# Recherche des facteurs génétiques contrôlant la réponse à l'infection par Mycobacterium tuberculosis et le développement d'une tuberculose maladie

Fabienne Jabot-Hanin

Université Paris Descartes

**Ecole doctorale EDSP**

*Laboratoire de génétique humaine des maladies infectieuses INSERM U1163*

# Recherche des facteurs génétiques contrôlant la réponse à l'infection par *Mycobacterium tuberculosis* et le développement d'une tuberculose maladie.

*Sommes-nous tous égaux face à l'infection par Mycobacterium tuberculosis et au développement d'une tuberculose pulmonaire ?*

Par **Fabienne Jabot-Hanin**

Thèse de doctorat de Génétique Statistique

Dirigée par Monsieur Laurent Abel

Présentée et soutenue publiquement le 12 octobre 2017

Devant un jury composé de :

Jacques FELLAY (rapporteur)

Lluis QUINTANA-MURCI (rapporteur)

Nabila BOUATIA-NAJI

Anne BOURGARIT-DURAND

Florence DEMENAIS

Sébastien GAGNEUX

Laurent ABEL

*A mes petits soleils*

« Il a mis devant mes yeux le livre de la nature et j'ai compris que toutes les fleurs qu'Il a créées sont belles, que l'éclat de la rose et la blancheur du Lys n'enlèvent pas le parfum de la petite violette ou la simplicité ravissante de la pâquerette... J'ai compris que si toutes les petites fleurs voulaient être des roses, la nature perdrait sa parure printanière, les champs ne seraient plus émaillés de fleurettes... »

Ste Thérèse de Lisieux, décédée de la tuberculose le 30 septembre 1897.

# Remerciements

On a toujours peur d'oublier de citer quelqu'un dans les remerciements, et je n'échappe pas à cette règle. Donc si je vous ai oublié, vous qui lisez ces lignes, et que vous trouviez cela injuste, veuillez bien me le pardonner.
En premier lieu, je tiens à remercier Jacques Fellay et Lluis Quintana-Murci pour avoir tenu le rôle important et chronophage de rapporteurs de ma thèse. Je remercie également Nabila Bouatia-Naji, Anne Bourgarit-Durand, Florence Demenais et Sébastien Gagneux pour avoir accepté de juger mon travail en tant qu'examinateurs. Je suis très flattée de pouvoir présenter ce travail devant un tel jury et je remercie chacun d'entre vous personnellement pour le temps que vous y avez consacré.

Ce travail de thèse a constitué une vraie aventure pour moi, dans le sens où j'ai pu découvrir le monde de la recherche de l'intérieur, un monde inconnu jusque-là, pavé d'émerveillements et de désillusions. Cette aventure m'a fait grandir (en humilité en premier lieu), et n'aurait pas été possible sans l'intervention de 3 personnes que je souhaite remercier particulièrement ici. Il s'agit du Pr Stanislas Lyonnet qui a pris le temps de m'écouter au tout début de ma quête et m'a confortée dans mon projet, du Dr Jean-Philippe Jaïs qui m'a guidée avec sa gentillesse et sa pédagogie légendaires durant mon stage de Master 2 et de qui j'ai beaucoup appris (et dont je continue d'apprendre) et enfin, du Dr Laurent Abel, qui m'a donné ma chance en acceptant de me confier ce projet et de diriger mon travail de thèse. Laurent, je te remercie pour avoir fait de ton mieux pour être disponible pour moi (avec un merci particulier pour ces derniers mois où article et manuscrit se sont chevauchés), et pour m'avoir fait découvrir ce que pouvait être la recherche. J'ai appris qu'il ne fallait pas s'enthousiasmer trop vite pour un premier résultat encourageant, qu'un article scientifique ne s'écrivait pas aussi vite qu'on pouvait le penser et que chaque mot employé avait son importance (ceci expliquant cela ….), et qu'il valait mieux arriver avec un bon résultat si on voulait te convaincre que la nouvelle méthode employée était la bonne ☺

Je remercie aussi Jean-Laurent Casanova, Emmanuelle Jouanguy, Anne Puel et Jacinta Bustamante pour leurs explications d'experts en immunologie et génétique de l'immunité, qui m'ont aidé à y voir plus clair lorsque je ne comprenais pas très bien mes résultats, et qui m'ont associée à certains de leurs projets. Je remercie Alexandre Alcaïs, le scientifique brillant et « un peu fou » du laboratoire, d'être tel qu'on imagine un chercheur dans les romans. Je le remercie pour sa bienveillance et son soutien, tout en discrétion.
Ces 4 années scientifiques ont été également riches de rencontres humaines, et j'ai été très heureuse de partager mon bureau, mes repas et les petits-déjeuners du jeudi matin avec les autres membres du « dry lab ». Un grand merci à Aziz et Jean pour leur chaleureux accueil à mon arrivée dans le laboratoire et pour leur amitié qui perdure au-delà du laboratoire. Merci Frédégonde pour toutes les discussions scientifiques et philosophiques, pour ta confiance aussi. Merci Aurélie pour ta simplicité et pour faire profiter tous les membres de l'équipe de ton expérience. Merci à Chaima, Gaspard, Vimel, Jérémy et Matthieu pour votre bonne humeur (la plupart du temps …), vos blagues et la bonne ambiance générale de l'équipe ; vous y êtes pour beaucoup ! Une pensée également pour Laurène, Maria-Emilia et Alix qui ont apporté leur fraîcheur pendant quelques mois dans le bureau épidémio, pour Audrey, Vincent et Quentin, qui m'ont précédée dans le laboratoire et pour Jocelyn qui prend la suite du projet.

Il ne m'est pas possible de terminer sans remercier les personnes qui me sont le plus chères ; merci à toute ma famille et mes proches amis d'avoir accueilli et compris ma démarche de

changement, de m'avoir encouragée et soutenue chacun à votre façon tout au long des 6 dernières années, d'avoir « subi » les looooongues discussions sur les relations thésards-directeurs de thèse au cours des déjeuners familiaux (n'est-ce pas Laura …), de vous être accrochés pour comprendre le lien entre génétique et tuberculose, et d'avoir accepté mon manque de disponibilité (en présence et en esprit) sans en avoir pris ombrage. Merci enfin à mes petites fleurs qui embellissent chacun de mes jours, et au compagnon de ma vie, pour toujours mon ami, sans qui rien de tout ça n'aurait été accompli.

# Table des matières

# Liste des tableaux

## Liste des Figures

# Liste de publications

Publications issues directement du travail de these :

**Jabot-Hanin F,** Cobat A, Feinberg J, Grange G, Remus N, Poirier C, Boland-Auge A, Besse C, Bustamante J, Boisson-Dupuis S, Casanova JL, Schurr E, Alcaïs A, Hoal EG, Delacourt C, Abel L. **Major Loci on Chromosomes 8q and 3q Control Interferon-γ Production Triggered by Bacillus Calmette-Guerin and 6-kDa Early Secretory Antigen Target, Respectively, in Various Populations** J Infect Dis. 2016 Apr 1;213(7):1173-9

**Jabot-Hanin F,** Cobat A, Feinberg J, Orlova M, Niay J, Deswarte C, Poirier C, Theodorou I, Bustamante J, Boisson-Dupuis S, Casanova JL, Alcaïs A, Hoal EG, Delacourt C, Schurr E, Abel L. **An eQTL variant of ZXDC is associated with IFN-γ production following Mycobacterium tuberculosis antigen-specific stimulation** Sci Rep. 2017 Oct 9;7(1):12800.

Publications issues de collaborations au sein du laboratoire pendant la thèse :

Cottineau J, Kottemann MC, Lach FP, Kang YH, Vély F, Deenick EK, Lazarov T, Gineau L, Wang Y, Farina A, Chansel M, Lorenzo L, Piperoglou C, Ma CS, Nitschke P, Belkadi A, Itan Y, Boisson B, **Jabot-Hanin** F, Picard C, Bustamante J, Eidenschenk C, Boucherit S, Aladjidi N, Lacombe D, Barat P, Qasim W, Hurst JA, Pollard AJ, Uhlig HH, Fieschi C, Michon J, Bermudez VP, Abel L, de Villartay JP, Geissmann F, Tangye SG, Hurwitz J, Vivier E, Casanova JL, Smogorzewska A, Jouanguy E. **Inherited GINS1 deficiency underlies growth retardation along with neutropenia and NK cell deficiency.** J Clin Invest. 2017 May 1;127(5):1991-2006.

Okada S, Markle JG, Deenick EK, Mele F, Averbuch D, Lagos M, Alzahrani M, Al-Muhsen S, Halwani R, Ma CS, Wong N, Soudais C, Henderson LA, Marzouqa H, Shamma J, Gonzalez M, Martinez-Barricarte R, Okada C, Avery DT, Latorre D, Deswarte C, **Jabot-Hanin F**, Torrado E, Fountain J, Belkadi A, Itan Y, Boisson B, Migaud M, Arlehamn CS, Sette A, Breton S, McCluskey J, Rossjohn J, de Villartay JP, Moshous D, Hambleton S, Latour S, Arkwright PD, Picard C, Lantz O, Engelhard D, Kobayashi M, Abel L, Cooper AM, Notarangelo LD, Boisson-Dupuis S, Puel A, Sallusto F, Bustamante J, Tangye SG, Casanova JL. **IMMUNODEFICIENCIES. Impairment of immunity to Candida and Mycobacterium in humans with bi-allelic RORC mutations**. Science. 2015

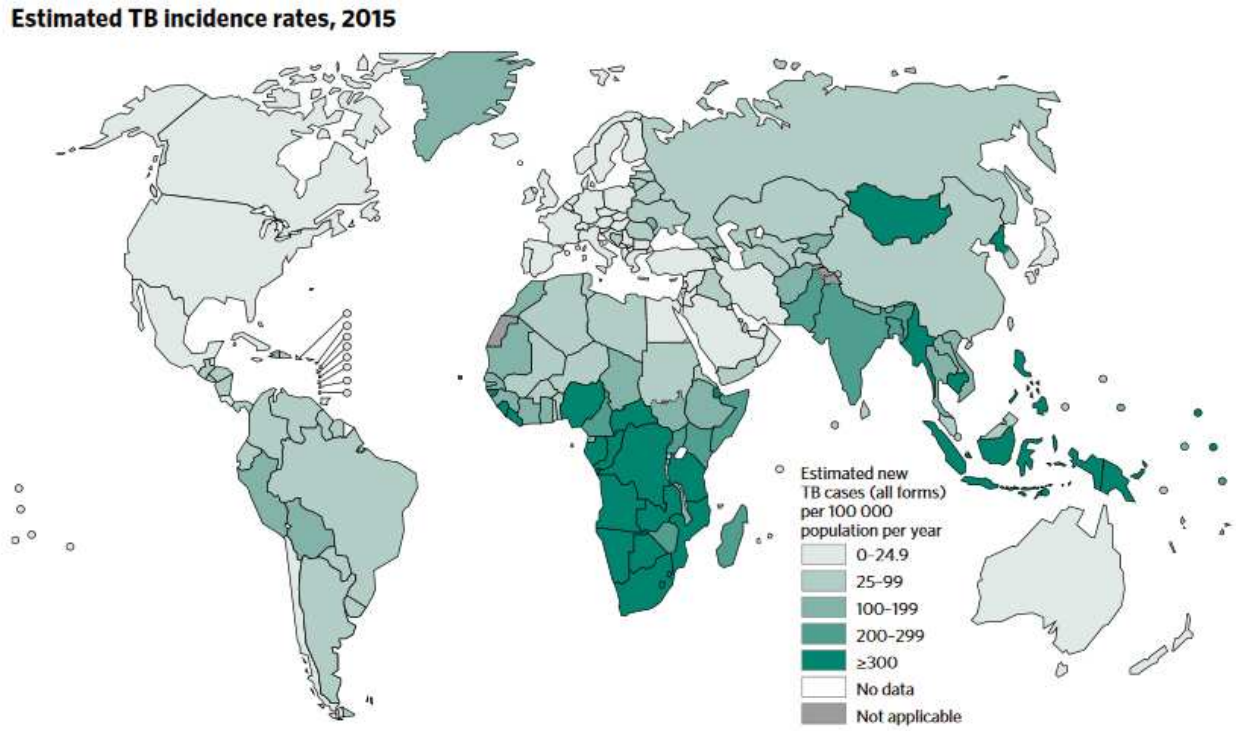
Baba LA, Ailal F, El Hafidi N, Hubeau M, **Jabot-Hanin F**, Benajiba N, Aadam Z, Conti F, Deswarte C, Jeddane L, Aglaguel A, El Maataoui O, Tissent A, Mahraoui C, Najib J, Martinez-Barricarte R, Abel L, Habti N, Saile R, Casanova JL, Bustamante J, Salih Alj H, Bousfiha AA. **Chronic granulomatous disease in Morocco: genetic, Immunological, and clinical features of 12 patients from 10 kindreds**. J Clin Immunol. 2014

# Introduction

## A. La tuberculose, une maladie mycobactérienne

Aujourd'hui encore, le complexe *Mycobacterium tuberculosis* (MTBC) fait de nombreuses victimes dans le monde, malgré la disponibilité de traitements antituberculeux depuis plus de 70 ans. On estime qu'environ un tiers de l'humanité en est infectée conduisant à 10,4 millions de nouveaux cas et 1,8 million de décès dus à la tuberculose dans le monde, en 2015 (1).

Figure 1 : Incidence de la tuberculose dans le monde en 2015 (1)




Ce complexe mycobactérien se compose de 6 espèces de bactéries (*M.africanum*, *M.bovis*, *M.caprae*, *M.microti*, *M.pinnipedii* et *M.tuberculosis*) qui infectent les mammifères, y compris les êtres humains, et qui sont à l'origine d'une maladie commune appelée

tuberculose. Une septième espèce *Mycobacterium canettii* initialement incluse dans le complexe *Mycobacterium tuberculosis* a été reclassée depuis car elle possède des caractéristiques différentes. Une phylogénie simplifiée du complexe MTBC est illustrée dans la figure 2.

*Mycobacterium tuberculosis,* qui a donné son nom au complexe, a été isolé par Robert Koch en 1882 chez un être humain et distingué du bacille infectant le bétail (*Mycobacterium bovis*) par Théobald Smith en 1898 (2). Son génome a été intégralement séquencé un siècle plus tard, en 1998, et est organisé en un chromosome circulaire comprenant environ 4000 gènes (3). De 1997 à 2010, grâce à la génétique, de nouvelles souches de mycobactéries appartenant au même complexe ont été caractérisées portant le nombre d'espèces du complexe MTBC au nombre actuel de 6 (4,5). Ce complexe mycobactérien est très homogène avec des espèces partageant 99.9% de leur génome, en particulier les ARNs ribosomaux 16S qui sont intégralement conservés (6). Il est ainsi émis l'hypothèse que les différents membres de MTBC sont des descendants d'un même ancêtre commun résultant d'un goulot d'étranglement de population relativement récent puisqu'ayant eu lieu entre 20 000 et 35 000 ans avant aujourd'hui (7).

En effet, jusqu'à récemment, il était supposé que *M.tuberculosis* dérivait de *M.bovis* et avait été transmis aux êtres humains lors du développement de l'agriculture au Proche Orient il y a environ 9000 ans, une hypothèse développée dans le livre « Guns, Germs and Steel » de Jared Diamond en 1997. Comme nous l'avons évoqué, des études plus récentes sur l'évolution de *M.tuberculosis* ont indiqué qu'en réalité *M.tuberculosis* serait plus proche d'une souche ancestrale commune qu'elle ne l'est de *M.bovis*. Cela laisse donc la porte ouverte à 2 possibilités : soit les êtres humains ont infecté le bétail avec *M.tuberculosis*, et créant ainsi par divergence une nouvelle espèce *M.bovis*, soit les 2 souches ont évolué en parallèle à partir d'une souche ancestrale commune infectant indifféremment le bétail et les hommes.

Bien que chaque souche du complexe MTBC ait été isolée dans un nombre restreint d'espèces voisines de mammifères, la spécificité de chaque souche à une espèce particulière n'est pas aussi grande que pour d'autres bactéries, car on observe des transmissions inter-espèces.

*M tuberculosis* est l'agent principal de la tuberculose humaine, mais sa transmission à des primates non-humains ainsi qu'à du bétail a été rapportée, ainsi que la transmission réciproque de *M.tuberculosis* d'animaux vers des êtres humains

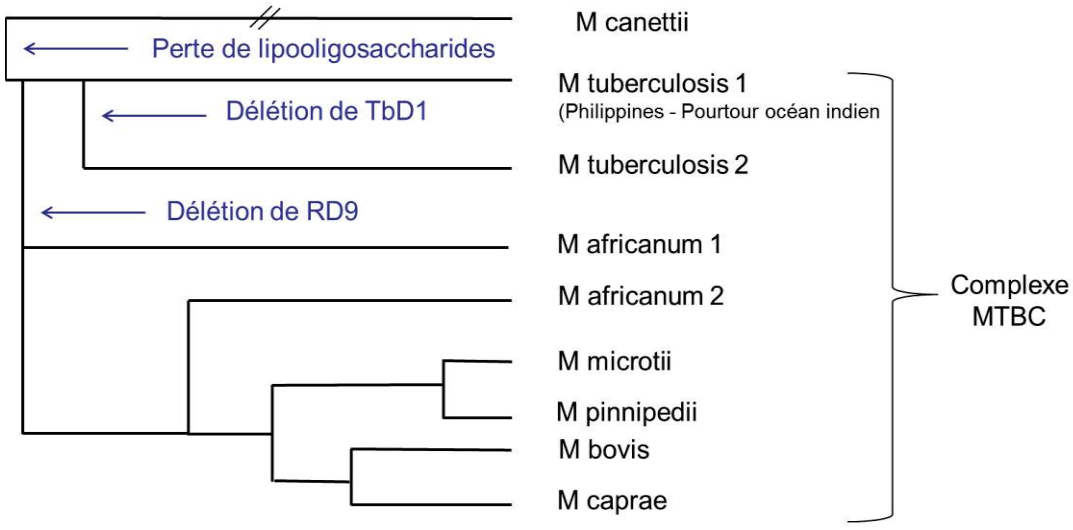
*M.bovis* peut infecter une palette particulièrement large d'espèces hôtes (animaux sauvages, bétail, primates) tandis que l'infection à *M.tuberculosis* n'a été retrouvée que chez des animaux en contact avec des hommes, comme les animaux en captivité (8). La souche étant transmissible à l'homme, les animaux infectés par *M.bovis* deviennent alors un réservoir de bacilles et un problème de santé publique.

*M.africanum* est une souche découverte en 1968 au Sénégal (9) qui possède des caractéristiques proches à la fois de *M.tuberculosis* et de *M.bovis*. Elle est localisée quasiment exclusivement en Afrique de l'Ouest et infecte principalement les êtres humains.

*M.caprae*, l'agent tuberculeux des moutons et des chèvres a été isolé chez des cerfs et des sangliers sauvages, et a été identifié dans 31% de cas de tuberculoses humaines initialement attribuées à *M.bovis* en Allemagne (10,11).

*M.microti* infecte de nombreuses espèces de petits rongeurs d'origine européenne comme les campagnols, les camélidés, les chats, les putois et les porcs. Des cas d'infection pulmonaire sévère chez des patients immunocompétents ont été rapportés (12), ainsi qu'un cas de transmission de *M.microti* d'un rongeur à un homme vivant dans une maison infestée de souris (13).

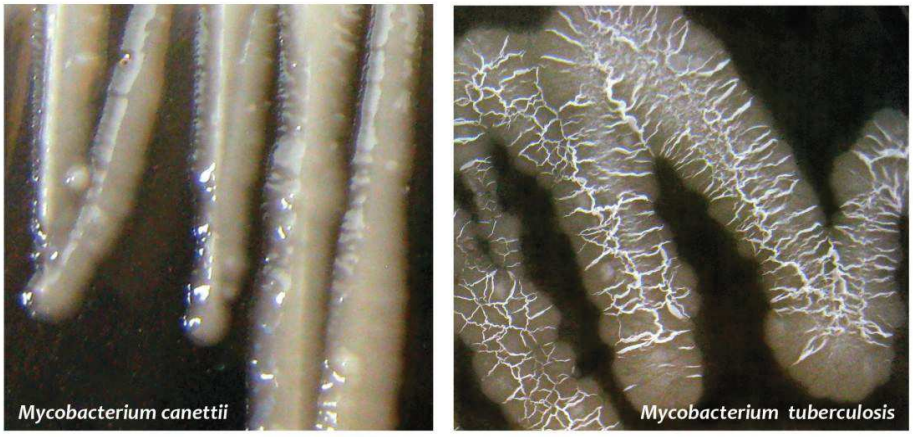
Figure 2 : Phylogénie simplifiée du complexe MTBC adaptée de (14,15)




*M.pinnipedii* a été trouvé chez des otaries et des Arctocephalinae en Australie, en Nouvelle-Zélande et en Amérique du sud. Des transmissions vers des mammifères terrestres ont été également décrites en Nouvelle Zélande (16). Il a été également retrouvé chez une momie pré-colombienne datant d'un millier d'années des traces d'ADN semblant provenir de *M.pinnipedii* (17).

*M.canettii*, placée en première intention dans le complexe MTBC, est également à l'origine de tuberculose humaine, mais celle-ci est beaucoup moins contagieuse que celle due à une souche du complexe MTBC. Depuis sa mise en évidence par Georges Canetti en 1969, moins de 100 isolats ont été décrits, ayant tous un lien avec la corne de l'Afrique.

Cette souche forme des colonies de morphologie différente de celles des 6 autres souches du complexe MTBC ; muqueuses et collantes tandis que celles de *M.tuberculosis* sont sèches, rugueuses et fripées. Des comparaisons génomiques suggèrent que *M.tuberculosis* aurait évolué par expansion clonale à partir d'un groupe de bacilles tuberculeux cousins de *M.canettii* en gagnant en virulence et persistance (18), ce gain en virulence venant de la perte de lipooligosaccharide à la surface de la mycobactérie (19).




© Roland Brosch, Institut Pasteur
*https://www.sciencesetavenir.fr/sante/tuberculose-pourquoi-la-maladie-fait-des-ravages_30045*

En résumé, toutes les souches du complexe MTBC semblent agir comme une espèce mycobactérienne génétiquement unique avec des écotypes qui se sont adaptés à leur hôte, adaptation laissant cependant possible la transmission inter-espèces.

Le complexe *M.tuberculosis* possède la propriété d'acido-alcoolo résistance, commune aux espèces du genre mycobacterium, dont la mise évidence repose sur la coloration de Ziehl-Neelsen. Il s'agit de mycobactéries à croissance lente dont la culture nécessite un milieu particulier dit de Löwenstein-Jensen. Après mise en culture, les colonies n'apparaissent qu'en 2 à 4 semaines (20). D'un point de vue clinique, le résultat de culture des micro-organismes présents dans les prélèvements du patient reste encore aujourd'hui le moyen de référence pour confirmer la maladie et identifier le traitement le plus efficace contre la souche portée par le patient. Depuis 2010, cependant, un test moléculaire rapide Xpert MTB/RIF® basé sur l'amplification d'acide nucléique s'est beaucoup répandu, suite aux recommandations de

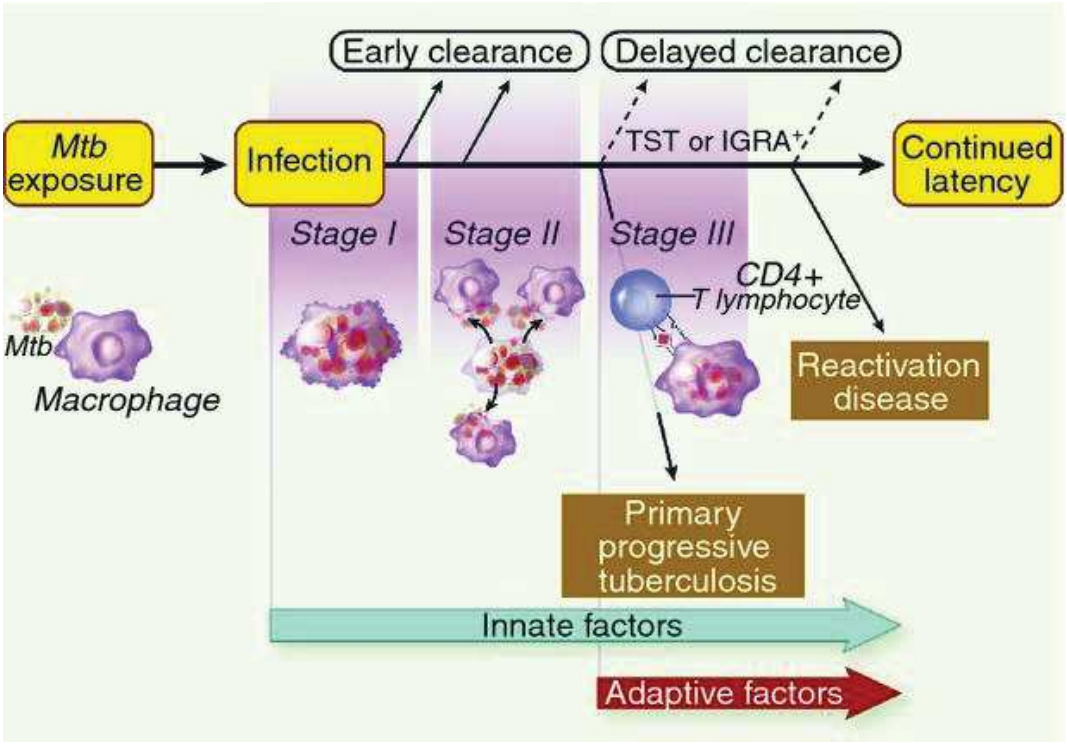
l'OMS. Ce test permet de détecter à la fois les mycobactéries du complexe *M.tuberculosis* et la résistance à la rifampicine, le médicament antituberculeux le plus important. Le diagnostic peut être posé en 2 heures et l'OMS recommande désormais cet essai comme test initial pour toute personne présentant des signes et symptômes de tuberculose. Plus d'une centaine de pays l'utilisent et 6,2 millions de cartouches ont été fournies dans le monde en 2015 (21).

## B. L'histoire naturelle de la tuberculose

Toutes les mycobactéries appartenant au complexe *M tuberculosis* sont à l'origine d'un unique type de maladie infectieuse, la tuberculose, touchant principalement le système respiratoire. Elle est caractérisée par une combinaison de fièvre, de perte de poids, de toux et d'hémoptysie lorsqu'elle touche les poumons, ainsi que par la formation de granulomes, signatures de la maladie, consistant en un noyau de macrophages infectés entouré de macrophages spumeux (dont le cytoplasme présente de petites vacuoles remplies de graisse) et d'autres cellules phagocytaires mononucléaires, ainsi que de lymphocytes (22). Ces manifestations sont retrouvées quel que soit l'hôte infecté. La transmission de l'infection se fait principalement par voie aérienne, par inhalation de bacilles en suspension dans l'air excrétés par des individus malades (gouttelettes de Pflügge). La physiopathologie de la maladie comporte encore de nombreuses zones d'ombre. En effet, bien que la plupart des êtres humains et des animaux expérimentaux développent une réponse immunitaire appropriée après avoir été infectés, cette réponse immunitaire n'éradique souvent pas complètement la bactérie. Cette réponse pousse les bacilles à adopter un état d'infection latent, cliniquement silencieux, dit quiescent, duquel ils peuvent sortir pour se réactiver (23). Sur la base de modèles expérimentaux et de travaux *in vivo* et *in vitro* chez l'homme, on peut dégager les principales étapes de l'infection par *M.tuberculosis* qui sont schématisées sur la figure 3.

Après inhalation, les premières cellules rencontrées par les bactéries dans les poumons sont les cellules phagocytaires, c'est-à-dire les macrophages alvéolaires, les monocytes, les neutrophiles et les cellules dendritiques (23,25). Les cellules phagocytaires expriment de nombreux récepteurs reconnaissant des motifs moléculaires des pathogènes (ou PRR pour « pattern recognition receptors »), permettant d'initier la réponse immunitaire. Les PRR organisent la phagocytose, la présentation antigénique et permettent l'activation de voies de signalisations intracellulaires et la production de cytokines (26).

Figure 3 : De l'exposition à *M.tuberculosis* à la maladie, issu de (24)

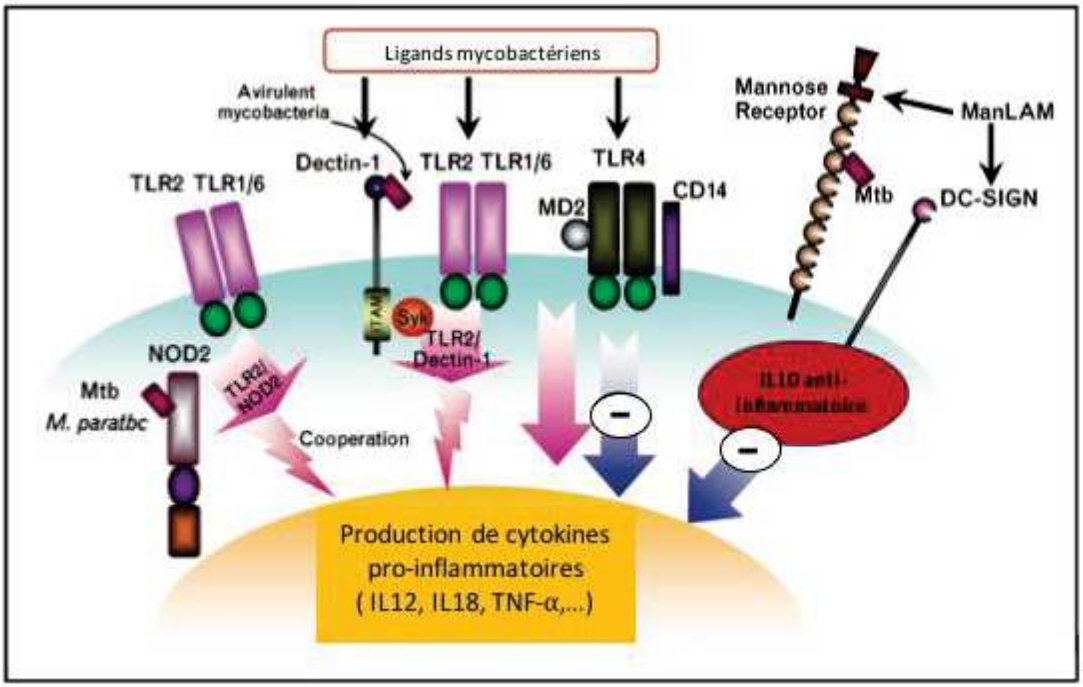



Alors que pour d'autres maladies infectieuses le recrutement de cellules phagocytaires limite voire même élimine les agents pathogènes, il semblerait que le recrutement de phagocytes vers le site d'infection des mycobactéries est favorable aux bacilles en début d'infection, en leur fournissant des niches cellulaires supplémentaires nécessaires à leur expansion (27). Les cellules phagocytaires interagissent avec *M.tuberculosis* par l'intermédiaire d'une grande diversité de PRR, comme les Toll-like récepteurs (TLRs), les récepteurs scavenger, les récepteurs du complément (CR), les récepteurs des protéines du surfactant, les lectines de type C (i.e « mannose receptor » [MR] ou « dendritic Cell-Specific Intercellular adhesion molecule-3-Grabbing Non-Integrin » [DC-SIGN]), et les récepteurs cytosoliques NOD (« nucleotide-binding oligomerization domain-containing proteins ») (28,29), schématisés sur la figure 4. La pénétration de *M.tuberculosis* dans les cellules phagocytaires semble principalement médiée par le MR et le CR3 pour les macrophages et par DC-SIGN pour les cellules dendritiques.

La reconnaissance de la mycobactérie par les TLRs induit une réponse pro-inflammatoire localisée, caractérisée par la production de cytokines comme les interleukines 12 (IL-12) et 18 (IL-18) et TNF-α, et par la production de chimiokines (CCL2 et CCL5) entraînant le

recrutement par vagues successives de neutrophiles, de cellules NK (pour « natural killer ») et de cellules T CD4+ et CD8+, chacune sécrétant des cytokines et chimiokines différentes qui amplifient le recrutement cellulaire et le remodelage du site d'infection.

Figure 4 :Interaction de la mycobactérie avec les récepteurs du macrophage et des cellules dendritiques adapté de (29)



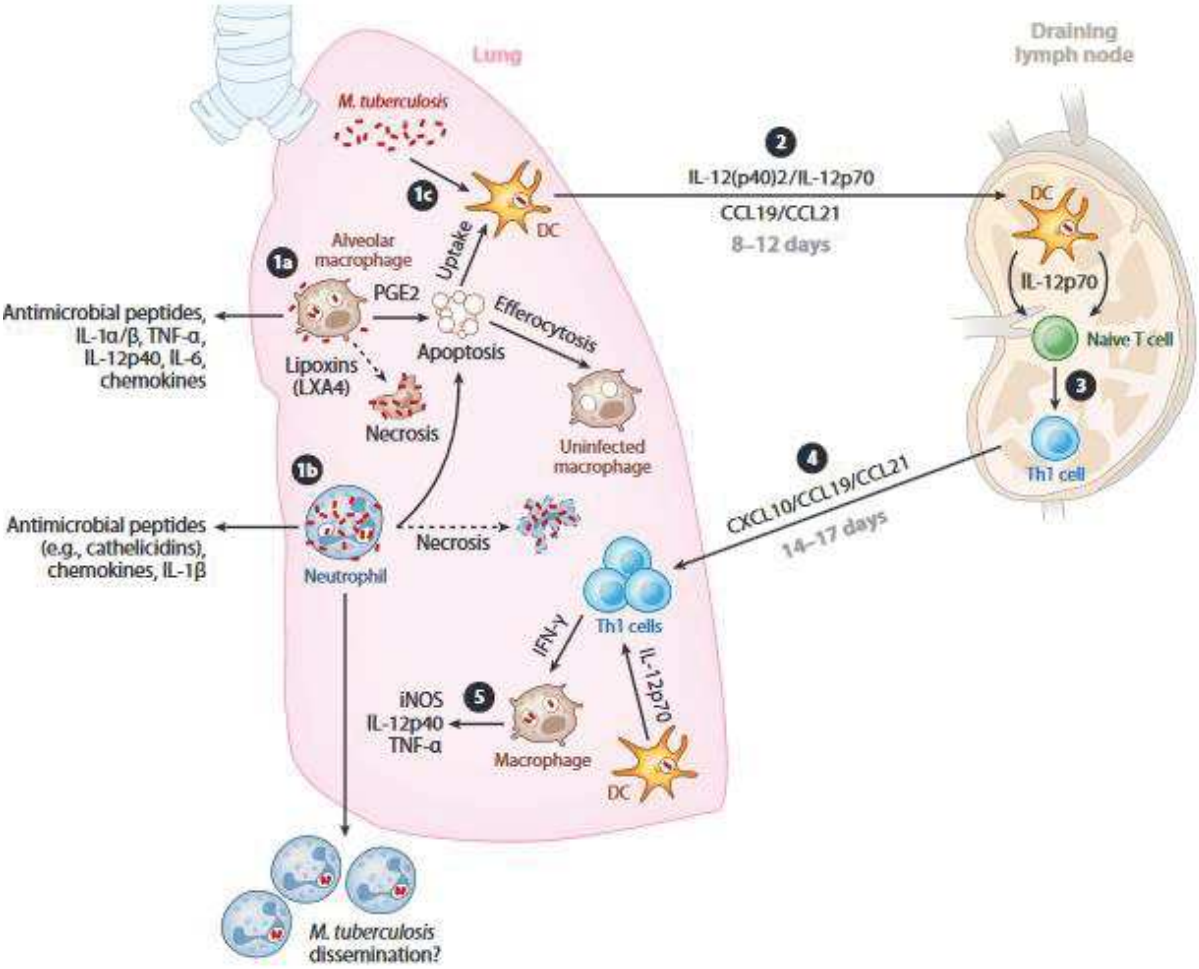

A contrario, l'interaction de la mycobactérie avec DC-SIGN et le MR semble stimuler la production d'interleukine 10 (IL-10), cytokine anti-inflammatoire, par les monocytes et les neutrophiles, contrebalançant ainsi la réponse immunitaire induite par les TLRs et profitant au bacille qui peut plus aisément se multiplier (30). Il est à noter également que les bacilles semblent avoir la faculté de retarder l'apoptose de leurs cellules hôtes, retardant d'une part l'activation de l'immunité adaptative et d'autre part se donnant ainsi le temps de croître suffisamment dans une cellule donnée avant d'être relâchés en nombre lors de la mort cellulaire (31).

A cette réaction inflammatoire locale succède le développement d'une immunité cellulaire T-spécifique caractérisée par la production d'interféron-γ (IFN-γ) (Figure 5). Les cellules dendritiques infectées migrent vers les ganglions lymphatiques régionaux où elles vont

pouvoir activer les lymphocytes T. Elles présentent les antigènes mycobactériens aux lymphocytes T CD8+ et CD4+ par le biais des molécules du complexe d'histocompatibilité (HLA) de classe I et II respectivement. En présence d'IL-12, les cellules T CD4+ vont se différencier en cellules de type Th1 caractérisées par la production d'IFN-γ. Une fois activées, les cellules T vont se multiplier et migrer vers le site d'infection primaire où elles vont recruter et activer des macrophages supplémentaires afin de détruire les bacilles intracellulaires. Il apparaît donc que l'IFN-γ est une cytokine clé de la réponse immunitaire dirigée contre les micro-organismes intracellulaires. Elle est principalement sécrétée par les cellules T et les cellules NK en réponse à l'IL-12, et active les macrophages pour permettre la destruction et l'élimination de la bactérie (32).

L'importance de l'axe IL-12/ IFN-γ dans la réponse immunitaire anti-mycobactérienne est particulièrement bien illustrée chez l'homme par le syndrome de susceptibilité mendélienne aux infections mycobactériennes (MSMD). Des mutations ont été identifiées dans 10 gènes impliqués dans l'immunité médiée par l'IFN-γ, et responsables d'une susceptibilité particulière aux infections sévères par des mycobactéries peu virulentes (*IFNGR1, IFNGR2, STAT1, IL12B, IL12RB1, NEMO, ISG15, IRF8, CYBB, TYK2*) (33,34). Certaines des mutations identifiées sont également responsables d'une susceptibilité à *M.tuberculosis* caractérisée par le développement de formes sévères de la maladie dans l'enfance (35–38).

Figure 5 : La réponse cellulaire à l'infection par *M.tuberculosis* d'après (30)




1a : macrophages des alvéoles pulmonaires - 1b : neutrophiles 1c : cellules dendritiques pulmonaires.
PGE2 = Prostaglandine E2 = médiateur lipidique pro-apoptotique – LXA4 = lipoxine A4 = médiateur lipidique pro-nécrotique.
2 : cellules dendritiques infectées migrent en 8 à 12 jours vers les ganglions lymphatiques locaux sous l'influence de l'IL-12(p40)2, del'IL-12p70 et des chimiokines CCL19 and CCL21.
3 : différentiation des cellules T naïves en lymphocytes Th1
4 : Migration des cellules Th1 vers les poumons14à 17 jours après le premier contact avec *M.tuberculosis* pour y produire de l'IFN-γ conduisant à l'activation des macrophages
5 : Production de facteurs microbicides par les macrophages pour contrôler la bactérie

La réponse immunitaire dirigée contre *M.tuberculosis* conduit à la formation du granulome tuberculeux, communément considéré comme une stratégie protective de l'hôte limitant la réplication bacillaire et prévenant la dissémination de l'infection. La réaction inflammatoire locale initiale est le point de départ pour la formation du granulome. L'apparition des lymphocytes spécifiques à *M.tuberculosis* 2 à 3 semaines après l'infection primaire marque la fin de la phase de réplication rapide des bacilles tuberculeux et le début de la phase de contrôle qui, chez la souris, est caractérisée par un nombre stable de bacilles (39). Cette population viable de mycobactéries peut persister dans un état dit quiescent, définissant l'infection tuberculeuse latente. La charge de bacilles durant cette période de latence reste inconnue en raison d'une incapacité à la mesurer jusqu'à présent, mais il apparaît que le bacille pourrait continuer à se répliquer et à accumuler des mutations « silencieusement » (Ernst, 2012; Ford et al., 2011).

Le développement d'une tuberculose clinique consistant à sortir de cet état latent est loin d'être systématique puisque parmi les individus exposés et infectés par *M.tuberculosis*, on estime que 90% d'entre eux ont une immunité efficace et durable contre la mycobactérie, et restent donc asymptomatiques tout au long de leur vie. En revanche, 5 à 15% des individus infectés développeront une tuberculose clinique dans un délai variable après la primo-infection (40). Dans certains cas, la réponse immunitaire initiale ne parvient pas à limiter la croissance bacillaire, et survient alors une tuberculose maladie dite primaire. Chez les jeunes enfants et les personnes immunodéprimées, les bacilles peuvent disséminer et engendrer des tuberculoses miliaires pulmonaires, mais aussi des tuberculoses extra-pulmonaires ou des méningites. Classiquement, on parle de tuberculose primaire lorsque la maladie se développe dans les 2 ans qui suivent l'infection initiale. Chez la plupart des individus, cependant, la réponse immunitaire parvient à contrôler l'infection, prévenant ainsi le développement d'une symptomatologie clinique. Des mois à des années après l'infection primaire, une tuberculose peut se développer du fait du « réveil » des bacilles ayant persisté à l'état quiescent (tuberculose dite de réactivation), ou de l'incapacité de l'hôte à contrôler une réinfection exogène. Même si des cas de tuberculoses cliniques ont été observés parfois plus de 30 ans après l'exposition et l'infection par *M.tuberculosis* (41), les signes cliniques semblent apparaître dans la grande majorité des cas dans les 5 ans suivant l'infection (42–44).

La forme habituelle de tuberculose de réactivation est la tuberculose pulmonaire commune, qui est la forme la plus contagieuse de la maladie. Le granulome tuberculeux devient cavitaire et les bacilles prolifèrent. La rupture de ce granulome dans les alvéoles pulmonaires libère de

nombreux bacilles viables dans les voies aériennes, résultant en l'apparition d'une toux qui facilite la dissémination bacillaire. Plus rarement, des localisations extra-pulmonaires peuvent aussi être observées du fait de la réactivation d'un foyer infectieux secondaire lorsqu'il y a eu initialement une dissémination hématogène.

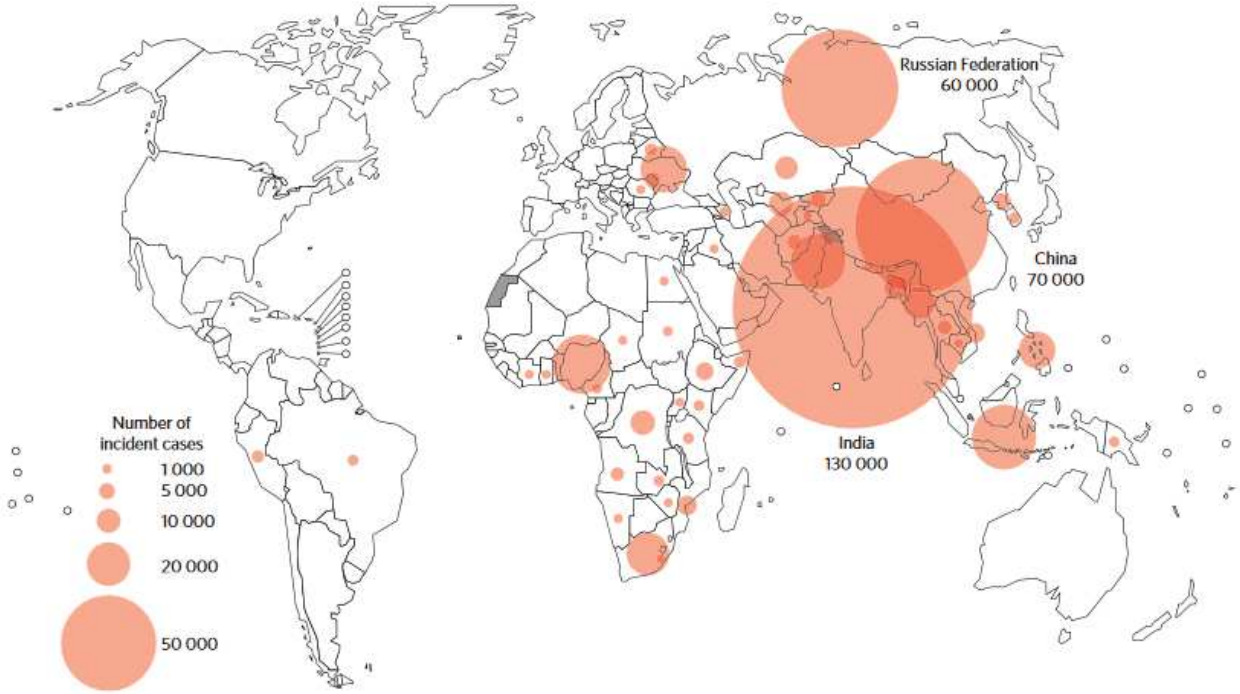
## C. Traitement et prévention de la tuberculose

Bien que certaines personnes semblent avoir réussi à éliminer complètement la mycobactérie de leur organisme sans aucun traitement (45), un traitement antibiotique est mis en place dès le diagnostic de la maladie. Le traitement de la tuberculose (curatif et préventif) a un objectif à la fois individuel et collectif. Il s'agit de traiter les patients symptomatiques, prévenir le développement d'une tuberculose clinique chez les personnes exposées à la mycobactérie, stopper la transmission du pathogène et limiter l'apparition de souches résistantes aux antibiotiques. La vaccination par le BCG largement répandue à travers le monde réduit l'incidence des formes graves de tuberculose de l'enfant mais la protection contre la tuberculose pulmonaire de l'adulte, qui est la forme la plus contagieuse, est très variable (46).

La souche historique du BCG (Bacille de Calmette et Guérin) est une souche mycobactérienne issue des travaux d'Albert Calmette et de Camille Guérin sur la tuberculose animale au début du XXème siècle. Ils ont cultivé pendant 13 ans des bactéries de *Mycobacterium bovis* sur des tranches de pommes de terre immergées dans de la bile de bœuf stérile pour en atténuer la virulence. Cette souche a été utilisée dès 1921 comme agent de vaccination chez l'homme, après 230 changements de milieu de culture. Comme à cette époque il n'était pas encore possible de conserver la souche mycobactérienne par lyophilisation ou congélation, elle a continué à être cultivée jusqu'à sa lyophilisation en 1961 après 1173 changements de milieu pour donner la souche BCG-Pasteur (47). A partir de 1924, des lots de BCG commencent à être distribués de par le monde pour que chaque laboratoire puisse fabriquer son propre vaccin à partir de souches fille du BCG. La durée de culture visant à l'atténuation de la virulence mycobactérienne variant d'un lot à l'autre, on comprend donc bien que des différences puissent alors exister entre les différents vaccins, différences qui ont pu être mises en évidence grâce aux techniques de génotypage (48,49). La principale différence entre *M.bovis* et *M.tuberculosis* d'une part, et toutes les souches de BCG d'autre part, réside dans une région génomique dénommée RD1, absente de BCG, qui contiendrait par conséquent des gènes responsables de la virulence des bacilles (50,51).

La tuberculose est traitée par antibiotiques depuis 1944, date de la découverte de la streptomycine (52). Le traitement a évolué au fil des découvertes pour aboutir aujourd'hui à un traitement standard chez l'adulte d'une durée minimale de 6 mois. Il associe 4 antibiotiques antituberculeux (isoniazide, rifampicine, pyrazinamide et éthambutol) qui doivent être pris quotidiennement pendant les 2 premiers mois, suivi d'une association isoniazide-rifampicine pendant 4 mois. L'observance du traitement est très importante pour éviter l'émergence de souches résistantes aux antibiotiques (1,53). On peut guérir l'immense majorité des cas si les médicaments sont fournis et pris correctement. Selon l'OMS, le diagnostic et le traitement de la tuberculose ont permis de sauver 49 millions de vies entre 2000 et 2015. Cependant, depuis plus de 70 ans que les médicaments antituberculeux sont utilisés, il a été mis en évidence des souches résistantes à un ou plusieurs médicaments. La résistance apparaît quand les médicaments antituberculeux ne sont pas utilisés comme il faut, du fait de prescriptions incorrectes de la part des professionnels de la santé, de médicaments de mauvaise qualité ou de la mauvaise observance du traitement par les patients. La tuberculose multirésistante (tuberculose-MDR) est une forme de la maladie due à un bacille ne réagissant pas à l'isoniazide et à la rifampicine, les deux médicaments antituberculeux de première intention les plus efficaces. On peut néanmoins soigner et guérir la tuberculose-MDR avec des médicaments de deuxième intention. Ces options thérapeutiques sont toutefois plus limitées et nécessitent une administration de longue durée (jusqu'à deux ans de traitement) de médicaments à la fois chers et toxiques. Dans certains cas, une résistance plus sévère peut se développer. La tuberculose ultrarésistante (tuberculose-XDR) est une forme encore plus grave de tuberculose-MDR due à des bacilles ne répondant pas aux médicaments de deuxième intention les plus efficaces, laissant souvent les patients sans aucune autre option thérapeutique. En 2015, environ 480 000 personnes ont eu une tuberculose-MDR dans le monde, dont près de la moitié des cas en Chine, dans la Fédération de Russie et en Inde (Figure 6). En 2015, près de 9,5 % des cas de tuberculose-MDR avaient en fait une tuberculose-XDR (1).

Figure 6 : Estimation de l'incidence des cas de tuberculose MDR ou résistant seulement à la rifampicine en 2015, pour les pays comportant au moins 1000 cas incidents.




Le traitement préventif des individus infectés est également un élément important pour limiter la diffusion de la maladie pouvant réduire de 65 à 93% l'apparition d'une tuberculose active chez les individus infectés (54,55). En pratique, le traitement le plus utilisé repose sur l'isoniazide en monothérapie, antituberculeux qui inhibe la croissance mycobactérienne mais n'a pas d'efficacité sur les bacilles quiescents. Il s'agit d'un traitement long (6 à 9 mois) dont l'observance est d'autant plus difficile que les individus sont asymptomatiques et qu'il n'est pas dénué de toxicité (principalement hépatique). Enfin, sa mise en route nécessite au préalable de porter le diagnostic d'infection tuberculeuse et d'éliminer formellement le diagnostic de tuberculose active afin de limiter l'apparition de souches multi-résistantes de *M.tuberculosis.*

# D. L'infection tuberculeuse

Après avoir été exposés à la mycobactérie, la plupart des êtres humains sont infectés par *M. tuberculosis,* ce qui signifie que la mycobactérie pénètre dans leur corps et y survit. Cette infection initiale reste la plupart du temps contrôlée et asymptomatique. Une fraction de la population est dite résistante à l'infection ce qui signifie que le pathogène ne pénètre pas dans le corps de l'hôte, ou bien son élimination y est tellement rapide qu'on ne peut identifier son passage (cf Figure 3).

Il n'existe pas de méthode diagnostique de référence de l'infection tuberculeuse permettant de mettre en évidence les bacilles et leur viabilité chez l'homme. Néanmoins, l'infection par *M.tuberculosis* laisse une « empreinte » immunologique sur le répertoire des lymphocytes T (56) qui peut être détectée par des tests immunologiques *in vivo* (test de Mantoux ou TST) ou *in vitro* (mesure de la production d'IFN-γ par les cellules sanguines ou IGRAs). Cette empreinte est à la base du diagnostic d'infection tuberculeuse. Il est important de noter que ces tests immunologiques sont des marqueurs d'infection par *M. tuberculosis* mais ne permettent pas de déterminer si la réponse immunitaire a conduit à une élimination complète du pathogène ou si celui-ci persiste dans l'organisme (56).

## 1. Test de Mantoux

Le plus ancien de ces tests immunologiques est l'intradermo-réaction à la tuberculine, ou test de Mantoux, qui reste l'examen diagnostic le plus largement utilisé de l'infection tuberculeuse. Il fut développé à la suite des travaux de Koch et de von Pirquet, en 1907, par le médecin français Charles Mantoux (1908 CR académie des sciences, Charles Mantoux + Mantoux C. L'intradermo-réaction à la tuberculine et son interprétation clinique. Presse Méd 1910;10-3). Il consiste en l'injection intradermique de 0.1 ml de tuberculine (dérivé protéinique purifié extrait de culture de bacilles tuberculeux [PPD]) à la face antérieure de l'avant-bras. Cette injection va entrainer une réaction cutanée (induration) traduisant une réaction d'hypersensibilité retardée induite par les antigènes mycobactériens chez les individus ayant une immunité cellulaire dirigée contre ces antigènes. La réaction au test de Mantoux commence quelques heures après l'injection et atteint un maximum d'induration à 48-72 heures (57). De fait, sa lecture se fait 72 heures plus tard en mesurant le diamètre d'induration de la réaction cutanée en millimètres (mm). Après l'injection intradermique de

tuberculine, les cellules présentatrices d'antigènes (i.e. les cellules de Langerhans) phagocytent et présentent les peptides microbiens aux cellules T mémoires locales, induisant la production de cytokines (telles que l'IFN-γ et le TNF-α) et de chimiokines, ainsi que l'afflux de monocytes, de macrophages et de lymphocytes au point d'injection (57). L'induration se caractérise au niveau microscopique par un œdème et une infiltration dense de cellules mononuclées, en particulier autour des petits vaisseaux sanguins (58).

Le plus souvent, le résultat du test de Mantoux est interprété de façon binaire ; un test de Mantoux positif témoignant d'une infection par *M.tuberculosis*. Cependant, le test de Mantoux n'est pas totalement spécifique de l'infection par *M.tuberculosis*. En effet, des réactions croisées existent avec le BCG et la plupart des mycobactéries environnementales qui expriment des antigènes communs avec ceux contenus dans la tuberculine. Il a été montré que l'effet des mycobactéries environnementales était faible dans les pays ayant une incidence de tuberculose intermédiaire ou élevée (59,60). En revanche, plus de 90% des individus vaccinés par le BCG développent une réaction au test de Mantoux ≥ 10mm dans les 8 à 12 semaines suivant la vaccination (61). Mais l'impact du BCG sur le résultat du test de Mantoux varie beaucoup avec l'âge à la vaccination d'une part et le délai entre la vaccination et la réalisation du test d'autre part. Dans une méta-analyse publiée en 2006 et portant sur plus de 240 000 individus vaccinés durant leur première année de vie, il a été estimé que respectivement 8.5% et 2.6% des individus seulement avaient une réaction attribuable au BCG supérieure à 10mm et 15mm, indépendamment du délai entre la vaccination et la réalisation du test de Mantoux (59). Au bout de 10 ans, l'effet de la vaccination par le BCG, lorsqu'elle a été réalisée dans la première année de vie, est quasi nul avec seulement 1% des individus ayant une réaction supérieure à 10mm attribuable au BCG. Si la vaccination a eu lieu plus tard, l'impact sur le résultat du TST est beaucoup plus fort ; et même plus de 10 ans après, 21% des individus ont toujours une réaction supérieure à 10 mm attribuable au BCG (59). Une étude récente réalisée chez des Amérindiens et des autochtones d'Alaska conforte ces résultats en montrant que le vaccin du BCG réalisé après la première année de vie pouvait avoir un effet sur la positivité du test de Mantoux jusqu'à 55 ans après (62).

Une autre limite de cette technique vient du fait que tout le monde ne monte pas de réaction immunitaire suite à l'injection intradermique de tuberculine, et qu'une anergie au test peut être observée dans certaines conditions. En particulier, parmi les patients atteints de tuberculose active, le taux de faux négatifs varie entre 10 et 25% (63). Un résultat faussement négatif peut également être observé en cas d'immunodépression, de malnutrition ou en phase

aigüe de maladie infectieuse (64). En pratique clinique, le diagnostic d'infection tuberculeuse doit donc tenir compte de la spécificité du test de Mantoux (en particulier si la personne a été vaccinée par le BCG), de sa sensibilité et de la prévalence de l'infection tuberculeuse dans la population étudiée. A spécificité et sensibilité données, la valeur prédictive positive du test de Mantoux (c'est à dire la probabilité qu'un test positif représente une vraie infection par *M. tuberculosis*) augmente avec la prévalence de l'infection tuberculeuse dans la population. Dans les pays industrialisés, où la prévalence de la tuberculose est faible, l'indication du test de Mantoux est limitée aux populations à risque et l'utilisation de seuils différentiels tenant compte des caractéristiques du test et du risque individuel d'infection est recommandée. Ainsi, le Center for Disease Control and prevention (CDC) sous la responsabilité du département des services de santé du gouvernement américain préconise 3 seuils différents pour établir le diagnostic d'infection tuberculeuse en fonction des individus. Un premier seuil d'induration à 5 mm est utilisé chez les individus les plus à risque de développer une tuberculose maladie en cas d'infection (par exemple les individus immunodéprimés). Un seuil d'induration à 10 mm est recommandé chez les individus ayant un risque accru d'infection ou de progression vers une tuberculose maladie. Enfin, chez les individus ayant un faible risque d'infection par *M. tuberculosis* un seuil de 15mm est recommandé (65). En France, où la couverture vaccinale par le BCG était quasi-totale jusqu'en 2007, le statut vaccinal et son ancienneté sont également pris en compte pour porter le diagnostic d'infection tuberculeuse.

Compte tenu de son manque de spécificité, le dépistage de l'infection tuberculeuse par la réalisation d'un test de Mantoux n'est recommandé que chez les individus à risque élevé d'infection. Dans le cadre des enquêtes réalisées autour d'un cas avéré de tuberculose (tuberculose confirmée bactériologiquement), l'utilisation de modèles prédictifs d'infection, intégrant les facteurs de risque d'infection tuberculeuse, peut permettre de mieux cibler les contacts à explorer de cette manière. Afin d'établir un tel modèle, une étude prospective a ainsi été réalisée dans la région parisienne du Val de Marne (66). Entre 2004 et 2005, un grand nombre de variables ont été systématiquement recueillies chez 325 cas incidents de tuberculose confirmée et chez 2009 contacts. Au total, huit facteurs de risque indépendants d'infection tuberculeuse (mesurée par le test de Mantoux) ont été identifiés et intégrés dans le modèle (la présence de caverne sur la radiographie pulmonaire du cas index, la forte concentration de bacilles à l'examen direct de ses crachats, le contact nocturne avec le cas index, la naissance dans un pays où l'incidence de la tuberculose est > 25 pour 100 000 habitants, le faible statut socioéconomique, l'augmentation de l'âge, le tabagisme et l'apparentement au premier degré avec le cas index). Ce modèle permet de réduire le nombre

de contacts à explorer de 26% tout en maintenant un taux de faux négatifs de 8%. Cet échantillon a été utilisé dans mon travail de thèse, comme nous le verrons plus loin.

## 2. Tests mesurant la production d'Interféron-gamma (IFN–γ)

Depuis une vingtaine d'année, des tests *in vitro* ont été développés pour le diagnostic de l'infection tuberculeuse : les tests de production d'IFN-γ (Interferon-γ release assays ou IGRA) (67–69). Leur principe est simple : les lymphocytes T d'un individu ayant déjà été exposé aux antigènes mycobactériens produisent de l'IFN-γ en réponse à une stimulation par ces mêmes antigènes mycobactériens *in vitro*. Ils mesurent donc une réponse immunitaire cellulaire, qui est la composante majeure de la réponse immunitaire vis-à-vis de *M.tuberculosis,* et peuvent permettre d'être plus spécifique que le test de Mantoux en choisissant bien les antigènes utilisés. Cette production d'IFN- γ peut être mesurée par les techniques ELISA ou ELISPOT. La dernière génération de ces tests utilise les antigènes 'early secretory antigenic target 6kDa' (ESAT-6) et 'culture filtrate protein 10 kDa' (CFP10) spécifiques de *M.tuberculosis*, codés par la région de différence 1 (RD1) du génome de *M.tuberculosis*, qui ne sont exprimés ni par le BCG, ni par la plupart des mycobactéries environnementales (50,63,69,70). Ces tests présentent donc une spécificité intéressante dans les populations vaccinées par le BCG. Il existe cependant une réaction croisée avec les antigènes analogues de *M.Leprae* (71–73) dont l'impact n'a pas été beaucoup étudié. Deux versions commerciales sont disponibles, T-Spot.TB ELISpot (Owford Immunotec, UK) qui utilise les antigènes ESAT-6 et CFP10 et le nouveau QuantiFERON-Gold In-Tube (Cellestis, Australia) qui utilise les antigènes ESAT-6, CFP10 et l'antigène TB7.7 codé par la région de différence 11 (RD11)(74).

Dans les tests commerciaux, l'incubation avec les antigènes mycobactériens est de 16 à 24h, et ils sont réalisés avec un contrôle négatif (incubation sans stimulus) et un contrôle positif (incubation avec un mitogène) internes qui aident à la détermination du résultat du test (Tableau 1). Par exemple, un individu ayant une différence de production d'IFN- γ à 5 UI/ml en l'absence de stimulation, à 5.9 UI/ml après stimulation par les antigènes de *M.tuberculosis* et à 10 UI/ml après stimulation par le mitogène sera considéré comme négatif au Quantiféron : 5.9-5=0.9 < 0.25 x 5=1.25, on se retrouve dans le cas de la seconde ligne du tableau 1.

Tableau 1 : Lecture du test Quantiféron® – TB Gold (QFT®) ELISA d'après (75)

| Valeur zéro (UI/ml)* | Antigène TB moins valeur zéro (UI/ml) | Mitogène moins valeur zéro (UI/ml)* | Résultat QFT | Rapport/interprétation |
|---|---|---|---|---|
| ≤ 8,0 | < 0,35 | ≥ 0,5 | Négatif | Infection à *M. tuberculosis* improbable |
| | ≥ 0,35 et < 25 % de la valeur zéro | ≥ 0,5 | Négatif | Infection à *M. tuberculosis* improbable |
| | ≥ 0,35 et ≥ 25 % de la valeur zéro | Tous | Positif† | Infection à *M. tuberculosis* probable |
| | < 0,35 | < 0,5 | Indéterminé‡ | Les résultats de la réponse des antigènes TB sont indéterminés |
| | ≥ 0,35 et < 25 % de la valeur zéro | < 0,5 | Indéterminé‡ | Les résultats de la réponse des antigènes TB sont indéterminés |
| > 8,0§ | Tous | Tous | Indéterminé‡ | Les résultats de la réponse des antigènes TB sont indéterminés |

*La valeur zéro correspond au contrôle négatif, valeur sans aucune stimulation. Le mitogène correspond au contrôle positif*

Des discordances sont fréquemment observées entre les tests IGRA et le test de Mantoux, et estimées entre 10 et 40% en fonction des populations testées ainsi que des seuils utilisés (67,76). La présence d'un test IGRA négatif et Mantoux positif peut être due à la vaccination par le BCG ou à des mycobactéries environnementales comme nous l'avons évoqué précédemment. Il est également avancé que les IGRAs ne seraient en mesure de détecter que les infections récentes (77,78), donc une infection ancienne ne serait détectée que par le test de Mantoux. Des discordances IGRA positif et test de Mantoux négatif sont également observées mais restent difficiles à expliquer (76,77,79–82).

En l'absence de 'gold standard' pour porter le diagnostic d'infection tuberculeuse, la sensibilité des tests immunologiques peut être évaluée chez des individus atteints de tuberculose clinique et estimée comme la proportion de tests positifs chez les cas de tuberculose active. La spécificité des tests immunologiques peut quant à elle être évaluée chez des individus non exposés à *M.tuberculosis* et estimée comme la proportion de tests négatifs chez des individus non exposés. De très nombreuses études ont évalué et comparé la sensibilité et spécificité des tests immunologiques *in vitro* (i.e. IGRAs) et *in vivo* (i.e. test de Mantoux) de cette manière. Leurs résultats ont été principalement synthétisés dans deux méta–analyses.

La première publiée en 2008 incluait 38 études et estimait par ELISA la spécificité des IGRAs à 99% (IC95% : 98 - 100%) dans les populations non exposées et non vaccinées par le BCG et à 96% (IC95% : 94 - 98%) dans les populations non exposées mais vaccinées par le BCG (68). La spécificité du test de Mantoux était très hétérogène selon le statut vaccinal par

le BCG mais élevée dans les populations non exposées et non vaccinées (97%, IC95% : 95 - 99%). Dans la plupart des études incluant des individus vaccinés par le BCG, on retrouve une proportion de positifs par le test de Mantoux dans la catégorie des individus les moins exposés supérieure à la proportion de positifs par les IGRAs. Il est toutefois à noter qu'une étude réalisée en Gambie, zone de forte endémie tuberculeuse, retrouvait une meilleure corrélation du test de Mantoux que des IGRAs avec le niveau d'exposition, malgré la vaccination par le BCG (83).

Dans la méta-analyse de Pai et al, la sensibilité des IGRAs était estimée entre 70% et 90% selon la version du test utilisée, avec une grande variabilité selon les études. De même, la sensibilité du test de Mantoux était très variable d'une étude à l'autre et estimée à 77% (IC95% : 71 - 82%) (68). Aucune comparaison deux à deux de sensibilité / spécificité n'était cependant réalisée dans cette méta-analyse et le nombre d'études ne permettait pas de distinguer les performances des tests dans les zones de forte endémie de celles dans les zones de faible endémie tuberculeuse. Il n'était pas non plus possible d'avoir des résultats selon le statut d'infection par le VIH ou l'expression clinique de la tuberculose. En 2010, une autre méta-analyse a été publiée (84), reprenant une partie des études incluses dans la première, ainsi que des études supplémentaires, toutes se limitant à la dernière version des tests commerciaux et aux cas confirmés de tuberculose. La sensibilité était alors estimée à 81% (IC95% : 78 - 83%), 88% (IC95% : 85 - 90%) et 70% (IC95% : 67 - 72%) pour le QuantiFERON, T-Spot.TB et le test de Mantoux, respectivement. En comparaison 2 à 2, les IGRAs avaient une sensibilité significativement plus élevée que le test de Mantoux. Toutefois, les études de sensibilité étaient très hétérogènes, en particulier dans les zones d'endémie tuberculeuse, et la sensibilité du test de Mantoux potentiellement affectée par les seuils de positivité utilisés dans chaque étude. Il est également important de noter que ces estimations ont été réalisées dans des populations à la fois adultes et pédiatriques, pour lesquelles les différents tests peuvent donner des résultats différents.

Il en ressort néanmoins que les IGRAs sont incontestablement plus spécifiques que le test de Mantoux dans les populations non exposées à *M. tuberculosis* de manière communautaire et vaccinées par le BCG et semblent plus sensibles pour détecter un cas de tuberculose active. Ces résultats restent cependant à confirmer dans les zones de forte endémie tuberculeuse et ne sont pas nécessairement transposables à l'infection tuberculeuse en tant que telle. En outre, le test de Mantoux a l'avantage de la simplicité par rapport aux IGRAs, en particulier en zone d'endémie tuberculeuse. Quelques études, dont les résultats ont été synthétisés par Menzies et

al. (79), ont évalué la sensibilité des IGRAs et du test de Mantoux en prenant comme indicateur d'infection un gradient d'exposition défini cliniquement sur la base de l'intensité du contact avec le cas de tuberculose. Bien que la définition des catégories d'exposition variât d'une étude à l'autre, globalement, la proportion de positifs par les IGRAs et le test de Mantoux dans la catégorie des individus les plus exposés était relativement similaire.

Au cours des dix dernières années, des études longitudinales ayant pour objectif de déterminer les performances des IGRAS et du test de Mantoux en prenant comme indicateur la progression vers une tuberculose maladie de sujets en contact avec un cas de tuberculose ont été réalisées. Une première méta-analyse de 15 études a été publiée en 2012 regroupant indistinctement des études provenant de régions du monde de faible ou forte endémie tuberculeuse, avec des taux variables de couverture vaccinale par le BCG (85). Les résultats sont très hétérogènes d'une étude à une autre et limitées par la faible incidence de la tuberculose chez les individus positifs aux IGRAs comme au TST (4 à 48 cas pour 1000 personnes-années pour les IGRA+ versus 6 à 26 cas pour 1000 personnes-années chez les TST+). Dans cette méta-analyse, le risque de développer une tuberculose active est estimé 2 à 3 fois plus élevé chez les individus ayant un résultat de test IGRA positif que chez ceux ayant un test IGRA négatif. L'ordre de grandeur est le même pour le TST, mais le pouvoir prédictif de progression vers une tuberculose clinique de tels tests n'est, de toute façon, pas très élevé, sachant que la valeur prédictive positive d'un test est proportionnelle à la prévalence de la maladie.

Même en imaginant que le TST ou le test IGRA ait une sensibilité de 95%, si on fait l'hypothèse qu'un tiers des individus testés est positif (infecté) et que la prévalence de la maladie est de 10% dans la population générale, la valeur prédictive positive du test correspondant à la probabilité de déclarer une tuberculose pulmonaire lorsqu'on a un test positif ne dépasserait pas 29%. D'autres méthodes basées sur la transcriptomique ont été utilisées afin d'identifier les différents groupes d'individus en fonction de leur statut vis-à-vis de la tuberculose (86–88). Une étude récente réalisée en Afrique du Sud a réussi à identifier une signature d'expression de 16 gènes dans le sang capable de prédire, avec une sensibilité entre 53 et 66% et une spécificité entre 80 et 83%, les individus ayant un risque de progresser vers une tuberculose pulmonaire dans les 12 mois à venir parmi ceux infectés par *M.tuberculosis* (avec TST > 10 mm ou un test Quantiferon positif) (89).

En pratique clinique, les résultats des tests IGRAs ou de Mantoux sont dichotomisés pour porter le diagnostic d'infection tuberculeuse. Les discordances apparentes sont à considérer

avec précaution car les résultats des tests de Mantoux et des IGRAs sont en premier lieu quantitatifs, et le taux de discordance varie en fonction des seuils utilisés (78,82). Si on considère les résultats des tests de Mantoux et des IGRAs de manière quantitative, leur corrélation est loin d'être nulle ; elle a été estimée entre 0.4 et 0.6 (67,90), même si dans un échantillon d'Afrique du Sud, l'analyse restreinte aux individus sensibilisés aux antigènes mycobactériens sur la base du test de Mantoux (i.e. induration ≥ 5mm) estimait une corrélation significativement diminuée entre 0.11 et 0.22. Ce dernier résultat suggère que le test de Mantoux et les IGRAs pourraient mesurer différents aspects de l'immunité anti–mycobactérienne (90).

En 2014, l'OMS a publié ses recommandations concernant la prise en charge de l'infection tuberculeuse latente, et elle préconise de ne pas utiliser les tests IGRA dans les pays à faibles revenus ou revenus intermédiaires ayant une incidence relative de tuberculose supérieure à 100 pour 100 000 habitants et par an (91), car le bénéfice d'utiliser un test IGRA en lieu et place du TST moins coûteux n'était pas fermement démontré. En revanche, dans certaines populations avec un fort risque de réaction croisée avec le TST (soit par le BCG soit par des mycobactéries environnementales), la proportion d'individus ayant un test IGRA positif est moins élevée que celle ayant test de Mantoux positif, ce qui peut s'avérer utile dans la politique d'administration d'un traitement préventif (85). Il est également intéressant de noter qu'une étude récente réalisée en Inde et portant sur 1511 contacts de cas de tuberculose confirmée dont 76 ayant développé une tuberculose clinique durant les 2 ans de suivi, n'a quant à elle pas pu clairement montrer que la positivité au test de Mantoux ou au test IGRA était associée à un risque plus fort de progression vers une tuberculose clinique après la prise en compte des facteurs de risques tels que le tabac, le BMI et la qualité de ventilation du domicile du cas index (92). Cependant, avant la prise en compte de facteurs de risque éventuels (qu'il est parfois difficile d'établir dans la vie réelle), le résultat positif au test IGRA était associé de manière significative à une plus forte progression vers une tuberculose active, contrairement au résultat du TST.

# E. Facteurs influant sur le processus d'infection et le développement d'une tuberculose active

Le développement d'une tuberculose active nécessite trois éléments :

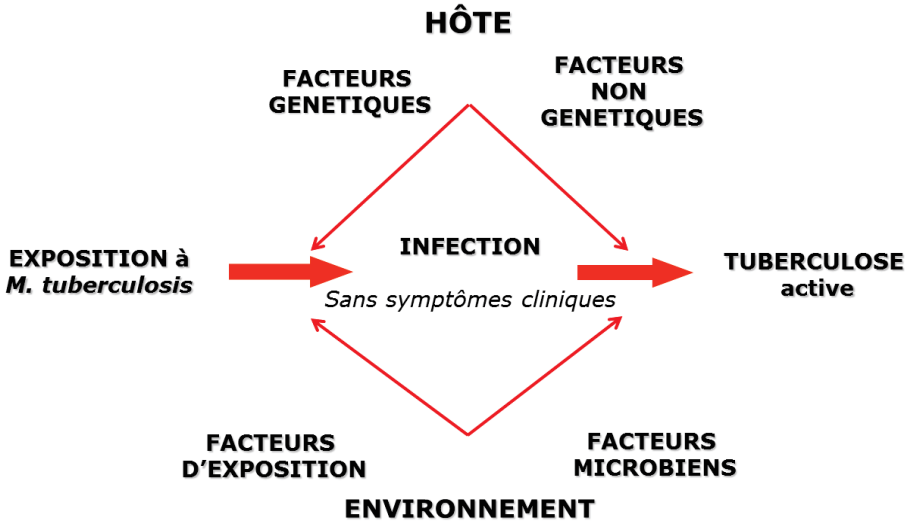
L'exposition à *M.tuberculosis*

L'infection par *M.tuberculosis*

La progression de l'infection vers une forme clinique de tuberculose

Des facteurs environnementaux (facteurs d'exposition et facteurs microbiens) et des facteurs de l'hôte (génétiques ou non spécifiquement génétiques) peuvent moduler le risque d'infection, le risque de progression vers la tuberculose active ainsi que sa présentation clinique (Figure 7).

Figure 7. Schéma général des facteurs modulant le risque de passage de l'exposition à l'infection et le développement d'une tuberculose active




## 1. Les facteurs d'exposition

Les facteurs environnementaux d'exposition à la mycobactérie modulent en premier lieu le risque d'infection tuberculeuse et sont relativement bien identifiés. Pour être infecté par la tuberculose, il faut avant tout rencontrer l'agent pathogène, c'est à dire être en contact avec des individus atteints de tuberculose active (et contagieuse). Au niveau populationnel, le

risque d'infection dépend de la prévalence de la tuberculose dans la population, et il est estimé par le risque annuel d'infection tuberculeuse ('annual risk of tuberculosis infection' ou ARTI) qui mesure l'incidence de l'infection tuberculeuse dans cette population (42). Ce risque est généralement estimé à 1% par an pour 50 à 60 cas de tuberculose pour 100 000 habitants, par an (93). Au niveau individuel, le risque d'infection dépend de la contagiosité du cas de tuberculose avec lequel on a été en contact, et de l'intensité d'exposition. De nombreuses études montrent que le risque d'infection tuberculeuse est élevé en cas de fortes concentrations de bacilles à l'examen direct des crachats du patient tuberculeux ou en présence de cavernes sur la radiographie pulmonaire de celui–ci (66,94–97). L'intensité du contact mesurée par la durée de l'exposition, la proximité du contact ou le confinement du lieu d'exposition joue également un rôle significatif dans le risque d'infection (66,94,96–99). D'un point de vue général, les facteurs d'exposition au niveau individuel jouent un rôle probablement moins important en zone d'hyperendémie tuberculeuse, car les sources de contamination sont alors multiples (100,101). Le risque d'infection tuberculeuse augmente également avec l'âge (66,94,97,102), probable reflet de l'exposition cumulée à la mycobactérie depuis le début de la vie.

## 2. Les facteurs mycobactériens

Comme nous l'avons vu précédemment, le complexe MTBC est composé de plusieurs espèces de mycobactéries dont *M.tuberculosis,* qui est la principale responsable de la tuberculose humaine. Grâce aux nouvelles techniques de biologie moléculaire, différentes souches de cette espèce mycobactérienne ont été identifiées, appartenant à 4 grandes lignées (103). Des études expérimentales *in vivo* et *in vitro,* ainsi que des études en population humaine, ont suggéré que l'espèce, la lignée et la souche de la mycobactérie pouvaient avoir un rôle dans la transmission et la présentation de la maladie (104). Dans le modèle murin, les souches de la famille Beijing (appartenant à la lignée d'Asie de l'Est) ont une virulence accrue qui se caractérise par une plus forte multiplication bacillaire, une plus grande dissémination des bacilles et un décès précoce de l'animal (105–107). Chez l'homme, une étude réalisée en Gambie a montré que le taux de progression vers une tuberculose active était significativement plus élevé chez les individus en contact avec un patient atteint de tuberculose causée par *M. tuberculosis*, et particulièrement par les souches de la famille Beijing, que chez les patients en contact avec un patient atteint de tuberculose causée par *M. africanum* (108). En revanche, le taux de transmission (mesuré par le test de Mantoux)

n'apparaissait pas différent. Les souches de la famille Beijing semblent également associées à une présentation plus sévère de tuberculose clinique, même si cette association n'est pas retrouvée dans toutes les études, et à une résistance aux antibiotiques (107,109).

Il est intéressant de noter que la même souche peut avoir des conséquences différentes en fonction de l'hôte qui l'héberge. Des résultats sur la souris le suggèrent (110), et une étude réalisée par Gagneux et al. a mis en évidence une association entre l'origine de la souche de *M. tuberculosis* et l'origine ethnique de patients atteints de tuberculose et nés à San Francisco, USA (111). Cette étude tend à montrer que certaines lignées de *M tuberculosis* seraient plus adaptées à certaines populations, bien qu'un rôle des facteurs sociaux n'ait pu être complètement écarté. A cela s'ajoutent cinq autres études récentes qui ont pu identifier une interaction entre l'espèce ou la lignée mycobactérienne retrouvée chez le patient et certains polymorphismes dans les gènes *TLR2*, *SLC11A1* (plus connu sous le nom de *NRAMP1*), *IRGM*, *ALOX5* et *MBL2,* dans le développement ou la présentation d'une tuberculose maladie (112–117).

Certaines souches pourraient également moduler le risque d'infection. Par exemple, lors d'une épidémie de tuberculose entre 1994 et 1996 au Tennessee et Kentucky (USA), due à la souche CDC1551, un taux particulièrement élevé de tests de Mantoux positifs (224 personnes soit 68% ayant un TST ≥ 10mm) a été retrouvé chez les sujets en contact avec le cas index en comparaison des 3% estimés dans la population locale n'ayant pas été en contact avec le cas index (118). L'étude d'une épidémie dans une école anglaise en 2001 a également montré qu'un seul cas index porteur de la souche CH avait infecté 254 personnes à lui tout seul sur 1128 enfants testés (chiffre basé sur le résultat du TST) (119), ce qui est bien supérieur au chiffre habituellement avancé de 16 à 20 personnes infectées par cas index. Toutefois, il est difficile de différencier le potentiel infectieux de la souche de la contagiosité du cas index (pouvant être due ou non à la souche) et l'absence de véritable référence pour porter le diagnostic d'infection tuberculeuse ne facilite pas l'interprétation des résultats.

## 3. Les facteurs de risque liés à l'hôte

Comme nous l'avons déjà dit, le développement d'une tuberculose active est loin d'être inévitable puisque parmi les individus infectés par *M.tuberculosis* on estime que 90% ont une immunité efficace et durable contre la maladie et restent asymptomatiques tout au long de leur vie. A l'inverse, 10% des individus infectés développeront une tuberculose active dans un délai variable après l'infection primaire ; environ la moitié développera une tuberculose dite

primaire dans les deux ans suivant l'infection tandis que l'autre moitié développera une tuberculose dite de réactivation dans un délai variable. En 2013, l'incidence de la tuberculose pédiatrique a été estimée aux alentours de 1 million de nouveaux cas dans le monde avec un nombre de décès s'élevant à 136000 (120). Les jeunes enfants sont les plus à risque de progresser vers une tuberculose active, particulièrement avant l'âge de 2 ans (121), et présentent plus volontiers des formes cliniques disséminées.

On observe depuis longtemps qu'après l'âge de 14 ans, le taux de notification des cas de tuberculose est deux fois plus élevé chez les hommes que chez les femmes à travers le monde (WHO 2009). De manière similaire, la prévalence de l'infection semble plus élevée chez les hommes que chez les femmes après l'âge de 15 ans (122–124). Aujourd'hui encore, il n'y a pas d'explication à ces observations, la part respective des facteurs génétiques, biologiques et socioculturels n'étant pas encore élucidée (Neyrolles and Quintana-Murci 2009), même si les hormones masculines et féminines ne semblent pas étrangères à ce déséquilibre de sexe-ratio (Nhamoyebonde and Leslie 2014). On notera que dans nos données issues d'une étude réalisée en banlieue parisienne du Val de Marne (France) sur près de 2000 contacts d'un cas de tuberculose, dont 75% âgés de plus de 15 ans, il n'a pas été mis en évidence de différence entre les hommes et les femmes concernant le risque d'infection (66,125).

Certaines situations médicales ont été associées à un risque accru de d'infection et réactivation de la tuberculose comme le diabète de type 2 (1,126–128), le traitement par des glucocorticoïdes (129) ou la maigreur et la malnutrition (1,130) sans que les mécanismes sous-jacents soient clairement établis. Le tabagisme (66,131) comme la consommation d'alcool (A. Davis et al. 2017; Francisco et al. 2017; "WHO - Global Tuberculosis Report 2016" 2017) ont été identifiés comme facteurs de risque à la fois d'infection et de tuberculose active. La carence en vitamine D a été évoquée dans les facteurs de risques possibles d'infection tuberculeuse, mais son impact reste controversé (134,135).

Comme nous l'avons mentionné précédemment, la vaccination par le BCG réduit l'incidence des formes disséminées de tuberculose de l'enfant, avec un effet protecteur compris entre 75% et 86% pour les formes méningées et miliaires (46,136,137). Récemment une méta-analyse a montré également un effet protecteur de 19% (IC : 8% - 29%) de la vaccination par le BCG sur l'infection tuberculeuse diagnostiquée sur la base des IGRAs chez les enfants de moins de 16 ans ayant été en contact avec un cas de tuberculose pulmonaire (102,138).

On observe d'autre part que l'immunodépression acquise est un facteur de risque important de progression vers une tuberculose clinique (1). Un déficit en lymphocytes T CD4+,

particulièrement chez les personnes infectées par le virus VIH, ainsi que le traitement par une thérapie neutralisant le TNF-α par des anticorps monoclonaux favorisent la réactivation bacillaire (23). En effet, le risque de développement d'une tuberculose clinique est 20 à 30 fois supérieur pour les individus infectés par le VIH que pour les non infectés, et ils sont également plus à risque de présenter une forme extrapulmonaire de la maladie (109). Il est en revanche plus difficile d'évaluer l'impact de l'immunosuppression, dont l'infection par le VIH, sur le risque d'infection tuberculeuse (par opposition au risque de progression vers une tuberculose active) dans la mesure où elle impacte le résultat des tests immunologiques à la base du diagnostic d'infection.

Les déficit immunitaires primaires sont également des facteurs de risque importants de tuberculose sévère (35) laissant entrevoir le lien entre tuberculose et génétique. C'est justement aux facteurs génétiques pouvant influer sur le processus d'infection par *M.tuberculosis* ou sur le développement d'une tuberculose active que nous nous sommes particulièrement intéressés dans ce travail de thèse, et que je vais détailler respectivement dans les parties I et II de ce manuscrit.

# F. Objectifs de la thèse

Mon travail de thèse concerne la recherche de facteurs génétiques jouant un rôle d'une part sur la réponse immunitaire après exposition à *M.tuberculosis*, mesurée par des phénotypes immunologiques *in vitro* relatifs à l'infection tuberculeuse (i.e. production d'IFN–γ après stimulation par des antigènes mycobactériens), et d'autre part sur le développement d'une tuberculose pulmonaire chez des individus préalablement infectés par la mycobactérie.

Le travail sur les phénotypes immunologiques a été réalisé en premier lieu dans un échantillon familial recruté au Val de Marne, en banlieue parisienne, dans l'entourage de cas déclarés et confirmés de tuberculose active. Cet échantillon a préalablement fait l'objet d'une étude épidémiologique sur les facteurs de risque d'infection tuberculeuse basée sur la réponse au test de Mantoux (66), ainsi que d'une analyse de liaison génétique génome-entier portant sur le trait binaire de réponse ou non-réponse à ce même test (139). Un second échantillon familial recruté dans une banlieue du Cap (Afrique du Sud) hyperendémique pour la tuberculose et avec un taux de séropositivité pour le VIH entre 2 et 4% chez les individus âgés de plus de 2 ans (140), a été utilisé en réplication pour confirmer nos résultats.

Après avoir étudié les facteurs non génétiques ayant une influence sur nos phénotypes d'intérêt, nous avons d'abord réalisé des analyses de liaison génétique génome-entier sur les traits quantitatifs de production d'interféron-gamma suite à différents stimuli mycobactériens. Ce premier travail est détaillé au chapitre I-B du manuscrit. Nous avons ensuite réalisé un 'fine-mapping' des régions de liaison trouvées afin d'essayer d'identifier de manière plus précise un ou des variants associés aux phénotypes étudiés. Cela a été réalisé grâce à un génotypage dense des régions de liaison dans les 2 échantillons sus-cités, suivi d'une imputation de milliers de variants toujours au sein de ces régions d'intérêt à partir des données du projet 1000 génomes. Ce second volet est détaillé au chapitre I-C du manuscrit.

Le second objectif de ma thèse concerne l'étude des facteurs de susceptibilité génétique à la tuberculose pulmonaire, et en particulier l'impact des variants génétiques dits « rares » (par opposition aux variants communs étudiés dans les études d'association génétiques classiques). En effet, grâce aux progrès technologiques et des connaissances en matière de génétique, de nombreuses études ont été conduites afin d'identifier les facteurs génétiques pouvant faire basculer les êtres humains d'une tuberculose latente à une tuberculose active. Des variants génétiques dits communs (d'une fréquence supérieure à 3-5% dans la population générale) ont pu être mis en évidence dans certaines populations grâce aux GWAS (*Genome Wide Association Studies*), mais n'ont pas toujours pu être répliqués dans d'autres régions du monde, et n'expliquent au total qu'une faible partie de l'héritabilité de la maladie. Nous avons testé une autre hypothèse en analysant des variants génétiques plus rares, ce qui a été rendu possible par les nouvelles technologies de séquençage. Un échantillon marocain de 120 patients tuberculeux et de 120 contrôles infectés par *M.tuberculosis* mais non malades a été séquencé sur l'exome entier (ensemble des parties codantes du génome), et nous avons alors analysé les variants en utilisant des méthodes d'agrégation par gène. Ce troisième volet de ma thèse est détaillé dans la seconde partie du manuscrit (partie II).

# I - Génétique humaine de l'infection tuberculeuse

## A. Introduction

L'existence de facteurs génétiques prédisposant à la déclaration d'une tuberculose active est à l'heure actuelle relativement bien établie, même si l'identification de tels facteurs est loin d'être terminée comme nous le verrons dans la seconde partie du manuscrit. En revanche, les facteurs génétiques influençant l'infection par *M.tuberculosis* après avoir été exposé à la mycobactérie n'ont été que plus rarement étudiés. Des études portant sur du personnel hospitalier ou sur l'entourage proche de patients tuberculeux ont montré que, sur la base du test de Mantoux ou des IGRAs, jusqu'à 50 % des personnes ayant été en contact avec la mycobactérie ne semblent pourtant pas être infectées (24,141–143). Dans des maisons de retraite de l'Arkansas aux Etats-Unis, il a été mis en évidence sur la base du test de Mantoux, que les Afro-américains étaient deux fois plus susceptibles à l'infection tuberculeuse que les américains d'origine européenne, cette différence ne pouvant être expliquée par des facteurs environnementaux ou sociaux (144). En 2008, une étude prospective réalisée en région parisienne dans le Val-de-Marne sur l'entourage familial de 325 cas incidents de tuberculose a montré notamment qu'indépendamment du temps de contact nocturne avec le malade et de la contagiosité du cas index, l'apparentement au premier degré entre le patient et le contact constituait un facteur de risque d'infection tuberculeuse (66). Tout cela suggère l'existence de facteurs de susceptibilité génétique à l'infection tuberculeuse.

### 1. Etudes sans marqueurs génétiques

La première façon de quantifier la contribution des facteurs génétiques à la variabilité observée du phénotype d'infection est de calculer l'héritabilité du trait, c'est-à-dire la proportion de variance phénotypique expliquée par la génétique. De façon générale, la variance d'un trait quantitatif peut être décomposée en la somme d'une composante génétique, d'une composante environnementale, de la covariance entre les 2 composantes (c'est-à-dire leur dépendance) et d'une composante d'interaction gène-environnement (145). La variance génétique peut elle-même se décomposer entre une composante additive transmissible (somme des effets moyens de tous les allèles transmis des parents aux enfants), une

composante de dominance (somme des effets correspondant à l'interaction entre les allèles à un locus donné) et une composante d'épistasie (interaction entre les allèles à différents loci). L'héritabilité est généralement estimée à partir d'un modèle plus simple, le modèle polygénique, qui omet à la fois la covariance et l'interaction gène-environnement, ainsi que l'épistasie génétique. Elle est alors estimée comme étant la proportion de variance attribuable à la composante génétique additive transmissible (héritabilité au sens strict) ou à la variance génétique totale, y compris la composante de dominance (héritabilité au sens large). Dans les 2 cas, le calcul de l'héritabilité est réalisé à partir de l'étude de la ressemblance phénotypique entre apparentés, donc dans des échantillons familiaux. Il est nécessaire de garder à l'esprit que dans une famille nucléaire composée des parents et de leurs enfants, il est difficile de faire la différence entre environnement partagé et facteurs génétiques en commun, et que cela peut conduire à une surestimation de l'héritabilité si on y inclut ces 2 composantes de variance. Il est cependant possible d'utiliser la corrélation entre époux pour mesurer l'importance de cet environnement familial partagé, car en l'absence de consanguinité, ceux-ci n'ont pas de matériel génétique en commun. Un cas particulier des études familiales est l'étude de jumeaux. Les jumeaux monozygotes partagent le même patrimoine génétique, alors que les jumeaux dizygotes partagent en moyenne la moitié de leur patrimoine génétique, comme des frères et sœurs classiques. On admet généralement que toute discordance phénotypique est d'origine environnementale chez des jumeaux monozygotes, et d'origine à la fois environnementale et génétique chez des jumeaux dizygotes. Si l'on fait l'hypothèse que l'environnement partagé est le même chez les 2 types de jumeaux, la comparaison de leurs corrélations permet d'estimer la composante génétique. Dans le cas des phénotypes liés aux maladies infectieuses, il est bien évidemment nécessaire que l'exposition à l'agent infectieux soit homogène au sein de la famille, sous peine de biaiser énormément le calcul d'héritabilité.

Dans le cadre de l'infection tuberculeuse, les facteurs de susceptibilité génétiques ont été recherchés en premier lieu en utilisant la réponse au test de Mantoux comme indicateur. La corrélation entre germains (frères et sœurs) du diamètre d'induration induite par le TST a été estimée à 0.46 dans une étude réalisée au Chili, chez des enfants âgés de moins de 14 ans, exposés à un cas de tuberculose, vaccinés par le BCG et ne présentant eux-mêmes pas de signes cliniques de tuberculose (146). L'héritabilité pourrait donc en être estimée à 92 % (deux fois la corrélation entre germains) en négligeant la composante d'environnement partagé (147). Deux études de jumeaux ont également estimé l'héritabilité de la réponse au TST à 28 % dans un petit échantillon de jumeaux âgés de moins de 3 ans et ayant été vaccinés à la naissance par le BCG, au Chili, zone de faible endémie tuberculeuse (148), et à 71 % en

Gambie, zone de forte endémie tuberculeuse, dans une population de jumeaux non malades âgés de 12 à 83 ans (149). Plus récemment, Cobat et al ont réalisé une étude de ségrégation dans un échantillon familial colombien exposé à un cas de tuberculose avéré et ont estimé l'héritabilité de la réponse au TST à 72 % en tenant compte de facteurs de confusion que sont l'âge, la vaccination par le BCG et le fait de partager le même lit qu'un patient tuberculeux (150). Les résultats des études semblent donc montrer une héritabilité importante de la réponse au test de Mantoux, cependant plus modérée dans le cas de très jeunes enfants ayant été vaccinés à la naissance par le BCG.

L'héritabilité de la production d'IFN-γ a été estimée en Afrique-du-Sud aux environs de 43% suite à une stimulation par le bacille du BCG et à 58% après stimulation par l'antigène ESAT-6 (151). En Ouganda, l'héritabilité de la production d'IFN-γ suite à stimulation par des antigènes de *M.tuberculosis* (dont l'ESAT-6) a été estimée entre 17 et 48% en fonction du statut TST, de la présence d'une forme clinique de tuberculose et de la sérologie HIV des individus testés, (152,153). En Gambie également, la quantité d'IFN-γ en réponse aux antigènes PPD a été estimée héritable à hauteur de 39 à 41% et la réponse à *M.tuberculosis* inactivée à hauteur de 39 % chez des jumeaux sains (149,154). Bien que ces études aient toutes été réalisées dans des zones de forte endémie tuberculeuse, leurs résultats ne sont pas directement comparables du fait des différentes stratégies utilisées pour estimer l'héritabilité, de l'hétérogénéité des facteurs de confusion pris en compte et de faibles tailles d'échantillon pour certaines qui peuvent conduire à des estimations imprécises de la variance génétique. Il s'en dégage malgré tout l'idée que la production d'IFN-γ, marqueur de l'intensité de la réponse immunitaire adaptative aux mycobactéries, connaît une part non négligeable de déterminisme génétique, comme c'est le cas pour la réponse au test de Mantoux.

L'étape suivant l'étude des corrélations familiales est de modéliser ces corrélations pour identifier la présence d'un gène majeur parmi l'ensemble des facteurs génétiques et environnementaux intervenant dans le déterminisme du trait considéré. C'est le principe de l'analyse de ségrégation. Le terme de gène majeur ne signifie pas qu'il s'agit du seul gène intervenant sur le phénotype étudié, mais que, parmi l'ensemble des gènes impliqués, il en existe un dont l'effet est suffisamment important pour être distingué des autres. Cette méthode repose sur la modélisation mathématique de la probabilité d'observer le phénotype d'un individu en fonction des différents facteurs génétiques et/ou environnementaux pouvant influencer ce phénotype. Elle permet de tester si la distribution familiale d'un phénotype

observé est compatible avec les distributions attendues sous différentes hypothèses de transmission, notamment celle de l'existence d'un gène majeur transmis de manière mendélienne, en prenant en compte simultanément des facteurs environnementaux et des corrélations familiales résiduelles pouvant refléter un environnement partagé et/ou un fond génétique commun. Si un gène majeur est mis en évidence, cette méthode permet d'en préciser les caractéristiques : fréquence allélique, moyennes et variances génotypiques dans le cas d'un trait quantitatif et pénétrance dans le cas d'un trait binaire. Dans l'étude colombienne de Cobat et al (Aurélie Cobat et al. 2012), l'existence d'un gène majeur d'expression co-dominante influençant l'intensité de la réponse au TST a été mise en évidence, avec une fréquence de l'allèle prédisposant aux valeurs élevées du TST estimée à 41 % dans la population étudiée.

## 2. Etudes avec marqueurs génétiques

Les facteurs génétiques semblant avoir une influence importante sur l'intensité de la réponse au TST, leur localisation sur le génome et l'identification de marqueurs génétiques précis font l'objet des études de type analyse de liaison et analyse d'association génétiques. L'analyse de liaison pangénomique vise à identifier et localiser sur le génome des régions chromosomiques contenant un ou plusieurs gène(s) régulant un phénotype d'intérêt. Elle étudie la co–ségrégation d'un locus influençant le phénotype et de marqueurs génétiques dont on connaît la localisation. Elle nécessite donc un échantillon familial. Les études d'association visent quant à elles à localiser plus précisément les variants génétiques influençant le phénotype étudié. Elles peuvent être réalisées en population générale (i.e. sur des individus non apparentés) ou en famille. Les marqueurs génétiques les plus utilisés en épidémiologie génétique à l'heure actuelle sont les polymorphismes mono–nucléotidiques (*single nucleotide polymorphism* en anglais ou SNP), qui correspondent au remplacement d'un nucléotide par un autre dans la séquence d'ADN. En épidémiologie génétique, on peut distinguer deux stratégies principales d'association avec marqueurs génétiques pour les traits complexes : l'approche dite "test d'hypothèse" et l'approche génome entier ou génération d'hypothèses. La première approche consiste à sélectionner des gènes candidats *a priori* en se fondant sur des données issues de modèles animaux ou de données humaines (études expérimentales *in vitro* ou observations *in vivo*) alors que la seconde permet de générer de nouvelles hypothèses par un criblage complet du génome. Historiquement, ce criblage du génome était réalisé par une analyse de liaison et les gènes candidats étaient définis sur la base de leur localisation sous un

pic de liaison. Cette stratégie a été abondamment utilisée et a donné de très bons résultats dans l'étude des maladies monogéniques ; c'est l'approche que nous avons privilégiée et que nous détaillerons dans les chapitres 1 et 2 de cette première partie du manuscrit. Plus récemment, des études d'association pangénomiques testant l'association entre le phénotype d'intérêt et plusieurs centaines de milliers de SNPs couvrant l'intégralité du génome se sont développées, ce sont les GWAS (*Genome Wide Association Study(ies)*). Avec le développement de technologies de séquençage de nouvelle génération, le séquençage systématique des régions codantes du génome (exome) devient également possible, et sera détaillé dans la seconde partie du manuscrit. Après identification de gènes et/ou de variants associés avec le phénotype d'intérêt, pour confirmer cette découverte basée sur les statistiques, une validation par des études fonctionnelles est nécessaire. Cette validation fonctionnelle est cependant difficile à mettre en œuvre dans le cadre des phénotypes complexes dans la mesure où les effets attendus sont subtils et les modèles expérimentaux appropriés difficiles à développer.

Concernant l'infection tuberculeuse, une analyse de liaison génome-entier réalisée sur un échantillon de 128 familles d'Afrique-du-Sud (en zone de très forte endémie tuberculeuse) a identifié 2 loci majeurs influençant la réponse au test de Mantoux ; le locus TST1 situé sans la région chromosomique 11p14 contrôlant la réponse intrinsèque au TST de manière binaire (réponse nulle versus réponse positive) et le locus TST2 situé dans la région chromosomique 5p15 contrôlant l'intensité de la réponse au TST (155). Le locus TST1 a ensuite été répliqué en zone de faible endémie tuberculeuse dans l'échantillon du Val de Marne que nous avons également étudié dans le cadre de cette thèse (139), renforçant l'idée de l'existence d'un ou plusieurs gènes contrôlant la résistance à l'infection tuberculeuse indépendamment de la réponse des lymphocytes T. Ce locus TST1 est tout proche d'une région chromosomique notée TNF1, identifiée par analyse de liaison et régulant la production de TNF-α suite à des stimulations mycobactériennes dans la même population sud-africaine (156). Ces résultats mettent en lumière un lien étroit entre la production de la cytokine et la résistance à l'infection par *M.tuberculosis*.

Le test de Mantoux ne peut cependant pas être considéré comme spécifique de l'infection par *M.tuberculosis* ainsi que nous l'avons détaillé précédemment dans l'introduction générale, et les tests IGRA sont aussi utilisés aujourd'hui dans le diagnostic de l'infection tuberculeuse, en particulier dans des zones où la vaccination par le BCG est généralisée et la prévalence de la tuberculose assez faible. Le test de Mantoux et les IGRA mesurant des aspects différents de

l'infection par *M.tuberculosis*, leurs résultats n'étant pas parfaitement corrélés, il paraît important de les étudier séparément d'un point de vue génétique pour progresser dans la compréhension du mécanisme infectieux. Jusqu'à présent, ces facteurs génétiques ont été moins étudiés que ceux influençant le TST, à notre connaissance aucune étude utilisant des marqueurs génétiques n'a été réalisée pour étudier la réponse aux IGRA. Le premier objectif de mon travail de thèse a donc consisté à essayer de localiser sur le génome les facteurs génétiques influant sur la production d'IFN-γ en réponse à une attaque mycobactérienne.

# B. Analyses de liaison de la production *in vitro* d'IFN-γ suite à des stimulations mycobactériennes (125)

Afin de localiser les facteurs génétiques influençant les résultats des IGRA après exposition à *M.tuberculosis*, nous avons réalisé une analyse de liaison génome entier de plusieurs phénotypes de production d'IFN-γ en réponse à des stimulations mycobactériennes dans des familles issues de l'entourage immédiat de patients tuberculeux recrutés près de Paris, à l'hôpital de Créteil (Val-de-Marne), puis en seconde intention, dans des familles vivant en banlieue du Cap, en Afrique du Sud, où la tuberculose est hyper-endémique.

## 1. Echantillons d'étude et marqueurs génétiques

### 1. Echantillon primaire

Une étude prospective a été réalisée dans le Val de Marne en banlieue parisienne concernant l'entourage familial immédiat de patients tuberculeux. Le Val de Marne est un département français avec une faible incidence de la tuberculose (22 cas pour 100 000 habitants à l'époque de l'étude), mais tout de même plus fortement touché que la moyenne des régions françaises présentant 8.8 cas pour 100 000 habitants. Cette étude, dont les premiers résultats ont été publiés en 2008 (66), a consisté à inclure entre avril 2004 et janvier 2009 des individus ayant été en contact avec un cas de tuberculose avéré (culture cellulaire positive). En effet, en France, chaque nouveau cas de tuberculose est rapporté aux autorités de santé publique, en l'occurrence au centre de lutte contre la tuberculose, afin de pouvoir réaliser des études épidémiologiques d'une part, et d'autre part, pour identifier des personnes ayant été exposées au bacille tuberculeux en vue de pouvoir prévenir la maladie. Dans cette étude, les individus

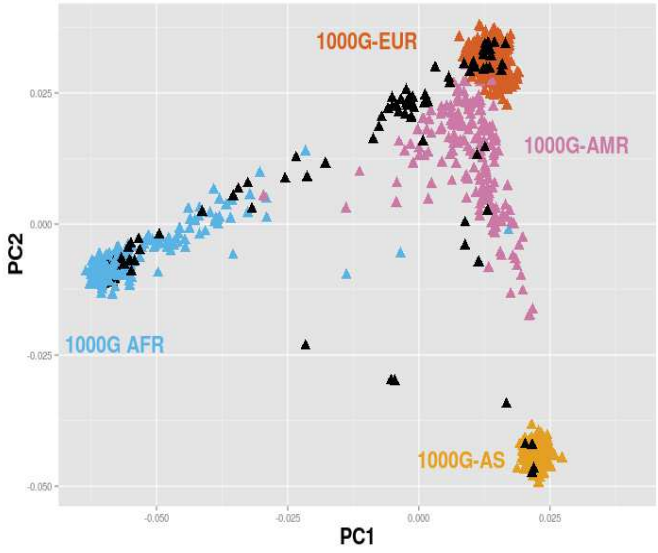
définis comme « contacts » sont tous les individus partageant le foyer d'un cas de tuberculose avéré durant les trois mois précédant le diagnostic de tuberculose. Cette étude a été approuvée par le Comité Consultatif de Protection des Personnes se prêtant à des recherches biomédicales de l'hôpital Henri Mondor (Créteil, Val de Marne) et un consentement écrit a été obtenu de tous les participants à l'étude ou de leurs parents quand il s'agissait d'enfants mineurs. Au total, 590 individus appartenant à 173 familles nucléaires ont été inclus dans cette étude. Un questionnaire a été complété par chaque individu afin d'établir précisément les relations familiales et les facteurs de risque liés à l'infection. De nombreuses covariables ont ainsi été recueillies dont parmi elles l'âge en années, le sexe, le statut vaccinal du BCG, la contagiosité du cas index caractérisée par la présence de cavernes sur la radiographie des poumons et la présence de bacilles à l'examen direct dans les crachats, le pays de naissance de l'individu et l'incidence de la tuberculose dans celui-ci, le nombre de pièces du logement du cas index et le nombre de personnes y vivant, l'estimation du nombre d'heures de contact entre chaque individu et le cas index au cours des 3 mois précédant le diagnostic, la présence d'une complémentaire santé, la catégorie socio-professionnelle, le statut tabagique ainsi que les antécédents de tuberculose.

Un criblage complet du génome a été réalisé par le Centre national de Génotypage (CNG) à l'aide du panel Illumina linkage V contenant 6056 SNPs, dont 5687 sur les chromosomes autosomes. Les SNPs de ce panel sont uniformément distribués sur le génome et espacés de 0.62 cM en moyenne. Ils ont également été sélectionnés sur des critères de fréquences alléliques pour fournir le maximum d'information en analyse de liaison (plus de 95% en moyenne sur l'ensemble du génome dans une population caucasienne) (http://www.illumina.com). Au total, 5376 SNPs autosomiques ont été retenus pour l'étude, après élimination des SNPs présentant un taux de génotypage < 90%. L'écart à l'équilibre d'Hardy Weinberg de ces SNPs autosomiques n'a pas été pris en compte dans les critères qualité du fait de la très grande hétérogénéité ethnique de la cohorte. Cette forte hétérogénéité ethnique a pu être illustrée grâce à une analyse de structure de population ou analyse en composantes principales réalisée à l'aide de 5350 marqueurs de la puce communs avec le panel de référence du projet 1000 Génomes. Le principe de la méthode est de déterminer les principaux axes de variation génétique de l'échantillon de référence et d'assigner à chaque individu de l'échantillon de référence comme de l'échantillon à analyser des coordonnées selon chacun des axes de variation. Les vecteurs propres ont tout d'abord été calculés sur les 1092 individus du projet 1000 Génomes à l'aide du logiciel EIGENSTRAT et de sa fonction SMARTPCA (157), puis dans un second temps chaque famille du Val de Marne représentée

par son enfant le plus âgé a été projetée sur les 2 premières composantes principales. On observe sur la figure 8 que les individus du Val de Marne (triangles noirs) n'appartiennent pas tous à la même population.

Figure 8 : Structure de population de l'échantillon du Val de Marne

Les points oranges représentent les européens du projet 1000 Génomes (89 GBR, 93 FIN, 85CEU, 14 IBS, 98 TSI), les points bleus représentent les africains sub-sahariens (61 ASW, 97 LWK, 88YRI), les points roses représentent les sud-américains (66 MXL, 60 CLM, 55PUR) et les points jaunes les asiatiques (97 CHB, 100 CHS and 89 JPT). Les individus du Val de Marne sont en noir.




2. Echantillon de réplication

Un échantillon de 662 individus appartenant à 155 familles nucléaires originaires de Ravensmead et Uitsig en banlieue du Cap (Afrique du Sud) a été utilisé en réplication, échantillon utilisé précédemment pour localiser les régions de liaison TST1 et TST2 (155), ainsi que pour étudier l'héritabilité de l'immunité anti-mycobactérienne (151). Brièvement, l'Afrique du Sud connaît le deuxième plus haut taux d'incidence de tuberculose de tous les pays du monde (1) et en particulier, la région où ont été recrutées les familles utilisées pour l'étude affichait un taux de déclaration annuelle de tuberculose à 761 pour 100000 habitants et par an, au début des années 2000 (100). L'hypothèse d'une exposition communautaire à *M.tuberculosis* dans la région semble donc justifiée. Ces individus ont été inclus de 2001 à 2004 dans l'étude, en privilégiant les grandes familles nucléaires pour assurer une meilleure

reconstruction des liens familiaux et augmenter la puissance de l'étude. Si le résultat du test de Mantoux était connu au moment de l'inclusion (dans le cadre de l'enquête ARTI de 1998), les familles comprenant des enfants phénotypiquement discordants ont été incluses en priorité afin d'augmenter la puissance de l'analyse de liaison du test de Mantoux (158) . Trois covariables ont été recueillies pour chaque individu : l'âge en années, le sexe et les antécédents de tuberculose datant de plus de deux ans (oui/non).

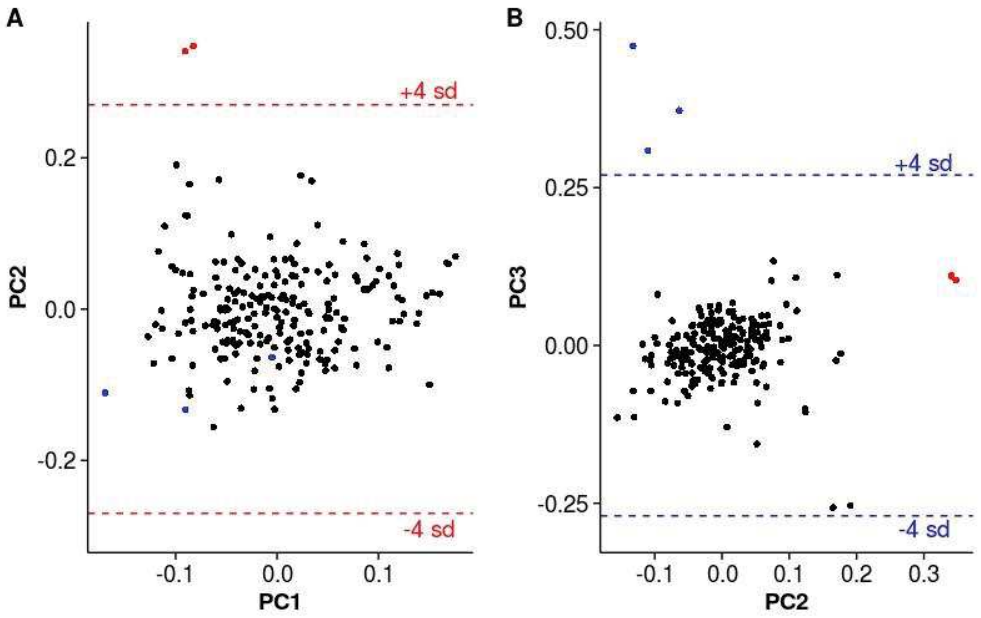
La vaccination par le BCG est systématique à la naissance depuis 1975 dans cette région du monde et la couverture vaccinale est estimée en 2005 à 99% (IC95% : 99–99.5%) dans la banlieue ouest du Cap chez des enfants âgés de 12 à 24 mois (159). Ainsi, bien que le statut vaccinal n'ait pas été formellement renseigné, il est probable que la vaste majorité des enfants inclus dans l'étude aient été vaccinés par le BCG. D'autre part, la séroprévalence du VIH est plus faible dans cette population qu'au niveau national, de l'ordre de 2% dans la population pédiatrique (160), ce qui ne peut donc constituer un biais majeur. Enfin, la population est composée d'une ethnie unique, les 'coloured', née du mélange des colons européens, des populations africaines locales et des descendants des esclaves déportés depuis l'Asie ou l'Afrique. Les analyses de structure de la population des coloured d'Afrique du Sud rapportent la contribution d'au moins quatre populations différentes (la population Khoisan, les Bantus, les Européens et des populations Asiatiques) (161,162). Ainsi, même si le fond génétique est mélangé, il est commun à toute la population. En conclusion, l'exposition différentielle, l'ethnie, la vaccination par le BCG et l'infection par le VIH n'apparaissent pas comme des facteurs de confusion dans cette étude.

Pour cet échantillon encore, un criblage complet du génome a été réalisé par le Centre national de Génotypage (CNG). Le génotypage a été effectué avec le panel Illumina linkage IVb contenant 6002 SNPs. Les SNPs de ce panel sont uniformément distribués sur le génome et espacés de 0.62 mégabases (Mb) en moyenne. Ils ont également été sélectionnés sur des critères de fréquences alléliques pour fournir le maximum d'information en analyse de liaison (plus de 95% en moyenne sur l'ensemble du génome dans une population caucasienne) (http://www.illumina.com). Au total, 5657 SNPs sont distribués sur les autosomes, dont 75% sont communs avec le panel utilisé pour l'échantillon du Val de Marne. Onze SNPs non polymorphes et 79 SNPs avec un taux de génotypage inférieur à 80% ont été exclus. L'écart à l'équilibre d'Hardy Weinberg des 5567 SNPs autosomiques restant a été testé à l'aide du logiciel PLINK (163,164) chez les fondateurs non apparentés. Aucun SNP ne montrait un écart à l'équilibre d'Hardy–Weinberg au seuil $\alpha = 0.0001$.

Comme cela avait été fait préalablement par Cobat et al dans le cadre de l'analyse de liaison du phénotype TST (155), une analyse de structure de population a été réalisée préalablement à l'analyse de liaison afin de minimiser l'hétérogénéité génétique au sein de l'échantillon. L'objectif de cette analyse était d'identifier, au sein de l'échantillon, la présence d'individus génétiquement distants de la majorité des autres. Pour ce faire, une analyse en composante principale des 5567 SNPs autosomiques sur les 220 fondateurs non apparentés et génotypés de notre échantillon a été réalisée à l'aide du logiciel EIGENSTRAT et de sa fonction SMARTPCA (157). La figure 9 montre les résultats de l'analyse en composantes principales. Les 220 fondateurs sont positionnés sur les trois principaux axes de variation. Cinq individus, appartenant à 4 familles nucléaires, ayant des valeurs extrêmes (> 4 écart–types) pour les trois premières composantes (représentés par des points rouges et bleus), ont été exclus (avec l'ensemble de leur famille) de l'analyse de liaison dans l'idée de diminuer au maximum le bruit de fond génétique pouvant diluer un signal potentiel spécifique de cette population.

Figure 9 : Analyse en composantes principales des 220 individus fondateurs de l'échantillon du Cap A) selon les deux premières composantes principales PC1 et PC2, (B) selon la seconde et la troisième composantes principales PC2 et PC3.

L'analyse a été réalisée sur 5567 snps autosomes. Chaque individu est représenté par un point, les points rouges représentent les individus ayant des valeurs extrêmes pour la seconde composante et les points bleus ceux ayant des valeurs extrêmes pour la troisième composante.



## 2. Phénotypes immunologiques

Dans l'échantillon du Val de Marne, des échantillons de sang ont été collectés pour chaque individu et des cellules mononucléaires du sang périphérique (PBMCs) ont été isolées par gradient de centrifugation à l'aide d'une plaque Ficoll-Plus (Amersham). Les PBMCs ont été lavées une fois dans du RPMI 1640 et comptées avec un compteur automatique de cellules (Beckmann). Au total, $2.10^6$ PBMC/mL ont été dispersées dans 200 µL de RPMI 1640 additionné de 10% de sérum de veau fœtal inactivé par la chaleur sur des plaques de 96 puits (Nunc), puis activées avec différents stimuli : l'antigène ESAT-6 spécifique de *M.tuberculosi*s (rdESAT6, Statens-Serum-Institute, Denmark; 2.5µg/mL), la tuberculine ou PPD (Statens-Serum-Institute, Danemark; 5µg/mL), le bacille du BCG (BCG-Pasteur; 20 BCG/leucocyte), un mitogène représenté par la phytohémagglutinine (PHA 6.25 µg/mL) comme contrôle positif ou du milieu seul comme contrôle négatif. Elles ont ensuite été cultivées à 37°C en présence de 5% de CO2. Le surnageant a été récupéré après 4 jours de stimulation, congelé et ensuite testé en IFN-γ par ELISA selon les recommandations du fabricant (Pelikin Compact; CLB). La densité optique a été déterminée à l'aide d'un lecteur cMR5000 ELISA (Thermolab Systems), et le résultat final standardisé par million de PBMCs dans l'unité pg/mL/$10^6$ de PBMCs.

Pour l'échantillon du Cap, une prise de sang a été réalisée chez les enfants uniquement. La production d'IFN–γ a été mesurée sur sang total après 3 ou 7 jours d'incubation avec du BCG, du PPD ou l'antigène ESAT–6. Pour chaque condition expérimentale, quatre mesures ont été effectuées pour chaque individu. Les cellules ont également été incubées sans antigènes mycobactériens (contrôle négatif) et en présence d'un mitogène, la phytohemagglutinine en contrôle positif. La production de d'IFN–γ a été mesurée en pg/ml par la technique ELISA. Afin d'être le plus cohérent possible entre les 2 échantillons d'étude, nous avons seulement considéré les phénotypes de production d'IFN–γ après 3 jours d'incubation.

## 3. Méthodes Statistiques

### 1. Ajustement des phénotypes sur des facteurs de confusion non génétiques

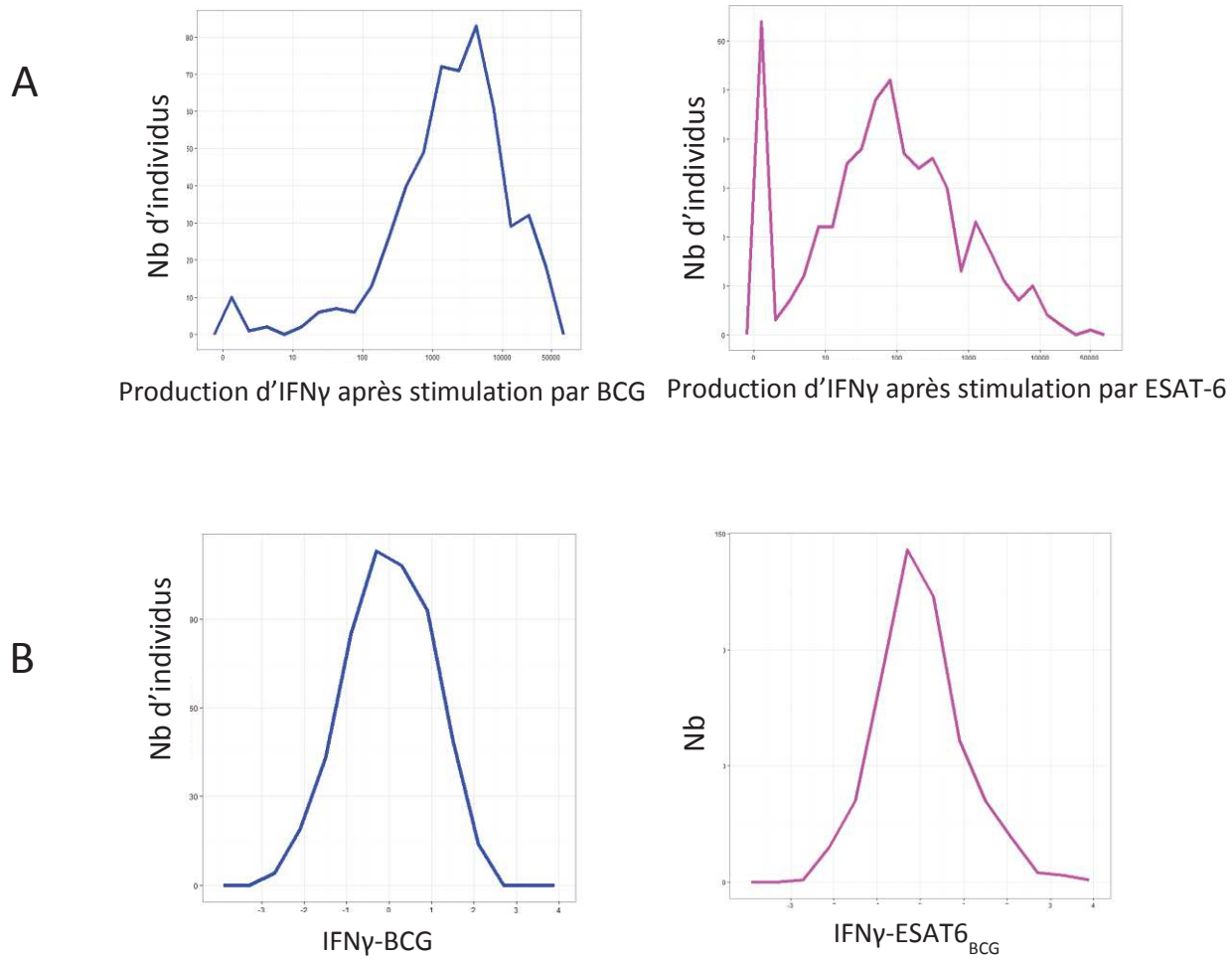
Les trois premiers phénotypes étudiés ont été la production d'IFN-γ après stimulation par le BCG, par le mélange d'antigènes PPD et par l'antigène ESAT-6. La distribution de chacun d'entre eux étant très asymétrique, nous avons choisi de leur faire subir une log-

transformation classique. Après cette transformation, la valeur contrôle non stimulée a été soustraite des valeurs stimulées. Afin de diminuer la variabilité non spécifiquement génétique de nos phénotypes d'intérêt, ils ont ensuite été ajustés par régression linéaire pour les facteurs de risque sélectionnés parmi ceux disponibles. Il est à noter que la régression linéaire classique telle qu'utilisée ici et implémentée dans le logiciel R ne tient pas compte de la dépendance des observations due à la structure familiale de nos échantillons. Bien que la corrélation des observations conduise à une mauvaise estimation de leur variance, elle ne change en théorie pas l'estimation des paramètres de régression (165). Dans le cas présent, nous n'étions pas intéressés par les tests d'hypothèses en tant que tels mais seulement par la prise en compte de l'effet moyen des covariables dans notre échantillon afin de pouvoir en tenir compte préalablement à l'analyse génétique. La corrélation des observations ne faussait donc pas les résultats attendus.

Pour l'échantillon du Val de Marne, les covariables ont en premier lieu été sélectionnées sur la base d'une association significative en analyse univariée avec au moins un des phénotypes étudiés, puis sur le meilleur modèle final multivarié en matière de critère d'Akaike (AIC). Ces phénotypes finaux ajustés seront référencés dans la suite du manuscrit comme IFNγ-BCG, IFNγ-ESAT6 et IFNγ-PPD. Nous avons également étudié un quatrième phénotype correspondant au phénotype IFNγ-ESAT6 ajusté sur le phénotype IFNγ-BCG. L'idée de ce phénotype est d'être très spécifique de l'ESAT-6 et de s'affranchir de la capacité intrinsèque de réponse, quant à la production d'IFNγ, suite à une stimulation mycobactérienne générale. Ce phénotype particulier sera noté IFNγ-ESAT6$_{BCG}$. La distribution de 2 de ces phénotypes avant et après ajustement est montrée sur la figure 10.

Les covariables pertinentes pour ces analyses incluaient le taux annuel d'incidence de tuberculose dans le pays de naissance des individus (en 2 catégories, plus ou moins de 100 nouveaux cas de tuberculose pour 100 000 habitants et par an), l'estimation de l'exposition totale de chaque individu au cas index quantifiée par le log(nombre d'heures de contact avec le cas index durant les 3 mois précédant le diagnostic de tuberculose), la contagiosité du cas index définie par la présence de cavernes sur sa radiographie des poumons ET la présence de bacilles dans ses crachats, la couverture par une assurance complémentaire de santé (OUI/NON), et l'âge. En ce qui concerne les phénotypes IFNγ-BCG et IFNγ-PPD, l'effet de la contagiosité du cas index était dépendante de l'âge du sujet exposé et un terme d'interaction entre les deux a été ajouté dans le modèle linéaire.

Figure 10 : Distribution des phénotypes IFNγ-BCG (bleu) and IFNγ-ESAT6 (magenta) avant et après ajustement. (A) Production d'IFNγ brute – en abscisses, les valeurs de l'échantillon du Val de Marne en pg/ml sur une échelle log décimale (B) Phénotypes standardisés après ajustement sur les covariables sélectionnées conduisant aux phénotypes IFNγ-BCG (bleu) et IFNγ-ESAT6$_{BCG}$(magenta).

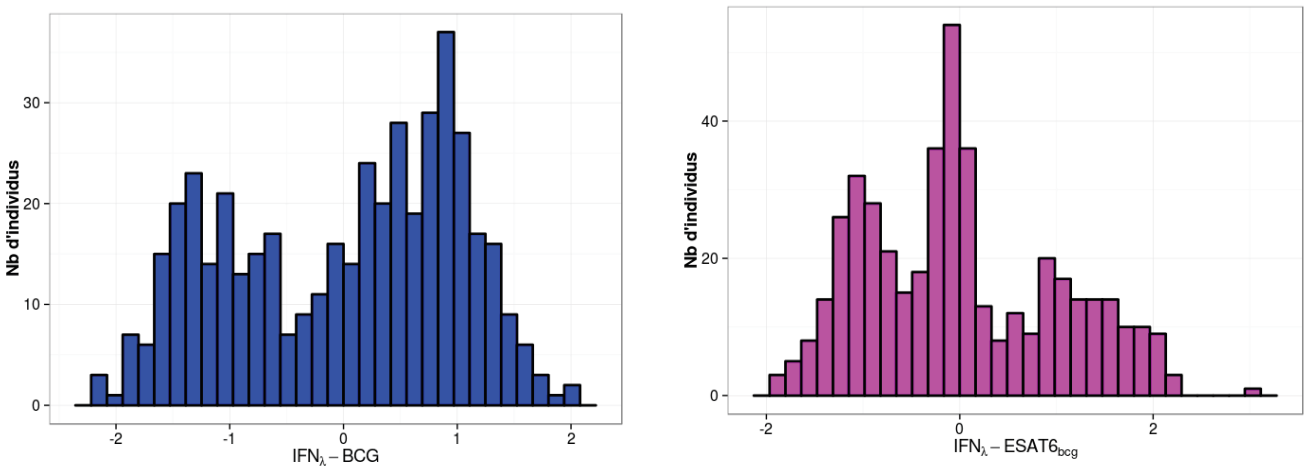



Comme la majorité des individus avaient été vaccinés par le BCG (89%), le vaccin BCG n'était associé à aucun des phénotypes étudiés. Nous avons également regardé l'effet de l'intervalle de temps entre l'administration du test de Mantoux et la prise de sang, car le TST peut provoquer une augmentation momentanée de la production d' IFN-γ (166,167). Comme cela a déjà été montré (166,168), nous avons confirmé le fait que la moyenne de nos phénotypes ne différait pas significativement entre les individus échantillonnés durant les 3 jours suivant le TST (35% des participants) et les individus échantillonnés avant l'administration du TST (32% des individus). Nous avons aussi vérifié que la moyenne des phénotypes des individus échantillonnés plus de 3 jours après le TST (33% des individus) ne différait pas de manière significative en fonction de l'intervalle de temps entre le TST et la

prise de sang. Pour ces raisons, nous avons utilisé un indicateur binaire de temps entre l'administration du test de Mantoux et la prise de sang (plus ou moins de 3 jours d'intervalle entre le TST et la prise de sang), pour finalement trouver une production d'IFN-γ plus grande chez les individus échantillonnés plus de 3 jours après l'administration du TST pour le phénotype IFNγ-BCG et dans une moindre mesure pour les phénotypes IFNγ-PPD et IFNγ-ESAT6. Ces 3 phénotypes ont donc été ajustés sur le délai d'échantillonnage en complément des autres covariables.

Pour l'échantillon d'Afrique du Sud, nous avons réalisé des régressions linéaires multivariées sur la moyenne géométrique des 4 mesures de production d'IFN-γ disponibles pour chaque individu et chaque stimulation, à laquelle nous avons soustrait le logarithme de la valeur obtenue en l'absence de toute stimulation. Les valeurs résultantes ont été ajustées sur le sexe, l'âge et l'antécédent de tuberculose active, comme cela avait été le cas dans les études précédentes réalisées sur cet échantillon (151). Le phénotype IFNγ-ESAT6 a été également ajusté sur le phénotype IFNγ-BCG comme pour l'échantillon du Val de Marne et la distribution des phénotypes est montrée sur la figure 11.

Figure 11 : Distribution des phénotypes IFNγ-BCG (bleu) and IFNγ-ESAT6$_{BCG}$ (magenta) après ajustement dans l'échantillon familial d'Afrique du sud



## 2. Analyse de liaison génétique modèle indépendante – nMLB-QTL

L'analyse de liaison génétique vise à identifier et localiser sur le génome des régions chromosomiques contenant un ou plusieurs gène(s) influant sur un phénotype d'intérêt, soit en étudiant un nombre restreint de régions candidates, soit en criblant le génome entier. Différentes méthodes existent et permettent l'analyse de phénotypes binaires ou quantitatifs. Sur le principe, l'analyse de liaison consiste à compter dans des familles le nombre d'enfants issus de gamètes ayant subi une recombinaison entre deux loci lors de la méiose et le nombre de ceux n'ayant pas eu de recombinaison entre ces deux mêmes loci, à estimer le taux de recombinaison, et à tester si ce taux diffère significativement de 0.5. En effet, en cas d'indépendance des deux loci (loci situés sur deux chromosomes différents ou à une grande distance l'un de l'autre sur le même chromosome), on s'attend à obtenir autant de gamètes recombinés que de gamètes parentaux et donc à ce que le taux de recombinaison $\theta$ soit égal à 0.5. Inversement, si les deux loci sont extrêmement proches, quasiment aucune recombinaison n'est possible et $\theta \cong 0$. En pratique, les génotypes au locus d'intérêt ne sont pas observés et le nombre d'individus porteurs d'une recombinaison ne peut pas être compté directement. Il est nécessaire de spécifier un modèle génétique (moyenne et variance du phénotype conditionnellement au génotype pour un phénotype quantitatif ou pénétrance pour un phénotype binaire et fréquence de l'allèle influençant le phénotype), habituellement défini par une analyse de ségrégation préalable, permettant d'inférer le génotype au locus d'intérêt à partir du phénotype observé, et donc d'estimer le nombre d'individus porteurs d'une recombinaison entre ce locus et un marqueur génétique pour lequel on connaît le génotype de chaque individu. Le test de liaison est un test de rapport de vraisemblance comparant la vraisemblance de l'échantillon familial sous l'hypothèse nulle d'absence de liaison ($\theta_0 = 0.5$) et la vraisemblance sous l'hypothèse alternative ($0 \leq \theta_1 < 0.5$). La statistique de test, notée $\lambda$, est donc :

$$\lambda = -2\mathrm{Ln}\ \frac{L(\theta_0=0.5)}{L(\theta_1)}$$

La statistique $\lambda$ est asymptotiquement distribuée comme un mélange 50% : 50% de $\chi^2$ à 0 et 1 degré de liberté puisque l'hypothèse alternative est $\theta < 0.5$ et non $\theta \neq 0.5$. Pour des raisons historiques, il est usuel de prendre le logarithme en base 10 du rapport de vraisemblance, et le LOD–score Z est alors défini par (169):

$$Z(\theta 1) = \log_{10} \frac{L(\theta 1)}{L(\theta 0 = 0.5)}$$

avec un LOD–score maximum $Z(\theta_{max})$ à l'estimateur du maximum de vraisemblance $\theta_{max}$.

Si le modèle génétique spécifié est correct, ces méthodes dites modèle–dépendantes sont extrêmement puissantes (170). Dans le contexte des maladies multifactorielles, il n'est souvent pas possible de définir correctement un modèle génétique. D'autres méthodes ne nécessitant pas de spécifier de modèle génétique, dites modèle–indépendantes (ou non paramétriques), ont été proposées. Leur principe commun est que, dans l'hypothèse de liaison génétique, des individus apparentés ayant des phénotypes similaires partagent plus de matériel génétique hérité d'un même ancêtre dans la région des gènes influençant ce phénotype que ne le laisserait supposer leur seul lien de parenté. Ainsi, au niveau d'un locus influençant le trait, on devrait observer une ressemblance génétique plus grande chez les apparentés de même degré ayant des phénotypes similaires et plus faible chez ceux ayant des phénotypes éloignés. Parmi les membres d'une même famille, la ressemblance génétique à un locus donné est mesurée par le nombre d'allèles hérités en commun d'un même ancêtre, ou allèles identiques par descendance [Identical By descent (IBD)]. C'est ce nombre d'allèles IBD tout au long des chromosomes que cherchent à calculer les différents algorithmes utilisés dans les méthodes d'analyse de liaison modèle-indépendantes.

L'analyse de liaison génétique des phénotypes quantitatifs de production d'IFN-γ a été réalisée à l'aide de la méthode nMLB-QTL (new Maximum Likelihood Binomial for Quantitative Trait Locus) initialement développée par L.Abel et A. Alcaïs en 1999 (Alcaïs and Abel 1999) puis améliorée par A.Cobat et al (172). La méthode MLB est une méthode d'analyse de liaison modèle–indépendante orientée fratrie, développée au départ pour des phénotypes binaires (individus atteints versus individus non atteints). L'idée générale de 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 AB (173). Pour étendre cette méthode aux traits quantitatifs, l'idée est d'introduire une variable latente binaire qui va capturer l'information de liaison entre le phénotype quantitatif et le marqueur étudié. Une fois cette variable introduite, la vraisemblance sera 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 latente binaire et le marqueur en utilisant la notion de distribution binomiale (171). La cohérence de la méthode nécessite de spécifier une

fonction de lien monotone entre les variables latentes binaires $y_i$ et les phénotypes quantitatifs observés $z_i$, i.e. une valeur plus élevée de $z_i$ devra être associée à une valeur plus élevée de $P(y_i=1/z_i)$ [remarquons que pour des raisons de symétrie, il serait équivalent d'associer une valeur élevée de $z_i$ avec une valeur élevée de $P(y_i=0/z_i)$], et une solution naturelle est d'utiliser une fonction cumulée. Une première approche consiste à utiliser une fonction paramétrique comme la fonction de répartition normale :

$$P(y_i = 1/z_i) = \Phi(z_i), \text{ et}$$

$$P(y_i = 0/z_i) = 1 - \Phi(z_i),$$

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 ou tout autre découpage en fonction des données disponibles. Si on note $d(z_i)$ le décile correspondant à la valeur $z_i$, $d(z_i) = \{1 ; 2 ; \ldots ; 10\}$, on définit :

$$P(y_i = 1/z_i) = 0.1\,[d(z_i)] - 0.05, \text{ et}$$

$$P(y_i = 0/z_i) = 1 - P(y_i = 1/z_i).$$

La méthode nMLB-QTL revisitée en 2011 (172) met en avant l'indépendance de transmission des allèles marqueurs parentaux sous l'hypothèse nulle d'absence de liaison génétique. Ainsi, pour une fratrie, la vraisemblance des observations au locus marqueur $M = (m_1, m_2, \ldots, m_n)$ sachant les phénotypes des enfants $Z = (z_1, z_2, \ldots, z_n)$, P(M/Z) est écrite

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

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 alors introduite dans $P(M_j/Z)$ en sommant sur les $2^n$ vecteurs Y possibles, n étant le nombre de phénotypes observés, puisque par définition, conditionnellement à Y, $M_j$ est indépendant de Z.

$$\begin{aligned} P(M_j / Z) &= \sum_{i=1}^{n} P(Mj\ /Yi, Z) P(Yi\ /Z) \\ &= \sum_{i=1}^{n} P(Mj\ /Yi) P(Yi\ /Z) \end{aligned}$$

Comme dans le cas binaire et dans la méthode MLB–QTL originale, on peut construire un test de liaison fondé sur un rapport de vraisemblance en testant l'écart par rapport à 0.5 de la

probabilité pour qu'un enfant avec $y_i=1$ ait reçu l'allèle marqueur transmis avec l'allèle prédisposant à une valeur élevée du phénotype. La statistique de test, $\lambda_{nMLB}$, a la même distribution que dans le cas binaire, i.e. un mélange 50% : 50% de $\chi^2$ à 0 et 1 ddl.

Dans nos échantillons du Val de Marne et d'Afrique du Sud, nous sommes en présence de fratries de différentes tailles et nous n'avons aucune idée du modèle génétique qui pourrait sous-tendre les phénotypes de production d'IFN-γ, ni des paramètres tels que la moyenne ou la variance du trait dans la population. La méthode nMLB-QTL étant robuste sur le plan de l'erreur de type I quelles que soient la distribution du phénotype étudié et la taille des fratries, et également plus puissante que la formulation originale (172), nous est apparue comme la méthode la plus adaptée à ce que nous souhaitions faire.

## 4. Résultats

### 1. Distribution des phénotypes d'intérêt et prise en compte des covariables

Sur les 590 individus contact inclus dans l'étude, 559 individus avaient des phénotypes disponibles, dont 528 sans aucune valeur manquante parmi les covariables d'intérêt. Nous nous sommes concentrés sur ces 528 individus provenant de 143 familles différentes et se composant de 268 femmes et 260 hommes âgés de moins d'un mois à 82 ans.

La distribution brute des productions d'IFN-γ suite aux stimulations des PBMCs par le BCG, le PPD et l'ESAT6 présentant une grande asymétrie, nous avons choisi de leur faire subir une transformation logarithmique et de leur soustraire la valeur d'IFN-γ trouvée sans aucune stimulation (Figure 10 paragraphe Méthodes).

Nous avons ajusté les phénotypes dans une régression linéaire multivariée sur les covariables les plus pertinentes, c'est-à-dire celles associées de manière significative avec au moins un des phénotypes étudiés et donnant le meilleur ajustement selon le critère d'information d'Akaike dans la régression finale. Cet ajustement avait pour but de ne garder que la part potentiellement génétique de la variabilité des phénotypes et d'éliminer tout autre facteur de confusion. Ces analyses ont montré des niveaux de production d'IFN-γ plus élevés lorsque l'incidence de la tuberculose dans le pays de naissance était supérieure à 100 pour 100 000 habitants et par an, lorsque l'exposition au cas index était plus longue et que celui-ci était plus contagieux (ce dernier effet étant restreint aux individus les plus jeunes pour les phénotypes IFNγ-BCG et IFNγ-PPD), et en l'absence de complémentaire santé. La production d'IFN-γ augmentait de manière significative avec l'âge pour les phénotypes IFNγ-ESAT6 et IFNγ-

PPD, mais par pour le phénotype IFNγ-BCG. Ceci tient vraisemblablement au fait que la quasi-totalité des individus était vaccinée par le BCG très jeune. Pour la suite de l'étude, les phénotypes ont donc été ajustés sur ces covariables, comme détaillé dans le tableau 2. On observe que le délai entre le test de Mantoux et la prise de sang n'est pas associé avec le phénotype IFNγ-ESAT6$_{BCG}$ comme attendu suite à l'ajustement sur IFNγ-BCG.

Tableau 2 : Significativité des covariables utilisées pour l'ajustement des phénotypes dans les analyses multivariées.

| **Phénotypes** | **IFNγ-BCG** | **IFNγ-PPD** | **IFNγ-ESAT6** | **IFNγ-ESAT6$_{BCG}$** |
|---|---|---|---|---|
| Incidence de la tuberculose dans le pays de naissance | 0.001 | 0.001 | 0.003 | - |
| Log(nombre d'heures de contact) | 0.09 | 0.05 | 0.0014 | 0.008 |
| Contagiosité du cas index | 0.015 [1] | 0.09 [1] | 0.089 | 0.0016 |
| Complémentaire santé | 0.098 | 0.02 | 0.04 | 0.12 |
| Age | 0.23 | 0.004 | 0.014 | 0.0001 |
| Echantillonnage > 3 jours après TST | 0.0005 | 0.035 | 0.06 | - |
| BCG | - | - | - | $<2.\ 10^{-16}$ |

[1] les valeurs de p incluent l'interaction avec l'âge.

Après ajustement, nous avons calculé la corrélation de ces phénotypes entre eux, ainsi que leur corrélation avec le résultat du test de Mantoux. Nous avons trouvé une forte corrélation (r=0.78) entre IFNγ-BCG et IFNγ-PPD, et une corrélation plus faible mais non nulle (r = 0.53) entre IFNγ-BCG et IFNγ-ESAT6 (cf Tableau 3). Cela laisse donc à penser que la capacité de réponse à l'antigène ESAT-6 n'est pas complètement indépendante de la capacité de réponse à BCG, bien que le bacille du BCG ne possède pas cet antigène. Cette corrélation pourrait signifier l'existence d'une capacité individuelle intrinsèque dans la production d'IFN-γ induite par la voie de signalisation des récepteurs T et indépendante de l'antigène présenté. Sous cette hypothèse, la composante très spécifique à l'antigène ESAT-6, marqueur de *M. tuberculosis*, indépendante de cette capacité intrinsèque de réponse, serait très intéressante à étudier. Afin de nous affranchir de la composante générale, nous avons donc ajusté le phénotype IFNγ-ESAT6 sur IFNγ-BCG (noté IFNγ-ESAT6$_{BCG}$), et ce nouveau phénotype apparaît comme indépendant de IFNγ-BCG (r=0.02) comme nous le souhaitions.

Tableau 3 : Coefficients de corrélation de Spearman des 4 phénotypes de production d'Interféron-γ utilisés dans les analyses de liaison génome-entier, et du test de Mantoux, dans l'échantillon du Val de Marne.

| Phénotype | IFNγ-BCG | IFNγ-PPD | IFNγ-ESAT 6 | IFNγ-ESAT6$_{BCG}$ | TST-bin[a] |
|---|---|---|---|---|---|
| **IFNγ-BCG** | 1 | 0.78 | 0.53 | 0.02 | 0.09 |
| **IFNγ-PPD** | 0.78 | 1 | 0.61 | 0.25 | 0.2 |
| **IFNγ-ESAT 6** | 0.53 | 0.61 | 1 | 0.86 | 0.07 |
| **IFNγ-ESAT6$_{BCG}$** | 0.02 | 0.25 | 0.86 | 1 | 0.02 |
| **TST-bin** | 0.09 | 0.2 | 0.07 | 0.02 | 1 |

a : réponse binaire au test de Mantoux ajustée comme dans (139)

## 2. Résultats de l'analyse de liaison

Les analyses de liaison génome entier ont été réalisées sur 97 familles informatives de l'échantillon du Val de Marne, chacune possédant de 2 à 6 enfants phénotypés, ce qui faisait un total de 240 enfants Ces analyses ont toutes été réalisées avec la méthode nMLB-QTL (172), en utilisant une distribution du trait en déciles. En se basant sur les 2 premières composantes principales de l'analyse en composantes principales détaillée dans le paragraphe B-1 décrivant les échantillons d'étude, les 97 familles étudiées pouvaient être scindées en 49 d'origine caucasienne, 36 provenant d'Afrique sub-saharienne et 12 ayant d'autres origines comprenant l'Asie (cf Figure 7). L'information multipoint (IC) était élevée sur tous les autosomes avec une valeur moyenne de 87.6% (de 69% à 94%) à l'échelle du génome. Les résultats des analyses de liaison sont présentés phénotype par phénotype.

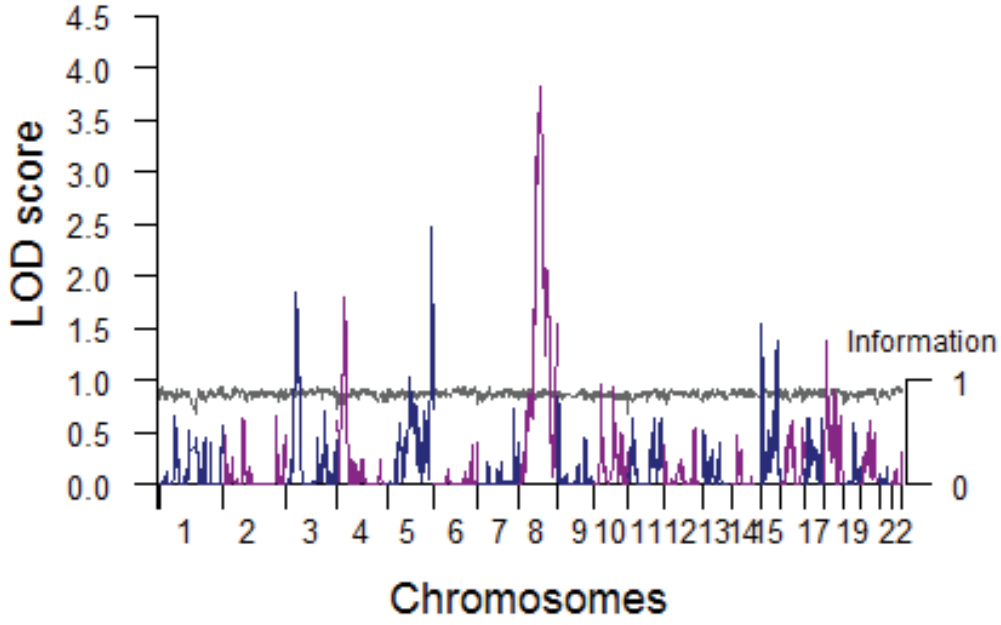
### a) Etude du phénotype IFNγ-BCG

Les résultats de l'analyse de liaison génétique du phénotype IFNγ-BCG (i.e. réponse ajustée sur les covariables) réalisée avec la méthode nMLB–QTL sont présentés sur la figure 12. Un signal de liaison significatif a été observé dans la région chromosomique 8q21.13 à la position 82.7Mb avec un LOD-score de 3.80 ($p=1.4.10^{-5}$) et une information à IC=86%. Ce signal

dépasse le seuil de significativité communément admis pour les analyses de liaison pangénomiques, à savoir un LOD–score ≥ 3.6 ($p \leq 2.10^{-5}$) (174).

Figure 12 : Analyse de liaison modèle indépendante du phénotype IFNγ-BCG pour l'échantillon du Val de Marne.

Résultats de l'analyse à l'échelle du génome pour l'échantillon du Val de Marne, présentant les LOD scores multipoints sur l'axe gauche des ordonnées et le niveau d'information (ligne horizontale grise, axe droit des ordonnées) sur les 22 autosomes en axe des abscisses.




Nous avons également trouvé un signal évocateur de liaison (LOD–score ≥ 2.2, $p \leq 7.10^{-4}$ selon (174)) dans la région chromosomique 5q35, avec un LOD score à 2.48 ($p=3.6.10^{-4}$) et une information à 86.4%, ainsi que 7 pics de liaison plus faibles avec des LOD scores >1.17 (soit $p < 0.01$). Tous ces signaux sont détaillés dans le tableau 4.

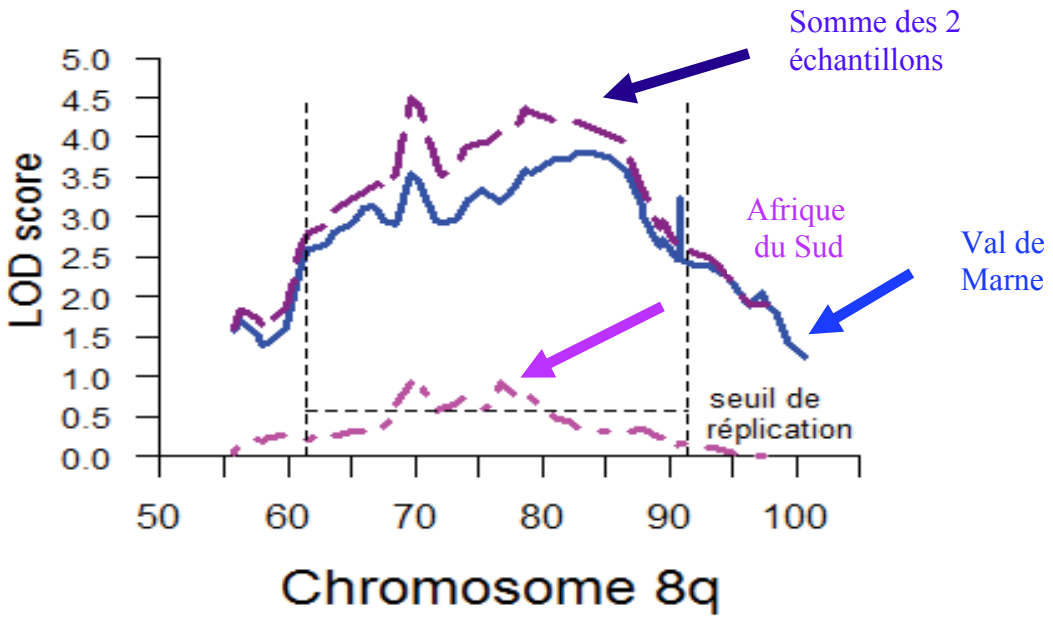
Nous avons ensuite réalisé une étude de réplication pour le phénotype IFNγ-BCG sur les 2 régions identifiées sur les chromosomes 8 et 5, dans l'échantillon d'Afrique du Sud. Le signal évocateur en 5q35 ne l'était plus du tout dans l'échantillon sud-africain (LOD-score maximal = 0.14), mais nous avons répliqué le signal de liaison de la région 8q21 avec un LOD-score de 0.98 (p =0.016) à la position de 70Mb. Ceci est illustré sur la figure 12. Aucun autre signal significatif de liaison n'a été trouvé dans la cohorte sud-africaine pour le phénotype IFNγ-BCG ; le plus haut LOD-score atteignait 2.06 à la position de 83Mb du chromosome 16. En sommant les LOD-scores des 2 échantillons, on a pu obtenir un LOD-score à 4.50 à la

position 69.7Mb (cf Figure 13), ce qui renforce l'hypothèse de liaison du locus avec le phénotype IFNγ-BCG.

Tableau 4 : Régions chromosomiques significatives au seuil de 1% pour le phénotype IFNγ-BCG, dans l'échantillon du Val de Marne.

| Région chromosomique | Intervalle (Mb) | LOD score maximum | p |
|---|---|---|---|
| **3p21** | 40-51 | 1.84 | 1.8 10-3 |
| **4p15.2** | 24-33 | 1.8 | 1.0 10-3 |
| **5q35** | 169-178 | 2.48 | 3.6 10-4 |
| **8q11.2-8q22** | 56-101 | 3.81 | 1.4 10-5 |
| **8q22-8q24** | 101-121 | 2.05 | 1.0 10-3 |
| **8q24.3** | 144-146 | 1.53 | 4.0 10-3 |
| **15q11.2-q12** | 26-30 | 1.54 | 3.8 10-3 |
| **15q25** | 82-90 | 1.39 | 5.7 10-3 |
| **18q11.2** | 8-11 | 1.37 | 6.0 10-3 |

Figure 13 : Zoom sur la région de liaison du chromosome 8



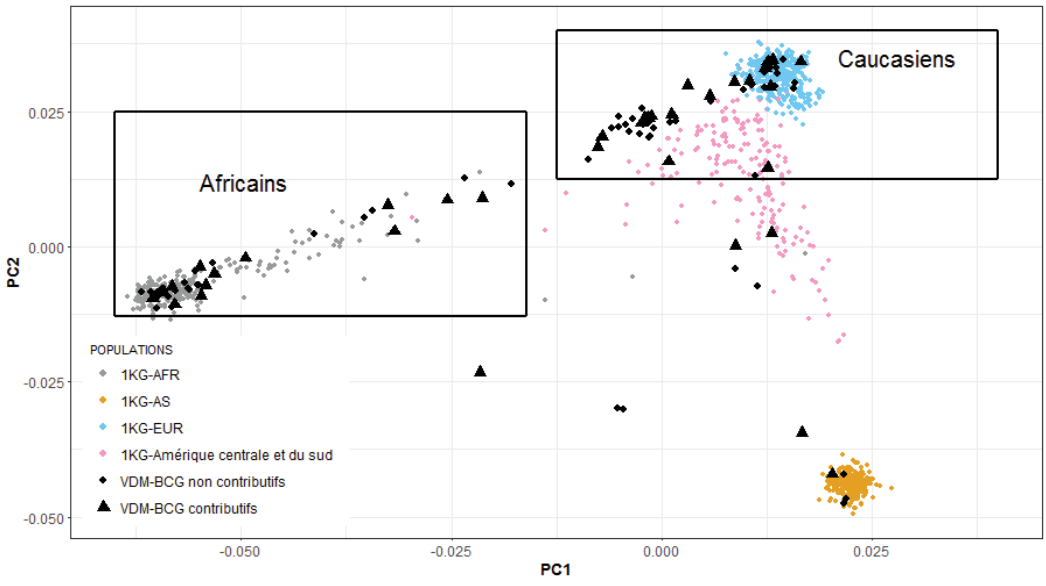
En tenant compte de la grande variabilité dans l'estimation de la position réelle d'un signal de liaison décrite par Roberts et al (175) et en n'oubliant pas que les phénotypes des 2 échantillons comportaient des petites différences, il a semblé prudent de considérer un intervalle de confiance plutôt large pour le positionnement du locus liés au phénotype IFNγ-BCG. En regardant la forme de la courbe de la somme des LOD-scores des 2 cohortes (Figure 12), nous avons défini un intervalle de confiance de 30Mb, de 61 à 91.5Mb. Cette région contient 117 gènes ou microARNs connus qui sont listés en annexe 1.

Il est intéressant de noter que ce locus se situe juste à la limite de la région 8q12-13 (55.1Mb-61.2Mb) qui a été rapportée comme étant liée à la tuberculose pulmonaire au Maroc (176) et incluant le gène TOX (59.7-60Mb) dont certains variants sont associés avec la tuberculose pulmonaire de survenue précoce (177).

Parmi les 97 familles de l'échanttillon du Val de Marne utilisées dans ces analyses, 39% des familles caucasiennes ont réellement contribué au signal de liaison observé pour le phénotype IFNγ-BCG, tout comme 36% des familles originaires d'Afrique sub-saharienne et 41% des familles ayant d'autres origines. La contribution réelle des familles au signal de liaison est définie ici comme l'apport d'un LOD-score > 0.1 au niveau de la famille prise isolément dans la région de liaison. Cette observation montre que le signal de liaison trouvé n'est pas propre à une seule population, mais qu'il est supporté, dans des proportions équivalentes, par des familles d'ethnies différentes de notre échantillon (Figure 14).

Figure 14 : Structure de population de l'échantillon du Val de Marne et IFNγ-BCG

Les triangles noirs représentent les familles contributives au signal de liaison pour le phénotype IFNγ-BCG (LOD score maximal > 0.1 entre 61 et 91.5Mb sur le chromosome 8), tandis que les points noirs représentent les autres familles du Val de Marne. Les individus de 1000 génomes sont représentés en couleur (bleu pour les caucasiens, gris pour le safricains, rose pour les américains du sud et orange pour les asiatiques.



b) Etude du phénotype IFNγ-PPD

Les résultats de l'analyse de liaison génétique du phénotype IFNγ-PPD (i.e. réponse ajustée sur les covariables) réalisée avec la méthode nMLB–QTL sont présentés sur la figure 13. Le signal de liaison le plus significatif à l'échelle du génome a été observé dans la même région 8q21.13 que pour le phénotype IFNγ-BCG, à la position 79Mb, avec un LOD-score à 3.02 et une information à 89.9% (cf figure 15). Ce résultat est tout à fait cohérent avec la forte corrélation des 2 phénotypes décrite dans le tableau 3.

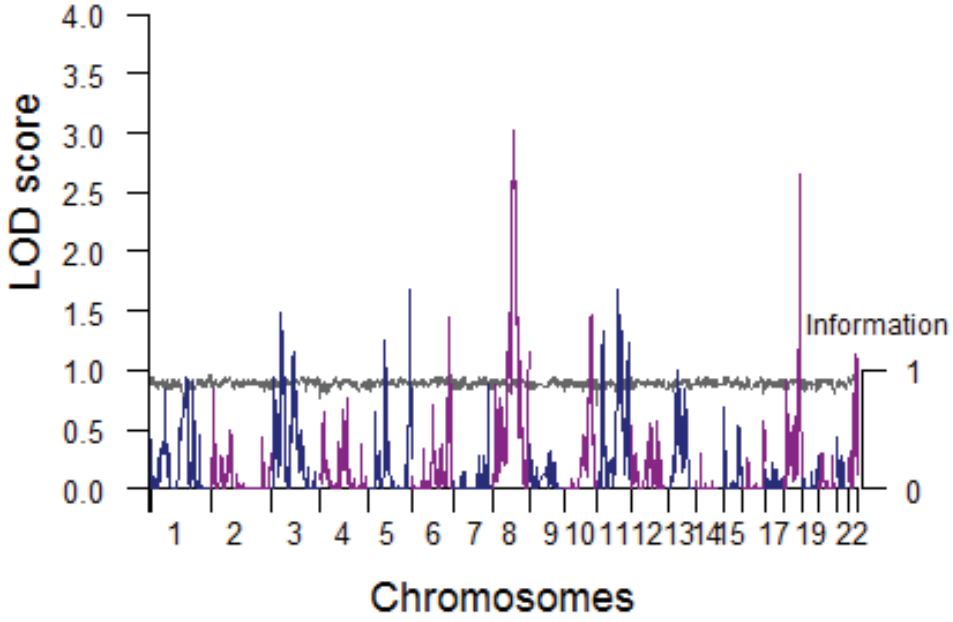
Un second signal évocateur de liaison (LOD score = 2.65) a été observé sur le chromosome 18 à 65Mb, ainsi que 9 pics de liaison plus faibles avec des LOD scores >1.17 (soit $p < 0.01$). Tous ces signaux sont détaillés dans le tableau 5. On peut noter que l'un de ces signaux se trouve dans la région 11p14-15 identifiée préalablement dans cet échantillon comme liée à la réponse binaire au test de Mantoux (139).

Tableau 5 : Régions chromosomiques significatives au seuil de 1% pour le phénotype IFNγ-PPD, dans l'échantillon du Val de Marne.

| Région chromosomique | Intervalle (Mb) | LOD score maximum | p |
|---|---|---|---|
| **3p21-22** | 41-47 | 1.49 | $4\ 10^{-3}$ |
| **5q13** | 74-75 | 1.24 | $8\ 10^{-3}$ |
| **5q35** | 172-175 | 1.68 | $3\ 10^{-3}$ |
| **6q25** | 148-150 | 1.45 | $5.4\ 10^{-3}$ |
| **8q12** | 61-62 | 1.47 | $5\ 10^{-3}$ |
| **8q21-22** | 71-97 | 3.02 | $5.09\ 10^{-5}$ |
| **10q25** | 108-113 | 1.46 | $5.2\ 10^{-3}$ |
| **11p14-15** | 21-23 | 1.33 | $7.5\ 10^{-3}$ |
| **11q14-21** | 81-97 | 1.68 | $3\ 10^{-3}$ |
| **11q24** | 124-125 | 1.22 | $1\ 10^{-2}$ |
| **18q21-22** | 61-69 | 2.65 | $2.5\ 10^{-4}$ |

Figure 15 : Résultats de l'analyse de liaison modèle indépendante du phénotype IFNγ-PPD à l'échelle du génome pour l'échantillon du Val de Marne.

Résultats de l'analyse à l'échelle du génome pour l'échantillon du Val de Marne, présentant les LOD scores multipoints sur l'axe gauche des ordonnées et le niveau d'information (ligne horizontale grise, axe droit des ordonnées) sur les 22 autosomes en axe des abscisses




c) Etude des phénotypes IFNγ-ESAT6 et IFNγ-ESAT6$_{BCG}$

L'analyse liaison du phénotype IFNγ-ESAT6 n'a pu identifier de signal de liaison significatif à l'échelle du génome, le LOD-score le plus élevé s'élevant à 2.19 ($p = 7.4.10^{-4}$) à la position 122 Mb sur le chromosome 3 (cf Figure 16).

Pour nous affranchir de la corrélation existant entre le phénotype IFNγ-BCG et IFNγ-ESAT6, et pouvoir nous intéresser à la part de la production d'IFN-γ propre à la réponse à l'antigène ESAT6, nous avons ajusté le phénotype IFNγ-ESAT6 sur celui noté IFNγ-BCG. En faisant cela, le signal de liaison en 3q13-22 est devenu significatif, avec un LOD-score à 3.72 ($p = 1.8.10^{-5}$) à la position 122.3Mb, avec une information à 90.9%. Aucun autre pic de liaison évocateur n'a été trouvé avec ce phénotype noté IFNγ-ESAT6$_{BCG,}$ mais 4 signaux plus faibles atteignant le seuil de p<1% ont été observés. Le résultat de ce phénotype ajusté est illustré par la figure 17 et les signaux trouvés sont listés dans le tableau 6.

Figure 16 : Résultats de l'analyse de liaison modèle indépendante du phénotype IFNγ-ESAT6 à l'échelle du génome pour l'échantillon du Val de Marne.

Ce résultat, exprimé en LOD score multipoint (lignes noires et bleues selon l'axe des ordonnées gauche), et son information (ligne rouge selon l'axe des ordonnées droit) sont représentés le long des 22 autosomes.

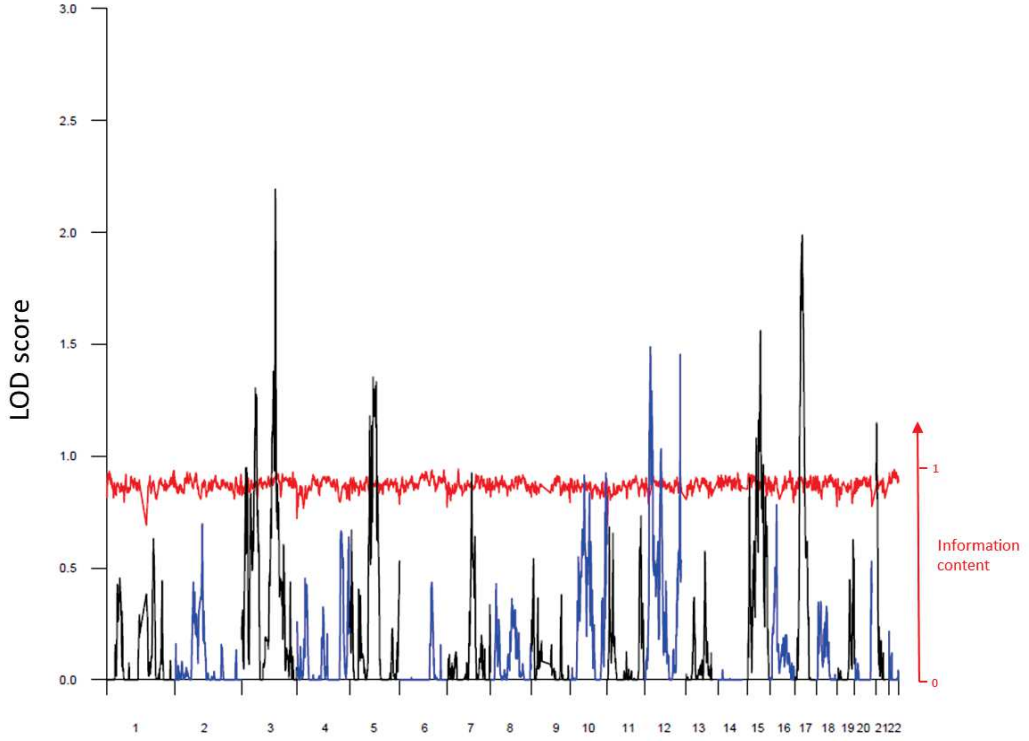



Figure 17 : Analyse de liaison modèle indépendante du phénotype IFNγ-ESAT6$_{BCG}$ à l'échelle du génome pour l'échantillon du Val de Marne

Résultats de l'analyse à l'échelle du génome pour l'échantillon du Val de Marne, présentant les LOD scores multipoints sur l'axe gauche des ordonnées et le niveau d'information (ligne horizontale grise, axe droit des ordonnées) sur les 22 autosomes en axe des abscisses.

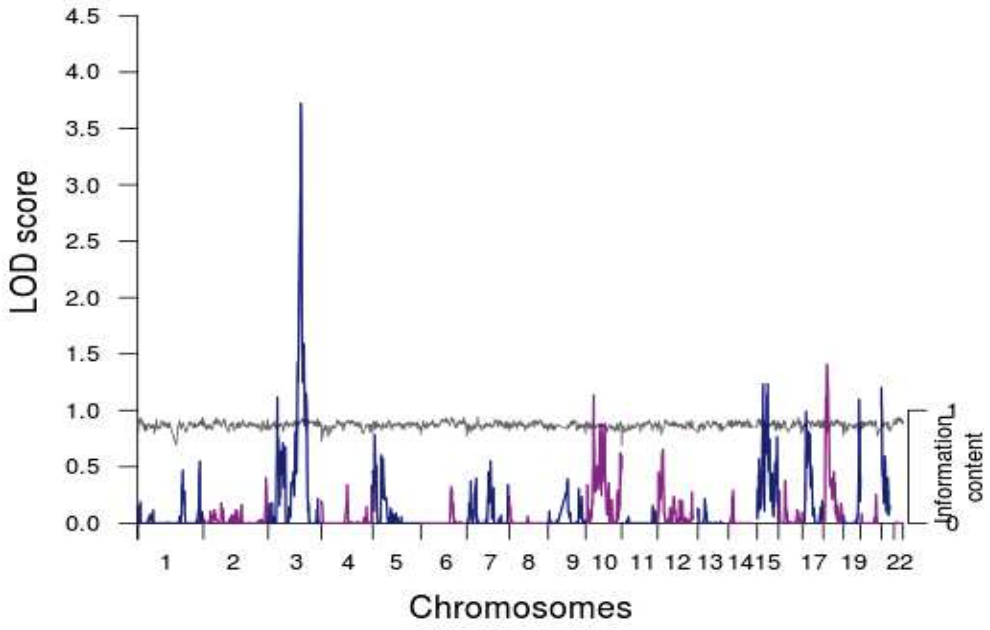
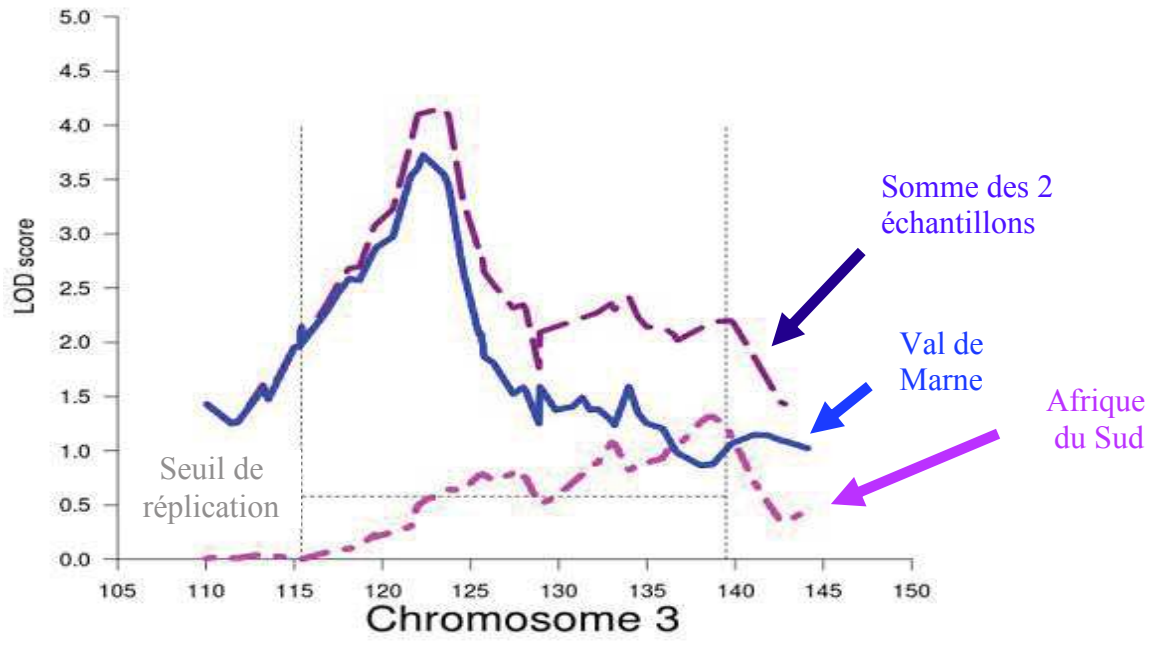


Tableau 6 : Régions chromosomiques significatives au seuil de 1% pour le phénotype IFNγ-ESAT6$_{BCG}$, dans l'échantillon du Val de Marne.

| Région chromosomique | Intervalle (Mb) | LOD score maximum | p |
|---|---|---|---|
| **3q13-3q22** | 109-135 | 3.72 | $1.8\ 10^{-5}$ |
| **15q15** | 42-43 | 1.23 | $9.7\ 10^{-3}$ |
| **15q22** | 59-60 | 1.23 | $9.7\ 10^{-3}$ |
| **18p11** | 13-18 | 1.4 | $6.0\ 10^{-3}$ |
| **21q21.1** | 15-16 | 1.19 | $1.0\ 10^{-2}$ |

Nous avons répliqué le signal majeur obtenu sur le chromosome 3q dans l'échantillon d'Afrique du Sud avec des LOD-scores de 0.78 (p = 0.028) à 125.7Mb et 1.31 (p = 0.007) à 138.7Mb. Aucun autre signal significatif n'a été trouvé pour le phénotype IFNγ-ESAT6$_{BCG}$ dans l'échantillon d'Afrique du Sud ; le meilleur LOD-score atteignait 1.8 à la position 4.5Mb du chromosome 19. En sommant les résultats des 2 échantillons dans la région de liaison, le LOD-score maximum s'élevait à 4.16 à 123.5Mb, comme on peut le voir sur la Figure 18.

Figure 18 : Zoom sur la région de liaison du chromosome 3

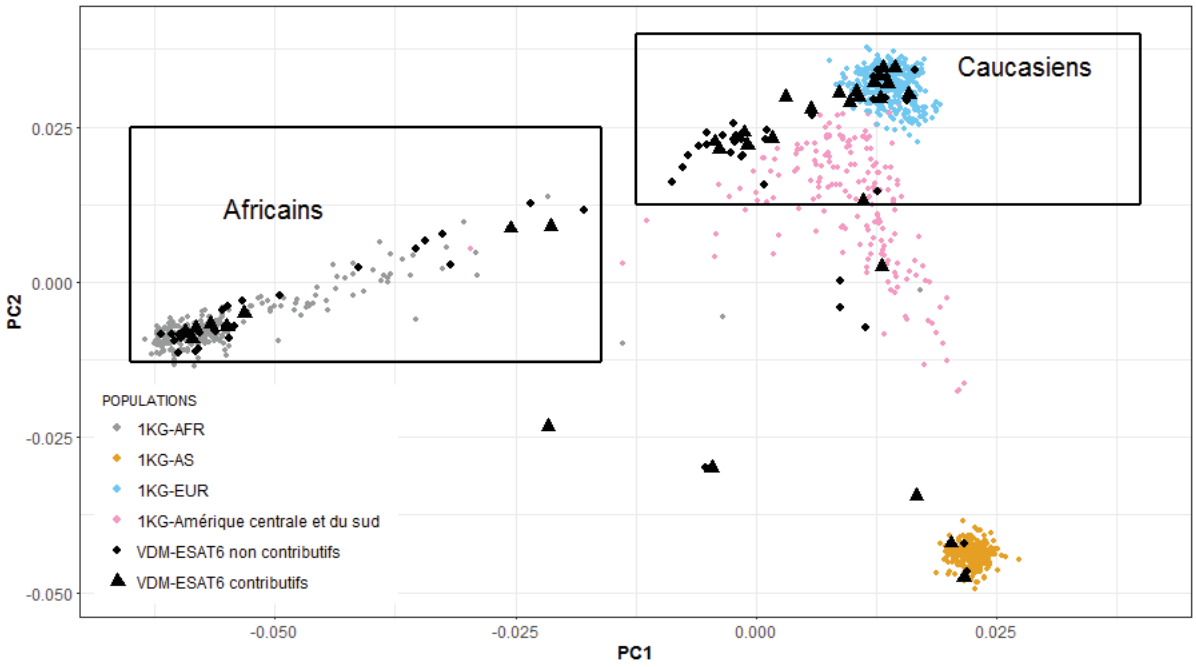


Comme précédemment, il était préférable de considérer une région de liaison assez large allant de 115 à 139 Mb sur le chromosome 3, et contenant 219 gènes ou microARNs connus qui sont listés en annexe 1. Ce résultat identifie donc un locus majeur dans la région chromosomique 3q13-22 qui contrôle la quantité d'IFN-γ produite en réponse à l'ESAT-6, l'un des antigènes spécifiques de *M.tuberculosis*, lorsque l'on tient compte de la production d'IFN-γ générée lors de la stimulation par le BCG.

Comme pour le phénotype IFNγ-BCG, nous avons regardé l'apport des différentes ethnies de l'échnatillon du Val de Marne au signal de liaison ; 39% des familles d'origine caucasienne, 25% des familles d'Afrique sub-saharienne et 41% des familles ayant d'autres origines contribuent à la liaison (Figure 19). Cela signifie, comme pour le phénotype IFNγ-BCG, que le signal de liaison ne semble pas restreint à une population humaine particulière, mais serait au contraire partagé par plusieurs ethnies.

Figure 19 : Structure de population de l'échantillon du Val de Marne et IFNγ-ESAT6$_{BCG}$

Les triangles noirs représentent les familles contributives au signal de liaison pour le phénotype IFNγ-ESAT6$_{BCG}$ (LOD score maximal > 0.1 entre 115 et 139Mb sur le chromosome 3), tandis que les points noirs représentent les autres familles du Val de Marne. Les individus de 1000 génomes sont représentés en couleur (bleu pour les caucasiens, gris pour le safricains, rose pour les américains du sud et orange pour les asiatiques.



Nous avons également réalisé une analyse de liaison du phénotype IFNγ-ESAT6 ajusté sur IFNγ-PPD. Ces 2 phénotypes étaient davantage corrélés que ne l'étaient les phénotypes IFNγ-ESAT6 et IFNγ-BCG (r=0.61 versus r=0.53), puisque PPD contient l'antigène ESAT6 contrairement au bacille du BCG. Avec ce nouvel ajustement, nous avons observé que le pic de liaison du chromosome 3q perdait son degré de significativité, avec un LOD-score à 2.37 qui n'était plus qu'évocateur de liaison. Cette observation va dans le sens de notre hypothèse, à savoir que le locus de liaison trouvé sur le chromosome 3 jouerait un rôle dans le contrôle de la production d'IFN-γ induite très spécifiquement par l'antigène ESAT6.

## 5. Conclusion et discussion

Grâce à des analyses de liaison pangénomiques des différents phénotypes d'IGRA, nous avons pu localiser 2 signaux de liaison significatifs à l'échelle du génome correspondant à 2 phénotypes d'immunité antibactérienne. Le premier, dans la région chromosomique 8q12-22, devrait correspondre à un ou plusieurs loci influençant la production d'IFN-γ générée par le bacille du BCG. Le second pic, en 3q13-22, éclaire le positionnement d'un ou plusieurs gènes pouvant avoir une influence sur la quantité d'IFN-γ libérée suite à une stimulation par l'antigène ESAT-6, après avoir ajusté sur la capacité intrinsèque de réponse à une stimulation mycobactérienne. Nous avons été en mesure de répliquer ces signaux dans un échantillon d'Afrique du Sud, avec des phénotypes proches bien que non totalement identiques (la production d'IFN-γ étant mesurée sur sang total et non uniquement sur les PBMCs, le temps d'incubation avec les stimuli étant de 3 jours au lieu de 4 jours pour le Val de Marne).

Les 2 populations étaient également très différentes en ce qui concerne l'exposition à *M.tuberculosis* ; les individus de l'étude d'Afrique du Sud vivaient dans une zone d'hyperendémie tuberculeuse au sein de laquelle l'exposition et la transmission de la mycobactérie ont lieu préférentiellement au niveau communautaire (100), tandis que l'étude française ciblait des contacts familiaux de cas de tuberculose, au sein d'une région où l'incidence de la maladie est relativement faible. Les deux cohortes différaient également en matière de fonds génétique. Les familles de l'échantillon du Val de Marne appartenaient à différents groupes ethniques que nous avons catégorisés en 3 sous-populations principales (les individus d'origine caucasienne, ceux venant d'Afrique sub-saharienne, et ceux ayant d'autres origines telles que l'Asie), et nous avons trouvé qu'au sein de chaque sous-population, une proportion similaire de familles contribuait aux 2 pics de liaison (Figures 13 et 18). A

l'inverse, tous les individus de l'échantillon de réplication provenaient du même groupe ethnique sud-africain nommé les « coloured », résultat du métissage de plusieurs ethnies à savoir les Khoisans (31%), les Bantus (33%), les Européens (16%) et les Asiatiques (20%) (178). Ainsi, la réplication des signaux de liaison trouvés pour les 2 loci dans des conditions aussi différentes est en faveur de leur authenticité dans le contrôle de la production d'IFN-γ généré par des mycobactéries chez l'homme.

Le phénotype IFNγ-BCG correspond à une réponse antimycobactérienne que l'on pourrait qualifier de généraliste. En effet, ce phénotype était fortement corrélé à la réponse induite par le mélange d'antigènes PPD (r=0.78), et le phénotype IFNγ-PPD était également lié au locus 8q avec un LOD score à 3.03. Cela laisse à penser que cette région du génome contrôle une composante non spécifique de la libération d'IFN-γ au cours d'une infection mycobactérienne. Le second locus sur le chromosome 3 est lié au phénotype IFNγ-ESAT6, lorsqu'on tient compte de la production d'IFN-γ induite par le BCG. En effet, bien que le bacille du BCG ne possède pas l'antigène ESAT-6, la corrélation des 2 phénotypes (r = 0.53) pourrait refléter une capacité générale et intrinsèque de production d'IFN-γ via la voie de signalisation du récepteur T des lymphocytes du même nom. En ajustant le phénotype IFNγ-ESAT6 sur IFNγ-BCG, on s'affranchit de cette capacité intrinsèque de réponse en IFN-γ pour isoler de manière très spécifique la capacité de réponse face à l'antigène ESAT-6, qui joue un rôle important dans l'infection par *M.tuberculosis*. Il a été montré chez la souris que l'injection de lymphocytes T CD4+ exprimant un récepteur spécifique à ESAT-6 conduisait à une résistance plus grande face à l'infection par voie aérienne de *M.tuberculosis* (179), montrant l'existence d'une réponse immunitaire spécifique face à l'infection par *M.tuberculosis* médiée par la réponse à l'antigène ESAT-6. L'utilisation du phénotype ajusté IFNγ-ESAT6$_{BCG}$ comme on vient de le décrire a fait émerger le pic de liaison du chromosome 3q, conduisant à la localisation d'un locus majeur. Nous avons également observé que ce pic de liaison diminuait substantiellement si on ajustait le phénotype IFNγ-ESAT6 sur la production d'IFN-γ après stimulation par le mélange PPD contenant notamment l'antigène ESAT-6. Tout cela contribue à la conclusion que le locus du chromosome 3q joue un rôle dans la production d'IFN-γ spécifiquement induite par l'antigène ESAT-6.

Comme on pouvait s'y attendre, les 2 loci trouvés sont différents des loci TST1 et TST2 contrôlant la réponse au test de Mantoux de manière binaire et quantitative (155). Ce constat est cohérent avec la faible corrélation observée entre les phénotypes IGRA et la réponse au

test de Mantoux (tableau 3), et corrobore l'hypothèse que le test de Mantoux et les IGRAs sont des marqueurs différents et complémentaires de l'immunité antimycobactérienne (90).

# C. Fine-mapping des régions de liaison (180)

Après avoir identifié 2 régions du génome influençant les résultats des IGRA après exposition à *M.tuberculosis* grâce à des analyses de liaison, nous avons souhaité aller plus loin dans l'identification de ces facteurs génétiques. A cette fin, nous avons réalisé des études d'association au sein des régions de liaison des phénotypes de production d'IFN-γ en réponse à des stimulations mycobactériennes dans les mêmes échantillons qu'au chapitre 1, c'est-à-dire, dans des familles issues de l'entourage immédiat de patients tuberculeux recrutés près de Paris, à l'hôpital de Créteil (Val-de-Marne), puis en seconde intention, dans des familles vivant en banlieue du Cap, en Afrique du Sud, où la tuberculose est hyper-endémique.

## 1. Echantillons d'étude et marqueurs génétiques

L'échantillon primaire utilisé dans cette étude est le même que celui décrit au chapitre B. Il s'agit d'individus issus d'une étude prospective réalisée dans le Val de Marne en banlieue parisienne entre 2004 et 2009, dans l'entourage familial immédiat de patients tuberculeux. Par définition, les individus qualifiés de « contacts » sont tous les individus partageant le foyer d'un cas de tuberculose avéré durant les trois mois précédant le diagnostic de la maladie. Au total, parmi tous les individus recrutés pour lesquels il nous restait de l'ADN, 576 parents et enfants ont été sélectionnés pour être génotypés avec la puce Illumina HumanOmniExpressExome sur la plateforme de génotypage de l'hôpital de la Pitié Salpétrière à Paris. Cette puce comporte 730 525 SNPs répartis uniformément sur tout le génome et capturant l'information de plus de 70% des variants de fréquence allélique supérieure à 5% dans les populations caucasiennes et asiatiques du projet 1000 génomes, ainsi que 240 000 variants exoniques choisis par le fabricant pour être spécifiques d'une population particulière, ou bien jouant un rôle dans des pathologies communes comme le diabète de type 2, des pathologies psychiatriques et métaboliques, ou certains cancers.

Ces 576 individus ont été choisis principalement sur des critères d'information familiale afin d'optimiser la puissance des études d'association envisagées. Les individus ayant un taux de génotypage inférieur à 90% et les duplicats repérés grâce au calcul d'IBS (« identity by state ») réalisé avec le logiciel PLINK1.9 (163,164) ont été supprimés de l'analyse. De manière similaire, les SNPs avec un taux de génotypage inférieur à 99% n'ont pas été analysés. Nous avons choisi de ne pas tester l'écart à l'équilibre d'Hardy Weinberg car notre échantillon était très hétérogène du point de vue des origines ethniques. Au total 743 735 SNPs autosomiques de haute qualité et 489 individus phénotypés issus de 232 familles nucléaires ont été retenus pour les analyses.

Pour la réplication, l'échantillon d'Afrique du Sud a été utilisé comme pour l'analyse de liaison du chapitre B. Il s'agit d'un échantillon familial recruté à Tygerberg, en banlieue du Cap, où l'incidence de la tuberculose est parmi les plus élevées au monde. Seuls les enfants ont été génotypés avec la puce Illumina HumanOmni2.5 comprenant plus de 75% des variants autosomiques de la puce Illumina HumanOmniExpressExome. Après un contrôle qualité semblable à celui de l'échantillon primaire du Val de Marne, nous avons retenu pour les analyses d'association 2 241 954 SNPs autosomiques génotypés chez 373 individus phénotypés appartenant à 157 familles nucléaires. A partir des SNPs génotypés, nous avons imputé des variants supplémentaires dans les 2 régions de liaison des chromosomes 3 et 8, et sur les 2 échantillons d'étude, en utilisant le panel de référence du projet 1000 génomes phase 1, afin d'augmenter la densité de marqueurs de la région. Les 2 régions d'intérêt étaient positionnées de 115 à 139 Mb sur le chromosome 3, et de 61 à 91.5Mb sur le chromosome 8, comme définies au chapitre 1. Les SNPs imputés de bonne qualité (cf paragraphe sur les méthodes statistiques) significativement associés avec l'un ou l'autre des phénotypes étudiés ont ensuite été génotypés grâce à la plateforme SEQUENOM iPLEX MassARRAY pour l'échantillon d'Afrique du Sud ou par la technologie TaqMan SNP genotyping assays d'Applied Biosystems pour les individus du Val de Marne.

## 2. Phénotypes immunologiques

Nous avons utilisé les mêmes phénotypes de production d'IFN-γ et les mêmes variables d'ajustement que pour les analyses de liaison décrites au chapitre B, et nous sommes concentrés sur les 2 phénotypes pour lesquels nous avions obtenu une liaison génétique significative. Le premier phénotype correspondait à la production d'IFN-γ suite à stimulation

par le BCG après une transformation logarithmique classique et la soustraction de la valeur contrôle non stimulée. Ce phénotype noté IFNγ-BCG a été ajusté par régression linéaire sur l'âge, le taux annuel d'incidence de tuberculose dans le pays de naissance des individus, l'estimation de l'exposition totale de chaque individu au cas index, la contagiosité du cas index définie par la présence de cavernes sur sa radiographie des poumons et la présence de bacilles dans ses crachats, ainsi que sur la couverture par une complémentaire de santé. Le second phénotype noté IFNγ-ESAT6$_{BCG}$ correspondait à la production d'IFN-γ induite par la stimulation par l'antigène ESAT-6, transformé et ajusté de la même manière que le phénotype précédent. Il a simplement été ajusté en plus sur le phénotype IFNγ-BCG afin d'isoler une réponse plus spécifique d'ESAT-6 comme détaillé au chapitre B.

## 3. Méthodes Statistiques

### 1. Phasage et imputation des régions de liaison

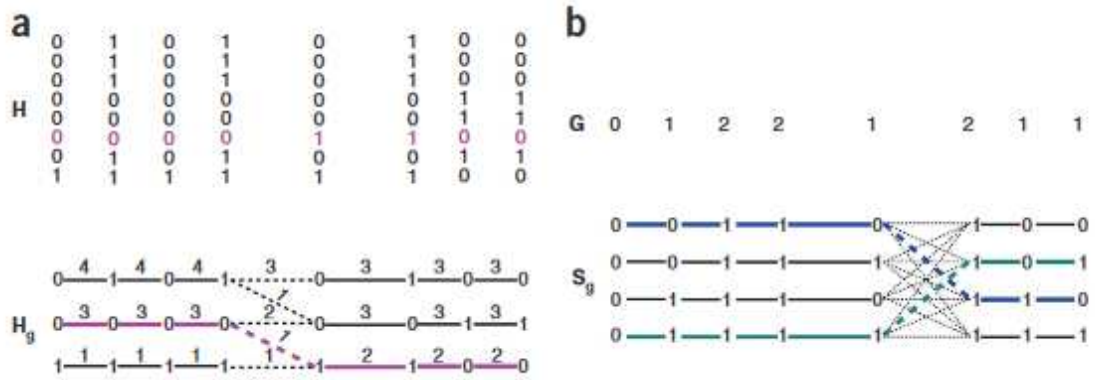
A partir des SNPs génotypés en première intention dans les 2 échantillons, nous avons cherché à prédire le génotype de marqueurs génétiques supplémentaires chez les sujets étudiés dans les régions chromosomiques d'intérêt. C'est ce qu'on appelle communément l'imputation de génotypes. Pour ce faire, il est nécessaire d'utiliser un panel d'haplotypes de référence pour lesquels le génotype de très nombreux variants est connu - 1000 Génomes phase 1 (181) dans notre cas – et sur lequel on se base pour imputer les génotypes absents chez nos sujets d'étude. Nous avons procédé en 2 étapes : le phasage des variants génotypés de très haute qualité à l'aide du logiciel SHAPEIT2 (182), puis l'imputation à partir du panel de référence avec le logiciel IMPUTE2 (183,184) qui semble être le meilleur compromis entre rapidité d'exécution et qualité d'imputation (185).

#### a) Principe du phasage

Pour pouvoir prédire le génotype de sujets à des positions non génotypées directement, il est nécessaire de connaître les haplotypes des individus. Les données génotypées à partir des puces Illumina ne comportent aucune information quant aux haplotypes des individus étudiés, la première étape consiste donc à phaser les génotypes des individus étudiés, c'est-à-dire à établir leurs haplotypes. Pour cela, parmi les différentes méthodes statistiques disponibles, nous avons choisi d'utiliser celle implémentée dans le logiciel SHAPEIT2 (182) que le projet 1000 génomes a lui-même sélectionné et qui est illustrée sur la figure 20. Cette méthode

partitionne les génotypes en groupes de marqueurs consécutifs (segments) comprenant au maximum 3 génotypes hétérozygotes, et établit la liste des 8 haplotypes possibles pour chaque segment. Toutes les paires d'haplotypes ne sont pas compatibles avec les génotypes observés, il existe donc des contraintes sur le choix des paires d'haplotypes compatibles, qui sont prises en compte dans la méthode. Une chaîne de Markov cachée calcule les probabilités de chaque paire d'haplotypes compatibles avec chaque segment, en fonction des haplotypes de référence fournis au modèle statistique au départ. Afin d'améliorer les performances computationelles de la méthode, ces haplotypes de référence sont eux-mêmes découpés en segments et les segments identiques regroupés et pondérés en fonction du nombre d'haplotypes de référence les possédant.

Figure 20 : Illustration du modèle utilisé dans SHAPEIT2 et des graphes associés sur un exemple simple extrait de (182)




(a,b) Dans cet exemple H contient K=8 haplotypes de référence (lignes de a) et les génotypes de l'individu que l'on souhaite phaser G contiennent 4 SNPs hétérozygotes parmi les M=8 marqueurs génotypés en colonnes (b) (0 correspond à un génotype homozygote sauvage et 2 à un génotype homozygote alternatif). Le graphe Hg est construit en découpant les haplotypes de H entre les marqueurs 4 et 5et contient donc ainsi 2 segments contenant chacun J=3 haplotypes distincts. Les nœuds du graphe sont étiquetés 0 ou 1 en fonction des allèles, et chaque arc est pondéré par le nombre d'haplotypes de H qui traversent cet arc. Un haplotype de H et son chemin correspondant dans Hg est coloré en violet.

En b, le graphe Sg est construit à partir de 2 segments de 5 et 3 marqueurs respectivement, chacun contenant 2 SNPs hétérozygotes de G. Chaque segment possède 4 haplotypes possibles et une paire de chemins compatible avec G est colorée en bleu et vert.

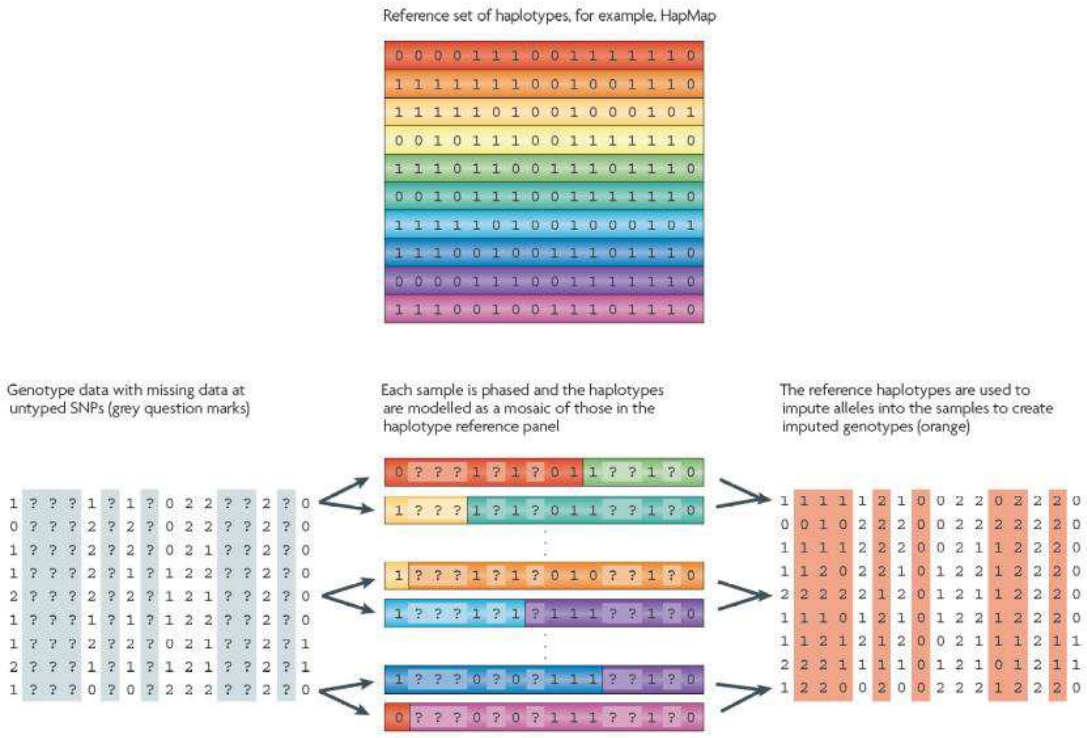
Pour notre étude, nous avons phasé les données génotypées des 2 échantillons d'étude sur les 2 chromosomes portant les régions de liaison qui nous intéressaient (les chromosomes 3 et 8) avec la version v2.r790 de SHAPEIT2 en n'utilisant que les variants de très bonne qualité : nous avons filtré les variants monomorphes, les variants avec taux de génotypage < 99% ainsi que les SNPs ambigus (A/T et C/G) concernant des bases nucléotidiques complémentaires. Nous avons utilisé les paramètres par défaut de la méthode avec les haplotypes du projet 1000 génomes phase 1 en panel de référence, et en utilisant l'option *duohmm* qui permet de prendre en compte les liens de parenté des individus dans les haplotypes trouvés.

b) Principe de l'imputation

Une fois les haplotypes des échantillons d'étude établis, la seconde étape consiste en l'imputation à proprement parler. L'imputation de données génétiques se base sur un panel d'haplotypes de référence pour lesquels le génotype de tous les variants que nous souhaitons imputer sont connus. Dans notre cas le panel comprend les 2 haplotypes des 1092 individus du projet de 1000 génomes phase 1 pour lesquels les génotypes de plus de 37 millions de variants ont été établis par séquençage haut débit et phasés à l'aide du logiciel SHAPEIT2 tel que décrit par Delaneau et al (186). Nous partons des génotypes connus de notre échantillon d'étude pré-phasés (cf paragraphe précédent) pour lesquels nous n'avons pas d'information pour les variants non génotypés ; nous disposons donc d'haplotypes « à trous » comme représenté sur la figure 21. En considérant les génotypes connus de l'échantillon d'étude, chaque haplotype « à imputer » est comparé à ceux du panel de référence et considéré comme un assemblage de morceaux d'haplotypes de référence. Une fois que les bornes de ces morceaux d'haplotypes sont définies par un algorithme de chaîne de Markov cachée, il ne reste plus qu'à imputer les génotypes manquants de l'haplotype étudié à partir des génotypes du morceau d'haplotype de référence pour lequel les génotypes de tous les variants sont connus. La dernière phase consiste à joindre les 2 haplotypes ainsi imputés de chaque individu pour connaître son génotype à chaque variant présent dans le panel de référence. Tout ce processus est expliqué en détails dans (184).

Les variants imputés ayant un critère d'information (mesure permettant de qualifier la fiabilité de l'imputation comprise entre 0 et 1) > 0.6 et une fréquence de leur allèle mineur (MAF) > 2% ont été retenus pour les analyses consécutives à cette imputation.

Figure 21 : Représentation du principe d'imputation repris de (183)




La sortie du logiciel IMPUTE2 peut se présenter sous 2 formes différentes. La première nommée « bestguess » consiste à attribuer à chaque variant imputé le génotype le plus probable d'après le logiciel. La seconde appelée dosage correspond à l'attribution pour chaque variant bi-allélique imputé de 3 probabilités correspondant aux probabilités d'observation de chacun des 3 génotypes possibles (pour un variant A/C par exemple, les génotypes AA, AC et CC), la somme des 3 probabilités étant égale à 1.

## 2. Analyses d'association

Les signaux de liaison génétique ont été identifiés en premier lieu dans l'échantillon du Val de Marne et ensuite répliqués dans l'échantillon d'Afrique du Sud. Nous avons donc opté pour une stratégie similaire en deux étapes pour l'analyse d'association. Nous avons d'abord réalisé une analyse d'association dans les régions de liaison sur le plus grand de nos échantillons, celui du Val de Marne, et ensuite testé en réplication les signaux les plus significatifs des 2 régions d'étude dans l'échantillon du Val de Marne. Ce choix a également été dicté par le fait que les phénotypes d'intérêt étaient proches dans les 2 échantillons mais pas complètement identiques, ce qui excluait une analyse conjointe des 2 échantillons.

Nous avons choisi dans une première étape d'analyser les variants imputés avec les données de dosage issues du logiciel IMPUTE2 selon 3 modèles génétiques : additif, récessif et dominant. Les données de dosage ont donc été agrégées en une seule valeur, différente selon le modèle génétique testé. Si on prend l'exemple d'un variant A/C imputé à 15% AA, 40% AC et 45% CC :

Pour un modèle additif pour l'allèle A, le génotype sera codé 0.15*2+0.4*1=0.70.

Pour un modèle récessif pour l'allèle A, le génotype sera codé 0.15.

Pour un modèle dominant pour l'allèle A, le génotype sera codé 0.15+0.4=0.55.

Les analyses d'association entre les SNPs de bonne qualité (génotypés ou imputés) et les 2 phénotypes auxquels nous nous sommes intéressés (IFNγ-BCG and IFNγ-ESAT6$_{BCG}$), ont été réalisées avec un modèle de régression linéaire mixte (LMM) implémenté dans le logiciel GEMMA (187) afin de prendre en compte les relations familiales existant au sein de nos échantillons. En effet, les approches LMM sont non seulement appropriées pour les études d'association en famille, mais également robustes et généralement plus puissantes que les méthodes traditionnelles basées sur le TDT (*Transmission Disequilibrium Test*) (188). Pour prendre en compte la dépendance des individus de nos échantillons, nous avons calculé la matrice de corrélation génétique ou matrice d'apparentement à partir des données génotypées de haute qualité (taux de génotypage > 99%, MAF >1%) et standardisées, et cette matrice a été introduite comme covariable dans le modèle de régression linéaire avec un effet aléatoire, permettant de corriger le modèle avec effets fixes, et de capturer la covariance des observations entre elles dues à leurs liens familiaux.

Nous avons en outre réalisé une analyse en composantes principales pour l'échantillon français avec la méthode EIGENSTRAT telle que nous l'avons décrite au chapitre B. Les 5 premières composantes principales de cette analyse ont été inclues dans le modèle de

régression utilisé pour les analyses d'association au Val de Marne en tant que covariables avec un effet fixe. En effet, il a été montré que l'introduction de la matrice de corrélation génétique ne suffisait pas toujours à corriger les effets de la structure de population en particulier en présence de facteurs de confusion environnementaux non identifiés mais liés à une localisation géographique particulière, et que l'ajustement sur les composantes principales utilisées en complément dans le modèle mixte permettait de mieux contrôler l'erreur de type I (soit le nombre de faux positifs) (189). A partir de l'analyse en composantes principales, nous avons également catégorisé les individus du Val de Marne en 3 sous populations : les caucasiens (en groupant ensemble les individus d'origine européenne et nord-africaine), les africains sub-sahariens et les asiatiques.

Chacune des 2 régions d'intérêt couvre environ 1% du génome entier. Par conséquent, nous avons considéré que $5.10^{-6}$ constituait un seuil de significativité raisonnable pour nos analyses, en nous basant sur un seuil de significativité de $5.10^{-8}$ à l'échelle du génome. Les SNPs atteignant une valeur de $p < 5.10^{-5}$ en association dans l'échantillon du Val de Marne avec au moins un des modèles génétiques testés (additif, récessif ou dominant) ont été évalués en réplication dans l'échantillon d'Afrique du Sud et nous avons regardé s'il existait une association dans le même sens pour les allèles sélectionnés. Nous avons exclu de l'analyse de réplication les SNPs imputés n'étant pas retrouvés dans panel du projet 1000 génomes Phase 3 (190), afin de nous focaliser sur les variants les plus fiables. Suite à l'analyse de réplication, les variants associés avec une valeur de $p < 0.05$ (en unilatéral) dans l'échantillon d'Afrique du Sud, ou avec un p initial $< 10^{-5}$ dans l'échantillon du Val de Marne et une tendance d'association en Afrique du Sud (c'est-à-dire une valeur $p < 0.5$ en unilatéral avec le même modèle génétique que celui du Val de Marne) ont été choisis pour être génotypés quand ils n'étaient qu'imputés. Une analyse d'association finale a ensuite été menée sur les SNPs génotypés dans les 2 cohortes sous le même modèle génétique.

### 3. Contributions au LOD-score

Nous avons ensuite cherché à savoir si les variants associés pouvaient, au moins en partie, expliquer les 2 signaux de liaison, en ajustant les phénotypes correspondants (IFNγ-BCG ou IFNγ-$ESAT6_{BCG}$) sur les génotypes des variants associés. Nous avons donc réalisé une analyse de liaison sur ce phénotype ajusté, en utilisant la méthode non paramétrique du nMLB-QTL (172) comme au chapitre B, avec une distribution du trait en déciles. Nous avons évalué si la baisse de LOD score observée après ajustement correspondait à une contribution significative des variants aux signaux de liaison en effectuant le même ajustement du

phénotype et la même analyse de liaison sur 1215 et 1109 variants génotypés avec la puce Ilumina HumanOmniExpressExome choisis respectivement et aléatoirement dans chacune des régions de liaison du chromosome 3 et du chromosome 8, avec une MAF > 2% et en relatif équilibre de liaison (coefficient de corrélation $r^2$ <0.5 entre chacun d'entre eux). Ces différentes analyses ont ainsi procuré des distributions empiriques de LOD scores auxquelles il était alors possible de comparer les valeurs obtenues après ajustement sur les variants associés.

### 4. Etude du déséquilibre de liaison

Pour étudier le profil de déséquilibre de liaison (LD) des SNPs les plus intéressants des analyses d'association, nous avons utilisé l'application web LDlink (https://analysistools.nci.nih.gov/LDlink/) (191) et regardé les variants en LD dans les 5 « super populations » du projet 1000 génomes phase 3, à savoir les européens, les africains, les individus originaires d'Amérique centrale ou du Sud, les sud-asiatiques et les individus originaires d'Asie de l'Est.

## 4. Résultats

### 1. Le phénotype IFNγ-BCG

Nous avons d'abord étudié la région génomique liée à la production d'IFN-γ dans les PBMCs suite à la stimulation par le bacille du BCG. Dans la région du chromosome 8 débutant à 61 Mb et se terminant à 91.5 Mb, 6 219 variants ont été génotypés dans l'échantillon du Val de Marne. Après un phasage avec le logiciel SHAPEIT (182), ces variants ont permis l'imputation de plus de 324 000 dans cette région de 30.5Mb, grâce à IMPUTE2 (183). Après contrôle qualité, nous avons retenu 117 354 variants avec une MAF > 2% et une information > 0.6 pour être analysés avec le logiciel GEMMA (187) afin de tester leur association avec le phénotype IFNγ-BCG.

Au total, 23 variants appartenant à 4 blocs différents de LD étaient associés au phénotype avec des $p < 5.10^{-5}$ (Tableau 7). Le manhattan plot de l'analyse est représenté sur la figure 22 et les plus forts signaux d'association ont été obtenus avec les variants rs202163431 ($p$=2.4x$10^{-6}$, bloc de LD 8-2, critère d'information = 0.65), et rs6981743 ($p$=2.7x$10^{-6}$, bloc de LD 8-4, critère d'information = 0.98). Ces 2 SNPs sont intergéniques et localisés à plus de 150 kb du gène codant le plus proche. Les 23 variants ont ensuite été analysés en réplication

dans l'échantillon d'Afrique du Sud. Seuls les 5 variants du bloc de LD 8.3 montraient une tendance à la réplication avec le même allèle associé aux hautes valeurs du phénotype IFNγ-BCG dans les 2 échantillons. Nous avons sélectionné un SNP parmi les 5 pour être génotypé, rs12056450, car il semblait représenter le meilleur compromis entre les résultats du Val de Marne et d'Afrique du sud. Nous avons obtenu les génotypes pour 368 individus de l'échantillon français et pour 236 de l'échantillon sud-africain. Chez les individus génotypés, la concordance entre les génotypes imputés dits « bestguess » (c'est-à-dire les plus probables selon IMPUTE2) et les génotypes réels était de 0.96 dans le Val de Marne et de 0.99 au Cap, confirmant la très bonne qualité de l'imputation.

Figure 22 : Manhattan plots montrant les résultats d'association génétique pour les 489 individus apparentés du Val de Marne après utilisation d'un modèle de régression linéaire mixte implémenté dans le logiciel GEMMA pour le phénotype IFNγ-BCG le long des 117 354 SNPs de la région du chromosome 8 comprise entre 61 Mb et 91.5 Mb.

La valeur minimale du logarithme décimal des valeurs p obtenues pour les modèles additif, récessif et dominant est visualisée en fonction de la position chromosomique en Mb, dans la région concernée. Une ligne horizontale à la valeur de −log10 (5.10-6) indique le seuil de significativité et les points en rouge représentent les SNPs appartenant aux groupes investigués plus en détails après les analyses de réplication.

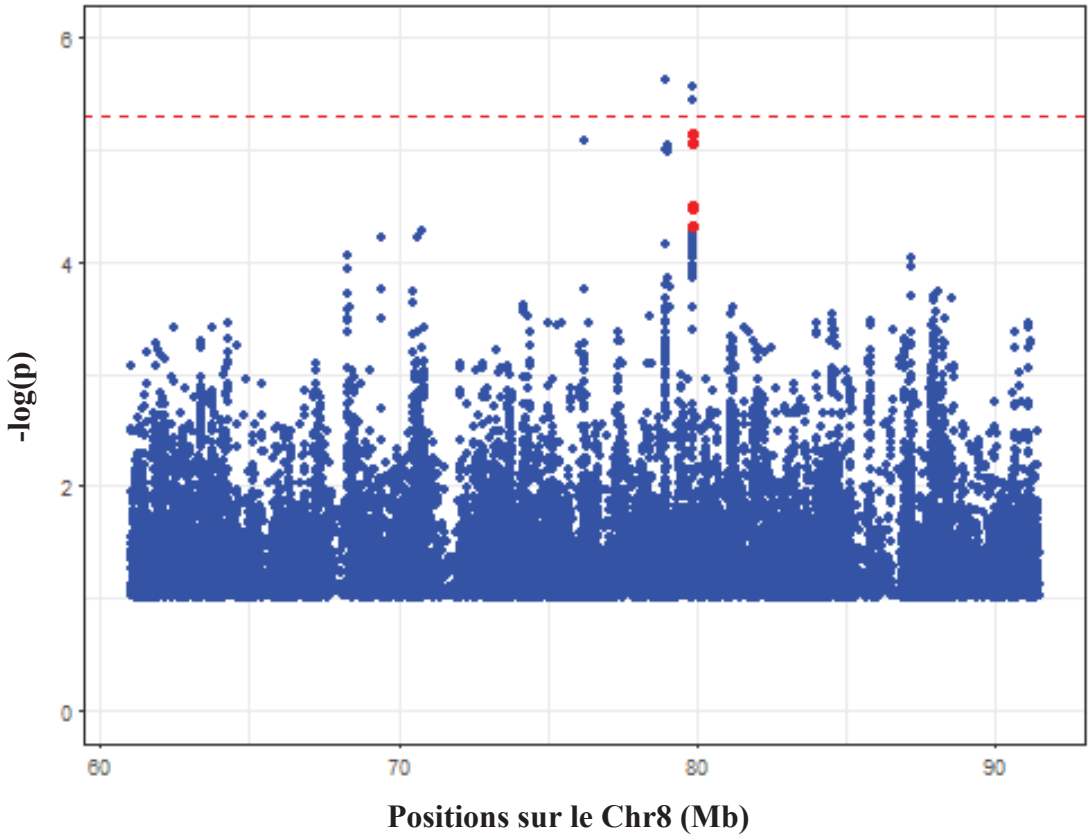


Tableau 7 : Résultats d'association sur les données imputées pour le phénotype IFNγ-BCG avec des $p < 5.10^{-5}$ dans l'échantillon du Val de Marne et dans l'échantillon d'Afrique du Sud.

| | | | | **Echantillon du Val de Marne** | | | | **Echantillon d'Afrique du Sud** | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| **Bloc de LD** | **Position (bp)** | **SNP** | **Allele*** | **Fréquence allélique** | **p** | **Modèle génétique**** | **Information §** | **Fréquence allélique** | **p***** | **Modèle génétique** | **Information§** |
| 8-1 | 76,236,341 | rs79392429 | C | 0.90 | 8.3E-06 | REC | 0.92 | 0.81 | >0.5 | - | 0.90 |
| 8-2 | 78,963,020 | rs117257553 | T | 0.04 | 9.9E-06 | ADD | 0.95 | 0.01 | >0.5 | - | 0.93 |
| | 78,963,622 | rs118187863 | G | 0.04 | 1.0E-05 | ADD | 0.95 | 0.01 | >0.5 | - | 0.93 |
| | 78,966,953 | rs202163431 | T | 0.06 | 2.4E-06 | ADD | 0.65 | 0.08 | >0.5 | - | 0.64 |
| | 78,978,131 | rs75231543 | A | 0.04 | 1.0E-05 | ADD | 0.95 | 0.01 | >0.5 | - | 0.91 |
| | 78,979,589 | rs141810731 | T | 0.04 | 1.0E-05 | ADD | 0.95 | 0.01 | >0.5 | - | 0.91 |
| | 78,981,194 | rs117762248 | C | 0.04 | 1.0E-05 | ADD | 0.95 | 0.01 | >0.5 | - | 0.91 |
| | 78,982,210 | rs77767122 | A | 0.04 | 1.0E-05 | ADD | 0.95 | 0.01 | >0.5 | - | 0.91 |
| | 78,984,646 | rs118084590 | A | 0.04 | 1.0E-05 | ADD | 0.95 | 0.01 | >0.5 | - | 0.91 |
| | 78,986,849 | rs117536094 | C | 0.04 | 1.0E-05 | ADD | 0.95 | 0.01 | >0.5 | - | 0.91 |
| | 78,987,821 | chr8:78987821:D | C | 0.04 | 1.0E-05 | ADD | 0.95 | 0.01 | >0.5 | - | 0.91 |
| | 78,991,235 | chr8:78991235:D | C | 0.04 | 1.0E-05 | ADD | 0.95 | 0.01 | >0.5 | - | 0.91 |
| | 78,995,179 | rs117746665 | C | 0.04 | 1.0E-05 | ADD | 0.95 | 0.01 | >0.5 | - | 0.91 |
| | 78,995,955 | rs78784982 | C | 0.04 | 1.0E-05 | ADD | 0.95 | 0.01 | >0.5 | - | 0.91 |
| | 78,998,300 | rs117528033 | A | 0.04 | 1.0E-05 | ADD | 0.95 | 0.01 | >0.5 | - | 0.91 |
| | 79,003,809 | rs147053766 | C | 0.04 | 9.2E-06 | ADD | 0.90 | 0.01 | >0.5 | - | 1.00 |

| | | | | Echantillon du Val de Marne | | | | Echantillon d'Afrique du Sud | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| **Bloc de LD** | **Position (bp)** | **SNP** | **Allele*** | **Fréquence allélique** | **p** | **Modèle génétique** | **Information** | **Fréquence allélique** | **p** | **Modèle génétique** | **Information** |
| 8-3 | 79,876,744 | rs6991466 | A | 0.25 | 7.1E-06 | ADD | 0.99 | 0.20 | 5.1E-02 | REC | 0.99 |
| | 79,885,854 | rs1427255 | C | 0.32 | 4.8E-05 | ADD | 0.99 | 0.23 | 2.2E-02 | REC | 1.00 |
| | 79,887,184 | rs12056761 | A | 0.32 | 3.3E-05 | ADD | 0.98 | 0.23 | 2.3E-02 | REC | 0.99 |
| | 79,887,368 | rs12056450 | G | 0.32 | 3.1E-05 | ADD | 0.98 | 0.23 | 2.3E-02 | REC | 0.99 |
| | 79,889,640 | rs11781015 | A | 0.32 | 8.6E-06 | DOM | 0.98 | 0.23 | 1.2E-01 | REC | 0.96 |
| 8-4 | 79,888,344 | rs7825423 | C | 0.41 | 3.6E-06 | DOM | 0.98 | 0.31 | >0.5 | - | 0.98 |
| | 79,888,908 | rs6981743 | T | 0.39 | 2.7E-06 | DOM | 0.98 | 0.30 | >0.5 | - | 0.97 |

* Allele associé avec les hautes valuers du phénotype

** Modèle génétique défini pour l'allèle associé aux hautes valeurs du phénotype

***p = valeur de p unilatérale issue des tests de rapport de vraisemblance dans l'échantillon d'Afrique du sud pour le meilleur modèle testé parmi les modèles additif, récessif et dominant. p > 0,5 signifie que l'allèle associé aux hautes valeurs du phénotype est différent en Afrique du Sud et dans le Val de Marne.

[§]critère qualité extrait du logiciel IMPUTE2. Information = 1 signifie que le SNp est génotypé tandis que information = 1.00 signifie que le SNP a été imputé.

Nous avons donc remplacé les génotypes imputés « bestguess » par les génotypes réels quand ils étaient disponibles, et avons regardé si l'association du variant avec le phénotype IFNγ-BCG en était modifiée (Tableau 8).

Tableau 8 : Résultats d'association pour le phénotype IFNγ-BCG sur les données génotypées du variant sélectionné d'après les critères détaillés dans le paragraphe des méthodes, pour les échantillons français et sud-africain.

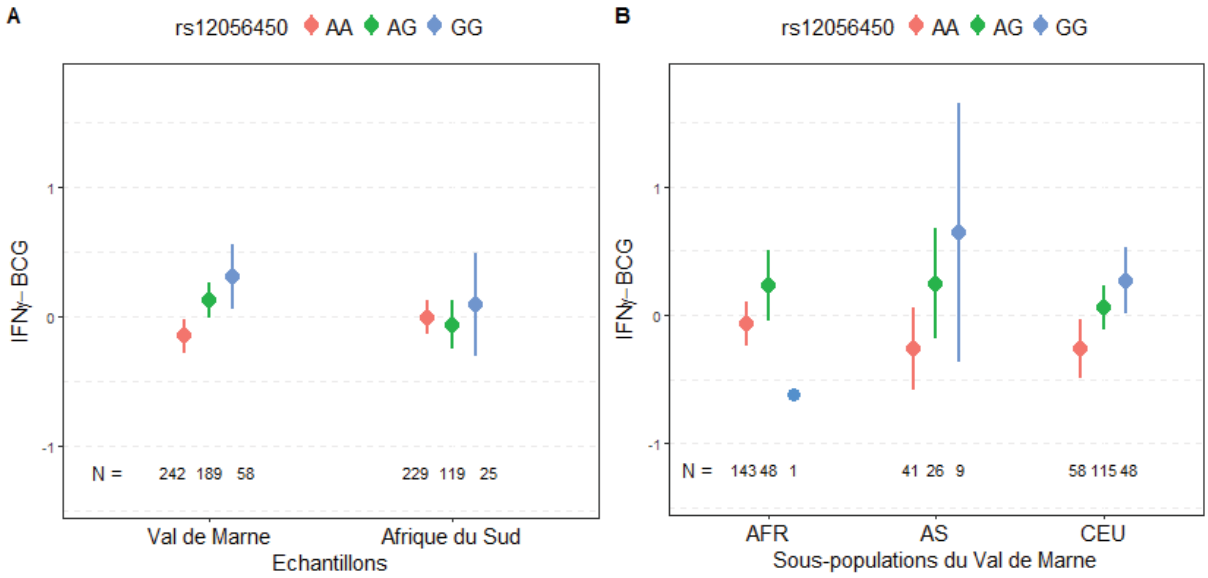
| | | | | | Val-de Marne | | Afrique du Sud | |
|---|---|---|---|---|---|---|---|---|
| **Bloc de LD*** | **Position (bp)** | **SNP** | **Allèle**** | **Modèle génétique** | **AF** | **p** | **AF** | **p** |
| 8-3 | 79887368 | rs12056450 | G | Additif | 0.31 | $1.2x10^{-5}$ | 0.23 | 0.25 |

* Bloc de LD tel que défini dans la table S1
** Allele associé aux hautes valeurs du phénotype
*** Modèle génétique pour l'allèle mentionné dans le tableau
AF = fréquence de l'allèle mentionné dans le tableau

Sous un modèle additif, le SNP rs12056450 renforce légèrement son association avec le phénotype par rapport au résultat sur les données imputées pour l'échantillon du Val de Marne ($p=1.16\ 10^{-5}$ versus $3.1\ 10^{-5}$ avec les données imputées), mais ce SNP n'est pas répliqué au seuil de 5% avec le même modèle génétique dans l'échantillon du Cap (p=0.25). Nous avons alors essayé de voir s'il existait une hétérogénéité d'effet entre les différentes sous-populations du Val de Marne définies à partir de l'analyse en composantes principales réalisée au chapitre B. La fréquence de l'allèle mineur G est comprise entre 0.13 chez les sujets d'origine africaine et 0.47 chez les caucasiens de l'échantillon français, avec une valeur intermédiaire de 0.23 dans l'échantillon sud-africain. L'effet du SNP, dont l'allèle G est associé avec les hautes valeurs du phénotype IFNγ-BCG, n'est pas homogène dans les différentes sous populations de l'échantillon du Val de Marne comme on peut le voir sur la figure 23 : le seul homozygote GG africain a une valeur de phénotype très basse dans le Val de Marne contrairement à ce qu'on pourrait attendre, et dans l'échantillon d'Afrique du Sud, les hétérozygotes AG ont un phénotype moyen légèrement plus bas que les homozygotes AA (ce qui explique la non significativité du SNP avec un modèle additif).

Figure 23 : Distribution du phénotype IFNγ-BCG en fonction des génotypes de rs12056450 dans les échantillons du Val de Marne et d'Afrique du Sud (**A**) et dans les différentes sous-populations du Val de Marne **(B).**

Les points correspondent aux moyennes et les barres d'erreur à l'intervalle de confiance à 95% de la moyenne, calculé sous l'hypothèse de normalité. Le phénotype IFNγ-BCG est standardisé.




Nous avons recherché, dans l'échantillon du Val de Marne, tous les variants en fort déséquilibre de liaison ($r^2$>0.8) avec rs12056450 et en avons trouvé 31. Ce bloc de 40kb se trouve dans une région intergénique débutant à la fin de l'ARN non-codant LOC105375914, et le gène codant le plus proche, IL7, en est distant de 140kb. Parmi ces variants en LD, rs12682556 ($r^2$ = 0.97 avec rs12056450 dans le Val-de-Marne, association avec IFNγ-BCG avec $p = 5.1.10^{-5}$) est référencé dans la base de données RegulomeDB comme appartenant à une région de liaison du facteur de transcription CTCF longue de 400bp (192).

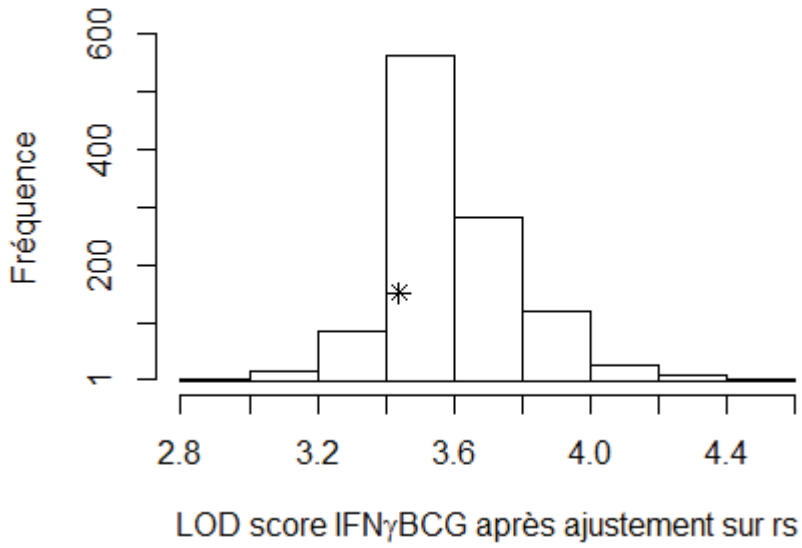
Nous avons également regardé les variants en LD dans les données du projet 1000 génomes phase 3, en particulier pour les super populations EUR et AFR. Nous avons trouvé 16 variants supplémentaires en $r^2 > 0.8$ avec rs12056450 dans au moins une de ces 2 super-populations mais aucun d'entre eux ne présentait un score fonctionnel supérieur à rs12682556 dans la base RegulomeDB.

Nous avons ensuite étudié la contribution du SNP rs12056450 au signal de liaison observé sur le chromosome 8 dans l'échantillon du Val de Marne, en ajustant les valeurs phénotypiques de IFNγ-BCG sur les génotypes dudit variant. L'ajustement sur rs12056450 a légèrement abaissé le LOD score de 3.80 à 3.44. Nous avons évalué la significativité de ce résultat en

calculant une distribution empirique de LOD scores telle que décrite dans le paragraphe des méthodes, mais avons trouvé que la baisse de LOD score observée pour rs12056450 n'était pas significative (p empirique = 0.14) (Figure 24).

Figure 24 : Distribution empirique de LOD scores utilisée pour évaluer la contribution des variants associés au pic de liaison.

La figure montre la distribution de LOD scores obtenue après ajustement du phénotype IFNγ-BCG sur 1109 variants imputés et tirés au hasard dans la région de liaison du chromosome 8 de 115 à 139 Mb, tous avec une MAF > 2% et une information >0.6. * correspond au LOD score obtenu après ajustement sur rs12056450.



### 2. Le phénotype IFNγ-ESAT6$_{BCG}$

Dans un second temps, nous nous sommes intéressés au locus impliqué dans la production d'IFN-γ en réponse à une stimulation par ESAT-6, antigène spécifique de *M.tuberculosis*, après ajustement sur la quantité d'IFN-γ générée en réponse au bacille du BCG. Au total, 5901 variants ont été génotypés et plus de 249 000 ont été imputés dans la région de liaison de 24 Mb du chromosome 3, en suivant la même stratégie que pour le chromosome 8. Des études d'association avec le phénotype IFNγ-ESAT6$_{BCG}$ ont été conduites dans l'échantillon du Val de Marne sur 93 218 variants ayant une MAF >2% et un critère d'information tel que défini par IMPUTE2 > 0.6. La figure 25 montre le résultat du meilleur des 3 modèles testés (additif, dominant et récessif) pour chaque variant testé.

Un total de 17 variants issus de 9 blocs de LD différents montrent une association avec un $p < 5\ 10^{-5}$ et sont résumés dans le tableau 9. Le signal d'association le plus fort est observé pour le SNP imputé rs116817490 (p= 5.10-6, information = 0.82) situé 14kb en aval du gène KLF15. Nous avons étudié ces 17 variants associés dans l'échantillon d'Afrique du Sud, et 3 d'entre eux appartenant à 2 blocs de LD différents (3-2 et 3-5) atteignaient les critères de réplication définis dans le paragraphe méthodes, à savoir une association suffisamment forte dans le Val de Marne ($p<10^{-5}$) avec le même allèle associé en Afrique du Sud ou bien une association en Afrique du Sud significative au seuil de 5% avec au moins des modèles génétiques testés. Deux de ces trois variants (rs9784373 et rs149692729) ont été imputés et sont en fort LD. Seulement un des deux (rs9784373) a été utilisé par la suite, ainsi que le SNP indépendant rs9828868 génotypé initialement dans l'échantillon du Val de Marne mais imputé au Cap. Suite au génotypage de ces variants, la concordance entre les génotypes imputés et les génotypes réels s'est avérée élevée, allant de 0.86 pour rs98228868 au Cap à 0.98 et 0.99 pour rs9784373 dans le Val de Marne et au Cap respectivement, confirmant la précision de l'imputation.

Figure 25 : Manhattan plot montrant les résultats d'association génétique pour les 489 individus apparentés du Val de Marne après utilisation d'un modèle de régression linéaire mixte implémenté dans le logiciel GEMMA pour le phénotype IFNγ-ESAT$_{bcg}$ le long de 93 218 SNPs de la région du chromosome 3 allant de 115 Mb à 139 Mb.

La valeur minimale du logarithme décimal des valeurs p obtenues pour les modèles additif, récessif et dominant est visualisée en fonction de la position chromosomique en Mb, dans la région concernée. Une ligne horizontale à la valeur de −log10 ($5.10^{-6)}$ indique le seuil de significativité et les points en rouge représentent les SNPs appartenant aux groupes investigués plus en détails après les analyses de réplication.

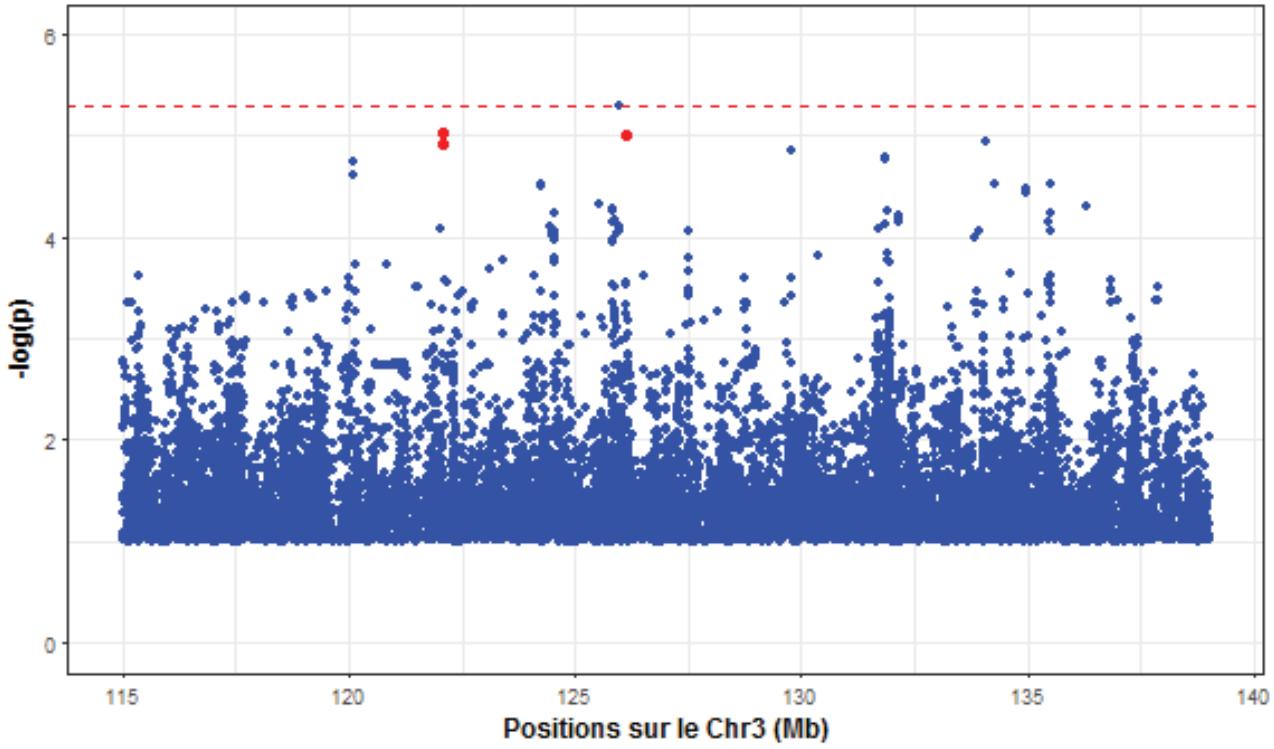


Tableau 9 : Résultats d'association avec un $p < 5.10^{-5}$ dans l'échantillon du Val de Marne sur les données imputées pour le phénotype IFNγ-$ESAT6_{BCG}$.

| | | | | Echantillon du Val de Marne | | | | Echantillon d'Afrique du Sud | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| **Bloc de LD** | **Position (bp)** | **SNP** | **Allele*** | **Fréquence allélique** | **p** | **Modèle génétique**** | **information§** | **Fréquence allélique** | **p***** | **Modèle génétique** | **information§** |
| 3-1 | 120,111,733 | chr3:120111733:D | T | 0.02 | 1.8E-05 | ADD | 0.96 | 0.01 | > 0,5 | - | 0.89 |
| | 120,117,992 | rs114720435 | T | 0.02 | 2.5E-05 | ADD | 0.97 | 0.01 | > 0,5 | - | 1 |
| | 120,129,794 | rs77250558 | T | 0.02 | 2.5E-05 | ADD | 0.96 | 0.02 | > 0,5 | - | 0.92 |
| 3-2 | 122,059,775 | rs9784373 | T | 0.05 | 9.5E-06 | DOM | 0.90 | 0.02 | 1.7E-05 | ADD | 0.93 |
| | 122,060,841 | rs149692729 | G | 0.05 | 1.2E-05 | DOM | 0.89 | 0.02 | 1.2E-05 | ADD | 0.94 |
| 3-3 | 124,248,975 | rs74677891 | C | 0.05 | 3.0E-05 | DOM | 0.96 | 0.02 | > 0,5 | - | 1 |
| | 124,253,483 | rs79289633 | G | 0.05 | 3.2E-05 | DOM | 0.97 | 0.02 | > 0,5 | - | 1 |
| | 124,253,581 | rs77945479 | G | 0.05 | 3.2E-05 | DOM | 0.97 | 0.02 | > 0,5 | - | 1.00 |
| 3-4 | 125,991,012 | rs116817490 | C | 0.03 | 5.0E-06 | DOM | 0.82 | 0.01 | > 0,5 | - | 0.66 |
| 3-5 | 126,129,646 | rs9828868 | T | 0.49 | 9.6E-06 | REC | 1 | 0.49 | 2.2E-01 | REC | 0.90 |
| 3-6 | 129,786,628 | rs57026314 | A | 0.03 | 1.4E-05 | DOM | 0.62 | 0.03 | 3.4E-01 | REC | 0.55 |
| 3-7 | 131,841,434 | rs140762555 | G | 0.05 | 1.7E-05 | ADD | 0.89 | 0.07 | > 0,5 | - | 0.84 |
| | 131,842,891 | rs59712276 | A | 0.03 | 1.7E-05 | ADD | 0.87 | 0.02 | > 0,5 | - | 0.98 |
| | 131,848,904 | rs75435362 | A | 0.03 | 1.7E-05 | ADD | 0.87 | 0.02 | > 0,5 | - | 0.99 |
| | 131,852,560 | rs57202631 | T | 0.03 | 1.7E-05 | ADD | 0.87 | 0.02 | > 0,5 | - | 0.98 |
| 3-8 | 135,484,040 | rs7431617 | G | 0.34 | 3.1E-05 | ADD | 0.98 | 0.43 | 4.8E-01 | ADD | 0.99 |
| 3-9 | 136,283,857 | rs60291974 | C | 0.84 | 4.9E-05 | REC | 0.66 | 0.89 | 2.7E-01 | ADD | 0.62 |

* Allele associé avec les hautes valeurs du phénotype

** Modèle génétique défini pour l'allèle associé aux hautes valeurs du phénotype

***p = valeur de p unilatérale issue des tests de rapport de vraisemblance dans l'échantillon d'Afrique du sud pour le meilleur modèle testé parmi les modèles additif, récessif et dominant. p > 0,5 signifie que l'allèle associé aux hautes valeurs du phénotype est différent en Afrique du Sud et dans le Val de Marne.

§critère qualité extrait du logiciel IMPUTE2. Information = 1 signifie que le SNp est génotypé tandis que information = 1.00 signifie que le SNP a été imputé.

Nous avons donc remplacé les génotypes imputés « bestguess » avec les génotypes réels lorsqu'ils étaient disponibles et refait les analyses d'association pour ces 2 variants. Les résultats sont synthétisés dans le tableau 10.

Le SNP rs9784373 donne avec les données génotypées des résultats similaires dans l'échantillon du Val de Marne avec un modèle dominant, mais l'association initialement trouvée dans l'échantillon d'Afrique du Sud avec les données imputées de dosage (p=1.7.10-5) est devenue beaucoup plus faible et n'est plus significative avec les données génotypées (p=0.14). Lorsqu'on regarde plus en détails les résultats pour essayer d'expliquer cette différence inattendue, on constate que les données de dosage ont favorisé le résultat de ce variant dans l'échantillon du Cap par rapport à celui obtenu avec les génotypes imputés les plus probables dits « bestguess » qui affichait un p à 0.17. La fréquence de ce variant est faible (MAF < 0.04) dans toutes les populations sauf chez les Africains de l'échantillon du Val de Marne où la MAF est de 0.1. On constate aussi un effet génétique différent chez les caucasiens du Val de Marne par rapport aux autres populations, comme on peut le voir sur la figure 26. Ce résultat d'association est donc difficile à interpréter.

Tableau 10 : Résultats d'association pour le phénotype IFNγ-ESAT6$_{bcg}$ sur les données génotypées pour les 2 variants sélectionnés d'après les critères détaillés dans le paragraphe des méthodes, pour les échantillons français et sud-africain.

| | | | | | Val-de Marne | | Afrique du Sud | |
|---|---|---|---|---|---|---|---|---|
| **Bloc de LD*** | **Position (bp)** | **SNP** | **Allèle**** | **Modèle génétique** | **AF** | **p** | **AF** | **p** |
| 3-2 | 122059775 | rs9784373 | T | Dominant | 0.05 | $1.9x10^{-5}$ | 0.02 | 0.14 |
| 3-5 | 126129646 | rs9828868 | T | Récessif | 0.49 | $9.6x10^{-6}$ | 0.46 | 0.19 |

* Bloc de LD tel que défini dans la table S2

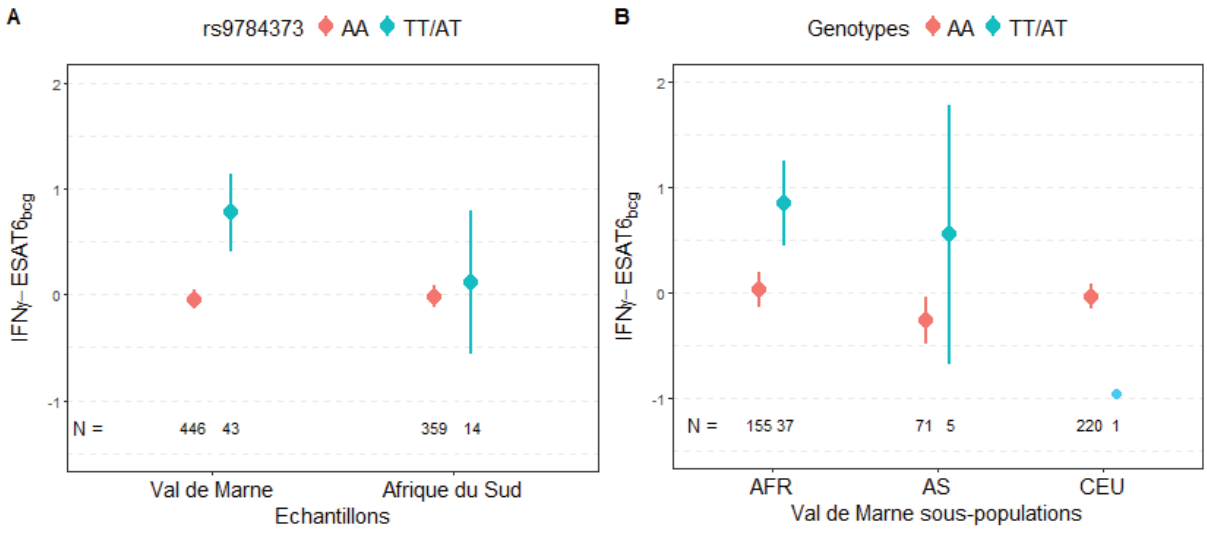
** Allele associé aux hautes valeurs du phénotype

*** Modèle génétique pour l'allèle mentionné dans le tableau

AF = fréquence de l'allèle mentionné dans le tableau

Figure 26 : Distribution du phénotype IFNγ-ESAT6$_{BCG}$ en fonction des génotypes de rs9784373 dans l'échantillon du Val-de-Marne et dans l'échantillon du Cap (**A**) et dans les différentes sous-populations du Val de Marne **(B).**

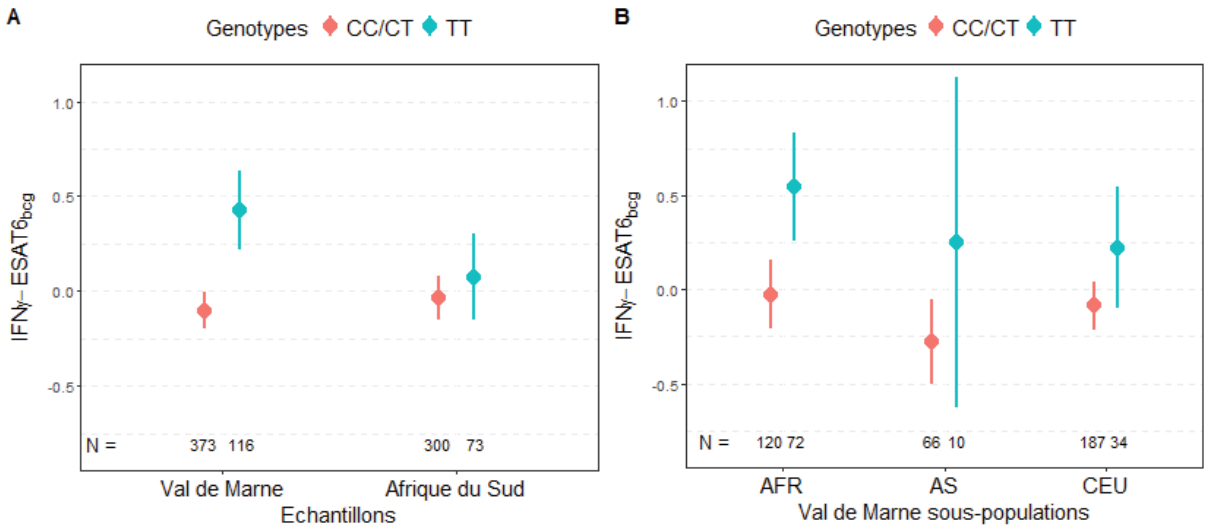
Les points correspondent aux moyennes et les barres d'erreur à l'intervalle de confiance à 95% de la moyenne, calculé sous l'hypothèse de normalité. Le phénotype IFNγ-ESAT6$_{BCG}$ est standardisé.




Le variant rs9828868 qui avait déjà été génotypé dans l'échantillon du Val de Marne (association avec p=9.6.10$^{-6}$ sous un modèle récessif) renforce légèrement son association dans l'échantillon d'Afrique du Sud après génotypage avec une valeur de p passant de 0.22 à 0.19 avec le même modèle récessif comme on peut le voir dans le tableau 9. La fréquence de son allèle mineur T est de 0.49 dans l'échantillon du Val de Marne et de 0.46 dans celui du Cap. Les sujets homozygotes TT ont des valeurs du phénotype IFNγ-ESAT$_{bcg}$ plus élevées que les sujets CT ou CC (la différence est de l'ordre d'un demi-écart type dans le Val de Marne), et on constate que cet effet est homogène dans les 3 sous populations principales du Val de Marne (caucasiens, africains subsahariens et asiatiques) (cf Figure 27).

Figure 27 : Distribution du phénotype IFNγ-ESAT6$_{BCG}$ en fonction des génotypes de rs9828868 dans l'échantillon du Val-de-Marne et dans l'échantillon du Cap (**A**) et dans les différentes sous-populations du Val de Marne **(B).**

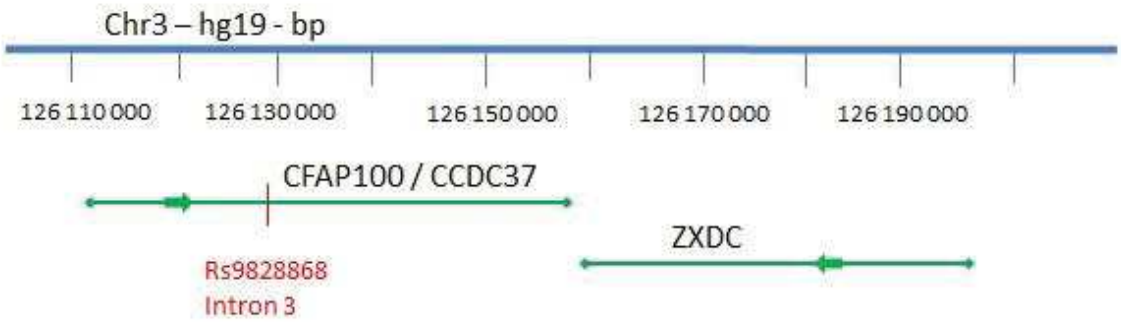
Les points correspondent aux moyennes et les barres d'erreur à l'intervalle de confiance à 95% de la moyenne, calculé sous l'hypothèse de normalité. Le phénotype IFNγ-ESAT6$_{BCG}$ est standardisé.




Nous avons également constaté que ce variant expliquait 15% de la variance génétique du phénotype IFNγ-ESAT$_{bcg}$ de l'échantillon du Val de Marne en comparant la proportion de variance expliquée par la génétique dans le modèle de régression avec et sans ajustement sur rs9828868. En regardant le profil de LD du variant dans le projet 1000 génomes phase 3, nous avons trouvé seulement un SNP en $r^2$ >0.8 dans les « super populations » d'Européens et d'Asiatiques, rs4679239. Ce SNP faisait partie des variants imputés avec IMPUTE2 mais n'était pas fortement associé au phénotype IFNγ-ESAT$_{bcg}$ dans l'échantillon du Val de Marne. L'association observée semble donc due à un unique variant rs9828868 situé dans un intron du gène CFAP100 (cilia and flagella associated protein 100 or CCDC37), comme on peut le voir sur la figure 28.

Figure 28 : Localisation de rs9828868 sur le chromosome 3 avec les coordonnées définies selon l'assemblage du génome humain hg19.

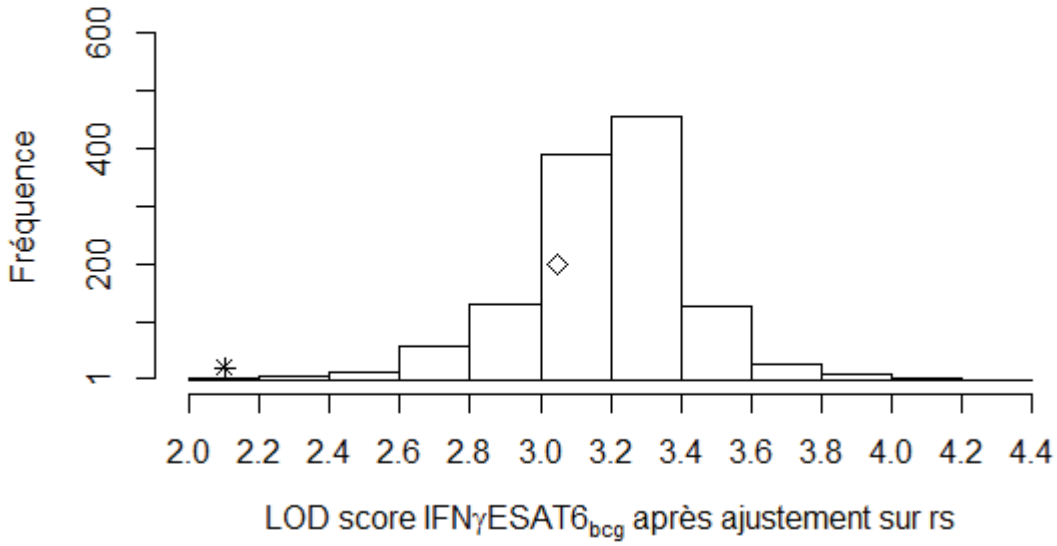
Le SNP est situé dans le gène *CFAP100* (cilia and flagella associated protein 100), à environ 30kb de son gène cible *ZXDC* (zinc finger X-linked duplicated family member C).




Nous avons ensuite cherché à voir si ces 2 variants potentiellement associés à IFNγ-ESAT$_{BCG}$ pouvaient expliquer le signal de liaison observé sur le chromosome 3. Pour ce faire, nous avons ajusté les valeurs du phénotype sur le génotype des SNPs correspondants, réalisé une analyse de liaison sur ces phénotypes ajustés, puis comparé les LOD scores ainsi obtenus à la distribution empirique de LOD scores établie comme décrit dans le paragraphe méthodes. Après ajustement sur le variant rs9784373, le LOD score était à 3.05 dans l'échantillon français, proche de la valeur initiale de 3.26 obtenue sur les mêmes 368 individus (p empirique = 0.22). En revanche, après ajustement sur rs9828868, le LOD score a baissé de 3.26 à 2.05. Cette chute s'est avérée très significative (p empirique <0.001), plus importante que toutes celles observées sur les 1215 variants indépendants sélectionnés aléatoirement dans la région de liaison (Figure 29). Cette dernière observation renforce l'intérêt du variant rs9828868 pour lequel l'évidence d'association avec le phénotype IFNγ-ESAT$_{BCG}$ est la plus grande. De façon tout à fait intéressante, dans une méta-analyse réalisée sur plus de 5000 individus d'origine caucasienne, ce variant a été rapporté comme régulant significativement l'expression du gène *ZXDC* (zinc finger X-linked duplicated family member C) dans les cellules sanguines (193). *ZXDC* joue un rôle dans la production d'interleukine 12 (IL12) par les monocytes, sachant que cette cytokine est particulièrement importante dans l'immunité anti-mycobactérienne et joue un rôle dans la production d'IFN-γ par les lymphocytes T (36,194).

Figure 29 : Distribution empirique de LOD scores utilisée pour évaluer la contribution des variants associés au pic de liaison.

La figure montre la distribution de LOD scores obtenue après ajustement du phénotype IFNγ-ESAT6$_{BCG}$ sur 1215 variants imputés et tirés au hasard dans la région de liaison du chromosome 3 de 115 à 139 Mb, tous avec une MAF > 2% et une information >0.6. * correspond au LOD score obtenu après ajustement sur rs9828868 et ◇ au LOD score obtenu après ajustement sur rs9784373.



## 5. Discussion du fine-mapping des régions de liaison

Dans cette étude, nous avons étudié en détails les régions de liaison décrites au chapitre I-B grâce à des analyses d'association génétique réalisées sur des données de génotypage et d'imputation extrêmement denses. Pour le locus 8q12-22 contrôlant la production d'IFN-γ par les PBMCs en réponse à une stimulation par le bacille du BCG, nous avons identifié, dans l'échantillon du Val de Marne, un groupe de SNPs intergéniques représenté par le variant rs12056450 potentiellement associé au phénotype étudié. Ce groupe de SNPs présentait la même direction d'association dans l'échantillon de réplication d'Afrique du Sud, avec le même modèle génétique que dans l'échantillon primaire. Cependant, l'effet sur les niveaux d'IFN-γ chez les individus hétérozygotes AG pour rs12056450 au Val de Marne n'était pas observé chez les individus hétérozygotes AG du Cap. Le groupe de variants n'expliquait pas non plus une part substantielle du signal de liaison du chromosome 8. Des études supplémentaires semblent donc nécessaires pour confirmer ou infirmer le rôle potentiel de ce groupe de SNPs dans la production d'IFN-γ en réponse au BCG.

Le second signal de liaison contrôlant la production d'IFN-γ induite par l'antigène ESAT-6 après avoir pris en compte la quantité sécrétée suite à une stimulation par le BCG a été localisé sur le chromosome 3q13-22. Nos analyses ont identifié un variant commun isolé rs9828868 associé au phénotype avec $p = 9.6.10^{-6}$ chez les individus du Val de Marne, et une tendance à l'association avec le même modèle génétique chez les individus sud-africains de notre cohorte de réplication. Cette association plus ténue observée dans l'échantillon du Cap est cohérente avec le signal de liaison également plus faible observé au chapitre I-B. Les individus homozygotes pour l'allèle T ont des valeurs plus élevées d'IFN-γ que ceux ayant un génotype C/T ou CC, de l'ordre d'un demi-écart type dans l'échantillon du Val de Marne, et cet effet est homogène pour les trois sous-populations principales de l'échantillon français (caucasienne, africaine et asiatique). Le SNP rs9828868 explique à lui seul 15% de la variance génétique du phénotype IFNγ-$ESAT6_{BCG,}$ et surtout contribue significativement au signal de liaison du chromosome 3, ce qui renforce l'hypothèse d'une authentique association entre le variant et le trait étudié.

Le SNP rs9828868 a été décrit comme régulant l'expression du gène *ZXDC* (eQTL) dans les cellules sanguines, avec son allèle T associé à une expression basse de *ZXDC* (193). Le produit de *ZXDC* a été d'abord décrit comme une protéine en doigt de zinc liant *CIITA* et contribuant à la transcription des gènes du complexe majeur d'histocompatibilité (MHC) de

classe II (195). Il régule aussi l'expression de gènes impliqués dans la fonction et la différenciation des monocytes. L'isoforme la plus grande en particulier, ZXDC1, active l'expression de *CCL2* (chemokine ligand 2, aussi connu sous le nom de *MCP-1,* monocyte chemoattractant protein 1) en prenant la place du répresseur de transcription *BCL6* (196). L'inhibition de *ZXDC* résulte en une augmentation du taux d'occupation du promoteur de *CCL2* par *BCL6* après stimulation par PMA (phorbol 12-myristate 13-acetate), et en une diminution de l'expression de *CCL2*. En se basant sur ces observations, les individus portant l'allèle T du variant rs9828868, et a fortiori les individus homozygotes TT, doivent donc exprimer faiblement le gène *ZXDC*, et par conséquent produire moins de protéine CCL2. Plusieurs études de cellules humaines *in vitro* ont rapporté que la protéine CCL2 inhibe la production d'IL12, en particulier dans les monocytes stimulés par *M.tuberculosis* (197,198). Cela conduit donc à l'hypothèse que les homozygotes TT de rs9828868 pourraient avoir des niveaux de phénotype IFNγ-ESAT6$_{BCG}$ plus élevés que les autres individus en raison d'une augmentation de la production d'IL-12 causée par une régulation négative de *CCL2* dépendant de *ZXDC*.

En conclusion, nous avons identifié un variant rs9828868 associé avec la quantité d'IFN-γ générée spécifiquement par l'antigène ESAT-6, dans les PBMCs, après avoir tenu compte de la capacité intrinsèque de réponse en IFN-γ de chaque individu, dans un échantillon du Val de Marne, très hétérogène en termes d'origines ethniques. Ce variant explique une part significative du pic de liaison du chromosome 3, identifié au chapitre I-B dans la même population d'étude, et semble impliquer dans l'expression du gène *ZXDC*, lui-même lié à la production d'IL-12. Dans l'échantillon de réplication du Cap, le même allèle est associé aux hautes valeurs du trait étudié, bien que l'association ne soit pas significative au seuil de 5%. En effet, les phénotypes utilisés pour la réplication en Afrique du sud sont proches mais pas identiques à ceux du Val de Marne (IFN-γ produit par les cellules du sang total vs seulement par les PBMCs, phénotypes mesurés après 3 jours d'incubation versus 4 jours, covariables d'ajustement différentes dans les 2 échantillons), et ces petites différences peuvent avoir conduit à un manque de puissance pour la réplication. D'autre part, les 2 populations diffèrent considérablement vis-à-vis de leur mode d'exposition à *M.tuberculosis* ; les individus d'Afrique du Sud vivent dans une région hyperendémique pour la tuberculose, au sein de laquelle la transmission a lieu préférentiellement au niveau de la communauté (100), au contraire des individus de l'étude française qui ciblait des personnes ayant été en contact avec la tuberculose dans leur entourage proche. Les 2 cohortes diffèrent également en terme de fonds génétique ; les familles du Val de Marne appartenaient à divers groupes ethniques,

tandis que tous les individus du Cap provenaient de la même population sud-africaine, les « coloured », résultat du métissage de plusieurs ethnies à savoir les Khoisans (31%), les Bantus (33%), les Européens (16%) et les Asiatiques (20%) (178). Dans ce contexte, le résultat obtenu pour rs9828868 malgré une hétérogénéité ethnique et environnementale manifeste est particulièrement prometteur.

# D. Discussion générale sur la génétique de l'infection tuberculeuse

A l'issue des travaux que je viens de détailler, des facteurs génétiques contrôlant la production d'IFN-γ en réponse à une attaque mycobactérienne ont été localisés sur le génome. La stratégie classique d'analyse de liaison génétique suivie d'une analyse d'association dans les régions de liaison, très efficace dans les études de maladies monogéniques, a également été fructueuse pour nous, pour des phénotypes d'infection tuberculeuse par définition plus complexes. Elle nous a permis d'identifier une région du chromosome 8 liée à la production d'IFN-γ en réponse à une agression mycobactérienne telle que celle représentée par le BCG, et une seconde région sur le chromosome 3 liée à la production d'IFN-γ en réponse à une stimulation très spécifique de l'antigène ESAT-6 au sein de laquelle un variant mono-nucléotidique en particulier, rs9828868, semble associer à cette réponse et expliquer une part significative du signal de liaison de cette région. La difficulté particulière d'un phénotype tel que celui de l'infection tuberculeuse repose sur le fait que de nombreux facteurs jouent un rôle dans l'expression de celui-ci et que la part de chacun est difficile à isoler et quantifier. L'hétérogénéité des facteurs génétiques est sans aucun doute importante, et dans le travail présenté ici, la quantité et la précision des informations disponibles pour chaque individu dans l'échantillon primaire, en particulier concernant l'exposition à l'agent infectieux, ont joué un rôle primordial dans nos résultats.

Pour aller plus loin, plusieurs voies pourraient être explorées. La première serait de rechercher un échantillon de réplication plus proche de notre échantillon primaire, en termes d'origines ethniques (ou tout au moins d'une de ses sous-populations) et de définitions de phénotype, afin d'optimiser les chances de pouvoir retrouver les mêmes signaux d'association dans les 2 groupes. En effet, les meilleurs signaux d'association trouvés dans l'échantillon primaire du

Val de Marne n'ont pas été répliqués dans celui du Cap et cela peut être simplement dû à un trop grand écart entre les 2 cohortes. Pour le phénotype lié à la réponse au BCG en particulier, 2 variants dépassaient le seuil de significativité, et l'un des 2, rs202163431, semblerait expliquer une part importante du signal de liaison (LOD score après ajustement sur données de dosage = 2.67 versus 3.80 initialement). Cet échantillon de réplication pourrait être familial si on souhaite consolider le résultat de liaison et se rapprocher au plus près du mode de recrutement des individus du Val de Marne, ou bien en population générale en vue de privilégier des résultats d'association potentiels (les analyses d'association sur données familiales étant en général plus conservatives que celles sur données indépendantes).

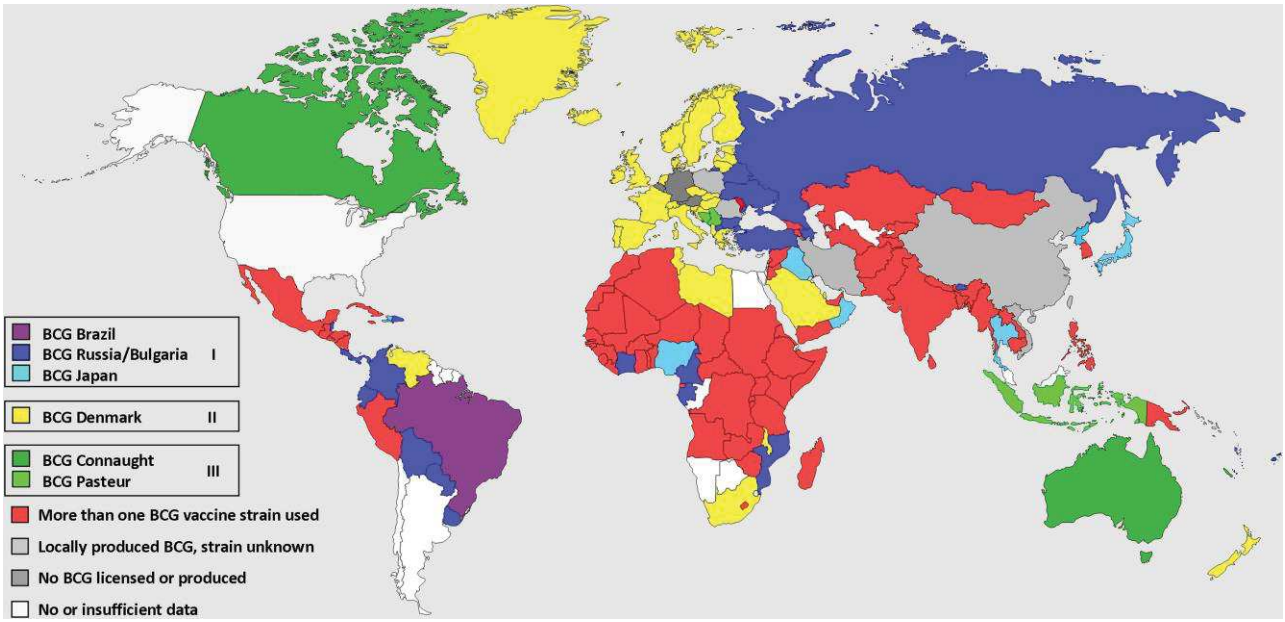
La seconde option, qui n'est pas incompatible avec la première, serait de séquencer entièrement les régions de liaison dans la cohorte primaire (et dans la cohorte de réplication) afin de pouvoir étudier l'impact éventuel de variants rares (MAF <2%) que nous n'avons pu capturer ni par génotypage, ni par imputation. Il n'est pas improbable que les signaux de liaison soient supportés par des variants rares ou familiaux, différents pour chacune des familles étudiées et contributives à la liaison, mais liés au même gène par exemple. Cette hypothèse pourrait expliquer le fait que l'analyse d'association du chromosome 8 n'ait pas donné de résultats aussi clairs qu'on aurait pu l'espérer au vu de l'analyse de liaison.

D'autre part, nous n'avons pas tenu compte dans nos analyses de l'influence potentielle de la souche de *M.tuberculosis* à laquelle les individus ont été exposés, ni de la souche de BCG avec laquelle les individus ont été vaccinés. Il existe en effet de nombreuses souches vaccinales de BCG à travers le monde (199) (cf Figure 30), et leurs pouvoirs immunogènes peut différer en fonction des milieux sur lesquels ils sont cultivés (200). La souche vaccinale pourrait donc avoir un impact sur la réponse en IFN-γ suite à une stimulation par BCG *in vitro* et donc constituer un facteur de confusion dans notre étude. Indépendamment de la vaccination par le BCG, les souches de *M.tuberculosis* avec lesquelles les individus étudiés ont été en contact sont fort probablement différentes dans le Val de Marne et en Afrique du Sud. Or il a été rapporté que le gène codant pour l'antigène ESAT-6 pouvait être plus moins exprimé selon la virulence des souches de *M.tuberculosis* (201). Il n'est donc pas impossible que la réponse immunitaire de l'hôte dépende pour une part de la souche précise de *M.tuberculosis* cherchant à l'infecter, et que le fait d'analyser les réponses en IFN-γ suite à une stimulation par ESAT-6 sans tenir compte de cette variabilité nous fasse perdre de la puissance et constitue un facteur explicatif de la difficulté de réplication des signaux d'un échantillon à un autre. Collecter des informations plus précises sur la nature de l'agent

infectieux pourrait donc également être une voie d'amélioration pour de futurs échantillons d'étude.

Figure 30 : Souches de BCG utilisées en vaccin à travers le monde entre 2003 et 2007 d'après (199).

Les cadres entourent les souches de BCG relativement similaires génétiquement .Le cadre I inclut les souches obtenues de l'institut Pasteur avant 1926. Les cadres II et III représentent les souches obtenues plus tard.




Pour poursuivre plus avant l'analyse du variant rs9828868 associé à la réponse en IFN-γ spécifique de l'antigène ESAT-6, qui représente le résultat le plus abouti de cette première partie, plusieurs pistes sont envisageables. Une première piste serait de pouvoir répliquer cette observation dans d'autres cohortes d'individus ayant été en contact avec *M.tuberculosis* et pour lesquelles un phénotype d'infection tuberculeuse serait disponible. Au vu des résultats obtenus dans le Val de Marne, une cohorte d'origine africaine optimiserait nos chances de réplication. Une autre piste, fonctionnelle cette fois, serait de vérifier notre hypothèse de lien entre rs9828868 et la production d'IL-12 dans les cellules sanguines, et en particulier dans les PBMCs, en mesurant la production d'IL-12 dans différents individus de notre cohorte primaire en fonction de leur génotype au SNP après stimulation par ESAT-6 et par BCG, puis de confirmer l'influence de *ZXDC* dans la production d'IL-12, en activant ou diminuant l'expression du gène chez les mêmes individus pour en étudier la conséquence sur la cytokine. Si tout est confirmé, il restera à établir si le variant rs9828868 influe uniquement sur la

réponse à l'ESAT-6, et si c'est le cas, pourquoi il n'influe pas sur la réponse aux autres antigènes.

La question restant à discuter concerne l'intérêt immunologique et clinique de cette recherche de facteurs génétiques influant sur la réponse immunitaire en IFN-γ suite à une agression mycobactérienne. L'étude des cas de tuberculose sévère de l'enfant a montré qu'une déficience de réponse immunitaire médiée par l'IL-12 et l'IFN-γ pouvait être la cause de la gravité de la maladie, et que l'injection d'IFN-γ recombinant chez les patients en complément des antibiotiques traditionnels pouvait être très bénéfique (202,203). La cytokine joue donc un rôle important dans la défense anti-mycobactérienne en général, et anti-tuberculeuse en particulier.

La norme aujourd'hui, utilisée dans l'utilisation des tests IGRA, est de considérer qu'une qu'une forte production d'IFN-γ après avoir été exposé à *M.tuberculosis* reflète un statut d'infection et constitue un argument en faveur d'un traitement antibiotique pour prévenir une éventulelle tuberculose clinique.

Mais que signifie une faible production ou une absence de production d'IFN-γ après stimulation des cellules sanguines par des antigènes mycobactériens ? On peut suggérer 3 explications :

- La première, c'est que l'individu n'a pas été exposé (ou pas assez) à la mycobactérie et qu'il n'est donc pas en mesure de monter une réponse immunitaire spécifique de celle-ci dans le temps d'observation imparti. Dans notre étude, nous avons essayé de minimiser ce cas de figure en quantifiant l'intensité d'exposition des individus à *M.tuberculosis* et en ajustant les réponses en IFN-γ sur cette mesure.
- La seconde c'est que bien qu'exposé, l'individu n'a pas été infecté par le bacille ; son corps l'a éliminé spontanément sans laisser de trace dans son système immunitaire (pas de lymphocytes T mémoire) et qu'il se retrouve donc comme dans le premier cas avec une impossibilité de réponse immunitaire spécifique dans le temps d'observation.
- La troisième explication, c'est que l'individu a bien été exposé et infecté par la mycobactérie, mais qu'il n'est capable de produire intrinsèquement qu'une petite quantité d'interféron gamma. Ce cas de figure conduirait à une tuberculose clinique si ce manque de réponse est dû à un déficit immunitaire inné ou acquis. Dans nos travaux, toute personne déclarant une tuberculose active dans le temps de l'étude a été écartée pour éviter ce cas-là.

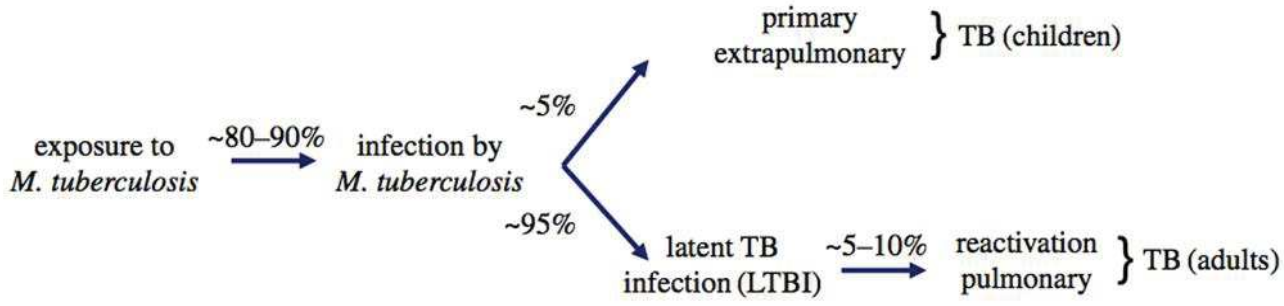
La recherche de facteurs génétiques influant sur la capacité de réponse en IFN-γ face aux mycobactéries, et à *M.tuberculosis* en particulier, vise in fine à identifier les individus appartenant à la seconde catégorie, ceux qui seraient résistants à l'infection et à comprendre ce qui, d'un point de vue génétique, les différencie des autres. La résistance à l'infection est une notion subjective dans la mesure où elle peut dépendre de l'intensité ponctuelle et/ou cumulée dans le temps de l'exposition à l'agent pathogène. On peut imaginer une notion de résistance simple et totale : la mycobactérie ne pénètrera jamais dans le corps de l'individu quelle que soit l'exposition même soutenue et continue pendant toute une vie à l'agent infectieux. Ces individus seraient extrêmement intéressants à étudier pour identifier chez eux la source de cette résistance innée à l'infection, mais ils sont difficiles à trouver. Il pourrait aussi exister une notion de résistance à l'infection tuberculeuse plus progressive qui modulerait le seuil d'exposition cumulée nécessaire pour que l'agent infectieux pénètre dans le corps et y laisse une trace. C'est ce type de résistance-là que des études comme celles que j'ai menées ici essaient de mettre en lumière, et le variant rs9828868 pourrait faire partie des facteurs modulant ce seuil de tolérance à l'antigène ESAT-6.

# II - Génétique humaine de la tuberculose pulmonaire

## A. Introduction

Après avoir étudié le processus d'infection par *M.tuberculosis*, le passage d'une infection tuberculeuse latente asymptomatique vers une forme pulmonaire de tuberculose clinique fait l'objet des travaux décrits dans la seconde partie de ce manuscrit. Comme le rappelle la figure 31, on estime en règle générale que 10% des individus infectés par la mycobactérie déclareront une tuberculose au cours de leur vie (204). Environ la moitié d'entre eux, particulièrement les jeunes enfants en zone d'endémie, développeront une tuberculose dite « primaire » dans les 2 ans suivant l'infection, souvent associée à des symptômes extra-pulmonaires (205). La seconde moitié des individus développera la maladie plus tardivement, sous une forme majoritairement pulmonaire et appelée tuberculose de réactivation lorsque l'infection jusqu'alors sous contrôle devient une maladie clinique reflétant une altération de la résistance naturelle de l'hôte à la mycobactérie (23,30,205). Des facteurs de risque environnementaux (tels que la mycobactérie elle-même) et médicaux (comme l'infection par le VIH ou la prise d'un traitement anti-TNFα) contribuent à la déclaration de la maladie comme nous l'avons détaillé dans l'introduction, mais ils ne peuvent expliquer à eux-seuls la totalité des cas observés, et une accumulation d'observations au fil des années tend à montrer que la génétique humaine joue un rôle dans la variabilité de progression de l'infection tuberculeuse latente vers un stade clinique de la maladie.

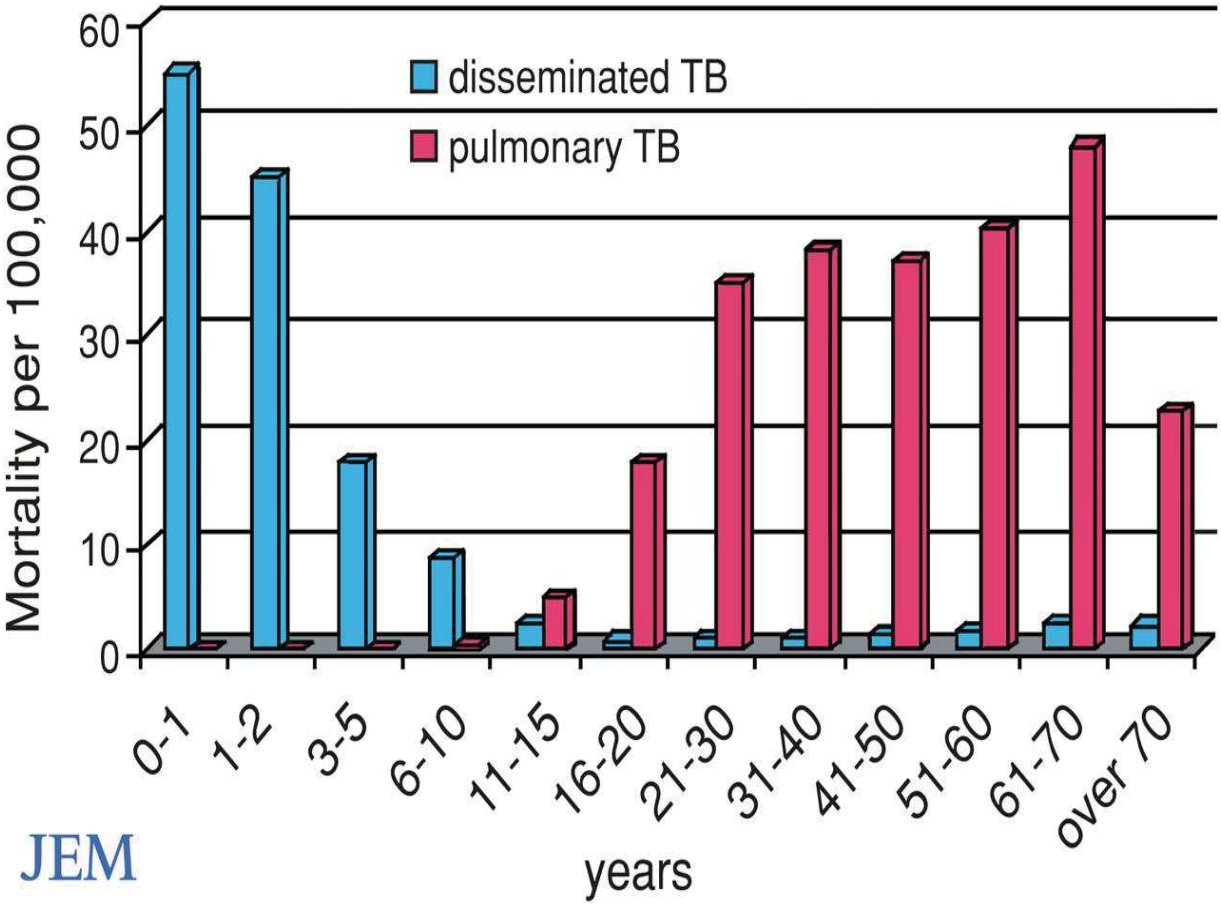
Figure 31: Histoire naturelle de l'infection humaine par M.tuberculosis et développement ultérieur de la tuberculose clinique d'après (206)



Dans une grande enquête réalisée au Pays-Bas, environ 2/3 des patients tuberculeux avaient développé leur maladie au cours des 2 ans ayant suivi leur infection (207), et dans les pays très endémiques pour la tuberculose, cette progression rapide se voit sur le nombre de cas de tuberculose chez les enfants, dont certains développent des formes disséminées de la maladie que l'on qualifie de tuberculose sévère (121). En reprenant les données historiques du début du siècle quand aucun traitement préventif ni curatif n'existait encore pour la tuberculose, on peut distinguer 2 pics d'incidence distincts pour la tuberculose sévère et pour la tuberculose pulmonaire (Figure 32), qui diffèrent non seulement dans leur présentation clinique mais également dans leur âge moyen de survenue. Il paraît donc intéressant d'étudier ces 2 formes de la maladie comme 2 maladies distinctes d'un point de vue génétique pour être en mesure de mieux comprendre leurs particularités.

Figure 32 : Distribution des taux de mortalité par tuberculose disséminée (en bleu) et par tuberculose pulmonaire (en rouge) pour 100 000 personnes non soignées vivant en Bavière en 1905, avant l'apparition du vaccin BCG, région endémique pour la tuberculose.

Figure reprise de (208)



L'existence d'une composante génétique dans la survenue d'une tuberculose clinique est supposée depuis plus d'un siècle. L'accident qui a eu lieu à Lübeck en Allemagne entre 1929 et 1930 illustre de manière malheureuse cette hypothèse. A cette époque, 251 nouveaux-nés ont été vaccinés avec une préparation de BCG malencontreusement contaminée par une souche virulente de *M.tuberculosis* (209). Parmi les enfants vaccinés, 72 moururent de la tuberculose, 23 ne déclarèrent aucun signe clinique de tuberculose et les autres survivants présentèrent une palette très large de symptômes cliniques. La première source de variabilité des présentations cliniques observées reposait sur la dose infectieuse injectée, du fait d'une contamination non uniforme des lots de vaccins. Cependant, même après avoir tenu compte de la dose de bactérie virulente injectée, la variabilité clinique résiduelle demeurait élevée, allant du décès de l'enfant à des symptômes très légers. En l'absence d'autres facteurs sociaux ou environnementaux connus, cette variabilité pourrait être due à des différences génétiques entre les nouveau-nés quant à leur capacité innée à combattre *M.tuberculosis*.

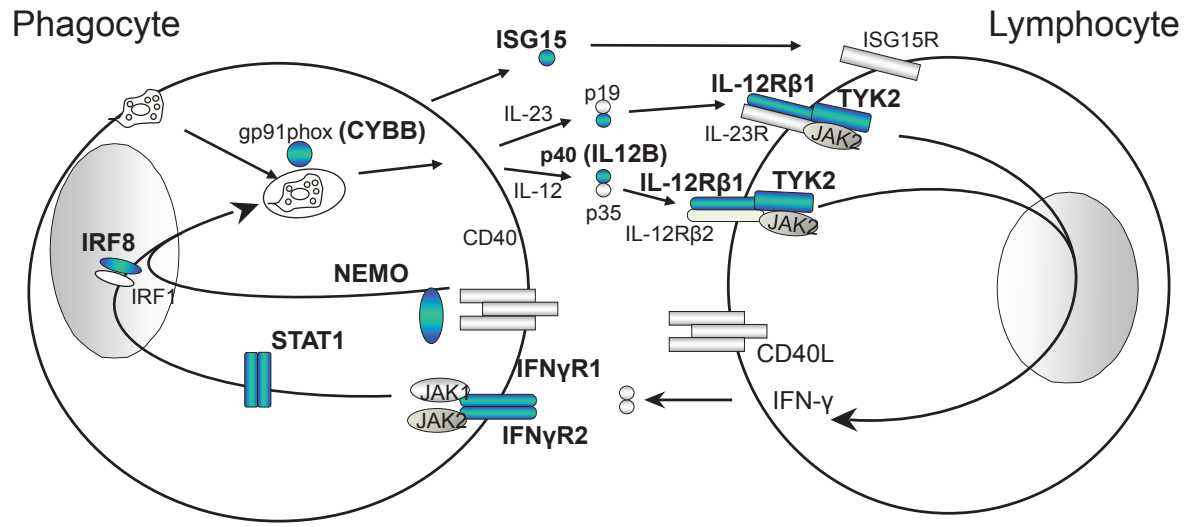
La base génétique de la résistance à la tuberculose s'appuie aussi sur le fait que des populations n'ayant jamais été ou très peu exposées à la mycobactérie sont beaucoup plus susceptibles à la maladie que les populations qui y sont exposées. Pour preuve, la tuberculose importée au Canada par les colons européens à la fin du XIXème siècle a causé la mort de très nombreux indiens à Qu'appelle dans la province de Saskatchewan, avec plus de la moitié des familles décimées sur les 3 premières générations concernées par l'épidémie, avant de constater une forte diminution du taux de mortalité, pouvant être vue comme la conséquence d'une très forte sélection négative vis-à-vis de gènes de susceptibilité (210). De la même manière, à force de contacts avec les explorateurs, les missionnaires et les baleiniers européens, la population Inuit du Nunavut au Canada a succombé à de plus en plus de maladies, y compris la tuberculose (211). Des études de jumeaux ont également souligné que la concordance de progression vers une tuberculose clinique était plus élevée chez les jumeaux monozygotes que chez les jumeaux dizygotes, les conditions environnementales et sociales étant égales par ailleurs. En étudiant 308 familles, Kallman et Reisner ont rapporté dès 1943 que les jumeaux monozygotes étaient concordants à 69.2% vis-à-vis de la tuberculose contre 26.3% pour des jumeaux dizygotes, lorsque les 2 frères ou sœurs avaient été exposés de manière avérée à *M.tuberculosis* (212,213). Une autre étude de 1978 ré-analysant des données du Prophit survey montre qu'après ajustement sur des facteurs de risque comme l'âge, le sexe ou la présence de bacilles dans les sécrétions des cas index, le taux de tuberculose était 2 fois plus élevé chez les individus ayant un jumeau monozygote atteint que chez ceux dont le jumeau atteint était dizygote (213). Au vu de ces diverses

observations, l'idée d'une composante génétique dans la susceptibilité à la tuberculose ne semble plus une simple hypothèse mais une réalité à investiguer.

La première preuve moléculaire que la tuberculose pouvait être due à une prédisposition génétique de type mendélien est venue de l'observation d'une tuberculose sévère chez des enfants souffrant d'immunodéficience primaire (35,214). Les plus gros progrès ont été réalisés grâce à l'étude du syndrome de susceptibilité mendélienne aux mycobactéries (MSMD), qui est défini par une vulnérabilité spécifique aux espèces mycobactériennes non tuberculeuses peu virulentes comme le bacille du BCG et les mycobactéries environnementales (33,215). Depuis 1996, des mutations germinales ont été trouvées dans 8 gènes autosomiques (*IFNGR1, IFNGR2, IL12B, IL12RB1, STAT1, IRF8, ISG15, TYK2*) et 2 gènes liés à l'X (*NEMO* et *CYBB*) chez des patients souffrant de MSMD (33,34,215). Ces défauts génétiques ne sont pas étrangers les uns aux autres car ils affectent tous l'immunité liée à l'IFN-γ (Figure 33).

Figure 33 : Schéma de la coopération entre les phagocytes et les lymphocytes T (ou NK) lors d'une infection mycobactérienne.

Les molécules en bleu sont mutées chez les patients souffrant de MSMD




De l'observation de tuberculose chez certains de ces patients est venue l'idée que leur tuberculose pouvait être également le résultat d'une prédisposition génétique de type monogénique (35), en particulier lorsqu'au sein d'une fratrie partageant le même défaut

génétique, le seul phénotype infectieux ségrégeant avec la mutation était une tuberculose sévère (216,217). Le défaut génétique le plus commun identifié chez des patients atteints de tuberculose sévère aujourd'hui est un déficit complet de la chaîne β1 du récepteur à l'interleukine-12 codée par le gène *IL12Rβ1*, observée dans plusieurs familles (218,219). Ces résultats valident le concept de prédisposition monogénique à la tuberculose sévère, et soulèvent l'hypothèse qu'une proportion importante de cas pourrait être expliquée par des défauts génétiques de l'immunité. Cette proportion a même été estimée jusqu'à 45% par des calculs théoriques (208).

Comme nous l'avons dit au début de cette seconde partie, la tuberculose pulmonaire de l'adulte pourrait être considérée comme une maladie différente de la tuberculose disséminée de l'enfant, et jusqu'à présent, aucune mutation dans les gènes identifiés pour la tuberculose sévère n'a été mise en évidence dans des cas de tuberculose pulmonaire de l'adulte. La quête de ces facteurs génétiques s'est en effet révélée plus difficile que prévue. La plupart des études d'association génétiques classiques concernant la tuberculose pulmonaire se sont d'abord concentrées sur des gènes candidats, et plusieurs variants à risque ayant des fréquences relativement élevées dans la population générale ont été rapportés dans des gènes de l'immunité tels que ceux codant pour DC-SIGN, les récepteurs de type Toll 1 et 2, le récepteur de la vitamine D, TNF, l'interleukine-1β, le facteur de transcription STAT4 ou d'autres molécules du système HLA de classe 2 (220–222). Cependant, ces découvertes n'ont été que rarement confirmées par des études indépendantes, en raison d'un manque de puissance desdites études, d'une grande hétérogénéité des contextes dans lesquels elles ont été effectuées et dans la définition de leur phénotype (220,223). L'un des résultats les plus convaincants issus de ces analyses de gènes candidats concerne le gène *NRAMP1* au sein duquel plusieurs polymorphismes (variants trouvés dans la population humaine à des fréquences élevées) ont été retrouvés associés à la tuberculose dans diverses populations avec des odds-ratio (OR) évalués par méta-analyse allant de 1.2 à 1.35 (224,225).

On peut noter cependant qu'une analyse de liaison réalisée au Brésil sur 164 familles comportant plusieurs membres atteints de tuberculose ou de lèpre a conduit à l'identification d'un signal de liaison conjoint aux 2 infections mycoactériennes, sur le chromosome 17 autour de 40 Mb (226), en partant d'une région candidate identifiée chez la souris. Une région très proche de celle-ci dans les 30 premières mégabases du chromosome 17 a également été identifiée comme évocatrice de liaison par une analyse pangénomique dans un groupe de 32 familles Thaïe où l'âge de déclaration de la tuberculose se situait entre 12 et 24 ans (227).

Dans cette même étude, un second signal de liaison a été identifié sur le chromosome 20 autour de 10Mb pour un ensemble de 30 familles avec un âge de survenue de la tuberculose inférieur à 23 ans. Une autre analyse de liaison pangénomique a été réalisée au Maroc et a mis en évidence un locus majeur sur le chromosome 8 conférant une prédisposition à la tuberculose pulmonaire (176). Cette région a ensuite été cartographiée par clonage positionnel, et c'est en considérant l'âge de survenue de la maladie que des variants du gène *TOX* ont été identifiés comme fortement associés au développement d'une tuberculose pulmonaire avant l'âge de 25 ans dans des populations du Maroc et de Madagascar (OR = 3.09) (177). *TOX* code pour un facteur nucléaire impliqué dans le développement des lymphocytes T, en particulier les lymphocytes T-CD4+ qui sont essentiels dans la lutte contre les mycobactéries (228).

L'avènement des GWAS (Genome Wide Association Studies) a suscité beaucoup d'espoirs dans la communauté scientifique, en particulier dans la recherche de facteurs explicatifs du développement d'une tuberculose pulmonaire. De nombreuses études ont été menées, mais leurs résultats restent globalement assez décevants. Des études au Ghana et en Gambie (229,230), et une autre en Russie (231) n'ont conduit qu'à l'identification de 3 signaux significatifs à l'échelle du génome ($p < 5.10^{-8}$). L'un des 2 variants identifiés en Afrique, rs4331426, est situé dans un désert de gènes du chromosome 18q11.2 (230), et le second, rs2057178, est situé près du gène *WT1* sur le chromosome 11p13 (229). Les 2 variants ont des tailles d'effet relativement modestes (OR=1.19 pour rs4331426 et OR=0.77 pour rs2057178). Le signal du chromosome 11 a été répliqué en Indonésie et en Russie dans l'étude originale, puis en Afrique du Sud (232) et au Maroc (233) dans des études indépendantes. Les tentatives de réplication du signal du chromosome 18 n'ont pas toutes abouties à des résultats probants, en particulier en Chine (234,235), bien qu'au Maroc, le signal semble répliqué (233). L'étude russe a pu identifier un groupe de variants introniques du gène *ASAP1* dans une population de plus de 15000 participants avec une taille d'effet relativement faible (OR=0.84 pour le SNP rs4733781) et un rôle potentiel dans la mobilité des cellules dendritiques (231). Une grande étude d'association pangénomiques réalisée en Islande a également rapporté des signaux dans la région HLA de classe II (236), mais le rôle précis de ces variants dans le processus d'infection tuberculeuse et/ou dans le développement d'une tuberculose pulmonaire clinique reste à définir. Afin d'affiner le phénotype étudié et d'augmenter la probabilité d'homogénéité génétique des individus, plusieurs approches ont été essayées. En Asie, sur des populations thaïe et japonaise, une stratification sur l'âge de survenue de la maladie en 2 groupes (plus ou moins de 45 ans) a permis d'identifier un signal d'association sur le chromosome 20q12 chez

les individus les plus jeunes (rs6071980, OR=1.73) (227), tout près de la région de liaison identifiée par les mêmes auteurs en 2009 et évoquée au paragraphe précédent (237). Enfin, un GWAS sur 581 patients infectés par le virus du VIH (dont 267 avec une tuberculose active) venant d'Ouganda et de Tanzanie a identifié le variant rs4921437 sur le chromosome 5q33, proche du gène *IL12B*, avec un effet protecteur important (OR=0.37) (238), beaucoup plus important que les effets des SNPs rapportés dans des populations non infectées par le VIH.

Dans toutes ces études à l'échelle du génome, l'absence de réplication des facteurs de susceptibilité génétique identifiés par les études centrées sur des gènes candidats est surprenante. Une explication possible serait que les variants communs, qui sont ceux étudiés dans les études d'association telles qu'on les connaît jusqu'à présent, auraient un impact limité sur les prédispositions individuelles à la tuberculose pulmonaire de l'adulte, tout au moins lorsqu'on étudie cette maladie comme un phénotype unique et homogène. On peut émettre plusieurs hypothèses pour expliquer ce phénomène d' « héritabilité manquante » constaté pour la tuberculose pulmonaire comme pour de nombreuses autres maladies dites complexes (239). Parmi elles, on peut citer le rôle de l'épigénétique régulant l'expression des gènes, les interactions entre différents variants génétiques affectant différents gènes du même complexe protéique ou de la même voie de signalisation, l'hétérogénéité des phénotypes étudiés, etc. Nous nous sommes attachés à explorer deux pistes simultanément. Nous avons choisi d'une part de considérer un phénotype extrême en nous intéressant à des patients marocains jeunes (la majorité ayant moins de 30 ans) déclarant une tuberculose pulmonaire causée par *M.tuberculosis*, et pour une partie d'entre eux, appartenant à des familles présentant plusieurs cas de tuberculose. Nous les avons comparés à des marocains en moyenne plus âgés, infectés par la tuberculose mais n'ayant aucun symptôme clinique. D'autre part, nous avons souhaité tester l'hypothèse que la tuberculose pulmonaire pouvait être due à des variants génétiques dits rares, c'est-à-dire de fréquence inférieure à 5% dans notre échantillon d'étude, avec des effets beaucoup plus forts que ceux trouvés jusque-là, et ne pouvant pas être étudiés simplement par des analyses d'association classiques. Pour cela, nous avons travaillé sur des données issues du séquençage de l'exome de ces individus, et c'est ce que je vais détailler dans cette seconde partie du manuscrit.

# B. Matériel et Méthodes

## 1. Echantillon d'étude

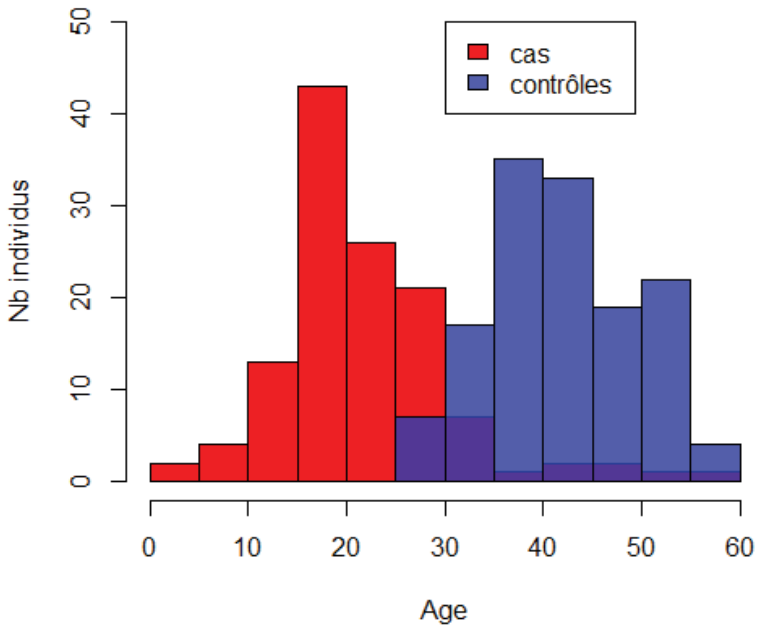
Les sujets d'étude ont été recrutés à l'hôpital Mohammed V de Rabat et dans les centres de diagnostic de tuberculose situés dans les régions endémiques de Casablanca et de Salé au Maroc, où le taux d'incidence annuel de la tuberculose est d'approximativement de 100 cas pour 100 000 habitants (240). Dans cette étude, seuls les cas présentant un diagnostic de tuberculose pulmonaire ont été retenus, diagnostic posé d'une part sur la base de symptômes cliniques et de signes pathologiques à l'examen de la radiographie des poumons, et d'autre part sur l'identification de *M.tuberculosis* dans les crachats des patients, à l'examen direct par microscopie après coloration de Ziehl-Neelsen et/ou par culture sur un milieu de Lowenstein-Jensen. Nous avons privilégié l'inclusion de patients jeunes et/ou souffrant de formes familiales de tuberculose en vue de mettre l'accent sur la composante génétique du développement de la maladie des individus étudiés. Nous avons comparé ces patients à des individus (dits contrôles) recrutés parmi des volontaires sains, ayant été en contact avec un cas de tuberculose au sein de leur famille ou faisant partie du personnel hospitalier. N'ont été retenus que les plus âgés d'entre eux ayant un examen clinique normal, sans historique de tuberculose ni d'autre maladie pulmonaire chronique, et répondant positivement au test de Mantoux (taille d'induration $\geq$ 10 mm) et au test Quantiféron, pour s'assurer de leur exposition et infection par *M.tuberculosis*. En augmentant l'âge des contrôles, nous souhaitions diminuer la probabilité que ceux-ci ne développent une tuberculose clinique après notre étude, ce qui fausserait leur statut de contrôle.

Un total de 120 patients de moyenne d'âge 21.8 ans et de 136 contrôles d'âge moyen 42.5 ans ont ainsi été sélectionnés (Figure 34). La plupart d'entre eux faisaient déjà partie des échantillons utilisés dans les précédentes études de tuberculose pulmonaire du laboratoire (177,233). Les procédures d'inclusion de tous les individus ont été approuvées par les comités d'éthiques ad hoc et des consentements écrits ont été obtenus pour chaque participant (via les parents lorsqu'il s'agissait d'enfants mineurs).

Afin de nous assurer que les individus étudiés étaient tous non apparentés et qu'aucun n'était trop différent des autres en terme d'origine ethnique, nous avons calculé le taux d'IBS entre les 256 individus séquencés et réalisé une analyse en composantes principales (ACP) de notre échantillon conjointement avec les individus de 1000 génomes phase 1, à l'aide du

logiciel PLINK (164) : pour l'ACP, nous avons utilisé 106 599 variants de MAF > 2% et sans aucune valeur manquante chez les 256 individus de l'échantillon et chez les 1092 individus du projet 1000 génomes Phase 1 (181). Pour le calcul de la matrice IBS des 256 individus de l'échantillon, nous avons utilisé tous les variants disponibles pour les individus du projet.

Figure 34 : Distribution de l'âge des 256 individus retenus pour l'étude en fonction de leur statut vis-à-vis de la tuberculose pulmonaire.




## 2. Séquençage des individus

### 1. Alignement et détection des variants

Tous les échantillons d'ADN ont été fragmentés et capturés avec le kit de capture d'exons 71 Mb SureSelect V4 + UTRs de la société Agilent Technologies, puis séquencés au New York Genome Center sur un séquenceur Illumina Hiseq 2000. Les fragments d'ADN ont été séquencés en « paired end », c'est-à-dire que les 2 extrémités du fragment étaient séquencées simultanément, et la longueur de lecture était d'environ 100 paires de bases. Tous les échantillons n'ont cependant pas été séquencés à la même période : 75 patients ont été

séquencés en octobre 2013, puis 94 contrôles en octobre 2015 et enfin 45 cas et 42 contrôles en mai 2016.

L'alignement des données de séquençage sur le génome humain de référence hg19 a été réalisé avec la version 0.7.12 du logiciel BWA (Burrows-Wheeler Aligner), en utilisant l'algorithme BWA-MEM optimisé pour les séquences de plus de 70 bp (241). Ces données de lecture alignées sur le génome humain et stockées sous forme de fichiers binaires, communément appelés fichiers BAM, ont ensuite été traitées en suivant les recommandations émises par les auteurs du logiciel GATK (Genome Analysis Toolkit) (242), et disponibles sur leur site (https://software.broadinstitute.org/gatk/best-practices/). Les « reads » identifiés comme duplicats (duplicats moléculaires de PCR ou duplicats optiques) ont ensuite été éliminés grâce au logiciel Picard (http://broadinstitute.github.io/picard), puis une seconde étape a ajusté localement l'alignement autours des petites insertions out délétions (indels) à l'aide de GATK, et enfin une dernière étape a permis de corriger les artefacts de séquençage en appliquant l'outil « Base Quality Score Recalibrator » de GATK. La détection des variants a été réalisée avec l'outil HaplotypeCaller de GATK pour chaque individu séparément, au sein des régions couvertes par le kit de capture auxquelles nous avons rajouté des marges de 200 paires de base en amont et en aval. Les données ont été ainsi générées dans le format de fichier gVCF (« Genome Variant Call Format ») permettant de conserver les informations de qualité de chaque position couverte, puis les gVCF de tous les individus ont été regroupés dans un seul fichier de format VCF (« Variant Call Format ») utilisé couramment pour la manipulation des données de séquençage. Dans ce dernier figurent les génotypes de chaque individu inclus dans l'étude pour chaque variant génétique identifié. Il est important de souligner que dans le cas où un génotype n'a pu être déterminé pour un (ou plusieurs) individu(s) séquencé(s) et que tous les autres individus portent un génotype homozygote sauvage, la position apparaît tout de même dans le fichier avec un génotype codé ./. quand il est manquant, permettant d'être sûr que toutes les positions absentes du fichier mais faisant partie du kit de capture ne comportent effectivement aucun variant pour aucun individu de l'étude.

### 2. Contrôle qualité

Une fois le fichier VCF global généré pour tous nos individus, nous avons réalisé un contrôle qualité des données de séquençage afin d'éliminer les variants de mauvaise qualité. En effet, les erreurs de génotypage sont relativement fréquentes à l'issue du séquençage d'exome ou WES (*Whole Exome Sequencing*) (243,244). Il est nécessaire de considérer l'existence de ces

erreurs potentielles puisqu'elles peuvent affecter les analyses d'associations en aval. Les erreurs non différentielles (pour lesquelles le taux d'erreur est identique entre les cas et les contrôles) ne perturbent pas l'erreur de type I des tests d'association mais diminuent considérablement la puissance statistique (243).

Notre contrôle qualité a comporté 2 principales étapes :

- Un filtre global appliqué au niveau des variants : il s'agissait ici d'éliminer les variants de mauvaise qualité sur l'ensemble de l'échantillon à l'aide de l'outil VQSR (« *Variant Quality Score Recalibration* ») de GATK. Celui-ci se base sur la distribution d'un ensemble de critères qualité et d'annotations disponibles à l'issue du génotypage des variants par HaplotypeCaller pour des variants supposés « vrais » car référencés dans des bases de données publiques de SNPs validés. Dans notre cas les bases de données utilisées étaient celles de dbSNP (245), des projets HapMap (phase 3 version 3) et 1000 Génomes (version 2.5). A partir de ces variants dits « vrais », l'outil génère un modèle multivarié auquel il compare l'ensemble des critères de chaque variant restant (ceux non répertoriés dans les bases de données publiques utilisées) afin de calculer la probabilité que chacun d'entre eux soit bien un variant réel et non un artefact de séquençage. VQSR permet ainsi d'éliminer les variants de mauvaise qualité. Comme conseillé dans le guide de bonnes pratiques de GATK, nous avons utilisé un seuil de sensibilité de détection des variants dits « vrais » de 99%. Nous avons également exclu tous les variants mono-alléliques (n'apportant aucune information dans notre étude) et les variants multi-alléliques (par souci de simplicité d'analyse).
- Un filtre appliqué au niveau du génotype de chaque variant afin de recoder les génotypes de mauvaise qualité en données manquantes. Ce filtre était basé sur des métriques de qualité du génotype telles que la profondeur de lecture (DP, correspondant au nombre de « reads » utilisés pour identifier le génotype à une position donnée), la qualité du génotype (GQ, valeur calibrée sur le score Phred représentant la probabilité que le génotype identifié à un site précis soit le vrai génotype) et le ratio du nombre de « reads » portant l'allèle le moins fréquent sur le nombre total de « reads » couvrant cette même position (« minor read ratio » noté MRR). Dans notre cas, nous avons considéré comme manquants les génotypes avec DP < 8, GQ < 20 (ce qui correspond à une probabilité supérieure à 1% d'avoir un génotype faux à ce site) ou MRR > 0.2. Il a été montré que l'application de ces filtres

améliorait substantiellement la qualité des variants et diminuait le nombre de faux-positifs (246).

- Dans un troisième temps, nous avons également appliqué un filtre sur le taux de génotypage des variants dans notre échantillon, afin de ne conserver dans les analyses que les variants comportant moins de 5% de génotypes manquants dans l'ensemble de notre échantillon d'étude.

Une attention particulière a été portée au suivi de l'évolution de la qualité des données aux cours des différentes étapes du contrôle qualité. Afin d'évaluer le gain de qualité, nous avons mesuré le ratio Ti/Tv à la fin de chaque étape. Ti correspond au nombre de transitions (mutations allant d'une purine vers une purine ou d'une pyrimidine vers une pyrimidine) alors que Ts représente le nombre de transversions (mutations allant d'une purine vers une pyrimidine et inversement). Plus le ratio Ti/Tv est élevé plus le taux de variants faussement positifs est faible (243). La valeur attendue de ce ratio est supérieure à 2.8 pour les exomes contre 2.1 pour le génome entier (242,247). Cette différence serait due au fait que les transversions induiraient des modifications plus conséquentes au niveau des protéines que les transitions (248).

### 3. Annotation des variants

L'identifiant de chaque variant a été actualisé en utilisant la base de données publique dbSNP (build 138) développée par le « *National Center for Biotechnology Information* » (NCBI) (245), qui recense la majorité des variations génétiques connues telles que les SNPs, les petites insertions/délétions, les marqueurs microsatellites, les séquences répétées en tandem et les MNPs (« Multinucleotide polymorphisms »). Les variants absents de cette base de données ont été considérés comme de nouveaux variants. Tous les variants ont ensuite été annotés à l'aide du logiciel snpEff (249) sur la base de leur localisation génomique par rapport à un ensemble de gènes et transcrits de référence du projet Ensembl (version GRCh37.75 - hg19). Ce logiciel permet de caractériser le positionnement du variant dans son contexte génique (exonique, intronique, 5' UTR, 3'UTR, intergénique…) ainsi que les conséquences des variants sur la séquence protéique (i.e. mutations synonymes, faux-sens, non-sens, insertions/délétions avec ou sans décalage du cadre de lecture, création d'un site d'épissage …). Ces annotations sont cruciales dans la définition des variants d'intérêt que l'on souhaite garder pour les analyses d'association réalisées par la suite.

D'autre part, pour prédire l'impact fonctionnel des variants, nous avons également utilisé le score CADD (« *Combined Annotation-Dependent Depletion* ») qui agrège de multiples informations et scores de prédiction dans un modèle de type support vecteur machine (SVM) et fournit un score de prédiction unique par variant génétique (250). Le gros avantage de cette méthode est d'être capable de fournir un score pour tous les types de SNPs du génome humain, y compris ceux qui sont synonymes ou en dehors des régions codantes. C'est actuellement le score le plus utilisé dans la littérature, même s'il ne semble pas être l'outil le plus performant quand il s'agit de prédire l'impact fonctionnel de variants codants non-synonymes (251). Le score CADD des mutations pathogènes connues est très variable d'un gène à l'autre et il peut s'avérer utile de considérer des seuils spécifiques par gène plutôt qu'un seuil universel visant à distinguer les variants délétères des variants sans grand impact fonctionnel. Pour ce faire, nous avons utilisé le MSC (« *Mutation Significance Cut-off* ») défini pour un gène donné comme la limite inférieure de l'intervalle de confiance à 99% du score CADD de toutes ses mutations pathogènes connues (252). Ainsi les variants avec un score CADD supérieur au MSC ont plus de chance d'entrainer un phénotype que les variants ayant un CADD inférieur au MSC.

## 3. Méthodes statistiques d'analyse des variants

Pour tester l'association des variants communs (MAF $\geq$ 5%) de notre échantillon avec le développement d'une tuberculose pulmonaire, nous avons utilisé l'approche classique consistant à comparer les fréquences de l'allèle mineur (ou des différents génotypes) chez les cas et chez les contrôles avec une régression logistique. Nous avons utilisé les méthodes implémentées dans le logiciel PLINK (163,164) et testé plusieurs modèles différents : un modèle allélique, un modèle dominant et un modèle récessif par un test du Chi2 à 1 degré de liberté (ddl), un modèle génotypique par un test du Chi2 à 2 ddl et un modèle additif par un test de tendance de Cochran-Armitage.

En revanche, cette stratégie n'est plus aussi pertinente en présence de variants dits rares (MAF < 5%). En effet, puisque les variants rares sont par définition présents à des fréquences très faibles dans la population, la puissance statistique de détection d'une association entre ces derniers et la maladie est elle aussi très faible à moins d'avoir une taille d'échantillon elle-même très grande (253,254). Afin de contourner ce problème majeur, de nombreuses méthodes statistiques ont été mises au point et publiées (255–260). Les plus simples d'entre

elles appelées « collapsing methods » consistent à agréger les variants rares au sein d'une unité (une unité étant le plus souvent définie comme un gène) de telle sorte que, même si les variants sont individuellement rares, ils peuvent devenir conjointement suffisamment fréquents pour être utilisés dans un test univarié. Dans notre analyse, nous avons choisi d'utiliser le « *Cohort Allelic Sum Test* » (CAST) (255) qui compare, au sein de l'unité de test (définie comme le gène), le nombre de porteurs d'au moins un variant chez les cas et chez les contrôles. Nous avons choisi cette méthode pour la simplicité de son interprétation d'une part, et d'autre part parce que supposant que le développement d'une tuberculose pulmonaire (concernant 5 à 10% de la population infectée) reposait plutôt sur un excès de variants à risque chez les cas, elle fait partie des méthodes les plus puissantes à condition de filtrer au préalable le maximum de variants neutres pour le phénotype d'intérêt (259,261).

Afin de tester la significativité de l'association entre les gènes porteurs de variants rares et le statut cas/témoin, la méthode originale proposait de réaliser un test du chi 2 ou de Fisher sur le tableau 2x2 suivant :

| | **Cas** | **Contrôles** |
|---|---|---|
| Nombre d'individus avec au moins 1 variant rare | N1 | N3 |
| Nombre d'individus sans aucun variant rare | N2 | N4 |

Pour permettre la prise en compte éventuelle de covariables, nous avons réalisé le même test d'association en utilisant un modèle de régression logistique : pour chaque individu et chaque gène, un score génétique est déterminé. Ainsi, pour le $i^{ème}$ gène et le $j^{ième}$ individu, le score génétique $X_{ij}$ vaut :

$$X_{ij} = \begin{cases} 1 & \text{Si au moins un variant rare est présent au sein de l'unité de test} \\ 0 & \text{Sinon} \end{cases}$$

Soit $Y_j$ le phénotype de l'individu $j$, $X_j$ le score génétique de l'individu j, $M_j$ la matrice des covariables et $\beta_0$, $\beta_1$ et $\beta_2$ les coefficients de régression, le modèle de régression s'écrit alors :

$$logit\left(P(Y_j = 1)\right) = ln\left(\frac{P(Y_j = 1)}{1 - P(Y_j = 1)}\right) = \beta_0 + \beta_1 X_j + \beta_2 M_j$$

Les valeurs de p ont ensuite été obtenues par un test de rapport de vraisemblance (LRT) consistant à comparer la vraisemblance des données observées entre le modèle saturé (incluant l'effet des covariables et du score génétique X) et le modèle nul (incluant uniquement l'effet des covariables).

Tous les tests ont été réalisés en unilatéral, c'est-à-dire que nous avons recherché un enrichissement en variants rares de susceptibilité à la tuberculose pulmonaire chez les cas par rapport aux contrôles. Les tests statistiques ont été effectués sous trois modèles de transmission différents : dominant (variant rare présent à l'état au moins hétérozygote), récessif (variant rare présent à l'état homozygote) et un modèle surnommé « bi-variants » (au moins 2 variants rares au sein du même gène), ce dernier modèle pouvant correspondre à des haplotypes géniques, à d'éventuels hétérozygotes composites ou bien des interactions entre variants du même gène.

Le modèle statistique utilisé nécessitait de définir un seuil de fréquence afin de déterminer les variants à inclure dans l'analyse. Autrement dit, il s'agissait de définir quels étaient les variants considérés comme rares parmi tous les variants issus du séquençage. Cette définition étant quelque peu arbitraire, nous avons choisi d'appliquer un seuil de MAF de 5% (calculée sur l'ensemble de notre cohorte) et d'exclure de l'analyse tous les variants dont la MAF excédait cette valeur seuil.

Nous avons également émis des hypothèses supplémentaires sur les variants rares qui nous intéressaient, basées sur l'annotation fonctionnelle donnée par SnpEff, sur la fréquence des variants dans les bases de données publiques ExAC « Exome Aggregation Consortium » (262) ou 1000 génomes (181), ainsi que sur l'impact prédit (estimé par le CADD). En effet, la puissance des méthodes d'analyse par gène telles que CAST dépend fortement de la proportion de variants ayant un effet sur le phénotype étudié parmi les variants inclus dans l'analyse (263) . En enrichissant les groupes de variants testés en variants potentiellement pathogènes, nous augmentons la puissance de nos analyses. Dans le cas d'une maladie dite complexe comme la tuberculose pulmonaire, il n'est pas évident d'inférer quel type de variant aura un effet sur le phénotype étudié. Pour cette raison, nous avons défini plusieurs jeux de variants différents sur lesquels effectuer nos tests d'association, avec en particulier les 2 jeux détaillés ci-dessous qui sont basés sur les annotations de snpEff :

- Un jeu de données noté HIGH IMPACT regroupant uniquement les variants annotés perte de fonction, c'est-à-dire les variants introduisant un codon STOP avant la fin de la séquence du gène, les variants faisant perdre le codon STOP du gène, les variants faisant perdre le codon START en début de gène, les variants décalant le cadre de lecture du gène ainsi que les variants positionnés sur des sites donneurs ou accepteurs d'épissage.
- Un jeu de données noté PRIORITY1 regroupant tous les variants HIGH IMPACT, complétés des variants non-synonymes et des insertions/délétions ne décalant pas le cadre de lecture mais modifiant un codon existant du gène.

A partir de ces 2 jeux de données, nous en avons créé 2 autres en rajoutant un critère sur le score CADD qui devait être supérieur au MSC pour que le variant soit inclus dans le groupe de variants à analyser.

Afin de prendre en compte la problématique de multiplicité des tests pour chaque jeu de données analysé, nous avons défini le seuil de significativité statistique par la méthode de correction de Bonferroni qui consiste à diviser le niveau d'erreur de type I souhaité (5%) par le nombre de gènes informatifs analysés. Nous avons considéré comme informatifs les gènes pour lesquels le nombre de porteurs de variants à analyser était supérieur ou égal à 3 et avec une proportion de porteurs de variants rares plus grande chez les cas que chez les témoins. Le seuil de significativité varie donc en fonction des groupes de variants et du modèle génétique considéré.

# C. Résultats

## 1. Identification et contrôle qualité des variants

Nous avons analysé et comparé les données capturées avec le kit 71Mb SureSelect V4+UTR (Agilent) et séquencées sur un Illumina HiSeq 2000 de 256 individus marocains, 120 atteints de tuberculose pulmonaire et 136 infectés par *M.tuberculosis*, mais ne présentant pas de symptômes cliniques.

A l'issue de la première étape d'alignement des séquences et de détection des variants, nous avons identifié 1 254 841 variants dont 1 100 596 SNPs pour 256 individus. Nous avons ensuite appliqué les différents filtres qualité détaillés dans le paragraphe méthodes et regardé l'impact de chacun sur le nombre de variants restants et sur la mesure de Ti/Tv. Ces différents résultats sont détaillés dans le tableau 11. Nous retrouvons une valeur de Ti/Tv finale inférieure au 2.8 espéré en raison de la présence des régions UTR dans le kit de séquençage de l'exome et des marges de 200 bp prises en compte de part et d'autre des intervalles de capture.

Tableau 11 : Nombres de variants détectés dans l'échantillon d'étude de 256 individus suivant les filtres appliqués.

| | Nombre total de SNPs | Nombre total d'indels | Nombre de variants privés | Ti/Tv moyen par individu |
|---|---|---|---|---|
| HaplotypeCaller (GATK) | 1 100 596 | 154 245 | 395 118 | 2.25 |
| VQSR (GATK) | 1 036 133* | 135 298* | 367 154 | 2.29 |
| Filtre DP, GQ et MRR | 1 036 133* | 135 298* | 360 826 | 2.31 |
| Valeurs manquantes < 5% | 591 931 | 54 452 | 224 825 | 2.44 |

*Le nombre de SNPs et d'Indels reste inchangé après le filtre DP, GQ et MRR car le filtre n'élimine aucun SNP directement, il met des valeurs manquantes aux génotypes des individus ne satisfaisant pas les critères du filtre.

Bien que tous séquencés au New York Genome Center et sur le même séquenceur, nos individus n'ont pas tous été séquencés en même temps. Nous avons donc cherché à vérifier s'il existait un biais de séquençage. Pour cela, nous avons comparé la distribution du nombre de variants des individus et de leur couverture moyenne en fonction de leur statut et de leur période de séquençage. A l'issue de la détection des variants et de l'application des filtres qualité VQSR et filtre fixe sur DP, GQ et MRR, en considérant l'ensemble des intervalles du kit de capture avec leurs marges de 200 pb en amont et aval, nous avons constaté une différence très nette dans le nombre de variants identifiés entre les premiers cas séquencés et les autres individus (cf Figure 35), à la fois sur les SNPs et sur les indels. Les premiers patients séquencés en 2013 possèdaient presque 30% de SNPs en moins en moyenne que les patients séquencés en 2016 (88 592 SNPs détectés par individu en 2013 versus 125 045 en 2016). Dans un modèle de régression linéaire, le nombre de SNPs détectés était très significativement expliqué par l'année de séquençage ($p < 2.10^{-16}$).

La différence s'est beaucoup estompée en ne gardant que les variants localisés au sein des régions du kit de capture après l'ensemble des filtres qualité (VQSR + filtre fixe sur DP, GQ et MRR). En comparant la médiane de distribution du nombre de SNPs détectés pour les patients séquencés en 2013 avec la même distribution pour les patients séquencés en 2016, la différence n'était plus que d'environ 2% (données non représentées). Dans un modèle de régression linéaire, cette différence restait cependant significative ($p=1.25.10^{-6}$ entre les cas séquencés en 2013 et ceux séquencés en 2016).

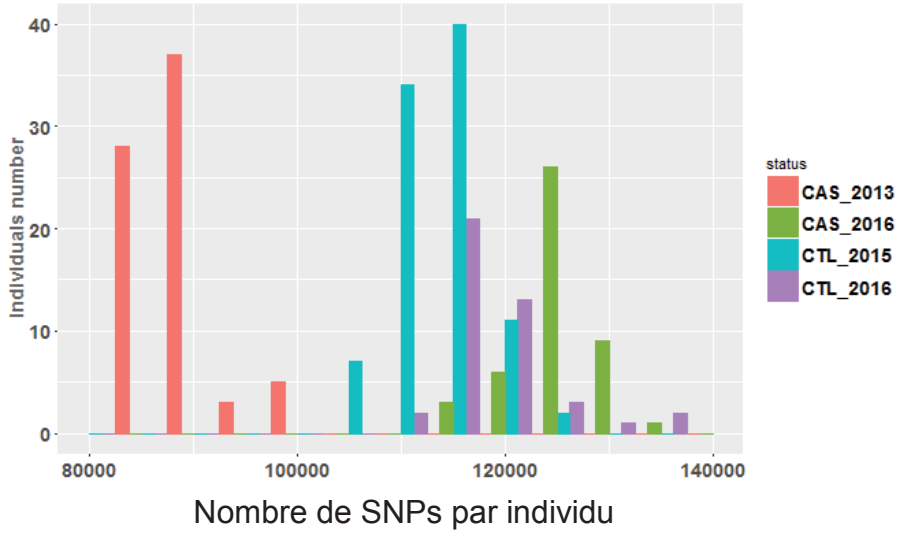
Notre étude s'intéressant avant tout aux variants rares, nous avons regardé si la tendance était la même sur les variants ayant une MAF < 5% dans notre échantillon et une forte probabilité d'avoir un impact fonctionnel au niveau du gène, c'est-à-dire les variants annotés comme non-synonymes, frameshift, disruptive inframe, non-sens, stop gain, start lost, site d'épissage donneur ou accepteur. Nous avons vu que l'écart s'était considérablement réduit en valeur absolue sur le nombre de variants entre les différents lots de séquençage. En comparant là encore les médianes de distribution du nombre de SNPs détectés chez les cas en fonction de leur année de séquençage, cette différence descendait en-dessous de 1%, et dans un modèle de régression linéaire, cette différence n'était plus significative ($p=0.92$ entre les cas séquencés en 2013 et ceux séquencés en 2016) (Figure 36).

Notre hypothèse principale étant la présence d'un excès de variants rares chez les patients atteints de tuberculose pulmonaire par rapport aux contrôles infectés non malades, la différence dans le nombre de variants au profit des contrôles était surtout un facteur de perte

de puissance dans notre étude, mais en aucun cas un facteur d'inflation de l'erreur de type I. A l'opposé, l'hypothèse d'un excès de variants rares chez les contrôles infectés mais non malades n'aurait pu être testée convenablement avec ces données, du fait du biais de séquençage.

Figure 35 : Distribution du nombre de variants (A) et de leur couverture moyenne (B) par individu à l'issue des différents filtres qualité (VQSR + filtre sur DP, GQ et MRR) sur les 256 individus séquencés.

A.




B.

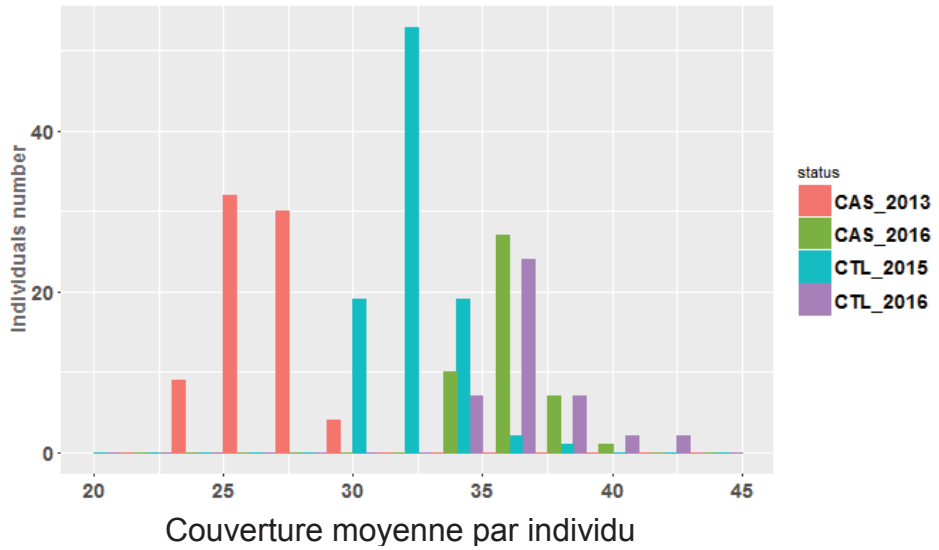
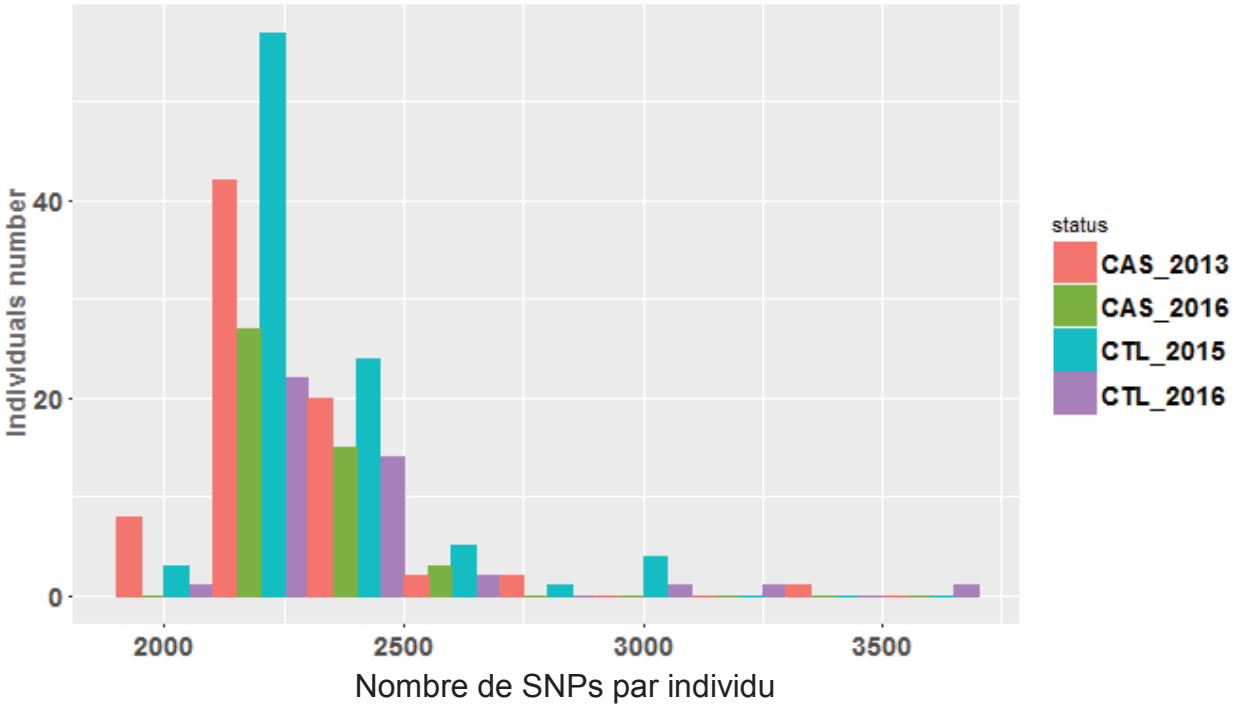
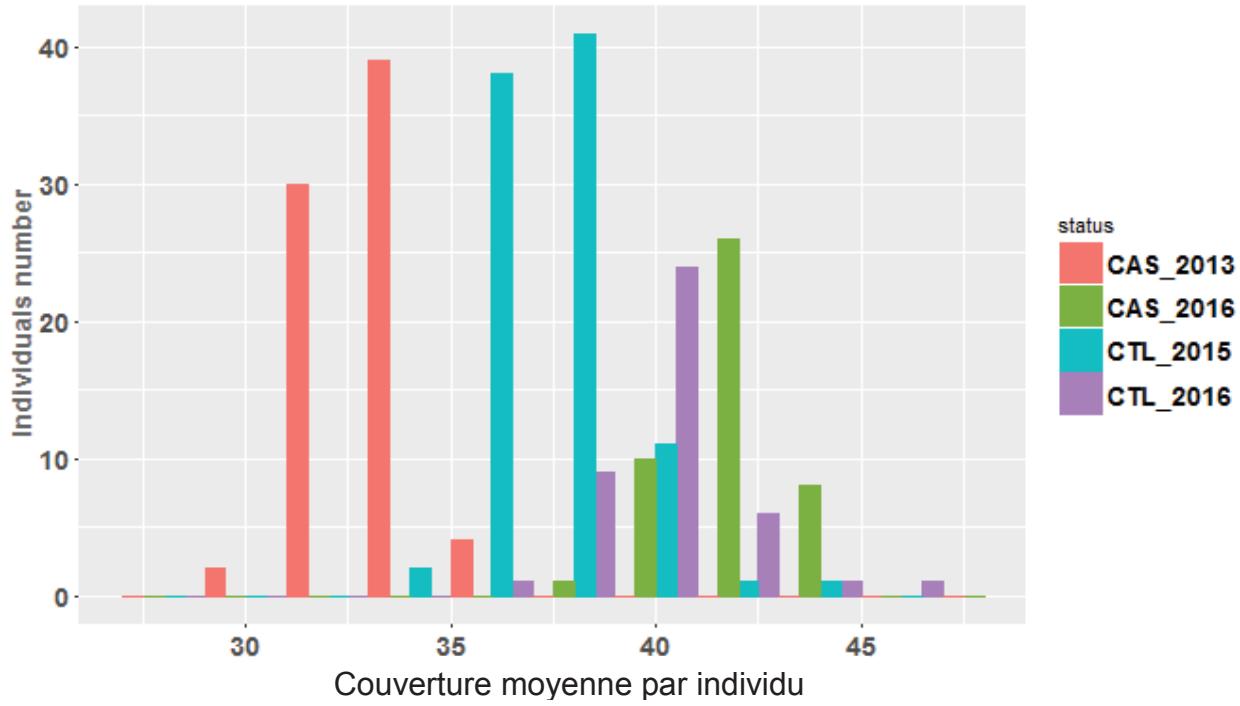


Figure 36 : Distribution du nombre des variants potentiellement délétères **avec une MAF <5%** et de leur couverture moyenne par individu dans l'échantillon d'étude sur les 256 individus séquencés.



## 2. Contrôle qualité des individus

Avant de réaliser les analyses d'association, il est nécessaire de s'assurer qu'il n'existe pas de stratification de population majeure entre les cas et les contrôles de l'échantillon d'étude. Cette problématique est en effet répandue dans les études GWAS testant les variants communs, mais elle existe tout autant voire davantage pour les études de variants rares car ils ont davantage tendance à subir des niveaux modérés de sélection négative, peuvent être apparus récemment et être spécifiques à une population ou à une région géographique particulière (264). Il est également nécessaire de vérifier le non-apparentement des individus, pour éviter tout biais dans les analyses. Pour la stratification de population, nous avons réalisé une analyse en composantes principales de notre échantillon avec les individus du projet 1000Génomes Phase 1. Le groupe d'individus de l'échantillon marocain se situe entre les Caucasiens et les Africains du projet 1000Génomes, comme on pouvait s'y attendre. On constate cependant que quelques individus se rapprochent davantage du groupe des africains que les autres et par souci d'homogénéité nous avons décidé de les exclure de l'échantillon d'étude soit 1 cas et 6 contrôles (Figure 37). L'analyse des taux d'IBS nous a permis de retrouver des individus dupliqués ou apparentés ; sur la base de la figure 38 représentant le taux d'IBS maximal partagé par chaque individu avec les autres sujets de la cohorte, nous avons défini arbitrairement un seuil d'IBS à 0.96 au-delà duquel les individus n'étaient plus considérés comme indépendants. Nous avons privilégié l'exclusion de contrôles dans ces procédures et avons finalement conservé 119 cas et 120 contrôles pour les analyses ultérieures.

La tuberculose ayant une incidence plus élevée chez les hommes que chez les femmes, les patients faisant partie de l'étude reflètent le sexe-ratio généralement observé d'approximativement 2 hommes malades pour 1 femme (1,265). Les contrôles ont été recrutés dans les mêmes proportions, comme indiqué dans le Tableau 12.

Tableau 12 : Répartition des cas et contrôles suivant leur sexe.

| | CAS | CONTROLES |
|---|---|---|
| Hommes | 75 | 78 |
| Femmes | 44 | 42 |
| TOTAL | 119 | 120 |

Figure 37 : Analyse en composantes principales des 256 individus séquencés et des 1092 individus du projet 1000 Génomes phase 1.

AFR = individus d'origine africaine – AS = individus d'origine asiatique –CEU = individus d'origine caucasienne

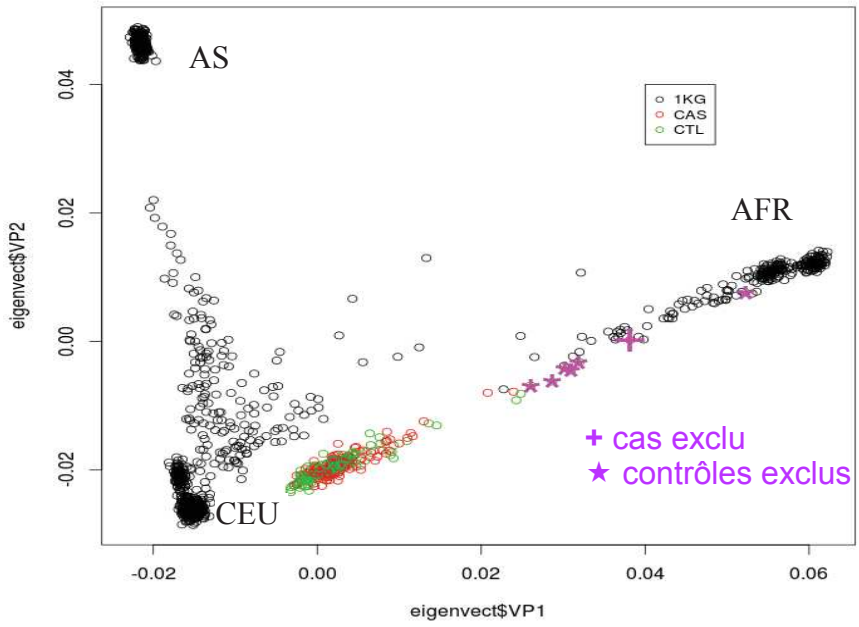
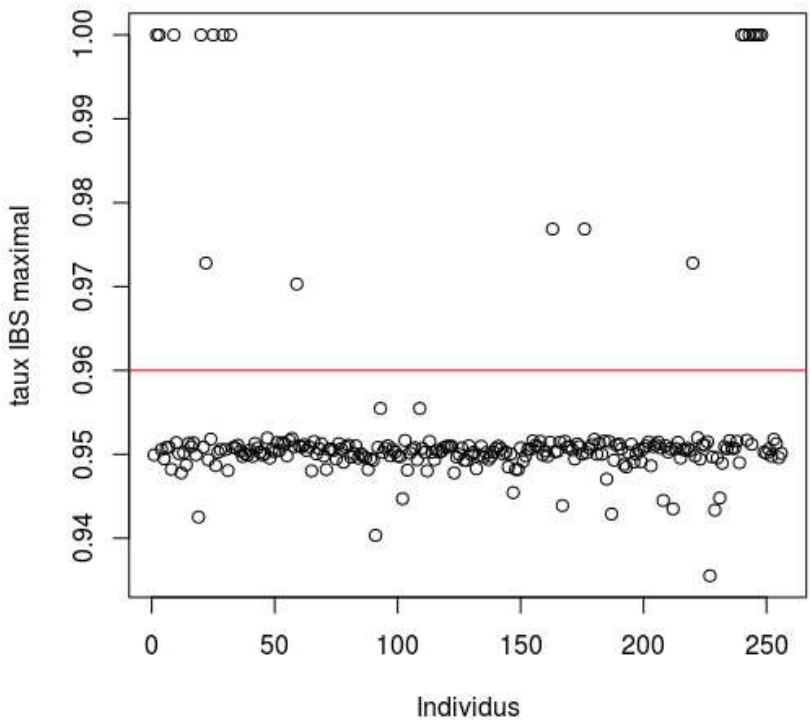


Figure 38 : Taux maximal d'IBS partagé par chaque individu avec les 255 autres sujets séquencés.




## 3. Analyses d'association

### 1. Analyse des variants communs

Nous avons donc analysé et comparé les données d'exomes après contrôle qualité de 119 patients marocains atteints de tuberculose pulmonaire et 120 contrôles marocains également infectés par *M.tuberculosis* mais non malades. Bien que le but principal de l'étude soit d'étudier l'impact des variants rares sur le développement d'une tuberculose pulmonaire, nous avons d'abord voulu éliminer l'hypothèse que des variants plus communs (MAF ≥ 5% dans notre échantillon d'étude) puissent jouer un rôle dans le développement du phénotype d'intérêt. Pour cela, nous avons réalisé une analyse par variant des 137 365 variants bi-alléliques de notre échantillon ayant moins de 5% de valeurs manquantes, une MAF ≥ 5% et en équilibre d'Hardy-Weinberg (au seuil de $p > 10^{-4}$) à l'aide du logiciel PLINK. Le meilleur résultat a été obtenu pour le modèle additif. Le QQplot (graphique quantile-quantile) de la figure 39 montre qu'il n'existait pas de biais systématiques dus à une stratification de population

quelconque, et a mis en lumière 2 variants présentant une valeur p aux alentours de $10^{-6}$. Ces 2 variants rs9906443 et rs739800 sont situés sur le chromosome 17, aux positions 30 185 565 et 30 190 083 respectivement, sont en complet déséquilibre de liaison ($r^2$=1) et la fréquence de leur allèle mineur est de 17% dans l'échantillon étudié. Lorsqu'on regarde la répartition des génotypes en fonction du statut cas/contrôles, c'est l'allèle G qui est associé au développement de la tuberculose (Tableau 13). En modèle additif, l'Odds ratio est estimé à 1.42 pour l'allèle G avec la fonction glm de R.

Tableau 13 : Génotypes des 239 individus étudiés pour rs9906443 en fonction du statut vis-à-vis de la tuberculose pulmonaire

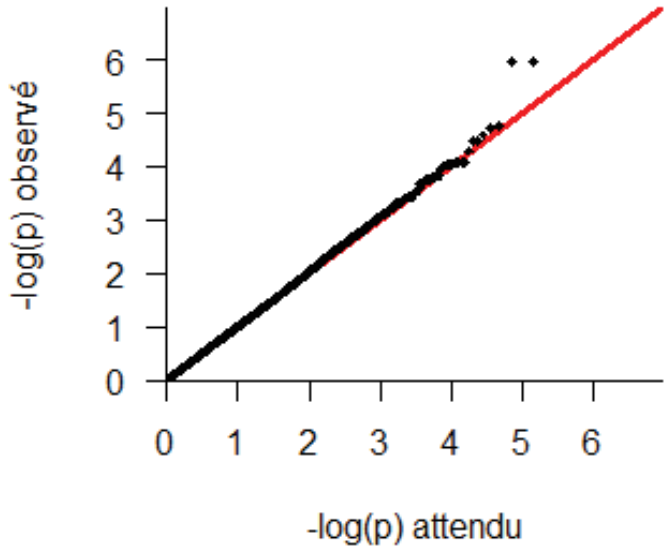
| rs9906443 | AA | AG | GG | TOTAL |
|---|---|---|---|---|
| CAS | 63 | 52 | 4 | 119 |
| CONTROLES | 98 | 22 | 0 | 120 |

Au sein de l'échantillon, aucun autre variant ayant moins de 5% de valeurs manquantes n'est en déséquilibre de liaison fort ($r^2$>0.8) avec ces 2 variants. Dans les super-populations européennes et africaines du projet 1000 génomes Phase 3, 32 autres variants sont en $r^2$>0.8 (définissant un bloc d'environ 50kb) avec rs9906443 et rs739800 chevauchant les gènes *COPRS* et *UTP6*.

Le groupe de variants défini par rs9906443 et rs739800 est situé dans la région chromosomique 17q11.2 rapportée préalablement comme possédant des groupes de gènes de susceptibilité à la tuberculose et à la lèpre dans une population brésilienne (226) et comme une région évocatrice de liaison génétique avec la tuberculose dans une population thaïe (237). Le variant rs9906443 est situé dans un intron du gène COPRS (*Coordinator Of PRMT5 And Differentiation Stimulator*). Ce gène induit une protéine requise pour l'activité de l'enzyme PMRT5 de la famille des méthyltransférases qui elle-même semble réguler l'activité de NF-κB (266). A moins de 5kb de rs9906443, le variant rs739800 est localisé à 500 bp en aval du gène *UTP6* codant pour une protéine appartenant à la sous-unité 90S du précurseur de la petite partie du ribosome. Parmi les 33 variants en LD, le premier variant rs9906443

possède le score le plus élevé dans la base RegulomeDB (192) avec un score de 1d, ce qui signifie qu'il a une forte probabilité de moduler la liaison d'un facteur de transcription et l'expression d'un gène cible. En effet, il est notamment rapporté comme eQTL du gène *COPRS* dans les cellules lymphoblastoïdes (267) et dans de nombreux autres tissus dans le projet GTEx (268), avec l'allèle G associé à une expression plus faible de *COPRS*.

Figure 39 : QQPLOT de l'analyse d'association par variant suivant le modèle additif testé par le test de tendance de Cochran Armitage pour 137 365 variants bi-alléliques ayant moins de 5% de valeurs manquantes, une MAF ≥ 5% et en équilibre d'Hardy-Weinberg (au seuil de $p > 10^{-4}$)



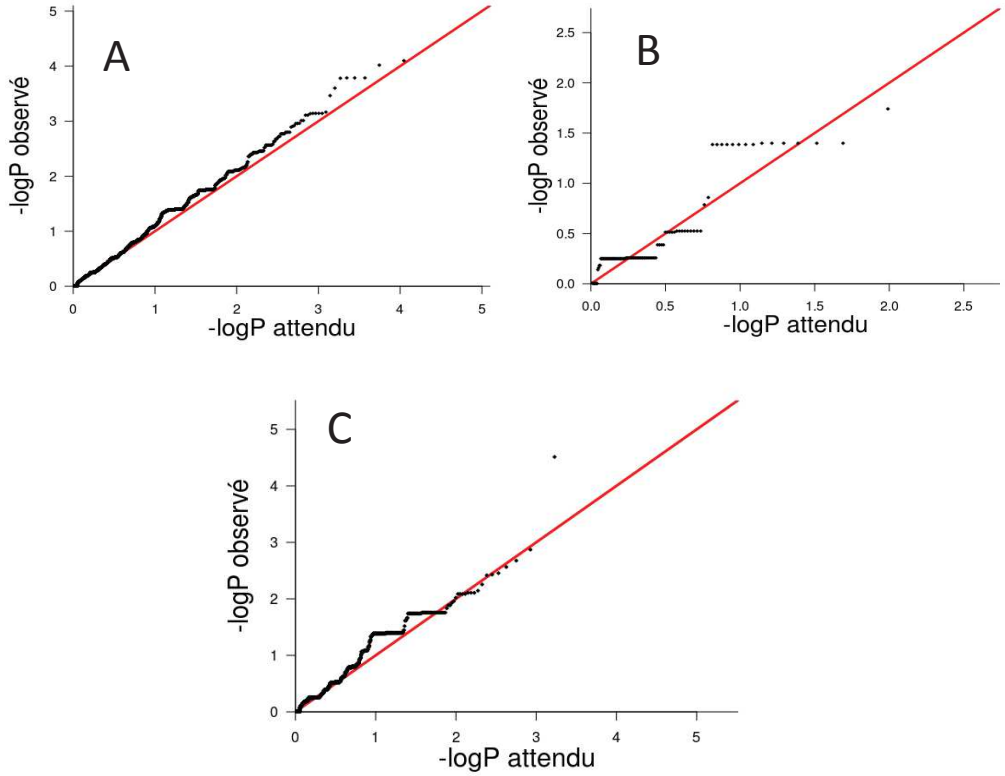

2. Analyse des variants rares

Nous nous sommes ensuite penchés sur l'impact potentiel des variants rares (MAF <5%) de l'exome sur le phénotype de tuberculose pulmonaire. Nous n'avons tout d'abord gardé que les variants bi-alléliques de bonne qualité ayant moins de 5% de valeurs manquantes dans notre échantillon d'étude final de 119 cas et 120 contrôles. Puis, nous avons filtré ces variants pour définir ceux devant être inclus dans l'étude en fonction de leur annotation fonctionnelle définie avec SnpEff (249) pour définir 2 groupes de variants ; un groupe restreint HIGH IMPACT incluant les variants ayant potentiellement un impact fonctionnel le plus fort au niveau du gène, et un groupe plus large PRIORITY1 auquel nous avons rajouté des variants

d'impact plus modéré (cf Méthodes). Pour chacun de ces groupes de variants, nous avons ensuite appliqué 1 critère supplémentaire pour créer 2 autres groupes incluant seulement les variants avec un score CADD supérieur au MSC. Pour chacun des 4 jeux de données, nous avons testé 3 modèles génétiques (dominant, récessif et bi-variants) avec une régression logistique agrégée représentant la méthode CAST. L'observation des QQplots des différents tests effectués montre que nos premières analyses ne comportaient pas de biais majeurs inflatant de manière évidente l'erreur de type I des tests, et nous n'avons donc pas eu besoin d'utiliser d'ajustement supplémentaire. La figure 40 l'illustre par un exemple pour un jeu de données parmi les 4, mais les courbes sont similaires d'un jeu de données à un autre.

Figure 40 : QQplots des analyses du jeu de données PRIORITY1 avec CADD > MSC pour les 3 modèles génétiques testés : A dominant, B récessif, C bi-variants.

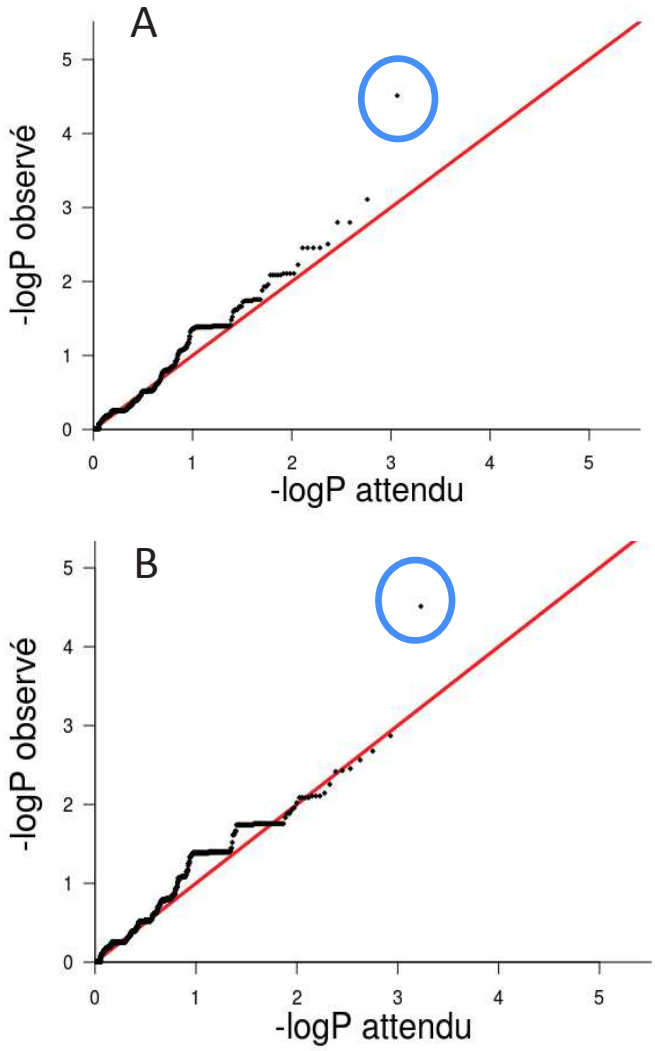
Seuls les gènes pour lesquels au moins 3 individus étaient porteurs d'un variant du jeu de données analysé ont été pris en compte : 11 116 gènes pour le modèle dominant, 97 gènes pour le modèle récessif et 1705 gènes pour le modèle bi-variants.




Parmi tous les jeux de données et modèles génétiques testés, un gène se détachait des autres avec une valeur de p inférieure d'un facteur 50 à celui des autres gènes les plus associés, et ceci à plusieurs reprises (Figure 41).

- Dans le jeu de données HIGH IMPACT et pour le modèle dominant, le gène *BTNL2* a obtenu un p = $1.5.10^{-5}$ (en unilatéral) avec 12 cas possédant au moins un variant du jeu testé versus 0 contrôles (Tableau 14).
- Dans le jeu de données PRORITY1 avec ou sans critère sur le score CADD et pour un modèle bi-variants, le gène *BTNL2* a obtenu un p = $1.5.10^{-5}$ (en unilatéral) avec 12 cas possédant au moins 2 variants du jeu testé versus 0 contrôles (Tableau 15)

Figure 41 : QQplots montrant le gène BTNL2 se détachant des autres pour 2 tests différents : A – HIGH IMPACT modèle dominant, PRIORITY1 CADD > MSC modèle bi-variants.




Dans le premier test en modèle dominant, 1155 gènes étaient informatifs (possédant au moins 3 porteurs de variants ayant un fort impact fonctionnel sur la protéine). En appliquant la méthode de Bonferroni, le seuil de significativité du test est 4.33 $10^{-5}$. Dans le second test

en modèle bi-variants, 2673 gènes sont informatifs pour le groupe de variants PRORITY 1 et 1705 gènes lorsqu'on rajoute aux variants la contrainte d'avoir un score CADD > MSC pour entrer dans le test, le seuil de significativité descendant alors respectivement à $1.87\ 10^{-5}$ et $2.93\ 10^{-5}$.

Tableau 14 : Liste des gènes les plus significatifs parmi les 1155 gènes informatifs testés sous l'hypothèse d'un modèle de transmission dominant pour le groupe de variants HIGH IMPACT (seuil de significativité = $4.33\ 10^{-5}$)

| Gene | Nb de variants testés | Nb de contrôles porteurs | Nb de cas porteurs | p |
|---|---|---|---|---|
| *BTNL2* | 2 | 0 | 12 | 1.5E-05 |
| *DPRX* | 2 | 0 | 7 | 7.95E-04 |
| *RP4-781K5.2* | 1 | 0 | 7 | 7.95E-04 |
| *RENBP* | 1 | 0 | 6 | 1.56E-03 |
| *CASP10* | 1 | 0 | 6 | 1.76E-03 |
| *CCDC67* | 3 | 0 | 6 | 1.76E-03 |
| *CYP4F2* | 3 | 0 | 6 | 1.76E-03 |
| *TRIM31* | 2 | 0 | 6 | 1.76E-03 |
| *ACOXL* | 2 | 0 | 5 | 3.90E-03 |
| *FAM126A* | 1 | 0 | 5 | 3.90E-03 |
| *HIST1H2BO* | 1 | 0 | 5 | 3.90E-03 |
| *ZNF589* | 1 | 0 | 5 | 3.90E-03 |

Tableau 15 : Liste des gènes les plus significatifs parmi les 1705 gènes informatifs testés sous l'hypothèse d'un modèle de transmission bi-variants pour le groupe de variants PRIORITY 1 CADD > MSC (seuil de significativité = $2.93\ 10^{-5}$)

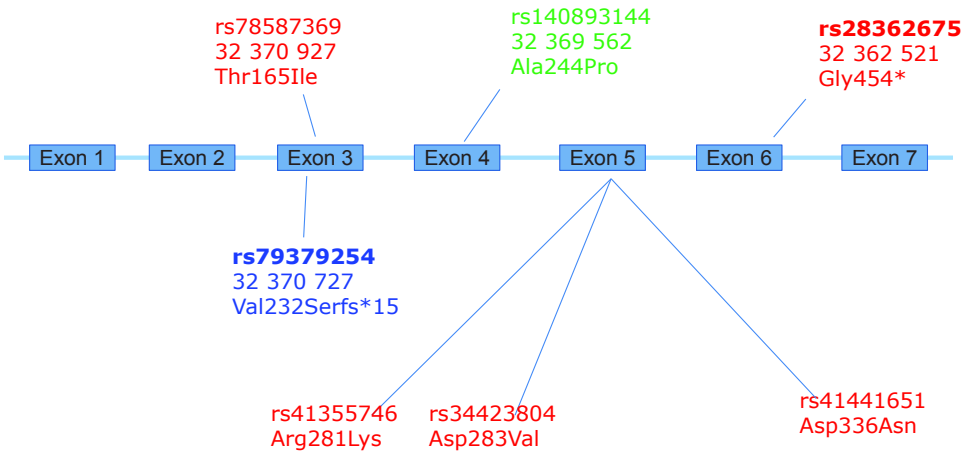
| Gene | Nb de variants testés | Nb de contrôles | Nb de cas | p |
|---|---|---|---|---|
| *BTNL2* | 10 | 0 | 12 | 1.5E-05 |
| *CEP250* | 8 | 1 | 11 | 6.7E-04 |
| *FAM208A* | 17 | 1 | 10 | 1.37E-03 |
| *IQCE* | 9 | 0 | 6 | 1.76E-03 |
| *WBSCR27* | 4 | 3 | 14 | 1.9E-03 |
| *DNAH6* | 30 | 5 | 17 | 2.8E-03 |
| *AFF1* | 14 | 0 | 5 | 3.90E-03 |
| *DSG1* | 9 | 0 | 5 | 3.90E-03 |
| *EPHA1* | 10 | 0 | 5 | 3.90E-03 |
| *FBN2* | 20 | 1 | 8 | 5.5E-03 |

En regardant la distribution des variants du gène entrant dans les tests effectués, on constate que 11 cas étaient communs aux 3 résultats car possédant le même variant STOP GAIN, rs28362675, associé à 4 autres variants non-synonymes ayant un score CADD > MSC au sein du gène et absents de tous les autres individus étudiés, pouvant laisser penser à la présence d'un haplotype commun. Le douzième cas de l'analyse du jeu de données HIGH IMPACT possédait un variant FRAMESHIFT, rs79379254, tandis que le douzième cas du jeu de données PRIORITY1 CADD > MSC était homozygote pour un variant non-synonyme, rs140893144, ayant un CADD > MSC (Figure 42).

Le variant frameshift rs79379254 retrouvé chez un des patients possède une fréquence de 0.25 % dans la base de données EXAC, de 2.2 $10^{-4}$ dans la sous-population européenne et de 2.8% dans la sous-population africaine. Le variant non-synonyme rs140893144 retrouvé à l'état homozygote chez un autres patient a une fréquence de 6.9 $10^{-4}$ dans la base de données EXAC, de 1.9 $10^{-4}$ dans la sous-population européenne et de 0.62% dans la sous-population africaine. Aucun individu n'a été retrouvé à l'état homozygote parmi les 59 161 individus de la cohorte. Le variant non-sens rs28362675 retrouvé chez 11 patients a une fréquence de 2.3% dans la base de données EXAC, mais seulement de 0.9% dans la sous-population européenne et 0.15% dans la population africaine. Au sein des 59 161 individus de la cohorte EXAC, seuls 97 sont homozygotes pour le variant dont 5 européens, les autres étant asiatiques. Dans notre échantillon de 119 patients, sa fréquence est de plus de 4.6%, soit 5 fois plus élevée que chez les européens et 30 fois plus que chez les africains.

Figure 42 : Représentation des variants du gène BTNL2 contribuant à la statistique des 2 tests pour lesquels il est significatif.

En rouge sont représentés les variants portés par les 11 mêmes patients, en bleu le variant frameshift porté par 1 patient et contribuant à la statistique du test HIGH IMPACT et en vert, le variant non-synonyme retrouvé à l'état homozygote chez un autre individu et entrant dans la statistique du jeu de données PRIORITY1 CADD > MSC.



La première question qui se pose est de savoir si le gène *BTNL2* pourrait avoir un effet direct et fort sur le développement de la tuberculose pulmonaire.

*BTNL2* (Butyrophilin Like 2) est un gène situé sur le chromosome 6, en bordure de région HLA de classe II, et code pour un récepteur de la famille B7 qui semble avoir un rôle dans la co-stimulation des lymphocytes T (269). Il a souvent été associé à la sarcoïdose, maladie inflammatoire systémique de cause inconnue, atteignant préférentiellement les poumons et se manifestant par la présence de granulomes (270–272), mais il a été également associé aux maladies inflammatoires du système digestif (273,274) . En raison des similitudes existant entre la sarcoïdose et la tuberculose, l'association de *BTNL2* avec la tuberculose a déjà été étudiée, mais les résultats n'ont pas été très convaincants (275,276). Un polymorphisme de *BTNL2* en particulier a été beaucoup étudié, il s'agit de rs2076530, situé dans une zone d'épissage et induisant un codon STOP prématuré au début de l'exon 6 du gène, soit en amont du variant non-sens porté par 11 des patients étudiés. Les 12 cas portant un variant faisant partie du groupe HIGH IMPACT sont également hétérozygotes pour rs2076530, mais le variant n'est aucunement associé à la tuberculose dans notre échantillon (Tableau 16).

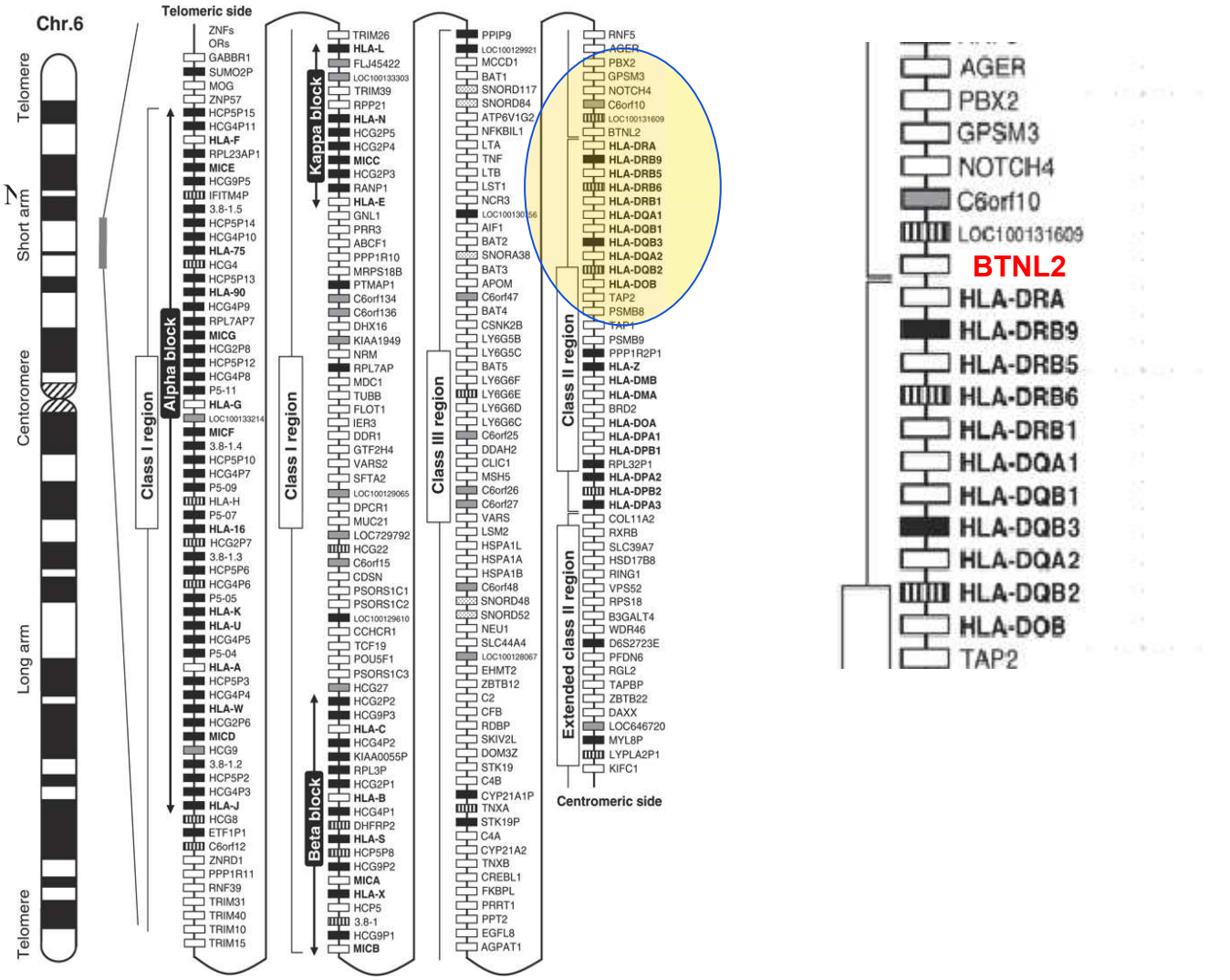
L'impact fonctionnel de rs2076530 a été particulièrement investigué dans un article datant de 2005 qui montre que l'allèle T du variant conduit à une protéine tronquée de *BTNL2* qui ne serait pas présentée à la membrane cellulaire, contrairement à la protéine longue (celle codée avec l'allèle C) et que l'épissage induit par l'allèle T est complet (270). Or l'allèle T du variant à une fréquence de 57.4% dans EXAC et 61.4% dans le projet 1000 Génomes. Cela signifie que près d'un tiers des individus seraient homozygotes pour l'allèle T et possèderaient seulement la version tronquée et non fonctionelle de la protéine. Au vu de ces observations, il paraît peu probable que la non fonctionnalité de la protéine codée par *BTNL2* puisse jouer un rôle dans le développement d'une tuberculose pulmonaire.

Tableau 16 : Génotypes du variant rs2076530 T / C au sein de notre échantillon

| **Cas** | | | **Contrôles** | | |
|---|---|---|---|---|---|
| CC | CT | TT | CC | CT | TT |
| 10 | 70 | 39 | 13 | 59 | 48 |

Cependant, le déséquilibre de liaison semble fort au sein du gène comme les résultats du test du jeu de données PRIORITY1 avec le modèle bi-variants nous l'ont montré. En observant l'ensemble des variants du gène, on constate que les 11 individus portant le variant rs28362675 possèdent en réalité 19 variants identiques dans *BTNL2* qui ne sont partagés par aucun autre individu de l'échantillon, ce qui suggère fortement l'existence d'un haplotype partagé par ces 11 individus. *BTNL2* est en effet situé en bordure de la région HLA de classe 2 (Figure 43), zone connue pour son très fort déséquilibre de liaison. L'haplotype commun aux 11 cas portant le variant rs28362675 dépasse donc certainement les limites du gène, et le lien avec la tuberculose est peut-être à rechercher ailleurs sur cet haplotype.

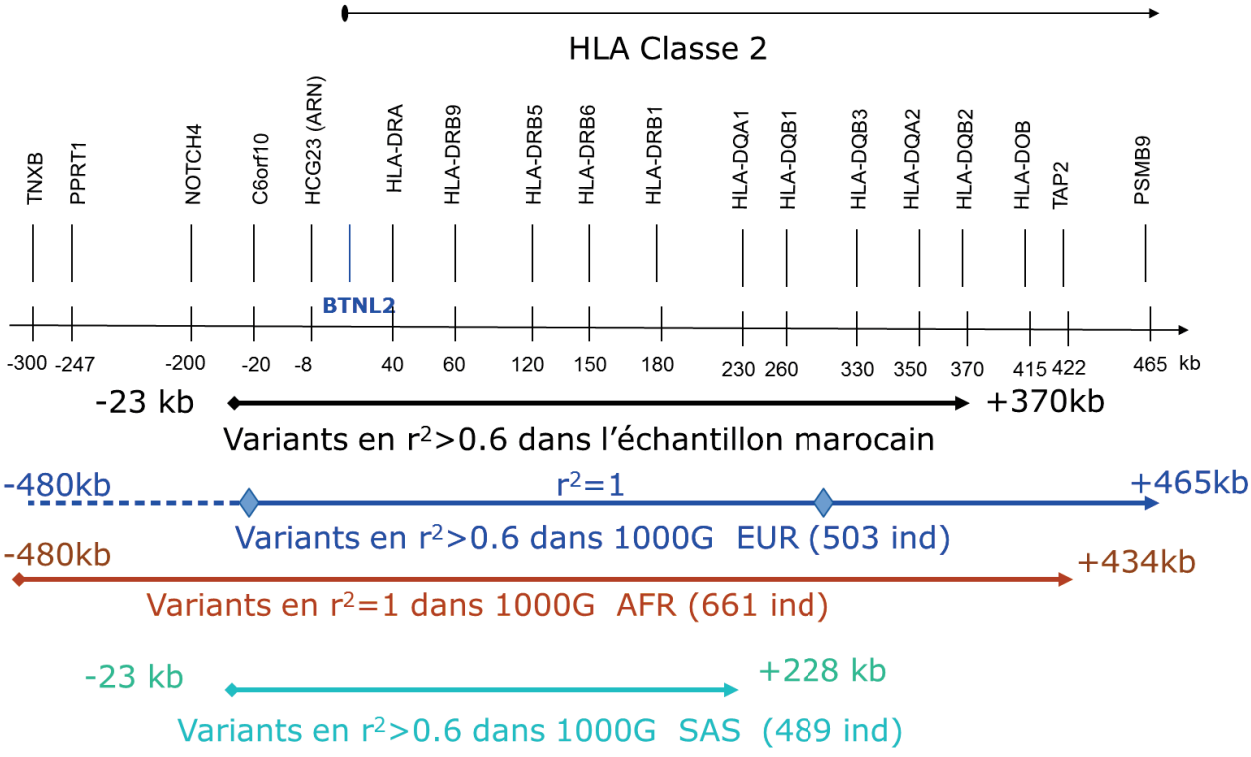
Figure 43 : Localisation de *BTNL2* sur le chromosome 6, adapté de (277)



Nous avons donc évalué le déséquilibre de liaison (LD) du variant rs28362675 dans notre échantillon à l'aide du logiciel PLINK (164) et l'avons comparé avec celui trouvé pour le projet 1000 Génomes et disponible sur LDlink (191). On constate sur la figure 44 que la région en LD avec rs28362675 s'étend bien au-delà du gène *BTNL2* et englobe plusieurs gènes de la région HLA de classe 2, aussi bien dans notre échantillon marocain que chez les individus européens, africains ou asiatiques du projet 1000 génomes. Il est donc tout à fait possible que nos 11 cas possèdent un ou des variants de réelle susceptibilité à la tuberculose au sein de la région HLA de classe 2. L'étude de cette hypothèse est en cours actuellement avec l'inférence des typages HLA à partir des données d'exomes des individus de l'échantillon d'étude et le génotypage HLA de référence de certains loci sur une partie des individus étudiés.

Figure 44 : Région de déséquilibre de liaison avec le variant rs28362675 au sein de l'échantillon d'étude (noir), des individus européens du projet 1000 génomes (bleu), des individus africains du projet 1000 génomes (orange) et des individus d'Asie du Sud du projet 1000 génomes (turquoise).

Entre les 2 losanges bleus, le $r^2$=1 chez les européens de 1000 Génomes



# D. Discussion et Perspectives

Dans cette étude, nous avons testé l'hypothèse que la tuberculose pulmonaire pouvait être due à des variants génétiques dits rares (MAF < 5%) et ayant des effets plus forts que les variants communs analysés dans les GWAS classiques. Les puces de génotypage ne pouvant fournir l'information de l'ensemble des variants codants de chaque individu, nous avons travaillé sur des données issues du séquençage de l'exome. Nous avons comparé les données génétiques de 119 patients marocains atteints de tuberculose pulmonaire avec 120 individus marocains infectés par *M.tuberculosis*, positifs à la fois pour le test de Mantoux et pour le Quantiféron. Bien que ce ne soit pas le but premier de l'étude, une analyse par variant réalisée sur les variants de MAF > 5% a mis en lumière 2 variants évocateurs d'association, rs9906443 et rs739800, en complet déséquilibre de liaison, situés dans une région du chromosome 17 rapportée préalablement comme liée à la tuberculose pulmonaire (226,237). D'après la base de données RegulomeDB, rs9906443 a une forte probabilité de réguler la liaison de facteurs de transcription et l'expression du gène *COPRS* auquel il appartient. La protéine codée par *COPRS* coopère avec l'enzyme PMRT5 de la famille des méthyltransférases qui semble réguler l'activité de NF-κB (266), facteur de transcription impliqué dans la réponse immunitaire. Ce résultat nécessite d'être répliqué dans une autre cohorte, mais si l'association s'avérait confirmée, la taille d'effet trouvée (OR=1.42) serait dans la fourchette haute de celle des variants identifiés jusqu'à présent par analyse d'association pangénomique.

En analysant les variants de faible fréquence (MAF < 5% dans notre cas), nous avons voulu tester l'hypothèse qu'un grand nombre de ces variants dits rares pouvait avoir un effet relativement fort sur le développement d'une tuberculose pulmonaire, et que ces variants, bien que nombreux, étaient tous localisés dans un même petit nombre de gènes, plus faciles à mettre en évidence statistiquement. Pour tester cette hypothèse, parmi toutes les méthodes disponibles, nous avons utilisé la méthode CAST d'agrégation de variants par gène nous permettant de comparer le nombre de porteurs d'au moins un variant rare au sein du même gène entre les cas et les contrôles. Nous avons également émis l'hypothèse que ces variants rares à effet fort avaient un impact fonctionnel sur le gène auquel ils appartenaient, c'est-à-dire qu'ils avaient un effet visible sur la protéine codée par celui-ci et y induisaient au minimum un changement d'acide aminé. Afin de limiter la prise en compte de variants ayant une faible probabilité de jouer un rôle dans le phénotype étudié, nous avons donc établi 4

groupes de variants à tester à partir de leur annotation fonctionnelle donnée par snpEff d'une part, et du score de prédiction de pathogénicité CADD d'autre part pour être encore plus restrictif.

A partir du groupe de variants ayant les plus forts impacts fonctionnels sur les protéines, nous avons identifié le gène *BTNL2* comme significativement associé au développement d'une tuberculose pulmonaire ($p = 1.5\ 10^{-5}$) ; 12 cas étaient porteurs d'un variant rare en son sein alors qu'aucun contrôle n'en portait. Parmi les 12 cas, 11 portaient le même variant non-sens rs28362675 à l'état hétérozygote absent chez les contrôles et entraînant un codon stop prématuré dans le 6ème exon du gène en comportant 7. La fréquence de l'allèle mineur de ce SNP était donc de 4.6% dans les patients marocains, alors que dans la base de données EXAC il affichait une MAF de 0.9% pour la population européenne (calculée sur 32 994 individus) et de 0.15% pour la population africaine (calculée sur 5 180 individus). Le gène *BTNL2* a été étudié dans le cadre de sa potentielle association avec la sarcoïdose, et il a été alors montré qu'une majorité d'êtres humains possèderaient un variant non-sens en amont de celui identifié dans notre étude conduisant à une perte de fonctionnalité de la protéine codée par le gène *BTNL2*, variant non associé à la tuberculose pulmonaire dans notre étude (meilleur p=0.21 pour un modèle dominant). Il paraît donc peu probable que le gène identifié dans notre étude joue un rôle direct dans notre phénotype d'intérêt.

En revanche, *BTNL2* se situe au bord de la région comportant les gènes du complexe majeur d'histocompatibilité (HLA) de classe 2, région connue pour son polymorphisme extrêmement important, mais aussi pour son faible taux de recombinaison conduisant à de longs segments haplotypiques (278). En regardant le déséquilibre de liaison du variant rs28362675 au sein de notre échantillon marocain d'une part, et également dans différentes sous-populations de 1000 Génomes, nous avons constaté que celui-ci pouvait s'étendre jusqu'à plus de 300kb du variant, englobant plusieurs gènes du système HLA de classe 2. Les variants trouvés dans le gène *BTNL2* pourraient donc être des marqueurs d'autres variants faisant partie du même haplotype et ayant peut-être eux un impact plus direct sur le développement d'une tuberculose ; c'est la piste principale qu'il reste à explorer dans ce travail.

Les gènes classiques du complexe majeur d'histocompatibilité de classe 2 codent pour des molécules présentes à la surface des cellules présentatrices d'antigène qui assurent la présentation de l'antigène aux lymphocytes T afin de les activer. Le rôle des gènes de cette région dans la susceptibilité aux maladies infectieuses et à la tuberculose en particulier a été étudié dans de nombreuses études ; certains allèles ont été mis en évidence que ce soit des

allèles protecteurs ou de susceptibilité à la tuberculose pulmonaire même si les tailles d'effet restent relativement modestes et hétérogènes d'une population à une autre (236,279). Pour continuer cette étude, il semblerait donc intéressant de regarder si les segments haplotypiques capturés par les variants perte de fonction de *BTNL2* dans notre population marocaine conduisent à des allèles identiques du complexe HLA de classe 2, qu'il faudra ensuite comparer aux allèles retrouvés chez les contrôles d'une part et dans la population nord-africaine d'autre part grâce à la base de données allelefrequencies.net (280). A partir des données d'exomes classiques, il est difficile de définir les allèles HLA des individus du fait de la très grande variabilité de ces régions et du mauvais alignement des « reads » sur le génome de référence (281,282). Il faudra trouver d'autres moyens pour exploiter les données de séquençage disponibles (par exemple, utilisation des « reads » non alignés sur le génome de référence pour essayer de les aligner sur les séquences des différents allèles HLA)(283,284) ou bien de réaliser des typages HLA avec les méthodes de référence pour les loci nous intéressant. Ce travail est en cours actuellement.

Grâce au séquençage de l'exome et à l'analyse de variants rares regroupés par gène, nous avons mis en évidence un gène significativement associé à la tuberculose pulmonaire. Au sein de ce gène, un variant probablement marqueur d'un haplotype particulier pourrait expliquer à lui seul 10% du nombre de cas de notre échantillon, si le mécanisme était démontré. Il est intéressant de voir dans le Tableau 17 que l'association du gène *BTNL2* avec la tuberculose pulmonaire existe pour au moins un modèle génétique testé dans plusieurs jeux de données pouvant être étudiés. Ce résultat est en cours de réplication dans des cohortes d'origine ethnique différente (africaine et asiatique). Notre hypothèse de départ était qu'un grand nombre de variants rares appartenant au(x) même(s) gène(s) pourraient jouer un rôle important dans la maladie. Or notre résultat est dû à la présence de seulement 2 variants rares différents, l'un effectivement très rare ayant une fréquence de 0.2% dans notre échantillon et l'autre une fréquence de 2.3%. Il est intéressant de noter qu'on retrouve ainsi le profil des gènes identifiés jusqu'à présent par des analyses de variants rares dans des petits échantillons (285–287) où finalement 2 variants expliquent à eux-seuls l'association du gène avec le phénotype étudié. Aurait-on pu identifier le variant principal rs28362675 sans passer par la méthode d'aggrégation des variants par gène ? Si on utilise les mêmes filtres basés sur les annotations fonctionnelles pour sélectionner les variants à analyser, en ne considérant que les 1071 variants à fort impact fonctionnel sur la protéine avec une MAF $\geq$ 2% et moins de 5% de données manquantes dans notre échantillon, l'analyse du variant rs28362675 pris isolément

n'aurait pas conclu à sa significativité ($p = 6.5.10^{-4}$ pour un seuil de $4.66.10^{-5}$), et on n'aurait pu inférer son intérêt potentiel.

Tableau 17 : Résultat de différentes analyses CAST pour le gène BTNL2 pour des variants bi-alléliques de MAF < 5% avec les 3 modèles génétiques testés

| Critères de sélection des variants | Gène | Nb de variants analysés | Modèle dominant | Modèle récessif | Modèle bi-variants |
|---|---|---|---|---|---|
| **HIGH IMPACT + CADD > MSC** | BTNL2 | 2 | **$1.5\ 10^{-5}$** | NA | NA |
| **HIGH IMPACT** | BTNL2 | 2 | **$1.5\ 10^{-5}$** | NA | NA |
| **PRIORITY1 + CADD > MSC** | BTNL2 | 10 | $1.7\ 10^{-3}$ | $1.2\ 10^{-1}$ | **$1.5\ 10^{-5}$** |
| **PRIORITY1** | BTNL2 | 13 | $2.01\ 10^{-3}$ | $1.2\ 10^{-1}$ | **$1.5\ 10^{-5}$** |
| **CADD > MSC** | BTNL2 | 24 | $1.5\ 10^{-2}$ | $1.2\ 10^{-1}$ | **$1.4\ 10^{-6}$** |
| **Tous** | BTNL2 | 58 | $1.5\ 10^{-2}$ | $1.2\ 10^{-1}$ | **$1.4\ 10^{-6}$** |

Lorsque les effectifs étudiés deviennent plus importants, le nombre de variants entrant en jeu augmente et on se retrouve dans la configuration attendue d'une hétérogénéité allélique plus grande (288–290). Dans tous les cas, les critères de sélection des variants devant être analysés sont très importants. Dans ce manuscrit, seuls quelques groupes de variants testés ont été présentés, mais d'autres critères de sélection pourraient être appliqués : le filtre de MAF a été arbitrairement fixé à 5% sur l'ensemble de l'échantillon, on pourrait d'une part faire varier ce seuil de fréquence (en particulier pour le modèle récessif, il serait envisageable de faire entrer dans l'analyse des variants plus fréquents à 10% ou 20% de fréquence allélique) et d'autre part appliquer ce seuil sur l'ensemble de l'échantillon ou seulement sur les contrôles. On pourrait également filtrer les variants sur leur fréquence dans les bases de données publiques pour cibler particulièrement des variants très rares dans la population générale mais présents dans notre échantillon. Ces analyses sont actuellement en cours, avec l'idée de générer de nouvelles hypothèses à étudier, plus que d'établir un résultat définitif basé uniquement sur un test statistique.

D'un point de vue global, le taux de mortalité des individus souffrant de tuberculose a beaucoup diminué grâce aux traitements médicamenteux, mais l'émergence de souches très résistantes aux antibiotiques existants constitue un signal d'alarme que d'autres approches seront nécessaires pour mettre fin à la tuberculose endémique dans certains pays du monde (291). Un vaccin plus efficace que le BCG, parvenant à empêcher le développement d'une tuberculose pulmonaire à l'âge adulte, serait utile pour diminuer le taux de transmission du bacille, mais malheureusement les essais cliniques vaccinaux récents sont plutôt décevants, signes sans doute de notre mauvaise compréhension de la vulnérabilité à *M.tuberculosis*. Par exemple, les lignées de souris possédant une certaine résistance génétique à la tuberculose bénéficient davantage des effets de la vaccination par le BCG que leurs congénères plus susceptibles à la mycobactérie (292). De même, les individus ayant déjà souffert d'une tuberculose clinique sont davantage sujets à des épisodes tuberculeux ultérieurs que l'incidence de la tuberculose dans la population générale ne le laisserait supposer (293). Ces observations suggèrent donc que les individus déclarant une tuberculose clinique ont une susceptibilité intrinsèque dont ils ne peuvent s'affranchir avec une vaccination traditionelle ou des antibiotiques, et que tout vaccin efficace devra probablement prendre en compte la susceptibilité génétique des patients si le but est de générer une réponse immunitaire pouvant les protéger par la suite. De la même manière, des approches thérapeutiques nouvelles basées sur la complémentation de déficiences immunitaires particulières identifiées par des études de génétique humaine pourraient avoir un effet positif conjoint avec les traitements classiques contre la tuberculose, comme c'est déjà le cas aujourd'hui dans le traitement par IFNγ recombinant des jeunes enfants souffrant de tuberculose disséminée due à un déficit de production de la cytokine (202,203). Les études de génétique humaine ont donc toute leur place dans la stratégie de lutte contre la tuberculose au niveau planétaire.

# III - Bibliographie


1. WHO - Global Tuberculosis Report 2016 [Internet]. [cited 2017 Feb 25]. Available from: http://www.who.int/tb/publications/global_report/high_tb_burdencountrylists2016-2020.pdf?ua=1

2. Smith T. A comparative study of bovine tubercle bacilli and of human bacilli from sputum. J Exp Med. 1898 Jul 1;3:451–511.

3. Cole ST, Brosch R, Parkhill J, Garnier T, Churcher C, Harris D, et al. Deciphering the biology of Mycobacterium tuberculosis from the complete genome sequence. Nature. 1998 Jun 11;393(6685):537–44.

4. Aranaz A, Cousins D, Mateos A, Domínguez L. Elevation of Mycobacterium tuberculosis subsp. caprae Aranaz et al. 1999 to species rank as Mycobacterium caprae comb. nov., sp. nov. Int J Syst Evol Microbiol. 2003 Nov;53(Pt 6):1785–9.

5. Cousins DV, Bastida R, Cataldi A, Quse V, Redrobe S, Dow S, et al. Tuberculosis in seals caused by a novel member of the Mycobacterium tuberculosis complex: Mycobacterium pinnipedii sp. nov. Int J Syst Evol Microbiol. 2003 Sep;53(Pt 5):1305–14.

6. Esteban J, Munoz-Egea M-C. Mycobacterium bovis and Other Uncommon Members of the Mycobacterium tuberculosis Complex. Microbiol Spectr [Internet]. 2016 Oct [cited 2017 May 22];4(6). Available from: http://www.asmscience.org.gate2.inist.fr/docserver/fulltext/microbiolspec/4/6/TNMI7-0021-2016.pdf?expires=1495457499&id=id&accname=esid054120&checksum=99F6FCF4AE6E4114C07851D0405546D9

7. Gutierrez MC, Brisse S, Brosch R, Fabre M, Omaïs B, Marmiesse M, et al. Ancient Origin and Gene Mosaicism of the Progenitor of Mycobacterium tuberculosis. PLOS Pathog. 2005 Aug 19;1(1):e5.

8. Vervenne RAW, Jones SL, Soolingen D van, van der Laan T, Andersen P, Heidt PJ, et al. TB diagnosis in non-human primates: comparison of two interferon-γ assays and the skin test for identification of Mycobacterium tuberculosis infection. Vet Immunol Immunopathol. 2004 Jul;100(1–2):61–71.

9. Castets M, Boisvert H, Grumbach F, Brunel M, Rist N. [Tuberculosis bacilli of the African type: preliminary note]. Rev Tuberc Pneumol (Paris). 1968 Mar;32(2):179–84.

10. Djelouadji Z, Raoult D, Drancourt M. Palaeogenomics of Mycobacterium tuberculosis: epidemic bursts with a degrading genome. Lancet Infect Dis. 2011 Aug;11(8):641–50.

11. Kubica T, Rüsch-Gerdes S, Niemann S. Mycobacterium bovis subsp. caprae caused one-third of human M. bovis-associated tuberculosis cases reported in Germany between 1999 and 2001. J Clin Microbiol. 2003 Jul;41(7):3070–7.

12. Niemann S, Richter E, Dalügge-Tamm H, Schlesinger H, Graupner D, Königstein B, et al. Two cases of Mycobacterium microti derived tuberculosis in HIV-negative immunocompetent patients. Emerg Infect Dis. 2000 Oct;6(5):539–42.

13. Xavier Emmanuel F, Seagar A-L, Doig C, Rayner A, Claxton P, Laurenson I. Human and animal infections with Mycobacterium microti, Scotland. Emerg Infect Dis. 2007 Dec;13(12):1924–7.

14. Comas I, Gagneux S. The Past and Future of Tuberculosis Research. PLOS Pathog. 2009 Oct 26;5(10):e1000600.

15. Smith NH, Hewinson RG, Kremer K, Brosch R, Gordon SV. Myths and misconceptions: the origin and evolution of Mycobacterium tuberculosis. Nat Rev Microbiol. 2009 Jul;7(7):537–44.

16. Loeffler SH, de Lisle GW, Neill MA, Collins DM, Price-Carter M, Paterson B, et al. The seal tuberculosis agent, Mycobacterium pinnipedii, infects domestic cattle in New Zealand: epidemiologic factors and DNA strain typing. J Wildl Dis. 2014 Apr;50(2):180–7.

17. Bos KI, Harkins KM, Herbig A, Coscolla M, Weber N, Comas I, et al. Pre-Columbian mycobacterial genomes reveal seals as a source of New World human tuberculosis. Nature. 2014 Oct 23;514(7523):494–7.

18. Supply P, Marceau M, Mangenot S, Roche D, Rouanet C, Khanna V, et al. Genomic analysis of smooth tubercle bacilli provides insights into ancestry and pathoadaptation of Mycobacterium tuberculosis. Nat Genet. 2013 Feb;45(2):172–9.

19. Boritsch EC, Frigui W, Cascioferro A, Malaga W, Etienne G, Laval F, et al. pks5-recombination-mediated surface remodelling in Mycobacterium tuberculosis emergence. Nat Microbiol. 2016 Jan 27;1(2):15019.

20. Samra Z, Kaufman L, Bechor J, Bahar J. Comparative Study of Three Culture Systems for Optimal Recovery of Mycobacteria from Different Clinical Specimens. Eur J Clin Microbiol Infect Dis. 2000 Nov 1;19(10):750–4.

21. WHO | A world free of tuberculosis (TB) [Internet]. WHO. [cited 2014 Dec 17]. Available from: http://www.who.int/tb/en/

22. Russell DG. Who puts the tubercle in tuberculosis? Nat Rev Microbiol. 2007 Jan;5(1):39–47.

23. Ernst JD. The immunological life cycle of tuberculosis. Nat Rev Immunol. 2012 Aug;12(8):581–91.

24. Verrall AJ, Netea MG, Alisjahbana B, Hill PC, van Crevel R. Early clearance of Mycobacterium tuberculosis: a new frontier in prevention. Immunology. 2014 Apr;141(4):506–13.

25. van Crevel R, Ottenhoff THM, van der Meer JWM. Innate immunity to Mycobacterium tuberculosis. Clin Microbiol Rev. 2002 Apr;15(2):294–309.

26. Neyrolles O, Gicquel B, Quintana-Murci L. Towards a crucial role for DC-SIGN in tuberculosis and beyond. Trends Microbiol. 2006 Sep;14(9):383–7.

27. Davis JM, Ramakrishnan L. The role of the granuloma in expansion and dissemination of early tuberculous infection. Cell. 2009 Jan 9;136(1):37–49.

28. Ehlers S. DC-SIGN and mannosylated surface structures of Mycobacterium tuberculosis: a deceptive liaison. Eur J Cell Biol. 2010 Jan;89(1):95–101.

29. Jo E-K. Mycobacterial interaction with innate receptors: TLRs, C-type lectins, and NLRs. Curr Opin Infect Dis. 2008 Jun;21(3):279–86.

30. O'Garra A, Redford PS, McNab FW, Bloom CI, Wilkinson RJ, Berry MPR. The Immune Response in Tuberculosis. Annu Rev Immunol. 2013 Mar 21;31(1):475–527.

31. Blomgran R, Desvignes L, Briken V, Ernst JD. Mycobacterium tuberculosis inhibits neutrophil apoptosis, leading to delayed activation of naive CD4 T cells. Cell Host Microbe. 2012 Jan 19;11(1):81–90.

32. Korbel DS, Schneider BE, Schaible UE. Innate immunity in tuberculosis: myths and truth. Microbes Infect. 2008 Jul;10(9):995–1004.

33. Bustamante J, Boisson-Dupuis S, Abel L, Casanova J-L. Mendelian susceptibility to mycobacterial disease : Genetic, immunological, and clinical features of inborn errors of IFN-γ immunity. Semin Immunol. 2014 Dec;26(6):454–70.

34. Kreins AY, Ciancanelli MJ, Okada S, Kong X-F, Ramírez-Alejo N, Kilic SS, et al. Human TYK2 deficiency: Mycobacterial and viral infections without hyper-IgE syndrome. J Exp Med. 2015 Sep 21;212(10):1641–62.

35. Boisson-Dupuis S, Bustamante J, El-Baghdadi J, Camcioglu Y, Parvaneh N, El Azbaoui S, et al. Inherited and acquired immunodeficiencies underlying tuberculosis in childhood. Immunol Rev. 2015 Mar 1;264(1):103–20.

36. Filipe-Santos O, Bustamante J, Chapgier A, Vogt G, de Beaucoudrey L, Feinberg J, et al. Inborn errors of IL-12/23- and IFN-γ-mediated immunity: molecular, cellular, and clinical features. Semin Immunol. 2006 Dec;18(6):347–61.

37. Tsumura M, Okada S, Sakai H, Yasunaga S, Ohtsubo M, Murata T, et al. Dominant-negative STAT1 SH2 domain mutations in unrelated patients with Mendelian susceptibility to mycobacterial disease. Hum Mutat. 2012 Sep;33(9):1377–87.

38. Dorman SE, Picard C, Lammas D, Heyne K, van Dissel JT, Baretto R, et al. Clinical features of dominant and recessive interferon gamma receptor 1 deficiencies. Lancet Lond Engl. 2004 Dec 11;364(9451):2113–21.

39. Russell DG, Barry CE, Flynn JL. Tuberculosis: what we don't know can, and does, hurt us. Science. 2010 May 14;328(5980):852–6.

40. Pai M, Behr M. Latent Mycobacterium tuberculosis Infection and Interferon-Gamma Release Assays. Microbiol Spectr. 2016 Oct;4(5).

41. Lillebaek T, Dirksen A, Baess I, Strunge B, Thomsen VØ, Andersen ÅB. Molecular Evidence of Endogenous Reactivation of Mycobacterium tuberculosis after 33 Years of Latent Infection. J Infect Dis. 2002 Feb 1;185(3):401–4.

42. Esmail H, Barry CE, Young DB, Wilkinson RJ. The ongoing challenge of latent tuberculosis. Phil Trans R Soc B. 2014 Jun 19;369(1645):20130437.

43. Borgdorff MW, Sebek M, Geskus RB, Kremer K, Kalisvaart N, van Soolingen D. The incubation period distribution of tuberculosis estimated with a molecular epidemiological approach. Int J Epidemiol. 2011 Aug 1;40(4):964–70.

44. McCarthy OR. Asian immigrant tuberculosis—the effect of visiting Asia. Br J Dis Chest. 1984;78:248–53.

45. Millington KA, Gooding S, Hinks TSC, Reynolds DJM, Lalvani A. Mycobacterium tuberculosis-Specific Cellular Immune Profiles Suggest Bacillary Persistence Decades after Spontaneous Cure in Untreated Tuberculosis. J Infect Dis. 2010 Jan 12;202(11):1685–16849.

46. McShane H, Jacobs WR, Fine PE, Reed SG, McMurray DN, Behr M, et al. BCG: Myths, realities, and the need for alternative vaccine strategies. Tuberculosis. 2012 May;92(3):283–8.

47. Behr MA. BCG — different strains, different vaccines? Lancet Infect Dis. 2002 Feb;2(2):86–92.

48. Fomukong NG, Dale JW, Osborn TW, Grange JM. Use of gene probes based on the insertion sequence IS986 to differentiate between BCG vaccine strains. J Appl Bacteriol. 1992 Feb;72(2):126–33.

49. Li H, Ulstrup JC, Jonassen TO, Melby K, Nagai S, Harboe M. Evidence for absence of the MPB64 gene in some substrains of Mycobacterium bovis BCG. Infect Immun. 1993 May;61(5):1730–4.

50. Mahairas GG, Sabo PJ, Hickey MJ, Singh DC, Stover CK. Molecular analysis of genetic differences between Mycobacterium bovis BCG and virulent M. bovis. J Bacteriol. 1996;178(5):1274–1282.

51. Pym AS, Brodin P, Brosch R, Huerre M, Cole ST. Loss of RD1 contributed to the attenuation of the live tuberculosis vaccines Mycobacterium bovis BCG and Mycobacterium microti. Mol Microbiol. 2002 Nov;46(3):709–17.

52. Murray JF. A Century of Tuberculosis. Am J Respir Crit Care Med. 2004 Jun 1;169(11):1181–6.

53. HAS-SANTE- Tuberculose maladie - Actes et prestations [Internet]. 2012 [cited 2017 Mar 10]. Available from: http://www.has-sante.fr/portail/upload/docs/application/pdf/actualisationlap_tuberculose__web_.pdf

54. Denholm JT, McBryde ES. The use of anti-tuberculosis therapy for latent TB infection. Infect Drug Resist. 2010;3:63.

55. Getahun H, Matteelli A, Abubakar I, Aziz MA, Baddeley A, Barreira D, et al. Management of latent Mycobacterium tuberculosis infection: WHO guidelines for low tuberculosis burden countries. Eur Respir J. 2015 Sep 24;ERJ-01245-2015.

56. Stewart GR, Robertson BD, Young DB. Tuberculosis: a problem with persistence. Nat Rev Microbiol. 2003 Nov;1(2):97–105.

57. Vukmanovic-Stejic M, Reed JR, Lacy KE, Rustin MHA, Akbar AN. Mantoux Test as a model for a secondary immune response in humans. Immunol Lett. 2006 Nov 15;107(2):93–101.

58. Lee E, Holzman RS. Evolution and current use of the tuberculin test. Clin Infect Dis Off Publ Infect Dis Soc Am. 2002 Feb 1;34(3):365–70.

59. Farhat M, Greenaway C, Pai M, Menzies D. False-positive tuberculin skin tests: what is the absolute effect of BCG and non-tuberculous mycobacteria? Int J Tuberc Lung Dis Off J Int Union Tuberc Lung Dis. 2006 Nov;10(11):1192–204.

60. Cobelens FGJ, Menzies D, Farhat M. False-positive tuberculin reactions due to non-tuberculous mycobacterial infections [Correspondence]. Int J Tuberc Lung Dis. 2007 Aug 1;11(8):934–5.

61. Menzies D. What does tuberculin reactivity after bacille Calmette-Guérin vaccination tell us? Clin Infect Dis Off Publ Infect Dis Soc Am. 2000 Sep;31 Suppl 3:S71-74.

62. Mancuso JD, Mody RM, Olsen CH, Harrison LH, Santosham M, Aronson NE. The Long-Term Effect of Bacille Calmette-Guérin Vaccination on Tuberculin Skin Testing: A 55-Year Follow-Up Study. Chest. 2017 Jan 10;

63. Andersen P, Munk ME, Pollock JM, Doherty TM. Specific immune-based diagnosis of tuberculosis. Lancet. 2000 Sep 23;356(9235):1099–104.

64. the Council of the Infectious Disease Society of America. Diagnostic Standards and Classification of Tuberculosis in Adults and Children. This official statement of the American Thoracic Society and the Centers for Disease Control and Prevention was adopted by the ATS Board of Directors, July 1999. This statement was endorsed by the Council of the Infectious Disease Society of America, September 1999. Am J Respir Crit Care Med. 2000 Apr;161(4 Pt 1):1376–95.

65. CDC | TB | LTBI - Diagnosis of Latent TB Infection [Internet]. [cited 2017 Mar 25]. Available from: https://www.cdc.gov/tb/publications/ltbi/diagnosis.htm

66. Aissa K, Madhi F, Ronsin N, Delarocque F, Lecuyer A, Decludt B, et al. Evaluation of a Model for Efficient Screening of Tuberculosis Contact Subjects. Am J Respir Crit Care Med. 2008 May;177(9):1041–7.

67. Pai M, Riley LW, Colford Jr JM. Interferon-γ assays in the immunodiagnosis of tuberculosis: a systematic review. Lancet Infect Dis. 2004;4(12):761–776.

68. Pai M, Zwerling A, Menzies D. Systematic review: T-cell-based assays for the diagnosis of latent tuberculosis infection: an update. Ann Intern Med. 2008 Aug 5;149(3):177–84.

69. Pai M, Denkinger CM, Kik SV, Rangaka MX, Zwerling A, Oxlade O, et al. Gamma Interferon Release Assays for Detection of Mycobacterium tuberculosis Infection. Clin Microbiol Rev. 2014 Jan 1;27(1):3–20.

70. Sørensen AL, Nagai S, Houen G, Andersen P, Andersen AB. Purification and characterization of a low-molecular-mass T-cell antigen secreted by Mycobacterium tuberculosis. Infect Immun. 1995 May;63(5):1710–7.

71. Geluk A, van Meijgaarden KE, Franken KLMC, Subronto YW, Wieles B, Arend SM, et al. Identification and characterization of the ESAT-6 homologue of Mycobacterium leprae and T-cell cross-reactivity with Mycobacterium tuberculosis. Infect Immun. 2002 May;70(5):2544–8.

72. Geluk A, van Meijgaarden KE, Franken KLMC, Wieles B, Arend SM, Faber WR, et al. Immunological crossreactivity of the Mycobacterium leprae CFP-10 with its homologue in Mycobacterium tuberculosis. Scand J Immunol. 2004 Jan;59(1):66–70.

73. Rendini T, Levis W. Quantiferon-Gold Tuberculosis Test Cannot Detect Latent Tuberculosis in Patients With Leprosy. Clin Infect Dis Off Publ Infect Dis Soc Am. 2015 Nov 1;61(9):1439–40.

74. Aagaard C, Brock I, Olsen A, Ottenhoff THM, Weldingh K, Andersen P. Mapping Immune Reactivity toward Rv2653 and Rv2654: Two Novel Low-Molecular-Mass Antigens Found Specifically in the Mycobacterium tuberculosis Complex. J Infect Dis. 2004 Jan 3;189(5):812–9.

75. Qiagen. Notice QuantiFERON®-TB Gold (QFT®) ELISA [Internet]. Available from: https://www.google.fr/url?sa=t&rct=j&q=&esrc=s&source=web&cd=1&ved=0ahUKEwiy1N_clu_UAhUHxxQKHZvEDGUQFggjMAA&url=https%3A%2F%2Fwww.researchgate.net%2Ffile.PostFileLoader.html%3Fid%3D571a186648954cd29170758c%26assetKey%3DAS%253A353649306947584%2540461327973998&usg=AFQjCNG9Qgj4C9i8Jg-C-rbjgFADJSLJBA&cad=rjt

76. Wilson FA, Miller TL, Stimpson JP. Mycobacterium Tuberculosis Infection, Immigration Status, and Diagnostic Discordance: A Comparison of Tuberculin Skin Test and QuantiFERON-TB Gold In-Tube Test Among Immigrants to the U.S. Public Health Rep Wash DC 1974. 2016 Apr;131(2):303–10.

77. Pai M, Kalantri S, Menzies D. Discordance between tuberculin skin test and interferon-gamma assays. Int J Tuberc Lung Dis Off J Int Union Tuberc Lung Dis. 2006 Aug;10(8):942–3.

78. Nienhaus A, Schablon A, Diel R. Interferon-gamma release assay for the diagnosis of latent TB infection--analysis of discordant results, when compared to the tuberculin skin test. PloS One. 2008 Jul 16;3(7):e2665.

79. Menzies D, Pai M, Comstock G. Meta-analysis: new tests for the diagnosis of latent tuberculosis infection: areas of uncertainty and recommendations for research. Ann Intern Med. 2007 Mar 6;146(5):340–54.

80. Mostafavi E, Nasehi M, Hashemi Shahraki A, Esmaeili S, Ghaderi E, Sharafi S, et al. Comparison of the tuberculin skin test and the QuantiFERON-TB Gold test in detecting latent tuberculosis in health care workers in Iran. Epidemiol Health. 2016;38:e2016032.

81. Tripodi D, Brunet-Courtois B, Nael V, Audrain M, Chailleux E, Germaud P, et al. Evaluation of the tuberculin skin test and the interferon-γ release assay for TB screening in French healthcare workers. J Occup Med Toxicol Lond Engl. 2009 Nov 30;4:30.

82. Ozdemir D, Annakkaya AN, Tarhan G, Sencan I, Cesur S, Balbay O, et al. Comparison of the tuberculin skin test and the quantiferon test for latent Mycobacterium tuberculosis infections in health care workers in Turkey. Jpn J Infect Dis. 2007 May;60(2–3):102–5.

83. Hill PC, Brookes RH, Fox A, Jackson-Sillah D, Lugos MD, Jeffries DJ, et al. Surprisingly high specificity of the PPD skin test for M. tuberculosis infection from recent exposure in The Gambia. PloS One. 2006 Dec 20;1:e68.

84. Diel R, Loddenkemper R, Nienhaus A. Evidence-based comparison of commercial interferon-gamma release assays for detecting active TB: a metaanalysis. Chest. 2010 Apr;137(4):952–68.

85. Rangaka MX, Wilkinson KA, Glynn JR, Ling D, Menzies D, Mwansa-Kambafwile J, et al. Predictive value of interferon-γ release assays for incident active tuberculosis: a systematic review and meta-analysis. Lancet Infect Dis. 2012 Jan;12(1):45–55.

86. Berry MPR, Graham CM, McNab FW, Xu Z, Bloch SAA, Oni T, et al. An interferon-inducible neutrophil-driven blood transcriptional signature in human tuberculosis. Nature. 2010 Aug 19;466(7309):973–7.

87. Anderson ST, Kaforou M, Brent AJ, Wright VJ, Banwell CM, Chagaluka G, et al. Diagnosis of childhood tuberculosis and host RNA expression in Africa. N Engl J Med. 2014 May 1;370(18):1712–23.

88. Blischak JD, Tailleux L, Myrthil M, Charlois C, Bergot E, Dinh A, et al. Predicting susceptibility to tuberculosis based on gene expression profiling in dendritic cells. Sci Rep. 2017 Jul 18;7(1):5702.

89. Zak DE, Penn-Nicholson A, Scriba TJ, Thompson E, Suliman S, Amon LM, et al. A blood RNA signature for tuberculosis disease risk: a prospective cohort study. Lancet Lond Engl. 2016 Mar 23;

90. Gallant CJ, Cobat A, Simkin L, Black GF, Stanley K, Hughes J, et al. Tuberculin skin test and in vitro assays provide complementary measures of antimycobacterial immunity in children and adolescents. Chest. 2010 May;137(5):1071–7.

91. WHO. Guidelines on the management of latent tuberculosis infection: the end TB strategy [Internet]. 2014 [cited 2017 May 13]. Available from: http://public.eblib.com/choice/publicfullrecord.aspx?p=1910127_0

92. Sharma SK, Vashishtha R, Chauhan LS, Sreenivas V, Seth D. Comparison of TST and IGRA in Diagnosis of Latent Tuberculosis Infection in a High TB-Burden Setting. PloS One. 2017;12(1):e0169539.

93. Dye C, Bassili A, Bierrenbach A, Broekmans J, Chadha V, Glaziou P, et al. Measuring tuberculosis burden, trends, and the impact of control programmes. Lancet Infect Dis. 2008 Apr;8(4):233–43.

94. Rathi SK, Akhtar S, Rahbar MH, Azam SI. Prevalence and risk factors associated with tuberculin skin test positivity among household contacts of smear-positive pulnionary tuberculosis cases in Umerkot, Pakistan. Int J Tuberc Lung Dis Off J Int Union Tuberc Lung Dis. 2002 Oct;6(10):851–7.

95. Madhi F, Fuhrman C, Monnet I, Atassi K, Poirier C, Housset B, et al. Transmission of tuberculosis from adults to children in a Paris suburb. Pediatr Pulmonol. 2002 Sep;34(3):159–63.

96. Rutherford ME, Hill PC, Maharani W, Apriani L, Sampurno H, van Crevel R, et al. Risk factors for Mycobacterium tuberculosis infection in Indonesian children living with a sputum smear-positive case. Int J Tuberc Lung Dis Off J Int Union Tuberc Lung Dis. 2012 Dec;16(12):1594–9.

97. Lienhardt C, Fielding K, Sillah J, Tunkara A, Donkor S, Manneh K, et al. Risk factors for tuberculosis infection in sub-Saharan Africa: a contact study in The Gambia. Am J Respir Crit Care Med. 2003 Aug 15;168(4):448–55.

98. MacIntyre CR, Kendig N, Kummer L, Birago S, Graham NM. Impact of tuberculosis control measures and crowding on the incidence of tuberculous infection in Maryland prisons. Clin Infect Dis Off Publ Infect Dis Soc Am. 1997 Jun;24(6):1060–7.

99. Almeida LM, Barbieri MA, Da Paixão AC, Cuevas LE. Use of purified protein derivative to assess the risk of infection in children in close contact with adults with tuberculosis in a population with high Calmette-Guérin bacillus coverage. Pediatr Infect Dis J. 2001 Nov;20(11):1061–5.

100. Verver S. Transmission of tuberculosis in a high incidence urban community in South Africa. Int J Epidemiol. 2004 Apr 1;33(2):351–7.

101. Glynn JR, Guerra-Assunção JA, Houben RMGJ, Sichali L, Mzembe T, Mwaungulu LK, et al. Whole Genome Sequencing Shows a Low Proportion of Tuberculosis Disease Is Attributable to Known Close Contacts in Rural Malawi. PloS One. 2015;10(7):e0132840.

102. Soysal A, Millington KA, Bakir M, Dosanjh D, Aslan Y, Deeks JJ, et al. Effect of BCG vaccination on risk of Mycobacterium tuberculosis infection in children with household tuberculosis contact: a prospective community-based study. The Lancet. 2005;366(9495):1443–1451.

103. Gagneux S, Small PM. Global phylogeography of Mycobacterium tuberculosis and implications for tuberculosis product development. Lancet Infect Dis. 2007 May;7(5):328–37.

104. Coscolla M, Gagneux S. Does M. tuberculosis genomic diversity explain disease diversity? Drug Discov Today Dis Mech. 2010 SPRING;7(1):e43.

105. Nicol MP, Wilkinson RJ. The clinical consequences of strain diversity in Mycobacterium tuberculosis. Trans R Soc Trop Med Hyg. 2008 Oct 1;102(10):955–65.

106. Dormans J, Burger M, Aguilar D, Hernandez-Pando R, Kremer K, Roholl P, et al. Correlation of virulence, lung pathology, bacterial load and delayed type hypersensitivity responses after infection with different Mycobacterium tuberculosis genotypes in a BALB/c mouse model. Clin Exp Immunol. 2004 Sep;137(3):460–8.

107. Parwati I, van Crevel R, van Soolingen D. Possible underlying mechanisms for successful emergence of the Mycobacterium tuberculosis Beijing genotype strains. Lancet Infect Dis. 2010 Feb;10(2):103–11.

108. de Jong BC, Hill PC, Aiken A, Awine T, Antonio M, Adetifa IM, et al. Progression to active tuberculosis, but not transmission, varies by Mycobacterium tuberculosis lineage in The Gambia. J Infect Dis. 2008 Oct 1;198(7):1037–43.

109. Malik AN, Godfrey-Faussett P. Effects of genetic variability of Mycobacterium tuberculosis strains on the presentation of disease. Lancet Infect Dis. 2005 Mar;5(3):174–83.

110. Di Pietrantonio T, Correa JA, Orlova M, Behr MA, Schurr E. Joint effects of host genetic background and mycobacterial pathogen on susceptibility to infection. Infect Immun. 2011 Jun;79(6):2372–8.

111. Gagneux S, DeRiemer K, Van T, Kato-Maeda M, de Jong BC, Narayanan S, et al. Variable host-pathogen compatibility in Mycobacterium tuberculosis. Proc Natl Acad Sci U S A. 2006 Feb 21;103(8):2869–73.

112. Di Pietrantonio T, Schurr E. Host-pathogen specificity in tuberculosis. Adv Exp Med Biol. 2013;783:33–44.

113. Caws M, Thwaites G, Dunstan S, Hawn TR, Lan NTN, Thuong NTT, et al. The influence of host and bacterial genotype on the development of disseminated disease with Mycobacterium tuberculosis. PLoS Pathog. 2008 Mar 28;4(3):e1000034.

114. van Crevel R, Parwati I, Sahiratmadja E, Marzuki S, Ottenhoff THM, Netea MG, et al. Infection with Mycobacterium tuberculosis Beijing genotype strains is associated with polymorphisms in SLC11A1/NRAMP1 in Indonesian patients with tuberculosis. J Infect Dis. 2009 Dec 1;200(11):1671–4.

115. Intemann CD, Thye T, Niemann S, Browne ENL, Amanua Chinbuah M, Enimil A, et al. Autophagy gene variant IRGM -261T contributes to protection from tuberculosis caused by Mycobacterium tuberculosis but not by M. africanum strains. PLoS Pathog. 2009 Sep;5(9):e1000577.

116. Herb F, Thye T, Niemann S, Browne ENL, Chinbuah MA, Gyapong J, et al. ALOX5 variants associated with susceptibility to human pulmonary tuberculosis. Hum Mol Genet. 2008 Apr 1;17(7):1052–60.

117. Thye T, Niemann S, Walter K, Homolka S, Intemann CD, Chinbuah MA, et al. Variant G57E of mannose binding lectin associated with protection against tuberculosis caused by Mycobacterium africanum but not by M. tuberculosis. PloS One. 2011;6(6):e20908.

118. Valway SE, Sanchez MP, Shinnick TF, Orme I, Agerton T, Hoy D, et al. An outbreak involving extensive transmission of a virulent strain of Mycobacterium tuberculosis. N Engl J Med. 1998 Mar 5;338(10):633–9.

119. Newton SM, Smith RJ, Wilkinson KA, Nicol MP, Garton NJ, Staples KJ, et al. A deletion defining a common Asian lineage of Mycobacterium tuberculosis associates with immune subversion. Proc Natl Acad Sci U S A. 2006 Oct 17;103(42):15594–8.

120. Jenkins HE. Global burden of childhood tuberculosis. Pneumonia [Internet]. 2016 Nov 24 [cited 2017 Jul 21];8. Available from: http://www.ncbi.nlm.nih.gov/pmc/articles/PMC5166554/

121. Cruz AT, Starke JR. Clinical manifestations of tuberculosis in children. Paediatr Respir Rev. 2007 Jun;8(2):107–17.

122. Fine PE, Bruce J, Ponnighaus JM, Nkhosa P, Harawa A, Vynnycky E. Tuberculin sensitivity: conversions and reversions in a rural African population. Int J Tuberc Lung Dis Off J Int Union Tuberc Lung Dis. 1999 Nov;3(11):962–75.

123. Mahomed H, Hughes EJ, Hawkridge T, Minnies D, Simon E, Little F, et al. Comparison of mantoux skin test with three generations of a whole blood IFN-gamma assay for tuberculosis infection. Int J Tuberc Lung Dis Off J Int Union Tuberc Lung Dis. 2006 Mar;10(3):310–6.

124. Lee HW, Lee YJ, Kim SJ, Park JS, Cho Y-J, Yoon HI, et al. Comparing tuberculin skin test and interferon γ release assay (T-SPOT.TB) to diagnose latent tuberculosis infection in household contacts. Korean J Intern Med. 2017 Jan 24;

125. Jabot-Hanin F, Cobat A, Feinberg J, Grange G, Remus N, Poirier C, et al. Major Loci on Chromosomes 8q and 3q Control Interferon γ Production Triggered by Bacillus Calmette-Guerin and 6-kDa Early Secretory Antigen Target, Respectively, in Various Populations. J Infect Dis. 2016 Apr 1;213(7):1173–9.

126. Harries AD, Lin Y, Satyanarayana S, Lönnroth K, Li L, Wilson N, et al. The looming epidemic of diabetes-associated tuberculosis: learning lessons from HIV-associated tuberculosis. Int J Tuberc Lung Dis Off J Int Union Tuberc Lung Dis. 2011 Nov;15(11):1436–1444, i.

127. Lee M-R, Huang Y-P, Kuo Y-T, Luo C-H, Shih Y-J, Shu C-C, et al. Diabetes mellitus and latent tuberculosis infection: a systemic review and meta-analysis. Clin Infect Dis Off Publ Infect Dis Soc Am. 2016 Dec 16;

128. Hensel RL, Kempker RR, Tapia J, Oladele A, Blumberg HM, Magee MJ. Increased risk of latent tuberculous infection among persons with pre-diabetes and diabetes mellitus. Int J Tuberc Lung Dis Off J Int Union Tuberc Lung Dis. 2016 Jan;20(1):71–8.

129. Jick SS, Lieberman ES, Rahman MU, Choi HK. Glucocorticoid use, other associated factors, and the risk of tuberculosis. Arthritis Rheum. 2006 Feb 15;55(1):19–26.

130. Schaible UE, Stefan HE. Malnutrition and infection: complex mechanisms and global impacts. PLoS Med. 2007;4(5):e115.

131. Lin H-H, Ezzati M, Murray M. Tobacco smoke, indoor air pollution and tuberculosis: a systematic review and meta-analysis. PLoS Med. 2007 Jan;4(1):e20.

132. Davis A, Terlikbayeva A, Aifah A, Hermosilla S, Zhumadilov Z, Berikova E, et al. Risks for tuberculosis in Kazakhstan: implications for prevention. Int J Tuberc Lung Dis Off J Int Union Tuberc Lung Dis. 2017 Jan 1;21(1):86–92.

133. Francisco J, Oliveira O, Felgueiras Ó, Gaio AR, Duarte R. How much is too much alcohol in tuberculosis? Eur Respir J. 2017 Jan;49(1).

134. Huang S-J, Wang X-H, Liu Z-D, Cao W-L, Han Y, Ma A-G, et al. Vitamin D deficiency and the risk of tuberculosis: a meta-analysis. Drug Des Devel Ther. 2017;11:91–102.

135. Hernández-Garduño E. Vitamin D deficiency and tuberculosis: what about body mass index? Drug Des Devel Ther. 2017 Apr 11;11:1193–4.

136. The role of BCG vaccine in the prevention and control of tuberculosis in the United States. A joint statement by the Advisory Council for the Elimination of Tuberculosis and the Advisory Committee on Immunization Practices. MMWR Recomm Rep Morb Mortal Wkly Rep Recomm Rep. 1996 Apr 26;45(RR-4):1–18.

137. Vaudry W. “To BCG or not to BCG, that is the question!”. The challenge of BCG vaccination: Why can’t we get it right? Paediatr Child Health. 2003 Mar;8(3):141.

138. Roy A, Eisenhut M, Harris RJ, Rodrigues LC, Sridhar S, Habermann S, et al. Effect of BCG vaccination against Mycobacterium tuberculosis infection in children: systematic review and meta-analysis. BMJ. 2014 Aug 5;349(aug04 5):g4643–g4643.

139. Cobat A, Poirier C, Hoal E, Boland-Auge A, Rocque F de L, Corrard F, et al. Tuberculin Skin Test Negativity Is Under Tight Genetic Control of Chromosomal Region 11p14-15 in Settings With Different Tuberculosis Endemicities. J Infect Dis. 2015 Jan 15;211(2):317–21.

140. Shisana O, Rehle T, Simbayi LC, Zuma K, Jooste S, Zungu N, et al. South African National HIV Prevalence, Incidence and Behaviour Survey, 2012 [Internet]. HSRC Press; 2014 [cited 2017 Mar 20]. Available from: http://repository.hsrc.ac.za/handle/20.500.11910/2490

141. Houk VN, Baker JH, Sorensen K, Kent DC. The epidemiology of tuberculosis infection in a closed environment. Arch Environ Health. 1968 Jan;16(1):26–35.

142. Kassim S, Zuber P, Wiktor SZ, Diomande FV, Coulibaly IM, Coulibaly D, et al. Tuberculin skin testing to assess the occupational risk of Mycobacterium tuberculosis infection among health care workers in Abidjan, Côte d'Ivoire. Int J Tuberc Lung Dis Off J Int Union Tuberc Lung Dis. 2000 Apr;4(4):321–6.

143. Pai M, Gokhale K, Joshi R, Dogra S, Kalantri S, Mendiratta DK, et al. Mycobacterium tuberculosis Infection in Health Care Workers in Rural India: Comparison of a Whole-Blood Interferon γ Assay With Tuberculin Skin Testing. JAMA. 2005 Jun 8;293(22):2746–55.

144. Stead WW, Senner JW, Reddick WT, Lofgren JP. Racial differences in susceptibility to infection by Mycobacterium tuberculosis. N Engl J Med. 1990 Feb 15;322(7):422–7.

145. Visscher PM, Hill WG, Wray NR. Heritability in the genomics era — concepts and misconceptions. Nat Rev Genet. 2008 Mar 4;9(4):255–66.

146. Sepulveda RL, Heiba IM, King A, Gonzalez B, Elston RC, Sorensen RU. Evaluation of tuberculin reactivity in BCG-immunized siblings. Am J Respir Crit Care Med. 1994 Mar;149(3 Pt 1):620–4.

147. Demenais FM, Bonney GE. Equivalence of the mixed and regressive models for genetic analysis. I. Continuous traits. Genet Epidemiol. 1989;6(5):597–617.

148. Sepulveda RL, Heiba IM, Navarrete C, Elston RC, Gonzalez B, Sorensen RU. Tuberculin reactivity after newborn BCG immunization in mono- and dizygotic twins. Tuber Lung Dis Off J Int Union Tuberc Lung Dis. 1994 Apr;75(2):138–43.

149. Jepson A, Fowler A, Banya W, Singh M, Bennett S, Whittle H, et al. Genetic regulation of acquired immune responses to antigens of Mycobacterium tuberculosis: a study of twins in West Africa. Infect Immun. 2001 Jun;69(6):3989–94.

150. Cobat A, Barrera LF, Henao H, Arbeláez P, Abel L, García LF, et al. Tuberculin Skin Test Reactivity Is Dependent on Host Genetic Background in Colombian Tuberculosis Household Contacts. Clin Infect Dis. 2012 Jan 4;54(7):968–71.

151. Cobat A, Gallant CJ, Simkin L, Black GF, Stanley K, Hughes J, et al. High Heritability of Antimycobacterial Immunity in an Area of Hyperendemicity for Tuberculosis Disease. J Infect Dis. 2010 Jan;201(1):15–9.

152. Stein CM, Guwatudde D, Nakakeeto M, Peters P, Elston RC, Tiwari HK, et al. Heritability analysis of cytokines as intermediate phenotypes of tuberculosis. J Infect Dis. 2003 Jun 1;187(11):1679–85.

153. Tao L, Zalwango S, Chervenak K, Thiel B, Malone LL, Qiu F, et al. Genetic and shared environmental influences on interferon-γ production in response to Mycobacterium tuberculosis antigens in a Ugandan population. Am J Trop Med Hyg. 2013 Jul;89(1):169–73.

154. Newport MJ, Goetghebuer T, Weiss HA, Whittle H, Siegrist C-A, Marchant A, et al. Genetic regulation of immune responses to vaccines in early life. Genes Immun. 2004 Mar;5(2):122–9.

155. Cobat A, Gallant CJ, Simkin L, Black GF, Stanley K, Hughes J, et al. Two loci control tuberculin skin test reactivity in an area hyperendemic for tuberculosis. J Exp Med. 2009 Nov 23;206(12):2583–91.

156. Cobat A, Hoal EG, Gallant CJ, Simkin L, Black GF, Stanley K, et al. Identification of a Major Locus, TNF1, That Controls BCG-Triggered Tumor Necrosis Factor Production by Leukocytes in an Area Hyperendemic for Tuberculosis. Clin Infect Dis. 2013 Jan 10;57(7):963–70.

157. Price AL, Patterson NJ, Plenge RM, Weinblatt ME, Shadick NA, Reich D. Principal components analysis corrects for stratification in genome-wide association studies. Nat Genet. 2006 Jul 23;38(8):904–9.

158. Risch N, Zhang H. Extreme discordant sib pairs for mapping quantitative trait loci in humans. Science. 1995 Jun 16;268(5217):1584–9.

159. Corrigall J, Coetzee D, Cameron N. Is the Western Cape at risk of an outbreak of preventable childhood diseases? Lessons from an evaluation of routine immunisation coverage. South Afr Med J Suid-Afr Tydskr Vir Geneeskd. 2008 Jan;98(1):41–5.

160. Kritzinger FE, den Boon S, Verver S, Enarson DA, Lombard CJ, Borgdorff MW, et al. No decrease in annual risk of tuberculosis infection in endemic area in Cape Town, South Africa. Trop Med Int Health. 2009 Feb;14(2):136–42.

161. de Wit E, Delport W, Rugamika CE, Meintjes A, Möller M, van Helden PD, et al. Genome-wide analysis of the structure of the South African Coloured Population in the Western Cape. Hum Genet. 2010 Aug;128(2):145–53.

162. Quintana-Murci L, Harmant C, Quach H, Balanovsky O, Zaporozhchenko V, Bormans C, et al. Strong maternal Khoisan contribution to the South African coloured population: a case of gender-biased admixture. Am J Hum Genet. 2010 Apr 9;86(4):611–20.

163. Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MAR, Bender D, et al. PLINK: a tool set for whole-genome association and population-based linkage analyses. Am J Hum Genet. 2007 Sep;81(3):559–75.

164. Chang CC, Chow CC, Tellier LC, Vattikuti S, Purcell SM, Lee JJ. Second-generation PLINK: rising to the challenge of larger and richer datasets. GigaScience. 2015;4:7.

165. Tregouet D-A. Les equations d'estimation : principes et applications a l'etude de la composante familiale et genetique des maladies multifactorielles (doctorat : epidemiologie genetique) [Internet]. Paris 11; 1999 [cited 2017 Jun 7]. Available from: http://www.theses.fr/1999PA11T001

166. van Zyl-Smit RN, Pai M, Peprah K, Meldau R, Kieck J, Juritz J, et al. Within-subject variability and boosting of T-cell interferon-gamma responses after tuberculin skin testing. Am J Respir Crit Care Med. 2009 Jul 1;180(1):49–58.

167. van Zyl-Smit RN, Zwerling A, Dheda K, Pai M. Within-subject variability of interferon-g assay results for tuberculosis and boosting effect of tuberculin skin testing: a systematic review. PloS One. 2009;4(12):e8517.

168. Detjen AK, Loebenberg L, Grewal HMS, Stanley K, Gutschmidt A, Kruger C, et al. Short-term reproducibility of a commercial interferon gamma release assay. Clin Vaccine Immunol CVI. 2009 Aug;16(8):1170–5.

169. Morton NE. Sequential tests for the detection of linkage. Am J Hum Genet. 1955 Sep;7(3):277–318.

170. Kleensang A, Franke D, Alcaïs A, Abel L, Müller-Myhsok B, Ziegler A. An extensive comparison of quantitative trait Loci mapping methods. Hum Hered. 2010;69(3):202–11.

171. Alcaïs A, Abel L. Maximum-Likelihood-Binomial method for genetic model-free linkage analysis of quantitative traits in sibships. Genet Epidemiol. 1999;17(2):102–17.

172. 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 Jan;35(1):46–56.

173. Abel L, Alcais A, Mallet A. Comparison of four sib-pair linkage methods for analyzing sibships with more than two affecteds: interest of the binomial maximum likelihood approach. Genet Epidemiol. 1998;15(4):371–90.

174. Lander E, Kruglyak L. Genetic dissection of complex traits: guidelines for interpreting and reporting linkage results. Nat Genet. 1995 Nov 1;11(3):241–7.

175. Roberts SB, MacLean CJ, Neale MC, Eaves LJ, Kendler KS. Replication of linkage studies of complex traits: an examination of variation in location estimates. Am J Hum Genet. 1999 Sep;65(3):876–84.

176. Baghdadi JE, Orlova M, Alter A, Ranque B, Chentoufi M, Lazrak F, et al. An autosomal dominant major gene confers predisposition to pulmonary tuberculosis in adults. J Exp Med. 2006 Jul 3;203(7):1679–84.

177. Grant AV, El Baghdadi J, Sabri A, El Azbaoui S, Alaoui-Tahiri K, Abderrahmani Rhorfi I, et al. Age-Dependent Association between Pulmonary Tuberculosis and Common TOX Variants in the 8q12–13 Linkage Region. Am J Hum Genet. 2013 Mar;92(3):407–14.

178. Chimusa ER, Daya M, Möller M, Ramesar R, Henn BM, van Helden PD, et al. Determining ancestry proportions in complex admixture scenarios in South Africa using a novel proxy ancestry selection method. PloS One. 2013;8(9):e73971.

179. Gallegos AM, Pamer EG, Glickman MS. Delayed protection by ESAT-6-specific effector CD4+ T cells after airborne M. tuberculosis infection. J Exp Med. 2008 Sep 29;205(10):2359–68.

180. Jabot-Hanin F. An eQTL variant of ZXDCis associated with IFN- γ production following Mycobacterium tuberculosis antigen-specific stimulation (under review)

181. 1000 Genomes Project Consortium, Abecasis GR, Auton A, Brooks LD, DePristo MA, Durbin RM, et al. An integrated map of genetic variation from 1,092 human genomes. Nature. 2012 Nov 1;491(7422):56–65.

182. Delaneau O, Marchini J, Zagury J-F. A linear complexity phasing method for thousands of genomes. Nat Methods. 2012 Feb;9(2):179–81.

183. Howie B, Fuchsberger C, Stephens M, Marchini J, Abecasis GR. Fast and accurate genotype imputation in genome-wide association studies through pre-phasing. Nat Genet. 2012 Aug;44(8):955–9.

184. Howie BN, Donnelly P, Marchini J. A flexible and accurate genotype imputation method for the next generation of genome-wide association studies. PLoS Genet. 2009 Jun;5(6):e1000529.

185. Marchini J, Howie B. Genotype imputation for genome-wide association studies. Nat Rev Genet. 2010 Jan 7;11(7):499–511.

186. Delaneau O, Marchini J, Consortium T 1000 GP, McVean GA, Donnelly P, Lunter G, et al. Integrating sequence and array data to create an improved 1000 Genomes Project haplotype reference panel. Nat Commun. 2014 Jun 13;5:ncomms4934.

187. Zhou X, Stephens M. Genome-wide efficient mixed-model analysis for association studies. Nat Genet. 2012 Jun 17;44(7):821–4.

188. Eu-Ahsunthornwattana J, Miller EN, Fakiola M, Wellcome Trust Case Control Consortium 2, Jeronimo SMB, Blackwell JM, et al. Comparison of methods to account for relatedness in genome-wide association studies with family-based data. PLoS Genet. 2014 Jul;10(7):e1004445.

189. Zhang Y, Pan W. Principal component regression and linear mixed model in association analysis of structured samples: competitors or complements? Genet Epidemiol. 2015 Mar;39(3):149–55.

190. 1000genomes_phase1_sites_missing_in_phase3 [Internet]. [cited 2017 Mar 17]. Available from: ftp://ftp.1000genomes.ebi.ac.uk/vol1/ftp/release/20130502/supporting/phase1_sites_missing_in_phase3/

191. Machiela MJ, Chanock SJ. LDlink: a web-based application for exploring population-specific haplotype structure and linking correlated alleles of possible functional variants. Bioinforma Oxf Engl. 2015 Nov 1;31(21):3555–7.

192. Boyle AP, Hong EL, Hariharan M, Cheng Y, Schaub MA, Kasowski M, et al. Annotation of functional variation in personal genomes using RegulomeDB. Genome Res. 2012 Sep;22(9):1790–7.

193. Westra H-J, Peters MJ, Esko T, Yaghootkar H, Schurmann C, Kettunen J, et al. Systematic identification of trans eQTLs as putative drivers of known disease associations. Nat Genet. 2013 Sep 8;45(10):1238–43.

194. Trinchieri G. Interleukin-12 and the regulation of innate resistance and adaptive immunity. Nat Rev Immunol. 2003 Feb;3(2):133–46.

195. Al-Kandari W, Koneni R, Navalgund V, Aleksandrova A, Jambunathan S, Fontes JD. The zinc finger proteins ZXDA and ZXDC form a complex that binds CIITA and regulates MHC II gene transcription. J Mol Biol. 2007 Jun 22;369(5):1175–87.

196. Ramsey JE, Fontes JD. The zinc finger transcription factor ZXDC activates CCL2 gene expression by opposing BCL6-mediated repression. Mol Immunol. 2013 Dec;56(4):768–80.

197. Braun MC, Lahey E, Kelsall BL. Selective suppression of IL-12 production by chemoattractants. J Immunol Baltim Md 1950. 2000 Mar 15;164(6):3009–17.

198. Flores-Villanueva PO, Ruiz-Morales JA, Song C-H, Flores LM, Jo E-K, Montaño M, et al. A functional promoter polymorphism in monocyte chemoattractant protein-1 is associated with increased susceptibility to pulmonary tuberculosis. J Exp Med. 2005 Dec 19;202(12):1649–58.

199. Ritz N, Curtis N. Mapping the global use of different BCG vaccine strains. Tuberc Edinb Scotl. 2009 Jul;89(4):248–51.

200. Venkataswamy MM, Goldberg MF, Baena A, Chan J, Jacobs WR, Porcelli SA. In vitro culture medium influences the vaccine efficacy of Mycobacterium bovis BCG. Vaccine. 2012 Feb 1;30(6):1038–49.

201. de Souza GA, Fortuin S, Aguilar D, Pando RH, McEvoy CRE, van Helden PD, et al. Using a label-free proteomics method to identify differentially abundant proteins in closely related hypo- and hypervirulent clinical Mycobacterium tuberculosis Beijing isolates. Mol Cell Proteomics MCP. 2010 Nov;9(11):2414–23.

202. Holland SM, Eisenstein EM, Kuhns DB, Turner ML, Fleisher TA, Strober W, et al. Treatment of refractory disseminated nontuberculous mycobacterial infection with interferon gamma. A preliminary report. N Engl J Med. 1994 May 12;330(19):1348–55.

203. Alangari AA, Al-Zamil F, Al-Mazrou A, Al-Muhsen S, Boisson-Dupuis S, Awadallah S, et al. Treatment of disseminated mycobacterial infection with high-dose IFN-γ in a patient with IL-12Rβ1 deficiency. Clin Dev Immunol. 2011;2011:691956.

204. Comstock GW, Livesay VT, Woolpert SF. The prognosis of a positive tuberculin reaction in childhood and adolescence. Am J Epidemiol. 1974 Feb;99(2):131–8.

205. Casanova J-L, Abel L. Genetic dissection of immunity to mycobacteria : The Human Model. Annu Rev Immunol. 2002;20(1):581–620.

206. Abel L, El-Baghdadi J, Bousfiha AA, Casanova J-L, Schurr E. Human genetics of tuberculosis: a long and winding road. Philos Trans R Soc B Biol Sci. 2014 Jun 19;369(1645):20130428.

207. Borgdorff MW, Sebek M, Geskus RB, Kremer K, Kalisvaart N, van Soolingen D. The incubation period distribution of tuberculosis estimated with a molecular epidemiological approach. Int J Epidemiol. 2011 Aug 1;40(4):964–70.

208. Alcaïs A, Fieschi C, Abel L, Casanova J-L. Tuberculosis in children and adults: two distinct genetic diseases. J Exp Med. 2005 Dec 19;202(12):1617–21.

209. Fox GJ, Orlova M, Schurr E. Tuberculosis in Newborns: The Lessons of the "Lübeck Disaster" (1929-1933). PLoS Pathog. 2016 Jan;12(1):e1005271.

210. Lux M. Perfect Subjects: Race, Tuberculosis, and the Qu?Appelle BCG Vaccine Trial. Can Bull Med Hist. 1998 Oct;15(2):277–95.

211. MacDonald N, Hébert PC, Stanbrook MB. Tuberculosis in Nunavut: a century of failure. CMAJ Can Med Assoc J J Assoc Medicale Can. 2011 Apr 19;183(7):741–3.

212. Kallmann FJ, Reisner D. Twin Studies on Genetic Variations in Resistance to Tuberculosis. J Hered. 1943 Jan 9;34(9):269–76.

213. Comstock GW. Tuberculosis in twins: a re-analysis of the Prophit survey. Am Rev Respir Dis. 1978 Apr;117(4):621–4.

214. Conti F, Lugo-Reyes SO, Blancas Galicia L, He J, Aksu G, Borges de Oliveira E, et al. Mycobacterial disease in patients with chronic granulomatous disease: A retrospective analysis of 71 cases. J Allergy Clin Immunol. 2016 Jul;138(1):241–248.e3.

215. Wu U-I, Holland SM. Host susceptibility to non-tuberculous mycobacterial infections. Lancet Infect Dis. 2015 Aug;15(8):968–80.

216. Jouanguy E, Lamhamedi-Cherradi S, Altare F, Fondanèche MC, Tuerlinckx D, Blanche S, et al. Partial interferon-gamma receptor 1 deficiency in a child with tuberculoid bacillus Calmette-Guérin infection and a sibling with clinical tuberculosis. J Clin Invest. 1997 Dec 1;100(11):2658–64.

217. Bustamante J, Arias AA, Vogt G, Picard C, Galicia LB, Prando C, et al. Germline CYBB mutations that selectively affect macrophages in kindreds with X-linked predisposition to tuberculous mycobacterial disease. Nat Immunol. 2011 Mar;12(3):213–21.

218. de Beaucoudrey L, Samarina A, Bustamante J, Cobat A, Boisson-Dupuis S, Feinberg J, et al. Revisiting human IL-12Rβ1 deficiency: a survey of 141 patients from 30 countries. Medicine (Baltimore). 2010 Nov;89(6):381–402.

219. Boisson-Dupuis S, El Baghdadi J, Parvaneh N, Bousfiha A, Bustamante J, Feinberg J, et al. IL-12Rβ1 deficiency in two of fifty children with severe tuberculosis from Iran, Morocco, and Turkey. PloS One. 2011 Apr 13;6(4):e18524.

220. Azad AK, Sadee W, Schlesinger LS. Innate Immune Gene Polymorphisms in Tuberculosis. Infect Immun. 2012 Oct 1;80(10):3343–59.

221. Sabri A, Grant AV, Cosker K, El Azbaoui S, Abid A, Rhorfi IA, et al. Association study of genes controlling IL-12-dependent IFN-γ immunity: STAT4 alleles increase risk of pulmonary tuberculosis in Morocco. J Infect Dis. 2014;jiu140.

222. Barreiro LB, Neyrolles O, Babb CL, Tailleux L, Quach H, McElreavey K, et al. Promoter variation in the DC-SIGN-encoding gene CD209 is associated with tuberculosis. PLoS Med. 2006 Feb;3(2):e20.

223. Möller M, de Wit E, Hoal EG. Past, present and future directions in human genetic susceptibility to tuberculosis. FEMS Immunol Med Microbiol. 2010 Feb;58(1):3–26.

224. Bellamy R, Ruwende C, Corrah T, McAdam KP, Whittle HC, Hill AV. Variations in the NRAMP1 gene and susceptibility to tuberculosis in West Africans. N Engl J Med. 1998 Mar 5;338(10):640–4.

225. Li X, Yang Y, Zhou F, Zhang Y, Lu H, Jin Q, et al. SLC11A1 (NRAMP1) polymorphisms and tuberculosis susceptibility: updated systematic review and meta-analysis. PloS One. 2011;6(1):e15831.

226. Jamieson SE, Miller EN, Black GF, Peacock CS, Cordell HJ, Howson JMM, et al. Evidence for a cluster of genes on chromosome 17q11-q21 controlling susceptibility to tuberculosis and leprosy in Brazilians. Genes Immun. 2004 Jan;5(1):46–57.

227. Mahasirimongkol S, Yanai H, Mushiroda T, Promphittayarat W, Wattanapokayakit S, Phromjai J, et al. Genome-wide association studies of tuberculosis in Asians identify distinct at-risk locus for young tuberculosis. J Hum Genet. 2012 Jun;57(6):363–7.

228. Aliahmad P, Seksenyan A, Kaye J. The many roles of TOX in the immune system. Curr Opin Immunol. 2012 Apr;24(2):173–7.

229. Thye T, Owusu-Dabo E, Vannberg FO, van Crevel R, Curtis J, Sahiratmadja E, et al. Common variants at 11p13 are associated with susceptibility to tuberculosis. Nat Genet. 2012 Feb 5;44(3):257–9.

230. Thye T, Vannberg FO, Wong SH, Owusu-Dabo E, Osei I, Gyapong J, et al. Genome-wide association analyses identifies a susceptibility locus for tuberculosis on chromosome 18q11.2. Nat Genet. 2010 Aug 8;42(9):739–41.

231. Curtis J, Luo Y, Zenner HL, Cuchet-Lourenço D, Wu C, Lo K, et al. Susceptibility to tuberculosis is associated with variants in the ASAP1 gene encoding a regulator of dendritic cell migration. Nat Genet. 2015 May;47(5):523–7.

232. Chimusa E, Hoal EG. GWAS of ancestry-specific TB risk in the South African Coloured population. Human Molecular Genetics [Internet]. 2013 Oct 8 [cited 2013 Oct 25]; Available from: http://hmg.oxfordjournals.org.gate2.inist.fr/content/early/2013/10/08/hmg.ddt462.full.pdf

233. Grant AV, Sabri A, Abid A, Abderrahmani Rhorfi I, Benkirane M, Souhi H, et al. A genome-wide association study of pulmonary tuberculosis in Morocco. Hum Genet. 2016 Mar;135(3):299–307.

234. Wang X, Tang NL-S, Leung CC, Kam KM, Yew WW, Tam CM, et al. Association of polymorphisms in the Chr18q11.2 locus with tuberculosis in Chinese population. Hum Genet. 2013 Mar 3;132(6):691–5.

235. Dai Y, Zhang X, Pan H, Tang S, Shen H, Wang J. Fine mapping of genetic polymorphisms of pulmonary tuberculosis within chromosome 18q11.2 in the Chinese population: a case-control study. BMC Infect Dis. 2011 Oct 22;11:282.

236. Sveinbjornsson G, Gudbjartsson DF, Halldorsson BV, Kristinsson KG, Gottfredsson M, Barrett JC, et al. HLA class II sequence variants influence tuberculosis risk in populations of European ancestry. Nat Genet. 2016 Feb 1;48(3):318–22.

237. Mahasirimongkol S, Yanai H, Nishida N, Ridruechai C, Matsushita I, Ohashi J, et al. Genome-wide SNP-based linkage analysis of tuberculosis in Thais. Genes Immun. 2009 Jan;10(1):77–83.

238. Sobota RS, Stein CM, Kodaman N, Scheinfeldt LB, Maro I, Wieland-Alter W, et al. A Locus at 5q33.3 Confers Resistance to Tuberculosis in Highly Susceptible Individuals. Am J Hum Genet. 2016 Mar;98(3):514–24.

239. Manolio TA, Collins FS, Cox NJ, Goldstein DB, Hindorff LA, Hunter DJ, et al. Finding the missing heritability of complex diseases. Nature. 2009 Oct 8;461(7265):747–53.

240. WHO | Tuberculosis country profiles [Internet]. [cited 2017 Jun 23]. Available from: http://who.int.gate2.inist.fr/tb/country/data/profiles/en/

241. Li H, Durbin R. Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinforma Oxf Engl. 2009 Jul 15;25(14):1754–60.

242. DePristo MA, Banks E, Poplin R, Garimella KV, Maguire JR, Hartl C, et al. A framework for variation discovery and genotyping using next-generation DNA sequencing data. Nat Genet [Internet]. 2011;43. Available from: http://dx.doi.org/10.1038/ng.806

243. Carson AR, Smith EN, Matsui H, Brækkan SK, Jepsen K, Hansen J-B, et al. Effective filtering strategies to improve data quality from population-based whole exome sequencing studies. BMC Bioinformatics. 2014;15(1):1–15.

244. Crowgey EL, Stabley DL, Chen C, Huang H, Robbins KM, Polson SW, et al. An Integrated Approach for Analyzing Clinical Genomic Variant Data from Next-Generation Sequencing. J Biomol Tech JBT. 2015 Apr;jbt.15-2601-002.

245. Sherry ST, Ward M-H, Kholodov M, Baker J, Phan L, Smigielski EM, et al. dbSNP: the NCBI database of genetic variation. Nucleic Acids Res. 2001 Jan 1;29(1):308–11.

246. Belkadi A, Bolze A, Itan Y, Cobat A, Vincent QB, Antipenko A, et al. Whole-genome sequencing is more powerful than whole-exome sequencing for detecting exome variants. Proc Natl Acad Sci. 2015 Apr 28;112(17):5473–8.

247. Freudenberg-Hua Y, Freudenberg J, Kluck N, Cichon S, Propping P, Nothen MM. Single nucleotide variation analysis in 65 candidate genes for CNS disorders in a representative sample of the European population. Genome Res [Internet]. 2003;13. Available from: http://dx.doi.org/10.1101/gr.1299703

248. Zhang J. Rates of conservative and radical nonsynonymous nucleotide substitutions in mammalian nuclear genes. J Mol Evol. 2000 Jan;50(1):56–68.

249. Cingolani P, Platts A, Wang LL, Coon M, Nguyen T, Wang L, et al. A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff. Fly (Austin). 2012 Apr 1;6(2):80–92.

250. Kircher M, Witten DM, Jain P, O'Roak BJ, Cooper GM, Shendure J. A general framework for estimating the relative pathogenicity of human genetic variants. Nat Genet. 2014 Mar;46(3):310–5.

251. Jabot-Hanin F, Varet H, Tores F, Alcais A, Jais J-P. RFPRED: a random forest approach for prediction of missense variants in human exome. bioRxiv. 2016;037127.

252. Itan Y, Shang L, Boisson B, Ciancanelli MJ, Markle JG, Martinez-Barricarte R, et al. The mutation significance cutoff: gene-level thresholds for variant predictions. Nat Methods. 2016 Feb;13(2):109–10.

253. Bansal V, Libiger O, Torkamani A, Schork NJ. Statistical analysis strategies for association studies involving rare variants. Nat Rev Genet. 2010 Nov;11(11):773–85.

254. Lee S, Abecasis GR, Boehnke M, Lin X. Rare-variant association analysis: study designs and statistical tests. Am J Hum Genet. 2014 Jul 3;95(1):5–23.

255. Morgenthaler S, Thilly WG. A strategy to discover genes that carry multi-allelic or mono-allelic risk for common diseases: A cohort allelic sums test (CAST). Mutat Res Mol Mech Mutagen. 2007 Feb;615(1–2):28–56.

256. Li B, Leal SM. Methods for Detecting Associations with Rare Variants for Common Diseases: Application to Analysis of Sequence Data. Am J Hum Genet. 2008 Sep;83(3):311–21.

257. Madsen BE, Browning SR. A Groupwise Association Test for Rare Mutations Using a Weighted Sum Statistic. Schork NJ, editor. PLoS Genet. 2009 Feb 13;5(2):e1000384.

258. Wu MC, Lee S, Cai T, Li Y, Boehnke M, Lin X. Rare-Variant Association Testing for Sequencing Data with the Sequence Kernel Association Test. Am J Hum Genet. 2011 Jul;89(1):82–93.

259. Lee S, Emond MJ, Bamshad MJ, Barnes KC, Rieder MJ, Nickerson DA, et al. Optimal Unified Approach for Rare-Variant Association Testing with Application to Small-Sample Case-Control Whole-Exome Sequencing Studies. Am J Hum Genet. 2012 Aug;91(2):224–37.

260. Hu H, Huff CD, Moore B, Flygare S, Reese MG, Yandell M. VAAST 2.0: Improved Variant Classification and Disease-Gene Identification Using a Conservation-Controlled Amino Acid Substitution Matrix: VAAST 2.0. Genet Epidemiol. 2013 Sep;37(6):622–34.

261. Basu S, Pan W. Comparison of Statistical Tests for Disease Association with Rare Variants. Genet Epidemiol. 2011 Nov;35(7):606–19.

262. Lek M, Karczewski KJ, Minikel EV, Samocha KE, Banks E, Fennell T, et al. Analysis of protein-coding genetic variation in 60,706 humans. Nature. 2016 Aug 18;536(7616):285–91.

263. Zuk O, Schaffner SF, Samocha K, Do R, Hechter E, Kathiresan S, et al. Searching for missing heritability: designing rare variant association studies. Proc Natl Acad Sci U S A. 2014 Jan 28;111(4):E455-464.

264. Babron M-C, de Tayrac M, Rutledge DN, Zeggini E, Génin E. Rare and Low Frequency Variant Stratification in the UK Population: Description and Impact on Association Tests. PLoS ONE [Internet]. 2012 Oct 5 [cited 2016 Jun 27];7(10). Available from: http://www.ncbi.nlm.nih.gov/pmc/articles/PMC3465327/

265. Nhamoyebonde S, Leslie A. Biological Differences Between the Sexes and Susceptibility to Tuberculosis. J Infect Dis. 2014 Jul 15;209(suppl_3):S100–6.

266. Kim JH, Yoo BC, Yang WS, Kim E, Hong S, Cho JY. The Role of Protein Arginine Methyltransferases in Inflammatory Responses. Mediators Inflamm [Internet]. 2016 [cited 2017 Jul 31];2016. Available from: https://www-ncbi-nlm-nih-gov.gate2.inist.fr/pmc/articles/PMC4793140/

267. Veyrieras J-B, Kudaravalli S, Kim SY, Dermitzakis ET, Gilad Y, Stephens M, et al. High-resolution mapping of expression-QTLs yields insight into human gene regulation. PLoS Genet. 2008 Oct;4(10):e1000214.

268. Consortium TGte. The Genotype-Tissue Expression (GTEx) pilot analysis: Multitissue gene regulation in humans. Science. 2015 May 8;348(6235):648–60.

269. Arnett HA, Escobar SS, Gonzalez-Suarez E, Budelsky AL, Steffen LA, Boiani N, et al. BTNL2, a butyrophilin/B7-like molecule, is a negative costimulatory molecule modulated in intestinal inflammation. J Immunol Baltim Md 1950. 2007 Feb 1;178(3):1523–33.

270. Valentonyte R, Hampe J, Huse K, Rosenstiel P, Albrecht M, Stenzel A, et al. Sarcoidosis is associated with a truncating splice site mutation in BTNL2. Nat Genet. 2005 Apr;37(4):357–64.

271. Tong X, Ma Y, Niu X, Yan Z, Liu S, Peng B, et al. The BTNL2 G16071A gene polymorphism increases granulomatous disease susceptibility: A meta-analysis including FPRP test of 8710 participants. Medicine (Baltimore). 2016 Jul;95(30):e4325.

272. Wennerström A, Pietinalho A, Lasota J, Salli K, Surakka I, Seppänen M, et al. Major histocompatibility complex class II and BTNL2 associations in sarcoidosis. Eur Respir J. 2013 Aug 1;42(2):550–3.

273. Prescott NJ, Lehne B, Stone K, Lee JC, Taylor K, Knight J, et al. Pooled sequencing of 531 genes in inflammatory bowel disease identifies an associated rare variant in BTNL2 and implicates other immune related genes. PLoS Genet. 2015 Feb;11(2):e1004955.

274. Mochida A, Kinouchi Y, Negoro K, Takahashi S, Takagi S, Nomura E, et al. Butyrophilin-like 2 gene is associated with ulcerative colitis in the Japanese under strong linkage disequilibrium with HLA-DRB1*1502. Tissue Antigens. 2007 Aug;70(2):128–35.

275. Möller M, Kwiatkowski R, Nebel A, van Helden PD, Hoal EG, Schreiber S. Allelic variation in BTNL2 and susceptibility to tuberculosis in a South African population. Microbes Infect Inst Pasteur. 2007 Apr;9(4):522–8.

276. Lian Y, Yue J, Han M, Liu J, Liu L. Analysis of the association between BTNL2 polymorphism and tuberculosis in Chinese Han population. Infect Genet Evol J Mol Epidemiol Evol Genet Infect Dis. 2010 May;10(4):517–21.

277. Shiina T, Hosomichi K, Inoko H, Kulski JK. The HLA genomic loci map: expression, interaction, diversity and disease. J Hum Genet. 2009 Jan 9;54(1):15–39.

278. Begovich AB, McClure GR, Suraj VC, Helmuth RC, Fildes N, Bugawan TL, et al. Polymorphism, recombination, and linkage disequilibrium within the HLA class II region. J Immunol. 1992 Jan 1;148(1):249–58.

279. Oliveira-Cortez A, Melo AC, Chaves VE, Condino-Neto A, Camargos P. Do HLA class II genes protect against pulmonary tuberculosis? A systematic review and meta-analysis. Eur J Clin Microbiol Infect Dis Off Publ Eur Soc Clin Microbiol. 2016 Oct;35(10):1567–80.

280. González-Galarza FF, Takeshita LYC, Santos EJM, Kempson F, Maia MHT, da Silva ALS, et al. Allele frequency net 2015 update: new features for HLA epitopes, KIR and disease and HLA adverse drug reaction associations. Nucleic Acids Res. 2015 Jan;43(Database issue):D784-788.

281. Warren RL, Choe G, Freeman DJ, Castellarin M, Munro S, Moore R, et al. Derivation of HLA types from shotgun sequence datasets. Genome Med. 2012 Dec 10;4(12):95.

282. Liu C, Yang X, Duffy B, Mohanakumar T, Mitra RD, Zody MC, et al. ATHLATES: accurate typing of human leukocyte antigen through exome sequencing. Nucleic Acids Res. 2013 Aug 1;41(14):e142–e142.

283. Huang Y, Yang J, Ying D, Zhang Y, Shotelersuk V, Hirankarn N, et al. HLAreporter: a tool for HLA typing from next generation sequencing data. Genome Med. 2015;7:25.

284. Dilthey AT, Gourraud P-A, Mentzer AJ, Cereb N, Iqbal Z, McVean G. High-Accuracy HLA Type Inference from Whole-Genome Sequencing Data Using Population Reference Graphs. PLOS Comput Biol. 2016 Oct 28;12(10):e1005151.

285. Emond MJ, Louie T, Emerson J, Zhao W, Mathias RA, Knowles MR, et al. Exome sequencing of extreme phenotypes identifies DCTN4 as a modifier of chronic Pseudomonas aeruginosa infection in cystic fibrosis. Nat Genet. 2012 août;44(8):886–9.

286. McLaren CE, Emond MJ, Subramaniam VN, Phatak PD, Barton JC, Adams PC, et al. Exome sequencing in HFE C282Y homozygous men with extreme phenotypes identifies a GNPAT variant associated with severe iron overload. Hepatol Baltim Md. 2015 Jan 20;

287. Zhou S, Ambalavanan A, Rochefort D, Xie P, Bourassa CV, Hince P, et al. RNF213 Is Associated with Intracranial Aneurysms in the French-Canadian Population. Am J Hum Genet. 2016 Nov 3;99(5):1072–85.

288. Schulte EC, Kousi M, Tan PL, Tilch E, Knauf F, Lichtner P, et al. Targeted Resequencing and Systematic In Vivo Functional Testing Identifies Rare Variants in MEIS1 as Significant Contributors to Restless Legs Syndrome. Am J Hum Genet. 2014 Jul 3;95(1):85–95.

289. Auer PL, Teumer A, Schick U, O'Shaughnessy A, Lo KS, Chami N, et al. Rare and low-frequency coding variants in CXCR2 and other genes are associated with hematological traits. Nat Genet. 2014;46(6):629–34.

290. Gao L, Emond MJ, Louie T, Cheadle C, Berger AE, Rafaels N, et al. Whole-exome sequencing identifies rare variants in ATP8B4 as a risk factor for systemic sclerosis. Arthritis Rheumatol Hoboken NJ. 2015 Oct 16;

291. Lienhardt C, Glaziou P, Uplekar M, Lönnroth K, Getahun H, Raviglione M. Global tuberculosis control: lessons learnt and future prospects. Nat Rev Microbiol. 2012 May 14;10(6):407–16.

292. Medina E, North RJ. Genetically susceptible mice remain proportionally more susceptible to tuberculosis after vaccination. Immunology. 1999 Jan;96(1):16–21.

293. Verver S, Warren RM, Beyers N, Richardson M, van der Spuy GD, Borgdorff MW, et al. Rate of reinfection tuberculosis after successful treatment is higher than rate of new tuberculosis. Am J Respir Crit Care Med. 2005 Jun 15;171(12):1430–5.

# Annexes 1

## Tableaux supplémentaires

Tableau S1 : Gènes et microARNs localisés sur le chromosome 8, dans la région de liaison du phénotype IFNγ-BCG.

| Début (bp) | Fin (bp) | Symbole | |
|---|---|---|---|
| 61101423 | 61193954 | CA8 | carbonic anhydrase VIII |
| 61429469 | 61536203 | RAB2A | RAB2A, member RAS oncogene family |
| 61591321 | 61780587 | CHD7 | chromodomain helicase DNA binding protein 7 |
| 62200514 | 62414204 | CLVS1 | clavesin 1 |
| 62413115 | 62627199 | ASPH | aspartate beta-hydroxylase |
| 62627347 | 62627418 | MIR4470 | microRNA 4470 |
| 63161501 | 63912211 | NKAIN3 | Na+/K+ transporting ATPase interacting 3 |
| 63927638 | 63951610 | GGH | gamma-glutamyl hydrolase (conjugase, folylpolygammaglutamyl hydrolase) |
| 63972047 | 63998612 | TTPA | tocopherol (alpha) transfer protein |
| 64081112 | 64125346 | YTHDF3 | YTH N(6)-methyladenosine RNA binding protein 3 |
| 65492795 | 65496191 | BHLHE22 | basic helix-loop-helix family, member e22 |
| 65508529 | 65711348 | CYP7B1 | cytochrome P450, family 7 subfamily B, polypeptide 1 |
| 66514691 | 66546452 | ARMC1 | armadillo repeat containing 1 |
| 66556888 | 66622798 | MTFR1 | mitochondrial fission regulator 1 |
| 66626569 | 66753969 | PDE7A | phosphodiesterase 7A |
| 66933791 | 67012755 | DNAJC5B | DnaJ (Hsp40) homolog, subfamily C, member 5 beta |
| 67039131 | 67087720 | TRIM55 | tripartite motif containing 55 |
| 67088612 | 67090846 | CRH | corticotropin releasing hormone |
| 67344693 | 67384260 | ADHFE1 | alcohol dehydrogenase, iron containing, 1 |
| 67405491 | 67430759 | C8orf46 | chromosome 8 open reading frame 46 |
| 67474410 | 67525484 | MYBL1 | |
| 67542488 | 67579452 | VCPIP1 | valosin containing protein (p97)/p47 complex interacting protein 1 |
| 67579787 | 67597797 | C8orf44 | chromosome 8 open reading frame 44 |
| 67579787 | 67774257 | C8orf44-SGK3 | readthrough |
| 67624653 | 67774257 | SGK3 | serum/glucocorticoid regulated kinase family, member 3 |
| 67782984 | 67834283 | MCMDC2 | minichromosome maintenance domain containing 2 |
| 67834165 | 67837778 | SNHG6 | small nucleolar RNA host gene 6 |
| 67834709 | 67834784 | SNORD87 | small nucleolar RNA, C/D box 87 |
| 67858736 | 67874825 | TCF24 | transcription factor 24 |
| 67876280 | 67940804 | PPP1R42 | protein phosphatase 1 regulatory subunit 42 |
| 67955314 | 67974562 | COPS5 | COP9 constitutive photomorphogenic homolog subunit 5 |
| 67976588 | 68108849 | CSPP1 | centrosome and spindle pole associated protein 1 |
| 68085746 | 68255912 | ARFGEF1 | ADP-ribosylation factor guanine nucleotide-exchange factor 1 (brefeldin A-inhibited) |
| 68334405 | 68658620 | CPA6 | carboxypeptidase A6 |
| 68864244 | 69143897 | PREX2 | phosphatidylinositol-3,4,5-trisphosphate-dependent Rac exchange factor 2 |
| 69242957 | 69731258 | C8orf34 | chromosome 8 open reading frame 34 |
| 70072340 | 70072414 | TRE-CTC14-1 | transfer RNA-Glu (CTC) 14-1 |
| 70378859 | 70573147 | SULF1 | sulfatase 1 |
| 70584110 | 70747299 | SLCO5A1 | solute carrier organic anion transporter family, member 5A1 |
| 70963886 | 70983562 | PRDM14 | PR domain containing 14 |

| | | | |
|---|---|---|---|
| 71021997 | 71316062 | NCOA2 | nuclear receptor coactivator 2 |
| 71485453 | 71520694 | TRAM1 | translocation associated membrane protein 1 |
| 71549501 | 71581447 | LACTB2 | lactamase, beta 2 |
| 71581600 | 71648177 | XKR9 | XK, Kell blood group complex subunit-related family, member 9 |
| 72109668 | 72459888 | EYA1 | EYA transcriptional coactivator and phosphatase 1 |
| 72753777 | 72756731 | MSC | hCG_18651 |
| 72933486 | 72987819 | TRPA1 | transient receptor potential cation channel, subfamily A, member 1 |
| 73449626 | 73850584 | KCNB2 | potassium channel, voltage gated Shab related subfamily B, member 2 |
| 73921097 | 73959987 | TERF1 | telomeric repeat binding factor (NIMA-interacting) 1 |
| 73976778 | 74005507 | SBSPON | somatomedin B and thrombospondin, type 1 domain containing |
| 74153623 | 74174150 | C8orf89 | chromosome 8 open reading frame 89 |
| 74202874 | 74205869 | RPL7 | ribosomal protein L7 |
| 74206837 | 74237520 | RDH10 | retinol dehydrogenase 10 (all-trans) |
| 74332604 | 74659943 | STAU2 | staufen double-stranded RNA binding protein 2 |
| 74692332 | 74791145 | UBE2W | ubiquitin-conjugating enzyme E2W (putative) |
| 74857373 | 74884522 | TCEB1 | transcription elongation factor B (SIII), polypeptide 1 (15kDa, elongin C) |
| 74888377 | 74895018 | TMEM70 | transmembrane protein 70 |
| 74903564 | 74941314 | LY96 | lymphocyte antigen 96 |
| 75146935 | 75233721 | JPH1 | junctophilin 1 |
| 75262618 | 75279345 | GDAP1 | ganglioside induced differentiation associated protein 1 |
| 75460778 | 75460852 | MIR5681A | 5681a |
| 75460785 | 75460844 | MIR5681B | 5681b |
| 75512101 | 75670587 | MIR2052HG | MIR2052 host gene |
| 75617928 | 75617982 | MIR2052 | 2052 |
| 75736772 | 75767264 | PI15 | peptidase inhibitor 15 |
| 75896708 | 75946793 | CRISPLD1 | cysteine-rich secretory protein LCCL domain containing 1 |
| 76135352 | 76190716 | CASC9 | cancer susceptibility candidate 9 (non-protein coding) |
| 76452203 | 76479069 | HNF4G | hepatocyte nuclear factor 4 gamma |
| 77593515 | 77779521 | ZFHX4 | zinc finger homeobox 4 |
| 77892494 | 77913280 | PEX2 | peroxisomal biogenesis factor 2 |
| 79428336 | 79517502 | PKIA | protein kinase (cAMP-dependent, catalytic) inhibitor alpha |
| 79578282 | 79632000 | ZC2HC1A | zinc finger, C2HC-type containing 1A |
| 79645007 | 79717758 | IL7 | interleukin 7 |
| 80523049 | 80578410 | STMN2 | stathmin 2 |
| 80676245 | 80680098 | HEY1 | hes-related family bHLH transcription factor with YRPW motif 1 |
| 80680377 | 80715036 | LINC01607 | intergenic non-protein coding RNA 1607 |
| 80830952 | 80942516 | MRPS28 | mitochondrial ribosomal protein S28 |
| 80947103 | 81083836 | TPD52 | tumor protein D52 |
| 81153624 | 81153708 | MIR5708 | 5708 |
| 81397854 | 81438500 | ZBTB10 | zinc finger and BTB domain containing 10 |
| 81540686 | 81787016 | ZNF704 | zinc finger protein 704 |
| 81880045 | 82024303 | PAG1 | phosphoprotein membrane anchor with glycosphingolipid microdomains 1 |
| 82192718 | 82197012 | FABP5 | fatty acid binding protein 5 (psoriasis-associated) |
| 82352561 | 82359719 | PMP2 | peripheral myelin protein 2 |
| 82370618 | 82373758 | FABP9 | fatty acid binding protein 9 testis |
| 82390732 | 82395473 | FABP4 | fatty acid binding protein 4 adipocyte |
| 82437216 | 82446056 | FABP12 | OTTHUMP00000227202 |
| 82569151 | 82599029 | IMPA1 | inositol(myo)-1(or 4)-monophosphatase 1 |
| 82605891 | 82607207 | SLC10A5 | solute carrier family 10 member 5 |
| 82613566 | 82633539 | ZFAND1 | zinc finger, AN1-type domain 1 |
| 82644688 | 82671750 | CHMP4C | Snf7 homologue associated with Alix 3 |
| 82711816 | 82754539 | SNX16 | sorting nexin 16 |
| 85095453 | 85834079 | RALYL | RALY RNA binding protein-like |
| 86019323 | 86058315 | LRRCC1 | leucine rich repeat and coiled-coil centrosomal protein 1 |

| | | | |
|---|---|---|---|
| 86089619 | 86126753 | E2F5 | E2F transcription factor 5 p130-binding |
| 86126288 | 86132643 | C8orf59 | chromosome 8 open reading frame 59 |
| 86157716 | 86196302 | CA13 | carbonic anhydrase XIII |
| 86240458 | 86290342 | CA1 | carbonic anhydrase I |
| 86351056 | 86361269 | CA3 | carbonic anhydrase III, muscle specific |
| 86376131 | 86393721 | CA2 | carbonic anhydrase II |
| 87060691 | 87081851 | PSKH2 | hCG_1651138 |
| 87111139 | 87166454 | ATP6V0D2 | H+ transporting, lysosomal 38kDa, V0 subunit d2 |
| 87226288 | 87242609 | SLC7A13 | solute carrier family 7 (anionic amino acid transporter), member 13 |
| 87354965 | 87480181 | WWP1 | WW domain containing E3 ubiquitin protein ligase 1 |
| 87479627 | 87526567 | RMDN1 | hRMD-1\|microtubule-associated protein\|regulator of microtubule dynamics 1\|regulator of microtubule dynamics protein 1 |
| 87493790 | 87495257 | NTAN1P2 | asparagine amidase pseudogene 2 |
| 87526656 | 87573726 | CPNE3 | copine III |
| 87586163 | 87755903 | CNGB3 | cyclic nucleotide gated channel beta 3 |
| 87878676 | 88394955 | CNBD1 | cyclic nucleotide binding domain containing 1 |
| 88882971 | 88886296 | DCAF4L2 | DDB1 and CUL4 associated factor 4-like 2 |
| 89049460 | 89339717 | MMP16 | matrix metallopeptidase 16 (membrane-inserted) |
| 90769335 | 90803292 | RIPK2 | receptor-interacting serine-threonine kinase 2 |
| 90914096 | 90940096 | OSGIN2 | oxidative stress induced growth inhibitor family member 2 |
| 90945564 | 90996952 | NBN | nibrin |
| 91013580 | 91064232 | DECR1 | 2,4-dienoyl CoA reductase 1 mitochondrial |
| 91070836 | 91095107 | CALB1 | calbindin 1 28kDa |

Tableau S2 : Gènes et microARNs localisés sur le chromosome 3, dans la région de liaison du phénotype IFNγ-ESAT6$_{BCG}$.

| Début (bp) | Fin (bp) | Symbole | |
|---|---|---|---|
| 111393523 | 111565294 | PLCXD2 | phosphatidylinositol-specific phospholipase C, X domain containing 2 |
| 111451327 | 111695364 | PHLDB2 | pleckstrin homology-like domain, family B, member 2 |
| 111697723 | 111712215 | ABHD10 | abhydrolase domain containing 10 |
| 111717586 | 111732735 | TAGLN3 | transgelin 3 |
| 111758465 | 111800116 | TMPRSS7 | matriptase-3 |
| 111805175 | 111837073 | C3orf52 | chromosome 3 open reading frame 52 |
| 111831648 | 111831745 | MIR567 | microRNA 567 |
| 111839688 | 111852453 | GCSAM | germinal center-associated, signaling and motility |
| 111859752 | 112013105 | SLC9C1 | solute carrier family 9, subfamily C (Na+-transporting carboxylic acid decarboxylase), member 1 |
| 112051194 | 112081659 | CD200 | CD200 molecule |
| 112182813 | 112218438 | BTLA | B and T lymphocyte associated |
| 112251354 | 112280810 | ATG3 | autophagy related 3 |
| 112280857 | 112303286 | SLC35A5 | solute carrier family 35, member A5 |
| 112323407 | 112359977 | CCDC80 | coiled-coil domain containing 80 |
| 112534556 | 112564797 | CD200R1L | CD200 receptor 1-like |
| 112641532 | 112693950 | CD200R1 | CD200 receptor 1 |
| 112709323 | 112720221 | GTPBP8 | GTP-binding protein 8 (putative) |
| 112721291 | 112738555 | C3orf17 | chromosome 3 open reading frame 17 |
| 112929850 | 113006306 | BOC | BOC cell adhesion associated, oncogene regulated |
| 113005777 | 113160986 | CFAP44 | cilia and flagella associated protein 44 |
| 113161565 | 113234034 | SPICE1 | spindle and centriole associated protein 1 |
| 113251218 | 113348422 | SIDT1 | SID1 transmembrane family, member 1 |
| 113313723 | 113313789 | MIR4446 | microRNA 4446 |
| 113367232 | 113415493 | KIAA2018 | KIAA2018 |
| 113435307 | 113465120 | NAA50 | N(alpha)-acetyltransferase 50, NatE catalytic subunit |
| 113465866 | 113530905 | ATP6V1A | ATPase, H+ transporting, lysosomal 70kDa, V1 subunit A |
| 113557671 | 113666021 | GRAMD1C | GRAM domain containing 1C |
| 113666748 | 113705706 | ZDHHC23 | zinc finger, DHHC-type containing 23 |
| 113682984 | 113775460 | KIAA1407 | KIAA1407 |
| 113775582 | 113807269 | QTRTD1 | queuine tRNA-ribosyltransferase domain containing 1 |
| 113847499 | 113918254 | DRD3 | dopamine receptor D3 |
| 113953370 | 113956425 | ZNF80 | zinc finger protein 80 |
| 114012833 | 114029135 | TIGIT | T cell immunoreceptor with Ig and ITIM domains |
| 114033348 | 114866132 | ZBTB20 | zinc finger and BTB domain containing 20 |
| 114035322 | 114035416 | MIR568 | microRNA 568 |
| 114462292 | 114462372 | MIR4796 | microRNA 4796 |
| 115342151 | 115440334 | GAP43 | growth associated protein 43 |
| 115521210 | 116164385 | LSAMP | limbic system-associated membrane protein |
| 116428635 | 116435887 | TUSC7 | tumor suppressor candidate 7 (non-protein coding) |

| | | | |
|---|---|---|---|
| 116569124 | 116569214 | MIR4447 | microRNA 4447 |
| 118619477 | 118864915 | IGSF11 | immunoglobulin superfamily, member 11 |
| 118864997 | 118879674 | C3orf30 | chromosome 3 open reading frame 30 |
| 118892425 | 118924000 | UPK1B | uroplakin 1B |
| 118930589 | 118959754 | B4GALT4 | UDP-Gal:betaGlcNAc beta 1,4- galactosyltransferase, polypeptide 4 |
| 119013220 | 119138323 | ARHGAP31 | Rho GTPase activating protein 31 |
| 119147807 | 119182529 | TMEM39A | transmembrane protein 39A |
| 119187785 | 119213555 | POGLUT1 | protein O-glucosyltransferase 1 |
| 119217324 | 119243128 | TIMMDC1 | translocase of inner mitochondrial membrane domain containing 1 |
| 119243140 | 119278481 | CD80 | CD80 molecule |
| 119298280 | 119308792 | ADPRH | ADP-ribosylarginine hydrolase |
| 119316695 | 119348658 | PLA1A | phospholipase A1 member A |
| 119360899 | 119379437 | POPDC2 | popeye domain containing 2 |
| 119388372 | 119396243 | COX17 | COX17 cytochrome c oxidase copper chaperone |
| 119421869 | 119485949 | MAATS1 | MYCBP-associated, testis expressed 1 |
| 119499331 | 119552295 | NR1I2 | nuclear receptor subfamily 1, group I, member 2 |
| 119540800 | 119813264 | GSK3B | glycogen synthase kinase 3 beta |
| 119884328 | 120003921 | GPR156 | G protein-coupled receptor 156 |
| 120043576 | 120068186 | LRRC58 | leucine rich repeat containing 58 |
| 120113061 | 120169918 | FSTL1 | follistatin-like 1 |
| 120114515 | 120114576 | MIR198 | microRNA 198 |
| 120315128 | 120321258 | NDUFB4 | NADH dehydrogenase (ubiquinone) 1 beta subcomplex, 4, 15kDa |
| 120347015 | 120401418 | HGD | OTTHUMP00000215350 |
| 120405528 | 120461384 | RABL3 | RAB, member of RAS oncogene family-like 3 |
| 120461558 | 120502216 | GTF2E1 | general transcription factor IIE, polypeptide 1, alpha 56kDa |
| 120627050 | 121143608 | STXBP5L | syntaxin binding protein 5-like |
| 120768487 | 120768562 | MIR5682 | microRNA 5682 |
| 121150273 | 121264853 | POLQ | polymerase (DNA directed), theta |
| 121281269 | 121309469 | ARGFX | arginine-fifty homeobox |
| 121312170 | 121349139 | FBXO40 | F-box protein 40 |
| 121350246 | 121379791 | HCLS1 | hematopoietic cell-specific Lyn substrate 1 |
| 121382046 | 121468662 | GOLGB1 | golgin B1 |
| 121488608 | 121553926 | IQCB1 | IQ motif containing B1 |
| 121554030 | 121605373 | EAF2 | ELL associated factor 2 |
| 121613171 | 121663034 | SLC15A2 | solute carrier family 15 (oligopeptide transporter), member 2 |
| 121706170 | 121741127 | ILDR1 | immunoglobulin-like domain containing receptor 1 |
| 121774209 | 121839990 | CD86 | CD86 molecule |
| 121902530 | 122005350 | CASR | calcium-sensing receptor |
| 122044011 | 122060816 | CSTA | cystatin A (stefin A) |
| 122078436 | 122102074 | CCDC58 | coiled-coil domain containing 58 |
| 122103023 | 122128961 | FAM162A | family with sequence similarity 162, member A |

| | | | |
|---|---|---|---|
| 122130700 | 122134882 | WDR5B | WD repeat domain 5B |
| 122140748 | 122233786 | KPNA1 | karyopherin alpha 1 (importin alpha 5) |
| 122246757 | 122283523 | PARP9 | poly (ADP-ribose) polymerase family, member 9 |
| 122283085 | 122294050 | DTX3L | deltex 3 like, E3 ubiquitin ligase |
| 122296435 | 122357894 | PARP15 | poly (ADP-ribose) polymerase family, member 15 |
| 122399672 | 122449687 | PARP14 | poly (ADP-ribose) polymerase family, member 14 |
| 122458844 | 122512666 | HSPBAP1 | HSPB (heat shock 27kDa) associated protein 1 |
| 122513901 | 122599986 | DIRC2 | disrupted in renal carcinoma 2 |
| 122628040 | 122747452 | SEMA5B | sema domain, seven thrombospondin repeats (type 1 and type 1-like), transmembrane domain (TM) and short cytoplasmic domain, (semaphorin) 5B |
| 122785856 | 122880953 | PDIA5 | protein disulfide isomerase family A, member 5 |
| 122920774 | 122992983 | SEC22A | SEC22 vesicle trafficking protein homolog A (S. cerevisiae) |
| 123001143 | 123167924 | ADCY5 | adenylate cyclase 5 |
| 123213363 | 123303924 | HACD2 | 3-hydroxyacyl-CoA dehydratase 2\|OTTHUMP00000215716\|protein-tyrosine phosphatase-like member B |
| 123331143 | 123603149 | MYLK | myosin light chain kinase |
| 123616152 | 123680255 | CCDC14 | coiled-coil domain containing 14 |
| 123687862 | 123711017 | ROPN1 | rhophilin associated tail protein 1 |
| 123813558 | 124440036 | KALRN | kalirin, RhoGEF kinase |
| 123851776 | 123851872 | MIR5002 | microRNA 5002 |
| 124449213 | 124468120 | UMPS | uridine monophosphate synthetase |
| 124451286 | 124451363 | MIR544B | microRNA 544b |
| 124480795 | 124606500 | ITGB5 | integrin, beta 5 |
| 124624289 | 124653595 | MUC13 | mucin 13, cell surface associated |
| 124684554 | 124774802 | HEG1 | heart development protein with EGF-like domains 1 |
| 124801480 | 124931609 | SLC12A8 | solute carrier family 12 (potassium/chloride transporters), member 8 |
| 124870309 | 124870396 | MIR5092 | microRNA 5092 |
| 124944513 | 125094198 | ZNF148 | zinc finger protein 148 |
| 125165488 | 125239058 | SNX4 | sorting nexin 4 |
| 125247702 | 125314381 | OSBPL11 | oxysterol binding protein-like 11 |
| 125509247 | 125509395 | MIR548I1 | microRNA 548i-1 |
| 125647459 | 125709383 | ALG1L | ALG1, chitobiosyldiphosphodolichol beta-mannosyltransferase-like |
| 125687987 | 125702297 | ROPN1B | OTTHUMP00000222455 |
| 125725200 | 125820398 | SLC41A3 | solute carrier family 41, member 3 |
| 125822404 | 125900029 | ALDH1L1 | aldehyde dehydrogenase 1 family, member L1 |
| 126061478 | 126076236 | KLF15 | Kruppel-like factor 15 |
| 126113751 | 126155399 | CCDC37 | coiled-coil domain containing 37 |
| 126156444 | 126194773 | ZXDC | ZXD family zinc finger C |
| 126200008 | 126236616 | UROC1 | urocanate hydratase 1 |
| 126243131 | 126262134 | CHST13 | carbohydrate (chondroitin 4) sulfotransferase 13 |
| 126245842 | 126277808 | C3orf22 | chromosome 3 open reading frame 22 |

| | | | |
|---|---|---|---|
| 126290622 | 126327398 | TXNRD3NB | thioredoxin reductase 3 neighbor |
| 126325895 | 126375056 | TXNRD3 | thioredoxin reductase 3 |
| 126423063 | 126679263 | CHCHD6 | coiled-coil-helix-coiled-coil-helix domain containing 6 |
| 126707437 | 126756235 | PLXNA1 | plexin A1 |
| 126911974 | 126917028 | C3orf56 | chromosome 3 open reading frame 56 |
| 127291907 | 127309602 | TPRA1 | transmembrane protein, adipocyte asscociated 1 |
| 127317200 | 127341279 | MCM2 | minichromosome maintenance complex component 2 |
| 127348002 | 127391653 | PODXL2 | podocalyxin-like 2 |
| 127391781 | 127399769 | ABTB1 | ankyrin repeat and BTB (POZ) domain containing 1 |
| 127407905 | 127542093 | MGLL | monoglyceride lipase |
| 127634075 | 127706514 | KBTBD12 | kelch repeat and BTB (POZ) domain containing 12 |
| 127770402 | 127790526 | SEC61A1 | Sec61 alpha 1 subunit (S. cerevisiae) |
| 127799800 | 127872752 | RUVBL1 | RuvB-like AAA ATPase 1 |
| 127872302 | 128127489 | EEFSEC | eukaryotic elongation factor, selenocysteine-tRNA-specific |
| 128181275 | 128186091 | DNAJB8 | DnaJ (Hsp40) homolog, subfamily B, member 8 |
| 128198265 | 128212030 | GATA2 | GATA binding protein 2 |
| 128256860 | 128257474 | TMED10P2 | transmembrane emp24-like trafficking protein 10 (yeast) pseudogene 2 |
| 128338813 | 128369719 | RPN1 | ribophorin I |
| 128444975 | 128533641 | RAB7A | RAB7A, member RAS oncogene family |
| 128598333 | 128631957 | ACAD9 | acyl-CoA dehydrogenase family, member 9 |
| 128720131 | 128759585 | EFCC1 | EF-hand and coiled-coil domain containing 1 |
| 128775459 | 128781254 | GP9 | glycoprotein IX (platelet) |
| 128806412 | 128840993 | RAB43 | RAB43, member RAS oncogene family |
| 128806412 | 128880073 | ISY1-RAB43 | ISY1-RAB43 readthrough |
| 128846259 | 128880409 | ISY1 | ISY1 splicing factor homolog (S. cerevisiae) |
| 128886658 | 128902810 | CNBP | CCHC-type zinc finger, nucleic acid binding protein |
| 128968453 | 128996616 | COPG1 | gamma-COP |
| 128997680 | 129025029 | HMCES | 5-hydroxymethylcytosine (hmC) binding, ES cell-specific |
| 129033614 | 129035120 | H1FX | H1 histone family, member X |
| 129120164 | 129147494 | EFCAB12 | EF-hand calcium binding domain 12 |
| 129149787 | 129159022 | MBD4 | methyl-CpG binding domain 4 DNA glycosylase |
| 129158671 | 129239350 | IFT122 | intraflagellar transport 122 |
| 129247482 | 129254187 | RHO | rhodopsin |
| 129262057 | 129270310 | H1FOO | H1 histone family, member O, oocyte-specific |
| 129274056 | 129325582 | PLXND1 | plexin D1 |
| 129366635 | 129612419 | TMCC1 | transmembrane and coiled-coil domain family 1 |
| 129693236 | 129696781 | TRH | thyrotropin-releasing hormone |
| 129800674 | 129817233 | ALG1L2 | ALG1, chitobiosyldiphosphodolichol beta-mannosyltransferase-like 2 |
| 130064359 | 130203690 | COL6A5 | collagen, type VI, alpha 5 |
| 130236198 | 130395890 | COL6A6 | collagen, type VI, alpha 6 |
| 130397778 | 130465696 | PIK3R4 | phosphoinositide-3-kinase, regulatory subunit 4 |
| 130569327 | 130735556 | ATP2C1 | ATPase, Ca++ transporting, type 2C, member 1 |
| 130732721 | 130745698 | ASTE1 | asteroid homolog 1 (Drosophila) |

| | | | |
|---|---|---|---|
| 130745694 | 131069309 | NEK11 | hCG_1817838 |
| 131100515 | 131107674 | NUDT16 | nudix (nucleoside diphosphate linked moiety X)-type motif 16 |
| 131181045 | 131221860 | MRPL3 | mitochondrial ribosomal protein L3 |
| 131252413 | 131759152 | CPNE4 | copine IV |
| 131704699 | 131704775 | MIR5704 | microRNA 5704 |
| 131947944 | 131948015 | TRC-GCA6-1 | transfer RNA-Cys (GCA) 6-1 |
| 131950642 | 131950713 | TRC-GCA9-1 | transfer RNA-Cys (GCA) 9-1 |
| 132036211 | 132087146 | ACPP | acid phosphatase, prostate |
| 132136361 | 132257876 | DNAJC13 | DnaJ (Hsp40) homolog, subfamily C, member 13 |
| 132276982 | 132378975 | ACAD11 | acyl-CoA dehydrogenase family, member 11 |
| 132276982 | 132441303 | NPHP3-ACAD11 | NPHP3-ACAD11 readthrough (NMD candidate) |
| 132316081 | 132321488 | ACKR4 | atypical chemokine receptor 4 |
| 132373290 | 132397467 | UBA5 | ubiquitin-like modifier activating enzyme 5 |
| 132399453 | 132441303 | NPHP3 | nephronophthisis 3 (adolescent) |
| 132757171 | 133116619 | TMEM108 | transmembrane protein 108 |
| 133118790 | 133194056 | BFSP2 | beaded filament structural protein 2, phakinin |
| 133292434 | 133309118 | CDV3 | CDV3 homolog (mouse) |
| 133319449 | 133380762 | TOPBP1 | topoisomerase (DNA) II binding protein 1 |
| 133464977 | 133497850 | TF | transferrin |
| 133502877 | 133540336 | SRPRB | signal recognition particle receptor, B subunit |
| 133543079 | 133614691 | RAB6B | RAB6B, member RAS oncogene family |
| 133646989 | 133648656 | C3orf36 | chromosome 3 open reading frame 36 |
| 133651540 | 133748920 | SLCO2A1 | solute carrier organic anion transporter family, member 2A1 |
| 133875978 | 133969586 | RYK | receptor-like tyrosine kinase |
| 134074187 | 134094321 | AMOTL2 | angiomotin like 2 |
| 134156669 | 134156748 | MIR4788 | microRNA 4788 |
| 134196546 | 134204865 | ANAPC13 | anaphase promoting complex subunit 13 |
| 134204575 | 134293855 | CEP63 | centrosomal protein 63kDa |
| 134318765 | 134370522 | KY | kyphoscoliosis peptidase |
| 134514099 | 134979309 | EPHB1 | EPH receptor B1 |
| 135684515 | 135866752 | PPP2R3A | PP2A, subunit B, B72/B130 isoforms |
| 135867760 | 135915522 | MSL2 | male-specific lethal 2 homolog (Drosophila) |
| 135969167 | 136056737 | PCCB | propionyl CoA carboxylase, beta polypeptide |
| 136055077 | 136471245 | STAG1 | stromal antigen 1 |
| 136537861 | 136574734 | SLC35G2 | solute carrier family 35, member G2 |
| 136581050 | 136670446 | NCK1 | NCK adaptor protein 1 |
| 136676707 | 136729927 | IL20RB | interleukin 20 receptor beta |
| 137483134 | 137485176 | SOX14 | SRY (sex determining region Y)-box 14 |
| 137717658 | 137752494 | CLDN18 | claudin 18 |
| 137780827 | 137834451 | DZIP1L | DAZ interacting zinc finger protein 1-like |
| 137820078 | 137821835 | KRT8P36 | keratin 8 pseudogene 36 |
| 137842560 | 137851229 | A4GNT | alpha-1,4-N-acetylglucosaminyltransferase |
| 137879830 | 137893791 | DBR1 | debranching RNA lariats 1 |

| | | | |
|---|---|---|---|
| 137906115 | 138017231 | ARMC8 | armadillo repeat containing 8 |
| 137980279 | 138049018 | NME9 | NME/NM23 family member 9 |
| 138066490 | 138124377 | MRAS | muscle RAS oncogene homolog |
| 138153415 | 138197910 | ESYT3 | extended synaptotagmin-like protein 3 |
| 138213186 | 138313187 | CEP70 | centrosomal protein 70kDa |
| 138327542 | 138352218 | FAIM | Fas apoptotic inhibitory molecule |
| 138371540 | 138553780 | PIK3CB | phosphatidylinositol-4,5-bisphosphate 3-kinase, catalytic subunit beta |
| 138663066 | 138665982 | FOXL2 | forkhead box L2 |
| 138666076 | 138672830 | FOXL2NB | FOXL2 neighbor |
| 138722804 | 138725110 | PRR23A | OTTHUMP00000218629 |
| 138737873 | 138739768 | PRR23B | proline rich 23B |
| 138760944 | 138763734 | PRR23C | proline rich 23C |
| 138823027 | 138844009 | BPESC1 | blepharophimosis, epicanthus inversus and ptosis, candidate 1 (non-protein coding) |
| 138951834 | 138952364 | PISRT1 | polled intersex syndrome regulated transcript 1 (non-protein coding RNA) |

# Annexe 2 : Articles issus du travail de thèse

The Journal of Infectious Diseases

MAJOR ARTICLE

# Major Loci on Chromosomes 8q and 3q Control Interferon γ Production Triggered by Bacillus Calmette-Guerin and 6-kDa Early Secretory Antigen Target, Respectively, in Various Populations

**Fabienne Jabot-Hanin,[1,2] Aurélie Cobat,[1,2] Jacqueline Feinberg,[1,2] Ghislain Grange,[1,2] Natascha Remus,[1,2] Christine Poirier,[5] Anne Boland-Auge,[6] Céline Besse,[6] Jacinta Bustamante,[1,2] Stéphanie Boisson-Dupuis,[1,2,7] Jean-Laurent Casanova,[1,2,3,7,8] Erwin Schurr,[9,10,11] Alexandre Alcaïs,[1,2,7] Eileen G. Hoal,[12] Christophe Delacourt,[4] and Laurent Abel[1,2,7]**

[1]Laboratory of Human Genetics of Infectious Diseases, Necker Branch, INSERM U1163, [2]Paris Descartes University, Sorbonne Paris Cité, Imagine Institute, [3]Pediatric Hematology-Immunology Unit and [4]Pediatric Pneumology Unit, Necker Hospital for Sick Children, AP-HP, Paris; [5]Centre de Lutte Anti-Tuberculeuse, Centre Hospitalier Intercommunal de Créteil, and [6]Centre National de Génotypage, Institut de Génomique, CEA, Evry, France; [7]St Giles Laboratory of Human Genetics of Infectious Diseases, Rockefeller Branch, Rockefeller University, and [8]Howard Hughes Medical Institute, New York, New York; [9]McGill International TB Centre, [10]Department of Human Genetics, and [11]Department of Medicine, McGill University, Montreal, Canada; and [12]Division of Molecular Biology and Human Genetics, MRC Centre for Molecular and Cellular Biology and DST/NRF Centre of Excellence for Biomedical TB Research, Faculty of Health Sciences, Stellenbosch University, Tygerberg, South Africa

***Background.*** Interferon γ (IFN-γ) release assays (IGRAs) provide an in vitro measurement of antimycobacterial immunity that is widely used as a test for *Mycobacterium tuberculosis* infection. IGRA outcomes are highly heritable in various populations, but the nature of the involved genetic factors remains unknown.

***Methods.*** We conducted a genome-wide linkage analysis of IGRA phenotypes in families from a tuberculosis household contact study in France and a replication study in families from South Africa to confirm the loci identified.

***Results.*** We identified a major locus on chromosome 8q controlling IFN-γ production in response to stimulation with live bacillus Calmette-Guerin (BCG; LOD score, 3.81; $P = 1.40 \times 10^{-5}$). We also detected a second locus, on chromosome 3q, that controlled IFN-γ levels in response to stimulation with 6-kDa early secretory antigen target, when accounting for the IFN-γ production shared with that induced by BCG (LOD score, 3.72; $P = 1.8 \times 10^{-5}$). Both loci were replicated in South African families, where tuberculosis is hyperendemic. These loci differ from those previously identified as controlling the response to the tuberculin skin test (*TST1* and *TST2*) and the production of TNF-α (*TNF1*).

***Conclusions.*** The identification of 2 new linkage signals in populations of various ethnic origins living in different *M. tuberculosis* exposure settings provides new clues about the genetic control of human antimycobacterial immunity.

**Keywords.** tuberculosis; genetic linkage analysis; interferon gamma release assays; mycobacteria; genetic control.

Tuberculosis remains a major public health problem, with *Mycobacterium tuberculosis* currently infecting an estimated one third of the world's population and approximately 9 million new cases of and 1.5 million deaths due to tuberculosis in 2013 [1, 2]. *M. tuberculosis* bacilli are transmitted by inhalation of aerosolized droplets generated by the coughing of patients with infectious tuberculosis. There is no direct proof of latent *M. tuberculosis* infection (hereafter, "latent infection") in exposed individuals, and the infection phenotype is inferred indirectly from quantitative measurements of antimycobacterial immunity, attesting to previous exposure to *M. tuberculosis* [3]. The tuberculin skin test (TST) is the most widely used method to test for latent infection [4], although it suffers from a lack of specificity, partly due to cross-reactions with bacillus Calmette-Guerin (BCG) and, to a lesser extent, environmental mycobacteria. Additional assays to detect latent infection, based on in vitro evaluations of T-cell antimycobacterial immunity, have been developed over the last 15 years [5]. They measure the secretion of interferon γ (IFN-γ) by circulating leukocytes in response to *M. tuberculosis* antigens, such as 6-kDa early secretory antigen target (ESAT-6) and 10-kDa culture filtrate protein [6]. ESAT-6 is encoded neither by the BCG strain used for vaccination nor by most environmental mycobacteria [7]. These IFN-γ release assays (IGRAs) yield results that are not fully concordant with TST findings, but they provide complementary information about infection status [8, 9].

Based on TST and IGRA results, an estimated 10%–20% of subjects do not become infected with *M. tuberculosis* despite sustained exposure and, hence, never develop disease [3, 10]. In addition, most infected subjects develop latent infection

Received 9 October 2015; accepted 11 December 2015; published online 21 December 2015.
Correspondence: L. Abel, Human Genetics of Infectious Diseases, Institut Imagine, 24 Bd du Montparnasse, 75015 Paris, France (laurent.abel@inserm.fr).

**The Journal of Infectious Diseases® 2016;213:1173–9**

DOI: 10.1093/infdis/jiv757

without ever developing clinical tuberculosis [2, 3, 10, 11]. There is accumulating evidence that human genetic factors play an important role in the development of clinical tuberculosis [3, 10, 12], particularly with the identification of single-gene inborn errors of immunity predisposing to at least some cases of severe childhood tuberculosis [13]. Several studies focusing on TST reactivity have also provided evidence for the role of human genetic factors in different steps of the latent infection process [14–16]. In particular, a linkage study in families from South Africa mapped 2 major loci controlling TST positivity per se (*TST1* on 11p14) and the intensity of TST reactivity (*TST2* on 5p15) [17]. The *TST1* locus was recently replicated in French families of various ethnic origins [18]. The genetic factors influencing IGRA phenotypes have been less thoroughly studied. The heritability of IFN-γ secretion has been estimated to be about 43% following BCG stimulation and 58% following ESAT-6 stimulation in South Africa [19] and to be 17%–48% following stimulation with *M. tuberculosis* antigens, including ESAT-6, in Uganda, depending on the TST status of those tested [20, 21]. In this study, we conducted a genome-wide linkage analysis (GWLA) of several phenotypes of IFN-γ production in response to mycobacterial stimulation, initially in families from household tuberculosis contacts in a suburb of Paris, France, and then in families from South Africa.

## MATERIALS AND METHODS

### Subjects and Families

A prospective study of household tuberculosis contacts was conducted in Val-de-Marne, in the suburbs of Paris, as previously described [22]. Val-de-Marne is an area of low tuberculosis endemicity with an annual tuberculosis incidence of 22.1 cases per 100 000 at the time of the study, compared with an overall incidence of 8.8 cases per 100 000 in France. From April 2004 to January 2009, household contacts exposed to a patient with culture-confirmed pulmonary tuberculosis were enrolled in the context of a general screening procedure, as detailed in the Supplementary Methods. This study was approved by the French Consultative Committee for Protecting Persons in Biomedical Research of Henri Mondor Hospital (Créteil, France). Written informed consent was obtained from all study participants and from parents of the enrolled minors/children.

As a replication cohort, we used 450 people from 135 nuclear families from Ravensmead and Uitsig, a suburban of Cape Town, South Africa, where tuberculosis is hyperendemic [23]. This sample had previously been used to map the *TST1* and *TST2* loci [17] and to study the heritability of antimycobacterial immunity [19].

### Measurement of IFN-γ Production

For the Val–de-Marne sample, blood samples were collected from each individual, and peripheral blood mononuclear cells (PBMCs) were isolated and activated with ESAT-6, purified protein derivative (PPD), live BCG, and phytohemagglutinin (PHA), as described in the Supplementary Methods. For the Cape Town sample, IGRAs were performed in quadruplicate on whole-blood specimens with BCG, PPD, ESAT-6, and PHA stimulations, as previously described [8]. IFN-γ levels were measured on days 3 and 7 after stimulation, but, for the sake of consistency with the primary cohort, we confined the analysis to the measurements made on day 3.

### Phenotypes and Covariates of Interest

Three phenotypes were studied: IFN-γ production after stimulation with BCG, PPD, and ESAT-6. The distributions of IFN-γ production after the various stimulations were strongly skewed to the left and were therefore subjected to classical log transformation. After this transformation, the nonstimulated control value was subtracted from the stimulated values. These transformed phenotypes were then adjusted by linear regression for risk factors selected from those recorded during recruitment [22]. The selected covariates were chosen on the basis of a significant association with at least one of the phenotypes studied in univariate analysis and to give the best fit in terms of the Akaike information criterion (AIC) in the final multivariate model. These adjusted phenotypes are referred to here as IFNγ-BCG, IFNγ-ESAT6, and IFNγ-PPD. The distribution of IFN-γ production before and after adjustment is shown in Supplementary Figure 1. We also studied a fourth phenotype corresponding to IFNγ-ESAT6 adjusted for IFNγ-BCG, to isolate a more specific response to the ESAT-6 antigen in terms of IFN-γ production, taking into account the effect shared between BCG and ESAT-6 stimulation. This phenotype is denoted IFNγ-$ESAT6_{BCG}$.

Relevant covariates for this analysis are detailed in the Supplementary Methods and Supplementary Table 1. These covariates include the annual incidence of tuberculosis in the country of birth; the estimated exposure to the index case; the infectivity of the index case; the presence or absence of complementary health insurance coverage, for use as a marker of socioeconomic status; age; and a binary indicator of the time between TST administration and blood sampling. In the South African cohort, we performed multivariate linear regression analyses in which we used the geometric mean value of the 4 measurements of IFN-γ production for the different types of stimulation, subtracted the logarithm of the value obtained in the absence of stimulation, and adjusted the resulting values for sex, age, and previous clinical tuberculosis, as previously described [19]. The IFNγ-ESAT6 phenotype in the South African cohort was also adjusted for IFNγ-BCG phenotype.

### Genetic Analysis

For the French sample, we used the Illumina linkage V panel to genotype children and their parents for the GWLA. Single-nucleotide polymorphisms (SNPs) with a call rate of <90%

were removed from the analysis, resulting in the use of 5376 autosomal informative SNPs for GWLA. The South African sample was genotyped with the Illumina linkage IVb panel, and, after quality control, 5657 autosomal SNPs were retained for linkage analyses [17]. Model-free GWLA of the adjusted IFN-γ production phenotypes was performed with the new maximum-likelihood binomial (MLB) method for quantitative traits (nMLB-QTL v.3.0) [24, 25]. The MLB approach considers the sibship as a whole and makes no assumptions about the distribution of the phenotype. The results of the linkage test can be expressed as a classical LOD score [24, 25]. We used LOD scores of 3.6 and 2.2 as genome-wide significant and suggestive thresholds, respectively [26]. A LOD score of 0.5875 (corresponding to a *P* value of .05) was used as the replication threshold, as previously suggested [27].

To investigate the population structure of our cohort, we performed a principal component analysis (PCA) on 5350 markers of the Illumina linkage IVb panel common between our sample and the 1000 Genomes Project multiethnic reference panel (Phase I interim release, 2011), using the EIGENSTRAT method [28]. Data only for individuals involved in the 1000 Genomes Project were used for computation of the principal components, and data for the individuals from Val-de-Marne were projected one by one onto the eigenvectors with the smartpca package of EIGENSOFT [28].

## RESULTS

A flow chart describing the selection of subjects for this study is presented in Supplementary Figure 2. Analyses of the covariates influencing the phenotypes of interest in the French sample were performed on 528 household tuberculosis contacts, of whom 268 were women and 260 were men from 143 pedigrees. Univariate and multivariate analyses (Supplementary Table 1) showed that higher levels of IFN-γ production were associated with a higher incidence of tuberculosis in the country of birth, a longer duration of exposure to the index case, a higher infectivity of the index case (this effect being restricted to young individuals for the IFNγ-BCG and IFNγ-PPD phenotypes), and an absence of complementary health insurance coverage. IFN-γ production increased significantly with age for the IFNγ-ESAT6 and IFNγ-PPD phenotypes but not for the IFNγ-BCG phenotype, probably because of the high rate of BCG vaccination. The IFN-γ production phenotypes were adjusted for the relevant covariates (Supplementary Table 1) for further analyses. After adjustment, a strong correlation was found ($r = 0.78$) between IFNγ-BCG and IFNγ-PPD, and a weaker correlation ($r = 0.53$) was detected between IFNγ-BCG and IFNγ-ESAT6 (Table 1). As described in "Methods" section, IFNγ-ESAT6 was also adjusted for IFNγ-BCG and for the covariates shown in Supplementary Table 1 (IFNγ-ESAT6$_{BCG}$).

GWLA was performed on 97 informative families, each including 2–6 offspring with available phenotypes and containing 240 siblings in total. Based on the 2 first principal components of the PCA, the 97 studied families could be divided into 49 from Europe or North Africa, 36 from sub-Saharan Africa (AFR), and 12 from other origins, including Asia (Supplementary Figure 3). Information content (IC) was high across all autosomes, with a mean genome-wide information value of 87.6% (range, 69%–94%). The results for the GWLA of the IFNγ-BCG phenotype are shown in Figure 1*A*. A significant linkage signal was observed on chromosome region 8q21.13, with a LOD score of 3.80 ($P = 1.4 \times 10^{-5}$), at 82.7 Mb (IC = 86%). In addition, we also found a suggestive linkage signal on chromosome 5q35 (LOD score = 2.48, IC = 86.4%), and seven weaker linkage peaks with LOD scores of >1.17 (ie, $P$ <.01; Supplementary Table 2). The most significant GWLA peak observed for the IFNγ-PPD phenotype was also on chromosome region 8q21.13, with a LOD score of 3.03 (IC = 89.9%), at 79 Mb (Supplementary Figure 4), consistent with the strong correlation between IFNγ-PPD and IFNγ-BCG. Among the 97 families used for the analysis, 39% of families from Europe or North Africa, 36% from sub-Saharan Africa, and 41% from other origins were contributing to the significant linkage signal observed for the IFNγ-BCG phenotype (ie, LOD score >0.1 in the linkage region). This result shows that the observed linkage signal was supported by families from the different ethnic backgrounds present in our sample.

We then performed a replication study for the 2 chromosomal regions identified, 8q21 and 5q35, in the South African sample, with the adjusted IFNγ-BCG phenotype. The suggestive 5q35 signal was not significant in the South African sample (maximal LOD score = 0.14), but we were able to replicate linkage to the 8q21 region with a LOD score of 0.98 ($P = .016$) at 70 Mb (Figure 1*B*). No other significant linkage signal for IFNγ-BCG phenotype was found in the Cape Town sample (the highest LOD score reached 2.06 at 83 Mb on chromosome 16). Additional support for linkage to the IFNγ-BCG phenotype was provided by summing the LOD scores of the 2 samples with a LOD score of 4.50 at 69.7 Mb (Figure 1*B*). Given the variability of estimates of location in linkage studies of complex traits [29] and the slight differences in phenotype definition between

**Table 1. Spearman Correlation Coefficients for the Relationships Between the 4 Phenotypes of Interferon γ Production Used for Genome-Wide Linkage Analysis and Tuberculin Skin Testing (TST) in the Val-de-Marne Sample**

| Phenotype | IFNγ-BCG | IFNγ-PPD | IFNγ-ESAT6 | IFNγ-ESAT6$_{BCG}$ | TST[a] |
|---|---|---|---|---|---|
| IFNγ-BCG | 1 | 0.78 | 0.53 | 0.02 | 0.09 |
| IFNγ-PPD | 0.78 | 1 | 0.61 | 0.25 | 0.2 |
| IFNγ-ESAT6 | 0.53 | 0.61 | 1 | 0.86 | 0.07 |
| IFNγ-ESAT6$_{BCG}$ | 0.02 | 0.25 | 0.86 | 1 | 0.02 |
| TST | 0.09 | 0.2 | 0.07 | 0.02 | 1 |

Phenotypes are described in "Materials and Methods" section.

Abbreviations: BCG, bacillus Calmette-Guerin; ESAT, early secretory antigen target; IFNγ, interferon γ; PPD, purified protein derivative.

[a] Adjusted as described by Cobat et al [18].

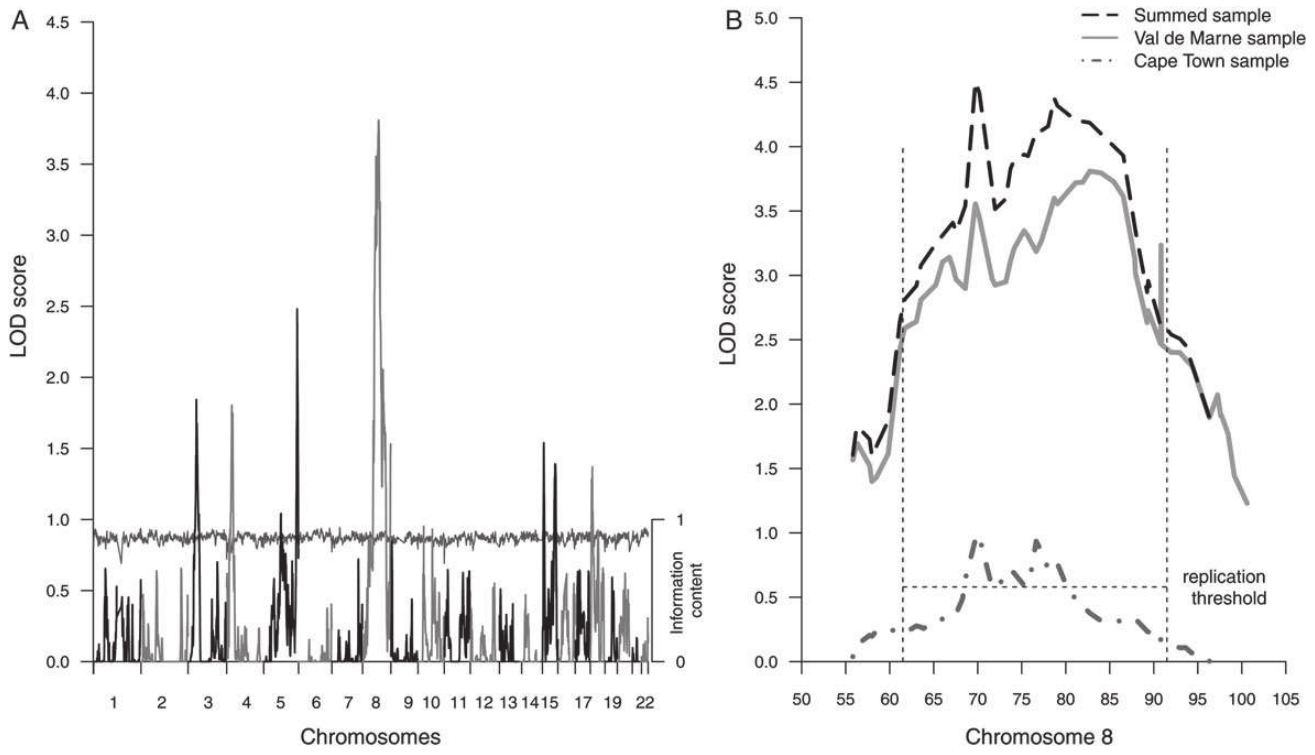



**Figure 1.** Model-free linkage analysis of the IFNγ-BCG phenotype described in "Materials and Methods" section. *A*, Results of genome-wide analysis for the Val de Marne sample, showing multipoint LOD scores (left *y*-axis) and information content (horizontal line; right *y*-axis) for the 22 autosomes (*x*-axis). *B*, Expanded view of the linked chromosome 8 region from 55 to 100 Mb (*x*-axis); the multipoint LOD score is shown on the *y*-axis for the Val-de-Marne sample (solid gray line), the Cape Town sample (double-dashed gray line), and the summed samples (black long-dash line). The horizontal dotted line indicates the significance threshold for replication, and the 2 vertical dotted lines delimit the confidence interval of the linked locus.

our samples, it seems reasonable to consider a rather large confidence interval for the locus influencing the IFNγ-BCG phenotype. Based on the curve of the summed LOD scores, we considered that this interval was between 61 and 91.5 Mb on chromosome 8 (Figure 1*B*). This region contains 108 known genes (Supplementary Table 3), including the gene (*IL7*) that encodes interleukin 7, which is required for the development and homeostasis of human T lymphocytes [30, 31]. Another interesting gene in this region is *LY96,* which encodes a protein that cooperates with Toll-like receptor 2 (TLR2) in the response to cell wall components from gram-positive and gram-negative bacteria [32]. Finally, this region borders the 8q12-13 region (55.1–61.2 Mb) that was previously reported to be linked to pulmonary tuberculosis in Morocco [33] and including the *TOX* gene (at 59.7–60 Mb), variants of which are associated with early onset pulmonary tuberculosis [34]. Overall, the present linkage analysis results highlight a major locus in chromosomal region 8q12-8q22 controlling the amount of IFN-γ produced in response to BCG.

The GWLA of the IFNγ-ESAT6 phenotype identified no significant linkage signal, with a maximum LOD score of 2.19 ($P = 7.4 \times 10^{-4}$) at 122 Mb on chromosome 3 (Supplementary Figure 5). However, when IFNγ-ESAT6 was adjusted for the IFNγ-BCG phenotype, the linkage signal on chromosome 3q13-22 became significant, with a LOD score of 3.72 ($P = 1.8 \times 10^{-5}$) at 122.3 Mb (IC = 90.9%; Figure 2). No other suggestive linkage peaks were found with the IFNγ-ESAT6$_{BCG}$ phenotype, and there were 4 weaker linkage signals with *P* values of <.01 (Supplementary Table 3). Among the 97 families used for the analysis, 39% from Europe or North Africa, 25% from sub-Saharan Africa, and 41% from other origins were contributing to the linkage signal observed for the IFNγ-ESAT6$_{BCG}$ phenotype, indicating that this linkage signal was also resulting from families of different ethnic origins. The chromosome 3q signal obtained with the IFNγ-ESAT6$_{BCG}$ phenotype was replicated in the South African sample, with LOD scores of 0.78 ($P = .028$) at 125.7 Mb and 1.31 ($P = .007$) at 138.7 Mb. No other significant linkage signal for IFNγ-ESAT6$_{BCG}$ phenotype was found in the Cape Town sample (the highest LOD score reached 1.8 at 4.5 Mb on chromosome 19). When the LOD scores for the 2 samples were summed, the maximum LOD score was 4.16 at 123.5 Mb. For the reasons given above, we considered that the linked region extended from 115 to 139 Mb (Figure 2).

The 180 known genes in this region (Supplementary Table 4) include *GATA2,* which encodes a transcription factor involved in the homeostasis of hematopoietic stem cells, haploinsufficiency

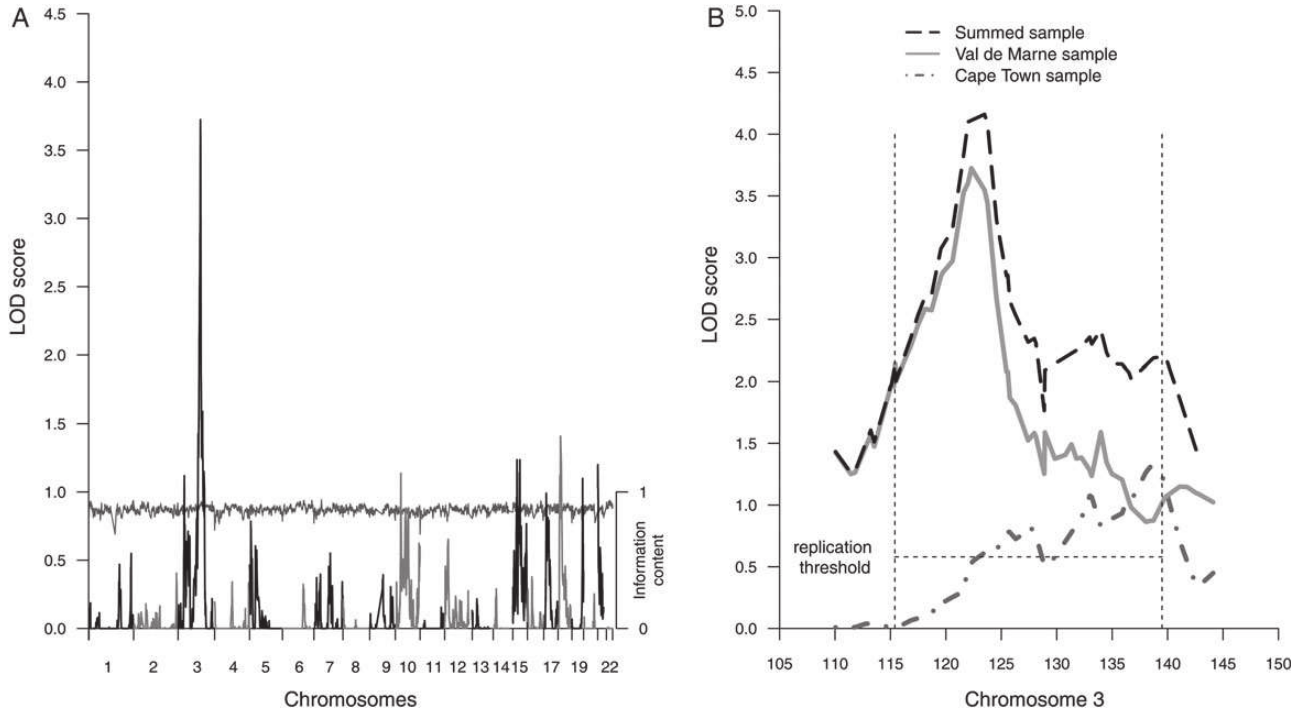



**Figure 2.** Model-free linkage analysis of the IFNγ-ESAT6$_{BCG}$ phenotype described in Materials and Methods. *A*, Results for genome-wide analysis for the Val-de-Marne sample, showing multipoint LOD scores (left *y*-axis) and information content (horizontal line; right *y*-axis) for the 22 autosomes (*x*-axis). *B*, Expanded view of the linked chromosome 3 region from 105 to 150 Mb (*x*-axis). The multipoint LOD score is indicated on the *y*-axis for the Val-de-Marne sample (solid gray line), the Cape Town sample (double-dashed gray line), and the summed samples (black long-dash line). The horizontal dotted line indicates the significance threshold for replication, and the 2 vertical dotted lines delimit the confidence interval of the linked locus.

of which is associated with mycobacterial infections, including tuberculosis [13]. *ITGB5* encodes the β chain of the integrin heterodimer $\alpha_v\beta_5$, which is involved in cell-cell adhesion and has been reported to be essential for the activation of dendritic cells by *M. tuberculosis*–exposed neutrophils [35]. *CD80* and *CD86* encode ligands expressed on antigen-presenting cells that contribute to the regulation of T-cell activation. Mice deficient in both B7.1 (CD80) and B7.2 (CD86) were found to have enhanced susceptibility to aerosol-mediated infection with *M. tuberculosis* [36], and these 2 molecules have been reported to be equally able to mediate host resistance to *M. tuberculosis* [37]. Furthermore, *CD80* is one of the genes displaying the highest degree of differential expression in primary human dendritic cells after *M. tuberculosis* infection [38].

This linkage result identifies a second major locus in chromosomal region 3q13-22 that controls the amount of IFN-γ production in response to ESAT-6, one of the specific antigenic proteins produced by *M. tuberculosis*, when taking into account the IFN-γ production, which is shared between the BCG and the ESAT-6 stimulation. We also performed a linkage analysis of the IFNγ-ESAT6 phenotype adjusted for IFNγ-PPD. The correlation of IFNγ-ESAT6 with IFNγ-PPD ($r = 0.61$) was stronger than that with IFNγ-BCG; PPD contains the ESAT-6 antigen, unlike BCG. Interestingly, the LOD score obtained for chromosome 3q fell to 2.37 after the adjustment of IFNγ-ESAT6 for IFNγ-PPD (data not shown). This suggests that the chromosome 3 locus is involved in controlling the IFN-γ production more specifically triggered by ESAT-6 stimulation.

## DISCUSSION

We identified 2 significant genome-wide linkage signals corresponding to 2 antimycobacterial immunity phenotypes. The first, on chromosome region 8q12-22, is expected to harbor 1 or several loci that influence IFN-γ production triggered by BCG. The second peak, on chromosome region 3q13-22, indicates the location of gene(s) influencing the amount of IFN-γ released after ESAT-6 stimulation, following adjustment for the effect common to stimulation with this antigen and with BCG. We were able to replicate the mapping of these loci in a sample from South Africa, using phenotypes that were similar although not identical (IFN-γ production in whole blood samples vs PBMCs, measured at 3 days vs 4 days). The 2 populations were also remarkably different in terms of exposure to *M. tuberculosis*. The individuals from South Africa studied live in an area of hyperendemic tuberculosis in which *M. tuberculosis* transmission occurs preferentially in the community [39]. By contrast, tuberculosis endemicity is low in France, and the design of the French study targeted household

tuberculosis contacts. In addition, the 2 cohorts also differed in terms of genetic background. The families in the Val-de-Marne sample belonged to several ethnic groups that we classified into 3 main subpopulations (individuals with a European or North African origin, those with a sub-Saharan African origin, and those with another origin, including Asia), and we found that, within each group, a similar proportion of families contributed to the 2 linkage peaks. By contrast, all individuals from the replication sample studied were from the South African Coloured ethnic group, a population resulting from an admixture of Khoesans (31%), Bantu-speaking Africans (33%), Europeans (16%), and Asians (20%) [40]. The French cohort displayed genetic diversity at the population level, whereas the South African cohort displayed genetic diversity at the individual level. Thus, the replication of linkage findings for these 2 loci in such different settings suggests a robust and, perhaps, universal role of these loci in the control of mycobacteria-triggered IFN-γ production in humans.

The IFNγ-BCG phenotype corresponds to a general antimycobacterial response. Indeed, IFN-γ production in response to BCG stimulation was highly correlated with the response to PPD antigens ($r = 0.78$), and the IFNγ-PPD phenotype was also linked to the 8q locus, with a LOD score of 3.03. This suggests that the 8q locus may control a nonspecific component of IFN-γ release during mycobacterial infection. The second locus on chromosome 3 corresponds to the IFNγ-ESAT6 phenotype analysis when taking into account the IFN-γ production which is common between the BCG and the ESAT6 stimulation ($r = 0.53$). This common element may reflect a general capacity for IFN-γ production via the T-cell receptor signaling pathway, whereas the IFNγ-ESAT6$_{BCG}$ phenotype is expected to be more specific to ESAT-6, as this antigen is absent from the BCG strain. ESAT-6 plays an important and specific role in *M. tuberculosis* infection. The adoptive transfer of $CD4^+$ T cells expressing an ESAT-6–specific T-cell receptor in mice has been reported to lead to strongly enhanced resistance to subsequent airborne *M. tuberculosis* infection [41], indicating the existence of a specific immune response against *M. tuberculosis* infection mediated by the response to ESAT-6. The use of the adjusted IFNγ-ESAT6$_{BCG}$ phenotype in our analysis strongly increased the linkage peak on chromosome 3q, leading to the mapping of a major locus. We also found that this linkage peak was substantially decreased by adjustment for IFN-γ production after stimulation by PPD that contains ESAT-6. Overall, these results are consistent with the view that the chromosome 3q locus plays a role in controlling the IFN-γ production more specifically induced by ESAT-6 stimulation.

Not surprisingly, these loci do not overlap with the *TST1* and the *TST2* loci controlling TST positivity per se and the intensity of TST reactivity, respectively [17]. This is consistent with the observed weak correlation between TST and in vitro measurements of *M. tuberculosis* infection (Table 1) and supports the hypothesis that TST and IFN-γ production by PBMCs are markers of different and complementary aspects of antimycobacterial immunity [8]. In particular, the *TST1* locus is thought to reflect T cell–independent resistance to *M. tuberculosis* infection. It is also likely that the functions of skin-homing cells are more diverse than the production of IFN-γ alone. For instance, the production of the proinflammatory cytokine tumor necrosis factor α (TNFα) is thought to play a major role in the initiation of the TST reaction, a hypothesis supported by the overlap of linkage regions between *TST1* and the *TNF1* locus controlling mycobacterium-driven tumor necrosis factor α production [18, 42]. In this context, the identification of 2 new loci controlling BCG- and ESAT-6–triggered IFN-γ production adds 2 new pieces to the puzzle of how human antimycobacterial immunity is assembled.

## Supplementary Data

Supplementary materials are available at http://jid.oxfordjournals.org. Consisting of data provided by the author to benefit the reader, the posted materials are not copyedited and are the sole responsibility of the author, so questions or comments should be addressed to the author.

## Notes

***Acknowledgments.*** We thank all members of the community who participated in this study; the Centre National de Génotypage, for conducting the genotyping; Aziz Belkadi, for bioinformatic support; and Emmanuelle Jouanguy, Anne Puel, and Capucine Picard, for helpful discussions.

***Financial support.*** This work was supported by the Programme Hospitalier de Recherche Clinique (AOR-04-003); the Legs Poix (Chancellerie des Universités de Paris); the French National Research Agency, under the "Investments for the future" program (grant ANR-10-IAHU-01); the European Research Council (ERC-2010-AdG-268777); the Rockefeller University; the Institut National de la Santé et de la Recherche Médicale; Paris Descartes University; the St. Giles Foundation; the Canadian Institutes of Health Research; the Sequella/Aeras Global Tuberculosis Foundation; and the Government of Canada (Banting postdoctoral fellowship 112932 to A. C.).

***Potential conflicts of interest.*** All authors: No reported conflicts. All authors have submitted the ICMJE Form for Disclosure of Potential Conflicts of Interest. Conflicts that the editors consider relevant to the content of the manuscript have been disclosed.

## References

1. WHO. A world free of tuberculosis (TB). http://www.who.int/tb/en/. Accessed 17 December 2014.
2. Zumla A, Raviglione M, Hafner R, von Reyn CF. Tuberculosis. N Engl J Med **2013**; 368:745–55.
3. O'Garra A, Redford PS, McNab FW, Bloom CI, Wilkinson RJ, Berry MPR. The immune response in tuberculosis. Annu Rev Immunol **2013**; 31:475–527.
4. Reichman LB. Tuberculin skin testing. The state of the art. Chest **1979**; 76(6 Suppl):764–70.
5. Pai M, Riley LW, Colford JM Jr. Interferon-γ assays in the immunodiagnosis of tuberculosis: a systematic review. Lancet Infect Dis **2004**; 4:761–76.
6. Mahairas GG, Sabo PJ, Hickey MJ, Singh DC, Stover CK. Molecular analysis of genetic differences between *Mycobacterium bovis* BCG and virulent M. bovis. J Bacteriol **1996**; 178:1274–82.
7. Andersen P, Munk ME, Pollock JM, Doherty TM. Specific immune-based diagnosis of tuberculosis. Lancet **2000**; 356:1099–104.
8. Gallant CJ, Cobat A, Hoal EG, Schurr E. Tuberculin Skin test and in vitro assays provide complementary measures of antimycobacterial immunity in children and adolescents. Chest **2010**; 137:1071–7.
9. Salgame P, Geadas C, Collins L, Jones-López E, Ellner JJ. Latent tuberculosis infection–Revisiting and revising concepts. Tuberc Edinb Scotl **2015**; 95:373–84.

10. Abel L, El-Baghdadi J, Bousfiha AA, Casanova J-L, Schurr E. Human genetics of tuberculosis: a long and winding road. Philos Trans R Soc B Biol Sci **2014**; 369:20130428.
11. Alcaïs A, Fieschi C, Abel L, Casanova J-L. Tuberculosis in children and adults: two distinct genetic diseases. J Exp Med **2005**; 202:1617–21.
12. Casanova J-L, Abel L. Genetic dissection of immunity to mycobacteria: the human model. Annu Rev Immunol **2002**; 20:581–620.
13. Boisson-Dupuis S, Bustamante J, El-Baghdadi J, et al. Inherited and acquired immunodeficiencies underlying tuberculosis in childhood. Immunol Rev **2015**; 264:103–20.
14. Jepson A, Fowler A, Banya W, et al. Genetic regulation of acquired immune responses to antigens of *Mycobacterium tuberculosis*: a study of twins in West Africa. Infect Immun **2001**; 69:3989–94.
15. Sepulveda RL, Heiba IM, King A, Gonzalez B, Elston RC, Sorensen RU. Evaluation of tuberculin reactivity in BCG-immunized siblings. Am J Respir Crit Care Med **1994**; 149(3 Pt 1):620–4.
16. Stein CM, Zalwango S, Malone LL, et al. Genome scan of *M. tuberculosis* infection and disease in Ugandans. PLoS One **2008**; 3:e4094.
17. Cobat A, Gallant CJ, Simkin L, et al. Two loci control tuberculin skin test reactivity in an area hyperendemic for tuberculosis. J Exp Med **2009**; 206:2583–91.
18. Cobat A, Poirier C, Hoal E, et al. Tuberculin skin test negativity is under tight genetic control of chromosomal region 11p14-15 in settings with different tuberculosis endemicities. J Infect Dis **2015**; 211:317–21.
19. Cobat A, Gallant CJ, Simkin L, et al. High heritability of antimycobacterial immunity in an area of hyperendemicity for tuberculosis disease. J Infect Dis **2010**; 201:15–9.
20. Tao L, Zalwango S, Chervenak K, et al. Genetic and shared environmental influences on interferon-γ production in response to *Mycobacterium tuberculosis* antigens in a Ugandan population. Am J Trop Med Hyg **2013**; 89:169–73.
21. Stein CM, Guwatudde D, Nakakeeto M, et al. Heritability analysis of cytokines as intermediate phenotypes of tuberculosis. J Infect Dis **2003**; 187:1679–85.
22. Aissa K, Madhi F, Ronsin N, et al. Evaluation of a model for efficient screening of tuberculosis contact subjects. Am J Respir Crit Care Med **2008**; 177:1041–7.
23. den Boon S, van Lill SWP, Borgdorff MW, et al. High prevalence of tuberculosis in previously treated patients, Cape Town, South Africa. Emerg Infect Dis **2007**; 13:1189–94.
24. Alcais A, Abel L. Maximum-likelihood-binomial method for genetic model-free linkage analysis of quantitative traits in sibships. Genet Epidemiol **1999**; 17:102–17.
25. 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**; 35:46–56.
26. Lander E, Kruglyak L. Genetic dissection of complex traits: guidelines for interpreting and reporting linkage results. Nat Genet **1995**; 11:241–7.
27. Nyholt DR. All LODs are not created equal. Am J Hum Genet **2000**; 67:282–8.
28. Price AL, Patterson NJ, Plenge RM, Weinblatt ME, Shadick NA, Reich D. Principal components analysis corrects for stratification in genome-wide association studies. Nat Genet **2006**; 38:904–9.
29. Roberts SB, MacLean CJ, Neale MC, Eaves LJ, Kendler KS. Replication of linkage studies of complex traits: an examination of variation in location estimates. Am J Hum Genet **1999**; 65:876–84.
30. Puel A, Ziegler SF, Buckley RH, Leonard WJ. Defective IL7R expression in T(-)B(+)NK(+) severe combined immunodeficiency. Nat Genet **1998**; 20:394–7.
31. Jacobs SR, Michalek RD, Rathmell JC. IL-7 is essential for homeostatic control of T cell metabolism in vivo. J Immunol **2010**; 184:3461–9.
32. Dziarski R, Wang Q, Miyake K, Kirschning CJ, Gupta D. MD-2 enables Toll-like receptor 2 (TLR2)-mediated responses to lipopolysaccharide and enhances TLR2-mediated responses to Gram-positive and Gram-negative bacteria and their cell wall components. J Immunol **2001**; 166:1938–44.
33. Baghdadi JE, Orlova M, Alter A, et al. An autosomal dominant major gene confers predisposition to pulmonary tuberculosis in adults. J Exp Med **2006**; 203:1679–84.
34. Grant AV, El-Baghdadi J, Sabri A, et al. Age-dependent association between pulmonary tuberculosis and common TOX variants in the 8q12–13 linkage region. Am J Hum Genet **2013**; 92:407–14.
35. Hedlund S, Persson A, Vujic A, Che KF, Stendahl O, Larsson M. Dendritic cell activation by sensing *Mycobacterium tuberculosis*-induced apoptotic neutrophils via DC-SIGN. Hum Immunol **2010**; 71:535–40.
36. Bhatt K, Uzelac A, Mathur S, McBride A, Potian J, Salgame P. B7 Costimulation Is Critical for Host Control of Chronic *Mycobacterium tuberculosis* Infection. J Immunol **2009**; 182:3793–800.
37. Bhatt K, Kim A, Kim A, Mathur S, Salgame P. Equivalent functions for B7.1 and B7.2 costimulation in mediating host resistance to *Mycobacterium tuberculosis*. Cell Immunol **2013**; 285:69–75.
38. Barreiro LB, Tailleux L, Pai AA, Gicquel B, Marioni JC, Gilad Y. Deciphering the genetic architecture of variation in the immune response to *Mycobacterium tuberculosis* infection. Proc Natl Acad Sci U S A **2012**; 109:1204–9.
39. Verver S, Warren RM, Munch Z, et al. Proportion of tuberculosis transmission that takes place in households in a high-incidence area. Lancet **2004**; 363:212–4.
40. Chimusa ER, Daya M, Möller M, et al. Determining ancestry proportions in complex admixture scenarios in South Africa using a novel proxy ancestry selection method. PLoS One **2013**; 8:e73971.
41. Gallegos AM, Pamer EG, Glickman MS. Delayed protection by ESAT-6-specific effector CD4+ T cells after airborne *M. tuberculosis* infection. J Exp Med **2008**; 205:2359–68.
42. Cobat A, Hoal EG, Gallant CJ, et al. Identification of a major locus, TNF1, that controls BCG-triggered tumor necrosis factor production by leukocytes in an area hyperendemic for tuberculosis. Clin Infect Dis **2013**; 57:963–70.

## Supplemental Methods

Subjects and families

From April 2004 to January 2009, household contacts exposed to a patient with culture-confirmed pulmonary tuberculosis (TB) were enrolled in the context of a general screening procedure in Val de Marne, suburb of Paris. In France, all new cases of TB are reported to the health authorities (Centre de Lutte contre la Tuberculose) to organize epidemiologic investigations and to identify contact subjects with the consent of the index case. A household contact was defined as any person sharing the residence of a TB index case during the three months preceding diagnosis of the case. A questionnaire was completed for each individual to assess the risk factors for infection and the familial relationships as detailed in [1,2]. This study was approved by the French Consultative Committee for Protecting Persons in Biomedical Research (CCPPRB) of Henri Mondor Hospital (Créteil, France). Written informed consent was obtained from all study participants, and from parents of the enrolled minors/children.

IFN-γ release assays details

For the Val-de-Marne sample, blood was drawn from each individual and peripheral blood mononuclear cells (PBMCs) were separated through Ficoll-Paque Plus (Amersham) gradient centrifugation. PBMCs were washed once in RPMI 1640 and counted with an automated cell counter (Beckmann). A total of $2.10^6$ PBMC/mL were dispensed in 200 µL RPMI1640 supplemented with 10 % heat inactivated fetal calf serum on 96-well flat-bottom plates (Nunc). They were activated with one of the following stimulations: ESAT-6 (rdESAT6, Statens-Serum-Institute, Denmark; 2.5µg/mL) antigen, PPD (Statens-Serum-Institute, Denmark; 5µg/mL), live BCG (BCG-Pasteur; Multiplicity of infection of 20 BCG/leukocyte), Phytohemagglutinin (PHA) (6.25 µg/mL) for positive control or medium alone for negative control. They were then cultured at 37°C with 5 % CO2. Supernatants were harvested after four days of stimulation, frozen and later assayed for IFN-γ by ELISA according to the manufacturer's recommendations (Pelikin Compact; CLB). Optical density was determined using an automated MR5000 ELISA reader (Thermolab Systems), and the final results were standardized per million PBMC's, in the unit pg/mL/$10^6$ PBMC's.

Covariates of interest and adjustment

After log-transformation the levels of IFN-γ production after stimulation by BCG, PPD and ESAT-6, denoted as IFNγ-BCG, IFNγ-ESAT6 and IFNγ-PPD, respectively, were adjusted

with a linear regression for risk factors selected from those recorded in the French sample during recruitment. The selected covariates were chosen to be significantly associated with at least one of the three studied phenotypes in univariate analysis and to give the best fit in terms of the Akaike's information criterion (AIC) in the multivariate final model.

Relevant covariates for these analyses included the annual incidence of TB in the country of birth (in two categories, more or less than 100 new TB cases per 100,000 per year), the estimated total exposure to the index case quantified as the log(number of contact hours with index case during the three months preceding the TB diagnosis), the infectivity of the index case defined as the presence of cavitation on chest radiography and the presence of bacilli in sputum smears, the complementary health insurance cover (yes/no), and age. Regarding IFNγ-BCG and IFNγ-PPD phenotypes, the effect of the index case infectivity was dependent on age, and an interaction term between infectivity and age was added in the regression model. As the majority of individuals were BCG vaccinated (89%), the BCG vaccine was not associated with any of the studied phenotype. We also investigated the effect of the timeframe between TST and blood sampling , which may introduce a potential boosting effect on IFN-γ production [3,4]. As already shown [3,5], we confirmed that our mean phenotypes were not significantly different between individuals sampled within the three days following the TST (35% of participants), and individuals sampled before TST (32% of participants). We also found that the mean phenotypes of individuals sampled later than three days after TST (33% of participants) did not differ significantly according to their time of sampling. Therefore, we used a binary indicator of the time between TST administration and blood sampling, and found a higher IFN-γ production in subjects sampled three days after TST for IFNγ-BCG phenotype, and to a lesser extent for IFNγ-PPD and IFNγ-ESAT6 (Table S1). These three phenotypes were subsequently adjusted for sampling time and the other covariates for further analyses. The sampling time was not associated with the IFNγ-ESAT6$_{BCG}$ phenotype as expected by the adjustment for IFNγ-BCG.

References


1. Aissa K, Madhi F, Ronsin N, et al. Evaluation of a Model for Efficient Screening of Tuberculosis Contact Subjects. Am. J. Respir. Crit. Care Med. **2008**; 177:1041–1047.

2. Cobat A, Poirier C, Hoal E, et al. Tuberculin skin test negativity is under tight genetic control of chromosomal region 11p14-15 in settings with different tuberculosis endemicities. J. Infect. Dis. **2015**; 211:317–321.

3. van Zyl-Smit RN, Pai M, Peprah K, et al. Within-subject variability and boosting of T-cell interferon-gamma responses after tuberculin skin testing. Am. J. Respir. Crit. Care Med. **2009**; 180:49–58.

4. van Zyl-Smit RN, Zwerling A, Dheda K, Pai M. Within-subject variability of interferon-g assay results for tuberculosis and boosting effect of tuberculin skin testing: a systematic review. PloS One **2009**; 4:e8517.

5. Detjen AK, Loebenberg L, Grewal HMS, et al. Short-term reproducibility of a commercial interferon gamma release assay. Clin. Vaccine Immunol. CVI **2009**; 16:1170–1175.

**Supplemental Tables**

Table S1- Covariates used for phenotypes adjustment and their significance level in multivariate analyses. The selected covariates were chosen to be significantly associated with at least one of the studied phenotypes in univariate analysis and to give the best fit in terms of Akaike's information criterion (AIC) in the multivariate final model.

| | IFNγ-BCG | IFNγ-PPD | IFNγ-ESAT6 | IFNγ-ESAT6$_{BCG}$ |
|---|---|---|---|---|
| TB incidence in the country of origin | 0.001 | 0.001 | 0.003 | - |
| Log(contact hours number) | 0.09 | 0.05 | 0.0014 | 0.008 |
| Index case infectivity | 0.015 [1] | 0.09 [1] | 0.089 | 0.0016 |
| Complementary insurance | 0.098 | 0.02 | 0.04 | 0.12 |
| Age | 0.23 | 0.004 | 0.014 | 0.0001 |
| Sampling > 3 days after TST | 0.0005 | 0.035 | 0.06 | - |
| BCG | - | - | - | $<2.\ 10^{-16}$ |

[1] the p-values include interaction with AGE

Table S2 – Chromosomal regions significant at the 0.01 level for the IFN-γ BCG phenotype in the Val-de-Marne sample

| Chromosomal region | Range (Mb) | Maximum LOD score | p-value |
|---|---|---|---|
| **3p21** | 40-51 | 1.84 | $1.8\ 10^{-3}$ |
| **4p15.2** | 24-33 | 1.8 | $2.0\ 10^{-3}$ |
| **5q35** | 169-178 | 2.48 | $3.6\ 10^{-4}$ |
| **8q11.2-8q22** | 56-101 | 3.81 | $1.4\ 10^{-5}$ |
| **8q22-8q24** | 101-121 | 2.05 | $1.0\ 10^{-3}$ |
| **8q24.3** | 144-146 | 1.53 | $4.0\ 10^{-3}$ |
| **15q11.2-q12** | 26-30 | 1.54 | $3.8\ 10^{-3}$ |
| **15q25** | 82-90 | 1.39 | $5.7\ 10^{-3}$ |
| **18p11.2** | 8-11 | 1.37 | $6.0\ 10^{-3}$ |

Table S3 - Chromosomal regions significant at the 0.01 level for the IFN-γ $ESAT6_{BCG}$ phenotype in the Val-de-Marne sample

| Chromosomal region | Range (Mb) | Maximum LOD score | p-value |
|---|---|---|---|
| **3q13-3q22** | 109-135 | 3.72 | $1.8\ 10^{-5}$ |
| **15q15** | 42-43 | 1.23 | $9.7\ 10^{-3}$ |
| **15q22** | 59-60 | 1.23 | $9.7\ 10^{-3}$ |
| **18p11** | 13-18 | 1.4 | $6.0\ 10^{-3}$ |
| **21q21.1** | 15-16 | 1.19 | $1.0\ 10^{-2}$ |

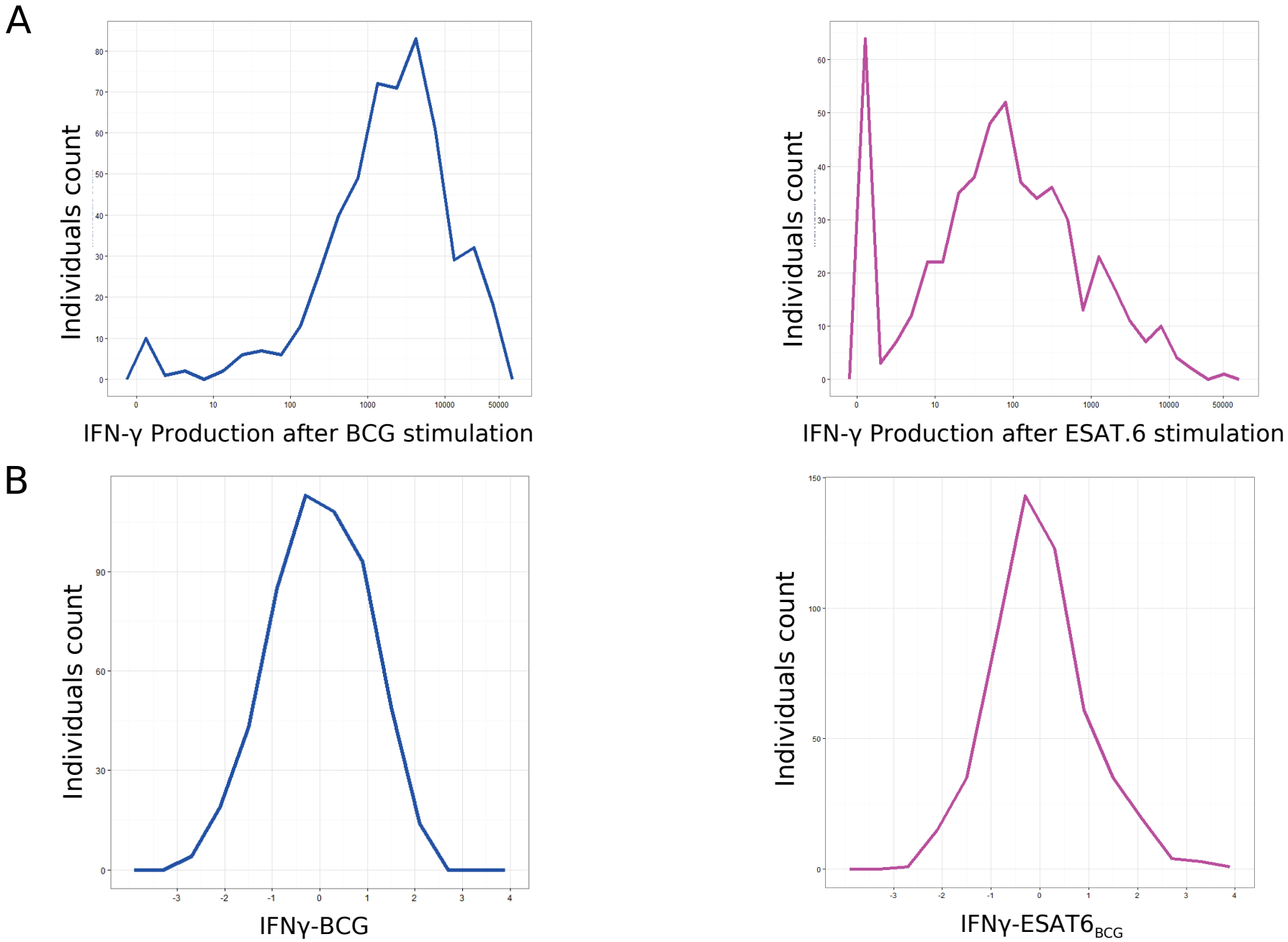



Figure S1 : IFNγ-BCG (blue) and IFNγ-ESAT6 (magenta) phenotypes distribution before and after adjustment
(A) Raw IFNγ production values of the Val de Marne sample in pg/ml on a log10 scaled x-axis (B) Standardized IFNγ phenotypes of the Val de Marne sample after adjustment for relevant covariates leading to the IFNγ-BCG (blue) and IFNγ-ESAT6$_{BCG}$(magenta) phenotypes

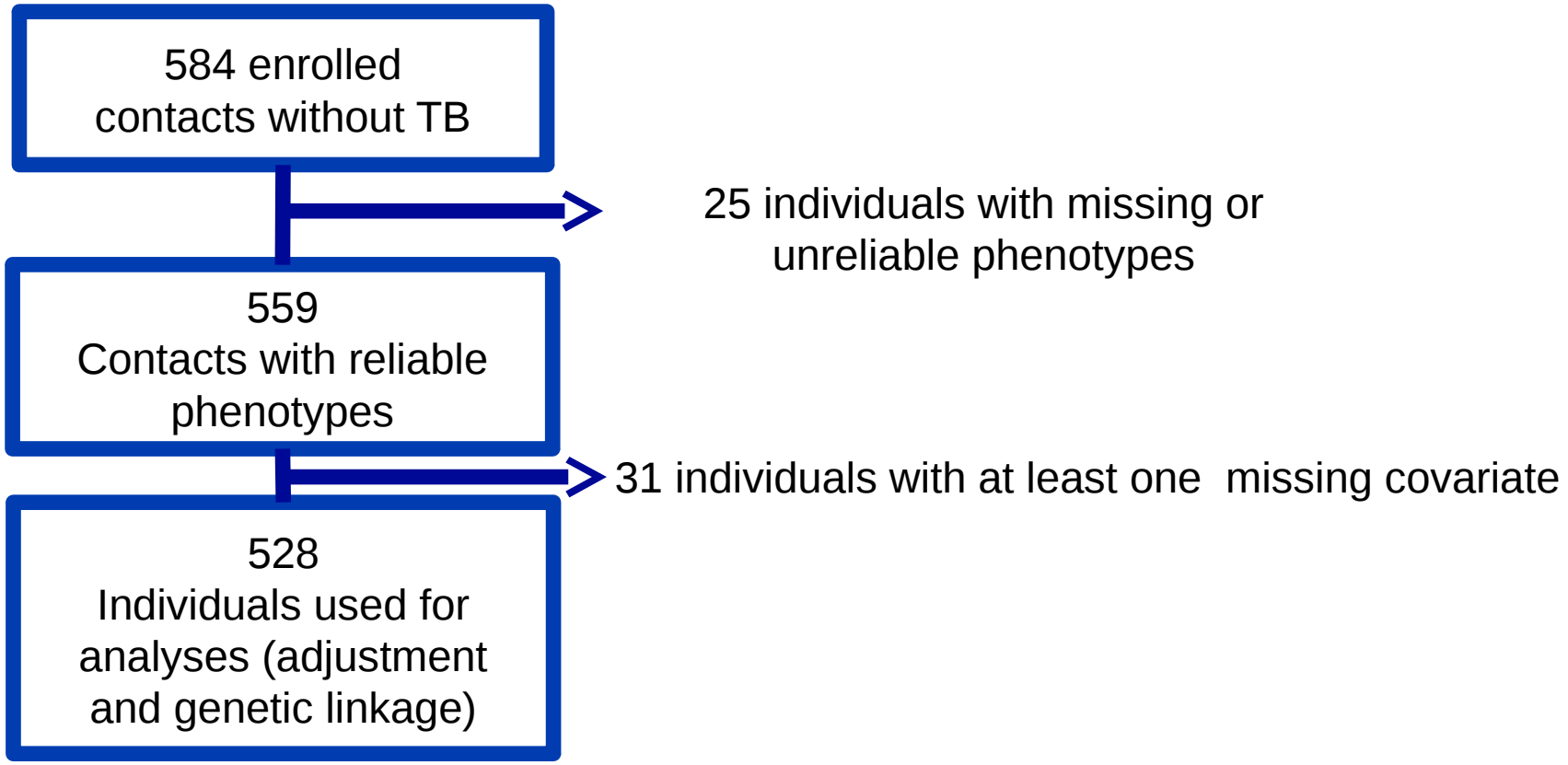



Figure S2 : Val de Marne sample selection flowchart

# Figure S3 : Population Structure of the  Val de Marne sample

We performed a principal component analysis based on 5350 markers of the Illumina linkage IVb panel common between our sample and the 1000 Genomes Project multi-ethnic reference panel. We first computed the main eigenvectors on 1092 individuals from the 1000 Genomes Project using the EIGENSOFT software [28]. The blue dots represent the Europeans (89 GBR, 93 FIN, 85CEU, 14 IBS, 98 TSI), the grey dots represent the sub-saharan Africans (61 ASW, 97 LWK, 88YRI), the pink dots represent the central and southern Americans (66 MXL, 60 CLM, 55PUR) and the orange dots represent the Asians (97 CHB, 100 CHS and 89 JPT). All these individuals are plotted on the figure along the two first principal components (PC1 along the x axis, and PC2 along the y axis).
In a second step, the Val-de-Marne individuals were projected one by one onto these principal components. Each of the 97 Val-de Marne nuclear families used for the Genome Wide linkage analysis is represented by its oldest offspring, who has been plotted on the graph and used to classify the families according to 3 arbitrary populations: 49 families are part of the Europe and North-Africa group (in the upper right rectangle), 36 are part of the Sub-saharan Africa group (in the bottom left rectangle), and 12 are considered as others (outside rectangles).
In panel (A) black triangles are representing the Val-de-Marne families contributing to the linkage signal for the IFNγ-BCG phenotype (i.e. Maximal LOD score > 0.1 between 61 and 91.5 Mb on chromosome 8) whereas the black dots are representing the other families.
In panel (B) black triangles are representing the families contributing to the linkage signal for the IFNγ-ESAT6$_{BCG}$ phenotype (i.e. Maximal LOD score > 0.1 between 115 and 140 Mb on chromosome 3) whereas the black dots are representing the other families

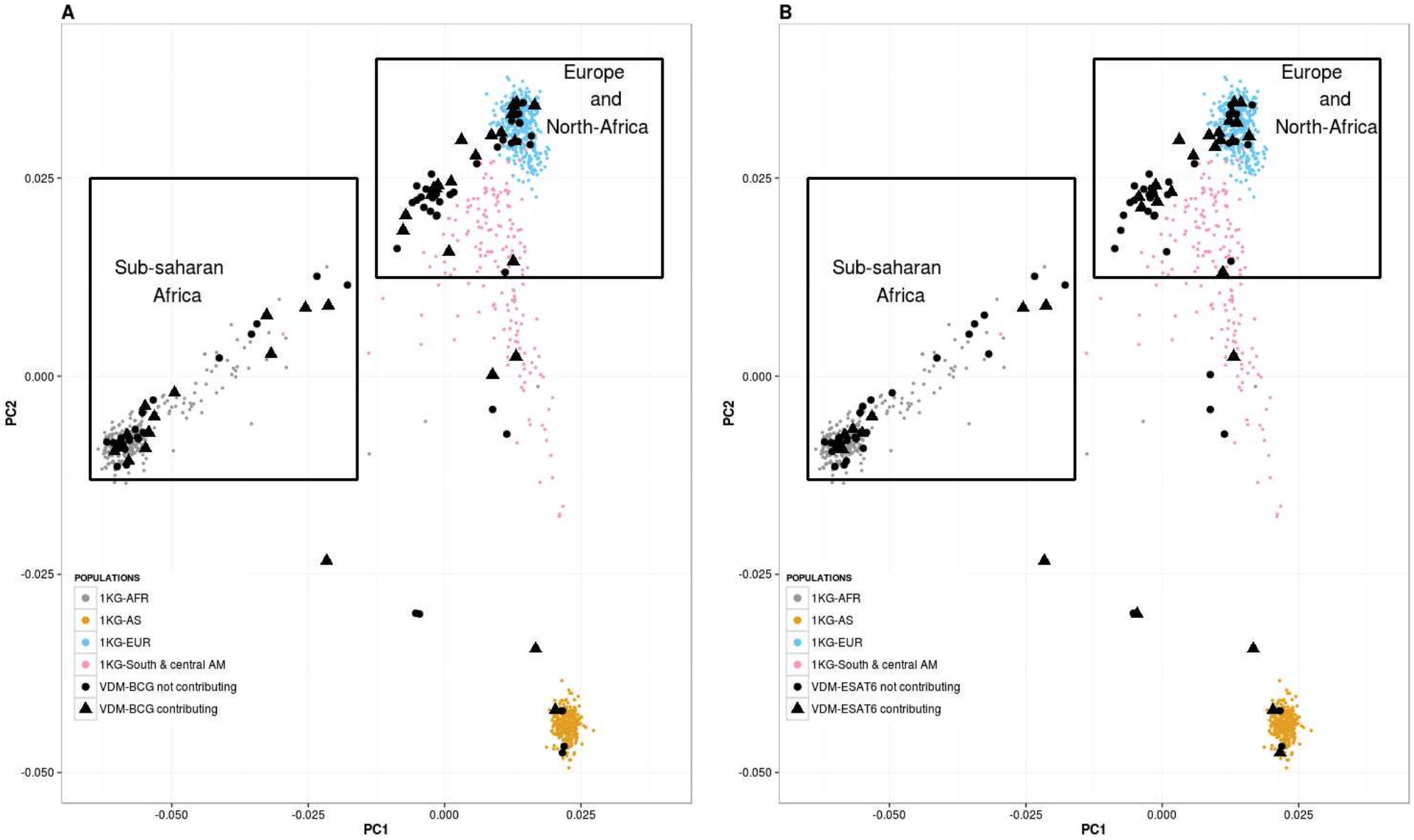



Figure S3 : Population Structure of the  Val de Marne sample

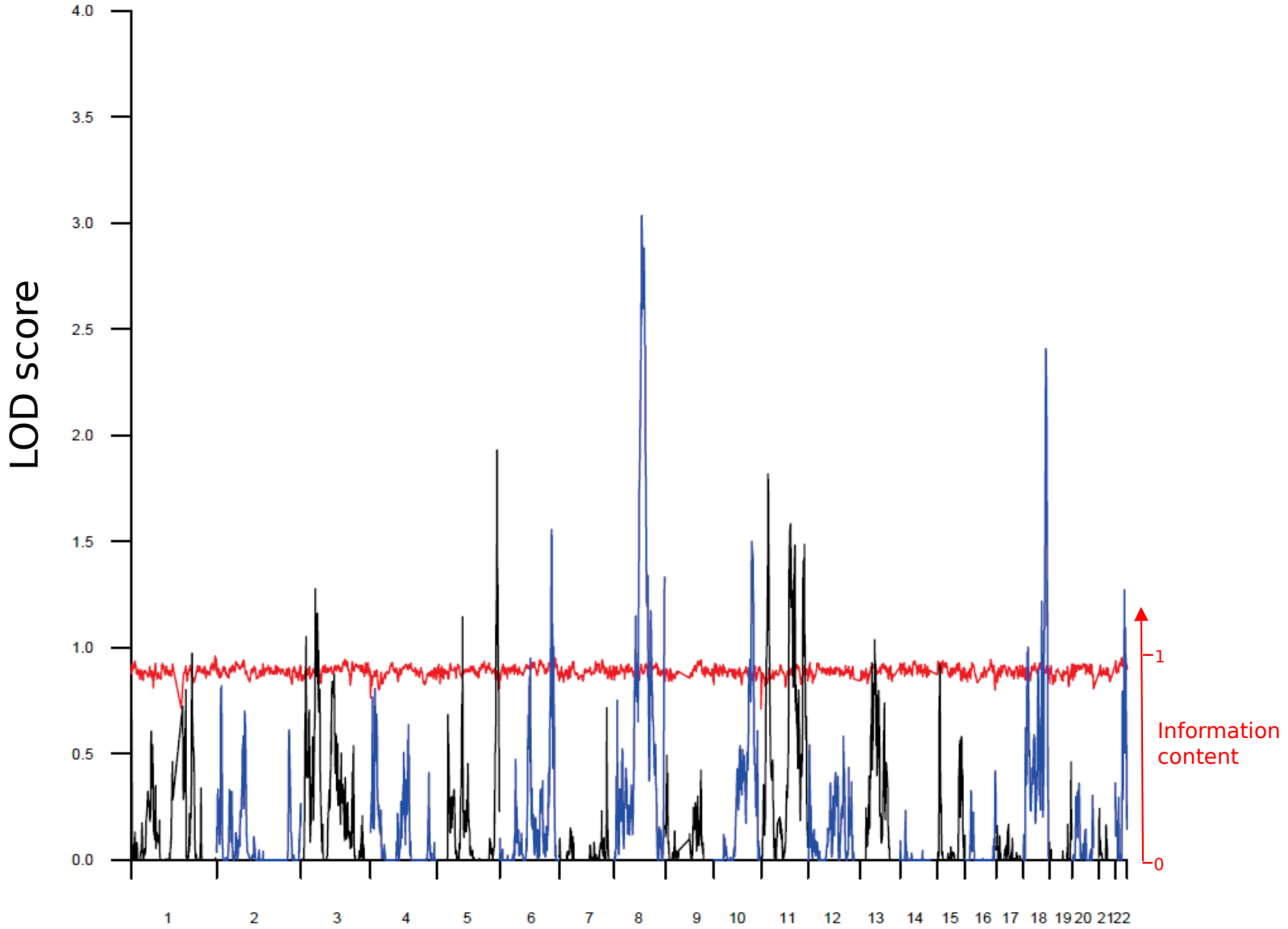



Figure S4 : Model-free genome wide linkage analysis of IFNy-PPD in the Val-de-Marne sample

Multipoint evidence of linkage expressed as a classical LOD score (black and blue lines; left y-axis) and the information content (red line; right y-axis) are plotted along the 22 autosomes.

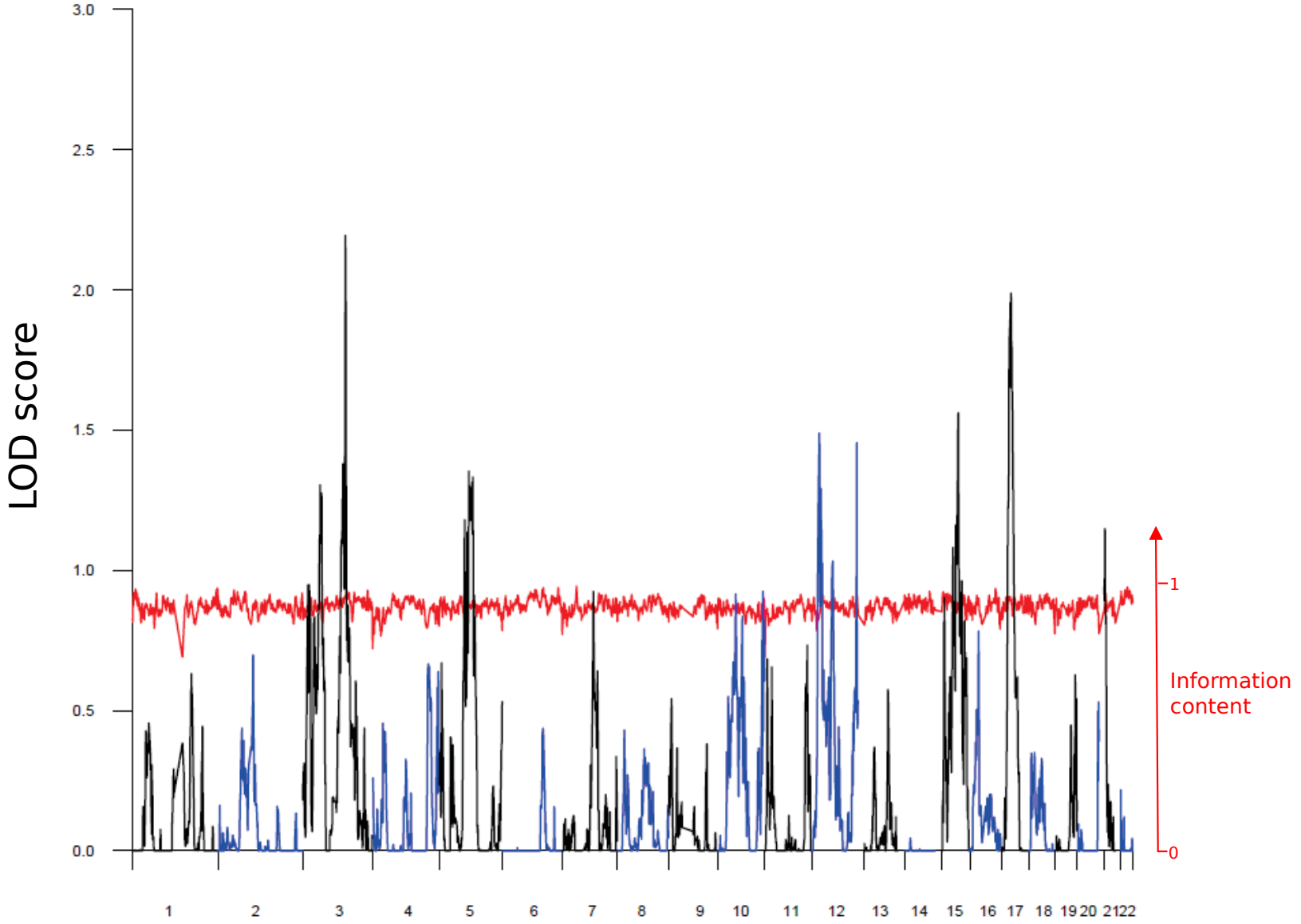



Figure S5 : Model-free genome wide linkage analysis of IFNy-ESAT6 in the Val-de-Marne sample

Multipoint evidence of linkage expressed as a classical LOD score (black and blue lines; left y-axis) and the information content (red line; right y-axis) are plotted along the 22 autosomes.

SCIENTIFIC REPORTS

OPEN

Received: 26 May 2017
Accepted: 12 September 2017
Published online: 09 October 2017

# An eQTL variant of *ZXDC* is associated with IFN-γ production following *Mycobacterium tuberculosis* antigen-specific stimulation

**Fabienne Jabot-Hanin[1,2], Aurélie Cobat[1,2], Jacqueline Feinberg[1,2], Marianna Orlova[3,4,5], Jonathan Niay[6], Caroline Deswarte[1,2], Christine Poirier[7], Ioannis Theodorou[6], Jacinta Bustamante[1,2], Stéphanie Boisson-Dupuis[1,2,8], Jean-Laurent Casanova[1,2,8,11,12], Alexandre Alcaïs[1,2], Eileen G. Hoal[9], Christophe Delacourt[2,10], Erwin Schurr[3,4,5] & Laurent Abel[1,2,8]**

**There is a large inter-individual variability in the response to *Mycobacterium tuberculosis* infection. In previous linkage analyses, we identified a major locus on chromosome region 8q controlling IFN-γ production after stimulation with live BCG (Bacillus Calmette-Guérin), and a second locus on chromosome region 3q affecting IFN-γ production triggered by the 6-kDa early secretory antigen target (ESAT-6), taking into account the IFN-γ production induced by BCG (IFNγ-ESAT6$_{BCG}$). High-density genotyping and imputation identified ~100,000 variants within each linkage region, which we tested for association with the corresponding IFN-γ phenotype in families from a tuberculosis household contact study in France. Significant associations were replicated in a South African familial sample. The most convincing association observed was that between the IFNγ-ESAT6$_{BCG}$ phenotype and rs9828868 on chromosome 3q ($p = 9.8 \times 10^{-6}$ in the French sample). This variant made a significant contribution to the linkage signal ($p < 0.001$), and a trend towards the same association was observed in the South African sample. This variant was reported to be an eQTL of the *ZXDC* gene, biologically linked to monocyte IL-12 production through CCL2/MCP1. The identification of rs9828868 as a genetic driver of IFNγ production in response to mycobacterial antigens provides new insights into human anti-tuberculosis immunity.**

Tuberculosis remains a major public health concern, with approximately 10.4 million new cases and 1.8 million deaths due to the disease in 2015[1]. While an estimated one third of the world population is estimated to be infected with *Mycobacterium tuberculosis*, only about 10% of infected individuals go on to develop clinical

[1]Laboratory of Human Genetics of Infectious Diseases, Necker Branch, INSERM U1163, Paris, France. [2]Paris Descartes University, Sorbonne Paris Cité, Imagine Institute, Paris, France. [3]Program in Infectious Diseases and Immunity in Global Health, The Research Institute of the McGill University Health Centre, Montreal, Canada. [4]McGill International TB Centre, McGill University, Montreal, Canada. [5]Department of Human Genetics and Department of Medicine, McGill University, Montreal, Canada. [6]Université Pierre et Marie Curie, UF d'Histocompatibilité et Immunogénétique, Département d'Immunologie, Groupe Hospitalier Pitié Salpêtrière - Charles Foix, Paris, France. [7]Centre de Lutte Anti-Tuberculeuse, Centre Hospitalier Intercommunal de Créteil, Créteil, France. [8]St Giles Laboratory of Human Genetics of Infectious Diseases, Rockefeller Branch, Rockefeller University, New York, NY, USA. [9]Molecular Biology and Human Genetics, MRC Centre for Molecular and Cellular Biology, DST/NRF Centre of Excellence for Biomedical TB Research, Faculty of Health Sciences, Stellenbosch University, Tygerberg, South Africa. [10]Pediatric Pneumology Unit, Necker Hospital for Sick Children, AP-HP, Paris, France. [11]Howard Hughes Medical Institute, New York, NY, USA. [12]Pediatric Hematology-Immunology Unit, Necker Hospital for Sick Children, AP-HP, Paris, France. Correspondence and requests for materials should be addressed to L.A. (email: laurent.abel@inserm.fr)

disease[2]. There is no direct proof of latent *M. tuberculosis* infection (hereafter referred to simply as LTBI) in exposed individuals, and the infection phenotype is inferred indirectly from quantitative measurements of anti-mycobacterial immunity[2]. The tuberculin skin test (TST) is the most widely used method[3], but additional assays testing for LTBI on the basis of *in vitro* evaluations of T-cell antimycobacterial immunity, have been developed over the last 15 years[4]. These tests measure the production of interferon–gamma (IFN-γ) by circulating leukocytes (IFN-γ release assays, IGRAs) in response to M. tuberculosis antigens, such as the 6 kDa early secretory antigen target (ESAT-6)[5].

Based on TST and IGRA results, an estimated 10%–20% of subjects do not become infected with *M. tuberculosis* despite sustained exposure and, hence, never develop disease[2,6]. Several studies focusing on TST reactivity have provided evidence for the role of human genetic factors in different steps of the infection process[7–11]. IGRA phenotypes have been less studied, but the heritability of IFN-γ secretion has been estimated at about 43% following BCG stimulation and 58% following ESAT-6 stimulation in South Africa[12], and at 17%–48% following stimulation with *M. tuberculosis* antigens, including ESAT-6, in Uganda, depending on the TST status of those tested[13,14].

In a recent linkage analysis, we identified two major loci controlling IFN-γ production induced by mycobacterial stimuli in populations of various ethnic origins living in different *M. tuberculosis* exposure settings[15]. A locus on chromosome 8q12-22 was implicated in IFN-γ production after live BCG stimulation, whereas a second locus on 3q13-22 was found to control IFN-γ levels upon ESAT6 stimulation, accounting for some of the IFN-γ production induced by BCG. In this study, we performed comprehensive fine mapping for these two loci, through high-density genotyping and imputation in the two familial samples used in our previous study[15].

## Materials and Methods

**Subjects and families.** A prospective study of household TB contacts was conducted in the Val-de-Marne, in the Greater Paris region, as previously described[15,16]. Val-de-Marne is an area of low TB endemicity, with an annual TB incidence of 22.1 cases per 100,000 at the time of study, versus an overall incidence of 8.8 per 100,000 in France. From April 2004 to January 2009, household contacts exposed to a patient with culture-confirmed pulmonary TB were enrolled in the context of a general screening procedure (Supplemental Methods). This study was approved by the French Consultative Committee for the Protection of Persons Involved in Biomedical Research (CCPPRB; an IRB) of Henri Mondor Hospital (Créteil, France). Written informed consent was obtained from all study participants, and from the parents of all minors/children enrolled. As a replication cohort, we used a familial sample from the Ravensmead and Uitsig suburbs near Tygerberg, Cape Town, South Africa, where TB is hyperendemic[17]. The Tygerberg families were part of the sample used to map the *TST1* and *TST2* loci[10], and to study the heritability of antimycobacterial immunity[12].

We confirm that all the methods used were performed in accordance with relevant guidelines and regulations.

**Measurement of IFN-γ production.** For the Val–de-Marne sample, blood samples were collected from each individual and peripheral blood mononuclear cells (PBMCs) were isolated and activated with ESAT-6, PPD, live BCG, and phytohemagglutinin (PHA), as previously described[15]. For the Cape Town sample, IGRAs were performed in quadruplicate on whole blood, with BCG, PPD, ESAT-6, and PHA used for stimulation, as described in our previous study[18]. IFN-γ levels were determined 3 and 7 days after stimulation, but, to ensure comparability with the French discovery sample, we restricted the analysis to the measurements made on day three, as previously discussed[15].

**Phenotypes and covariates of interest.** We used the same phenotypes and covariates as for the linkage analyses[15], and we focused on the two phenotypes for which significant evidence for linkage was obtained. The first phenotype corresponds to IFN-γ production following BCG stimulation after classical log-transformation and subtraction of the non-stimulated control value. This phenotype, IFNγ-BCG, was adjusted by linear regression for age and covariates relating to individual levels of exposure to *M. tuberculosis*, as previously described[15]. The second phenotype, IFNγ-ESAT6$_{BCG}$, corresponds to IFN-γ production after ESAT-6 stimulation, which was assessed with the same strategy as for the first phenotype. It was further adjusted for IFNγ-BCG, to isolate the more specific response to the ESAT-6 antigen, taking into account the overlapping effects of BCG and ESAT-6 stimulation. The distributions of the two adjusted phenotypes were close to normality (Figure S1).

**Genotyping and Imputation.** For the French sample, we used the Illumina HumanOmniExpressExome BeadChip to genotype children and their parents for the genetic association analysis. Individuals with a call rate <90% and duplicates based on identity-by-descent statistics calculated with PLINK1.9 software[19,20] were removed from the analysis. Single-nucleotide polymorphisms (SNPs) with a call rate <99% were also removed from the analysis. Following quality control filtering, 743,735 high-quality autosomal SNPs and 489 individuals (out of the 528 individuals previously used in the linkage analyses[15]) from 232 families for whom phenotyping data were available were retained for the analyses. The Tygerberg sample was genotyped with the Illumina HumanOmni2.5 BeadChip and, after quality control according to the same criteria as for the French sample, we retained a total of 2,241,954 autosomal SNPs from 373 individuals from 157 families for whom phenotyping data were available, for association analyses.

From the genotyped SNPs, we imputed additional SNPs across the two linkage regions on chromosomes 3 and 8, using the 1000 Genomes Phase 1 reference panel to increase the density of markers in these regions. The two regions of interest extended from 115 Mb to 139 Mb on chromosome 3, and from 61 Mb to 91.5 Mb on chromosome 8, as defined in our previous study[15]. Indeed, given the variability of estimates of location in linkage studies of complex traits, and the slight differences in phenotype definition between the samples used to determine the linkage loci position, it seems reasonable to consider a rather large confidence interval based on the summed LOD scores curves obtained for Val-de-Marne and Tygerberg sample, in our original linkage analyses[15]. For

imputation, we first used SHAPEIT software[21] to pre-phase separately the Illumina HumanOmniExpressExome and HumanOmni2.5 M genotype data for SNPs that passed quality control. We then used IMPUTE2[22,23] with the 1000 Genomes Phase 1 integrated reference panel to impute the SNP genotypes for the two samples. Imputed SNPs with an information criterion >0.6 and a minor allele frequency (MAF) >0.02 were retained for further analyses. Imputed SNPs significantly associated with either of the two phenotypes of interest were genotyped in the two samples with the high-throughput SEQUENOM iPLEX MassARRAY platform or TaqMan SNP genotyping assays (Applied Biosystems Inc., Foster City, CA).

**Association analysis.** Linkage signals were mostly and primarily identified in the Val-de-Marne sample and replicated in the Tygerberg sample. We therefore also used a two-step strategy for association analysis. We first performed a region-wide association analysis on the larger Val-de-Marne sample, and we then tested the replication of the most significant signals in the two regions of interest in the Tygerberg sample. This strategy was also driven by the fact that phenotypes were very similar in the two samples, but not identical, precluding a combined analysis.

Analyses of association between the high-quality SNPs and the two phenotypes (IFNγ-BCG and IFNγ-ESAT6$_{BCG}$) were performed with linear mixed models (LMM) in GEMMA software[24] to take into account the familial relationships within our samples. The LMM approach is appropriate and robust for family-based association studies, and generally provides a higher power than traditional family-based methods[25]. The relationship matrix used in the regression model was estimated with genotyped genome-wide SNP data and the imputed dosage data were used first-line in the association analyses. In addition, we also performed a principal component analysis (PCA) for the French sample, with the EIGENSTRAT method[26] as previously described[15]. The five first principal components were used as fixed covariates for adjustment in the association analyses for the Val-de-Marne cohort, to take into account the ethnic heterogeneity of the cohort in LMM analyses, as previously described[27]. Based on this PCA, we classified the individuals of the French sample into three subpopulations: Caucasian individuals (grouping together all individuals of European or North African origin), individuals originating from sub-Saharan African and those of Asian origin.

Each of the two regions of interest has a length corresponding to about 1% of the whole genome. We therefore considered $5 \times 10^{-6}$ to be a reasonable region-wide significance threshold in our analyses, based on a genome-wide threshold of $5 \times 10^{-8}$. We checked that this threshold was appropriate by estimating more accurate region-wide thresholds based on the effective number of independent markers in each region[28], and by taking into account the observed genomic inflation factors for each phenotype (Supplemental Methods).

SNPs yielding a p-value for association $< 5 \times 10^{-5}$ in the Val-de-Marne sample with at least one of the inheritance models tested (additive, recessive or dominant) were assessed in replication analyses in the Tygerberg sample, in which we checked for associations in the same direction for the selected alleles. We excluded the imputed SNPs missing from the 1000 Genomes project Phase 3[29] from the replication analysis, to focus on the most reliable variants. Following the replication analysis, we selected SNPs for genotyping if they were associated with a $p$-value $< 0.05$ (one-tailed) in the Tygerberg sample or if they had an initial $p$-value $< 10^{-5}$ in the Val-de-Marne sample and a trend for association was observed in the Tygerberg sample (i.e. a one-tailed p-value <0.5 with the same genetic model as for the French sample). A final association analysis was conducted on the genotyped SNPs in the two samples, with the same genetic model.

**LOD score contributions.** We investigated whether the associated variants could, at least to some extent, explain the two linkage signals, by adjusting the corresponding phenotypes (IFNγ-BCG or IFNγ-ESAT6$_{BCG}$) according to the genotypes at the associated SNPs. Using the original Illumina Linkage IVb markers, we then performed a linkage analysis on this adjusted phenotype, using the maximum likelihood binomial (MLB) model-free method[30] with a trait distribution in deciles, as previously described[15]. We assessed whether the decrease in LOD-score observed after adjustment corresponded to a significant contribution of the variants to the linkage signals, by carrying out the same phenotype adjustment and linkage analysis on 1215 and 1109 randomly selected variants belonging to the OmniExpress beadchip, with a MAF >0.02 and a pairwise correlation coefficient $r^2$ <0.5 from the same linkage regions on chromosome 3 and chromosome 8, respectively. Analyses of these variants provided empirical distributions of LOD-scores for comparison with the values obtained for the associated SNPs. We investigated the LD pattern of the most interesting associated SNPs within the five superpopulations of the 1000 genomes project Phase 3 (Europeans, East Asians, South Asians, Africans and Admixed Americans) with the LDlink web application (https://analysistools.nci.nih.gov/LDlink/)[31].

## Results

**The IFNγ-BCG phenotype.** We first investigated the genomic region linked to the IFNγ production in PBMCs following live BCG stimulation. In the region of interest on chromosome 8 extending from 61 Mb to 91.5 Mb, 6219 variants were genotyped in the French sample. After phasing with SHAPEIT[21], more than 324,000 variants were imputed with IMPUTE2[23] in this 30.5 Mb region. After quality control, we retained a total of 117,354 variants with a MAF >2% and an information criterion >0.6 for analysis with GEMMA software[24] for association with the IFNγ-BCG phenotype (Fig. 1A). In total, 23 variants from four different LD clusters had $p$-values for association $< 5 \times 10^{-5}$. The strongest association signals were obtained with rs202163431 ($p = 2.4 \times 10^{-6}$, LD cluster 8-2, information criterion = 0.65), and rs6981743 ($p = 2.7 \times 10^{-6}$, LD cluster 8-4, information criterion = 0.98). Both these SNPs are intergenic and located more than 150 kb away from the nearest protein-coding gene (Table S1).

We carried out a replication analysis for these 23 associated SNPs in the South African sample. Only the five SNPs of cluster 8.3 met the criteria for replication. One of these five SNPs, rs12056450, was selected for genotyping, as it was also one of the most significant SNPs in the South African sample. Genotyping was successful in 368

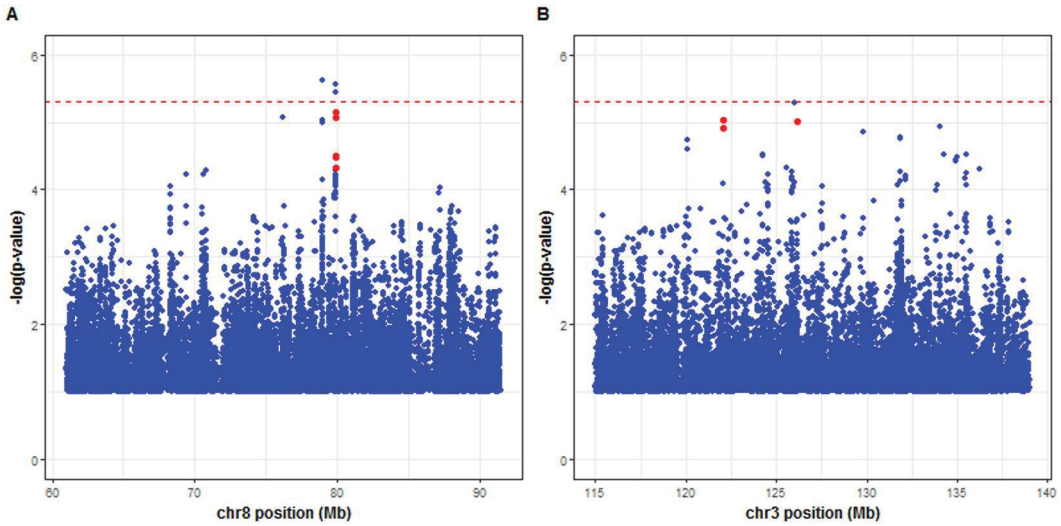



**Figure 1.** Manhattan plots displaying genetic association results for 489 related individuals from the Val-de-Marne sample using a linear mixed model approach implemented in GEMMA software (**A**) The IFNγ-BCG phenotype across 117354 SNPs in the chromosome 8 region from 61 Mb to 91.5 Mb, and (**B**) The IFNγ-ESAT6$_{bcg}$ phenotype across 93218 SNPs in the chromosome 3 region from 115 Mb to 139 Mb. The −log 10 value of the minimum *p*-value obtained in the additive, dominant and recessive tests, is displayed against chromosomal position, in Mb, in the chromosomal region concerned. A horizontal line at a −log10 *p* value of $5 \times 10^{-6}$ indicates the significance threshold, and points in red represent the SNPs belonging to clusters investigated in more detail after replication analyses.

| LD cluster* | Position (bp) | SNP | Alleles** | Genetic Model*** | Val-de Marne Sample | | | Tygerberg Sample | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | | | | | AF ** | Estimated effect(SE)# | p-value | AF | Estimated effect(SE)# | p-value |
| **IFNγ_BCG** | | | | | | | | | | |
| 8-3 | 79887368 | rs12056450 | G/A | Additive | 0.31 | 0.36 (0.08) | $1.2 \times 10^{-5}$ | 0.23 | 0.06 (0.09) | 0.25 |
| **IFNγ-ESAT6$_{bcg}$** | | | | | | | | | | |
| 3-2 | 122059775 | rs9784373 | T/A | Dominant | 0.05 | 0.72 (0.17) | $1.9 \times 10^{-5}$ | 0.02 | 0.31 (0.30) | 0.14 |
| 3–5 | 126129646 | rs9828868 | T/C | Recessive | 0.49 | 0.49 (0.11) | $9.6 \times 10^{-6}$ | 0.46 | 0.12 (0.13) | 0.19 |

**Table 1.** Association results for IFNγ-BCG and IFNγ-ESAT6$_{bcg}$ phenotypes, based on genotyping data for the 3 selected variants according to the criteria described in the Methods, for the French and South African samples. *LD cluster as defined in Tables S1 and S3. **The first mentioned allele is associated with high phenotype values and AF = allele frequency for the first allele mentioned. ***Genetic model for the allele mentioned in the Table. #Estimated effect = regression coefficient with its standard error (SE).

individuals from the French sample and in 236 individuals from the South African sample. In the genotyped individuals, the concordance between the imputed genotypes and the real genotypes was 0.96 for the Val-de-Marne sample and 0.99 for the Tygerberg sample, confirming the high quality of imputation (Table S2). We therefore replaced the imputed dosage data with the real genotypes when available or with best-guess genotypes otherwise, and repeated the association analyses for this SNP. The results of the association study are shown in Table 1. With an additive model, rs12056450 SNP had a slightly lower *p*-value ($1.16 \times 10^{-5}$) in the Val-de-Marne sample, but this SNP was not significantly replicated in the Cape Town sample ($p = 0.25$) with the same genetic model. The frequency of the minor G allele ranged from 0.13 in subjects of African origin to 0.47 in Caucasians from the French sample, with an intermediate value of 0.23 for the South African sample. The effect of the SNP, with the allele G associated with high IFNγ-BCG values, displayed some heterogeneity between populations, with African GG homozygotes having a very low phenotype value in the French sample, and AG heterozygotes having slightly lower mean phenotype values than AA homozygotes in the South African sample (Figure S2).

The cluster of SNPs tagged by rs12056450 included 31 variants with $r^2$ values >0.8 in the French sample. This 40 kb block is located in an intergenic region starting at the end of the LOC105375914 non-coding RNA gene, and the nearest protein-coding gene, encoding IL7, is located 140 kb away. Among the bin SNPs, rs12682556 ($r^2 = 0.97$ with rs12056450 in the Val-de-Marne sample, *p*-value for association with IFNγ-BCG of $5.1 \times 10^{-5}$) is referenced as corresponding to a 400 bp binding region of the CTCF transcription factor in the RegulomeDB database[32]. Finally, we investigated whether rs12056450 contributed to the linkage signal observed on chromosome 8 for the French sample, by adjusting the IFNγ-BCG values for the corresponding SNP. Adjustment for rs12056450 decreased the LOD score from 3.80 to 3.44. We assessed the significance of this result, by calculating an empirical

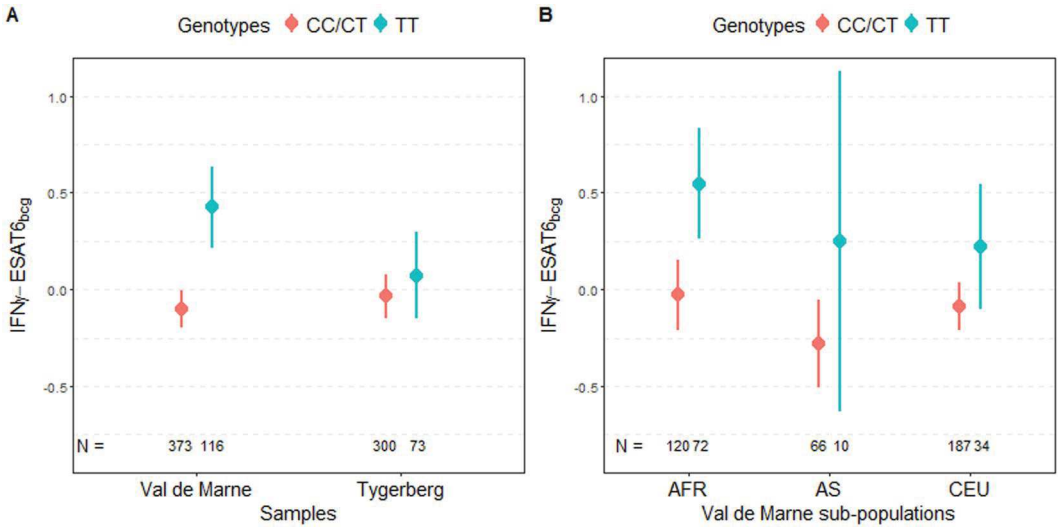



**Figure 2.** Distribution of IFNγ-ESAT6$_{bcg}$ means by rs9828868 genotype in (**A**) Val-de-Marne and Tygerberg samples, and in (**B**) different subpopulations of the Val-de-Marne sample. The dots correspond to the means and the error bars correspond to the 95% confidence interval of the mean calculated under an assumption of normality. The IFNγ-ESAT6$_{bcg}$ phenotype was standardized.

distribution of LOD scores (see methods). We found that the decrease in LOD score observed with rs12056450 was not significant (empirical *p*-value of 0.14) (Figure S3A).

**The IFNγ-ESAT6$_{BCG}$ phenotype.** Next, we focused on the locus impacting the IFN-γ production after *M. tuberculosis* specific ESAT-6 stimulation adjusted for the IFN-γ amount triggered by live BCG. In total, 5901 variants were genotyped and more than 249,000 were imputed within the 24 Mb linkage region of chromosome 3, by the same strategy as used for chromosome 8. Studies of association with the IFNγ-ESAT6$_{BCG}$ phenotype were conducted in the French sample, with 93,218 variants having a MAF >2% and an information criterion >0.6. Figure 1B shows the results for the most significant association for the three genetic models tested (additive, recessive, dominant). In total, 17 variants from nine independent LD clusters provided a *p*-value for association $< 5 \times 10^{-5}$ (Table S3). The strongest association signal was that for the imputed SNP rs116817490 ($p = 5 \times 10^{-6}$, information criterion = 0.82) located 14 kb downstream from the *KLF15* gene. We investigated these 17 associated SNPs in the Tygerberg sample: three, from two independent clusters, met the criteria for replication described in the methods (Table S3).

Two of these three selected variants (rs9784373 and rs149692729) had been imputed, and were found to be in strong LD. Only one of these two SNPs (rs9784373) was used in subsequent analyses, together with the independent SNP rs9828868, which had already been genotyped in the French sample. The concordance between the imputed and real genotypes was high, ranging from 0.86 to 0.99, confirming the accuracy of imputation (Table S2). We therefore replaced the imputed dosage data with the real genotypes when available or with best-guess genotypes otherwise, and repeated the association analyses for these two SNPs (Table 1). SNP rs9784373 yielded similar results with the dominant model in the French sample, but the initial evidence of association in the Tygerberg sample ($p = 1.7 \times 10^{-5}$) became much weaker and was no longer significant after genotyping ($p = 0.14$). The frequency of this variant was low (<0.04) in most populations other than the African subpopulation of the Val-de-Marne sample (MAF = 0.1), with also some heterogeneity of the genetic effect in the Caucasian subpopulation of the Val de Marne sample (Figure S4), making this association result more difficult to interpret.

SNP rs9828868, which had already been genotyped in the French sample (association *p*-value of $9.6 \times 10^{-6}$ under a recessive model), displayed a slight improvement in its *p*-value for association after genotyping in the Tygerberg sample, from 0.22 to 0.19 under the same recessive model (Table 1). Its minor allele T had a frequency of 0.49 in the French sample and 0.46 in the Tygerberg sample. Homozygous TT individuals had higher IFNγ-ESAT6$_{bcg}$ values than CC and CT individuals (difference of ~0.5 standard deviations in the French sample) (Fig. 2A). This effect was homogeneous in the three main populations (Caucasian, African, and Asian) of the French sample (Fig. 2B). Overall, this SNP accounted for 15% of the genetic variance of the distribution of the IFNγ-ESAT6$_{bcg}$ phenotype in the French sample. We investigated the LD pattern of rs9828868 in the populations of the 1000 Genomes project Phase 3. We found only one SNP with an $r^2 > 0.8$, rs4679239, in European and Asian superpopulations. This SNP was imputed, and was not strongly associated with the IFNγ-ESAT6$_{bcg}$ phenotype in the French sample. This association therefore appears to be driven by a single SNP, rs9828868, located in intron of the *CFAP100* gene (cilia and flagella associated protein 100 or *CCDC37*) (Fig. 3).

Finally, we investigated whether these two putative associated variants could account, at least in part, for the chromosome 3 linkage signal, by adjusting the IFNγ-ESAT6$_{bcg}$ values for the corresponding SNPs. Following the same strategy as for the IFNγ-BCG phenotype, we computed an empirical distribution of LOD scores (see

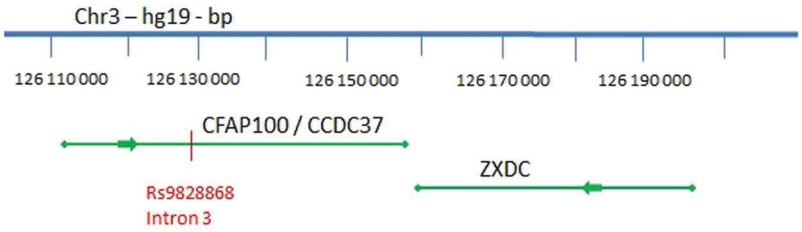



**Figure 3.** Localization of rs9828868 on chromosome 3 according to hg19 coordinates. The SNP is located within the *CFAP100* gene (cilia and flagella associated protein 100), approximately 30 kb from its target gene *ZXDC* (zinc finger X-linked duplicated family member C).

methods). After adjustment for rs9784373, the LOD score in the French sample was 3.05, close to the initial value of 3.26 obtained with the same individuals (empirical *p*-value of 0.22). By contrast, when we adjusted for rs9828868, the LOD score decreased from 3.26 to 2.05. This fall in LOD score was highly significant (empirical $p$-value $< 10^{-3}$), and was larger than those obtained with 1215 randomly chosen independent variants (Figure S3B). Overall, our analyses of the chromosome 3 region identified rs9828868 as the SNP for which the evidence for association with the IFN$\gamma$-ESAT6$_{bcg}$ phenotype was the strongest. Interestingly, rs9828868 has also been reported to be a significant expression quantitative trait locus (eQTL) associated with expression of the nearby *ZXDC* (zinc finger X-linked duplicated family member C) gene in whole blood cells[33].

## Discussion

In this study, we conducted fine mapping of two previously described linkage loci[15] through high-density genotyping and imputation. For the 8q12-22 locus controlling the production of IFN-$\gamma$ by PBMCs following stimulation with live BCG, we identified, in the French sample, a suggestive association with a cluster of intergenic SNPs tagged by rs12056450. This cluster presented the same trend of association in the Tygerberg sample, with the same genetic model as in the Val-de-Marne. However, the effect on IFN-$\gamma$ levels observed in AG heterozygotes in the French sample was not observed in AG heterozygotes from the South African sample. This cluster of SNPs did not explain a substantial part of the linkage signal for chromosome 8, and further studies are required to confirm or rule out a role for this cluster in IFN-$\gamma$ production in response to BCG stimulation.

The second major locus on chromosome 3q13-22 was found to control the IFN-$\gamma$ production induced by ESAT-6 antigen after taking into account the amount shared with that induced by BCG. Our analyses identified a single common C/T variant, rs9828868, with a $p$-value of $9.6 \times 10^{-6}$ in the French sample and a trend for association, under the same genetic model, in the Tygerberg sample. This weaker association in the Tygerberg sample was not unexpected given the weaker linkage signal obtained in the primary study. Subjects homozygous for the T allele had higher values than individuals with C/T and CC genotypes, with a difference of ~0.5 SD in the French sample, accounting for 15% of the genetic variance of the IFN$\gamma$-ESAT6$_{BCG}$ phenotype. This effect was homogeneous across the three main subpopulations of the Val-de-Marne sample (Caucasians, Africans, and Asians) (Fig. 2B). This variant also made a significant contribution to the linkage signal on chromosome 3, providing strong evidence for a genuine association.

SNP rs9828868 was reported to be an eQTL of the *ZXDC* gene in blood cells, with the T allele of the variant being associated with low levels of *ZXDC* expression[33]. The product of *ZXDC* was first described as a zinc finger protein that binds CIITA and contributes to the transcription of MHC class II genes[34]. It also regulates the expression of genes involved in monocyte differentiation and function. In particular, the largest isoform, ZXDC1, activates the expression of *CCL2* (chemokine ligand 2, also known as monocyte chemoattractant protein 1, MCP-1) by evicting the transcriptional repressor BCL6[35]. *ZXDC* knockdown leads to an increase in the occupancy of the *CCL2* promoter by BCL6 following PMA induction, and to lower levels of *CCL2* expression. Individuals carrying the T allele of rs9828868, and TT homozygotes in particular, may have lower levels of *ZXDC* expression, resulting in lower levels of CCL2 induction. Several studies of human cells *in vitro* studies have reported that CCL2 inhibits IL-12 production[36], particularly in *M. tuberculosis*-stimulated monocytes[37]. All these observations are consistent with the view that TT homozygotes have higher IFN$\gamma$-ESAT6$_{BCG}$ levels due to an increase in IL-12 production triggered by the *ZXDC*-dependent downregulation of CCL2.

In conclusion, we identified rs9828868 as associated with the production of IFN-$\gamma$ by PBMCs following stimulation with the ESAT-6 antigen, in an ethnically heterogeneous sample from Val-de-Marne, after adjustment for the levels of IFN-$\gamma$ production common to BCG and ESAT-6 stimulation. This common element may reflect a general capacity for IFN-$\gamma$ production via the TCR signaling pathway, whereas the IFN$\gamma$-ESAT6$_{BCG}$ phenotype is thought to be more specific to ESAT-6, and, consequently, to *M. tuberculosis*, as this antigen is absent from the BCG strain. This variant explains a significant part of the linkage peak previously identified for the same sample, and is involved in expression of the *ZXDC* gene, which is biologically linked to IL-12 production. In the Tygerberg replication sample, the same allele was associated with high values of the studied trait, although this association with not significant at the 5% level. The phenotypes used for replication in South Africa were similar, but not identical to those used in the French sample (IFN-$\gamma$ production in whole-blood samples vs. PBMCs, measured at 3 days vs. 4 days, respectively), and these slight differences may have led to a loss of replication power.

Moreover, the two populations differed considerably in terms of their exposure to *M. tuberculosis*. The studied individuals from South Africa live in an area of hyperendemic tuberculosis, in which *M. tuberculosis* transmission occurs preferentially in the community[38]. By contrast, tuberculosis endemicity is low in France, and the design of

the French study targeted household tuberculosis contacts. The two cohorts also differed in terms of genetic background. The families included in the French sample belonged to several different ethnic groups, whereas all the individuals from the replication sample studied were from the South African Coloured ethnic group, a population resulting from an admixture of Khoesans (31%), Bantu-speaking Africans (33%), Europeans (16%), and Asians (20%)[39]. In this context, the result obtained for rs9828868, which seems to be robust to ethnic and environmental heterogeneity, is particularly promising, and provides new clues to the mechanisms of anti-tuberculosis immunity in humans.

## References


1. WHO - Global Tuberculosis Report 2016. Available at: http://www.who.int/tb/publications/global_report/high_tb_burdencountrylists2016-2020.pdf?ua=1. (Accessed: 25th February 2017)
2. O'Garra, A. *et al.* The Immune Response in Tuberculosis. *Annu. Rev. Immunol.* **31**, 475–527 (2013).
3. Reichman, L. B. Tuberculin skin testing. *The state of the art. Chest* **76**, 764–770 (1979).
4. Pai, M., Riley, L. W. & Colford, J. M. Jr Interferon-γ assays in the immunodiagnosis of tuberculosis: a systematic review. *Lancet Infect. Dis.* **4**, 761–776 (2004).
5. Mahairas, G. G., Sabo, P. J., Hickey, M. J., Singh, D. C. & Stover, C. K. Molecular analysis of genetic differences between Mycobacterium bovis BCG and virulent M. bovis. *J. Bacteriol.* **178**, 1274–1282 (1996).
6. Abel, L., El-Baghdadi, J., Bousfiha, A. A., Casanova, J.-L. & Schurr, E. Human genetics of tuberculosis: a long and winding road. *Philos. Trans. R. Soc. B Biol. Sci.* **369**, 20130428 (2014).
7. Jepson, A. *et al.* Genetic regulation of acquired immune responses to antigens of Mycobacterium tuberculosis: a study of twins in West Africa. *Infect. Immun.* **69**, 3989–3994 (2001).
8. Sepulveda, R. L. *et al.* Evaluation of tuberculin reactivity in BCG-immunized siblings. *Am. J. Respir. Crit. Care Med.* **149**, 620–624 (1994).
9. Stein, C. M. *et al.* Genome scan of M. tuberculosis infection and disease in Ugandans. *PloS One* **3**, e4094 (2008).
10. Cobat, A. *et al.* Two loci control tuberculin skin test reactivity in an area hyperendemic for tuberculosis. *J. Exp. Med.* **206**, 2583–2591 (2009).
11. Cobat, A. *et al.* Tuberculin Skin Test Negativity Is Under Tight Genetic Control of Chromosomal Region 11p14-15 in Settings With Different Tuberculosis Endemicities. *J. Infect. Dis.* **211**, 317–321 (2015).
12. Cobat, A. *et al.* High Heritability of Antimycobacterial Immunity in an Area of Hyperendemicity for Tuberculosis Disease. *J. Infect. Dis.* **201**, 15–19 (2010).
13. Tao, L. *et al.* Genetic and shared environmental influences on interferon-γ production in response to Mycobacterium tuberculosis antigens in a Ugandan population. *Am. J. Trop. Med. Hyg.* **89**, 169–173 (2013).
14. Stein, C. M. *et al.* Heritability analysis of cytokines as intermediate phenotypes of tuberculosis. *J. Infect. Dis.* **187**, 1679–1685 (2003).
15. Jabot-Hanin, F. *et al.* Major Loci on Chromosomes 8q and 3q Control Interferon γ Production Triggered by Bacillus Calmette-Guerin and 6-kDa Early Secretory Antigen Target, Respectively, in Various Populations. *J. Infect. Dis.* **213**, 1173–1179 (2016).
16. Aissa, K. *et al.* Evaluation of a Model for Efficient Screening of Tuberculosis Contact Subjects. *Am. J. Respir. Crit. Care Med.* **177**, 1041–1047 (2008).
17. den Boon, S. *et al.* High Prevalence of Tuberculosis in Previously Treated Patients, Cape Town, South Africa. *Emerg. Infect. Dis.* **13**, 1189–1194 (2007).
18. Gallant, C. J. *et al.* Tuberculin skin test and *in vitro* assays provide complementary measures of antimycobacterial immunity in children and adolescents. *Chest* **137**, 1071–1077 (2010).
19. Purcell, S. *et al.* PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am. J. Hum. Genet.* **81**, 559–575 (2007).
20. Chang, C. C. *et al.* Second-generation PLINK: rising to the challenge of larger and richer datasets. *GigaScience* **4**, 7 (2015).
21. Delaneau, O., Marchini, J. & Zagury, J.-F. A linear complexity phasing method for thousands of genomes. *Nat. Methods* **9**, 179–181 (2012).
22. Howie, B., Fuchsberger, C., Stephens, M., Marchini, J. & Abecasis, G. R. Fast and accurate genotype imputation in genome-wide association studies through pre-phasing. *Nat. Genet.* **44**, 955–959 (2012).
23. Howie, B. N., Donnelly, P. & Marchini, J. A flexible and accurate genotype imputation method for the next generation of genome-wide association studies. *PLoS Genet.* **5**, e1000529 (2009).
24. Zhou, X. & Stephens, M. Genome-wide efficient mixed-model analysis for association studies. *Nat. Genet.* **44**, 821–824 (2012).
25. Eu-Ahsunthornwattana, J. *et al.* Comparison of methods to account for relatedness in genome-wide association studies with family-based data. *PLoS Genet.* **10**, e1004445 (2014).
26. Price, A. L., Zaitlen, N. A., Reich, D. & Patterson, N. New approaches to population stratification in genome-wide association studies. *Nat. Rev. Genet.* **11**, 459–463 (2010).
27. Zhang, Y. & Pan, W. Principal component regression and linear mixed model in association analysis of structured samples: competitors or complements? *Genet. Epidemiol.* **39**, 149–155 (2015).
28. Sobota, R. S. *et al.* Addressing population-specific multiple testing burdens in genetic association studies. *Ann. Hum. Genet.* **79**, 136–147 (2015).
29. 1000genomes_phase1_sites_missing_in_phase3. Available at: ftp://ftp.1000genomes.ebi.ac.uk/vol1/ftp/release/20130502/supporting/phase1_sites_missing_in_phase3/. (Accessed: 17th March 2017)
30. 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.* **35**, 46–56 (2011).
31. Machiela, M. J. & Chanock, S. J. LDlink: a web-based application for exploring population-specific haplotype structure and linking correlated alleles of possible functional variants. *Bioinforma. Oxf. Engl.* **31**, 3555–3557 (2015).
32. Boyle, A. P. *et al.* Annotation of functional variation in personal genomes using RegulomeDB. *Genome Res.* **22**, 1790–1797 (2012).
33. Westra, H.-J. *et al.* Systematic identification of trans eQTLs as putative drivers of known disease associations. *Nat. Genet.* **45**, 1238–1243 (2013).
34. Al-Kandari, W. *et al.* The zinc finger proteins ZXDA and ZXDC form a complex that binds CIITA and regulates MHC II gene transcription. *J. Mol. Biol.* **369**, 1175–1187 (2007).
35. Ramsey, J. E. & Fontes, J. D. The zinc finger transcription factor ZXDC activates CCL2 gene expression by opposing BCL6-mediated repression. *Mol. Immunol.* **56**, 768–780 (2013).
36. Braun, M. C., Lahey, E. & Kelsall, B. L. Selective suppression of IL-12 production by chemoattractants. *J. Immunol. Baltim. Md 1950* **164**, 3009–3017 (2000).
37. Flores-Villanueva, P. O. *et al.* A functional promoter polymorphism in monocyte chemoattractant protein-1 is associated with increased susceptibility to pulmonary tuberculosis. *J. Exp. Med.* **202**, 1649–1658 (2005).
38. Verver, S. Transmission of tuberculosis in a high incidence urban community in South Africa. *Int. J. Epidemiol.* **33**, 351–357 (2004).
39. Chimusa, E. R. *et al.* Determining ancestry proportions in complex admixture scenarios in South Africa using a novel proxy ancestry selection method. *PloS One* **8**, e73971 (2013).

## Acknowledgements

This work was supported by the Programme Hospitalier de Recherche Clinique (AOR-04-003); the Legs Poix (Chancellerie des Universités de Paris); the French National Research Agency (grant ANR TBPATHGEN-ANR-14-CE14-0007-01), and under the "Investments for the future" program (grant ANR-10-IAHU-01); the European Research Council (ERC-2010-AdG-268777); the Rockefeller University; the Institut National de la Santé et de la Recherche Médicale; Paris Descartes University; the St. Giles Foundation; the Canadian Institutes of Health Research; the Sequella/Aeras Global Tuberculosis Foundation; and the Government of Canada (Banting postdoctoral fellowship 112932 to A. C.). We are grateful to the Centre de Ressources Biologiques (CRB, CHI Créteil) for DNA management. We thank all members of the community who participated in this study and members of the lab of Human Genetics of Infectious diseases for helpful discussions.

## Author Contributions

E.H., C.D., E.S. and L.A. conceived and designed the study. J.N., C.D., and M.O. carried out the genotyping of individuals. A.C., J.F., M.O., C.P., M.O., I.T., J.B., S.B.-D., and A.A. contributed reagents/materials/analysis tools. F.J.-H. performed the data analysis. F.J.-H. and L.A. interpreted the results and wrote the first draft of the manuscript. J.-L.C., E.S. and E.H. contributed to the manuscript in its final form. All authors reviewed the manuscript.

## Additional Information

**Supplementary information** accompanies this paper at https://doi.org/10.1038/s41598-017-13017-8.

**Competing Interests:** The authors declare that they have no competing interests.

**Publisher's note:** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

# An eQTL variant of *ZXDC* is associated with IFN-γ production following *Mycobacterium tuberculosis* antigen-specific stimulation

Fabienne Jabot-Hanin[1,2], Aurélie Cobat[1,2], Jacqueline Feinberg[1,2], Marianna Orlova[3,4], Jonathan Niay[5], Caroline Deswarte[1,2], Christine Poirier[6], Ioannis Theodorou[5], Jacinta Bustamante[1,2], Stéphanie Boisson-Dupuis[1,2,7], Jean-Laurent Casanova[1,2,7,10,11], Alexandre Alcaïs[1,2], Eileen G. Hoal[8], Christophe Delacourt[9], Erwin Schurr[3,4], and Laurent Abel*[1,2,]

**Supplementary Data**

**Methods - Subjects and families from Val de Marne**

From April 2004 to January 2009, household contacts exposed to a patient with culture confirmed pulmonary tuberculosis (TB) were enrolled in the context of a general screening procedure in Val de Marne, suburb of Paris. In France, all new cases of TB are reported to the health authorities (Centre de Lutte contre la Tuberculose) to organize epidemiologic investigations and to identify contact subjects with the consent of the index case. A household contact was defined as any person sharing the residence of a TB index case during the three months preceding diagnosis of the case. A questionnaire was completed for each individual to assess the risk factors for infection and the familial relationships as detailed in (1,2).

**Methods - Significance thresholds and inflation factors**

For each of the two studied phenotypes, we computed the genomic inflation factor λ on imputed variants with a MAF > 2% and pruned on the basis of $r^2 < 0.1$ using the GenABEL package(6). For the IFNγ-BCG phenotype analysis, we did not use variants located on the chromosome 8, and found a $\lambda = 1.032$. For the IFNγ-$ESAT6_{BCG}$ phenotype analysis, we removed variants located on chromosome 3, and found a $\lambda = 1.025$. We also computed a precise region-wide significance threshold based on the effective number of independent markers in each region to check if our estimation of $5.10^{-6}$ was a reasonable significance threshold for each of the 2 regions. For that, we thinned the imputed markers with a MAF > 2% thanks to PLINK software (3,4) using a window of 5000 SNPs, a shift of 5 SNPs and a $r^2 < 0.2$, as suggested in (5). The thresholds thus calculated were of $1.15\ 10^{-5}$ for the chromosome 8 region and of $1.3\ 10^{-5}$ for the chromosome 3 region. When these thresholds were corrected by the corresponding genomic inflation factor, they became $8.3\ 10^{-6}$ and $1.0\ 10^{-5}$ for chromosomes 8 and 3, respectively. These thresholds were therefore close to our approximated thresholds of $5.10^{-6}$ which appeared quite reasonable and appropriate.

References

1. Aissa K, Madhi F, Ronsin N, Delarocque F, Lecuyer A, Decludt B, et al. Evaluation of a Model for Efficient Screening of Tuberculosis Contact Subjects. Am J Respir Crit Care Med. 2008 May;177(9):1041–7.

2. Cobat A, Poirier C, Hoal E, Boland-Auge A, Rocque F de L, Corrard F, et al. Tuberculin Skin Test Negativity Is Under Tight Genetic Control of Chromosomal Region 11p14-15 in Settings With Different Tuberculosis Endemicities. J Infect Dis. 2015 Jan 15;211(2):317–21.

3. Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MAR, Bender D, et al. PLINK: a tool set for whole-genome association and population-based linkage analyses. Am J Hum Genet. 2007 Sep;81(3):559–75.

4. Chang CC, Chow CC, Tellier LC, Vattikuti S, Purcell SM, Lee JJ. Second-generation PLINK: rising to the challenge of larger and richer datasets. GigaScience. 2015;4:7.

5. Sobota RS, Shriner D, Kodaman N, Goodloe R, Zheng W, Gao Y-T, et al. Addressing Population-Specific Multiple Testing Burdens in Genetic Association Studies. Ann Hum Genet. 2015 Mar 1;79(2):136–47.

6. Karssen LC, van Duijn CM, Aulchenko YS. The GenABEL Project for statistical genomics. F1000Research. 2016 May 19;5:914.

**Supplementary Figures**

Figure S1: Distributions of adjusted standardized phenotypes IFNγ-BCG (A) and IFNγ-ESAT6$_{bcg}$ (B) used for association analyses in the Val-de-Marne sample.

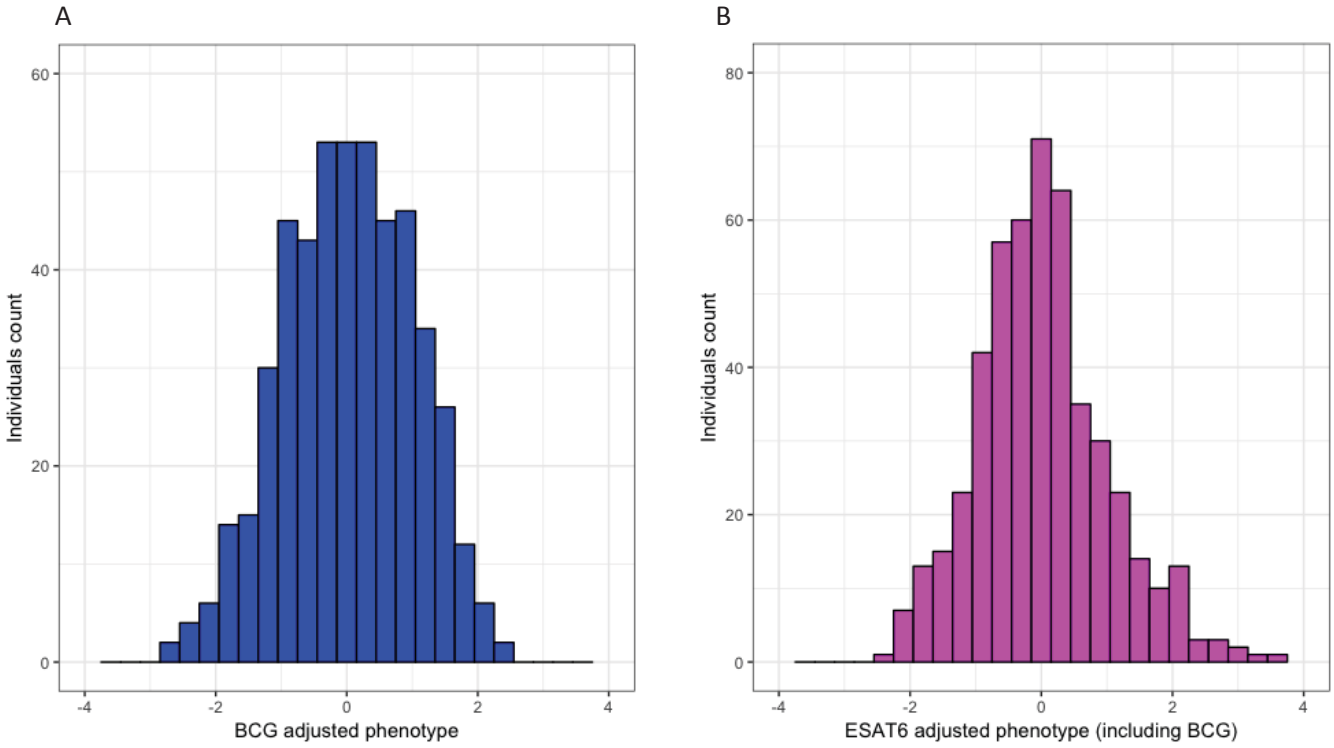


Figure S2: Distribution of the IFNγ-BCG means according to rs12056450 genotypes in Val-de-Marne sample and in Tygerberg sample (**A**) and in the different sub-populations of the Val de Marne sample **(B).** The dots correspond to the means and the error bar to the 95% confidence interval of the mean computed under the normality assumption. IFNγ-BCG phenotype is standardized.

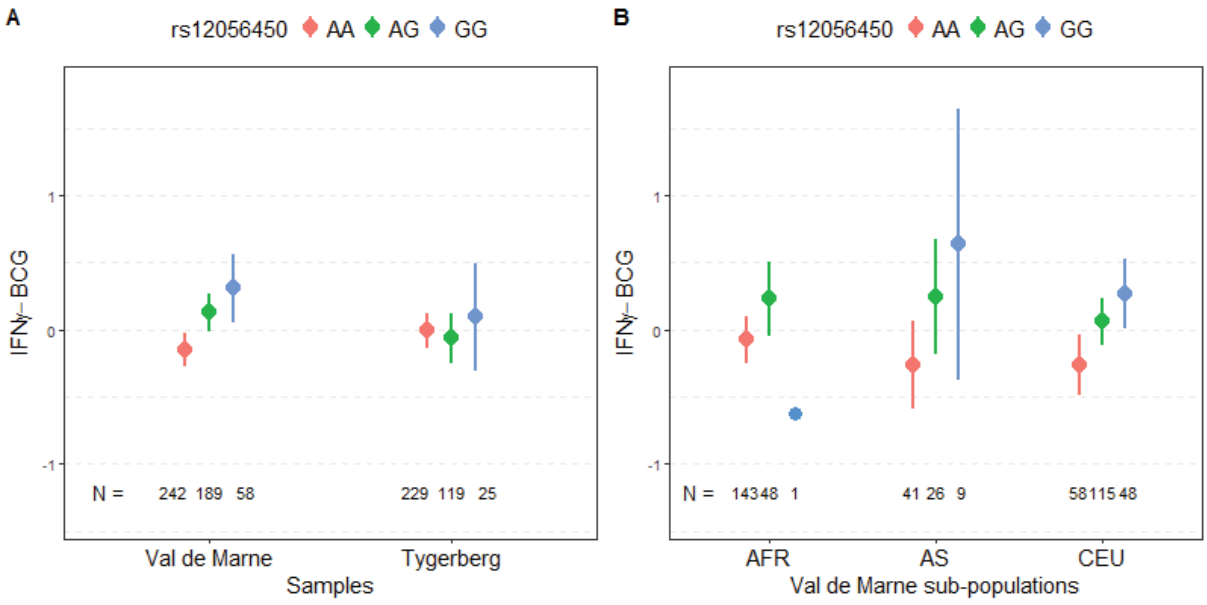


Figure S3: Empirical LOD score distributions used for assessing the contribution of the associated SNPs to the linkage peak. Figure shows the distribution of LOD scores obtained after adjustment of IFNγ-BCG phenotype on 1109 randomly selected imputed variants belonging the chromosome 8 linkage region (115 – 139 Mb) **(A)** and adjustment of IFNγ-ESAT6$_{bcg}$ phenotype on 1215 randomly selected imputed variants belonging the chromosome 3 linkage region (115 – 139 Mb) **(B)**, all with a MAF > 2% and an info criteria >0.6. * corresponds to the LOD score obtained after adjustment on rs12056450 **(A)** and rs9828868 **(B)**. corresponds to the LOD score obtained after adjustment on rs9784373 **(B).**

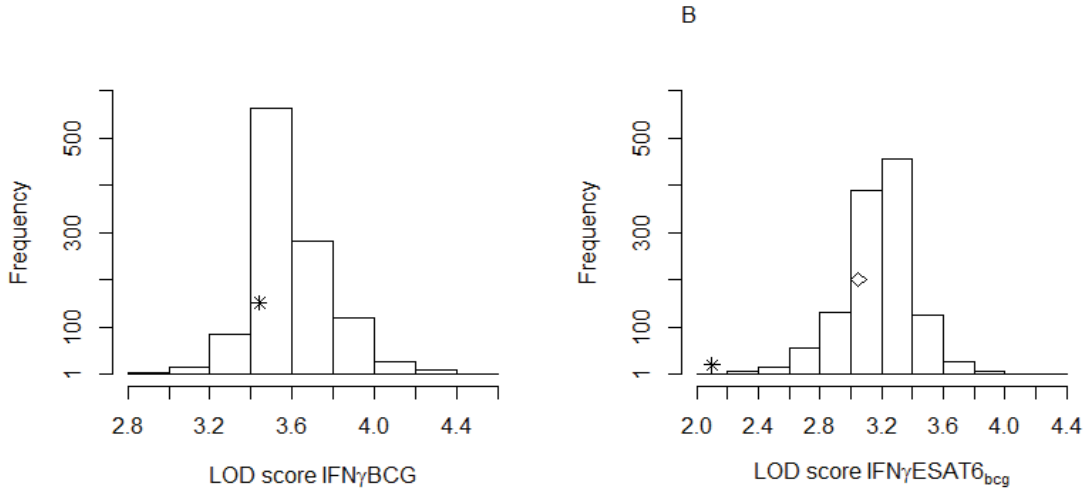
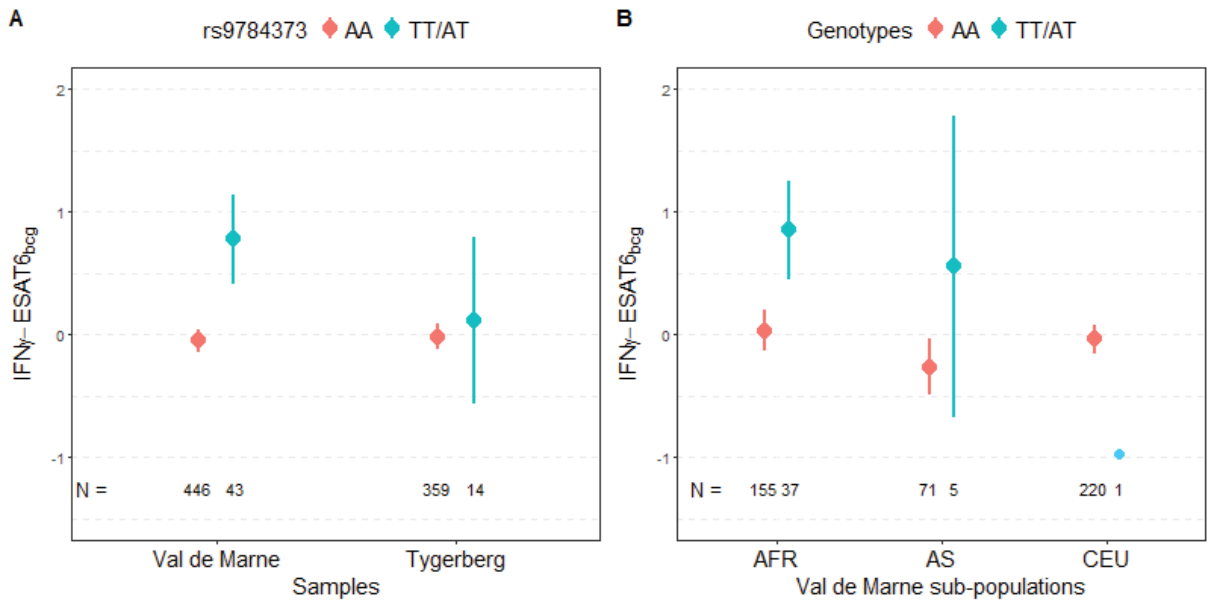


Figure S4: Distribution of the IFNγ-ESAT6$_{BCG}$ means according to rs9784373 genotypes in Val-de-Marne sample and in Tygerberg sample (**A**) and in the different sub-populations of the Val de Marne sample **(B).** The dots correspond to the means and the error bar to the 95% confidence interval of the mean computed under the normality assumption. IFNγ-ESAT6$_{bcg}$ phenotype is standardized.



**Supplementary Tables**

Table S1: Association results on imputed data for IFNγ-BCG phenotype for p-value < $5.10^{-5}$ in Val-de-Marne sample

Table S2: Comparison of the bestguess genotypes coming from IMPUTE2 software versus the real genotypes on 368 individuals from Val-de-Marne and 236 individuals from Tygerberg sample

Table S3: Association results on imputed data for IFNγ-ESAT6$_{bcg}$ phenotype for p-value < $5.10^{-5}$ in Val-de-Marne sample

| | | | | Val-de-Marne sample | | | | | Tygerberg sample | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| LD cluster | position (bp) | SNP | Alleles* | Allele frequency* | Estimated Effect (SE)# | P-value | Genetic model** | information § | Allele frequency* | Estimated Effect (SE)# | P-value*** | Genetic model | information § |
| 8-1 | 76,236,341 | rs79392429 | C / T | 0.90 | 0.59 (0.13) | 8.3E-06 | REC | 0.92 | 0.81 | - | >0.5 | - | 0.90 |
| 8-2 | 78,963,020 | rs117257553 | T / A | 0.04 | 0.80 (0.18) | 9.9E-06 | ADD | 0.95 | 0.01 | - | >0.5 | - | 0.93 |
| | 78,963,622 | rs118187863 | G / A | 0.04 | 0.80 (0.18) | 1.0E-05 | ADD | 0.95 | 0.01 | - | >0.5 | - | 0.93 |
| | 78,966,953 | rs202163431 | T / TA | 0.06 | 0.82 (0.17) | 2.4E-06 | ADD | 0.65 | 0.08 | - | >0.5 | - | 0.64 |
| | 78,978,131 | rs75231543 | A / G | 0.04 | 0.80 (0.18) | 1.0E-05 | ADD | 0.95 | 0.01 | - | >0.5 | - | 0.91 |
| | 78,979,589 | rs141810731 | T / C | 0.04 | 0.80 (0.18) | 1.0E-05 | ADD | 0.95 | 0.01 | - | >0.5 | - | 0.91 |
| | 78,981,194 | rs117762248 | C / A | 0.04 | 0.80 (0.18) | 1.0E-05 | ADD | 0.95 | 0.01 | - | >0.5 | - | 0.91 |
| | 78,982,210 | rs77767122 | A / G | 0.04 | 0.80 (0.18) | 1.0E-05 | ADD | 0.95 | 0.01 | - | >0.5 | - | 0.91 |
| | 78,984,646 | rs118084590 | A / G | 0.04 | 0.80 (0.18) | 1.0E-05 | ADD | 0.95 | 0.01 | - | >0.5 | - | 0.91 |
| | 78,986,849 | rs117536094 | C / T | 0.04 | 0.80 (0.18) | 1.0E-05 | ADD | 0.95 | 0.01 | - | >0.5 | - | 0.91 |
| | 78,987,821 | chr8:78987821:D | C / CT | 0.04 | 0.80 (0.18) | 1.0E-05 | ADD | 0.95 | 0.01 | - | >0.5 | - | 0.91 |
| | 78,991,235 | chr8:78991235:D | C / CATAAGTCTGGG | 0.04 | 0.80 (0.18) | 1.0E-05 | ADD | 0.95 | 0.01 | - | >0.5 | - | 0.91 |
| | 78,995,179 | rs117746665 | C / T | 0.04 | 0.80 (0.18) | 1.0E-05 | ADD | 0.95 | 0.01 | - | >0.5 | - | 0.91 |
| | 78,995,955 | rs78784982 | C / T | 0.04 | 0.80 (0.18) | 1.0E-05 | ADD | 0.95 | 0.01 | - | >0.5 | - | 0.91 |
| | 78,998,300 | rs117528033 | A / G | 0.04 | 0.80 (0.18) | 1.0E-05 | ADD | 0.95 | 0.01 | - | >0.5 | - | 0.91 |
| | 79,003,809 | rs147053766 | C / T | 0.04 | 0.84 (0.19) | 9.2E-06 | ADD | 0.90 | 0.01 | - | >0.5 | - | 1.00 |
| 8-3 | 79,876,744 | rs6991466 | A / G | 0.25 | 0.38 (0.09) | 7.1E-06 | ADD | 0.99 | 0.20 | 0.37 (0.23) | 5.1E-02 | REC | 0.99 |
| | 79,885,854 | rs1427255 | C / T | 0.32 | 0.33 (0.08) | 4.8E-05 | ADD | 0.99 | 0.23 | 0.36 (0.21) | 2.2E-02 | REC | 1.00 |
| | 79,887,184 | rs12056761 | A / G | 0.32 | 0.34 (0.08) | 3.3E-05 | ADD | 0.98 | 0.23 | 0.36 (0.21) | 2.3E-02 | REC | 0.99 |
| | 79,887,368 | rs12056450 | G / A | 0.32 | 0.34 (0.08) | 3.1E-05 | ADD | 0.98 | 0.23 | 0.36 (0.21) | 2.3E-02 | REC | 0.99 |
| | 79,889,640 | rs11781015 | A / G | 0.32 | 0.35 (0.08) | 8.6E-06 | ADD | 0.98 | 0.23 | 0.26 (0.22) | 1.2E-01 | REC | 0.96 |
| 8-4 | 79,888,344 | rs7825423 | C / T | 0.41 | 0.49 (0.10) | 3.6E-06 | DOM | 0.98 | 0.31 | - | >0.5 | - | 0.98 |
| | 79,888,908 | rs6981743 | T / C | 0.39 | 0.48 (0.10) | 2.7E-06 | DOM | 0.98 | 0.30 | - | >0.5 | - | 0.97 |

Table S2 : Association results on imputed data for IFNγ-BCG phenotype for p-value < $5.10^{-5}$ in Val de Marne sample

* The first mentioned allele is associated with high values of the phenotype, and the frequency is given for this allele

** Genetic model defined for Allele associated with high values of phenotype

***p-value for Tygerberg sample = best p-value of the LRT test among the 3 genetic models additive, recessive or dominant, one tailed. p-value > 0,5 means that the allele associated with the highest phenotypes is not the same as in the french sample after looking carefully at the phenotype levels by genotypes

§quality criterion extracted from IMPUTE2 software. Information=1 means that the SNP is genotyped wheras information=1.00 means that the SNP has been imputed.

# Regression coefficient with its standard error (SE)

**Table S3** : Comparison of the bestguess genotypes coming from IMPUTE2 software versus the real genotypes on 368 individuals from Val-de-Marne and 236 individuals from Tygerberg sample

| chr | SNP | Val de Marne | Tygerberg |
|---|---|---|---|
| 3 | rs9784373 | 0.98 | 0.99 |
| 3 | rs9828868 | * | 0.86 |
| 8 | rs12056450 | 0.96 | 0.99 |

* the SNP was part of the genotyped variants in the Illumina HumanOmniExpressExome BeadChip in Val-de-Marne individuals

| | | | | Val-de-Marne sample | | | | | Tygerberg sample | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| LD cluster | position (bp) | SNP | Alleles* | Allele frequency* | Estimated Effect (SE)# | P-value | Genetic model** | information§ | Allele frequency* | Estimated Effect (SE)# | P-value*** | Genetic model | information§ |
| 3-1 | 120,111,733 | chr3:120111733:D | T / TA | 0.02 | 1.02 (0.24) | 1.8E-05 | ADD | 0.96 | 0.01 | - | > 0,5 | - | 0.89 |
| | 120,117,992 | rs114720435 | T / C | 0.02 | 1.00 (0.24) | 2.5E-05 | ADD | 0.97 | 0.01 | - | > 0,5 | - | 1 |
| | 120,129,794 | rs77250558 | T / C | 0.02 | 1.00 (0.24) | 2.5E-05 | ADD | 0.96 | 0.02 | - | > 0,5 | - | 0.92 |
| 3-2 | 122,059,775 | rs9784373 | T / A | 0.05 | 0.81 (0.18) | 9.5E-06 | DOM | 0.90 | 0.02 | 0.32 (0.29) | 1.3E-01 | DOM | 0.93 |
| | 122,060,841 | rs149692729 | G / A | 0.05 | 0.80 ( 0.18) | 1.2E-05 | DOM | 0.89 | 0.02 | 0.38 (0.29) | 9.2E-02 | DOM | 0.94 |
| 3-3 | 124,248,975 | rs74677891 | C / T | 0.05 | 0.66 (0.16) | 3.0E-05 | DOM | 0.96 | 0.02 | - | > 0,5 | - | 1 |
| | 124,253,483 | rs79289633 | G / A | 0.05 | 0.66 (0.16) | 3.2E-05 | DOM | 0.97 | 0.02 | - | > 0,5 | - | 1 |
| | 124,253,581 | rs77945479 | G / A | 0.05 | 0.66 (0.16) | 3.2E-05 | DOM | 0.97 | 0.02 | - | > 0,5 | - | 1.00 |
| 3-4 | 125,991,012 | rs116817490 | C / A | 0.03 | 1.12 (0.25) | 5.0E-06 | DOM | 0.82 | 0.01 | - | > 0,5 | - | 0.66 |
| 3-5 | 126,129,646 | rs9828868 | T / C | 0.49 | 0.49 (0.11) | 9.6E-06 | REC | 1 | 0.49 | 0.11 (0.14) | 2.2E-01 | REC | 0.90 |
| 3-6 | 129,786,628 | rs57026314 | A / G | 0.03 | 1.21 (0.28) | 1.4E-05 | DOM | 0.62 | 0.03 | 0.12 (0.31) | 3.4E-01 | DOM | 0.55 |
| 3-7 | 131,841,434 | rs140762555 | G / A | 0.05 | 0.68 (0.16) | 1.7E-05 | ADD | 0.89 | 0.07 | - | > 0,5 | - | 0.84 |
| | 131,842,891 | rs59712276 | A / G | 0.03 | 0.93 (0.22) | 1.7E-05 | ADD | 0.87 | 0.02 | - | > 0,5 | - | 0.98 |
| | 131,848,904 | rs75435362 | A / T | 0.03 | 0.93 (0.22) | 1.7E-05 | ADD | 0.87 | 0.02 | - | > 0,5 | - | 0.99 |
| | 131,852,560 | rs57202631 | T / C | 0.03 | 0.93 (0.22) | 1.7E-05 | ADD | 0.87 | 0.02 | - | > 0,5 | - | 0.98 |
| 3-8 | 135,484,040 | rs7431617 | G / A | 0.34 | 0.29 (0.07) | 3.1E-05 | ADD | 0.98 | 0.43 | 0.02 (0.08) | 4.2E-01 | ADD | 0.99 |
| 3-9 | 136,283,857 | rs60291974 | C / A | 0.84 | 0.51 ( 0.12) | 4.9E-05 | REC | 0.66 | 0.89 | - | >0,5 | - | 0.62 |

Table S4 : Association results on imputed data for IFN$\gamma$-ESAT6$_{BCG}$ phenotype for p-value < $5.10^{-5}$ in Val de Marne sample

* The first mentioned allele is associated with high values of the phenotype, and the frequency is given for this allele

** Genetic model defined for Allele associated with high values of phenotype

***p-value for Tygerberg sample = best p-value of the LRT test among the 3 genetic models additive, recessive or dominant, one tailed. P min CT > 0,5 means that the allele associated with the highest phenotypes is not the same as in the french sample after looking carefully at the phenotype levels by genotypes

§quality criterion extracted from IMPUTE2 software. Information=1 means that the SNP is genotyped wheras information=1.00 means that the SNP has been imputed.

# Regression coefficient with its standard error (SE)