openSNP
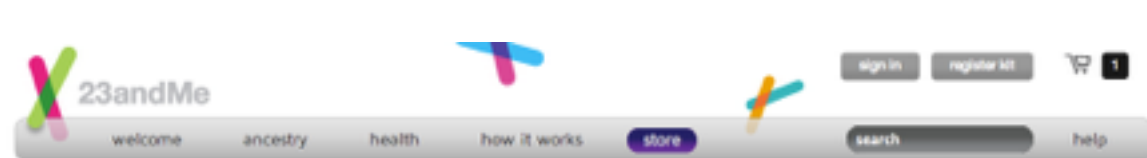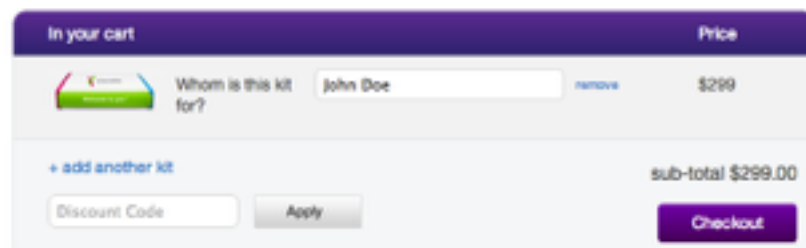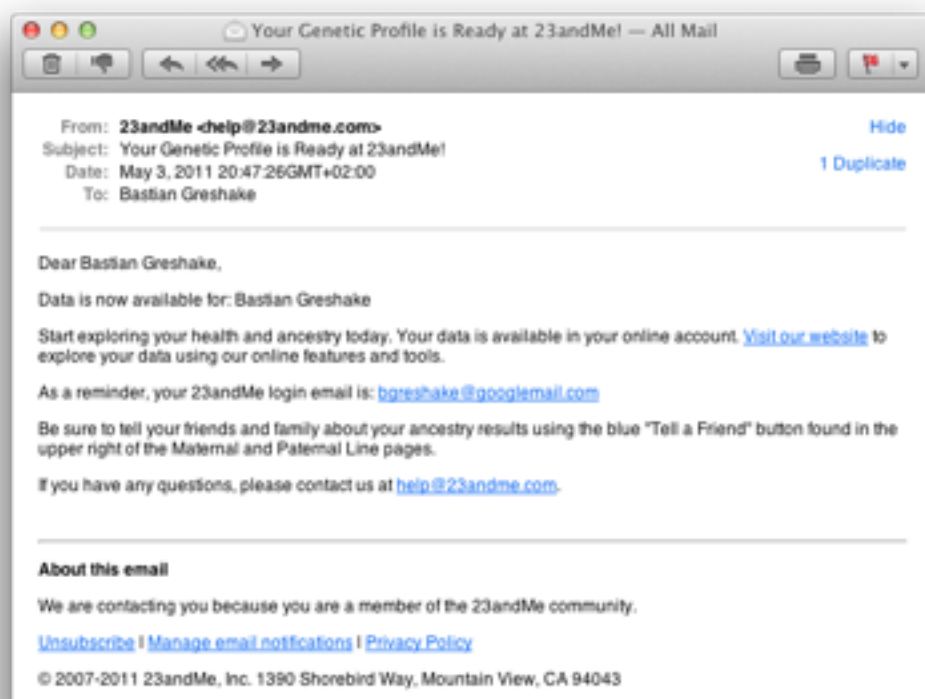
personalized genotype & phenotype data in the public domain

DTC
Genetic
Testing

Potential

- 23andMe already close to 1 mio customers

- AncestryDNA projected to reach 1.3 mio customers at the end of 2015

- What if (at least some) people share their data?

gedankenstuecke / **genotyping_23andme**

🗋 **Pull Request**	◉ **Unwatch** ▾	★ **Unstar** 8	⑂ **Fork** 3

Code	Network	Pull Requests 0	Issues 0	Wiki	Graphs	Admin

My raw genotyping data from 23andme — Read more

🍎 Clone in Mac	☁ ZIP	HTTP	SSH	Git Read-Only	git@github.com:gedankenstuecke/genotyping_23and	🔒 Read+Write access

⑂ branch: **master** ▾ | **Files** | Commits | Branches 1 | | Tags | Downloads

genotyping_23andme /   ⊕ **16 commits**

Merge pull request #1 from sauloal/master ...

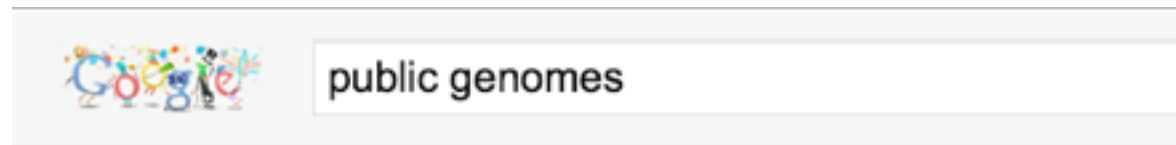👤 **gedankenstuecke** authored 7 months ago   📋 latest commit e3489a0c5f

📄 .gitignore	2 years ago	last change to readme [gedankenstuecke]
📄 23andme_feed.xml	2 years ago	added example-RSS [gedankenstuecke]
📄 README.mediawiki	a year ago	edit on readme [gedankenstuecke]
📄 description.data	2 years ago	added pubmed-crawler [gedankenstuecke]
📄 genome_Bastian_Greshake_Full_2011050...	2 years ago	added Promethease-report [gedankenstuecke]
📄 genome_Bastian_Greshake_Full_2011050...	2 years ago	added first data [gedankenstuecke]
📄 genotypes.data	2 years ago	added pubmed-crawler [gedankenstuecke]
📄 homozygot_dominants.txt	2 years ago	added pubmed-crawler [gedankenstuecke]
📄 homozygots.txt	2 years ago	added data about homozygots & homozygot-dominants [gedankenstuecke]
📄 pubmed_crawler.py	7 months ago	adding pickle to paper_getter to allow continuation in case of accide... [sauloal]
📄 snp.py	a year ago	added other genotypes to snp-class [gedankenstuecke]
📄 snpedia.data	2 years ago	added pubmed-crawler [gedankenstuecke]
📄 snpedia.py	a year ago	added allel-frequencies [gedankenstuecke]

More Data?
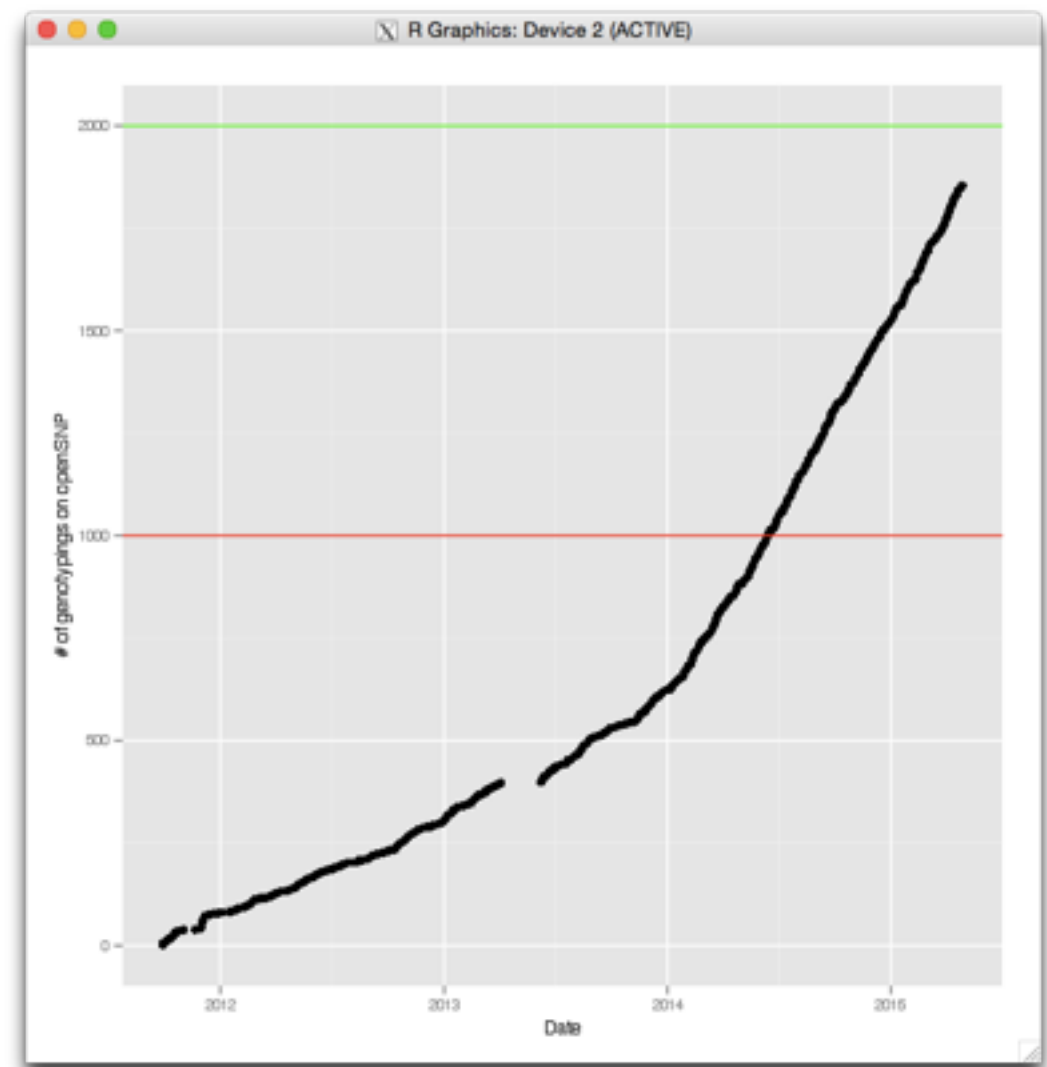
It's a mess!

- Genotypes

- Phenotypes

- Metadata about SNPs

Genotypes

- close to 2000 data sets

- formats:

  - 23andMe

  - AncestryDNA,

  - FamilyTreeDNA

  - (deCODEme)

Phenotypes

Phenotype: Eye color

Known Phenotypes:

- **Brown** Download genotyping-files of all corresponding users
- **Brown-green** Download genotyping-files of all corresponding users
- **Blue-green** Download genotyping-files of all corresponding users
- **Blue-grey** Download genotyping-files of all corresponding users
- **Green** Download genotyping-files of all corresponding users
- **Blue** Download genotyping-files of all corresponding users
- **Hazel** Download genotyping-files of all corresponding users
- **Mixed** Download genotyping-files of all corresponding users
- **Gray-blue** Download genotyping-files of all corresponding users
- **Blue-grey; broken amber collarette** Download genotyping-files of all corresponding users
- **Dark blue** Download genotyping-files of all corresponding users
- **Dark brown** Download genotyping-files of all corresponding users
- **Blue-green** Download genotyping-files of all corresponding users
- **Light-mixed green** Download genotyping-files of all corresponding users
- **Hazel/light brown** Download genotyping-files of all corresponding users

- completely crowdsourced

- so far: free text entries (messy) + picture uploads

- ~350 different categories

- additionally: Fitbit data (weight, sleep, steps) + pictures

Metadata for SNPs

- Effect of a given SNP according to SNPedia

- Effect of a SNP according to NHGRI-EBI GWAS catalog

- Effect according to GET Evidence of Personal Genome Project Harvard

- Papers that mention a given Rs-ID

Access: Genotypes/SNPs

- Distributed Annotation System

- Ugly JSON-"API":

  - downloadlinks for complete files per user

  - genotypes at a SNP for a given user/all users

Data Points:
- Rs-ID

- Chromosome & Location

- Allele & Genotype-Frequency in our DB

- Individual Genotypes

Access: Phenotypes

- JSON-API:

  - By Phenotype

  - All phenotypes for a given user

Data Points:
- Phenotype Name

- Phenotype Description

- Known Variations

- Users & their variations

Access: Metadata

Papers:

- Title

- First Author

- Publication Date

- DOI / URL

- Open Access?

- # of Readers

Others:

- Gene

- Trait

- Impact

- Summary

- Inheritance

- p-value

- CI

missing so far

- API access to Images & Fitbit data

- curated phenotypes as additional level to crowdsourced phenotypes

Thx!