Using epigenetic marks for biological interpretation of complex trait associations

wellcome trust sanger institute

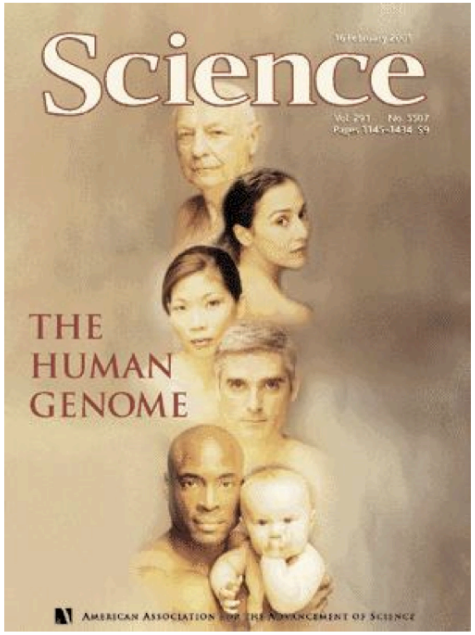
Gosia Trynka

15 January 2016

Understanding disease biology

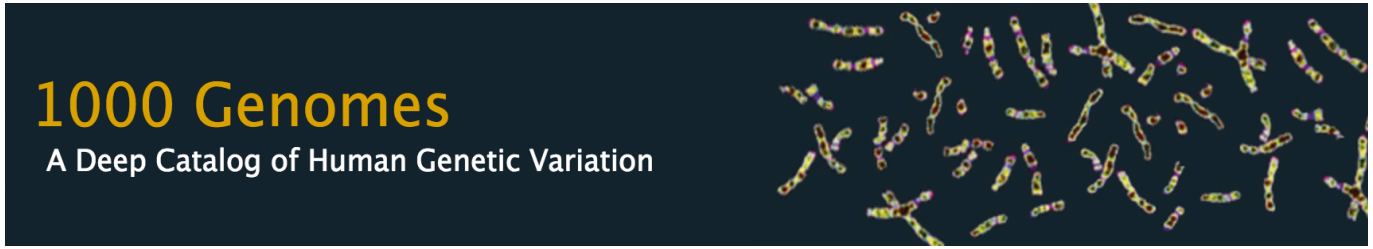
- Using SNPs to map disease loci for complex traits
  - Immune-related diseases, coeliac disease
- From associated variants to function
  - Correlation with gene epxression
  - Causal variants, tissues and pathways

- Sequencing of the human genome



- HapMap project

- 1000 Genomes Project

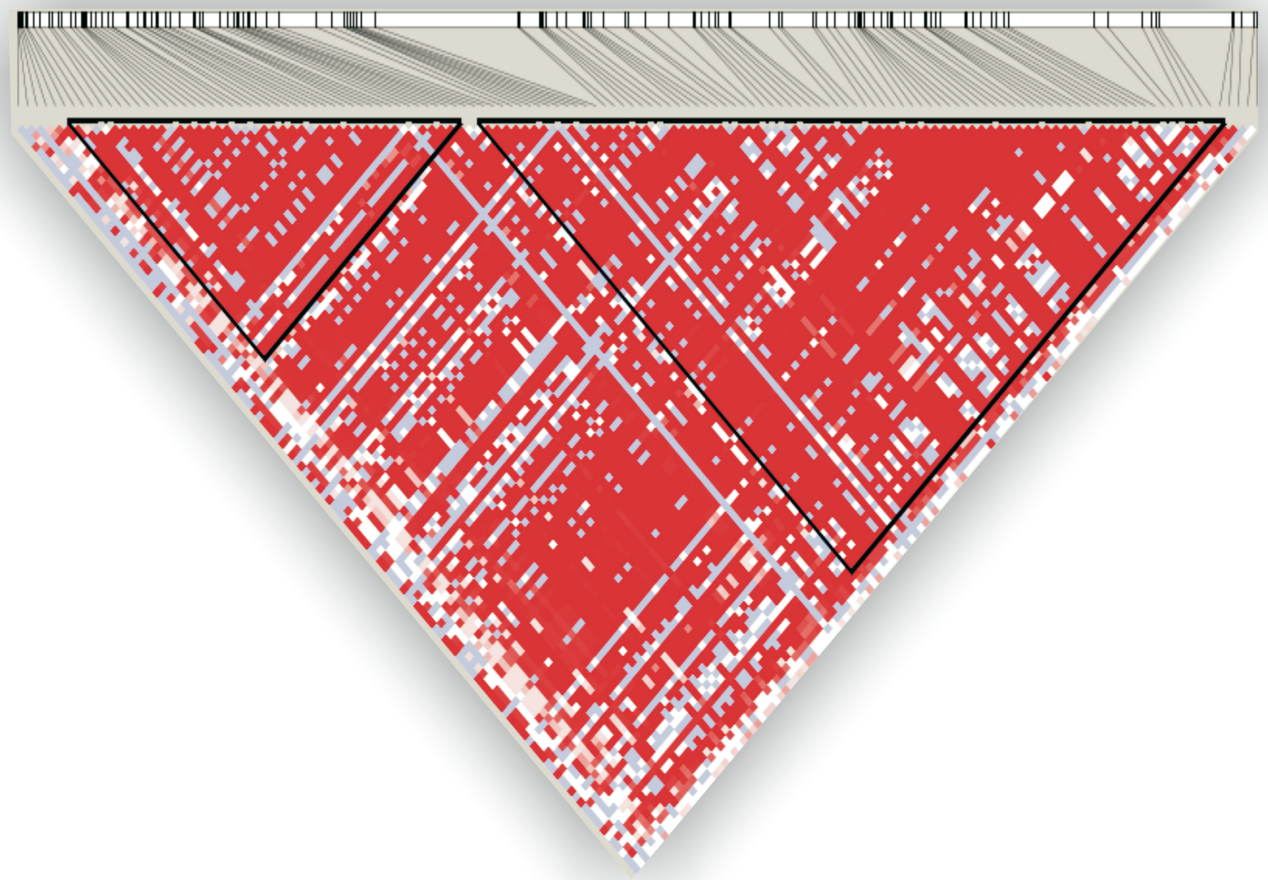


wellcome trust sanger institute

Refernce panels of human variation: HapMap

- Goal: **provide a haplotype map of human genome using common variants**
  - Genotyped 1.6 million common SNPs
  - 1,184 reference individuals
  - from 11 global populations
    - Northern and western Europe (CEU); Han Chinesein Beijing, China (CHB); Japanese in Tokyo, Japan (JPT); Yoruba in Ibadan, Nigeria (YRI); African ancestry in the southwestern USA (ASW); Chinese in metropolitan Denver, Colorado, USA (CHD); Gujarati Indians in Houston, Texas, USA (GIH); Luhya in Webuye, Kenya (LWK); Maasai in Kinyawa, Kenya (MKK); Mexican ancestry in Los Angeles, California, USA (MXL); Tuscans in Italy (TSI)

Correlation of genetic variants - LD blocks



LD – your best friend, your greatest enemy

wellcome trust sanger institute

3 billion base pairs in the human genome

wellcome trust sanger institute

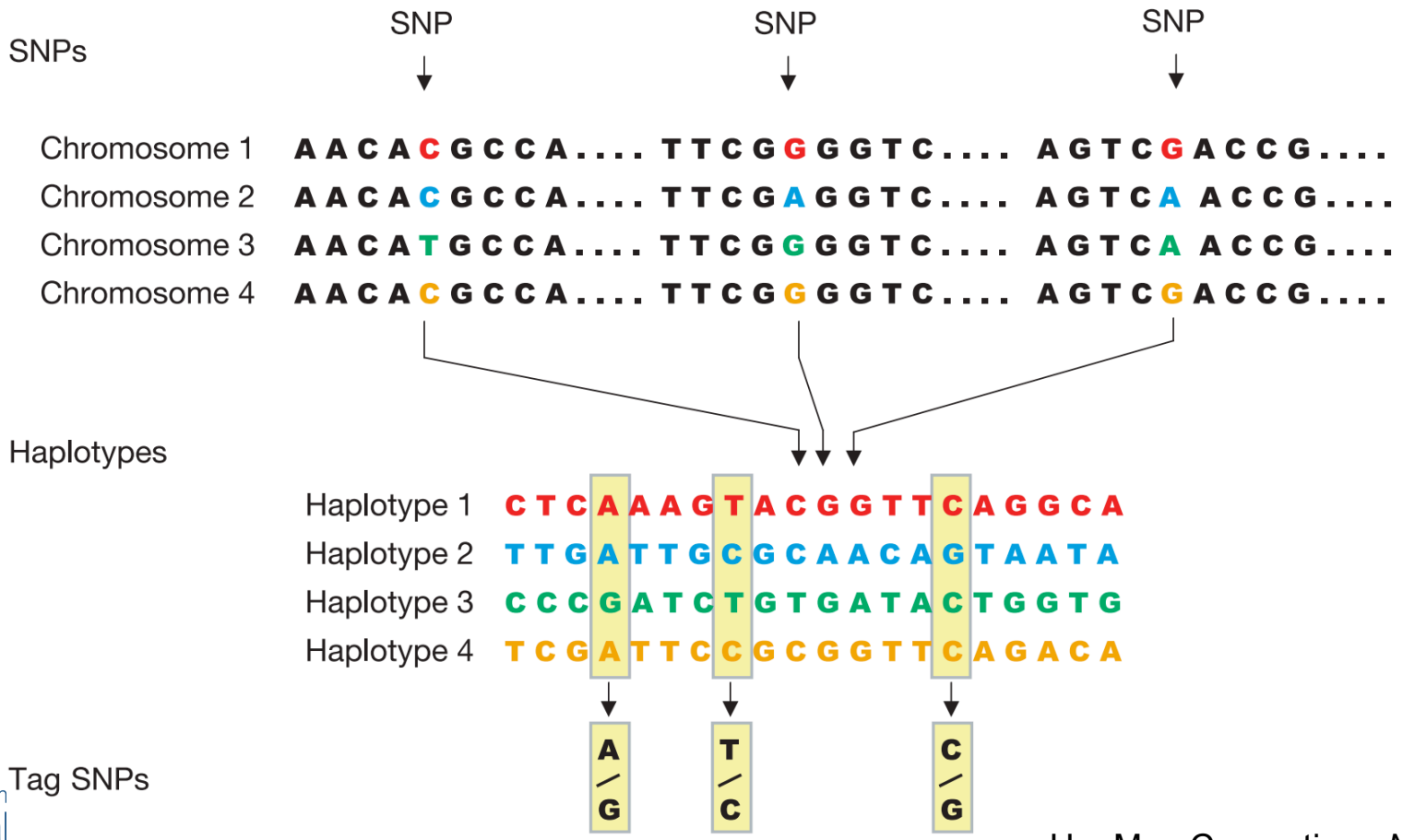
4 million common

SNPs

	SNP ↓	SNP ↓	SNP ↓
Chromosome 1	A A C A **C** G C C A	T T C G **G** G G T C	A G T C **G** A C C G . . . .
Chromosome 2	A A C A **C** G C C A	T T C G **A** G G T C	A G T C **A** A C C G . . . .
Chromosome 3	A A C A **T** G C C A	T T C G **G** G G T C	A G T C **A** A C C G . . . .
Chromosome 4	A A C A **C** G C C A	T T C G **G** G G T C	A G T C **G** A C C G . . . .

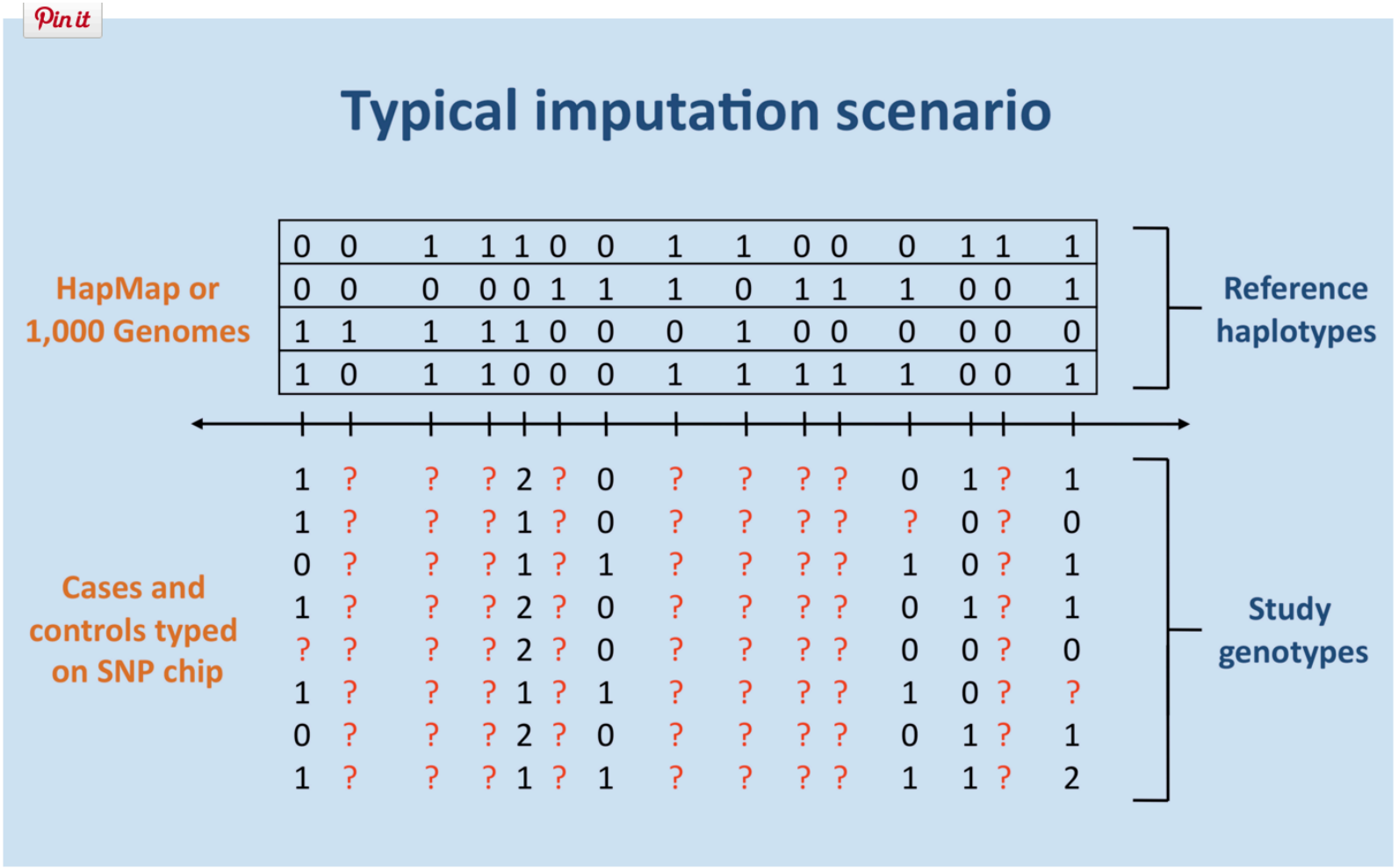
wellcome trust sanger institute

500k tagging SNPs



HapMap Consortium, *Nature* 2003

Imputing missing genotypes



wellcome trust sanger institute

https://mathgen.stats.ox.ac.uk/impute/impute_v2.html

Refernce panels of human variation: 1000 Genomes Project

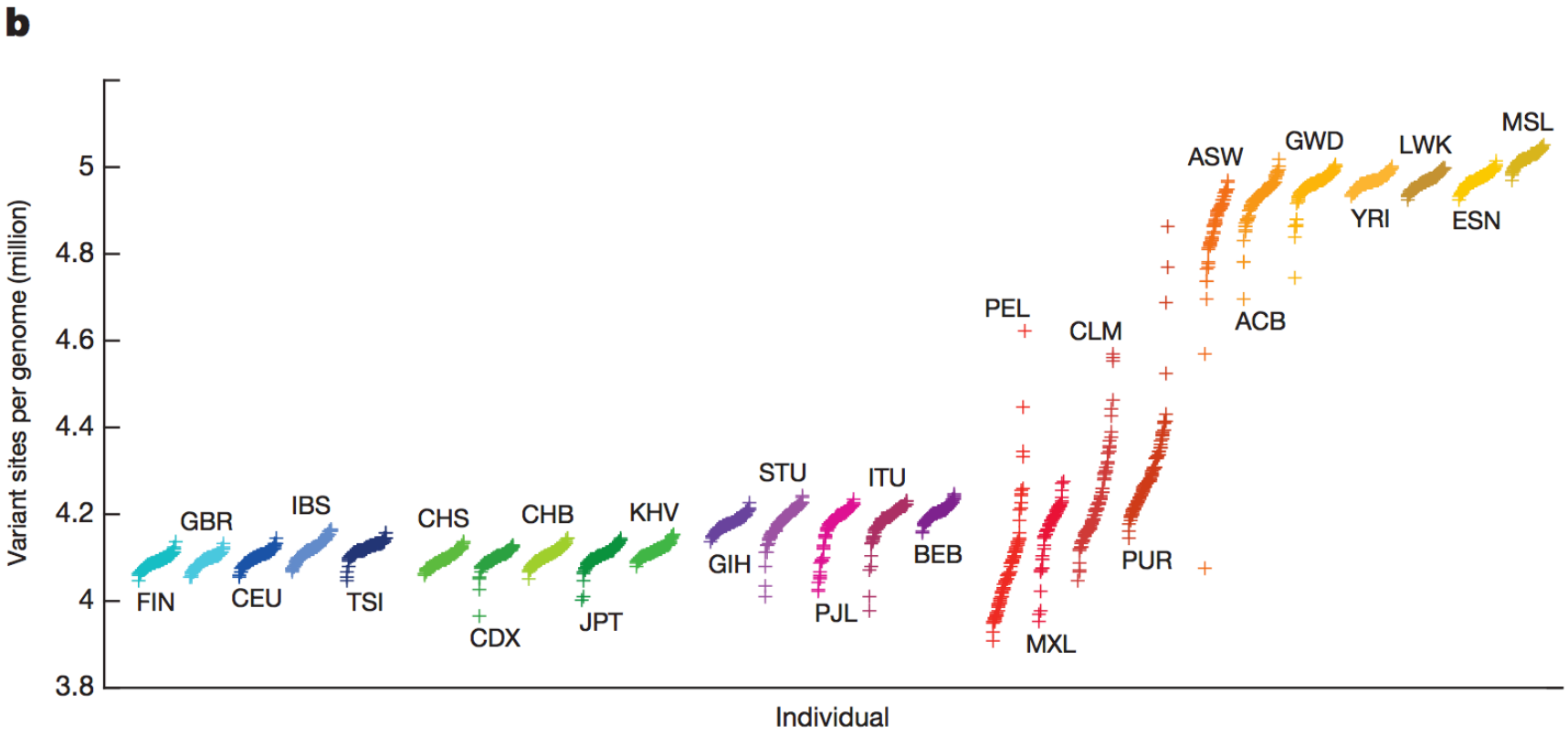
- 2,504 individuals from 26 populations using a combination of low-coverage whole-genome sequencing, deep exome sequencing, and dense microarray genotyping

- > 88 million SNPs, 3.6 million indels, and 60,000 structural variants
- phased onto high-quality haplotypes
- >99% of SNP variants with a frequency of >1% for a variety of ancestries

wellcome trust sanger institute

1000 Genomes Project, *Nature* 2015

The total number of observed non-reference sites differs greatly among populations



1000 Genomes Project, *Nature* 2015

Putatively fuctional variation

- The majority of variants in the data set are rare but the majority of variants observed in a single genome are common
- A typical genome contained 149–182 sites with protein truncating variants
- ~2,000 variants per genome associated with complex traits through genome-wide association studies (GWAS)
- 24–30 variants per genome implicated in rare disease through ClinVar

wellcome trust sanger institute

1000 Genomes Project, *Nature* 2015

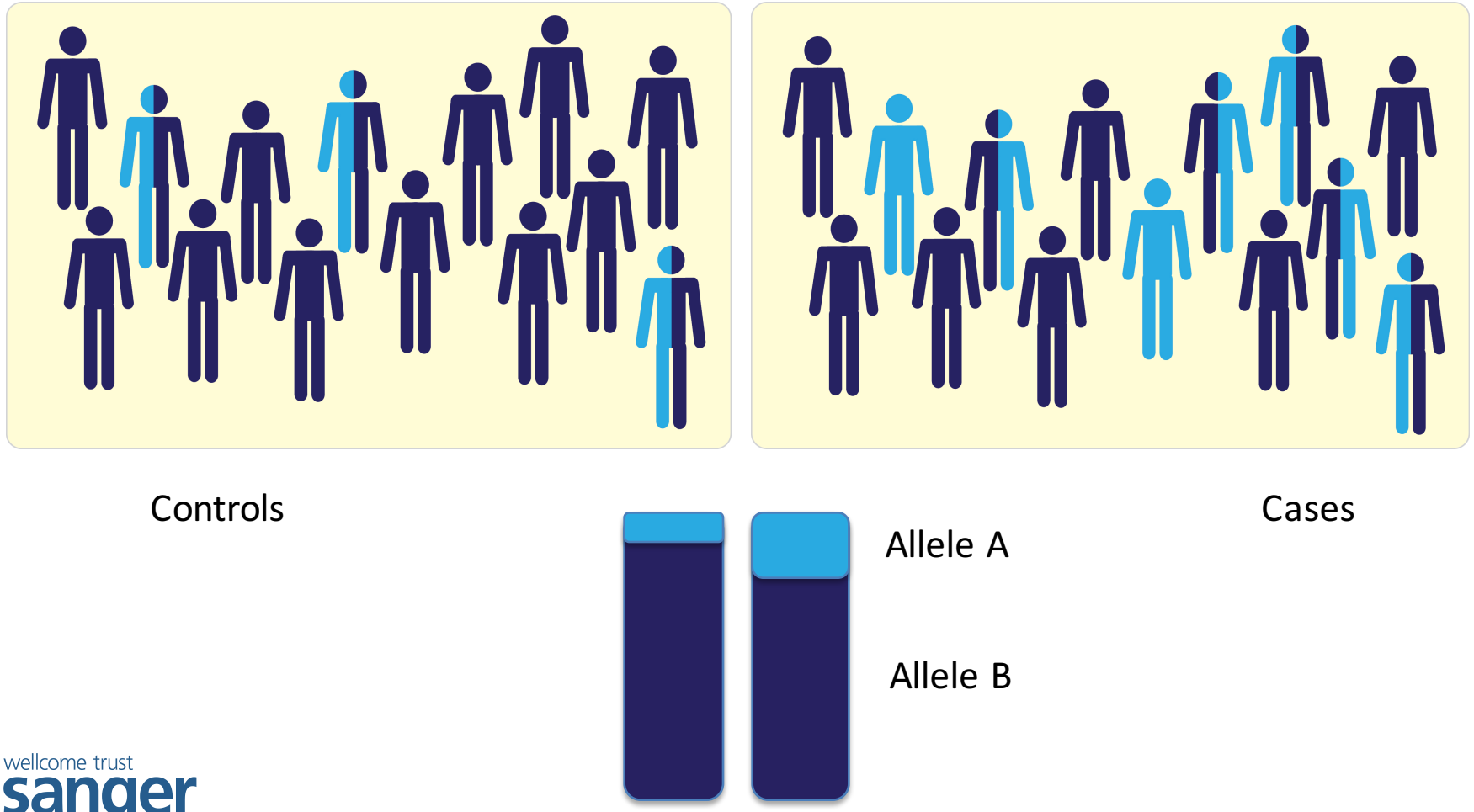
Haplotype Reference Consortium

- http://www.haplotype-reference-consortium.org
- The first release consists of 64,976 haplotypes at 39,235,157 SNPs, all with an estimated minor allele count of >= 5

wellcome trust sanger institute

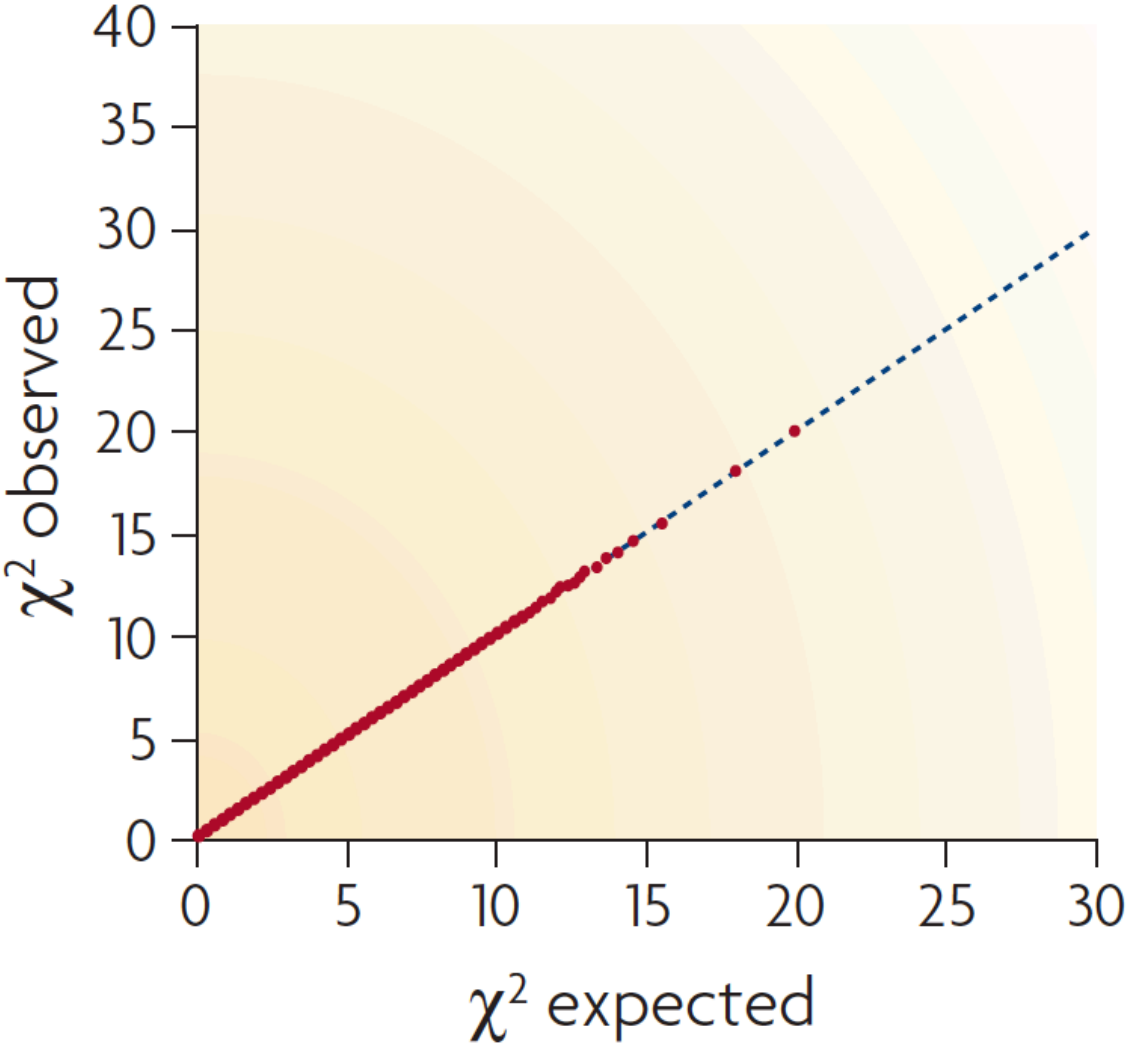
Genome wide association study (GWAS)

- GWAS compares allele frequency at common SNPs between cohorts of cases and controls



wellcome trust sanger institute

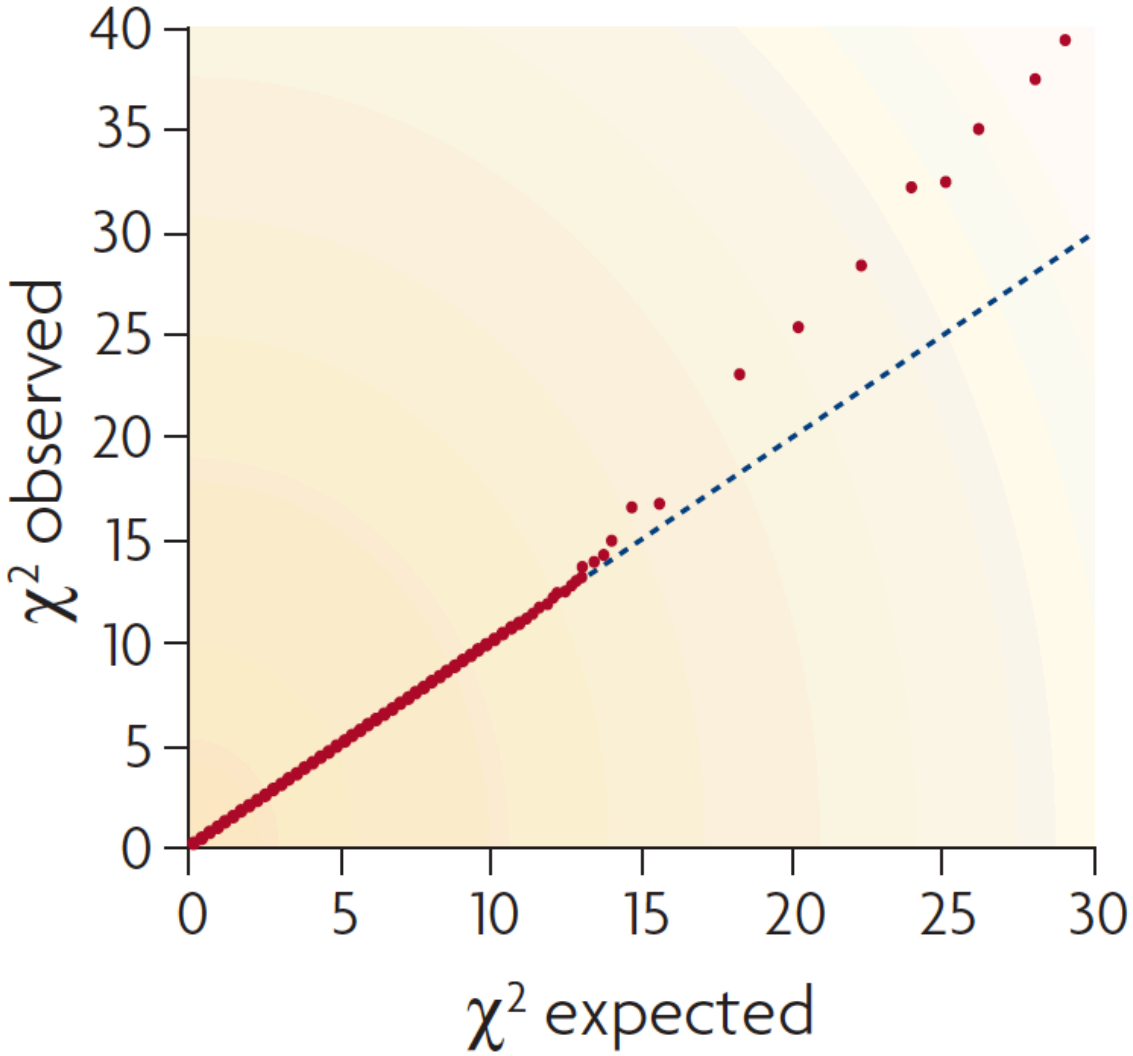
Quantile-quantile (Q-Q) plots



McCarthy et al., *Nat Rev Genet 2008*

wellcome trust sanger institute

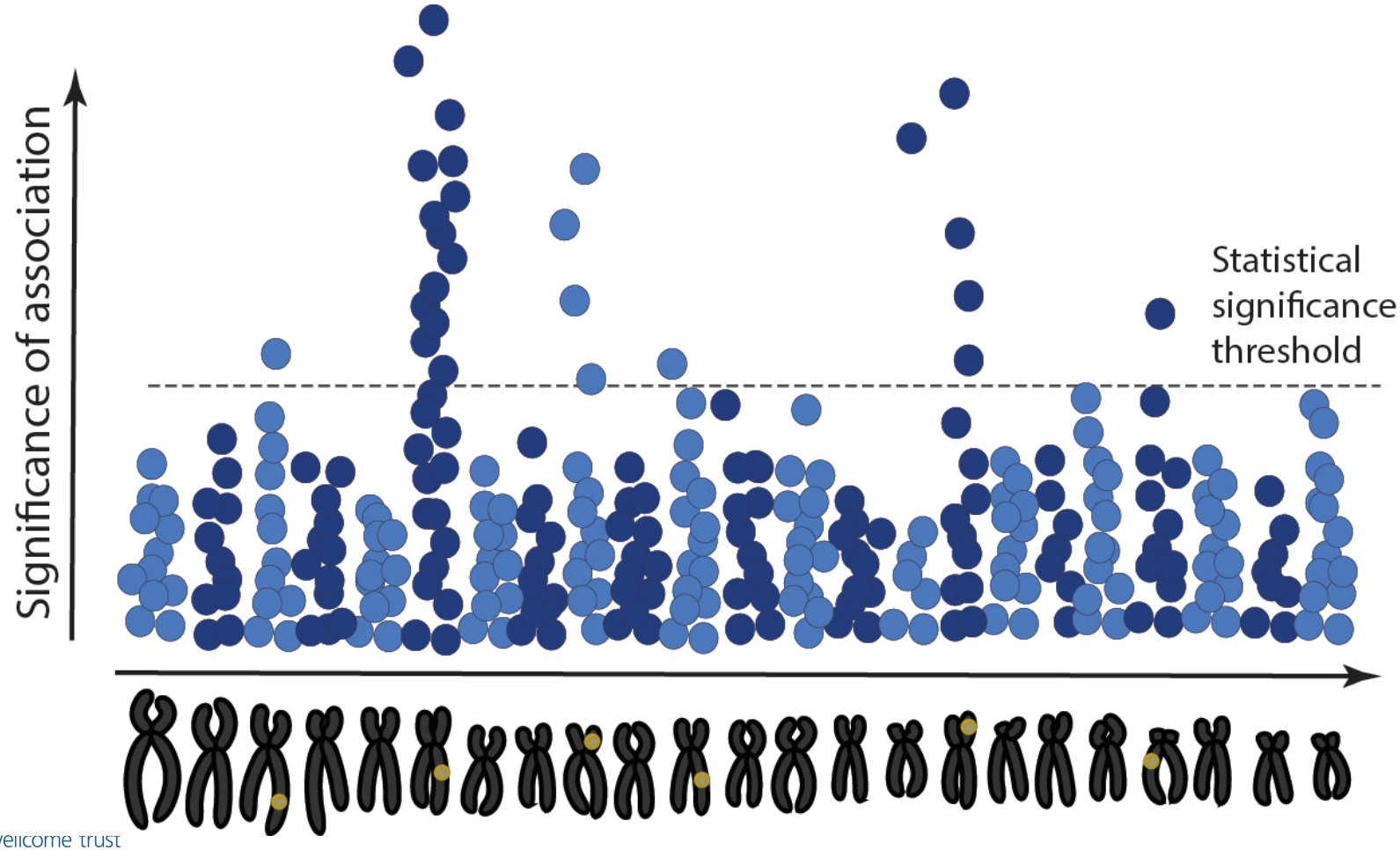
Quantile-quantile (Q-Q) plots



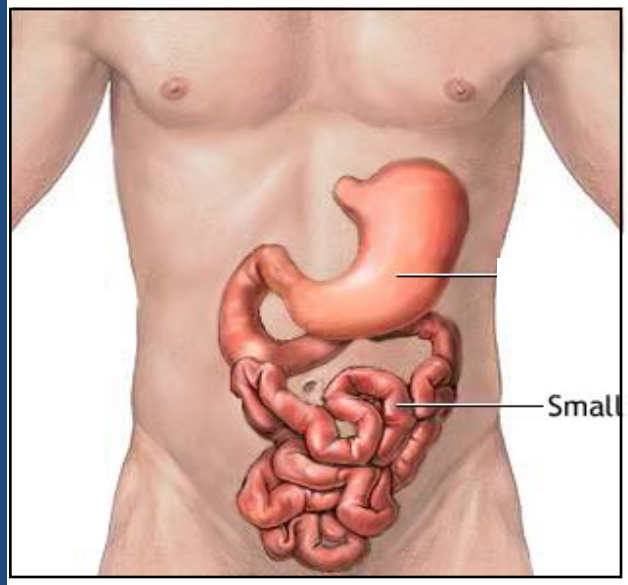
McCarthy et al., *Nat Rev Genet 2008*

wellcome trust sanger institute

‘Manhattan’ plot represents significantly associated regions in the genome



Celiac disease – autoimmune disease of the small intestine



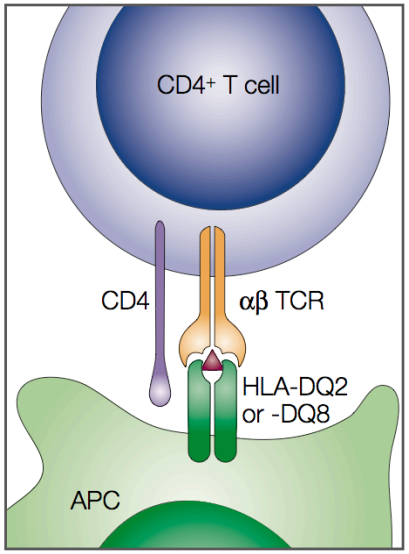
The most common food intolerance in western populations **(~1%)**

wellcome trust sanger institute

Unique understanding of pathogenesis of coeliac disease

Wheat, Barley, Rye

+



HLA DQ2/DQ8

+ More genes? = Celiac disease

HLA accounts for about 35% of the genetic variance
The same genotypes are present in 30-40% of general population

wellcome trust sanger institute

Mapping risk variants for coeliac disease

- 800 cases and 1,400 controls from UK

nature genetics

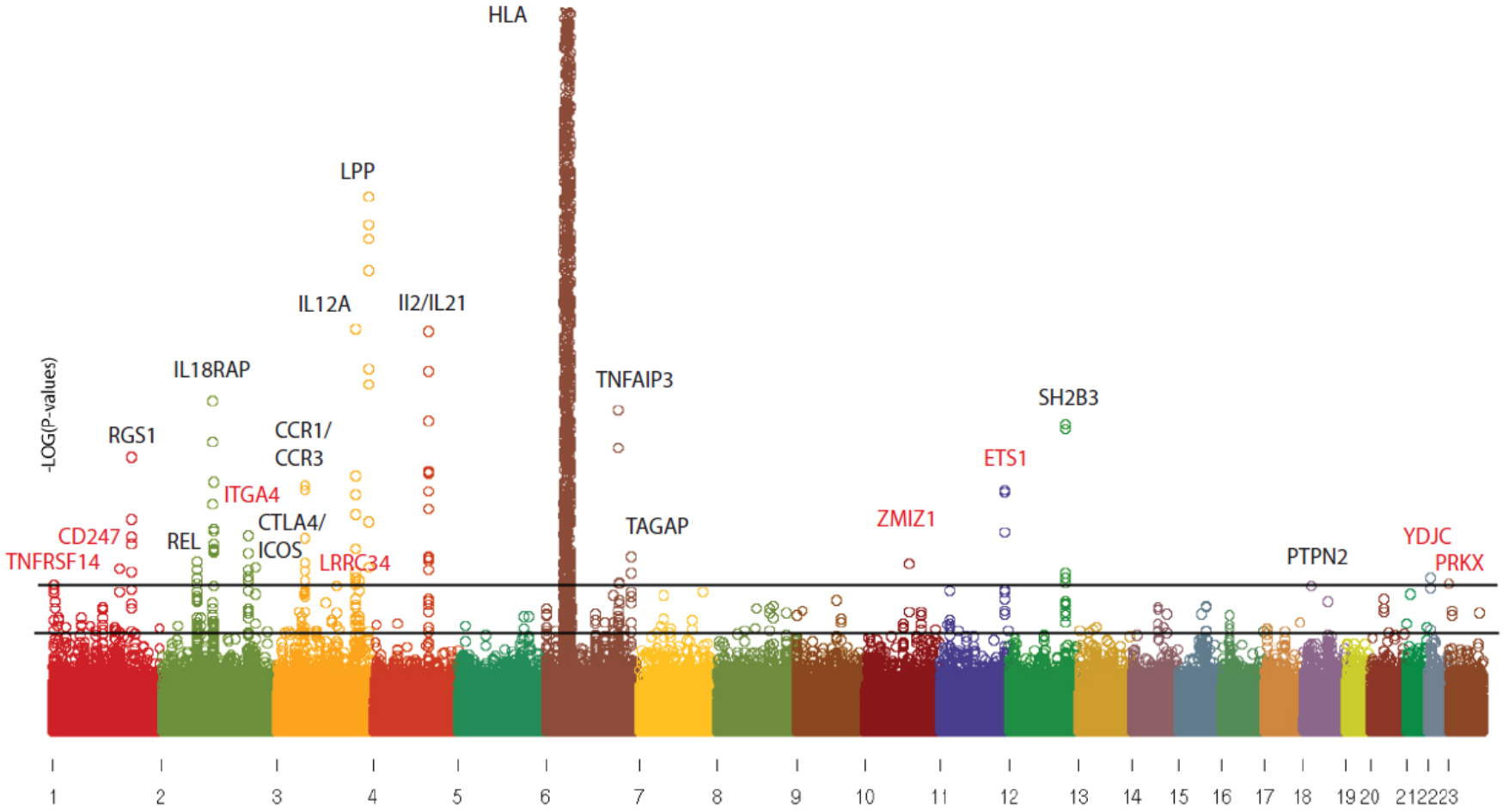
A genome-wide association study for celiac disease identifies risk variants in the region harboring *IL2* and *IL21*

Van Heel, *Nat Genet* 2007

- Increased power with increased sample size and denser SNP genotyping
  - 4,533 cases and 10,750 controls from four populations
  - Top 131 SNPs genotyped in a replication cohort (10,602 samples, 7 populations)

wellcome trust sanger institute

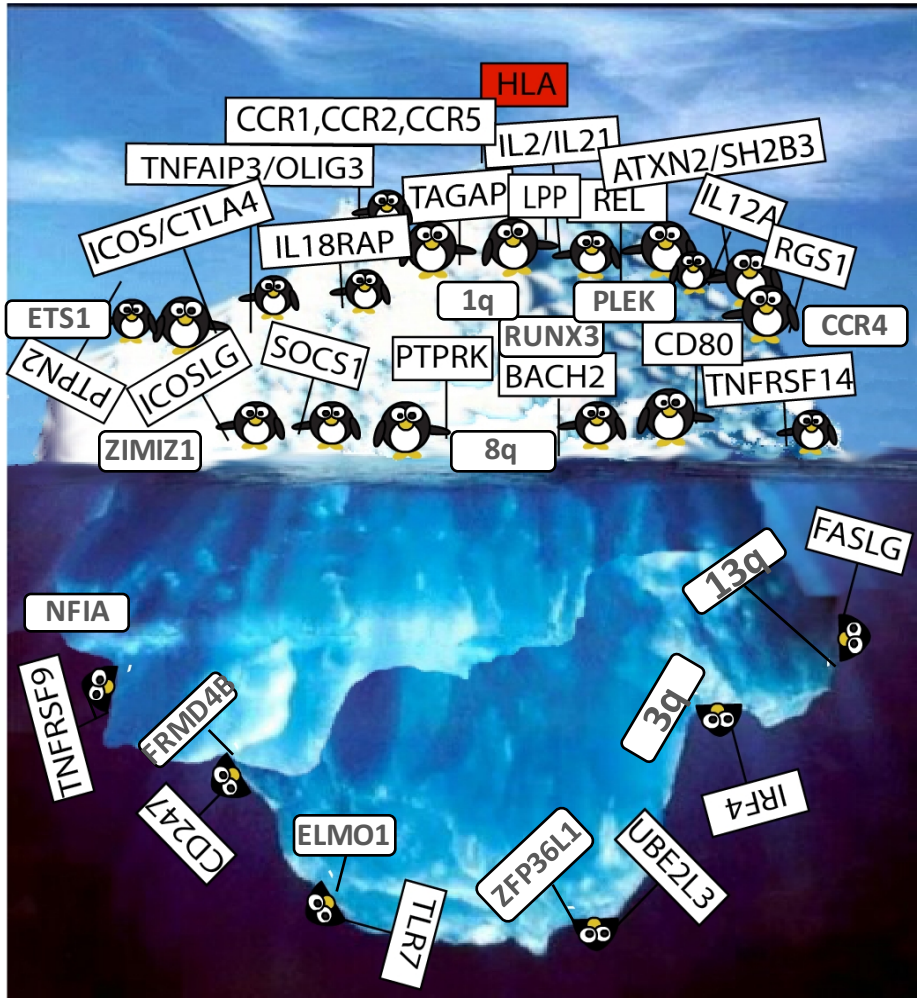
GWAS on 10k cases and 16k controls from 10 populations – 26 associated regions



wellcome trust sanger institute

Trynka and Dubois *et al.*, *Nat Genetics* 2010

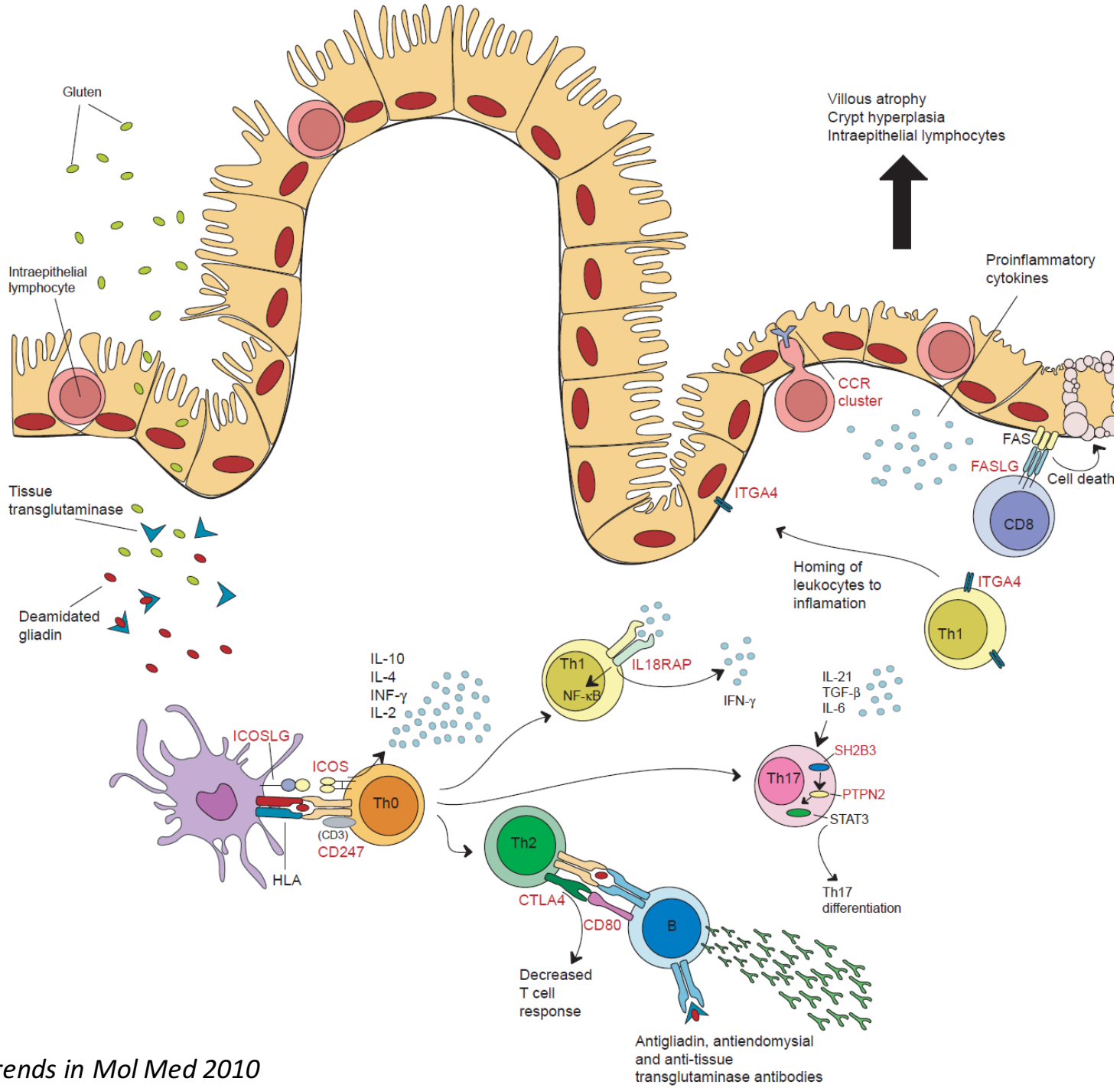
HLA and 26 loci explain 40% of genetic risk



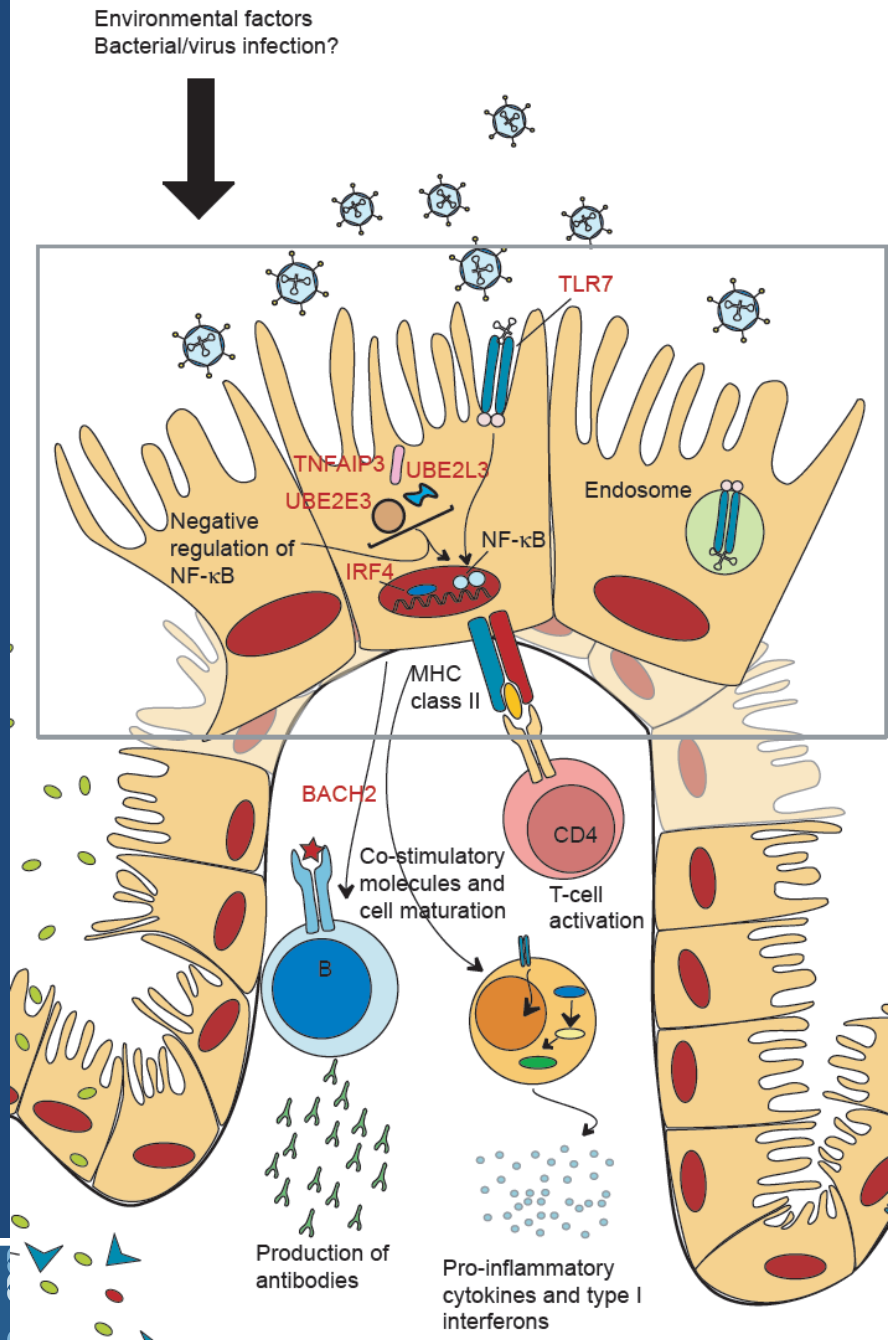
wellcome trust sanger institute

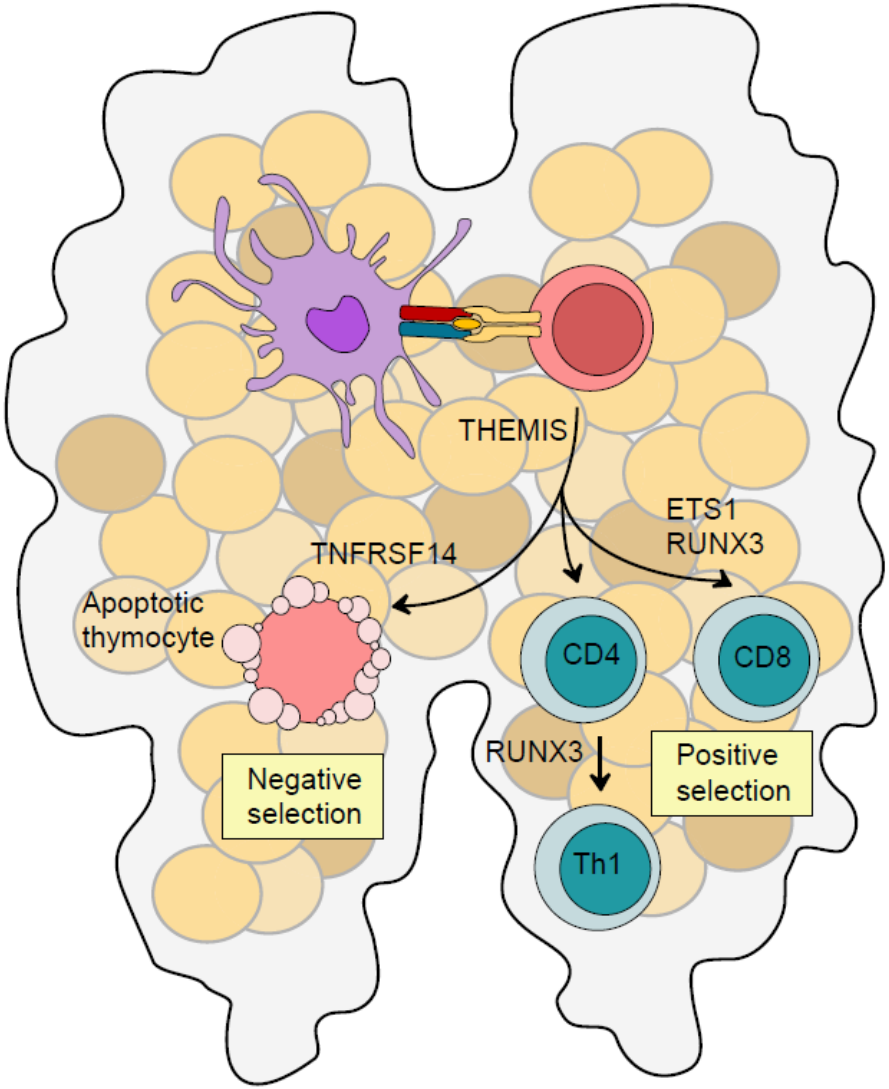
- Genes in proximity of associated variants confirm known and implicate novel pathways

wellcome trust sanger institute



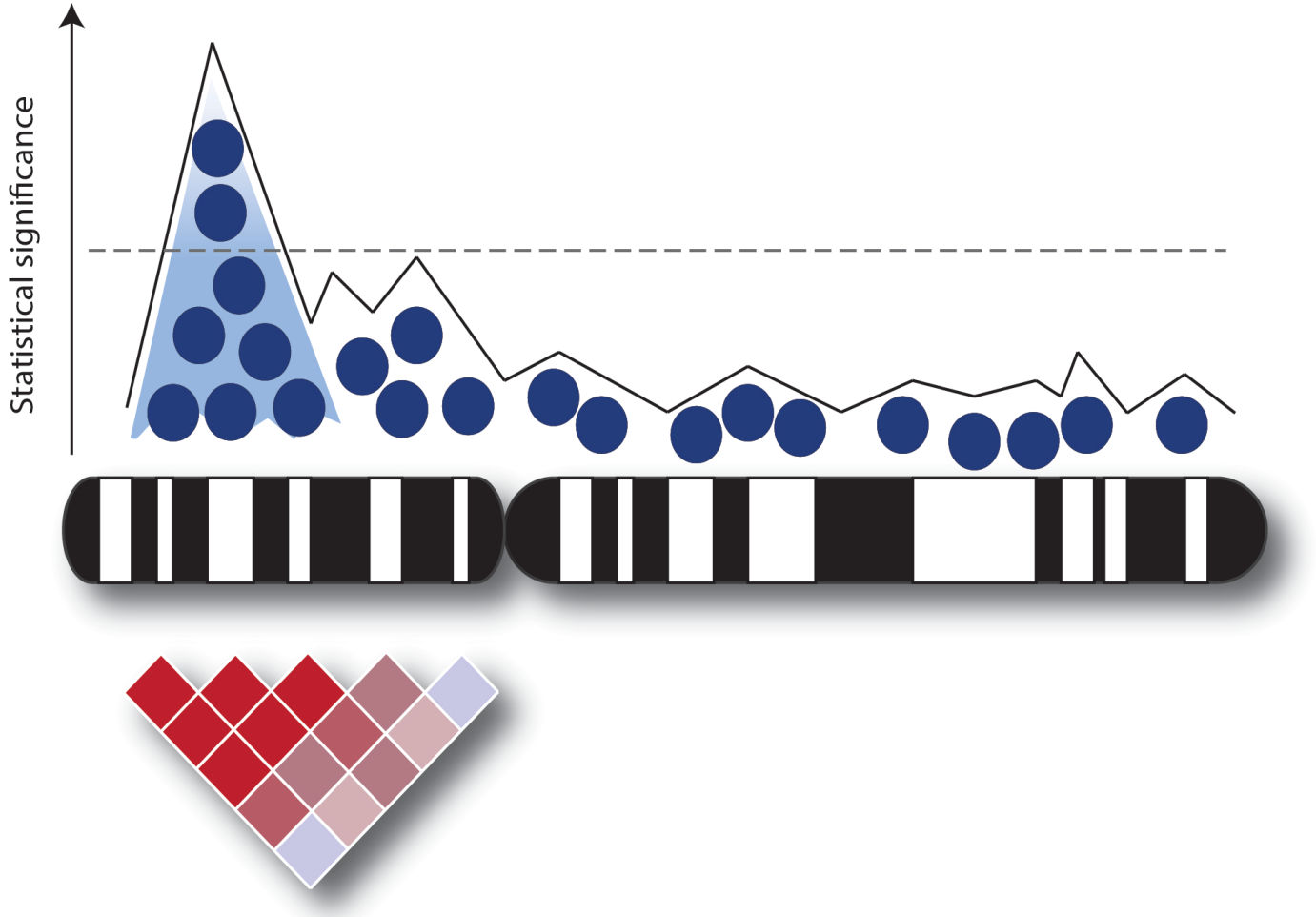
Trynka et al Trends in Mol Med 2010





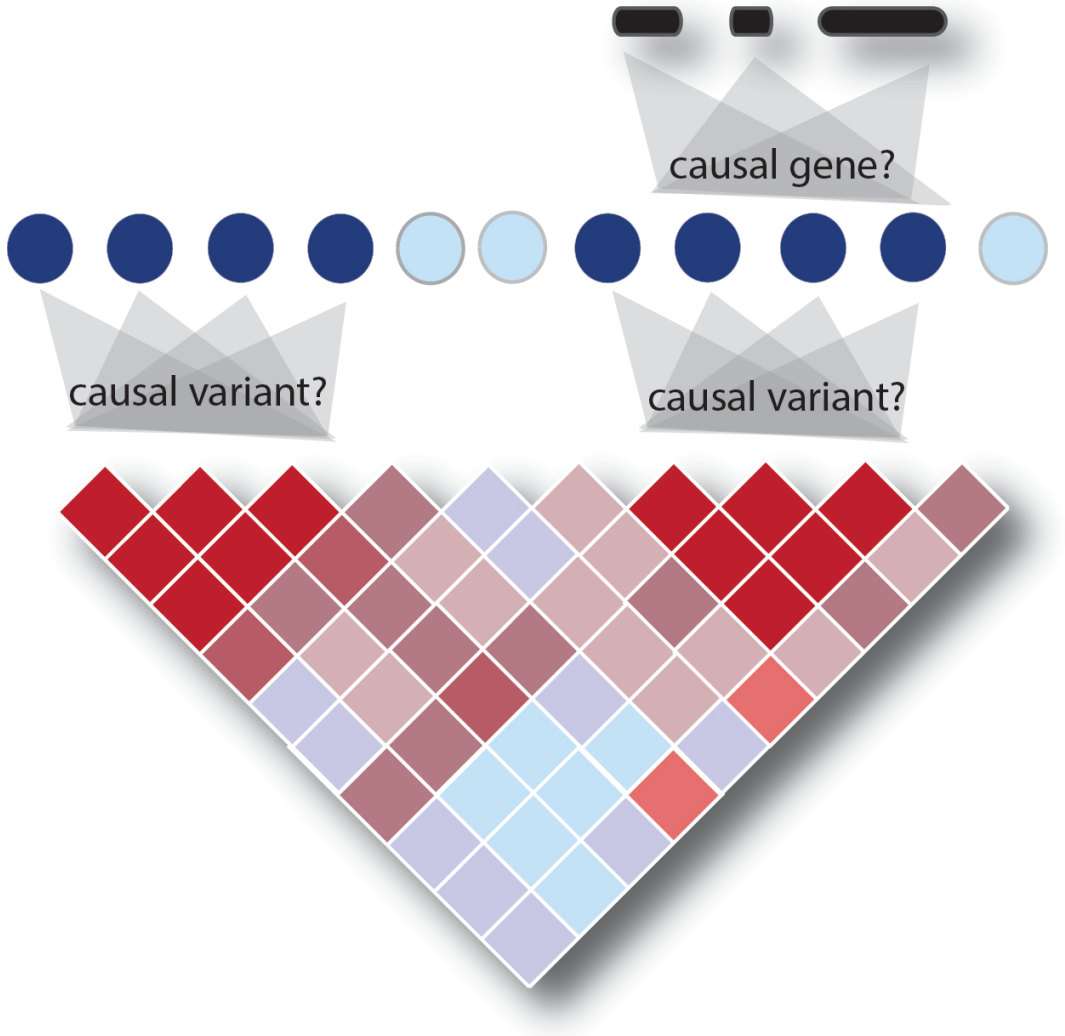
Trynka et al. Trends in Mol Med 2010

Associated variants fall in the regions of extended LD



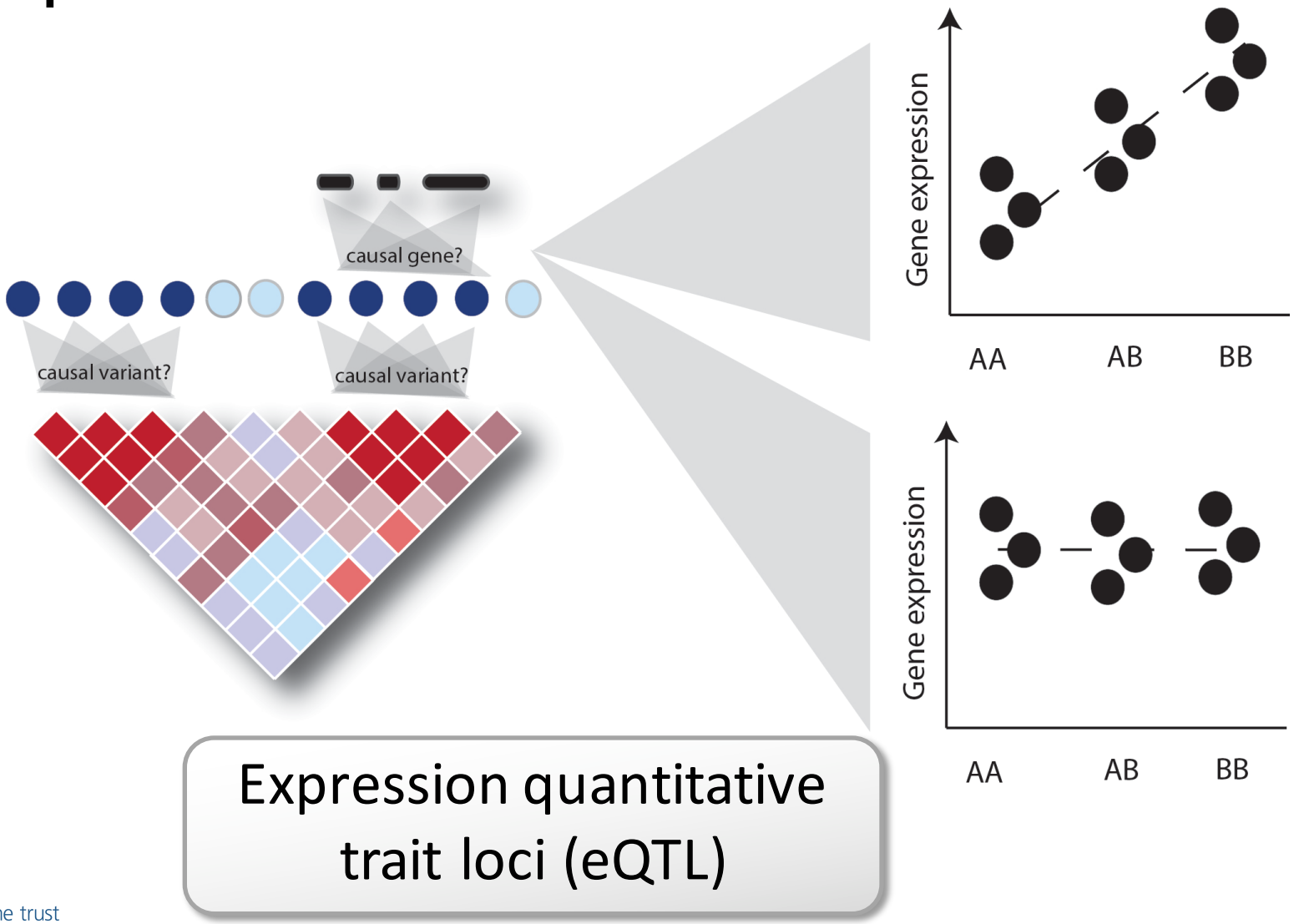
wellcome trust sanger institute

The challenge to identify disease relevant genes



wellcome trust sanger institute

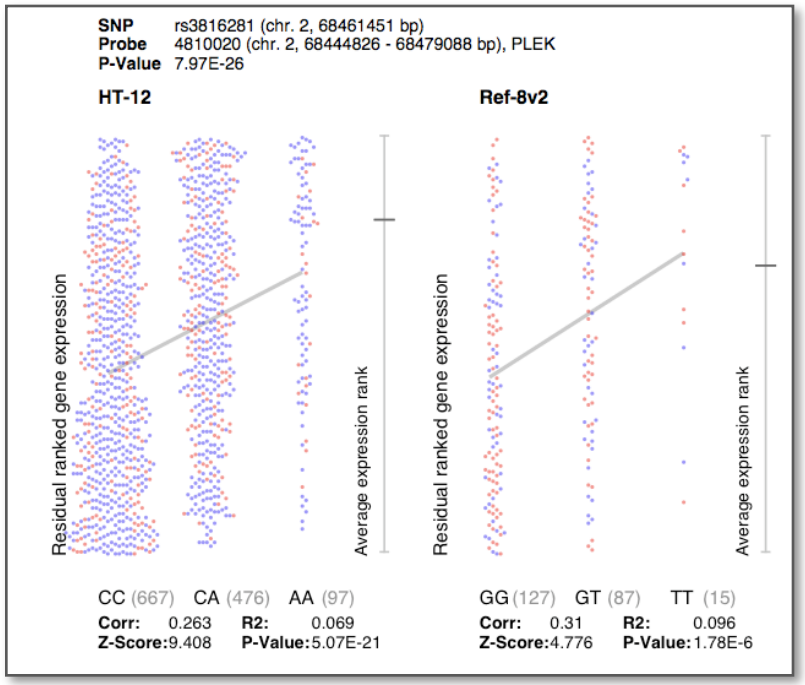
Correlating SNP genotypes with gene expression

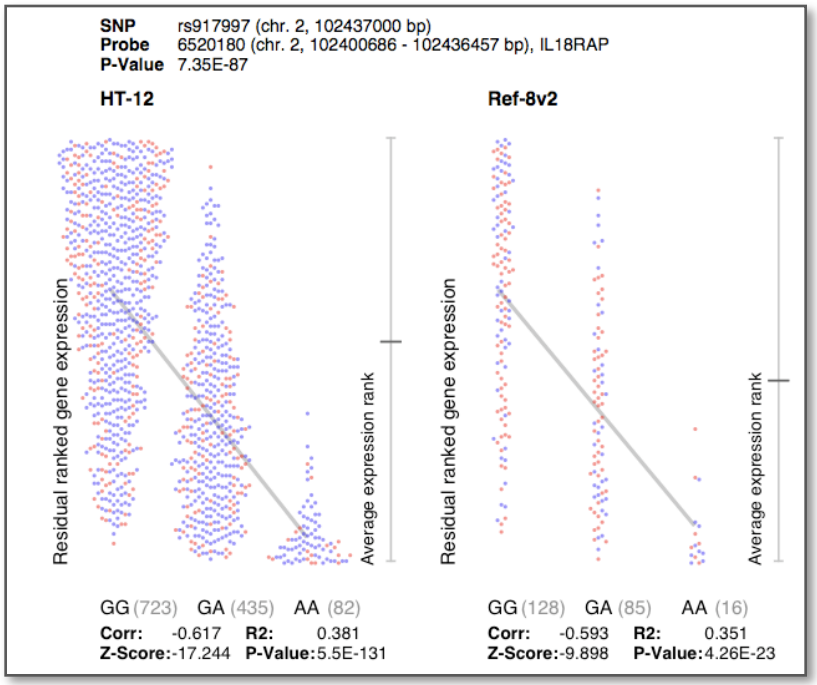


Expression quantitative trait loci (eQTL)

wellcome trust sanger institute

Genotype-gene expression correlation across 1500 samples



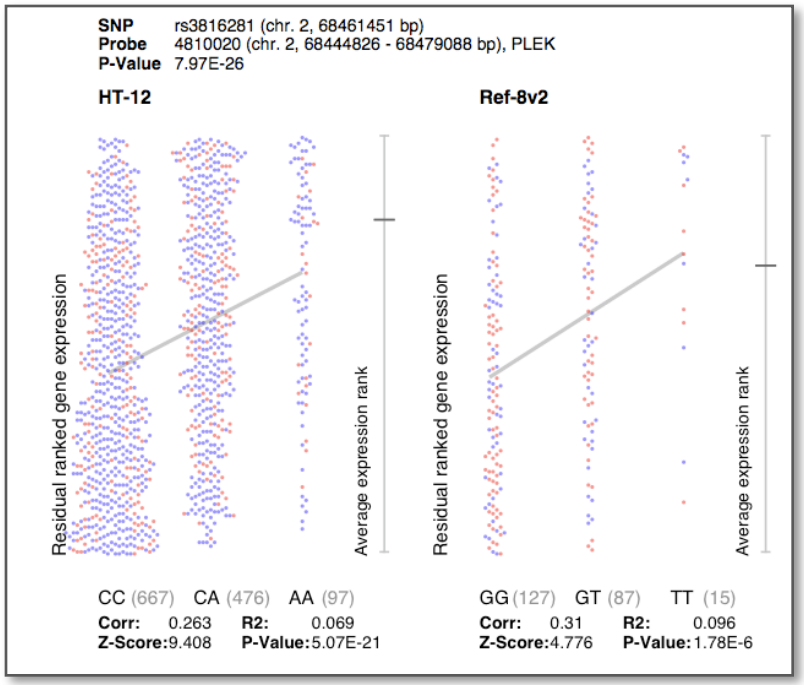


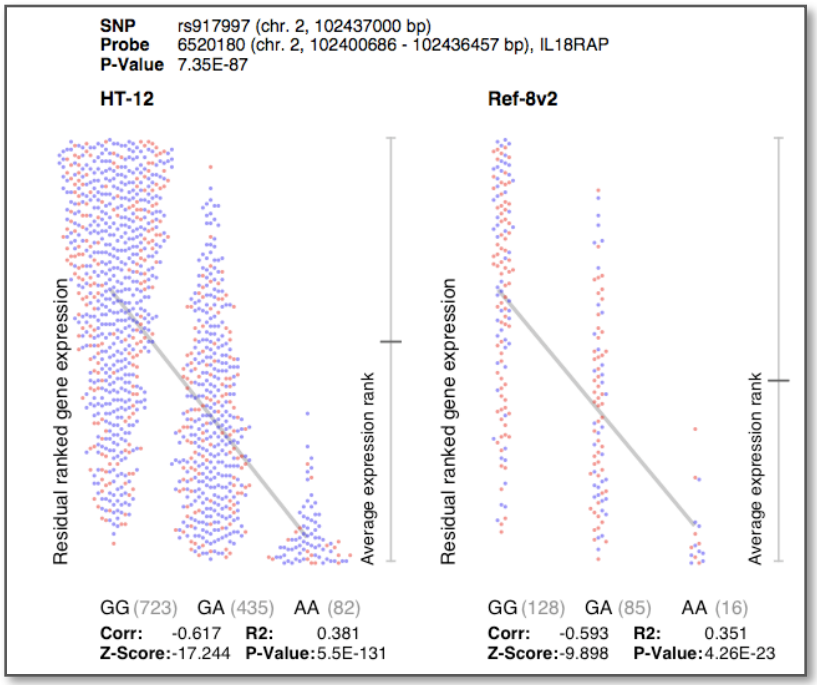
wellcome trust sanger institute

Lude Franke, UMCG, the Netherlands

http://genenetwork.nl/

Genotype-gene expression correlation across 1500 samples

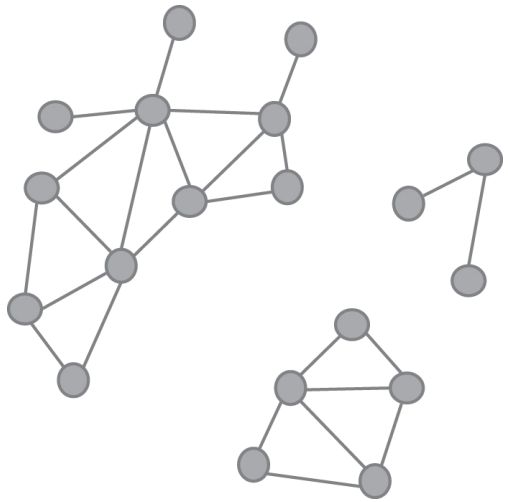




53% of associated variants influence gene expression

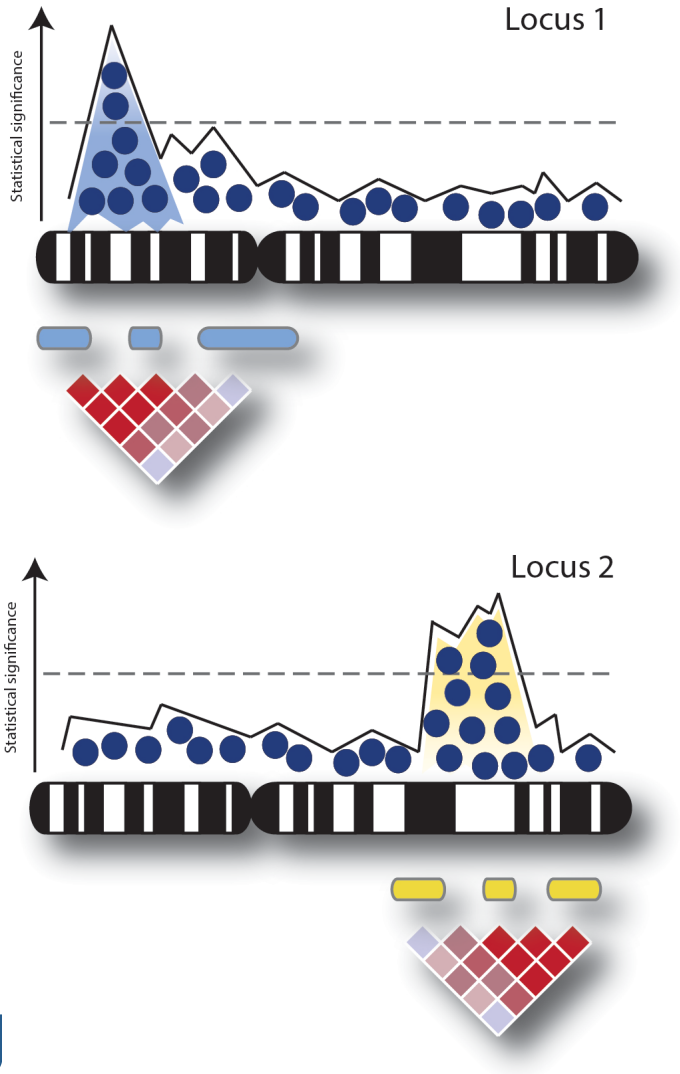
wellcome trust sanger institute

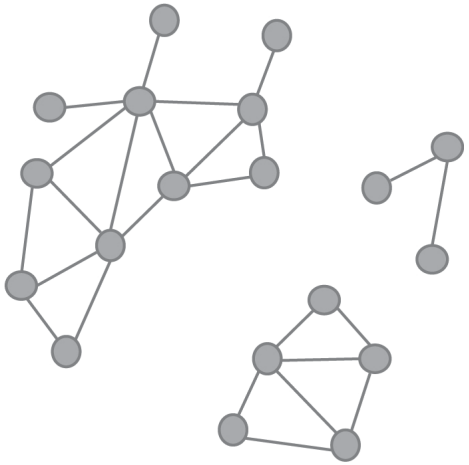
Genes co-expressed can indicate biologically functional modules



wellcome trust sanger institute

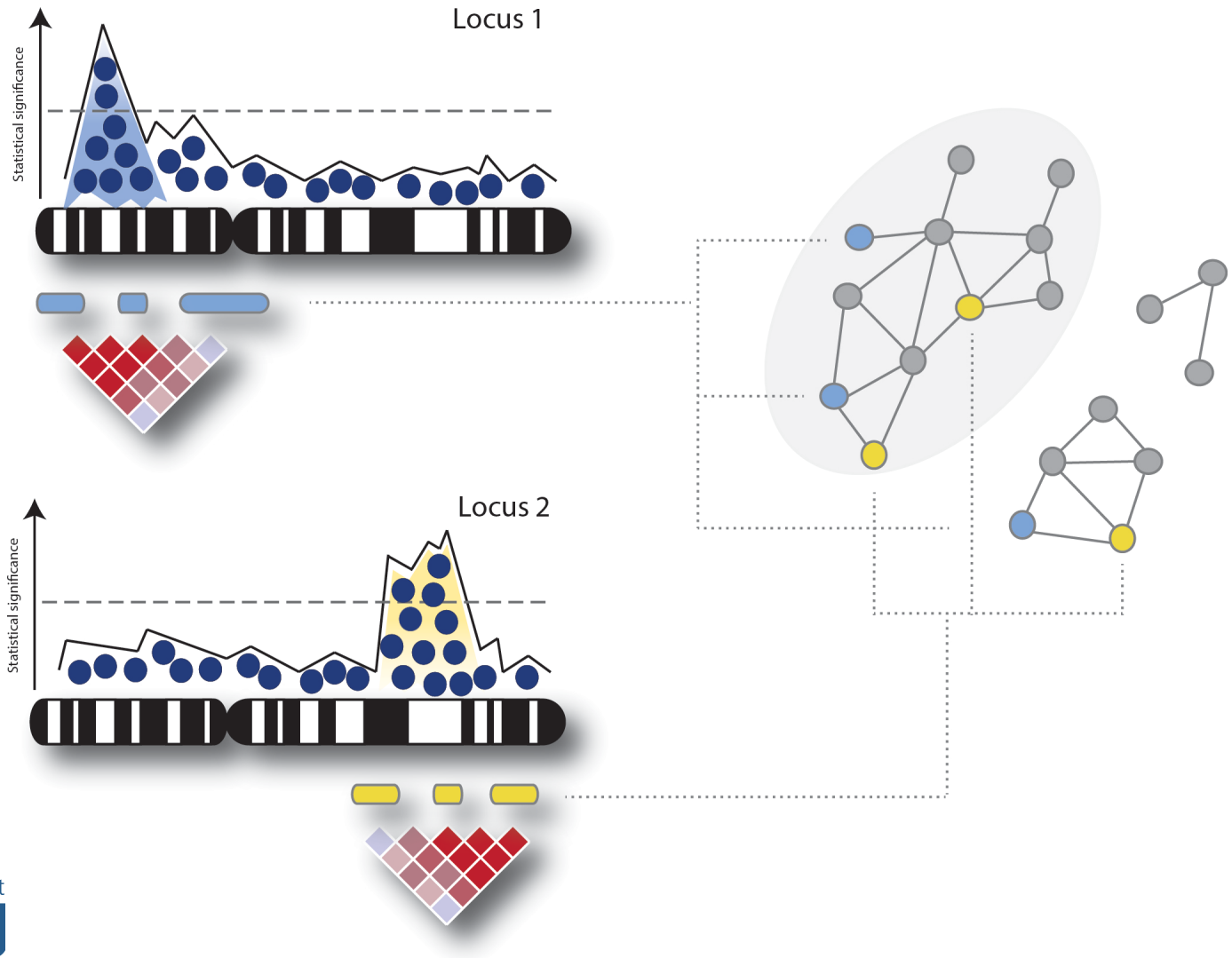
Genes co-expressed can indicate biologically functional modules



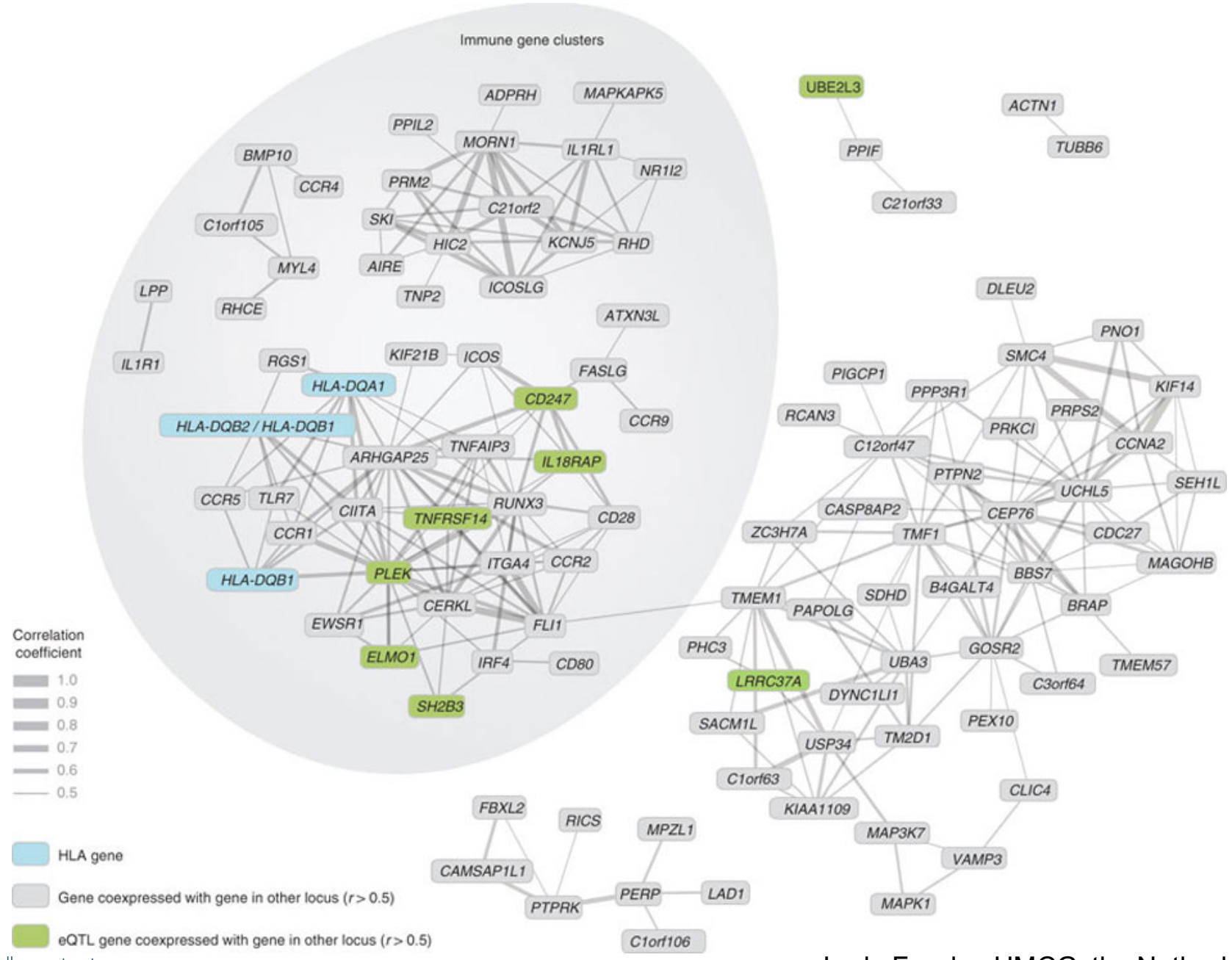


wellcome trust sang institute

Genes co-expressed can indicate biologically functional modules



wellcome trust sang institute



wellcome trust sanger institute

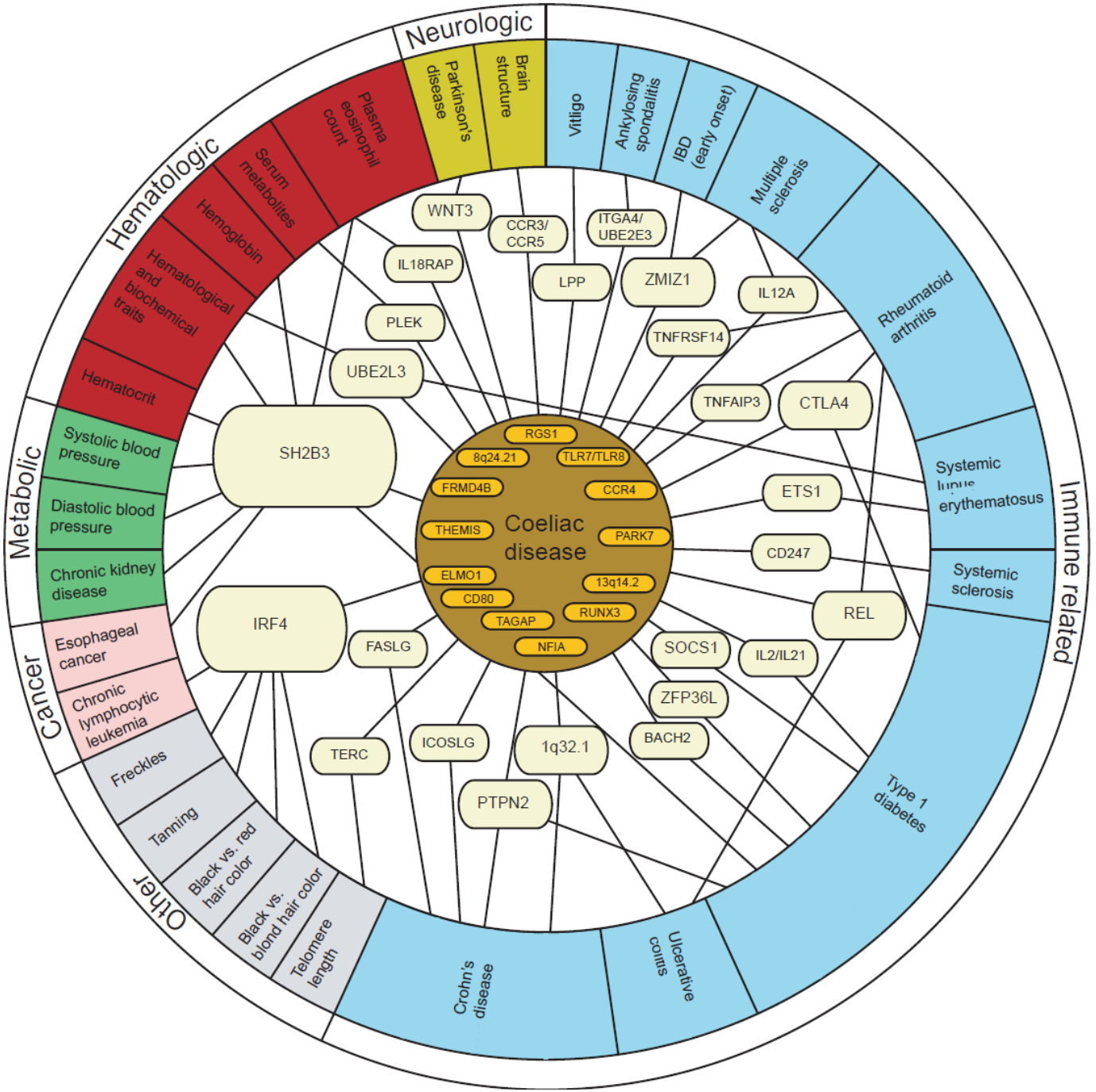
Lude Franke, UMCG, the Netherlands

http://genenetwork.nl/

Genetic sharing across immune diseases, motivation for Immunochip consortium

wellcome trust sanger institute

Associated loci are shared with other diseases



Trynka et al. Trends in Mol Med 2010

wellcome trust sanger institute

Immunochip

- Follow-up on results from 12 autoimmune GWASs
- 196,524 manufactured SNPs

Design

- Replication of signals with intermediate significance

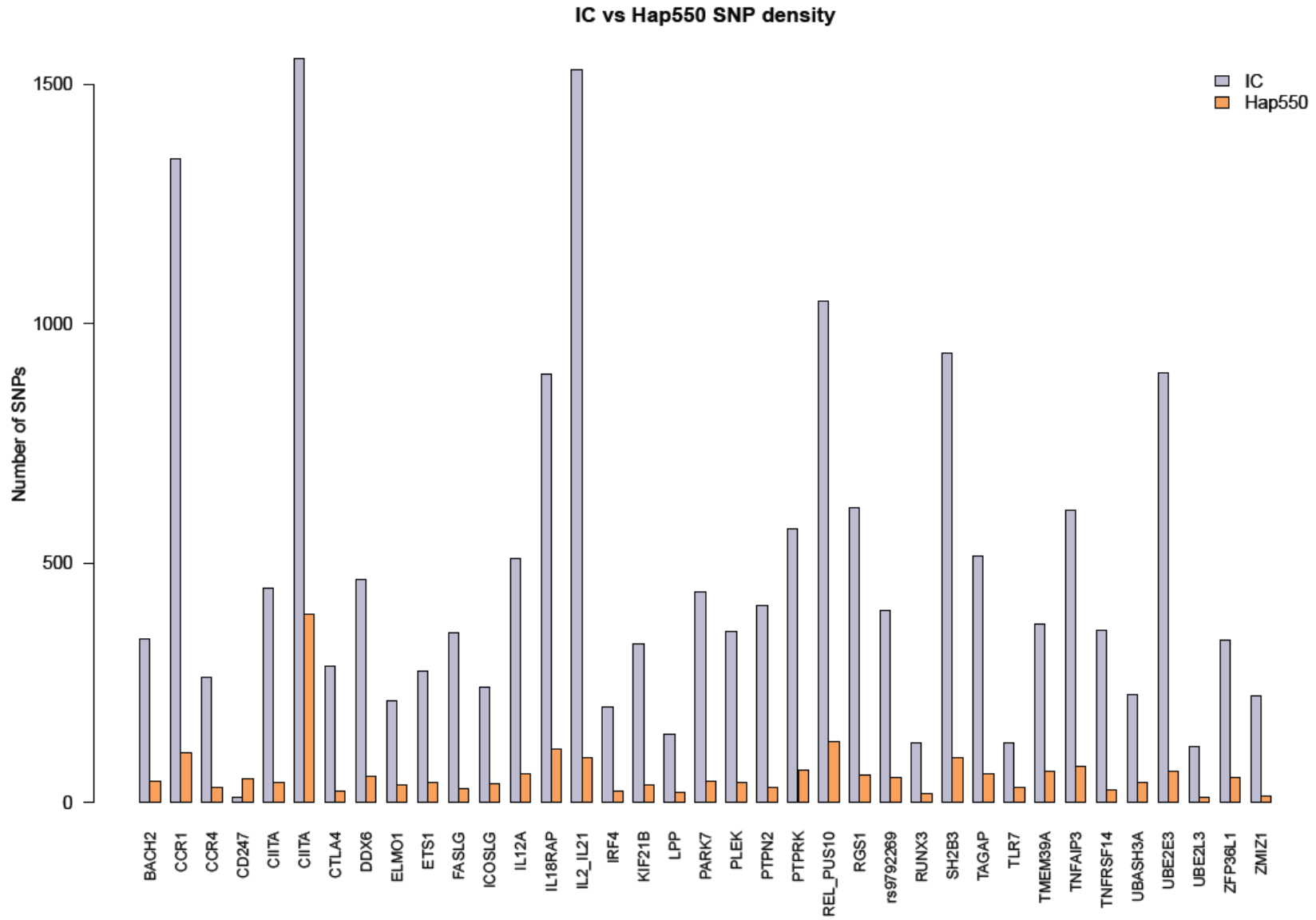
Replication

- Dense genotyping at 186 loci associated to immune-related diseases (10-20x greater marker density)
- Sequence variants (early release of 1000 Genomes Project and individual sequencing efforts)

Fine-mapping

wellcome trust sanger institute

10-20x greater SNP density than Hap550 (post-QC)



wellcome trust sanger institute

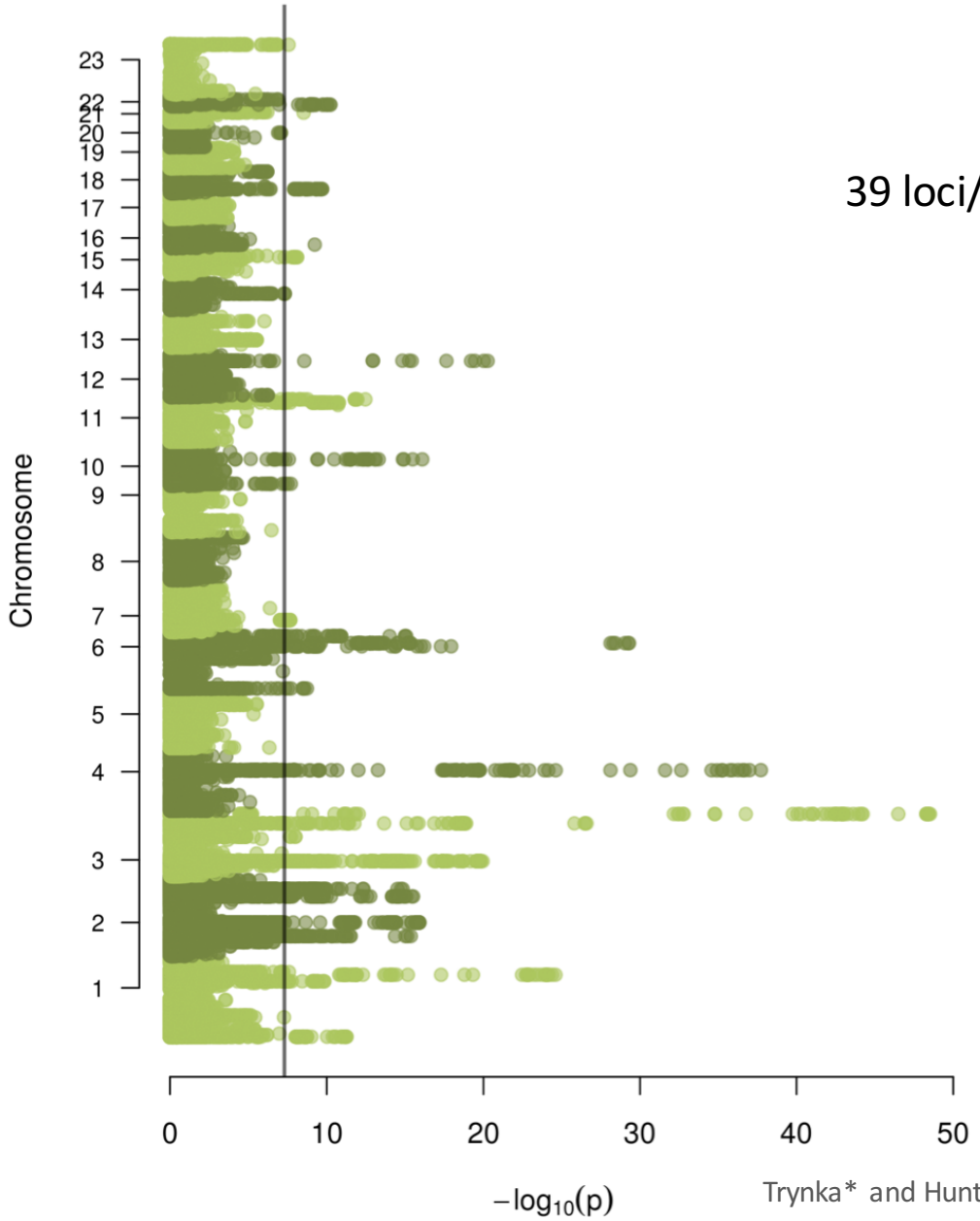
Sample collections – ultra clean dataset

Population	Celiac cases	Controls
UK	7728	8274[b]
The Netherlands	1123	1147
Poland	505	533
Spain - Basque[a]	545	308
Spain - Madrid[a]	537	320
Italy - Rome, Milan, Naples	1374	1255
India - Punjab	229	391
Total	**12041**	**12228**

- Sample call rate > 99.5%
- SNP call rate > 99.95%
- Overall 0.008% missing genotype calls
- 139,553 polymorphic SNPs (excluded15,657 NP variants)
- 24,661 variants: 5%> MAF > 0.05% and 22,941 MAF<0.05%

wellcome trust sanger institute

Identification of 39 risk loci



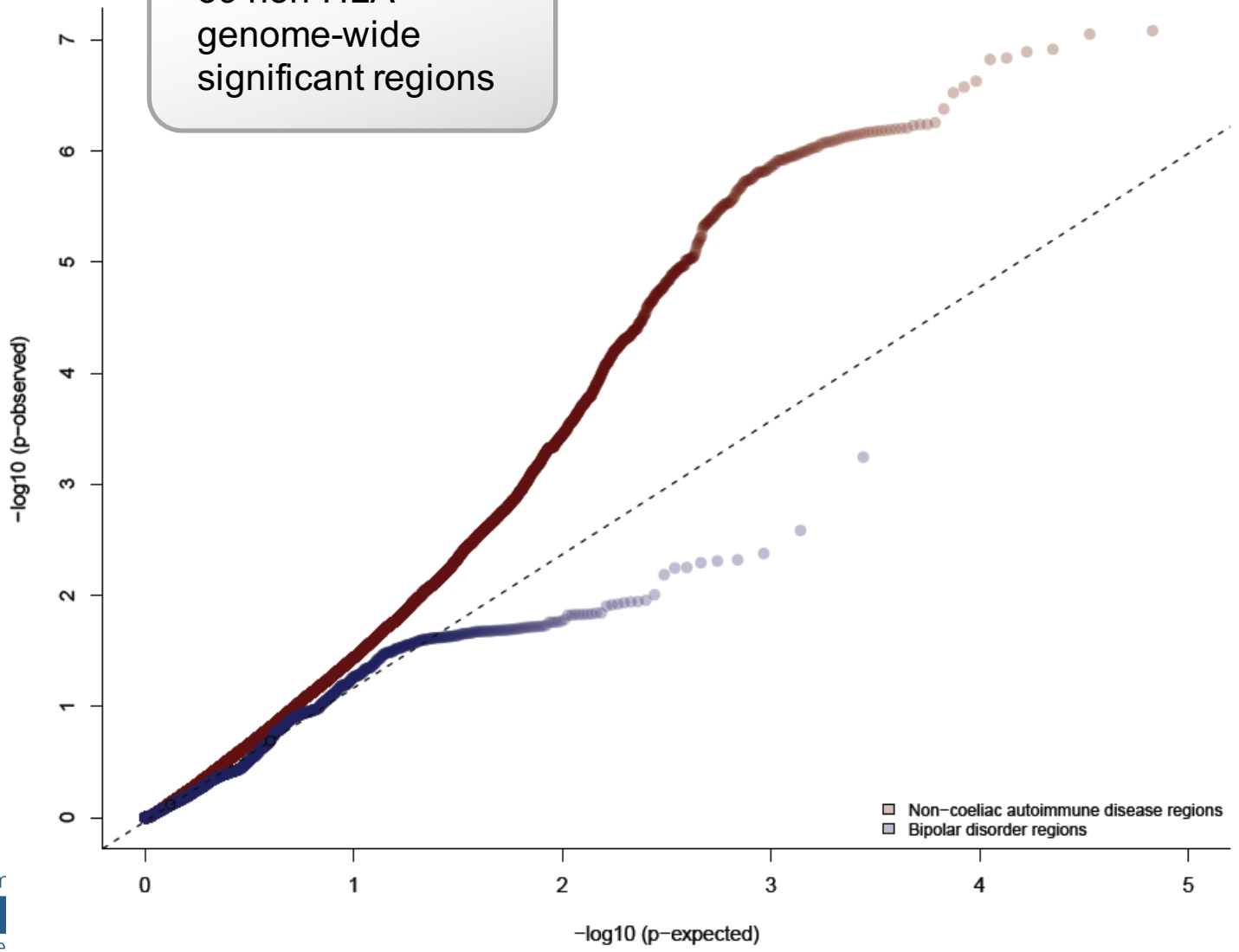
39 loci/57 SNPs

Trynka* and Hunt* et al. *Nat Genet*, 2011

wellcome trust sanger institute

Excessive enrichment for association signals in autoimmune loci

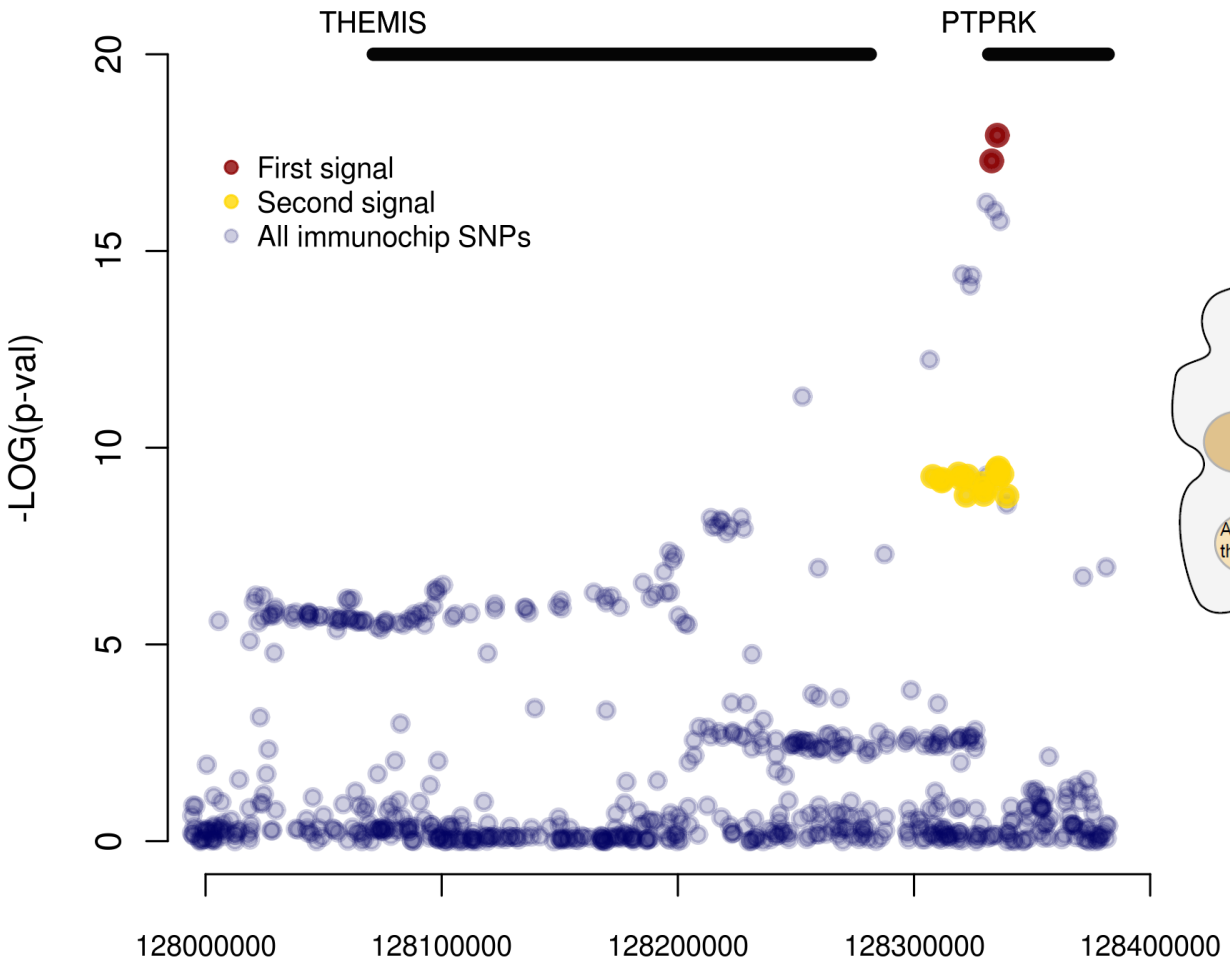
39 non-HLA genome-wide significant regions

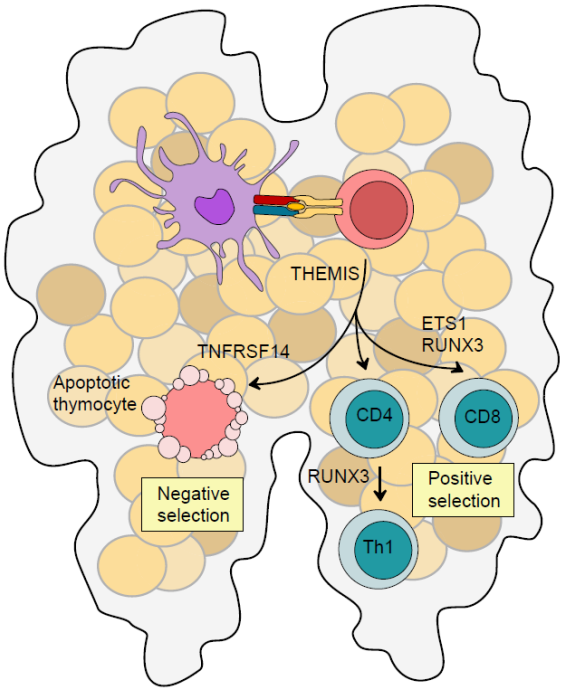


wellcome sanger institute

- Dense SNP map should allow us to carry out statistical fine-mapping to refine association signals

Refining the association signals

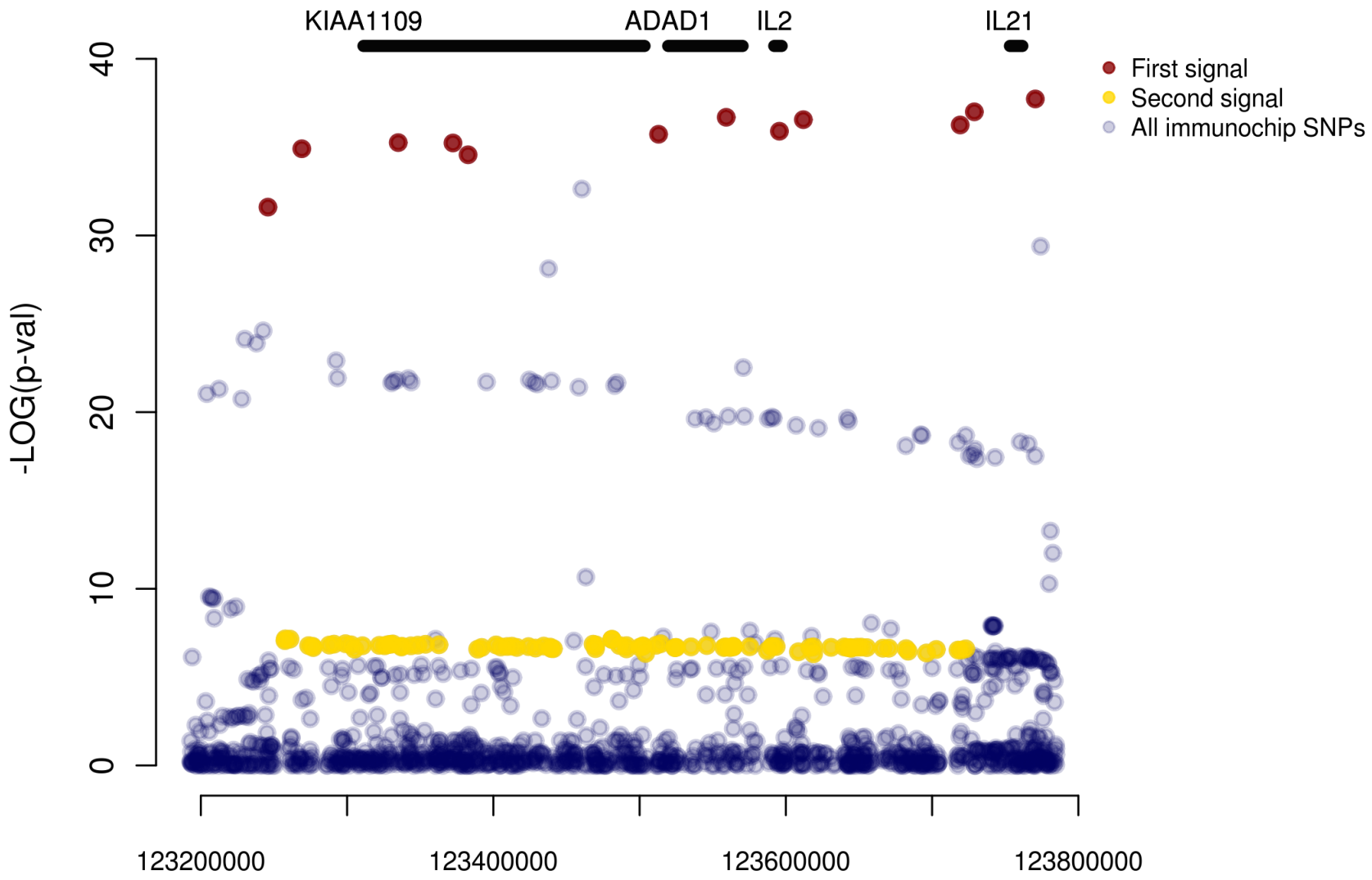




wellcome trust sanger institute

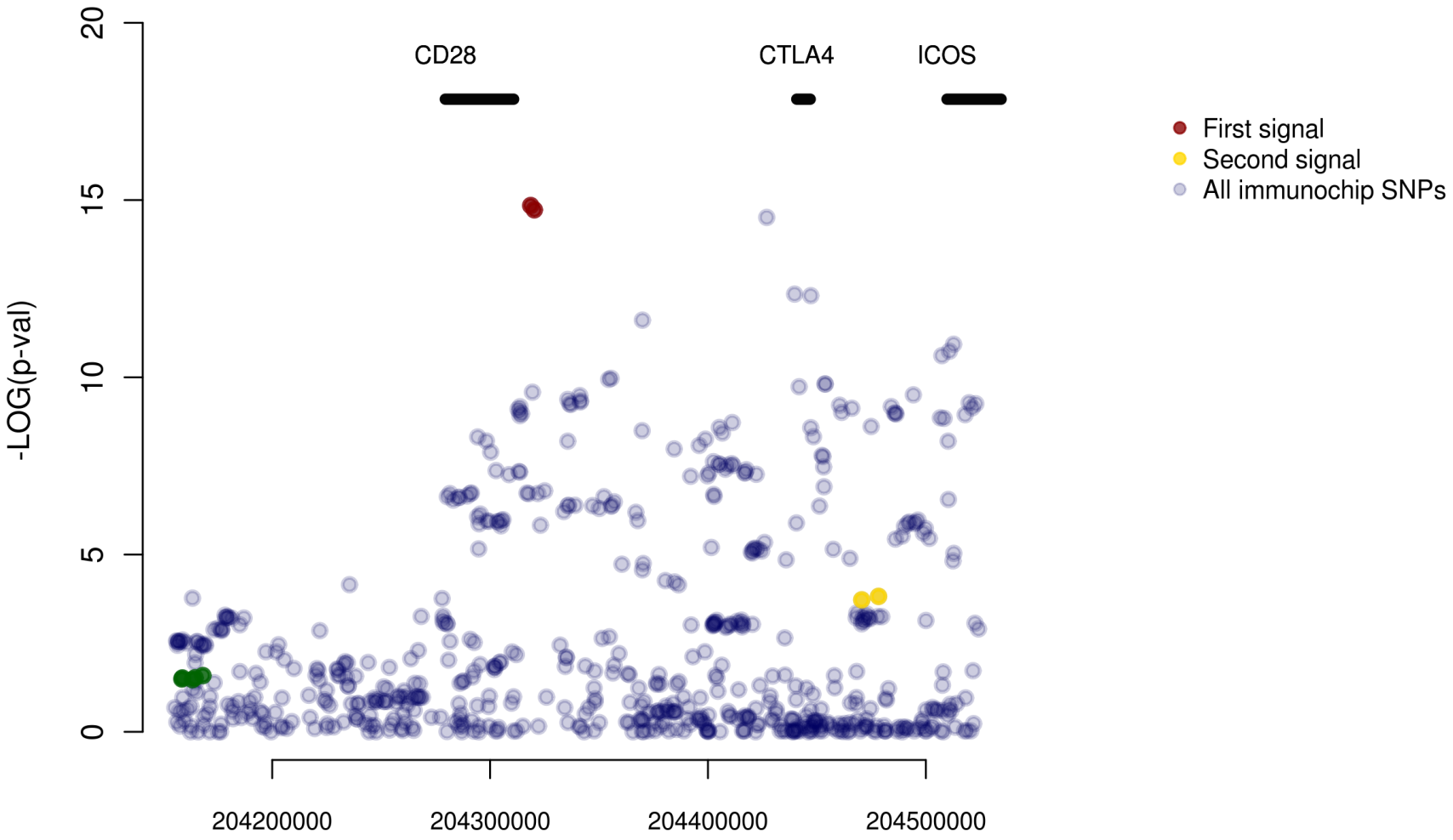
Trynka and Hunt et al., Nat Genet 2011

Unsuccessful fine-mapping

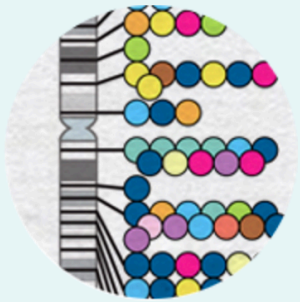


Trynka and Hunt et al., Nat Genet 2011

Multiple independent signals



Trynka and Hunt *et al., Nat Genet 2011*



GWAS Catalog

The NHGRI-EBI Catalog of published genome-wide association studies

Search the catalog

Examples: breast cancer, rs7329174, Yang, 2q37.1, HBS1L

https://www.ebi.ac.uk/fgpt/gwas/

genome.gov
National Human Genome Research Institute
National Institutes of Health

Google™ Search SEARCH

Research Funding | Research at NHGRI | Health | Education | Issues in Genetics | Newsroom | Careers & Training | About | For You

Home › Research Funding › Research Funding Divisions › Division of Genomic Medicine › GWAS Catalog

Share Print

Division of Genomic Medicine

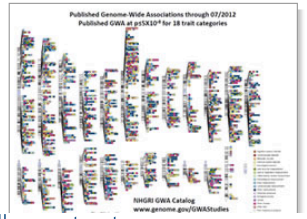
A Catalog of Published Genome-Wide Association Studies

Division Staff ¦ Funding Opportunities ¦ Genomic Medicine Activities ¦ GWAS Catalog ¦ Meetings & Workshops ¦ Potential Sample Collections for Sequencing ¦ Programs ¦ Publications ¦ Trans-NIH Sequencing Inventory

Additional information has been added to the HTML catalog columns below. For a description of column headings for the HTML catalog, go to: Catalog Heading Descriptions new

Potential etiologic and functional implications of genome-wide association loci for human diseases and traits
Click here to read our recent *Proceedings of the Academy of Sciences (PNAS)* article on catalog methods and analysis.

View the Full Catalog new Download the Catalog new Search the Catalog new



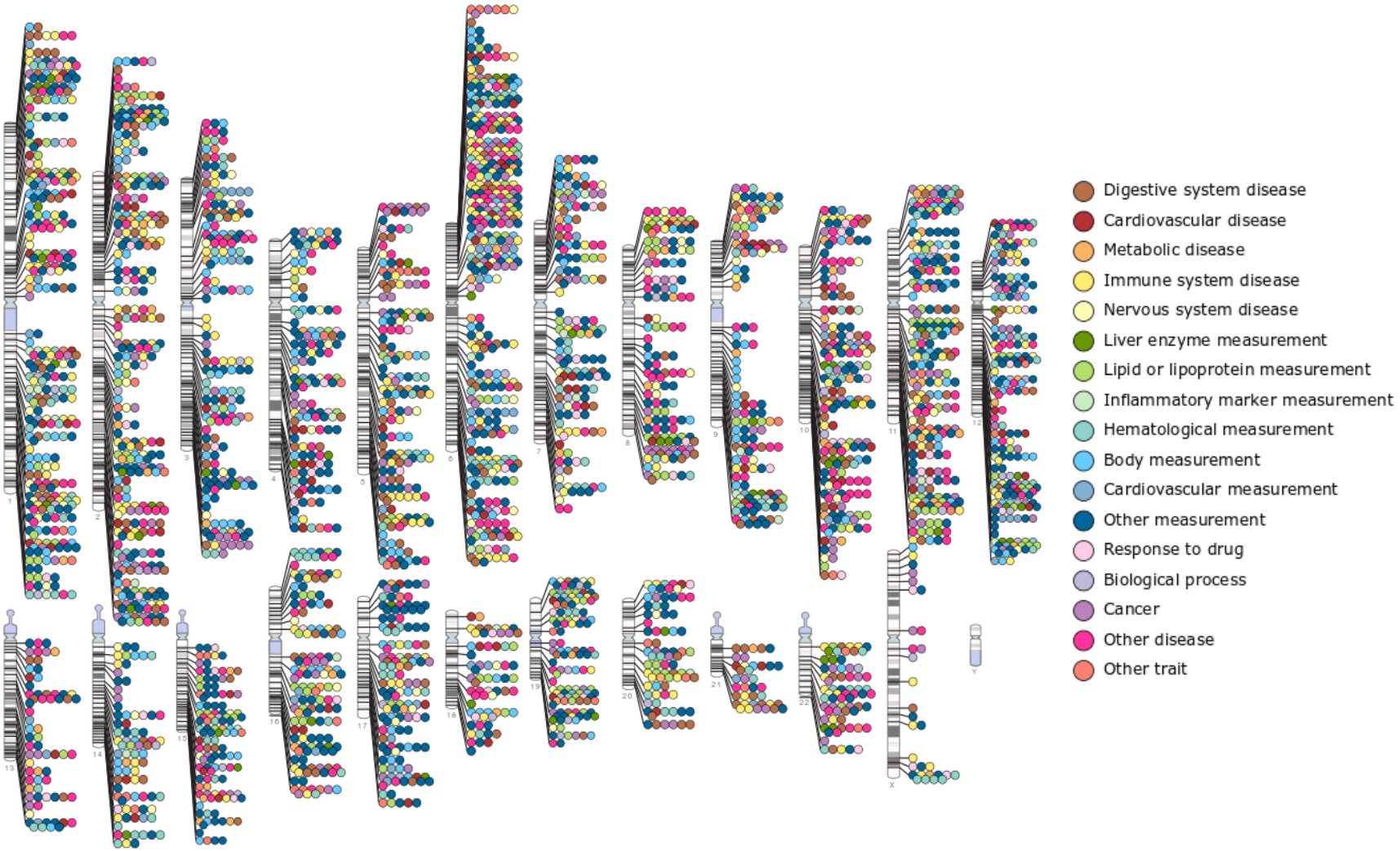
The genome-wide association study (GWAS) publications listed here include only those attempting to assay at least 100,000 single nucleotide polymorphisms (SNPs) in the initial stage. Publications are organized from most to least recent date of publication, indexing from online publication if available. Studies focusing only on candidate genes are excluded from this catalog. Studies are identified through weekly PubMed literature searches, daily NIH-distributed compilations of news and media reports, and occasional comparisons with an existing database of GWAS literature (HuGE Navigator).

SNP-trait associations listed here are limited to those with p-values < 1.0 x 10-5 (see full methods for additional details). Multipliers of powers of 10 in p-values are rounded to the nearest single digit; odds ratios and allele frequencies are rounded to two decimals. Standard errors are converted to 95 percent confidence intervals where applicable. Allele frequencies, p-values, and odds ratios derived from the largest sample size, typically a combined analysis (initial plus replication studies), are recorded below if reported; otherwise statistics from the initial study sample are recorded. For quantitative traits, information on % variance explained, SD increment, or unit difference is reported where available. Odds ratios < 1 in the original paper are converted to OR > 1 for the alternate allele. Where results from multiple genetic models are available, we prioritized effect sizes (OR's or beta-coefficients) as follows: 1) genotypic model, per-allele estimate; 2) genotypic model, heterozygote estimate, 3) allelic model, allelic estimate.

wellcome trust sanger institute

https://www.genome.gov/gwastudies/

We have mapped thousands of disease loci!



wellcome trust sanger institute

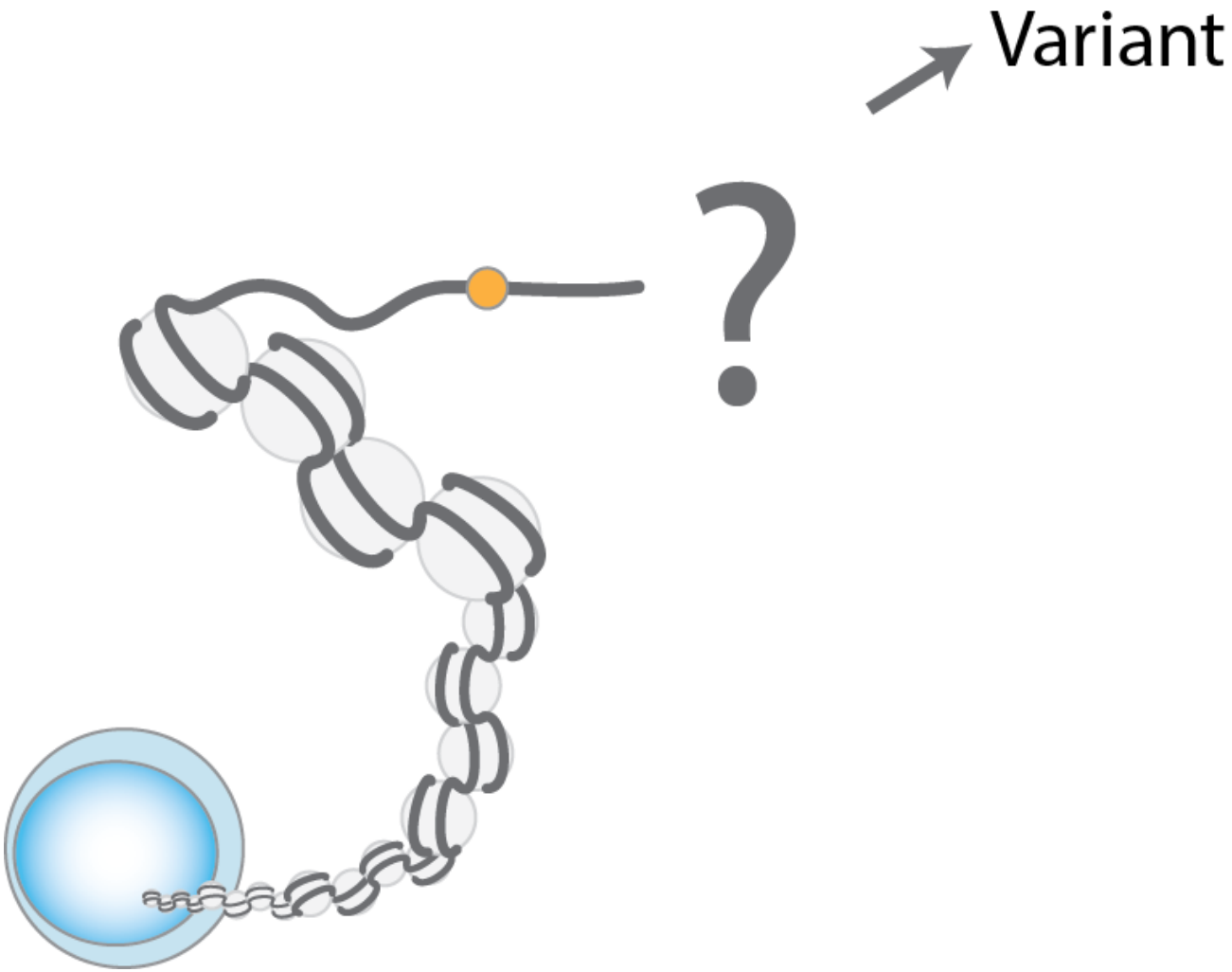
https://www.ebi.ac.uk/fgpt/gwas/

And then we got stuck…

The path from disease association to function remains a challenge

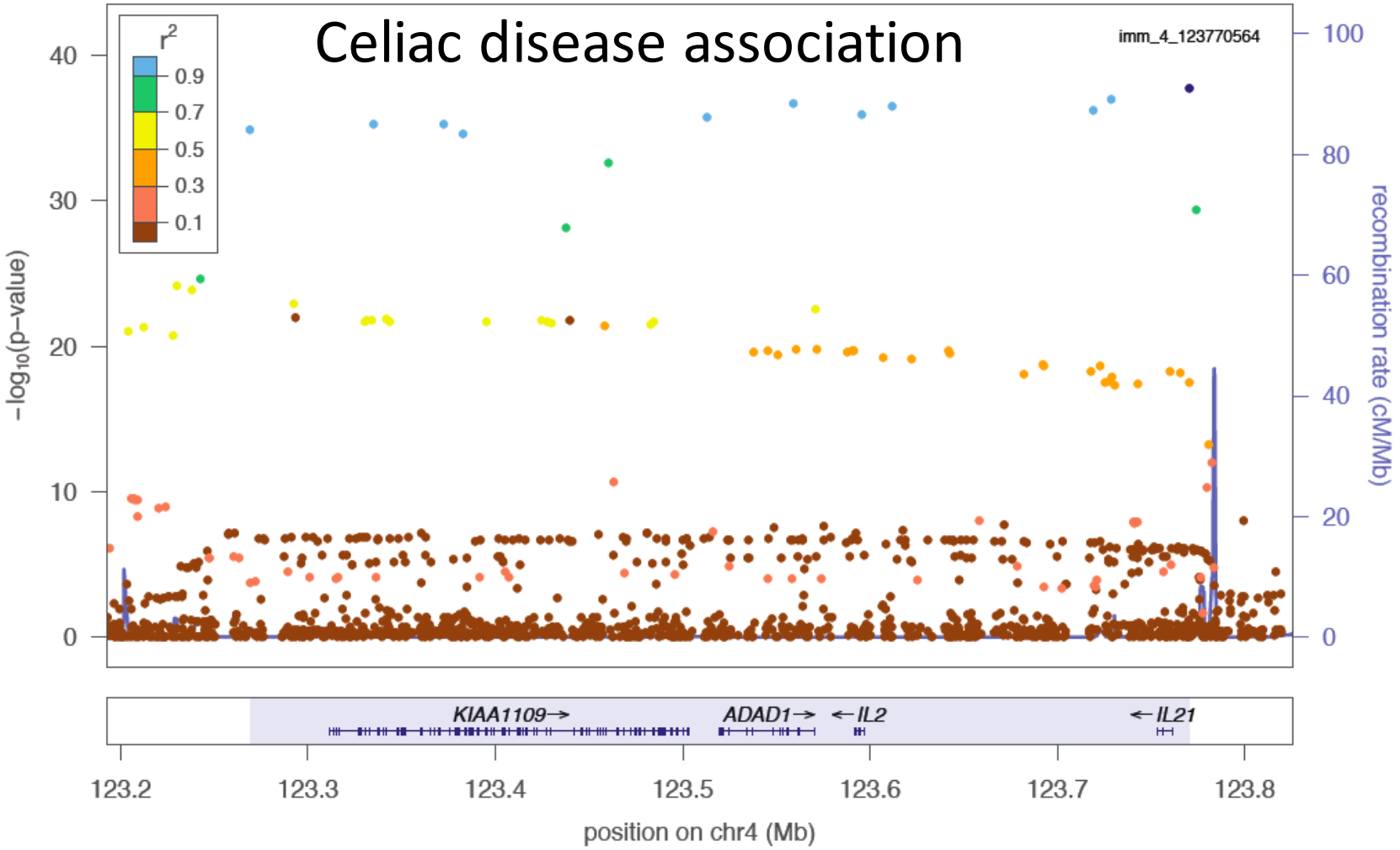
wellcome trust sanger institute

Interpreting disease variants



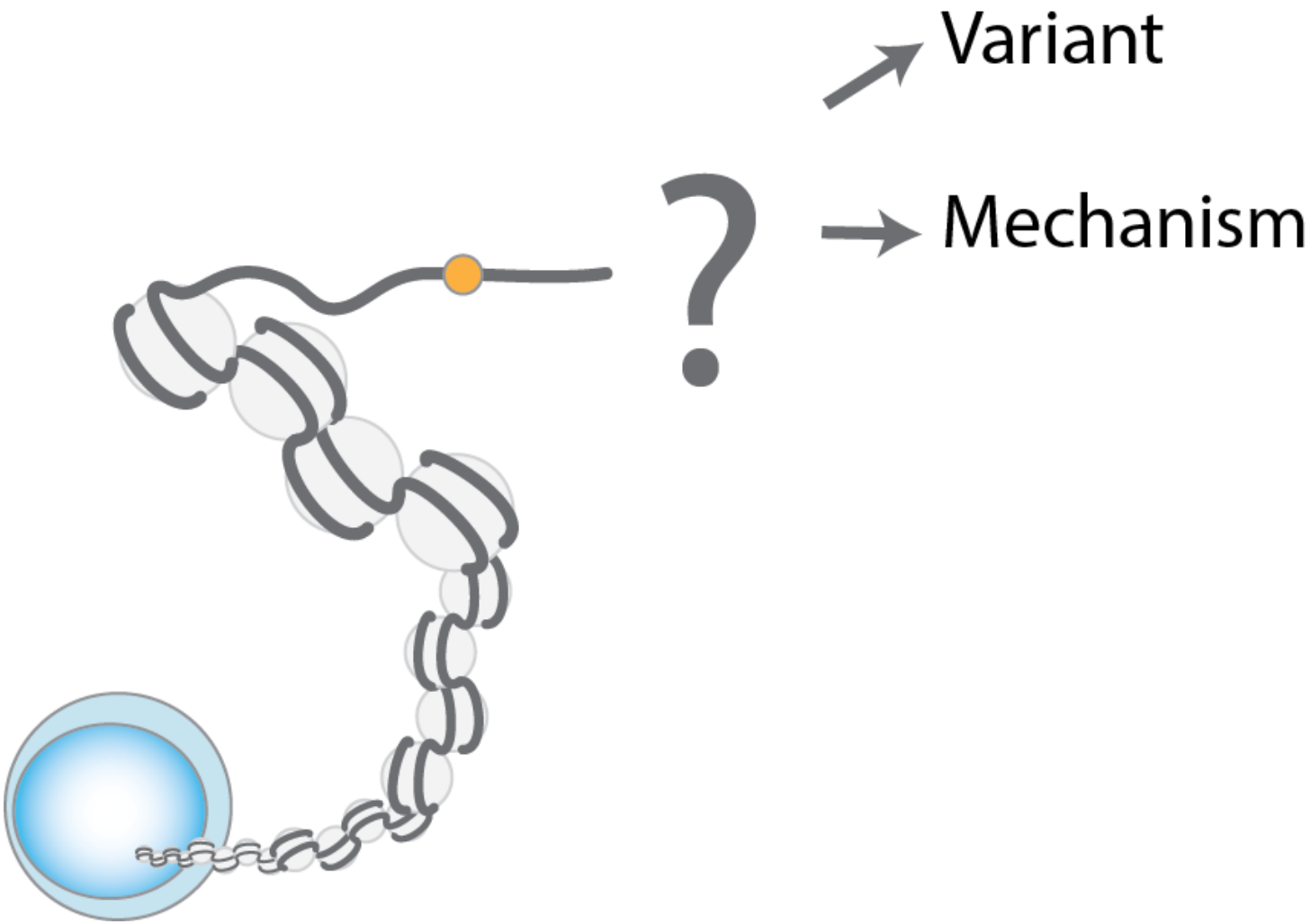
wellcome trust sanger institute

The challenge to identify relevant variants – extended LD



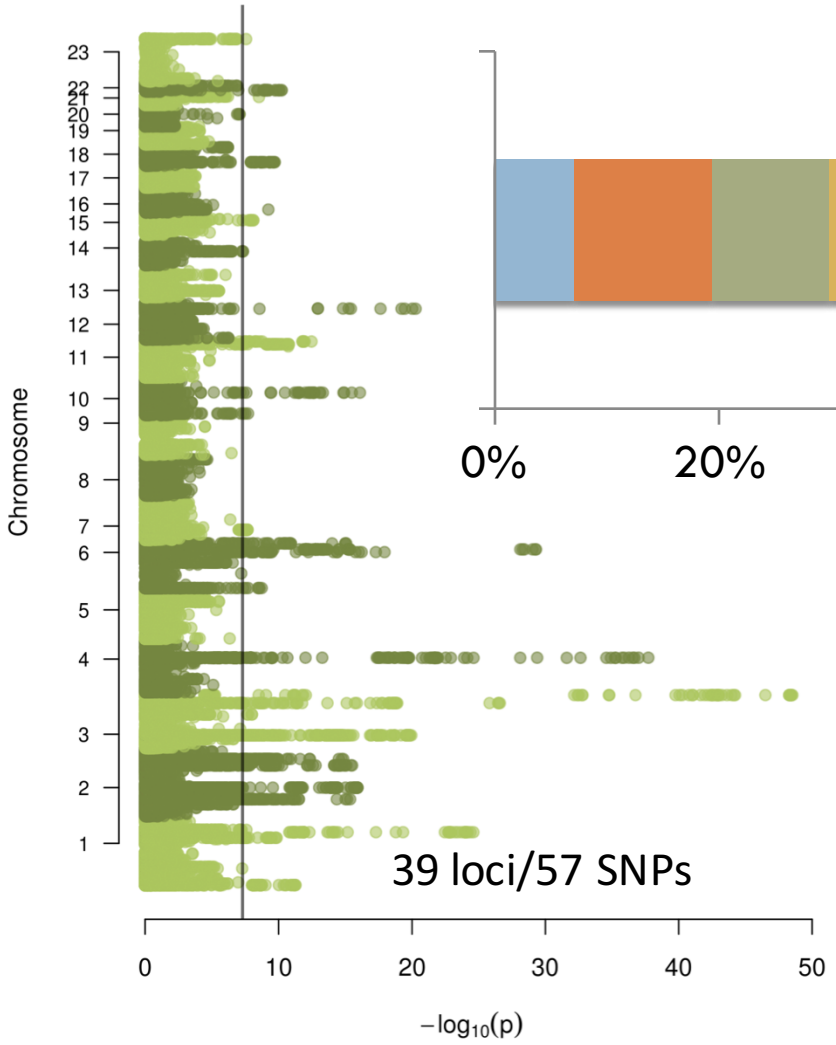
Trynka* and Hunt* et al. *Nat Genet*, 2011

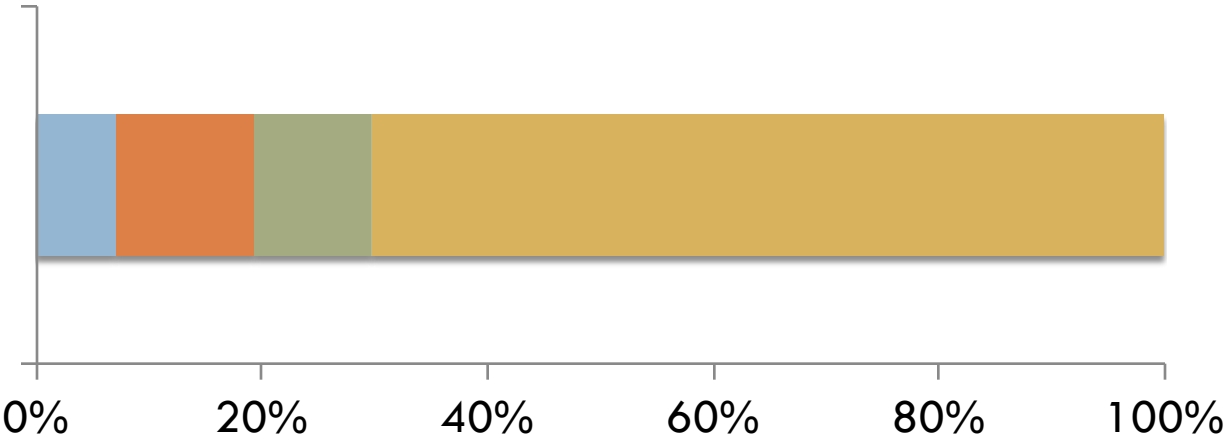
Interpreting disease variants



wellcome trust sanger institute

Celiac disease alleles implicate gene regulatory function



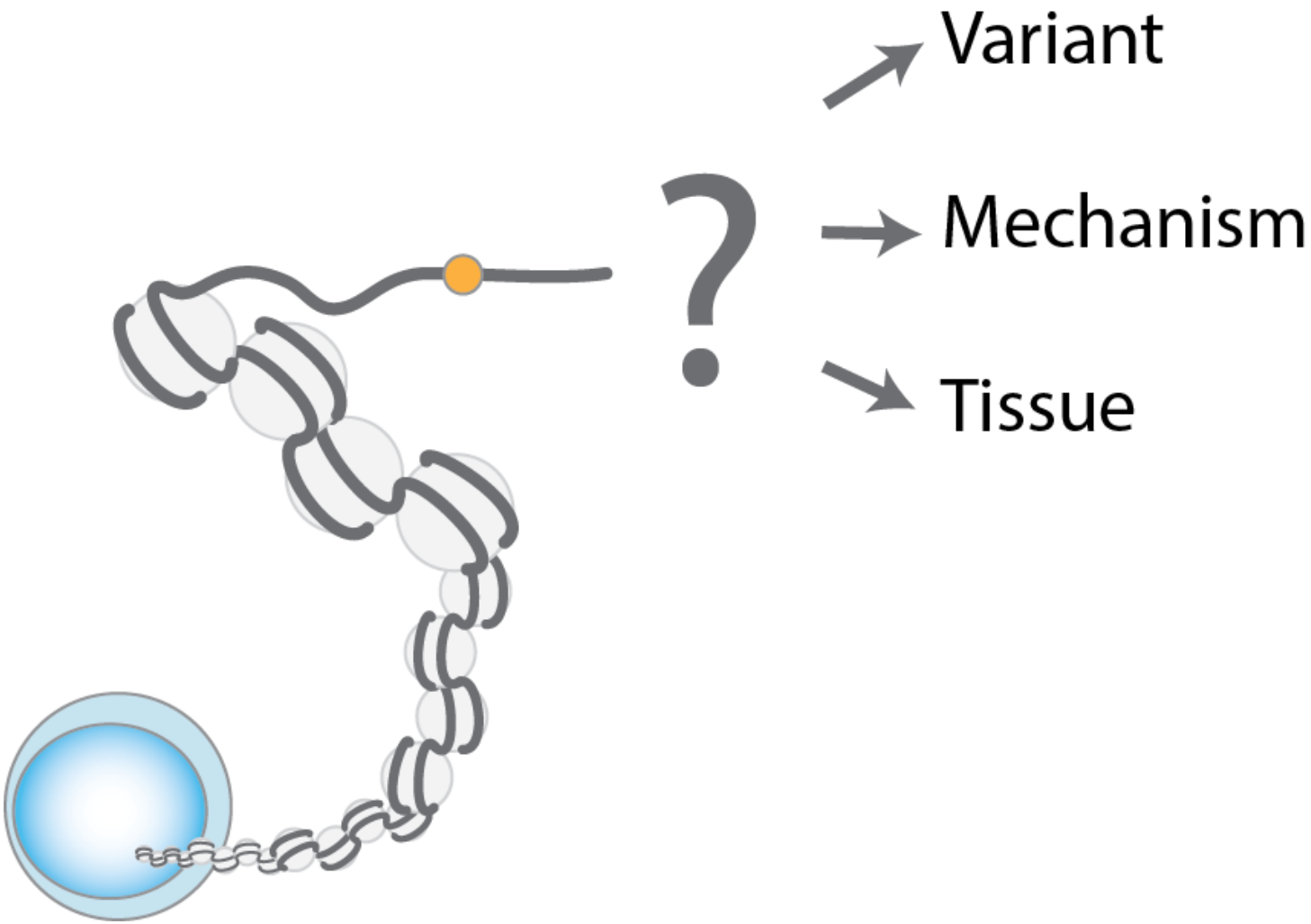


- Coding
- Transcribed Non-Coding
- Regulatory
- Other

wellcome trust sanger institute

Trynka* and Hunt* et al. *Nat Genet*, 2011

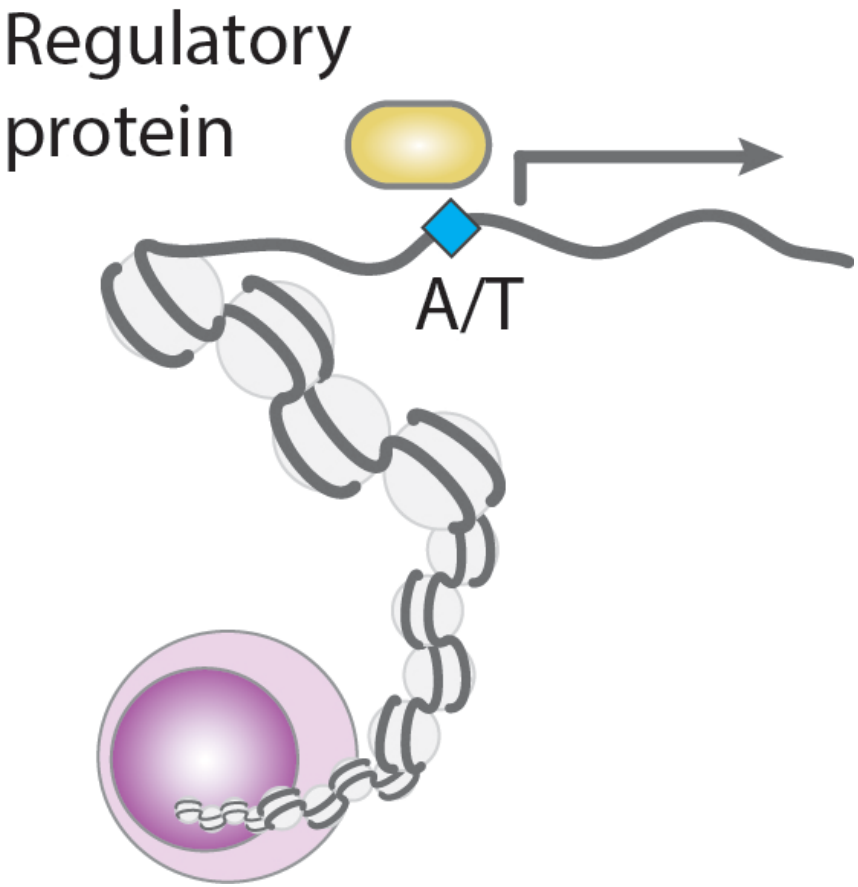
Interpreting disease variants



wellcome trust sanger institute

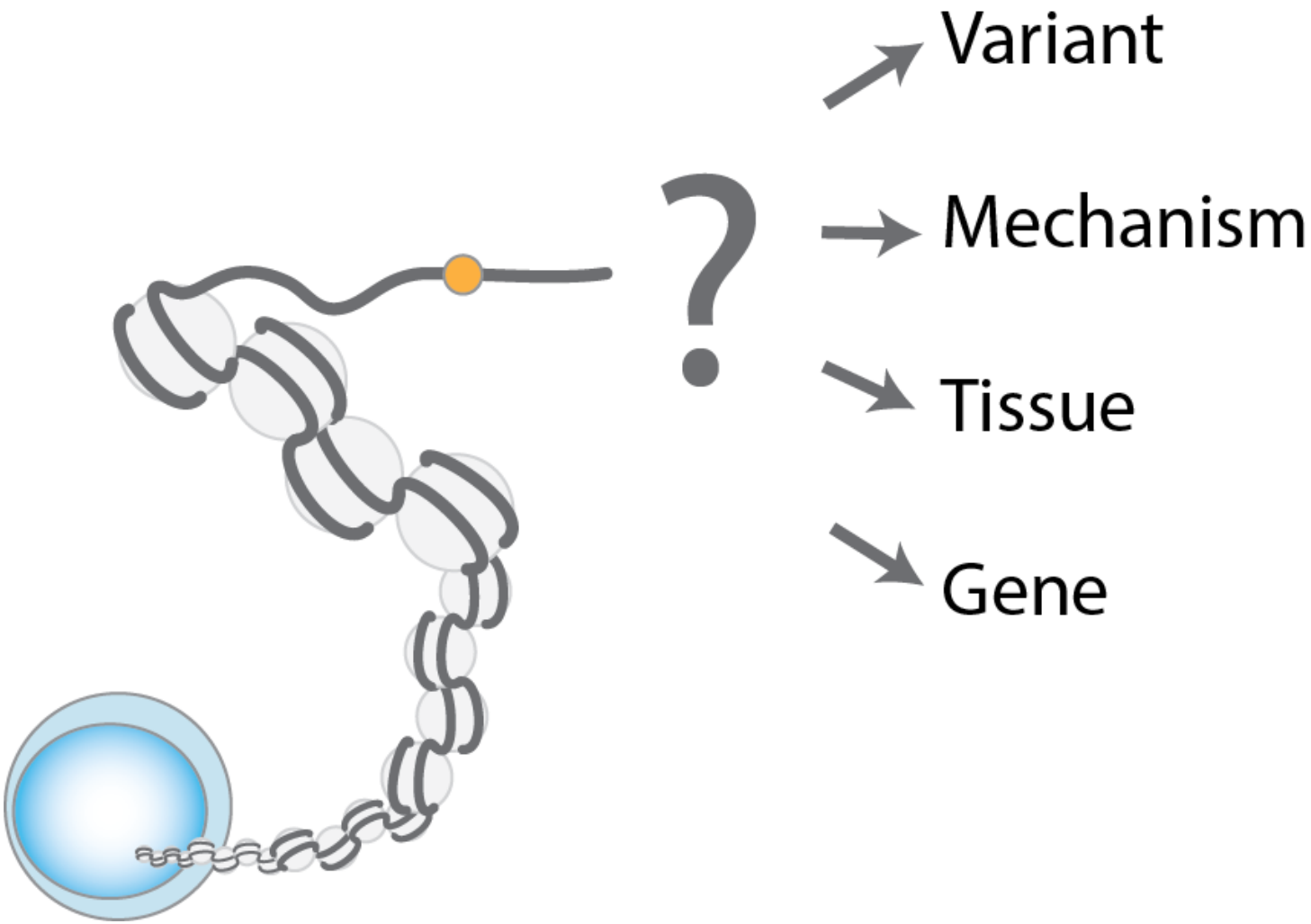
A handful of studies describing function of non-coding disease variants

- E.g. LDL (Musunuru et al., *Nature* 2010)
- LDL associated non-coding variant creates transcription factor binding site and alters the hepatic expression of the *SORT1*



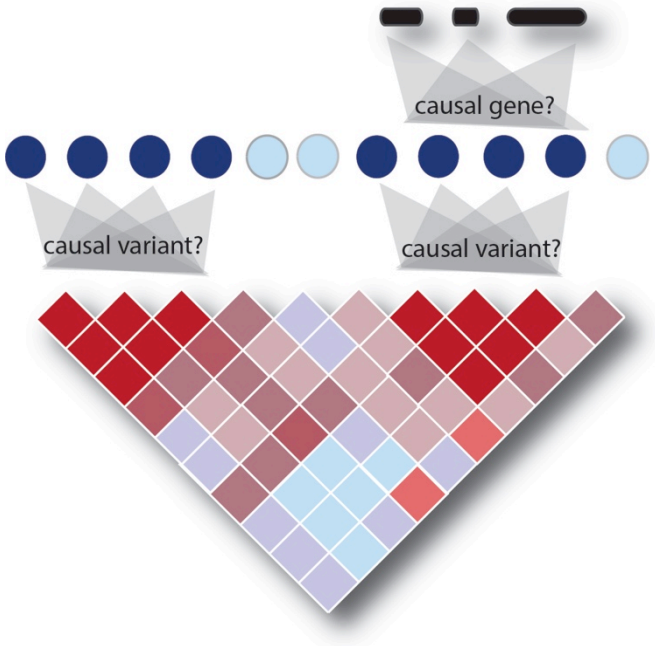
wellcome trust sanger institute

Interpreting disease variants



wellcome trust sanger institute

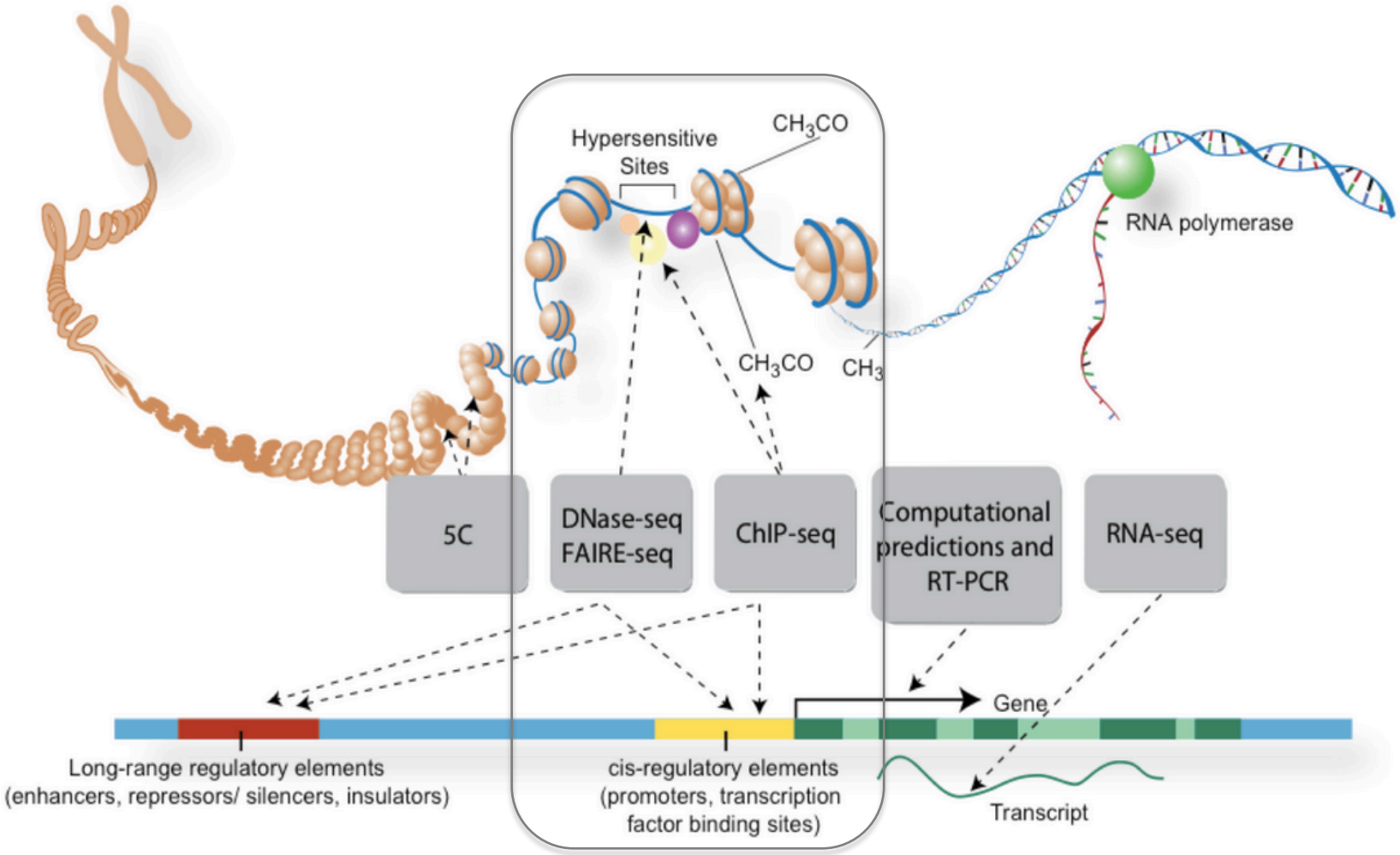
The challenge to identify relevant variants and genes



- Multiple genes within associated LD block

wellcome trust sanger institute

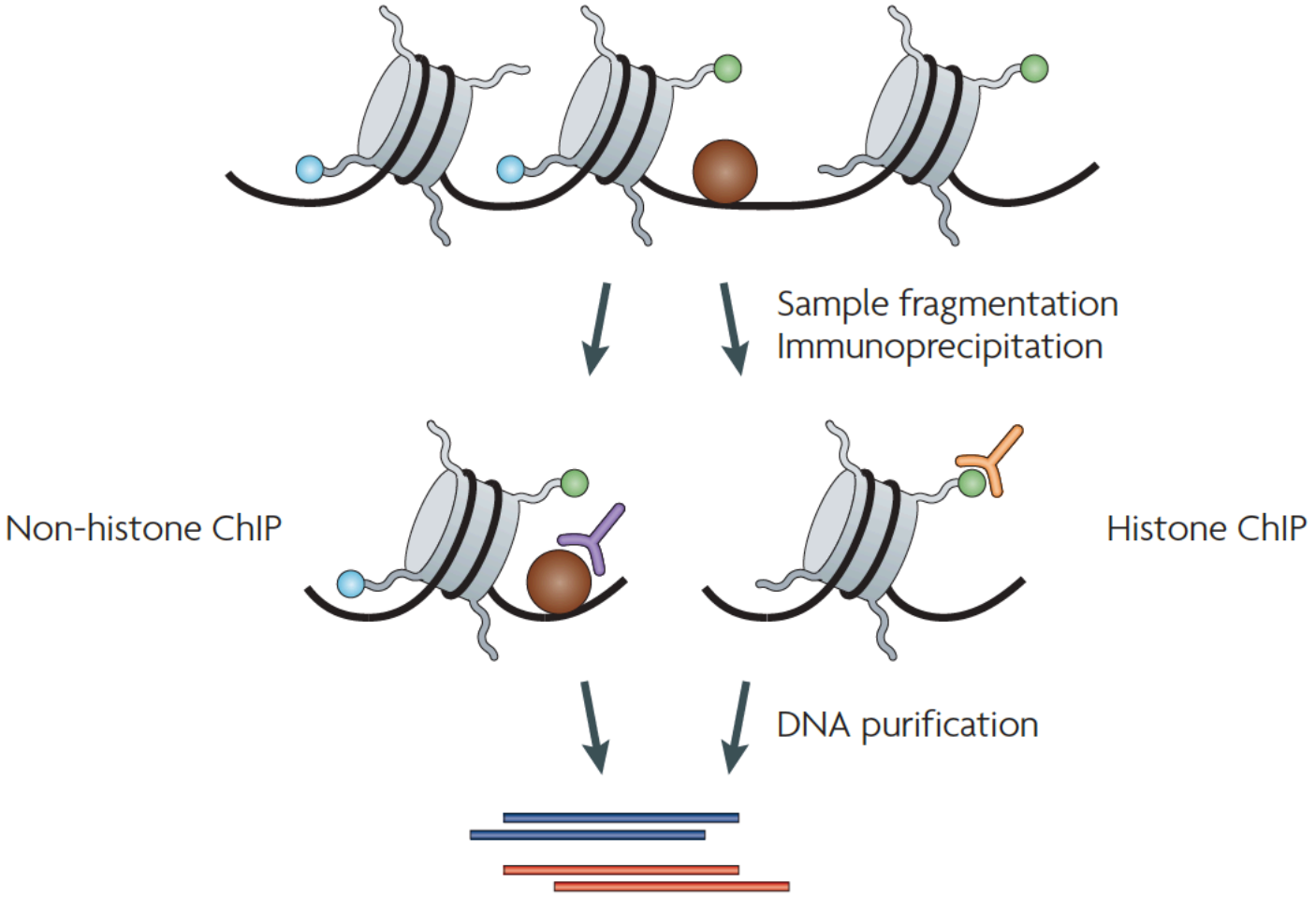
Mapping gene regulation



wellcome trust sanger institute

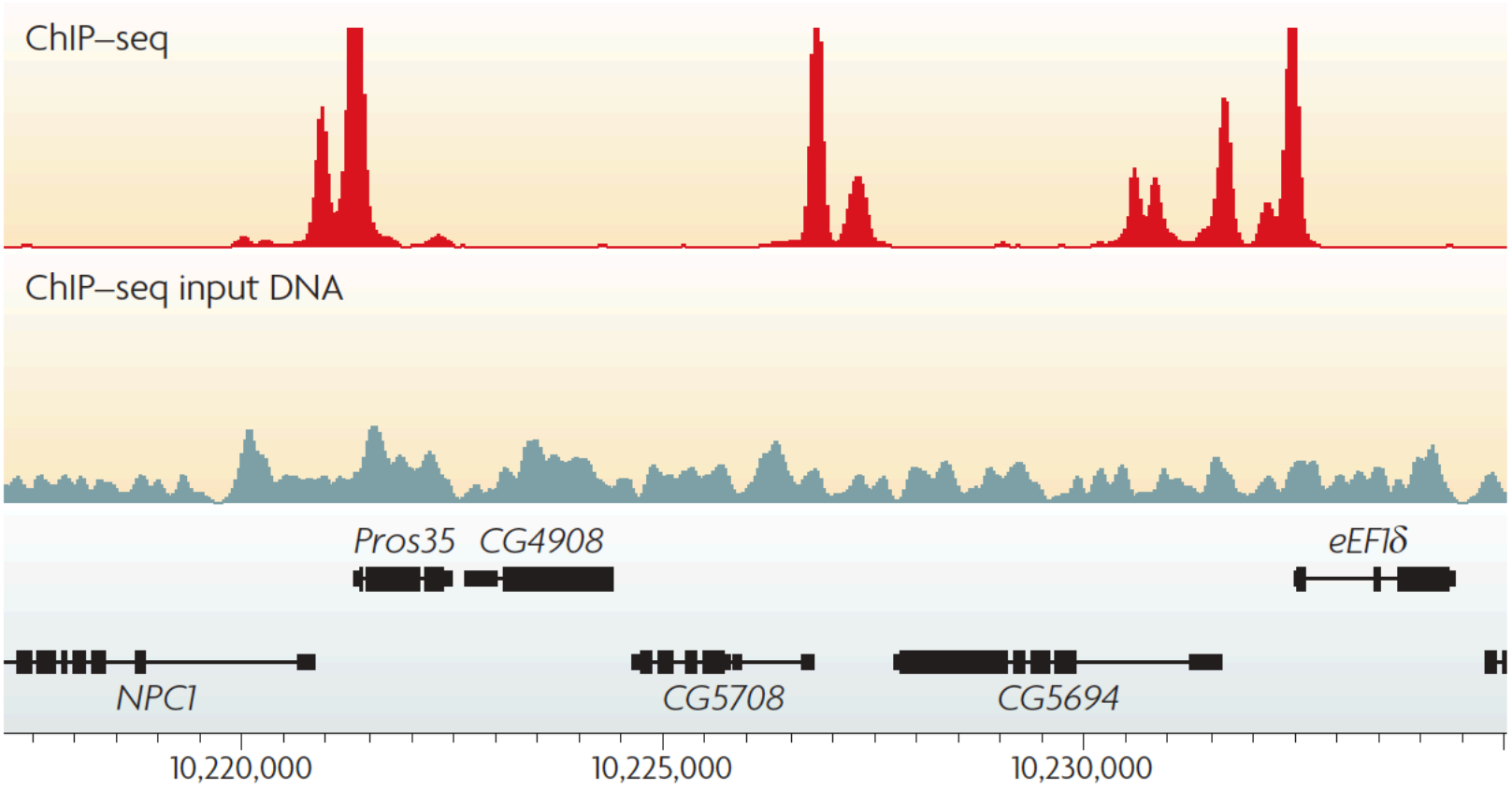
http://genome.ucsc.edu/ENCODE/

ChIP-seq



Park P, *Nat Rev Genet 2009*

ChIP-seq



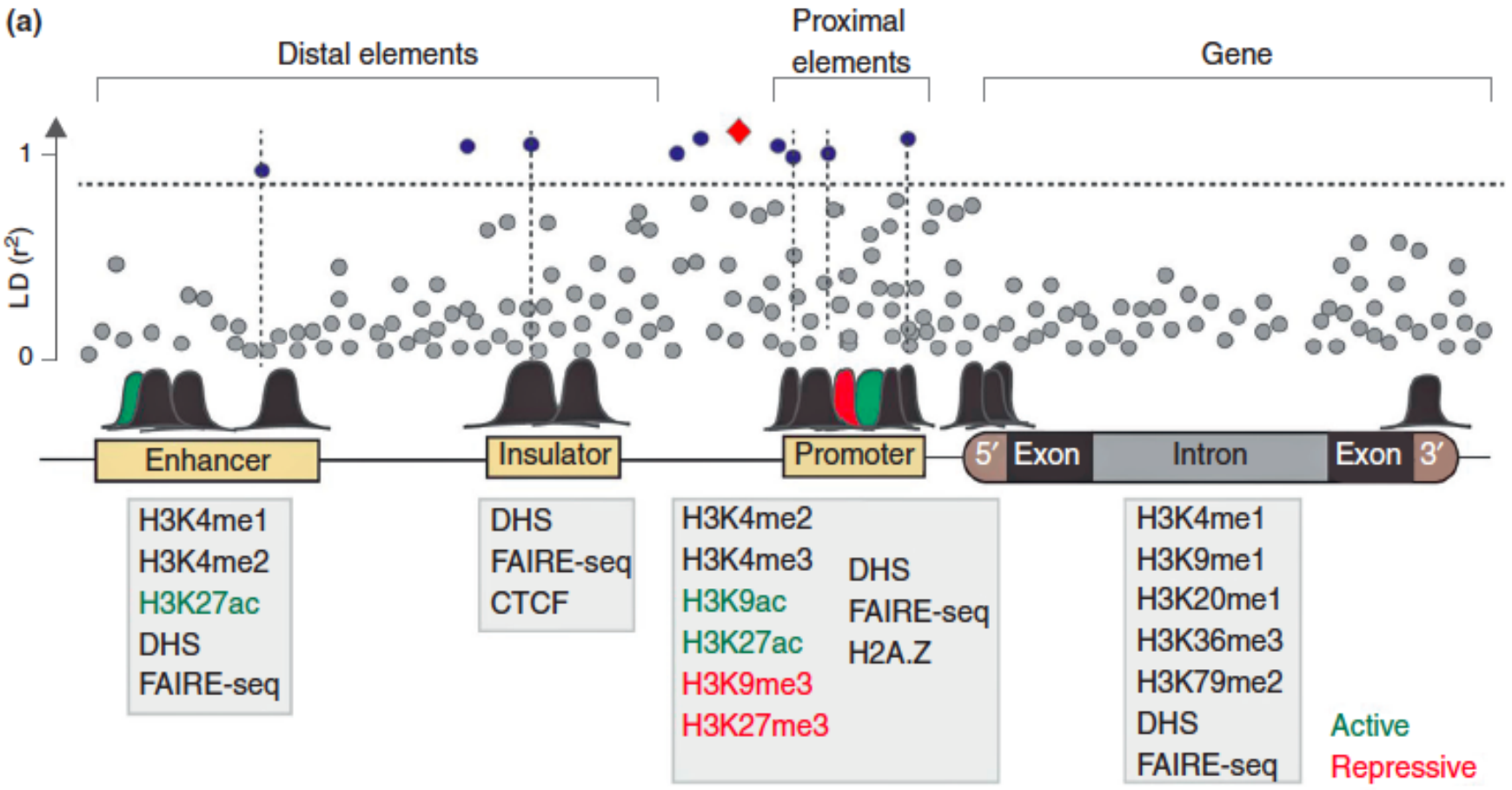
wellcome trust sanger institute

Park P, *Nat Rev Genet 2009*

Resources

- ENCODE: cell types and cell lines, broad range of assays
- NIH Roadmap Epigenome Project: human tissue atlas, limited number of assays
- BLUEPRINT epigenome: >50 blood cells from healthy and patient samples, variome

Can we interestect associated variants with epigenetic data to learn about their function?



Trynka and Raychaudhuri, *Curr Opin Genet Dev* 2013

wellcome trust sanger institute

Challenges to functionally follow-up GWAS variants

- For many diseases the causal cell type is unknown
- Disease variants can express functionality in a cell type and a cell state specific context

Outline

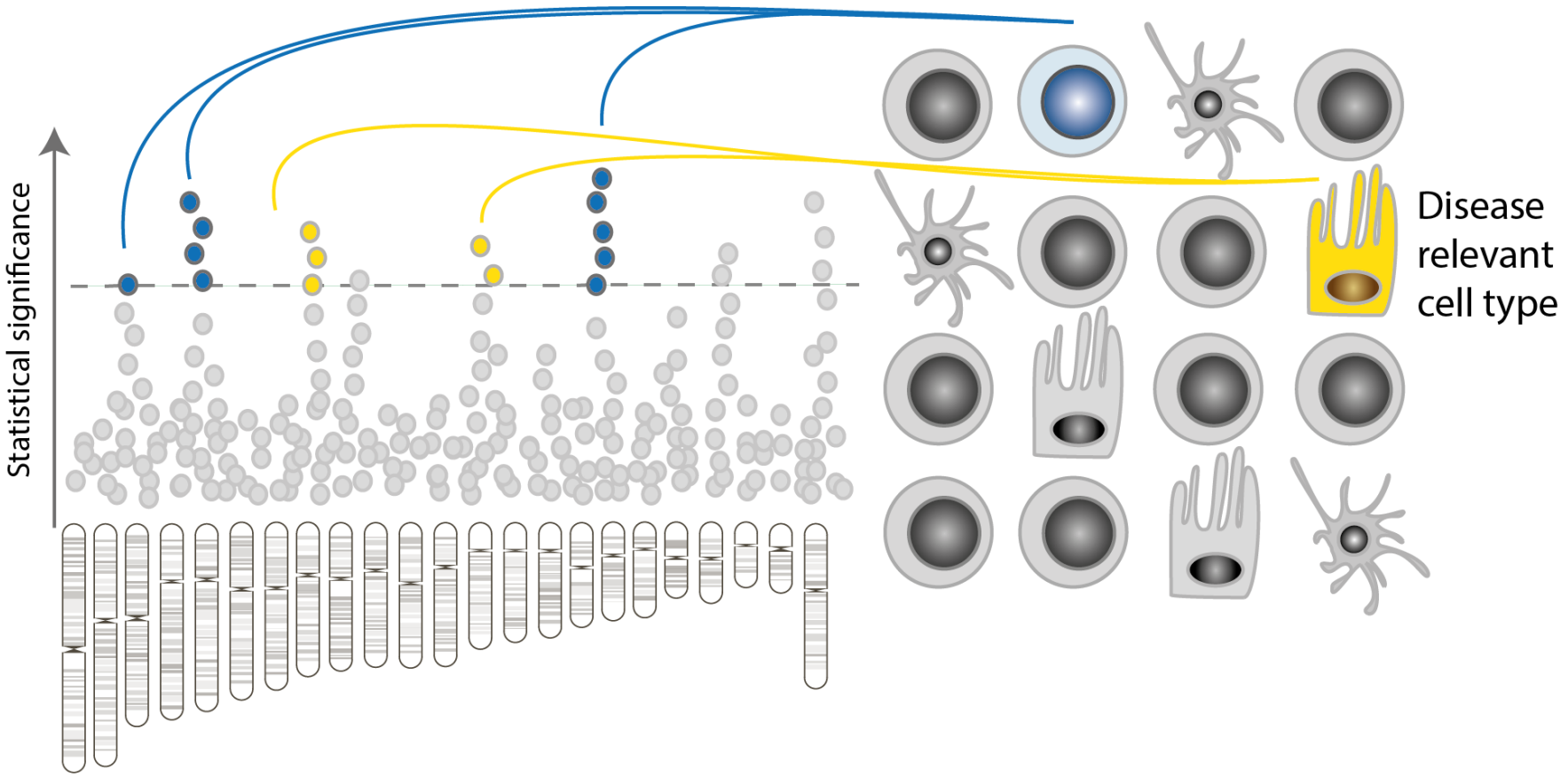
- Identifying informative chromatin marks to perform quantitative genomic assays at scale
- Trynka et al., Nature Genet 2013

- Developing a robust test to infer enrichment of associated variants across a range of genomic annotations
- Trynka et al., AJHG 2015

- Where pathways and histone marks meet: increasing specificity to identify disease relevant cell types

wellcome trust sanger institute

Leveraging cell type specific gene regulation with GWAS SNPs can pinpoint relevant cell types



wellcome trust sanger institute

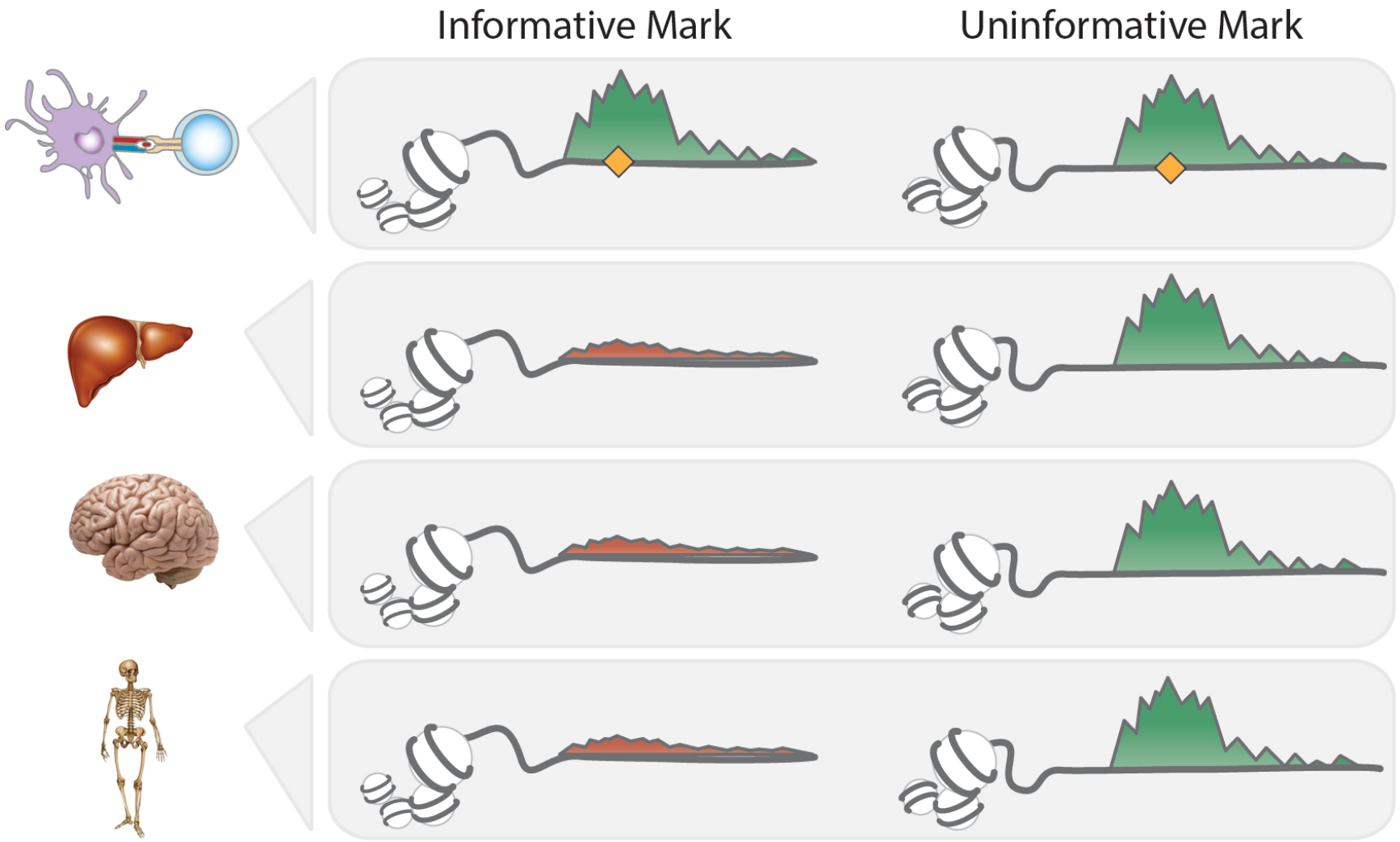
Motivation

1) Is there a **single** informative chromatin mark?

2) Can it be used to:

- identify cell-types relevant to the phenotype
- fine-map to suggest causative variant

Cell-type specific signature of disease associated variants

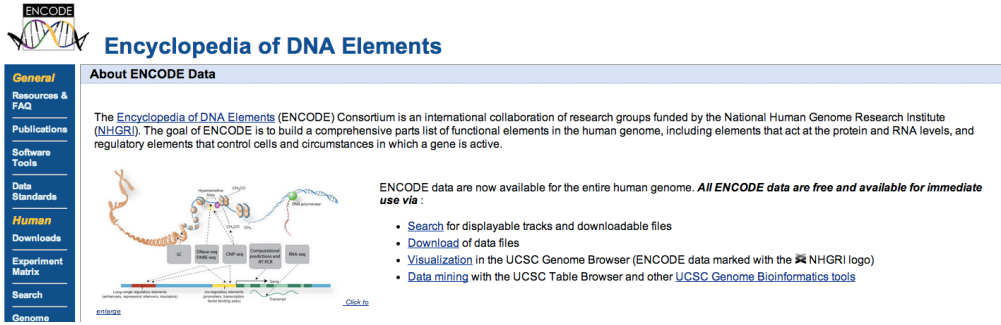


wellcome trust sanger institute

Datasets – publically available

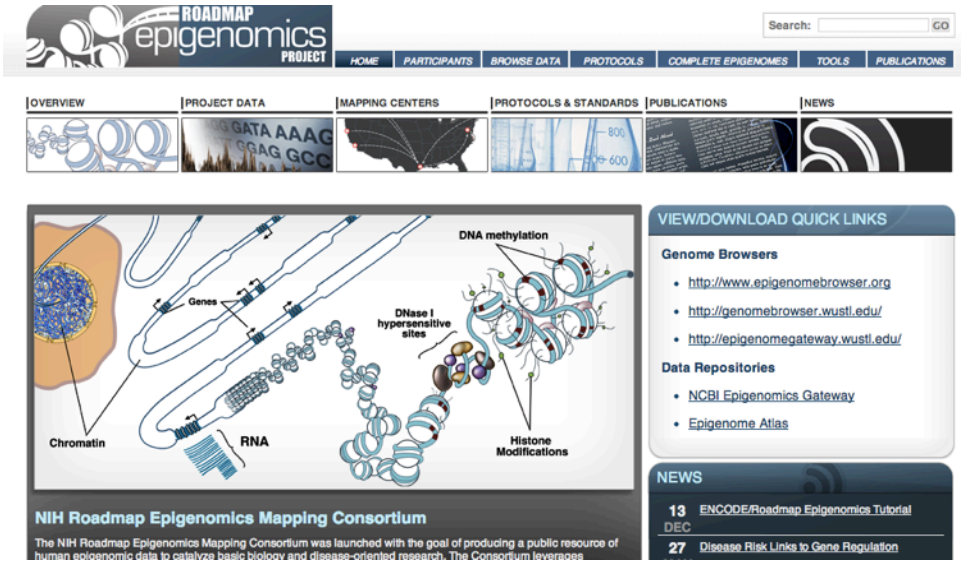
ENCODE:

- n = 14 cell lines
- 15 histone marks



NIH Epigenome Project:

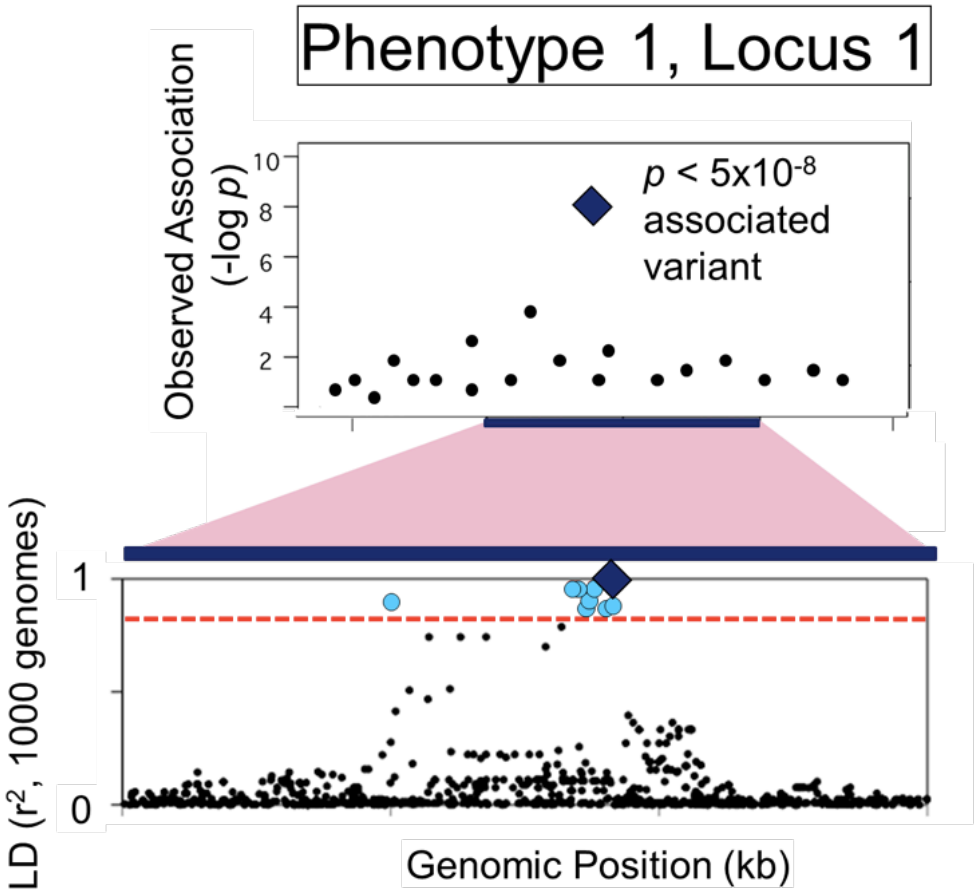
- n = 34 primary cells
- 6 histone marks



wellcome trust sanger institute

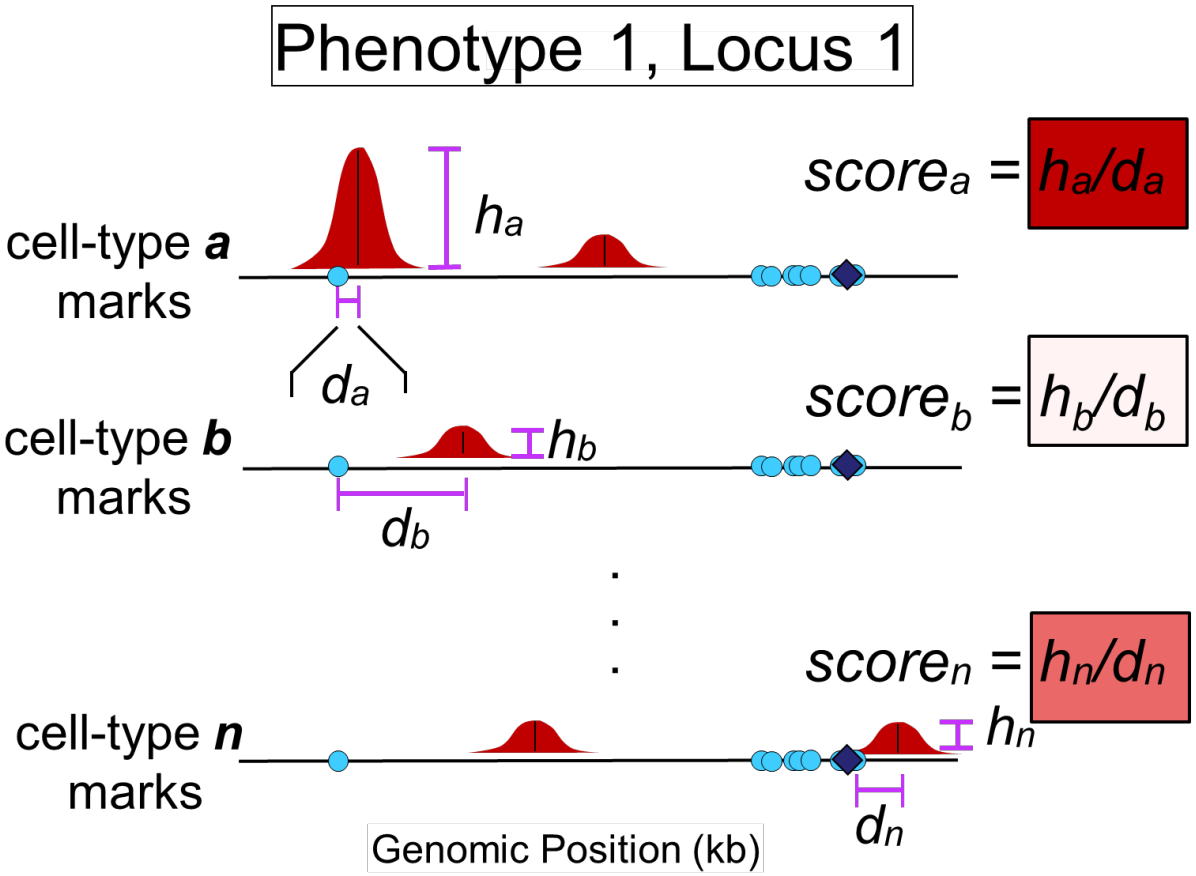
1) Utilise LD information

A. For phenotypically associated variants, find other variants in tight LD (r^2>0.8)



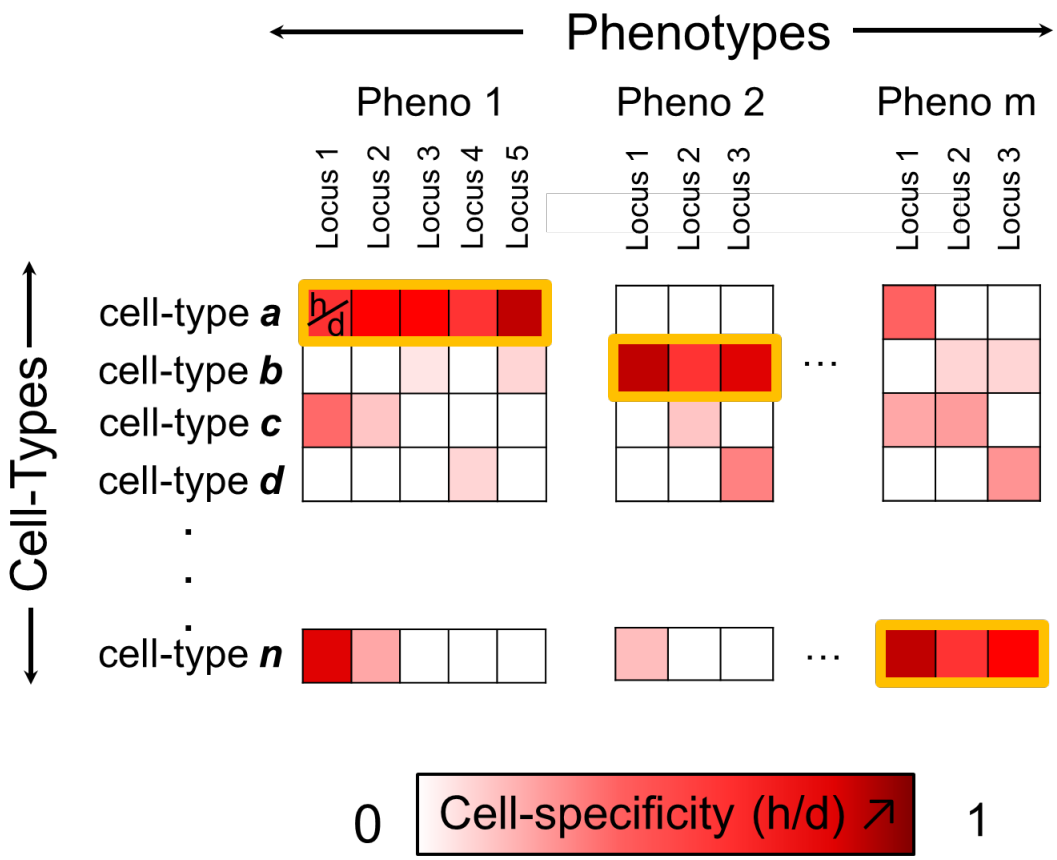
Define best scoring variant at each locus

B. Score each locus based on height and distance of the nearest peak to a variant in tight LD



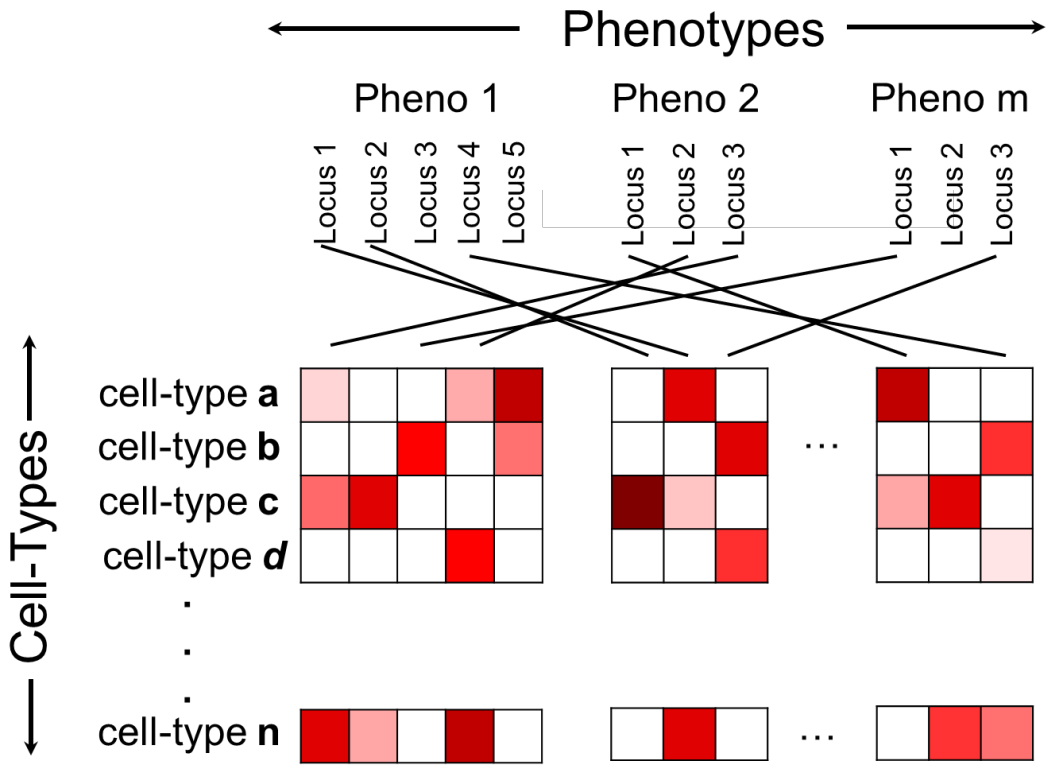
Quantify cell-type specificity of each histone mark

C. Across many phenotypes, assess if marks overlap alleles in specific cell-types



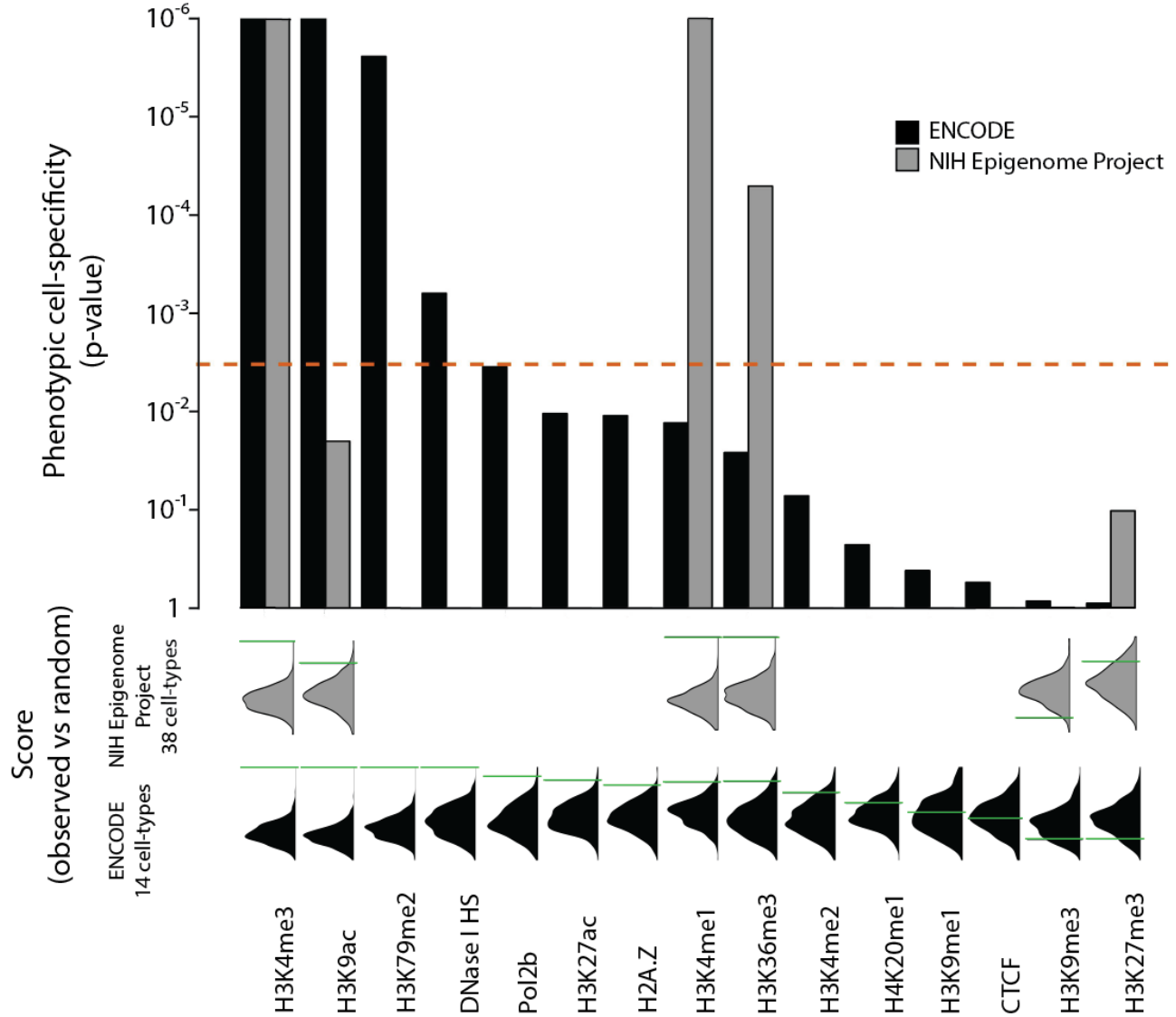
Assess the significance of cell type sepcific signal

D. Permute to assess significance of phenotypic cell-specificity



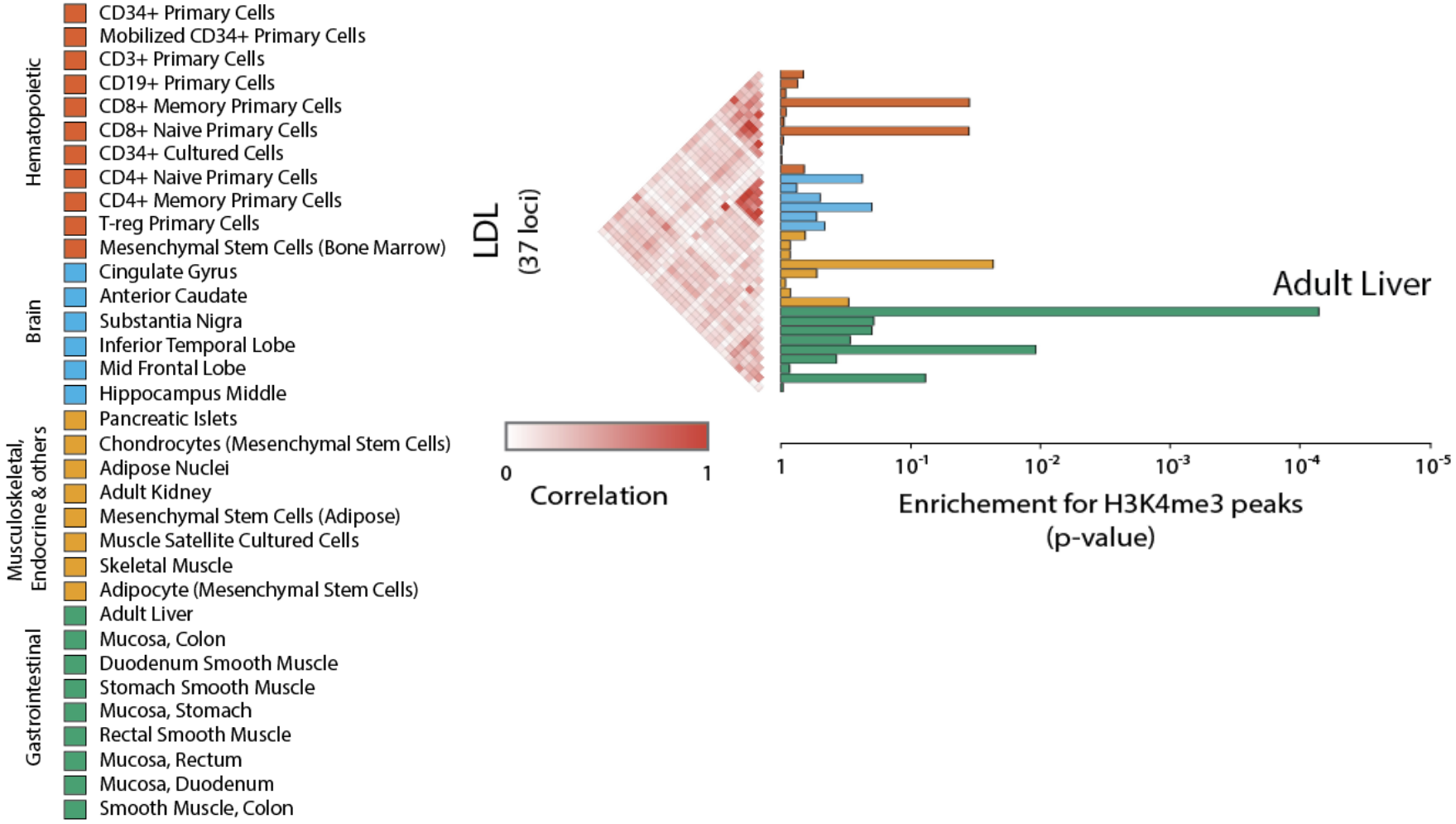
wellcome trust sanger institute

Informative marks – active gene regulation



Trynka*, Sandor* et al., *Nat Genet* 2013

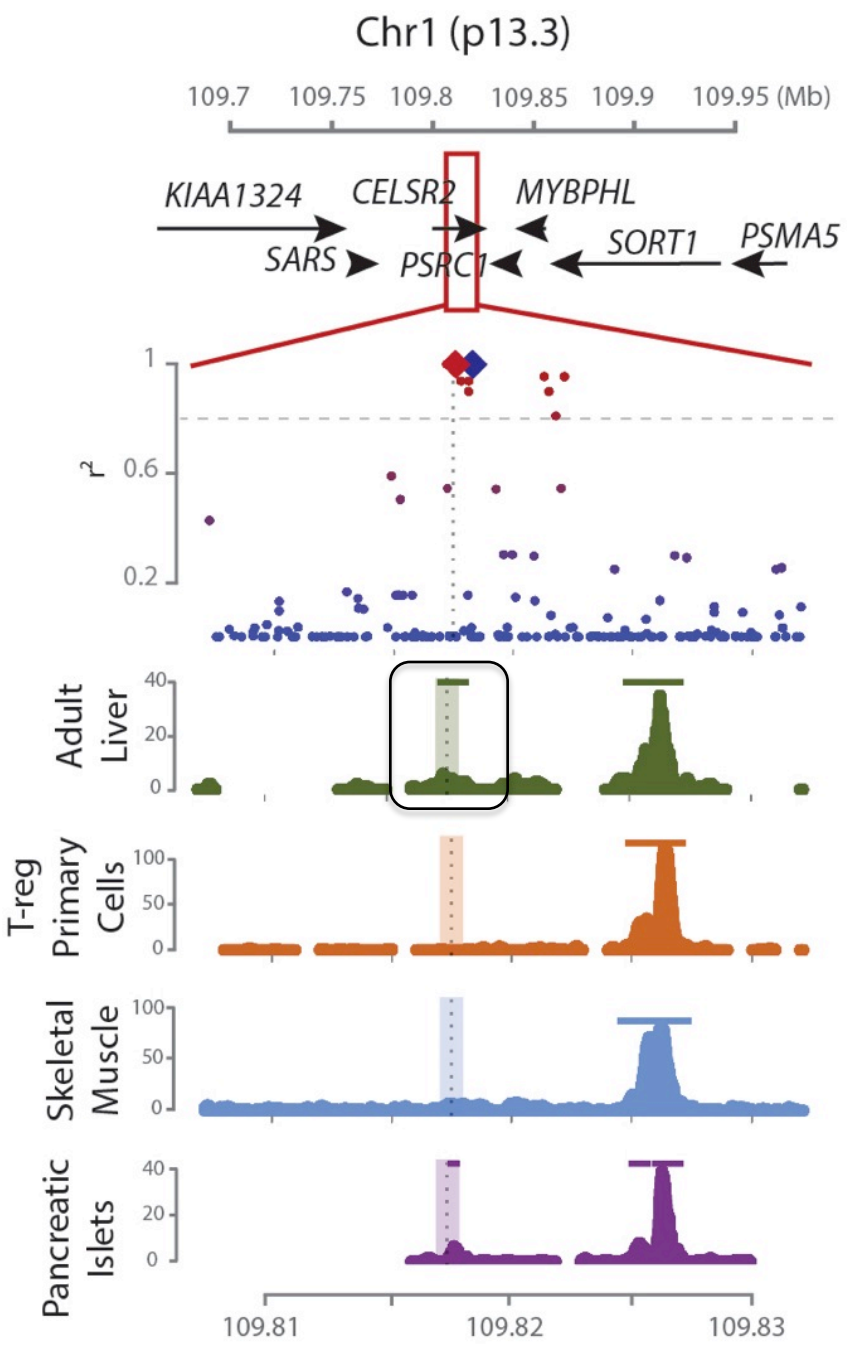
Cell type specific H3K4me3 overlap with LDL SNPs



wellcome trust sanger institute

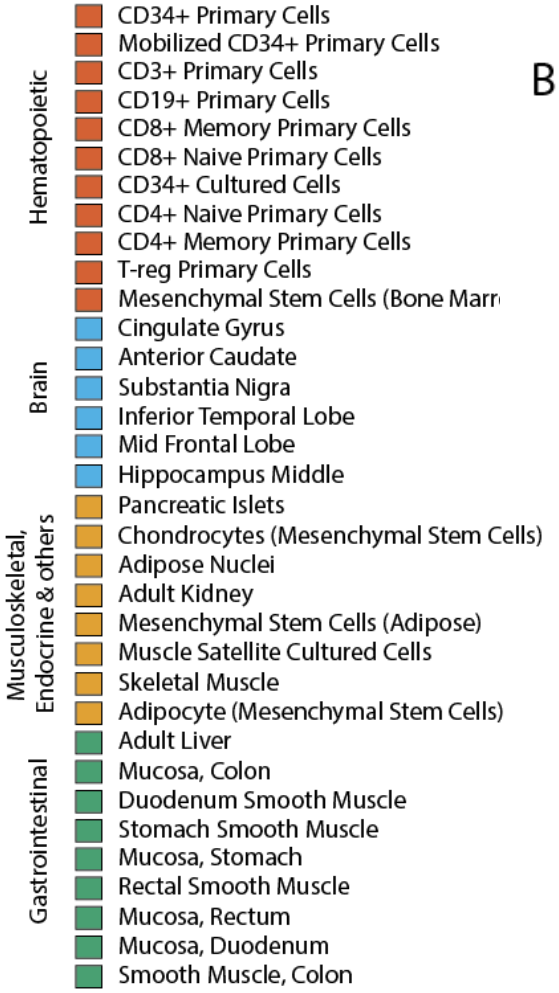
LDL – *SORT1* locus

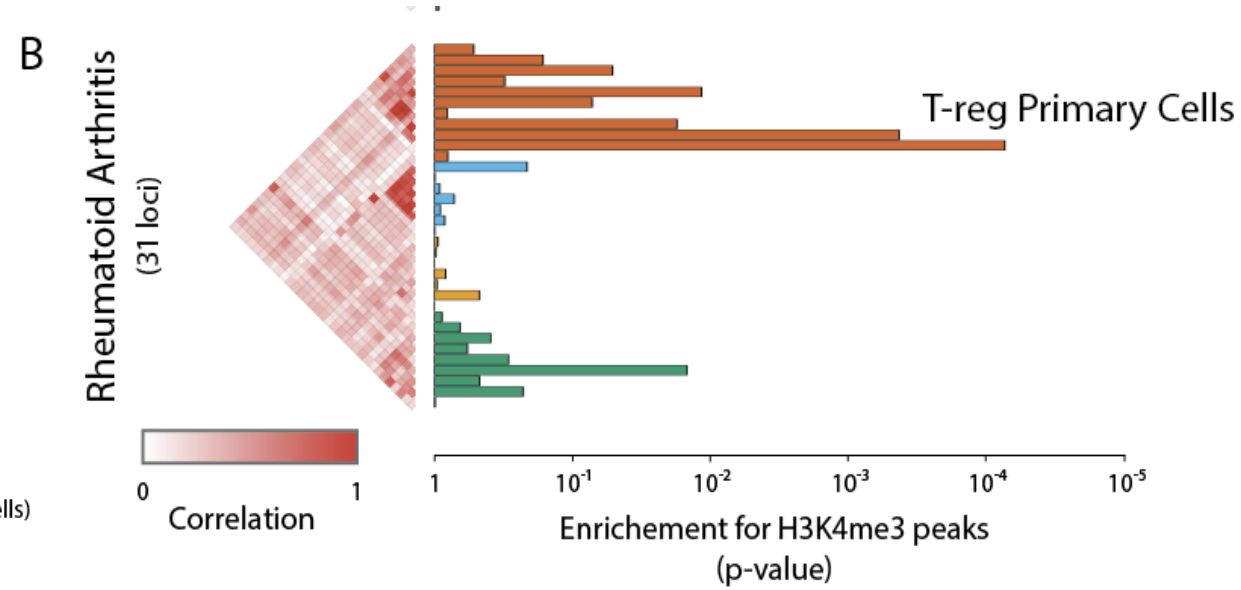
- SNP closest to H3K4me3 liver specific peak is rs12740374 (87 bp away from the summit)
- **Known** functional variant! (Musunuru et al. *Nature* 2010)



wellcome trust sanger institute

RA: CD4+ cells

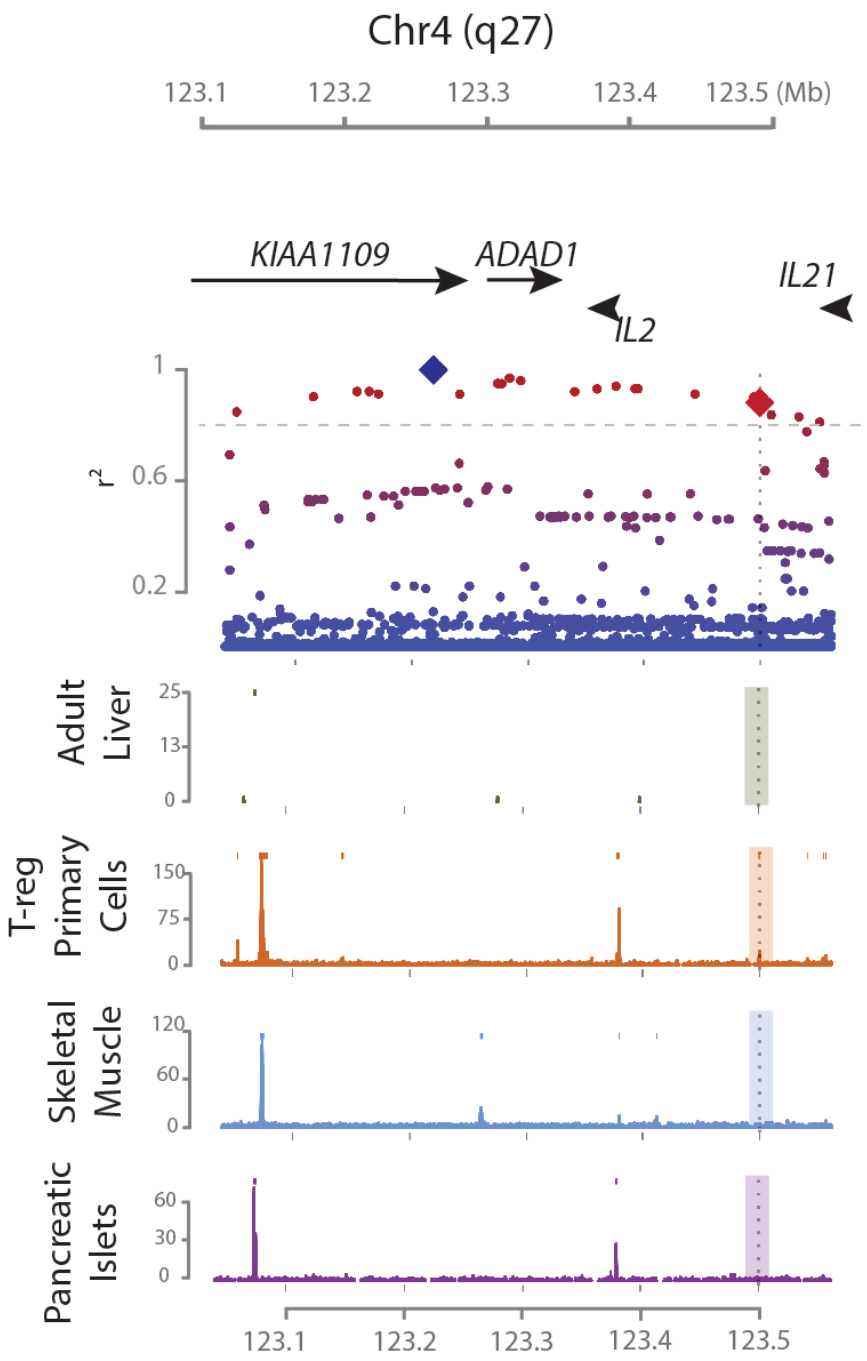




wellcome trust sanger institute

RA – *IL2/IL21* locus

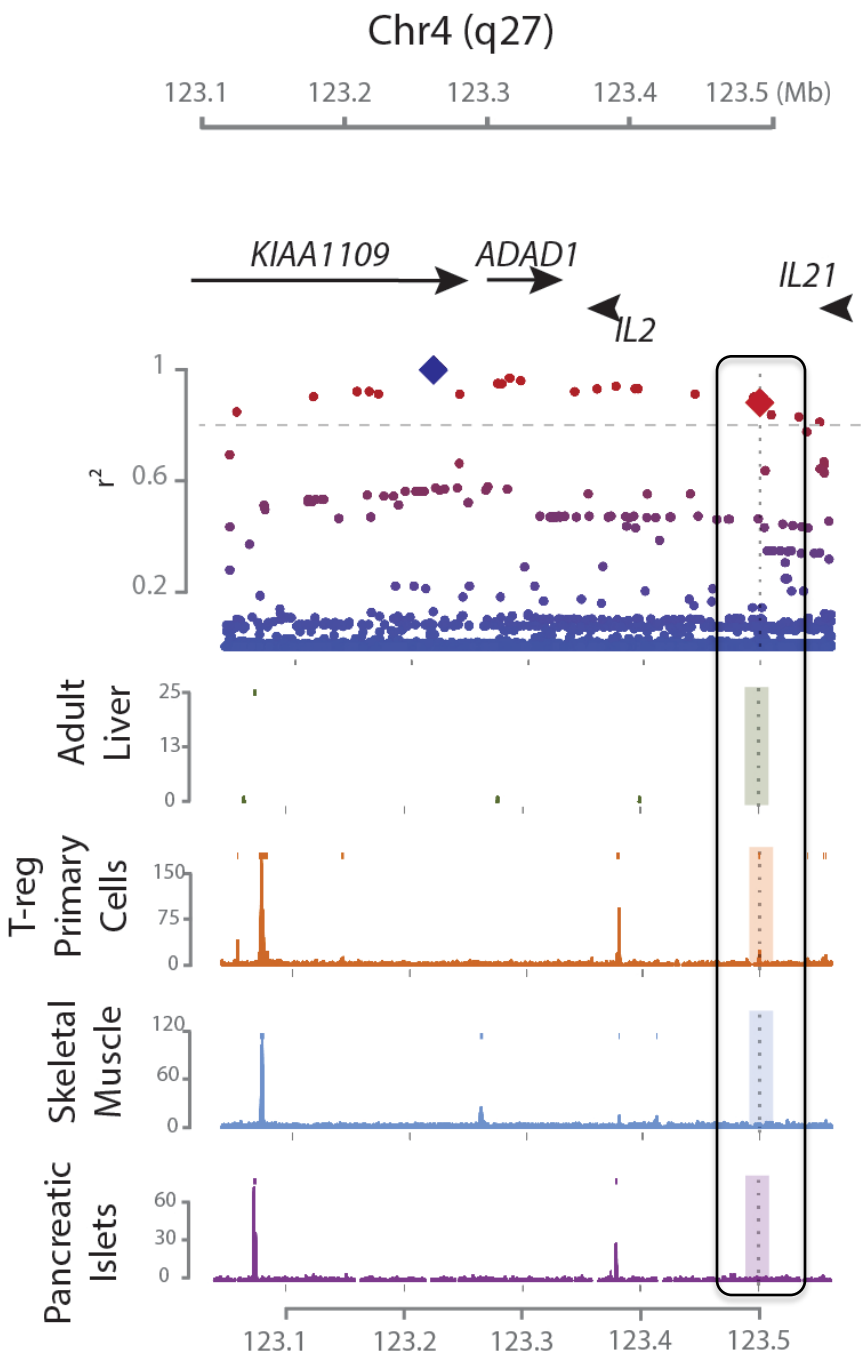
- Reported associated rs13119723 SNP is within *KIAA1109*
- LD block of over 500-kb
- rs13140464, 116 bp from the summit of the peak highly specific to the T-reg cells, located between *IL2* and *IL21* genes



wellcome trust sanger institute

RA – *IL2/IL21* locus

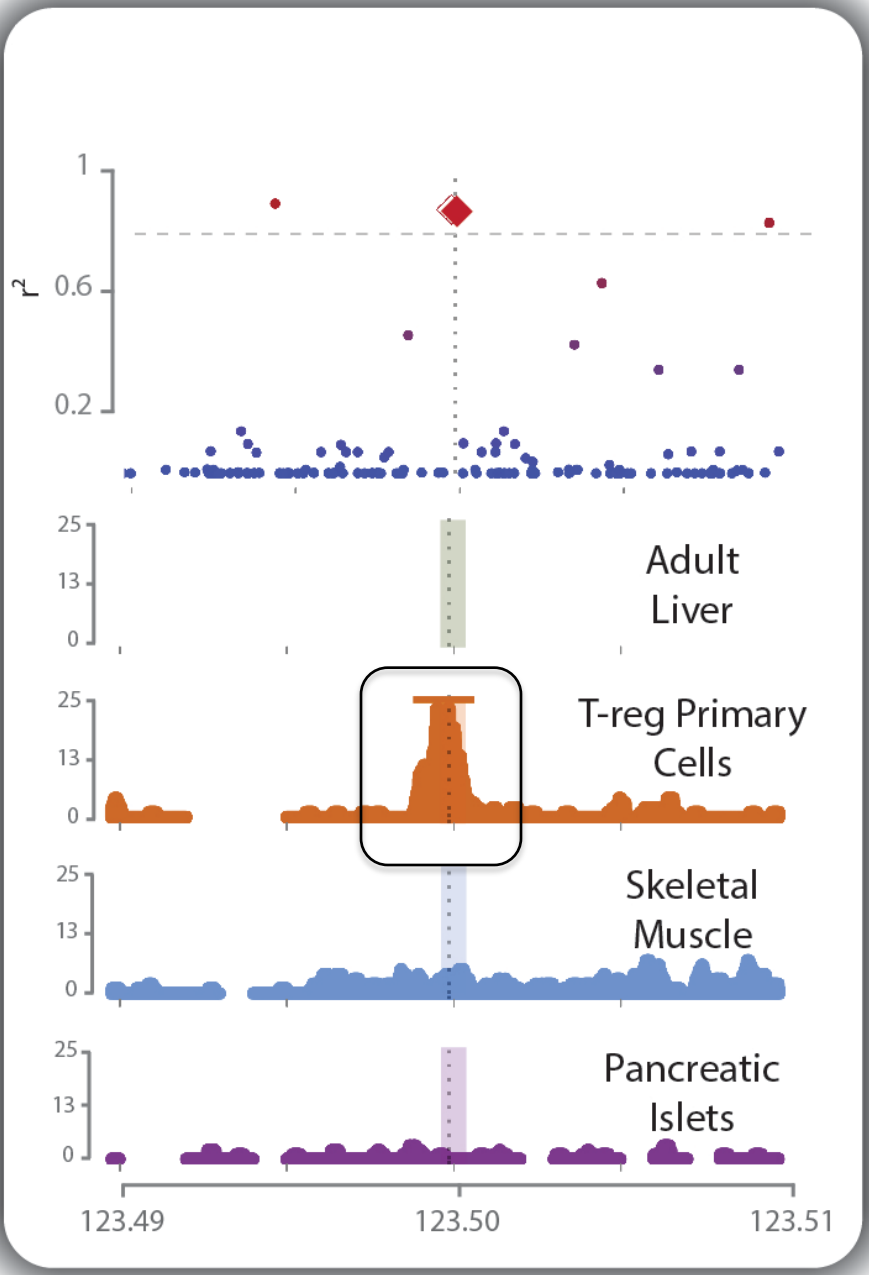
- Reported associated rs13119723 SNP is within *KIAA1109*
- LD block of over 500-kb
- rs13140464, 116 bp from the summit of the peak highly specific to the T-reg cells, located between *IL2* and *IL21* genes



wellcome trust sanger institute

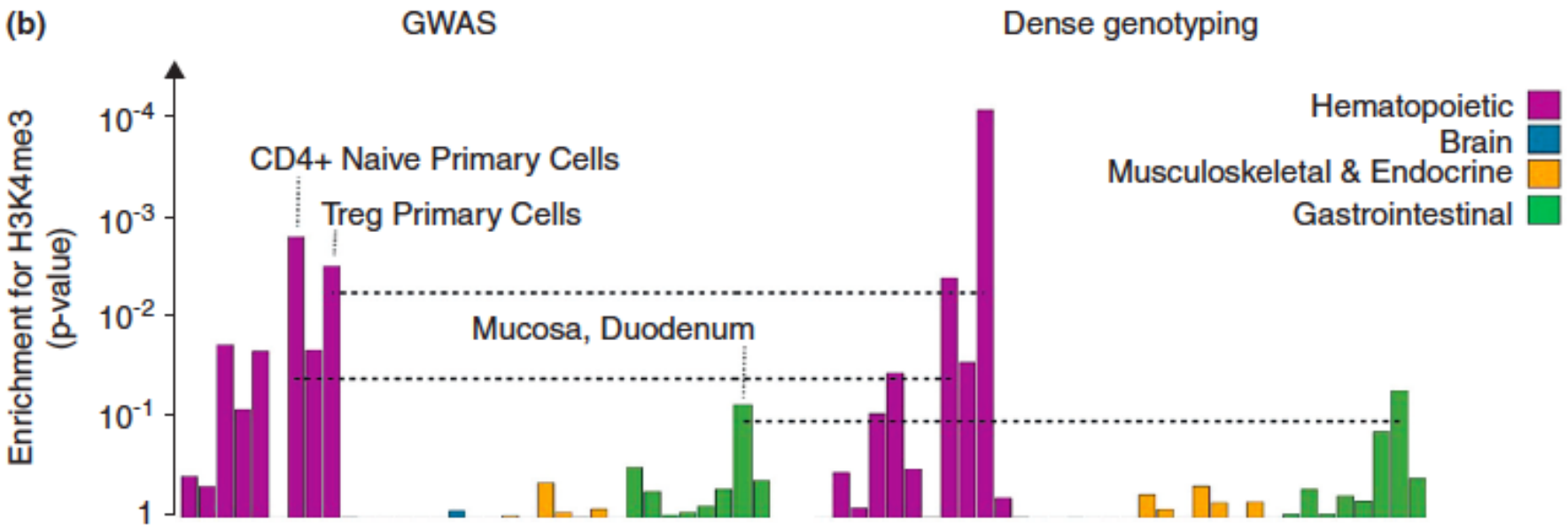
RA – *IL2/IL21* locus

- Reported associated rs13119723 SNP is within *KIAA1109*
- LD block of over 500-kb
- rs13140464, 116 bp from the summit of the peak highly specific to the T-reg cells, located between *IL2* and *IL21* genes



wellcome trust sanger institute

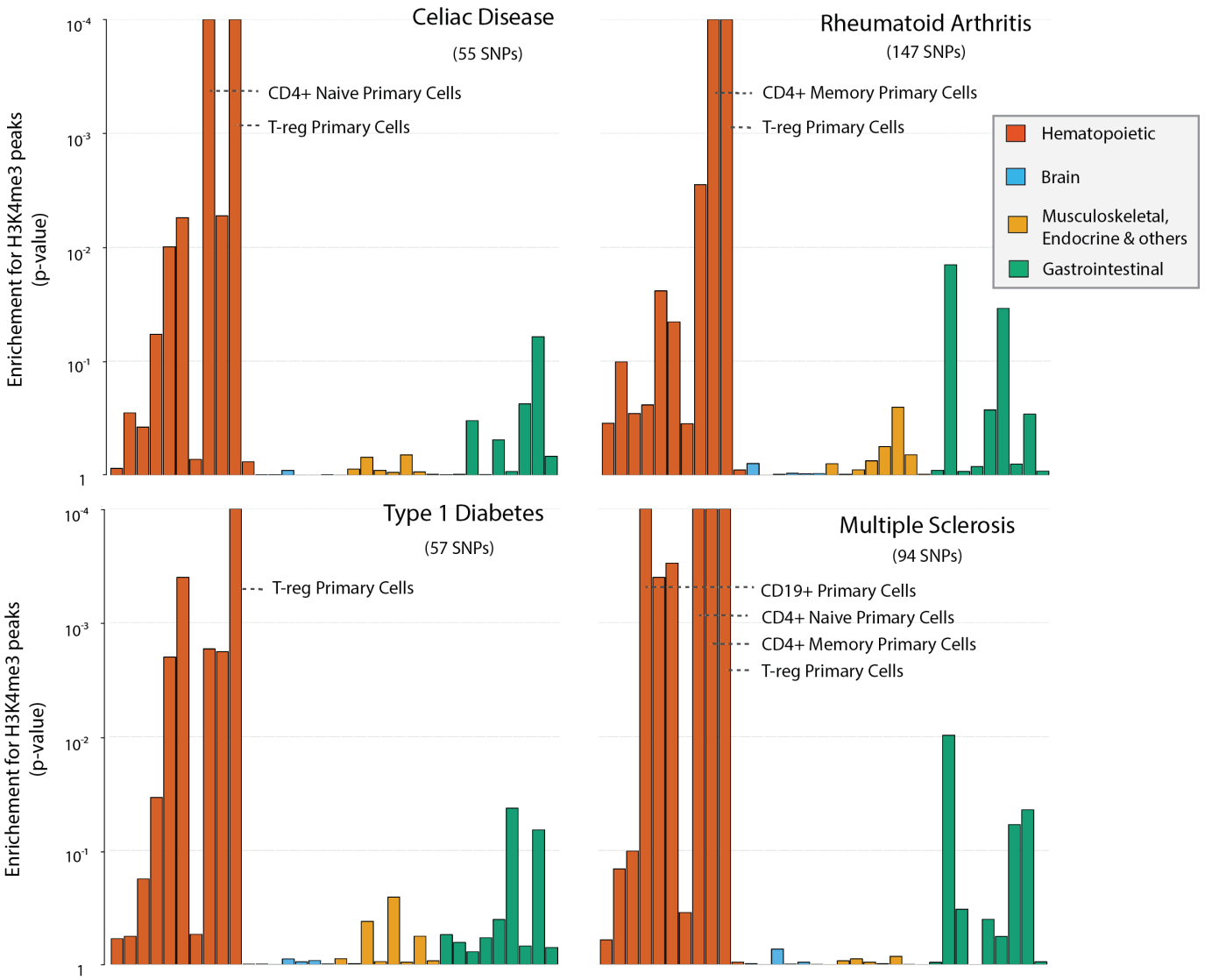
Statistical refinement of association signals improves tissue specific signal



Trynka and Raychaudhuri, *Curr Opin Genet Dev* 2013

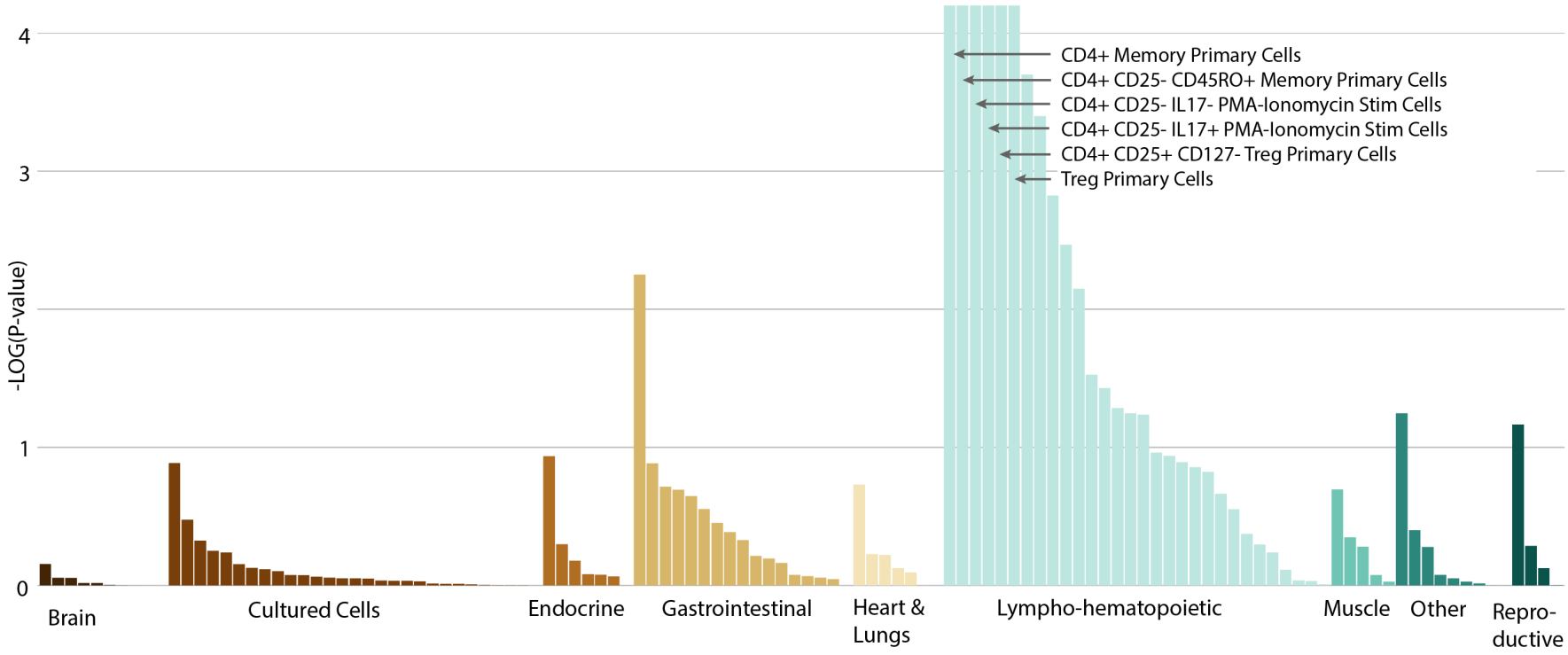
wellcome trust sanger institute

The role of CD4+ subsets in autoimmune diseases



wellcome trust sanger institute

T1D variants implicate a limited set of cell types



wellcome trust sanger institute

Predicting Cell Types and Genetic Variations Contributing to Disease by Combining GWAS and Epigenetic Data

Anna Gerasimova*, Lukas Chavez, Bin Li, Gregory Seumois, Jason Greenbaum, Anjana Rao, Pandurangan Vijayanand, Bjoern Peters

La Jolla Institute for Allergy and Immunology, La Jolla, California, United States of America

Breast cancer risk–associated SNPs modulate the affinity of chromatin for FOXA1 and alter gene expression

Richard Cowper-Sal·lari[1,2,7], Xiaoyang Zhang[1,2,7], Jason B Wright[1], Swneke D Bailey[3,4], Jerome Eeckhoute[5,6], Jason H Moore[1,2] & Mathieu Lupien[3,4]

Maps of Open Chromatin Guide the Functional Follow-Up of Genome-Wide Association Signals: Application to Hematological Traits

[1,2,3], Augusto Rendon[2,3,4], Katta [2,5], Suthesh Sivapalaratnam[6], Hanneke R Consortium[¶], Berthold Göttgens[2,5], Nicole

Systematic functional regulatory assessment of disease-associated variants

Konrad J. Karczewski[a,b,1], Joel T. Dudley[a,c,1,2], Kimberly R. Kukurba[b], Rong Chen[c], Atul J. Butte[c], Stephen B. Montgomery[b,d], and Michael Snyder[b,3]

Linking disease associations with regulatory information in the human genome

Marc A. Schaub,[1] Alan P. Boyle,[2] Anshul Kundaje,[1] Serafim Batz

Systematic Localization of Common Disease-Associated Variation in Regulatory DNA

Matthew T. Maurano,[1*] Richard Humbert,[1*] Eric Rynes,[1*] Robert E. Thurman,[1] Eric Haugen,[1] Hao Wang,[1] Alex P. Reynolds,[1] Richard Sandstrom,[1] Hongzhu Qu,[1,2] Jennifer Brody,[3] Anthony Shafer,[1] Fidencio Neri,[1] Kristen Lee,[1] Tanya Kutyavin,[1] Sandra Stehling-Sun,[1] Audra K. Johnson,[1] Theresa K. Canfield,[1] Erika Giste,[1] Morgan Diegel,[1] Daniel Bates,[1] R. Scott Hansen,[4] Shane Neph,[1] Peter J. Sabo,[1] Shelly Heimfeld,[5] Antony Raubitschek,[6] Steven Ziegler,[6] Chris Cotsapas,[7,8] Nona Sotoodehnia,[3,9] Ian Glass,[10] Shamil R. Sunyaev,[11] Rajinder Kaul,[4] John A. Stamatoyannopoulos[1,12]†

ARTICLE

doi:10.103

An integrated encyclopedia of DNA elements in the human genome

The ENCODE Project Consortium*

sanger institute

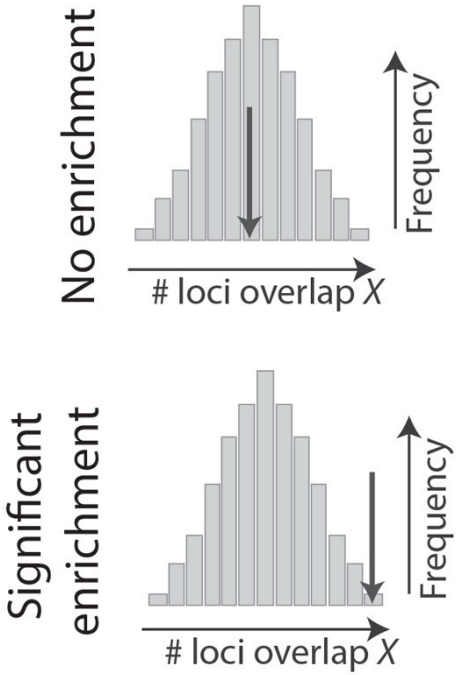
GWAS variants and enrichment across a range of genomic annotations

- A robust test to assess enrichment of disease SNPs with any genomic annotation

wellcome trust sanger institute

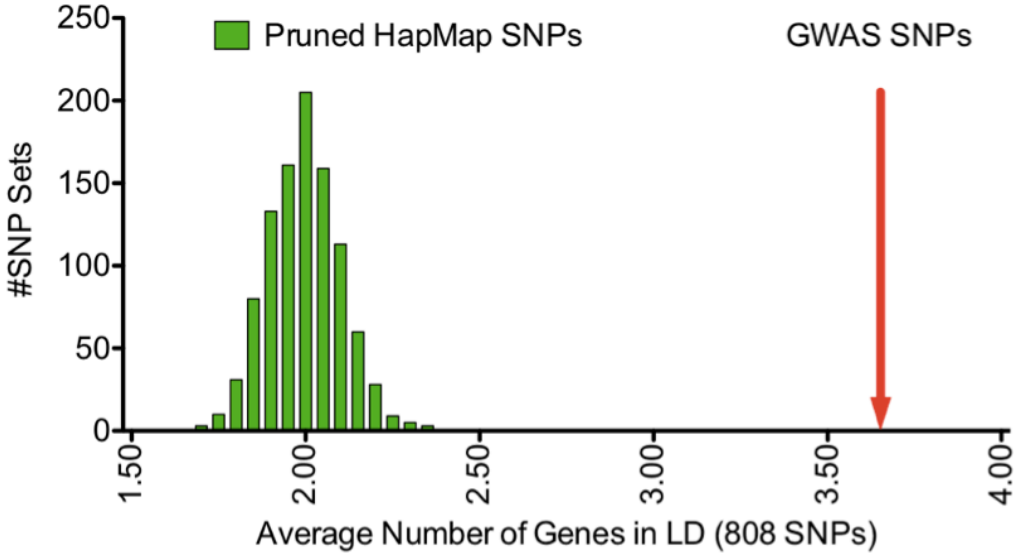
Caveats in enrichment testing

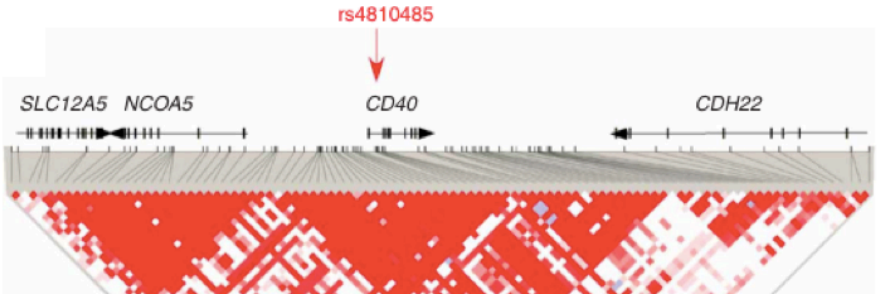
- Matching based enrichment tests can be biased – require a prior knowledge of relevant matching parameters



wellcome trust sanger institute

Distinct properties at GWAS associated loci

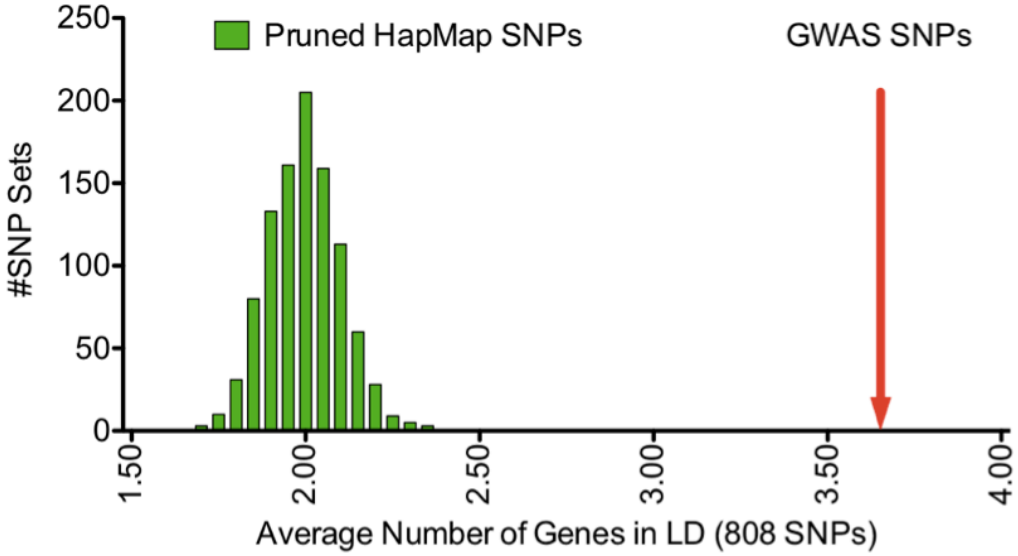


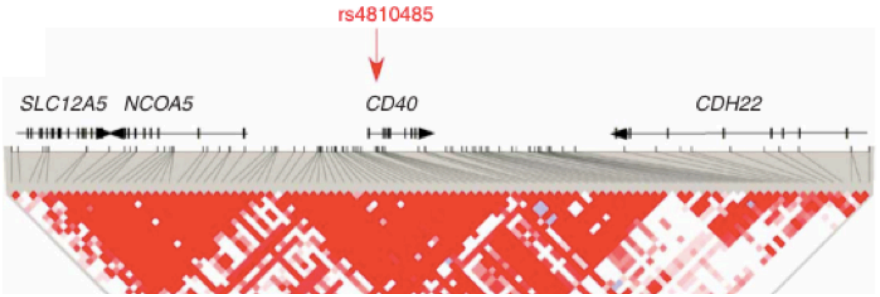


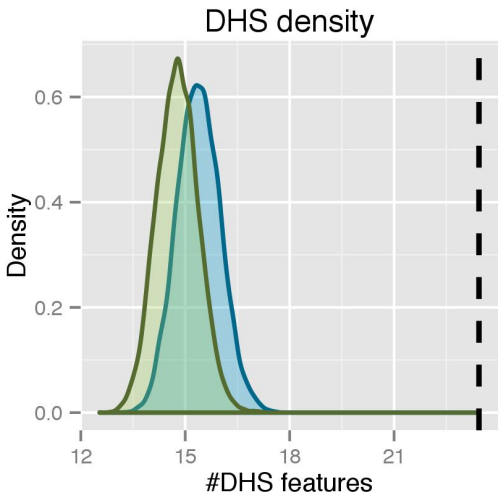
wellcome trust sanger institute

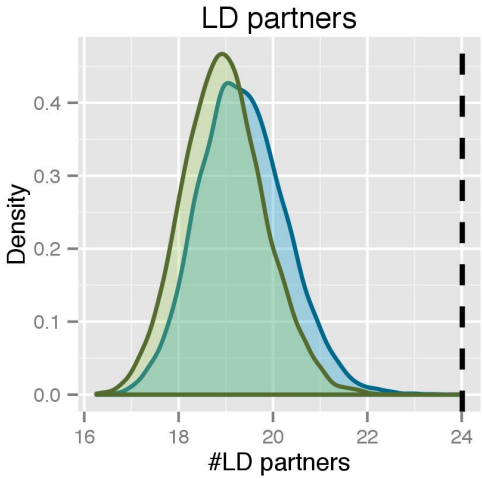
Distinct properties at GWAS associated loci

Matched
Random



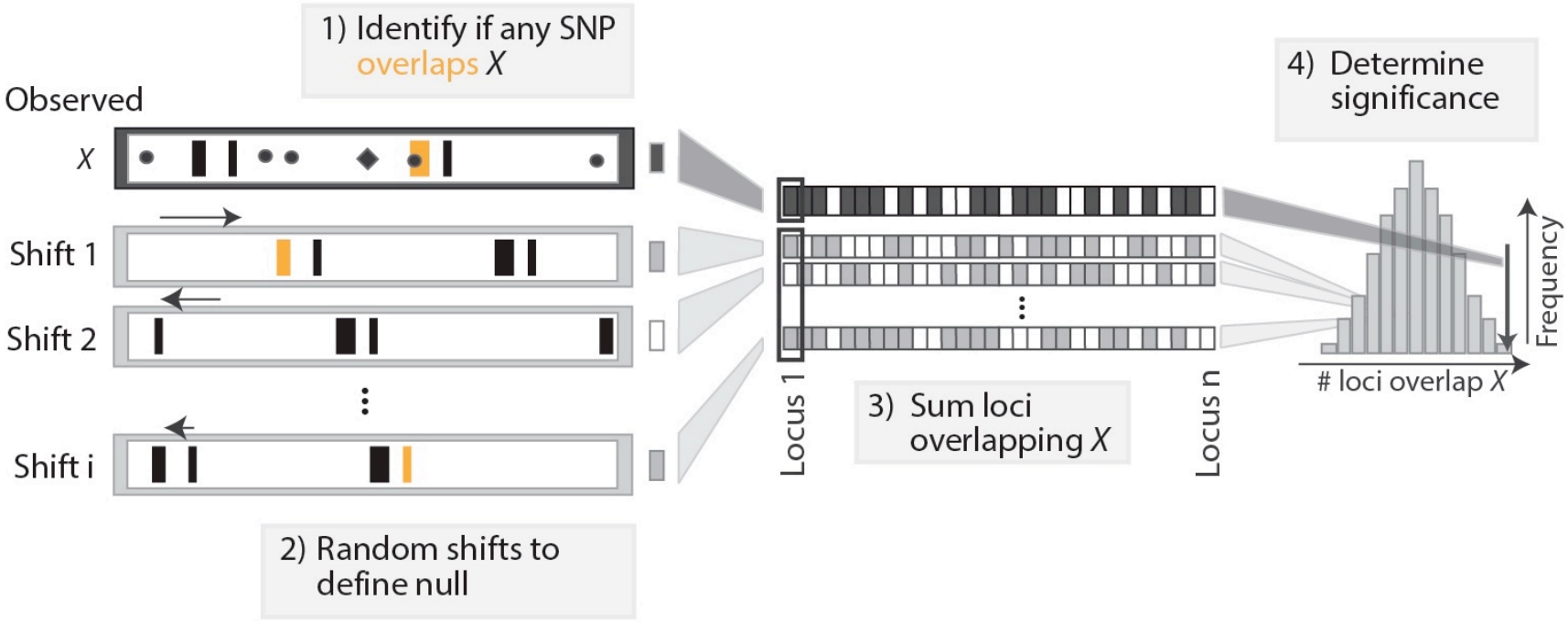






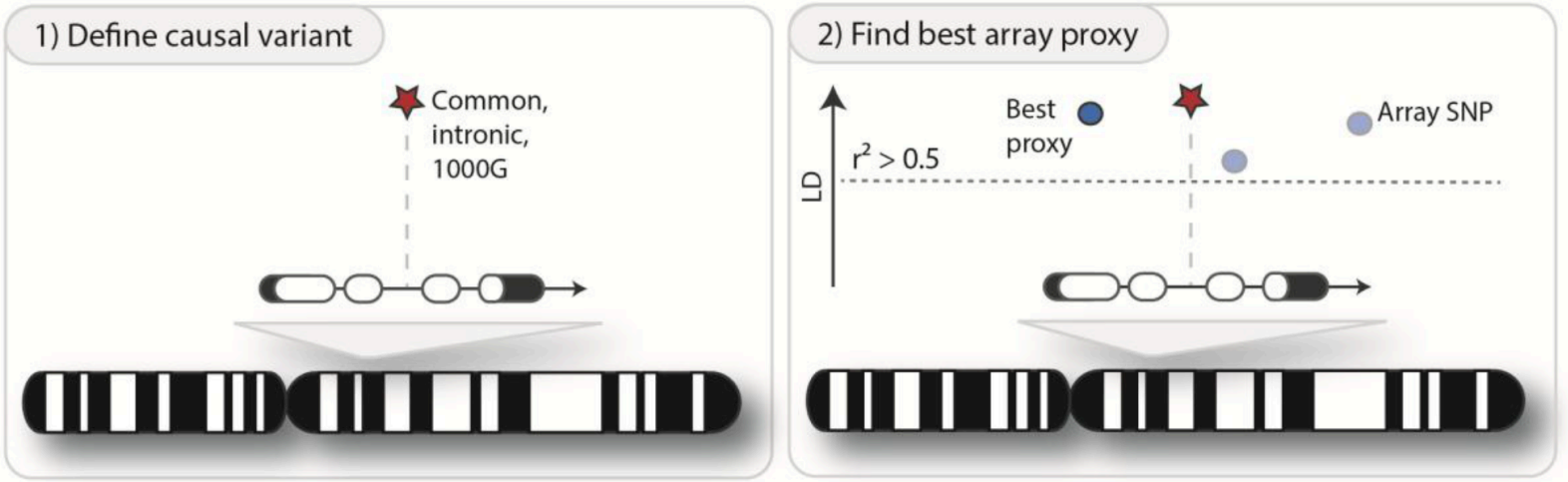
wellcome trust sanger institute

Genomic Annotation Shifter (*GoShifter*)



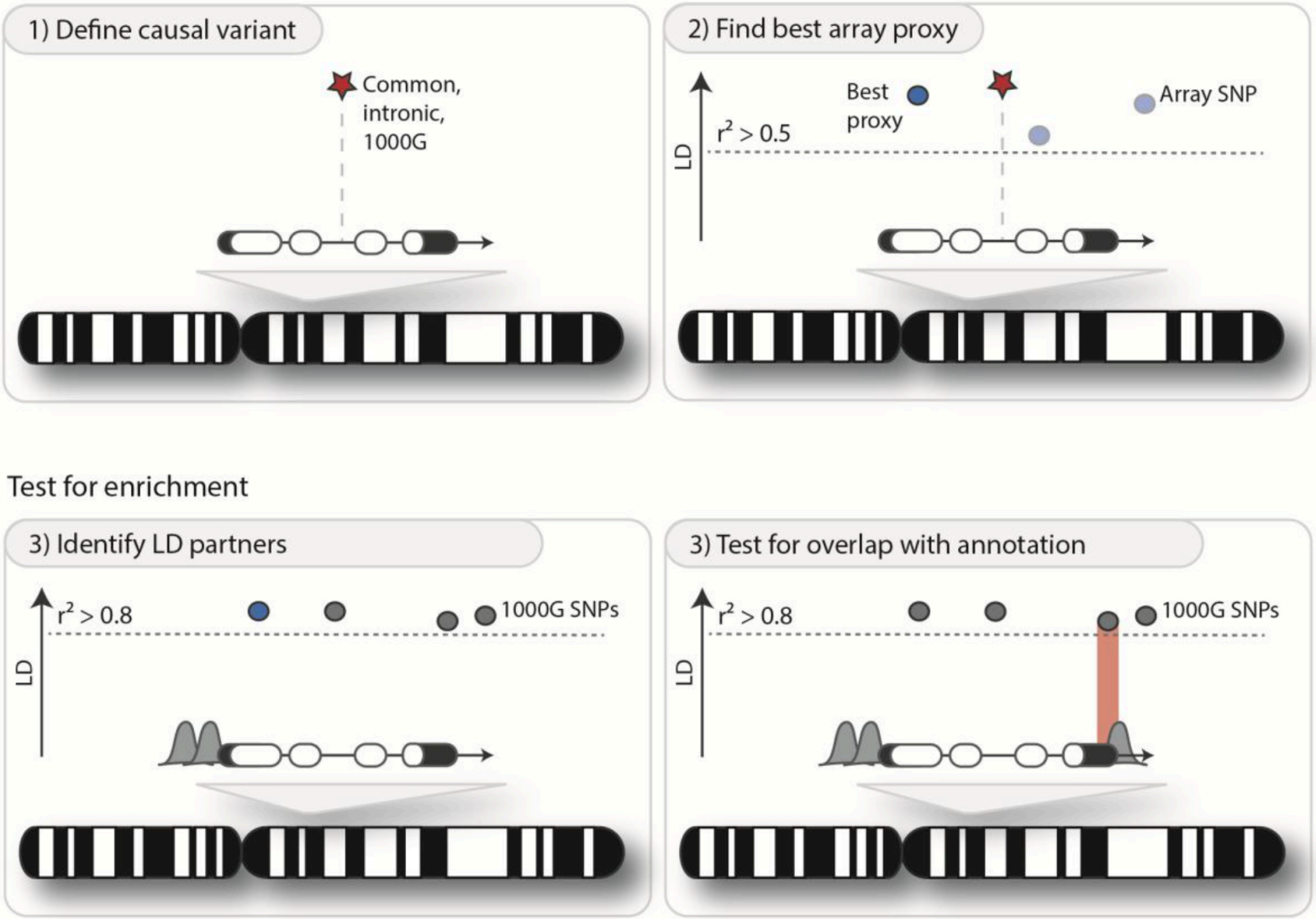
Trynka et al. *AJHG*, 2015

Simulate GWAS results

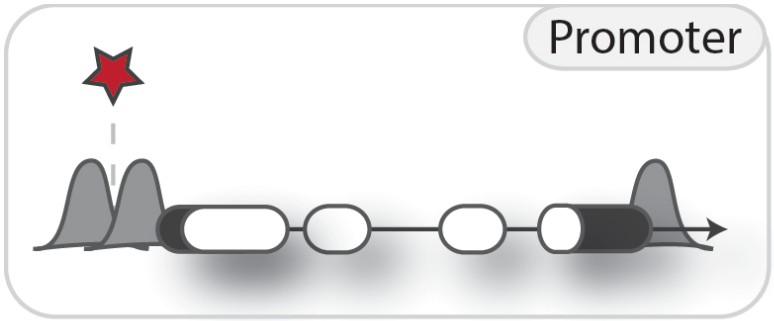


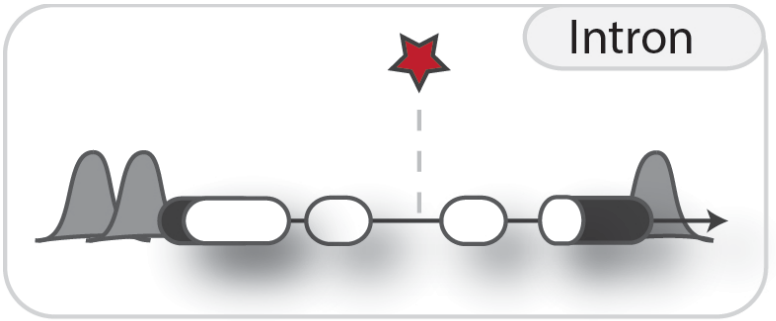
wellcome trust sanger institute

Simulate GWAS results



Assumption





- Promoter catalogs should enrich for DHS peaks

- Intron catalogs should not enrich for DHS peaks

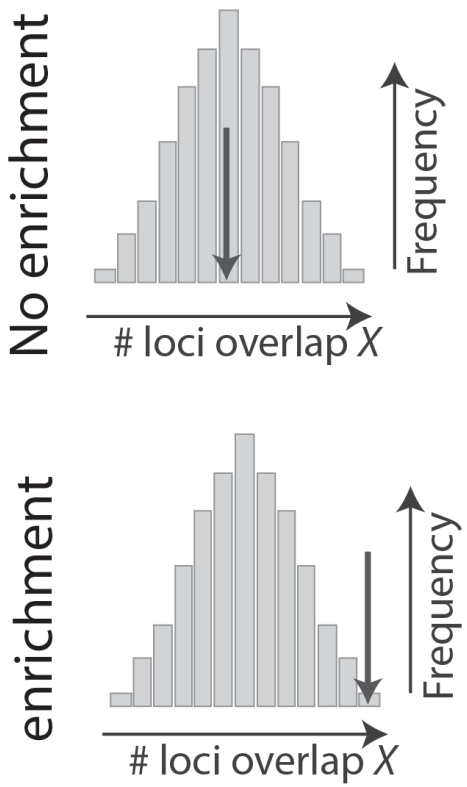
wellcome trust sanger institute

Dataset

- DHS peaks from the ENCODE and Roadmap Epigenomics Projects

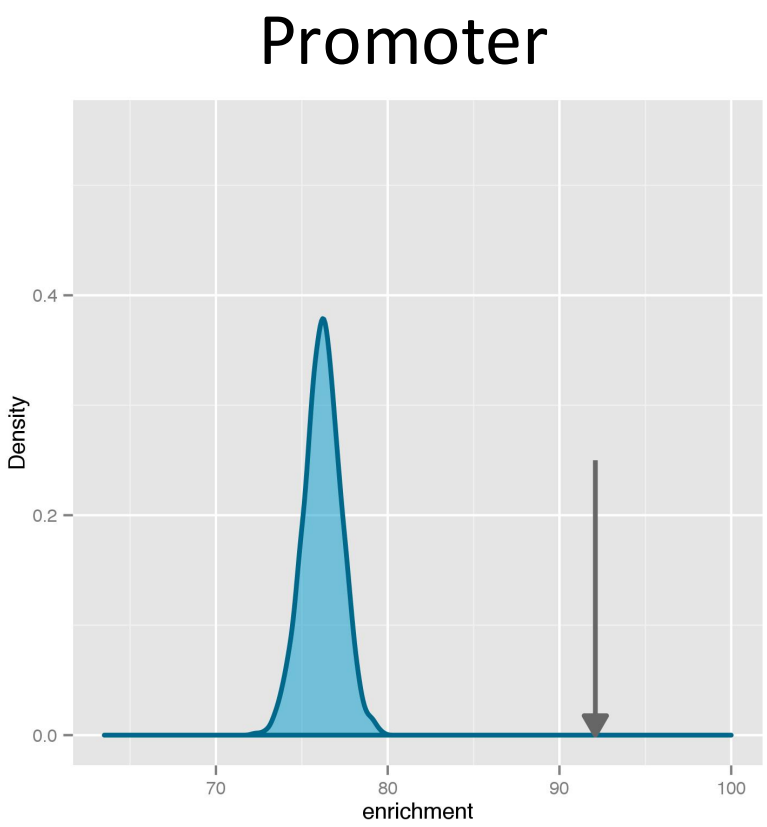
- DNase hypersensitivity data from:
  - 217 cell and tissue samples
  - 1,376,301 distinct DHS positions
  - Collectively spanning 16.4% of the genome

Interpretation of the results



wellcome trust sanger institute

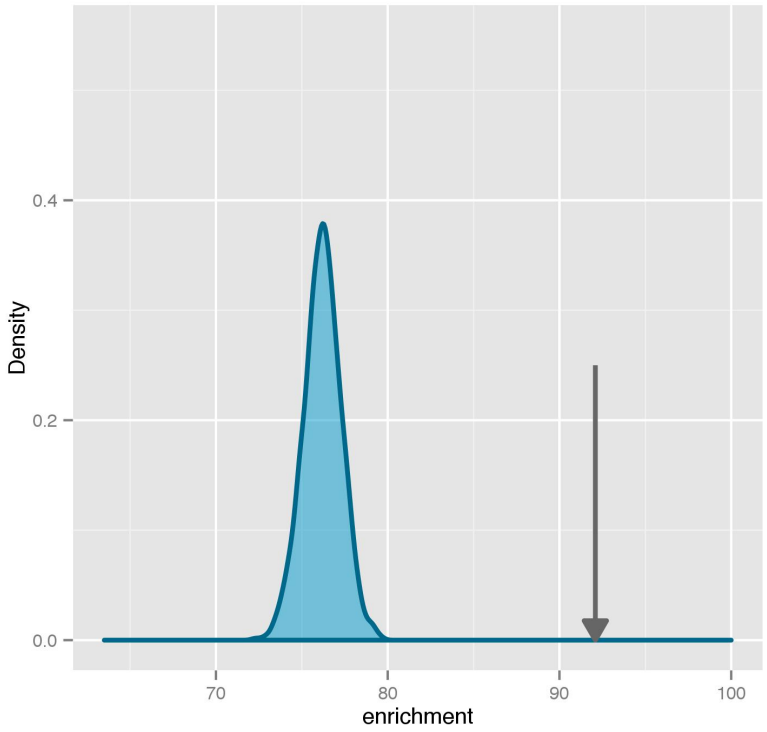
Standard matching on GEN, MAF and TSS



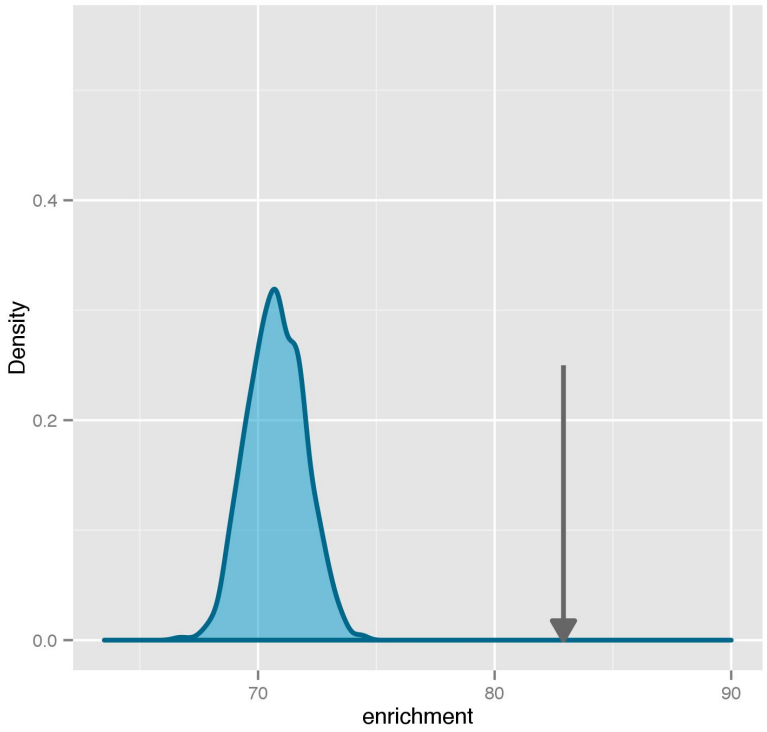
wellcome trust sanger institute

Standard matching on GEN, MAF and TSS

Promoter

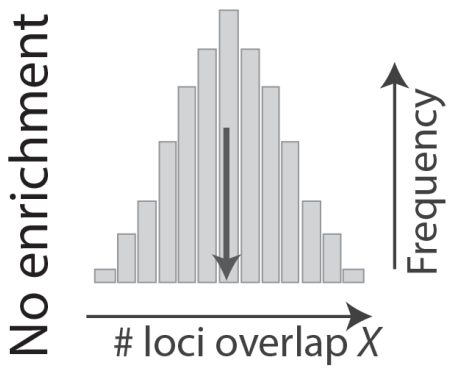


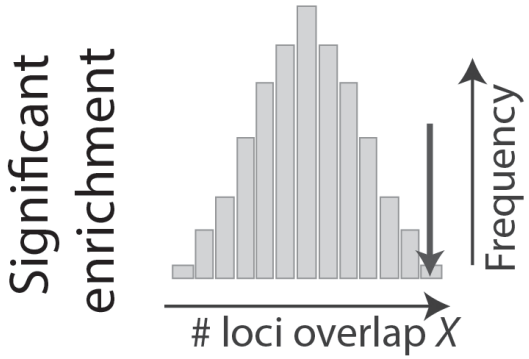
Intron

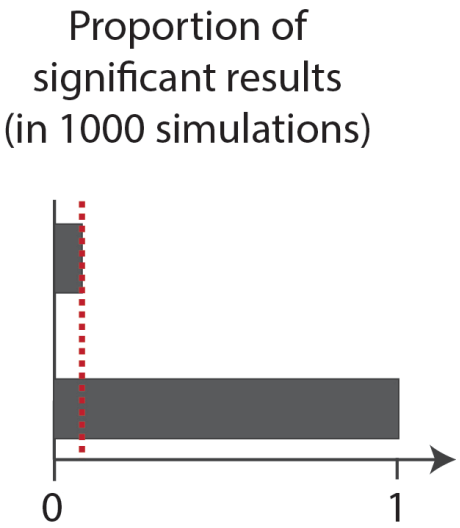


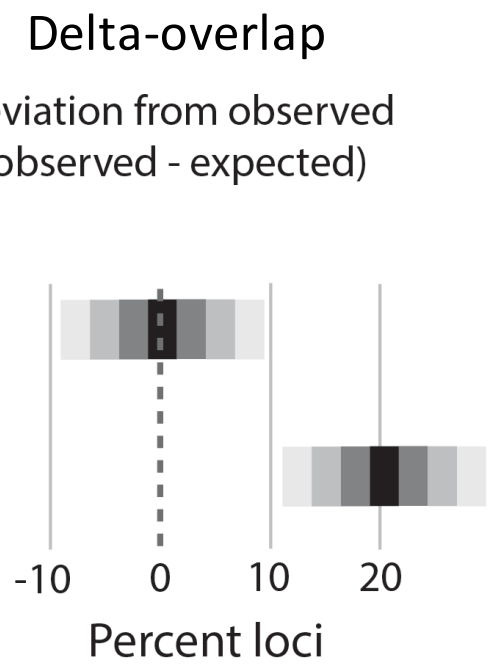
wellcome trust sanger institute

Interpretation of the results



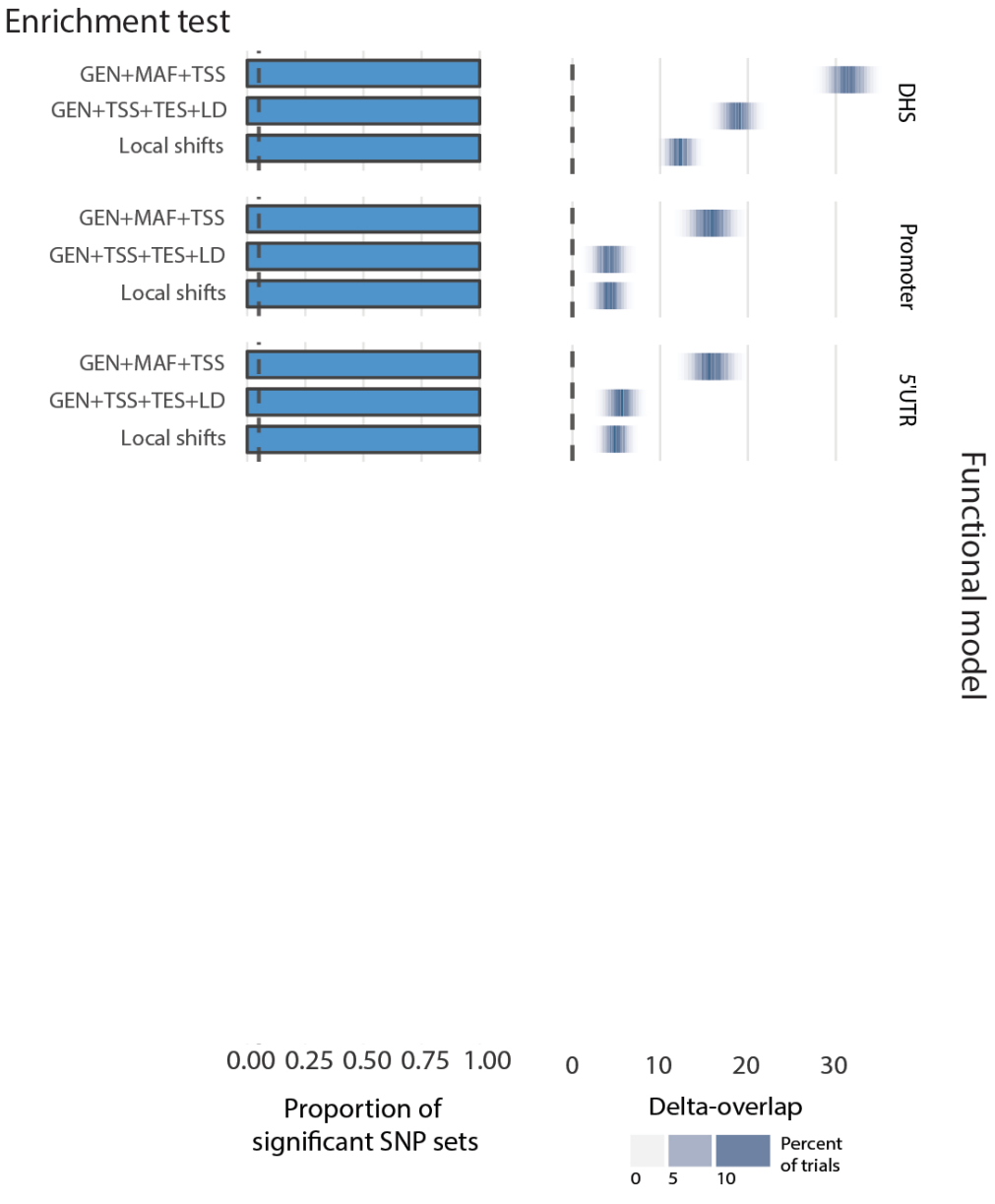




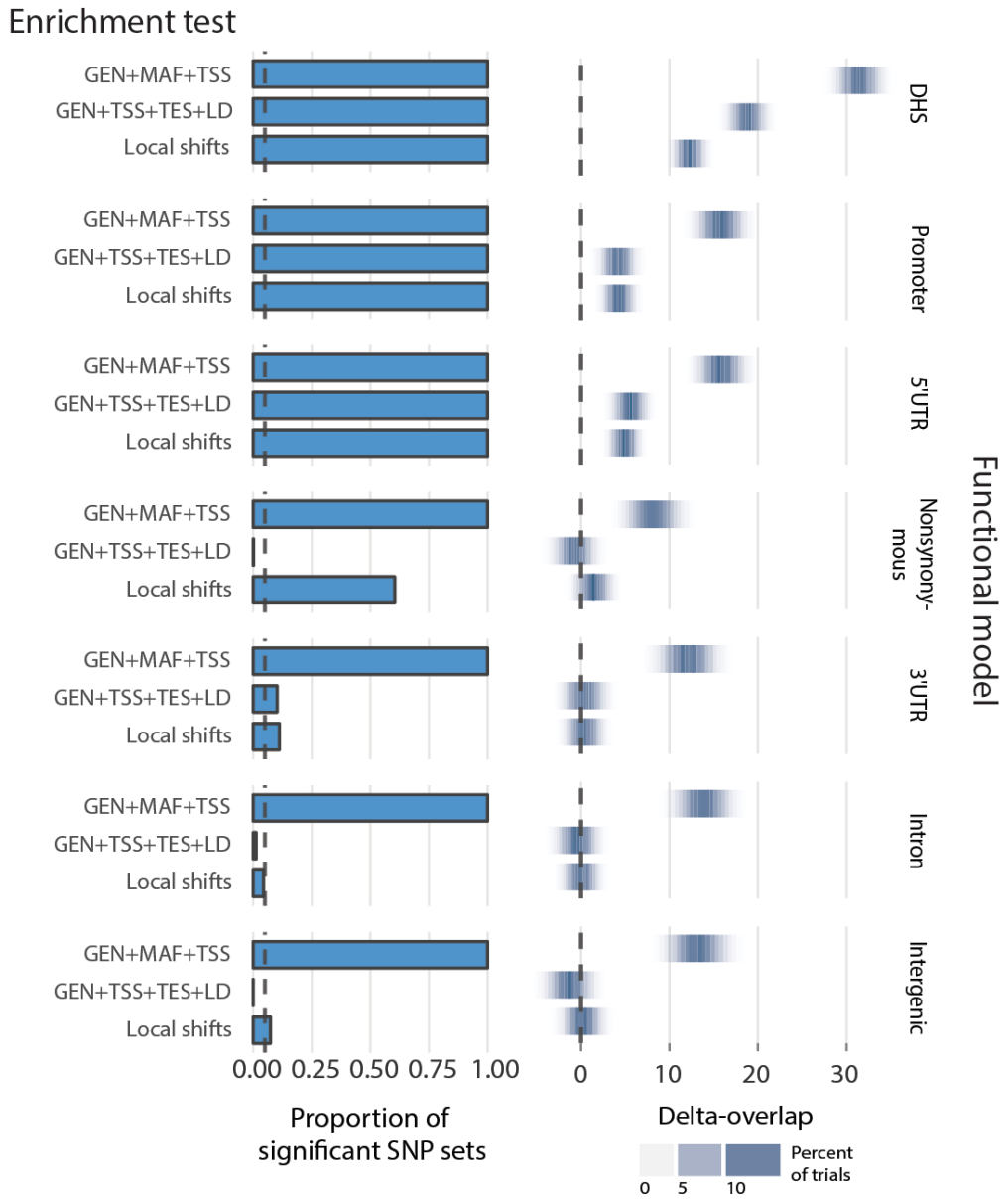


wellcome trust sanger institute

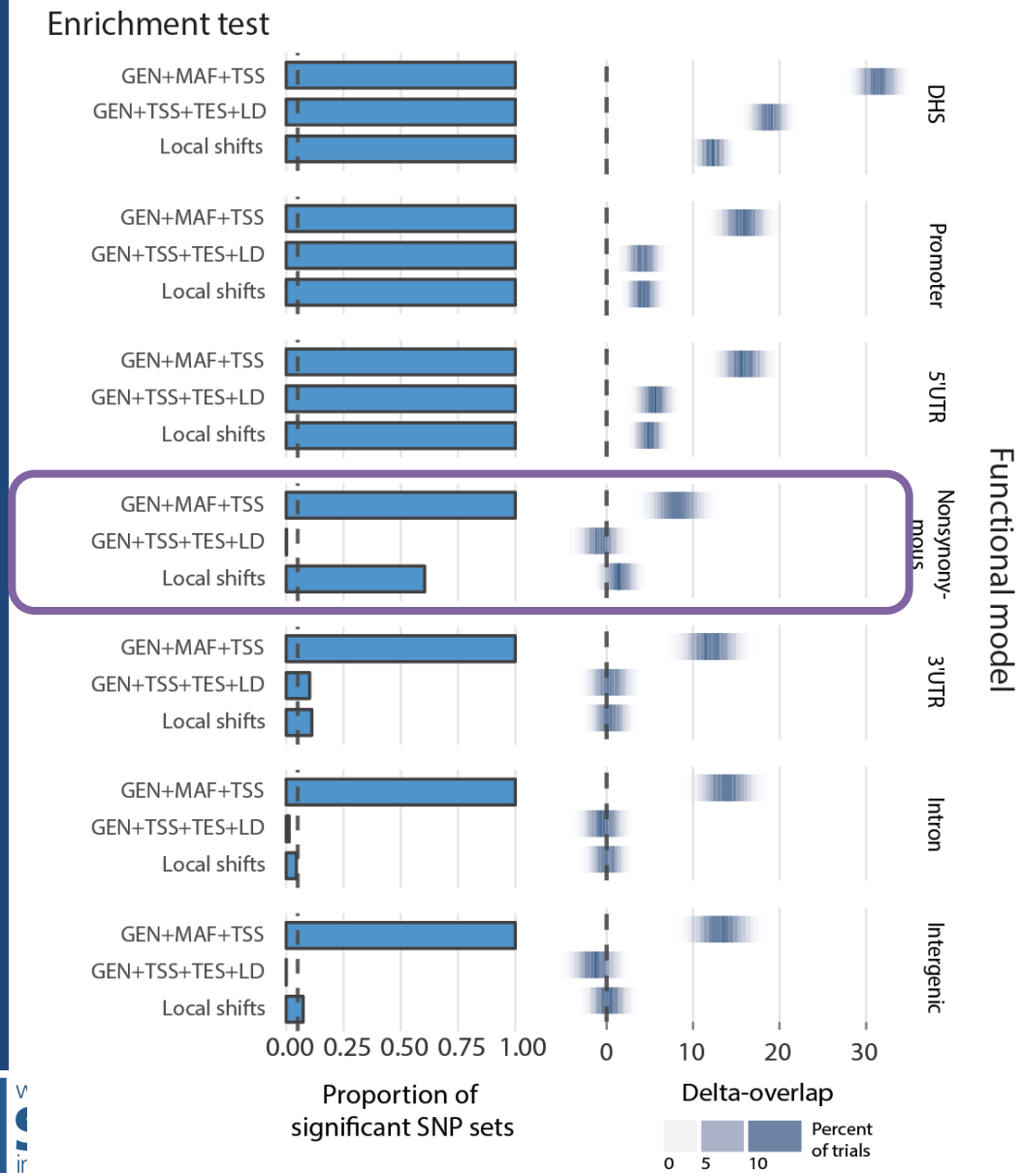
Benchmarking enrichment methods using simulated sets of SNPs



Benchmarking enrichment methods using simulated sets of SNPs

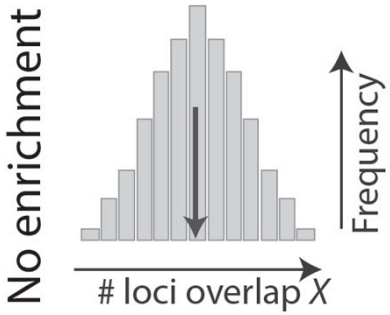


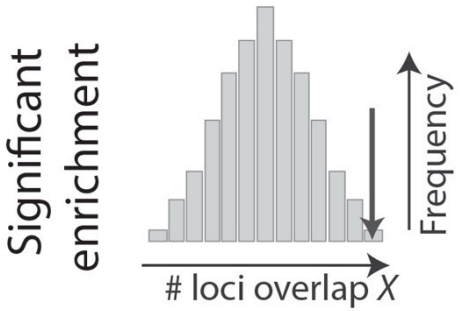
Benchmarking enrichment methods using simulated sets of SNPs



Caveats in enrichment testing

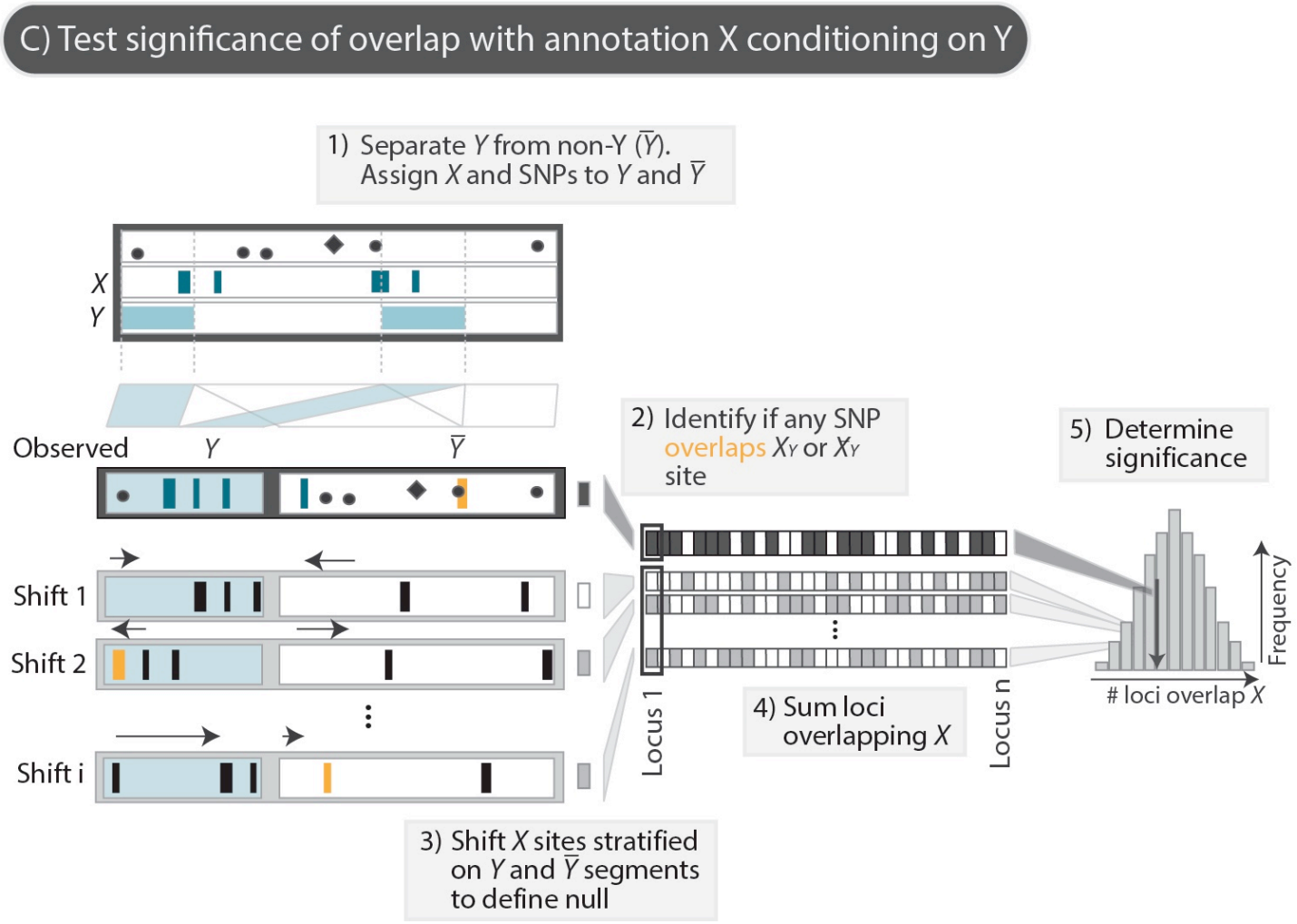
- Matching based enrichment tests can be biased – require a prior knowledge of relevant matching parameters





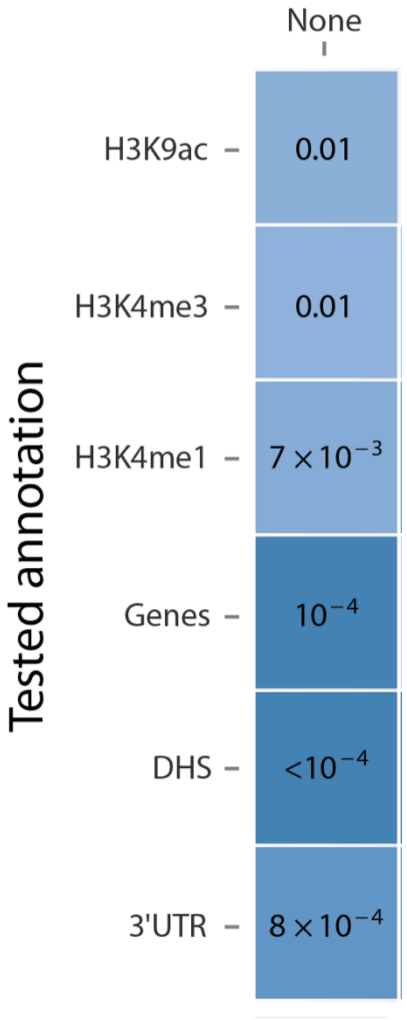
- Genomic annotations colocalize

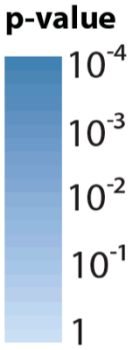
Differentiating effects of colocalizing annotations



Trynka et al. *AJHG*, 2015

eQTLs are independently enriched for DHS and 3'UTR

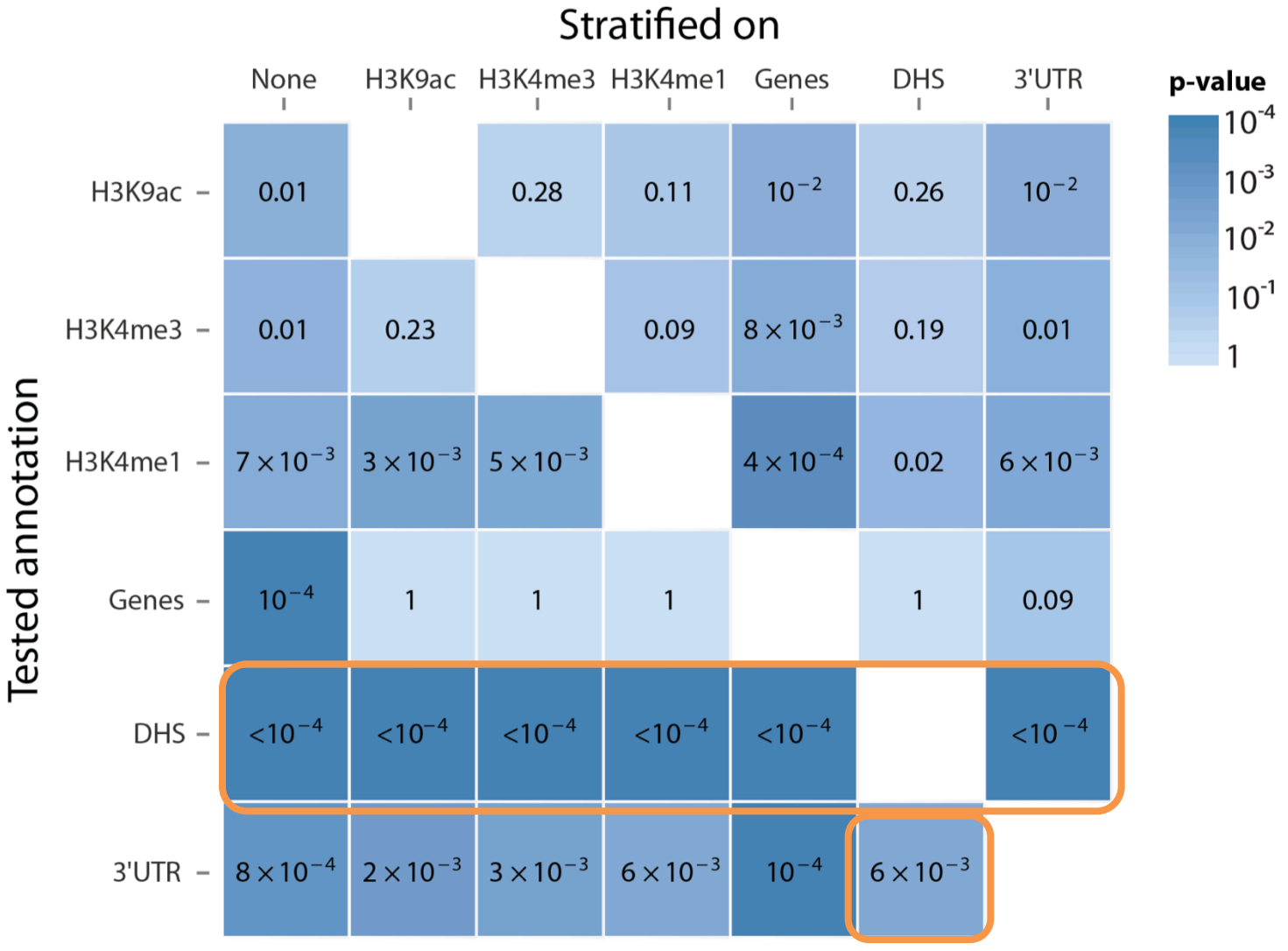




wellcome trust sanger institute

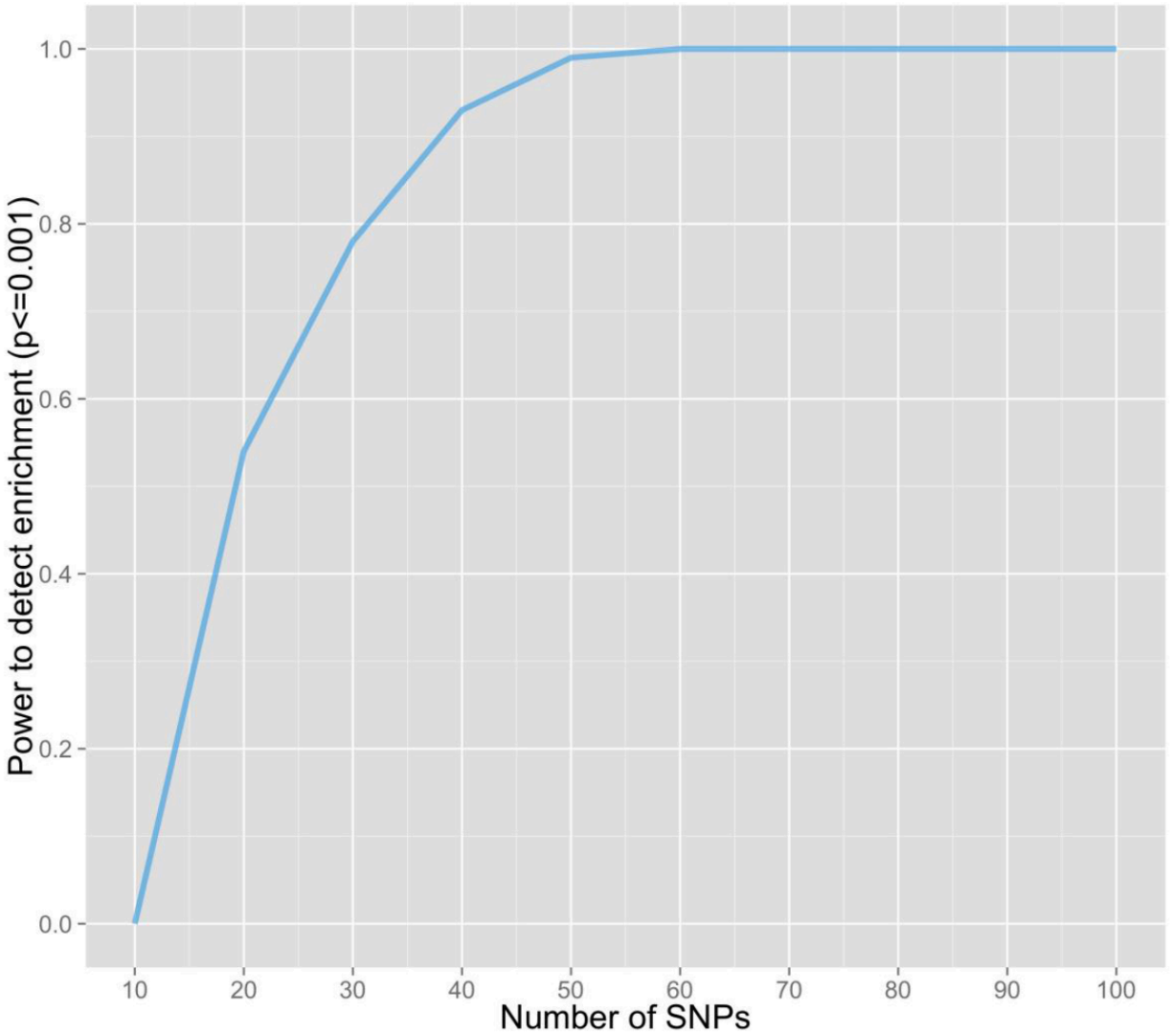
Trynka et al. *AJHG*, 2015

eQTLs are independently enriched for DHS and 3'UTR



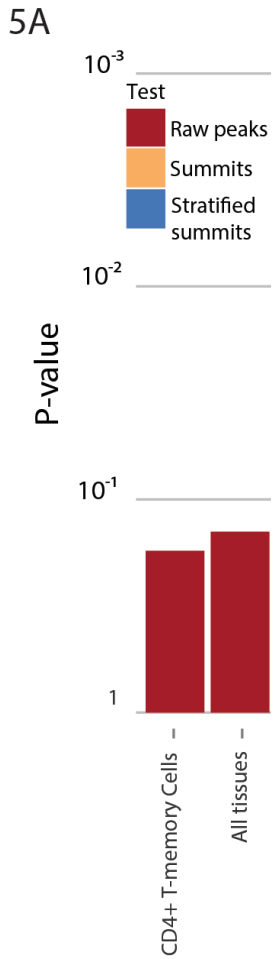
Trynka et al. *AJHG*, 2015

With 30 SNPs GoShifter has 80% power to detect significant enrichment



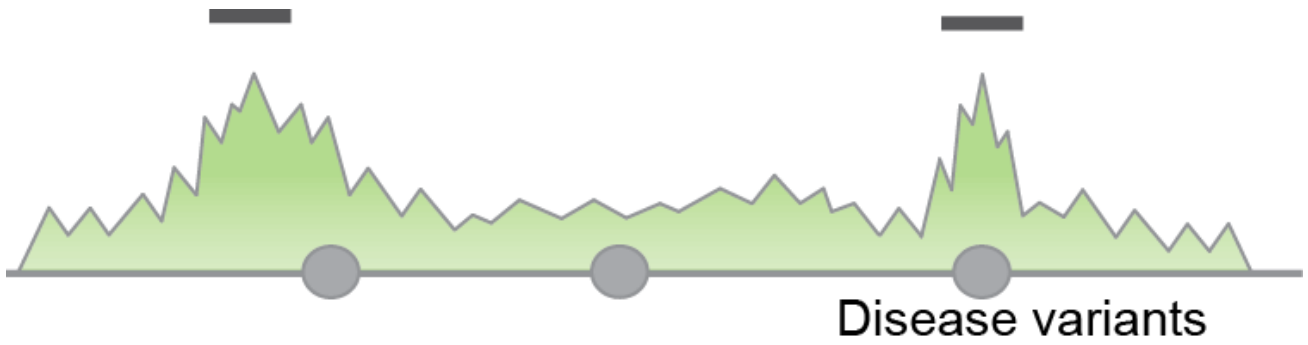
wellcome trust sanger institute

RA variants are not enriched in raw H3K4me3 peaks in CD4+ memory cells

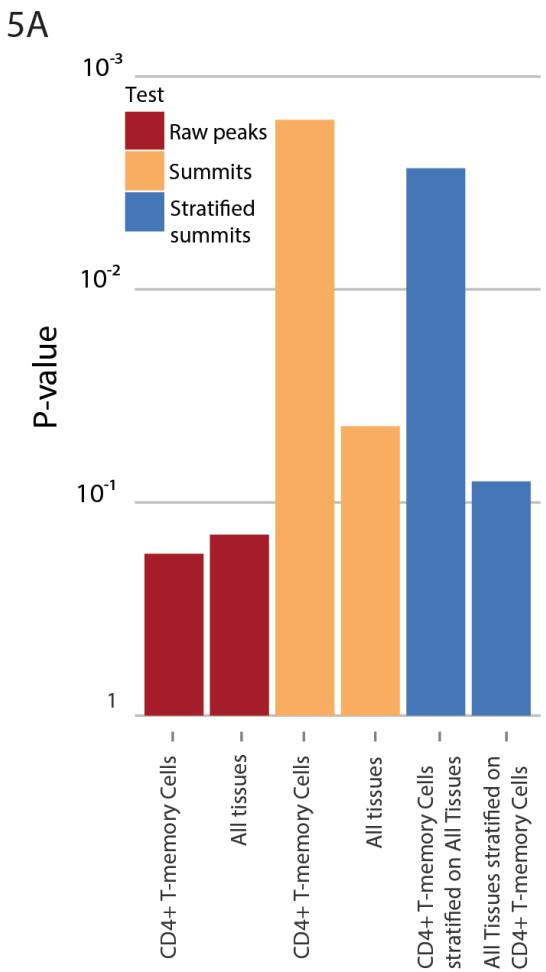


wellcome trust sanger institute

Restricting chromatin peaks to summit regions

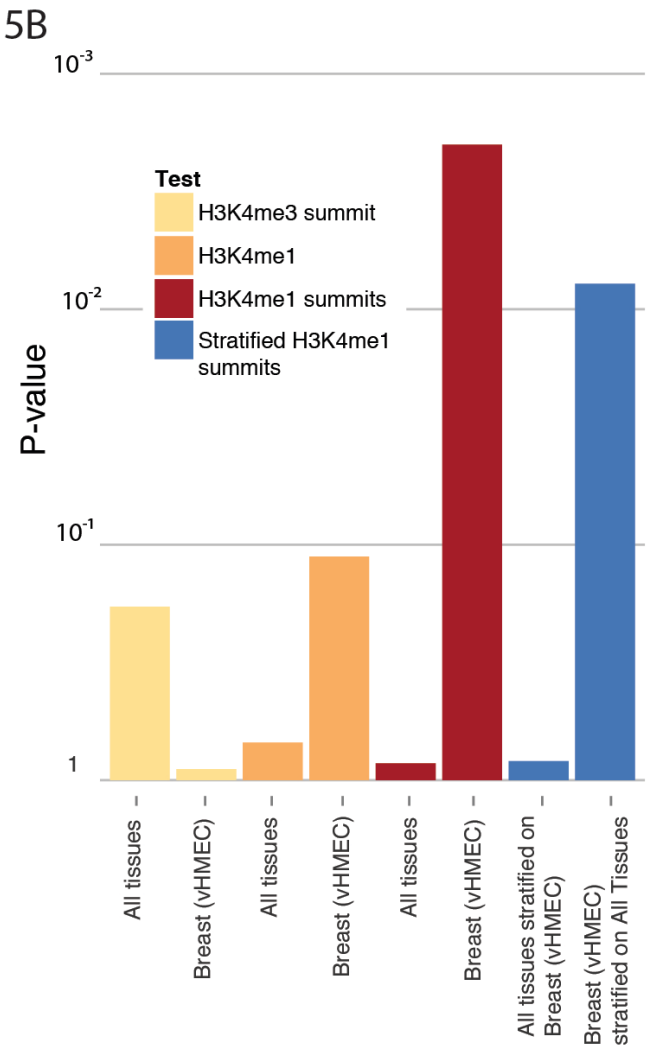


Summits of H3K4me3 peaks better capture cell type specific signatures

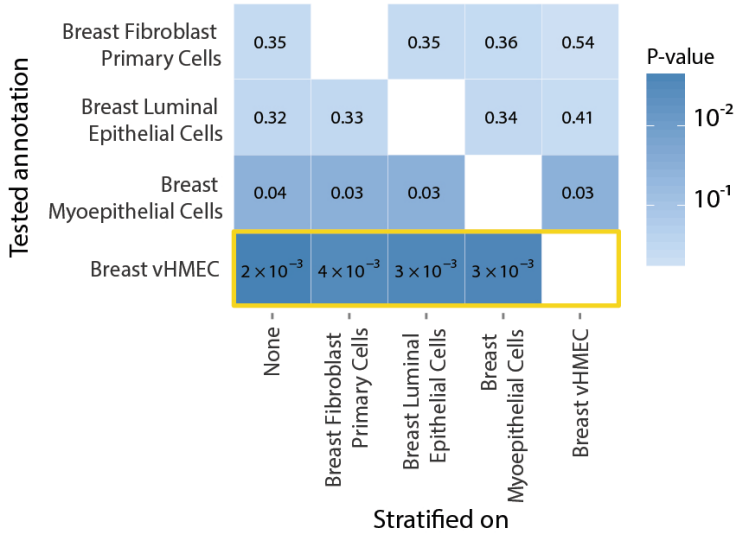


wellcome trust sanger institute

Disease enrichment at the summits of cell-type specific histone marks

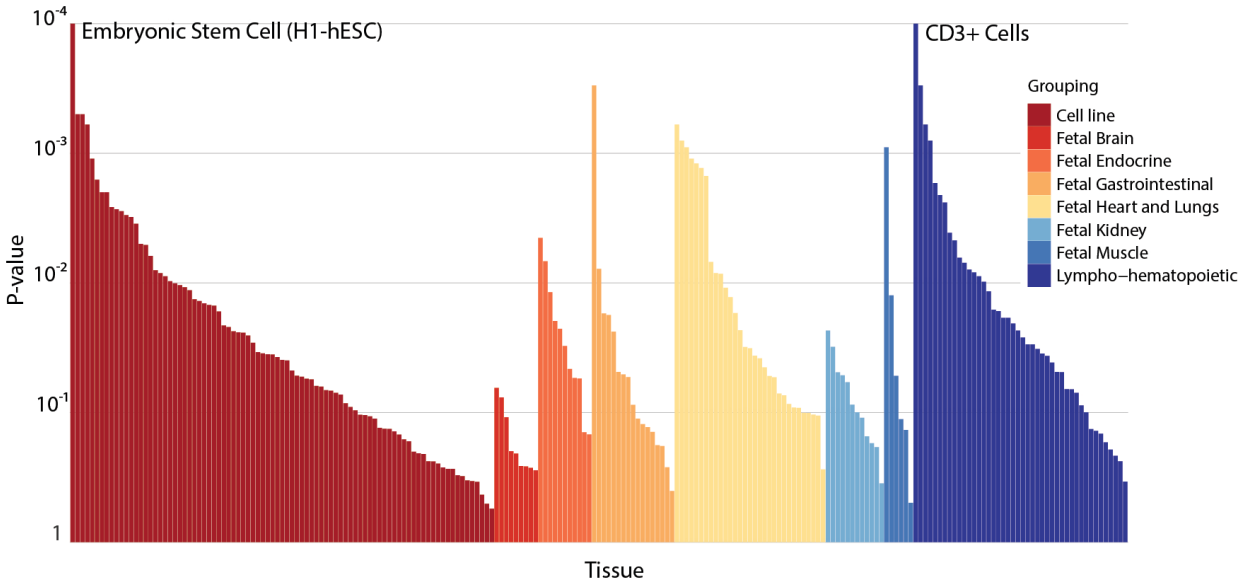


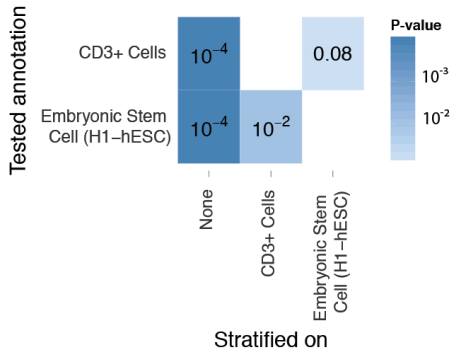
Breast cancer SNPs are enriched in the summits of H3K4me1 peaks of mammary epithelial cells (vHMEC)



Stratified analysis distinguishes relevant cell types for height

5C



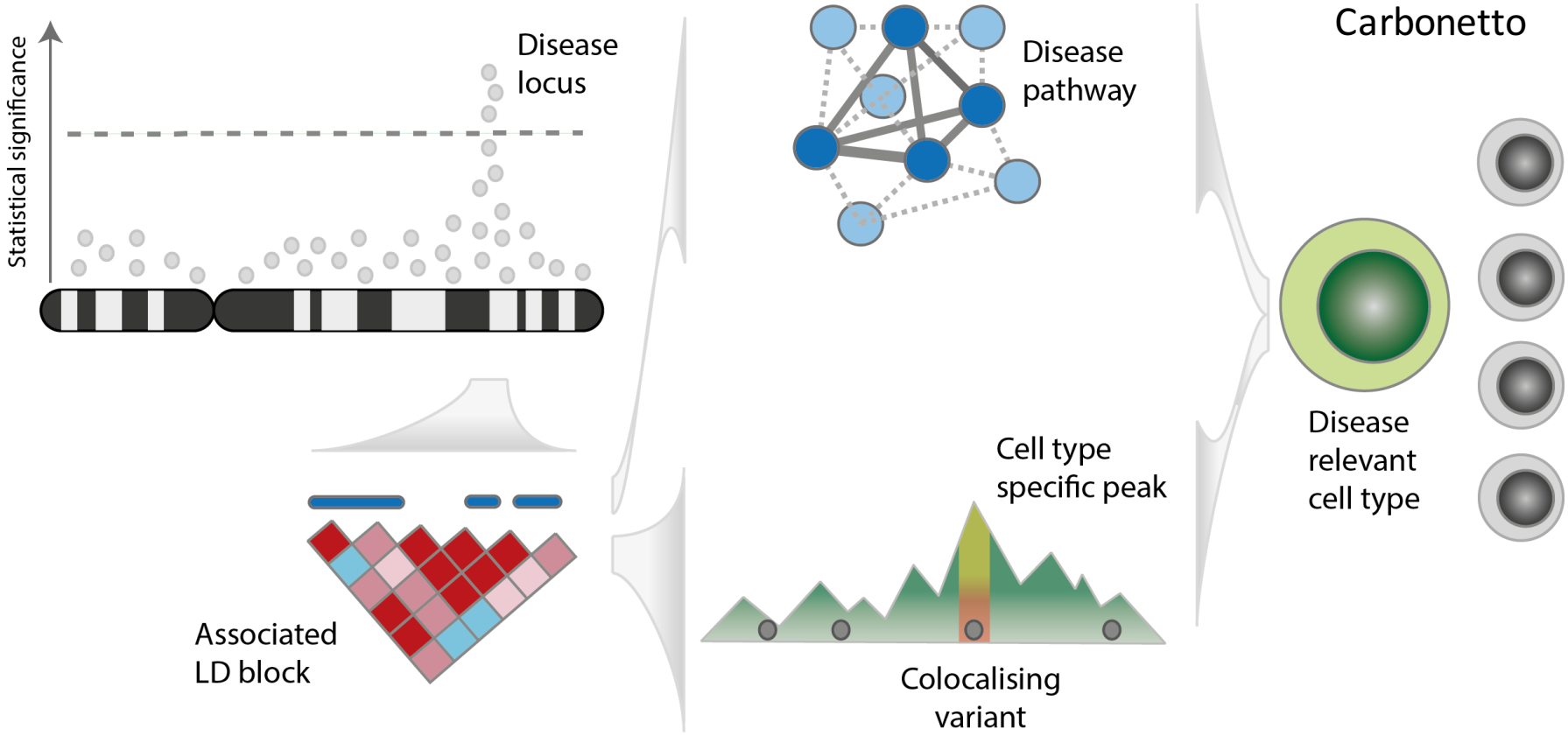


- Gene pathways represent cell type specific modules of gene expression

- If disease variants act through regulation of genes in specific pathways, stratifying enrichment within histone marks on disease pathway could improve identification of disease relevant cell types

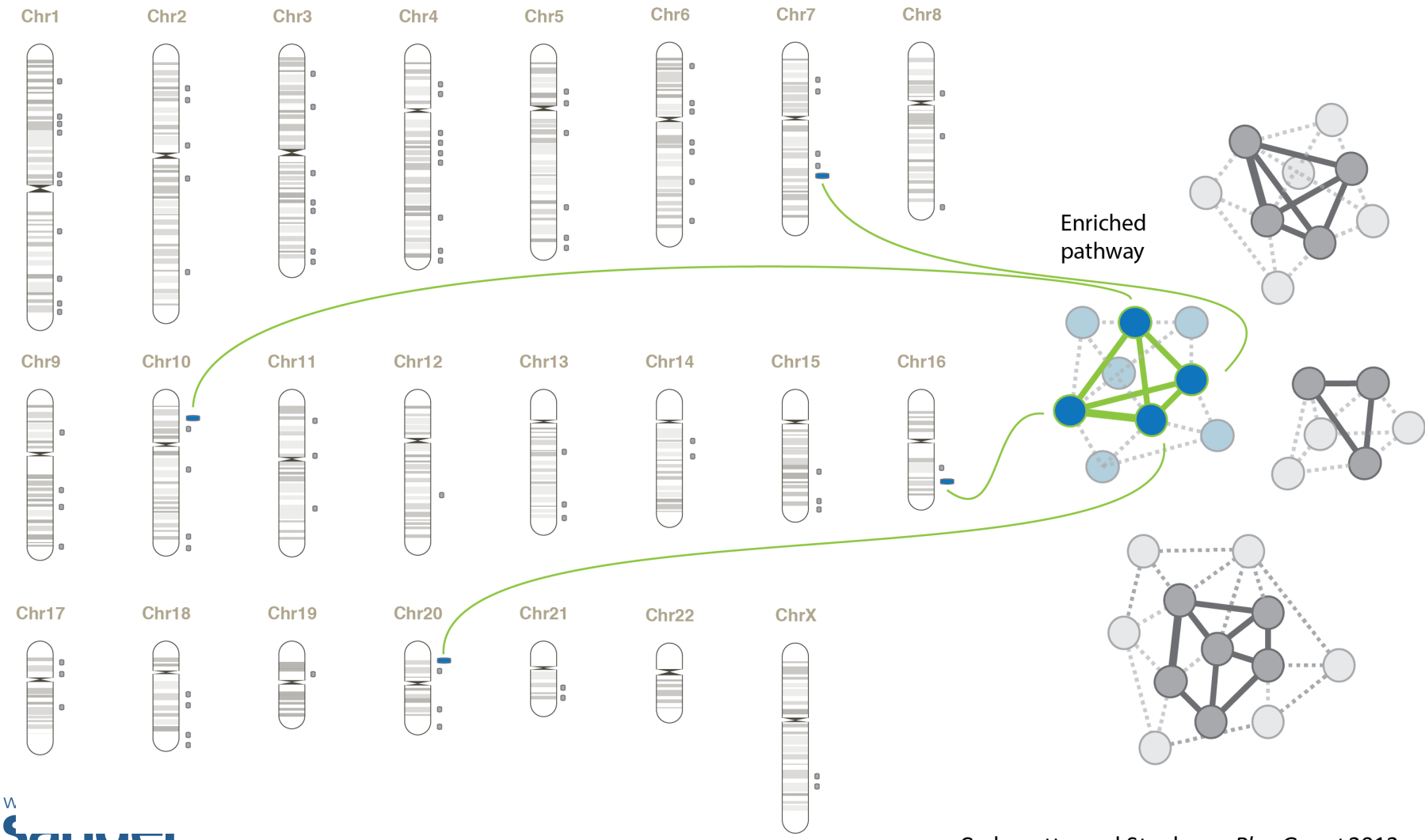
Linking disease variants to pathways and causal cell types

Peter Carbonetto



wellcome trust sanger institute

Assign SNPs to genes and test for association assuming pathway enrichment



Carbonetto and Stephens, *Plos Genet* 2013

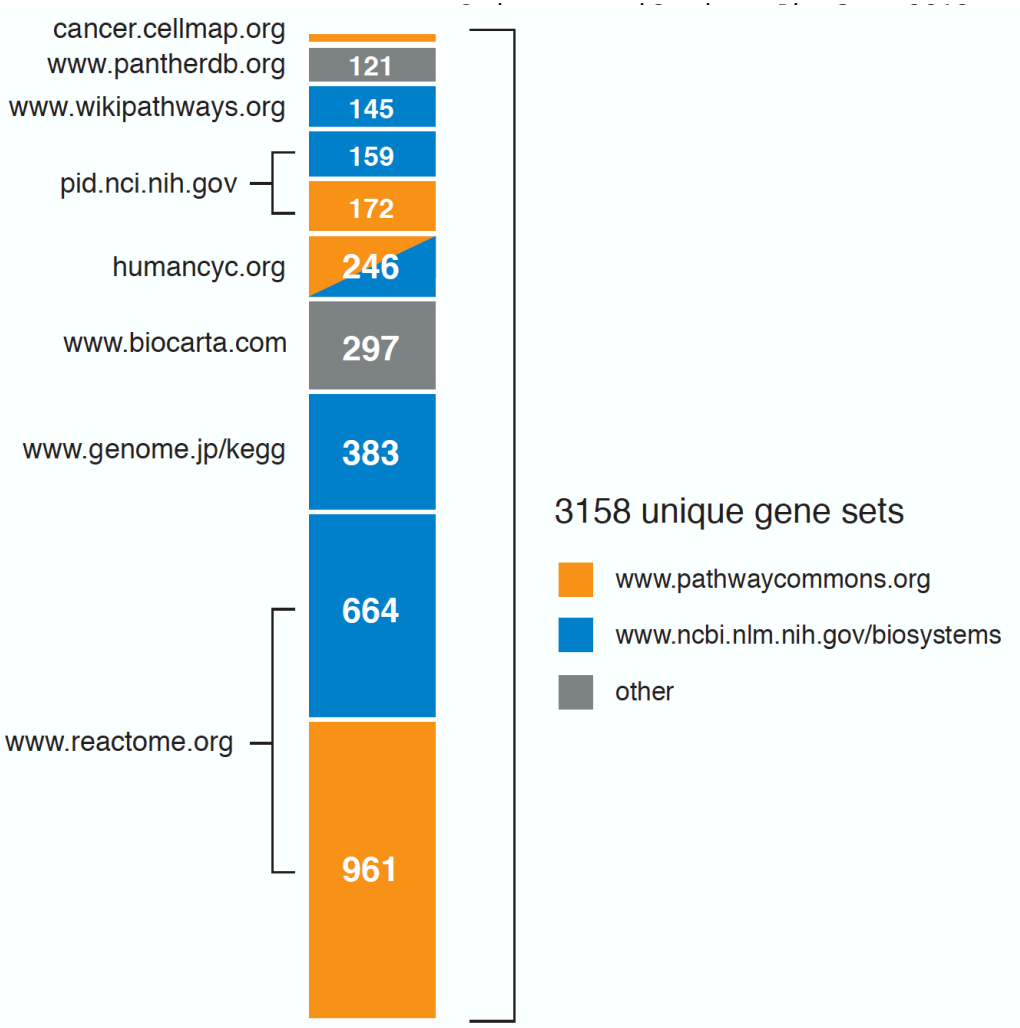
Dataset

3,149 coeliac disease cases

6,325 controls

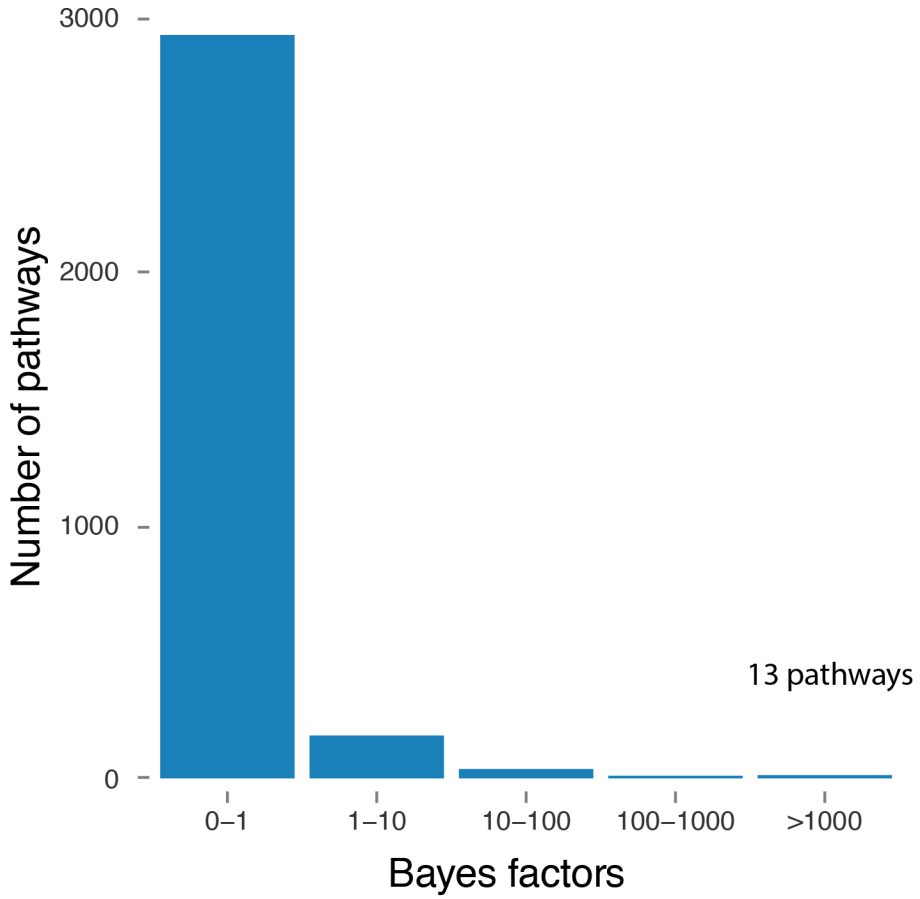
Dubois et al. Nat Genet 2010

Test for enrichment 3159 candidate pathways from 8 pathway databases



wellcome trust sanger institute

Most pathways show low evidence for enrichment



Enriched pathways
(Bayes factor > 10,000)

- Expression of chemokine receptors
- IL12 signaling
- Cytokine-cytokine receptor
- E-cadherin adherens junction
- Th1/Th2 differentiation

wellcome trust sanger institute

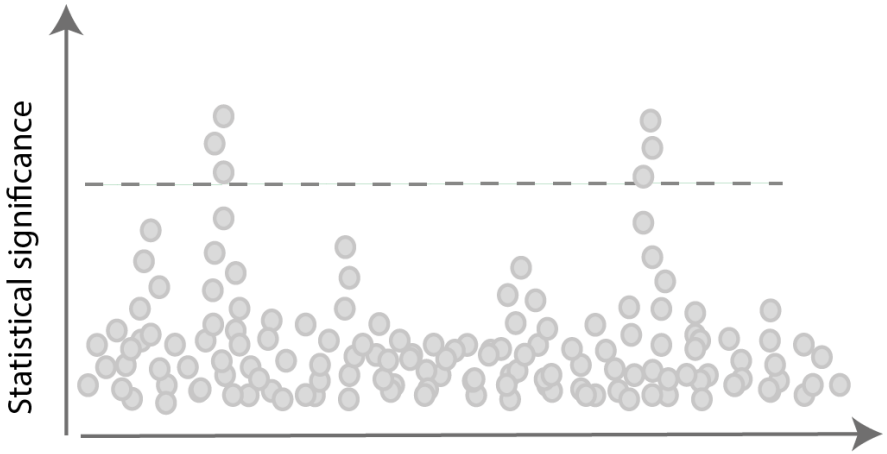
Ranking doesn't tell the full story

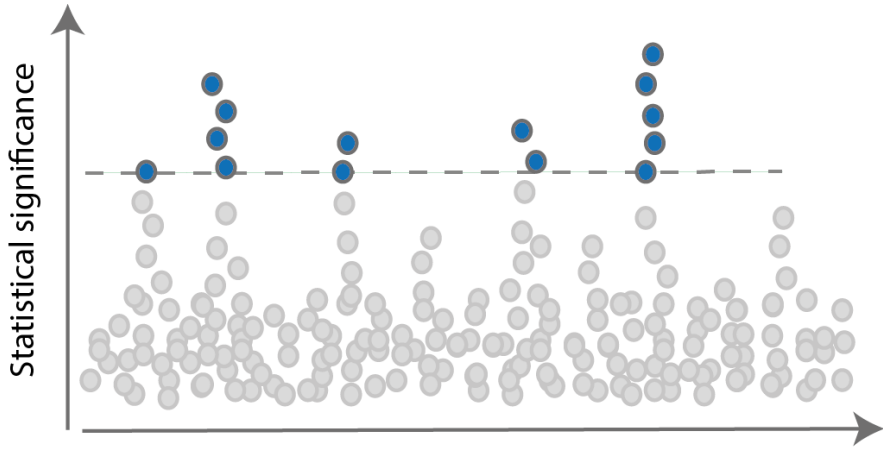
enriched pathway	database	source	genes	SNPs	Bayes factor
Expr of chemokine receptors	**BioCarta**	**BioCarta**	**29**	**887**	**6.6e7**
IL-12 signaling	PID	BioSystems	62	2,087	1.8e6
Cytokine-cytokine receptor	KEGG	BioSystems	265	8,190	1.8e5
IL-12 signaling	PID	PC	113	4,915	1.2e5
E-cadherin adherens junction	**PID**	**BioSystems**	**40**	**2,114**	**4.2e4**
Th1/Th2 differentiation	BioCarta	BioCarta	19	771	1.7e4

Bayes factor	enriched pathway(s)		
8.5e11	**Expr of chemokine**	**+**	**E-cadherin adherens junction**
8.7e10	IL-12 signaling	+	E-cadherin adherens junction
6.8e10	Cytokine-cytokine	+	E-cadherin adherens junction
	...		

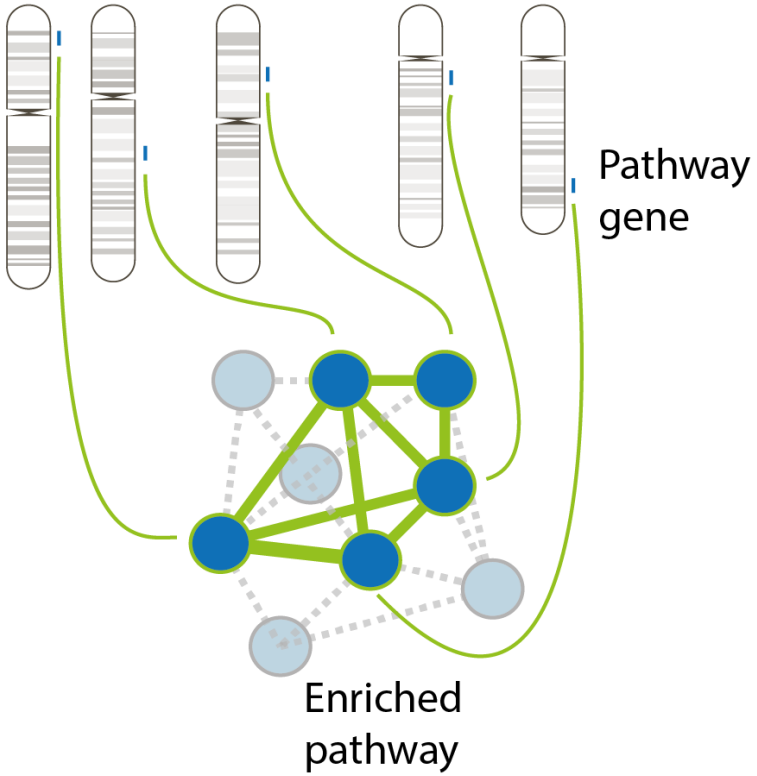
wellcome trust sanger institute

Prioritize variants within the enriched pathways

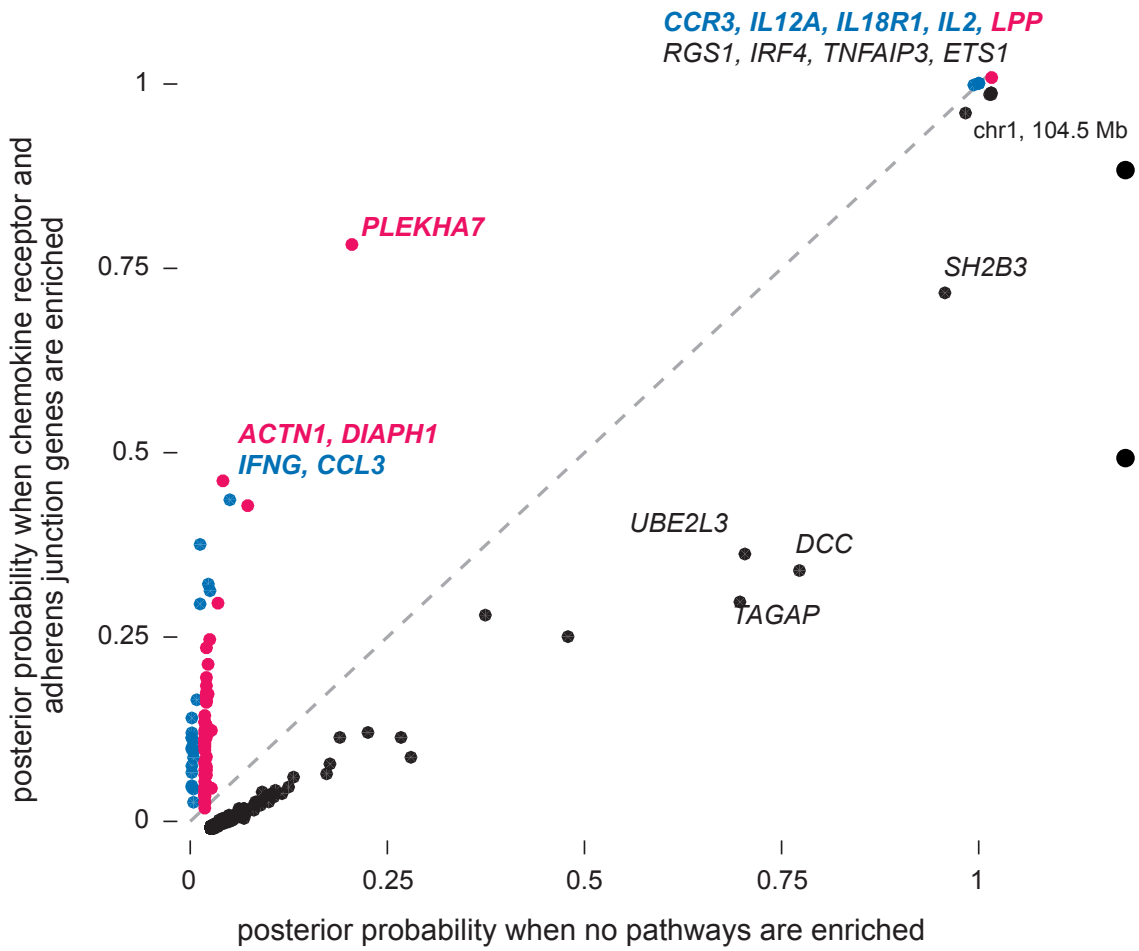




Promote new associations assigned to enriched pathways



Prioritisation of SNPs in enriched pathways increases support for risk factors at many loci



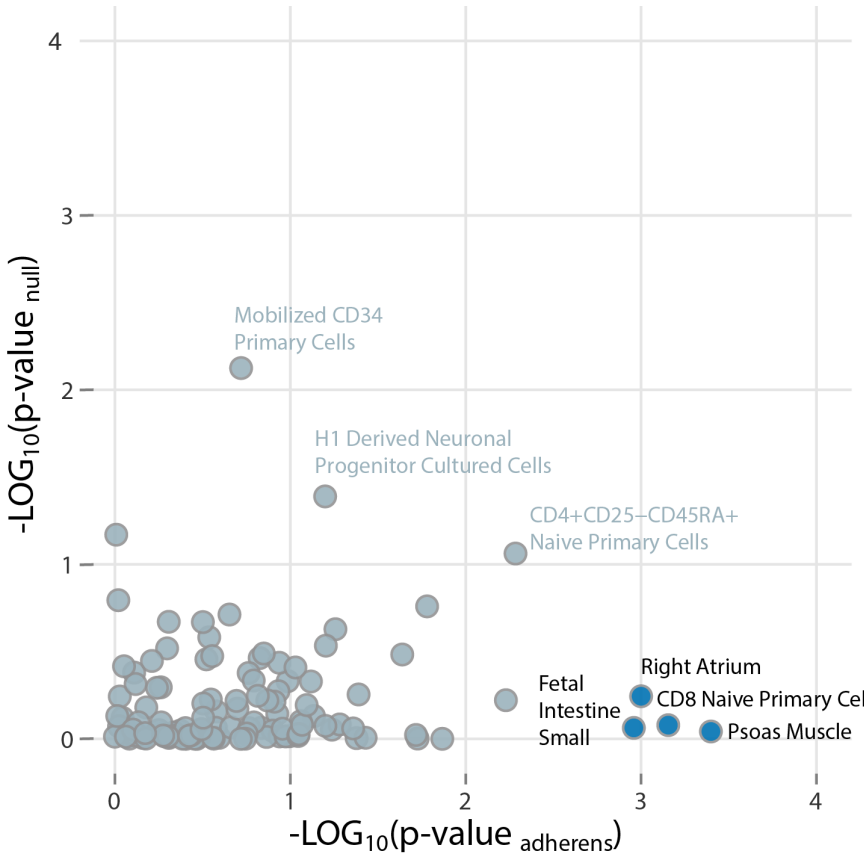
- 887 SNPs map to **chemokine receptor**
- 2,113 SNPs in **adherens junction** pathway

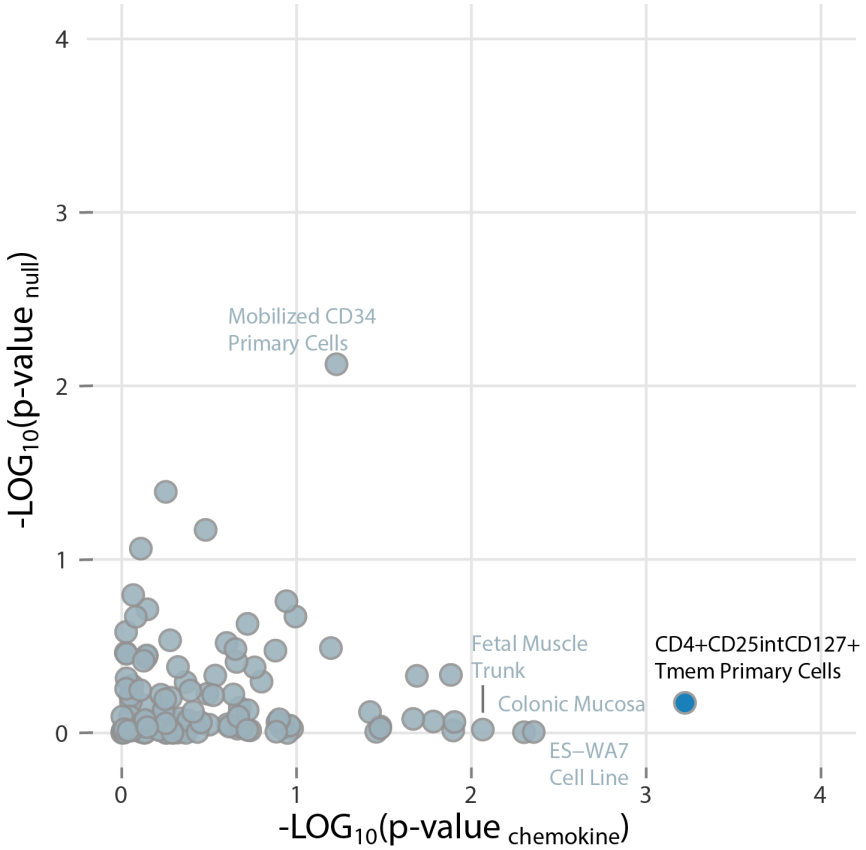
wellcome trust sanger institute

Linking pathways to effector cell types

- SNPs with posterior probability > 0.01
  - 77 SNPs assigned to chemokine receptor pathway
  - 114 SNPs assigned to adherens junction pathway
  - 158 SNPs under null of no pathway enrichment
- Total of 118 cell types assayed for H3K4me3

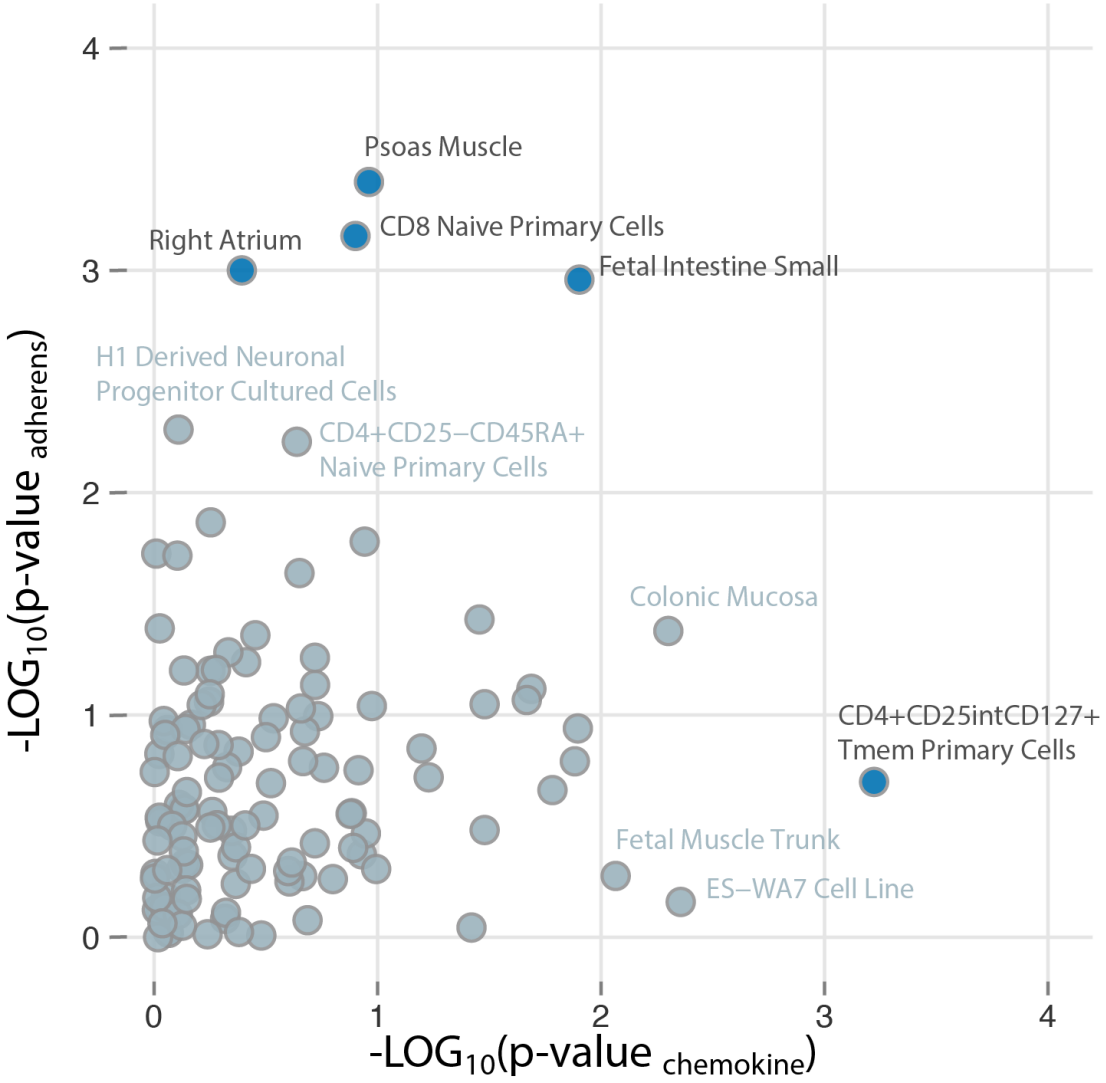
Pathway prioritise variants point to functional effects in specific cell types





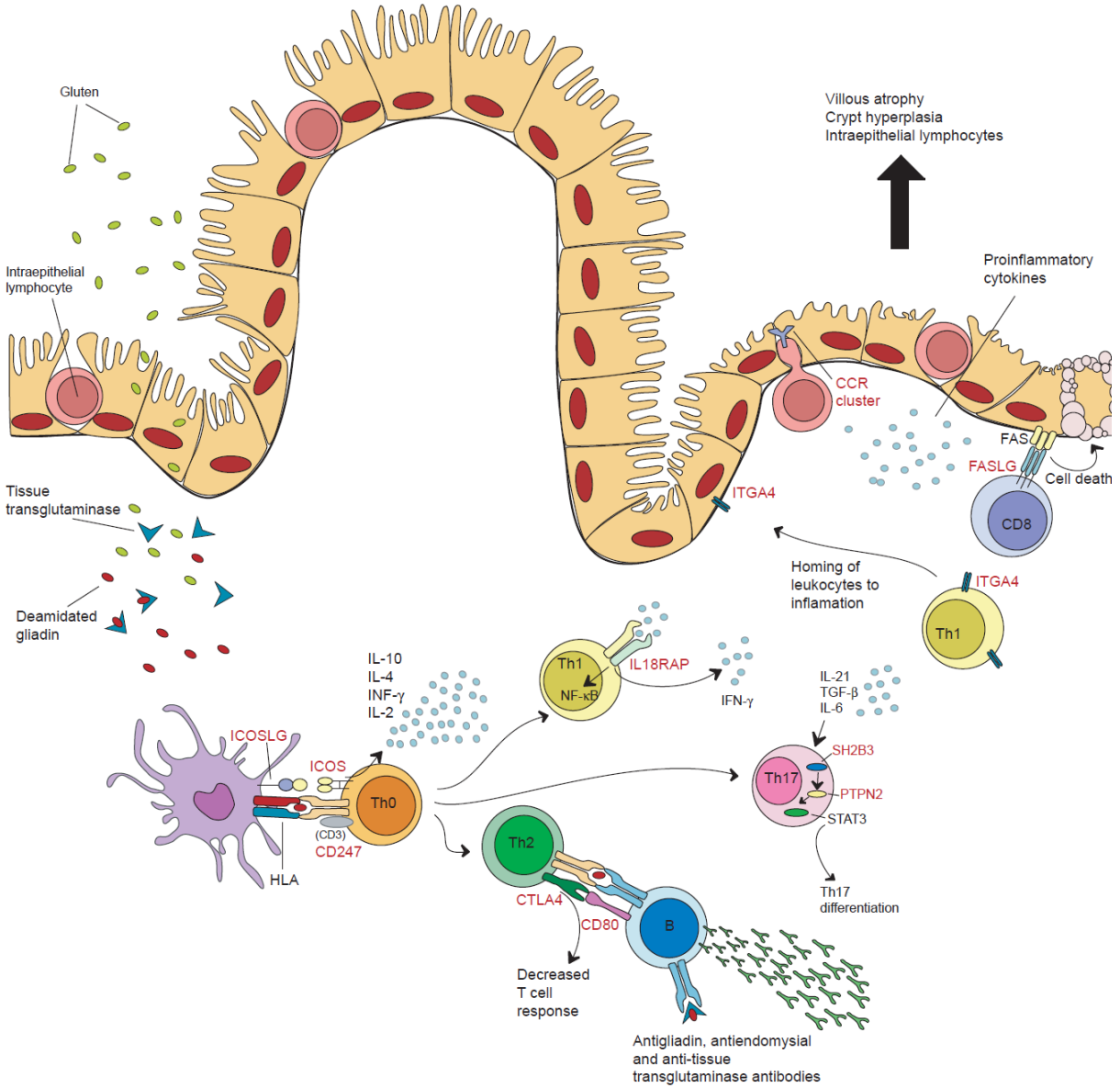
wellcome trust sanger institute

Distinct pathways point to activity in different cell types



wellcome trust sanger institute

Adherens junction could point towards impaired epithelial gut barrier

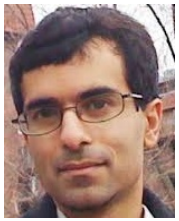


wellcome trust sanger institute

Cisca Wijmenga
Groningen University, Netherlands

Peter Carbonetto
AncestryDNA, San Francisco

Soumya Raychaudhuri
Brigham and Women's Hospital
Broad Institute



Matthew Stephens
University of Chicago

wellcome trust sanger institute