From personal genomics to personal health

Daniel MacArthur

Wellcome Trust Sanger Institute

1000 Genomes Project Consortium

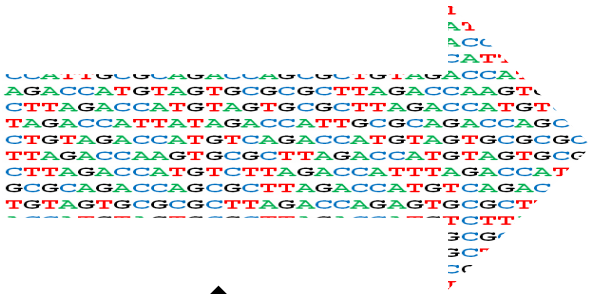
Genomes Unzipped Project

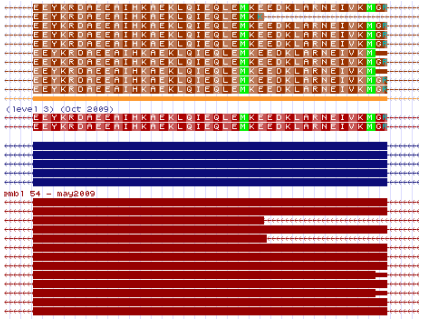
wellcome trust sanger institute

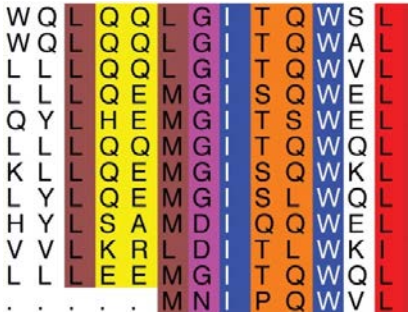
A THOUSAND GENOMES

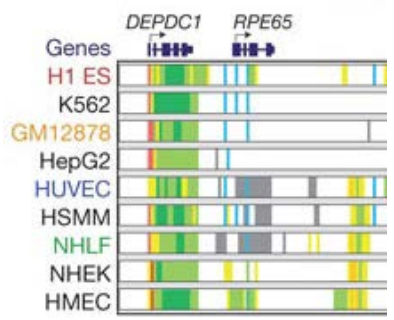
genomes unzipped public personal genomics

The sequence-function intersection







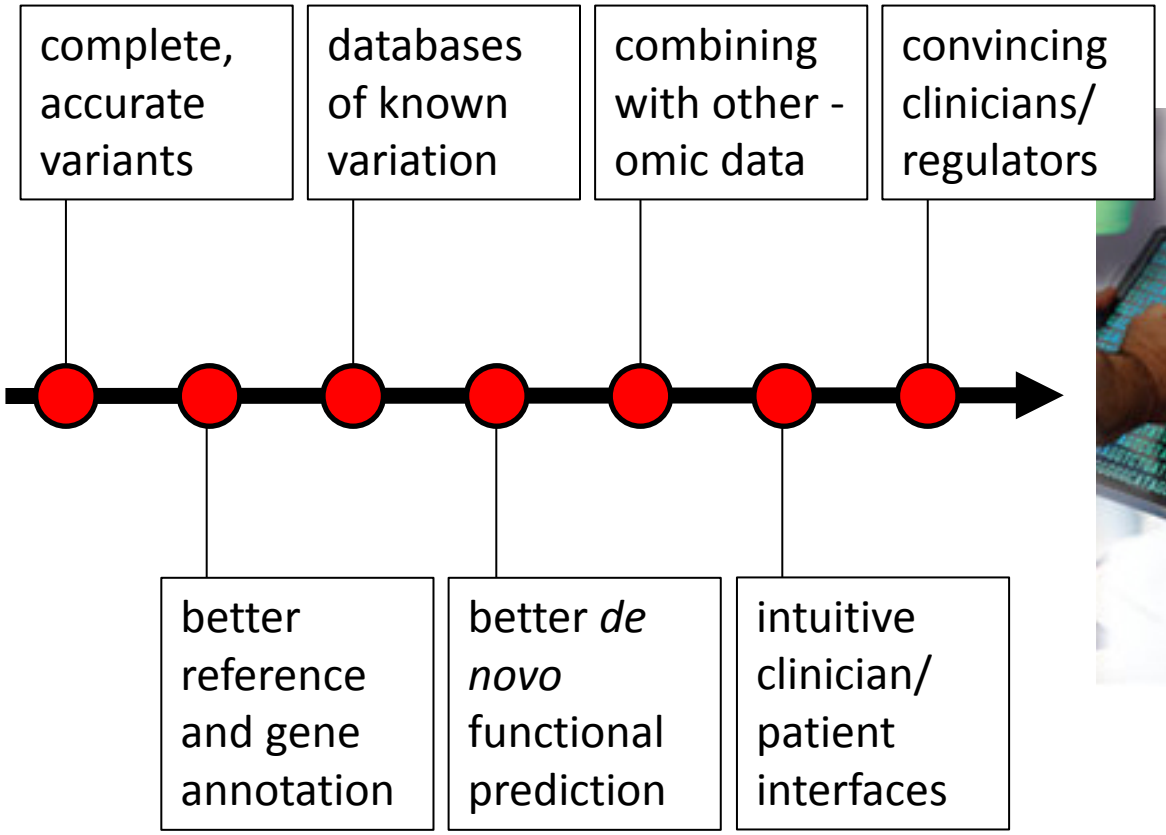


How will genomics transform healthcare?

- Carrier testing for severe recessive disease mutations (prenatal counselling)
- Cancer sequencing to determine prognosis/treatment
- Pharmacogenomics and nutrigenomics
- Risk prediction for complex diseases: both common and rare variants
- **Routine inclusion of *useful* genetic results in healthcare decisions**

Getting from here to there

AGCGCTGGAT
AGGCTCGGAT
CGGCTAGAGA
TTGCTAGCCA
TATAGGGATC
GCCTGGATAG
GCCGATCGGC
TAGGATCTAC
CATTAGGATA



Data-sets discussed

Functional annotation group

Lumigenix™

Counsyl

deCODE

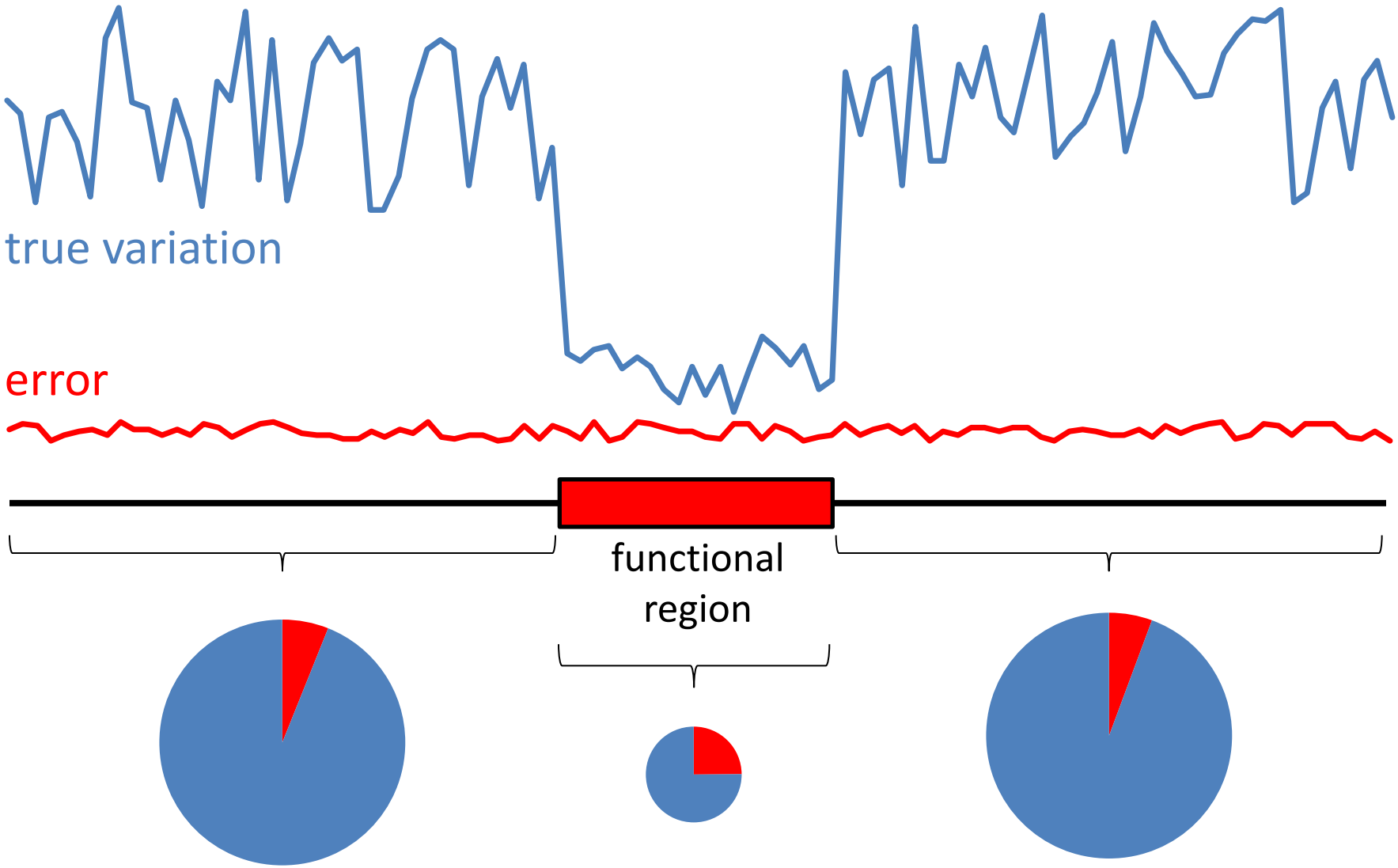
23andMe

Genomes Unzipped

Sequencing is not perfect

Errors and challenges ahead in the functional interpretation of sequence data

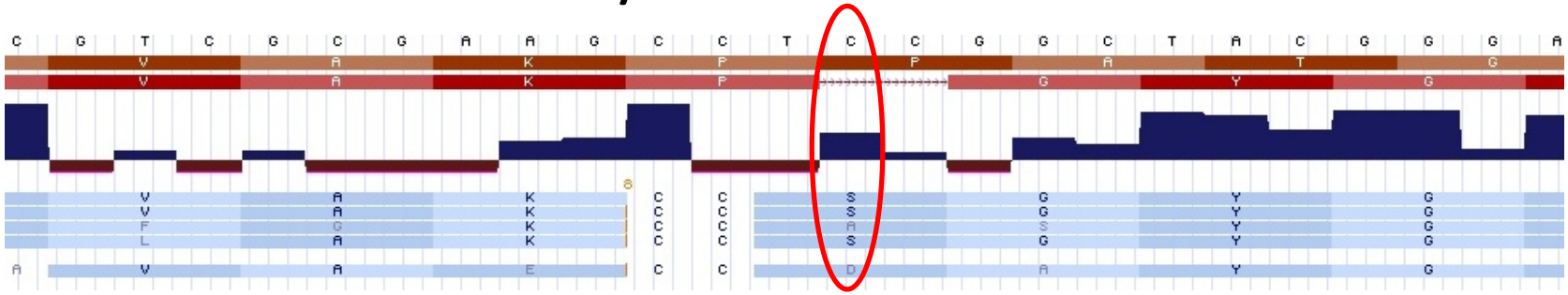
Functional variants are enriched for artefacts



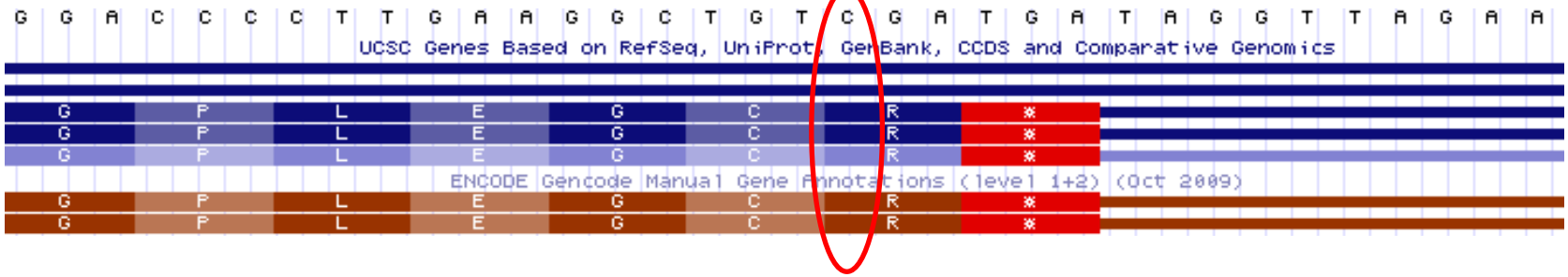
Hunting mistakes

Sequencing/genotype calling error

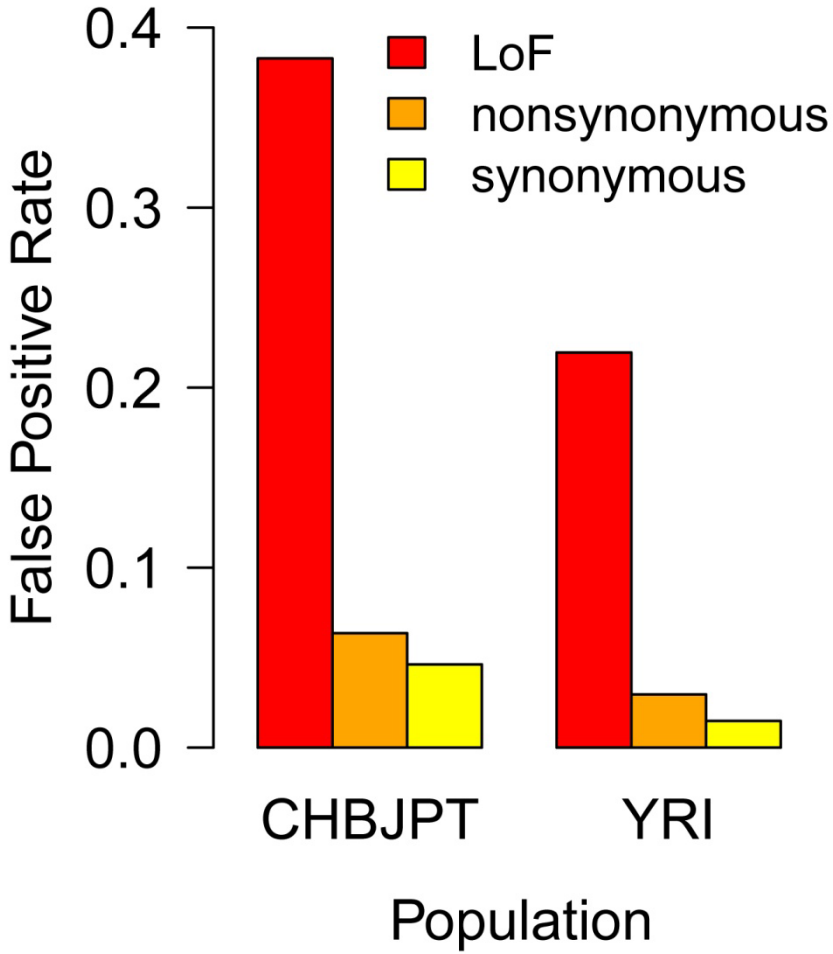
Reference/annotation artefact



Incorrect functional prediction



Functional variants are massively enriched for sequencing errors...

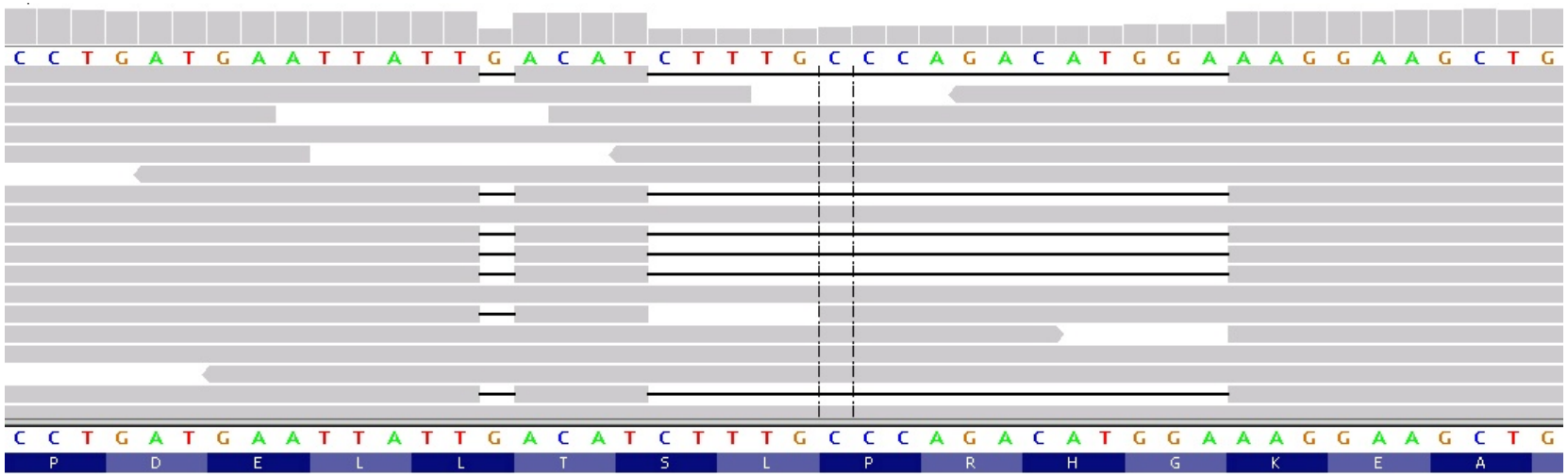


- Final low-coverage calls vs Illumina 1KGP12 chip
- Errors are mostly (but not entirely!) removed by new read-based filters

...and also massively enriched for annotation artefacts

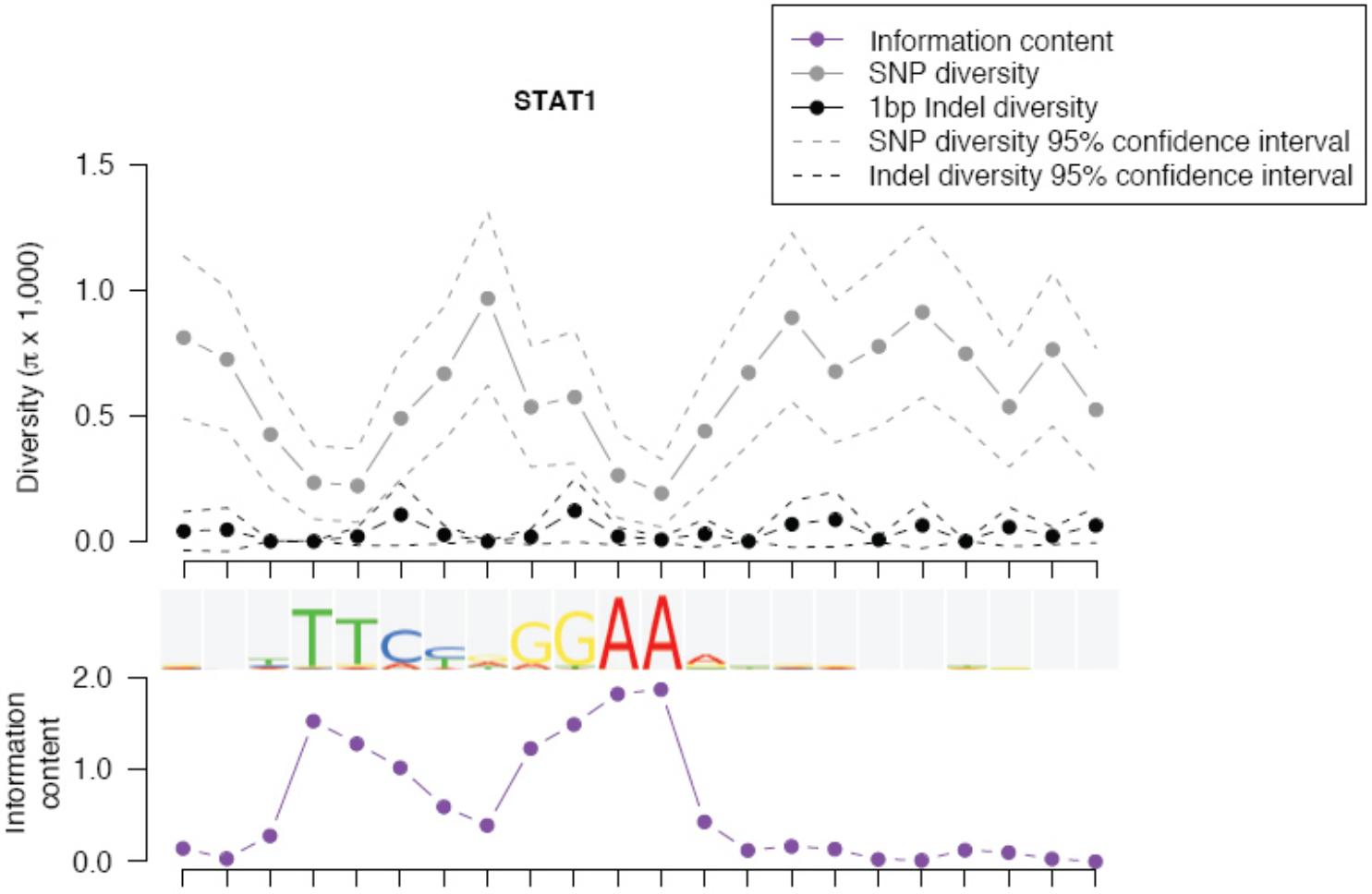
- Complete manual reannotation of ~700 LoF-containing genes by HAVANA team
- Developed automated pipeline to find other common errors
- Identified many artefacts:
  - 244 likely reference errors
  - 566 likely annotation errors
  - 313 variants unlikely to cause real LoF
- Errors fixed in Gencode, submitted to GRC

Predicting function will require including all variants



- Two apparent frameshift indels in the *CASP8AP2* gene (one 17 bp, one 1 bp) are in fact on the same haplotype
- Overall effect is in-frame deletion of six amino acids

Moving beyond coding regions



Mark Gerstein, Yale

Translating genomics

Some lessons from the personal genomics industry

Accurate genome interpretation is just the beginning

- Most individuals (including clinicians) have very little understanding of genetics
- We need good interfaces for effectively conveying genetic data to non-experts
- To date, academics have done a fairly poor job of building such interfaces


Lumigenix™


Counsyl


Navigenics®

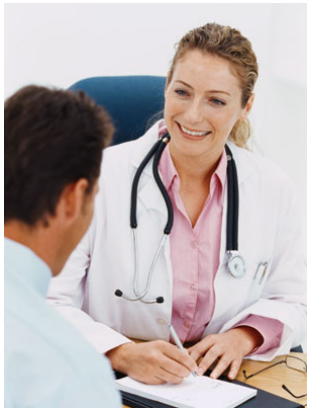

PATHWAY GENOMICS™
The Value of Knowing

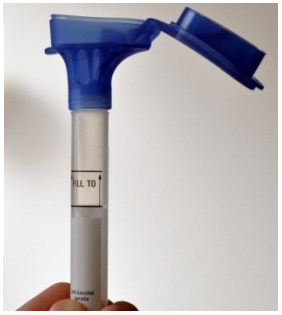

deCODE
ME


23andMe


Knome®

How it works

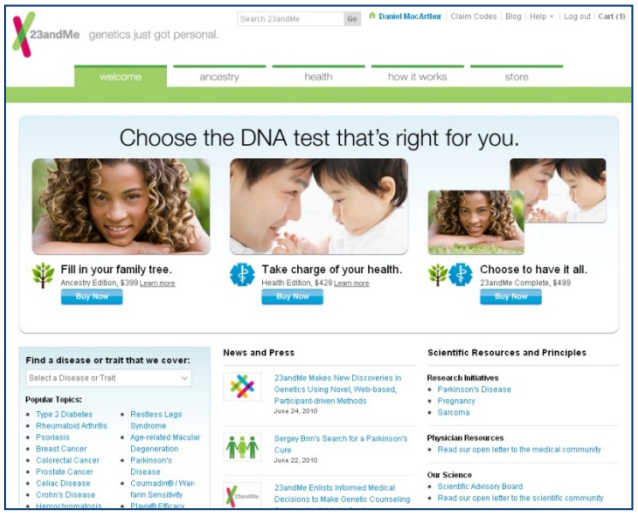


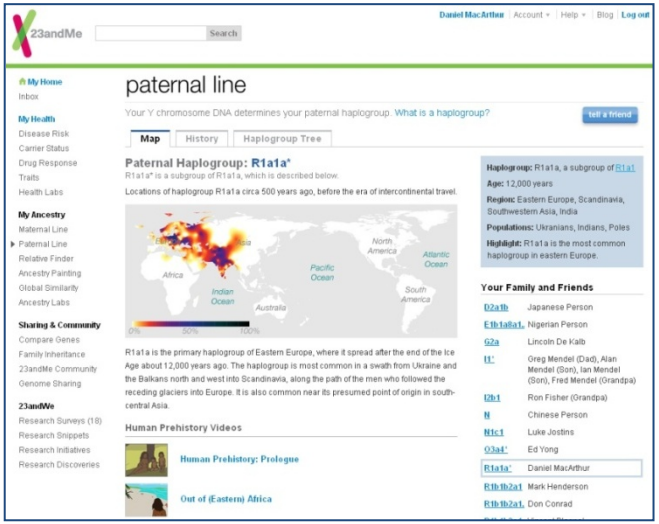


FedEx

FedEx

4-6 weeks





How much is your genome worth?

1 million SNPs
$99 + $9/month
disease/ancestry

illumina®

whole genome
$9,500
raw data only

genomes unzipped

public personal genomics

www.genomesunzipped.org

genomes unzipped
public personal genomics

Home | Project | Members | Resources | Data/Code | Browser | Subscribe | Contact

Search

« Last chance to submit comments to the FDA about DTC genetics

My Genome Online - A Challenge To You »

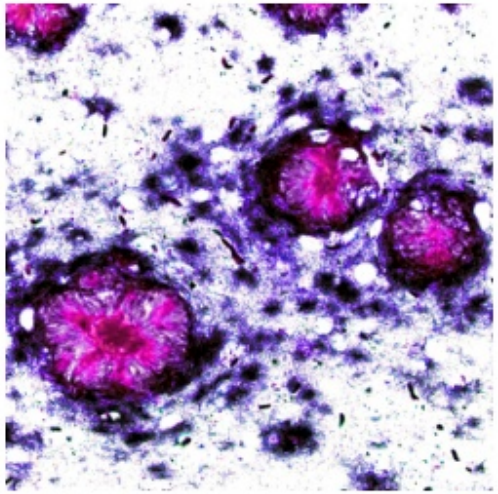
Calculating your Alzheimer's risk

Edit

13/05/2011

Categories: Disease Risk and DIY

Written by Luke Jostins



For many diseases we have very little ability to determine who is at high or low risk; the risk factors are unreplicated, complicated, or understudied. However, for other diseases we can do much better. Alzheimer's disease is a form of senile dementia that is characterised by abnormal clustering of proteins in the brain (right). We know a number of important risk factors for Alzheimer's, and knowing your own risk factors may seriously change your estimate of the chance of developing the disease. But how can you calculate this risk?

About

Genomes Unzipped is a group blog providing expert, independent commentary on the personal genomics industry.

About The Project

About The Contributors

Subscribe

RSS

Email

Twitter

Recent Posts

Cracking non-coding variation, carrying cystic fibrosis, and more

Alzheimer's prediction

Notes on the evidence for extensive RNA editing in humans

www.genomesunzipped.org

www.genomesunzipped.org

chr	pos	rsID	DGM001	LXJ001	DBV001	CFW001	KIM001	VXP001	JCB001	JXA001	JKP001	DFC001
chr2	1000713	rs6707252	AA	AG	AG	AG	AA	AA	AG	AA	AA	AG
chr2	1003294	rs11894899	GG	GT	GG	GG	GG	GG	GT	GG	GT	GG
chr2	1004757	rs7574097	GG	GG	GG	GG	AG	GG	GG	AG	AG	GG
chr2	1004837	rs7574207	GG	AG	GG	GG	AG	GG	AG	AG	AA	GG
chr2	1009753	rs9309723	CC	CT	CT	CT	CC	CC	CT	CT	CC	CT
chr2	1017830	rs4971336	AA	CC	AC	AC	AC	AA	CC	CC	CC	AC
chr2	1027731	rs8179874	AA	AG	AG	AG	AA	AA	AG	AA	AA	AG
chr2	1027783	rs4971329	AA	CC	AC	AC	AA	AC	CC	CC	CC	AC
chr2	1033352	rs4246558	CC	CC	CC	CC	CC	CC	CC	CT	CC	CC
chr2	1034617	rs4488692	AA	AG	AG	AG	AA	AA	AG	AA	AA	AG
chr2	1036996	rs9309724	AA	AG	AA	AA	AA	AG	AG	AA	AG	AA
chr2	1038118	rs4629189	AA	AG	AG	AG	AA	AA	AG	AA	AA	AG
chr2	1048688	rs4640416	CC	TT	CT	TT	CC	TT	TT	CT	CT	TT
chr2	1054700	rs7426276	GG	GG	GG	GG	AG	GG	GG	AG	GG	GG
chr2	1054885	rs4555377	AA	AG	AG	GG	AA	AA	AG	AA	AA	GG
chr2	1059949	rs4586665	TT	CC	CC	CT	TT	CC	CC	CT	CT	CC
chr2	1069320	rs4971432	TT	CT	CT	CT	TT	TT	CT	TT	TT	CC
chr2	1071594	rs10865542	TT	GT	GT	GT	TT	TT	GT	TT	TT	GG
chr2	1073995	rs7578498	TT	CT	CT	TT	TT	CT	CT	CT	CT	TT
chr2	1081787	rs11686452	TT	CC	CC	CT	TT	CC	CC	TT	CC	CC
chr2	1087795	rs6716119	GG	AG	AG	AG	GG	GG	AG	GG	GG	AA
chr2	1089597	rs4971441	TT	CC	CC	CT	TT	CC	CC	TT	CC	CC
chr2	1091080	rs4482521	CC	AA	AA	AC	CC	AA	AA	CC	AA	AA
chr2	1091127	rs12469446	AA	GG	GG	AG	AA	GG	GG	AA	GG	GG
chr2	1106648	rs4971463	GG	AG	AG	AG	GG	GG	AG	GG	GG	AA
chr2	1109365	rs7565112	TT	CC	CC	CT	TT	CC	CC	TT	CC	CC
chr2	1110566	rs4971465	AA	GG	GG	AG	AA	GG	GG	AA	GG	GG
chr2	1118570	rs4491751	CC	AC	AC	AC	CC	AC	AC	CC	CC	AA
chr2	1133477	rs4998208	AA	AG	GG	GG	AA	AG	AG	GG	AG	GG
chr2	1140095	rs7579535	CC	CC	CC	AC	CC	CC	CC	AC	AC	CC
chr2	1142434	rs4246561	TT	TT	CT	TT	CT	TT	TT	CT	CT	TT
chr2	1143552	rs4971396	GG	GG	GT	GT	GG	GG	GG	TT	GT	GG
chr2	1147068	rs4971397	TT	TT	GT	GT	GT	TT	TT	GG	GG	TT
chr2	1189920	rs4461289	AG	GG	AG	GG	AG	AA	GG	GG	AA	GG
chr2	1193337	rs4446093	GG	AG	AG	GG	GG	GG	AG	GG	GG	AA
chr2	1194796	rs6751090	CC	CT	CT	CC	CC	CC	CC	CC	CC	CC
chr2	1206395	rs4971418	AG	GG	AG	AA	AG	AA	AG	AG	AA	GG
chr2	1257125	rs7568265	CT	CT	TT	TT	CC	TT	TT	CT	TT	TT
chr2	1258693	rs7369772	CC	CC	CC	CC	CT	CC	CC	CT	CC	CC
chr2	1262038	rs11674954	TT	TT	TT	TT	TT	GG	TT	TT	GG	TT
chr2	1263027	rs7424980	CC	TT	CT	CC	CC	TT	CT	CC	TT	TT

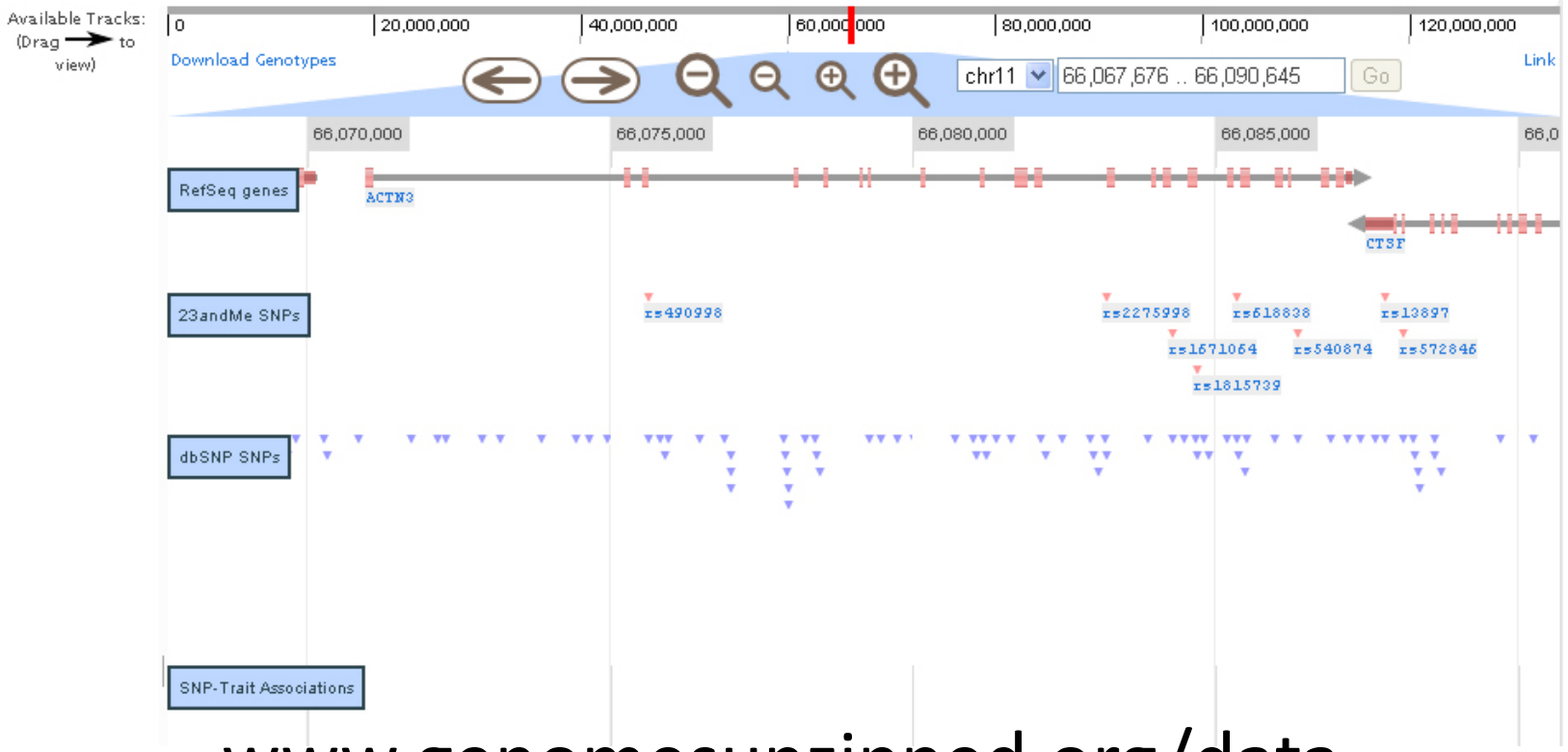
www.genomesunzipped.org/data

genomes unzipped
public personal genomics

Home | Project | Members | Resources | Data/Code | Browser | Subscribe | Contact Us | Search

The Genomes Unzipped Browser

The Unzipped Browser was developed using JBrowse by Joe Pickrell and Luke Jostins. Click here for information on how to use it. All chromosome positions are relative to build 36.



www.genomesunzipped.org/data

Lessons learned so far

- Raw data: extremely accurate for SNP genotyping tests
  - same individual, 3 tests: **1.4-4.5 errors/100K SNPs**
  - 23andMe trios: **2.6-4.5 errors/100K SNPs**
- Risk prediction: good, but could be improved
  - some potentially worrying variation in algorithms between companies
  - larger variation in background risk estimates
  - SNP lists can be overly conservative/out of date
- Interfaces: generally intuitive, responsible

Risks and benefits

Table 3. Primary Outcome Measures before and after Receipt of Results of Genetic Testing for 2037 Subjects Who Completed Follow-up.*

Outcome Measure	Baseline Score	Follow-up Score	P Value†
Anxiety	35.2±9.6	34.6±10.0	0.80
Dietary fat intake	16.0±7.9	15.2±7.5	0.89
Exercise‡	28.6±23.0	28.6±22.9	0.61

* Plus–minus values are means ±SD. The assessment tools and ranges of scores for each category are listed in Table 2.
† All P values were calculated with the use of the Wilcoxon signed-rank test after adjustment for covariates.
‡ A total of 1943 subjects were included in this analysis.

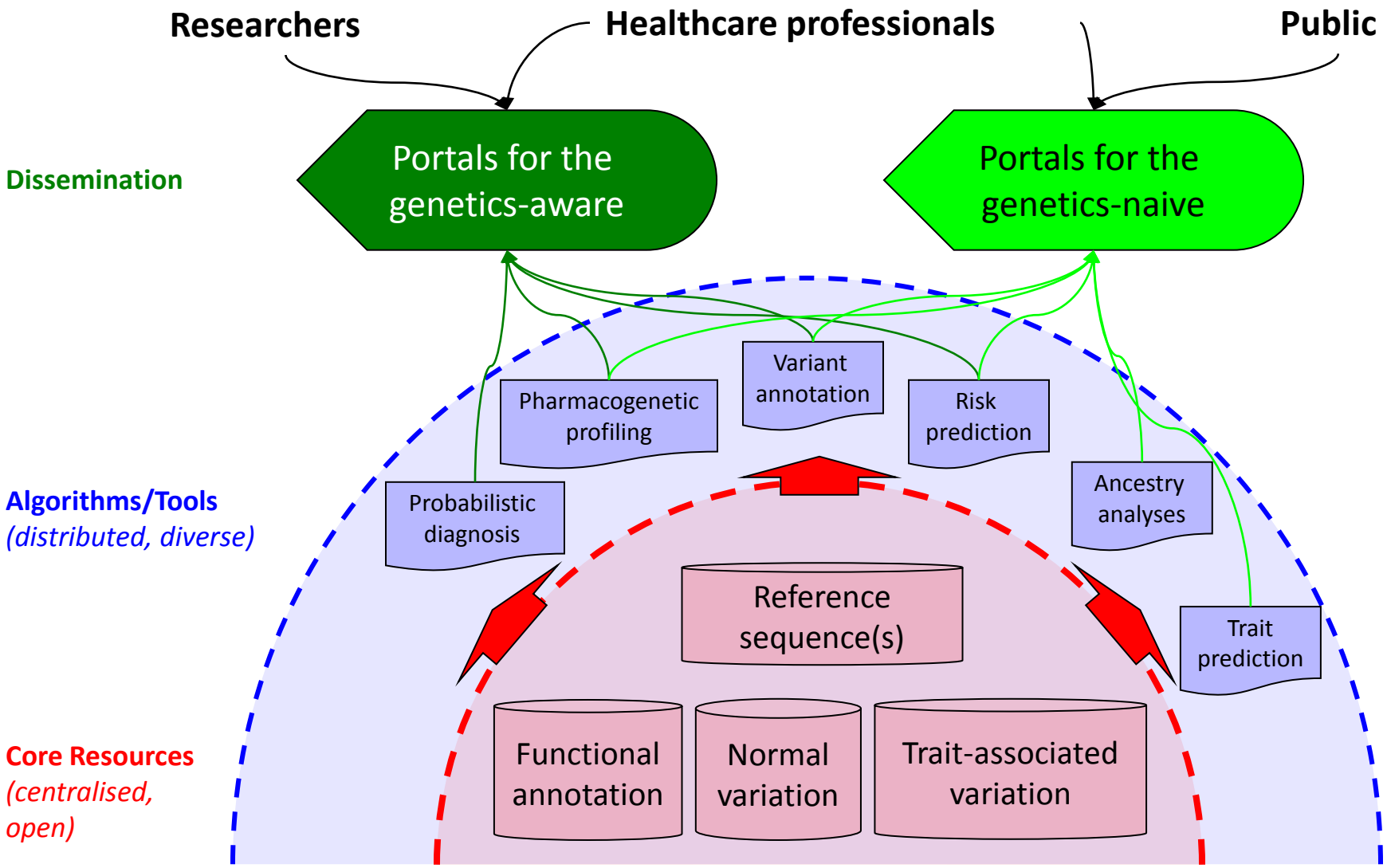
- Early adopters are well-educated about genetics
- Genetic risk predictions don't increase anxiety...
- ...but they don't seem to positively change behaviour either
- More study needed, especially for more serious conditions and later adopters

Bloss *et al.* 2011. *NEJM* **473:**43–49

Ethical challenges

- **Clinicians as gatekeepers *vs* empowering individuals**
  - doctors more likely to create lifestyle change?
  - can health system survive with gatekeeper model?
  - importance of individual autonomy
- **Regulation *vs* innovation**
  - some companies offer bogus products
  - strong regulation would hinder innovation
  - could just better enforce existing legislation

Moving forward: what happens next?



adapted from Matt Hurles

Thanks!

1000 Genomes Functional Annotation

Chris Tyler-Smith, **Sanger**
Suganthi Balasubramanian, Lukas Habegger, Mark Gerstein, **Yale**
Adam Frankish and Jennifer Harrow, **Sanger**
James Morris, Luke Jostins and Jeff Barrett, **Sanger**
Ni Huang, Klaudia Walter, Don Conrad and Matt Hurles, **Sanger**
Kees Albers and Richard Durbin**, Sanger**
Joe Pickrell and Jonathan Pritchard, **U. Chicago**
Stephen Montgomery and Manolis Dermitzakis, **U. Geneva**
Mark DePristo and Eric Banks, **Broad Institute**

Genomes Unzipped

Luke Jostins, Dan Vorhaus, Kate Morley, Caroline Wright,
Vincent Plagnol, Jeff Barrett, Joe Pickrell, Don Conrad,
Carl Anderson, Ilana Fisher, Jan Aerts

wellcome trust sanger institute

A THOUSAND GENOMES

genomes unzipped
public personal genomics

Daniel MacArthur

www.genomesunzipped.org
Twitter: @dgmacarthur