes des enfants, et celle obtenue en comparant les fréquences des génotypes attendus en fonction des données parentales entre les cas et les contrôles. Ces deux tests sont indépendants. Le second test est sensible à la stratification de population, mais pas le premier, et la combinaison est telle que le test global ne l'est pas (Ionita-Laza et al., 2007; Lasky-Su et al., 2010; Won et al., 2009). Enfin, la troisième stratégie est de réaliser une étude cas-témoins avec tous les phénotypes disponibles (parents et enfants) et de tenir compte des relations familiales dans le cadre d'un modèle mixte, comme proposé dans le logiciel GEMMA (Zhou and Stephens, 2012).

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. Il existe une extension du MLB-QTL pour tester cette empreinte (Vincent et al., 2006). Cette analyse pourra donc être réalisée prochainement, sans coût supplémentaire.

Enfin, nous n'avons pas testé l'association sur les chromosomes sexuels. Certes, la majorité des GWAS publiés ne rapportent pas l'association sur les chromosomes X et Y, qui posent des problématiques statistiques particulières (Wise et al., 2013). 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'hétérozygocité à ces loci. Cependant, des outils robustes sont maintenant disponibles, y compris pour des données familiales (König et al., 2014).

Perspectives de recherche originales

Recherche d'une hérédité mendélienne dans certaines familles

Nous avons jusqu'alors plutôt envisagé la lèpre comme une maladie « complexe », pour laquelle l'objectif est de rechercher des variants génétiques qui ne sont pas individuellement responsables du phénotype, mais qui augmentent chacun un 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. 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 (Manolio et al., 2009).

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 bientôt du génome entier, est maintenant commun.

Une idée future de projet de recherche serait d'étudier notre phénotype, la polarisation de la lèpre, sous l'angle mendélien, et d'aller plus loin que le seul contraste de deux grands groupes aux phénotypes extrêmes (comme les TT vs. les LL). Nous pourrions séquencer l'exome de deux ou trois fratries dont les cas présentent 1) des phénotypes très discordants (TT vs. LL), 2) un âge jeune au diagnostic et éventuellement 3) un sexe opposé à celui connu pour être un facteur de risque de la forme clinique en question. Nous disposons d'une famille qui remplit ces critères : une fratrie dans laquelle la sœur a été diagnostiquée LL à l'âge de 13 ans, et le frère TT à l'âge de 17 ans. Une étude récente suggère que le séquençage du génome est plus performant pour détecter les variants des régions codantes que le séquençage de l'exome (Belkadi et al., 2014). Avec les coûts décroissants de ces technologies, le séquençage du génome serait sans doute le matériel de choix pour cette étude.

Le tatou à 9 bandes

Le tatou à neuf bandes est un mammifère de la taille du chat qui peut vivre jusqu'à 12 ans. Il vit essentiellement dans les Amériques et sa température corporelle fraîche (32-35°C) explique sans doute qu'il soit un hôte naturel de *M. leprae* (Job, 2003; Clark et al., 2008). Il peut aussi être utilisé pour l'étude de la multiplication *in vivo* des bacilles (Kirchheimer and Storrs, 1971; Adams et al., 2012; Truman et al., 2014). La majorité des tatous à neuf bandes développe une forme disséminée de la maladie après une inoculation par voie intraveineuse de *M. leprae*. La clinique ressemble à la lèpre MB de l'homme. Les lésions nerveuses sont en particulier sévères (Sharma et al., 2013). Cette observation se retrouve lorsque le tatou s'infecte dans son milieu naturel. Ainsi, une étude sur 392 tatous sauvages malades aux États-Unis a montré que

plus de 80 % étaient MB selon Ridley-Jopling (72 % LL et 11 % BL). Cette proportion était comparable entre la Louisiane – région de zoonose, et la Floride – région où les cas sont sporadiques (Adams et al., 2012). Cette étude a donc mis en évidence que 17 % de tatous pouvaient être classés PB. L'animal, généralement considéré comme un modèle pour la lèpre MB, pourrait donc aussi servir à l'étude de la forme PB de la lèpre, et de sa polarisation.

Le génome du tatou a été séquencé et annoté (Lindblad-Toh et al., 2011; Lynch et al., 2011). A ce jour, un seul tatou a été entièrement séquencé. Le séquençage d'autres 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. Adams *et al* ont 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 (Adams et al., 2012).

La pléiotropie : lien avec les maladies inflammatoires de l'intestin

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) (Zhang et al., 2009a; Alter et al., 2011; Grant et al., 2012) dont l'origine bactérienne est toujours débattue (Greenstein, 2003; Sartor, 2005; Uzoigwe et al., 2007; Pierce et al., 2010). L'analogie peut être poursuivie en comparant les deux principales formes de MICI, la MC et la rectocolite hémorragique (RCH), avec les deux pôles cliniques de la lèpre.

En effet, la forme PB, comme la MC, implique préférentiellement une signalisation Th1 tandis que la forme MB, comme la RCH, implique une réponse Th2 (Brand, 2009). Les GWAS sur les MICI ont identifié 163 loci associés aux MICI : 110 étaient communs aux deux maladies tandis que 30 étaient spécifiques pour la MC et 23 pour la RCH (Jostins et al., 2012). Trois gènes spécifiquement associés à la MC (*RIPK2*, *LACC1* et *NOD2*) étaient aussi associés à la lèpre : *NOD2* à la lèpre *per se*, *RIPK2* à la forme PB et *LACC1* à la forme MB.

Dans le premier GWAS sur la lèpre, quatre gènes (*NOD2*, *TNFSF15*, *RIPK2* et *HLA DR-DQ*) appartiennent à la voie de signalisation impliquant *NOD2* (Zhang et al., 2009a). *NOD2* est d'ailleurs le premier gène de susceptibilité découvert par une approche pangénomique dans la maladie de Crohn. Les odds-ratio de 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 (Hugot et al., 2001).

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 (Strober and Watanabe, 2011). 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 actuellement 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 (Greenstein, 2003), 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 (Alter et al., 2011).

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 lymphotoxine α participe au recrutement des macrophages et des lymphocytes aux sites d'infection par des bactéries intracellulaires (Roach et al., 2001). 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-contrôle) (Alcaïs et al., 2007; Alter et al., 2008).

Les analogies entre les MICI et la lèpre *per se*, entre la MC et la forme PB, et entre la RCH et la forme MB 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, et aussi de sa polarisation.

Conclusion

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 et de sa polarisation. Notre thèse a illustré que de nombreuses associations sont bien spécifiques de la survenue d'une forme clinique de lèpre préférentiellement à une autre. Notre modeste contribution aura été l'identification d'une région de liaison pour la polarisation, située sur le chromosome 19p12. Nous suggérons que l'identification d'autres liaisons ou associations avec la polarisation sera optimisée par une caractérisation précise du phénotype et par une approche méthodologique spécifique. Par ailleurs, 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, dont la polarisation est un aspect important. Les avancées nécessiteront une étroite collaboration entre les médecins, biologistes, épidémiologistes et les généticiens. L'identification de marqueurs associés à la polarisation de 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.

Bibliographie

- 1000 Genomes Project Consortium, Abecasis, G.R., Auton, A., Brooks, L.D., DePristo, M.A., Durbin, R.M., Handsaker, R.E., Kang, H.M., Marth, G.T., and McVean, G.A. (2012). An integrated map of genetic variation from 1,092 human genomes. *Nature* 491, 56–65.
- Abel, L., and Demenais, F. (1988). Detection of major genes for susceptibility to leprosy and its subtypes in a Caribbean island: Desirade island. *Am. J. Hum. Genet.* 42, 256–266.
- Abel, L., and Muller-Myhsok, B. (1998). Robustness and power of the maximum-likelihood-binomial and maximum-likelihood-score methods, in multipoint linkage analysis of affected- sibship data. *Am J Hum Genet* 63, 638–647.
- Abel, L., Alcais, A., and Mallet, A. (1998). Comparison of four sib-pair linkage methods for analyzing sibships with more than two affecteds: interest of the binomial maximum likelihood approach. *Genet. Epidemiol.* 15, 371–390.
- Adams, L.B., Pena, M.T., Sharma, R., Hagge, D.A., Schurr, E., and Truman, R.W. (2012). Insights from animal models on the immunogenetics of leprosy: a review. *Mem. Inst. Oswaldo Cruz* 107 Suppl 1, 197–208.
- Alcais, A., Philippi, A., and Abel, L. (1999). Genetic model-free linkage analysis using the maximum-likelihood-binomial method for categorical traits. *Genet Epidemiol* 17 Suppl 1, S467–S472.
- Alcais, A., Plancoulaine, S., and Abel, L. (2001). An autosome-wide search for loci underlying wheezing age of onset in German asthmatic children identifies a new region of interest on 6q24-q25. *Genet Epidemiol* 21 Suppl 1, S168–S173.
- Alcaïs, A., Mira, M., Casanova, J.-L., Schurr, E., and Abel, L. (2005). Genetic dissection of immunity in leprosy. *Curr. Opin. Immunol.* 17, 44–48.
- Alcaïs, A., Alter, A., Antoni, G., Orlova, M., Nguyen, V.T., Singh, M., Vanderborght, P.R., Katoch, K., Mira, M.T., Vu, H.T., et al. (2007). Stepwise replication identifies a low-producing lymphotoxin-alpha allele as a major risk factor for early-onset leprosy. *Nat. Genet.* 39, 517–522.
- Ali, S., Srivastava, A.K., Chopra, R., Aggarwal, S., Garg, V.K., Bhattacharya, S.N., and Bamezai, R.N.K. (2013). IL12B SNPs and copy number variation in IL23R gene associated with susceptibility to leprosy. *J. Med. Genet.* 50, 34–42.
- Alter, A., Alcais, A., Abel, L., and Schurr, E. (2008). Leprosy as a genetic model for susceptibility to common infectious diseases. *Hum Genet.*
- Alter, A., de Léséleuc, L., Van Thuc, N., Thai, V.H., Huong, N.T., Ba, N.N., Cardoso, C.C., Grant, A.V., Abel, L., Moraes, M.O., et al. (2010). Genetic and functional analysis of common MRC1 exon 7 polymorphisms in leprosy susceptibility. *Hum. Genet.* 127, 337–348.
- Alter, A., Grant, A., Abel, L., Alcaïs, A., and Schurr, E. (2011). Leprosy as a genetic disease. *Mamm. Genome Off. J. Int. Mamm. Genome Soc.* 22, 19–31.

- Alter, A., Fava, V.M., Huong, N.T., Singh, M., Orlova, M., Van Thuc, N., Katoch, K., Thai, V.H., Ba, N.N., Abel, L., et al. (2012). Linkage disequilibrium pattern and age-at-diagnosis are critical for replicating genetic associations across ethnic groups in leprosy. *Hum. Genet.*
- Anderson, C.A., Pettersson, F.H., Clarke, G.M., Cardon, L.R., Morris, A.P., and Zondervan, K.T. (2010). Data quality control in genetic case-control association studies. *Nat. Protoc.* 5, 1564–1573.
- Anonymous (1953). [Propositions of a group of leprologists of Madrid regarding the classification of leprosy]. *Actas Dermo-Sifiliográficas* 44, 706–708.
- Antunes, D.E., Araujo, S., Ferreira, G.P., Cunha, A.C.S.R. da, Costa, A.V. da, Gonçalves, M.A., and Goulart, I.M.B. (2013). Identification of clinical, epidemiological and laboratory risk factors for leprosy reactions during and after multidrug therapy. *Mem. Inst. Oswaldo Cruz* 108, 901–908.
- Aubry, P. *Lèpre ou maladie de Hansen.*
- Bakker, M.I., Hatta, M., Kwenang, A., Van Benthem, B.H.B., Van Beers, S.M., Klatser, P.R., and Oskam, L. (2005). Prevention of leprosy using rifampicin as chemoprophylaxis. *Am. J. Trop. Med. Hyg.* 72, 443–448.
- Bakker, M.I., Hatta, M., Kwenang, A., Van Mosseveld, P., Faber, W.R., Klatser, P.R., and Oskam, L. (2006). Risk factors for developing leprosy--a population-based cohort study in Indonesia. *Lepr. Rev.* 77, 48–61.
- Beiguelman, B., da Silva, E., and Dall’Aglia, F.F. (1968). [Leprosy and sex]. *Rev. Paul. Med.* 72, 120–129.
- Belkadi, A., Bolze, A., Itan, Y., Vincent, Q.B., Antipenko, A., Boisson, B., Casanova, J.-L., and Abel, L. (2014). Whole-genome sequencing is more powerful than whole-exome sequencing for detecting exome variants. *bioRxiv* 010363.
- Blake, L.A., West, B.C., Lary, C.H., and Todd, J.R., 4th (1987). Environmental nonhuman sources of leprosy. *Rev. Infect. Dis.* 9, 562–577.
- Bleharski, J.R., Li, H., Meinken, C., Graeber, T.G., Ochoa, M.-T., Yamamura, M., Burdick, A., Sarno, E.N., Wagner, M., Rölinghoff, M., et al. (2003). Use of genetic profiling in leprosy to discriminate clinical forms of the disease. *Science* 301, 1527–1530.
- Bochud, P.-Y., Hawn, T.R., Siddiqui, M.R., Saunderson, P., Britton, S., Abraham, I., Argaw, A.T., Janer, M., Zhao, L.P., Kaplan, G., et al. (2008). Toll-like receptor 2 (TLR2) polymorphisms are associated with reversal reaction in leprosy. *J. Infect. Dis.* 197, 253–261.
- Bochud, P.-Y., Sinsimer, D., Aderem, A., Siddiqui, M.R., Saunderson, P., Britton, S., Abraham, I., Tadesse Argaw, A., Janer, M., Hawn, T.R., et al. (2009). Polymorphisms in Toll-like receptor 4 (TLR4) are associated with protection against leprosy. *Eur. J. Clin. Microbiol. Infect. Dis. Off. Publ. Eur. Soc. Clin. Microbiol.* 28, 1055–1065.
- Brand, S. (2009). Crohn’s disease: Th1, Th17 or both? The change of a paradigm: new immunological and genetic insights implicate Th17 cells in the pathogenesis of Crohn’s disease. *Gut* 58, 1152–1167.
- Britton, W.J., and Lockwood, D.N.J. (2004). Leprosy. *Lancet* 363, 1209–1219.

- Caws, M., Thwaites, G., Dunstan, S., Hawn, T.R., Lan, N.T., Thuong, N.T., Stepniewska, K., Huyen, M.N., Bang, N.D., Loc, T.H., et al. (2008). The influence of host and bacterial genotype on the development of disseminated disease with *Mycobacterium tuberculosis*. *PLoS Pathog* 4, e1000034.
- Chaitanya, V.S., Jadhav, R.S., Lavania, M., Singh, M., Valluri, V., and Sengupta, U. (2013). Interleukin-17F single-nucleotide polymorphism (7488T>C) and its association with susceptibility to leprosy. *Int. J. Immunogenet.*
- Chakravarti, M.R., and Vogel, F. (1973). *A twin study on leprosy* (Georg Thieme Publishers, Stuttgart, Germany).
- Clark, B.M., Murray, C.K., Horvath, L.L., Deye, G.A., Rasnake, M.S., and Longfield, R.N. (2008). Case-control study of armadillo contact and Hansen's disease. *Am. J. Trop. Med. Hyg.* 78, 962–967.
- Cobat, A. (2010). *Génétique épidémiologique de l'infection tuberculeuse*. Paris XI.
- Cobat, A., Abel, L., and Alcaïs, A. (2011a). The Maximum-Likelihood-Binomial method revisited: a robust approach for model-free linkage analysis of quantitative traits in large sibships. *Genet Epidemiol* 35, 46–56.
- Cobat, A., Abel, L., and Alcaïs, A. (2011b). The Maximum-Likelihood-Binomial method revisited: a robust approach for model-free linkage analysis of quantitative traits in large sibships. *Genet. Epidemiol.* 35, 46–56.
- Cole, S.T., Brosch, R., Parkhill, J., Garnier, T., Churcher, C., Harris, D., Gordon, S.V., Eiglmeier, K., Gas, S., Barry, C.E., 3rd, et al. (1998). Deciphering the biology of *Mycobacterium tuberculosis* from the complete genome sequence. *Nature* 393, 537–544.
- Cole, S.T., Eiglmeier, K., Parkhill, J., James, K.D., Thomson, N.R., Wheeler, P.R., Honoré, N., Garnier, T., Churcher, C., Harris, D., et al. (2001). Massive gene decay in the leprosy bacillus. *Nature* 409, 1007–1011.
- Convit, J., Sampson, C., Zúñiga, M., Smith, P.G., Plata, J., Silva, J., Molina, J., Pinardi, M.E., Bloom, B.R., and Salgado, A. (1992). Immunoprophylactic trial with combined *Mycobacterium leprae*/BCG vaccine against leprosy: preliminary results. *Lancet* 339, 446–450.
- Coscolla, M., and Gagneux, S. (2014). Consequences of genomic diversity in *Mycobacterium tuberculosis*. *Semin. Immunol.* 26, 431–444.
- Dupuis, J., and Siegmund, D. (1999). Statistical methods for mapping quantitative trait loci from a dense set of markers. *Genetics* 151, 373–386.
- Fajardo, T.T., Villahermosa, L.G., Cruz, E.C.D., Cellona, R.V., Balagon, M.V.F., Abalos, R.M., and Gelber, R.H. (2004). A clinical trial of pefloxacin and ofloxacin in lepromatous leprosy. *Lepr. Rev.* 75, 389–397.
- Feitosa, M.F., Borecki, I., Krieger, H., Beiguelman, B., and Rao, D.C. (1995). The genetic epidemiology of leprosy in a Brazilian population. *Am. J. Hum. Genet.* 56, 1179–1185.
- Fine, P.E. (1982). Leprosy: the epidemiology of a slow bacterium. *Epidemiol. Rev.* 4, 161–188.

- Fitness, J., Floyd, S., Warndorff, D.K., Sichali, L., Mwaungulu, L., Crampin, A.C., Fine, P.E.M., and Hill, A.V.S. (2004). Large-scale candidate gene study of leprosy susceptibility in the Karonga district of northern Malawi. *Am. J. Trop. Med. Hyg.* **71**, 330–340.
- Flageul, B. (2011). Le diagnostic de la lèpre. *Rev. Francoph. Lab.* **431**, 37–42.
- Franceschi, D.S.A., Mazini, P.S., Rudnick, C.C.C., Sell, A.M., Tsuneto, L.T., de Melo, F.C., Braga, M.A., Peixoto, P.R.F., and Visentainer, J.E.L. (2008). Association between killer-cell immunoglobulin-like receptor genotypes and leprosy in Brazil. *Tissue Antigens* **72**, 478–482.
- Franceschi, D.S.A., Mazini, P.S., Rudnick, C.C.C., Sell, A.M., Tsuneto, L.T., Ribas, M.L., Peixoto, P.R., and Visentainer, J.E.L. (2009). Influence of TNF and IL10 gene polymorphisms in the immunopathogenesis of leprosy in the south of Brazil. *Int. J. Infect. Dis. IJID Off. Publ. Int. Soc. Infect. Dis.* **13**, 493–498.
- García, V.E., Uyemura, K., Sieling, P.A., Ochoa, M.T., Morita, C.T., Okamura, H., Kurimoto, M., Rea, T.H., and Modlin, R.L. (1999). IL-18 promotes type 1 cytokine production from NK cells and T cells in human intracellular infection. *J. Immunol. Baltim. Md 1950* **162**, 6114–6121.
- Gaschignard, J., Scurr, E., and Alcaïs, A. (2013). La lèpre, pilier de la génétique des maladies infectieuses. *Pathol. Biol.* **61**, 120–128.
- Gaschignard, J., Vincent, Q., Cobat, A., and Alcais, A. Implicit null hypothesis are hidden power droppers in family.
- Gillis, T.P., Scollard, D.M., and Lockwood, D.N.J. (2011). What is the evidence that the putative *Mycobacterium lepromatosis* species causes diffuse lepromatous leprosy? *Lepr. Rev.* **82**, 205–209.
- Le Grand, A. (1997). Women and leprosy: a review. *Lepr. Rev.* **68**, 203–211.
- Grant, A.V., Alter, A., Huong, N.T., Orlova, M., Van Thuc, N., Ba, N.N., Thai, V.H., Abel, L., Schurr, E., and Alcais, A. (2012). Crohn's disease susceptibility genes are associated with leprosy in the Vietnamese population. *J. Infect. Dis.* **206**, 1763–1767.
- Grant, A.V., Cobat, A., Van Thuc, N., Orlova, M., Huong, N.T., Gaschignard, J., Alter, A., Ba, N.N., Thai, V.H., Abel, L., et al. (2014). CUBN and NEBL common variants in the chromosome 10p13 linkage region are associated with multibacillary leprosy in Vietnam. *Hum. Genet.*
- Greenstein, R.J. (2003). Is Crohn's disease caused by a mycobacterium? Comparisons with leprosy, tuberculosis, and Johne's disease. *Lancet Infect. Dis.* **3**, 507–514.
- Greiner, J., Weber, F.J., Mauff, G., and Baur, M. (1980). 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* **158**, 134–138.
- Guerra-Silveira, F., and Abad-Franch, F. (2013). Sex bias in infectious disease epidemiology: patterns and processes. *PloS One* **8**, e62390.
- Gupte, M.D., Vallishayee, R.S., Anantharaman, D.S., Nagaraju, B., Sreevatsa, Balasubramanyam, S., de Britto, R.L., Elango, N., Uthayakumaran, N., Mahalingam, V.N., et al. (1998). Comparative leprosy vaccine trial in south India. *Indian J. Lepr.* **70**, 369–388.

- Han, X.Y., Seo, Y.-H., Sizer, K.C., Schoberle, T., May, G.S., Spencer, J.S., Li, W., and Nair, R.G. (2008). A new *Mycobacterium* species causing diffuse lepromatous leprosy. *Am. J. Clin. Pathol.* **130**, 856–864.
- Han, X.Y., Sizer, K.C., Thompson, E.J., Kabanja, J., Li, J., Hu, P., Gómez-Valero, L., and Silva, F.J. (2009). Comparative sequence analysis of *Mycobacterium leprae* and the new leprosy-causing *Mycobacterium lepromatosis*. *J. Bacteriol.* **191**, 6067–6074.
- Han, X.Y., Aung, F.M., Choon, S.E., and Werner, B. (2014). Analysis of the leprosy agents *Mycobacterium leprae* and *Mycobacterium lepromatosis* in four countries. *Am. J. Clin. Pathol.* **142**, 524–532.
- Hansen, G.A., and Looft, C. (1895). Leprosy: in its clinical and pathological aspects.
- Van Heel, D.A., Dechairo, B.M., Dawson, G., McGovern, D.P.B., Negoro, K., Carey, A.H., Cardon, L.R., Mackay, I., Jewell, D.P., and Lench, N.J. (2003). The IBD6 Crohn's disease locus demonstrates complex interactions with CARD15 and IBD5 disease-associated variants. *Hum. Mol. Genet.* **12**, 2569–2575.
- Hugot, J.P., Chamaillard, M., Zouali, H., Lesage, S., Cézard, J.P., Belaiche, J., Almer, S., Tysk, C., O'Morain, C.A., Gassull, M., et al. (2001). Association of NOD2 leucine-rich repeat variants with susceptibility to Crohn's disease. *Nature* **411**, 599–603.
- Ilanguvaran, S., Ramanathan, S., Shankernarayan, N., Ramu, G., and Muthukarauppan, V. (1996). Immunological profiles of leprosy patients and healthy family contacts toward *M. leprae* antigens. *Int. J. Lepr. Mycobact. Dis. Off. Organ Int. Lepr. Assoc.* **64**, 6–14.
- Ionita-Laza, I., McQueen, M.B., Laird, N.M., and Lange, C. (2007). Genomewide weighted hypothesis testing in family-based association studies, with an application to a 100K scan. *Am. J. Hum. Genet.* **81**, 607–614.
- Jamieson, S.E., Miller, E.N., Black, G.F., Peacock, C.S., Cordell, H.J., Howson, J.M.M., Shaw, M.-A., Burgner, D., Xu, W., Lins-Lainson, Z., et al. (2004). Evidence for a cluster of genes on chromosome 17q11-q21 controlling susceptibility to tuberculosis and leprosy in Brazilians. *Genes Immun.* **5**, 46–57.
- Jannot, A.-S., Ehret, G., and Perneger, T. $P < 5 \times 10^{-8}$ has emerged as a standard of statistical significance for genome-wide association studies. *J. Clin. Epidemiol.*
- Job, C.K. (2003). Nine-banded armadillo and leprosy research. *Indian J. Pathol. Microbiol.* **46**, 541–550.
- Jonas, E., Schreinemachers, H.-J., Kleinwächter, T., Un, C., Oltmanns, I., Tetzlaff, S., Jennen, D., Tesfaye, D., Ponsuksili, S., Murani, E., et al. (2008). QTL for the heritable inverted teat defect in pigs. *Mamm. Genome Off. J. Int. Mamm. Genome Soc.* **19**, 127–138.
- De Jong, R., Janson, A.A., Faber, W.R., Naafs, B., and Ottenhoff, T.H. (1997). IL-2 and IL-12 act in synergy to overcome antigen-specific T cell unresponsiveness in mycobacterial disease. *J. Immunol. Baltim. Md 1950* **159**, 786–793.
- Jostins, L., Ripke, S., Weersma, R.K., Duerr, R.H., McGovern, D.P., Hui, K.Y., Lee, J.C., Schumm, L.P., Sharma, Y., Anderson, C.A., et al. (2012). Host-microbe interactions have shaped the genetic architecture of inflammatory bowel disease. *Nature* **491**, 119–124.
- Kang, T.J., and Chae, G.T. (2001). Detection of Toll-like receptor 2 (TLR2) mutation in the lepromatous leprosy patients. *FEMS Immunol. Med. Microbiol.* **31**, 53–58.

- Kennerson, M.L., Zhu, D., Gardner, R.J., Storey, E., Merory, J., Robertson, S.P., and Nicholson, G.A. (2001). Dominant intermediate Charcot-Marie-Tooth neuropathy maps to chromosome 19p12-p13.2. *Am. J. Hum. Genet.* 69, 883–888.
- Kirchheimer, W.F., and Storrs, E.E. (1971). Attempts to establish the armadillo (*Dasypus novemcinctus* Linn.) as a model for the study of leprosy. I. Report of lepromatoid leprosy in an experimentally infected armadillo. *Int. J. Lepr. Mycobact. Dis. Off. Organ Int. Lepr. Assoc.* 39, 693–702.
- Knapp, M. (1999). The transmission/disequilibrium test and parental-genotype reconstruction: the reconstruction-combined transmission/ disequilibrium test. *Am. J. Hum. Genet.* 64, 861–870.
- König, I.R., Loley, C., Erdmann, J., and Ziegler, A. (2014). How to include chromosome X in your genome-wide association study. *Genet. Epidemiol.* 38, 97–103.
- Kruglyak, L., Daly, M.J., Reeve-Daly, M.P., and Lander, E.S. (1996). Parametric and nonparametric linkage analysis: a unified multipoint approach. *Am. J. Hum. Genet.* 58, 1347–1363.
- Laird, N.M., and Lange, C. (2006). Family-based designs in the age of large-scale gene-association studies. *Nat. Rev. Genet.* 7, 385–394.
- Laird, N.M., Horvath, S., and Xu, X. (2000). Implementing a unified approach to family-based tests of association. *Genet. Epidemiol.* 19 Suppl 1, S36–S42.
- Lake, S.L., Blacker, D., and Laird, N.M. (2000). Family-based tests of association in the presence of linkage. *Am. J. Hum. Genet.* 67, 1515–1525.
- Lange, C., DeMeo, D., Silverman, E.K., Weiss, S.T., and Laird, N.M. (2003). Using the noninformative families in family-based association tests: a powerful new testing strategy. *Am. J. Hum. Genet.* 73, 801–811.
- Languillon, J. (1999). *Précis de léprologie*.
- Lasky-Su, J., Won, S., Mick, E., Anney, R.J.L., Franke, B., Neale, B., Biederman, J., Smalley, S.L., Loo, S.K., Todorov, A., et al. (2010). On genome-wide association studies for family-based designs: an integrative analysis approach combining ascertained family samples with unselected controls. *Am. J. Hum. Genet.* 86, 573–580.
- De Léséleuc, L., Orlova, M., Cobat, A., Girard, M., Huong, N.T., Ba, N.N., Thuc, N.V., Truman, R., Spencer, J.S., Adams, L., et al. (2013). PARK2 mediates interleukin 6 and monocyte chemoattractant protein 1 production by human macrophages. *PLoS Negl. Trop. Dis.* 7, e2015.
- Lindblad-Toh, K., Garber, M., Zuk, O., Lin, M.F., Parker, B.J., Washietl, S., Kheradpour, P., Ernst, J., Jordan, G., Mauceli, E., et al. (2011). A high-resolution map of human evolutionary constraint using 29 mammals. *Nature* 478, 476–482.
- Liu, J.Z., Almarri, M.A., Gaffney, D.J., Mells, G.F., Jostins, L., Cordell, H.J., Ducker, S.J., Day, D.B., Heneghan, M.A., Neuberger, J.M., et al. (2012). Dense fine-mapping study identifies new susceptibility loci for primary biliary cirrhosis. *Nat. Genet.* 44, 1137–1141.
- Lockwood, D.N.J., Shetty, V., and Penna, G.O. (2014). Hazards of setting targets to eliminate disease: lessons from the leprosy elimination campaign. *BMJ* 348, g1136.

- Lucena-Silva, N., Teixeira, M.A.G., Ramos, A. de L., de Albuquerque, R.S., Diniz, G.T.N., Mendes-Junior, C.T., Castelli, E.C., and Donadi, E.A. (2013). The +3187A/G HLA-G polymorphic site is associated with polar forms and reactive reaction in leprosy. *Mol. Genet. Genomic Med.* 1, 123–130.
- Lynch, V.J., Leclerc, R.D., May, G., and Wagner, G.P. (2011). Transposon-mediated rewiring of gene regulatory networks contributed to the evolution of pregnancy in mammals. *Nat. Genet.* 43, 1154–1159.
- Manolio, T.A., Collins, F.S., Cox, N.J., Goldstein, D.B., Hindorff, L.A., Hunter, D.J., McCarthy, M.I., Ramos, E.M., Cardon, L.R., Chakravarti, A., et al. (2009). Finding the missing heritability of complex diseases. *Nature* 461, 747–753.
- Manzanillo, P.S., Ayres, J.S., Watson, R.O., Collins, A.C., Souza, G., Rae, C.S., Schneider, D.S., Nakamura, K., Shiloh, M.U., and Cox, J.S. (2013). The ubiquitin ligase parkin mediates resistance to intracellular pathogens. *Nature*.
- Marcinek, P., Jha, A.N., Shinde, V., Sundaramoorthy, A., Rajkumar, R., Suryadevara, N.C., Neela, S.K., van Tong, H., Balachander, V., Valluri, V.L., et al. (2013). LRRK2 and RIPK2 Variants in the NOD 2-Mediated Signaling Pathway Are Associated with Susceptibility to *Mycobacterium leprae* in Indian Populations. *PLoS ONE* 8, e73103.
- Martínez Domínguez, V., Gallego Garbajosa, P., Mg Gyi, M., Tamondong, C.T., Sundaresan, T., Bechelli, L.M., Lwin, K., Sansarricq, H., Walter, J., and Noussitou, F.M. (1980). Epidemiological information on leprosy in the Singu area of Upper Burma. *Bull. World Health Organ.* 58, 81–89.
- Masaki, T., Qu, J., Cholewa-Waclaw, J., Burr, K., Raaum, R., and Rambukkana, A. (2013). Reprogramming Adult Schwann Cells to Stem Cell-like Cells by Leprosy Bacilli Promotes Dissemination of Infection. *Cell* 152, 51–67.
- Meisner, S.J., Mucklow, S., Warner, G., Sow, S.O., Lienhardt, C., and Hill, A.V. (2001). Association of NRAM1 polymorphism with leprosy type but not susceptibility to leprosy per se in west Africans. *Am. J. Trop. Med. Hyg.* 65, 733–735.
- Merle, C.S.C., Cunha, S.S., and Rodrigues, L.C. (2010). BCG vaccination and leprosy protection: review of current evidence and status of BCG in leprosy control. *Expert Rev. Vaccines* 9, 209–222.
- De Messias-Reason, I.J., Boldt, A.B.W., Moraes Braga, A.C., Von Rosen Seeling Stahlke, E., Dornelles, L., Pereira-Ferrari, L., Kremsner, P.G., and Kun, J.F.J. (2007). The association between mannan-binding lectin gene polymorphism and clinical leprosy: new insight into an old paradigm. *J. Infect. Dis.* 196, 1379–1385.
- Miller, E.N., Jamieson, S.E., Joberty, C., Fakiola, M., Hudson, D., Peacock, C.S., Cordell, H.J., Shaw, M.-A., Lins-Lainson, Z., Shaw, J.J., et al. (2004). Genome-wide scans for leprosy and tuberculosis susceptibility genes in Brazilians. *Genes Immun.* 5, 63–67.
- Mira, M.T., Alcais, A., Di Pietrantonio, T., Thuc, N.V., Phuong, M.C., Abel, L., and Schurr, E. (2003a). Segregation of HLA/TNF region is linked to leprosy clinical spectrum in families displaying mixed leprosy subtypes. *Genes Immun* 4, 67–73.

- Mira, M.T., Alcaïs, A., Van Thuc, N., Thai, V.H., Huong, N.T., Ba, N.N., Verner, A., Hudson, T.J., Abel, L., and Schurr, E. (2003b). Chromosome 6q25 is linked to susceptibility to leprosy in a Vietnamese population. *Nat. Genet.* 33, 412–415.
- Mira, M.T., Alcaïs, A., Di Pietrantonio, T., Thuc, N.V., Phuong, M.C., Abel, L., and Schurr, E. (2003c). Segregation of HLA/TNF region is linked to leprosy clinical spectrum in families displaying mixed leprosy subtypes. *Genes Immun* 4, 67–73.
- Mira, M.T., Alcaïs, A., Nguyen, V.T., Moraes, M.O., Di Flumeri, C., Vu, H.T., Mai, C.P., Nguyen, T.H., Nguyen, N.B., Pham, X.K., et al. (2004). Susceptibility to leprosy is associated with PARK2 and PACRG. *Nature* 427, 636–640.
- Misch, E.A., Berrington, W.R., Vary, J.C., Jr, and Hawn, T.R. (2010). Leprosy and the human genome. *Microbiol. Mol. Biol. Rev. MMBR* 74, 589–620.
- Modlin, R.L. (2010). The innate immune response in leprosy. *Curr. Opin. Immunol.* 22, 48–54.
- Moet, F.J., Meima, A., Oskam, L., and Richardus, J.H. (2004). Risk factors for the development of clinical leprosy among contacts, and their relevance for targeted interventions. *Lepr. Rev.* 75, 310–326.
- Moet, F.J., Pahan, D., Schuring, R.P., Oskam, L., and Richardus, J.H. (2006). Physical distance, genetic relationship, age, and leprosy classification are independent risk factors for leprosy in contacts of patients with leprosy. *J. Infect. Dis.* 193, 346–353.
- Monot, M., Honoré, N., Garnier, T., Araoz, R., Coppée, J.-Y., Lacroix, C., Sow, S., Spencer, J.S., Truman, R.W., Williams, D.L., et al. (2005). On the origin of leprosy. *Science* 308, 1040–1042.
- Monot, M., Honoré, N., Garnier, T., Zidane, N., Sherafi, D., Paniz-Mondolfi, A., Matsuoka, M., Taylor, G.M., Donoghue, H.D., Bouwman, A., et al. (2009). Comparative genomic and phylogeographic analysis of *Mycobacterium leprae*. *Nat. Genet.* 41, 1282–1289.
- Moraes, M.O., Pacheco, A.G., Schonkeren, J.J.M., Vanderborght, P.R., Nery, J.A.C., Santos, A.R., Moraes, M.E., Moraes, J.R., Ottenhoff, T.H.M., Sampaio, E.P., et al. (2004). Interleukin-10 promoter single-nucleotide polymorphisms as markers for disease susceptibility and disease severity in leprosy. *Genes Immun.* 5, 592–595.
- Mukherjee, P.K., Das, P., and Rao, P.S.S. (2013). Time trends in MB-PB ratio among untreated leprosy patients attending a referral hospital in UP, India during 2001 to 2010. *Indian J. Lepr.* 85, 59–64.
- Murray, R.A., Siddiqui, M.R., Mendillo, M., Krahenbuhl, J., and Kaplan, G. (2007). *Mycobacterium leprae* inhibits dendritic cell activation and maturation. *J. Immunol. Baltim. Md 1950* 178, 338–344.
- Newell, K.W. (1966). A epidemiologist's view of leprosy. *Bull. World Health Organ.* 34, 827.
- Novembre, J., Johnson, T., Bryc, K., Kutalik, Z., Boyko, A.R., Auton, A., Indap, A., King, K.S., Bergmann, S., Nelson, M.R., et al. (2008). Genes mirror geography within Europe. *Nature* 456, 98–101.
- Ohyama, H., Ogata, K., Takeuchi, K., Namisato, M., Fukutomi, Y., Nishimura, F., Naruishi, H., Ohira, T., Hashimoto, K., Liu, T., et al. (2005). Polymorphism of the 5' flanking region of the IL-12 receptor beta2 gene partially determines the clinical types of leprosy through impaired transcriptional activity. *J. Clin. Pathol.* 58, 740–743.

- Ohyama, H., Kato-Kogoe, N., Nishimura, F., Takeuchi-Hatanaka, K., Matsushita, S., Yamanegi, K., Yamada, N., Hata, M., Nakasho, K., and Terada, N. (2008). Differential effects of polymorphisms in the 5' flanking region of IL12RB2 on NK- and T-cell activity. *J. Interferon Cytokine Res. Off. J. Int. Soc. Interferon Cytokine Res.* 28, 563–569.
- Oskam, L., and Mi, B. (2007). 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.* 78, 173–185.
- Ouederni, M., Vincent, Q.B., Frange, P., Touzot, F., Scerra, S., Bejaoui, M., Bousfiha, A., Levy, Y., Lisowska-Grospierre, B., Canioni, D., et al. (2011). Major histocompatibility complex class II expression deficiency caused by a RFXANK founder mutation: a survey of 35 patients. *Blood* 118, 5108–5118.
- Parascandola, J. (1996). Sulfones and the miracle at Carville. *Rev. Hist. Pharm.* 44, 409–412.
- Pardillo, F.E.F., Fajardo, T.T., Abalos, R.M., Scollard, D., and Gelber, R.H. (2007). Methods for the classification of leprosy for treatment purposes. *Clin. Infect. Dis. Off. Publ. Infect. Dis. Soc. Am.* 44, 1096–1099.
- Pardillo, F.E.F., Burgos, J., Fajardo, T.T., Dela Cruz, E., Abalos, R.M., Paredes, R.M.D., Andaya, C.E.S., and Gelber, R.H. (2008). Powerful bactericidal activity of moxifloxacin in human leprosy. *Antimicrob. Agents Chemother.* 52, 3113–3117.
- Pierce, S.B., Walsh, T., Chisholm, K.M., Lee, M.K., Thornton, A.M., Fiumara, A., Opitz, J.M., Levy-Lahad, E., Klevit, R.E., and King, M.C. (2010). Mutations in the DBP-deficiency protein HSD17B4 cause ovarian dysgenesis, hearing loss, and ataxia of Perrault Syndrome. *Am J Hum Genet* 87, 282–288.
- Di Pietrantonio, T., Hernandez, C., Girard, M., Verville, A., Orlova, M., Belley, A., Behr, M.A., Loredó-Osti, J.C., and Schurr, E. (2010). Strain-specific differences in the genetic control of two closely related mycobacteria. *PLoS Pathog.* 6, e1001169.
- Portaels, F., and de Jong, B. (2014). Nouveaux défis dans la lutte contre la lèpre. *R. Acad. Overseas Sci. communication.*
- Portevin, D., Gagneux, S., Comas, I., and Young, D. (2011). Human macrophage responses to clinical isolates from the Mycobacterium tuberculosis complex discriminate between ancient and modern lineages. *PLoS Pathog.* 7, e1001307.
- Prado-Montes de Oca, E., Velarde-Félix, J.S., Ríos-Tostado, J.J., Picos-Cárdenas, V.J., and Figuera, L.E. (2009). SNP 668C (-44) alters a NF-kappaB1 putative binding site in non-coding strand of human beta-defensin 1 (DEFB1) and is associated with lepromatous leprosy. *Infect. Genet. Evol. J. Mol. Epidemiol. Evol. Genet. Infect. Dis.* 9, 617–625.
- Purcell, S., Neale, B., Todd-Brown, K., Thomas, L., Ferreira, M.A.R., Bender, D., Maller, J., Sklar, P., de Bakker, P.I.W., Daly, M.J., et al. (2007). PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am. J. Hum. Genet.* 81, 559–575.
- Quaresma, J. a. S., Aarão, T.L.S., Sousa, J.R., Botelho, B.S., Barros, L.F., Araujo, R.S., Rodrigues, J.L.P., Prudente, D.L., Pinto, D.S., Carneiro, F.R.O., et al. (2014). Th17 cytokines expression in leprosy skin lesions. *Br. J. Dermatol.*

- Rabinowitz, D. (1997). A transmission disequilibrium test for quantitative trait loci. *Hum. Hered.* 47, 342–350.
- Rabinowitz, D., and Laird, N. (2000). A unified approach to adjusting association tests for population admixture with arbitrary pedigree structure and arbitrary missing marker information. *Hum. Hered.* 50, 211–223.
- Rajalingam, R., Mehra, N.K., and Singal, D.P. (2000). Polymorphism in heat-shock protein 70-1 (HSP70-1) gene promoter region and susceptibility to tuberculoid leprosy and pulmonary tuberculosis in Asian Indians. *Indian J. Exp. Biol.* 38, 658–662.
- Ranque, B. (2007a). Recherche de facteurs génétiques influençant deux phénotypes liés à la lèpre : la réaction granulomateuse de Mitsuda et les réactions de réversion.
- Ranque, B. (2007b). La lèpre, une maladie éliminée... ou négligée ? *Antibiotiques* 99–114.
- Ranque, B., Alcais, A., Thuc, N.V., Woynard, S., Thai, V.H., Huong, N.T., Ba, N.N., Khoa, P.X., Schurr, E., and Abel, L. (2005). A recessive major gene controls the mitsuda reaction in a region endemic for leprosy. *J Infect Dis* 192, 1475–1482.
- Ranque, B., Nguyen, V.T., Vu, H.T., Nguyen, T.H., Nguyen, N.B., Pham, X.K., Schurr, E., Abel, L., and Alcaïs, A. (2007a). Age is an important risk factor for onset and sequelae of reversal reactions in Vietnamese patients with leprosy. *Clin. Infect. Dis. Off. Publ. Infect. Dis. Soc. Am.* 44, 33–40.
- Ranque, B., Alter, A., Mira, M., Thuc, N.V., Thai, V.H., Huong, N.T., Ba, N.N., Khoa, P.X., Schurr, E., Abel, L., et al. (2007b). Genomewide linkage analysis of the granulomatous mitsuda reaction implicates chromosomal regions 2q35 and 17q21. *J. Infect. Dis.* 196, 1248–1252.
- Ranque, B., Alter, A., Schurr, E., Abel, L., and Alcais, A. (2008). [Leprosy: a paradigm for the study of human genetic susceptibility to infectious diseases]. *Médecine Sci.* MS 24, 491–497.
- Rao, P.S., Karat, A.B., Kaliaperumal, V.G., and Karat, S. (1972). Prevalence of leprosy in Gudiyatham Taluk, South India. I. Specific rates with reference to age, sex and type. *Int. J. Lepr. Mycobact. Dis. Off. Organ Int. Lepr. Assoc.* 40, 157–163.
- R Development Core Team (2008). R: A language and environment for statistical computing. R Found. Stat. Comput. Vienna Austria.
- Revelle, W. (2014). psych: Procedures for Personality and Psychological Research. Northwest. Univ. Evanst. Ill. USA *Version* = 1.4.8.
- Reynard, M.P., Turner, D., Junqueira-Kipnis, A.P., Ramos de Souza, M., Moreno, C., and Navarrete, C.V. (2003). Allele frequencies for an interferon-gamma microsatellite in a population of Brazilian leprosy patients. *Eur. J. Immunogenetics Off. J. Br. Soc. Histocompat. Immunogenetics* 30, 149–151.
- Richard, J., Bushanam, R.S., Samuel, P., and Ebenezer, M. (2010). Recent trends in new case detection rates in leprosy by age and sex in Gudiyatham Taluk, Tamil Nadu, India. *Indian J. Lepr.* 82, 131–135.
- Ridley, D.S., and Jopling, W.H. (1966). Classification of leprosy according to immunity. A five-group system. *Int. J. Lepr. Mycobact. Dis. Off. Organ Int. Lepr. Assoc.* 34, 255–273.

- Ridley, D.S., and Waters, M.F. (1969). Significance of variations within the lepromatous group. *Lepr. Rev.* 40, 143–152.
- Risch, N., and Zhang, H. (1995). Extreme discordant sib pairs for mapping quantitative trait loci in humans. *Science* 268, 1584–1589.
- Roach, D.R., Briscoe, H., Saunders, B., France, M.P., Riminton, S., and Britton, W.J. (2001). Secreted lymphotoxin- α is essential for the control of an intracellular bacterial infection. *J. Exp. Med.* 193, 239–246.
- Roeder, K., Bacanu, S.-A., Wasserman, L., and Devlin, B. (2006). Using linkage genome scans to improve power of association in genome scans. *Am. J. Hum. Genet.* 78, 243–252.
- Roy, S., McGuire, W., Mascie-Taylor, C.G., Saha, B., Hazra, S.K., Hill, A.V., and Kwiatkowski, D. (1997). Tumor necrosis factor promoter polymorphism and susceptibility to lepromatous leprosy. *J Infect Dis* 176, 530–532.
- Roy, S., Frodsham, A., Saha, B., Hazra, S.K., Mascie-Taylor, C.G., and Hill, A.V. (1999). Association of vitamin D receptor genotype with leprosy type. *J. Infect. Dis.* 179, 187–191.
- Do Sacramento, W.S., Mazini, P.S., Franceschi, D.A.S., de Melo, F.C., Braga, M.A., Sell, A.M., Tsuneto, L.T., and Visentainer, J.E.L. (2012). Frequencies of MICA alleles in patients from southern Brazil with multibacillary and paucibacillary leprosy. *Int. J. Immunogenet.* 39, 210–215.
- Saini, C., Ramesh, V., and Nath, I. (2013). 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.* 7, e2338.
- Santos, A.R., Suffys, P.N., Vanderborght, P.R., Moraes, M.O., Vieira, L.M.M., Cabello, P.H., Bakker, A.M., Matos, H.J., Huizinga, T.W.J., Ottenhoff, T.H.M., et al. (2002). Role of tumor necrosis factor- α and interleukin-10 promoter gene polymorphisms in leprosy. *J. Infect. Dis.* 186, 1687–1691.
- Dos Santos, D.S., Duppre, N.C., Sales, A.M., Nery, J.A. da C., Sarno, E.N., and Hacker, M.A. (2013). Kinship and Leprosy in the Contacts of Leprosy Patients: Cohort at the Souza Araújo Outpatient Clinic, Rio de Janeiro, RJ, 1987-2010. *J. Trop. Med.* 2013, 596316.
- Sapkota, B.R., Macdonald, M., Berrington, W.R., Misch, E.A., Ranjit, C., Siddiqui, M.R., Kaplan, G., and Hawn, T.R. (2010). Association of TNF, MBL, and VDR polymorphisms with leprosy phenotypes. *Hum. Immunol.* 71, 992–998.
- Sartor, R.B. (2005). Does *Mycobacterium avium* subspecies paratuberculosis cause Crohn's disease? *Gut* 54, 896–898.
- Schepers, K., Schandené, L., Bustamante, J., Van Vooren, J.-P., de Suremain, M., Casanova, J.-L., Yombi, J.C., Jacobs, F., Mascart, F., and Goffard, J.-C. (2013). IL-12R β 1 deficiency and disseminated *Mycobacterium tuberculosis* disease. *J. Clin. Immunol.* 33, 1285–1288.
- Scollard, D.M. (2005). Leprosy research declines, but most of the basic questions remain unanswered. *Int. J. Lepr. Mycobact. Dis. Off. Organ Int. Lepr. Assoc.* 73, 25–27.
- Scollard, D.M., Adams, L.B., Gillis, T.P., Krahenbuhl, J.L., Truman, R.W., and Williams, D.L. (2006). The continuing challenges of leprosy. *Clin. Microbiol. Rev.* 19, 338–381.

- Setia, M.S., Steinmaus, C., Ho, C.S., and Rutherford, G.W. (2006). The role of BCG in prevention of leprosy: a meta-analysis. *Lancet Infect. Dis.* 6, 162–170.
- Sharma, P.V. (1999). *Sushruta-Samhita: with the English Translation of Text and Dalhana's Commentary Along with Critical Notes* (Varanasi, India).
- Sharma, R., Lahiri, R., Scollard, D.M., Pena, M., Williams, D.L., Adams, L.B., Figarola, J., and Truman, R.W. (2013). The armadillo: a model for the neuropathy of leprosy and potentially other neurodegenerative diseases. *Dis. Model. Mech.* 6, 19–24.
- Shepard, C.C. (1960). The experimental disease that follows the injection of human leprosy bacillus into foot-pads of mice. *J. Exp. Med.* 112, 445–454.
- Shugart, Y.Y., Silverberg, M.S., Duerr, R.H., Taylor, K.D., Wang, M.-H., Zarfes, K., Schumm, L.P., Bromfield, G., Steinhart, A.H., Griffiths, A.M., et al. (2008). An SNP linkage scan identifies significant Crohn's disease loci on chromosomes 13q13.3 and, in Jewish families, on 1p35.2 and 3q29. *Genes Immun.* 9, 161–167.
- Siddiqui, M.R., Meisner, S., Tosh, K., Balakrishnan, K., Ghei, S., Fisher, S.E., Golding, M., Shanker Narayan, N.P., Sitaraman, T., Sengupta, U., et al. (2001). A major susceptibility locus for leprosy in India maps to chromosome 10p13. *Nat Genet* 27, 439–441.
- Sieling, P.A., and Modlin, R.L. (1994). Cytokine patterns at the site of mycobacterial infection. *Immunobiology* 191, 378–387.
- Singh, P., and Cole, S.T. (2011). *Mycobacterium leprae*: genes, pseudogenes and genetic diversity. *Future Microbiol.* 6, 57–71.
- Smith, D.G. (1979). The genetic hypothesis for susceptibility to lepromatous leprosy. *Hum. Genet.* 50, 163–177.
- Spielman, R.S., McGinnis, R.E., and Ewens, W.J. (1993). Transmission test for linkage disequilibrium: the insulin gene region and insulin-dependent diabetes mellitus (IDDM). *Am. J. Hum. Genet.* 52, 506–516.
- Strober, W., and Watanabe, T. (2011). NOD2, an intracellular innate immune sensor involved in host defense and Crohn's disease. *Mucosal Immunol.* 4, 484–495.
- Teles, R.M.B., Graeber, T.G., Krutzik, S.R., Montoya, D., Schenk, M., Lee, D.J., Komisopoulou, E., Kelly-Scumpia, K., Chun, R., Iyer, S.S., et al. (2013). Type I interferon suppresses type II interferon-triggered human anti-mycobacterial responses. *Science* 339, 1448–1453.
- Thakur, G.A., Sengupta, S.M., Grizenko, N., Choudhry, Z., and Joobee, R. (2012). Comprehensive phenotype/genotype analyses of the norepinephrine transporter gene (SLC6A2) in ADHD: relation to maternal smoking during pregnancy. *PloS One* 7, e49616.
- Thuc, N.V., Abel, L., Lap, V.D., Oberti, J., and Lagrange, P.H. (1994). Protective effect of BCG against leprosy and its subtypes: a case-control study in southern Vietnam. *Int. J. Lepr. Mycobact. Dis. Off. Organ Int. Lepr. Assoc.* 62, 532–538.

- Tobin, D.M., Vary, J.C., Jr, Ray, J.P., Walsh, G.S., Dunstan, S.J., Bang, N.D., Hagge, D.A., Khadge, S., King, M.-C., Hawn, T.R., et al. (2010). The *Ita4h* locus modulates susceptibility to mycobacterial infection in zebrafish and humans. *Cell* **140**, 717–730.
- Tosh, K., Meisner, S., Siddiqui, M.R., Balakrishnan, K., Ghei, S., Golding, M., Sengupta, U., Pitchappan, R.M., and Hill, A.V.S. (2002). A region of chromosome 20 is linked to leprosy susceptibility in a South Indian population. *J. Infect. Dis.* **186**, 1190–1193.
- Truman, R.W., Ebenezer, G.J., Pena, M.T., Sharma, R., Balamayooran, G., Gillingwater, T.H., Scollard, D.M., McArthur, J.C., and Rambukkana, A. (2014). The armadillo as a model for peripheral neuropathy in leprosy. *ILAR J. Natl. Res. Counc. Inst. Lab. Anim. Resour.* **54**, 304–314.
- Turner, S. (2014). qqman: Q-Q and manhattan plots for GWAS data. R package version 0.1.2.
- Uzoigwe, J.C., Khaita, M.L., and Gibbs, P.S. (2007). Epidemiological evidence for *Mycobacterium avium* subspecies paratuberculosis as a cause of Crohn's disease. *Epidemiol. Infect.* **135**, 1057–1068.
- Vejbæsya, S., Mahaisavariya, P., Luangtrakool, P., and Sermduangprateep, C. (2007). TNF alpha and NRAM1 polymorphisms in leprosy. *J. Med. Assoc. Thail. Chotmaihet Thangphaet* **90**, 1188–1192.
- Vera-Cabrera, L., Escalante-Fuentes, W.G., Gomez-Flores, M., Ocampo-Candiani, J., Busso, P., Singh, P., and Cole, S.T. (2011). Case of diffuse lepromatous leprosy associated with "Mycobacterium lepromatosis." *J. Clin. Microbiol.* **49**, 4366–4368.
- Vincent, Q. (2014). *Epidémiologie et Génétique Humaine de l'ulcère de Buruli*. Thèse *Université Paris Descartes*.
- Vincent, Q., Alcais, A., Alter, A., Schurr, E., and Abel, L. (2006). Quantifying genomic imprinting in the presence of linkage. *Biometrics* **62**, 1071–1080.
- Voorend, C.G.N., and Post, E.B. (2013). A Systematic Review on the Epidemiological Data of Erythema Nodosum Leprosum, a Type 2 Leprosy Reaction. *PLoS Negl. Trop. Dis.* **7**, e2440.
- De Vries, J.L., and Perry, B.H. (1985). Leprosy case detection rates by age, sex, and polar type under leprosy control conditions. *Am. J. Epidemiol.* **121**, 403–413.
- Wade, H.W. (1935). Tuberculoid changes in leprosy IV. Classification of Tuberculoid leprosy. *Int. J. Lepr.* **3**, 121–136.
- Wade, H.W. (1953). The original publications on the Mitsuda test. *Int. J. Lepr.* **21**, 363–364.
- Wagener, D.K., Schauf, V., Nelson, K.E., Scollard, D., Brown, A., and Smith, T. (1988). Segregation analysis of leprosy in families of northern Thailand. *Genet. Epidemiol.* **5**, 95–105.
- Waldman, I.D., Robinson, B.F., and Rowe, D.C. (1999). A logistic regression based extension of the TDT for continuous and categorical traits. *Ann Hum Genet* **63**, 329–340.
- Wallace, C., Fitness, J., Hennig, B., Sichali, L., Mwaungulu, L., Pönnighaus, J.M., Warndorff, D.K., Clayton, D., Fine, P.E.M., and Hill, A.V.S. (2004). Linkage analysis of susceptibility to leprosy type using an IBD regression method. *Genes Immun.* **5**, 221–225.

- Wang, L.M., Kimura, A., Satoh, M., and Mineshita, S. (1999). HLA linked with leprosy in southern China: HLA-linked resistance alleles to leprosy. *Int. J. Lepr. Mycobact. Dis. Off. Organ Int. Lepr. Assoc.* 67, 403–408.
- WHO (1982). Chemotherapy of leprosy control programmes.
- WHO (2008). Trends in the epidemiology of leprosy - Viet Nam, 1983-2006. *Wkly. Epidemiol. Rec.* 83, 217–224.
- WHO (2013). Global leprosy: update on the 2012 situation. *Wkly. Epidemiol. Rec.* 365–380.
- WHO (2014a). Fact sheet n°101.
- WHO (2014b). Global leprosy updates, 2013 : reducing disease burden. *Wkly. Epidemiol. Rec.* 36, 389–400.
- WHO Expert Committee on Leprosy (1988). Sixth Report.
- WHO Expert Committee on Leprosy (1998). Seventh report. *World Health Organ. Tech. Rep. Ser.* 874, 1–43.
- Wibawa, T., Soebono, H., and Matsuo, M. (2002). Association of a missense mutation of the laminin alpha2 gene with tuberculoid type of leprosy in Indonesian patients. *Trop. Med. Int. Health TM IH* 7, 631–636.
- Wise, A.L., Gyi, L., and Manolio, T.A. (2013). eXclusion: toward integrating the X chromosome in genome-wide association analyses. *Am. J. Hum. Genet.* 92, 643–647.
- Won, S., Wilk, J.B., Mathias, R.A., O'Donnell, C.J., Silverman, E.K., Barnes, K., O'Connor, G.T., Weiss, S.T., and Lange, C. (2009). On the analysis of genome-wide association studies in family-based designs: a universal, robust analysis approach and an application to four genome-wide association studies. *PLoS Genet.* 5, e1000741.
- Wong, S.H., Gochhait, S., Malhotra, D., Pettersson, F.H., Teo, Y.Y., Khor, C.C., Rautanen, A., Chapman, S.J., Mills, T.C., Srivastava, A., et al. (2010). Leprosy and the adaptation of human toll-like receptor 1. *PLoS Pathog.* 6, e1000979.
- Yang, T.-L., Xiong, D.-H., Guo, Y., Recker, R.R., and Deng, H.-W. (2006). Association analyses of CYP19 gene polymorphisms with height variation in a large sample of Caucasian nuclear families. *Hum. Genet.* 120, 119–125.
- Zhang, F., Liu, H., Chen, S., Low, H., Sun, L., Cui, Y., Chu, T., Li, Y., Fu, X., Yu, Y., et al. (2011). Identification of two new loci at IL23R and RAB32 that influence susceptibility to leprosy. *Nat. Genet.* 43, 1247–1251.
- Zhang, F.-R., Huang, W., Chen, S.-M., Sun, L.-D., Liu, H., Li, Y., Cui, Y., Yan, X.-X., Yang, H.-T., Yang, R.-D., et al. (2009a). Genomewide association study of leprosy. *N. Engl. J. Med.* 361, 2609–2618.
- Zhang, F.-R., Huang, W., Chen, S.-M., Sun, L.-D., Liu, H., Li, Y., Cui, Y., Yan, X.-X., Yang, H.-T., Yang, R.-D., et al. (2009b). Genomewide association study of leprosy. *N. Engl. J. Med.* 361, 2609–2618.

- Zhang, F.-R., Liu, H., Irwanto, A., Fu, X.-A., Li, Y., Yu, G.-Q., Yu, Y.-X., Chen, M.-F., Low, H.-Q., Li, J.-H., et al. (2013). HLA-B*13:01 and the dapsone hypersensitivity syndrome. *N. Engl. J. Med.* 369, 1620–1628.
- Zhou, X., and Stephens, M. (2012). Genome-wide efficient mixed-model analysis for association studies. *Nat. Genet.* 44, 821–824.

Annexes

Démonstrations

Démonstration 1. $FBAT_{+1/-1}$ est un test d'hétérogénéité Q de Cochran

$$\begin{aligned} FBAT_{+1/-1} &= \frac{((S_1 - E_1) - (S_2 - E_2))^2}{V_1 + V_2} \\ &= \frac{V_1 V_2 ((S_1 - E_1)^2 - 2(S_1 - E_1)(S_2 - E_2) + (S_2 - E_2)^2)}{V_1 V_2 (V_1 + V_2)} \\ &= \frac{(V_1 + V_2)(V_2(S_1 - E_1)^2 + V_1(S_2 - E_2)^2) - (V_2(S_1 - E_1) + V_1(S_2 - E_2))^2}{V_1 V_2 (V_1 + V_2)} \\ &= \frac{(V_1 + V_2)(V_2(S_1 - E_1)^2 + V_1(S_2 - E_2)^2)}{V_1 V_2 (V_1 + V_2)} - \frac{V_1^2 V_2^2 \left(\frac{S_1 - E_1}{V_1} + \frac{S_2 - E_2}{V_2} \right)^2}{V_1 V_2 (V_1 + V_2)} \\ &= \frac{(S_1 - E_1)^2}{V_1} + \frac{(S_2 - E_2)^2}{V_2} - \frac{\left(\frac{S_1 - E_1}{V_1} + \frac{S_2 - E_2}{V_2} \right)^2}{\frac{1}{V_1} + \frac{1}{V_2}} \\ &= \text{Cochran's } Q(S_1, S_2) \end{aligned}$$

Démonstration 2. Equivalences de codage dans $FBAT$ sous un modèle additif

Soient N_1 (N_2) le nombre de trios dont l'enfant est de phénotype D_1 (D_2) ; puis N_{id} (N_{is}) le nombre de trios avec deux (d) ou un seul (s) parent(s) hétérozygote(s). Soit n_i le nombre de parents hétérozygotes. Alors

$$n_i = 2N_{id} + N_{is}$$

Soient V_s et V_d les variances unitaires d'un trio avec un ou deux parent(s) hétérozygote(s).

Pour $FBAT$ et $FBAT_{het}$, $V_s = 1/4$ et $V_d = 1/2 = 2V_s$. Alors

$$V_1 = \sum_{\substack{\text{pheno 1} \\ \& \text{ 1 parent}}} \text{Var}(S_j) + \sum_{\substack{\text{pheno 1} \\ \& \text{ 2 parents}}} \text{Var}(S_j) = N_{1s}V_s + N_{1d}V_d = \frac{N_{1s}}{4} + \frac{N_{1d}}{2} = \frac{n_1}{4}$$

$$\text{et } V_2 = \sum_{\substack{\text{pheno 2} \\ \& \text{ 1 parent}}} \text{Var}(S_j) + \sum_{\substack{\text{pheno 2} \\ \& \text{ 2 parents}}} \text{Var}(S_j) = \frac{n_2}{4}$$

Etant donné que $\text{FBAT}(T_1 = x, T_2 = y) = \text{FBAT}(T_1 = kx, T_2 = ky)$, coder les phénotypes

$T_1 = \frac{1}{v_1}$ et $T_2 = -\frac{1}{v_2}$ est équivalent à les coder $T_1 = \frac{1}{n_1}$ et $T_2 = -\frac{1}{n_2}$ (idem pour FBAT_{het}).

Pour $\text{FBAT}_{\text{het free}}$, $V_s = \left(1 - \frac{n_A}{N}\right) \frac{n_A}{N}$ et $V_d = 2 \left(1 - \frac{n_A}{N}\right) \frac{n_A}{N} = 2V_s$. Alors

$$V_1 = \sum_{\substack{\text{pheno 1} \\ \& \text{ 1 parent}}} \text{Var}(S_j) + \sum_{\substack{\text{pheno 1} \\ \& \text{ 2 parents}}} \text{Var}(S_j) = (N_{1s} + 2N_{1d}) \left(1 - \frac{n_A}{N}\right) \frac{n_A}{N} = n_1 \left(1 - \frac{n_A}{N}\right) \frac{n_A}{N}$$

$$\text{et } V_2 = \sum_{\substack{\text{pheno 2} \\ \& \text{ 1 parent}}} \text{Var}(S_j) + \sum_{\substack{\text{pheno 2} \\ \& \text{ 2 parents}}} \text{Var}(S_j) = n_2 \left(1 - \frac{n_A}{N}\right) \frac{n_A}{N}$$

Donc coder les phénotypes $T_1 = \frac{1}{v_1}$ et $T_2 = -\frac{1}{v_2}$ est équivalent à les coder $T_1 = \frac{1}{n_1}$ et $T_2 = -\frac{1}{n_2}$ pour $\text{FBAT}_{\text{het free}}$.

Démonstration 3. Offset qui minimise la variance sous un modèle additif

Nous nous plaçons dans la situation d, un phénotype binaire codé $T_1 = 1$ et $T_2 = 0$.

Soit μ l'offset. Alors

$$T_1 = 1 - \mu \text{ et } T_2 = -\mu$$

Avec les mêmes notations que dans la démonstration 2,

$$\begin{aligned} \text{Var} &= \sum_{\substack{\text{pheno 1} \\ \& \text{ 1 parent}}} \text{Var}(S_j) + \sum_{\substack{\text{pheno 1} \\ \& \text{ 2 parents}}} \text{Var}(S_j) + \sum_{\substack{\text{pheno 2} \\ \& \text{ 1 parent}}} \text{Var}(S_j) + \sum_{\substack{\text{pheno 2} \\ \& \text{ 2 parents}}} \text{Var}(S_j) \\ &= N_{1s}T_1^2V_s + N_{1d}T_1^2V_d + N_{2s}T_2^2V_s + N_{2d}T_2^2V_d \\ &= (1 - \mu)^2(N_{1s}V_s + N_{1d}V_d) + (-\mu)^2(N_{2s}V_s + N_{2d}V_d) \end{aligned}$$

Pour FBAT et FBAT_{het}, $V_s = 1/4$, $V_d = 1/2 = 2V_s$ et

$$\text{Var} = \frac{1}{4}((1 - \mu)^2 n_1 + \mu^2 n_2)$$

et $\min_{\mu}(\text{Var}) = \min_{\mu}((1 - \mu)^2 n_1 + \mu^2 n_2)$ est obtenu pour $\mu = \frac{n_1}{n_1 + n_2}$

Pour FBAT_{het free}, $V_s = \left(1 - \frac{n_A}{N}\right) \frac{n_A}{N}$, $V_d = 2 \left(1 - \frac{n_A}{N}\right) \frac{n_A}{N} = 2V_s$ et

$$\text{Var} = \left(1 - \frac{n_A}{N}\right) \frac{n_A}{N} ((1 - \mu)^2 n_1 + \mu^2 n_2)$$

et $\min_{\mu}(\text{Var}) = \min_{\mu}((1 - \mu)^2 n_1 + \mu^2 n_2)$ est aussi obtenu pour $\mu = \frac{n_1}{n_1 + n_2}$

Démonstration 4. FBAT_{het free} est un test du χ^2 de Pearson

Avec les notations du manuscrit, écrivons le tableau de contingence de la transmission des allèles A et a à deux groupes de phénotype D₁ et D₂.

Transmission	A	a	Total
D ₁	n_{A1}	n_{a1}	n_1
D ₂	n_{A2}	n_{a2}	n_2
	n_A	n_a	N

χ^2 de Pearson($n_{A1}, n_{a1}, n_{a2}, n_{A2}$)

$$\begin{aligned}
 &= \frac{(n_{A1}N - n_A n_1)^2}{N n_A n_1} + \frac{(n_{a1}N - n_a n_1)^2}{N n_a n_1} + \frac{(n_{A2}N - n_A n_2)^2}{N n_A n_2} + \frac{(n_{a2}N - n_a n_2)^2}{N n_a n_2} \\
 &= \left(\frac{1}{N n_A n_1} + \frac{1}{N n_a n_1} + \frac{1}{N n_A n_2} + \frac{1}{N n_a n_2} \right) (n_{A1} n_{a2} - n_{a1} n_{A2})^2 \\
 &= \frac{N}{n_1 n_2 n_A n_a} (n_{A1} n_{a2} - n_{a1} n_{A2})^2 \\
 &= \text{FBAT}_{\text{het free}}
 \end{aligned}$$

Démonstration 5. $FBAT_{het} = \rho FBAT_{het\ free}$

Avec les notations du manuscrit, nous avons, pour $FBAT_{het}$,

$$S - E = \frac{1}{n_1} \left(n_{A1} \left(1 - \frac{1}{2} \right) + n_{a1} \left(0 - \frac{1}{2} \right) \right) - \frac{1}{n_2} \left(n_{A2} \left(1 - \frac{1}{2} \right) + n_{a2} \left(0 - \frac{1}{2} \right) \right)$$

$$\text{et } V = \frac{1}{4} \left(\frac{1}{n_1} \right)^2 (n_{A1} + n_{a1}) + \frac{1}{4} \left(\frac{1}{n_2} \right)^2 (n_{A2} + n_{a2}) = \left(\frac{1}{n_1} \right)^2 \frac{n_1}{4} + \left(\frac{1}{n_2} \right)^2 \frac{n_2}{4}$$

$$\begin{aligned} \text{Alors } FBAT_{het} &= \frac{(S - E)^2}{V} = \frac{\left(\frac{n_{A1}}{n_1} \left(1 - \frac{1}{2} \right) + \frac{n_{a1}}{n_1} \left(0 - \frac{1}{2} \right) - \frac{n_{A2}}{n_2} \left(1 - \frac{1}{2} \right) - \frac{n_{a2}}{n_2} \left(0 - \frac{1}{2} \right) \right)^2}{\frac{1}{4} \frac{n_1}{n_1^2} + \frac{1}{4} \frac{n_2}{n_2^2}} \\ &= \frac{4(n_{A1}n_{a2} - n_{a1}n_{A2})^2}{Nn_1n_2} = \frac{4}{N^2} \frac{n_A n_a}{n_A n_a n_1 n_2} \frac{N(n_{A1}n_{a2} - n_{a1}n_{A2})^2}{n_A n_a n_1 n_2} = \rho FBAT_{het\ free} \\ &\quad \text{avec } \rho = \frac{4n_A(N-n_A)}{N^2} \end{aligned}$$

Protocole d'implémentation de $FBAT_{het\ free}$

Notre protocole décrit les différentes étapes à effectuer à partir d'un fichier nomdefichier au format PLINK (.bed, .bim, .fam), dans le répertoire rep. Nous utilisons les logiciels R et PLINK sous un système d'exploitation LINUX. Pour calculer la statistique $FBAT_{het\ free}$, nous devons calculer n_{A1} , n_{a1} , n_{A2} et n_{a2} , qui sont le nombre de transmissions d'allèles A et a par les parents à leurs enfants de phénotype D_1 et D_2 . $FBAT_{het\ free}$ sera alors simplement la statistique du χ^2 issue du tableau de contingence à 4 cases : n_{A1} , n_{a1} , n_{A2} et n_{a2} (cf démonstration 4). Le logiciel PLINK calcule facilement ces 4 valeurs.

1^{ère} étape : créer deux fichiers, un pour chaque sous-phénotype

Créer un fichier pour chaque phénotype D_1 et D_2 , **pheno1** et **pheno2**, à 3 colonnes sans en-têtes. Les deux premières colonnes sont les identifiants familial et individuel, et la troisième le phénotype, codé comme par défaut dans PLINK : 2 pour un enfant atteint (qu'il soit de phénotype D_1 ou D_2), 1 pour un non-atteint, 0 pour un phénotype manquant.

2^{ème} étape : lancer la fonction TDT de PLINK

Commandes LINUX :

```
plink --noweb --bfile nomdefichier --pheno pheno1 --tdt --out pheno1
```

```
plink --noweb --bfile nomdefichier --pheno pheno2 --tdt --out pheno2
```

En sortie, deux fichiers pheno1.tdt et pheno2.tdt.

3^{ème} étape : combiner les deux sorties pour obtenir la valeur de p du test " $FBAT_{het\ free}$ "

Transformer les fichiers de sortie PLINK en des tableaux compatibles avec R en utilisant le script `./tdt_to_R` (voir ci-dessous), à l'aide des commandes LINUX :

```
./tdt_to_R.sh pheno1.tdt
```

```
./tdt_to_R.sh pheno2.tdt
```

en sortie, 2 fichiers *pheno1.txt* et *pheno2.txt*.

Puis, dans R :

```
tdt1 <- read.table(file = "rep/pheno1.txt", sep = "\t", header = T)
```

```
tdt2 <- read.table(file = "rep/pheno2.txt", sep = "\t", header = T)
```

```
tdtmerge <- merge(subset(tdt1, select = c(SNP, nA, na)), \
```

```

subset(tdt2,select = c(SNP,nA,na)),by = "SNP")
nA1 <- tdtmerge$NA.x
nA2 <- tdtmerge$NA.y
na1 <- tdtmerge$na.x
na2 <- tdtmerge$na.y
Chi <- (nA1 + nA2 + na1 + na2)*(nA1*na2 - nA2*na1)^2 \
      /((nA1 + nA2)*(na1 + na2)*(nA1 + na1)*(nA2 + na2))
P <- 1 - pchisq(Chi,1)

```

P est le vecteur des valeurs de p pour $FBAT_{het\ free}$.

Script bash: tdt_to_R.sh

```

# Ce script convertit les fichiers .tdt (issu de la fonction --tdt de PLINK)
# en des fichiers .txt à 3 colonnes :
# - colonne 1 : nom du SNP
# - colonne 2 : nA (= nombre de transmissions de l'allèle majeur)
# - colonne 3 : na (= nombre de transmissions de l'allèle mineur)
#!/bin/bash
for nom in `ls *.tdt`                # transforme les fichiers .tdt du répertoire
do
    prefixe = $(echo "$nom" | cut -f1 -d '.') # prefixe est le nom du fichier *.tdt
    echo conversion of $nom in process        # affiche le fichier en cours de création
    sed 's/ */ /g' $nom > int1                # supprime les espaces multiples
    sed 's/ //' int1 > int2                   # supprime le premier espace
    sed -e 's/ ^/g' int2 > $prefixe.txt       # convertit les espaces en tabulations
done
rm int1 int2                               # efface les fichiers intermédiaires

```

Article 1

Pauci- and multibacillary leprosy: two distinct genetically neglected diseases

J. Gaschignard, A.V. Grant, N.V. Thuc, M. Orlova, N.T. Huong, N.N. Ba, V.H. Thai, L. Abel,
E. Schurr and A. Alcaïs

soumis à Plos Neglected Tropical Diseases

Pauci- and multibacillary leprosy: two distinct genetically neglected diseases

Jean Gaschignard ^{1,2}, Audrey V. Grant ^{1,2,3}, Nguyen Van Thuc ⁴, Marianna Orlova ⁵,
Nguyen Thu Huong ⁴, Nguyen Ngoc Ba ⁴, Vu Hong Thai ⁴, Laurent Abel ^{1,2},
Erwin Schurr ^{5,6}, Alexandre. Alcaïs ^{1,2,7 *}

¹ Laboratory of Human Genetics of Infectious Diseases, Necker Branch, INSERM U1163, Necker Hospital for Sick Children, INSERM, Paris, France, EU.

² Paris Descartes University, Imagine Institute, Paris, France, EU.

³ Unité de Génétique fonctionnelle des maladies infectieuses, Institut Pasteur, Paris, France

⁴ Hospital for Dermato-Venerology, Ho Chi Minh City, Vietnam

⁵ Program in Infectious Diseases and Immunity in Global Health, The Research Institute of the McGill University Health Centre, Montreal, Quebec, Canada

⁶ The McGill International TB Centre, Departments of Human Genetics and Medicine, McGill University, Montreal, Quebec, Canada

⁷ URC, CIC, Necker and Cochin Hospitals, Paris, France

* Corresponding author. Phone: +33 1 42 75 43 13, E-mail: alexandre.alcais@inserm.fr

Abstract

After sustained exposure to *Mycobacterium leprae* and despite quasi-clonality of global isolates, only a subset of exposed individuals develop clinical leprosy. Moreover, leprosy patients show a wide spectrum of clinical manifestations that extend from the paucibacillary (PB) to the multibacillary (MB) form of the disease. This ‘polarization’ of leprosy has long been a major focus of investigation for immunologists because it is an excellent example of the clinical impact in natura of differential immune responses of B and T lymphocytes on the form of disease. But while there is strong evidence to support that the development of leprosy per se is under tight human genetic control, remarkably few epidemiological or genetic studies have focused on leprosy subtypes.

Here, we collected available data on the epidemiology of leprosy polarization and the possible role of human genetics in its pathophysiology. In addition, we took advantage of a large and long-lasting field study carried out in Vietnam to validate some hypotheses and propose future research directions. For example, we show that males who developed leprosy at an older age display a significantly higher proportion of MB forms as compared to younger affected individuals and that this impact of age is more pronounced in men than in women. Multivariate analysis revealed that the effects of these two factors (age and sex) on leprosy polarization were independent of the ethnic background of patients.

At the genetic level, linkage and association studies suggested that a number of genes modulate the onset of a particular leprosy subtype (MB or PB) or influence the process of leprosy polarization (impact on all clinical forms but not leprosy per se). To date, however, only five genes have been convincingly associated with leprosy subtypes, *i.e.* found significant in at least two independent samples: MBL2, MRC1, TLR2, TNF and MICA. We assembled a list of 36 genes from the literature that are associated with leprosy subtypes or implicated in the polarization process. This list provided proof-of-principle that, among infected individuals, some genes act specifically in driving the immune response towards PB or MB forms of leprosy.

A review of reported genetic associations with leprosy subtypes revealed that improved study designs are needed for a better understanding of the genetic architecture. The two key points are that physicians and geneticists need to jointly design studies to identify genetic factors involved in leprosy polarization, and future investigations should not be restricted to a sub-analysis of leprosy per se studies, but instead contrast MB to PB individuals. We show the latter the latter approach to be the most powerful design for the identification of genetic polarization

determinants. Finally, we bring to light the important resource represented by the nine-banded armadillo model, a unique animal model for leprosy.

Introduction

Leprosy is a chronic infectious disease caused by *Mycobacterium leprae*, with almost 200,000 new cases detected each year, often leading to severe disabilities and social stigma. After sustained exposure to *M. leprae*, only a subset of individuals develops clinical leprosy. From the early observations of familial aggregation of leprosy cases to the most recent genome-wide association studies identifying genetic polymorphisms associated with leprosy, there is strong evidence that the development of the disease is under tight human genetic control [1,2]. The interest in the genetics of leprosy is reflected by the continuous increase in the number of genetic publications (host and pathogen), while publications on immunology of leprosy have decreased significantly after 1980 [3].

However, there is another level of inter-individual variability in leprosy susceptibility, which is considerable clinical variability among leprosy patients, with a clinical spectrum that extends from the polar ‘tuberculoid’ to the polar ‘lepromatous’ form of the disease [4,5]. The tuberculoid form is characterized by a small number of hypopigmented, well-bordered, anesthetic skin lesions with a low bacillary load (therefore this form is commonly referred to as ‘paucibacillary’), early peripheral nerve impairment, and a Th1 mediated immune response. In contrast, the lepromatous form is characterized by numerous, infiltrated skin lesions displaying high bacillary loads (therefore referred to as ‘multibacillary’), impaired peripheral nerves, possible involvement of internal organs and a Th2 mediated immune response. Leprosy has long been a major focus of investigation for immunologists because it is an excellent example in natura of the clinical impact of a differential immune response of B and T lymphocytes on the form of disease, but it has not been thoroughly considered by geneticists. Some genetic studies have focused specifically on either ‘paucibacillary’ (PB) leprosy or ‘multibacillary’ (MB) leprosy; however, very few have aimed at the process that we call ‘polarization’, which is the process that drives the differentiation to different subtypes and therefore impacts on all or at least multiple subtypes of leprosy.

From a genetic perspective, polarization can thus be regarded as neglected phenotype. To address this shortcoming, we have conducted a systematic review of epidemiological studies addressing leprosy subtypes and polarization. In addition, we took advantage of a large and long-lasting field study in Vietnam to illustrate important concepts discussed in this review. We argue that the poor understanding of the genetic architecture of leprosy polarization is mainly caused by inappropriate study designs. We provide suggestions for future research on PB and MB leprosy, and propose that the use of the armadillo model will be of pivotal importance for the study of these phenotypes.

Methodology

Systematic review

On December 15, 2013, we searched PubMed with the search terms “(leprosy OR *Mycobacterium leprae* OR multibacillary OR paucibacillary OR lepromatous OR tuberculoid NOT tuberculosis) AND (genetic OR association OR risk factors)” and identified ~ 1,400 publications on leprosy (~ 1,000 on genetics and ~ 400 on epidemiology). Epidemiological data were also gathered from the Weekly Epidemiological Records released by the World Health Organization (WHO).

Personal data

We also took advantage in this review of research conducted by the authors. Over the past twenty years (1990-2009) and in close collaboration with the hospital for Dermato-Venereology of Ho-Chi-Minh City (Vietnam), we recruited a large sample of families (parents and children) with at least one child diagnosed with leprosy. The diagnosis of leprosy patients and their subtype classification according to Ridley and Jopling were based on clinical and histological criteria [6]. Besides, physicians also determined a clinical ‘multibacillary’ or ‘paucibacillary’ status to deliver the appropriate treatment. Lepromin testing was not systematically used in the classification. Bacillary index was used when available and diagnosis (of leprosy and its subtype) was systematically established by two independent physicians. Informed consent was obtained from all study participants. The study was approved by institutional review boards and health authorities in Ho Chi Minh City, Vietnam, and The Research Institute of the McGill University Health Centre, Montreal, Canada.

Leprosy classification

As early as 600 BC, the Sushruta-Samhita (the most ancient Indian medical book), distinguished kushta, a lepromatous form of the disease, from rakta, a neural form [7]. In 1881, Danielssen and Boeck distinguished a nodular and an anesthetic leprosy [8], and in 1895, Hansen and Looft pinpointed the term ‘anesthetic’ to ‘maculo-anesthetic’ [9]. In 1935, Wade noticed that some patients could move from a tuberculoid state to a lepromatous one, and named ‘borderline’ the form of leprosy of these patients [10]. Nowadays, once the diagnosis of leprosy is made, the identification of the clinical form is critical because it will determine the length of multidrug therapy [11]. Similarly, the clinical determination of leprosy subtype provides major pathophysiological information, which was early explored by immunologists but as already mentioned somewhat disregarded by the genetics community.

International Leprosy Conferences in Cairo (1938), La Havana (1948) and Madrid (1953) paved the way for the sub-classification of the disease in three classes, based on clinical (aspect of lesions), bacteriological (identification of bacilli), immunological (lepromin test) and histological (nature of the infiltrate, grouping of bacilli) criteria, known as the Madrid definition. The three Madrid classes are: Tuberculoid (T), Borderline (or Dimorphic) (B) and Lepromatous (or Virchowian) (L). The Madrid classification scheme is still used in some countries like Brazil [12]. A major development came with the leprosy classification by Ridley and Jopling in 1966, that aimed to better characterize the borderline group that could comprise patients with different clinical, immunological, bacteriological or histological findings. Ridley and Jopling proposed five subtypes: Tuberculoid Tuberculoid (TT), Borderline Tuberculoid (BT), Borderline Borderline (BB), Borderline Lepromatous (BL) and Lepromatous Lepromatous (LL) [6]. The latter form may itself be refined in ‘subpolar’ (LLs) and ‘polar’ (LLp) on a clinico-pathologic basis [13].

To facilitate diagnosis and treatment of leprosy in the field, the World Health Organization (WHO) proposed in 1982 a simplified classification. The WHO-82 classification distinguishes PB leprosy, which includes TT and BT patients with bacillary index (BI) ≤ 1 , from MB leprosy which comprises BB, BL and LL patients, as well as those with a BI ≥ 2 [11]. In 1988, WHO changed its bacteriological criteria and any patient with a positive BI was now classified as MB (‘WHO-88’ definition) [14]. Finally, in 1996, WHO switched for an operational definition to optimize treatment allocation: patients with less than five skin lesions are diagnosed as PB, while all others are diagnosed as MB (‘WHO-96’ definition). A side effect is that this

operationally driven classification has eclipsed pathophysiological aspects differentiating the leprosy forms. A summary of all classifications is given in Box 1.

To which extent the evolution of the WHO classification system impacted on the counts of PB and MB forms is unclear. Therefore, we took advantage of our large sample from Vietnam to assess the concordance between these classifications. Between 1990 and 2009, we enrolled a total of 2,784 individuals, including 1,182 cases of leprosy. Data was available for the number of skin and nerve lesions, BI, Ridley-Jopling classification as well as the two physician's classification that in practice determined which multidrug regimen each patient should receive. From the 1,182 leprosy cases, 1,127, 1,074 and 1,051 could be classified according to 'WHO-82' (mainly relying on Ridley and Jopling criteria), 'WHO-96' (which only relies on the number of lesions) and the physicians, respectively. The proportion of MB patients was 66% (695/1,051), 60% (672/1,127) and 41% (435/1,074) with the final physician's, WHO-82, and WHO-96 classifications, respectively. Figure 1 shows the Venn diagram of MB and PB patients classified according to the three definitions. The concordance analysis of the three classifications was quantified by means of the Kappa coefficient for the subset of 943 individuals classified for all three definitions: it was higher between 'WHO-82' and 'physician' than between 'WHO-82' and 'WHO-96' or between 'WHO-96' and 'physician' (0.88 [0.85;0.91], 0.62 [0.58;0.67] and 0.59 [0.54;0.64], respectively – a common interpretation of the Kappa coefficient is: [0;0.4] poor ; [0.4;0.6]: medium ; [0.6;0.8]: good ; [0.8;1]: excellent).

Apart from defining MB and PB forms according to various criteria, one can be interested in the phenotype that precisely consists in the contrast between these two forms. We define leprosy polarization as the process that drives the development of leprosy subtypes which needs to be differentiated from factors that impact on only one subtype. In our paper, we will refer to polarization whenever PB and MB forms are opposed, irrespective of the classification used to define PB and MB leprosy.

Epidemiology of leprosy subtypes

No epidemiological study specifically dealing with MB or PB leprosy has been published. In parallel, a systematic review of the epidemiological data on erythema nodosum leprosum, another leprosy phenotype, revealed the great difficulty to synthesize the data when the main phenotype is not leprosy per se [15]. Therefore we extracted such information from

more general epidemiological studies focusing on leprosy per se that included information on MB and PB forms, aware of a potential bias that could result from such a selection. As the vast majority of recent epidemiological studies focused on the contacts of an index leprosy case to identify risk factors for the occurrence of secondary cases [16–20], subgroup analysis of clinical forms is not very powerful, given the low number of such secondary cases: indeed, the detection rate of new cases is often less than 1% after 4 years of follow-up, requiring 10,000 contacts to ultimately study 100 secondary leprosy cases [17,18]. Since there is minimal variability within the genome of *M. leprae*, both at the local and global scale [21–23], pathogen factors are unlikely to contribute to clinical variability. Therefore, we restricted our explorations to host-related risk factors: gender, age, geographical origin, and BCG vaccination status.

Sex

The number of leprosy cases reported in adults is consistently higher among men, with a male to female sex-ratio ranging between 1.5 and 2 [24]. Few exceptions of a sex-ratio < 1 have been described in some African countries such as Burkina Faso, Kenya or Uganda [25]. Exposure, reporting or access to care biases cannot fully explain such a difference: during a massive outbreak on the island of Nauru in the Pacific (where leprosy affected a third of the population during the 1920s and where these biases are expected to play a minor role), the sex-ratio was still 1.5 [26]. Concerning the sex-ratio for MB or PB forms of the disease, a very large Brazilian study that used census data from the Ministry of Health between 2006 and 2010 (40,544 PB and 29,764 MB cases) showed a sex-ratio larger than 2 for MB patients but slightly lower than 1 for PB cases (Madrid classification) [27]. This might explain why, according to the relative proportion of MB and PB forms in a population, the overall sex-ratio varies widely, and why in some African countries, where the PB form is predominant, women are overrepresented among leprosy cases. In contrast to the Brazilian data, the proportion of MB was similar in males and in females in the Vietnamese sample, whatever the classification used (respectively 60% and 60% for WHO-82, 65% and 63% for physician, and 42 and 41% for WHO-96).

Age

Generally, the mean age of MB cases is higher than the one of PB cases; the common explanation being one of a longer period of incubation for the MB form (reviewed in [28]). A study in Andhra Pradesh (India) in 1962 showed that the proportion of MB cases – corresponding to L/(L+T) with the Madrid classification – increased from less than 5% between

5 and 9 years of age to over 25% for 25 to 64 year old patients [29]. These proportions were similar to those observed during the same period in another region from India (Tamil Nadu) [30] and in Burma (both using the Madrid definition) [27] (Figure 1). An independent study conducted in Uttar Pradesh (India) between 2005 and 2010 found the proportion of MB cases to be 54% in children against 77% in adults ($p < 0.001$) [31]. Finally, we used the data of the Brazilian study mentioned above [27] to show that the proportion of MB leprosy continuously increased with age, from less than 10% before 10 years of age to more than 50% for patients older than 60 years (Figure 1) [27]. The influence of age on the proportion of MB patients was greater in men than in women and multivariate analysis showed that the effects of these two factors (age and sex) on leprosy polarization were independent (Figure 1). Although the MB proportion was similar in males and females in the Vietnamese sample, we also observed that age was correlated with a higher proportion of MB patients, and this correlation was stronger for males than for females.

BCG

Two meta-analyses have shown that BCG vaccination significantly reduced the risk of developing clinical leprosy [32,33]. They included 26 and 28 studies, respectively, that used various classifications of MB and PB leprosy (Madrid, Ridley-Jopling, WHO-82, WHO-88 and WHO-96); 22 studies were common to both meta-analyses. The proportion of the observed protection ranged from 26% to 61% depending on the studies included. In the first meta-analysis, the reduction was greater for MB than for PB patients (78% vs. 59%, $p = 0.04$). In contrast, in the second study, where an adjustment was made on the number of immunizations with BCG and on the type of study (observational vs. interventional), the impact of BCG was not significantly different on the onset of one leprosy form or the other. Few reports suggested that the LL form is insensitive to vaccination (as these patients are not able to develop a granulomatous response) [34], others that BCG vaccination may shift MB to PB forms [35,36].

Geography

The proportion of MB cases varies greatly across continents, but also within each continent, between neighboring countries: among the countries that reported more than 1,000 cases in 2012, the proportion of MB cases (WHO-96) varied between 45% (Bangladesh) and 92% (Philippines) [24]. It is unlikely that demographic data such as age and gender explain the

observed variability, given that population sex-ratio are similar across countries and the patient age distribution is relatively homogeneous within a continent.

In 1966, Kenneth Newell, an epidemiologist self-defined as ‘naïve’ about leprosy, published a seminal paper entitled ‘An Epidemiologist's View of Leprosy’. He focused on a somewhat disregarded observation that he considered to be very informative: in countries of high endemicity, the prevalence of lepromatous cases (similar to MB cases as defined by Madrid classification) was relatively stable (between 5 and 10/1000). This observation enabled him to disregard some debated hypotheses about leprosy pathogenesis: 1) the lepromatous form is caused by a more virulent strain of *M. leprae*, 2) lepromatous cases are associated with a greater inoculum and/or greater exposure, 3) the polarization is linked to an environmental factor common to all human populations. Conversely, the data were consistent with the hypothesis that a fixed proportion of the infected population presents one or some characteristics that make certain persons more likely to develop a lepromatous form of leprosy. It is therefore reasonable to consider host genetics as a plausible explanation [37].

Human genetics of leprosy subtypes

There are many methodological tools in genetic epidemiology that can be used according to the hypotheses to be tested and the nature of the data. Their main characteristics are summarized in Table 1 [38]. Almost forgotten today, the field of genetic epidemiology used to be written with a capital E. Indeed, quite a number of approaches do not require any biological material and rely only on the information and the knowledge of the familial relationships. It is also possible to address the issue of human genetic contribution to the disease by simply considering familial links without having biological samples. These DNA-independent studies are of particular interest in the context of Neglected Tropical Diseases and knowledge transfer to endemic countries: they are easier to implement and intrinsically technology independent (*i.e.* their cost has remained low). The following section illustrates how such DNA-independent methods have contributed to our current understanding of the genetics of leprosy subtypes.

DNA-independent analyses: familial correlations, twin and segregation studies

Numerous studies have established that among close contacts of index cases (e.g. individuals living in the same house or sharing the same bed as the case), those biologically related to the index case are significantly more at risk of developing leprosy per se [16,17,20]. Intriguingly, no study has focused specifically on the subtype of leprosy developed by these secondary cases according to their demographics (age, gender etc...) or those of the index case, further substantiating that these phenotypes are neglected. Conversely, twin studies focused on polarization in addition to leprosy per se [39,40]. Among them, a study of twin pairs where at least one twin suffered from the MB subtype of leprosy (Madrid classification) found a higher concordance rate of the MB subtype in monozygotic (MZ) as compared to dizygotic (DZ) pairs (70% in MZ vs. 20% in DZ) [39]. On the opposite, for the PB form, concordance rates between MZ and DZ twins were very close. As we mentioned earlier, there is a greater proportion of MB leprosy among males [24] and this probably marginally increased the concordance rate in MZ twins vs. DZ ones (which may be of opposite sex). Finally, several familial segregation studies in the Philippines, La Desirade Island, Thailand and Brazil support a significant host genetic contribution to both MB and PB forms of the disease, although the genetic models were not always concordant across studies [40–43].

DNA-based analyses: linkage and association studies

Candidate-region or genome-wide linkage studies

Early linkage studies using DNA were performed on small chromosomal segments around a particular gene of interest chosen for its known function ('candidate-gene' or more precisely 'candidate-region' hypothesis). While the vast majority of these studies focused on leprosy per se, one analyzed the genetic control of leprosy polarization as defined by Ridley and Jopling [44]. It included 118 sibs (*i.e.* brothers and/or sisters) from 20 Vietnamese families and covered six chromosomal regions of interest: 1p36 around Tumor Necrosis Factor Receptor 2 (TNFR2), 5q31-33 around a group of cytokines, 6p21 around TNF, 10p13, 12p13 around TNFR1 and 12q13 around Vitamin D Receptor (VDR). Polarization was coded as a 'quantitative' trait, giving an increasing weight throughout the Jopling spectrum (TT<BT<BB<BL<LL). The study detected significant linkage of the 6p21 region including TNF with leprosy polarization [44].

Studies from Jamieson and Miller in a Brazilian sample reported linkage of leprosy per se with chromosomal regions 17q11-q21 [45], 6p21, 17q22 and 20p13 [46]. These studies

subsequently defined two groups T* and L*. The T* group included patients classified as TT, BT and BB according to the Ridley-Jopling scale while the L* group comprised BL and LL patients. Using this uncommon classification (BB usually come along with BL and LL in the MB group), significant linkage of the 17q11 and 20p13 chromosomal regions was detected for the L* phenotype but not for the T* group. Finally, a linkage study from Malawi found a significant linkage signal between 'polarization' and genetic markers on the 21q22 [47] chromosomal region (WHO-88 definition). However, as for most regions identified by linkage, the functional importance of specific genes located in this region awaits confirmation.

The first genome-wide (*i.e.* no restriction to a particular genomic region) linkage study of leprosy was performed in India in 2001. The sample included 224 families comprising 245 sib pairs affected by leprosy [48]. A significant linkage hit ($p < 2 \cdot 10^{-5}$) for leprosy was observed with genetic markers located on chromosomal region 10p13. However, because of the very high proportion of PB cases (98%; Madrid definition) it was not possible to decide whether the mapped locus influenced leprosy per se or was specific for the PB form. A strong argument for the impact of the 10p13 region on leprosy subtypes rather than leprosy per se is the significant heterogeneity found in a Vietnamese sample between PB and MB individuals (WHO-88, $p = 0.03$) [49]. A second study in India added some families to the sample of the first linkage study (243 families in total, including 233 with only PB children) and identified a second linkage signal on the 20p12 chromosomal region [50]. This region has been replicated for polarization but not for leprosy per se in the Vietnamese study (unpublished data). Overall, the results of the linkage studies in Indian and Vietnamese leprosy patients suggest that the 10p13 and 20p12 regions are differentially implicated according to the subtype considered.

Candidate-gene, candidate-region and genome-wide association studies

Numerous studies have tested the association between leprosy per se and genetic variants located in biologically relevant candidate genes. They have mainly focused on genes from the major histocompatibility complex, the Toll like receptors (TLRs) and the cytokines family [1,2]. Similarly to linkage studies, only a subset of the candidate-gene studies have considered the clinical forms of the disease. Their results are summarized in Table 2.

The strategy of association studies with a 'candidate gene' faces the classical problem of 'multiple testing' (several genes tested and several variants within each gene) which can result in inflated type I errors if not properly corrected. To avoid inflated type I errors genetic association studies, replication (*i.e.* confirmation in an independent sample from the same

population) and validation studies (confirmation in a sample from a different population) are critical. The problem of inflated type I errors in candidate gene association studies is not specific to leprosy but extends to the whole field of complex traits. The only validated studies for one of leprosy's subtypes (whether PB, MB or polarization) is provided by single nucleotide polymorphisms (SNPs) in genes for Toll Like Receptor 2 (TLR2) in Ethiopia [51] and Malawi [52], Tumor Necrosis Factor (TNF) in India [53] and Thailand [54], Mannose Binding Lectin 2 (MBL2) in Brazil [55] and Nepal [56], Mannose Receptor C type 1 (MRC1) in Brazil and Vietnam [57], and MHC class I chain-related gene A (MICA) in China [58] and Brazil [59]. And the association of one gene, Interleukin 10 (IL-10), with polarization, has been replicated in two independent Brazilian samples [60,61].

Other associations have only been found once and the implication of the tested genes in leprosy pathogenesis must be considered with caution: IFN γ and HLA-G in Brazil [62,63], TLR4 and Leukotriene A4 Hydrolase (LTA4H) in Ethiopia [64,65], Vitamin D receptor (VDR) [66], Killer Cell Immunoglobulin-like Receptor – 2 domains – short tail 3 (KIR2DS3) [67], Heat Shock Protein 1A (HSPA1A) and IL-23R in India [68,69], Lymphotoxin- α (LTA) in Malawi [52], Complement fraction 4 (C4A) in Thailand [70], IL-12B in Mexico [71], IL-12RB2 in Japan [72], Natural Resistance Associated Macrophage protein 1 (NRAMP1) in Mali [73], Laminin 2 (LAMA2) in Indonesia [74] and Defensin- β 1 (DEFB1) in Mexico [75]. Finally, a Brazilian study illustrated the impact of clinical classification. Here, the association of genetic variants in the gene coding for IL-17F was significant for leprosy polarization based on the WHO-88 classification (88 PB and 52 MB patients, OR 3.2, $p < 0.05$) but not for the WHO-96 definition (53 PB and 87 MB individuals) [76].

As mentioned above, linkage of the 10p13 chromosomal region with the PB form of leprosy in India [48] has been confirmed in Vietnam [49]. To characterize the molecular identity of the genes responsible for the signal in the region, a fine-mapping association study of this 'candidate region' was conducted using 1,522 markers and spanning over nine megabases (Mb) excluding MRC1 [77]. The primary sample included 294 nuclear families from Southern Vietnam with 303 sibs affected by leprosy, of whom 63% were classified as MB (WHO-96 definition). The replication sample from the same area consisted of 192 trios (2 parents and one child with leprosy), with 55% MB children. Three SNPs localized in the genes coding for Nebulette (NEBL) and Cubilin (CUBN, intrinsic factor-cobalamin receptor), located 3 Mb, apart were specifically and independently associated with the MB form [77].

The first Genome-Wide Association Study (GWAS) for leprosy was published in 2009 [78]. It included 706 cases and 1,225 controls from China; 4,367 additional controls were added in a subsequent study published in 2011 [79]. Five hundred thousand SNPs spanned over the entire genome were analyzed in this sample. The most promising signals were tested in a replication sample including 3,254 cases and 5,955 controls, mainly from Eastern China but also from Chinese ethnic minorities. In total, significant association of leprosy was detected with 18 polymorphisms assigned to eight genes: HLA-DR-DQ, Receptor-Interacting Protein Kinase 2 (RIPK2), Tumor Necrosis Factor SuperFamily member 15 (TNFSF15), Coiled-Coil Domain Containing 122 (CCDC122), LACCase Containing Domain 1 (LACC1), Nucleotide-binding Oligomerization Domain containing 2 (NOD2), Rab32 protein (RAB32) and IL-23R [78,79]. Heterogeneity in strength of association between the MB and PB groups (WHO-88) was observed for four polymorphisms assigned to four genes (LACC1, LRRK2, NOD2 and RIPK2). Validation studies of the observed associations were conducted in Vietnamese and Indian samples. In a familial sample from Vietnam including 286 MB and 188 PB leprosy patients (WHO-88), signals located in HLA-DR-DQ, RIPK2, CCDC122, LACC1 and NOD2 were validated ($p < 0.05$) and the effect was stronger in MB patients for HLA-DR-DQ, CCDC122 and LACC1 [80]. In the Indian case-control study using a smaller population (137 MB and 74 PB cases - WHO-88), signals in LRRK2 and RIPK2 genes were validated, and the observed effect was consistently more significant for PB patients [81]. A second study, using 48,000 markers located in 2,092 genes, was performed in an Indian population in 2010. The primary sample included only 258 cases and 300 controls from North India. Markers in the TLR1 and HLA-DRB1/DQA1 genes were significantly associated with leprosy per se. The test for heterogeneity between MB and PB groups (WHO-88) was not significant for the association with markers of these two genes [82]. Nonetheless, in both GWAS, leprosy per se was the main phenotype, and only SNPs that were significant with this phenotype were subsequently tested for heterogeneity between MB and PB groups. This is not an optimal design (see next section) and may well have missed SNPs with significant impact on polarization.

Study design: benefits of optimal hypothesis testing

The list of 36 genes implicated in one of leprosy subtypes or its polarization is shorter than the 78-genes list for leprosy per se (Figure 3). Several scenarios may explain the difference.

First since most studies were conducted in samples used for leprosy per se studies, by design patient numbers for subtype analysis were smaller than for leprosy per se resulting in lower power to detect genetic variants impacting on leprosy subtype. This is consistent with the fact that most reported associations concerned the MB form, which is more common in endemic countries except India. Second, an inadequate definition of MB and PB forms can also lead to a loss of power, as illustrated by the study of IL-17F [76]. Third, most studies analyzing leprosy polarization did not account for demographic risk factors: yet the study of leprosy per se has revealed that the inclusion of these covariates was essential (e.g. the association of leprosy per se with PARK2 or LTA was 3-5 times stronger in young patients [83,84]).

Schematically, most studies first look for association by comparing leprosy cases to controls, derive a p-value for a given genetic marker and a given genetic model (pper se). Only in a second step, additional tests for significance of association by comparing MB to all controls (pMB) and PB to all controls (pPB) are conducted. If $p_{MB} < p_{per\ se} < p_{PB}$, one might conclude that this SNP is associated with leprosy and has a major effect on MB patients (if $p_{MB} < p_{per\ se} < \text{threshold}$) or that this SNP is associated with MB leprosy (if $p_{MB} < \text{threshold} < p_{per\ se}$). Such a general strategy has several major shortcomings. For example a problem inherent to the above stepwise approach is that p-values contain little meaningful information and it is ORs that should be compared instead. We show in Figure 4 that, when the aim is to identify variants that specifically differentiate MB from PB patients, the optimal strategy is to refine the phenotype, balance the number in each class and then test the polarization phenotype as a whole, without partitioning into sub-groups since partitioning results in a significant loss of power.

Clearly, it is vital to use specific experiment designs according to the hypotheses to be tested, and not to limit future investigations to the simple sub-analysis of leprosy per se studies. In particular, comparing MB (or PB) individuals with controls does not allow determining if tagged genes are implicated in the infection, the occurrence of the particular form of leprosy or even leprosy per se. By contrast, testing MB vs. PB individuals is a powerful strategy to identify ‘specific’ polarization genetic determinants. Another part of the study design is the choice of the genetic variants to be tested (DNA – SNP or structural variant such as Copy Number Variant-, RNA, epigenetics - like methylation or glycosylation status). Finally, the size of the genotyping effort (candidate gene, candidate region, GWAS) is critical since the multiple testing will be much less an issue when testing few strong variants in a candidate gene as opposed to several millions of SNPs in a GWAS. In the face of limited resources and relative small sample sizes a focus on strong candidates can provide a very substantial increase in power. In the next section,

we argue that several powerful strategies able to generate appealing candidate genes have been surprisingly overlooked in the case of leprosy and its sub-phenotypes.

Generation of candidate genes

The human track

Because *M. leprae* cannot be cultivated in vitro, human studies have emerged as the only source for genetic studies. ‘Candidate genes’, are mainly chosen from a list of genes involved in the human immune system: major histocompatibility complex, TLR and cytokine family genes have been the most popular choices [1,2]. By contrast, ‘candidate pathways’ have never been tested. Definitions of pathways can be derived from the screening of in silico databases such as the Kyoto Encyclopedia of Genes and Genomes (KEGG), Gene Ontology (GO) or Search Tool for the Retrieval of Interacting Genes (STRING). Pathway-based approaches allow to extend the search for genetic markers (most typically SNPs) significantly associated with a phenotype from a single gene (~ 2-10 SNPs) to the pathway level (~ 100-1,000 SNPs depending on the size of the pathway). Similarly, pathway-based approaches can be used in the context of GWAS. Such strategies have been used successfully in cancer research [85,86], but have not been reported yet for leprosy.

Large GWAS remain a promising approach for the study of leprosy [87]. However, GWAS rarely have sufficient power to identify genetic variants that are rare at the population level. Such rare variants may potentially have strong biological effects and contribute strongly to the onset of the disease in a given individual. Such variants, although by themselves of minor importance at the population level can be critical to highlight pathways, as illustrated by the implication of the IFN γ - IL12/23 pathway in tuberculosis through the detection of private mutations causing susceptibility to mycobacterial diseases in a limited number of individuals. The knowledge of such ‘Mendelian/monogenic’ causes of leprosy is currently lacking. Yet, the identification of such cases should be of high priority due to the dearth of relevant in vitro and in vivo experimental models for leprosy.

The armadillo track

The nine-banded armadillo is a natural host of *M. leprae* that can be used for the study of the in vivo propagation of the bacilli [93–95]. Armadillos are cat-sized mammals found only

in the Americas that can live for up to 12 years and have a cool body temperature (32-35°C). The majority of nine-banded armadillos inoculated intravenously with *M. leprae* develop a disseminated form of the disease which is similar to multibacillary leprosy in man, and in particular, exhibits severe nerve lesions [96]. A study that included 392 wild armadillos affected by the disease in the United States showed that more than 80% were MB according to Ridley-Jopling (72% LL and 11% BL), regardless of their geographic origin (either from Louisiana, an enzootic region, or from Florida, a non-zoonotic area). Still, 20% were classified as PB. Hence, while armadillo is usually considered as a model for multibacillary leprosy, it can also be used to study polarization. The armadillo genome has been sequenced and annotated [97,98]. Although to date only one armadillo has been fully sequenced, additional sequencing will enable comparison of genetic variants between mammals that develop leprosy (only armadillo and human) and those that do not. Any variants unique to armadillo and humans might offer valuable insight into susceptibility while genes tagged for comparisons of MB and PB armadillos might provide useful insight into leprosy polarization.

Armadillos also have a unique characteristic that makes them particularly interesting for genetic studies, which is polyembryony (always giving birth to monozygotic quadruplets). This peculiarity can help us distinguish between environmental and genetic factors that favor the phenotype. Adams et al observed that among sets of sibling armadillos experimentally infected with *M. leprae* and bred in the same environment, the results of bacteria harvests were similar: this is a strong argument in favor of a major role of host genetics in the ability to fight the bacteria [93].

The pleiotropic track

In the absence of functional studies specific to leprosy, some genes have been explored in other diseases that share a common genetic background with leprosy. Inflammatory bowel disease (IBD) is a striking example. Some studies have revealed common signaling pathways between leprosy and Crohn's disease (CD) [99,80,2] fostering an ongoing debate about a possible contribution of mycobacteria to CD [100–103]. The analogy can be pursued further by comparing the two main forms of IBD, CD and ulcerative colitis (UC) with the two clinical forms of leprosy. Indeed, PB leprosy, like CD, preferably involves a Th1 signaling while MB leprosy, like UC, involves a Th2 response [104]. GWAS on IBD have identified 163 loci associated with IBD: 110 were common to both diseases, 30 specific for CD and 23 for UC [105]. Three genes specific for CD, RIPK2, LACC1 and NOD2 have been shown to be

associated in leprosy studies: NOD2 with leprosy per se, RIPK2 with PB leprosy and LACC1 with MB leprosy. This analogy also invites scientists working on leprosy to use advances in the genetics of IBD to derive an architecture for the genetic determinants of leprosy.

Conclusion

Given the results of the past 30 years, the scientific community may legitimately keep hope in the study of human genetics of leprosy and of its polarization. We showed in our review that evidence accumulates to isolate genes acting at the polarization level and that to not influence the onset of the disease. We have highlighted that a precise characterization of the phenotype is crucial to identify new genes driving polarization. Moreover, *M. leprae* can still not be cultured in vitro, which limits conventional biological studies. Therefore, the “reverse genetics” approach through genetic epidemiology remains not only attractive but determinant to decipher the physiopathology of this neglected tropical disease, of which polarization is a major aspect. The next step would be to use some of these genes to adapt the duration of treatment at an individual level. Progress can only come from a close collaboration between physicians, biologists, epidemiologists and geneticists. Leprosy, so long a cause of patient’s isolation and of division between men, will only be eradicated by a joint effort of all components of the scientific community.

Acknowledgments

We thank Dr Abad-Franch for sharing his data, Laurent Abel, Jean-Laurent Casanova and all members of the Epidemiological Team for their support and constructive criticism.. Special thanks go to Frédégonde About and Fabienne Jabot-Hanin for their assistance in designing the figures and their careful reading of the manuscript. JG is funded by the Fondation pour la Recherche Médicale, This work was supported by the Programme Blanc de l’Agence National de la Recherche, the Programme MALTALEP de l’Ordre de Malte and by the Canadian Institutes of Health Research (CIHR).

Bibliography

1. Misch EA, Berrington WR, Vary JC Jr, Hawn TR (2010) Leprosy and the human genome. *Microbiol Mol Biol Rev* MMBR 74: 589–620. doi:10.1128/MMBR.00025-10.
2. Gaschignard J, Scurr E, Alcaïs A (2013) La lèpre, pilier de la génétique des maladies infectieuses. *Pathol Biol* 61: 120–128. doi:10.1016/j.patbio.2013.03.003.
3. Scollard DM (2005) Leprosy research declines, but most of the basic questions remain unanswered. *Int J Lepr Mycobact Dis Off Organ Int Lepr Assoc* 73: 25–27. doi:10.1489/1544-581X(2005)73[25:LRDBMO]2.0.CO;2.
4. Scollard DM, Adams LB, Gillis TP, Krahenbuhl JL, Truman RW, et al. (2006) The continuing challenges of leprosy. *Clin Microbiol Rev* 19: 338–381. doi:10.1128/CMR.19.2.338-381.2006.
5. Scollard DM (2004) Classification of leprosy: a full color spectrum, or black and white? *Int J Lepr Mycobact Dis Off Organ Int Lepr Assoc* 72: 166–168. doi:10.1489/1544-581X(2004)072<0166:COLAFC>2.0.CO;2.
6. Ridley DS, Jopling WH (1966) Classification of leprosy according to immunity. A five-group system. *Int J Lepr Mycobact Dis Off Organ Int Lepr Assoc* 34: 255–273.
7. Sharma PV (1999) *Sushruta-Samhita: with the English Translation of Text and Dalhana's Commentary Along with Critical Notes*. Varanasi, India.
8. Languillon J (1999) *Précis de léprologie*.
9. Hansen GA, Looft C (1895) Leprosy: in its clinical and pathological aspects.
10. Wade HW (1935) Tuberculoid changes in leprosy IV. Classification of Tuberculoid leprosy. *Int J Lepr* 3: 121–136.
11. WHO (1982) Chemotherapy of leprosy control programmes.
12. Classification of leprosy at the Madrid Congress (1948). *Int J Lepr*: 504.
13. Ridley DS, Waters MF (1969) Significance of variations within the lepromatous group. *Lepr Rev* 40: 143–152.
14. WHO Expert Committee on Leprosy (1988) Sixth Report.
15. Voorend CGN, Post EB (2013) A Systematic Review on the Epidemiological Data of Erythema Nodosum Leprosum, a Type 2 Leprosy Reaction. *PLoS Negl Trop Dis* 7: e2440. doi:10.1371/journal.pntd.0002440.
16. Moet FJ, Meima A, Oskam L, Richardus JH (2004) Risk factors for the development of clinical leprosy among contacts, and their relevance for targeted interventions. *Lepr Rev* 75: 310–326.
17. Moet FJ, Pahan D, Schuring RP, Oskam L, Richardus JH (2006) Physical distance, genetic relationship, age, and leprosy classification are independent risk factors for leprosy in contacts of patients with leprosy. *J Infect Dis* 193: 346–353. doi:10.1086/499278.
18. Bakker MI, Hatta M, Kwenang A, Van Mosseveld P, Faber WR, et al. (2006) Risk factors for developing leprosy--a population-based cohort study in Indonesia. *Lepr Rev* 77: 48–61.
19. Antunes DE, Araujo S, Ferreira GP, Cunha ACSR da, Costa AV da, et al. (2013) Identification of clinical, epidemiological and laboratory risk factors for leprosy reactions during and after multidrug therapy. *Mem Inst Oswaldo Cruz* 108: 901–908. doi:10.1590/0074-0276130222.
20. Dos Santos DS, Duppre NC, Sales AM, Nery JA da C, Sarno EN, et al. (2013) Kinship and Leprosy in the Contacts of Leprosy Patients: Cohort at the Souza Araújo Outpatient Clinic, Rio de Janeiro, RJ, 1987-2010. *J Trop Med* 2013: 596316. doi:10.1155/2013/596316.
21. Cole ST, Eiglmeier K, Parkhill J, James KD, Thomson NR, et al. (2001) Massive gene decay in the leprosy bacillus. *Nature* 409: 1007–1011. doi:10.1038/35059006.
22. Monot M, Honoré N, Garnier T, Zidane N, Sherafi D, et al. (2009) Comparative genomic and phylogeographic analysis of *Mycobacterium leprae*. *Nat Genet* 41: 1282–1289. doi:10.1038/ng.477.
23. Singh P, Cole ST (2011) *Mycobacterium leprae*: genes, pseudogenes and genetic diversity. *Future Microbiol* 6: 57–71. doi:10.2217/fmb.10.153.
24. WHO (2013) Global leprosy: update on the 2012 situation. *Wkly Epidemiol Rec*: 365–380.
25. Le Grand A (1997) Women and leprosy: a review. *Lepr Rev* 68: 203–211.
26. WADE HW, LEDOWSKY V (1952) The leprosy epidemic at Nauru; a review, with data on the status since 1937. *Int J Lepr* 20: 1–29.
27. Guerra-Silveira F, Abad-Franch F (2013) Sex bias in infectious disease epidemiology: patterns and processes. *PloS One* 8: e62390. doi:10.1371/journal.pone.0062390.
28. Fine PE (1982) Leprosy: the epidemiology of a slow bacterium. *Epidemiol Rev* 4: 161–188.
29. De Vries JL, Perry BH (1985) Leprosy case detection rates by age, sex, and polar type under leprosy control conditions. *Am J Epidemiol* 121: 403–413.

30. Rao PS, Karat AB, Kaliaperumal VG, Karat S (1972) Prevalence of leprosy in Gudiyatham Taluk, South India. I. Specific rates with reference to age, sex and type. *Int J Lepr Mycobact Dis Off Organ Int Lepr Assoc* 40: 157–163.
31. Mukherjee PK, Das P, Rao PSS (2013) Time trends in MB-PB ratio among untreated leprosy patients attending a referral hospital in UP, India during 2001 to 2010. *Indian J Lepr* 85: 59–64.
32. Setia MS, Steinmaus C, Ho CS, Rutherford GW (2006) The role of BCG in prevention of leprosy: a meta-analysis. *Lancet Infect Dis* 6: 162–170. doi:10.1016/S1473-3099(06)70412-1.
33. Merle CSC, Cunha SS, Rodrigues LC (2010) BCG vaccination and leprosy protection: review of current evidence and status of BCG in leprosy control. *Expert Rev Vaccines* 9: 209–222. doi:10.1586/erv.09.161.
34. Thuc NV, Abel L, Lap VD, Oberti J, Lagrange PH (1994) Protective effect of BCG against leprosy and its subtypes: a case-control study in southern Vietnam. *Int J Lepr Mycobact Dis Off Organ Int Lepr Assoc* 62: 532–538.
35. Muliyl J, Nelson KE, Diamond EL (1991) Effect of BCG on the risk of leprosy in an endemic area: a case control study. *Int J Lepr Mycobact Dis Off Organ Int Lepr Assoc* 59: 229–236.
36. Schuring RP, Richardus JH, Pahan D, Oskam L (2009) Protective effect of the combination BCG vaccination and rifampicin prophylaxis in leprosy prevention. *Vaccine* 27: 7125–7128. doi:10.1016/j.vaccine.2009.09.054.
37. Newell KW (1966) A epidemiologist's view of leprosy. *Bull World Health Organ* 34: 827.
38. Ranque B, Alter A, Schurr E, Abel L, Alcais A (2008) [Leprosy: a paradigm for the study of human genetic susceptibility to infectious diseases]. *Médecine Sci MS* 24: 491–497. doi:10.1051/medsci/2008245491.
39. Chakravarti MR, Vogel F (1973) A twin study on leprosy. Georg Thieme Publishers, Stuttgart, Germany.
40. Smith DG (1979) The genetic hypothesis for susceptibility to lepromatous leprosy. *Hum Genet* 50: 163–177.
41. Abel L, Demenais F (1988) Detection of major genes for susceptibility to leprosy and its subtypes in a Caribbean island: Desirade island. *Am J Hum Genet* 42: 256–266.
42. Wagener DK, Schauf V, Nelson KE, Scollard D, Brown A, et al. (1988) Segregation analysis of leprosy in families of northern Thailand. *Genet Epidemiol* 5: 95–105. doi:10.1002/gepi.1370050205.
43. Feitosa MF, Borecki I, Krieger H, Beiguelman B, Rao DC (1995) The genetic epidemiology of leprosy in a Brazilian population. *Am J Hum Genet* 56: 1179–1185.
44. Mira MT, Alcais A, Di Pietrantonio T, Thuc NV, Phuong MC, et al. (2003) Segregation of HLA/TNF region is linked to leprosy clinical spectrum in families displaying mixed leprosy subtypes. *Genes Immun* 4: 67–73.
45. Jamieson SE, Miller EN, Black GF, Peacock CS, Cordell HJ, et al. (2004) Evidence for a cluster of genes on chromosome 17q11-q21 controlling susceptibility to tuberculosis and leprosy in Brazilians. *Genes Immun* 5: 46–57. doi:10.1038/sj.gene.6364029.
46. Miller EN, Jamieson SE, Joberty C, Fakiola M, Hudson D, et al. (2004) Genome-wide scans for leprosy and tuberculosis susceptibility genes in Brazilians. *Genes Immun* 5: 63–67. doi:10.1038/sj.gene.6364031.
47. Wallace C, Fitness J, Hennig B, Sichali L, Mwaungulu L, et al. (2004) Linkage analysis of susceptibility to leprosy type using an IBD regression method. *Genes Immun* 5: 221–225. doi:10.1038/sj.gene.6364062.
48. Siddiqui MR, Meisner S, Tosh K, Balakrishnan K, Ghei S, et al. (2001) A major susceptibility locus for leprosy in India maps to chromosome 10p13. *Nat Genet* 27: 439–441.
49. Mira MT, Alcais A, Van Thuc N, Thai VH, Huang NT, et al. (2003) Chromosome 6q25 is linked to susceptibility to leprosy in a Vietnamese population. *Nat Genet* 33: 412–415. doi:10.1038/ng1096.
50. Tosh K, Meisner S, Siddiqui MR, Balakrishnan K, Ghei S, et al. (2002) A region of chromosome 20 is linked to leprosy susceptibility in a South Indian population. *J Infect Dis* 186: 1190–1193. doi:10.1086/343806.
51. Bochud P-Y, Hawn TR, Siddiqui MR, Saunderson P, Britton S, et al. (2008) Toll-like receptor 2 (TLR2) polymorphisms are associated with reversal reaction in leprosy. *J Infect Dis* 197: 253–261. doi:10.1086/524688.
52. Fitness J, Floyd S, Warndorff DK, Sichali L, Mwaungulu L, et al. (2004) Large-scale candidate gene study of leprosy susceptibility in the Karonga district of northern Malawi. *Am J Trop Med Hyg* 71: 330–340.
53. Roy S, McGuire W, Mascie-Taylor CG, Saha B, Hazra SK, et al. (1997) Tumor necrosis factor promoter polymorphism and susceptibility to lepromatous leprosy. *J Infect Dis* 176: 530–532.
54. Vejbaesya S, Mahaisavariya P, Luangtrakool P, Sermduangprateep C (2007) TNF alpha and NRAMP1 polymorphisms in leprosy. *J Med Assoc Thai Chotmaihet Thangphaet* 90: 1188–1192.
55. De Messias-Reason IJ, Boldt ABW, Moraes Braga AC, Von Rosen Seeling Stahlke E, Dornelles L, et al. (2007) The association between mannan-binding lectin gene polymorphism and clinical leprosy: new insight into an old paradigm. *J Infect Dis* 196: 1379–1385. doi:10.1086/521627.
56. Sapkota BR, Macdonald M, Berrington WR, Misch EA, Ranjit C, et al. (2010) Association of TNF, MBL, and VDR polymorphisms with leprosy phenotypes. *Hum Immunol* 71: 992–998. doi:10.1016/j.humimm.2010.07.001.

57. Alter A, de Léséleuc L, Van Thuc N, Thai VH, Huong NT, et al. (2010) Genetic and functional analysis of common MRC1 exon 7 polymorphisms in leprosy susceptibility. *Hum Genet* 127: 337–348. doi:10.1007/s00439-009-0775-x.
58. Wang LM, Kimura A, Satoh M, Mineshita S (1999) HLA linked with leprosy in southern China: HLA-linked resistance alleles to leprosy. *Int J Lepr Mycobact Dis Off Organ Int Lepr Assoc* 67: 403–408.
59. Do Sacramento WS, Mazini PS, Franceschi DAS, de Melo FC, Braga MA, et al. (2012) Frequencies of MICA alleles in patients from southern Brazil with multibacillary and paucibacillary leprosy. *Int J Immunogenet* 39: 210–215. doi:10.1111/j.1744-313X.2011.01074.x.
60. Moraes MO, Pacheco AG, Schonkeren JJM, Vanderborght PR, Nery JAC, et al. (2004) Interleukin-10 promoter single-nucleotide polymorphisms as markers for disease susceptibility and disease severity in leprosy. *Genes Immun* 5: 592–595. doi:10.1038/sj.gene.6364122.
61. Santos AR, Suffys PN, Vanderborght PR, Moraes MO, Vieira LMM, et al. (2002) Role of tumor necrosis factor-alpha and interleukin-10 promoter gene polymorphisms in leprosy. *J Infect Dis* 186: 1687–1691. doi:10.1086/345366.
62. Reynard MP, Turner D, Junqueira-Kipnis AP, Ramos de Souza M, Moreno C, et al. (2003) Allele frequencies for an interferon-gamma microsatellite in a population of Brazilian leprosy patients. *Eur J Immunogenetics Off J Br Soc Histocompat Immunogenetics* 30: 149–151.
63. Lucena-Silva N, Teixeira MAG, Ramos A de L, de Albuquerque RS, Diniz GTN, et al. (2013) The +3187A/G HLA-G polymorphic site is associated with polar forms and reactive reaction in leprosy. *Mol Genet Genomic Med* 1: 123–130. doi:10.1002/mgg3.14.
64. Bochud P-Y, Sinsimer D, Aderem A, Siddiqui MR, Saunderson P, et al. (2009) Polymorphisms in Toll-like receptor 4 (TLR4) are associated with protection against leprosy. *Eur J Clin Microbiol Infect Dis Off Publ Eur Soc Clin Microbiol* 28: 1055–1065. doi:10.1007/s10096-009-0746-0.
65. Tobin DM, Vary JC Jr, Ray JP, Walsh GS, Dunstan SJ, et al. (2010) The *Ita4h* locus modulates susceptibility to mycobacterial infection in zebrafish and humans. *Cell* 140: 717–730. doi:10.1016/j.cell.2010.02.013.
66. Roy S, Frodsham A, Saha B, Hazra SK, Mascie-Taylor CG, et al. (1999) Association of vitamin D receptor genotype with leprosy type. *J Infect Dis* 179: 187–191. doi:10.1086/314536.
67. Franceschi DSA, Mazini PS, Rudnick CCC, Sell AM, Tsuneto LT, et al. (2008) Association between killer-cell immunoglobulin-like receptor genotypes and leprosy in Brazil. *Tissue Antigens* 72: 478–482. doi:10.1111/j.1399-0039.2008.01127.x.
68. Rajalingam R, Mehra NK, Singal DP (2000) Polymorphism in heat-shock protein 70-1 (HSP70-1) gene promoter region and susceptibility to tuberculoid leprosy and pulmonary tuberculosis in Asian Indians. *Indian J Exp Biol* 38: 658–662.
69. Ali S, Srivastava AK, Chopra R, Aggarwal S, Garg VK, et al. (2013) IL12B SNPs and copy number variation in IL23R gene associated with susceptibility to leprosy. *J Med Genet* 50: 34–42. doi:10.1136/jmedgenet-2012-101214.
70. Greiner J, Weber FJ, Mauff G, Baur M (1980) 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* 158: 134–138. doi:10.1016/S0171-2985(80)80053-2.
71. Alvarado-Navarro A, Montoya-Buelna M, Muñoz-Valle JF, López-Roa RI, Guillén-Vargas C, et al. (2008) The 3'UTR 1188 A/C polymorphism in the interleukin-12p40 gene (IL-12B) is associated with lepromatous leprosy in the west of Mexico. *Immunol Lett* 118: 148–151. doi:10.1016/j.imlet.2008.03.015.
72. Ohyama H, Ogata K, Takeuchi K, Namisato M, Fukutomi Y, et al. (2005) Polymorphism of the 5' flanking region of the IL-12 receptor beta2 gene partially determines the clinical types of leprosy through impaired transcriptional activity. *J Clin Pathol* 58: 740–743. doi:10.1136/jcp.2004.023903.
73. Meisner SJ, Mucklow S, Warner G, Sow SO, Lienhardt C, et al. (2001) Association of NRAMP1 polymorphism with leprosy type but not susceptibility to leprosy per se in west Africans. *Am J Trop Med Hyg* 65: 733–735.
74. Wibawa T, Soebono H, Matsuo M (2002) Association of a missense mutation of the laminin alpha2 gene with tuberculoid type of leprosy in Indonesian patients. *Trop Med Int Health TM IH* 7: 631–636.
75. Prado-Montes de Oca E, Velarde-Félix JS, Ríos-Tostado JJ, Picos-Cárdenas VJ, Figuera LE (2009) SNP 668C (-44) alters a NF-kappaB1 putative binding site in non-coding strand of human beta-defensin 1 (DEFB1) and is associated with lepromatous leprosy. *Infect Genet Evol J Mol Epidemiol Evol Genet Infect Dis* 9: 617–625. doi:10.1016/j.meegid.2009.03.006.
76. Chaitanya VS., Jadhav RS, Lavania M, Singh M, Valluri V, et al. (2013) Interleukin-17F single-nucleotide polymorphism (7488T>C) and its association with susceptibility to leprosy. *Int J Immunogenet.* doi:10.1111/iji.12097.

77. Grant AV, Cobat A, Van Thuc N, Orlova M, Huong NT, et al. (2014) CUBN and NEBL common variants in the chromosome 10p13 linkage region are associated with multibacillary leprosy in Vietnam. *Hum Genet*. doi:10.1007/s00439-014-1430-8.
78. Zhang F-R, Huang W, Chen S-M, Sun L-D, Liu H, et al. (2009) Genomewide association study of leprosy. *N Engl J Med* 361: 2609–2618. doi:10.1056/NEJMoa0903753.
79. Zhang F, Liu H, Chen S, Low H, Sun L, et al. (2011) Identification of two new loci at IL23R and RAB32 that influence susceptibility to leprosy. *Nat Genet* 43: 1247–1251. doi:10.1038/ng.973.
80. Grant AV, Alter A, Huong NT, Orlova M, Van Thuc N, et al. (2012) Crohn's disease susceptibility genes are associated with leprosy in the Vietnamese population. *J Infect Dis* 206: 1763–1767. doi:10.1093/infdis/jis588.
81. Marcinek P, Jha AN, Shinde V, Sundaramoorthy A, Rajkumar R, et al. (2013) LRRK2 and RIPK2 Variants in the NOD 2-Mediated Signaling Pathway Are Associated with Susceptibility to Mycobacterium leprae in Indian Populations. *PLoS ONE* 8: e73103. doi:10.1371/journal.pone.0073103.
82. Wong SH, Gochhait S, Malhotra D, Pettersson FH, Teo YY, et al. (2010) Leprosy and the adaptation of human toll-like receptor 1. *PLoS Pathog* 6: e1000979. doi:10.1371/journal.ppat.1000979.
83. Alcais A, Alter A, Antoni G, Orlova M, Van Thuc N, et al. (2007) Stepwise replication identifies a low-producing lymphotoxin-alpha allele as a major risk factor for early-onset leprosy. *Nat Genet* 39: 517–522.
84. Alter A, Fava VM, Huong NT, Singh M, Orlova M, et al. (2012) Linkage disequilibrium pattern and age-at-diagnosis are critical for replicating genetic associations across ethnic groups in leprosy. *Hum Genet*. doi:10.1007/s00439-012-1227-6.
85. Vinayagamoorthy N, Hu H-J, Yim S-H, Jung S-H, Jo J, et al. (2014) New Variants Including ARG1 Polymorphisms Associated with C-Reactive Protein Levels Identified by Genome-Wide Association and Pathway Analysis. *PloS One* 9: e95866. doi:10.1371/journal.pone.0095866.
86. Lee YH, Kim J-H, Song GG (2014) Genome-wide pathway analysis of breast cancer. *Tumour Biol J Int Soc Oncodevelopmental Biol Med*. doi:10.1007/s13277-014-2027-5.
87. Stranger BE, Stahl EA, Raj T (2011) Progress and promise of genome-wide association studies for human complex trait genetics. *Genetics* 187: 367–383. doi:10.1534/genetics.110.120907.
88. Blair DR, Lytle CS, Mortensen JM, Bearden CF, Jensen AB, et al. (2013) A nondegenerate code of deleterious variants in Mendelian loci contributes to complex disease risk. *Cell* 155: 70–80. doi:10.1016/j.cell.2013.08.030.
89. Edwards SL, Beesley J, French JD, Dunning AM (2013) Beyond GWASs: Illuminating the Dark Road from Association to Function. *Am J Hum Genet* 93: 779–797. doi:10.1016/j.ajhg.2013.10.012.
90. De Léséleuc L, Orlova M, Cobat A, Girard M, Huong NT, et al. (2013) PARK2 mediates interleukin 6 and monocyte chemoattractant protein 1 production by human macrophages. *PLoS Negl Trop Dis* 7: e2015. doi:10.1371/journal.pntd.0002015.
91. Manzanillo PS, Ayres JS, Watson RO, Collins AC, Souza G, et al. (2013) The ubiquitin ligase parkin mediates resistance to intracellular pathogens. *Nature*. doi:10.1038/nature12566.
92. Mira MT, Alcais A, Nguyen VT, Moraes MO, Di Flumeri C, et al. (2004) Susceptibility to leprosy is associated with PARK2 and PACRG. *Nature* 427: 636–640. doi:10.1038/nature02326.
93. Adams LB, Pena MT, Sharma R, Hagge DA, Schurr E, et al. (2012) Insights from animal models on the immunogenetics of leprosy: a review. *Mem Inst Oswaldo Cruz* 107 Suppl 1: 197–208.
94. Kirchheimer WF, Storrs EE (1971) Attempts to establish the armadillo (*Dasypus novemcinctus* Linn.) as a model for the study of leprosy. I. Report of lepromatoid leprosy in an experimentally infected armadillo. *Int J Lepr Mycobact Dis Off Organ Int Lepr Assoc* 39: 693–702.
95. Truman RW, Ebenezer GJ, Pena MT, Sharma R, Balamayooran G, et al. (2014) The armadillo as a model for peripheral neuropathy in leprosy. *ILAR J Natl Res Counc Inst Lab Anim Resour* 54: 304–314. doi:10.1093/ilar/ilt050.
96. Sharma R, Lahiri R, Scollard DM, Pena M, Williams DL, et al. (2013) The armadillo: a model for the neuropathy of leprosy and potentially other neurodegenerative diseases. *Dis Model Mech* 6: 19–24. doi:10.1242/dmm.010215.
97. Lindblad-Toh K, Garber M, Zuk O, Lin MF, Parker BJ, et al. (2011) A high-resolution map of human evolutionary constraint using 29 mammals. *Nature* 478: 476–482. doi:10.1038/nature10530.
98. Lynch VJ, Leclerc RD, May G, Wagner GP (2011) Transposon-mediated rewiring of gene regulatory networks contributed to the evolution of pregnancy in mammals. *Nat Genet* 43: 1154–1159. doi:10.1038/ng.917.
99. Alter A, Grant A, Abel L, Alcais A, Schurr E (2011) Leprosy as a genetic disease. *Mamm Genome Off J Int Mamm Genome Soc* 22: 19–31. doi:10.1007/s00335-010-9287-1.
100. Greenstein RJ (2003) Is Crohn's disease caused by a mycobacterium? Comparisons with leprosy, tuberculosis, and Johne's disease. *Lancet Infect Dis* 3: 507–514.

101. Sartor RB (2005) Does *Mycobacterium avium* subspecies paratuberculosis cause Crohn's disease? *Gut* 54: 896–898. doi:10.1136/gut.2004.055889.
102. Uzoigwe JC, Khaita ML, Gibbs PS (2007) Epidemiological evidence for *Mycobacterium avium* subspecies paratuberculosis as a cause of Crohn's disease. *Epidemiol Infect* 135: 1057–1068. doi:10.1017/S0950268807008448.
103. Pierce ES (2010) Ulcerative colitis and Crohn's disease: is *Mycobacterium avium* subspecies paratuberculosis the common villain? *Gut Pathog* 2: 21. doi:10.1186/1757-4749-2-21.
104. Brand S (2009) Crohn's disease: Th1, Th17 or both? The change of a paradigm: new immunological and genetic insights implicate Th17 cells in the pathogenesis of Crohn's disease. *Gut* 58: 1152–1167. doi:10.1136/gut.2008.163667.
105. Jostins L, Ripke S, Weersma RK, Duerr RH, McGovern DP, et al. (2012) Host-microbe interactions have shaped the genetic architecture of inflammatory bowel disease. *Nature* 491: 119–124. doi:10.1038/nature11582.
106. Ohyama H, Kato-Kogoe N, Nishimura F, Takeuchi-Hatanaka K, Matsushita S, et al. (2008) Differential effects of polymorphisms in the 5' flanking region of *IL12RB2* on NK- and T-cell activity. *J Interferon Cytokine Res Off J Int Soc Interferon Cytokine Res* 28: 563–569. doi:10.1089/jir.2008.0133.
107. Franceschi DSA, Mazini PS, Rudnick CCC, Sell AM, Tsuneto LT, et al. (2009) Influence of TNF and IL10 gene polymorphisms in the immunopathogenesis of leprosy in the south of Brazil. *Int J Infect Dis IJID Off Publ Int Soc Infect Dis* 13: 493–498. doi:10.1016/j.ijid.2008.08.019.

Box 1. Different classifications of leprosy

Madrid (1953): T (Tuberculoid), B (Borderline) and L (Lepromatous), according to clinical (skin and nerves lesions), bacteriological (bacillary index - BI, a measure of the number of acid-fast bacilli in the dermis expressed in a logarithmic scale), immunological (lepromin reaction) and histological (infiltrate's nature) criteria.

Ridley-Jopling (1966): TT (Tuberculoid Tuberculoid), BT (Borderline Tuberculoid), BB (Borderline Borderline), BL (Borderline Lepromatous) and LL (Lepromatous Lepromatous), and according to clinical, bacteriological, immunological and histological criteria.

WHO (1982): PB (Paucibacillary) or MB (Multibacillary) according to Ridley-Jopling (or Madrid) classifications and the BI. PB = TT + BT (or T) and $BI \leq 1$, MB = BB + BL + LL (or B + L) or $BI \geq 2$.

WHO (1988): PB = TT + BT and $BI = 0$, MB = BB + BL + LL or $BI \geq 1$.

WHO (1996): PB or MB according to the sole number of skin lesions ($\leq / > 5$).

Polarization: process that drives an infected individual towards the PB or the MB form (in our manuscript, irrespective of the classification used to define PB and MB cases).

Table 1. Main types of analyses used in genetic epidemiology

	Sample	Nb of markers	Aim	Design	Limitations
DNA-independent studies	Familial aggregation Families	No genetic markers	Highlight an excess in the familial risk for the disease under study	Comparison of the risk of developing the disease among relatives of affected individuals or of healthy subjects	No possible differentiation between genetic and environmental factors (common familial lifestyle)
	Twins Mono and dizygotic twins		Distinguish the relative contribution of genetic factors from other familial factors that contribute to the familial aggregation	Comparison of the phenotypic concordance rate in and monozygotic and dizygotic twins	Rarity of cases Ascertainment bias of mono and dizygotic twin pairs in old studies
	Adoptees Adopted children and their siblings			Comparison of the correlation between children and their natural or adopted parents	Rarity of cases Age of adoption
	Segregation Families ≥ 2 generations		Determine the origin of familial correlations, and in particular investigate the effect of a major gene among all genetic and environmental risk factors	Mathematical modeling of the disease risk according to the effects of a major gene, familial correlations and environmental factors and their interactions	Large samples required Simple genetic models only, otherwise computationally intensive
DNA-dependent studies	Linkage Families ≥ 2 affected siblings	Candidate-region (2-10)	Map chromosomal regions on the genome containing one or more gene(s) of susceptibility	Study of the correlation between the phenotypic similarity of two related individuals and their genotypic similarity. In the example of affected sib pairs, we test whether, for a given marker, two affected siblings share more than 50% of alleles inherited from their parents	No possible fine localization of the genes of interest (usual region size is 5-10 megabases in the context of a complex phenotype)
		Genome-wide (400-10,000)			
	Association <i>Case-control design:</i> unmatched cases and controls from a same ethnicity <i>Familial design:</i> the unit is the 'trio' = an affected child and his parents	Candidate-gene (1-10)	Identify in a particular region a genetic variant responsible for an increased risk of disease or in linkage disequilibrium with the functional genetic variant	<i>Case-control design:</i> Comparison of allele/genotype frequencies between cases and unrelated controls for a given marker	<i>Case-control design:</i> Delicate choice of controls Spurious associations (population stratification & linkage disequilibrium)
		GWAS (500,000-3M)		<i>Familial design:</i> Comparison of allele transmission from parents to their affected children	<i>Familial design:</i> Genotyping of relatives required
		Sequencing (30M-3G)		Study of genetic imprinting	Bias towards younger age

Table 2. Genes associated with clinical forms of leprosy or its polarization

The table is divided into two panels. Upper panel groups studies that analyzed the polarization phenotype, and lower panel those which compared MB or PB patients to controls. Studies are sorted according to their decreased sample size. Bold lines correspond to genes that were identified by GWAS; red lines to studies whose main outcome was to identify genes specific to leprosy polarization or one polar form of leprosy. Hence all black lines correspond to studies designed to study leprosy *per se* and a subsequent subgroup analysis for MB, PB or polarization phenotypes.

Comparison of MB against PB individuals						
Gene	Country	Phenotype	Definition	Nb of individuals (MB/PB/Controls)	p	Marker (allele, genetic model)
<i>MBL2</i> (Sapkota et al., 2010)	Nepal	MB vs. PB	WHO-88	581/343/101	0.01	polymorphism G161A (recessive)
<i>LTA4H</i> (Tobin et al., 2010)	Ethiopia	MB vs. PB	WHO-88	443/335/0	0.001	rs1978331 T/C and rs2660898 T/G
<i>LACC1</i> (Zhang et al., 2009a)	China	MB vs. PB	WHO-88	305/397/1225	<0.00001	rs3764147 and rs10507522 (allele ? additive)
<i>LRRK2</i> (Zhang et al., 2009a)	China	MB vs. PB	WHO-88	305/397/1225	0.0008	rs1491938 (allele ? additive)
<i>NOD2</i> (Zhang et al., 2009a)	China	MB vs. PB	WHO-88	305/397/1225	0.00004	rs9302752 (allele ? additive)
<i>RIPK2</i> (Zhang et al., 2009a)	China	MB vs. PB	WHO-88	305/397/1225	0.0008	rs42490 (allele ? additive)
<i>TLR4</i> (Bochud et al., 2009)	Ethiopia	MB vs. PB	WHO-88	298/138/197	0.05	polymorphism T1196C (allele C, dominant)
<i>NRAMP1</i> (Meisner et al., 2001)	Mali	MB vs. PB	WHO-88	181/92/201	0.003	3'-UTR 4-bp insertion/deletionTGTG (het vs del/del)
<i>MBL2</i> (de Messias-Reason et al., 2007)	Brazil	MB vs. PB	Madrid	150/36/214	0.012	
<i>IL-10</i> (Moraes et al., 2004)	Brazil	MB vs. PB	WHO-96	166/131/283	0.006	Haplotype -375A/-2849G/-2763C
<i>VDR</i> (Roy et al., 1999)	India	MB vs. PB	Madrid	124/107/165	0.001	TaqI T -> C
<i>IL-10</i> (Santos et al., 2002)	Brazil	MB vs. PB	WHO-96	143/59/62	<0.01	Haplotype -375A/-2849G/-2763C

<i>MICA</i> (do Sacramento et al., 2012)	Brazil	MB vs. PB	Madrid	153/45/201	0.01	MICA*010 and MICA*027
<i>IL-12RB2</i> (Ohyama et al., 2008)	Japan	MB vs. PB	Madrid	130/46/68	<0.01	A-1035G (allele G, allelic) A-1023G (allele G, allelic) -650delG (allelic G, allelic) A-464G (allele G, allelic) Haplotype -1035A/-1023A/-650G/-464A
<i>TLR2</i> (Fitness et al., 2004)	Malawi	MB vs. PB	WHO-96	26/184/379	0.04	microsatellite 224pb (in intron 2) (undefined model)
<i>IL-17F</i> (Chaitanya et al., 2013)	India	MB vs. PB	WHO-88	88/52/84	<0.05	T7488C (TT vs TC)
<i>IL-10</i> (Franceschi et al., 2009)	Brazil	MB vs. PB	Madrid	65(LL)/43(TT)/240	<0.01	haplotype -1082G/-819C/-592C
<i>LAMA2</i> (Wibawa et al., 2002)	Indonesia	MB vs. PB	WHO-88	27/26/58	<0.005	missense mutation (T7809C) substituting valine with alanine (V2587A) (TC vs TT)
<i>KIR2DS3</i> (Franceschi et al., 2008)	Brazil	LL vs. TT	RJ	65(LL)/42(TT)/289	0.04	
Comparison of one leprosy subtype (MB or PB) individuals against controls						
Gene	Country	Phenotype	Definition	Nb of individual (MB/PB/Controls)	p	
<i>IL-23R</i> (Prado-Montes de Oca et al., 2009)	India	PB vs. CTL	WHO-88	416/427/1502	0.02	CNV - 77 base amplicon located at exon 11 of the gene
<i>MRC1</i> (Alter et al., 2010)	Brazil	MB vs. CTL	WHO-88	373/331/396	0.001	rs1926736 (G/A= G396S, allele G recessive) rs2437257 (C/G = F407L, allele C recessive) haplotype G396-F407
<i>CCDC122</i> (Grant et al., 2012)	Vietnam	MB vs. CTL	WHO-88	286/188	0.0064	rs3088362 (C/A, A allele, additive)
<i>HLA-DR-DQ</i> (Grant et al., 2012)	Vietnam	MB vs. CTL	WHO-88	286/188	0.0003	rs602875 (A/G, A allele, additive)
<i>LACC1</i> (Grant et al., 2012)	Vietnam	MB vs. CTL	WHO-88	286/188	0.00073	rs3764147 (A/G, G allele, additive) rs10507522 (A/G, A allele, additive)
<i>TLR2</i> (Bochud et al., 2008)	Ethiopia	MB vs. CTL	WHO-88	298/138/187	0.02	microsatellite 288bp (dominant)

<i>TNF</i> (Roy et al., 1997)	India	MB vs. CTL	Madrid	121/107/160	0.01	G308A (allele A, allelic)
<i>RIPK2</i> (Marcinek et al., 2013)	India	PB vs. CTL	WHO-96	137/74/230	0.02	hapotype rs40457 G - rs42490 A
<i>LRRK2</i> (Marcinek et al., 2013)	India	PB vs. CTL	WHO-88	137/74/230	0.00001	rs1873613A/G (allele A, allelic)
<i>LTA</i> (Fitness et al., 2004)	Malawi	MB vs. CTL	WHO-96	26/184/379	0.003	5' UTR microsatellite (AC/GT) _n at -3.5kb (101-bp allele)
<i>HLA-G</i> (Lucena-Silva et al., 2013)	Brazil	MB vs. CTL	WHO-88	76/70/128	0.018	polymorphism + A3187G (allele A, dominant)
<i>MICA</i> (Wang et al., 1999)	China	MB vs. CTL	X	50/19/112	<0.05	MICA-A5 (allelic)
<i>IFNγ</i> (Reynard et al., 2003)	Brazil	PB vs. CTL	Madrid	59/10/98	0.01	CA repeat in intron 1 (< 122bp = "short", 122-126bp = "long") at risk = "long"
<i>DEFB1</i> (Prado-Montes de Oca et al., 2009)	Mexico	MB vs. CTL	Madrid	46/18/151	0.02	G668C (allele C, dominant)
<i>TNFα</i> (Vejbaesya et al., 2007)	Thailand	MB vs. CTL	WHO-96	24/13/140	0.04	G308A (allele A, undefined model)
<i>C4A</i> (Greiner et al., 1980)	Thailand	LL vs. CTL	RJ	71(LL)/27(TT)/201	<0.01	F1 allele ("functionally inactive")
<i>HSPA1A</i> (Rajalingam et al., 2000)	India	TT vs. CTL	RJ	0/49(TT)/38	0.03	allele at risk HSP70-1A (allelic)

Figure legends

Figure 1. Concordance of leprosy polarization phenotypes according to WHO-82, WHO-96 or physician definitions

Venn diagram for the number of patients with PB leprosy (left) and MB leprosy (right) according to three definitions: WHO-82, WHO-96 and physician. Only patients for which the three definitions were available appear in the diagram (534 PB and 623 MB individuals). The surfaces of overlap are approximately proportional to the number of individuals identically classified by one, two or three definitions. Note that there was no individual classified PB by a physician and MB by WHO-82 + WHO-96 as well as no individual classified MB by a physician and PB by WHO-82 + WHO-96.

Figure 2. Proportion of multibacillary patients according to age and sex

Brazil: 2006-2010 period, Madrid classification: 40,544 PB and 29,764 MB cases (total population of 189,000,000 inhabitants) (Guerra-Silveira and Abad-Franch, 2013).

Tamil Nadu (India): 1962-1970 period, Madrid classification: 3,963 PB and 1,258 MB cases (out of a total of 276,568 persons screened for leprosy) (Rao et al., 1972).

Ho Chi Minh region (Vietnam): 1990-2009 period, physician classification, 380 PB and 693 MB and cases.

Figure 3. Distribution of 82 published studies reporting a significant association between at least one genetic variant and leprosy *per se*, any leprosy subtype (*i.e.* MB or PB) or polarization

The + and – symbols refer to the presence or the absence of association of a given gene with one of leprosy subtypes in the study. The red quadrant refers to studies for which the primary outcome was the association with leprosy polarization or the PB subtypes; the blue quadrant to studies that could not conclude to an association with leprosy *per se* but with one of its subtypes; the green quadrant to studies that identified an association for both leprosy *per se* and one of its subtypes; and the grey quadrant to studies for which an association with the sole leprosy *per se* was reported.

Figure 4. Statistical impact of the hypothesis used to test for an association between a genetic variant and leprosy polarization

In the common case-control study design, a standard test of association between a genetic variant and the phenotype under study is to compare allelic frequencies (f) between the group of cases and the group of controls. Here, we consider a sample including 1,000 controls and 1,000 leprosy cases of whom 500 are PB and 500 are MB. In addition, we fixed the frequency of the allele of interest to 0,5 among the controls ($f_{\text{controls}} = 0,5$).

Two different strategies to test for association are represented. The first one directly compares MB to PB cases (Panel A). The second first tests PB vs. controls ((Panel B) and MB vs. controls (Panel C) before performing a heterogeneity test PB vs. MB, possibly after testing leprosy *per se* vs. controls (panel D). In each panel, f_{PB} (X-axis) is the allele frequency among PB individuals and f_{MB} (Y-axis), the allele frequency among MB individuals. The grey line represents the null hypothesis for each test. The green zone corresponds to values of f_{PB} and f_{MB} for which the association test is significant.

A yellow circle exemplifies a situation where a real difference exists between MB and PB, as well as for PB subtype; a white circle the case where a real difference exists between MB and PB, as well as for MB subtype. A red circle illustrates a situation where a real difference exists between MB and PB. Finally, a blue circle exemplifies the case where no difference exists between MB and PB, as well as for both MB and PB cases when compared to controls.

One can note with the red circle that some SNPs will be found significant with the first strategy but not with the second; indeed, the latter is hampered both by multiple testing (testing MB vs. PB after testing MB vs. controls and PB vs. controls) as well as the usual low power of heterogeneity tests.

Figure 1.

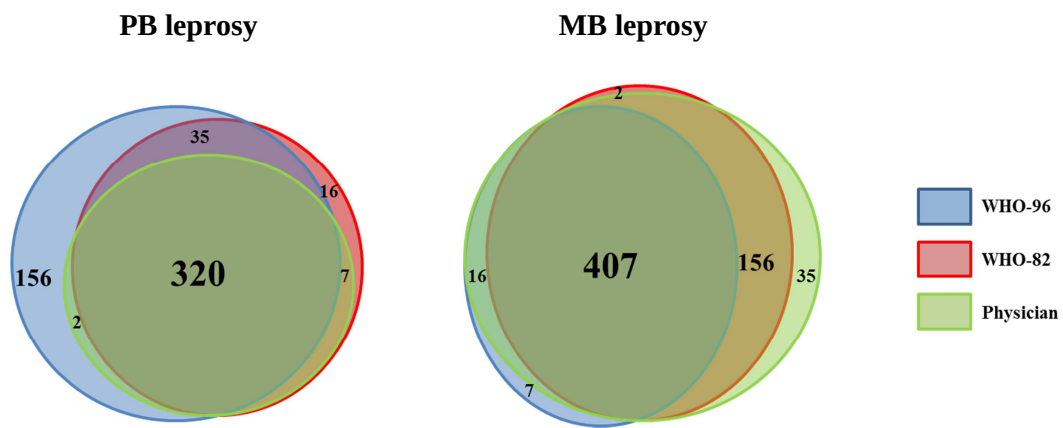


Figure 2.

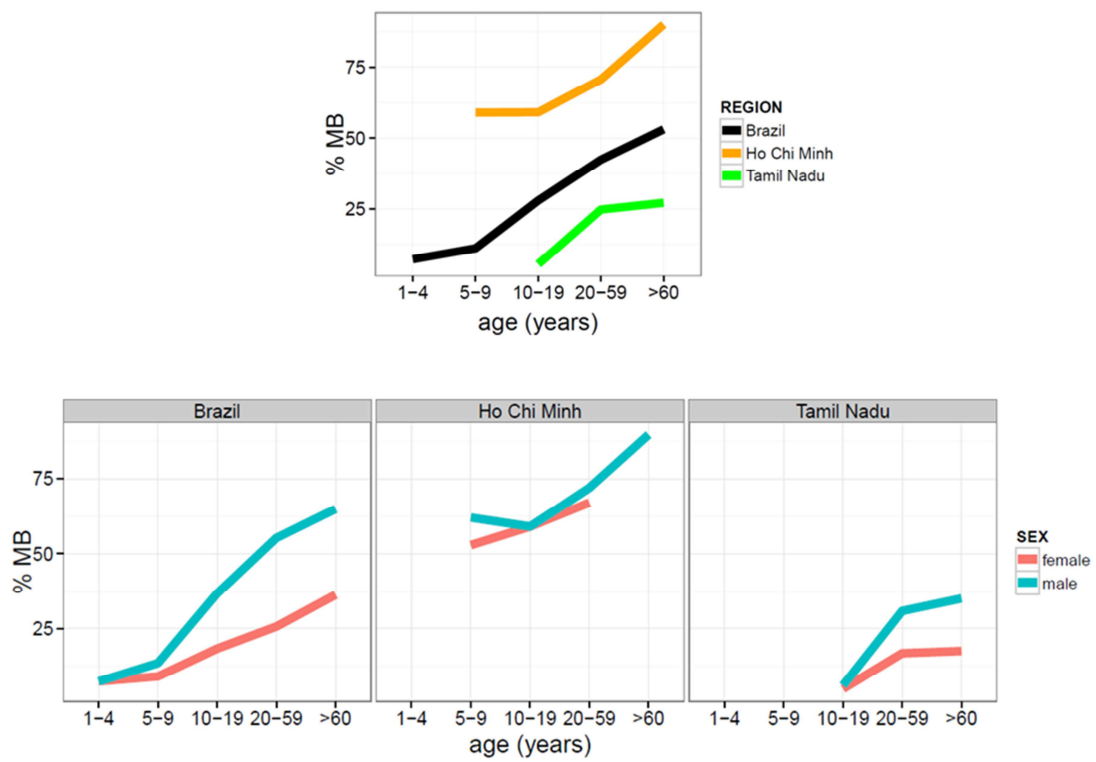


Figure 3.

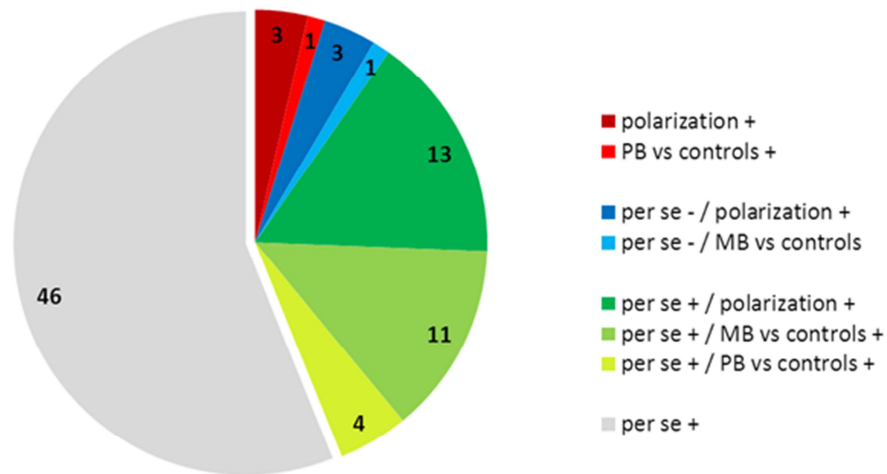
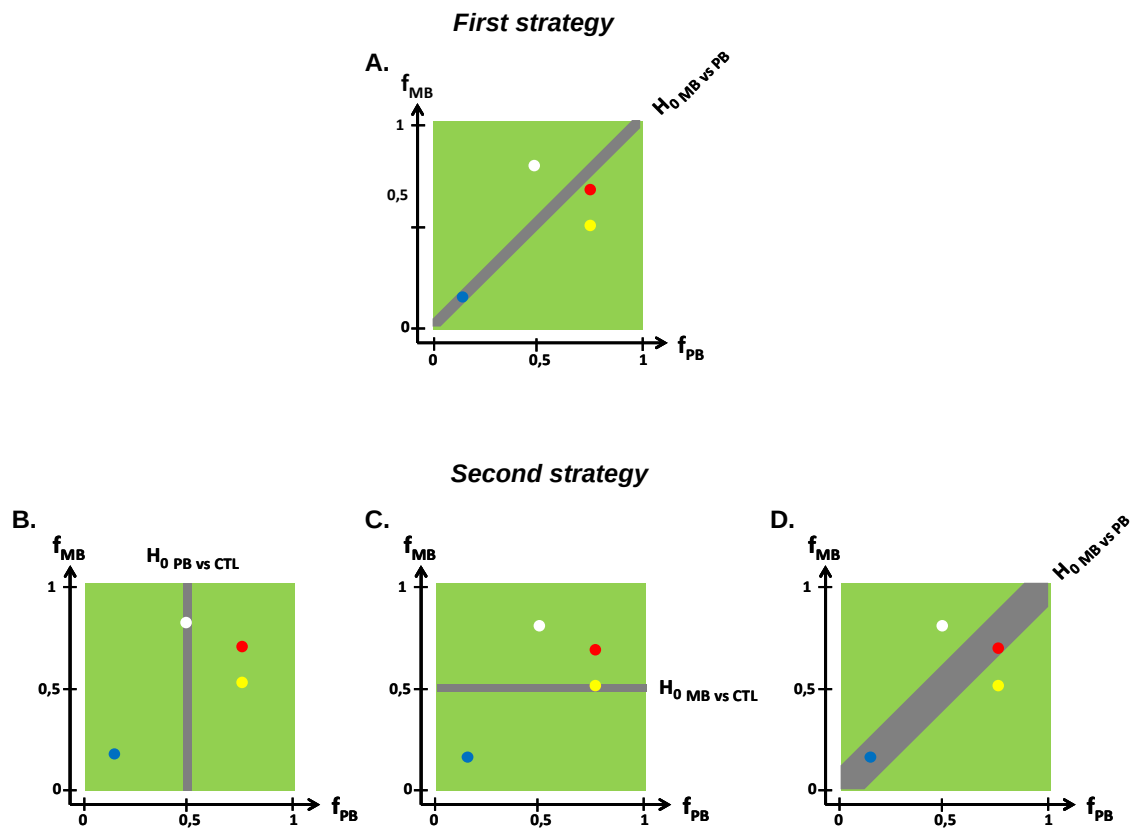


Figure 4.



Article 2

Chromosome 19p12-p13 is linked to susceptibility to leprosy polarization in a Vietnamese population

J. Gaschignard, A. Cobat, N.V. Thuc, M. Orlova, A. Boland-Auge, N.T. Huong, N.N. Ba, V.H. Thai, E. Schurr and A. Alcaïs

en cours de soumission à Journal of infectious Diseases

**Genomewide linkage analysis of leprosy polarization
implicates chromosomal region 19p12-p13**

Jean Gaschignard^{1,2}, Aurélie Cobat^{1,2}, Nguyen Van Thuc³, Marianna Orlova⁴, Anne Boland-Auge⁵, Nguyen Thu Huong³, Vu Hong Thai³, Erwin Schurr^{4,6}, Alexandre Alcaïs^{1,2,7*}

¹ Laboratory of Human Genetics of Infectious Diseases, Necker Branch, INSERM U1163, Necker Hospital for Sick Children, INSERM, Paris, France, EU

² Paris Descartes University, Imagine Institute, Paris, France, EU

³ Hospital for Dermato-Venerology, Ho Chi Minh City, Vietnam

⁴ Program in Infectious Diseases and Immunity in Global Health, The Research Institute of the McGill University Health Centre, Montreal, Quebec, Canada

⁵ National Center for Genotyping, Genomic Institute, CEA, Evry, France, EU

⁶ The McGill International TB Centre, Departments of Human Genetics and Medicine, McGill University, Montreal, Quebec, Canada

⁷ URC, CIC, Necker and Cochin Hospitals, Paris, France

* Corresponding author. Phone: +33 1 42 75 43 13, E-mail: alexandre.alcais@inserm.fr

Abstract

Lepromatous and tuberculoid leprosy are two distinct clinical forms of a disease for which there is strong evidence of a tight human genetic control. Still, few studies tried or succeeded in identifying genes influencing leprosy polarization, *i.e.* genes driving the disease towards its lepromatous or a tuberculoid form. We performed a genome-wide linkage scan for leprosy polarization in 72 families from Southern Vietnam with a total of 74 tuberculoid and 88 lepromatous children, using 45,947 independent single nucleotide polymorphisms (SNPs) covering the genome. Linkage was found with leprosy polarization on chr19p12-13 (maximal LOD score 3.57, $p=2.5 \cdot 10^{-5}$).

Keywords: leprosy, polarization, lepromatous, tuberculoid, genome-wide linkage, 19p12-13

Introduction

In 2013, more than 200,000 new cases of leprosy were declared worldwide. The cause of leprosy is *Mycobacterium leprae*. However, there is considerable clinical variability, with a clinical spectrum that extends from the polar ‘tuberculoid’ to the polar ‘lepromatous’ form of the disease (2014). The tuberculoid form is characterized by a small number of hypopigmented, well-bordered, anesthetic skin lesions and early peripheral nerve impairment, and involves a Th1 mediated immune response. In contrast, the lepromatous form is characterized by numerous, infiltrated skin lesions with high bacillary load and impaired peripheral nerves; it may even affect internal organs, and the immune-response is Th2-dominated. Within these two extremes, leprosy can present with a gradient of impaired Th1-mediated immunity, distinct histopathological manifestation, and variable bacillary counts in the lesions.

Mainly based on these clinical signs, leprosy is usually classified in 5 groups according to Ridley and Jopling: Tuberculoid Tuberculoid (TT), Borderline Tuberculoid (BT), Borderline Borderline (BB), Borderline Lepromatous (BL) and Lepromatous Lepromatous (LL) (Ridley and Jopling, 1966). Given the strong correlation of Th1 and Th2 responses with clinical subtype, leprosy polarization is an excellent example *in natura* of the impact of a differential immune response of B and T lymphocytes on clinical outcome of an infectious disease. While

considerable evidence has been accumulated for the implication of human genetics in the development of leprosy *per se* less is known about the genetic control of leprosy subtypes (Misch et al., 2010).

Genetic association between a gene and one of leprosy subtypes is usually accomplished by comparing patients from one subtype to unaffected controls. This does not allow to say whether the association is specific of the subtype or simply the reflect of the association with leprosy *per se*. We will refer to polarization when directly comparing individuals with different leprosy subtypes. Genes linked or associated with polarization are those that specifically drive the disease towards one subtype or the other among infected individuals.

Still, there is evidence for the impact of human genetic factors on leprosy polarization. Among linkage studies, one report focused on six chromosomal regions of interest (1p36 around Tumor Necrosis Factor Receptor 2 - *TNFR2*, 5q31-33 around a group of cytokines, 6p21 around *TNF*, 10p13, 12p13 around *TNFR1* and 12q13 around Vitamin D Receptor - *VDR*). The study detected significant linkage of the 6p21 HLA region with leprosy polarization, defined as a 'quantitative' trait (giving an increasing weight throughout the Jopling spectrum - TT<BT<BB<BL<LL) (Mira et al., 2003c). Another linkage study from Malawi found a significant linkage signal between polarization and genetic markers on the 21q22 chromosomal region, by grouping patients in a 2-categorical trait (LL + BL + BB vs BT + TT) (Wallace et al., 2004). Among candidate-gene tested for association, only Interleukin 10 (*IL-10*) and Mannose Binding Lectin 2 (*MBL2*) have been associated in at least two studies when opposing patients from the two poles, which is what we refer to as polarization (Misch et al., 2010).

In the large GWAS performed in China, heterogeneity between the lepromatous (LL + BL + BB) and the tuberculoid (BT + TT) groups was observed for four polymorphisms assigned to four genes (LACcase Containing Domain 1 - *LACC1*, Leucine-Rich Repeat Kinase 2 - *LRRK2*, Nucleotide-binding Oligomerization Domain containing 2 - *NOD2*, and Receptor-Interacting Protein Kinase 2 - *RIPK2*) (Zhang et al., 2009b). Additional evidence for leprosy polarization is provided by whole genome gene expression studies. Modlin et al. identified a genetic signature opposing lepromatous to tuberculoid forms, and showed that LIR-7 suppressed innate host defense mechanisms by shifting monocyte production from IL-12 toward IL-10 and by blocking antimicrobial activity triggered by Toll-like receptors (Bleharski et al., 2003). More recently, using the same array data, the authors showed that the expression of interferon family genes differs between tuberculoid and lepromatous patients (Teles et al., 2013). Given these results supporting the human genetic control of leprosy polarization, we performed a hypothesis-

free analysis of the polarization through a model-free genome-wide linkage study in a Vietnamese familial sample.

Materials and methods

Families and phenotype

Families were identified from records available at the Dermato-Venerology Hospital, Ho Chi Minh City. Patients that were not of Kinh ethnicity were excluded. Diagnosis of leprosy was made by at least two independent and experienced physicians and patients were classified in 5 categories according to Ridley and Jopling's classification (2). The number of lesions also allowed to classify patients according to the latest WHO field-oriented classification: paucibacillary (PB) if 5 skin lesions or less, multibacillary (MB) otherwise (WHO Expert Committee on Leprosy, 1998). Sex and age-of-onset of the disease were also recorded. Ethical approval of the study was obtained from the local Health Authorities in Ho Chi Minh City, Vietnam and the Research Ethics Committee of the McGill University Health Centre, Montreal, Canada.

Genotyping

Genotyping of children and parents for the genome-wide linkage analysis was performed by the Centre National de Génomique (Paris) with the Illumina HumanW660 chip, containing 592,633 Single Nucleotide Polymorphisms (SNPs). Quality control (QC) for the data was performed using PLINK software (<http://pngu.mgh.harvard.edu/~purcell/plink/>). For genetic markers, we applied the following QC criteria: call rate > 99.5%, minor allele frequency (MAF) ≥ 0.05 , p-value for Hardy-Weinberg equilibrium > 10^{-3} and no Mendelian errors. In total, 342,568 high-quality SNPs were considered for further analysis. Individuals with call-rate < 95% were excluded. For linkage analysis, we selected 45,947 common SNPs (MAF ≥ 0.1) displaying low levels of linkage disequilibrium ($r^2 < 0.2$). Mean Information content (IC) was higher than 99% in all autosomes (from 99.2% on chromosome 20 to 99.6% on chromosome 9). Less than 5% of the genome (159 Mb) was covered by markers spaced by more than 1 Mb. Only 8 SNPs were in chromosomal regions that had an IC below 95%.

Statistical analysis

Linkage analysis was performed on the quantitative phenotype ‘polarization’ coded as a categorical trait, using 5 categories (TT, BT, BB, BL and LL). Alternatively, 2 categories were used either by grouping BT and TT patients in a ‘tuberculoid’ group and LL, BL and BB patients in a ‘lepromatous’ group, or by using WHO’s classification of ‘paucibacillary’ vs ‘multibacillary’ patients. We used the new maximum-likelihood binomial (nMLB) model-free method for quantitative traits (nMLB-QTL v3.0) (Cobat et al., 2011b). This approach considers sibships as a whole. The categorical extension of the MLB method introduces a latent binary variable [0;1] that captures the linkage information between the observed categorical phenotype and the genetic marker. We fixed the probability to have the value 1 (i.e. to be lepromatous) at 0, 0.1, 0.8, 0.9 and 1 for TT, BT, BB, BL and LL patients, respectively; or at 0 and 1 for the ‘tuberculoid’ and ‘lepromatous’ (or ‘paucibacillary’ and ‘multibacillary’) groups when the trait was coded in two categories. The linkage test is asymptotically distributed as a 50:50 mixture of χ^2 distributions with 0 and 1 degree of freedom and the result can be expressed as a logarithm of odds (LOD) score: LOD scores > 3.6/2.2 are considered significant/suggestive for linkage at a genome-wide level (these correspond to nominal p-values of $7.3 \cdot 10^{-4}$ and $2.3 \cdot 10^{-5}$, respectively) (Kruglyak et al., 1996). The 90% and 95% confidence intervals correspond to a 1 and 1.5 LOD decrease from the linkage peak.

Results

We collected from our registries 373 individuals in 82 families with 2 or more children affected by leprosy, all belonging to the Kinh ethnic group. We decided to restrict our analysis to the 72 most informative families that displayed two genotyped parents, and included a total of 306 individuals. Figure 1 shows the distribution of tuberculoid (N=74) and lepromatous (N=88) cases among siblings. Children’s mean age at diagnosis was 17 years. The male to female ratio was 1.9 (106 males and 56 females). Age and sex were not significantly associated with subtypes in our sample (either in univariate or multivariate analysis, Supplementary Table 1). Thus, we did not adjust the phenotypes on these variables. The proportion of ‘concordant’ families (displaying only lepromatous or only tuberculoid children) was significantly higher than expected by chance given the proportion of lepromatous cases of 0.54 among children ($p=0.005$). When using WHO’s classification, 68 families (291 individuals) were available for analysis, including 103 paucibacillary and 52 multibacillary children.

Results of the genome-wide linkage for leprosy polarization, analyzed as a two-categorical trait (tuberculoid vs lepromatous), are shown in Figure 2A. We identified 1 region on chromosome 19p suggestive for linkage, peaking at rs2049051 (position 18,315,831 ; LOD score of 3.19, $p=6\cdot10^{-5}$, figure 2B and 2C, left). The 90% confidence interval defined a 3.6Mb-wide region extending from 17,1 to 20.7Mb (not shown). One other region almost reached suggestive linkage on chromosome 16, peaking at rs198174 (positions 24,247,031 ; LOD score 1.99, $p=1\cdot10^{-3}$, figure 2B). The significance of the peak on chromosome 19 was increased by refining the phenotype into 5 categories (with probabilities to be lepromatous of 0, 0.1, 0.8, 0.9 and 1 for TT, BT, BB, BL and LL children, respectively) (peak for the same SNP rs2049051 ; LOD score 3.57, $p=2.5\cdot10^{-5}$) and the 90% confidence interval could be narrowed to a 2.9Mb-wide region extending from 17,6 to 20.5 Mbs while the 95% confidence interval was between 17.0 and 20.8Mbs (Figure 2C, right). On the opposite, the peak on chromosomes 16 was less significant (LOD score 1.63 for rs1976194, 180kb away from rs198174, for which the LOD score was 1.59, $p=3\cdot10^{-4}$). Otherwise, the expansion of 2 to 5 categories of the quantitative phenotype did not detect any other suggestive or significant linkage regions on the other autosomes (Figure S1A).

When using WHO's operational definition to classify patients ('paucibacillary' vs 'multibacillary'), neither suggestive nor significant evidence for linkage was observed: the highest LOD score was 2.07 on chromosome 20 (for rs10891749 ; position 58,406,750) (Figure S1B). With this phenotype definition, the LOD score was 0 for rs1976194 on chromosomes 16, and 0.40 for rs2049051 on chromosome 19.

Discussion

Our study identified a region on chromosome 19p12-p13 linked to leprosy polarization, defined according to the Ridley and Jopling's classification. Within the 3.8 Mb confidence interval of this region, 180 genes are located of which 6 deserve further notion (Supplementary Table 2) : UNC-13 homolog A (*UNC13A*, 19:17,601,328- 17,688,314), Janus kinase 3 (*JAK3*, 19: 17,824,782-17,848,071), Interleukin 12 receptor Beta 1 (*IL12RB1*, 19:18059561-18086996), Interferon gamma-Inducible protein 30 (*IFI30*, 19:18,173,780-18,178,124), member RAS oncogene family *RAB3A* (19: 18,196,801-18,204,064) and Regulatory Factor X-associated Ankyrin-containing protein (*RFXANK*, 19:19192199-19201869). In particular, rs2049051 is located 230kb from *IL12RB1* and 1 Mb from *RFXANK*. A homozygous mutation in the latter gene is responsible for a Major Histocompatibility Complex Class II deficiency (Ouederni et al.,

2011). *IL12RB1* has been associated with Mendelian forms of infection with *Mycobacterium tuberculosis* or *Mycobacterium tuberculosis* (Schepers et al., 2013). Furthermore, one study has shown association between leprosy polarization and *IL12RB2* located on chromosome 1 (Ohyama et al., 2005).

An interesting finding is the weaker evidence for linkage when polarization was defined using the WHO operational classification. This may be a result of the weaker biological justification of the WHO operational classification as compared to the more solid clinically rooted subtypes defined by Ridley and Jopling. It seems important to keep in mind that operationally defined and successfully used phenotypes are much less suited for scientific studies addressing the pathophysiology of a trait. Results of the present study clearly highlight the benefits of a clinically more robust classification to increase the power of studies focused on leprosy polarization. While fewer families had been classified according to WHO classification scheme (68 families) as compared to Ridley and Jopling's one (72 families), this slight reduction in numbers was unlikely to cause the loss of linkage signal since it did not affect those families that contributed strongest to the linkage peak.

An additional important finding is that chromosomal regions previously shown to be linked with leprosy *per se* – 2p14, 6p21 and 6q25, 10p13, 17q11-21, 20p12-13, or with leprosy subtype – 10p13, were not found to be linked with leprosy polarization in our study (Misch et al., 2010). This can comfort the concept of a 'two-stage disease' model, with genes predisposing to the development of leprosy *per se* on the one-hand, and different genes driving the polarization of the disease on the other hand, as developed in another paper (Alter et al., 2011).

Finally, ongoing association studies in the region of linkage on chromosome 19 will eventually identify the genes underlying polarization which will be of considerable importance for our understanding of leprosy pathogenesis.

Acknowledgments

We thank all members of the Human Genetics of Infectious Diseases laboratory, particularly Frédégonde About and Fabienne Jabot-Hanin and all members of the Schurr lab for fruitful discussions. We thank the Centre National de Génotypage for providing the genotypes.

Financial support

Work in the lab of ES was supported by a grant from the Canadian Institutes of Health Research (CIHR). JG is funded by the Fondation pour la Recherche Médicale. This work was supported by the Programme Blanc de l'Agence National de la Recherche.

Potential conflicts of interest

All authors: no reported conflict.

Bibliography

1. Global leprosy update, 2013; reducing disease burden. *Relevé Épidémiologique Hebd Sect Hygiène Secrétariat Société Nations Wkly Epidemiol Rec Health Sect Secr Leag Nations*. 2014; 89(36):389–400.
2. Ridley DS, Jopling WH. Classification of leprosy according to immunity. A five-group system. *Int J Lepr Mycobact Dis Off Organ Int Lepr Assoc*. 1966; 34(3):255–273.
3. Misch EA, Berrington WR, Vary JC Jr, Hawn TR. Leprosy and the human genome. *Microbiol Mol Biol Rev MMBR*. 2010; 74(4):589–620.
4. Mira MT, Alcais A, Pietrantonio T Di, Thuc NV, Phuong MC, Abel L, et al. Segregation of HLA/TNF region is linked to leprosy clinical spectrum in families displaying mixed leprosy subtypes. *Genes Immun*. 2003; 4(1):67–73.
5. Wallace C, Fitness J, Hennig B, Sichali L, Mwaungulu L, Pönnighaus JM, et al. Linkage analysis of susceptibility to leprosy type using an IBD regression method. *Genes Immun*. 2004; 5(3):221–225.
6. Zhang F-R, Huang W, Chen S-M, Sun L-D, Liu H, Li Y, et al. Genomewide association study of leprosy. *N Engl J Med*. 2009; 361(27):2609–2618.
7. Bleharski JR, Li H, Meinken C, Graeber TG, Ochoa M-T, Yamamura M, et al. Use of genetic profiling in leprosy to discriminate clinical forms of the disease. *Science*. 2003; 301(5639):1527–1530.
8. Teles RMB, Graeber TG, Krutzik SR, Montoya D, Schenk M, Lee DJ, et al. Type I interferon suppresses type II interferon-triggered human anti-mycobacterial responses. *Science*. 2013; 339(6126):1448–1453.
9. WHO Expert Committee on Leprosy. *World Health Organ Tech Rep Ser*. 1998; 874:1–43.
10. Cobat A, Abel L, Alcaïs A. The Maximum-Likelihood-Binomial method revisited: a robust approach for model-free linkage analysis of quantitative traits in large sibships. *Genet Epidemiol*. 2011; 35(1):46–56.
11. Kruglyak L, Daly MJ, Reeve-Daly MP, Lander ES. Parametric and nonparametric linkage analysis: a unified multipoint approach. *Am J Hum Genet*. 1996; 58(6):1347–1363.
12. Ouederni M, Vincent QB, Frange P, Touzot F, Scerra S, Bejaoui M, et al. Major histocompatibility complex class II expression deficiency caused by a RFXANK founder mutation: a survey of 35 patients. *Blood*. 2011; 118(19):5108–5118.
13. Schepers K, Schandené L, Bustamante J, Vooren J-P Van, Suremain M de, Casanova J-L, et al. IL-12R β 1 deficiency and disseminated *Mycobacterium tuberculosis* disease. *J Clin Immunol*. 2013; 33(8):1285–1288.
14. Ohya H, Ogata K, Takeuchi K, Namisato M, Fukutomi Y, Nishimura F, et al. Polymorphism of the 5' flanking region of the IL-12 receptor beta2 gene partially determines the clinical types of leprosy through impaired transcriptional activity. *J Clin Pathol*. 2005; 58(7):740–743.
15. Alter A, Grant A, Abel L, Alcaïs A, Schurr E. Leprosy as a genetic disease. *Mamm Genome Off J Int Mamm Genome Soc*. 2011; 22(1-2):19–31.

Figure legends

Figure 1. Clustering of the distribution of tuberculoid and lepromatous children in 72 families from Southern Vietnam.

Each dot represents a child, either tuberculoid (blue) or lepromatous (red). Families with 2,3,4 or 5 children are sorted according to the concordance of their sibships: ‘concordant’ if all children from the same leprosy subtype, ‘discordant’ otherwise. The numbers in front of the dots correspond to the number of such families in the sample, and totals are given at the bottom.

Figure 2. Genome-wide linkage of leprosy polarization using Ridley-Jopling’s classification.

A: Multimarker LOD score (y-axis) with the nMLB method for linkage of leprosy polarization along each chromosome (x-axis) employing Ridley and Jopling’s classification in 2 categories (TT+BT+BB vs BL+LL. The red line indicates the information content ($0 \leq IC \leq 1$). **B:** Zoom of chromosomes 16 and 19 showing suggestive evidence for linkage, using Ridley and Jopling’s classification in 2 categories. Chromosomal location is indicated in million base pairs (Mb) on the x-axis. **C:** LOD scores for chromosome 19 region, using Ridley and Jopling’s classification in 2 categories (left) or 5 categories (TT<BT<BB<BL<LL, right). The red dot represents the peak LOD score values for rs2049051: 3.19 with the classification in 2 categories (left) and 3.57 with the classification in 5 categories (right). The 90% and 95% confidence intervals for location of the polarization locus correspond to a 1 and 1.5 LOD decrease from the linkage peak, that is 2.57 and 2.07, respectively. The limit for significant ($LOD=3.6$) genome-wide evidence for linkage is indicated by a green dotted line.

Figure S1. Genome-wide linkage of leprosy polarization.

Multimarker LOD score (y-axis) with the nMLB method for linkage of leprosy polarization along each chromosome (x-axis) employing Ridley and Jopling’s classification in 5 categories (TT<BT<BB<BL<LL, **panel A**) or employing WHO’s classification based on lesion number (PB vs MB, **panel B**). The red line indicates the information content ($0 \leq IC \leq 1$).

Figure 1.

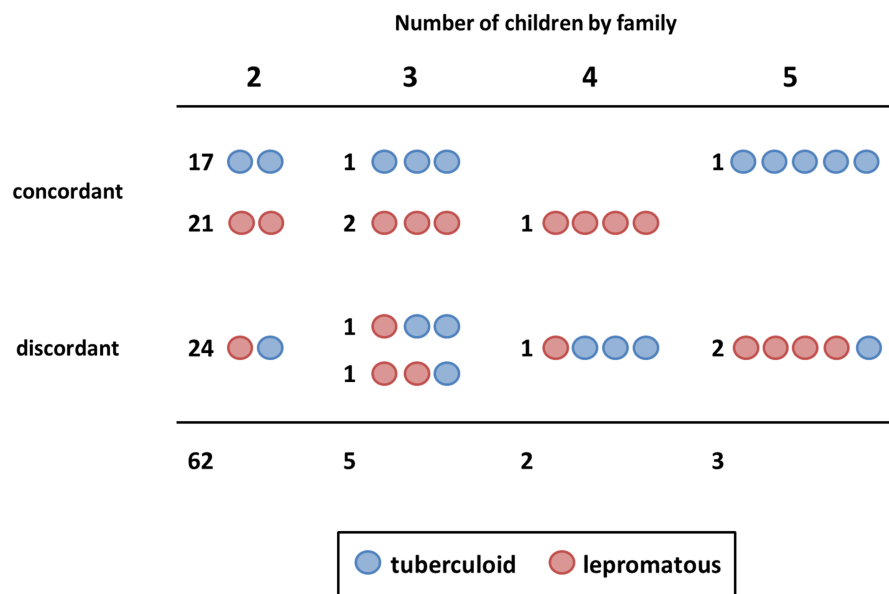


Figure 2.

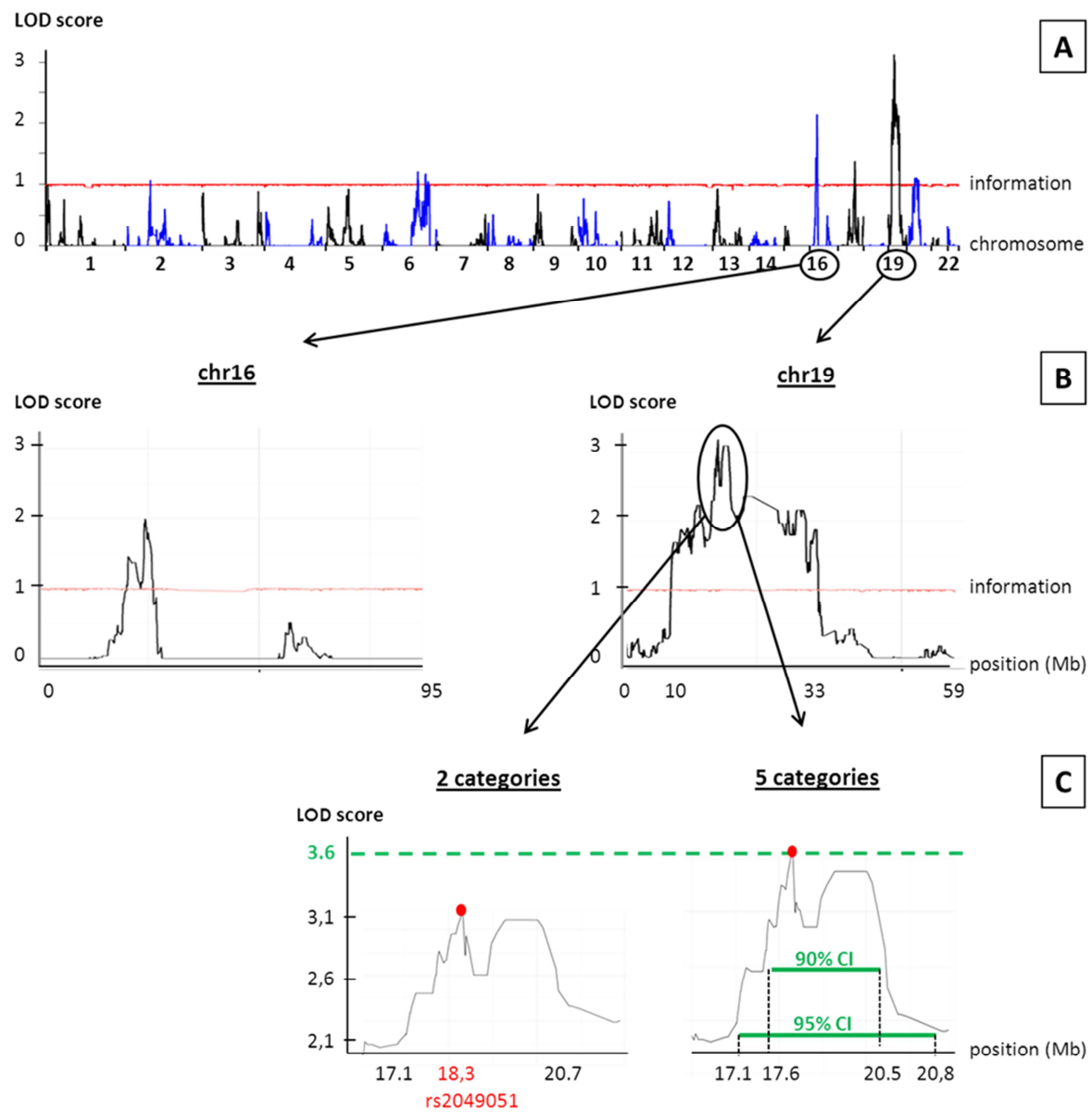


Figure S1.

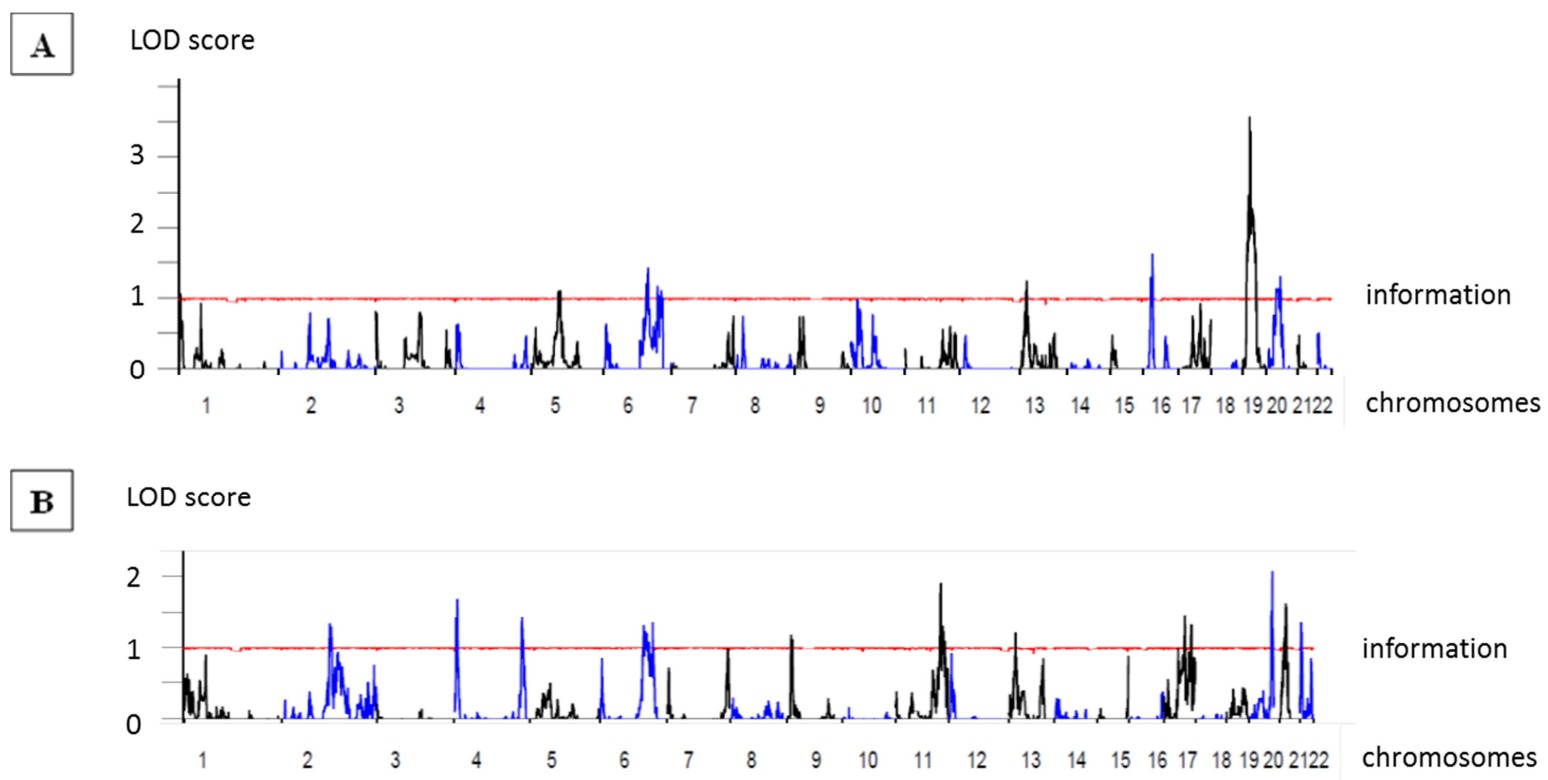


Table S1. Characteristics of the 162 siblings in the 72 families used for the linkage analysis.

	Tuberculoid (N = 74)	Lepromatous (N = 88)	p
Male	52	54	
Female	22	34	
Sex ratio	2.4	1.6	0.31
Age (mean [min;max])	17 [4;35]	17 [5;37]	0.43
Lesion number (mean [min;max])	2 [1;10]	9 [1;>20]	$< 10^{-6}$
> 5 lesions	4/69	47/85	$< 10^{-6}$
BI > 0	1/66	22/86	$2 \cdot 10^{-5}$
BCG scarr	19/63	21/82	0.67

Table S2. List of genes in the linkage region on chr19p12-13.

Loc	Start	Stop	Symbol	Description
19p13	17215346	17219829	USE1	unconventional SNARE in the ER 1 homolog (S. cerevisiae)
	17226221	17229219	OCEL1	occludin/ELL domain containing 1
	17231885	17245342	NR2F6	nuclear receptor subfamily 2, group F, member 6
	17249728	17264796	USHBP1	Usher syndrome 1C binding protein 1
	17267376	17279353	BABAM1	BRISC and BRCA1 A complex member 1
	17281645	17287646	ANKLE1	ankyrin repeat and LEM domain containing 1
	17292131	17303473	ABHD8	abhydrolase domain containing 8
	17305668	17306843	MRPL34	mitochondrial ribosomal protein L34
	17309528	17323298	DDA1	DET1 and DDB1 associated 1
	17323223	17334829	ANO8	anoctamin 8
	17334982	17342731	GTPBP3	GTP binding protein 3 (mitochondrial)
	17351455	17377328	PLVAP	plasmalemma vesicle associated protein
	17402939	17405648	BST2	bone marrow stromal cell antigen 2
	17405703	17506168	SLC27A1	solute carrier family 27 (fatty acid transporter), member 1
	17417031	17418288	HMGB3P29	high mobility group box 3 pseudogene 29
	17420103	17425331	MVB12A	family with sequence similarity 125, member A
	17435509	17448567	TMEM221	transmembrane protein 221
	17455425	17460916	NXNL1	nucleoredoxin-like 1
	17511469	17521288	PGLS	6-phosphogluconolactonase
	17523301	17553839	FAM129C	family with sequence similarity 129, member C
	17555702	17583162	COLGALT1	collagen beta(1-O)galactosyltransferase 1
	17589570	17590109	RPL21P130	ribosomal protein L21 pseudogene 130
	17601328	17688314	UNC13A	unc-13 homolog A (C. elegans)
	17719494	17734515	MAP1S	microtubule-associated protein 1S
	17747718	17788568	FCHO1	FCH domain only 1
	17795110	17813576	B3GNT3	UDP-GlcNAc:betaGalbeta-1,3-N-acetylglucosaminyltransferase 3

17816513	17821574	INSL3	insulin-like 3 (Leydig cell)
17824782	17848071	JAK3	Janus kinase 3
17859878	17863324	RPL18A	hCG_36873
17862588	17862720	SNORA68	small nucleolar RNA, H/ACA box 68
17871832	17895174	SLC5A5	solute carrier family 5 (sodium/iodide cotransporter), member 5
17933015	17943985	CCDC124	coiled-coil domain containing 124
17951302	17999324	KCNN1	potassium intermediate/small conductance calcium-activated channel, subfamily N, member 1
18001132	18014102	ARRDC2	arrestin domain containing 2
18030512	18033436	LOC440514	citrate synthase, mitochondrial-like
18040338	18041145	LOC646666	apolipoprotein A-I binding protein pseudogene
18048629	18049134	RPS18P13	ribosomal protein S18 pseudogene 13
18059561	18086996	IL12RB1	interleukin 12 receptor, beta 1
18097775	18151692	MAST3	microtubule associated serine/threonine kinase 3
18153178	18170533	PIK3R2	phosphoinositide-3-kinase, regulatory subunit 2 (beta)
18173780	18178124	IFI30	interferon, gamma-inducible protein 30
18193230	18196742	MPV17L2	MPV17 mitochondrial membrane protein-like 2
18196801	18204064	RAB3A	RAB3A, member RAS oncogene family
18207961	18248200	PDE4C	phosphodiesterase 4C, cAMP-specific
18249950	18255419	LOC729966	uncharacterized LOC729966
18257096	18274509	KIAA1683	KIAA1683
18271352	18285819	LOC102724325	uncharacterized LOC102724325
18279694	18281656	JUND	jun D proto-oncogene
18282077	18282161	MIR3188	microRNA 3188
18285166	18285545	RPL39P38	ribosomal protein L39 pseudogene 38
18306230	18323191	LSM4	LSM4 homolog, U6 small nuclear RNA associated (S. cerevisiae)
18309883	18315752	LOC102724342	uncharacterized LOC102724342
18340595	18369953	PGPEP1	pyroglutamyl-peptidase I
18386158	18389176	GDF15	growth differentiation factor 15
18386562	18386634	MIR3189	microRNA 3189
18391137	18397618	LRRC25	leucine rich repeat containing 25
18418869	18434562	SSBP4	single stranded DNA binding protein 4
18434388	18438649	ISYNA1	inositol-3-phosphate synthase 1
18442663	18522127	ELL	elongation factor RNA polymerase II
18531758	18543573	FKBP8	FK506 binding protein 8, 38kDa
18557762	18569387	KXD1	KxDL motif containing 1
18563766	18577460	UBA52	ubiquitin A-52 residue ribosomal protein fusion product 1
18588685	18592337	C19orf60	chromosome 19 open reading frame 60
18593225	18606850	CRLF1	cytokine receptor-like factor 1
18607435	18621039	TMEM59L	transmembrane protein 59-like
18637028	18670514	KLHL26	kelch-like family member 26

18683615	18782333	CRTC1	CREB regulated transcription coactivator 1
18782773	18791305	COMP	cartilage oligomeric matrix protein
18831935	18868230	UPF1	UPF1 regulator of nonsense transcripts homolog (yeast)
18868552	18896144	CERS1	ceramide synthase 1
18868552	18896144	GDF1	GDF-1
18899514	18919390	COPE	coatomer protein complex, subunit epsilon
18919675	18928633	DDX49	DEAD (Asp-Glu-Ala-Asp) box polypeptide 49
18929201	18941258	HOMER3	homer homolog 3 (Drosophila)
18946200	18946831	LOC102724360	uncharacterized LOC102724360
18990888	19034038	SUGP2	SURP and G patch domain containing 2
19033578	19058178	ARMC6	armadillo repeat containing 6
19063994	19113030	SLC25A42	solute carrier family 25, member 42
19119616	19138501	TMEM161A	transmembrane protein 161A
19145567	19170289	MEF2B	myocyte enhancer factor 2B
19145567	19192591	MEF2BNB-MEF2B	MEF2BNB-MEF2B readthrough
19176904	19192591	MEF2BNB	MEF2B neighbor
19192199	19201869	RFXANK	regulatory factor X-associated ankyrin-containing protein
19201414	19203429	NR2C2AP	nuclear receptor 2C2-associated protein
19211964	19252252	NCAN	neurocan
19255641	19262787	HAPLN4	hyaluronan and proteoglycan link protein 4
19264365	19273265	TM6SF2	transmembrane 6 superfamily member 2
19276511	19320512	SUGP1	SURP and G patch domain containing 1
19320687	19358754	MAU2	MAU2 sister chromatid cohesion factor
19385833	19508932	GATAD2A	GATA zinc finger domain containing 2A
19435063	19435158	MIR640	microRNA 640
19514219	19515660	TSSK6	testis-specific serine kinase 6
19516210	19528204	NDUFA13	NADH dehydrogenase (ubiquinone) 1 alpha subcomplex, 13
19528861	19537584	YJEFN3	YjeF N-terminal domain containing 3
19538265	19546659	CILP2	cartilage intermediate layer protein 2
19561707	19618916	PBX4	hCG_38572
19587606	19588439	PHF5CP	PHD finger protein 5C pseudogene
19623655	19628230	LPAR2	lysophosphatidic acid receptor 2
19629473	19643667	GMIP	GEM interacting protein
19645198	19663703	ATP13A1	ATPase type 13A1
19668153	19680926	ZNF101	zinc finger protein 101
19710472	19733112	ZNF14	zinc finger protein 14
19750617	19752539	LOC100533625	sorting nexin 18 pseudogene
19751495	19752661	LOC101060187	putative uncharacterized protein FLJ25328-like
19756371	19776413	LINC00663	long intergenic non-protein coding RNA 663
19786621	19816005	LOC101928926	uncharacterized LOC101928926
19792711	19821751	ZNF506	zinc finger protein 506

19p12	19827014	19827476	ISCA1P5	iron-sulfur cluster assembly 1 pseudogene 5
	19828054	19837420	ZNF56	zinc finger protein 56
	19865905	19893484	ZNF253	zinc finger protein 253
	19870724	19871283	BNIP3P10	BCL2/adenovirus E1B 19kDa interacting protein 3 pseudogene 10
	19900913	19935573	ZNF93	zinc finger protein 93
	19943703	20009237	BNIP3P12	BCL2/adenovirus E1B 19kDa interacting protein 3 pseudogene 12
	19979291	19979610	VN1R77P	vomer nasal 1 receptor 77 pseudogene
	20004418	20039468	ZNF682	zinc finger protein 682
	20033828	20034501	LOC101928960	BCL2/adenovirus E1B 19 kDa protein-interacting protein 3-like
	20052054	20063325	BNIP3P13	BCL2/adenovirus E1B 19kDa interacting protein 3 pseudogene 13
	20077994	20121168	ZNF90	zinc finger protein 90
	20079859	20080539	LOC100419301	ribosomal protein S8 pseudogene
	20087016	20087576	BNIP3P14	BCL2/adenovirus E1B 19kDa interacting protein 3 pseudogene 14
	20124980	20125550	RPS16P10	ribosomal protein S16 pseudogene 10
	20167214	20200490	ZNF486	zinc finger protein 486
	20198433	20199423	LOC100419841	zinc finger protein 107 pseudogene
	20221878	20222461	BNIP3P17	BCL2/adenovirus E1B 19kDa interacting protein 3 pseudogene 17
	20248415	20250387	LOC102724427	uncharacterized LOC102724427
	20257647	20262883	LOC284441	actin-related protein 2 pseudogene
	20295093	20303960	LOC100129265	zinc finger protein 66-like
	20351508	20352077	BNIP3P20	BCL2/adenovirus E1B 19kDa interacting protein 3 pseudogene 20
	20382406	20382978	BNIP3P21	BCL2/adenovirus E1B 19kDa interacting protein 3 pseudogene 21
	20397434	20398579	LOC102724436	zinc finger protein 93-like
	20398658	20424965	ZNF826P	zinc finger protein 826, pseudogene
	20399272	20399354	MIR1270	microRNA 1270
	20416514	20417066	BNIP3P22	BCL2/adenovirus E1B 19kDa interacting protein 3 pseudogene 22
	20443213	20443747	BNIP3P23	BCL2/adenovirus E1B 19kDa interacting protein 3 pseudogene 23
	20531261	20565820	ZNF737	zinc finger protein 737
	20582930	20583647	VN1R78P	vomer nasal 1 receptor 78 pseudogene
	20619939	20661596	ZNF626	zinc finger protein 626
	20667549	20686752	LOC102724456	uncharacterized LOC102724456
	20695212	20697232	LOC100533638	zinc finger protein 85 pseudogene
	20758464	20759138	VN1R79P	vomer nasal 1 receptor 79 pseudogene
	20776270	20837476	ZNF66	zinc finger protein 66
	20784599	20785152	BNIP3P24	BCL2/adenovirus E1B 19kDa interacting protein 3 pseudogene 24

Article 3

Implicit hypotheses are hidden power droppers in family-based association studies of secondary outcomes

J. Gaschignard, Q.B. Vincent, A. Cobat, J.P. Jaïs, and A. Alcaïs

Open Journal of Statistics

Implicit Hypotheses Are Hidden Power Droppers in Family-Based Association Studies of Secondary Outcomes

Jean Gaschignard^{1,2*}, Quentin B. Vincent^{1,2}, Jean-Philippe Jaïs^{1,2,3}, Aurélie Cobat^{1,2}, Alexandre Alcaïs^{1,2,4}

¹Laboratoire de Génétique des Maladies Infectieuses, Institut National de la Santé et de la Recherche Médicale, Paris, France

²Université Paris Descartes, Sorbonne Paris Cité, Institut Imagine, Paris, France

³Biostatistique et Informatique Médicale, Hôpital Necker, Paris, France

⁴URC, CIC, Necker and Cochin Hospitals, Paris, France

Email: jean.gaschignard@inserm.fr, alexandre.alcais@inserm.fr

Received 8 January 2015; accepted 26 January 2015; published 30 January 2015

Copyright © 2015 by authors and Scientific Research Publishing Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

Abstract

Family-based tests of association between a genetic marker and a disease constitute a common design to dissect the genetic architecture of complex traits. The FBAT software is one of the most popular tools to perform such studies. However, researchers are also often interested in the genetic contribution to a more specific manifestation of the phenotype (e.g. severe vs. non-severe form) known as a secondary outcome. Here, what we demonstrate is the limited power of the classical formulation of the FBAT statistic to detect the effect of genetic variants that influence a secondary outcome, in particular when these variants also impact on the onset of the disease, the primary outcome. We prove that this loss of power is driven by an implicit hypothesis, and we propose a derivation of the original FBAT statistic, free from this implicit hypothesis. Finally, we demonstrate analytically that our new statistic is robust and more powerful than FBAT for the detection of association between a genetic variant and a secondary outcome.

Keywords

Family-Based Association Test, FBAT, Genetic Association Studies, Null Hypothesis, Secondary Outcome, Homogeneity Test

*Corresponding author.

However, because of the bivariate nature of the phenotype under study (*i.e.* disease AND severe form or disease AND non-severe form), rejection of the null hypothesis cannot distinguish between alleles associated with the disease *per se* (*i.e.* independently of its severity) or alleles specifically associated with the severity of the disease. FBAT offers no immediate solution to study such secondary outcomes, *i.e.* to distinguish between alleles impacting the primary (e.g. disease *per se*) or the secondary (e.g. severe vs. non-severe) outcome. Below we propose two new tests denoted as FBAT_{het} and $\text{FBAT}_{\text{het free}}$ that can be used to directly assess the association between a marker allele and a secondary outcome.

3.1. The FBAT_{het} Test

A first straightforward idea is to perform a homogeneity test of the allelic transmission rate between the two subgroups D_1 and D_2 .

$$\begin{aligned} \text{Let } \text{FBAT}_{\text{het}}(D_1, D_2) &= \text{homogeneity}(S_1, S_2) \\ &= \frac{(S_1 - E_1)^2}{V_1} + \frac{(S_2 - E_2)^2}{V_2} - \frac{(S_1 - E_1 + S_2 - E_2)^2}{V_1 + V_2} \\ \text{FBAT}_{\text{het}} &= \frac{\left(\frac{S_1 - E_1}{V_1} - \frac{S_2 - E_2}{V_2} \right)^2}{\frac{1}{V_1} + \frac{1}{V_2}} = \frac{\left(\frac{S_1 - E_1}{V_1} - \frac{S_2 - E_2}{V_2} \right)^2}{\frac{1}{V_1^2} V_1 + \frac{1}{V_2^2} V_2} \end{aligned}$$

FBAT_{het} = FBAT with the phenotypes coded as $T = \frac{1}{V_1}$ for individuals D_1 and $T = -\frac{1}{V_2}$ for individuals D_2 .

Indeed,

$$\begin{aligned} S\left(T_1 = \frac{1}{V_1}, T_2 = -\frac{1}{V_2}\right) &= \sum_{i \in 1} \frac{1}{V_1} X_i - \sum_{i \in 2} \frac{1}{V_2} X_i = \frac{1}{V_1} S_1 - \frac{1}{V_2} S_2 \\ E &= \frac{1}{V_1} E_1 - \frac{1}{V_2} E_2 \quad \text{and} \quad V = \frac{1}{V_1^2} V_1 + \frac{1}{V_2^2} V_2 \\ \text{and } \text{FBAT}\left(T_1 = \frac{1}{V_1}, T_2 = -\frac{1}{V_2}\right) &= \frac{\left(\frac{S_1 - E_1}{V_1} - \frac{S_2 - E_2}{V_2} \right)^2}{\frac{1}{V_1} + \frac{1}{V_2}} = \text{FBAT}_{\text{het}}. \end{aligned}$$

The two hypotheses can then be written as:

$$\begin{aligned} H_0 : p_1 &= p_2 = \frac{1}{2} \\ H_1 : p_1 &\neq \frac{1}{2} \cup p_2 \neq \frac{1}{2}. \end{aligned}$$

Note that under an additive genetic model and in a sample of trios with no missing parental data, coding $T_1 = \frac{1}{V_1}$ and $T_2 = -\frac{1}{V_2}$ is equivalent to coding $T_1 = \frac{1}{n_1}$ and $T_2 = -\frac{1}{n_2}$, where n_1 and n_2 are the number of heterozygote parents of children with phenotype D_1 and D_2 (see Appendix A)².

² FBAT_{het} can be implemented in FBAT by using the offset option “-o” while coding $T_1 = 1$ and $T_2 = 0$: the software then calculates, for each allele, an offset μ used to transform the phenotypic values in $T_1 = 1 - \mu$ and $T_2 = -\mu$ that minimizes the variance of the statistics.

We show in Appendix B that using the offset option is equivalent to coding $T_1 = \frac{1}{V_1}$ and $T_2 = -\frac{1}{V_2}$, thus testing for secondary outcome.

Here, one should not code unaffected individuals as 0 but as missing to avoid that the controls interfere in the calculation of the statistics. FBAT software can be downloaded from: <http://www.biostat.harvard.edu/fbat/fbat.htm>.

3.1. The PMAFusion Test

being that carrying one copy of the allele is sufficient to develop the disease, just as for that carrying two alleles will be associated with developing a severe form of the disease.

and α) and denote $^{\circ}N_i$ the number of nonisotopes of alkyl i from the homogeneous polymer i .

Let us homogenize the differential equation relative to λ , so

under the null hypothesis of no association is $\chi^2_{(1)} = 2 \ln \frac{L_{\text{alt}}}{L_{\text{null}}}$. It is $\frac{1}{2} \chi^2_{(1)}$. We can calculate $\hat{U}$, $\hat{V}$ and $\hat{F}$

The space is 1:1 in $\text{H}^0(\Sigma)$ and $\text{H}^0(\Sigma)_{\text{even}}$, and $\frac{n}{2}$ in $\text{H}^0(\Sigma)_{\text{odd}}$. Similarly, its contribution to V is 1:1 in

that f and $f|_{\text{Stell}_\infty}$ and $\left(1, \frac{f|_{\text{Stell}_\infty}}{f}\right)_{\frac{f|_{\text{Stell}_\infty}}{f}}$ is $f|_{\text{Stell}_\infty}$. (Figure 1) Note that for all these modules, the exponential

and variance of a site including two heterozygous parents are twice those of a site with only one heterozygous parent. Consequently, an informative marker association study is more powerful for a biallelic and $\sum_{i=1}^n \frac{1}{n_i}$

is a $\mathbb{C}$ -algebra, and for $i \in \mathbb{N}$ and $j \in \mathbb{N}$ the elements $\{x_{ij} \mid i, j \in \mathbb{N}\}$ satisfy the relations

$$\xi_1 = \frac{1}{\sqrt{2}} \text{ and } \xi_2 = -\frac{1}{\sqrt{2}} \text{ in base}$$
$$(\frac{a}{b}, \frac{a}{b}, \frac{a}{b}, \frac{a}{b}, \frac{a}{b}, \frac{a}{b}, \frac{a}{b}, \frac{a}{b})$$

Table 1

$$\frac{1}{2} \log \frac{1}{2}$$

It is shown in Appendix B that $\text{Hilb}(\mathcal{H})_{\text{max}}$ is a Putnam's algebra and, in summary, the hypothesis of the Putnam's algebra is satisfied.

As expected in $\text{Pb}(\text{H}_2\text{O})_2^2+$ and $\text{Pb}(\text{H}_2\text{O})_6^{2+}$, the implicit addition of a coordinate has disappeared.

To illustrate the magnitude of the differential impact of $\text{Pb}(\text{Pb})_{\text{low}}$ and $\text{Pb}(\text{Pb})_{\text{high}}$, we could have used the following:

$$\|f(t) - g(t)\|_1 \leq \sum_{i=1}^n c_i \frac{1 - e^{-\lambda_i t}}{\lambda_i} \leq c \sum_{i=1}^n \frac{1 - e^{-\lambda_i t}}{\lambda_i} \leq c \int_0^t \sum_{i=1}^n e^{-\lambda_i s} ds \leq c \int_0^t e^{-\lambda_1 s} ds \leq \frac{c}{\lambda_1} (1 - e^{-\lambda_1 t}) \leq \frac{c}{\lambda_1}.$$

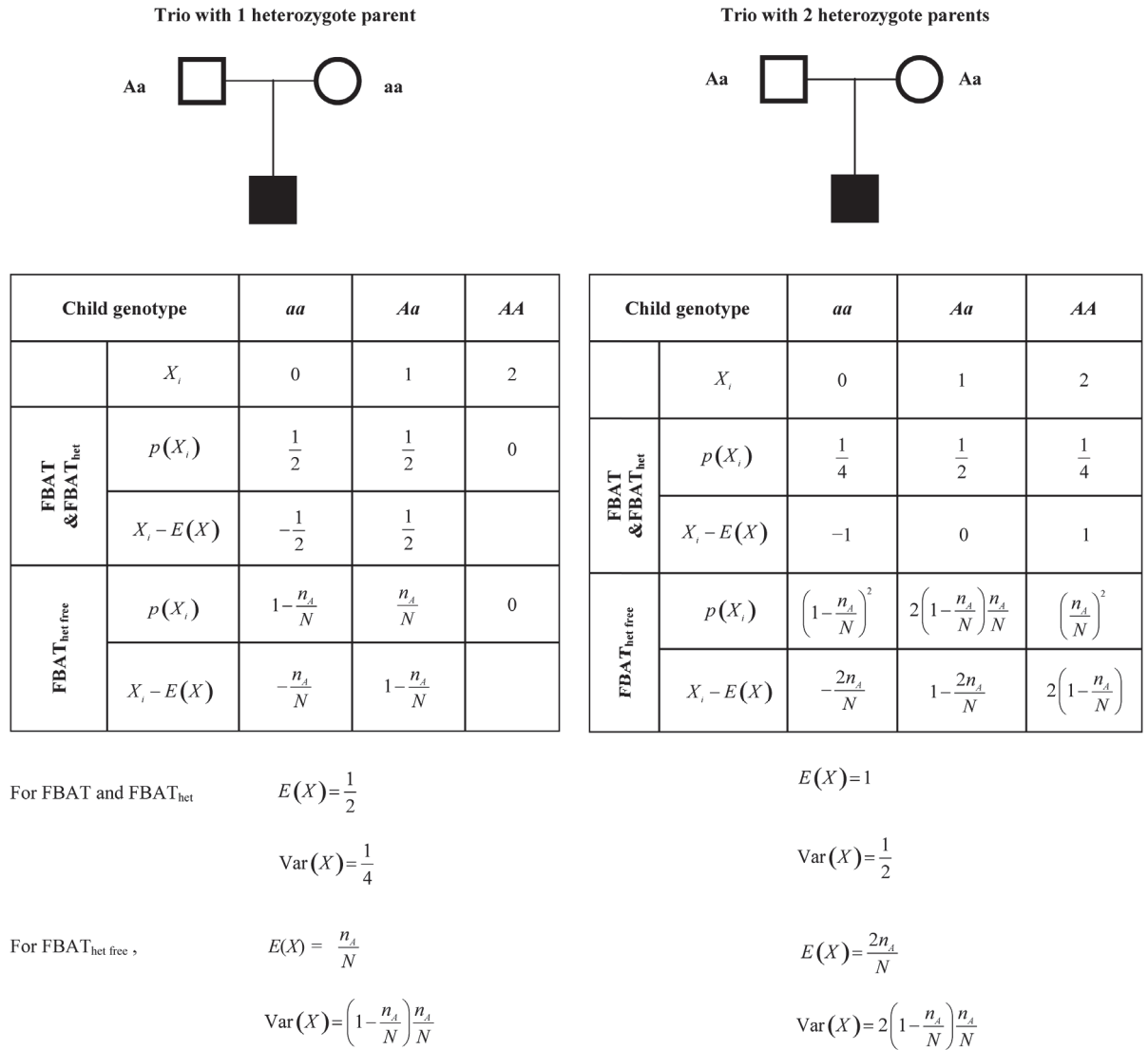
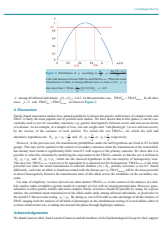


Figure 1. Contribution of a trio to FBAT, FBAT_{het} and FBAT_{het free} according to the number of heterozygote parents. In a trio with one (left panel) and two (right panel) heterozygote parents, the expected genotypes *aa*, *Aa* and *AA* of the child will vary according to the statistics used. In FBAT and FBAT_{het}, the transmission probability of an allele *A* from an heterozygote parent is $\frac{1}{2}$, whereas it is $\frac{n_A}{N}$ for FBAT_{het free} (with N denoting the total number of alleles transmitted from heterozygote parents in the whole sample, n_A the number of alleles *A* transmitted, and $\frac{n_A}{N}$ the mean transmission of allele *A*).

The distribution of ρ according to $\frac{n_A}{N}$ is shown in **Figure 2**. As an example, consider a sample of 300 trios with an affected child (150 D_1 and 150 D_2), all with one heterozygote parent. Consider the mean transmission of allele *A* is 0.7 in D_1 and 0.8 in D_2 . Then $\frac{n_A}{N} = 0.75$, $\rho = 0.75$, FBAT_{het} = 3 and FBAT_{het free} = 4, $p(\text{FBAT}_{\text{het}}) = 0.083$ and $p(\text{FBAT}_{\text{het free}}) = 0.046$.

When there is an equivalent number of transmissions of alleles *A* and *a* from *Aa* heterozygote parents to their children, $n_A = n_a = \frac{N}{2}$ and $\rho = 1$. In practice, this is observed when the mean transmission of allele



Titre : Génétique humaine des formes cliniques de la lèpre

Résumé : La lèpre est une maladie tropicale négligée qui atteint près de 200 000 personnes chaque année et dont l'agent causal est *Mycobacterium leprae*. La susceptibilité génétique de l'hôte à la maladie est bien établie, et a permis de comprendre certains mécanismes de la physiopathologie de la maladie. Il existe par ailleurs une grande variabilité inter-individuelle des manifestations cliniques de la maladie, qui s'étendent d'un pôle dit tuberculoïde à un pôle dit lépromateux. Nous avons cherché à identifier les facteurs de susceptibilité génétique à cette polarisation de la maladie. Nous avons tout d'abord décrit que le sexe et l'âge sont des facteurs non-génétiques associés à ce phénotype. Notre travail s'est ensuite appuyé sur les méthodes classiques de l'épidémiologie génétique, c'est-à-dire les études de liaison et d'association, pour identifier des variants génétiques qui influencent la polarisation de la lèpre. Nous avons utilisé une puce à ADN pangénomique avec plus de 500 000 marqueurs pour génotyper un échantillon de familles vietnamiennes comprenant 939 malades, dont 692 enfants. Nous avons identifié une liaison de la région 19p12 avec la polarisation de la lèpre. L'étude d'association n'a pas permis d'identifier de signal significatif à l'échelle du génome. Nous avons développé un nouveau test d'association pour des données familiales qui a permis d'améliorer les résultats sans atteindre la significativité. Notre travail sera prolongé par des études d'association dans deux populations cas-témoins du Vietnam et du Brésil. Nous chercherons à identifier les marqueurs causaux au sein de la région de liaison 19p12 d'une part, et à découvrir de nouveaux variants d'autre part. L'identification de marqueurs associés à la polarisation de la lèpre permettra de mieux prévoir l'évolution de la maladie et de proposer des traitements plus ciblés selon le risque génétique individuel.

Mots clés : Epidémiologie génétique, génétique humaine, lèpre, *Mycobacterium leprae*, tuberculoïde, paucibacillaire, lépromateux, multibacillaire, polarisation, tatou, transmission familiale, analyse de liaison génétique, analyse d'association, association familiale, trait catégoriel.

Title: Human genetics of clinical forms of leprosy

Abstract: Leprosy is a neglected tropical disease that affects nearly 200,000 people each year and caused by *Mycobacterium leprae*. Genetic host susceptibility to the disease is well established, and helped to understand some of the mechanisms of the disease's physiopathology. There is also a wide inter-individual variability of the clinical manifestations of the disease, which runs from a so-said tuberculoid to a so-said lepromatous pole. We sought to identify genetic susceptibility factors for this polarization of the disease. We initially described gender and age as non-genetic factors associated with this phenotype. We then based our analysis on the classical tools from the field of genetic epidemiology, namely linkage and association studies, to identify genetic variants that influence leprosy polarization. We used a DNA-microarray with 500,000 markers spanning over the whole genome to genotype a sample of Vietnamese families including 939 patients, of which 692 were children. We have found a region linked to leprosy polarization on chromosome 19p12. The association study could not identify a significant signal across the genome. We developed a new test for association designed for familial data that improved the results without reaching significance. Our work will be pursued by association studies in two case-control populations from Vietnam and Brazil. We will try to identify the causal markers within the 19p12 linkage region on the one hand, and to discover new variants on the other hand. The identification of markers associated with leprosy polarization could help us to better predict the evolution of the disease, and to offer more targeted treatments to patients, based on their individual genetic risk.

Keywords : Genetic epidemiology, human genetics, leprosy, *Mycobacterium leprae*, tuberculoid paucibacillary, lepromatous, multibacillary, polarization, armadillo, familial transmission, linkage analysis, association analysis, familial association, categorical trait.