

Proposition de stage de master 2 – Janvier à Juin 2025

Analyse des microsatellites en population générale française, à partir des données de séquençage de génome du projet PopGen

Encadrant :

Dr Kevin Uguen (kevin.uguen@univ-brest.fr)

Praticien Hospitalier Universitaire

Unité Inserm U1078 – GGB (Université de Bretagne Occidentale)



Equipe d'accueil :

Unité Inserm U1078- GGB, Equipe BIGG

Université de Bretagne Occidentale, Brest

Faculté de Médecine – Bâtiment IBRBS, 22 Avenue Camille Desmoulin, 29238 Brest

Contexte de l'étude : Les short-tandem repeats (STR), aussi appelés microsatellites, sont des motifs répétés de 1 à 9 paires de bases (pb), qui représentent environ 3 % de la longueur totale de notre ADN génomique. Ils représentent (avec les minisatellites) les loci les plus variables du génome humain affectant principalement le nombre de répétitions de ces séquences. Bien que la grande majorité des STR soit située hors des séquences codantes du génome, près de 17 % des gènes humains, y compris un large ensemble de gènes ayant une fonction primordiale au cours du développement, contiennent des STR dans leurs séquences codantes ou leurs régions régulatrices. Depuis les années 90, et la description du syndrome de l'X fragile, près de 60 pathologies, principalement neurologiques, ont été associées à des expansions de nucléotides. Ces découvertes ont notamment été possibles par l'arrivée des techniques de Next Generation Sequencing (NGS), et la mise en place d'algorithmes de détection de ces expansions de nucléotides.

L'enjeu majeur actuel est, à l'instar des variations nucléotidiques, du nombre de copies, et de structure, de disposer de bases de données de populations pour permettre d'utiliser ces outils de manière efficace et d'interpréter les résultats dans un contexte clinique. Dans le cadre du Plan France Médecine Génomique 2025 (PFMG2025), le projet POPGEN a pour objectif de réaliser une cartographie exhaustive de la variabilité génomique au sein de la population de France Métropolitaine (<https://lysine.univ-brest.fr/popgen/>). Pour ce faire, le projet prévoit la construction d'une base de données répertoriant les variations génomiques et leurs fréquences sur le territoire national. A ce jour, le séquençage de génome de près de 4000 individus sains de la cohorte CONSTANCES est disponible pour ce projet.

Objectif du projet : Au cours de ce stage, l'objectif premier sera la mise en place d'un pipeline bio-informatique permettant l'analyse des STR chez les individus sains séquencés dans le cadre du projet POPGEN. Secondairement, il s'agira d'analyser des données de séquençage de ces régions, ainsi que de débiter la construction de la base de données devant centraliser tous les résultats issus de cette étude.

Profil recherché : Etudiant en master 2 en bio-informatique.

Techniques/méthodes pouvant être mises en œuvre : ExpansionHunter, TandemRepeatFinder, Snakemake, Python, R.