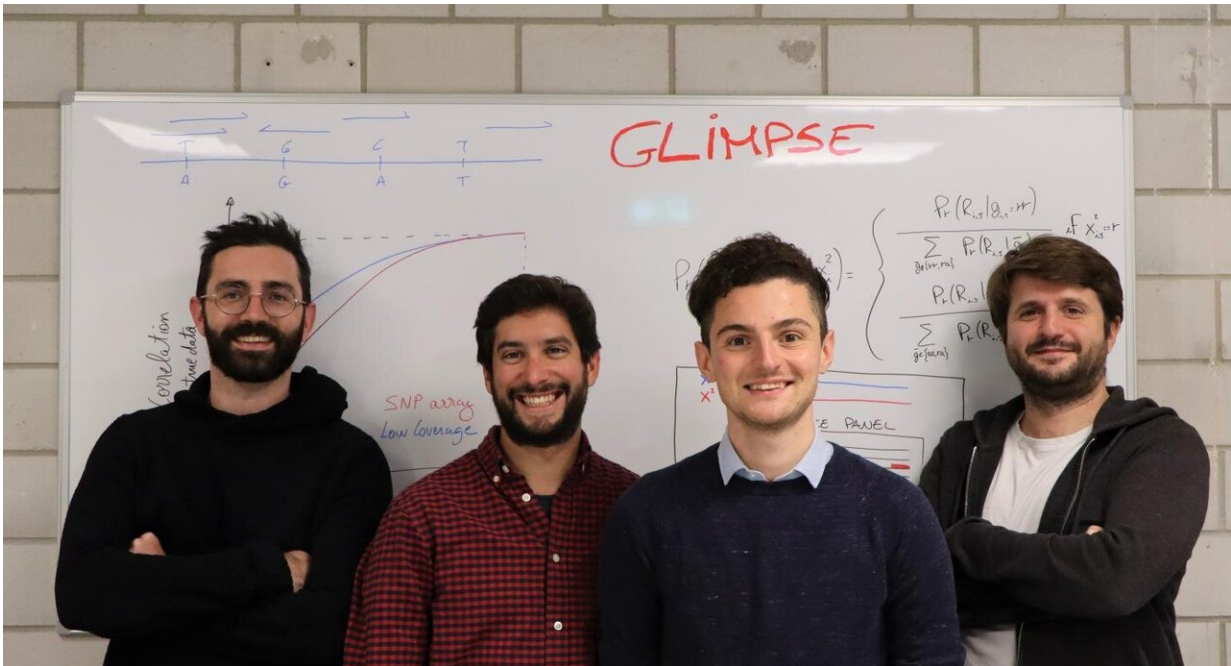


Inferring human genomes at a fraction of the cost promises to boost biomedical research

January 13 2021



From left to right: Robin Hofmeister, Diogo Ribeiro, Simone Rubinacci and Olivier Delaneau Credit: Delaneau Group

Thousands of genetic markers have already been robustly associated with complex human traits, such as Alzheimer's disease, cancer, obesity, or height. To discover these associations, researchers need to compare the genomes of many individuals at millions of genetic locations or markers, and therefore require cost-effective genotyping technologies. A new

statistical method, developed by Olivier Delaneau's group at the SIB Swiss Institute of Bioinformatics and the University of Lausanne (UNIL), offers game-changing possibilities. For less than \$1 in computational cost, GLIMPSE is able to statistically infer a complete human genome from a very small amount of data. The method offers a first realistic alternative to current approaches relying on a predefined set of genetic markers, and so allows a wider inclusion of underrepresented populations. The study, which suggests a paradigm shift for data generation in biomedical research, is published in *Nature Genetics*.

A cost-effective approach to probing genetic markers

Low-coverage whole [genome](#) sequencing (LC-WGS) followed by genotype imputation is a method by which a whole genome can be inferred statistically from a very low sequencing effort. It has been proposed as a less biased and more powerful alternative to SNP arrays (see box), but its high computational cost has prevented it from becoming a widely used alternative. The team of scientists led by Olivier Delaneau, Group Leader at SIB and UNIL, has developed an [open-source software](#), called GLIMPSE, that finally overcomes these issues. "GLIMPSE provides a framework that is 10-1,000 times faster, and thus cheaper, than other LC-WGS methods, while being much more accurate for rare genetic markers" explains Olivier Delaneau. "GLIMPSE is able to greatly enhance a low-coverage genome at millions of markers for less than \$1 in computational cost, making it the first real alternative to SNP arrays."

From unbiased data to unbiased healthcare Genome-wide association studies have so far mostly focused on Europeans: 80% of all GWAS participants are individuals of European descent, yet these make up only 16% of the world population. This is an important ethical issue in terms of healthcare inclusiveness and equitable access to the benefits of

[biomedical research](#), as the way genetic markers contribute to disease susceptibility varies across human populations. LC-WGS naturally circumvents the bias inherent to pre-established sets of [genetic markers](#) (SNP arrays). It can thus be successfully applied to underrepresented populations, as shown in this study for an African-American population as a proof-of-concept. "In addition to breaking down the financial barrier to enable GWAS studies based on LC-WGS, what is really exciting about this approach is that it enables researchers to efficiently uncover associations in understudied populations" says Simone Rubinacci, Postdoctoral Researcher in Olivier Delaneau's Group and first author of the paper.

Taking advantage of genomes already sequenced

"Our original thinking was: can we make use of the wealth of sequenced genomes to improve those that are newly sequenced? In other words, more for less: this is exactly what GLIMPSE does," explains Diogo Ribeiro, Postdoctoral Researcher in Olivier Delaneau's Group and co-author of the paper. How does it work? By building on the idea that we all share relatively recent common ancestors, from which small portions of our DNA are inherited. Briefly, GLIMPSE mines large collections of human genomes that have been very accurately sequenced (high-coverage WGS) to identify portions of DNA that are shared with newly sequenced genomes. In this way, GLIMPSE can reliably fill in the gaps in the low-coverage data.

A new paradigm for future genomic studies with far-ranging applications

Made available as part of an open-source suite of tools, GLIMPSE paves the way for wide adoption of low-coverage WGS, promoting a [paradigm shift](#) in data generation for future genomic studies. Since the first release

of the software as a preprint in April 2020, ongoing research has already started to use the tool, for instance to reconstruct the genomes of people living thousands of years ago from ancient DNA, or of COVID-19 patients from SARS-CoV-2 nasopharyngeal swabs as part of a GWAS study.

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

Provided by Swiss Institute of Bioinformatics

Citation: Inferring human genomes at a fraction of the cost promises to boost biomedical research (2021, January 13) retrieved 27 April 2024 from <https://medicalxpress.com/news/2021-01-inferring-human-genomes-fraction-boost.html>

<p>This document is subject to copyright. Apart from any fair dealing for the purpose of private study or research, no part may be reproduced without the written permission. The content is provided for information purposes only.</p>
--