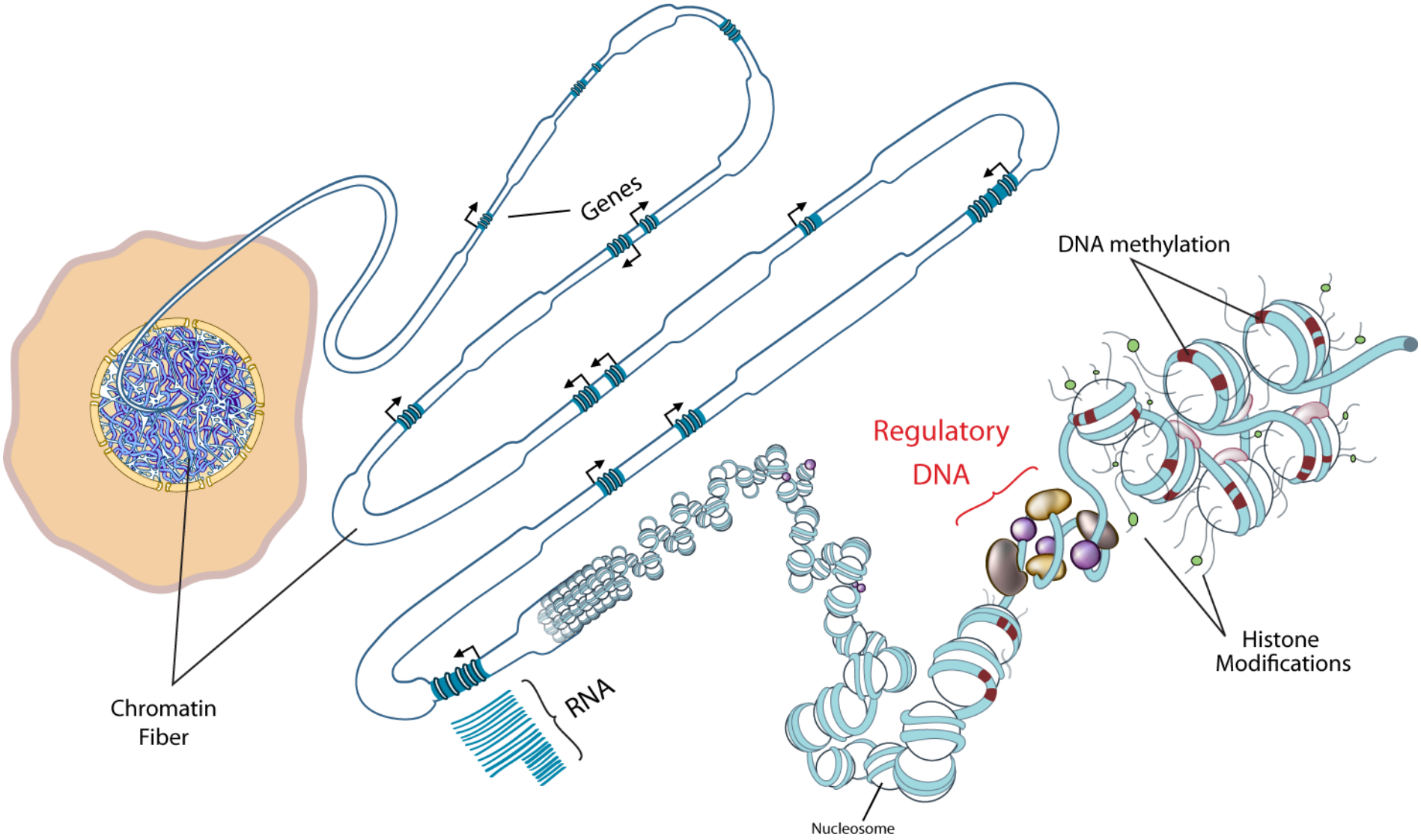
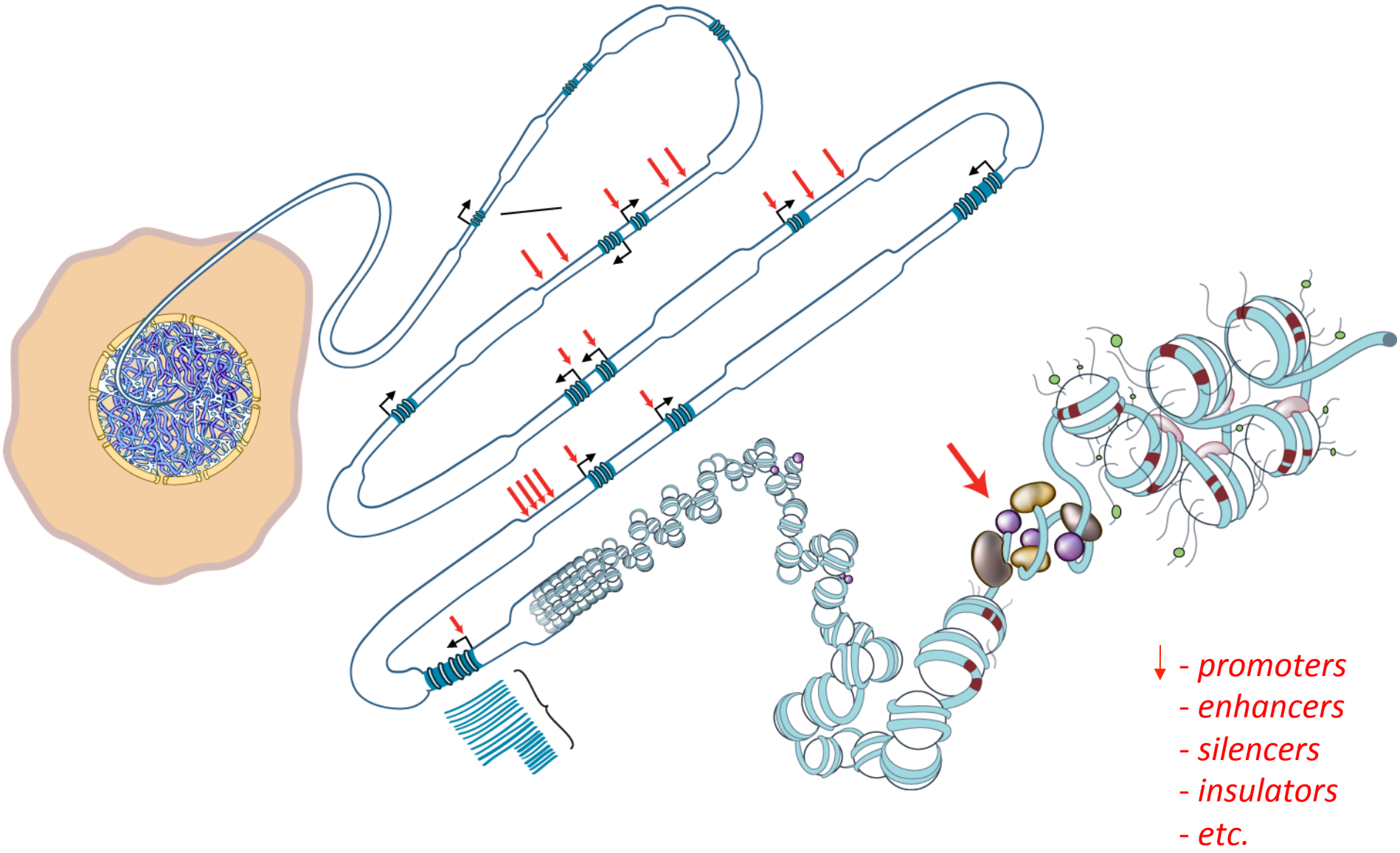
Using ENCODE data to illuminate disease-and trait-associated variation

John A. Stamatoyannopoulos, M.D.
Depts. of Genome Sciences & Medicine
University of Washington

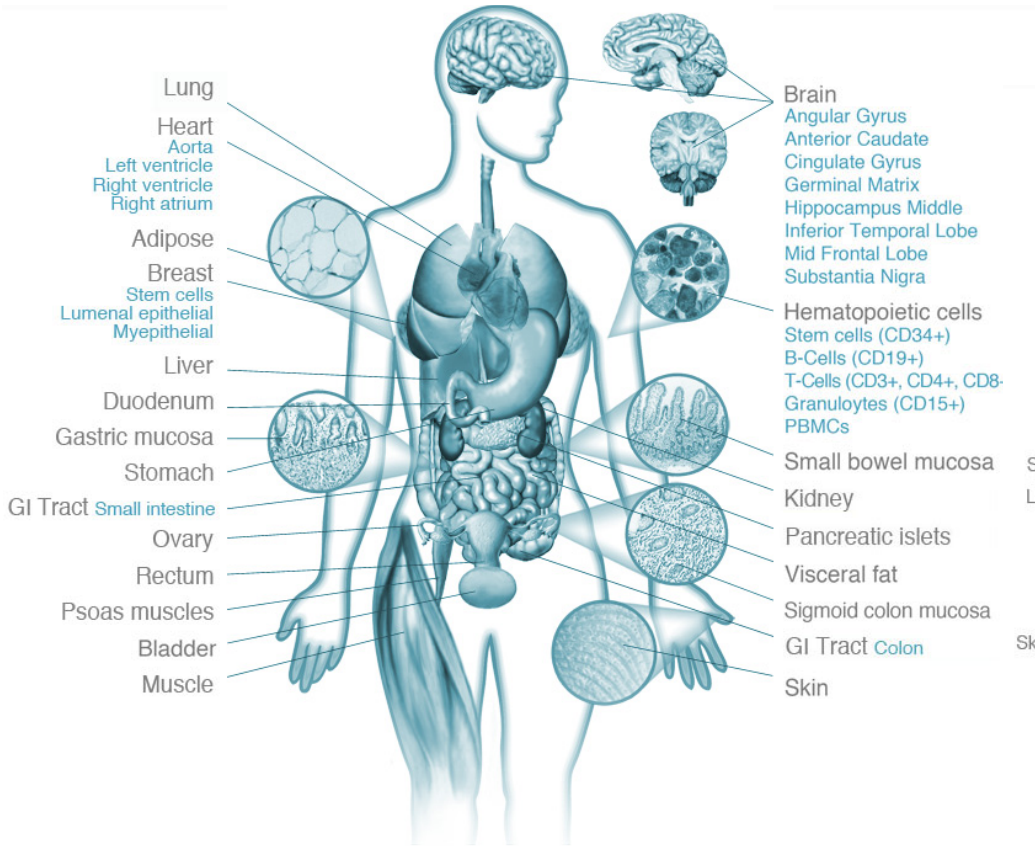
Genes, regulatory DNA, and epigenetic features

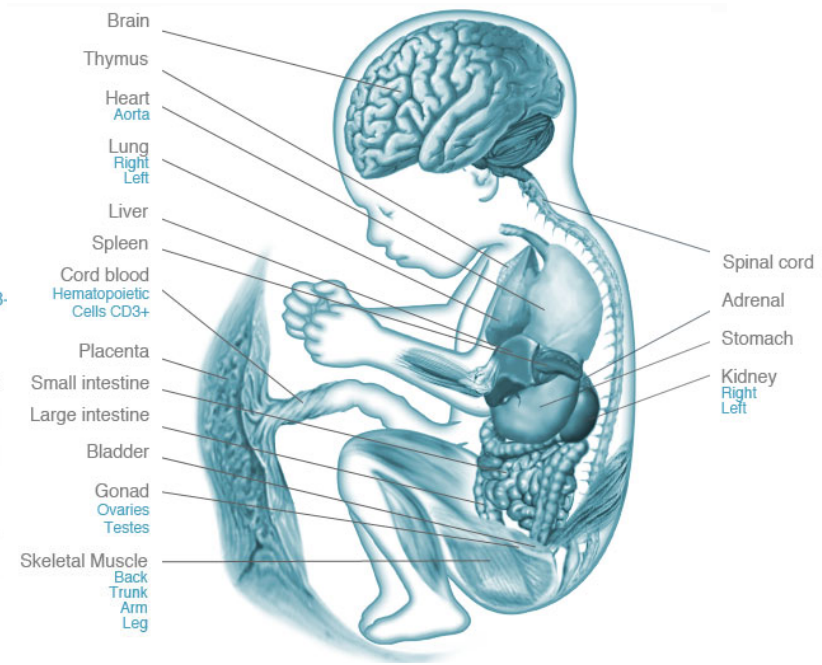


Genes, regulatory DNA, and ‘epigenetic’ features



Creating deep maps of human regulatory DNA





www.encodeproject.org

www.roadmapepigenomics.org

www.encode-roadmap.org

ENCODE XXX + ROADMAP epigenomics PROJECT **COMBINED** DATA BROWSER

Using this page

Roadmap ENCODE Combined Selected

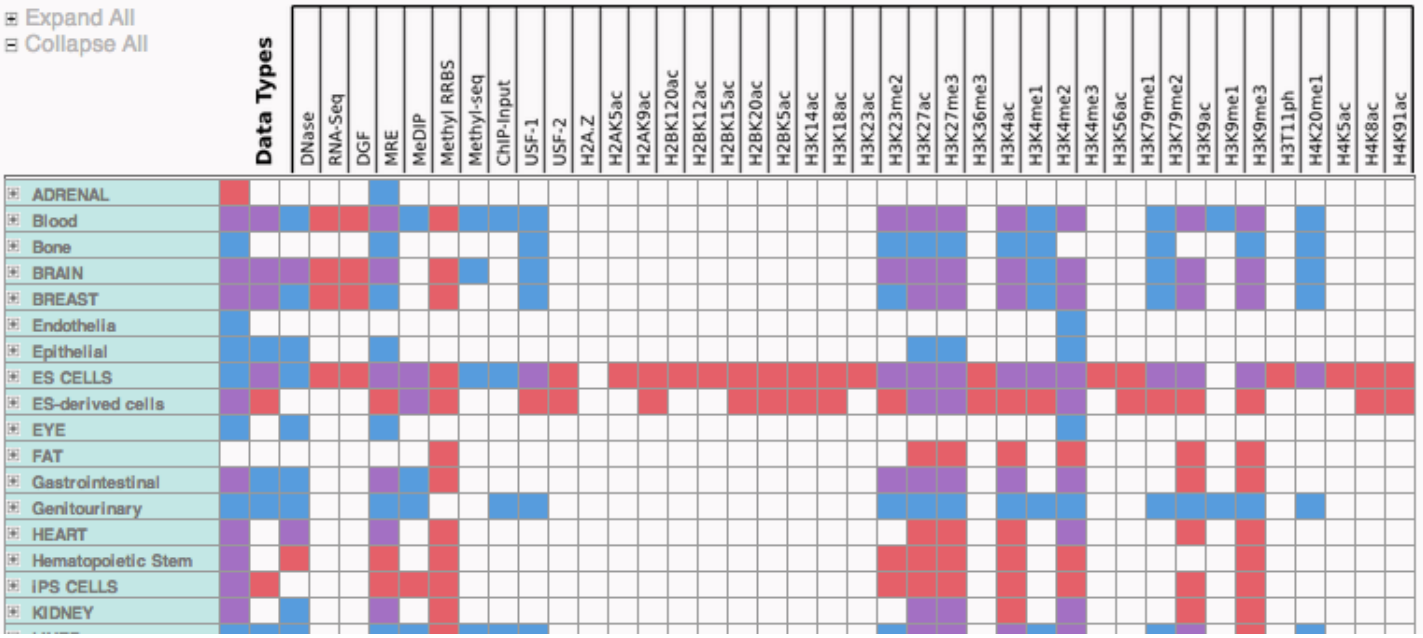
The tables below show which cell types have browser tracks for different data types in the combine Encode Roadmap browser. By default all, cell type rows are collapse into functional groups. Clicking on the group name will expand out all the cell types. The legend above shows what the colors of different squares mean.

Data can be viewed by clicking on data types at the top of the table, the squares for cell type groups and data types, the names of cell types in an expanded cell type group, or individual cell type and data type squares.

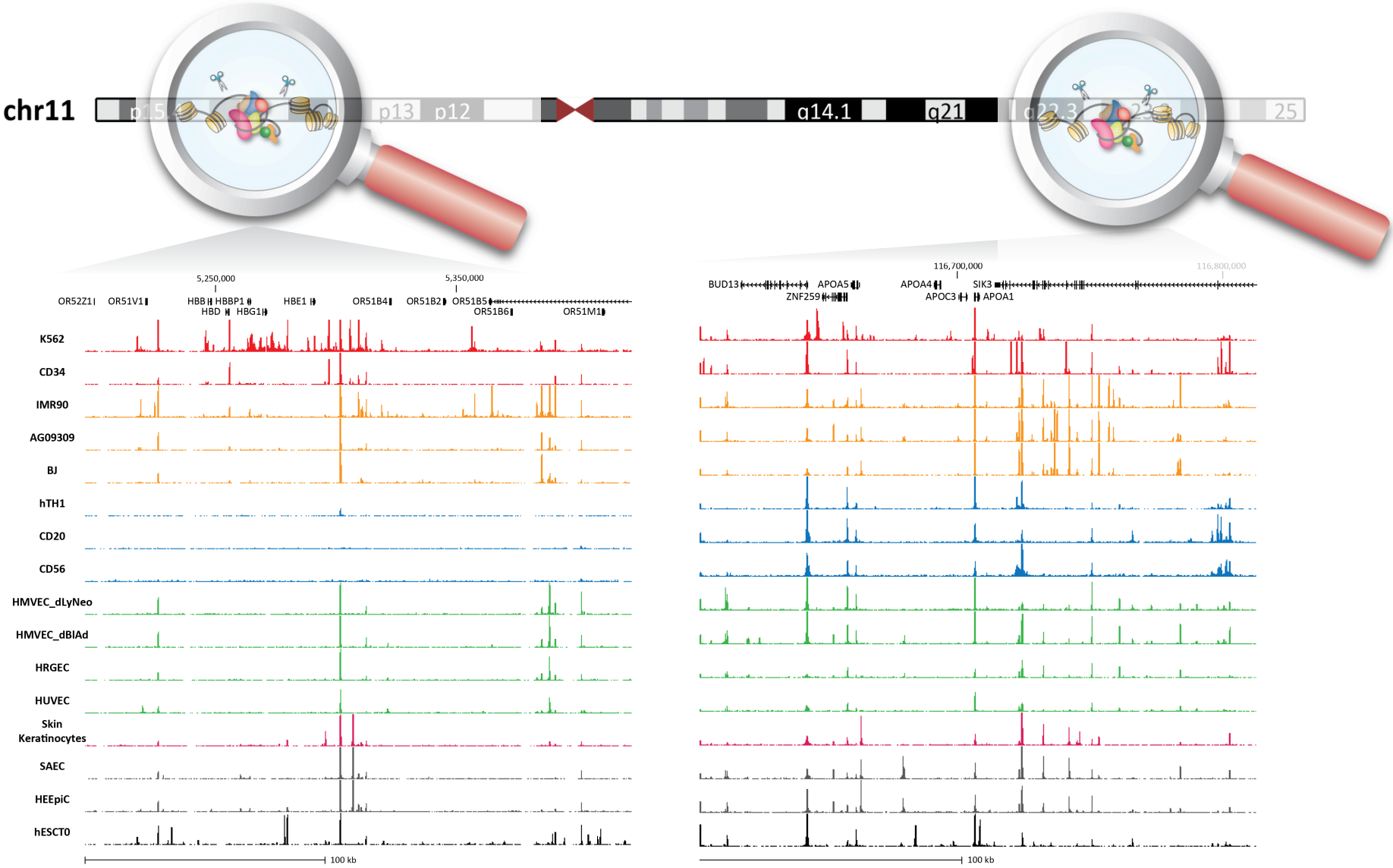
After selecting data, go to the bottom of the page to review what is selected and then click on the "Go to Browser" button.

See Selected Data

Expand All
Collapse All

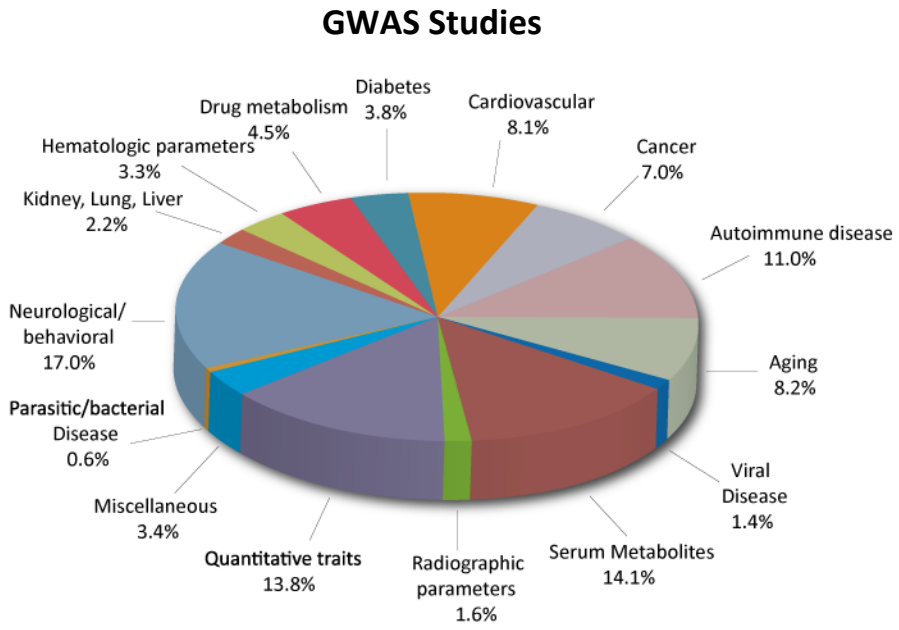


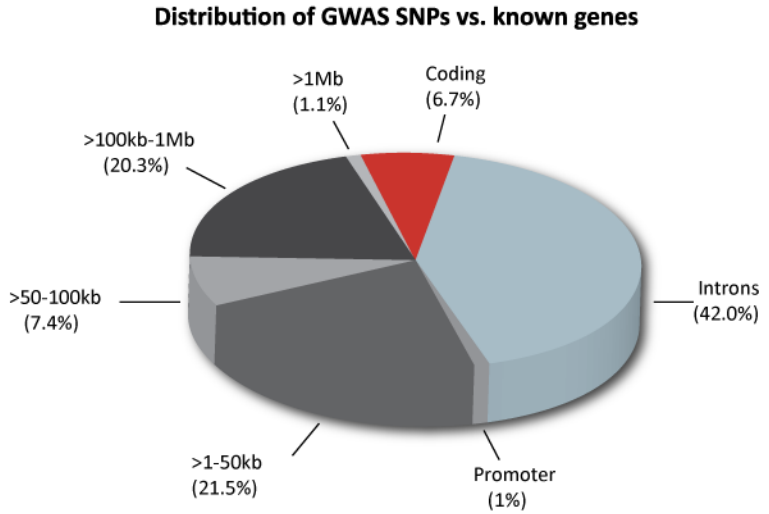
Closing in on a comprehensive atlas of human regulatory DN



Regulatory DNA variation associated with common diseases and traits

Identification of disease- and trait-associated variation by GWAS





GWAS disease/trait associated variants

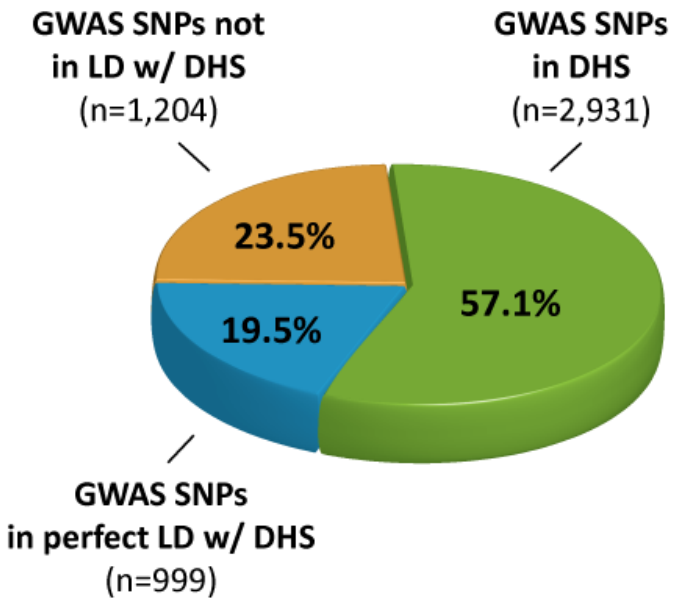
x

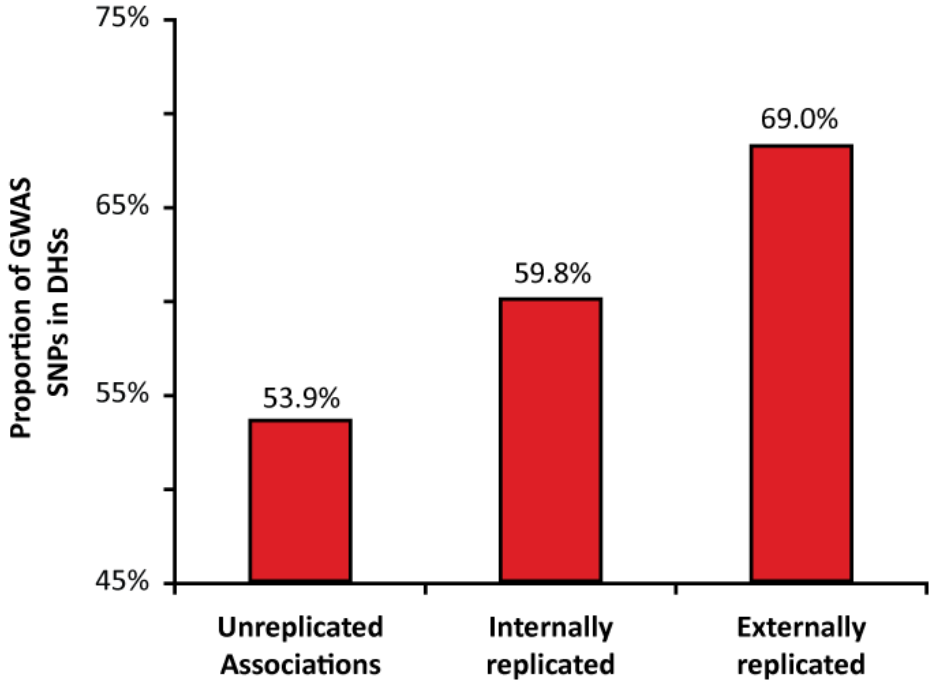
Maps of regulatory DNA in >300 diverse cell and tissue types

Maurano et al., Science 2012

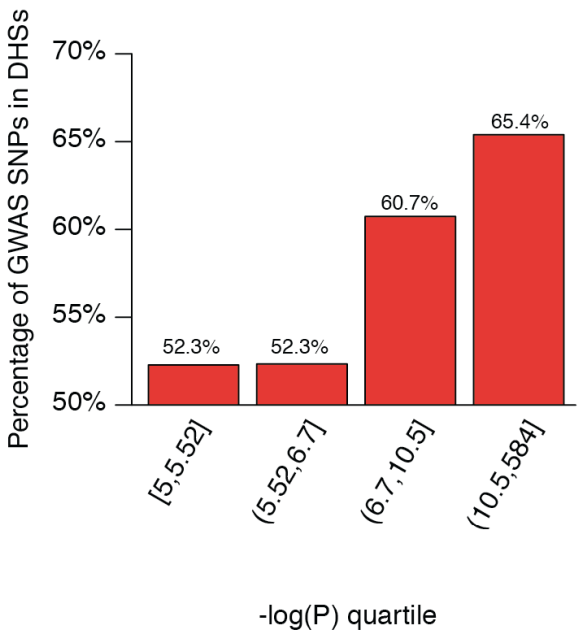
#1
Disease-associated variation is concentrated in regulatory DNA

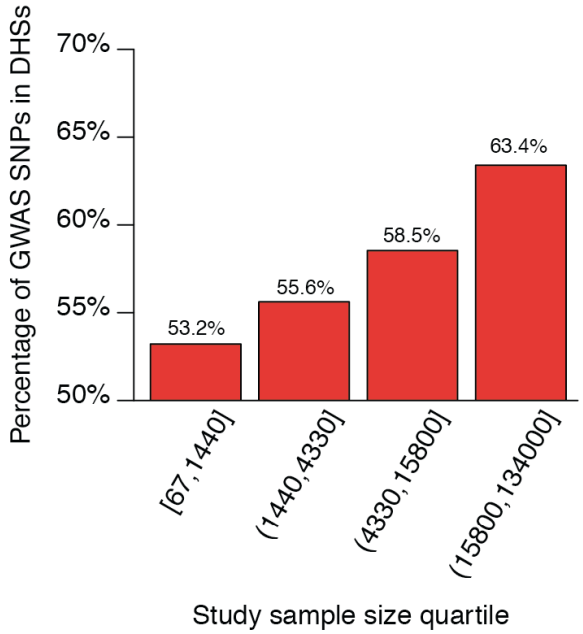
Disease- and trait-associated SNPs are concentrated in regulatory DNA





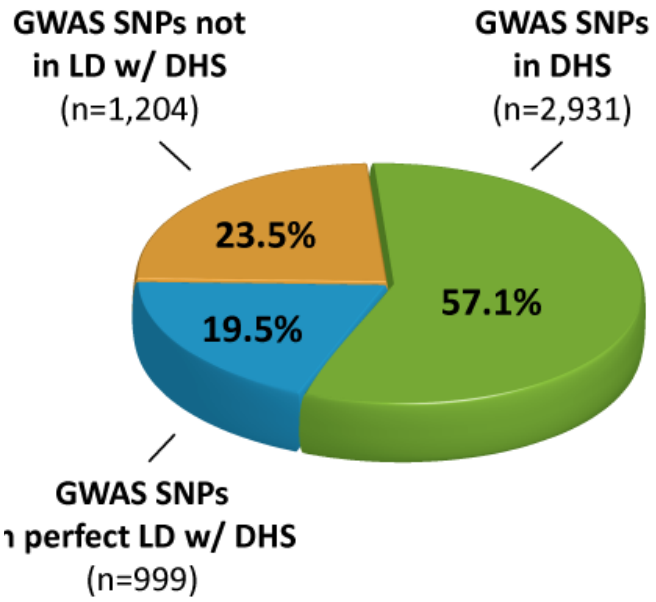
The effect increases monotonically with other measures of higher quality associations

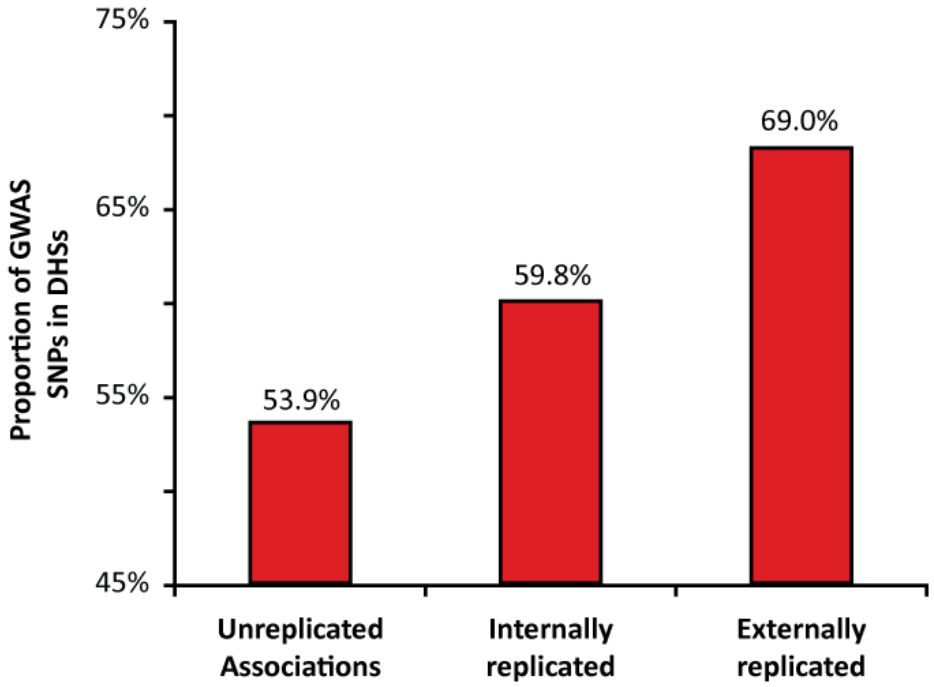




Maurano et al., Science 2012

Disease- and trait-associated SNPs are concentrated in regulatory DNA



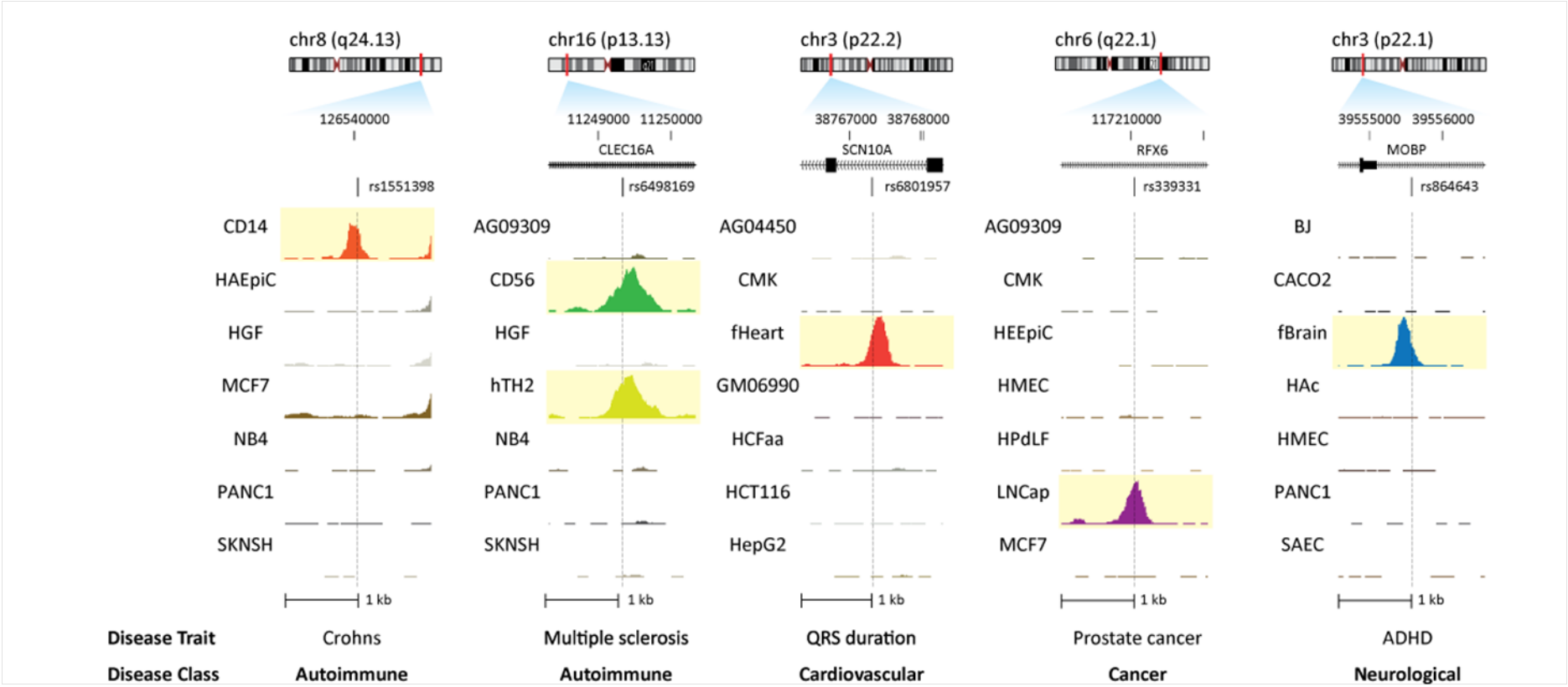


~1.8-fold for all replicated variants in all disorders

>10-fold for specific disese-cell type pairings

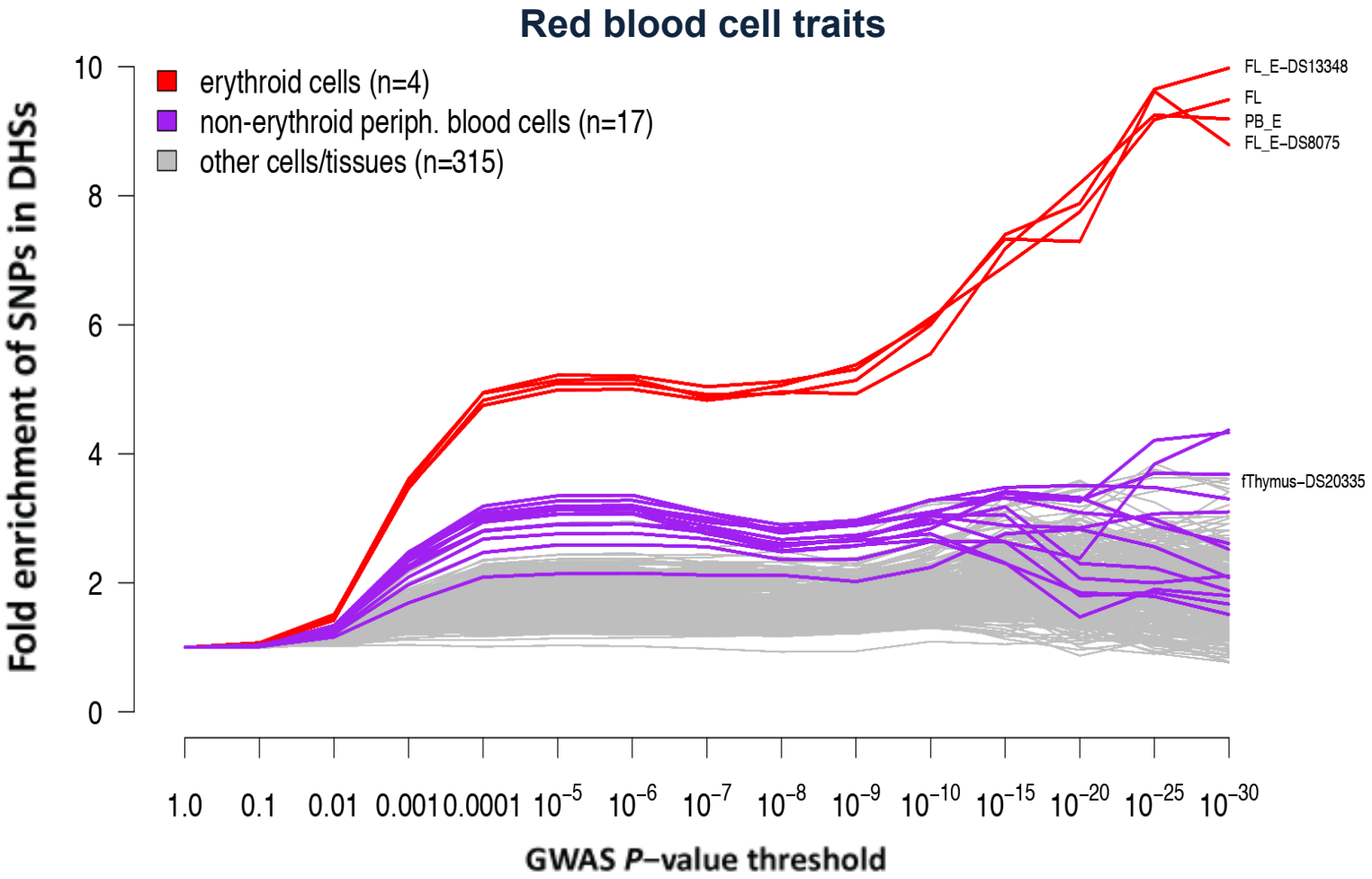
#2
GWAS variants selectively localize in pathologically relevant cell types

Disease-associated variation clusters in pathogenic or target cell types



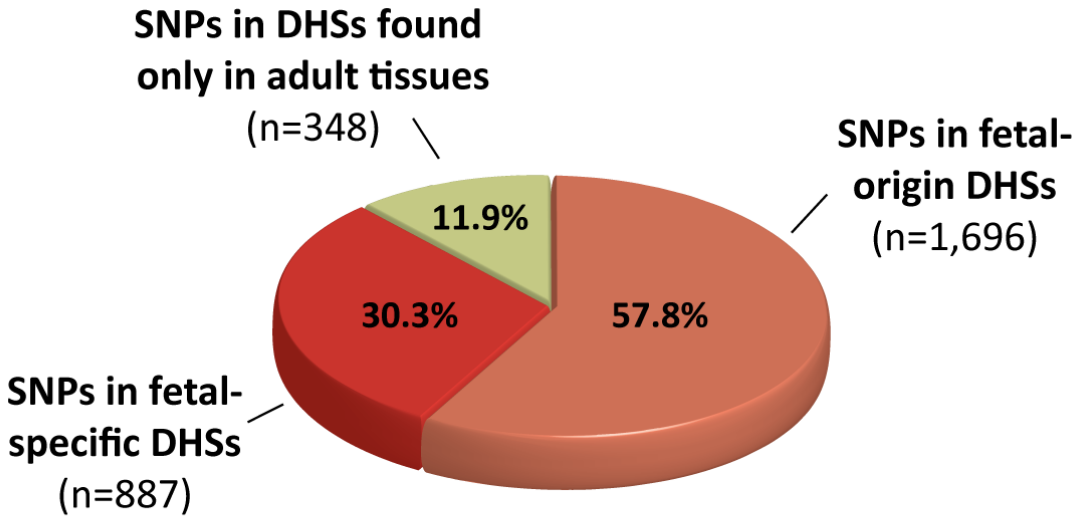
Maurano et al., Science 2012

Cell-selective enrichment of trait-associated variants



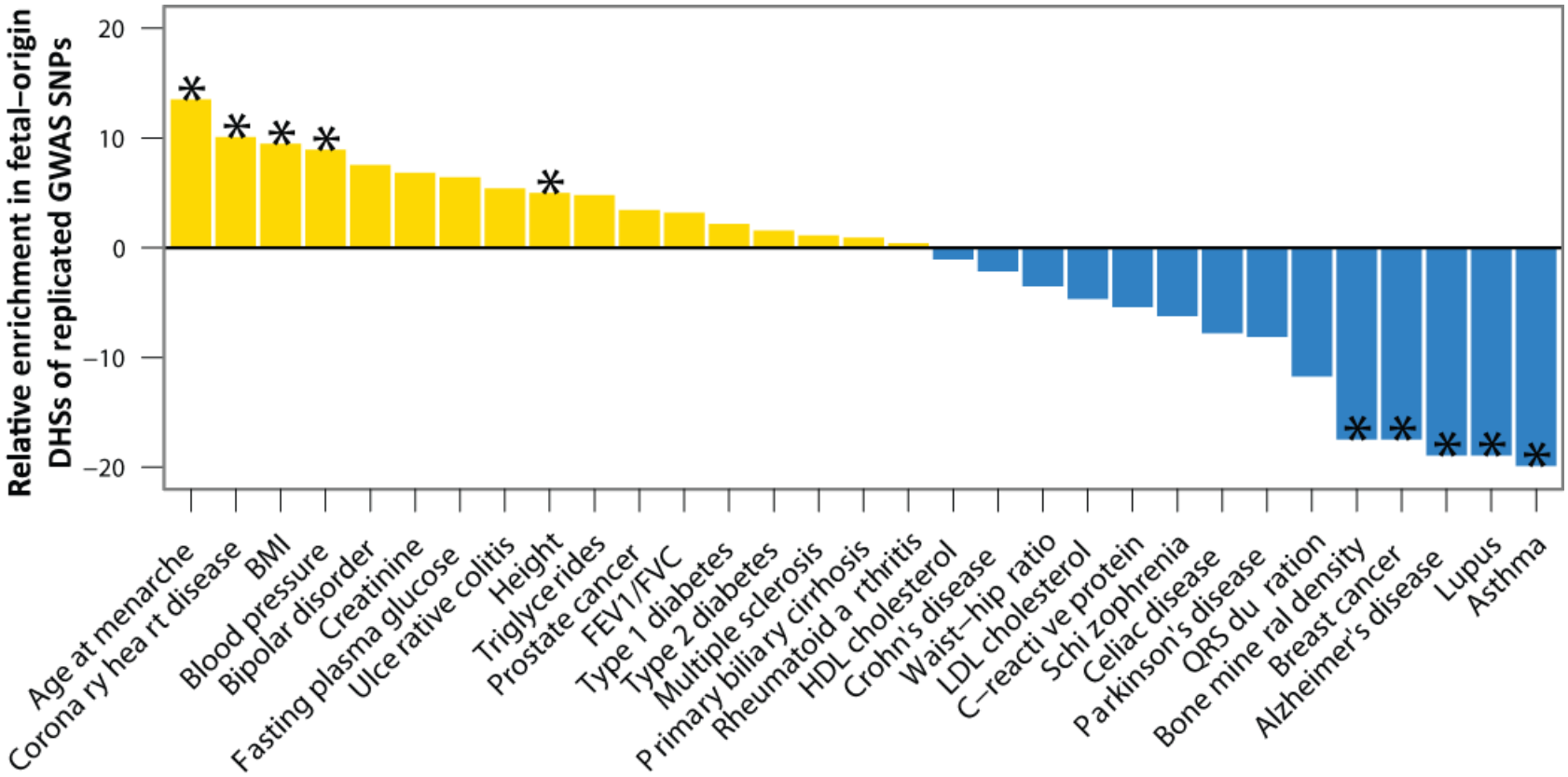
#3
Diseases and traits with developmental contributions preferentially localize in fetal regulatory DNA

Most variants lie in regulatory DNA of fetal origin



Maurano et al., Science 2012

Fetal regulatory variants are enriched in traits & diseases with known links to intrauterine exposures

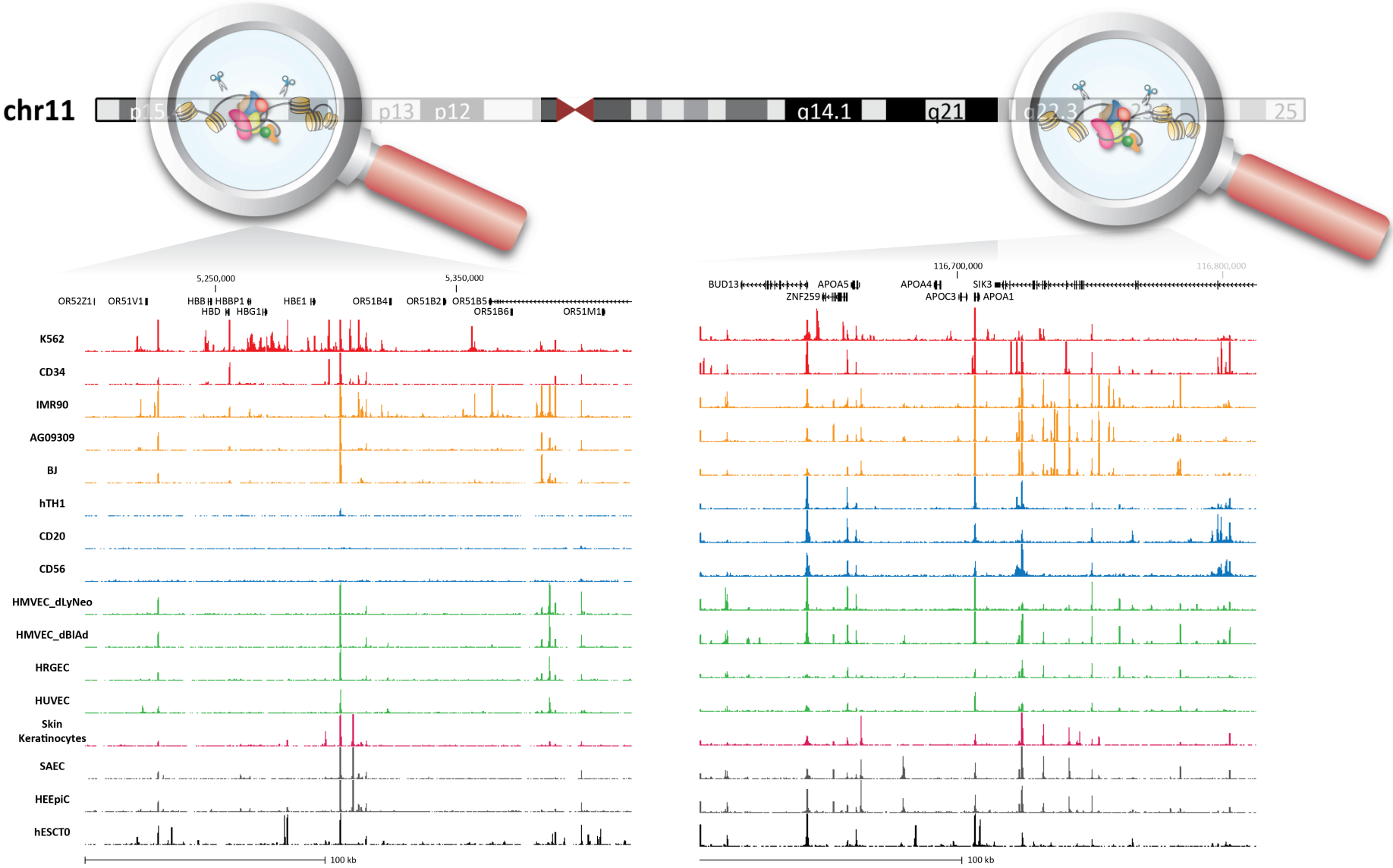


Maurano et al., Science 2012

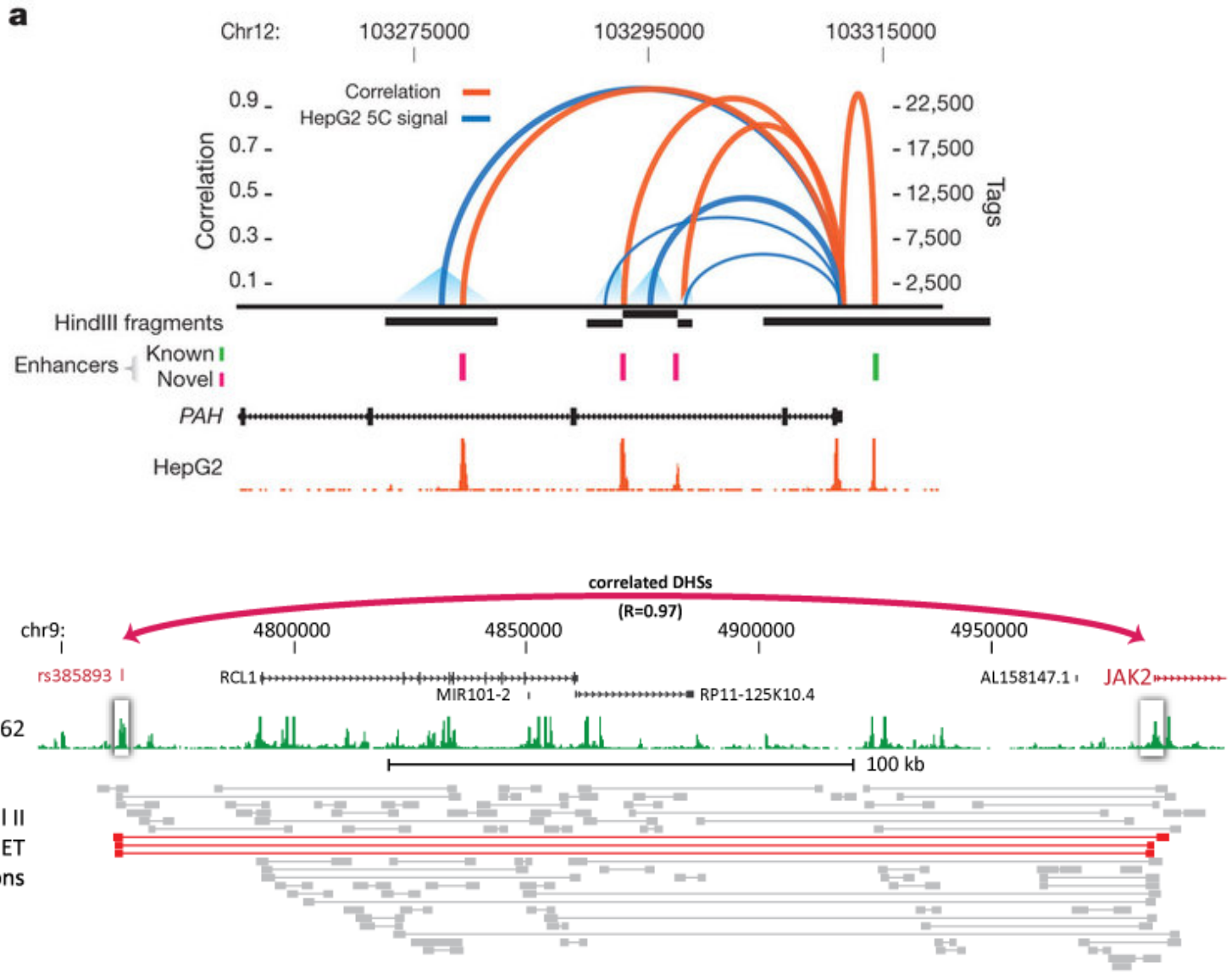
#4
Don’t assume local effects

#4
Regulatory DNA harboring GWAS variants mainly controls distant genes

Closing in on a comprehensive atlas of human regulatory DN



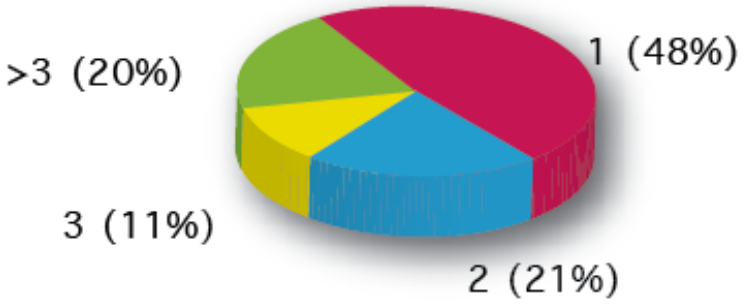
How is regulatory DNA 'wired' *in cis?*



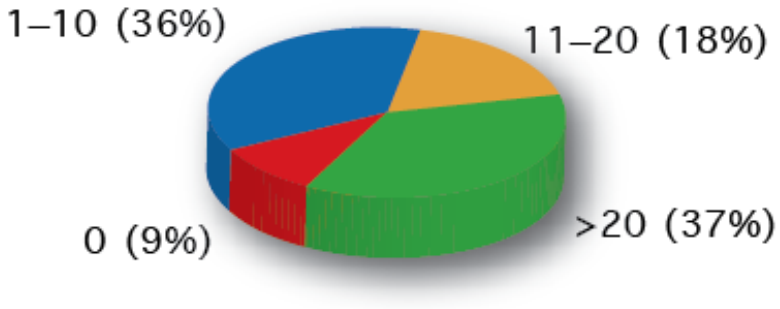
Thurman et al. Nature 2012
Maurano et a., Science 2012

How is regulatory DNA 'wired' *in cis?*

Promoter DHSs connected per distal DHS (n = 578,905)

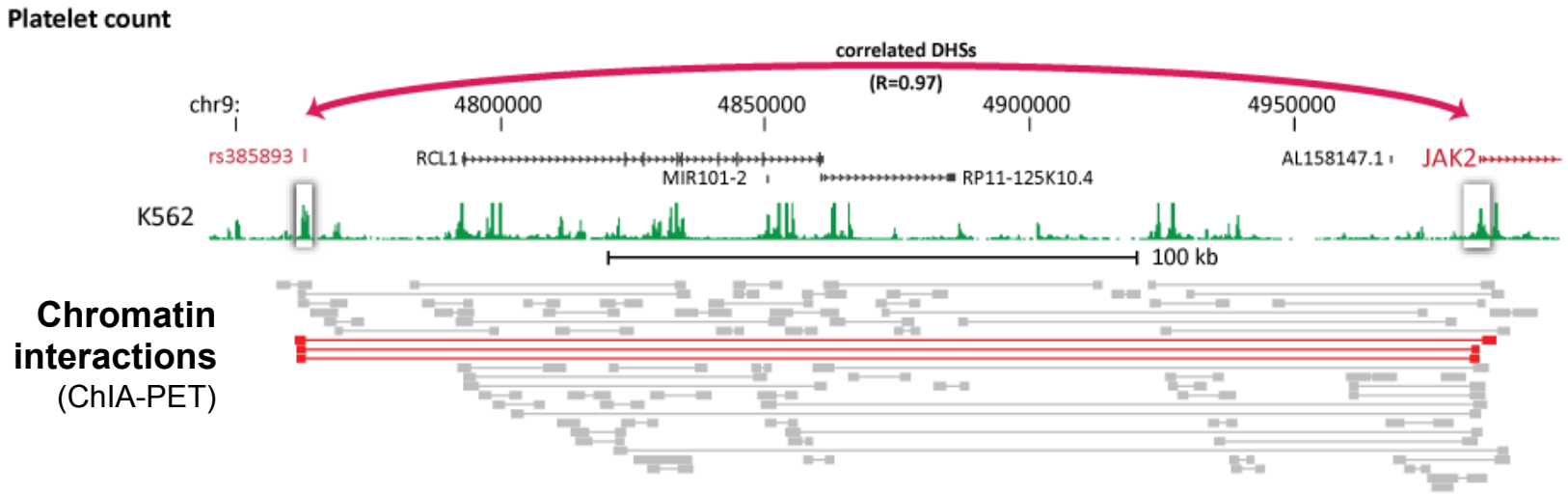


Distal DHSs connected per promoter DHS (n = 69,965)

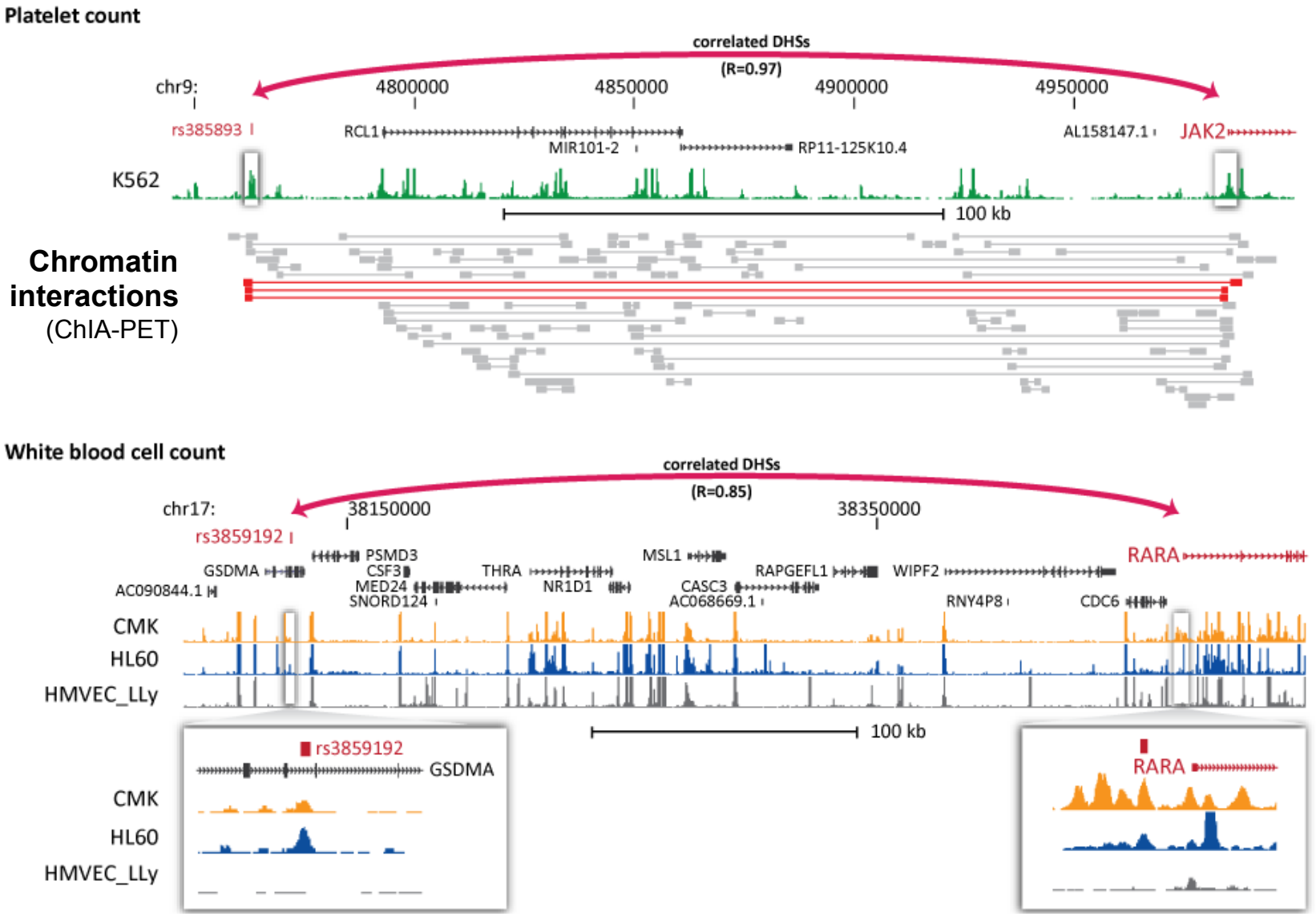


Thurman et al., Nature 2012

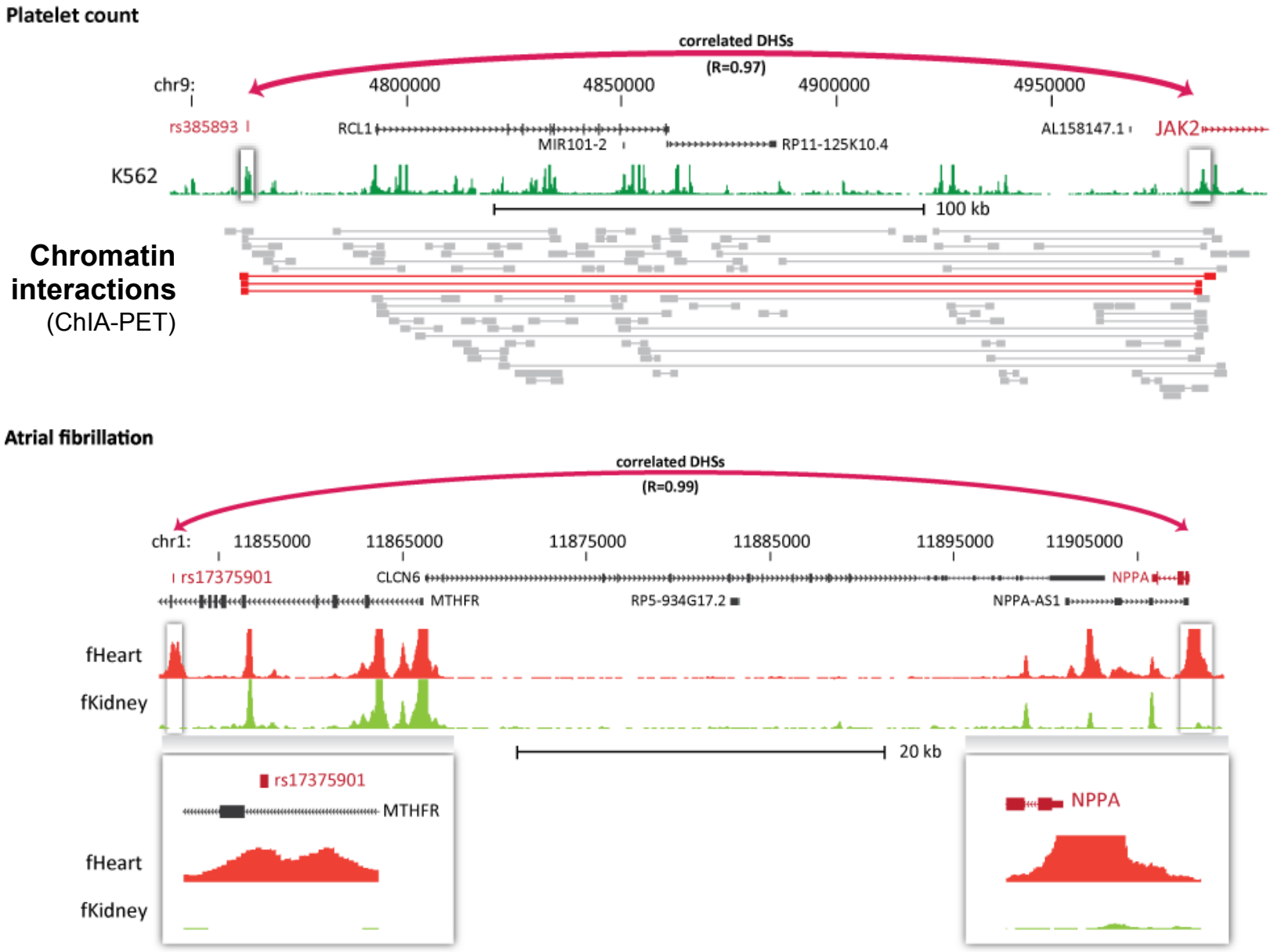
Regulatory GWAS variants linked to distant genes with causative potential



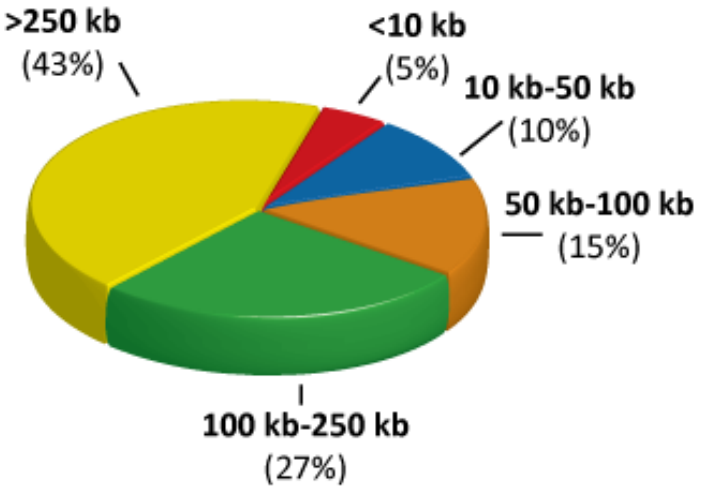
Regulatory GWAS variants linked to distant genes with causative potential



Regulatory GWAS variants linked to distant genes with causative potential



Regulatory GWAS variants linked to distant genes with pathogenic potentia

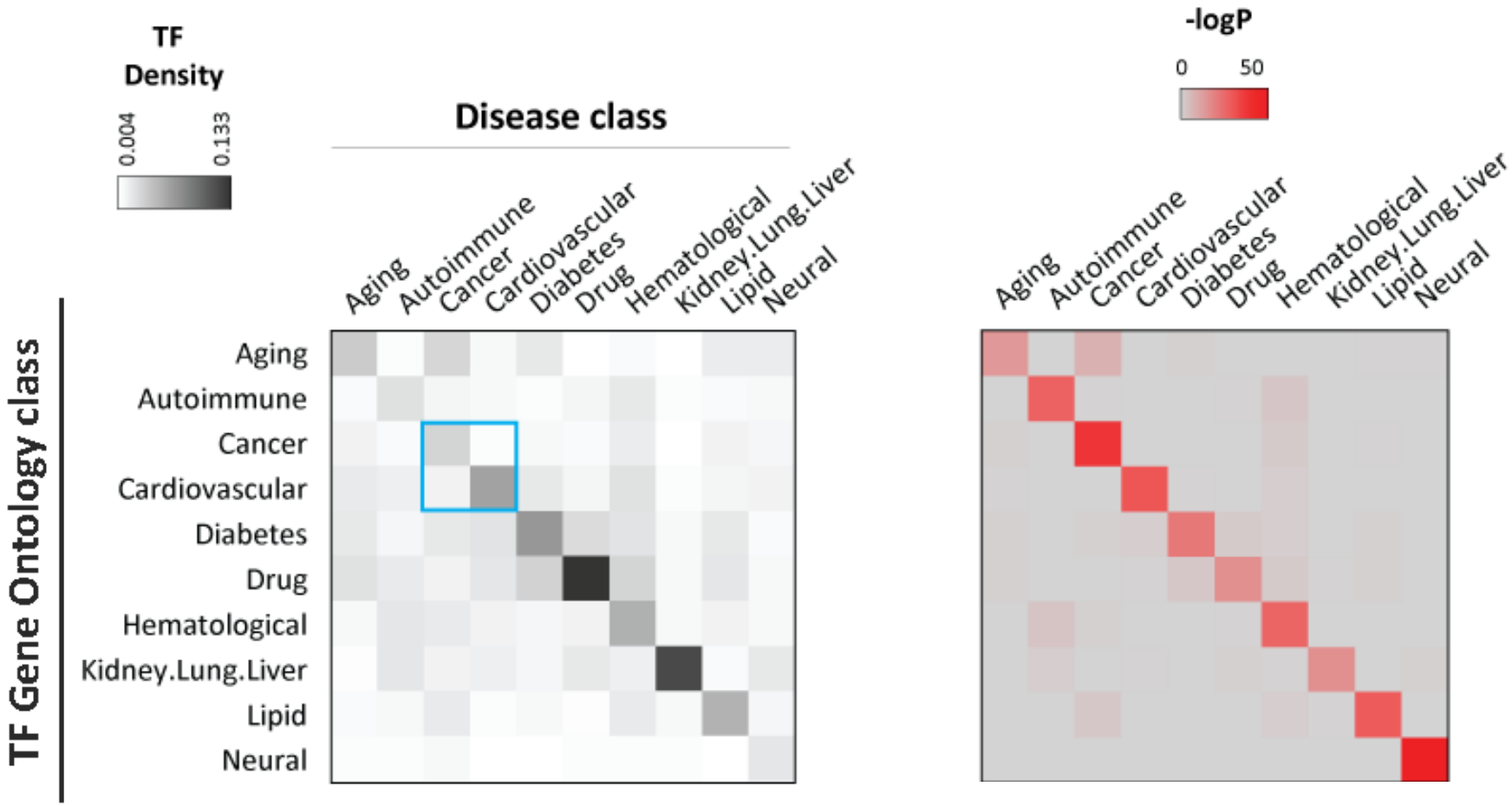


Disease or trait	*r*	Target gene	Function	Distance (kb)
Amyotrophic lateral sclerosis	1	SYNGAP1*	Axon formation; component of NMDA complex	411
Crohn's disease	1	TRIB1*	NF-κB regulation	95
Time to first primary tooth	0.99	PRDM1*	Craniofacial development	452
C-reactive protein	0.99	NLRP3	Response to bacterial pathogens	20
Multiple sclerosis	0.98	AHI1*	White matter abnormalities	149
QRS duration	0.96	SCN10A*	Sodium channel involved in cardiac conduction	181
Breast cancer	0.96	TACC2*	Tumor suppressor	411
Schizophrenia/brain imaging	0.95	KIF1A*	Neuron-specific kinesin involved in axonal transport	428
Brain structure	0.94	CXCR6*	Chemokine receptor involved in glial migration	357
Rheumatoid arthritis	0.94	CTSB*	Cysteine proteinase linked to articular erosion	359
Ovarian cancer	0.93	HSPG2*	Ovarian tumor suppressor	268
Multiple sclerosis	0.93	ZP1*	Known autoantigen	153
ADHD	0.93	PDLIM5*	Neuronal calcium signaling	328
Breast cancer	0.88	MAP3K1*	Response to growth factors	158
Amyotrophic lateral sclerosis	0.88	CNTN4	Neuronal cell adhesion	306
Schizophrenia	0.81	FXR1*	Cognitive function	120
Type 1 diabetes	0.75	ACAD10*	Mitochondrial oxidation of fatty acids	343
Lupus	0.74	STAT4	Mediates IL-12 immune response and T_H1 differentiation	113

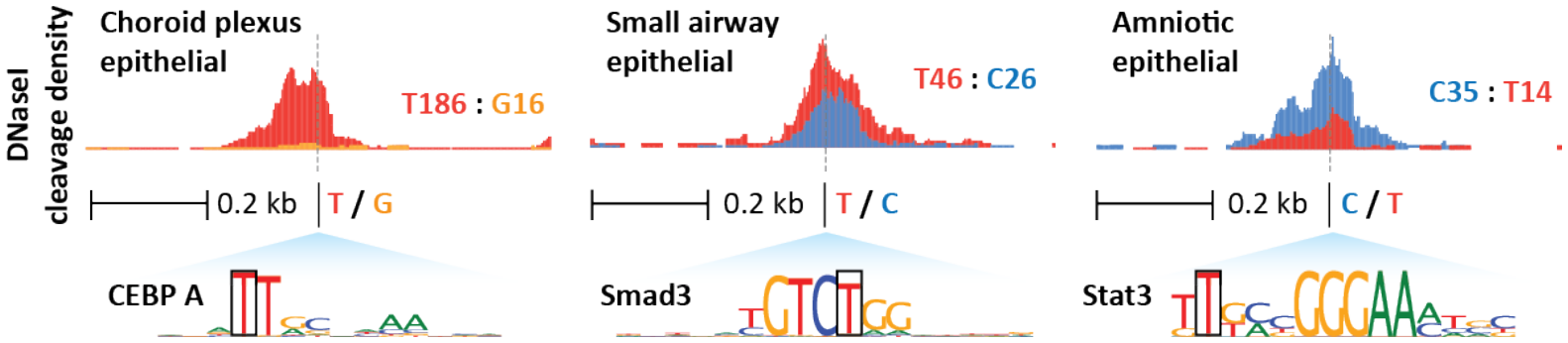
Maurano et al., Science 2012

#5
GWAS variants in regulatory DNA selectively localize to relevant TF recognition sites and many directly affect TF occupancy

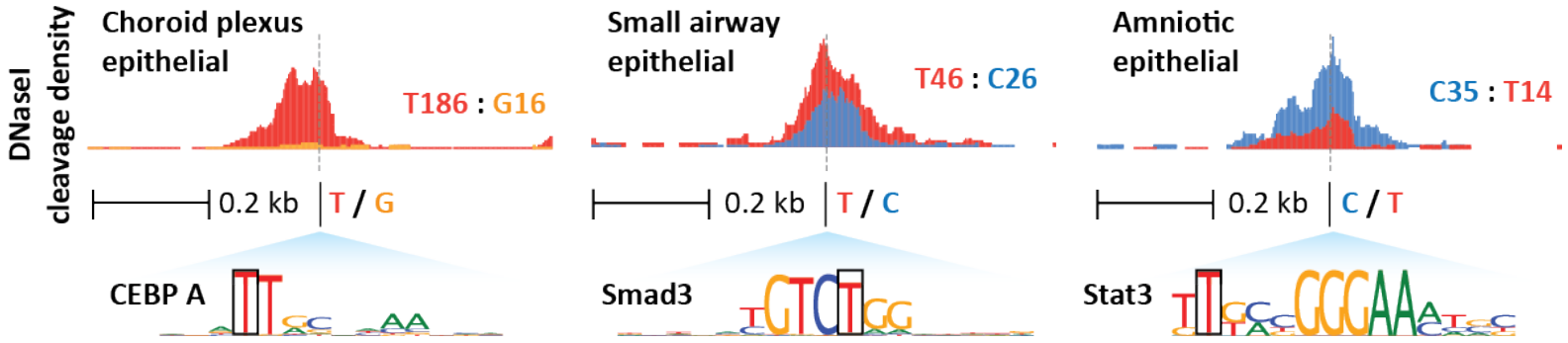
Within regulatory DNA, disease-associated variants systematically localize within relevant TF recognition sites



Disease/trait variants specify allelic chromatin states



Disease/trait variants specify allelic chromatin states



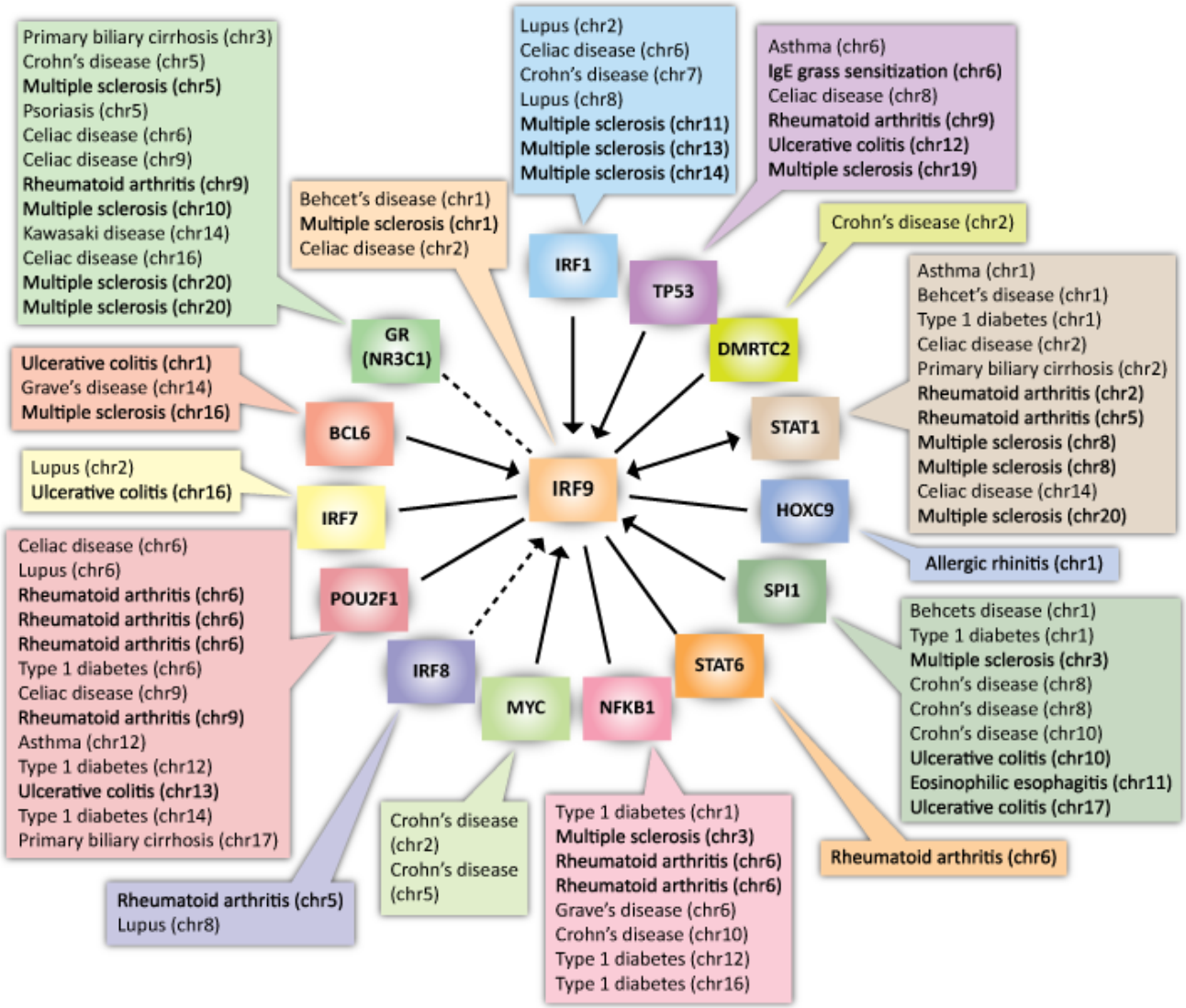
Overall, 20.5% of GWAS SNPs exhibit significant allelic imbalance

For those with high sequencing depth (>200x), 38.7%

Maurano et al., Science 2012

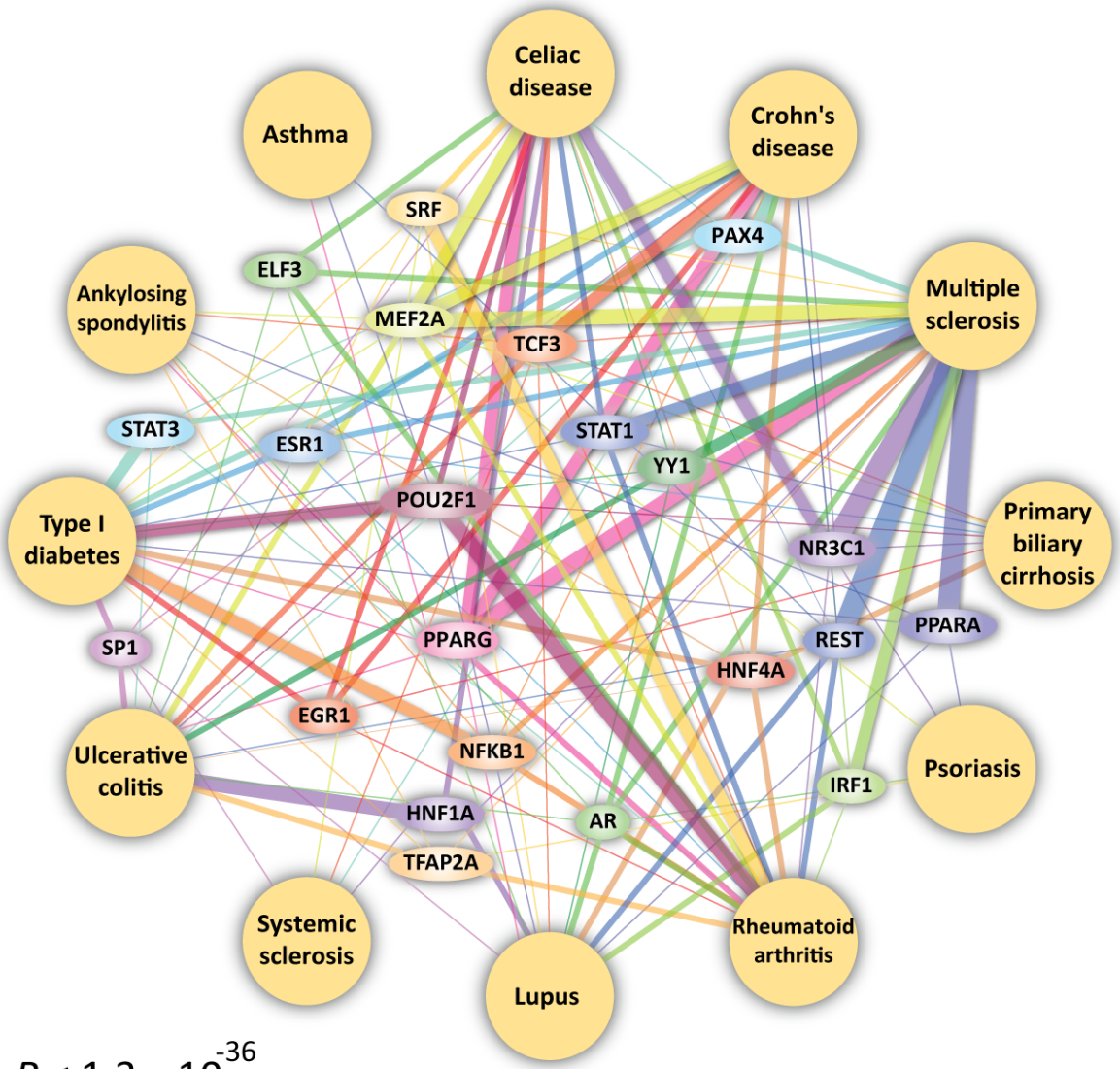
#6
GWAS variants cluster in regulatory pathways and form regulatory networks

~25% of rGWAS inflammatory disease variants lie in IRF9 pathway



$P < 1.6 \times 10^{-13}$

A common regulatory network for autoimmune disease

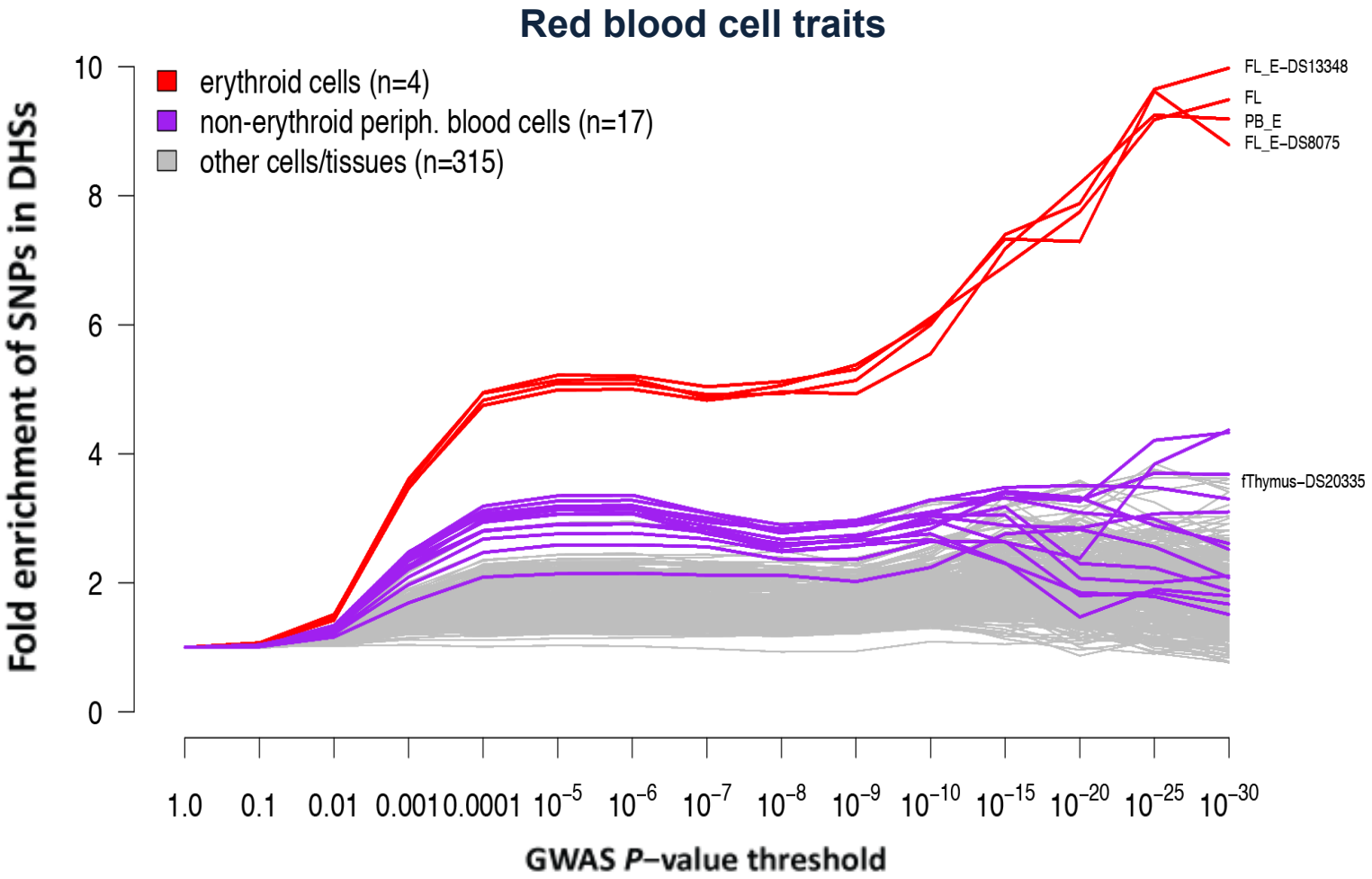


$P < 1.3 \times 10^{-36}$

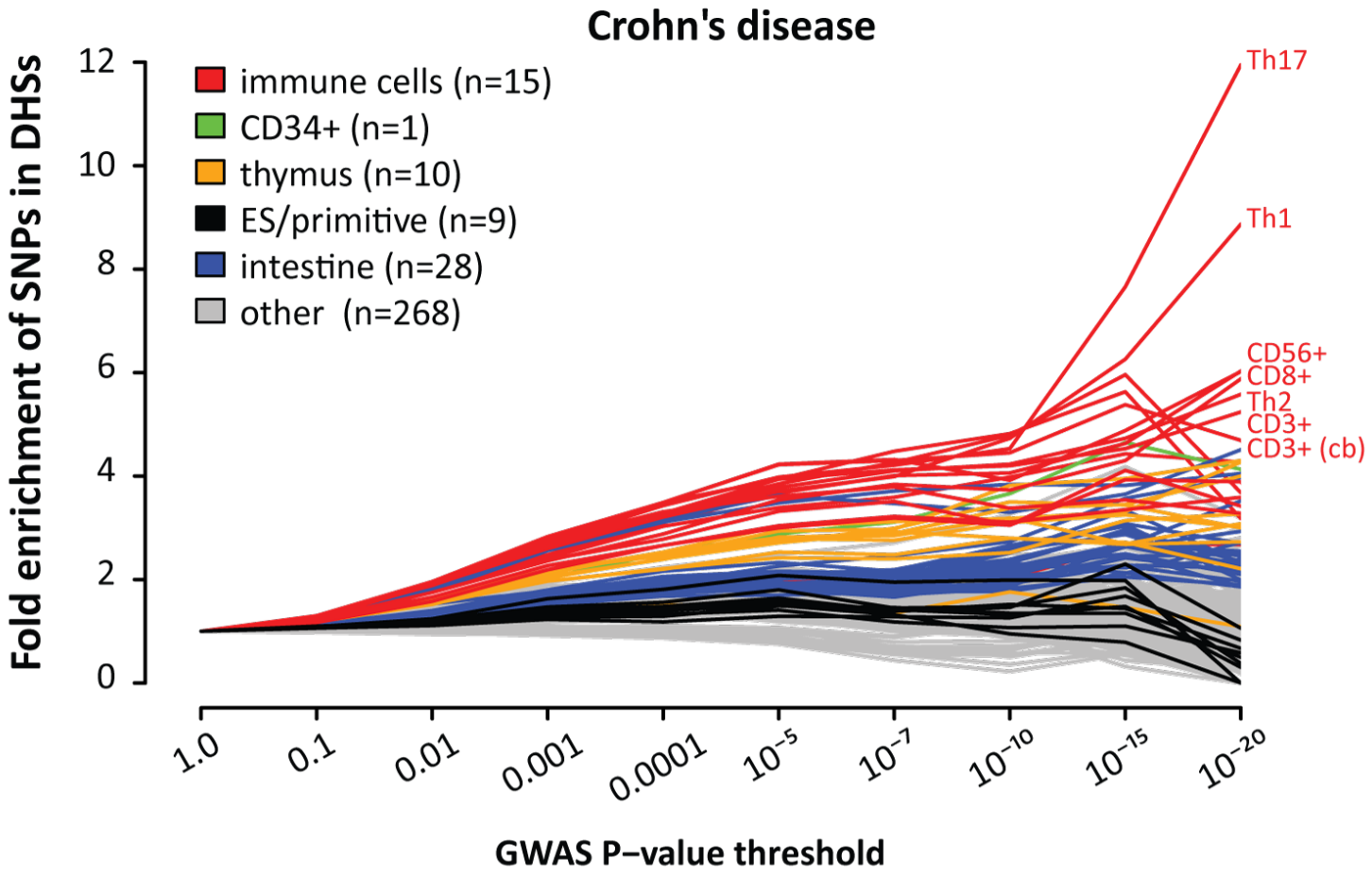
Maurano et al., Science 2012

#7
ENCODE and GWAS data can be combined to pinpoint disease/trait-relevant cell types

Cell-selective enrichment of trait-associated variants



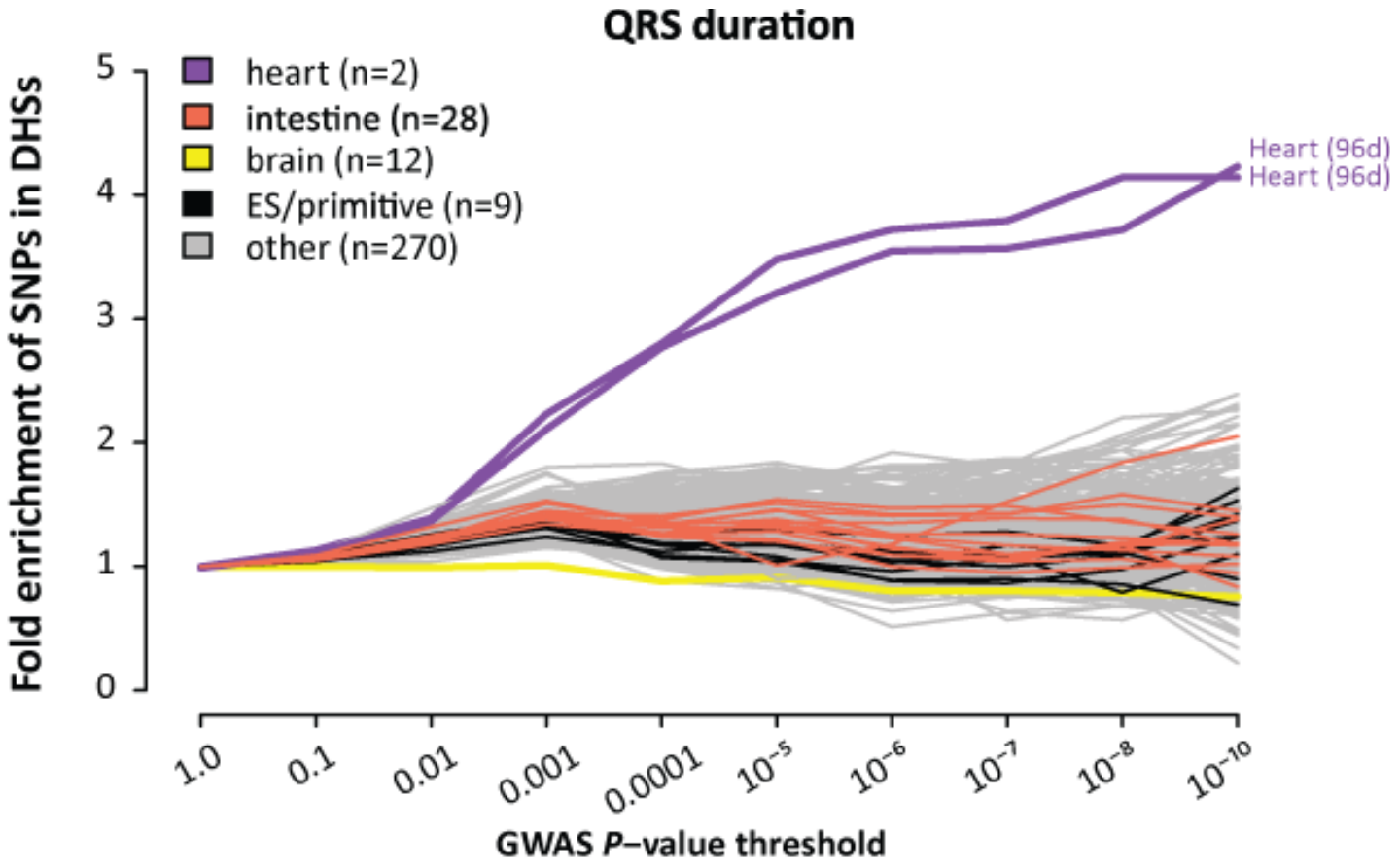
Selective enrichment of GWAS variants in pathogenic cell types



#8

Many, many more variants show these effects than caconical ‘genome-wide significant’ SNPs

Selective enrichment of GWAS variants in pathogenic cell types



Maurano et al., Science 2012

Acknowledgements

Analysis

Matthew Maurano, Rich Humbert, Eric Rynes
Eric Haugen, Richard Sandstrom, Audra Johnson,
Eric Haugen, Alex Reynolds, Bob Thurman

Data production

Raj Kaul, Scott Hansen, Peter Sabo,
Molly Weaver, Theresa Cantwell, Kristin
Lee, Shinny Vong, Vaughan Roach, Erica
Gist, Sandra Stehling-Sun

Collaborators

Ian Glass (UW Birth Def. Res. Lab)
Shelly Heimfeld (FHCRC)
Nona Sotoodehia and Jen Brody (UW Cardiology)
Chris Cotsapas (Yale)
Shamil Sunyaev (Harvard)

Funding: **NHGRI (ENCODE)**

NIH Common Fund (Roadmap Epigenomics)