



## L'intégration de la grande diversité génétique de la population humaine grâce au pangénome

### Résumé

Le consortium du pangénome humain de référence (*Human Pangenome Reference Consortium*) présente ici sa première version [1]. Ce pangénome contient 47 assemblages diploïdes phasés provenant d'une cohorte d'individus d'origines génétiques diverses. Ces assemblages couvrent plus de 99 % de la séquence attendue dans chaque génome et sont précis à plus de 99 % au niveau de la structure et des paires de bases. Sur la base des alignements des assemblages, le pangénome capture les variants et haplotypes déjà connus, mais révèle aussi de nouveaux allèles à des loci structurellement complexes. Par rapport au génome de référence GRCh38, 119 millions de paires de bases de séquences polymorphes euchromatiques et 1115 duplications de gènes ont été ajoutées. Environ 90 millions de paires de bases supplémentaires proviennent de variations structurelles. L'utilisation du projet de pangénome pour analyser les données à lecture courte a permis de réduire les erreurs de découverte de petits variants de 34 % et d'augmenter le nombre de variants structurels détectés par haplotype de 104 % par rapport au génome de référence GRCh38, ce qui a conduit à la caractérisation de la grande majorité des allèles de variants structurels par échantillon.

### Commentaire

Le *Human Pangenome Reference Consortium* (HPRC) est un projet fondé par le *National Human Genome Research Institute* (NHGRI). Il a pour but de générer un « pangénome » regroupant 350 génomes de populations différentes afin de prendre en compte la diversité génétique de toutes les populations.



© L. Benarroch

Sorbonne Université, Inserm, Centre de recherche en Myologie, Institut de Myologie, Paris, France

[louise.benarroch@inserm.fr](mailto:louise.benarroch@inserm.fr)

Le premier génome complet a été publié en 2013. Il avait été généré à partir du séquençage d'un seul individu, puis il a été mis à jour continuellement au cours des vingt dernières années. Cet assemblage contient de nombreuses erreurs et reste incomplet malgré les nombreuses mises à jour, générant des biais dans les analyses génétiques sous-jacentes. Plus récemment, l'assemblage T2T (*Telomere-to-telomere*, T2T-CHM13v2.0) a permis d'améliorer les analyses génétiques, en apportant la séquence complète de tous les autosomes et du chromosome X. Mais cet assemblage ne représente toujours pas la diversité génétique de la population humaine.

En 2023, Liao *et al.* [1] ont publié la première version du pangénome humain, générée à partir de l'assemblage de 47 génomes de 13 populations différentes. Basé sur le séquençage à « longues lectures » (*long read*), cet assemblage permet une cartographie plus précise du génome humain, en donnant accès à des régions complexes comme celles hautement répétées ou encore les variants structuraux polymorphes. Cette première version est une avancée majeure dans le monde de la génomique. L'aboutissement de ce projet promet de grandes découvertes pour la prévention, le diagnostic et le traitement des maladies génétiques. ♦

**The pangenome integrates the vast genetic diversity of the human population**

### RÉFÉRENCE

1. Liao WW, Asri M, Ebler J, *et al.* A draft human pangenome reference. *Nature*. 2023 ; 617 (7960) : 312-324.

### LIENS D'INTÉRÊTS

L'auteure déclare n'avoir aucun lien d'intérêt concernant les données publiées dans cet article.

### TIRÉS À PART

L. Benarroch