# openSNP - personal genomics and the public domain

Bastian Greshake[1], Philipp Bayer[2], Helge Rausch, Julia Reda

[1]Department for Applied Bioinformatics, Goethe University, Frankfurt am Main, Germany Email: bgreshake@googlemail.com

[2] School of Plant Biology, University of Western Australia, Perth, Australia

**Project Website**: https://openSNP.org
**Source Code**: https://github.com/gedankenstuecke/snpr
**License**: Code: MIT, Data: Creative Commons Zero

Direct-To-Consumer (DTC) genetic testing is still a rather recent but growing phenomenon, with some companies now having as many as 850k paying customers (1). DTC genetic tests are still largely based on the analysis of Single Nucleotide Polymorphisms (SNPs) with micro arrays. The analysis of such SNP data sets in general, using Genome-Wide Association Studies (GWAS), has lead to the discovery of many genotype-phenotype associations for a wide array of phenotypic traits and diseases (2,3). The growing number of customers also enabled providers of DTC genetic testing to perform their own GWAS (4).

By and large the data generated in the field of DTC genetic testing is not openly available to third parties, like academic researchers, citizen scientists, and hobbyist genealogists. This is partially due to privacy and ethical concerns (5,6), but also due to financial interests of the DTC providers (7,8). With ***openSNP*** we created an easily accessible platform to enable customers of DTC to dedicate their own DTC genetic testing data along with their phenotypes into the public domain using Creative Commons Zero. Additionally the platform offers annotations for those SNPs generated by mining different public and open data sets, such as the *Public Library of Science*, the *SNPedia* (9), *Mendeley* and more.

Since *openSNP* started at the end of 2011 people have used it to release over 1700 genotyping files into the public domain and included information for over 289 phenotypes, thus creating a free and growing resource for scientists and citizen scientists alike. With this the project actively contributes to the discussion on open human genetic data, such as bioethical implications and privacy research (10,11), genealogy, teaching, pharmacogenomics (12) and even art (13).

**REFERENCES**

(1) http://mediacenter.23andme.com/blog/2015/02/19/fdabloomupdate/
(2) Johnson A, O'Donnell C. BMC Med. Genet. (2009); 10: 6.
(3) Hindorff LA, Sethupathy P, Junkins HA, Ramos EM, Mehta JP, et al. Proc Nat Acad Sci USA. (2009); 106: 9362-9367.
(4) Do CB, Tung JY, Dorfman E, Kiefer AK, et al. PLoS Genet. (2011); 7(6): e1002141.
(5) Caulfield T, McGuire AL. Annu. Rev. Med. (2011); 63: 1.1-1.11.
(6) Hogarth S, Javitt G, Melzer D. Annu. Rev. Genomics Hum. Genet. (2008); 9: 161–82.
(7) http://mediacenter.23andme.com/blog/2015/01/06/23andme-genentech-pd/
(8) http://mediacenter.23andme.com/blog/2015/01/12/23andme-pfizer-research-platform/
(9) Cariaso M and Lennon G, Nucl. Acids Res. (2012) 40 (D1):D1308-D1312.
(10) Angrist M, PLoS ONE (2014) 9(3): e92060.
(11) http://seclab.soic.indiana.edu/GenomePrivacy/papers/Genome%20Privacy-paper4.pdf
(12) Samwald et al. BMC Medical Informatics and Decision Making (2015) 15:12
(13) https://soundcloud.com/thesoundofpeople/the-sound-of-bastian-greshake-6-channels-remixed