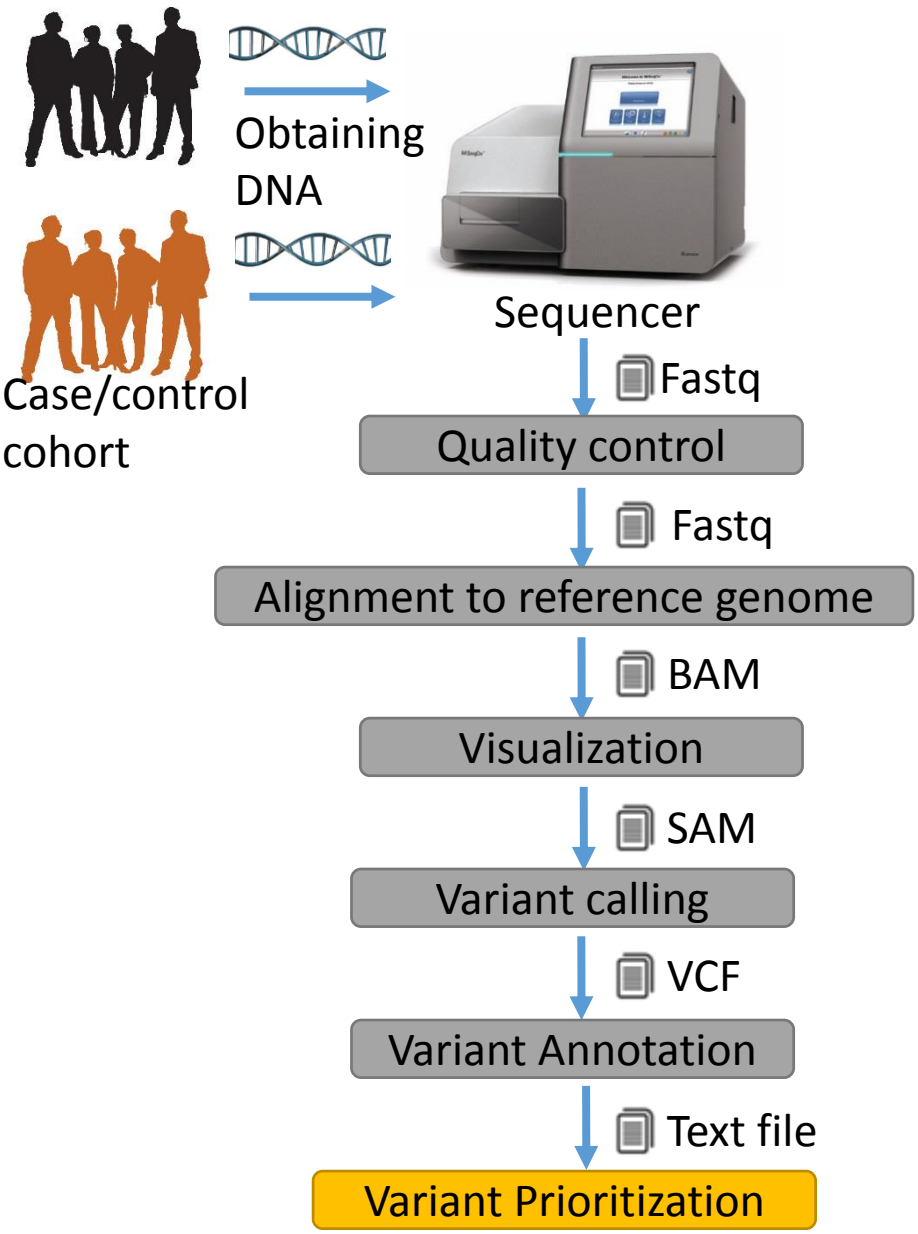
Variant prioritization

- The pipeline so far
- Introduction to variant prioritization
- A proposed prioritization scheme
- Population genetic approach to variant prioritization
- Why “bad mutations” are sometimes helpful.
- Beyond genetic variants

The pipeline so far - Files



Fastq - storing the nucleotide sequence and its corresponding quality scores

BAM (Binary Alignment/Map) - a binary format for storing sequence data

SAM (Sequence Alignment/Map format) - a tab delimited text format consisting of the aligned data

VCF (Variant Call Format) – a text file, contains information about a position in the genome.

The pipeline so far - Tools

Obtaining DNA

Case/control cohort

Sequencer

Fastq

Quality control

Fastq

Alignment to reference genome

BAM

Visualization

SAM

Variant calling

VCF

Variant Annotation

Text file

Variant Prioritization

FastQC

Bowtie2

BAM-TO-SAM, IGV

GATK,SAMtools

Accessible through Galaxy

Galaxy

Annovar, wAnnovar, Variant Effect Predictor, dbSNP, ExAC browser

Why prioritize variants?

Most of the SNPs are probably not causative. Can we know which of them are?

1) Biological validation

- Re-sequencing with *different* methods (e.g., Sanger)
- But you cannot do that for hundreds of SNPs.

2) Experimental valuation on animal models

- You can test whether gene knock-out experiment will produce a similar phenotype.
- But its expensive, slow, and has its own problems.

3) Bioinformatic validation

- A proxy to the reality.
- Cheap! but may be wrong!

A simplistic approach to variant prioritization

We can prioritize variants in several ways giving more weight to SNPs based on certain characteristics. To have an idea where to look for you are required to have an *a priori* hypothesis or knowledge of the disease you study.

For example, in psychiatric disorders, the priority is rare variants (at the time being). Therefore, minor allele frequency is considered as one prioritization characteristic. Variants that show expression in the brain are also favored.

In diabetes, we look at functionality.

What is the problem with this approach?

A proposed prioritization scheme

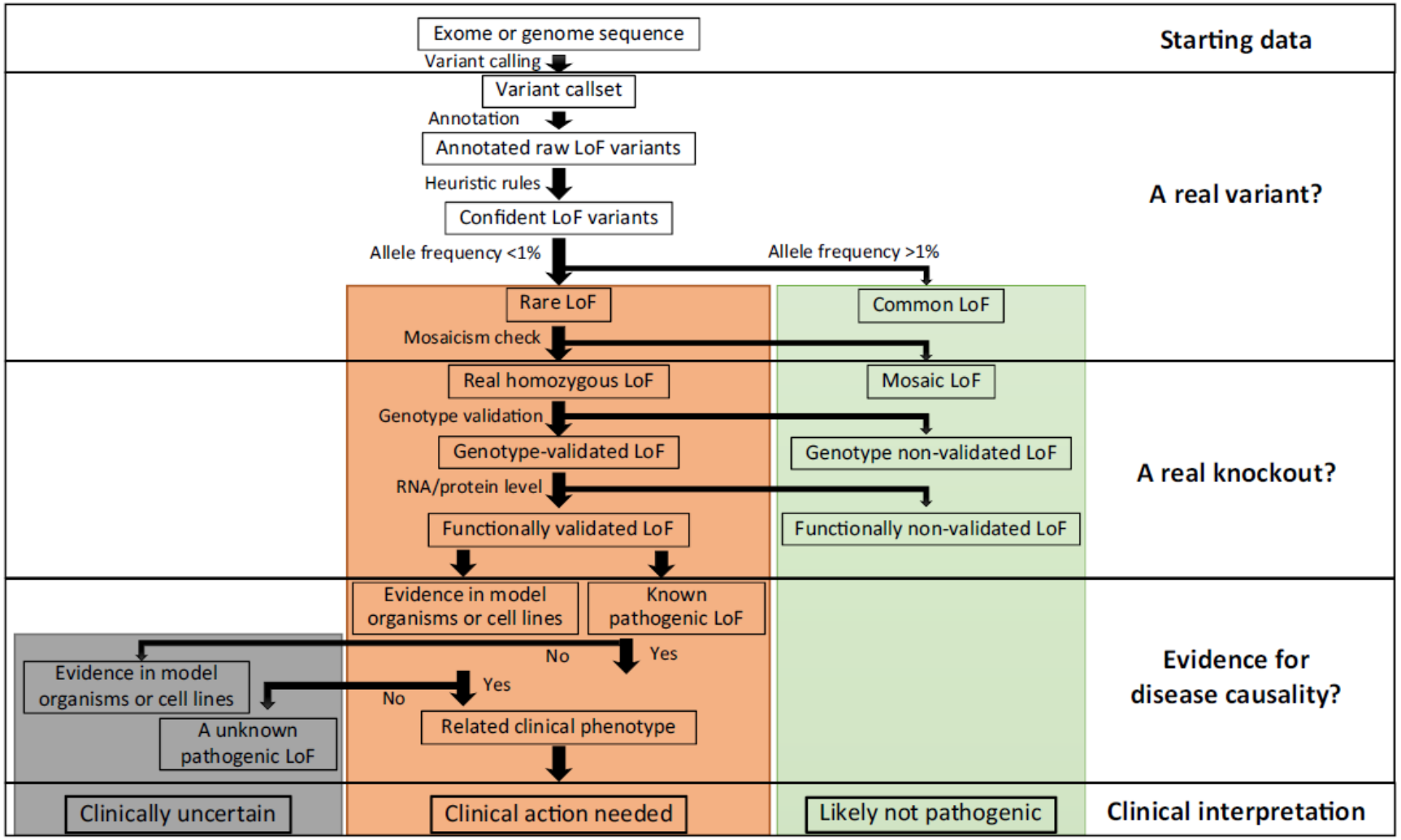
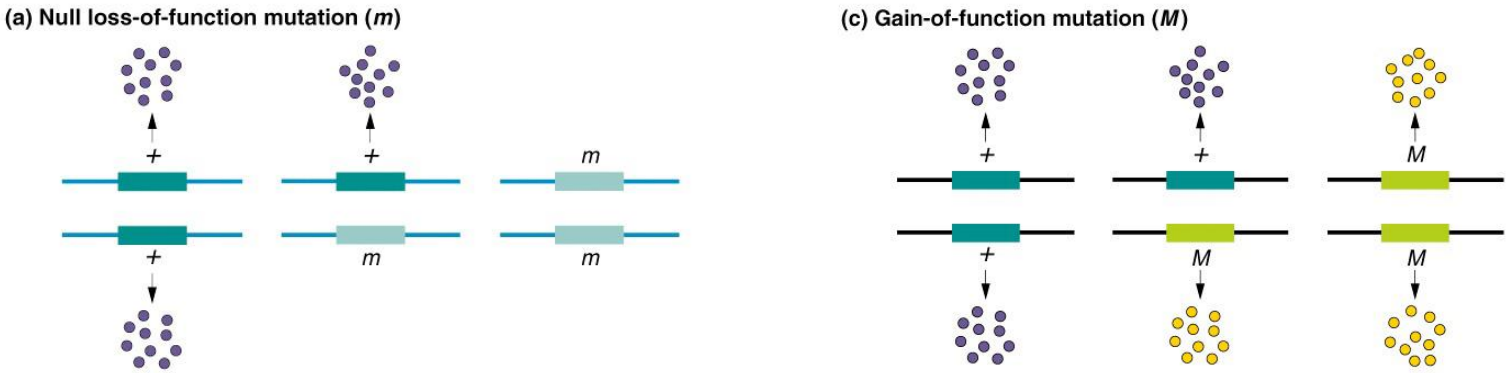


Figure 1. In increasing order of complexity, decisions must be made about whether or not (yes/no) (i) the variant itself is real, (ii) really leads to the knockout of the gene, and (iii) there is evidence that it is likely to cause disease. As a result, the interpretation may be that clinical action is needed, that the variant is not likely to be pathogenic, or that the clinical implications are uncertain. Abbreviation: LoF, loss of function.

Narasimhan VM, et al. 2016. Human Knockout Carriers: Dead, Diseased, Healthy, or Improved? Trends Mol Med.

A proposed prioritization scheme

a real variant?



- LoF variants are genetic variants that are predicted to severely disrupt the function of human protein-coding genes and are therefore prime candidates for follow-up.
- LoFs of some genes in humans cause genetic diseases.
- In other genes the consequences depend on the genetic background or environment
- Some LoFs may have no detectable effect, or may even be beneficial.

Significant difficulties remain with:

a) Calling DNA variants (Why?)

b) The annotation (Why?)

http://www.mun.ca/biology/desmid/brian/BIOL2250/Week_Five/1GeneW5b.html

There are more problems that affect the reliability of our LoF inference

- Calling insertions and deletions (indels) is still problematic, particularly in exome sequencing.

- Large structural variants are less frequent than indels and difficult to infer due to differences in **coverage** as well as their inability to span **breakpoints**.

- Non-SNP variants cause LoF and may still have high error rates.

Solution

- A validation of variants of interest using Sanger sequencing/Sequenom genotyping is ALWAYS needed.

Predicting the LoF effect on a specific transcript/protein and the phenotype is challenging

- Measuring transcript levels is helpful in the case of large LoF (transcript is fully deleted), however small LoF may produce nonsense transcript that may not reduce the protein level.

- Alternative splicing may lead to partial LoF variants that affect only some transcripts so functional protein may be produced by the remaining ones.

- It is currently impossible to effectively measure the functional importance of different transcripts for most genes. So, partial LoF variants can cause Mendelian Diseases.

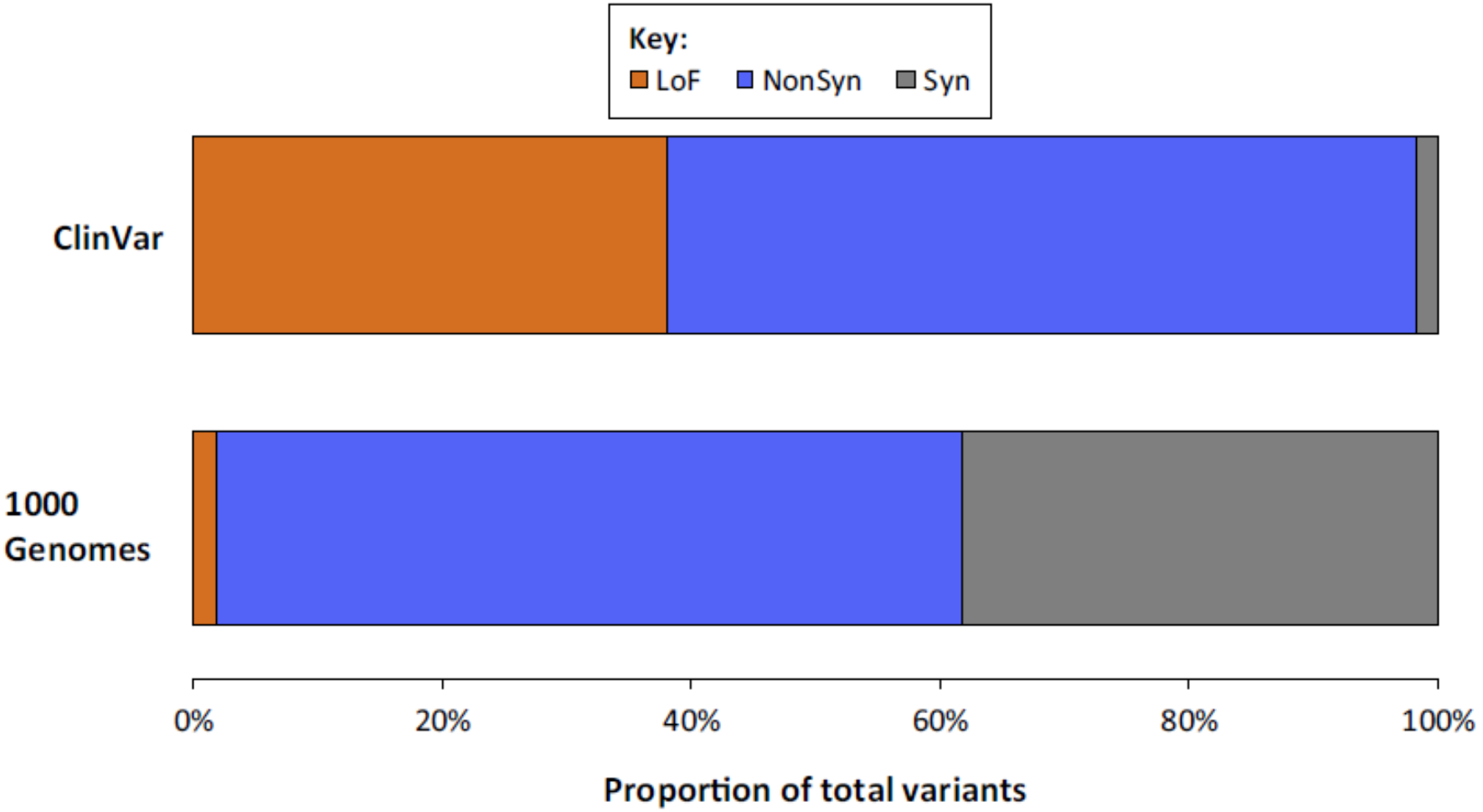
Solution

- In diagnostic setting, confirmation can be obtained by measuring the absence of the protein product/activity in the sample.

A proposed prioritization scheme

a pathogenic variant?

- Many proteins are **not essential** for good health.
- To infer the phenotypic effect of a LoF we need to examine database of disease causing SNPs.
- Traditionally, gene/variants segregating in families were identified and followed with additional patients with similar phenotypes. After assessing the mode of inheritance causality was established.
- Interpreting LoF in the Clinic requires consulting online databases like: OMIM, HGMD, ClinVar – however these are all ascertained from individuals and their penetrance is poorly understood.



Trends in Molecular Medicine

Figure 4. Proportions of Different Variant Classes in the General Population. The graph provides data from the 1000 Genomes Project, Phase 3 (lower bar), and the ClinVar database of disease-associated variants (ClinVar; upper bar). Non-synonymous variants (NonSyn; blue) are abundant in both samples; synonymous variants (Syn; grey) are abundant in the general population, but seldom cause disease; LoF variants are scarce in the general population but form a high proportion of ClinVar entries (LoF, orange). This shows that, although knockout variation is present at low frequency in the general population, it has a substantial impact on disease.

Narasimhan VM, et al. 2016. Human Knockout Carriers: Dead, Diseased, Healthy, or Improved? Trends Mol Med.

So what if you found a deleterious alleles?

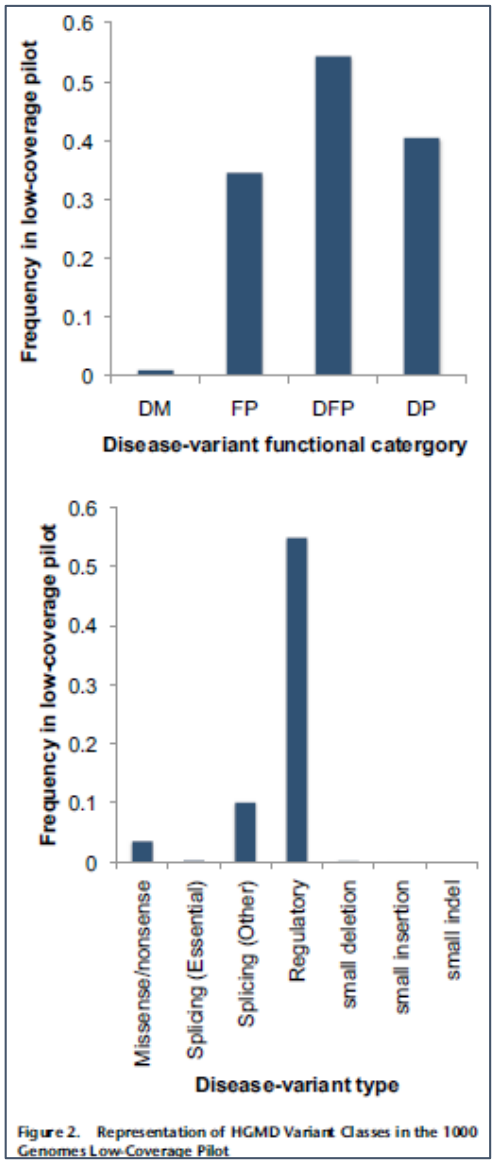


Figure 2. Representation of HGMD Variant Classes in the 1000 Genomes Low-Coverage Pilot

Xue et al. 2012. AJHG. 91, 1022–1032,

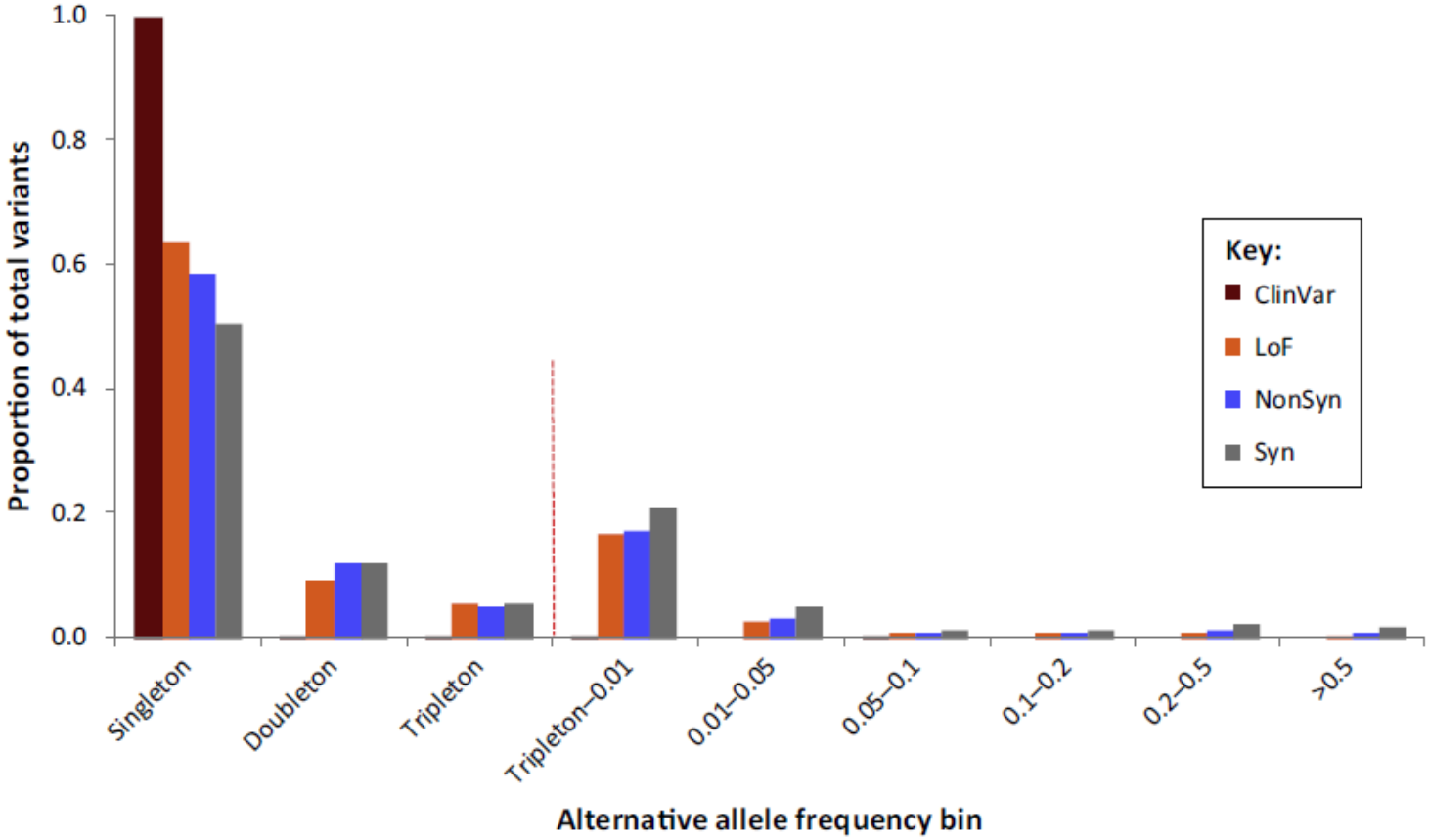
CAUTION!

On average, each healthy person carries:

- ~11,000 synonymous variants
- ~11,000 non-synonymous variants
- 250 to 300 loss-of-function variants in annotated genes
- 50 to 100 variants previously implicated in inherited disorders

1000 Genomes Project Consortium. *A map of human genome variation from population-scale sequencing.* **Nature**. 2010 Oct 28;467(7319):1061-73. PubMed PMID: 20981092

DM, disease-causing mutation in HGMD.
FP, in vitro or in vivo functional polymorphism but with no disease association reported as yet.
DFP, disease-associated polymorphism with additional supporting functional evidence with disease and that has evidence of being of direct functional importance.
DP, disease-associated polymorphism with a disease or phenotype and that is assumed to be functional, although there might not yet be any direct evidence of a functional effect.



Trends in Molecular Medicine

Figure 2. Allele Frequency Spectrum of Different Classes of Variants in the 1000 Genomes Project Data. Alleles were assigned to a bin according to their frequency in the study population, and the bins plotted in order of increasing frequency on the horizontal axis, with the functional classes being indicated by different colors within each bin. Singleton, doubleton, or tripleton variants refer to those seen only once, twice, or three times in the data, respectively. In this sample from apparently healthy populations, variants seen in disease databases such as ClinVar (ClinVar; dark red) are observed almost exclusively in single individuals. Loss-of-function variants (LoF; orange), which knock out genes and represent the most damaging functional class of variant, are also seen most often in only a single individual, although some are more frequent. Non-synonymous variants (NonSyn; blue), which change an amino acid in the protein, are on average present at higher frequency in the population, and are thus shifted towards the right-hand side of the plot. Synonymous variants (Syn; grey), which do not change an amino acid, have on average the highest allele frequencies.

Narasimhan VM, et al. 2016. Human Knockout Carriers: Dead, Diseased, Healthy, or Improved? Trends Mol Med.

OMIM - Online Mendelian Inheritance in Man

The OMIM Gene Map (https://www.omim.org/search/advanced/geneMap) and Morbid Map (http://www.mad-cow.org/00/omim_cds.html) present the cytogenetic locations of genes and disorders, respectively, that are described in OMIM.

Only OMIM entries for which a cytogenetic location has been published in the cited references are represented in the Gene Map and Morbid Map.

The OMIM Gene Map can be searched by gene symbol (e.g., "SOD1"), chromosomal location (e.g., "5", "1pter", "Xq"), or by disorder keyword (e.g., "alzheimer").

Exercise I

1. What condition is associated with the *nfkb2* and *CFTR* genes? What are the inheritance modes?

2. Which genes are associated with *Bipolar affective disorder* (*Major affective disorder 1*)?

3. What is the function of *RELA*? What conditions is it associated with?

A population genetic approach to prioritize variants

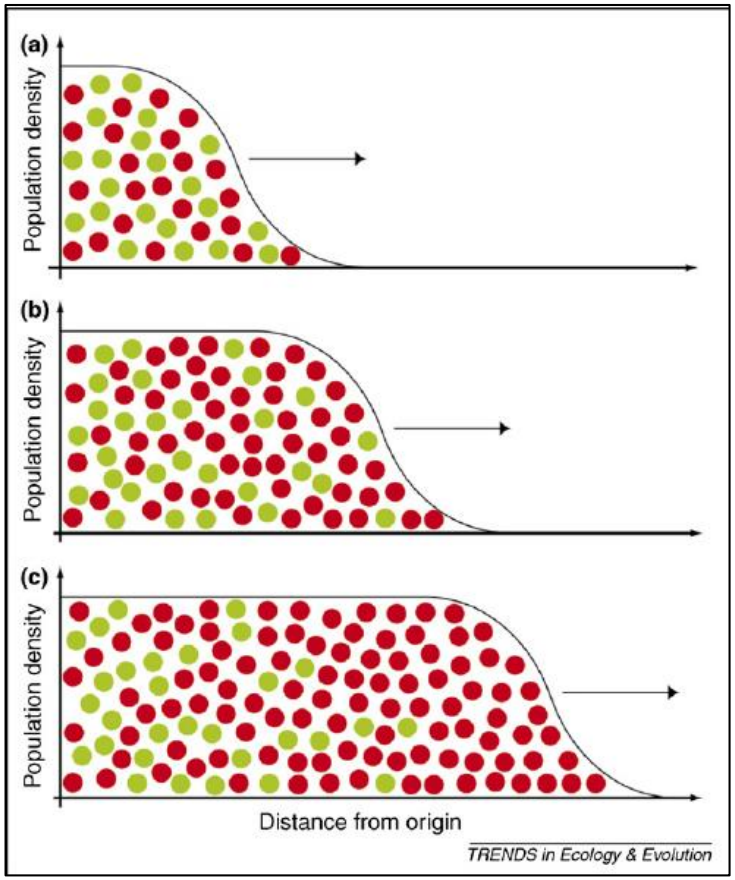
So what if you found an association?

Genetic drift leads to changes in allele frequencies causing false association

(a) Initial conditions show an equal proportion of two alleles (red and green).

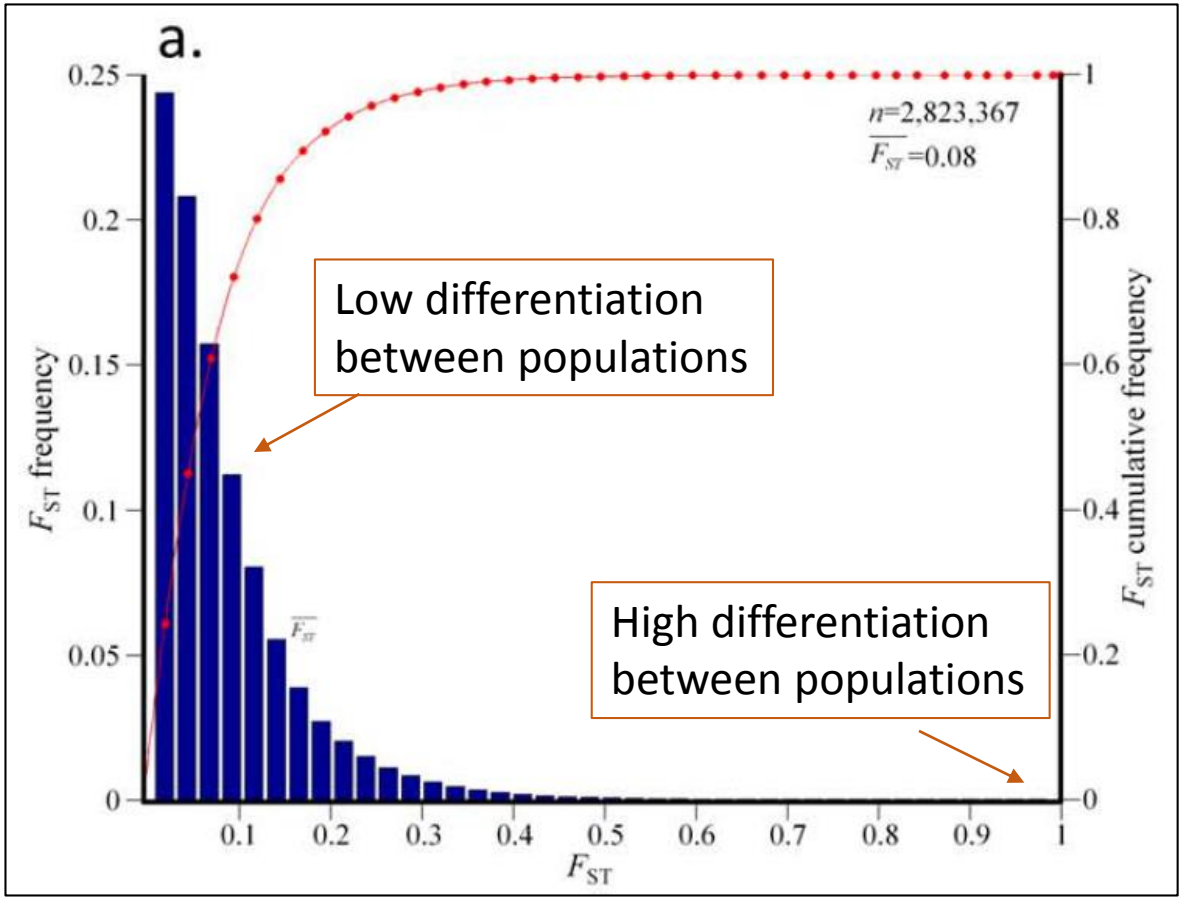
(b) The red allele increases in frequency owing to local drift.

(c) The red allele has become fixed by drift. Populations will only carry this allele.



Excoffier L, Ray N.. Surfing during population expansions promotes genetic revolutions and structuration. Trend Ecol Evol 23: 347-351

Allelic changes due to genetic drift can be misleading



Distribution of locus-specific F_{ST} in three continental populations (Europeans, Africans, and Asians). F_{ST} values were obtained for 2.8M autosomal SNPs.

Adaptation Genetic drift

Pop1: ACG**C**TCA**G**CTAGCATAG
Pop2: ACG**A**TCA**C**CTAGCATAG

F_{ST}

Variants with opposite directions of effect in populations, will have high F_{ST} and higher levels of *P*-values which may be mistaken for true association

Rare variants are the worst!

Solutions: account for population structure

Elhaik et al. 2012. Plos one. 7.
Zawistowski et al. 2014. EJHG. 22. 1137–1144

Complexity of population structure may be difficult to account for

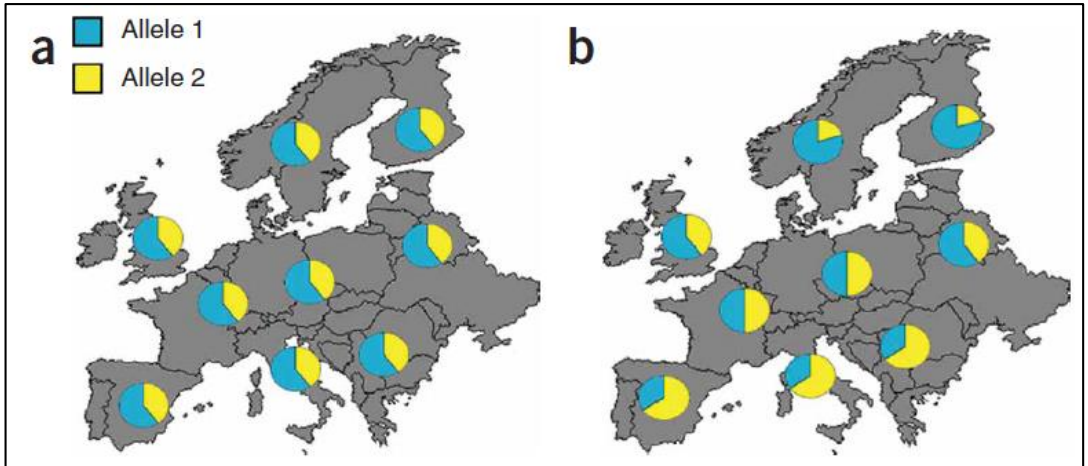


Figure 2 A complementary approach for assessing the likelihood of an association being explained by stratification. If the allele frequencies of an associated marker are similar throughout a range of closely related possible source populations, in this case, Europe (**a**), then the association is less likely to be a false positive due to stratification. If allele frequencies differ widely among closely related populations that could be represented in the study sample (**b**), then the possibility of a false positive association due to stratification must be strongly considered.

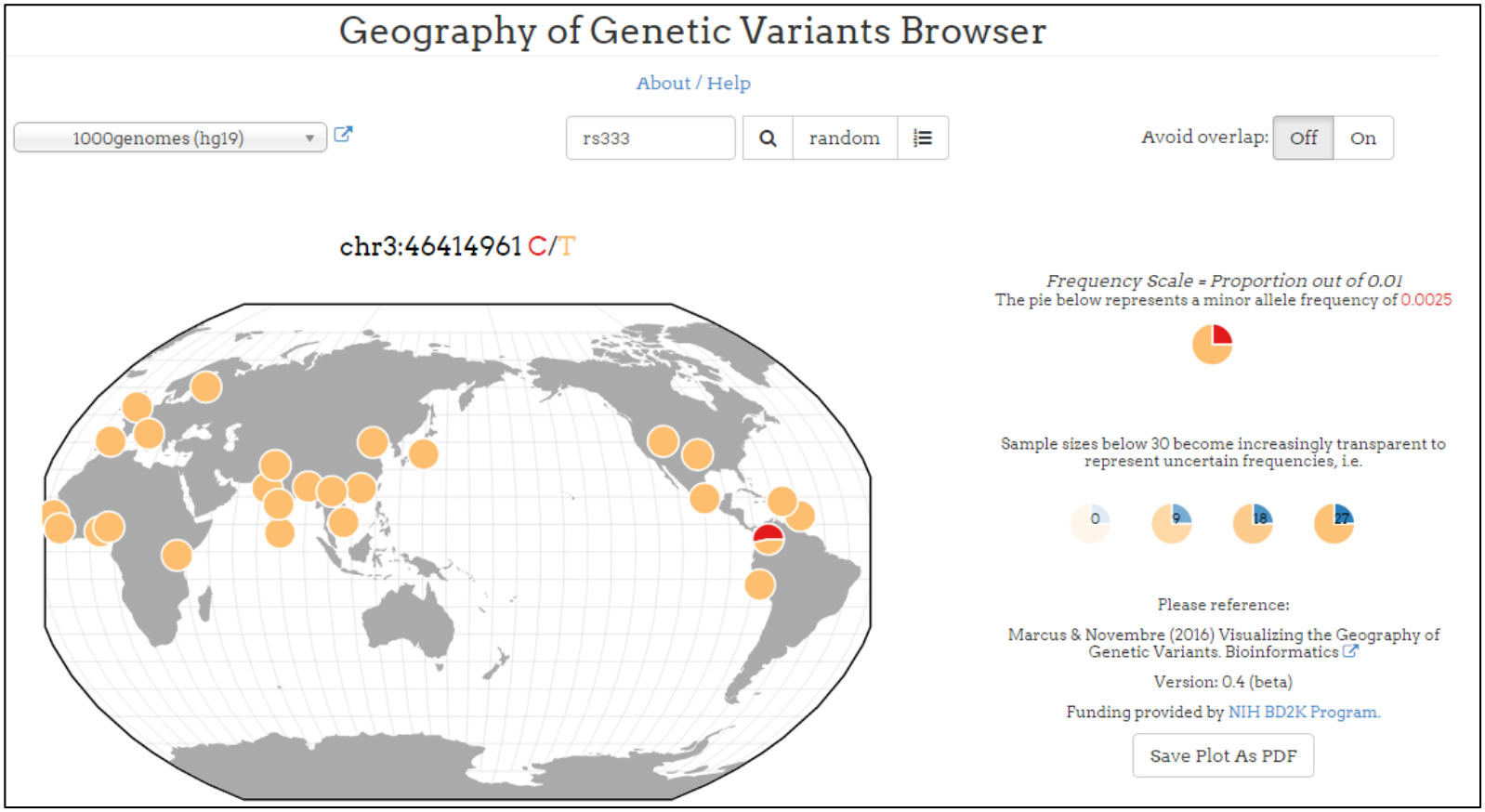
Questions to consider:

- How to define a population? (e.g., European-Americans, Scots, Ashkenazic Jews)
- How to model population structure? (e.g., PCA, Structure)
- How to account for population structure? (e.g., study design, statistical correction)

Campbell et al. Demonstrating stratification in a European American population. Nat Genet. 2005;37(8):868-72.

Global Allele Frequency Data using the Geography of Genetic Variants Browser

http://popgen.uchicago.edu/ggv/



Exercise II

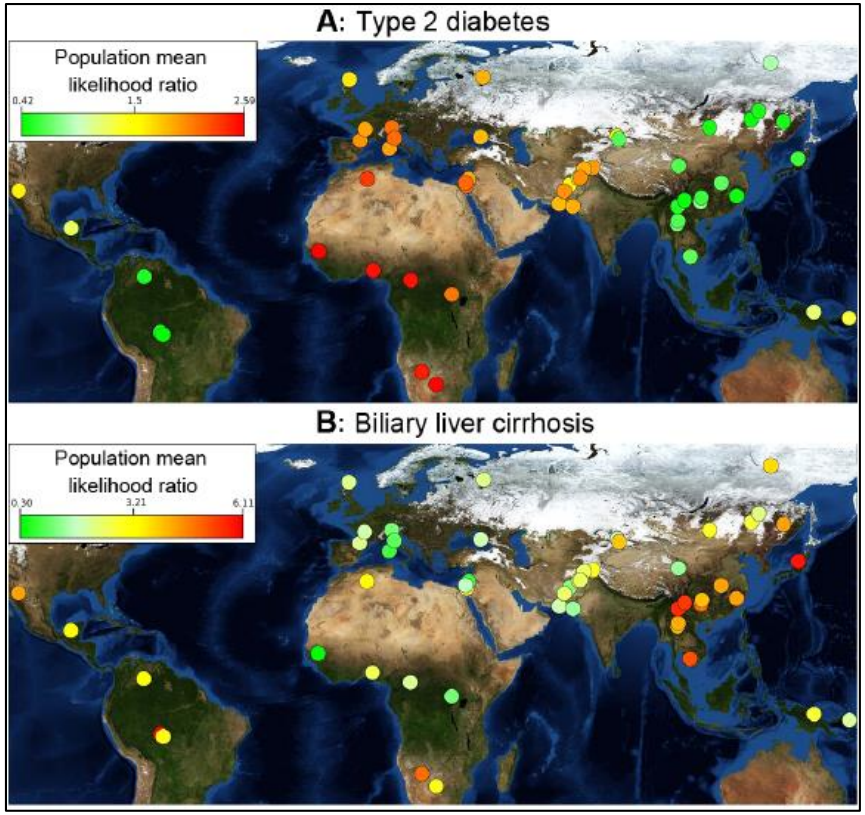
Let us annotate and prioritize the following variants:

- rs121908792
- rs1800079
- rs121908752
- rs121908760

1. Which gene/s are these variants associated with?
2. Which condition/s are these variants associated with?
3. Considering the functional annotation and their allele frequencies prioritize these variants for a follow up study (i.e., rank them by potential relevance for the condition).

Human groups exhibit real differences in risk for diseases

Disease susceptibility



Corona et al.. *PLoS Genet.* 2013;9(5):e1003447

Drug response

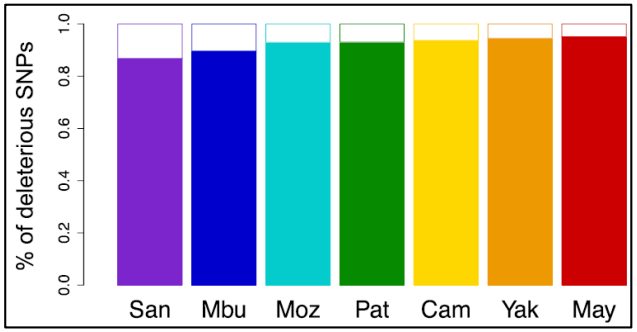
Subject group	Reference group	Effect	Reference
Chinese	Caucasian, other S. Asian	Chinese patients are at substantially lower risk than European patients for cardiovascular complications after diabetes diagnosis, whereas South Asian patients were at comparable risk. Mortality after diabetes diagnosis is markedly lower for both minority populations.	Shah et al., 2013
Hispanic	Hispanic	The prevalence of hypertension and diabetes varies significantly among Hispanics by country of origin.	Pabon-Nau, et al., 2010
S. Asian, African	Caucasian	Type 2 diabetes 3- to 6-fold more likely in Africans and Asians, depending on ethnicity	Diabetes UK reports
Asian, African	Caucasian	Endometriosis is more common in Asian women and less common in African women as compared to Caucasian women	Gerlinge et al. 2012
Chinese	Caucasian	In glaucoma, multiple observable eye parameters differ between these racial groups; suggest potentially different mechanisms in occludable angle development in the two racial groups	Wang et al., 2013

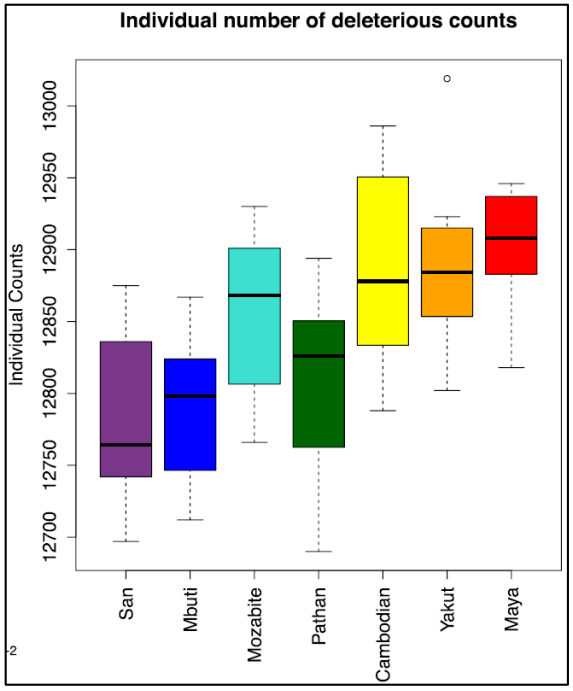
B. R. Shah *et al.*, *Diabetes Care* **36**, 2670-2676 (2013).
L. P. Pabon-Nau *et al.*, *J. Gen. Intern. Med.* **25**, 847-852 (2010).
C. Gerlinger *et al.*, *BMC Womens Health* **12**, 9 (2012).
Y. E. Wang *et al.*, *Invest. Ophthalmol. Vis. Sci.* **54**, 7717-7723 (2013).

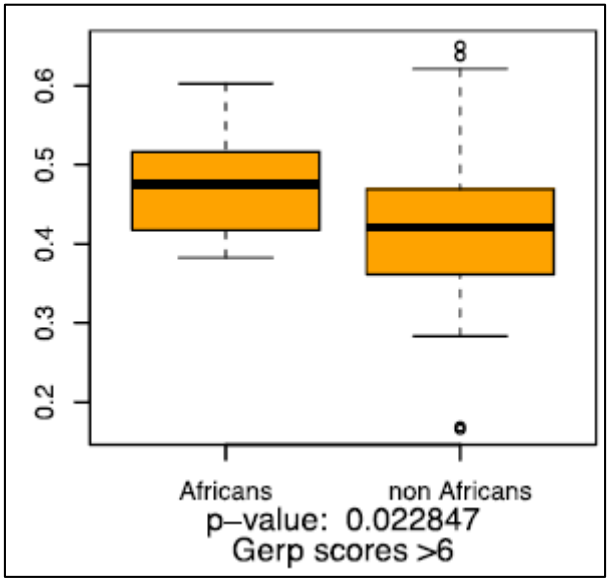
Population groups differ in mutational load

Distance from sub-Saharan Africa predicts mutational load in diverse human genomes

Brenna M. Henn[a,1,2], Laura R. Botigué[a,1], Stephan Peischl[b,c,d,1], Isabelle Dupanloup[b], Mikhail Lipatov[a], Brian K. Maples[e], Alicia R. Martin[e], Shaila Musharoff[e], Howard Cann[f,3], Michael P. Snyder[e], Laurent Excoffier[b,c,4], Jeffrey M. Kidd[g,4], and Carlos D. Bustamante[e,2,4]







Genomic Evolutionary Rate Profiling (GERP). This is a conservation score. Variants with a score of 6 are extremely conserved

Henn BM, et al. 2016. Distance from sub-Saharan Africa predicts mutational load in diverse human genomes. Proc. Natl. Acad. Sci. U.S.A. 113:E440-E449.

Identifying the population/s of origin
global vs local approaches

Knowledge of your population is critical.

You can identify the geographical origin of a population using:

- Global ancestry tools, which provide a single result for the entire genome: (STRCUTRUE/ADMITURE, PCA, or GPS).

- Local ancestry tool, which partition the genome into different ancestries (e.g., *Lamp*)

Advantages? Disadvantages?

Admixture: Alexander DH, Lange K. 2011. Enhancements to the ADMIXTURE algorithm for individual ancestry estimation. BMC Bioinformatics. 12:246.
PCA: Price AL, et al. 2006. Principal components analysis corrects for stratification in genome-wide association studies. Nat. Genet. 38:904-909.
GPS: E. Elhaik *et al.*, *Nat Commun* **5**, (2014).
Lamp: Sankararaman S, et al. 2008. Estimating local ancestry in admixed populations. Am. J. Hum. Genet. 82:290-303.

nature COMMUNICATIONS

ARTICLE

Received 17 Apr 2013 | Accepted 26 Feb 2014 | Published xx xxx

Online attention

388

Altmetric score (what's this?)

Tweeted by 88
On 3 Facebook pages
Mentioned in 3 Google+ posts
Picked up by 30 news outlets
1 Reddit

View more

This Altmetric score means that the article is:

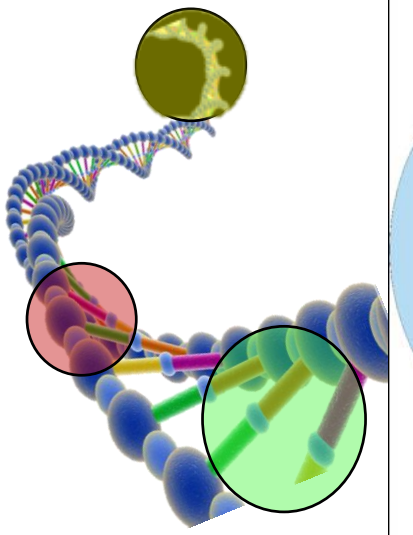
- in the 99 percentile (ranked 203rd) of the 128,174 tracked articles of a similar age in all journals
- in the 98 percentile (ranked 8th) of the 469 tracked articles of a similar age in *Nature Communications*

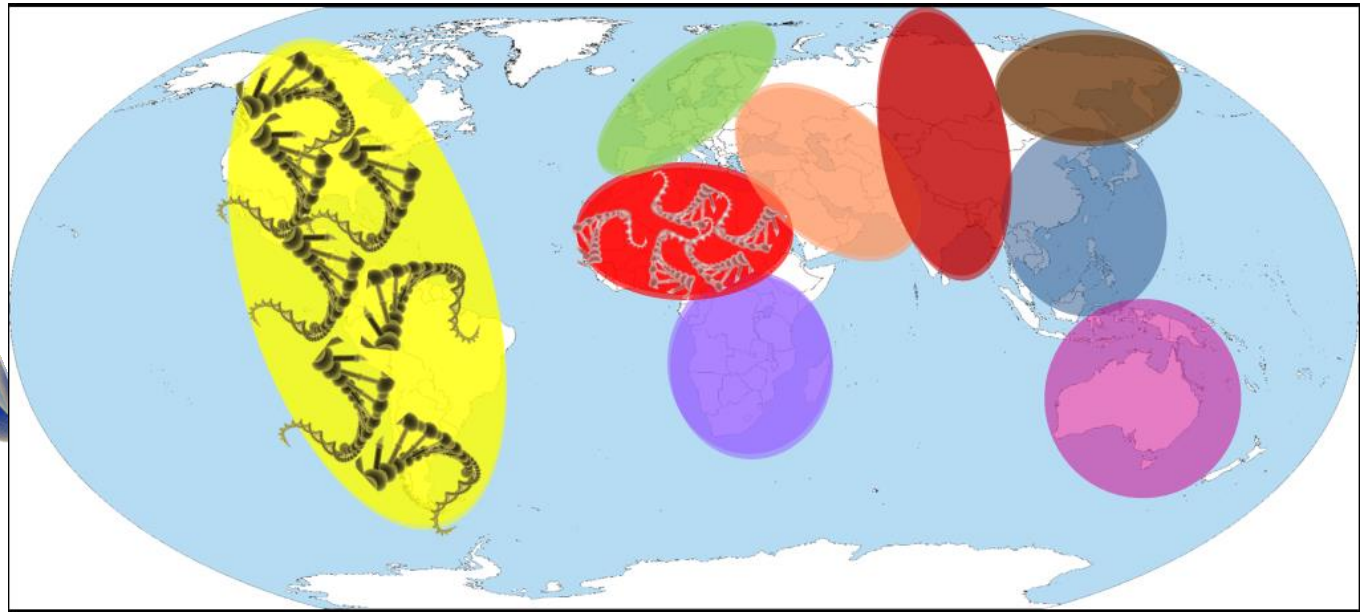
Geographic population structure analysis of worldwide human populations infers their biogeographical origins

Eran Elhaik[1,2,*], Tatiana Tatarinova[3,*], Dmitri Chebotarev[4], Ignazio S. Piras[5], Antonella De Montis[6], Manuela Atzori[6], Monica Marini[6], Sergio Tofanelli[7], Chris Tyler-Smith[9], Yali Xue[9], Francesco Cucca[5], Theodore G. Schurr[10], Jill B Miguel G. Vilar[10], Amanda C. Owings[10], Rocío Gómez[11], Ricardo Fujita[12], Fa Oleg Balanovsky[15,16], Elena Balanovska[16], Pierre Zalloua[17], Himla Soodyall[18], ArunKumar GaneshPrasad[19], Michael Hammer[20], Lisa Matisoo-Smith[21], Spe & The Genographic Consortium[24]

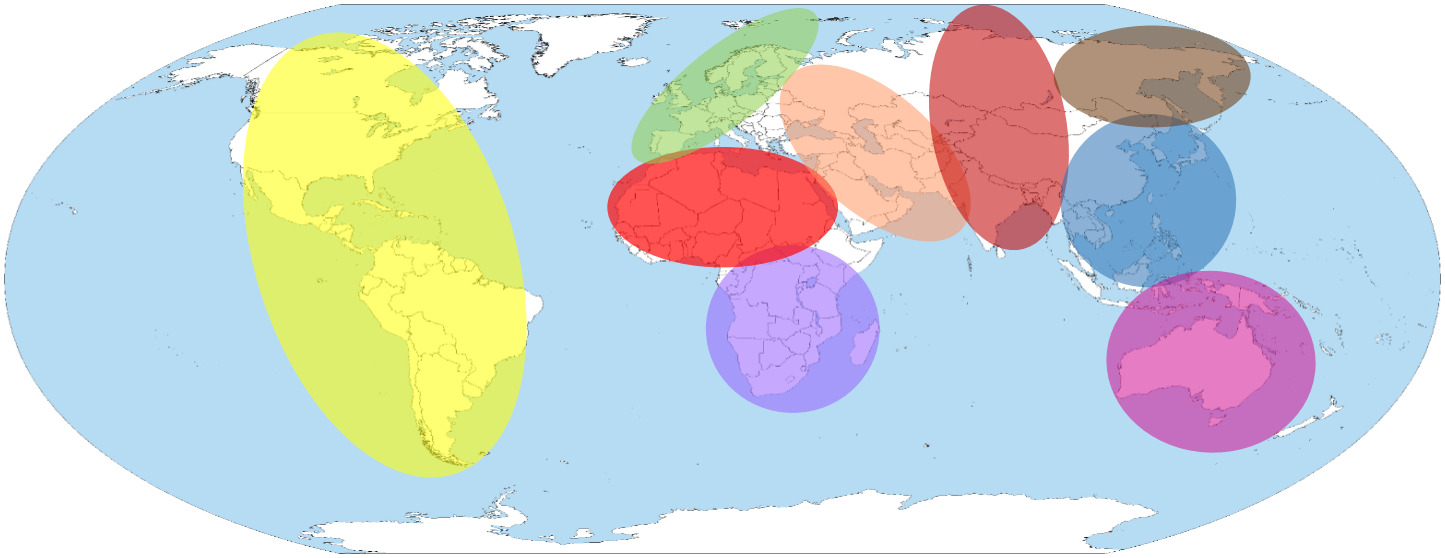
Nine genetic soups (modern DNA)

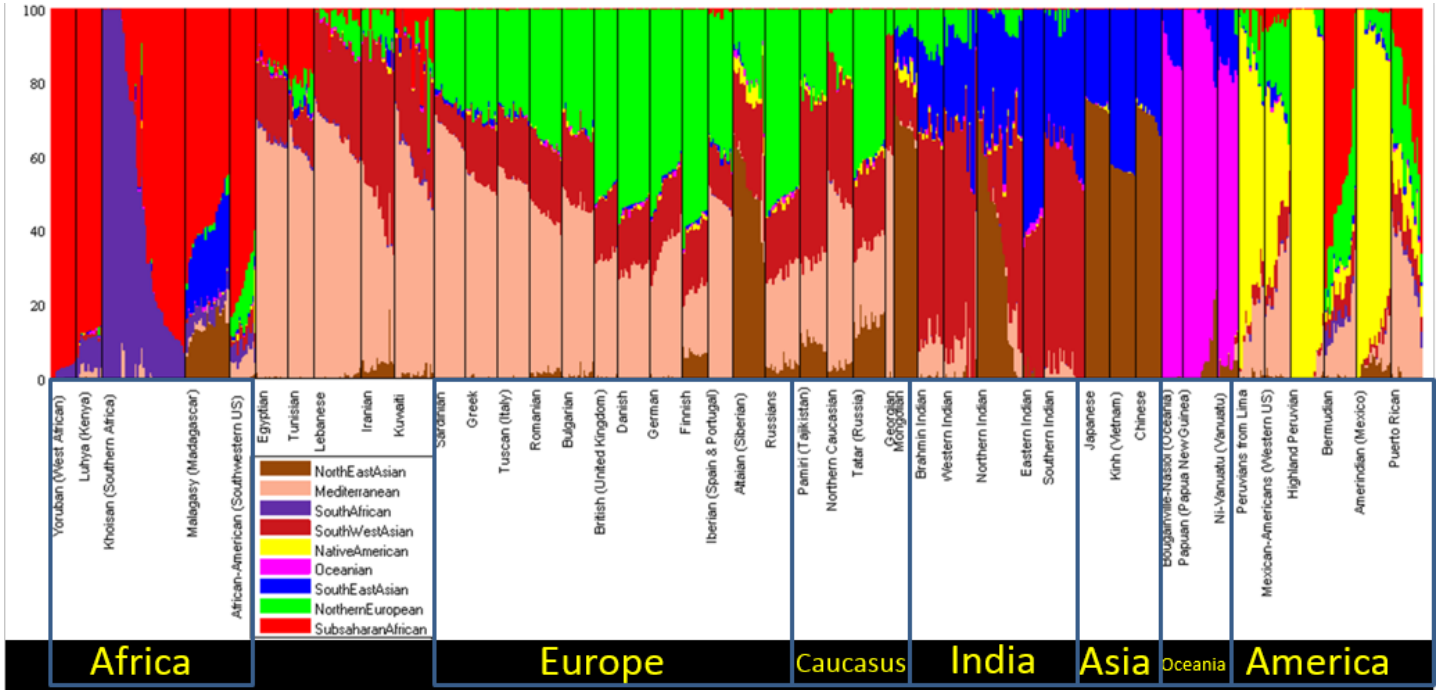
A Match!

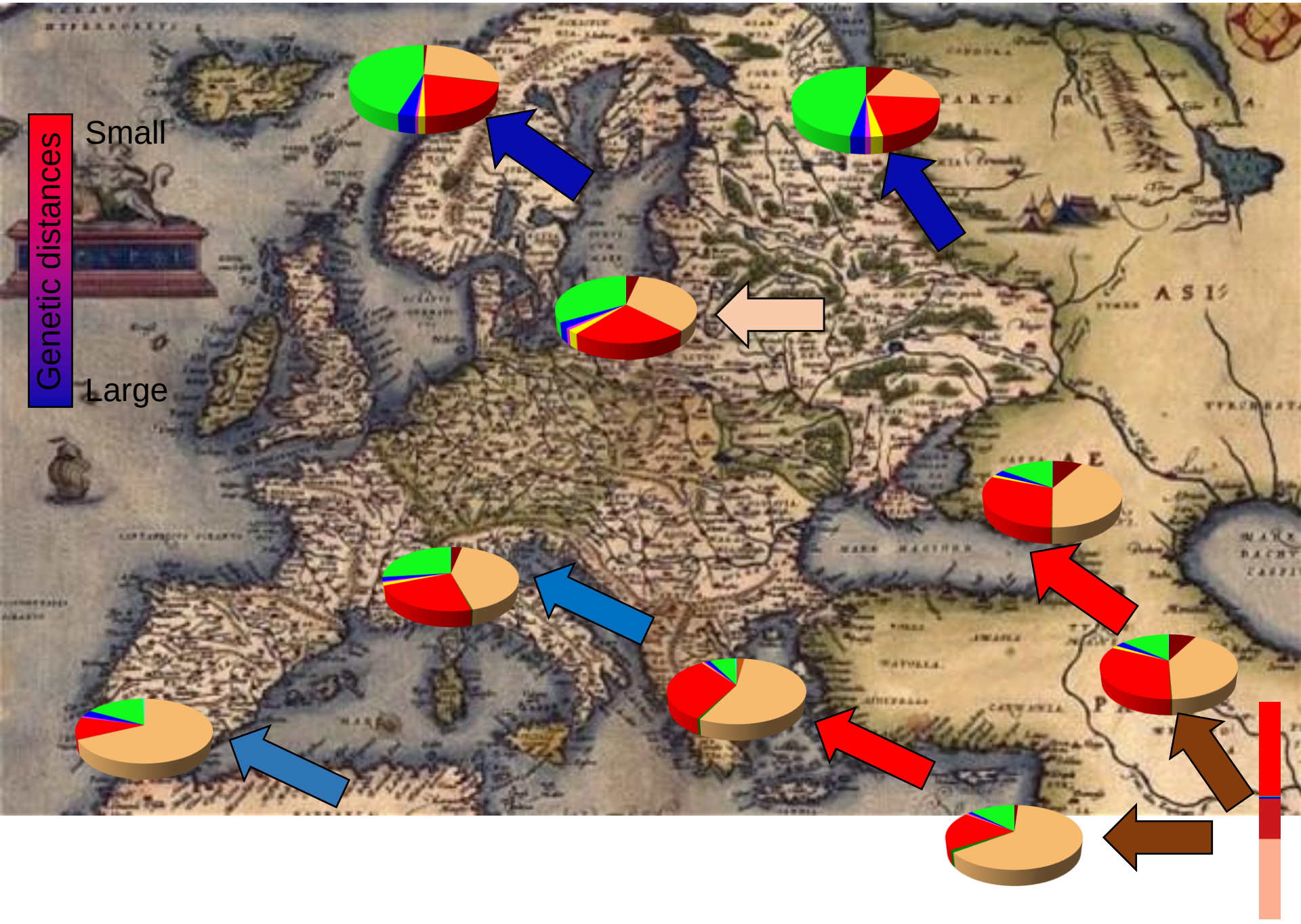




Modelling population mixtures

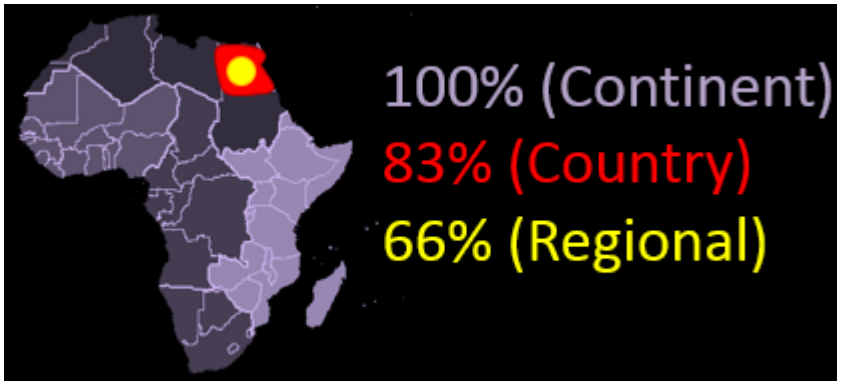




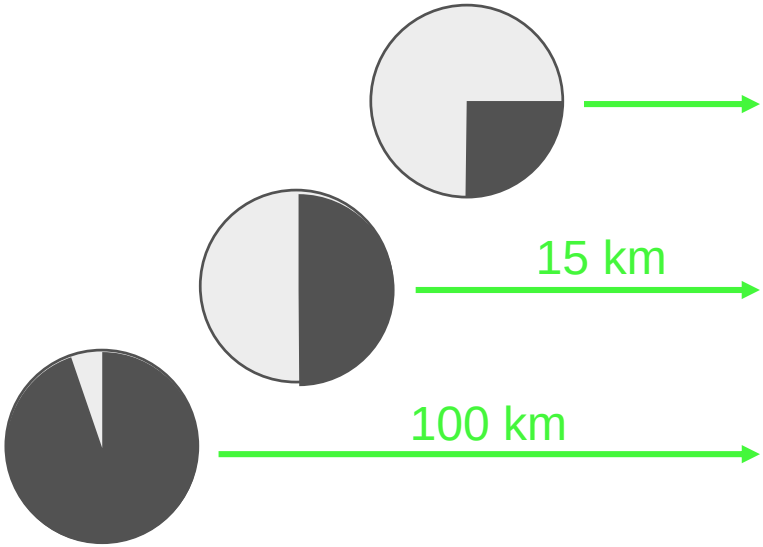

Genetic distances
Small
Large

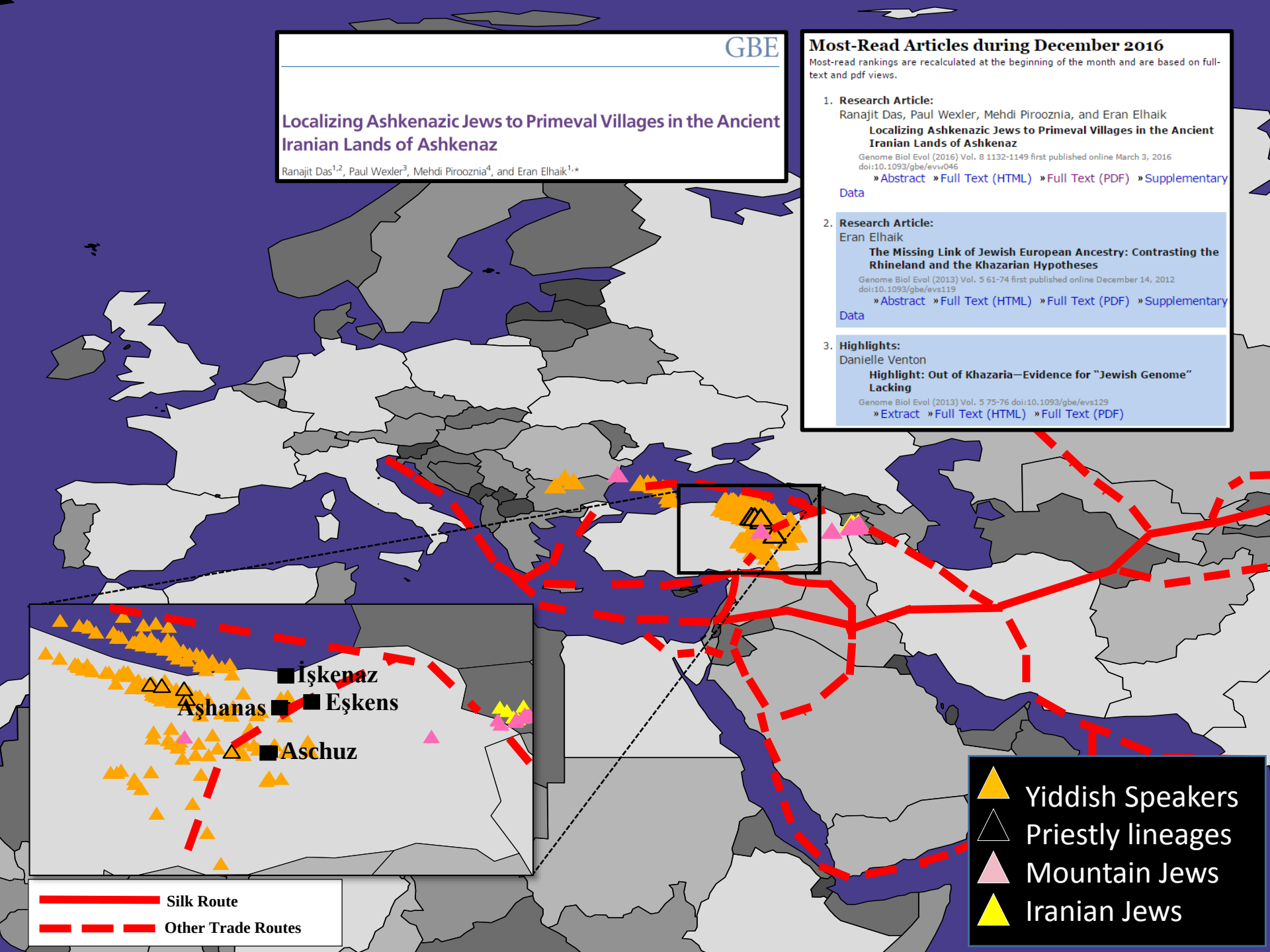
GPS's accuracy

Genographic's global dataset



Sardinian villagers dataset




GBE
Localizing Ashkenazic Jews to Primeval Villages in the Ancient Iranian Lands of Ashkenaz
Ranajit Das[1,2], Paul Wexler[3], Mehdi Pirooznia[4], and Eran Elhaik[1,*]
Most-Read Articles during December 2016
Most-read rankings are recalculated at the beginning of the month and are based on full-text and pdf views.
1. Research Article:
Ranajit Das, Paul Wexler, Mehdi Pirooznia, and Eran Elhaik
Localizing Ashkenazic Jews to Primeval Villages in the Ancient Iranian Lands of Ashkenaz
Genome Biol Evol (2016) Vol. 8 1132-1149 first published online March 3, 2016 doi:10.1093/gbe/evw046
»Abstract »Full Text (HTML) »Full Text (PDF) »Supplementary Data
2. Research Article:
Eran Elhaik
The Missing Link of Jewish European Ancestry: Contrasting the Rhineland and the Khazarian Hypotheses
Genome Biol Evol (2013) Vol. 5 61-74 first published online December 14, 2012 doi:10.1093/gbe/evs119
»Abstract »Full Text (HTML) »Full Text (PDF) »Supplementary Data
3. Highlights:
Danielle Venton
Highlight: Out of Khazaria—Evidence for "Jewish Genome" Lacking
Genome Biol Evol (2013) Vol. 5 75-76 doi:10.1093/gbe/evs129
»Extract »Full Text (HTML) »Full Text (PDF)
Işkenaz
Eşkens
Aşhanas
Aschuz
Silk Route
Other Trade Routes
Yiddish Speakers
Priestly lineages
Mountain Jews
Iranian Jews

GPS ORIGINS™
PINPOINT YOUR ANCESTRY

GET STARTED | LOGIN

THOMAS MACENTEE'S RESULTS

HUMAN ORIGINS : OUR SHARED HISTORY TO YOUR STORY

The questions of who we are and where we come from have been asked for throughout our history. Once we explained our origins with mythology and folklore but now we utilize modern science to answer them. Genetics help us tell the story of our origins from the beginning, through the formation of the human gene pools and to the last 2000 years of history. The test results you have just received, along with the following information, will help you understand your personal story, from the shared history of all humans to your unique family story. *Read More* ▶

GENE POOL PERCENTAGES

TOP 3 GENE POOLS

#1 Fennoscandia 21.5%

#2 Orkney Islands 15.2%

#3 Southern France 15.2%

% All Gene Pools

GENE POOL REGIONS

https://gpsorigins.com/share/d96678f189b5340efba9f2305472bf05

Real life example

You are studying a very rare disorder.
It is a devastating disorder. No known variants. No treatment. No cure.
You were lucky enough to find 7 families and obtained their complete-genome data.

One of the families is known to be inbred. You excluded it from the study.
Looking at the literature for Europeans, Africans, and Asians, you interpret your inbreeding statistic (e.g., *pi hat*) of the other families as an indication of no inbreeding.

Due to genetic privacy you have no demographic information about the patients.
You know that the inbred family was provided by a Middle Eastern country.
The 6 other families are from different American hospitals.

- You found a **deleterious variant** that segregated in all families.
- The gene's annotation is **highly suggestive** of causation.
- **A replication study** in unrelated patients from Arizona **confirmed the** segregation of the variant compared to European-American controls.
- **All the results are highly statistically significant.**

You are **very excited** and want to submit your findings to *Nature*.
Your scientific Spidey-sense urges you to rethink your study design.

Loss-of-function (Lof) mutations are a lot more dangerous than you think

The Biology Complicated by Genetic Analysis

Xionglei He*[1]

[1]The State Key Laboratory of Bio-control, College of Ecology and Evolution, School of Life Sciences, Sun Yat-sen University, Guangzhou, China

***Corresponding author:** E-mail: hexiongl@mail.sysu.edu.cn.

Associate Editor: Jianzhi Zhang

Abstract

Genetics is used as a tool to study living systems because of a key assumption that the phenotypes of loss-of-function mutations on a gene indicate the gene's normal/native functions. I propose that inactivation of a gene not only suppresses the gene's native functions but may also create spurious functions that cause phenotypes irrelevant to the gene's native functions. Such spurious functions represent the otherwise dormant physical/chemical potentials of a living system, do not follow the existing rules built by natural selection, and can hardly be integrated with other functions using empirical data. Thus, the rationale of using loss-of-function phenotypic data to understand a living system is challenged. Fortunately, spurious functions are expected to be evolutionarily unstable while native functions should be conserved, suggesting a means of separating them. I argue that current biology is confused by the undiscerned use of genetic data and suggest a solution.

Solutions: experimental validation, larger cohorts, replication

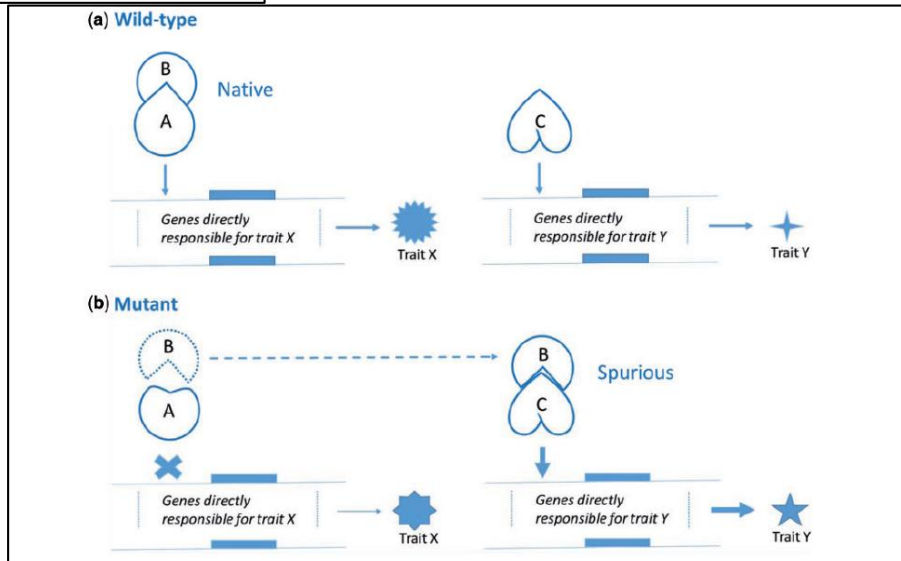


FIG. 1. Inactivation of gene A affects traits X and Y. The former is due to the loss of A–B dimer, a native function of gene A, while the latter is due to the gain of B–C dimer, a spurious function associated with gene A.

He X. 2016. The Biology Complicated by Genetic Analysis. Mol. Biol. Evol. 33:2177-2181.

Nonsense genetic variants play a role in translation plasticity

Research

Translational plasticity facilitates the accumulation of nonsense genetic variants in the human population

Sujatha Jagannathan[1,2,3] and Robert K. Bradley[1,2]

[1]Computational Biology Program, Public Health Sciences Division, Fred Hutchinson Cancer Research Center, Seattle, Washington 98109, USA; [2]Basic Sciences Division, Fred Hutchinson Cancer Research Center, Seattle, Washington 98109, USA; [3]Human Biology Division, Fred Hutchinson Cancer Research Center, Seattle, Washington 98109, USA

Genetic variants that disrupt protein-coding DNA are ubiquitous in the human population, with about 100 such loss-of-function variants per individual. While most loss-of-function variants are rare, a subset have risen to high frequency and occur in a homozygous state in healthy individuals. It is unknown why these common variants are well tolerated, even though some affect essential genes implicated in Mendelian disease. Here, we combine genomic, proteomic, and biochemical data to demonstrate that many common nonsense variants do not ablate protein production from their host genes. We provide computational and experimental evidence for diverse mechanisms of gene rescue, including alternative splicing, stop codon readthrough, alternative translation initiation, and C-terminal truncation. Our results suggest a molecular explanation for the mild fitness costs of many common nonsense variants and indicate that translational plasticity plays a prominent role in shaping human genetic diversity.

- Many common nonsense variants have only modest impacts upon the levels of total protein produced from their seemingly disabled parent genes.
- Permissive RNA processing has the potential to convert loss-of-function variants from genetic nulls into hypomorphic, silent, or even neomorphic alleles.
- This explains why healthy individuals carry homozygous putative LOF variants within genes implicated in Mendelian disease

Jagannathan S, Bradley RK. 2016. Translational plasticity facilitates the accumulation of nonsense genetic variants in the human population. Genome Res.

The intolerance to functional genetic variation of protein domains

Gussow et al. (2016) created a ranking of gene sub-regions (exons/protein domains) that would reflect their intolerance to functional variation based on variation in the human population.

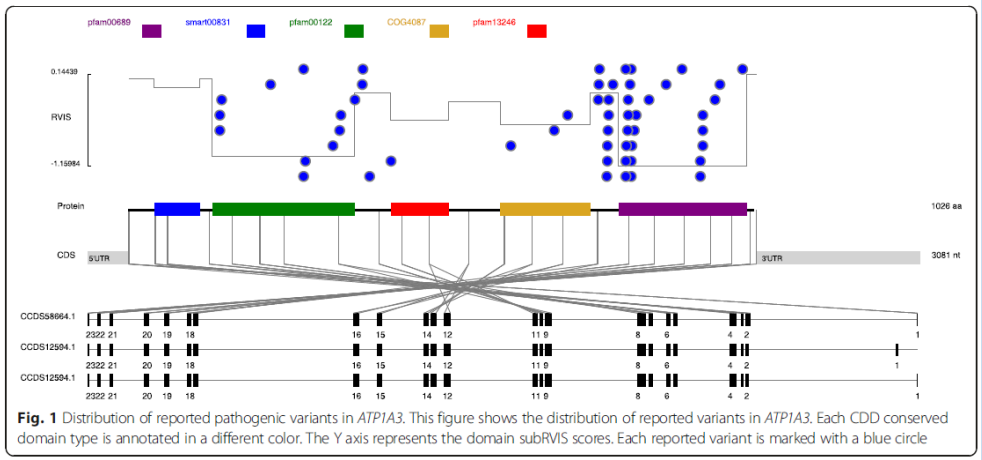


Fig. 1 Distribution of reported pathogenic variants in *ATP1A3*. This figure shows the distribution of reported variants in *ATP1A3*. Each CDD conserved domain type is annotated in a different color. The Y axis represents the domain subRVIS scores. Each reported variant is marked with a blue circle

In some cases, genes are somewhat evenly divided in more and less tolerant regions. One example in this category is the ATP1A3 gene (Fig. 1). Overall, ATP1A3 is a highly intolerant gene.

The MAPT gene (Fig. 2) is highly tolerant (98th percentile) despite carrying mutations that cause frontotemporal dementia.
Only a small proportion (26 %) of the gene is very intolerant relative to the majority of the gene.

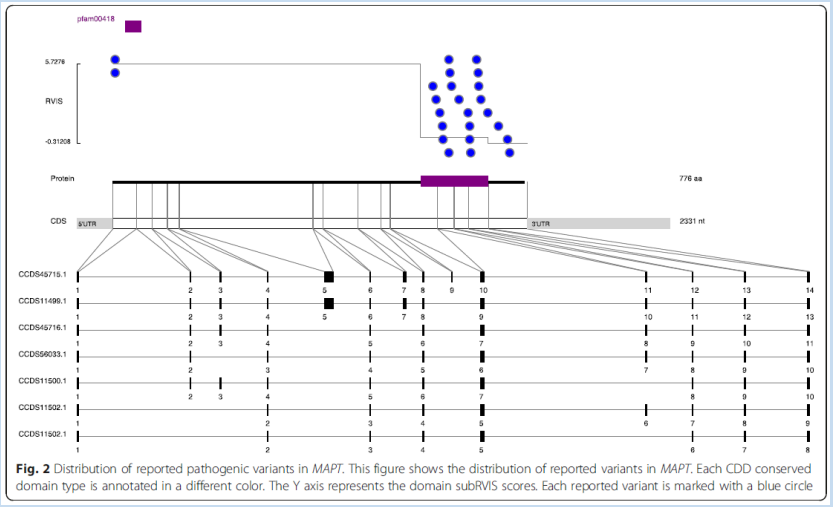


Fig. 2 Distribution of reported pathogenic variants in *MAPT*. This figure shows the distribution of reported variants in *MAPT*. Each CDD conserved domain type is annotated in a different color. The Y axis represents the domain subRVIS scores. Each reported variant is marked with a blue circle

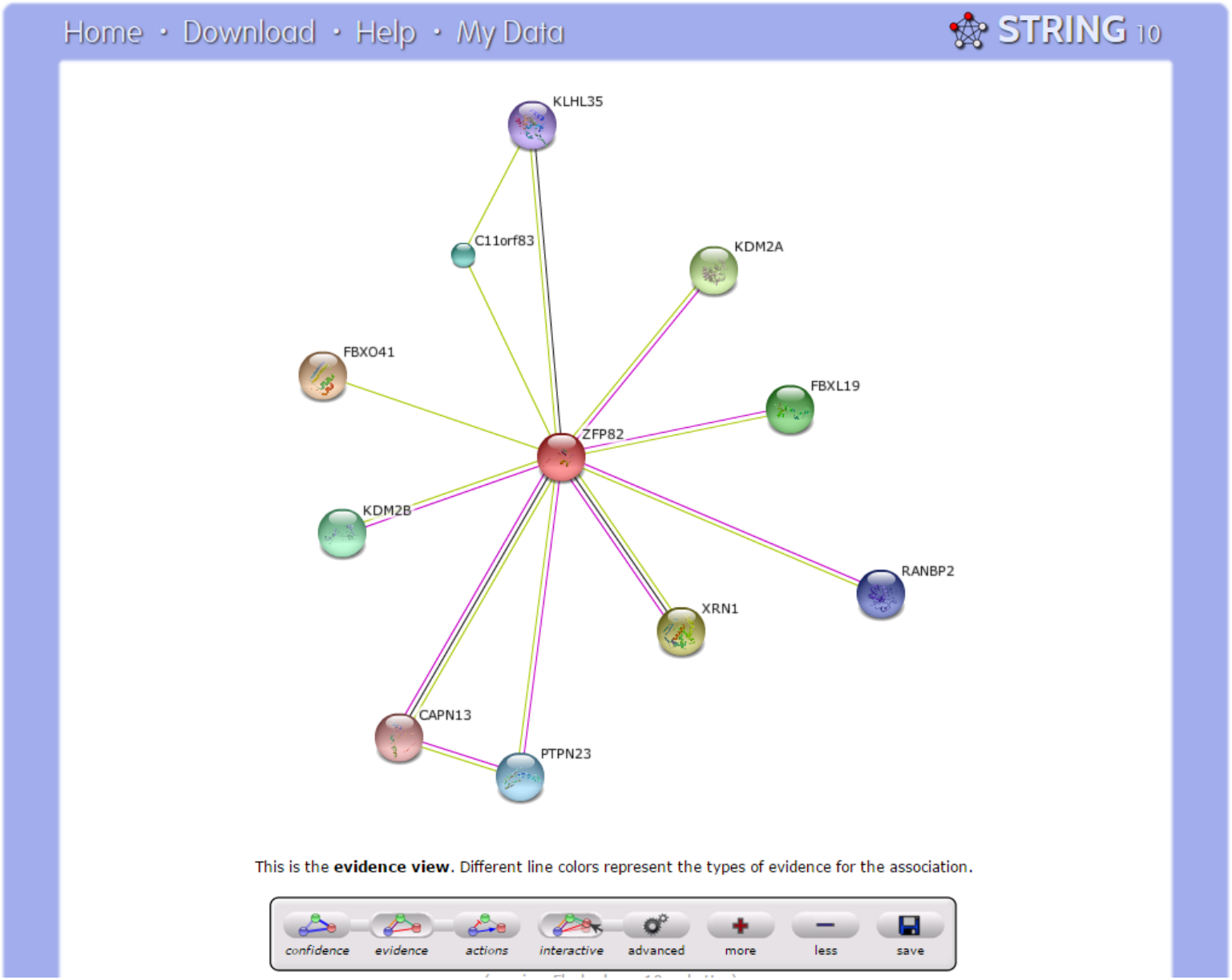
https://genomebiology.biomedcentral.com/articles/10.1186/s13059-016-0869-4

Beyond genetic variants

STRING

(http://string-db.org/)

Known and Predicted Protein-Protein Interactions



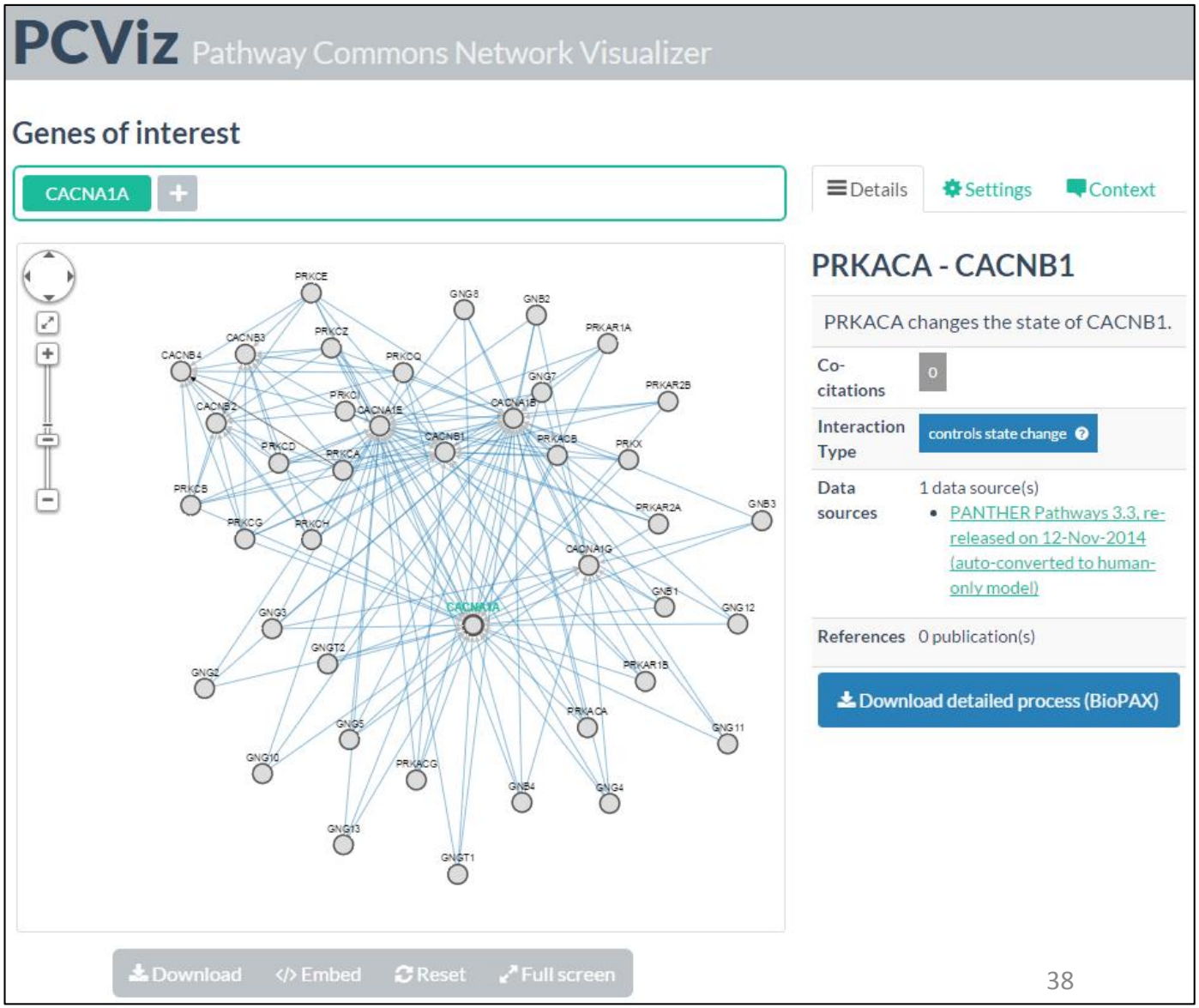
Pathway Commons

(http://www.pathwaycommons.org/)

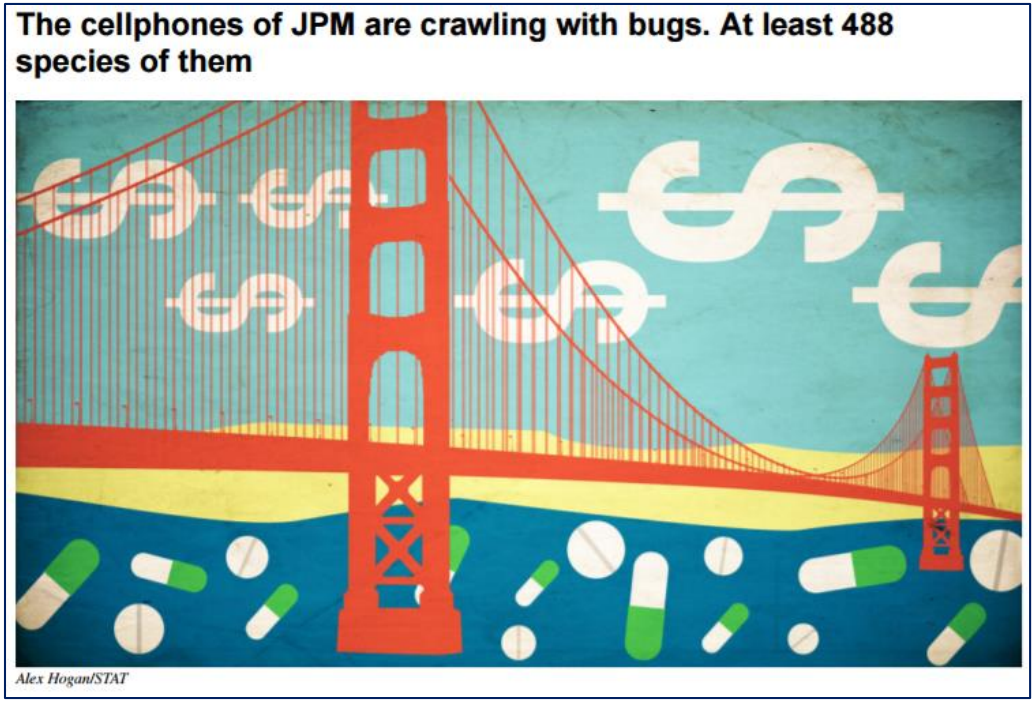
A SNP may interfere with a gene whose function is unknown, but it is associated with a well studied gene.

Important questions:

1. What is the nature of that "association?"
2. Is there a biological confirmation for the association?
3. What type of "pathway" is this?
4. Who defined the "pathway" and was it validated?



Microbiome

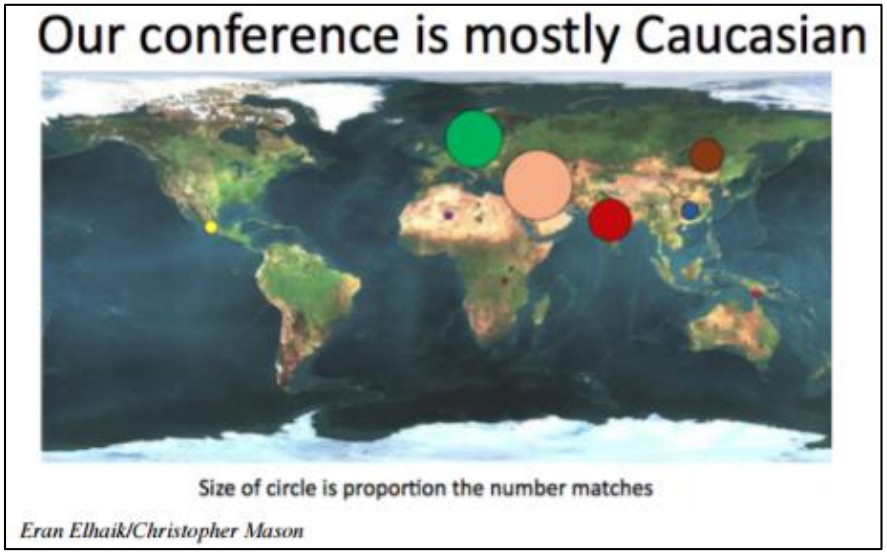


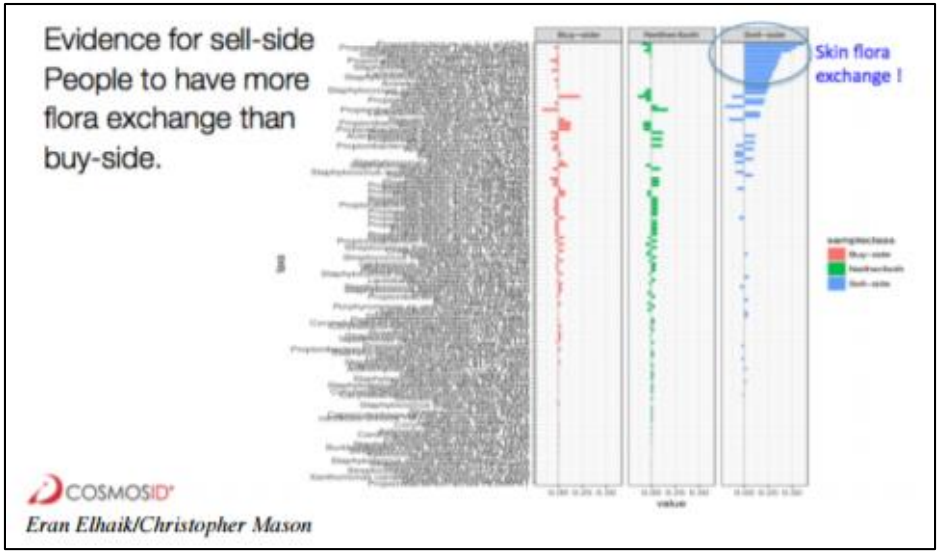
REVIEW

The microbiome and innate immunity

Christoph A. Thaiss[1*], Niv Zmora[1,2,3*], Maayan Levy[1*] & Eran Elinav[1]

doi:10.1038/nature18847





The pipeline so far - Tools

Obtaining DNA

Case/control cohort

Sequencer

Fastq

Quality control

Fastq

Alignment to reference genome

BAM

Visualization

SAM

Variant calling

VCF

Variant Annotation

Text file

Variant Prioritization

FastQC

Bowtie2

BAM-TO-SAM, IGV

GATK,SAMtools

Accessible through Galaxy

Galaxy

Annovar, wAnnovar, Variant Effect Predictor, dbSNP, ExAC browser

OMIM, ClinVAR, Geography of Genetic Variants Browser, GPS