

Titre : Etude populationnelle des gènes *HLA* de 4 ethnies d'Afrique de l'Ouest (Bénin) et leur impact dans l'imputation *HLA* des populations africaines

Encadrant : Nicolas Vince (nicolas.vince@univ-nantes.fr), Laure Gineau (laure.gineau@ird.fr)

Lieu du stage : équipe 3 ITHINK, CR2TI, Nantes, France

Objectif :

L'objectif du stage est l'étude populationnelle des gènes *HLA* (classiques et non-classiques) de 4 ethnies (n ~ 500 individus) du Bénin à partir de données de séquençages et de génotypage. Ce jeu de données est exceptionnel du fait de la rareté des études génétiques chez les populations africaines. De plus, le but de cette analyse est d'améliorer l'imputation *HLA* des populations d'origine africaine.

Résumé :

Le système *HLA* (Human Leukocyte Antigen) est au centre de la réponse immunitaire et joue un rôle majeur dans la transplantation et les maladies immunologiques telles que les maladies infectieuses et l'auto-immunité.

Les populations d'origine non-européenne restent sous étudiées en génétique. La description en population des gènes *HLA* des populations africaines est manquante alors que la diversité y est très importante et les enjeux de santé publique conséquents. La connaissance de cette diversité permettra de mieux appréhender ces populations au niveau *HLA* et ainsi d'apporter une meilleure compréhension lors des études d'association *HLA* avec différentes pathologies. Cette partie fera ainsi appel à des outils classiques en génétique des populations.

Depuis une dizaine d'années, une ère génomique a débuté et la disponibilité de données SNP (Single Nucleotide Polymorphism) sur l'ensemble du génome a augmenté de manière très importante. Malgré les progrès biotechnologiques, la caractérisation de la molécule *HLA* reste complexe expérimentalement, longue et onéreuse, et des alternatives basées sur l'inférence statistique ont été récemment développées, c'est l'imputation *HLA*. Celle-ci permet d'inférer (imputer) le génotype *HLA* d'individus à partir des données SNP simples en s'appuyant sur des panels de référence (modèles). La qualité de l'imputation du *HLA* est très dépendante du panel de référence : plus celui-ci est grand et en adéquation avec le type de données à imputer (par exemple, même origine ancestrale), plus la qualité de l'imputation est bonne. Il est donc indispensable de pouvoir disposer de jeux de données en grand nombre et variés afin de réaliser la meilleure imputation possible.

Afin d'accompagner ce nouvel élan, nous avons mis en place le SNP-HLA Reference Consortium (SHLARC), un projet international collaboratif visant à (1) augmenter le nombre et la diversité des données de référence, (2) améliorer les méthodes mathématiques d'imputation HLA, et (3) rendre ce service accessible à l'ensemble de la communauté scientifique.

La personne recrutée devra déterminer la composition et la fréquence des différents allèles présents dans ces populations béninoises afin d'en déterminer les spécificités et ressemblance avec d'autres populations mondiales. Un second objectif sera d'évaluer les performances de l'imputation *HLA* chez cette population. De nouveaux panels de référence basés sur des jeux de données (HLA+SNPs) seront à réaliser. Enfin, à l'aide des panels de référence nouvellement créés, il faudra s'assurer de la qualité de l'imputation *HLA*.

L'étudiant devra ainsi prendre en main plusieurs outils (package R HIBAG, Plink, serveur de calcul avec GPU) et des données génétiques diverses (séquençage *HLA*, génotypes SNPs). Les outils d'inférence statistique ainsi que différents algorithmes de gestion des données ont déjà été implémentés par l'équipe. Le stagiaire sera hébergé au sein de l'équipe 3 du CR2TI à la Nantes Université (site du CHU de Nantes, Hôtel Dieu) sous la direction de Nicolas Vince et sera co-encadré par Laure Gineau, chercheur de l'UMR MERIT situé à Paris.