



**HAL**  
open science

# Facteurs de risque génétiques et environnementaux de l'infection par *Mycobacterium tuberculosis*

Jocelyn Quistrebert

► **To cite this version:**

Jocelyn Quistrebert. Facteurs de risque génétiques et environnementaux de l'infection par *Mycobacterium tuberculosis*. Médecine humaine et pathologie. Université Paris Cité, 2021. Français. NNT : 2021UNIP5229 . tel-04548957

**HAL Id: tel-04548957**

**<https://theses.hal.science/tel-04548957>**

Submitted on 16 Apr 2024

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

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

# Université de Paris

École doctorale Pierre Louis de Santé Publique  
ED 393 Epidémiologie et Sciences de l'Information Biomédicale  
IHU Imagine (Institut des Maladies Génétiques) - UMR\_S 1163

## **Facteurs de risque génétiques et environnementaux de l'infection par *Mycobacterium tuberculosis***

Par Jocelyn QUISTREBERT

Thèse de doctorat d'épidémiologie génétique

Dirigée par Dr Laurent ABEL

Et par Dr Aurélie COBAT

Présentée et soutenue publiquement le 10 février 2021

Devant un jury composé de :

Prof Jacques FELLAY, professeur associé, Université de Lausanne, Suisse (Rapporteur)  
Dr Etienne PATIN, chercheur HDR, Institut Pasteur (Rapporteur)  
Prof Anne BOURGARIT-DURAND, PU-PH, Université Sorbonne Paris Nord (Examinatrice)  
Dr Audrey SABBAGH, MCU, Université de Paris (Examinatrice)  
Dr Laurent ABEL, DR, Université de Paris (Directeur de thèse)  
Dr Aurélie COBAT, CR-HDR, Université de Paris (Co-directrice de thèse)



**Titre :** Facteurs de risque génétiques et environnementaux de l'infection par *Mycobacterium tuberculosis*

**Résumé :** La tuberculose est une maladie infectieuse chronique causée par *Mycobacterium tuberculosis* (*Mtb*). C'est l'une des 10 premières causes de mortalité dans le monde et la première par un unique agent infectieux. Après exposition à *Mtb*, une majorité d'individus deviennent infectés. L'infection est asymptomatique et ne conduit à la tuberculose maladie que dans ~10% des cas. Le diagnostic d'infection tuberculeuse repose sur des tests immunologiques *in vivo* (intradermoréaction, IDR, ou test de Mantoux) ou *ex vivo* (test de production d'interféron-gamma ou IGRA pour *interferon-gamma release assay*). L'objectif global de cette thèse était d'explorer les facteurs de risques génétiques et non génétiques associés à l'infection tuberculeuse.

Mon premier travail s'est focalisé sur une population à haut-risque d'infection: le personnel de santé. Un large échantillon de 631 employés a été recruté au Maroc, un pays où la tuberculose est moyennement endémique et la couverture par le vaccin BCG (Bacille Calmette et Guérin) presque complète. Ces employés ont été testés à la fois par IDR et IGRA. Une forte prévalence de l'infection a pu être observée, allant de 41% si l'on utilisait un test IGRA positif à 52% si l'on utilisait l'IDR avec un seuil de positivité de 10 mm. En définissant un statut d'infection plus robuste par la positivité aux 2 tests, le sexe masculin, l'âge, un antécédent familial de tuberculose et travailler en pneumologie étaient des facteurs de risque associés à l'infection. La consommation de tabac était associée à la positivité à l'IGRA mais pas à l'IDR. Cette étude a confirmé que les résultats d'IDR et d'IGRA ne sont pas pleinement concordants et que, en pratique, un diagnostic d'infection doit tenir compte des propriétés de chaque test et du contexte dans lequel il est réalisé.

Dans un deuxième travail, je me suis intéressé aux facteurs génétiques associés au phénotype d'infection. Il existe en effet de nombreuses preuves épidémiologiques qu'une partie des individus fortement exposés à *Mtb* ne s'infectent pas et la grande variabilité de réponse à une exposition à *Mtb* à environnement égal par ailleurs suggère un rôle important de ces facteurs génétiques. Une cohorte de contacts familiaux a été établie dans une région très endémique du Vietnam à partir du recensement de cas index atteints d'une tuberculose maladie. Ces contacts ont donc été intensément exposés pendant plusieurs semaines/mois et ont été testés par IDR et IGRA. J'ai réalisé une étude d'association pangénomique sur plus de 5 millions de polymorphismes nucléotidiques entre des individus considérés comme résistants à l'infection testés doublement négatifs par IDR et IGRA ( $n = 185$ ), et des individus infectés testés

doublement positifs ou atteints d'une tuberculose pulmonaire ( $n = 353$ ). J'ai pu mettre en évidence une association significative au seuil pangénomique sur un locus du bras long du chromosome 10 (10q26.2) dont les polymorphismes confèrent une forte protection contre l'infection par *Mtb* (pour le marqueur génotypé rs17155120, odds ratio, OR = 0.42,  $P = 3.71 \times 10^{-8}$ ). Cette association a été validée dans une cohorte indépendante constituée de contacts familiaux très exposés en France, en région parisienne, et dans un second échantillon familial recruté en banlieue du Cap, Afrique du Sud, région hyper-endémique pour la tuberculose, avec un OR combiné de 0.50 ( $P = 1.26 \times 10^{-9}$ ) pour le marqueur rs17155120. Ces polymorphismes associés au phénotype sont situés dans l'intron du gène *C10Orf90* (ou *FATS*) codant une ubiquitine ligase E3 qui active le facteur de transcription p53. L'analyse *in silico* a montré que les allèles protecteurs sont associés à une baisse d'expression du gène voisin *ADAM12*, pouvant conduire à une réponse exacerbée de lymphocytes Th17.

En conclusion, ce travail a permis d'identifier des facteurs de risques de l'infection tuberculeuse sur diverses populations fortement exposées. La partie d'épidémiologie descriptive a confirmé des résultats connus dans le contexte non-étudié du personnel de santé au Maroc. Le volet épidémiologie génétique a quant à lui confirmé que la composante génétique de l'hôte contribue au phénotype d'infection en mettant en évidence un locus conférant une forte protection contre *Mtb* dans 3 populations distinctes intensément exposées. Ces résultats n'ont pu être observés qu'à partir d'une définition stringente de l'infection tuberculeuse et la conception rigoureuse de cohortes exposées. Ils soulignent l'importance de ces facteurs de risque dans la recherche de nouvelles interventions pour prévenir, diagnostiquer ou traiter la tuberculose.

**Mots clefs :** Tuberculose ; infection tuberculeuse ; résistance ; intradermoréaction ; IDR ; test de Mantoux ; test de production d'interféron-gamma ; Interferon Gamma Release Assays ; IGRA ; prévalence ; concordance ; Maroc ; étude d'association pangénomique ; *C10Orf90* ; *FATS* ; *ADAM12* ; 10q26.2 ; Vietnam ; France ; Afrique du Sud

# Sommaire

Remerciements .....	6
Liste des figures .....	7
Liste des tableaux .....	9
Liste des publications .....	10
1 Introduction .....	11
1.1 La tuberculose .....	11
1.1.1 Contexte .....	11
1.1.2 Microbiologie .....	14
1.1.3 Histoire naturelle de la tuberculose .....	17
1.2 L'infection par <i>M. tuberculosis</i> .....	19
1.2.1 Physiopathologie .....	19
1.2.2 Diagnostic de l'infection .....	21
1.2.3 Hétérogénéité du phénotype d'infection .....	25
1.2.4 Traitement et prévention.....	29
1.2.5 Facteurs de risque de l'infection .....	32
1.2.6 Facteurs de risque non génétiques.....	33
1.2.7 Facteurs de risque génétiques.....	37
1.3 Objectifs de la thèse .....	40
2 Epidémiologie descriptive et analytique : Prévalence et facteurs de risque de l'infection <i>M. tuberculosis</i> parmi du personnel de santé au Maroc .....	42
2.1 Introduction .....	42
2.2 Matériel et méthodes .....	43
2.3 Résultats .....	44
2.3.1 Description de l'échantillon .....	44

2.3.2 Accord entre les résultats QFT et IDR .....	46
2.3.3 Facteurs de risque associés à l'infection <i>M. tuberculosis</i> .....	47
2.3.4 Facteurs de risque associés aux résultats QFT et IDR discordants .....	50
2.4 Discussion .....	51
3 Épidémiologie génétique : Analyse d'association pangénomique de la résistance à l'infection par <i>M. tuberculosis</i> dans 3 populations distinctes .....	55
3.1 Introduction .....	55
3.2 Description des cohortes .....	55
3.2.1 Vietnam .....	55
3.2.2 France .....	56
3.2.3 Afrique du Sud .....	57
3.3 Génotypage, contrôle qualité et imputation .....	57
3.4 Définition du phénotype d'infection par <i>M. tuberculosis</i> .....	70
3.5 Analyse d'association pangénomique .....	78
3.5.1 Méthodes statistiques .....	78
3.5.2 Analyse d'association pangénomique au Vietnam.....	83
3.5.3 Réplication dans 2 cohortes indépendantes de France et d'Afrique du Sud .....	92
3.5.4 Méta-analyse .....	98
3.6 Analyse <i>in silico</i> et annotation fonctionnelle du locus 10q26.2.....	103
3.7 Réplication d'associations avec l'infection par <i>M. tuberculosis</i> .....	110
3.8 Discussion .....	113
4 Conclusion.....	118
Bibliographie.....	120
Annexes.....	141

# Remerciements

Je tiens à remercier en premier lieu Laurent Abel et Aurélie Cobat pour m'avoir encadré et offert l'opportunité de faire cette thèse.

Je remercie bien évidemment les membres de ce jury pour le temps passé et l'attention portée à ce manuscrit, en espérant qu'ils y trouveront un intérêt. Je remercie donc Jacques Fellay et Etienne Patin d'avoir endossé le rôle chronophage de rapporteur ainsi qu'Anne Bourgarit-Durand et Audrey Sabbagh pour avoir accepté de juger mon travail en tant qu'examinatrices.

Merci à Jean-Laurent Casanova et l'ensemble des membres du Laboratoire de Génétique Humaine des Maladies Infectieuses de m'avoir accueilli en leur sein.

Ce travail n'a pu être effectué qu'à partir de données d'une qualité rare. J'adresse un merci particulier à nos collaborateurs canadiens, Erwin Schurr et Marianna Orlova mais également à nos collaborateurs marocains, vietnamiens, sud-africains et français.

Je souhaiterais remercier les deux premières personnes avec qui j'ai entamé mon cursus de recherche, Philippe Broët et Hervé Perdry, ainsi que toute l'ancienne équipe Génostat de Villejuif.

Je remercie enfin la Fondation pour la Recherche Médicale pour avoir financé ces trois années de travail.

# Liste des figures

Figure 1. Planches extraites de Die Ätiologie der Tuberkulose, Robert Koch.....	13
Figure 2. Top 10 des causes de décès dans le monde en 2016 .....	13
Figure 3. Incidence de la tuberculose en 2018 .....	14
Figure 4. Arbre phylogénétique et répartition géographique du complexe <i>Mycobacterium tuberculosis</i> .....	16
Figure 5. Histoire naturelle de la tuberculose .....	18
Figure 6. Etablissement de l'infection tuberculeuse .....	20
Figure 7. Test de Mantoux.....	23
Figure 8. Tests de production d'interféron-gamma .....	24
Figure 9. Hétérogénéité du phénotype d'infection tuberculeuse.....	28
Figure 10. Prévalence de l'infection tuberculeuse chez des contacts familiaux. ....	28
Figure 11. Vaccins candidats en évaluation en 2019 .....	32
Figure 12. Facteurs de risques associés à la susceptibilité à l'infection tuberculeuse puis au développement d'une maladie active. ....	33
Figure 13. Règle de Styblo.....	34
Figure 14. Nuage de points représentant les valeurs quantitatives de QFT et d'IDR des 631 participants marocains.....	46
Figure 15. Assignation correcte d'un marqueur génotypé .....	58
Figure 16. Comparatif de la qualité d'imputation .....	61
Figure 17. Analyse en composantes principales, Vietnam.....	63
Figure 18. Analyse en composantes principales, France et Afrique du Sud.....	64
Figure 19. Proportions des différentes populations ancestrales, Vietnam .....	66
Figure 20. Proportions des différentes populations ancestrales, France .....	67
Figure 21. Proportions des différentes populations ancestrales, Afrique du Sud.....	67
Figure 22. Structures familiales au Vietnam, en France et en Afrique du Sud.....	69
Figure 23. Organigramme général de l'étude.....	71
Figure 24. Histogramme représentant la distribution de l'IDR stratifiée par les résultats IGRA .....	73
Figure 25. Définition du phénotype dans la cohorte de France .....	75
Figure 26. Définition du phénotype dans la cohorte d'Afrique du Sud .....	77
Figure 27. QQ-plot de l'analyse d'association pangénomique au Vietnam.....	85
Figure 28. Analyse d'association pangénomique pour la résistance à l'infection par <i>Mtb</i> au Vietnam .....	85
Figure 29. Déséquilibre de liaison du locus 10q26.2 au Vietnam.....	86
Figure 30. Manhattan plot de la GWAS entre 185 résistants et 353 infectés avec ajustement sur âge, sexe et 3 PCs au Vietnam. ....	86
Figure 31. Manhattan plot de la GWAS entre 185 résistants et 201 infectés au Vietnam.....	88
Figure 32. Manhattan plot de la GWAS ajustée sur âge, sexe, vaccination BCG, nombre d'individus par foyer, durée de la toux du cas index et 3 PCs entre 167 résistants et 171 infectés au Vietnam.....	89
Figure 33. Manhattan plot de la GWAS entre 185 résistants et 152 PTB au Vietnam. ....	90
Figure 34. Manhattan plot de la GWAS entre 250 résistants et 338 infectés au Vietnam (au seuil IDR de 10 mm). ....	92
Figure 35. Déséquilibre de liaison du locus 10q26.2 en France et Afrique du Sud.....	94

Figure 36. Manhattan plots pour les GWASs de la résistance à l'infection par <i>Mtb</i> en France et en Afrique du Sud.....	94
Figure 37. Distribution géographique des allèles de rs17155120 parmi les populations de 1000G Phase 3.....	95
Figure 38. Déséquilibre de liaison dans une fenêtre de 30kb autour du marqueur rs17155120 .....	96
Figure 39. Méta-analyse de la résistance à l'infection par <i>Mtb</i> .....	101
Figure 40. QQ-plot de la méta-analyse de la résistance à l'infection par <i>Mtb</i> .....	101
Figure 41. Manhattan plots des méta-analyses de la résistance à l'infection par <i>Mtb</i> par effets fixes et par effets aléatoires classiques.....	102
Figure 42. Diagramme d'une GWAS, de l'analyse d'association à l'annotation <i>in silico</i> .....	104
Figure 43. Analyses post-GWAS et leurs difficultés .....	105
Figure 44. Annotation génomique du locus 10q26.2 .....	108
Figure 45. Niveaux d'expression d' <i>ADAM12</i> dans les monocytes.....	108
Figure 46. Méta-analyse de la résistance à l'infection par <i>Mtb</i> dans la région HLA.....	112
Figure 47. Étude d'association du QFT en trait quantitatif.....	117

## Liste des tableaux

Tableau 1. Détermination du résultat d'un test QuantiFERON-TB Gold In-Tube .....	25
Tableau 2. Traitement de l'infection tuberculeuse selon les recommandations de l'Organisation Mondiale de la Santé .....	30
Tableau 3. Caractéristiques de la cohorte marocaine.....	45
Tableau 4. Accord entre les résultats QFT et IDR chez les 631 participants marocains .....	47
Tableau 5. Facteurs de risque associés à un résultat QFT positif.....	48
Tableau 6. Facteurs de risque associés à un résultat IDR+ .....	49
Tableau 7. Facteurs de risque associés à un résultat QFT+/IDR+ .....	50
Tableau 8. Facteurs de risque associés à des résultats QFT et IDR discordants.....	51
Tableau 9. Covariables associées avec le phénotype d'infection au Vietnam .....	72
Tableau 10. Association des covariables avec le phénotype d'infection en France.....	74
Tableau 11. Association des covariables avec le phénotype d'infection en Afrique du Sud...	76
Tableau 12. Description des cohortes .....	78
Tableau 13. Association entre les marqueurs du locus 10q26.2 et la résistance à l'infection par <i>Mtb</i> .....	87
Tableau 14. Marqueurs du locus 10q26.2 associés avec la résistance à l'infection par <i>Mtb</i> (185 résistants et 201 infectés) .....	89
Tableau 15. Marqueurs du locus 10q26.2 associés avec la résistance à l'infection par <i>Mtb</i> (185 résistants et 152 PTB) .....	91
Tableau 16. Analyse de puissance statistique .....	91
Tableau 17. Marqueurs associés avec la résistance à l'infection par <i>Mtb</i> avec $P < 5.0 \times 10^{-8}$ dans la méta-analyse (333 résistants et 616 infectés).....	102
Tableau 18. Marqueurs associés avec la résistance à l'infection par <i>Mtb</i> en 11p14-15.....	111
Tableau 19. Réplication de marqueurs précédemment associés avec une IDR négative dans les 3 cohortes Vietnam, France et Afrique du Sud. ....	111

## Liste des publications

1. **J. Quistrebert**, M. Orlova, G. Kerner, L. T. Ton, N. T. Luong, N. T. Danh, Q. B. Vincent, F. Jabot-Hanin, Y. Seeleuthner, J. Bustamante, S. Boisson-Dupuis, N. T. Huong, N. N. Ba, J.-L. Casanova, C. Delacourt, E. G. Hoal, A. Alcaïs, V. H. Thai, L. T. Thành, L. Abel, E. Schurr, A. Cobat, Genome-wide association study of resistance to Mycobacterium tuberculosis infection identifies a locus at 10q26.2 in three distinct populations. *PLOS Genetics*. **17**, e1009392 (2021). doi: 10.1371/journal.pgen.1009392
2. A. Sabri\*, **J. Quistrebert\***, H. N. Amrani\*, A. Abid, A. Zegmout, I. A. Ghorfi, H. Souhi, A. Boucaid, A. Benali, R. Abilkassem, M. Kmari, A. Hassani, B. Lahcen, S. Siah, E. Schurr, S. Boisson-Dupuis, J.-L. Casanova, A. Lahlou, A. Laatiris, L. Louzi, A. Ouarssani, A. Bourazza, A. Aouragh, B. Mustapha, N. Messaoudi, A. Agader, A. Cobat, L. Abel, J. E. Baghdadi, Prevalence and risk factors for latent tuberculosis infection among healthcare workers in Morocco. *PLOS ONE*. **14**, e0221081 (2019). doi: 10.1371/journal.pone.0221081

\*co-premier auteur

# 1 Introduction

## 1.1 La tuberculose

### 1.1.1 Contexte

La tuberculose est une maladie infectieuse chronique causée principalement par *Mycobacterium tuberculosis* (*Mtb*), dont la transmission interhumaine se fait par voie aérienne à partir d'un patient contagieux. C'est en 1882 que Robert Koch mit en évidence le bacille [1] (Figure 1). En 2018, la tuberculose reste un problème majeur de santé publique : c'est l'une des 10 premières causes de mortalité dans le monde et la première par un unique agent infectieux. (Figure 2). Ainsi en 2018, on estime qu'environ 10 millions de personnes ont été infectées par *Mtb* et environ 1.5 millions sont décédées de la maladie, dont 250 000 chez des personnes positives au virus de l'immunodéficience humaine (VIH). L'Asie du Sud-Est et l'Afrique comptent environ deux tiers des cas de tuberculose (44 et 24% respectivement) et l'Inde compte à elle seule plus d'un quart du total mondial (27%) [2] (Figure 3).

La tuberculose demeure un tel fléau (« la peste blanche » comme nommée au cours du XVIIIème siècle [3]) que dès 1993 l'Organisation Mondiale de la Santé (OMS) déclarait une urgence mondiale [4]. Vingt ans plus tard, l'ambitieuse « End TB Strategy » est adoptée dans le but d'éradiquer la pandémie pour 2035 [5]. Il faut dire que les tendances dans l'épidémiologie et la gestion de la tuberculose révèlent de fortes disparités [6]. Si l'incidence décline globalement d'environ 2% par an actuellement et si la baisse d'un facteur 100 du nombre de décès dus à la maladie a été spectaculaire en Europe et aux Etats-Unis au cours du XXème siècle, et ce même avant l'apparition d'antibiothérapies dans les années 1950, les progrès sont plus mitigés ailleurs. En Afrique, l'épidémie de VIH a catalysé celle de tuberculose ; en Inde, ce sont les difficultés au niveau du diagnostic ou d'accès aux traitements qui sont problématiques [7]. D'une manière générale, là où l'attention est portée et les moyens mis en œuvre avec une vraie volonté politique de combattre cette maladie, des progrès sont observés à moyen/long terme à une échelle locale comme à New York [8] ou nationale comme au Pérou ou en Chine [9],[10].

Entre le moment où cette thèse a débuté, novembre 2017, et où elle a été rédigée, été/automne 2020, la recherche sur la tuberculose a aussi connu de beaux succès : un nouveau candidat vaccin prometteur contre le développement de la tuberculose maladie [11], le vaccin centenaire BCG (bacille Calmette et Guérin) revisité en revaccination ou administration par voie intraveineuse [12],[13], une antibiothérapie d'un mois unique pour traiter l'infection tuberculeuse au lieu de 3 mois minimum [14] ou une nouvelle antibiothérapie de 6 mois, par voie orale, incluant bédaquiline, linézolide et prétomanide, efficace contre la tuberculose ultrarésistante [15]. Malheureusement, l'apparition récente de souches résistantes à la bédaquiline et la clofazimine en Afrique est préoccupante [16]. Pire, la pandémie de Covid-19 est en train de réduire à néant les efforts de ces dernières années et risque d'avoir un impact très négatif aussi bien en terme de santé publique (mortalité, accès aux soins) que dans la recherche (concentration des ressources humaines et financières) [17],[18].

Le contrôle de la pandémie de tuberculose est donc entravé par : un manque de ressources et d'investissements publics-privés ; des outils de prévention (vaccin BCG), des moyens diagnostics (examen microscopique des crachats) ou des traitements (antibiotiques) datés et moins efficaces ; des facteurs de risque mal maîtrisés (coinfections, diabète, pollution, malnutrition, pauvreté etc.) ; et enfin une physiopathologie de la maladie loin d'être élucidée.



Figure 1. Planches extraites de Die Ätiologie der Tuberkulose, Robert Koch. A gauche, dessins de cultures de bacilles tuberculeux dans un tube à essai. A droite, de nombreux bacilles, en bleu, entourent une cellule géante de poumon.

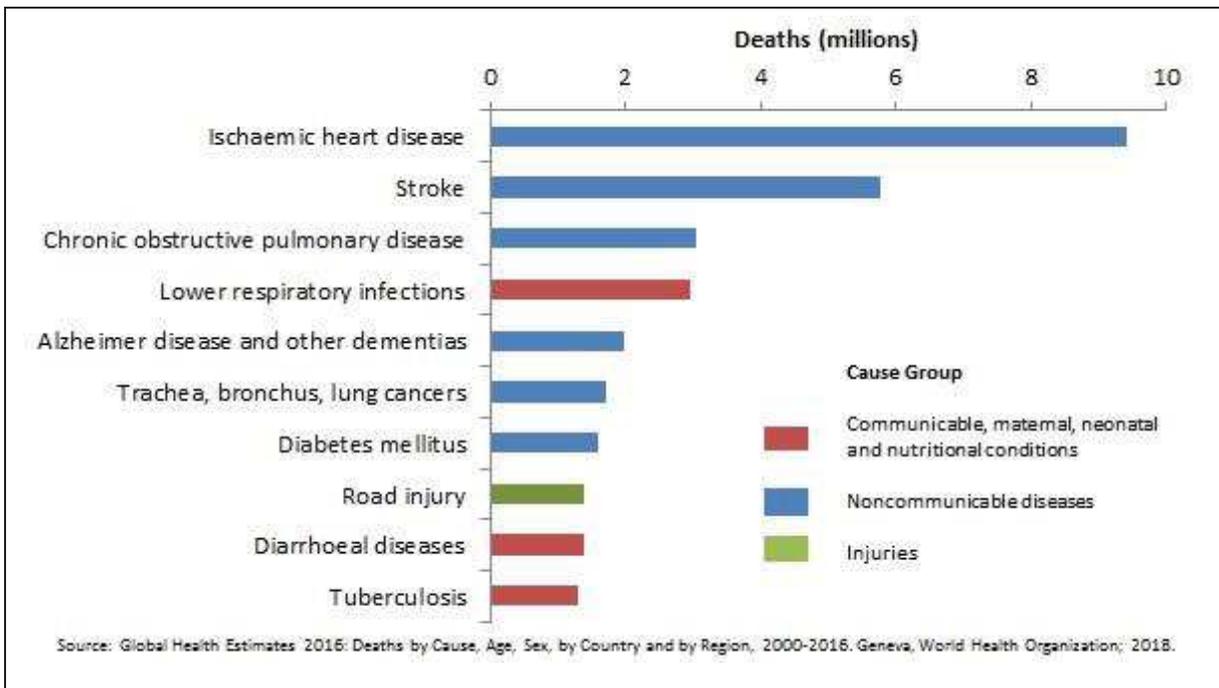


Figure 2. Top 10 des causes de décès dans le monde en 2016.

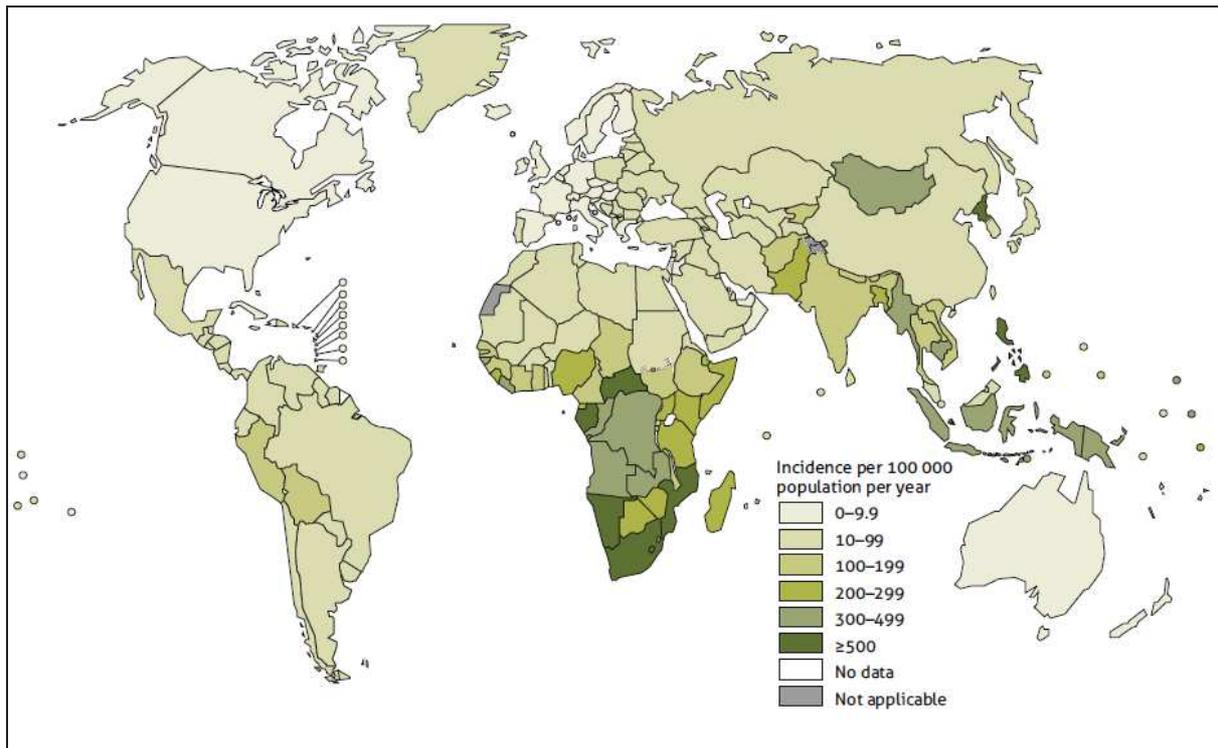


Figure 3. Incidence de la tuberculose en 2018 [2]. Les pays à haut revenus (Europe de l’Ouest, Etats-Unis, Canada, Australie etc.) présentent les plus faibles taux d’incidence, contrairement aux pays émergents ou à moyen ou faible revenus.

### 1.1.2 Microbiologie

*Mtb* est un bacille aérobie qui appartient à l’embranchement *Actinobacteria* et particulièrement au genre *Mycobacterium*, l’unique membre de la famille *Mycobacteriaceae* comprenant 187 autres espèces connues à ce jour. *Mtb* est un pathogène obligatoire qui a évolué pendant des milliers d’années pour ne persister exclusivement que chez l’humain [19],[20]. Le bacille a donné son nom au complexe *Mycobacterium tuberculosis* (CMTB) constitué de différentes espèces de mycobactéries partageant 99.9% d’identité de séquence de leur ADN. Le CMTB comprend notamment 7 lignées qui sont adaptées à l’humain et sont responsables de l’immense majorité des tuberculoses dans le monde (L1 à L4 et L7 représentant *Mtb stricto sensu* et L5-L6 représentant *M. africanum*). Tandis que certaines lignées dites « modernes » (L2 à L4) se sont largement diffusées, les autres dites « anciennes » sont très restreintes géographiquement. Cette classification évolue rapidement puisque 2 nouvelles lignées retrouvées en Afrique viennent d’être récemment découvertes [21],[22] (Figure 4).

L'Afrique est le continent où toutes les lignées du CMTB peuvent être observées et leur diversité s'estompe à mesure que l'on s'en éloigne, suggérant qu'un ancêtre commun au CMTB y a émergé. Si l'origine géographique en Afrique de l'Est laisse peu de doutes, l'origine temporelle du bacille est plus controversée. Plusieurs études, s'appuyant sur le séquençage du génome de diverses souches modernes, proposaient une évolution du CMTB, depuis 70 000 ans avant le présent, époque Pléistocène, qui aurait ainsi suivi les migrations d'*Homo sapiens* hors d'Afrique [23],[24]. Ce scénario est remis en question depuis quelques années par d'autres études s'appuyant sur le séquençage d'ADN ancien et datant l'origine d'un ancêtre commun au CMTB aux alentours de 4000 ans avant le présent, soit une émergence plutôt au cours du Néolithique [25],[26].

En comparaison avec d'autres pathogènes, *Mtb* n'est pas si infectieux : en moyenne, un individu bacillifère infecte 5 à 15 personnes par an, parmi lesquelles une minorité développera la maladie et se retrouvera contagieuse [2]. En effet, environ 10% des individus infectés souffrira d'une tuberculose maladie à n'importe quel moment de sa vie. Ainsi, si l'on considère qu'un quart de la population mondiale est infectée [27], *Mtb* est donc plus efficace pour causer une infection asymptomatique qu'une maladie active. Cette observation se confirme au niveau microbiologique grâce aux techniques récentes de séquençage de génome. Le génome de *Mtb* est relativement petit (environ 4.4 millions de bases, 4 000 gènes, un seul chromosome) et la possibilité de son séquençage a fait grandement progresser la recherche : surveillance épidémiologique, résistance aux antibiotiques, phylogénétique etc. [28]. Comas *et al.* ont montré que (i) les gènes essentiels de *Mtb* (c'est-à-dire ceux ayant des fonctions fondamentales comme réplication de l'ADN, synthèse d'ARN ou d'acides aminés) étaient plus évolutivement conservés que les gènes non essentiels, et (ii) plus surprenant, les épitopes de *Mtb* (devant se lier aux molécules du complexe majeur d'histocompatibilité (MHC) pour être présentés aux lymphocytes T de l'hôte) étaient hyper-conservés et plus de 95% d'entre eux ne présentaient aucun changement d'acide aminé [29]. Ce peu de variation implique que *Mtb* a évolué non pas pour échapper à l'immunité de l'hôte mais pour se faire reconnaître, déclencher une réponse immune, former un granulome dans le tissu pulmonaire et perpétuer un cycle de transmission. En dehors des épitopes, les lignées du CMTB affichent en fait une grande diversité génétique qui a des conséquences sur leur virulence ou leur immunogénicité [23]. Cette diversité explique en partie la distribution géographique du CMTB (Figure 4).

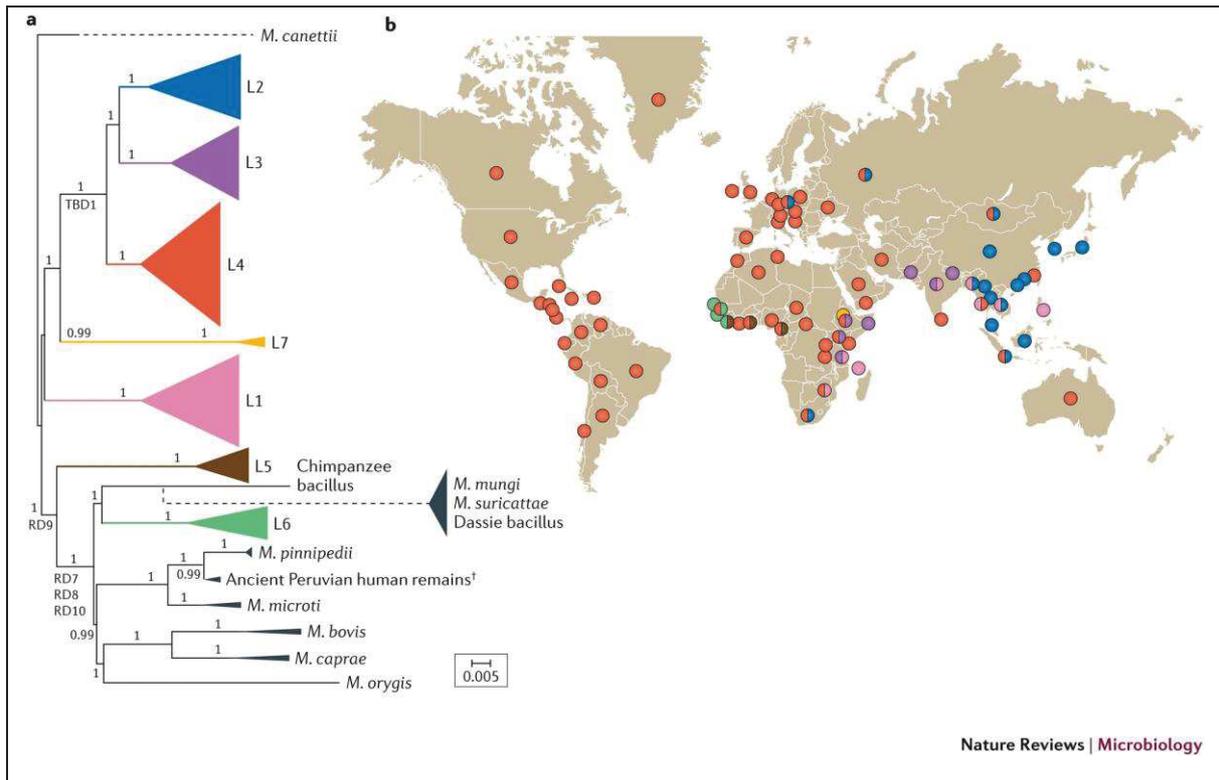


Figure 4. Arbre phylogénétique et répartition géographique du complexe *Mycobacterium tuberculosis* [20]. Le complexe comprend 7 lignées adaptées à l'humain, en couleur : L1, L2, L3, L4 et L7 appartiennent à *Mtb* et L5-L6 à *M. africanum*. Deux récentes lignées, L8 et L9, découvertes très récemment ne sont pas représentées. Les lignées *Mtb* sont classiquement catégorisées en « modernes » (L2, L3 et L4, partageant la délétion TBD1 de leur génome) et « anciennes » (L1, L5, L6 et L7). En gris sont représentées les lignées adaptées aux animaux sauvages ou domestiques (caractérisées par la délétion RD9) mais capables d'infecter parfois l'humain.

En théorie, la virulence est positivement corrélée à la transmission et l'accès à de nombreux hôtes favorise une plus grande virulence associée à un plus court délai de l'infection à la maladie. Les lignées « anciennes », apparues aux temps des chasseurs-cueilleurs, auraient donc été d'abord favorisées : densités de population faibles et peu d'hôtes disponibles donc faible virulence avec une longue période de latence avant une réactivation de l'infection et le développement mortel de la maladie. L'expansion démographique en Europe au XIX<sup>ème</sup> siècle puis en Inde ou en Chine, a, au contraire, privilégié des lignées plus virulentes et moins gênées par la disparition de leur hôte/réservoir. Ces hypothèses pourraient expliquer pourquoi une lignée ancienne comme *M. africanum* est resté cantonnée en Afrique de l'Ouest et n'a pas suivi les mouvements de la traite d'esclaves jusqu'en Amérique. Ces observations d'adaptation des lignées « modernes » à leur nouvel environnement ont été confirmées depuis par de nombreuses études, aussi bien expérimentales [30] qu'épidémiologiques [31],[32].

Si ce paragraphe s'est focalisé sur le bacille, cette thèse se concentre sur l'hôte. Mais la biologie et l'épidémiologie de la tuberculose ont donc été façonnées par une longue coévolution *Mtb* – humain et c'est probablement par l'intégration des 2 composantes dans les futures études que de nouvelles réponses seront apportées.

### 1.1.3 Histoire naturelle de la tuberculose

L'histoire naturelle de la tuberculose se caractérise par sa transmission aérienne et l'extrême variabilité de ses manifestations cliniques [6]. La maladie se présente comme un spectre continu allant de la simple infection asymptomatique à une maladie mortelle (Figure 5). Pour plus de praticité en clinique ou santé publique, les patients sont souvent catégorisés comme ayant soit (i) une infection tuberculeuse latente (ITL), asymptomatique et non-transmissible, soit (ii) une tuberculose maladie, transmissible et symptomatique, classiquement pulmonaire mais pouvant atteindre d'autres organes (ganglions, os [mal de Pott], méninges entre autres). Ce n'est qu'à ce stade que la bactérie est directement détectable et que le diagnostic peut se faire à partir d'un échantillon provenant des voies respiratoires. De nombreux pays emploient encore une vieille méthode, l'examen des expectorations au microscope à la recherche de bacilles acido-alcoolo-résistants révélés par la technique de coloration Ziehl-Neelsen. La mise en culture en milieu solide de Lowenstein-Jensen pendant 3 à 4 semaines ou en milieu liquide n'est pas toujours possible. Depuis 2010, un test moléculaire rapide, Xpert MTB/RIF, a été développé. Basé sur l'amplification d'acide nucléique, il est suffisamment simple et robuste pour pouvoir être utilisé en dehors du cadre classique des laboratoires. Il permet de détecter *Mtb* ainsi que les mutations conférant la résistance à la rifampicine.

L'exposition à *Mtb* peut aboutir à l'élimination du pathogène par une réponse immunitaire innée immédiate ou bien adaptative médiée par les lymphocytes T et B. Les individus qui sont restés non-infectés ou ont éliminé la bactérie présentent une intradermoréaction (IDR ou TST pour *tuberculin skin test*) ou un test de production d'interféron-gamma (IFN- $\gamma$ ) (IGRA pour *interferon-gamma release assay*) négatifs. Certains individus peuvent avoir éliminé la bactérie tout en gardant une mémoire acquise par les lymphocytes T, cela se traduit par une IDR ou un IGRA positif. Si la bactérie n'est pas éliminée et persiste, l'ITL se met en place, se traduisant par une positivité à l'IDR ou l'IGRA. Après établissement de l'infection, certains individus peuvent rapidement progresser vers une maladie active ; on parle de forme primaire, classiquement retrouvée chez l'enfant dans l'année suivant l'infection. Dans la plupart des cas,

l'infection est contenue et l'individu maintient une relation dynamique avec *Mtb* via son système immunitaire et la régulation de nutriments.

De nouveaux stades de la maladie ont été récemment définis pour caractériser ce processus continu entre infection latente et maladie active (tuberculose « naissante » puis sub-clinique) [33]. Finalement, après plusieurs années, on observe une réactivation chez une minorité d'individus ; c'est la tuberculose classique de l'adulte, souvent pulmonaire, caractérisée par sa symptomatologie (fièvre, perte poids, sueurs nocturnes, toux etc.), ses anomalies radiologiques (cavités, adénopathies etc.) et la mise en évidence du bacille microbiologiquement. Si l'on a longtemps pensé que la maladie pouvait se réactiver après des décennies de latence, il est en fait très probable qu'une proportion (large ?) soit en fait due à une réinfection récente [34],[35].

Cette thèse est avant tout centrée sur l'infection tuberculeuse, première étape clé de la maladie et l'on n'évoquera pas (ou très peu) ce qui se déroule après cette phase (Figure 5).

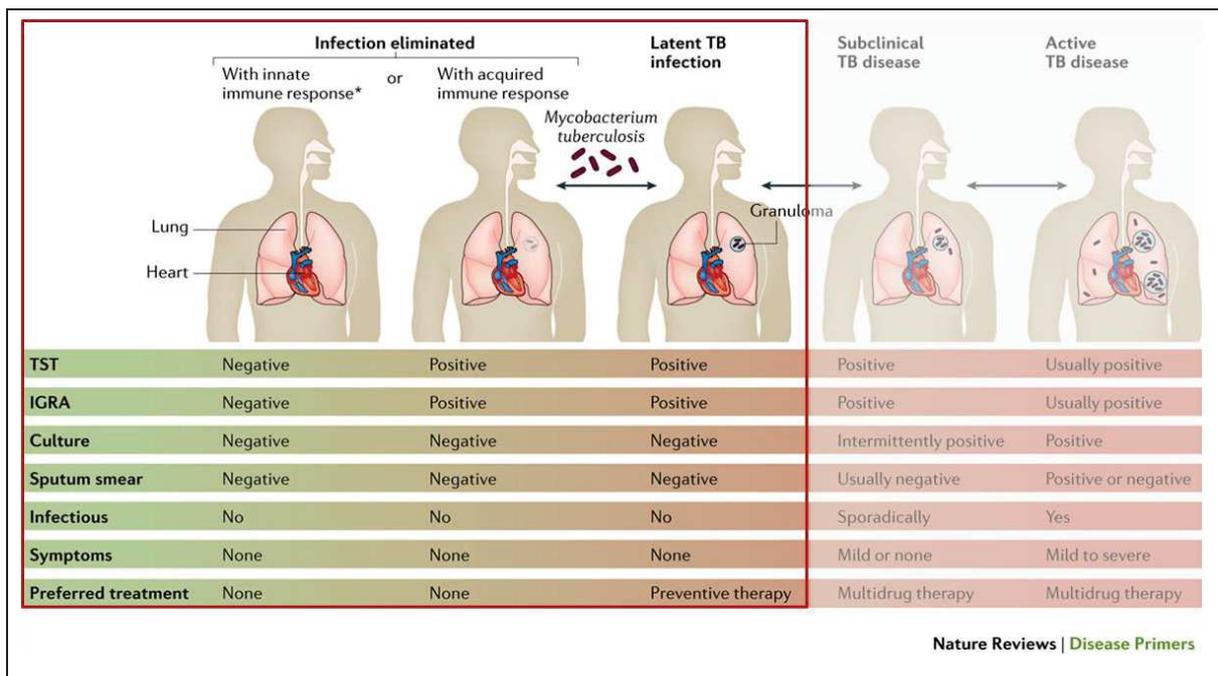


Figure 5. Histoire naturelle de la tuberculose [6]. L'exposition d'un individu au bacille *Mtb* conduit soit à son élimination soit à l'établissement d'une infection latente.

## 1.2 L'infection par *M. tuberculosis*

### 1.2.1 Physiopathologie

Après inhalation, les premières cellules rencontrées par *Mtb* sont les macrophages alvéolaires. Les macrophages sont un réservoir majeur pour l'infection et sont au premier rang pour contrôler et éliminer le bacille. La reconnaissance de *Mtb* par de nombreux récepteurs à la surface ou à l'intérieur des macrophages aboutit à la libération d'une multitude de médiateurs de l'immunité innée [36]. Ces récepteurs, appelés PRR pour *pattern recognition receptors*, incluent les *Toll-like* récepteurs (TLR), les récepteurs *scavenger*, les récepteurs du complément (CR), les récepteurs des protéines du surfactant, les lectines de type C (*mannose receptors*, [MR], *dendritic Cell-Specific Intercellular adhesion molecule-3-Grabbing Non-Integrin*, [DC-SIGN]), et les récepteurs cytosoliques NOD (*nucleotide-binding oligomerization domain-containing proteins*). Les études chez l'humain ou la souris ont montré que les principaux médiateurs de l'immunité innée étaient le facteur de nécrose tumorale (TNF pour *tumor necrosis factor*), l'interleukine-1 (IL-1), IL-12, IL-18 et, qu'en synergie avec l'IFN- $\gamma$  produit par les lymphocytes cytotoxiques naturels (NK pour *natural killer*), activaient les premières défenses antibactériennes. Les neutrophiles, les cellules lymphoïdes innées, les lymphocytes MAIT pour *Mucosal-Associated Invariant T cells* et les lymphocytes T  $\gamma\delta$  jouent aussi un rôle dans la défense précoce contre l'infection.

Après l'établissement d'une infection primaire, les cellules dendritiques et les monocytes transportent *Mtb* vers les ganglions lymphatiques pulmonaires pour initier l'activation des lymphocytes T naïfs et la mise en action de l'immunité adaptative (Figure 6). Les lymphocytes T et B sont recrutés et se regroupent dans le parenchyme pulmonaire pour former un granulome. Les lymphocytes T auxiliaires CD4<sup>+</sup> (Th1) produisant de l'IFN- $\gamma$  sont primordiaux dans le contrôle de l'infection. Le rôle central de l'IFN- $\gamma$  et de l'axe Th1 dans la protection contre *Mtb* a notamment été mis en évidence grâce à la caractérisation des immunodéficiences dans la cascade de signalisation de l'IFN- $\gamma$  [37]. D'autres sous-populations de lymphocytes CD4<sup>+</sup>, notamment celles produisant de l'IL-17 (lymphocytes Th17) et IL-21 participent à la résistance de l'hôte. Les lymphocytes T cytotoxiques CD8<sup>+</sup> vont quant à eux lyser les macrophages infectés et produire d'autres molécules cytotoxiques s'attaquant à *Mtb*. Le bacille provoque

donc à la fois des lymphocytes T mémoires et effecteurs. Enfin, les lymphocytes B produisant des anticorps contribuent également à l'immunité [38].

Ainsi, L'axe Th1-IFN- $\gamma$ , bien que nécessaire, est insuffisant pour une protection totale. Des individus, très exposés à *Mtb* mais résistants à l'infection et continuellement négatifs aux tests diagnostiques, présentent une signature immunologique particulière, caractérisée par une réponse des lymphocytes T indépendante de l'IFN- $\gamma$  et une forte activité humorale [39]. Ce phénotype de résistance sera l'objet de la deuxième partie de cette thèse.

Malgré toutes ces armes déployées par le corps humain, au vu de la prévalence globale de l'ITL, l'évasion bactérienne et l'établissement d'une infection persistante sans élimination complète apparaissent comme la caractéristique essentielle de l'infection par *Mtb*. La question de la durée de cette infection reste à ce jour non élucidée et cruciale dans le combat contre *Mtb* [34]. On peut observer qu'une majorité d'individus positifs pour une ITL et traités préventivement par antibiotiques continue d'être testée positive ; ou bien des individus positifs en situation d'immunodépression ne développent pas de maladie. Cela prouve que l'infection ne s'établit pas à vie et donc la prévalence globale d'ITL est largement surévaluée [34]. Les tests diagnostiques décrits plus en précision ci-après ne détectent que l'empreinte immunologique laissée par *Mtb*. Cette question est centrale et oriente la recherche dans 2 voies opposées si l'on considère que l'infection s'installe pour la vie (tolérance) ou peut-être guérie (élimination).

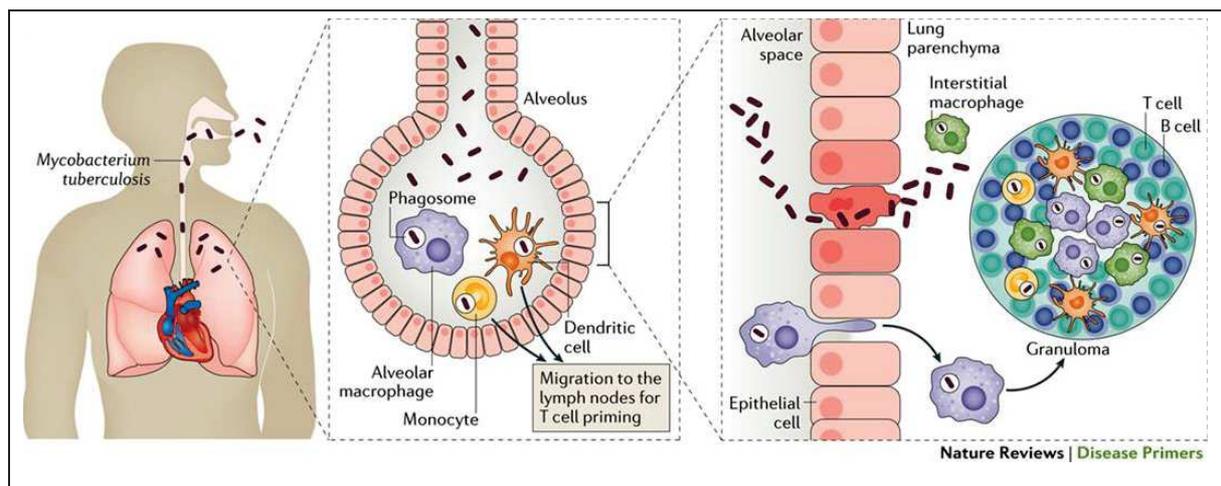


Figure 6. Etablissement de l'infection tuberculeuse [6]. *Mtb* entre dans les poumons après inhalation et rencontre les macrophages alvéolaires. Si ceux-ci ne parviennent pas à éliminer la bactérie, elle s'évade du tissu pulmonaire. Par la suite, les monocytes et les cellules dendritiques transportent *Mtb* jusqu'aux ganglions lymphatiques adjacents pour l'amorçage des lymphocytes T. Cet évènement conduit au recrutement de cellules immunitaires dans le parenchyme pulmonaire et à la formation d'un granulome.

### 1.2.2 Diagnostic de l'infection

Le diagnostic d'infection est déduit, indirectement, d'une mesure quantitative de l'immunité anti-mycobactérienne. À ce jour, il n'existe pas de méthode diagnostique de référence et seulement 2 types de tests immunologiques sont utilisés. Ils servent de marqueurs d'infection par *Mtb* mais ne permettent pas de déterminer si la réponse immunitaire a conduit à une élimination du pathogène ou non.

Le plus ancien de ces tests immunologiques est l'intradermoréaction à la tuberculine (IDR), ou test de Mantoux, qui reste l'examen diagnostique 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. Il consiste en l'injection intradermique de tuberculine (un dérivé protéinique purifié extrait de culture de bacilles tuberculeux [PPD]) à la face antérieure de l'avant-bras. Cette injection va entraîner 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. De fait, sa lecture se fait 72 heures plus tard en mesurant le diamètre d'induration (et non de l'érythème provoqué) de la réaction cutanée en millimètres (mm) (Figure 7). Après l'injection intradermique de tuberculine, les cellules de Langerhans, présentatrices d'antigènes, phagocytent et présentent les peptides microbiens aux lymphocytes T mémoires, induisant la production de cytokines ainsi que l'afflux de monocytes, de macrophages et de lymphocytes au point d'injection. L'induration se caractérise au niveau microscopique par un œdème et une infiltration dense de cellules mononuclées, en particulier autour des petits vaisseaux sanguins [40].

Le plus souvent, le résultat d'une IDR est interprété de façon binaire, un test positif témoignant d'une infection par *Mtb*. Il existe 3 seuils différents pour établir le diagnostic d'infection tuberculeuse en fonction des individus [41]:

- 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 ou ceux exposés récemment à un cas contact).
- 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 (jeunes enfants, personnes incarcérées ou originaires de régions très endémiques).

- chez les individus ayant un faible risque d'infection par *Mtb*, un seuil de 15 mm est recommandé.

Cependant, l'IDR n'est pas spécifique à 100%. En effet, des réactions croisées existent avec le BCG et la plupart des autres mycobactéries environnementales, qui expriment des antigènes communs avec ceux contenus dans la tuberculine [42]. Il a été montré que l'effet des mycobactéries environnementales était faible dans les pays ayant une incidence de tuberculose intermédiaire ou élevée. En revanche, plus de 90% des individus vaccinés par le BCG présentent une induration  $\geq 10$  mm dans les 8 à 12 semaines suivant la vaccination [43]. Mais l'impact du BCG sur l'IDR 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 [42]. Dans cette 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 10 mm et 15 mm, indépendamment du délai entre la vaccination et la réalisation du test de Mantoux [42]. 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 10 mm attribuable au BCG. Si la vaccination a eu lieu plus tard, l'impact sur l'IDR est plus fort. Ainsi, 21% des individus ont toujours une réaction supérieure à 10 mm attribuable au BCG plus de 10 ans après vaccination. Enfin, une autre limite de cette technique vient du fait que tout le monde ne présente pas de réaction immunitaire contre la tuberculine. Une anergie au test peut être observée dans certaines conditions (immunodépression, malnutrition ou en phase aiguë de maladie infectieuse).

En pratique clinique, le diagnostic d'infection tuberculeuse à partir de l'IDR doit donc tenir compte de la spécificité du test, 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 (probabilité qu'un test positif représente une vraie infection) augmente avec la prévalence de l'infection dans la population. Dans les pays où la prévalence de la tuberculose est faible, l'indication de l'IDR est limitée aux populations à risque et l'utilisation des seuils décrits plus haut tenant compte des caractéristiques du test et du risque individuel d'infection est recommandée.

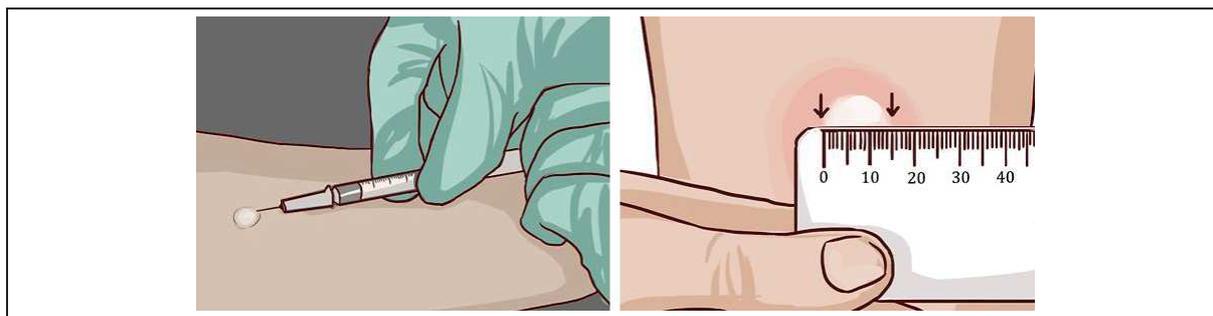


Figure 7. Test de Mantoux. Injection de tuberculine puis lecture à 72h de l'induration.

Plus récemment, des tests *ex vivo* ont été développés: les tests de production d'IFN- $\gamma$  (IGRAs pour *Interferon-gamma release assays*) [44]. Leur principe est simple : les lymphocytes T d'un individu ayant déjà été exposés aux antigènes *Mtb* produisent de l'IFN- $\gamma$  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 *Mtb*, et peuvent permettre d'être plus spécifique que l'IDR en choisissant bien les antigènes utilisés. Cette production d'IFN- $\gamma$  peut être mesurée par les techniques ELISA (pour *enzyme-linked immunosorbent assay*) ou ELISpot (pour *enzyme-linked immunospot*, dérivée d'ELISA) (Figure 8). La dernière génération de ces tests utilise les antigènes ESAT-6 (*early secretory antigenic target 6kDa*), CFP10 (*culture filtrate protein 10 kDa*) et/ou TB7.7, qui sont spécifiques de *Mtb* et ne sont exprimés ni par le BCG, ni par la plupart des mycobactéries environnementales. Ces tests présentent donc une meilleure spécificité dans les populations vaccinées par le BCG. Deux versions commerciales sont disponibles, T-Spot.TB ELISpot (Oxford Immunotec, UK) qui utilise ESAT-6 et CFP10 et le nouveau QuantiFERON-TB Gold In-Tube (QFT) (Qiagen, Hilden, Germany) qui utilise ESAT-6, CFP10 et TB7.7. Dans ces tests commerciaux, l'incubation dans un tube avec les antigènes mycobactériens est de 16 à 24h, et ils sont réalisés avec 2 autres tubes : un contrôle négatif (incubation sans stimulus) et un contrôle positif (incubation avec un mitogène) qui aident à la détermination du résultat du test (Tableau 1). Le contrôle négatif indique la présence résiduelle d'une production d'IFN- $\gamma$  puisqu'il n'y a pas de stimulation antigénique mycobactérienne dans ce tube. Une forte valeur est souvent le signe d'une infection en cours reflétant une réponse immunitaire médiée par les lymphocytes T et peut donc donner un résultat faux-positif. Le contrôle positif permet quant à lui de vérifier la compétence des cellules immunitaires de l'individu testé. Par exemple, un individu immunodéprimé (sous thérapie, séropositif au VIH ou bien simple lymphopénie) ne répondra pas à la stimulation par mitogène, indiquant une incapacité à réagir aux antigènes mycobactériens. Une petite valeur donne au test un résultat indéterminé quand l'échantillon

sanguin a aussi une réponse négative pour le tube contenant les antigènes *Mtb*. Le plus souvent, le tube mitogène permet de détecter une erreur technique avec la collection de l'échantillon sanguin comme un temps de transport trop long ou une mauvaise incubation. Ces tests commerciaux peuvent souffrir d'une mauvaise reproductibilité (un même individu testé plusieurs fois successivement peut présenter des résultats différents si les conditions de prélèvement diffèrent légèrement ou la production d'IFN- $\gamma$  se situe dans une zone grise autour du seuil de positivité) et d'une incertitude quant au seuil optimal de positivité [45].

Comme pour l'IDR, les résultats de l'IGRA doivent donc être interprétés dans le contexte clinique dans lequel s'effectue le test. D'après les dernières recommandations de l'OMS [46], dans les pays à revenu élevé et où l'incidence est  $< 100$  pour 100 000, IDR ou IGRA peuvent être utilisés indifféremment. Dans les pays à faible revenu ou les autres pays à revenu intermédiaire, l'IDR, simple et peu coûteuse, doit être privilégiée.

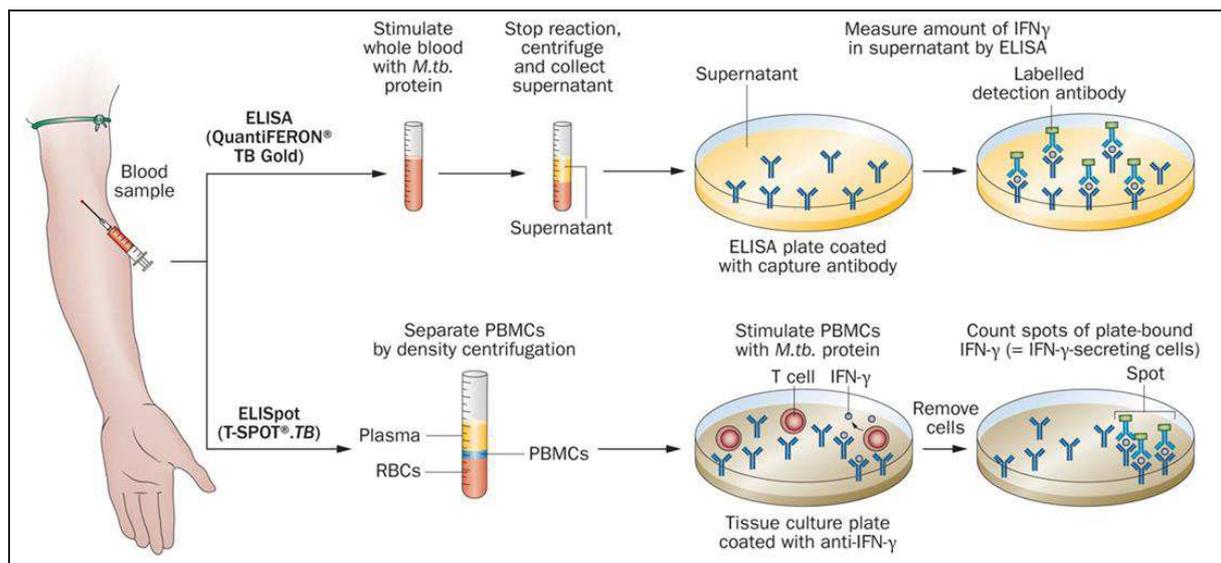


Figure 8. Tests de production d'interféron-gamma (IGRA) [47]. Le QuantiFERON-TB Gold In-Tube mesure la production d'IFN- $\gamma$  par méthode immuno-enzymatique ELISA. Le T-Spot.TB ELISpot compte le nombre de cellules sécrétant de l'IFN- $\gamma$ .

Tableau 1. Détermination du résultat d'un test QuantiFERON-TB Gold In-Tube (QFT) (Qiagen, Hilden, Germany). En fonction des valeurs de production d'interféron-gamma (*TB Antigen*), du contrôle positif (*Mitogen*) et du contrôle négatif (*Nil*), le résultat peut être positif, négatif ou indéterminé.

Nil (IU/ml)	TB Antigen minus Nil (IU/ml)	Mitogen minus Nil (IU/ml)*	QFT result	Report/Interpretation
≤8.0	< 0.35	≥ 0.5	Negative	<i>M. tuberculosis</i> infection NOT likely
	≥ 0.35 and < 25% of Nil value	≥ 0.5	Negative	<i>M. tuberculosis</i> infection NOT likely
	≥ 0.35 and ≥ 25% of Nil value	Any	Positive†	<i>M. tuberculosis</i> infection likely
	< 0.35	< 0.5	Indeterminate‡	Results are indeterminate for TB-Antigen responsiveness
	≥ 0.35 and < 25% of Nil value	< 0.5	Indeterminate‡	Results are indeterminate for TB-Antigen responsiveness
> 8.0§	Any	Any	Indeterminate‡	Results are indeterminate for TB-Antigen responsiveness

### 1.2.3 Hétérogénéité du phénotype d'infection

On peut observer qu'une large proportion de contacts proches d'un patient tuberculeux contagieux convertit une IDR (c'est-à-dire de négative à positive) après exposition alors que seulement 1 ou 2% développent une tuberculose maladie. Cela suggère que les non-réactifs à l'IDR représentent des individus qui : (i) n'ont pas (jamais) été infectés, ou (ii) ont été infectés puis ont rapidement éliminé l'infection [48].

Dans le premier cas, on pourrait supposer qu'il existe une barrière physique (par exemple mucoale) ayant empêchée *Mtb* d'infiltrer les poumons ou bien que l'immunité innée a immédiatement neutralisé la bactérie au site d'infection sans stimulation de l'immunité acquise. Ces individus, naturellement résistants ou « *innate resisters* », présentent des tests IDR et IGRA négatifs, de façon persistante, et ce malgré une exposition intense et documentée à *Mtb* [49].

Dans le deuxième cas, les individus « *adaptive resisters* » ont déclenché une réponse adaptative (impliquant peut-être les lymphocytes B et les lymphocytes T non-conventionnels) et ont éliminé très tôt l'infection (laissant dans l'impossibilité les tests de devenir positifs) [50] ; une autre explication étant que ces individus, effectivement infectés, ont contenu l'infection, de façon très localisée et indétectable (absence de bactéries viables dans le granulome) et ne

progresseront jamais vers une maladie active. Ces individus également restent négatifs aux tests.

De récentes études montrent que les cas contacts très exposés et négatifs aux IDR et IGRA présentent des anticorps spécifiques à *Mtb* [39],[51]. Ce phénomène avait également pu être observé chez du personnel soignant exposé [52]. Cela signifie que l'exposition a bien été réelle et une infection probablement éliminée. La résistance à *Mtb* est très sûrement relative et a un seuil au dessus duquel l'infection est inévitable à partir d'un niveau d'exposition suffisante. L'ITL n'est donc pas nécessairement définie par la présence effective de *Mtb* et se caractérise par une très grande hétérogénéité (Figure 9).

Sur un aspect microbiologique d'abord, l'ITL a longtemps été considérée comme paucibacillaire, *Mtb* entrant dans un état « dormant », car on ne peut pas retrouver de bacille à partir des crachats de la personne infectée. Cependant, il est probable que l'ITL englobe différents types de populations bactériennes allant de celles métaboliquement quiescentes à d'autres plus actives dans leur réplication [53]. Il a aussi été montré par *whole-genome sequencing* le développement de mutations chez *Mtb* isolé de macaques infectés [54].

L'anatomopathologie classique du début du XX<sup>ème</sup> siècle et des séries d'autopsies chez des personnes décédées de causes non liées à la tuberculose ont montré l'existence de bacille viable dans des biopsies de poumon en apparence sain ou de tissu lymphatique [53]. Dans une célèbre expérience de 1890, Loomis démontrait que des ganglions lymphatiques contenaient des bacilles après que des lapins inoculés par des extraits de ces tissus aient développé une tuberculose [55]. L'avènement de nouvelles techniques (PCR in situ de biopsies pour détecter de l'ADN bactérien [56], *barcoding* de *Mtb* inoculé chez des macaques pour suivre la trajectoire de l'infection [57], approche computationnelle pour tracer la dynamique temporelle et spatiale des granulomes [58] etc.) a permis de montrer cette variété des populations bactériennes, des lésions rencontrées ainsi que de leur localisation (dans et en dehors des poumons).

Ensuite, de nombreuses études épidémiologiques ont aussi montré l'existence d'un panel varié de phénotypes. Par exemple, l'épidémie de VIH durant les années 1990 dans les mines d'or d'Afrique du Sud a entraîné une rupture du contrôle de la tuberculose, entraînant des pics d'incidence excédant 4000 pour 100 000 habitants par an (4%). Néanmoins, au beau milieu de cette explosion, 13% des mineurs VIH- présentaient une IDR nulle (0 mm) [59]. Des épidémies de tuberculose à bord de l'US Navy ont fait ressortir des constatations semblables, avec 5 à 10% des membres d'équipage restant négatifs à l>IDR sous une intense exposition [60].

La caractérisation de ce phénotype de résistance a eu un regain d'intérêt ces dernières années, preuve en est, ces récentes revues sur le sujet : Verrall et al. (2014) [61], Möller et al.

(2018) [48], Simmons et al. (2018) (49) et de nombreuses autres, aussi publiées en cours de thèse [62],[63] ou [64]. Selon la définition basée sur l'un ou l'autre de ces tests IDR ou IGRA, ou les 2, l'intensité de l'exposition et la durée de suivi, la prévalence de la résistance variait de ~7 à 25% dans le contexte des contacts familiaux exposés en pays endémiques comme en Indonésie [50], en Inde [65] ou en Ouganda [66]. Une méta-analyse observait ~50% de contacts familiaux infectés (définition à partir de la positivité de l'IDR) dans une vingtaine d'études, et avec une très forte hétérogénéité des résultats [67] (Figure 10).

Le modèle expérimental de tuberculose du macaque cynomolgus a permis également de mettre en évidence cette hétérogénéité phénotypique et le spectre complet de la maladie, ce qui est impossible chez l'humain [68]. Dans ce modèle, les macaques étaient infectés par des doses de *Mtb* directement instillées dans les poumons par bronchoscopie. Cette expérience permet de contrôler le temps de l'infection et la quantité de *Mtb*. L'infection, contrôlée par IDR positive, a abouti à de multiples signes pathologiques allant du tissu pulmonaire stérile à la cavité contenant des hautes doses de bacilles. Environ 40% des macaques n'ont pas présenté de symptômes d'une maladie active et étaient donc juste infectés. Plus de la moitié ont présenté des signes cliniques ou radiographiques, là encore avec une grande variabilité : pendant que certains macaques progressaient rapidement à une maladie sévère et décédaient, un autre macaque a spontanément récupéré d'une maladie apparue après 9 mois. La différence principale avec le modèle humain est qu'ici la plupart des primates ont développé la maladie. Cela peut s'expliquer par le mode d'infection assez direct et la quantité de bacilles inoculée. Quoiqu'il en soit, ce modèle synthétise parfaitement le spectre continu de la maladie. D'une manière plus générale, les modèles animaux ont beaucoup apporté (et apporteront encore) à la compréhension de la maladie (souris, cochons d'Inde, poissons zèbre et autres primates) [69].

La caractérisation de facteurs génétiques qui contribuent à protéger de l'infection par *Mtb* est une priorité car cela pourrait se traduire par de nouvelles solutions pour prévenir, diagnostiquer ou traiter la tuberculose.

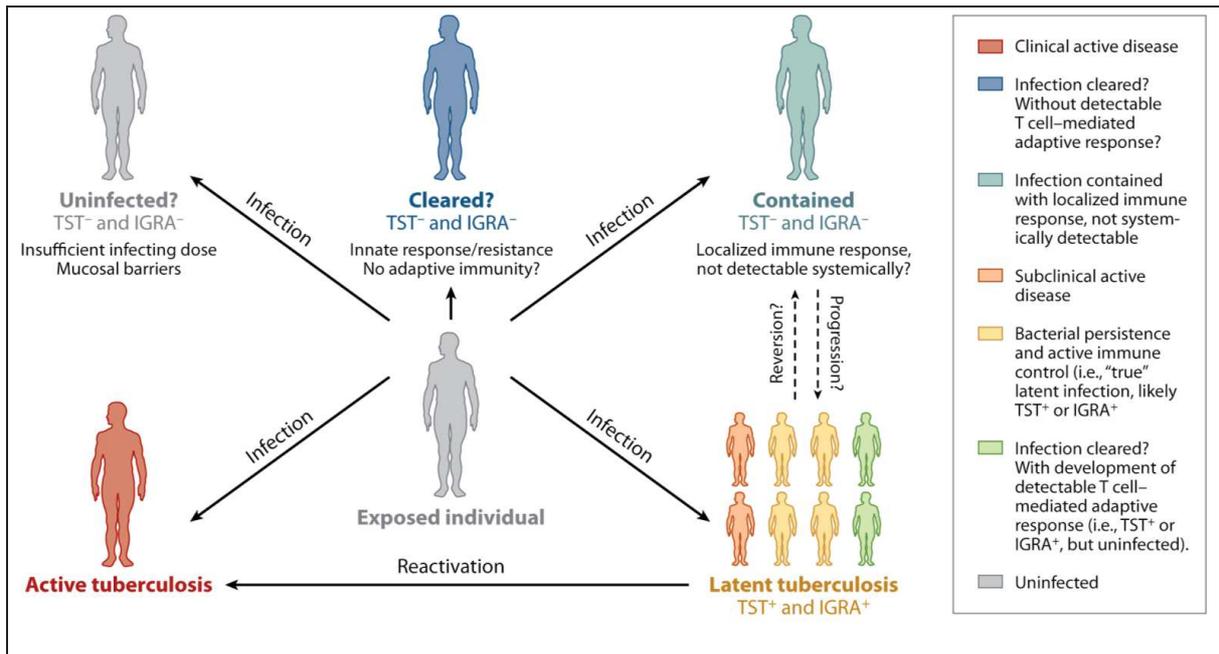


Figure 9. Hétérogénéité du phénotype d'infection tuberculeuse [36]. Un individu exposé à *Mtb* peut rester non-infecté (tests négatifs), développer une réponse immunitaire pour éliminer l'infection (tests négatifs qui auraient pu être positifs à un instant  $t$ ), être infecté mais contenir l'infection (tests négatifs incapables de détecter une réponse immunitaire ?) ou être infecté et asymptomatique (infection latente tuberculeuse, tests positifs).

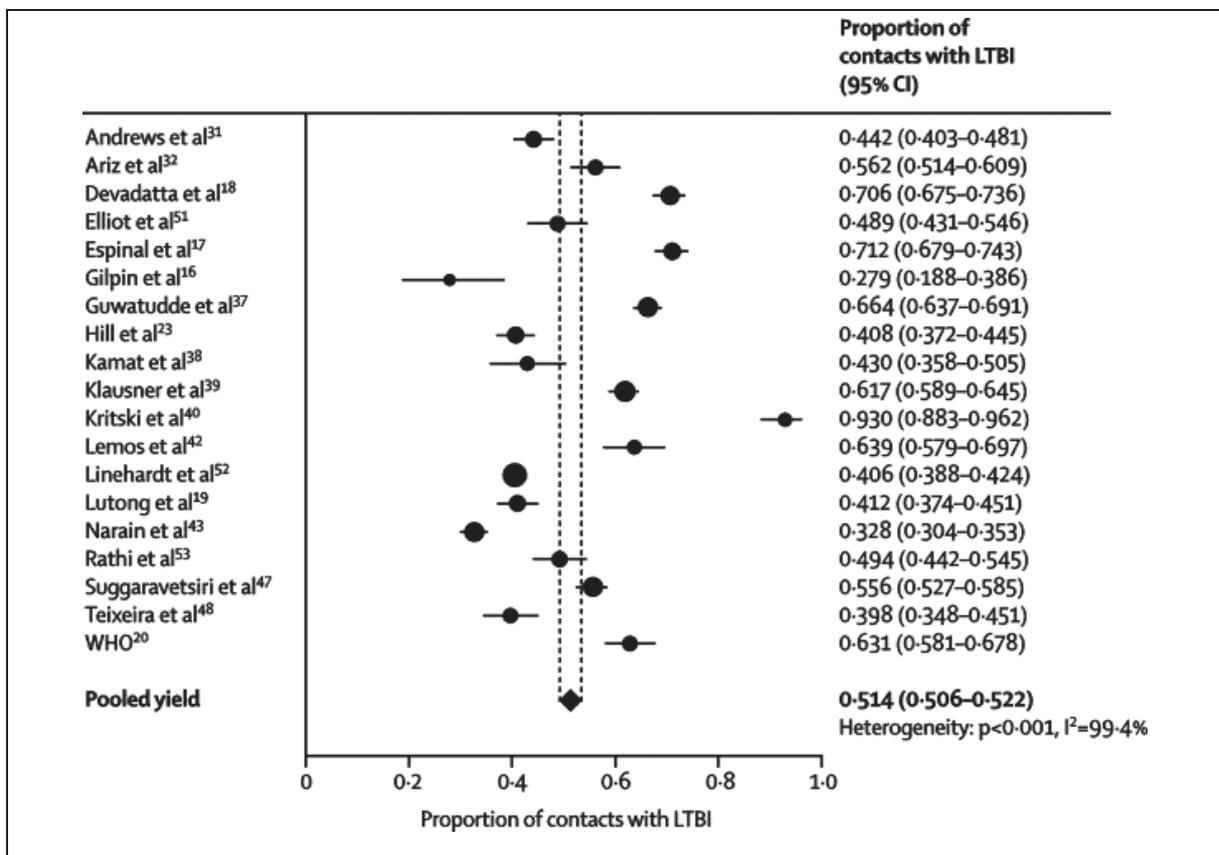


Figure 10. Prévalence de l'infection tuberculeuse chez des contacts familiaux [67].

### 1.2.4 Traitement et prévention

Tous les individus infectés ne développeront pas une tuberculose maladie. Un traitement préventif engendre des coûts et des risques (événements indésirables, inefficacité par pharmacorésistance ou mauvaise observance) et il doit donc être prescrit seulement à des populations à haut risque de développer une maladie [46]. La détermination de ces populations à risque dans les programmes de contrôle de la tuberculose doit notamment prendre en compte l'épidémiologie et les modes de transmission de la tuberculose dans le pays/la région concernés. Au niveau individuel, comme déjà mentionné dans l'interprétation d'un résultat IDR ou IGRA, le contexte clinique et l'évaluation de la balance bénéfices/risques sont essentiels. Les recommandations de l'OMS évoluent régulièrement. Les plus récentes identifient ces populations à risque, pour lesquelles un traitement est très recommandé lorsqu'une infection tuberculeuse est diagnostiquée. Il faut notamment rechercher et traiter systématiquement une ITL chez les sujets VIH+, les contacts récents des cas de tuberculose pulmonaire (adultes et enfants), les patients qui commencent un traitement anti-TNF, les patients en dialyse ou qui se préparent à une greffe d'organe. Le personnel soignant, les détenus, les personnes en provenance de pays où l'incidence de la tuberculose est élevée, les sans-abris et les consommateurs de drogues représentent également des populations à risque où dépistage et traitement doivent être faits selon le contexte.

Les options thérapeutiques suivantes sont recommandées (et avec une efficacité équivalente) pour traiter une ITL : 6 mois sous isoniazide ; 9 mois sous isoniazide ; rifapentine plus isoniazide chaque semaine pendant 3 mois ; isoniazide plus rifampicine pendant 3 à 4 mois ou isoniazide seule pendant 3 à 4 mois (Tableau 2). Il est intéressant de noter qu'en cas de contact avec une tuberculose multi-résistante, il est recommandé de soumettre les contacts à une surveillance clinique rigoureuse plutôt qu'à un traitement préventif.

Tableau 2. Traitement de l'infection tuberculeuse selon les recommandations de l'Organisation Mondiale de la Santé [46].

Drug regimen	Dose per kg body weight	Maximum dose
Isoniazid alone, daily for 6 or 9 months	Adults, 5 mg Children, 10 mg (range, 7-15 mg)	300 mg
Daily rifampicin alone for 3-4 months	Adults, 10 mg Children, 15 mg (range, 10-20 mg)	600 mg
Daily isoniazid plus rifampicin for 3-4 months	Isoniazid: Adults, 5 mg Children, 10 mg (range, 7-15 mg) Rifampicin Adults, 10 mg Children, 15 mg (range, 10-20 mg)	Isoniazid, 300 mg Rifampicin, 600 mg
Weekly rifapentine plus isoniazid for 3 months (12 doses)	Individuals aged $\geq 12$ years: Isoniazid: 15 mg Individuals aged 2-11 years: isoniazid: 25 mg Rifapentine: 10.0-14.0 kg = 300 mg 14.1-25.0 kg = 450 mg 25.1-32.0 kg = 600 mg 32.1-50.0 kg = 750 mg > 50 kg = 900 mg	Isoniazid, 900 mg Rifapentine, 900 mg

Pour clore ce paragraphe et cette introduction générale sur la tuberculose et en particulier l'infection, il convient d'évoquer la vaccination comme moyen de prévention [70]. À ce jour, le vaccin BCG est le seul vaccin utilisé contre la tuberculose. Développé en 1921 par Albert Calmette et Camille Guérin à partir de *Mycobacterium bovis*, il est surtout efficace contre les formes sévères et extra-pulmonaires de l'enfant [71]. Son principe repose sur l'activation d'une immunité naturelle contre les antigènes mycobactériens et l'imitation d'une infection naturelle afin d'établir une mémoire immune capable d'assurer une réponse rapide et efficace lors d'une éventuelle exposition à *Mtb*. Malgré une forte réponse des lymphocytes auxiliaires Th1, le BCG ne protège que très partiellement contre la tuberculose pulmonaire et il est plus efficace dans les pays à faible incidence [72]. Dans les pays hyper-endémiques, le pic d'incidence se situe chez les jeunes adultes : il y a donc une altération de son efficacité durant l'enfance, la puberté ou la fin de l'adolescence (le BCG étant donné à la naissance). Une hypothèse serait que le BCG induit surtout une réponse de lymphocytes T à mémoire insuffisante à long-terme ou sujette à une usure à la suite d'une infection chronique ou d'une exposition très répétée, inévitable dans ces régions. Une méta-analyse a montré que le BCG était plus efficace en l'absence d'une infection préalable à *Mtb* ou une sensibilisation aux mycobactéries non-tuberculeuses (partageant une partie de leurs antigènes avec *M. bovis*) [71]. Autres limitations, comparé à *Mtb*, un quart des épitopes T sont absents dans les souches BCG (il en existe plusieurs

à travers le monde : Pasteur, danoise, Glaxo, Tokyo, russe, Moreau), tout comme le système ESX-1, un facteur clé de virulence sécrétant ESAT-6 ou CFP10 ainsi que d'autres lipides de surface. D'une part, cette hétérogénéité des souches entraîne des réponses différentes [73] ; d'autre part, les systèmes ESX étant un moyen essentiel pour *Mtb* d'échapper à l'hôte et leurs antigènes étant très reconnus par les lymphocytes T, ce sont devenus des candidats très attractifs dans le développement des sous-unités d'un vaccin comme le récent MTBVAC [70].

Le BCG a été récemment repensé. Un essai de phase 2 en Afrique du Sud a montré une réduction du taux de conversion d'IGRA chez des adolescents vaccinés à la naissance et revaccinés pour l'étude [12]. Dans une autre étude, l'administration du BCG par voie intraveineuse a offert une forte protection contre l'infection chez des macaques chez qui on a inoculé *Mtb* 6 mois plus tard [13]. Un nouveau vaccin reste quoiqu'il en soit indispensable pour endiguer la pandémie.

En 2019, 14 candidats ont été évalués en essai clinique (Figure 11). La plupart d'entre eux étaient destinés à la prévention de la tuberculose maladie. L'une des grandes questions dans le développement d'un candidat est de savoir s'il faudrait privilégier un vaccin en prévention de l'infection ou de la maladie. Il peut sembler intuitif ou évident que la première approche serait la meilleure stratégie d'un point de vue de santé publique. Le fait que le phénotype de résistance à l'infection (des individus fortement exposés mais non-infectés) ait été décrit, et que le BCG peut offrir une certaine protection contre l'infection [74] justifient qu'elle soit d'ailleurs explorée. Dans des pays hyper-endémiques où les taux annuels d'infection sont très supérieurs à ceux de la maladie, les essais peuvent être plus petits, plus courts et moins coûteux. Une difficulté majeure réside, comme on l'a vu précédemment, dans l'impossibilité de mesurer directement l'acquisition, la persistance ou l'élimination d'une infection asymptomatique. En outre, un candidat vaccin qui ne peut pas prévenir l'infection pourrait tout de même protéger contre la progression vers une maladie active en contenant *Mtb*, ce qui serait un grand progrès. À l'inverse, un candidat qui protégerait de l'infection chez des individus qui, s'ils étaient infectés, progresseraient quoi qu'il arrive vers une maladie n'a que peu d'intérêt. Le candidat idéal dans le scénario de prévention de l'infection doit en effet surtout cibler les 10% d'individus infectés qui pourraient développer la maladie et non les 90% d'autres capables de contenir une infection asymptomatique. Le récent essai chez des adolescents sud-africains testant la revaccination par BCG ou le vaccin H4:IC31 a été en ce sens encourageant mais les objectifs principaux de réduction du risque d'infection n'ont pas été atteints par aucune des 2 méthodes [12].

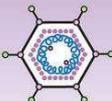
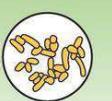
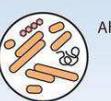
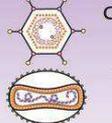
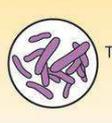
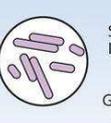
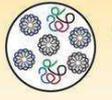
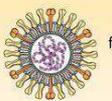
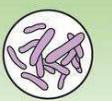
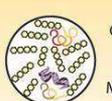
Phase 1	Phase 2a	Phase 2b	Phase 3
 <p><b>Ad5 Ag85A</b> McMasterUniversity, CanSino</p>	 <p><b>RUTI</b> Archivel Farma, S.L.</p>	 <p><b>DAR-901</b> Dartmouth University</p>	 <p><b>Vaccac™</b> Ahui Zhifei Longcom</p>
 <p><b>ChAdOx1.85A/ MVA85A (ID/IM/Aerosol)</b> University of Oxford, University of Birmingham</p>	 <p><b>MTBVAC</b> Biofabri, Tuberculosis Vaccine Initiative, Zaragoza</p>	 <p><b>M72:AS01<sub>E</sub></b> GlaxoSmith Kline, Aeras</p>	 <p><b>VPM1002</b> Serum Institute of India, Max Planck, Vakzine Projekt Management GmbH, Tuberculosis Vaccine Initiative</p>
<p><b>Vaccac™</b>: <i>M. vaccae</i> lysate  <b>VPM1002</b>: Live recombinant BCG with urease C deletion and lysteriolysin insertion  <b>MIP</b>: Inactivated <i>M. indicus pranii</i>  <b>DAR-901</b>: Heat killed <i>M. obuense</i>  <b>M72:AS01E</b>: Recombinant fusion protein (M72) of Rv1196 (PPE protein) and Rv0125 (serine proteinase) in AS01E adjuvant; a liposomal formulation of Toll-like receptor (TLR)-4 ligand (MPL-A) and saponin-fraction QS21  <b>H56:IC31</b>: Recombinant fusion protein (Hybrid-56) of three antigens expressed during early (Ag85B/mycolyl transferase), presumably late (Rv2660c) or all (ESAT-6/EsxA) stages of <i>Mtb</i> infection in IC31 adjuvant that signal through TLR-9  <b>RUTI</b>: Liposomized killed <i>M. tuberculosis</i> fragments  <b>MTBVAC</b>: Live attenuated <i>M. tuberculosis</i> vaccine with PhoP and FadD26 deletions  <b>BCG revaccination</b>: Live <i>M. bovis</i> BCG vaccine  <b>ID93:GLA-SE</b>: Recombinant fusion protein (ID93) of four antigens associated with virulence (Rv2608, Rv3619c/EsxV, and Rv3620c/EsxW) or latency (Rv1813) in the oil-in-water emulsion (GLA-SE) of MPL and glucopyranosyl lipid</p>	 <p><b>ID93:GLA-SE</b> Infectious Disease Research Institute, Aeras</p>	 <p><b>H56:IC31</b> Statens Serum Institute, Valneva, Aeras</p>	 <p><b>MIP</b> Cadila ICMR</p>
	 <p><b>TB/ FLU-04L</b> Research Institute for Biological Safety Problems</p>	 <p><b>BCG Revaccination</b></p>	
	 <p><b>GamTBvac</b> Gamaleya Research Institute of Epidemiology &amp; Microbiology, Russia</p>	<p><b>GamTBvac</b>: Two <i>Mtb</i> antigen fusions (Ag85A &amp; ESAT-6-CFP-10) with dextran-binding domain and mixed with DEAE-dextran &amp; CpG (TLR-9 agonist) adjuvant  <b>TB/FLU-04L</b>: Attenuated influenza viral vector expressing <i>Mtb</i> Ag85A and ESAT-6/EsxA  <b>Ad5 Ag85A</b>: Replication-deficient human adenovirus serotype-5 vector expressing <i>Mtb</i> Ag85A  <b>ChAdOx1.85A/MVA85A</b>: Replication-deficient Chimpanzee adenovirus/ modified vaccinia Ankara virus vector expressing <i>Mtb</i> Ag85A</p>	

Figure 11. Vaccins candidats en évaluation en 2019 [70].

### 1.2.5 Facteurs de risque de l'infection

L'issue de l'infection par *Mtb* et si celle-ci sera contrôlée par l'individu ou bien progressera en maladie active est donc complexe et déterminée en grande partie par des facteurs liés à l'environnement, notamment l'exposition, au pathogène en lui-même et enfin à l'hôte (génétiques ou non-génétiques) (Figure 12).

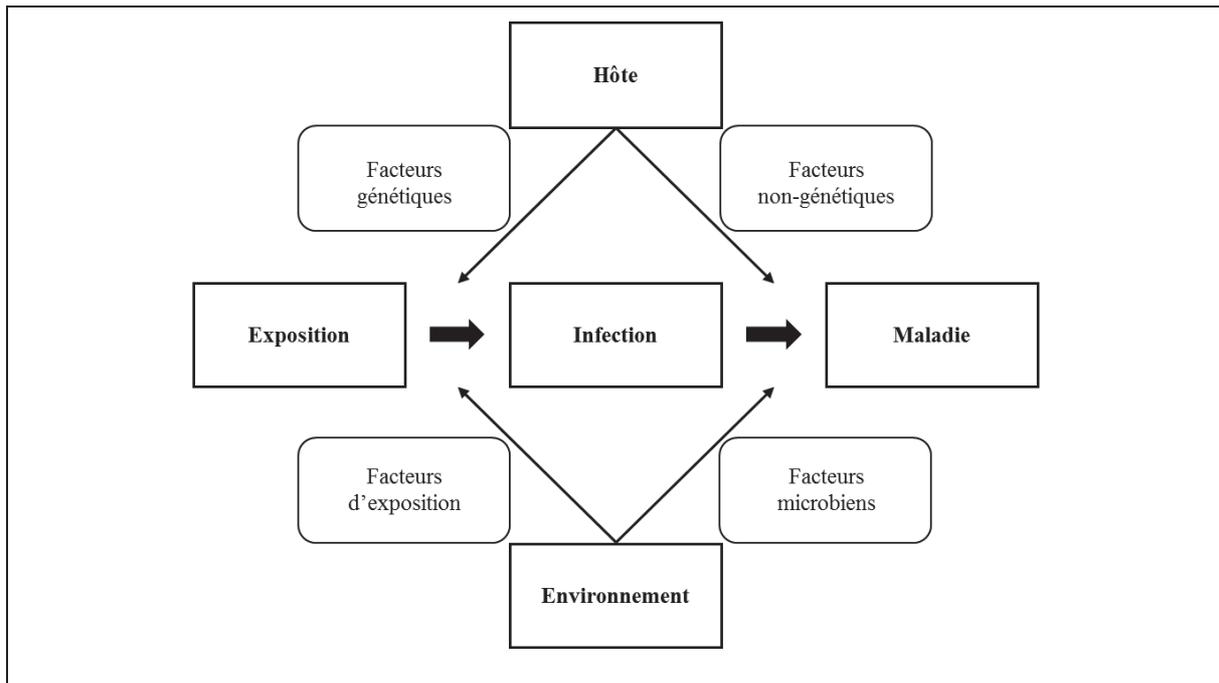


Figure 12. Facteurs de risques associés à la susceptibilité à l'infection tuberculeuse puis au développement d'une maladie active.

### 1.2.6 Facteurs de risque non génétiques

Les facteurs environnementaux d'exposition à *Mtb* modulent en premier lieu le risque d'infection tuberculeuse et sont relativement bien identifiés. Pour être infecté, il faut avant tout rencontrer le pathogène.

Au niveau populationnel, le risque d'infection dépend de la prévalence de la tuberculose dans la population. Le risque annuel d'infection tuberculeuse (RAIT), représentant la proportion d'individus d'une population qui deviennent infectés pendant l'année, était initialement estimé à partir de larges enquêtes sur l'IDR [75]. Ce RAIT a longtemps été utilisé par les programmes de lutte antituberculeuse pour dériver l'incidence de tuberculose maladie (un nombre essentiel dans le contrôle de la maladie qui permet par exemple de mesurer le succès d'un programme mis en place) [76]. En effet, il est difficile de suivre prospectivement pendant des années des cohortes de populations générales et compter le nombre de nouveaux malades par an, notamment dans les pays avec des ressources limitées. La règle de Styblo a permis de définir une relation mathématique entre l'incidence de la tuberculose maladie, sa prévalence et le RAIT (Figure 13) ; c'est-à-dire que 50 nouveaux cas de tuberculose pour 100 000 habitants par an correspondait à un RAIT de 1%. L'apparition des antibiotiques puis de la pandémie du VIH,

ainsi que des politiques de santé publique de contrôle de la maladie ont cependant rendu cette règle imparfaite [77]. Par exemple, des programmes nationaux de dépistage et de traitement de malades réduisent les durées moyennes d'infection et interrompent les transmissions du bacille. De même, des individus VIH+ deviennent symptomatiques plus rapidement que des individus VIH-, menant à des détections des cas plus précoces et donc une interruption des transmissions. Dans ce contexte, ainsi que celui d'un accroissement démographique lors des dernières décennies, en particulier dans des régions très endémiques, le fardeau de l'infection tuberculeuse était probablement très surestimé (le classique « un tiers de la population mondiale infectée » [78]). Houben et Dodd ont récemment réévalué cette estimation [27]. Les estimations du RAIT ont été révisées (i) par une approche directe, en se reposant sur de larges études de prévalence de positivité à l'IDR (le RAIT est alors égal à  $1 - (1 - \text{prévalence})^{1/\text{âge}}$ ) et (ii) par l'approche indirecte de Styblo via la prévalence de la maladie mais en tenant compte des effets possibles du VIH et de l'âge. Finalement, la prévalence mondiale de l'infection a été légèrement revue à la baisse, à 23%, soit un peu moins d'un quart de la population.

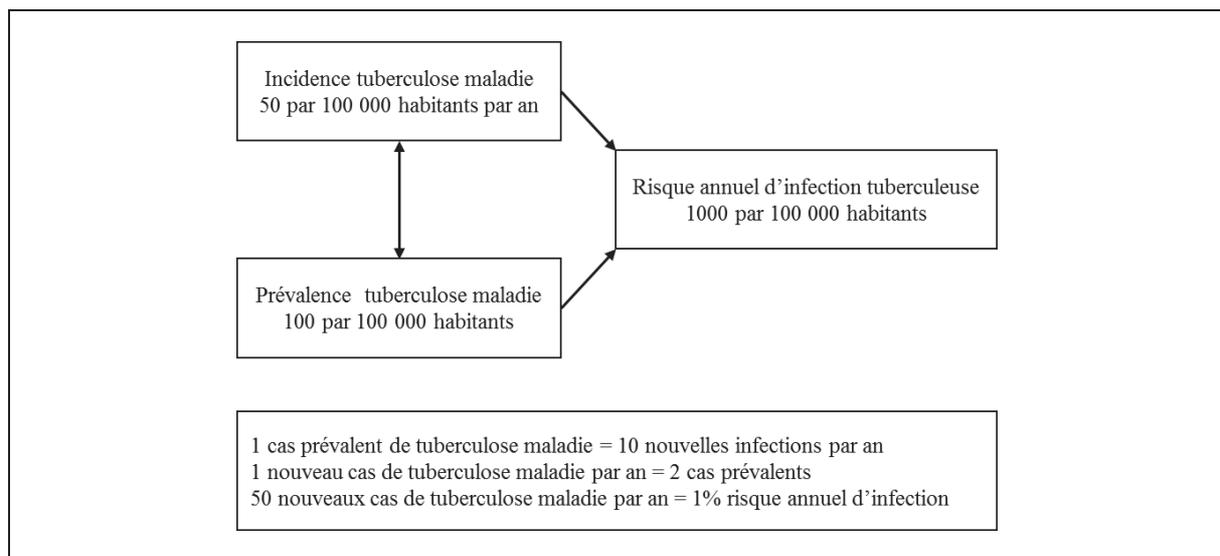


Figure 13. Relation entre risque annuel d'infection, prévalence de tuberculose maladie et incidence de tuberculose maladie selon la règle de Styblo [77].

Au niveau individuel, le risque d'infection dépend beaucoup du cas index contagieux. Dès le milieu du XX<sup>ème</sup> siècle, des études épidémiologiques ont observé que des malades avec crachats positifs étaient plus contagieux que les autres malades [79]. De même, la concentration de bacille dans les crachats d'un malade est corrélée positivement avec sa contagiosité [80]. Des patients avec frottis négatifs présentent peu de bacilles dans leurs expectorations mais ils

restent tout de même contagieux. Dans une étude menée à Vancouver, Canada, au moins 1/6<sup>ème</sup> des nouvelles transmissions de tuberculose étaient dues à des cas avec crachats négatifs [81]. Des résultats similaires ont été observés à San Francisco ou aux Pays-Bas [82],[83].

Les contacts proches de patients tuberculeux, comme les contacts familiaux ou le personnel soignant ont plus de risques d'être infectés. Les premières études épidémiologiques conduites au début du XX<sup>ème</sup> siècle l'ont confirmé [84],[85]. De même, l'intensité du contact estimée par la durée de l'exposition, la proximité du contact ou l'aération du lieu d'exposition jouent un rôle significatif dans le risque d'infection [86],[87]. Une démonstration de l'influence de ces facteurs a été faite dans une enquête portant sur un foyer épidémique apparu lors d'un trajet en avion : les passagers assis jusqu'à 2 rangs autour du cas index avaient un risque 8.5 fois plus élevés d'avoir une IDR positive comparé aux autres passagers [88]. D'un point de vue général, la promiscuité augmentant la probabilité d'infection, la recherche de contacts infectés a longtemps ciblés les membres de la famille. Mais l'importance d'une transmission communautaire, en dehors du cercle familial et la définition d'un contact font débat depuis longtemps [89]. D'anciennes études épidémiologiques avaient remarqué qu'une majorité d'enfants avec IDR positive ne reportait aucun contact familial avec un cas index, et que par conséquent l'infection s'était probablement produite en dehors du foyer familial [90],[91]. Cette transmission communautaire s'observe aisément en zone hyper-endémique ; de récentes études d'épidémiologie moléculaire où l'on peut identifier les souches de *Mtb* de malades et remonter jusqu'au cas index source ont aussi montré que la tuberculose se transmettait dans des zones de moyenne ou basse incidence [92],[93],[94]. De nombreuses méta-analyses ont depuis confirmé l'importance du risque d'infection en communauté par rapport aux transmissions intra-familiales [95],[96],[97].

On l'a évoqué dans un paragraphe précédent, des facteurs spécifiques à la bactérie jouent sûrement un rôle sur le risque d'infection. La diversité du complexe *Mtb* influe sur la diversité de la réponse immune et inflammatoire de l'hôte. Si la virulence a un effet marqué sur la pathogénie d'une tuberculose maladie, notamment via le système de sécrétion ESX-1, le risque d'infection est probablement plus modulé par le degré de transmissibilité de telle ou telle lignée. Les études de génomique comparative et d'épidémiologie moléculaire ont démontré que les lignées de *Mtb* se sont adaptées au fil des millénaires/siècles/années aux populations rencontrées et que si une lignée est désormais rarement retrouvée dans une population, c'est simplement car elle ne s'est pas adaptée pour être transmise d'un hôte à l'autre [98].

D'autres facteurs spécifiques à l'individu sont associés à l'infection tuberculeuse. L'âge est un facteur connu et d'une manière générale le risque augmente avec les années, ce qui

reflèterait une exposition cumulée à *Mtb* au cours de la vie. Cependant, cette relation n'est pas complètement linéaire : le risque est en effet plus élevé chez les adolescents et les jeunes adultes. Dans une récente étude conduite au Pérou [99], le risque d'infection augmentait depuis la naissance jusqu'à l'âge de 20 ans, en ligne avec ce qui pouvait être observé en Afrique du Sud (prévalence de l'infection chez les enfants de 5 ans autour de 20% puis 50% chez les adolescents de 15 ans et 75 % chez les adultes de 25 ans) [100] ou en Ouganda (presque 20% d'adolescents infectés contre moins de 10% d'enfants) [101]. Les plus jeunes enfants et nouveau-nés sont certes à risque de s'infecter mais surtout de développer rapidement une tuberculose maladie (particulièrement avant 2 ans et souvent dans l'année suivant la primo-infection). La présence d'un cas index avec crachats positifs dans un ménage demeure le facteur de risque le plus important pour des nouvelles infections chez l'enfant [90],[101]. Cette tendance s'estompe à partir de l'adolescence où les transmissions semblent d'origine communautaire sans qu'il n'y ait d'explications [99],[101]. Une plus grande mobilité et un développement accru des relations sociales chez les jeunes adultes pourraient en être une mais l'étude péruvienne apporte un autre éclairage : les auteurs n'ont pas simplement comparé des prévalences d'infections par classes d'âges mais ont modélisé un risque d'infection âge-spécifique familial (la probabilité qu'une personne non-infectée d'âge  $x$  s'infecte après contact avec un cas malade de sa famille) et communautaire (la probabilité qu'une personne non-infectée d'âge  $x$  s'infecte après une année d'exposition communautaire) et ont constaté que des pics du risque d'infection familiale et communautaire arrivaient au même moment chez les jeunes adultes. Ce résultat serait plus en faveur d'une influence biologique que sociale (puberté ?) sinon le risque communautaire aurait montré un grand pic pendant que le risque familial serait resté constant. Enfin, les taux d'infection semblent décroître chez les plus âgés, mais cela reflète probablement les limitations de l>IDR et l'augmentation de faux négatifs à mesure que l'âge avance (l>IDR devient moins réactive avec l'âge et est incapable de diagnostiquer une réinfection) [102].

Le sexe masculin est un facteur de risque établi depuis longtemps pour le développement d'une tuberculose maladie (le ratio homme:femme est  $\sim 2:1$  et très prononcé chez les jeunes adultes) mais moins certain quant à une plus grande susceptibilité à l'infection [103],[104]. Des facteurs génétiques, biologiques et socioculturels expliquent probablement cette disparité entre les sexes, sans que leur part respective n'ait été élucidée.

De la même façon que le sexe, les comorbidités telles que l'immunosuppression acquise (coinfection VIH ou traitement anti-TNF), la malnutrition (qui entraîne une immunodépression) ou le diabète de type 2 ainsi que la consommation de tabac ou d'alcool sont clairement associés

à un risque accru de progresser vers une tuberculose maladie, mais leur rôle dans la susceptibilité à l'infection tuberculeuse est beaucoup moins évident [105],[106],[107],[108],[109]. Il est en particulier difficile d'évaluer leur impact car l'immunodépression biaise les résultats des tests IDR ou IGRA.

Enfin, la carence en vitamine D est un facteur de risque souvent évoquée mais qui prête à débat. Dans une très récente étude randomisée conduite chez des écoliers d'Oulan-Bator en Mongolie, la supplémentation en vitamine D ne protégeait pas de l'infection comparée à la prise d'un placebo [110].

### 1.2.7 Facteurs de risque génétiques

Il convient enfin d'évoquer la contribution des facteurs génétiques de l'hôte. Si l'avènement de la théorie microbienne à la fin du XIX<sup>ème</sup> siècle l'a un peu occultée, la théorie génétique des maladies infectieuses est pourtant ancienne : Hippocrate au V<sup>ème</sup> av. J.-C. observait que la phtisie (ancien nom de la tuberculose) avait tendance à se transmettre au sein des familles comportant déjà un malade [111]. L'idée que la tuberculose était héréditaire, et non contagieuse, a donc longtemps prédominé. Georges Sand s'en offusquait d'ailleurs dans ses *Correspondances* [112]. En 1839, alors en voyage aux Baléares avec Frédéric Chopin (l'Europe du Sud adoptait alors des mesures drastiques à l'encontre des malades), elle écrivait : « *Au bout d'un mois, le pauvre Chopin qui, depuis Paris, allait toujours toussant, tomba plus malade et nous fîmes appeler un médecin, deux médecins, trois médecins, tous plus ânes les uns que les autres, et qui allèrent répandre, dans l'île, la nouvelle que le malade était poitrinaire au dernier degré. Sur ce, grande épouvante ! La phtisie est rare dans ces climats et passe pour contagieuse.* » Ce n'est que 25 ans plus tard que Jean-Antoine Villemin démontra la nature infectieuse de la tuberculose en inoculant chez un lapin un échantillon prélevé dans une cavité d'un malade tout juste décédé [113].

L'existence de facteurs génétiques prédisposant au développement d'une tuberculose active est aujourd'hui bien établie et leur identification a été faite au fur et à mesure des progrès de la génétique humaine : des études de jumeaux (montrant une plus grande concordance phénotypique entre jumeaux monozygotes, qui partagent 100% de leur patrimoine génétique, par rapport aux jumeaux dizygotes [114]) aux études de liaison localisant des régions contenant les gènes influant la susceptibilité (6 de ces études ont été contributives mais les régions

identifiées étaient souvent larges et ne contenaient pas des gènes d'intérêt particulier) [115]. Ces études utilisaient des données familiales et combinaient l'information phénotypique et l'information des transmissions des allèles à un locus d'intérêt. L'idée était de tester la présence d'une corrélation (liaison) entre le phénotype et le marqueur étudié, indiquant une cotransmission du locus d'intérêt avec le trait. Cependant, ces études ne permettaient que d'identifier des régions candidates susceptibles de contenir un locus causal.

Le développement du génotypage a permis d'explorer des signaux d'association entre des polymorphismes nucléotidiques (SNPs pour *single-nucleotide polymorphisms*), fréquents en population générale, et un phénotype d'intérêt (ici, cas de tuberculose pulmonaire et contrôles). Les études gènes-candidats se focalisant sur des gènes connus *a priori* ont pu parfois identifier des polymorphismes associés à la tuberculose, mais ils ont rarement été validés dans des études indépendantes [116].

Des études d'association pangénomique (GWASs pour *genome-wide association studies*) ont alors pris le relai (permettant d'avoir une vue sans hypothèse *a priori*). Ces études consistent à tester l'association de chaque SNP avec le phénotype, de façon indépendante, puis de corriger le test utilisé afin d'éviter un trop grand nombre de faux positifs. Une douzaine de GWASs concernant la tuberculose à ce jour ont été menées (ce qui est extrêmement peu en comparaison avec d'autres maladies), là aussi avec des résultats mitigés [117]. Ces études, qui nécessitaient de larges échantillons, n'ont pas abouti à la découverte de marqueurs avec des effets majeurs sur la maladie.

L'avènement des techniques de séquençage et la recherche de polymorphismes plus rares a cependant grandement fait avancer la compréhension des bases génétiques et immunologiques de la maladie. 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 [37]. Les plus gros progrès ont été réalisés grâce à l'étude du syndrome de susceptibilité mendélienne aux mycobactéries (MSMD), causé par une vulnérabilité spécifique aux espèces mycobactériennes non tuberculeuses peu virulentes comme le bacille du BCG et les mycobactéries environnementales et des défauts de l'immunité liée à l'IFN- $\gamma$  [118]. Une quinzaine de gènes causant ce syndrome ont pu être identifiés [119]. En outre, à partir de l'observation du développement d'une tuberculose chez certains de ces patients, est venue l'hypothèse que leur tuberculose pouvait être également le résultat d'une prédisposition génétique de type monogénique, en particulier lorsqu'au sein d'une fratrie partageant le même défaut génétique, le seul phénotype infectieux ségrégant avec la mutation était une tuberculose sévère [120]. Le défaut génétique le plus

commun identifié chez des patients atteints de tuberculose sévère aujourd'hui est un déficit complet de la chaîne  $\beta 1$  du récepteur à l'interleukine-12 codée par le gène *IL12R $\beta$ 1*, observée dans plusieurs familles [121]. Ces résultats valident le concept de prédisposition monogénique, tout comme le récent travail de mon collègue Gaspard Kerner qui a pu identifier une cause monogénique fréquente de susceptibilité à la tuberculose par analyse d'exome (homozygotie d'un polymorphisme du gène *TYK2*) [122],[123].

De la même manière que pour les facteurs de risques non-génétiques, le phénotype d'infection tuberculeuse a été beaucoup moins étudié, probablement lié à la complexité à le définir correctement. La réactivité de l'IDR est corrélée entre jumeaux, mais pas entre des enfants non-apparentés qui vivent dans le même foyer [124]. De nombreuses études ont montré une forte héritabilité de l'IDR et des IGRA en traits quantitatifs (et non binaires, infectés/non-infectés). L'héritabilité représente la proportion de variance phénotypique expliquée par la composante génétique au sein d'une population donnée. La variance d'un trait quantitatif peut être décomposée en la somme d'une composante génétique, d'une composante environnementale, de la covariance entre ces 2 et d'une composante d'interaction gène-environnement [125]. La variance génétique peut elle-même se décomposer en plusieurs composantes mais l'héritabilité est généralement estimée à partir d'un modèle plus simple et calculée à partir de l'étude de la ressemblance phénotypique entre apparentés. Ainsi, des études conduites dans différentes populations exposées ont montré des héritabilités de 92% pour l'IDR au Chili [124], 71% pour l'IDR et 39% pour l'IGRA en Gambie [126], 58% pour IGRA stimulé uniquement par ESAT-6 en Afrique du Sud [127], ou bien 27% pour un phénotype similaire en Ouganda [128].

À la suite de ces analyses d'héritabilités, démontrant une vraie contribution génétique, et même si probablement surestimées, peu d'études ont tenté d'identifier les gènes ou polymorphismes génétiques pouvant expliquer ces résultats. Une analyse de liaison pangénomique conduite en Ouganda a montré que des régions chromosomiques 2q21-2q24 (affinée plus tard vers le gène *ZEB2*) et 5p13-5q22 étaient liées à une IDR négative de façon persistante [129],[130]. Les associations gène-candidat les plus intéressantes ont identifié des polymorphismes d'*IL10* (au Ghana et au Brésil) et d'*ULK1* comme susceptibles d'accroître le risque d'infection en prenant un phénotype binaire IDR négatif/positif [131],[132],[133]. Jusqu'à présent, 3 GWASs se sont intéressées à ce phénotype. Une première conduite chez des participants de 23andMe était très peu informative (phénotype d'infection mal défini et populations non exposées notamment) [134]. La suivante, réalisée en Islande sur une population un peu plus exposée d'après les auteurs, explorait différents phénotypes tuberculose dont la

positivité à l'IDR (seuil non défini par ailleurs) et mettait en évidence un signal dans la région HLA (*Human Leukocyte Antigen*) [135]. La troisième et dernière GWAS est la plus intéressante : en recrutant des participants VIH+ dans des zones endémiques d'Ouganda et Tanzanie, les auteurs ont émis l'hypothèse que ceux ne convertissant pas l'IDR malgré une forte exposition et leur immunodéficiences étaient résistants à l'infection tuberculeuse. Avec une taille d'échantillon modeste, un locus conférant une forte protection (un odds ratio, OR, à 0.27) a ainsi pu être découvert à côté du gène *IL9* sur le chromosome 5 [136].

Le laboratoire qui m'accueille a été pionnier dans l'identification de facteurs génétiques des maladies infectieuses. Une analyse de la réponse IDR, menée par ma directrice de thèse Aurélie Cobat, identifiait dans un échantillon familial sud-africain 2 loci sur les chromosomes 11p14 (TST1) et 5p15 (TST2) influençant respectivement (i) la réponse binaire (0 versus > 0 mm), et (ii) l'intensité de la réaction [137]. Le locus TST1 a été par la suite validé dans une cohorte de contacts familiaux du Val-de-Marne en France [138]. En outre, un locus ayant un effet sur la production de TNF après stimulation par BCG a été trouvé à proximité de TST1 [139]. Si TST1 est effectivement un indicateur de la résistance innée à l'infection, alors il existe une possibilité pour que celle-ci implique une voie de réponse médiée par le TNF. Poursuivant ces travaux, ma collègue Fabienne Jabot-Hanin a ensuite mis en évidence 2 loci sur les bras longs des chromosomes 3 et 8 contrôlant la production d'IFN- $\gamma$  après stimulation par BCG et ESAT-6, toujours dans les cohortes sud-africaine et française [140]. Le travail de cette thèse, notamment la deuxième partie, s'inscrit ainsi dans la lignée de ces projets par la recherche de polymorphismes génétiques associés à l'infection en s'appuyant sur un phénotype unique et spécifique, basé à la fois sur les résultats IDR et IGRA.

### 1.3 Objectifs de la thèse

L'objectif général de ce travail consiste à identifier des facteurs de risque de l'infection par *Mtb*. Comme on a pu le voir dans cette introduction, ces facteurs ont finalement été peu explorés en comparaison avec ceux de la tuberculose maladie, en particulier la tuberculose pulmonaire. En outre, la définition du phénotype d'infection repose sur des tests indirects qui mesurent une immunoréactivité aux antigènes mycobactériens et incapables de détecter la bactérie vivante. Il est donc compliqué de faire la part entre vraie infection et infection ancienne. Enfin, les 2 types de tests présentent chacun des limitations qui leur sont propres.

Dans ce contexte, on se propose dans une première partie d'explorer les facteurs de risque non-génétiques d'une population à risque et plus exposée que la population générale, le personnel de santé. Je me suis appuyé sur des données recueillies par des collègues marocains, le fruit d'une longue collaboration avec le laboratoire. Un grand échantillon de plus de 600 employés a été ainsi recruté dans 2 hôpitaux de Rabat et Meknès et ont été testés à la fois par IDR et IGRA. J'ai examiné la prévalence d'ITL et les facteurs de risques associés en définissant le phénotype d'infection par différents seuils et combinaisons.

Dans une seconde partie, j'ai exploré les facteurs de risque génétiques dans une autre population très exposée, les contacts familiaux d'un malade tuberculeux ainsi que des individus vivant en zone hyper-endémique. Comme énoncé dans l'introduction, la grande variabilité de réponse chez des individus exposés à *Mtb* peut être en partie expliquée par des facteurs génétiques. L'identification de ces facteurs s'est faite, dans la continuité des travaux du laboratoire, par la recherche de marqueurs génétiques communs (avec une fréquence > 5% en population générale) associés au phénotype d'infection. Plus particulièrement, nos collaborateurs de l'Université McGill, Montréal, Canada et leurs collègues vietnamiens ont établi une large cohorte de contacts familiaux dans des zones endémiques du sud du Vietnam. Le design familial permet d'assurer une forte exposition à *Mtb*. Les participants ont ensuite été testés à la fois par IDR et IGRA et le phénotype d'infection a été défini à partir des résultats des 2 tests, permettant de minimiser des erreurs de classifications du statut infectieux. J'ai ainsi réalisé une étude d'association pangénomique tout en prenant en compte l'apparentement entre les sujets de l'étude. J'ai enfin validé les résultats obtenus en utilisant un phénotype similaire dans 2 autres cohortes, évoquées dans cette introduction, établies en France dans la région Val-de-Marne (région peu endémique pour la tuberculose) et en banlieue du Cap en Afrique du Sud (une des zones présentant les plus hautes incidences de tuberculose de la planète). Les caractéristiques différentes de ces cohortes en terme d'ascendance génétique et de contexte épidémiologique, tout en utilisant un phénotype très spécifique, ont permis d'assurer la robustesse des résultats observés.

## **2 Epidémiologie descriptive et analytique : Prévalence et facteurs de risque de l'infection par *M. tuberculosis* parmi du personnel de santé au Maroc**

### **2.1 Introduction**

Le personnel soignant est plus à risque de s'infecter et développer une tuberculose par rapport à la population générale du fait de leur exposition à des malades à l'hôpital [141]. La prévalence d'ITL est en effet assez élevée, variant de 37 à 63% selon de récentes méta-analyses dans des pays endémiques [141],[142],[143],[144]. Ces pourcentages varient selon le contexte, richesse du pays, couverture BCG et bien sûr définition du statut d'infection. J'ai largement évoqué les caractéristiques et limitations des 2 types de tests diagnostic dans l'introduction, or l'OMS recommande une gestion sur-mesure de l'ITL en fonction de la prévalence nationale et des ressources à disposition [145].

Dans cette première partie, j'ai donc exploré la prévalence et les facteurs de risques de l'infection tuberculeuse en utilisant des résultats IDR et IGRA chez du personnel de santé marocain. Le Maroc est un pays à revenus intermédiaires, avec une incidence moyenne (stable autour de 100 nouveaux cas par an par 100 000 habitants au moment de l'étude) et une couverture vaccinale BCG presque complète (proche de 100% depuis la fin des années 1980) [146].

## 2.2 Matériel et méthodes

Une étude transversale a été conduite entre mars 2012 et septembre 2016 dans 2 villes du Maroc, la capitale Rabat et Meknès au nord-Maroc. Le personnel de santé a été recruté dans 2 centres, l'Hôpital Militaire Mohammed V de Rabat et l'Hôpital Militaire Moulay Ismail de Meknès. Tous les employés âgés de 18 à 60 ans ayant accepté de participer à l'étude ont été inclus. Les critères d'exclusion étaient : grossesse en cours, infections connues par VIH, hépatite B ou C, cancer et insuffisance rénale chronique. Les informations sur les variables suivantes ont été récoltées par un questionnaire : sexe, âge, comorbidités (diabète, asthme ; oui/non), tabagisme (oui/non), antécédent familial de tuberculose (oui/non), secteur de travail et statut professionnel. La variable secteur de travail était catégorisée ainsi: non-clinique (par exemple services administratifs ou les laboratoires, avec une exposition indirecte à des patients), chirurgie (exposition à des patients mais faible risque de transmission aérosol), services cliniques (exposition directe à des patients) et service de pneumologie (exposition directe à des patients tuberculeux). Le statut professionnel a aussi été catégorisé en personnel non-médical (comme le personnel administratif ou les secrétaires), personnel paramédical (comme techniciens de laboratoire) et personnel médical (médecins et infirmières). Tous les participants ont déclaré avoir été vaccinés par le BCG.

Un test IGRA QuantiFERON-TB Gold In-Tube (QFT) a été réalisé sur les échantillons de sang collectés avant que l'IDR ne soit effectuée. Les résultats QFT ont ensuite été catégorisés à partir de la valeur TB antigène (UI/mL) (voir Tableau 2) en positifs ( $\geq 0.35$  UI/mL, QFT+), négatifs ( $< 0.35$  UI/mL, QFT-) ou indéterminés. Pour explorer plus en détails les résultats discordants avec l'IDR, les résultats QFT ont été par la suite catégorisés en négatifs forts ( $< 0.2$  UI/mL), négatifs faibles (0.20–0.34 UI/mL), positifs faibles (0.35–0.70 UI/mL) et positifs forts ( $> 0.70$  UI/mL). Ces seuils ne sont pas totalement arbitraires et sont basés sur deux études importantes portant sur la variabilité du test QuantiFERON [147],[148].

L'IDR a été réalisée de façon classique par un personnel entraîné à effectuer l'injection de tuberculine et à interpréter la réaction 72 heures plus tard. J'ai considéré les 3 seuils de positivité  $\geq 5$ ,  $\geq 10$  et  $\geq 15$  mm (IDR+) pour les comparaisons avec QFT. Ensuite dans l'analyse des facteurs de risque, j'ai considéré le seuil de positivité de 10 mm qui était le plus pertinent après l'analyse de concordance et le seuil utilisé dans d'autres études sur les travailleurs hospitaliers [142], ainsi que des recommandations officielles par exemple du Center for Disease Control and Prevention américain.

La concordance entre les résultats QFT et IDR a été calculée à partir du nombre de double-positifs (QFT+/IDR+) et double-négatifs (QFT-/IDR-) divisé par le nombre total de résultats de tests. L'accord entre les 2 tests a été estimé par le coefficient Kappa de Cohen ( $\kappa$ ). Le test exact de Fisher a été utilisé pour comparer les proportions des sous-catégories de QFT entre résultats concordants et discordants QFT/IDR. La comparaison de l'induration de l'IDR entre les résultats QFT-/IDR+ et QFT+/IDR+ ainsi qu'entre QFT+/IDR- et QFT-/IDR- a été faite avec un test de Mann-Whitney. Les facteurs de risque associés à l'infection tuberculeuse ont été étudiés en utilisant 5 définitions possibles du statut infectieux : (i) QFT+ vs. QFT-, (ii) IDR+ vs. IDR -, (iii) QFT+/ IDR + vs. QFT-/ IDR -, (iv) résultat discordant IDR positif QFT-/IDR+ vs. QFT-/IDR- et (v) résultat discordant QFT positif QFT+/IDR- vs. QFT-/IDR-. Ces analyses ont été faites en utilisant un modèle de régression logistique à effets aléatoires : la variable à expliquer est donc le statut infectieux binaire, les effets fixes des variables explicatives ont été estimés une variable à la fois puis toutes les covariables ensemble, l'effet aléatoire permet de prendre en compte un possible effet de l'hôpital.

## 2.3 Résultats

### 2.3.1 Description de l'échantillon

Un total de 662 employés a été recruté dans les hôpitaux de Rabat ( $n = 328$ ) et Meknès ( $n = 334$ ) et 31 ont été exclus car ils présentaient un résultat QFT indéterminé. Les caractéristiques des 631 participants inclus sont présentées dans le

Tableau 3. Brièvement, 54.0% des participants étaient des hommes, 47.1% avaient moins de 35 ans et 44.1% travaillaient dans un service non-clinique. La prévalence d'un test QFT positif était de 40.7% (IC95% 36.9–44.5%) tandis que celle d'une IDR positive  $\geq 10$  mm était de 52.1% (95%IC 48.2–56.0%). En considérant le statut infectieux chez les participants avec des résultats uniquement concordants (QFT+/IDR+ et QFT-/IDR-), la prévalence de l'infection était de 45.2 % (95%IC 40.7–49.5%). Les valeurs des résultats QFT et IDR sont visualisées ci-après sur la Figure 14.

Tableau 3. Caractéristiques de la cohorte marocaine.

<b>Caractéristiques</b>	<b>n (%)</b>
Sexe	
Masculin	341 (54.0)
Féminin	290 (46.0)
Age	
18-34	297 (47.1)
35-44	186 (29.5)
45-60	148 (23.4)
Comorbidités	
Non	586 (92.9)
Oui	45 (7.1)
Tabagisme	
Non	549 (87.0)
Oui	82 (13.0)
Antécédent familial de TB	
Non	588 (93.2)
Oui	43 (6.8)
Secteur de travail	
Non-clinique	278 (44.0)
Chirurgie	118 (18.7)
Clinique	184 (29.2)
Pneumologie	51 (8.1)
Activité	
Non-médicale	167 (26.4)
Paramédicale	143 (22.7)
Médicale	321 (50.9)
QFT	
Négatif	374 (59.3)
Positif	257 (40.7)
IDR	
<10 mm	302 (47.9)
≥10 mm	329 (52.1)

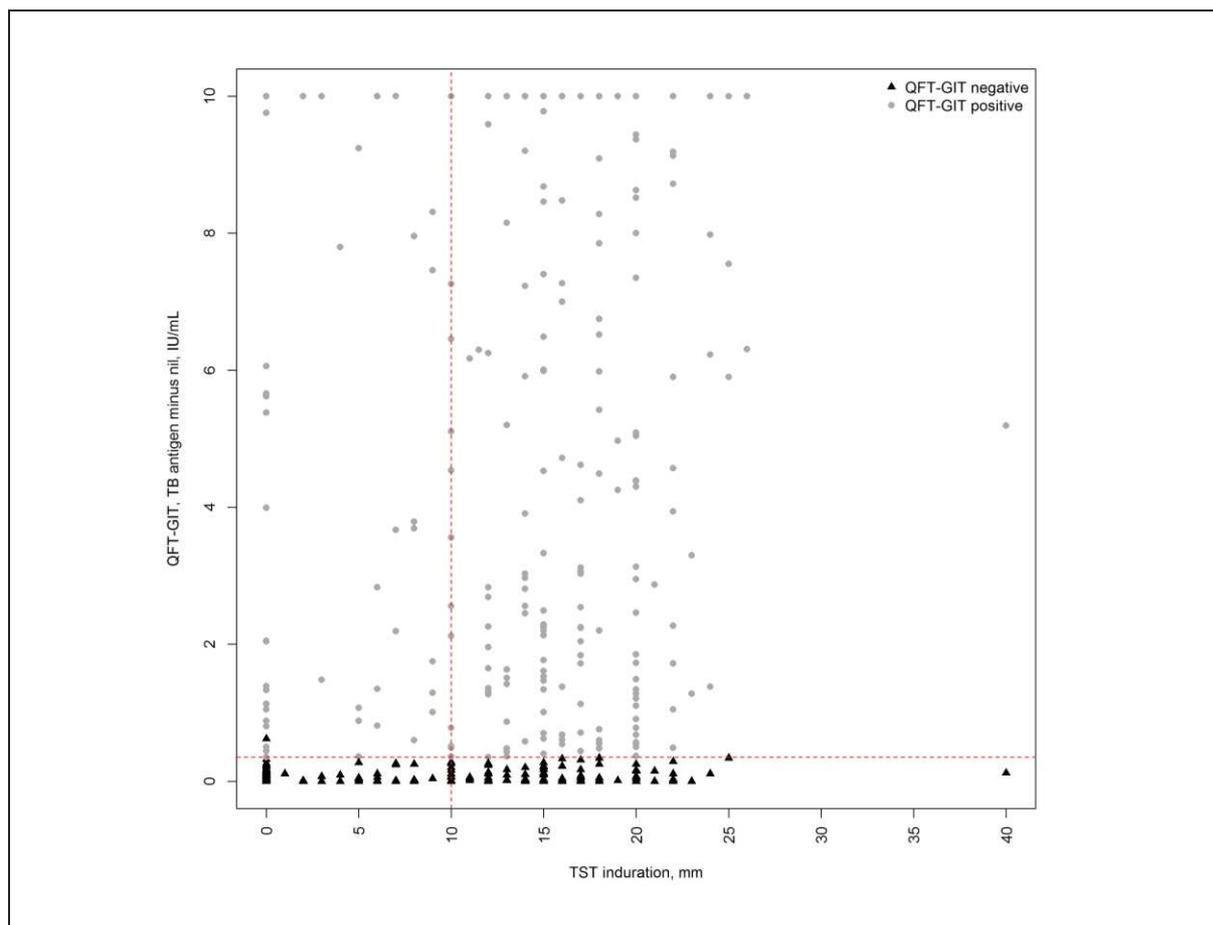


Figure 14. Nuage de points représentant les valeurs quantitatives de QFT et d’IDR des 631 participants marocains. Les lignes rouges représentent le seuil de positivité de 0.35 UI/mL pour le test QFT et de 10 mm pour l’IDR (TST).

### 2.3.2 Accord entre les résultats QFT et IDR

L’accord entre les résultats QFT et IDR était sensiblement meilleur avec un seuil de positivité de 10 mm ( $\kappa = 0.50$ ; 95%IC 0.43–0.56; concordance de 74.6 %) par rapport aux seuils de 5 mm ( $\kappa = 0.48$ ; 95%IC 0.41–0.54; concordance de 72.7%) et 15 mm ( $\kappa = 0.44$ ; 95%IC 0.37–0.51; concordance de 73.3%) (Tableau 4).

En s’intéressant d’un peu plus près aux niveaux de QFT parmi 374 sujets QFT-, la proportion de résultats QFT faiblement négatifs (0.20–0.34 UI/mL) était plus grande ( $P = 0.0005$ ) chez les sujets QFT-/IDR+ (17/116, 14.7%) que chez les sujets QFT-/IDR- (10/258, 3.9%). Parmi 257 individus QFT+, la proportion de résultats QFT faiblement positifs (0.35–0.70 UI/mL) n’était pas significativement différente entre les sujets QFT+/IDR- (5/44, 11.4%) et les double positifs QFT+/IDR+ (26/213, 12.2%).

Parmi 302 participants IDR-, la médiane du diamètre de l'induration était significativement plus haute ( $P = 3 \times 10^{-9}$ ) chez les discordants QFT+/IDR- (médiane de 3 mm) que chez les double négatifs QFT-/IDR- (médiane de 0 mm). À l'inverse, parmi 329 participants IDR+, la médiane du diamètre de l'induration était significativement plus petite ( $P = 2 \times 10^{-4}$ ) chez les discordants QFT-/IDR+ (médiane de 15 mm) que chez les doubles positifs (médiane de 17 mm).

Tableau 4. Accord entre les résultats QFT et IDR chez les 631 participants marocains

	IDR 5 mm	IDR 10 mm	IDR 15 mm
QFT-/IDR-	226	258	303
QFT+/IDR+	233	213	160
QFT-/IDR+	148	116	701
QFT+/IDR-	24	44	97
Concordance, %	72.7	74.6	73.3
Kappa (95% IC)	0.48 (0.41-0.54)	0.50 (0.43-0.56)	0.44 (0.37-0.51)

### 2.3.3 Facteurs de risque associés à l'infection par *M. tuberculosis*

En considérant le statut infectieux défini uniquement par les résultats QFT, l'analyse univariante a montré que le statut QFT+ était associé avec le sexe masculin (OR = 2.14; 95%IC 1.54–2.97,  $P = 6 \times 10^{-6}$ ), les groupes d'âge plus avancé 35–44 (OR = 1.80; 95%IC 1.23–2.63;  $P = 3 \times 10^{-3}$ ) et 45–60 ans (OR = 3.02; 95%IC 1.99–4.57;  $P = 2 \times 10^{-7}$ ), le tabagisme (OR = 2.57; 95%IC 1.59–4.15;  $P = 0.0001$ ) et un antécédent familial de tuberculose (OR = 3.16; 95%IC 1.63–6.12;  $P = 7 \times 10^{-4}$ ). Toutes ces variables restaient significativement associées en analyse multivariante (Tableau 5).

En considérant le statut infectieux défini par l'IDR (Tableau 6), les mêmes variables étaient associées avec l'infection dans l'analyse univariante; avec en plus le fait de travailler dans un service de pneumologie (OR = 2.31; 95%IC 1.21–4.40;  $P = 0.01$ ). Dans l'analyse multivariante, toutes les variables restaient associées à l'exception du tabagisme (OR = 1.27; 95%IC 0.71–2.26;  $P = 0.42$ ), probablement dû au fait que tous les fumeurs ( $n = 82$ ) étaient des hommes.

J'ai ensuite étudié le statut infectieux avec une définition plus stricte qui prenait en compte les double positifs QFT+/IDR+ ( $n = 213$ ) et les double négatifs QFT-/IDR- ( $n = 258$ ). Les mêmes facteurs de risque étaient retrouvés en analyse multivariante (Tableau 7) : sexe masculin (OR = 2.21; 95%IC 1.40–3.49;  $P = 7 \times 10^{-4}$ ), les groupes d'âge plus entre 35–44 ans (OR = 2.43; 95%IC 1.45–4.06;  $P = 7 \times 10^{-4}$ ) et 45–60 ans (OR = 4.81; 95%IC 2.72–8.52;  $P =$

$7 \times 10^{-8}$ ), antécédent familial de tuberculose (OR = 6.62; 95%IC 2.59–16.94;  $P = 8 \times 10^{-5}$ ) et travailler en service de pneumologie (OR = 3.64; 95%IC 1.44–9.23;  $P = 6 \times 10^{-3}$ ). De nouveau, fumer n'était plus associé au statut QFT+/IDR+ en incluant toutes les autres covariables (OR = 1.83; 95%IC 0.90–3.74;  $P = 0.10$ ).

Tableau 5. Facteurs de risque associés à un résultat QFT positif

Variabiles	Total n=631	QFT- n=374	QFT+ n=257	OR <sub>uni</sub> (95%IC)	P <sub>uni</sub>	OR <sub>mult</sub> (95%IC)	P <sub>mult</sub>
Sexe							
F	290	200	90	Ref		Ref	
M	341	174	167	<b>2.14 (1.54-2.97)</b>	<b>6×10<sup>-6</sup></b>	<b>1.69 (1.17-2.46)</b>	<b>6×10<sup>-3</sup></b>
Age							
18-34	297	207	90	Ref		Ref	
35-44	186	104	82	<b>1.80 (1.23-2.63)</b>	<b>0.003</b>	<b>1.57 (1.04-2.37)</b>	<b>0.03</b>
45-60	148	63	85	<b>3.02 (1.99-4.57)</b>	<b>2×10<sup>-7</sup></b>	<b>2.74 (1.74-4.31)</b>	<b>1×10<sup>-5</sup></b>
Comorbidités							
Non	586	348	238	Ref		Ref	
Oui	45	26	19	1.08 (0.58-1.99)	0.82	0.95 (0.49-1.85)	0.88
Tabagisme							
Non	549	342	207	Ref		Ref	
Oui	82	32	50	<b>2.57 (1.59-4.15)</b>	<b>1×10<sup>-4</sup></b>	<b>1.89 (1.12-3.21)</b>	<b>0.02</b>
Antécédent familial de TB							
Non	588	360	228	Ref		Ref	
Oui	43	14	29	<b>3.16 (1.63-6.12)</b>	<b>7×10<sup>-4</sup></b>	<b>3.41 (1.70-6.85)</b>	<b>6×10<sup>-4</sup></b>
Secteur de travail							
Non-clinique	278	160	118	Ref		Ref	
Chirurgie	118	79	39	0.94 (0.64-1.38)	0.74	1.40 (0.84-2.34)	0.19
Clinique	184	106	78	0.68 (0.43-1.07)	0.10	0.73 (0.41-1.27)	0.26
Pneumologie	51	29	22	1.09 (0.59-2.01)	0.77	1.80 (0.88-3.69)	0.11
Activité							
Non-médicale	167	89	78	Ref		Ref	
Paramédicale	143	80	63	0.91 (0.58-1.43)	0.68	1.00 (0.58-1.74)	0.99
Médicale	321	205	116	0.65 (0.44-1.05)	0.10	0.70 (0.41-1.19)	0.19

Tableau 6. Facteurs de risque associés à un résultat IDR positif

Variables	Total n=631	IDR- n=302	IDR+ n=329	OR <sub>uni</sub> (95%IC)	P <sub>uni</sub>	OR <sub>mult</sub> (95%IC)	P <sub>mult</sub>
Sexe							
F	290	165	125	Ref		Ref	
M	341	137	204	<b>2.10 (1.50-2.94)</b>	<b>2×10<sup>-5</sup></b>	<b>1.70 (1.15-2.51)</b>	<b>8×10<sup>-3</sup></b>
Age							
18-34	297	185	112	Ref		Ref	
35-44	186	73	113	<b>2.58 (1.74-3.84)</b>	<b>3×10<sup>-6</sup></b>	<b>2.80 (1.82-4.30)</b>	<b>3×10<sup>-6</sup></b>
45-60	148	44	104	<b>3.50 (2.25-5.45)</b>	<b>3×10<sup>-8</sup></b>	<b>3.68 (2.26-5.99)</b>	<b>2×10<sup>-7</sup></b>
Comorbidités							
Non	586	280	306	Ref		Ref	
Oui	45	32	23	0.98 (0.52-1.85)	0.95	0.82 (0.41-1.65)	0.58
Tabagisme							
Non	549	273	276	Ref		Ref	
Oui	82	29	53	<b>1.83 (1.10-3.05)</b>	<b>0.02</b>	1.27 (0.71-2.26)	0.42
Antécédent familial de TB							
Non	588	294	294	Ref		Ref	
Oui	43	8	35	<b>4.04 (1.79-9.10)</b>	<b>8×10<sup>-4</sup></b>	<b>4.89 (2.04-11.74)</b>	<b>4×10<sup>-4</sup></b>
Secteur de travail							
Non-clinique	278	144	134	Ref		Ref	
Chirurgie	118	57	61	1.07 (0.72-1.59)	0.75	1.41 (0.83-2.41)	0.21
Clinique	184	81	103	1.28 (0.81-2.02)	0.29	1.29 (0.73-2.29)	0.39
Pneumologie	51	20	31	<b>2.31 (1.21-4.40)</b>	<b>0.01</b>	<b>3.79 (1.77-8.13)</b>	<b>6×10<sup>-4</sup></b>
Activité							
Non-médicale	143	60	83	Ref		Ref	
Paramédicale	167	82	85	1.46 (0.91-2.35)	0.12	1.51 (0.84-2.71)	0.17
Médicale	321	160	161	1.02 (0.69-1.51)	0.93	1.07 (0.61-1.88)	0.81

Tableau 7. Facteurs de risque associés à un résultat QFT+/IDR+

Variables	Total n=471	QFT- IDR- n=258	QFT+ IDR+ n=213	OR <sub>mult</sub> (95%CI)	P <sub>mult</sub>
Sexe					
F	229	152	77	Ref	
M	242	106	136	<b>2.21 (1.40-3.49)</b>	<b>7×10<sup>-4</sup></b>
Age					
18-34	245	170	75	Ref	
35-44	121	56	65	<b>2.43 (1.45-4.06)</b>	<b>0.0007</b>
45-60	105	32	73	<b>4.81 (2.72-8.52)</b>	<b>7×10<sup>-8</sup></b>
Comorbidités					
Non	440	241	199	Ref	
Oui	31	17	14	0.80 (0.34-1.91)	0.62
Tabagisme					
Non	414	240	174	Ref	
Oui	57	18	39	1.83 (0.90-3.74)	0.10
Antécédent familial de TB					
Non	436	251	185	Ref	
Oui	35	7	28	<b>6.62 (2.59-16.94)</b>	<b>8×10<sup>-5</sup></b>
Secteur de travail					
Non-clinique	204	115	89	Ref	
Chirurgie	90	54	36	1.46 (0.78-2.72)	0.23
Clinique	43	73	70	0.87 (0.44-1.72)	0.68
Pneumologie	34	16	18	<b>3.64 (1.44-9.23)</b>	<b>7×10<sup>-4</sup></b>
Activité					
Non-médicale	118	61	57	Ref	
Paramédicale	107	52	55	1.31 (0.66-2.62)	0.44
Médicale	246	145	101	0.82 (0.43-1.56)	0.54

### 2.3.4 Facteurs de risque associés aux résultats QFT et IDR discordants

Finalement, je me suis intéressé aux participants présentant des résultats QFT et IDR discordants, c'est-à-dire QFT-/IDR+ (n = 116) et QFT+/IDR- (n = 44), en les comparant au groupe de référence non infecté QFT-/IDR- (n = 258) (Tableau 8).

De façon intéressante, pour le statut QFT-/IDR+, l'analyse multivariable était similaire à celle du statut IDR+, avec des associations significatives pour le sexe masculin (OR = 1.73; 95%IC: 1.02–2.93;  $P = 0.04$ ), groupes d'âge entre 35–44 (OR = 4.18; 95%IC: 2.34–7.44;  $P = 1 \times 10^{-6}$ ) et 45–60 ans (OR = 4.35; 95%IC 2.16–8.75;  $P = 2 \times 10^{-7}$ ) et travailler en pneumologie (OR = 3.95; 95%IC 1.43–10.88;  $P = 8 \times 10^{-3}$ ); toujours sans effet du tabac (OR = 0.98; 95%IC 0.40–2.38;  $P = 0.97$ ).

Enfin, pour le statut QFT+/IDR-, l'analyse multivariable était proche de celle du statut QFT+ seul, quoiqu'avec des p-values moins significatives, reflétant une taille d'échantillon plus petite. On retrouvait ainsi le sexe masculin (OR = 2.37; 95%IC 1.07–5.25;  $P = 0.03$ ), les groupes d'âge 35–44 (OR = 2.41; 95%IC 0.99–5.88;  $P = 0.05$ ) et 45–60 ans (OR = 3.17; 95%IC 1.18–8.50;  $P = 0.02$ ) et le tabagisme (OR = 3.35; 95%IC 1.17–9.60;  $P = 0.02$ ). On notait également de façon plus anecdotique un effet protecteur du travail en secteur clinique (OR = 0.17; 95%IC 0.04–0.72;  $P = 0.02$ ).

Tableau 8. Facteurs de risque associés à des résultats QFT et IDR discordants

Variables	QFT-IDT- n=258	QFT-IDR+ n=116	OR <sub>mult</sub> (95%CI)	$P_{\text{mult}}$	QFT+ IDR- n=44	OR <sub>mult</sub> (95%CI)	$P_{\text{mult}}$
Sexe							
F	152	48	Ref		13	Ref	
M	106	68	<b>1.73 (1.02-2.93)</b>	<b>0.04</b>	31	<b>2.37 (1.07-5.25)</b>	<b>0.03</b>
Age							
18-34	170	37	Ref		15	Ref	
35-44	56	48	<b>4.18 (2.34-7.44)</b>	$1 \times 10^{-6}$	17	2.41 (0.99-5.88)	0.05
45-60	32	31	<b>4.35 (2.16-8.75)</b>	$4 \times 10^{-5}$	12	<b>3.17 (1.18-8.50)</b>	<b>0.02</b>
Comorbidités							
Non	241	107	Ref		39	Ref	
Oui	17	9	0.90 (0.35-2.36)	0.84	5	1.23 (0.35-4.40)	0.75
Tabagisme							
Non	240	102	Ref		33	Ref	
Oui	18	14	0.98 (0.40-2.38)	0.97	11	<b>3.35 (1.17-9.60)</b>	<b>0.02</b>
Antécédent familial de TB							
Non	251	109	Ref		43	Ref	
Oui	7	7	1.89 (0.52-6.91)	0.33	1	1.65 (0.18-15.56)	0.66
Secteur de travail							
Non-clinique	115	45	Ref		29	Ref	
Chirurgie	54	25	1.27 (0.60-2.66)	0.53	3	0.80 (0.83-2.41)	0.68
Clinique	73	33	1.22 (0.56-2.67)	0.62	8	<b>0.17 (0.04-0.72)</b>	<b>0.02</b>
Pneumologie	16	13	<b>3.95 (1.43-10.88)</b>	$8 \times 10^{-3}$	4	1.70 (0.43-6.72)	0.46
Activité							
Non-médicale	61	28	Ref		21	Ref	
Paramédicale	52	28	1.56 (0.69-3.55)	0.29	8	0.85 (0.29-2.53)	0.78
Médicale	145	60	1.22 (0.55-2.68)	0.62	15	0.54 (0.20-1.49)	0.24

## 2.4 Discussion

Cette étude a permis d'explorer les facteurs de risque non-génétiques de l'infection tuberculeuse dans une population plus à risque que la population générale. Bien qu'il y ait de nombreuses études sur le personnel de santé, il s'agit de la première faite au Maroc, un pays où la tuberculose est moyennement endémique et où la couverture vaccinale par BCG est presque complète.

Dans ce contexte, la prévalence de l'infection allait de 40.7% (36.9–44.5%) avec QFT+ à 52.1% (48.2–56.0%) avec IDR+ et était de 45.2 % (40.7–49.5%) en considérant uniquement les résultats concordants. Ces observations sont conformes à celles reportées chez le personnel de santé dans des pays très touchés comme le Brésil (37%), l'Inde (43%) ou la Chine (54%) [143] (estimations basées sur l'IDR). La plus grande proportion de participants IDR+ comparés à ceux QFT+ dans notre étude est d'ailleurs quelque chose qui se retrouve dans une méta-analyse comparant l'efficacité des 2 tests parmi 24 études sur 25 incluses [149]. Cela peut s'expliquer en partie par la plus grande spécificité des tests IGRA et peut-être la généralisation de la vaccination BCG. L'accord entre les résultats QFT et IDR ( $\kappa = 0.50$  au seuil IDR de 10 mm) est également similaire aux estimations faites dans des pays de haute incidence ( $\kappa = 0.38$  en moyenne dans une récente méta-analyse avec des variations allant de 0.19 en Chine à 0.61 en Inde [150]).

De façon intéressante, des facteurs de risque comme le sexe masculin, un âge avancé ou un antécédent familial de tuberculose étaient associés à l'infection quelle que soit sa définition. Comme évoqué lors de l'introduction, le sexe masculin est un facteur souvent retrouvé sans qu'il n'y ait pour l'instant d'explication définitive. Dans le cadre spécifique du personnel de santé, de nombreuses études ont donc confirmé cette observation [151],[152],[153]. Dans une étude conduite à Taiwan (incluant du personnel de santé testé par IGRA), le sexe masculin était associé à l'infection en univariable mais l'association disparaissait en prenant en compte d'autres covariables [154]. La disparité hommes/femmes augmentait d'ailleurs fortement avec l'âge. En stratifiant leur analyse par sexe, les auteurs ont identifié des associations différentes, notamment le tabac chez l'homme (association non retrouvée chez les femmes). Cependant, la proportion d'infectés n'était pas différente parmi les fumeurs et les non-fumeurs. L'augmentation de la prévalence de l'infection est également largement observée, que ce soit avec des résultats positifs QFT [155],[156] ou IDR [157]. Dans notre étude, j'ai observé que l'âge était très corrélé à la durée de l'emploi ; le personnel de santé dans les hôpitaux militaires

est recruté à l'âge de jeunes adultes. Le risque accru d'infection dans les groupes d'âge 35-44 et 45-60 ans reflète probablement une exposition cumulée au cours de la vie, sans qu'il soit possible de déterminer quelle part est directement liée au travail. Également développé dans l'introduction, les facteurs de risque individuels liés à l'exposition à des cas contagieux sont très importants. Sans surprise, certains ont pu être identifiés dans mes analyses, comme le contact familial (antécédent de cas de tuberculose dans la famille) ou l'exposition professionnelle (travailler dans un service de pneumologie). L'exposition professionnelle est bien documentée [151],[155],[158]. Enfin, le tabagisme a pu être identifié comme un facteur de risque en fonction de la définition du statut d'infection. L'effet du tabac d'abord observé en univariable diminuait (statut QFT+) voire disparaissait (statut IDR+) en incluant d'autres covariables, notamment le sexe, probablement car tous les fumeurs étaient des hommes. En outre, l'effet du tabagisme est apparu plus fort pour le statut discordant QFT+/IDR- que pour le statut concordant QFT+/IDR+. Le tabagisme a déjà été reporté comme un facteur de risque d'infection, qu'elle soit définie par l>IDR ou l'IGRA, [151],[152],[159] et il a été montré que le tabac pouvait altérer la performance du test QFT [160]. Il faudra conduire des études de plus grande taille, longitudinales pour examiner les interactions entre l'infection, l'âge, le sexe, le tabagisme et d'autres facteurs socio-économiques.

L'analyse des résultats discordants entre QFT et IDR s'est montrée intéressante. Ce genre d'analyses mériterait d'ailleurs d'être approfondi, avec un focus particulier sur les valeurs proches des seuils de décision, puisqu'il n'y a pas de *gold standard* pour le diagnostic d'infection. Les résultats QFT-/IDR+ étaient plus fréquents que les QFT+/IDR- (116/631, 18.4% contre 44/631, 7.0%), ce qui avait déjà été noté dans une méta-analyse [149]. La discordance QFT-/IDR+ peut être attribuée à la vaccination BCG ou aux mycobactéries non-tuberculeuses qui rendent l>IDR faussement positive. De façon intéressante, les participants QFT-/IDR+ partageaient les mêmes facteurs de risque que ceux identifiés dans l'analyse pour les participants QFT+/IDR+. En outre, j'ai observé une grande proportion de résultats QFT faiblement négatifs parmi les QFT-/IDR+ comparé aux QFT-/IDR- (14.7% contre 3.9%), ce qui voudrait dire que certains individus pourraient présenter un résultat négatif en partie à cause de la simple variabilité des IGRA [44]. J'ai aussi observé une plus grande proportion de résultats QFT-/IDR+ dans le groupe d'âge 35-60 comparé aux 18-34 ans (79/334, 23.7% contre 37/297, 12.5%), ce qui est en phase avec l'hypothèse que, les IGRA sont plus efficaces à diagnostiquer des infections récentes [161]. Ainsi, ce *pool* d'individus contient très probablement de réels infectés mais que seuls des tests QFT n'auraient pu identifier. Concernant l'autre type de discordance QFT+/IDR-, j'ai également pu observer des facteurs de risque similaires à ceux

trouvés avec le phénotype QFT+ seul. Les individus discordants QFT+/IDR- présentaient par ailleurs des réactions IDR plus larges que les individus QFT-/IDR- et il n'est pas non plus à exclure que certains d'entre eux présentaient une anergie cutanée expliquant l'IDR négative [162]. Ainsi, de la même manière que le groupe QFT-/IDR+, il est fort probable que le groupe QFT+/IDR- contienne de nombreux individus réellement infectés.

En conclusion de ce premier travail réalisé chez du personnel soignant marocain, une forte prévalence a pu être mise en évidence, confirmant le risque professionnel plus élevé qu'en population générale. On aura pu observer comme les 2 types de tests IDR et QFT ne sont pas complètement concordants même si les facteurs de risque associés à leur positivité sont assez similaires (et étaient attendus). D'un point de vue de santé publique, il n'existe pas de politique particulière pour le personnel de santé. Un dépistage pourrait être justifié afin d'identifier des infections qui seraient récentes et qui pourraient progresser rapidement en maladie mais le faire de façon systématique ou périodique semble inadapté. En revanche, avoir un test de référence à l'embauche (ici IDR ou QFT pourraient convenir), appliquer des mesures de prévention efficaces (isolement des patients à risque, ventilation, port de masques) et cibler le dépistage du personnel plus exposé (poste de travail, existence de facteurs de risque, documentation de tous les contacts directs) semble plus approprié.

# **3 Épidémiologie génétique : Analyse d'association pangénomique de la résistance à l'infection par *M. tuberculosis* dans 3 populations distinctes**

## **3.1 Introduction**

Dans cette partie, je me propose d'explorer les facteurs de risque génétiques associés à la résistance à l'infection par *Mtb*. La solidité de ce travail repose sur 2 points essentiels, (i) la définition d'un phénotype très spécifique et homogène et (ii) des individus intensément exposés à *Mtb*. Comme je l'ai évoqué dans l'introduction puis observé dans la première partie, les tests IDR et IGRA ne sont pas complètement concordants et souffrent de nombreuses limitations. Pour éviter toute discussion sur quel test choisir ou quel seuil définir, le phénotype d'infection a été ici déterminé en prenant en compte les résultats des 2 tests effectués sur les participants. En outre, ceux-ci ont été recrutés (i) chez des contacts familiaux qui ont côtoyé un patient tuberculeux contagieux dans le foyer ou bien (ii) en zone hyper-endémique où l'exposition à *Mtb* est très élevée, même dehors du foyer.

## **3.2 Description des cohortes**

### **3.2.1 Vietnam**

Le Vietnam est un pays à revenus intermédiaires d'Asie du Sud-Est avec une incidence annuelle de tuberculose élevée au moment de l'étude, environ 130 pour 100 000 habitants par an [163]. De 2010 à 2015, des collaborateurs de l'Université McGill ont entrepris de recruter

avec leurs collègues vietnamiens des adultes atteints d'une tuberculose pulmonaire. Ces cas index présentaient une toux chronique (en moyenne de 3 mois) avant le début d'un traitement, et le diagnostic de tuberculose pulmonaire a pu être fait à partir de la présentation clinique, d'une radiographie pulmonaire et d'une culture positive pour *Mtb*. À partir de 466 cas index, un total de 1108 contacts familiaux âgés de plus de 18 ans ont été invités à participer à l'étude. Leur statut d'infection a été déterminé à la fois par (i) une IDR (test classique de Mantoux: une dose standard de 5 unités de tuberculine (0.1 mL, Tubertest, Sanofi Pasteur, Lyon, France) injectée dans l'avant-bras des participants avec une lecture à 48-72h par un personnel entraîné) et (ii) un IGRA (QuantiFERON-TB Gold In-Tube test (QFT) réalisé sur des échantillons sanguins, prélevés avant l'IDR).

### 3.2.2 France

Comme première cohorte de réplique, j'ai pu compter sur les données d'une étude de contacts familiaux établie dans le Val-de-Marne, en périphérie de Paris. Cette cohorte a déjà fait l'objet d'analyses dans le laboratoire, notamment une analyse de liaison mettant en évidence 2 loci sur les chromosomes 3 et 8 contrôlant la production d'IFN- $\gamma$  [140]. Le Val-de-Marne est une région de faible endémie (environ 22 nouveaux cas de tuberculose pour 100 000 habitants par an) [164]. En France, la tuberculose est obligatoirement déclarée auprès de l'Agence Régionale de Santé et signalée au Centre de Lutte Anti-Tuberculeuse au niveau départemental. Le CLAT est notamment chargé d'enquêter autour du cas signalé (à la recherche d'éventuels cas secondaires ou bien du vrai cas index) et d'évaluer le risque environnemental et de transmission (proximité des contacts familiaux, temps passé au contact du sujet infecté, environnement confiné etc.). Ainsi, 664 contacts familiaux liés à 132 cas tuberculeux avaient été recrutés et suivis sur 2 visites de dépistage, conformément aux recommandations en vigueur. À la première visite (V1), les contacts étaient soumis à un examen physique, une radiographie thoracique ainsi qu'aux 2 tests IDR et IGRA. Si les IDR avaient été réalisées de la même façon qu'au Vietnam, des IGRAs réalisés à partir de 3 différentes stimulations (ESAT-6, PPD ou BCG) étaient disponibles. Les IGRAs déterminant la production d'IFN- $\gamma$  à partir de la stimulation par ESAT-6 sur du sang périphérique ont été retenus car ils sont plus spécifiques à *Mtb* et équivalents au QuantiFERON-TB Gold In-Tube QFT (fait à partir de stimulations par ESAT-6, CFP10 et TB7.7). Une seconde visite (V2) pouvait avoir lieu 8 à 12 semaines plus tard, si le diagnostic d'infection n'était pas retenu à V1, et une seconde IDR était réalisée.

### 3.2.3 Afrique du Sud

Comme seconde cohorte de réplication, j'ai pu m'appuyer sur les données recueillies dans un échantillon de 153 familles sud-africaines, résidant en banlieue du Cap. Cette cohorte, constituée d'individus dits « Coloured », issus d'un brassage génétique complexe [165], a aussi fait l'objet d'analyses dans ce laboratoire, notamment la mise en évidence des loci contrôlant l'IDR sur les chromosomes 5 et 11 [137]. Cette cohorte n'est pas une cohorte de contacts familiaux à proprement parler et les individus n'étaient pas spécifiquement définis comme des contacts d'un patient tuberculeux. Il faut dire que la tuberculose y était (et l'est toujours) particulièrement présente, à l'époque du recrutement, l'incidence frôlait les 1 000 pour 100 000 [166]; et dans ce contexte, la transmission de la tuberculose est surtout communautaire, en dehors des ménages [167]. Le recrutement initial ciblait majoritairement des enfants/adolescents dans le but de reconstituer de larges familles puis d'en extraire des familles nucléaires pour l'analyse de liaison originelle [137]. Un sous-ensemble a par la suite été génomisé. Les participants furent testés à la fois par IDR et IGRA. Comme en France, les IGRAs ont été réalisés après stimulation des échantillons de sang total par ESAT-6.

## 3.3 Génomique, contrôle qualité et imputation

Un total de 724 Vietnamiens et 573 Français ont été génomisés avec une puce Illumina Infinium OmniExpressExome-8-v1 couvrant près de 1M de polymorphismes; 374 Sud-Africains ont quant à eux été génomisés avec une puce Illumina HumanOmni2.5-8 BeadChip, couvrant environ 2M de SNPs. Les 2 puces ne sont pas très différentes hormis pour le nombre absolu de marqueurs couverts ; la Omni2.5 BeadChip ayant l'avantage de couvrir ~65% des marqueurs (dont l'allèle mineur a une fréquence > 5%) d'un génome africain Yoruba du panel de référence 1000 Génomes (1000G) contre 43% pour l'OmniExpressExome [168]. Le principe du génomage reste le même. Brièvement, une lame va servir de support à des puces ; chaque puce contenant des millions de puits (*wells*) ; chaque puits contenant une bille (*bead*). Chaque SNP est représenté par plusieurs puits contenant la même bille, sur laquelle il y a plusieurs copies d'une sonde de 50 nucléotides (*probe*), spécifique d'un endroit du génome. La sonde reconnaît la séquence juste en amont du SNP. Les 2 allèles de la séquence d'ADN non marqué s'hybrident alors sur la même sonde. Le nucléotide de l'ADN hybridé à la sonde est recopié au

bout de la sonde avec un marqueur fluorescent ; 2 marqueurs, un rouge et un vert, sont utilisés. Les intensités des 2 types de fluorescence vont déterminer le génotype d'un individu donné à un locus donné. L'assignation d'un génotype pour un ensemble d'individus génotypés se fait de manière graphique par la représentation des intensités lumineuses, après normalisation; en particulier, les individus avec un même génotype vont s'agréger et 3 types de *clusters*, représentant les 3 génotypes possibles, seront observables (Figure 15). Les génotypes ont été déterminés par l'algorithme GenTrain dans l'environnement logiciel GenomeStudio d'Illumina puis convertis en format PLINK pour permettre toutes les analyses suivantes et notamment un contrôle qualité des données. Un échec du génotypage d'un marqueur peut se manifester de plusieurs façons : (i) la répartition des intensités lumineuses peut aboutir à une définition peu précise des groupes, et donc un taux de génotypage faible, (ii) certaines valeurs extrêmes d'intensité peuvent tromper l'algorithme et aboutir à la fusion du groupe hétérozygote et d'un groupe homozygote ; l'excès d'hétérozygotes provoquant un écart à l'équilibre de Hardy-Weinberg, ou (iii) les marqueurs ayant une faible fréquence de l'allèle mineur sont caractérisés par l'absence fréquente du groupe des homozygotes pour l'allèle mineur. On s'attache donc à regarder de près le taux de génotypage, la fréquence de l'allèle mineur (MAF pour *minor allele frequency*) ou un écart majeur à l'équilibre de Hardy-Weinberg; les seuils exacts de ces filtres étant déterminés par l'observation des distributions. Au niveau des individus, on contrôle également la proportion de génotypes identiques par état (IBD pour *identity by descent*) pour toutes les paires d'échantillons, à la recherche d'individus dupliqués ainsi que la correspondance entre sexe déclaré et inféré génétiquement, représentant un critère de qualité générale pour l'ensemble de l'échantillon.

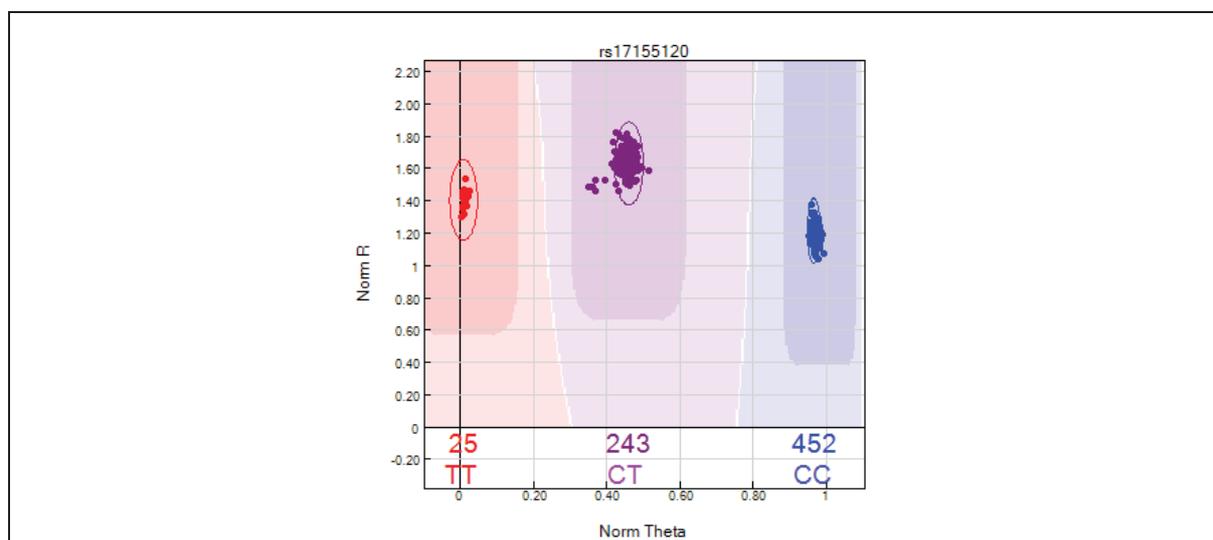


Figure 15. Assignation correcte d'un marqueur génotypé dans la cohorte vietnamienne.

Ainsi, pour chaque cohorte, ont été gardés les marqueurs avec une MAF > 1%, un taux de génotypage > 99% et une *p-value* du test de l'écart par rapport à la loi de Hardy-Weinberg >  $1 \times 10^{-5}$ . Les individus présentant un taux de génotypage < 95% (n = 2) ou dupliqués (n = 1, correspondant au membre d'une paire de duplicata avec le plus faible taux de génotypage) ont aussi été exclus. Au final, ont été utilisés pour l'imputation : 720 individus et 598 090 marqueurs pour la cohorte Vietnam, 573 individus et 886 471 marqueurs pour la cohorte France et 374 individus et 1 347 846 marqueurs pour la cohorte Afrique du Sud.

L'imputation de génotypes consiste à inférer le génotype de marqueurs non directement génotypés à partir d'un panel d'haplotypes de référence extrêmement dense en marqueurs. Cette étape incontournable des GWASs permet d'augmenter considérablement le nombre de marqueurs testés et se réalise en 2 temps. Tout d'abord, un phasage permet d'établir les haplotypes des individus de nos cohortes. Ces haplotypes sont reconstitués par des méthodes statistiques basées sur des chaînes de Markov cachées et des algorithmes d'estimation itératifs de type espérance-maximisation implémentées dans différents logiciels. Puis, au moment de l'imputation à proprement parler, les génotypes sont imputés sous la forme d'un vecteur de probabilité estimant la probabilité que l'individu soit hétérozygote ou homozygote pour l'allèle mineur/de référence sachant ses génotypes connus voisins et connaissant les haplotypes de référence. Là encore, plusieurs méthodes existent. D'une manière générale, la qualité de l'imputation va dépendre du software utilisé, des marqueurs couverts initialement par la puce de génotypage et surtout du panel de référence, en termes de taille et de représentativité génétique [169].

Dans le cas des SNPs imputés, la contribution de chaque individu sera représentée par (i) un vecteur de 3 probabilités postérieures, par exemple pour un marqueur AB, {Probabilité du génotype AA,  $P_{AA}$  ; Probabilité du génotype AB,  $P_{AB}$  ; Probabilité du génotype BB,  $P_{BB}$ }, (ii) le génotype avec la plus haute probabilité (ou *best-guess*) et (iii) le dosage de ce SNP, correspondant à  $2 * P_{AA} + 1 * P_{AB} + 0 * P_{BB}$  (en prenant l'allèle A comme référence). La qualité de l'imputation est classiquement évaluée par un score d'information, qui correspond au coefficient de corrélation  $r^2$  entre le vrai dosage d'un allèle et celui estimé pour tous les individus imputés ; l'avantage du  $r^2$  étant qu'on peut l'estimer à partir des probabilités postérieures [169]. Des variantes existent, comme le  $R^2$  allélique qui se calcule avec le *best-guess* ou le '*info score*' calculé par IMPUTE2 à partir d'une estimation de la fréquence allélique. Ces estimations sont très similaires (voire équivalentes, par exemple  $r^2$  et *info score* sont égaux sous Hardy-Weinberg). Le seuil retenu varie d'une étude à l'autre, descendant parfois jusqu'à 0.4. Si les génotypes d'une population européenne sont imputés avec une très

grande qualité à partir d'un panel de référence comme 1000 Génomes [170], la précision peut diminuer chez des populations moins représentées, typiquement africaines ou métisses. La Figure 16 montre que l'imputation de la cohorte sud-africaine est de meilleure qualité avec le panel de référence African Genome Resources qui contient ~2400 individus africains en plus du panel 1000G Phase 3. Le phasage et l'imputation sont par ailleurs des processus très lourds au niveau computationnel.

Mon choix s'est finalement porté vers l'utilisation de serveurs web (Michigan [171] et Sanger [172]) qui permettent, d'une part, une imputation rapide (les données sont prêtes en quelques jours voire quelques heures au lieu d'1 à 2 semaines localement) et, d'autre part, de choisir un panel de référence plus adapté aux cohortes. Ainsi, l'imputation a été réalisée sur le serveur Michigan (panel de référence 1000G Phase 3, phasage avec le logiciel Eagle2 et imputation avec Minimac3) sur 720 individus / 598 090 SNPs pour le Vietnam et 573 individus / 886 471 SNPs pour la France. Pour l'Afrique du Sud (374 individus / 1 347 846 SNPs), l'imputation a été réalisée sur le serveur Sanger (panel de référence African Genome Resources, phasage avec Eagle 2 et imputation à partir de méthodes *Positional Burrows-Wheeler Transform*). Au final, les SNPs avec un score de qualité d'imputation  $> 0.8$  et une MAF  $> 5\%$  ont été retenus pour les analyses à suivre (5 591 951 marqueurs Vietnam, 7 737 070 marqueurs France et 6 922 541 marqueurs Afrique du Sud).

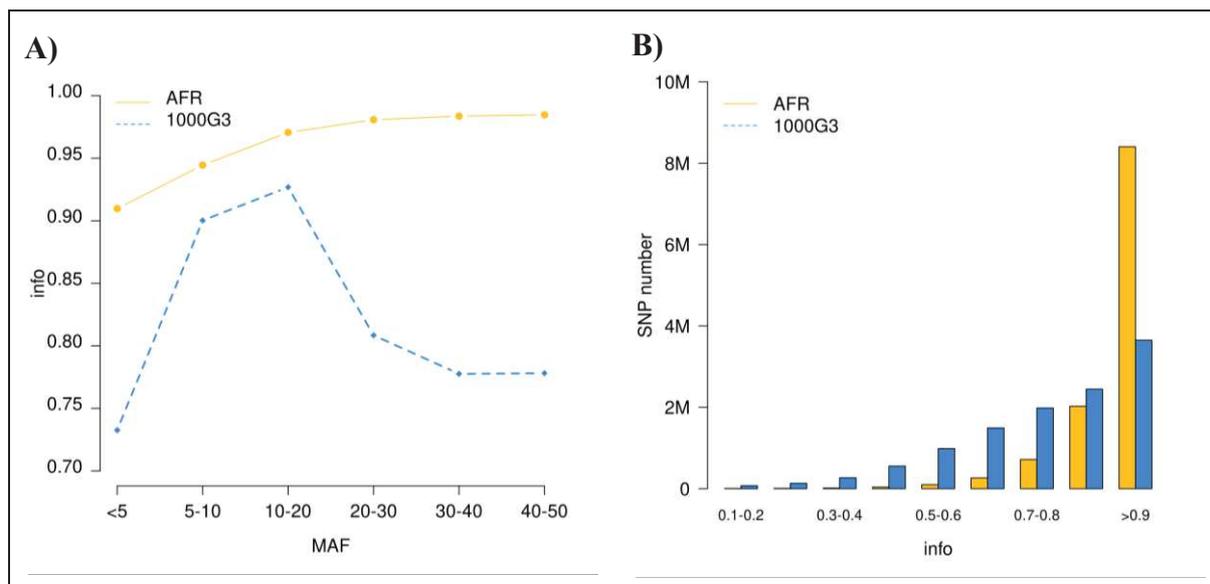


Figure 16. Comparatif de la qualité d'imputation dans la cohorte d'Afrique du Sud en utilisant comme panels de référence 1000G Phase 3 (bleu) ou le panel African Genome Resources disponible sur le serveur Sanger (jaune). La figure A) montre le nombre de marqueurs imputés par bin de  $r^2$  (le score info de qualité) ; la figure B) montre ce  $r^2$  en fonction de la fréquence de l'allèle mineur. On observe une différence sensible avec beaucoup plus de marqueurs imputés avec une très haute qualité ( $> 0.9$ ) avec le panel AFR et en moyenne le score info reste très élevé quelle que soit la fréquence allélique ; plus étonnant, avec le panel 1000G Phase 3, l'info a même tendance à diminuer à mesure que la fréquence augmente (passant de  $\sim 0.9$  en moyenne entre 5 et 20% à  $\sim 0.8$  entre 20 et 50%).

La prise en compte des structures génétiques des populations est devenue une question centrale dans les études génétiques d'association. En effet, la stratification de population, c'est-à-dire des différences dans les fréquences alléliques entre les groupes ou sous-groupes d'une étude, peut rapidement mener à une inflation de l'erreur de type I et donc à observer des fausses associations ou bien à une diminution de la puissance [173],[174]. Les structures génétiques sont d'autant plus difficiles à contrôler quand des structures familiales voire une parenté cryptique sont présentes. Il est ainsi classique et nécessaire de faire le point sur la présence de structures génétiques au sein des cohortes étudiées. Cela étant dit, dans le cadre de notre étude, l'apparentement (déjà connu) entre les individus nous orientera pour le choix de la méthode statistique à employer et que je détaillerai un peu plus loin, le modèle mixte.

L'analyse en composantes principales (ACP) d'une matrice mesurant la similarité génétique entre individus est la méthode de choix pour observer la structure génétique des cohortes à l'échelle populationnelle (mondiale). Pour résumer [175], la ressemblance génétique de  $n$  individus sur  $m$  marqueurs (matrice  $X$  de dimension  $m \times n$  des génotypes centrés et réduits), une réduction de dimensions est obtenue en estimant la covariance des génotypes pour chaque

paire d'individu (matrice de variance-covariance  $\Sigma$ , proportionnelle à  $X^T X$ , de dimension  $n \times n$ ). Deux individus, dont les génotypes sont proches, auront une valeur de covariance fortement positive et inversement. La matrice  $X^T X$  peut ensuite être reformulée comme le produit d'une matrice  $V$  de  $n$  vecteurs propres indépendants et d'une matrice diagonale  $S^2$  de  $n$  valeurs propres ou *eigenvalues* ( $X^T X = VS^2V^T$ ). La matrice  $V$  contient les coordonnées de chaque individu sur les  $n$  composantes principales et la matrice  $S^2$  contient l'importance relative des différentes composantes. La matrice d'origine  $X$  peut elle-même être exprimée comme le produit  $USV^T$ , avec  $U$  la matrice de dimension  $m \times n$  des coordonnées de chaque marqueur sur les  $n$  composantes principales (appelés poids ou *loadings* du marqueur sur chaque composante principale). Cette ACP ou décomposition en valeur singulière de la matrice  $X$ , recherche l'axe de projection qui conserve au maximum la dispersion des données (première composante principale), puis d'un second axe de projection orthogonal conservant au maximum la dispersion des données (seconde composante principale) et ainsi de suite. Il s'agit d'un changement de base dans un espace à  $n$  dimensions, tel que les premiers axes soient les plus informatifs. Chaque individu, initialement caractérisé par  $m$  génotypes résumés en  $n$  covariances, est maintenant caractérisé par  $n$  coordonnées sur  $n$  composantes principales, combinaisons linéaires des covariances initiales (elles-mêmes combinaisons des génotypes initiaux). Les composantes principales sont ordonnées en fonction de leur valeur propre, une mesure de la proportion de la dispersion initiale qu'elles capturent. La visualisation des 2 premières composantes principales peut déjà être très informative sur la structure d'un échantillon, notamment en le fusionnant avec des populations de référence comme 1000G Phase 3. À une échelle locale, la présence de quelques individus fortement reliés, comme les membres d'une même grande famille, peut parfois accaparer les premières composantes principales et il est préférable d'identifier a priori les individus apparentés entre eux pour n'en choisir qu'un seul puis réaliser l'ACP sur ce sous-échantillon des individus indépendants.

L'ACP a été ici réalisée pour chaque cohorte à partir des génotypes des individus de 1000G Phase 3 et de ceux des cohortes Vietnam (Figure 17); France et Afrique du Sud (Figure 18). La cohorte vietnamienne, établie chez des individus appartenant au groupe des Kinh, se regroupe de façon très homogène avec les populations est-asiatiques de 1000G, et notamment les Kinh de 1000G, comme attendu. La cohorte française, au contraire, montre une certaine hétérogénéité à l'échelle continentale avec une majorité d'individus d'ascendance européenne, nord-africaine et sub-saharienne, que l'on observe avec un long gradient nord-sud. Enfin, la cohorte sud-africaine est plus homogène mais forme un groupe distinct à côté des populations africaines de 1000G, où les « Coloured » ne sont pas représentés.

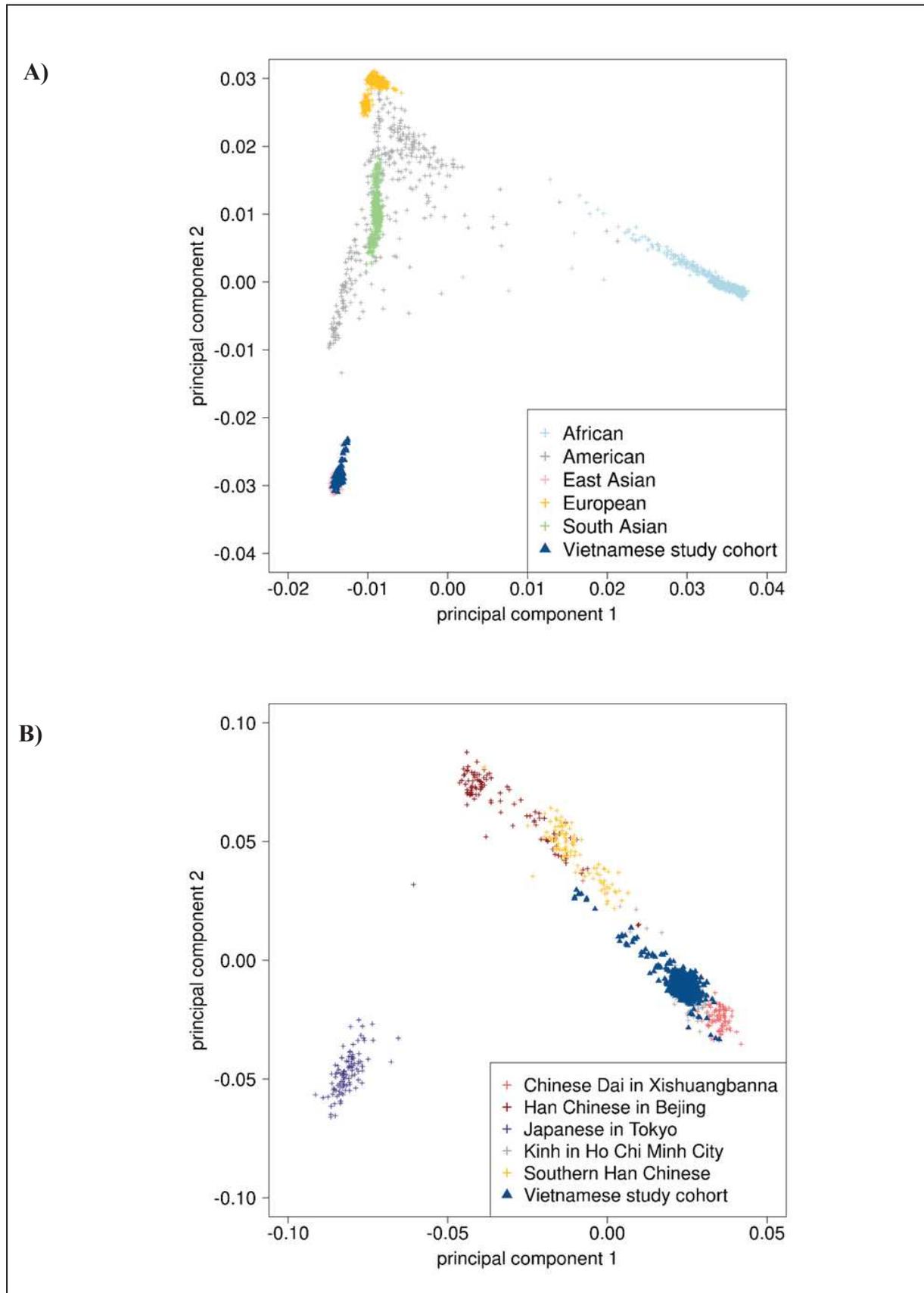


Figure 17. Analyse en composantes principales des 720 individus de la cohorte Vietnam (en bleu foncé) avec les grandes populations de 1000G Phase 3 (A) et leurs sous-populations d'Asie de l'Est (B).

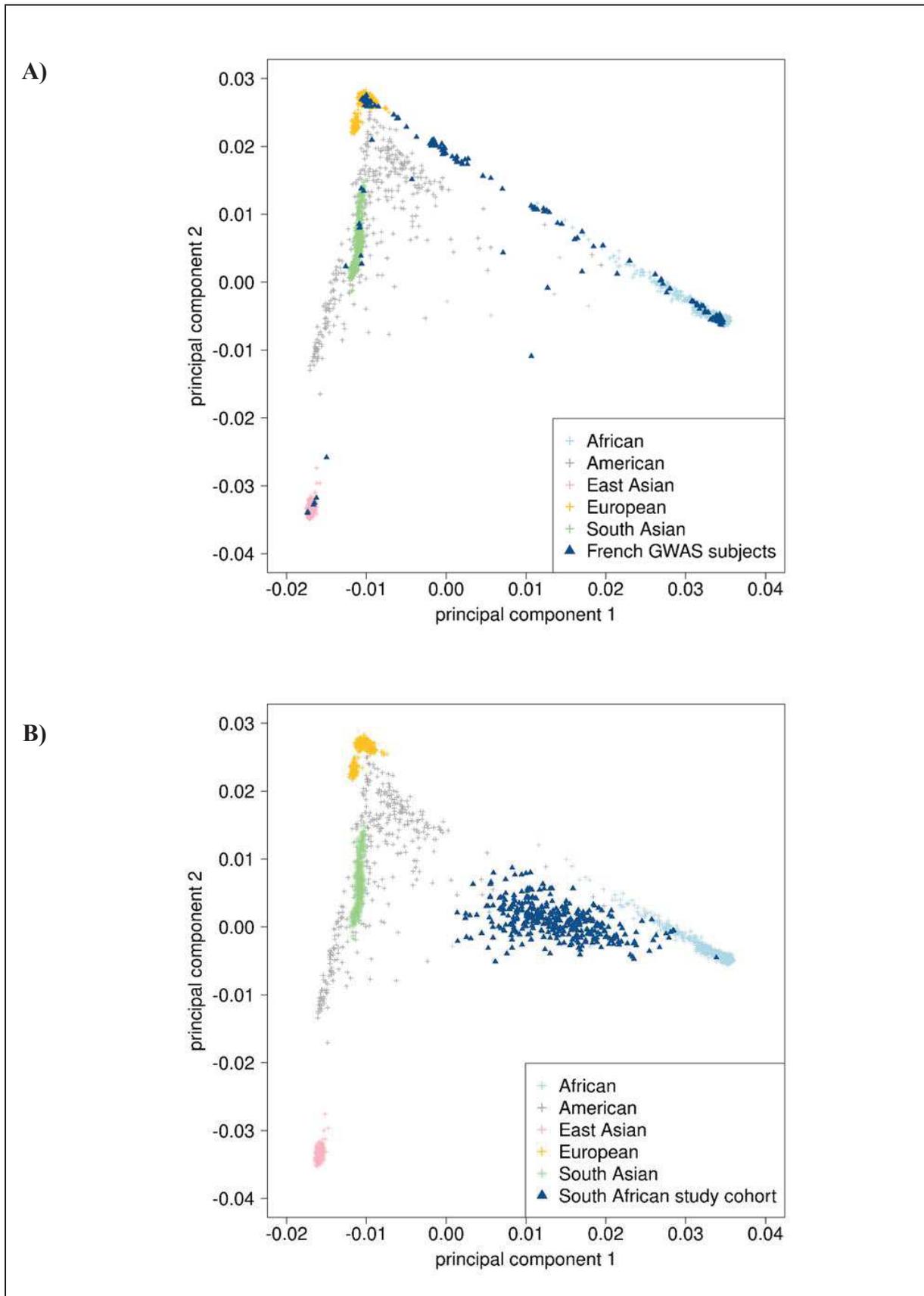


Figure 18. Analyse en composantes principales des 157 individus de la cohorte du Val-de-Marne, France retenus pour la GWAS (A) et des 374 de la cohorte Afrique du Sud (B) avec les grandes populations de 1000G Phase 3.

D'autres méthodes se sont développées pour inférer la structure génétique d'une population, notamment en déterminant des clusters génétiques qui sont définis comme des groupes, divergents génétiquement, d'individus issus de la même population ancestrale et en calculant la probabilité pour chaque individu d'appartenir à un cluster [176],[177]. Comme un individu métisse peut appartenir à plusieurs clusters, l'objectif est aussi d'inférer les contributions relatives de chaque population ancestrale distincte à un génome métisse. Ces méthodes d'apprentissage non-supervisé reposent en général sur un modèle de vraisemblance (le modèle d'*admixture*) et plus précisément des méthodes bayésiennes qui appliquent des estimations de Monte-Carlo par chaînes de Markov. Ces processus commencent par assigner aléatoirement les individus à un nombre pré-déterminé de clusters  $K$  (le nombre supposé de populations ancestrales) puis les fréquences alléliques sont estimées dans chaque cluster et les individus ré-assignés à partir des nouvelles estimations des fréquences alléliques. Ces processus sont répétés des milliers de fois pour converger vers des estimations fiables des fréquences alléliques au sein de chaque population et des probabilités pour chaque individu d'appartenir à un groupe. Les modèles d'*admixture* font de nombreuses hypothèses sur l'histoire démographique des individus analysés (les individus sont censés être issus de  $K$  populations ancestrales homogènes) et le choix optimal du nombre de clusters  $K$  est critique avec là aussi de nombreuses stratégies pour l'estimer, allant de la simple estimation par connaissance de l'histoire, à une validation croisée comme le propose le logiciel ADMIXTURE que j'utiliserai ici [177]. L'histoire génétique des populations peut ainsi être reconstruite finement, particulièrement celles issues d'un métissage récent entre 2 populations très différentes (la population afro-américaine par exemple). Mais d'une manière générale, il faut faire preuve de beaucoup de précaution dans l'interprétation des données résultant de ces modèles aussi bien d'un point de vue statistique [178] que historique et éthique.

Dans le contexte des maladies complexes, et infectieuses en particulier, l'inclusion de populations diverses, métisses, permet d'explorer les liens entre mouvements des populations et adaptation à de nouveaux environnements ou pathogènes qui conduisent à différents niveaux de susceptibilité à certaines maladies, par exemple ascendance africaine protectrice contre la dengue chez des Cubains [179], ou ascendance péruvienne native associée à un risque accru de tuberculose [180]. Pour cette thèse, de telles approches d'*admixture-mapping* n'étaient pas adaptées notamment à cause de l'homogénéité de la cohorte principale (population vietnamienne Kinh) et la présence de données familiales. Dans la continuité de l'ACP, on peut néanmoins explorer par curiosité la structure génétique des 3 cohortes et du panel 1000G Phase 3 réalisée à partir d'un groupe d'individus indépendants et des génotypes filtrés et élagués

(Figure 19, Figure 20, Figure 21). En supposant 2 populations ancestrales, on peut déjà observer une distinction entre populations africaines et les autres populations de 1000G, reflétant le « *out-of-Africa bottleneck* ». Avec 5 clusters pré-définis, la distinction entre les grandes populations continentales est nette et l'on peut constater que la cohorte vietnamienne reste très homogène. On peut retrouver dans la cohorte française le gradient observé sur l'ACP par cette démarcation « diagonale », reflétant des origines européennes, nord-africaines et subsahariennes. Enfin, on note clairement le métissage des individus sud-africains avec l'apparition de 4 groupes correspondant aux ascendances africaines (la partie violette), européenne (en rouge), sud-asiatique (en bleu) et est-asiatique (en vert). Avec d'autres populations africaines en référence, on pourrait affiner la composante africaine de ces *Cape-townians*.

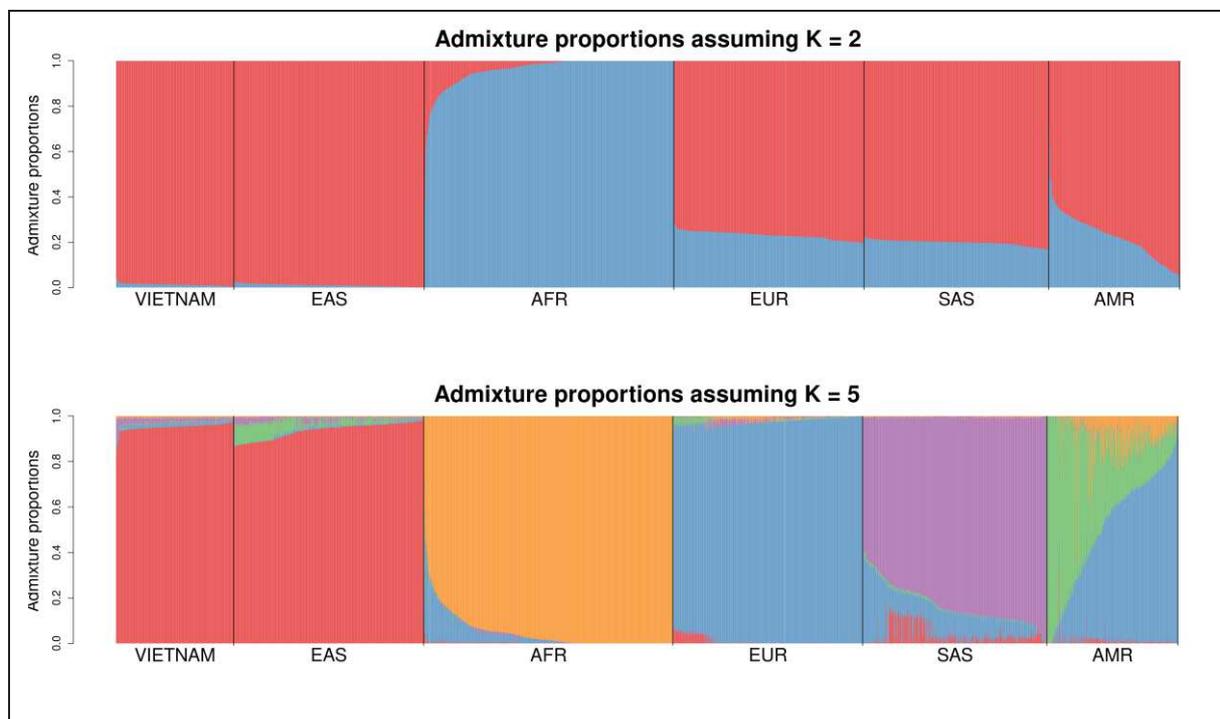


Figure 19. Proportions des différentes populations ancestrales ( $K = 2$  ou  $5$ ) chez les participants du Vietnam et 1000G Phase 3.

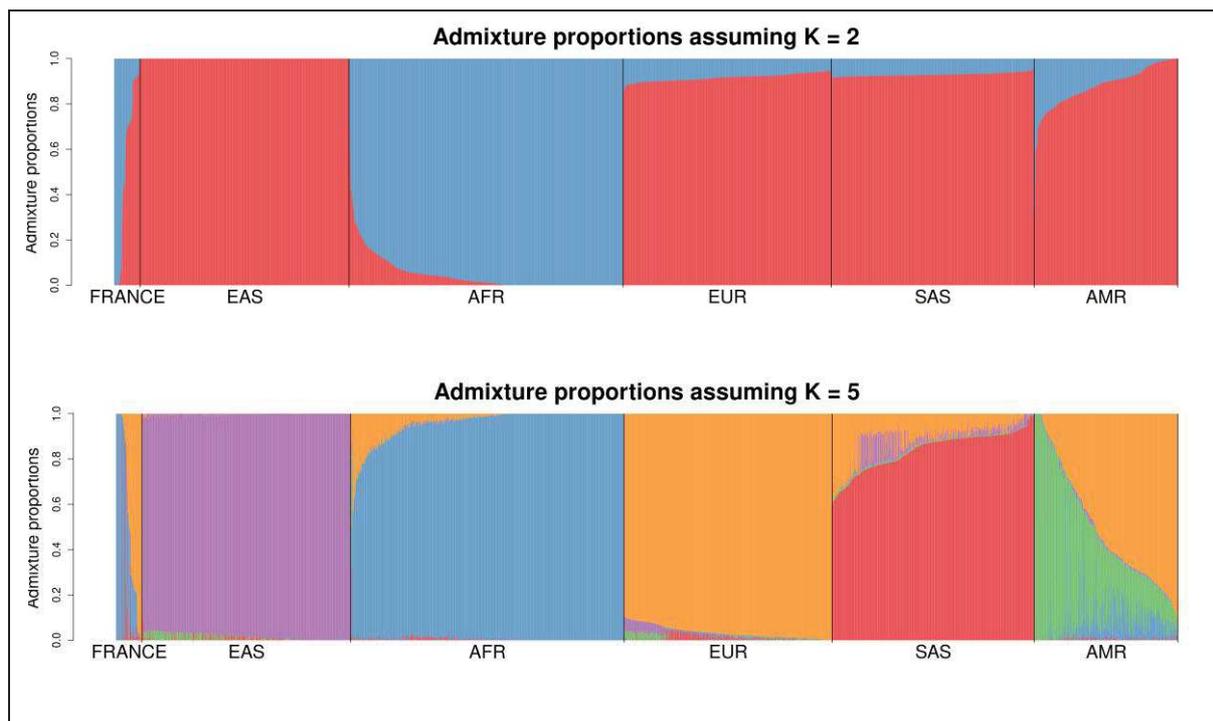


Figure 20. Proportions des différentes populations ancestrales ( $K = 2$  ou  $5$ ) chez les participants de la cohorte du Val-de-Marne, France et 1000G Phase 3.

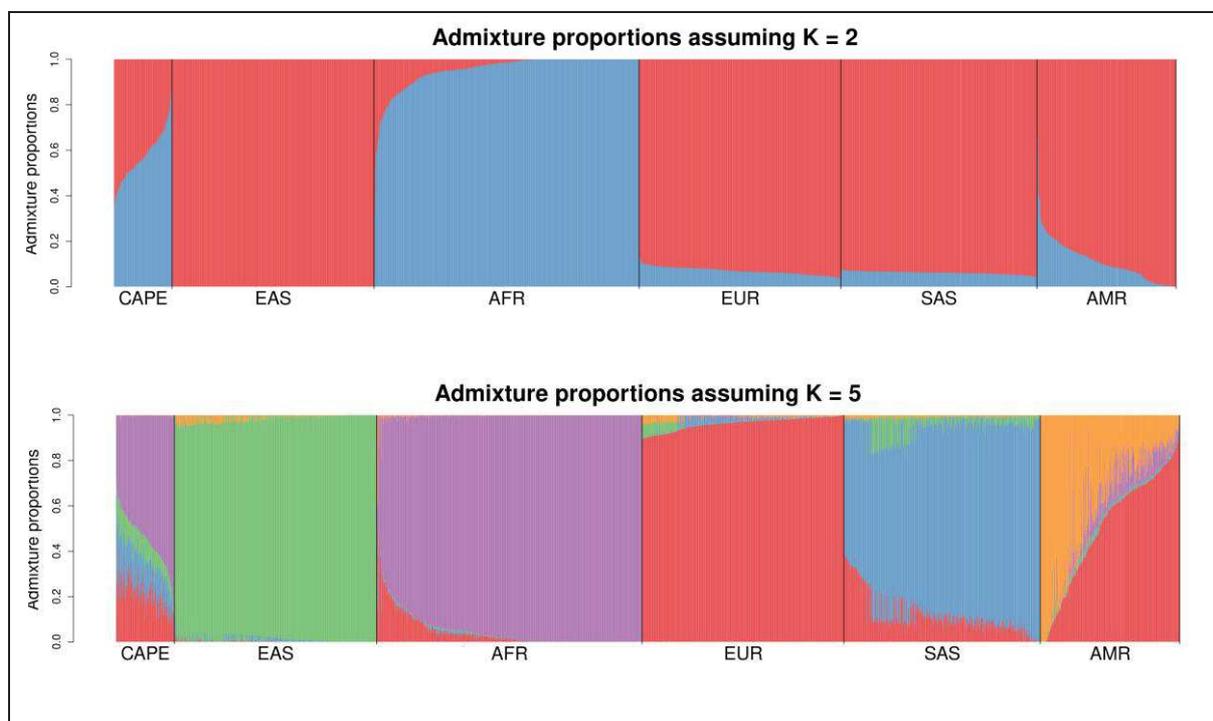


Figure 21. Proportions des différentes populations ancestrales ( $K = 2$  ou  $5$ ) chez les participants de la cohorte du Cap, Afrique du Sud et 1000G Phase 3.

Après cette visualisation des structures génétiques globales des cohortes, il est intéressant d'observer les liens familiaux dans chaque échantillon lors du contrôle qualité. Dans notre étude, les apparentements sont connus (les liens familiaux étaient même renseignés) et aucun lien cryptique ne viendra perturber par la suite l'étude d'association. Tout d'abord le logiciel PLINK permet d'inspecter les inconsistances mendéliennes, indiquant des erreurs de génotypage lorsque les génotypes des parents et des descendants sont connus (certaines transmissions d'allèles sont en effet impossibles). En outre, il est possible de vérifier certaines inconsistances des liens familiaux reportés. Ainsi, au Vietnam, une vingtaine d'individus assignés à une famille étaient en réalité indépendants. Toujours à partir des génotypes, PLINK permet d'estimer les coefficients IBD partagés par chaque paire d'individus d'un échantillon et notamment  $Z_0$ ,  $Z_1$  et  $Z_2$ .  $Z_0$  est la probabilité qu'à un locus donné, 0 allèle n'est IBD; autrement dit, une paire d'individus indépendants aura un  $Z_0$  proche de 1. Une paire parent/enfant aura théoriquement  $\{Z_0, Z_1, Z_2\} = \{0, 1, 0\}$ , et une paire de frères/soeurs aura  $\{Z_0, Z_1, Z_2\} = \{1/4, 1/2, 1/4\}$ , soit 25% de loci avec 0 allèle IBD, 50% avec 1 allèle IBD et 25% avec 2 allèles IBD. Les structures familiales pour chaque cohorte sont visualisées dans la Figure 22.

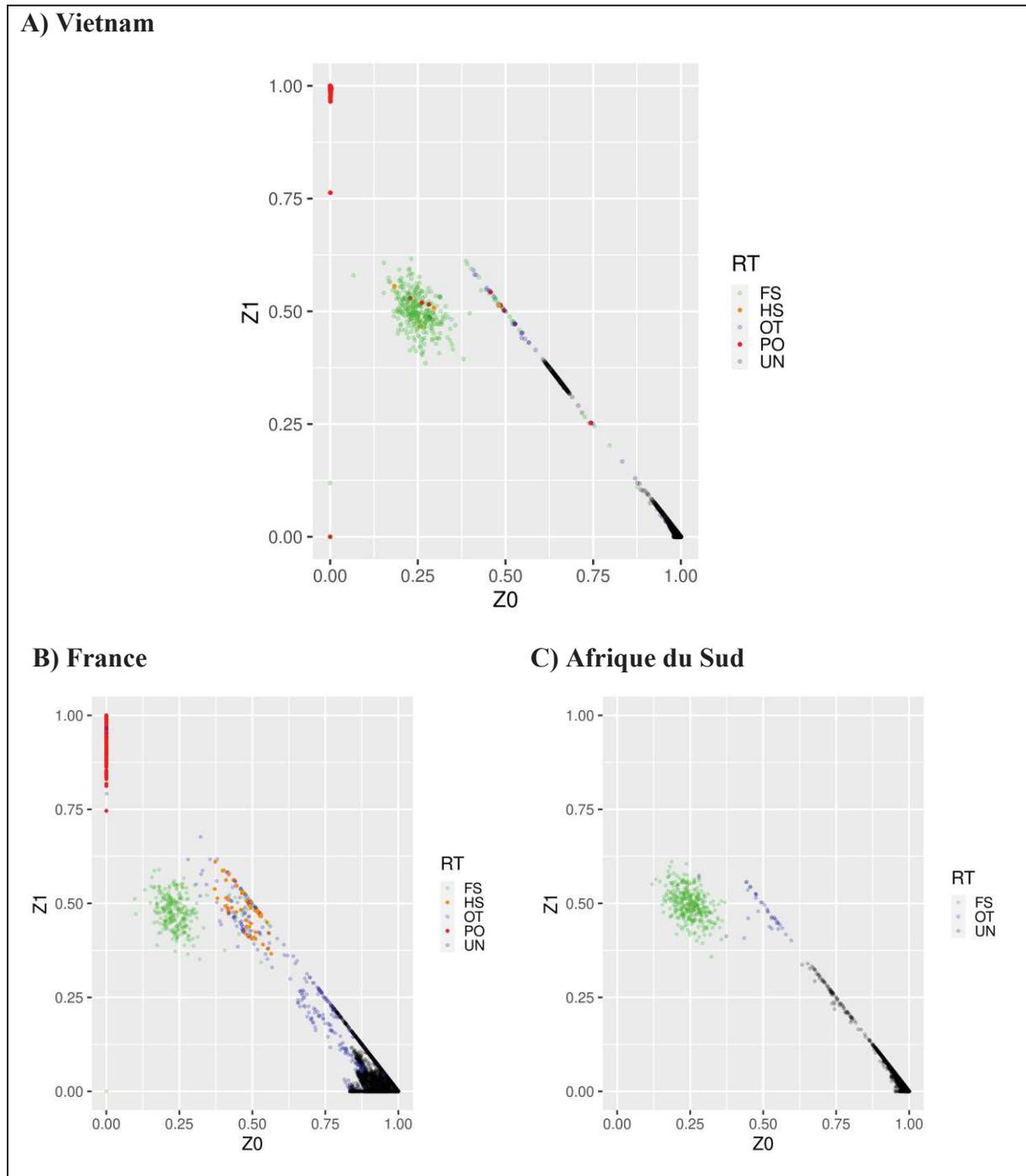


Figure 22. Structures familiales des cohortes du Vietnam (A), de France (B) et d'Afrique du Sud (C). Chaque point représente Z0 et Z1 pour une paire d'individus, et la couleur correspond aux relations familiales reportées dans le fichier .ped (en rouge PO, *parent/offspring* ; en vert, FS, *full siblings* ; en orange, HS, *half siblings* ; en violet et noir, autres relations et non-apparentés). Au Vietnam, on aperçoit en bas à gauche une paire d'individus FS avec Z0 = 0 et Z1 = 0 (autrement dit, cette paire partage 2 allèles identiques par descendance à chaque locus du génome et correspond à de vrais jumeaux). Dans chacune des cohortes, on peut observer un cluster de points verts correspondant aux fratries puis un grand nuage de points s'étendant du bas à droite (individus non apparentés en noir) jusqu'au milieu du graphique (individus apparentés au 2<sup>ème</sup> ou 3<sup>ème</sup> en violet ou demi-frères/sœurs). On peut également constater que la cohorte sud-africaine est essentiellement constituée de fratries. Enfin, l'hétérogénéité de la cohorte française se retrouve avec différents niveaux d'IBD reflétant différentes populations.

### 3.4 Définition du phénotype d'infection par *M. tuberculosis*

La définition du phénotype d'infection se basait à la fois sur les résultats de l'IDR et de l'IGRA. Le seuil de positivité de l'IDR a été fixé pour les 3 cohortes à 5 mm car c'est le seuil recommandé chez les personnes très exposées lors d'un contact récent avec un individu contagieux [181]. On retrouve également ce seuil dans la précédente GWAS sur la réactivité de l'IDR en Ouganda/Tanzanie et dans les études épidémiologiques de la résistance à l'infection [62],[136]. En outre, le phénotype d'intérêt étant la résistance à l'infection, on pourrait plutôt parler de seuil de négativité de 5 mm, qui, de fait, permet de définir des individus avec une IDR en moyenne de 0 mm dans les 3 cohortes. Les covariables associées au statut d'infection double négatif/double positif, défini plus en détail ci-dessous, ont été investiguées à chaque fois chez l'ensemble des contacts recrutés afin de maximiser le nombre d'individus, mais seule une partie d'entre eux ayant été génotypés a été retenue pour la GWAS (Figure 23).

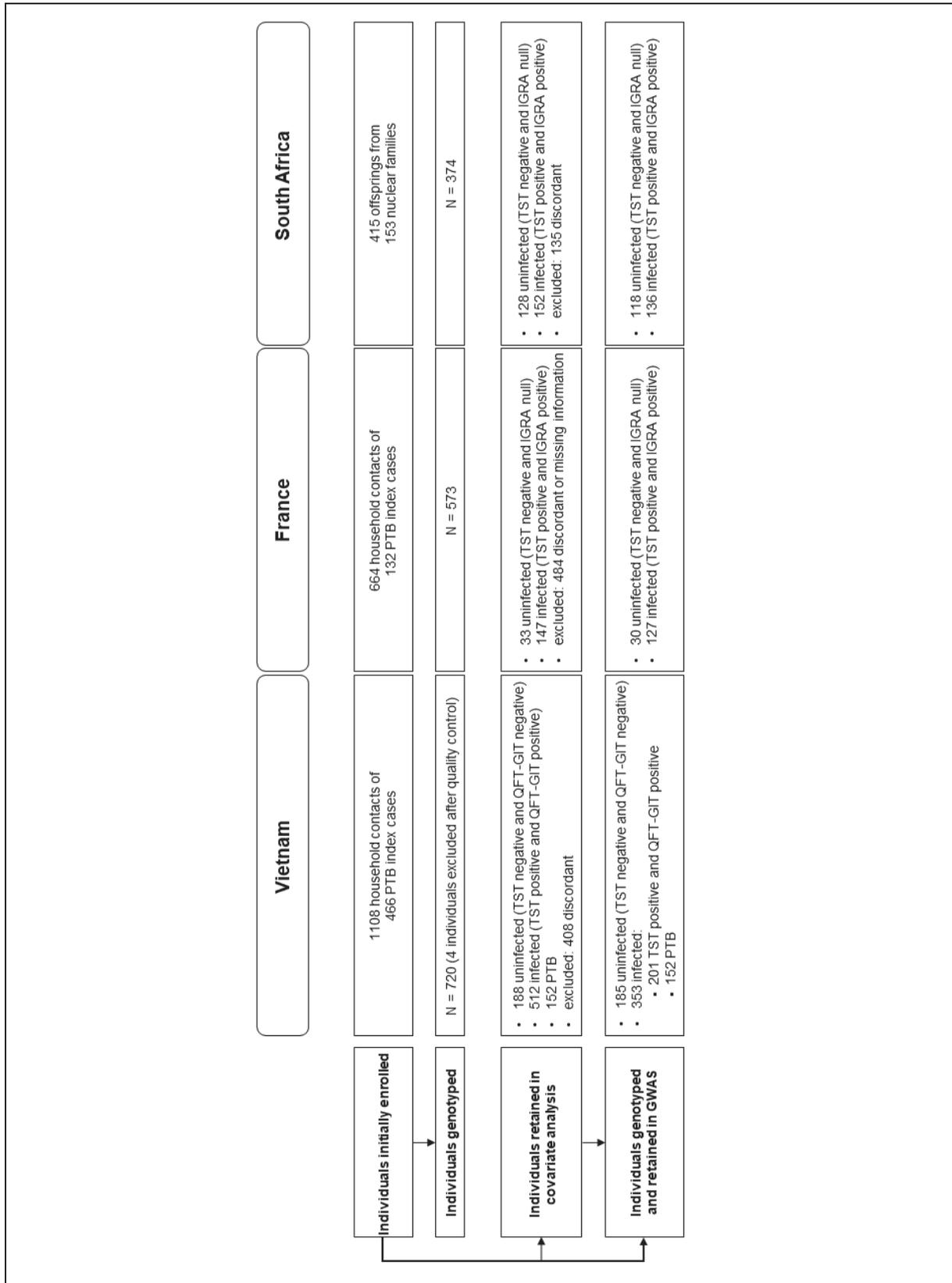


Figure 23. Organigramme général de l'étude.

Pour la cohorte du Vietnam, la résistance à l'infection a été définie par la présence d'une IDR < 5 mm ainsi qu'un résultat QFT négatif. Les individus infectés quant à eux devaient présenter une IDR  $\geq$  5mm et un résultat QFT positif (Figure 24). Un total de 188 contacts a été donc défini comme double négatifs ou résistants et 512 comme double positifs ou infectés, parmi lesquels 185 et 201 contacts ont été génotypés, respectivement. Ainsi, sur un total de 1108 contacts familiaux examinés, ~17% des contacts étaient résistants (en ligne avec les ~7 à 25% de résistants observés dans des cohortes de contacts en pays endémiques). Des covariables pouvant être associées à ce phénotype ont été explorées dans une régression logistique mixte (notamment vaccination BCG, nombre d'individus en moyenne par pièce dans la maison, durée de la toux du cas index avant début d'un traitement) mais aucune association n'a été observée (Tableau 9). Afin d'augmenter la taille de l'échantillon, ont été ajoutés 152 patients TB pulmonaire (PTB) génotypés (146 cas index et 6 individus ayant un antécédent de tuberculose maladie) au groupe des infectés. Ces individus PTB, étant malades, sont bien sûr infectés par définition ; mais l'infection et la maladie pouvant relever de différents mécanismes génétiques, je discuterai par la suite de l'influence des PTB dans les résultats.

Tableau 9. Covariables associées avec le phénotype d'infection (doubles positifs QFT et IDR, et avec et sans les 152 patients PTB comparés aux doubles négatifs) au Vietnam. Odds ratios par régression logistique mixte, analyse multivariable sur les données non manquantes.

Characteristics	QFT-/IDR-	QFT+/IDR+	OR (95%CI)	QFT+/IDR+/	OR (95%CI)
	n=188	n=512		PTB n=664	
Gender					
Male	61	189	ref	223	ref
Female	127	323	0.84 (0.53-1.33)	441	0.96 (0.80-1.08)
Age, years, mean(sd)	39(16)	32(14)	0.78 (0.30-1.62)	40(13)	0.93 (0.53-1.25)
BCG vaccination					
No	99	260	ref	NA	NA
Yes	69	190	1.01 (0.64-1.62)	NA	NA
Number of HHC per room, mean(sd)	2.4(1.4)	2.4(1.3)	0.99 (0.81-1.20)	NA	NA
Duration of index cough before treatment, months					
<2	63	193	ref	NA	NA
2-3	84	195	1.64 (0.95-2.84)	NA	NA
>3	40	120	1.07 (0.57-2.02)	NA	NA

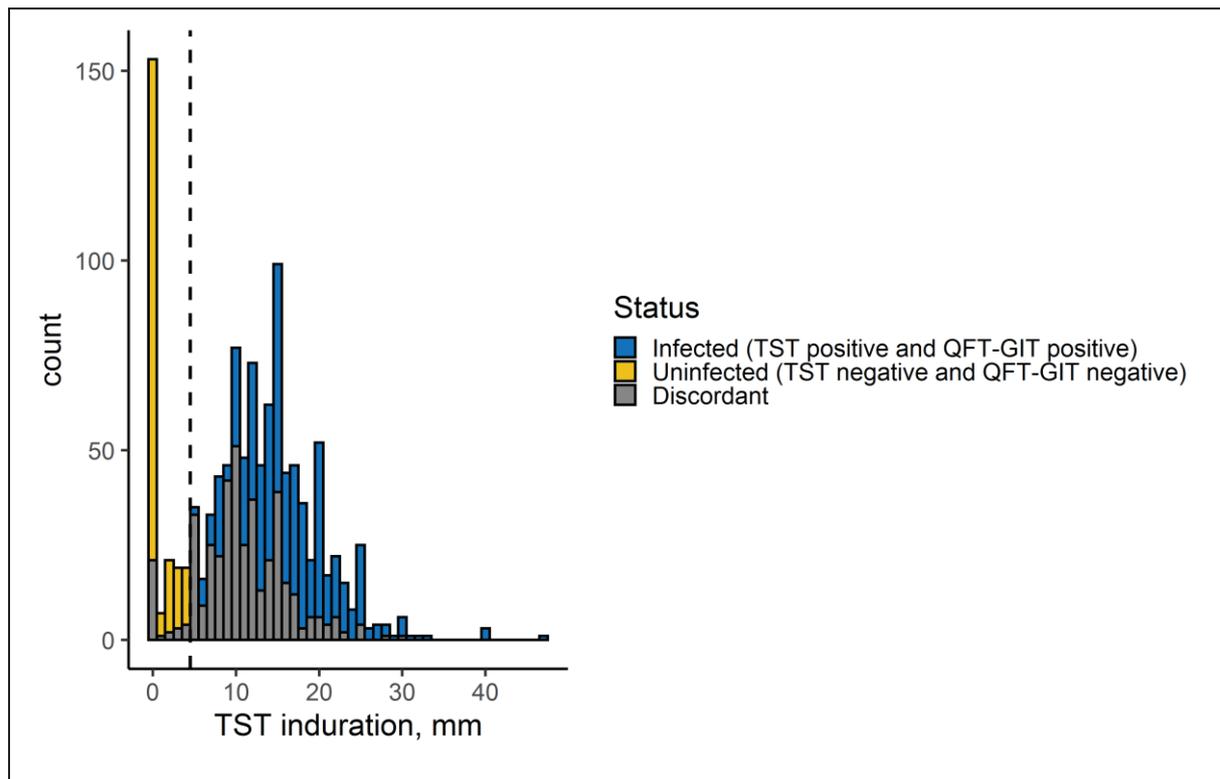


Figure 24. Histogramme représentant la distribution de l'IDR (TST) stratifiée par les résultats IGRA (QFT) dans la cohorte du Vietnam. La ligne verticale représente le seuil de 5 mm. Les individus résistants, double négatifs sont en jaune et ceux infectés, double positifs, en bleu. Une certaine proportion d'individus, en gris, avait des résultats discordants (positifs pour un test mais négatif pour l'autre) et n'ont pas été retenus pour la GWAS. Un sous-groupe de ces individus double négatifs/double positifs (185 et 201, respectivement) avaient été génotypés et ont été retenus pour la GWAS.

Pour la cohorte française, les contacts pouvaient avoir 1 ou 2 visites (V1 et V2) avec une mesure de l'IDR à chaque visite ; l'IGRA n'étant réalisé qu'à V1 par ailleurs. L'IDR était considérée comme négative si (i)  $< 5$  mm à la fois à V1 et V2 ou (ii)  $< 5$  mm à V1 quand une seule visite était faite. L'IDR était considérée comme positive si (i)  $\geq 5$  mm à V1 et V2 ou (ii)  $< 5$  mm à V1 puis  $\geq 10$  mm à V2, ce qui correspondrait à de vraies conversions à l'infection. Le résultat du test IGRA utilisé ici n'indiquait pas un statut catégoriel (positif ou négatif) mais directement la quantité d'IFN- $\gamma$  produite par les cellules sanguines après stimulation par ESAT6. J'ai donc dû déterminer un seuil de positivité et je suis resté assez conservateur. Ainsi, étaient considérés comme négatifs, les IGRAs avec une production nulle (= 0) d'IFN- $\gamma$ . Pour définir les IGRAs positifs, une courbe ROC (pour *receiver operating characteristic*) a été construite à partir du statut IDR comme la vraie observation et les quantités d'IFN- $\gamma$  (corrigées par le contrôle négatif) comme la valeur à prédire. La valeur présentant la plus grande somme sensibilité + spécificité était retenue comme seuil de positivité, ici 175 pg/mL. Au final, 33

contacts étaient définis comme résistants et présentaient une IDR négative et un IGRA nul ; et 147 contacts étaient définis comme infectés et présentaient une IDR positive et un IGRA positif > 175 pg/mL (Figure 25). Un sous-groupe de ces individus (30 et 127, respectivement) avaient été génotypés et ont été retenus pour la GWAS. Pour cette cohorte, de nombreux facteurs pouvant être associés au statut d'infection ont pu être explorés (une estimation de la durée de contact avec l'index, niveau de contagiosité de l'index défini par présence de cavernes à la radiographie, incidence de la tuberculose dans le pays de naissance, présence d'une couverture social) mais seul l'âge a été retenu pour la GWAS (Tableau 10).

Tableau 10. Association des covariables avec le phénotype d'infection (doubles positifs IDR et IGRA comparés aux doubles négatifs) chez les contacts de la cohorte de France. Odds ratios par régression logistique mixte, analyse multivariable sur les données non manquantes.

Characteristics	IGRA-/IDR- n=33	IGRA+/IDR+ n=147	OR (95%CI)
Gender			
Male	14	72	ref
Female	19	75	0.69 (0.10-4.60)
Age, years, mean(sd)	17(21)	28(18)	<b>1.07 (1.01-1.15)</b>
Contact duration with index, log hours, mean(sd)	5(2)	5(2)	0.87 (0.46-1.51)
Index contagiousness			
High	15	79	ref
Low-intermediate	18	68	0.57 (0.02-11.11)
Tuberculosis incidence of country of birth			
High	5	51	ref
Low-intermediate	28	96	0.14 (0.05-1.37)
Complementary health insurance			
No	10	45	ref
Yes	23	99	0.58 (0.01-8.33)

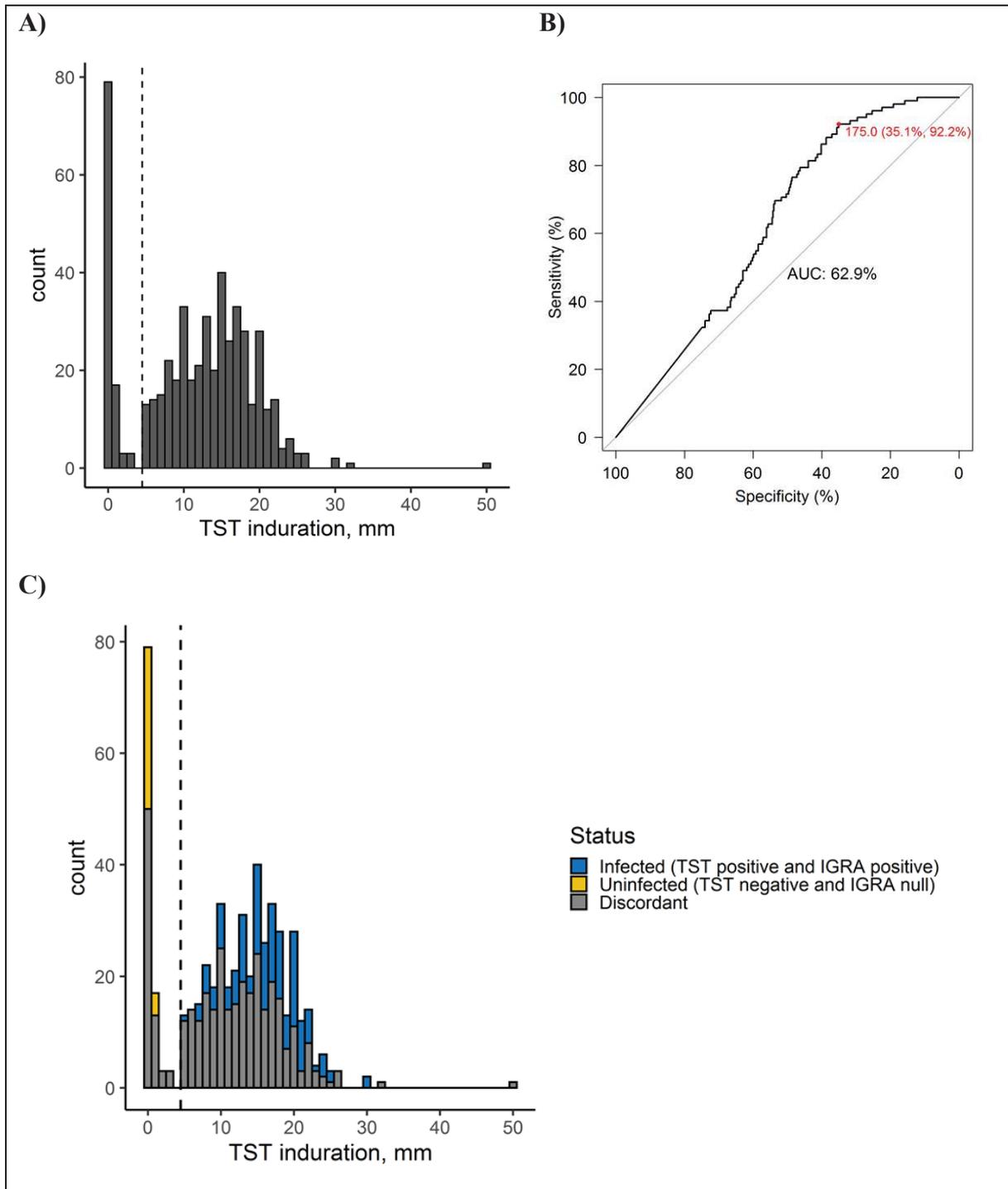


Figure 25. Définition du phénotype dans la cohorte française. **(A)** Distribution de l'IDR (TST) parmi 516 contacts de la cohorte (il s'agit de la valeur à V1 ou de la moyenne des 2 visites). La ligne verticale représente le seuil de 5 mm. **(B)** Construction de la courbe ROC à partir du statut IDR négatif/positif. La meilleure valeur discriminante du test IGRA était égale à 175. Un résultat IGRA était donc considéré positif si  $> 175$  ; un résultat égal à 0 était lui considéré négatif. **(C)** Histogramme représentant la distribution de l'IDR stratifiée par IGRA : en jaune les individus définis comme résistants (double négatifs) et en bleu ceux définis comme infectés (double positifs). Un sous-groupe de ces individus (30 et 127, respectivement) avaient été génotypés et ont été retenus pour la GWAS.

Enfin, j'ai procédé de la même façon que précédemment pour la définition du phénotype dans la cohorte d'Afrique du Sud : d'abord détermination d'un seuil de 5 mm pour l'IDR puis construction d'une courbe de ROC pour déterminer le seuil IGRA, ce qui a donné une valeur de 20.9 pg/mL. Au final, 128 contacts étaient définis comme résistants et présentaient une IDR négative et un IGRA nul ; et 152 contacts étaient définis comme infectés et présentaient une IDR positive et un IGRA positif > 21 pg/mL (Figure 26). Un sous-groupe de ces individus (118 et 136, respectivement) avaient été génotypés et ont été retenus pour la GWAS, qui a été ajustée uniquement sur l'âge (Tableau 11).

Tableau 11. Association des covariables avec le phénotype d'infection (doubles positifs IDR et IGRA comparés aux doubles négatifs) chez les contacts de la cohorte d'Afrique du Sud. Odds ratios par régression logistique mixte, analyse multivariable.

Characteristics	IGRA-/IDR- n=128	IGRA+/IDR+ n=152	OR (95%CI)
Gender			
Male	60	69	ref
Female	58	83	1.22 (0.67-2.17)
Age, years, mean(sd)	10(6)	15(6)	<b>1.30 (1.20-1.44)</b>

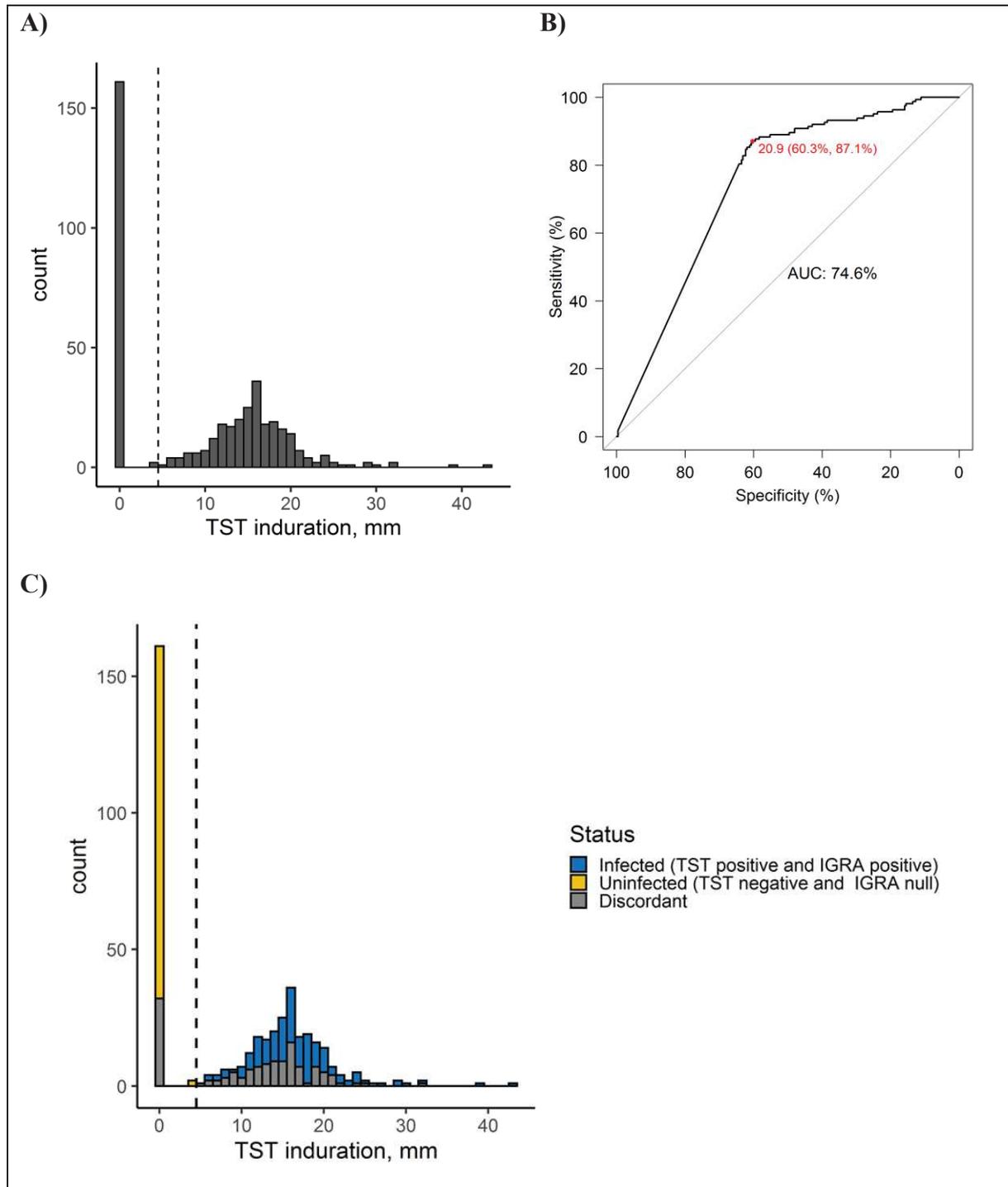


Figure 26. Définition du phénotype dans la cohorte d'Afrique du Sud. **(A)** Distribution de l'IDR (TST) parmi 415 participants de la cohorte. La ligne verticale représente le seuil de 5 mm. **(B)** Construction de la courbe ROC à partir du statut IDR négatif/positif. La meilleure valeur discriminante du test IGRA était égale à 20.9. Un résultat IGRA était donc considéré comme positif si  $> 20.9$ ; un résultat égal à 0 était lui considéré comme négatif. **(C)** Histogramme représentant la distribution de l'IDR stratifiée par IGRA : en jaune les individus définis comme résistants (double négatifs) et en bleu ceux définis comme infectés (double positifs). Un sous-groupe de ces individus (118 et 136, respectivement) avaient été génotypés et ont été retenus pour la GWAS.

Pour résumer, la GWAS a d'abord été conduite au Vietnam (entre 185 individus résistants et 353 infectés) puis dans 2 cohortes de réplication en France (entre 30 individus résistants et 127 infectés) et en Afrique du Sud (entre 118 résistants et 136 infectés) (Tableau 12).

Tableau 12. Description des cohortes

	Vietnam			France		Afrique du Sud	
	Résistants	Infectés		Résistants	Infectés	Résistants	Infectés
		ITL	PTB				
<b>n</b>	185	201	152	30	127	118	136
<b>Age moy</b>	36(16)	32(14)	48(12)	17(19)	28(18)	10(4)	15(6)
<b>Sexe, %F</b>	68	59	22	57	52	49	56
<b>IDR moy, mm</b>	1(1)	16(6)	na	0(1)	16(4)	0(0)	17(5)
<b>IGRA</b>	QFT-	QFT+	na	0	>175	0	>21

## 3.5 Analyse d'association pangénomique

### 3.5.1 Méthodes statistiques

Le caractère binaire du phénotype d'intérêt (non-infecté/infecté) orienterait naturellement la méthode d'analyse vers une régression logistique, où le logarithme de la cote de la probabilité d'être résistant à l'infection par *Mtb* est modélisé par une combinaison linéaire de facteurs explicatifs que sont le génotype et des variables d'ajustement comme les composantes principales.

La structure génétique de l'échantillon de type stratification de population pourrait être justement prise en compte par l'introduction des composantes principales dans la régression logistique. Cependant, la considération de liens familiaux n'est pas assurée par l'ajustement sur les composantes principales, en particulier quand celle-ci a été réalisée sur le sous-échantillon des individus « indépendants » pour permettre une meilleure détection des populations. Cela peut donc provoquer une inflation de l'erreur de type I [174],[182]. Leur prise en compte est possible dans le cadre de modèles linéaires mixtes incorporant une matrice de corrélation génétique (GRM pour *genetic relatedness matrix*) entre individus. Les modèles mixtes sont très utilisés dans le domaine de l'analyse de données répétées (plusieurs mesures chez un individu

à différents temps, par exemple), où ils permettent la prise en compte naturelle d'un effet « sujet » en attribuant à chaque individu une ordonnée à l'origine (*l'intercept*) propre, issue du tirage d'une variable aléatoire dont on modélise la variance. Dans le cadre d'une GWAS, les groupes d'individus appartenant à une même famille s'apparentent à ces données répétées. Le modèle mixte modélise l'effet du marqueur sur le phénotype comme un effet fixe, à l'instar de la régression linéaire classique, mais modélise les erreurs résiduelles comme issues du tirage d'un vecteur aléatoire dont la structure de covariance est contrainte par la spécification d'une matrice de corrélation génétique. Ainsi, le modèle attribue à chaque individu une ordonnée à l'origine qui lui est propre en contraignant les individus proches à avoir des ordonnées à l'origine proches. En plus de prendre en compte la structure familiale, les modèles linéaires mixtes peuvent adapter un phénotype binaire 0/1 (pour notre étude, résistant/infecté) en phénotype continu. Le cas échéant, une approche alternative consisterait à extraire d'abord les résidus d'une régression logistique classique du phénotype sur des covariables d'intérêt puis d'utiliser ces résidus comme nouveau phénotype dans la GWAS. Les modèles linéaires mixtes sont quoiqu'il en soit beaucoup plus avantageux que les modèles mixtes généralisés de type modèles logistiques. Pour les rendre accessibles aux approches pangénomiques, de nombreuses implémentations ont été proposées, dont le logiciel GEMMA que j'utiliserai ici, et elles se trouvent être très efficaces computationnellement [183].

L'impact d'un allèle protecteur sur la résistance à l'infection peut suivre différents mécanismes aboutissant à différents modèles génétiques. Un allèle récessif ne modifierait la protection contre l'infection que s'il est présent de manière homozygote. Un allèle dominant modifierait la protection contre l'infection dès sa présence à l'état hétérozygote et cette modification est identique à l'état homozygote. La protection pourrait également n'exister qu'en présence d'hétérozygotie (l'homozygotie pour n'importe lequel des 2 allèles étant alors sans effet ou délétère). Ce type de mécanisme a été observé chez l'homme, dans le domaine des maladies infectieuses où il peut être associé à de meilleures capacités de reconnaissance des pathogènes, notamment au niveau du complexe majeur d'histocompatibilité [184], conférant ainsi un avantage sélectif certain. Enfin, l'allèle à effet pourrait également augmenter la protection contre l'infection de façon additive, en fonction de sa présence à l'état hétérozygote ou homozygote. Ce dernier modèle est probablement celui le plus utilisé dans les GWASs. Le modèle additif a l'avantage d'être asymptotiquement équivalent en cas d'équilibre de Hardy-Weinberg à un modèle simple allélique qui comparerait directement la fréquence de l'allèle chez les résistants et les infectés, ce qui simplifie la description de l'allèle comme « facteur de

risque » au sens de l'épidémiologie classique. Pour beaucoup de marqueurs, la fréquence des homozygotes pour l'allèle mineur est faible, ce qui rend en pratique l'estimation d'un OR pour un effet dominant ou un avantage aux hétérozygotes proche de l'OR estimé par le modèle additif. Le seul modèle réellement négligé par cette approche est le modèle récessif, mais dans le contexte de notre étude et une taille d'échantillon de quelques centaines d'individus, la puissance pour détecter un effet récessif serait bien trop faible, en raison de la rareté globale des résistants et des infectés homozygotes. En pratique, je me propose de présenter le modèle additif (les autres modèles n'ayant pas apportés plus d'information). L'intégration du génotype d'un SNP AB dans la régression sur ce modèle est assez directe par rapport au vecteur des 3 probabilités assignées à chaque génotype puisqu'il s'agit de la somme  $2 * P_{AA} + 1 * P_{AB} + 0 * P_{BB}$  en prenant l'allèle A en référence (un modèle dominant donnant  $1 * P_{AA} + 1 * P_{AB}$ ).

La GWAS a donc été conduite entre les individus résistants et ceux infectés, tels que définis plus haut, dans les 3 cohortes, avec un modèle linéaire mixte et sur un modèle additif. La GRM a été calculée pour chaque cohorte à partir des génotypes filtrés après le contrôle qualité détaillé ci-dessus. De taille ( $n \times n$  individus), elle est calculée à partir de la matrice des génotypes centrés  $X$  de taille ( $n \times m$  marqueurs). Si nous notons  $G_{ij} = 0, 1$  ou  $2$  le génotype de l'individu  $i$  au SNP  $j$  codé comme le nombre d'allèles alternatifs, le même génotype centré vaut  $X_{ij} = G_{ij} - 2 * p_j$ , où  $2 * p_j$  est l'espérance des génotypes au SNP  $j$  et  $p_j$  la fréquence de l'allèle mineur. La GRM est donc la matrice de variance-covariance empirique des génotypes centrés des individus et est égale à  $(1/m) * XX^T$ .

Là aussi, d'autres stratégies sont possibles mais n'ont pas été plus informatives pour les résultats présentés ci-après. Le logiciel GEMMA propose par exemple l'estimation d'une GRM standardisée en centrant et réduisant les génotypes, c'est-à-dire en divisant  $G_{ij} - 2 * p_j$  par la variance des génotypes. Il a aussi été montré que l'inclusion des SNPs causaux dans la GRM pouvait conduire à des résultats d'association beaucoup plus conservateurs, réduisant la puissance des GWAS dans la découverte de vraies associations de [185]. Dans cette optique, une approche du *leave one chromosome out* (LOCO) a été proposée et consiste à retirer les SNPs d'un chromosome dans le calcul de la GRM qui sera utilisée pour l'association des marqueurs sur ce même chromosome. Cela revient à générer 22 GRM et réaliser 22 analyses, chacune incluant la GRM estimée sans le chromosome analysé. Une dernière stratégie consisterait à non pas utiliser la GRM à proprement parler mais une matrice de *kinship*, calculée à partir des liens familiaux et qui équivaut  $\sim 2 \times$  la matrice de corrélation génétique. La matrice de *kinship* donne le coefficient d'apparentement entre chaque paire d'individus, c'est-à-dire la

probabilité que 2 allèles d'un même gène soit tiré au hasard parmi la paire d'individus soient IBD. Une paire de germains ou parents/enfants aura un coefficient de 0.25, une paire de demi-frères/soeurs ou oncle-tante/neveu-nièce aura un coefficient de 0.125 etc. Cependant, les liens familiaux recueillis lors du recrutement d'une cohorte peuvent parfois être imprécis ou faux (voir Figure 22) pouvant amener à des imprécisions dans les résultats d'associations.

Toujours concernant l'utilisation de modèles linéaires mixtes, les effets d'un SNP estimés en régression linéaire sont obtenus sur l'échelle observée 0/1, différente de l'échelle logarithmique. Si, dans un modèle logistique, l'OR peut être estimé directement à partir de l'exponentiel du coefficient de régression, une transformation des coefficients estimés par le modèle linéaire est nécessaire [186],[187]. L'OR compare les cotes de la maladie quand un individu est porteur ou non-porteur de l'allèle à risque. Si  $p_0$  et  $p_1$  représentent sa fréquence chez les contrôles et les cas, l'OR est égal à [188]:

$$\frac{p_1}{1 - p_1} \times \frac{p_0}{1 - p_0}$$

Dans les GWASs, les cas sont souvent sur-représentés par rapport au contrôle, ce qui biaise la probabilité qu'un individu ait la maladie dans cet échantillon sélectionné par rapport à sa vraie distribution en population. En définissant  $k$ , la proportion de cas dans l'étude et une expression de la fréquence allélique,  $p = (1 - k)p_0 + kp_1$

Sous le modèle linéaire classique et s'appuyant sur la méthode des moindres-carrés pour trouver les coefficients de régression  $\beta$ , une transformation de l'OR peut être dérivée comme suit:

$$\text{OR} = \frac{[k + \beta(1-p)][1 - k + \beta p]}{[k - \beta p][1 - k - \beta(1-p)]}$$

En pratique, les coefficients, leur erreur standard et la proportion de cas (ici, les résistants à l'infection) dans la cohorte ont été utilisés pour obtenir les ORs et leur intervalle de confiance.

La formulation d'une fonction de vraisemblance des données en fonction du coefficient de régression permet son estimation au maximum de vraisemblance. Le caractère significativement différent de 0 de ce coefficient peut être évalué par différents tests asymptotiques classiques. En effet, en statistique, on teste des hypothèses. Il faut donc définir une hypothèse nulle à tester (dans une GWAS, l'absence d'effet d'un marqueur génétique), et une hypothèse alternative. Pour tester l'hypothèse nulle, il faut ensuite définir une statistique

dont la loi est connue. Enfin, on regarde si la valeur observée de cette statistique est trop différente de ce que l'on attendait sous l'hypothèse nulle. Si oui, l'hypothèse nulle est rejetée et la *p-value* correspond à la probabilité, sous l'hypothèse nulle, d'obtenir une statistique de test au moins aussi différente que celle observée. Ainsi, une petite *p-value* est révélatrice d'une vraie association. La *p-value* est un élément clé des GWAS car l'hypothèse nulle (l'effet d'un SNP est nul) est une hypothèse réaliste pour la large majorité des SNPs. Une association objectivée entre un SNP et un phénotype pointe un mécanisme biologique sous-jacent et la taille de l'effet peut paraître secondaire derrière l'identification de l'existence d'une association par la *p-value*. De plus, les tailles d'effet sont tellement petites dans les GWASs typiques qu'elles n'ont aucune réelle utilité en pratique. Il existe 3 types de tests préférentiellement utilisés (test du score, test de Wald et test du rapport de vraisemblance) mais le test du rapport de vraisemblance (LRT pour *likelihood ratio test*) qui regarde la différence des log-vraisemblances évaluées pour le coefficient sous hypothèse nulle et pour le coefficient estimé, reste le plus communément utilisé et sera conservé pour les résultats.

Enfin, pour terminer cette section, (il y a beaucoup à dire et cela pourrait faire l'objet d'une thèse à part entière [189]), les modèles linéaires mixtes permettent d'estimer l'héritabilité du phénotype d'intérêt (initialement, un trait quantitatif), c'est-à-dire la proportion de variance phénotypique expliquée par la composante génétique (notamment les facteurs génétiques avec effets additifs, on parle aussi d'héritabilité restreinte) pour une population donnée [190],[191]. L'estimation de l'héritabilité a longtemps été faite sur des données familiales, notamment sur des études de jumeaux car cela permettait de prendre facilement en compte l'environnement partagé dans une famille malgré des biais induits par la différence dans l'environnement partagé par des jumeaux monozygotes ou dizygotes. Dans le contexte des GWASs, les SNPs associés n'expliquent pas toute la variance génétique estimée antérieurement. Cette portion non-expliquée a été appelée héritabilité manquante [192], et a déclenché une avalanche d'études et de développement méthodologiques pour tenter de l'expliquer. Dorénavant, les estimations d'héritabilité reposent sur les données pangénomiques de (dizaines/centaines de) milliers d'individus non-apparentés, qui, en l'absence d'environnement partagé, seraient moins biaisées et ce, malgré les problèmes de stratification de population.

Dans le cas d'un phénotype binaire, la définition de l'héritabilité est moins évidente : il faut supposer l'existence d'un trait quantitatif sous-jacent (*liability*),  $L$ , dont la valeur ou seuil  $S$  déterminerait le phénotype binaire ( $Y$ ). Ainsi,  $Y = 1$  si  $L > S$ , sinon  $Y = 0$ . Il doit donc y avoir une transformation de l'héritabilité estimée. Sans rentrer dans les détails mathématiques [193],

cette conversion repose sur les prévalences du phénotype dans la population d'étude et en population générale. On voit donc que cela peut être très problématique. D'une part, dans une étude cas-témoins, les individus sont très sélectionnés en fonction de leur statut et ne sont pas représentatifs de la population générale. D'autre part, la prévalence du phénotype dans cette dernière peut largement être biaisée, c'est ce que l'on constate avec l'infection par *Mtb* par exemple.

Concrètement, dans le cadre de notre GWAS, mettant de côté l'aspect philosophique du calcul d'une héritabilité (après tout, si l'on réalise cette GWAS, c'est que l'on suppose qu'il existe des facteurs génétiques, a-t-on besoin de les quantifier sachant que l'on sait déjà que le trait est complexe ?), les estimations faites ici sont ininterprétables. Outre les difficultés de correctement supposer les prévalences d'infection, les échantillons étant de petite taille, les erreurs-type font que les héritabilités ne sont pas statistiquement différentes de zéro. Enfin, les structures familiales apportent une énième couche de complexité.

### 3.5.2 Analyse d'association pangénomique au Vietnam

Une GWAS a donc été réalisée dans la cohorte principale du Vietnam en comparant 185 individus résistants et 353 infectés sur 5 591 951 marqueurs génétiques. Le diagramme quantile-quantile (QQ-plot) permet de visualiser les *p-values* du test du rapport de vraisemblance aux *p-values* théoriques pour une loi du  $\chi^2$  à 1 degré de liberté (Figure 27). Le facteur d'inflation génomique ( $\lambda$ ) était de 0.997. Il est usuellement calculé en divisant la médiane observée des tests statistiques du  $\chi^2$  à 1 degré de liberté par la médiane attendue de cette même distribution. Une forte déviation du QQ-plot ou  $\lambda$  très supérieur à 1 suggèrent en premier lieu des problèmes de stratification de population. Ici, on peut donc être confiant quant au contrôle de la structure familiale de la cohorte. Une visualisation des résultats de la GWAS est donnée Figure 28 A. Il s'agit d'un Manhattan plot, un graphe représentant le logarithme négatif de la *p-value* issue de l'association entre chaque SNP et le phénotype le long des 22 chromosomes autosomaux. On peut observer, au niveau du chromosome 10, une association statistiquement significative au seuil pangénomique de  $5 \times 10^{-8}$  (seuil communément utilisé dans les GWASs pour prendre en compte les millions de tests indépendants effectués ;  $5 \times 10^{-8}$  correspond au seuil classique de 0.05 divisé par 1 million). Un zoom sur cette région du génome, au niveau du locus 10q26.2, plus précisément en région intronique et légèrement en

amont du gène *C10orf90* (ou aussi nommé *FATS*, HGNC: 26563) est donné Figure 28 B grâce au logiciel LocusZoom [194]. Il s'agit d'un groupe de 18 SNPs (12 significatifs et 6 autres associés avec  $P < 5 \times 10^{-7}$ ) en fort déséquilibre de liaison (LD pour *linkage disequilibrium*), comme on peut le constater en Figure 29 grâce au logiciel Haploview [195]. Le LD est une forme de corrélation entre 2 marqueurs, plus le déséquilibre est important, plus les 2 marqueurs dépendent l'un de l'autre. Le LD est également lié à la recombinaison : plus la recombinaison est probable entre 2 marqueurs du génome, plus le déséquilibre tendra à être faible. Les recombinaisons génétiques permettent de remanier les haplotypes et donc, avec du temps, de diminuer ce LD. Le LD est très important car il permet d'avoir une meilleure couverture sur tout le génome en génotypant seulement une petite partie des marqueurs, qui serviront de balise ou tag pour d'autres marqueurs en fort LD.

Le marqueur le plus fortement associé, rs11245088, était d'ailleurs un SNP imputé (OR = 0.42, intervalle de confiance à 95% (95%CI) 0.39-0.45,  $P = 1.58 \times 10^{-8}$ ). On pouvait également observer un marqueur génotypé, rs17155120 ( $P = 3.71 \times 10^{-8}$ ). Ce SNP a été génotypé avec une bonne qualité comme le montrait l'assignation des intensités en Figure 15 avec 3 groupes distincts. La qualité d'imputation des autres marqueurs était aussi excellente avec un *info score* > 0.9. Chaque copie de l'allèle mineur T du rs17155120 (que l'on retrouve à une fréquence de 20% chez les vietnamiens) confère une protection contre l'infection par *Mtb* avec un OR pour le risqué d'être infecté quand un individu est CT vs CC ou bien TT vs CT de 0.42 (95%CI 0.35-0.49). On peut apprécier l'effet additif de cet allèle T sur la Figure 28 C qui représente la proportion d'individus résistants à l'infection pour chacun des 3 génotypes possibles. Les ORs ont été transformés à partir des coefficients estimés dans le modèle linéaire mixte. L'OR obtenu à partir d'une régression logistique mixte effectuée dans R (directement l'exponentiel du coefficient) pour rs17155120 était de 0.41.

En conduisant la même analyse ajustée sur âge, sexe et les 3 premières PCs, on ne notait pas de différence nette, les *p-values* étaient même un peu plus significatives (rs17155120,  $P = 1.26 \times 10^{-9}$ ) (Figure 30).

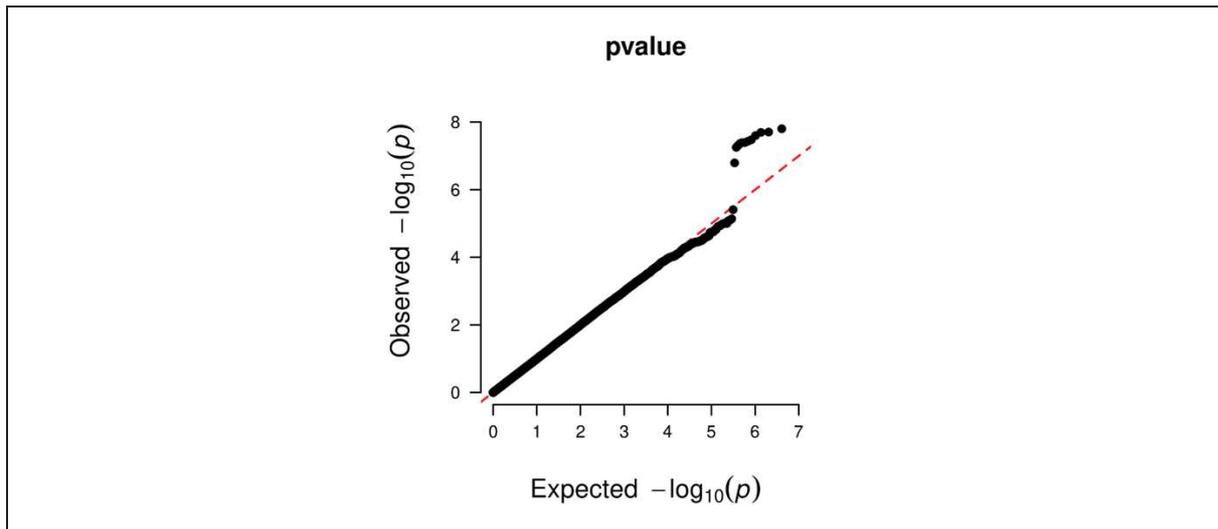


Figure 27. QQ-plot de l'analyse d'association pangénomique au Vietnam.

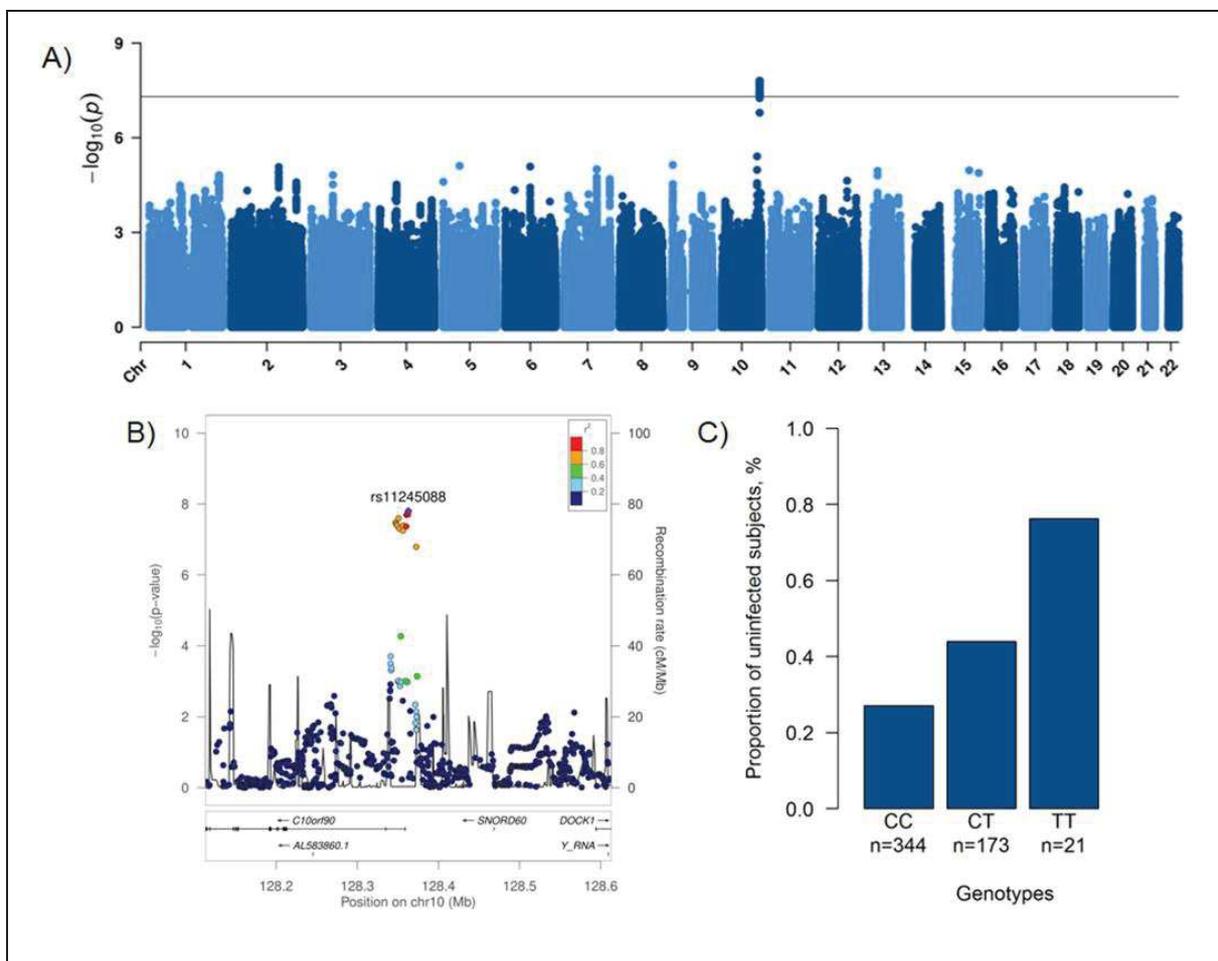


Figure 28. Analyse d'association pangénomique pour la résistance à l'infection par *Mtb* au Vietnam **A)** Manhattan plot pour 5 591 951 SNPs (MAF > 5% et info > 0.8) sous un modèle additif. La ligne horizontale représente le seuil de significativité  $P = 5 \times 10^{-8}$ . **B)** Locus zoom au niveau de 10q26.2, dans une fenêtre de 500kb entourant le top SNP rs11245088 (en violet). La couleur correspond au LD entre chaque marqueur et rs11245088 calculé pour la population Kinh de 1000G Phase 3. **C)** Proportion d'individus résistants à l'infection par *Mtb* pour chaque génotype du SNP génotypé rs17155120. (16/21 TT, 76/173 CT et 93/344 CC sont résistants).

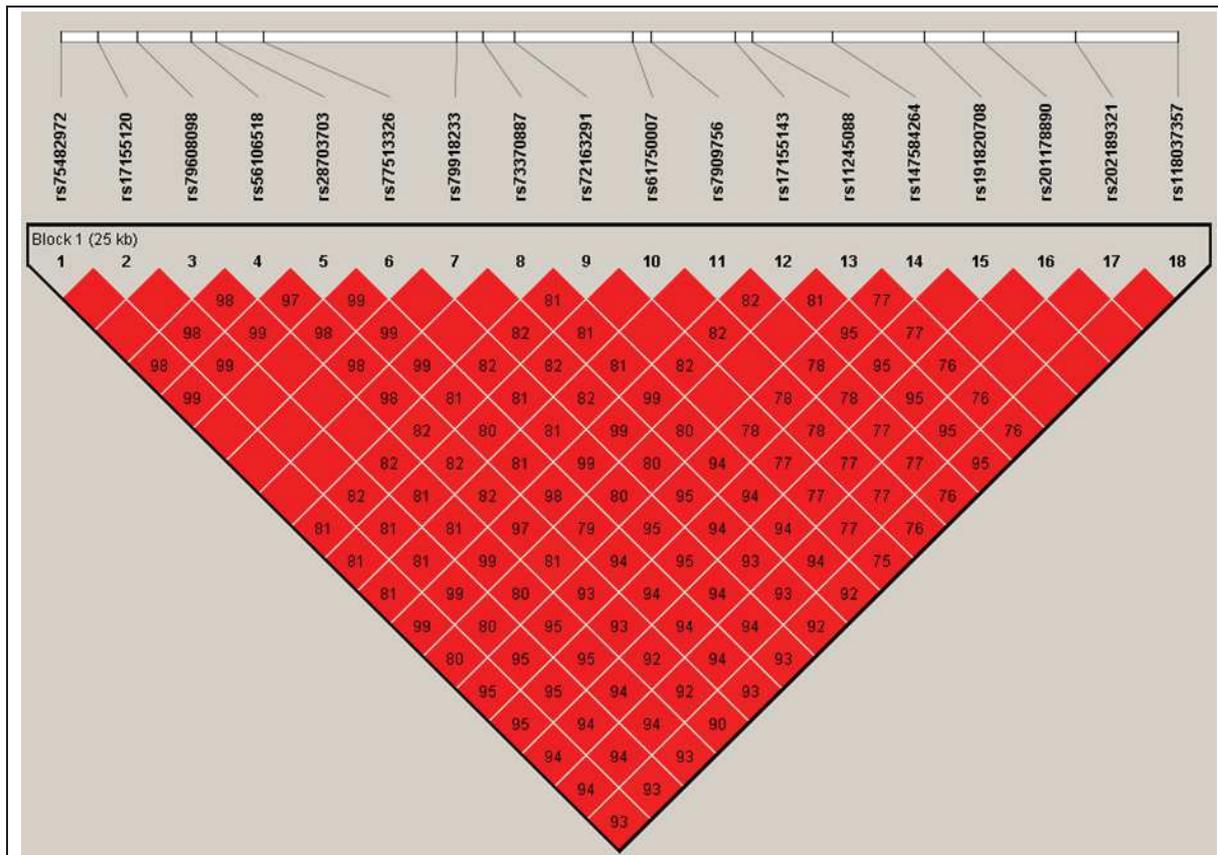


Figure 29. Déséquilibre de liaison entre chaque paire des 18 marqueurs associés au Vietnam.

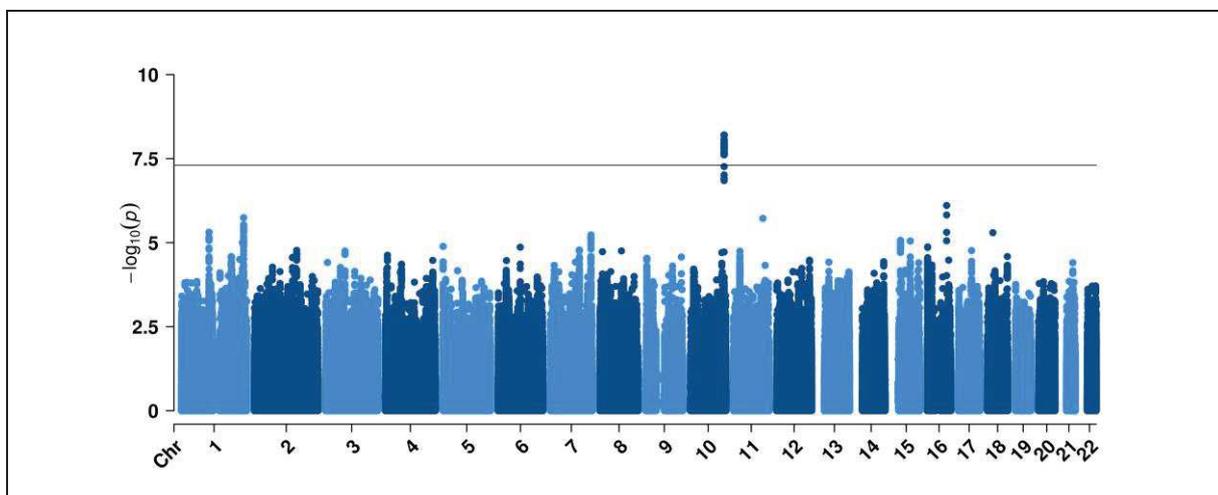


Figure 30. Manhattan plot de la GWAS entre 185 résistants et 353 infectés avec ajustement sur âge, sexe et 3 PCs au Vietnam.

Tableau 13. Association entre les marqueurs du locus 10q26.2 et la résistance à l'infection par *Mtb*. Pour chaque marqueur, sont donnés sa position par rapport au gène *C10orf90*, l'allèle de référence, sa fréquence, son odd ratio et sa *p-value* pour chacune des 3 cohortes.

Variant	C10orf90	EA	Vietnam			France			South Africa		
			EAF	OR (95%CI)	<i>P</i> value	EAF	OR (95%CI)	<i>P</i> value	EAF	OR (95%CI)	<i>P</i> value
rs11245088	upstream	C	0.25	0.42 (0.39-0.45)	$1.58 \times 10^{-8}$	-	-	-	0.44	0.99 (0.91-1.07)	$4.75 \times 10^{-1}$
rs72163291	intron	ins	0.24	0.42 (0.35-0.49)	$1.94 \times 10^{-8}$	0.37	0.88 (0.78-0.98)	$3.54 \times 10^{-1}$	-	-	-
rs17155143	upstream	A	0.20	0.39 (0.31-0.47)	$1.98 \times 10^{-8}$	0.18	0.63 (0.50-0.77)	$1.29 \times 10^{-1}$	0.15	0.60 (0.48-0.72)	$1.72 \times 10^{-2}$
rs7909756	upstream	G	0.24	0.42 (0.35-0.49)	$2.03 \times 10^{-8}$	0.37	0.83 (0.73-0.93)	$2.87 \times 10^{-1}$	0.36	0.90 (0.81-0.99)	$2.77 \times 10^{-1}$
rs28703703	intron	G	0.20	0.41 (0.34-0.49)	$2.52 \times 10^{-8}$	0.20	0.59 (0.48-0.69)	$4.95 \times 10^{-2}$	0.17	0.66 (0.55-0.77)	$2.95 \times 10^{-2}$
rs56106518	intron	C	0.23	0.41 (0.33-0.48)	$3.07 \times 10^{-8}$	0.30	0.40 (0.30-0.51)	$2.98 \times 10^{-3}$	-	-	-
rs75482972	intron	A	0.20	0.42 (0.34-0.49)	$3.35 \times 10^{-8}$	0.20	0.58 (0.47-0.68)	$4.38 \times 10^{-2}$	0.17	0.68 (0.57-0.78)	$3.48 \times 10^{-2}$
rs17155120	intron	T	0.20	0.42 (0.35-0.49)	$3.71 \times 10^{-8}$	0.18	0.48 (0.36-0.59)	$1.51 \times 10^{-2}$	0.16	0.62 (0.52-0.73)	$1.74 \times 10^{-2}$
rs73370887	intron	A	0.20	0.40 (0.33-0.48)	$4.05 \times 10^{-8}$	0.31	0.78 (0.71-0.85)	$1.43 \times 10^{-1}$	0.28	0.75 (0.67-0.83)	$3.69 \times 10^{-2}$
rs79608098	intron	T	0.20	0.42 (0.35-0.49)	$4.06 \times 10^{-8}$	0.20	0.58 (0.48-0.69)	$4.50 \times 10^{-2}$	0.17	0.66 (0.55-0.76)	$2.71 \times 10^{-2}$
rs61750007	upstream	C	0.24	0.43 (0.36-0.50)	$4.30 \times 10^{-8}$	0.24	0.53 (0.42-0.64)	$3.49 \times 10^{-2}$	0.20	0.73 (0.62-0.84)	$7.62 \times 10^{-2}$
rs77513326	intron	A	0.20	0.42 (0.34-0.49)	$4.93 \times 10^{-8}$	0.17	0.47 (0.36-0.59)	$1.61 \times 10^{-2}$	0.16	0.63 (0.52-0.75)	$2.35 \times 10^{-2}$
rs79918233	intron	A	0.20	0.41 (0.33-0.49)	$5.61 \times 10^{-8}$	0.17	0.51 (0.39-0.63)	$3.00 \times 10^{-2}$	0.16	0.63 (0.52-0.74)	$1.95 \times 10^{-2}$
rs147584264	upstream	C	0.19	0.41 (0.33-0.49)	$1.21 \times 10^{-7}$	0.22	0.77 (0.62-0.93)	$3.00 \times 10^{-1}$	-	-	-
rs191820708	upstream	A	0.19	0.41 (0.33-0.49)	$1.31 \times 10^{-7}$	0.19	0.63 (0.50-0.77)	$1.40 \times 10^{-1}$	-	-	-
rs201178890	upstream	T	0.19	0.41 (0.33-0.49)	$1.36 \times 10^{-7}$	0.22	0.79 (0.62-0.97)	$3.36 \times 10^{-1}$	-	-	-
rs202189321	upstream	T	0.20	0.41 (0.33-0.49)	$1.37 \times 10^{-7}$	0.18	0.65 (0.52-0.78)	$1.44 \times 10^{-1}$	-	-	-
rs118037357	upstream	A	0.19	0.41 (0.33-0.49)	$1.62 \times 10^{-7}$	0.15	0.47 (0.33-0.60)	$2.83 \times 10^{-2}$	0.14	0.55 (0.43-0.67)	$7.38 \times 10^{-3}$

Cette première analyse a été réalisée avec des sujets soit simplement infectés (double positifs à l'IDR et au QFT) soit ayant/ayant eu une tuberculose pulmonaire. Afin de confirmer que l'association était bien le fait de l'infection par Mtb et non de la maladie, la GWAS a été répétée entre les 185 individus résistants et les 201 double positifs, en excluant donc les 152 malades. Malgré une taille d'échantillon plus petite, les 28 SNPs du locus restaient associés avec la résistance à l'infection avec  $P < 5 \times 10^{-6}$  et des ORs similaires (pour rs17155120, OR = 0.40, 95%CI 0.32-0.49,  $P = 2.55 \times 10^{-7}$ ) (Figure 31 et Tableau 14). En plus du locus 10q26.2, un groupe de SNPs au niveau du chromosome 7 se détachait avec cette configuration (top marqueur rs2055017,  $P = 6.9 \times 10^{-8}$ ). Cependant, cette association n'existait quasiment plus dans la GWAS principale incluant les PTB ( $P = 2.6 \times 10^{-5}$ ).

La même analyse ajustée sur tous les facteurs d'exposition (âge, sexe, vaccination BCG, nombre d'individus par foyer, durée de la toux du cas index) et les 3 premières PCs était similaire aux résultats initiaux (rs17155120,  $P = 3.55 \times 10^{-7}$ ) (Figure 32).

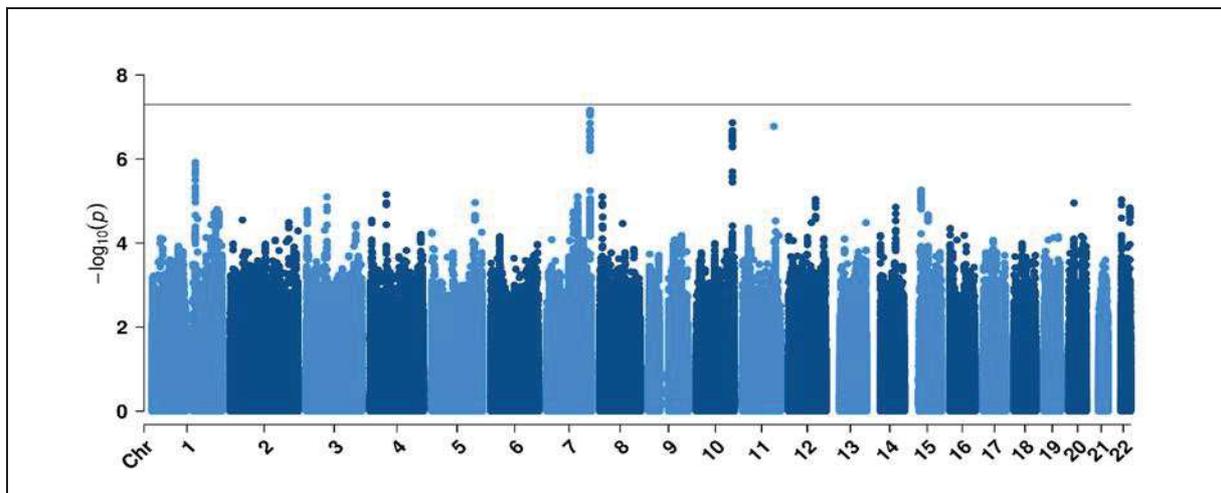


Figure 31. Manhattan plot de la GWAS entre 185 résistants et 201 infectés au Vietnam.

Tableau 14. Marqueurs du locus 10q26.2 associés avec la résistance à l'infection par *Mtb* (185 résistants et 201 infectés).

SNP	OR (95%CI)	P value
rs56106518	0.39 (0.30-0.47)	$1.37 \times 10^{-7}$
rs72163291	0.41 (0.33-0.49)	$2.09 \times 10^{-7}$
rs28703703	0.40 (0.32-0.49)	$2.40 \times 10^{-7}$
rs17155120	0.40 (0.32-0.49)	$2.55 \times 10^{-7}$
rs11245088	0.41 (0.40-0.42)	$2.72 \times 10^{-7}$
rs7909756	0.41 (0.33-0.49)	$2.73 \times 10^{-7}$
rs75482972	0.41 (0.32-0.49)	$2.82 \times 10^{-7}$
rs17155143	0.39 (0.30-0.47)	$3.02 \times 10^{-7}$
rs79608098	0.41 (0.32-0.49)	$3.28 \times 10^{-7}$
rs77513326	0.40 (0.32-0.49)	$3.56 \times 10^{-7}$
rs73370887	0.39 (0.31-0.48)	$3.82 \times 10^{-7}$
rs61750007	0.42 (0.34-0.50)	$4.91 \times 10^{-7}$
rs79918233	0.40 (0.31-0.49)	$5.13 \times 10^{-7}$
rs147584264	0.41 (0.32-0.50)	$2.01 \times 10^{-6}$
rs191820708	0.41 (0.32-0.50)	$2.59 \times 10^{-6}$
rs201178890	0.41 (0.32-0.50)	$2.71 \times 10^{-6}$
rs202189321	0.41 (0.32-0.50)	$2.72 \times 10^{-6}$
rs118037357	0.42 (0.32-0.51)	$3.54 \times 10^{-6}$

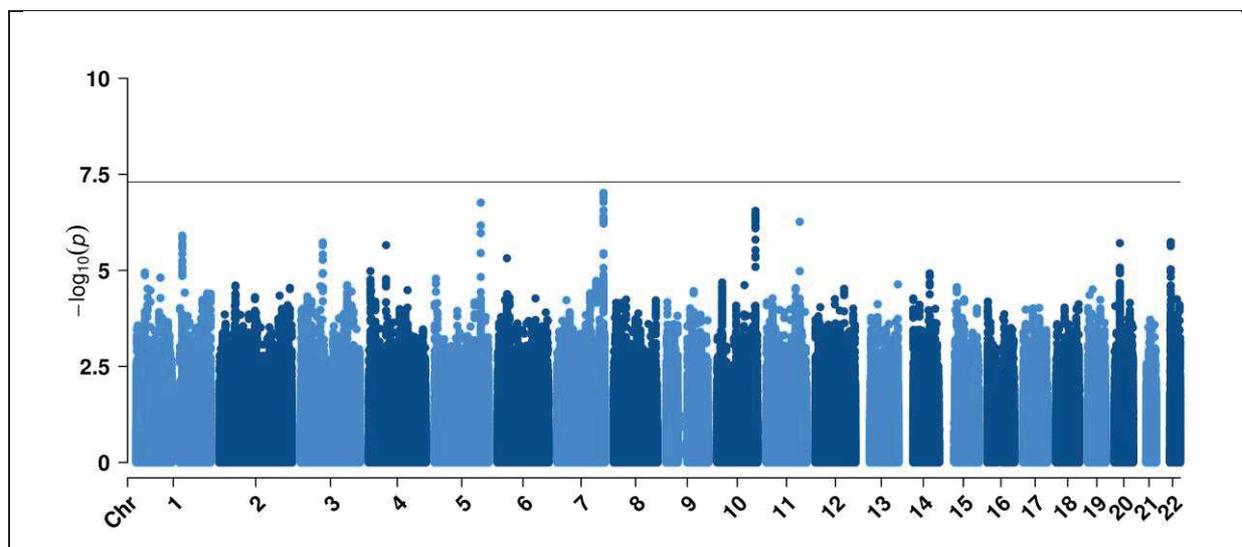


Figure 32. Manhattan plot de la GWAS ajustée sur âge, sexe, vaccination BCG, nombre d'individus par foyer, durée de la toux du cas index et 3 PCs entre 167 résistants et 171 infectés au Vietnam.

En prenant les 152 PTB comme infectés (et excluant les 201 double positifs cette fois-ci) et en ajustant sur l'âge et le sexe (les PTB étant plus âgés et plus souvent des hommes), on retrouvait une association significative, mais moins forte que celle en présence uniquement des doubles positifs, pour le rs17155120, OR = 0.50 (95%CI 0.41-0.59,  $P = 2.10 \times 10^{-4}$ ) (Figure 33 et Tableau 15).

Enfin, en comparant les 201 individus infectés double positifs aux 152 malades PTB (ce qui revient en fait à faire une GWAS sur la susceptibilité à la maladie) montrait en plus qu'il n'y avait pas de différence entre ces 2 types de sujets (pour le rs17155120,  $P = 0.87$ ).

On constate donc que les PTB étaient des individus appropriés pour être inclus dans le groupe infection mais que l'association observée est bien liée à l'infection latente seule, et non à la maladie. C'est une observation importante car il est plus probable que la biologie et les gènes impliqués dans la résistance à l'infection ou à la maladie soient assez différents.

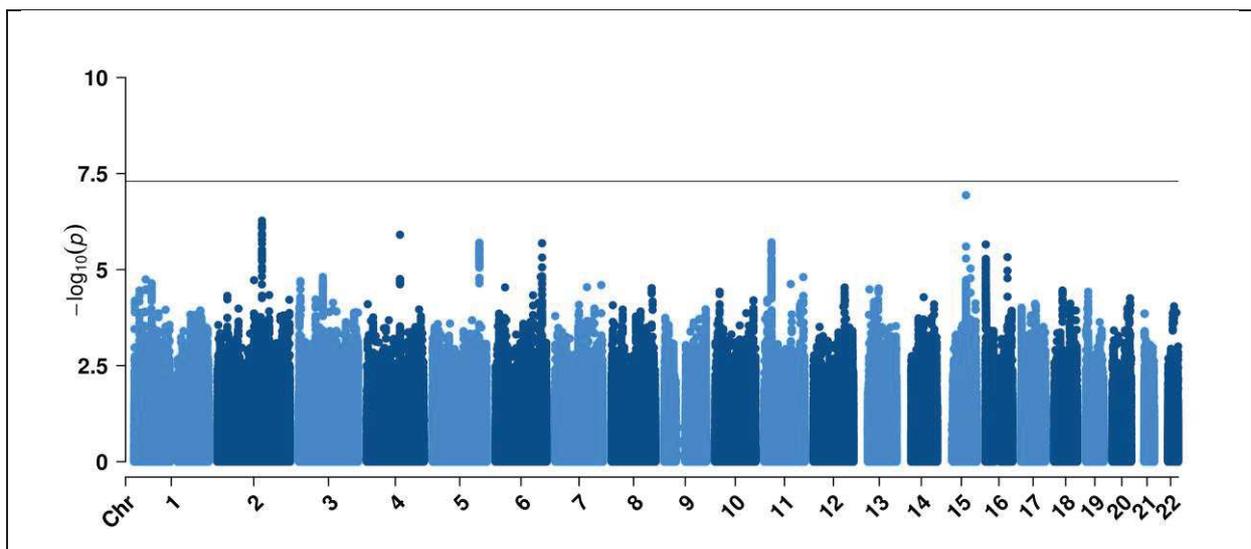


Figure 33. Manhattan plot de la GWAS entre 185 résistants et 152 PTB au Vietnam.

Tableau 15. Marqueurs du locus 10q26.2 associés avec la résistance à l'infection par *Mtb* (185 résistants et 152 PTB).

SNP	OR (95%CI)	P value
rs11245088	0.48 (0.40-0.56)	$6.30 \times 10^{-5}$
rs7909756	0.48 (0.39-0.57)	$6.39 \times 10^{-5}$
rs72163291	0.48 (0.40-0.56)	$6.56 \times 10^{-5}$
rs28703703	0.49 (0.41-0.56)	$1.01 \times 10^{-4}$
rs56106518	0.48 (0.39-0.56)	$1.12 \times 10^{-4}$
rs17155143	0.45 (0.38-0.53)	$1.22 \times 10^{-4}$
rs79918233	0.48 (0.40-0.57)	$1.45 \times 10^{-4}$
rs61750007	0.48 (0.40-0.56)	$1.53 \times 10^{-4}$
rs73370887	0.48 (0.40-0.57)	$1.55 \times 10^{-4}$
rs75482972	0.48 (0.40-0.57)	$1.89 \times 10^{-4}$
rs79608098	0.50 (0.41-0.59)	$1.99 \times 10^{-4}$
rs202189321	0.46 (0.39-0.53)	$2.03 \times 10^{-4}$
rs17155120	0.50 (0.41-0.59)	$2.10 \times 10^{-4}$
rs77513326	0.50 (0.41-0.59)	$2.65 \times 10^{-4}$
rs147584264	0.48 (0.40-0.57)	$2.90 \times 10^{-4}$
rs191820708	0.48 (0.40-0.57)	$3.23 \times 10^{-4}$
rs118037357	0.48 (0.40-0.57)	$3.26 \times 10^{-4}$
rs201178890	0.48 (0.40-0.57)	$3.31 \times 10^{-4}$

Comme on peut le voir avec les Manhattans plots (Figure 28 et Figure 31), l'ajout de ces patients PTB au groupe infecté était néanmoins nécessaire pour avoir la possibilité de détecter un effet de cette taille. Le Tableau 16 montre les estimations de puissance avec une taille d'échantillon de 538 pour différentes fréquences alléliques et tailles d'effet (en supposant une prévalence de la résistance à l'infection de 10% et un seuil de significativité à  $5 \times 10^{-8}$ ).

Tableau 16. Analyse de puissance statistique de la GWAS Vietnam.

Freq	Effet									
	1.1	1.5	2	2.5	3	3.5	4	4.5	5	6
<b>0.01</b>	0	0	0	0	0.001	0.003	0.008	0.017	0.035	0.100
<b>0.05</b>	0	0	0.005	0.054	0.222	0.484	0.726	0.879	0.954	0.995
<b>0.1</b>	0	0.001	0.043	0.310	0.692	0.910	0.98	0.996	0.999	1
<b>0.2</b>	0	0.003	0.179	0.654	0.920	0.987	0.998	1	1	1
<b>0.3</b>	0	0.006	0.246	0.707	0.926	0.984	0.997	0.999	1	1
<b>0.4</b>	0	0.007	0.226	0.634	0.868	0.955	0.984	0.994	0.997	0.999
<b>0.5</b>	0	0.006	0.159	0.476	0.720	0.854	0.921	0.955	0.973	0.988
<b>0.6</b>	0	0.003	0.083	0.274	0.469	0.617	0.718	0.787	0.834	0.891
<b>0.7</b>	0	0.001	0.029	0.104	0.199	0.289	0.367	0.430	0.483	0.615
<b>0.8</b>	0	0	0.006	0.020	0.041	0.063	0.085	0.105	0.123	0.154
<b>0.9</b>	0	0	0	0.001	0.002	0.003	0.004	0.005	0.006	0.008

Toutes ces analyses ont été faites avec un phénotype double IDR/IGRA avec un seuil d'IDR fixé à 5 mm. Même si le phénotype ne semblait pas être biaisé par le BCG par exemple, il était intéressant de tester un seuil de 10 mm, toujours couplé au QFT. Cela conduisait à analyser 250 individus dans le groupe non-infecté (IDR < 10 mm et QFT négatif) et 338 dans le groupe infecté ( $\geq 10$  mm et QFT positif). L'association sur le locus 10q26.2 était toujours présente, même si non significative au seuil pangénomique (rs17155120,  $P = 1.17 \times 10^{-7}$ ) (Figure 34). Cette perte de puissance s'explique possiblement par une plus mauvaise classification du statut d'infection avec un seuil de 10 mm: une partie des contacts présentant une IDR à 8 ou 9 mm par exemple et définis comme résistants avec ce nouveau seuil pourraient être en réalité des individus infectés.

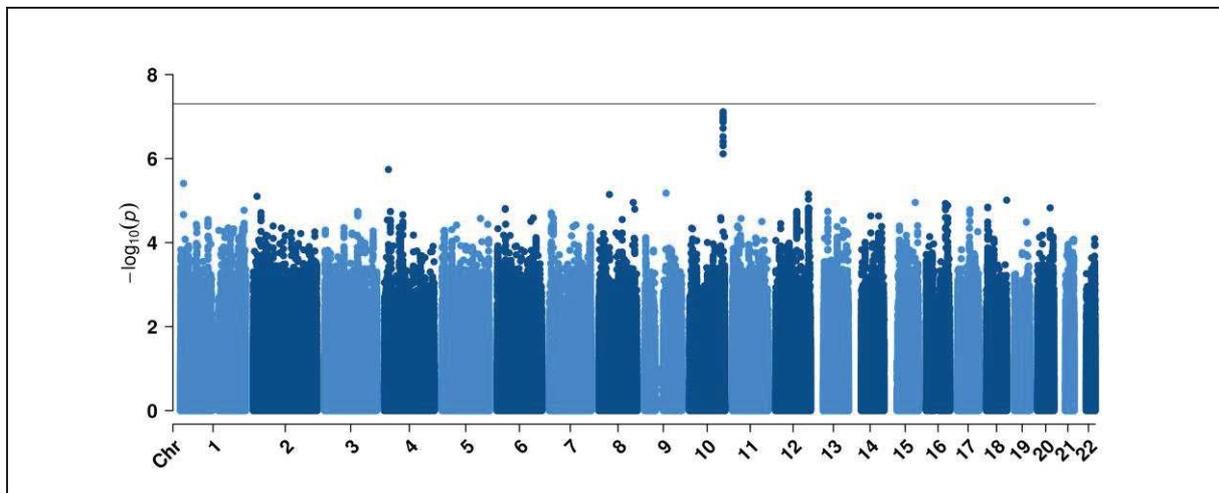


Figure 34. Manhattan plot de la GWAS entre 250 résistants et 338 infectés au Vietnam (au seuil IDR de 10 mm).

### 3.5.3 Réplication dans 2 cohortes indépendantes de France et d'Afrique du Sud

Après avoir identifié cette association significative dans notre cohorte de découverte, j'ai cherché à répliquer ce signal. En effet, les GWASs interrogent la problématique des tests multiples: les corrections faisant l'hypothèse de tests indépendants comme Bonferroni peuvent sembler trop sévères dans une perspective de génération d'hypothèses, et aboutir à l'élimination excessive de faux négatifs. Dans le même temps, de nombreux artefacts (problèmes de génotypage, stratification de population, problèmes méthodologiques etc.) peuvent aboutir à

l'identification de fausses associations. La manière la plus raisonnable de se départir des faux positifs tout en contrôlant en partie l'erreur de type II est le passage des meilleurs signaux au révéléteur de l'étude de réplication [196]. La réplication consiste à tester les meilleurs signaux dans un nouvel échantillon indépendant et de ne valider que les signaux significatifs et avec une direction d'effet uniforme à 2 reprises. Ce nouvel échantillon indépendant peut être issu d'une même population (réplication à proprement parler) voire d'une population totalement différente en terme d'ascendance génétique (le terme transférabilité est alors parfois utilisé en lieu de réplication). Il n'existe pas de consensus quant à la conduite d'une réplication. Tout dépend des signaux initiaux identifiés, leur *p-value* initiale et des échantillons disponibles pour la réplication.

Les 12 marqueurs du locus 10q26.2 avec  $P < 5 \times 10^{-8}$  ont donc été évalués dans 2 cohortes indépendantes de France (contacts familiaux, 30 individus résistants et 157 infectés) et d'Afrique du Sud (individus fortement exposés en communauté, 118 résistants et 136 infectés). Après inspection du LD entre les marqueurs du locus présents dans ces 2 cohortes (sur le groupe de 18 SNPs identifiés au Vietnam, 17 étaient imputés/génotypés avec succès en France et 12 en Afrique du Sud, Figure 35), 2 blocs pouvaient être distincts, pouvant correspondre à la réalisation de 2 tests indépendants. Un seuil nominal de réplication avec  $P$  unilatérale  $< 0.025$  ( $0.05/2$ ) a été choisi. Ainsi, en France, 3 des SNPs significatifs au Vietnam étaient validés dans la cohorte française dont le rs17155120, également génotypé (OR = 0.48, 95%CI 0.36-0.59,  $P = 1.51 \times 10^{-2}$ ). Même constat en Afrique du Sud, 3 SNPs étaient répliqués, y compris le rs17155120, là encore génotypé (OR = 0.62, 95%CI 0.52-0.73,  $P = 1.74 \times 10^{-2}$ ). Par ailleurs, au niveau pangénomique, aucune association n'était observée, comme on pouvait s'y attendre compte tenu de la taille des échantillons (Figure 36).

En ajustant l'analyse avec 3 PCs et les facteurs d'exposition (France: âge, sexe, durée du contact, contagiosité du cas index défini par la présence de cavernes à la radiographie, incidence de la tuberculose dans le pays de naissance et couverture sociale ; Afrique du Sud: âge et sexe), les résultats étaient là aussi similaires à la réplication principale: pour rs17155120,  $P = 8.98 \times 10^{-3}$  en France et  $P = 1.88 \times 10^{-2}$  en Afrique du Sud.

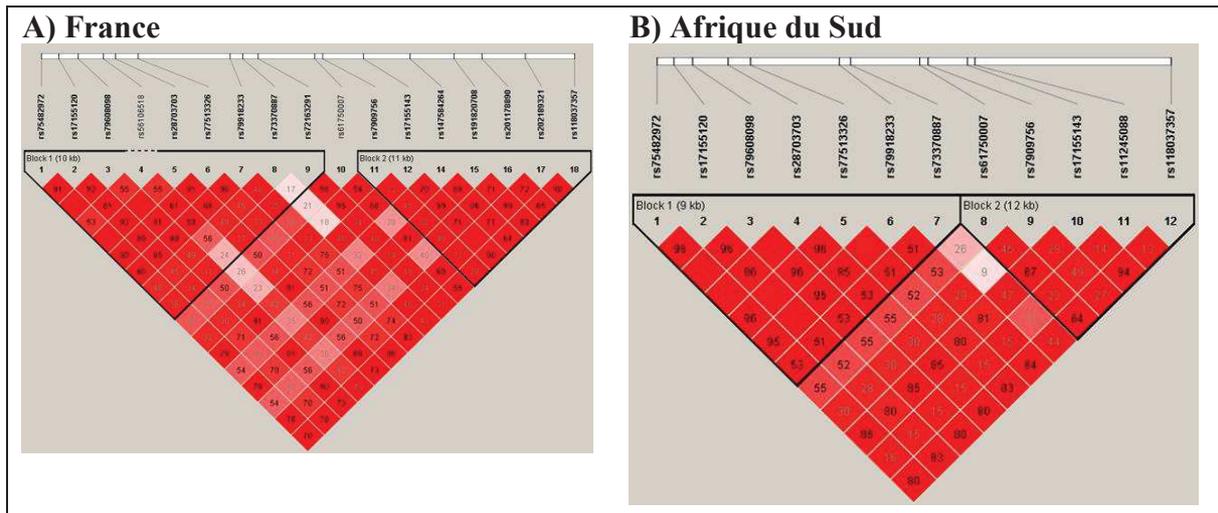


Figure 35. Déséquilibre de liaison entre chaque paire de marqueurs du locus 10q26.2 en France (A) et Afrique du Sud (B).

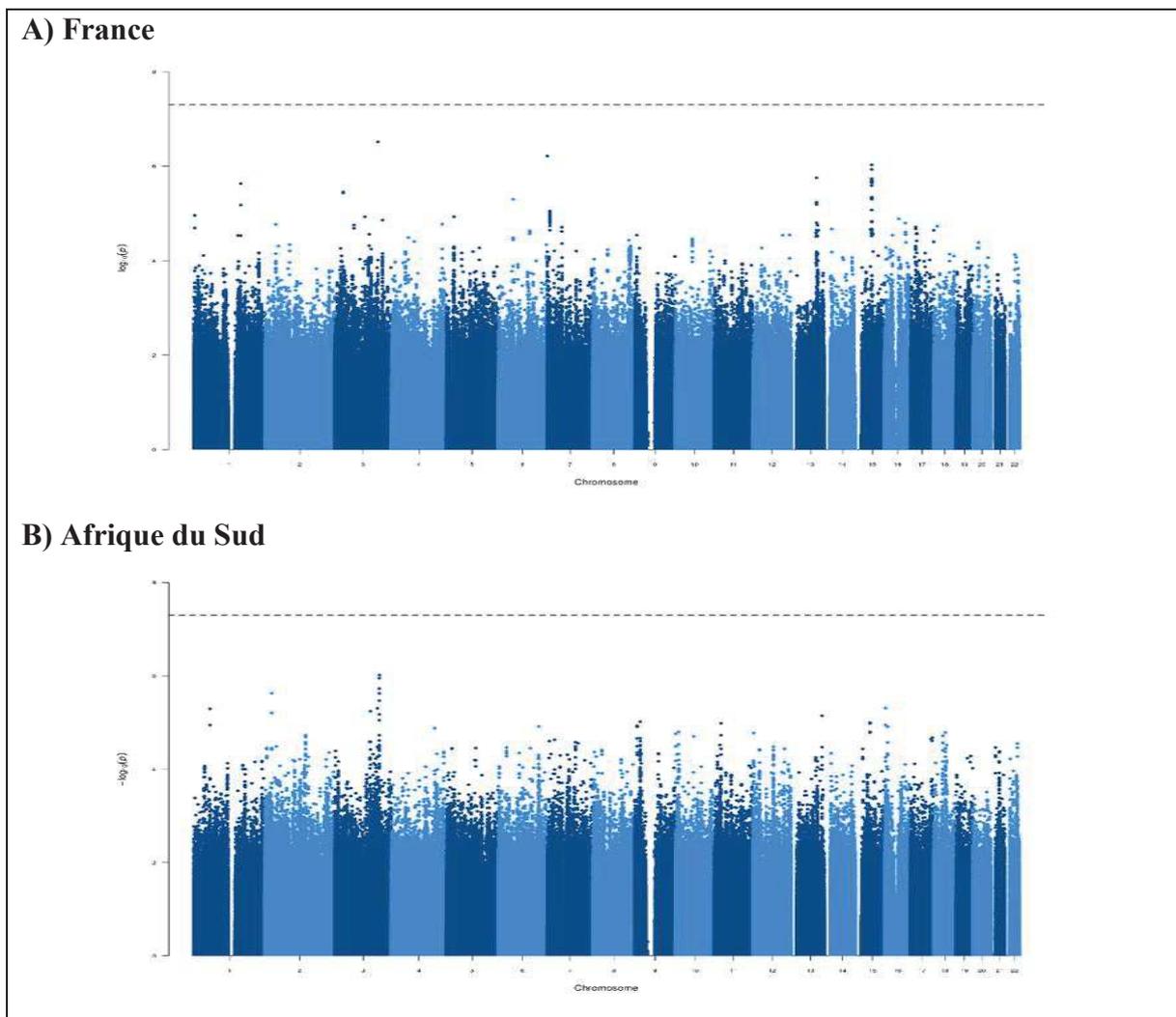


Figure 36. Manhattan plots pour les GWASs de la résistance à l'infection par *Mtb* en France (A) et en Afrique du Sud (B).

De façon intéressante, les marqueurs de ce locus 10q26.2 sont aussi en fort LD les uns avec les autres dans les 2 cohortes de France et d'Afrique du Sud, et proches du LD observé au Vietnam. La MAF du SNP géotypé, rs17155120 T, qui est l'allèle protecteur, était également similaire pour les 3 cohortes, entre 16 et 20%. Plus précisément, la fréquence de cet allèle au Vietnam (20%) est similaire à celle de la population Kinh de 1000G Phase 3 (18%) (Figure 37). De la même façon, la fréquence retrouvée dans la cohorte française, qui incluait des individus de toutes les origines, correspondait à la fréquence moyenne retrouvée chez toutes les populations de 1000G Phase 3 (18 et 16%, respectivement) (Figure 37). Cependant, pour la cohorte sud-africaine, la MAF de rs17155120 T (16%) était supérieure à celles que l'on peut observer parmi les populations africaines de 1000G Phase 3 (allant de 5 à 11%) (Figure 37). Cette différence se confirme à l'inspection du LD en Figure 38 : dans une région de ~30kb autour de rs17155120, on observe un LD moins fort et des blocs plus petits parmi les populations africaines comparé aux sud-africains. Ce n'est pas vraiment une surprise car les « Coloured » sud-africains présentent un grand brassage génétique avec des ascendances notamment européennes et sud-asiatiques en plus de la composante africaine. L'ACP et l'analyse d'admixture avaient déjà permis de noter ces différences.

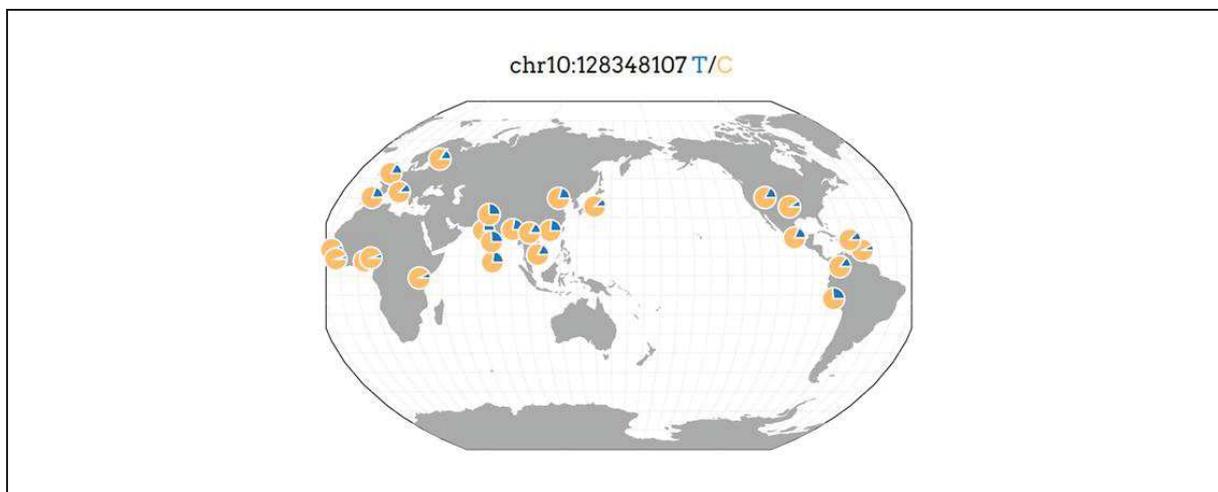


Figure 37. Distribution géographique des allèles de rs17155120 parmi les populations de 1000G Phase 3. L'allèle protecteur contre l'infection par *Mtb*, T, est en bleu. (<http://www.popgen.uchicago.edu/ggv>).

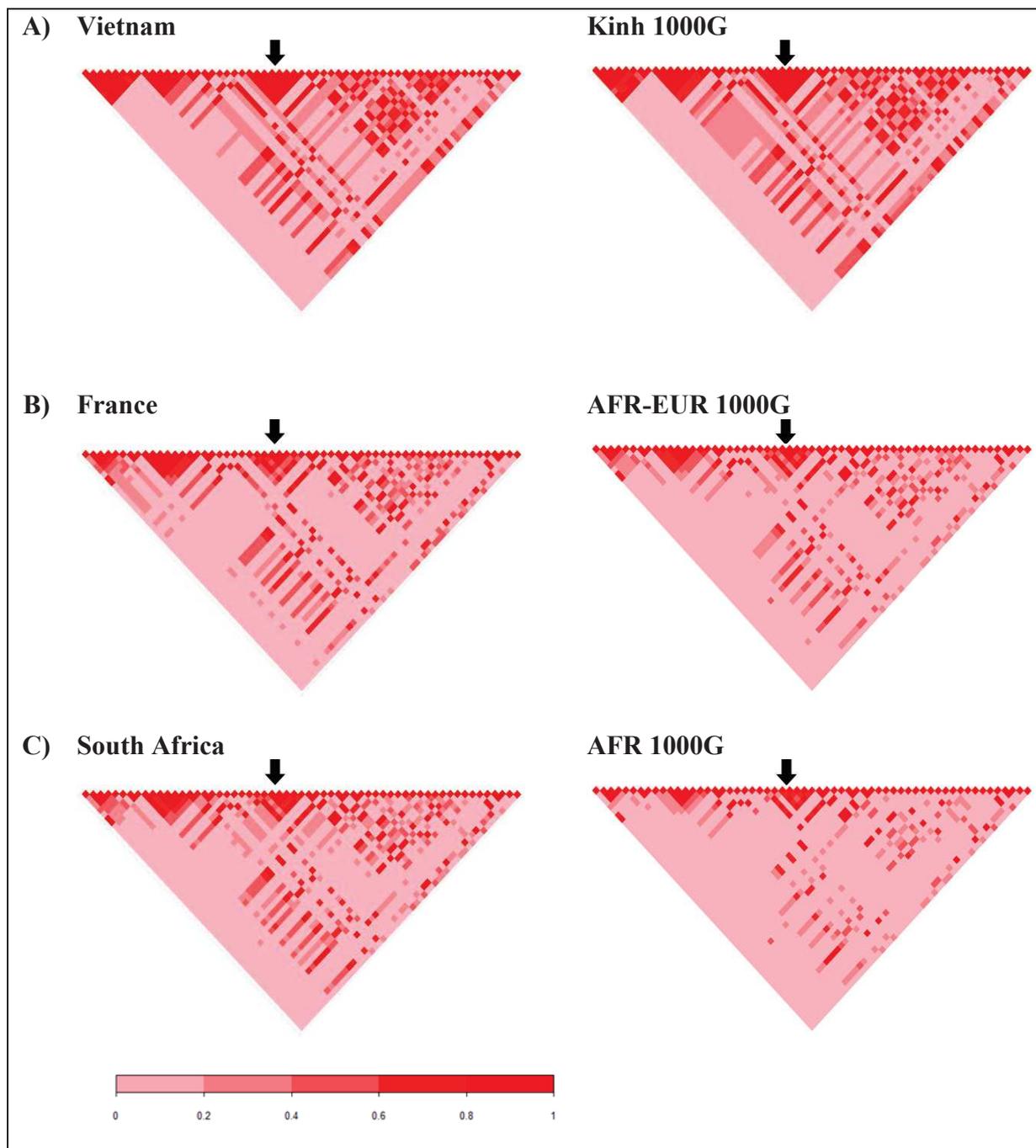


Figure 38. Déséquilibre de liaison dans une fenêtre de 30kb autour du marqueur rs17155120 (flèche) dans les 3 cohortes de l'étude et les populations correspondantes dans 1000G Phase 3.

Cette structure de LD similaire dans différentes populations et une fréquence allélique moyennement élevée dans plusieurs populations exceptées africaines interrogent sur les forces évolutives qui ont pu aboutir à ces observations. Est-ce que les allèles protecteurs contre l'infection par *Mtb* ont pu être soumis par exemple à une sélection positive ? Dans une population exposée à un environnement particulier, la mutation sélectionnée positivement va

augmenter en fréquence mais continuer d'évoluer sous neutralité dans les autres populations. En particulier, la recherche de loci positivement sélectionnés qui n'ont pas encore atteint la fixation montre que ce genre de signatures sont souvent trouvées à travers plusieurs populations différentes [197].

L'indice de différenciation  $F_{ST}$ , par exemple, mesure une évolution différentielle des fréquences alléliques entre populations qui va entraîner une augmentation de la différenciation génétique à ce locus [198]. Chaque mutation qui apparaît au sein d'un haplotype est en LD avec les mutations qui l'entourent. Au cours des générations, dans un modèle neutre, ce LD va disparaître sous l'effet de la recombinaison. Une nouvelle mutation avantageuse dont la fréquence augmente rapidement sera portée par un haplotype long car la recombinaison n'aura pas le temps de casser l'haplotype sur lequel elle est apparue. Des tests ont été développés sur le calcul de l'EHH (*extended haplotype homozygosity*) mesurant l'homozygotie autour de chaque allèle à un SNP central. En comparant l'EHH autour des allèles ancestraux et dérivés à un locus donné, l'iHS (*integrated haplotype score*) permet la détection d'haplotypes anormalement longs [199],[200]. Pour explorer cette hypothèse de sélection pour le marqueur génotypé rs17155120, des *p-values* empiriques de l'indice de différenciation  $F_{ST}$  et de l'iHS avaient été calculées dans le laboratoire à partir des populations de 1000G Phase 3 ( $P = 0.14$  et  $P = 0.55$ , respectivement), indiquant que le marqueur suivait la trajectoire que l'on attendrait sous neutralité.

Un autre type de force pourrait être envisagé, comme une sélection balancée, mais il est plus probable que la résistance à l'infection ne soit pas un phénotype si avantageux dans l'évolution. Après tout, l'infection est simplement asymptomatique et n'est pas délétère comme la maladie. Allant dans ce sens, aucun signal de sélection n'avait été décrit pour l'association identifiée dans la GWAS de réactivité à l'IDR chez des sujets VIH+ en Ouganda et Tanzanie [136] tandis que dans une GWAS « jumelle » par les mêmes auteurs portant cette-fois sur la tuberculose maladie [201], le marqueur conférant une résistance à la tuberculose maladie présentait un signal fort de sélection positive. Dans la même idée, l'allèle P1104A de TYK2 prédisposant à la tuberculose a fait l'objet d'une forte sélection négative [122].

### 3.5.4 Méta-analyse

Pour compléter ces 3 GWASs, il est utile de combiner les résultats obtenus dans les 3 cohortes via une méta-analyse. La méta-analyse de GWASs permet d'augmenter considérablement la taille de l'étude (à la fois en nombre d'individus et en nombre de SNPs couverts) et la puissance statistique et améliore ainsi la détection de marqueurs génétiques à effet faible que l'on aurait pu manquer dans une étude simple. La création de consortia et la mise à disposition des résultats des GWASs (*summary statistics*) dans le domaine public a aidé à considérablement augmenter le nombre d'études impliquées dans une méta-analyse, le nombre total d'individus analysés et le nombre de traits examinés. Il n'est plus rare de voir des méta-GWASs de plus d'1 million d'individus [202],[203] ou de regrouper des phénotypes similaires (types de cancers, maladies autoimmunes, maladies neurologiques) à la recherche de marqueurs pléiotropiques [204]. Tout cela s'accompagne de son flot de problèmes computationnels, méthodologiques et conceptuels [205].

Une des principales difficultés des méta-analyses est la variabilité des effets observés qui peut soit venir de l'incertitude intra-étude, soit de la variabilité inter-études, c'est-à-dire d'une réelle hétérogénéité entre les différentes études combinées. C'est généralement le cas lorsque les études ont été conduites au sein de populations différentes comme dans cette thèse.

En terme de méthodes, on distingue 3 grandes familles de méta-analyses qui ont été développées autour de cette difficulté [205]: (i) celles à effets fixes où la variance des tailles d'effet est suppose égale à 0, (ii) celles à effets aléatoires qui estiment la variance inter-études et la prennent en compte et (iii) les bayésiennes qui intègrent des *a priori* sur des paramètres comme la variance inter-études, les tailles d'effet ou le modèle génétique.

Ces-dernières méthodes semblaient de prime abord les plus adaptées pour ce travail afin de réaliser une méta-analyse « trans-ethnique ». Certaines méthodes comme MANTRA [206] prennent en considération une hétérogénéité allélique parmi les différentes populations. MANTRA groupe les populations en fonction des fréquences alléliques moyennes ou un indice de fixation  $F_{ST}$ , ce qui autorise une variation des effets entre les populations en fonction de leur « distance génétique » et génère un facteur de Bayes pour chaque marqueur et sa probabilité postérieure d'hétérogénéité. Cependant, ce genre de méthode suppose que les populations soient homogènes, ce qui n'est pas du tout le cas de la cohorte française par exemple.

Les méthodes à effets fixes sont basées sur l'hypothèse qu'un unique effet commun (fixe) pour chaque marqueur est sous-jacent aux études incluses dans la méta-analyse. Autrement dit, si chaque étude avait une taille infinie, les résultats de chacune d'entre elles seraient identiques en l'absence d'hétérogénéité. Ce type de méthode teste l'hypothèse nulle d'absence d'association dans aucune des études analysées. Dans l'optique d'augmenter la puissance de détecter une association, c'est la méthode de choix mais en présence d'hétérogénéité, cette puissance devient limitée.

À l'inverse, les méthodes à effets aléatoires font l'hypothèse que chaque étude estime différents effets, qui ont une distribution avec une moyenne et une certaine variance. En l'absence d'une variabilité inter-études, les estimations par effets fixes ou aléatoires ont tendance à être identiques alors qu'en présence d'hétérogénéité, les estimations d'effets aléatoires sont conservatrices et vont avoir des *p-values* plus grandes et une variance plus large (des intervalles de confiance plus grands donc). Cet aspect très conservateur a été en partie corrigé par Han et Eskin et leur implémentation dans le logiciel METASOFT [207]. Traditionnellement, l'analyse à effets aléatoires estime dans un premier temps la taille de l'effet et son intervalle de confiance en tenant compte de l'hétérogénéité (une variance inter-études estimée par maximum de vraisemblance par exemple est ajoutée au calcul qui consiste à pondérer la somme des effets du SNP pour toutes les études par l'inverse de sa variance). Cette taille d'effet est normalisée en un *z-score* puis transformée en *p-value*. Cette étape correspond en fait à faire l'hypothèse d'une hétérogénéité sous l'hypothèse nulle. Or, sous l'hypothèse nulle (pas d'association à travers les études), il ne devrait pas y avoir d'hétérogénéité puisque tous les effets devraient être égal à 0. Cela se traduit par des *p-values* trop conservatrices. À la place de la transformation en *z-score*, Han et Eskin ont proposé d'appliquer un test du rapport de vraisemblance et de faire l'hypothèse d'absence d'hétérogénéité sous l'hypothèse nulle et d'autoriser cette hétérogénéité sous l'hypothèse alternative.

Dans le cadre de ce travail, la méthode à effets aléatoires était la plus appropriée *a priori*, étant donné la diversité des 3 cohortes étudiées, et les résultats principaux ont été obtenus avec METASOFT et leur méthode améliorée. J'ai utilisé comme effets à combiner les coefficients de la régression linéaire mixte et leur erreur-type et non les ORs dérivés. Cela apporte un peu plus de précision [187]. Les effets estimés ont donc été de nouveau transformés pour obtenir des ORs interprétables. L'hétérogénéité a été calculée par la statistique Q de Cochran, une méthode communément utilisée obtenue par la somme pondérée du carré des différences entre

les effets individuels de chaque étude et l'effet combiné de toutes les études.  $Q$  suit une loi du  $\chi^2$  à 1 degré de liberté.

L'analyse combinée a été réalisée avec 333 individus résistants et 616 infectés. Un unique signal était observé, celui du locus 10q26.2 (Figure 39 et Figure 40). Le signal le plus significatif était observé pour le marqueur génotypé rs17155120 (OR = 0.50, 95%CI = 0.45-0.55,  $P = 1.26 \times 10^{-9}$ ) et aucune hétérogénéité n'était constatée à travers les 3 études du Vietnam, de la France et d'Afrique du Sud ( $P_{\text{het}} = 0.31$ ) (Tableau 17).

En réalisant cette méta-analyse avec les autres méthodes évoquées ci-dessus, on s'aperçoit que la méthode à effets fixes (METAL [208]) livrait des *p-values* plus significatives comme l'on pouvait s'y attendre en l'absence d'hétérogénéité tandis que la méthode à effets aléatoires était effectivement très conservatrice (Figure 41). Enfin, la conduite d'une analyse combinée sans les PTB vietnamiens (au total, 333 résistants et 464 infectés) menaient à des résultats similaires avec rs17155120 et une discrète baisse de la *p-value* (OR = 0.51, 95%CI = 0.46-0.57,  $P = 1.21 \times 10^{-8}$ )

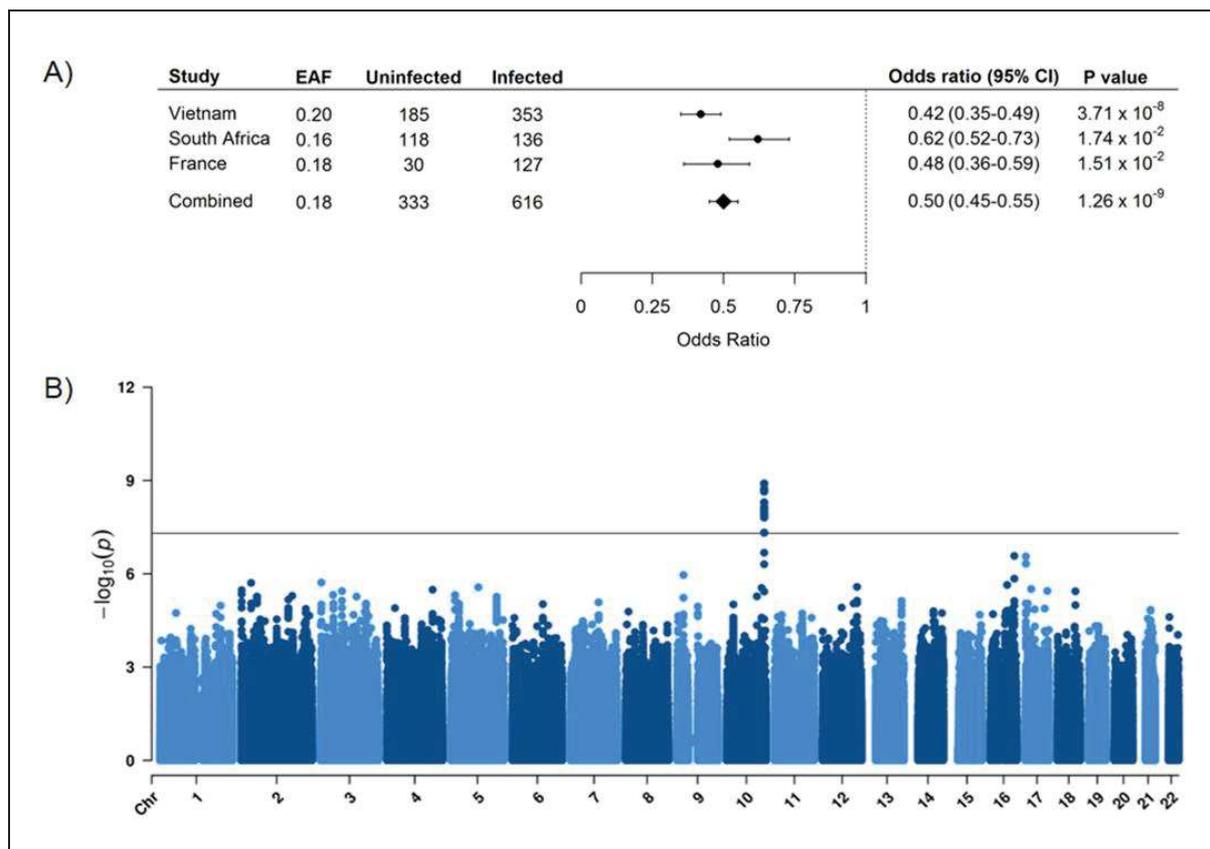


Figure 39. Méta-analyse de la résistance à l'infection par *Mtb* au Vietnam, en France et en Afrique du Sud (333 résistants et 616 infectés). **A)** Forest plot pour le marqueur rs17155120 génotypé dans les 3 cohortes. **B)** Manhattan plot pour 3 967 482 marqueurs.

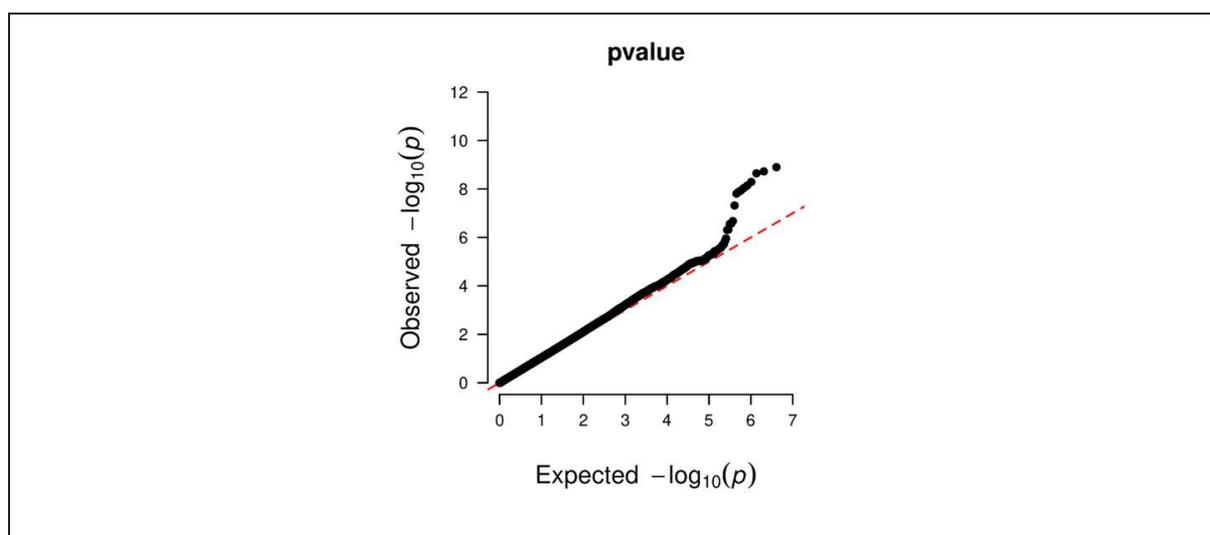


Figure 40. QQ-plot de la méta-analyse de la résistance à l'infection par *Mtb* ( $\lambda = 1.06$ )

Tableau 17. Marqueurs associés avec la résistance à l'infection par *Mtb* avec  $P < 5.0 \times 10^{-8}$  dans la méta-analyse (333 résistants et 616 infectés).

SNP	EA	OR (95%CI)	P value	$P_{het}$	Gène
rs17155120	T	0.50 (0.45-0.55)	$1.26 \times 10^{-9}$	0.31	<i>C10orf90</i>
rs118037357	A	0.47 (0.41-0.53)	$1.89 \times 10^{-9}$	0.50	<i>C10orf90</i>
rs77513326	A	0.50 (0.45-0.56)	$2.26 \times 10^{-9}$	0.29	<i>C10orf90</i>
rs79918233	A	0.51 (0.46-0.56)	$5.12 \times 10^{-9}$	0.24	<i>C10orf90</i>
rs17155143	A	0.53 (0.48-0.58)	$7.30 \times 10^{-9}$	0.10	<i>C10orf90</i>
rs28703703	G	0.55 (0.50-0.60)	$9.18 \times 10^{-9}$	0.11	<i>C10orf90</i>
rs79608098	T	0.55 (0.50-0.60)	$1.15 \times 10^{-8}$	0.14	<i>C10orf90</i>
rs75482972	A	0.55 (0.50-0.60)	$1.32 \times 10^{-8}$	0.11	<i>C10orf90</i>
rs61750007	C	0.56 (0.51-0.61)	$1.55 \times 10^{-8}$	0.12	<i>C10orf90</i>

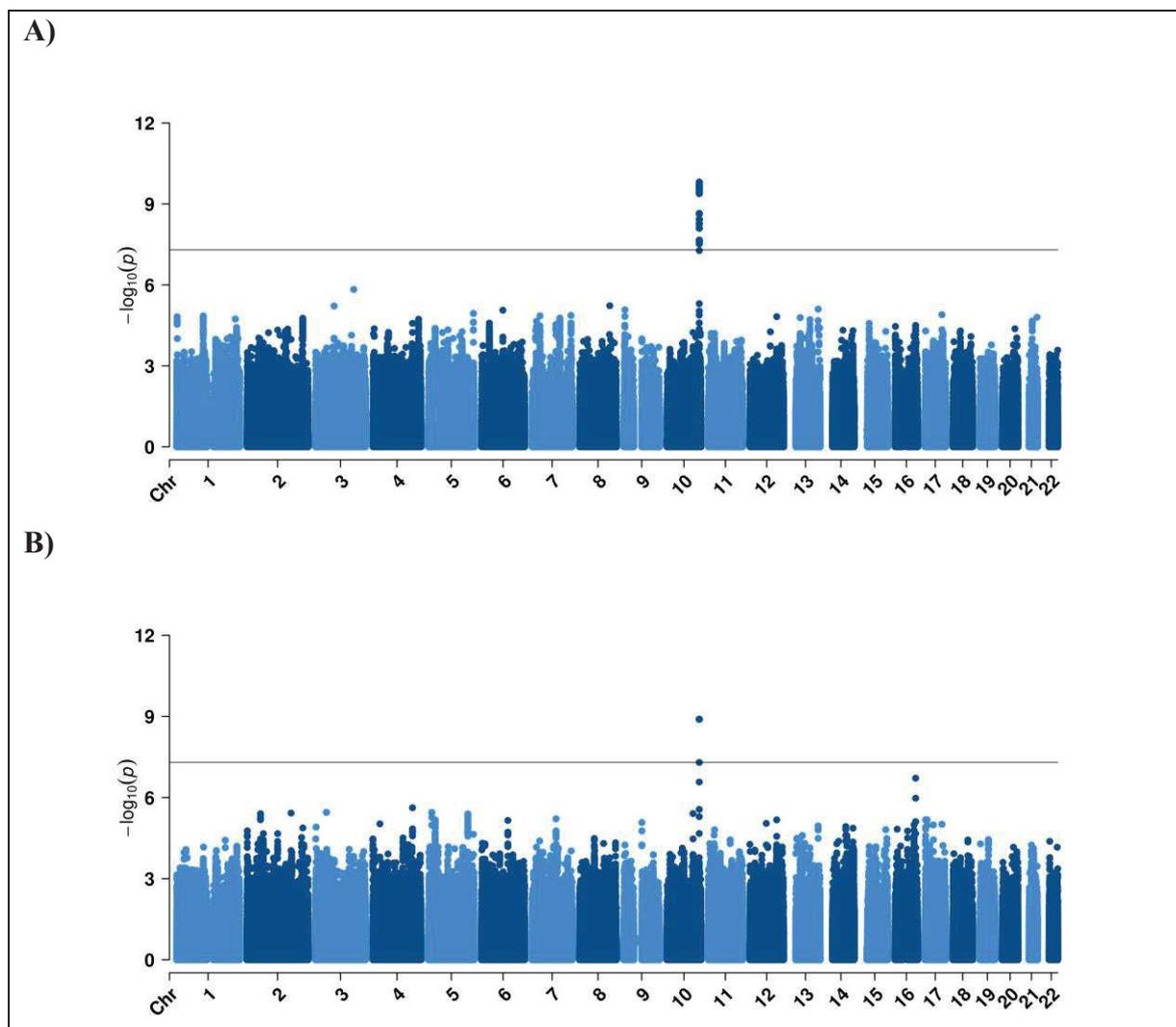


Figure 41. Manhatan plots des méta-analyses de la résistance à l'infection par *Mtb* par effets fixes (METAL) (A) et par effets aléatoires classiques (proposé dans METASOFT en plus de leur méthode améliorée) (B).

### 3.6 Analyse *in silico* et annotation fonctionnelle du locus 10q26.2

Une fois une association entre phénotype et marqueurs génétiques identifiée, il reste à élucider le mécanisme biologique sous-jacent à ce résultat [209],[210],[211]. Les GWASs ne sont pas une fin en soi puisqu'elles n'indiquent pas une relation de cause à effet entre le locus et le trait d'intérêt. On l'oublie vite, mais les GWASs sont censées décortiquer l'architecture génétique d'un trait/maladie, identifier statistiquement des marqueurs/gènes/cascades biologiques et donner des pistes pour des études fonctionnelles postérieures. Jusqu'à présent, les GWASs ont identifié des milliers de SNPs associés à des traits ou des maladies avec de nombreuses réplifications à travers cohortes et populations, validant la robustesse de ces observations statistiques. Cependant, seule une très petite fraction ont fait l'objet d'études fonctionnelles permettant de déterminer (i) quel(s) est(sont) le(s) polymorphisme(s) causal(aux), (ii) quelles sont les fonctions moléculaires du variant causal, (iii) quel gène est affecté et (iv) comment les changements dans la régulation ou la fonction du gène affecté influencent le phénotype [209]. Cette passerelle difficile à franchir entre association statistique et biologie est probablement la plus grande critique adressée aux GWASs. Il faut dire que la génétique humaine et la biologie moléculaire se sont bâties sur l'étude approfondie de gènes codants pour une protéine et comment des mutations pouvaient causer des maladies. La mutation était séquencée, la protéine formée changeait et sa fonction était annulée puis la supplémentation par un allèle sauvage restaurait cette fonction. À l'inverse des maladies mendéliennes, la génétique des maladies complexes est beaucoup moins directe que le modèle « un gène, une mutation, une conséquence ».

Les GWASs des maladies complexes ou d'un trait biologique peuvent identifier de nombreuses régions du génome associées au phénotype, chacune constituée de nombreux SNPs avec des *p-values* statistiquement significatives. La cartographie fine d'une région génomique, ou *fine-mapping*, consiste à déterminer le ou les SNP(s) responsable(s) d'un phénotype complexe, compte-tenu de l'association statistique observée dans une région du génome (Figure 42) [211]. En effet, l'association d'un locus avec un phénotype ne permet pas de distinguer le ou les SNP(s) causal (causaux) ni quel(s) gène(s) sera ou seront affectés : il existe plusieurs polymorphismes en LD avec le SNP présentant le plus fort effet statistique et formant ainsi un haplotype, duquel il est compliqué de déterminer lequel est le « vrai » polymorphisme fonctionnel. De plus, une grande majorité des SNPs associés avec un trait se situent dans des régions du génome qui ne codent pas une protéine, parfois très loin d'un gène codant [212].

Identifier le rôle de ces SNPs en régions non-codantes est devenue l’une des tâches les plus importantes et compliquées des analyses *in silico* « post-GWAS ». (Figure 43) [210].

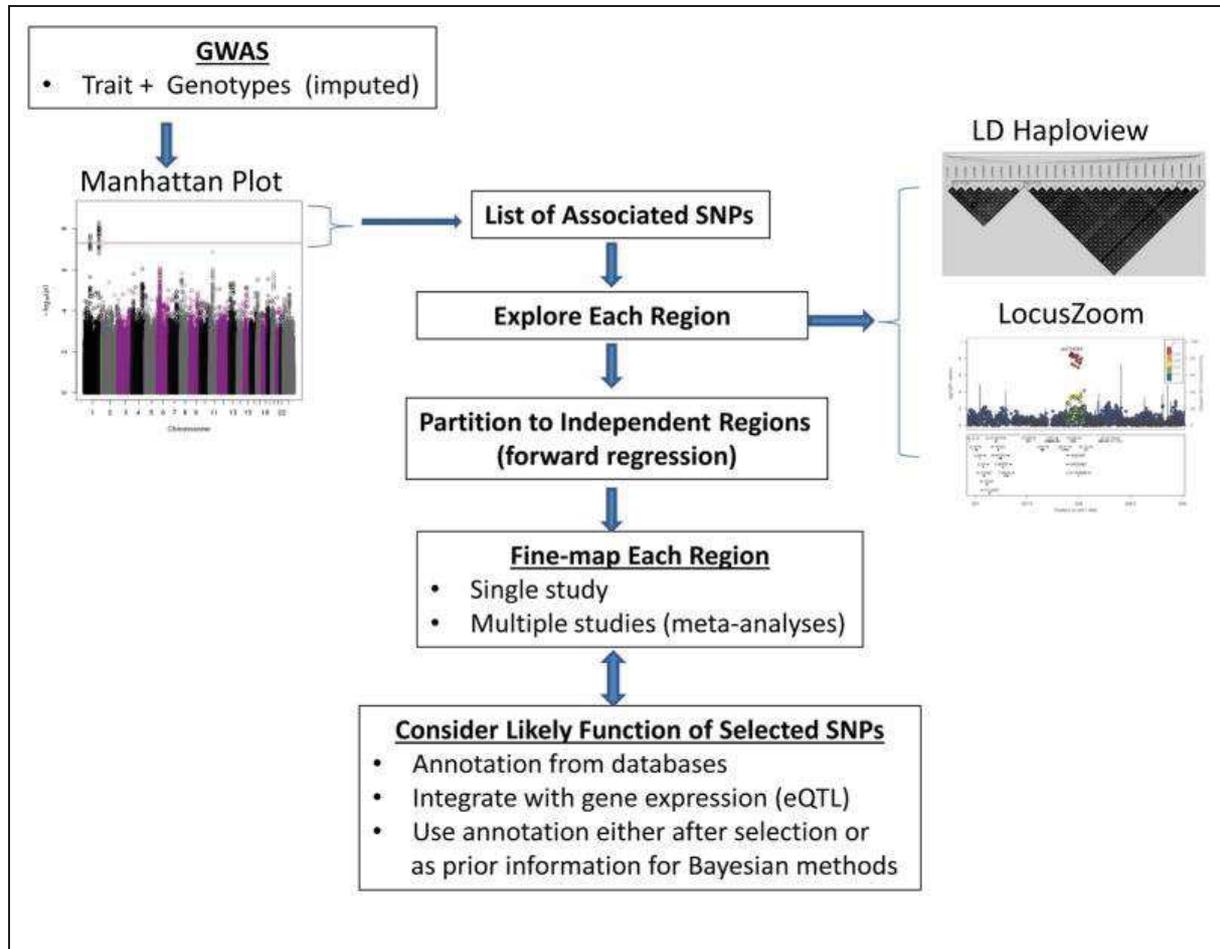


Figure 42. Diagramme d’une GWAS, de l’analyse d’association à l’annotation *in silico* des SNPs associés au phénotype d’intérêt [211]. À partir des *p-values* de l’association entre les SNPs et le phénotype, seuls ceux atteignant un seuil statistiquement significatif sont gardés pour des analyses post-GWAS. Chaque région est explorée en terme de déséquilibre de liaison et notamment comment les marqueurs sont corrélés avec le top SNP. Une cartographie fine de la région, *fine-mapping*, permet d’identifier un ou plusieurs signaux indépendants dans une même région puis de déterminer quel pourrait-être le(s) SNP(s) causal(causeux). Ces SNPs seront annotés avec des caractéristiques génomiques pour évaluer leur potentiel régulateur et prioriser des études fonctionnelles *in vivo* futures.

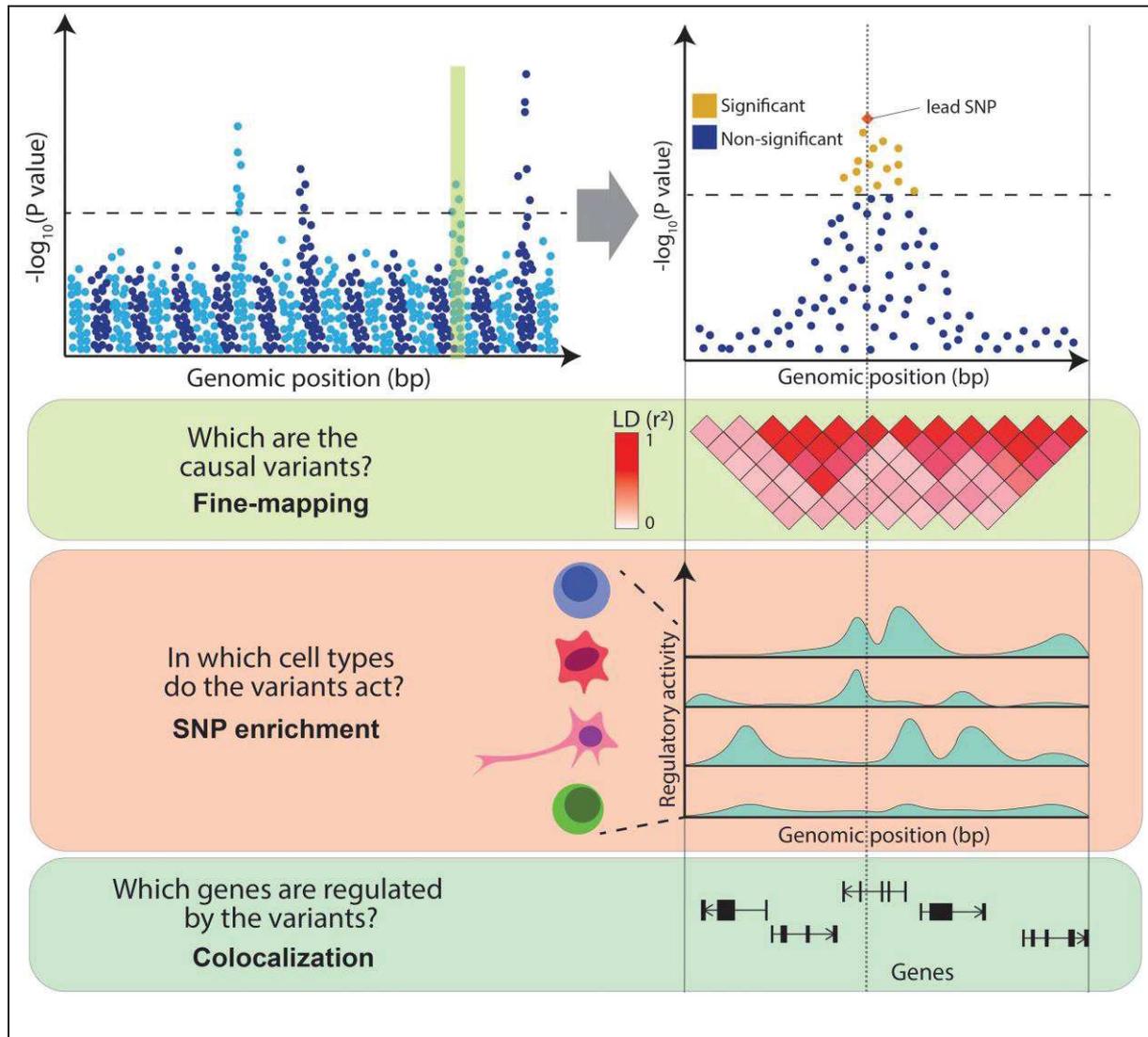


Figure 43. Analyses post-GWAS et leurs difficultés [210]. Une association entre des marqueurs génétiques (SNPs) et un trait/une maladie est mise en évidence par un Manhattan plot (en haut à gauche) et zoom sur un locus associé significativement (en haut à droite). Les polymorphismes identifiés sont corrélés les uns avec les autres et il est difficile de déterminer celui qui est responsable du phénotype observé. Ils peuvent avoir différentes fonctions régulatrices et dans différents types cellulaires.

La principale hypothèse est que ces polymorphismes régulent la transcription des gènes voisins dits cibles et changent leur niveau d'expression. De nombreuses études l'ont depuis confirmée, montrant que beaucoup de SNPs associés à un trait se concentraient dans des régions prédites pour être régulatrices par accessibilité à la chromatine, fixation de facteurs de transcription ou marqueurs d'histone. En outre, cet enrichissement de SNPs dans des régions cis-régulatrices a tendance à se produire dans des types cellulaires concernés par le trait/la maladie [210]. En accord avec ces observations, les SNPs des GWAS ont d'avantage tendance à chevaucher des loci de caractères quantitatifs d'expression (eQTLs pour expression

quantitative trait loci) et sont plus souvent associés avec les niveaux d'expression d'un ou plusieurs gènes par rapport à ce que l'on attendrait par le fait du hasard [213].

Une multitude d'approches *in silico* ont été développées ces dernières années afin de réduire la taille d'un locus associé à un nombre restreint de SNPs probablement causaux, qui, une fois, évalués pour leur potentielle action régulatrice sur des gènes cibles par une annotation extensive, serviront de bases aux futures études fonctionnelles *in vivo* [211]. Le *fine-mapping* s'est considérablement amélioré par (i) l'intégration des données statistiques venant de multiples cohortes, de tailles de plus en plus grandes, augmentant ainsi la puissance, (ii) l'inclusion de cohortes aux ascendances génétiques diverses, et notamment africaines qui présentent des blocs haplotypiques plus courts (au contraire de populations européennes et asiatiques) et (iii) l'utilisation d'annotations génomiques toujours plus larges grâce aux bases de données publiques.

Dans le cadre de notre étude, l'UCSC Genome Browser [214] a été interrogé pour déterminer si les SNPs associés à la résistance à l'infection par *Mtb* pouvaient chevaucher des régions du génome potentiellement régulatrices comme: des marqueurs d'histone catalogués par le projet ENCODE [215], des états de la chromatine ou de son accessibilité dans les cellules immunitaires [216],[217],[218]. Les SNPs ont également été recherchés dans les bases de données eQTL axées sur l'expression des gènes dans des monocytes, des macrophages, des lymphocytes T et d'autres types de cellules immunitaires [219],[220],[221],[222],[223].

Les marqueurs du locus 10q26.2 se situent en région intronique et légèrement en amont (5') du gène *C10orf90* (ou *FATS*, pour Fragile-site Associated Tumor Suppressor, HGNC: 26563). *C10orf90* est un gène suppresseur de tumeur localisé dans un site commun de fragilité chromosomale. Ces sites de fragilité sont des régions du génome très susceptibles à la dégradation de l'ADN et des zones à risque élevé de réarrangements chromosomiques conduisant à la tumorigénèse lors des cancers [224]. Ils sont aussi plus susceptibles de voir apparaître des variabilités de copies d'un gène [225]. La protéine codée par *C10orf90* participe à l'activation de p53 en réponse à une dégradation de l'ADN grâce à une activité d'ubiquitine ligase [226],[227]. Ce type de protéines est par lui-même intéressant car leur rôle dans la défense contre *Mtb* a déjà été démontré, notamment via des mécanismes d'autophagie [228],[229],[230]. Le facteur de transcription p53 n'est pas seulement un régulateur de l'autophagie mais aussi un activateur d'apoptose, un autre mécanisme biologique courant par lequel les cellules de l'hôte infecté limitent la propagation de pathogènes comme *Mtb* [231],[232]. Il a été démontré expérimentalement que l'apoptose induite par p53 jouait un rôle important dans l'inhibition de la survie de la mycobactérie et la résistance des macrophages

[233],[234]. Au vu de ces observations, *C10orf90* pourrait être un gène candidat attractif même si un rôle direct durant l'infection par *Mtb* ou tout autre processus immunitaire reste inconnu.

De fait, l'annotation des marqueurs du locus 10q26.2, visualisée en Figure 44, amènerait à penser que le gène *ADAM12*, voisin de *C10orf90* à ~280 kb en aval, pourrait être en réalité le gène cible.

Tout d'abord, selon les données épigénomiques d'ENCODE, certains SNPs du locus comme rs28703703 ou rs77513326 chevauchent une région caractérisée par des marqueurs d'histone H3K4me1 et H3K27ac (des modifications par méthylation ou acétylation), susceptibles de modifier leur interaction avec l'ADN pour le rendre plus ou moins accessible à la machinerie transcriptionnelle (en fonction du degré de modification et de son emplacement). Plus précisément H3K4me1 et H3K27ac ont tendance à favoriser la transcription et être associées à des fonctions amplificatrices [235],[236].

En phase avec ces modifications, cette région spécifique est aussi annotée comme un amplificateur actif dans les lymphocytes Th17. De plus, le marqueur rs77513326 chevauche aussi des pics d'ATAC-seq, caractéristiques de zones accessibles de la chromatine, dans des cellules immunitaires comme les lymphocytes Th17, les lymphocytes cytotoxiques naturels, les lymphocytes T CD8<sup>+</sup> ou les lymphocytes T mémoire.

Enfin, certains marqueurs, en particulier rs28703703 de nouveau ou le SNP génotypé rs17155120 étaient associés avec l'expression d'*ADAM12* dans les monocytes ( $P = 2.10 \times 10^{-3}$  et  $P = 4.70 \times 10^{-3}$  respectivement) (Figure 45). Par exemple, l'allèle T de rs17155120, celui protecteur contre l'infection par *Mtb*, est associé à une diminution des niveaux d'expression d'*ADAM12* avec un effet additif.

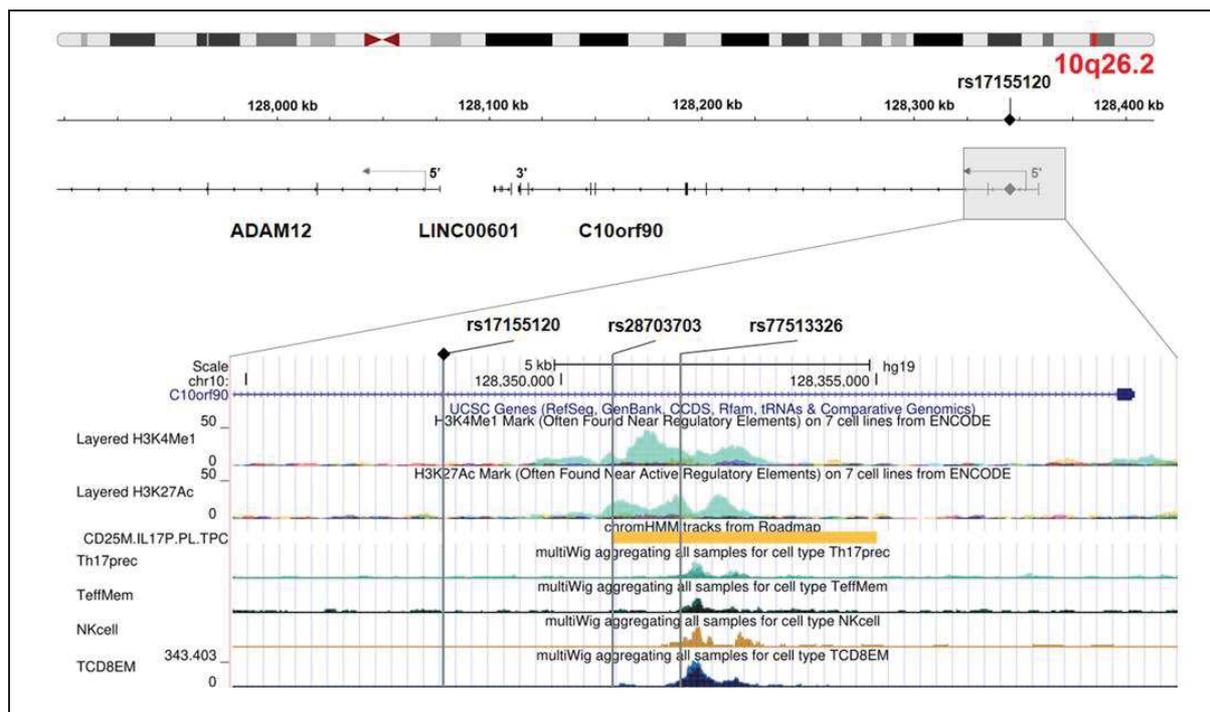


Figure 44. Annotation génomique du locus 10q26.2 dans l’UCSC Genome Browser avec de haut en bas: H3K4me1 et H3K27ac marqueurs d’histone (ENCODE), amplificateur dans les lymphocytes Th17 (chromHMM, ROADMAP), pics ATAC-seq dans différents types cellulaires (lymphocytes Th17, T mémoire, cytotoxiques naturels et T CD8<sup>+</sup>) [218].

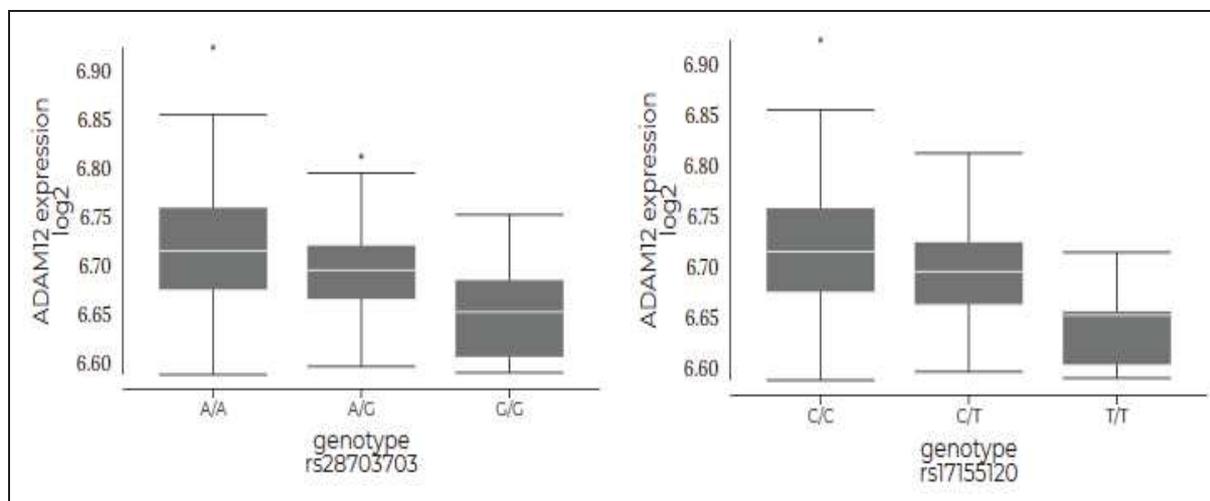


Figure 45. Niveaux d’expression d’*ADAM12* dans les monocytes selon les génotypes de 2 marqueurs associés à la résistance à l’infection par *Mtb* (<http://immunpop.com/kim/eQTL>).

*ADAM12* code une métalloprotéase matricielle impliquée dans de nombreux processus biologiques [237]. De façon très intéressante, l’expression d’*ADAM12* spécifiquement dans les lymphocytes Th17 a déjà été reportée [238],[239] et son *knockdown* dans des lymphocytes T humains entraîne une hausse de la production de cytokines Th17 (IL-17A, IL-17F et IL-22)

[238]. Si l'on revient à des observations purement épidémiologiques, des individus très exposés en Inde avec une IDR négative produisaient significativement plus de cytokines Th17 que les individus positifs [240]. De la même façon, en Gambie, des niveaux de production de ces cytokines de l'axe IL-17 étaient significativement plus élevés chez des individus IGRA négatif persistant par rapport à ceux convertissant positivement le test [241].

Ces observations corroborent nos résultats montrant que rs28703703 G ou rs17155120 T, protecteurs contre l'infection par *Mtb*, sont associés à une diminution des niveaux d'expression d'*ADAM12*, qui pourrait en cascade mener à une plus grande production de cytokines Th17. *In fine*, cela voudrait dire que ces cytokines ont un rôle protecteur important contre l'infection par *Mtb* à un stage précoce. Cette précision sur la probable nécessité d'une réponse Th17 à un stade précoce a son importance car, à l'autre bout du spectre de la tuberculose, le déclenchement de la maladie se caractérise entre autre par une réponse IL-17 exacerbée favorisant l'inflammation, la sécrétion de chémokines et le recrutement de neutrophiles [242],[243],[244].

Évidemment, cette partie repose beaucoup sur de l'interprétation et des suppositions, mais les possibles fonctions biologiques du locus 10q26.2, mis en évidence dans un premier temps par des méthodes purement statistiques, viennent renforcer son intérêt dans l'infection par *Mtb*. De façon intéressante, si la réponse par les lymphocytes Th17 est bel et bien impliquée, ce locus semble jouer un rôle dans l'infection par *Mtb* indépendamment de l'IFN- $\gamma$ . Cette cytokine est une pièce maitresse de la défense anti-mycobactérienne en général, et anti-tuberculeuse en particulier comme a pu le montrer l'étude des cas de tuberculose sévère chez l'enfant atteint d'une déficience de réponse immunitaire médiée par l'IL12 et l'IFN- $\gamma$ . Les individus résistants, eux, ont une production faible ou nulle d'IFN- $\gamma$  en présence d'antigènes mycobactériens alors qu'ils sont capables d'en produire (les tubes contrôles des IGRAs en attestent). Si l'on ajoute notre découverte à (i) celle de la GWAS en Ouganda/Tanzanie (individus VIH+ résistants), qui avait identifié *IL9* comme possible gène causal [136], et (ii) l'étude des profils immunologiques d'individus résistants sur le long terme, qui avait révélé une réponse humorale robuste et une forte immunité médiée par les lymphocytes T CD4<sup>+</sup> mais en l'absence d'IFN- $\gamma$  [39], alors cela confirme que des facteurs indépendants de la classique immunité par IFN- $\gamma$  et la réponse Th1 assurent une protection contre *Mtb*.

Découvrir le mécanisme moléculaire et valider fonctionnellement les allèles protecteurs en 10q26.2 ne sera pas chose aisée [245]. Les SNPs se situant en région non-codante, un test de gène rapporteur pourrait être utilisé pour évaluer l'impact des changements de la régulation sur l'expression du gène cible. L'élément régulateur est inséré dans un vecteur d'expression

comme un plasmide, en amont d'un promoteur qui induit l'expression d'une gène reporteur détectable comme la luciférase. La capacité et la force de l'élément régulateur à induire l'expression est ainsi mesurable dans des cellules transfectées avec le reporteur. Cependant, trouver un organisme modèle (comme des souris avec une perte de fonction de *C10orf90* ou *ADAM12*) ou un modèle de lignées cellulaires *in vitro* (de cellules souches pluripotentes induites à l'organoïde reproduisant l'épithélium des voies respiratoires) qui ressemblent au mieux au phénotype humain d'infection me semble quasi-impossible, notamment à cause du facteur d'exposition : si inoculer *Mtb* à faible dose rend compte du phénomène d'infection, quelle expérience permet de correctement modéliser une résistance sous exposition ? Enfin, d'autres approches par des technologies *omics* (transcriptome, protéome, séquençage du répertoire des récepteurs des lymphocytes T et B etc.) permettront sans doute d'apporter de nouveaux éléments de compréhension.

### **3.7 Réplication d'associations avec l'infection par *M. tuberculosis* précédemment identifiées**

Considérant les résultats de cette GWAS, j'ai cherché à répliquer les associations identifiées dans de précédentes études. Comme évoqué dans l'introduction, le phénotype d'infection par *Mtb* a d'abord été étudié par des études de liaison, et notamment 2 pics liés à la négativité de l>IDR avait été identifiés sur le chromosome 11 (11p14-15) dans 2 cohortes de France (locus majeur à 14.3 Mb), et d'Afrique du Sud (locus majeur à 26.4 Mb) [137],[138], à partir desquelles d'ailleurs se sont basées par la suite nos 2 cohortes de la GWAS. En examinant une fenêtre de 1 Mb autour de chaque pic, un groupe de SNPs en régions introniques des gènes *ARNTL* dans la cohorte française et *ANO3* et un lincRNA dans la cohorte sud-africaine (mais avec des effets en sens opposé) étaient associés avec  $P < 5.0 \times 10^{-3}$  (Tableau 18). Il est néanmoins difficile d'interpréter ces résultats, le sens réel de l'association initiale n'étant pas connu.

Tableau 18. Marqueurs associés avec la résistance à l'infection par *Mtb* en 11p14-15 (les 5 premiers marqueurs sont montrés). La direction indique celle de l'effet du marqueur (+ associé à la résistance et – inversement associé).

Cohorte	SNP	Chr:position	Allèle	MAF	Direction	P value	Gène
France	rs4757139	11:13300456	C	0.44	-	$2.2 \times 10^{-3}$	<i>ARNTL</i>
	rs4757140	11:13300540	G	0.44	-	$2.2 \times 10^{-3}$	<i>ARNTL</i>
	rs1351525	11:13301548	A	0.44	-	$2.2 \times 10^{-3}$	<i>ARNTL</i>
	rs1236189		C	0.44	-		<i>ARNTL</i>
	3	11:13301114				$2.2 \times 10^{-3}$	
	rs1481892	11:13301921	G	0.44	-	$2.2 \times 10^{-3}$	<i>ARNTL</i>
Afrique du Sud	rs3446365		T	0.23	-		<i>LINC02699</i>
	0	11:25684712				$7.2 \times 10^{-4}$	
	rs445875	11:26570372	C	0.34	+	$2.3 \times 10^{-3}$	<i>ANO3</i>
	rs1228777		T	0.39	-		<i>LINC02699</i>
	8	11:25453646				$2.8 \times 10^{-3}$	
	rs2641389	11:26565113	G	0.33	+	$4.0 \times 10^{-3}$	<i>ANO3</i>
	rs398615	11:26566441	C	0.33	+	$4.6 \times 10^{-3}$	<i>ANO3</i>

J'ai ensuite exploré une dizaine de SNPs associés (avec  $P < 5.0 \times 10^{-5}$ ) avec l'IDR de sujets VIH+ en Ouganda et Tanzanie [136], ainsi qu'un SNP du chromosome 2 associé avec une IDR négative de façon persistante [130]. Là aussi, cette investigation s'est révélée infructueuse (Tableau 19). La différence dans la définition du phénotype (uniquement l'IDR) et dans les populations étudiées (VIH+) peuvent expliquer cette absence de réplication dans notre GWAS.

Tableau 19. Réplication des marqueurs associés avec une IDR négative [136],[130] dans les 3 cohortes Vietnam, France et Afrique du Sud.

SNP	Chr.	$P_{\text{Vietnam}}$	$P_{\text{France}}$	$P_{\text{AfriqueDuSud}}$
rs877356	5	0.68	0.86	0.80
rs17062122	6	0.41	0.92	0.10
rs654718	11	0.60	0.92	0.53
rs7944514	11	0.89	0.94	0.96
rs7837658	8	0.43	0.61	0.34
rs8142256	22	0.16	0.21	0.51
rs7808481	7	0.48	0.66	0.03*
rs10998959	10	0.39	0.69	0.27
rs1880386	10	NA	0.57	0.71
rs11736841	4	0.68	0.14	0.45
rs7568133	2	0.42	0.68	0.86

\*effets en sens opposé

Enfin, une GWAS conduite en Islande avait notamment identifié 2 SNPs de la région HLA dont l'allèle mineur était associé avec l'infection par *Mtb* [135]. L'infection était uniquement définie par l'IDR (à un seuil non précisé) et l'association n'était significative au seuil pangénomique qu'en ajoutant de façon déséquilibrée des individus TB (plus de 8000) au groupe infecté (~6500). Le groupe non-infecté était lui issu de la population générale, non exposée. Ces SNPs n'étaient pas répliqués dans notre GWAS et par ailleurs la région HLA ne présentait pas de signal particulier (Figure 46).

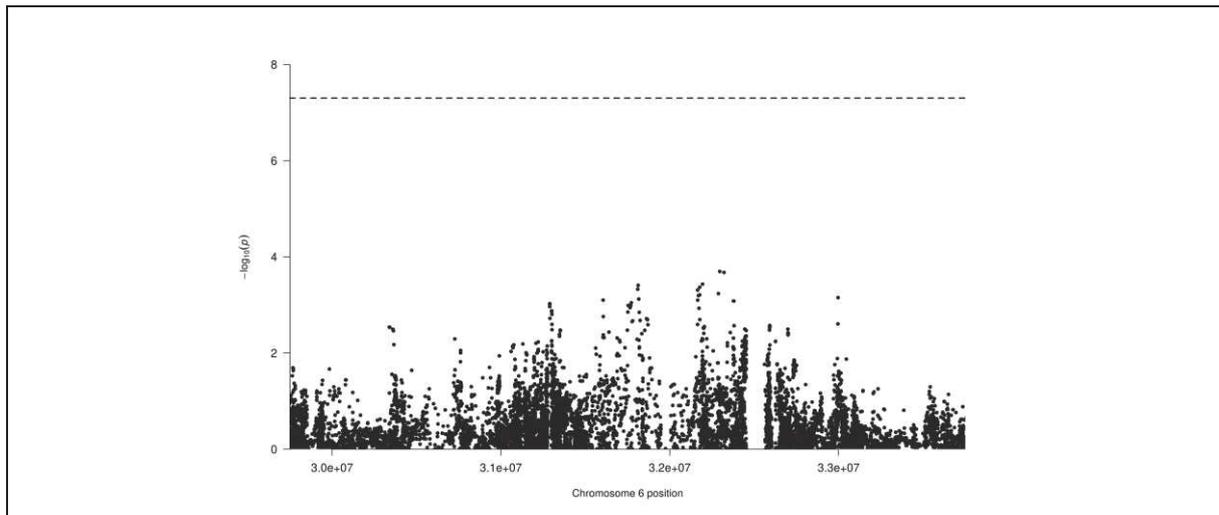


Figure 46. Méta-analyse de la résistance à l'infection par *Mtb* au Vietnam, en France et en Afrique du Sud (333 résistants et 616 infectés), région HLA.

### 3.8 Discussion

Cette deuxième partie s'est focalisée sur les facteurs de risque génétiques associés à la résistance naturelle à l'infection par *Mtb* après une exposition intense. Il n'existe pas de tests directs rendant compte d'une infection car l'IDR et l'IGRA ne mesurent en fait qu'une réponse immunitaire cellulaire adaptative déclenchée après la rencontre avec *Mtb*. Aucun de ces tests ne permet non plus de savoir si cette réponse immunitaire est le reflet d'une infection récente impliquant des bacilles dormants, contrôlés par le système immunitaire ou bien d'une infection plus ancienne sans bacilles vivants. L'IDR et l'IGRA ont chacun de nombreuses limitations [246]. Ainsi, pour éviter au maximum les erreurs de classification des individus résistants ou infectés, un phénotype a été défini très rigoureusement et de façon homogène en utilisant à la fois les tests IDR et IGRA dans les 3 cohortes.

Cependant, il est vrai que le *testing* transversal pourrait constituer une limite et seul un suivi longitudinal avec tests répétés permettrait de distinguer avec encore plus de précision les individus négatifs de manière persistente ou ceux convertissant un résultat négatif en positif. Le suivi longitudinal est donc attrayant mais il pose également des questions quant à la répétition des tests (variabilité des IGRAs, surtout ceux ayant une valeur de production d'IFN- $\gamma$  autour du seuil de positivité [147], ou bien effet booster concernant l'IDR [247],[248]) et la documentation de l'exposition pendant de longues périodes (est-ce qu'un contact toujours négatif 2 ans après le début de l'exposition à son cas index ne signifie pas simplement qu'il n'est plus exposé car l'index n'est plus contagieux après traitement ?). Dans l'une des rares études épidémiologiques longitudinales (plus de 2 ans) investigant la résistance à l'infection par *Mtb* chez des contacts familiaux en Ouganda, près de 90% des conversions se passaient dans les 6 mois après le début de l'exposition du contact à un cas index, et environ 75% dans les 3 mois [66].

L'exposition au bacille est difficilement quantifiable et pourtant est l'élément clé dans la conduite d'études sur l'infection [249]. Le meilleur moyen de prendre en compte ce paramètre reste probablement de se focaliser sur des contacts familiaux récemment exposés à un cas index contagieux pendant assez de temps pour pouvoir laisser l'opportunité à une infection de se produire (comme au Vietnam et en France) ou bien des individus vivants dans des zones où le risque de rencontrer *Mtb* est très élevé (comme en Afrique du Sud).

Le statut VIH des participants des 3 études n'était pas connu mais l'immunodéficience n'avait que peu de chance de biaiser les résultats compte tenu de la faible prévalence du VIH dans les populations étudiées et du fait que le VIH est surtout trouvé en présence de la tuberculose maladie, pas de l'infection (au moment des études, prévalence estimée à 0.4% en population générale au Vietnam et 5% chez les PTB [250],[163], 0.3% en population générale en France et 7% chez les PTB du Val-de-Marne [251],[252], 5% dans la zone concernée du Cap en Afrique du Sud et seulement 2% chez les enfants [138]). De plus, le contrôle qualité des tubes mitogènes des IGRAs était conforme à ce que l'on attend d'individus immunocompétents.

La collection de données permettant de quantifier l'exposition via un score de risque d'infection robuste et adapté à chaque environnement (endémicité, âge des participants etc.) devrait accompagner un suivi longitudinal. La nature du cas index (une mère au contact de ses enfants et son époux ou un parent éloigné ne vivant dans le foyer que quelques heures par jour), son niveau de contagiosité (durée de la toux, présence de crachats positifs ou niveau de charge bactérienne dans les aérosols), degré de ventilation seront autant de facteurs à intégrer dans les futures études épidémiologiques.

Dans ce travail, une association significative entre un groupe de marqueurs sur le chromosome 10 et la résistance à l'infection par *Mtb* a pu être mise en évidence dans une cohorte principale spécifiquement établie pour l'occasion. Ce locus a pu être en outre répliqué dans 2 autres cohortes indépendantes, établies dans un cadre épidémiologique différent en terme d'endémicité (haute au Vietnam, très basse en France, et très haute en Afrique du Sud), du type de participants (adultes au Vietnam, très jeunes en Afrique du Sud, les 2 en France) ou du type d'exposition. Ce dernier facteur explique peut-être la réplication moins prononcée en Afrique du Sud où l'exposition des sujets était communautaire et non familiale. Quoiqu'il en soit, une association robuste (à travers les populations et la multitude d'analyses possibles, toutes non présentées, suivant la définition du phénotype, les méthodes employées, l'ajustement à des facteurs d'exposition etc.) était mise en évidence entre un marqueur génotypé rs17155120 et l'infection par *Mtb* avec un OR combiné de 0.50 (95%CI 0.45-0.55) pour l'allèle mineur T protecteur.

Les 3 GWASs étaient de taille modeste, ce qui n'a pas empêché de détecter cette association significative avec un large effet. Mais l'ajout de patients PTB dans la cohorte initiale du Vietnam a permis d'augmenter la puissance. De telles tailles d'effet ne sont pas communes pour la plupart des GWASs excepté dans le contexte des maladies infectieuses. Le SNP identifié dans la GWAS de réactivité à l>IDR chez les individus VIH+ en Ouganda avait un OR de 0.29

[136], un autre SNP associé à une clairance du virus de l'hépatite C avait un OR de 0.45 [253], ou encore une GWAS de susceptibilité à l'infection par *Cryptosporidium* trouvait un SNP associé avec un OR de 2.4 [254]. Ces tailles d'effets sont surtout dues aux petits effectifs et à une sélection rigoureuse des cas et des contrôles. En outre, on a pu constater que le SNP rs17155120 était présent avec une fréquence allélique et dans une région au LD similaire dans les 3 populations de cette étude, aux ascendances génétiques diverses et variées (vietnamiens Kinh, européens, nord-africains, africains sub-sahariens, « Coloured » du Cap). De telles populations sont ultra-minoritaires dans la recherche en génétique où 80% des participants sont européens [255]. Ce biais commence, enfin, à être reconnu [256],[257], et nul doute que leur inclusion apporteront de nombreux réponses et de nouvelles perspectives dans la compréhension des maladies complexes [258]. À ce titre, il est dommageable de ne pas avoir davantage de moyens mis à disposition dans le contexte des maladies infectieuses, et notamment la tuberculose, qui touchent en priorité des populations défavorisées et sous-représentées.

Cela m'amène à conclure ce travail en évoquant des perspectives possibles dans l'étude de la résistance naturelle à l'infection par *Mtb*. Évidemment plus de populations diverses et en plus grand nombre seront nécessaires. On le constate avec les tailles d'effectif dans notre étude : pour pouvoir analyser  $x$  individus que l'on définirait comme résistant, il faut recruter  $\sim 10x$  plus de participants. Ajouter à cela la logistique pour procéder au testing et un suivi longitudinal, à une documentation exhaustive de l'exposition et d'autres covariables, on comprend qu'il est difficile d'établir de telles cohortes, *a fortiori* dans des pays en voie de développement. Les études de contacts familiaux ont en cela l'avantage de faciliter le recrutement des participants et la collection d'un spectre complet de phénotypes tout en contrôlant l'environnement.

Les facteurs génétiques explorés dans ce travail concernaient des polymorphismes fréquents en population générale. Ici, la notion de résistance est différente de celle rencontrée chez les individus homozygotes pour *DARC* ou *CCR5* complètement résistants à la malaria ou au VIH [259],[260]. Il sera intéressant d'étudier des polymorphismes rares ou structuraux qui pourraient conférer une résistance innée de type mendélienne à l'infection. Là encore, l'inclusion de populations isolées, de larges familles ou d'individus au phénotype extrême comme des individus immunodéficients mais constamment non-infectés [51] apporteront sans doute de nouveaux éclairages. Une récente étude menée au Vietnam identifiait 85 maladies récessives à partir d'un échantillon de moins de 1 000 exomes, démontrant la grande prévalence de ce genre de maladies, très probablement à cause d'effets fondateurs [261]. De larges

populations, allant jusqu'à 1 million d'individus, descendantes d'événements fondateurs extrêmes, sont par exemple présentes en grand nombre sur le sous-continent indien, l'une des régions du monde la plus touchée par la tuberculose [262]. De nombreuses découvertes viendront sans nul doute de leurs études.

La longue histoire entre l'Homme et *Mtb* a probablement mené à des signatures de coévolution. De récentes approches intégrant à la fois les génomes de l'hôte et du pathogène dans le contexte des maladies virales VIH et hépatite C ont identifié des interactions indiquant que des polymorphismes de l'hôte au niveau de HLA ou *IFNL4* respectivement exerçaient une forte pression sur les génomes des virus [263],[264]. Deux récentes études ont tenté d'intégrer la diversité de *Mtb* avec des approches différentes. Dans la première, les auteurs ont réalisé une GWAS chez des PTB par une régression logistique multinomiale, la variable dépendante (d'habitude cas/contrôle) étant définie par l'occurrence d'une lignée *Mtb*, mais sans résultats probants [265]. Dans la deuxième, les auteurs faisaient l'hypothèse que la coévolution entre humains et *Mtb* conduit à une maladie moins sévère quand l'hôte est porteur de l'allèle ancestral et en même temps infecté par la lignée historique de *Mtb* présente dans la région, dans leur cas, une sous lignée L4.6 Ouganda. En génotypant quelques gènes candidats et en codant le phénotype de façon binaire (L4.6 Ouganda versus autres lignées), une interaction entre un SNP de *SLCA11* (*NRAMP1*) et *Mtb* a pu être mise en évidence [266]. Des analyses à l'échelle des génomes entiers posent des défis très intéressants à relever (tests multiples, structures des populations etc.). Concernant le phénotype d'infection, il est moins sûr d'en attendre quelque chose, notamment car il est impossible de mettre en culture *Mtb* chez des individus asymptomatiques.

Une alternative serait d'affecter la lignée/le génome *Mtb* du sujet infecté à son cas index, ce qui serait possible dans les régions faiblement endémiques où l'on est certain de la chaîne de transmission. S'il est impossible de détecter *Mtb* dans des crachats d'un individu infecté sain, il serait néanmoins possible de récupérer des bacilles viables à partir de prélèvements de tissu pulmonaire lors d'autopsies de personnes décédées de cause non-tuberculeuse ou d'identifier de l'ADN de *Mtb* persistant par PCR [53],[56]. En collectant des séries d'échantillons dans des régions endémiques, cela permettrait (i) d'avoir une estimation de la prévalence réelle de l'infection par *Mtb* et (ii) de classer correctement les individus pour une étude d'association en s'affranchissant des résultats des tests. À considérer la résistance à *Mtb* comme un phénotype quantitatif relatif au niveau d'exposition (passé un certain seuil, l'infection est inévitable), les

individus auraient en plus la caractéristique d'avoir eu une exposition cumulée tout au long de leur vie. La quantité de bacilles recouvrée pourrait être aussi analysée en trait quantitatif.

Enfin, à défaut de travailler sur le phénotype d'infection *per se*, il serait intéressant d'explorer les résultats d'IDR et d'IGRAs en traits quantitatifs. Les premières études d'héritabilité et de liaison, notamment dans le laboratoire, ont été réalisées sur ces traits quantitatifs et peuvent nous renseigner sur les processus plus directement liés à la réponse immunitaire à *Mtb* et son intensité. Un exemple en cours est donné Figure 47 et permet de mettre en évidence une association de la région HLA avec la production d'IFN- $\gamma$ . Le phénotype binaire de la GWAS n'avait montré aucune implication de la région HLA alors qu'on aurait pu s'attendre à ce que le complexe majeur d'histocompatibilité de classe II contribue à la résistance à l'infection en réduisant par exemple la présentation des antigènes *Mtb* aux lymphocytes T.

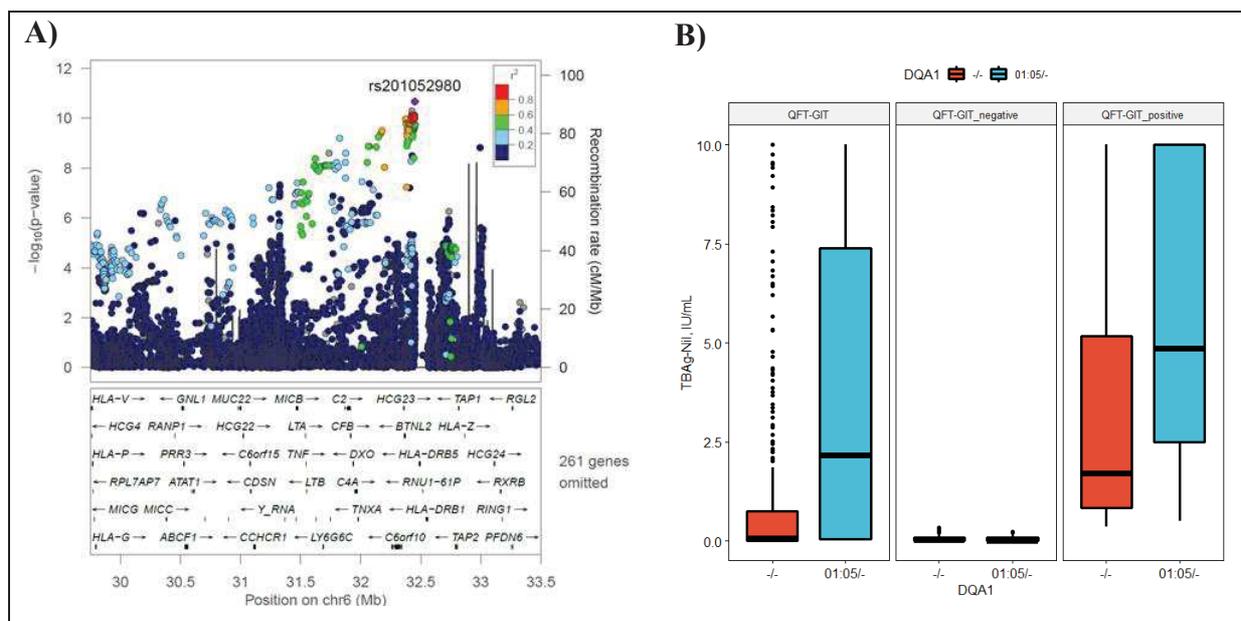


Figure 47. Étude d'association du QFT en trait quantitatif (TB Antigen corrigé par la valeur nil) dans la cohorte du Vietnam. **A)** Locus zoom de la région HLA où un signal fort est mis en évidence. **B)** L'imputation des allèles HLA a permis d'affiner l'association à DQA1\*01:05. Les porteurs d'un allèle DQA1\*01:05 ont des niveaux de production d'IFN- $\gamma$  supérieurs aux non-porteurs (panel de gauche). De façon intéressante, cette différence est surtout observable chez les individus avec un résultat positif (panel de droite) alors que les individus ayant un test négatif ne présentent pas de différence (panel du milieu).

## 4 Conclusion

Près de 140 ans après la découverte de l'agent responsable de la tuberculose, la maladie demeure un problème de santé publique majeur. L'infection tuberculeuse constitue le principal réservoir de cette pandémie et son diagnostic repose sur 2 types de tests, l'intradermoréaction (IDR) ou le test de libération d'interferon-gamma (IGRA), dont les performances varient suivant leur contexte d'utilisation. En particulier, les facteurs de risque non génétiques peuvent confondre leur interprétation. Par ailleurs, il existe une grande variabilité de réponse à l'infection par *Mtb*, et le fait que certains individus soient résistants malgré une longue et intense exposition au bacille suggèrent l'implication de facteurs génétiques associés à ce phénotype complexe.

Dans une première partie d'épidémiologie classique, je me suis focalisé sur le personnel de santé à risque d'infection. Un large échantillon d'employés, qui ont été testés à la fois par IDR et IGRA, a été recruté au Maroc. Une forte prévalence de l'infection a pu être observée (de 40% à 50% selon le test utilisé). Le sexe masculin, l'âge, un antécédent familial de tuberculose, travailler en pneumologie et la consommation de tabac étaient des facteurs de risque associés à l'infection selon différentes définitions. Cette étude a confirmé que les résultats d'IDR et d'IGRA ne sont pas pleinement concordants et que, en pratique, un diagnostic d'infection doit tenir compte des propriétés de chaque test et du contexte dans lequel il est réalisé.

Dans une deuxième partie, je me suis intéressé aux facteurs génétiques associés à l'infection. L'incertitude du statut infectieux a été contournée en maximisant l'exposition des individus recrutés et en définissant un phénotype strict basé sur les 2 types de tests. Par une approche de GWAS, j'ai pu identifier une association entre la résistance à l'infection et des polymorphismes communs sur le chromosome 10 dans une cohorte initiale vietnamienne de contacts familiaux fortement exposés à un proche tuberculeux contagieux. À partir de ce même phénotype très spécifique, j'ai pu répliquer ce résultat dans 2 autres cohortes indépendantes de France et d'Afrique du Sud aux caractéristiques génétiques et épidémiologiques différentes. Les SNPs sont situés en région intronique du gène *C10orf90* (ou *FATS*) codant une ubiquitine ligase. Les allèles protecteurs contre l'infection sont associés à une baisse d'expression du gène voisin *ADAMI2*, et pourraient conduire à une réponse exacerbée de lymphocytes Th17.

Le phénotype d'infection par *Mtb* est largement moins étudié que la susceptibilité à la maladie. Avec le manque crucial d'un vaccin efficace et la propagation de souches résistantes aux antibiotiques, la résistance naturelle à *Mtb* constitue un phénotype extrême dont la compréhension pourrait révéler les corrélats de l'immunité et de la protection transposables dans la recherche d'un vaccin ou de nouvelles thérapies ciblées par l'hôte. De larges études intégrant épidémiologie rigoureuse des populations incluses, génétique humaine et génomique bactérienne, immunologie et génétique fonctionnelle devront être menées dans ce combat pour éradiquer *Mtb*.

# Bibliographie

1. Koch R. Die Ätiologie der Tuberkulose. doi:10.25646/5088
2. WHO | Global tuberculosis report 2019. In: WHO [Internet]. Available: [http://www.who.int/tb/publications/global\\_report/en/](http://www.who.int/tb/publications/global_report/en/)
3. Barberis I, Bragazzi NL, Galluzzo L, Martini M. The history of tuberculosis: from the first historical records to the isolation of Koch's bacillus. *J Prev Med Hyg.* 2017;58: E9–E12.
4. Nakojimo H, Brown P, Nunn P, Kochi A, Porter J, McAdam K, et al. Tuberculosis: a global emergency 3. *World Health.* 1993; 32.
5. Uplekar M, Weil D, Lonroth K, Jaramillo E, Lienhardt C, Dias HM, et al. WHO's new end TB strategy. *Lancet.* 2015;385: 1799–1801. doi:10.1016/S0140-6736(15)60570-0
6. Pai M, Behr MA, Dowdy D, Dheda K, Divangahi M, Boehme CC, et al. Tuberculosis. *Nature Reviews Disease Primers.* 2016;2: 16076. doi:10.1038/nrdp.2016.76
7. Subbaraman R, Nathavitharana RR, Satyanarayana S, Pai M, Thomas BE, Chadha VK, et al. The Tuberculosis Cascade of Care in India's Public Sector: A Systematic Review and Meta-analysis. *PLOS Medicine.* 2016;13: e1002149. doi:10.1371/journal.pmed.1002149
8. Frieden TR, Fujiwara PI, Washko RM, Hamburg MA. Tuberculosis in New York City — Turning the Tide. *New England Journal of Medicine.* 1995;333: 229–233. doi:10.1056/NEJM199507273330406
9. Suárez PG, Watt CJ, Alarcón E, Portocarrero J, Zavala D, Canales R, et al. The dynamics of tuberculosis in response to 10 years of intensive control effort in Peru. *J Infect Dis.* 2001;184: 473–478. doi:10.1086/322777
10. Wang L, Zhang H, Ruan Y, Chin DP, Xia Y, Cheng S, et al. Tuberculosis prevalence in China, 1990–2010; a longitudinal analysis of national survey data. *The Lancet.* 2014;383: 2057–2064. doi:10.1016/S0140-6736(13)62639-2
11. Van Der Meeren O, Hatherill M, Nduba V, Wilkinson RJ, Muyoyeta M, Van Brakel E, et al. Phase 2b placebo-controlled trial of M72/AS01E candidate vaccine to prevent active tuberculosis in adults. *N Engl J Med.* 2018;379: 1621–1634. doi:10.1056/NEJMoa1803484
12. Nemes E, Geldenhuys H, Rozot V, Rutkowski KT, Ratangee F, Bilek N, et al. Prevention of *M. tuberculosis* Infection with H4:IC31 Vaccine or BCG Revaccination. *New England Journal of Medicine.* 2018;379: 138–149. doi:10.1056/NEJMoa1714021

13. Darrah PA, Zeppa JJ, Maiello P, Hackney JA, Wadsworth MH, Hughes TK, et al. Prevention of tuberculosis in macaques after intravenous BCG immunization. *Nature*. 2020;577: 95–102. doi:10.1038/s41586-019-1817-8
14. Swindells S, Ramchandani R, Gupta A, Benson CA, Leon-Cruz J, Mwelase N, et al. One Month of Rifapentine plus Isoniazid to Prevent HIV-Related Tuberculosis. *N Engl J Med*. 2019;380: 1001–1011. doi:10.1056/NEJMoa1806808
15. Conradie F, Diacon AH, Ngubane N, Howell P, Everitt D, Crook AM, et al. Treatment of Highly Drug-Resistant Pulmonary Tuberculosis. *N Engl J Med*. 2020;382: 893–902. doi:10.1056/NEJMoa1901814
16. Nimmo C, Millard J, van Dorp L, Brien K, Moodley S, Wolf A, et al. Population-level emergence of bedaquiline and clofazimine resistance-associated variants among patients with drug-resistant tuberculosis in southern Africa: a phenotypic and phylogenetic analysis. *The Lancet Microbe*. 2020;1: e165–e174. doi:10.1016/S2666-5247(20)30031-8
17. Hogan AB, Jewell BL, Sherrard-Smith E, Vesga JF, Watson OJ, Whittaker C, et al. Potential impact of the COVID-19 pandemic on HIV, tuberculosis, and malaria in low-income and middle-income countries: a modelling study. *Lancet Glob Health*. 2020;8: e1132–e1141. doi:10.1016/S2214-109X(20)30288-6
18. Pai M. Covidization of research: what are the risks? *Nature Medicine*. 2020;26: 1159–1159. doi:10.1038/s41591-020-1015-0
19. Brites D, Gagneux S. Co-evolution of *Mycobacterium tuberculosis* and *Homo sapiens*. *Immunological Reviews*. 2015;264: 6–24. doi:10.1111/imr.12264
20. Gagneux S. Ecology and evolution of *Mycobacterium tuberculosis*. *Nature Reviews Microbiology*. 2018;16: 202–213. doi:10.1038/nrmicro.2018.8
21. Ngabonziza JCS, Loiseau C, Marceau M, Jouet A, Menardo F, Tzfidia O, et al. A sister lineage of the *Mycobacterium tuberculosis* complex discovered in the African Great Lakes region. *Nature Communications*. 2020;11: 2917. doi:10.1038/s41467-020-16626-6
22. Coscolla M, Brites D, Menardo F, Loiseau C, Borrell S, Otchere ID, et al. Phylogenomics of *Mycobacterium africanum* reveals a new lineage and a complex evolutionary history. *bioRxiv*. 2020; 2020.06.10.141788. doi:10.1101/2020.06.10.141788
23. Hershberg R, Lipatov M, Small PM, Sheffer H, Niemann S, Homolka S, et al. High Functional Diversity in *Mycobacterium tuberculosis* Driven by Genetic Drift and Human Demography. *PLOS Biology*. 2008;6: e311. doi:10.1371/journal.pbio.0060311
24. Comas I, Coscolla M, Luo T, Borrell S, Holt KE, Kato-Maeda M, et al. Out-of-Africa migration and Neolithic co-expansion of *Mycobacterium tuberculosis* with modern humans. *Nat Genet*. 2013;45: 1176–1182. doi:10.1038/ng.2744
25. 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;514: 494–497. doi:10.1038/nature13591

26. Sabin S, Herbig A, Vågane ÅJ, Ahlström T, Bozovic G, Arcini C, et al. A seventeenth-century *Mycobacterium tuberculosis* genome supports a Neolithic emergence of the *Mycobacterium tuberculosis* complex. *Genome Biology*. 2020;21: 201. doi:10.1186/s13059-020-02112-1
27. Houben RMGJ, Dodd PJ. The Global Burden of Latent Tuberculosis Infection: A Re-estimation Using Mathematical Modelling. *PLOS Medicine*. 2016;13: e1002152. doi:10.1371/journal.pmed.1002152
28. Meehan CJ, Goig GA, Kohl TA, Verboven L, Dippenaar A, Ezewudo M, et al. Whole genome sequencing of *Mycobacterium tuberculosis* : current standards and open issues. *Nature Reviews Microbiology*. 2019;17: 533–545. doi:10.1038/s41579-019-0214-5
29. Comas I, Chakravarti J, Small PM, Galagan J, Niemann S, Kremer K, et al. Human T cell epitopes of *Mycobacterium tuberculosis* are evolutionarily hyperconserved. *Nat Genet*. 2010;42: 498–503. doi:10.1038/ng.590
30. Mitchison DA, Wallace JG, Bhatia AL, Selkon JB, Subbaiah TV, Lancaster MC. A comparison of the virulence in guinea-pigs of South Indian and British tubercle bacilli. *Tubercle*. 1960;41: 1–22. doi:10.1016/s0041-3879(60)80019-0
31. 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;198: 1037–1043. doi:10.1086/591504
32. Holt KE, McAdam P, Thai PVK, Thuong NTT, Ha DTM, Lan NN, et al. Frequent transmission of the *Mycobacterium tuberculosis* Beijing lineage and positive selection for the EsxW Beijing variant in Vietnam. *Nat Genet*. 2018;50: 849–856. doi:10.1038/s41588-018-0117-9
33. Drain PK, Bajema KL, Dowdy D, Dheda K, Naidoo K, Schumacher SG, et al. Incipient and Subclinical Tuberculosis: a Clinical Review of Early Stages and Progression of Infection. *Clin Microbiol Rev*. 2018;31. doi:10.1128/CMR.00021-18
34. Behr MA, Edelstein PH, Ramakrishnan L. Is *Mycobacterium tuberculosis* infection life long? *BMJ*. 2019;367. doi:10.1136/bmj.l5770
35. Cardona P-J. Reactivation or reinfection in adult tuberculosis: Is that the question? *Int J Mycobacteriol*. 2016;5: 400–407. doi:10.1016/j.ijmyco.2016.09.017
36. 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. doi:10.1146/annurev-immunol-032712-095939
37. 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;264: 103–120. doi:10.1111/imr.12272
38. Lu LL, Chung AW, Rosebrock T, Ghebremichael M, Yu WH, Grace PS, et al. A functional role for antibodies in tuberculosis. *Cell*. 2016;167: 433-443.e14. doi:10.1016/j.cell.2016.08.072

39. Lu LL, Smith MT, Yu KKQ, Luedemann C, Suscovich TJ, Grace PS, et al. IFN- $\gamma$ -independent immune markers of Mycobacterium tuberculosis exposure. *Nat Med*. 2019;25: 977–987. doi:10.1038/s41591-019-0441-3
40. Goldstein EJC, Lee E, Holzman RS. Evolution and Current Use of the Tuberculin Test. *Clin Infect Dis*. 2002;34: 365–370. doi:10.1086/338149
41. Lewinsohn DM, Leonard MK, LoBue PA, Cohn DL, Daley CL, Desmond E, et al. Official American Thoracic Society/Infectious Diseases Society of America/Centers for Disease Control and Prevention Clinical Practice Guidelines: Diagnosis of Tuberculosis in Adults and Children. *Clinical Infectious Diseases*. 2017;64: e1–e33. doi:10.1093/cid/ciw694
42. 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*. 2006;10: 1192–1204.
43. Menzies D. What does tuberculin reactivity after bacille Calmette-Guérin vaccination tell us? *Clin Infect Dis*. 2000;31 Suppl 3: S71-74. doi:10.1086/314075
44. 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;27: 3–20. doi:10.1128/CMR.00034-13
45. Nemes E, Rozot V, Geldenhuys H, Bilek N, Mabwe S, Abrahams D, et al. Optimization and Interpretation of Serial QuantiFERON Testing to Measure Acquisition of Mycobacterium tuberculosis Infection. *Am J Respir Crit Care Med*. 2017;196: 638–648. doi:10.1164/rccm.201704-0817OC
46. WHO | Guidelines on the management of latent tuberculosis infection. In: WHO [Internet]. Available: <http://www.who.int/tb/publications/latent-tuberculosis-infection/en/>
47. Ermann J, Rao DA, Teslovich NC, Brenner MB, Raychaudhuri S. Immune cell profiling to guide therapeutic decisions in rheumatic diseases. *Nat Rev Rheumatol*. 2015;11: 541–551. doi:10.1038/nrrheum.2015.71
48. Möller M, Kinnear CJ, Orlova M, Kroon EE, van Helden PD, Schurr E, et al. Genetic Resistance to Mycobacterium tuberculosis Infection and Disease. *Front Immunol*. 2018;9: 2219. doi:10.3389/fimmu.2018.02219
49. Simmons JD, Stein CM, Seshadri C, Campo M, Alter G, Fortune S, et al. Immunological mechanisms of human resistance to persistent Mycobacterium tuberculosis infection. *Nat Rev Immunol*. 2018;18: 575–589. doi:10.1038/s41577-018-0025-3
50. Verrall AJ, Alisjahbana B, Apriani L, Novianty N, Nurani AC, van Laarhoven A, et al. Early Clearance of Mycobacterium tuberculosis: The INFECT Case Contact Cohort Study in Indonesia. *J Infect Dis*. 2020;221: 1351–1360. doi:10.1093/infdis/jiz168
51. Kroon EE, Kinnear CJ, Orlova M, Fischinger S, Shin S, Boolay S, et al. An observational study identifying highly tuberculosis-exposed, HIV-1-positive but persistently TB, tuberculin and IGRA negative persons with M. tuberculosis specific antibodies in Cape

- Town, South Africa. medRxiv. 2020; 2020.07.07.20147967. doi:10.1101/2020.07.07.20147967
52. Li H, Wang X, Wang B, Fu L, Liu G, Lu Y, et al. Latently and uninfected healthcare workers exposed to TB make protective antibodies against *Mycobacterium tuberculosis*. *Proc Natl Acad Sci U S A*. 2017;114: 5023–5028. doi:10.1073/pnas.1611776114
  53. Dutta NK, Karakousis PC. Latent Tuberculosis Infection: Myths, Models, and Molecular Mechanisms. *Microbiol Mol Biol Rev*. 2014;78: 343–371. doi:10.1128/MMBR.00010-14
  54. Ford CB, Lin PL, Chase MR, Shah RR, Iartchouk O, Galagan J, et al. Use of whole genome sequencing to estimate the mutation rate of *Mycobacterium tuberculosis* during latent infection. *Nat Genet*. 2011;43: 482–486. doi:10.1038/ng.811
  55. Loomis HP. Some facts in the etiology of tuberculosis, evidenced by thirty autopsies and experiments upon animals. *M. Rec*. 1890;38:689–698
  56. Hernández-Pando R, Jeyanathan M, Mengistu G, Aguilar D, Orozco H, Harboe M, et al. Persistence of DNA from *Mycobacterium tuberculosis* in superficially normal lung tissue during latent infection. *Lancet*. 2000;356: 2133–2138. doi:10.1016/s0140-6736(00)03493-0
  57. Martin CJ, Cadena AM, Leung VW, Lin PL, Maiello P, Hicks N, et al. Digitally Barcoding *Mycobacterium tuberculosis* Reveals In Vivo Infection Dynamics in the Macaque Model of Tuberculosis. *mBio*. 2017;8. doi:10.1128/mBio.00312-17
  58. Wessler T, Joslyn LR, Borish HJ, Gideon HP, Flynn JL, Kirschner DE, et al. A computational model tracks whole-lung *Mycobacterium tuberculosis* infection and predicts factors that inhibit dissemination. *PLoS Comput Biol*. 2020;16. doi:10.1371/journal.pcbi.1007280
  59. Hanifa Y, Grant AD, Lewis J, Corbett EL, Fielding K, Churchyard G. Prevalence of latent tuberculosis infection among gold miners in South Africa. *Int J Tuberc Lung Dis*. 2009;13: 39–46.
  60. V. N. Houk MC U, J. H. Baker MC U, K. Sorensen MC U, D. C. Kent MC U. The Epidemiology of Tuberculosis Infection in a Closed Environment. *Archives of Environmental Health: An International Journal*. 1968;16: 26–35. doi:10.1080/00039896.1968.10665011
  61. Verrall AJ, Netea MG, Alisjahbana B, Hill PC, van Crevel R. Early clearance of *Mycobacterium tuberculosis*: a new frontier in prevention. *Immunology*. 2014;141: 506–513. doi:10.1111/imm.12223
  62. Kaipilyawar V, Salgame P. Infection resisters: targets of new research for uncovering natural protective immunity against *Mycobacterium tuberculosis*. *F1000Res*. 2019;8. doi:10.12688/f1000research.19805.1
  63. Meermeier EW, Lewinsohn DM. Early clearance versus control: what is the meaning of a negative tuberculin skin test or interferon-gamma release assay following exposure to *Mycobacterium tuberculosis*? *F1000Res*. 2018;7. doi:10.12688/f1000research.13224.1

64. Koeken V a. CM, Verrall AJ, Netea MG, Hill PC, van Crevel R. Trained innate immunity and resistance to *Mycobacterium tuberculosis* infection. *Clin Microbiol Infect.* 2019;25: 1468–1472. doi:10.1016/j.cmi.2019.02.015
65. Mave V, Chandrasekaran P, Chavan A, Shivakumar SVBY, Danasekaran K, Paradkar M, et al. Infection free “resisters” among household contacts of adult pulmonary tuberculosis. *PLoS ONE.* 2019;14: e0218034. doi:10.1371/journal.pone.0218034
66. Stein CM, Zalwango S, Malone LL, Thiel B, Mupere E, Nsereko M, et al. Resistance and Susceptibility to *Mycobacterium tuberculosis* Infection and Disease in Tuberculosis Households in Kampala, Uganda. *Am J Epidemiol.* 2018;187: 1477–1489. doi:10.1093/aje/kwx380
67. Morrison J, Pai M, Hopewell PC. Tuberculosis and latent tuberculosis infection in close contacts of people with pulmonary tuberculosis in low-income and middle-income countries: a systematic review and meta-analysis. *Lancet Infect Dis.* 2008;8: 359–368. doi:10.1016/S1473-3099(08)70071-9
68. Capuano SV, Croix DA, Pawar S, Zinovik A, Myers A, Lin PL, et al. Experimental *Mycobacterium tuberculosis* infection of cynomolgus macaques closely resembles the various manifestations of human *M. tuberculosis* infection. *Infect Immun.* 2003;71: 5831–5844. doi:10.1128/iai.71.10.5831-5844.2003
69. Zhan L, Tang J, Sun M, Qin C. Animal Models for Tuberculosis in Translational and Precision Medicine. *Front Microbiol.* 2017;8. doi:10.3389/fmicb.2017.00717
70. Sable SB, Posey JE, Scriba TJ. Tuberculosis Vaccine Development: Progress in Clinical Evaluation. *Clinical Microbiology Reviews.* 2019;33. doi:10.1128/CMR.00100-19
71. Mangtani P, Abubakar I, Ariti C, Beynon R, Pimpin L, Fine PEM, et al. Protection by BCG vaccine against tuberculosis: a systematic review of randomized controlled trials. *Clin Infect Dis.* 2014;58: 470–480. doi:10.1093/cid/cit790
72. Dockrell HM, Smith SG. What Have We Learnt about BCG Vaccination in the Last 20 Years? *Front Immunol.* 2017;8. doi:10.3389/fimmu.2017.01134
73. Ritz N, Dutta B, Donath S, Casalaz D, Connell TG, Tebruegge M, et al. The influence of bacille Calmette-Guerin vaccine strain on the immune response against tuberculosis: a randomized trial. *Am J Respir Crit Care Med.* 2012;185: 213–222. doi:10.1164/rccm.201104-0714OC
74. 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;349. doi:10.1136/bmj.g4643
75. Esmail H, Barry CE, Young DB, Wilkinson RJ. The ongoing challenge of latent tuberculosis. *Philos Trans R Soc Lond B Biol Sci.* 2014;369. doi:10.1098/rstb.2013.0437
76. Dye C. Breaking a law: tuberculosis disobeys Styblo’s rule. *Bull World Health Organ.* 2008;86: 4. doi:10.2471/BLT.07.049510

77. van Leth F, van der Werf MJ, Borgdorff MW. Prevalence of tuberculous infection and incidence of tuberculosis: a re-assessment of the Styblo rule. *Bull World Health Organ.* 2008;86: 20–26. doi:10.2471/blt.06.037804
78. Dye C, Scheele S, Dolin P, Pathania V, Raviglione MC. Consensus statement. Global burden of tuberculosis: estimated incidence, prevalence, and mortality by country. WHO Global Surveillance and Monitoring Project. *JAMA.* 1999;282: 677–686. doi:10.1001/jama.282.7.677
79. Shaw JB, Wynn-Williams N. Infectivity of pulmonary tuberculosis in relation to sputum status. *Am Rev Tuberc.* 1954;69: 724–732. doi:10.1164/art.1954.69.5.724
80. Espinal MA, Pérez EN, Baéz J, Hénriquez L, Fernández K, Lopez M, et al. Infectiousness of *Mycobacterium tuberculosis* in HIV-1-infected patients with tuberculosis: a prospective study. *Lancet.* 2000;355: 275–280. doi:10.1016/S0140-6736(99)04402-5
81. Hernández-Garduño E, Cook V, Kunimoto D, Elwood RK, Black WA, FitzGerald JM. Transmission of tuberculosis from smear negative patients: a molecular epidemiology study. *Thorax.* 2004;59: 286–290. doi:10.1136/thx.2003.011759
82. Behr MA, Warren SA, Salamon H, Hopewell PC, Ponce de Leon A, Daley CL, et al. Transmission of *Mycobacterium tuberculosis* from patients smear-negative for acid-fast bacilli. *Lancet.* 1999;353: 444–449. doi:10.1016/s0140-6736(98)03406-0
83. Tostmann A, Kik SV, Kalisvaart NA, Sebek MM, Verver S, Boeree MJ, et al. Tuberculosis transmission by patients with smear-negative pulmonary tuberculosis in a large cohort in the Netherlands. *Clin Infect Dis.* 2008;47: 1135–1142. doi:10.1086/591974
84. Opie EL, McPHEDRAN FM. SPREAD OF TUBERCULOSIS WITHIN FAMILIES. *JAMA.* 1926;87: 1549–1551. doi:10.1001/jama.1926.02680190025007
85. McPHEDRAN FM, Opie EL. THE SPREAD OF TUBERCULOSIS IN FAMILIES. *Am J Epidemiol.* 1935;22: 565–643. doi:10.1093/oxfordjournals.aje.a118191
86. 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.* 2012;16: 1594–1599. doi:10.5588/ijtld.12.0389
87. 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;168: 448–455. doi:10.1164/rccm.200212-1483OC
88. Kenyon TA, Valway SE, Ihle WW, Onorato IM, Castro KG. Transmission of multidrug-resistant *Mycobacterium tuberculosis* during a long airplane flight. *N Engl J Med.* 1996;334: 933–938. doi:10.1056/NEJM199604113341501
89. Blomquist ET. Tuberculosis casefinding, 1961. *Public Health Rep.* 1961;76: 871–876.

90. Brailey M. A STUDY OF TUBERCULOUS INFECTION AND MORTALITY IN THE CHILDREN OF TUBERCULOUS HOUSEHOLDS. *Am J Epidemiol.* 1940;31-SectionA: 1–43. doi:10.1093/oxfordjournals.aje.a118607
91. Andersen S, Geser A. The distribution of tuberculous infection among households in African communities. *Bull World Health Organ.* 1960;22: 39–60.
92. Bradford WZ, Koehler J, El-Hajj H, Hopewell PC, Reingold AL, Agasino CB, et al. Dissemination of *Mycobacterium tuberculosis* across the San Francisco Bay Area. *J Infect Dis.* 1998;177: 1104–1107. doi:10.1086/517405
93. Diel R, Schneider S, Meywald-Walter K, Ruf C-M, Rüsç-Gerdes S, Niemann S. Epidemiology of tuberculosis in Hamburg, Germany: long-term population-based analysis applying classical and molecular epidemiological techniques. *J Clin Microbiol.* 2002;40: 532–539. doi:10.1128/jcm.40.2.532-539.2002
94. Cacho Calvo J, Astray Mochales J, Pérez Meixeira A, Ramos Martos A, Hernando García M, Sánchez Concheiro M, et al. Ten-year population-based molecular epidemiological study of tuberculosis transmission in the metropolitan area of Madrid, Spain. *Int J Tuberc Lung Dis.* 2005;9: 1236–1241.
95. Martinez L, Shen Y, Mupere E, Kizza A, Hill PC, Whalen CC. Transmission of *Mycobacterium Tuberculosis* in Households and the Community: A Systematic Review and Meta-Analysis. *Am J Epidemiol.* 2017;185: 1327–1339. doi:10.1093/aje/kwx025
96. Morrison J, Pai M, Hopewell PC. Tuberculosis and latent tuberculosis infection in close contacts of people with pulmonary tuberculosis in low-income and middle-income countries: a systematic review and meta-analysis. *The Lancet Infectious Diseases.* 2008;8: 359–368. doi:10.1016/S1473-3099(08)70071-9
97. Fox GJ, Barry SE, Britton WJ, Marks GB. Contact investigation for tuberculosis: a systematic review and meta-analysis. *Eur Respir J.* 2013;41: 140–156. doi:10.1183/09031936.00070812
98. 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 USA.* 2006;103: 2869–2873. doi:10.1073/pnas.0511240103
99. Zelner JL, Murray MB, Becerra MC, Galea J, Lecca L, Calderon R, et al. Age-Specific Risks of Tuberculosis Infection From Household and Community Exposures and Opportunities for Interventions in a High-Burden Setting. *Am J Epidemiol.* 2014;180: 853–861. doi:10.1093/aje/kwu192
100. Wood R, Liang H, Wu H, Middelkoop K, Oni T, Rangaka MX, et al. Changing prevalence of TB infection with increasing age in high TB burden townships in South Africa. *Int J Tuberc Lung Dis.* 2010;14: 406–412.
101. Marquez C, Atukunda M, Balzer LB, Chamie G, Kironde J, Ssemmondo E, et al. The age-specific burden and household and school-based predictors of child and adolescent tuberculosis infection in rural Uganda. *PLOS ONE.* 2020;15: e0228102. doi:10.1371/journal.pone.0228102

102. Byng-Maddick R, Noursadeghi M. Does tuberculosis threaten our ageing populations? *BMC Infect Dis.* 2016;16. doi:10.1186/s12879-016-1451-0
103. Horton KC, MacPherson P, Houben RMGJ, White RG, Corbett EL. Sex Differences in Tuberculosis Burden and Notifications in Low- and Middle-Income Countries: A Systematic Review and Meta-analysis. *PLoS Med.* 2016;13. doi:10.1371/journal.pmed.1002119
104. Neyrolles O, Quintana-Murci L. Sexual Inequality in Tuberculosis. *PLOS Medicine.* 2009;6: e1000199. doi:10.1371/journal.pmed.1000199
105. Walker NF, Meintjes G, Wilkinson RJ. HIV-1 and the immune response to TB. *Future Virol.* 2013;8: 57–80. doi:10.2217/fvl.12.123
106. Schaible UE, Kaufmann SHE. Malnutrition and infection: complex mechanisms and global impacts. *PLoS Med.* 2007;4: e115. doi:10.1371/journal.pmed.0040115
107. 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 Metaanalysis. *Clin Infect Dis.* 2017;64: 719–727. doi:10.1093/cid/ciw836
108. Slama K, Chiang C-Y, Enarson DA, Hassmiller K, Fanning A, Gupta P, et al. Tobacco and tuberculosis: a qualitative systematic review and meta-analysis. *Int J Tuberc Lung Dis.* 2007;11: 1049–1061.
109. Imtiaz S, Shield KD, Roerecke M, Samokhvalov AV, Lönnroth K, Rehm J. Alcohol consumption as a risk factor for tuberculosis: meta-analyses and burden of disease. *Eur Respir J.* 2017;50. doi:10.1183/13993003.00216-2017
110. Ganmaa D, Uyanga B, Zhou X, Gantsetseg G, Delgerekh B, Enkhmaa D, et al. Vitamin D Supplements for Prevention of Tuberculosis Infection and Disease. *N Engl J Med.* 2020;383: 359–368. doi:10.1056/NEJMoa1915176
111. Herzog H. History of tuberculosis. *Respiration.* 1998;65: 5–15. doi:10.1159/000029220
112. LEJARD C. NOTIONS ACTUELLES SUR LA CONTAGION DE LA TUBERCULOSE. *Revue des Deux Mondes (1829-1971).* 1950; 308–317.
113. Daniel TM. Jean-Antoine Villemin and the infectious nature of tuberculosis. *Int J Tuberc Lung Dis.* 2015;19: 267–268. doi:10.5588/ijtld.06.0636
114. Kallmann FJ, Reisner D. Twin Studies on the Significance of Genetic Factors in Tuberculosis. *am rev tuberc.* 1943;47: 549–574. doi:10.1164/art.1943.47.6.549
115. McHenry ML, Williams SM, Stein CM. Genetics and evolution of tuberculosis pathogenesis: New perspectives and approaches. *Infection, Genetics and Evolution.* 2020;81: 104204. doi:10.1016/j.meegid.2020.104204
116. Stein CM, Sausville L, Wejse C, Sobota RS, Zetola NM, Hill PC, et al. Genomics of human pulmonary tuberculosis: from genes to pathways. *Curr Genet Med Rep.* 2017;5: 149–166. doi:10.1007/s40142-017-0130-9

117. Abel L, Fellay J, Haas DW, Schurr E, Srikrishna G, Urbanowski M, et al. Genetics of human susceptibility to active and latent tuberculosis: present knowledge and future perspectives. *Lancet Infect Dis.* 2018;18: e64–e75. doi:10.1016/S1473-3099(17)30623-0
118. 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- $\gamma$  immunity. *Semin Immunol.* 2014;26: 454–470. doi:10.1016/j.smim.2014.09.008
119. Bustamante J. Mendelian susceptibility to mycobacterial disease: recent discoveries. *Hum Genet.* 2020;139: 993–1000. doi:10.1007/s00439-020-02120-y
120. 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;100: 2658–2664. doi:10.1172/JCI119810
121. Boisson-Dupuis S, El Baghdadi J, Parvaneh N, Bousfiha A, Bustamante J, Feinberg J, et al. IL-12R $\beta$ 1 Deficiency in Two of Fifty Children with Severe Tuberculosis from Iran, Morocco, and Turkey. *PLoS One.* 2011;6. doi:10.1371/journal.pone.0018524
122. Boisson-Dupuis S, Ramirez-Alejo N, Li Z, Patin E, Rao G, Kerner G, et al. Tuberculosis and impaired IL-23-dependent IFN- $\gamma$  immunity in humans homozygous for a common TYK2 missense variant. *Sci Immunol.* 2018;3. doi:10.1126/sciimmunol.aau8714
123. Kerner G, Ramirez-Alejo N, Seeleuthner Y, Yang R, Ogishi M, Cobat A, et al. Homozygosity for TYK2 P1104A underlies tuberculosis in about 1% of patients in a cohort of European ancestry. *Proc Natl Acad Sci U S A.* 2019;116: 10430–10434. doi:10.1073/pnas.1903561116
124. 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: 620–624. doi:10.1164/ajrccm.149.3.8118628
125. Visscher PM, Hill WG, Wray NR. Heritability in the genomics era--concepts and misconceptions. *Nat Rev Genet.* 2008;9: 255–266. doi:10.1038/nrg2322
126. 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;69: 7964.
127. 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;201: 15–19. doi:10.1086/648611
128. Tao L, Zalwango S, Chervenak K, Thiel B, Malone LL, Qiu F, et al. Genetic and Shared Environmental Influences on Interferon- $\gamma$  Production in Response to Mycobacterium tuberculosis Antigens in a Ugandan Population. *Am J Trop Med Hyg.* 2013;89: 169–173. doi:10.4269/ajtmh.12-0670

129. Stein CM, Zalwango S, Malone LL, Won S, Mayanja-Kizza H, Mugerwa RD, et al. Genome scan of *M. tuberculosis* infection and disease in Ugandans. *PLoS ONE*. 2008;3: e4094. doi:10.1371/journal.pone.0004094
130. Igo RP, Hall NB, Malone LL, Hall JB, Truitt B, Qiu F, et al. Fine-mapping Analysis of a Chromosome 2 Region Linked to Resistance to *Mycobacterium tuberculosis* Infection in Uganda Reveals Potential Regulatory Variants. *Genes Immun*. 2019;20: 473–483. doi:10.1038/s41435-018-0040-1
131. Thye T, Browne EN, Chinbuah MA, Gyapong J, Osei I, Owusu-Dabo E, et al. IL10 haplotype associated with tuberculin skin test response but not with pulmonary TB. *PLoS ONE*. 2009;4: e5420. doi:10.1371/journal.pone.0005420
132. Zembruski VM, Basta PC, Callegari-Jacques SM, Santos RV, Coimbra CEA, Salzano FM, et al. Cytokine genes are associated with tuberculin skin test response in a native Brazilian population. *Tuberculosis (Edinb)*. 2010;90: 44–49. doi:10.1016/j.tube.2009.11.002
133. Horne DJ, Graustein AD, Shah JA, Peterson G, Savlov M, Steele S, et al. Human ULK1 Variation and Susceptibility to *Mycobacterium tuberculosis* Infection. *J Infect Dis*. 2016;214: 1260–1267. doi:10.1093/infdis/jiw347
134. Tian C, Hromatka BS, Kiefer AK, Eriksson N, Noble SM, Tung JY, et al. Genome-wide association and HLA region fine-mapping studies identify susceptibility loci for multiple common infections. *Nat Commun*. 2017;8. doi:10.1038/s41467-017-00257-5
135. 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;48: 318–322. doi:10.1038/ng.3498
136. Sobota RS, Stein CM, Kodaman N, Maro I, Wieland-Alter W, Igo RP, et al. A chromosome 5q31.1 locus associates with tuberculin skin test reactivity in HIV-positive individuals from tuberculosis hyper-endemic regions in east Africa. *PLoS Genet*. 2017;13: e1006710. doi:10.1371/journal.pgen.1006710
137. 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;206: 2583–2591. doi:10.1084/jem.20090892
138. Cobat A, Poirier C, Hoal E, Boland-Auge A, de La Rocque F, Corrad 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;211: 317–321. doi:10.1093/infdis/jiu446
139. 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;57: 963–970. doi:10.1093/cid/cit438
140. Jabot-Hanin F, Cobat A, Feinberg J, Grange G, Remus N, Poirier C, et al. Major Loci on Chromosomes 8q and 3q Control Interferon  $\gamma$  Production Triggered by *Bacillus*

- Calmette-Guerin and 6-kDa Early Secretory Antigen Target, Respectively, in Various Populations. *J Infect Dis.* 2016;213: 1173–1179. doi:10.1093/infdis/jiv757
141. Menzies D, Joshi R, Pai M. Risk of tuberculosis infection and disease associated with work in health care settings. *Int J Tuberc Lung Dis.* 2007;11: 593–605.
142. Joshi R, Reingold AL, Menzies D, Pai M. Tuberculosis among health-care workers in low- and middle-income countries: a systematic review. *PLoS Med.* 2006;3: e494. doi:10.1371/journal.pmed.0030494
143. Nasreen S, Shokoohi M, Malvankar-Mehta MS. Prevalence of Latent Tuberculosis among Health Care Workers in High Burden Countries: A Systematic Review and Meta-Analysis. *PLoS One.* 2016;11. doi:10.1371/journal.pone.0164034
144. Uden L, Barber E, Ford N, Cooke GS. Risk of Tuberculosis Infection and Disease for Health Care Workers: An Updated Meta-Analysis. *Open Forum Infect Dis.* 2017;4: ofx137. doi:10.1093/ofid/ofx137
145. Getahun H, Matteelli A, Abubakar I, Hauer B, Pontali E, Migliori GB. Advancing global programmatic management of latent tuberculosis infection for at risk populations. *European Respiratory Journal.* 2016;47: 1327–1330. doi:10.1183/13993003.00449-2016
146. WHO | UNICEF coverage estimates: Immunization, Vaccines And Biologicals. Vaccine preventable diseases Vaccines monitoring system 2018 Global Summary Reference Time Series: BCG. Available: [http://apps.who.int/immunization\\_monitoring/globalsummary/timeseries/tswucoverage/bcg.html](http://apps.who.int/immunization_monitoring/globalsummary/timeseries/tswucoverage/bcg.html)
147. 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;180: 49–58. doi:10.1164/rccm.200811-1704OC
148. Nemes E, Rozot V, Geldenhuys H, Bilek N, Mabwe S, Abrahams D, et al. Optimization and Interpretation of Serial QuantiFERON Testing to Measure Acquisition of Mycobacterium tuberculosis Infection. *Am J Respir Crit Care Med.* 2017;196: 638–648. doi:10.1164/rccm.201704-0817OC
149. Zwerling A, Hof S van den, Scholten J, Cobelens F, Menzies D, Pai M. Interferon-gamma release assays for tuberculosis screening of healthcare workers: a systematic review. *Thorax.* 2012;67: 62–70. doi:10.1136/thx.2010.143180
150. Lamberti M, Uccello R, Monaco MGL, Muoio M, Feola D, Sannolo N, et al. Tuberculin skin test and Quantiferon test agreement and influencing factors in tuberculosis screening of healthcare workers: a systematic review and meta-analysis. *J Occup Med Toxicol.* 2015;10. doi:10.1186/s12995-015-0044-y
151. Souza FM de, Prado TN do, Pinheiro J dos S, Peres RL, Lacerda TC, Loureiro RB, et al. Comparison of Interferon- $\gamma$  Release Assay to Two Cut-Off Points of Tuberculin Skin Test to Detect Latent Mycobacterium tuberculosis Infection in Primary Health Care Workers. *PLOS ONE.* 2014;9: e102773. doi:10.1371/journal.pone.0102773

152. He G, Li Y, Zhao F, Wang L, Cheng S, Guo H, et al. The Prevalence and Incidence of Latent Tuberculosis Infection and Its Associated Factors among Village Doctors in China. *PLOS ONE*. 2015;10: e0124097. doi:10.1371/journal.pone.0124097
153. He GX, Wang LX, Chai SJ, Klena JD, Cheng SM, Ren YL, et al. Risk factors associated with tuberculosis infection among health care workers in Inner Mongolia, China. *Int J Tuberc Lung Dis*. 2012;16: 1485–1491. doi:10.5588/ijtld.12.0193
154. Ting W-Y, Huang S-F, Lee M-C, Lin Y-Y, Lee Y-C, Feng J-Y, et al. Gender disparities in latent tuberculosis infection in high-risk individuals: a cross-sectional study. *PLoS ONE*. 2014;9: e110104. doi:10.1371/journal.pone.0110104
155. Harada N, Nakajima Y, Higuchi K, Sekiya Y, Rothel J, Mori T. Screening for tuberculosis infection using whole-blood interferon-gamma and Mantoux testing among Japanese healthcare workers. *Infect Control Hosp Epidemiol*. 2006;27: 442–448. doi:10.1086/504358
156. Jo K-W, Hong Y, Park JS, Bae I-G, Eom JS, Lee S-R, et al. Prevalence of Latent Tuberculosis Infection among Health Care Workers in South Korea: A Multicenter Study. *Tuberc Respir Dis (Seoul)*. 2013;75: 18–24. doi:10.4046/trd.2013.75.1.18
157. Rutanga C, Lowrance DW, Oeltmann JE, Mutembayire G, Willis M, Uwizeye CB, et al. Latent Tuberculosis Infection and Associated Factors among Health Care Workers in Kigali, Rwanda. *PLoS One*. 2015;10. doi:10.1371/journal.pone.0124485
158. Lien LT, Hang NTL, Kobayashi N, Yanai H, Toyota E, Sakurada S, et al. Prevalence and Risk Factors for Tuberculosis Infection among Hospital Workers in Hanoi, Viet Nam. *PLOS ONE*. 2009;4: e6798. doi:10.1371/journal.pone.0006798
159. Zhang H, Xin H, Li X, Li H, Li M, Lu W, et al. A dose-response relationship of smoking with tuberculosis infection: A cross-sectional study among 21008 rural residents in China. *PLoS One*. 2017;12. doi:10.1371/journal.pone.0175183
160. Aabye MG, Hermansen TS, Ruhwald M, PrayGod G, Faurholt-Jepsen D, Jeremiah K, et al. Negative effect of smoking on the performance of the QuantiFERON TB gold in tube test. *BMC Infect Dis*. 2012;12: 379. doi:10.1186/1471-2334-12-379
161. Pai M, Kalantri S, Menzies D. Discordance between tuberculin skin test and interferon-gamma assays. *Int J Tuberc Lung Dis*. 2006;10: 942–943.
162. Slovis BS, Plitman JD, Haas DW. The case against anergy testing as a routine adjunct to tuberculin skin testing. *JAMA*. 2000;283: 2003–2007.
163. WHO | Global tuberculosis report 2015. In: WHO [Internet]. Available: <http://apps.who.int/iris/handle/10665/191102>
164. Antoine D. Les cas de tuberculose maladie déclarés en France en 2006. *Bull Epidemiol Hebd*. 2008;10-11.
165. de Wit E, Delpont 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;128: 145–153. doi:10.1007/s00439-010-0836-1

166. Naidoo P, Theron G, Rangaka MX, Chihota VN, Vaughan L, Brey ZO, et al. The South African Tuberculosis Care Cascade: Estimated Losses and Methodological Challenges. *J Infect Dis.* 2017;216: S702–S713. doi:10.1093/infdis/jix335
167. Mathema B, Andrews JR, Cohen T, Borgdorff MW, Behr M, Glynn JR, et al. Drivers of Tuberculosis Transmission. *J Infect Dis.* 2017;216: S644–S653. doi:10.1093/infdis/jix354
168. Microarray Kits. Available: <https://www.illumina.com/products/by-type/microarray-kits.html>
169. Das S, Abecasis GR, Browning BL. Genotype Imputation from Large Reference Panels. *Annu Rev Genomics Hum Genet.* 2018;19: 73–96. doi:10.1146/annurev-genom-083117-021602
170. 1000 Genomes Project Consortium, Auton A, Brooks LD, Durbin RM, Garrison EP, Kang HM, et al. A global reference for human genetic variation. *Nature.* 2015;526: 68–74. doi:10.1038/nature15393
171. Das S, Forer L, Schönherr S, Sidore C, Locke AE, Kwong A, et al. Next-generation genotype imputation service and methods. *Nat Genet.* 2016;48: 1284–1287. doi:10.1038/ng.3656
172. McCarthy S, Das S, Kretzschmar W, Delaneau O, Wood AR, Teumer A, et al. A reference panel of 64,976 haplotypes for genotype imputation. *Nat Genet.* 2016;48: 1279–1283. doi:10.1038/ng.3643
173. Voight BF, Pritchard JK. Confounding from Cryptic Relatedness in Case-Control Association Studies. *PLoS Genet.* 2005;1. doi:10.1371/journal.pgen.0010032
174. Price AL, Zaitlen NA, Reich D, Patterson N. New approaches to population stratification in genome-wide association studies. *Nat Rev Genet.* 2010;11: 459–463. doi:10.1038/nrg2813
175. Vincent Q. *Epidémiologie et génétique humaine de l’ulcère de Buruli.* These de doctorat, Paris 5. 2014. Available: <http://www.theses.fr/2014PA05S019>
176. FALUSH D, STEPHENS M, PRITCHARD JK. Inference of population structure using multilocus genotype data: dominant markers and null alleles. *Mol Ecol Notes.* 2007;7: 574–578. doi:10.1111/j.1471-8286.2007.01758.x
177. Alexander DH, Novembre J, Lange K. Fast model-based estimation of ancestry in unrelated individuals. *Genome Res.* 2009;19: 1655–1664. doi:10.1101/gr.094052.109
178. Lawson DJ, van Dorp L, Falush D. A tutorial on how not to over-interpret STRUCTURE and ADMIXTURE bar plots. *Nat Commun.* 2018;9: 3258. doi:10.1038/s41467-018-05257-7
179. Sierra B, Triska P, Soares P, Garcia G, Perez AB, Aguirre E, et al. OSBPL10, RXRA and lipid metabolism confer African-ancestry protection against dengue haemorrhagic fever in admixed Cubans. *PLoS Pathog.* 2017;13: e1006220. doi:10.1371/journal.ppat.1006220

180. Asgari S, Luo Y, Slowikowski K, Huang C-C, Calderon R, Jimenez J, et al. Higher Native Peruvian ancestry proportion is associated with tuberculosis progression risk. medRxiv. 2020; 2020.09.09.20191437. doi:10.1101/2020.09.09.20191437
181. 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;161: 1376–1395. doi:10.1164/ajrccm.161.4.16141
182. Voight BF, Pritchard JK. Confounding from cryptic relatedness in case-control association studies. *PLoS Genet.* 2005;1: e32. doi:10.1371/journal.pgen.0010032
183. 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;10: e1004445. doi:10.1371/journal.pgen.1004445
184. Penn DJ, Damjanovich K, Potts WK. MHC heterozygosity confers a selective advantage against multiple-strain infections. *Proc Natl Acad Sci U S A.* 2002;99: 11260–11264. doi:10.1073/pnas.162006499
185. Listgarten J, Lippert C, Kadie CM, Davidson RI, Eskin E, Heckerman D. Improved linear mixed models for genome-wide association studies. *Nat Methods.* 2012;9: 525–526. doi:10.1038/nmeth.2037
186. Zhou X, Carbonetto P, Stephens M. Polygenic modeling with bayesian sparse linear mixed models. *PLoS Genet.* 2013;9: e1003264. doi:10.1371/journal.pgen.1003264
187. Cook JP, Mahajan A, Morris AP. Guidance for the utility of linear models in meta-analysis of genetic association studies of binary phenotypes. *Eur J Hum Genet.* 2017;25: 240–245. doi:10.1038/ejhg.2016.150
188. Lloyd-Jones LR, Robinson MR, Yang J, Visscher PM. Transformation of Summary Statistics from Linear Mixed Model Association on All-or-None Traits to Odds Ratio. *Genetics.* 2018;208: 1397–1408. doi:10.1534/genetics.117.300360
189. Dandine-Roulland C. Modélisation de la composante génétique des maladies humaines : Données familiales et Modèles Mixtes. These de doctorat, Université Paris-Saclay (ComUE). 2016. Available: <http://www.theses.fr/2016SACLS259>
190. Visscher PM, Hill WG, Wray NR. Heritability in the genomics era--concepts and misconceptions. *Nat Rev Genet.* 2008;9: 255–266. doi:10.1038/nrg2322
191. Tenesa A, Haley CS. The heritability of human disease: estimation, uses and abuses. *Nat Rev Genet.* 2013;14: 139–149. doi:10.1038/nrg3377
192. Manolio TA, Collins FS, Cox NJ, Goldstein DB, Hindorff LA, Hunter DJ, et al. Finding the missing heritability of complex diseases. *Nature.* 2009;461: 747–753. doi:10.1038/nature08494

193. Golan D, Lander ES, Rosset S. Measuring missing heritability: inferring the contribution of common variants. *Proc Natl Acad Sci U S A*. 2014;111: E5272-5281. doi:10.1073/pnas.1419064111
194. Pruim RJ, Welch RP, Sanna S, Teslovich TM, Chines PS, Gliedt TP, et al. LocusZoom: regional visualization of genome-wide association scan results. *Bioinformatics*. 2010;26: 2336–2337. doi:10.1093/bioinformatics/btq419
195. Barrett JC, Fry B, Maller J, Daly MJ. Haploview: analysis and visualization of LD and haplotype maps. *Bioinformatics*. 2005;21: 263–265. doi:10.1093/bioinformatics/bth457
196. Hunter DJ, Kraft P. Drinking from the fire hose--statistical issues in genomewide association studies. *N Engl J Med*. 2007;357: 436–439. doi:10.1056/NEJMp078120
197. Johnson KE, Voight BF. Patterns of shared signatures of recent positive selection across human populations. *Nat Ecol Evol*. 2018;2: 713–720. doi:10.1038/s41559-018-0478-6
198. Holsinger KE, Weir BS. Genetics in geographically structured populations: defining, estimating and interpreting  $F_{ST}$ . *Nat Rev Genet*. 2009;10: 639–650. doi:10.1038/nrg2611
199. Sabeti PC, Varilly P, Fry B, Lohmueller J, Hostetter E, Cotsepas C, et al. Genome-wide detection and characterization of positive selection in human populations. *Nature*. 2007;449: 913–918. doi:10.1038/nature06250
200. Voight BF, Kudaravalli S, Wen X, Pritchard JK. A Map of Recent Positive Selection in the Human Genome. *PLoS Biol*. 2006;4. doi:10.1371/journal.pbio.0040072
201. 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;98: 514–524. doi:10.1016/j.ajhg.2016.01.015
202. Evangelou E, Warren HR, Mosen-Ansorena D, Mifsud B, Pazoki R, Gao H, et al. Genetic analysis of over 1 million people identifies 535 new loci associated with blood pressure traits. *Nat Genet*. 2018;50: 1412–1425. doi:10.1038/s41588-018-0205-x
203. Liu M, Jiang Y, Wedow R, Li Y, Brazel DM, Chen F, et al. Association studies of up to 1.2 million individuals yield new insights into the genetic etiology of tobacco and alcohol use. *Nat Genet*. 2019;51: 237–244. doi:10.1038/s41588-018-0307-5
204. Solovieff N, Cotsepas C, Lee PH, Purcell SM, Smoller JW. Pleiotropy in complex traits: challenges and strategies. *Nat Rev Genet*. 2013;14: 483–495. doi:10.1038/nrg3461
205. Zeggini E, Ioannidis JPA. Meta-analysis in genome-wide association studies. *Pharmacogenomics*. 2009;10: 191–201. doi:10.2217/14622416.10.2.191
206. Morris AP. Transethnic meta-analysis of genomewide association studies. *Genet Epidemiol*. 2011;35: 809–822. doi:10.1002/gepi.20630
207. Han B, Eskin E. Random-Effects Model Aimed at Discovering Associations in Meta-Analysis of Genome-wide Association Studies. *Am J Hum Genet*. 2011;88: 586–598. doi:10.1016/j.ajhg.2011.04.014

208. Willer CJ, Li Y, Abecasis GR. METAL: fast and efficient meta-analysis of genomewide association scans. *Bioinformatics*. 2010;26: 2190–2191. doi:10.1093/bioinformatics/btq340
209. Gallagher MD, Chen-Plotkin AS. The Post-GWAS Era: From Association to Function. *Am J Hum Genet*. 2018;102: 717–730. doi:10.1016/j.ajhg.2018.04.002
210. Cano-Gamez E, Trynka G. From GWAS to Function: Using Functional Genomics to Identify the Mechanisms Underlying Complex Diseases. *Front Genet*. 2020;11. doi:10.3389/fgene.2020.00424
211. Schaid DJ, Chen W, Larson NB. From genome-wide associations to candidate causal variants by statistical fine-mapping. *Nat Rev Genet*. 2018;19: 491–504. doi:10.1038/s41576-018-0016-z
212. Maurano MT, Humbert R, Rynes E, Thurman RE, Haugen E, Wang H, et al. Systematic Localization of Common Disease-Associated Variation in Regulatory DNA. *Science*. 2012;337: 1190–1195. doi:10.1126/science.1222794
213. Westra H-J, Franke L. From genome to function by studying eQTLs. *Biochim Biophys Acta*. 2014;1842: 1896–1902. doi:10.1016/j.bbadis.2014.04.024
214. Haeussler M, Zweig AS, Tyner C, Speir ML, Rosenbloom KR, Raney BJ, et al. The UCSC Genome Browser database: 2019 update. *Nucleic Acids Res*. 2019;47: D853–D858. doi:10.1093/nar/gky1095
215. ENCODE Project Consortium. An integrated encyclopedia of DNA elements in the human genome. *Nature*. 2012;489: 57–74. doi:10.1038/nature11247
216. Ernst J, Kellis M. Chromatin-state discovery and genome annotation with ChromHMM. *Nat Protoc*. 2017;12: 2478–2492. doi:10.1038/nprot.2017.124
217. Kundaje A, Meuleman W, Ernst J, Bilenky M, Yen A, Heravi-Moussavi A, et al. Integrative analysis of 111 reference human epigenomes. *Nature*. 2015;518: 317–330. doi:10.1038/nature14248
218. Calderon D, Nguyen MLT, Mezger A, Kathiria A, Müller F, Nguyen V, et al. Landscape of stimulation-responsive chromatin across diverse human immune cells. *Nat Genet*. 2019;51: 1494–1505. doi:10.1038/s41588-019-0505-9
219. Kim-Hellmuth S, Bechheim M, Pütz B, Mohammadi P, Nédélec Y, Giangreco N, et al. Genetic regulatory effects modified by immune activation contribute to autoimmune disease associations. *Nat Commun*. 2017;8: 1–10. doi:10.1038/s41467-017-00366-1
220. Nédélec Y, Sanz J, Baharian G, Szpiech ZA, Pacis A, Dumaine A, et al. Genetic Ancestry and Natural Selection Drive Population Differences in Immune Responses to Pathogens. *Cell*. 2016;167: 657–669.e21. doi:10.1016/j.cell.2016.09.025
221. Kasela S, Kisand K, Tserel L, Kaleviste E, Remm A, Fischer K, et al. Pathogenic implications for autoimmune mechanisms derived by comparative eQTL analysis of CD4<sup>+</sup> versus CD8<sup>+</sup> T cells. *PLOS Genetics*. 2017;13: e1006643. doi:10.1371/journal.pgen.1006643

222. Alasoo K, Rodrigues J, Mukhopadhyay S, Knights AJ, Mann AL, Kundu K, et al. Shared genetic effects on chromatin and gene expression indicate a role for enhancer priming in immune response. *Nat Genet.* 2018;50: 424–431. doi:10.1038/s41588-018-0046-7
223. Schmiedel BJ, Singh D, Madrigal A, Valdovino-Gonzalez AG, White BM, Zapardiel-Gonzalo J, et al. Impact of Genetic Polymorphisms on Human Immune Cell Gene Expression. *Cell.* 2018;175: 1701-1715.e16. doi:10.1016/j.cell.2018.10.022
224. Ma K, Qiu L, Mrasek K, Zhang J, Liehr T, Quintana LG, et al. Common Fragile Sites: Genomic Hotspots of DNA Damage and Carcinogenesis. *Int J Mol Sci.* 2012;13: 11974–11999. doi:10.3390/ijms130911974
225. Zack TI, Schumacher SE, Carter SL, Cherniack AD, Saksena G, Tabak B, et al. Pan-cancer patterns of somatic copy-number alteration. *Nat Genet.* 2013;45: 1134–1140. doi:10.1038/ng.2760
226. Yan S, Qiu L, Ma K, Zhang X, Zhao Y, Zhang J, et al. FATS is an E2-independent ubiquitin ligase that stabilizes p53 and promotes its activation in response to DNA damage. *Oncogene.* 2014;33: 5424–5433. doi:10.1038/onc.2013.494
227. Song F, Zhang J, Qiu L, Zhao Y, Xing P, Lu J, et al. A functional genetic variant in fragile-site gene FATS modulates the risk of breast cancer in triparous women. *BMC Cancer.* 2015;15: 559. doi:10.1186/s12885-015-1570-9
228. Manzanillo PS, Ayres JS, Watson RO, Collins AC, Souza G, Rae CS, et al. The ubiquitin ligase parkin mediates resistance to intracellular pathogens. *Nature.* 2013;501: 512–516. doi:10.1038/nature12566
229. Franco LH, Nair VR, Scharn CR, Xavier RJ, Torrealba JR, Shiloh MU, et al. The Ubiquitin Ligase Smurf1 Functions in Selective Autophagy of Mycobacterium tuberculosis and Anti-tuberculous Host Defense. *Cell Host Microbe.* 2017;21: 59–72. doi:10.1016/j.chom.2016.11.002
230. Wang L, Wu J, Li J, Yang H, Tang T, Liang H, et al. Host-mediated ubiquitination of a mycobacterial protein suppresses immunity. *Nature.* 2020;577: 682–688. doi:10.1038/s41586-019-1915-7
231. Kasthuber ER, Lowe SW. Putting p53 in Context. *Cell.* 2017;170: 1062–1078. doi:10.1016/j.cell.2017.08.028
232. Behar SM, Martin CJ, Booty MG, Nishimura T, Zhao X, Gan H-X, et al. Apoptosis is an innate defense function of macrophages against Mycobacterium tuberculosis. *Mucosal Immunol.* 2011;4: 279–287. doi:10.1038/mi.2011.3
233. Lim Y-J, Lee J, Choi J-A, Cho S-N, Son S-H, Kwon S-J, et al. M1 macrophage dependent-p53 regulates the intracellular survival of mycobacteria. *Apoptosis.* 2020;25: 42–55. doi:10.1007/s10495-019-01578-0
234. Cruz A, Ludovico P, Torrado E, Gama JB, Sousa J, Gaifem J, et al. IL-17A Promotes Intracellular Growth of Mycobacterium by Inhibiting Apoptosis of Infected Macrophages. *Front Immunol.* 2015;6. doi:10.3389/fimmu.2015.00498

235. spicuglia salvatore, Vanhille L. Chromatin signatures of active enhancers. *Nucleus*. 2012;3: 126–131. doi:10.4161/nucl.19232
236. Heintzman ND, Stuart RK, Hon G, Fu Y, Ching CW, Hawkins RD, et al. Distinct and predictive chromatin signatures of transcriptional promoters and enhancers in the human genome. *Nat Genet*. 2007;39: 311–318. doi:10.1038/ng1966
237. Kveiborg M, Albrechtsen R, Couchman JR, Wewer UM. Cellular roles of ADAM12 in health and disease. *Int J Biochem Cell Biol*. 2008;40: 1685–1702. doi:10.1016/j.biocel.2008.01.025
238. Zhou AX, El Hed A, Mercer F, Kozhaya L, Unutmaz D. The Metalloprotease ADAM12 Regulates the Effector Function of Human Th17 Cells. *PLoS One*. 2013;8. doi:10.1371/journal.pone.0081146
239. Arlehamn CL, Seumois G, Gerasimova A, Huang C, Fu Z, Yue X, et al. Transcriptional profile of TB antigen-specific T cells reveals novel multifunctional features. *J Immunol*. 2014;193: 2931–2940. doi:10.4049/jimmunol.1401151
240. Babu S, Bhat SQ, Kumar NP, Kumaraswami V, Nutman TB. Regulatory T Cells Modulate Th17 Responses in Patients with Positive Tuberculin Skin Test Results. *J Infect Dis*. 2010;201: 20–31. doi:10.1086/648735
241. Coulter F, Parrish A, Manning D, Kampmann B, Mendy J, Garand M, et al. IL-17 Production from T Helper 17, Mucosal-Associated Invariant T, and  $\gamma\delta$  Cells in Tuberculosis Infection and Disease. *Front Immunol*. 2017;8. doi:10.3389/fimmu.2017.01252
242. Shen H, Chen ZW. The crucial roles of Th17-related cytokines/signal pathways in M. tuberculosis infection. *Cell Mol Immunol*. 2018;15: 216–225. doi:10.1038/cmi.2017.128
243. Das S, Khader S. Yin and yang of interleukin-17 in host immunity to infection. *F1000Res*. 2017;6. doi:10.12688/f1000research.10862.1
244. Pollara G, Turner CT, Tomlinson GS, Bell LC, Khan A, Peralta LF, et al. Exaggerated in vivo IL-17 responses discriminate recall responses in active TB. *bioRxiv*. 2019; 516690. doi:10.1101/516690
245. Werren EA, Garcia O, Bigham AW. Identifying adaptive alleles in the human genome: from selection mapping to functional validation. *Hum Genet*. 2020. doi:10.1007/s00439-020-02206-7
246. 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. doi:10.1371/journal.pone.0169539
247. Friedland IR. The booster effect with repeat tuberculin testing in children and its relationship to BCG vaccination. *S Afr Med J*. 1990;77: 387–389.
248. Moreno S, Blázquez R, Novoa A, Carpena I, Menasalvas A, Ramírez C, et al. The effect of BCG vaccination on tuberculin reactivity and the booster effect among hospital employees. *Arch Intern Med*. 2001;161: 1760–1765. doi:10.1001/archinte.161.14.1760

249. Stein CM, Hall NB, Malone LL, Mupere E. The household contact study design for genetic epidemiological studies of infectious diseases. *Front Genet.* 2013;4: 61. doi:10.3389/fgene.2013.00061
250. Prevalence of HIV, total (% of population ages 15-49) - Vietnam | Data. Available: <https://data.worldbank.org/indicator/SH.DYN.AIDS.ZS?locations=VN>
251. Prevalence of HIV, total (% of population ages 15-49) - France | Data. Available: <https://data.worldbank.org/indicator/SH.DYN.AIDS.ZS?locations=FR>
252. 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;177: 1041–1047. doi:10.1164/rccm.200711-1756OC
253. Duggal P, Thio CL, Wojcik GL, Goedert JJ, Mangia A, Latanich R, et al. Genome wide association study of spontaneous resolution of hepatitis C virus infection. *Ann Intern Med.* 2013;158: 235–245. doi:10.7326/0003-4819-158-4-201302190-00003
254. Wojcik GL, Korpe P, Marie C, Mentzer AJ, Carstensen T, Mychaleckyj J, et al. Genome-Wide Association Study of Cryptosporidiosis in Infants Implicates PRKCA. *mBio.* 2020;11. doi:10.1128/mBio.03343-19
255. Sirugo G, Williams SM, Tishkoff SA. The Missing Diversity in Human Genetic Studies. *Cell.* 2019;177: 26–31. doi:10.1016/j.cell.2019.02.048
256. Gurdasani D, Barroso I, Zeggini E, Sandhu MS. Genomics of disease risk in globally diverse populations. *Nat Rev Genet.* 2019;20: 520–535. doi:10.1038/s41576-019-0144-0
257. Ben-Eghan C, Sun R, Hleap JS, Diaz-Papkovich A, Munter HM, Grant AV, et al. Don't ignore genetic data from minority populations. *Nature.* 2020;585: 184–186. doi:10.1038/d41586-020-02547-3
258. Wojcik GL, Graff M, Nishimura KK, Tao R, Haessler J, Gignoux CR, et al. Genetic analyses of diverse populations improves discovery for complex traits. *Nature.* 2019;570: 514–518. doi:10.1038/s41586-019-1310-4
259. Miller LH, Mason SJ, Clyde DF, McGinniss MH. The resistance factor to *Plasmodium vivax* in blacks. The Duffy-blood-group genotype, FyFy. *N Engl J Med.* 1976;295: 302–304. doi:10.1056/NEJM197608052950602
260. Liu R, Paxton WA, Choe S, Ceradini D, Martin SR, Horuk R, et al. Homozygous defect in HIV-1 coreceptor accounts for resistance of some multiply-exposed individuals to HIV-1 infection. *Cell.* 1996;86: 367–377. doi:10.1016/s0092-8674(00)80110-5
261. Tran NH, Thi T-HN, Tang H-S, Hoang L-P, Nguyen T-HL, Tran N-T, et al. Genetic landscape of recessive diseases in the Vietnamese population from large-scale clinical exome sequencing. *bioRxiv.* 2020; 2020.10.27.358358. doi:10.1101/2020.10.27.358358
262. Nakatsuka N, Moorjani P, Rai N, Sarkar B, Tandon A, Patterson N, et al. The promise of discovering population-specific disease-associated genes in South Asia. *Nat Genet.* 2017;49: 1403–1407. doi:10.1038/ng.3917

263. Bartha I, Carlson JM, Brumme CJ, McLaren PJ, Brumme ZL, John M, et al. A genome-to-genome analysis of associations between human genetic variation, HIV-1 sequence diversity, and viral control. *Elife*. 2013;2: e01123. doi:10.7554/eLife.01123
264. Ansari MA, Aranday-Cortes E, Ip CL, da Silva Filipe A, Lau SH, Bamford C, et al. Interferon lambda 4 impacts the genetic diversity of hepatitis C virus. *eLife*. 8. doi:10.7554/eLife.42463
265. Müller SJ, Schurz H, Tromp G, Spuy G van der, Hoal EG, Helden PD van, et al. A multi-phenotype genome-wide association study of clades causing tuberculosis in a Ghanaian- and South African cohort. *medRxiv*. 2020; 2020.07.27.20162925. doi:10.1101/2020.07.27.20162925
266. McHenry ML, Bartlett J, Igo RP, Wampande EM, Benchek P, Mayanja-Kizza H, et al. Interaction between host genes and *Mycobacterium tuberculosis* lineage can affect tuberculosis severity: Evidence for coevolution? *PLoS Genet*. 2020;16: e1008728. doi:10.1371/journal.pgen.1008728

# Annexes

RESEARCH ARTICLE

# Prevalence and risk factors for latent tuberculosis infection among healthcare workers in Morocco

Ayoub Sabri<sup>1</sup>, Jocelyn Quistrebert<sup>2,3</sup>, Hicham Naji Amrani<sup>4,5</sup>, Ahmed Abid<sup>6,7</sup>, Adil Zegmout<sup>6</sup>, Ismail Abderrhamani Ghorfi<sup>6,7</sup>, Hicham Souhi<sup>6,7</sup>, Abdelhalim Boucaid<sup>6</sup>, Anas Benali<sup>6</sup>, Rachid Abilkassem<sup>7,8</sup>, Mohamed Kmari<sup>8</sup>, Amal Hassani<sup>7,8</sup>, Belyamani Lahcen<sup>7,9</sup>, Samir Siah<sup>7,10</sup>, Erwin Schurr<sup>11</sup>, Stéphanie Boisson-Dupuis<sup>2,3,12</sup>, Jean-Laurent Casanova<sup>2,3,12,13,14</sup>, Amine Lahlou<sup>7,15</sup>, Abdelkader Laatiris<sup>7,15</sup>, Lhoussain Louzi<sup>7,16</sup>, Aziz Ouarsani<sup>4,5</sup>, Ahmed Bourazza<sup>7,17</sup>, Aziz Aouragh<sup>7,18</sup>, Bengshir Mustapha<sup>7,19</sup>, Nezha Messaoudi<sup>7,20</sup>, Aomar Agader<sup>7,8</sup>, Aurélie Cobat<sup>2,3,12</sup>, Laurent Abel<sup>2,3,12</sup>, Jamila El Baghdadi<sup>1\*</sup>



**1** Genetics Unit, Military Hospital Mohamed V, Rabat, Morocco, **2** Laboratory of Human Genetics of Infectious Diseases, Necker Branch, INSERM U1163, Paris, France, **3** Imagine Institute, Paris Descartes University, Paris, France, **4** Department of Pulmonology, Military Hospital Moulay Ismail, Meknes, Morocco, **5** Medical and Pharmacological College, Mohamed Benabdellah University, Fes, Morocco, **6** Department of Pulmonology, Military Hospital Mohammed V, Rabat, Morocco, **7** Medical and Pharmacy School of Rabat, Mohammed V University, Rabat, Morocco, **8** Department of Paediatrics, Military Hospital Mohammed V, Rabat, Morocco, **9** Emergency Department, Mohammed V Military Hospital, Rabat, Morocco, **10** Department of Restorative Plastic Surgery and Burns, Military Hospital Mohammed V, Rabat, Morocco, **11** Infectious Diseases and Immunity in Global Health Program, Research Institute of the McGill University Health Centre, Montreal, Quebec, Canada, **12** St. Giles Laboratory of Human Genetics of Infectious Diseases, Rockefeller Branch, The Rockefeller University, New York, New York, United States of America, **13** Howard Hughes Medical Institute, New York, New York, United States of America, **14** Paediatric Haematology-Immunology Unit, Necker Hospital for Sick Children, Paris, France, **15** Center of Virology and of Infectious and Tropical Diseases, Mohammed V Military Hospital, Rabat, Morocco, **16** Department of Microbiology, Military Hospital Moulay Ismail, Meknes, Morocco, **17** Department of Neurology, Military Hospital Mohammed V, Rabat, Morocco, **18** Department of Gastroenterology, Military Hospital Mohammed V, Rabat, Morocco, **19** Department of Anesthesiology and Resuscitation, Military Hospital Mohammed V, Rabat, Morocco, **20** Laboratory of Hematology and Immunohematology, Military Hospital Mohammed V, Rabat, Morocco

**OPEN ACCESS**

**Citation:** Sabri A, Quistrebert J, Naji Amrani H, Abid A, Zegmout A, Abderrhamani Ghorfi I, et al. (2019) Prevalence and risk factors for latent tuberculosis infection among healthcare workers in Morocco. PLoS ONE 14(8): e0221081. <https://doi.org/10.1371/journal.pone.0221081>

**Editor:** Joan A Caylà, Agencia de Salut Publica de Barcelona, SPAIN

**Received:** March 12, 2019

**Accepted:** July 30, 2019

**Published:** August 15, 2019

**Copyright:** © 2019 Sabri et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

**Data Availability Statement:** All relevant data are within the paper and its Supporting Information files.

**Funding:** The research of the co-first author JQ was supported by a grant from the Fondation pour la Recherche Médicale (FRM) number FDM20170637664. This work was supported by grants from the French National Agency for Research (ANR, grant number ANR-14-CE14-0007-01), the European Research Council (ERC-2010-AdG-268777), the St. Giles Foundation, the

© These authors contributed equally to this work.  
\* [baghdadijamila@yahoo.fr](mailto:baghdadijamila@yahoo.fr)

## Abstract

Increased prevalence of latent tuberculosis infection (LTBI) has been observed among high-risk populations such as healthcare workers (HCWs). The results may depend on the method of LTBI assessment, interferon-gamma release assay (IGRA) and/or tuberculin skin test (TST). Here, we investigated the prevalence and risk factors for LTBI assessed by both IGRAs and TST in HCWs living in Morocco, a country with intermediate tuberculosis (TB) endemicity and high BCG vaccination coverage. HCWs were recruited in two Moroccan hospitals, Rabat and Meknes. All the participants underwent testing for LTBI by both IGRA (QuantiFERON-TB Gold In-Tube, QFT-GIT) and TST. Different combinations of IGRA and TST results defined the LTBI status. Risk factors associated with LTBI were investigated using a mixed-effect logistic regression model. The prevalence of LTBI among 631 HCWs (age range 18–60 years) varied from 40.7% (95%CI 36.9–44.5%) with QFT-GIT to 52% (95%CI 48.2–56.0%) with TST using a 10 mm cut-off. The highest agreement between

National Institute of Allergy and Infectious Diseases (grant number U01AI088685), the Rockefeller University, Institut National de la Santé Et de la Recherche Médicale (INSERM), the Howard Hughes Medical Institute, Paris Descartes University and CNRST.

**Competing interests:** The authors have declared that no competing interests exist.

QFT-GIT and TST ( $\kappa = 0.50$ ; 95%CI 0.43–0.56) was observed with the 10 mm cut-off for a positive TST. For a definition of LTBI status using a double positive result for both QFT-GIT and TST, significant associations were found with the following risk factors: being male (OR = 2.21; 95%CI 1.40–3.49;  $p = 0.0007$ ), belonging to age groups 35–44 years (OR = 2.43; 95%CI 1.45–4.06;  $p = 0.0007$ ) and even more 45–60 years (OR = 4.81; 95%CI 2.72–8.52;  $p = 7.10^{-8}$ ), having a family history of TB (OR = 6.62; 95%CI 2.59–16.94;  $p = 8.10^{-5}$ ), and working at a pulmonology unit (OR = 3.64; 95%CI 1.44–9.23;  $p = 0.006$ ). Smoking was associated with LTBI status when defined by a positive QFT-GIT result (OR = 1.89; 95%CI 1.12–3.21;  $p = 0.02$ ). A high prevalence of LTBI was observed among HCWs in two Moroccan hospitals. Male gender, increased age, family history of TB, and working at a pulmonology unit were consistent risk factors associated with LTBI.

## Introduction

Approximately one quarter of the world population is estimated to have latent tuberculosis infection (LTBI) [1]. At this stage, infected subjects do not experience clinical symptoms and are not contagious. However they have a 5–10% life-time risk of developing active tuberculosis (TB) [2], which is most pronounced during the early stage of infection [3], thereby also becoming an important source of TB contagion. LTBI screening in targeted populations is thus pivotal in order to prevent the spread of TB. Healthcare workers (HCWs) have a higher risk of contracting TB than the general population through occupational exposure [4]. Several studies have investigated the prevalence of and the risk factors for LTBI in HCWs [4–7] in countries highly endemic for TB. Reported pooled prevalences ranged from 37 to 63% in these recent meta-analyses [4–7] with a variability depending on important factors such as countries' income, Bacillus Calmette-Guérin (BCG) vaccination coverage, and LTBI status definition.

The diagnosis of LTBI is inferred indirectly from quantitative measurements of anti-mycobacterial immunity, attesting to induction of acquired immunity to *Mycobacterium tuberculosis* [8]. The tuberculin skin test (TST) is done *in vivo* and consists of an intradermal injection of purified protein derivative (PPD) that provokes a delayed hypersensitivity reaction at the site of injection. It is a sensitive test widely used but it lacks specificity, partly due to cross-reactions with BCG, and, to a lesser extent, non-tuberculous mycobacteria [9]. Another type of tests, interferon-gamma release assays (IGRAs), are performed *in vitro* and detect the secretion of interferon-gamma (IFN- $\gamma$ ) by leukocytes that recognize *M. tuberculosis*-specific antigens. In particular, the QuantiFERON-TB Gold In-Tube test (QFT-GIT) is based on the specific *M. tuberculosis* antigens ESAT-6, CFP10 and TB7.7 (Qiagen, Hilden, Germany). The specificity of these tests is higher than TST and is not affected by BCG vaccination [10]. IGRA results are not fully concordant with TST findings, but they provide complementary information about infection status [11–13]. The World Health Organization recommends tailored LTBI management based on tuberculosis burden and resource availability using either the TST or IGRAs [14]. However, the performance of the two tests is not known in some specific settings. In this study, we investigated the prevalence and risk factors for LTBI assessed by both IGRAs and TST in HCWs from Morocco, a low-middle-income country with an intermediate annual incidence of TB (99/100,000 in 2017), and high BCG vaccination coverage [15–16].

## Materials and methods

### Study design and population

A cross-sectional study was conducted from March 2012 to September 2016 in two Moroccan cities, Rabat and Meknes. HCWs were recruited from two hospitals, the Military Hospital Mohammed V of Rabat and the Military Hospital Moulay Ismail of Meknes. During the four years period of enrollment, the TB incidence in Morocco was stable around 100/100,000. We included in the study all adult HCWs with an age between 18 and 60 years who accepted to participate. Exclusion criteria were ongoing pregnancy, immunosuppressive treatment, history of clinical TB disease, known chronic HIV, hepatitis B or C infections, as well as known cancer or chronic renal failure. Information on the following variables was obtained using a questionnaire: gender, age, comorbidities such as diabetes or asthma (yes/no), smoking status (yes/no), history of TB in family members (yes/no), BCG vaccination, working area and occupation. The working area variable consisted of four relevant categories: non-clinical department (such as administrative sectors or laboratories, with indirect exposure to patients), surgical department (direct exposure to patients but low risk of airborne transmission), clinical department (direct exposure to patients) and the pulmonology unit (direct exposure to high-risk patients). The occupation variable was divided into three categories: the non-medical staff (such as administrative personnel, secretaries), the paramedical staff (lab technicians and auxiliaries), and the medical staff (physicians and nurses). All HCWs declared to have been vaccinated with BCG at birth. The sample size was determined by the number of HCWs who fulfilled the inclusion criteria and responded favorably to the questionnaire. The study was approved by the ethics committee of the Faculty of Medicine of Casablanca, Morocco. Written informed consent was obtained from each participant.

### QuantIFERON-TB Gold In-Tube test (QFT-GIT) assays

QFT-GIT testing was performed on blood samples collected before the TST was carried out, in accordance with the kit manufacturer's instructions. Briefly, 1 ml of peripheral blood was drawn directly into each of three blood collection tubes under vacuum: the nil tube or negative control, the mitogen tube containing phytohaemagglutinin or positive control and the TB specific antigen tube containing specific antigens (early secretory antigenic target 6, culture filtrate protein 10 and TB 7.7 antigen) for *M. tuberculosis*. The contents of each tube were mixed and the tubes were incubated at 37°C for 16 to 24 hours within 16 hours of collection. Then they were centrifuged at 2500 rpm for 15 minutes at room temperature. The supernatant was stored at -20°C. The concentration of IFN- $\gamma$  in each sample was determined by QFT-GIT ELISA (enzyme-linked immunosorbent assay) kit. QFT-GIT results, calculated as TB antigen minus nil value (IU/mL), were categorized as positive ( $\geq 0.35$  IU/mL, QFT+), negative ( $< 0.35$  IU/mL, QFT-) or indeterminate following the software provided by the manufacturer. To further investigate discordant results with TST, QFT-GIT results were sub-classified as strong negative ( $< 0.2$  IU/mL), weak negative (0.2–0.34 IU/mL), weak positive (0.35–0.7 IU/mL) and strong positive ( $> 0.7$  IU/mL) [17–18].

### Tuberculin skin test

The tuberculin skin test (TST) was performed by the medical staff, in accordance with routine hospital procedure. Two units of standardized purified protein derivative (PPD) solution (2 T. U Tuberculin PPD RT 23, Statens Serum Institut, Copenhagen, Denmark) were injected intradermally into the forearm of all the participants using the Mantoux method. Results were evaluated after 72 hours of the intradermal injection. The transversal diameter of the induration

was measured in millimeters. We considered three different cut-offs ( $\geq 5$ , 10 and 15 mm) for a positive result (TST+) to compare with QFT-GIT results. Based on our agreement results, previous studies in similar settings [5] and published guidelines [19], the 10 mm cut-off was used to assess LTBI risk factors.

### Statistical analysis

The concordance between QFT-GIT and TST results was calculated as the number of double positive (QFT+/TST+) plus double negative (QFT-/TST-) test results divided by the total number of test results. The agreement between QFT-GIT and TST results was also calculated using Cohen's kappa coefficient ( $\kappa$ ) and its 95% confidence intervals (CIs). Fisher's exact test was used to compare the proportion of QFT-GIT subcategories between concordant QFT/TST and discordant QFT/TST results. The TST induration between QFT-/TST+ and QFT+/TST+ results and between QFT+/TST- and QFT-/TST- was compared using Mann-Whitney test. We assessed the risk factors associated with TB infection using five different definitions of the LTBI status: (i) QFT+ vs. QFT-, (ii) TST+ vs. TST-, (iii) QFT+/TST+ vs. QFT-/TST-, (iv) discordant TST positive QFT-/TST+ vs. QFT-/TST- and (v) discordant QFT-GIT positive QFT+/TST- vs. QFT-/TST-. All analyses were conducted with a univariable and a multivariable mixed-effects logistic regression with a random effect per hospital to investigate the risk factors associated with LTBI. Statistical significance was considered to be a p-value under 0.05. All analyses were carried out using R software (version 3.4.0) and related packages "psych" and "lme4" [20–22].

## Results

### Healthcare workers

A total of 662 healthy HCWs were recruited from the hospitals of Rabat ( $n = 328$ ) and Meknes ( $n = 334$ ). We excluded 31 participants because of an indeterminate QFT-GIT result. Baseline characteristics of the 631 HCWs with reliable QFT-GIT results are displayed in [Table 1](#). Overall, 54% of HCWs were male ( $n = 341$ ), 47.1% were younger than 35 years old ( $n = 297$ ) and 44.1% of them worked in a non-clinical department ( $n = 278$ ). The majority of HCWs belonged to the medical staff ( $n = 321$ , 50.9%). The prevalence of QFT positivity was 40.7% (95%CI 36.9–44.5%) and that of TST positivity at the 10 mm cut-off was 52.1% (95%CI 48.2–56.0%) ([Fig 1](#)). When considering the LTBI status among individuals with concordant positive (QFT+/TST+) and concordant negative (QFT-/TST-) results, the prevalence of QFT+/TST+ was 45.2% (95%CI 40.7–49.5%).

### Agreement between QFT-GIT and TST results

The agreement between QFT-GIT and TST results was slightly higher with a 10 mm TST cut-off [ $\kappa = 0.50$ ; 95%CI 0.43–0.56; concordance of 74.6%] than with a 5 mm cut-off [ $\kappa = 0.48$ ; 95%CI 0.41–0.54; concordance of 72.7%] or a 15 mm cut-off [ $\kappa = 0.44$ ; 95%CI 0.37–0.51; concordance of 73.3%] ([Table 2](#)). Therefore, we kept the 10 mm cut-off for all further analyses on TST.

We further investigated the levels of the QFT-GIT and TST results in discordant and concordant subjects ([S1 Table](#)). Among 374 HCWs with QFT- results, the proportion of weak negative QFT-GIT result (0.2–0.34 IU/mL) was significantly higher ( $p = 0.0005$ ) in discordant QFT-/TST+ subjects (17/116, 14.7%) than in double negative QFT-/TST- subjects (10/258, 3.9%). When focusing on 257 HCWs with QFT+ results, we did not observe any significant difference in the proportions of weak positive QFT-GIT result (0.35–0.7 IU/mL) between

**Table 1. Baseline characteristics of 631 healthcare workers.**

Characteristics	n (%)
Gender	
Male	341 (54.0)
Female	290 (46.0)
Age group (years)	
18–34	297 (47.1)
35–44	186 (29.5)
45–60	148 (23.4)
Comorbidities	
No	586 (92.9)
Yes	45 (7.1)
Smoking status	
No	549 (87.0)
Yes	82 (13.0)
Family history of TB	
No	588 (93.2)
Yes	43 (6.8)
Work area	
Non-clinical department	278 (44.0)
Surgical department	118 (18.7)
Clinical department	184 (29.2)
Pulmonology unit	51 (8.1)
Occupation	
Non-medical staff	167 (26.4)
Paramedical staff	143 (22.7)
Medical staff	321 (50.9)
QFT-GIT	
Negative	374 (59.3)
Positive	257 (40.7)
TST	
<10 mm	302 (47.9)
≥10 mm	329 (52.1)

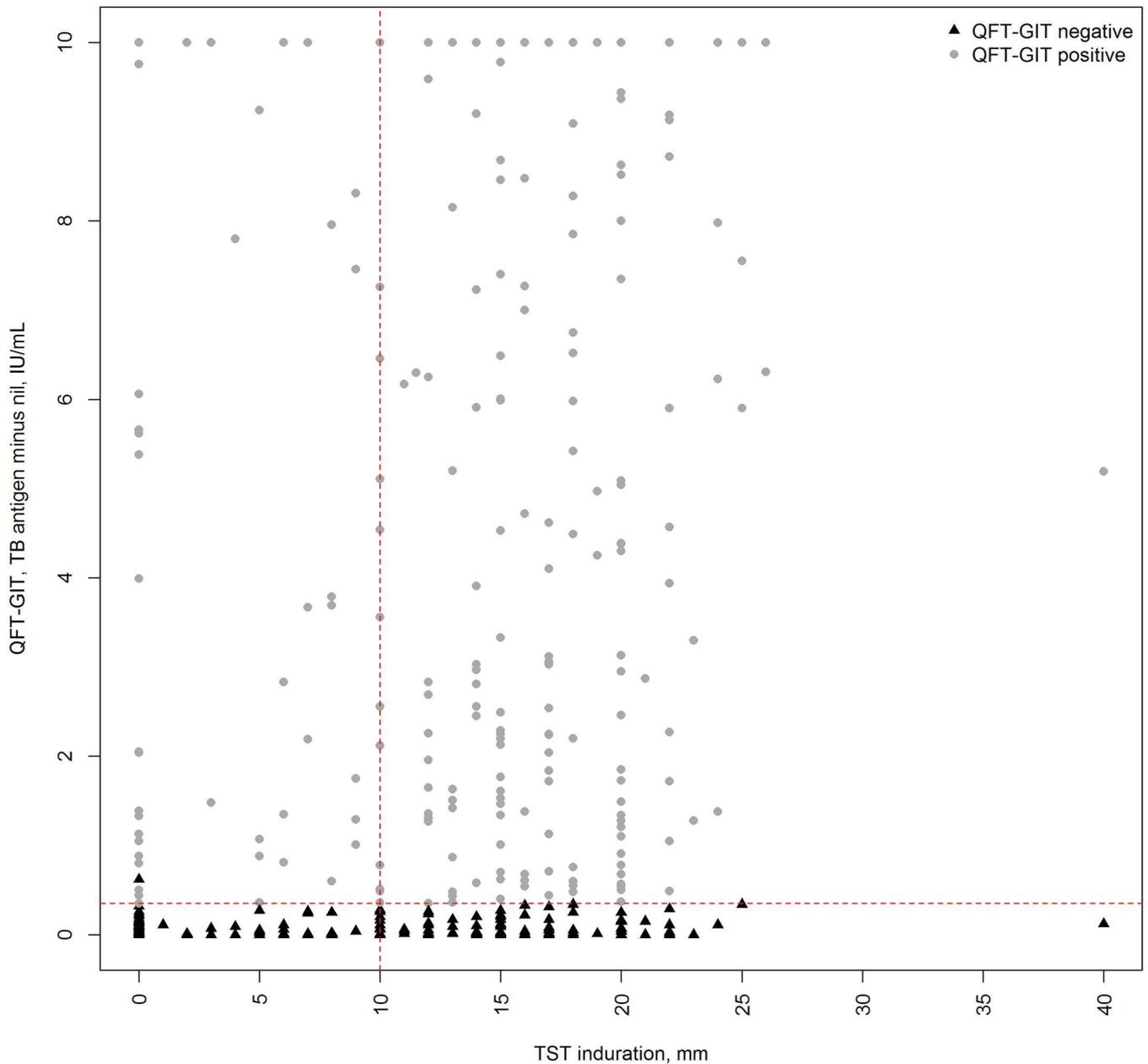
QFT-GIT: QuantiFERON-TB Gold In-Tube; TB: tuberculosis; TST: tuberculin skin test

<https://doi.org/10.1371/journal.pone.0221081.t001>

discordant QFT+/TST- subjects (5/44, 11.4%) and double positive QFT+/TST+ subjects (26/213, 12.2%). Among 302 HCWs with TST- results, the median induration size was significantly higher ( $p = 3.10^{-9}$ ) in discordant QFT+/TST- subjects (median = 3 mm) than in double negative QFT-/TST- subjects (median = 0 mm). Finally, among 329 HCWs with TST+ results, the median induration size was significantly lower ( $p = 0.0002$ ) in discordant QFT-/TST+ subjects (median = 15 mm) than in double positive QFT+/TST+ subjects (median = 17 mm).

### Risk factors associated with LTBI status

When considering the LTBI status defined only by QFT-GIT results, univariable analyses showed that QFT+ was significantly associated with male gender (OR = 2.14; 95%CI 1.54–2.97,  $p = 6.10^{-6}$ ), older age between 35–44 years (OR = 1.80; 95%CI 1.23–2.63;  $p = 0.003$ ) and 45–60 years (OR = 3.02; 95%CI 1.99–4.57;  $p = 2.10^{-7}$ ), smoking (OR = 2.57; 95%CI 1.59–4.15;



**Fig 1. Scatterplot of quantitative QuantiFERON-TB Gold In-Tube (QFT-GIT) and tuberculin skin test (TST) results of 631 healthcare workers.** The horizontal red dashed line represents a QFT-GIT cut-off of 0.35 IU/mL and the vertical one a TST cut-off of 10 mm. Samples with a QFT-GIT TB Antigen value inferior to the nil value were truncated to 0 (n = 122) and samples with a QFT-GIT TB Antigen value > 10 IU/mL were truncated to 10 (n = 63). One sample had a QFT-GIT TB Antigen minus nil value superior to 0.35 IU/mL but inferior to 25% of the nil value leading to a negative test result.

<https://doi.org/10.1371/journal.pone.0221081.g001>

p = 0.0001) and family history of TB (OR = 3.16; 95%CI 1.63–6.12; p = 0.0007). All these variables remained significant in multivariable analysis (Table 3).

When the LTBI status was defined only by TST, the same variables as those found in the previous analysis were significantly associated with TST+ in univariable analysis, with an additional one corresponding to working at a pulmonology unit (OR = 2.31; 95%CI 1.21–4.40; p = 0.01) (Table 4). In multivariable analysis, male gender (OR = 1.70; 95%CI 1.15–2.51;

**Table 2. Agreement between QFT-GIT and TST results in 631 healthcare workers according to various TST cut-offs (5 mm, 10 mm and 15 mm).**

	TST 5 mm	TST 10 mm	TST 15 mm
QFT-/TST-	226	258	303
QFT+/TST+	233	213	160
QFT-/TST+	148	116	71
QFT+/TST-	24	44	97
Concordance, %	72.7	74.6	73.3
Kappa (95% CI)	0.48 (0.41–0.54)	0.50 (0.43–0.56)	0.44 (0.37–0.51)

QFT-GIT: QuantiFERON-TB Gold In-Tube; TST: tuberculin skin test

<https://doi.org/10.1371/journal.pone.0221081.t002>

**Table 3. Risk factors associated with a positive QFT-GIT result in 631 healthcare workers.**

Variables	Total n = 631	QFT- n = 374	QFT+ n = 257	OR <sub>uni</sub> <sup>a</sup> (95%CI)	p <sub>uni</sub> <sup>b</sup>	OR <sub>mult</sub> <sup>c</sup> (95%CI)	p <sub>mult</sub> <sup>d</sup>
Gender							
Female	290	200	90	Ref		Ref	
Male	341	174	167	<b>2.14 (1.54–2.97)</b>	<b>6.10<sup>-6</sup></b>	<b>1.69 (1.17–2.46)</b>	<b>0.006</b>
Age group (years)							
18–34	297	207	90	Ref		Ref	
35–44	186	104	82	<b>1.80 (1.23–2.63)</b>	<b>0.003</b>	<b>1.57 (1.04–2.37)</b>	<b>0.03</b>
45–60	148	63	85	<b>3.02 (1.99–4.57)</b>	<b>2.10<sup>-7</sup></b>	<b>2.74 (1.74–4.31)</b>	<b>1.10<sup>-5</sup></b>
Comorbidities							
No	586	348	238	Ref		Ref	
Yes	45	26	19	1.08 (0.58–1.99)	0.82	0.95 (0.49–1.85)	0.88
Smoking status							
No	549	342	207	Ref		Ref	
Yes	82	32	50	<b>2.57 (1.59–4.15)</b>	<b>0.0001</b>	<b>1.89 (1.12–3.21)</b>	<b>0.02</b>
Family history of TB							
No	588	360	228	Ref		Ref	
Yes	43	14	29	<b>3.16 (1.63–6.12)</b>	<b>0.0007</b>	<b>3.41 (1.70–6.85)</b>	<b>0.0006</b>
Work area							
Non-clinical department	278	160	118	Ref		Ref	
Surgical department	118	79	39	0.94 (0.64–1.38)	0.74	1.40 (0.84–2.34)	0.19
Clinical department	184	106	78	0.68 (0.43–1.07)	0.10	0.73 (0.41–1.27)	0.26
Pulmonology unit	51	29	22	1.09 (0.59–2.01)	0.77	1.80 (0.88–3.69)	0.11
Job category							
Non-medical staff	167	89	78	Ref		Ref	
Paramedical staff	143	80	63	0.91 (0.58–1.43)	0.68	1.00 (0.58–1.74)	0.99
Medical staff	321	205	116	0.65 (0.44–1.05)	0.10	0.70 (0.41–1.19)	0.19

CI: confidence interval; QFT-GIT: QuantiFERON-TB Gold In-Tube; TB: tuberculosis

<sup>a</sup>OR<sub>uni</sub>: odds ratio obtained in a univariable regression

<sup>b</sup>p<sub>uni</sub>: p-value corresponding to the coefficient obtained in a univariable regression

<sup>c</sup>OR<sub>mult</sub>: odds ratio obtained in a multivariable regression including all the variables explored in univariable analyses

<sup>d</sup>p<sub>mult</sub>: p-value corresponding to the coefficient obtained in a multivariable regression including all the variables explored in univariable analyses

Bold text indicates a statistically significant result

<https://doi.org/10.1371/journal.pone.0221081.t003>

**Table 4. Risk factors associated with a positive TST result in 631 healthcare workers.**

Variables	Total n = 631	TST- n = 302	TST+ n = 329	OR <sub>uni</sub> <sup>a</sup> (95%CI)	P <sub>uni</sub> <sup>b</sup>	OR <sub>mult</sub> <sup>c</sup> (95%CI)	P <sub>mult</sub> <sup>d</sup>
Gender							
Female	290	165	125	Ref		Ref	
Male	341	137	204	<b>2.10 (1.50–2.94)</b>	<b>2.10<sup>-5</sup></b>	<b>1.70 (1.15–2.51)</b>	<b>0.008</b>
Age group (years)							
18–34	297	185	112	Ref		Ref	
35–44	186	73	113	<b>2.58 (1.74–3.84)</b>	<b>3.10<sup>-6</sup></b>	<b>2.80 (1.82–4.30)</b>	<b>3.10<sup>-6</sup></b>
45–60	148	44	104	<b>3.50 (2.25–5.45)</b>	<b>3.10<sup>-8</sup></b>	<b>3.68 (2.26–5.99)</b>	<b>2.10<sup>-7</sup></b>
Comorbidities							
No	586	280	306	Ref		Ref	
Yes	45	32	23	0.98 (0.52–1.85)	0.95	0.82 (0.41–1.65)	0.58
Smoking status							
No	549	273	276	Ref		Ref	
Yes	82	29	53	<b>1.83 (1.10–3.05)</b>	<b>0.02</b>	1.27 (0.71–2.26)	0.42
Family history of TB							
No	588	294	294	Ref		Ref	
Yes	43	8	35	<b>4.04 (1.79–9.10)</b>	<b>0.0008</b>	<b>4.89 (2.04–11.74)</b>	<b>0.0004</b>
Work area							
Non-clinical department	278	144	134	Ref		Ref	
Surgical department	118	57	61	1.07 (0.72–1.59)	0.75	1.41 (0.83–2.41)	0.21
Clinical department	184	81	103	1.28 (0.81–2.02)	0.29	1.29 (0.73–2.29)	0.39
Pulmonology unit	51	20	31	<b>2.31 (1.21–4.40)</b>	<b>0.01</b>	<b>3.79 (1.77–8.13)</b>	<b>0.0006</b>
Job category							
Non-medical staff	143	60	83	Ref		Ref	
Paramedical staff	167	82	85	1.46 (0.91–2.35)	0.12	1.51 (0.84–2.71)	0.17
Medical staff	321	160	161	1.02 (0.69–1.51)	0.93	1.07 (0.61–1.88)	0.81

CI: confidence interval; TB: tuberculosis; TST: tuberculin skin test

<sup>a</sup>OR<sub>uni</sub>: odds ratio obtained in a univariable regression

<sup>b</sup>P<sub>uni</sub>: p-value corresponding to the coefficient obtained in a univariable regression

<sup>c</sup>OR<sub>mult</sub>: odds ratio obtained in a multivariable regression including all the variables explored in univariable analyses

<sup>d</sup>P<sub>mult</sub>: p-value corresponding to the coefficient obtained in a multivariable regression including all the variables explored in univariable analyses

Bold text indicates a statistically significant result

<https://doi.org/10.1371/journal.pone.0221081.t004>

p = 0.008), older age between 35–44 years (OR = 2.80; 95%CI 1.82–4.30; p = 3.10<sup>-6</sup>) or 45–60 years (OR = 3.68; 95%CI 2.26–5.99; p = 2.10<sup>-7</sup>), family history of TB (OR = 4.89; 95%CI 2.04–11.74; p = 0.0004) and working at a pulmonology unit (OR = 3.79; 95%CI 1.77–8.13; p = 0.0006] were still associated with TST+. However smoking was no more a risk factor when adjusting for other covariates (OR = 1.27; 95%CI 0.71–2.26; p = 0.42) in particular gender, probably explained by the fact that all smokers were males (n = 82).

We then considered a more stringent definition of the LTBI status by focusing on the double positive QFT+/TST+ (n = 213) vs. the double negative QFT-/TST- subjects (n = 258). The same risk factors as those identified in the previous analysis focusing on TST only were found to be significantly associated with LTBI in a multivariable analysis (Table 5): being a male (OR = 2.21; 95%CI 1.40–3.49; p = 0.0007), having an older age between 35–44 years (OR = 2.43; 95%CI 1.45–4.06; p = 0.0007) or 45–60 years (OR = 4.81; 95%CI 2.72–8.52; p = 7.10<sup>-8</sup>), family history of TB (OR = 6.62; 95%CI 2.59–16.94; p = 8.10<sup>-5</sup>) and working at a

**Table 5. Risk factors from multivariable regressions associated with positive concordant results (QFT+/TST+) results compared to the reference group (negative concordant results QFT-/TST-) in healthcare workers.**

Variables	Total n = 471	QFT- TST- n = 258	QFT+ TST+ n = 213	OR <sub>mult</sub> <sup>a</sup> (95%CI)	p <sub>mult</sub> <sup>b</sup>
Gender					
Female	229	152	77	Ref	
Male	242	106	136	<b>2.21 (1.40–3.49)</b>	<b>0.0007</b>
Age group (years)					
18–34	245	170	75	Ref	
35–44	121	56	65	<b>2.43 (1.45–4.06)</b>	<b>0.0007</b>
45–60	105	32	73	<b>4.81 (2.72–8.52)</b>	<b>7.10<sup>-8</sup></b>
Comorbidities					
No	440	241	199	Ref	
Yes	31	17	14	0.80 (0.34–1.91)	0.62
Smoking status					
No	414	240	174	Ref	
Yes	57	18	39	1.83 (0.90–3.74)	0.10
Family history of TB					
No	436	251	185	Ref	
Yes	35	7	28	<b>6.62 (2.59–16.94)</b>	<b>8.10<sup>-5</sup></b>
Work area					
Non-clinical department	204	115	89	Ref	
Surgical department	90	54	36	1.46 (0.78–2.72)	0.23
Clinical department	43	73	70	0.87 (0.44–1.72)	0.68
Pulmonology unit	34	16	18	<b>3.64 (1.44–9.23)</b>	<b>0.006</b>
Job category					
Non-medical staff	118	61	57	Ref	
Paramedical staff	107	52	55	1.31 (0.66–2.62)	0.44
Medical staff	246	145	101	0.82 (0.43–1.56)	0.54

CI: confidence interval; QFT: QuantiFERON-TB Gold In-Tube; TB: tuberculosis; TST: tuberculin skin test

<sup>a</sup>OR<sub>mult</sub>: odds ratio obtained in a multivariable regression including all the variables explored in univariable analyses

<sup>b</sup>p<sub>mult</sub>: p-value corresponding to the coefficient obtained in a multivariable regression including all the variables explored in univariable analyses

Bold text indicates a statistically significant result

<https://doi.org/10.1371/journal.pone.0221081.t005>

pulmonology unit (OR = 3.64; 95%CI 1.44–9.23; p = 0.006). Again, smoking was not a significant risk factor for QFT+/TST+ when adjusting for other covariates (OR = 1.83; 95%CI 0.90–3.74; p = 0.10).

### Risk factors associated with discordant QFT-GIT and TST results

Finally, we performed a specific analysis on the individuals with discordant results, QFT-/TST+ (n = 116) and QFT+/TST- (n = 44) that we compared to QFT-/TST- subjects (n = 258) (Table 6). For QFT-/TST+ individuals, we obtained in multivariable analysis results similar to those observed when considering all TST+ subjects with significant associations for male gender (OR = 1.73; 95% CI: 1.02–2.93; p = 0.04), older age between 35–44 years (OR = 4.18; 95% CI: 2.34–7.44; p = 1.10<sup>-6</sup>) or 45–60 years (OR = 4.35; 95%CI 2.16–8.75; p = 2.10<sup>-7</sup>) and working at a pulmonology unit (OR = 3.95; 95%CI 1.43–10.88; p = 0.008) without any effect of smoking (OR = 0.98; 95%CI 0.40–2.38; p = 0.97).

**Table 6. Risk factors from multivariable regressions associated with discordant results (QFT-/TST+, QFT+/TST-) compared to the reference group (negative concordant results QFT-/TST-) in healthcare workers.**

Variables	QFT-TST- n = 258	QFT-TST+ n = 116	OR <sub>mult</sub> <sup>a</sup> (95%CI)	P <sub>mult</sub> <sup>b</sup>	QFT+TST- n = 44	OR <sub>mult</sub> <sup>a</sup> (95%CI)	P <sub>mult</sub> <sup>b</sup>
Gender							
Female	152	48	Ref		13	Ref	
Male	106	68	<b>1.73 (1.02–2.93)</b>	<b>0.04</b>	31	<b>2.37 (1.07–5.25)</b>	<b>0.03</b>
Age group (years)							
18–34	170	37	Ref		15	Ref	
35–44	56	48	<b>4.18 (2.34–7.44)</b>	<b>1.10<sup>-6</sup></b>	17	2.41 (0.99–5.88)	0.05
45–60	32	31	<b>4.35 (2.16–8.75)</b>	<b>4.10<sup>-5</sup></b>	12	<b>3.17 (1.18–8.50)</b>	<b>0.02</b>
Comorbidities							
No	241	107	Ref		39	Ref	
Yes	17	9	0.90 (0.35–2.36)	0.84	5	1.23 (0.35–4.40)	0.75
Smoking status							
No	240	102	Ref		33	Ref	
Yes	18	14	0.98 (0.40–2.38)	0.97	11	<b>3.35 (1.17–9.60)</b>	<b>0.02</b>
Family history of TB							
No	251	109	Ref		43	Ref	
Yes	7	7	1.89 (0.52–6.91)	0.33	1	1.65 (0.18–15.56)	0.66
Work area							
Non-clinical department	115	45	Ref		29	Ref	
Surgical department	54	25	1.27 (0.60–2.66)	0.53	3	0.80 (0.83–2.41)	0.68
Clinical department	73	33	1.22 (0.56–2.67)	0.62	8	<b>0.17 (0.04–0.72)</b>	<b>0.02</b>
Pulmonology unit	16	13	<b>3.95 (1.43–10.88)</b>	<b>0.008</b>	4	1.70 (0.43–6.72)	0.46
Job category							
Non-medical staff	61	28	Ref		21	Ref	
Paramedical staff	52	28	1.56 (0.69–3.55)	0.29	8	0.85 (0.29–2.53)	0.78
Medical staff	145	60	1.22 (0.55–2.68)	0.62	15	0.54 (0.20–1.49)	0.24

CI: confidence interval; QFT: QuantiFERON-TB Gold In-Tube; TB: tuberculosis; TST: tuberculin skin test

<sup>a</sup>OR<sub>mult</sub>: odds ratio obtained in a multivariable regression including all the variables explored in univariable analyses

<sup>b</sup>P<sub>mult</sub>: p-value corresponding to the coefficient obtained in a multivariable regression including all the variables explored in univariable analyses

Bold text indicates a statistically significant result

<https://doi.org/10.1371/journal.pone.0221081.t006>

For QFT+/TST- individuals, we observed in multivariable analysis results close to those with the QFT+ analysis, with less significant p-values due to the smaller sample size. An association was observed with male gender (OR = 2.37; 95%CI 1.07–5.25; p = 0.03), older age which is borderline non-significant for the group 35–44 years (OR = 2.41; 95%CI 0.99–5.88; p = 0.05) and significant for the group 45–60 years (OR = 3.17; 95%CI 1.18–8.50; p = 0.02) and smoking (OR = 3.35; 95%CI 1.17–9.60; p = 0.02). We also noted a protective effect of working at a clinical department (OR = 0.17; 95%CI 0.04–0.72; p = 0.02).

## Discussion

We report here the first study of LTBI conducted in HCWs from Morocco, a middle-income country with intermediate endemicity for TB and high BCG vaccination coverage. The prevalence of LTBI ranged from 40.7% (36.9–44.5%) with QFT+ to 52.1% (48.2–56.0%) with TST+ and was 45.2% (40.7–49.5%) when considering the double positive QFT+/TST+ status.

These values are consistent with prevalences observed in HCWs living in high TB burden countries such as Brazil (37%), China (54%), and India (43%) [6]. The higher proportion of TST+ compared to QFT+ in Morocco is consistent with a meta-analysis showing the same pattern in 24 out of 25 studies comparing IGRAs with TST in HCWs [23]. This observation may be explained in part by the higher specificity of IGRAs [9], and an extensive BCG vaccination coverage as it is the case in Morocco. The overall agreement between QFT-GIT and TST results ( $\kappa = 0.50$  at a 10 mm cut-off) was also similar to previous estimates in high TB incidence countries ( $\kappa = 0.38$  in a recent meta-analysis ranging from 0.19 in China to 0.61 in India [24]).

Interestingly, we identified risk factors that were consistently associated with LTBI, whatever its definition, namely gender, age, and familial history of TB. Male gender is a well-known risk factor for active TB disease due at least in part to socio-cultural causes as well as biological differences between sexes [25–26] but this gender disparity has not been fully investigated for LTBI. A significant higher proportion of LTBI in adult men than in adult women was reported in several studies [27–29] except one in a moderate TB burden country (Taiwan) that did not find gender as a risk factor while taking into account potential confounders [30]. The increase of LTBI prevalence with age is a well-known observation which has been previously reported either with QFT-GIT [31–32] or TST [33]. In our study, age was extremely correlated with the length of employment as almost all HCWs in military hospitals are recruited when they are young adults (between 18–21 years). Therefore, the increased risk of LTBI in the age groups belonging to 35–44 and 45–60 years is probably explained by a longer cumulative exposure that could be occupational and/or non-occupational.

We also identified other risk factors that reflect a close contact to infectious TB patients either through household exposure with a family TB history or an occupational one as working at a pulmonology unit, as already reported [27,31,34]. Beside the work area, we did not find any association with the job category and other potential occupational exposures may have not been captured. We also observed some interesting effects related to smoking. First, the smoking effect observed in univariable analyses decreased (QFT-GIT) or disappeared (TST) when adjusting for other covariates, in particular gender, probably because all smokers were males. In addition, when defining LTBI status by QFT-GIT, the smoking effect appeared to be stronger when considering discordant QFT+/TST- than concordant QFT+/TST+. Smoking was previously reported to be a risk factor for LTBI when assessed by QFT-GIT and/or TST [27–28,35], and also to affect the performance of the QFT-GIT test itself [36]. Further specific investigations should be conducted to investigate the relationship between smoking and IGRAs results with particular attention to confounding factors such as age, gender or socioeconomic status.

Finally, we focused on the discordant results between the two tests. As expected, QFT-/TST+ results were more frequent (116/631, 18.4%) than QFT+/TST- (44/631, 7.0%) [23]. The QFT-/TST+ discordance is often attributed in part to BCG vaccination and non-tuberculous mycobacteria. Interestingly, QFT-/TST+ subjects shared the same risk factors as those identified in the analysis for the QFT+/TST+ subjects. Moreover, we found a stronger proportion of weak negative QFT-GIT results in QFT-/TST+ than in QFT-/TST- (14.7% vs. 3.9%) indicating that some of these individuals may have a negative result due to the substantial variability of IGRA tests [10]. We also observed that the proportion of QFT-/TST+ was higher (79/334, 23.7%) in the oldest age group 35–60 years, when compared with the younger age group 18–34 (37/297, 12.5%) (S2 Table), consistent with the view that discordant QFT-/TST+ results may be due in part to temporally more remote infections [37]. Regarding the other discordant group QFT+/TST-, we also observed risk factors similar to those obtained with the QFT+ outcome, and this group had larger TST reactions than those of the QFT-/TST- group. Also, we

cannot exclude that some of these QFT+/TST- subjects displayed a complete nonspecific cutaneous anergy that may explain a negative TST [38]. All these observations suggest that a substantial proportion of QFT+/TST- and QFT-/TST+ discordant HCWs may be truly infected individuals. Overall, the results we observed in terms of prevalence and risk factors for LTBI in HCWs living in Morocco were similar when using TST or QFT-GIT, suggesting that either could be used for a screening.

## Supporting information

**S1 Table. Comparison of discordant QFT-/TST+ vs. double negative QFT-/TST- results and QFT+/TST- vs. double positive QFT+/TST+ results according to QFT-GIT TB antigen minus nil categories.**

(DOC)

**S2 Table. Agreement between QFT-GIT and TST results in 631 healthcare workers according to age category and a TST cut-off at 10 mm.**

(DOC)

## Acknowledgments

The authors thank all the participating HCWs from Rabat and Meknes for their time and cooperation in the study.

## Author Contributions

**Conceptualization:** Jamila El Baghdadi.

**Data curation:** Jamila El Baghdadi.

**Formal analysis:** Jamila El Baghdadi.

**Funding acquisition:** Laurent Abel, Jamila El Baghdadi.

**Investigation:** Laurent Abel, Jamila El Baghdadi.

**Methodology:** Ayoub Sabri, Hicham Naji Amrani, Ahmed Abid, Adil Zegmout, Ismail Abderhamani Ghorfi, Hicham Souhi, Abdelhalim Boucaid, Anas Benali, Rachid Abilkassem, Mohamed Kmari, Amal Hassani, Belyamani Lahcen, Samir Siah, Amine Lahlou, Lhoussain Louzi, Aziz Ouarssani, Ahmed Bourazza, Aziz Aouragh, Bensghir Mustapha, Nezha Mes-saudi, Laurent Abel, Jamila El Baghdadi.

**Project administration:** Jamila El Baghdadi.

**Resources:** Jamila El Baghdadi.

**Software:** Jocelyn Quistrebert, Aurélie Cobat.

**Supervision:** Ayoub Sabri, Hicham Naji Amrani, Stéphanie Boisson-Dupuis, Jamila El Baghdadi.

**Validation:** Jocelyn Quistrebert, Aurélie Cobat, Jamila El Baghdadi.

**Visualization:** Ahmed Abid, Erwin Schurr, Jean-Laurent Casanova, Abdelkader Laatiris, Aomar Agader, Aurélie Cobat, Jamila El Baghdadi.

**Writing – original draft:** Jocelyn Quistrebert, Laurent Abel, Jamila El Baghdadi.

## References

1. Houben RMGJ, Dodd PJ. The Global Burden of Latent Tuberculosis Infection: A Re-estimation Using Mathematical Modelling. *PLoS Med* 2016; 13:e1002152 <https://doi.org/10.1371/journal.pmed.1002152> PMID: 27780211
2. Vynnycky E, Fine PE. The natural history of tuberculosis: the implications of age-dependent risks of disease and the role of reinfection. *Epidemiol Infect* 1997; 119:183–201 <https://doi.org/10.1017/S0950268897007917> PMID: 9363017
3. Behr MA, Edelstein PH, Ramakrishnan L. Revisiting the timetable of tuberculosis. *BMJ* 2018; 362:k2738 <https://doi.org/10.1136/bmj.k2738> PMID: 30139910
4. Menzies D, Joshi R, Pai M. Risk of tuberculosis infection and disease associated with work in health care settings. *Int J Tuberc Lung Dis* 2007; 11:593–605 PMID: 17519089
5. Joshi R, Reingold AL, Menzies D, Pai M. Tuberculosis among health-care workers in low- and middle-income countries: a systematic review. *PLoS Med* 2006; 3:e494 <https://doi.org/10.1371/journal.pmed.0030494> PMID: 17194191
6. Nasreen S, Shokoohi M, Malvankar-Mehta MS. Prevalence of Latent Tuberculosis among Health Care Workers in High Burden Countries: A Systematic Review and Meta-Analysis. *PLoS One* 2016; 11:e0164034 <https://doi.org/10.1371/journal.pone.0164034> PMID: 27711155
7. Uden L, Barber E, Ford N, Cooke GS. Risk of Tuberculosis Infection and Disease for Health Care Workers: An Updated Meta-Analysis. *Open Forum Infect Dis* 2017; 4:ofx137
8. Getahun H, Matteelli A, Chaisson RE, Raviglione M. Latent Mycobacterium tuberculosis infection. *N Engl J Med* 2015; 372:2127–2135 <https://doi.org/10.1056/NEJMra1405427> PMID: 26017823
9. 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* 2006; 10:1192–1204 PMID: 17131776
10. Pai M, Denkinger CM, Kik SV, Rangaka M, Zwerling A, Oxlade O et al. Gamma Interferon Release Assays for Detection of Mycobacterium tuberculosis Infection. *Clin Microbiol Rev* 2014; 27:3–20 <https://doi.org/10.1128/CMR.00034-13> PMID: 24396134
11. Pai M, Gokhale K, Joshi R, Dogra S, Kalantri S, Mendiratta D et al. Mycobacterium tuberculosis infection in health care workers in rural India: comparison of a whole-blood interferon gamma assay with tuberculin skin testing. *JAMA* 2005; 293:2746–2755 <https://doi.org/10.1001/jama.293.22.2746> PMID: 15941804
12. 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; 137:1071–7 <https://doi.org/10.1378/chest.09-1852> PMID: 20040612
13. 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; 146:340–354 <https://doi.org/10.7326/0003-4819-146-5-200703060-00006> PMID: 17339619
14. Getahun H, Matteelli A, Abubakar I, Hauer B, Pontali E, Migliori GB. Advancing global programmatic management of latent tuberculosis infection for at risk populations. *Eur Respir J*. 2016; 47:1327–1330 <https://doi.org/10.1183/13993003.00449-2016> PMID: 27132266
15. World Health Organization WHO. Global tuberculosis report 2018. [http://www.who.int/tb/publications/global\\_report/en/](http://www.who.int/tb/publications/global_report/en/). Accessed 2 Nov 2018
16. World Health Organization WHO. WHO vaccine-preventable diseases: monitoring system 2018 global summary. [http://apps.who.int/immunization\\_monitoring/globalsummary/timeseries/tswucoveragebcg.html](http://apps.who.int/immunization_monitoring/globalsummary/timeseries/tswucoveragebcg.html) Accessed 2 Nov 2018
17. van Zyl-Smit RN, Pai M, Pehrah 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; 180:49–58 <https://doi.org/10.1164/rccm.200811-1704OC> PMID: 19342414
18. Nemes E, Rozot V, Geldenhuys H, Bilek N, Mabwe S, Abrahams D et al. Optimization and Interpretation of Serial QuantiFERON Testing to Measure Acquisition of Mycobacterium tuberculosis Infection. *Am J Respir Crit Care Med* 2017; 196:638–648 <https://doi.org/10.1164/rccm.201704-0817OC> PMID: 28737960
19. American Thoracic Society; Centers for Disease Control and Prevention. Targeted tuberculin testing and treatment of latent tuberculosis infection. 2000; 49(RR-6): 1–51.
20. R: The R Project for Statistical Computing. <https://www.r-project.org/>. Accessed 5 Nov 2018
21. Revelle W. (2018) psych: Procedures for Psychological, Psychometric, and Personality Research Northwestern University, Evanston, Illinois, USA, <https://CRAN.R-project.org/package=psychVersion=1.8.12>

22. Bates D, Maechler M, Bolker B, Walker S. Fitting Linear Mixed-Effects Models Using lme4. *J Stat Softw* 2018; 67:1–48
23. Zwerling A, van den Hof S, Scholten J, Cobelens F, Menzies D, Pai M. Interferon-gamma release assays for tuberculosis screening of healthcare workers: a systematic review. *Thorax* 2012; 67:62–70 <https://doi.org/10.1136/thx.2010.143180> PMID: 21228420
24. Lamberti M, Uccello R, Monaco MGL, Muoio M, Feola D, Sannolo N et al. Tuberculin skin test and QuantiFERON test agreement and influencing factors in tuberculosis screening of healthcare workers: a systematic review and meta-analysis. *J Occup Med Toxicol*. 2015; 10:2 <https://doi.org/10.1186/s12995-015-0044-y> PMID: 25670962
25. Horton KC, MacPherson P, Houben RMGJ, White RG, Corbett EL. Sex Differences in Tuberculosis Burden and Notifications in Low- and Middle-Income Countries: A Systematic Review and Meta-analysis. *PLoS Med*. 2016; 13:e1002119 <https://doi.org/10.1371/journal.pmed.1002119> PMID: 27598345
26. Neyrolles O, Quintana-Murci L. Sexual Inequality in Tuberculosis. *PLoS Med* 2009; 6:e1000199 <https://doi.org/10.1371/journal.pmed.1000199> PMID: 20027210
27. de Souza FM, do Prado TN, dos S Pinheiro J, Peres RL, Lacerda TC, Loureiro RB et al. Comparison of Interferon- $\gamma$  Release Assay to Two Cut-Off Points of Tuberculin Skin Test to Detect Latent Mycobacterium tuberculosis Infection in Primary Health Care Workers. *PLoS One* 2014; 9:e102773 <https://doi.org/10.1371/journal.pone.0102773> PMID: 25137040
28. He G, Li Y, Zhao F, Wang L, Cheng S, Guo H et al. The Prevalence and Incidence of Latent Tuberculosis Infection and Its Associated Factors among Village Doctors in China. *PLoS One* 2015; 10:e0124097 <https://doi.org/10.1371/journal.pone.0124097> PMID: 25996960
29. He GX, Wang LX, Chai SJ, Klana JD, Cheng SM, Ren YL et al Risk factors associated with tuberculosis infection among health care workers in Inner Mongolia, China. *Int J Tuberc Lung Dis* 2012; 16:1485–1491 <https://doi.org/10.5588/ijtld.12.0193> PMID: 22964074
30. Ting WY, Huang SF, Lee MC, Lin YY, Lee YC, Feng JY et al. Gender disparities in latent tuberculosis infection in high-risk individuals: a cross-sectional study. *PLoS One* 2014; 9:e110104 <https://doi.org/10.1371/journal.pone.0110104> PMID: 25369472
31. Harada N, Nakajima Y, Higuchi K, Sekiya Y, Rothel J, Mori T Screening for tuberculosis infection using whole-blood interferon-gamma and Mantoux testing among Japanese healthcare workers. *Infect Control Hosp Epidemiol* 2006; 27:442–448 <https://doi.org/10.1086/504358> PMID: 16671023
32. Jo KW, Hong Y, Park JS, Bae IG, Eom JS, Lee SR et al. Prevalence of Latent Tuberculosis Infection among Health Care Workers in South Korea: A Multicenter Study. *Tuberc Respir Dis (Seoul)* 2013; 75:18–24
33. Rutanga C, Lowrance DW, Oeltmann JE, Mutembayire G, Willis M, Uwizeye CB et al. Latent Tuberculosis Infection and Associated Factors among Health Care Workers in Kigali, Rwanda. *PLoS One* 2015; 10:e0124485 <https://doi.org/10.1371/journal.pone.0124485> PMID: 25919759
34. Lien LT, Hang NTL, Kobayashi N, Yanai H, Toyota E, Sakurada S et al. Prevalence and Risk Factors for Tuberculosis Infection among Hospital Workers in Hanoi, Viet Nam. *PLoS One* 2009; 4:e6798 <https://doi.org/10.1371/journal.pone.0006798> PMID: 19710920
35. Zhang H, Xin H, Li X, Li H, Li M, Lu W et al. A dose-response relationship of smoking with tuberculosis infection: A cross-sectional study among 21008 rural residents in China. *PLoS One*. 2017; 12: e0175183 <https://doi.org/10.1371/journal.pone.0175183> PMID: 28384350
36. Aabye MG, Hermansen TS, Ruhwald M, PrayGod G, Faurholt-Jepsen M, Range N et al. Negative effect of smoking on the performance of the QuantiFERON TB gold in tube test. *BMC Infect Dis* 2012; 12:379 <https://doi.org/10.1186/1471-2334-12-379> PMID: 23270417
37. Pai M, Kalantri S, Menzies D. Discordance between tuberculin skin test and interferon-gamma assays. *Int J Tuberc Lung Dis* 2006; 10:942–943 PMID: 16898382
38. Slovis BS, Plitman JD, Haas DW. The case against anergy testing as a routine adjunct to tuberculin skin testing. *JAMA* 2000; 283:2003–2007 <https://doi.org/10.1001/jama.283.15.2003> PMID: 10789669

**S1 Table.** Comparison of discordant QFT-/TST+ vs. double negative QFT-/TST- results and QFT+/TST- vs. double positive QFT+/TST+ results according to QFT-GIT TB Antigen minus nil categories

	QFT-/TST+ (n=116)	QFT-/TST- (n=258)	p <sup>a</sup>	QFT+/TST- (n=44)	QFT+/TST+ (n=213)	p <sup>a</sup>
TB Antigen minus nil (IU/mL)			0.0005			1
<0.2	99	247		0	0	
0.2–0.34	17	10		0	0	
0.35–0.7	0	1		5	26	
>0.7	0	0		39	187	

QFT-GIT: QuantiFERON-TB Gold In-Tube ; TST: tuberculin skin test

<sup>a</sup>p-value of Fisher's exact test

**S2 Table.** Agreement between QFT-GIT and TST results in 631 healthcare workers according to age category and a TST cut-off at 10 mm

	<b>TST 10 mm</b>
Age group 18-34, n=297	
QFT-/TST-	170
QFT+/TST+	75
QFT-/TST+	37
QFT+/TST-	15
Concordance, %	82.5
Kappa (95%CI)	0.61 (0.52-0.71)
Age group 35-60, n=334	
QFT-/TST-	88
QFT+/TST+	138
QFT-/TST+	79
QFT+/TST-	29
Concordance, %	67.7
Kappa (95%CI)	0.35 (0.26-0.45)

QFT-GIT: QuantiFERON-TB Gold In-Tube ; TST: tuberculin skin test

RESEARCH ARTICLE

# Genome-wide association study of resistance to *Mycobacterium tuberculosis* infection identifies a locus at 10q26.2 in three distinct populations

Jocelyn Quistrebert<sup>1,2</sup>, Marianna Orlova<sup>3,4,5</sup>, Gaspard Kerner<sup>1,2,6</sup>, Le Thi Ton<sup>7</sup>, Nguyễn Trọng Luong<sup>7</sup>, Nguyễn Thanh Danh<sup>7</sup>, Quentin B. Vincent<sup>1,2,11</sup>, Fabienne Jabot-Hanin<sup>8,9,10</sup>, Yoann Seeleuthner<sup>1,2</sup>, Jacinta Bustamante<sup>1,2,11</sup>, Stéphanie Boisson-Dupuis<sup>12</sup>, Nguyen Thu Huong<sup>13</sup>, Nguyen Ngoc Ba<sup>13</sup>, Jean-Laurent Casanova<sup>1,2,12,14</sup>, Christophe Delacourt<sup>15</sup>, Eileen G. Hoal<sup>16</sup>, Alexandre Alcaïs<sup>1,2</sup>, Vu Hong Thai<sup>13</sup>, Lai The Thành<sup>7</sup>, Laurent Abel<sup>1,2,12‡</sup>, Erwin Schurr<sup>3,4,5‡</sup>, Aurélie Cobat<sup>1,2‡\*</sup>



**1** Laboratory of Human Genetics of Infectious Diseases, Necker Branch, INSERM U1163, Paris, France, **2** University of Paris, Imagine Institute, Paris, France, **3** Infectious Diseases and Immunity in Global Health Program, Research Institute of the McGill University Health Centre, Montreal, Canada, **4** McGill International TB Centre, McGill University, Montreal, Canada, **5** Departments of Medicine and Human Genetics, McGill University, Montreal, Canada, **6** Human Evolutionary Genetics Unit, Institut Pasteur, UMR2000, CNRS, Paris, France, **7** Center for Social Disease Control, Binh Duong, Vietnam, **8** Bioinformatics Core Facility, Paris Descartes University, Paris, France, **9** Bioinformatics Platform, INSERM UMR U1163, Imagine Institute, Paris, France, **10** Necker Federative Research Structure, INSERM US24, CNRS UMS3633, Paris, France, **11** Study Center of Immunodeficiencies, Necker Hospital for Sick Children, Paris, France, **12** St Giles Laboratory of Human Genetics of Infectious Diseases, Rockefeller Branch, Rockefeller University, New York City, New York, United States of America, **13** Hospital for Dermato-Venereology, Ho Chi Minh City, Vietnam, **14** Howard Hughes Medical Institute, New York City, New York, United States of America, **15** Paediatric Pulmonology and Allergology Department, Necker Hospital for Sick Children, Paris, France, **16** South African Medical Research Council Centre for Tuberculosis Research, DST-NRF Centre of Excellence for Biomedical Tuberculosis Research, Division of Molecular Biology and Human Genetics, Faculty of Medicine and Health Sciences, Stellenbosch University, Cape Town, South Africa

**OPEN ACCESS**

**Citation:** Quistrebert J, Orlova M, Kerner G, Ton LT, Luong NT, Danh NT, et al. (2021) Genome-wide association study of resistance to *Mycobacterium tuberculosis* infection identifies a locus at 10q26.2 in three distinct populations. *PLoS Genet* 17(3): e1009392. <https://doi.org/10.1371/journal.pgen.1009392>

**Editor:** Cathy Stein, Case Western Reserve University School of Medicine, UNITED STATES

**Received:** October 11, 2020

**Accepted:** February 2, 2021

**Published:** March 4, 2021

**Peer Review History:** PLOS recognizes the benefits of transparency in the peer review process; therefore, we enable the publication of all of the content of peer review and author responses alongside final, published articles. The editorial history of this article is available here: <https://doi.org/10.1371/journal.pgen.1009392>

**Copyright:** © 2021 Quistrebert et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

**Data Availability Statement:** All summary statistics are available in the manuscript or at <https://doi.org/10.5281/zenodo.3942125>. Because

‡ These authors are joint senior authors on this work.  
\* [aurelie.cobat@inserm.fr](mailto:aurelie.cobat@inserm.fr)

## Abstract

The natural history of tuberculosis (TB) is characterized by a large inter-individual outcome variability after exposure to *Mycobacterium tuberculosis*. Specifically, some highly exposed individuals remain resistant to *M. tuberculosis* infection, as inferred by tuberculin skin test (TST) or interferon-gamma release assays (IGRAs). We performed a genome-wide association study of resistance to *M. tuberculosis* infection in an endemic region of Southern Vietnam. We enrolled household contacts (HHC) of pulmonary TB cases and compared subjects who were negative for both TST and IGRA (n = 185) with infected individuals (n = 353) who were either positive for both TST and IGRA or had a diagnosis of TB. We found a genome-wide significant locus on chromosome 10q26.2 with a cluster of variants associated with strong protection against *M. tuberculosis* infection (OR = 0.42, 95%CI 0.35–0.49, P = 3.71 × 10<sup>-8</sup>, for the genotyped variant rs17155120). The locus was replicated in a French multi-ethnic HHC cohort and a familial admixed cohort from a hyper-endemic area of South Africa, with an overall OR for rs17155120 estimated at 0.50 (95%CI 0.45–0.55, P = 1.26 × 10<sup>-9</sup>). The variants are

of the IRB restriction on the data from Vietnam, France and South Africa, individual level data are only available in the context of a tuberculosis research project upon request from the VFSA TB infection Access Committee by contacting the committee chair Pr Stanislas Lyonnet ([stanislas.lyonnet@inserm.fr](mailto:stanislas.lyonnet@inserm.fr)), Director of the Imagine Institute in Paris, France. Stanislas Lyonnet was not involved with the research reported in this article.

**Funding:** This work was supported in part by the French Foundation for Medical Research (FRM) (EQU201903007798; FDM20170637664 to JQ), the Programme Hospitalier de Recherche Clinique (AOR-04-003); the Legs Poix (Chancellerie des Universités de Paris, <https://www.sorbonne.fr/en/> to AA), the European Research Council (ERC-2010-AdG-268777 to LA), the National Institute of Allergy and Infectious Diseases of the National Institutes of Health (R01AI12434), the French National Research Agency (ANR) under the “Investments for the future” program (grant ANR-10-IAHU-01), the Integrative Biology of Emerging Infectious Diseases Laboratory of Excellence (ANR-10-LABX-62-IBEID), and the TBPATGEN project (ANR-14-CE14-0007-01 to LA), the SCOR Corporate Foundation for Science (<https://www.scor.com/fr/la-fondation-dentreprise-scor-pour-la-science> to LA), the Rockefeller University (<https://www.rockefeller.edu/>), the Institut National de la Santé et de la Recherche Médicale (INSERM, <https://www.inserm.fr/>), Université de Paris (<https://u-paris.fr/en/>), the St. Giles Foundation (<http://www.stgilesfoundation.org/>), the Canadian Institutes of Health Research (CIHR, FDN-143332 to ES) and the Sequella/Aeras Global Tuberculosis Foundation (<https://www.aeras.org>, to ES). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

**Competing interests:** The authors have declared that no competing interests exist.

located in intronic regions and upstream of *C10orf90*, a tumor suppressor gene which encodes an ubiquitin ligase activating the transcription factor p53. *In silico* analysis showed that the protective alleles were associated with a decreased expression in monocytes of the nearby gene *ADAM12* which could lead to an enhanced response of Th17 lymphocytes. Our results reveal a novel locus controlling resistance to *M. tuberculosis* infection across different populations.

## Author summary

There is strong epidemiological evidence that a proportion of highly exposed individuals remain resistant to *M. tuberculosis* infection, as shown by a negative result for Tuberculin Skin Test (TST) or IFN- $\gamma$  Release Assays (IGRAs). We performed a genome-wide association study between resistant and infected individuals, which were carefully selected employing a household contact design to maximize exposure by infectious index patients. We employed stringently defined concordant results for both TST and IGRA assays to avoid misclassifications. We discovered a locus at 10q26.2 associated with resistance to *M. tuberculosis* infection in a Vietnamese discovery cohort. This locus could be replicated in two independent cohorts from different epidemiological settings and of diverse ancestries enrolled in France and South Africa.

## Introduction

Tuberculosis (TB) remains a major public health threat worldwide [1]. An estimated 10 million people developed TB disease in 2018, of whom 1.45 million died. The causative agent of TB is *Mycobacterium tuberculosis* which is transmitted by aerosol from contagious TB patients. However, not all persons encountering infectious aerosols will become infected with *M. tuberculosis*, defining the first line of human resistance against TB [2,3]. Infection is inferred from the presence of anti-mycobacterial immunoreactivity, as shown by a positive result in tuberculin skin test (TST) and/or interferon-gamma (IFN- $\gamma$ ) release assay (IGRA). TST is done *in vivo* and consists of an intradermal injection of purified protein derivative (PPD) that provokes a delayed hypersensitivity reaction at the site of injection. IGRAs are performed *ex vivo* and measure the secretion of IFN- $\gamma$  by leukocytes in response to *M. tuberculosis*-specific antigens. Both tests have their own limitations and results are not fully concordant [4–6]. Individuals who score positive by TST and/or IGRA are considered to suffer from asymptomatic latent TB infection (LTBI). Conversely, persons who score negative despite documented exposure to *M. tuberculosis* are considered resistant to infection. Based on TST and/or IGRA results, the intensity of exposure or the duration of follow-up, from 7% to 25% of subjects display the *M. tuberculosis* infection resistance phenotype [2,7–9].

The large inter-individual variability in exposure outcomes supports a major role for human genetic factors [10]. Various genome-wide approaches have confirmed this hypothesis by either considering TST and IGRA results as quantitative traits or relying on TST reactivity (positive/negative) as a surrogate marker for infection [11–14]. Regarding this latter phenotype, persistent TST negativity was linked to loci on 2q21–2q24, further fine-mapped to *ZEB2*, and 5p13–5q22 in an Ugandan population [14,15]. A major locus, named *TST1* on chromosome 11p14, employing stringently defined TST negativity (0 mm vs. > 0 mm) as phenotype, was identified in a linkage analysis conducted in South Africa [11]. *TST1* was later replicated in a

household contact (HHC) study of French families [16]. A genome-wide association study (GWAS) among highly *M. tuberculosis* exposed HIV-seropositive individuals from East Africa, identified a locus in the 5q31.1 region near *IL9* associated with negative TST [17].

Here, we performed a GWAS of resistance to *M. tuberculosis* infection using a robust phenotype based on both TST and IGRA information. In addition, we used a HHC study design guaranteeing shared environmental effects and high intensity of exposure to *M. tuberculosis*. We found a locus on chromosome 10q26.2 associated with resistance to *M. tuberculosis* infection in an East Asian population from Southern Vietnam. Importantly, this locus was replicated in two other cohorts from France and South Africa, representing different ancestries and epidemiological settings.

## Results

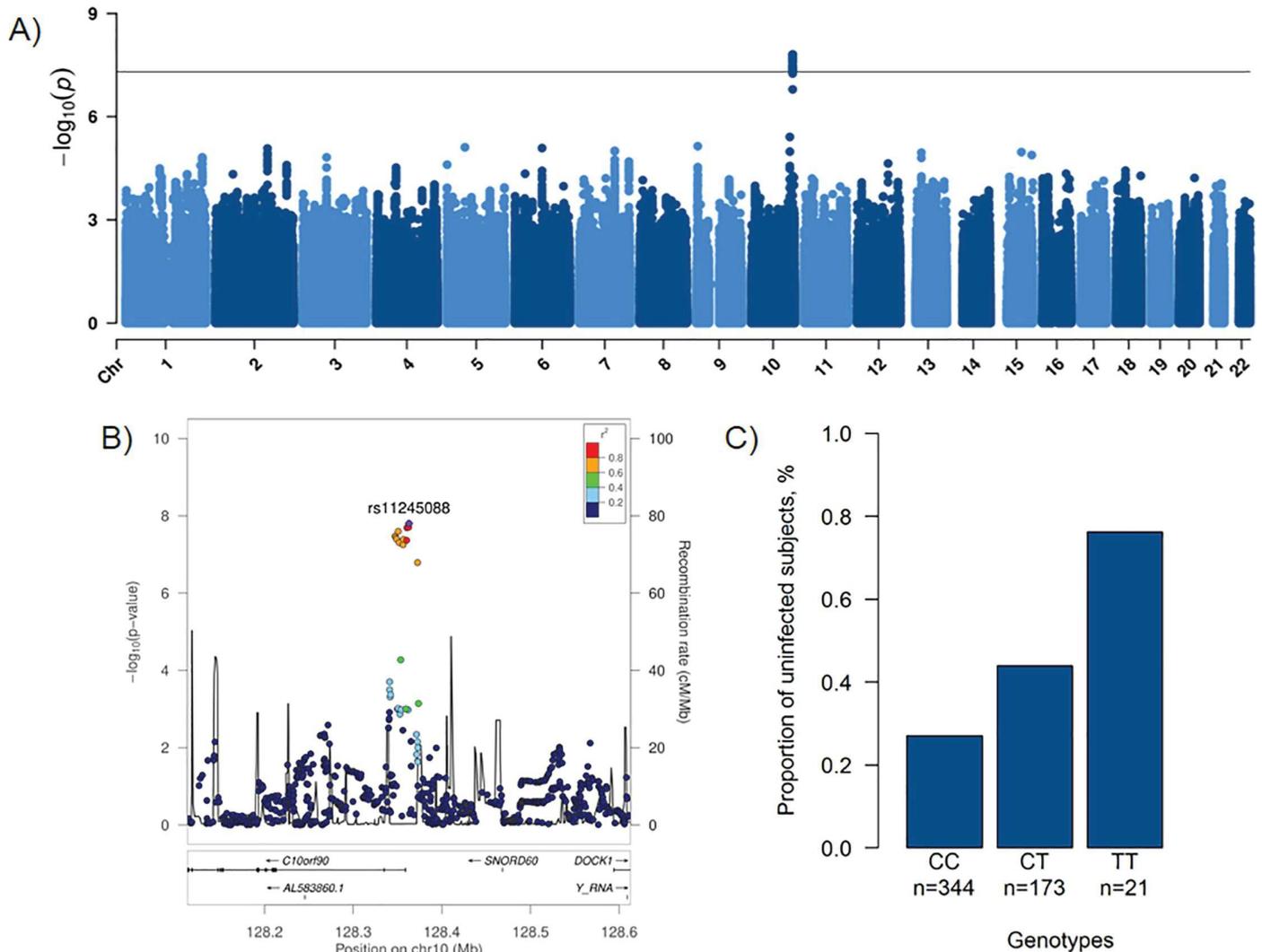
### Genome-wide association study of resistance to *M. tuberculosis* infection in Vietnam

First, we conducted a GWAS in the Vietnamese sample using 185 uninfected and 353 infected subjects, consisting of 201 individuals with positive TST/Quantiferon-TB Gold In-Tube test (QFT-GIT) results and 152 pulmonary TB (PTB) patients. A total of 5,591,951 high quality variants were tested with a genomic inflation factor ( $\lambda$ ) at 0.997, suggesting that effects from the familial study design were well controlled (S1 Fig). The corresponding Manhattan plot is shown in Fig 1A. We observed a genome-wide significant association on chromosome 10q26.2, corresponding to a cluster of 12 variants and 6 additional variants in high linkage disequilibrium (LD) with  $P < 5 \times 10^{-7}$  in the intronic regions and upstream of *C10orf90* (or *FATS*, HGNC: 26563) (Fig 1B). The top-associated variant was the imputed rs11245088 (odds ratio (OR) = 0.42, 95% confidence interval (CI) 0.39–0.45,  $P = 1.58 \times 10^{-8}$ ) while the top-associated genotyped variant was rs17155120 ( $P = 3.71 \times 10^{-8}$ ) (Table 1). Each copy of the minor allele T of rs17155120 conferred protection against *M. tuberculosis* infection with an OR of being infected for CT vs. CC or TT vs. CT at 0.42 (95%CI 0.35–0.49) (Fig 1C). The intensity cluster plot for rs17155120 showed that the genotype calling was of high quality and separated clearly into 3 genotype groups (S2 Fig). Since all 18 variants in the locus were in high LD (S3 Fig), the imputed variants were likely to have a high imputation quality as suggested by their info score (S1 Table).

We also performed a GWAS between the 185 uninfected and the 201 infected subjects, excluding the 152 PTB patients (S4 Fig). Despite a smaller sample size, all the 18 variants of the locus were still associated with protection against infection with  $P < 5.0 \times 10^{-6}$ , with similar ORs (for rs17155120, OR = 0.40, 95%CI 0.32–0.49,  $P = 2.55 \times 10^{-7}$ ) (S2 Table). Similar findings were also observed when considering only the 152 PTB patients as infection reference, with an OR for rs17155120 estimated at 0.50 (95%CI 0.41–0.59,  $P = 2.10 \times 10^{-4}$ ). These results indicate that PTB patients are an appropriate infection reference group in this analysis.

### Replication of variants associated with resistance to *M. tuberculosis* infection in France and South Africa

We tested the effects of the variants of the 10q26.2 locus in a French multi-ethnic HHC cohort (30 uninfected vs. 157 infected subjects) and an admixed familial sample from South Africa (118 uninfected vs. 136 infected subjects). In the French cohort, 17 variants of the cluster could be genotyped or imputed (including 11 genome-wide significant variants), and 3 were replicated at the  $P < 0.025$  level with effect sizes in the same direction as in the Vietnamese cohort. The most significant associated variant was rs56106518 (OR = 0.40, 95%CI 0.30–0.51,



**Fig 1. Genome-wide association study of resistance to *M. tuberculosis* infection in Vietnam.** A) Manhattan plot showing results from a genome-wide association study between 185 uninfected subjects (negative for both tuberculin skin test and QuantiFERON-TB Gold In-Tube test) and 353 infected subjects (201 infected individuals positive for both tests and 152 patients with a history of pulmonary tuberculosis) for 5,591,951 variants (minor allele frequency > 5% and info > 0.8) with an unadjusted additive genetic model. The  $-\log_{10}(P)$  value for each variant (y-axis) is presented according to its chromosomal position (x-axis, build hg19). The dashed line indicates the genome-wide significant threshold at  $P = 5 \times 10^{-8}$ . B) Locus zoom plot showing association for the 10q26.2 locus, in a 500 kb window surrounding the top imputed variant rs11245088 (purple diamond). Colors represent pairwise linkage disequilibrium ( $r^2$ ) with rs11245088 as calculated for the Vietnamese Kinh population of 1000 Genomes phase 3. C) Proportion of Vietnamese individuals resistant to *M. tuberculosis* infection by genotype for the variant rs17155120. Each bar represents the proportion of uninfected subjects among CC individuals ( $n = 93/344$ ), CT individuals ( $n = 76/173$ ) and TT individuals ( $n = 16/21$ ) for the variant rs17155120 in Vietnam.

<https://doi.org/10.1371/journal.pgen.1009392.g001>

$P = 2.98 \times 10^{-3}$ ) (Table 1). In the South African population, 12 variants of the locus were genotyped or successfully imputed (including 10 genome-wide significant variants) and 3 of them showed evidence for replication. The most significant associated variant was rs118037357 (OR = 0.55, 95%CI 0.43–0.67,  $P = 7.38 \times 10^{-3}$ ) (Table 1). The top genotyped variant rs17155120 was also replicated in the two cohorts ( $P_{France} = 1.51 \times 10^{-2}$  and  $P_{SouthAfrica} = 1.74 \times 10^{-2}$ ) (Table 1 and Fig 2A).

Interestingly, the variants across the two replication cohorts were in high LD with each other presenting similar LD patterns as in Vietnam (S5 Fig). The frequencies of the

**Table 1. Association between an additive genetic effect of variants on chromosome region 10q26.2 and resistance to *M. tuberculosis* infection in Vietnam, France and South Africa.**

Variant	<i>C10orf90</i>	EA	Vietnam			France			South Africa		
			EAF	OR (95%CI)	P value	EAF	OR (95%CI)	P value	EAF	OR (95%CI)	P value
rs11245088	upstream	C	0.25	0.42 (0.39–0.45)	$1.58 \times 10^{-8}$	-	-	-	0.44	0.99 (0.91–1.07)	$4.75 \times 10^{-1}$
rs72163291	intron	ins	0.24	0.42 (0.35–0.49)	$1.94 \times 10^{-8}$	0.37	0.88 (0.78–0.98)	$3.54 \times 10^{-1}$	-	-	-
rs17155143	upstream	A	0.20	0.39 (0.31–0.47)	$1.98 \times 10^{-8}$	0.18	0.63 (0.50–0.77)	$1.29 \times 10^{-1}$	0.15	0.60 (0.48–0.72)	$1.72 \times 10^{-2}$
rs7909756	upstream	G	0.24	0.42 (0.35–0.49)	$2.03 \times 10^{-8}$	0.37	0.83 (0.73–0.93)	$2.87 \times 10^{-1}$	0.36	0.90 (0.81–0.99)	$2.77 \times 10^{-1}$
rs28703703	intron	G	0.20	0.41 (0.34–0.49)	$2.52 \times 10^{-8}$	0.20	0.59 (0.48–0.69)	$4.95 \times 10^{-2}$	0.17	0.66 (0.55–0.77)	$2.95 \times 10^{-2}$
rs56106518	intron	C	0.23	0.41 (0.33–0.48)	$3.07 \times 10^{-8}$	0.30	0.40 (0.30–0.51)	$2.98 \times 10^{-3}$	-	-	-
rs75482972	intron	A	0.20	0.42 (0.34–0.49)	$3.35 \times 10^{-8}$	0.20	0.58 (0.47–0.68)	$4.38 \times 10^{-2}$	0.17	0.68 (0.57–0.78)	$3.48 \times 10^{-2}$
rs17155120	intron	T	0.20	0.42 (0.35–0.49)	$3.71 \times 10^{-8}$	0.18	0.48 (0.36–0.59)	$1.51 \times 10^{-2}$	0.16	0.62 (0.52–0.73)	$1.74 \times 10^{-2}$
rs73370887	intron	A	0.20	0.40 (0.33–0.48)	$4.05 \times 10^{-8}$	0.31	0.78 (0.71–0.85)	$1.43 \times 10^{-1}$	0.28	0.75 (0.67–0.83)	$3.69 \times 10^{-2}$
rs79608098	intron	T	0.20	0.42 (0.35–0.49)	$4.06 \times 10^{-8}$	0.20	0.58 (0.48–0.69)	$4.50 \times 10^{-2}$	0.17	0.66 (0.55–0.76)	$2.71 \times 10^{-2}$
rs61750007	upstream	C	0.24	0.43 (0.36–0.50)	$4.30 \times 10^{-8}$	0.24	0.53 (0.42–0.64)	$3.49 \times 10^{-2}$	0.20	0.73 (0.62–0.84)	$7.62 \times 10^{-2}$
rs77513326	intron	A	0.20	0.42 (0.34–0.49)	$4.93 \times 10^{-8}$	0.17	0.47 (0.36–0.59)	$1.61 \times 10^{-2}$	0.16	0.63 (0.52–0.75)	$2.35 \times 10^{-2}$
rs79918233	intron	A	0.20	0.41 (0.33–0.49)	$5.61 \times 10^{-8}$	0.17	0.51 (0.39–0.63)	$3.00 \times 10^{-2}$	0.16	0.63 (0.52–0.74)	$1.95 \times 10^{-2}$
rs147584264	upstream	C	0.19	0.41 (0.33–0.49)	$1.21 \times 10^{-7}$	0.22	0.77 (0.62–0.93)	$3.00 \times 10^{-1}$	-	-	-
rs191820708	upstream	A	0.19	0.41 (0.33–0.49)	$1.31 \times 10^{-7}$	0.19	0.63 (0.50–0.77)	$1.40 \times 10^{-1}$	-	-	-
rs201178890	upstream	T	0.19	0.41 (0.33–0.49)	$1.36 \times 10^{-7}$	0.22	0.79 (0.62–0.97)	$3.36 \times 10^{-1}$	-	-	-
rs202189321	upstream	T	0.20	0.41 (0.33–0.49)	$1.37 \times 10^{-7}$	0.18	0.65 (0.52–0.78)	$1.44 \times 10^{-1}$	-	-	-
rs118037357	upstream	A	0.19	0.41 (0.33–0.49)	$1.62 \times 10^{-7}$	0.15	0.47 (0.33–0.60)	$2.83 \times 10^{-2}$	0.14	0.55 (0.43–0.67)	$7.38 \times 10^{-3}$

CI, confidence intervals; EA, effect allele; EAF, effect allele frequency; OR, odds ratio; ins, insertion

<https://doi.org/10.1371/journal.pgen.1009392.t001>

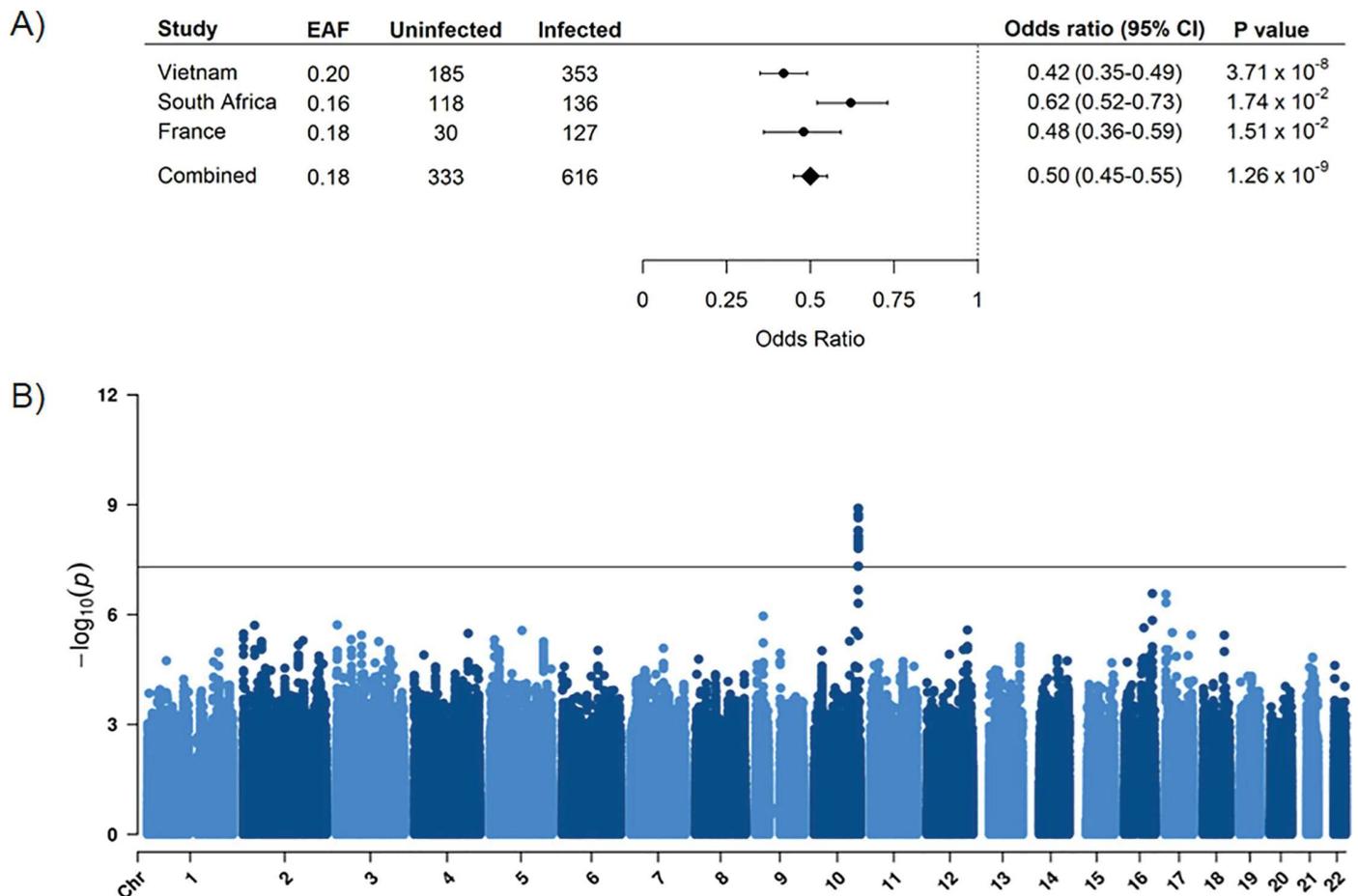
rs17155120 T allele for all 3 cohorts were also similar, ranging from 0.16 to 0.20. The rs17155120 T allele frequency of 0.20 in the Vietnamese cohort was consistent with the Kinh allele frequency of 0.18 in 1000 Genomes (1000G) phase 3 (S6 Fig [18], and S3 Table). The rs17155120 T allele frequency of 0.18 in the multi-ethnic French cohort was also close to the global frequency of 0.16 in all 1000G populations. However, the frequency of 0.16 for rs17155120 T allele in the South African cohort was higher than in any 1000G African population that ranged from 0.05 to 0.11. Discrepancy of allele frequencies between South African subjects and African populations of 1000G was confirmed by different patterns of LD in a 30 kb region around rs17155120 (S7 Fig), which could be explained by the specific ethnic origin of the South African subjects.

### Trans-ethnic meta-analysis

Next, we performed a meta-analysis of resistance to *M. tuberculosis* infection using the GWAS data of the 3 cohorts. The combined analysis was carried out in 333 uninfected and 616 infected subjects. The only genome-wide significant result was observed with the variants of the 10q26.2 region (Figs 2B and S8). The most significant signal was observed at the genotyped variant rs17155120 (OR = 0.50, 95%CI = 0.45–0.55,  $P = 1.26 \times 10^{-9}$ ) and no heterogeneity was observed across the 3 studies ( $P_{\text{het}} = 0.31$ ) (Fig 2A and S4 Table).

### Functional annotation

The variants of the 10q26.2 region map to the intronic regions and upstream of the tumor suppressor gene *C10orf90* (or *FATS*, for Fragile-site Associated Tumor Suppressor, HGNC: 26563) (Fig 3). According to ENCODE, two associated intronic variants, rs28703703 and



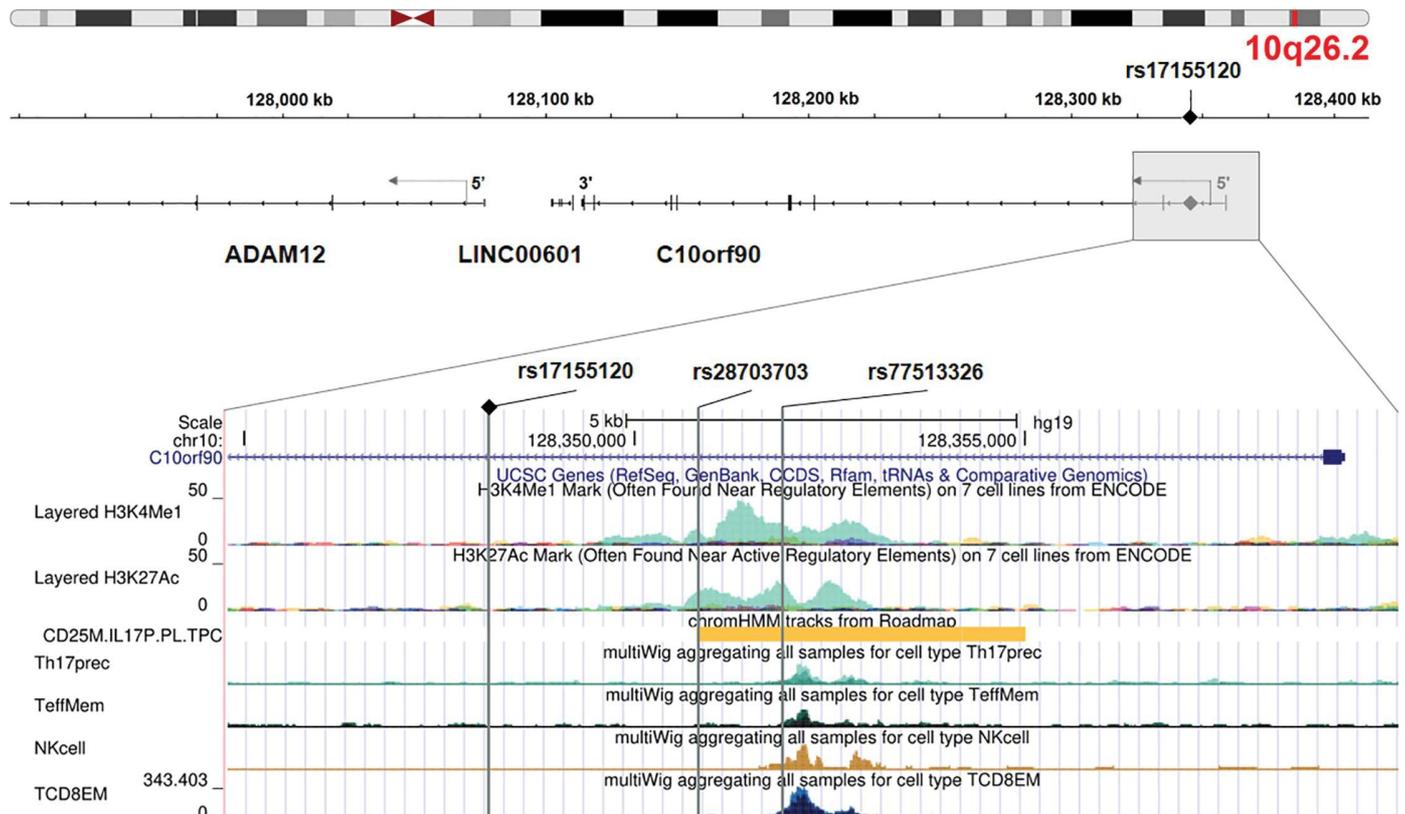
**Fig 2. Genome-wide association study of resistance to *M. tuberculosis* infection in 3 cohorts from Vietnam, France and South Africa.** A) Forest plot of the association between an additive genetic effect of rs17155120 on chromosome region 10q26.2 and resistance to *M. tuberculosis* infection. Odds ratios and 95% confidence intervals derived from a linear mixed model, *P* values, sample sizes and frequency of the effect allele (EAF) are reported by individual cohort and for the random effects meta-analysis. B) Manhattan plot showing results from a genome-wide association study between 333 uninfected subjects and 616 infected subjects for 3,967,482 variants (minor allele frequency > 5% and *info* > 0.8) with an unadjusted additive genetic model. The  $-\log_{10}(P)$  value for each variant (y-axis) is presented according to its chromosomal position (x-axis, build hg19). The dashed line indicates the genome-wide significant threshold at  $P = 5 \times 10^{-8}$ .

<https://doi.org/10.1371/journal.pgen.1009392.g002>

rs77513326, are located in a regulatory genomic region characterized by H3K4me1 and H3K27ac histone marks and an active enhancer signature in lymphocytes T helper 17 (Th17). The variant rs77513326 also overlaps ATAC peaks in Th17 cells, memory T cells, natural killer cells and CD8<sup>+</sup> T cells. We further explored the variants in various expression quantitative trait loci (eQTLs) databases of relevant tissues for the phenotype and found an association between them and expression of the nearby gene *ADAM12*. In particular, rs28703703 and the genotyped variant rs17155120 displayed decreasing expression of *ADAM12* with each minor allele (having a protective effect against *M. tuberculosis* infection) in monocytes ( $P = 2.10 \times 10^{-3}$  and  $P = 4.70 \times 10^{-3}$  respectively) (from <https://immunopop.com/kim/eQTL> [19]). No association was observed in other immune cell types.

## Discussion

In this study, we explored the genetic determinants of natural resistance to *M. tuberculosis* infection after intense exposure. There are no direct tests for established infection because TST



**Fig 3. Genomic annotation of the locus on chromosome region 10q26.2.** The upper panel is adapted from Integrative Genomics Viewer (<http://igv.org/app/>) and the lower panel is adapted from UCSC Genome Browser (<http://genome.ucsc.edu/>). The 3 vertical grey lines represent the associated variants rs17155120, rs28703703 and rs77513326 (from left to right) that overlap regulatory regions. From top to bottom: H3K4me1 and H3K27ac histone marks from ENCODE, active enhancer in Th17 cells (chromHMM annotation from ROADMAP), chromatin accessibility as represented by ATAC peaks in Th17 cells, memory T cells, NK cells and CD8<sup>+</sup> T cells [20].

<https://doi.org/10.1371/journal.pgen.1009392.g003>

and IGRAs measure an immune response that does not allow to distinguish between past or present infection with *M. tuberculosis* bacilli. Nevertheless, our results show a strong genetic effect on resistance to *M. tuberculosis* infection irrespective of infection being persistent or temporary. TST and IGRA tests have well-known limits [4]. To minimize misclassification of uninfected and infected subjects in a cross-sectional setting, we therefore relied on negative results in both tests and defined stringent cut-offs. As exposure to *M. tuberculosis* is difficult to quantitate, yet a critical feature of TB studies [21], we focused on individuals at high risk of infection. We recruited household contacts recently exposed for extended periods to a contagious PTB index, some of whom remained infection-free. We discovered a genome-wide significant association between a cluster of variants on chromosome 10q26.2 and resistance to *M. tuberculosis* infection in well-characterized Vietnamese subjects. Strikingly, this locus could be replicated in two independent cohorts with different epidemiological settings from France and South Africa, resulting in association of the C/T variant rs17155120 across the 3 populations at an estimated combined odds ratio of 0.50 (95%CI 0.45–0.55) for becoming infected for TC vs. CC or TT vs. TC individuals.

The cluster of associated variants overlaps intronic and 5' regions of *C10orf90* (or *FATS*, HGNC: 26563), a tumor suppressor gene. This gene lies within a common-fragile site, which is an evolutionarily conserved region among mammals and susceptible to DNA damage [22]. The protein encoded by *C10orf90* has been shown to promote p53 activation in response to

DNA damage through an E3 ubiquitin ligase activity [23,24]. Several E3 ubiquitin ligases have already been shown to participate in the defense against *M. tuberculosis*, in particular through autophagy [25–27]. The p53 transcription factor is not only a master regulator of autophagy but also activates apoptosis, another key process for host infected cells to limit the spread of pathogens as *M. tuberculosis* [28,29]. Recent studies demonstrated that p53-induced apoptosis plays a critical role in the inhibition of mycobacteria survival and the macrophage resistance, possibly mediated by IL-17 [30,31]. However, possible role of *C10orf90* during *M. tuberculosis* infection or any related immune process remains unknown.

*In silico* functional annotation of the 10q26.2 associated variants revealed additional regulatory features related to lymphocytes Th17. In particular, the A/G variant rs28703703 was identified as a likely cis-eQTL for *ADAM12* in monocytes, with decreased expression of *ADAM12* associated with the G allele, protective against *M. tuberculosis* infection. *ADAM12* is located ~280 kb downstream *C10orf90*, and encodes a matrix metalloprotease linked to a broad range of biological processes [32]. *ADAM12* expression correlates with lung inflammation as it is overexpressed in cells issued from asthmatic sputum and in the airway epithelium during allergic inflammatory reaction [33,34]. Previous studies also reported *ADAM12* expression in Th17 cells [35,36], and *ADAM12* knockdown in human T cells was found to increase Th17 cytokine production (IL-17A, IL-17F, and IL-22) [35]. In South India, TST-negative individuals produced significantly higher levels of Th17 cytokines than TST-positive individuals [37]. Similarly, significant higher levels of Th17 cytokines were observed in persistent negative IGRA individuals as compared with IGRA converters in a recent study conducted in the Gambia [38]. These findings are consistent with our present results showing that the rs28703703 G allele protective against *M. tuberculosis* infection is associated with lower *ADAM12* expression that could lead to higher Th17 cytokine production. Overall, these observations support the view that Th17 cytokines may have a protective role against early stages of *M. tuberculosis* infection.

The 3 enrolled samples were of modest size. However, the design of the 3 studies, and the definition of a stringent and homogenous phenotype across them enabled us to detect a significant association with large effects. Interestingly, adding the PTB Vietnamese patients, who are by definition infected, increased the power of the analysis. Strikingly, the top associated variants displayed large effects with similar allele frequencies across populations and were in high LD across multiple ancestries. This observation was not expected because of the various epidemiological settings and the genetic diversity of the cohorts that included Kinh Vietnamese, Europeans, North Africans, Sub-Saharan Africans and admixed Western Cape Coloureds. Such populations are usually under-represented in studies of genetic association while they could provide valuable insights to understand complex diseases [39].

In conclusion, we demonstrated that rigorous epidemiological design and phenotype definition with seemingly limited sample sizes can reveal novel genetic factors that offer protection from major pathogens such as *M. tuberculosis*. We found that *C10orf90* and *ADAM12* are promising candidate genes involved in the natural resistance to *M. tuberculosis*. Further investigations are needed to elucidate their role in the process of the initial infection, which could be a major step to provide new opportunities in the fight against TB.

## Subjects and methods

### Ethics statement

Signed informed consent was obtained from all the participants, and from the parents of enrolled minors. The study was approved by the regulatory authorities in Binh Duong, Vietnam (1366/UBND-VX), the French Consultative Committee for Protecting Persons in

Biomedical Research of Henri Mondor Hospital (Créteil, France) and the Stellenbosch University Health Research Ethics Committee (Tygerberg, South Africa).

## Study settings and populations

**Vietnam.** Vietnam is a middle-income country in South-East Asia with a high annual incidence of TB (130/100,000 at the time of the study [40]), high Bacille Calmette-Guérin (BCG) vaccination coverage (with reported rates exceeding 95% [41]) and a low population prevalence of HIV (0.4% in 2015 in the general population and 5% among TB cases [40,42]). From 2010 to 2015, we recruited, in a *M. tuberculosis* endemic region of Southern Vietnam, index PTB adults with persistent cough before the start of treatment (mean duration of coughing = 3.2 months). PTB diagnosis was assessed by clinical presentation, chest X-ray and/or positive cultures for *M. tuberculosis*. A total of 1108 HHCs of 466 PTB index cases were invited to participate in the study and underwent both TST and QFT-GIT (S1 Text). HHCs did not undergo any further follow-up.

**France.** As a first replication cohort, we used household TB contacts studied in Val-de-Marne, a suburban region of Paris, in the context of a general screening procedure. This multi-ethnic cohort has been previously described [12,16]. Val-de-Marne is an area of low TB endemicity, displaying an annual incidence of 22/100,000 at the time of the study [43] and BCG vaccination rates are high [41]. HIV seroprevalence is low in France in the general population (0.3% at the time of the study [44]) and was estimated at 7% in the index cases of the study [45]. For this study, 664 HHCs of 132 PTB index cases were investigated according to national guidelines that required two screening visits. HHCs were individuals sharing residence with an index during the 3 months before diagnosis. Briefly, the first visit (V1) included a physical examination, a chest radiograph, TST and in-house IGRA [12,16] (S1 Text). These investigations, except for IGRA, were repeated 8–12 weeks later (V2) if the contact subject did not meet the criteria for infection at V1.

**South Africa.** As a second replication cohort, we used a large sample ( $n = 415$ ) of 153 nuclear families from a suburban area of Cape Town, South Africa, which has been previously described [6,11,12,16]. All individuals belonged to the South African Coloured group, a unique multi-way admixed population [46]. There was no specific requirement for subjects to be HHCs of PTB patients. Indeed, TB is hyper-endemic in this area with an incidence of ~800/100,000 at the time of the study [47] and TB transmission occurs more often outside the household [48]. BCG vaccination at birth is routine in this area [49]. HIV seroprevalence was estimated at 5.2% in the overall population and less than 2% in the pediatric population at the time of the study [11]. In addition, individuals who were known to be HIV positive, pregnant, or using immunomodulatory chemotherapy were excluded at the time of enrollment [11]. TST and in-house IGRA were performed as previously described (S1 Text). Subjects who had clinical TB disease in the two years preceding the study were excluded. Thus, only healthy children and young adults from the area were included and tested for infection.

## Genotyping, quality control and imputation

A total of 724 individuals from Vietnam (S9 Fig) and 573 from France were genotyped using the Illumina Infinium OmniExpressExome-8-v1 chip (960,212 single nucleotide polymorphisms, SNPs). For the South African cohort ( $n = 374$ ), the Illumina HumanOmni2.5–8 Bead-Chip (~2 million of SNPs) was used. All quality control steps were done in each cohort with PLINK v1.9 [50]. Autosomal SNPs with a minor allele frequency (MAF)  $> 0.01$ , a genotype call rate  $> 0.99$  and a Hardy-Weinberg (HWE) equilibrium  $P > 1.00 \times 10^{-5}$  were retained. Individuals with a call rate  $< 95\%$  were excluded ( $n = 2$ ). Identity-by-descent (IBD) analysis

was done to detect duplicated individuals and the members of the pairs with the lower call rate were excluded ( $n = 1$ ). After the quality control, imputation was performed on 720 individuals and 598,090 variants from Vietnam and 573 individuals and 886,471 variants from France using the Michigan Imputation Server [51] with Eagle2 [52] for the pre-phasing and the 1000G Project as reference panel [53]. For the South African cohort (374 individuals and 1,347,846 variants), the imputation was done on the Sanger Imputation Server with Eagle2 for the pre-phasing and the African Genome Resources as reference panel which includes ~2000 African samples in addition to the individuals from 1000G [54]. Imputed SNPs with an imputation quality info score  $> 0.8$  and MAF  $> 0.05$  were retained for further association analyses (5,591,951 variants in Vietnam, 7,737,070 variants in France and 6,922,541 variants in South Africa).

For each cohort, principal component analysis (PCA) was conducted to evaluate population structure. Genotypes of the individuals from 1000G were used to calculate principal components and data for subjects from the cohorts were projected onto the eigenvectors. The Vietnamese cohort, which was sampled from the Vietnamese Kinh group, was homogenous and clustered with the 1000G East Asian populations (S10 Fig). By contrast, the families in the Val-de-Marne sample showed genetic diversity at the population level with a majority of individuals of European, North African and Sub-Saharan ancestries (S11 Fig). The admixed South African subjects, who exhibited genetic diversity at the individual level, were forming a distinct cluster close to the African populations of 1000G (S12 Fig).

### Definition of the *M. tuberculosis* infection phenotype

The definition of the *M. tuberculosis* infection phenotype relied on both TST and IGRA results. In particular, we used a 5 mm cut-off to determine TST status, based on previous studies in similar settings [7,17] and published guidelines [55]. We explored covariates associated with our infection definition in the entire cohort of enrolled individuals and the subset of those with genotype information were retained for the GWAS (S9 Fig).

For the Vietnamese study, resistance to *M. tuberculosis* infection was defined by the presence of a negative TST  $< 5$ mm and a negative QFT-GIT test result following a protocol suggested by the manufacturer (S1 Text). Infected individuals were defined as subjects presenting both a positive TST  $\geq 5$ mm and a positive QFT-GIT test result. A total of 188 subjects were classified as double negative and 512 as double positive (S13 Fig), among which 185 and 201 subjects were genotyped, respectively. In order to increase the sample size, we also added 152 genotyped PTB patients, consisting of 146 index cases and 6 subjects with a history of PTB, to the infected group (S5 Table). We investigated covariates associated with our infection definition and no significant association was found (S6 Table). Therefore, we conducted an unadjusted genetic association analysis of our binary infection phenotype.

For the French study, which included only HHCs and none of their PTB index cases, contacts could have had one or two screening visits (V1 and V2) with a TST measurement (S14 Fig). A TST was considered negative when the skin induration was (i)  $< 5$  mm at both V1 and V2, (ii)  $< 5$  mm at V1, when only one visit was done. A TST was considered positive when the skin induration was (i)  $\geq 5$  mm at both V1 and V2, (ii)  $< 5$  mm at V1 and  $\geq 10$  mm at V2, which reflected true conversions. In-house IGRA was used in this study and provided quantitative levels of IFN- $\gamma$  production upon early secretory antigenic target 6 (ESAT6) stimulation (S1 Text). A negative IGRA result was defined by a null production of IFN- $\gamma$ . To determine the optimal positivity cut-off, we built a receiver operating characteristic (ROC) curve with TST status as the observed outcome and the corrected IFN- $\gamma$  levels (ESAT6 response minus non-stimulated control value) as the predicted outcome among all the contacts enrolled. We

selected as positivity threshold the highest sum of sensitivity plus specificity, which was equal to 175 pg/mL (S14 Fig). Then, we defined uninfected subjects as HHCs with a negative TST and a null IFN- $\gamma$  production ( $n = 33$ ) and infected subjects as HHCs who presented both a positive TST and a positive IGRA result (IFN- $\gamma$  production  $> 175$  pg/mL) ( $n = 147$ ) (S14 Fig). We also looked for covariates associated with our infection definition in this sample of 180 individuals (S7 Table). Age was the only factor significantly associated with our infection definition and was included as covariate in the genetic association analysis that finally included 30 genotyped uninfected and 127 infected subjects (S8 Table).

For the South African study, a 5 mm cut-off was used to identify negative and positive TSTs (S15 Fig). In-house IGRA was also based on the production of IFN- $\gamma$  upon ESAT6 stimulation (S1 Text). We determined the optimal IGRA positivity cut-off similarly to the French study by building a ROC curve, leading to a threshold of 20.9 pg/mL (S15 Fig). Then, we defined uninfected subjects as HHCs with a negative TST ( $< 5$  mm) and a null IFN- $\gamma$  production ( $n = 128$ ), and infected subjects as those with both positive TST and IGRA result (IFN- $\gamma$  production  $> 20.9$  pg/mL) ( $n = 152$ ) (S15 Fig). Age was the only factor significantly associated with our infection definition in the whole cohort (S9 Table) and was included as covariate in the genetic association analysis that finally included 118 genotyped uninfected and 136 infected subjects (S10 Table).

Covariates associated with our infection definition were investigated using mixed-effects logistic regression with a random effect per family in each cohort. All the analyses were carried out using R software (version 3.5.2) and related packages “pROC” and “lme4” [56–58].

### Genetic association analyses

We conducted genetic association analyses of uninfected vs. infected subjects in the 3 cohorts using a linear mixed-model (LMM) assuming an additive genetic model as implemented in GEMMA v0.98 [59]. To account for the familial relationships, a genetic relatedness matrix (GRM) was used as random effects. In each cohort, the GRM was estimated using centered genotypes after the quality control described above.  $P$  values from the likelihood ratio test were reported. For better interpretability, we reported odds ratios (OR) and their 95% confidence intervals (95%CI) after transforming the regression coefficients of the LMM [60]. Manhattan plots of the  $-\log_{10}(P$  values) and quantile-quantile (QQ) plots were generated using “CMplot” package in R [61]. Regional plots were generated using LocusZoom Standalone v1.4 [62]. Haplotype plots were generated using Haploview [63]. Replication of genome-wide associated variants ( $P < 5 \times 10^{-8}$ ) in the primary cohort from Vietnam was assessed in the two cohorts from France and South Africa. The observed LD in France and South Africa was slightly weaker as compared with Vietnam, and two LD blocks were inferred by Haploview (S5 Fig). We therefore considered variants at a nominal one sided  $P$  value  $< 0.025$  and with a consistent direction of the effect size as replicated.

We also conducted a trans-ethnic meta-analysis by using summary statistics (i.e. beta estimates and their standard errors) from the Vietnamese discovery cohort and the two replication datasets. We used the random-effects model of Han and Eskin implemented in METASOFT [64]. This model assumes effect sizes of exactly zero in all the studies (i.e. no heterogeneity) under the null hypothesis of no associations and allows the effect sizes to vary among studies (i.e. heterogeneity) under the alternative hypothesis. The effect size consistency across studies were determined using the Cochran’s Q statistic. Allelic effect estimates were also derived on the log-odds scale.

### Functional annotation

We used the UCSC Genome Browser [65] to identify associated variants which may overlap with known regulatory regions: 1) histone marks from the ENCODE project [66], 2)

chromatin state annotated by ChromHMM on the basis of ROADMAP [67,68] and 3) chromatin accessibility determined by assay for transposase-accessible chromatin using sequencing (ATAC-seq) from immune cell-types [20]. We also looked at the associated variants in eQTL databases which focused on gene expression in monocytes [19,69], T cells [70], macrophages [71], and various types of immune cells [72].

## Supporting information

**S1 Text. Supplemental Methods.** Immune assays in Vietnam, France and South Africa. (PDF)

**S1 Fig. Quantile-quantile plot of GWAS resistance to *M. tuberculosis* infection in Vietnam.** Plot of the expected distribution of association test statistics (x axis) for 5,591,951 variants compared to the observed values (y axis). Red line is the null hypothesis of no association ( $y = x$ ). (PDF)

**S2 Fig. Intensity cluster plot for rs17155120 of the Illumina Infinium OmniExpressExome-8-v1 chip in the 720 Vietnamese individuals who were genotyped.** (PDF)

**S3 Fig. Haploview LD graph of the top associations on chromosome 10q26.2 in Vietnam.** Pairwise LD coefficients  $r^2$  are shown in each cell between 18 variants with  $P < 5.0 \times 10^{-6}$ . Top genotyped variant rs17155120 is second from left. (PDF)

**S4 Fig. Manhattan plot for the GWAS of resistance to *M. tuberculosis* infection in Vietnam (185 uninfected vs 201 infected participants).** (PDF)

**S5 Fig. Haploview LD graphs of the locus on chromosome 10q26.2 in replication cohorts.** Pairwise LD coefficients  $r^2$  are shown in each cell A) between the 17 variants of the locus in the French cohort, and B) between 12 out of the 18 variants of the locus in the South African cohort. Top genotyped variant rs17155120 is second from left in each figure. (PDF)

**S6 Fig. Geographic distribution of the variant rs17155120 in 1000G phase 3 populations.** The T allele (in blue) is protective against *M. tuberculosis* infection. Map generated from the Geography of Genetic Variants Browser (<http://www.popgen.uchicago.edu/ggv>). (PDF)

**S7 Fig. Pairwise linkage disequilibrium ( $r^2$ ) in A) Vietnamese, B) French and C) South African study cohorts (left panels) between 60 variants in a 30 kb window around rs17155120 (arrow) as compared to 1000G populations (right panels).** (PDF)

**S8 Fig. Quantile-quantile plot of GWAS of resistance to *M. tuberculosis* infection in the 3 cohorts from Vietnam, France and South Africa (333 uninfected vs 616 infected subjects) ( $\lambda = 1.06$ ).** (PDF)

**S9 Fig. Flowchart of the study.** (PDF)

**S10 Fig. Principal component analysis of the Vietnamese study cohort.** Plot of the first and second principal components of the 720 individuals from the Vietnamese cohort after projection A) on the 1000 Genomes phase 3 population, B) on the East Asian 1000 Genomes Phase 3 populations only.

(PDF)

**S11 Fig. Principal component analysis of the French cohort.** Plot of the first and second principal components of A) 573 individuals from the French cohort and B) the 157 subjects subsequently analyzed in the GWAS, after projection on the 1000 Genomes phase 3 populations.

(PDF)

**S12 Fig. Principal component analysis of the South African cohort.** Plot of the first and second principal components of 374 individuals from the South African cohort after projection on the 1000 Genomes phase 3 populations.

(PDF)

**S13 Fig. Stacked histogram of the tuberculin skin test (TST) distribution, stratified by our infection definition among 1108 household contacts in Vietnam.** The dashed line represents a 5 mm cut-off for TST. Uninfected subjects (in yellow) were negative for TST and QuantiFERON-TB Gold In-Tube (QFT-GIT) and infected subjects (in blue) positive for both.

(PDF)

**S14 Fig. Definition of the GWAS phenotype in the household contacts (HHCs) study in France.** A) Distribution of the tuberculin skin test (TST) induration among 516 HHCs (mean induration when 2 screening visits were done). The dashed line represents a 5 mm cut-off. TST was considered negative when the skin induration was i)  $< 5$  mm at both V1 and V2, ii)  $< 5$  mm at V1, when only one visit was done. TST was considered positive when the skin induration was i)  $\geq 5$  mm at both V1 and V2, ii)  $< 5$  mm at V1 and  $\geq 10$  mm at V2. B) Construction of a ROC curve based on the defined TST status to determine the optimal interferon- $\gamma$  release assay (IGRA) cut-off (175 pg/mL). C) Stacked histogram of the TST distribution stratified by our infection definition. Uninfected subjects (in yellow) presented a negative TST and a null production of IFN- $\gamma$ . Infected subjects (in blue) presented a positive TST and a positive IGRA (IFN- $\gamma$  production  $> 175$  pg/mL).

(PDF)

**S15 Fig. Definition of the GWAS phenotype in the family-based study in South Africa.** A) Distribution of the tuberculin skin test (TST) induration among 415 participants. The dashed line represents a 5 mm cut-off. B) Construction of a ROC curve based on the TST status to determine the optimal interferon- $\gamma$  release assay (IGRA) cut-off (20.9 pg/mL). C) Stacked histogram of the TST distribution stratified by our infection definition. Uninfected subjects (in yellow) presented a negative TST and a null production of IFN- $\gamma$ . Infected subjects (in blue) presented a positive TST and a positive IGRA (IFN- $\gamma$  production  $> 20.9$  pg/mL).

(PDF)

**S1 Table. Info score of the associated variants on 10q26.2.**

(PDF)

**S2 Table. Variants in the locus on chromosome 10q26.2 associated with resistance to *M. tuberculosis* infection in Vietnam (185 uninfected and 201 double positive infected subjects).**

(PDF)

**S3 Table. Frequency of rs17155120 minor allele T in 1000 Genomes phase 3 populations.**  
(PDF)

**S4 Table. Variants associated with resistance to *M. tuberculosis* infection in the 3 cohorts from Vietnam, France and South Africa (333 uninfected vs 616 infected subjects) with  $P < 5.0 \times 10^{-8}$ .**  
(PDF)

**S5 Table. GWAS cohort in Vietnam.**  
(PDF)

**S6 Table. Risk factors associated with (i) double positive tuberculin skin test (5 mm cut-off) and QuantiFERON-TB Gold In-Tube results or (ii) double positive TST/QFT-GIT and pulmonary TB patients, compared to the reference group (double negative TST and QFT-GIT results) in household contacts in Vietnam.**  
(PDF)

**S7 Table. Risk factors associated with both positive tuberculin skin test (5 mm cut-off) and positive interferon- $\gamma$  release assay (IGRA) results compared to the reference group (negative TST and null IGRA) in household contacts in France.**  
(PDF)

**S8 Table. GWAS cohort in France.**  
(PDF)

**S9 Table. Risk factors associated with both positive tuberculin skin test (5 mm cut-off) and interferon- $\gamma$  release assay (IGRA) results compared to the reference group (both negative TST and IGRA results) in the family-based study in South Africa.**  
(PDF)

**S10 Table. GWAS cohort in South Africa.**  
(PDF)

## Acknowledgments

We are grateful for the skilled technical assistance of Ho thi Ngoc, Binh Duong Center for Social Diseases Control, and Hanh La, Hospital for Dermato-Venereology, Ho Chi Minh City. We are indebted to the laboratory staff of the Hospital for Dermato-Venereology, Ho Chi Minh City, and the tuberculosis control teams of the Thu dau Mot, Thuận An, Ben Cat, and Di An districts. We sincerely thank all members of the community who participated in this work in South Africa and France and members of the lab of Human Genetics of Infectious Diseases for helpful discussions.

## Author Contributions

**Conceptualization:** Erwin Schurr.

**Data curation:** Jocelyn Quistrebert, Marianna Orlova.

**Formal analysis:** Jocelyn Quistrebert.

**Funding acquisition:** Jocelyn Quistrebert, Alexandre Alcaïs, Laurent Abel, Erwin Schurr.

**Investigation:** Le Thi Ton, Nguyễn Trong Luong, Nguyễn Thanh Danh.

**Methodology:** Jocelyn Quistrebert, Aurélie Cobat.

**Project administration:** Marianna Orlova, Christophe Delacourt, Eileen G. Hoal, Alexandre Alcaïs, Vu Hong Thai, Lai The Thành.

**Resources:** Gaspard Kerner, Le Thi Ton, Nguyễn Trong Luong, Nguyễn Thanh Danh, Quentin B. Vincent, Fabienne Jabot-Hanin, Yoann Seeleuthner, Jacinta Bustamante, Stéphanie Boisson-Dupuis, Nguyen Thu Huong, Nguyen Ngoc Ba, Jean-Laurent Casanova, Christophe Delacourt, Eileen G. Hoal, Alexandre Alcaïs, Vu Hong Thai, Lai The Thành.

**Software:** Gaspard Kerner, Fabienne Jabot-Hanin, Yoann Seeleuthner.

**Supervision:** Laurent Abel, Erwin Schurr, Aurélie Cobat.

**Validation:** Laurent Abel, Erwin Schurr, Aurélie Cobat.

**Visualization:** Jocelyn Quistrebert.

**Writing – original draft:** Jocelyn Quistrebert.

**Writing – review & editing:** Jocelyn Quistrebert, Marianna Orlova, Laurent Abel, Erwin Schurr, Aurélie Cobat.

## References

1. Global tuberculosis report 2019. World Health Organization; 2019. [cited 25 Oct 2019]. Available: [http://www.who.int/tb/publications/global\\_report/en/](http://www.who.int/tb/publications/global_report/en/)
2. Verrall AJ, Netea MG, Alisjahbana B, Hill PC, van Crevel R. Early clearance of *Mycobacterium tuberculosis*: a new frontier in prevention. *Immunology*. 2014; 141: 506–513. <https://doi.org/10.1111/imm.12223> PMID: 24754048
3. Simmons JD, Stein CM, Seshadri C, Campo M, Alter G, Fortune S, et al. Immunological mechanisms of human resistance to persistent *Mycobacterium tuberculosis* infection. *Nat Rev Immunol*. 2018; 18: 575–589. <https://doi.org/10.1038/s41577-018-0025-3> PMID: 29895826
4. 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. <https://doi.org/10.1371/journal.pone.0169539> PMID: 28060926
5. Pai M, Menzies D. The new IGRA and the old TST: making good use of disagreement. *Am J Respir Crit Care Med*. 2007; 175: 529–531. <https://doi.org/10.1164/rccm.200701-024ED> PMID: 17341646
6. 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; 137: 1071–1077. <https://doi.org/10.1378/chest.09-1852> PMID: 20040612
7. Mave V, Chandrasekaran P, Chavan A, Shivakumar SVBY, Danasekaran K, Paradkar M, et al. Infection free “resisters” among household contacts of adult pulmonary tuberculosis. *PLoS ONE*. 2019; 14: e0218034. <https://doi.org/10.1371/journal.pone.0218034> PMID: 31318864
8. Verrall AJ, Alisjahbana B, Apriani L, Novianty N, Nurani AC, van Laarhoven A, et al. Early Clearance of *Mycobacterium tuberculosis*: The INFECT Case Contact Cohort Study in Indonesia. *J Infect Dis*. 2019. <https://doi.org/10.1093/infdis/jiz168> PMID: 31298280
9. Stein CM, Zalwango S, Malone LL, Thiel B, Mupere E, Nsereko M, et al. Resistance and Susceptibility to *Mycobacterium tuberculosis* Infection and Disease in Tuberculosis Households in Kampala, Uganda. *Am J Epidemiol*. 2018; 187: 1477–1489. <https://doi.org/10.1093/aje/kwx380> PMID: 29304247
10. Abel L, Fellay J, Haas DW, Schurr E, Srikrishna G, Urbanowski M, et al. Genetics of human susceptibility to active and latent tuberculosis: present knowledge and future perspectives. *Lancet Infect Dis*. 2018; 18: e64–e75. [https://doi.org/10.1016/S1473-3099\(17\)30623-0](https://doi.org/10.1016/S1473-3099(17)30623-0) PMID: 29111156
11. 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; 206: 2583–2591. <https://doi.org/10.1084/jem.20090892> PMID: 19901083
12. Jabot-Hanin F, Cobat A, Feinberg J, Grange G, Remus N, Poirier C, et al. Major Loci on Chromosomes 8q and 3q Control Interferon  $\gamma$  Production Triggered by *Bacillus Calmette-Guérin* and 6-kDa Early Secretory Antigen Target, Respectively, in Various Populations. *J Infect Dis*. 2016; 213: 1173–1179. <https://doi.org/10.1093/infdis/jiv757> PMID: 26690346

13. Jabot-Hanin F, Cobat A, Feinberg J, Orlova M, Niay J, Deswarte C, et al. An eQTL variant of ZXDC is associated with IFN- $\gamma$  production following *Mycobacterium tuberculosis* antigen-specific stimulation. *Sci Rep*. 2017; 7: 1–8. <https://doi.org/10.1038/s41598-016-0028-x> PMID: [28127051](https://pubmed.ncbi.nlm.nih.gov/28127051/)
14. Stein CM, Zalwango S, Malone LL, Won S, Mayanja-Kizza H, Mugerwa RD, et al. Genome scan of *M. tuberculosis* infection and disease in Ugandans. *PLoS ONE*. 2008; 3: e4094. <https://doi.org/10.1371/journal.pone.0004094> PMID: [19116662](https://pubmed.ncbi.nlm.nih.gov/19116662/)
15. Igo RP, Hall NB, Malone LL, Hall JB, Truitt B, Qiu F, et al. Fine-mapping Analysis of a Chromosome 2 Region Linked to Resistance to *Mycobacterium tuberculosis* Infection in Uganda Reveals Potential Regulatory Variants. *Genes Immun*. 2019; 20: 473–483. <https://doi.org/10.1038/s41435-018-0040-1> PMID: [30100616](https://pubmed.ncbi.nlm.nih.gov/30100616/)
16. Cobat A, Poirier C, Hoal E, Boland-Auge A, de La Rocque F, Corrad 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; 211: 317–321. <https://doi.org/10.1093/infdis/jiu446> PMID: [25143445](https://pubmed.ncbi.nlm.nih.gov/25143445/)
17. Sobota RS, Stein CM, Kodaman N, Maro I, Wieland-Alter W, Igo RP, et al. A chromosome 5q31.1 locus associates with tuberculin skin test reactivity in HIV-positive individuals from tuberculosis hyper-endemic regions in east Africa. *PLoS Genet*. 2017; 13: e1006710. <https://doi.org/10.1371/journal.pgen.1006710> PMID: [28628665](https://pubmed.ncbi.nlm.nih.gov/28628665/)
18. Marcus JH, Novembre J. Visualizing the geography of genetic variants. *Bioinformatics*. 2017; 33: 594–595. <https://doi.org/10.1093/bioinformatics/btw643> PMID: [27742697](https://pubmed.ncbi.nlm.nih.gov/27742697/)
19. Kim-Hellmuth S, Bechheim M, Pütz B, Mohammadi P, Nédélec Y, Giangreco N, et al. Genetic regulatory effects modified by immune activation contribute to autoimmune disease associations. *Nat Commun*. 2017; 8: 1–10. <https://doi.org/10.1038/s41467-016-0009-6> PMID: [28232747](https://pubmed.ncbi.nlm.nih.gov/28232747/)
20. Calderon D, Nguyen MLT, Mezger A, Kathiria A, Müller F, Nguyen V, et al. Landscape of stimulation-responsive chromatin across diverse human immune cells. *Nat Genet*. 2019; 51: 1494–1505. <https://doi.org/10.1038/s41588-019-0505-9> PMID: [31570894](https://pubmed.ncbi.nlm.nih.gov/31570894/)
21. Stein CM, Hall NB, Malone LL, Mupere E. The household contact study design for genetic epidemiological studies of infectious diseases. *Front Genet*. 2013; 4: 61. <https://doi.org/10.3389/fgene.2013.00061> PMID: [23641253](https://pubmed.ncbi.nlm.nih.gov/23641253/)
22. Ma K, Qiu L, Mrasek K, Zhang J, Liehr T, Quintana LG, et al. Common Fragile Sites: Genomic Hotspots of DNA Damage and Carcinogenesis. *Int J Mol Sci*. 2012; 13: 11974–11999. <https://doi.org/10.3390/ijms130911974> PMID: [23109895](https://pubmed.ncbi.nlm.nih.gov/23109895/)
23. Yan S, Qiu L, Ma K, Zhang X, Zhao Y, Zhang J, et al. FATS is an E2-independent ubiquitin ligase that stabilizes p53 and promotes its activation in response to DNA damage. *Oncogene*. 2014; 33: 5424–5433. <https://doi.org/10.1038/onc.2013.494> PMID: [24240685](https://pubmed.ncbi.nlm.nih.gov/24240685/)
24. Song F, Zhang J, Qiu L, Zhao Y, Xing P, Lu J, et al. A functional genetic variant in fragile-site gene FATS modulates the risk of breast cancer in triparous women. *BMC Cancer*. 2015; 15: 559. <https://doi.org/10.1186/s12885-015-1570-9> PMID: [26223354](https://pubmed.ncbi.nlm.nih.gov/26223354/)
25. Manzanillo PS, Ayres JS, Watson RO, Collins AC, Souza G, Rae CS, et al. The ubiquitin ligase parkin mediates resistance to intracellular pathogens. *Nature*. 2013; 501: 512–516. <https://doi.org/10.1038/nature12566> PMID: [24005326](https://pubmed.ncbi.nlm.nih.gov/24005326/)
26. Franco LH, Nair VR, Scharn CR, Xavier RJ, Torrealba JR, Shiloh MU, et al. The Ubiquitin Ligase Smurf1 Functions in Selective Autophagy of *Mycobacterium tuberculosis* and Anti-tuberculous Host Defense. *Cell Host Microbe*. 2017; 21: 59–72. <https://doi.org/10.1016/j.chom.2016.11.002> PMID: [28017659](https://pubmed.ncbi.nlm.nih.gov/28017659/)
27. Wang L, Wu J, Li J, Yang H, Tang T, Liang H, et al. Host-mediated ubiquitination of a mycobacterial protein suppresses immunity. *Nature*. 2020; 577: 682–688. <https://doi.org/10.1038/s41586-019-1915-7> PMID: [31942069](https://pubmed.ncbi.nlm.nih.gov/31942069/)
28. Kastenhuber ER, Lowe SW. Putting p53 in Context. *Cell*. 2017; 170: 1062–1078. <https://doi.org/10.1016/j.cell.2017.08.028> PMID: [28886379](https://pubmed.ncbi.nlm.nih.gov/28886379/)
29. Behar SM, Martin CJ, Booty MG, Nishimura T, Zhao X, Gan H-X, et al. Apoptosis is an innate defense function of macrophages against *Mycobacterium tuberculosis*. *Mucosal Immunol*. 2011; 4: 279–287. <https://doi.org/10.1038/mi.2011.3> PMID: [21307848](https://pubmed.ncbi.nlm.nih.gov/21307848/)
30. Lim Y-J, Lee J, Choi J-A, Cho S-N, Son S-H, Kwon S-J, et al. M1 macrophage dependent-p53 regulates the intracellular survival of mycobacteria. *Apoptosis*. 2020; 25: 42–55. <https://doi.org/10.1007/s10495-019-01578-0> PMID: [31691131](https://pubmed.ncbi.nlm.nih.gov/31691131/)
31. Cruz A, Ludovico P, Torrado E, Gama JB, Sousa J, Gaifem J, et al. IL-17A Promotes Intracellular Growth of *Mycobacterium* by Inhibiting Apoptosis of Infected Macrophages. *Front Immunol*. 2015; 6. <https://doi.org/10.3389/fimmu.2015.00498> PMID: [26483789](https://pubmed.ncbi.nlm.nih.gov/26483789/)

32. Kveiborg M, Albrechtsen R, Couchman JR, Wewer UM. Cellular roles of ADAM12 in health and disease. *Int J Biochem Cell Biol.* 2008; 40: 1685–1702. <https://doi.org/10.1016/j.biocel.2008.01.025> PMID: [18342566](https://pubmed.ncbi.nlm.nih.gov/18342566/)
33. Paulissen G, Rocks N, Quesada-Calvo F, Gosset P, Foidart J-M, Noel A, et al. Expression of ADAMs and Their Inhibitors in Sputum from Patients with Asthma. *Mol Med.* 2006; 12: 171–179. <https://doi.org/10.2119/2006-00028.Paulissen> PMID: [17088949](https://pubmed.ncbi.nlm.nih.gov/17088949/)
34. Estrella C, Rocks N, Paulissen G, Quesada-Calvo F, Noël A, Vilain E, et al. Role of A Disintegrin And Metalloprotease-12 in Neutrophil Recruitment Induced by Airway Epithelium. *Am J Respir Cell Mol Biol.* 2009; 41: 449–458. <https://doi.org/10.1165/rcmb.2008-0124OC> PMID: [19213876](https://pubmed.ncbi.nlm.nih.gov/19213876/)
35. Zhou AX, El Hed A, Mercer F, Kozhaya L, Unutmaz D. The Metalloprotease ADAM12 Regulates the Effector Function of Human Th17 Cells. *PLoS One.* 2013; 8. <https://doi.org/10.1371/journal.pone.0081146> PMID: [24363794](https://pubmed.ncbi.nlm.nih.gov/24363794/)
36. Arlehamn CL, Seumois G, Gerasimova A, Huang C, Fu Z, Yue X, et al. Transcriptional profile of TB antigen-specific T cells reveals novel multifunctional features. *J Immunol.* 2014; 193: 2931–2940. <https://doi.org/10.4049/jimmunol.1401151> PMID: [25092889](https://pubmed.ncbi.nlm.nih.gov/25092889/)
37. Babu S, Bhat SQ, Kumar NP, Kumaraswami V, Nutman TB. Regulatory T Cells Modulate Th17 Responses in Patients with Positive Tuberculin Skin Test Results. *J Infect Dis.* 2010; 201: 20–31. <https://doi.org/10.1086/648735> PMID: [19929695](https://pubmed.ncbi.nlm.nih.gov/19929695/)
38. Coulter F, Parrish A, Manning D, Kampmann B, Mendy J, Garand M, et al. IL-17 Production from T Helper 17, Mucosal-Associated Invariant T, and  $\gamma\delta$  Cells in Tuberculosis Infection and Disease. *Front Immunol.* 2017; 8. <https://doi.org/10.3389/fimmu.2017.01252> PMID: [29075255](https://pubmed.ncbi.nlm.nih.gov/29075255/)
39. Gurdasani D, Barroso I, Zeggini E, Sandhu MS. Genomics of disease risk in globally diverse populations. *Nat Rev Genet.* 2019; 20: 520–535. <https://doi.org/10.1038/s41576-019-0144-0> PMID: [31235872](https://pubmed.ncbi.nlm.nih.gov/31235872/)
40. Global tuberculosis report 2015. World Health Organization; 2015. Available: <http://apps.who.int/iris/handle/10665/191102>
41. Immunization Country Profile. World Health Organization [cited 1 Oct 2019]. Available: [https://apps.who.int/immunization\\_monitoring/globalsummary/countries?countrycriteria%5Bcountry%5D%5B%5D=VNM&commit=OK](https://apps.who.int/immunization_monitoring/globalsummary/countries?countrycriteria%5Bcountry%5D%5B%5D=VNM&commit=OK)
42. Prevalence of HIV, total (% of population ages 15–49)—Vietnam | Data. [cited 1 Oct 2019]. Available: <https://data.worldbank.org/indicator/SH.DYN.AIDS.ZS?locations=VN>
43. Antoine D. Les cas de tuberculose maladie déclarés en France en 2006. *Bull Epidemiol Hebd.* 2008;10–11.
44. Prevalence of HIV, total (% of population ages 15–49)—France | Data. [cited 20 Nov 2020]. Available: <https://data.worldbank.org/indicator/SH.DYN.AIDS.ZS?locations=FR>
45. 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; 177: 1041–1047. <https://doi.org/10.1164/rccm.200711-1756OC> PMID: [18263798](https://pubmed.ncbi.nlm.nih.gov/18263798/)
46. 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; 128: 145–153. <https://doi.org/10.1007/s00439-010-0836-1> PMID: [20490549](https://pubmed.ncbi.nlm.nih.gov/20490549/)
47. Naidoo P, Theron G, Rangaka MX, Chihota VN, Vaughan L, Brey ZO, et al. The South African Tuberculosis Care Cascade: Estimated Losses and Methodological Challenges. *J Infect Dis.* 2017; 216: S702–S713. <https://doi.org/10.1093/infdis/jix335> PMID: [29117342](https://pubmed.ncbi.nlm.nih.gov/29117342/)
48. Mathema B, Andrews JR, Cohen T, Borgdorff MW, Behr M, Glynn JR, et al. Drivers of Tuberculosis Transmission. *J Infect Dis.* 2017; 216: S644–S653. <https://doi.org/10.1093/infdis/jix354> PMID: [29112745](https://pubmed.ncbi.nlm.nih.gov/29112745/)
49. Hesselting AC, Caldwell J, Cotton MF, Eley BS, Jaspan HB, Jennings K, et al. BCG vaccination in South African HIV-exposed infants—risks and benefits. *S Afr Med J.* 2009; 99: 88–91. PMID: [19418667](https://pubmed.ncbi.nlm.nih.gov/19418667/)
50. 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; 81: 559–575. <https://doi.org/10.1086/519795> PMID: [17701901](https://pubmed.ncbi.nlm.nih.gov/17701901/)
51. Das S, Forer L, Schönherr S, Sidore C, Locke AE, Kwong A, et al. Next-generation genotype imputation service and methods. *Nat Genet.* 2016; 48: 1284–1287. <https://doi.org/10.1038/ng.3656> PMID: [27571263](https://pubmed.ncbi.nlm.nih.gov/27571263/)
52. Loh P-R, Danecek P, Palamara PF, Fuchsberger C, A Reshef Y, K Finucane H, et al. Reference-based phasing using the Haplotype Reference Consortium panel. *Nat Genet.* 2016; 48: 1443–1448. <https://doi.org/10.1038/ng.3679> PMID: [27694958](https://pubmed.ncbi.nlm.nih.gov/27694958/)

53. 1000 Genomes Project Consortium, Auton A, Brooks LD, Durbin RM, Garrison EP, Kang HM, et al. A global reference for human genetic variation. *Nature*. 2015; 526: 68–74. <https://doi.org/10.1038/nature15393> PMID: 26432245
54. McCarthy S, Das S, Kretzschmar W, Delaneau O, Wood AR, Teumer A, et al. A reference panel of 64,976 haplotypes for genotype imputation. *Nat Genet*. 2016; 48: 1279–1283. <https://doi.org/10.1038/ng.3643> PMID: 27548312
55. Standards Diagnostic 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; 161: 1376–1395. <https://doi.org/10.1164/ajrccm.161.4.16141> PMID: 10764337
56. R: The R Project for Statistical Computing. [cited 5 Nov 2018]. Available: <https://www.r-project.org/>
57. Robin X, Turck N, Hainard A, Tiberti N, Lisacek F, Sanchez J-C, et al. pROC: an open-source package for R and S+ to analyze and compare ROC curves. *BMC Bioinformatics*. 2011; 12: 77. <https://doi.org/10.1186/1471-2105-12-77> PMID: 21414208
58. Bates D, Maechler M, Bolker B, Walker S, Christensen RHB, Singmann H, et al. lme4: Linear Mixed-Effects Models using “Eigen” and S4. 2018. Available: <https://CRAN.R-project.org/package=lme4>
59. Zhou X, Stephens M. Genome-wide Efficient Mixed Model Analysis for Association Studies. *Nat Genet*. 2012; 44: 821–824. <https://doi.org/10.1038/ng.2310> PMID: 22706312
60. Lloyd-Jones LR, Robinson MR, Yang J, Visscher PM. Transformation of Summary Statistics from Linear Mixed Model Association on All-or-None Traits to Odds Ratio. *Genetics*. 2018; 208: 1397–1408. <https://doi.org/10.1534/genetics.117.300360> PMID: 29429966
61. GitHub—YinLiLin/CMplot: Circular and Rectangular Manhattan Plot. [cited 24 Jun 2020]. Available: <https://github.com/YinLiLin/CMplot>
62. Pruim RJ, Welch RP, Sanna S, Teslovich TM, Chines PS, Gliedt TP, et al. LocusZoom: regional visualization of genome-wide association scan results. *Bioinformatics*. 2010; 26: 2336–2337. <https://doi.org/10.1093/bioinformatics/btq419> PMID: 20634204
63. Barrett JC, Fry B, Maller J, Daly MJ. Haploview: analysis and visualization of LD and haplotype maps. *Bioinformatics*. 2005; 21: 263–265. <https://doi.org/10.1093/bioinformatics/bth457> PMID: 15297300
64. Han B, Eskin E. Random-Effects Model Aimed at Discovering Associations in Meta-Analysis of Genome-wide Association Studies. *Am J Hum Genet*. 2011; 88: 586–598. <https://doi.org/10.1016/j.ajhg.2011.04.014> PMID: 21565292
65. Haeussler M, Zweig AS, Tyner C, Speir ML, Rosenbloom KR, Raney BJ, et al. The UCSC Genome Browser database: 2019 update. *Nucleic Acids Res*. 2019; 47: D853–D858. <https://doi.org/10.1093/nar/gky1095> PMID: 30407534
66. ENCODE Project Consortium. An integrated encyclopedia of DNA elements in the human genome. *Nature*. 2012; 489: 57–74. <https://doi.org/10.1038/nature11247> PMID: 22955616
67. Ernst J, Kellis M. Chromatin-state discovery and genome annotation with ChromHMM. *Nat Protoc*. 2017; 12: 2478–2492. <https://doi.org/10.1038/nprot.2017.124> PMID: 29120462
68. Kundaje A, Meuleman W, Ernst J, Bilenky M, Yen A, Heravi-Moussavi A, et al. Integrative analysis of 111 reference human epigenomes. *Nature*. 2015; 518: 317–330. <https://doi.org/10.1038/nature14248> PMID: 25693563
69. Nédélec Y, Sanz J, Baharian G, Szpiech ZA, Pacis A, Dumaine A, et al. Genetic Ancestry and Natural Selection Drive Population Differences in Immune Responses to Pathogens. *Cell*. 2016; 167: 657–669. e21. <https://doi.org/10.1016/j.cell.2016.09.025> PMID: 27768889
70. Kasela S, Kisand K, Tserel L, Kaleviste E, Remm A, Fischer K, et al. Pathogenic implications for autoimmune mechanisms derived by comparative eQTL analysis of CD4+ versus CD8+ T cells. *PLOS Genetics*. 2017; 13: e1006643. <https://doi.org/10.1371/journal.pgen.1006643> PMID: 28248954
71. Alasoo K, Rodrigues J, Mukhopadhyay S, Knights AJ, Mann AL, Kundu K, et al. Shared genetic effects on chromatin and gene expression indicate a role for enhancer priming in immune response. *Nat Genet*. 2018; 50: 424–431. <https://doi.org/10.1038/s41588-018-0046-7> PMID: 29379200
72. Schmiedel BJ, Singh D, Madrigal A, Valdovino-Gonzalez AG, White BM, Zapardiel-Gonzalo J, et al. Impact of Genetic Polymorphisms on Human Immune Cell Gene Expression. *Cell*. 2018; 175: 1701–1715.e16. <https://doi.org/10.1016/j.cell.2018.10.022> PMID: 30449622

## Supplemental Methods

### Immune assays in Vietnam

For tuberculin skin test (TST), 5 units (TU) of purified protein derivative (PPD) solution per 0.1 mL test dose (Tubertest, Sanofi Pasteur, Lyon, France) were injected intradermally into the forearm of the household contacts (HHCs) using the Mantoux method. The diameter of the induration was measured by trained personal 48-72 hours after the injection. QuantiFERON-TB Gold In-Tube test (QFT-GIT) assay was performed on blood samples collected before the TST was carried out, in accordance with the kit manufacturer's instructions. Briefly, peripheral blood was drawn directly into each of three blood collection tubes under vacuum: the nil tube or negative control, the mitogen tube containing phytohaemagglutinin (PHA) as a positive control and the TB antigen tube containing specific antigens for *M. tuberculosis* (early secretory antigenic target 6, ESAT6, culture filtrate protein 10 and TB 7.7 antigen). The contents of each tube were mixed and the tubes were incubated at 37°C for 16 to 24 hours within 8 hours of collection. Tubes were centrifuged at 2500 g for 15 minutes at room temperature, plasma was collected and stored at -20°C. The concentration of interferon-gamma (IFN- $\gamma$ ) in each sample was determined with the QFT-GIT ELISA (enzyme-linked immunosorbent assay) kit. Results were categorized by the manufacturer's software as positive, negative or indeterminate.

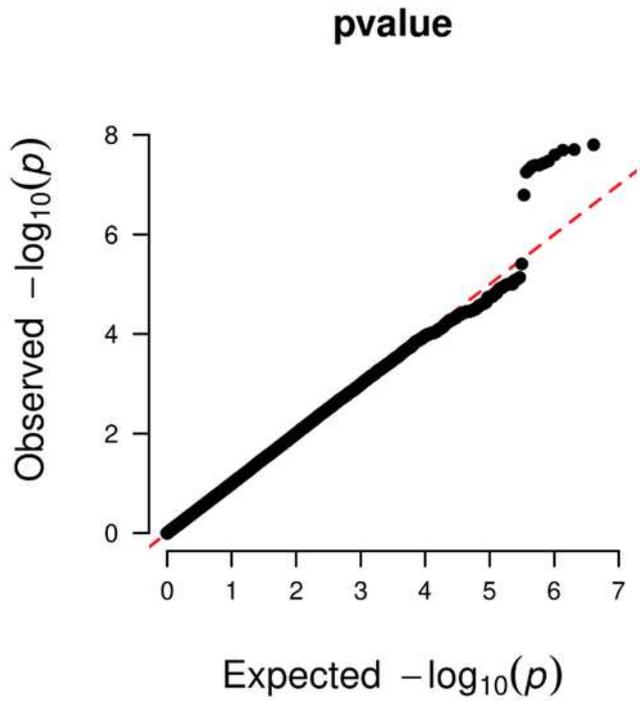
### Immune assays in France

TSTs were performed on the volar surface of the forearm by intradermal injection of 0.1 mL (5 TU) of purified protein derivative tuberculin obtained from a human *M. tuberculosis* strain (Tubertest, Sanofi Pasteur, Lyon, France). The skin induration was read within 72 hours. IGRAs were performed once either at V1 or V2. Blood was drawn from each individual and peripheral blood mononuclear cells (PBMCs) were separated. They were activated with one of the following stimulations: ESAT-6 (rdESAT6, Statens-Serum-Institute, Denmark; 2.5 $\mu$ g/mL),

PHA (6.25 µg/mL) for positive control and medium alone for negative control. After 4 days of stimulation, supernatants were assayed for IFN-γ by ELISA according to the manufacturer's recommendations (Pelikin Compact; CLB). The final results were standardized per million PBMC's, in the unit pg/mL/10<sup>6</sup> PBMC's.

### **Immune assays in South Africa**

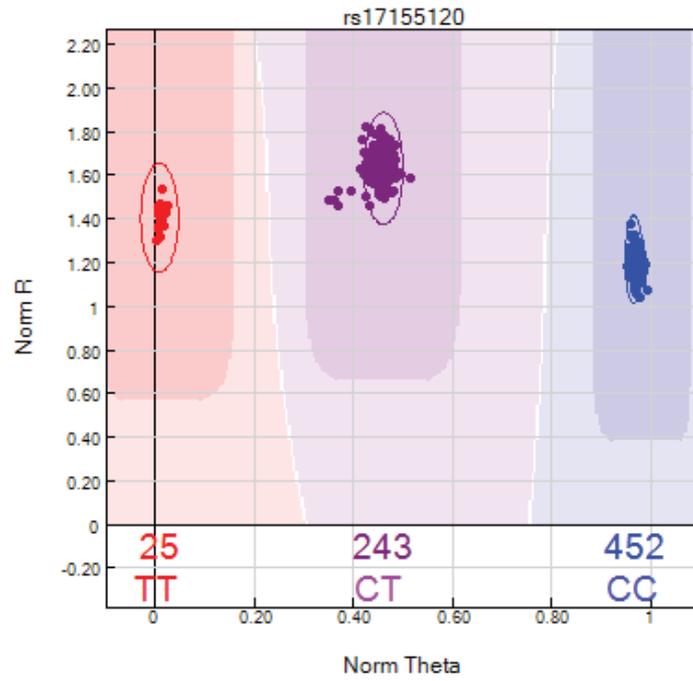
TSTs were performed by specially trained health care providers using the Mantoux method with PPD solution (2 T.U Tuberculin PPD RT 23, Statens Serum Institut, Copenhagen, Denmark) and skin induration was read 48-72 hours later. Whole-blood specimens were stimulated by ESAT-6 (rdESAT6, Statens-Serum-Institute, Denmark; 20µg/mL) and PHA (5 µg/mL) for positive control. Cultures incubated without stimulation in serum-free medium provided the negative control. IFN-γ levels were measured on day 3 after stimulation (limit of detection at 31 pg/mL).



**S1 Figure. Quantile-quantile plot of GWAS resistance to tuberculosis infection in Vietnam.** Plot of the expected distribution of association test statistics (x axis) for 5,591,951 variants compared to the observed values (y axis). Red line is the null hypothesis of no association ( $y = x$ ).

**S1 Table. Info score of the associated variants on 10q26.2.**

<b>Variant</b>	<b>chr:pos</b>	<b>Vietnam</b>	<b>France</b>	<b>South Africa</b>
rs11245088	10:128362986	0.91	-	0.92
rs72163291	10:128357575	0.92	0.92	-
rs17155143	10 :128362587	0.93	0.94	0.93
rs7909756	10:128360681	0.92	0.92	0.92
rs28703703	10:128350796	0.99	0.98	0.99
rs56106518	10:128350239	0.86	0.83	-
rs75482972	10:128347258	0.99	0.98	0.96
rs17155120	10:128348107	genotyped	genotyped	genotyped
rs73370887	10:128356858	0.95	0.96	0.93
rs79608098	10:128349019	0.99	0.99	0.99
rs61750007	10:128360275	0.94	0.95	0.94
rs77513326	10:128351876	0.99	0.99	genotyped
rs79918233	10:128356265	0.97	0.98	0.97
rs147584264	10:128364818	0.92	0.94	-
rs191820708	10:128366904	0.91	0.93	-
rs201178890	10:128368249	0.92	0.94	-
rs202189321	10:128370320	0.90	0.93	-
rs118037357	10:128372646	0.93	0.97	0.95

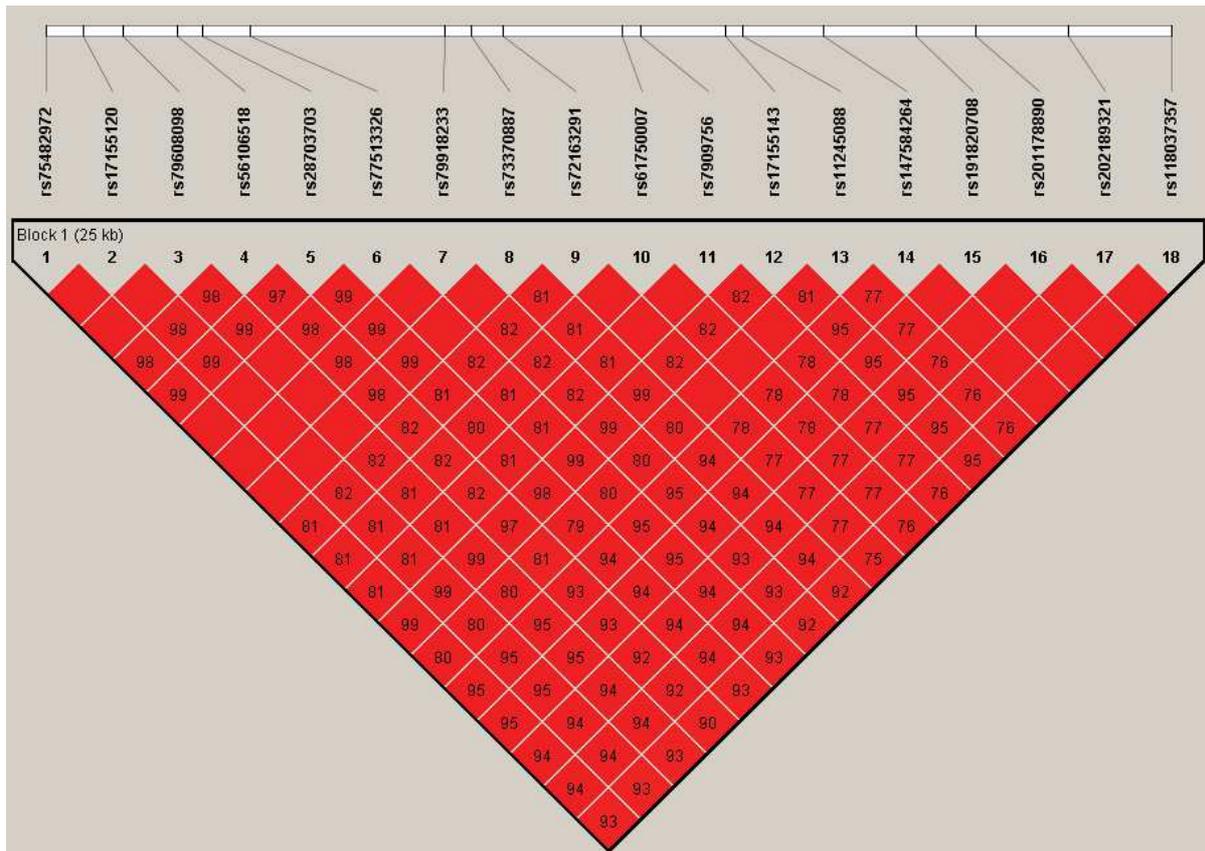


**S2 Figure. Intensity cluster plot for rs17155120 of the Illumina Infinium OmniExpressExome-8-v1 chip in the 720 initial Vietnamese individuals who were genotyped**

**S2 Table. Variants in the locus on chromosome 10q26.2 associated with resistance to *M. tuberculosis* infection in Vietnam (185 uninfected and 201 double positive infected subjects).**

<b>Variant</b>	<b>EA</b>	<b>EAF</b>	<b>OR (95%CI)</b>	<b>P value</b>
rs56106518	C	0.24	0.39 (0.30-0.47)	$1.37 \times 10^{-7}$
rs72163291	ins	0.26	0.41 (0.33-0.49)	$2.09 \times 10^{-7}$
rs28703703	G	0.21	0.40 (0.32-0.49)	$2.40 \times 10^{-7}$
rs17155120	T	0.21	0.40 (0.32-0.49)	$2.55 \times 10^{-7}$
rs11245088	C	0.26	0.41 (0.40-0.42)	$2.72 \times 10^{-7}$
rs7909756	G	0.26	0.41 (0.33-0.49)	$2.73 \times 10^{-7}$
rs75482972	A	0.21	0.41 (0.32-0.49)	$2.82 \times 10^{-7}$
rs17155143	A	0.21	0.39 (0.30-0.47)	$3.02 \times 10^{-7}$
rs79608098	T	0.21	0.41 (0.32-0.49)	$3.28 \times 10^{-7}$
rs77513326	A	0.21	0.40 (0.32-0.49)	$3.56 \times 10^{-7}$
rs73370887	A	0.21	0.39 (0.31-0.48)	$3.82 \times 10^{-7}$
rs61750007	C	0.25	0.42 (0.34-0.50)	$4.91 \times 10^{-7}$
rs79918233	A	0.21	0.40 (0.31-0.49)	$5.13 \times 10^{-7}$
rs147584264	C	0.21	0.41 (0.32-0.50)	$2.01 \times 10^{-6}$
rs191820708	A	0.21	0.41 (0.32-0.50)	$2.59 \times 10^{-6}$
rs201178890	T	0.21	0.41 (0.32-0.50)	$2.71 \times 10^{-6}$
rs202189321	T	0.21	0.41 (0.32-0.50)	$2.72 \times 10^{-6}$
rs118037357	A	0.20	0.42 (0.32-0.51)	$3.54 \times 10^{-6}$

CI, confidence intervals; EA, effect allele; EAF, effect allele frequency; OR, odds ratio; ins, insertion



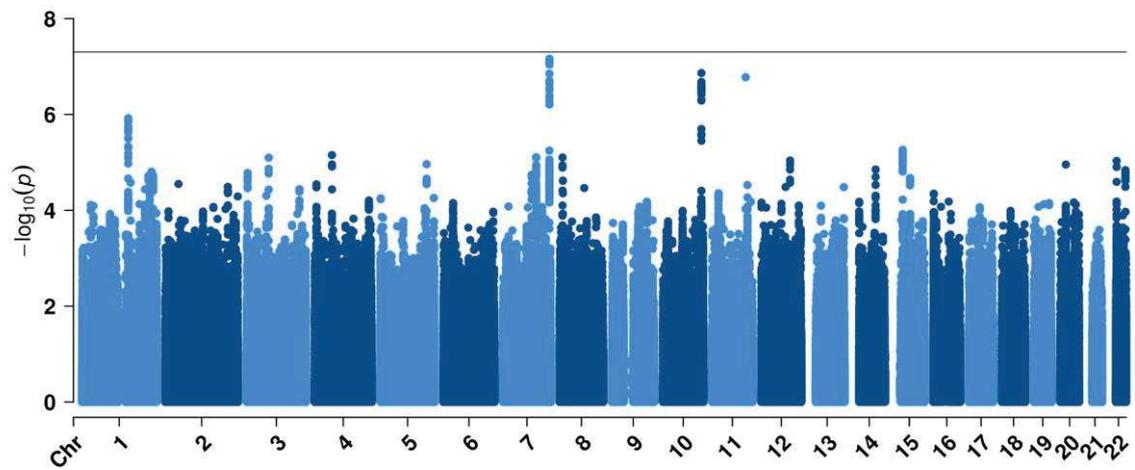
**S3 Figure. Haploview LD graph of the top associations on chromosome 10q26.2 in Vietnam.** Pairwise LD coefficients  $r^2$  are shown in each cell between 18 variants with  $P < 5.0 \times 10^{-6}$ . Top genotyped variant rs17155120 is second from left.

**S3 Table. Frequency of rs17155120 minor allele T in 1000 Genomes phase 3 populations<sup>1</sup>.**

1000 Genomes population	rs17155120 T frequency
ALL	0.16
African	0.07
Yoruba in Ibadan, Nigeria	0.07
Luhya in Webuye, Kenya	0.08
Gambian in Western Divisions in the Gambia	0.06
Mende in Sierra Leone	0.05
Esan in Nigeria	0.06
Americans of African Ancestry in SW USA	0.11
African Caribbeans in Barbados	0.08
American	0.18
Mexican Ancestry from Los Angeles USA	0.19
Puerto Ricans from Puerto Rico	0.12
Colombians from Medellin, Colombia	0.15
Peruvians from Lima, Peru	0.27
East Asian	0.18
Han Chinese in Beijing, China	0.20
Japanese in Tokyo, Japan	0.12
Southern Han Chinese	0.22
Chinese Dai in Xishuangbanna, China	0.16
Kinh in Ho Chi Minh City, Vietnam	0.18
European	0.16
Utah Residents with Northern and Western European Ancestry	0.18
Toscani in Italia	0.14
Finnish in Finland	0.16
British in England and Scotland	0.16
Iberian Population in Spain	0.19
South Asian	0.24
Gujarati Indian from Houston, Texas	0.22
Punjabi from Lahore, Pakistan	0.26
Bengali from Bangladesh	0.22
Sri Lankan Tamil from the UK	0.23
Indian Telugu from the UK	0.25

#### Reference

1. Machiela, M.J., and Chanock, S.J. (2015). LDlink: a web-based application for exploring population-specific haplotype structure and linking correlated alleles of possible functional variants. *Bioinformatics* 31, 3555–3557



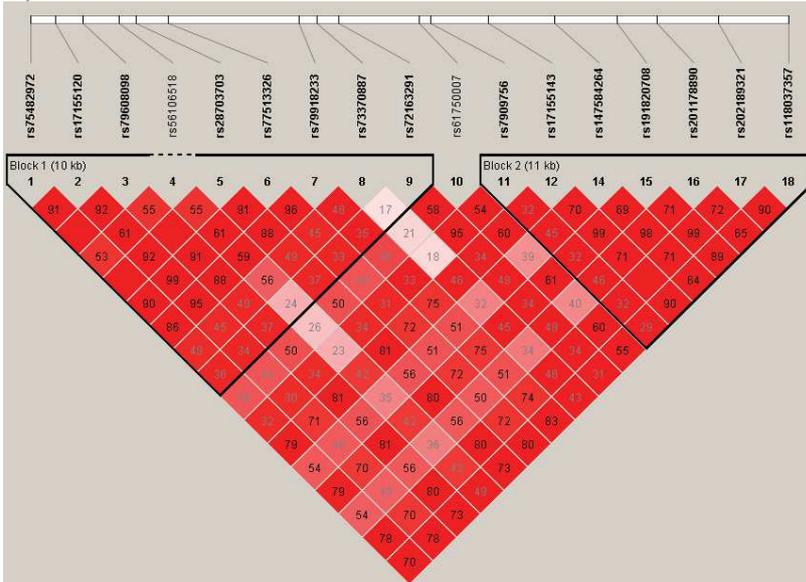
**S4 Figure.** Manhattan plot for the GWAS of resistance to *M. tuberculosis* infection in Vietnam (185 uninfected vs 201 infected participants).

**S4 Table. Variants associated with resistance to *M. tuberculosis* infection in the 3 cohorts from Vietnam, France and South Africa (333 uninfected vs 616 infected subjects) with  $P < 5.0 \times 10^{-8}$ .**

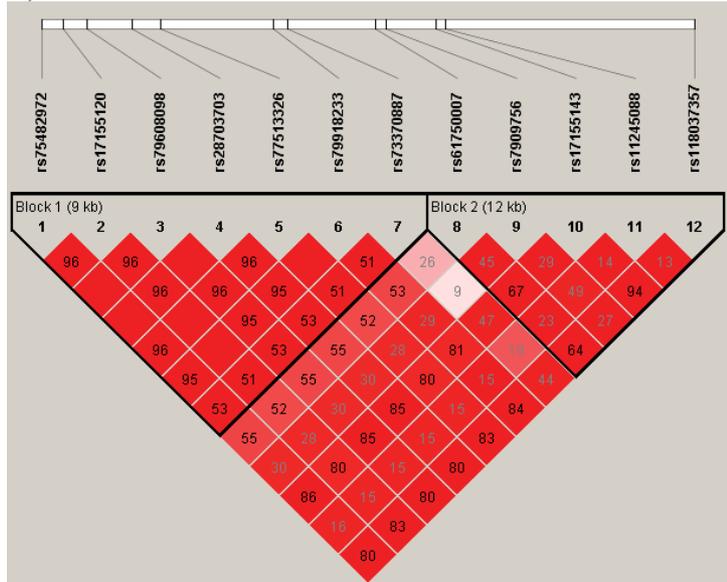
<b>Variant</b>	<b>EA</b>	<b>OR (95%CI)</b>	<b><i>P</i> value</b>	<b><i>P</i><sub>het</sub></b>	<b>Chromosome</b>	<b>Nearest Gene</b>
rs17155120	T	0.50 (0.45-0.55)	$1.26 \times 10^{-9}$	0.31	10	<i>C10orf90</i>
rs118037357	A	0.47 (0.41-0.53)	$1.89 \times 10^{-9}$	0.50	10	<i>C10orf90</i>
rs77513326	A	0.50 (0.45-0.56)	$2.26 \times 10^{-9}$	0.29	10	<i>C10orf90</i>
rs79918233	A	0.51 (0.46-0.56)	$5.12 \times 10^{-9}$	0.24	10	<i>C10orf90</i>
rs17155143	A	0.53 (0.48-0.58)	$7.30 \times 10^{-9}$	0.10	10	<i>C10orf90</i>
rs28703703	G	0.55 (0.50-0.60)	$9.18 \times 10^{-9}$	0.11	10	<i>C10orf90</i>
rs79608098	T	0.55 (0.50-0.60)	$1.15 \times 10^{-8}$	0.14	10	<i>C10orf90</i>
rs75482972	A	0.55 (0.50-0.60)	$1.32 \times 10^{-8}$	0.11	10	<i>C10orf90</i>
rs61750007	C	0.56 (0.51-0.61)	$1.55 \times 10^{-8}$	0.12	10	<i>C10orf90</i>

CI, confidence intervals; EA, effect allele; EAF, effect allele frequency; OR, odds ratio

**A) France**



**B) South Africa**



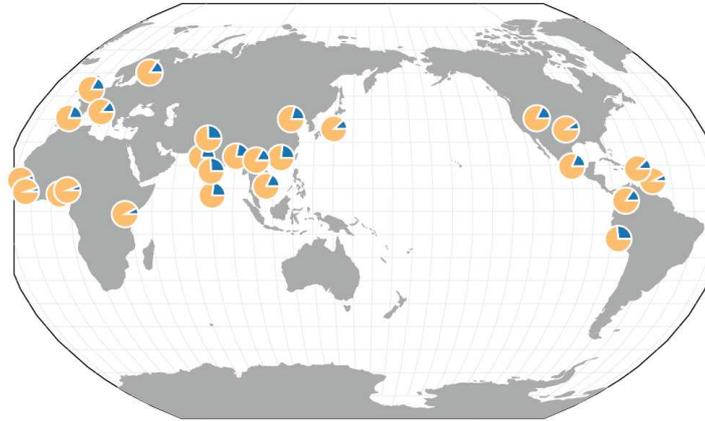
**S5 Figure. Haploview LD graphs of the locus on chromosome 10q26.2 in replication cohorts. Pairwise LD coefficients  $r^2$  are shown in each cell **A)** between the 17 variants of the locus in the French cohort, and **B)** between 12 out of the 18 variants of the locus in the South African cohort. Top genotyped variant rs17155120 is second from left in each figure.**

**S5 Table. GWAS cohort in Vietnam.**

	<b>Uninfected</b>	<b>Infected</b>	
	<b>TST-/QFT-GIT-</b>	<b>TST+/QFT-GIT+</b>	<b>PTB</b>
n	185	201	152
Age, years, mean(sd)	36(16)	32(14)	48(12)
Sex, %F	68	59	22
TST, mm, mean(sd)	1(1)	16(6)	-
QFT-GIT	negative	positive	-

PTB, pulmonary tuberculosis; QFT-GIT, QuantiFERON-TB Gold In-Tube; TST, tuberculin skin test

chr10:128348107 T/C



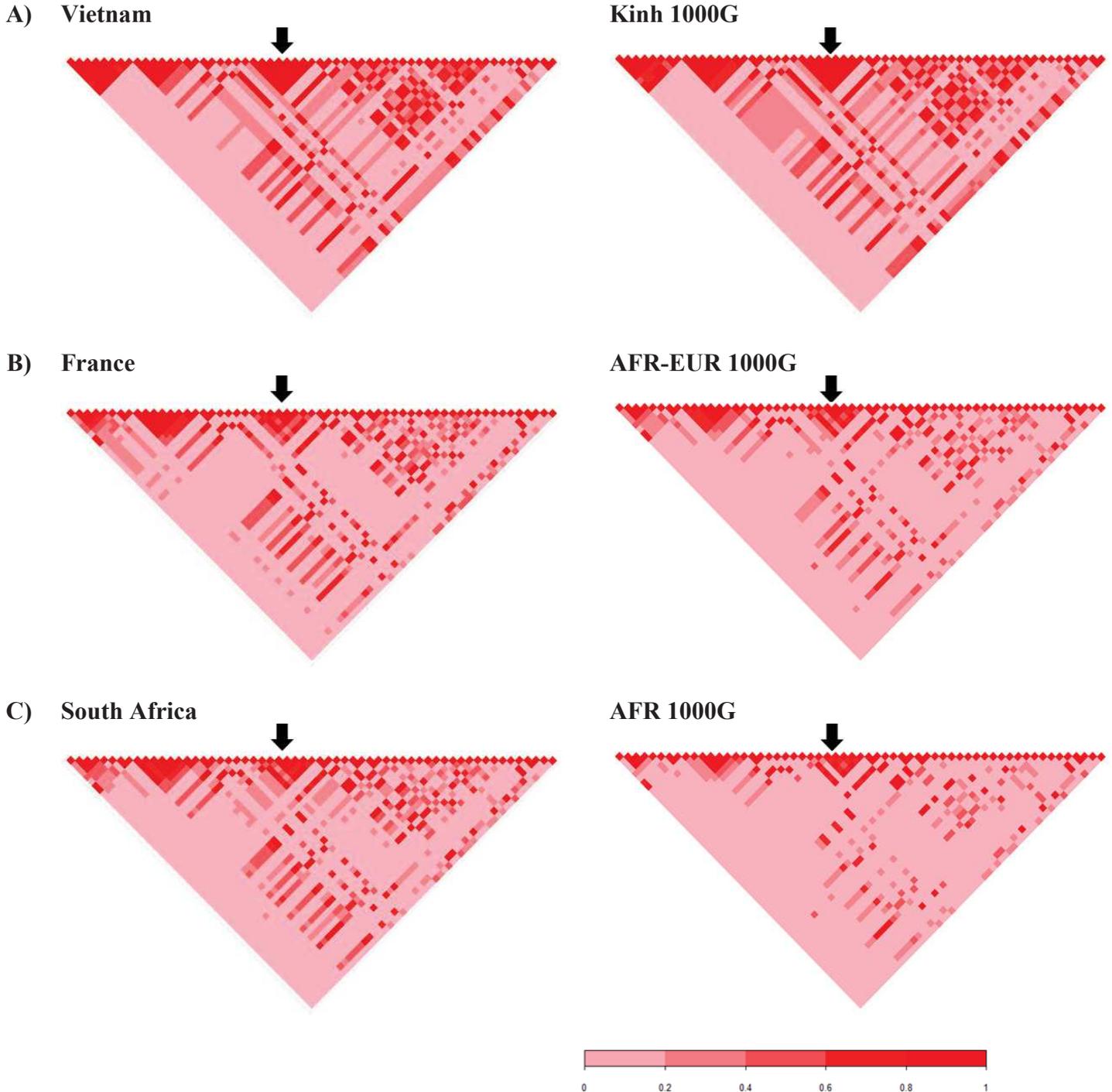
**S6 Figure. Geographic distribution of the variant rs17155120 in 1000G phase 3 populations.** The T allele (in blue) is protective against *M. tuberculosis* infection. Map generated from the Geography of Genetic Variants Browser (<http://www.popgen.uchicago.edu/ggv>).

**S6 Table. Risk factors associated with (i) double positive tuberculin skin test (5 mm cut-off) and QuantiFERON-TB Gold In-Tube results or (ii) double positive TST/QFT-GIT and pulmonary TB patients, compared to the reference group (double negative TST and QFT-GIT results) in household contacts in Vietnam.**

Characteristics	TST-/QFT-GIT-	TST+/QFT-GIT+	OR (95%CI)	TST+/QFT-GIT+/ PTB	OR (95%CI)
	n=188	n=512		n=664	
Gender					
Male	61	189	ref	223	ref
Female	127	323	0.84 (0.53-1.33)	441	0.96 (0.80-1.08)
Age, years, mean(sd)	39(16)	32(14)	0.78 (0.30-1.62)	40(13)	0.93 (0.53-1.25)
BCG vaccination <sup>a</sup>					
No	99	260	ref	NA	NA
Yes	69	190	1.01 (0.64-1.62)	NA	NA
Number of HHC per room, mean(sd)	2.4(1.4)	2.4(1.3)	0.99 (0.81-1.20)	NA	NA
Duration of index cough before treatment, months <sup>b</sup>					
<2	63	193	ref	NA	NA
2-3	84	195	1.64 (0.95-2.84)	NA	NA
>3	40	120	1.07 (0.57-2.02)	NA	NA

<sup>a</sup>missing information for 62 individuals

<sup>b</sup>missing information for 5 individuals

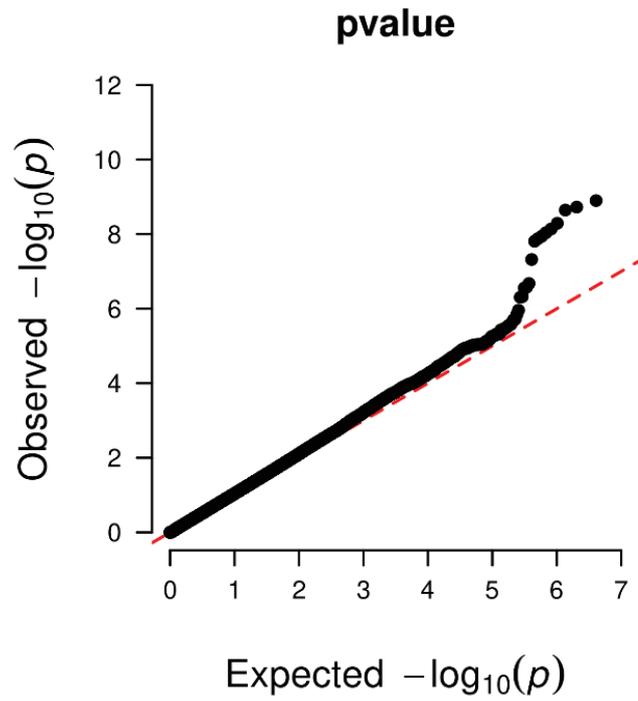


**S7 Figure. Pairwise linkage disequilibrium ( $r^2$ ) in A) Vietnamese, B) French and C) South African study cohorts (left panels) between 60 variants in a 30 kb window around rs17155120 (arrow) as compared to 1000G populations (right panels).**

**S7 Table. Risk factors associated with both positive tuberculin skin test (5 mm cut-off) and positive interferon- $\gamma$  release assay (IGRA) results compared to the reference group (negative TST and null IGRA) in household contacts in France.**

<b>Characteristics</b>	<b>TST-/IGRA- n=33</b>	<b>TST+/IGRA+ n=147</b>	<b>OR (95%CI)</b>
Gender			
Male	14	72	ref
Female	19	75	0.69 (0.10-4.60)
Age, years, mean(sd)	17(21)	28(18)	<b>1.07 (1.01-1.15)</b>
Contact duration with index, log hours, mean(sd)	5(2)	5(2)	0.87 (0.46-1.51)
Index contagiousness			
High	15	79	ref
Low-intermediate	18	68	0.57 (0.02-11.11)
Tuberculosis incidence of country of birth			
High	5	51	ref
Low-intermediate	28	96	0.14 (0.05-1.37)
Complementary health insurance <sup>a</sup>			
No	10	45	ref
Yes	23	99	0.58 (0.01-8.33)

<sup>a</sup>information missing for 3 individuals

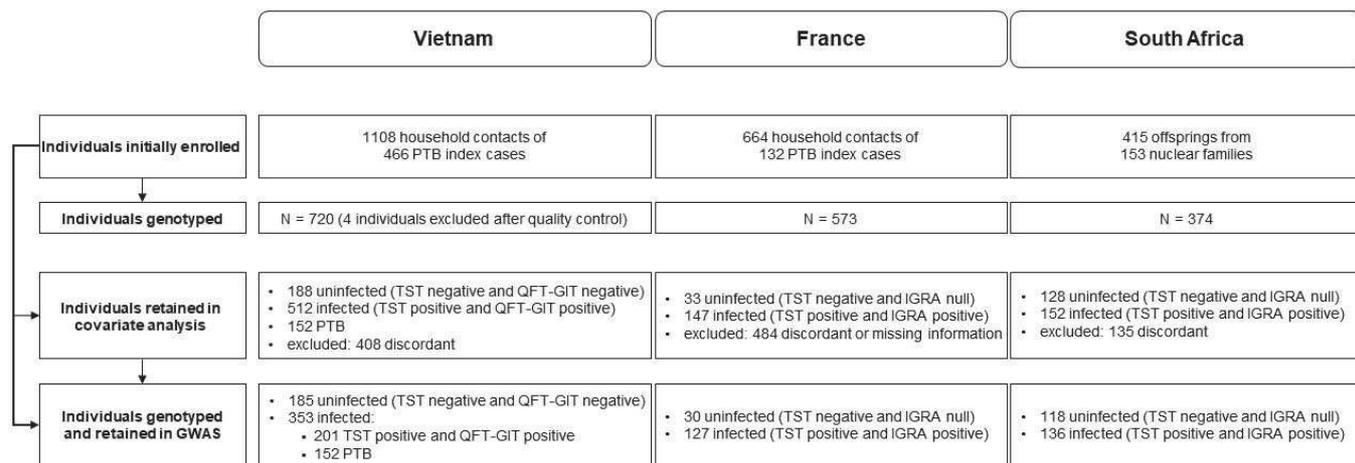


**S8 Figure.** Quantile-quantile plot of GWAS of resistance to *M. tuberculosis* infection in the 3 cohorts from Vietnam, France and South Africa (333 uninfected vs 616 infected subjects) ( $\lambda = 1.06$ ).

**S8 Table. GWAS cohort in France.**

	<b>TST-/IGRA-</b>	<b>TST+/IGRA+</b>
n	30	127
Age, years, mean(sd)	17(19)	28(18)
Sex, %F	57	52
TST, mm, mean(sd)	0(1)	16(4)
IGRA, IFN- $\gamma$ , pg/mL	0	>175

IGRA, interferon- $\gamma$  release assay; TST, tuberculin skin test

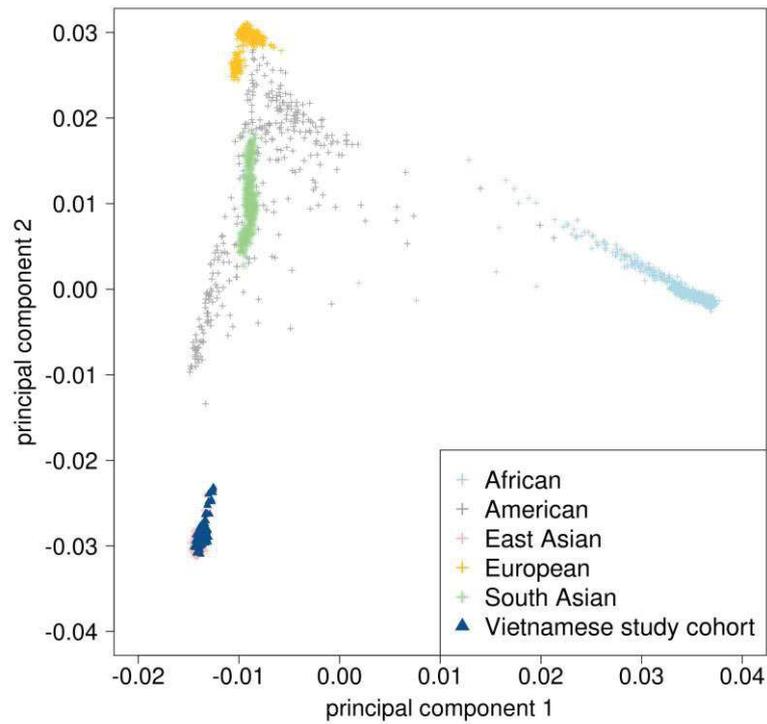


**S9 Figure. Flowchart of the study.**

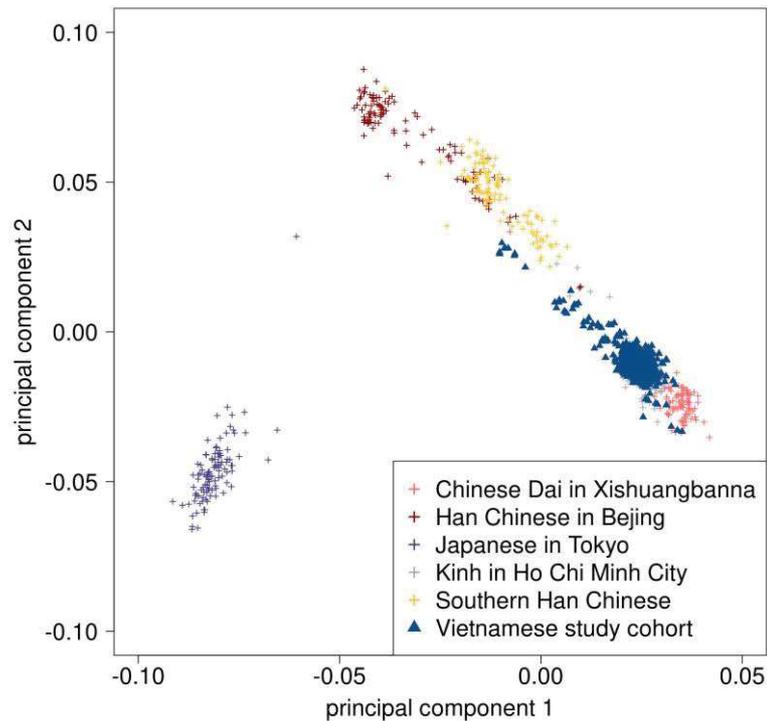
**S9 Table. Risk factors associated with both positive tuberculin skin test (5 mm cut-off) and interferon- $\gamma$  release assay (IGRA) results compared to the reference group (both negative TST and IGRA results) in the family-based study in South Africa.**

<b>Characteristics</b>	<b>TST-/IGRA- n=128</b>	<b>TST+/IGRA+ n=152</b>	<b>OR (95%CI)</b>
Gender			
Male	60	69	ref
Female	58	83	1.22 (0.67-2.17)
Age, years, mean(sd)	10(6)	15(6)	<b>1.30 (1.20-1.44)</b>

A)



B)



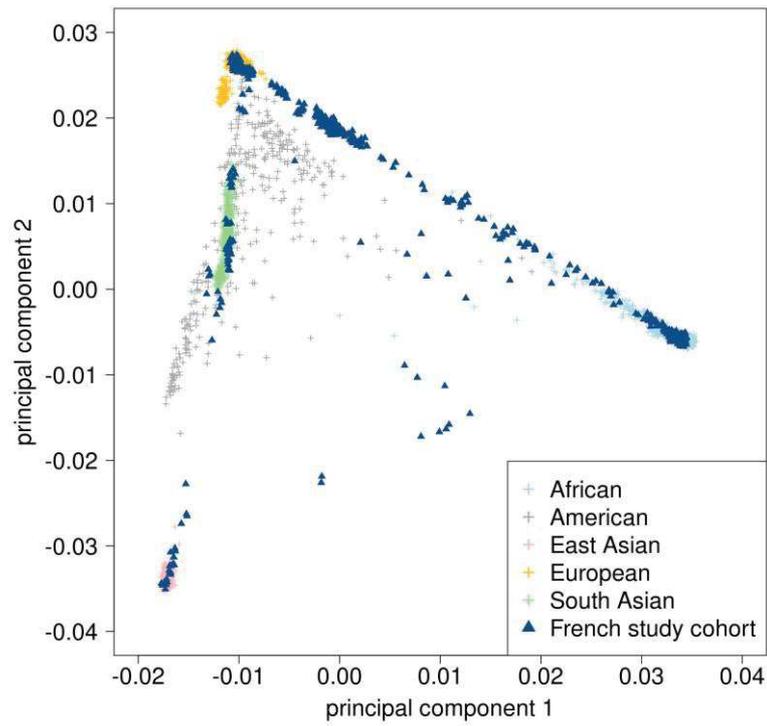
**S10 Figure. Principal component analysis of the Vietnamese study cohort.** Plot of the first and second principal components of the 720 individuals from the Vietnamese cohort after projection **A)** on the 1000 Genomes phase 3 population, **B)** on the East Asian 1000 Genomes Phase 3 populations only.

**S10 Table. GWAS cohort in South Africa.**

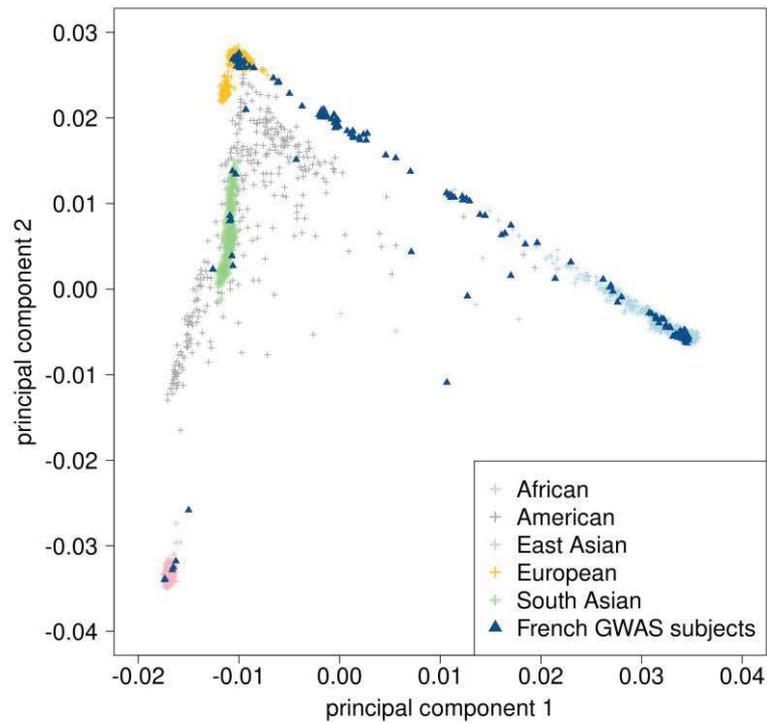
	<b>TST-/IGRA-</b>	<b>TST+/IGRA+</b>
n	118	136
Age, years, mean(sd)	10(4)	15(6)
Sex, %F	49	56
TST, mm, mean(sd)	0(0)	17(5)
IGRA, IFN- $\gamma$ , pg/mL	0	>20.9

IGRA, interferon- $\gamma$  release assay; TST, tuberculin skin test

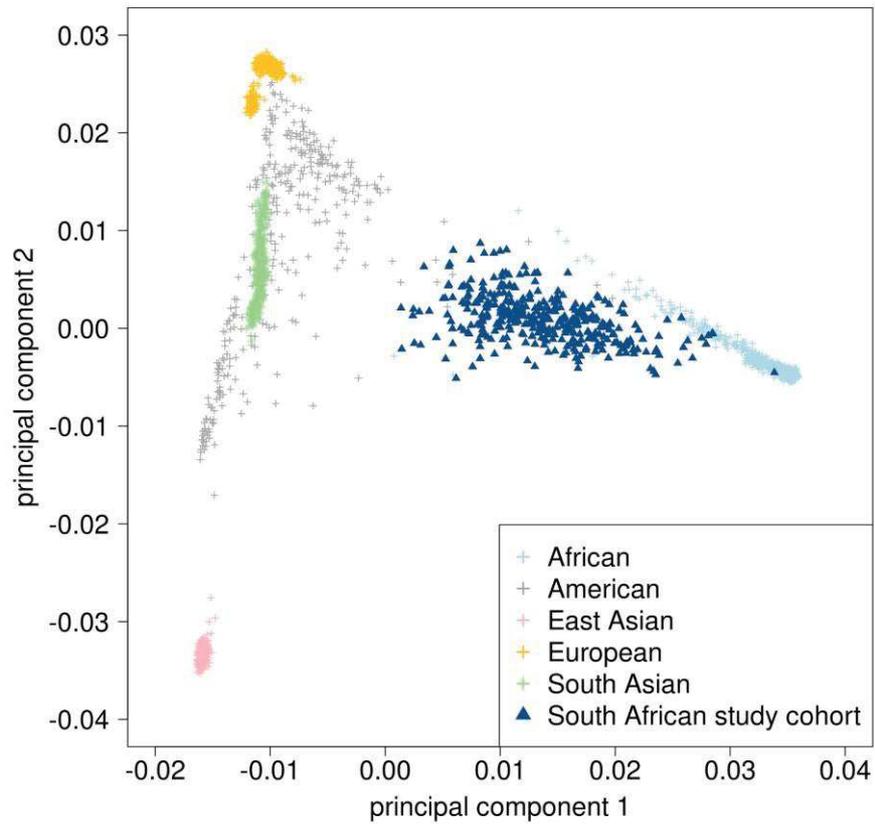
**A)**



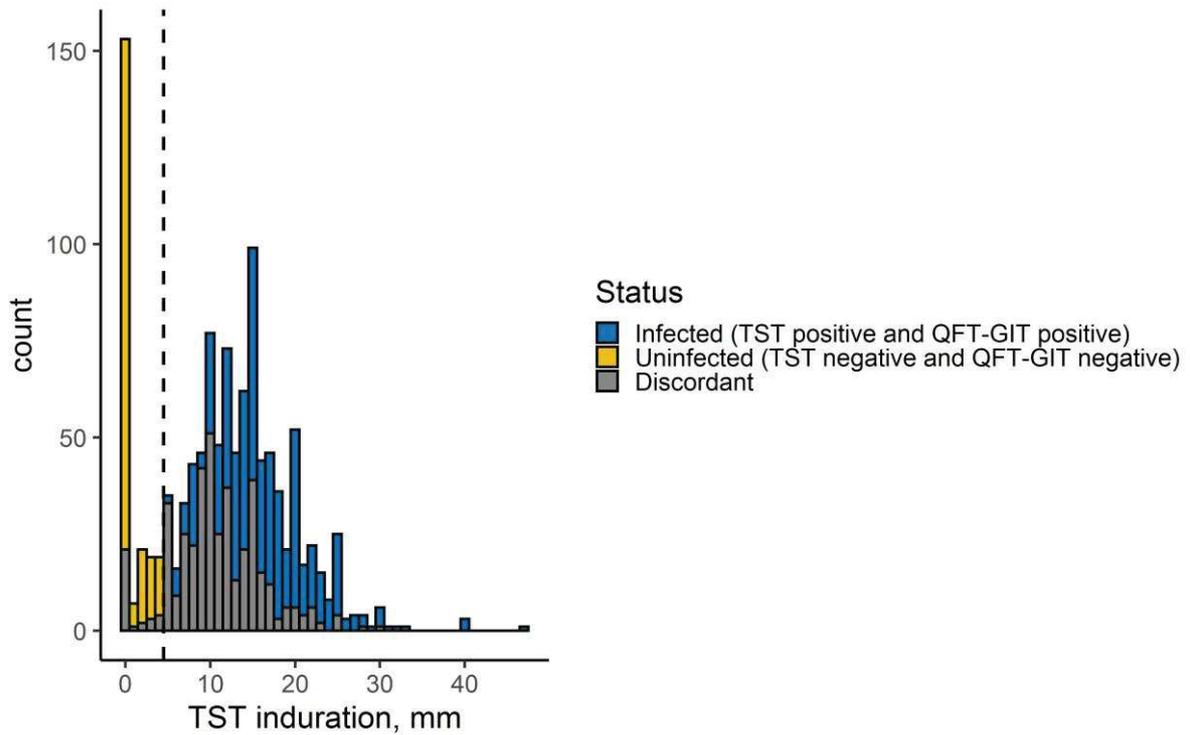
**B)**



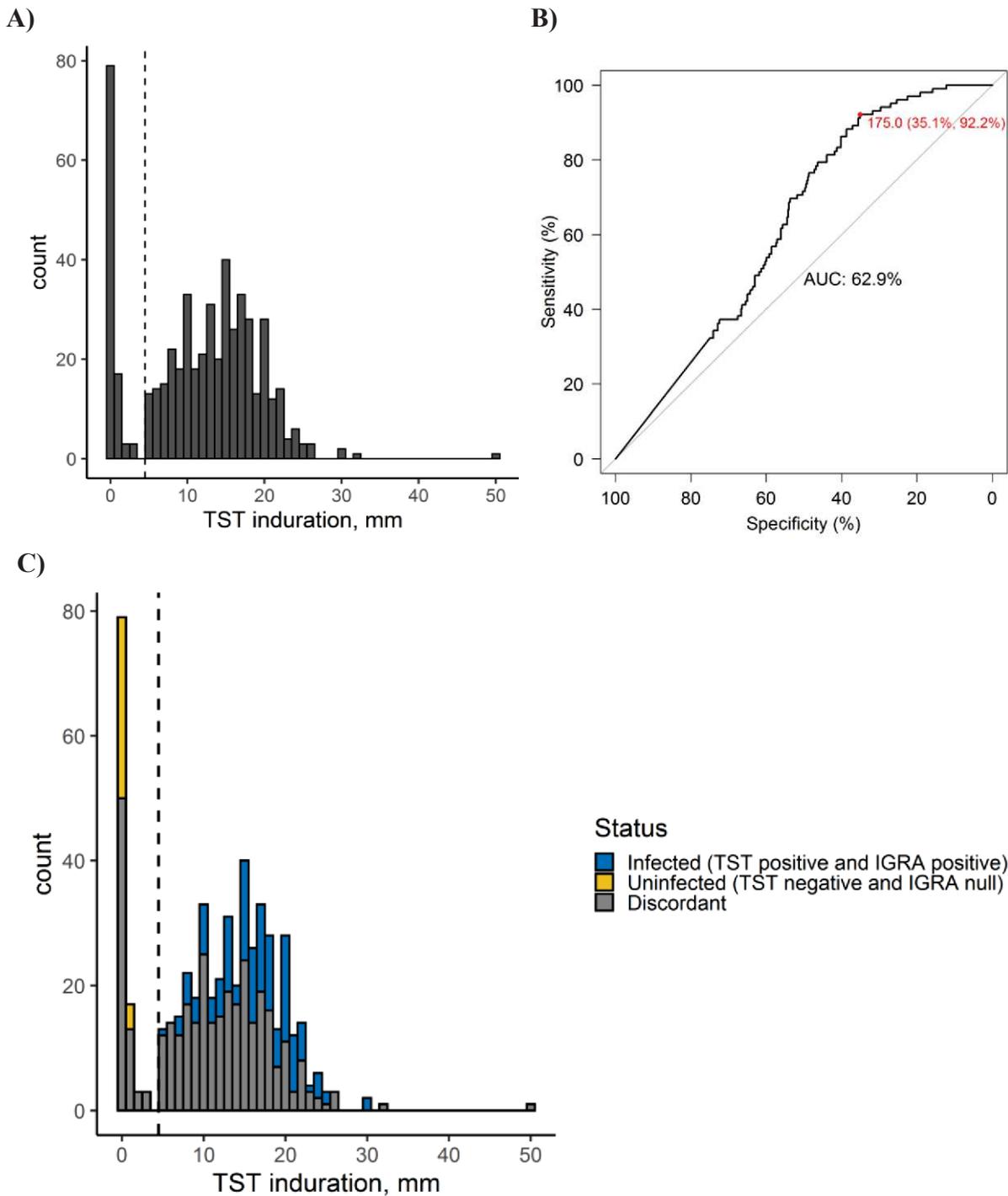
**S11 Figure. Principal component analysis of the French cohort.** Plot of the first and second principal components of **A)** 573 individuals from the French cohort and **B)** the 157 subjects subsequently analyzed in the GWAS, after projection on the 1000 Genomes phase 3 populations.



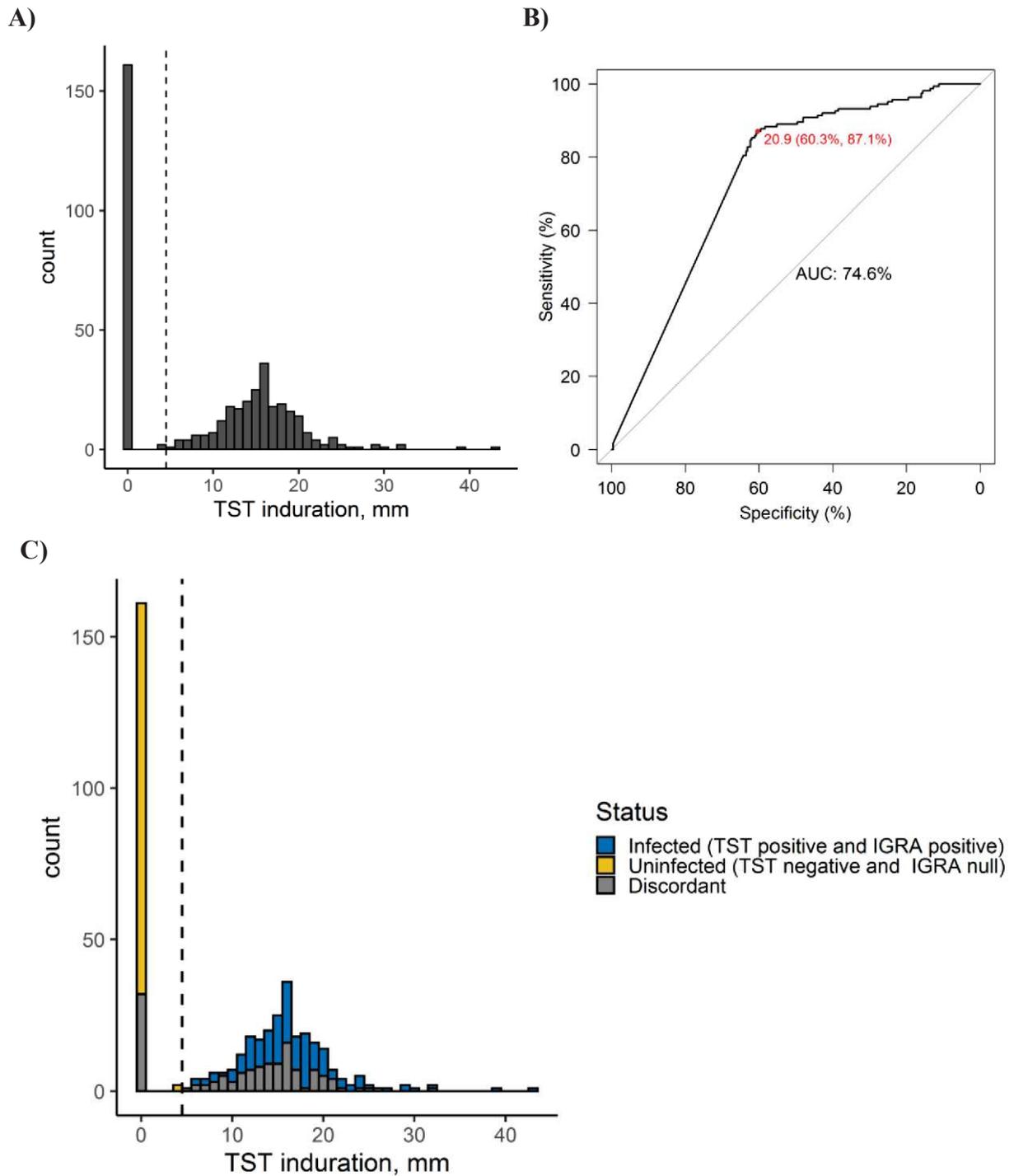
**S12 Figure. Principal component analysis of the South African cohort.** Plot of the first and second principal components of 374 individuals from the South African cohort after projection on the 1000 Genomes phase 3 populations.



**S13 Figure. Stacked histogram of the tuberculin skin test (TST) distribution, stratified by our infection definition among 1108 household contacts in Vietnam.** The dashed line represents a 5 mm cut-off for TST. Uninfected subjects (in yellow) were negative for TST and QuantiFERON-TB Gold In-Tube (QFT-GIT) and infected subjects (in blue) positive for both.



**S14 Figure. Definition of the GWAS phenotype in the household contacts (HHCs) study in France. A)** Distribution of the tuberculin skin test (TST) induration among 516 HHCs (mean induration when 2 screening visits were done). The dashed line represents a 5 mm cut-off. TST was considered negative when the skin induration was i)  $< 5$  mm at both V1 and V2, ii)  $< 5$  mm at V1, when only one visit was done. TST was considered positive when the skin induration was i)  $\geq 5$  mm at both V1 and V2, ii)  $< 5$  mm at V1 and  $\geq 10$  mm at V2. **B)** Construction of a ROC curve based on the defined TST status to determine the optimal interferon- $\gamma$  release assay (IGRA) cut-off (175 pg/mL). **C)** Stacked histogram of the TST distribution stratified by our infection definition. Uninfected subjects (in yellow) presented a negative TST and a null production of IFN- $\gamma$ . Infected subjects (in blue) presented a positive TST and a positive IGRA (IFN- $\gamma$  production  $> 175$  pg/mL).



**S15 Figure. Definition of the GWAS phenotype in the family-based study in South Africa.**

**A)** Distribution of the tuberculin skin test (TST) induration among 415 participants. The dashed line represents a 5 mm cut-off. **B)** Construction of a ROC curve based on the TST status to determine the optimal interferon- $\gamma$  release assay (IGRA) cut-off (20.9 pg/mL). **C)** Stacked histogram of the TST distribution stratified by our infection definition. Uninfected subjects (in yellow) presented a negative TST and a null production of IFN- $\gamma$ . Infected subjects (in blue) presented a positive TST and a positive IGRA (IFN- $\gamma$  production > 20.9 pg/mL).