

Université de Paris

Ecole doctorale 393 - Santé publique :
épidémiologie et sciences de l'information biomédicale

Laboratoire de génétique humaine des maladies infectieuses INSERM U1163

Recherche de mutations pathogéniques par analyses d'exomes
de larges échantillons de patients : application à la tuberculose

Par Gaspard Kerner

Thèse de doctorat en Epidémiologie Génétique
Dirigée par Monsieur Laurent Abel

Présentée et soutenue publiquement le 10 janvier 2020

Devant un jury composé de :

Jacques Fellay, Pr., Ecole Polytechnique Fédérale de Lausanne, rapporteur

Erwin Schurr, Pr., McGill University, rapporteur

Lluis Quintana-Murci, Pr., Collège de France, examinateur

France Mentré, Pr., Université de Paris, Président

Laurent Abel, DR, Université de Paris, directeur de thèse



Except where otherwise noted, this is work licensed under <https://creativecommons.org/licenses/by-nc-nd/3.0/fr/>

Titre : Recherche de mutations pathogéniques par analyses d'exomes de larges échantillons de patients : application à la tuberculose

Résumé : Le cadre général de mon projet se situe dans la recherche de gènes contribuant à la prédisposition ou à la résistance à la tuberculose en se servant des exomes non seulement de patients tuberculeux mais aussi de ceux de patients MSMD. Il s'agit en particulier de comprendre pourquoi, lorsque des individus sont infectés par *M. tuberculosis*, certains (~10%) vont développer la maladie clinique (qui peut être plus ou moins sévère) alors que d'autres (~90%) ne présentent aucune symptomatologie (infection latente). Nous décrivons dans cette thèse la découverte d'une nouvelle étiologie génétique du MSMD au sein du gène *IFNG*, gouvernant le seul axe immunologique de défense anti-mycobactérienne connu. Nous rapportons aussi la première étiologie monogénique relativement commune de la tuberculose. Cette découverte est un des premiers exemples faisant le lien entre étiologies monogéniques à effets forts et variants communs obtenus par analyses GWAS. Compte tenu de l'histoire de la tuberculose, ce variant aurait pu être responsable à lui tout seul de la mort de 10 millions d'individus au cours des 2,000 dernières années. Finalement, nous étendons l'horizon d'étude de la tuberculose par la recherche d'étiologies digéniques à partir du développement d'une nouvelle méthode statistique. Elle est appropriée à l'étude de maladies dites monogéniques qui ne sont pourtant pas complètement expliquées par une seule lésion monogénique. La robustesse et la puissance de cette approche à identifier des interactions entre deux régions génétiques indépendantes ont été démontrées et réalisées tant sur des données simulées comme sur des données réelles (craniosynostose).

Mots clefs : Tuberculose, exome, analyses d'associations, monogénique, digénique, MSMD, pénétrance incomplète, cas-seulement

Title : Investigation of pathogenic mutations by whole exome sequencing analyses of large cohorts of patients : application to tuberculosis

Abstract : The thesis aimed at finding new genetic etiologies of tuberculosis disease using whole exome sequencing data of tuberculosis patients as well as MSMD patients. Since the beginning of the XXth century researchers have tried to understand why infected individuals would (~10%) in some cases develop diseases while others (~90%) would remain asymptomatic. We herein describe the first genetic etiology of MSMD in the *IFNG* gene, the core gene controlling host's immune system against mycobacterial infections. We also describe the first common monogenic etiology of tuberculosis in the *TYK2* gene, bridging the gap between monogenic variants with strong effect and more common variants obtained with GWAS. Looking back at the history of TB, it would be estimated that about 10 million people died due to this one *TYK2* mutation. We finally expanded the study of the genetics of tuberculosis by developing a new methodology to study digenic effects on disease susceptibility. It is in particular appropriate to the study of so far assumed monogenic diseases that are not fully explained by a monogenic lesion. Robustness and power of this approach to find interaction between independent genomic regions were shown in both extensive simulated data and real exome data (craniosynostosis).

Keywords : Tuberculosis, exome, association analyses, monogenic, digenic, MSMD, incomplete penetrance, case-only

Sommaire

I	Introduction.....	11
	A. Le défi de la tuberculose maladie.	11
	B. Le scénario évolutif de <i>M. tuberculosis</i> et le genre <i>Mycobacterium</i>	15
	C. La tuberculose, une maladie ancestrale chez l'humain.	19
	D. L'histoire naturelle de la tuberculose	24
	E. Formes cliniques et diagnostic de la tuberculose	27
	F. Prévention et traitement de la tuberculose.	30
	G. Les facteurs de risque de la tuberculose.	34
	H. La génétique humaine de la tuberculose maladie.	37
	H.1. Historique.....	37
	H.2. Etudes en population.	41
	H.3. Le syndrome MSMD et les formes monogéniques de tuberculose.	44
	I. Objectifs de la thèse.	47
II	Dissection des bases génétiques de la tuberculose par analyse d'exomes : approche monogénique.	50
	A. La cohorte d'étude: HGID.....	50
	A.1. Séquençage des individus, leur contrôle qualité et celui des variants.	51
	A.2. Annotation des variants.	52
	A.3. Analyse de la cohorte HGID en composantes principales.....	53
	A.4. Etudes d'association par analyse d'exome : l'exemple de la pneumopathie grippale sévère.	54
	B. L'étude spécifique d'une famille consanguine d'origine libanaise par une approche combinant analyse de liaison et analyse d'exome.	57
	B.1. Echantillon d'étude.	58
	B.2. Méthode statistique d'analyse.....	59
	B.3. Résultats.	60
	B.4. Conclusion.	68
	C. Homozygotie d'un variant fréquent faux-sens du gène <i>TYK2</i> : première étiologie génétique relativement fréquente de la tuberculose.	68
	C.1. Analyse de la cohorte HGID.....	69
	C.2. Analyse de la cohorte UK Biobank.	76
	C.3. Discussion.....	84
III	Une approche statistique pour la détection d'une hérédité digénique à l'échelle du génome entier à partir de données WES.	88

A. Introduction.	88
B. Une méthode statistique à l'échelle du génome entier pour la recherche de DI.....	91
B.1. Modèle d'agrégation de variants rares.	92
B.2. Modèle cas-témoin pour une analyse monogénique	93
B.3. Modèle cas-témoin pour le test d'interaction	93
B. 4. Modèle <i>case-only</i> pour le test d'interaction.	95
C. Echantillons d'étude.....	95
D. Contrôle qualité des variants.....	97
E. Analyses de simulation.....	98
E.1. Analyse de l'erreur de type I.	98
E.2. Etude de puissance.....	105
E.3. Analyse de données réelles	112
F. Conclusions et perspectives de l'approche digénique.	116
IV Discussion générale sur le travail de thèse.....	119
V Bibliographie	125
VI Annexe : Articles issus du travail de thèse	145

Liste des figures

Figure 1: Incidence de la tuberculose dans le monde en 2012 d'après (12)	14
Figure 2: Courbes d'incidence et de mortalité estimées nécessaires pour atteindre l'objectif de l'OMS face à la TB, 2015–2035	15
Figure 3: Lignées de MTBC (16)	17
Figure 4: La répartition mycobactérienne du MTBC selon les lignées issus de (18)	17
Figure 5: La tuberculose a tué plus que toute autre maladie infectieuse dans l'Histoire. Plus d'un milliard de personnes au cours des deux milles dernières années	24
Figure 6: Principales étapes de l'infection par <i>M. tuberculosis</i> d'après (44).....	25
Figure 7: La réponse cellulaire à l'infection par <i>M. tuberculosis</i> d'après (42).....	27
Figure 8: Distribution des taux de mortalité par tuberculose disséminée (en bleu) et par tuberculose pulmonaire (en rouge) pour 100,000 personnes non soignées vivant en Bavière en 1905, avant l'apparition du vaccin BCG, région endémique pour la tuberculose (51).....	29
Figure 9: Estimation de l'incidence des cas de tuberculose MDR ou résistant seulement à la rifampicine en 2017, pour les pays comportant au moins 1000 cas incidents	32
Figure 10: Pays ayant utilisé la bédaquiline pour le traitement de la tuberculose XDR dans le cadre d'un accès élargi, dans des conditions de programmation normales pendant l'année 2017	34
Figure 11: Schéma pour un système moderne de prévention et traitement de la tuberculose issu de (53)	34
Figure 12: Schéma général du passage de l'exposition à l'infection et de l'infection au développement d'une tuberculose active en fonction des facteurs associés.....	35
Figure 13: Taux de morbidité (en %) liée à la tuberculose chez les apparentés de cas index d'après l'étude de Kallmann & Reisner	40
Figure 14: Schéma de la coopération entre les phagocytes et les lymphocytes T (ou NK) lors d'une infection mycobactérienne	45
Figure 15: PCA des 2,504 individus "sains" de 1000G d'origines ancestrales connues (en fonction des cinq populations de 1000G, Africaine (AFR), Américaine (AMR), Asiatique de l'Est (EAS), Européenne (EUR), Sud Asiatique (SAS)), et des patients avec (n=917, en rouge) ou sans (n=2,835, en gris) des infections par mycobactéries de la cohorte HGID.....	54
Figure 16: Arbre généalogique de la famille consanguine montrant la ségrégation de l'allèle mutant c.354_357del (p.T119Ifs*4) du gène <i>IFNG</i>	59
Figure 17: Résultat graphique des régions chromosomiques obtenues à partir de l'analyse de liaison	61
Figure 18: Analyse d'exome qui a permis d'identifier un variant rare homozygote d' <i>IFNG</i> dans une des régions de liaison génétique.....	62
Figure 19: Analyse par western blot de l'IFN- γ	63
Figure 20: Distribution à l'échelle du génome entier de l'intensité de la sélection évolutive agissant sur les gènes <i>IFNG</i> , <i>IFNGR1</i> et <i>IFNGR2</i> par le paramètre <i>f</i>	65
Figure 21: Distribution à l'échelle du génome entier de l'intensité de la sélection évolutive agissant sur les gènes <i>IFNG</i> , <i>IFNGR1</i> et <i>IFNGR2</i> par le paramètre GDI.....	66
Figure 22: PCA des 2,504 individus "sains" de 1000G et des individus de la cohorte HGID	71

Figure 23: Distribution des familles des dix porteurs homozygotes du variant TYK2 P1104A ayant une tuberculose ou le syndrome MSMD	72
Figure 24: Carte de fréquences de l'allèle P1104A autour du globe	73
Figure 25: Distribution des ORs obtenus lors du test d'association de la tuberculose avec tous les variants codants fréquents faux-sens ou potentiellement perte de fonction	73
Figure 26: Analyse des taux d'IFN- γ en sang total pour les patients homozygotes P1104A en comparaison à des témoins.....	74
Figure 27: Distributions des fréquences alléliques actuelles des variants qui ont ségrégués il y a 4,000 ans à des fréquences similaires à celles des variants P1104A et I684S TYK2, M694V MEFV et C282Y HFE.....	75
Figure 28: Année de survenue de la tuberculose pour tous les patients du champs #22137 de la base de donnée UK Biobank	78
Figure 29: PCA des 115,093 participants de la sous-cohorte UK Biobank utilisée pour notre analyse et des 2,504 individus non apparentés de 1000G	80
Figure 30: Estimation de l'OR de développer une tuberculose après infection chez les homozygotes P1104A	83
Figure 31 : Diagramme Manhattan d'une analyse d'association pan-génomique sur le phénotype « tuberculose diagnostiquée par un médecin » de la ressource UK Biobank.....	86
Figure 32: PCA des individus Européens utilisés dans l'étude de simulation de puissance	97
Figure 33: QQ-plots pour le test case-only conduit sur la population IBS + TSI, en fonction de la distance entre les deux gènes de chaque paire.....	101
Figure 34: QQ-plot pour le test case-only mené à l'échelle du génome entier sur la population IBS + TSI avec (A) > 5% de porteurs de variants rares pour le premier et le second gène de chaque paire ou bien (B) >1% de porteurs de variants rares pour le premier et >15% pour le deuxième	104
Figure 35: Puissance des tests case-only et cas-témoins pour l'analyse de toutes les paires de gènes (schéma A)	108
Figure 36: Puissance des tests case-only et cas-témoins pour l'analyse de deux paires de gènes spécifiques en absence (schéma 2G) ou en présence (schéma 2GS) de stratification de la population.....	111
Figure 37: Puissance des tests case-only et cas-témoins pour l'analyse d'une paire de gènes avec différentes proportions de porteurs de variants rares (schéma 2GR).....	112
Figure 38: QQ-plot pour le test de case-only à l'échelle du génome mené sur les 191 patients atteints de craniosynostose	114

Liste des tableaux

Tableau 1: Nombres d'individus appartenant aux principaux groupes phénotypiques de la cohorte HGID	51
Tableau 2: Classement des cinq gènes les plus enrichis en variants rares liés à <i>IRF7</i> dans la cohorte ARDS par rapport à 1,485 témoins sans infections virales.....	57
Tableau 3: Distribution des variants dans les régions codantes et non codantes (couvertes par WES) de l'IFNG et de ses deux récepteurs	67
Tableau 4: Synthèse des porteurs des variants TYK2 étudiés dans la cohorte HGID	72
Tableau 5: Distribution des variants TYK2 P1104A, I684S et V362F parmi cas et témoins et estimation des ORs pour les modes génétiques de transmission dominant et récessif	81
Tableau 7: Auto-identification des 115,093 individus génotypés au locus TYK2 P1104A de la cohorte UK Biobank en termes d'origines ethniques, pour lesquels une information sur le champ tuberculose « diagnostiquée par un médecin » a été notée.....	81
Tableau 8: Auto-identification des 234 individus homozygotes pour le variant TYK2 P1104A de la cohorte UK Biobank en termes d'origines ethniques, pour lesquels une information sur le champ tuberculose « diagnostiquée par un médecin » a été notée.....	81
Tableau 9: Tableau de contingence des porteurs de variants dans Sk ou Sj pour une paire donnée de gènes k et j et pour des individus atteints ou non atteints.....	94
Tableau 10: Erreurs de type I empiriques à une valeur nominale $\alpha = 0.1\%$ pour le test case-only en fonction de la distance minimale entre les deux gènes d'une paire	100
Tableau 11: Erreurs de type I empiriques à une valeur nominale $\alpha = 0.1\%$ pour les tests case-only et cas-témoins en l'absence de stratification de population.....	102
Tableau 12: Erreurs de type I empiriques à une valeur nominale $\alpha = 0.1\%$ pour les tests case-only et cas-témoins en l'absence de stratification de population.....	103
Tableau 13: Erreurs de type I empiriques à une valeur nominale $\alpha = 0.1\%$ pour les tests case-only et cas-témoins en présence de PS.....	105
Tableau 14: Valeurs de pénétrance des combinaison digéniques G1 vs G2 lors de l'analyse de deux paires de gènes spécifiques, l'une avec des proportions de porteurs de variants rares similaires pour les deux gènes qui la composent (schéma 2G) et l'autre avec des proportions très différentes (schéma 2GR).....	106
Tableau 15: Description des schémas utilisés dans l'étude de simulation de la section « Puissance »	107
Tableau 16: Proportion (en %) de porteurs de variants rares pour les gènes testés dans la section III_E.2, pour la population et les sous-populations Européennes, telles que définies en III_C	109
Tableau 17: Tableau de contingence pour la combinaison digénique <i>rs1884302 (BMP2) – SMAD6</i> basée sur les 191 proposants de l'étude sur la craniosynostose.....	114
Tableau 18: Tableau de contingence pour la combinaison digénique Q613P (<i>MST1R</i>) – <i>IL12RB1</i>	116

Liste de publications

Liste de publications issues directement du travail de thèse :

Boisson-Dupuis S, Noe Ramirez-Alejo, Zhi Li*, Etienne Patin*, Geetha Rao*, Gaspard Kerner* et al. (2018) Tuberculosis and impaired IL-23-dependent IFN- γ immunity in humans homozygous for a common TYK2 missense variant. *Sci Immunol* 3(30). doi:10.1126/sciimmunol.aau8714.

Kerner, G., Ramirez-Alejo, N., Seeleuthner, Y., Yang, R., Ogishi, M., Cobat, A., Patin, E., Quintana-Murci, L., Boisson-Dupuis, S., Casanova, J.-L., et al. (2019). Homozygosity for TYK2 P1104A underlies tuberculosis in about 1% of patients in a cohort of European ancestry. *PNAS* 201903561.

Lim, H.K.*, Huang, S.X.L.*, Chen, J.*, **Kerner, G.***, Gilliaux, O., Bastard, P., Dobbs, K., Hernandez, N., Goudin, N., Hasek, M.L., et al. (2019). Severe influenza pneumonitis in children with inherited TLR3 deficiency. *Journal of Experimental Medicine* 216, 2038–2056.

Gaspard Kerner, Matthieu Bouaziz, Aurélie Cobat, Benedetta Bigio, Andrew T Timberlake, Jacinta Bustamante, Richard P Lifton, Jean-Laurent Casanova and Laurent Abel. A genome-wide case-only test for the detection of digenic inheritance in human exomes. Submitted.

Gaspard Kerner*, Jérémie Rosain*, Antoine Guérin*, Waleed Al-Herz*, Carmen Oleaga-Quintas, Franck Rappaport, Michel J. Massaad, Ahmed Al-Khabaz, Jing-Ya Ding, Fatima Al Ali. Inherited IFN- γ deficiency in humans with mycobacterial disease. Submitted.

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

Guérin, A., **Kerner, G.**, Marr, N., Markle, J.G., Fenollar, F., Wong, N., Boughorbel, S., Avery, D.T., Ma, C.S., Bougarn, S., et al. (2018). IRF4 haploinsufficiency in a family with Whipple's disease. *Elife* 7,.

Martínez-Barricarte, R., Markle, J.G., Ma, C.S., Deenick, E.K., Ramírez-Alejo, N., Mele, F., Latorre, D., Mahdaviyani, S.A., Aytakin, C., Mansouri, D., et al. (2018). Human IFN- γ immunity to mycobacteria is governed by both IL-12 and IL-23. *Sci Immunol* 3,.

Ogunjimi, B., Zhang, S.-Y., Sørensen, K.B., Skipper, K.A., Carter-Timofté, M., **Kerner, G.**, Luecke, S., Prabakaran, T., Cai, Y., Meester, J., et al. (2017). Inborn errors in RNA polymerase III underlie severe varicella zoster virus infections. *J. Clin. Invest.* 127, 3543–3556.

Rosain, J., Oleaga-Quintas, C., Deswarte, C., Verdin, H., Marot, S., Syridou, G., Mansouri, M., Mahdaviyani, S.A., Venegas-Montoya, E., Tsolia, M., et al. (2018). A Variety of Alu-Mediated Copy Number Variations Can Underlie IL-12R β 1 Deficiency. *J. Clin. Immunol.*

Wang, Y., Ma, C.S., Ling, Y., Bousfiha, A., Camcioglu, Y., Jacquot, S., Payne, K., Crestani, E., Roncagalli, R., Belkadi, A., et al. (2016). Dual T cell- and B cell-intrinsic deficiency in humans with biallelic RLTPR mutations. *J. Exp. Med.* 213, 2413–2435.

* co-paternité

Introduction

A. Le défi de la tuberculose maladie.

“The struggle [against tuberculosis] has caught hold along the whole line and enthusiasm for the lofty aim runs so high that a slackening is no longer to be feared. If work goes on in this powerful way, then the victory must be won.”

Robert Koch’s 1905 Nobel Prize speech

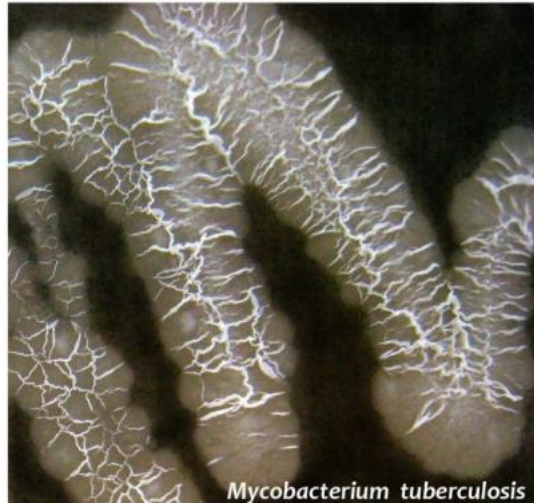
Les êtres humains sont sujets à une multitude de maladies infectieuses endémiques et épidémiques (1) qui furent leurs principales causes de décès au cours de l’Histoire. Le fardeau de celles-ci était trop important et l’espérance de vie était en moyenne de 25 ans jusqu’à la révolution industrielle (2). Même en Europe, seulement 35% des personnes atteignaient l’âge de 40 ans à la fin du XIX^e siècle (3). En Angleterre, durant la même période, 60% des décès étaient dus à des maladies infectieuses, en particulier lors d’épidémies. L’augmentation de l’espérance de vie, qui atteint aujourd’hui 80 ans dans ces régions, est récente et contribue largement à l’explosion démographique actuelle. Elle reflète principalement les progrès accomplis dans la lutte contre les maladies infectieuses sur la base de trois facteurs: le développement de l’hygiène à partir du milieu du XIX^e siècle (prévention de la transmission de l’infection); la mise au point de vaccins à partir du début du XX^e siècle (prévention de la maladie chez les individus infectés); et la mise au point de médicaments anti-infectieux à partir de la première moitié du XX^e siècle (prévention de la mort chez les patients atteints d’une maladie clinique) (3).

Cependant, les maladies infectieuses restent de nos jours un enjeu majeur de santé publique. Elles sont la deuxième cause de mortalité et de morbidité dans le monde mais avec des degrés différents selon le niveau de développement des pays atteints. L’épidémie de fièvre hémorragique Ebola qui sévit en 2014 dans quatre pays de l’Afrique de l’Ouest est un exemple récent et médiatique de maladie infectieuse, qui sans traitement depuis sa découverte en 1976, a causé la mort de plus de 11,000 individus au cours de deux ans de croissance de l’épidémie (OMS, *Ebola Situation Report, 10 June 2016*, <https://www.who.int/csr/disease/ebola/en/>). Si les pays du Nord semblaient plus à l’abri d’une épidémie dévastatrice ils ont cependant été témoin il y a quelques années seulement de la diffusion mondiale de l’infection à VIH et d’autres infections virales chroniques comme l’hépatite B et C ou encore de la pandémie H1N1 en 2009. Il en est de même pour les infections dites « émergentes » ou « ré-émergentes » dont certaines sont liées à l’introduction de nouveaux vecteurs (par exemple, l’épidémie de maladie à virus zika, imputable en partie à une plus grande densité des *Aedes* vecteurs, particulièrement en milieu

urbain) et d'autres à l'apparition de souches résistantes, comme c'est le cas des souches tuberculeuses multi-résistantes. Aujourd'hui, en Europe, la tuberculose maladie est plus souvent évoquée dans les livres d'Histoire que dans les discours politiques. Cependant, elle est toujours responsable de plus de 10 millions de nouveaux cas par an (source OMS 2017).



Robert Koch



L'agent pathogène de la tuberculose, *Mycobacterium tuberculosis* (*M. tuberculosis*), appelé communément bacille de Koch du nom de son découvreur allemand décédé il y a plus de cent ans, infecte toujours en ce début du XXI^e siècle, environ un quart de l'humanité. Elle est aujourd'hui la première cause de mortalité due à une maladie infectieuse parmi les adultes en comptant plus de 1.3 millions de morts par an au total. Il y eut un temps, pas si lointain, où l'éradication de la tuberculose fut envisagée, supposée et même annoncée, en particulier, lors de la découverte des antibiotiques antituberculeux après la deuxième guerre mondiale. René et Jean Dubos, auteurs du livre *The White Plague* (1952), avaient cependant déjà prévenu au beau milieu du XX^e siècle contre « la logique scientifique et sociale qui prétend soigner la maladie plutôt que de modifier les causes sous-jacentes de la vulnérabilité » (4). Ceci dit, il y a cent ans, alors que la France comptait 40 millions d'habitants, plus de 100,000 personnes mourraient chaque année de la tuberculose. Aujourd'hui, le nombre de morts y est réduit par un facteur 300, essentiellement grâce à la découverte des antibiotiques, marquant un tournant dans la progression de la maladie chez les atteints. Elles ne peuvent cependant pas changer l'évolution des cycles d'épidémie qui dépend en partie des contributions politiques et institutionnelles inter- et intra-pays pour prévenir les risques d'infection. Des raisons plus complexes peuvent être à l'origine de ces cycles épidémiques. Dans leur livre *The White Plague*, les auteurs ont estimé que la prévalence de l'infection par la tuberculose avait

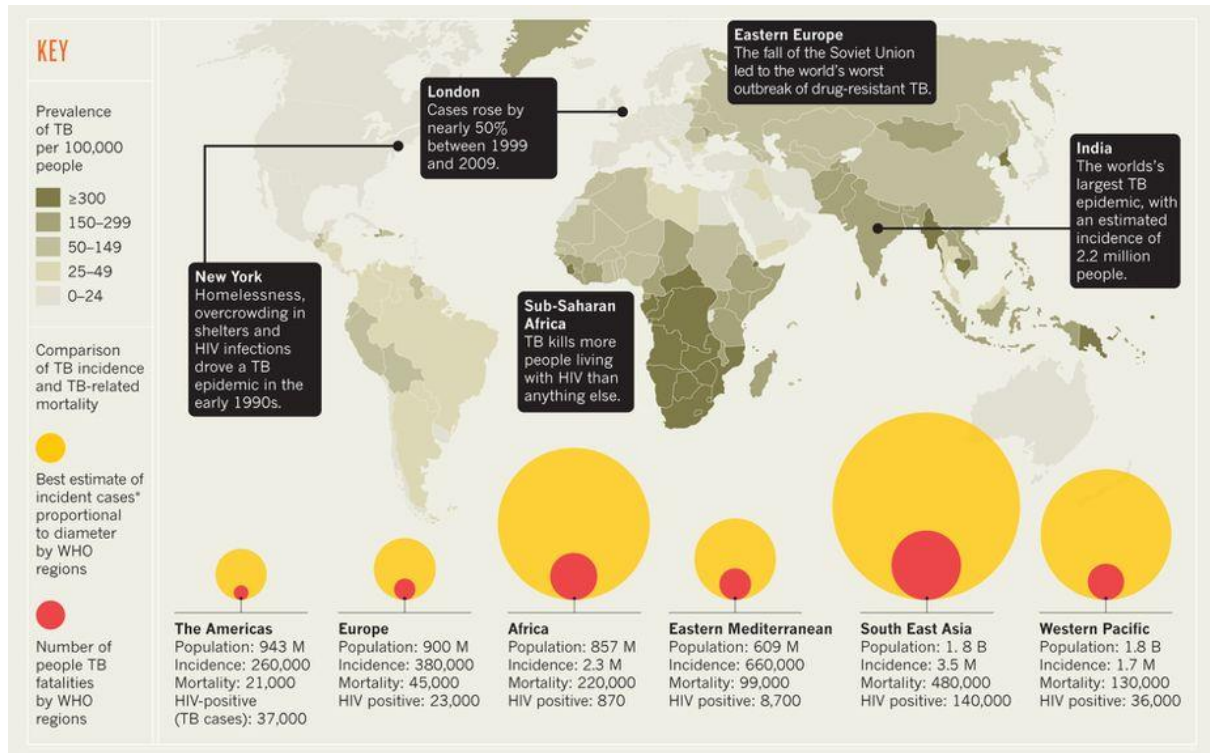
diminuée d'un facteur 2 dans la deuxième moitié du XIX^e siècle, avant même la découverte de l'agent pathogène (4), postulant l'existence d'un cycle épidémique indépendant des traitements antituberculeux.

Depuis les années 1980 une recrudescence de la tuberculose a été enregistrée dans de nombreuses régions du monde (Figure 1), y compris les pays industrialisés, imputable à divers facteurs, notamment la réduction du temps et de l'argent dédiés aux programmes de lutte antituberculeuse, l'augmentation de la migration, la propagation du VIH, l'augmentation des souches multi-résistantes, l'augmentation du nombre de certaines maladies non transmissibles qui sont des facteurs de risque de la tuberculose, en particulier le diabète de type 2 associé à la pandémie mondiale d'obésité, ainsi que les crises économiques et l'inégalité croissante des richesses et de l'accès à la santé (5). Ceci a conduit dans les années 1990 à une prise de conscience de la part de l'OMS, qui, depuis les 25 dernières années, considère la tuberculose comme une urgence mondiale de santé publique. Une étiquette similaire a été associée à l'épidémie de fièvre hémorragique Ebola en 2014, attirant cependant un plus grand public (ainsi que les efforts internationaux pour son éradication) que la tuberculose malgré ses 1.5 million de morts durant la période où l'épidémie Ebola fut responsable de 11,310 décès (6).

Koch a résumé l'essentiel de son travail sur la tuberculose, pour lequel il a reçu le prix Nobel de 1905, dans un manuscrit qu'il publia dans le *Berliner Klinische Wochenschrift* peu de temps après l'annonce de la découverte du bacille le 24 mars 1882 à la Société physiologique de Berlin: « À l'avenir, la lutte contre ce terrible fléau de l'humanité ne portera plus sur un fait indéterminé, mais sur un parasite tangible, dont les conditions de vie sont pour la plupart connues et sur lesquelles on peut aller plus vers l'avant » (7). Mais ce ne fut que 39 ans plus tard que le premier traitement préventif contre la tuberculose fut mis au point par Calmette et Guérin, le vaccin du BCG. En 1943, la streptomycine, le premier antibiotique antituberculeux efficace, fut isolé dans le laboratoire de Selman Waksman de l'Université Rutgers. En 1951, l'hydrazide d'acide isonicotinique (isoniazide) fut testé à l'hôpital Sea View de New York. Il améliora considérablement les résultats cliniques et fut bientôt introduit pour une utilisation plus large. Il fut suivi par le développement de la pyrazinamide (1952), de la cycloserine (1952), de l'éthionamide (1956), de la rifampicine (1957) et de l'éthambutol (1962). Grâce à son haut niveau d'efficacité et à la facilité de son administration, la rifampicine révolutionna le traitement de la tuberculose (8,9). Cependant, l'usage répandu de ces antibiotiques conduisirent à la sélection de mutations lui conférant une résistance. Entre 1970 et 1990, se déclenchèrent de nombreuses épidémies de tuberculose pharmaco-résistante impliquant des souches résistantes à deux médicaments ou plus (10). Ces complexités, qui

vont maintenant de la tuberculose pan-résistante à la maladie pédiatrique non diagnostiquée, représentent une fraction non négligeable des 10 millions de nouveaux cas et des 1.3 million de décès dus à la tuberculose chaque année dans le monde (11).

Figure 1: Incidence de la tuberculose dans le monde en 2012 d'après (12)

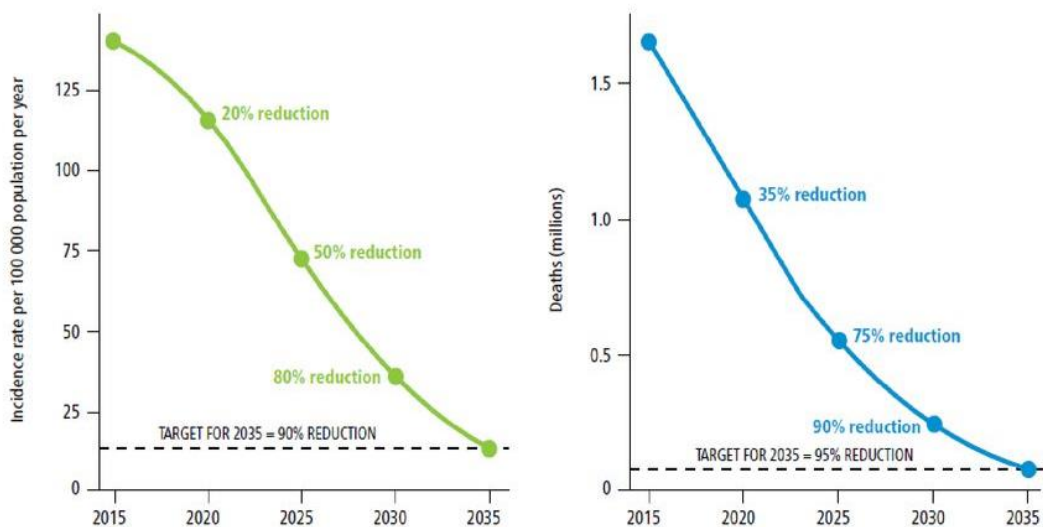


WORLD TUBERCULOSIS REPORT 2012

L'humanité a probablement disposé vers la deuxième partie du XX^e siècle des antibiotiques nécessaires pour éradiquer la tuberculose, mais n'a pas toujours su s'en servir correctement. Elle a vu émerger des formes multi-résistantes et ultra-résistantes et a été confrontée à une recrudescence de la tuberculose dans le monde industrialisé. Dans les pays où la tuberculose est très endémique, la tuberculose dite primaire (survenant dans les deux ans qui suivent l'infection initiale) est particulièrement fréquente chez les enfants, qui développent souvent une maladie potentiellement mortelle (13,14). Selon le rapport annuel de l'OMS (https://www.who.int/tb/publications/global_report/en/), à l'échelle mondiale, la tuberculose est l'une des 10 principales causes de décès et la principale cause d'un seul agent infectieux (au-dessus du VIH / sida). Elle est probablement par ailleurs la maladie infectieuse la plus mortelle dans l'histoire de l'humanité (12,15). Si les agents politiques et les aides humanitaires au niveau mondial sont des acteurs essentiels dans le contrôle de la tuberculose pour garantir l'objectif de l'OMS pour 2030, à savoir, la réduction de 90% du nombre absolu

de décès dus à la tuberculose et de 80% de l'incidence de la maladie (nouveaux cas pour 100,000 habitants par an) par rapport au niveau de 2015, jusqu'ici les mesures de contrôle n'ont pas été suffisantes pour éviter qu'un quart de la population mondiale soit infecté (Figure 2). Les cas et les décès continuent à se produire. Empêcher tout contact avec les bacilles tuberculeux semble relever de l'utopie. Il sera donc important de déployer d'autres lignes d'attaques, par exemple, sur la base d'une connaissance plus affûtée de sa physiopathologie, permettant, entre autre, de comprendre pourquoi certaines personnes développent la maladie tandis que d'autres ne le font pas.

Figure 2: Courbes d'incidence et de mortalité estimées nécessaires pour atteindre l'objectif de l'OMS face à la TB, 2015–2035



WORLD TUBERCULOSIS REPORT 2018

B. Le scénario évolutif de *M. tuberculosis* et le genre *Mycobacterium*.

La tuberculose a harcelé l'humanité au cours des siècles. Elle a, par ailleurs, probablement accompagné les *Homo sapiens* modernes hors d'Afrique (16). L'histoire évolutive de l'hôte est étroitement associée à celle de son agent pathogène et vice-versa. La tuberculose ne peut être donc pleinement comprise qu'à la lumière des origines, des migrations et de la démographie d'*Homo sapiens*. Quand en 1882 Robert Koch annonce la découverte de l'agent pathogène de la tuberculose, il présente une nouvelle mycobactérie au monde, qui plus tard, avec les progrès de la biologie moléculaire, permettant l'étude de l'ADN bactérien et éclaircissant les structures génétiques à petite échelle, pris sa place dans un grand complexe mycobactérien dit complexe *M. tuberculosis* (MTBC). Le complexe inclut *M. africanum*, *M. bovis*, *M. caprae*, *M. microti*, *M. pinnipedii*, *M. orygis* ainsi que d'autres

souches telles que les souches BCG (issues du vaccin BCG). L'homogénéité de ce complexe mycobactérien, avec des espèces partageant 99.9% de leur génome, en particulier les ARNs ribosomiaux 16S qui sont intégralement conservés (17), a conduit à la génération de l'hypothèse que les différents membres du complexe MTBC étaient des descendants d'un même ancêtre commun. Aujourd'hui, les mycobactéries qui le composent se différencient, par exemple, par leurs tropismes. *M. tuberculosis* est le principal agent infectieux de l'être humain dans le monde tandis que *M. africanum* ne se retrouve chez l'espèce humaine qu'en Afrique de l'Ouest. *M. caprae* et *M. orygis* ont un tropisme exclusivement animal, n'infectant que des moutons et des chèvres en Europe. *M. microti* apparaît chez des petits rongeurs et *M. pinnipedii* chez des otaries et des Arctocephalinae en Australie, Nouvelle Zélande et Amérique du Sud. Finalement, *M. bovis* peut infecter une palette particulièrement large d'espèces hôtes (animaux sauvages, bétail, primates), et étant transmissible à l'humain, les animaux infectés par *M. bovis* deviennent alors un réservoir de bacilles et un problème de santé publique.

Toutes les souches du complexe MTBC semblent agir comme une espèce mycobactérienne génétiquement unique avec des écotypes qui se sont adaptés à leur hôte, laissant cependant possible la transmission inter-espèces. Sept lignées majeures sont reconnues dans le MTBC (Figure 3) avec trois lignées dites modernes (présentant la délétion TbD1) et plus proches génétiquement (lignées 2,3, et 4) et quatre plus anciennes (1, 5, 6 et 7 sans la délétion TbD1) (16). Comme le montre la Figure 4, les lignées modernes ainsi que les lignées 1 et 7 n'incluent que *M. tuberculosis sensu strictu* tandis que les lignées 5 et 6 n'incluent que *M. africanum*. Les autres mycobactéries du complexe, incluant les différentes souches BCG, sont considérées comme des souches adaptées aux animaux et ne font pas partie des sept lignées décrites. La Figure 4 indique par ailleurs un représentant animal pour chaque souche ainsi que les délétions génétiques de la mycobactérie qui caractérisent les différentes souches ou lignées.

Figure 3: Lignées de MTBC (16)

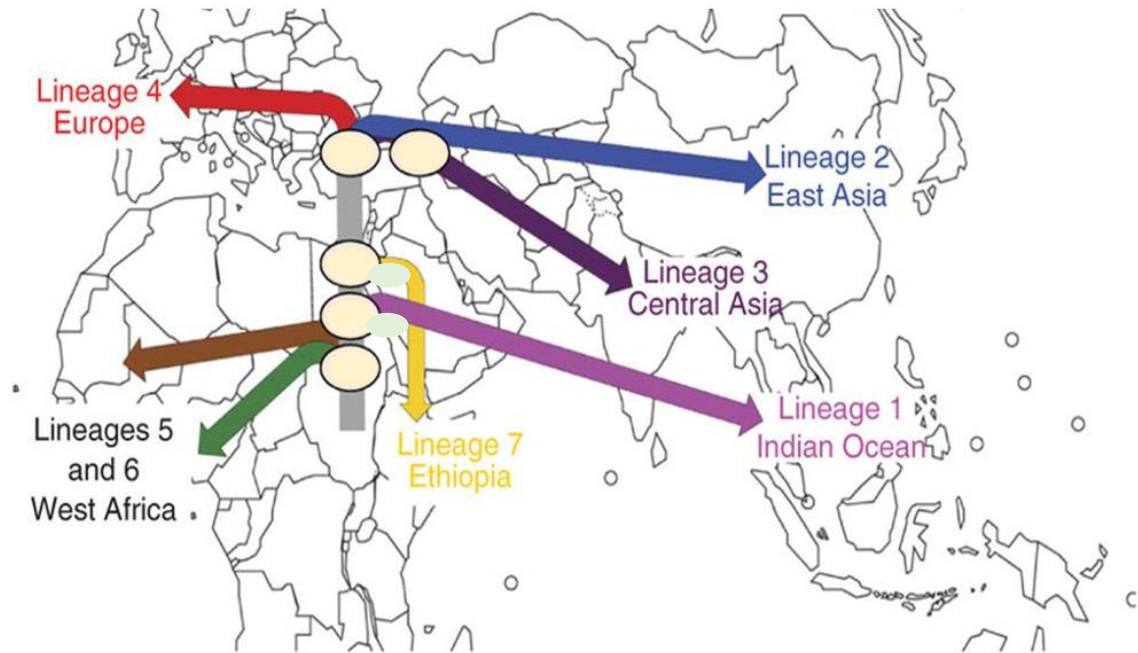
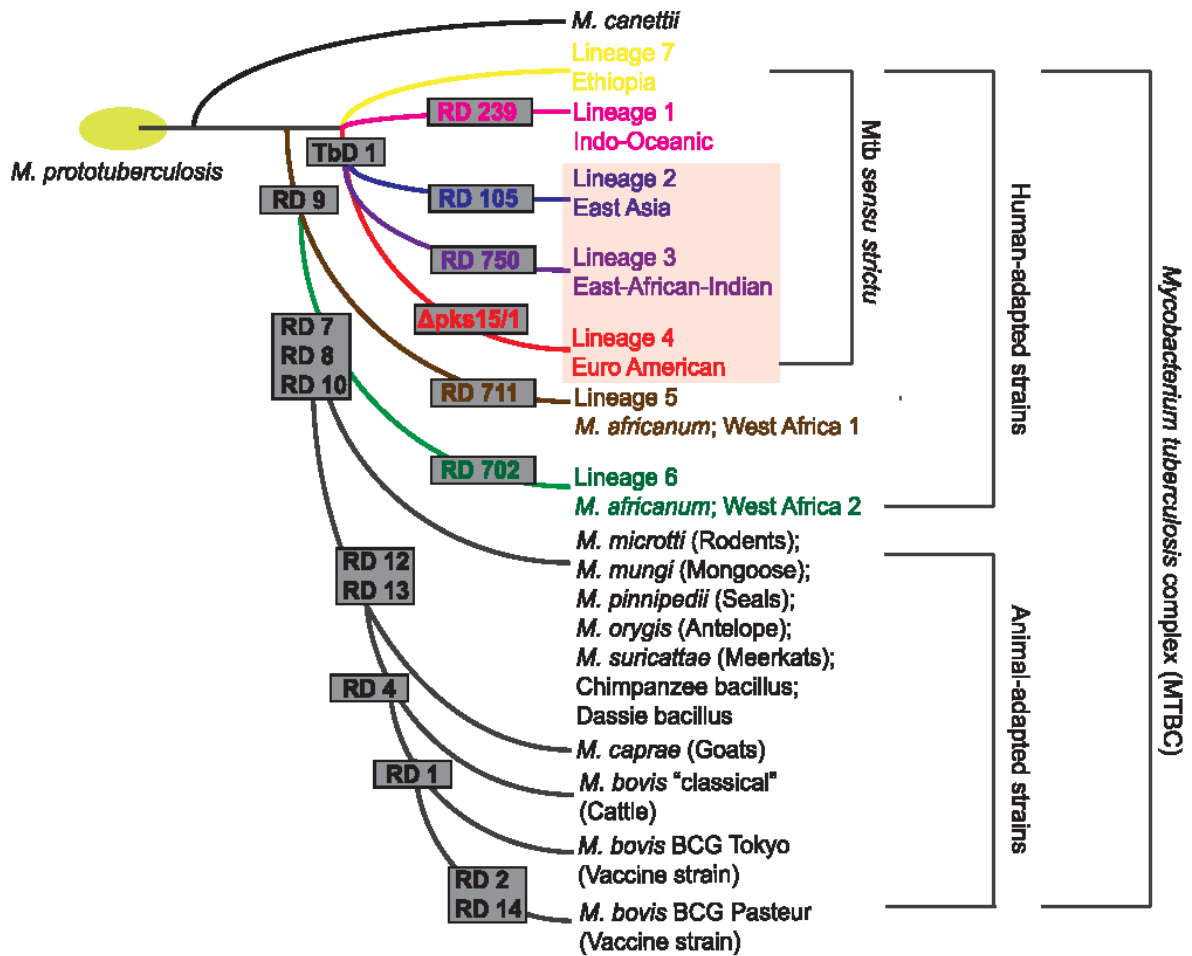


Figure 4: La répartition mycobactérienne du MTBC selon les lignées issues de (18)



Les relations observées entre les lignées MTBC sont similaires à celles observées dans l'espèce humaine, ce qui suggère que l'humain aurait pu être porteur de la tuberculose pendant des millénaires et que la répartition géographique actuelle de la tuberculose aurait été façonnée par des migrations humaines anciennes, si ce n'est la première, hors d'Afrique (19). En effet, Comas et ses collègues (16) suggérèrent, à partir d'un modèle de coalescence, que MTBC aurait émergé au moins 70,000 ans auparavant étant sorti d'Afrique avec les premières vagues de migration de notre espèce vers l'est. *M. tuberculosis* aurait commencé à infecter l'humain bien avant la transition démographique néolithique (*neolithic demographic transition*, NDT) comme cela a été souvent proposé (20,21). Elle n'aurait donc pas suivi une transition animal-humain pendant la période Néolithique en Europe ni aurait évolué de *M. bovis* (souche infectant le bétail). A contrario, les humains auraient plutôt transmis la tuberculose au bétail compte tenu du fait que le clade des souches anciennes de MTBC, incluant les souches animales, infectent aussi les humains et que les souches modernes n'infectent que ces derniers.

Les maladies infectieuses sont souvent divisées en deux grandes catégories: celles étant apparues pendant la période de transition démographique néolithique (NDT) (souvent appelées maladies de foule), et celles antérieures à la période NDT (20). La tuberculose se caractérise par une progression chronique, de latence et de réactivation, propre à une maladie ancienne, antérieure à la période NDT, ce qui s'adapte parfaitement à la théorie précédemment mentionnée. Cependant, elle s'assimile aussi aux maladies de foule dans le mode de transmission (forme d'aérosol) et dans sa virulence. En effet, l'analyse des séquences MTBC en (16) suggère que l'augmentation majeure de la taille de la population du complexe MTBC commença il y a environ 10,000 ans. Ces résultats indiquent que la période néolithique a contribué au succès de MTBC, non pas en augmentant la probabilité de transfert zoonotique à l'homme comme proposé précédemment, mais en raison de l'augmentation de la taille et de la densité de la population hôte.

De façon plus générale, *M. tuberculosis* fait partie du genre *Mycobacterium* qui, avec *Corynebacterium* et *Nocardia*, forme un taxon monophylétique au sein de la famille des Actinomycètes (22). Ces bactéries à Gram positif ont des parois cellulaires très cireuses en raison de la présence d'acides mycoliques. Cela les rend acido-alcool-résistantes, une caractéristique qui les distingue spécifiquement des autres bactéries et leur confère des voies biochimiques uniques. Le genre *Mycobacterium* est très diversifié avec 85 espèces différentes identifiées depuis la découverte de *M. leprae* en 1873 (23). Il existe également un nombre

considérable de sous-souches du vaccin BCG, dérivées d'une souche atténuée de *M. bovis* obtenue en 1921 (24).

La grande majorité des espèces de mycobactéries sont des saprophytes vivant en liberté dans l'environnement (25) et tous les êtres humains sont exposés aux mycobactéries environnementales (*environmental mycobacteria*, EM) d'origine hydrique et aérienne, qui entrent fréquemment en contact avec la peau et les muqueuses (principalement les épithéliums digestif et respiratoire). En outre, la vaste majorité (85%) des enfants dans le monde sont vaccinés par voie intradermique ou sous-cutanée avec le vaccin BCG (qui protège contre la tuberculose). Une proportion importante d'individus, principalement dans les régions les plus pauvres du monde, sont exposés à des mycobactéries plus virulentes à tropisme humain et à transmissions inter-humaines, telles que *M. tuberculosis* et *M. leprae*. Bien que certaines EM puissent infecter des animaux (26) et que certaines espèces de mycobactéries semblent être principalement à tropisme animal (27), l'amélioration de l'hygiène a permis de réduire le niveau d'exposition humaine à ces espèces. Il n'est pas clair si les humains et les mycobactéries tirent des avantages indirects l'un de l'autre mais il est plutôt clair que les mycobactéries sont à l'origine d'un large spectre de maladies épidémiologiques, cliniques et pathologiques chez l'humain.

M. leprae (causant la lèpre), *M. tuberculosis* (la tuberculose) et *M. ulcerans* (l'ulcère de Buruli) sont les trois mycobactéries pathogènes les plus prévalentes dans l'espèce humaine. Cependant, certaines EM (par exemple, *M. avium*) ainsi que les vaccins BCG (qui sont inoculés) peuvent occasionnellement, bien que très rarement car étant peu virulentes, causer des maladies cliniques localisées ou disséminées (25), que nous détaillerons en **I_H.4**. En fait, l'immunité innée peut suffire à contrôler l'infection et seule une minorité des personnes infectées développent une maladie clinique. Cela vaut non seulement pour les EM et le BCG, légèrement virulents, mais aussi pour les espèces plus virulentes, *M. tuberculosis* (28) et *M. leprae* (29), qui provoquent une maladie clinique chez moins de 10% des individus infectés. Ainsi, la vulnérabilité aux mycobactéries est l'exception plutôt que la règle chez l'humain.

C. La tuberculose, une maladie ancestrale chez l'humain.

Les échanges ancestraux entre MTBC et *Homo sapiens*, intimes et riches, ont probablement façonné l'histoire génétique de l'un et de l'autre. Les premiers documents écrits décrivant la tuberculose, remontant à 3,300 et 2,300 ans, ont été découverts en Inde et en Chine, respectivement (30). Dans l'Égypte du Royaume Ancien, précédant l'an 2,400 av. J.C.,

sur les portraits funéraires des murs des chambres funéraires ont été retrouvées des traces de lésions causées par la tuberculose (31). Divers rapports techniques des corps momifiés des anciens Égyptiens eux-mêmes, rédigés par des anatomistes professionnels, attestent aussi de la présence de la tuberculose dans l'Égypte de l'Antiquité. Entre 410 et 400 av. J.C. Hippocrate, dans son livre *Book 1, Of the Epidemics*, la décrit comme une maladie de « faiblesse des poumons ». Le mot attribué par les Grecs pour la décrire, *phthisis*, phtisie en français, désignait le dépérissement et la fonte du corps causés par une maladie des poumons. Les phases initiales de la tuberculose étant généralement inaperçues pour le patient et la maladie ne faisait l'objet de soins médicaux que sous sa forme très avancée jusqu'aux temps modernes. En effet, les connaissances anciennes sur la maladie découlaient presque exclusivement de l'observation des symptômes.

Après Koch, il est devenu facile de reconnaître que les différentes formes de tuberculose dans les divers organes du corps possédaient un élément commun, à savoir, les dommages causés par la multiplication des bacilles dans les tissus infectés. Mais avant que ce concept unificateur ne devienne disponible, le médecin était confronté à une gamme confuse de signes et de symptômes, sans relation évidente les uns avec les autres. Plus son acuité clinique était fine, plus il était habile à différencier ces signes et symptômes, et plus il avait tendance à les considérer comme l'expression de maladies différentes. La tuberculose fut appelée par vastes critères liés à son spectre symptomatologique. Par exemple, selon la terminologie anglaise, le nom de *consumption* (consommation en français) accompagna la tuberculose durant la période de la révolution industrielle en Europe tandis que *scrofula*, écrouelles en français, du latin *scrofa*, voulant dire truie d'élevage en français, fut le nom utilisé durant le Moyen-Âge pour décrire une forme répandue de tuberculose (même si le lien entre phtisie et écrouelles n'avait pas été fait), qui produisait un gonflement ganglionnaire similaire à celui observé chez les cochons.

Du temps d'Hippocrate, la tuberculose était déjà considérée la maladie la plus commune de l'époque ; on évoquait sa fatalité, surtout chez les jeunes adultes, et on relevait son caractère héréditaire (30). Les patients étaient soignés dans des temples et suivaient un régime strict et copieux (buvaient en particulier du lait d'ânesse) accompagné d'exercice physique. Des bains de mer étaient de même suggérés pour les patients. Remarquablement, jusqu'au XX^e siècle, le traitement contre la tuberculose n'évolua pas considérablement, du moins dans sa forme. La mort par tuberculose du jeune poète anglais John Keats (1795-1821) en est un clair exemple. Les médecins qui le soignaient combattaient l'évolution de sa maladie en le maintenant sous un régime alimentaire hypocalorique. On l'avait fait partir pour l'Italie

en septembre 1820 à la recherche de la chaleur et du soleil. Le voyage de Londres à Naples prit six semaines. Le commandant de bord essaya même de faire monter une chèvre, le lait de chèvre étant alors très réputé pour le traitement de la tuberculose. Sur les conseils du médecin le plus distingué de la colonie anglaise, le docteur James Clark, et en compagnie d'autres compatriotes malades, Keats commença à faire de l'exercice à cheval et à passer beaucoup de temps sur la colline de Pincio à Rome. Plusieurs hémorragies et une forte fièvre persistante affaiblirent la force de Keats au point de le conduire à l'inactivité. Il mourra finalement, dans son sommeil, un an après le début de cette panoplie de traitements (4).

Ce fut au tournant du XVI^e siècle que les médecins commencèrent à creuser plus profondément sur les aspects pathologiques de la tuberculose, par exemple par la dissection des corps de patients morts. Ils découvrirent bientôt l'association de phtisie avec la présence de curieuses lésions au niveau des organes intérieurs. En 1679, Franciscus Delaboe Sylvius rapporta dans son *Opera-medica* qu'il avait trouvé, en particulier dans les poumons mais également dans d'autres parties du corps, des nodules caractéristiques qu'il appelait des "tubercules", qui suppuraient et provoquaient des ulcères (4). Lors de l'autopsie d'un jeune homme en 1700, Manget observa des tubercules si petits qu'ils ressemblaient à des "graines de mil" présentes de façon ubiquitaire, dans les poumons, la rate, le mésentère, l'intestin. C'est pour commémorer la description de Manget que ce type de maladie disséminée s'appelle de nos jours tuberculose «miliaire». Avec le nombre croissant d'observations, il fut évident que les tubercules, ulcères et cavités pouvaient être représentés chez différents individus par tailles ou caractéristiques bien différentes. La plupart des médecins avaient tendance à considérer ces diverses manifestations pathologiques comme des signes de maladies non liées et essayaient de mettre de l'ordre dans la masse confuse des connaissances anatomiques en les séparant avec soin. Ce genre de pensée continua tout au long d'un siècle. Laënnec (1781-1826), un clinicien français et chercheur médical, dévoué à la tâche de dissection de cadavres, étudiant de Corvisart et de Bayle qui en l'occurrence enseignaient que la phtisie était à l'origine d'au moins six maladies différentes, fut un des premiers à postuler une théorie unifiée de la tuberculose. En 1804, à l'occasion de sa célèbre conférence, il exposa ses vues révolutionnaires sur la genèse de la phtisie pulmonaire, indiquant que les tubercules et les cavités étaient l'expression d'une maladie unique et que les différentes formes de phtisie n'étaient que des aspects différents de la tuberculose pulmonaire. Ce fut Schönlein, professeur de médecine à Zurich qui proposa, d'après le rapport de Franciscus Delaboe Sylvius, d'utiliser le terme « tuberculose » comme terme générique pour décrire toute manifestation de phtisie.

D'une certaine façon, Laënnec mit ainsi un point final à une longue discussion à travers les siècles.

En 1720, le clinicien anglais Benjamin Marten conjectura, pour la première fois, la nature infectieuse de la tuberculose dans son manuscrit *A new theory of Consumption*, c'est-à-dire, 162 ans avant la découverte du bacille par Koch (30). Pendant la révolution industrielle, l'hypothèse fut mise en valeur. Les conditions sociales problématiques, notamment dans les lieux de travail, des logements mal ventilés et surpeuplés, des installations sanitaires primitives, la malnutrition et d'autres facteurs de risque, furent intimement liée à la maladie. En 1799, Thomas Beddoes, un médecin anglais, déclara que la tuberculose pulmonaire était, et de loin, la plus importante cause de mortalité en Grande Bretagne. Seize ans plus tard, ce fut Thomas Young, membre de l'Académie Royale, qui estima qu'un quart des morts étaient dus à la tuberculose. Elle fut bientôt connue dans l'Europe occidentale comme *white plague* (peste blanche en français) dû à la pâleur anémique extrême qui caractérisait les individus atteints. *Youth grows pale, and spectre thin, and dies* (La jeunesse pâlit, devient spectrale et meurt) écrivit Keats durant le printemps 1819. Elle fut particulièrement dévastatrice chez les jeunes adultes. Un tiers de ceux âgés entre 15 et 34 ans et la moitié de ceux compris entre 20 et 24 ans sont morts à cause de la tuberculose pendant la deuxième moitié du XIX^e siècle en Grande Bretagne (4). La tuberculose fut aussi nommée *the robber of youth* ou bien *the captain of all these men of death* suivant le nom que John Bunyan lui donna dans son livre *The life and death of Mr. Badman*.

La nature infectieuse de la tuberculose fut finalement démontrée en 1865 par Jean-Antoine Villemin, médecin à l'hôpital militaire de Val de Grace à Paris. Les expériences de Villemin consistèrent à inoculer à un lapin une petite quantité de liquide purulent d'une cavité tuberculeuse prélevée à l'autopsie d'un individu décédé de tuberculose (32). Le lapin inoculé resta en vie et aucun signe de maladie ne fut découvert, mais à l'autopsie trois mois plus tard, une tuberculose disséminée fut mise en évidence (33). Avec l'isolation du bacille de la tuberculose par Koch en 1882 une nouvelle ère s'ouvrit pour la tuberculose. Dans les décennies qui suivirent, l'intradermo-réaction à la tuberculine (*tuberculin skin test*, TST) de Von Pirquet et Mantoux, le vaccin BCG, la streptomycine Selman Waksman et d'autres médicaments antituberculeux furent développés. Nous rapporterons plus de détails sur les traitements antituberculeux qui accompagnèrent le XX^e siècle en **I_F**, mais globalement, ils contribuèrent au fort déclin des courbes de mortalité et morbidité de la tuberculose en Europe.

Il est évident que la tuberculose eut un impact fort sur l'humain dans un sens évolutif, vu les conséquences démographiques et sociales des derniers millénaires. Cependant, l'étude

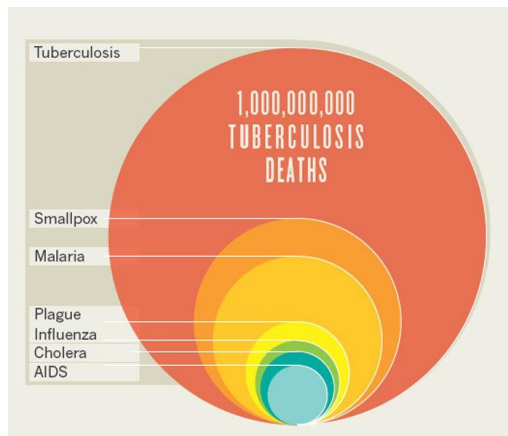
de la prévalence de la maladie au long de l'Histoire relève d'une difficulté majeure car de nombreuses maladies du thorax pouvaient être confondues avec la tuberculose. D'autant plus que, à cause de nombreux préjugés et d'autres raisons pratiques, la tuberculose était souvent négligée des registres publics (pouvant signifier défauts héréditaires obscurs, pauvreté, diminution dans l'admissibilité au mariage, ...). Malgré ces difficultés, grossièrement, il est estimé que la tuberculose aurait tué au moins un milliard de personnes au cours de l'histoire de l'humanité (15,34) (Figure 5). Pour évaluer si cette estimation semble correcte, nous avons effectué les calculs suivants. Nous avons utilisé les estimations effectuées par le « Population Reference Bureau » (www.prb.org/howmanypeoplehaveeverhivedonearth/) pour estimer le nombre de personnes nées sur Terre au cours de deux périodes pour lesquelles la plupart des décès dus à la tuberculose ont eu lieu en Europe: 0-1650 et 1650-1900. Ensuite, nous avons estimé la proportion d'Européens parmi la population mondiale vivant à ces deux périodes à l'aide des données fournies en (35) et disponibles à l'adresse suivante: (www.ggdnc.net/maddison/historical_statistics/horizontal-file_03-2007.xls).

Période	Nombre d'individus nés ^a	Proportion d'Européens nés (%)	Nombre d'Européens nés ^a	Nombre de morts dus à la TB	
				Estimation inférieure	Estimation supérieure
0-1650	39,373	16	6,300	315	630
1650-1900	10,118	23	2,327	465	581
0-1900	49,491	17	8,627	780	1,211

^a En millions.

La proportion de tous les décès dus à la tuberculose en Europe au cours de la période allant de 1650 à 1900 est estimée entre 20% et 25% (4,36), soit un total de décès compris entre 465 et 581 millions d'Européens. Il n'existe pas d'estimation fiable de la proportion de décès liés à la maladie avant 1650. Cependant, des estimations prudentes allant de 5 à 10% refléteraient la mort de 315 et 630 millions de personnes. La mortalité due à la tuberculose en Europe entre 0 et 1900 varierait entre 780 et 1,211 millions de personnes. Des calculs similaires fournissent un nombre supplémentaire d'environ 35 millions d'Européens décédés de la tuberculose après 1900. Globalement, le chiffre d'un milliard de décès au cours des 2,000 dernières années semble être une estimation tout à fait raisonnable.

Figure 5: La tuberculose a tué plus que toute autre maladie infectieuse dans l'Histoire. Plus d'un milliard de personnes au cours des deux mille dernières années



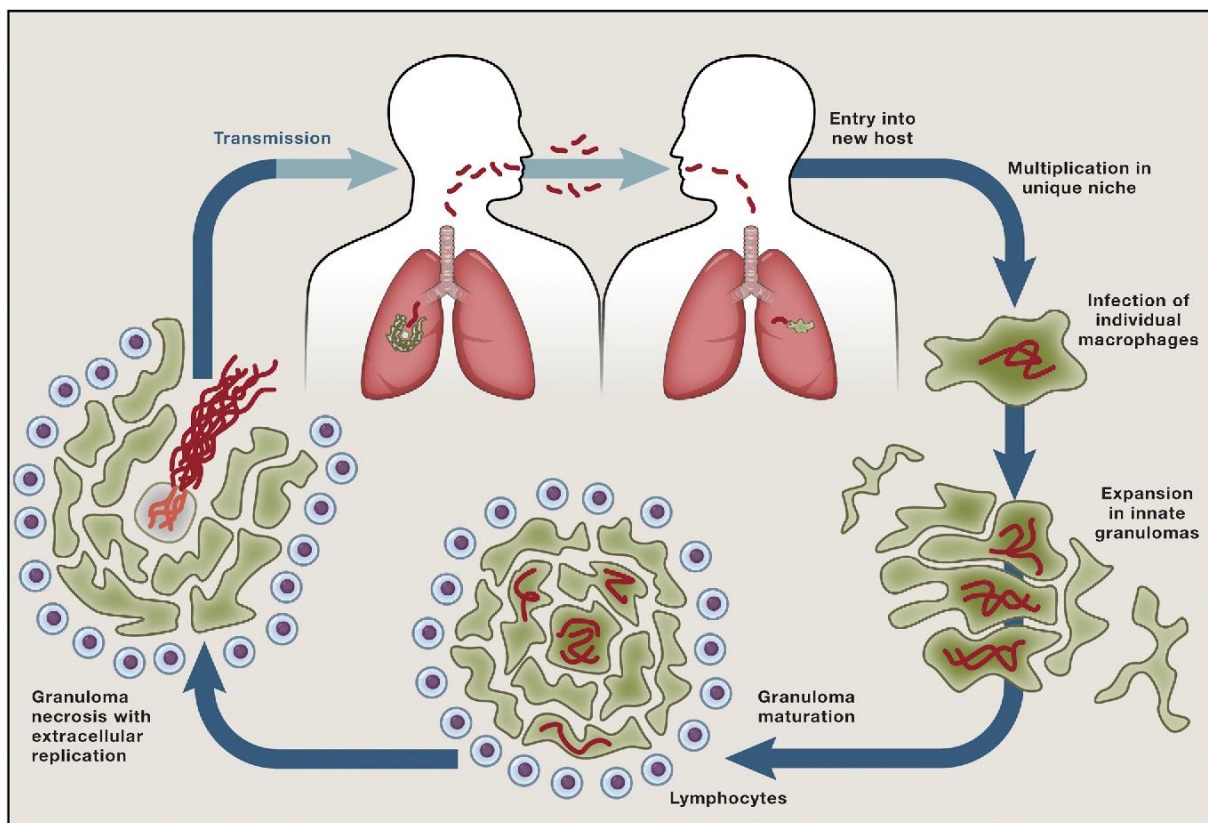
WORLD TUBERCULOSIS REPORT 2012

D. L'histoire naturelle de la tuberculose

Les principales étapes de l'infection par *M. tuberculosis* et de la réponse cellulaire à celle-ci, schématisées sur la Figure 6 et la Figure 7, respectivement, peuvent être dégagées sur la base de modèles expérimentaux et de travaux *in vivo* et *in vitro* chez l'humain. Après inhalation de l'agent pathogène de la tuberculose, transmis généralement par l'expulsion de particules contenant les bactéries contractées par un individu déjà infecté (Figure 6), les premières cellules rencontrées par les bactéries dans les poumons sont les cellules phagocytaires, c'est-à-dire les macrophages alvéolaires, les monocytes, les neutrophiles et les cellules dendritiques (37,38). Les cellules phagocytaires expriment de nombreux récepteurs reconnaissant des motifs moléculaires des pathogènes (ou PRR pour *pattern recognition receptors*), permettant d'initier la réponse immunitaire. Les PRR organisent la phagocytose, la présentation antigénique et permettent l'activation de voies de signalisations intracellulaires et la production de cytokines (39). Les cellules phagocytaires interagissent avec *M. tuberculosis* par l'intermédiaire d'une grande diversité de PRR, comme les Toll-like récepteurs (TLRs), les récepteurs scavenger, les récepteurs du complément (*complement receptor*, CR), les récepteurs des protéines du surfactant, les lectines de type C (i.e *mannose receptor* [MR] ou *dendritic Cell-Specific Intercellular adhesion molecule-3-Grabbing Non-Integrin* [DC-SIGN]), et les récepteurs cytosoliques NOD (*nucleotide-binding oligomerization domain-containing proteins*) (40,41). La reconnaissance de la mycobactérie par les TLRs induit une réponse pro-inflammatoire localisée, caractérisée par la production de cytokines comme les interleukines 12 (IL-12) et 18 (IL-18) et TNF- α , et par celle de chimiokines (CCL2 et CCL5)

entraînant le recrutement par vagues successives de neutrophiles, de cellules NK (pour *natural killer*) et de cellules T CD4⁺ et CD8⁺, chacune sécrétant des cytokines et chimiokines différentes qui amplifient le recrutement cellulaire et le remodelage du site d'infection. A contrario, l'interaction de la mycobactérie avec DC-SIGN et le MR semble stimuler la production d'interleukine 10 (IL-10), cytokine anti-inflammatoire, par les monocytes et les neutrophiles, contrebalançant ainsi la réponse immunitaire induite par les TLRs et profitant au bacille qui peut plus aisément se multiplier (42). Il est à noter également que les bacilles semblent avoir la faculté de différer l'apoptose de leurs cellules hôtes, retardant d'une part l'activation de l'immunité adaptative et d'autre part se donnant ainsi le temps de croître suffisamment dans une cellule donnée avant d'être relâchés en nombre lors de la mort cellulaire (43).

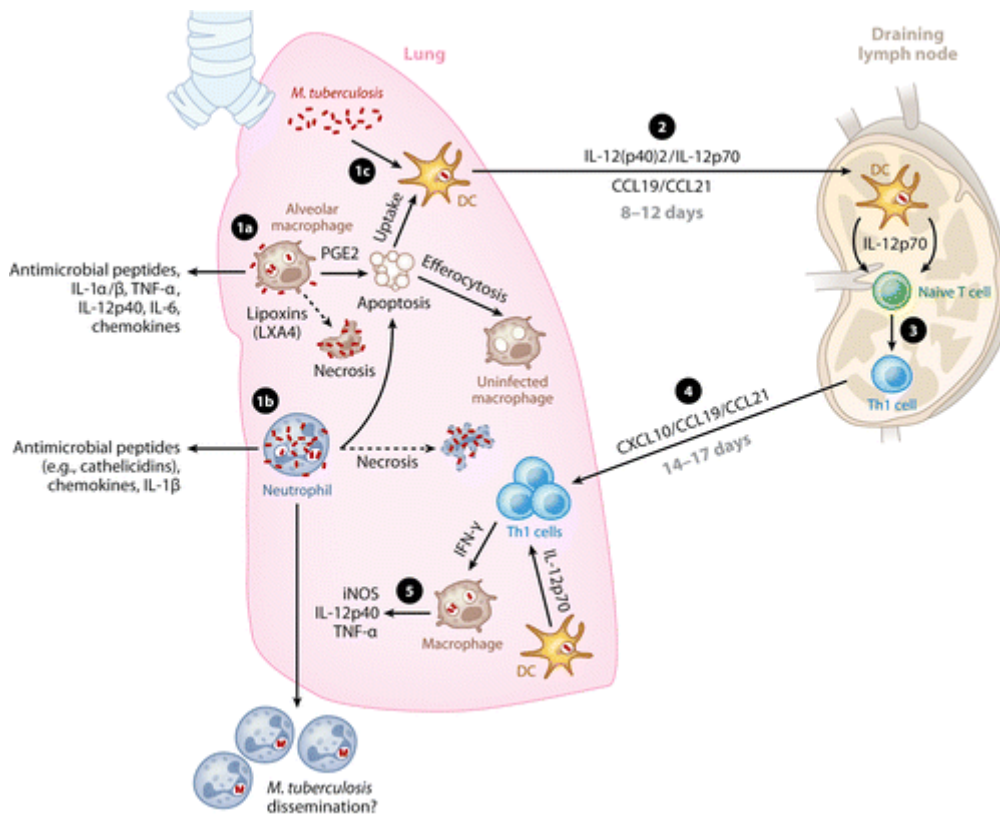
Figure 6: Principales étapes de l'infection par *M. tuberculosis* d'après (44)



A cette réaction inflammatoire locale succède le développement d'une immunité cellulaire T spécifique caractérisée par la production d'interféron- γ (IFN- γ) (Figure 7). Les cellules dendritiques infectées migrent vers les ganglions lymphatiques régionaux où elles vont pouvoir activer les lymphocytes T. Elles présentent les antigènes mycobactériens aux lymphocytes T CD8⁺ et CD4⁺ par le biais des molécules du complexe d'histocompatibilité (HLA) de classe I et II respectivement. En présence d'IL-12, les cellules T CD4⁺ vont se

différencier en cellules de type Th1 caractérisées par la production d'IFN- γ . Une fois activées, les cellules T vont se multiplier et migrer vers le site d'infection primaire où elles vont recruter et activer des macrophages supplémentaires afin de détruire les bacilles intracellulaires. Il apparaît donc que l'IFN- γ est une cytokine clé de la réponse immunitaire dirigée contre les micro-organismes intracellulaires. Elle est principalement sécrétée par les cellules T et les cellules NK en réponse à l'IL-12, et active les macrophages pour permettre la destruction et l'élimination de la bactérie (45). En outre, les études sur le syndrome de susceptibilité mendélienne aux infections mycobactériennes (*Mendelian Susceptibility to Mycobacterial Disease*, MSMD), qui est décrit en détail en **I_H**, ont clairement montré que l'immunité humaine médiée par l'IFN- γ est un trait presque continu, contrôlé génétiquement, influant sur le devenir des infections à mycobactéries (46). Par la suite, la réponse immunitaire dirigée contre *M. tuberculosis* conduit à la formation du granulome tuberculeux (Figure 6), qui est probablement une stratégie protectrice de l'hôte limitant la réplication bacillaire et prévenant la dissémination de l'infection. Une population stable de mycobactéries peut persister dans un état dit quiescent, définissant l'infection tuberculeuse latente.

Figure 7: La réponse cellulaire à l'infection par *M. tuberculosis* d'après (42)



1a : macrophages des alvéoles pulmonaires - 1b : neutrophiles 1c : cellules dendritiques pulmonaires. PGE2 = Prostaglandine E2 = médiateur lipidique pro-apoptotique – LXA4 = lipoxine A4 = médiateur lipidique pro-nécrotique. 2 : cellules dendritiques infectées migrent en 8 à 12 jours vers les ganglions lymphatiques locaux sous l'influence de l'IL-12(p40)2, de l'IL-12p70 et des chimiokines CCL19 and CCL21. 3 : différenciation des cellules T naïves en lymphocytes Th1 4 : Migration des cellules Th1 vers les poumons 14 à 17 jours après le premier contact avec *M. tuberculosis* pour y produire de l'IFN- γ conduisant à l'activation des macrophages. 5 : Production de facteurs microbicides par les macrophages pour contrôler la bactérie.

E. Formes cliniques et diagnostic de la tuberculose

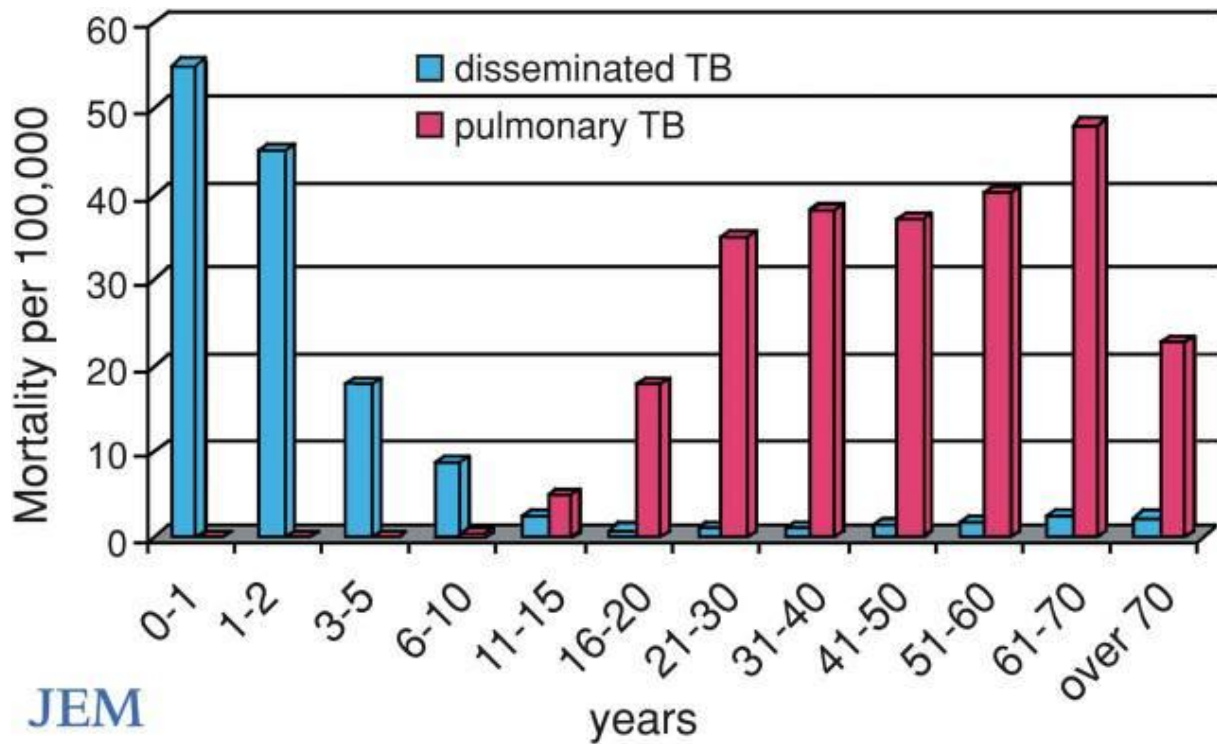
Après avoir été exposés à la mycobactérie, la plupart des êtres humains sont infectés par *M. tuberculosis*, ce qui signifie que la mycobactérie pénètre dans leur corps, déclenche tout le processus détaillé précédemment, et y survit (Figure 6). Cette infection initiale, qui reste la plupart du temps contrôlée et asymptomatique, peut, soit être neutralisée par l'élimination définitive de l'agent pathogène, soit devenir une infection latente (une population de mycobactérie persiste dans un état quiescent), soit encore devenir un cas de maladie. Dans tous les cas, le diagnostic des différentes évolutions possibles après exposition initiale, nécessite des méthodes de détection directes ou indirectes de la mycobactérie, dont certaines sont listées par la suite. Les différentes formes de tuberculose maladie et certains tests de diagnostic d'une tuberculose active sont aussi mentionnées dans ce chapitre.

Le travail que nous présentons ici est focalisé sur la tuberculose maladie. Cependant, les méthodes de diagnostic de l'infection tuberculose motivent aussi cette section. Il n'existe pas de méthode de diagnostic de référence de l'infection tuberculeuse permettant de mettre en

évidence les bacilles et leur viabilité chez l'humain. Néanmoins, l'infection par *M. tuberculosis* laisse une « empreinte » immunologique sur le répertoire des lymphocytes T (47) qui peut être détectée par des tests immunologiques *in vivo* (test de Mantoux ou IDR) ou *in vitro* (mesure de la production d'IFN- γ par les cellules sanguines ou IGRAs). Cette empreinte est à la base du diagnostic d'infection tuberculeuse. Par contre, ces tests immunologiques sont des marqueurs d'infection par *M. tuberculosis* mais ne permettent pas de déterminer si la réponse immunitaire a conduit à une élimination complète du pathogène ou si celui-ci persiste dans l'organisme (47). De plus, une étude récente a mis en évidence la présence d'anticorps contre les protéines ESAT6 et CFP10 spécifiques à *M. tuberculosis* chez des patients négatifs aux deux tests, TST et IGRA, (mais très exposés à l'agent pathogène) révélant un nouveau profil immunitaire chez des sujets exposés à la mycobactérie (48).

La grande majorité (~90%) de ces individus infectés ne présentera jamais de symptomatologie clinique (28). Parmi les 10% qui développeront la maladie, environ la moitié le fera dans les deux années suivant l'infection, ce qui est en général considéré comme une forme primaire de tuberculose (s'agissant souvent des formes pédiatriques de la tuberculose). La mycobactérie peut infecter n'importe quelle partie du corps, mais se limite le plus souvent aux poumons (connue sous le nom de tuberculose pulmonaire). La tuberculose extra-pulmonaire survient lorsque la maladie se développe en dehors des poumons, surtout observé dans les formes primaires, bien que celle-ci puisse coexister avec une tuberculose pulmonaire. Il existe, en effet, deux formes cliniques majeures de tuberculose, correspondant à deux pics d'incidence dépendants de l'âge (Figure 8). Chez l'adulte, l'infection se limite souvent aux poumons et reflète classiquement la réactivation d'une tuberculose latente à partir d'une infection primaire silencieuse. Chez les jeunes enfants, ou les personnes avec un système immunitaire affaibli, la tuberculose est souvent disséminée en raison de la propagation hématogène précoce de la mycobactérie après une infection pulmonaire primaire. En guise d'exemple, chez les personnes à VIH séropositives cela se produit dans plus de 50% des cas (49). Les sites d'infection extra-pulmonaires incluent la plèvre (dans la pleurésie tuberculeuse), le système nerveux central (dans la méningite tuberculeuse), le système lymphatique (dans les écrouelles du cou), le système génito-urinaire (dans la tuberculose urogénitale) et les os et les articulations (dans le mal de Pott par exemple), entre autres. Une forme de tuberculose largement répandue et potentiellement plus grave est appelée "tuberculose disséminée" ou tuberculose miliaire (50).

Figure 8: Distribution des taux de mortalité par tuberculose disséminée (en bleu) et par tuberculose pulmonaire (en rouge) pour 100,000 personnes non soignées vivant en Bavière en 1905, avant l'apparition du vaccin BCG, région endémique pour la tuberculose (51)



Le diagnostic de certitude de la tuberculose maladie repose sur la mise en évidence du bacille, souvent plus facile à réaliser dans les formes pulmonaires de tuberculose. Un examen direct peut être fait à partir de l'étude des crachats des patients, par un examen microscopique ou par la mise en culture du bacille, qui est le moyen le plus fiable pour faire le diagnostic d'une tuberculose. La mise en culture peut se faire à partir d'un milieu solide, qui est la méthode la plus classique (milieu Lowenstein-Jensen), ou bien à partir d'un milieu liquide, qui est une méthode plus récente permettant de diminuer le délai de réponse (qui peut varier, en moyenne, de 8 à 30 jours) (52). Cependant, alors que ces méthodes de laboratoire sont globalement très lentes, des méthodes de biologie moléculaire (ou de diagnostic génotypique) ont été développées pour accélérer le processus de diagnostic. Ces méthodes consistent à amplifier un fragment d'ADN de la mycobactérie puis à l'identifier par hybridation avec des sondes marquées spécifiques. Cette technique est appelée la "réaction polymérase en chaîne" ou PCR. Elle permet une détection directe, rapide et spécifique des bacilles de la tuberculose dans le produit pathologique. Elle est cependant moins sensible que la culture et nécessite un équipement spécifique et coûteux. Le test Xpert MTB/RIF (Cepheid, CA, USA), dédié à la détection du matériel génétique de *M. tuberculosis* (la détection dure en moyenne deux heures) ainsi qu'aux mutations causant des résistances à l'antibiotique rifampicine, est depuis

quelque temps le test de diagnostic génotypique de référence (53). Une version plus rapide, appelée Xpert MTB/RIF Ultra, employant moins de ressources que sa version précédente, a été récemment développée. Sa spécificité dans la détection de la mycobactérie semble être moindre (54), mais il est actuellement déjà en usage en Afrique du Sud. Une version étendue de ces derniers, Xpert XDR, qui permet la détection de souches résistantes à l'isoniazide, s'est montré tout aussi efficace dans une importante étude de validation (54) et serait commercialisée durant cette année 2019. De façon générale, les techniques de séquençage de génome entier (*whole genome sequencing*, WGS) sont aussi utilisées aujourd'hui pour la détection de souches pharmaco-résistantes par l'identification de mutations spécifiquement associés à une résistance phénotypique aux médicaments dans le génome de *M. tuberculosis* (55).

Sans évidence bactériologique, le diagnostic de la tuberculose repose sur un ensemble d'arguments basés sur l'histoire clinique des patients (contact avec des cas tuberculeux, forte exposition au bacille, etc.), leurs études histologiques, l'imagerie, en particulier, des évidences radiologiques ou encore l'évolution des patients sous traitement (53). L'OMS recommande toujours la radiologie des poumons pour le dépistage et le diagnostic de la tuberculose dans certaines populations, cependant, d'autres techniques comme les radiologies digitales des poumons sont en cours de développement pour la détection de la tuberculose (56). Un autre domaine en développement est le champ des biomarqueurs tuberculeux pour lequel quelques candidats prometteurs associés au risque d'infection et au risque et à la protection de la maladie existent déjà (57). Cependant, à ce jour, aucun test clinique basé sur des biomarqueurs n'est disponible sur le marché.

F. Prévention et traitement de la tuberculose.

Tant que la tuberculose n'était que diagnostiquée tardivement, les traitements ciblaient le soulagement des symptômes pénibles de la maladie. Avec l'avènement de la théorie microbienne de nouvelles lignes d'attaque s'ouvrirent contre la tuberculose. Le vaccin BCG eut, depuis la première moitié du XX^e siècle, un rôle important dans la prévention de la maladie mais ce ne fut qu'à partir de la deuxième moitié du XX^e siècle qu'un traitement efficace par antibiotiques fut établi.

À ce jour, le BCG reste le vaccin le plus utilisé dans le monde ayant été administré à plus de 4 milliards d'individus (58). Les diverses analyses réalisées sur la vaccination par le BCG n'ont vraisemblablement pas menées à des conclusions uniformes. Il existe un

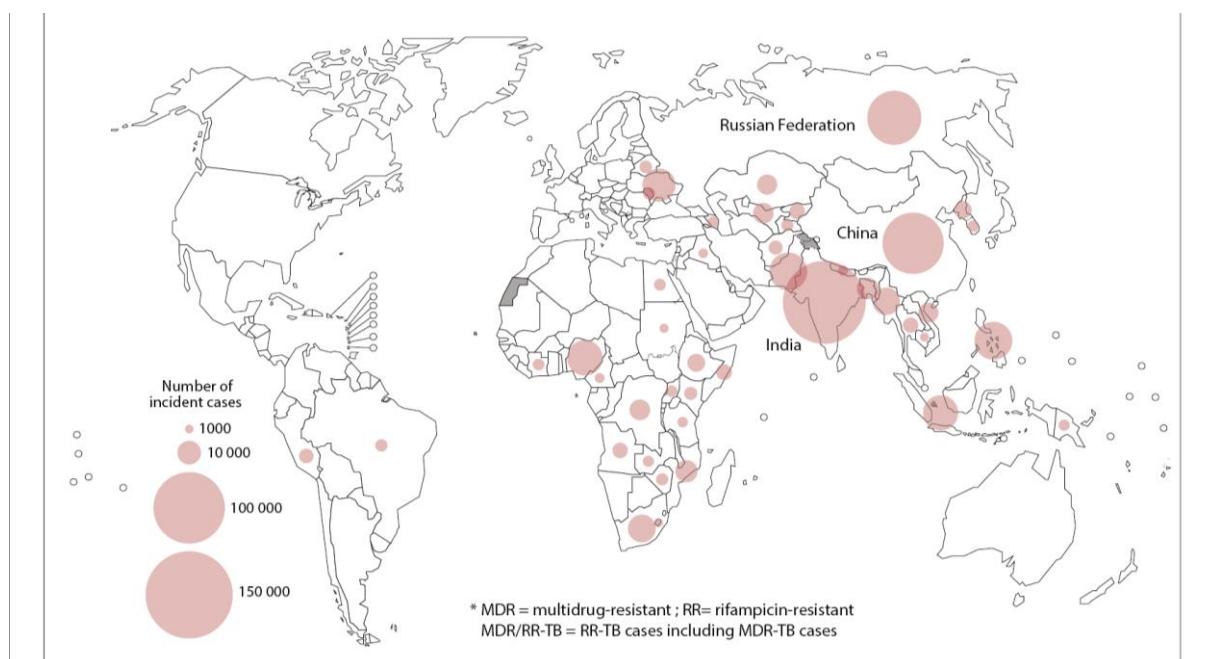
consensus selon lequel le BCG offre une certaine protection, en particulier contre la tuberculose sévère (formes extra-pulmonaires) chez les enfants (59), en réduisant le taux d'incidence de la maladie de 15% à 50% selon la région du monde évaluée (60,61). La protection contre la tuberculose pulmonaire de l'adulte, qui est la forme la plus contagieuse, reste néanmoins très variable (62). D'abord isolé d'un cas de tuberculose chez des bovins, la souche de *M. bovis* initialement utilisée par Calmette et Guérin pour développer le vaccin a été progressivement atténuée pendant 13 années. Des expériences menées dans le monde entier ont confirmé la perte de virulence et la rétention de son pouvoir immunisant. En 1921, Calmette prit la décision de mener un premier essai de vaccination chez l'humain et ce fut par voie orale à un enfant qui venait de naître d'une mère décédée de tuberculose quelques heures après accouchement. Aucun effet indésirable ne fut rapporté. Déjà en 1924, un rapport décrivait 664 vaccinations orales de BCG chez des enfants (63). Entre 1924 et 1928, 114,000 enfants avaient été vaccinés (64) et des lots de BCG furent distribués de par le monde pour que chaque laboratoire puisse fabriquer son propre vaccin à partir de souches dérivées du BCG.

La lutte contre la tuberculose s'est accélérée après la fin de la deuxième Guerre mondiale et l'administration du BCG fut encouragée. La procédure utilisée au Royaume-Uni (souche de Copenhague, administrée aux âgés de 13 ans) se montra très efficace contre la tuberculose (65), alors qu'aux États-Unis (souche Tice, administrée à un grand rang d'âge d'individus), elle ne montra pas d'avantage remarquable (66). En effet, comme la durée de culture visant à l'atténuation de la virulence de la mycobactérie est variable d'un lot à l'autre, les vaccins diffèrent. Ces différences furent mises en évidence plus tard grâce aux techniques de génotypage (67,68). Il s'avère que la plupart des pays du monde suivirent l'exemple de l'Europe et introduisirent la vaccination systématique du BCG, tandis que les États-Unis se prononcèrent contre l'utilisation systématique du BCG et fondèrent leur stratégie de lutte antituberculeuse sur l'utilisation de la tuberculine pour identifier les personnes à traiter (69). Aujourd'hui le vaccin BCG demeure le seul vaccin utilisé pour prévenir la tuberculose même si au cours des deux dernières décennies un nombre considérable de vaccins alternatifs ont été étudiés, ne montrant jusqu'ici pas plus d'efficacité que le BCG lui-même (70).

Les premières avancées importantes dans le traitement contre la tuberculose vinrent avec l'introduction du premier antibiotique efficace en 1944, date de la découverte de la streptomycine. Le traitement a évolué au fil des découvertes pour aboutir aujourd'hui à un traitement standard chez l'adulte d'une durée minimale de 6 mois. Il associe 4 antibiotiques antituberculeux (isoniazide, rifampicine, pyrazinamide et éthambutol) qui doivent être pris

quotidiennement pendant les 2 premiers mois, suivi d'une association isoniazide-rifampicine pendant 4 mois. De nombreuses études visant à évaluer des schémas thérapeutiques antituberculeux plus courts sont en cours, notamment des schémas thérapeutiques contenant de la rifapentine, de la clofazimine et deux nouveaux médicaments, à savoir, la bédaquiline et la delamanide (53). Il est à noter que depuis quelque temps, le vaccin n'étant pas suffisamment immunogène pour induire une immunité à long terme, des schémas thérapeutiques de prévention de la tuberculose nettement plus courts ont été mis au point et se sont avérés efficaces chez les adultes et les enfants, y compris ceux vivant avec le VIH (71). Un schéma thérapeutique à la rifampicine d'une durée de 4 mois a été testé chez l'adulte et l'enfant et s'est révélé aussi efficace que l'isoniazide pris pendant 9 mois (71). Cependant, le développement d'une pharmino-résistance lors de l'utilisation d'un médicament unique chez des individus sans une tuberculose active diagnostiquée demeure un danger préoccupant.

Figure 9: Estimation de l'incidence des cas de tuberculose MDR ou résistant seulement à la rifampicine en 2017, pour les pays comportant au moins 1000 cas incidents



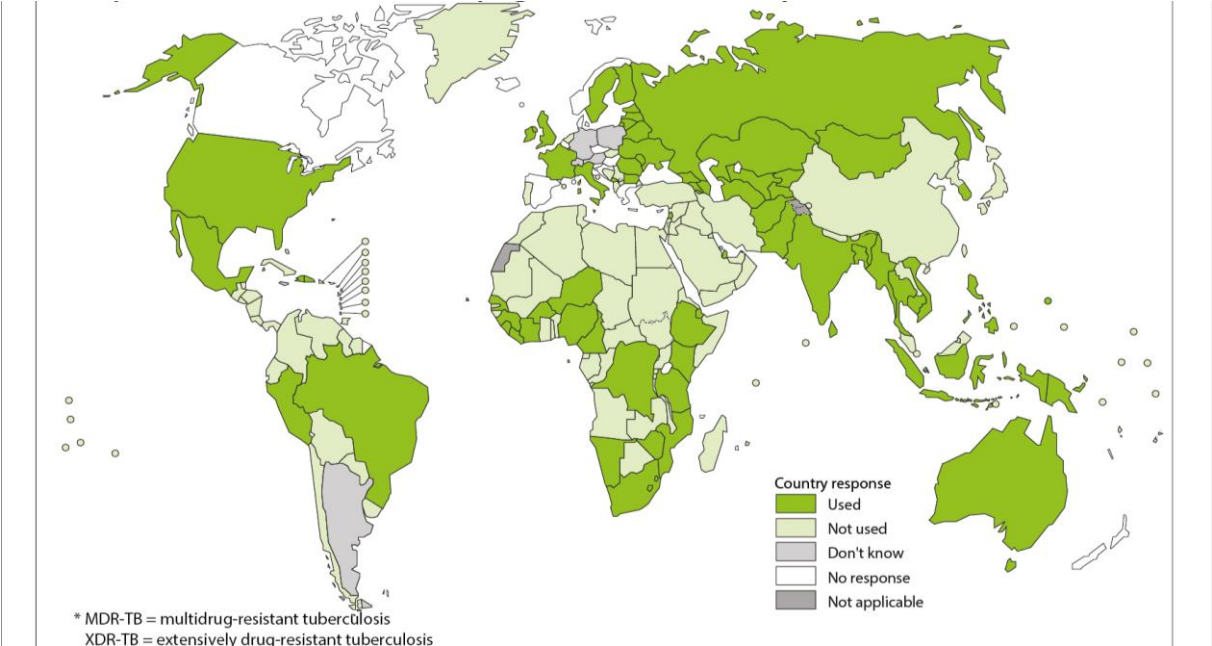
Source : OMS 2018

La résistance apparaît quand les médicaments antituberculeux ne sont pas utilisés adéquatement, du fait de prescriptions incorrectes de la part des professionnels de santé, de médicaments de mauvaise qualité ou de la mauvaise observance du traitement par les patients. La tuberculose multi-résistante (tuberculose-MDR) est une forme de la maladie due à un bacille ne réagissant ni à l'isoniazide ni à la rifampicine, les deux médicaments antituberculeux de première intention les plus efficaces (Figure 9). On peut néanmoins

soigner et guérir la tuberculose-MDR avec des médicaments de deuxième intention. Ces options thérapeutiques sont toutefois plus limitées et nécessitent une administration de longue durée (jusqu'à deux ans de traitement) de médicaments à la fois chers et toxiques. Dans certains cas, une résistance plus sévère peut se développer. La tuberculose ultra-résistante (tuberculose-XDR) est une forme encore plus grave de tuberculose-MDR due à des bacilles ne répondant pas aux médicaments de deuxième intention les plus efficaces, laissant souvent les patients sans aucune autre option thérapeutique. Les formes de tuberculose résistantes à l'isoniazide sont les formes de tuberculose pharmaco-résistantes les plus courantes dans le monde, bien qu'il n'existe pas de directives sur le moment de comment ou quand tester cette forme de résistance. Par contre, le traitement de la tuberculose résistante à la rifampicine eut un bouleversement très positif avec l'introduction de bédaquiline et de delamanide. La bédaquiline a été recommandée pour la première fois par l'OMS en 2013 et à nouveau en 2017 (53).

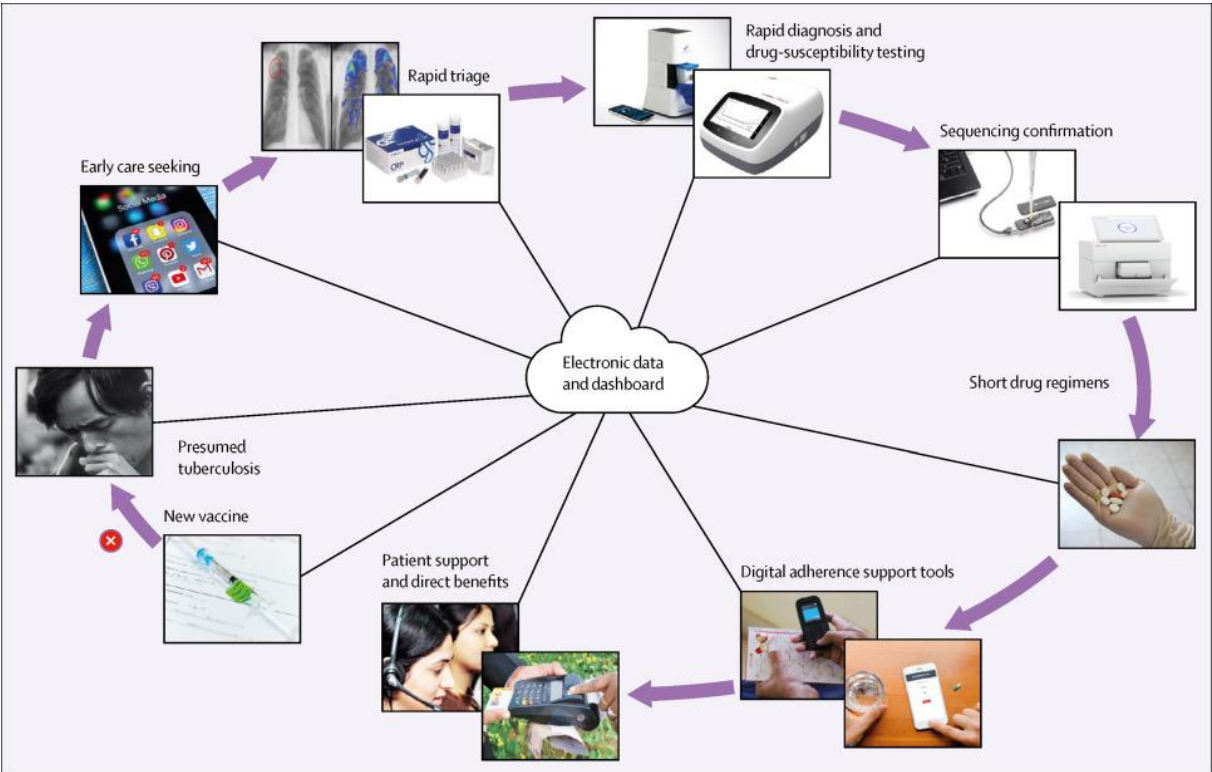
En somme, les stratégies présentées consistent à traiter les patients symptomatiques, à prévenir le développement d'une tuberculose clinique chez les personnes exposées à la mycobactérie, à stopper la transmission du pathogène et à limiter l'apparition de souches résistantes aux antibiotiques. Enfin, la prévention de la tuberculose qui a longtemps reposé sur des outils peu efficaces nécessite la mise au point d'un meilleur vaccin ou d'une thérapie plus courte pour moderniser le traitement de la tuberculose.

Figure 10: Pays ayant utilisé la bédaquiline pour le traitement de la tuberculose XDR dans le cadre d'un accès élargi, dans des conditions de programmation normales pendant l'année 2017



Source : OMS 2018

Figure 11: Schéma pour un système moderne de prévention et traitement de la tuberculose issu de (53)

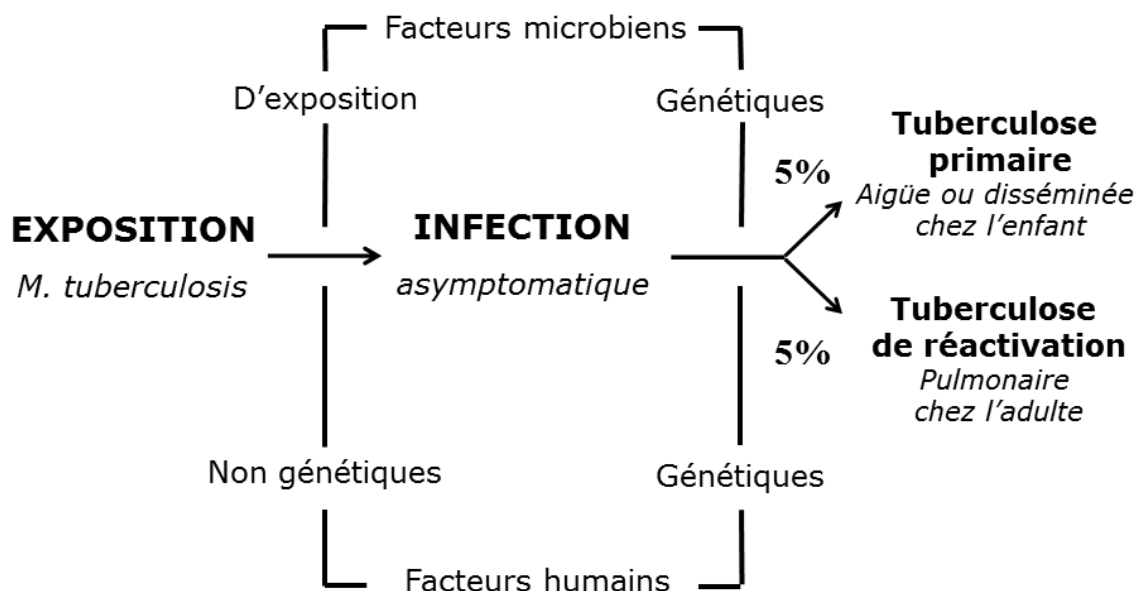


G. Les facteurs de risque de la tuberculose.

« Il est certain que les facteurs qui déterminent la susceptibilité ou la résistance à la tuberculose sont contrôlés par l'hérédité, modifiés par les conditions de vie et reposent probablement sur certains processus de l'immunité qui agissent dans d'autres maladies infectieuses. » Pearson (1912)

Le risque de progression vers une maladie active est un processus régi par des facteurs de risque exogènes et endogènes. Des facteurs environnementaux (facteurs d'exposition et facteurs microbiens) et des facteurs de l'hôte (génétiques ou non spécifiquement génétiques) peuvent moduler le risque d'infection, le risque de progression vers une tuberculose active ainsi que sa présentation clinique (Figure 12), pour lesquels nous essaierons de donner quelques exemples.

Figure 12: Schéma général du passage de l'exposition à l'infection et de l'infection au développement d'une tuberculose active en fonction des facteurs associés



Pour développer un cas clinique de tuberculose il faut tout d'abord être infecté par la mycobactérie et donc avoir été exposé à celle-ci. L'urbanisation rapide (72,73) observée dans les pays en développement ainsi que le statut socio-économique de certains individus promouvant la promiscuité augmentent le risque d'infection. Des études de contacts domestiques chez des patients tuberculeux du début du XX^e siècle (74,75) et de grandes enquêtes épidémiologiques (76–78) ont montré cet effet. D'autres, plus spécifiques, ont montré que le niveau de concentration de bacilles dans les expectorations d'un cas de tuberculose était positivement corrélé au niveau de contagiosité du patient (79). Il est à noter néanmoins qu'il existe des individus fortement exposés à *M. tuberculosis* qui ne deviennent

pas infectés. Nous ne traiterons cependant pas dans ce travail le cas spécifique et très intéressant de ces individus dits « résistants » à l'infection.

Après les facteurs d'exposition à l'agent pathogène, des facteurs génétiques liés au microbe peuvent également contribuer au risque du développement de la maladie. Dans le modèle murin, les souches de la famille Beijing (appartenant à la lignée 2, Asie de l'Est) ont une virulence accrue qui se caractérise par une plus forte multiplication bacillaire, une plus grande dissémination des bacilles et un décès précoce de l'animal (80–82). Chez l'humain, une étude réalisée en Gambie a montré que le taux de progression vers une tuberculose active était significativement plus élevé chez les individus en contact avec un patient atteint de tuberculose causée par *M. tuberculosis*, et particulièrement par les souches de la famille Beijing, que chez les patients en contact avec un patient atteint de tuberculose causée par *M. africanum* (83).

Avec les facteurs microbiens susmentionnés, des facteurs propres à l'hôte peuvent aussi moduler la survenue suite à l'infection. Nous pouvons distinguer ceux ayant une composante génétique (que nous détaillerons dans la section **I_H**) et ceux non spécifiquement génétiques, pour lesquels nous donnons quelques exemples ci-après. La malnutrition est un facteur connu pour augmenter le risque de tuberculose en raison d'une réponse immunitaire altérée (10,84). L'association entre la malnutrition et la tuberculose a été démontrée lors d'essais de vaccination du BCG menés aux États-Unis à la fin des années 1960, estimant que les enfants malnutris étaient deux fois plus susceptibles de contracter la maladie que ceux bien nourris (66). L'association entre le tabagisme et la tuberculose a été également étudiée dans plusieurs revues systématiques (85,86). Dans leur méta-analyse de 24 études sur les effets du tabagisme sur la tuberculose, Bates et ses collègues ont montré que le risque relatif d'infection par la tuberculose ($RR = 2.3-2.7$) était élevé chez les fumeurs par rapport aux non-fumeurs, entraînant un risque supplémentaire de décès chez les personnes atteintes de tuberculose active (87).

D'autres comorbidités peuvent contribuer au développement d'une tuberculose active. Le diabète en est sûrement une. Les premiers rapports sur l'association entre diabète et tuberculose incluent la description du philosophe persan Avicenne au X^e siècle et d'un siddhar indien, Yugimahamuni, qui décrivait la progression de l'obésité vers l'impuissance, la soif, la glycosurie et finalement vers la tuberculose (88,89). En 2008, une revue systématique comparant 13 études focalisées sur l'association entre le diabète et la tuberculose a révélé que les patients diabétiques présentaient un risque environ trois fois plus élevé de développer une tuberculose par rapport aux patients non diabétiques (90). Les immunodéficiences acquises

sont probablement, parmi toutes les comorbidités, celles ayant le plus fort impact dans la tuberculose. Par exemple, la co-infection par le virus du VIH est le facteur de risque immunosuppresseur le plus récent dans le développement de la tuberculose active (91), en exacerbant la gravité de la maladie (92). Plusieurs études ont montré que le nombre de cellules CD4 (ou la diminution du nombre) chez les adultes infectés par le virus du VIH, qui ne suivent pas de traitement antirétroviral, est fortement corrélé au risque de développer une tuberculose (93–95). Les individus sous traitements immunosuppresseurs sont, de façon générale à risque pour la tuberculose, ceci étant par exemple le cas dans les maladies systémiques sous corticothérapie au long cours. Dans ce contexte, l'utilisation d'inhibiteurs du TNF- α est connue pour augmenter le risque de développer une tuberculose active (96,97).

Les facteurs de risque tels les déficits immunitaires acquis laissent entrevoir un lien entre tuberculose et génétique humaine. C'est justement aux facteurs génétiques ayant une influence sur le développement d'une tuberculose active que nous nous sommes particulièrement intéressés dans ce travail de thèse. Pour mieux comprendre la motivation agissant sur celui-ci, le prochain chapitre tente de résumer, très brièvement, quelques aspects qui ont permis d'établir les bases d'une théorie génétique de la tuberculose.

H. La génétique humaine de la tuberculose maladie.

H.1. Historique

La découverte d'un grand nombre d'infections asymptomatiques au début du XX^e siècle ouvrit une interrogation majeure dans le monde scientifique, avide de comprendre la variabilité interindividuelle observée au cours de l'infection. Charles Nicolle fut le premier à démontrer formellement, entre 1911 et 1917, la coexistence d'infections symptomatiques et asymptomatiques dans les populations humaines (98). Cette variabilité fut souvent attribuée au rôle de l'immunité acquise, comme le démontrent en quelque sorte les premières expériences de vaccination, cependant, elle ne put pas expliquer la variabilité interindividuelle observée dans les infections primaires chez l'enfant. La description des premières immunodéficiences primaires (*primary immunodeficiencies*, PIDs) dans les années 1950 dues à des mutations génétiques instaura l'idée, parmi de nombreux généticiens, que la prédisposition au développement des maladies infectieuses pouvait être transmise de génération en génération. Cependant l'hypothèse génétique eut du mal à trouver des adeptes. Le paradigme établi par Pasteur et Koch avec leur théorie microbienne des maladies

infectieuses ne fut pas souvent remise en cause, répétant le célèbre aphorisme de Pasteur : « une maladie, un microbe, un vaccin ». D'autre part, l'impossibilité de tester concrètement l'hypothèse génétique due à l'absence de marqueurs génétiques jusque dans les années 1990 fut un obstacle majeur pour le soutien de cette théorie.

Dans la deuxième moitié du XX^e siècle, certains exemples agirent tout de même comme pierres fondatrices d'une théorie génétique des maladies infectieuses. L'étude des formes sévères de *Plasmodium falciparum* malaria par Anthony Allison au tout début de cette période est un des plus remarquables. Il montra que le développement de ces formes de malaria avait lieu plus fréquemment dans les zones de forte prévalence d'une variation génétique, en l'occurrence, de la variation Hbs du gène Hb β (HBB), causant le trait drépanocytaire, et que les porteurs hétérozygotes développaient souvent des formes moins sévères. Durant la même période, les premières descriptions de PIDs associés à un seul agent infectieux tel que l'épidermodysplasie verruciforme (prédisposant au papillomavirus humain) et la maladie lymphoproliférative liée au chromosome X (prédisposant au virus Epstein-Barr) furent rapportées. Ces observations contribuèrent indirectement à la découverte, dans les années 1990, des premières mutations prédisposant à un seul agent infectieux, telles les infections par mycobactérie ou l'encéphalite herpétique (99).

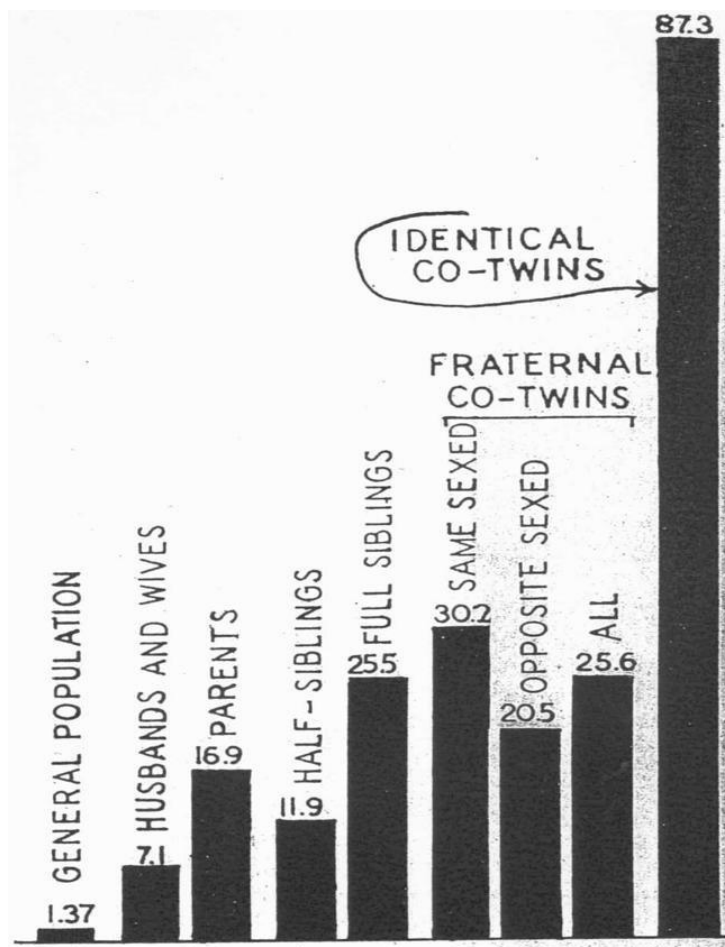
Parmi le groupe de maladies infectieuses qui furent soumises à l'examen d'une théorie génétique de l'hôte on retrouve aussi la tuberculose. En 1944, R. R. Puffer, dans son livre *Familial Susceptibility to Tuberculosis* fit un résumé des études scientifiques focalisées sur deux facteurs du développement de la tuberculose : la susceptibilité familiale et l'exposition au bacille de la tuberculose (100). Plus tôt, en 1912, Karl Pearson, lors d'une conférence au laboratoire Francis Galton Eugenics de l'université de Londres avait rappelé : « Pour les grands rédacteurs médicaux de la première partie du siècle dernier, le sol signifiait tout et le microbe peu, peut-être [...] parce qu'ils n'en savaient rien. Avec la découverte du bacille tuberculeux, le pendule a basculé immédiatement de l'autre côté - l'idée de l'infection dominait tout le monde, [...]. On ne saurait surestimer l'importance de la découverte de Koch, mais en conséquence, une ligne d'action très précise fut adoptée sans enquête scientifique adéquate sur l'importance relative: 1) du facteur héréditaire, 2) du facteur environnemental et 3) du risque d'infection. » Ainsi, très tôt dans l'histoire et bien avant la consolidation d'une théorie génétique de l'hôte, Pearson reconnut qu'il était souhaitable de reprendre en compte le sol (dit autrement, l'hôte, l'humain) tout autant que le microbe dans le développement de la tuberculose.

Plusieurs observations ont suggéré fortement le rôle de la génétique humaine dans la résistance ou la susceptibilité face aux infections par mycobactéries, par exemple, à partir d'études interethniques, en particulier, la constatation d'une prévalence plus élevée de la tuberculose chez les populations d'origine africaine que chez celles d'origine caucasienne (101). D'autres études menées au Canada et en Australie ont montré que, en comparaison aux non-autochtones, les autochtones ou les aborigènes étaient plus susceptibles au risque de tuberculose (102,103). Cependant, la preuve expérimentale irréfutable fut malheureusement fournie par une tragédie qui se produisit en 1926 à Lübeck, Allemagne. Au tout début de l'application du BCG en Europe, 251 bébés furent injectés avec d'importantes doses de bacilles tuberculeux vivants et virulents par erreur (le vaccin BCG avait été contaminé par des quantités variables de *M. tuberculosis* complètement virulent). Parmi les enfants vaccinés, 228 développèrent des signes cliniques ou radiologiques de tuberculose variés. Un an après l'accident, 77 décédèrent et 69 d'entre eux moururent de causes reliées à la tuberculose (104,105). Remarquablement, 23 enfants ne présentèrent aucun signe clinique de la maladie, et 68% des cas confirmés de tuberculose guérirent spontanément (104,105). La première source de variabilité des présentations cliniques observées reposait sur la dose infectieuse injectée, du fait d'une contamination non uniforme des lots de vaccins. Cependant, même après avoir tenu compte de la dose de bactérie virulente injectée, la variabilité clinique résiduelle demeura élevée, allant du décès de l'enfant à des symptômes très légers. En l'absence d'autres facteurs sociaux ou environnementaux connus, cette variabilité suggérerait des différences génétiques entre les nouveau-nés quant à leur capacité innée à combattre *M. tuberculosis*, en particulier, chez ceux ayant reçu des faibles doses.

D'autres études montrèrent que la composante familiale était un déterminant important dans la susceptibilité à certaines maladies infectieuses posant le problème de la distinction entre les causes environnementales (facteurs familiaux augmentant le risque d'infection) et génétiques de l'agrégation familiale. Des études de jumeaux soulignèrent que la concordance de progression vers une tuberculose clinique était plus élevée chez des jumeaux monozygotes que chez des jumeaux dizygotes, les conditions environnementales et sociales étant par ailleurs égales. En étudiant 308 familles de jumeaux, Kallman et Reisner ont rapporté dès 1943 que le taux de morbidité chez les jumeaux monozygotes était de 87.3% vis-à-vis de la tuberculose contre 26.3% pour les jumeaux dizygotes, lorsque les deux frères ou sœurs avaient été exposés de manière avérée à *M. tuberculosis* (106) (Figure 13). L'étude montra par ailleurs que cette concordance était d'autant plus élevée chez des individus reliés par des liens familiaux que chez des individus sans lien sanguin direct (par exemple, 1% dans la

population générale et 7% chez les époux contre 16.9% chez les parents et 11.9% chez les demi-frères ou sœurs, respectivement) (Figure 13). Une autre étude de 1978 ré-analysant des données du Prophit survey montra qu'après ajustement sur des facteurs de risque comme l'âge, le sexe ou la présence de bacilles dans les sécrétions des cas index, le taux de tuberculose était deux fois plus élevé chez les individus ayant un jumeau monozygote atteint que chez ceux dont le jumeau atteint était dizygote (107).

Figure 13: Taux de morbidité (en %) liée à la tuberculose chez les apparentés de cas index d'après l'étude de Kallmann & Reisner



L'immunité aux mycobactéries a fait aussi l'objet de nombreuses recherches sur plusieurs modèles animaux. Les études les plus approfondies sur la tuberculose chez le lapin ont été réalisées par Max B. Lurie (108–111). Il mena des études détaillées sur une période de sept ans sur des lignées de lapins consanguins afin d'étudier leur résistance naturelle face à la tuberculose. Lurie constata que les lignées présentaient soit une résistance, soit une susceptibilité à l'infection par *M. bovis* virulent, et que, bien que les lignées résistantes développaient une tuberculose cavitaire, le plus sensibles développaient une maladie disséminée. Les lapins résistants vivaient d'ailleurs environ deux fois plus longtemps que les

lapins sensibles. Ces conclusions suggérèrent que la résistance était principalement déterminée génétiquement. Cependant, dans ce contexte, les progrès les plus importants ont été réalisés à partir du modèle murin. Un des exemples les plus emblématiques fut rapporté au début des années 1990. Dans ce travail, le phénotype utilisé pour cartographier le premier gène qui chez la souris contribua à des différences de réponse à *M. bovis*/BCG fut la permissivité à la réplication microbienne dans la rate après injection intraveineuse. Il fut constaté que certaines lignées de souris étaient permissives ou non permissives à la réplication de *M. bovis*/BCG dans la rate aux stades précoces de l'infection (112). La susceptibilité était héritée comme un trait récessif sur un seul locus (112–114) contrôlant l'étendue de la réplication microbienne intracellulaire dans les macrophages (115–117). Une approche de clonage positionnel identifia le gène associé à ce locus, qui fut appelé *Nramp1* (protéine 1 de macrophage associée à la résistance naturelle, maintenant connu sous le nom de *SLC11A1*) (118). Les lignées de souris sensibles à l'infection par *M. bovis*/BCG portaient une seule mutation dans la région impactant le repliement des protéines, altérant leur maturation et provoquant une dégradation rapide (118–120). Malgré l'impact considérable de ces études sur la recherche de facteurs pathogéniques associés aux mycobactéries, les modèles animaux souffrent des limites inhérentes aux modèles expérimentaux d'infection. Par exemple, les espèces de mycobactéries avec lesquelles les souris sont inoculées ne sont pas des agents pathogènes naturels pour elle.

H.2. Etudes en population.

Dans les années 1990 vint l'avènement des marqueurs génétiques. Les études de génétique épidémiologique qui s'en servirent eurent pour but l'identification de gènes ayant un impact sur un phénotype donné mais furent tout d'abord très limités par le nombre relativement faible de marqueurs génétiques disponibles à tester. Compte tenu de cette limite, les premiers travaux essayant d'associer gènes et pathologie (études d'association) furent des études dites « gènes candidats par hypothèse » ou tout simplement « gènes-candidats », se focalisant sur des gènes connus au préalable pour leurs implications dans la pathologie analysée ou dans une voie physiopathologique associée (en se basant souvent sur des études murines). Pour la tuberculose, certains gènes, en particulier ceux du complexe d'histocompatibilité (HLA) de classe I et II, remplissant les deux critères, furent l'objet de plusieurs études (121). Il fut de même le cas pour un polymorphisme du gène du récepteur de la vitamine D (122) ou pour certains polymorphismes du gène de l'interleukine-1 β (123). Cependant, ces découvertes n'ont été que rarement confirmées par des études indépendantes,

en raison d'un manque de puissance desdites études, d'une grande hétérogénéité des contextes dans lesquels elles ont été effectuées, de la définition de leur phénotype, ou tout simplement de l'implication réelle du locus sur le phénotype. L'une des découvertes les plus convaincantes fut l'identification de polymorphismes associés au gène *NRAMP1*, impliquant par ailleurs un effet hétérogène au niveau des populations, des paramètres épidémiologiques, de l'âge et des phénotypes cliniques (124–126). Le gène est spécialement intéressant car l'homologue murin du gène était connu pour participer dans le contrôle de certaines infections mycobactériennes (127) (voir section **I_H.1**). Il a été de même par la suite associé à la lèpre dans l'étude de familles vietnamiennes (128).

Avec le développement des puces de génotypage, les analyses de liaison, développées par Fisher et Penrose dans les années 1930 (129,130), décrites pour un cas spécifique en **II_B.2**, furent utilisées dans le but d'identifier des régions chromosomiques contenant des gènes associés à la maladie étudiée. De façon très générale, elles analysent la co-ségrégation de marqueurs génétiques et de la maladie au sein de familles. Parmi les analyses basées sur les techniques susmentionnées, une grande étude réalisée au Canada, après une épidémie de tuberculose qui toucha spécialement des individus d'origine amérindienne, mis en évidence la région chromosomique 2q35 contenant le gène *NRAMP1* (125). Une autre étude réalisée au Maroc mis en évidence un locus majeur sur le chromosome 8 conférant une prédisposition à la tuberculose pulmonaire (131). Cette région a ensuite été cartographiée par clonage positionnel. En considérant l'âge de survenue de la maladie, des variants du gène *TOX* ont été identifiés comme fortement associés au développement d'une tuberculose pulmonaire avant l'âge de 25 ans dans des populations du Maroc et de Madagascar (132). *TOX* code pour un facteur nucléaire impliqué dans le développement des lymphocytes T, en particulier les lymphocytes T CD4⁺ qui sont essentiels dans la lutte contre les mycobactéries (133). D'autre part, certains papiers récents focalisés sur les cellules T CD8⁺, soulignent l'importance du gène dans l'immunité à long terme de ces cellules dans la lutte contre des infections virales chroniques (134–136). *TOX* apparaît comme un élément régulateur central du développement de certains lymphocytes T dits « épuisés » (*exhausted*).

L'avènement des études d'association pan-génomiques (Genome Wide Association Studies, GWAS) à l'aide de données de génotypage au début des années 2000 suscita beaucoup d'espoirs dans la communauté scientifique, en particulier dans la recherche de facteurs explicatifs du développement de la tuberculose. Ces études sont basées sur la recherche d'associations statistiques entre des polymorphismes fréquents répartis tout au long du génome, et un phénotype donné (par exemple, statut binaire malade / non-malade). Des

études à grande échelle au Ghana et en Gambie (137,138), et une autre en Russie (139) n'ont cependant conduit qu'à l'identification de trois signaux significatifs à l'échelle du génome ($p < 5 \times 10^{-8}$). L'un des deux polymorphismes mononucléotidique (*single nucleotide polymorphism*, SNP) identifiés en Afrique, *rs4331426*, est situé dans un désert de gènes du chromosome 18q11.2, et le second, *rs2057178*, est situé près du gène WT1 sur le chromosome 11p13. Les deux variants ont des tailles d'effet modestes (OR = 1.19 pour *rs4331426* et OR = 0.77 pour *rs2057178*). Le signal du chromosome 11 a été répliqué en Indonésie et en Russie dans l'étude originale, puis en Afrique du Sud (140) et au Maroc (141) dans des études indépendantes. Les tentatives de réplication du signal du chromosome 18 n'ont pas toutes abouties à des résultats probants, en particulier en Chine (142,143), bien qu'au Maroc, le signal semble être répliqué (141). L'étude russe a pu identifier un groupe de variants introniques du gène ASAP1 dans une population de plus de 15,000 participants avec aussi une taille d'effet faible (OR=0.84 pour le SNP *rs4733781*) et un rôle potentiel dans la mobilité des cellules dendritiques (139). Une grande étude d'association réalisée en Islande a également rapporté des signaux dans la région HLA de classe II (144), mais le rôle précis de ces variants dans le développement d'une tuberculose clinique reste à définir.

Malgré les découvertes mentionnées, ces analyses d'associations à l'échelle du génome entier, focalisées sur des variants largement représentés dans la population d'étude, nécessitant un grand nombre de cas et de témoins pour faire valoir la puissance statistique des tests sous-jacents, se sont montrées décevantes. Une première surprise fut l'absence de réplication des facteurs de susceptibilité génétique identifiés par les études centrées sur des gènes candidats, témoignant probablement du fait que ces loci avaient impact négligeable sur le phénotype. Mais ces GWAS furent décevants car ils n'aboutirent pas à la découverte de variants génétiques avec un effet majeur dans la tuberculose. Ceci peut être expliqué par plusieurs phénomènes. En premier lieu, les variants communs peuvent avoir un impact faible sur la tuberculose ou hétérogène par rapport au phénotype clinique (les patients inclus ayant souvent un large éventail de manifestations cliniques de la tuberculose maladie). Dans ce contexte, des variants rares (pas capturés par les données de puces de génotypage), peuvent avoir des conséquences sur le phénotype bien plus importantes que les variants communs. D'autre part, ces analyses sont souvent limitées à l'étude d'un mode génétique de transmission additif, l'impact d'allèles à risque uniquement sous un mode de transmission récessif, pouvant donc être indétectable. Parmi ces différents phénomènes, nous pouvons aussi mentionner par exemple le rôle de l'épigénétique, qui étudie les mécanismes moléculaires capables de réguler l'expression des gènes sans pour autant modifier la séquence

nucléotidique dont ils découlent. Finalement, les interactions entre différents variants génétiques affectant différents gènes du même complexe protéique, de la même voie de signalisation ou d'autres, et qui ne sont pas étudiés dans les études GWAS, peuvent aussi être sous-jacentes à des natures ou des hérédités plus compliquées de la tuberculose. Dans ce contexte, des hérédités digéniques, que nous traiterons en détail dans la section **III**, ou encore des modèles oligogéniques peuvent dévoiler de nouvelles pistes sur les bases génétiques de la tuberculose.

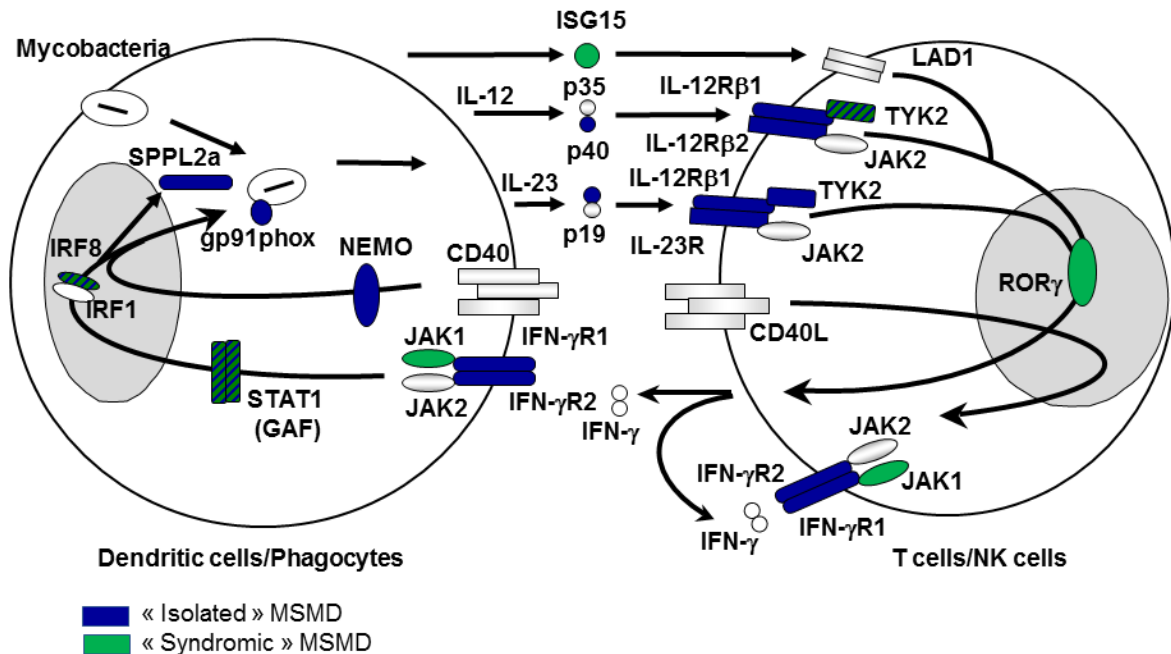
H.3. Le syndrome MSMD et les formes monogéniques de tuberculose.

La tuberculose maladie ne représente pas un phénotype uniforme et homogène. Par exemple, un phénotype plus extrême et plus grave est souvent observé chez les enfants. C'est justement à partir de la description des formes les plus sévères de tuberculose, représentant des cas pédiatriques, que le rôle de la génétique humaine dans la tuberculose maladie a été démontré par l'action directe de mutations génétiques (145–151). En 1996, des infections sévères par mycobactéries chez des patients par ailleurs sains sont montrées être causées par des lésions monogéniques innées du système immunitaire (152), suggérant le rôle d'une voie génétique liée à l'IFN- γ dans le contrôle de ces infections chez l'humain. C'est autour de la dissection génétique d'un syndrome rare appelé MSMD pour *Mendelian susceptibility to mycobacterial disease*, que cette voie immunologique, non redondante, est confirmée, caractérisant ces infections sévères causées par des mycobactéries peu virulentes.

Le syndrome MSMD débute généralement pendant l'enfance, plus rarement à l'adolescence ou à l'âge adulte, et présente diverses manifestations, allant d'infections localisées à disséminées sous l'action d'une ou de plusieurs espèces de mycobactéries. Le syndrome peut être causé par des mycobactéries environnementales, que l'on retrouve partout (l'eau, la terre, etc.), et auxquelles nous sommes tous exposés, ou encore par le vaccin BCG. Environ la moitié d'entre eux présentent aussi des infections par salmonelle. Comme nous le détaillerons par la suite, les individus de familles vivant dans des régions endémiques pour la tuberculose sont aussi susceptibles à *M. tuberculosis*. D'autres infections graves ont été rapportées chez quelques patients sporadiques, notamment des infections causées par diverses bactéries « intra-macrophagiques » (listériose, nocardiose, klebsiellose), ainsi que par des champignons (candidose, histoplasmosse, paracoccidioïdomycose, coccidioïdomycose) ou des parasites (leishmaniose, toxoplasmose). Ces infections concomitantes ont conduit à la distinction de deux sous-types de MSMD : le MSMD isolé (limité aux infections par mycobactéries) et le MSMD syndromique (plus large dans son spectre infectieux). Toutefois,

ce terme reste utile car les maladies mycobactériennes sont de loin les infections les plus courantes chez ces patients.

Figure 14: Schéma de la coopération entre les phagocytes et les lymphocytes T (ou NK) lors d'une infection mycobactérienne



En bleu, les gènes impliqués dans le syndrome MSMD dit isolé et en vert ceux impliqués dans le syndrome MSMD dit syndromique.

Depuis 1996, 13 gènes ont été identifiés dans le syndrome MSMD intervenant tous dans la voie immunitaire de la cytokine IFN- γ (153) (Figure 14). Les premières étiologies génétiques découvertes du MSMD furent des mutations bi-alléliques du gène *IFNGR1*, codant pour la chaîne de liaison au ligand du récepteur de l'IFN- γ (IFN- γ R1) (154). Depuis, des mutations au niveau des gènes *IL12B*, *IL12RB1*, *ISG15*, *TYK2*, *IRF8*, *SPPL2A*, *CYBB*, *IFNGR1*, *IFNGR2*, *STAT1*, *NEMO*, *IL23R* et *IL12RB2* ont été décrites dans la pathogénie du syndrome, certaines suivant un mode de transmission génétique dominant (par exemple, *IRF8*), d'autres récessif (par exemple, *TYK2*), d'autres dominant et récessif à la fois (par exemple, *IFNGR1*) ou encore d'autres suivant un mode de transmission récessif lié au chromosome X (par exemple, *NEMO*) (46,153). Les produits de tous ces gènes sont impliqués soit dans la production d'IFN- γ (*IL-12p40*, *IL-12R β 1*, *TYK2*, *SPPL2a*, *ISG15*), soit dans la réponse à l'IFN- γ (*IFN- γ R1*, *IFN- γ R2*, *STAT1*, *gp91phox*), soit dans les deux (*IRF8* et *NEMO*) (155–157). Parmi les lésions génétiques causales on retrouve des variants mononucléotidiques (*single nucleotide variants*, SNVs), des petites délétions, des duplications, des insertions et des grandes délétions, insertions ou duplications dites CNV (*copy number variation*). Les étiologies génétiques les plus courantes du MSMD se retrouvent notamment au niveau des gènes *IL12RB1*, *IL12B*, *IFNGR1*, *IFNGR2* et *STAT1* (153,155).

Comme nous avons mentionné avant, les patients MSMD peuvent présenter d'autres infections, en particulier, par *M. tuberculosis*. Cependant, les observations les plus intéressantes concernant la connexion entre la physiopathologie de la tuberculose et du MSMD vinrent au début des années 2000. Certaines mutations responsables du MSMD furent retrouvées chez des patients ayant des infections uniquement par *M. tuberculosis*. Les premiers cas furent décrits dans quatre études familiales indépendantes dans lesquelles, des patients atteints uniquement d'une tuberculose, ont été rapportés avec un déficit au niveau de la sous-unité $\beta 1$ du récepteur de l'interleukine-12 (IL-12R $\beta 1$). La première étude, rapportée en 2001, se focalisa sur le cas de deux frères et sœurs issus d'une famille consanguine originaire du Maroc avec un déficit en IL-12R $\beta 1$. Le frère s'était présenté avec une infection par BCG disséminée après inoculation du vaccin dans sa petite enfance, comme attendu pour ce déficit. Sa sœur fut, par contre, diagnostiquée d'une tuberculose abdominale à l'âge de 18 ans, sans antécédents d'une maladie par infection du BCG, malgré trois inoculations avec du BCG vivant durant la même période que son frère (158). Deux ans plus tard, en 2003, fut publiée l'étude de trois frères et sœurs présentant aussi un déficit en IL-12R $\beta 1$, sans diagnostic MSMD (159). La plus petite des sœurs, âgée de cinq ans, fut diagnostiquée d'une tuberculose disséminée. Une autre sœur, âgée de 13 ans, avait présenté une infection par Salmonelle à l'âge de 15 mois et fut aussi diagnostiquée d'une tuberculose à l'âge de 13 ans, tandis que l'aînée était restée asymptomatique à ses 15 ans. Une autre étude présenta le cas d'une fille âgée de onze ans qui, présentant la même déficience, fut hospitalisée en 2005 à cause d'une tuberculose disséminée, en absence d'antécédents personnels ou familiaux d'infection par d'autres espèces mycobactériennes (150). Finalement, le cas d'un homme âgé de 33 ans, déficient au niveau du gène *IL12RB1*, ayant décédé d'une tuberculose disséminée résistante à l'isoniazide et à la rifampicine, fut rapporté en 2011 (151).

Dans ce contexte, et dans le but d'estimer la proportion d'enfants développant une tuberculose sévère due à un déficit en IL-12R $\beta 1$ dans les zones où la tuberculose est endémique et où la consanguinité parentale est courante, une étude publiée en 2011, chercha à identifier des mutations au niveau du gène *IL12RB1* sur 50 patients non reliés provenant d'Iran du Maroc et de la Turquie (146). Cette étude identifia deux autres patients avec une déficience en IL-12R $\beta 1$ qui, comme pour les 50 analysés, étaient résistants au vaccin BCG et aux mycobactéries peu virulentes. Par ailleurs, d'autres défauts génétiques communs à la tuberculose et au syndrome MSMD ont été rapportés. La déficience partielle de l'*IFNGR1*, retrouvée chez un enfant atteint d'une infection disséminée par le BCG et chez son frère, atteint d'une tuberculose clinique (n'ayant pas reçu du BCG), en est une (160). Des patients

hétérozygotes de certaines mutations *STAT1*, altérant l'immunité liée à l'IFN- γ , ont aussi développé une tuberculose (161). Des mutations germinales du gène *CYBB*, le gène humain codant pour la sous-unité gp91(phox) de la NADPH oxydase du phagocyte, entraînant une maladie granulomateuse chronique liée au chromosome X, ont été retrouvées chez des familles de cas index MSMD avec certains membres atteints d'une tuberculose clinique (162). Finalement, des défauts au niveau du gène *TYK2*, sur lesquels nous reviendrons, ont aussi été retrouvés chez des patients tuberculeux (149).

En somme, ces observations reflètent la connexion physiopathologique entre la tuberculose et le syndrome MSMD, et apportent la preuve du principe que la tuberculose a une forte composante génétique de l'hôte. La découverte des premières formes monogéniques de tuberculose à partir de la dissection génétique du syndrome MSMD en est bien sûr la preuve la plus convaincante. Cependant, ce syndrome reste un syndrome très rare, tandis que la tuberculose est bien plus commune dans le monde, avec des pays qui sont endémiques pour la maladie. Les allèles associés aux cadres sévères de tuberculose sont trop peu fréquents pour expliquer un pourcentage significatif des cas de tuberculose dans le monde. Dans ce travail nous nous sommes donc intéressés à la recherche d'autres défauts génétiques impliqués dans le développement de la tuberculose maladie, dans l'espoir de contribuer à une meilleure compréhension de la physiopathologie de celle-ci.

I. Objectifs de la thèse.

De façon générale, le « laboratoire de génétique humaine des maladies infectieuses » dans lequel je réalise ma thèse conduit des recherches visant à identifier les variants génétiques impliqués dans le développement des maladies infectieuses. Le laboratoire a en particulier identifié la plupart des causes génétiques du MSMD (46,163–168) ainsi que le rôle, comme précédemment mentionné, de certains variants rares dans le développement de la tuberculose maladie (147,169,170). L'ensemble de ces recherches a été considérablement transformé avec l'arrivée des technologies de séquençage de nouvelle génération (*next generation sequencing*, NGS), en particulier au niveau des régions codantes du génome appelé exome (171). L'utilisation de données de séquençage à haut débit, notamment le séquençage d'exome entier (*whole exome sequencing*, WES), a conduit à des découvertes majeures dans le domaine de la génétique humaine, spécialement pour les pathologies monogéniques (172–175). Le laboratoire a réalisé plus de 5,000 exomes sur des patients présentant des infections particulièrement sévères. Les recherches menées sur ces exomes ont suivi jusqu'à présent une hypothèse monogénique ayant permis l'identification de mutations

dans des gènes nouveaux dans des pathologies mycobactériennes, bactériennes, virales et fongiques (176–181).

Le cadre général de mon projet se situe dans la recherche de gènes contribuant à la prédisposition ou à la résistance à la tuberculose en se servant des exomes non seulement de patients tuberculeux mais aussi de ceux de patients MSMD. Comme mentionné tout au long de cette introduction, il s'agit en particulier de comprendre pourquoi, lorsque des individus sont infectés par *M. tuberculosis*, certains (~10%) vont développer la maladie clinique (qui peut être plus ou moins sévère) alors que d'autres (~90%) ne présentent aucune symptomatologie (infection latente). En se basant sur les découvertes majeures dans le domaine de la génétique humaine utilisant les données récentes de NGS (172–175), notre principale hypothèse est qu'il existe des causes génétiques de la tuberculose impliquant d'autres variants rares ou à fréquence intermédiaire (<5% de porteurs du génotype à risque) avec des effets individuels forts (OR >5) et identifiables à partir de données d'exomes.

Le laboratoire dispose d'une cohorte de 475 exomes (WES) de patients atteints de tuberculose. Il dispose aussi depuis quelques années d'une cohorte unique de 479 exomes de patients MSMD dont les analyses initiales n'ont pas permis d'identifier des causes génétiques connues. Il s'avère que pour environ la moitié des patients MSMD, l'étiologie génétique sous-jacente demeure inconnue. L'objectif général de ma thèse consiste à identifier, à partir de ces données d'exomes, des nouvelles causes génétiques de la tuberculose et/ou du MSMD, suivant l'hypothèse émise précédemment. La recherche de nouvelles étiologies génétiques du syndrome MSMD peut par ailleurs améliorer la compréhension actuelle des bases génétiques de la tuberculose en éclaircissant certains ou en faisant apparaître de nouveaux mécanismes biologiques sous-jacents. Spécifiquement, je recherche le rôle de ces variants par des analyses suivant deux hypothèses générales de travail :

- Monogénique, par une approche gène candidat, basée notamment sur la recherche de gènes impliqués dans la voie immunologique de l'IFN- γ et des études précédentes du syndrome MSMD, en particulier du gène *TYK2*.
- Digénique, nécessitant le développement d'une méthode originale et son application aux données de tuberculose ; dans cette approche la cohorte de patients MSMD sera aussi utilisé pour tester la méthode dans une pathologie avec des bases génétiques mieux connues.

Si les méthodes d'analyses déployées relèvent principalement de la statistique génétique et de la bio-informatique, l'étroite collaboration avec notre équipe de génétique immunologique nous permet ensuite de valider expérimentalement nos découvertes.

Dissection des bases génétiques de la tuberculose par analyse d'exomes : approche monogénique.

Les données NGS ont révolutionné l'étude des maladies mendéliennes et celle des maladies communes, en particulier, par l'analyse systématique de variants rares (< 1%) dans le contexte d'analyses d'associations. Sur la base de ces données, mon travail de thèse a porté de façon générale sur l'étude d'infections mycobactériennes, à la fois sur une maladie rare, par l'étude d'une famille consanguine de membres atteints du syndrome MSMD (décrite dans la section **II_B**), à partir d'une approche combinant analyse de liaison et analyse d'exome, et sur une maladie bien plus commune, la tuberculose, par des études d'association de variants (section **II_C**) au sein d'une grande cohorte WES (décrite dans la section **II_A**).

A. La cohorte d'étude: HGID.

Avec l'arrivée des données WES, le laboratoire mit au point de nouvelles méthodes de travail dans la recherche de variants pathogéniques. Dans un premier temps, les méthodes furent basées sur des approches gènes-candidats, permettant l'identification de mutations au sein de nouveaux gènes dans des pathologies mycobactériennes, bactériennes, virales et fongiques (177,178,180,181). Cependant, pour la plupart des individus séquencés, sélectionnés à partir d'échantillons provenant de diverses régions du globe et recrutés à l'aide de cliniciens autour du monde, la (ou les) supposée(s) mutation(s) génétique(s) impliquée(s) dans le développement de la maladie sous-jacente n'a (n'ont) pas su être mis en évidence.

Le séquençage des individus donna lieu à la création d'une base de données WES, regroupant tous les exomes des plus de 20 cohortes de pathologies infectieuses du laboratoire. Je débutai la recherche de nouvelles étiologies génétiques du syndrome MSMD et/ou de la tuberculose à partir de cette base de données, dénommée *Human Genetics of Infectious Diseases* (HGID), qui regroupait plus de 4,000 individus au début de ma thèse. Elle était composée, en particulier, de 463 patients MSMD, de 291 enfants atteints de tuberculose provenant de diverses régions du monde, et de 163 adultes d'origine Nord-Africaine atteints d'une tuberculose pulmonaire précoce (< 40 ans). Aucune mutation pathogène dans les gènes connus responsables du MSMD et de la tuberculose (Figure 14) n'avait été décrite pour ces patients. Parmi le restant des exomes sans étiologie génétique connue de la cohorte HGID, 2,835 patients, atteints de diverses infections non-mycobactériennes et appartenant à diverses origines ethniques (en particulier, celles des cas), furent choisis comme témoins pour les analyses d'association présentées en **II_C**. Les groupes phénotypiques majeurs des 3,752

individus analysés sont inscrits dans le Tableau 1. Il est à noter que malgré l'hétérogénéité phénotypique et ethnique (voir **II_A.3**) des 2,835 témoins, ce groupe servit raisonnablement aux analyses de la section **II_C**. Cette cohorte fut d'ailleurs utilisée dans les analyses de simulation de la section **III**.

Tableau 1: Nombres d'individus appartenant aux principaux groupes phénotypiques de la cohorte HGID

	Infections virales	Infections par bactéries	Infections fongiques	Plusieurs maladies du système immunitaire	Autres maladies	Individus sains	Total Témoins	TB	MSMD
Total	1,290	423	278	395	320	129	2,835	454	463

Les témoins correspondent aux témoins utilisés pour les analyses d'association pour des patients avec des infections par mycobactéries.

A.1. Séquençage des individus, leur contrôle qualité et celui des variants.

Les individus de la cohorte HGID ont été séquençés tout au long des huit dernières années, suivant l'évolution de différents kits de capture. Le séquençage des exomes a été réalisé selon les procédures suivantes. L'ADN génomique des cellules mononuclées du sang périphérique a été d'abord extrait et fragmenté avec un *ultrasonicator Covaris S2*. La procédure s'est suivie par la capture d'exome, sachant qu'elle a pu être réalisée alternativement avec les kits SureSelect Human All Exon de 37-, 50- ou 71-Mb (Agilent Technologies), selon l'année où celle-ci a été faite. Le tout a été suivi d'une parallélisation massive WES sur une machine HiSeq 2000 ou 2500 (Illumina), générant des lectures de 72, 100 ou 125 bases, respectivement. Les lectures ont été alignées sur le génome de référence humain hg19 en utilisant l'algorithme de correspondances exactes maximales de Burrows – Wheeler Aligner (BWA). Ces données de lecture, alignées sur le génome humain et stockées sous forme de fichiers binaires, communément appelés fichiers BAM, ont ensuite été traitées en suivant les recommandations du logiciel GATK (*Genome Analysis Toolkit*) (182), versions 3.4 à 46, disponibles sur leur site (<https://software.broadinstitute.org/gatk/best-practices/>). Les « reads » identifiés comme duplicats (duplicats moléculaires de PCR ou duplicats optiques) ont été éliminés grâce au logiciel Picard (<http://broadinstitute.github.io/picard>), permettant la conduction d'une seconde étape d'ajustement local, consistant à réaliser l'alignement autour des petites insertions ou délétions (indels) à l'aide du logiciel GATK. Une dernière étape a permis de corriger les artefacts de séquençage en appliquant l'outil « Base Quality Score Recalibrator » de GATK. GATK HaplotypeCaller a été utilisé pour identifier les appels de variants au sein des régions couvertes par le kit de capture auxquelles des marges de 200

paires de base en amont et en aval ont été rajoutées. Les données ont été finalement générées dans le format de fichier gVCF (*Genome Variant Call Format*) qui permet de conserver les informations de qualité de chaque position couverte.

Ces gVCF sont regroupés dans un seul fichier de format VCF (*Variant Call Format*) sur lequel, en particulier, nous appliquons un processus de contrôle qualité. Spécifiquement pour mes analyses, les variants avec peu de lectures (*read depth*, $DP < 8$), avec une importante différence dans le nombre de « reads » entre un allèle et l'autre au même locus quand les deux allèles ont été lus, mesurée par la proportion de « reads » de l'allèle moins lu (*minor read ratio*, $MRR < 0.2$), de basse qualité (*genotype quality*, $GQ < 40$) ou filtrés par gnomAD r2.0.2 (Exome Aggregation Consortium, Broad Institute), ont été exclus. La recalibration VQSR a été appliquée sur notre cohorte avec VariantRecalibrator et ApplyRecalibration pour les SNPs, et avec *ts_filter_level* défini sur 99.0 et d'autres paramètres spécifiés par les recommandations de GATK pour les indels. Les SNPs présentant un taux de génotypage $< 80\%$ ont été éliminés. Il est à noter que le seuil choisi pour le taux de génotypage minimum accepté est en accord avec la diversité des kits de capture utilisés dans l'ensemble de la cohorte. L'écart à l'équilibre d'Hardy-Weinberg de ces SNPs autosomiques n'a pas été pris en compte dans les critères qualité en raison de la forte hétérogénéité d'origines ethniques de la cohorte. Précédant la sélection des 3,752 cas et témoins pour les analyses d'enrichissement, le software PLINK 1.9 (183) a été utilisé pour détecter des potentielles contaminations, des échanges ou duplications d'échantillons, ainsi que des relations familiales inconnues (par exemple, des paires de frères et sœurs ou enfant / parent). A l'aide des estimations « identité par état » (*identity-by-state*, IBS) à l'échelle du génome entier, pour toutes les paires d'individus, nous avons pu garantir que notre cohorte cas-témoin soit constituée d'individus partageant un très faible degré d'apparentement (les doublons et les individus apparentés au moins au deuxième degré ont été exclus).

A.2. Annotation des variants.

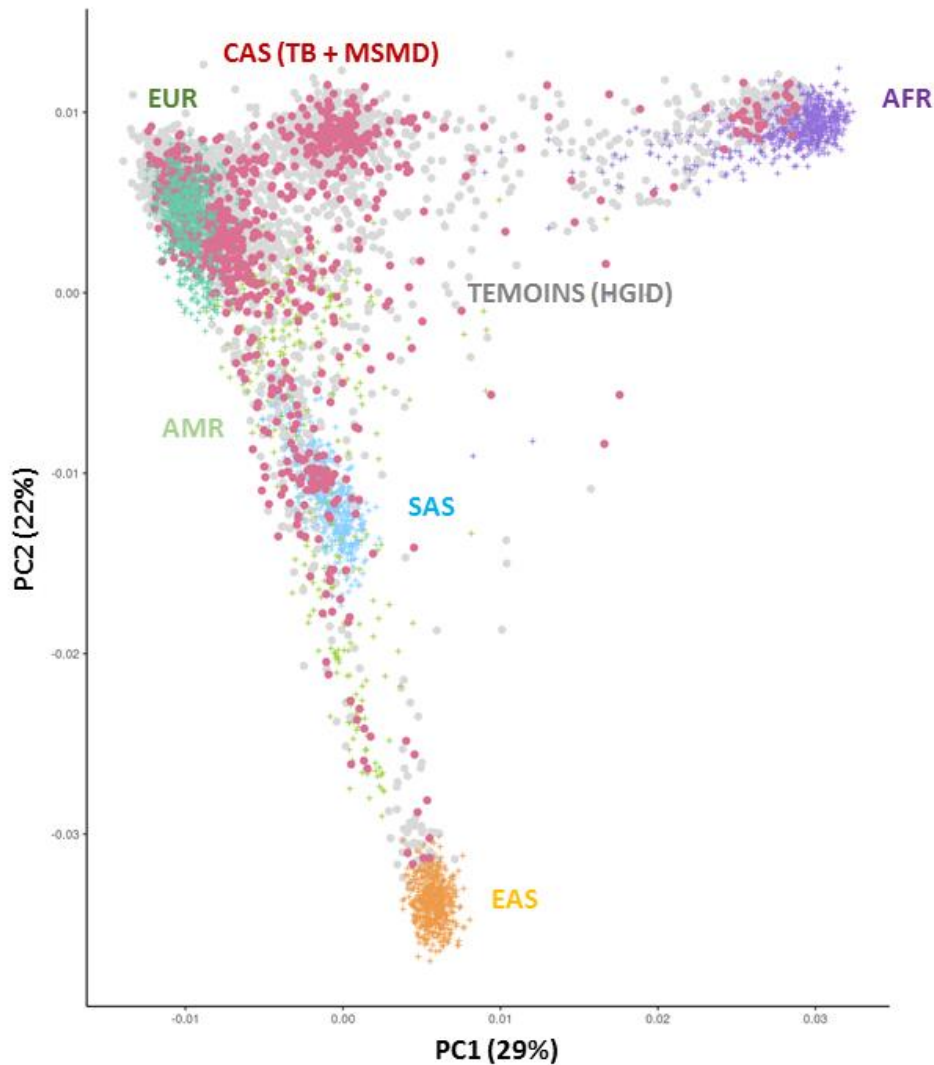
L'identifiant de chaque variant a été déterminé en utilisant la base de données publique dbSNP (build 138), développée par le « National Center for Biotechnology Information » (NCBI) (184) qui recense la majorité des variations génétiques connues, telles que les SNPs, les indels, les marqueurs microsatellites, les séquences répétées en tandem et les MNPs (« Multinucleotide polymorphisms »). Les variants absents de cette base de données ont été considérés comme de nouveaux variants. Tous les variants ont ensuite été annotés à l'aide du logiciel SnpEff (185) sur la base de leur localisation génomique par rapport à un ensemble de

gènes et transcrits de référence du projet Ensembl (version GRCh37.75 - hg19). Ce logiciel permet de caractériser le positionnement du variant dans son contexte génique (exonique, intronique, 5' UTR, 3'UTR, intergénique...) ainsi que les conséquences des variants sur la séquence protéique (i.e. mutations synonymes, faux-sens, non-sens, insertions/délétions avec ou sans décalage du cadre de lecture, création d'un site d'épissage ...). Ces annotations sont par exemple cruciales dans la définition des groupes de variants d'intérêt que l'on souhaitera garder pour des analyses d'association réalisées en particulier dans la section **III**.

A.3. Analyse de la cohorte HGID en composantes principales

La forte hétérogénéité d'origines ethniques présentes dans la cohorte HGID a pu être illustrée grâce à une analyse de structure de population ou analyse en composantes principales (*principal component analysis*, PCA), réalisée à l'aide de 22,612 marqueurs fréquents (> 5%) et communs avec le panel de référence du projet 1000 Génomes (www.internationalgenome.org/, 1000G). Par élimination des variants en fort LD ($r^2 > 0.6$) à l'aide de fenêtres glissantes de 100 Kb l'analyse a été restreinte à des variants indépendants. De façon générale, le principe de la méthode est de déterminer les principaux axes de variation génétique (composantes principales) de l'échantillon propre en le combinant à l'échantillon de référence et d'assigner à chaque individu de l'échantillon de référence comme de l'échantillon à analyser des coordonnées selon chacun des axes de variation. Encore une fois, PLINK 1.9 a été utilisé pour estimer la PCA basée sur la matrice de relations standardisée par variance. La Figure 15 montre la structure de population observée dans notre cohorte quand celle-ci est combinée à celle de 1000G, composée de données de séquençage de génome entier (*whole genome sequencing*, WGS) de 2,504 individus « sains » et non apparentés, appartenant à cinq groupes ethniques majeures, à savoir, Européen, Africain, Américain, de l'Asie du Sud et de l'Est. Une estimation faite à partir des données extraites des composantes principales a montré que la cohorte HGID était composée de deux groupes ethniques majeurs : 43% des individus étaient descendants d'Européens tandis que 31% avaient des ancêtres Nord-Africains. Ces pourcentages étaient par ailleurs équitablement répartis entre les cas (patients avec des infections par mycobactéries) et les témoins (patients avec d'autres infections non-mycobactériennes) de l'analyse en **II_C**, avec 48% des cas et 42% des témoins d'origine Européenne, respectivement, et 30% des cas et 31% des témoins d'origine Nord-Africaine, respectivement. Le plus des 20% restant des individus de la cohorte sont visuellement regroupés, sur la PCA de la Figure 15, autour des quatre autres groupes ethniques majeurs de la base de données 1000G.

Figure 15: PCA des 2,504 individus "sains" de 1000G d'origines ancestrales connues (en fonction des cinq populations de 1000G, Africaine (AFR), Américaine (AMR), Asiatique de l'Est (EAS), Européenne (EUR), Sud Asiatique (SAS)), et des patients avec (n=917, en rouge) ou sans (n=2,835, en gris) des infections par mycobactéries de la cohorte HGID



Les proportions de la variance expliquées par la première (PC1) et la deuxième (PC2) composante principale sont indiquées entre parenthèses.

A.4. Etudes d'association par analyse d'exome : l'exemple de la pneumopathie grippale sévère.

Parmi les diverses sous-cohortes de pathologies infectieuses de la cohorte HGID, je me suis principalement intéressé à celles regroupant des pathologies mycobactériennes, que je décris dans les sections II_B, II_C et III_E.3. Cependant, j'ai également eu l'opportunité de réaliser des analyses d'association sur des données d'exome sur une cohorte présentant une autre pathologie pulmonaire infectieuse. Plus spécifiquement, il s'agissait de l'étude de 25

enfants non reliés de la cohorte HGID atteints d'une pneumopathie grippale sévère. A ce jour, trois types de virus de la grippe infectent les humains: le virus de la grippe A (VGA), le virus de la grippe B et le virus de la grippe C. Deux sous-types de VGA, qui est le plus virulent parmi les trois virus de la grippe, H1N1 et H3N2, circulent actuellement chez l'humain (186), causant des symptomatologies grippales classiques. Rarement, les épisodes de grippe peuvent entraîner une pneumopathie sévère, se traduisant par un syndrome de détresse respiratoire aiguë (*acute respiratory distress syndrome* [ARDS]; (187)) dont la prévalence est de l'ordre de 4 sur 100,000 (OMS 2018). Des études précédentes ont rapporté le cas d'individus en apparence sains et résistants à d'autres agents infectieux, développant un ARDS associé au VGA (188). Cependant, la pathogenèse des syndromes décrits demeure largement méconnue.

En 2015, le laboratoire a identifié le déficit autosomique récessif complet en IRF7 comme la première étiologie génétique humaine de l'ARDS de la grippe chez un enfant par ailleurs en bonne santé (189). Lors d'une infection par le VGA, le déficit en IRF7 altère la production d'IFN- α/β et d'IFN- λ (189). De plus, IRF7 est requis pour l'amplification des deux types d'interférons antiviraux. Nous avons donc émis l'hypothèse que l'intégralité du circuit associé à l'IFN- α/β et - λ était requis pour l'immunité protectrice contre les virus de la grippe dans les poumons (190). Pour tester notre hypothèse, nous avons analysé les exomes de 25 enfants non apparentés atteints de l'ARDS de la grippe, sans mutations en *IRF7*, en nous focalisant sur des gènes liés à l'immunité du circuit IFN- α/β et - λ .

Pour l'analyse d'exome nous avons considéré uniquement des variants codants non-synonymes rares (MAF <0.001 dans la base de données publique gnomAD et dans la cohorte HGID), prédits délétères par le score CADD (CADD > MSC, (191)) et présents dans des gènes étroitement reliés à *IRF7*. Pour cette dernière condition je me suis basé sur le « connectome » de gènes humains (192), qui est lui-même construit à partir d'un algorithme cherchant la plus courte distance entre paires de gènes (dans un sens biologique). J'ai retenu ceux avec le plus haut degré de connectivité avec *IRF7* (*p value* <0.01), réduisant l'analyse à l'étude de 158 gènes. Suivant les filtres décrits et sous les hypothèses d'homogénéité génétique et de mode génétique de transmission dominant, j'ai comparé, pour chaque gène, la proportion de variants présents chez les cas avec celle de 1,485 individus sans infection virale de la cohorte HGID (Tableau 2). Les résultats ont été ajustés à l'aide des trois premières composantes principales de la PCA pour tenir compte de l'hétérogénéité génétique de la cohorte analysée.

Cette approche a identifié *TLR3* comme étant le meilleur gène candidat ($p = 0.00052$; $OR=24.82$, $IC = [5.05,96.83]$; Tableau 2). Parmi les trois cas porteurs de variants du gène *TLR3*, deux sont hétérozygotes pour la mutation faux-sens P554S ($MAF = 0.0004$ dans gnomAD) et un troisième l'est pour la mutation faux-sens P680L (absent de gnomAD). Suite à la reconnaissance de motifs viraux, TLR3 déclenche une voie de signalisation aboutissant à l'induction de l'IFN- α/β et $-\lambda$ (193). Il est à noter que le laboratoire avait précédemment identifié le rôle pathogénique de certains variants du gène *TLR3* chez des patients atteints d'une encéphalite herpétique, qui est causée par un autre virus, en l'occurrence, HSV-1 (*human simplex virus 1*; (194–196)). En effet, P554S a été aussi retrouvé chez quatre patients de la cohorte HGID avec une encéphalite herpétique. Des analyses expérimentales précédentes ont prouvé que ce variant était perte de fonction (*loss of function*, LOF) agissant de manière dominante négative (194). Sa récurrence chez deux enfants non apparentés atteints de l'ARDS renforce encore son rôle causal. Il est aussi à noter que dans l'encéphalite herpétique le variant avait été associé à une pénétrance incomplète, ce qui est également observé dans cette étude où cinq membres des familles impliquées sont porteurs asymptomatiques. Pour le variant P680L, qui n'est observé que chez le patient avec un ARDS de la grippe, le groupe de génétique immunologique du laboratoire a prouvé qu'il était LOF pour l'induction des interférons antiviraux (197).

En somme, cette étude suggère la déficience autosomique dominante de TLR3 comme une nouvelle étiologie génétique des pneumopathies grippales sévères. Des erreurs innées de l'immunité associée à l'IFN- α/β et $-\lambda$ peuvent donc être sous-jacentes à l'ARDS de la grippe chez l'humain. Notre résultat s'avère de même intéressant suite à la description de six patients atteints d'une encéphalite herpétique présentant un déficit en TLR3 (194–196). Tous semblent être normalement résistants à d'autres virus courants, y compris le VGA (194,195,198–201). De plus, les trois patients ici décrits, présentant un déficit autosomique dominant en TLR3 atteints de l'ARDS de la grippe, n'ont pas développé une encéphalite herpétique, même si l'un d'entre eux était séropositif pour l'HSV-1. L'observation selon laquelle un déficit en TLR3 peut être à la base d'une encéphalite herpétique ou d'un ARDS de la grippe chez des enfants normalement en bonne santé, respectivement résistants au VGA et à l'HSV-1, est en accord non seulement avec la pénétrance clinique incomplète observée pour les deux infections virales chez les hétérozygotes *TLR3*, mais également avec les deux maladies infectieuses étant typiquement sporadiques et mutuellement exclusives dans la

population générale. Cette étude élargit le spectre de mutations TLR3, d'une part du fait de leur implication sur un autre virus, et d'autre part par la description de la déficience sur un autre tissu, en l'occurrence, les poumons. Elle confirme également le spectre étroit de la déficience TLR3 chez l'humain, même si la spécificité virale et tissulaire reste largement à explorer. D'autres études génétiques et immunologiques humaines de diverses maladies infectieuses isolées seront nécessaires pour clarifier les rôles de TLR3, de l'IFN- α/β et - λ et d'autres voies de défense de l'hôte dans l'immunité aux infections virales et leur spécificité tissulaire.

Tableau 2: Classement des cinq gènes les plus enrichis en variants rares liés à *IRF7* dans la cohorte ARDS par rapport à 1,485 témoins sans infections virales.

Gène	Témoins porteurs (1,485 exomes)	Cas porteurs (25 exomes)	<i>P value</i>	OR	CI_2.5%	CI_97.5%
<i>TLR3</i>	11	3	0.00052	24.82	5.05	96.83
<i>HLA- DRB5</i>	16	2	0.0312	8.52	1.27	33.99
<i>PML</i>	56	3	0.0687	3.87	0.88	11.93
<i>MX1</i>	31	2	0.0738	5.35	0.82	20.22
<i>TRIM21</i>	6	1	0.1129	9.75	0.48	70.50

B. L'étude spécifique d'une famille consanguine d'origine libanaise par une approche combinant analyse de liaison et analyse d'exome.

J'ai débuté la recherche de nouveaux gènes responsables de pathologies mycobactériennes avec l'étude du cas spécifique de deux patients MSMD issus d'une même famille de parents consanguins et résidant au Koweït. Je me suis focalisé sur une approche combinant analyse de liaison, méthode classique de la génétique épidémiologique, sur laquelle nous reviendrons, et analyse d'exome. Il est à noter que la consanguinité représente une situation dans laquelle des individus se marient avec des apparentés. Les mariages consanguins sont pratiqués largement encore dans certaines populations et régions du monde. Ces familles nous intéressent spécialement car la descendance des unions consanguines peut être exposée à un risque accru de pathologies génétiques en raison de l'expression de

mutations génétiques récessives héritées d'un ancêtre commun. Plus la relation biologique entre parents est étroite, plus grande est la probabilité que leur descendance ait reçu une ou deux copies d'un variant délétère (202). Par exemple, les cousins germains partagent en moyenne 12.5% (1/8) de leurs matériels génétiques, ce qui fait que leur descendance est homozygote en moyenne sur 6.25% (1/16) de loci (c'est-à-dire qu'ils recevront des copies génétiques identiques de chaque parent à ces sites dans leur génome) (203).

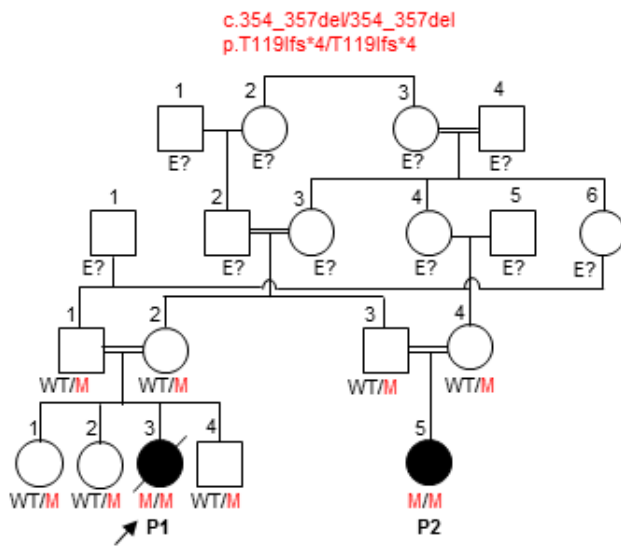
B.1. Echantillon d'étude.

La famille étudiée dans cette section est originaire du Liban et provient du Koweït. Deux membres de cette famille développèrent le syndrome MSMD suite à la vaccination avec le BCG. La Figure 16 représente la structure familiale et les mariages consanguins à chaque génération. Les cas P1 et P2, au bas de l'arbre généalogique, sont les membres atteints de la maladie.

P1 est le troisième enfant né d'une famille consanguine sans aucune complication périnatale. Elle a été vaccinée avec le BCG à l'âge de 3 mois, et trois semaines plus tard, elle a commencé à avoir une fièvre prolongée associée à une masse axillaire gauche importante (8 x 8 cm), érythémateuse et sensible. Elle a été traitée avec plusieurs antibiotiques sans signes d'amélioration. La biopsie par excision des ganglions lymphatiques a révélé des bacilles acido-résistants confirmant le diagnostic de maladie mycobactérienne disséminée (BCG-ose) pour laquelle elle avait été traitée initialement avec de l'amikacine et de la lévofloxaciline. Ceux-ci ont ensuite été modifiés en isoniazide, rifampicine et éthambutol. Aucune autre infection grave causée par d'autres bactéries ou virus n'a été notée. Cependant, elle a été testée séropositive par VirScan pour l'adénovirus humain C, le virus de l'herpès humain 7, le métapneumovirus humain, l'entérovirus C et le rhinovirus A.

P2, née sans complications périnatales, est cousine germaine de P1. Six semaines après avoir été vaccinée par le BCG, administré à l'âge de 3 mois, elle a présenté une fièvre associée à un gonflement axillaire et à une lésion ulcérée au site d'injection. Elle a de même été testée séropositive par VirScan pour plusieurs infections virales, notamment les adénovirus humains C et E, le virus de l'herpès humain (HHV) 7, le virus syncytial respiratoire humain, l'Enterovirus C, le Rhinovirus A et le *Mycoplasma pneumoniae*. Une biopsie du ganglion lymphatique effectuée par excision a confirmé la présence de bacilles acido-résistants. En conséquence, le patient a été diagnostiqué d'une BCG-ose. Elle a été traitée avec de l'isoniazide, de la rifampicine et de l'éthambutol.

Figure 16: Arbre généalogique de la famille consanguine montrant la ségrégation de l'allèle mutant c.354_357del (p.T119Ifs*4) du gène *IFNG*.



Les patients atteints sont représentés par des cercles noirs remplis. Une flèche indique le cas index. Les individus de génotype inconnu sont désignés par « E? ».

B.2. Méthode statistique d'analyse.

L'étude de la famille consanguine s'est faite par une approche combinant analyse de liaison, appliquée à des données de puce de génotypage, et analyse d'exome, appliquée à des données WES de la cohorte HGID. L'analyse de liaison est une approche traditionnelle de la génétique épidémiologique utilisée pour la cartographie de mutations pathogéniques. Elle consiste à identifier les régions chromosomiques contenant des gènes associés à la maladie étudiée, en analysant la co-ségrégation de marqueurs génétiques et de la maladie, au sein de familles avec des individus atteints (204). La co-ségrégation correspond à la tendance des variants génétiques étroitement liés à être hérités ensemble au niveau chromosomique. Pour notre analyse, nous avons utilisé une méthode de cartographie par homozygotie (*homozygosity mapping*), spécialement pertinente pour la cartographie de traits récessifs issus de familles consanguines. L'analyse est basée sur l'estimation d'haplotypes, permettant la recherche de régions d'homozygotie (des régions suffisamment larges de loci homozygotes) qui ne sont partagées que par les membres atteints de la famille. Elle est basée sur une analyse de liaison à l'échelle du génome entier en se servant de SNPs en tant que marqueurs génétiques. Avec le logiciel Merlin (205), nous avons utilisé une méthode paramétrique nécessitant la spécification d'un modèle génétique (ici, modèle récessif) ainsi que d'un vecteur de pénétrance pour les trois génotypes du variant causal (ici, modèle à pénétrance complète pour le génotype homozygote de l'allèle mineur).

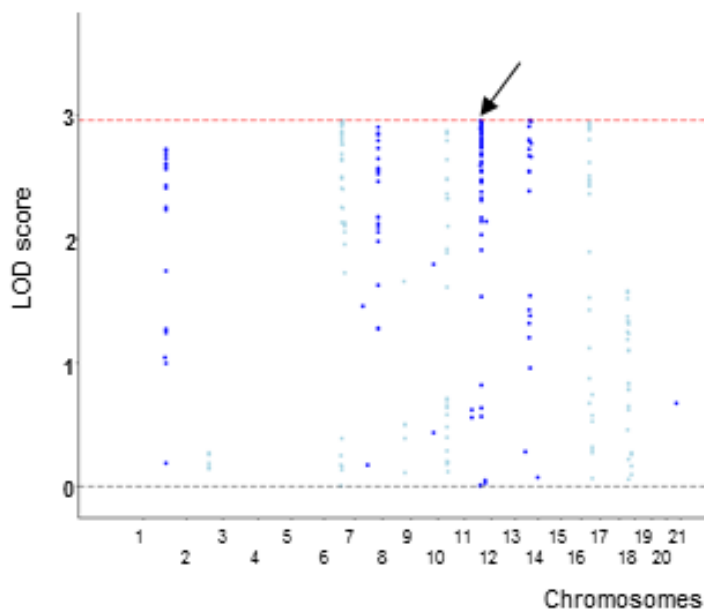
Concrètement, l'analyse de liaison a été réalisée selon les étapes décrites ci-après. Les deux patients ont été génotypés avec la puce Affymetrix Genome-wide Mapping 6.0 tandis que les parents de chaque famille nucléaire ont été génotypés avec la puce Affymetrix Genome-wide 250K. L'appel des génotypes a été réalisé avec le logiciel Affymetrix Power Tools

(http://www.affymetrix.com/estore/partners_programs/programs/developer/tools/powertools.aff). Les SNPs présentant plus d'une incohérence mendélienne ont été exclus. L'analyse de liaison s'est servie des 163,807 marqueurs autosomiques restants et communs aux patients et aux parents. Nous avons supposé une transmission autosomique récessive avec pénétrance complète et une fréquence de l'allèle causal à 0.001. Les fondateurs de la famille et les Européens non apparentés de HapMap CEU (<http://www.hapmap.org>) ont été utilisés pour estimer les fréquences alléliques ainsi que les blocs alléliques en déséquilibre de liaison avec un seuil de r^2 à 0.4. Les régions de liaison chromosomiques ont été définies à partir du score maximale (LOD_{max}) obtenu par la statistique de test appelée *LOD-score*. Elles incluent tous les variants avec un *LOD-score* supérieur à $0.9 \times LOD_{max}$, c'est-à-dire dans notre cas, 2.6587. Après l'analyse de liaison, nous avons conduit une analyse d'exome sur les régions chromosomiques obtenues qui est présentée ci-dessous.

B.3. Résultats.

Tout d'abord, nous avons confirmé le degré d'apparentement des parents de la famille par étude du taux d'homozygotie des patients : 6.8% pour P1 et 4.8% pour P2. A partir de l'analyse de liaison, six régions chromosomiques, s'étendant sur un total de 12.16 Mb, ont été identifiées sur les chromosomes 7, 8, 11, 12, 14 et 17, avec un *LOD-score* proche (> 2.6587) de la valeur maximale attendue (2.959) (Figure 17).

Figure 17: Résultat graphique des régions chromosomiques obtenues à partir de l'analyse de liaison

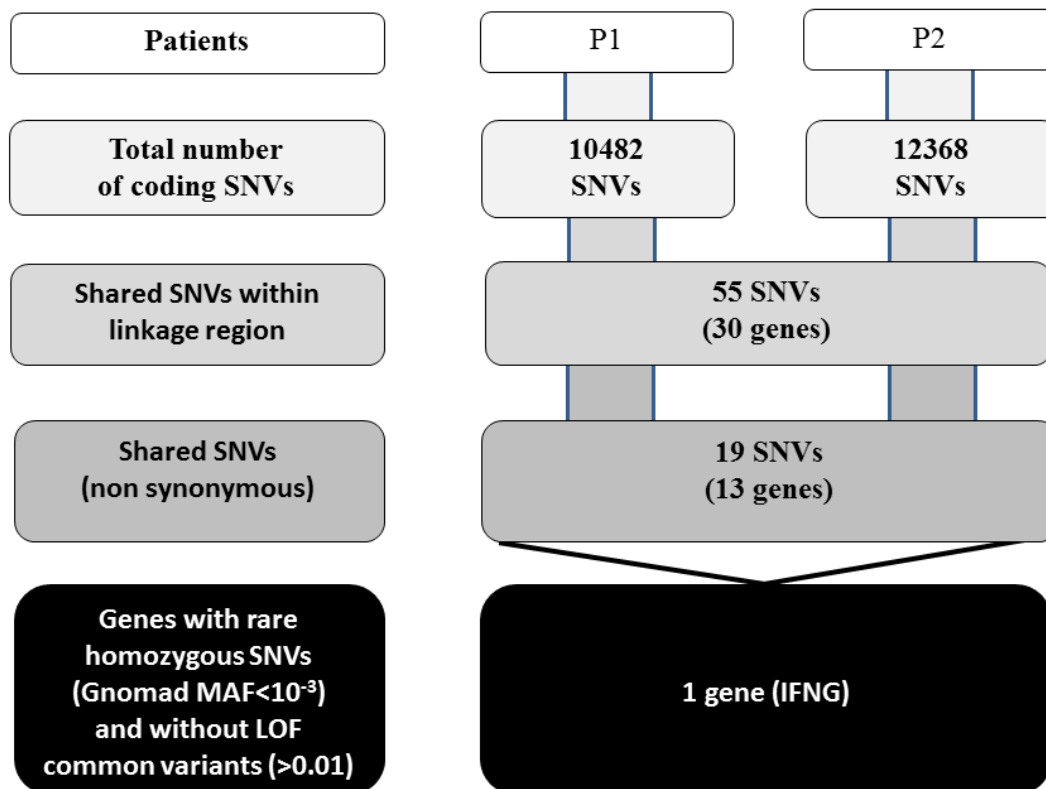


L'analyse a permis d'identifier six régions avec un LOD-score proche du LOD-score maximal (incluant la région contenant *IFNG* indiquée par une flèche).

L'analyse WES ne se focalisant que sur les variants codants non-synonymes ou affectant un site d'épissage essentiel, uniquement présents en deux copies chez les patients, permit d'identifier, parmi ces six régions, 19 variants (Figure 18). Etant donné que la prévalence de la maladie est très rare (sûrement inférieure à 10^{-6} , en accord avec les autres causes génétiques du syndrome), nous n'avons considéré, parmi les 19 variants, que les deux avec une fréquence de l'allèle mineur (*minor allele frequency*, MAF) < 0.001 . L'un d'eux, situé sur le chromosome 7, correspond à la mutation, c.494C>A ou p.S165* (rs3694405900, fréquence gnomAD de 4.06×10^{-6}) du gène *CD36*, qui est un récepteur intervenant dans l'absorption phagocytaire de plusieurs agents pathogènes, y compris Plasmodium (206). La mutation est prédite, par le logiciel snpEff, de créer un *codon stop* prématuré, pouvant résulter dans un déficit au niveau de la protéine. Cependant, d'autres variants pouvant entraîner une perte de fonction sont relativement fréquents dans *CD36*. En particulier, le variant faux-sens Y325X, présent en Afrique à 9% de fréquence dans la base de données gnomAD (dont 103 homozygotes), suggère que la déficience complète en *CD36* serait compatible avec une vie saine en Afrique (207). D'après ces observations nous avons donc décidé d'écartier ce variant. Le deuxième variant identifié chez les deux patients est situé sur le chromosome 12 et correspond à la mutation *frameshift*, c.354_357del de l'exon 3 d'*IFNG*, qui code pour la protéine IFN- γ . L'IFN- γ , est, comme largement décrit en **I**, une cytokine cruciale pour l'immunité contre les mycobactéries. Elle se situe au centre de l'axe immunologique de tous les autres défauts génétiques décrits pour le syndrome MSMD, qui altèrent soit la réponse à

cette cytokine, soit sa production. Paradoxalement, à ce jour, aucune mutation sur le gène *IFNG* n'avait été décrite sur des patients présentant des infections par mycobactéries. Se basant sur nos travaux précédents, cette mutation est d'un intérêt particulier.

Figure 18: Analyse d'exome qui a permis d'identifier un variant rare homozygote d'*IFNG* dans une des régions de liaison génétique

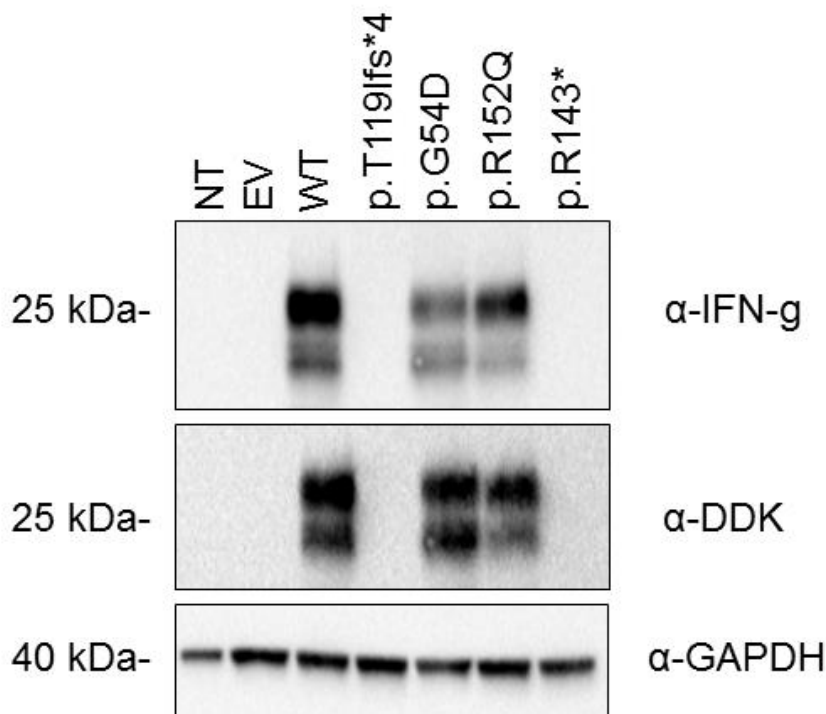


Nous avons confirmé la mutation *IFNG* c.354_357del par séquençage Sanger ainsi que l'hétérozygotie des parents des deux patients et de la sœur de P1. Cette observation est compatible avec le trait autosomique récessif à pénétrance complète que nous avons supposé dans l'analyse de liaison (Figure 16). Le variant est en outre absent de notre cohorte WES interne et des bases de données publiques gnomAD et BRAVO, indiquant que la MAF de l'allèle devrait être inférieure à 10^{-6} dans la population globale. Le score CADD (<https://cadd.gs.washington.edu/>), utilisé pour estimer la sévérité de la variation génétique, est de 25 (c.354_357del) ce qui est bien supérieur au seuil de significativité de la mutation (*mutation significance cutoff*, MSC) de 13.41 (191). Il n'y a pas de porteurs homozygotes de variants prédits LOF d'*IFNG* dans les bases de données publiques gnomAD ou BRAVO. Il y a pourtant deux homozygotes pour deux variant faux-sens, p.G54D et p.R152Q. Ils ont été désormais inclus dans les analyses expérimentales que nous décrivons par la suite. Au total, tant la ségrégation familiale de l'allèle *IFNG* c.354_357del (Figure 16) comme les valeurs

généétiques indiquées (MAF, CADD, etc.), suggèrent fortement que ces deux patients atteints de MSMD présentent un déficit complet autosomique récessif associé à l'IFN- γ .

Pour vérifier cette hypothèse, le groupe de génétique immunologique du laboratoire a conduit une étude approfondie de l'allèle *frameshift* retrouvé, en le comparant aux deux autres variants codants p.G54D et p.R152Q. Dans cette analyse, ils ont observé que l'expression de l'allèle mutant c.354_357del de l'*IFNG* était nulle, tandis que celle des autres deux mutants p.G54D et p.R152Q était comparable à celle de l'*IFNG wild type* (Figure 19). Une troisième mutation (p.R143*), étudiée précédemment dans une étude fonctionnelle, absente des bases de données publiques, a été aussi incluse dans cette étude comme contrôle négatif, ne menant pas non plus à une expression de la protéine (Figure 19). Les conséquences fonctionnelles de la mutation *frameshift* furent étudiées dans les lymphocytes primaires, en mesurant la sécrétion de cytokines dans des analyses de sang total. Par exemple, après stimulation avec du BCG et de l'IL-12 (cytokine pouvant induire l'IFN- γ), le sang total des témoins sains produit de grandes quantités d'IFN- γ ce qui n'est pas le cas pour les cellules des patients. Ces études, ainsi que d'autres études qui ne sont pas ici décrites, prouvèrent *in fine* que P1 et P2 avaient un déficit complet autosomique récessif associé à l'IFN- γ .

Figure 19: Analyse par western blot de l'IFN- γ



Analyse réalisée dans des cellules HEK293T non transfectées (NT) ou transfectées avec un plasmide vide (EV), WT-IFNG, IFNG muté, le tout inséré dans CNV6 avec une étiquette DDK C-terminale.

Nous nous sommes aussi intéressés au fait, plus que remarquable, qu'un seul cas de MSMD avec un déficit autosomique récessif en IFN- γ ait été diagnostiqué à ce jour, alors que des centaines de familles avec un déficit au niveau des récepteurs 1 et 2 de l'IFN- γ (IFN- γ R1 et l'IFN- γ R2) aient été rapportés au cours des 23 dernières années. Dans ce but, nous nous sommes focalisés sur deux paramètres de sélection pour estimer la force de pression évolutive qui a pu agir sur les gènes *IFNG*, *IFNGR1* et *IFNGR2*. En premier lieu, nous avons considéré le paramètre f , qui est basé en partie sur le rapport du taux de substitutions sur des sites synonymes et non-synonymes d'un gène (dénnoté par dN/dS), quantifiant la force de sélection agissant sur les protéines humaines. Ce paramètre prend des valeurs entre 0 et 1, où 0 correspond à une sélection fortement négative et 1 à une sélection neutre. Estimé par la méthode SnIPRE (208), *IFNG* se situe à 0.318, ce qui le classe parmi les 8.9% des gènes sous la plus forte pression évolutive (Figure 20) selon le paramètre f . En revanche, *IFNGR1* et *IFNGR2* se situent au-dessus des 28% et des 65% des gènes avec la sélection négative la plus forte, respectivement (Figure 20). Un résultat similaire a été obtenu avec l'indice d'endommagement génétique (*gene damage index*, GDI) (209) qui est également beaucoup plus faible pour *IFNG* (classé dans les 1.3% des gènes présentant le plus faible GDI) que pour *IFNGR1* et *IFNGR2* (classés dans le top 62% et 59% de ces gènes, respectivement) (Figure 21).

Figure 20: Distribution à l'échelle du génome entier de l'intensité de la sélection évolutive agissant sur les gènes *IFNG*, *IFNGR1* et *IFNGR2* par le paramètre f

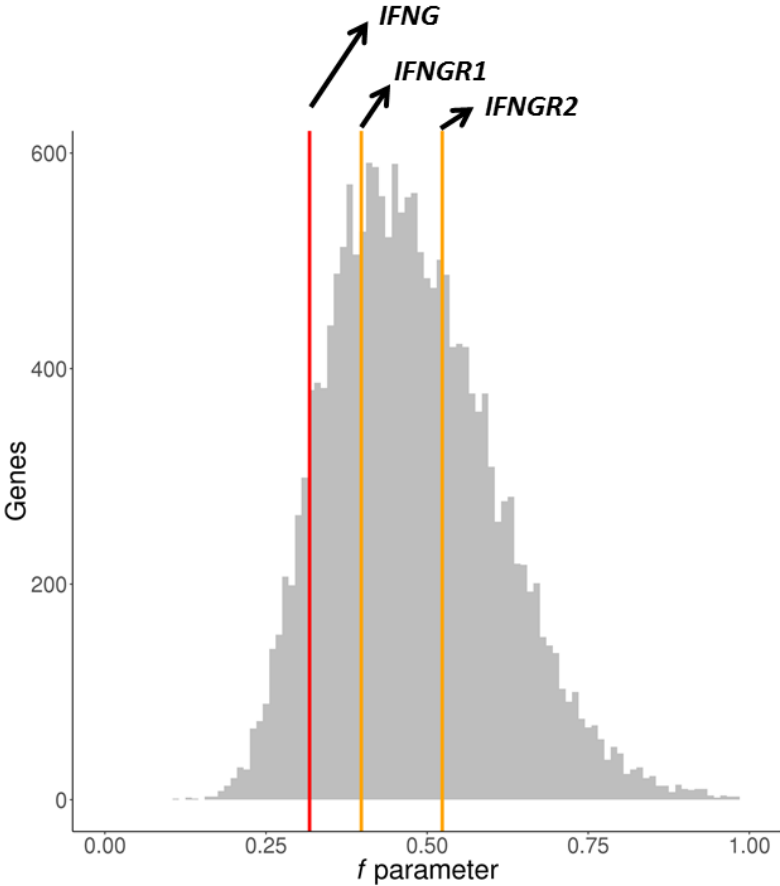
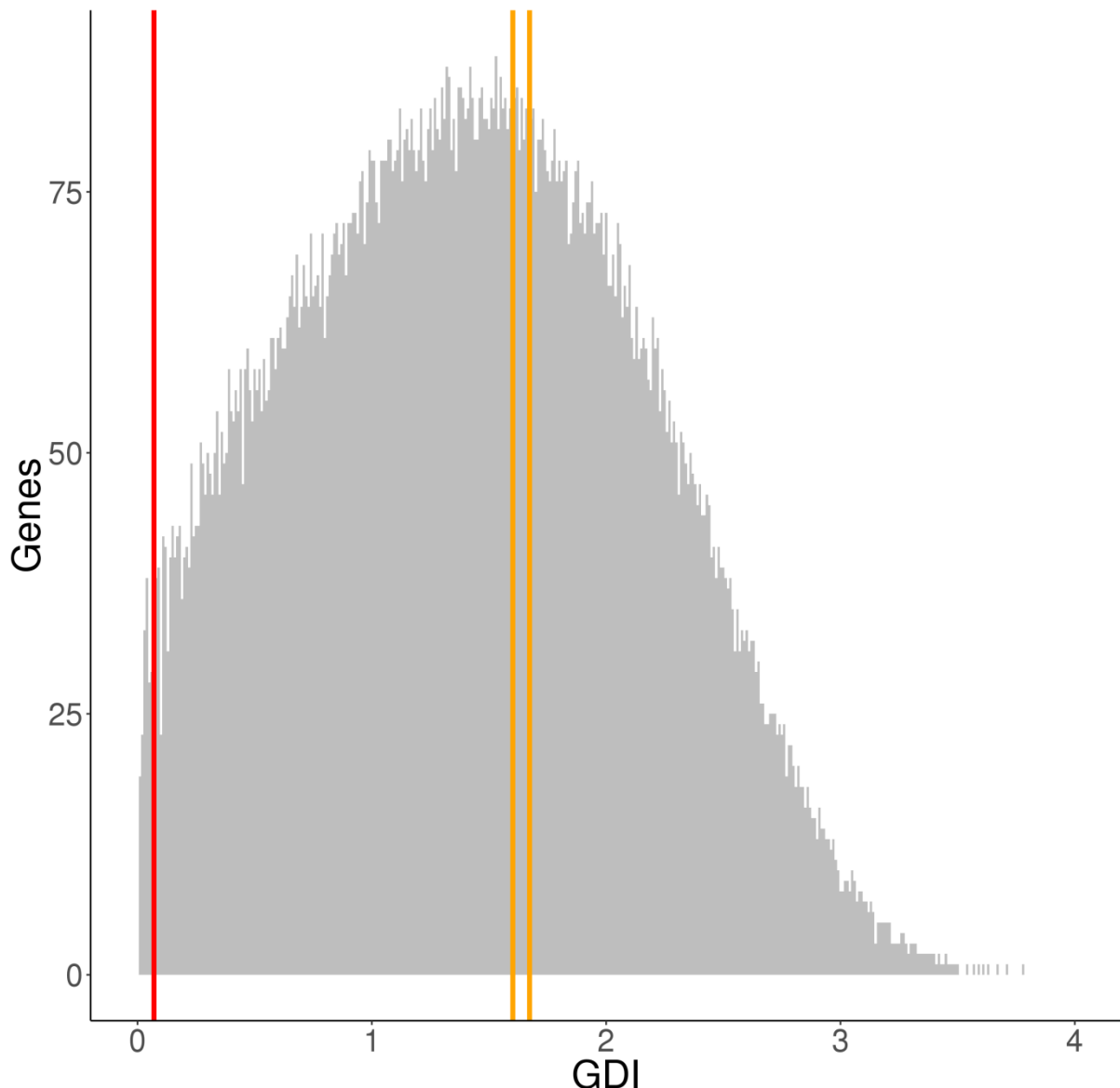


Figure 21: Distribution à l'échelle du génome entier de l'intensité de la sélection évolutive agissant sur les gènes *IFNG*, *IFNGR1* et *IFNGR2* par le paramètre GDI



IFNG est représenté par un trait vertical rouge tandis que, dans l'ordre, *IFNGR1* et *IFNGR2* sont représentés par un trait vertical orange.

De plus, de nombreux variants faux-sens ou prédits LOF sont présents à l'état hétérozygote dans *IFNGR1* et *IFNGR2*, souvent avec une MAF > 0.001, alors que beaucoup moins de variants de cette nature sont observés dans *IFNG*. La fréquence individuelle maximale d'un variant faux-sens ou prédit LOF en *IFNG* est de 6×10^{-5} tandis que la fréquence cumulée maximale de ces variants est de 5.3×10^{-4} . Cependant, pour *IFNGR1* et *IFNGR2* des variants faux-sens fréquents à plus de 1% sont présents dans les bases de données publiques. Enfin, en tenant compte de la taille respective de leurs régions codante, les rapports des variants non-synonymes en fonction des paires de bases (pb) est deux fois plus faibles pour *IFNG* (0.09 / pb) que pour *IFNGR1* (0.20 / pb) et *IFNGR2* (0.17 / pb) (Tableau 3). Par contre, en

considérant les régions adjacentes non codantes couvertes par WES, ce rapport de variant par pb est plus élevé dans *IFNG* (0.12 / pb) que dans *IFNGR1* (0.06 / pb) et *IFNGR2* (0.08 / pb), ce qui indique qu'*IFNG* n'est pas dans une région avec un taux spécialement faible de mutation.

Ce modèle de sélection, plus important au locus *IFNG* qu'aux loci *IFNGR1* et *IFNGR2*, pourrait être aussi influencé par la physiologie de l'immunité à l'IFN- γ . Les deux chaînes du récepteur sont ubiquitaires, en particulier IFN- γ R1, dont l'expression est constitutive, et même IFN- γ R2, dont l'expression peut être régulée dans certains types cellulaires pour augmenter ou diminuer leurs réponses à l'IFN- γ . En revanche, la production d'IFN- γ est strictement régulée. Seuls quelques types de cellules produisent cette cytokine en abondance (cellules NK, $\gamma\delta$ T, CD4 T et MAIT). Il est à noter de même que le déficit autosomique dominant en IFN- γ R1 n'est provoqué que par quelques mutations spécifiques avec un effet de dominance négative (153,155), alors que le déficit autosomique dominant en IFN- γ R2, par haplo-insuffisance, montre une très faible pénétrance pour le MSMD (157), le récepteur étant abondamment exprimé, se retrouvant donc en excès. Globalement, ces résultats indiquent que le gène *IFNG* a probablement évolué sous une sélection négative beaucoup plus forte qu'*IFNGR1* et qu'*IFNGR2*, ce qui explique par ailleurs qu'*IFNG* soit beaucoup moins tolérant aux mutations hétérozygotes qu'*IFNGR1* et qu'*IFNGR2*. Ces résultats sont cohérents avec la rareté du déficit autosomique récessif associé à l'IFN- γ , en particulier, quand il est comparé aux déficiences en IFN- γ R1 et en IFN- γ R2. Ils suggèrent également que les patients hétérozygotes pour des mutations délétères en *IFNG* pourraient être susceptibles à une mycobactérie plus virulente, telle la tuberculose.

Tableau 3: Distribution des variants dans les régions codantes et non codantes (couvertes par WES) de l'*IFNG* et de ses deux récepteurs

			<i>IFNG</i>	<i>IFNGR1</i>	<i>IFNGR2</i>
Gène	Codants (non-synonymes)	Variants	47	286	172
		Taille (bp)	498	1467	1011
		Rapport (var/bp)	0.09	0.20	0.17
	Non-codants	Variants	101	237	237
		Taille (bp)	866	4146	2809
		Rapport (var/bp)	0.12	0.06	0.08

B.4. Conclusion.

Le travail que nous avons conduit a permis de décrire une nouvelle étiologie génétique du MSMD, le déficit complet autosomique récessif en IFN- γ , chez deux patients d'une grande famille consanguine originaire du Liban. Ces patients ne produisent pas de l'IFN- γ et sont donc candidats pour un traitement par IFN- γ recombinant. IFN- γ se lie à IFN- γ R1 et à IFN- γ R2. Au cours des 23 dernières années, des études sur le MSMD ont caractérisé cinq formes alléliques de déficit en IFN- γ R1 et cinq formes alléliques de déficiences en IFN- γ R2, dont huit sont récessives. Des centaines de ces patients ont été retrouvés dans le monde entier, sans appartenance ethnique particulière. En revanche, nous rapportons ici les deux seuls premiers patients avec des mutations bi-alléliques en *IFNG*. Une explication plausible avancée dans ce travail est que le gène *IFNG* ait évolué sous une sélection de purification plus stricte, empêchant l'accumulation de variants délétères à ce locus. Dans tous les cas, les mutations d'*IFNG* doivent être considérées dans l'étude de patients MSMD ou avec d'autres infections apparentées.

C. Homozygotie d'un variant fréquent faux-sens du gène *TYK2* : première étiologie génétique relativement fréquente de la tuberculose.

Le défaut complet autosomique récessif de la tyrosine kinase 2 (TYK2) est un des quelques défauts génétiques rares décrits dans la tuberculose maladie. Le premier patient déficient en TYK2 à être identifié a été signalé en 2006 (210). Ce patient était japonais et présentait les signes caractéristiques du syndrome de l'hyper-IgE (HIES), mais également des infections bactériennes intracellulaires, en particulier, une lymphadénite causée par le vaccin BCG. Par la suite, sept autres patients déficients en TYK2 appartenant à cinq familles non apparentées originaires de la Turquie, du Maroc, de l'Iran et de l'Argentine et présentant des infections bactériennes intracellulaires furent rapportés (156). Il convient de remarquer que deux de ces patients (l'un Marocain et l'autre Iranien) ne présentaient que des infections par *M. tuberculosis*, avec des signes cliniques d'une tuberculose abdominale et miliaire, respectivement. Cette étude a fourni la preuve d'une nouvelle cause monogénique de la tuberculose, s'ajoutant à celles décrites en **I_H**. Cependant, la prévalence de la tuberculose est bien plus importante dans le monde et les rares formes monogéniques (qui sont elles-mêmes rares) décrites à partir de la dissection des bases génétiques du syndrome MSMD (souvent chez des familles avec des membres atteints du MSMD et vivant dans des zones endémiques

pour la tuberculose), ne peuvent expliquer qu'une petite minorité des cas de tuberculose dans le monde.

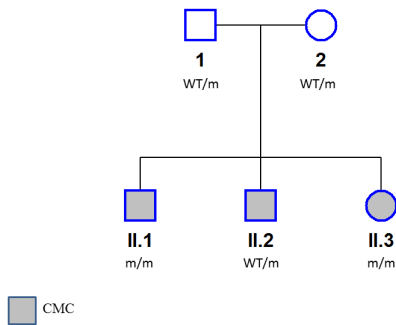
Suivant une approche polymorphisme-candidat et basé sur les études précédentes, nous avons testé l'hypothèse que des variants plus communs de *TYK2* pourraient prédisposer à la tuberculose. Spécifiquement, nous avons analysé deux variants faux-sens indépendants du gène *TYK2*, pour lesquels leurs homozygoties respectives avaient été prouvées être inactives sur le plan catalytique, P1104A (*rs34536443*) et I684S (*rs12720356*). Des études avaient déjà suggéré que l'activité catalytique de *TYK2* était nécessaire pour les réponses des lymphocytes T à l'IL-12 et à l'IL-23 induisant à leur tour l'IFN- γ (211). Dans ce travail j'ai été directement en charge d'une première analyse d'enrichissement du variant *TYK2* P1104A dans la cohorte d'étude HGID, contribuant également à l'étude de sélection négative du variant qui a été fait en collaboration avec l'équipe de Lluís Quintana-Murci. Ce travail, qui comporte également une importante partie expérimentale, réalisée par le groupe de génétique immunologique du laboratoire, est décrit dans la section **II_C.1**. Par la suite, j'ai conduit l'analyse d'association de ce même variant dans une autre cohorte d'étude dénommée UK Biobank, qui est décrite dans la section **II_C.2**.

C.1. Analyse de la cohorte HGID.

Pour une première analyse d'association des variants P1104A et I684S du gène *TYK2*, nous nous sommes focalisés sur la cohorte HGID, qui est décrite dans la section **II_A**. Nous avons considéré en tant que cas, soit les 463 MSMD de la cohorte, soit les 454 tuberculeux, soit le total des deux groupes. Afin d'augmenter la puissance statistique de cette analyse, nous avons considéré pour les témoins, en plus des 2,835 individus atteints d'autres infections sévères et non-mycobactériennes de la cohorte HGID, décrits en **II_A**, les 2,504 individus « sains » et non-apparentés de 1000G, aussi décrits en **II_A**, menant à un total de 5,339 témoins. L'association des deux variants *TYK2* avec le syndrome MSMD et/ou la tuberculose a été testée en utilisant un modèle de régression logistique ajusté sur les trois premières composantes principales de la PCA de la Figure 15 (voir section **II_C.2**). Pour le cas spécifique d'un mode génétique de transmission récessif et d'un variant donné, le modèle compare la proportion de porteurs homozygotes chez le groupe des cas et chez celui des témoins à l'aide d'un test de rapport de vraisemblance (*likelihood ratio test*, LRT).

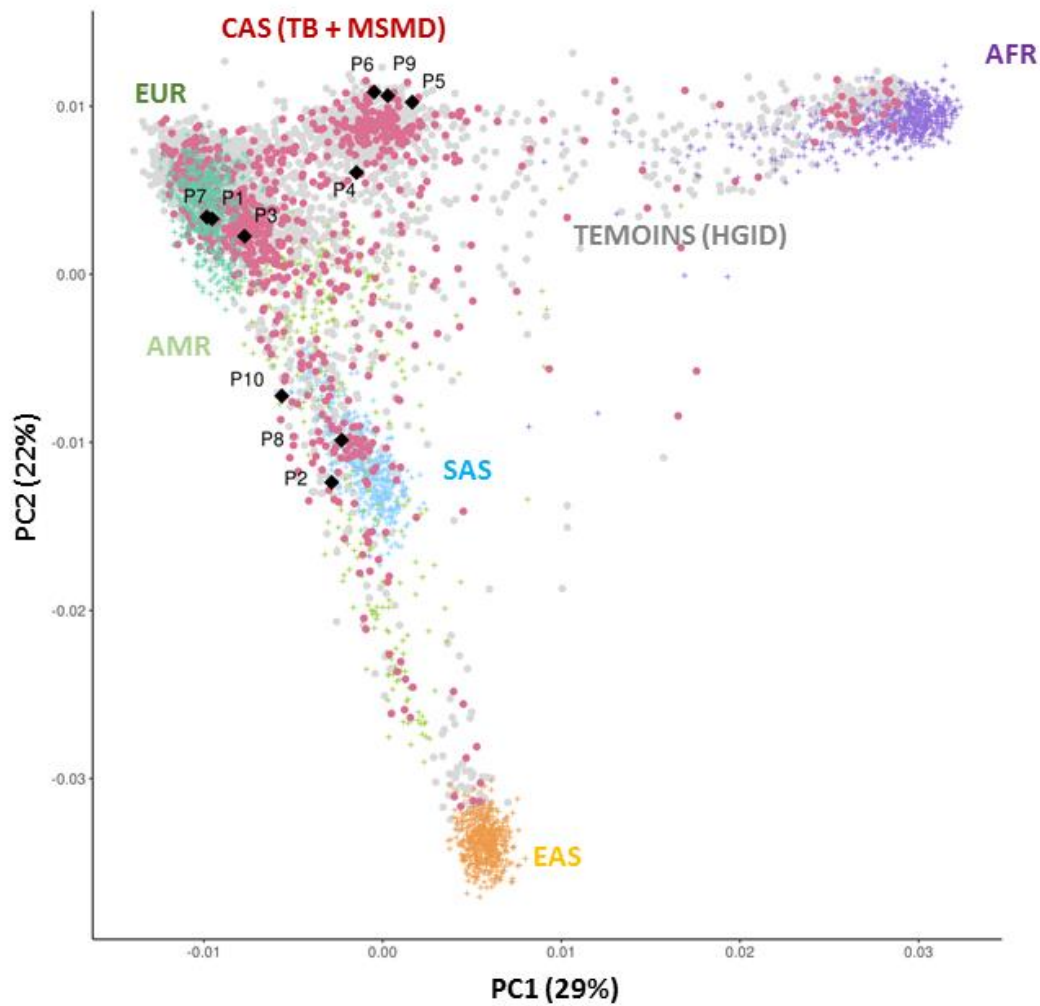
En appliquant cette méthode, nous avons observé un fort enrichissement de porteurs homozygotes du variant P1104A chez les patients tuberculeux ($P = 8.37 \times 10^{-8}$; OR = 89.31; intervalle de confiance (IC) à 95% = [14.7 ; 1,725]) et un enrichissement plus modéré chez les

patients atteints de MSMD ($P = 3.27 \times 10^{-3}$; OR = 23.53; IC à 95% = [2.9 ;483]). Onze homozygotes P1104A ont été identifiés et confirmés par séquençage Sanger: sept atteints de tuberculose (trois enfants de moins de 15 ans et quatre adultes de moins de 40 ans), trois de MSMD (tous âgés de moins de 3 ans) et un patient atteint d'une candidose cutanéomuqueuse chronique (*chronic mucocutaneous candidiasis*, CMC) (âgé de 1 an). Outre les dix patients atteints du syndrome MSMD ou d'une tuberculose (voir Tableau 4), un seul autre homozygote P1104A a été identifié par WES dans la cohorte HGID (aucun homozygote P1104A n'a été identifié parmi les 2,504 individus de 1000G) (voir Figure 22). Cependant, l'homozygotie pour P1104A ne semble pas être à l'origine de la maladie CMC pour ce dernier patient, la ségrégation familiale n'étant pas en faveur de cette théorie (voir diagramme ci-dessous).



Ce patient vit par ailleurs aux États-Unis où les enfants ne reçoivent pas le vaccin du BCG et où *M. tuberculosis* n'est pas endémique. Finalement, le modèle récessif pour I684S ainsi que les modèles dominants et hétérozygote composite pour les deux allèles ne conduisirent pas à des résultats significatifs, la proportion de porteurs entre les différents groupes étant relativement similaire (Tableau 4).

Figure 22: PCA des 2,504 individus "sains" de 1000G et des individus de la cohorte HGID

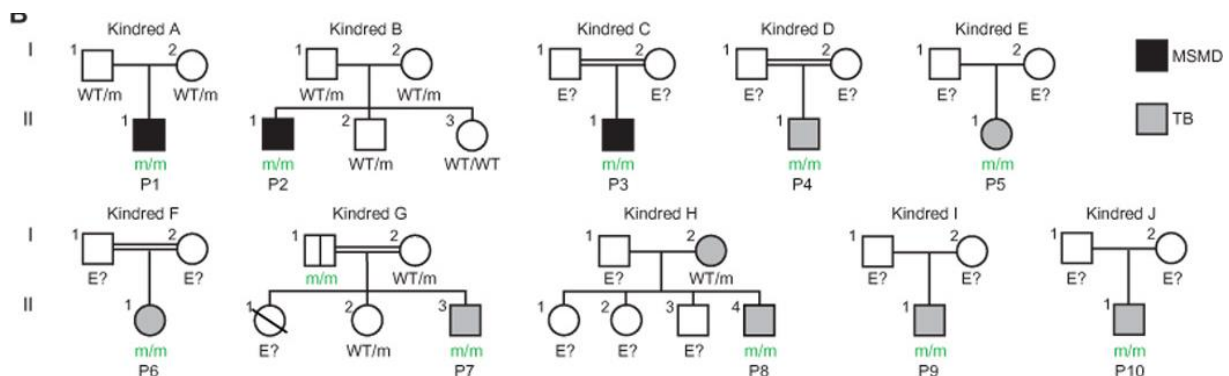


Les individus de 1000G sont d'origines ancestrales connues : Africaine (AFR), Américaine (AMR), Asiatique de l'Est (EAS), Européenne (EUR), Sud Asiatique (SAS). Les patients avec (n=917, en rouge) ou sans (n=2,835, en gris) des infections par mycobactéries de la cohorte HGID sont représentés dans le même graphique ainsi que les dix porteurs homozygote du variant TYK2 P1104A (carreaux noirs) atteints de tuberculose ou du syndrome MSMD. Les proportions de la variance expliquées par la première (PC1) et la deuxième (PC2) composante principale sont indiquées entre parenthèses.

Tableau 4: Synthèse des porteurs des variants TYK2 étudiés dans la cohorte HGID

		Témoins (n=2,835)	TB (n=454)	MSMD (n=463)	TB+MSMD (n=917)
I684S	Hétérozygote (%)	277 (9.77)	40 (8.81)	49 (10.58)	89 (9.7)
	Homozygote (%)	15 (0.52)	1 (0.22)	2 (0.43)	3 (0.32)
P1104A	Hétérozygote (%)	128 (4.51)	21 (4.66)	19 (4.1)	40 (4.36)
	Homozygote (%)	1 (0.03)	7 (1.54)	3 (0.64)	10 (1.09)
I684S+P1104A	Composite (%)	5 (0.17)	1 (0.22)	0 (0)	1 (0.10)

Figure 23: Distribution des familles des dix porteurs homozygotes du variant TYK2 P1104A ayant une tuberculose ou le syndrome MSMD



Le variant P1104A est présent en état homozygote chez $\sim 1/600$ Européens même si en dehors de l'Europe il est bien plus rare (Figure 24). Il se retrouve en état homozygote chez $\sim 1/5,000$ individus en moyenne ailleurs, à l'exception de l'Asie de l'Est où sa MAF est inférieure à 10^{-5} , et de l'Afrique sub-saharienne où le variant est absent des ~ 400 individus appartenant aux populations exclusivement Africaines de 1000G. Il est cependant observé en état hétérozygote sur deux individus Afro-Américains de cette base de données. Nous avons valorisé notre découverte par une analyse d'association à l'échelle de l'exome entier, conduite au sein de notre cohorte finale, à l'aide d'un modèle récessif. L'allèle P1104A a donné le plus grand OR possible parmi les 6,231 variants fréquents ($>1\%$) annotés faux-sens ou LOF, observés dans notre cohorte (Figure 25).

Figure 24: Carte de fréquences de l'allèle P1104A autour du globe

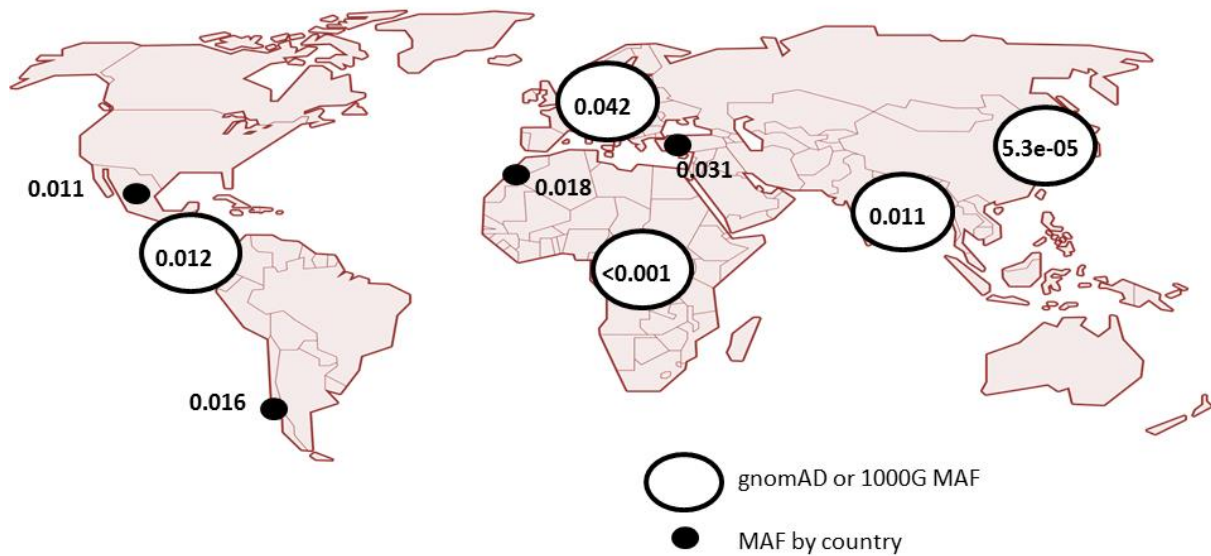
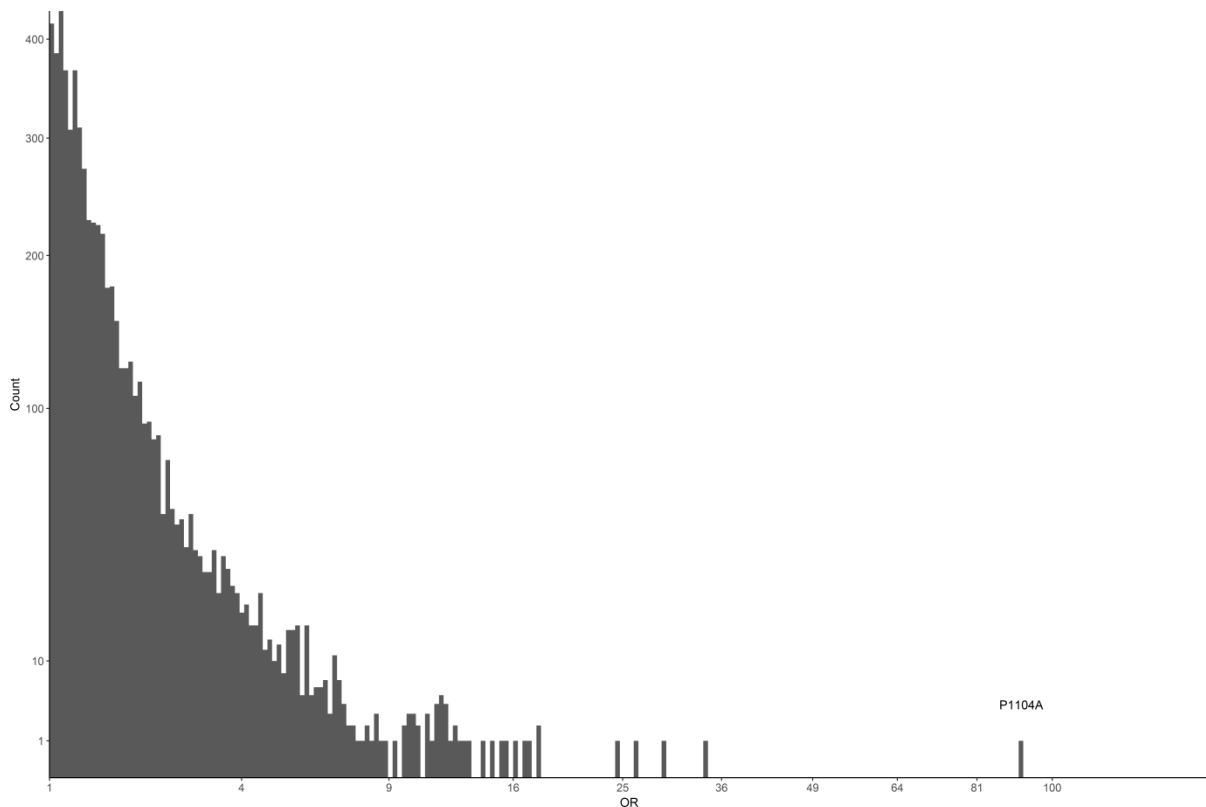


Figure 25: Distribution des ORs obtenus lors du test d'association de la tuberculose avec tous les variants codants fréquents faux-sens ou potentiellement perte de fonction

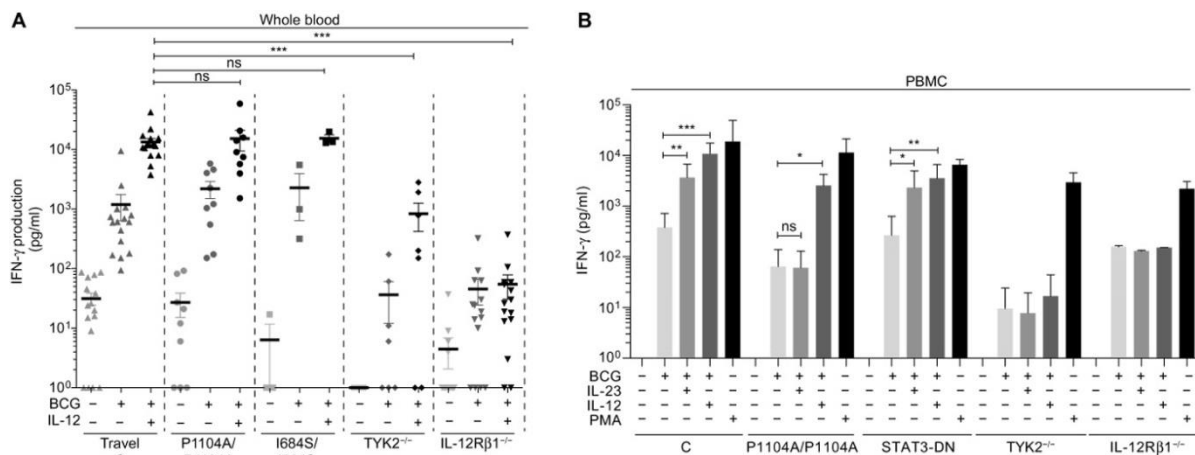


L'analyse a été conduite en supposant un mode de transmission récessif.

De nombreuses analyses expérimentales ont contribué à cette étude pour valider le rôle fonctionnel du variant, réalisées par le groupe de génétique immunologique du laboratoire.

Ces études ont montré un défaut spécifique de production d'IFN- γ en réponse à l'IL-23. En particulier, la réponse à l'IL-12 a été normale contrairement au cas des patients présentant un déficit complet en TYK2 pour lesquels les réponses cellulaires à l'IL-12 et à l'IL-23 sont sévèrement altérées (156,212,213), ce qui est illustré par la Figure 26. En effet, les homozygotes P1104A ont une réponse à l'IL-12 normale en comparaison aux témoins sains de l'analyse, tandis que la réponse est moindre ou abolie pour les patients avec une déficience complète en TYK2 ou les patients avec une déficience complète en IL12-R β 1, respectivement (Figure 26A). D'autre part, tout comme pour les patients présentant des déficiences complètes en TYK2 et en IL-12R β 1, les homozygotes P1104A ne répondent pas à l'IL-23 en termes de production d'IFN- γ , comme le montre la comparaison avec les témoins sains (Figure 26B). Pour les patients homozygotes P1104A, l'immunité médiée par l'IFN- γ et induite par l'IL-23 est donc altérée. L'homozygotie pour le variant P1104A du gène *TYK2* perturberait donc sélectivement l'immunité anti-mycobactérienne médiée par l'IFN- γ de manière dépendante de l'IL-23.

Figure 26: Analyse des taux d'IFN- γ en sang total pour les patients homozygotes P1104A en comparaison à des témoins



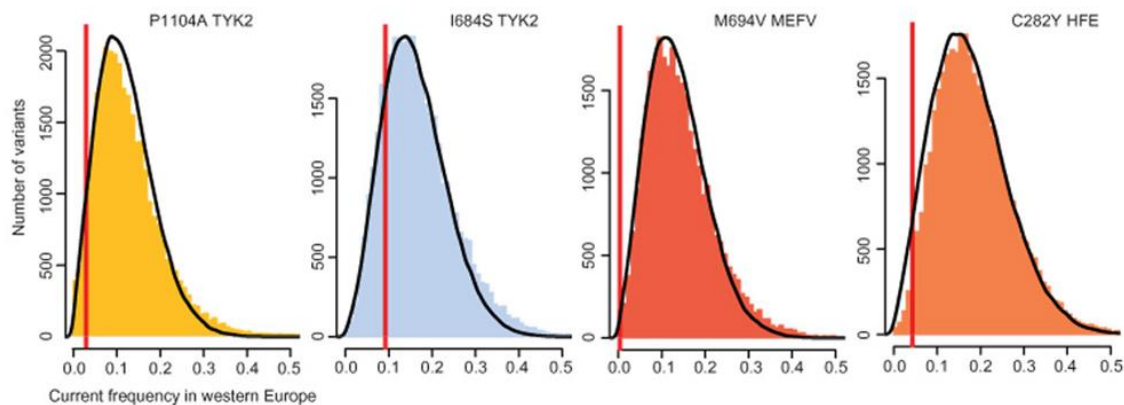
A) Analyse des taux d'IFN- γ en sang total après stimulation par le BCG, ou le BCG plus d'IL-12, chez des témoins sains, des patients déficients en TYK2 et en IL-12R β 1, et des patients homozygotes pour les allèles TYK2 P1104A ou TYK2 I684S. B) Production d'IFN- γ à partir de *peripheral blood mononuclear cells* (PBMC) stimulées par le BCG, le BCG plus l'IL-12 ou le BCG plus l'IL-23 chez des témoins sains, des patients homozygotes TYK2 P1104A, des patients hyper-IgE présentant des mutations hétérozygotes STAT3 et des patients présentant des déficiences complètes en TYK2 et en IL-12R β 1.

Sélection négative de P1104A en Europe

Le risque plus élevé de tuberculose chez les homozygotes P1104A suggère la possibilité que ce variant ait été soumis à une pression de sélection négative au cours du temps, dans les zones où cette maladie fut endémique, comme ce fut le cas de l'Europe (214). Nous avons donc étudié la variation des fréquences des variants P1104A et I684S dans des populations Européennes des temps anciens aux temps modernes (215). Seuls trois variants

TYK2 non synonymes - P1104A, I684S et V362F - ont été retrouvés dans un échantillon de 22 individus d'Europe centrale ayant vécu à la fin du Néolithique, il y a environ 4,000 ans (215). Cette analyse a montré une diminution de la fréquence de P1104A dans cette période chez les Européens, passant d'environ 9 à 4.2% (Figure 27). Parmi les 31,276 variants avec une fréquence comprise entre 8 et 10% il y a 4,000 ans, P1104A fait partie des 5% ayant subi la plus forte diminution de fréquence (p empirique = 0.048; Figure 27). De plus, un modèle d'évolution neutre a été rejeté de manière significative pour l'allèle à partir de simulations de Wright-Fisher (p = 0.050; Figure 28), suggérant une absence de biais dans les analyses empiriques. Par contre, la fréquence de I684S n'a pas diminué de manière significative au cours de cette période (p empirique = 0.181; Figure 27). En outre, la fréquence de V362F est restée stable (de 25% à 26.2%) dans le temps (non montré).

Figure 27: Distributions des fréquences alléliques actuelles des variants qui ont ségrégués il y a 4,000 ans à des fréquences similaires à celles des variants P1104A et I684S *TYK2*, M694V *MEFV* et C282Y *HFE*



Les lignes verticales rouges indiquent la fréquence actuelle des quatre variants d'intérêt. Les barres de couleur indiquent la distribution des fréquences actuelles des variants ayant des fréquences dans l'ADN d'Européens anciens similaires à celles des quatre variants étudiés, issus de la base de données 1000 Genomes Project. Les lignes noires indiquent la distribution des fréquences simulées, dans la génération actuelle, pour les allèles ayant une fréquence ancienne similaire à celle des quatre variants étudiés, avec une propagation sur 160 générations (correspondant à une période de ~4,000 ans) sous le modèle neutre de Wright-Fisher.

Nous avons ensuite analysé, en tant que contrôles, deux mutations relativement fréquentes, connues pour causer des pathologies autosomiques récessives et potentiellement sévères, présentes chez des Européens ayant vécu il y a environ 4,000 ans: le variant *MEFV* M694V prédisposant à la fièvre méditerranéenne (216), pouvant en plus réduire la capacité de reproduction du fait de la sévérité de la maladie, et le variant *HFE* C282Y sous-jacent à l'hémochromatose, diminuant par ailleurs la fertilité (217). La fréquence de ces deux variants a significativement diminué au cours de la même période, passant d'environ 11% à 0.4% pour le *MEFV* M694V et de 16% à 5.7% pour le *HFE* C282Y (p empirique = 0.016 pour les deux variants; Figure 27). Nos évaluations préliminaires suggèrent que *TYK2* P1104A, *MEFV*

M694V et HFE C282Y ont été soumis à une sélection négative chez les Européens, contrairement à TYK2 I684S. Une plus forte sélection opérant sur MEFV M694V, et dans une moindre mesure sur HFE C282Y, que sur TYK2 P1104A, est compatible avec le fait que ces variants ne nécessitent pas un agent extérieur pour avoir un effet pathogène. Leurs conséquences sont relativement indépendantes d'une exposition particulière. Ces résultats suggèrent que, contrairement à I684S, P1104A a subi une purge en Europe depuis la période néolithique en raison du caractère endémique persistant de la tuberculose (214). Aucune autre infection « intramacrophagique », dont le contrôle dépend de l'IFN- γ , n'a été endémique depuis si longtemps en Europe (2,4). Les effets de purge devraient être beaucoup moins efficaces en l'absence d'une exposition continue (218,219) ce qui a été le cas pour d'autres infections ayant tué une proportion non négligeable d'Européens, sans toutefois dépasser plusieurs décennies ou quelques siècles, comme la peste (220). De plus, la purge est fortement dépendante du mode de dominance du génotype à risque. Le déclin de la fréquence de l'allèle P1104A est compatible avec un mode de transmission récessif du génotype à risque. Cette diminution serait beaucoup plus prononcée pour un trait dominant. Ces résultats suggèrent que l'homozygotie pour P1104A, qui est toujours présente chez environ 1/600 Européens, a été un déterminant génétique majeur de la tuberculose au cours de l'histoire humaine.

C.2. Analyse de la cohorte UK Biobank.

L'analyse d'enrichissement du variant TYK2 P1104A dans la cohorte HGID, son étude évolutive, et les analyses immunologiques associées (221), suggèrent fortement que l'homozygotie pour P1104A est une première étiologie génétique relativement commune de la tuberculose. Dans cette première étude (221), les patients porteurs du génotype à risque proviennent de diverses régions du monde comme le montre la PCA de la Figure 23. Aucun de ces porteurs n'était d'origine Européenne, région du monde où le variant est plus fréquent (~ 1/600 Européens sont homozygotes) mais où l'exposition à l'agent pathogène de la tuberculose est actuellement faible (222–225). La cohorte HGID est par ailleurs très hétérogène, non seulement au niveau génétique (dû à la grande diversité d'origines ethniques de ses individus), mais aussi par rapport aux diverses sources de pressions environnementales agissant sur les individus de la cohorte. Nous avons donc conduit une nouvelle analyse du variant dans une cohorte plus homogène et moins biaisée par rapport au recrutement de ses individus. La cohorte publique UK Biobank composée de 500,000 volontaires recrutés au Royaume-Uni, incluant en particulier, des individus ayant été diagnostiqués d'une tuberculose, nous sembla particulièrement pertinente à nos propos.

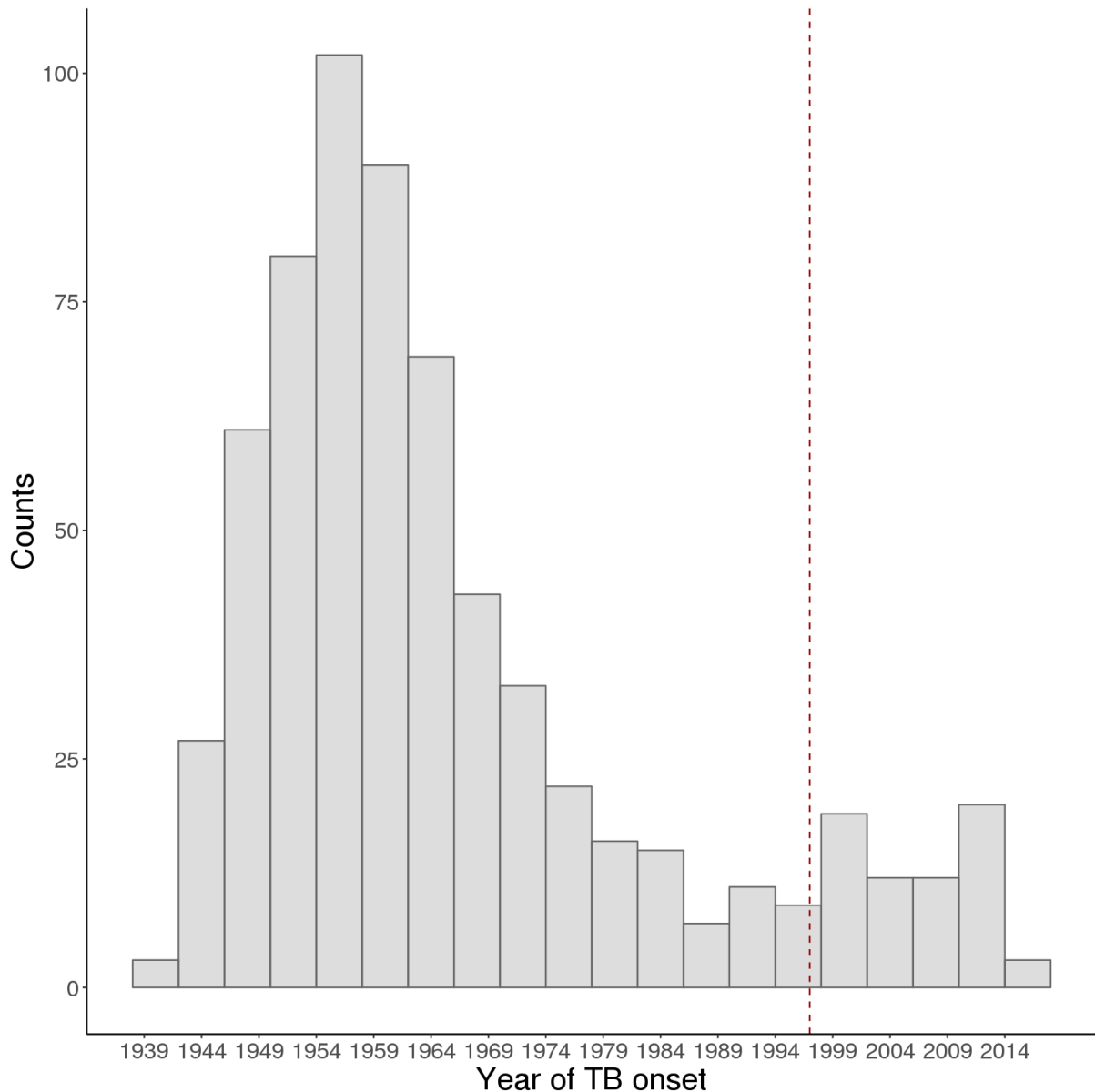
Description de la cohorte d'étude UK Biobank

La cohorte UK Biobank contient environ 500,000 volontaires nés entre 1940 et 1970, qui ont été recrutés entre 2006 et 2010, et qui sont originaires, pour la grande majorité, du Royaume-Uni (226). La ressource dispose de données de génotypage à l'échelle du génome entier (utilisées notamment dans le cadre de GWAS) pour tous les participants, couvrant notamment la position du variant TYK2 P1104A. Elle inclut par ailleurs des informations liées au phénotype tuberculose pour un grand nombre de participants. Parmi ces données nous nous sommes focalisés sur le champ le plus pertinent de la ressource (#22137), fournissant des informations sur la tuberculose « diagnostiquée par un médecin » (comme cela est indiqué par la ressource elle-même), et qui sont disponibles pour 121,284 participants au format binaire. Au total, 654 personnes de ce groupe sont ou ont été des cas confirmés de tuberculose (code 1), alors que les 120,630 autres qui ne l'ont jamais été (code 0) ont été utilisés comme témoins pour nos analyses d'enrichissement. Le variant P1104A a été génotypé dans 620 de ces cas et 114,473 de ces témoins.

Deux autres champs de la ressource UK Biobank sont aussi liés au phénotype tuberculose. Le premier correspond à ce qui est appelé statistiques d'épisodes hospitaliers (champs #41202 et #41204), qui sont codés selon la Classification internationale des maladies (*International Classification of Diseases, ICD*) 10. Ces données sont accessibles, mais ne sont disponibles que pour les patients qui ont été hospitalisés après 1996. Etant donné que le diagnostic de 581 (89%) cas de tuberculose « diagnostiqués par un médecin » du champ #22137 a été fait avant 1996 (Figure 28), la plupart de ces patients tuberculeux n'apparaissent pas dans ces derniers champs. Un autre champ lié à la tuberculose, extrait du champ majeur *Non-cancer illness code, self-reported field* (#20002), correspond à la « tuberculose auto-déclarée », totalisant 2,629 individus. Nous avons constaté un très faible chevauchement avec le champ #22137, car seuls 308 (47%) des 654 cas de tuberculose « diagnostiqués par un médecin » se sont auto-déclarés tuberculeux. Le chevauchement est encore plus faible avec les données d'épisodes hospitaliers (#41202 et #41204) étant donné que moins de 35% d'individus avec un code ICD-10 associé à la tuberculose maladie se sont auto-déclarés tuberculeux. Une autre étude focalisée sur les volontaires de la ressource UK Biobank qui ont déclaré avoir eu des accidents vasculaires cérébraux (227) avait déjà pu constater une faible fiabilité des données auto-déclarées. Dans l'ensemble, le champ #22137 (individus diagnostiqués par un médecin) demeure le champ lié à la tuberculose le plus pertinent. En effet, les 654 cas de tuberculose ont renseigné un âge de survenue de la maladie (ceci n'étant

pas le cas pour les autres champs) et un nombre spécifique de 120,630 personnes ont rentré un code 0 (c'est-à-dire, pas de tuberculose) dans ce champ.

Figure 28: Année de survenue de la tuberculose pour tous les patients du champs #22137 de la base de donnée UK Biobank



Un trait pointillé rouge est inscrit à l'année 1996 pour laquelle les données ICD-10 ont commencé à être enregistrées.

En se focalisant donc uniquement sur le champ #22137, nous avons étudié les distributions d'âges et de sexe des 115,093 individus appartenant à cette sous-cohorte. L'étude n'a montré pas de différences majeures entre le groupe des cas et des témoins. Par rapport aux hommes, un léger excès de femmes chez les cas (rapport homme / femme = 0.83) comme chez les témoins (0.80), a été observé, correspondant au même excès observé dans toute la cohorte UK Biobank (0.84). L'âge moyen au recrutement a été de 56.5 ans (ET = 7.7)

pour les témoins et de 61.1 ans (ET = 6.5) pour les patients atteints de tuberculose, dont l'âge moyen est de 18.5 ans (ET = 16.0). Nous n'avons donc pas envisagé des corrections statistiques liées à l'âge ou au sexe pour nos analyses ultérieures.

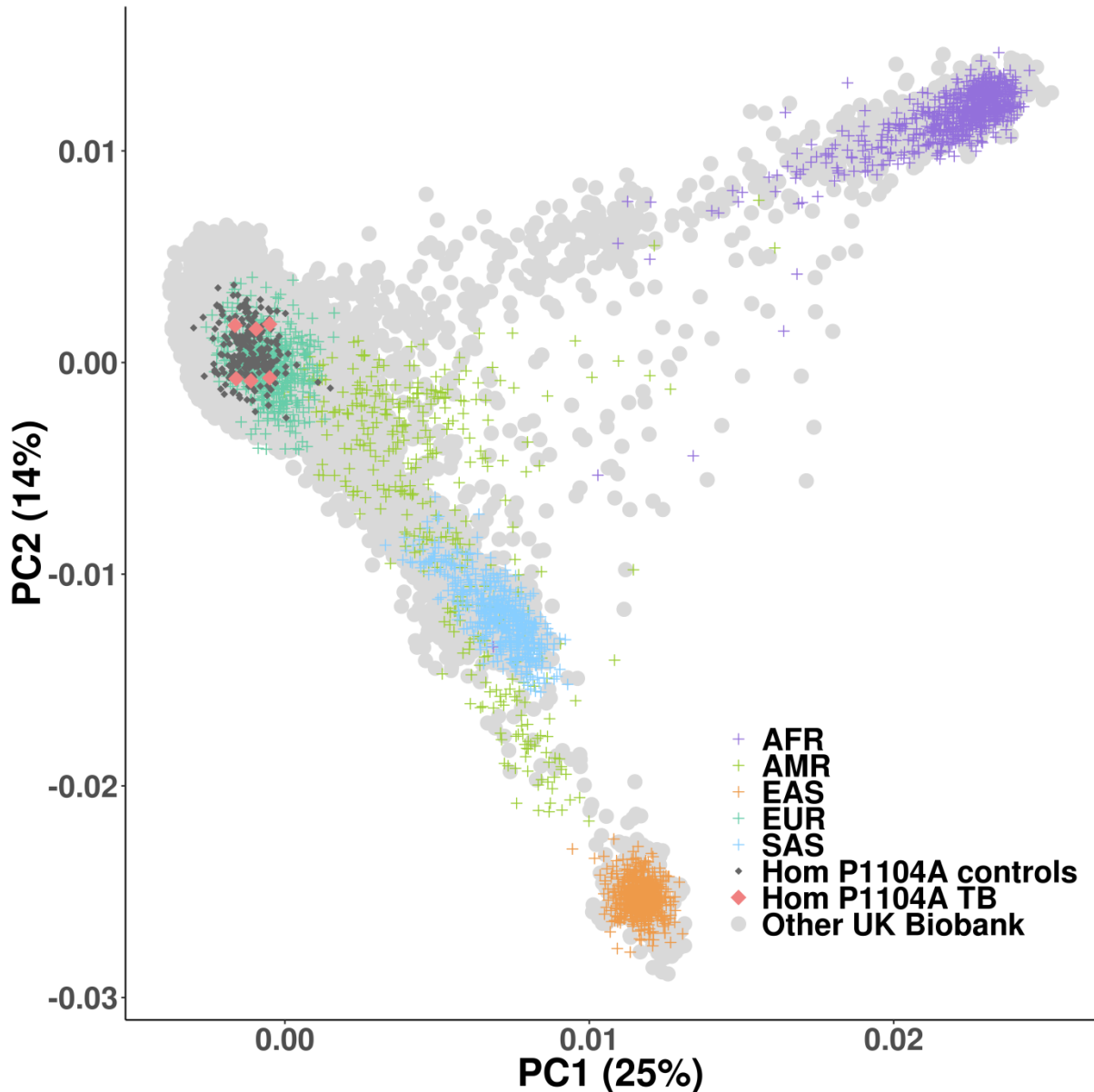
Analyse d'enrichissement

Nous avons d'abord observé que la fréquence des porteurs homozygotes de P1104A était plus élevée chez les patients tuberculeux (6/620 = 0.97%) que chez les témoins (228/114,473, 0.2%) [OR = 4.90 (IC 95%: 1.93-10.10); P = 2×10^{-3}] (Tableau 5) de la sous-cohorte UK Biobank. Parmi les six patients homozygotes, l'âge de survenue de la tuberculose était variable, une moitié précédant l'âge de 15 ans (deux nés en 1944 et un autre en 1946), et l'autre dépassant l'âge de 50 ans (nés en 1943, 1944 et 1946) (Tableau 5). Aucune différence significative n'a été observée entre la MAF de P1104A calculée chez les témoins de cette cohorte (10,663/228,946 allèles ; 4.7%) et la MAF de 4.2% donnée par la base de données gnomAD pour les Européen non-Finlandais. Comme dans l'analyse de la cohorte HGID, nous n'avons pas observé de différence significative dans la proportion de porteurs hétérozygotes de P1104A entre les deux groupes (Tableau 5). Aucune association statistique, suivant les mêmes trois modèles de transmission génétique, entre le variant TYK2 I684S et la tuberculose n'a été trouvée (Tableau 5). La même situation que pour I684S a été observée pour le troisième variant commun, faux-sens et génotypé de la ressource UK Biobank, V362F (Tableau 5).

Pour être plus fin dans l'analyse du variant, et pour confirmer les origines ethniques auto-déclarés des individus de la ressource, une PCA des 115,093 volontaires et des 2,504 personnes d'origines connues de la base de données 1000G a été effectuée. La PCA résultante et la distribution des homozygotes P1104A sont illustrées sur la Figure 29. Plus de 97% des 115,093 participants, y compris les 234 homozygotes P1104A, sont d'origine Européenne selon la PCA. Ce pourcentage est en accord avec le fait que 91% des participants se soient auto-déclarés Britanniques, 2% Irlandais et 4% appartenant à un groupe ethnique «blanc» autre, tel que défini dans le champ #21000 de la ressource UK Biobank. Les 3% restants, qui ont déclaré être d'une origine non-Européenne ou mixte (Tableau 6), se situent dans la PCA de la Figure 29 parmi les quatre autres groupes ethniques majeurs de la base de données 1000G. Parmi les 234 homozygotes P1104A, 212 témoins et cinq patients tuberculeux ont déclaré être d'origine Britannique, quatre témoins et le sixième tuberculeux se sont déclarés d'origine Irlandaise, et les douze témoins restants ont déclaré appartenir à d'autres origines « blanches » (Tableau 7). Une analyse d'enrichissement ajustée sur les trois premières composantes principales a été conduite pour tenir compte de cette légère variabilité génétique

de la cohorte, potentiellement significative dans le cas où par exemple certains homozygotes auraient été d'origines non-Européennes. Après ajustement sur les composantes principales, l'OR obtenu fut de 5.0 (1.96-10.31 ; $P = 2 \times 10^{-3}$), plus proche de l'OR brut de 4.98 estimé sur les 589 et les 111,345 cas et témoins d'origine uniquement Européenne de la cohorte, reflétant une forte prédominance Européenne dans l'analyse.

Figure 29: PCA des 115,093 participants de la sous-cohorte UK Biobank utilisée pour notre analyse et des 2,504 individus non apparentés de 1000G



Les pourcentages de variance expliqués par PC1 et PC2 sont indiqués entre parenthèses. Les croix de couleur représentent les cinq principaux groupes ethniques de 1000G: les Africains (AFR, violet), les Américains (AMR, vert clair), les Asiatiques de l'Est (EAS, orange), les Européens (EUR, vert foncé) et les Asiatiques du Sud (SAS, bleu). Les cercles gris clair représentent les individus de la cohorte UK Biobank. Les porteurs homozygotes de l'allèle TYK2 P1104A sont indiqués par des losanges rouges (cas) ou gris foncé (témoins).

Tableau 5: Distribution des variants TYK2 P1104A, I684S et V362F parmi cas et témoins et estimation des ORs pour les modes génétiques de transmission dominant et récessif

		Cas (%)	Témoins (%)	OR (IC à 95%) ^a			
				Dom	Dom (aj)	Rec	Rec (aj)
P1104A	GG	567 (91)	104,038 (91)	0.93	0.94	4.90	5.00
	CG	47 (8)	10,207 (9)	(0.69-1.22)	(0.71-1.24)	(1.93-10.10)	(1.96-10.31)
	CC	6 (1)	228 (0.2)	P=0.62	P=0.71	P=0.002	P=0.002
I684S	AA	525 (82)	96,158 (82)	1.01	1.03	0.89	0.82
	CA	113 (18)	20,463 (18)	(0.82-1.22)	(0.84-1.25)	(0.32-1.92)	(0.32-1.96)
	CC	5 (0.8)	1,028 (0.9)	P=0.96	P=0.80	P=0.79	P=0.90
V362F	CC	318 (49)	60,050 (51)	1.07	1.06	1.01	1.00
	AC	275 (43)	48,434 (41)	(0.91-1.25)	(0.91-1.24)	(0.75-1.33)	(0.74-1.31)
	AA	53 (8)	9,580 (8)	P=0.41	P=0.43	P=0.93	P=0.97

Note: Les données proviennent de la ressource génétique UK Biobank.

^a Une régression logistique a été utilisée pour obtenir ORs et p-values (suivant un test de ratio de vraisemblances), et les trois premières composantes principales de la PCA ont été ajoutées pour obtenir les ORs et les p-values ajustés.

Tableau 6: Auto-identification des 115,093 individus génotypés au locus TYK2 P1104A de la cohorte UK Biobank en termes d'origines ethniques, pour lesquels une information sur le champ tuberculose « diagnostiquée par un médecin » a été notée

	Cas (%) n=620	Témoins (%) n=114,473	Total (%) n=115,093
Britannique	537 (87)	104,752 (91)	105,289 (91)
Irlandais	25 (4)	2,518 (2)	2,543 (2)
“Blanc” autre	27 (4)	4,039 (4)	4,066 (4)
Autre	31 (5)	3,128 (3)	3,159 (3)

Tableau 7: Auto-identification des 234 individus homozygotes pour le variant TYK2 P1104A de la cohorte UK Biobank en termes d'origines ethniques, pour lesquels une information sur le champ tuberculose « diagnostiquée par un médecin » a été notée

	Cas (%) n=6	Témoins (%) n=228	Total (%) n=234
Britannique	5 (83)	212 (93)	217 (93)
Irlandais	1 (17)	4 (2)	6 (2)
“Blanc” autre	0 (0)	12 (5)	12 (5)
Autre	0 (0)	0 (0)	0 (0)

Néanmoins, l'OR estimé n'est probablement qu'une forte sous-estimation de l'OR de développer une tuberculose après infection. Les participants sont nés entre 1940 et 1970, lorsque l'incidence annuelle moyenne de la tuberculose passait de 75 pour 100,000 dans les années 1950 à moins de 20 pour 100,000 dans les années 1970 (5). Même si aujourd'hui une estimation de la prévalence de l'infection tuberculeuse chez UK Biobank n'est pas disponible, elle est probablement inférieure à 10%, ceci étant le cas, selon des tests de production d'IFN- γ (*IFN- γ release assays*, IGRA), chez un groupe d'infirmières (228) et de prisonniers (229) Britanniques, soumis naturellement à un fort risque d'exposition à *M. tuberculosis*. Par exemple, la prévalence de l'infection tuberculeuse rapportée par Gray et ses collègues est de 6.9% chez des prisonniers âgés de plus de 45 ans (229) et de 7.6% chez les infirmières (225). Ces résultats sont en accord avec la constante baisse de l'incidence de la tuberculose dans la population née au Royaume-Uni, qui était de 3.2 pour 100,000 en 2016. L'incidence globale est cependant de 10.2 pour 100,000 (224) aujourd'hui, reflétant la forte incidence de 49.4 pour 100,000 au sein de la population née en dehors du Royaume-Uni mais vivant au Royaume-Uni. Globalement, ces observations indiquent que la grande majorité (> 90%) des témoins d'origine Européenne de la base de données UK Biobank a peu de chances d'avoir été exposé, et bien sûr donc, infecté, par *M. tuberculosis*.

Pour obtenir une estimation plus précise de l'OR de développer une tuberculose après infection, un scénario idéal dans lequel tous les individus de la ressource auraient été infectés a été considéré. Nous avons donc estimé le nombre d'individus, qui dans ce groupe idéal, aurait développé une tuberculose. En premier lieu, nous avons fixé la probabilité que des individus infectés, non porteurs du génotype P1104A à risque, aient développé une tuberculose à 5%, qui est le chiffre généralement admis pour la population générale. Pour le restant des témoins, c'est-à-dire, les homozygotes P1104A, nous avons dû estimer la pénétrance. Pour cela, nous nous sommes basés sur l'OR obtenu dans l'étude menée sur la cohorte HGID (221). Nous avons défini F_{TYK2TB} comme la probabilité de développer une tuberculose chez les homozygotes P1104A (c'est-à-dire la pénétrance) et F_{TB} comme la probabilité de développer une tuberculose chez le restant des individus. Par définition, l'OR de développer une tuberculose pour les homozygotes P1104A peut s'écrire comme suit:

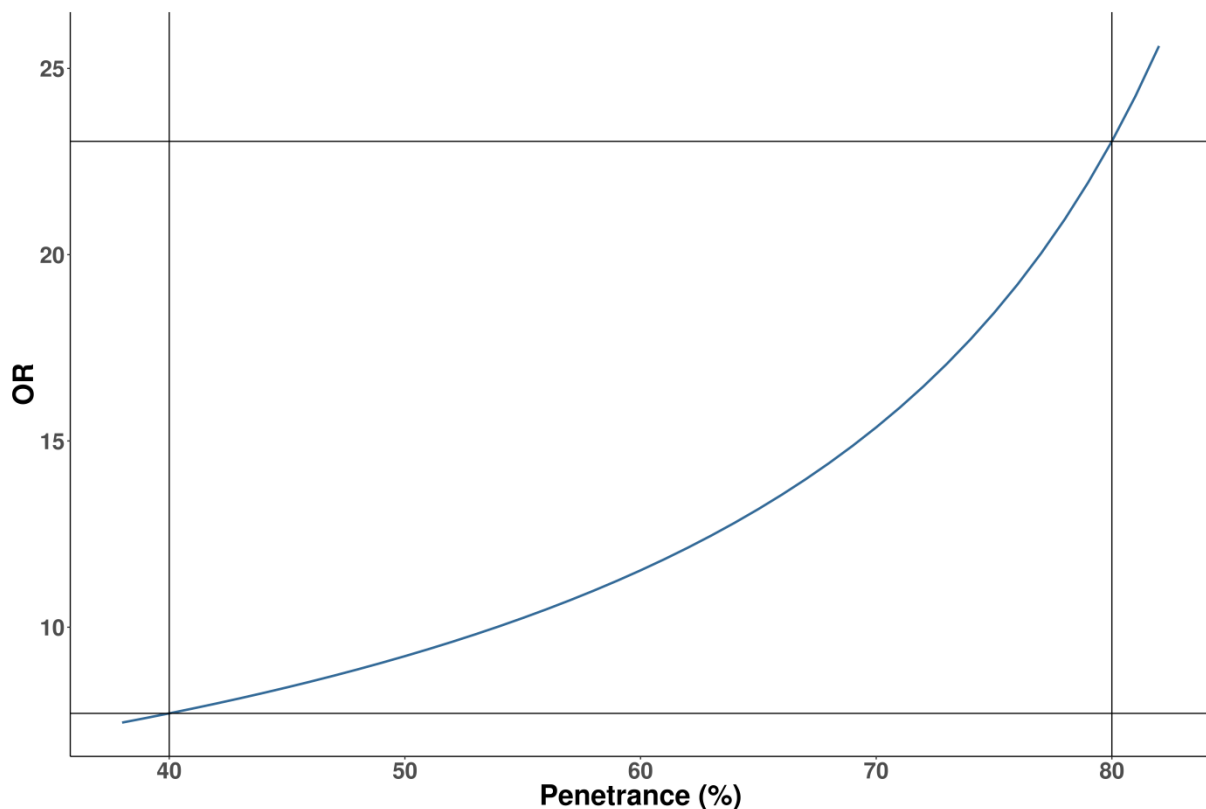
$$OR = \frac{F_{TYK2TB} (1-F_{TB})}{F_{TB} (1-F_{TYK2TB})}$$

Ce qui nous emmène à :

$$F_{TYK2TB} = \frac{OR \times F_{TB}}{[(OR \times F_{TB}) + 1 - F_{TB}]}$$

En fixant F_{TB} à 5%, comme expliqué précédemment, et l'OR à 89.31, F_{TYK2TB} est désormais égal à 82.4%. Sur la base de ces chiffres, nous avons donc retiré aléatoirement 5% des individus appartenant au groupe des témoins non homozygotes, qui auraient théoriquement développé une tuberculose, et, pour rester conservateurs, 40 à 80% des témoins homozygotes qui auraient, sur la base de la pénétrance estimée, développé aussi une tuberculose. Les ORs estimés sous cette simulation sont montrés sur la Figure 30, sous la forme d'une courbe en fonction de la pénétrance des homozygotes. Par exemple, il a été estimé à 7.7 et à 23.0 pour une pénétrance de 40 et 80%, respectivement. Malgré l'utilisation d'estimations brutes pour cette simulation, il est clair que l'effet de l'homozygotie pour P1104A sur le risque de développer une tuberculose en cas d'infection est encore bien plus fort.

Figure 30: Estimation de l'OR de développer une tuberculose après infection chez les homozygotes P1104A



L'estimation des ORs est représentée en fonction de la pénétrance de l'homozygotie pour P1104A allant de 40% à 80%.

Les analyses ici présentées confirment l'hypothèse que l'homozygotie pour P1104A est une étiologie génétique relativement commune de la tuberculose, compte tenu de sa MAF. Nous avons cherché à estimer la proportion d'homozygotes P1104A parmi les cas de tuberculose dans les régions du monde où nous avons des patients. Les sept patients homozygotes P1104A rapportés dans l'étude de la cohorte HGID (221) étaient originaires du

Brésil, du Chili, de l'Algérie, du Maroc et de la Turquie, pays où l'incidence de la tuberculose est comprise entre 17 pour 100,000 (Chili) et 99 pour 100,000 (Maroc). Sur la base cette étude et de la MAF moyenne du variant dans ces pays (1.4%), nous avons estimé le nombre de cas de tuberculose dus à l'homozygotie pour P1104A à 0.33% (IC à 95%: 0.17 à 0.40%). La cohorte UK Biobank, bien plus homogène, a été utilisée pour garantir une meilleure précision dans ce calcul. Sur la base des 589 cas de tuberculose et des 111,345 témoins d'origine Britannique, Irlandaise et d'autres origines Européennes, correspondant à 97% de l'échantillon initial (Tableau 6), les homozygotes P1104A représentent 1.02% des cas de tuberculose d'origine européenne (6/589), et seulement 0.2% des témoins (228/111,345). Globalement, ces résultats indiquent que l'homozygotie pour l'allèle TYK2 P1104A représente environ 1% des cas de tuberculose en Grande-Bretagne et, par inférence, probablement un pourcentage similaire dans d'autres populations d'origine Européenne.

C.3. Discussion

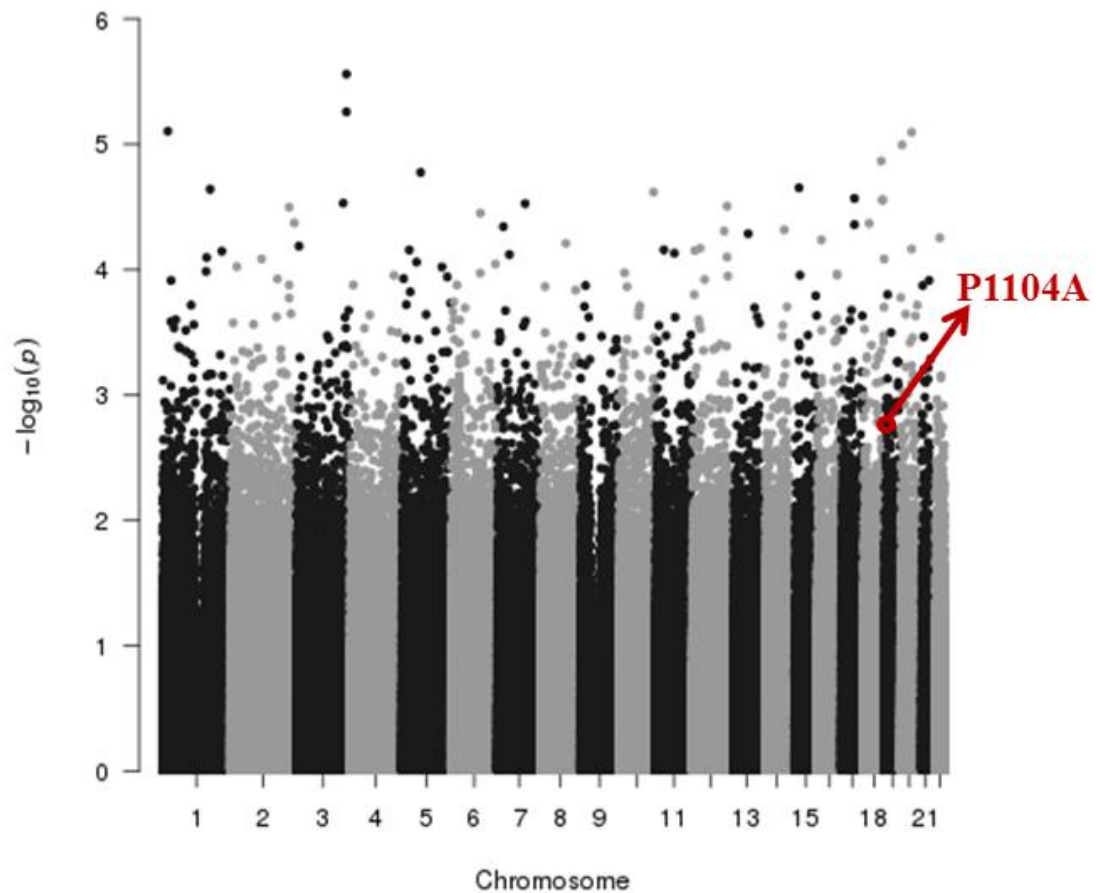
Le variant TYK2 P1104A affecte sélectivement l'immunité anti-mycobactérienne médiée par l'IFN- γ et dépendante de l'IL-23, ce qui explique la susceptibilité des homozygotes aux infections par mycobactéries. L'homozygotie pour TYK2 P1104A confère en effet une forte prédisposition à la tuberculose (et en moindre mesure au syndrome MSMD), chez des personnes d'origines ethniques diverses (étude de la cohorte HGID), originaires de pays où la tuberculose est plus ou moins endémique, mais aussi, spécifiquement chez des personnes d'origine Européenne (étude de la cohorte UK Biobank), originaires du - et résidant au - Royaume-Uni après la Seconde Guerre mondiale. Cette dernière étude est d'autant plus intéressante car les participants vivent tous dans un pays de faible endémicité pour la tuberculose. Il est peu probable que plus de 10% des Britanniques soient infectés par *M. tuberculosis* comme cela a été expliqué avant, impliquant, en particulier, que l'OR estimé pour les homozygotes P1104A de développer une tuberculose après infection (OR = 5) soit sous-estimé. Il est plus probable, selon nos estimations, qu'il soit supérieur à 10.

Cet OR est par ailleurs bien supérieur à celui rapporté pour les quelques SNPs identifiés par des GWAS dans le cadre d'études reliées au phénotype tuberculose (voir **Introduction**). Aucun de ces GWAS, ni, en particulier, les douze GWAS cités en **I**, n'a détecté l'effet de P1104A. Ces derniers n'ont pas signalé dans leurs études avoir testé un mode génétique de transmission récessif, ce qui, compte tenu de la MAF du variant et les tailles des échantillons analysées, aurait été fondamentale pour capturer l'effet du génotype à risque. Ceci fut également le cas pour le GWAS réalisé sur le phénotype « tuberculose auto-

déclarée » de la ressource UK Biobank et accessible dans Gene Atlas (230). En outre, la plupart de ces études, en particulier les dix études réalisées en Amérique latine ($n = 1$), en Afrique ($n = 5$) et en Asie ($n = 4$), pour lesquelles la fréquence de l'allèle P1104A est plus faible (~1% en Amérique latine et en Asie du Sud-Est et vraisemblablement absent en Afrique sub-saharienne et en Asie de l'Est), auraient eu un manque considérable (ou absolu) de puissance statistique pour le détecter.

Pour creuser en plus de détail la question qui se pose par rapport à la capacité des GWAS à identifier P1104A sans aucune considération a priori, nous avons conduit une analyse d'association pan-génomique sur le phénotype « tuberculose diagnostiquée par un médecin » de la ressource UK Biobank à l'aide des données de génotypage disponibles, supposant bien sûr un mode génétique de transmission récessif (Figure 31). La *p value* de 2×10^{-3} obtenue pour P1104A est bien inférieure au seuil de significativité du génome (5×10^{-8}). Cependant, une importante limite de notre étude est le nombre relativement petit de cas de tuberculose diagnostiqués par un médecin ($n = 620$). Un simple calcul montre que dans une étude cas-témoins équilibrée, supposant 1% d'homozygotes parmi les cas et 0.2% parmi les témoins, il nous aurait fallu 5,550 cas et 5,550 témoins pour atteindre une *p value* significative à l'échelle du génome entier de 5×10^{-8} . Alors qu'aucun allèle n'a atteint une *p value* $< 2 \times 10^{-6}$ dans notre GWAS, 121 variants génétiques avec une *p value* plus significative que P1104A ($< 2 \times 10^{-3}$) et une MAF $> 2\%$, ont rapporté des ORs supérieurs à 5, suggérant que d'autres étiologies génétiques récessives et relativement fréquentes de la tuberculose pourraient être cachées dans cette cohorte.

Figure 31 : Diagramme Manhattan d'une analyse d'association pan-génomique sur le phénotype « tuberculose diagnostiquée par un médecin » de la ressource UK Biobank



L'analyse a été réalisée sur 718,232 allèles génotypés dans le cadre d'un modèle génétique de transmission récessif.

L'homozygotie pour TYK2 P1104A semble être à l'origine d'une proportion non négligeable de cas de tuberculose, représentant peut-être environ 1% des cas de tuberculose en Europe et environ 0.33% des cas dans la plupart des autres régions du monde (en dehors de l'Europe, de l'Afrique subsaharienne, ou de l'Asie de l'Est). Bien que ces pourcentages soient estimés à partir d'un petit nombre de patients, il correspond à des millions d'individus au cours des derniers millénaires. En effet, l'estimation présentée en **I_H.1** d'un milliard de décès dus à la tuberculose en Europe au cours des 2,000 dernières années impliquerait qu'environ dix millions de ces personnes seraient décédées à cause de la susceptibilité accordée par le génotype TYK2 P1104A à risque. Des futures études basées sur des données d'ADN ancien vont permettre d'étudier la fréquence de cet allèle, qui semble avoir diminué au cours des 4,000 dernières années, à travers le temps et l'espace. Si cette dernière observation est confirmée, l'impact du génotype à risque pendant les dernières 2,000 années devrait être réévalué, étant très probable que celui-ci ait pu être encore plus important.

Dans un contexte très différent, des GWAS effectuées sur de larges échantillons de population, cette fois, pour des phénotypes autres que la tuberculose, ont montré que l'homozygotie pour TYK2 P1104A avait un effet protecteur considérable (OR allant de 0.1 à 0.3) contre diverses affections auto-inflammatoires ou auto-immunes (231–237). À la lumière de ces résultats, les potentiels avantages pharmacologiques des inhibiteurs de TYK2 dans le traitement des affections auto-inflammatoires ou auto-immunes sont en cours d'évaluation (238–240). Nos résultats indiquent que si de tels traitements devaient être généralisés, il serait important d'évaluer le risque de tuberculose avant et pendant le traitement, comme c'est actuellement le cas pour l'immunothérapie anti-TNF (241).

Un grand nombre de cas de tuberculose liés à P1104A aurait aussi des implications majeures pour la prévention et le traitement de la tuberculose. Il devrait être plus facile de cibler les personnes à risque élevé de développer la maladie lors de la définition des cohortes optimales d'essais de vaccins contre la tuberculose. Les stratégies de vaccination devraient viser à protéger les 5% des personnes qui ne sont pas génétiquement « résistantes » à la tuberculose. Des tests génétiques pour détecter le génotype à risque pourraient également être justifiés avant de voyager dans des pays où la tuberculose est fortement endémique. De plus, le diagnostic d'homozygotie pour P1104A chez les patients atteints de tuberculose pourrait ouvrir la voie au conseil génétique dans leurs familles. Encore, des injections d'IFN- γ recombinant seraient probablement bénéfiques pour ces patients, comme pour les patients présentant un déficit en IL-12R β 1 (46,242,243) surtout dans le contexte actuel d'une augmentation accrue et préoccupante de la résistance aux médicaments antituberculeux (voir **Introduction**). Enfin, cette découverte fait le lien théorique entre étiologies monogéniques rares (variants rares avec un effet fort sur le phénotype) et facteurs de risque communs (variants communs avec un effet modeste) (244–251).

Une approche statistique pour la détection d'une hérédité digénique à l'échelle du génome entier à partir de données WES.

A. Introduction.

L'introduction des données WES (et des données NGS en général) a permis, entre autres, l'étude systématique des variants rares et à fréquence intermédiaire. Elles ont conduit, au long des deux dernières décennies, à des découvertes majeures dans le domaine de la génétique humaine, en particulier dans le contexte de pathologies monogéniques rares (172–175), comme le syndrome MSMD. Elles ont contribué à la description d'un nombre non négligeable des 3,766 gènes responsables de pathologies monogéniques décrites en OMIM (<https://www.omim.org>). Cependant, dans ce même contexte de maladies dites monogéniques, l'existence de porteurs asymptomatiques de variants pathogéniques, souvent repérés à partir d'études familiales, relève un défi qui n'a pas été jusqu'ici largement étudié sur la base de données NGS. La probabilité qu'un individu porteur d'une mutation ou d'un génotype particulier, qui a été défini comme pathogène, présente des signes cliniques au cours d'une période spécifique est communément appelé la pénétrance. Si cette probabilité est égale à 1, le génotype pathogène est dit à pénétrance complète tandis que dans le cas contraire, la pénétrance est dite incomplète. Des cas de pénétrance incomplète peuvent être associés à des combinaisons de facteurs génétiques et environnementaux, mais ils demeurent souvent très mal compris.

Au début du séquençage des premiers génomes humains, il est devenu évident que des individus en bonne santé pouvaient porter un nombre assez important de variants génétiques potentiellement délétères, sans subir néanmoins d'effets néfastes évidents (252–255). Les raisons sous-jacentes inter- et intra-individuelles peuvent être nombreuses et variées. La mutation en question peut ne provoquer qu'un phénotype clinique léger. L'âge d'apparition de la pathologie étudiée peut également être tardif, l'expression étant par exemple dépendante de l'âge. Une dernière possibilité que nous citons ici, qui nous intéresse spécialement dans ce travail, est que certains défauts génétiques ne s'expriment cliniquement qu'en présence d'autres facteurs génétiques et/ou environnementaux. Avec la base de données 1000G vinrent les premières observations de mutations pathogéniques portées par des individus « sains ». Un des exemples classiques est l'identification par Xue et ses collègues (253) d'un porteur homozygote du variant HBB (Glu7Val), qui est responsable de la drépanocytose, parmi le groupe des Africains Yoruba (YRI). La mutation Cys282Tyr (*rs1800562*) du gène de

l'hémochromatose (*HFE*) fait aussi l'exemple d'une situation de pénétrance incomplète associée à une pathologie récessive (256) et son cas est assez remarquable. Le génotype homozygote de la mutation est présent chez environ une personne sur 200 d'origine Nord-Européenne et est responsable de 80 à 90% des hémochromatoses héréditaires (257,258). Bien que l'homozygotie du variant présente une pénétrance biologique relativement élevée (c'est-à-dire une accumulation de fer), sa pénétrance clinique est faible (259). Les données disponibles suggèrent que 38 à 50% des homozygotes développent une surcharge en fer et que 10 à 25% développent des symptômes cliniques associés à l'hémochromatose (260). Cependant, des études supplémentaires à large échelle ont montré un effet important du sexe dans ces statistiques: chez des homozygotes du variant *rs1800562*, testés spécifiquement pour une maladie du foie, la maladie s'est manifestée chez 24 à 43% des hommes, mais seulement chez 1 à 14% des femmes (261). Certains modificateurs liés à l'environnement, eux-mêmes liés au sexe, tels le régime alimentaire ou la consommation d'alcool ont été étudiés dans ce contexte. Par contre, la pénétrance du gène *HFE* est aussi modulée par d'autres composantes génétiques influençant l'expression clinique de l'hémochromatose. Parmi ces composantes on retrouve des mutations dans les gènes *HAMP*, *HFE2* et *TFR2* et des polymorphismes dans les gènes *BMP2*, *BMP4*, *CYBRD1*, *HP*, *LTA*, *MPO*, *TMPRSS6* et *TNF* (257,262,263).

La liste des pathologies monogéniques associées à une pénétrance incomplète peut s'avérer très longue, renforcée ces derniers temps par le nombre croissant d'individus séquencés dans le monde. Par ailleurs, certaines mutations responsables d'un trait clinique défini, peuvent déterminer un large éventail de manifestations cliniques (allant du cas du porteur asymptomatique, au cas d'une maladie potentiellement mortelle). Ceci est d'ailleurs le cas dans la déficience associée au gène *IL12RB1*, causant le syndrome MSMD (153,155). Globalement, ces observations, au sein de certaines pathologies dites monogéniques, suggèrent fortement que des contributions supplémentaires, autres que les lésions mono- ou bi-alléliques, jouent un rôle important dans les manifestations cliniques de ces patients. Dans ce contexte, nous nous sommes spécifiquement focalisés sur les contributions supplémentaires de nature génétique (264,265). L'hérédité digénique (*digenic inheritance*, DI) est, à cet égard, le modèle le plus simple (266,267). Il est défini théoriquement par la présence d'allèles à deux loci différents, qui ne sont pas en déséquilibre de liaison (LD), et qui sont nécessaires et suffisants pour définir un phénotype. Les DI peuvent, en particulier, expliquer des situations de pénétrance incomplète pour des pathologies dites monogéniques, mais pour lesquelles seul un des deux loci sous-jacents à la maladie est connu. En pratique, des cas de DI sont aussi rapportés dans des situations où un locus modificateur est responsable d'une variabilité de

l'expression phénotypique (par exemple, par rapport au degré de sévérité de la maladie) chez des individus porteurs d'une lésion monogénique définissant un phénotype établi (268,269). Cette situation correspond par exemple au cas déjà décrit de l'hémochromatose ou encore à celui de la mucoviscidose, pour laquelle de nombreux variants génétiques modificateurs de l'expression phénotypique ont été rapportés, sur lesquels nous reviendrons. Dans ce travail, les composantes génétiques déterminant cette variabilité phénotypique sont dénommés « gènes modificateurs » et notés GM. Au cours des 18 dernières années, le nombre de DI décrites dans des maladies rares a considérablement augmenté (266). Aujourd'hui, des informations détaillées concernant 258 combinaisons digéniques dans 54 pathologies sont accessibles sur <http://dida.ibsquare.be/>, une base de données récente qui est appelée DIDA.

Les exemples de DI rapportées dans la littérature restent rares. Ceux où l'interaction biologique des gènes impliqués est connue le sont encore plus. Un tel cas a été rapporté dans la dystrophie musculaire facioscapulohumérale de type 2, qui nécessite à la fois un fond génétique permettant des transcriptions stables du gène *DUX4* (sur le chromosome 4) et des mutations au locus *SMCHD1* (sur le chromosome 18) entraînant une hypométhylation et l'expression de *DUX4* (270). Plus récemment, un exemple remarquable de DI expliquant une pénétrance incomplète a été fourni dans l'étude de 191 probands atteints de craniosynostose, maladie causant la fusion prématurée des articulations du crâne chez des nouveau-nés. Timberlake et ses collègues (2016) ont trouvé un enrichissement de mutations rares et délétères du gène *SMAD6* chez ces patients, présentant toutefois une pénétrance incomplète (<60%). Dans ce même travail, l'étude d'un variant commun proche de *BMP2*, un gène lié à *SMAD6*, a élucidé la quasi-totalité de la variation phénotypique observée, suggérant un défaut digénique pour ces individus. Sinon, la majorité des DI rapportés sont liés à des GM, influençant l'expression d'un phénotype clinique, déterminé principalement par une lésion monogénique. Par exemple, une étude basée sur des données WES a montré que deux variants faux-sens et peu fréquents du gène *DCTN4* étaient associés à la sévérité des infections pulmonaires par *Pseudomonas aeruginosa* (268) chez des patients atteints de mucoviscidose.

Malgré ces découvertes, seulement 1% des 5,442 pathologies dites monogéniques listés dans OMIM présentent également une DI, comme cela a été indiqué par DIDA, contrastant avec le nombre plus élevé de pathologies associées à une pénétrance incomplète. Ceci peut s'expliquer en partie par le fait que toutes les lésions digéniques rapportées dans la littérature ont été découvertes par des études de gènes candidats sur au moins un des loci impliqués. Ces découvertes sont biaisées dans le sens où elles se sont basées sur des *a priori* (connaissance biologique de la maladie, résultats sur d'études précédentes, ...) pour exclure la

plupart des gènes du génome dans l'étude d'au moins un des deux loci. Elles n'ont pas utilisé des tests statistiques à l'échelle du génome entier pour les deux loci, ce qui aurait permis potentiellement la mise en évidence d'autres DI ou la confirmation, dans un sens génétique, du DI retrouvé. Certaines DI, comme dans l'exemple de la mucoviscidose présenté ci-dessus, ont été découvertes par des analyses monogéniques de patients porteurs de variants pathogéniques responsables d'un phénotype principal (268). L'exemple de la craniosynostose est unique car il s'est servi d'une analyse monogénique à l'échelle du génome entier (détection des variants *SMAD6* rares) et d'un seul variant commun (variant situé proche de *BMP2*) rapporté par une étude GWAS précédente (271,269). Pour les maladies génétiquement hétérogènes, telles le syndrome d'Alport pour lequel trois gènes pathogènes ont été décrit, ou le syndrome du QT long ou encore le syndrome Bardet-Biedl avec plus d'une douzaine de gènes pathogènes chacun, des DI impliquant ces mêmes gènes pathogènes ont été rapportées (272,273). Cependant, d'autres variants modificateurs peuvent être cachés dans des gènes ayant jusqu'ici un impact fonctionnel méconnu sur la maladie, voire aucun effet détectable. De la même manière, de nombreuses affections héréditaires sans étiologie génétique connue, masquées parfois sous forme de cas sporadiques, peuvent être dues à une DI.

Dans l'étude que je présente par la suite, j'ai développé une méthode statistique originale, que je décris en **III_B**, servant à analyser l'hérédité digénique, en particulier, de maladies peu fréquentes. Pour valider cette méthode, j'ai conduit de nombreuses études de simulations, à la fois sous l'hypothèse nulle d'absence d'interaction, par l'analyse de l'erreur de type I, et sous l'hypothèse alternative d'interaction génétique au niveau d'une paire de gènes, par l'étude de sa puissance. J'ai prouvé la fiabilité de la méthode par la réplication d'une lésion digénique décrite préalablement dans la littérature chez des patients atteints de craniosynostose (269), impliquant des variants rares du gène *SMAD6* et un variant commun situé proche du gène *BMP2*. J'ai par ailleurs débuté l'analyse de DI au sein des patients MSMD de notre cohorte, rapportant un premier GM candidat dans la déficience la plus commune du syndrome.

B. Une méthode statistique à l'échelle du génome entier pour la recherche de DI.

Le développement de méthodes statistiques adaptées à l'échelle du génome entier sur deux loci et visant à rechercher une DI dans des pathologies rares à l'aide de données NGS a acquis un intérêt particulièrement important dans le domaine de la génétique

épidémiologique. Dans ce contexte, les données WES apparaissent comme les plus pertinentes, étant donné que les variants délétères à effet fort sont plus susceptibles d'endommager les séquences génomiques codantes (274,275). Cependant, deux défis majeurs se posent dans la conduction d'analyses d'interactions par paires à partir de données WES. Le premier est le manque de puissance statistique pour détecter des interactions entre variants rares car celles-ci sont encore plus rares que les variants qui les composent. L'autre est la longueur du temps de calcul dû au nombre de combinaisons possibles deux à deux. Le test que nous proposons ici agrège des variants rares au sein d'un gène qu'il utilise en tant qu'unité d'analyse, réduisant par-là notamment le nombre de tests à réaliser et le temps de calcul utilisé. C'est un test dit *case-only* (cas-seulement) qui ne nécessite pas le recrutement de témoins, évitant par là en plus le potentiel biais associé à leur inclusion.

B.1. Modèle d'agrégation de variants rares.

Les méthodes basées sur l'agrégation de variants sont classées dans la littérature comme des tests *burden*, des tests *adaptive burden*, des tests *variance-component* ou des combinaisons de ces trois classes (276). Nous proposons ici d'agréger des variants « rares » ou peu fréquents (par exemple, < 5%) au sein de l'unité d'un gène dans une variable binaire qu'on appelle « porteur » et qui définit le statut principal d'un individu dans l'analyse. Cette méthode est connue pour optimiser la puissance statistique sous une hypothèse d'homogénéité génétique, tout en permettant de tester l'interaction gène x gène en n'analysant que le nombre de combinaisons de gènes deux à deux (et non pas toutes les combinaisons de variants deux à deux). La méthode, définie formellement par la suite, est basée sur la méthodologie dénommée test de somme allélique au niveau d'une cohorte (*cohort allelic sum test*, CAST). Pour chaque gène j et un sous-ensemble de variants observés S_j dans ce gène, si n correspond au nombre d'individus étudiés, nous désignons le vecteur « porteur » par (g_{j1}, \dots, g_{jn}) ou tout simplement G_j . Les composantes g_{ji} du vecteur qui désignent, pour chaque $i = 1, \dots, n$, le statut principal d'un individu au sein d'un gène sont définis comme suit:

$$g_{ji} = \begin{cases} 1 & \text{Si l'individu } i \text{ "porte" au moins un variant dans } S_j \\ 0 & \text{autrement} \end{cases}$$

Dans ce contexte, le statut « porteur » d'un individu dépendra du modèle génétique choisi pour conduire l'analyse. Par exemple, à partir d'un modèle dominant, pour un gène j , $g_{ji} = 1$ si et seulement si l'individu i est porteur d'au moins une copie d'au moins un allèle des variants de S_j .

B.2. Modèle cas-témoin pour une analyse monogénique

La variable g_{ji} peut tout d'abord être interprété dans le contexte d'une étude monogénique utilisant des cas et des témoins. Dans ce contexte, le rôle potentiel des variants du gène j chez des cas phénotypiques peut être étudié par l'analyse du tableau de contingence 2 x 2 suivant, par exemple, à partir d'un test χ^2 ou de Fisher.

	Cas	Témoins
Porteurs d'au moins un variant dans S_j	a	b
Autres	c	d

Pour permettre la prise en compte éventuelle de covariables, le même test d'association peut être défini en utilisant un modèle de régression logistique. Si D_j est un vecteur de phénotypes binaire, G_j le score génétique défini préalablement, C_j la matrice des covariables et β_0 , β_1 et β_2 les coefficients de régression, le modèle de régression logistique s'écrit alors:

$$\text{logit}(P(D_j = 1)) = \ln\left(\frac{P(D_j=1)}{1-P(D_j=1)}\right) = \beta_0 + \beta_1 G_j + \beta_2 C_j \quad (1).$$

Les *p values* sont ensuite obtenues par un LRT consistant à comparer la vraisemblance des données observées entre le modèle saturé (incluant l'effet des covariables et du score génétique G) et le modèle nul (incluant uniquement l'effet des covariables). Dans ce contexte, la puissance du test est étroitement liée à l'efficacité avec laquelle les filtres appliqués éliminent au maximum le nombre de variants neutres par rapport au phénotype d'intérêt. En d'autres termes, la définition de S_j est clé dans la réussite du modèle conduit. Les filtres peuvent être choisis en fonction de différentes caractéristiques au niveau du variant, telles que la MAF ou leurs annotations définis dans la section **II_A**.

B.3. Modèle cas-témoin pour le test d'interaction

L'extension au modèle digénique suivant la même méthode peut se faire par l'étude de l'apport de deux gènes en simultané au phénotype d'intérêt. Les données relatives à deux gènes k et j dans un ensemble de données cas-témoins et un statut binaire phénotypique D peuvent être résumées en deux tableaux de contingence 2 x 2, un pour les individus atteints (cas, $D = 1$) et un autre pour les individus non atteints (témoins, $D = 0$) comme dans le Tableau 8.

Tableau 8: Tableau de contingence des porteurs de variants dans Sk ou Sj pour une paire donnée de gènes k et j et pour des individus atteints ou non atteints

		Gène k	
		Porteurs	Non porteurs
Gène j	Porteurs	$n_{kj,11}^i$	$n_{kj,01}^i$
	Non porteurs	$n_{kj,10}^i$	$n_{kj,00}^i$

^a $i = \{a, u\}$. Lorsque $i = a$, n représente le nombre d'individus atteints, lorsque $i = u$, n représente le nombre d'individus non atteints.

Soit $N_{kj}^a = (n_{kj,00}^a, n_{kj,10}^a, n_{kj,01}^a, n_{kj,11}^a)$ un vecteur du nombre observé de porteurs pour le gène k et le gène j parmi les cas, de sorte que, par exemple, $n_{kj,11}^a = \sum_{i \text{ in cases}} (g_{ki} \times g_{ji})$. De même, soit $N_{kj}^u = (n_{kj,00}^u, n_{kj,10}^u, n_{kj,01}^u, n_{kj,11}^u)$ un vecteur d'observation du nombre de porteurs pour le gène k et le gène j parmi les témoins. Les ORs pour les cas et les témoins associés aux gènes k et j sont donc respectivement définis comme suit:

$$OR_{kj}^a = \frac{n_{kj,11}^a \times n_{kj,00}^a}{n_{kj,10}^a \times n_{kj,01}^a}, \quad OR_{kj}^u = \frac{n_{kj,11}^u \times n_{kj,00}^u}{n_{kj,10}^u \times n_{kj,01}^u}.$$

Les analyses statistiques d'interaction reposent généralement sur la comparaison de OR_{kj}^a et de OR_{kj}^u . Plus spécifiquement, un modèle de régression logistique classique cas-témoin utilisé pour tester une interaction, et qui n'est qu'une extension de l'équation (1), est défini comme suit:

$$\text{logit } P(D = 1) = \beta_0 + \beta_k G_k + \beta_j G_j + \beta_I G_k \times G_j \quad (2).$$

Il peut être démontré que le coefficient d'interaction β_I est égal à $\log\left(\frac{OR_{kj}^a}{OR_{kj}^u}\right)$. De plus, ce modèle prend en compte les effets principaux de chaque gène indépendamment de l'autre, par la considération des coefficients β_k et β_j . Encore une fois, des covariables spécifiques, par exemple, des composantes principales (*principal component*, PC), peuvent être introduites dans le modèle. En incluant une matrice de covariables C et un vecteur X de coefficients, le modèle de régression logistique complet prend la forme suivante:

$$\text{logit } P(D = 1) = \beta_0 + \beta_j G_j + \beta_k G_k + \beta_I G_j \times G_k + XC \quad (3).$$

L'hypothèse nulle d'absence d'interaction définie par $\beta_I = 0$ peut être testée par un LRT à un degré de liberté, en présence ou non d'effets génétiques principaux et/ou de covariables.

B. 4. Modèle *case-only* pour le test d'interaction.

L'étude de l'interaction peut se faire également en ne se focalisant que sur les cas, de sorte que l'analyse repose uniquement sur le tableau de contingence 2 x 2 des personnes atteintes du Tableau 8. Les études *case-only* ont été développées dans le contexte de variants communs et étudiées largement au long des deux dernières décennies (273,277–281). Dans cette situation, le modèle de régression logistique destiné à tester l'interaction entre les gènes G_k et G_j s'écrit désormais comme suit:

$$\text{logit } P(G_k = 1) = \gamma_0 + \gamma_I G_j + XC \quad (4),$$

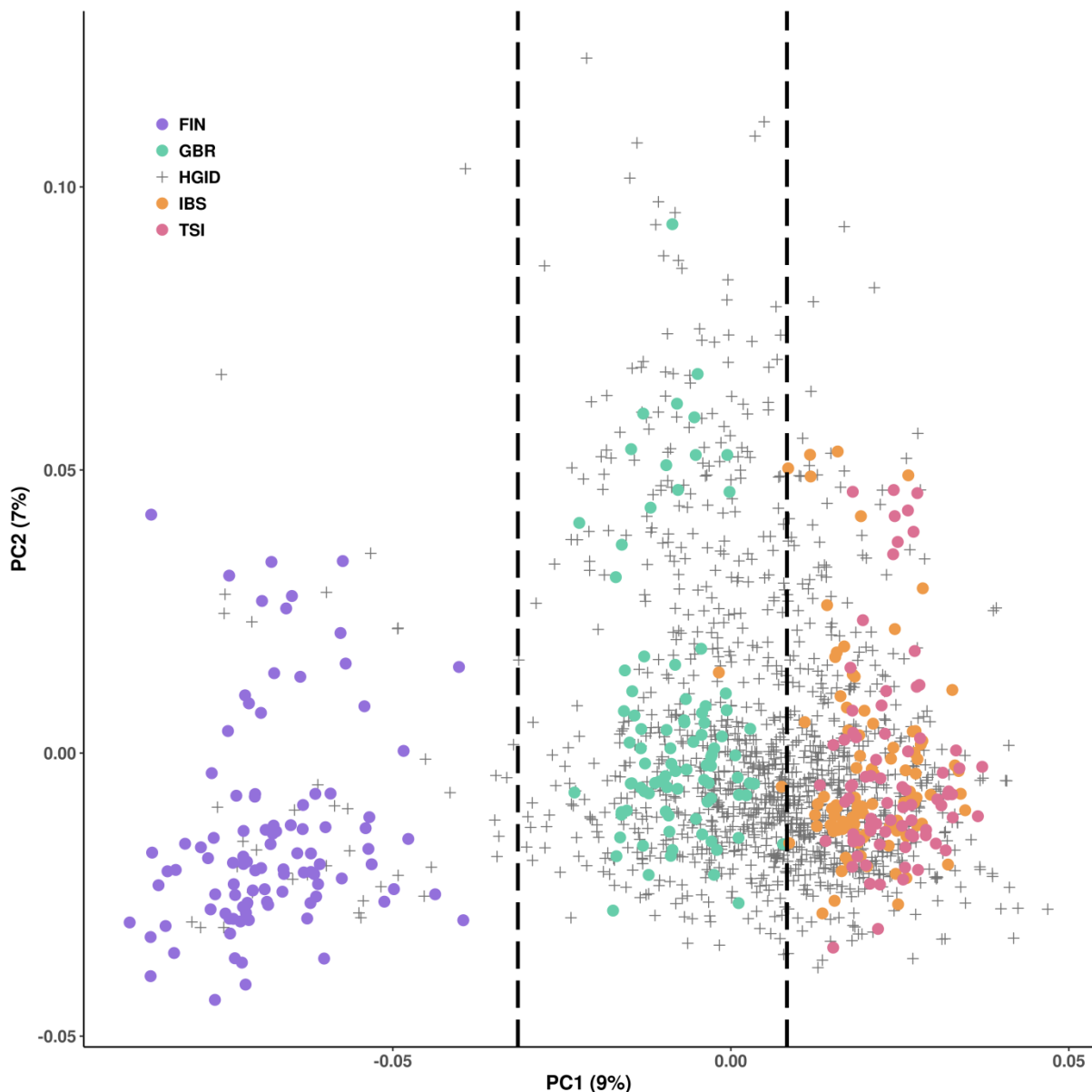
où γ_I est égal à $\log(OR_{kj}^a)$, C est une matrice de covariables et X un vecteur de coefficients. Comme précédemment, un LRT peut être utilisé pour tester l'hypothèse nulle $\gamma_I = 0$.

Quand les vecteurs G_k et G_j ne sont pas corrélés, ce qui signifie notamment que les variants des deux gènes ne sont pas en LD, une interaction existe si OR_{kj}^a est différent de 1. En plus, si la maladie est rare, OR_{kj}^u est proche de 1 et par conséquent β_I est approximativement égal à γ_I . Les avantages de ce test par rapport à un test cas-témoins ont été largement étudiés théoriquement par de nombreux auteurs dans le contexte de variants communs (280–284), en particulier, le gain de puissance. Ce gain provient de la nature des estimateurs des coefficients d'interaction des deux méthodes. Ces estimateurs dépendent soit du rapport (OR_{kj}^a/OR_{kj}^u) pour le test cas-témoin, soit uniquement d' OR_{kj}^a pour le test de *case-only*. Les variances asymptotiques des estimateurs sont la somme des comptes réciproques du Tableau 8, soit pour les cas et les témoins (modèle cas-témoin), soit uniquement pour les cas (*case-only*) (282). Par conséquent, la variance de l'estimateur du coefficient d'interaction cas-témoin est plus grande, menant à un test moins puissant. Parmi les avantages de ce test on peut mentionner également le fait qu'il ne nécessite pas le recrutement de témoins, ce qui, mis à part le temps et les avantages économiques, évite le problème de classification erronée des individus avec le phénotype non atteint et l'introduction par-là d'un potentiel biais. Néanmoins, ce test repose sur l'hypothèse d'indépendance des variants testés dans la population générale. Dans des termes statistiques, les variants ne doivent pas être corrélés la population globale. C'est par le moyen de simulations sous l'hypothèse nulle, évaluant l'erreur de type I, que nous avons étudié les potentielles sources de violation de cette hypothèse dans le contexte de données WES.

C. Echantillons d'étude.

Pour l'étude de simulation, nous avons utilisé des données réelles d'exome afin d'étudier l'erreur de type I et la puissance statistique de la méthode décrite en **III_B**, utilisant notamment des échantillons des populations de 1000G et un sous-ensemble de la cohorte HGID décrite en **II_A**. Spécifiquement, six populations de la base de données 1000G ont été utilisées : quatre populations Européennes - Espagnols (IBS, $n = 107$), Italiens (TSI, $n = 107$), Anglais et Ecossais (GBR, $n = 91$) et Finlandais (FIN, $n = 99$), - et deux populations asiatiques d'origine chinoise - les Chinois Han (CHS, $n = 105$) et les Dai Chinois à Xishuangbanna (CDX, $n = 93$). Uniquement pour les études de puissance nous avons inclus aux données 1000G un sous-ensemble de la base de données HGID, comprenant 1,331 personnes d'origine Européenne, telles que définies par la PCA de la Figure 15 de la section **II_A**. Pour des études de stratification de population nous avons défini par la suite trois sous-populations Européennes sur la base de la PCA de la Figure 32: les «Européens du Nord» (N, y compris les Finlandais), les «Européens du centre-ouest» (M, y compris les Anglais et les Ecossais) et les «Européens du Sud» (S, y compris les Italiens et les Espagnols).

Figure 32: PCA des individus Européens utilisés dans l'étude de simulation de puissance



La figure montre la PCA (PC1 vs PC2) réalisée à partir des données WES de 1,735 Européens, dont 404 sujets 1000G (FIN, GBR, IBS et TSI) et 1,331 de notre base de données HGID. Les couleurs sont utilisées pour représenter les quatre sous-populations du projet 1000G (cercles) et les individus de la cohorte HGID (croix grises). Deux lignes verticales en pointillés ont été tracées le long de la PC1 sur la base du gradient classique nord-sud d'ascendance Européenne, définissant en outre trois sous-populations: les «Européens du Nord» (y compris FIN), « les Européens du centre-ouest » (y compris les GBR) et les « Européens du Sud » (y compris les IBS et les TSI).

D. Contrôle qualité des variants.

Pour l'étude de simulation et afin de travailler avec des données homogènes d'exome, l'analyse des variants a été restreinte au même kit de capture que celui utilisé pour la cohorte HGID, *SureSelect Human All Exon V4+UTR* (<https://agilent.com>). Les mêmes critères de qualité appliqués à la cohorte HGID et qui sont décrits en **II_A.1.1** ont été utilisés ici. Seuls les variants bialléliques, autosomiques et prédits faux-sens ou LOF par le logiciel SnpEff ont été conservés tandis que les variants filtrés par la base de donnée gnomAD r2.0.2 (indiqué

autrement par le statut bioinformatique « PASS ») ont été exclus. Afin de travailler avec des variants rares, sauf indication contraire explicite, seuls les variants avec une MAF $< \alpha$ dans la base de données 1000G ont été retenus. Toutes les analyses présentées dans l'étude de simulation ont été conduites sous un modèle de transmission génétique dominant pour chaque gène, c'est-à-dire, pour chaque gène j , les porteurs sont définis comme des individus porteurs d'au moins une copie d'au moins un allèle du groupe S_j défini en **III_B.1**.

E. Analyses de simulation.

Les propriétés de notre test *case-only* ont été étudiées au moyen de simulations à partir de données réelles WES provenant des populations 1000G et d'un sous-ensemble de notre base de données interne HGID. Comme mentionné précédemment, dans cette section sont décrites les analyses conduites sous l'hypothèse nulle d'absence d'interactions digéniques ainsi que les analyses de puissance. Tout au long de ces analyses, l'approche *case-only* a été comparée à l'approche cas-témoins correspondante, qui est décrite en **III_B**, sous différents scénarios impliquant divers effets génétiques et divers degrés de stratification de population.

E.1. Analyse de l'erreur de type I.

Méthode de simulation

L'analyse sous l'hypothèse nulle de l'erreur de type I de notre méthode *case-only* a été effectuée avec les six populations de la base de données 1000G décrites en **III_C**. Nous avons considéré deux niveaux de stratification de population pour nos analyses : une avec un effet très modeste impliquant des populations de l'Europe du Sud et l'autre avec un effet fort, mélangeant une moitié d'individus d'origine Européenne et une moitié d'origine Chinoise. Pour une première étude du test *case-only*, nous avons utilisé tous les individus de l'échantillon IBS + TSI ($n = 214$, voir **III_C**), présentant un effet faible de stratification de population. Nous nous sommes focalisés sur des tailles d'échantillon d'environ 200 individus par groupe phénotypique, ce nombre correspondant à la taille d'une éventuelle cohorte de patients avec une maladie rare, comme c'est par ailleurs le cas pour l'analyse présentée dans la section **III_E.3** sur la craniosynostose. Dans une deuxième analyse, nous avons considéré pour les cas, définis par le total des individus de l'échantillon utilisé, la population IBS + CHS ($n = 212$, voir **III_C**), présentant de forts niveaux de structure de population. Les potentiels problèmes liés à la taille de l'échantillon (relativement petit pour des études digéniques) ont été pris en compte en se focalisant uniquement sur les gènes pour lesquels le pourcentage de

porteurs de variants rares d'intérêt dans la population étudiée était d'au moins 15%. Par conséquent, après l'application des filtres décrits en **III_D**, 1,588 gènes ont été inclus dans les analyses IBS + TSI et 1,248 dans les analyses IBS + CHS, menant, respectivement, à 1,260,067 et à 776,879 tests d'interaction.

Nous avons comparé le test *case-only* au test cas-témoins correspondant, décrit en **III_B**. Deux combinaisons d'échantillons 1000G, IBS + TSI + GBR + FIN ($n = 404$) et IBS + TSI + CHS + CDX ($n = 412$), composés de quatre populations Européennes (échantillon EUROP), et de deux Européennes et deux Asiatiques (EURAS), respectivement, ont été définies pour cette tâche. La moitié des individus a été simulée comme des cas et l'autre comme témoins. Encore une fois, deux degrés de stratification de population ont été considérés dans la définition de deux scénarios : un scénario dit « équilibré » et un « non équilibré », à partir de la définition du statut cas/témoin de l'échantillon EURAS. Dans le scénario équilibré, 50% des individus de chacun des quatre échantillons constitutifs ont été considérés au hasard comme des cas ou des témoins. Dans le scénario non équilibré, 5/6 des individus des échantillons IBS + TSI a été considéré comme des cas (et 1/6 des témoins), et 1/6 des individus CHS + CDX a été simulé en tant que cas (et 5/6 en tant que témoins). En se focalisant uniquement sur les gènes dont au moins 15% des individus sont porteurs de variants rares d'intérêt et qui sont séparés d'au moins 2 Mb (pour tenir compte de l'effet potentiel du LD décrit par la suite), 1,563 gènes ont été retenus pour l'analyse équilibrée cas-témoins EUROP, tandis que 1,173 gènes ont été utilisés dans les analyses équilibrées et non équilibrées EURAS.

Résultats de simulation sans stratification de population

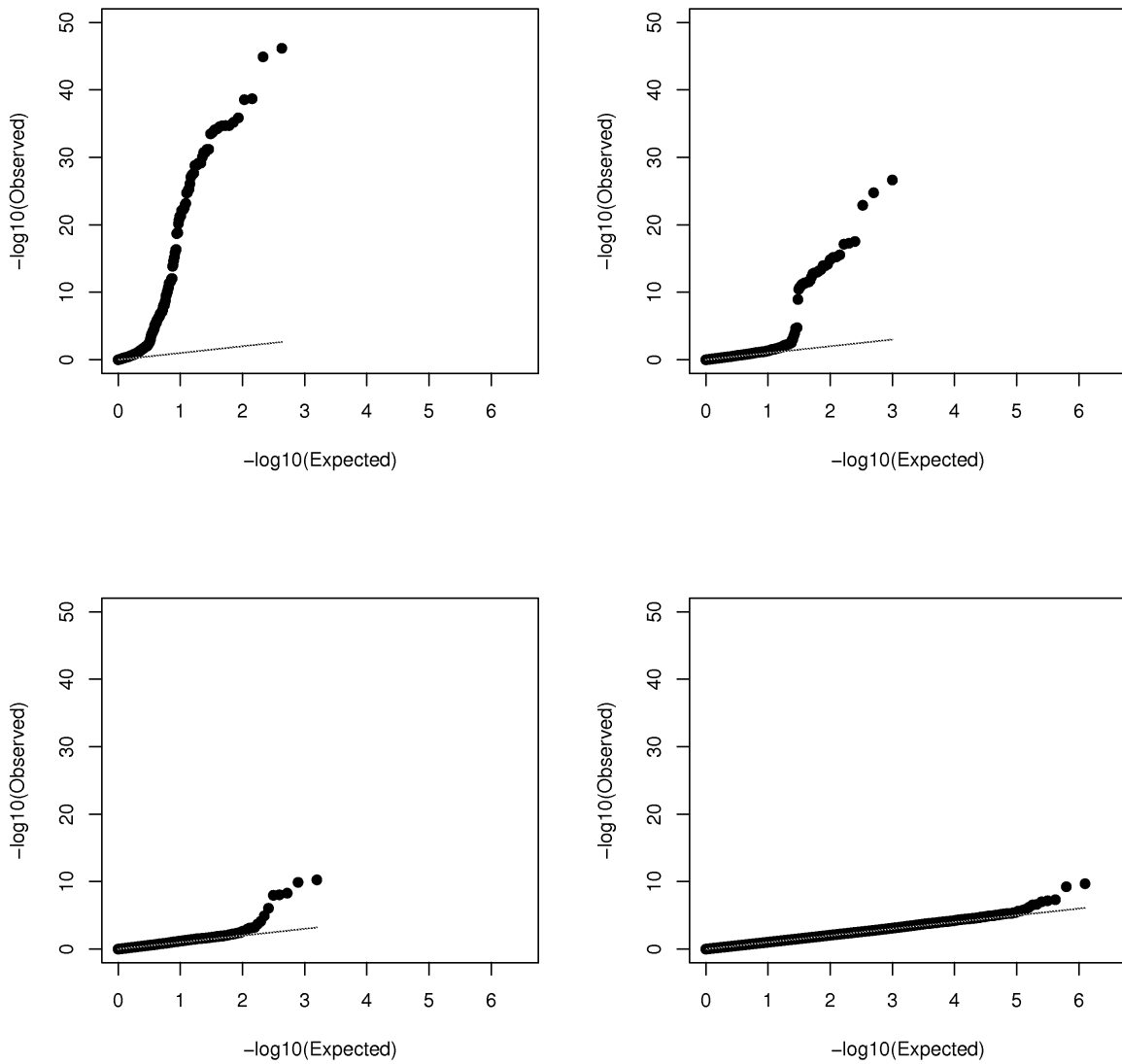
En testant toutes les paires de gènes possibles nous avons observé une inflation modérée de l'erreur de type I, soit 0.00147 pour un $\alpha = 0.1\%$ (Tableau 9). Pour identifier la source de cette inflation nous avons étudié le LD, précédemment associé à l'inflation de l'erreur de type I dans les tests *case-only* (277). Dans ce but nous sommes restreints à l'analyse des paires de gènes séparés physiquement par une distance minimale δ (mesurée en Mb). En faisant augmenter δ de 0.1 à 2 Mb, nous avons remarqué une diminution des erreurs de type I empiriques (Tableau 9), jusqu'à atteindre une erreur de type I de 0.00122 pour une valeur nominale α de 0.1% et un $\delta = 2$ Mb (Tableau 9). Nous avons observé une inflation spécialement importante dans la distribution des *p values* des tests n'incluant que des paires de gènes avec un $\delta < 2$ Mb (Figure 33). En particulier, les 204 *p values* $< 10^{-10}$ de l'analyse globale IBS+TSI proviennent des paires de gènes avec un $\delta < 2$ Mb. Aucune amélioration importante n'est observée pour un $\delta > 2$ Mb. Globalement, ces résultats ont prouvé que le LD

était à l'origine des *p values* les plus basses du test *case-only*. Cependant, après correction du LD, les erreurs de type I, bien que meilleures, sont restées légèrement au-dessus de la borne supérieure de l'intervalle de confiance (Tableau 10).

Tableau 9: Erreurs de type I empiriques à une valeur nominale $\alpha = 0.1\%$ pour le test *case-only* en fonction de la distance minimale entre les deux gènes d'une paire

<i>Case-only</i> (IBS+TSI)					
Distance minimales (Mb) entre les gènes d'une paire					
Erreur de type I (%)	0	0.1	0.5	1	2
0.1%	0.00147	0.00137	0.00126	0.00123	0.00122

Figure 33: QQ-plots pour le test case-only conduit sur la population IBS + TSI, en fonction de la distance entre les deux gènes de chaque paire



Des *QQ-plots* sont montrés pour le test *case-only* sur la population IBS + TSI, avec des distances entre les deux gènes d'une paire (A) $< 0,1$ Mb, (B) entre 0.5 et 1 Mb, (C) entre 1 et 2 Mb, et (D) > 2 Mb.

Tableau 10: Erreurs de type I empiriques à une valeur nominale $\alpha = 0.1\%$ pour les tests case-only et cas-témoins en l'absence de stratification de population

Méthode	Modèle	
	Pg_0^a	Pg_2^b
<i>Case-only</i> (IBS+TSI)	<i>0.00147</i> [0.0009-0.00110]	<i>0.00122</i> [0.0009-0.00110]
Cas-témoin (IBS+TSI+GBR+FIN)	<i>0.00128</i> [0.0009-0.00110]	<i>0.00128</i> [0.0009-0.00110]

Note: Les limites de l'intervalle de confiance à 95% est montré entre crochets. Les erreurs de type I se situant en dehors de ces limites sont marqués en italique.

^a Toutes les paires de gènes avec des variants avec une MAF<5% et des gènes avec >15% de porteurs de variants rares.

^b Paires de gènes comme dans Pg_0 mais séparés d'au moins 2 Mb.

Une étude analogue a été menée sous le modèle cas-témoin pour une population Européenne élargie comprenant les 404 échantillons IBS + TSI + GBR + FIN de 1000G, afin d'avoir environ 200 cas et environ 200 témoins. Aucune inflation due au LD (comme prévu pour un modèle cas-témoin) ou à la stratification de population (comme prévu pour un scénario équilibré) n'a été observée. Néanmoins, l'observation d'une erreur de type I de 0.00128 pour un $\alpha = 0.1\%$ suggère que l'inflation observée après correction du LD pour le test *case-only*, bien que modeste, est aussi retrouvée dans l'analyse cas-témoin (Tableau 11). Toutes les analyses *case-only* et cas-témoin ici présentées ont été par ailleurs conduites pour un $\alpha = 5\%$, reflétant les mêmes tendances observées pour un $\alpha = 0.1\%$ (non montré). Dans le but de mieux comprendre le phénomène observé, nous avons émis l'hypothèse que cette inflation pouvait être due, au moins en partie, à des petits nombres dans les cellules du tableau de contingence donné en **III_B** (Tableau 8). Nous l'avons testé par la répétition de nos analyses, *case-only* et cas-témoin, avec des variants plus fréquents et des taux de porteurs au niveau des gènes plus élevés (c'est-à-dire des variants avec une MAF <10% et des gènes dont 25% des individus étaient porteurs de variants rares d'intérêt, et des variants avec une MAF <15% et des gènes dont 35% des individus étaient porteurs de variants rares d'intérêt; Tableau 11). L'erreur de type I a nettement diminué avec l'augmentation de la fréquence des variants, permettant d'obtenir des erreurs de type I compris dans les intervalles de confiance attendus pour un $\alpha = 0.1\%$.

Tableau 11: Erreurs de type I empiriques à une valeur nominale $\alpha = 0.1\%$ pour les tests case-only et cas-témoins en l'absence de stratification de population

Méthode	Modèle			
	Pg_0^a	Pg_2^b	$Pg_2 + C_{25}^c$	$Pg_2 + C_{35}^d$
<i>Case-only</i> (IBS+TSI)	0.00147 [0.0009-0.00110]	0.00121 [0.0009-0.00110]	0.00109 [0.0009-0.00113]	0.00108 [0.0009-0.00114]
Cas-témoins (IBS+TSI+GBR+FIN)	0.00128 [0.0009-0.00110]	0.00128 [0.0009-0.00110]	0.00107 [0.0009-0.00113]	0.00103 [0.0009-0.00114]

Note: Les limites de l'intervalle de confiance à 95% est montré entre crochets. Les erreurs de type I se situant en dehors de ces limites sont marqués en italique.

^a Toutes les paires de gènes avec des variants avec une MAF<5% et des gènes avec >15% de porteurs de variants rares.

^b Paires de gènes comme dans Pg_0 mais séparés d'au moins 2 Mb.

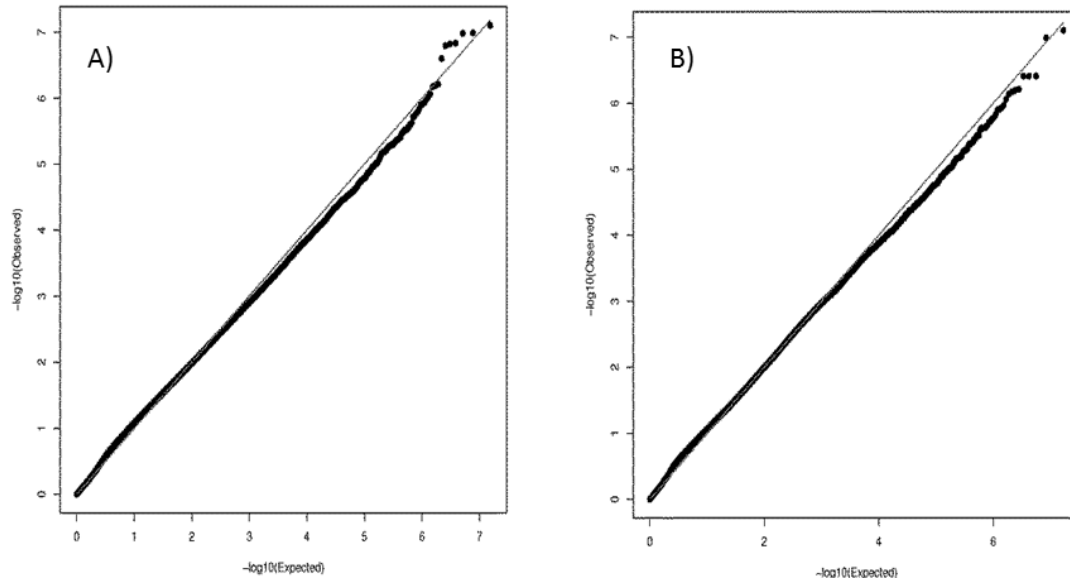
^c Paires de gènes comme dans Pg_2 incluant des variants plus fréquents (MAF<10%) et des gènes avec plus de porteurs (>25%).

^d Paires de gènes comme dans Pg_2 incluant des variants plus fréquents (MAF<15%) et des gènes avec plus de porteurs (>35%).

A part l'impact dû à la taille des échantillons nous avons aussi analysé l'impact que le nombre de tests conduits sur l'approche *case-only* pouvait avoir sur son erreur de type I. Pour cela nous avons élargi le scénario précédent par la considération de seuils de MAF moins rigoureux. En premier lieu, nous avons appliqué le test *case-only* à tous les gènes avec > 5% de porteurs de variants rares plutôt que 15%, parmi les individus de la population IBS + TSI. Cette stratégie a permis d'incorporer 5,563 gènes dans l'analyse, soit 3,975 gènes de plus, ce qui, après élimination des gènes en LD, nous a mené à tester un total de 15,465,141 de paires et à générer le QQ-plot de la Figure 34A. Nous avons observé une erreur de type I légèrement au-dessus du niveau attendu (0.057) pour un $\alpha = 5\%$ et légèrement conservatrice (0.00085) pour un $\alpha = 0.1\%$. Dans cette analyse, impliquant une forte augmentation (un ordre de grandeur de plus) du nombre de tests réalisés, les taux d'inflation dus aux petites tailles d'échantillons (encore plus petites qu'avant) sont plus faibles. Une dernière analyse, considérant pour chaque paire, un gène « rare » (au moins 1% de porteurs, 11,470 gènes au total) et un autre « commun » (au moins 15% de porteurs, soit au total 1,588 gènes) a été conduite de façon systématique sur tout le génome. Dans ce scénario, les 16,951,106 de paires de gènes testées sont représentées sur le QQ-plot de la Figure 34B. Les erreurs de type I obtenues de 0.053 et de 0.00097 sont encore plus proches des valeurs attendues de 5 et 0.1%, respectivement. Ces résultats suggèrent que le test *case-only* est fiable pour analyser des DI

sur une large gamme de fréquences de porteurs au niveau des gènes à condition, bien sûr, que le LD soit pris en compte.

Figure 34: QQ-plot pour le test case-only mené à l'échelle du génome entier sur la population IBS + TSI avec (A) > 5% de porteurs de variants rares pour le premier et le second gène de chaque paire ou bien (B) >1% de porteurs de variants rares pour le premier et >15% pour le deuxième



Les gènes sont écartés d'au moins 2 Mb.

Résultats de simulation avec stratification de population

Nous avons par la suite étudié l'effet de la stratification de population sur le test *case-only* en utilisant les 212 échantillons IBS + CHS. Nous avons observé des erreurs de type I particulièrement élevées (0.0143 pour un $\alpha = 0.1\%$) (Tableau 12). L'ajustement sur les trois premières composantes principales a ramené les erreurs de type I empiriques à des niveaux très proches de ceux observés précédemment (0.0013 pour un $\alpha = 0.1\%$). Pour le test cas-témoin, les 412 échantillons EURAS ont été utilisés dans un scénario dit « non équilibré » (décrit en **Méthode de simulation**), menant à une inflation de l'erreur de type I moins importante que pour le test *case-only* (0.0026 pour un $\alpha = 0.1\%$). L'ajustement sur les composantes principales (0.0013 pour un $\alpha = 0.1\%$) a ramené les erreurs de type I à des valeurs similaires à celles d'une situation sans stratification de population (Tableau 12). Ainsi, à condition que l'espace de recherche soit limité à des paires de gènes suffisamment éloignés pour éviter les effets non désirables du LD et qu'un ajustement soit effectué en cas de structure de population, notre test *case-only* est convenable (en termes d'erreur de type I) pour l'étude de DI, et comparable en termes d'erreur de type I à la méthode cas-témoins correspondante.

Tableau 12: Erreurs de type I empiriques à une valeur nominale $\alpha = 0.1\%$ pour les tests case-only et cas-témoins en présence de PS

Méthode	Ajustement PC	
	No adjustment	3PC
Case-only ^a (IBS+CHS)	<i>0.01432</i> [0.0009-0.00113]	<i>0.00135</i> [0.0009-0.00113]
Cas-témoin équilibré (IBS+TSI+CHS+CDX)	<i>0.00132</i> [0.0009-0.00113]	<i>0.00136</i> [0.0009-0.00113]
Cas-témoin Non équilibré (IBS+TSI+CHS+CDX)	<i>0.00257</i> [0.0009-0.00113]	<i>0.00126</i> [0.0009-0.00113]

Note: Les limites de l'intervalle de confiance à 95% est montré entre crochets. Les erreurs de type I se situant en dehors de ces limites sont marqués en italique.

^a En utilisant des paires de gènes séparées d'au moins 2 Mb.

E.2. Etude de puissance.

Méthode de simulation

Pour l'étude de puissance nous avons élargi notre base de données afin d'échantillonner des cas indépendants. Nous avons considéré 1,735 individus Européens en combinant quatre populations Européennes de 1000G (IBS, TSI, GBR, FIN) et les 1,331 individus de la base de données HGID décrits en III_C. Pour une paire de gènes donnés, nous avons attribué des phénotypes de la façon suivante. En premier lieu, la probabilité d'être atteint (pénétrance) a été calculée à l'aide du modèle de régression logistique décrit en Equation (1). Pour un triplé prédéfini (β_j , β_k , β_l), prenant compte des effets marginaux et d'interaction, le paramètre de «pénétrance de base» β_0 a été calculé pour assurer 15% de cas (par rapport au total d'individus). Des exemples de pénétrances pour des schémas décrits ultérieurement sont présentés dans le Tableau 13. En deuxième instance, comme les individus ont été classés en fonction de leurs génotypes au niveau du premier et du deuxième gène (quatre groupes), un phénotype a été attribué au hasard en fonction de la pénétrance associée au groupe génotypique d'appartenance de l'individu, menant à une moyenne de 260 cas par réplicat. Les tests *case-only* et cas-témoins ont été conduits en simultanément sur les mêmes cas, de sorte que, pour chaque réplicat, le même nombre de témoins que de cas simulés a été échantillonné de manière aléatoire à partir du groupe témoin. Les effets génétiques simulés

ont été appliqués sur l'ensemble des variants rares des gènes impliqués, c'est-à-dire, sur le vecteur de porteurs G_j .

Tableau 13: Valeurs de pénétrance des combinaison digéniques G1 vs G2 lors de l'analyse de deux paires de gènes spécifiques, l'une avec des proportions de porteurs de variants rares similaires pour les deux gènes qui la composent (schéma 2G) et l'autre avec des proportions très différentes (schéma 2GR)

		Pénétrances							
Schéma	2G				2GR				
Porteurs	$\overline{G_1} \times \overline{G_2}^c$	G_1	G_2	$G_1 \times G_2$	$\overline{G_1} \times \overline{G_2}$	G_1	G_2	$G_1 \times G_2$	
$OR_1 = 1; OR_2^a = 1$									
$OR_I^b = 2$	0.14	0.14	0.14	0.25	0.15	0.15	0.15	0.26	
$OR_I = 3$	0.13	0.13	0.13	0.32	0.15	0.15	0.15	0.34	
$OR_I = 5$	0.12	0.12	0.12	0.41	0.15	0.15	0.15	0.46	
$OR_I = 10$	-	-	-	-	0.14	0.14	0.14	0.63	
$OR_1 = 1; OR_2 = 2$									
$OR_I = 2$	0.11	0.11	0.20	0.33	0.14	0.14	0.25	0.40	
$OR_I = 3$	0.10	0.10	0.19	0.41	0.14	0.14	0.25	0.50	
$OR_I = 5$	0.09	0.09	0.17	0.51	0.14	0.14	0.25	0.62	
$OR_I = 10$	-	-	-	-	0.14	0.14	0.24	0.76	

Note: Sans effet marginal du gene 1 ($OR_1 = 1$).

^a OR_2 est le OR marginal du gène 2 de chaque paire.

^b OR_I est le OR du terme d'interaction de Eq. (1).

^c Individus non porteurs de variants rares ni dans le gène 1 ni dans le gène 2.

Les gènes du schéma 2G ont des proportions de porteurs de variants rares de 26% et de 33% dans la population générale, respectivement, tandis que les gènes du schéma 2GR ont des proportions de porteurs de variants rares de 26% et de 5%, respectivement.

Résultats de puissance « moyenne »

Nous avons d'abord estimé une puissance « moyenne » par l'analyse de toutes les paires de gènes possibles (schéma A, Tableau 14), en nous focalisant toujours sur les gènes pour lesquels au moins 15% des individus de l'échantillon de 1,735 individus sont porteurs de variants rares. Seuls les gènes séparés d'au moins 2 Mb ont été pris en compte, ce qui a permis d'analyser un total de 253 gènes (37,053 combinaisons). Chacune de ces 37,053 combinaisons a été successivement considérée comme la paire sous-jacente à l'hérédité

digénique. Par la réplication de chaque simulation dix fois nous avons conduits un nombre total de 370,530 tests. De plus, plusieurs modèles ont été considérés en fonction du niveau de l'effet d'interaction (OR de 1 à 5) et de la présence ou de l'absence d'effets marginaux de l'un ou des deux gènes (Tableau 14). La Figure 35 donne les courbes de puissances obtenues quand un seul ou aucun effet génétique marginal n'est considéré, correspondant aux situations les plus pertinentes pour la recherche d'une interaction gène x gène. Nous observons que les courbes ajustées et non ajustées sont superposées, indiquant que cette analyse, au sein d'une population Européenne homogène, n'a pas été affectée par la structure de population. Dans tous les cas analysés, la puissance du test *case-only* a été supérieure à celle du test cas-témoin. Par exemple, une puissance de 65% pour un $\alpha = 0.1\%$ a été obtenue pour $OR_I = 5$ et aucun effet marginal, comparé à une puissance de 40% pour le test cas-témoin correspondant dans les mêmes conditions. Des tendances similaires ont été observées lorsqu'un effet marginal a été considéré (Figure 35) ou lorsque la puissance a été évaluée pour une valeur nominale $\alpha = 5\%$ (non montré).

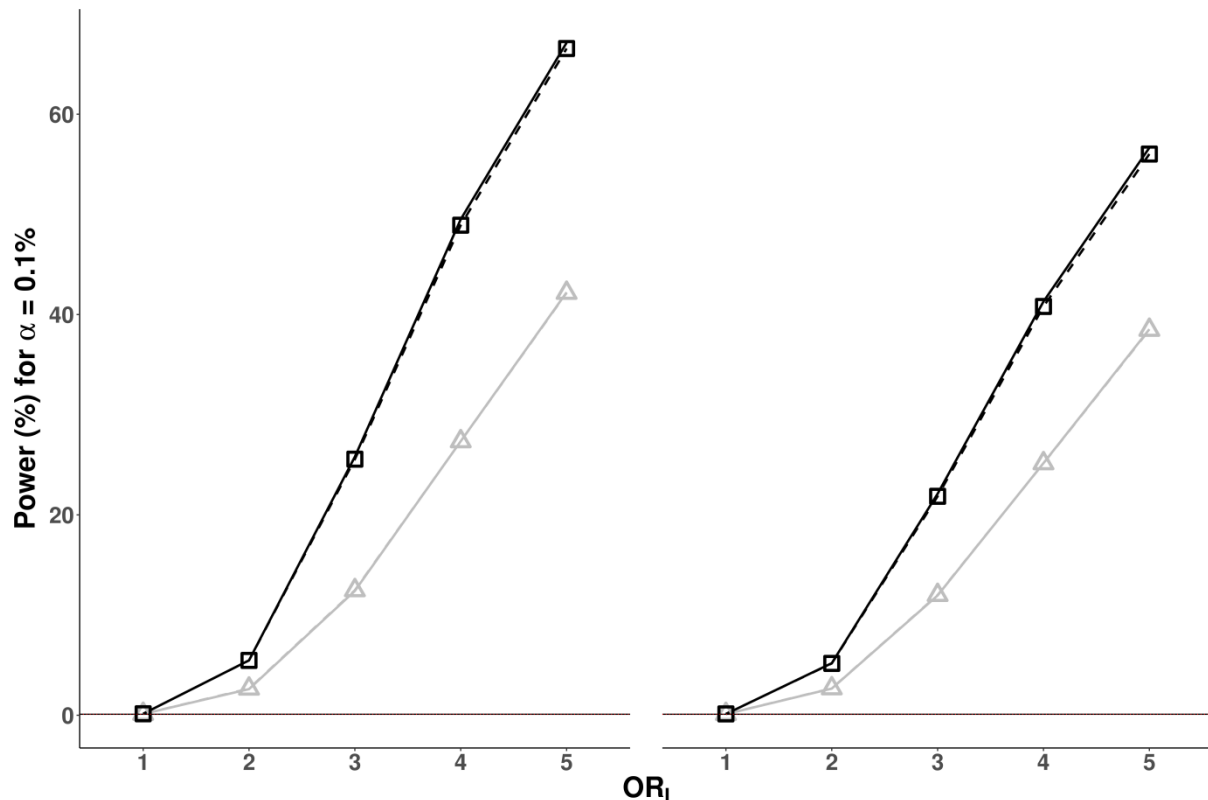
Tableau 14: Description des schémas utilisés dans l'étude de simulation de la section « Puissance »

	Schémas			
	<i>A</i>	<i>2G</i>	<i>2GS</i>	<i>2GR</i>
Gènes testés	Génome entier		2 gènes	
Caractéristiques des gènes	Tous	Les deux communs et non stratifiés	Les deux communs et stratifiés	Un commun et l'autre rare et non stratifiés
OR_j^a	{1,2}	{1,2}	{1,2}	{1,2}
OR_k^a	{1,2}	{1,2}	{1,2}	{1,2}
OR_I^b	{1, ..., 5}	{1, ..., 5}	{1, ..., 5}	{1, ..., 10}

^a OR_j et OR_k sont les ORs des effets marginaux du premier et du deuxième gene de chaque paire, respectivement.

^b OR_I est le OR du terme d'interaction de l'Eq. (1).

Figure 35: Puissance des tests case-only et cas-témoins pour l'analyse de toutes les paires de gènes (schéma A)



Les valeurs de puissance sont présentées en % pour une erreur de type I de 0.1%, en fonction de l'OR de l'interaction (OR_i), pour les tests *case-only* (courbes sombres) et cas-témoins (courbes claires) avec (traits coupés avec symboles) ou sans (traits pleins sans symboles) ajustement pour les trois premières composantes principales. Le panneau de gauche est obtenu quand aucun effet marginal n'est présent, alors que le panneau de droite montre les résultats quand un effet marginal au niveau du deuxième gène ($OR = 2$) est appliqué. Notez que les résultats avec et sans ajustement sont très similaires de sorte que les courbes sont fortement superposées.

Résultats de puissance pour une paire de gènes donnée

Nous nous sommes ensuite focalisés sur deux paires de gènes spécifiques (*AHNAK*, *PKHD1L1*, schéma 2G, voir Tableau 14) et (*ARPP21*, *MACF1*, schéma 2GS, voir Tableau 15) composées de gènes pour lesquels 26%, 33%, 17% et 36%, des individus sont porteurs de variants rares d'intérêt, respectivement (Tableau 14). Ces gènes ont été choisis sur la base de leur niveau de stratification dans les trois sous-populations Européennes définies en III_C (Figure 32). Comme le montre le Tableau 15, les gènes composant la première paire (*AHNAK*, *PKHD1L1*) avaient des fréquences similaires d'une population à l'autre, ceci n'étant pas le cas pour les gènes de la seconde paire (*ARPP21*, *MACF1*). Une précision plus fine a été obtenue à l'aide d'un test d'homogénéité de fréquence pour chaque gène j (Tableau 15). Enfin, nous avons considéré un cinquième gène avec une fréquence cumulée de variants rares plus faible (5%) (Tableau 15) pour tester une interaction entre un gène « commun » et un gène plus « rare ». *AHNAK* et *MPC1* (schéma 2GR, voir Tableau 14) ont été respectivement choisis pour cette tâche. Dans ce dernier schéma, les tests ont été effectués suivant les paramètres décrits précédemment, mais avec une gamme plus large d'OR d'interaction (de 1 à 10 plutôt

que de 1 à 5, Tableau 14). Pour chacune de ces paires nous avons effectué 10,000 répliques des analyses associées.

Tableau 15: Proportion (en %) de porteurs de variants rares pour les gènes testés dans la section III_E.2, pour la population et les sous-populations Européennes, telles que définies en III_C

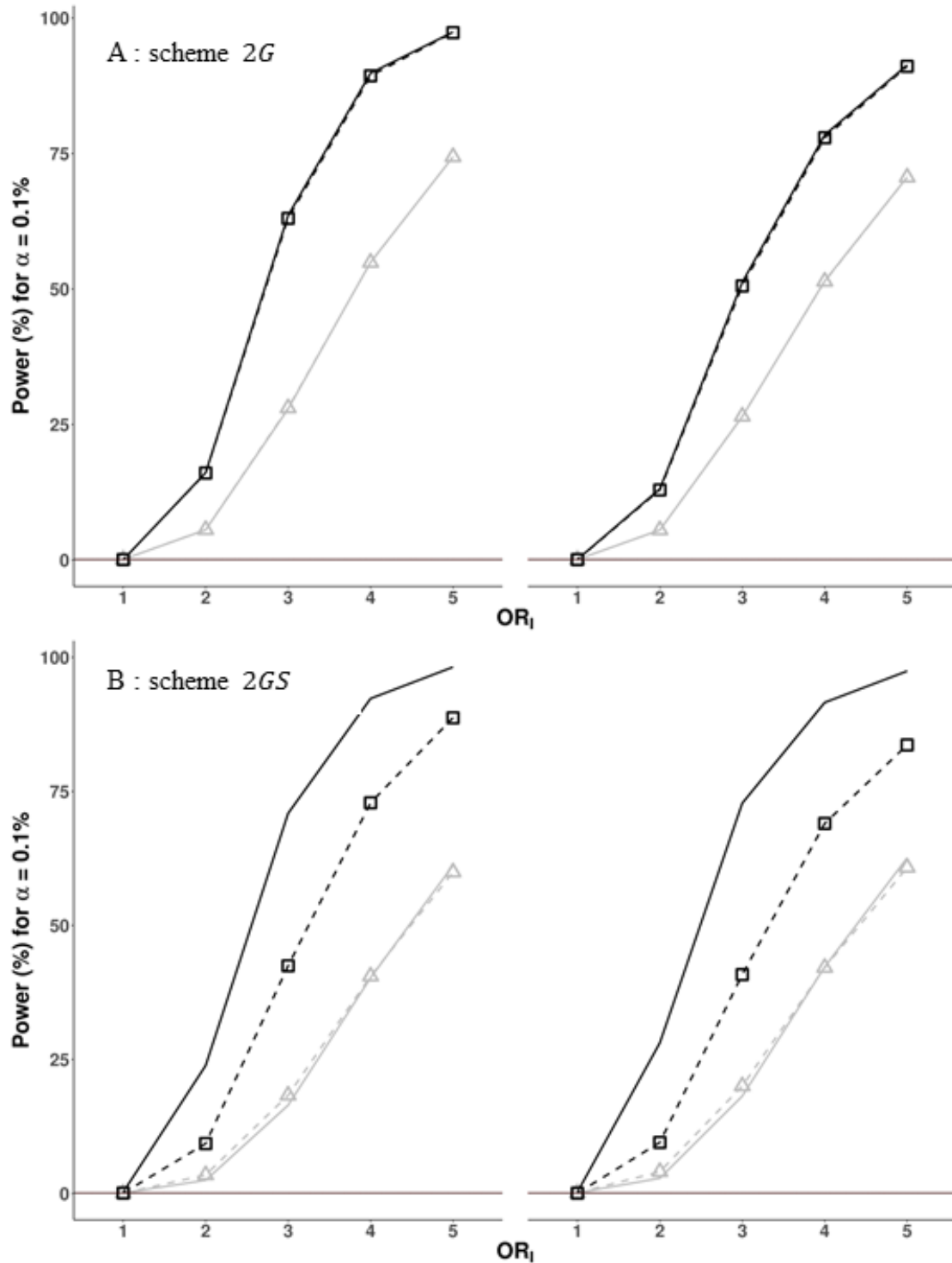
P-value ^a	Gene	Carriers of European populations (%)			
		Global	Southern	Middle	Northern
0.9551	<i>PKHD1L1</i>	33	32	34	29
0.8937	<i>AHNAK</i>	26	26	27	23
2.05E-05	<i>MACF1</i>	36	33	37	51
3.25E-32	<i>ARPP21</i>	17	12	16	45
0.0049	<i>MPC1</i>	5	6	5	1

^a Test d'homogénéité des fréquence cumulées de porteurs pour les trois sous-populations Européennes.

Dans l'analyse du schéma 2G, le test *case-only* s'est montré globalement plus performant en termes de puissance (Figure 36, figure du haut). Par exemple, en absence d'effets marginaux, pour $OR_1 = 3$ et un $\alpha = 0.1\%$, une valeur de puissance de 62% a été obtenue pour le test *case-only*, contre 27% seulement pour le test cas-témoins. De nouveau, les courbes ajustées et non ajustées du test *case-only* étaient confondues confirmant l'absence de stratification de population de notre échantillon. Par contre, pour le schéma 2GS, les courbes de puissance pour les tests *case-only* ajustés et non ajustés étaient bien distinctes, indiquant l'effet de la structure de population (Figure 36, figure du bas). Le test *case-only* ajusté a été donc le seul utilisé pour la comparaison. Comme attendu, le test cas-témoin n'a pas été affecté par la stratification de population (0.0009 pour un $\alpha = 0.1\%$) et a présenté des valeurs d'erreur de type I similaires à celles du test *case-only* ajusté (0.0011 pour un $\alpha = 0.1\%$). Après ajustement, la puissance observée du test *case-only* a été par exemple de 90% pour un $OR_1 = 5$, dans un contexte dépourvu d'effet marginal, alors que la puissance correspondante pour le test cas-témoin n'a été que de 60%. Pour le dernier scénario correspondant au test digénique d'un gène « commun » (26% de porteurs) et d'un gène « rare » (5% de porteurs) (schéma 2GR, voir le Tableau 14), le test *case-only* a été encore plus puissant que le test cas-témoins (Figure 37), en particulier en absence d'effets marginaux,

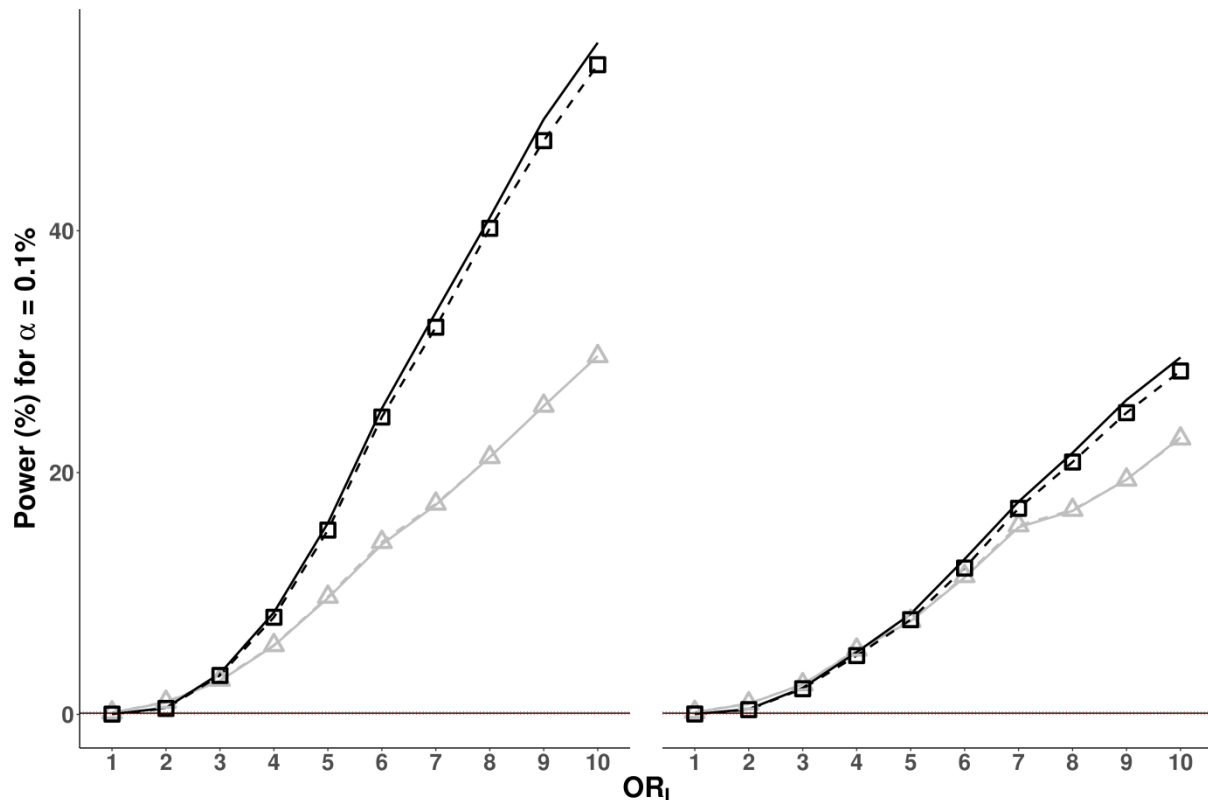
menant à une différence de puissance absolue de près de 30% lorsque $OR_I = 10$. Comme attendu, les courbes ajustées et non ajustées étaient presque confondues. En présence d'un effet marginal sur le gène rare, les différences de puissance entre les deux tests se sont montrées moins fortes, surtout pour les plus petits effets d'interaction ($OR_I < 7$). Cette situation avec une faible fréquence cumulée de variants rares et un OR fort convient mieux à une hypothèse de maladie rare avec une composante génétique forte. Elle est particulièrement intéressante pour l'application de cette approche aux données réelles présentées ci-dessous.

Figure 36: Puissance des tests case-only et cas-témoins pour l'analyse de deux paires de gènes spécifiques en absence (schéma 2G) ou en présence (schéma 2GS) de stratification de la population



Les valeurs de puissance sont présentées comme dans la Figure ?. Les résultats sont présentés pour l'analyse de A) les deux gènes non stratifiés PKHD1L1 et AHNAK (schéma 2G, figure du haut) et B) les deux gènes stratifiés ARPP21 et MACF1 (schéma 2GS, figure du bas). Le panneau de gauche est obtenu lorsqu'aucun effet marginal n'est présent, tandis que le panneau de droite montre des résultats avec un effet marginal ($OR = 2$) du second gène, à savoir respectivement AHNAK et MACF1.

Figure 37: Puissance des tests case-only et cas-témoins pour l'analyse d'une paire de gènes avec différentes proportions de porteurs de variants rares (schéma 2GR)



Les courbes de puissance sont présentées comme dans la Figure ?. Les résultats sont présentés pour l'analyse d'un gène « commun » (AHNAK) et d'un gène « rare » (MPC1) (schéma 2GR). Le panneau de gauche est obtenu quand aucun effet marginal n'est présent, alors que le panneau de droite montre les résultats avec un effet marginal (OR = 2) du deuxième gène, c'est-à-dire MPC1.

E.3. Analyse de données réelles

E.3.1. Craniosynostose.

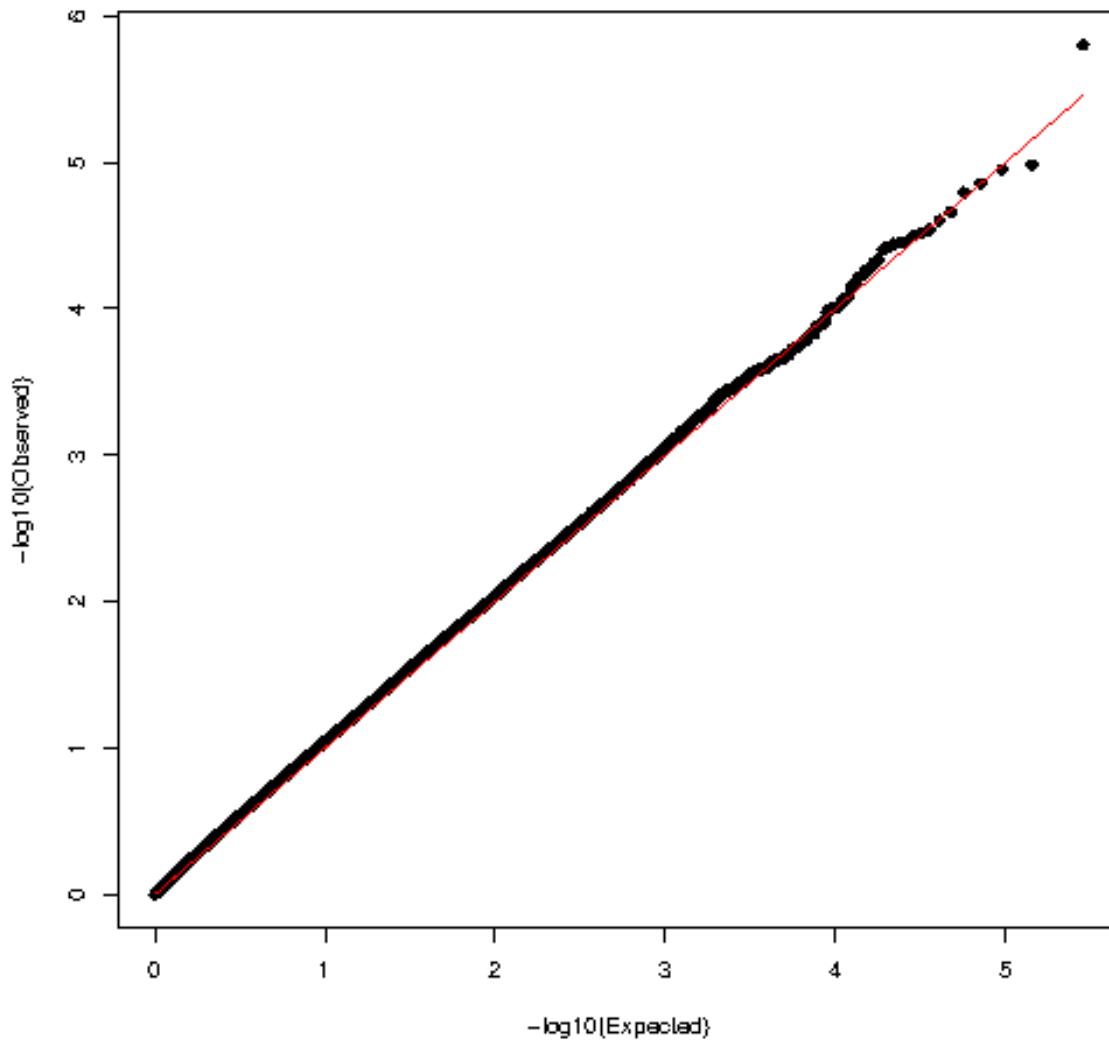
Nous avons utilisé les données WES des 191 proposants de l'étude réalisée par Timberlake et ses collègues (269) sur la craniosynostose, introduite en **III_A**, dans le cadre d'une analyse digénique. Dans l'étude originale conduite par Timberlake et ses collègues (269), un fort enrichissement de mutations hétérozygotes rares *SMAD6* avait été observé parmi les 191 cas (13 porteurs). Cependant, le séquençage des mutations *SMAD6* des membres des familles des 13 porteurs avait mis en évidence une forte pénétrance incomplète (>60%). Le rôle d'un variant commun *rs1884302* (MAF = 0.33 dans les populations Européennes de gnomAD), situé près du gène *BMP2*, un gène lié à *SMAD6*, et précédemment associé à la craniosynostose via GWAS (271), avait donc été étudié dans le but d'expliquer le phénomène de pénétrance incomplète observé. En effet, celui-ci avait résolu la quasi-totalité de la variation phénotypique présente dans la cohorte. Tous les doubles porteurs (onze des 13 proposants *SMAD6*, également porteurs de *rs1884302*) étaient atteints, tandis que les porteurs sains *SMAD6* ne l'étaient pas. Nous avons utilisé ces données pour évaluer la puissance de

notre test *case-only* et sa capacité à détecter cette association dans le contexte d'une recherche à l'échelle du génome entier en données WES (c'est-à-dire sans connaissances préalables du rôle de variants ou gènes dans la maladie étudiée). L'approche a été conduite sous l'hypothèse que des variants rares situés à un premier locus interagissaient avec un variant commun situé à un deuxième locus.

Pour cette approche nous n'avons inclus que les variants faux-sens ou prédits LOF. Seuls des variants de très bonne qualité ont été retenus, en appliquant les mêmes filtres qui ont été décrits en **II_A.1.1** pour l'étude de la cohorte HGID. Au niveau du premier locus nous avons agrégé des variants très rares (MAF gnomAD $<10^{-4}$) que nous avons testé contre tous les variants communs (MAF gnomAD >0.1) au sein d'un deuxième locus, en rajoutant en plus, le variant commun situé près de *BMP2* (*rs1884302*), qui n'est pas capturé par les données WES. Pour le deuxième locus, nous avons utilisé des fenêtres glissantes de 100 Kb pour ne retenir que des variants en faible LD ($r^2 < 0.6$). Seuls les gènes avec $>5\%$ de porteurs de variants rares au premier locus ont été retenus, afin d'assurer une puissance statistique minimale, et seuls des paires de gènes et variants séparés d'au moins 2 Mb ont été retenus. En accord avec la découverte du papier original, un modèle de transmission génétique dominant pour les deux loci a été testé, tandis que les *p values* ont été ajustées sur les trois premières composantes principales.

Au total, 285,216 tests (83 gènes et 8,102 variants) ont été effectués sur les données WES des 191 proposant suite à l'application des filtres décrits. Dans le *QQ-plot* représentant les *p values* de tous ces tests, une seule paire se différencie clairement de la distribution attendue (Figure 38). Ce dernier ($p = 1.58 \times 10^{-6}$, OR = 30.95) correspond à la combinaison digénique *SMAD6-rs1884302* et se situe un ordre de grandeur au-dessus du deuxième résultat ($p = 1.04 \times 10^{-5}$), qui est bien plus proche de la distribution attendue des *p values*. Le tableau de contingence 2 x 2 pour la combinaison digénique *SMAD6-rs1884302* est présenté dans le Tableau 16 et correspond bien sûr à la distribution des porteurs dans le papier original (269). Cette analyse est la preuve de concept que notre test statistique permet de détecter une DI sans recourir à des hypothèses biologiques concernant la maladie étudiée, même lorsque la maladie est très rare.

Figure 38: QQ-plot pour le test de case-only à l'échelle du génome mené sur les 191 patients atteints de craniosynostose



QQ-plot pour une analyse à l'échelle du génome entier sous un mode de transmission dominant, ajusté sur les trois premières composantes principales, en ne considérant que des paires séparés par au moins 2 Mb avec > 5% de porteurs de variants rares pour le premier locus et une fréquence globale > 10% pour le variant situé au deuxième locus ($n = 285,216$ paires).

Tableau 16: Tableau de contingence pour la combinaison digénique *rs1884302* (*BMP2*) – *SMAD6* basée sur les 191 proposants de l'étude sur la craniosynostose

		<i>SMAD6</i>		
		Carriers	Non carriers	Total
<i>BMP2</i> <i>rs1884302</i>	Carriers	11	38	49
	Non carriers	2	140	142
Total		13	178	191

E.3.2. MSMD.

Nous avons de plus appliqué le test proposé à la sous-cohorte de patients MSMD de notre cohorte HGID. Comme ça a déjà été mentionné, certains défauts génétiques du syndrome, tels que ceux associés au déficit autosomique récessif de la sous-unité $\beta 1$ des récepteurs de l'interleukine 12 et -23 (IL-12R $\beta 1$), présentent une pénétrance incomplète et, plus particulièrement, un large éventail de manifestations cliniques. Cependant, aucune étude menée sur le syndrome n'a recherché l'existence de GMs pour expliquer la variabilité clinique du déficit en IL-12R $\beta 1$ (IL12RB1 -/-), qui est la cause génétique la plus commune du MSMD (environ 35% des patients dans le monde).

Nous avons spécifiquement émis l'hypothèse qu'un ou plusieurs GMs pourraient être à l'origine de l'expression phénotypique variable observée chez les patients IL12RB1 -/-. Pour tester cette hypothèse, nous avons tout d'abord combiné les 49 patients MSMD IL12RB1 -/- non apparentés de la cohorte HGID aux 479 patients MSMD sans étiologie génétique connue de la même cohorte, en ne retenant, encore une fois, que des variants de très bonne qualité par l'application des filtres décrits en II_A.1.1. Sur ce groupe de patients nous avons conduit notre test digénique *case-only* sous l'hypothèse que des variants relativement communs pourraient interagir avec le groupe de variants pathogéniques des 49 patients IL12RB1 -/-. Nous avons donc considéré toujours le gène *IL12RB1* au premier locus, suivant un mode de transmission récessif, en accord avec la nature de la déficience. Comme les allèles pathogéniques du déficit en IL-12R $\beta 1$ sont extrêmement rares au niveau populationnel, les potentiels GMs pourraient être très fréquents dans la population générale, comme dans le cas de la craniosynostose. Par conséquent, nous avons considéré, pour le deuxième locus à l'étude, tous les variants codants faux-sens ou prédits LOF avec une MAF gnomAD globale >1%. Pour ne retenir que des variants indépendants à ce deuxième locus nous avons appliqué la même technique utilisée dans l'étude de la craniosynostose pour éliminer les variants en fort LD. Finalement, nous avons considéré un mode de transmission génétique dominant pour le potentiel GM.

Un total de 10,691 variants avec une MAF >1% ont été testés en association avec *IL12RB1*. Aucune *p value* n'a atteint le seuil de significativité donné par la méthode de correction de Bonferroni de 4.68×10^{-6} . Cependant, le meilleur résultat ($p = 1.22 \times 10^{-5}$; *p* corrigée = 0.13) a été observé pour un variant faux-sens du gène *MST1R*. Ce variant, Q613P, a une fréquence globale de 1.04% et n'est observé que parmi les patients IL12RB1 -/- (5 sur 49, 10% des porteurs). Aucun porteur de Q613P n'est présent chez les 479 patients restants (Tableau 17). Il est à noter que les 6 personnes asymptomatiques IL12RB1 -/- qui font partie de la cohorte HGID, mais qui sont absentes de cette analyse, focalisée uniquement sur les

patients symptomatiques, ne sont pas porteuses du variant *MST1R*. Cette observation est compatible avec le rôle présumé du variant pour le syndrome MSMD bien qu'aucune analyse statistique formelle incluant ces individus asymptomatiques ne puisse être effectuée, leur nombre étant trop petit.

MST1R, également appelé « Macrophage Stimulating 1 Receptor » ou « RON », code pour un récepteur de surface cellulaire de protéines stimulées par des macrophages avec activité de la tyrosine kinase et joue un rôle dans la réponse immunitaire innée de l'humain en régulant la migration et l'activité phagocytaire des macrophages (285,286). Trois variants faux-sens rares de ce gène ($0 < \text{MAFs} < 9 \times 10^{-4}$) ont été découverts chez quatre patients de trois familles distinctes atteints d'un syndrome connu sous le nom de Lady Windermere (LW), qui est une maladie non-pulmonaire causée par des infections par mycobactéries non tuberculeuses (287). Il convient de noter que les quatre porteurs des variants rares *MST1R* de l'étude susmentionnée (287) étaient hétérozygotes et qu'un grand nombre important de membres de la famille étaient des porteurs asymptomatiques, suggérant que si les variants *MST1R* prédisposaient en effet au syndrome LW, ils présenteraient une pénétrance incomplète.

Tableau 17: Tableau de contingence pour la combinaison digénique Q613P (*MST1R*) – *IL12RB1*

		<i>IL12RB1</i>		
		Carriers	Non carriers	Total
Q613P (<i>MST1R</i>)	Carriers	5	0	5
	Non carriers	44	479	523
Total		49	479	528

F. Conclusions et perspectives de l'approche digénique.

L'architecture génétique de nombreuses affections, en particulier, des pathologies supposées monogéniques, est aujourd'hui remise en cause par la présence d'éventuelles contributions génétiques supplémentaires, en particulier de DI. Dans ce contexte, le test *case-only* proposée est une approche appropriée pour le test d'interactions, permettant d'identifier des lésions digéniques à l'échelle du génome entier. Il convient en particulier à l'analyse de variants rares, ne nécessite pas d'un sous-ensemble de gènes candidats et présente plusieurs avantages par rapport à une approche classique cas-témoins. Les études de simulation présentées dans ce travail, basées sur des données réelles WES, ont montré que le test *case-only* était plus puissant que son analogue cas-témoin dans tous les scénarios testés. Les tests

ne sont comparables en termes de puissance qu'en présence d'effets marginaux; en d'autres termes, lorsque nous avons simulé un impact marginal sur la maladie pour l'un des deux gènes d'une paire, indépendamment de l'autre gène, et ceci que dans le cas d'une maladie rare avec une composante génétique forte. Néanmoins, la différence de puissance entre les deux tests est fortement corrélée à l'effet d'interaction génétique (Figure 38), étant plus importante pour des effets forts.

Le test *case-only* est conçu pour capturer l'effet d'une association entre des variants de deux gènes. En d'autres termes, il est sensible à la non indépendance de variants situés sur deux loci différents. Il peut donc détecter des interactions de type digénique, comme nous l'avons montré, mais également du LD. Le problème du LD peut être résolu en éliminant les paires de gènes situés à forte proximité les uns des autres (<2 Mb). L'investigation ponctuelle des paires de gènes avec un $\delta < 2$ Mb et statistiquement significatives (680 avec $p < 0.05$ sur 4,082 paires dans la cohorte IBS + TSI) nécessiterait par exemple d'une approche cas-témoin. La comparaison des éventuelles corrélations entre les variants de deux gènes au sein des cas et des témoins permettrait de distinguer entre une véritable interaction et du LD, le LD étant aussi observable chez les témoins. Il est à noter que même un tout petit nombre de témoins pourrait aider à révéler la vraie nature de ces signaux statistiques au moyen d'une approche analogue n'utilisant que des témoins (*control-only*), qui ne détecterait que du LD.

Notre test *case-only* est donc pertinent dans le contexte d'une pénétrance incomplète. En particulier, il peut directement évaluer le rôle de variants communs en association à des variants rares situés à un autre locus. Ce cas ponctuel du modèle digénique, dans lequel l'effet d'un seul variant est testé pour le deuxième locus, a été utilisé dans notre étude de DI dans la craniosynostose, impliquant des mutations rares *SMAD6* et un variant commun situé près de *BMP2*. Il a été aussi utilisé dans notre analyse des patients MSMD *IL12RB1* -/- de la cohorte HGID. Une observation importante découle par ailleurs de cette dernière analyse. Alors que la fréquence des porteurs que nous avons considérés (allant de 1 à 15%) dans nos analyses de simulations peut sembler trop élevée lors de l'étude d'une pathologie rare, deux points importants doivent être soulignés. Premièrement, ces seuils correspondent à la fréquence cumulée des variants contribuant à la maladie, qui peuvent individuellement être beaucoup plus rares. Le second point est que les allèles pathogènes devraient être fortement enrichis dans un échantillon composé exclusivement de patients. Par exemple, dans notre échantillon total de 633 patients MSMD avec et sans étiologie génétique décrites, les fréquences de certains allèles LOF sur *IL12RB1*, tels que c.1791 + 2T > G et c.783 + 1G > A, atteignent 1.1%

et 1.4%, tandis que dans des grandes bases de données publiques, comme gnomAD, elles sont respectivement à 9.57×10^{-5} et 0.

Ce test est basée sur l'hypothèse d'homogénéité génétique, c'est-à-dire, sur l'hypothèse qu'un nombre considérable de variants ont un effet similaire (voire même identique) sur la maladie (par exemple, effet de prédisposition). C'est une hypothèse raisonnable pour des maladies rares comme la craniosynostose, mais il sera intéressant d'étendre le test pour qu'il puisse capturer l'effet hétérogène de variants à la fois protecteurs et délétères, comme dans le cas par exemple du test d'association SKAT (288). D'autres études seront nécessaires pour évaluer les propriétés de cette extension, qui devra par ailleurs faciliter l'intégration de variants et phénotypes plus communs dans l'analyse. Le test *case-only* proposé semble déjà être un outil approprié et puissant pour l'étude de DI basé sur des données NGS dans le contexte de pathologies rares qui ne sont pas entièrement expliquées par une lésion monogénique.

Discussion générale sur le travail de thèse.

Le travail ici présenté a été principalement motivé par la recherche de nouvelles étiologies génétiques de la tuberculose. Cette maladie infectieuse, considérée la plus létale de l'histoire humaine, tue encore chaque année plus d'un million d'individus, en particulier, des milliers d'enfants, avec une prédisposition génétique à la tuberculose. Cependant, pour la vaste majorité d'entre eux, les causes génétiques sous-jacentes à leur prédisposition est méconnue. Ici, nous rapportons des résultats qui contribuent à cette recherche, en nous basant en partie sur la découverte des premières formes monogéniques de tuberculose au long des deux dernières décennies. Nous nous sommes servis de la cohorte d'exomes HGID pour conduire nos premières analyses, qui ont abouti à la description de la première forme monogénique relativement commune de la tuberculose, au niveau du gène *TYK2*. Cette découverte renforce le lien connu entre la physiopathologie de la tuberculose et du syndrome MSMD du fait, qu'en particulier, les patients déficients en *TYK2* sont atteints de diverses infections par des bactéries intracellulaires. De façon plus générale, les gènes décrits dans les formes monogéniques de tuberculose sont aussi responsables des causes génétiques du MSMD (152).

Ces gènes sont impliqués dans la voie immunologique non-redondante de la défense anti-mycobactérienne, dépendante de l' $\text{IFN-}\gamma$. Nous nous sommes donc intéressés à élargir les connaissances sur cette voie par l'analyse du cas spécifique d'une famille consanguine d'origine Libanaise pouvant héberger une nouvelle étiologie génétique du MSMD. L'étude, combinant liaison génétique et analyse d'exome, permet la description de la première déficience autosomique récessive de l' $\text{IFN-}\gamma$ au sein du MSMD. La prédisposition au syndrome MSMD des individus déficients au niveau du gène était attendue. Dans ce contexte, l'analyse de sélection évolutive, conduite pour le gène *IFNG*, fut certainement plus intéressante. Elle servit à comprendre pourquoi cette déficience n'a été que récemment observée dans le MSMD alors que des centaines de cas ont été décrits au niveau des récepteurs de l'*IFNG*. Nous avons observé l'absence quasi-totale de mutations potentiellement délétères pour l'*IFNG*, surtout par rapport à ses deux récepteurs, au sein des bases de données publiques gnomAD et BRAVO. L'analyse spécifique du taux de mutation global au niveau du locus *IFNG* a montré que ce taux était comparable à celui de ces deux récepteurs. Nous avons donc émis l'hypothèse que ce locus aurait été soumis à une forte pression de sélection négative, que nous avons confirmée à partir de l'analyse de deux critères de sélection différents. En outre, ces analyses soutiennent l'hypothèse que des mutations

hétérozygotes sur l'*IFNG* pourraient avoir des conséquences importantes sur la santé des porteurs, les rendant par exemple susceptibles aux infections par mycobactéries, particulièrement à celles plus virulentes, telle la tuberculose.

Mon travail de thèse s'est poursuivi par l'analyse spécifique des données WES de tuberculose de notre base de données interne HGID, suivant des approches variant-candidat. Nous avons cherché à mieux comprendre la physiopathologie d'une maladie infectieuse ancienne, intrinsèquement relié à l'humain par ses probables plus de 70,000 années d'échanges hôte/pathogène. Dévastatrice en Europe jusque dans la première moitié du XX^e siècle, la tuberculose semblait avoir « disparue » du vieux continent. Elle se voit néanmoins revenir depuis quelques décennies en Europe, accompagnant en partie l'épidémie du VIH. Ailleurs dans le monde, la tuberculose est encore plus létale, avec plusieurs pays endémiques pour la maladie. L'étude de la cohorte HGID a permis de mettre en évidence le rôle pathogène d'un variant qui est le plus fréquent en Europe (> 4%) et qui est présent (> 1%) dans d'autres régions du monde où la tuberculose est endémique (Inde, Afrique du Nord, Amérique latine ...). L'homozygotie pour le variant P1104A a un fort impact sur la susceptibilité aux infections par mycobactéries, ce qui n'est pas le cas pour les porteurs hétérozygotes. Les homozygotes, bien que globalement nombreux, n'ont pas pu être identifiés par GWAS dans des analyses de cohortes de patients tuberculeux de tailles moyennes (<5000 cas). Les explications sous-jacentes potentielles sont soit le manque de puissance statistique (dû à la méthode employée ou à la fréquence de l'allèle dans la cohorte étudiée), soit l'utilisation d'un modèle génétique inapproprié pour la découverte du génotype en question (par exemple, un modèle additif). Nos calculs ont confirmé que l'homozygotie pour le variant P1104A pouvait représenter jusqu'à environ 1% des cas de tuberculose en Europe, et 0.33% des cas dans les pays de provenance des porteurs homozygotes de l'étude de la cohorte HGID. Cette dernière observation pourrait avoir des conséquences majeures dans le traitement et la prévention de ces patients, car ils pourraient par exemple bénéficier d'un traitement avec de l'IFN- γ recombinant, ce qui revêt une importance particulière dans le contexte actuel d'augmentation de souches de *M. tuberculosis* résistantes aux antibiotiques.

La baisse significative de la fréquence du variant TYK2 P1104A depuis 4,000 ans en Europe ainsi que son implication dans la tuberculose maladie suggèrent l'hypothèse que ce locus ait été soumis à une pression de sélection depuis des milliers d'années. Une étude complète de l'histoire géographique et temporelle du variant TYK2 P1104A pourrait révéler des caractéristiques qui lui sont propres, pouvant faciliter la découverte de nouvelles étiologies génétiques de la tuberculose ou d'autres maladies infectieuses anciennes. Dans ce

but, il sera intéressant de dater la mutation, tout en estimant le degré de sélection subit. Dû à l'ancienneté de la tuberculose maladie chez l'humain, il est raisonnable d'estimer que cette pression est restée constante. Elle a été probablement aussi modeste. En effet, le variant n'a pas été complètement purgé de la population humaine. Son caractère pathogène dépend exclusivement d'un facteur externe qui est la présence de la mycobactérie. Plus important, la pression sélective n'est effective que sur les homozygotes P1104A car les autres génotypes du locus ne confèrent pas une prédisposition à la tuberculose. Cette pression dépend en outre de la taille effective des sous-populations qui ont hébergé le variant qui, en raison d'effets migratoires et d'effets *bottlenecks*, a souvent été petite, permettant probablement la dérive du variant et l'augmentation ou le maintien de sa fréquence. L'homozygotie pour P1104A est aussi connu aujourd'hui pour avoir un rôle protecteur dans certaines maladies auto-inflammatoires ou auto-immunes. Cependant, ces maladies sont trop récentes dans l'histoire de l'être humain pour être responsables d'une force de pression de sélection opposée à celle qui a été estimée due à la tuberculose. Les études de larges cohortes WES de patients tuberculeux ainsi que les études de génétique évolutive, utilisant notamment des données d'ADN ancien (humaines ou microbiennes), seront certainement importantes dans l'élargissement des connaissances actuelles des bases génétiques de la tuberculose.

La tuberculose est aujourd'hui une maladie monogénique, dans le sens où les seules étiologies génétiques qui lui sont connues ont été décrites sur un seul gène à effet fort. Cependant, le nombre de maladies supposées monogéniques mais pour lesquelles une étiologie plus complexe, en particulier digénique, a été retrouvée, s'est accru depuis les deux dernières décennies. La découverte de modèles génétiques plus complexes pour ces maladies provient souvent d'observations de pénétrance incomplète au sein des familles étudiées. Dans ce contexte, la pénétrance incomplète a été souvent expliquée par la présence d'autres variants génétiques à un locus différent. L'étude de la tuberculose peut se servir du modèle digénique ou d'un modèle oligogénique pour la découverte de nouvelles étiologies génétiques. En effet, la tuberculose n'est pas une maladie rare. Sa prévalence suggère que des combinaisons de variants plus ou moins fréquents, impliquant notamment les voies immunologiques connues dans la défense anti-mycobactérienne, peuvent être à l'origine d'un nombre non négligeable de cas. Des combinaisons digéniques plus rares peuvent par ailleurs être responsables de cas plus sévères de tuberculose, en particulier de cas pédiatriques. Pour analyser cette dernière hypothèse il sera intéressant d'appliquer notre méthodologie *case-only* à notre sous-cohorte WES de patients atteints de tuberculose comme cela a déjà été fait avec la sous-cohorte MSMD (et pour laquelle d'autres analyses avec d'autres modes de dominance doivent être

encore testées). L'application du test *case-only* à notre cohorte WES de patients tuberculeux serait par ailleurs la première analyse digénique conduite dans la tuberculose maladie à partir de données NGS.

L'analyse de réplication des lésions digéniques observés originellement par Timberlake et ses collègues (269) dans la craniosynostose a permis de valider non seulement notre approche *case-only* pour l'étude de DI, mais aussi la découverte initiale du groupe de Richard Lifton. En effet, l'étude systématique des combinaisons digéniques au niveau de l'exome entier, permet de réduire le biais de l'étude de Timberlake et ses collègues qui n'ont testé qu'un seul variant commun dans l'étude. D'autre part, l'analyse préliminaire réalisée sur les patients *IL12RB1* *-/-* de notre cohorte HGID, mettant en évidence l'existence d'un potentiel variant modificateur au niveau du gène *MSTIR*, suggère le rôle de GMs pour expliquer la pénétrance incomplète. Le recrutement de plus de porteurs asymptomatiques (six ont été recrutés jusqu'ici appartenant à cinq familles différentes) pourrait aider à confirmer le rôle du variant candidat. En effet, si l'analyse *case-only* que nous avons conduite sur les données de craniosynostose avait été faite sans tenir compte des résultats rapportés par Timberlake et ses collègues, elle aurait pu se servir des porteurs asymptomatiques des variants *SMAD6* de la cohorte original pour valider la découverte (aucun n'étant aussi porteur du variant commun proche de *BMP2*). Un autre point pouvant contribuer à la recherche de GMs dans la déficience en IL-12R β 1 serait la définition de sous-groupes phénotypiques au sein des patients *IL12RB1* *-/-* (suivant par exemple un classement défini en fonction de la sévérité de la maladie, ou par rapport au type ou au nombre d'infections). Parmi ces individus, environ une moitié est aussi atteinte d'une salmonellose. Il sera intéressant, comme cela a été fait par Emond et ses collègues pour l'étude d'infections pulmonaires par *Pseudomonas aeruginosa* dans la mucoviscidose (268), d'analyser l'apport potentiel de GMs dans cette situation.

Nous avons mentionné la longue histoire des échanges hôte/pathogène au cours de multiples cycles de sélection et d'évasion entre l'humain et *M. tuberculosis*, menant très probablement à des signatures génétiques de co-évolution. Nous avons aussi mentionné les problèmes sous-jacents aux analyses GWAS focalisées sur un phénotype défini sur la base de la présentation clinique. Dans le contexte d'études d'interactions, des travaux récents focalisés sur des données génétiques appariées à l'échelle du génome entier, provenant d'hôtes et d'agents pathogènes, ont permis de tester l'hypothèse de co-signatures génomiques pour des infections virales. Pour le virus du VIH-1, Bartha et ses collègues ont retrouvé, par exemple, de forts signaux d'association entre des allèles HLA de classe I et des mutations virales dans les épitopes CTL correspondants (289). Une autre étude pour le virus de l'hépatite C a montré

que des allèles HLA de classe I et de classe II ainsi que certains gènes du système immunitaire inné associés à l'interféron lambda étaient associés à des polymorphismes du génome viral (290). Ces analyses qui ont permis de mieux comprendre les facteurs de l'hôte associés à l'adaptation virale ont utilisé les empreintes apparentes laissées par l'humain dans le génome du virus. Cette nouvelle approche dans les études de susceptibilité ou résistance aux infections pourrait avoir des implications tout aussi pertinentes pour l'étude de la tuberculose, notamment en raison de son ancienneté. En effet, la répartition géographique des lignées du complexe MTBC est associée à des populations humaines spécifiques, reflétant la co-migration et la co-expansion probable de l'hôte et de l'agent pathogène (16,291,292). Par ailleurs, des études sur des populations humaines ont avancé des premiers résultats suggérant une interaction entre le génome humain et celui de certaines lignées spécifiques du complexe (279,280). Cependant, les résultats rapportés nécessiteront d'études plus approfondies pour mieux comprendre la fonction des allèles impliqués.

Aujourd'hui, le séquençage de l'exome ou du génome entier d'individus avec ou sans un trait phénotypique particulier (par rapport à la population générale) est devenu une pratique répandue. Les coûts des WES sont en baisse (~ 300 dollars aujourd'hui) et l'intérêt des chercheurs à mieux comprendre les bases génétiques d'un grand nombre de traits ou maladies s'ajoutent à l'attitude bienveillante de certaines personnes à donner leurs ADN pour être étudiés. La base de données gnomAD qui ajoute à la description de plus de 120,000 exomes, distribués au sein de groupes ethniques très larges, un nombre considérable d'individus définis comme « témoins » (appartenant à différentes cohortes d'études) en est un des meilleurs exemples. Cependant, gnomAD ne permet pas l'accès aux données phénotypiques et génotypiques individuelles. Tout récemment, la base de données UK Biobank, qui accumule de nombreuses données biologiques et statistiques en relation à la santé de ses participants, a ouvert l'accès à ses utilisateurs à une sous-cohorte de 50,000 exomes des 500,000 individus génotypés qui la composent, ayant pour but d'atteindre les 500,000 WES dans les années à suivre. Cette base de données, unique dans le monde par sa taille et sa composition, sera importante pour la détection de nouveaux variants pathogéniques (comme le cas du variant TYK2 P1104A), dans un nombre important de maladies (rapportés par la ressource UK Biobank), qui dans l'analyse de cohortes plus petites seraient indétectables ou moins discernables. L'effort de différentes communautés scientifiques, travaillant en collaboration pour la mise en place de cette (en particulier, par l'étude de la qualité des données génétiques (293,294)) ou d'autres cohortes publiques NGS, sera bien sûr

fondamentale pour atteindre une meilleure compréhension de la physiopathologie des plus récentes ou des plus vieilles pathologies humaines.

Bibliographie

1. Anderson RM, May RM, Anderson B. *Infectious Diseases of Humans: Dynamics and Control*. Revised ed. edition. Oxford: Oxford University Press; 1992. 768 p.
2. Cairns J, Singer AL. *Matters Of Life And Death: Perspectives On Public Health, Molecular Biology, Cancer, And The Prospects For The Human Race*. Diane Pub Co; 1997. 257 p.
3. Casanova J-L, Abel L. Inborn errors of immunity to infection: the rule rather than the exception. *Journal of Experimental Medicine*. 18 juill 2005;202(2):197-201.
4. Dubos RJ, Dubos J. *The White Plague: Tuberculosis, Man, and Society*. Rutgers University Press; 1987. 320 p.
5. Glaziou P, Floyd K, Raviglione M. Trends in tuberculosis in the UK. *Thorax*. 2018;73(8):702-3.
6. Nathavitharana RR, Friedland JS. A tale of two global emergencies: tuberculosis control efforts can learn from the Ebola outbreak. *Eur Respir J*. août 2015;46(2):293-6.
7. Koch R. The etiology of tuberculosis. :7.
8. Daniel TM. Rifampin--a major new chemotherapeutic agent for the treatment of tuberculosis. *N Engl J Med*. 13 mars 1969;280(11):615-6.
9. Vall-Spinosa A, Lester W, Moulding T, Davidson PT, McClatchy JK. Rifampin in the treatment of drug-resistant mycobacterium tuberculosis infections. *N Engl J Med*. 17 sept 1970;283(12):616-21.
10. Cegielski JP. Extensively drug-resistant tuberculosis: « there must be some kind of way out of here ». *Clin Infect Dis*. 15 mai 2010;50 Suppl 3:S195-200.
11. Keshavjee S, Farmer PE. Tuberculosis, Drug Resistance, and the History of Modern Medicine. *New England Journal of Medicine*. 6 sept 2012;367(10):931-6.
12. Paulson T. Epidemiology: A mortal foe. *Nature*. 10 oct 2013;502(7470):S2-3.
13. Abel L, El-Baghdadi J, Bousfiha AA, Casanova J-L, Schurr E. Human genetics of tuberculosis: a long and winding road. *Philos Trans R Soc Lond, B, Biol Sci*. 2014;369(1645):20130428.
14. Marais BJ, Gie RP, Schaaf HS, Beyers N, Donald PR, Starke JR. Childhood pulmonary tuberculosis: old wisdom and new challenges. *Am J Respir Crit Care Med*. 15 mai 2006;173(10):1078-90.
15. Dheda K, Barry CE, Maartens G. Tuberculosis. *Lancet*. 19 mars 2016;387(10024):1211-26.
16. Comas I, Coscolla M, Luo T, Borrell S, Holt KE, Kato-Maeda M, et al. Out-of-Africa migration and Neolithic coexpansion of *Mycobacterium tuberculosis* with modern humans. *Nat Genet*. oct 2013;45(10):1176-82.
17. Esteban J, Muñoz-Egea M-C. *Mycobacterium bovis* and Other Uncommon Members of the *Mycobacterium tuberculosis* Complex. *Microbiol Spectr*. 2016;4(6).

18. Wiens KE, Woyczynski LP, Ledesma JR, Ross JM, Zenteno-Cuevas R, Goodridge A, et al. Global variation in bacterial strains that cause tuberculosis disease: a systematic review and meta-analysis. *BMC Med.* 30 oct 2018;16(1):196.
19. Barbier M, Wirth T. The Evolutionary History, Demography, and Spread of the Mycobacterium tuberculosis Complex. *Microbiology Spectrum* [Internet]. 12 août 2016 [cité 27 juin 2019];4(4). Disponible sur: <http://www.asmscience.org/content/journal/microbiolspec/10.1128/microbiolspec.TBTB2-0008-2016>
20. Wolfe ND, Dunavan CP, Diamond J. Origins of major human infectious diseases. *Nature.* 17 mai 2007;447(7142):279-83.
21. Smith NH, Hewinson RG, Kremer K, Brosch R, Gordon SV. Myths and misconceptions: the origin and evolution of Mycobacterium tuberculosis. *Nat Rev Microbiol.* 2009;7(7):537-44.
22. Embley TM, Stackebrandt E. The molecular phylogeny and systematics of the actinomycetes. *Annu Rev Microbiol.* 1994;48:257-89.
23. Rastogi N, Legrand E, Sola C. The mycobacteria: an introduction to nomenclature and pathogenesis. *Rev - Off Int Epizoot.* avr 2001;20(1):21-54.
24. Behr MA, Wilson MA, Gill WP, Salamon H, Schoolnik GK, Rane S, et al. Comparative genomics of BCG vaccines by whole-genome DNA microarray. *Science.* 28 mai 1999;284(5419):1520-3.
25. Falkinham JO. Epidemiology of infection by nontuberculous mycobacteria. *Clin Microbiol Rev.* avr 1996;9(2):177-215.
26. Bercovier H, Vincent V. Mycobacterial infections in domestic and wild animals due to Mycobacterium marinum, M. fortuitum, M. chelonae, M. porcinum, M. farcinogenes, M. smegmatis, M. scrofulaceum, M. xenopi, M. kansasii, M. simiae and M. genavense. *Rev - Off Int Epizoot.* avr 2001;20(1):265-90.
27. Collins CH. The bovine tubercle bacillus. *Br J Biomed Sci.* 2000;57(3):234-40.
28. Bloom BR, Small PM. The evolving relation between humans and Mycobacterium tuberculosis. *N Engl J Med.* 5 mars 1998;338(10):677-8.
29. Jacobson RR, Krahenbuhl JL. Leprosy. *Lancet.* 20 févr 1999;353(9153):655-60.
30. 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.* mars 2017;58(1):E9-12.
31. Cave AJE, Demonstrator A. The evidence for the incidence of tuberculosis in ancient Egypt. *British Journal of Tuberculosis.* 1 juill 1939;33(3):142-52.
32. Daniel TM. The history of tuberculosis. *Respir Med.* nov 2006;100(11):1862-70.
33. texte A nationale de médecine (France) A du. *Bulletin de l'Académie nationale de médecine* [Internet]. Gallica. 1865 [cité 2 juill 2019]. Disponible sur: <https://gallica.bnf.fr/ark:/12148/bpt6k4086577>

34. Paulson T. Epidemiology: A mortal foe. *Nature*. 9 oct 2013;502:S2-3.
35. Maddison A. *The World Economy*. OECD Publishing; 2006. 657 p.
36. Murray JF. The Industrial Revolution and the decline in death rates from tuberculosis. *Int J Tuberc Lung Dis*. mai 2015;19(5):502-3.
37. Ernst JD. The immunological life cycle of tuberculosis. *Nat Rev Immunol*. 13 juill 2012;12(8):581-91.
38. van Crevel R, Ottenhoff THM, van der Meer JWM. Innate immunity to *Mycobacterium tuberculosis*. *Clin Microbiol Rev*. avr 2002;15(2):294-309.
39. Neyrolles O, Gicquel B, Quintana-Murci L. Towards a crucial role for DC-SIGN in tuberculosis and beyond. *Trends Microbiol*. sept 2006;14(9):383-7.
40. Ehlers S. DC-SIGN and mannosylated surface structures of *Mycobacterium tuberculosis*: a deceptive liaison. *Eur J Cell Biol*. janv 2010;89(1):95-101.
41. Jo E-K. Mycobacterial interaction with innate receptors: TLRs, C-type lectins, and NLRs. *Curr Opin Infect Dis*. juin 2008;21(3):279-86.
42. 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.
43. Blomgran R, Desvignes L, Briken V, Ernst JD. *Mycobacterium tuberculosis* inhibits neutrophil apoptosis, leading to delayed activation of naive CD4 T cells. *Cell Host Microbe*. 19 janv 2012;11(1):81-90.
44. Cambier CJ, Falkow S, Ramakrishnan L. Host Evasion and Exploitation Schemes of *Mycobacterium tuberculosis*. *Cell*. 2014;159:1497-509.
45. Korbelt DS, Schneider BE, Schaible UE. Innate immunity in tuberculosis: myths and truth. *Microbes Infect*. juill 2008;10(9):995-1004.
46. Bustamante J, Boisson-Dupuis S, Abel L, Casanova J-L. Mendelian susceptibility to mycobacterial disease: genetic, immunological, and clinical features of inborn errors of IFN- γ immunity. *Semin Immunol*. déc 2014;26(6):454-70.
47. Stewart GR, Robertson BD, Young DB. Tuberculosis: a problem with persistence. *Nat Rev Microbiol*. nov 2003;1(2):97-105.
48. Lu LL, Smith MT, Yu KKQ, Luedemann C, Suscovich TJ, Grace PS, et al. IFN- γ -independent immune markers of *Mycobacterium tuberculosis* exposure. *Nat Med*. juin 2019;25(6):977-87.
49. Golden MP, Vikram HR. Extrapulmonary tuberculosis: an overview. *Am Fam Physician*. 1 nov 2005;72(9):1761-8.
50. Bennett JE, Dolin R, Bennett JE. *Mandell, Douglas y Bennet Enfermedades infecciosas. Principios y práctica + acceso online*. Elsevier España; 2011. 4128 p.

51. Alcaïs A, Fieschi C, Abel L, Casanova J-L. Tuberculosis in children and adults: two distinct genetic diseases. *Journal of Experimental Medicine*. 19 déc 2005;202(12):1617-21.
52. Naveen G, Peerapur BV. Comparison of the Lowenstein-Jensen Medium, the Middlebrook 7H10 Medium and MB/BacT for the Isolation of Mycobacterium Tuberculosis (MTB) from Clinical Specimens. *J Clin Diagn Res*. déc 2012;6(10):1704-9.
53. Furin J, Cox H, Pai M. Tuberculosis. *Lancet*. 20 avr 2019;393(10181):1642-56.
54. Xie YL, Chakravorty S, Armstrong DT, Hall SL, Via LE, Song T, et al. Evaluation of a Rapid Molecular Drug-Susceptibility Test for Tuberculosis. *N Engl J Med*. 14 2017;377(11):1043-54.
55. Satta G, Lipman M, Smith GP, Arnold C, Kon OM, McHugh TD. Mycobacterium tuberculosis and whole-genome sequencing: how close are we to unleashing its full potential? *Clin Microbiol Infect*. juin 2018;24(6):604-9.
56. Pande T, Cohen C, Pai M, Ahmad Khan F. Computer-aided detection of pulmonary tuberculosis on digital chest radiographs: a systematic review. *Int J Tuberc Lung Dis*. 2016;20(9):1226-30.
57. Goletti D, Lee M-R, Wang J-Y, Walter N, Ottenhoff THM. Update on tuberculosis biomarkers: From correlates of risk, to correlates of active disease and of cure from disease. *Respirology*. 2018;23(5):455-66.
58. McShane H. Tuberculosis vaccines: beyond bacille Calmette-Guerin. *Philos Trans R Soc Lond, B, Biol Sci*. 12 oct 2011;366(1579):2782-9.
59. Maartens G, Wilkinson RJ. Tuberculosis. *The Lancet*. 15 déc 2007;370(9604):2030-43.
60. 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*. févr 2014;58(4):470-80.
61. Trunz BB, Fine P, Dye C. Effect of BCG vaccination on childhood tuberculous meningitis and miliary tuberculosis worldwide: a meta-analysis and assessment of cost-effectiveness. *The Lancet*. 8 avr 2006;367(9517):1173-80.
62. McShane H, Jacobs WR, Fine PE, Reed SG, McMurray DN, Behr M, et al. BCG: myths, realities, and the need for alternative vaccine strategies. *Tuberculosis (Edinb)*. mai 2012;92(3):283-8.
63. Calmette A, Guerin C, Weill-Halle B. Essai d'immunisation contre l'infection tuberculeuse. *Bull Acad Med*. 1924;91:787-96.
64. Calmette A, Guerin C, Negre L, et al. Sur la vaccination préventive des enfants nouveau-nés contre la tuberculose par le BCG. *Ann Inst Pasteur*. 1927;41:201-32.
65. Hart PD, Sutherland I. BCG and vole bacillus vaccines in the prevention of tuberculosis in adolescence and early adult life. *Br Med J*. 30 juill 1977;2(6082):293-5.
66. Comstock GW, Palmer CE. Long-term results of BCG vaccination in the southern United States. *Am Rev Respir Dis*. févr 1966;93(2):171-83.

67. Fomukong NG, Dale JW, Osborn TW, Grange JM. Use of gene probes based on the insertion sequence IS986 to differentiate between BCG vaccine strains. *J Appl Bacteriol.* févr 1992;72(2):126-33.
68. Li H, Ulstrup JC, Jonassen TO, Melby K, Nagai S, Harboe M. Evidence for absence of the MPB64 gene in some substrains of *Mycobacterium bovis* BCG. *Infect Immun.* mai 1993;61(5):1730-4.
69. LUCA S, MIHAESCU T. History of BCG Vaccine. *Maedica (Buchar).* mars 2013;8(1):53-8.
70. Brennan MJ, Thole J. Tuberculosis Vaccines: A Strategic Blueprint for the Next Decade. *Tuberculosis.* 1 mars 2012;92:S6-13.
71. Menzies D, Adjobimey M, Ruslami R, Trajman A, Sow O, Kim H, et al. Four Months of Rifampin or Nine Months of Isoniazid for Latent Tuberculosis in Adults. *N Engl J Med.* 02 2018;379(5):440-53.
72. 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.* 18 août 1999;282(7):677-86.
73. Eisenberg JNS, Desai MA, Levy K, Bates SJ, Liang S, Naumoff K, et al. Environmental determinants of infectious disease: a framework for tracking causal links and guiding public health research. *Environ Health Perspect.* août 2007;115(8):1216-23.
74. Frost WH. Risk of Persons in Familial Contact with Pulmonary Tuberculosis. *Am J Public Health Nations Health.* mai 1933;23(5):426-32.
75. Opie EL, McPHEDRAN FM. SPREAD OF TUBERCULOSIS WITHIN FAMILIES. *JAMA.* 6 nov 1926;87(19):1549-51.
76. Andrews RH, Devadatta S, Fox W, Radhakrishna S, Ramakrishnan CV, Velu S. Prevalence of tuberculosis among close family contacts of tuberculous patients in South India, and influence of segregation of the patient on early attack rate. *Bull World Health Organ.* 1960;23:463-510.
77. Devadatta S, Dawson JJ, Fox W, Janardhanam B, Radhakrishna S, Ramakrishnan CV, et al. Attack rate of tuberculosis in a 5-year period among close family contacts of tuberculous patients under domiciliary treatment with isoniazid plus PAS or isoniazid alone. *Bull World Health Organ.* 1970;42(3):337-51.
78. Kamat SR, Dawson JJ, Devadatta S, Fox W, Janardhanam B, Radhakrishna S, et al. A controlled study of the influence of segregation of tuberculous patients for one year on the attack rate of tuberculosis in a 5-year period in close family contacts in South India. *Bull World Health Organ.* 1966;34(4):517-32.
79. Espinal MA, Pérez EN, Baéz J, Hénriquéz L, Fernández K, Lopez M, et al. Infectiousness of *Mycobacterium tuberculosis* in HIV-1-infected patients with tuberculosis: a prospective study. *Lancet.* 22 janv 2000;355(9200):275-80.
80. Dormans J, Burger M, Aguilar D, Hernandez-Pando R, Kremer K, Roholl P, et al. Correlation of virulence, lung pathology, bacterial load and delayed type hypersensitivity responses after infection with different *Mycobacterium tuberculosis* genotypes in a BALB/c mouse model. *Clin Exp Immunol.* sept 2004;137(3):460-8.

81. Nicol MP, Wilkinson RJ. The clinical consequences of strain diversity in *Mycobacterium tuberculosis*. *Trans R Soc Trop Med Hyg.* oct 2008;102(10):955-65.
82. Parwati I, van Crevel R, van Soolingen D. Possible underlying mechanisms for successful emergence of the *Mycobacterium tuberculosis* Beijing genotype strains. *Lancet Infect Dis.* févr 2010;10(2):103-11.
83. 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.* 1 oct 2008;198(7):1037-43.
84. Chandra RK, Kumari S. Nutrition and immunity: an overview. *J Nutr.* 1994;124(8 Suppl):1433S-1435S.
85. Maurya V, Vijayan VK, Shah A. Smoking and tuberculosis: an association overlooked. *Int J Tuberc Lung Dis.* nov 2002;6(11):942-51.
86. Pai M, Mohan A, Dheda K, Leung CC, Yew WW, Christopher DJ, et al. Lethal interaction: the colliding epidemics of tobacco and tuberculosis. *Expert Rev Anti Infect Ther.* juin 2007;5(3):385-91.
87. Bates MN, Khalakdina A, Pai M, Chang L, Lessa F, Smith KR. Risk of tuberculosis from exposure to tobacco smoke: a systematic review and meta-analysis. *Arch Intern Med.* 26 févr 2007;167(4):335-42.
88. Sharma OA. Avicenna's description of tuberculosis. *Bull Indian Inst Hist Med Hyderabad.* 1981;1-4:83-6.
89. Rajalakshmi S, Veluchamy G. Yugi's pramegam and diabetes mellitus: an analogue. *Bull Indian Inst Hist Med Hyderabad.* janv 1999;29(1):83-7.
90. Jeon CY, Murray MB. Diabetes mellitus increases the risk of active tuberculosis: a systematic review of 13 observational studies. *PLoS Med.* 15 juill 2008;5(7):e152.
91. Corbett EL, Watt CJ, Walker N, Maher D, Williams BG, Raviglione MC, et al. The growing burden of tuberculosis: global trends and interactions with the HIV epidemic. *Arch Intern Med.* 12 mai 2003;163(9):1009-21.
92. Collins KR, Quiñones-Mateu ME, Toossi Z, Arts EJ. Impact of tuberculosis on HIV-1 replication, diversity, and disease progression. *AIDS Rev.* sept 2002;4(3):165-76.
93. Ellis PK, Martin WJ, Dodd PJ. CD4 count and tuberculosis risk in HIV-positive adults not on ART: a systematic review and meta-analysis. *PeerJ [Internet].* 14 déc 2017 [cité 15 oct 2019];5. Disponible sur: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5733368/>
94. Williams BG, Dye C. Antiretroviral drugs for tuberculosis control in the era of HIV/AIDS. *Science.* 12 sept 2003;301(5639):1535-7.
95. Williams BG, Granich R, De Cock KM, Glaziou P, Sharma A, Dye C. Antiretroviral therapy for tuberculosis control in nine African countries. *Proc Natl Acad Sci USA.* 9 nov 2010;107(45):19485-9.

96. Smith R, Cattamanchi A, Steingart KR, Denkinger C, Dheda K, Winthrop KL, et al. Interferon- γ release assays for diagnosis of latent tuberculosis infection: evidence in immune-mediated inflammatory disorders. *Curr Opin Rheumatol*. juill 2011;23(4):377-84.
97. Winthrop KL. Risk and prevention of tuberculosis and other serious opportunistic infections associated with the inhibition of tumor necrosis factor. *Nat Clin Pract Rheumatol*. nov 2006;2(11):602-10.
98. Nicolle C. *Destin des maladies infectieuses*. France Lafayette; 1993.
99. Casanova J-L, Abel L. The genetic theory of infectious diseases: a brief history and selected illustrations. *Annu Rev Genomics Hum Genet*. 2013;14:215-43.
100. *Familial Susceptibility to Tuberculosis: Its Importance as a Public Health Problem*: By Ruth Rige Puffer. The Harvard University Press. Price \$2.00. *British Journal of Tuberculosis and Diseases of the Chest*. 1 oct 1945;39(4):107-8.
101. Stead WW, Senner JW, Reddick WT, Lofgren JP. Racial differences in susceptibility to infection by *Mycobacterium tuberculosis*. *N Engl J Med*. 15 févr 1990;322(7):422-7.
102. Wang L, Noertjojo K, Elwood RK, FitzGerald JM. Tuberculosis among aboriginal and nonaboriginal persons in British Columbia. *Can Respir J*. avr 2000;7(2):151-7.
103. Clark M, Vynnycky E. The use of maximum likelihood methods to estimate the risk of tuberculous infection and disease in a Canadian First Nations population. *Int J Epidemiol*. juin 2004;33(3):477-84.
104. Correa-Macedo W, Cambri G, Schurr E. The Interplay of Human and *Mycobacterium Tuberculosis* Genomic Variability. *Front Genet* [Internet]. 2019 [cité 8 nov 2019];10. Disponible sur: <https://www.frontiersin.org/articles/10.3389/fgene.2019.00865/full>
105. Fox GJ, Orlova M, Schurr E. Tuberculosis in Newborns: The Lessons of the “Lübeck Disaster” (1929–1933). *PLoS Pathog* [Internet]. 21 janv 2016 [cité 4 juill 2019];12(1). Disponible sur: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4721647/>
106. Kallmann FJ, Reisner D. Twin Studies on the Significance of Genetic Factors in Tuberculosis. *American Review of Tuberculosis and Pulmonary Diseases*. 1943;47(6):549-71.
107. Comstock GW. Tuberculosis in twins: a re-analysis of the Proffit survey. *Am Rev Respir Dis*. avr 1978;117(4):621-4.
108. Lurie MB. THE CORRELATION BETWEEN THE HISTOLOGICAL CHANGES AND THE FATE OF LIVING TUBERCLE BACILLI IN THE ORGANS OF TUBERCULOUS RABBITS. *J Exp Med*. 1 janv 1932;55(1):31-54.
109. Lurie MB, Dannenberg AM. Macrophage Function in Infectious Disease with Inbred Rabbits. *Bacteriol Rev*. déc 1965;29(4):466-76.
110. Lurie MB. Studies on the Mechanism of Immunity in Tuberculosis: The Rôle of Extracellular Factors and Local Immunity in the Fixation and Inhibition of Growth of Tubercle Bacilli. *Journal of Experimental Medicine*. 1 avr 1939;69(4):555-78.

111. Lurie MB, Zappasodi W the C of P. Studies on the Mechanism of Immunity in Tuberculosis: The Fate of Tubercle Bacilli Ingested by Mononuclear Phagocytes Derived from Normal and Immunized Animals. *Journal of Experimental Medicine*. 1 mars 1942;75(3):247-68.
112. Malo D, Vidal SM, Hu J, Skamene E, Gros P. High-resolution linkage map in the vicinity of the host resistance locus Bcg. *Genomics*. juin 1993;16(3):655-63.
113. Skamene E, Gros P, Forget A, Kongshavn PAL, Charles CS, Taylor BA. Genetic regulation of resistance to intracellular pathogens. *Nature*. juin 1982;297(5866):506-9.
114. Gros P, Skamene E, Forget A. Genetic control of natural resistance to *Mycobacterium bovis* (BCG) in mice. *J Immunol*. déc 1981;127(6):2417-21.
115. Arias M, Rojas M, Zabaleta J, Rodríguez JI, París SC, Barrera LF, et al. Inhibition of virulent *Mycobacterium tuberculosis* by Bcg(r) and Bcg(s) macrophages correlates with nitric oxide production. *J Infect Dis*. déc 1997;176(6):1552-8.
116. de Chastellier C, Fréhel C, Offredo C, Skamene E. Implication of phagosome-lysosome fusion in restriction of *Mycobacterium avium* growth in bone marrow macrophages from genetically resistant mice. *Infect Immun*. sept 1993;61(9):3775-84.
117. Lissner CR, Swanson RN, O'Brien AD. Genetic control of the innate resistance of mice to *Salmonella typhimurium*: expression of the *Ity* gene in peritoneal and splenic macrophages isolated in vitro. *J Immunol*. déc 1983;131(6):3006-13.
118. Vidal SM, Malo D, Vogan K, Skamene E, Gros P. Natural resistance to infection with intracellular parasites: isolation of a candidate for Bcg. *Cell*. 7 mai 1993;73(3):469-85.
119. Vidal SM, Pinner E, Lepage P, Gauthier S, Gros P. Natural resistance to intracellular infections: *Nramp1* encodes a membrane phosphoglycoprotein absent in macrophages from susceptible (*Nramp1* D169) mouse strains. *J Immunol*. 15 oct 1996;157(8):3559-68.
120. Malo D, Vogan K, Vidal S, Hu J, Cellier M, Schurr E, et al. Haplotype mapping and sequence analysis of the mouse *Nramp* gene predict susceptibility to infection with intracellular parasites. *Genomics*. 1 sept 1994;23(1):51-61.
121. Goldfeld AE, Delgado JC, Thim S, Bozon MV, Ugliarolo AM, Turbay D, et al. Association of an HLA-DQ allele with clinical tuberculosis. *JAMA*. 21 janv 1998;279(3):226-8.
122. Bellamy R, Ruwende C, Corrah T, McAdam KP, Thursz M, Whittle HC, et al. Tuberculosis and chronic hepatitis B virus infection in Africans and variation in the vitamin D receptor gene. *J Infect Dis*. mars 1999;179(3):721-4.
123. Wilkinson RJ, Patel P, Llewelyn M, Hirsch CS, Pasvol G, Snounou G, et al. Influence of polymorphism in the genes for the interleukin (IL)-1 receptor antagonist and IL-1beta on tuberculosis. *J Exp Med*. 21 juin 1999;189(12):1863-74.
124. Bellamy R, Ruwende C, Corrah T, McAdam KP, Whittle HC, Hill AV. Variations in the *NRAMP1* gene and susceptibility to tuberculosis in West Africans. *N Engl J Med*. 5 mars 1998;338(10):640-4.

125. Greenwood CM, Fujiwara TM, Boothroyd LJ, Miller MA, Frappier D, Fanning EA, et al. Linkage of tuberculosis to chromosome 2q35 loci, including NRAMP1, in a large aboriginal Canadian family. *Am J Hum Genet.* août 2000;67(2):405-16.
126. Malik S, Abel L, Tooker H, Poon A, Simkin L, Girard M, et al. Alleles of the NRAMP1 gene are risk factors for pediatric tuberculosis disease. *Proc Natl Acad Sci USA.* 23 août 2005;102(34):12183-8.
127. Canonne-Hergaux F, Gruenheid S, Govoni G, Gros P. The Nramp1 protein and its role in resistance to infection and macrophage function. *Proc Assoc Am Physicians.* août 1999;111(4):283-9.
128. Abel L, Sánchez FO, Oberti J, Thuc NV, Hoa LV, Lap VD, et al. Susceptibility to leprosy is linked to the human NRAMP1 gene. *J Infect Dis.* janv 1998;177(1):133-45.
129. Fisher RA. The Detection of Linkage with "Dominant" Abnormalities. *Annals of Eugenics.* 1935;6(2):187-201.
130. Penrose LS. The Detection of Autosomal Linkage in Data Which Consist of Pairs of Brothers and Sisters of Unspecified Parentage. *Annals of Eugenics.* 1935;6(2):133-8.
131. Baghdadi JE, Orlova M, Alter A, Ranque B, Chentoufi M, Lazrak F, et al. An autosomal dominant major gene confers predisposition to pulmonary tuberculosis in adults. *J Exp Med.* 10 juill 2006;203(7):1679-84.
132. Grant AV, El Baghdadi J, Sabri A, El Azbaoui S, Alaoui-Tahiri K, Abderrahmani Rhorfi I, et al. Age-dependent association between pulmonary tuberculosis and common TOX variants in the 8q12-13 linkage region. *Am J Hum Genet.* 7 mars 2013;92(3):407-14.
133. Aliahmad P, Seksenyan A, Kaye J. The many roles of TOX in the immune system. *Curr Opin Immunol.* avr 2012;24(2):173-7.
134. Khan O, Giles JR, McDonald S, Manne S, Ngiow SF, Patel KP, et al. TOX transcriptionally and epigenetically programs CD8 + T cell exhaustion. *Nature.* juill 2019;571(7764):211.
135. Yao C, Sun H-W, Lacey NE, Ji Y, Moseman EA, Shih H-Y, et al. Single-cell RNA-seq reveals TOX as a key regulator of CD8 + T cell persistence in chronic infection. *Nature Immunology.* juill 2019;20(7):890.
136. Alfei F, Kanev K, Hofmann M, Wu M, Ghoneim HE, Roelli P, et al. TOX reinforces the phenotype and longevity of exhausted T cells in chronic viral infection. *Nature.* juill 2019;571(7764):265.
137. Thye T, Owusu-Dabo E, Vannberg FO, van Crevel R, Curtis J, Sahiratmadja E, et al. Common variants at 11p13 are associated with susceptibility to tuberculosis. *Nat Genet.* 5 févr 2012;44(3):257-9.
138. Thye T, Vannberg FO, Wong SH, Owusu-Dabo E, Osei I, Gyapong J, et al. Genome-wide association analyses identifies a susceptibility locus for tuberculosis on chromosome 18q11.2. *Nat Genet.* sept 2010;42(9):739-41.

139. Curtis J, Luo Y, Zenner HL, Cuchet-Lourenço D, Wu C, Lo K, et al. Susceptibility to tuberculosis is associated with variants in the ASAP1 gene encoding a regulator of dendritic cell migration. *Nat Genet.* mai 2015;47(5):523-7.
140. Chimusa ER, Zaitlen N, Daya M, Möller M, van Helden PD, Mulder NJ, et al. Genome-wide association study of ancestry-specific TB risk in the South African Coloured population. *Hum Mol Genet.* 1 févr 2014;23(3):796-809.
141. Grant AV, Sabri A, Abid A, Abderrahmani Rhorfi I, Benkirane M, Souhi H, et al. A genome-wide association study of pulmonary tuberculosis in Morocco. *Hum Genet.* mars 2016;135(3):299-307.
142. Dai Y, Zhang X, Pan H, Tang S, Shen H, Wang J. Fine mapping of genetic polymorphisms of pulmonary tuberculosis within chromosome 18q11.2 in the Chinese population: a case-control study. *BMC Infect Dis.* 22 oct 2011;11:282.
143. Wang X, Tang NL-S, Leung CC, Kam KM, Yew WW, Tam CM, et al. Association of polymorphisms in the Chr18q11.2 locus with tuberculosis in Chinese population. *Hum Genet.* juin 2013;132(6):691-5.
144. 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.* mars 2016;48(3):318-22.
145. Altare F, Lammas D, Revy P, Jouanguy E, Döffinger R, Lamhamedi S, et al. Inherited interleukin 12 deficiency in a child with bacille Calmette-Guérin and Salmonella enteritidis disseminated infection. *J Clin Invest.* 15 déc 1998;102(12):2035-40.
146. Boisson-Dupuis S, El Baghdadi J, Parvaneh N, Bousfiha A, Bustamante J, Feinberg J, et al. IL-12R β 1 deficiency in two of fifty children with severe tuberculosis from Iran, Morocco, and Turkey. *PLoS ONE.* 13 avr 2011;6(4):e18524.
147. 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.* mars 2015;264(1):103-20.
148. Caragol I, Raspall M, Fieschi C, Feinberg J, Larrosa MN, Hernández M, et al. Clinical tuberculosis in 2 of 3 siblings with interleukin-12 receptor beta1 deficiency. *Clin Infect Dis.* 15 juill 2003;37(2):302-6.
149. Kreins AY, Ciancanelli MJ, Okada S, Kong X-F, Ramírez-Alejo N, Kilic SS, et al. Human TYK2 deficiency: Mycobacterial and viral infections without hyper-IgE syndrome. *Journal of Experimental Medicine.* 21 sept 2015;212(10):1641-62.
150. Ozbek N, Fieschi C, Yilmaz BT, de Beaucoudrey L, Demirhan B, Feinberg J, et al. Interleukin-12 receptor beta 1 chain deficiency in a child with disseminated tuberculosis. *Clin Infect Dis.* 15 mars 2005;40(6):e55-58.
151. Tabarsi P, Marjani M, Mansouri N, Farnia P, Boisson-Dupuis S, Bustamante J, et al. Lethal tuberculosis in a previously healthy adult with IL-12 receptor deficiency. *J Clin Immunol.* août 2011;31(4):537-9.

152. Jouanguy E, Altare F, Lamhamedi S, Revy P, Emile JF, Newport M, et al. Interferon-gamma-receptor deficiency in an infant with fatal bacille Calmette-Guérin infection. *N Engl J Med*. 26 déc 1996;335(26):1956-61.
153. Rosain J, Kong X-F, Martinez-Barricarte R, Oleaga-Quintas C, Ramirez-Alejo N, Markle J, et al. Mendelian susceptibility to mycobacterial disease: 2014-2018 update. *Immunol Cell Biol*. 28 sept 2018;
154. Jouanguy E, Altare F, Lamhamedi S, Revy P, Emile JF, Newport M, et al. Interferon-gamma-receptor deficiency in an infant with fatal bacille Calmette-Guérin infection. *N Engl J Med*. 26 déc 1996;335(26):1956-61.
155. Bustamante J, Boisson-Dupuis S, Abel L, Casanova J-L. Mendelian susceptibility to mycobacterial disease: genetic, immunological, and clinical features of inborn errors of IFN- γ immunity. *Semin Immunol*. déc 2014;26(6):454-70.
156. Kreins AY, Ciancanelli MJ, Okada S, Kong X-F, Ramírez-Alejo N, Kilic SS, et al. Human TYK2 deficiency: Mycobacterial and viral infections without hyper-IgE syndrome. *Journal of Experimental Medicine*. 21 sept 2015;212(10):1641-62.
157. Kong X-F, Martinez-Barricarte R, Kennedy J, Mele F, Lazarov T, Deenick EK, et al. Disruption of an antimycobacterial circuit between dendritic and helper T cells in human SPPL2a deficiency. *Nat Immunol*. 2018;19(9):973-85.
158. Altare F, Ensser A, Breiman A, Reichenbach J, Baghdadi JE, Fischer A, et al. Interleukin-12 receptor beta1 deficiency in a patient with abdominal tuberculosis. *J Infect Dis*. 15 juill 2001;184(2):231-6.
159. Caragol I, Raspall M, Fieschi C, Feinberg J, Larrosa MN, Hernández M, et al. Clinical tuberculosis in 2 of 3 siblings with interleukin-12 receptor beta1 deficiency. *Clin Infect Dis*. 15 juill 2003;37(2):302-6.
160. 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*. 1 déc 1997;100(11):2658-64.
161. Chapgier A, Boisson-Dupuis S, Jouanguy E, Vogt G, Feinberg J, Prochnicka-Chalufour A, et al. Novel STAT1 alleles in otherwise healthy patients with mycobacterial disease. *PLoS Genet*. 18 août 2006;2(8):e131.
162. Bustamante J, Arias AA, Vogt G, Picard C, Galicia LB, Prando C, et al. Germline CYBB mutations that selectively affect macrophages in kindreds with X-linked predisposition to tuberculous mycobacterial disease. *Nat Immunol*. mars 2011;12(3):213-21.
163. Boisson-Dupuis S, Kong X-F, Okada S, Cypowyj S, Puel A, Abel L, et al. Inborn errors of human STAT1: allelic heterogeneity governs the diversity of immunological and infectious phenotypes. *Curr Opin Immunol*. août 2012;24(4):364-78.
164. Bustamante J, Arias AA, Vogt G, Picard C, Galicia LB, Prando C, et al. Germline CYBB mutations that selectively affect macrophages in kindreds with X-linked predisposition to tuberculous mycobacterial disease. *Nat Immunol*. mars 2011;12(3):213-21.

165. de Beaucoudrey L, Samarina A, Bustamante J, Cobat A, Boisson-Dupuis S, Feinberg J, et al. Revisiting human IL-12R β 1 deficiency: a survey of 141 patients from 30 countries. *Medicine (Baltimore)*. nov 2010;89(6):381-402.
166. Filipe-Santos O, Bustamante J, Haverkamp MH, Vinolo E, Ku C-L, Puel A, et al. X-linked susceptibility to mycobacteria is caused by mutations in NEMO impairing CD40-dependent IL-12 production. *J Exp Med*. 10 juill 2006;203(7):1745-59.
167. Picard C, Casanova J-L, Puel A. Infectious diseases in patients with IRAK-4, MyD88, NEMO, or I κ B α deficiency. *Clin Microbiol Rev*. juill 2011;24(3):490-7.
168. Prando C, Samarina A, Bustamante J, Boisson-Dupuis S, Cobat A, Picard C, et al. Inherited IL-12p40 deficiency: genetic, immunologic, and clinical features of 49 patients from 30 kindreds. *Medicine (Baltimore)*. mars 2013;92(2):109-22.
169. 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(3):e64-75.
170. Grant AV, El Baghdadi J, Sabri A, El Azbaoui S, Alaoui-Tahiri K, Abderrahmani Rhorfi I, et al. Age-Dependent Association between Pulmonary Tuberculosis and Common TOX Variants in the 8q12-13 Linkage Region. *Am J Hum Genet*. 7 mars 2013;92(3):407-14.
171. Meyts I, Bosch B, Bolze A, Boisson B, Itan Y, Belkadi A, et al. Exome and genome sequencing for inborn errors of immunity. *J Allergy Clin Immunol*. 2016;138(4):957-69.
172. Brastianos PK, Taylor-Weiner A, Manley PE, Jones RT, Dias-Santagata D, Thorner AR, et al. Exome sequencing identifies BRAF mutations in papillary craniopharyngiomas. *Nat Genet*. févr 2014;46(2):161-5.
173. Kaiser J. Human genetics. Affordable « exomes » fill gaps in a catalog of rare diseases. *Science*. 12 nov 2010;330(6006):903.
174. Meyts I, Bosch B, Bolze A, Boisson B, Itan Y, Belkadi A, et al. Exome and genome sequencing for inborn errors of immunity. *J Allergy Clin Immunol*. 2016;138(4):957-69.
175. Zastrow DB, Zornio PA, Dries A, Kohler J, Fernandez L, Waggott D, et al. Exome sequencing identifies de novo pathogenic variants in FBN1 and TRPS1 in a patient with a complex connective tissue phenotype. *Cold Spring Harb Mol Case Stud*. janv 2017;3(1):a001388.
176. Boisson-Dupuis S, Ramirez-Alejo N, Li Z, Patin E, Rao G, Kerner G, et al. Tuberculosis and impaired IL-23-dependent IFN- γ immunity in humans homozygous for a common TYK2 missense variant. *Sci Immunol*. 21 déc 2018;3(30).
177. Boisson B, Wang C, Pedergnana V, Wu L, Cypowyj S, Rybojad M, et al. An ACT1 mutation selectively abolishes interleukin-17 responses in humans with chronic mucocutaneous candidiasis. *Immunity*. 17 oct 2013;39(4):676-86.
178. Bustamante J, Boisson-Dupuis S, Abel L, Casanova J-L. Mendelian susceptibility to mycobacterial disease: genetic, immunological, and clinical features of inborn errors of IFN- γ immunity. *Semin Immunol*. déc 2014;26(6):454-70.

179. Casanova J-L, Abel L. The genetic theory of infectious diseases: a brief history and selected illustrations. *Annu Rev Genomics Hum Genet.* 2013;14:215-43.
180. Ciancanelli MJ, Huang SXL, Luthra P, Garner H, Itan Y, Volpi S, et al. Infectious disease. Life-threatening influenza and impaired interferon amplification in human IRF7 deficiency. *Science.* 24 avr 2015;348(6233):448-53.
181. Guérin A, Kerner G, Marr N, Markle JG, Fenollar F, Wong N, et al. IRF4 haploinsufficiency in a family with Whipple's disease. *Elife.* 14 mars 2018;7.
182. DePristo MA, Banks E, Poplin R, Garimella KV, Maguire JR, Hartl C, et al. A framework for variation discovery and genotyping using next-generation DNA sequencing data. *Nat Genet.* mai 2011;43(5):491-8.
183. 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.* sept 2007;81(3):559-75.
184. Sherry ST, Ward MH, Kholodov M, Baker J, Phan L, Smigielski EM, et al. dbSNP: the NCBI database of genetic variation. *Nucleic Acids Res.* 1 janv 2001;29(1):308-11.
185. Cingolani P, Platts A, Wang LL, Coon M, Nguyen T, Wang L, et al. A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff: SNPs in the genome of *Drosophila melanogaster* strain w1118; iso-2; iso-3. *Fly (Austin).* juin 2012;6(2):80-92.
186. Pulendran B, Maddur MS. Innate immune sensing and response to influenza. *Curr Top Microbiol Immunol.* 2015;386:23-71.
187. Jaber S, Conseil M, Coisel Y, Jung B, Chanques G. [ARDS and influenza A (H1N1): patients' characteristics and management in intensive care unit. A literature review]. *Ann Fr Anesth Reanim.* févr 2010;29(2):117-25.
188. Glaser CA, Winter K, DuBray K, Harriman K, Uyeki TM, Sejvar J, et al. A population-based study of neurologic manifestations of severe influenza A(H1N1)pdm09 in California. *Clin Infect Dis.* août 2012;55(4):514-20.
189. Ciancanelli MJ, Huang SXL, Luthra P, Garner H, Itan Y, Volpi S, et al. Life-threatening influenza and impaired interferon amplification in human IRF7 deficiency. *Science.* 24 avr 2015;348(6233):448-53.
190. Zhang S-Y, Jouanguy E, Zhang Q, Abel L, Puel A, Casanova J-L. Human inborn errors of immunity to infection affecting cells other than leukocytes: from the immune system to the whole organism. *Current Opinion in Immunology.* 1 août 2019;59:88-100.
191. Itan Y, Shang L, Boisson B, Ciancanelli M, Markle J, Martinez-Barricarte R, et al. The mutation significance cutoff (MSC): gene-level thresholds for variant-level predictions. *Nat Methods.* févr 2016;13(2):109-10.
192. Itan Y, Zhang S-Y, Vogt G, Abhyankar A, Herman M, Nitschke P, et al. The human gene connectome as a map of short cuts for morbid allele discovery. *PNAS.* 2 avr 2013;110(14):5558-63.

193. Alexopoulou L, Holt AC, Medzhitov R, Flavell RA. Recognition of double-stranded RNA and activation of NF-kappaB by Toll-like receptor 3. *Nature*. 18 oct 2001;413(6857):732-8.
194. Zhang S-Y, Jouanguy E, Ugolini S, Smahi A, Elain G, Romero P, et al. TLR3 Deficiency in Patients with Herpes Simplex Encephalitis. *Science*. 14 sept 2007;317(5844):1522-7.
195. Lim HK, Huang SXL, Chen J, Kerner G, Gilliaux O, Bastard P, et al. Severe influenza pneumonitis in children with inherited TLR3 deficiency. *Journal of Experimental Medicine*. 2 sept 2019;216(9):2038-56.
196. Guo Y, Audry M, Ciancanelli M, Alsina L, Azevedo J, Herman M, et al. Herpes simplex virus encephalitis in a patient with complete TLR3 deficiency: TLR3 is otherwise redundant in protective immunity. *Journal of Experimental Medicine*. 26 sept 2011;208(10):2083-98.
197. Lim HK, Seppänen M, Hautala T, Ciancanelli MJ, Itan Y, Lafaille FG, et al. TLR3 deficiency in herpes simplex encephalitis: high allelic heterogeneity and recurrence risk. *Neurology*. 18 nov 2014;83(21):1888-97.
198. Casrouge A, Zhang S-Y, Eidenschenk C, Jouanguy E, Puel A, Yang K, et al. Herpes Simplex Virus Encephalitis in Human UNC-93B Deficiency. *Science*. 13 oct 2006;314(5797):308-12.
199. Pérez de Diego R, Sancho-Shimizu V, Lorenzo L, Puel A, Plancoulaine S, Picard C, et al. Human TRAF3 adaptor molecule deficiency leads to impaired Toll-like receptor 3 response and susceptibility to herpes simplex encephalitis. *Immunity*. 24 sept 2010;33(3):400-11.
200. Herman M, Ciancanelli M, Ou Y-H, Lorenzo L, Klaudel-Dreszler M, Pauwels E, et al. Heterozygous TBK1 mutations impair TLR3 immunity and underlie herpes simplex encephalitis of childhood. *Journal of Experimental Medicine*. 27 août 2012;209(9):1567-82.
201. Andersen LL, Mørk N, Reinert LS, Kofod-Olsen E, Narita R, Jørgensen SE, et al. Functional IRF3 deficiency in a patient with herpes simplex encephalitis. *Journal of Experimental Medicine*. 24 août 2015;212(9):1371-9.
202. Shawky RM, Elsayed SM, Zaki ME, Nour El-Din SM, Kamal FM. Consanguinity and its relevance to clinical genetics. *Egyptian Journal of Medical Human Genetics*. 1 avr 2013;14(2):157-64.
203. Bennett RL, Motulsky AG, Bittles A, Hudgins L, Uhrich S, Doyle DL, et al. Genetic Counseling and Screening of Consanguineous Couples and Their Offspring: Recommendations of the National Society of Genetic Counselors. *J Genet Couns*. avr 2002;11(2):97-119.
204. Baron M. The search for complex disease genes: fault by linkage or fault by association? *Mol Psychiatry*. mars 2001;6(2):143-9.
205. Abecasis GR, Cherny SS, Cookson WO, Cardon LR. Merlin--rapid analysis of dense genetic maps using sparse gene flow trees. *Nat Genet*. janv 2002;30(1):97-101.
206. Cabrera A, Neculai D, Kain KC. CD36 and malaria: friends or foes? A decade of data provides some answers. *Trends in Parasitology*. 1 sept 2014;30(9):436-44.
207. Casanova J-L, Abel L. Human genetics of infectious diseases: Unique insights into immunological redundancy. *Semin Immunol*. avr 2018;36:1-12.

208. Eilertson KE, Booth JG, Bustamante CD. SnIPRE: Selection Inference Using a Poisson Random Effects Model. *PLOS Computational Biology*. 6 déc 2012;8(12):e1002806.
209. Itan Y, Shang L, Boisson B, Patin E, Bolze A, Moncada-Vélez M, et al. The human gene damage index as a gene-level approach to prioritizing exome variants. *PNAS*. 15 oct 2015;201518646.
210. Minegishi Y, Saito M, Morio T, Watanabe K, Agematsu K, Tsuchiya S, et al. Human tyrosine kinase 2 deficiency reveals its requisite roles in multiple cytokine signals involved in innate and acquired immunity. *Immunity*. nov 2006;25(5):745-55.
211. Sohn SJ, Barrett K, Abbema AV, Chang C, Kohli PB, Kanda H, et al. A Restricted Role for TYK2 Catalytic Activity in Human Cytokine Responses Revealed by Novel TYK2-Selective Inhibitors. *The Journal of Immunology*. 1 sept 2013;191(5):2205-16.
212. Fuchs S, Kaiser-Labusch P, Bank J, Ammann S, Kolb-Kokocinski A, Edelbusch C, et al. Tyrosine kinase 2 is not limiting human antiviral type III interferon responses. *Eur J Immunol*. 2016;46(11):2639-49.
213. Kilic SS, Hacimustafaoglu M, Boisson-Dupuis S, Kreins AY, Grant AV, Abel L, et al. A patient with tyrosine kinase 2 deficiency without hyper-IgE syndrome. *J Pediatr*. juin 2012;160(6):1055-7.
214. Pósa A, Maixner F, Mende BG, Köhler K, Osztás A, Sola C, et al. Tuberculosis in Late Neolithic-Early Copper Age human skeletal remains from Hungary. *Tuberculosis (Edinb)*. juin 2015;95 Suppl 1:S18-22.
215. Mathieson I, Lazaridis I, Rohland N, Mallick S, Patterson N, Roodenberg SA, et al. Genome-wide patterns of selection in 230 ancient Eurasians. *Nature*. 24 déc 2015;528(7583):499-503.
216. Rigante D, Frediani B, Cantarini L. A Comprehensive Overview of the Hereditary Periodic Fever Syndromes. *Clin Rev Allergy Immunol*. juin 2018;54(3):446-53.
217. Powell LW, Seckington RC, Deugnier Y. Haemochromatosis. *Lancet*. 13 août 2016;388(10045):706-16.
218. Cvijović I, Good BH, Jerison ER, Desai MM. Fate of a mutation in a fluctuating environment. *PNAS*. 8 sept 2015;112(36):E5021-8.
219. Karlsson EK, Kwiatkowski DP, Sabeti PC. Natural selection and infectious disease in human populations. *Nat Rev Genet*. juin 2014;15(6):379-93.
220. Cohn SK. Epidemiology of the Black Death and successive waves of plague. *Med Hist Suppl*. 2008;(27):74-100.
221. Boisson-Dupuis S, Ramirez-Alejo N, Li Z, Patin E, Rao G, Kerner G, et al. Tuberculosis and impaired IL-23-dependent IFN- γ immunity in humans homozygous for a common TYK2 missense variant. *Science Immunology*. 21 déc 2018;3(30):eaau8714.
222. Glaziou P, Floyd K, Raviglione M. Trends in tuberculosis in the UK. *Thorax*. 2018;73(8):702-3.
223. Gray BJ, Perrett SE, Gudgeon B, Shankar AG. Investigating the prevalence of latent Tuberculosis infection in a UK remand prison. *J Public Health (Oxf)* [Internet]. [cité 20 févr

- 2019]; Disponible sur: <https://academic.oup.com/jpubhealth/advance-article/doi/10.1093/pubmed/fdy219/5273169>
224. Public Health England. Tuberculosis in England [Internet]. 2017. Disponible sur: <https://www.gov.uk/government/organisations/public-health-england>
 225. Khanna P, Nikolayevskyy V, Warburton F, Dobson E, Drobniowski F. Rate of latent tuberculosis infection detected by occupational health screening of nurses new to a london teaching hospital. *Infect Control Hosp Epidemiol*. juin 2009;30(6):581-4.
 226. Bycroft C, Freeman C, Petkova D, Band G, Elliott LT, Sharp K, et al. The UK Biobank resource with deep phenotyping and genomic data. *Nature*. oct 2018;562(7726):203.
 227. Woodfield R, UK Biobank Stroke Outcomes Group, UK Biobank Follow-up and Outcomes Working Group, Sudlow CLM. Accuracy of Patient Self-Report of Stroke: A Systematic Review from the UK Biobank Stroke Outcomes Group. *PLoS ONE*. 2015;10(9):e0137538.
 228. Khanna P, Nikolayevskyy V, Warburton F, Dobson E, Drobniowski F. Rate of latent tuberculosis infection detected by occupational health screening of nurses new to a london teaching hospital. *Infect Control Hosp Epidemiol*. juin 2009;30(6):581-4.
 229. Gray BJ, Perrett SE, Gudgeon B, Shankar AG. Investigating the prevalence of latent Tuberculosis infection in a UK remand prison. *J Public Health (Oxf)*. 4 janv 2019;
 230. Canela-Xandri O, Rawlik K, Tenesa A. An atlas of genetic associations in UK Biobank. *Nat Genet*. nov 2018;50(11):1593-9.
 231. Diogo D, Bastarache L, Liao KP, Graham RR, Fulton RS, Greenberg JD, et al. TYK2 protein-coding variants protect against rheumatoid arthritis and autoimmunity, with no evidence of major pleiotropic effects on non-autoimmune complex traits. *PLoS ONE*. 2015;10(4):e0122271.
 232. International Genetics of Ankylosing Spondylitis Consortium (IGAS), Cortes A, Hadler J, Pointon JP, Robinson PC, Karaderi T, et al. Identification of multiple risk variants for ankylosing spondylitis through high-density genotyping of immune-related loci. *Nat Genet*. juill 2013;45(7):730-8.
 233. International Multiple Sclerosis Genetics Consortium (IMSGC), Beecham AH, Patsopoulos NA, Xifara DK, Davis MF, Kempainen A, et al. Analysis of immune-related loci identifies 48 new susceptibility variants for multiple sclerosis. *Nat Genet*. nov 2013;45(11):1353-60.
 234. Jostins L, Ripke S, Weersma RK, Duerr RH, McGovern DP, Hui KY, et al. Host-microbe interactions have shaped the genetic architecture of inflammatory bowel disease. *Nature*. 1 nov 2012;491(7422):119-24.
 235. Tsoi LC, Spain SL, Knight J, Ellinghaus E, Stuart PE, Capon F, et al. Identification of 15 new psoriasis susceptibility loci highlights the role of innate immunity. *Nat Genet*. déc 2012;44(12):1341-8.
 236. Fodil N, Langlais D, Gros P. Primary Immunodeficiencies and Inflammatory Disease: A Growing Genetic Intersection. *Trends Immunol*. févr 2016;37(2):126-40.

237. Langlais D, Fodil N, Gros P. Genetics of Infectious and Inflammatory Diseases: Overlapping Discoveries from Association and Exome-Sequencing Studies. *Annu Rev Immunol.* 26 2017;35:1-30.
238. Gadina M, Johnson C, Schwartz D, Bonelli M, Hasni S, Kanno Y, et al. Translational and clinical advances in JAK-STAT biology: The present and future of jakinibs. *J Leukoc Biol.* sept 2018;104(3):499-514.
239. He X, Chen X, Zhang H, Xie T, Ye X-Y. Selective Tyk2 inhibitors as potential therapeutic agents: a patent review (2015-2018). *Expert Opin Ther Pat.* févr 2019;29(2):137-49.
240. Papp K, Gordon K, Thaçi D, Morita A, Gooderham M, Foley P, et al. Phase 2 Trial of Selective Tyrosine Kinase 2 Inhibition in Psoriasis. *N Engl J Med.* 04 2018;379(14):1313-21.
241. Kaya F. Anti-TNF Treatment and Tuberculosis. *Austin J Pulm Respir Med.* sept 2017;4(2).
242. Fieschi C, Dupuis S, Catherinot E, Feinberg J, Bustamante J, Breiman A, et al. Low Penetrance, Broad Resistance, and Favorable Outcome of Interleukin 12 Receptor β 1 Deficiency: Medical and Immunological Implications. *Journal of Experimental Medicine.* 17 févr 2003;197(4):527-35.
243. Holland SM. Immunotherapy of mycobacterial infections. *Semin Respir Infect.* mars 2001;16(1):47-59.
244. Cohen JC, Hobbs HH. Simple Genetics for a Complex Disease. *Science.* 10 mai 2013;340(6133):689-90.
245. Chakravarti A, Clark AG, Mootha VK. Distilling pathophysiology from complex disease genetics. *Cell.* 26 sept 2013;155(1):21-6.
246. Florez JC, Hirschhorn J, Altshuler D. The inherited basis of diabetes mellitus: implications for the genetic analysis of complex traits. *Annu Rev Genomics Hum Genet.* 2003;4:257-91.
247. Katsanis N. The continuum of causality in human genetic disorders. *Genome Biol.* 17 2016;17(1):233.
248. McClellan J, King M-C. Genetic Heterogeneity in Human Disease. *Cell.* 16 avr 2010;141(2):210-7.
249. McLaren PJ, Carrington M. The impact of host genetic variation on infection with HIV-1. *Nat Immunol.* juin 2015;16(6):577-83.
250. Casanova J-L, Abel L. Genetic dissection of immunity to mycobacteria: the human model. *Annu Rev Immunol.* 2002;20:581-620.
251. Lifton RP. Genetic dissection of human blood pressure variation: common pathways from rare phenotypes. *Harvey Lect.* 2005 2004;100:71-101.
252. 1000 Genomes Project Consortium, Abecasis GR, Altshuler D, Auton A, Brooks LD, Durbin RM, et al. A map of human genome variation from population-scale sequencing. *Nature.* 28 oct 2010;467(7319):1061-73.

253. Xue Y, Chen Y, Ayub Q, Huang N, Ball EV, Mort M, et al. Deleterious- and disease-allele prevalence in healthy individuals: insights from current predictions, mutation databases, and population-scale resequencing. *Am J Hum Genet.* 7 déc 2012;91(6):1022-32.
254. Shen H, Li J, Zhang J, Xu C, Jiang Y, Wu Z, et al. Comprehensive characterization of human genome variation by high coverage whole-genome sequencing of forty four Caucasians. *PLoS ONE.* 2013;8(4):e59494.
255. MacArthur DG, Balasubramanian S, Frankish A, Huang N, Morris J, Walter K, et al. A systematic survey of loss-of-function variants in human protein-coding genes. *Science.* 17 févr 2012;335(6070):823-8.
256. Beutler E. The HFE Cys282Tyr mutation as a necessary but not sufficient cause of clinical hereditary hemochromatosis. *Blood.* 1 mai 2003;101(9):3347-50.
257. Rochette J, Le Gac G, Lassoued K, Férec C, Robson KJH. Factors influencing disease phenotype and penetrance in HFE haemochromatosis. *Hum Genet.* sept 2010;128(3):233-48.
258. Weiss G. Genetic mechanisms and modifying factors in hereditary hemochromatosis. *Nat Rev Gastroenterol Hepatol.* janv 2010;7(1):50-8.
259. McCune CA, Ravine D, Carter K, Jackson HA, Hutton D, Hedderich J, et al. Iron loading and morbidity among relatives of HFE C282Y homozygotes identified either by population genetic testing or presenting as patients. *Gut.* avr 2006;55(4):554-62.
260. Whitlock EP, Garlitz BA, Harris EL, Beil TL, Smith PR. Screening for hereditary hemochromatosis: a systematic review for the U.S. Preventive Services Task Force. *Ann Intern Med.* 1 août 2006;145(3):209-23.
261. Rossi E, Olynyk JK, Jeffrey GP. Clinical penetrance of C282Y homozygous HFE hemochromatosis. *Expert Rev Hematol.* déc 2008;1(2):205-16.
262. Milet J, Dehais V, Bourgain C, Jouanolle AM, Mosser A, Perrin M, et al. Common variants in the BMP2, BMP4, and HJV genes of the hepcidin regulation pathway modulate HFE hemochromatosis penetrance. *Am J Hum Genet.* oct 2007;81(4):799-807.
263. Valenti L, Fracanzani AL, Rametta R, Fraquelli M, Soverini G, Pelusi S, et al. Effect of the A736V TMPRSS6 polymorphism on the penetrance and clinical expression of hereditary hemochromatosis. *J Hepatol.* déc 2012;57(6):1319-25.
264. Bell CJ, Dinwiddie DL, Miller NA, Hateley SL, Ganusova EE, Mudge J, et al. Carrier Testing for Severe Childhood Recessive Diseases by Next-Generation Sequencing. *Sci Transl Med.* 12 janv 2011;3(65):65ra4.
265. Cooper DN, Krawczak M, Polychronakos C, Tyler-Smith C, Kehrer-Sawatzki H. Where genotype is not predictive of phenotype: towards an understanding of the molecular basis of reduced penetrance in human inherited disease. *Hum Genet.* 2013;132(10):1077-130.
266. Deltas C. Digenic inheritance and genetic modifiers. *Clin Genet.* mars 2018;93(3):429-38.
267. Schäffer AA. Digenic inheritance in medical genetics. *J Med Genet.* oct 2013;50(10):641-52.

268. Emond MJ, Louie T, Emerson J, Zhao W, Mathias RA, Knowles MR, et al. Exome sequencing of extreme phenotypes identifies DCTN4 as a modifier of chronic *Pseudomonas aeruginosa* infection in cystic fibrosis. *Nat Genet.* 8 juill 2012;44(8):886-9.
269. Timberlake AT, Choi J, Zaidi S, Lu Q, Nelson-Williams C, Brooks ED, et al. Two locus inheritance of non-syndromic midline craniosynostosis via rare SMAD6 and common BMP2 alleles. *Elife.* 8 sept 2016;5.
270. Lemmers RJLF, Tawil R, Petek LM, Balog J, Block GJ, Santen GWE, et al. Digenic inheritance of an SMCHD1 mutation and an FSHD-permissive D4Z4 allele causes facioscapulohumeral muscular dystrophy type 2. *Nat Genet.* déc 2012;44(12):1370-4.
271. Justice CM, Yagnik G, Kim Y, Peter I, Jabs EW, Erazo M, et al. A genome-wide association study identifies susceptibility loci for nonsyndromic sagittal craniosynostosis near BMP2 and within BBS9. *Nat Genet.* déc 2012;44(12):1360-4.
272. Mencarelli MA, Heidet L, Storey H, van Geel M, Knebelmann B, Fallerini C, et al. Evidence of digenic inheritance in Alport syndrome. *J Med Genet.* mars 2015;52(3):163-74.
273. Westenskow Peter, Splawski Igor, Timothy Katherine W., Keating Mark T., Sanguinetti Michael C. Compound Mutations. *Circulation.* 20 avr 2004;109(15):1834-41.
274. Botstein D, Risch N. Discovering genotypes underlying human phenotypes: past successes for mendelian disease, future approaches for complex disease. *Nat Genet.* mars 2003;33 Suppl:228-37.
275. Majewski J, Schwartzenruber J, Lalonde E, Montpetit A, Jabado N. What can exome sequencing do for you? *J Med Genet.* sept 2011;48(9):580-9.
276. Lee S, Abecasis GR, Boehnke M, Lin X. Rare-variant association analysis: study designs and statistical tests. *Am J Hum Genet.* 3 juill 2014;95(1):5-23.
277. Ueki M, Cordell HJ. Improved statistics for genome-wide interaction analysis. *PLoS Genet.* 2012;8(4):e1002625.
278. Yadav P, Freitag-Wolf S, Lieb W, Dempfle A, Krawczak M. Allowing for population stratification in case-only studies of gene-environment interaction, using genomic control. *Hum Genet.* 1 oct 2015;134(10):1117-25.
279. Yadav P, Freitag-Wolf S, Lieb W, Krawczak M. The role of linkage disequilibrium in case-only studies of gene-environment interactions. *Hum Genet.* janv 2015;134(1):89-96.
280. Yang Q, Khoury MJ, Flanders WD. Sample size requirements in case-only designs to detect gene-environment interaction. *Am J Epidemiol.* 1 nov 1997;146(9):713-20.
281. Yang Q, Khoury MJ, Sun F, Flanders WD. Case-only design to measure gene-gene interaction. *Epidemiology.* mars 1999;10(2):167-70.
282. Albert PS, Ratnasinghe D, Tangrea J, Wacholder S. Limitations of the case-only design for identifying gene-environment interactions. *Am J Epidemiol.* 15 oct 2001;154(8):687-93.

283. Gauderman WJ. Sample size requirements for association studies of gene-gene interaction. *Am J Epidemiol.* 1 mars 2002;155(5):478-84.
284. Kraft P, Yen Y-C, Stram DO, Morrison J, Gauderman WJ. Exploiting gene-environment interaction to detect genetic associations. *Hum Hered.* 2007;63(2):111-9.
285. Nanney LB, Skeel A, Luan J, Polis S, Richmond A, Wang MH, et al. Proteolytic cleavage and activation of pro-macrophage-stimulating protein and upregulation of its receptor in tissue injury. *J Invest Dermatol.* oct 1998;111(4):573-81.
286. Sakamoto O, Iwama A, Amitani R, Takehara T, Yamaguchi N, Yamamoto T, et al. Role of macrophage-stimulating protein and its receptor, RON tyrosine kinase, in ciliary motility. *J Clin Invest.* 15 févr 1997;99(4):701-9.
287. Becker KL, Arts P, Jaeger M, Plantinga TS, Gilissen C, van Laarhoven A, et al. MST1R mutation as a genetic cause of Lady Windermere syndrome. *Eur Respir J.* 2017;49(1).
288. Wu MC, Lee S, Cai T, Li Y, Boehnke M, Lin X. Rare-variant association testing for sequencing data with the sequence kernel association test. *Am J Hum Genet.* 15 juill 2011;89(1):82-93.
289. 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. *McVean G, éditeur. eLife.* 29 oct 2013;2:e01123.
290. Ansari MA, Pedergnana V, Ip C, Magri A, Von Delft A, Bonsall D, et al. Genome-to-genome analysis highlights the impact of the human innate and adaptive immune systems on the hepatitis C virus. *Nat Genet.* mai 2017;49(5):666-73.
291. 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.* 21 févr 2006;103(8):2869-73.
292. Luo T, Comas I, Luo D, Lu B, Wu J, Wei L, et al. Southern East Asian origin and coexpansion of *Mycobacterium tuberculosis* Beijing family with Han Chinese. *Proc Natl Acad Sci USA.* 30 juin 2015;112(26):8136-41.
293. Karczewski KJ, Gauthier LD, Daly MJ. Technical artifact drives apparent deviation from Hardy-Weinberg equilibrium at CCR5-Δ32 and other variants in gnomAD. *bioRxiv.* 2 oct 2019;784157.
294. Maier R, Akbari A, Wei X, Patterson N, Nielsen R, Reich D. No statistical evidence for an effect of CCR5-Δ32 on lifespan in the UK Biobank cohort. *bioRxiv.* 2 oct 2019;787986.

Annexe : Articles issus du travail de thèse