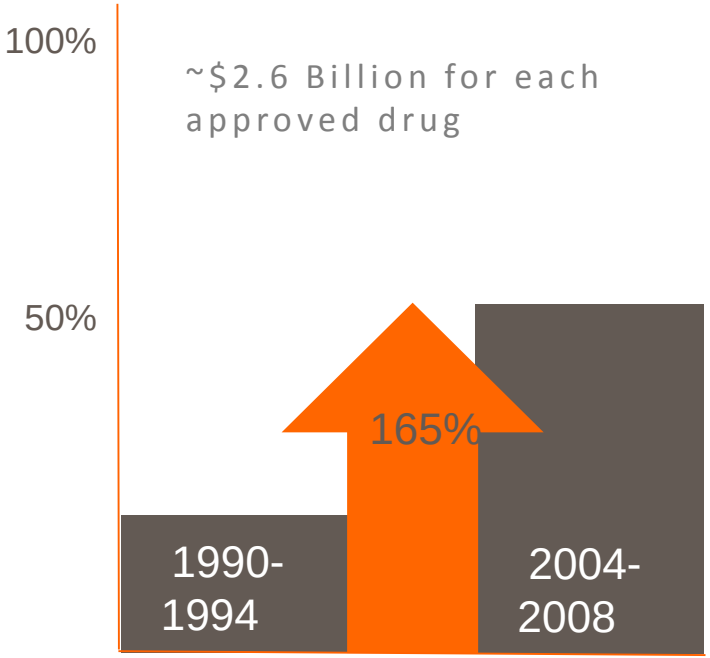
Using GWAS and Sequence Data to Identify and Validate Drug Targets

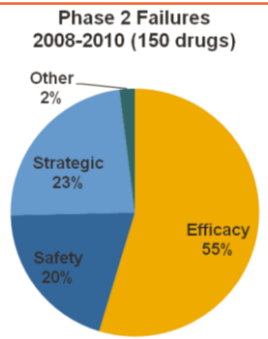
October 15, 2018

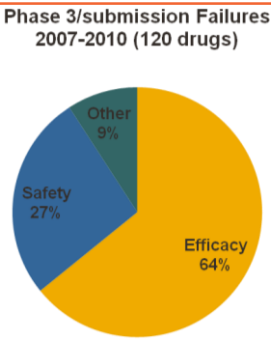
Matt Nelson
Head, Genetics

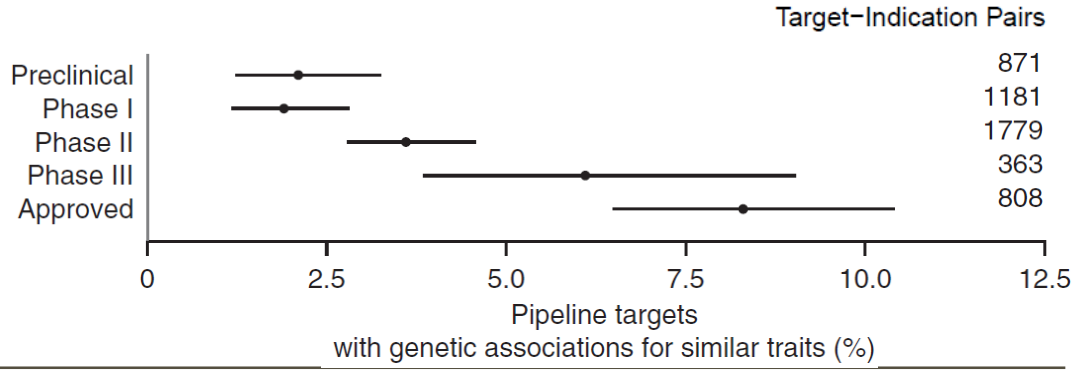
The drug discovery dilemma

Drugs with human genetic evidence >2x more likely to be successful







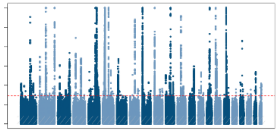


DiMasi JA *et al.* (2016) J Health Econ 47:20-33; Nelson *et al.* (2015) Nat Genet 47:846-60

From genetic association to target selection

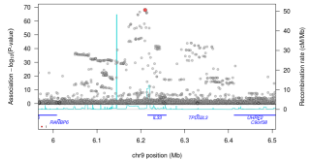
Find genetic associations

- 23andMe, UKB, FinnGen, EBI
- UKB analysis capabilities
- UKB exome sequence



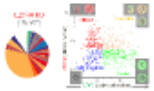
Variant-to-gene mapping

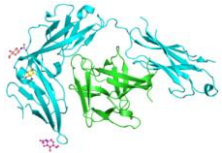
- Effector gene
- Cellular context
- Analytical capabilities

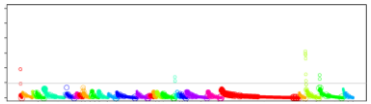


In vitro elucidation

- Gene editing
- High throughput assays

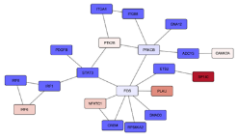






Understanding genetic impacts

- PheWAS
- MR applications
- Direction of effects



Pathway/network expansion

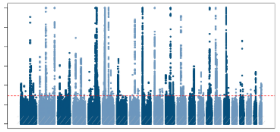
- Pathway analysis
- Network propagation

Synthesis & decisions

- MOA
- Probability of success
- Target filters
- Portfolio view

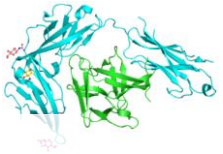
From genetic association to target selection

Find genetic associations
- 23andMe, UKB, FinnGen, EBI
- UKB analysis capabilities
- UKB exome sequence



Variant-to-gene mapping
- Effector gene
- Cellular context
- Analytical capabilities

In vitro elucidation
- Gene editing
- High throughput assays

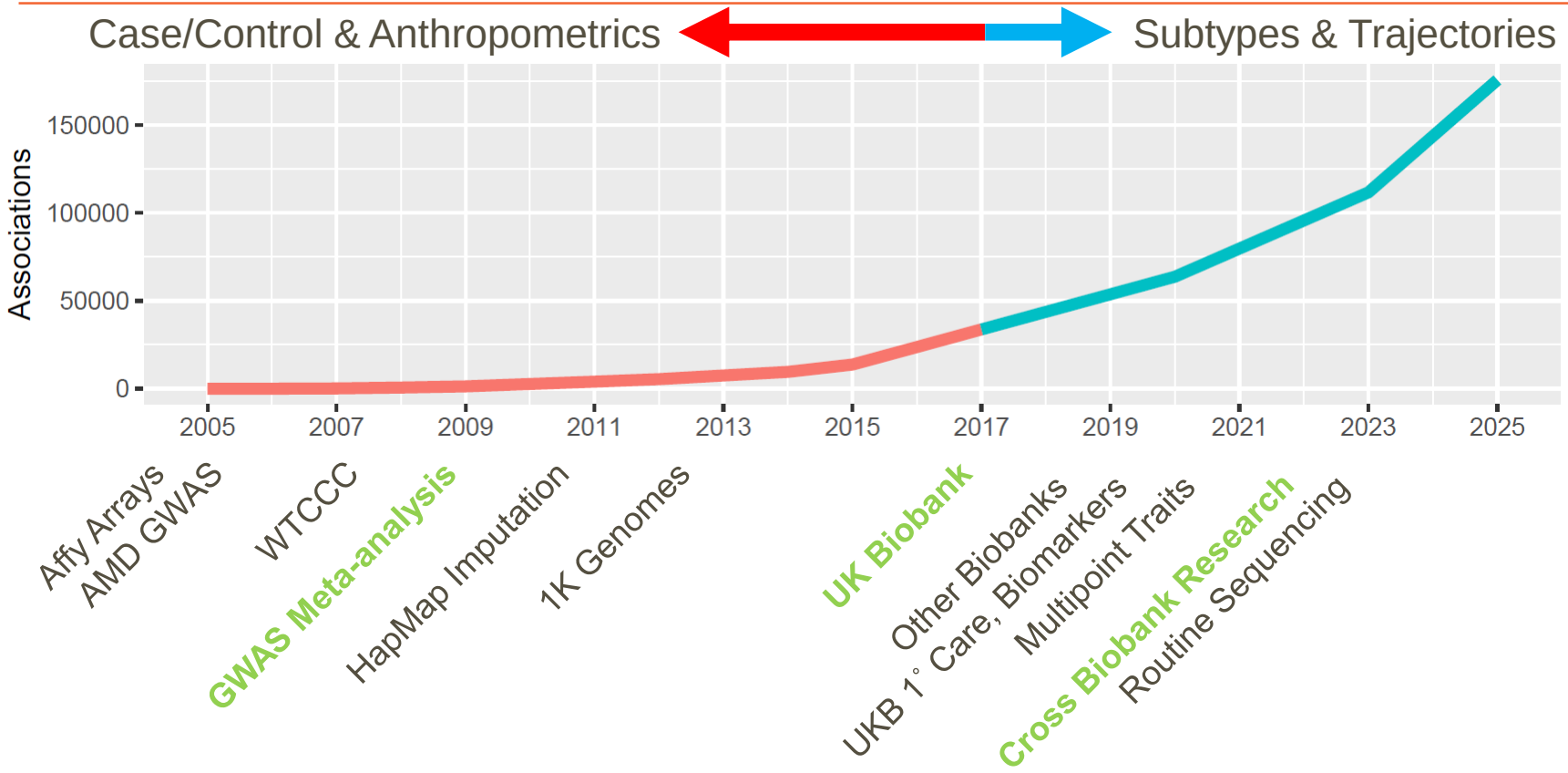


Understanding genetic impacts
- PheWAS
- MR applications
- Direction of effects

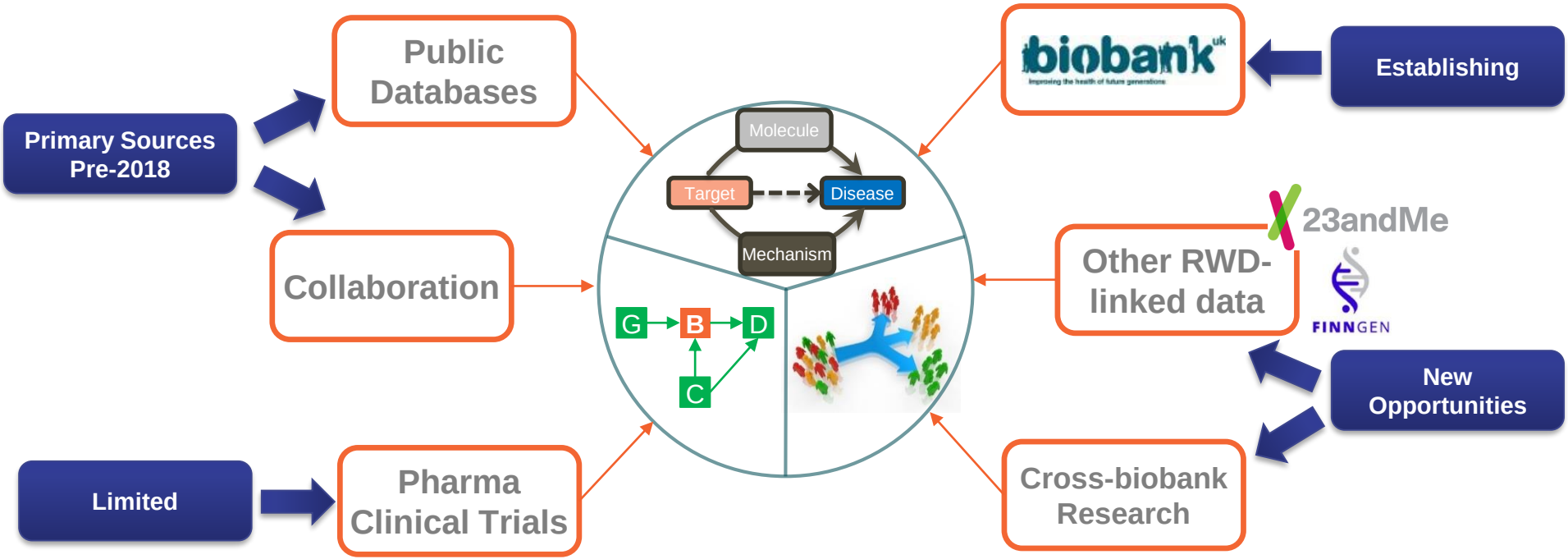
Pathway/network expansion
- Pathway analysis
- Network propagation

Synthesis & decisions
- MOA
- Probability of success
- Target filters
- Portfolio view

The discovery of genetic contributions to complex traits is just getting started



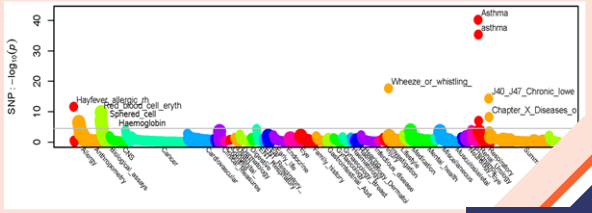
Sources of genetic evidence



UK Biobank is changing the way we do genetic research

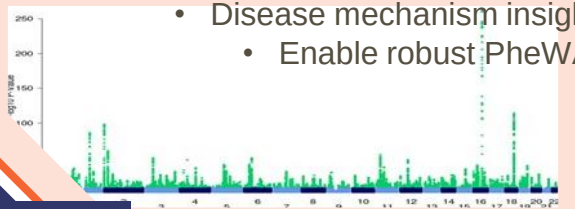
PheWAS of every target

- Integrate into early target validation



GWAS for Key Traits

- Support new target discovery
- Disease mechanism insights
- Enable robust PheWAS



Exome Sequencing

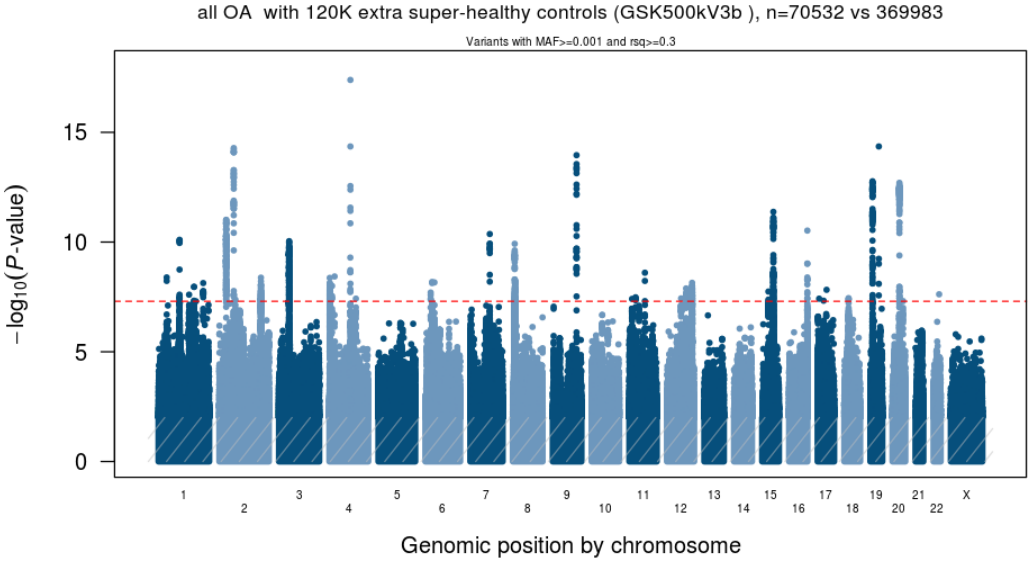
- Selected phenotype variant/gene discovery
- Characterize consequence of impactful variants
- **50K data landed early June: analyses ongoing for August sprint**

Diversify Phenotypes

Expand genetic investigation to include:

- All disease phenotypes in UKBB
- Disease sub-types
- Stratified models
- Disease severity
- “New” diseases

GWAS analysis of osteoarthritis in biobank uk



64 Loci

- $P < 3.1x10^{-8}$ across hip, knee, hip or knee, any OA
- 52 novel loci

216 Genes

- With V2G evidence: fine-mapping and annotation, eQTL colocalisation, rare disease, OA joint 'omics, animal model
- 4 genes are targets for OA drugs in development

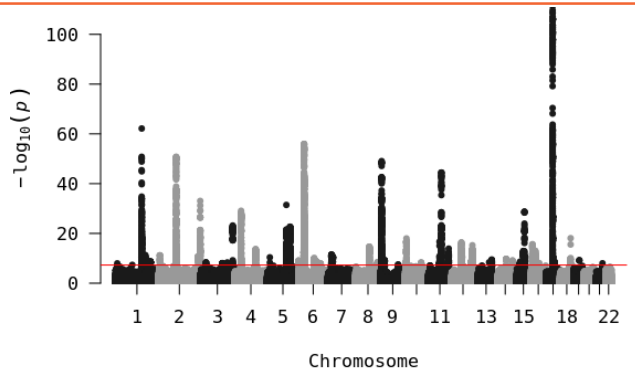
Target ID

- Tool molecule ID
- Triaging genes for ELT screen / tractability
- Test molecules in OA joint explant model

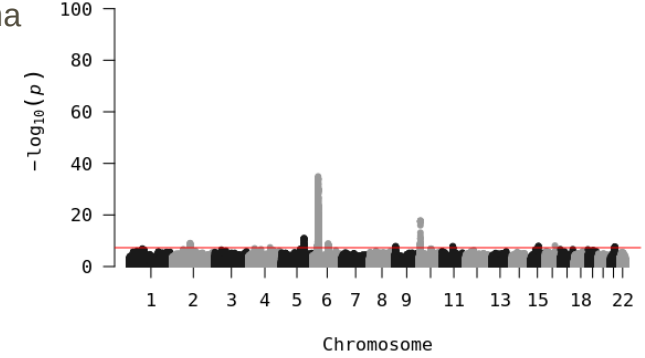
Linda McCarthy et al., Program #2307W; Poster session on Wednesday 2-3 PM
Ioanna Tachmazidou et al., Session #27, Wednesday 4:15pm-5:45pm

GWAS and pathway analyses in childhood and adult-onset asthma in biobank uk

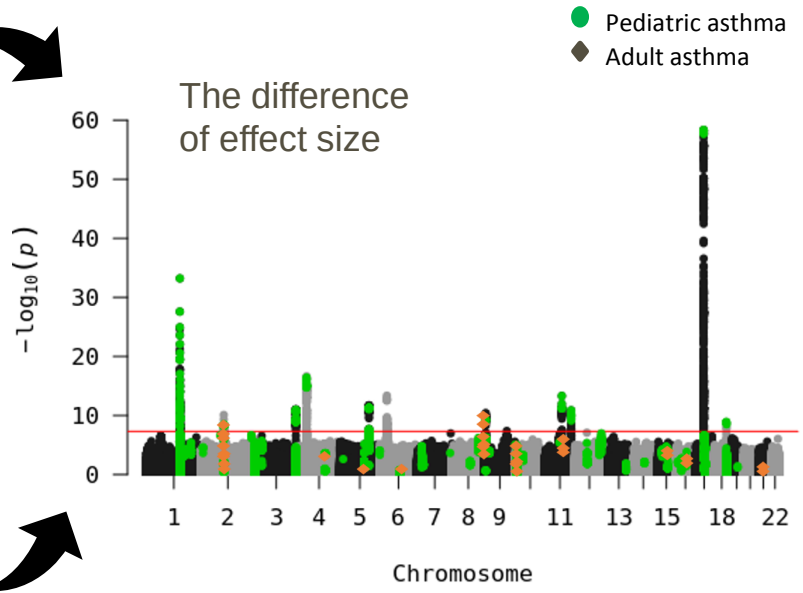
Childhood Asthma
≤12 years
10,920 cases
163,978 controls



Adult-onset Asthma
≥40 years
16,659 cases
163,978 controls



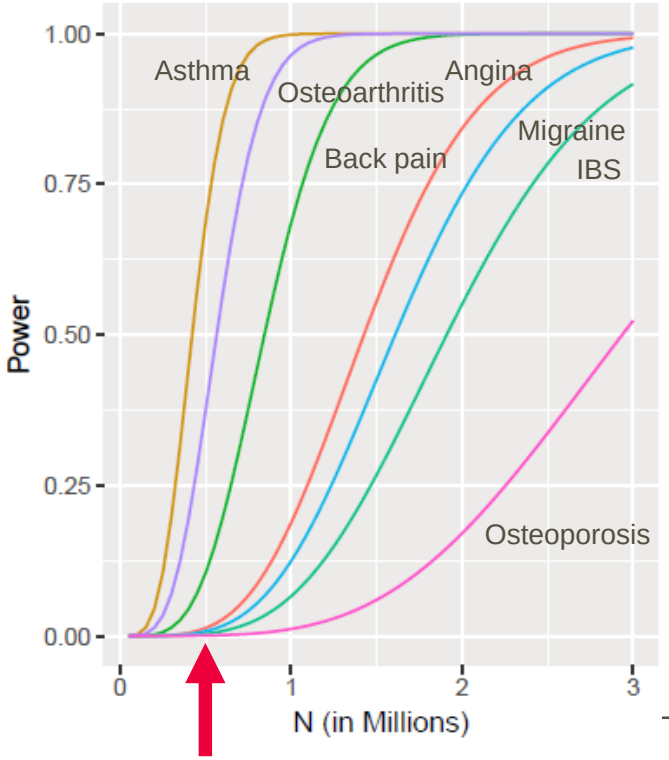
The difference of effect size



Kijoung Song et al., Session #56, Thursday at 11 AM

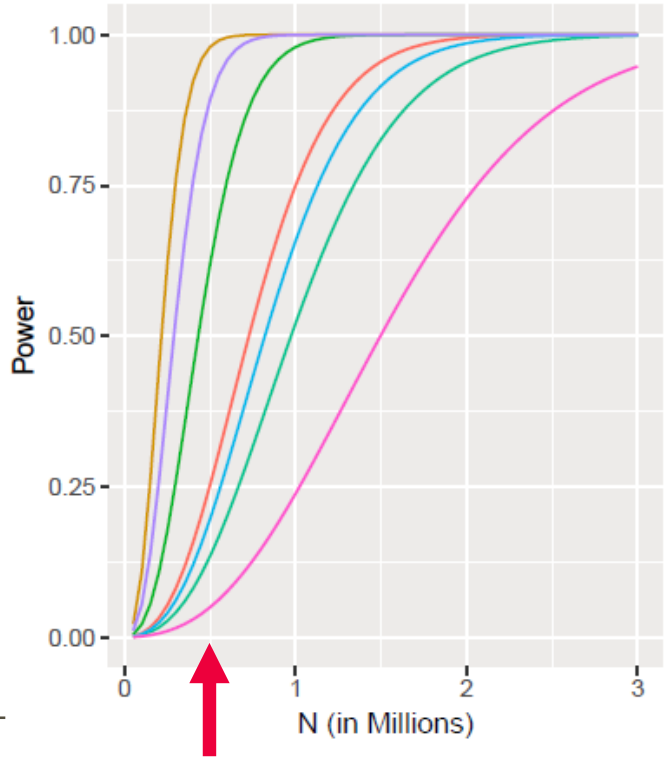
Broad discovery power in population samples requires very large numbers

GWAS Significance Threshold



UK Biobank

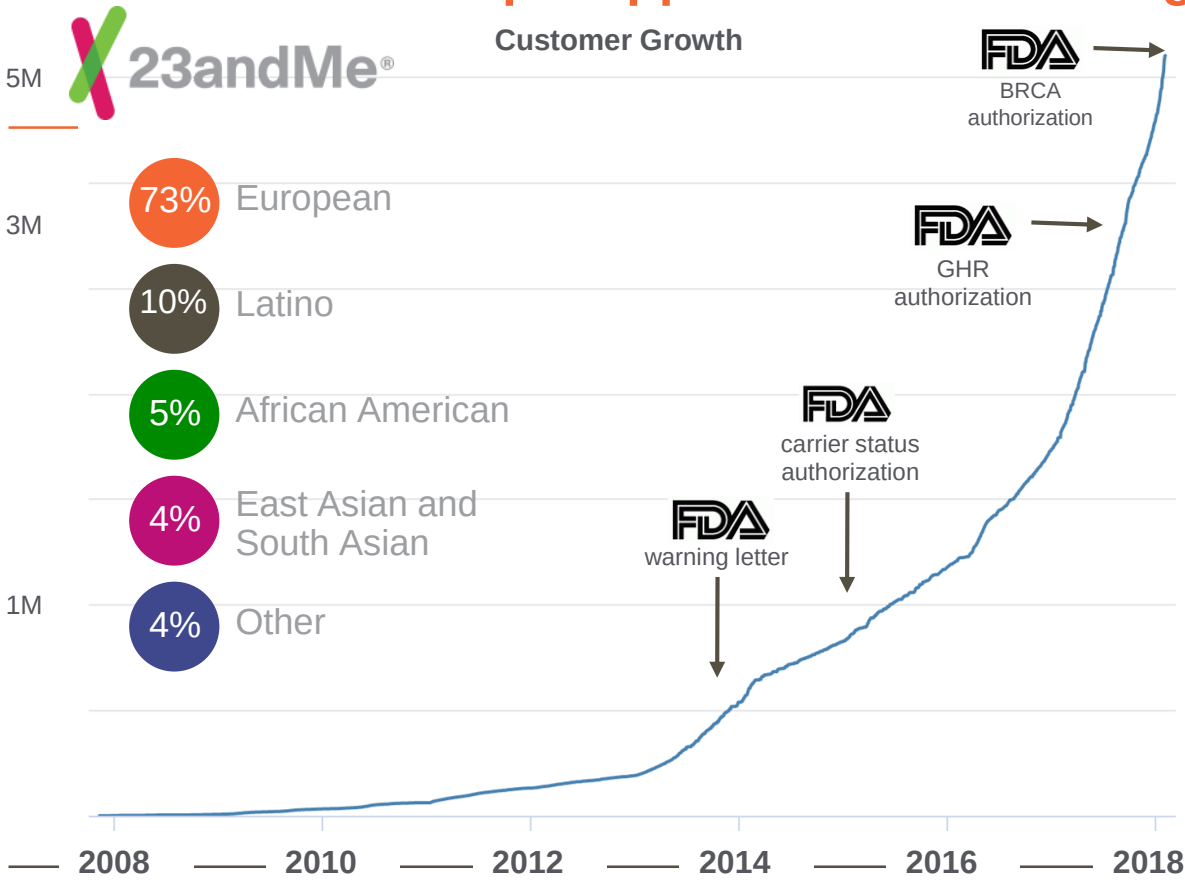
PheWAS Significance Threshold



UK Biobank

Assumptions
UKB case #s
MAF = 0.05, OR = 1.1
P = 5e-8 and 1e-4

23andMe offers unique opportunities to scale genetic research



5M+ customers

>80% consent to research & recontact

1.5B+ survey questions answered

To genotype or sequence…

VS

....A.....A....A...
....G.....C....A...

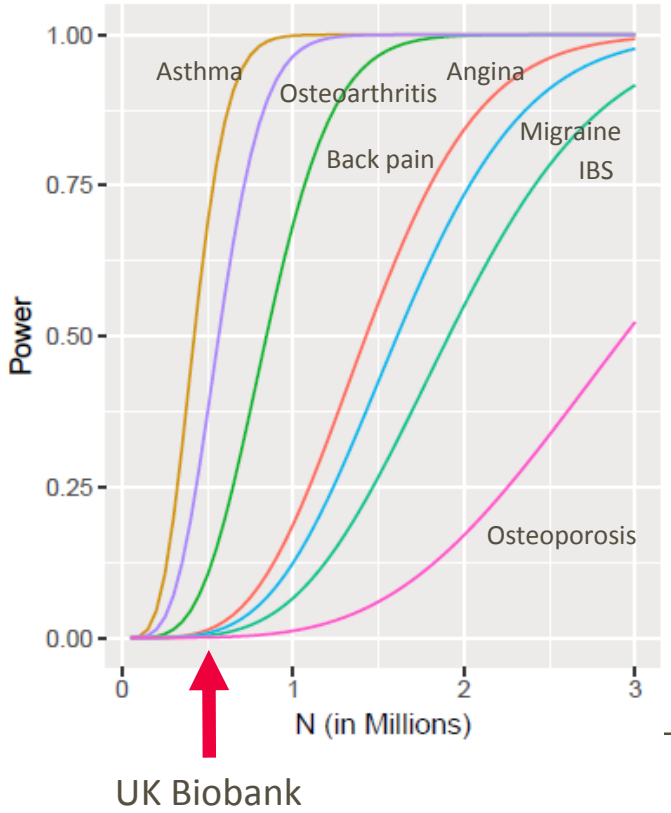
CGAGATCTCCCGACCTCATGG
CCAAGCTCTTTTCTTCTGTGC

Impact of cost on scale

Approach	Cost*	N = 100K	$10M (N)
Genotype array	$50	$5M	200K
Exome sequence	$300	$30M	33K
Genome sequence	$800	$80M	12.5K

*Informal, ballpark estimates for high volume (N >50K) initiatives

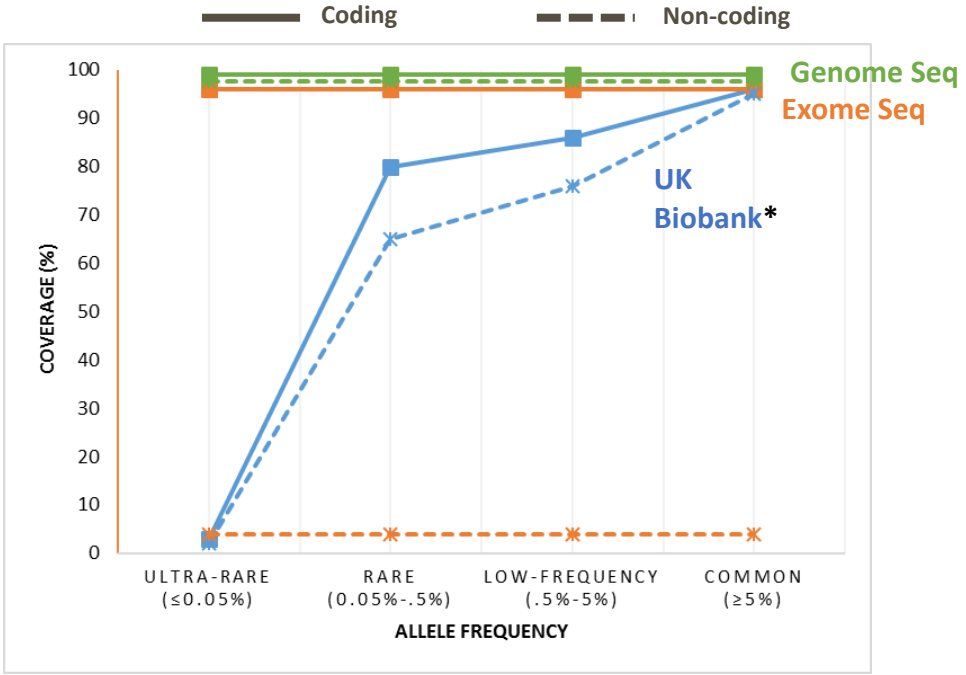
Power assumptions:
MAF = 0.05, OR = 1.1
P = 5e-8
Based on UKB case counts



Coverage trade-offs

Sequence (extrapolated from ExAC) versus UK Biobank genotypes

- Imputation from a large reference substantially improves array genotype value
  - Enrichment of low frequency coding variants improves coverage
- Whole genome sequence generally outperforms exome even within captured regions
- The contribution of rare non-coding variation to complex traits is largely unexplored
  - Need to learn how to identify rLoF variants (regulatory loss of function) to fully exploit WGS



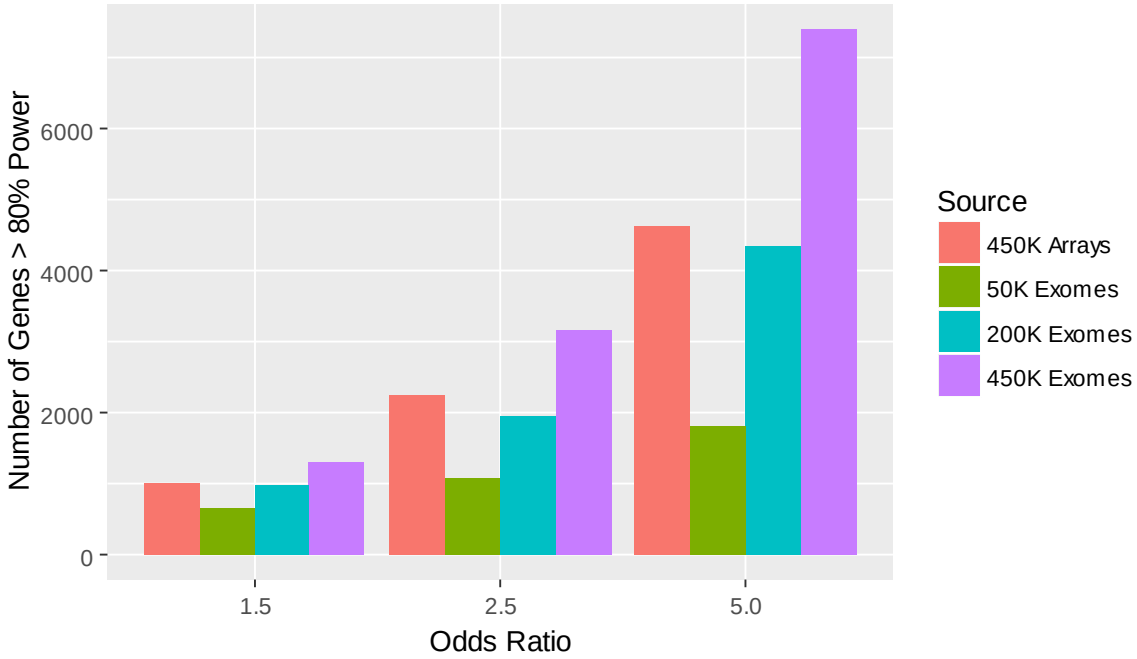
*Coverage reported as imputation quality (r^2) with 1KG/UK10K imputation

Capture of pLoF variants

Sequence (extrapolated from ExAC) versus UK Biobank genotypes

For testing effect of pLoF variants, sequencing catches up with arrays at around a 1:2 ratio

Capture of coding *indels* directly or via imputation in UKB is ~40% less efficient than SNPs of the same frequency



Sequencing in consanguineous populations identifies disproportionately more knock-outs

LETTER

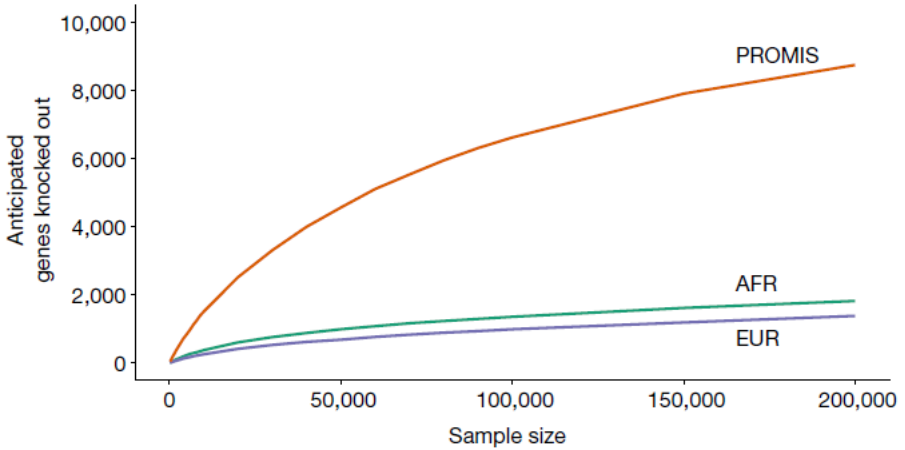
doi:10.1038/nature22034

Human knockouts and phenotypic analysis in a cohort with a high rate of consanguinity

Danish Saleheen[1,2]*, Pradeep Natarajan[3,4]*, Irina M. Armean[4,5], Wei Zhao[1], Asif Rasheed[2], Sumeet A. Khetarpal[6], Hong-Konrad J. Karczewski[4,5], Anne H. O'Donnell-Luria[4,5,8], Kaitlin E. Samocha[4,5], Benjamin Weisburd[4,5], Namrata Gupta[4], Mozzam Zaidi[2], Maria Samuel[2], Atif Imran[2], Shahid Abbas[9], Faisal Majeed[2], Madiha Ishaq[2], Saba Akhtar[2], Kevin Trindel Megan Mucksavage[6], Nadeem Qamar[10], Khan Shah Zaman[10], Zia Yaqoob[10], Tahir Saghir[10], Syed Nadeem Hasan Rizvi[1] Anis Memon[10], Nadeem Hayyat Mallick[11], Mohammad Ishaq[12], Syed Zahed Rasheed[12], Fazal-ur-Rehman Memon[13], Khalid Mahmood[14], Naveeduddin Ahmed[15], Ron Do[16,17], Ronald M. Krauss[18], Daniel G. MacArthur[4,5], Stacey Gabriel[4], Eric S. Lander[4], Mark J. Daly[4,5], Philippe Frossard[2]§, John Danesh[19,20]§, Daniel J. Rader[6,21]§ & Sekar Kathiresan[3,4]§

Offspring of consanguineous unions are more likely to be homozygous for loss-of-function (LoF) mutations than non-consanguineous populations (i.e. human knock-outs)

Estimated number of genes with knock-outs for PROMIS versus LRM populations

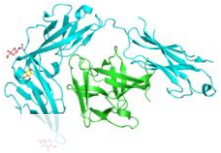


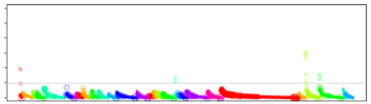
From genetic association to target selection

Find genetic associations
- 23andMe, UKB, FinnGen, EBI
- UKB analysis capabilities
- UKB exome sequence

Variant-to-gene mapping
- Effector gene
- Cellular context
- Analytical capabilities

In vitro elucidation
- Gene editing
- High throughput assays





Understanding genetic impacts
- PheWAS
- MR applications
- Direction of effects

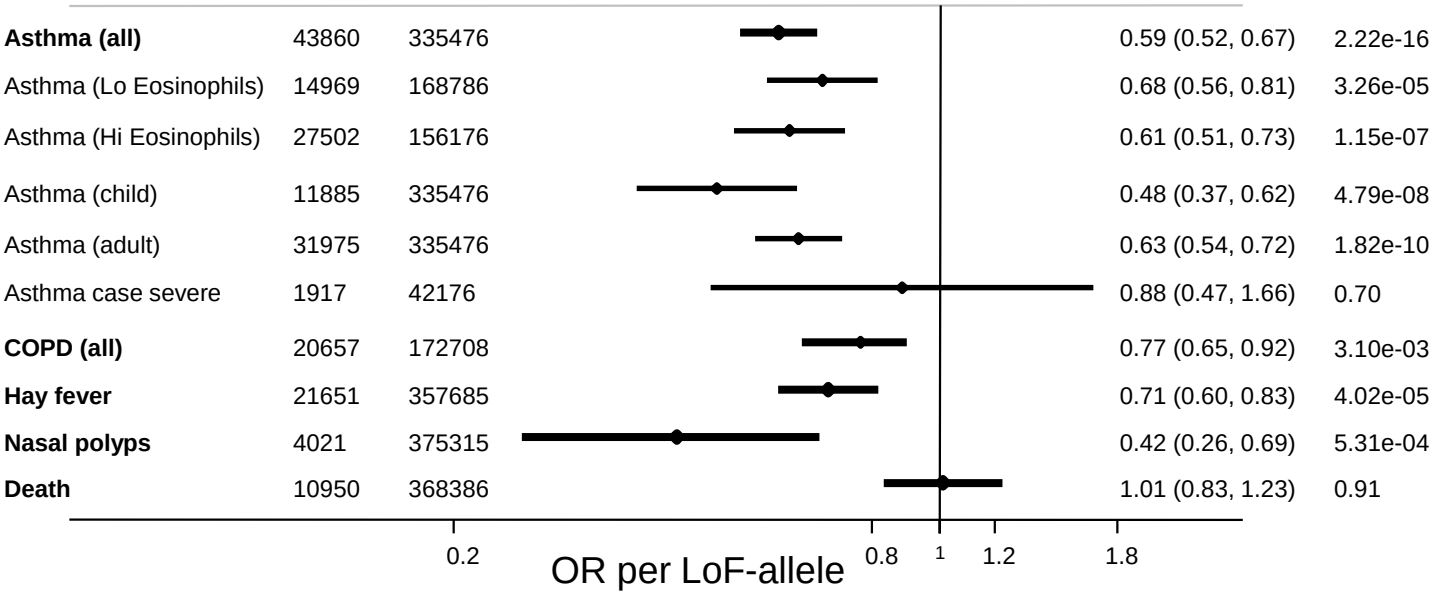
Pathway/network expansion
- Pathway analysis
- Network propagation

Synthesis & decisions
- MOA
- Probability of success
- Target filters
- Portfolio view

Access to individual level data allow detailed follow-up analyses

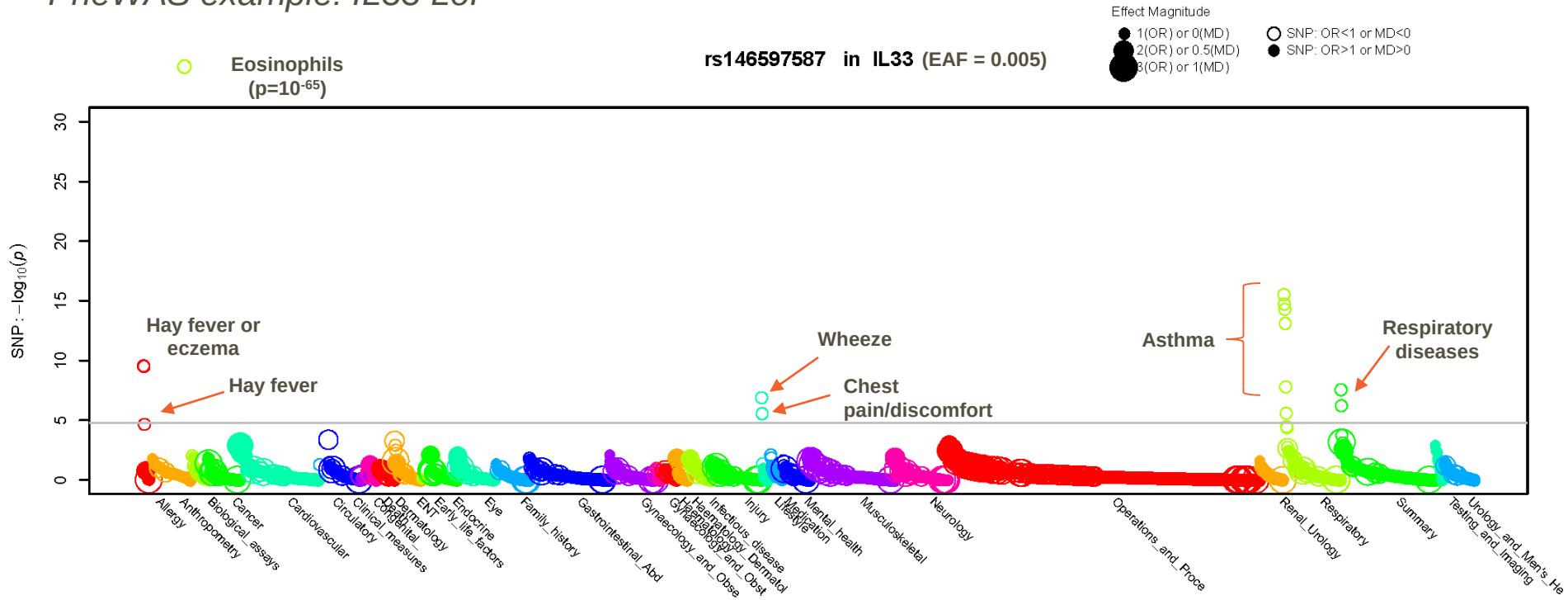
IL33 LoF disease association

Outcome	Case	Control	OR (95% CI)	P-value
Asthma (all)	43860	335476	0.59 (0.52, 0.67)	2.22e-16
Asthma (Lo Eosinophils)	14969	168786	0.68 (0.56, 0.81)	3.26e-05
Asthma (Hi Eosinophils)	27502	156176	0.61 (0.51, 0.73)	1.15e-07
Asthma (child)	11885	335476	0.48 (0.37, 0.62)	4.79e-08
Asthma (adult)	31975	335476	0.63 (0.54, 0.72)	1.82e-10
Asthma case severe	1917	42176	0.88 (0.47, 1.66)	0.70
COPD (all)	20657	172708	0.77 (0.65, 0.92)	3.10e-03
Hay fever	21651	357685	0.71 (0.60, 0.83)	4.02e-05
Nasal polyps	4021	375315	0.42 (0.26, 0.69)	5.31e-04
Death	10950	368386	1.01 (0.83, 1.23)	0.91



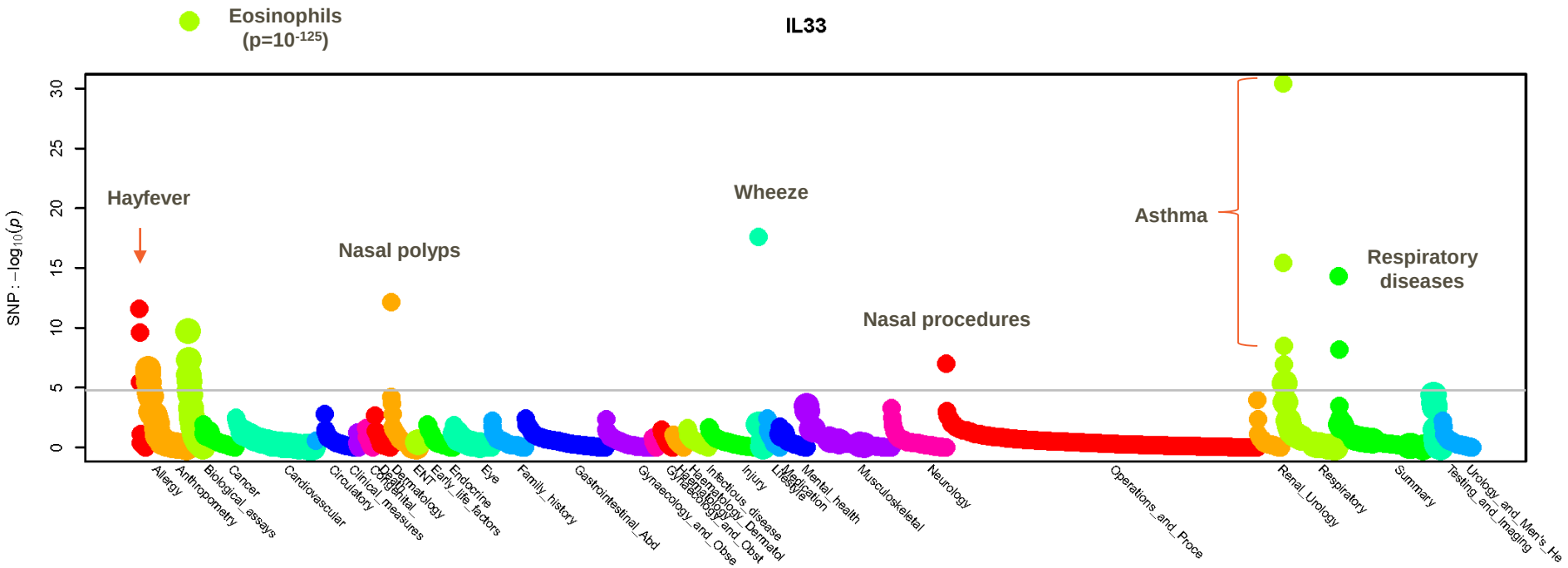
SNVs as proxies for target perturbation

PheWAS example: IL33 LoF



When we don't have impactful SNVs, predicted expression as an approach to instrument genes

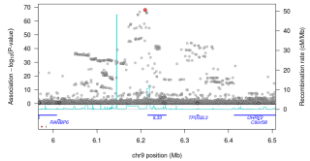
We can "instrument" ~17000 genes in 1 tissue, ~7000 in > 10 tissues



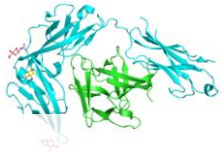
From genetic association to target selection

Find genetic associations
- 23andMe, UKB, FinnGen, EBI
- UKB analysis capabilities
- UKB exome sequence

Variant-to-gene mapping
- Effector gene
- Direction of effect
- Cellular context
- Analytical capabilities



In vitro elucidation
- Gene editing
- HT assays



Understanding genetic impacts
- PheWAS
- Direction of effects
- MR applications

Pathway/network expansion
- Pathway analysis
- Network propagation

Synthesis & decisions
- MOA
- Probability of success
- Target filters
- Portfolio view

Can we use the growing set of functional genomic data types to inform V2G mapping?

Functional annotation of variants

Protein-coding
- LoF/GoF mutations
- SIFT/PolyPhen/PROVEAN/etc…

Open chromatin
- DNaseI hypersensitivity (DHS)
- ATAC-seq

Regulatory function
- Chromatin accessibility (DHS)
- eRNA expression (FANTOM5)
- Histone modifications (ChIP-seq)
- DNA methylation
- Chromatin state prediction (ChromHMM/IDEAS)

Variant-to-gene mapping

Protein-coding

Genetics-based methods
- eQTL/pQTL

Correlation-based methods
- Correlated epigenetics (i.e. DHSs)
- FANTOM5 enhancers-to-genes

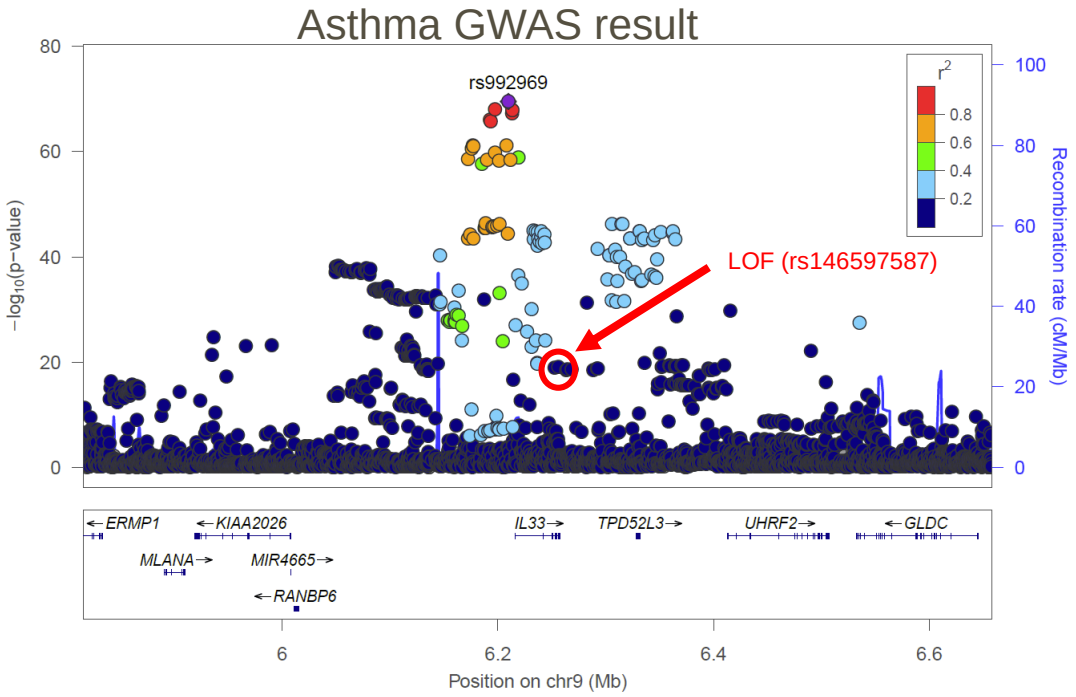
DNA-looping
- Chromosome conformation (i.e. *C or HiC)
- Promoter-capture HiC (enrich for promoters)
- HiChIP/ChIA-PET (enrich for functional marks)

Opportunity: A large amount of this data is being generated

Challenges: Different methods tell you different things about transcriptional regulation
Different methods have different levels of noise and sensitivity/specificity
Considerable biases in the available tissues/cell-types (often an immune bias)

For a given locus, the weight of evidence for each gene is the combination of these data

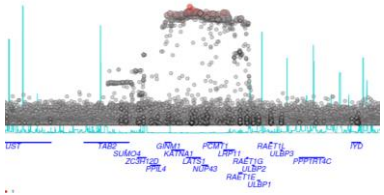
Teasing apart associations and identifying effector genes



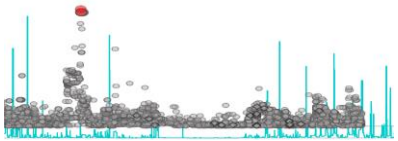
- Conditional analyses implicate 4 putatively causal SNVs in the region
  - Including the LoF
- Individual level data allow us to perform conditional analyses and fine-mapping even for rare variants

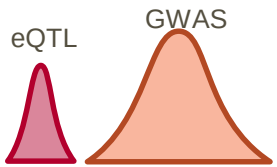
Identify disease causal genes through inference: colocalization testing

UKB Grip Strength GWAS



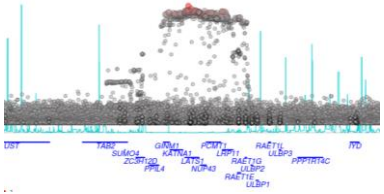
Gene 1 eQTL



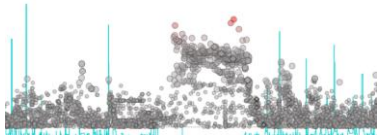


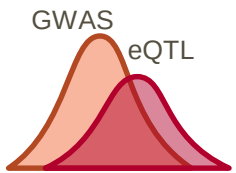
colocalization: 0%

UKB Grip Strength GWAS



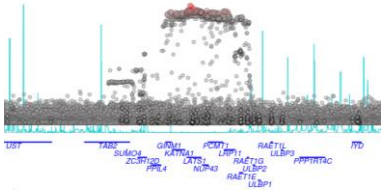
Gene 2 eQTL



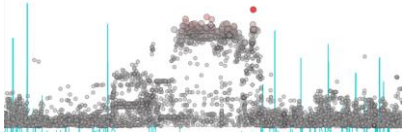


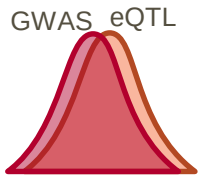
colocalization: 60%

UKB Grip Strength GWAS



Gene 3 eQTL



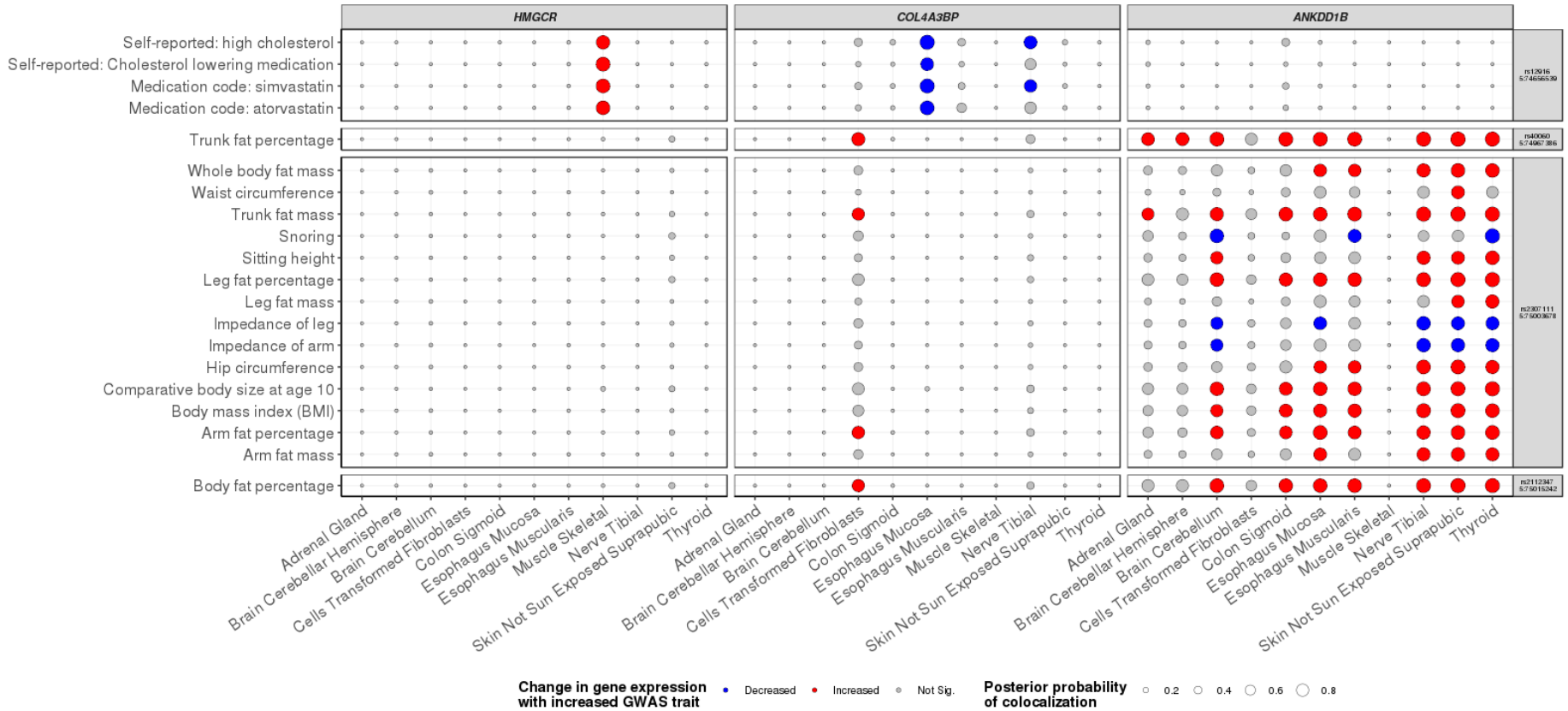


colocalization: 95%

Giambartolomei *et al.,* Bayesian Test for Colocalisation between Pairs of Genetic Association Studies Using Summary Statistics. *PLoS Genetics*, 2014

Dawn Waterworth, Session #32, Wednesday at 9:45 AM

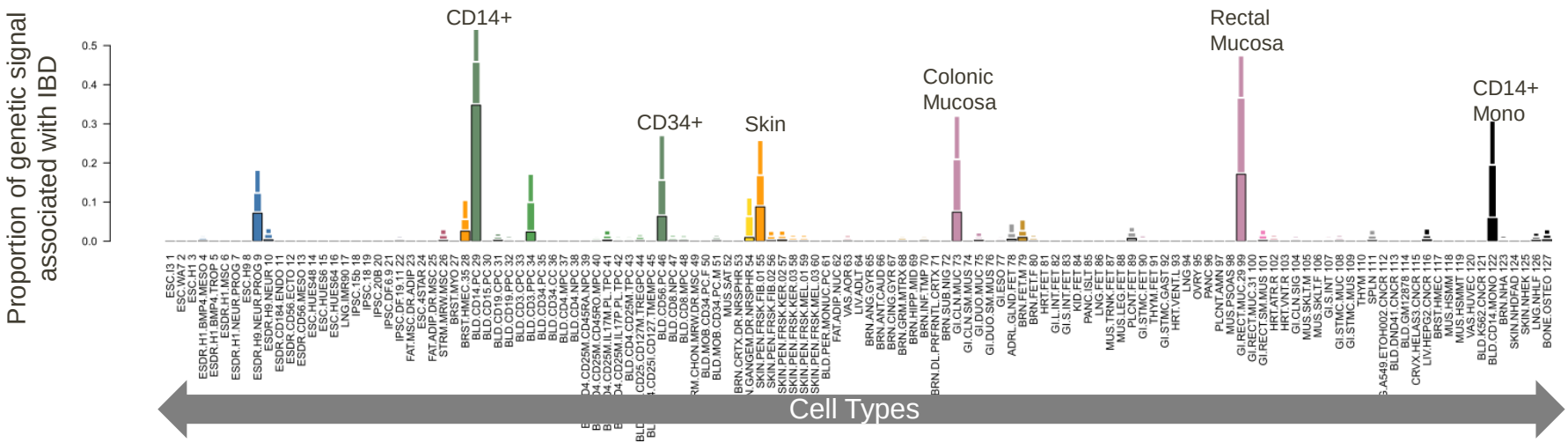
Comprehensive colocalization can differentiate causal genes



Karl Guo, Session #56, Thursday at 11 AM
Karsten Sieber, Poster #3314, Friday at 3 PM

Identifying disease relevant tissues by combining GWAS with chromatin states across cell types: An IBD example

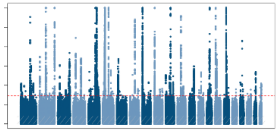
For IBD ~48% of signals are best explained by immune/blood cells, while ~40% are best explained by epithelial cells



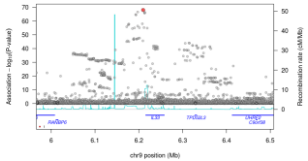
Collaboration with Yu Zhang, Penn State

From genetic association to target selection

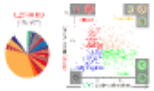
Find genetic associations
- 23andMe, UKB, FinnGen, EBI
- UKB analysis capabilities
- UKB exome sequence

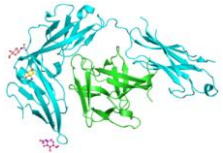


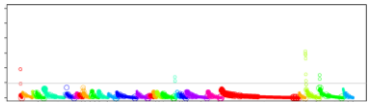
Variant-to-gene mapping
- Effector gene
- Cellular context
- Analytical capabilities



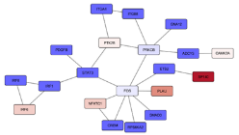
In vitro elucidation
- Gene editing
- High throughput assays







Understanding genetic impacts
- PheWAS
- MR applications
- Direction of effects



Pathway/network expansion
- Pathway analysis
- Network propagation

Synthesis & decisions
- MOA
- Probability of success
- Target filters
- Portfolio view

Special thanks

Kijoung Song

Josh Hoffman

Ioanna Tachmazidou

Ashutosh Pandey

Giovanni Dall'Olio

Mathias Chiano

Karl Guo

John Whittaker

Toby Johnson

Meg Ehm

Dawn Waterworth

Laura Yerges-Armstrong

Robert Scott

Linda McCarthy