



UNIVERSITÉ
DE GENÈVE

Unil

UNIL | Université de Lausanne



Swiss Institute of
Bioinformatics

SHAPEIT4 : un algorithme de nouvelle génération pour la génomique à grande échelle

Des chercheurs de l'UNIL, de l'UNIGE et du SIB mettent à la disposition de tous un outil informatique extrêmement puissant pour faciliter l'interprétation des Big Data du génome.

COMMUNIQUÉ DE PRESSE

Genève | 19 décembre 2019

Les haplotypes constituent un ensemble de variations génétiques qui, situés côte à côte sur un même chromosome, sont transmis en un seul groupe à la génération suivante. Leur examen permet de comprendre l'héritabilité de certains traits complexes, comme par exemple le risque de développer plus tard une maladie. Cependant, pour effectuer cette analyse, il faut généralement disposer du génome des membres d'une même famille (les parents et leur enfant), un procédé long et cher. Pour contourner ce problème, des chercheurs des universités de Genève (UNIGE) et de Lausanne (UNIL), ainsi qu'au SIB Institut Suisse de Bioinformatique, ont mis au point SHAPEIT4, un puissant algorithme informatique permettant d'identifier très rapidement les haplotypes de centaines de milliers d'individus sans lien familiaux, avec un résultat aussi fin que dans le cadre d'analyses familiales impossibles à mener à si large échelle. Leur outil est maintenant disponible en ligne sous licence open source, à la disposition libre de l'ensemble de la communauté des chercheurs. A découvrir dans la revue *Nature Communications*.

A l'heure actuelle, l'analyse des données génétiques prend de plus en plus d'importance, notamment dans le champ de la médecine personnalisée. Preuve en est : le nombre de génomes humains séquencés chaque année croît de manière exponentielle et les grandes bases de données comptent maintenant près d'un million d'individus. Cette masse de données est d'une extrême richesse pour mieux comprendre le destin génétique de l'humanité, qu'il s'agisse de déterminer le poids génétique dans telle ou telle maladie ou de mieux comprendre l'histoire des migrations humaines. Pour faire sens, cependant, ces données de masse doivent être traitées informatiquement. «Or, la puissance de calcul des ordinateurs reste relativement stable, contrairement à la croissance ultra-rapide de ces Big Data du génome», souligne Olivier Delaneau, professeur boursier FNS au Département de biologie computationnelle de la Faculté de biologie et médecine de l'UNIL et au SIB, qui a dirigé ces travaux. «Notre algorithme vise ainsi à optimiser le traitement des données génétiques de façon à pouvoir absorber cette quantité d'information et les rendre exploitables par les scientifiques, malgré le décalage entre leur quantité et la puissance comparativement limitée des ordinateurs.»

Mieux comprendre le rôle des haplotypes

Le génotypage permet de connaître les allèles d'un individu, c'est-à-dire les variations génétiques reçues de ses parents. Par contre, sans connaître le génome parental, on ignore quels allèles sont simultanément transmis aux enfants, et selon quelles combinaisons. «Cette information – les haplotypes – est cruciale si l'on veut réellement comprendre les bases génétiques de la variation humaine, explique Emmanouil Dermitzakis, professeur à la Faculté de médecine de l'UNIGE et au SIB, qui a co-supervisé ces travaux. Et ce, que ce soit



© UNIL

Olivier Delaneau, professeur boursier FNS au Département de biologie computationnelle de la Faculté de biologie et médecine de l'UNIL et au SIB.

Illustrations haute définition



Emmanouil Dermitzakis, professeur à la Faculté de médecine de l'UNIGE et au SIB.

en génétique des populations ou dans l'optique de la médecine de précision.»

Pour déterminer le risque génétique de maladie, par exemple, les scientifiques évaluent si une variation génétique est plus ou moins présente chez les individus ayant développé cette maladie, afin de déterminer le rôle de cette variation dans la maladie étudiée. «En connaissant les haplotypes, nous conduisons le même type d'analyse, souligne Emmanouil Dermitzakis. Par contre, nous passons d'un seul variant à une combinaison de nombreux variants, ce qui permet de déterminer quelles combinaisons alléliques sur un même chromosome ont le plus d'impact sur le risque de maladie. C'est beaucoup plus précis!»

La méthode développée par les chercheurs permet de traiter un nombre de génomes extrêmement important, de l'ordre de 500 000 à 1 000 000 d'individus, et de déterminer leurs haplotypes sans connaître leur ascendance ni leur descendance, tout en utilisant une puissance de calcul standard. L'outil **SHAPEIT4** a été testé avec succès sur les 500 000 génomes individuels que compte la UK Biobank, une banque de données scientifiques développée au Royaume-Uni. «Nous avons ici un exemple typique de ce que sont les Big Data, indique Olivier Delaneau. Une quantité si importante de données permet de construire des modèles statistiques de très haute précision, pour autant que l'on puisse les interpréter sans s'y noyer.»

Une licence open source en gage de transparence

Les chercheurs ont décidé de rendre leur outil accessible à tous sous licence open source MIT: le code entier est disponible et modifiable à l'envie, selon les besoins des chercheurs, et ce principalement pour des questions de transparence et de reproductibilité, mais aussi pour stimuler les chercheurs du monde entier. «Toutefois attention ! Nous donnons accès à l'outil d'analyse, mais en aucun cas à un corpus de données, tient à préciser Olivier Delaneau. A chacun, ensuite, de l'utiliser sur les données dont il ou elle dispose.»

Cet outil se révèle ainsi beaucoup plus efficace que ses prédécesseurs, plus rapide et moins cher. Il permet aussi de limiter l'impact environnemental numérique. Les ordinateurs puissants utilisés pour traiter les Big Data sont en effet très énergivores; réduire leur utilisation permet alors de minimiser leur impact négatif.

contact

Olivier Delaneau

Professeur assistant FNS, Département de biologie computationnelle
Faculté de biologie et médecine UNIL, SIB
+41 21 692 40 55
Olivier.Delaneau@unil.ch

Emmanouil Dermitzakis

Professeur ordinaire au Département de médecine génétique et développement
Faculté de médecine UNIGE, SIB
+41 22 379 54 83
Emmanouil.Dermitzakis@unige.ch

DOI: 10.1038/s41467-019-13225-y

UNIVERSITÉ DE GENÈVE

Service de communication

24 rue du Général-Dufour
CH-1211 Genève 4

Tél. +41 22 379 77 17

media@unige.ch

www.unige.ch