

# Part II: Rare variants

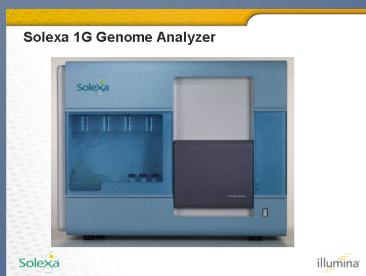
- 1. Advances in genomic technologies
  - From genotype array to DNA sequencing
- 2. How to design a sequencing study
- 3. Predicting the functional consequences of genetic variation
- 4. Models for rare variant associations
- 5. Examples
- 6. Other data types

# Advances in genomic sequencing technologies

454



Solexa



Illumina GA2

Ion Torrent



Illumina HiSeq

2000

2007 2008 2009 2010 2011 2012 2013 2014 2015



Helicos

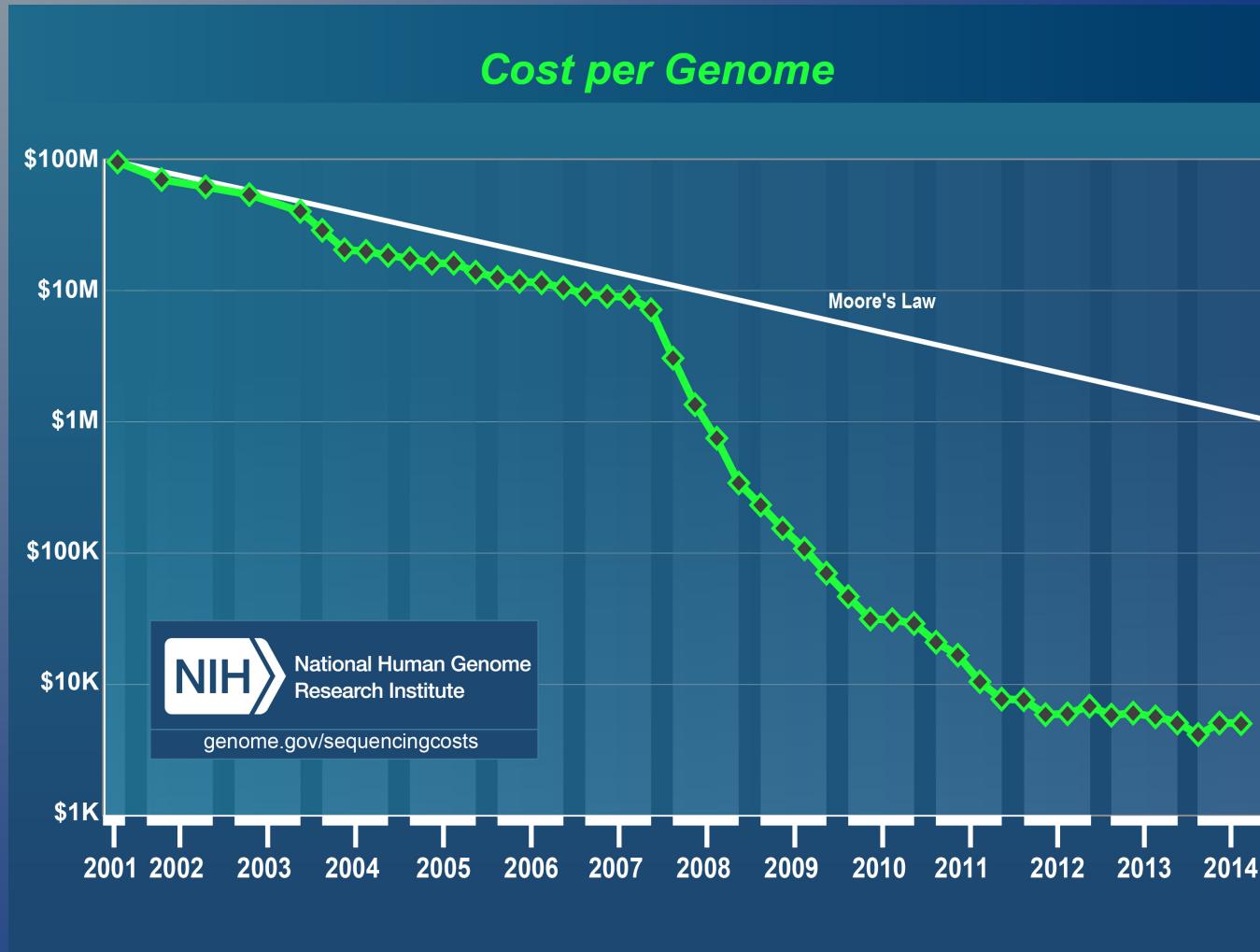


Illumina MiSeq



Oxford Nanopore

# Genome sequencing costs dropped...



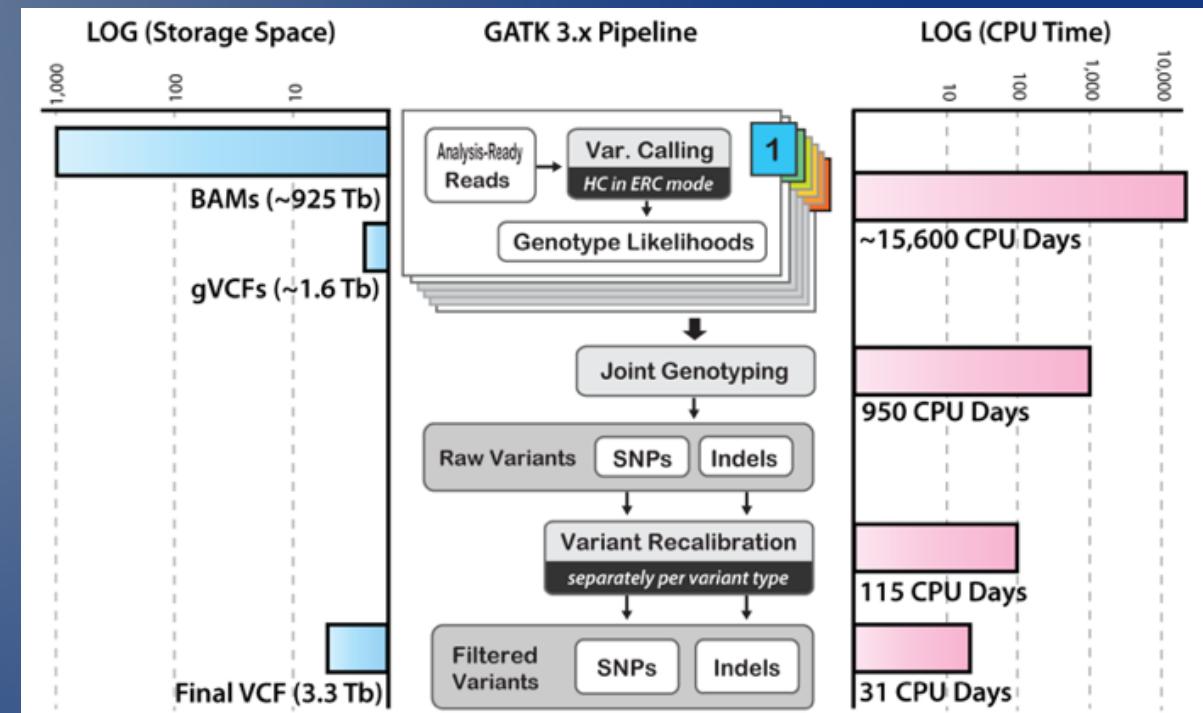
...enabling the exploration of rare variants throughout the genome

# Processing data

## Standard File Formats

SAM/BAM  
Sequencing  
Alignment and  
Mapping file  
(Raw read data)

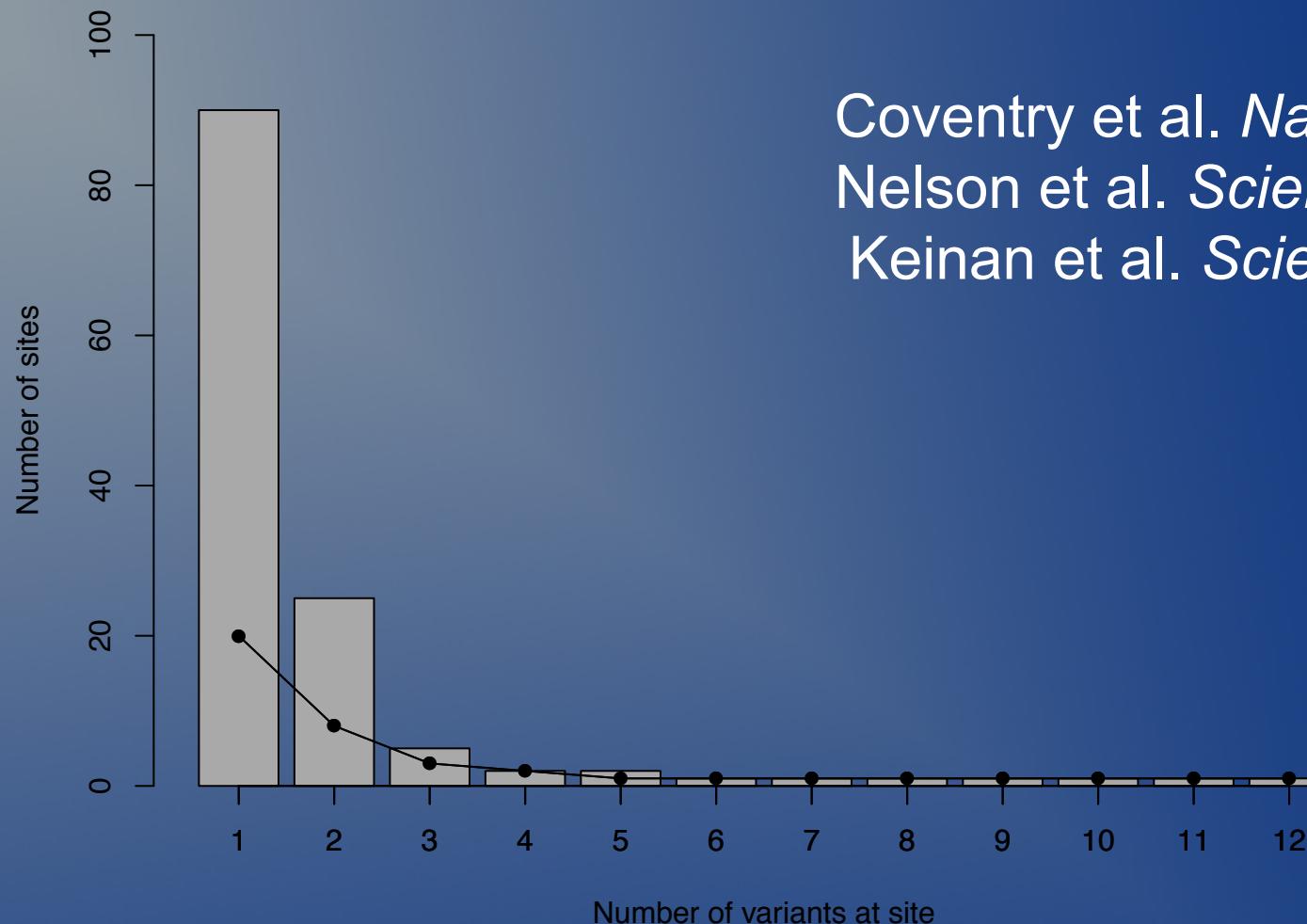
VCF  
Variant Call Format



Monkol Lek

Terabytes close to Petabytes of data

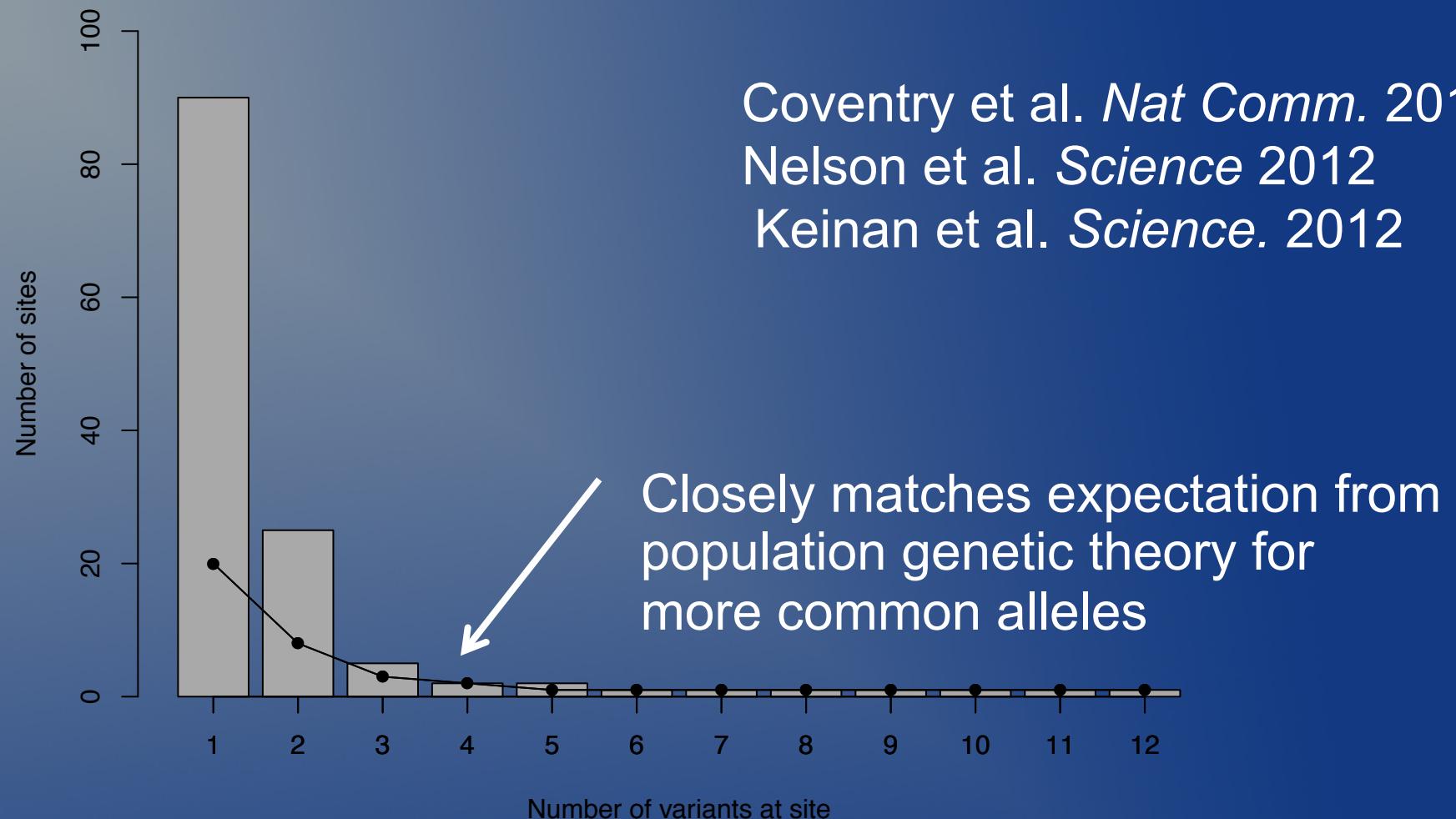
# Surveys of rare genetic variation



Coventry et al. *Nat Comm.* 2010  
Nelson et al. *Science* 2012  
Keinan et al. *Science*. 2012

Initial surveys of rare genetic variation found the number of rare variant sites detected in data were about 4x larger than expected by population genetic theory (black line).

# Surveys of rare genetic variation

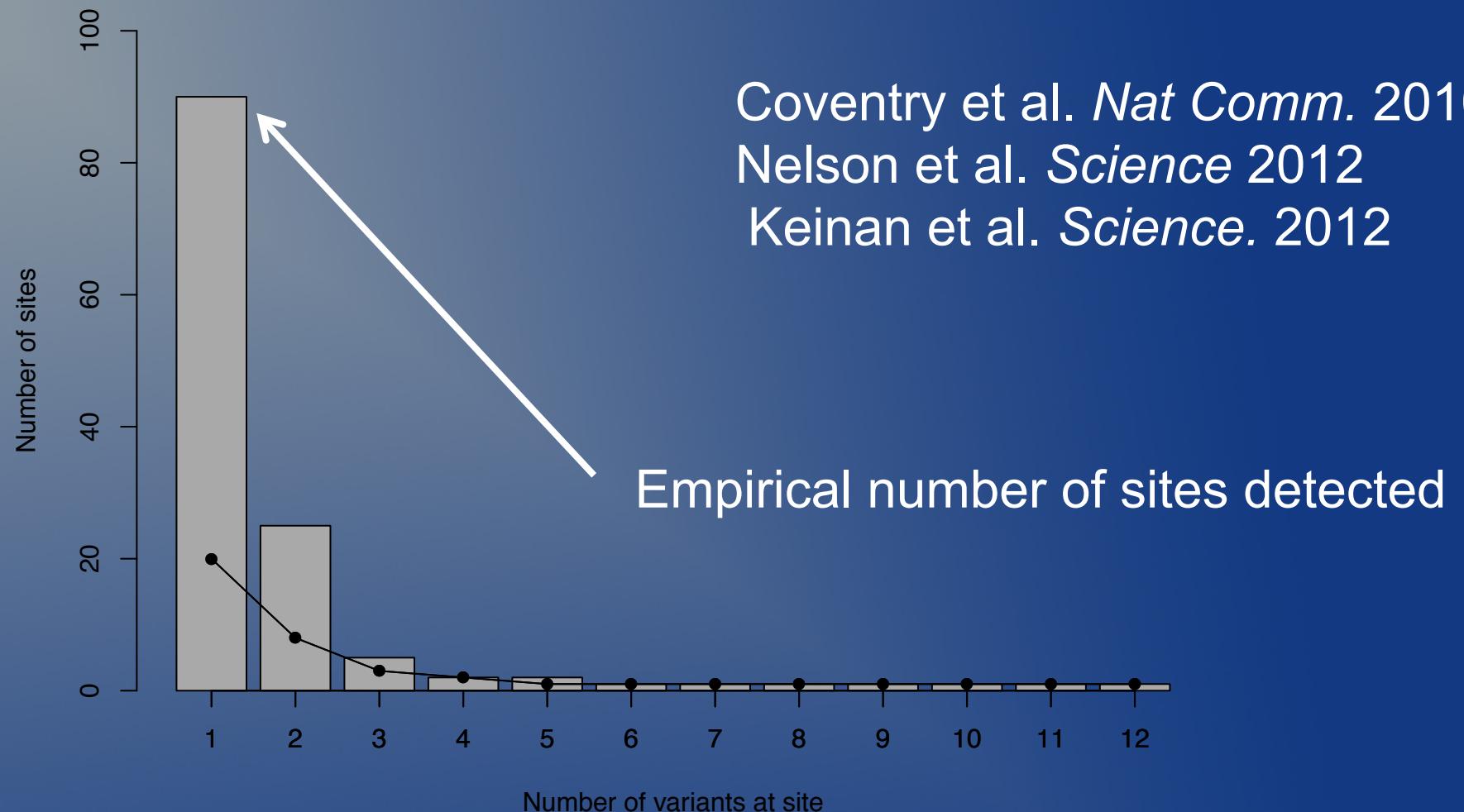


Initial surveys of rare genetic variation found the number of rare variant sites detected in data were about 4x larger than expected by population genetic theory (black line).

Coventry et al. *Nat Comm.* 2010  
Nelson et al. *Science* 2012  
Keinan et al. *Science*. 2012

Closely matches expectation from population genetic theory for more common alleles

# Surveys of rare genetic variation



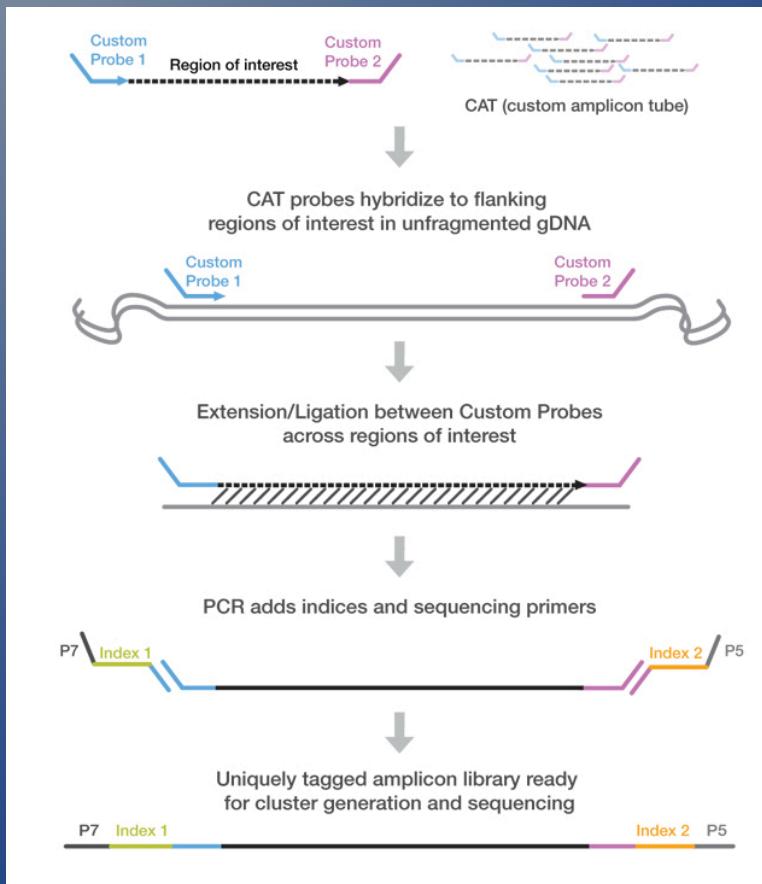
Initial surveys of rare genetic variation found the number of rare variant sites detected in data were about 4x larger than expected by population genetic theory (black line).

# Designing a sequencing study: properties to consider

Targeted, exome, genome sequencing?

# Designing a sequencing study: properties to consider

Targeted, exome, genome sequencing?



Region of interest  
in the genome, e.g. gene

Design “amplicons”  
or “baits”

Capture>Select DNA

Library gets made  
For sequencing

Illumina protocol

# Designing a sequencing study: properties to consider

Targeted, exome, genome sequencing?

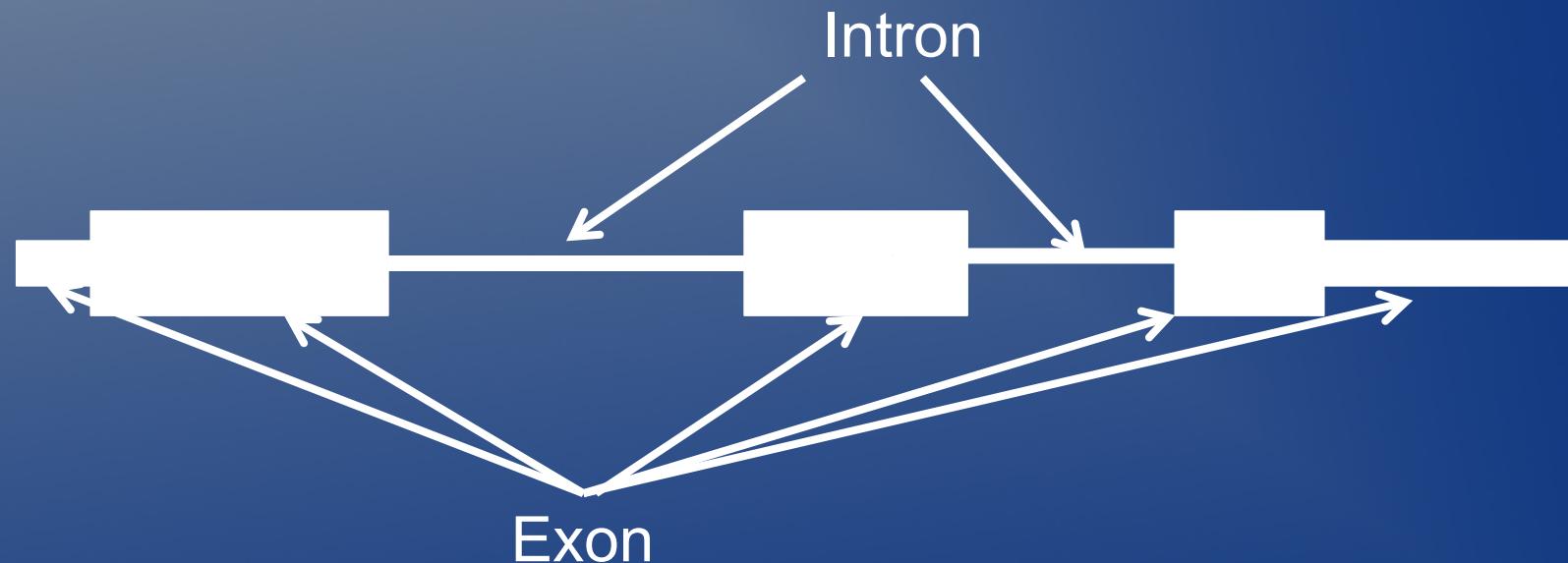
Pros: Large numbers of samples, Inexpensive, run on smaller faster  
Machines, usually used for diagnostic gene panels

Cons: Don't cover the rest of the genome

# Designing a sequencing study: properties to consider

Targeted, **exome**, genome sequencing?

Exome: collection of exon sequence of the human genome  
~1% of the genome (~30 million bases [megabases])



# Designing a sequencing study: properties to consider

Targeted, **exome**, genome sequencing?

Pros: Modest numbers of samples (~10k samples: 5-10M USD),  
capture important part of the genome

Cons: Expensive and 1% of the genome

# Designing a sequencing study: properties to consider

Targeted, exome, **genome** sequencing?

Pros: All regions of the genome

Cons: Difficult to analyze, expensive, other DNA contaminants may be introduced, e.g. foreign DNA from microbes

# Designing a sequencing study: properties to consider

## Rare or common disease?

For rare diseases this usually takes only a handful number of samples for gene discovery (that is about the number of cases Available)

Different analytical strategy: focus on sites that have not been previously discovered, annotation, and focus on mode of inheritance (textbook genetics!)

# Designing a sequencing study: properties to consider

Rare or **common** disease?

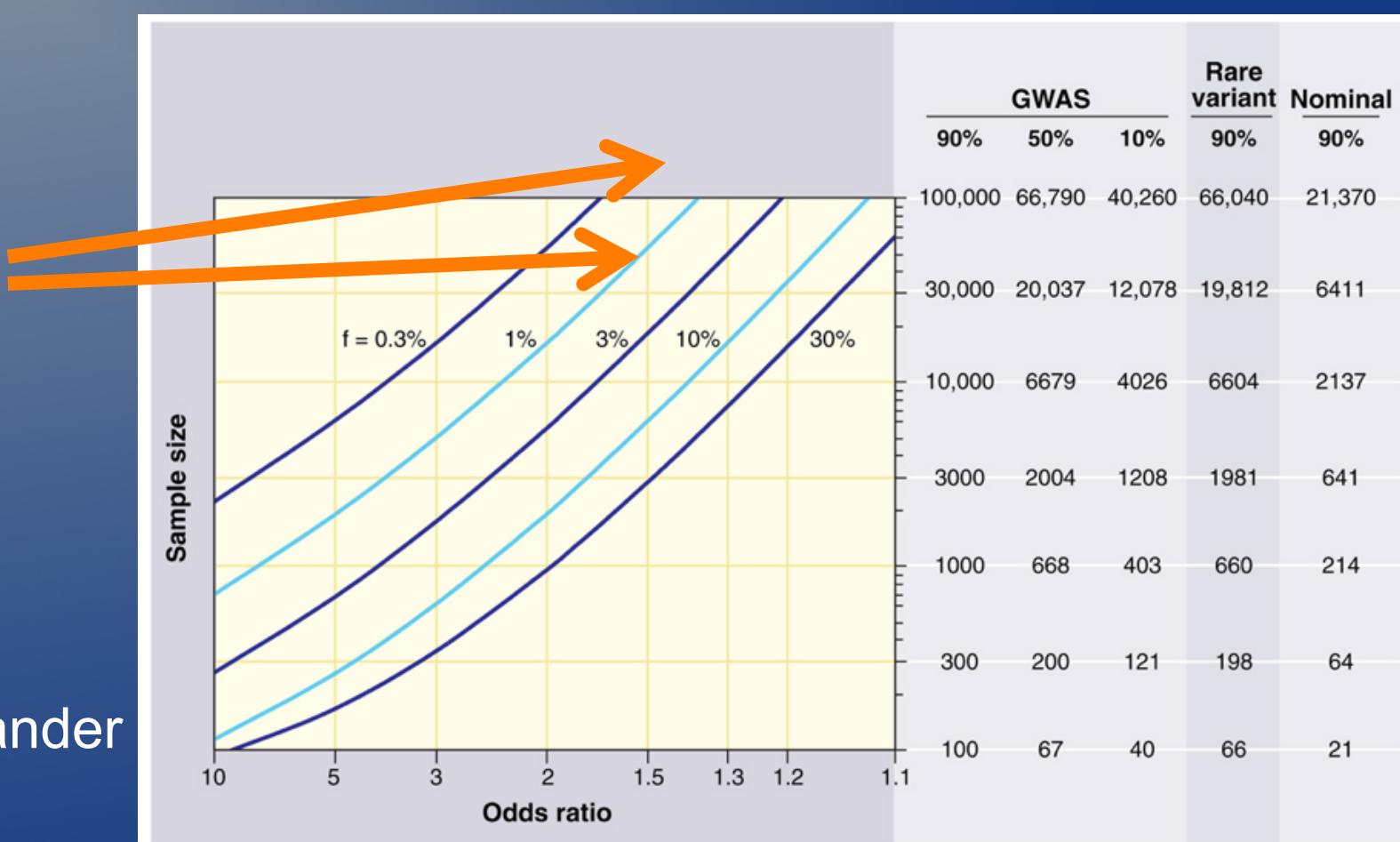
Challenging!

# Designing a sequencing study: properties to consider

Rare or common disease?

## Sample size

Odds ratio of 1.5  
For variants < 1%  
> 30k samples  
required



Altshuler, Daly, Lander  
*Science* 2009.

# Designing a sequencing study: properties to consider

Rare or **common** disease?

**Population**

Population isolates



Wikipedia

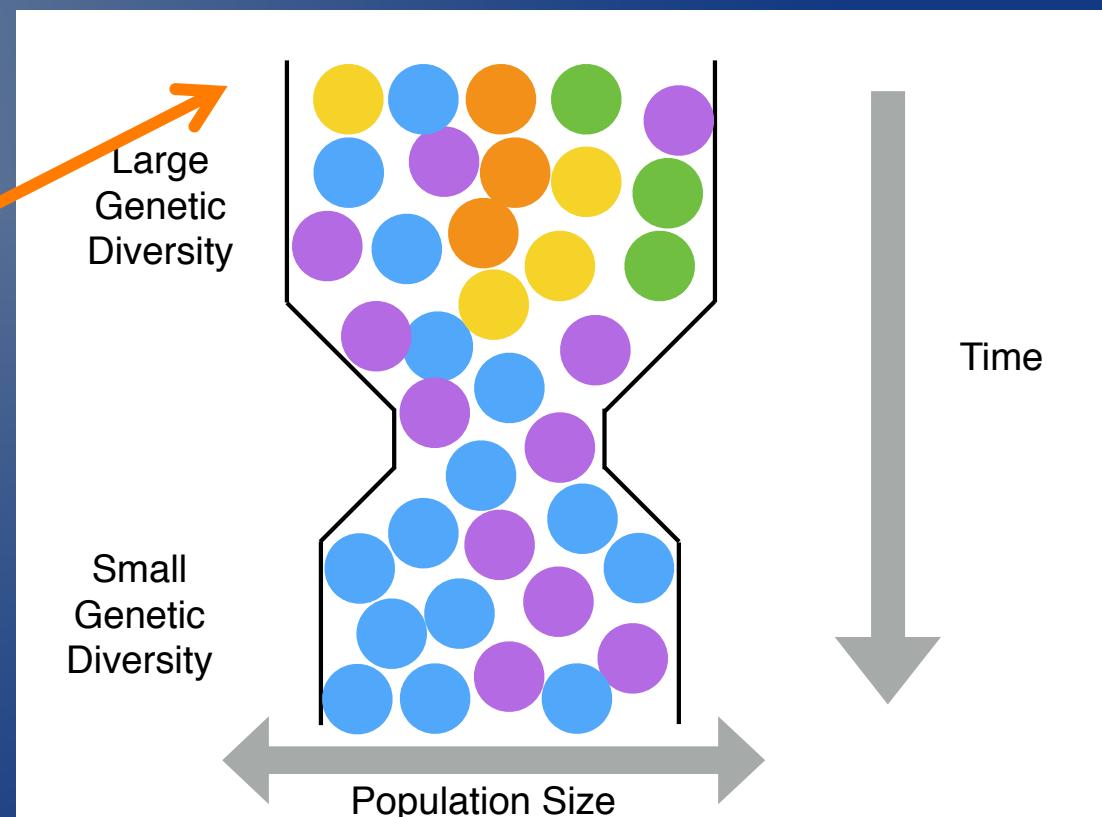
# Designing a sequencing study: properties to consider

Rare or **common** disease?

**Population**

Population isolates

Begin with large  
number  
of alleles, i.e.  
Large genetic  
diversity



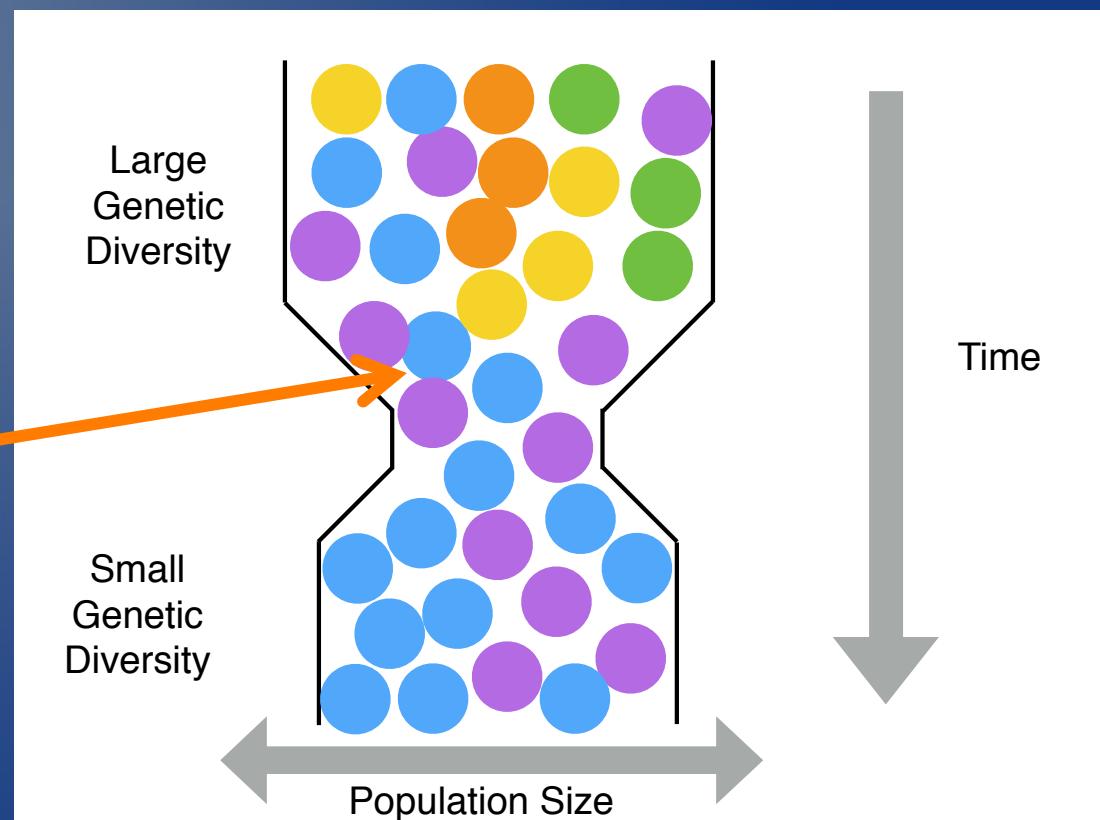
# Designing a sequencing study: properties to consider

Rare or **common** disease?

**Population**

Population isolates

Population gets  
Smaller (migration,  
war, plague, etc)



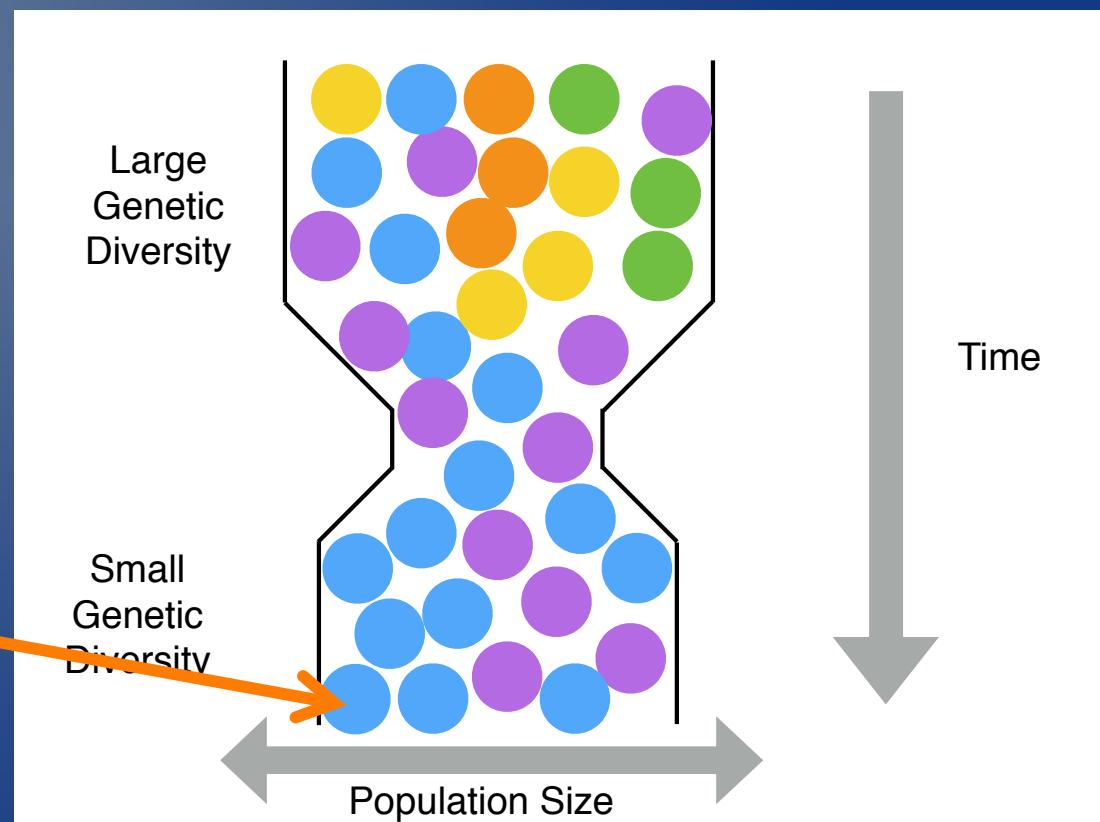
# Designing a sequencing study: properties to consider

Rare or **common** disease?

**Population**

Population isolates

Modern day: some alleles get enriched in the bottleneck, other alleles do exist

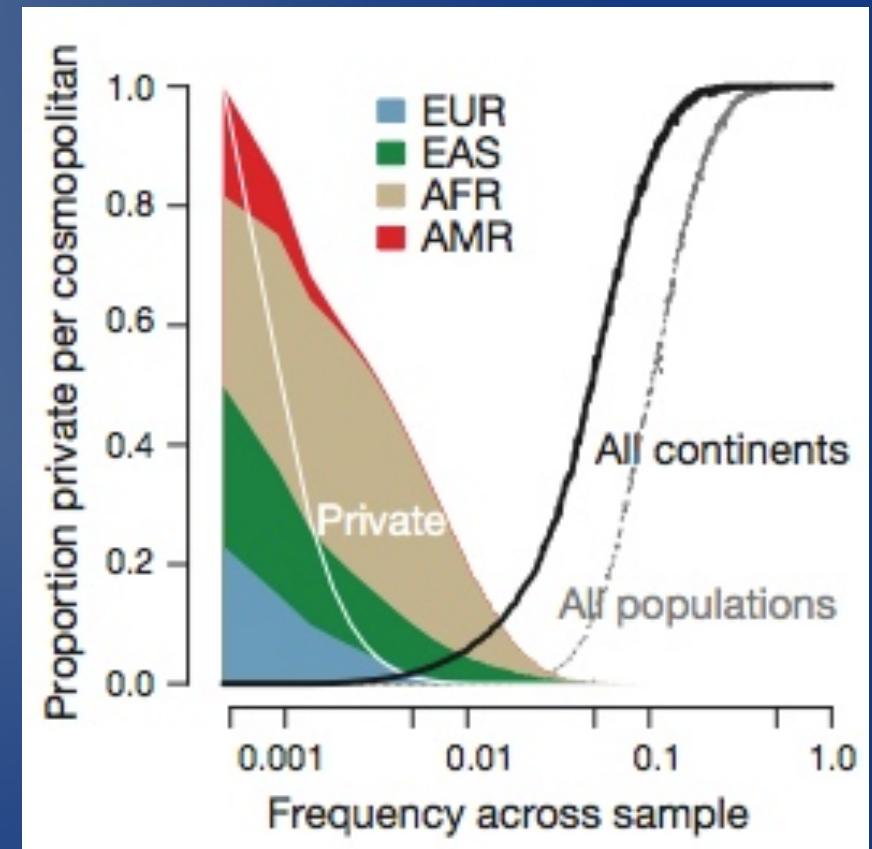


# Designing a sequencing study: properties to consider

Rare or **common** disease?

Population

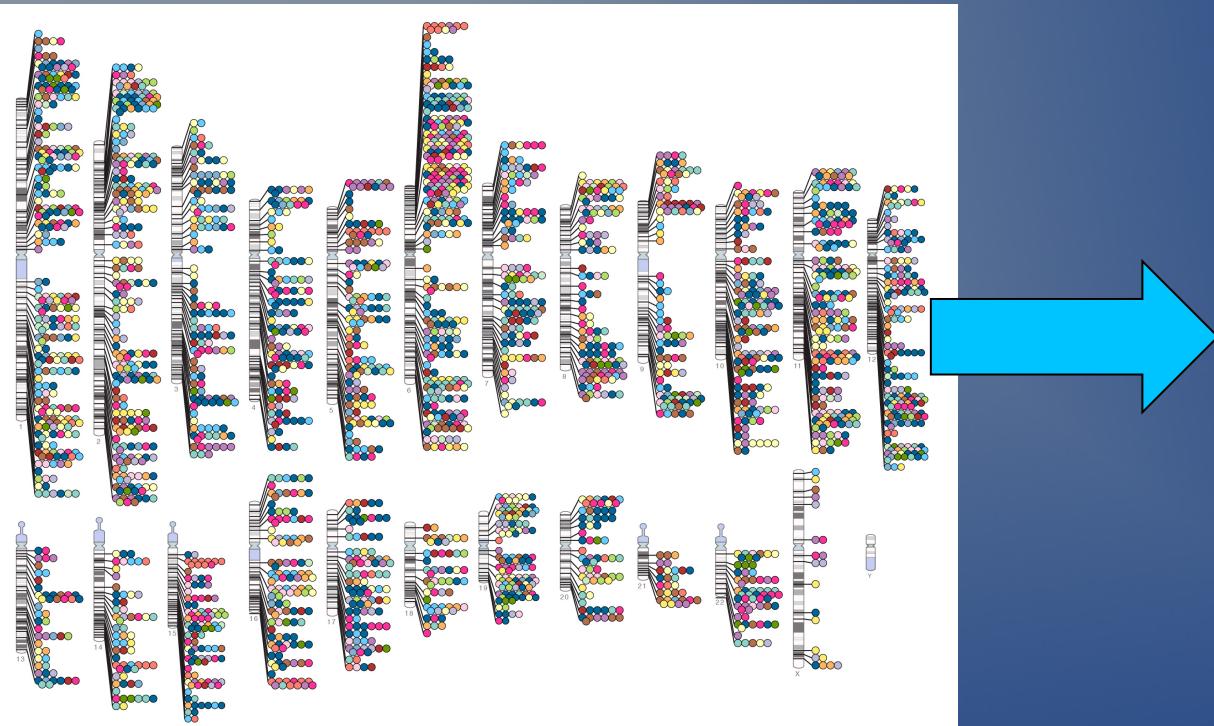
Larger genetic diversity



1000 Genomes Project

Wikipedia

# From common variant association studies to rare variant association studies

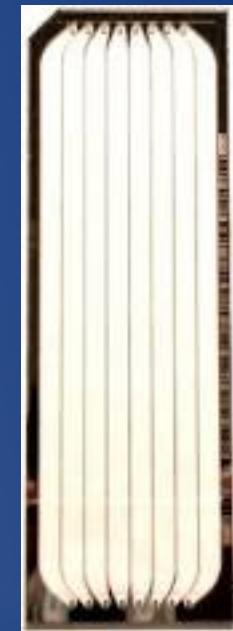


NHGRI GWAS  
catalog

Common  
variants

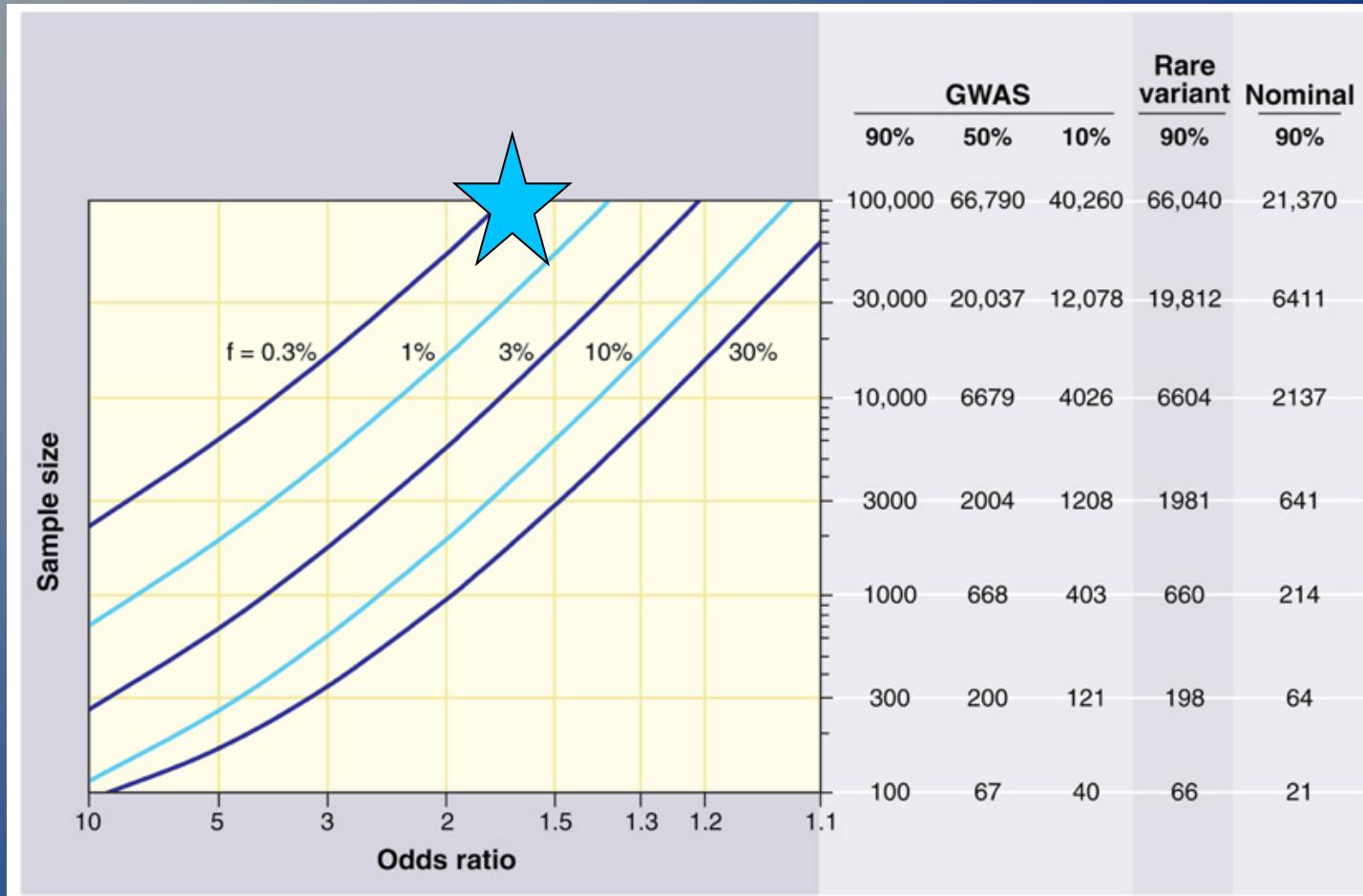


288mm<sup>2</sup>



Common, rare,  
and private  
variants

# How do we analyze rare variants?



Altshuler, Daly, Lander *Science*  
2009.

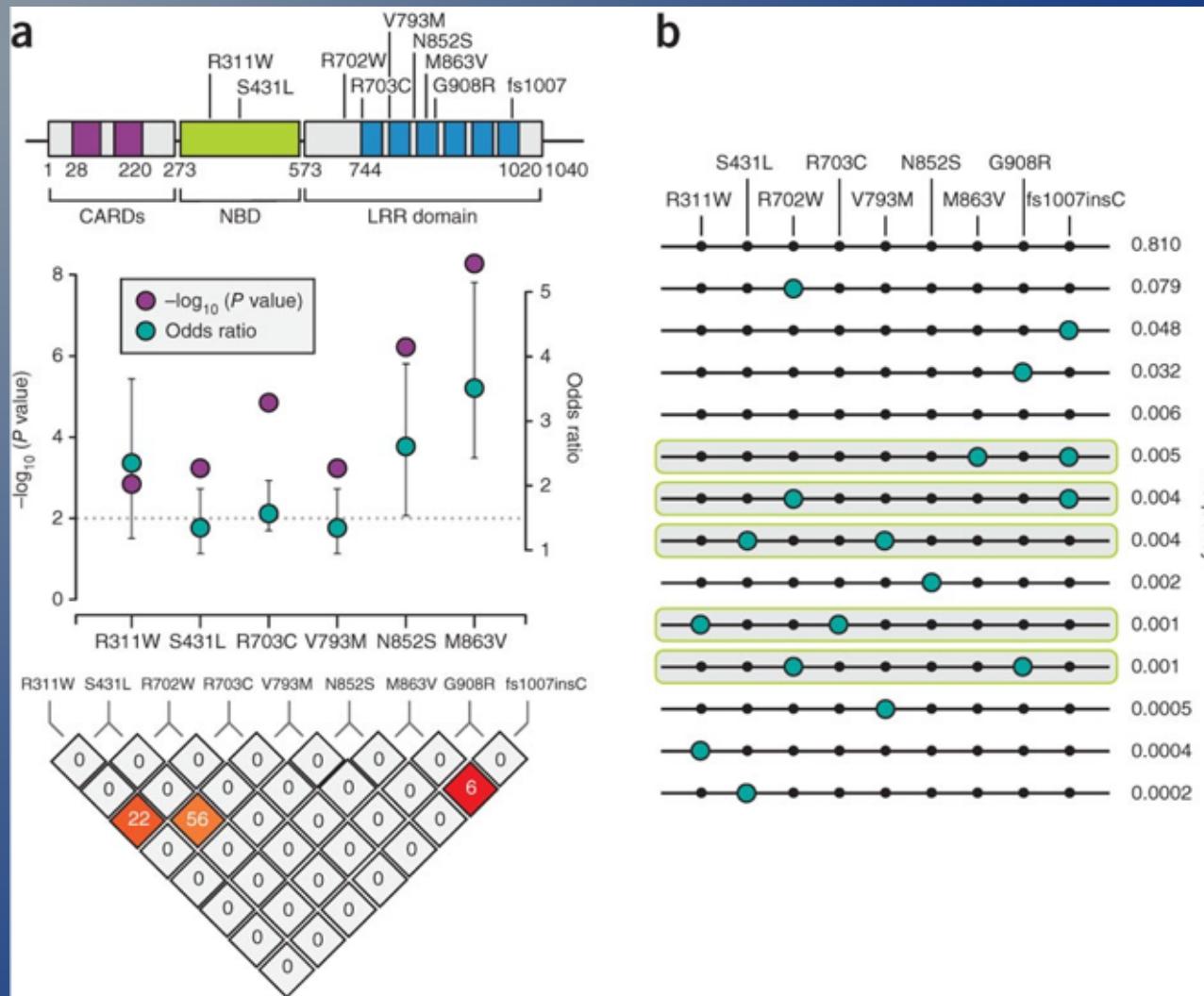
For a single variant with .3% MAF  
we achieve 90% power to detect  
association at 100K samples for  
OR > 1.6.

# Exploit allelic heterogeneity

Multiple independent disease alleles in close proximity

e.g. protein-coding gene

# Allelic heterogeneity



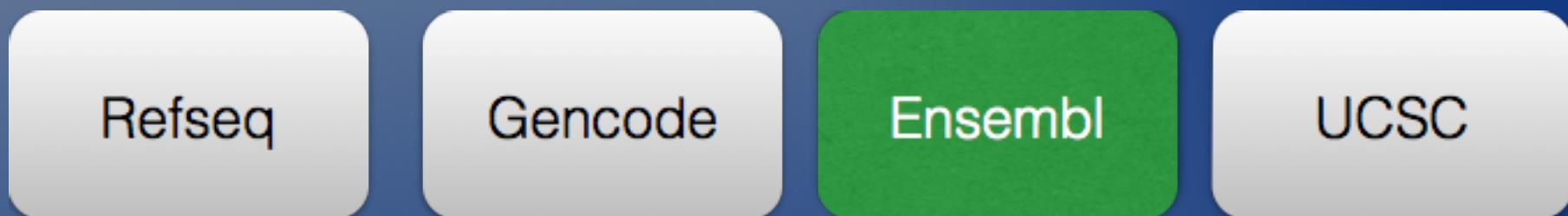
Multiple independent rare variants in *NOD2* confer risk to Crohn's disease

# Predicting the functional consequences of genetic variants

Take a genomic variant



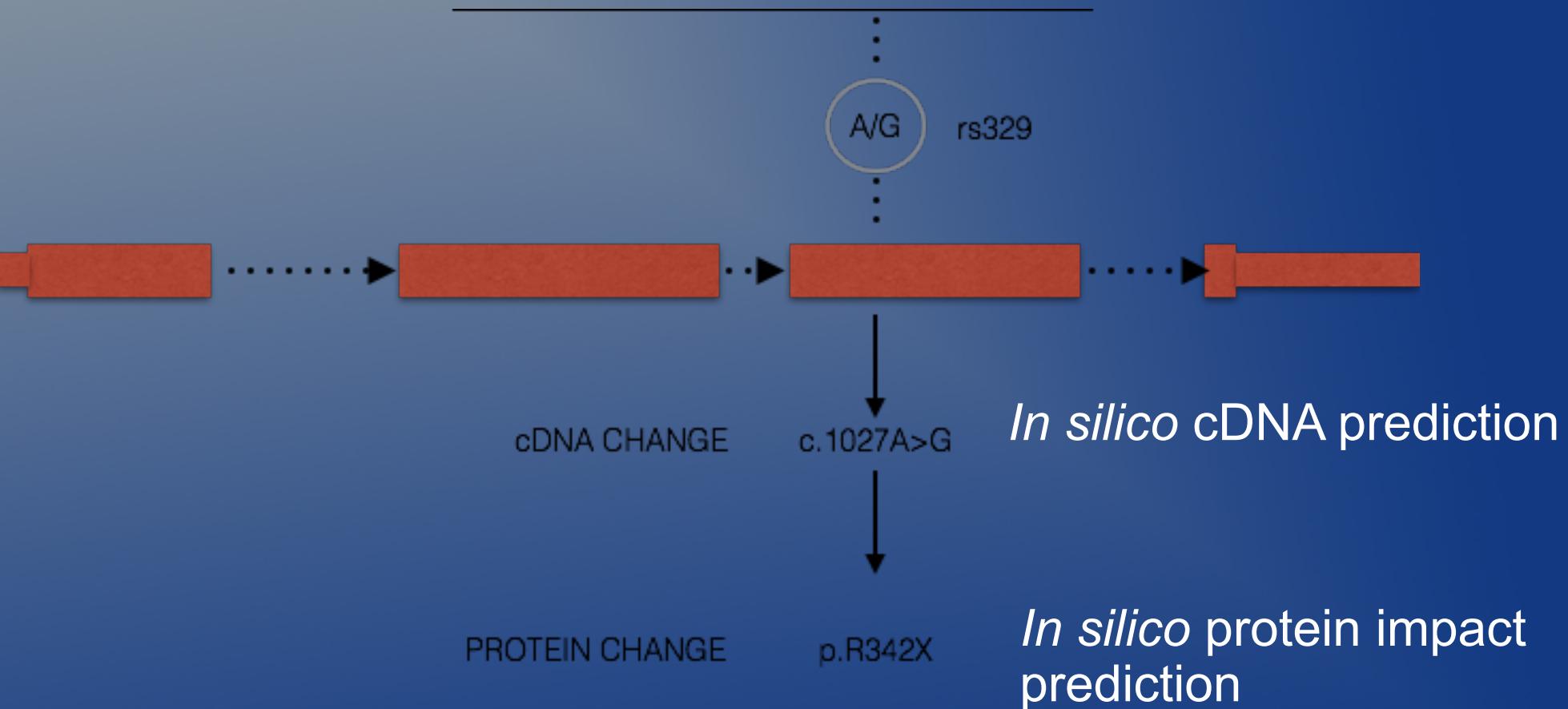
Take a reference transcript set



Perform *in silico* predictions of impact on transcript and translation with transcript set models



# Predicting the functional consequences of genetic variants



# PLINK/SEQ

a toolset for next generation sequencing (NGS) data sets

The screenshot shows the PLINK/SEQ website homepage. The header includes the logo "PLINK/SEQ" and the tagline "A library for the analysis of genetic variation data". The top navigation bar has links for Home, Overview, Download, and Installation. The main content area features a section titled "A toolset for working with human genetic variation data". Below this, there is a paragraph about the library's purpose, a "Downloads" section, a "Getting Started" section with a bulleted list of links, and a "Support" section. On the right side, there is a sidebar with sections for "Getting started" (links to PLINK/SEQ 101 and Extended tutorial), "Key concepts" (links to Project structure, Variants and samples, Meta-information, and Masks), and "PSEQ documentation" (links to Basic syntax, Project management, Main data input, Auxiliary data input, Viewing data, Eval expressions, Data output, Summary statistics, Association tests, Locus DBs, Other reference DBs, and Misc. commands).

- VCF as primary input
- Focus on analysis of rare variants
- Extensible meta-information on locus, genotypes, individuals
- Bundled with key reference databases that can be directly intersected with one's own data
- Command-line, R, and Python library

<http://atgu.mgh.harvard.edu/plinkseq/>

# DNA sequence variant annotation

```
In [9]: pyplinkseq.annotate_load('refseq')
```

Load a locus set

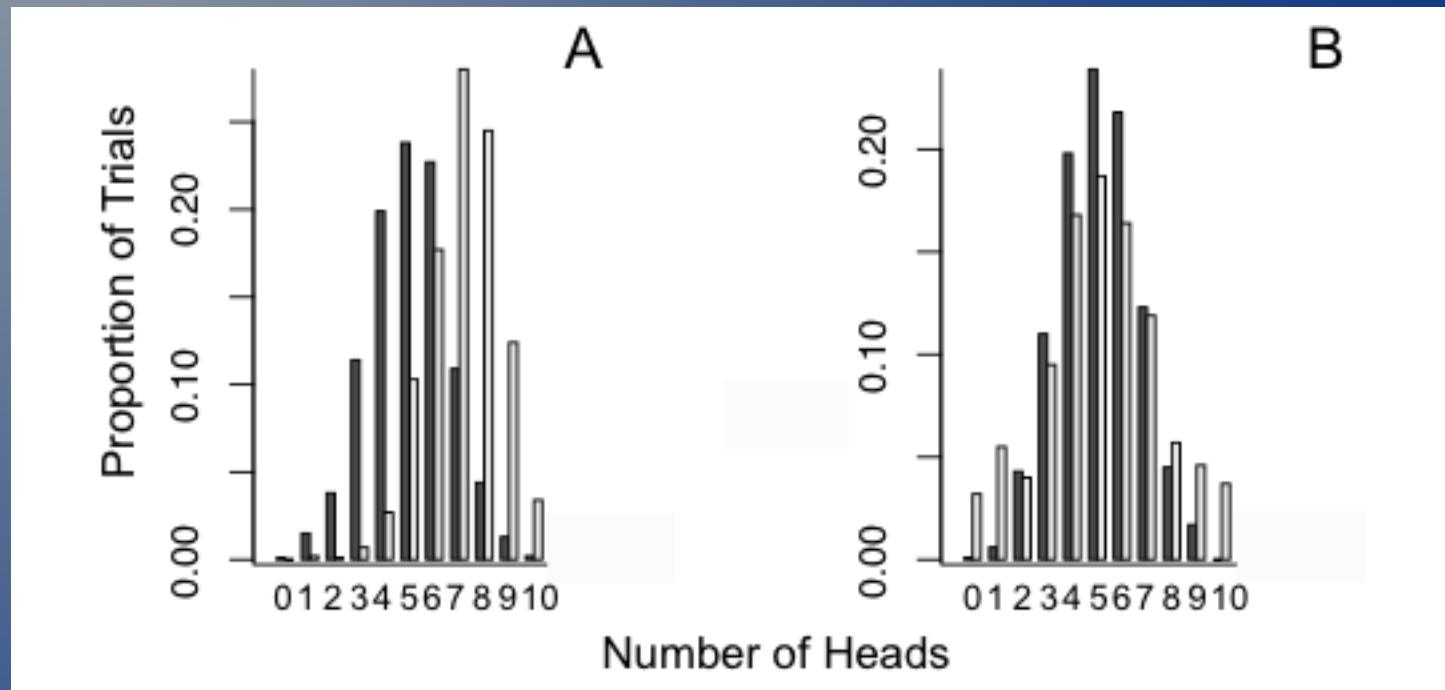
```
In [10]: pyplinkseq.annotate(9,125391241,'G','A','annot',"")  
Out[10]: 'nonsense'
```

Annotate a variant

## pyplinkseq

Python package with extensions  
to the PLINK/SEQ library

# Statistical models for rare variants



Focus on the overall shift in the mean of the distribution

Focus on the overall shift in the variance of the distribution

Aggregate signal across multiple variants

# A very active area of research

OPEN  ACCESS Freely available online

PLOS GENETICS

## A Groupwise Association Test for Rare Mutations Using a Weighted Sum Statistic

Bo Es  
1 Bioinfo

All  
Re  
wi  
ge  
lar  
of  
re  
is p  
a r  
wh  
on  
as

### ARTICLE

#### Pooled Association Tests for Rare Variants in Exon-Resequencing Studies

Alkes L.  
Lee-Jen

#### Rare-Variant Association Testing for Sequencing Data with the Sequence Kernel Association Test

 Like  ARTICLE

Michael C.

<sup>1</sup> Department  
<sup>2</sup> Department  
<sup>3</sup> Department  
<sup>4</sup> Department  
Stud<sup>1,2,3,4</sup>

 Corresp  Like

FOC Seunggi  
shift Nickers  
the Wurfel<sup>8</sup>  
1 Departm

Aggre

B

OPEN  ACCESS Freely available online

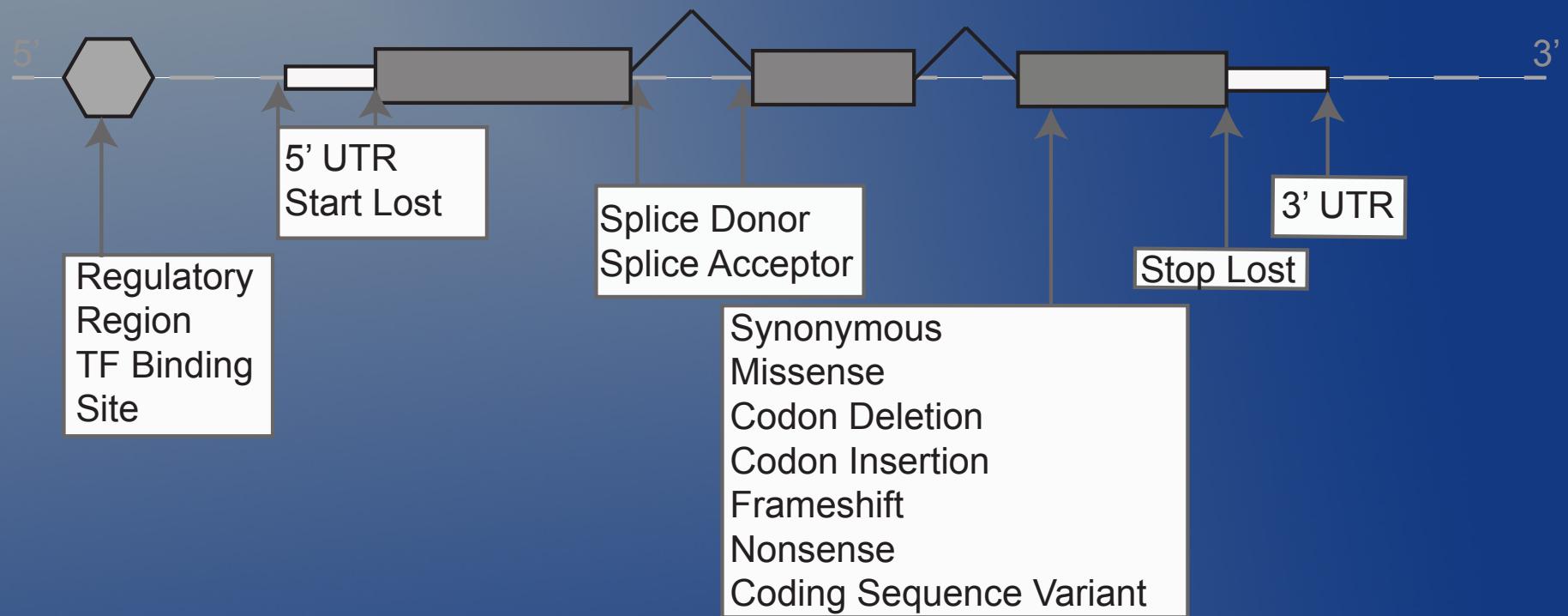
PLOS GENETICS

## Testing for an Unusual Distribution of Rare Variants

Benjamin M. Neale<sup>1,2\*</sup>, Manuel A. Rivas<sup>1,2\*</sup>, Benjamin F. Voight<sup>1,2</sup>, David Altshuler<sup>2,3,4</sup>, Bernie Devlin<sup>5</sup>, Marju Orho-Melander<sup>6</sup>, Sekar Kathiresan<sup>1,2,7,8</sup>, Shaun M. Purcell<sup>2,9</sup>, Kathryn Roeder<sup>10\*</sup>, Mark J. Daly<sup>1,2\*</sup>

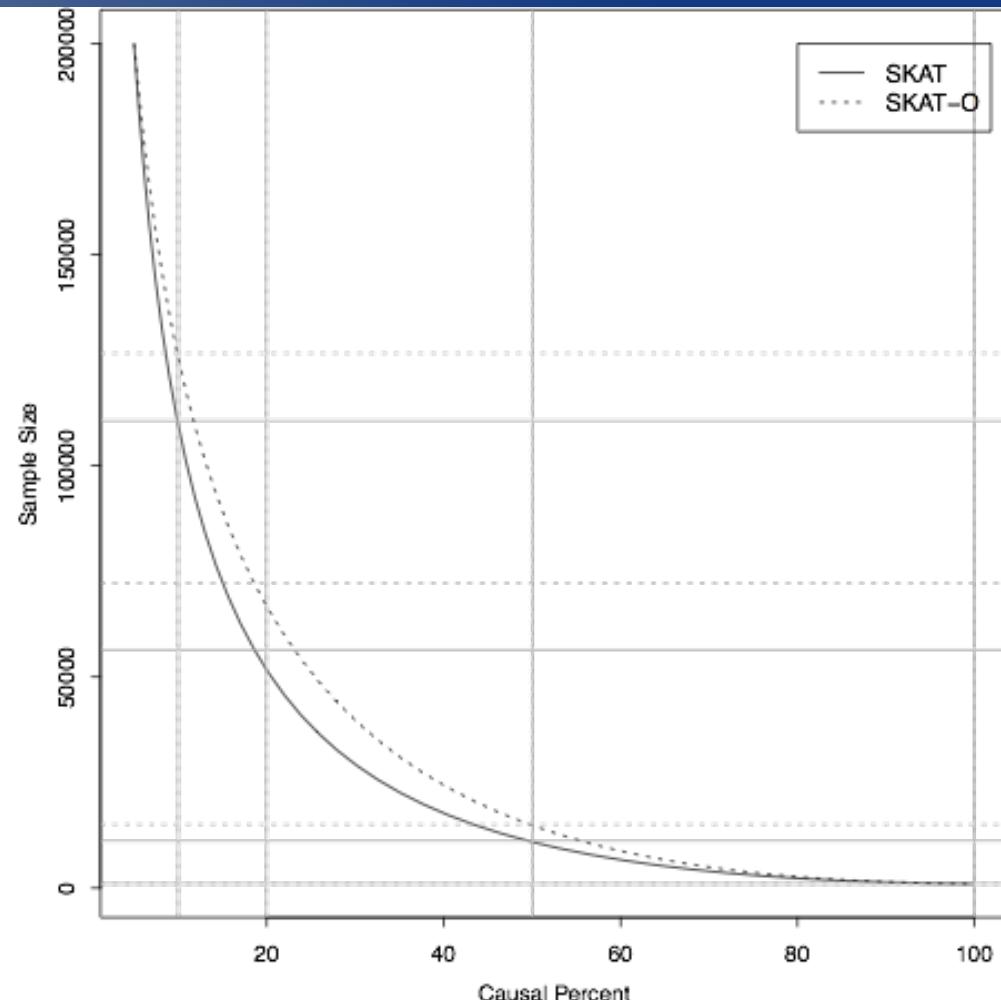
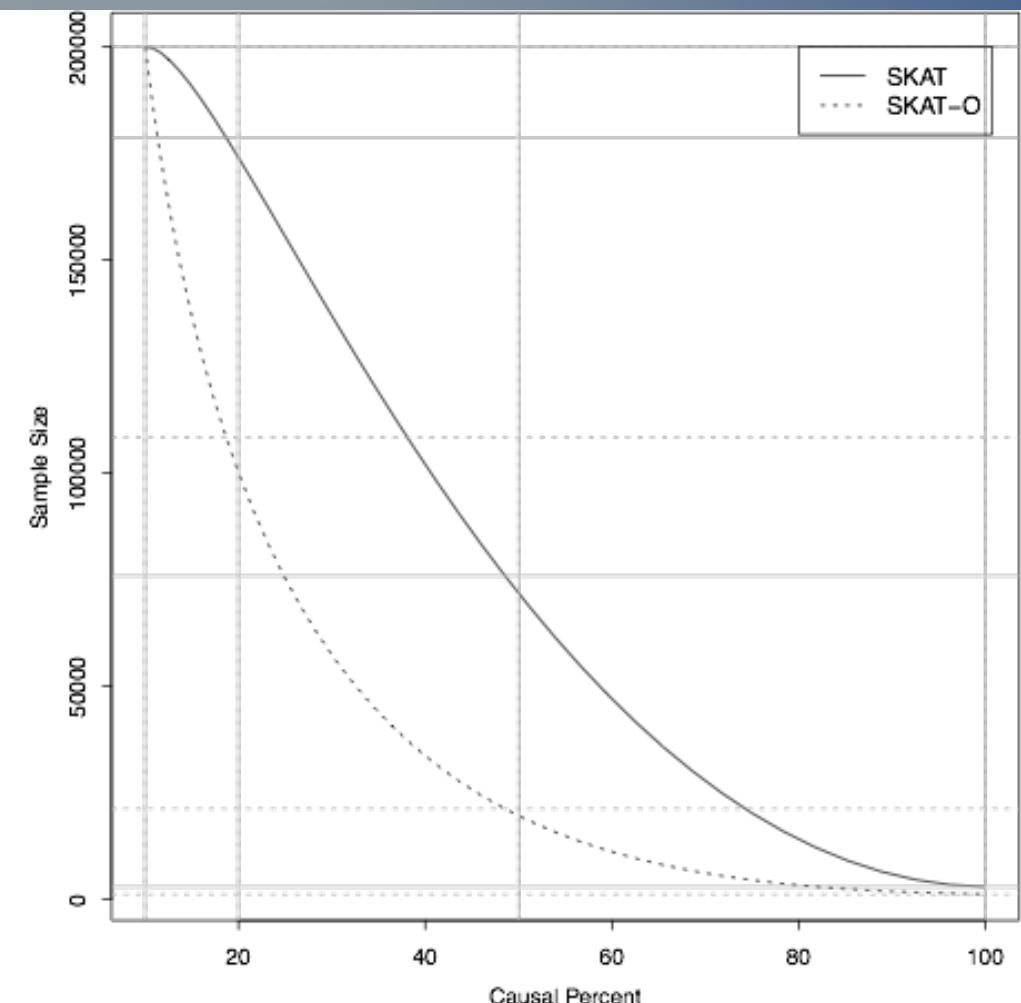
<sup>1</sup> The Center for Human Genetic Research, Massachusetts General Hospital, Boston, Massachusetts, United States of America, <sup>2</sup> The Broad Institute of Harvard and MIT, Cambridge, Massachusetts, United States of America, <sup>3</sup> Department of Genetics, Harvard Medical School, Boston, Massachusetts, United States of America, <sup>4</sup> Department of Molecular Biology, Massachusetts General Hospital, Boston, Massachusetts, United States of America, <sup>5</sup> Department of Psychiatry, University of Pittsburgh School of Medicine, Pittsburgh, Pennsylvania, United States of America, <sup>6</sup> Department of Clinical Sciences Malmö, Diabetes and Cardiovascular Diseases, Genetic Epidemiology CRC, University Hospital Malmö, Malmö, Sweden, <sup>7</sup> Cardiovascular Research Center, Massachusetts General Hospital, Boston, Massachusetts, United States of America, <sup>8</sup> Department of Medicine, Harvard Medical School, Boston, Massachusetts, United States of America, <sup>9</sup> Psychiatric and Neurodevelopmental Genetics Unit, Massachusetts General Hospital,

# How do we analyze rare variants?



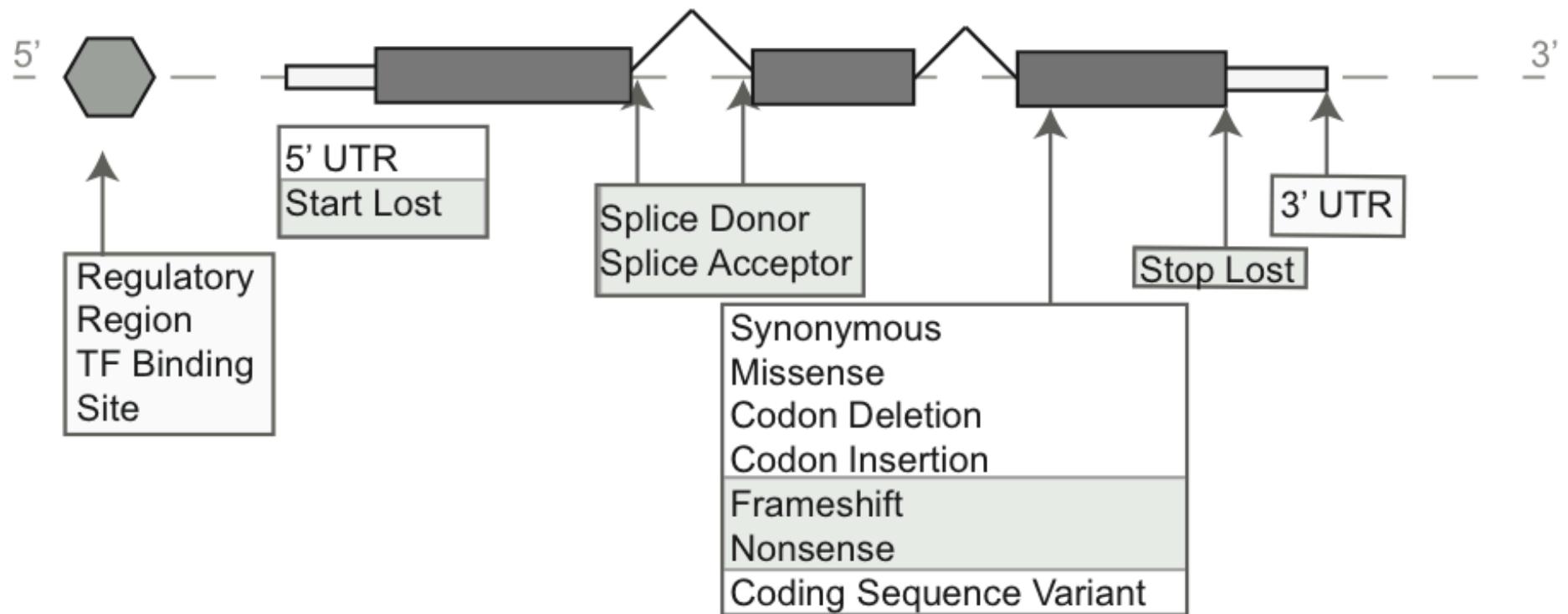
Consider variant annotation

# Power analysis



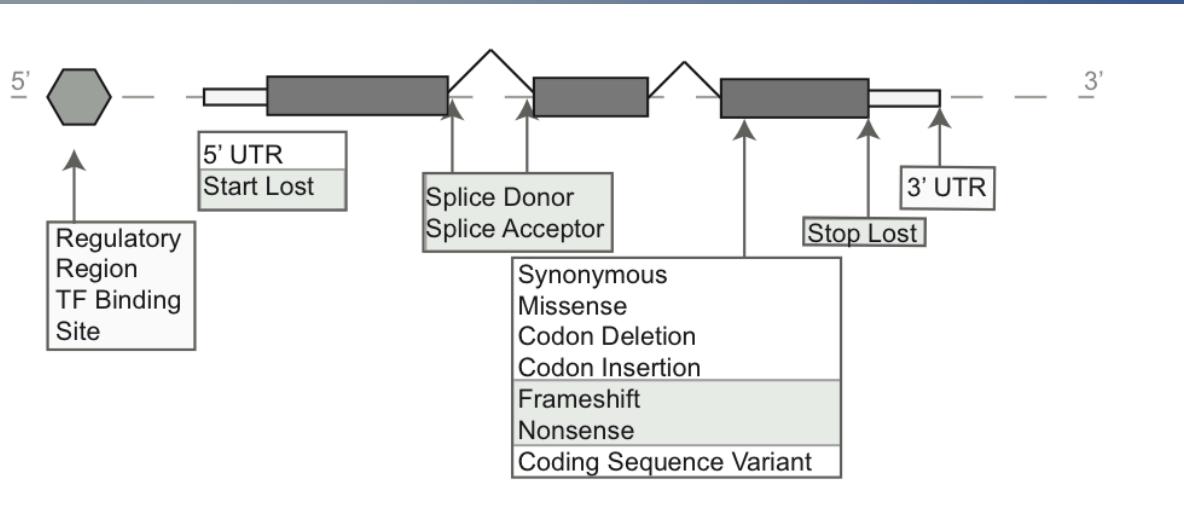
Increasing the proportion of null variants tested, quickly increases the sample size required to identify association.

# Focusing on high impact variants

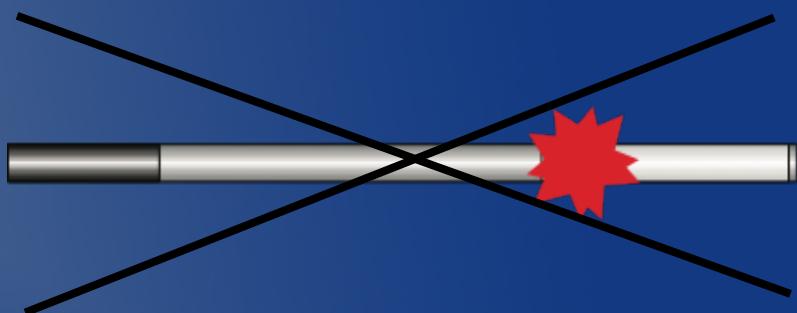


A common approach is to focus on predicted high impact variants. Formally called protein truncating variants.

# Focusing on high impact variants



Potentially functional  
or toxic truncated  
protein product



NMD - Nonsense Mediated Decay

No protein product

Annotation program now part of PLINK/SEQ:  
<http://atgu.mgh.harvard.edu/plinkseq/>

# Disruptive protein truncating variants

Ample evidence from studies of mendelian phenotypes, common diseases, and quantitative traits that protein truncating variants play an important role in the genetic architecture of human traits.

**nature genetics**

**ARTICLES**

**Deep resequencing of GWAS loci identifies independent rare variants associated with inflammatory bowel disease**

Manuel A Rivas<sup>1-3</sup>, Mélissa Beaudoin<sup>4,23</sup>, Agnes Gardet<sup>5,23</sup>, Christine Stevens<sup>2,23</sup>, Yashoda Sharma<sup>6</sup>, Clarence K Zhang<sup>6</sup>, Gabrielle Boucher<sup>4</sup>, Stephan Ripke<sup>1,2</sup>, David Ellinghaus<sup>7</sup>, Noel Burtt<sup>8</sup>, Tim Fennell<sup>2</sup>, Andrew Kirby<sup>1,2</sup>, Anna Latiano<sup>8</sup>, Phillippe Goyette<sup>6</sup>, Todd Green<sup>2</sup>, Jonas Halfvarson<sup>9</sup>, Talin Harutunians<sup>10</sup>, Joshua M Korn<sup>2</sup>, Finny Kuruvilla<sup>2,11</sup>, Caroline Lagace<sup>6</sup>, Benjamin Neale<sup>1,2</sup>, Ken Sin Lo<sup>4</sup>, Phil Schumm<sup>12</sup>, Leif Törkvist<sup>13</sup>, National Institute of Diabetes and Digestive Kidney Diseases Inflammatory Bowel Disease Genetics Consortium (NIDDK IBDGC)<sup>14</sup>, United Kingdom Inflammatory Bowel Disease Genetics Consortium<sup>14</sup>, International Inflammatory Bowel Disease Genetics Consortium<sup>14</sup>, Marla C Dubinsky<sup>15</sup>, Steven R Brant<sup>16,17</sup>, Mark S Silverberg<sup>18</sup>, Richard H Duerr<sup>19,20</sup>, David Altschuler<sup>1,2</sup>, Stacey Gabriel<sup>3</sup>, Guillaume Lettre<sup>4</sup>, Andre Franke<sup>2</sup>, Mauro D'Amato<sup>21</sup>, Dermot P B McGovern<sup>10,22</sup>, Judy H Cho<sup>6</sup>, John D Rioux<sup>4</sup>, Ramnik J Xavier<sup>1,2,5</sup> & Mark J Daly<sup>1,2</sup>

More than 1,000 susceptibility loci have been identified through genome-wide association studies (GWAS) of common variants; however, the specific genes and full allelic spectrum of causal variants underlying these findings have not yet been defined. Here we used pooled next-generation sequencing to study 56 genes from regions associated with Crohn's disease in 350 cases and 350 controls. Through follow-up genotyping of 70 rare and low-frequency protein-altering variants in nine independent case-control series (16,054 Crohn's disease cases, 12,153 ulcerative colitis cases and 17,575 healthy controls), we identified four additional independent risk factors in NOD2, two additional protective variants in NLRP1, a highly significant association with a protective

The NEW ENGLAND JOURNAL of MEDICINE

ORIGINAL ARTICLE

## Sequence Variations in PCSK9, Low LDL, and Protection against Coronary Heart Disease

Jonathan C. Cohen, Ph.D., Eric Boerwinkle, Ph.D., Thomas H. Mosley, Jr., Ph.D., and Helene H. Hsueh, M.D.

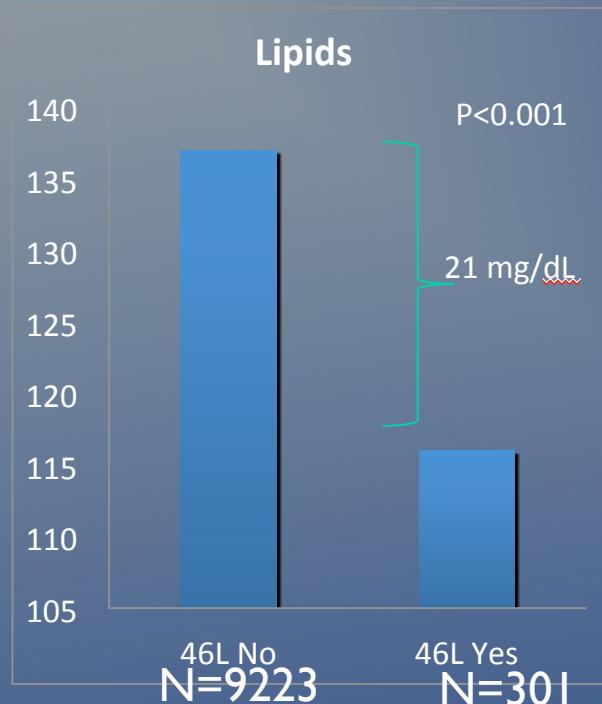
The NEW ENGLAND JOURNAL of MEDICINE

ORIGINAL ARTICLE

## Effect of a Monoclonal Antibody to PCSK9 on LDL Cholesterol

Evan A. Stein, M.D., Ph.D., Scott Mellis, M.D., Ph.D., George D. Yancopoulos, M.D., Ph.D., Neil Stahl, Ph.D., Douglas Logan, M.D., William B. Smith, M.D., Eleanor Lisbon, M.D., M.P.H., Maria Gutierrez, M.D., Cheryle Webb, M.D., Richard Wu, Ph.D., Yunling Du, Ph.D.,

# PCSK9 loss-of-function variant lowers LDL and protects against coronary artery disease

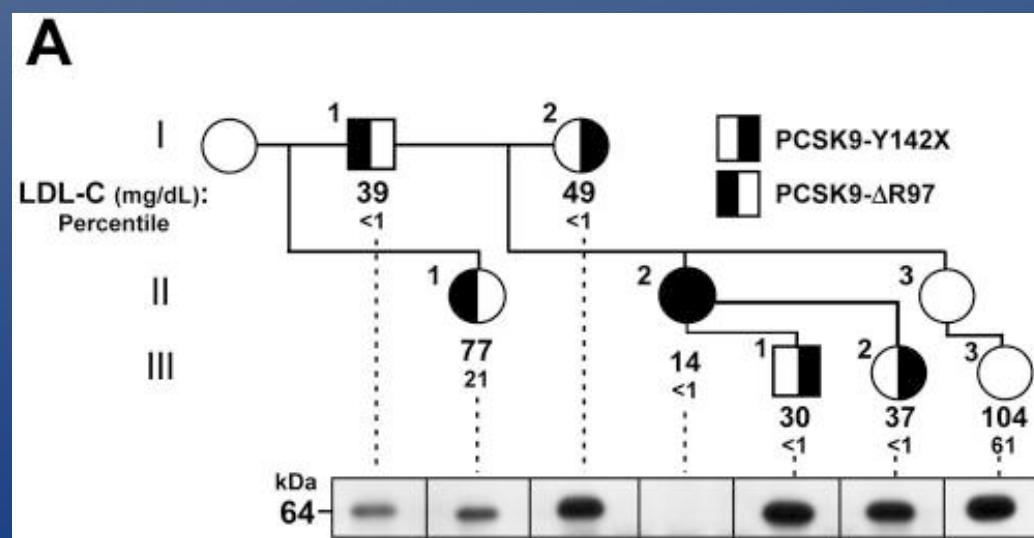


## Outcome

Hazard Ratio for incident CHD:  
0.50 (0.32 – 0.79)  
P=0.003  
1108 events/9524

Cohen, *N Engl J Med* 2006  
Kathiresan, *N Engl J Med* 2008

A



Zhao, Tuakli-Wosornu, et al (2006) American Journal Human Genetics

From Mark Daly

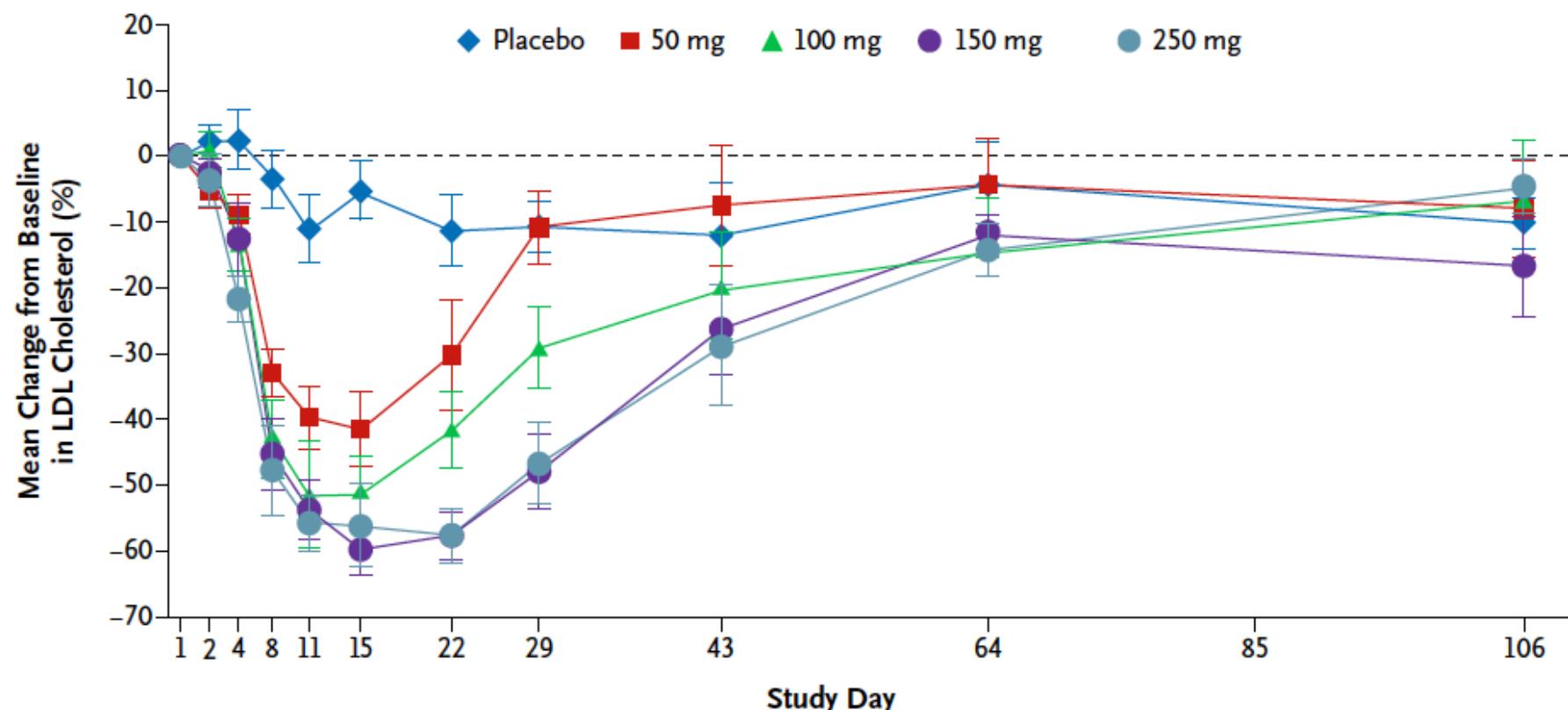
## ORIGINAL ARTICLE

## Effect of a Monoclonal Antibody to PCSK9 on LDL Cholesterol

Evan A. Stein, M.D., Ph.D., Scott Mellis, M.D., Ph.D.,  
George D. Yancopoulos, M.D., Ph.D., Neil Stahl, Ph.D., Douglas Logan, M.D.,  
William B. Smith, M.D., Eleanor Lisbon, M.D., M.P.H., Maria Gutierrez, M.D.,  
Cheryle Webb, M.D., Richard Wu, Ph.D., Yunling Du, Ph.D.,  
Therese Kranz, R.N., M.B.A., Evelyn Gasparino, B.S.,  
and Gary D. Swerdlow, M.D., Ph.D.

As predicted by genetics – PCSK9 inhibition appears safe and effective

## B Subcutaneous Administration



# Take advantage of polygenic nature of diseases

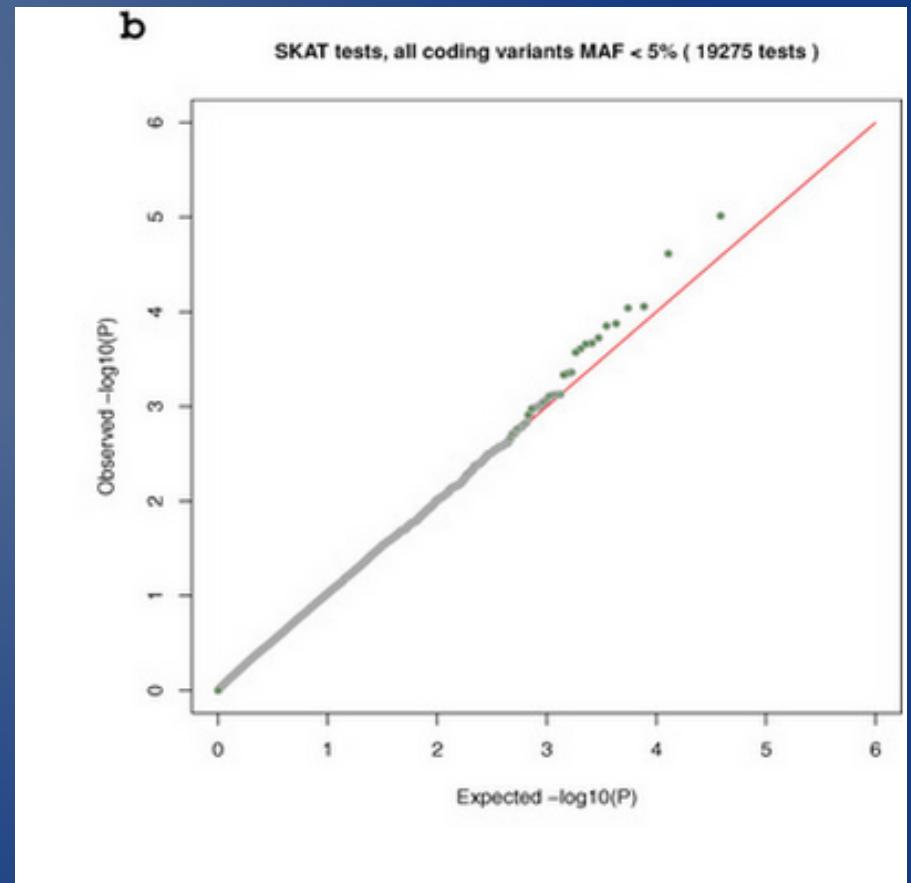
Apply model to sets of genes with prior established relationships

e.g. GWAS, sets of genes causal for monogenic forms of disease, pathways?

Polygenic signatures in rare variants

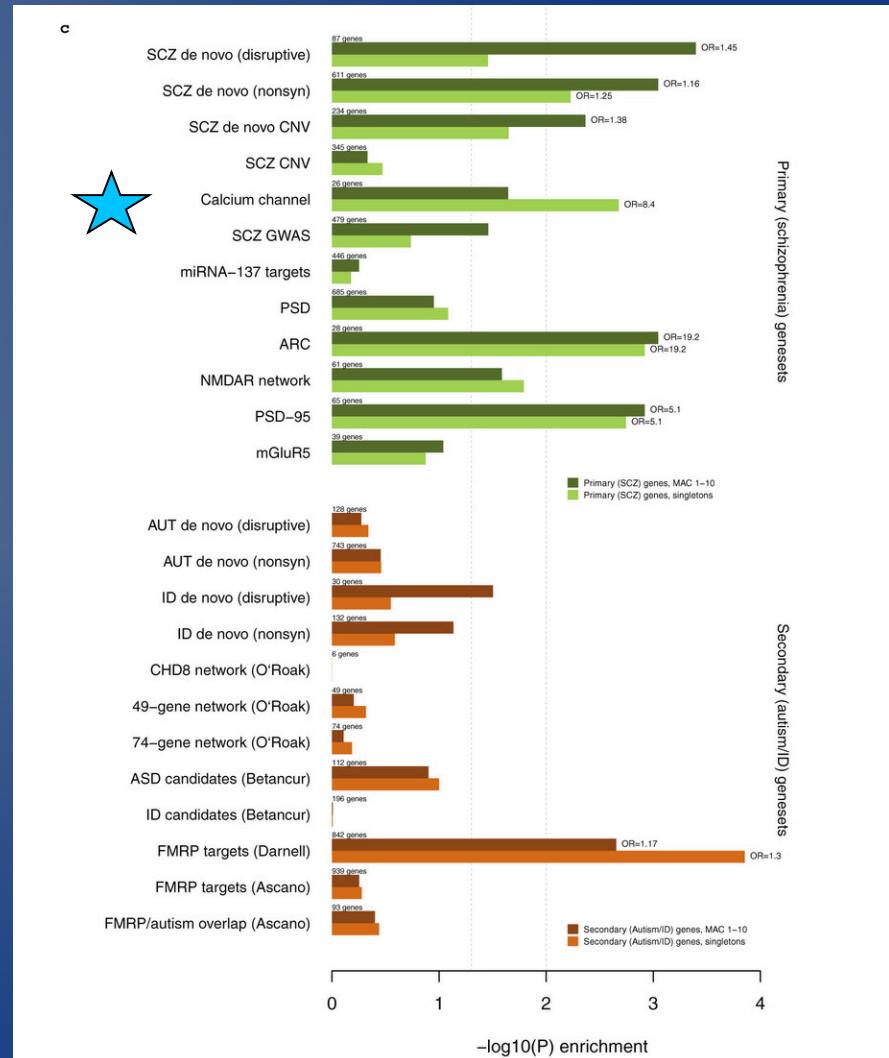
# Take advantage of polygenic nature of diseases

Q-Q plot that follows pattern of null distribution, i.e. no signal.



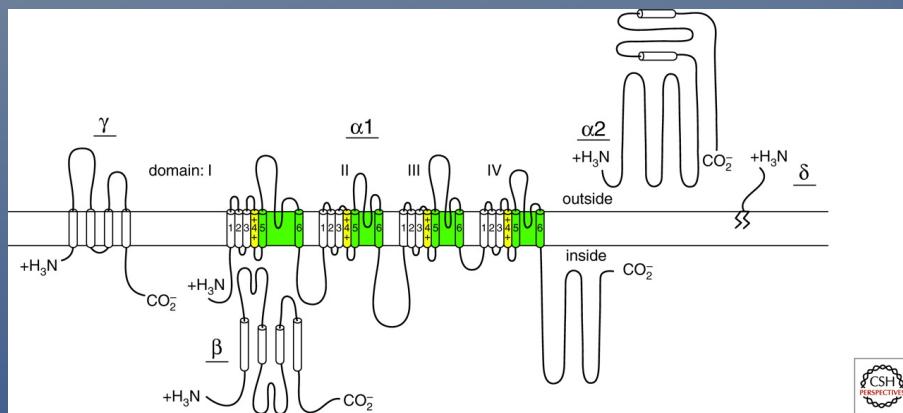
# Take advantage of polygenic nature of diseases

Schizophrenia results from aggregating disruptive PTVs identified in genes in sets of genes

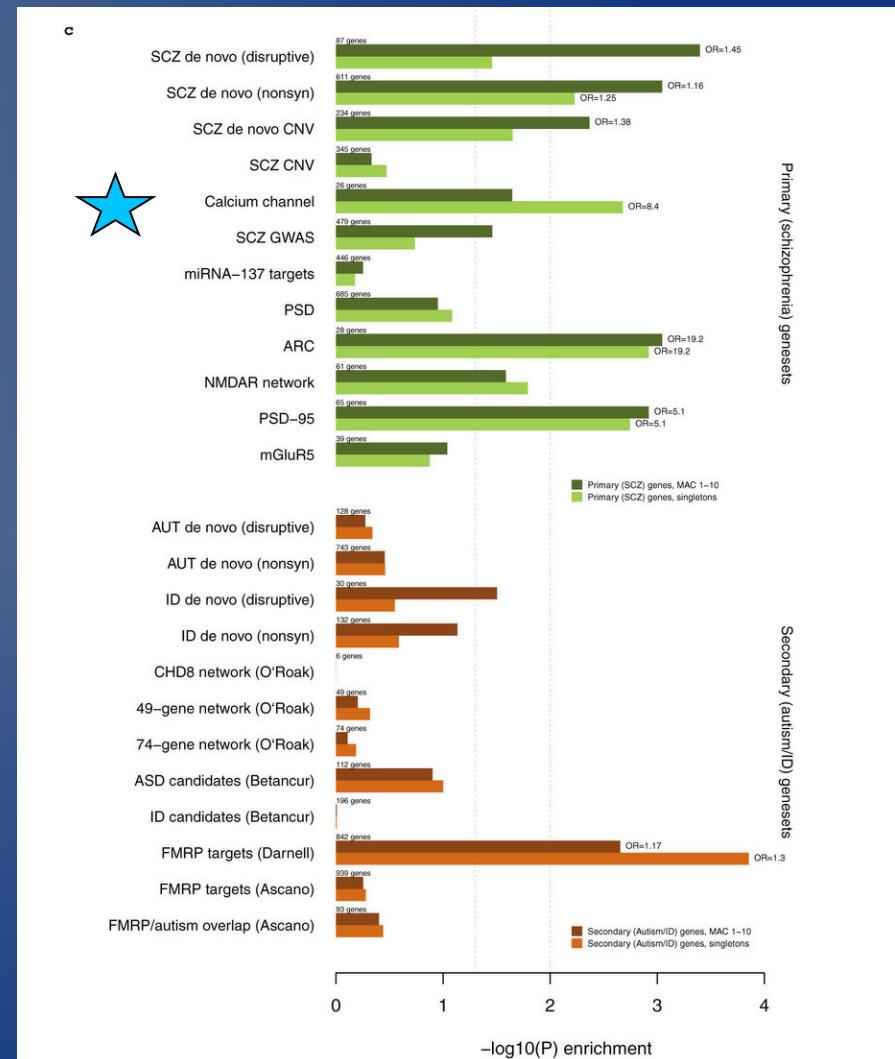


# Take advantage of polygenic nature of diseases

## Voltage gated calcium channels



Catterall W A Cold  
Spring Harb Perspect  
Biol 2011;3:a003947



Purcell et al. *Nature* 2014.

# Aggregate rare variant association tests

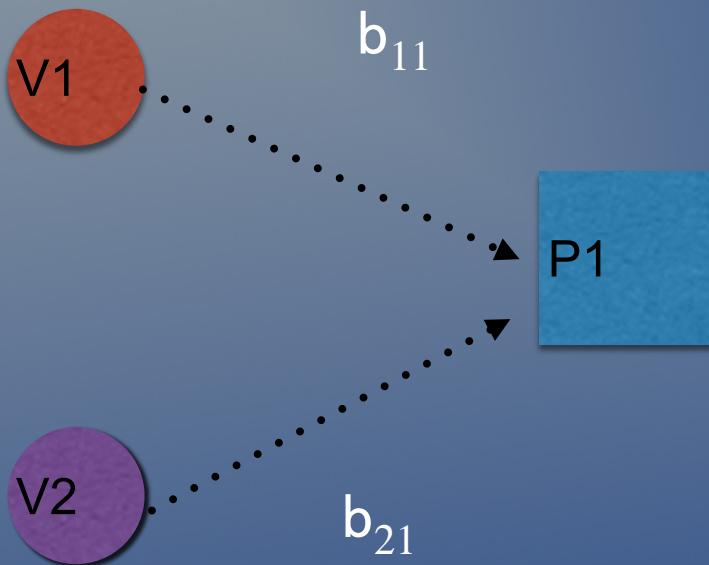
Implicit assumption about the correlation of genetic effects on a phenotype.

## Recap

Aggregate tests: “burden” – **Assume similar effects (highly correlated)**

Aggregate tests: “dispersion” – **Assume no correlation of genetic effects (independent)**

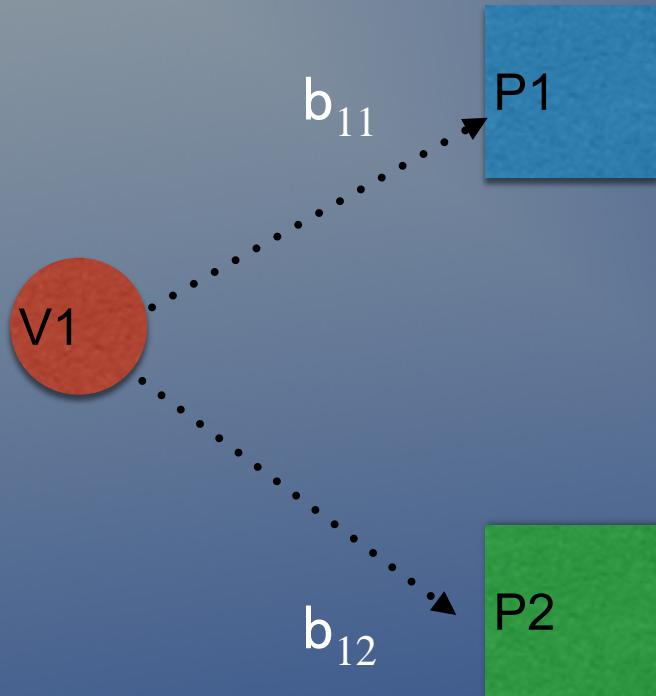
# Take advantage of the concept of correlation of effects



If we assume the correlation of effects ( $V1 \rightarrow P1$ ) and ( $V2 \rightarrow P1$ ) to be 1, then similar to collapsing.

Correlation of effects between two variants and one trait.

# Take advantage of the concept of correlation of effects



We can have a similar treatment on multiple traits for any variant we can make an assumption about the correlation of effects on multiple traits.

Correlation of effects between one variant and two traits.

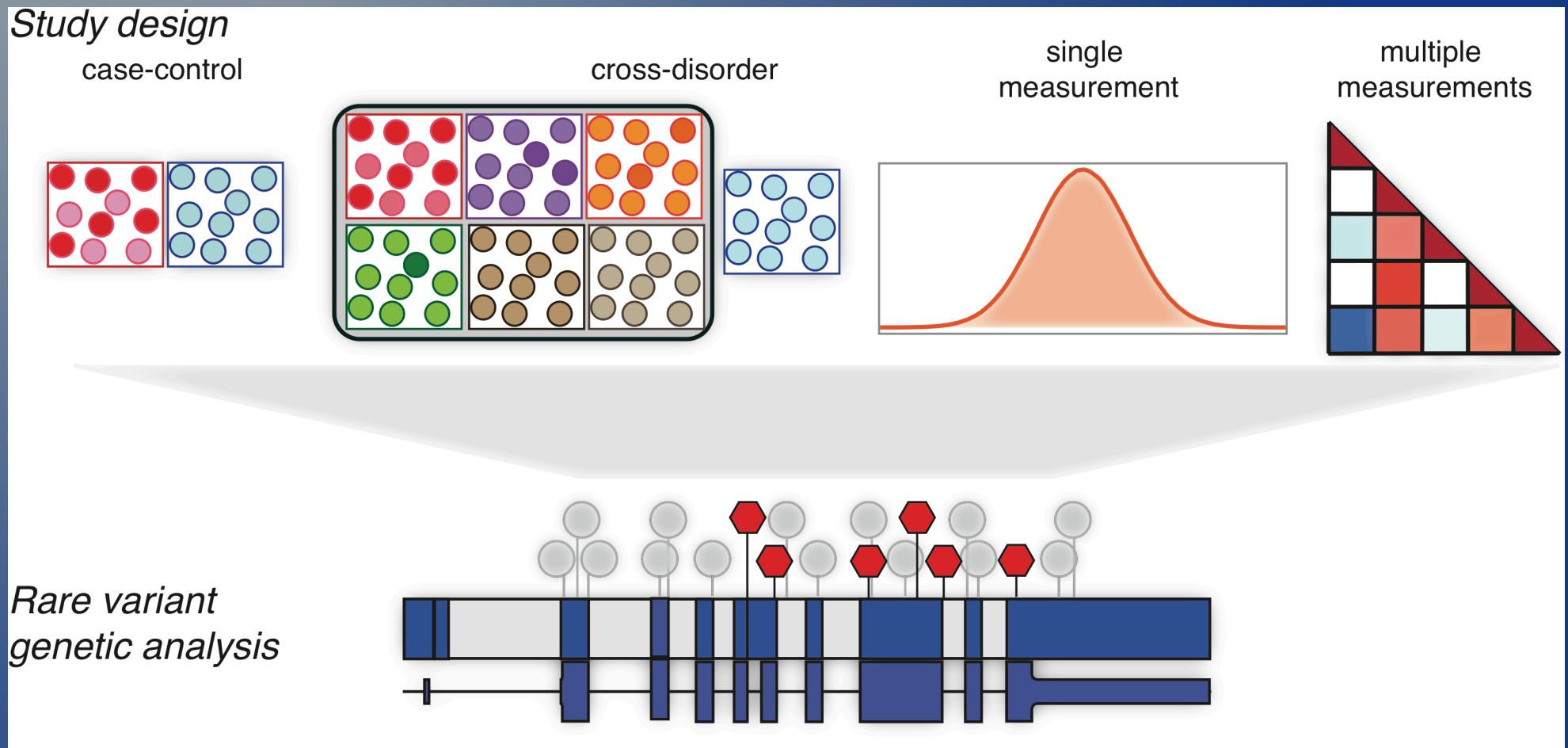
# Derive a test statistic that allows incorporation of the correlation of genetic effects

We can apply the treatment in any framework: SKAT, Bayesian, C-alpha. We choose **C-alpha** and derive a score statistic on the multivariate (cross-phenotype [continuous] analyses) and multinomial likelihoods (cross-disorder analyses).

$$T = \text{sum} \left( (\mathbf{R}_{\text{var}} \otimes \mathbf{R}_{\text{phen}}) \odot \mathbf{A} \right).$$

Binomial C-alpha, Generalised C-alpha, Burden, and Multivariate Wald test are special cases.

# Applied across a broad range of study designs



Statistical framework for  
rare variant association studies

# Application of C-alpha MRP tests to rare variant autoimmune disease study

6 disease groups; 1 shared control group.  
Exon resequencing in over 40k individuals.

## LETTER

doi:10.1038/nature12170

### Negligible impact of rare autoimmune-locus coding-region variants on missing heritability

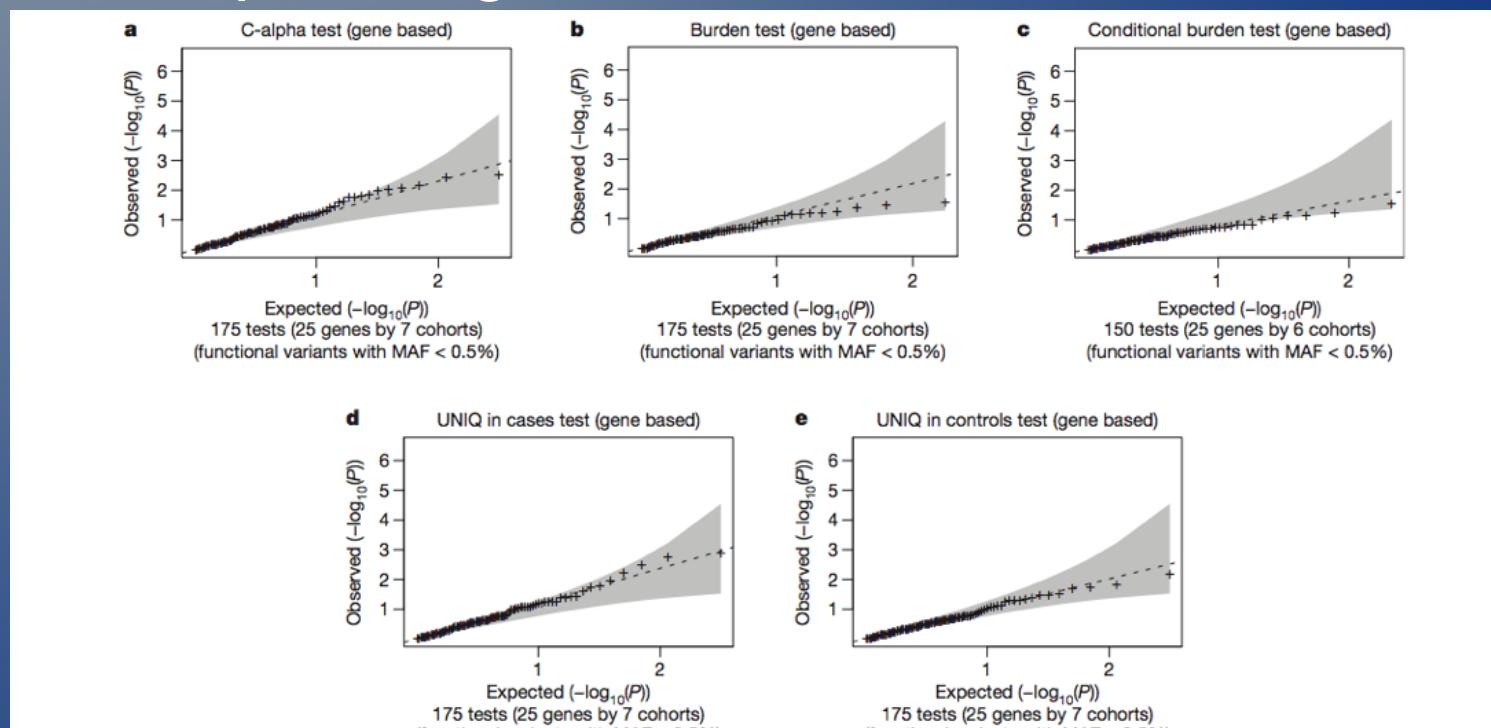
Karen A. Hunt<sup>1</sup>, Vanisha Mistry<sup>1</sup>, Nicholas A. Bockett<sup>1</sup>, Tariq Ahmad<sup>2</sup>, Maria Ban<sup>3</sup>, Jonathan N. Barker<sup>4</sup>, Jeffrey C. Barrett<sup>5</sup>, Hannah Blackburn<sup>5</sup>, Oliver Brand<sup>6</sup>, Oliver Burren<sup>7</sup>, Francesca Capon<sup>4</sup>, Alastair Compston<sup>3</sup>, Stephen C. L. Gough<sup>6</sup>, Luke Jostins<sup>8</sup>, Yong Kong<sup>9</sup>, James C. Lee<sup>10</sup>, Monkol Lek<sup>11</sup>, Daniel G. MacArthur<sup>11</sup>, John C. Mansfield<sup>12</sup>, Christopher G. Mathew<sup>4</sup>, Charles A. Mein<sup>13</sup>, Muddassar Mirza<sup>4</sup>, Sarah Nutland<sup>7</sup>, Suna Onengut-Gumuscu<sup>14</sup>, Efterpi Papouli<sup>4</sup>, Miles Parkes<sup>10</sup>, Stephen S. Rich<sup>14</sup>, Steven Sawcer<sup>3</sup>, Jack Satsangi<sup>15</sup>, Matthew J. Simmonds<sup>6</sup>, Richard C. Trembath<sup>16</sup>, Neil M. Walker<sup>7</sup>, Eva Wozniak<sup>13</sup>, John A. Todd<sup>7</sup>, Michael A. Simpson<sup>4</sup>, Vincent Plagnol<sup>17</sup> & David A. van Heel<sup>1</sup>

Genome-wide association studies (GWAS) have identified common variants of modest effect size at hundreds of loci for common autoimmune diseases; however, a substantial fraction of heritability remains unexplained, to which rare variants may contribute<sup>1,2</sup>. To discover rare variants and test them for association with a phenotype, most studies re-sequence a small initial sample size and then genotype the discovered variants in a larger sample set<sup>3–5</sup>. This approach fails to analyse a large fraction of the rare variants present in the entire sample set. Here we perform simultaneous amplicon sequencing for rare variants across six autoimmune diseases and a control population for analysis.

There are only a handful of published examples of rare coding-region variants associated with common autoimmune diseases (although many examples in familial/Mendelian immune-mediated diseases). Coding-region variants in *IFIH1* associated with type 1 diabetes (MAF in controls = 0.67–2.2%)<sup>3</sup>, *TYK2* with multiple autoimmune diseases<sup>15</sup> and *IL23R* with inflammatory bowel disease<sup>6</sup>, for example, are low frequency (which we define as MAF = 0.5–5%) rather than particularly rare. In other examples, the existing evidence for association, and/or the effect sizes, are relatively weak (for example, *CARD14* and *IL2RA*<sup>16</sup>, *IL2RA* and *IL2RB* and *IL2RB* and *IL2RA*<sup>17</sup>).

# Application of C-alpha MRP tests to rare variant autoimmune disease study

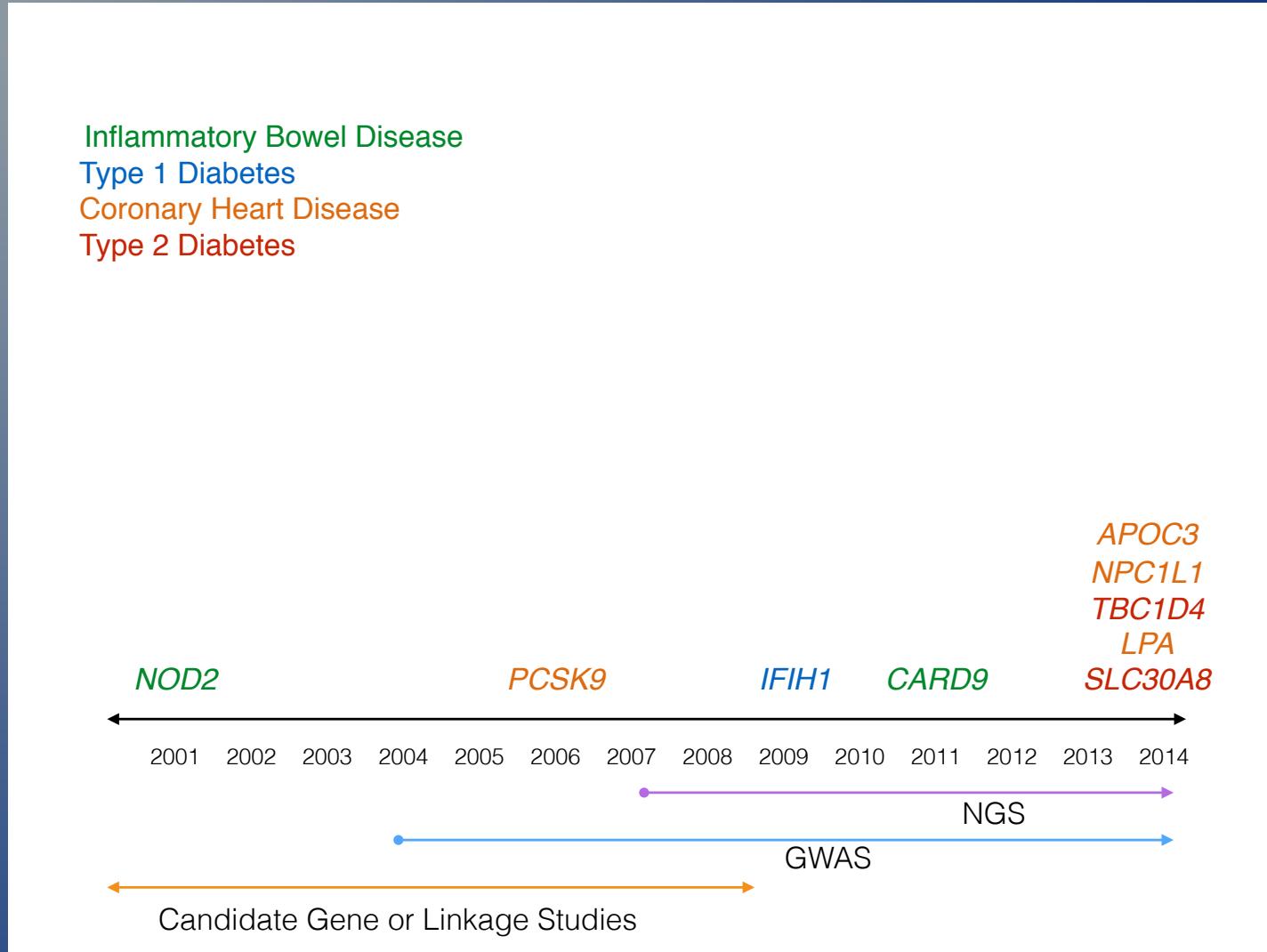
6 disease groups; 1 shared control group.  
Exon resequencing in over 40k individuals.



**Figure 1 | Association analyses of discovered rare functional variants in autoimmune diseases.** We define rare functional variants as MAF < 0.5% in 17,019 controls and predicted nonsynonymous, premature-stop or splice-site annotation. Quantile-quantile plots compare observed versus expected test-statistic distributions, with shading indicating 99% confidence intervals. Full results are available in Supplementary Data. Each of six individual diseases, and all autoimmune diseases combined, were tested as phenotypes. **a**, Gene-based C-alpha test (25 genes by 7 phenotypes,  $n = 41,911$  subjects) allowing for both risk and protective effects for rare functional variants. Singleton variants pooled into a single binomial count per phenotype. **b**, Gene-based burden tests (25 genes by 7 phenotypes,  $n = 41,911$  subjects) comparing summed allele counts

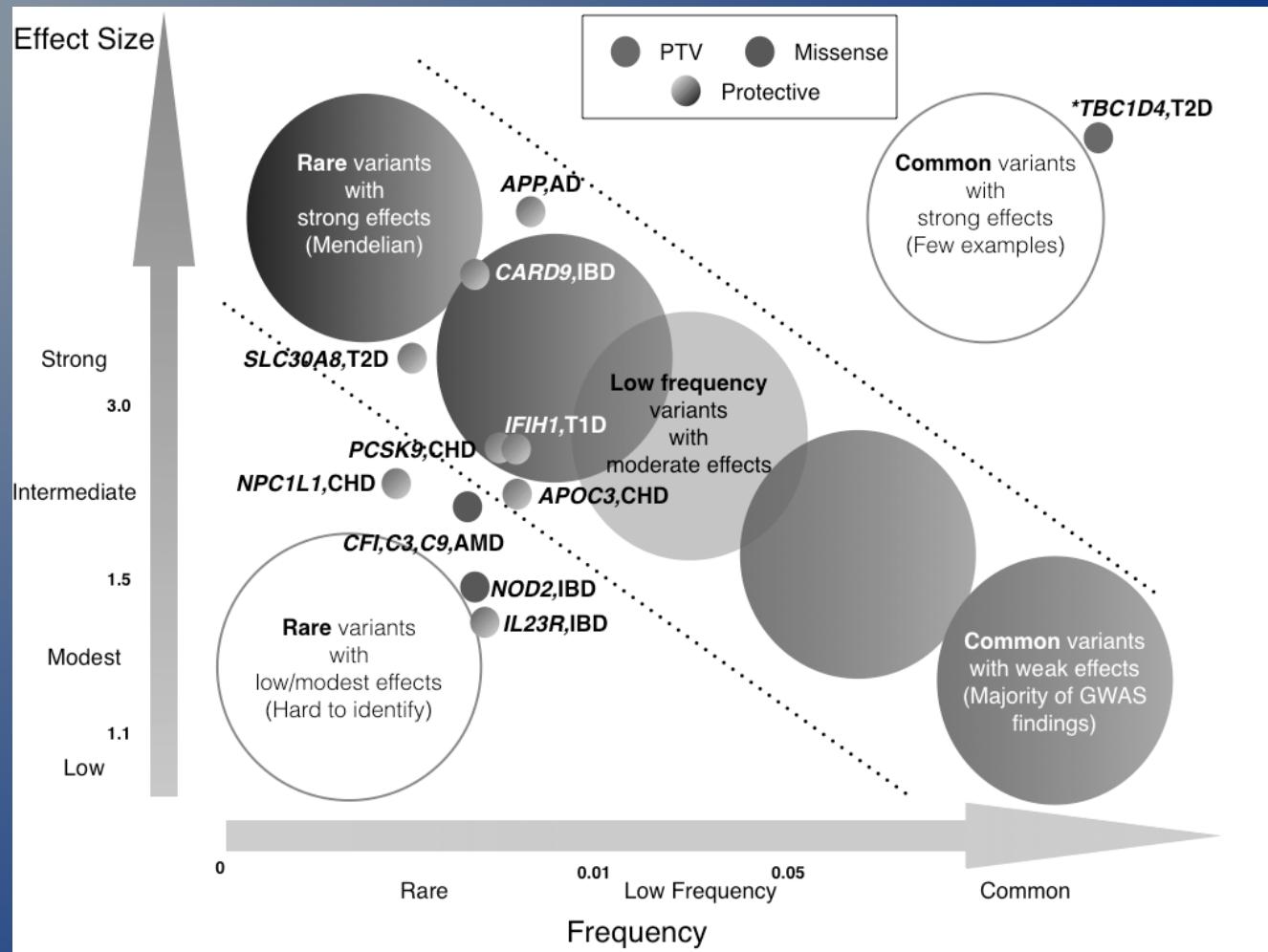
**c**, Conditional gene-based burden test (25 genes by 6 phenotypes,  $n = 32,806$  subjects): rare functional-variant allele counts are summed for each individual per gene and introduced in a logistic regression, including ImmunoChip covariates for multiple independent top (common) variant signals selected on the basis of a stepwise regression (down to  $P > 10^{-4}$ ). The psoriasis phenotype was not tested as most samples do not have ImmunoChip data. **d**, Count of case-unique rare alleles (UNIQ) tests (25 genes by 7 phenotypes,  $n = 41,911$  subjects): compares the number of rare functional variants only observed in cases with the distribution of this value upon random permutation (10,000 times) of the phenotypes. **e**, Count of control-unique rare alleles (UNIQ) tests: same as **d** but for rare functional variants uniquely observed in controls.

# High impact PTVs discovered so far



Timeline of PTV association

# Genetic variants: their frequency and effects



The expected relationship in the field of human genetics  
between allele frequency and effect size  
Transition from CVAS to RVAS

# Other data types of interest

RNA-seq – collection of expressed transcripts

Microbiome – quantification of organisms living in the gut, colon, etc.

Proteomics – collection of proteins

Metabolomics – collection of measurements related to metabolism

# Other data types of interest

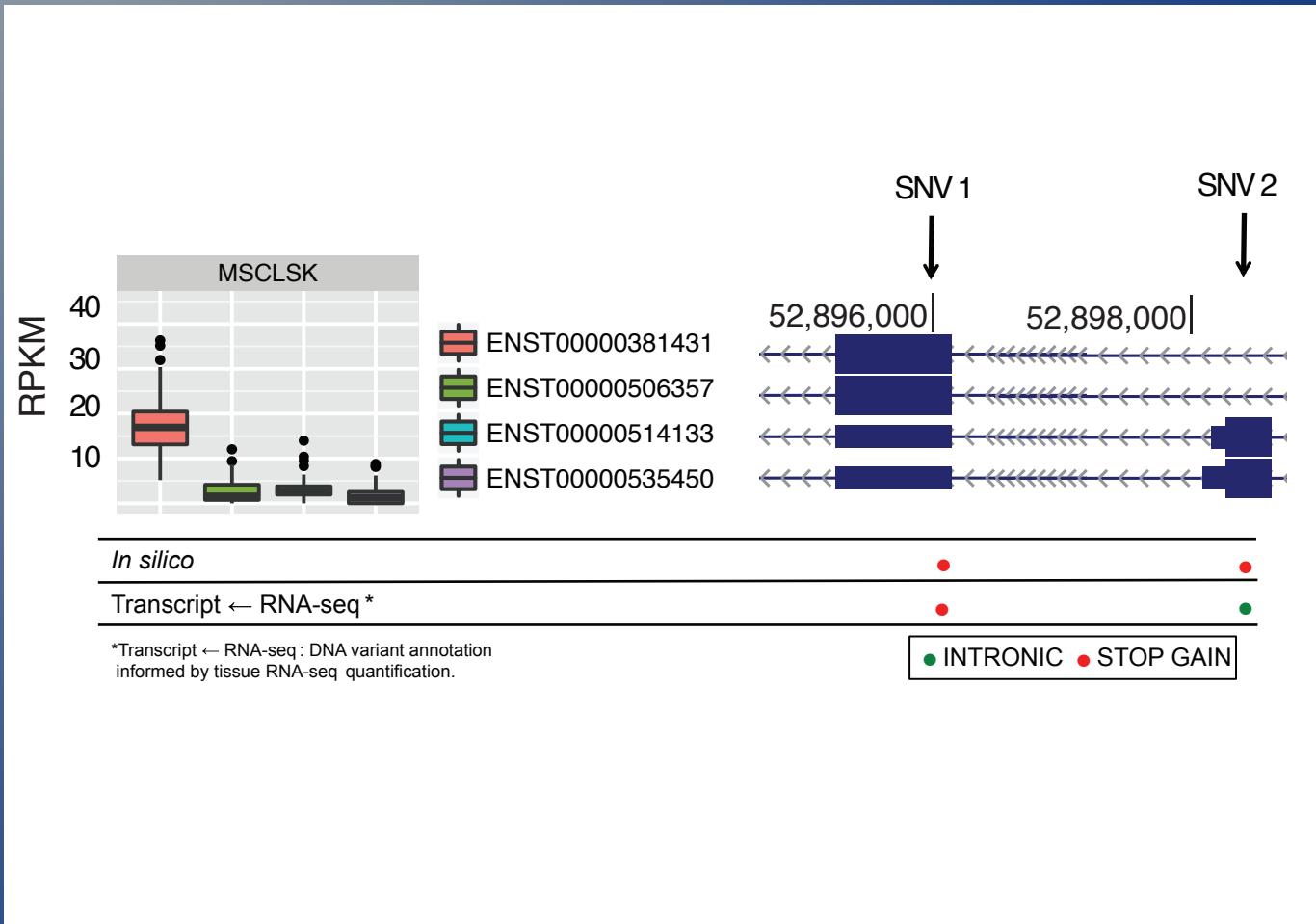
RNA-seq – collection of expressed transcripts

Microbiome – quantification of organisms living in the gut, colon, etc.

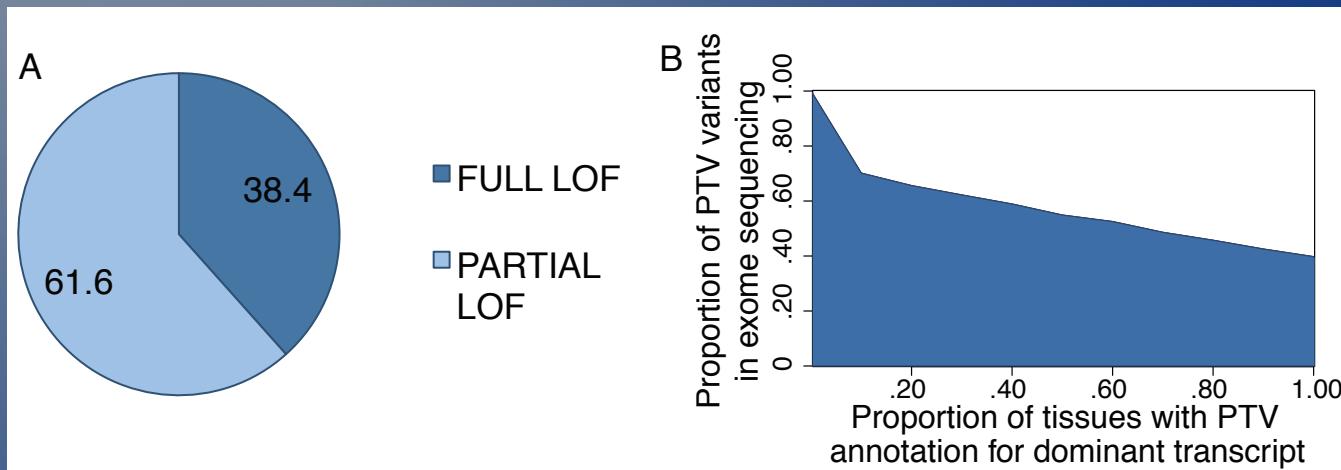
Proteomics – collection of proteins

Metabolomics – collection of measurements related to metabolism

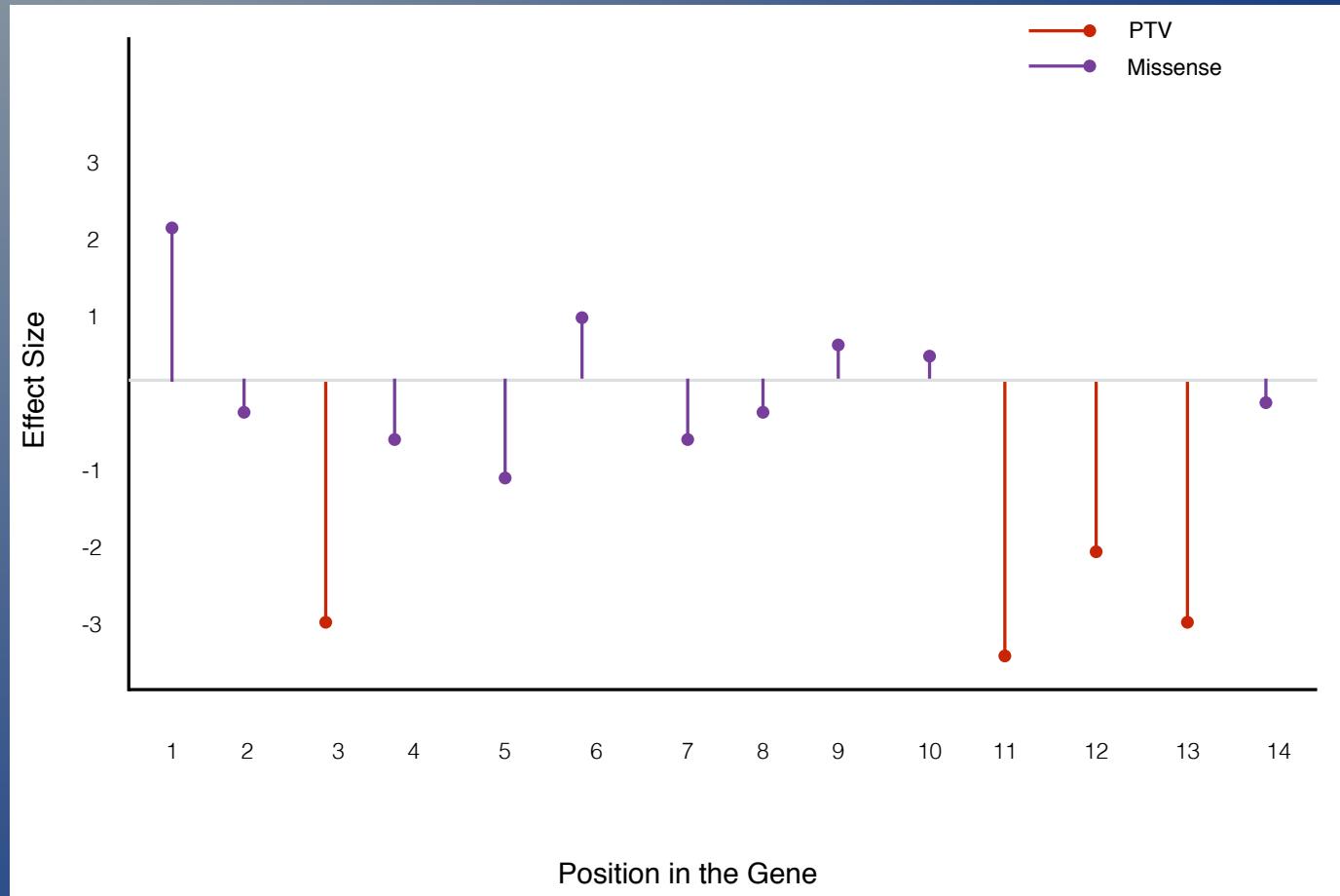
# RNA-seq data can be used to inform rare variant annotation



# RNA-seq data can be used to inform rare variant annotation



# Studying the direct functional consequences of genetic variants



CRISPR technology enables editing of any variant of interest in cell lines/tissues of interest

# Other data types of interest

RNA-seq – collection of expressed transcripts

Microbiome – quantification of organisms living in the gut, colon, etc.

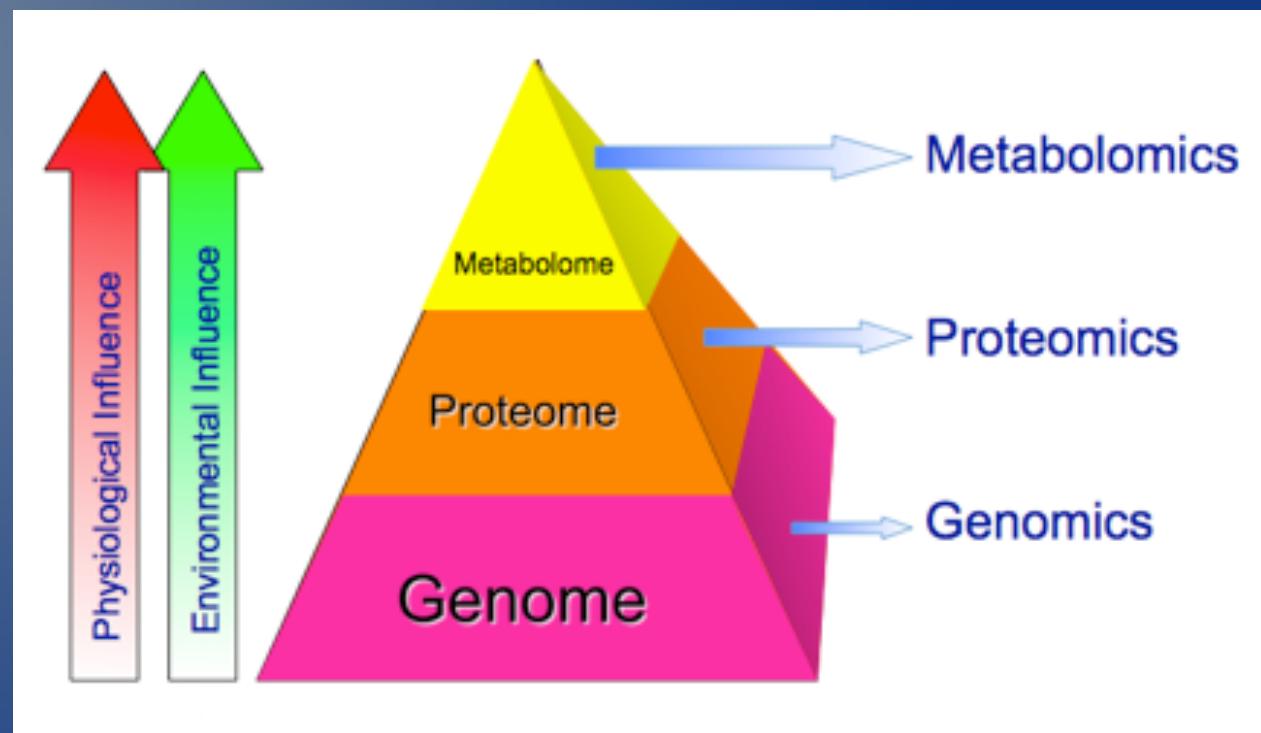
Proteomics – collection of proteins

Metabolomics – collection of measurements related to metabolism

# Other data types of interest

## METABOLOME

Collection of small-molecule chemicals found within a biological sample, e.g nucleic acids, sugars, lipoproteins



# Interested in exploring further?

## Download lecture materials from



<https://github.com/marivascruz/pydata2014genomics>

Presentation, iPython Notebook

Contact us: [matti.pirinen@helsinki.fi](mailto:matti.pirinen@helsinki.fi)  
[rivas@broadinstitute.org](mailto:rivas@broadinstitute.org)

# Interested in exploring further?

## MAMBA software

<http://www.well.ox.ac.uk/~rivas/mamba/>



tackling problems in medical genomics.

## PLINK/SEQ software

<https://atgu.mgh.harvard.edu/plinkseq/tutorial.shtml>

# PLINK/SEQ

A library for the analysis of genetic variation data