

Personal Genomics Enters a New Phase

Scott Edmunds 

Published October 11, 2016

Citation

Edmunds, S. (n.d.). In *GigaBlog*. GigaBlog. <https://doi.org/10.59350/1zm58-mc892>

Keywords

Medicine, Technology, Genomics, GigaDB, Open Data
Feature Image

Copyright

Copyright © Scott Edmunds 2016. Distributed under the terms of the [Creative Commons Attribution 4.0 International License](#), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Open data is a critical component of the scientific method, but genomes are both identifiable and predictive. As a result, most studies choose to withhold data from participants and restrict access to researchers, hampering the connections and sample sizes needed for precision medicine to work (see our previous blog covering [this](#)). The [Personal Genome Project \(PGP\)](#) has pioneered using detailed and portable "open consent" procedures to move beyond these restrictions for the greater good, using volunteers willing to donate diverse personal information to become a public resource. Founded by George Church of Harvard Medical School, and spawning a network of regional offshoots such as [PGP UK](#) (founded by our editorial board member Stephan Beck), PGP aims to produce a unique resource for humans, providing open access to genes, environments and traits. Starting with George being their first personal genomics volunteer (see #PGP1's profile [here](#)) they now have over 5,000 registered users that have passed the "open consent" procedure and have started updating diverse phenotypic information.

[Complete Genomics](#) (a subsidiary of our co-publishers BGI), and the [PGP](#) today [publish a Data Note](#) releasing and describing of over 100 individual whole genome sequences with experimental haplotype phasing. This set of personal genomics data was generated using Complete Genomics Long Fragment Read (LFR) technology and represents the largest set of high coverage whole human genome assemblies with comprehensive experimentally determined haplotypes. "The vast majority of genomic data that has been generated to date is without experimentally derived haplotypes" explained Dr. Brock Peters, Senior Director of Research and project leader for Complete Genomics. "This represents a very unique set of data that is freely available for anyone to use through open access data publication." A total of 184 individuals, recruited by the PGP, took part in the project. Each individual consented to have their identity, their genome, and their phenotype data made freely and publicly available. Blood samples were collected by the PGP team and sent to Complete Genomics for DNA isolation, LFR library generation, and whole genome sequencing. Currently 114 genome assemblies are available with the remaining 70 expected to be released in the coming few months after the sign off of the donors.

"In 2011, we made freely available a [set of 69 whole human genome sequence assemblies](#) which quickly became a highly utilized resource and benchmark for the genetics community," stated Dr. Radoje Drmanac, CSO of Complete Genomics. "We are proud to continue the tradition by releasing this set of experimentally haplotyped whole human genome sequence assemblies. This represents the largest and most accurate set of human haplotypes currently available." The terabytes of sequencing data and detailed phenotypic information is available from dbGap (phs000905.v1.p1), [the PGP website](#) and the *GigaScience* GigaDB repository ([doi:10.5524/100242](#)).

"Combining Complete Genomics' advanced WGS with the PGP's informed consent policy which allows for unrestricted access and *GigaScience*'s open access data publication method enables the full release of a large data set with exceptional scientific value. We expect it will be used by many researchers around the world", explained Dr. Church.

The technology used to generate this dataset, LFR, was previously described by Complete Genomics in a [2012 *Nature* publication](#). In our new publication LFR was again shown to be highly accurate and complete. Each sample was sequenced to 100X coverage allowing for the detection of most variants with high confidence. This allowed for over 98% of heterozygous variants to be placed into long contigs approaching 1 Mb in length. On average, over 85% of haplotypes contained no errors with the majority of the remaining 15% having only a single phasing error.

We encourage use of this data by the academic community and beyond, as George Church says in the [above video](#), empowering the credential less out-of-the-box thinkers who usually would not get access to this type of data. On top of the high quality of this phased data, the large number and politics free and open nature of these datasets will make them a priceless reference in enabling genome-driven precision medicine to succeed.

Save

The post [Personal Genomics Enters a New Phase](#) appeared first on [GigaBlog](#).