Assessing colocalisation of causal variants for complex and molecular traits

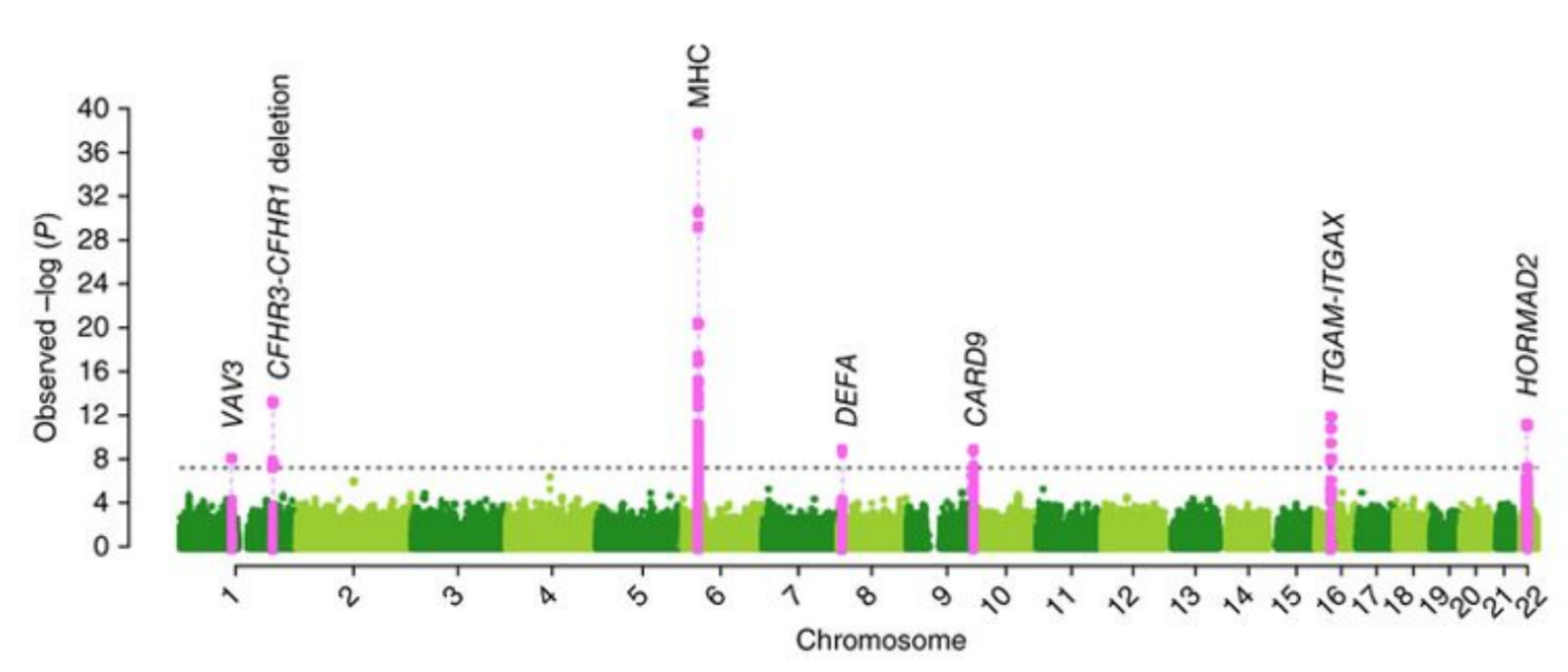
Chris Wallace

chr1swallace chr1swallace.github.io

18th July 2019

University of Cambridge

Identifying molecular traits that mediate a disease association

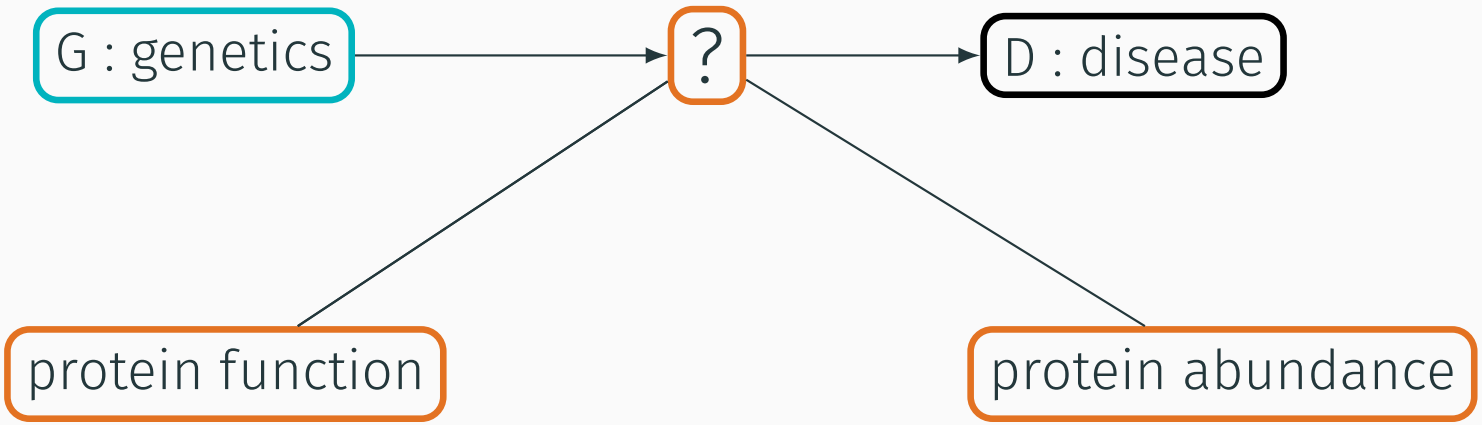


GWAS of IgA Nephropathy (Kiryluk et al, Nat Genet 2014)

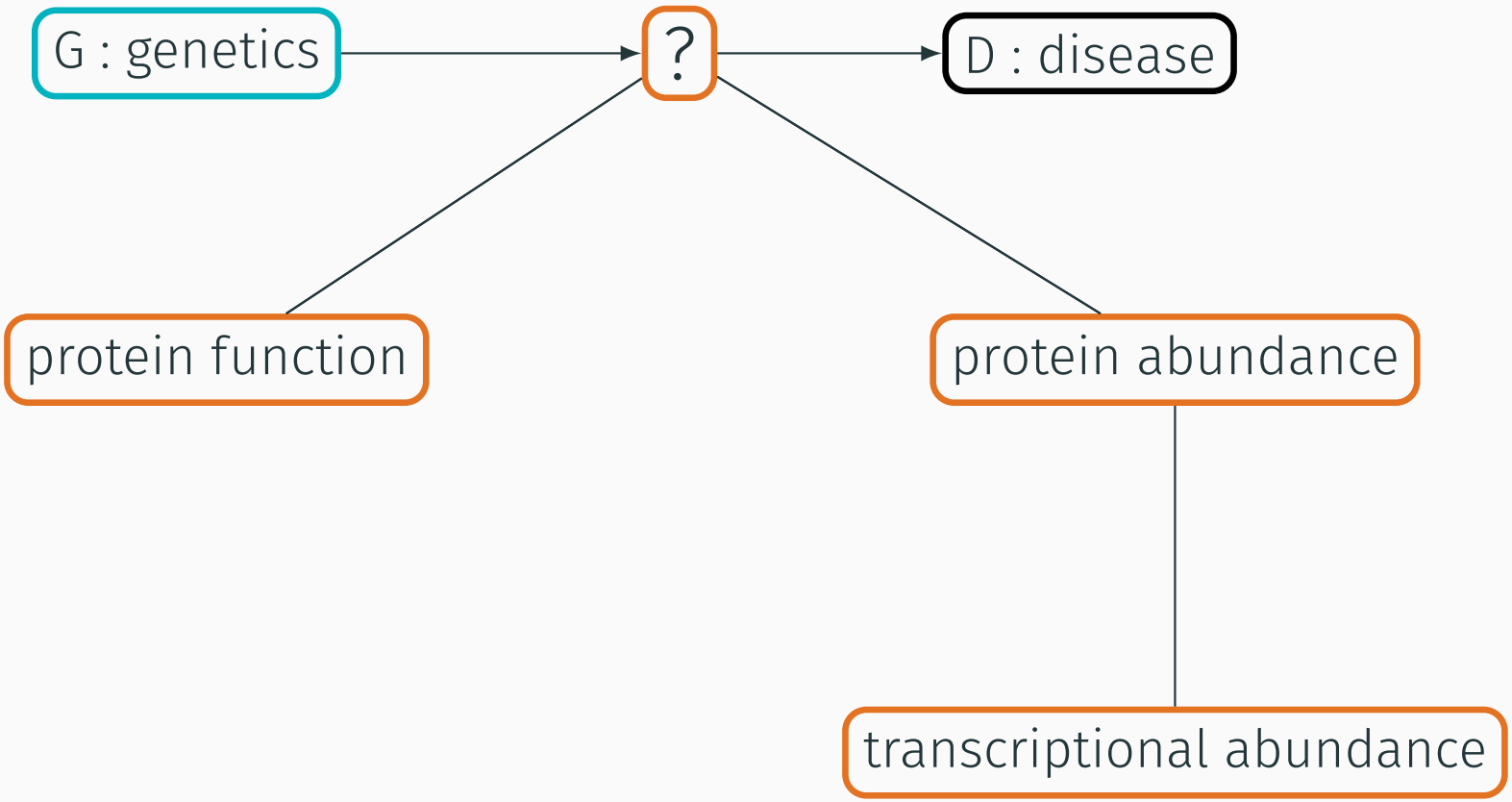
Testing whether a trait is causal for disease

Testing whether a trait is causal for disease

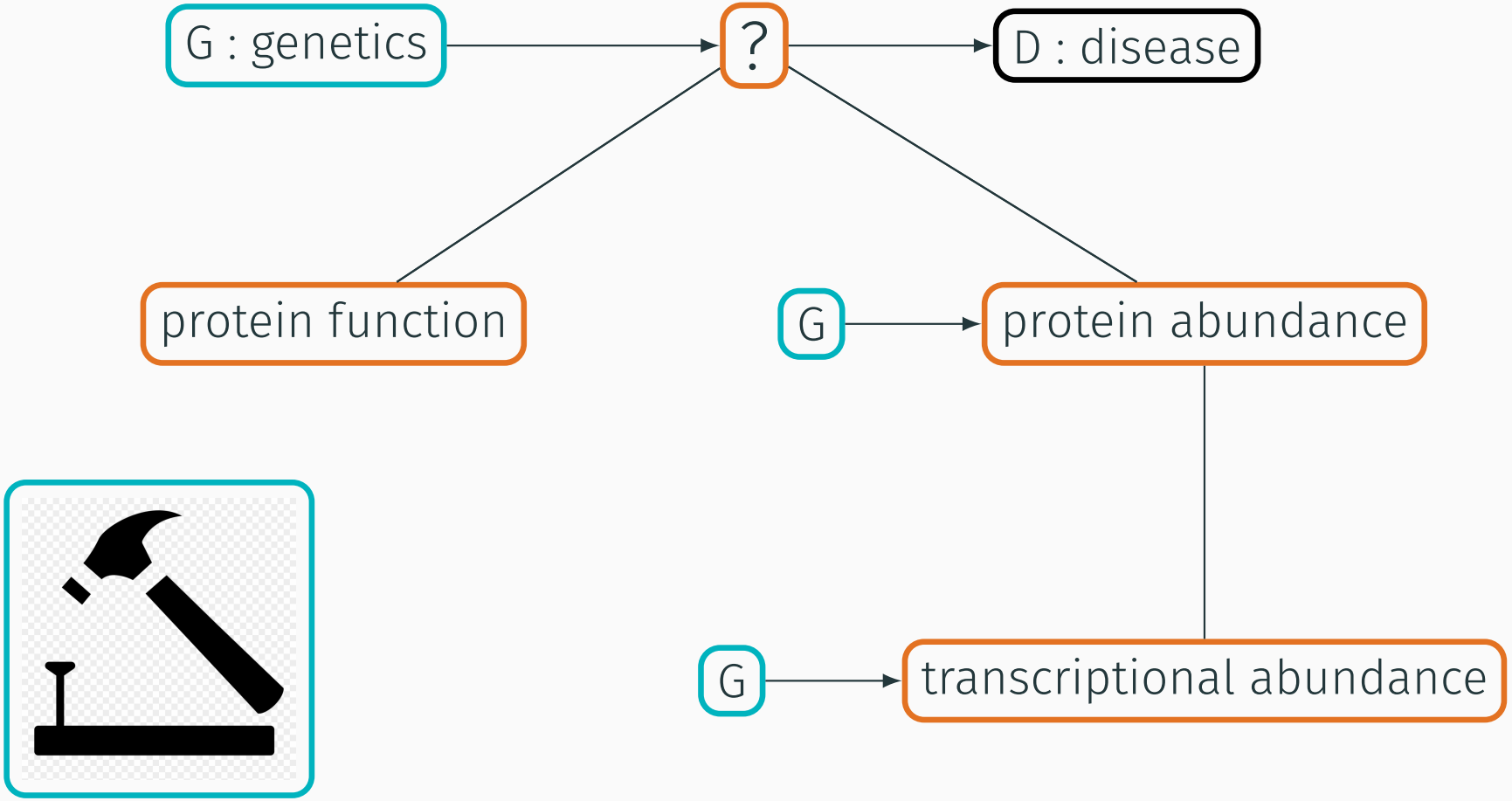
Testing whether a trait is causal for disease



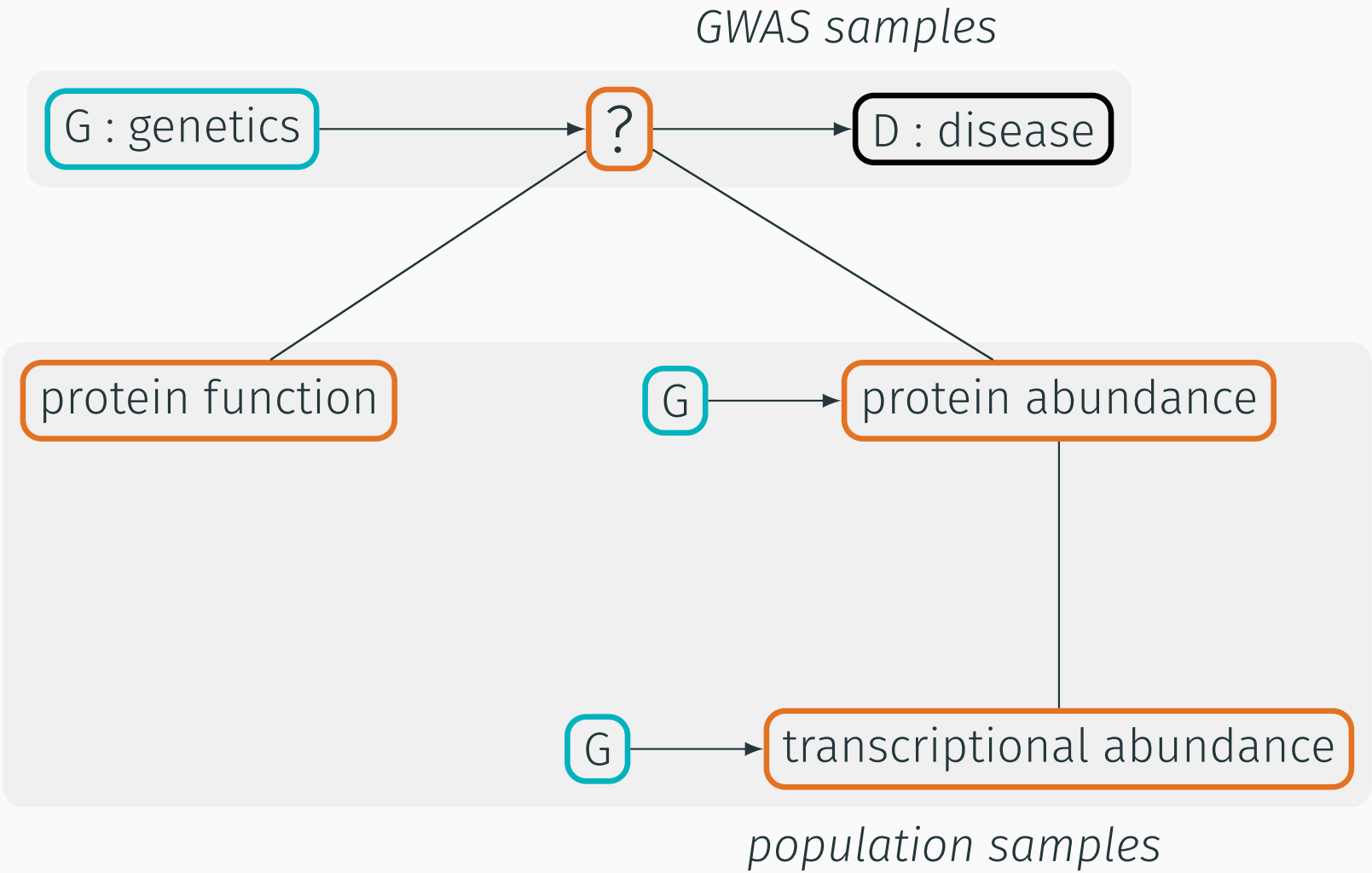
Testing whether a trait is causal for disease



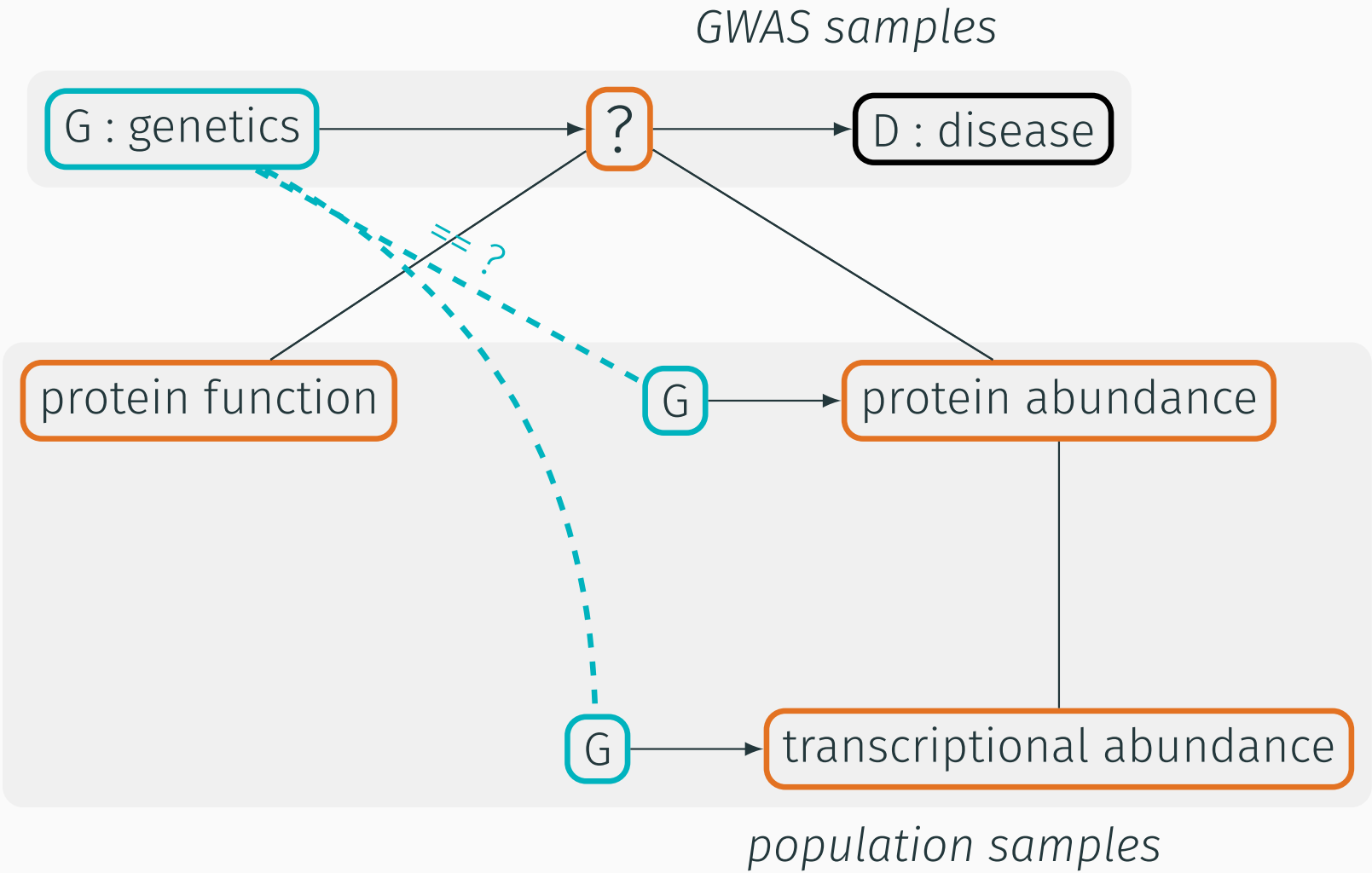
Testing whether a trait is causal for disease



Testing whether a trait is causal for disease



Testing whether a trait is causal for disease

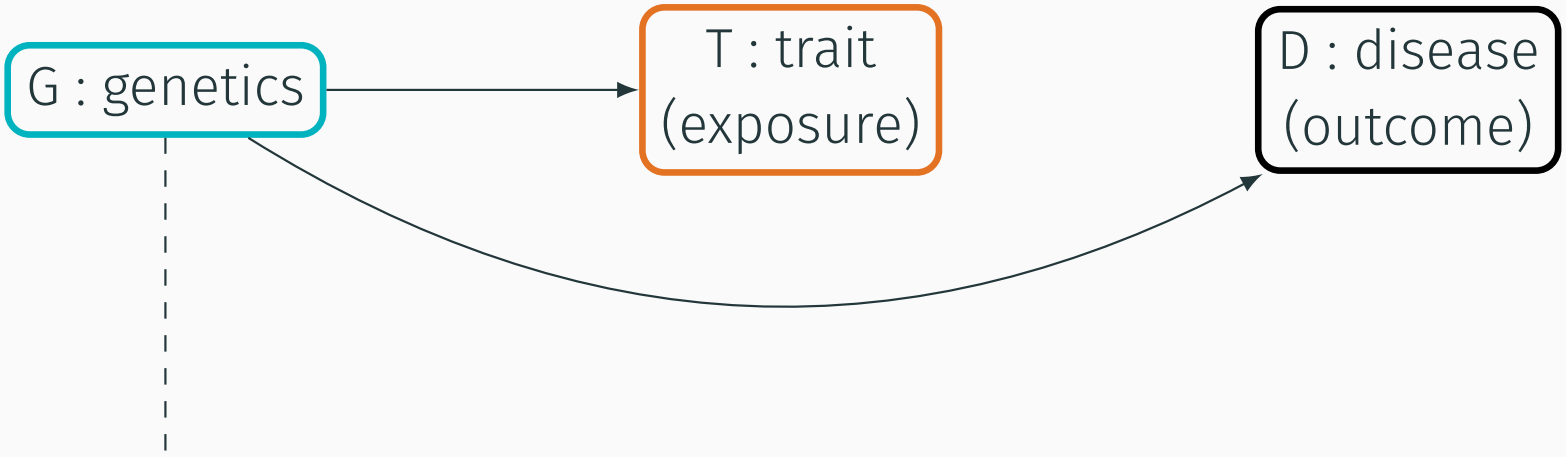


Testing whether an exposure causally affects outcome

Mendelian Randomisation

Testing whether an exposure causally affects outcome

Mendelian Randomisation



Instrumental variable

- potentially multiple IVs

Test significance of effects

$G \rightarrow D$

Particular issues for finding mediating molecular traits

trait/IV not generally known in advance

- Need to use the same data for discovery and testing

May only be one variant affecting molecular trait

Need to distinguish

- molecular variant in LD with disease variant

OR

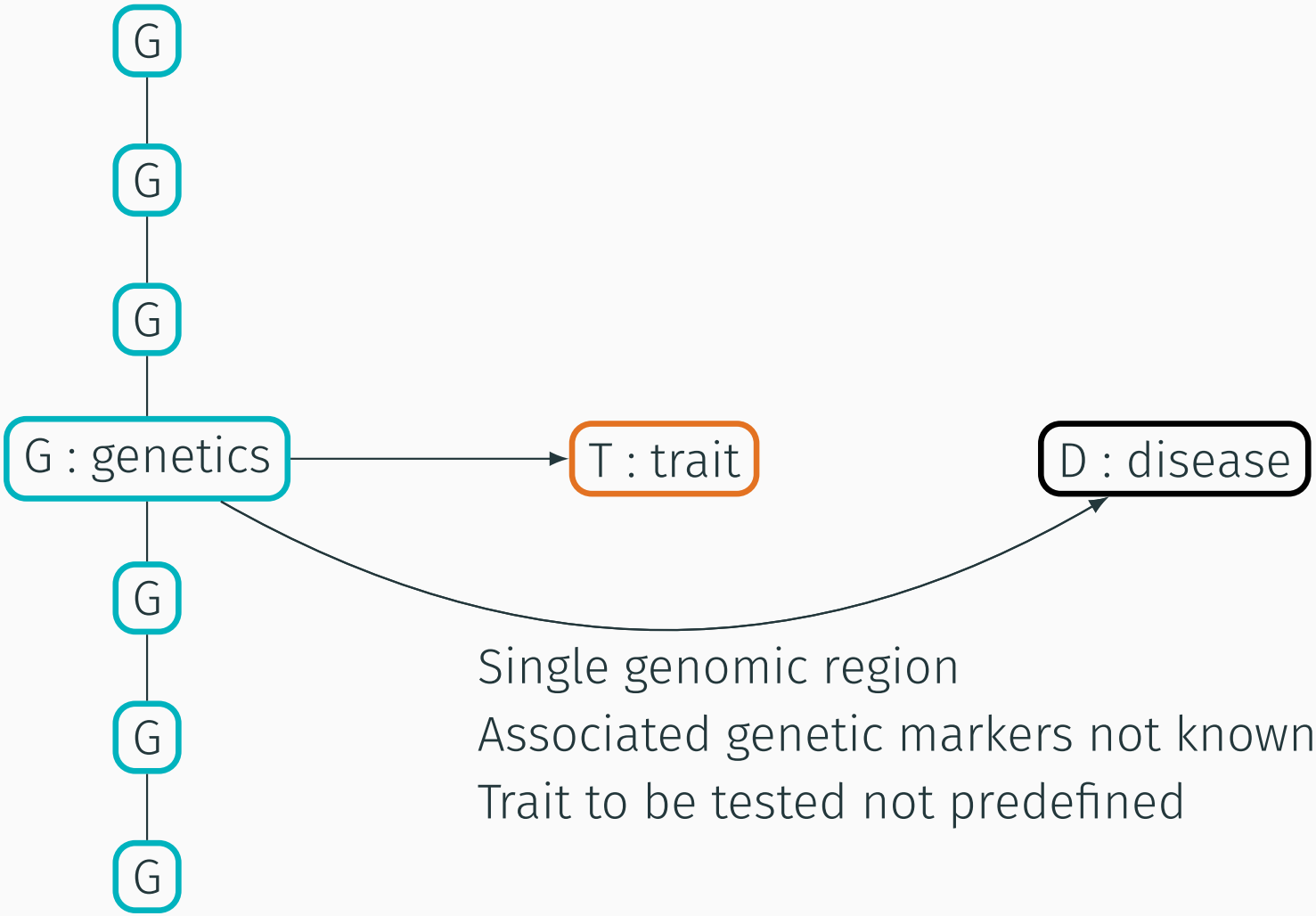
- molecular variant coincident with disease variant

Identifying molecular traits that mediate a disease association

Identifying molecular traits that mediate a disease association

Colocalisation

Genetic discovery **and** test joint relationship



Single genomic region

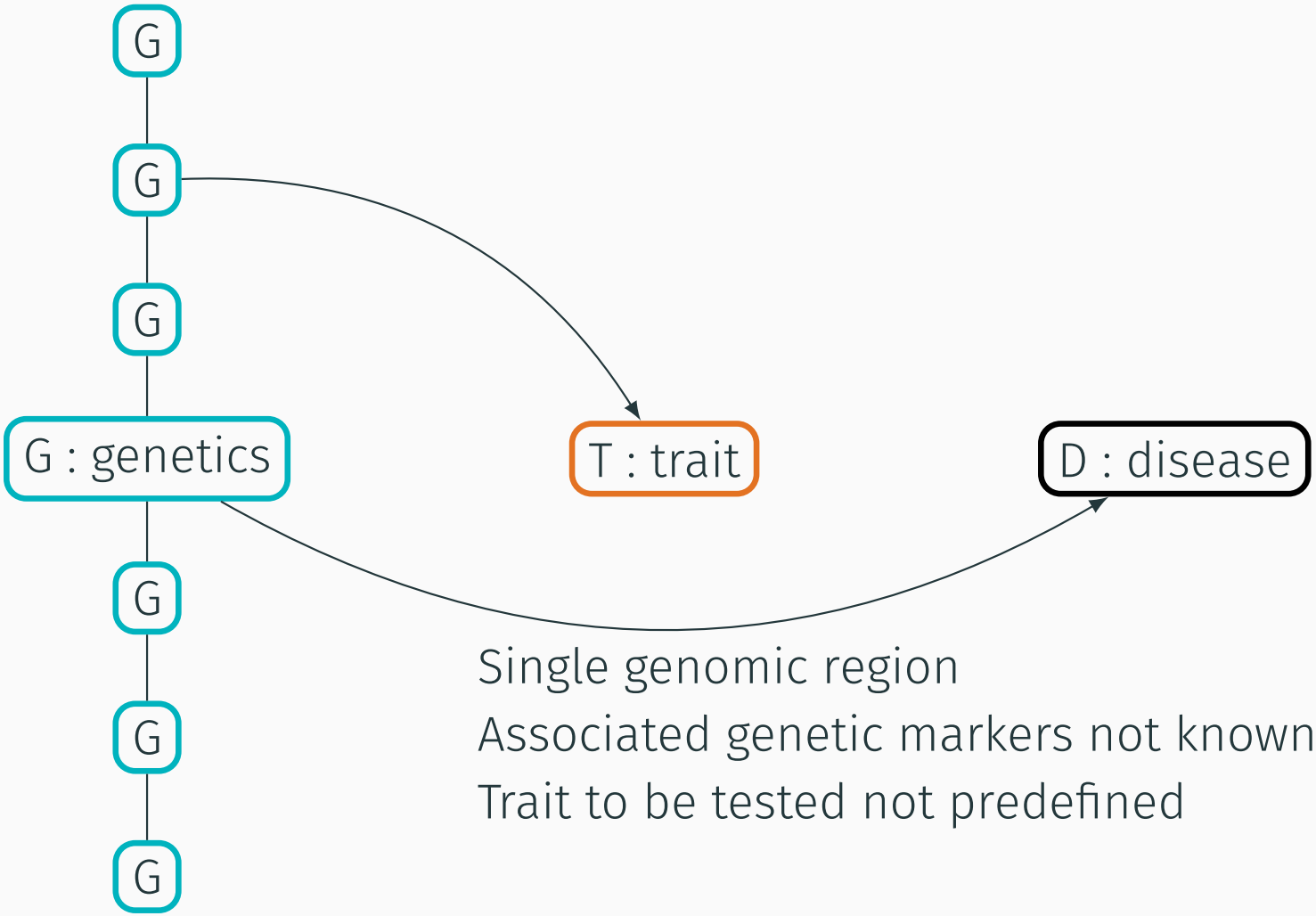
Associated genetic markers not known *a priori*

Trait to be tested not predefined

Identifying molecular traits that mediate a disease association

Colocalisation

Genetic discovery **and** test joint relationship



Single genomic region

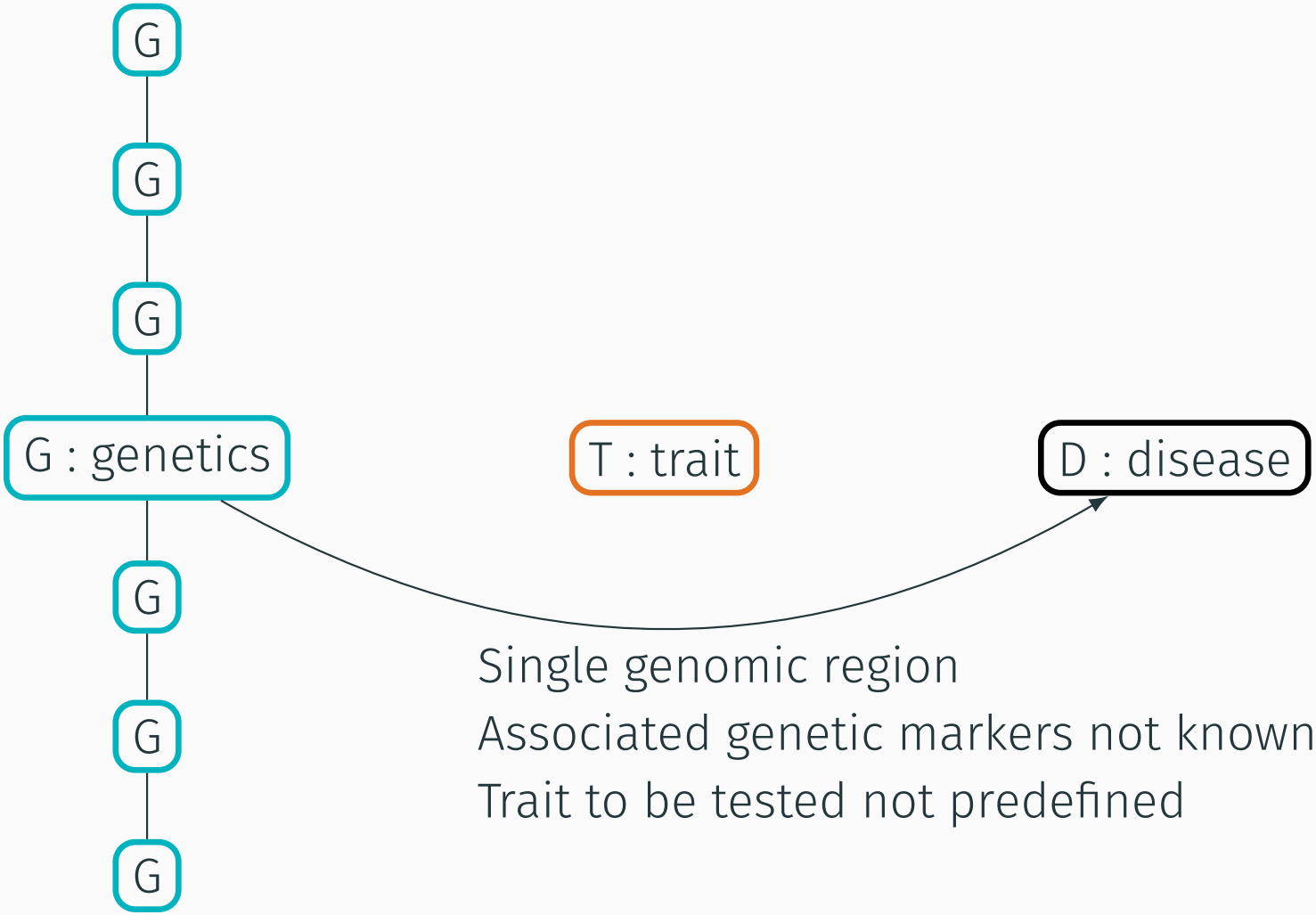
Associated genetic markers not known *a priori*

Trait to be tested not predefined

Identifying molecular traits that mediate a disease association

Colocalisation

Genetic discovery **and** test joint relationship



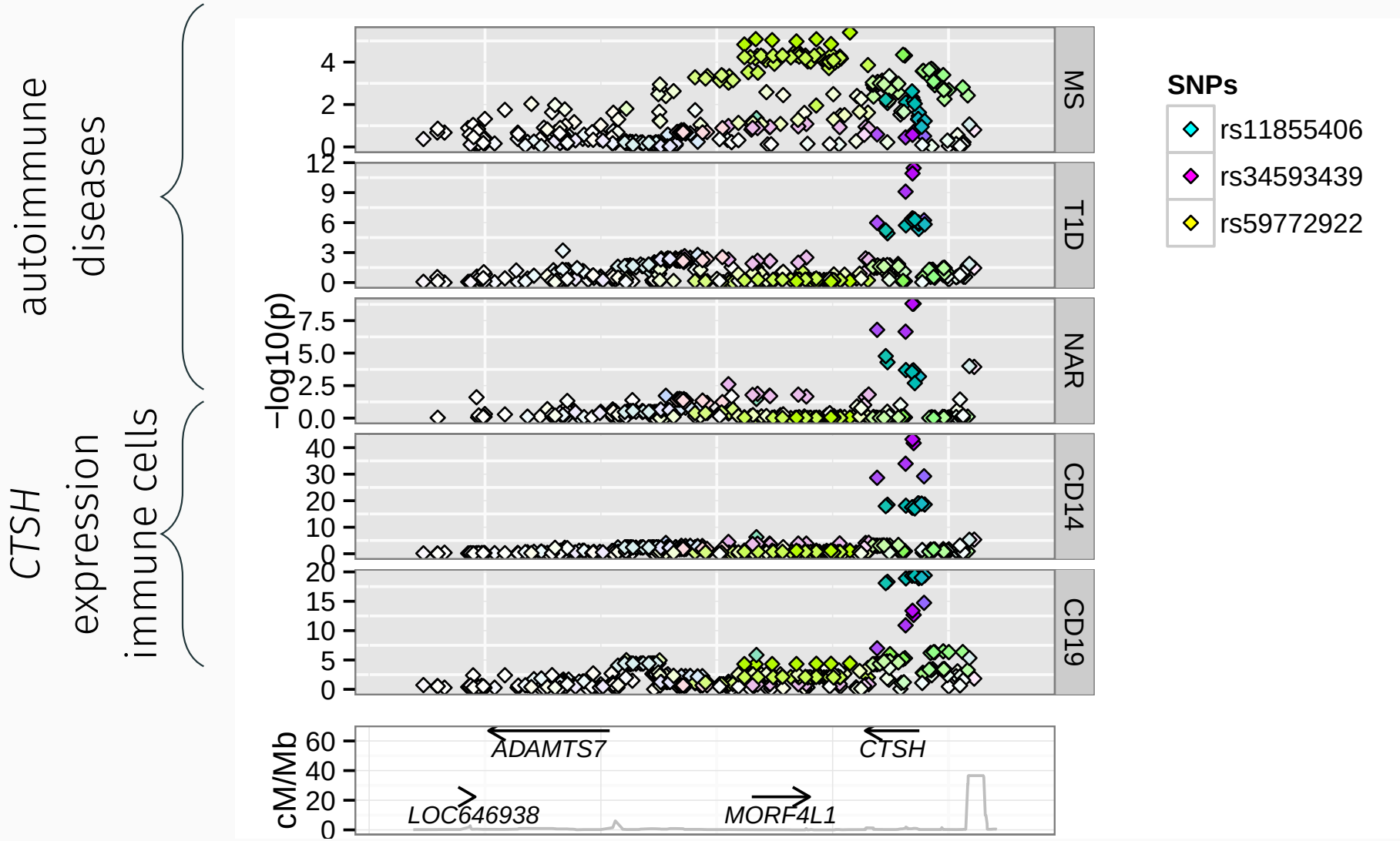
Single genomic region

Associated genetic markers not known *a priori*

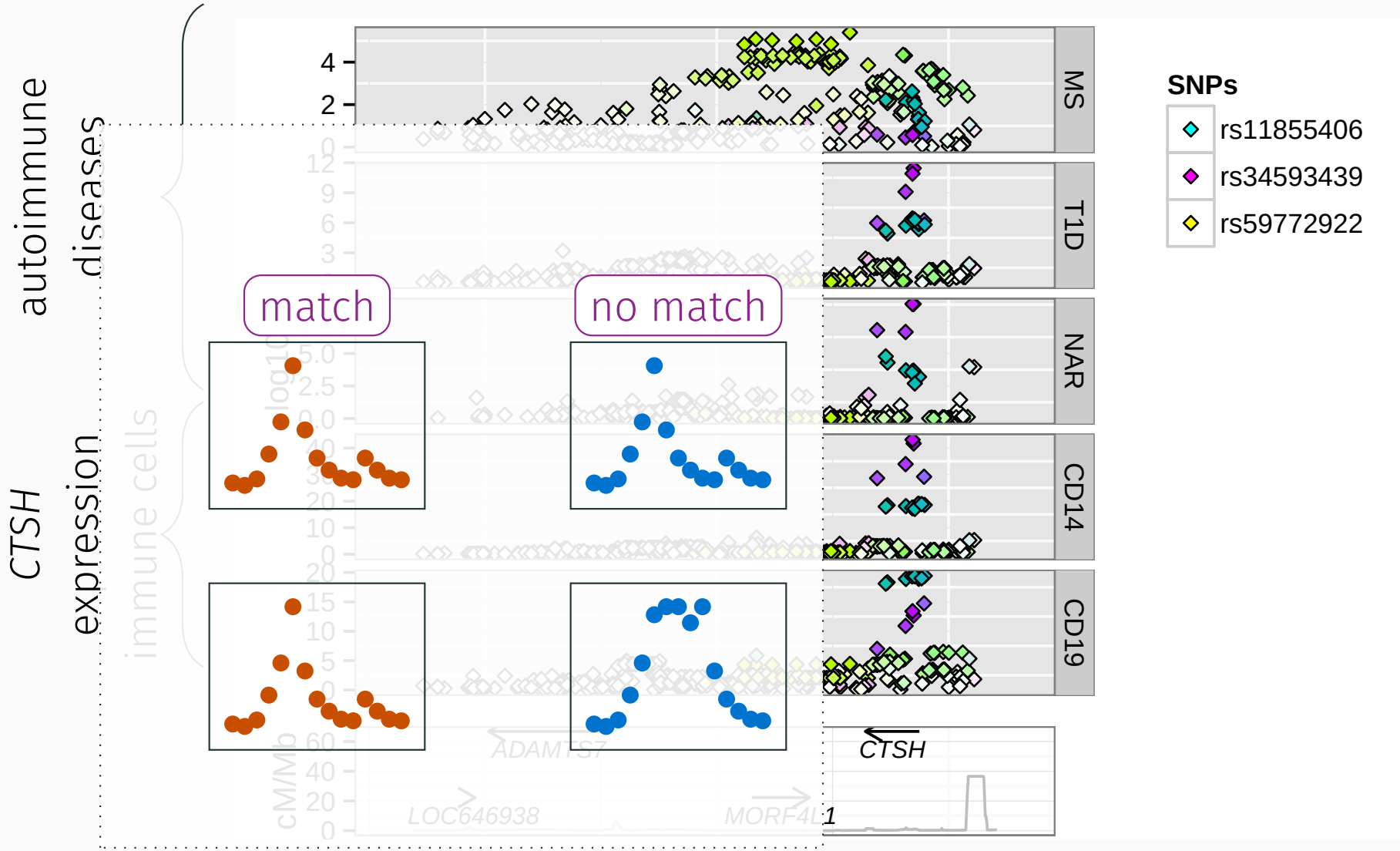
Trait to be tested not predefined

Colocalisation:
the coloc approach

Coloc: at essence a pattern matcher



Coloc: at essence a pattern matcher



Giambartolomei et al, 2014. Guo et al, 2016. chr1swallace/coloc

Coloc enumerates hypotheses

We have a pair of traits.

For a single, LD-defined, genomic region, assume at most one association per trait.

Then exactly 5 possibilities:

H_0 : no association

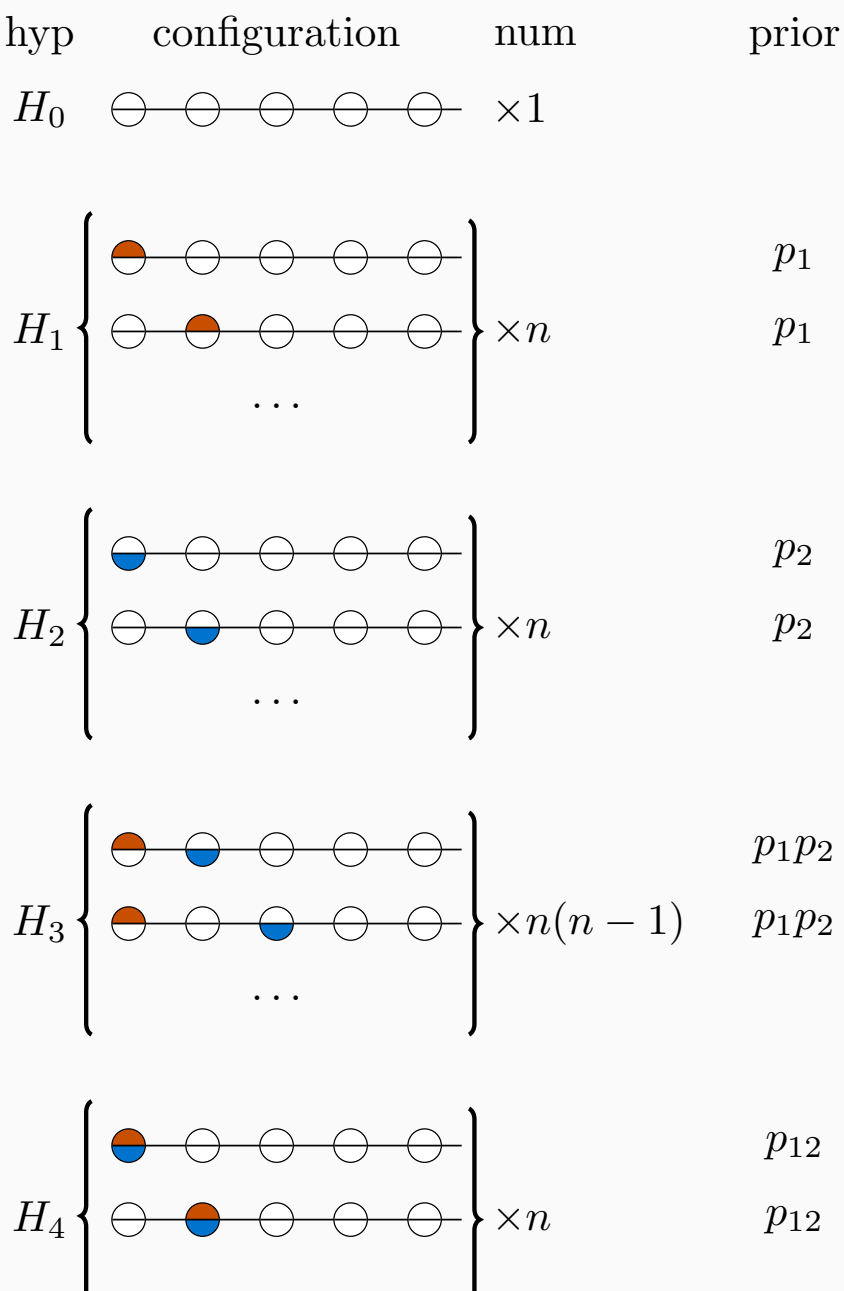
H_1 : association to trait 1 only

H_2 : association to trait 2 only

H_3 : association to both traits, distinct causal variants

H_4 : association to both traits, shared causal variant

Coloc enumerates hypotheses



Values of prior parameters

Original context:

- GWAS of lipid traits (100,000 individuals)
- Liver eQTLs (1000 individuals)

i.e. two large studies, *a priori* biologically relevant to each other

$$p_1 = p_2 = 10^{-4} \qquad p_{12} = 10^{-5}$$

For a 1000 SNP region, this gives

$$P(H_1) = 0.1 \qquad P(H_2) = 0.1$$
$$P(H_3) = 0.01 \qquad P(H_4) = 0.01$$

Giambartolomei et al, PLoS Gen 2014

Review of current practice

Out of 25 papers which used coloc in 2018 ...

... 22 used software default priors

Review of current practice

Out of 25 papers which used coloc in 2018 ...

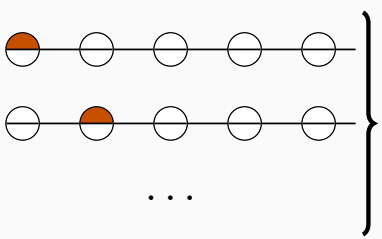
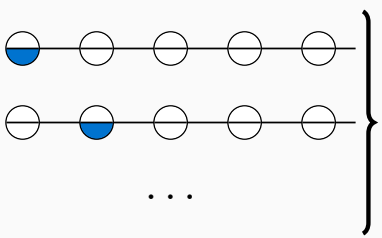
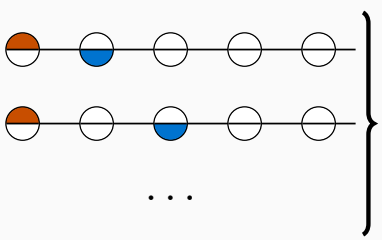
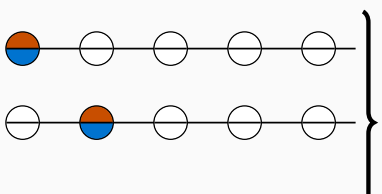
... 22 used software default priors

Varied trait pairs:

GWAS-eQTL	GWAS-meQTL	eQTL-pQTL	GWAS-chromatinQTL
15	3	2	2
GWAS-GWAS	GWAS-molecular QTL	GWAS-pQTL	
1	1	1	

Marginal priors: q_1, q_2

Colocalisation priors

hyp	configuration	num	prior
H_0		$\times 1$	
H_1		$\times n$	p_1 p_1
H_2		$\times n$	p_2 p_2
H_3		$\times n(n-1)$	p_1p_2 p_1p_2
H_4		$\times n$	p_{12} p_{12}

Marginal prior

$q_1 = p_1 + p_{12}$

$q_2 = p_2 + p_{12}$

Prior explorer for coloc

Input: Type of Per SNP priors

select one

Raw
Marginal/Conditional

Input: Parameter values

Number of SNPs in region 1000

p12 0.000001

p1 0.0001

p2 0.0001

nsnps > 0
p1 * p2 < p12 < min(p1,p2)
0 < p1 < 1/nsnps
0 < p2 < 1/nsnps

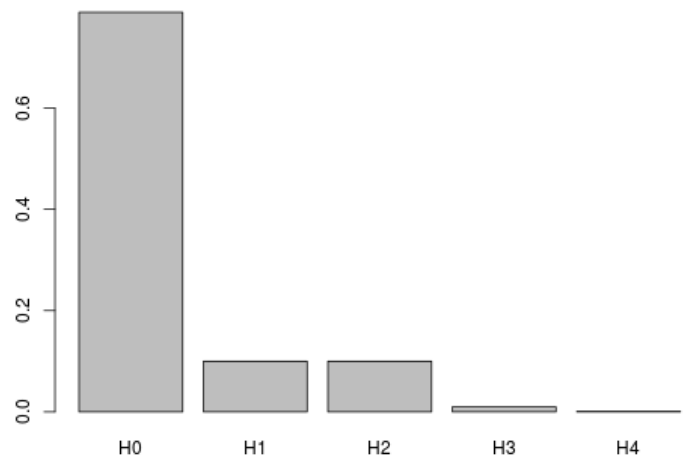
Result

Per-SNP priors

p12	p1	p2	q1|2	q1	q2
1.0E-06	1.0E-04	1.0E-04	9.9E-03	1.0E-04	1.0E-04

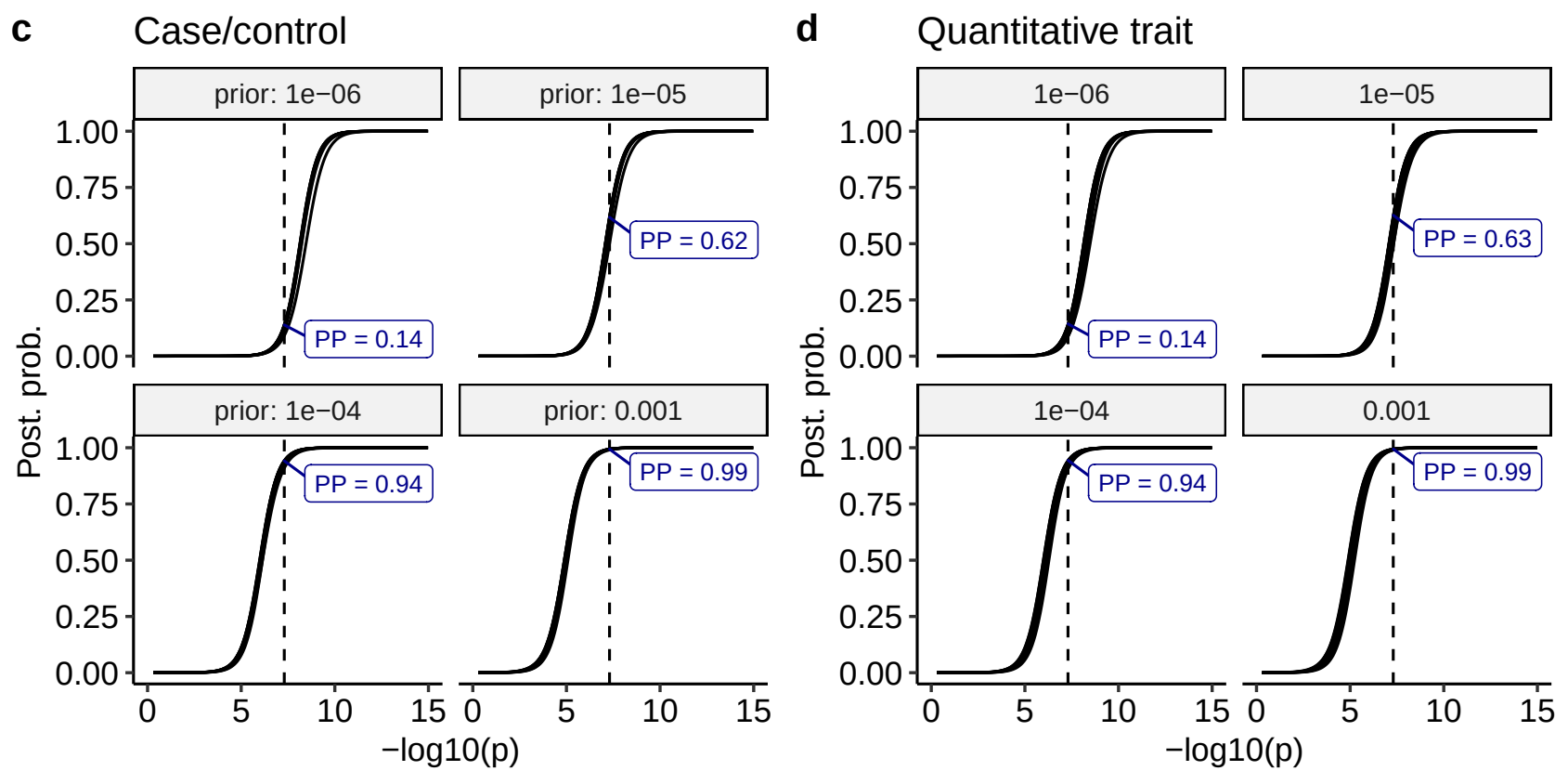
Per-hypothesis priors

H0	H1	H2	H3	H4
0.789	0.100	0.100	0.010	0.001



https://chr1swallace.shinyapps.io/coloc-priors/

Pragmatic prior for single trait



Empirical prior for single trait

Largest relevant GWAS disease studies:

eg IBD, 60,000 subjects[†]
$\sim$ 200 "hits" from 2 million common SNPs $\rightarrow$

$$q_i = \frac{200}{2,000,000} = \frac{1}{10,000}$$

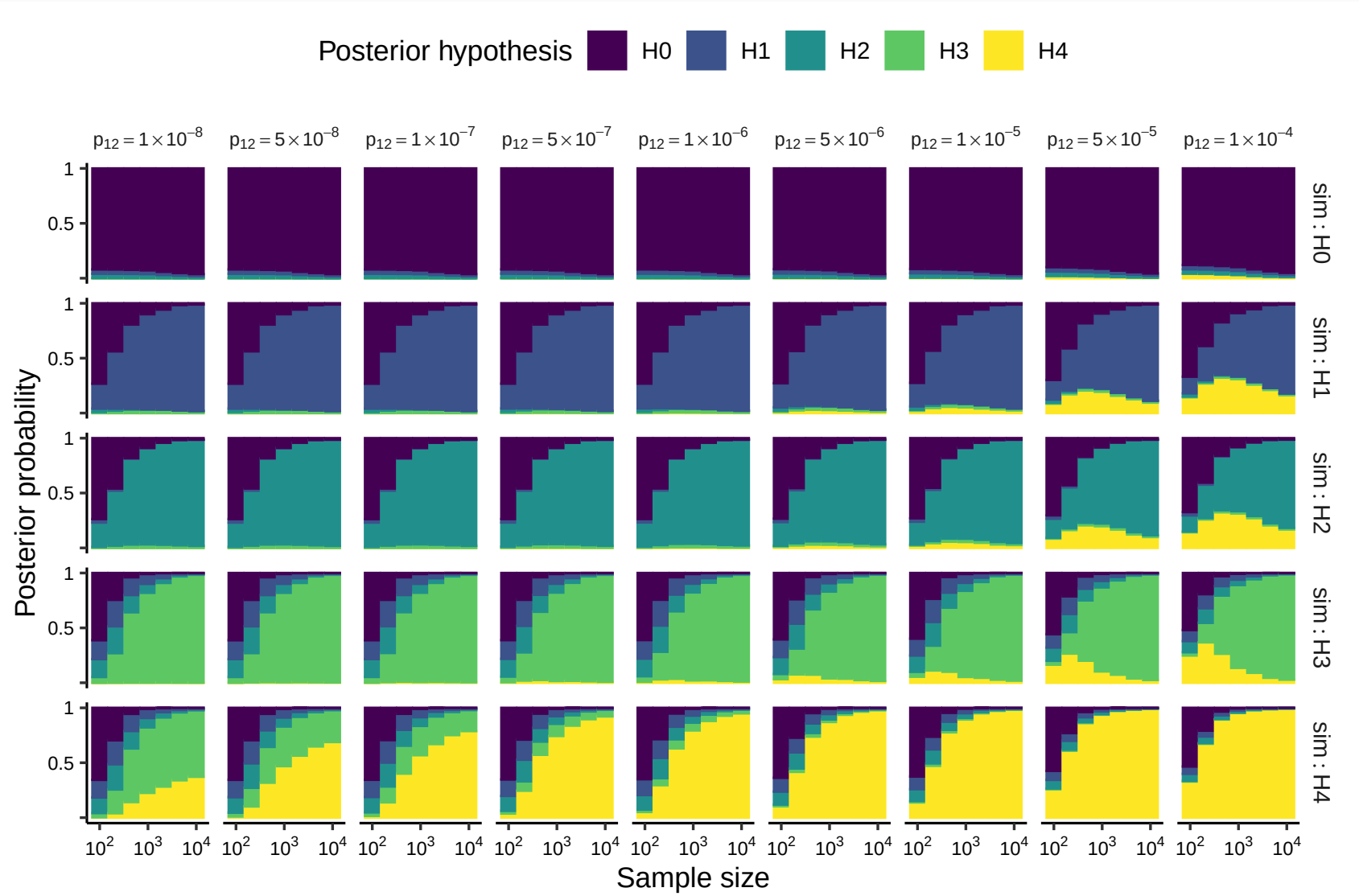
GTeX whole blood:

Fraction of SNPs in 1 mb window around TSS that are genome-wide significant $\approx \frac{1}{10,000}$

[†] *de Lange et al, Nat Genet 2017*

Joint prior p_{12}

Pragmatic p_{12}

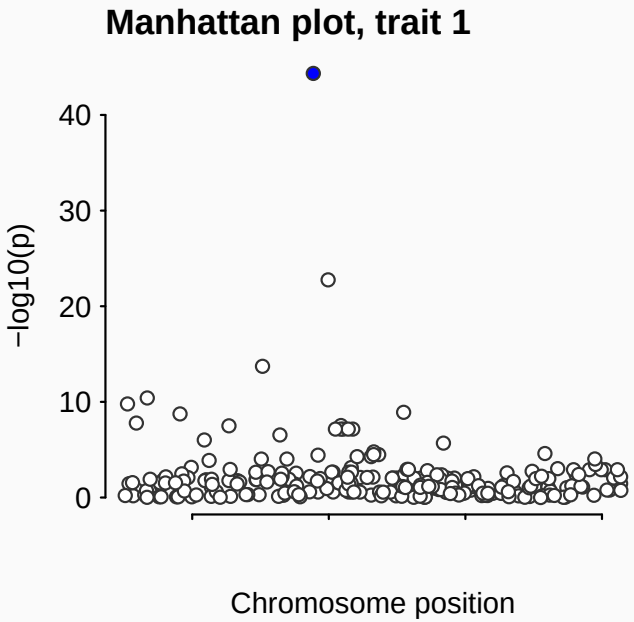


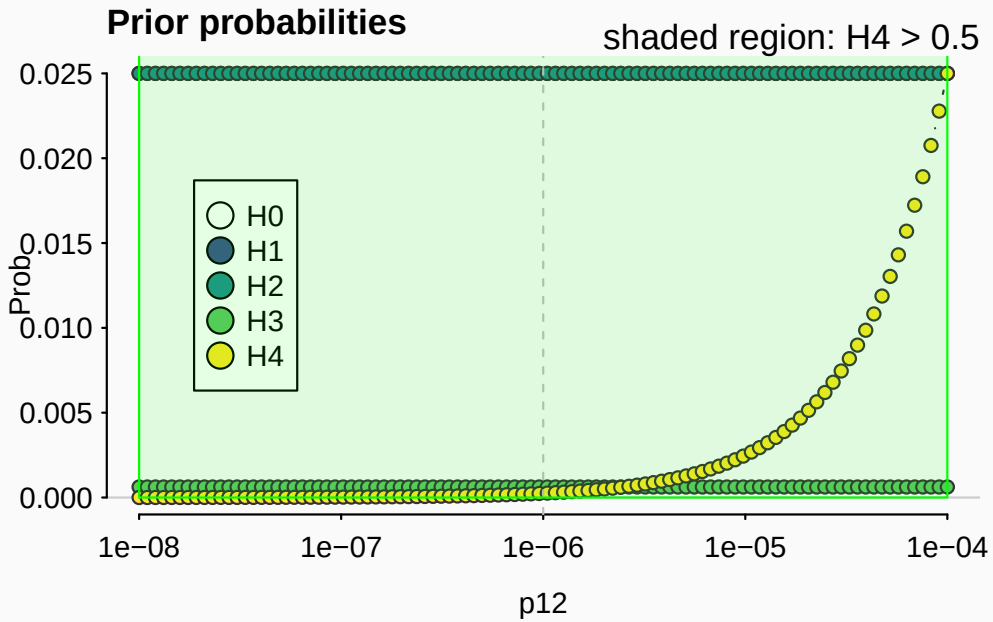
Empirical prior for p_{12}

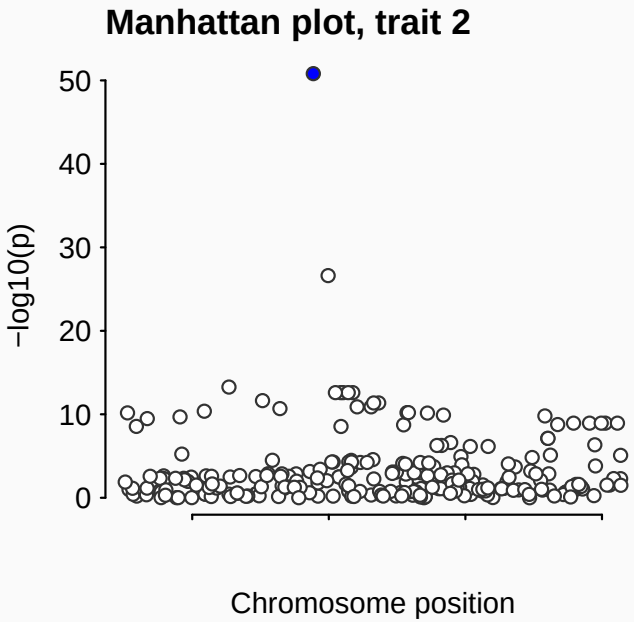
Empirical prior for p_{12}

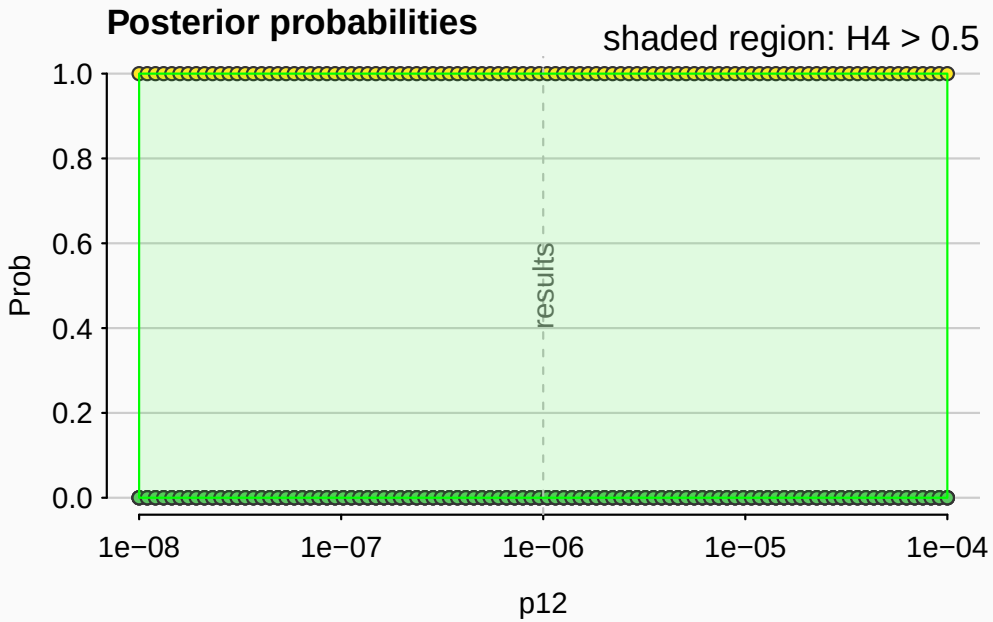
?

Sensitivity analysis

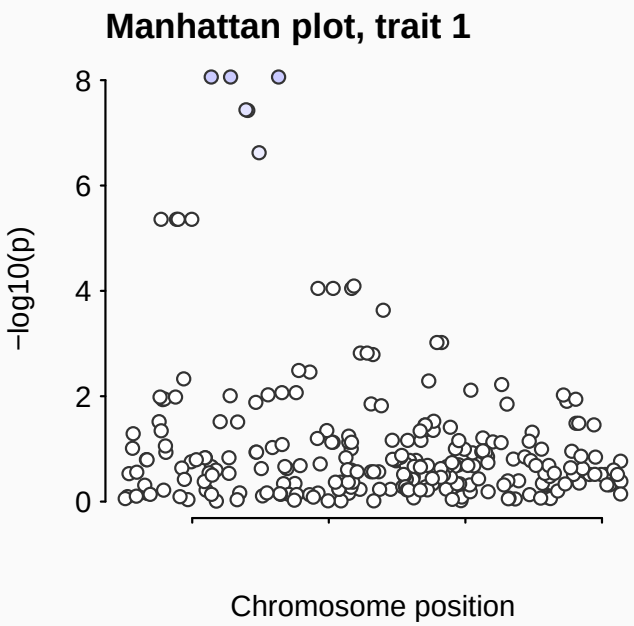


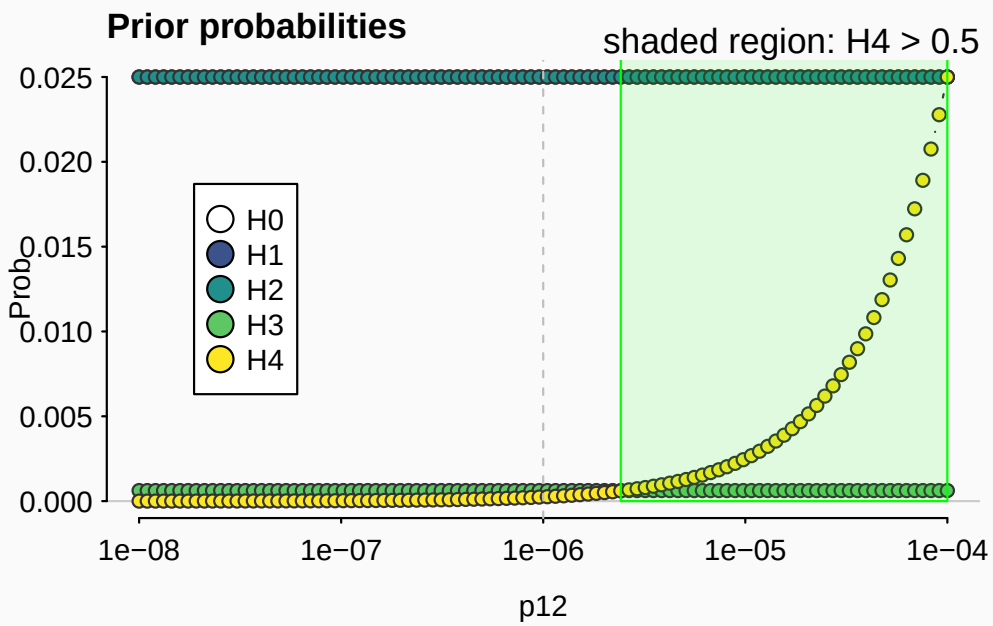


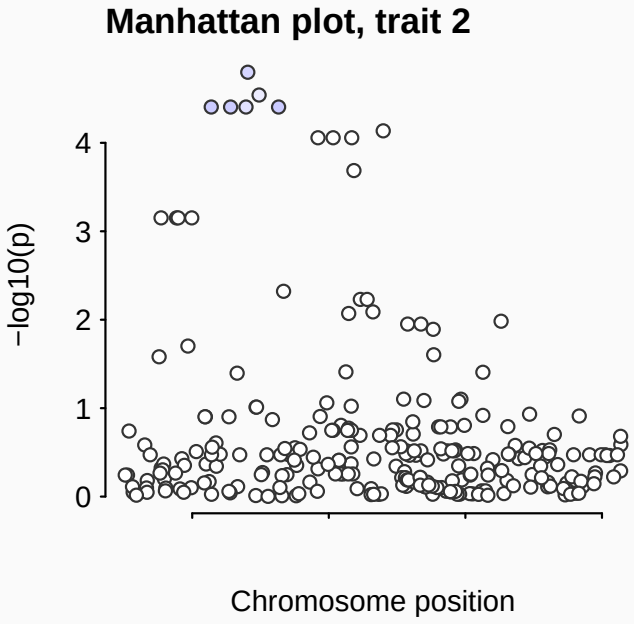


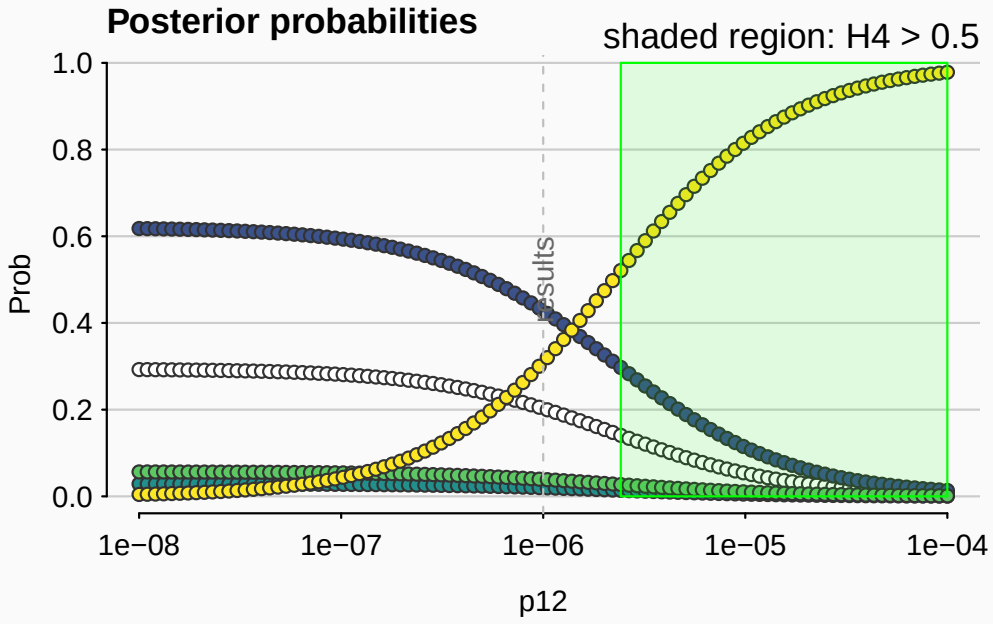


Sensitivity analysis



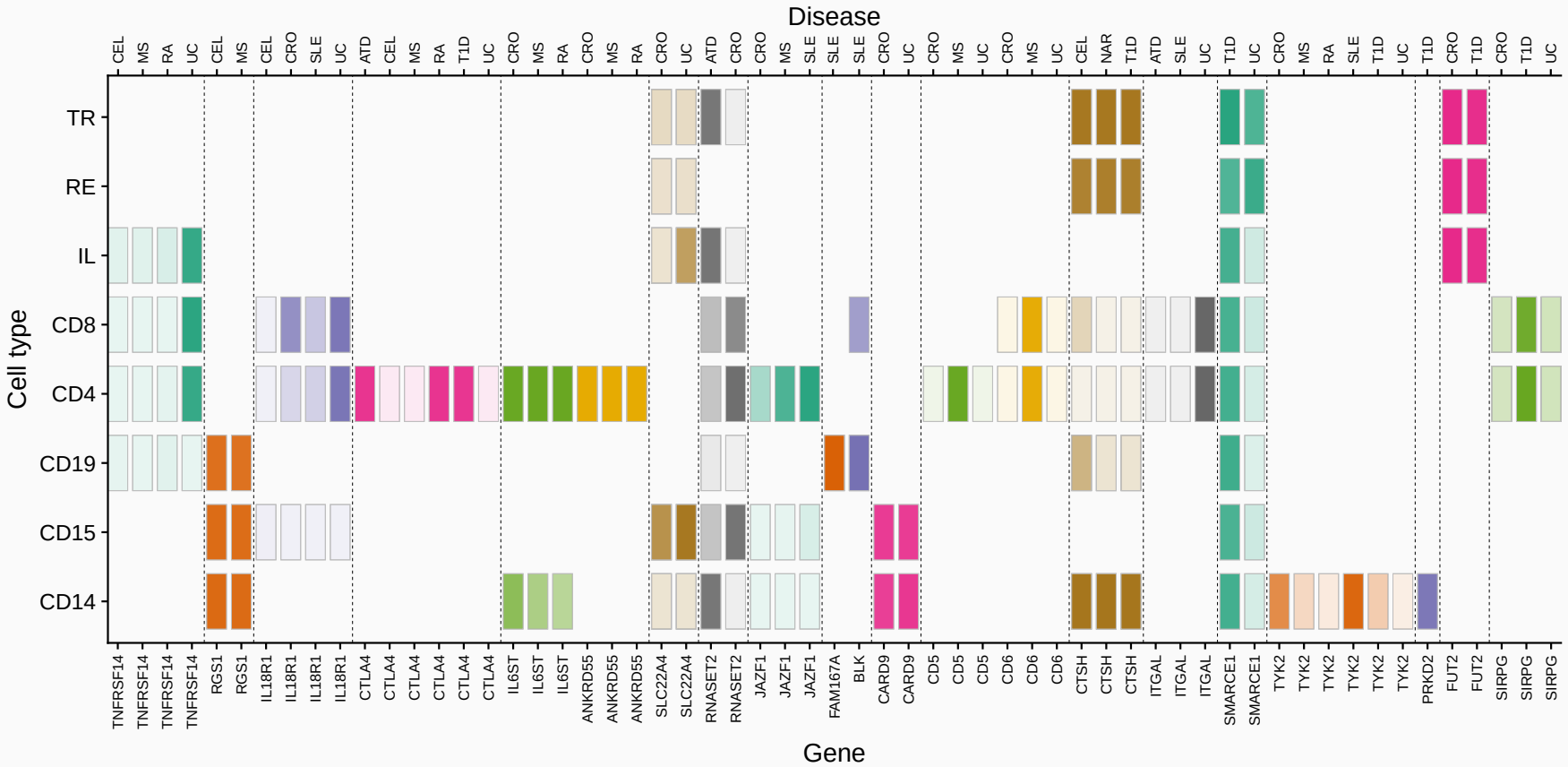






Transcripts colocalising with immune mediated disease risk

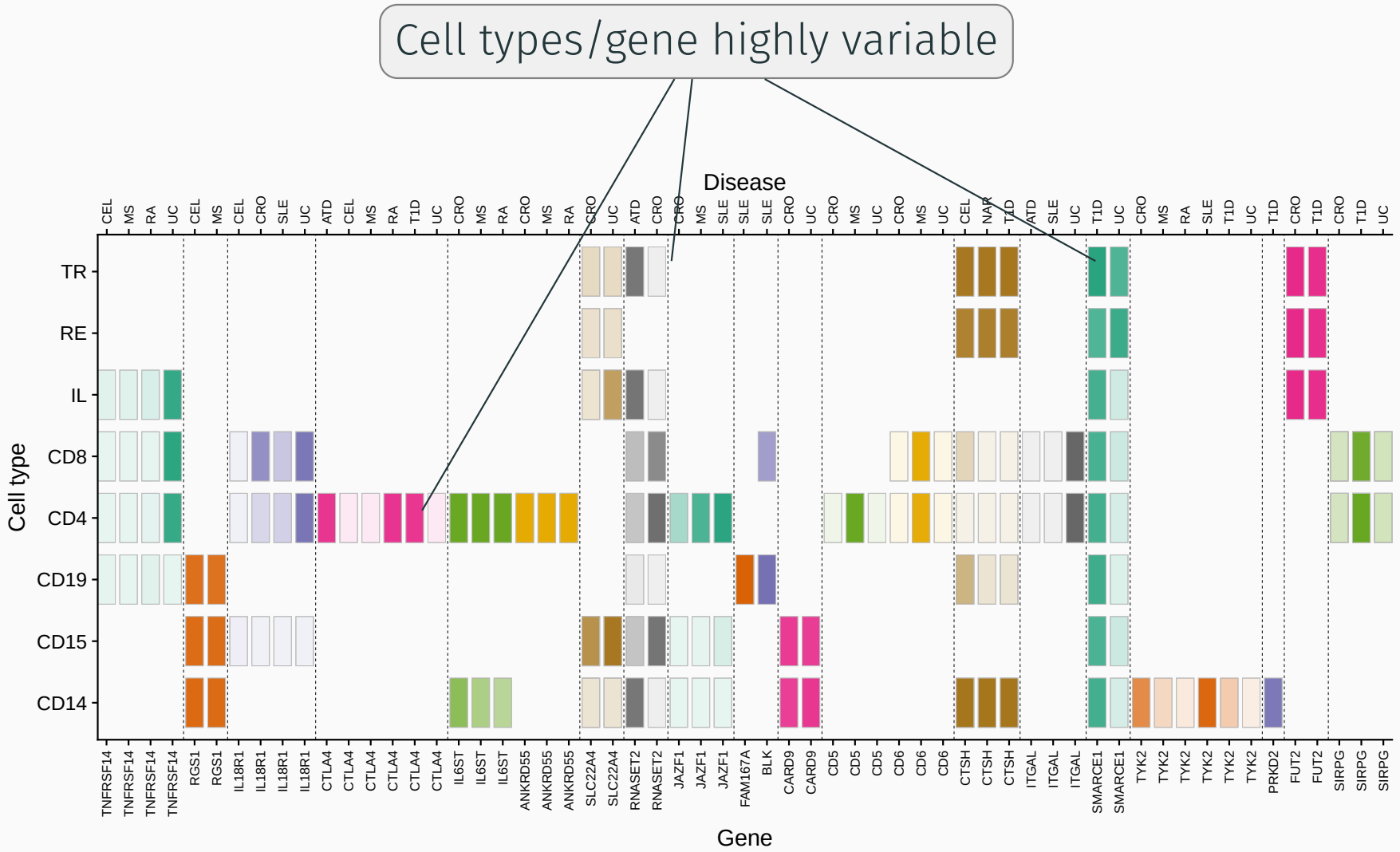
Colocalisation of eQTL (8 cell types) vs 11 diseases



All genes with $P(H_4) > 0.9$ in at least one cell type/disease shown.

Momozawa et al, Nat Genet 2018

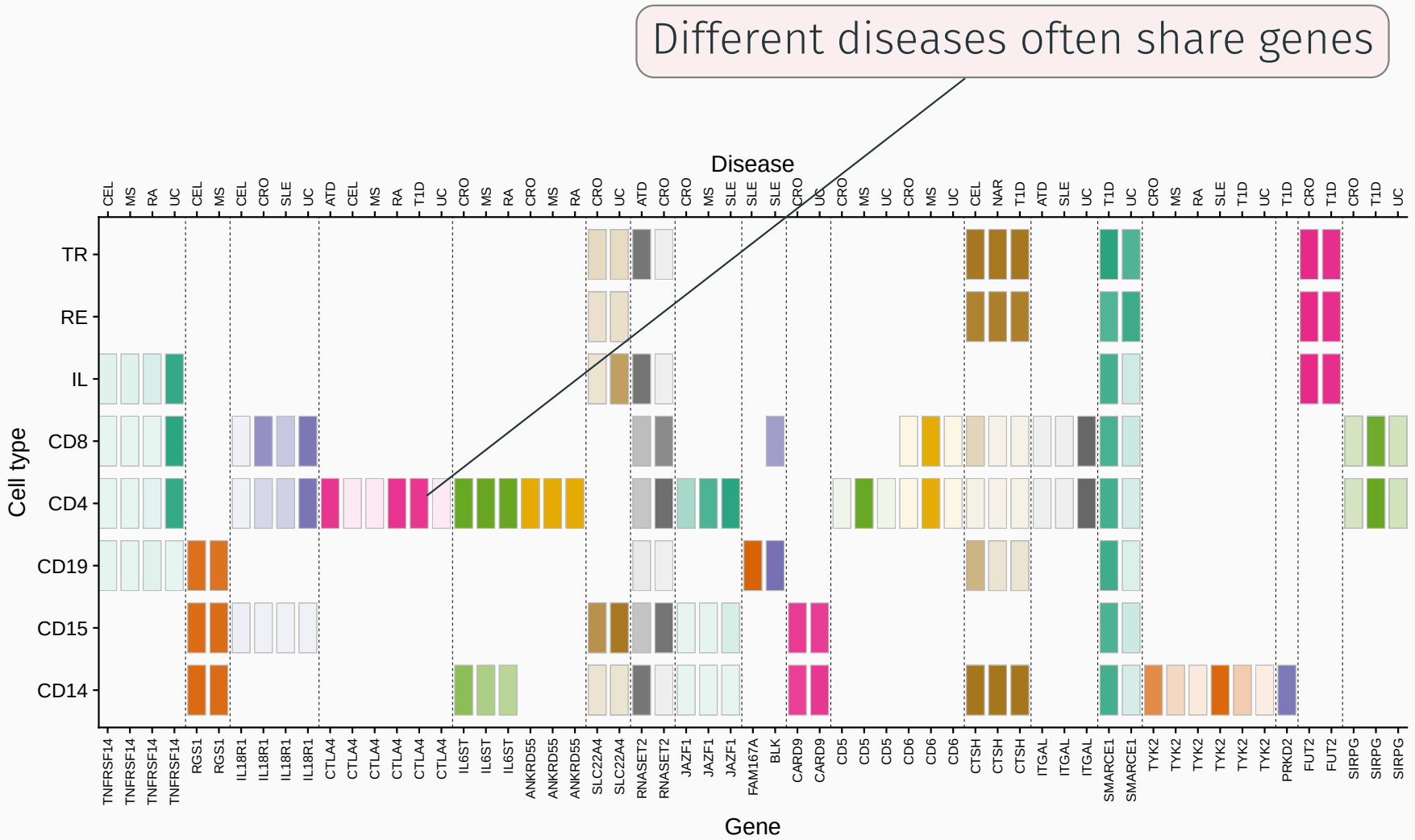
Colocalisation of eQTL (8 cell types) vs 11 diseases



All genes with $P(H_4) > 0.9$ in at least one cell type/disease shown.

Momozawa et al, Nat Genet 2018

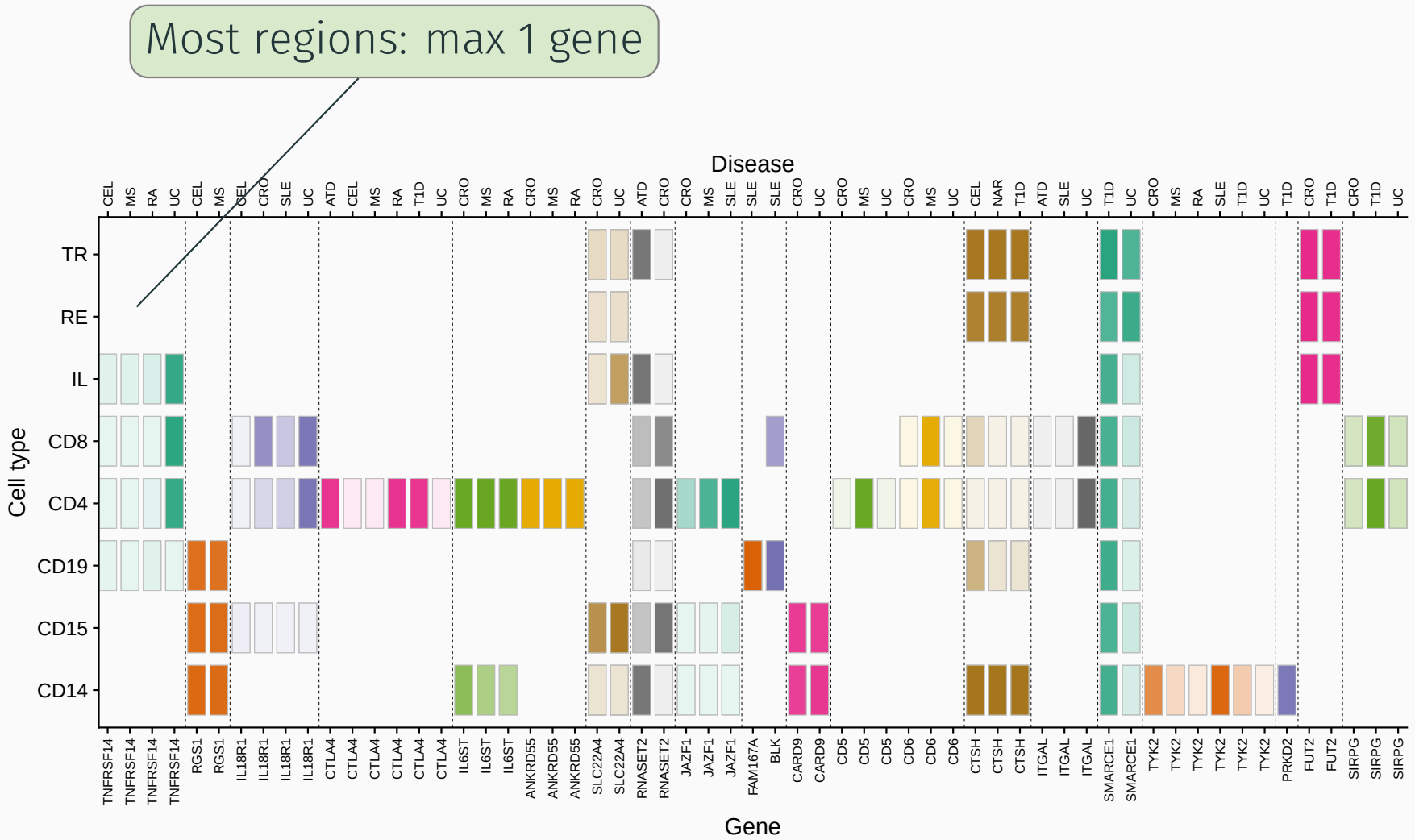
Colocalisation of eQTL (8 cell types) vs 11 diseases



All genes with $P(H_4) > 0.9$ in at least one cell type/disease shown.

Momozawa et al, Nat Genet 2018

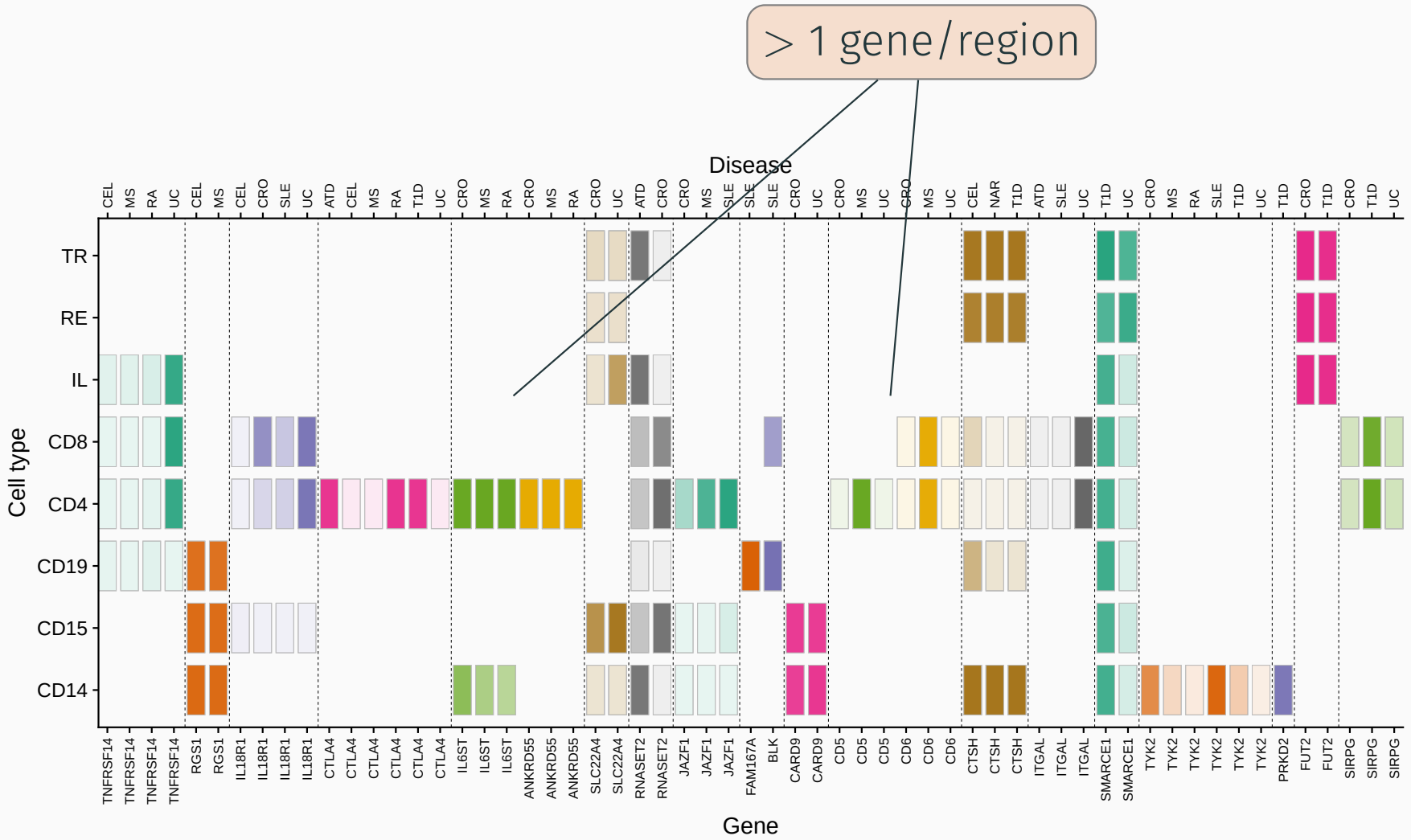
Colocalisation of eQTL (8 cell types) vs 11 diseases



All genes with $P(H_4) > 0.9$ in at least one cell type/disease shown.

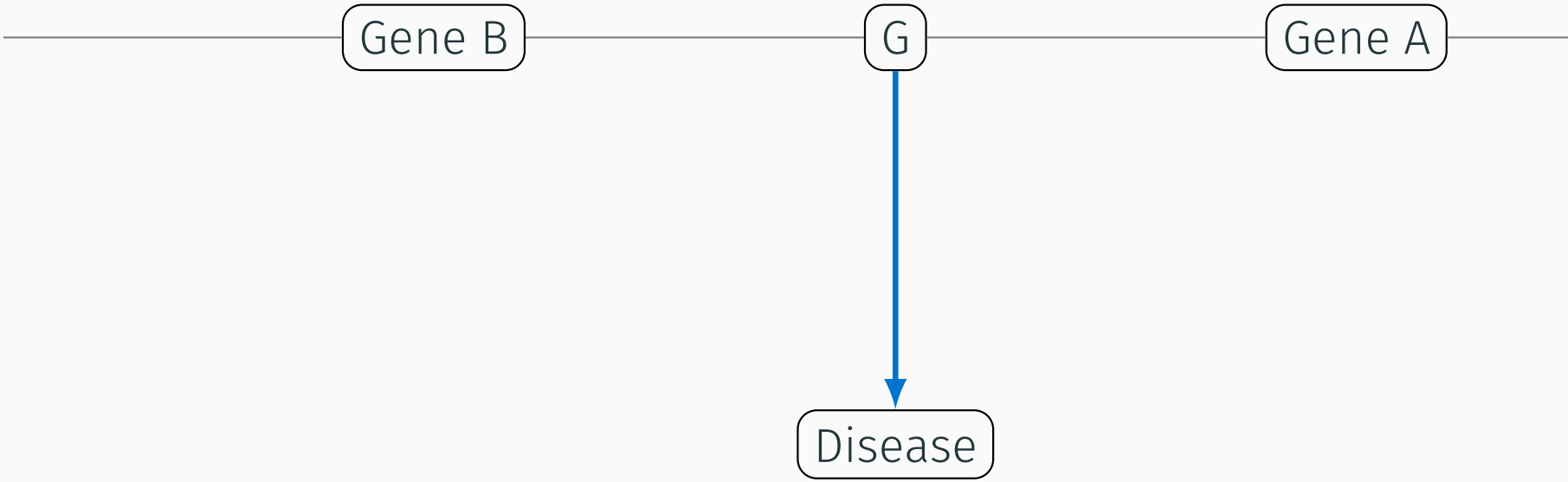
Momozawa et al, Nat Genet 2018

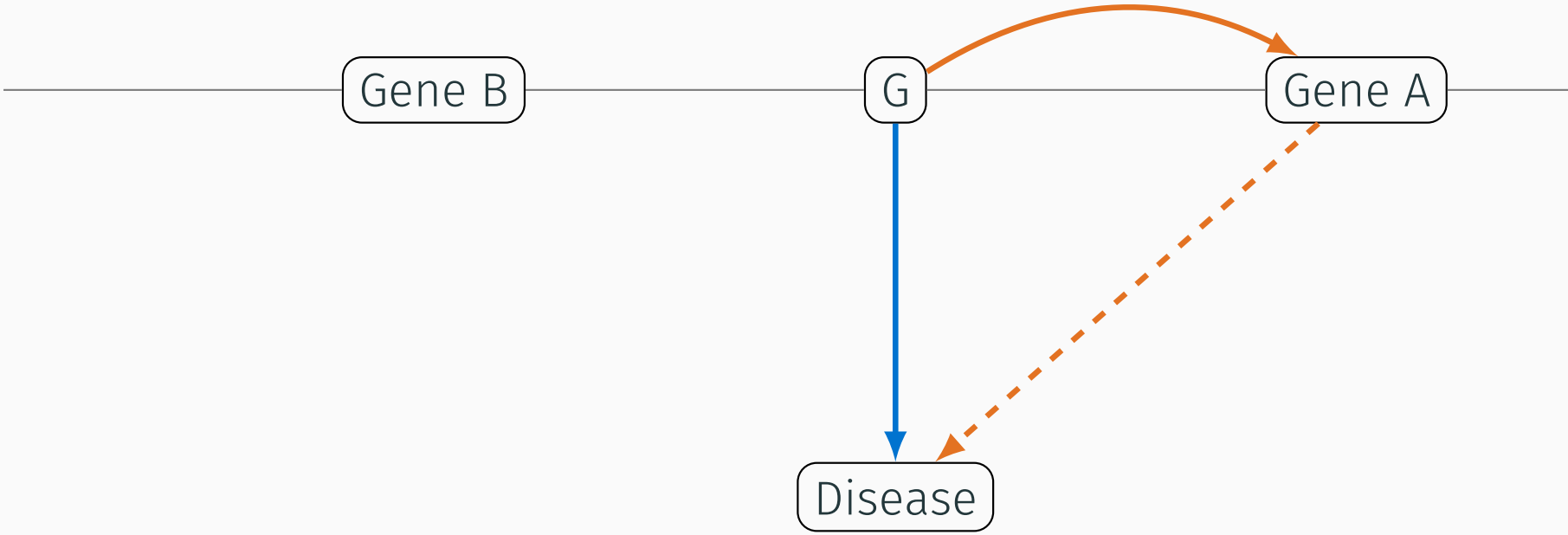
Colocalisation of eQTL (8 cell types) vs 11 diseases

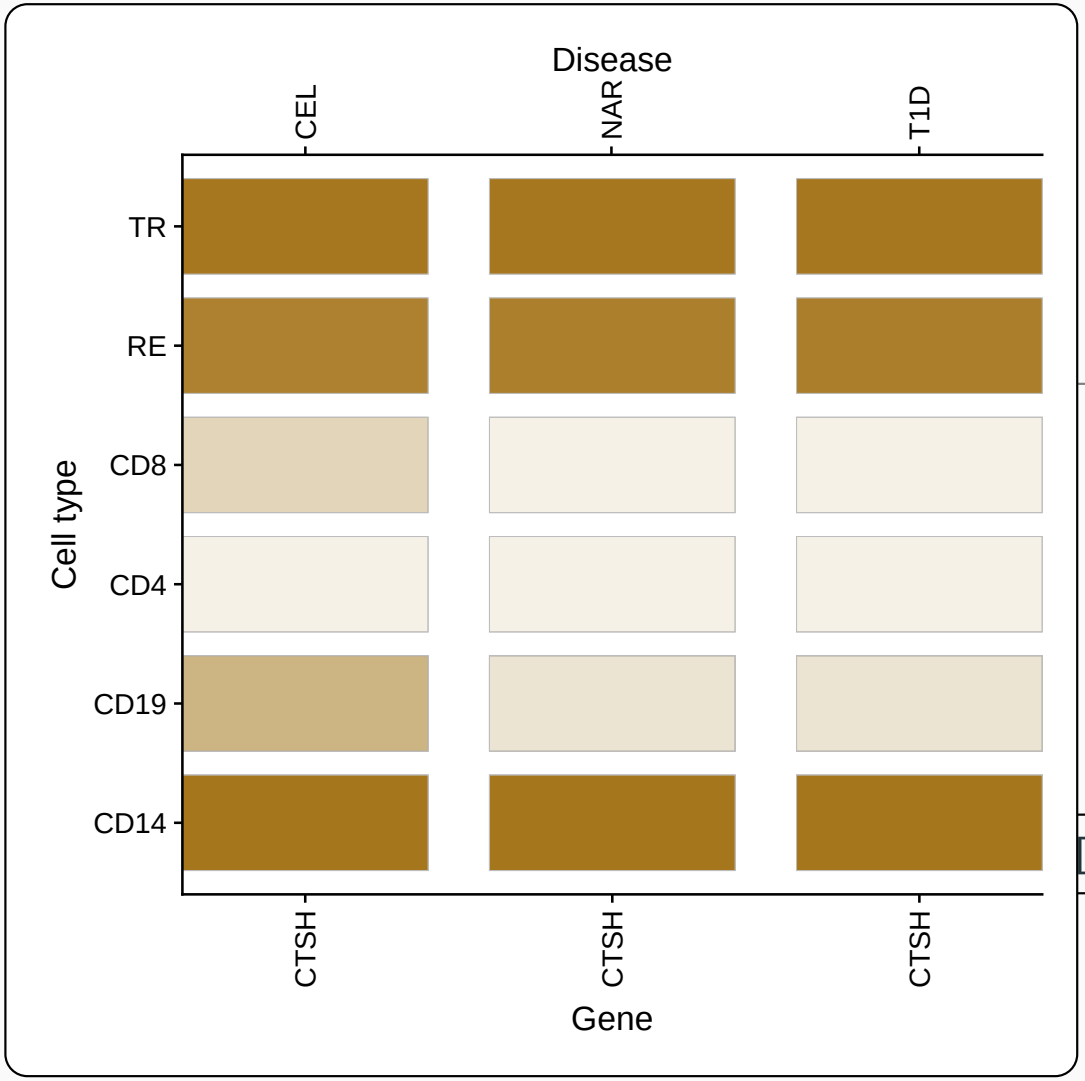


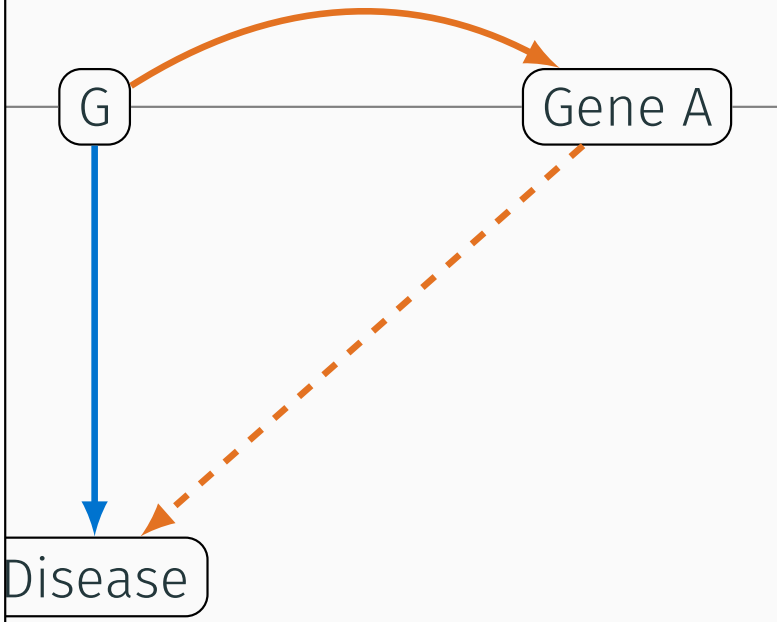
All genes with $P(H_4) > 0.9$ in at least one cell type/disease shown.

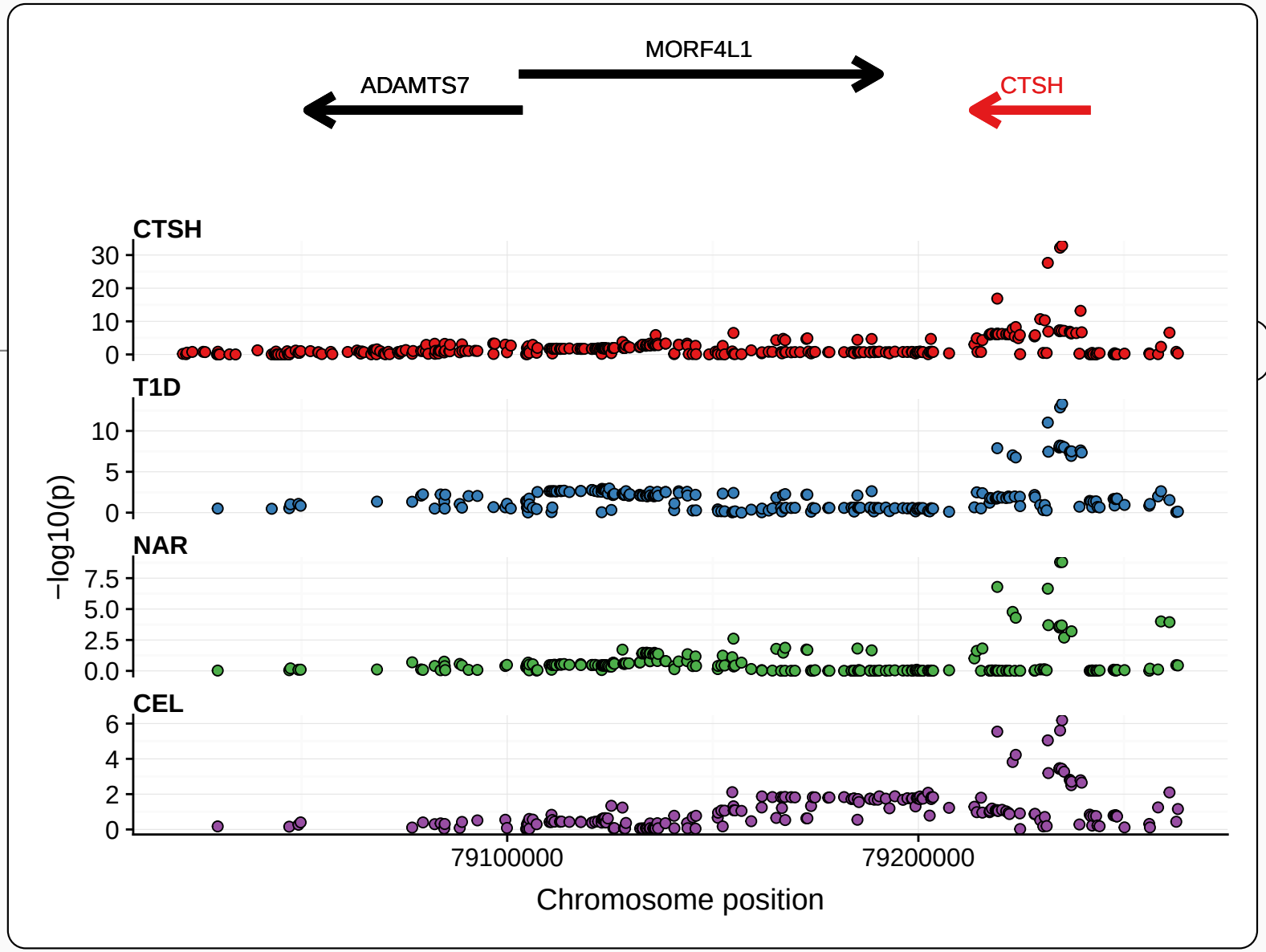
Momozawa et al, Nat Genet 2018

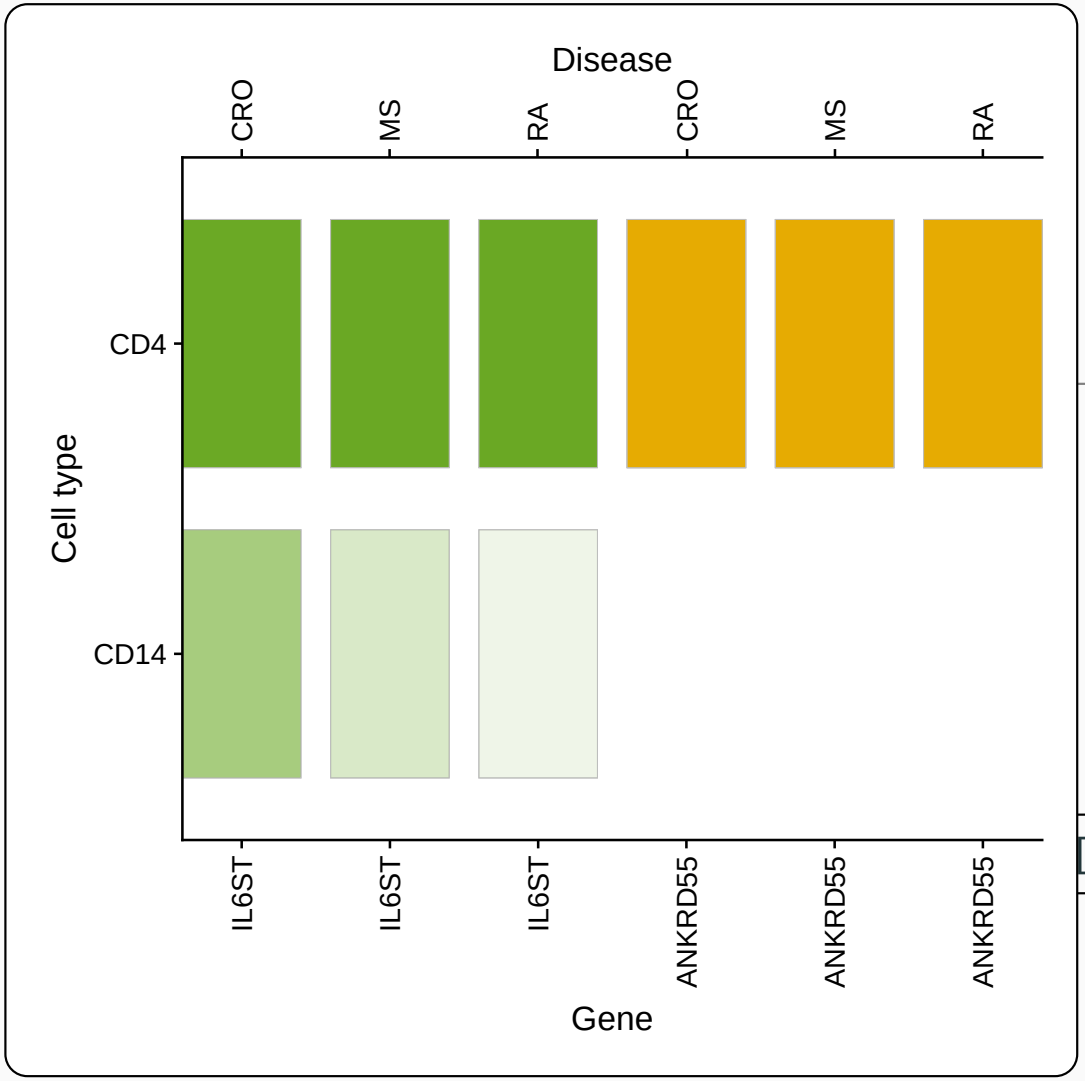

Gene B
G
Gene A
Disease


Gene B
G
Gene A
Disease


Disease
CEL
NAR
T1D
TR
RE
CD8
CD4
CD19
CD14
Cell type
CTSH
CTSH
CTSH
Gene


G
Gene A
Disease

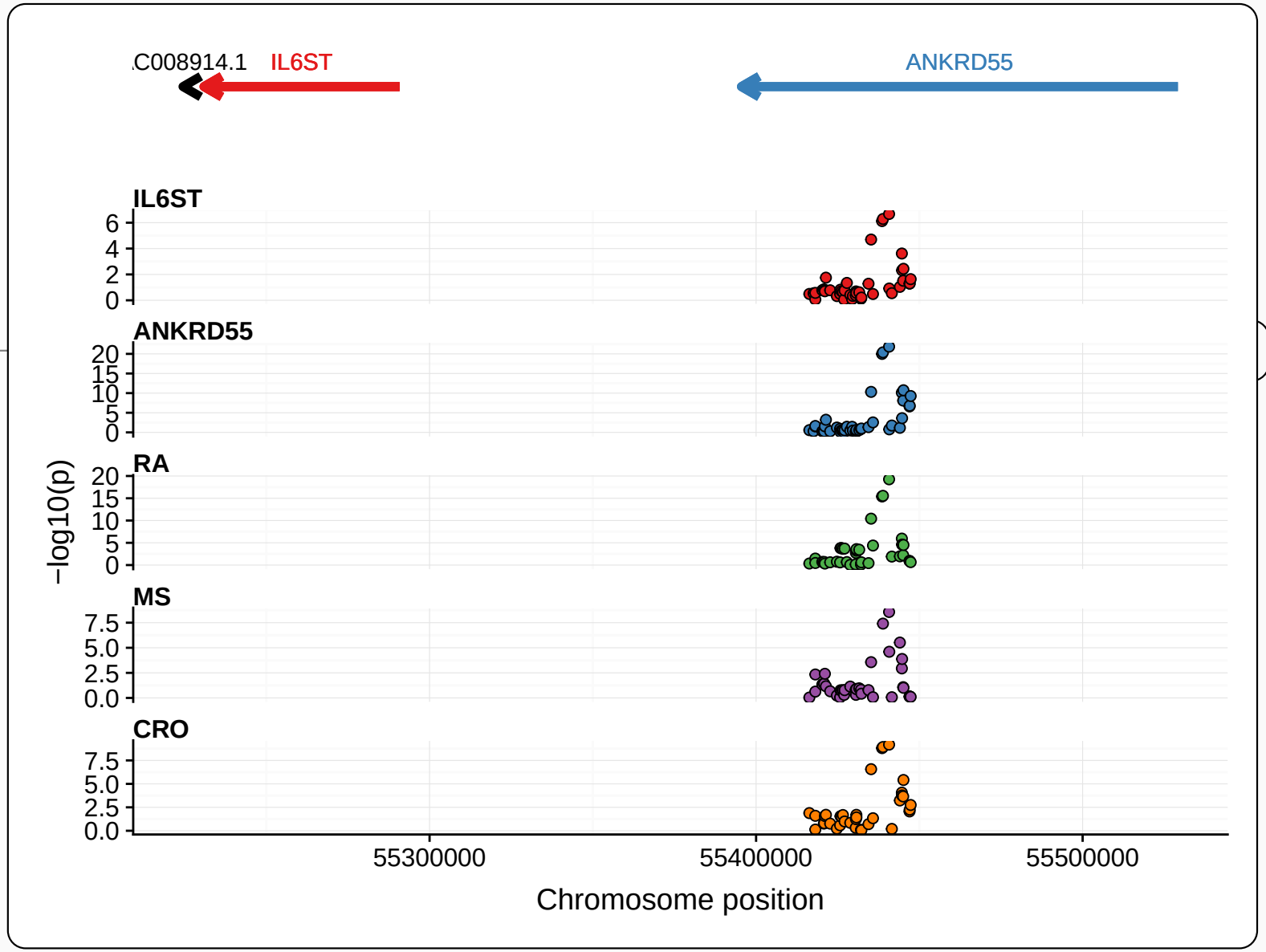

MORF4L1
ADAMTS7
CTSH
CTSH
30
20
10
0
T1D
10
5
0
NAR
7.5
5.0
2.5
0.0
CEL
6
4
2
0
−log10(p)
79100000
79200000
Chromosome position

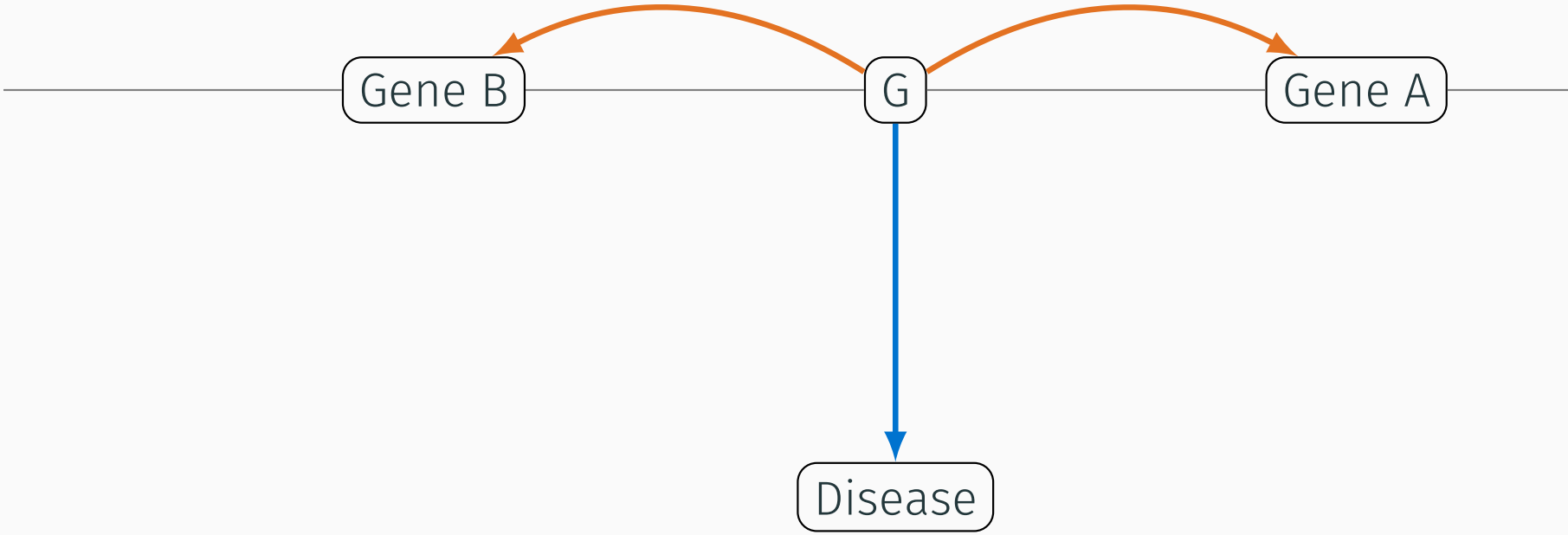


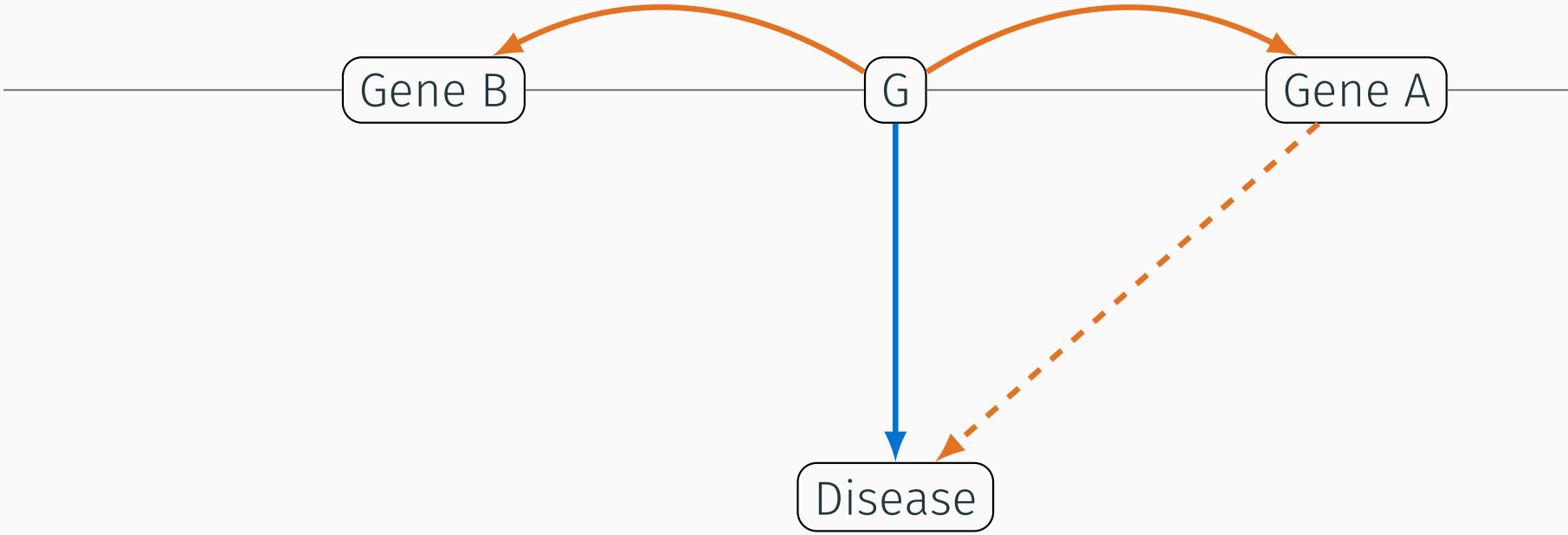
G

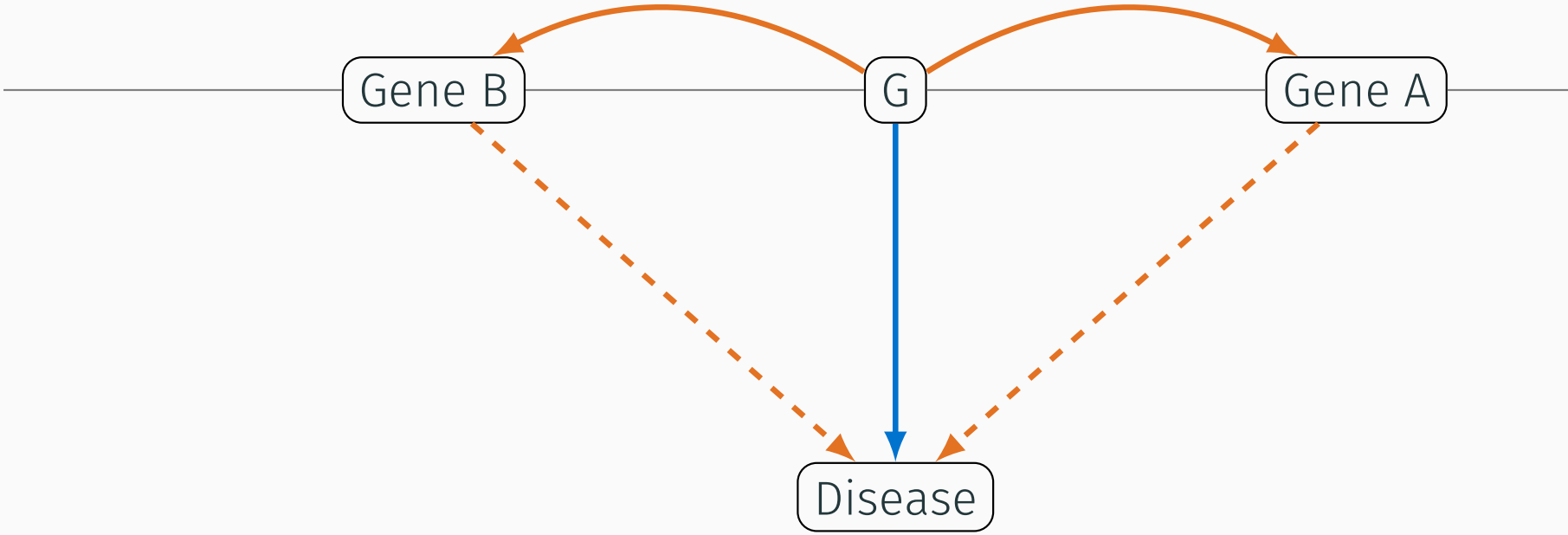
Gene A

Disease


.C008914.1
IL6ST
ANKRD55
IL6ST
ANKRD55
RA
MS
CRO
−log10(p)
55300000
55400000
55500000
Chromosome position


Gene B
G
Gene A
Disease


Gene B
G
Gene A
Disease


Gene B
G
Gene A
Disease

Summary

Colocalisation looks for shared genetic causes

- Search for 1+ traits that colocalise with trait of interest – potential mediators
- Or test for shared genetics for specific trait pair
- Pleitropy (one variant $\rightarrow$ two phenotypes) **is** colocalisation
- Co-regulation (one variant $\rightarrow$ two genes) detectable – interpretation: “more experiments/data needed”
- Single coloc gene might reflect only partial search

Bayesian method - informative priors

- Default values for marginal priors seem reasonable
- Always reasonable priors for joint causality not known (also situation dependent)

Thanks to...

Stasia Grinberg 

Cooper *HMG* 2012	Trynka *NG* 2011	Eyre *NG* 2012	Onengut-Gumuscu *NG* 2015
Faraco *PlosG* 2013	Liu *NG* 2012	Liu *NG* 2013	Momozawa *NG* 2018
Tsoi *NG* 2012	Hinks *NG* 2013	IMSGC *NG* 2013	Liu *NG* 2015

wellcome

UNIVERSITY OF CAMBRIDGE

MRC | Biostatistics Unit