



Swiss Institute of
Bioinformatics

COMMUNIQUÉ DE PRESSE

Reconstruire des génomes humains à une fraction du coût actuel promet de doper la recherche biomédicale

Lausanne, 13 janvier 2021 – Des milliers de marqueurs génétiques ont déjà été associés à des traits humains complexes tels que la maladie d'Alzheimer, le cancer, l'obésité ou la taille. Pour découvrir ces associations, les chercheurs doivent comparer le génome de nombreux individus à des millions de sites – ou marqueurs – génétiques, ce qui nécessite des technologies de génotypage efficaces et bon marché. Une nouvelle méthode statistique, développée par le groupe d'Olivier Delaneau au SIB Institut Suisse de Bioinformatique et à l'Université de Lausanne (UNIL), offre des possibilités qui changent la donne. Pour moins d'un dollar de coût de calcul, GLIMPSE est en mesure de reconstruire statistiquement un génome humain complet à partir d'une très petite quantité de données. La méthode offre une première alternative réaliste aux approches actuelles, basées sur un ensemble prédéfini de marqueurs génétiques, et permet ainsi d'inclure plus largement des populations sous-représentées. L'étude, qui suggère un changement de paradigme pour la production de données dans la recherche biomédicale, est publiée dans la revue Nature Genetics.

Une approche efficace et bon marché pour examiner les marqueurs génétiques

Le séquençage du génome entier à faible couverture (LC-WGS) est une méthode consistant à reconstruire statistiquement le génome à partir d'un très faible effort de séquençage. Cette méthode a été proposée comme alternative aux technologies actuelles, reposant sur des ensembles de marqueurs prédéfinis (voir encadré ci-dessous), car elle est moins biaisée et plus puissante. Cependant son coût de calcul élevé a jusqu'ici freiné son adoption comme alternative de choix. [L'équipe de chercheurs dirigée par Olivier Delaneau](#), chef de groupe au SIB et à l'UNIL, a mis au point un logiciel appelé GLIMPSE, qui remédie à cet obstacle majeur. «GLIMPSE permet une approche 10 à 1000 fois plus rapide, et donc moins chère, que d'autres méthodes LC-WGS, tout en étant bien plus précis pour les marqueurs génétiques rares», explique Olivier Delaneau. «GLIMPSE est capable d'améliorer considérablement la qualité d'un génome généré par une technique de séquençage à faible couverture pour moins d'un dollar de frais de calcul, ce qui en fait la première alternative réaliste aux approches actuelles».

De données non biaisées à des soins de santé plus inclusifs

Jusqu'à présent, les études d'association pangénomiques ont surtout porté sur des Européens: 80 % des participants à ces études sont des individus d'origine européenne alors que ceux-ci ne représentent que 16 % de la population mondiale. C'est une question éthique importante en termes de soins de santé inclusifs et d'accès équitable aux avancées de la recherche biomédicale, car la manière dont les marqueurs génétiques contribuent à la prédisposition aux maladies varie selon les populations humaines. Le LC-WGS contourne naturellement le biais inhérent aux groupes prédéfinis de marqueurs génétiques. Il peut donc être appliqué avec succès à des populations sous-représentées, comme l'équipe de chercheurs le montre en guise de 'preuve de concept' sur une population afro-américaine. «En plus de briser l'obstacle économique permettant des études d'association pangénomiques basées sur le LC-WGS, ce qui est vraiment intéressant dans cette approche est cette possibilité de découvrir des associations dans des populations sous-étudiées»

indique Simone Rubinacci, chercheur postdoctoral du groupe de Delaneau et premier auteur de l'article.

Tirer parti de génomes déjà séquencés

«Notre réflexion initiale était: pouvons-nous utiliser la profusion de génomes déjà séquencés pour améliorer les suivants? En d'autres termes, obtenir plus pour moins: c'est exactement ce que fait GLIMPSE,» explique Diogo Ribeiro, chercheur postdoctoral du groupe de Delaneau et co-auteur de l'article. Comment cela fonctionne-t-il? En s'appuyant sur l'idée que nous avons tous des ancêtres communs relativement récents dont nous avons hérité de petites portions d'ADN. En bref, GLIMPSE exploite de larges collections de génomes humains qui ont été séquencés avec une grande précision (séquençage de génomes entiers à large couverture) pour identifier des portions d'ADN partagées avec des génomes récemment séquencés. De cette façon, GLIMPSE peut combler de manière fiable les lacunes des données de faible couverture.

Un nouveau paradigme pour les futures études génomiques avec des applications très variées

Disponible dans une suite d'outils open-source, GLIMPSE ouvre la voie à une adoption étendue du séquençage du génome entier à faible couverture, en promouvant un changement de paradigme dans la génération de données pour les futures études génomiques. Depuis la première version du logiciel sous forme de préprint en avril 2020, des recherches ont déjà été lancées à l'aide de cet outil, par exemple pour reconstruire le génome de personnes ayant vécu il y a des milliers d'années à partir d'ADN ancien ou pour reconstituer l'ADN de patients COVID-19 à partir de prélèvements nasopharyngés dans le cadre d'une étude d'association pangénomique.

Études d'association génétique

Aujourd'hui pour trouver des associations entre des marqueurs génétiques et des traits complexes, les chercheurs étudient les données croisées des dossiers médicaux et données génétiques de milliers de participants (études d'association pangénomiques). Pour ce faire, une technologie communément utilisée repose sur des ensembles prédéfinis de marqueurs génétiques appelés SNP (pour polymorphismes nucléotidiques simples). Il s'agit d'éléments constitutifs individuels de l'ADN dans le génome qui varient d'une personne à l'autre. Toutefois, cette technique, bien que relativement rapide et peu coûteuse, présente des inconvénients majeurs puisque des variants, nouveaux ou rares, tels que ceux présents dans des populations sous-étudiées (lire plus haut), peuvent passer inaperçus.

RÉFÉRENCE

Rubinacci *et al.*, Efficient phasing and imputation of low-coverage sequencing data using large reference panels, Nat. Genet. 2021. DOI : [10.1038/s41588-020-00756-0](https://doi.org/10.1038/s41588-020-00756-0)

À PROPOS DU SIB INSTITUT SUISSE DE BIOINFORMATIQUE

Le [SIB](#) est une organisation à but non lucratif reconnue au niveau international et spécialisée en science des données biologiques et biomédicales. Ses scientifiques traduisent les données en savoirs et se passionnent pour les questions complexes dans de nombreux domaines des sciences du vivant, de l'écologie à la médecine. Ils fournissent des bases de données et des logiciels essentiels à la recherche ainsi qu'une large palette d'expertises en bioinformatique utiles aux institutions publiques et privées. Le SIB fédère la communauté bioinformatique suisse, qui compte quelque 800 scientifiques, en encourageant les collaborations et le partage de connaissances. L'Institut contribue ainsi à maintenir la Suisse à la pointe de l'innovation en favorisant les avancées de la recherche biologique et du domaine de la santé.



CONTACTS MÉDIA

Olivier Delaneau – Chef de groupe – olivier.delaneau@unil.ch – 021 692 40 55

Maia Berman – Communication manager – maia.berman@sib.swiss – 021 692 40 54

MATÉRIEL POUR LES MÉDIAS

[Photo](#) des auteurs, de gauche à droite: Robin Hofmeister, Diogo Ribeiro, Simone Rubinacci et Olivier Delaneau.