openSNP An open data resource for personalized genotype & phenotype data

Bastian Greshake[1], Philipp Bayer[2], Helge Rausch, Julia Reda
[1]Department of Applied Bioinformatics Goethe University Frankfurt/Main, Germany
[2]Applied Bioinformatics Group, School of Agriculture and Food Sciences, University of Queensland, Oz

Introduction

Direct-To-Consumer (DTC) genetic testing, as offered by companies such as *23andMe*, *FamilyTreeDNA* or *Ancestry,* has been on the rise in the last couple of years: *23andMe* alone has over 1 million customers worldwide who have had their genomes tested [1]. Those services usually check for Single Nucleotide Polymorphisms (SNPs) using microarrays, though Exome and Whole Genome sequencing services are also currently entering the market [2]. Many of those DTC customers allow their service providers to use the data for in-house research.

If only a small fraction of DTC customers would make their data publicly available, a rich and free data resource for scientists all over the world could be created. With openSNP we created a platform where people can donate their genetic & phenotypic data into the public domain.

Data Sources

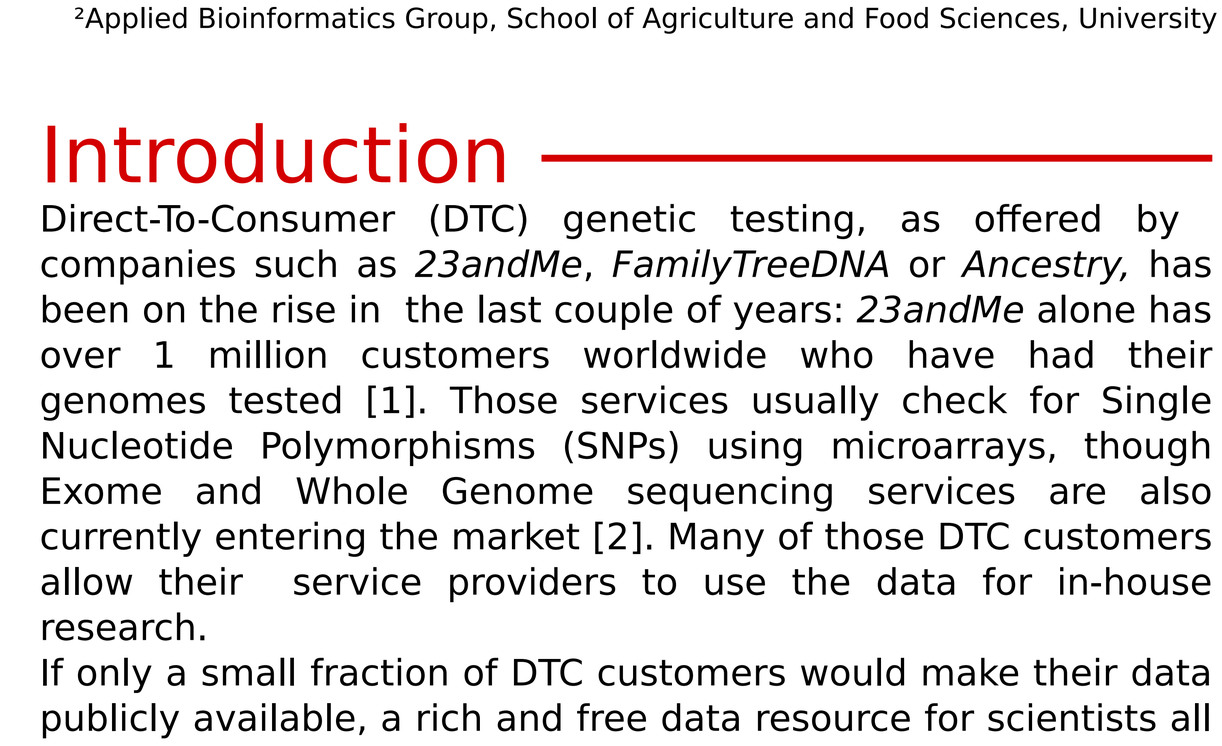

Figure 1 The different data inputs of *openSNP*

People can upload their **personal genetic data**, coming from sources like 23andMe, FamilyTreeDNA, Ancestry.com and deCODEme without any need to convert their input files. Additionally support for generic VCF files is offered.

People can also **collect phenotypes.** They are entered using either text or by uploading pictures. New phenotype categories are generated by the users themselves. Additionally people can publish their personal activity data, as recorded by Fitbit [3].

To annotate the individual variants present, we mine external services for primary literature (PLOS, Mendeley) and known associations (SNPedia [4], GWAS Catalog [5], GET Evidence System [6]).

The Web Platform

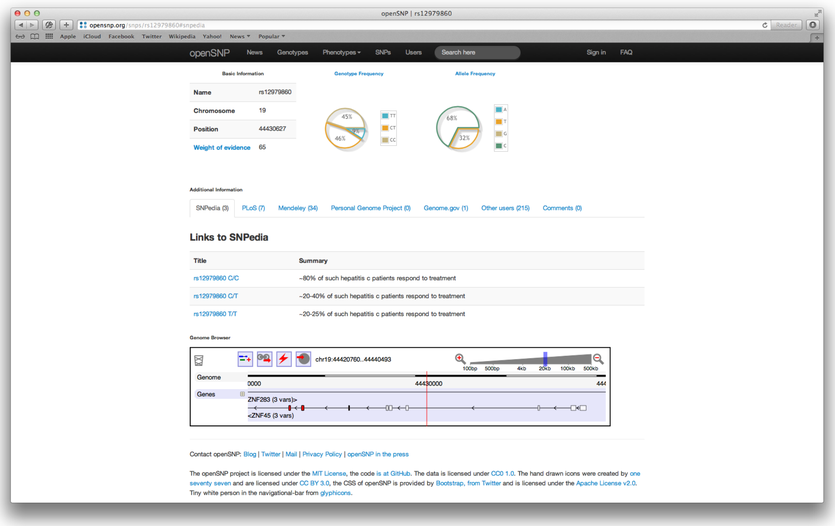

Figure 2 Screenshot of detailed SNP view. Showing genome browser & annotations

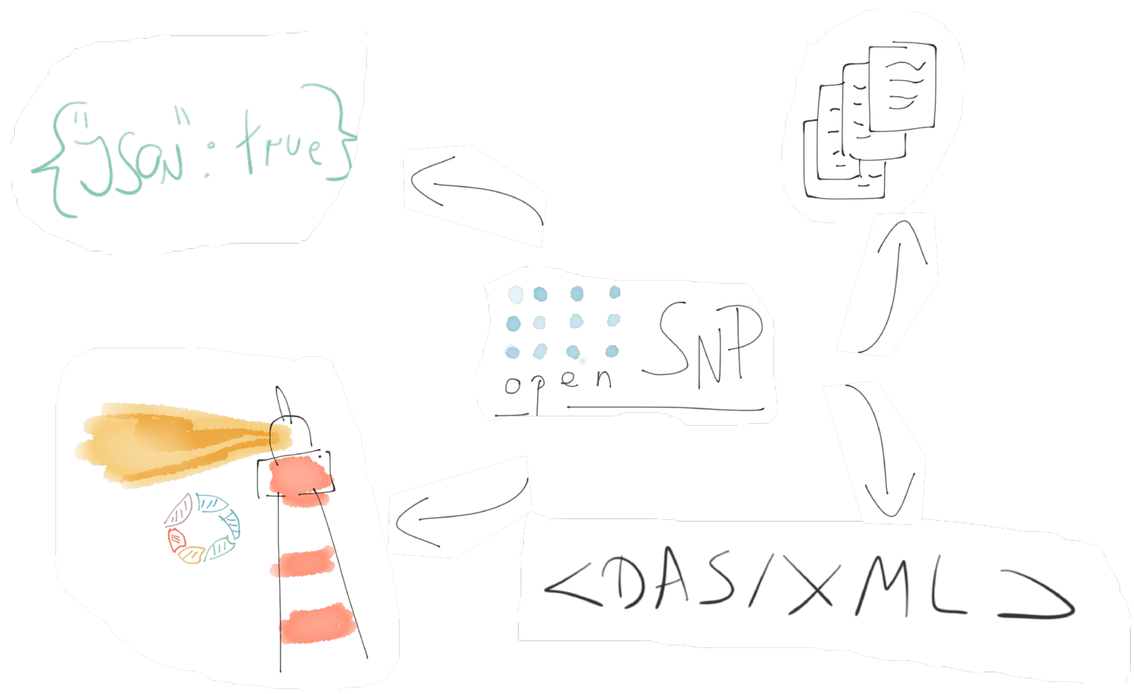

Figure 3 The different output methods *openSNP* offers

The genotype and phenotype data can be downloaded using different strategies. People can download all data sets in a single zip-archive or get only data sets associated with a given phenotype. Additionally ***openSNP* offers different APIs**: a custom *JSON*-API, the *Distributed Annotation System* (DAS) [7] and it also works as a Beacon for the *Global Alliance for Genomics and Health*.

Usage & Outlook

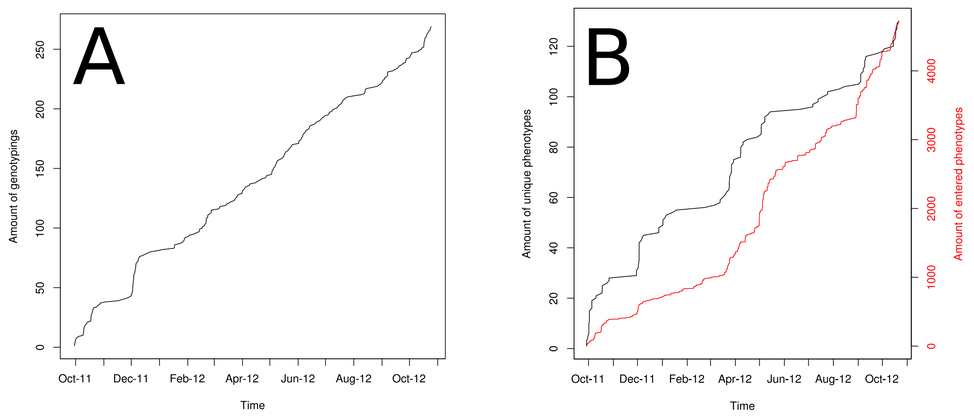

Figure 4 Growth of *openSNP*. **A** shows number of genetic data sets. **B** shows the number of phenotypes & variations **GRAPHS NEED TO BE REDONE!**

Between the start of openSNP on 27.09.2011 and 19.06.2015 3841 people have signed up for an account. In total 1984 genetic data sets have been uploaded by the users. Those users also have provided a total of 4743 entries on 360 different phenotypes, which range from hair- and eye-colour to political beliefs.

The application can be found at https://opensnp.org and the code is released under MIT License. All data uploaded to openSNP is released under Creative Commons Zero, which is equivalent to the Public Domain in many countries.

References

[1] http://blog.23andme.com/news/one-in-a-million/
[2] https://www.23andme.com/exome/
[3] http://www.fitbit.com
[4] Cariaso M and Lennon G. (2012) Nucl. Acids Res. 40 (D1): D1308-D1312
[5] Welter D, MacArthur J, Morales J et al. (2014) Nucl. Acids Res V42 (Database issue): D1001-D1006
[6] Ball MP, Thakuria JV, Zaranek AW, et al. (2012) PNAS 109: 11920- 11927.
[7] Dowell R, Jokerst R, Day A et al. (2001) BMC Bioinformatics 2: 7.
[8] http://genomicsandhealth.org/