Value of genomic information and how to gather it

International Cohorts Summit
March 26, 2018

Matt Nelson
Head, Genetics

Overview

The genetics value proposition: a brief view from pharma

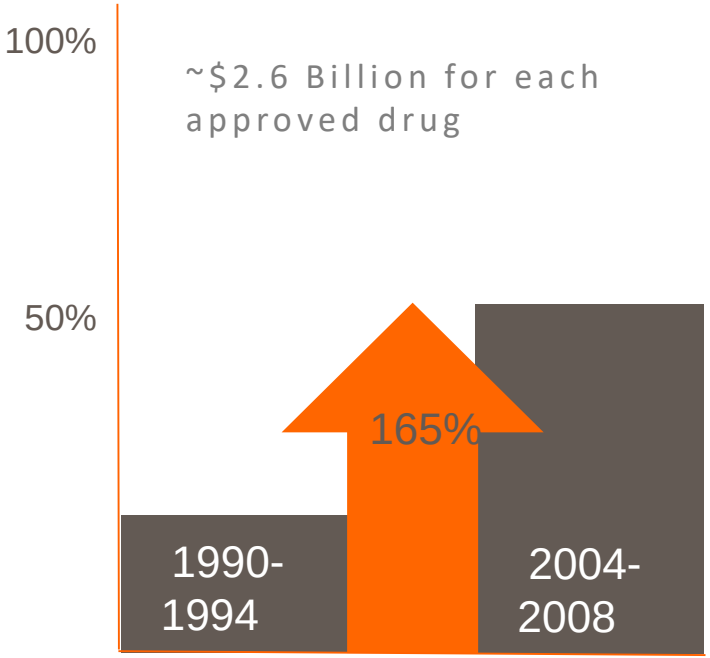
Genotyping versus sequencing

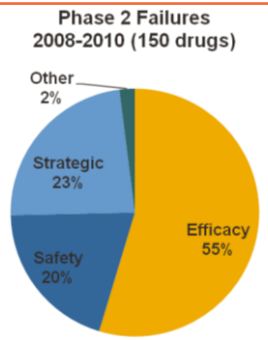
The importance of standards and careful QC

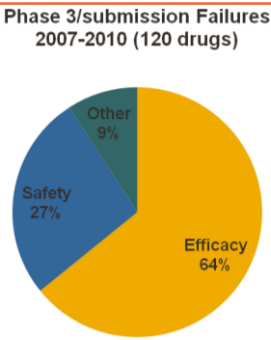
Changes in scientific culture drive the biggest gains

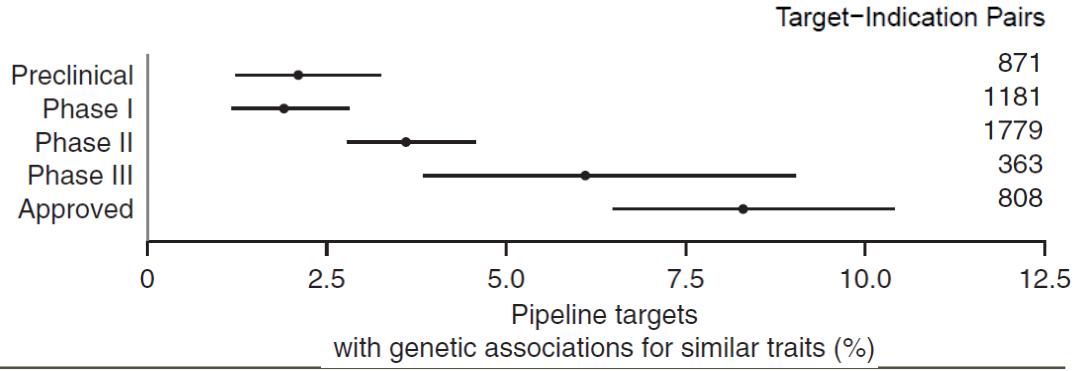
The drug discovery dilemma

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



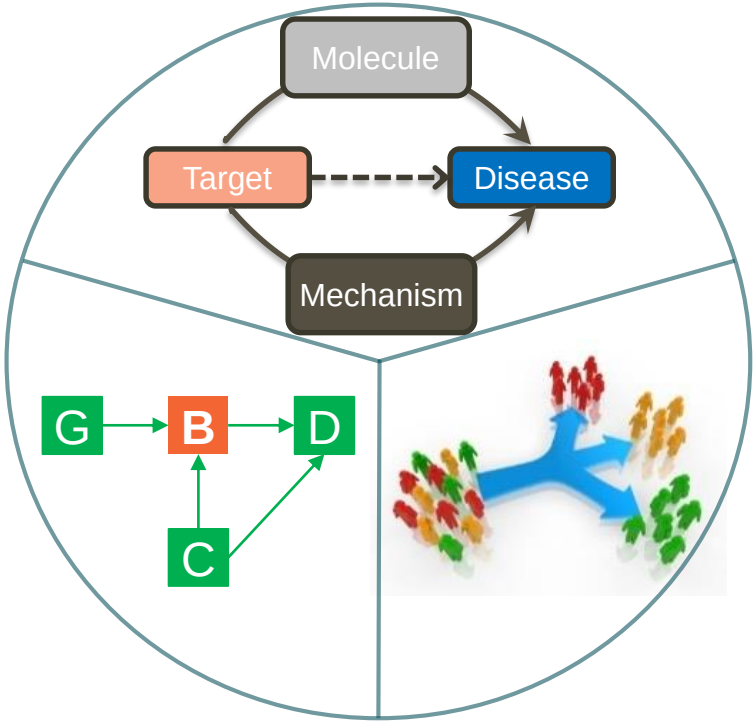




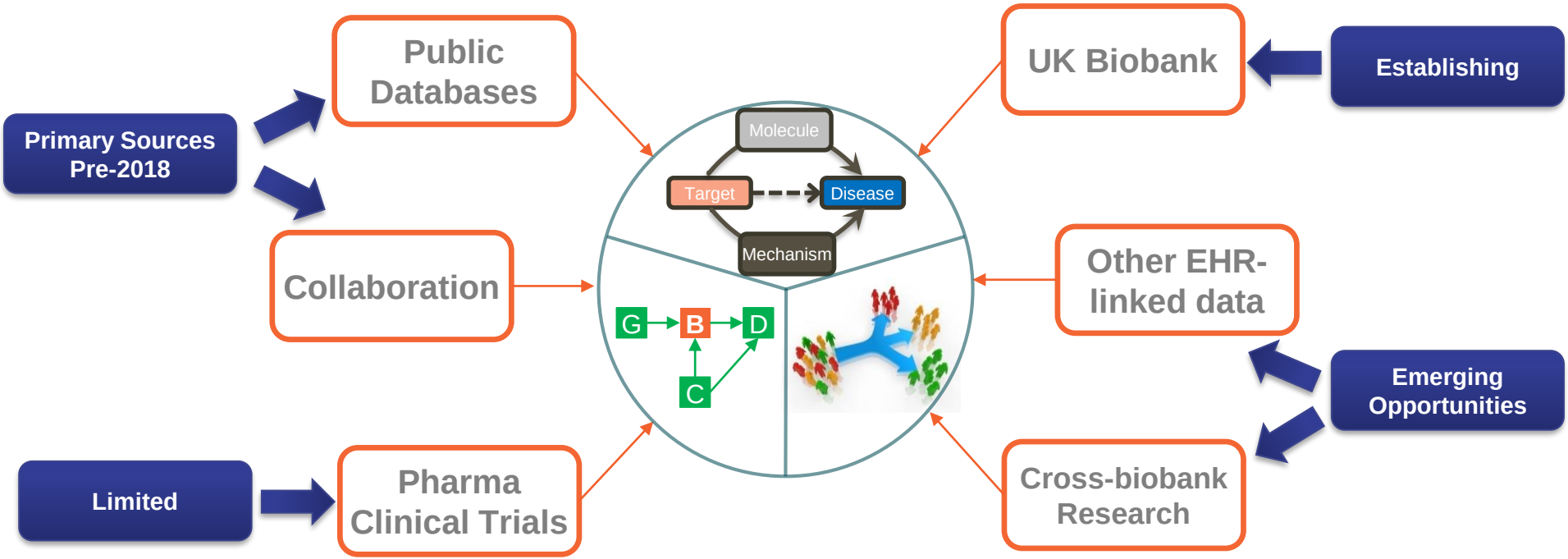


DiMasi JA *et al.* (2016) J Health Econ 47:20-33; Nelson *et al.* (2015) Nat Genet 47:846-60

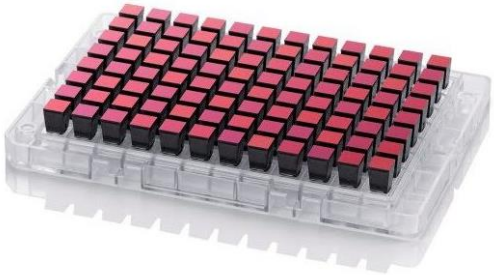
Applying genetics to drug discovery and development

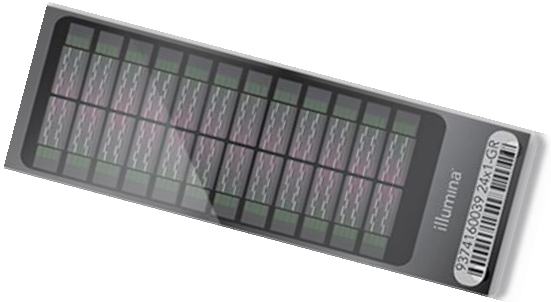


Applying genetics to drug discovery and development



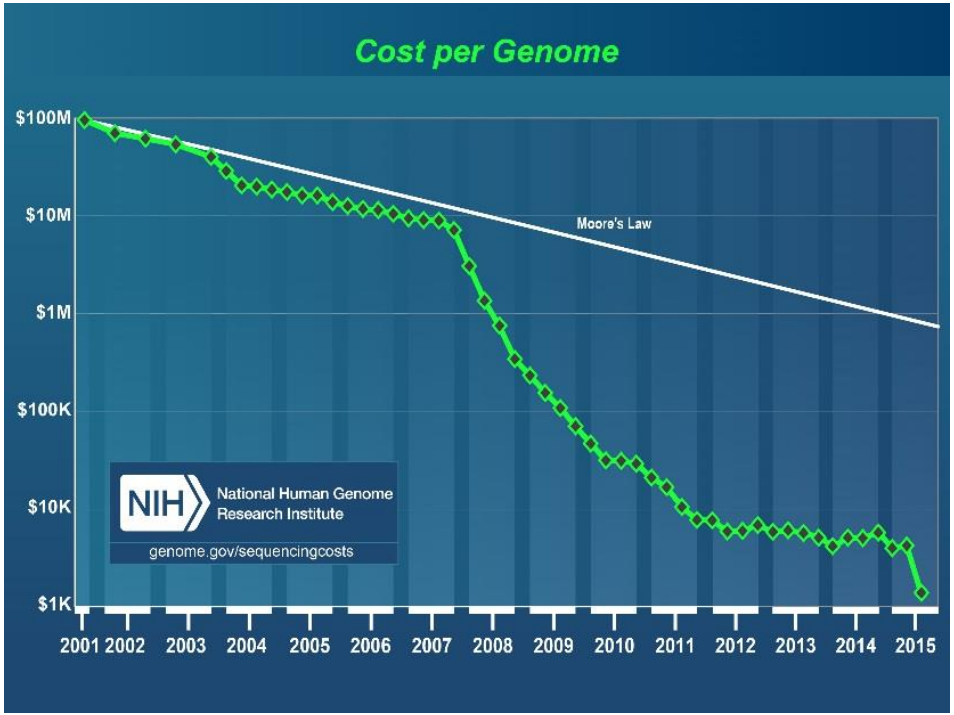
To genotype or sequence





VS

Why is this even a debate? COST



Approach	Cost*
Genotype array	$50
Exome sequence	$300
Genome sequence	$600

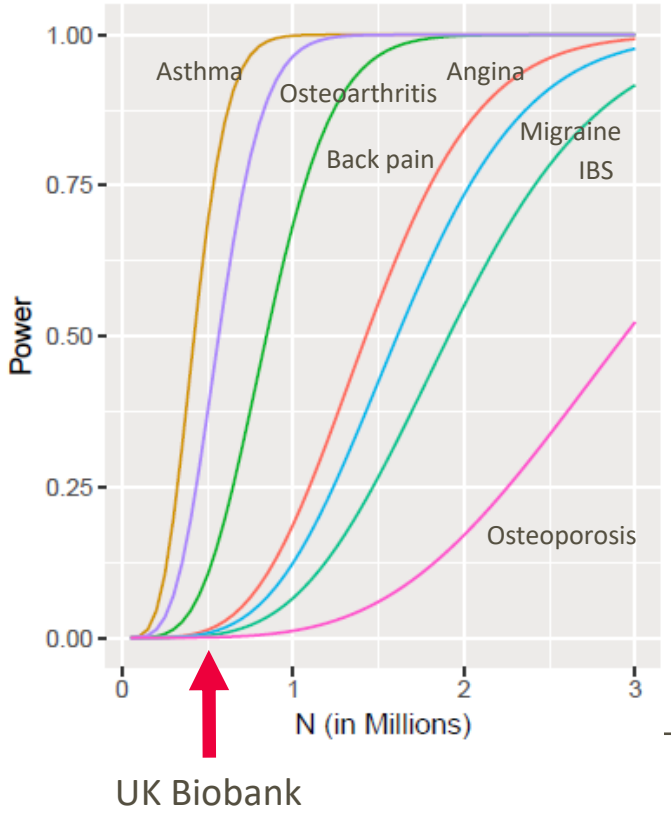
*Ballpark estimates for high throughput (N >50K) initiatives

Impact of cost on scale

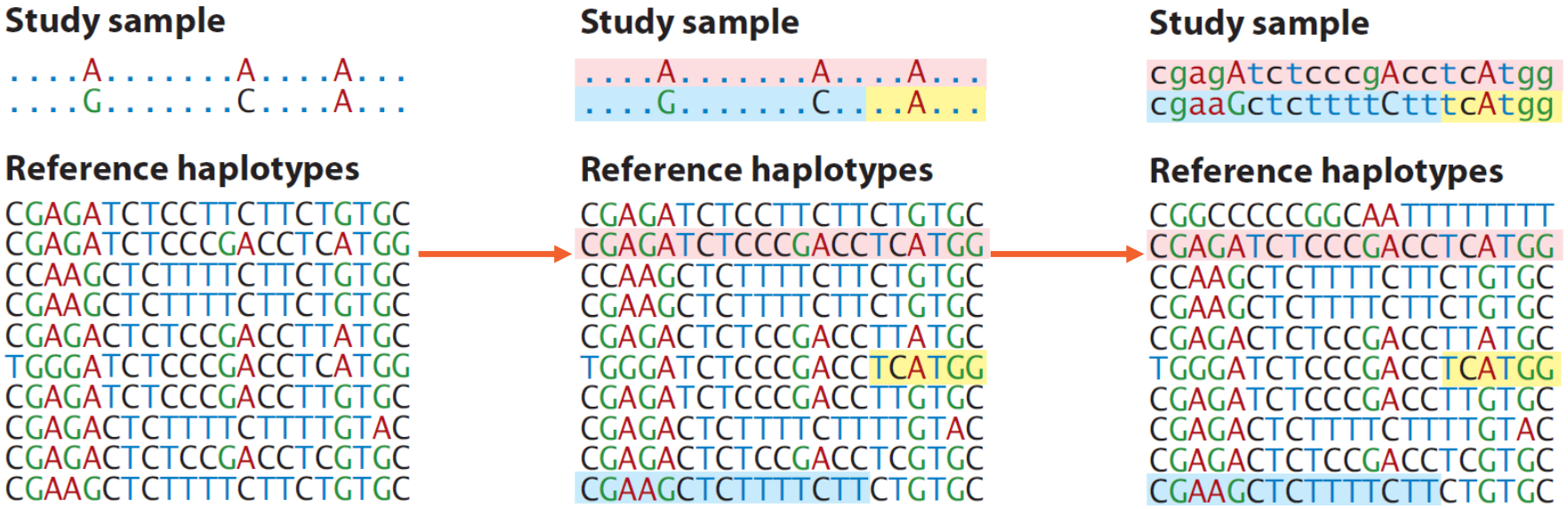
Approach	Cost*	N = 100K	$10M (N)
Genotype array	$50	$5M	200K
Exome sequence	$300	$30M	33K
Genome sequence	$600	$60M	17K

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

Power assumptions:
MAF = 0.05, OR = 1.1
P = 5e-8



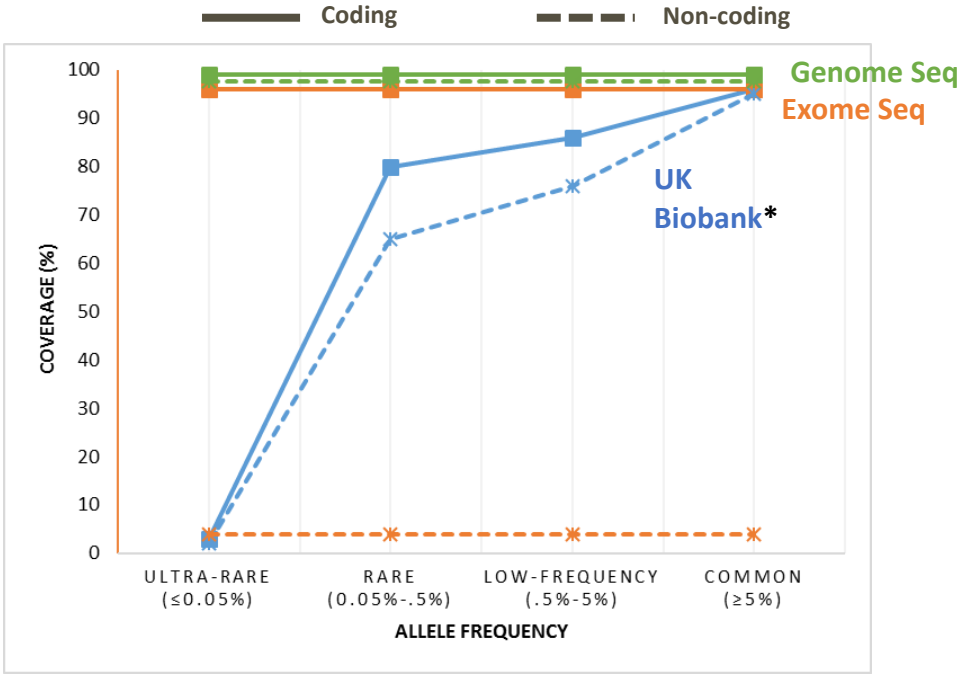
Imputation allows us to fill in unmeasured genotypes



Coverage trade-offs

Exomes (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

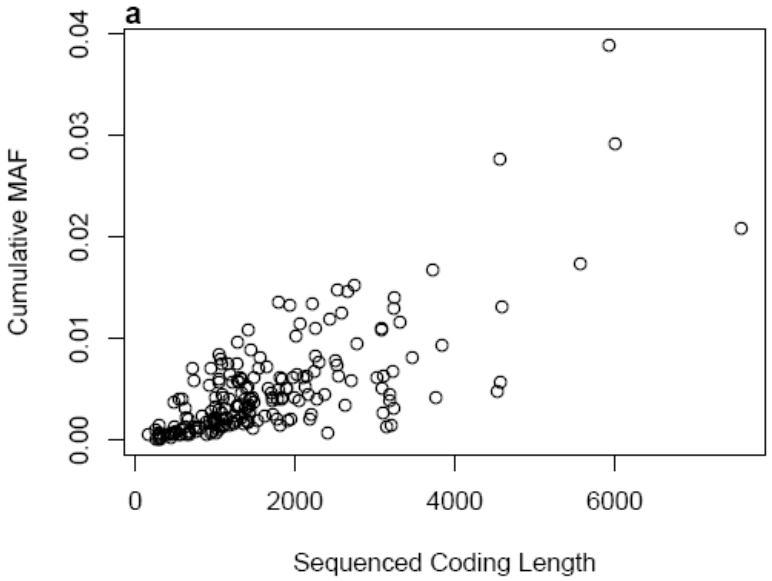


*Coverage reported as imputation quality (r^2)

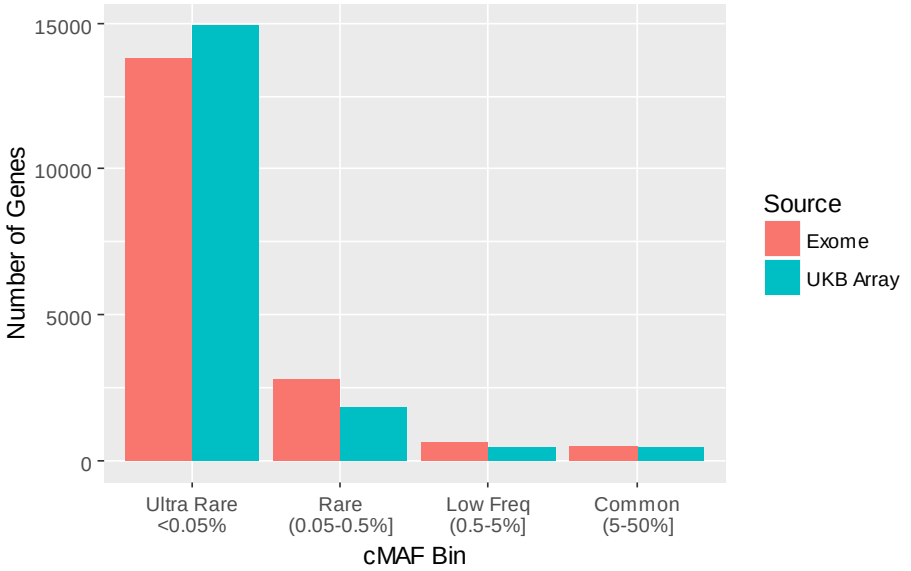
Capture of pLoF variants

Sequence (extrapolated from ExAC) versus UK Biobank genotypes

For most genes, cumulative frequency of pLoF variants is proportional to coding length



Sequencing modestly improves the cMAF of pLoF variants -> statistical power

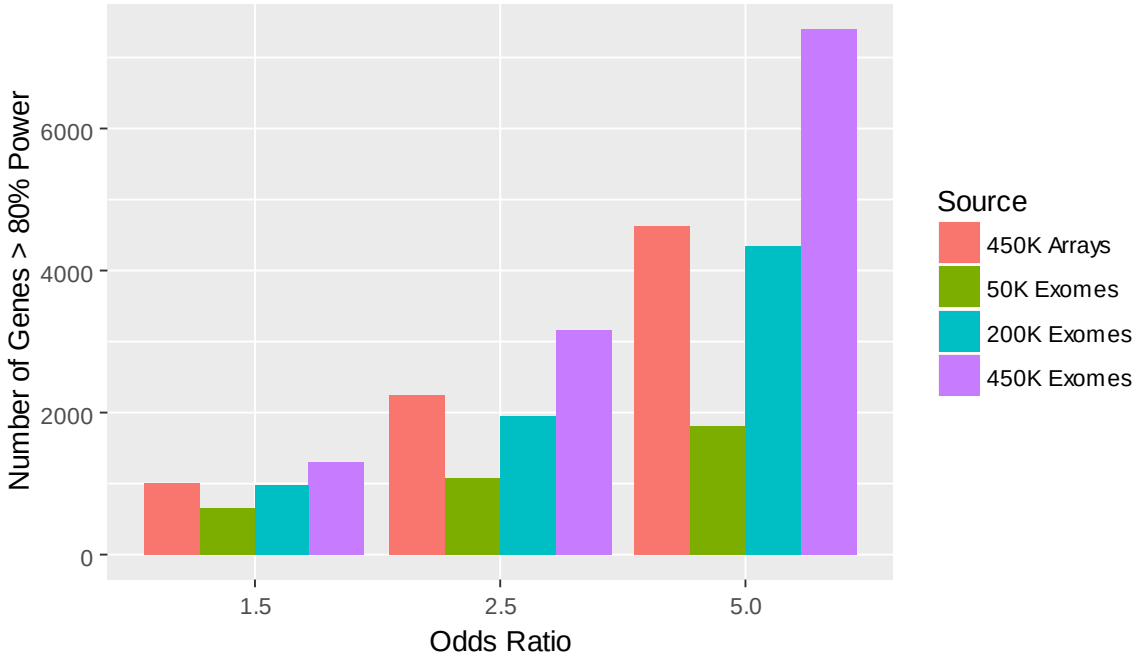


Capture of pLoF variants

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

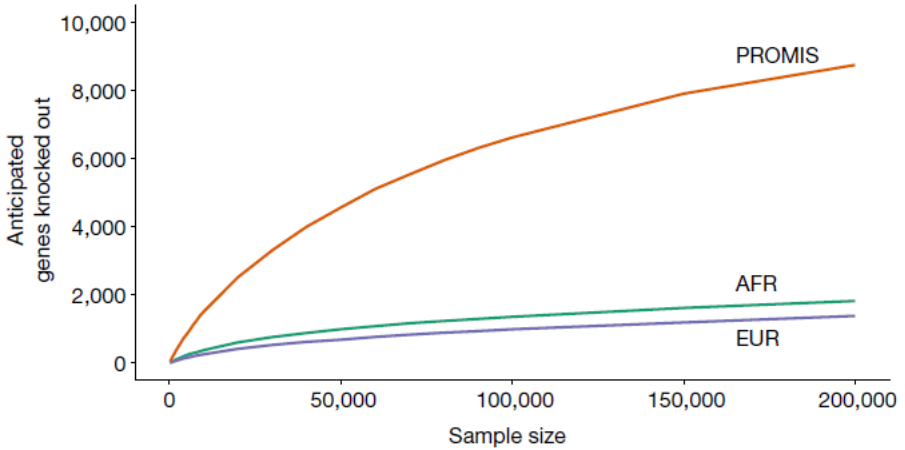
doi:10.1038/nature22034

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 Trindade[6], 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]§

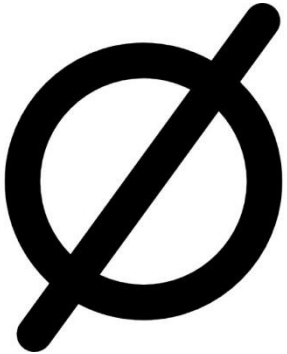
Offspring of consanguineous unions are more likely to be homozygous for loss-of-function (LoF) mutations than non-consanguineous populations (i.e. human knock-outs)

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

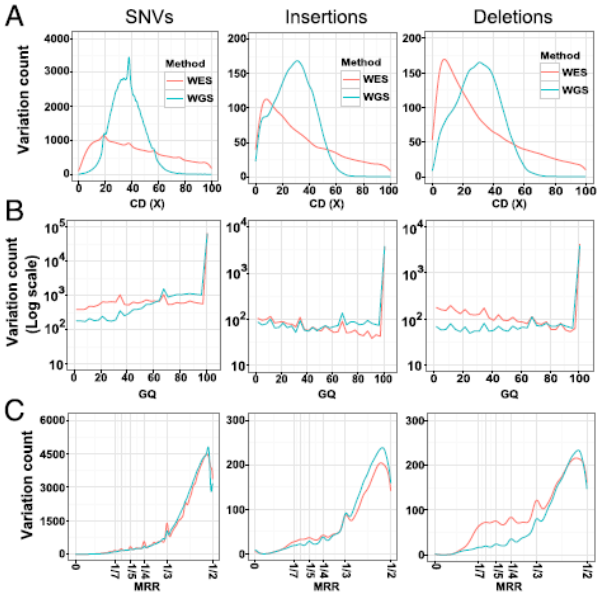


Whole genome versus exome

Rare noncoding variants

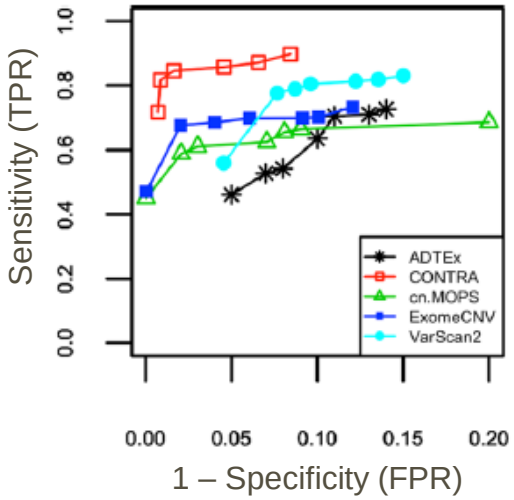


Rare Coding variants



Structural variants

WES calls of duplications relative to WGS: comparison of 5 methods

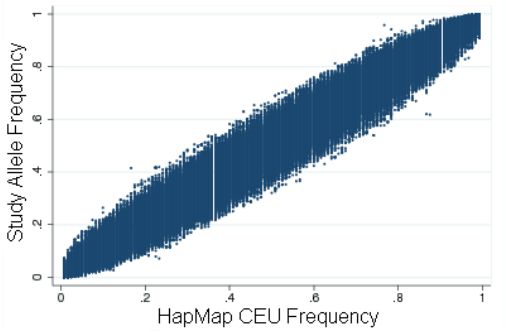


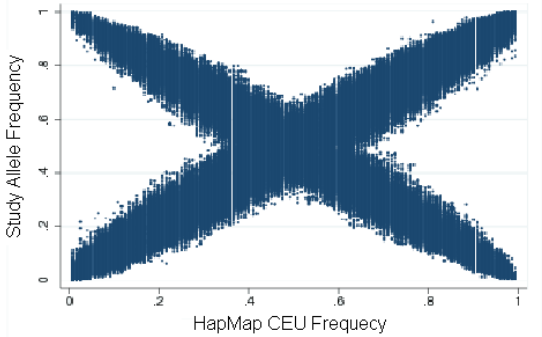
Belkadi *et al.* (2015) PNAS 112:5473-8

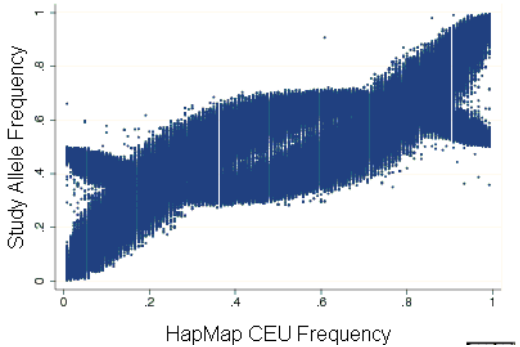
Zare *et al.* (2017) BMC Bioinf 18:286

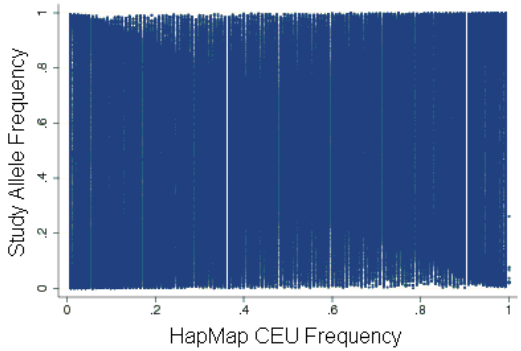
Standards: when good data go bad

Real world horror stories in data sharing

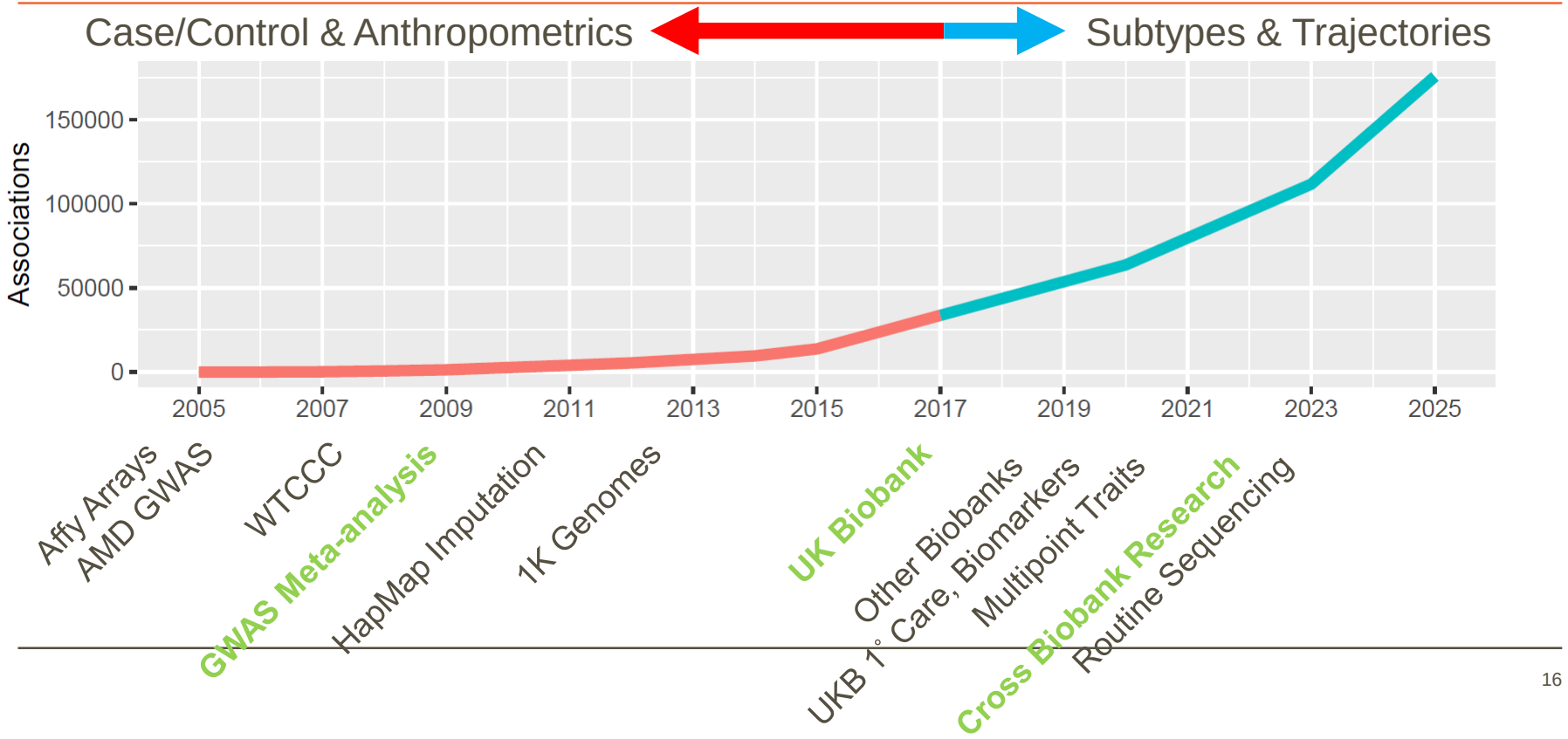








Advances in genetics research driven by technology and scientific culture



Special thanks

Kijoung Song

Josh Hoffman

Ioanna Tachmazidou

Ashutosh Pandey

Giovanni Dall'Olio

Mathias Chiano

John Whittaker

Toby Johnson

Meg Ehm

Matt Nelson

Laura Yerges-Armstrong

Robert Scott

Backup

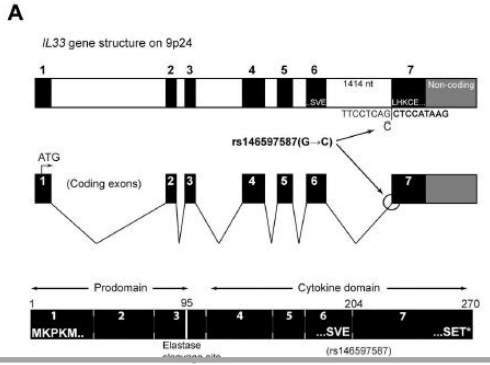
Large-scale sequencing efforts have identified informative LoF variants and new associations

PLOS GENETICS

RESEARCH ARTICLE

A rare *IL33* loss-of-function mutation reduces blood eosinophil counts and protects from asthma

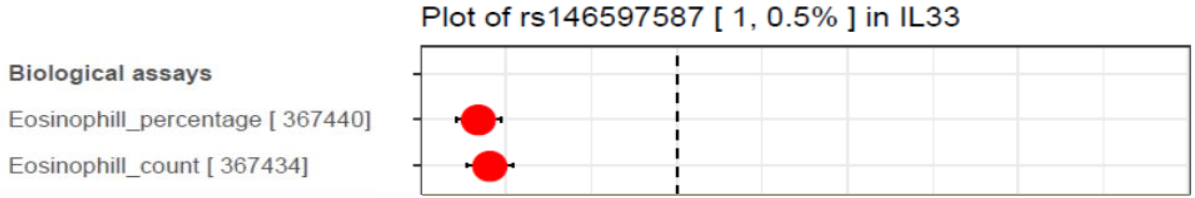
Dirk Smith[1☯¤a], Hannes Helgason[2,3☯], Patrick Sulem[2], Unnur Steina Bjornsdottir[4], Ai Ching Lim[1], Gardar Sveinbjornsson[2], Haruki Hasegawa[1], Michael Brown[1¤b], Randal R. Ketchem[1¤c], Monica Gavala[1], Logan Garrett[1¤d], Adalbjorg Jonasdottir[2], Aslaug Jonasdottir[2], Asgeir Sigurdsson[2], Olafur T. Magnusson[2], Gudmundur I. Eyjolfsson[5], Isleifur Olafsson[6], Pall Torfi Onundarson[7,8], Olof Sigurdardottir[8,9], David Gislason[10], Thorarinn Gislason[8,10], Bjorn Runar Ludviksson[8,11], Dora Ludviksdottir[8,10], H. Marike Boezen[12,13], Andrea Heinzmann[14], Marcus Krueger[14], Celeste Porsbjerg[15], Tarunveer S. Ahluwalia[16], Johannes Waage[16], Vibeke Backer[15], Klaus A. Deichmann[14], Gerard H. Koppelman[12,17], Klaus Bønnelykke[16], Hans Bisgaard[16], Gisli Masson[2], Unnur Thorsteinsdottir[2,8], Daniel F. Gudbjartsson[2,3], James A. Johnston[1], Ingileif Jonsdottir[2,8,11]*, Kari Stefansson[2,8]*



What else can we do?

Do we see a similar association with eosinophils in UKB?

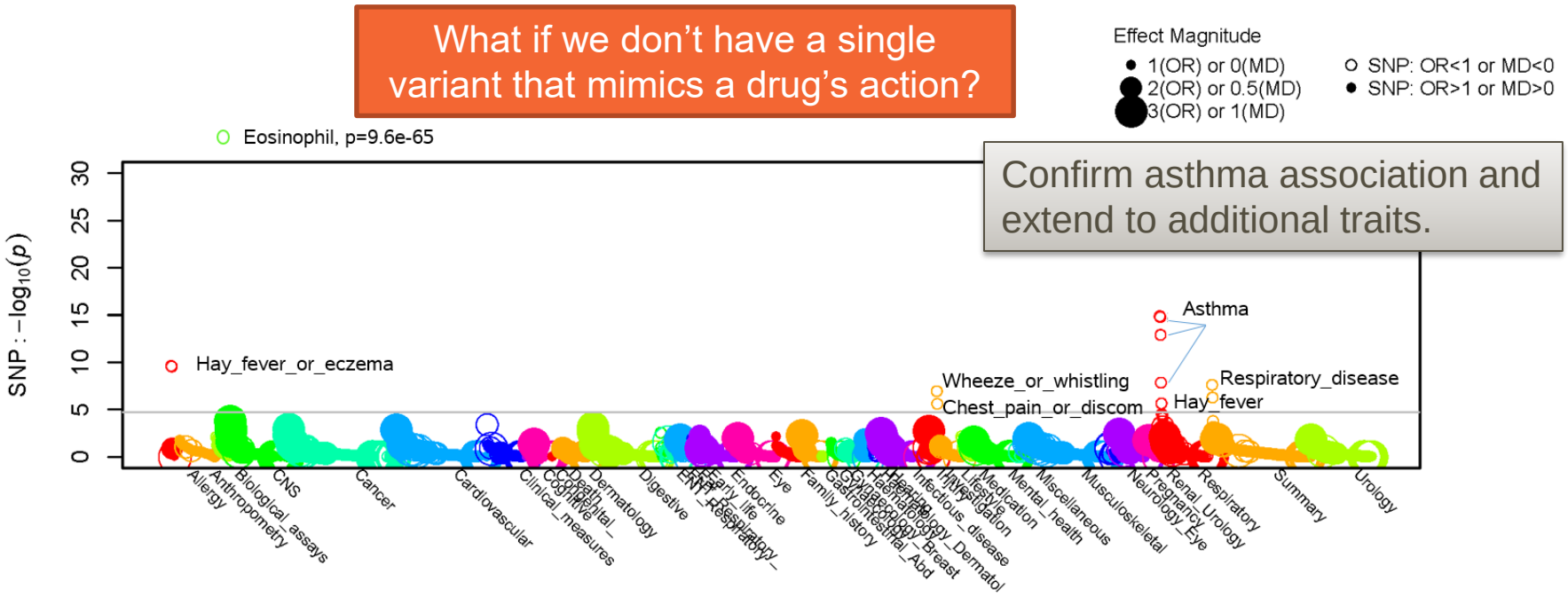
YES! ($p=9.6x10^{-65}$)



Smith D, Helgason H, Sulem P, Bjornsdottir US, Lim AC, et al. (2017) A rare IL33 loss-of-function mutation reduces blood eosinophil counts and protects from asthma. PLOS Genetics 13(3): e1006659. https://doi.org/10.1371/journal.pgen.1006659

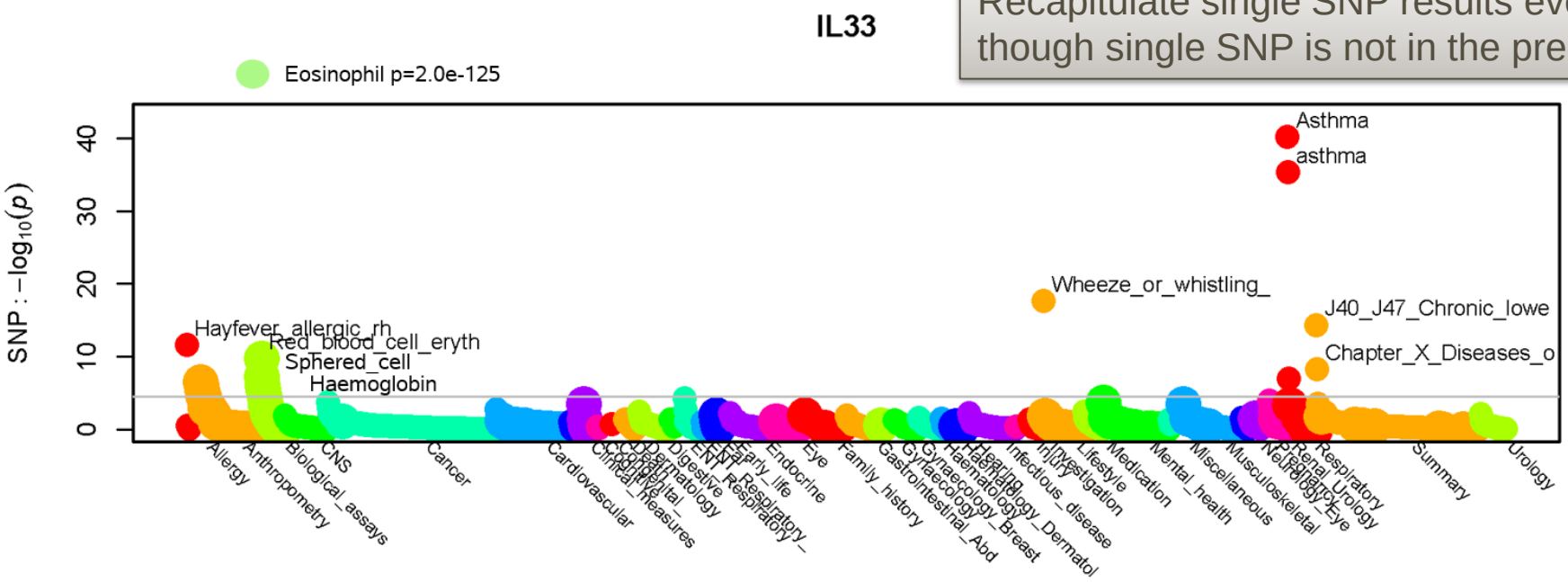
PheWAS of IL33 LoF allele in UKBB recapitulates known associations and identifies new ones

What if we don't have a single variant that mimics a drug's action?



Confirm asthma association and extend to additional traits.

Phenome-wide predixcan analysis recapitulate those from LoF variant



Recapitulate single SNP results even though single SNP is not in the predictor.