



NOTE INTERDISCIPLINAIRE DE SYNTHÈSE

---

**Bases génétiques de l'infection par *Mycobacterium tuberculosis* et évaluation expérimentale des scores *in silico* de prédiction de l'impact des variants faux-sens chez l'homme**

Réalisée au sein du laboratoire de génétique humaine des maladies infectieuses,

du réseau doctoral de l'EHESP et

de l'École Doctorale ED 393 Pierre Louis de Santé Publique

Université Paris Cité

Directrice de thèse : Dr Aurélie Cobat

## 1. Résumé

La tuberculose (TB), causée par *Mycobacterium tuberculosis* (Mtb), est un problème majeur de santé publique mondiale, exacerbée par la co-infection avec le VIH et l'émergence de souches résistantes aux antibiotiques. Malgré les outils de diagnostic existants (TST et IGRA), la détection de l'infection latente reste imparfaite. La contribution des facteurs génétiques de l'hôte à la TB est bien établie, mais le phénotype d'infection tuberculeuse est moins étudié. Les études d'épidémiologie génétique permettent d'analyser l'association entre des variants génétiques et un phénotype d'intérêt par une approche cas-témoins. Toutefois, l'impact fonctionnel de la majorité des variants reste inconnu, car non testés expérimentalement. Pour cette raison, de nombreux outils bioinformatiques ont été développés (CADD, AlphaMissense, SIFT, etc.) afin de prioriser les variants potentiellement délétères. Bien que largement utilisés en recherche et en clinique, ces outils présentent des divergences de prédiction, et leur évaluation est souvent biaisée par des jeux de données non indépendants. Des validations expérimentales indépendantes restent donc essentielles pour garantir leur fiabilité et soutenir les décisions médicales.

Dans le cadre de cette thèse, j'ai réalisé une étude d'association pangénomique (GWAS) du phénotype d'infection tuberculeuse et identifié un variant génétique fortement associé avec la résistance à la conversion aux tests TST et IGRA et ce dans deux cohortes indépendantes. J'ai également réalisé une étude comparative d'outils de prédiction de sévérité des variants faux-sens, ce qui a permis d'identifier ceux plus adaptés à un contexte exploratoire de recherche (forte sensibilité) et ceux plus adaptés un contexte de clinique (forte spécificité).

Les objectifs de l'OMS dans le cadre du programme ambitieux ENDTB ne sont pas encore atteints et les objectifs de financement pour la recherche et pour l'accessibilité au traitement sont mis à mal par une diminution globale des investissements dans la lutte contre la TB.

L'étude des déterminants génétiques de la conversion aux tests TST et IGRA permet d'améliorer notre compréhension des mécanismes de pathogénèse de la TB, ainsi que de guider le design de de nouveaux tests de détection afin d'améliorer la détection de la tuberculose latente et ainsi mieux prévenir les cas de tuberculose maladie. L'utilisation de données issues d'études génétiques et de l'intelligence artificielle pour guider les politiques publiques pour l'identification des individus les plus à risque représente un potentiel considérable pour la santé publique. Cependant, cela nécessite une réflexion éthique commune et globale sur la protection de ces données personnelles sensibles.

## **2. Contexte**

### **2.1. La tuberculose, un problème de santé publique majeur**

La tuberculose (TB), causée par l'agent infectieux *Mycobacterium tuberculosis* (*Mtb*), est une menace de santé publique mondiale malgré l'existence de la vaccination par le BCG et les traitements antibiotiques. Son contrôle est freiné par la co-infection avec le VIH qui multiplie le risque de TB par 15 à 21 [1] et par l'émergence de souches multi- et ultrarésistantes. Ainsi, la TB fait partie des 10 principales causes de mortalité dans le monde et se situe au premier rang des causes de mortalité par un agent infectieux unique (devant le VIH/SIDA depuis 2007) (Figure 1). En 2022, on estimait à 10.6 millions le nombre de nouveaux cas de TB symptomatique dans le monde, dont 6.3% chez des personnes vivant avec le VIH (PVVIH), et le nombre de mort à 1.3 million [1].

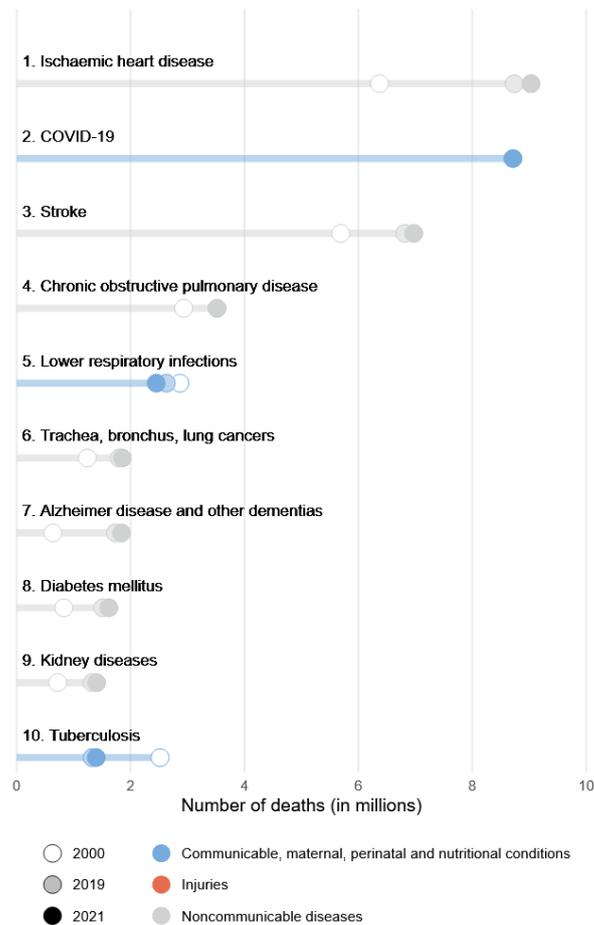


Figure 1 : Principales causes de décès dans le monde en 2021. la tuberculose se situe en dixième position.

La transmission de *Mtb* se fait par voie aérienne via les microgouttelettes expulsées par des personnes atteintes de la forme pulmonaire active. Parmi les personnes infectées, seule une minorité (environ 10 %) développera une tuberculose clinique, alors que la majorité contrôle l'infection sans symptôme. Pour ces derniers, on parle alors de tuberculose latente (LTB), détectée uniquement par des tests immunologiques. Si 90 à 95 % des personnes LTB ne développeront jamais la maladie, 5 à 10 % peuvent évoluer vers une tuberculose active tardive (PTB), souvent pulmonaire, due à une réactivation [2,3] ou à une nouvelle infection [4,5].

## **2.2. Diagnostic de l'infection par *Mycobacterium tuberculosis***

Le diagnostic de la TB pulmonaire repose sur plusieurs méthodes (radiographie, analyse des expectorations), mais il n'existe pas de test de référence permettant de détecter directement une infection tuberculeuse latente ni d'évaluer la viabilité des bacilles. Celle-ci est diagnostiquée par des tests immunologiques fondés sur la mémoire immunitaire : le test cutané à d'intradermoréaction (TST) et les tests in vitro de mesure de la production d'IFN- $\gamma$  (IGRA). Le TST, utilisé depuis 1907, consiste en une injection intradermique de tuberculine sur l'avant-bras, suivie de la mesure de l'induration après 48 à 72 heures. Le résultat dépend de la réponse immunitaire de l'hôte, mais la tuberculine n'étant pas spécifique à *Mtb*, le test peut être faussement positif chez les personnes vaccinées par le BCG ou exposées à d'autres mycobactéries. À l'inverse, il peut être faussement négatif chez les personnes immunodéprimées, notamment chez les PWH. Malgré ses limites, l'IDR est simple, peu coûteux et ne nécessite pas d'infrastructure de laboratoire. Les tests IGRA, plus récents, mesurent la production d'IFN- $\gamma$  par les lymphocytes T après exposition à des antigènes spécifiques de *Mtb*. Plus spécifiques que le TST, les IGRA présentent néanmoins des contraintes logistiques, notamment la nécessité d'un laboratoire et d'un traitement rapide des échantillons, ce qui limite leur utilisation dans les milieux à faibles ressources.

## **2.3. Prédispositions génétiques à la tuberculose et à la réponse aux tests IDR et IGRA**

La contribution des facteurs génétiques de l'hôte au développement de la TB est bien établie. Le syndrome de prédisposition mendélienne aux infections mycobactériennes (MSMD), une immunodéficiences rare liée à des mutations dans 22 gènes affectant la voie de l'IFN- $\gamma$ , en fournit une preuve de principe. Certaines de ces mutations ont aussi été identifiées chez des patients atteints uniquement de TB sévère, suggérant une base monogénique dans certains cas. Plus récemment, un variant du gène *TYK2*, à l'état homozygote, a été identifié comme première

cause fréquente de prédisposition monogénique à la TB, expliquant environ 1 % des cas en Europe. En revanche, l'étude génétique du phénotype d'infection (détection immunologique sans maladie) a longtemps été limitée, en partie en raison de la nature indirecte et quantitative des tests IDR et IGRA. Néanmoins, plusieurs études ont mis en évidence une forte héritabilité de la réponse immunitaire à *Mtb*. Par exemple, des héritabilités de 92 % (IDR, Chili), 71 % (IDR, Gambie), 58 % (IGRA, Afrique du Sud) ou 27 % (Ouganda) ont été rapportées, soulignant une influence génétique notable dans la susceptibilité à l'infection.

#### **2.4. Identification des variants génétiques associés aux maladies**

Les prédispositions génétiques aux maladies sont généralement étudiées en épidémiologie génétique sous deux angles : les prédispositions communes, généralement étudiées via des études d'association pangénomique, ou Genome-Wide Association Study (GWAS) [6] ; et les prédispositions rares, généralement étudiées via des méthodes d'agrégation des variants par gène pour pallier le manque de puissance statistique [7]. Les variations candidates causales peuvent être identifiées et priorisées selon leur impact fonctionnel sur la protéine. Cependant, l'impact fonctionnel de la majorité de ces variants est inconnue, et leur nombre important chez chaque individu ne permet pas de tester fonctionnellement l'impact de chacun d'entre eux. En effet, plus de 10 millions de variants faux-sens ont été identifiés chez l'humain [8,9]. Ainsi, des outils de prédictions ont été développés pour aider les chercheurs et cliniciens à identifier les variants génétiques les plus à même d'avoir un impact physiologique. Ces outils (CADD, AlphaMissense, Sift, PolyPhen2, etc.) sont largement utilisés pour prioriser les variants en vue d'investigations fonctionnelles approfondies [10] et sont couramment employés dans le cadre du diagnostic clinique [11]. Les recommandations de l'American College of Medical Genetics and Genomics (ACMG) et de l'Association for Molecular Pathology (AMP) préconisent l'utilisation simultanée de plusieurs outils et exigent une concordance totale entre les

prédictions afin de constituer une preuve à l'appui des décisions médicales [12]. Ces outils reposent sur des données variées (évolution, conservation, structure etc.), mais leurs prédictions divergent souvent. Leur évaluation est souvent biaisée par le chevauchement entre données d'entraînement et de test. Des évaluations indépendantes, avec des variants validés en laboratoire, sont donc cruciales pour juger leur fiabilité [13].

### **3. Objectifs de la thèse et principaux résultats**

Dans le cadre de cette thèse, j'ai abordé la question de l'identification des bases génétiques des maladies infectieuses sous deux angles complémentaires : d'une part, l'étude des déterminants génétiques de la résistance à la conversion au TST et aux IGRA, et d'autre part, l'évaluation des performances des outils de prédiction de la sévérité des variants faux-sens.

Le premier objectif consistait à étudier les facteurs génétiques impliqués dans le phénotype de résistance à la conversion aux TST et IGRA via un séquençage pangénomique (WGS) chez des individus présentant un phénotype de résistance extrême. Deux cohortes de personnes vivant avec le VIH (PWH), fortement exposées à la tuberculose, ont été recrutées en Afrique du Sud et en Haïti, deux régions où la TB est hyperendémique. Ces individus avaient un antécédant de faible taux de cellules T CD4+, les rendant extrêmement susceptible à l'infection par *Mtb* et à la TB. Leur taux a ensuite rétabli par thérapie antirétrovirale au moment de l'étude. À partir des données de WGS, nous avons défini un ensemble complet de variants candidats, utilisés pour mener une GWAS du phénotype de résistance, en comparant les individus restant durablement négatifs au TST et à l'IGRA à des témoins présentant une positivité aux deux tests ou une TB. Dans le cadre du premier objectif, nous avons identifié un variant génétique situé dans la région du chromosome 12q15 fortement associé au phénotype de résistance dans les deux cohortes étudiées. Des analyses *in silico* ont montré que ce variant est localisé à proximité du gène *YEATS4* et qu'il est prédit augmenter la traduction de ce dernier en supprimant un élément

régulateur appelé « upstream open reading frame » (uORF). Nous avons confirmé l'effet de ce variant sur la traduction du gène par une étude fonctionnelle réalisée dans des cellules HEK293T. *YEATS4* a été montré comme jouant un rôle dans la différenciation des cellules lymphoïdes innées, un sous-type de leucocytes.

Le deuxième objectif consistait à évaluer la performance d'outils de prédiction à l'aide de 4 913 variants issus de 12 gènes ayant été testés fonctionnellement dans notre laboratoire ou dans des études publiées. Nous avons comparé les performances de 11 scores de prédiction de la sévérité largement utilisés pour classer ces variants. En utilisant la base de données HGMD [14] comme référence de variants probablement pathogènes, nous avons également analysé l'impact du mode de transmission et de la fréquence des variants sur les performances de prédiction. Dans le cadre du second objectif, nous avons observé que certains outils se distinguaient par une meilleure sensibilité (notamment CADD et MutationAssessor), tandis que d'autres présentaient une meilleure spécificité (comme AlphaMissense et PolyPhen2). L'utilisation de méta-prédicteurs, combinant plusieurs scores, permettait d'améliorer la performance globale. Les outils les plus performants variaient selon la base de données utilisée (qu'elle contienne des variants à effet délétère sur la protéine ou des variants cliniquement pathogènes) suggérant que certains algorithmes sont mieux adaptés à prédire une altération biochimique plutôt qu'un impact phénotypique. Enfin, nous avons montré que la performance de ces outils était fortement influencée par la fréquence des variants dans la population générale, reflet probable d'un biais de conception les amenant à considérer les variants fréquents comme neutres.

#### **4. Regard interdisciplinaire**

La TB touche de manière inégale les différentes régions du monde. En Afrique du Sud en 2023, l'un des pays les plus touchés par la TB, l'incidence s'élevait à 427 nouveaux cas pour 100 000 habitants [1]. Pour l'Union européenne en 2022, cet indicateur était de 8.0 pour 100

000 habitants [15]. Cela peut s'expliquer par différents facteurs épidémiologiques, dont une exacerbation du fardeau dû aux conflits, au changement climatique, et à la distribution des principaux facteurs de risque comme le diabète, l'infection à VIH, la malnutrition et la consommation de tabac et d'alcool. Particulièrement, la co-infection VIH-*Mtb* forment une synergie meurtrière où chacune accélère l'évolution de l'autre, ce qui fait de la tuberculose la première cause de décès parmi les PWH [1]. Les pays ayant une forte prévalence de l'infection à VIH ont donc un fardeau très élevé de la tuberculose, et ces pays se concentrent sur le continent Africain et en Asie du Sud-Est (Figure 2). Les pays où l'accessibilité au traitement et au suivi médical est rendue difficile à cause du manque de moyens ou d'infrastructure ont également un fardeau plus élevé. A cela s'ajoute le développement de souches multi et ultra résistantes (résistante au traitement de première et deuxième intention respectivement) du aux problèmes d'observance du traitement (prise d'antibiotique quotidienne pendant plusieurs mois), dont la pénibilité et la durée favorise les arrêts prématurés.

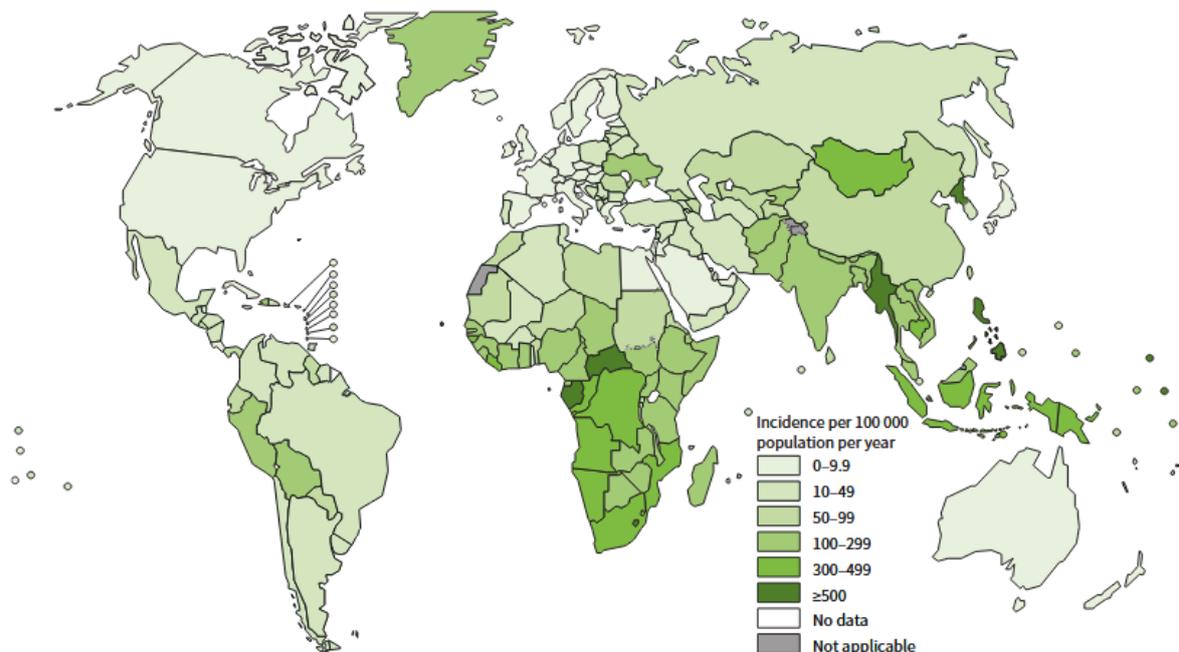


Figure 2: Incidence estimée de la TB dans le monde en 2023. World Health Organization. Global Tuberculosis Report 2024

L'organisation mondiale de la santé (OMS) a développé le programme ENDTB avec le but de réduire pour chaque pays de 80% l'incidence de la TB, de 90% les décès liés à la TB et de 100% les impacts catastrophiques pour les ménages affectés par la TB, et ce d'ici 2030 (Figure 3). Chaque pays doit adapter sa stratégie aux particularités locales et aux enjeux qui lui sont propres [16]. En 2023, l'Organisation des Nations Unies (ONU) a organisé une réunion de haut niveau sur le thème de la lutte contre la tuberculose. Cette réunion a permis de fixer des objectifs concernant la prévention, le diagnostic, le traitement et la prise en charge de la tuberculose. Parmi ces objectifs, on retrouve la préparation des systèmes de santé ; l'augmentation des financements liés à la prévention/diagnostic/traitement ; et la réduction des déficits de financement pour la recherche sur cette pathologie d'ici 2027 [17]. En 2022, le financement de la recherche et de l'innovation concernant la TB au niveau mondial s'élevait à 1 milliard USD, bien loin de l'objectif des 5 milliards USD par an, freinant ainsi les avancées scientifiques nécessaires pour accélérer la lutte contre la maladie [1]. Aussi, l'OMS estime qu'il faudra mobiliser 22 milliards de dollars par an pour assurer la prévention, le diagnostic, le traitement et la prise en charge de la tuberculose afin d'atteindre ces objectifs d'ici 2027 [1]. En 2023, 80 % des dépenses liées aux services de lutte contre la tuberculose provenaient de financements nationaux. Toutefois, dans les pays à revenu faible ou intermédiaire, l'appui des donateurs internationaux demeure indispensable. Entre 2019 et 2023, les financements domestiques ont chuté de 1,2 milliard USD, tandis que les contributions internationales n'ont augmenté que de 0,1 milliard USD. Le recul des financements nationaux s'explique en grande partie par une diminution des investissements dans les pays des BRICS (Brésil, Russie, Inde, Chine, Afrique du Sud). Les récentes politiques américaines sur le développement international et les coupures budgétaires appliquées aux organisations comme l'ONUSIDA et l'USAID vont rendre encore plus difficile l'atteinte de ces objectifs.

## Global TB milestones and targets: latest status<sup>a</sup> of progress

### End TB Strategy, 2025 milestones

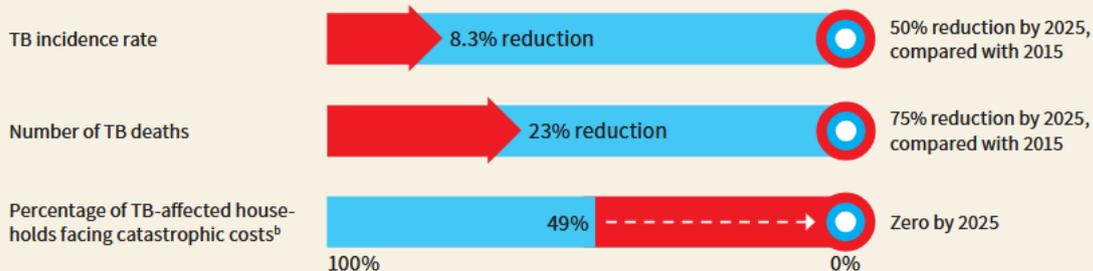


Figure 3: Accomplissements relatifs à l'incidence, la mortalité et la proportions de foyers grandement affectés dans le cadre du programme ENDTB, par rapport aux objectifs intermédiaires visés pour 2025 [1].

L'augmentation des investissements dans les pays les plus touchés par la TB permettra d'optimiser la prévention avec la mise en place de campagnes de vaccination, dépistage et traitement préventif des cas contacts.

L'apport de la recherche dans la lutte contre la TB est important pour plusieurs raisons. Les études épidémiologiques permettent d'identifier les facteurs de risques liés à l'hôte (génétiques, comportementaux, sociaux, co-infections etc.) et non liés à l'hôte (environnementaux, démographiques etc.) et d'identifier les populations les plus pertinentes à cibler lors des campagnes préventives et par les politiques de santé. Comprendre les interactions entre déterminants génétiques et l'environnement permettrait de mieux cibler la prévention dans des contextes professionnels ou environnementaux à risque (ex. travailleurs exposés, zones à forte prévalence etc.) en identifiant les personnes plus à risque. De plus, les études épidémiologiques visant à identifier les déterminants génétiques de la susceptibilité ou de la résistance à l'infection tuberculeuse jouent un rôle essentiel pour orienter la recherche biomédicale. En

mettant en lumière les mécanismes génétiques impliqués dans la réponse à l'infection, elles permettent de guider le développement de nouvelles cibles thérapeutiques, que ce soit pour la conception de vaccins ou pour la mise au point de traitements innovants contre la tuberculose. Aussi, l'identification des individus résistants à la conversion aux tests de dépistage TST et IGRA permettrait d'améliorer la puissance des essais clinique évaluant des vaccins candidats contre la TB en excluant ces participants moins susceptibles d'être informatifs pour mesurer l'efficacité du vaccin. De la même façon, l'identification des déterminants de la réponse aux TST et IGRA permettra de mieux guider le design de nouveaux tests de détection afin d'améliorer la détection de la tuberculose latente et ainsi mieux prévenir les cas de tuberculose maladie.

L'utilisation des données génétiques dans une optique de prévention et de développement de politiques publiques offre un potentiel considérable pour améliorer le ciblage des individus à risque. Toutefois, cette pratique soulève d'importantes questions éthiques. Il est indispensable qu'elle s'inscrive dans un cadre juridique solide, garantissant la protection des personnes contre toute utilisation de leurs données non consentie ou contraire à leurs intérêts, notamment à des fins commerciales. L'émergence récente des données génétiques à grande échelle s'accompagne d'un cadre législatif encore en construction, évoluant lentement et de manière hétérogène selon les pays. Cette situation appelle une vigilance particulière et une réflexion collective, impliquant l'ensemble des parties prenantes [18]. La masse d'informations générée par les données de séquençage génomique conduit à un recours croissant à l'intelligence artificielle pour leur interprétation. Comme l'a montré mon projet de thèse concernant les scores de prédiction, même les outils les plus sophistiqués et puissants présentent des limites. Une mauvaise compréhension de ces limites peut conduire à une surestimation de leurs capacités et à une confiance excessive dans leurs résultats. L'utilisation de ces outils pour la recherche, la clinique et la prise de décision politique soulève d'importants enjeux éthiques, notamment en

raison du manque de transparence des algorithmes, souvent perçus comme des « boîtes noires ». Cela amène une réflexion sur les protocoles de prise de décision, les risques d'erreur, la responsabilité en cas de préjudice, ainsi que sur la manipulation et la protection des données personnelles sensibles [19].

L'identification de déterminants génétiques associés à la conversion aux tests TST et IGRA, à la TB, ou à d'autres maladies, ne doit pas induire une vision déterministe de la santé ni minimiser l'importance des facteurs environnementaux et comportementaux. Si les déterminants génétiques jouent un rôle important, il reste essentiel que les messages de prévention continuent de promouvoir les comportements et réflexes protecteurs accessibles à tous, qui permettent de réduire activement le risque de contracter une maladie.

Enfin, les outils de prédiction peuvent être utilisés en clinique pour guider le processus de décision. Cependant, l'utilisation de ces outils dans le cadre des tests génétiques restent limités chez les patients d'ascendance africaine et de populations sous représentées dans la littérature scientifique. Le manque de données génomiques représentatives de ces populations et la diversité génétique propre aux populations africaines ont conduit à la création de bases de données cliniques et de recommandations pour l'interprétation des variants majoritairement biaisées en faveur des populations européennes [20]. Dans une étude récente, Zhou et al. montrent que les outils de prédiction de la sévérité des variants sont en moyenne moins sensibles lorsqu'ils sont appliqués à des variants oncogéniques spécifiques aux populations africaines, comparativement à ceux spécifiques aux populations européennes. Par ailleurs, ils soulignent qu'il existe environ deux fois plus de variants pathogènes spécifiques aux populations européennes référencés dans la base ClinVar, en comparaison des variants spécifiques aux populations africaines [21]. Sur les 48 pays de la région d'Afrique Subsaharienne, seulement 15 sont représentés dans des études génétiques publiées, la grande

majorité de ces travaux étant limitée aux gènes bien connus du cancer du sein, *BRCA1* et *BRCA2* [22]. Ce manque de données concernant les populations africaines a pour conséquence un risque majoré d'erreur de diagnostic pour cette population [23] et des conséquences cliniques notables [24]. Pour des raisons éthiques et scientifiques, il est important que la communauté génétique mondiale surmonte les obstacles ayant jusqu'à présent freiné les avancées des études génétiques en Afrique, tout en assurant une inclusion réelle des chercheurs africains. Une recherche plus inclusive et un renforcement des capacités locales offrent en effet des opportunités majeures pour accélérer les découvertes génomiques et favoriser leur traduction en bénéfices concrets pour l'ensemble des populations [20].

## 5. Références

1. World Health Organization. Global Tuberculosis Report 2024. 2024.
2. Abel L, El-Baghdadi J, Bousfiha AA, Casanova J-L, Schurr E. Human genetics of tuberculosis: a long and winding road. *Phil Trans R Soc B*. 2014;369:20130428.
3. Alcaïs A, Fieschi C, Abel L, Casanova J-L. Tuberculosis in children and adults. *J Exp Med*. 2005;202:1617–21.
4. Cardona P-J. Reactivation or reinfection in adult tuberculosis: Is that the question? *International Journal of Mycobacteriology*. 2016;5:400–7.
5. Behr MA, Edelstein PH, Ramakrishnan L. Is *Mycobacterium tuberculosis* infection life long? *BMJ*. 2019;l5770.
6. Uffelmann E, Huang QQ, Munung NS, De Vries J, Okada Y, Martin AR, et al. Genome-wide association studies. *Nat Rev Methods Primers* [Internet]. 2021 [cited 2025 Jul 28];1. Available from: <https://www.nature.com/articles/s43586-021-00056-9>
7. Lee S, Abecasis GR, Boehnke M, Lin X. Rare-variant association analysis: study designs and statistical tests. *Am J Hum Genet*. 2014;95:5–23.
8. Sun KY, Bai X, Chen S, Bao S, Zhang C, Kapoor M, et al. A deep catalogue of protein-coding variation in 983,578 individuals. *Nature*. 2024;631:583–92.
9. Genome Aggregation Database Consortium, Karczewski KJ, Francioli LC, Tiao G, Cummings BB, Alföldi J, et al. The mutational constraint spectrum quantified from variation in 141,456 humans. *Nature*. 2020;581:434–43.

10. Domené S, Scaglia PA, Gutiérrez ML, Domené HM. Applying Bioinformatic Platforms, In Vitro, and In Vivo Functional Assays in the Characterization of Genetic Variants in the GH/IGF Pathway Affecting Growth and Development. *Cells*. 2021;10:2063.
11. Ernst C, Hahnen E, Engel C, Nothnagel M, Weber J, Schmutzler RK, et al. Performance of in silico prediction tools for the classification of rare BRCA1/2 missense variants in clinical diagnostics. *BMC Med Genomics*. 2018;11:35.
12. Richards S, Aziz N, Bale S, Bick D, Das S, Gastier-Foster J, et al. Standards and guidelines for the interpretation of sequence variants: a joint consensus recommendation of the American College of Medical Genetics and Genomics and the Association for Molecular Pathology. *Genetics in Medicine*. 2015;17:405–24.
13. Grimm DG, Azencott C-A, Aicheler F, Gieraths U, MacArthur DG, Samocha KE, et al. The evaluation of tools used to predict the impact of missense variants is hindered by two types of circularity. *Hum Mutat*. 2015;36:513–23.
14. Stenson PD, Mort M, Ball EV, Chapman M, Evans K, Azevedo L, et al. The Human Gene Mutation Database (HGMD®): optimizing its use in a clinical diagnostic or research setting. *Hum Genet*. 2020;139:1197–207.
15. ECDC. Tuberculosis : Annual Epidemiological Report for 2022. 2024.
16. World Health Organization. The end TB Strategy, A world free of TB. Zero deaths, disease and suffering due to TB. 2015.
17. Organisation des Nations Unies. L'ONU et les Etats membres s'engagent à mettre fin à la tuberculose [Internet]. Available from: <https://news.un.org/fr/story/2023/09/1138962>
18. Rial-Sebbag E. Données génétiques massives : quelles règles juridiques et éthiques pour leur usage en santé publique ? *Actualité et dossier en santé publique*. 2020;
19. Coghlan S, Gyngell C, Vears DF. Ethics of artificial intelligence in prenatal and pediatric genomic medicine. *J Community Genet*. 2024;15:13–24.
20. Martin AR, Teferra S, Möller M, Hoal EG, Daly MJ. The critical needs and challenges for genetic architecture studies in Africa. *Curr Opin Genet Dev*. 2018;53:113–20.
21. Zhou K, Gheybi K, Soh PXY, Hayes VM. Evaluating variant pathogenicity prediction tools to establish African inclusive guidelines for germline genetic testing. *Commun Med*. 2025;5:157.
22. Shain J, Michel A, May MS, Qunaj L, El-Sadr W, Chung WK, et al. Cancer genetic mutation prevalence in sub-Saharan Africa: A review of existing data. *Semin Oncol*. 2023;50:123–30.
23. Manrai AK, Funke BH, Rehm HL, Olesen MS, Maron BA, Szolovits P, et al. Genetic Misdiagnoses and the Potential for Health Disparities. *N Engl J Med*. 2016;375:655–65.

24. Petrovski S, Goldstein DB. Unequal representation of genetic variation across ancestry groups creates healthcare inequality in the application of precision medicine. *Genome Biol.* 2016;17:157.