Crowdsourcing Genome Wide Association Studies

Bastian Greshake and Philipp Bayer

28.12.2011

Overview

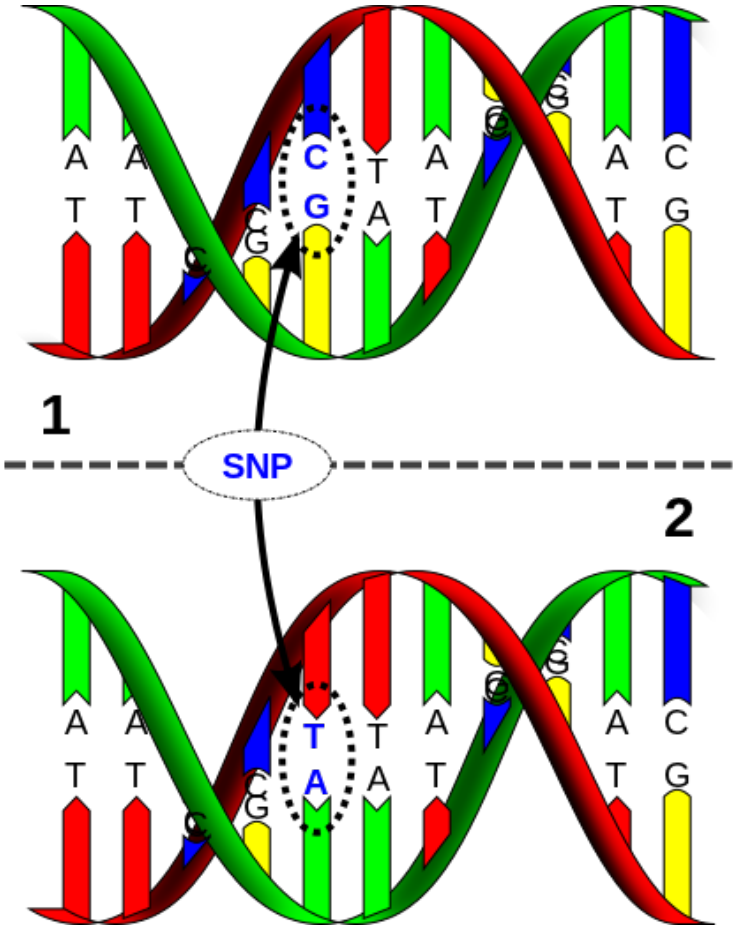
What are GWAS?

- Genome-wide Association Studies

What are GWAS?

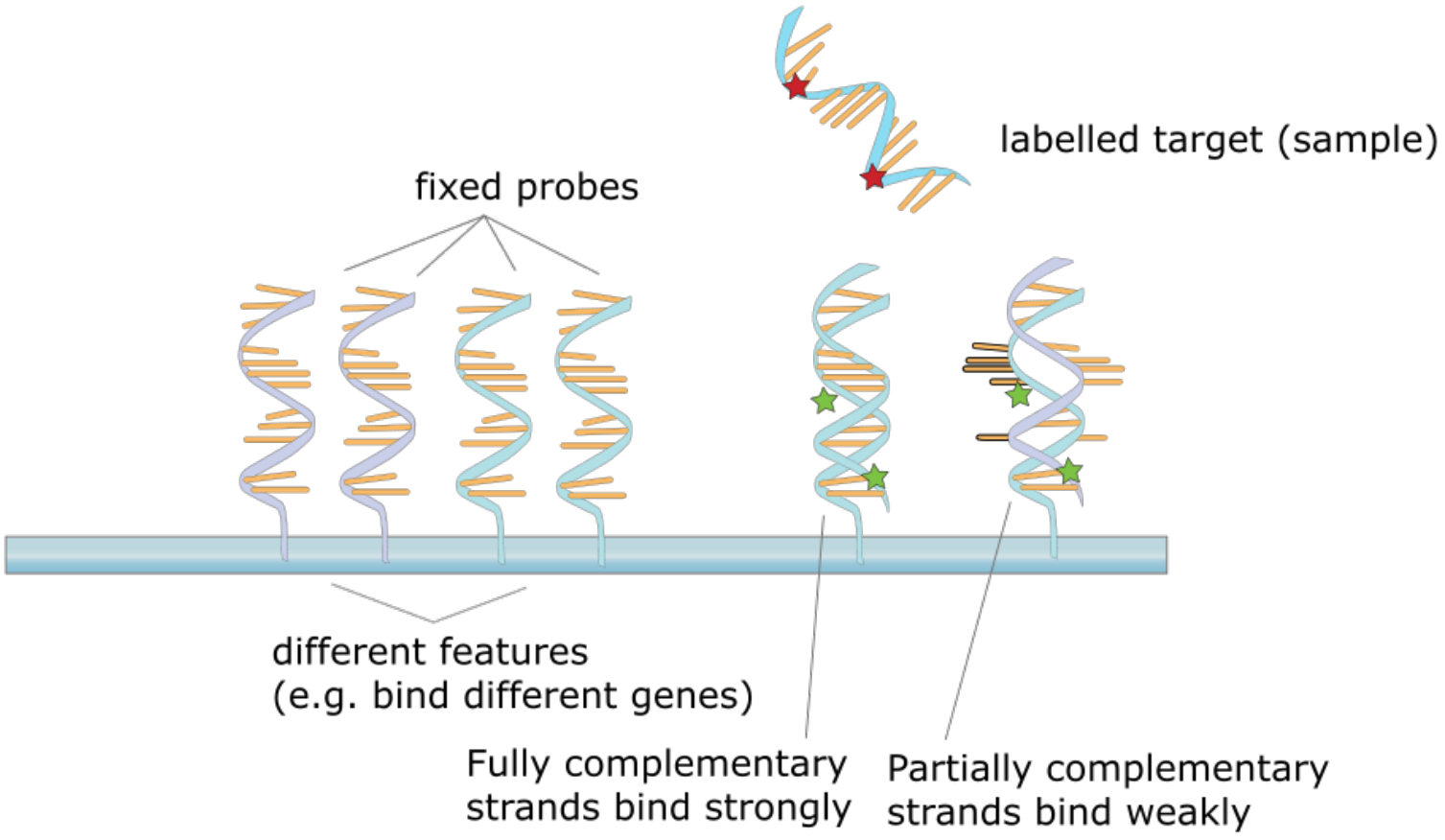
- Genome-wide Association Studies
- Link genetic variants (SNPs) to certain traits like eye or hair colour or to diseases like Diabetes, types of cancer

Single Nucleotide Polymorphism



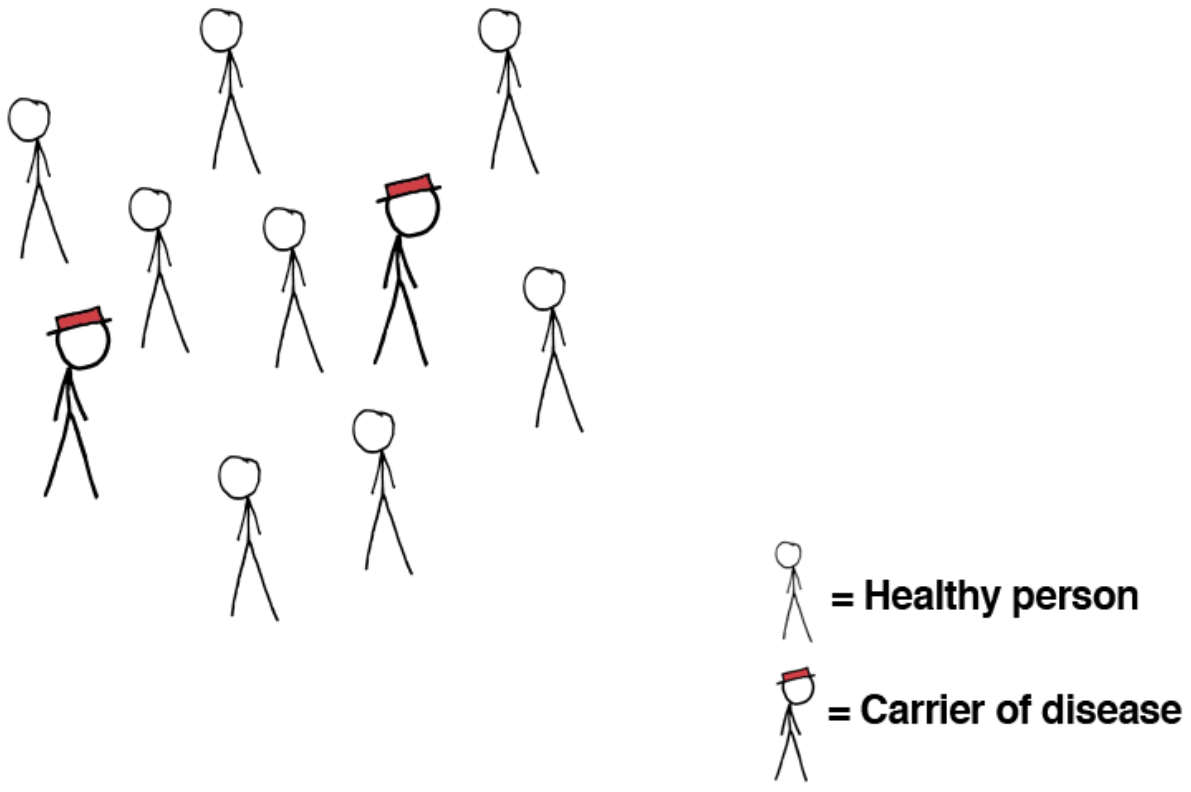
Source: http://en.wikipedia.org/wiki/File:Dna-SNP.svg

How to analyse SNPs?

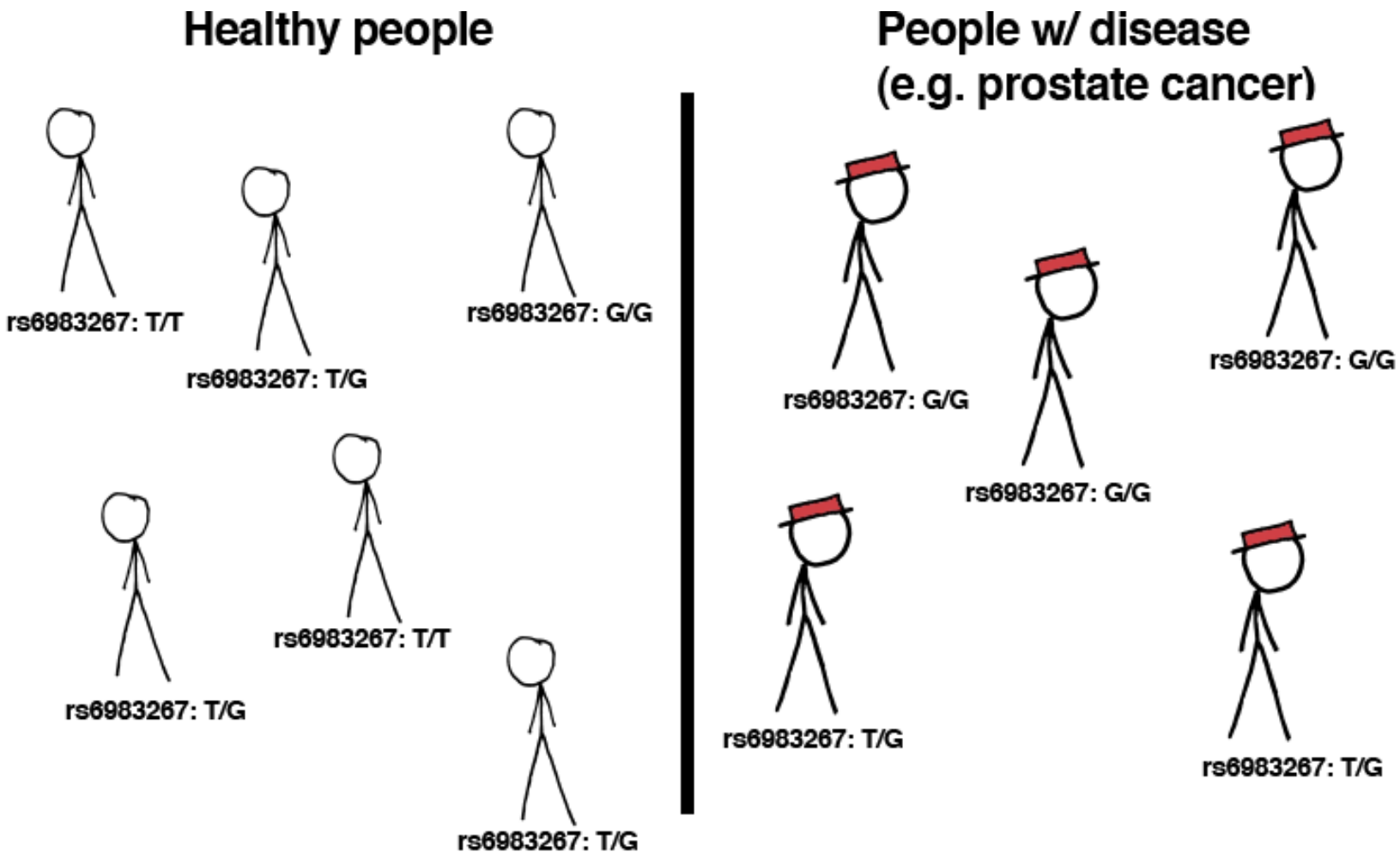


Source: http://en.wikipedia.org/wiki/File:NA_hybrid.svg

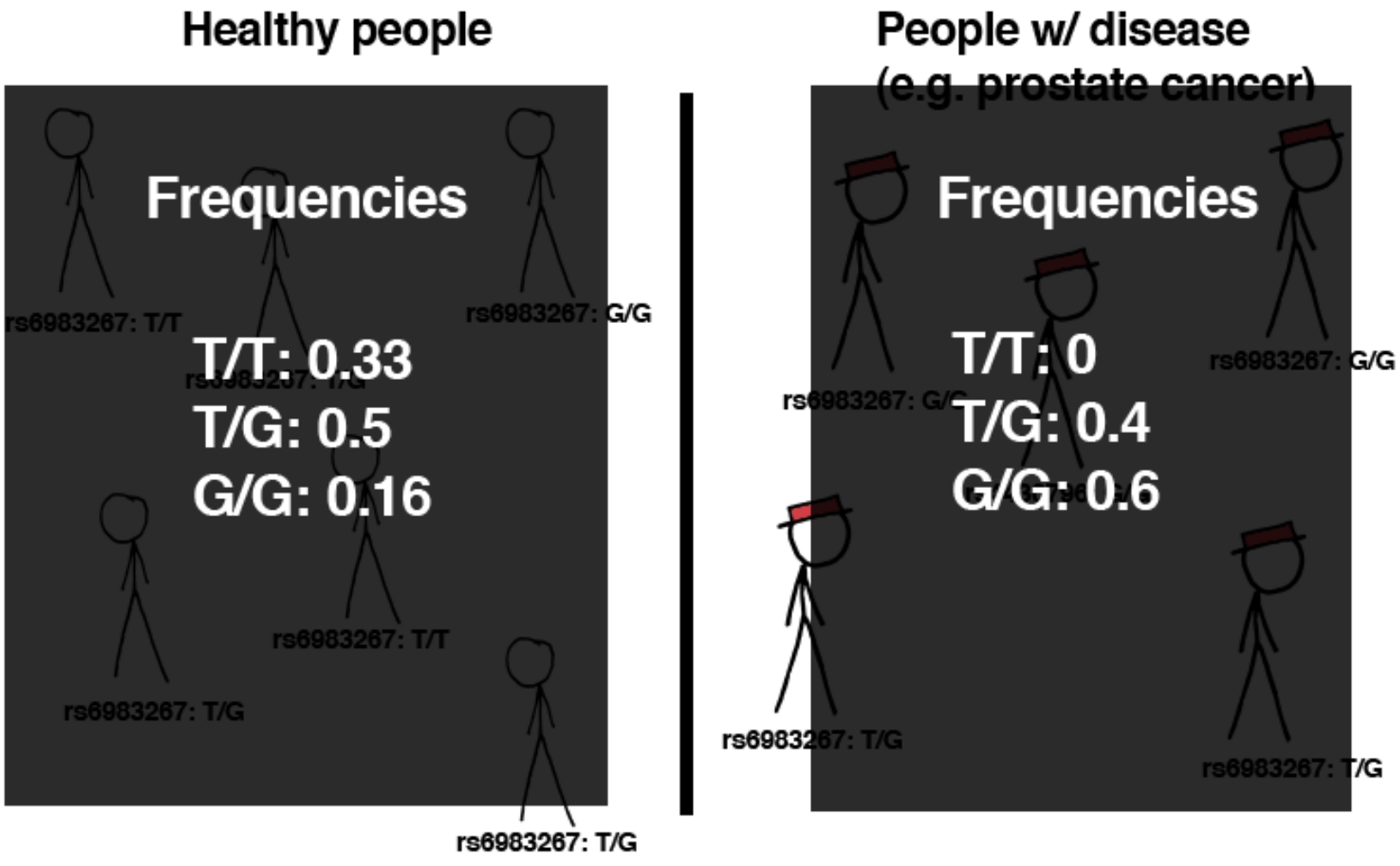
How do GWAS work?



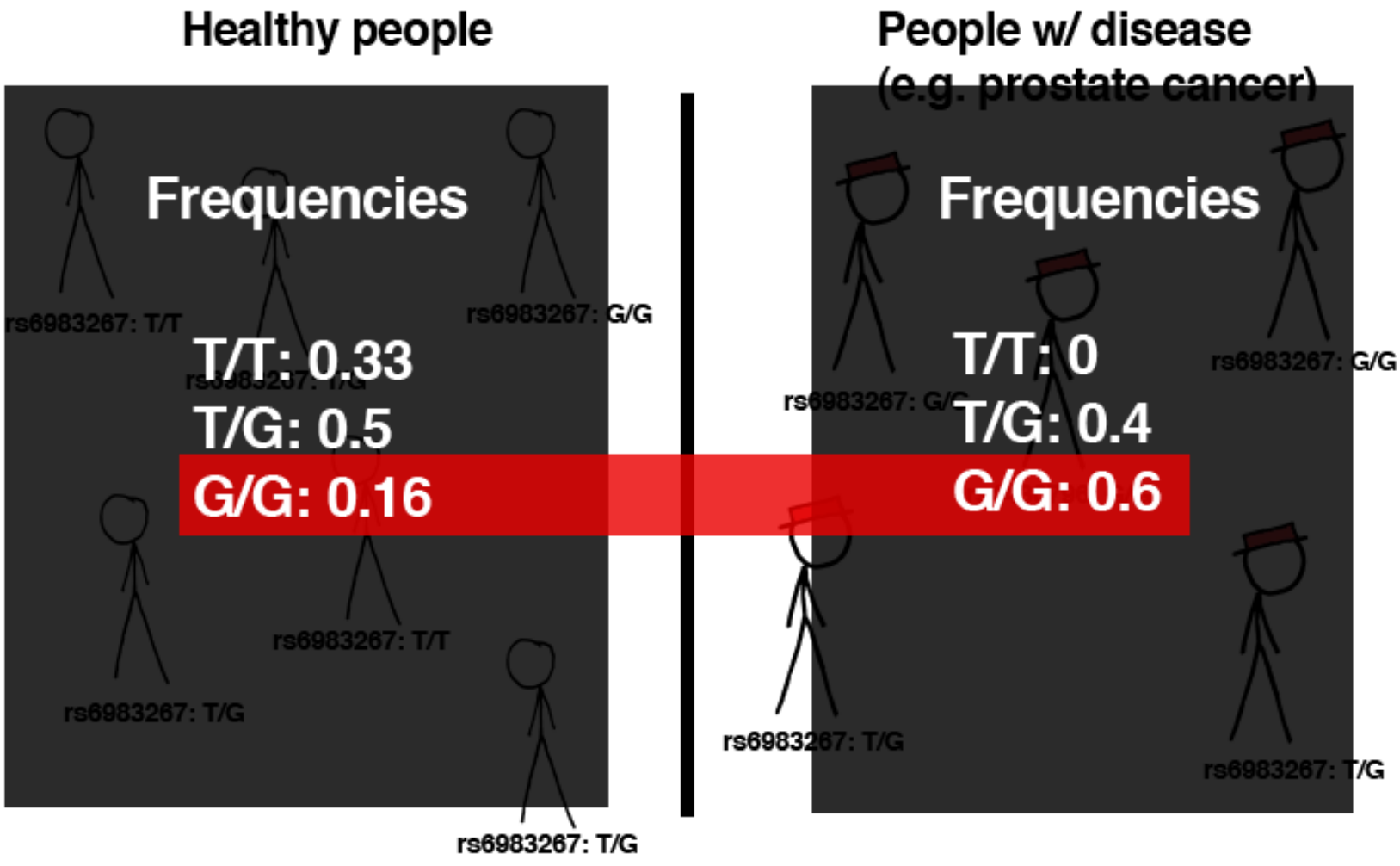
How do GWAS work?



How do GWAS work?



How do GWAS work?



Some GWAS-examples

- Sladek *et al.* (2007) identified four gene locations linked to heightened type 2 diabetes risk

Some GWAS-examples

- Sladek *et al.* (2007) identified four gene locations linked to heightened type 2 diabetes risk
- Kogan *et al.* (2011) linked rs53576 (G:G) to pro-social behaviour

Some GWAS-examples

- Sladek *et al.* (2007) identified four gene locations linked to heightened type 2 diabetes risk
- Kogan *et al.* (2011) linked rs53576 (G:G) to pro-social behaviour
- The Wellcome Trust Case Control Consortium (2007) linked 24 locations to 7 major diseases

Problems with GWAS

- Large enough sample size

Problems with GWAS

- Large enough sample size
- Correcting for multiple testing

Problems with GWAS

- Large enough sample size
- Correcting for multiple testing
- Correlation != Causation

Putting GWAS to use

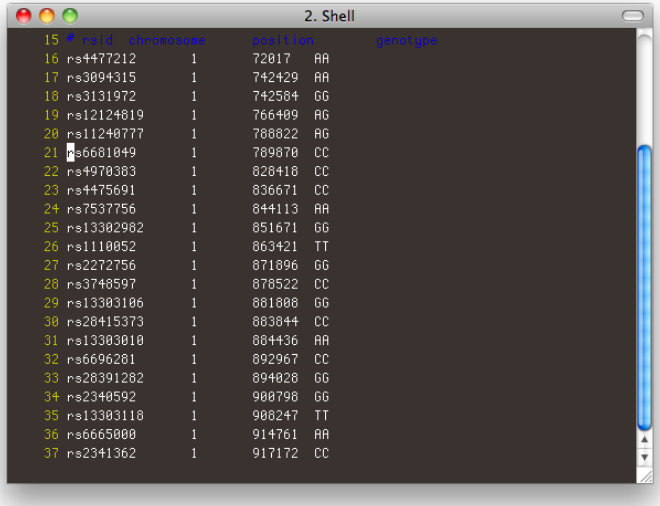
- Direct-To-Consumer genetic testing
- Analyse about 1 million SNPs and provide summary of disease risks & ancestry
- About $200 for a genotyping

Putting GWAS to use

- Direct-To-Consumer genetic testing
- Analyse about 1 million SNPs and provide summary of disease risks & ancestry
- About $200 for a genotyping
- Providers: 23andMe, deCODEme, FamilyTree DNA, ...

Putting GWAS to use

- Direct-To-Consumer genetic testing
- Analyse about 1 million SNPs and provide summary of disease risks & ancestry
- About $200 for a genotyping
- Providers: 23andMe, deCODEme, FamilyTree DNA, ...
- You get access to the raw data!



Numbers on DTC

- 23andMe alone has over 100.000 customers

Numbers on DTC

- 23andMe alone has over 100.000 customers
- 76 % of their customers agree to participate in research

Numbers on DTC

- 23andMe alone has over 100.000 customers
- 76 % of their customers agree to participate in research
- 59 % of them share phenotypic information with 23andMe

Research in company labs

- 23andMe published results of studies with up to 30.000 participants

Research in company labs

- 23andMe published results of studies with up to 30.000 participants
- Replication of older GWAS

Research in company labs

- 23andMe published results of studies with up to 30.000 participants
- Replication of older GWAS
- Finding new associations for Parkinsons disease

Data sharing

- People are already sharing the raw data of DTC tests

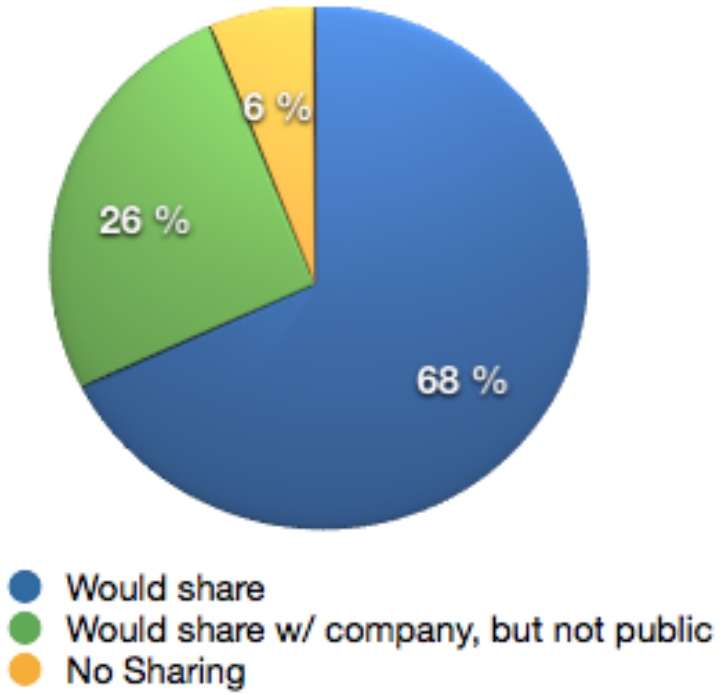
Data sharing

- People are already sharing the raw data of DTC tests
- 1-5 % of 23andMe customers would be enough to perform simple GWAS

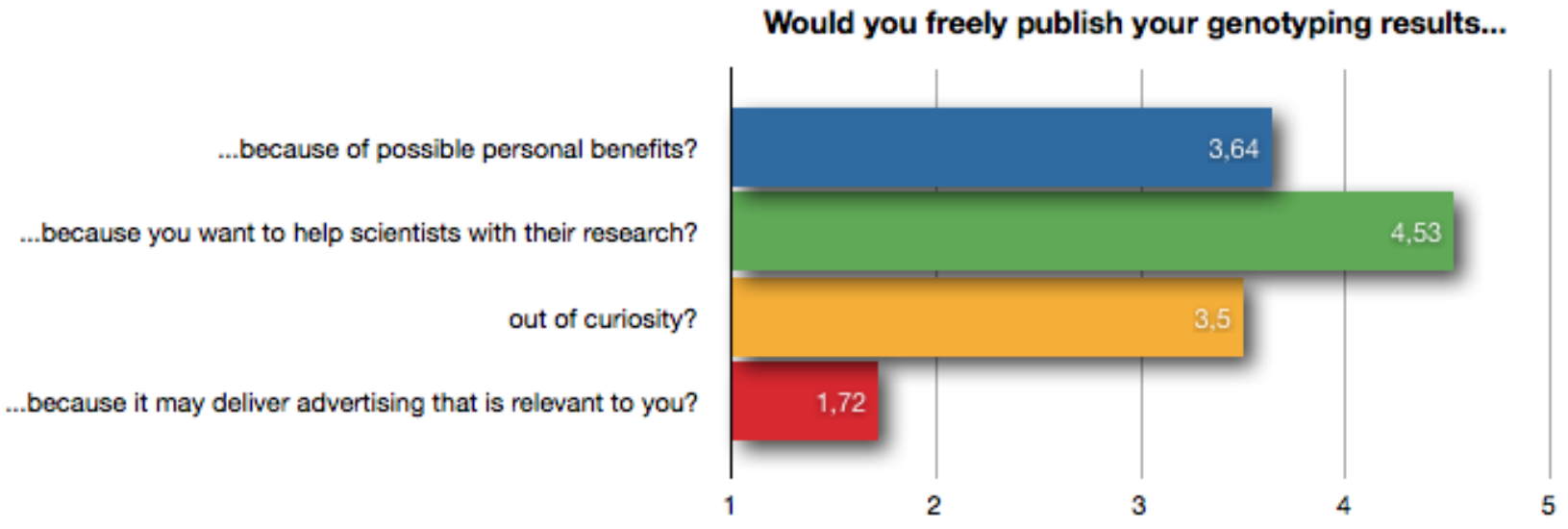
Data sharing

- People are already sharing the raw data of DTC tests
- 1-5 % of 23andMe customers would be enough to perform simple GWAS
- The *Personal Genome Project*: Open data, but closed participation

Willing to share?



Willing to share?



What can happen to your open data?

- Positive and negative consequences

What can happen to your open data?

- Positive and negative consequences
  - Possibly extremely bad consequences

What can happen to your open data?

- Positive and negative consequences
  - Possibly extremely bad consequences
- Up to you to decide whether you want to open your data

Positive consequences

- More knowledge about yourself

Positive consequences

- More knowledge about yourself
- Cheap, open science

Positive consequences

- More knowledge about yourself
- Cheap, open science
- Great data-source for citizen scientists

Negative consequences

- People know more about you than you might like

Negative consequences

- People know more about you than you might like
  - Including your boss, insurance company, government...

Negative consequences

- People know more about you than you might like
  - Including your boss, insurance company, government...
- Knowledge isn't static: Future research could show new, negative (or positive) associations.

Negative consequences

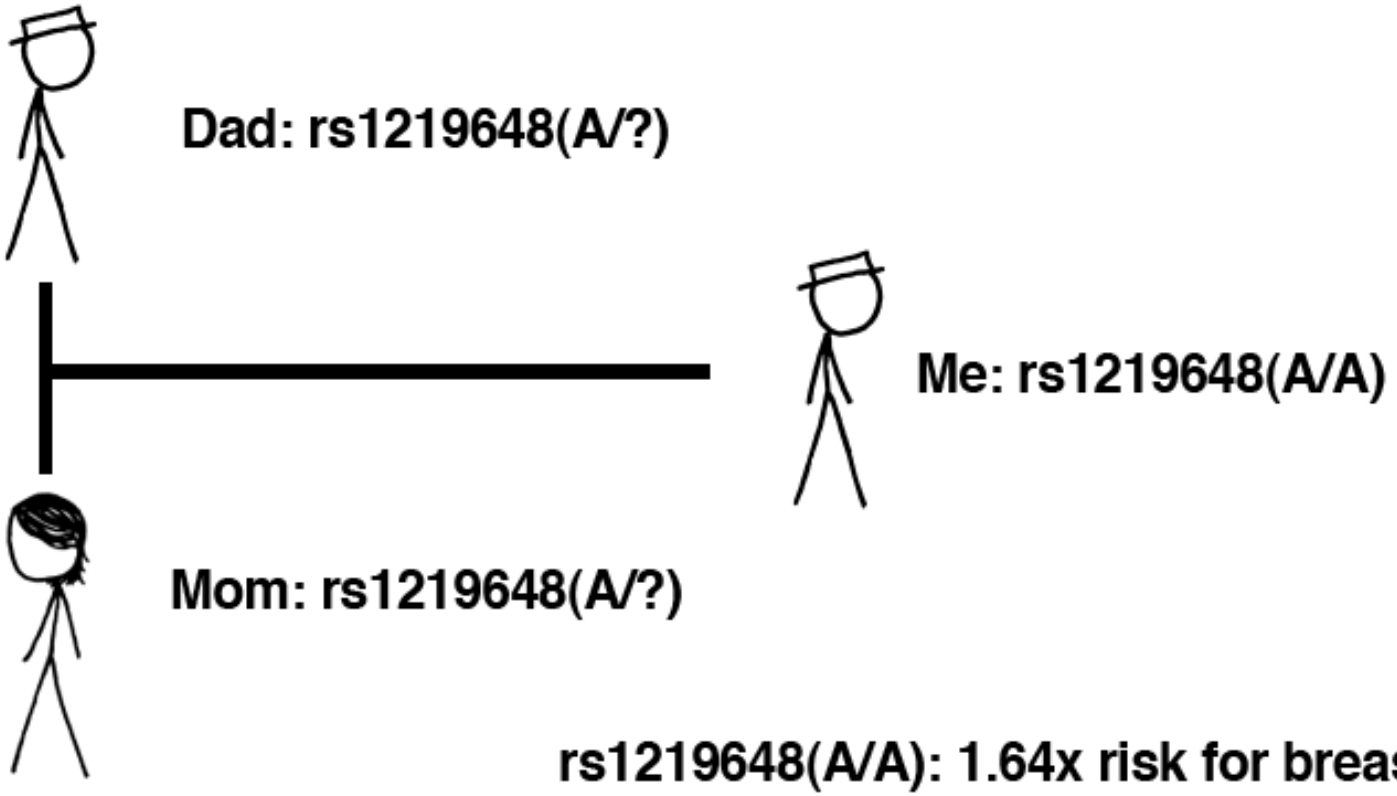
- People know more about you than you might like
  - Including your boss, insurance company, government...
- Knowledge isn't static: Future research could show new, negative (or positive) associations.
- Personal SNPs very similar to parents and relatives

Somebody Else's Problem? A case study

Me: rs1219648(A/A)

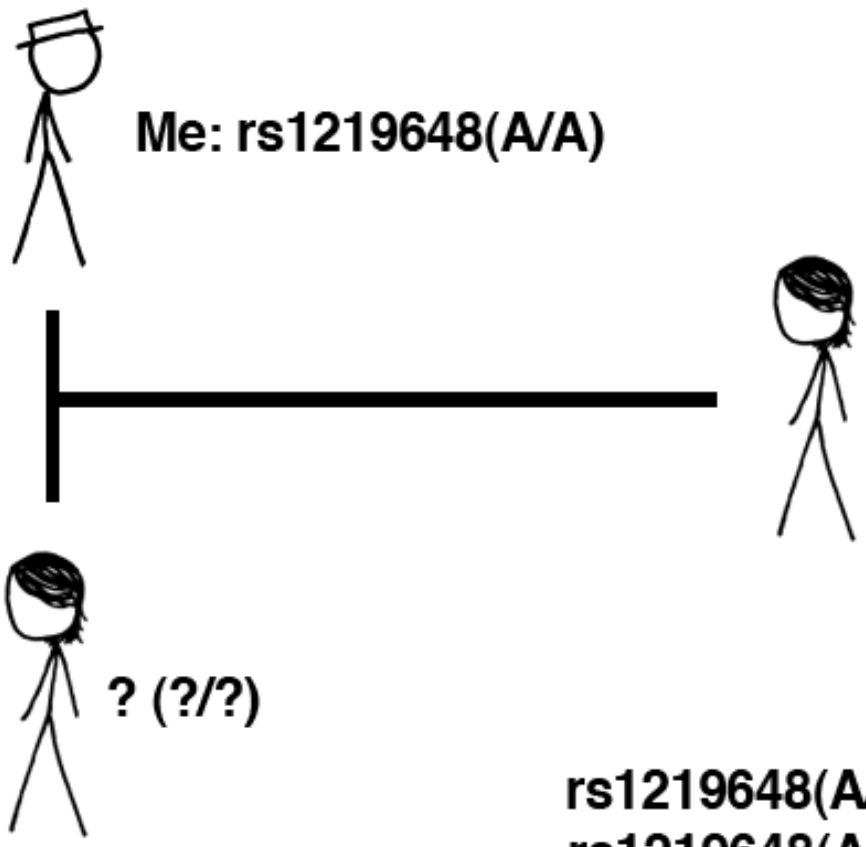
rs1219648(A/A): 1.64x risk for breast cancer
rs1219648(A/G): 1.20x risk for breast cancer
rs1219648(G/G): "normal" risk

Somebody Else's Problem? A case study



rs1219648(A/A): 1.64x risk for breast cancer
rs1219648(A/G): 1.20x risk for breast cancer
rs1219648(G/G): "normal" risk

Somebody Else's Problem? A case study



rs1219648(A/A): 1.64x risk for breast cancer
rs1219648(A/G): 1.20x risk for breast cancer
rs1219648(G/G): "normal" risk

Possible Solutions

- What about laws?

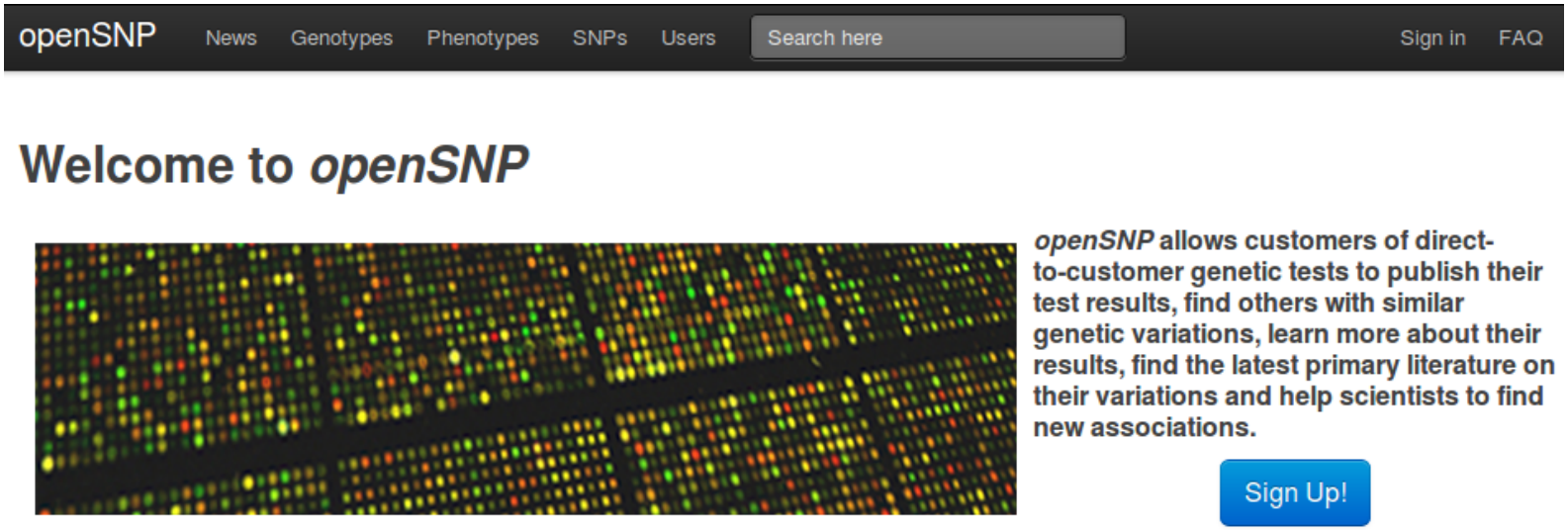
Possible Solutions

- What about laws?
  - US: Genetic Information Nondiscrimination Act (GINA, 2008)

Possible Solutions

- What about laws?
  - US: Genetic Information Nondiscrimination Act (GINA, 2008)
  - Germany: Gendiagnostikgesetz (GenDG, 2010)

For those who still want to share: Open GWAS



openSNP

- No central repository for open genotypings!

openSNP

- No central repository for open genotypings!
- We've created openSNP.org

openSNP

- No central repository for open genotypings!
- We've created openSNP.org
- open source repository for CC0-genotypings from 23andme, deCODEme and others

... continued

- Allows users to annotate with phenotypes (hair colour, nicotine dependence...)

... continued

- Allows users to annotate with phenotypes (hair colour, nicotine dependence...)
- Everybody can download everything

... continued

- Allows users to annotate with phenotypes (hair colour, nicotine dependence...)
- Everybody can download everything
- So far: 81 genotypings and 207 users

Conclusions

- Open GWAS are the future of personalised medicine

Conclusions

- Open GWAS are the future of personalised medicine
- It's in the hands of users to make or break the situation

Conclusions

- Open GWAS are the future of personalised medicine
- It's in the hands of users to make or break the situation
- Chance to take science into our own hands

Future of openSNP

- We've won the PLoS/Mendeley Binary Battle

Future of openSNP

- We've won the PLoS/Mendeley Binary Battle
- Trying got some funding to get more people (who are willing to share) genotyped

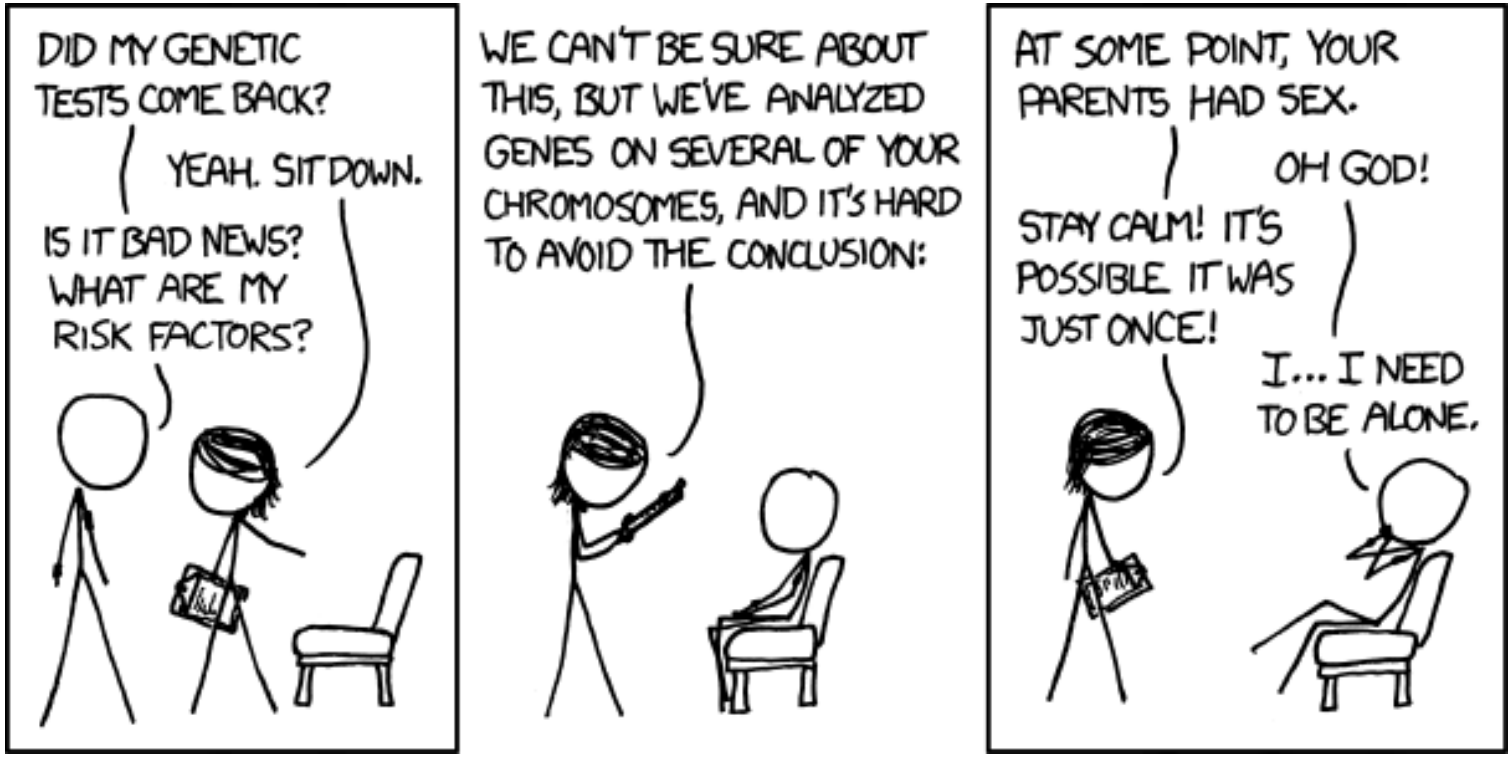
Future of openSNP

- We've won the PLoS/Mendeley Binary Battle
- Trying got some funding to get more people (who are willing to share) genotyped
  - Details on this will be released at the start of the next year

Future of openSNP

- We've won the PLoS/Mendeley Binary Battle
- Trying got some funding to get more people (who are willing to share) genotyped
  - Details on this will be released at the start of the next year
- Constantly improving the project (and are happy if somebody wants to help)

The end



Thanks for listening. Any questions?

References

Do *et al.* (2011) Web-Based Genome-Wide Association Study Identifies Two Novel Loci and a Substantial Genetic Component for Parkinson's Disease. PLoS Genetics 7(6): e1002141. doi:10.1371/journal.pgen.1002141

Eriksson *et al.* (2010) Web-Based, Participant-Driven Studies Yield Novel Genetic Associations for Common Traits. PLoS Genet 6(6): e1000993. doi:10.1371/journal.pgen.1000993

Kogan, *et al.* (2011): Thin-slicing study of the oxytocin receptor (OXTR) gene and the evaluation and expression of the prosocial disposition. Proceedings of the National Academy of Sciences

Sladek *et al.* (2007): A genome-wide association study identifies novel risk loci for type 2 diabetes. Nature 445 (7130): 881-5.

The Wellcome Trust Case Control Consortium (2007): Genome-wide association study of 14,000 cases of seven common diseases and 3,000 shared controls. Nature 447: 661-678.