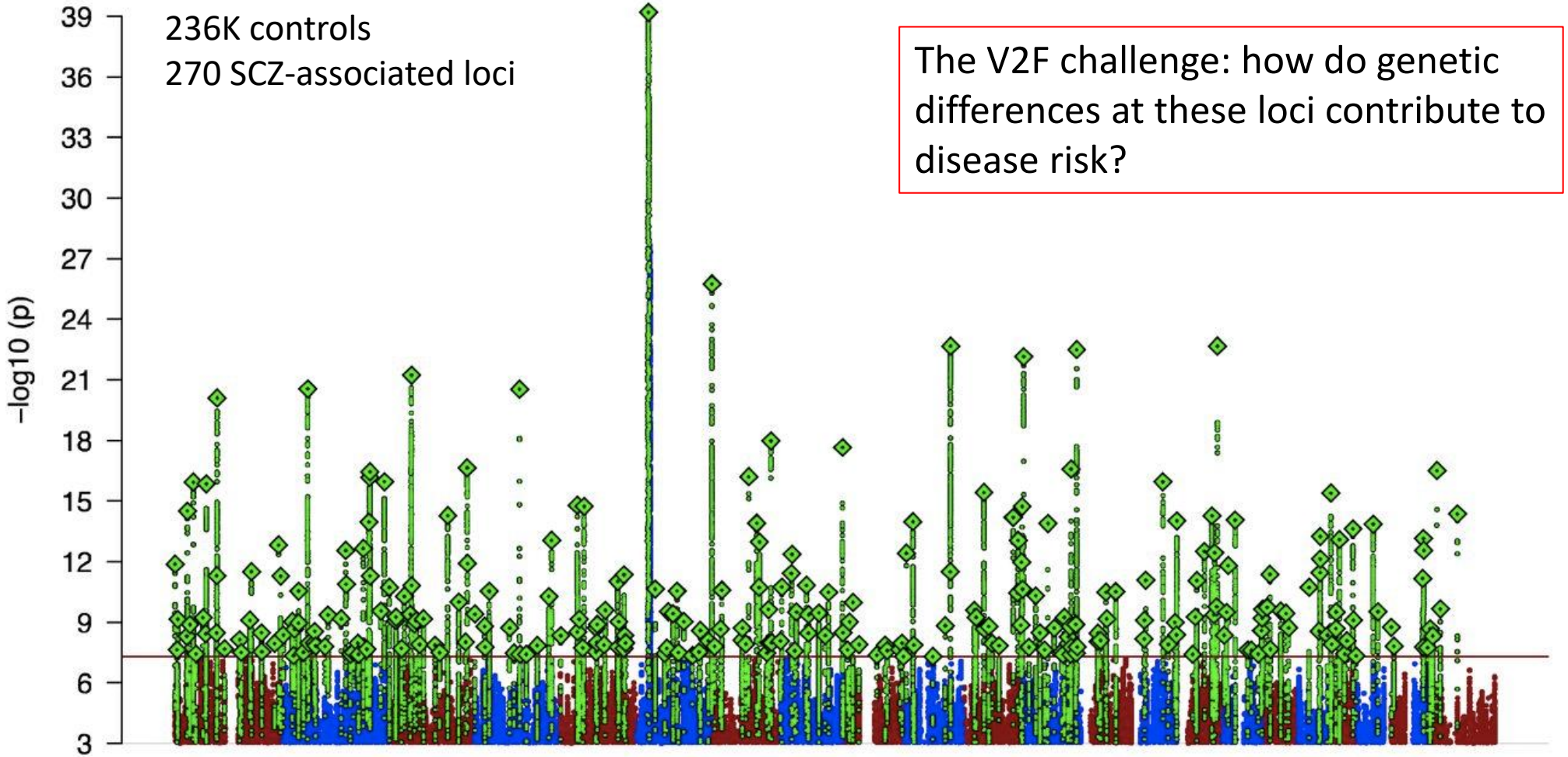
Insights from complex trait fine-mapping across diverse biobanks

Hilary Finucane
Broad Institute and Massachusetts General Hospital
NASEM Workshop
October 5, 2021

270 SCZ-associated loci

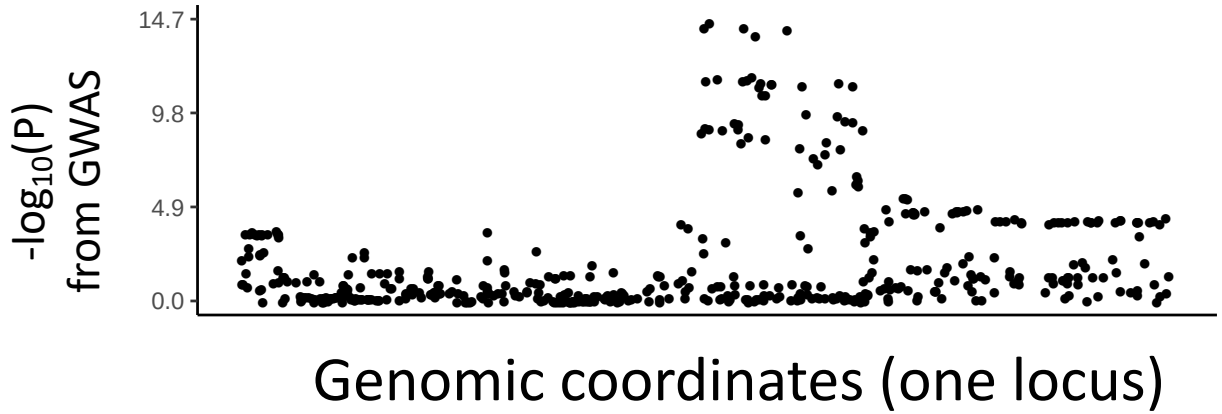
69K cases
236K controls
270 SCZ-associated loci



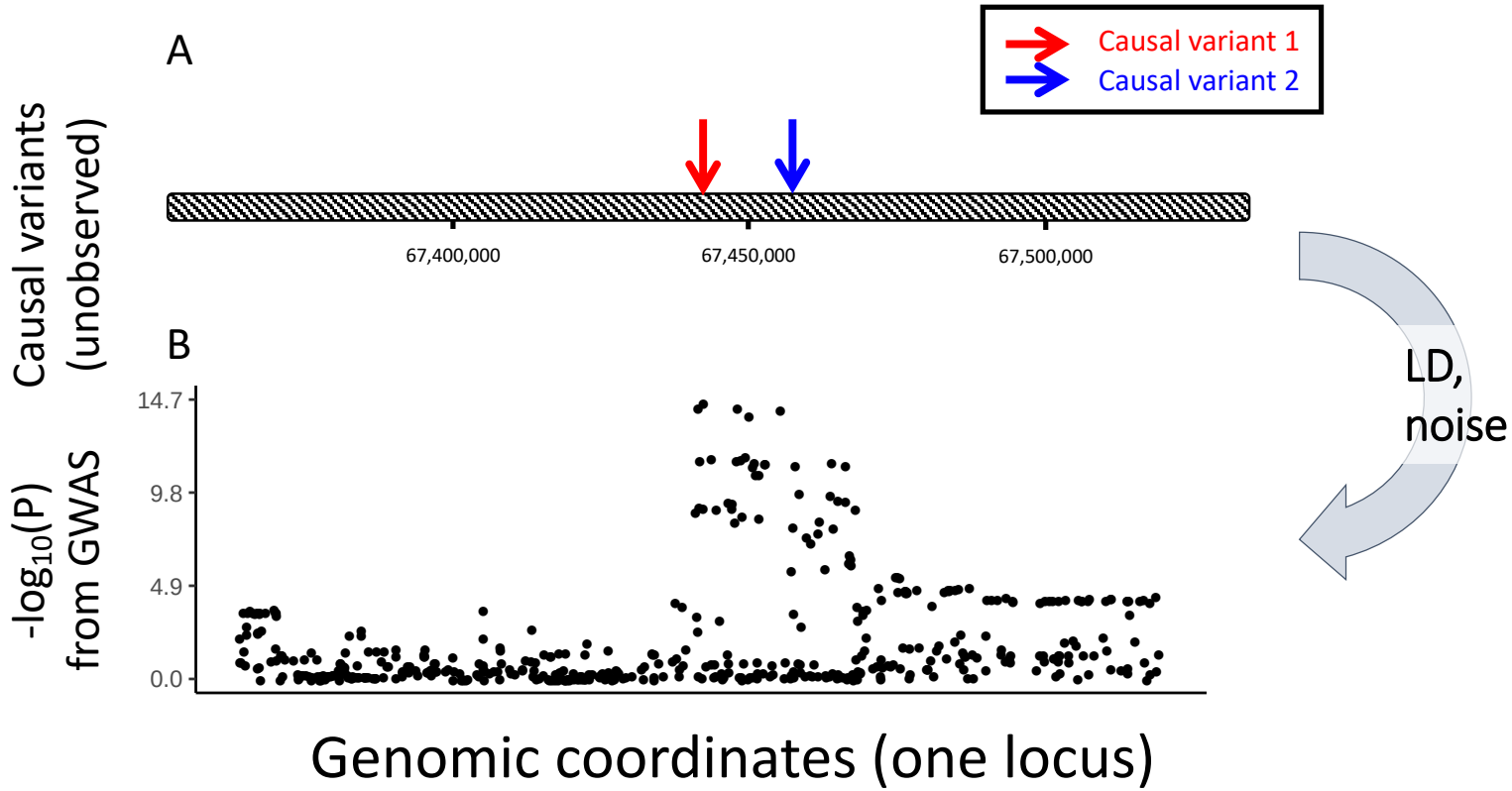
The V2F challenge: how do genetic differences at these loci contribute to disease risk?

Schizophrenia Working Group of the PGC, 2020 *medRxiv*

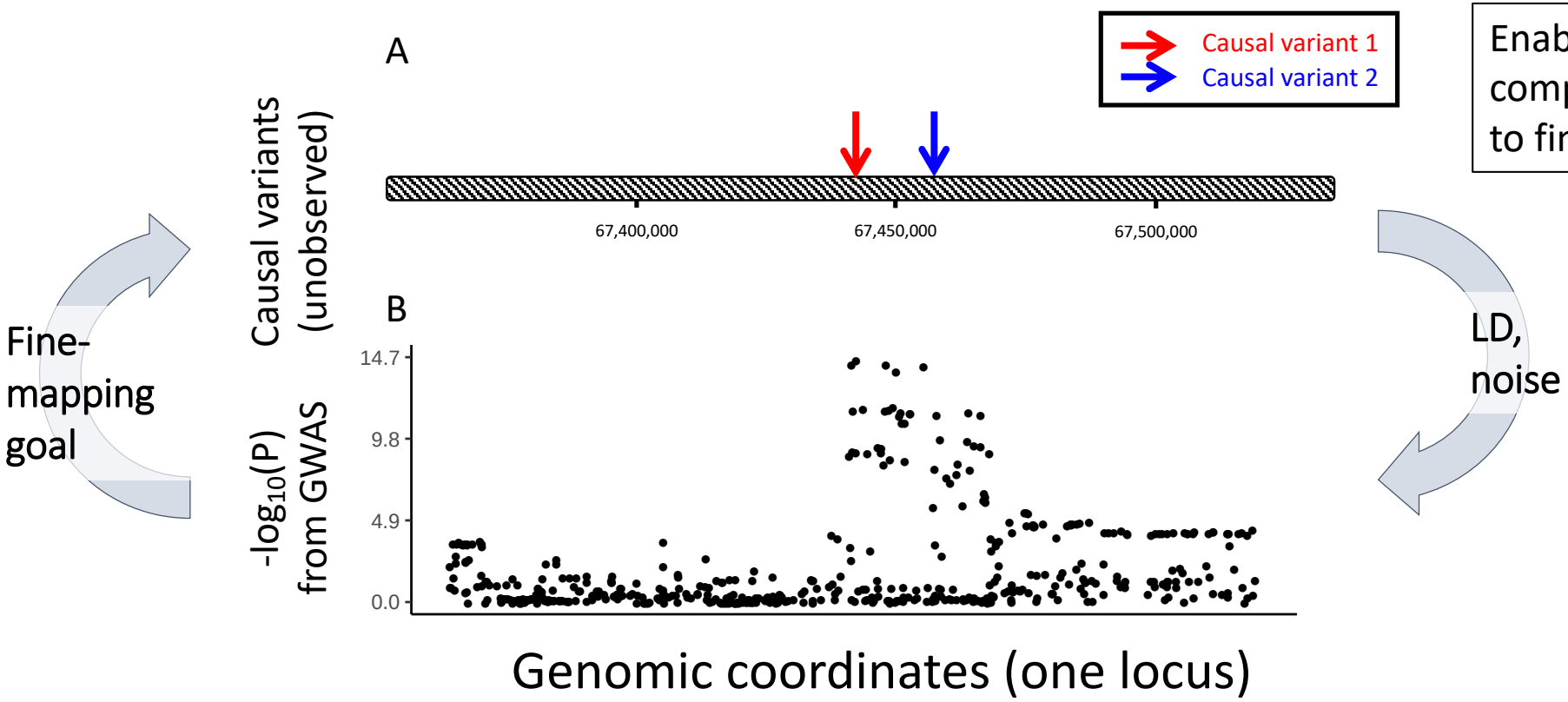
Fine-mapping: what are the causal SNPs?



Fine-mapping: what are the causal SNPs?

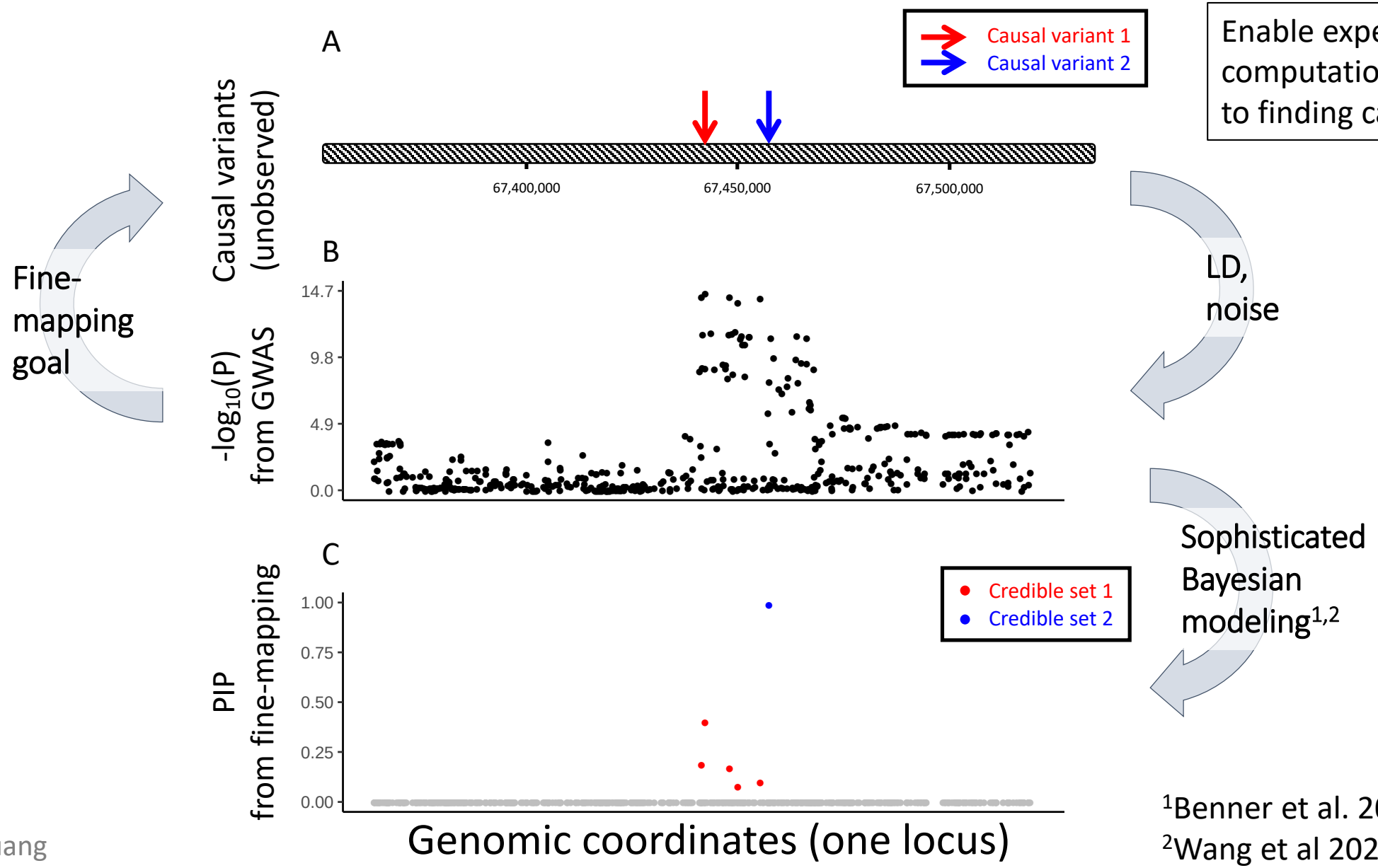


Fine-mapping: what are the causal SNPs?



Enable experimental and computational approaches to finding causal genes

Bayesian fine-mapping outputs PIPs and credible sets



Enable experimental and computational approaches to finding causal genes

[1]Benner et al. 2016 *Bioinformatics*
[2]Wang et al 2020 *J R Stat Soc B*

We performed fine-mapping in three biobanks

Biobank Japan (BBJ)
79 complex traits

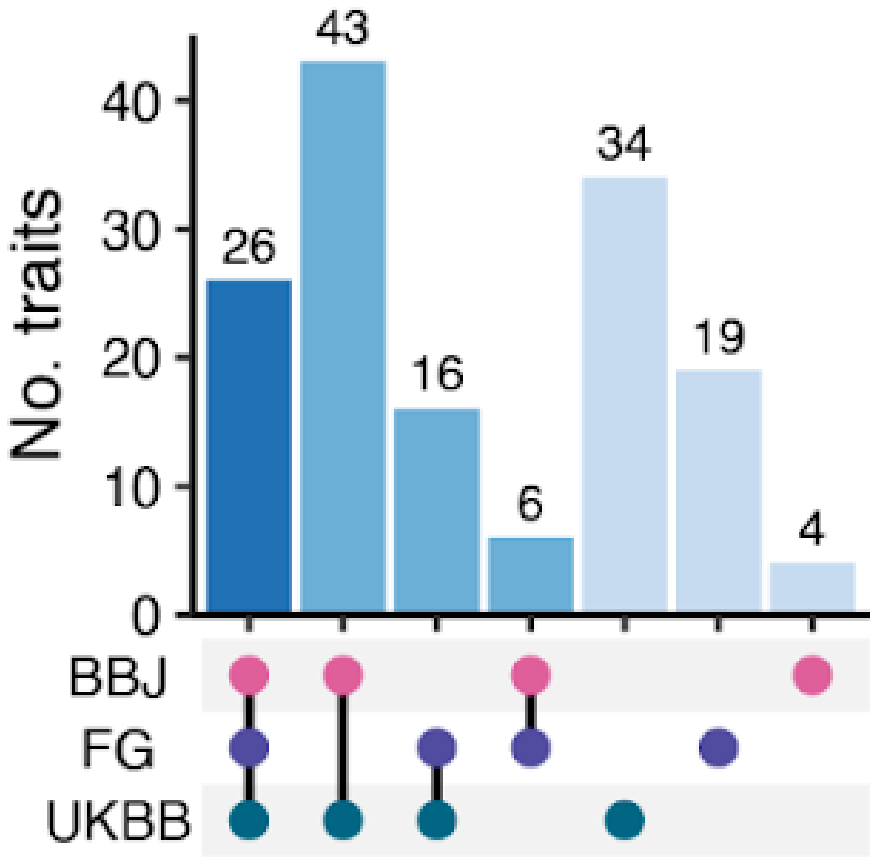
Japanese
(*n* = 178,726)

FinnGen (FG) release 6
67 complex traits

Finnish
(*n* = 271,341)

UK Biobank (UKBB)
119 complex traits

White British
(*n* = 361,194)



Masahiro Kanai

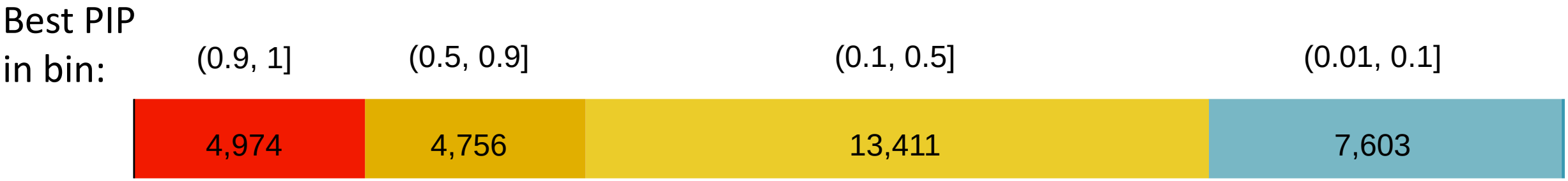
Our pipeline

- *Association*: mixed linear models
- *Fine-mapping:* FINEMAP[1] and SuSiE[2]
- *Simulations* first to set parameters

[1]Benner et al. 2016 *Bioinformatics*
[2]Wang et al 2020 *J R Stat Soc B*

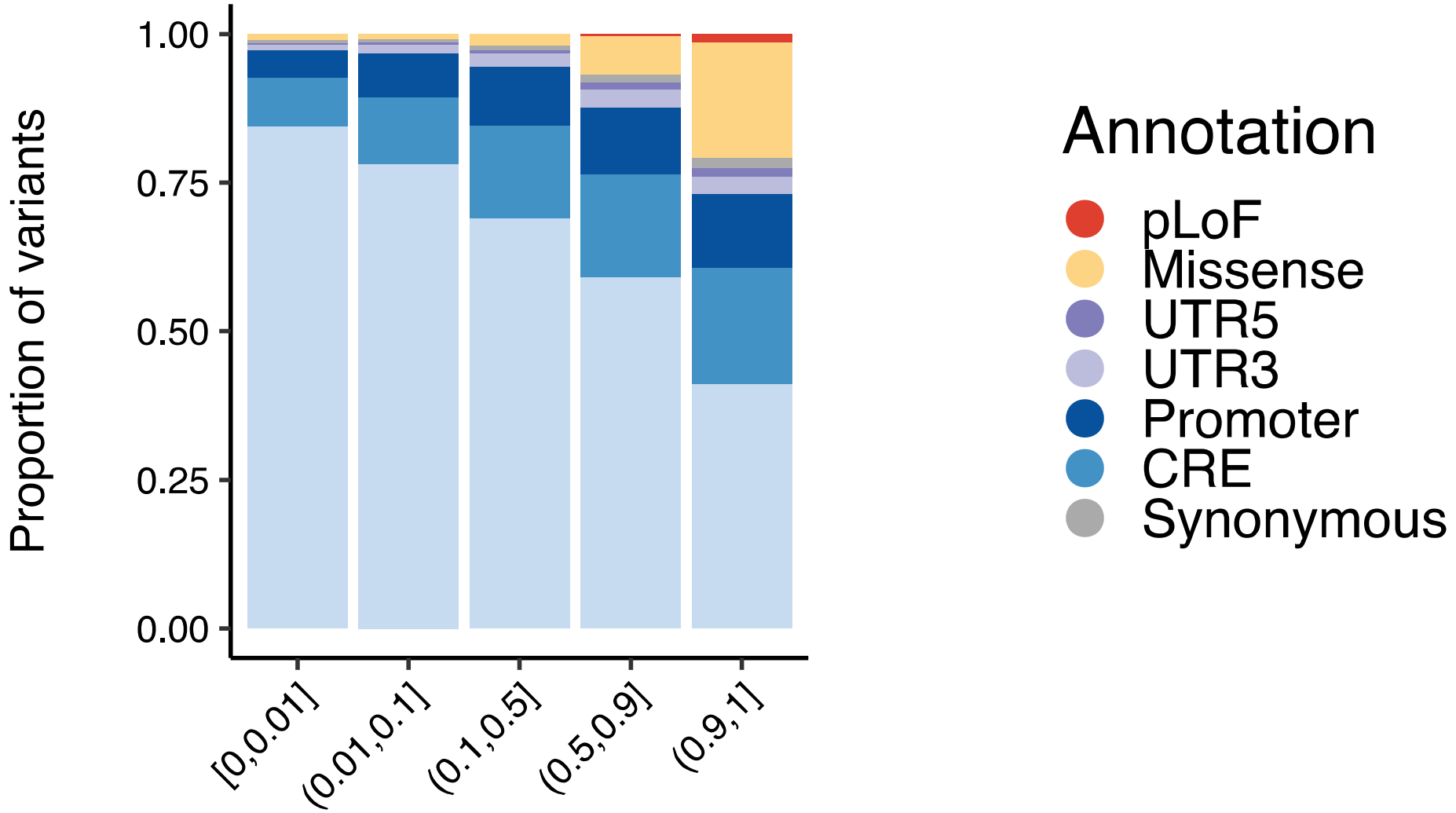
Kanai et al. 2021 medRxiv

We identified 30,812 credible sets



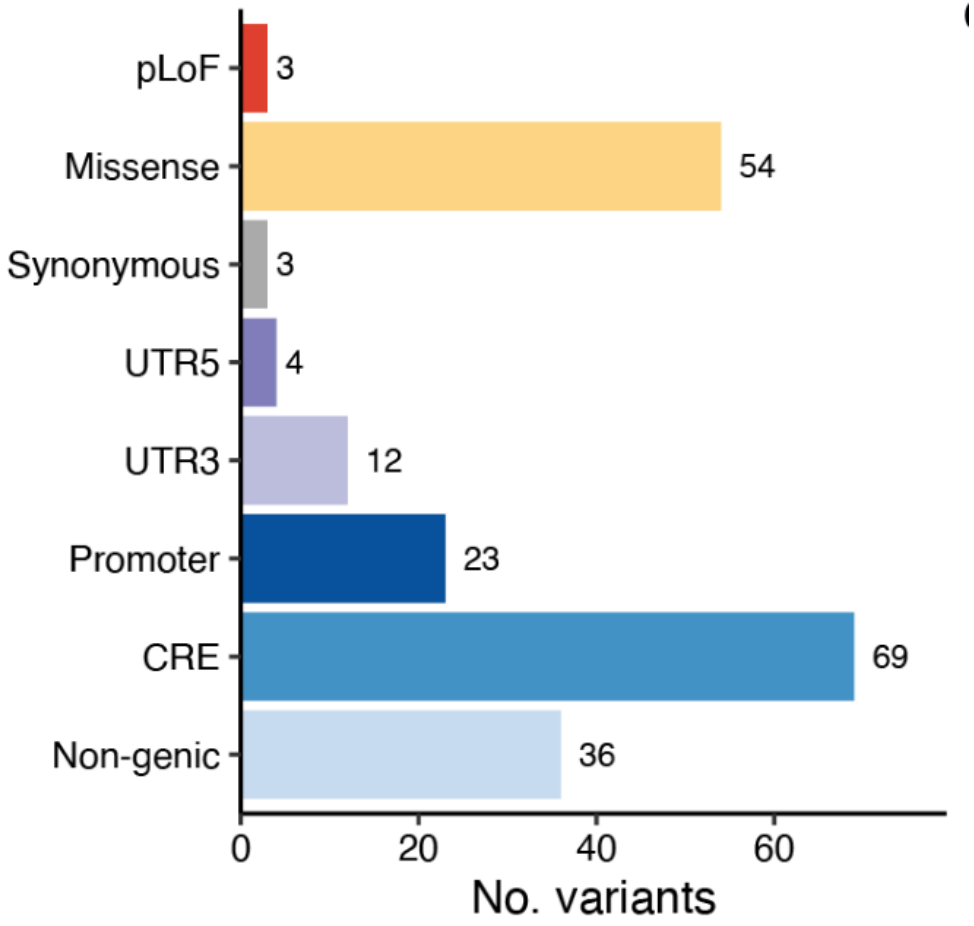
- 75% have at least one variant with PIP>0.1
- 16% have at least one variant with PIP>0.9

We found thousands of high-PIP variants, enriched for functional annotations



Kanai et al. 2021 medRxiv

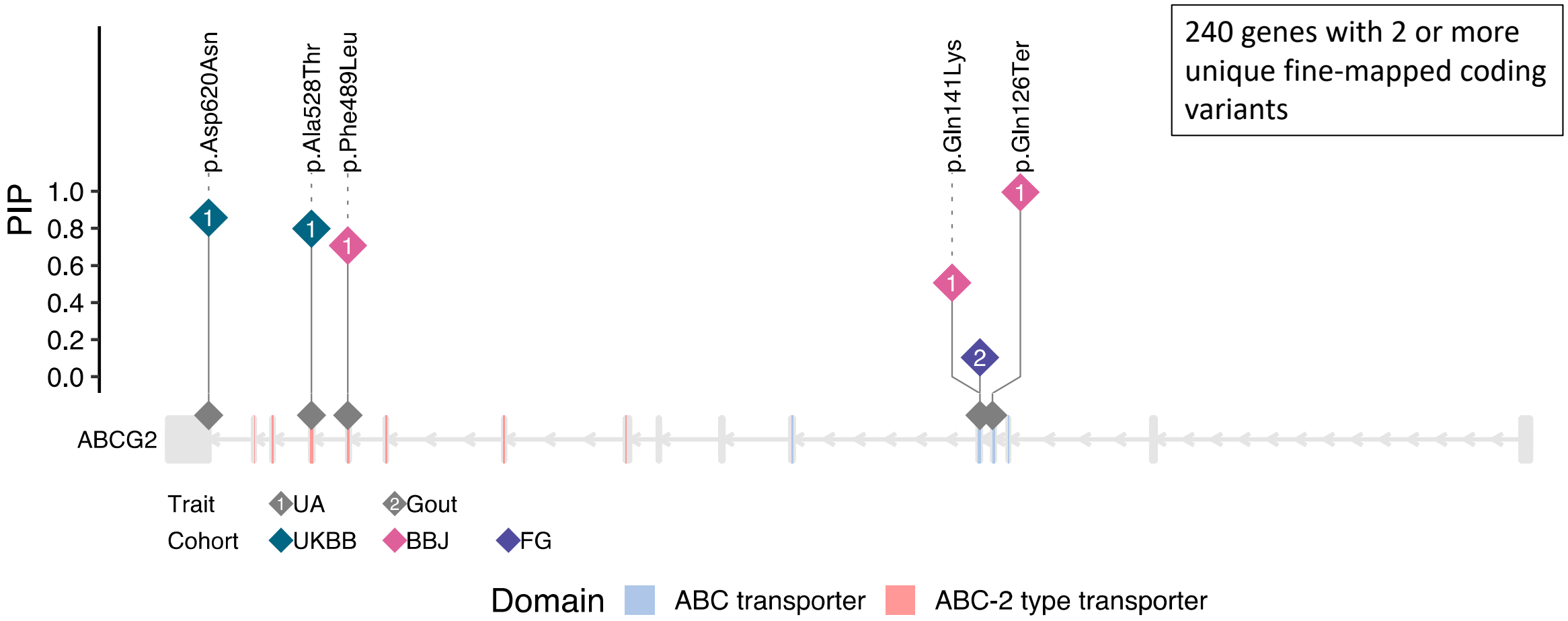
We identified 285 high-confidence putative causal variants that replicate across biobanks



Kanai et al. 2021 medRxiv

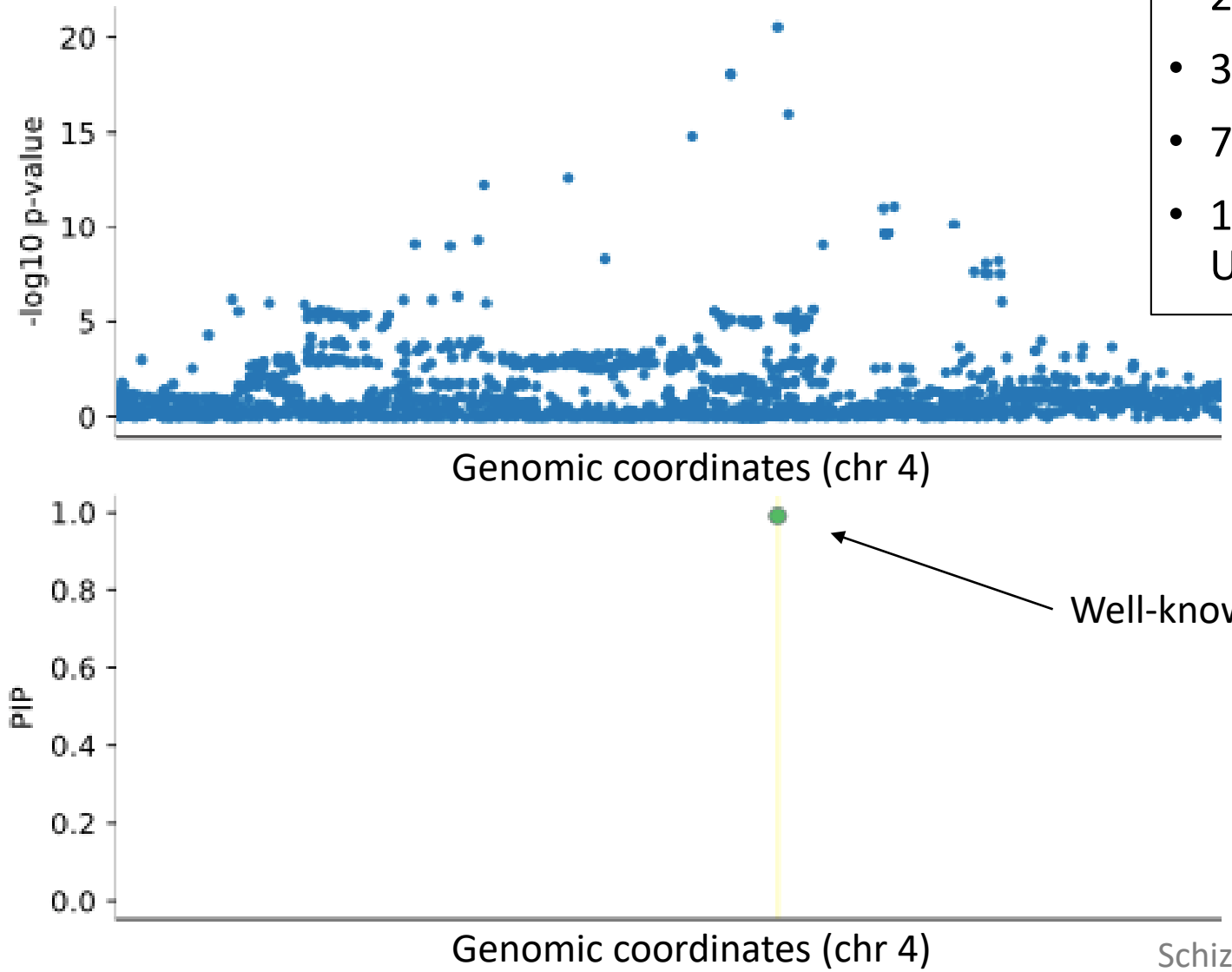
Why so few? Details in our preprint.

Fine-mapping in multiple populations identifies allelic series



Kanai et al. 2021 medRxiv

The PGC has performed fine-mapping for SCZ



- 288 credible sets
- 32 credible sets have 5 or fewer SNPs
- 7 SNPs with PIP>0.95
- 19 genes with fine-mapped coding or UTR variant

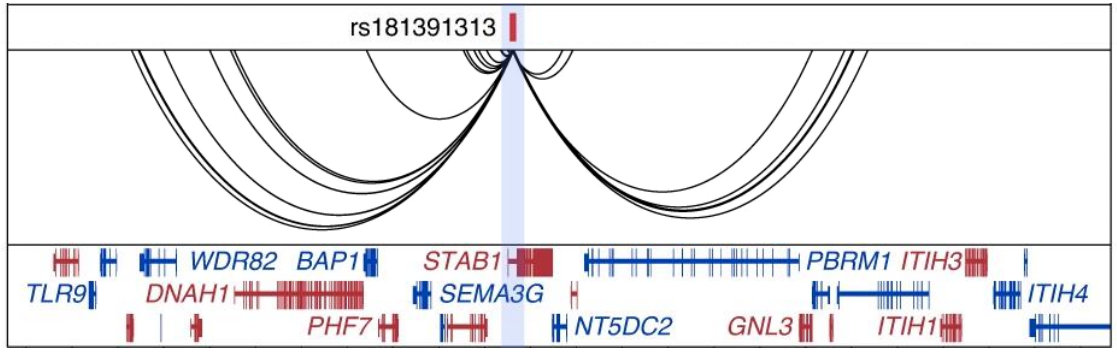
Well-known missense variant in SLC39A8

Schizophrenia Working Group of the PGC, 2020 *medRxiv*

Fine-mapping aids gene prioritization and benchmarking

Many ways to nominate a gene:

- Coding variant
- Local epigenetics
- eQTL-based analysis
- Closest gene
- Similarity-based

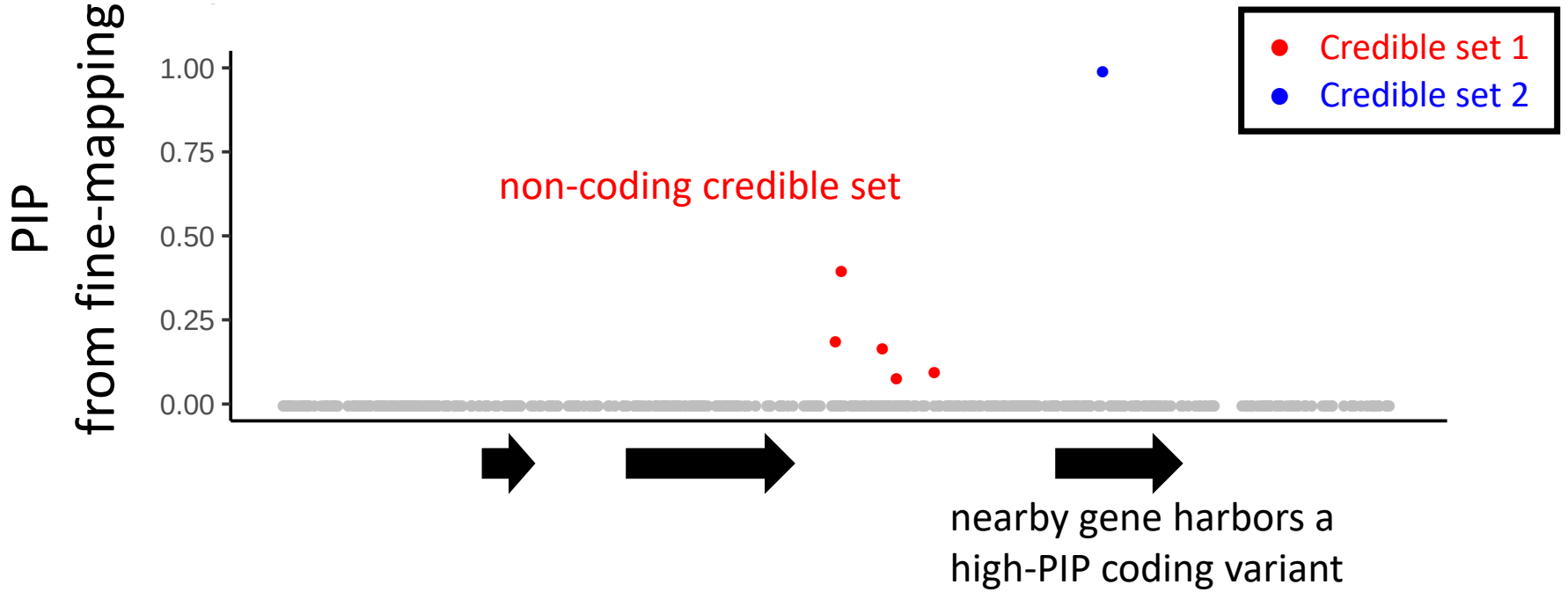


Corces et al. 2020 Nat Genet

E.g.: deLeeuw et al. 2015 *PLoS Comp Bio,* Ulirsch et al. 2019 *Nat Genet*, The FANTOM Consortium et al. 2014 *Nature,* Liu et al 2017 *Genome Biol,* Jung et al. 2019 *Nat Genet,* Javierre et al. 2016 *Cell,* Fulco et al. 2019 *Nat Genet,* Gusev et al. 2016 *Nat Genet,* Hormozdiari et al. 2016 *Am J Hum Genet,* Corces et al. 2020 Nat Genet, Pers et al. 2015 *Nat Commun*, Stacey et al. 2018 *Nucleic Acids Res*, Greene et al 2015 *Nat Genet,* Pers et al 2015 Nat Commun, Greene et al. 2015 *Nat Genet*, …

Large-scale fine-mapping in UKBB provides a new evaluation set of genes

Jesse Engreitz



>1K cases of a non-coding credible set near a gene with a high-PIP coding variant

We can estimate precision and recall with our new evaluation set

Q1: How confident should we be in what we prioritize?

A1:

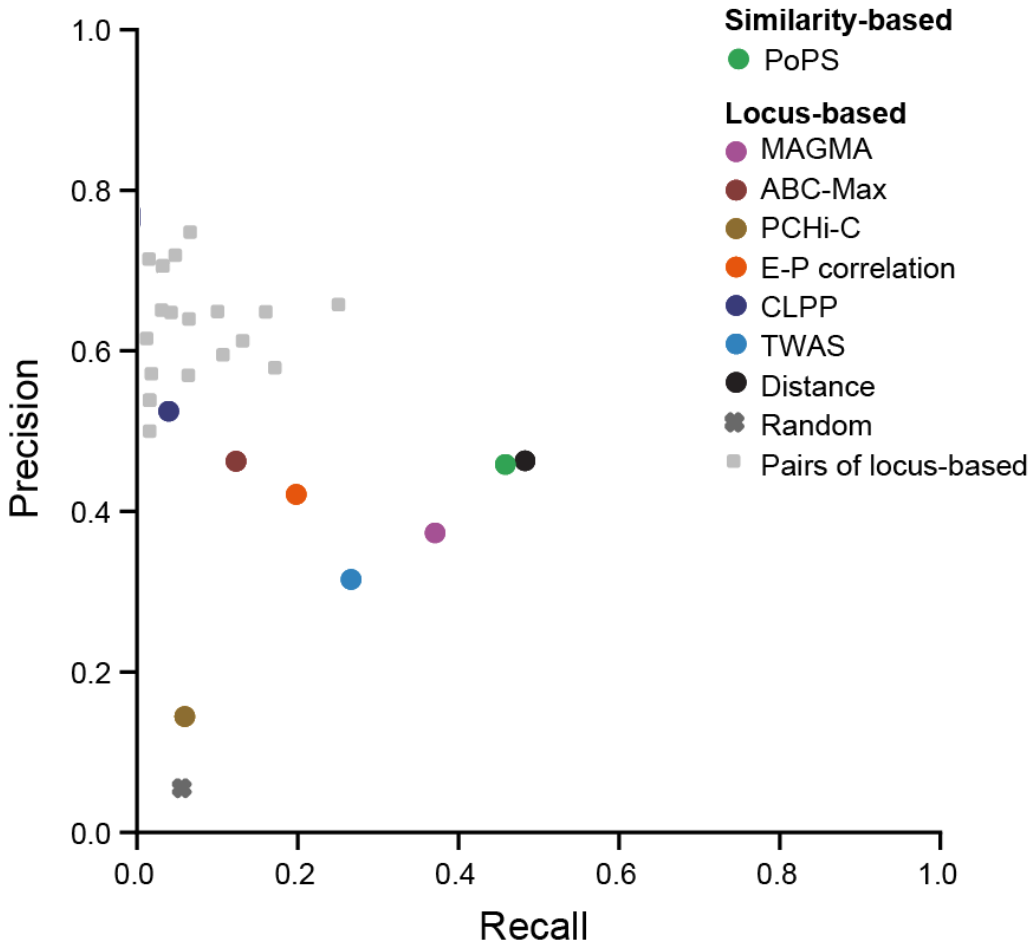
$$\text{Precision} = \Pr(\text{in the evaluation set} \mid \text{prioritized})$$

Q2: How much do we find?

A2:

$$\text{Recall} = \Pr(\text{prioritized} \mid \text{in the evaluation set})$$

No individual method achieves very high precision or recall

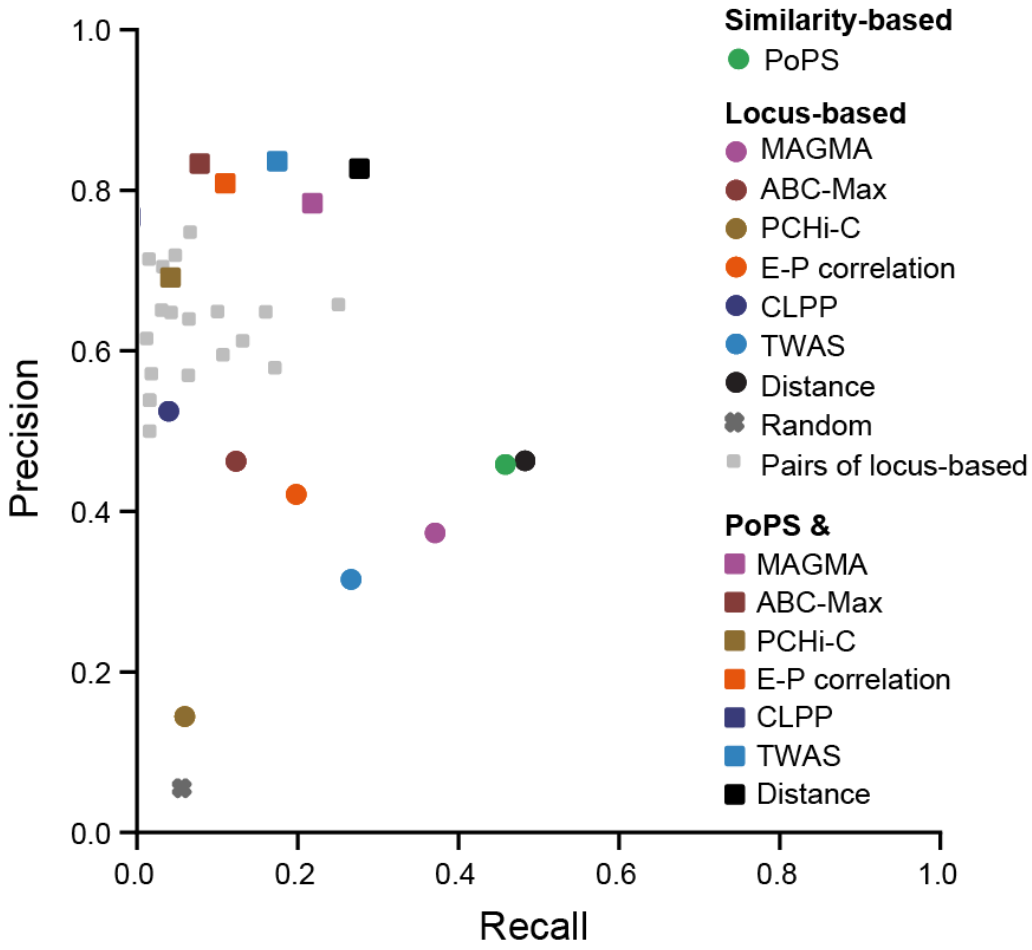


Uses expression, pathways, etc

Use local GWAS signal: proximity, eQTLs, or epigenetics

Weeks et al. 2020 medRxiv

Combining similarity-based and locus-based methods increases precision



Uses expression, pathways, etc

Use local GWAS signal: proximity, eQTLs, or epigenetics

Require both types of evidence

Weeks et al. 2020 medRxiv

Masahiro Kanai

Mark Daly

Elle Weeks

Osaka University
Yukinori Okada
Saori Sakaue

Biobank Japan
Yoichiro Kamatani
Masato Akiyama
Kazuyoshi Ishigaki
Koichi Matsuda
Michiaki Kubo
Yoshinori Murakami

FinnGen
Aarno Palotie
Mitja Kurki
Juha Karjalainen
Sina Rüeger
Arto Lehisto

Broad Institute
Jacob Ulirsch
Zack McCaw
Ben Neale
Alicia Martin
Wei Zhou
Jon Bloom
Cotton Seed
Carlos Albors
Elle Weeks
Ran Cui
Qingbo Wang
Pardis Sabeti
Steve Reilly
Ryan Tewhey
Brian Trippe
Rebecca Fine
Jenkai Miao
Tejal Patwardhan
Joseph Nasser
Charles Fulco
Katherine Tashman
Francois Aguet
Taibo Li
Jose Ordovas-Montanes
Christopher Smillie
Moshe Biton
Alex Shalek
Ashwin Ananthakrishnan
Ramnik Xavier
Aviv Regev
Rajat Gupta
Kasper Lage
Kristin Ardlie
Joel Hirschhorn
Jesse Engreitz
Eric Lander

Funding:
The Broad Institute
ENCODE
NIH Early Independence Award
Nakajima Foundation

Thank you to participants of:
Biobank Japan
FinnGen
UK Biobank

Preprint: Kanai et al. 2021 medRxiv.
Data: www.finucanelab.com/data