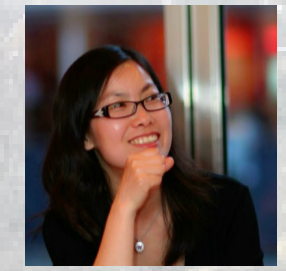
Integrating multiple sources and shapes of genomic data to prioritise autoimmune disease-candidate genes

Chris Wallace chr1swallace chr1swallace.github.io cew54@cam.ac.uk



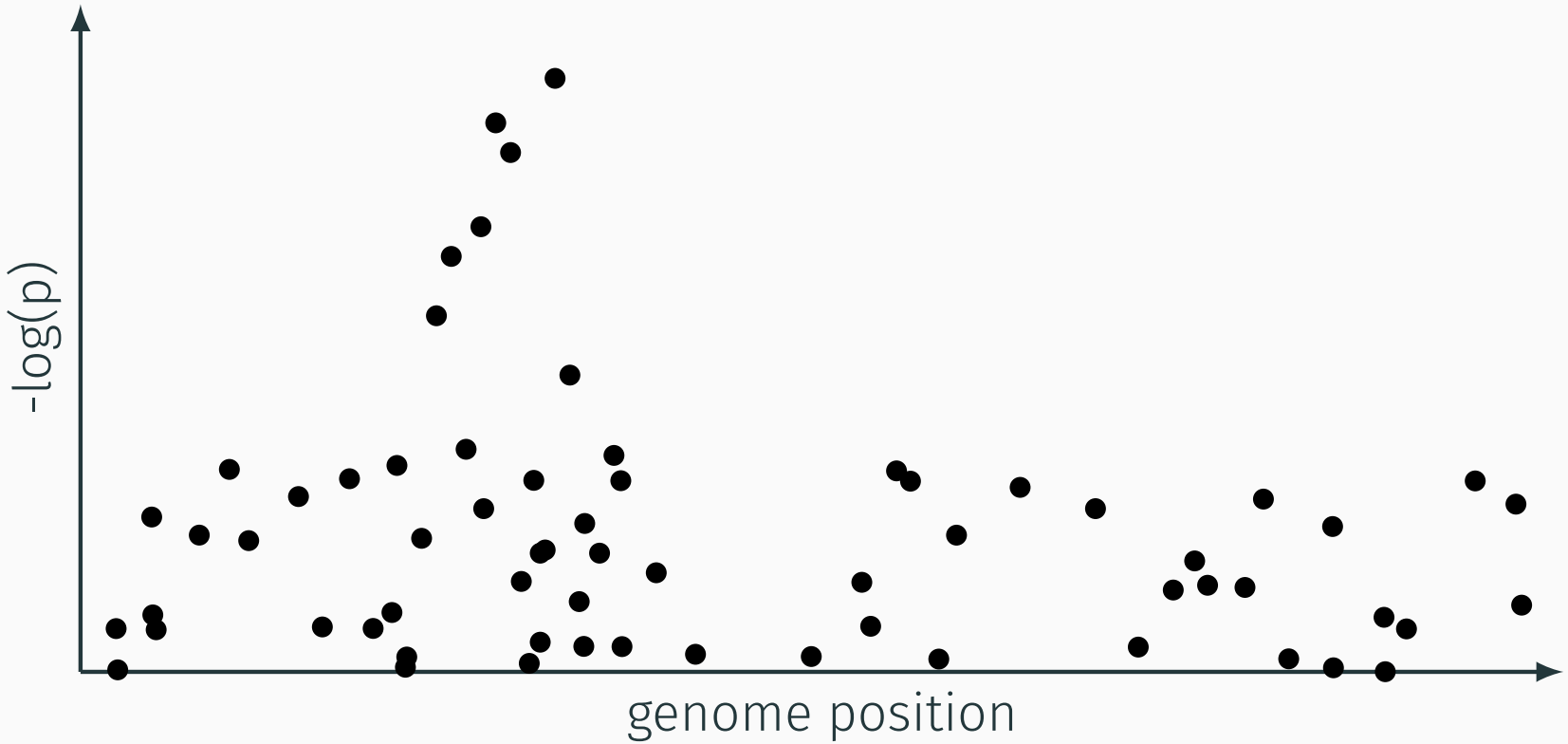
Hui Guo

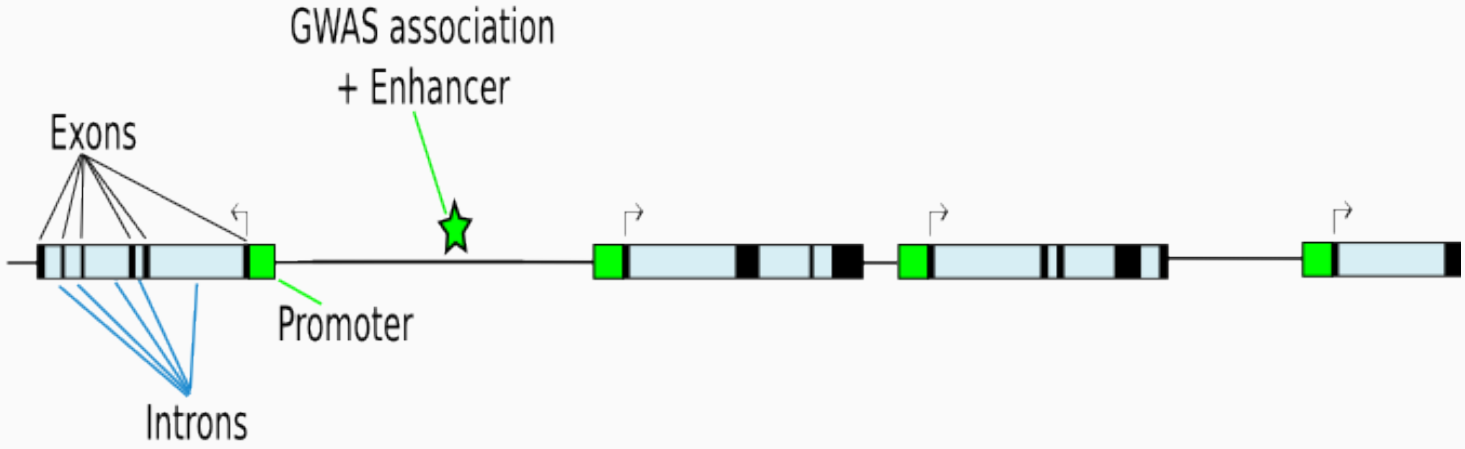
Stasia Grinberg

Olly Burren

Dan Rainbow

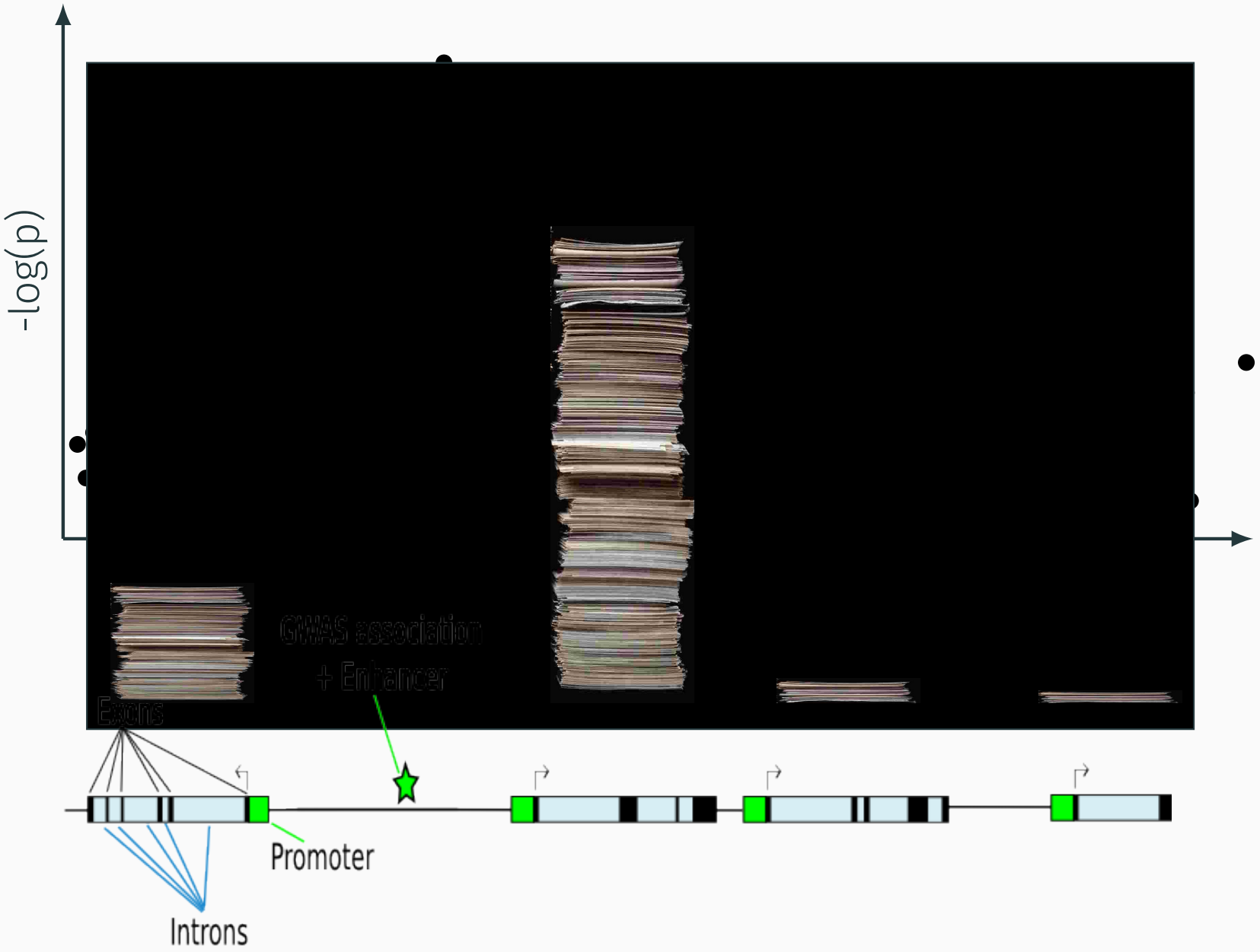
Which *gene* alters disease risk?



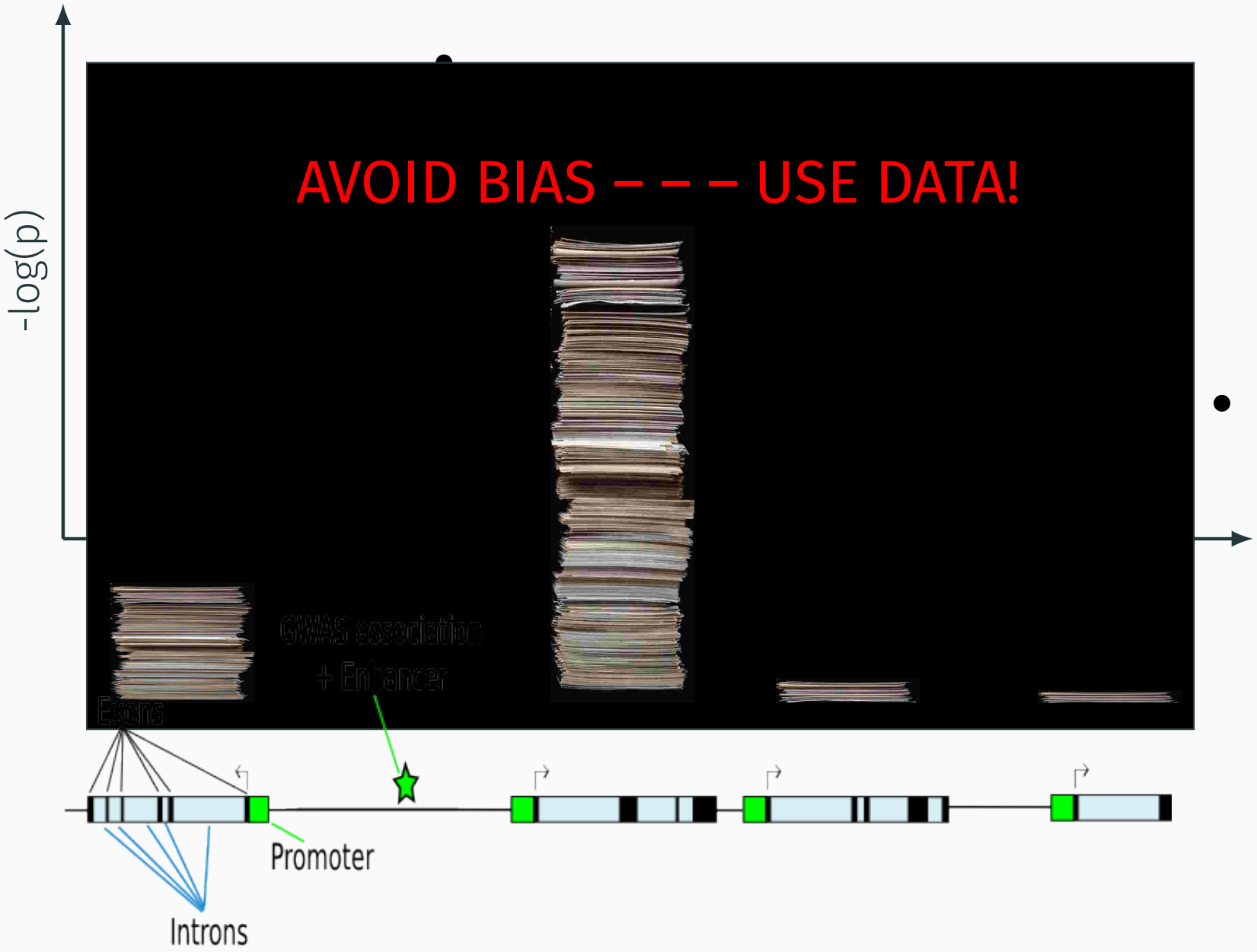


Which *gene* alters disease risk?

Which *gene* alters disease risk?

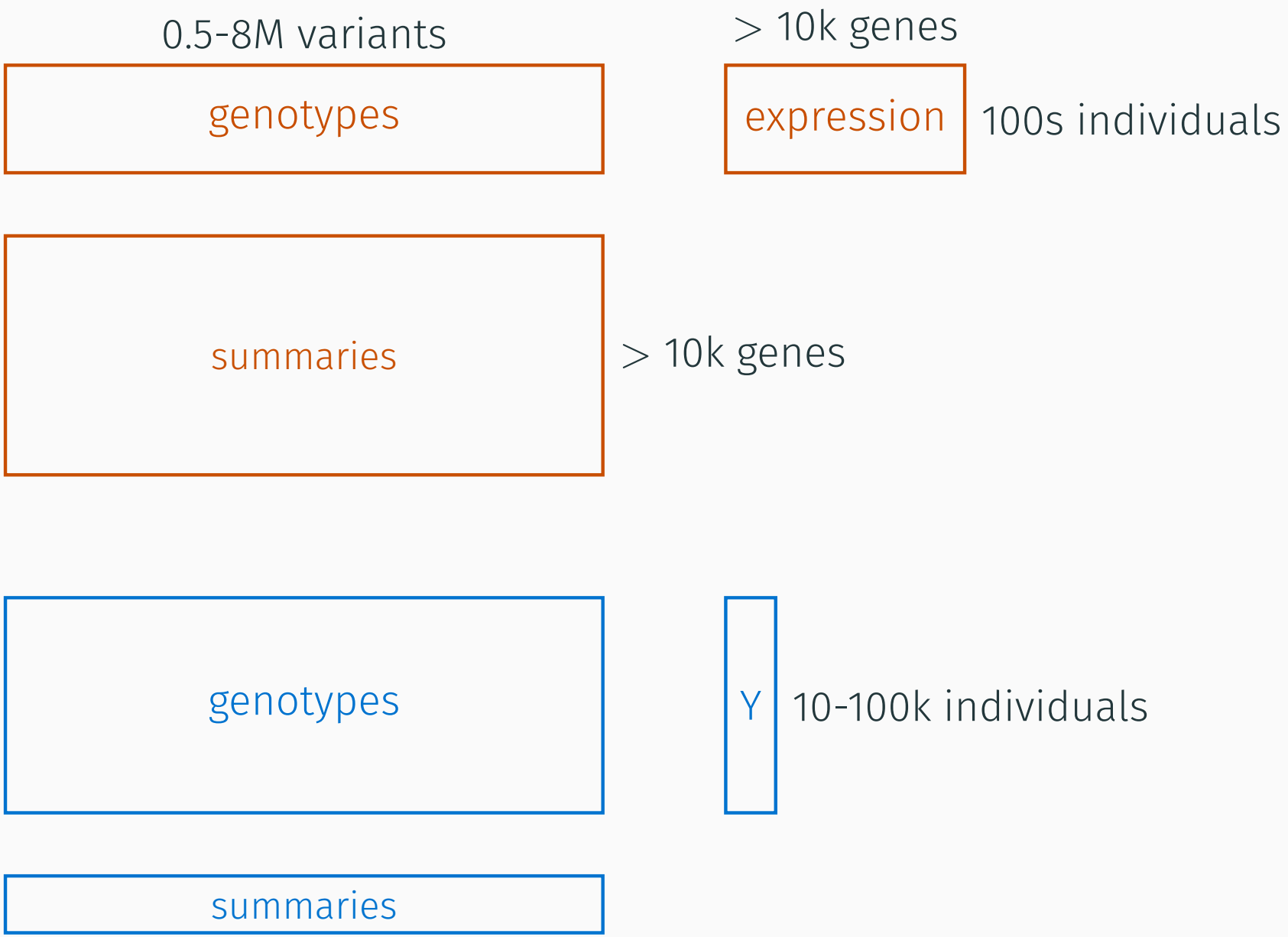


Which *gene* alters disease risk?

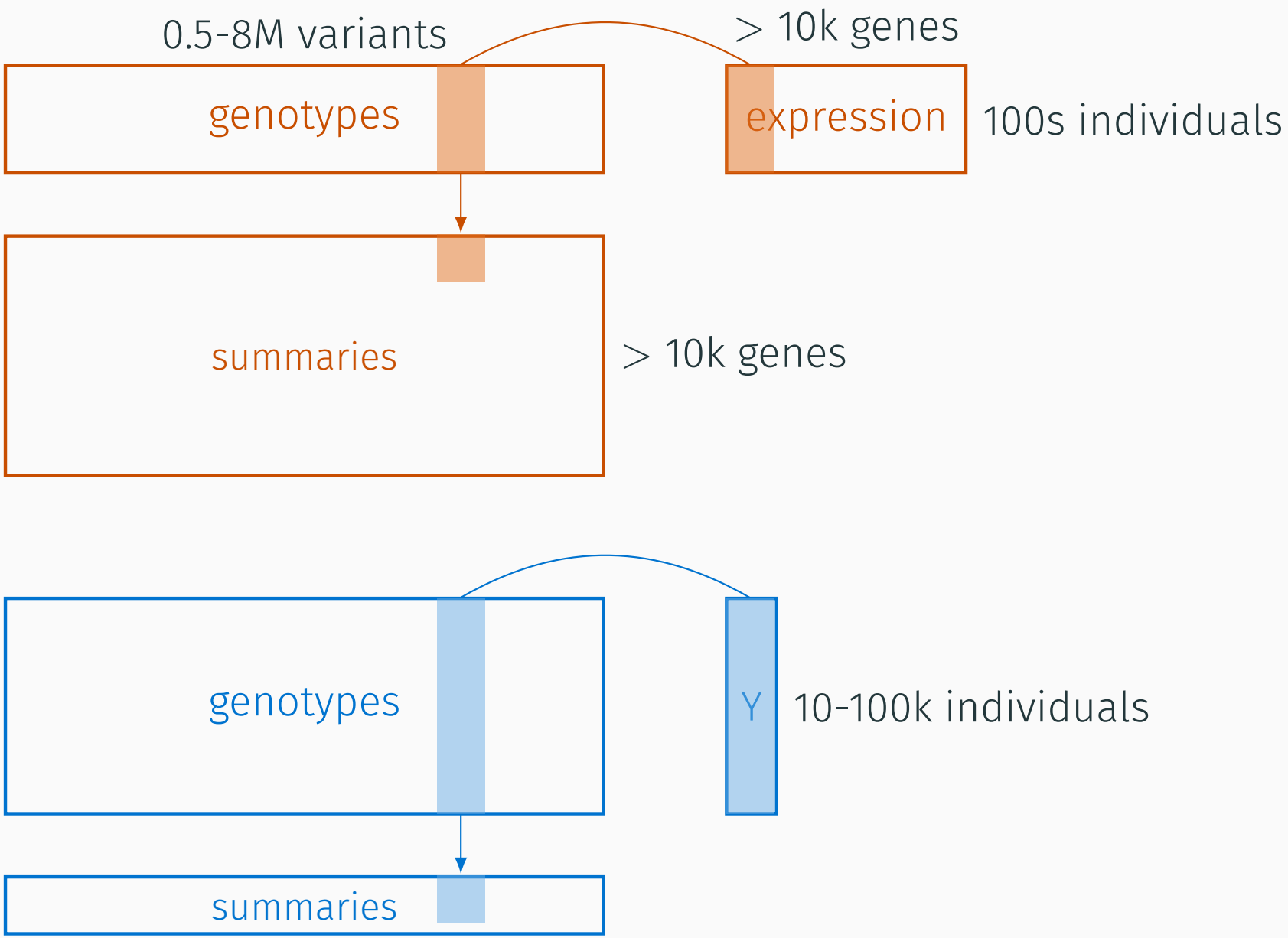


1. Colocalisation of disease and gene expression variants

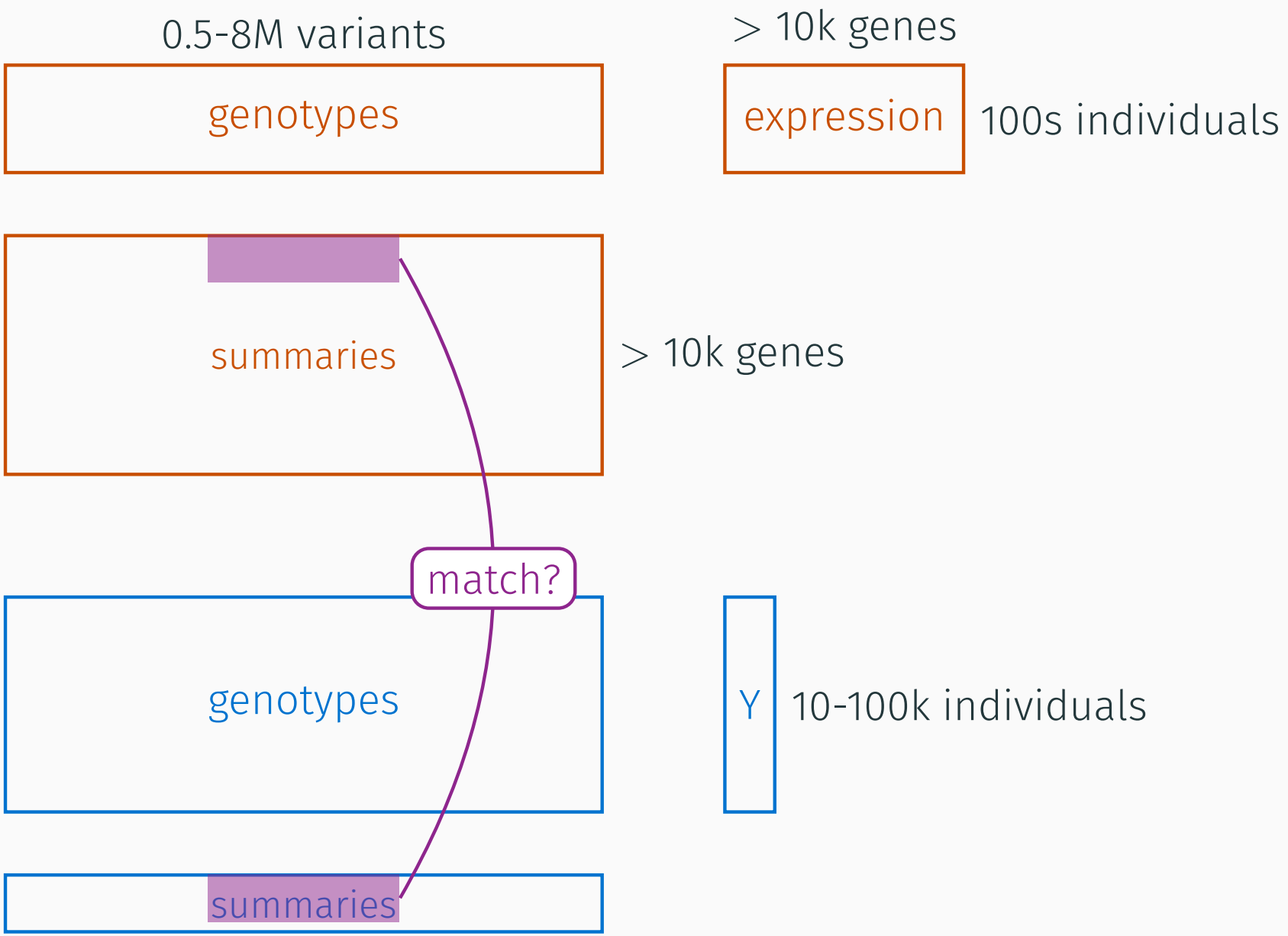
Integrate GWAS and eQTL datasets via colocalisation



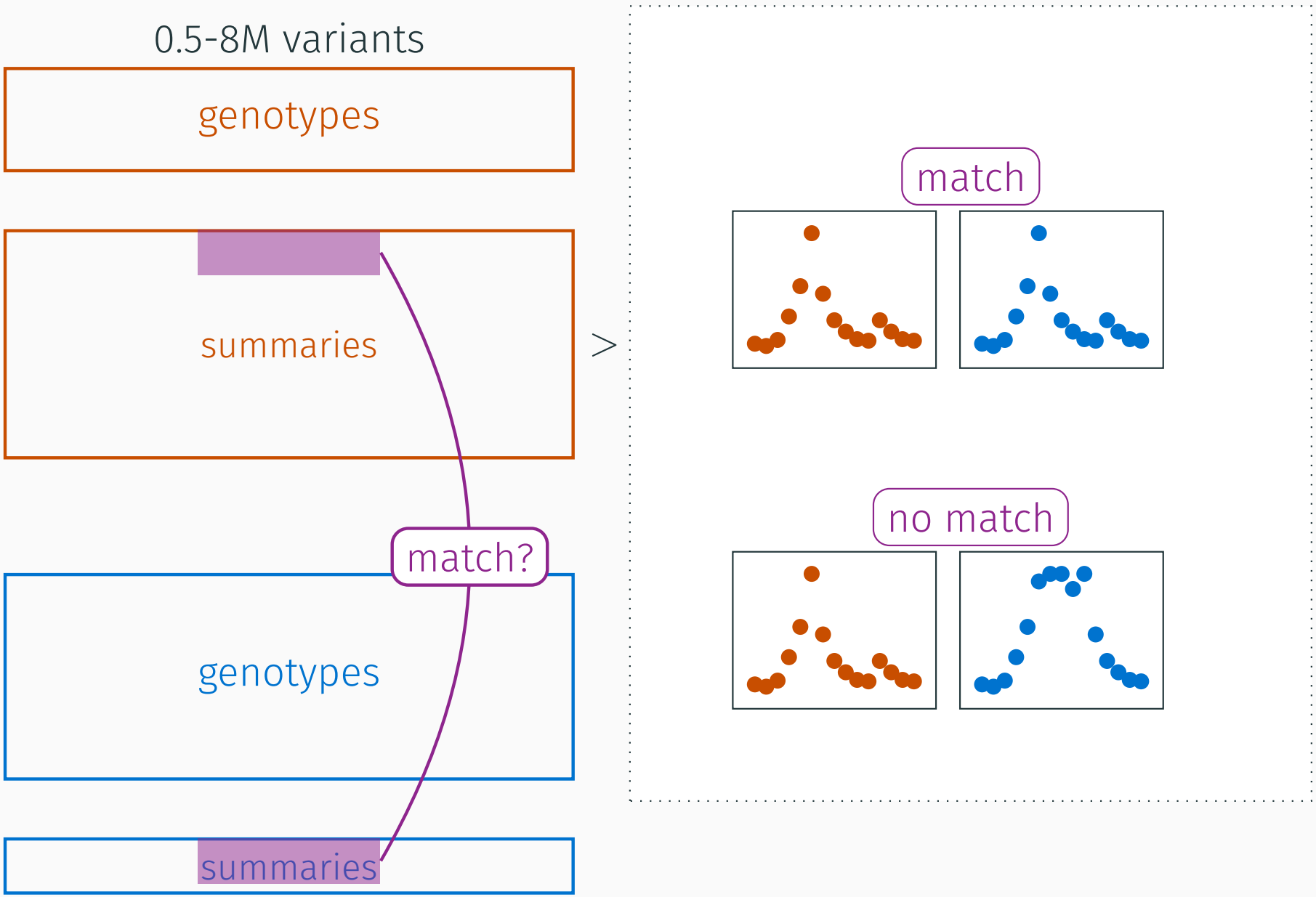
Integrate GWAS and eQTL datasets via colocalisation



Integrate GWAS and eQTL datasets via colocalisation

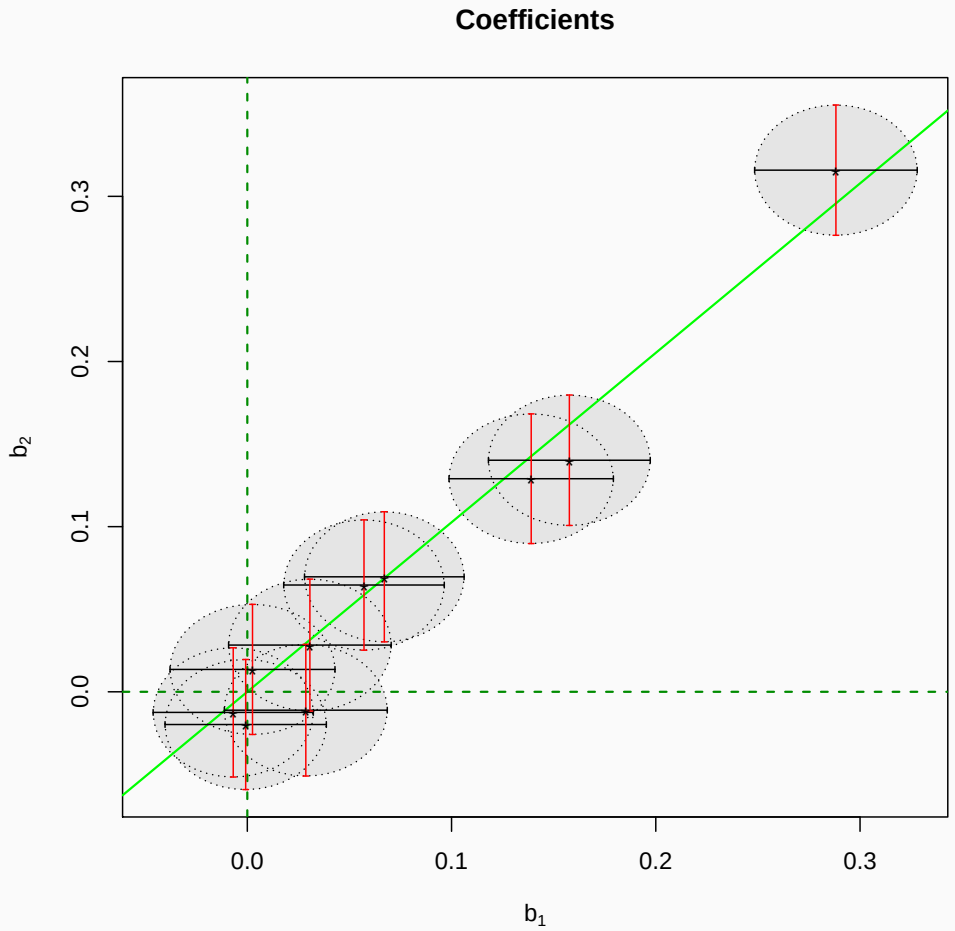


Integrate GWAS and eQTL datasets via colocalisation



Colocalisation methods (1) testing proportionality of effects

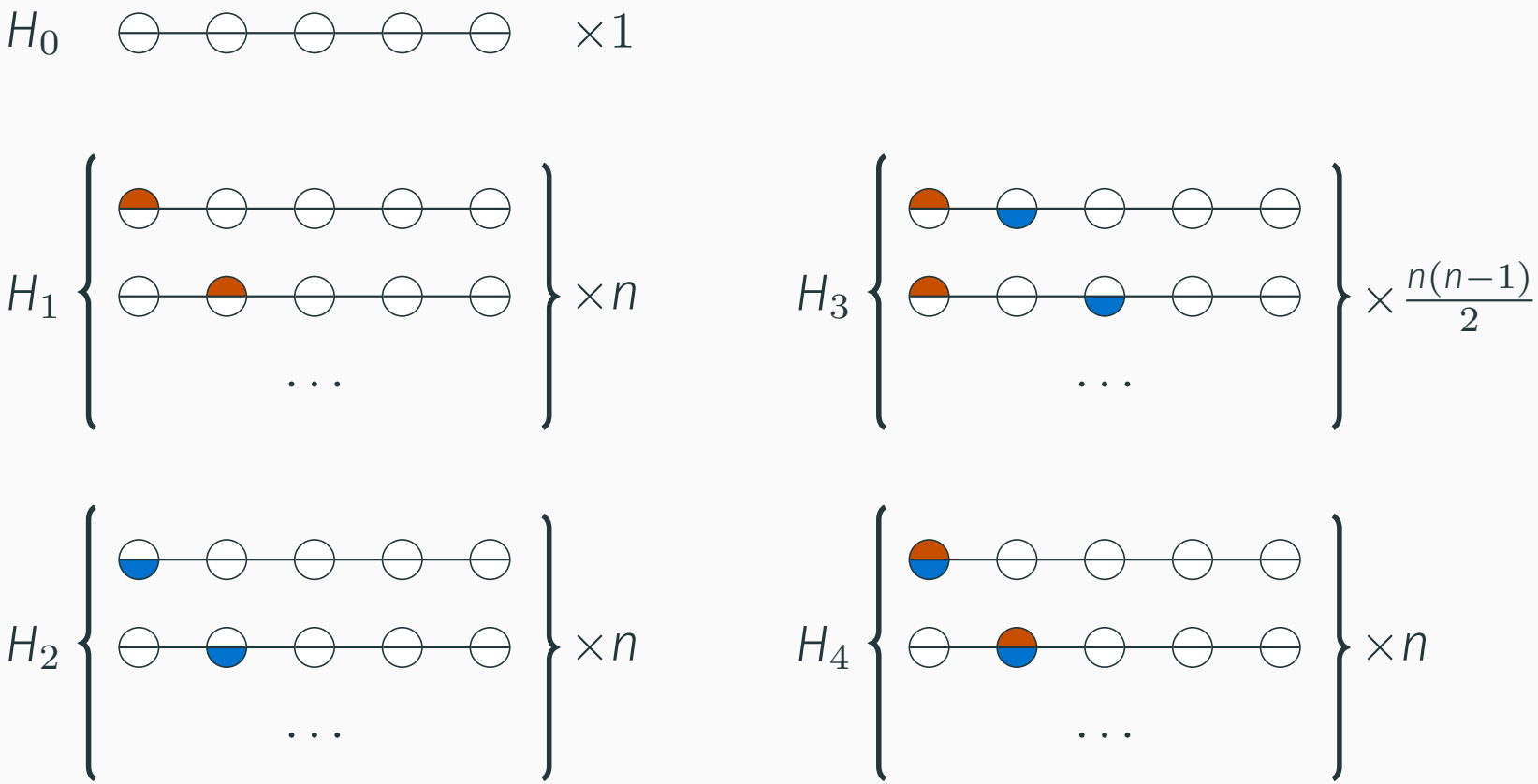
If two traits share one or more causal variants, then regression coefficients for the traits against any set of variants *in the neighbourhood of those causal variants* should be proportional.



Plagnol et al, Biostat 2012; Wallace, Genet Epidemiol 2014

Colocalisation methods (2) enumeration of possibilities

- Assume at most one causal variant in a region
- Partition possible configurations of association into hypotheses



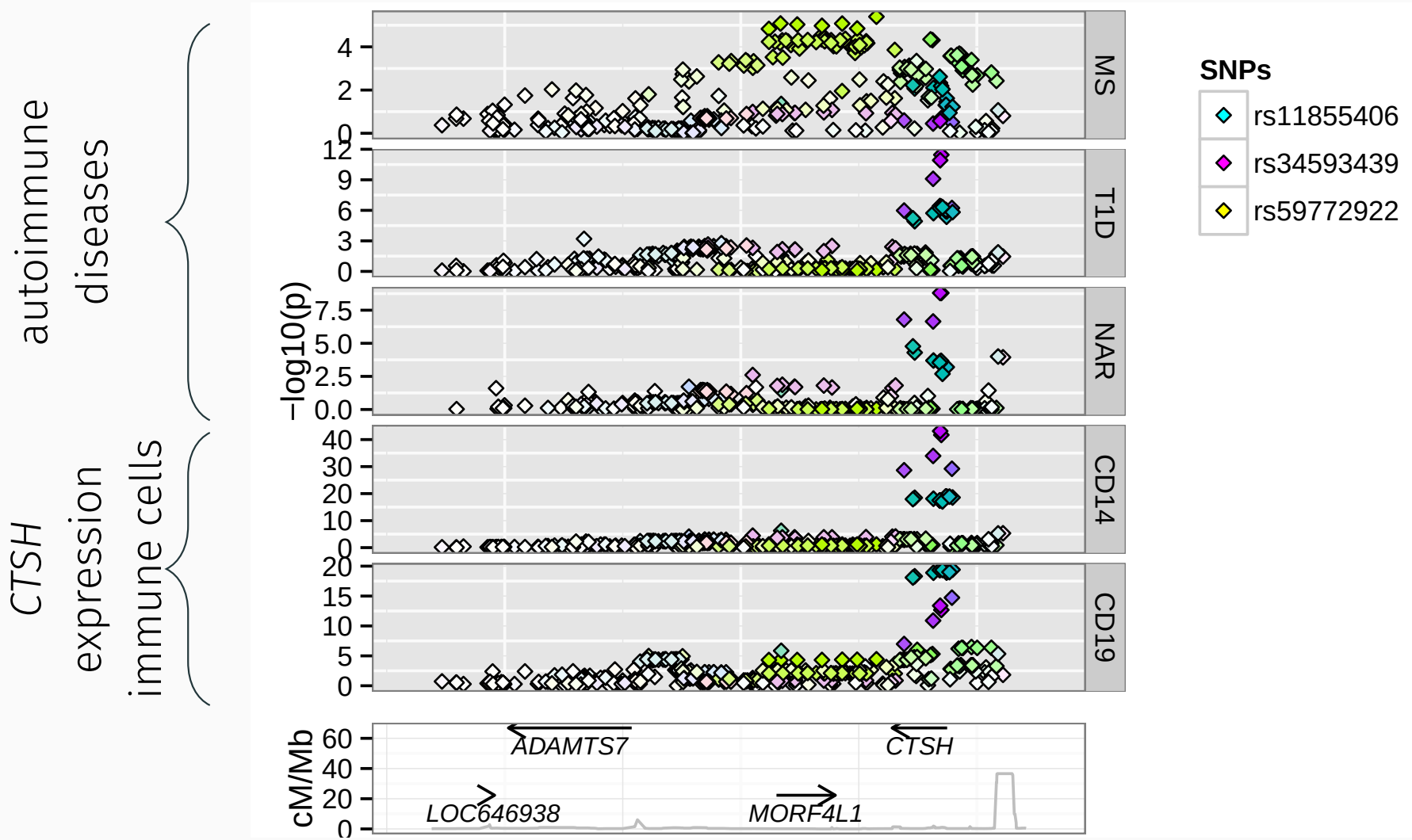
Comparison of two approaches

	Proportionality	Enumeration
N. causal variants	any	one
Density required	any	dense
Data required	full	summary
Variant selection	complicated	–

Selection of variants for proportionality testing

Too many → too many degrees of freedom (loss of power)
Most significant → biased coefficients (type 1 errors)
Solutions: dimension reduction on genotype matrix, model averaging

Colocalisation points to *CTSH* in CD14^{+} monocytes



Guo et al, HMG 2015

Colocalisation anaysis of ten autoimmune diseases with B cell, monocyte, stimulated monocyte eQTLs

+

B cells, monocytes (~ 300 samples)

Genetics of gene expression in primary immune cells identifies cell type–specific master regulators and roles of HLA alleles

Benjamin P Fairfax[1], Seiko Makino[1], Jayachandran Radhakrishnan[1], Katharine Plant[1], Stephen Leslie[2], Alexander Dilthey[3], Peter Ellis[4], Cordelia Langford[4], Fredrik O Vannberg[1,5] & Julian C Knight[1]

Trans-acting genetic vari
expression. Using paired
trans expression quantit
B cells, respectively. Add
disease locus, with *IP6K*
suggesting roles for cell
pathogenesis. We also fi
AOAH and *ARHGAP24*
cell populations identifie
disease susceptibility.

Defining the genetic deter
understanding the biologic
iation. This is particularly
variants underlying obser
wide association studies
functional activity of many
the context of relevant cell
state[2–4]. This context-dep

Innate Immune Activity Conditions the Effect of Regulatory Variants upon Monocyte Gene Expression

Benjamin P. Fairfax,* Peter Humburg, Seiko Makino, Vivek Naranbhai, Daniel Wong, Evelyn Lau, Luke Jostins, Katharine Plant, Robert Andrews, Chris McGee, Julian C. Knight*

Introduction: Many genetic variants associated with common disease susceptibility occur close to immune-related genes in noncoding DNA, suggestive of a regulatory function. The definition of functional variants and the specific genes that they regulate remains challenging and in many cases is unresolved. We hypothesized that a significant proportion of variants, including those implicated in disease, may show activity in a context-specific manner and therefore only be identifiable upon triggering of immune responses.

Methods: We mapped interindividual variation in gene expression as a quantitative trait, defining expression quantitative trait loci (eQTLs). To investigate the effect of innate immune stimuli on eQTLs, we exposed primary $CD14^+$ human monocytes from 432 European volunteers to the inflammatory proxies interferon-γ (IFN-γ) or differing durations (2 or 24 hours) of lipopolysaccharide (LPS). eQTL mapping was performed on a genome-wide basis with an additive linear model. A subset of 228 individuals with expression data available for all experimental conditions enabled cross-treatment comparisons.

Results: Stimulation with LPS or IFN-γ resulted in profound effects across monocyte eQTLs, with hundreds of genes and associated pathways demonstrating context-specific eQTLs dependent on the type and duration of stimulus. Context-specific eQTLs frequently intersected established canonical pathways of monocyte signaling and included key nodal genes and effector molecules. These eQTLs are typically more distal to the transcriptional start site and, in some cases, showed reversal of

on context, paralleling the expression of class II genes. Induced eQTLs were enriched for disease-risk loci with context-specific associations to many putative causal genes including at *ATM*, *IRF8*,

READ THE FULL ARTICLE ONLINE
http://dx.doi.org/10.1126/science.1246

Cite this article as B. P. Fairfax *et al.*, *Science* **343**, 1246949 (2014).
DOI: 10.1126/science.1246949

FIGURES IN THE FULL ARTICLE

Fig. 1. Genotype modulates the gene expression response to innate immune stimuli in monocytes.

Fig. 2. Trans-eQTL demonstrate context specificity and identify master regulatory loci after treatment.

Fig. 3. Temporal effects for a stimulus-specific trans-eQTL.

Fig. 4. Cis regulation of IRF2 at rs13149
has profound transcriptional consequenc
in trans.

Fig. 5. Stimulus-specific eQTL and GWAS.

Fig. 6. Examples of context-specific eQTL
informative for disease risk.

SUPPLEMENTARY MATERIALS

Materials and Methods
Figs. S1 to S12

signaling.
VB1 expressio
a single–nucleotide polymorphism (rs2275
revealed after 2 hours of LPS stimulation of

Activated monocytes (~ 300 samples)

=

?

candidate causal autoimmune genes

Guo et al., Hum Mol Genet, *2015*

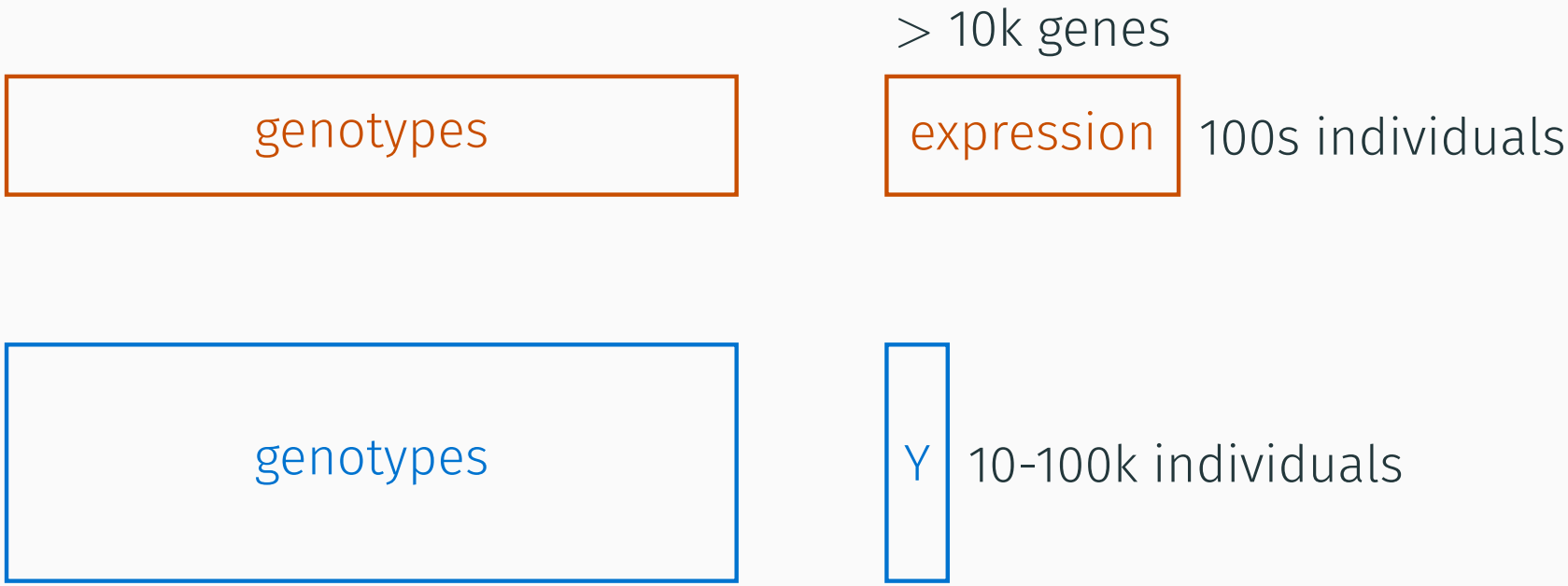
Colocalisation identified six candidate causal autoimmune disease genes

Gene	Disease(s)	Direction
Resting B cells + monocytes		
RGS1	Celiac, MS	-
SYNGR1	Primary Biliary cirrhosis	+
Resting + activated monocytes		
ADAM15	Crohn's	?
CARD19	Crohn's, ulcerative colitis	+
LTBR	Primary Biliary cirrhosis	+
CTSH	T1D, narcolepsy	-

Guo et al., Hum Mol Genet, *2015*

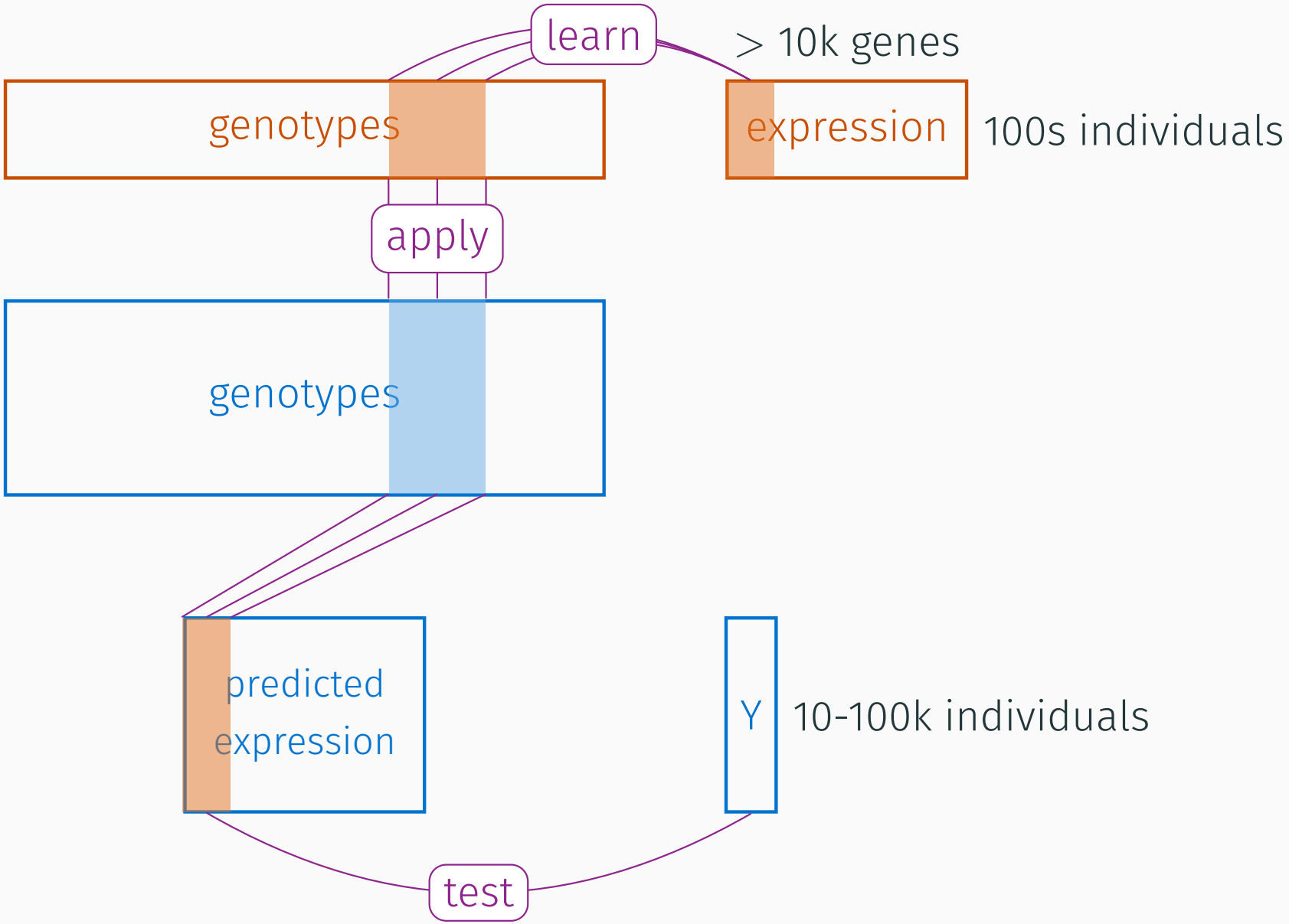
2. Transcriptome-wide association study

Integrate GWAS and eQTL datasets via TWAS



Gamazon et al, Nat Genet, 2015; Gusev et al, Nat Genetics, 2016

Integrate GWAS and eQTL datasets via TWAS



Gamazon et al, Nat Genet, 2015; Gusev et al, Nat Genetics, 2016

TWAS of T1D with B cell, monocyte, stimulated monocyte eQTLs

+

B cells, monocytes (~ 300 samples)

Genetics of gene expression in primary immune cells identifies cell type–specific master regulators and roles of HLA alleles

Benjamin P Fairfax[1], Seiko Makino[1], Jayachandran Radhakrishnan[1], Katharine Plant[1], Stephen Leslie[2], Alexander Dilthey[3], Peter Ellis[4], Cordelia Langford[4], Fredrik O Vannberg[1,5] & Julian C Knight[1]

Trans-acting genetic vari
expression. Using paired
trans expression quantit
B cells, respectively. Add
disease locus, with *IP6K*
suggesting roles for cell
pathogenesis. We also fi
AOAH and *ARHGAP24* i
cell populations identifie
disease susceptibility.

Defining the genetic deter
understanding the biologic
iation. This is particularly
variants underlying obser
wide association studies
functional activity of many
the context of relevant cell
state[2–4]. This context-dep

Innate Immune Activity Conditions the Effect of Regulatory Variants upon Monocyte Gene Expression

Benjamin P. Fairfax,* Peter Humburg, Seiko Makino, Vivek Naranbhai, Daniel Wong, Evelyn Lau, Luke Jostins, Katharine Plant, Robert Andrews, Chris McGee, Julian C. Knight*

Introduction: Many genetic variants associated with common disease susceptibility occur close to immune-related genes in noncoding DNA, suggestive of a regulatory function. The definition of functional variants and the specific genes that they regulate remains challenging and in many cases is unresolved. We hypothesized that a significant proportion of variants, including those implicated in disease, may show activity in a context-specific manner and therefore only be identifiable upon triggering of immune responses.

Methods: We mapped interindividual variation in gene expression as a quantitative trait, defining expression quantitative trait loci (eQTLs). To investigate the effect of innate immune stimuli on eQTLs, we exposed primary CD14+ human monocytes from 432 European volunteers to the inflammatory proxies interferon-γ (IFN-γ) or differing durations (2 or 24 hours) of lipopolysaccharide (LPS). eQTL mapping was performed on a genome-wide basis with an additive linear model. A subset of 228 individuals with expression data available for all experimental conditions enabled cross-treatment comparisons.

Results: Stimulation with LPS or IFN-γ resulted in profound effects across monocyte eQTLs, with hundreds of genes and associated pathways demonstrating context-specific eQTLs dependent on the type and duration of stimulus. Context-specific eQTLs frequently intersected established canonical pathways of monocyte signaling and included key nodal genes and effector molecules. These eQTLs are typically more distal to the transcriptional start site and, in some cases, showed reversal of

on context, paralleling the expression of class II genes. Induced eQTLs were enriched for disease-risk loci with context-specific associations to many putative causal genes including at *ATM*, *IRF8*,

READ THE FULL ARTICLE ONLINE
http://dx.doi.org/10.1126/science.1246

Cite this article as B. P. Fairfax *et al.*, *Science* **343**, 1246949 (2014). DOI: 10.1126/science.1246949

FIGURES IN THE FULL ARTICLE

Fig. 1. Genotype modulates the gene expression response to innate immune stimuli in monocytes.

Fig. 2. Trans-eQTL demonstrate context specificity and identify master regulatory loci after treatment.

Fig. 3. Temporal effects for a stimulus-specific trans-eQTL.

Fig. 4. Cis regulation of IRF2 at rs13149 has profound transcriptional consequenc in trans.

Fig. 5. Stimulus-specific eQTL and GWAS.

Fig. 6. Examples of context-specific eQTL informative for disease risk.

SUPPLEMENTARY MATERIALS

Materials and Methods
Figs. S1 to S12

signaling.
VB1 expressio
a single–nucleotide polymorphism (rs2275
revealed after 2 hours of LPS stimulation of

Activated monocytes (~ 300 samples)

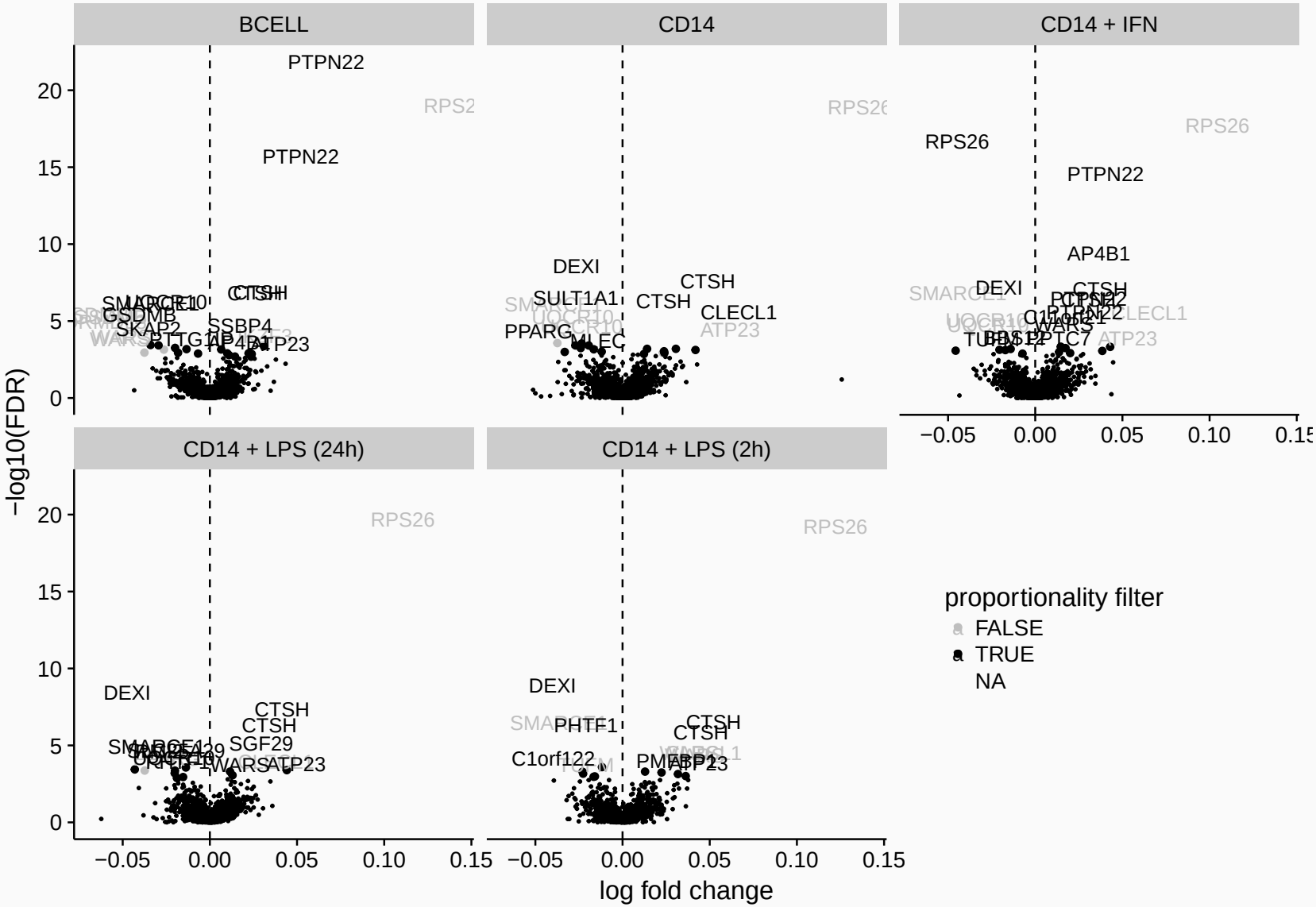
=

?

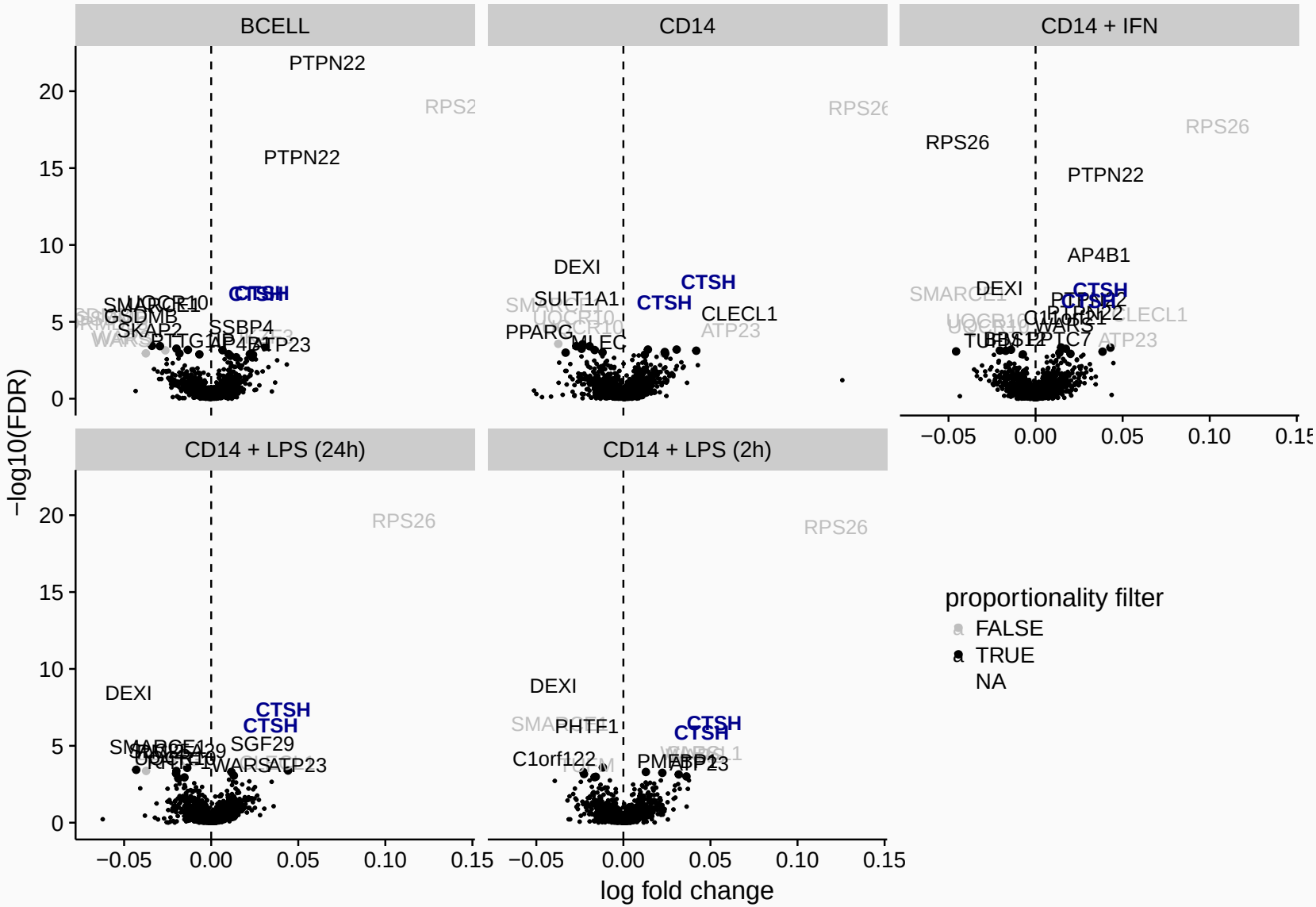
candidate causal autoimmune genes

Guo et al., Hum Mol Genet, *2015*

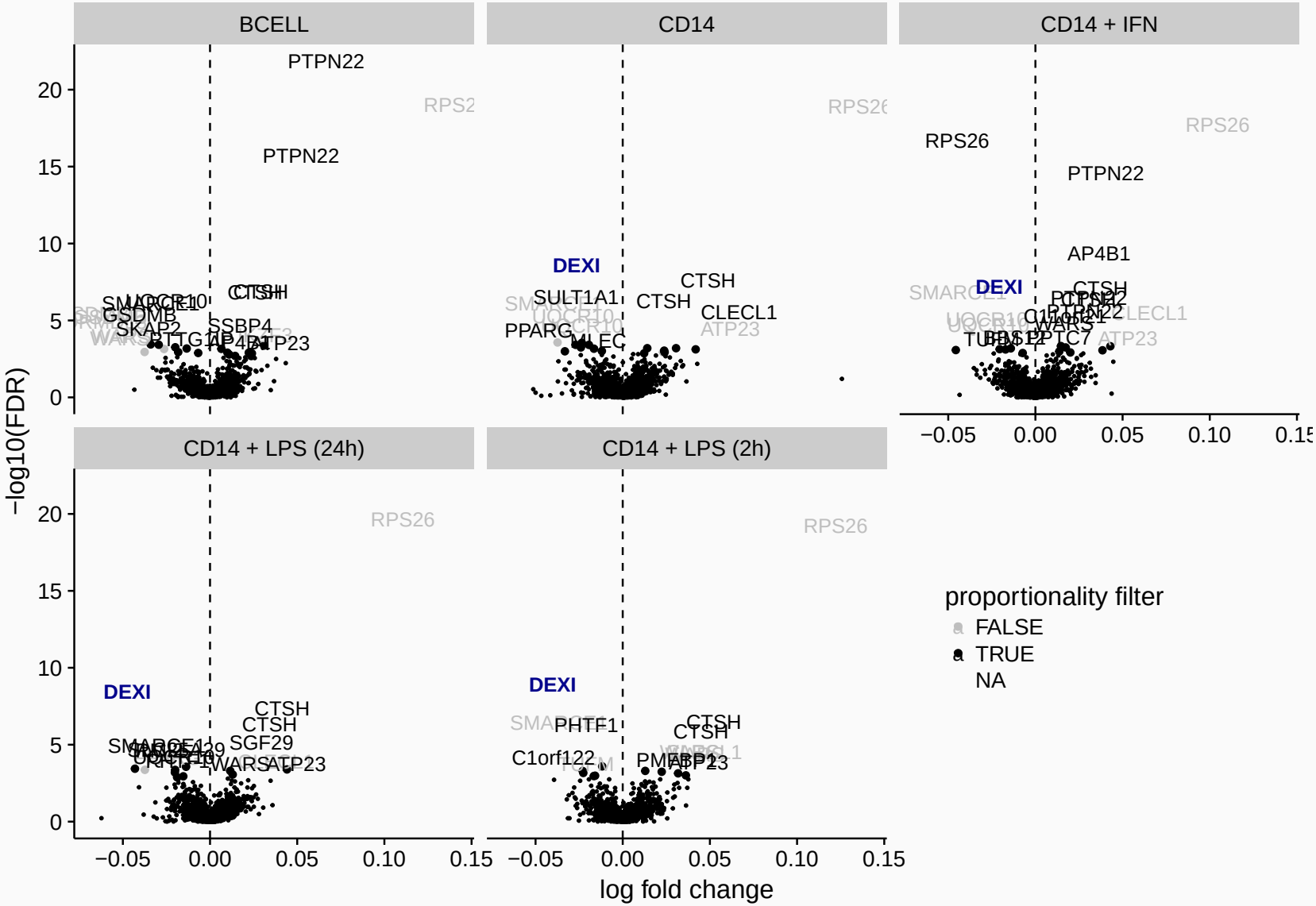
TWAS identified 88 T1D candidate genes, 61 after filtering



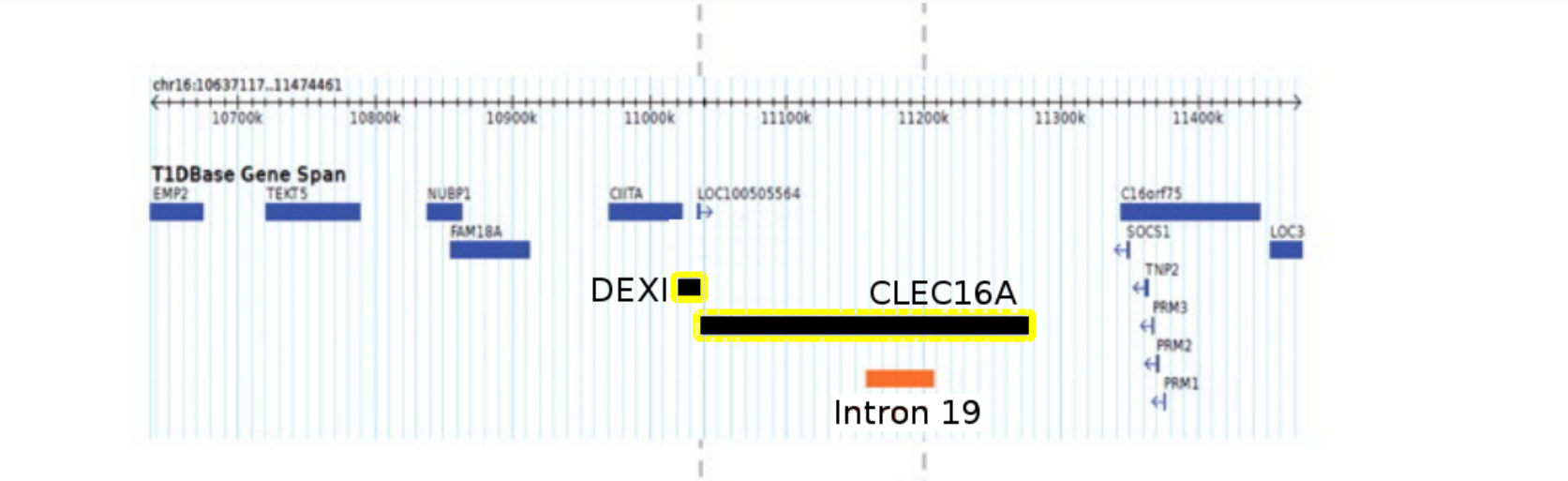
TWAS identified 88 T1D candidate genes, 61 after filtering



TWAS identified 88 T1D candidate genes, 61 after filtering

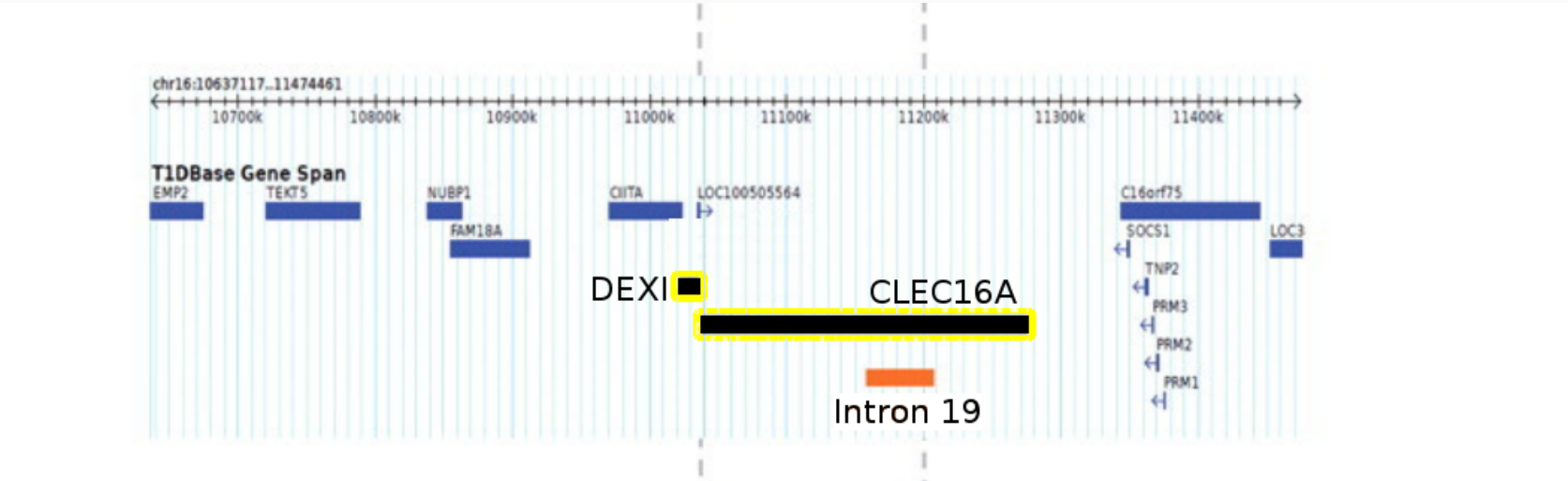


TWAS candidate *DEXI* supported by 3C experiments

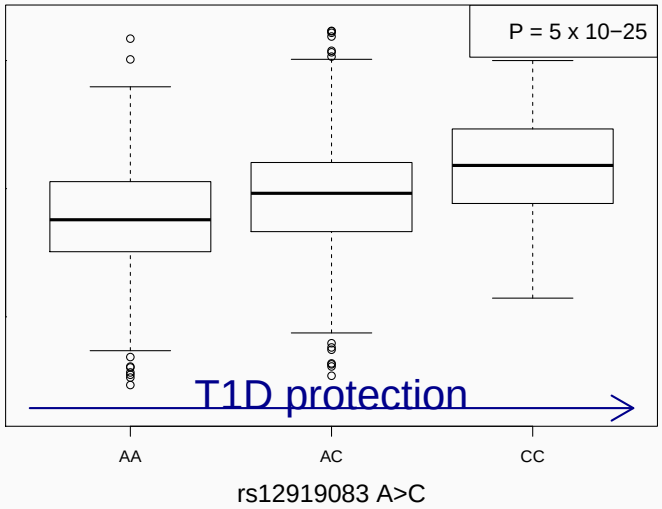


Wallace et al., Hum Mol Genet *2012*, *Davison et al.*, Hum Mol Genet *2012*

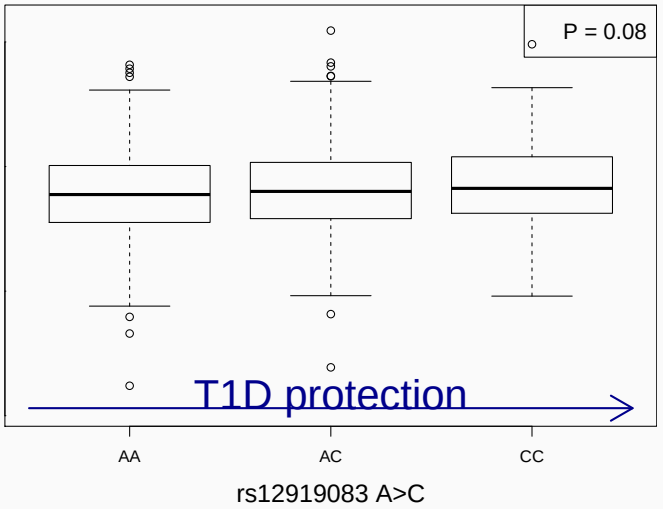
TWAS candidate *DEXI* supported by 3C experiments



DEXI expression

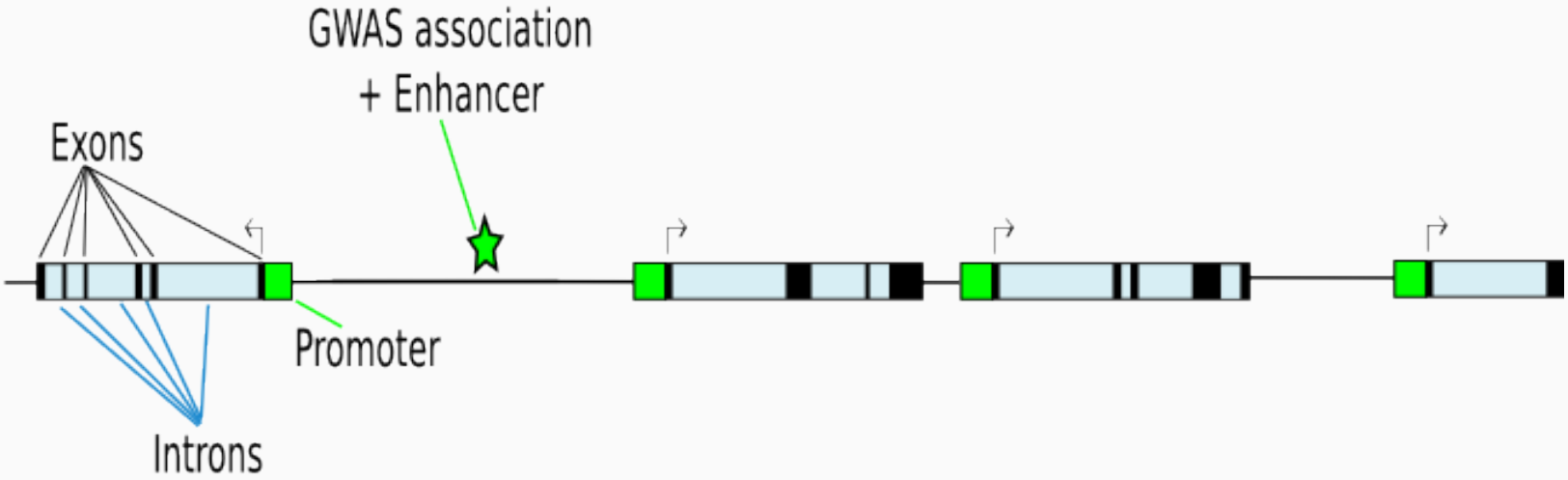


CLEC16A expression

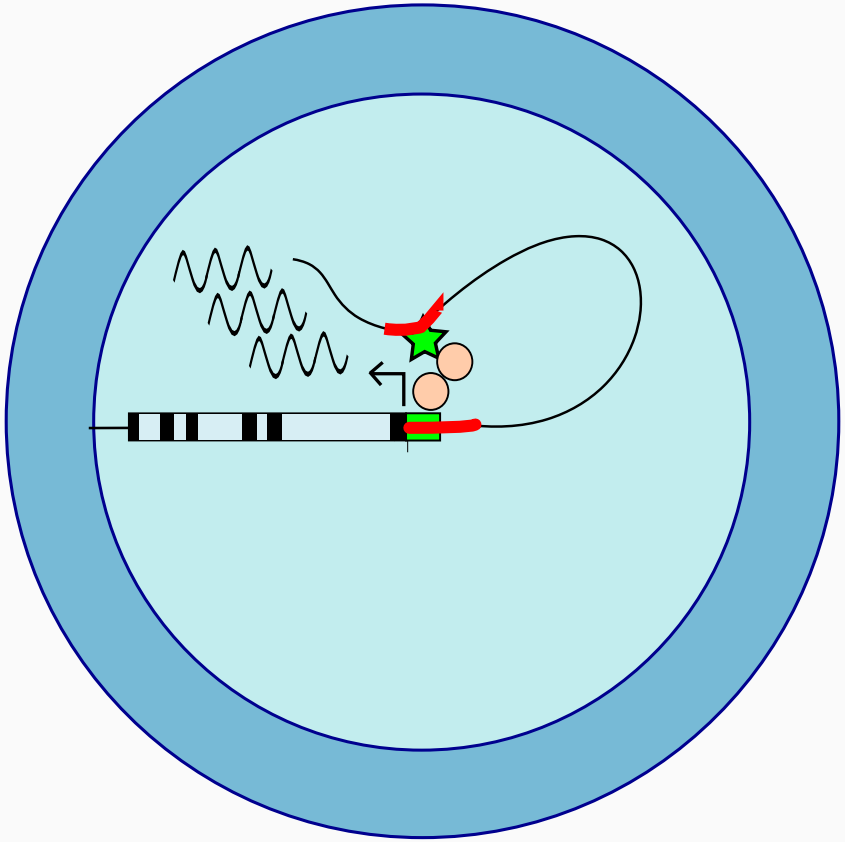


Wallace et al., Hum Mol Genet *2012*, *Davison et al.*, Hum Mol Genet *2012*

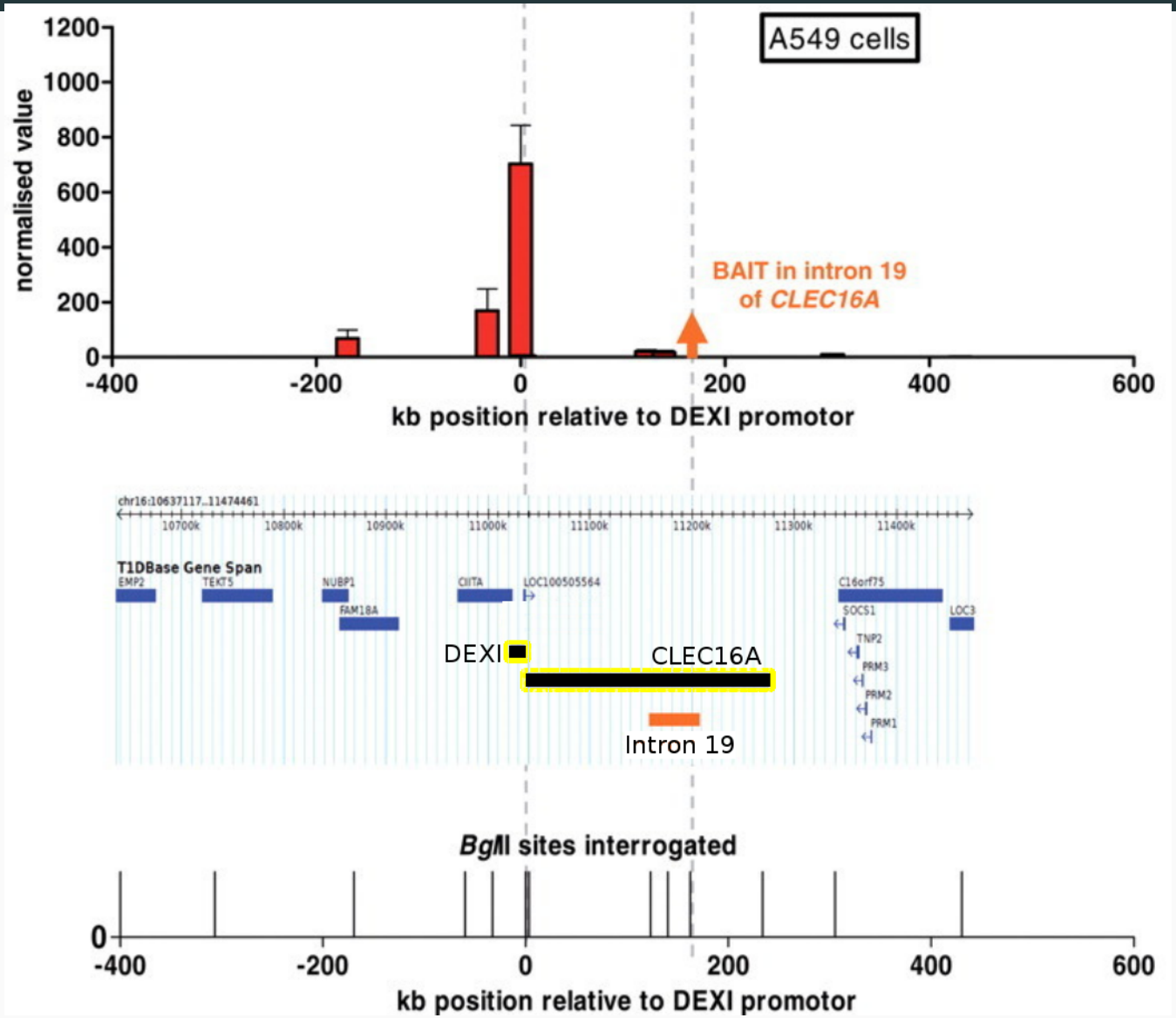
3D folding of DNA in nucleus connects enhancers to promoters



3D folding of DNA in nucleus connects enhancers to promoters



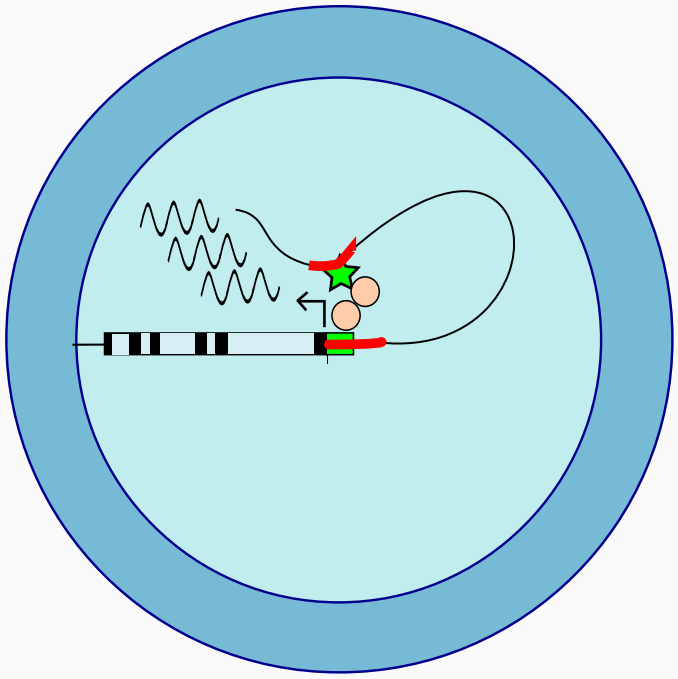
TWAS candidate *DEXI* supported by 3C experiments



Davison et al., Hum Mol Genet *2012*

3. Link promoters and regulatory elements through 3D DNA structure

Promoter capture Hi-C in 17 primary human blood cell types



Endothelial precursors

Erythroblasts

Neutrophils + precursors

Megakaryocytes

Macrophages (M0,M1,M2)

$CD4^+$ T cells, naïve and total

$CD8^+$ T cells, naïve and total

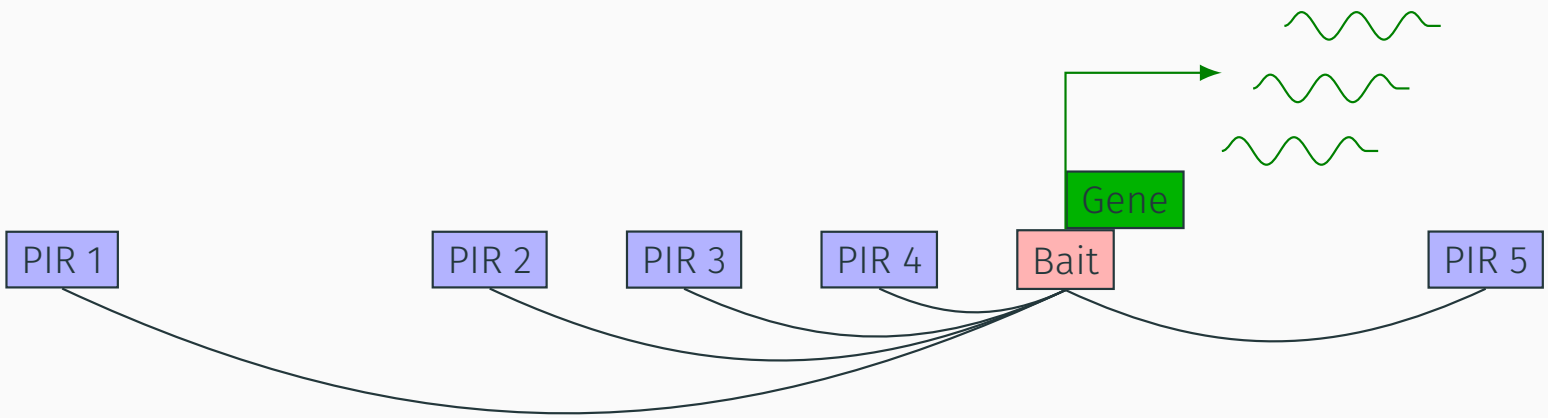
Monocytes

Fetal thymus

B cells, naïve and total

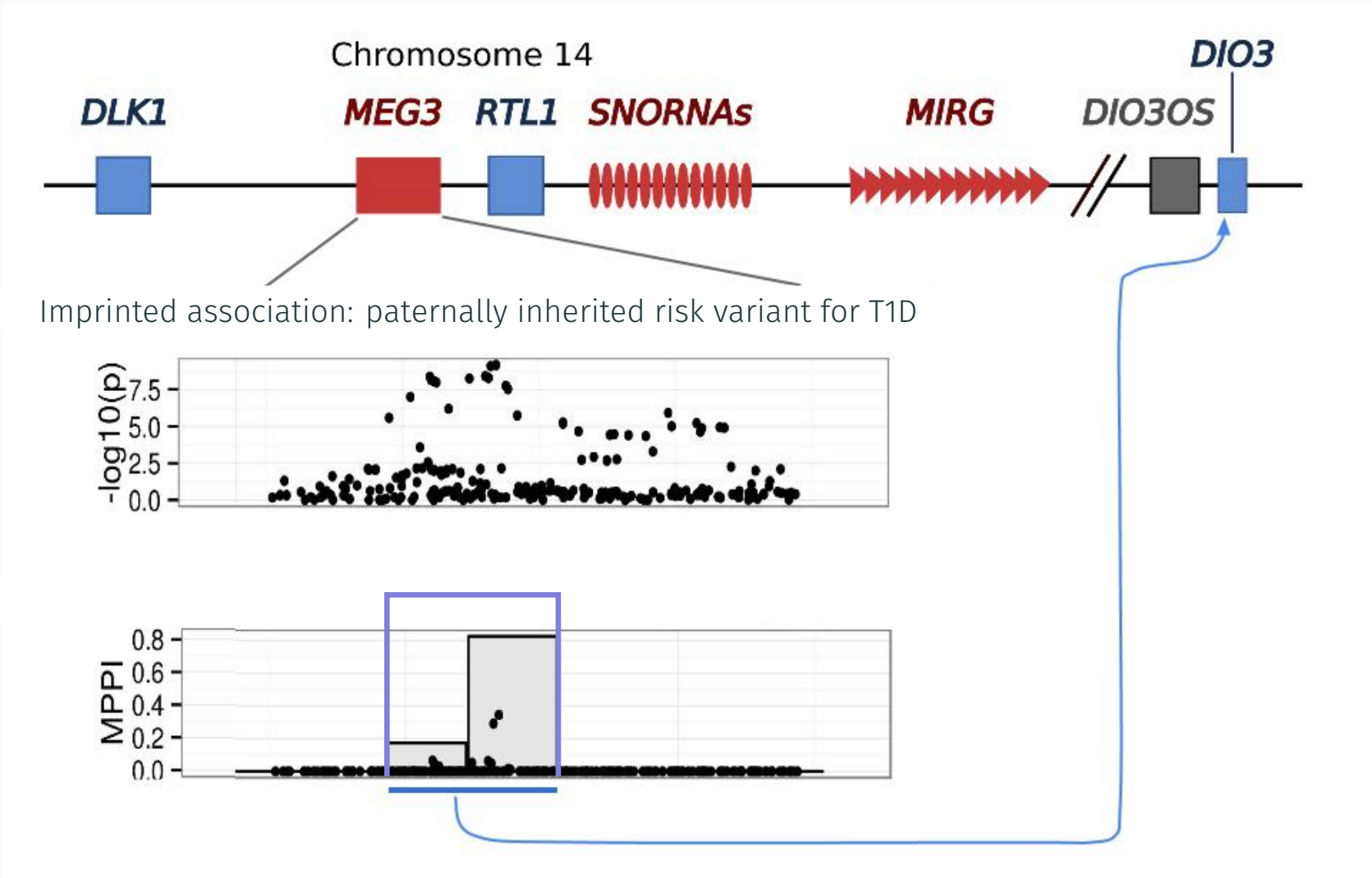
Total $CD4^+$, non-activated and activated at 4 hours

Promoter capture Hi-C in 17 primary human blood cell types



Javierre et al, Cell 2016

Integrate GWAS + PCHi-C to prioritise candidate disease genes



Imprinted association: paternally inherited risk variant for T1D

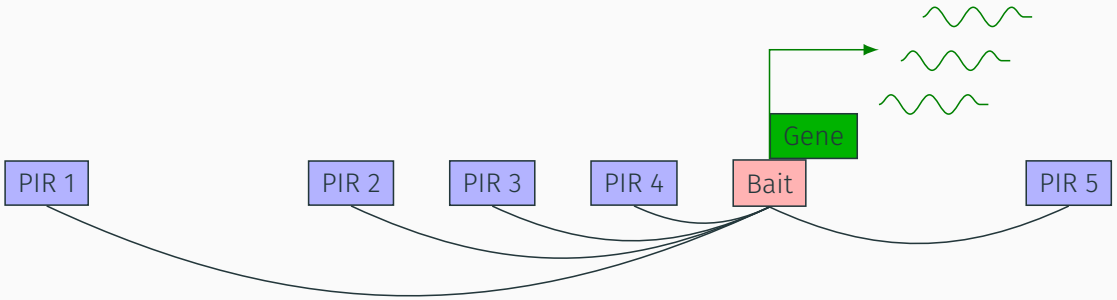
Aggregate by *Hind*III fragment or set of *Hind*III fragments

Wallace et al, Nat Genet, 2010

PCHi-C prioritisation in six autoimmune diseases

six diseases

\+



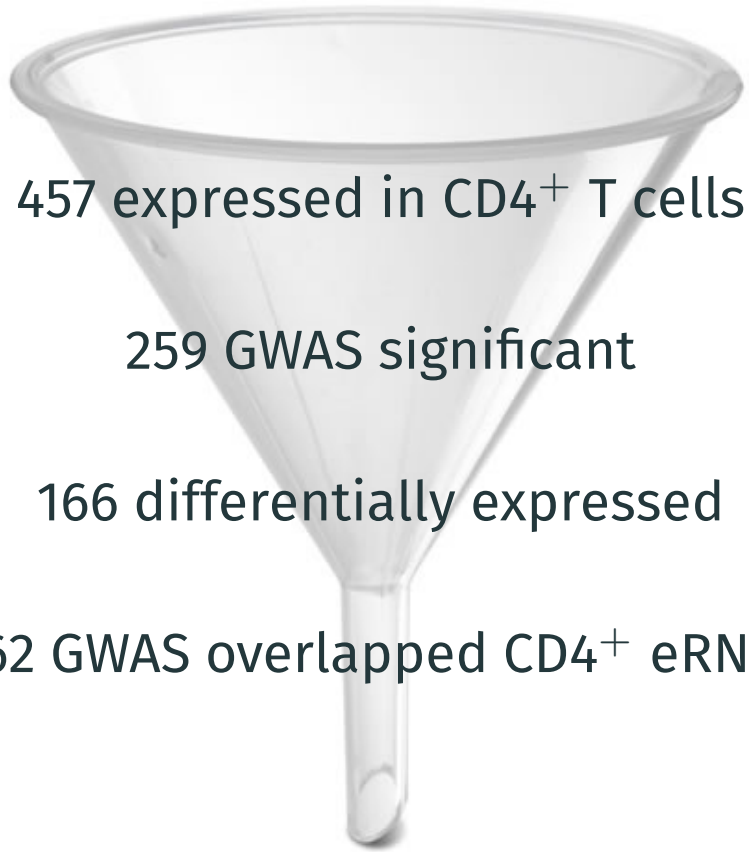
CD4$^+$ T cells, activated and non-activated

=

?
candidate causal
autoimmune genes

602 autoimmune disease prioritised genes in CD4$^+$ T cells

602 prioritised genes



457 expressed in CD4$^+$ T cells

259 GWAS significant

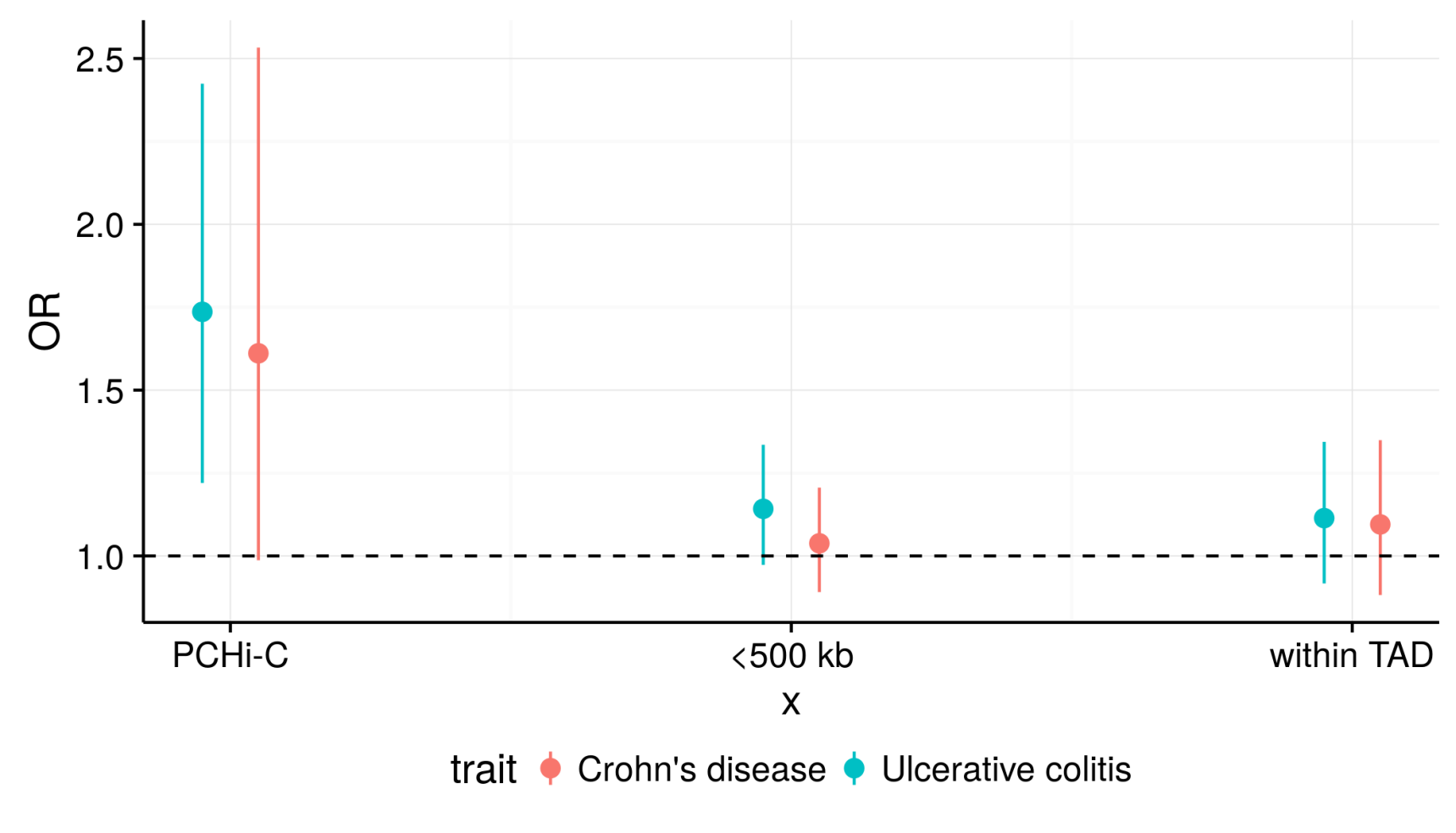
166 differentially expressed

62 GWAS overlapped CD4$^+$ eRNA

Require functional validation experiments

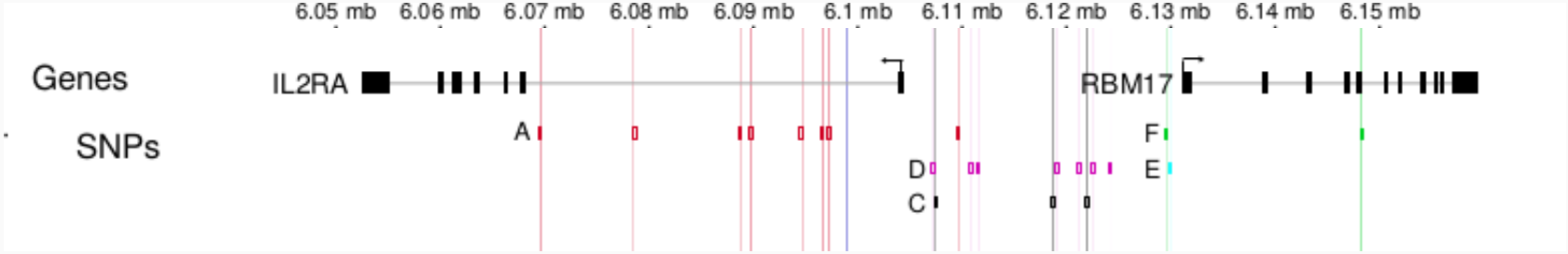
Burren et al., Gen Biol 2017

Prioritised IBD genes are enriched for differential expression in IBD patients



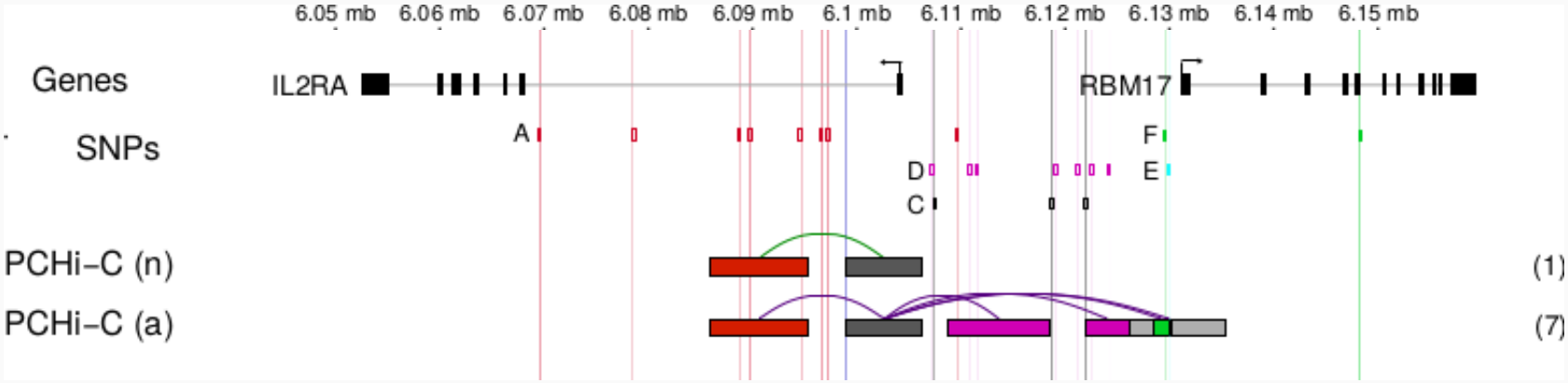
Peters et al, PLoS Genet 2016; Javeierre et al, Cell 2016

IL2RA prioritised in both activated and non-activated cells



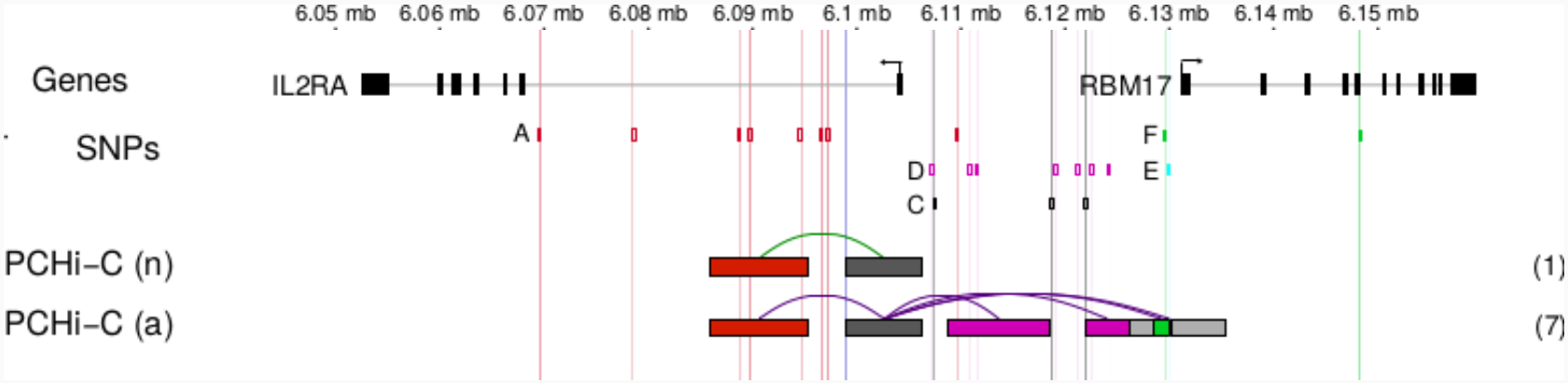
Wallace et al, PLoS Genet 2015; Burren et al, Gen Biol 2017

IL2RA prioritised in both activated and non-activated cells

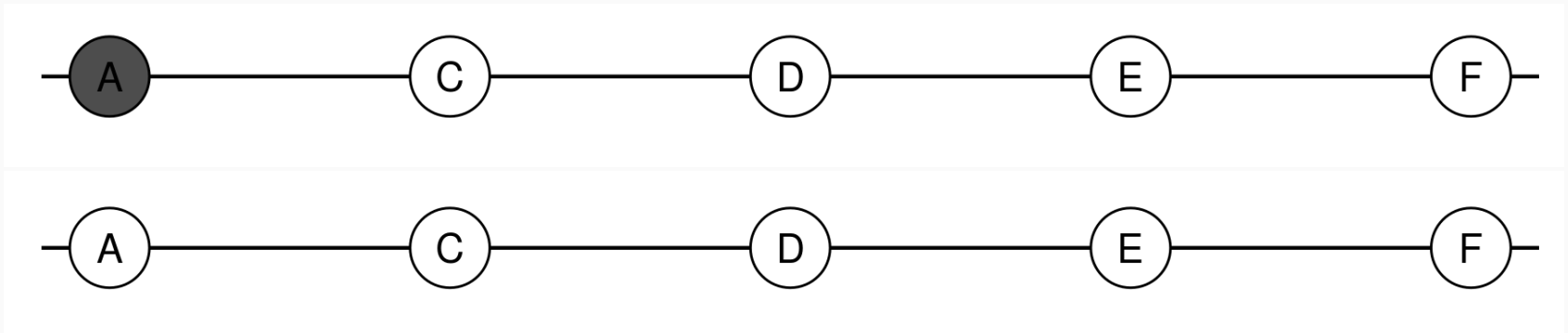


Wallace et al, PLoS Genet 2015; Burren et al, Gen Biol 2017

IL2RA prioritised in both activated and non-activated cells

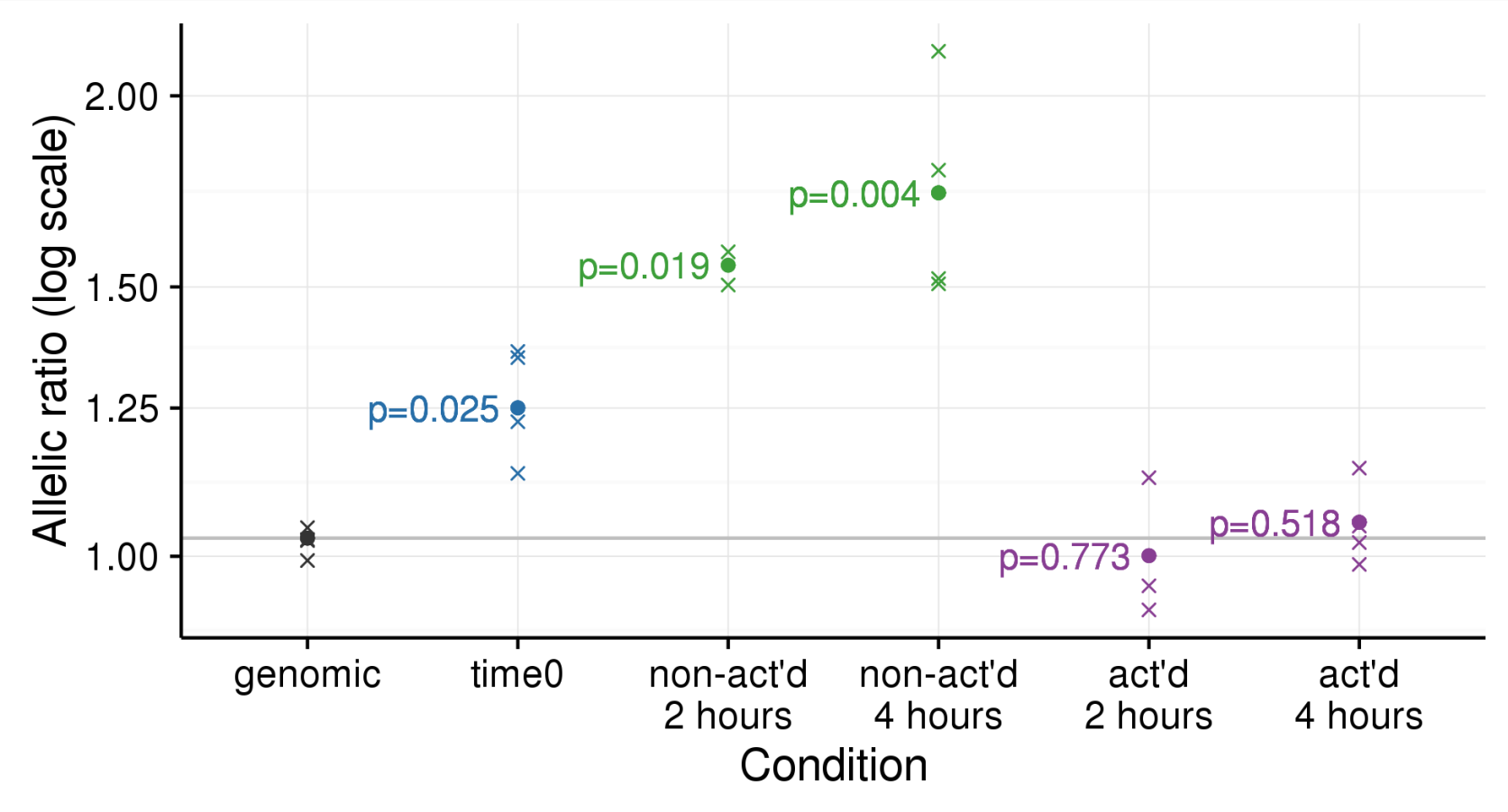


Allele specific expression: quantify relative expression of two chromosomes using targetted PCR and sequencing



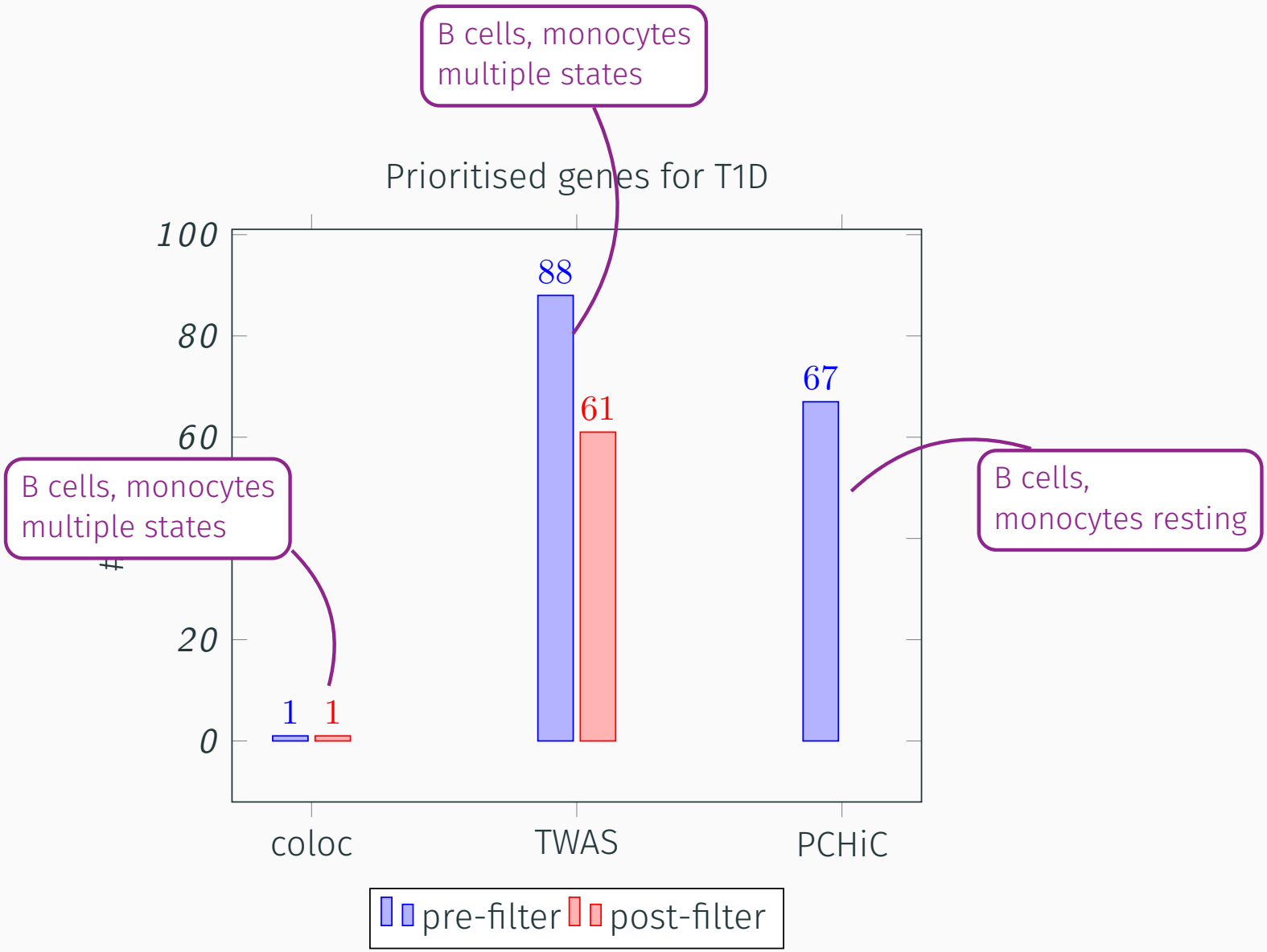
Wallace et al, PLoS Genet 2015; Burren et al, Gen Biol 2017

Allele specific expression confirms effect of group A variants

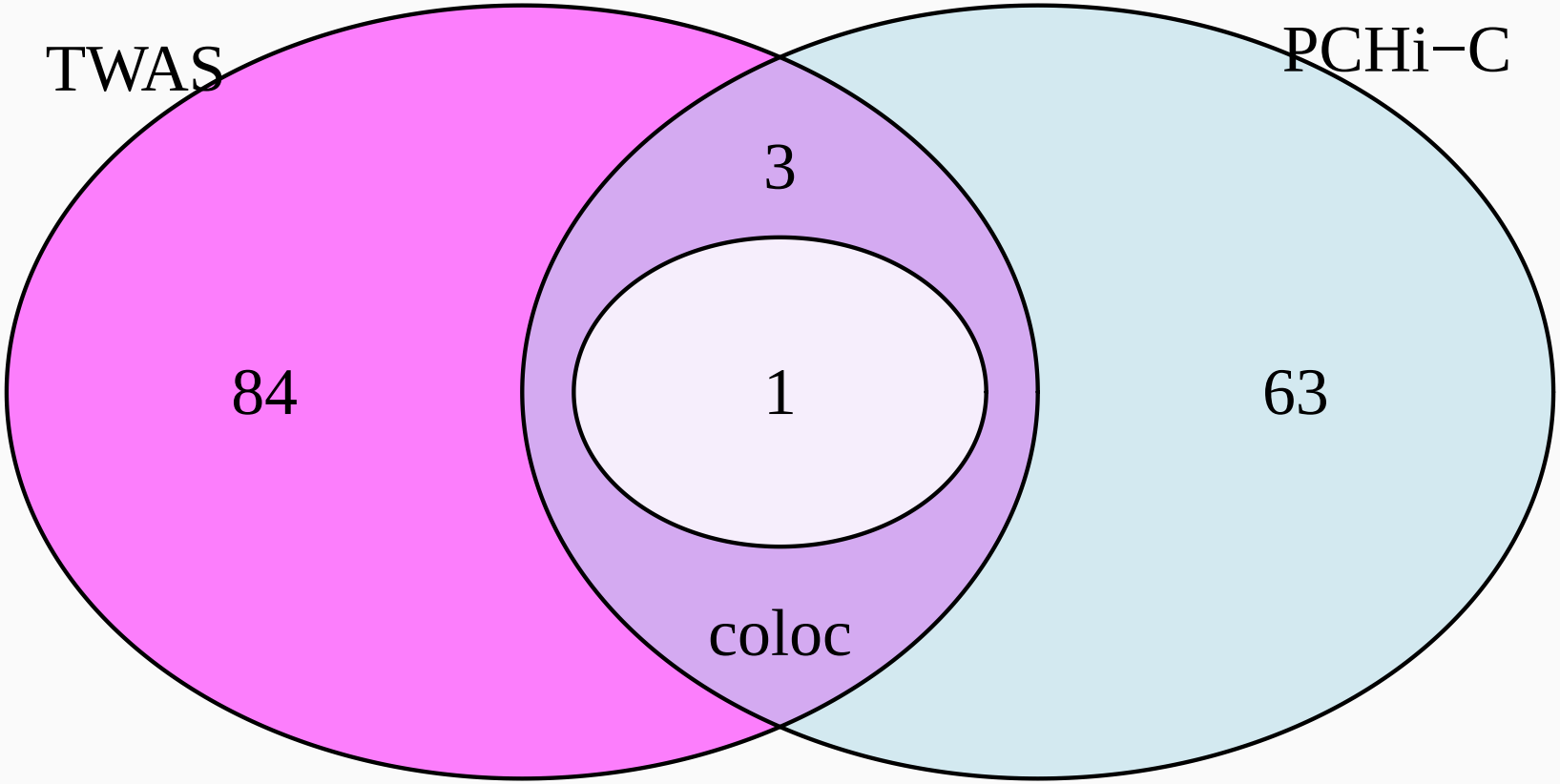


Burren et al, Gen Biol 2017

Summary numbers: one disease, similar cell types



Summary numbers: one disease, similar cell types



Summary of methods

Colocalisation

are GWAS and eQTL signals “the same”?

Close to sufficient for causality

very stringent hypothesis, possibility of false negatives

TWAS

what genes are (predicted to be) differentially expressed in patients? appear to be lots!

Possibly necessary, **NOT** sufficient evidence of causality

PCHi-C methods

does a GWAS variant lie in a region which contacts a gene promoter?

useful circumstantial evidence; **NOT** causality

Remaining challenges and opportunities

Cell diversity

- Amazing diversity in gene expression, chromatin binding and chromatin conformation - we only study common, accessible cells!
- eQTL datasets need to be larger, more diverse
- PCHi-C *already* covers greater diversity than eQTL

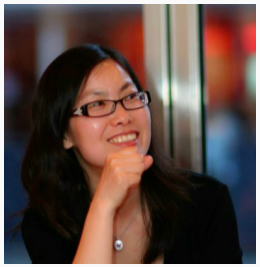
Evidence synthesis

- Different methods, different assumptions $\rightarrow$ different genes

Wider adoption of methods

- Datasets need to be more available
- Methods need to be easier to apply

Thanks to


Hui Guo


Stasia Grinberg


Olly Burren


Dan Rainbow

Diabetes and Inflammation Laboratory
Xaquin Castro Dopico, Ellen Schofield, John Todd, Linda Wicker, Tony Cutler

PCHi-C Collaboration
Biola Javierre, Jonathan Cairns, Mattia Frontini, Willem Ouwehand, Peter Fraser, Mikhail Spivakov

Cambridge NIHR BioResource

Capture Hi-C Plotter: `http://www.chicp.org`

wellcome

UNIVERSITY OF CAMBRIDGE

MRC | Biostatistics Unit