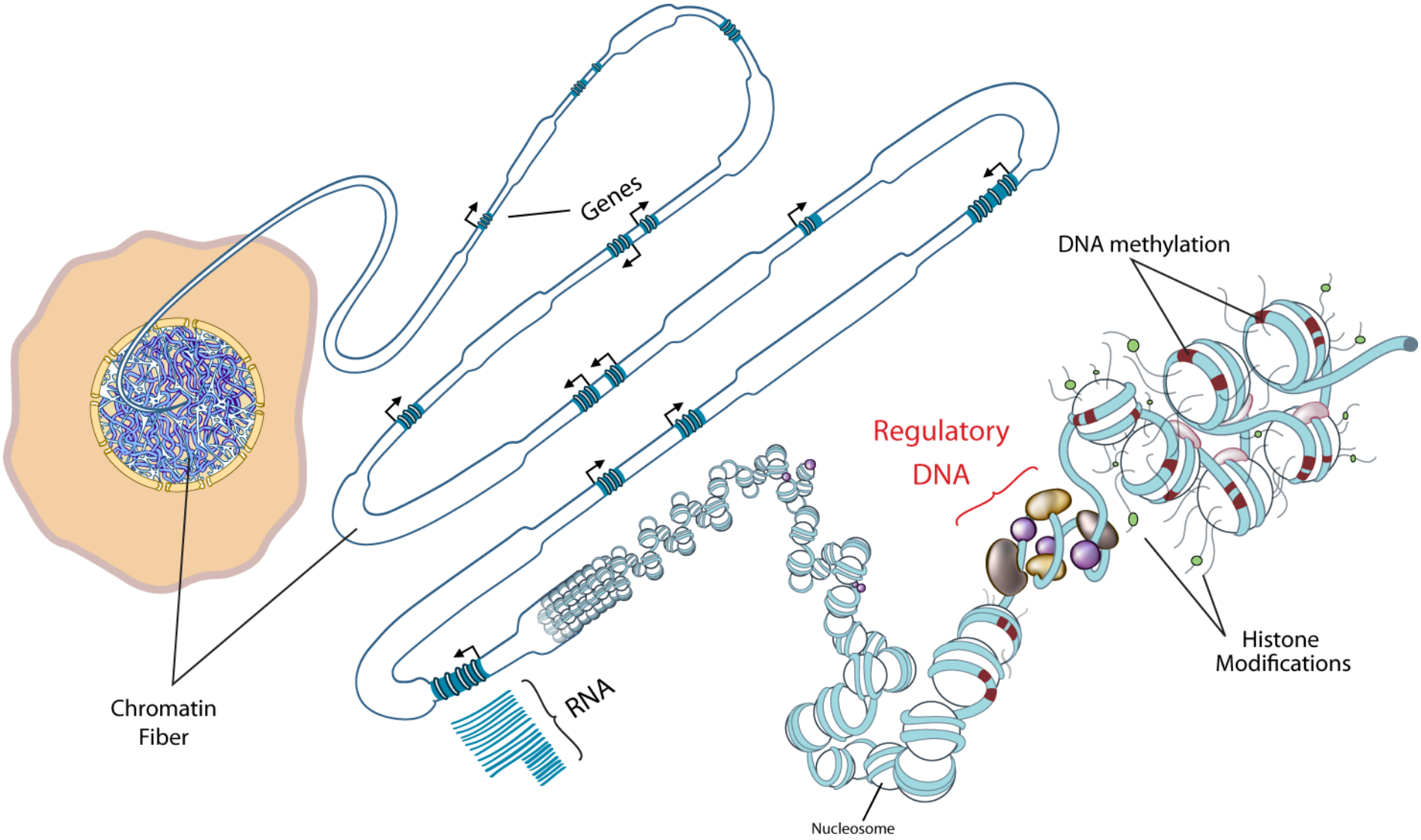
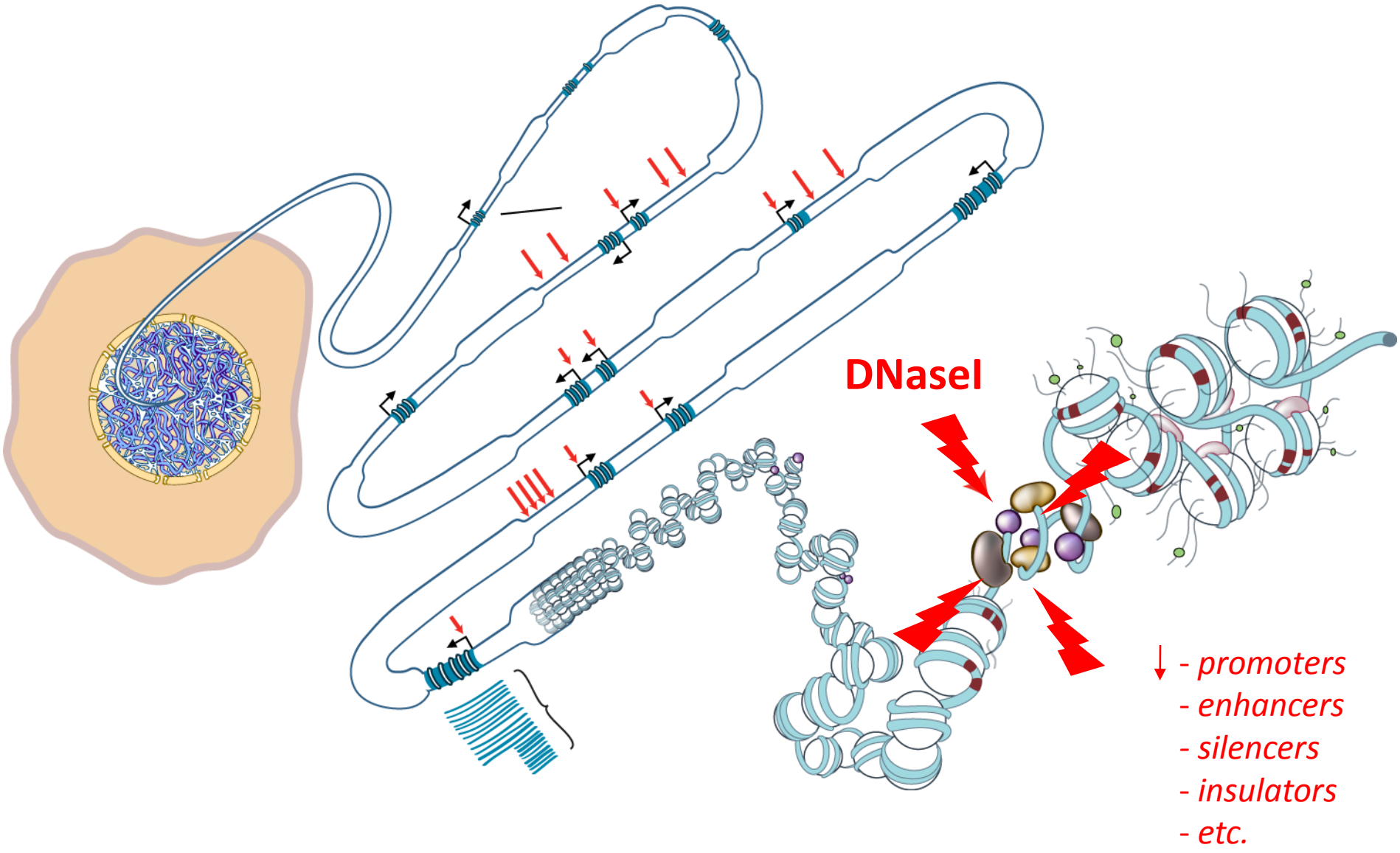
# Illuminating disease-associated variation with ENCODE & Roadmap Data

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

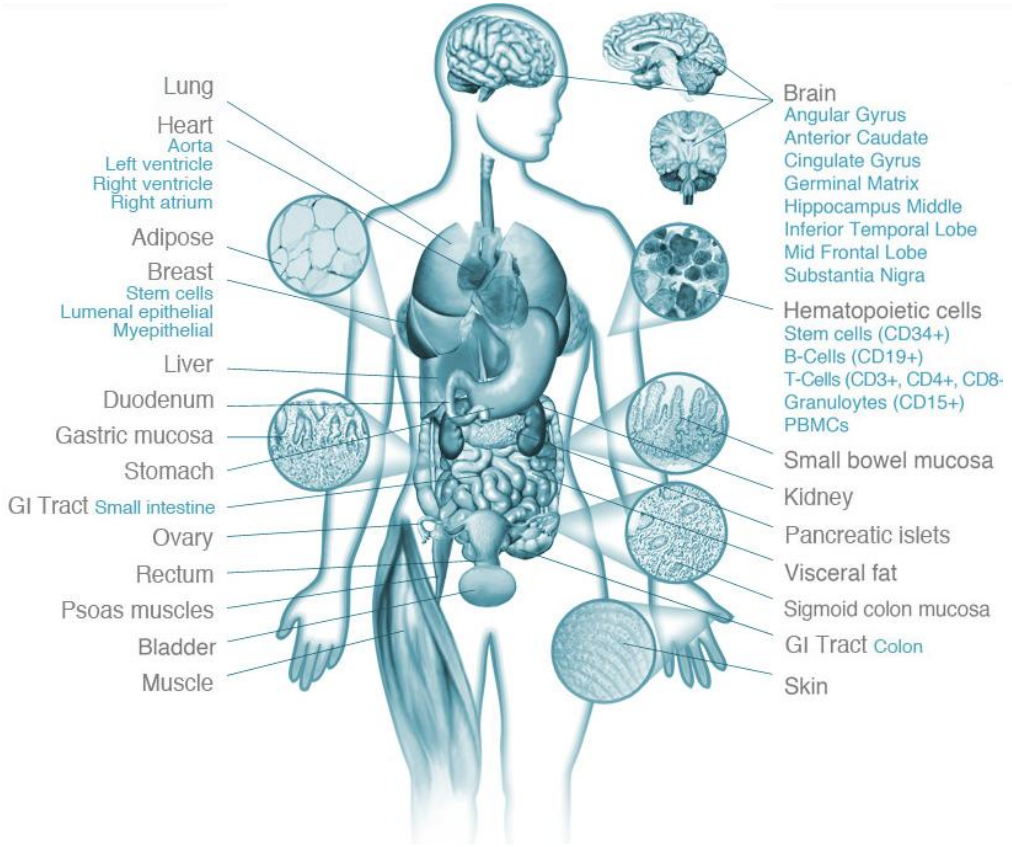
# Genes, regulatory DNA, and epigenetic features

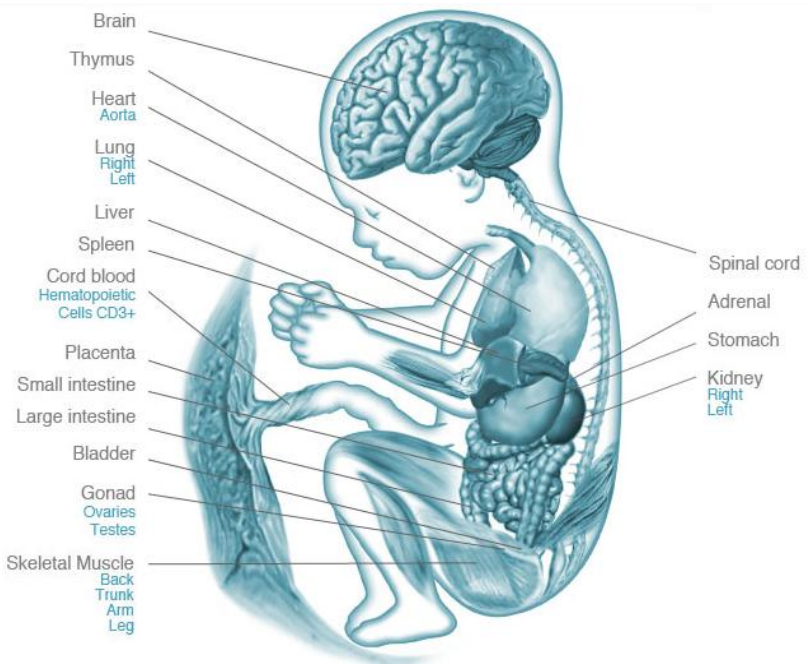


# Genes, regulatory DNA, and epigenetic features



# Creating deep maps of human regulatory DNA




genome.ucsc.edu
www.uwencode.org

www.roadmapepigenomics.org
vishub.wustl.edu

www.encode-roadmap.org

# A global atlas of human regulatory DNA

- **>340 cell types, tissues, and developmental stages**
  *Model cell lines, Primary cells in culture*
  Ex vivo *hematopoietic cells, Fetal tissues* *(late 1st – late 2nd trimest*

- **~100,000 - >250,000 DHSs per cell type** (0.5-1.5% of genom

- **Collectively >4 million distinct elements**
  ~800,000 cell type-specific, ~3,500 constitutive

- **Comprehensive annotation of known regulatory DNA**
  99% of experimentally-validated enhancers, silencers, insulators, locus control regions

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

## Selected data

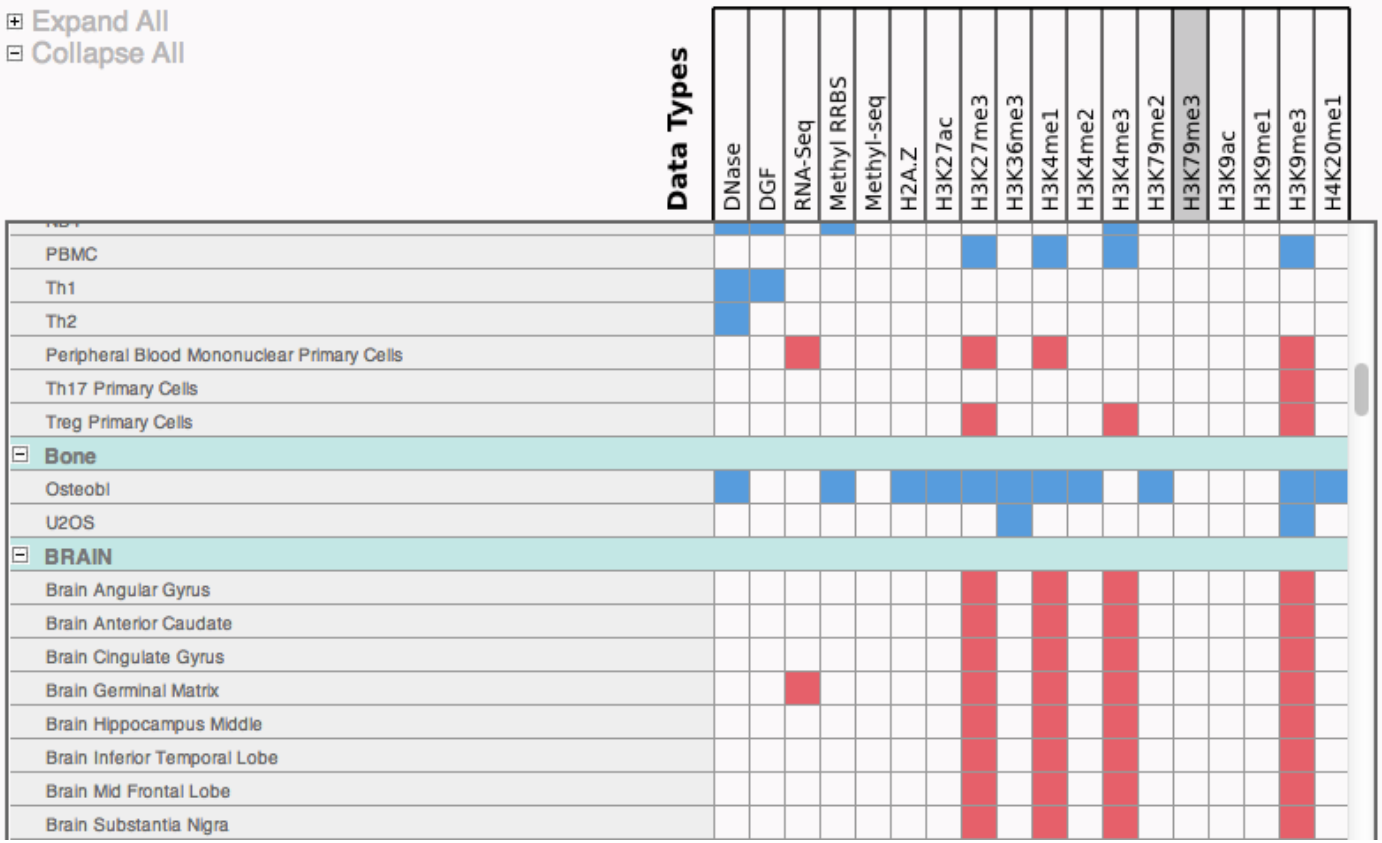
Roadmap ENCODE Combined Selected

Data points you select in the tables above will appear below for preview before viewing them in the browser.

Go to Browser

Unselect all data cell points

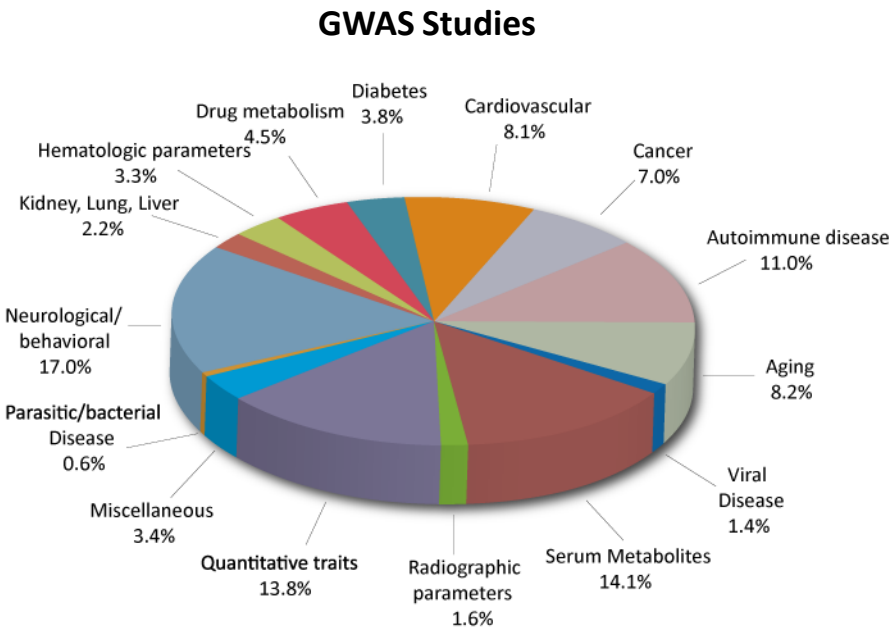
⊞ Expand All
⊟ Collapse All

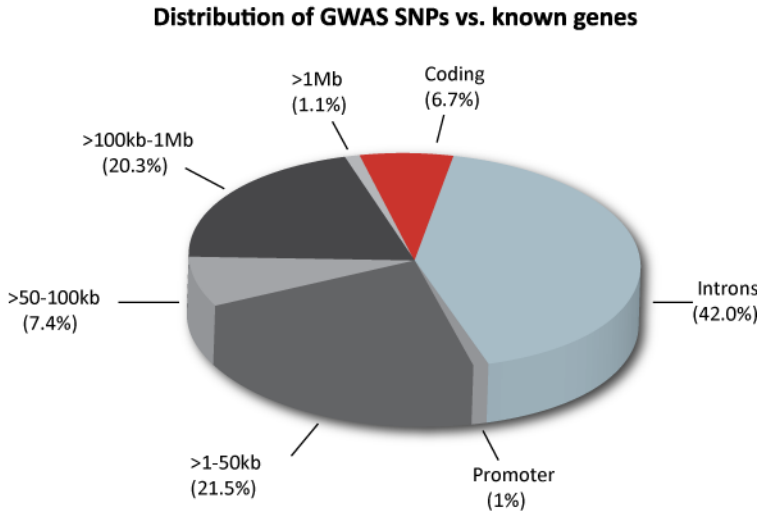



**www.encode-roadmap.org**

# Regulatory DNA variation associated with common diseases and traits

# Identification of disease- and trait-associated variation by GWAS




**GWAS disease/trait associated variants**

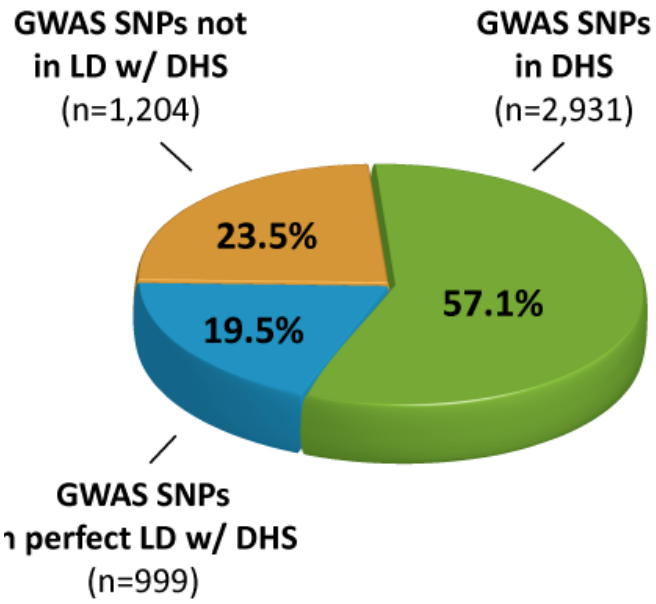
**x**

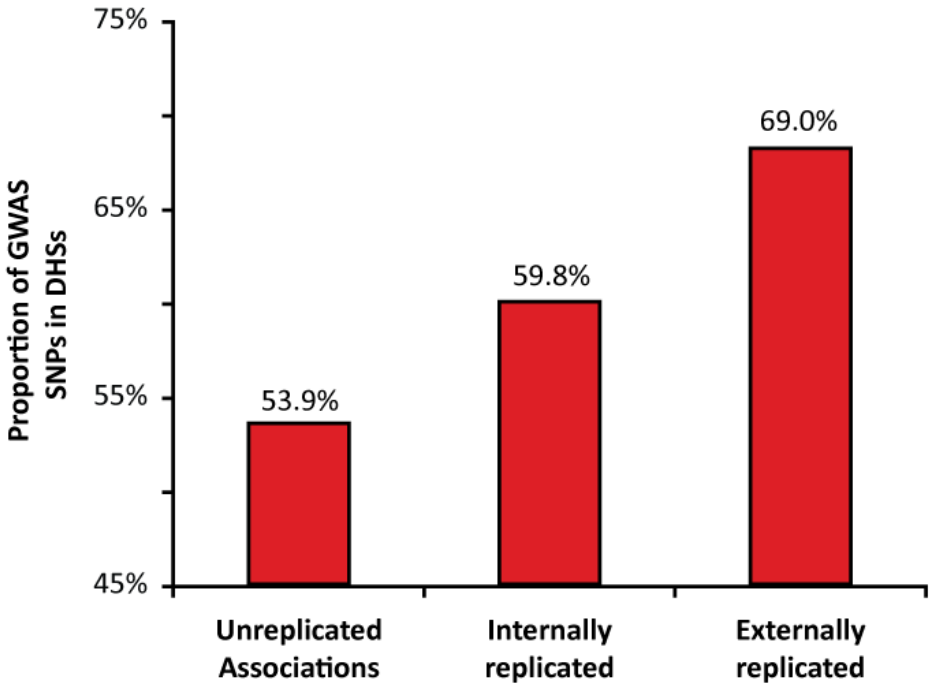
**Maps of regulatory DNA in >300 diverse cell and tissue types**

*Maurano et al., Science 2012*

# #1
# Disease-associated variation is concentrated in regulatory DNA

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



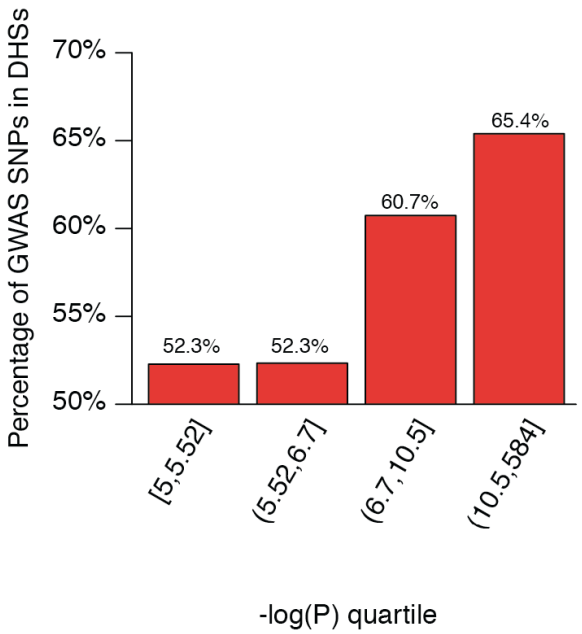



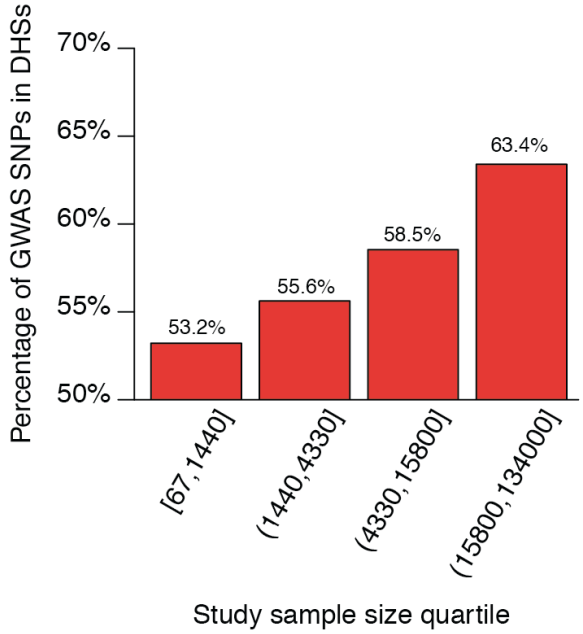

~1.8-fold for all replicated variants in all disorders

**>10-fold for specific disese-cell type pairings**

*Maurano et al., Science 2012*

# The effect increases monotonically with other measures of higher quality associations

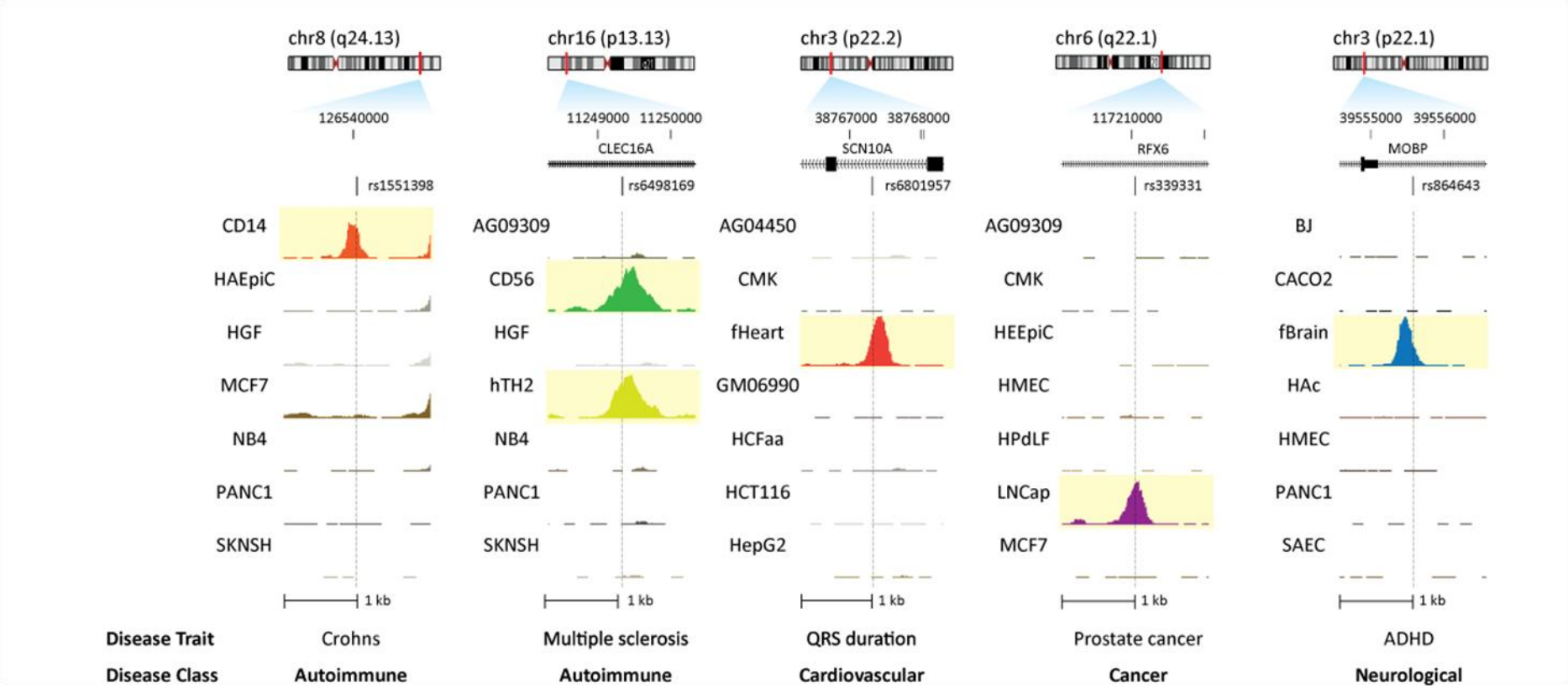



*Maurano et al., Science 2012*

# #2
# GWAS variants selectively localize in pathologically relevant cell types

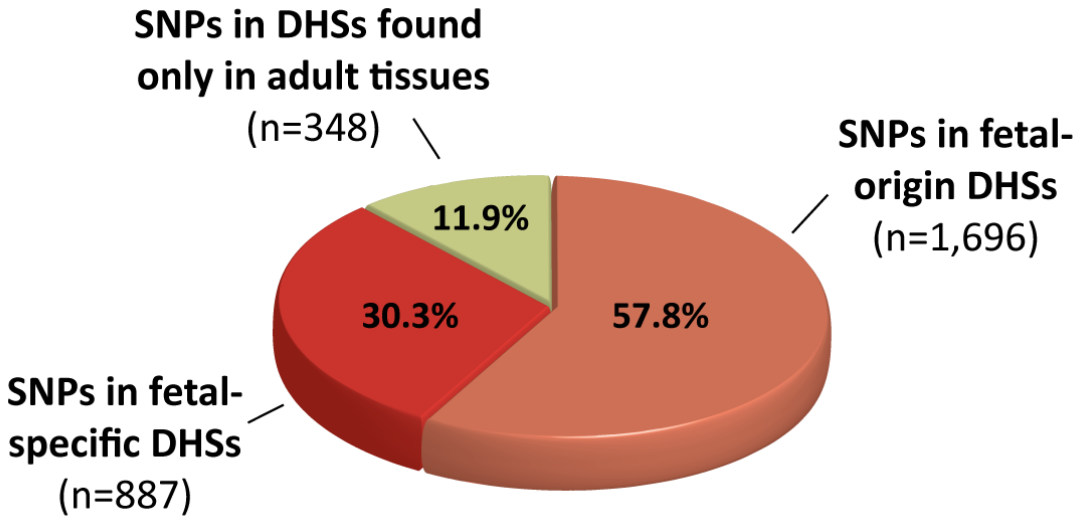
# Disease-associated variation clusters in pathogenic or target cell types




*Maurano et al., Science 2012*

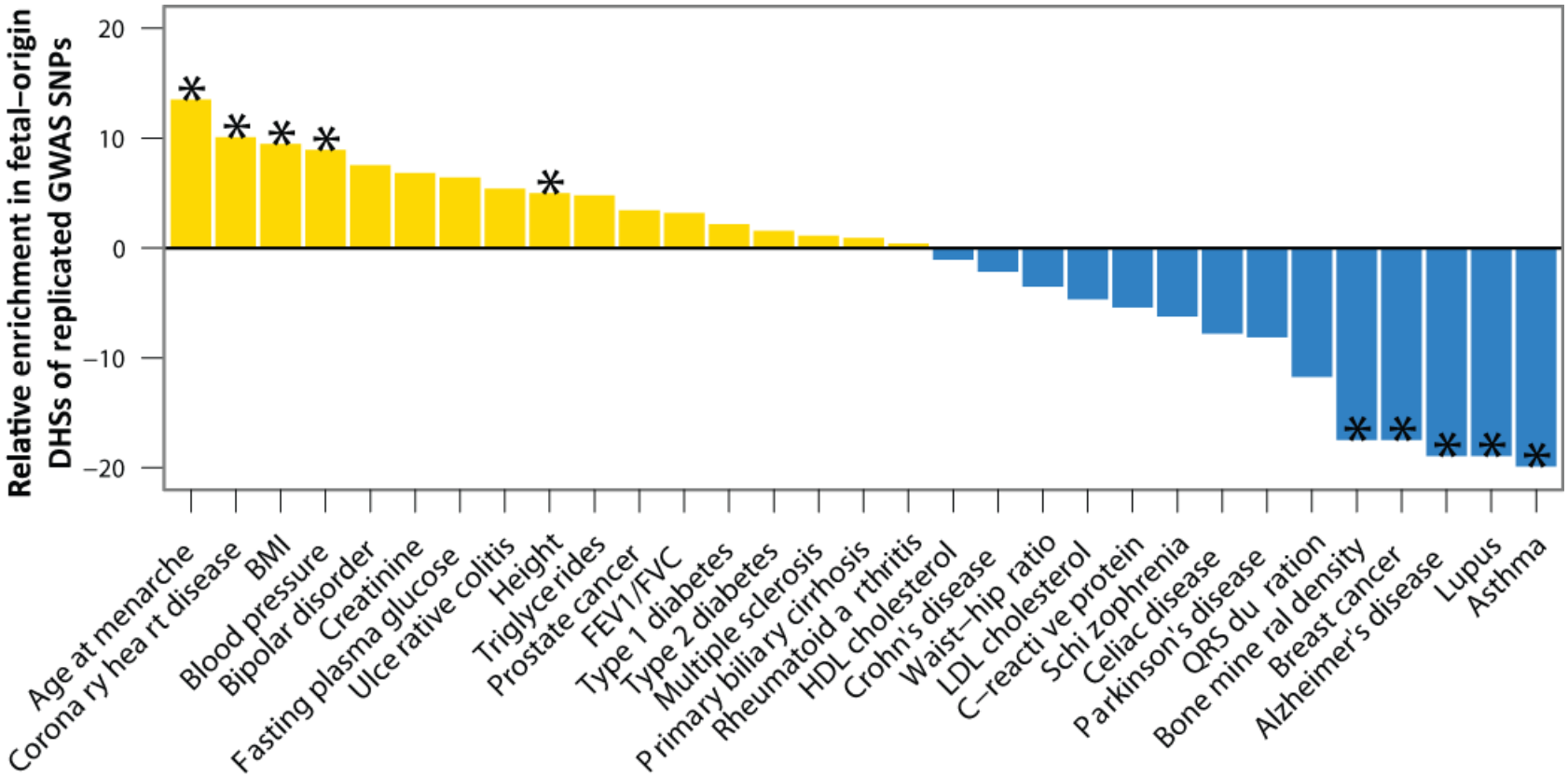
# #3
# Common diseases and traits with known or suspected developmental contributions preferentially localize in fetal regulatory DNA

# 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

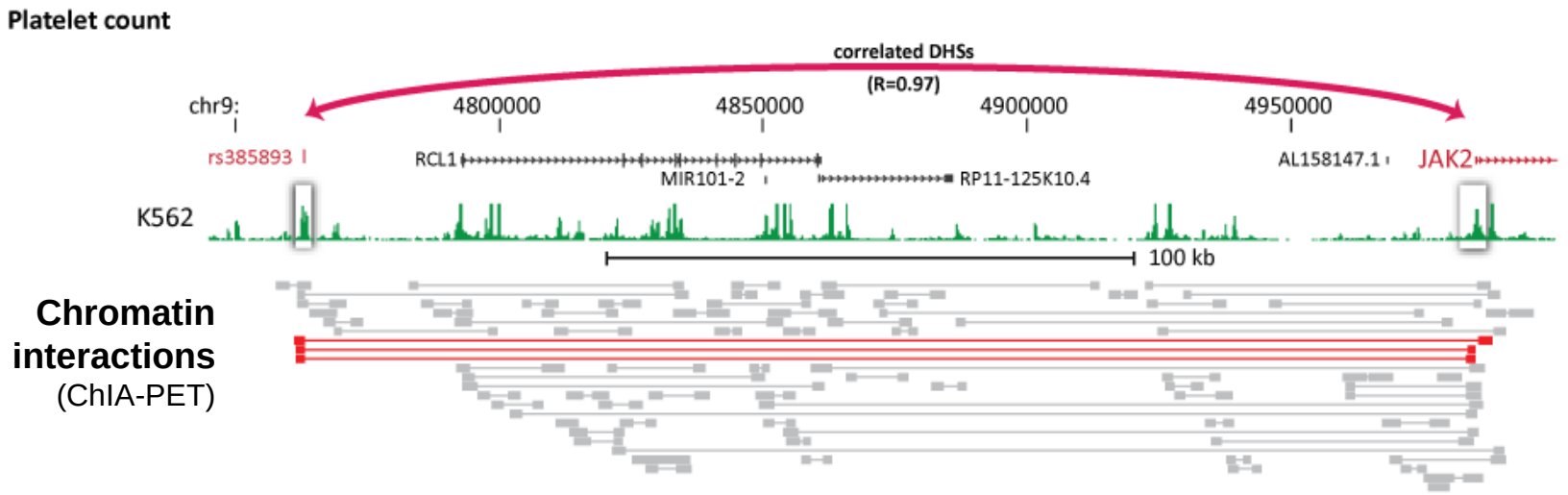



*Maurano et al., Science 2012*

# #4
# Don’t assume local effects

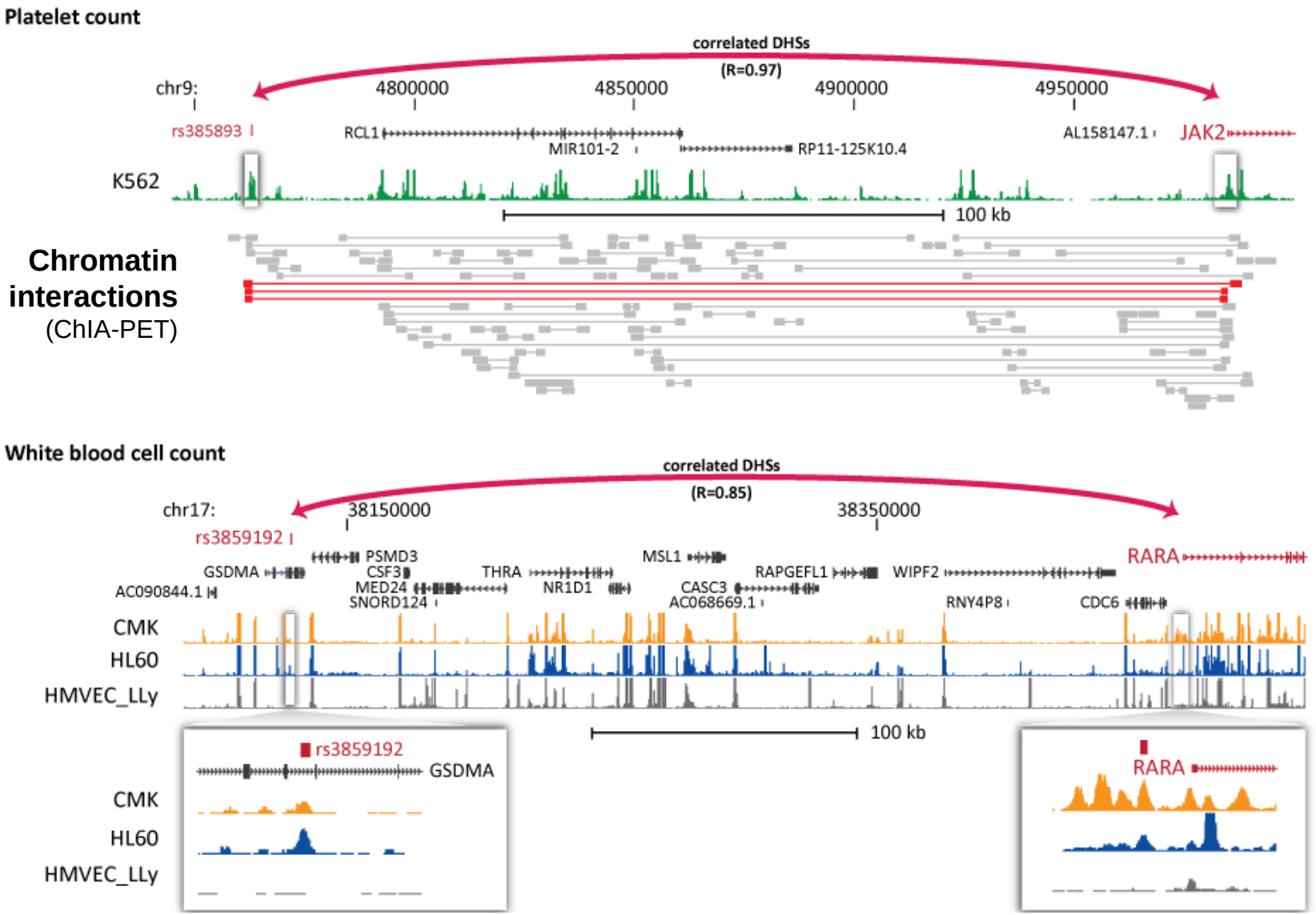
# #4
# GWAS variants in regulatory DNA control distant disease-related genes

# Regulatory GWAS variants linked to distant genes with causative potential



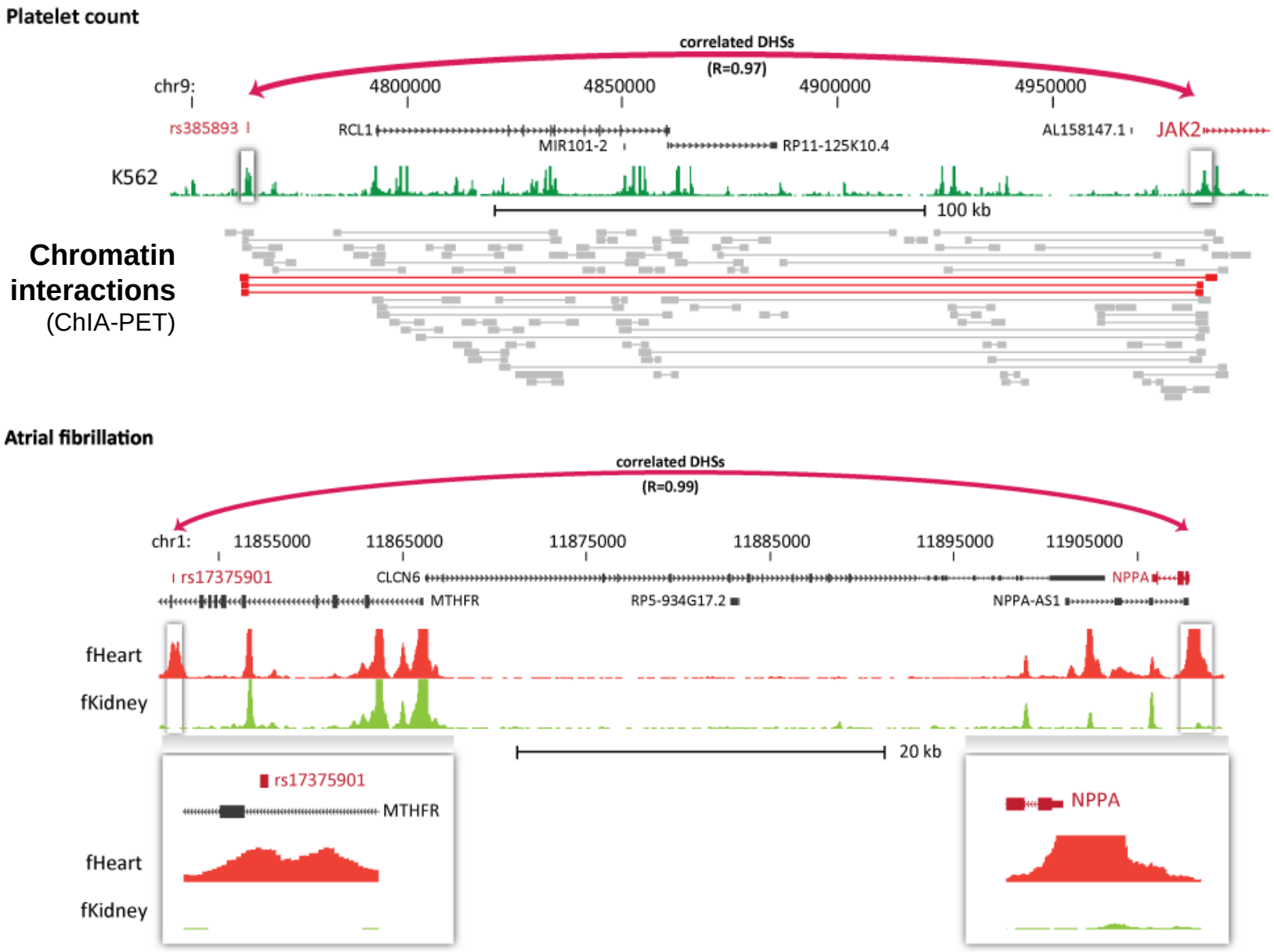

*Maurano et al., Science 2012*

# Regulatory GWAS variants linked to distant genes with causative potential



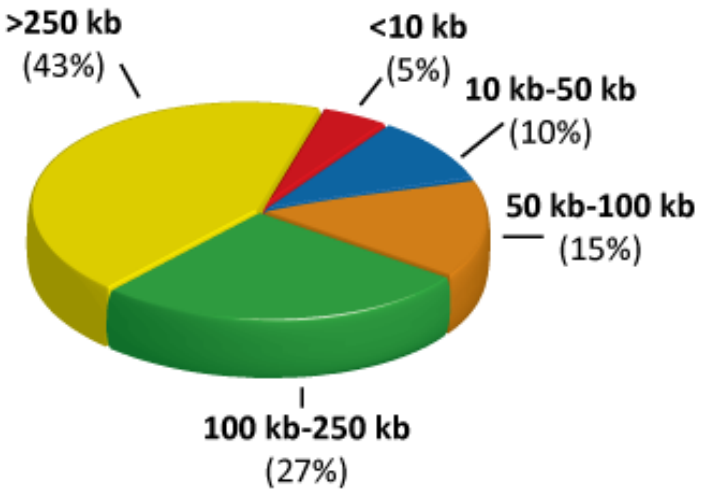

*Maurano et al., Science 2012*

# Regulatory GWAS variants linked to distant genes with causative potential




*Maurano et al., Science 2012*

# Regulatory GWAS variants linked to distant genes with causative potential

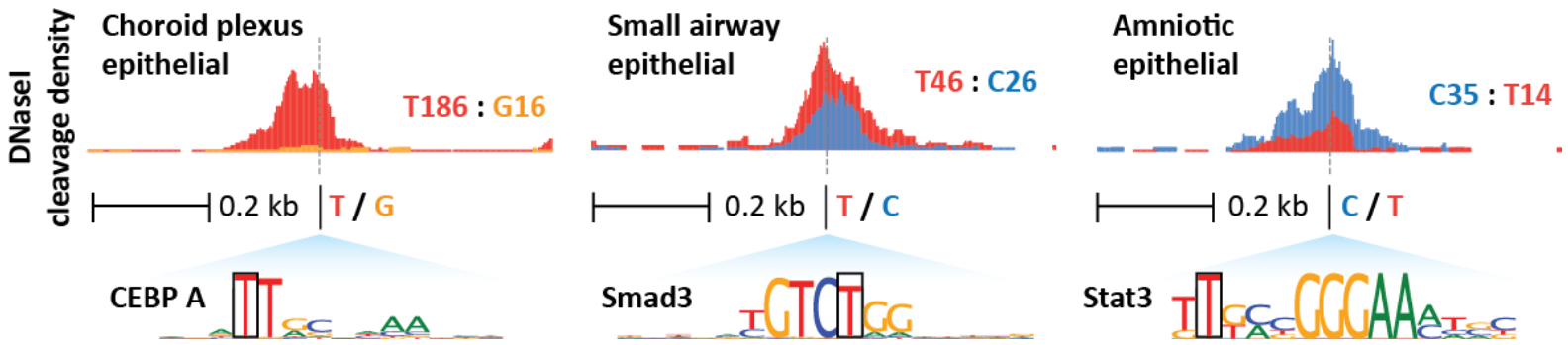



| Disease or trait | *r* | Target gene | Function | Distance (kb) |
|---|---|---|---|---|
| Amyotrophic lateral sclerosis | 1 | SYNGAP1* | Axon formation; component of NMDA complex | 411 |
| Crohn's disease | 1 | TRIB1* | NF-κB regulation | 95 |
| Time to first primary tooth | 0.99 | PRDM1* | Craniofacial 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 |

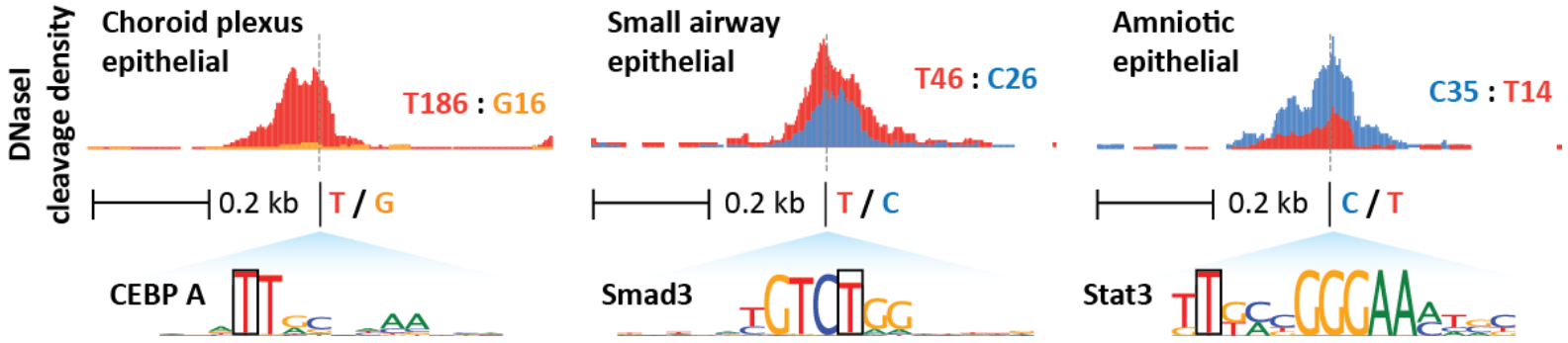
*Maurano et al., Science 2012*

#5
A significant fraction of GWAS variants falling in regulatory DNA directly affect TF occupancy

# Disease/trait variants specify allelic chromatin states



# Disease/trait variants specify allelic chromatin states



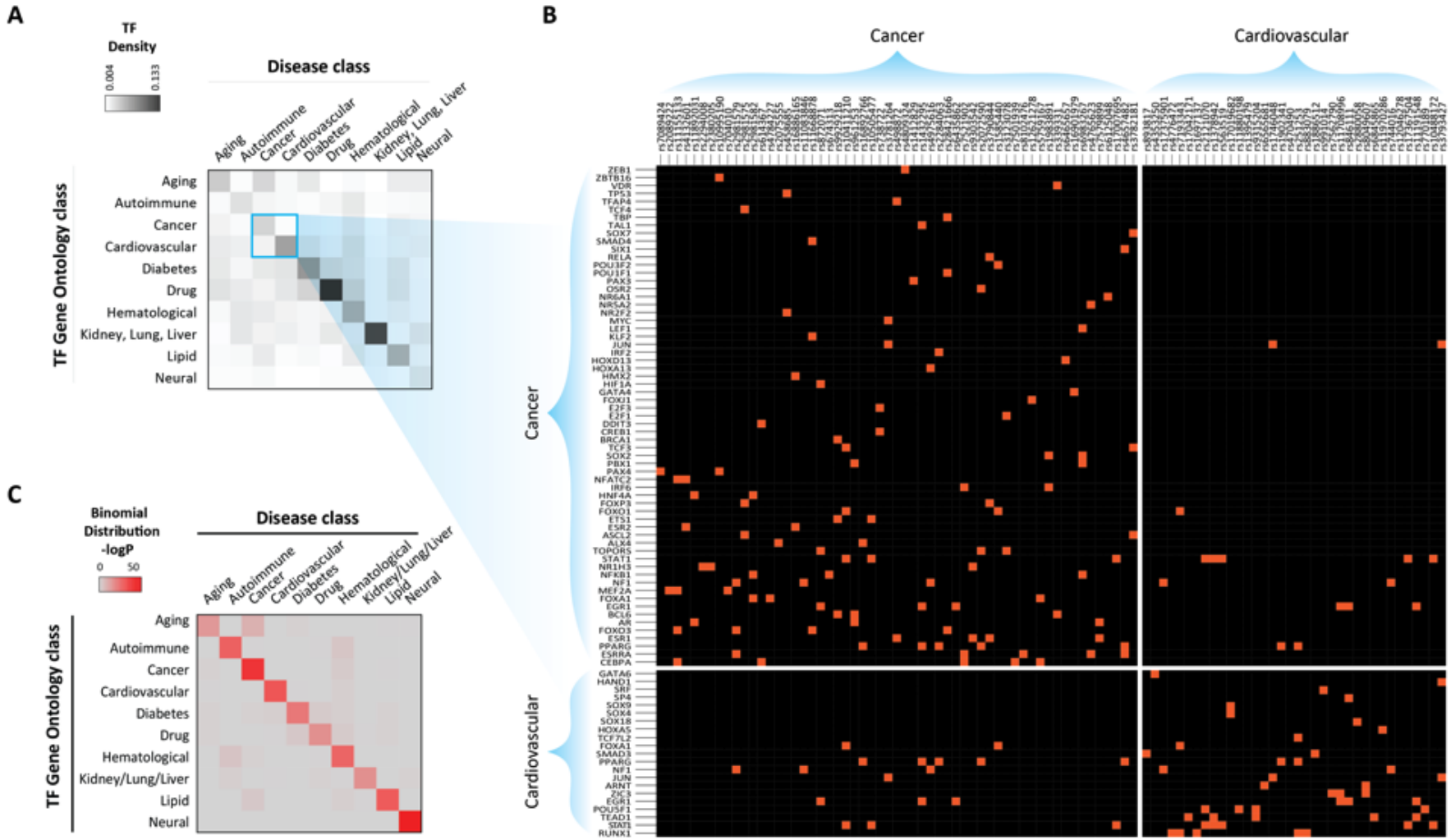

**Overall, 20.5% of GWAS SNPs exhibit significant allelic imbalance**

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

*Maurano et al., Science 2012*

# #6
# Disease-associated variants cluster in regulatory pathways and form regulatory networks

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




*Maurano et al., Science 2012*

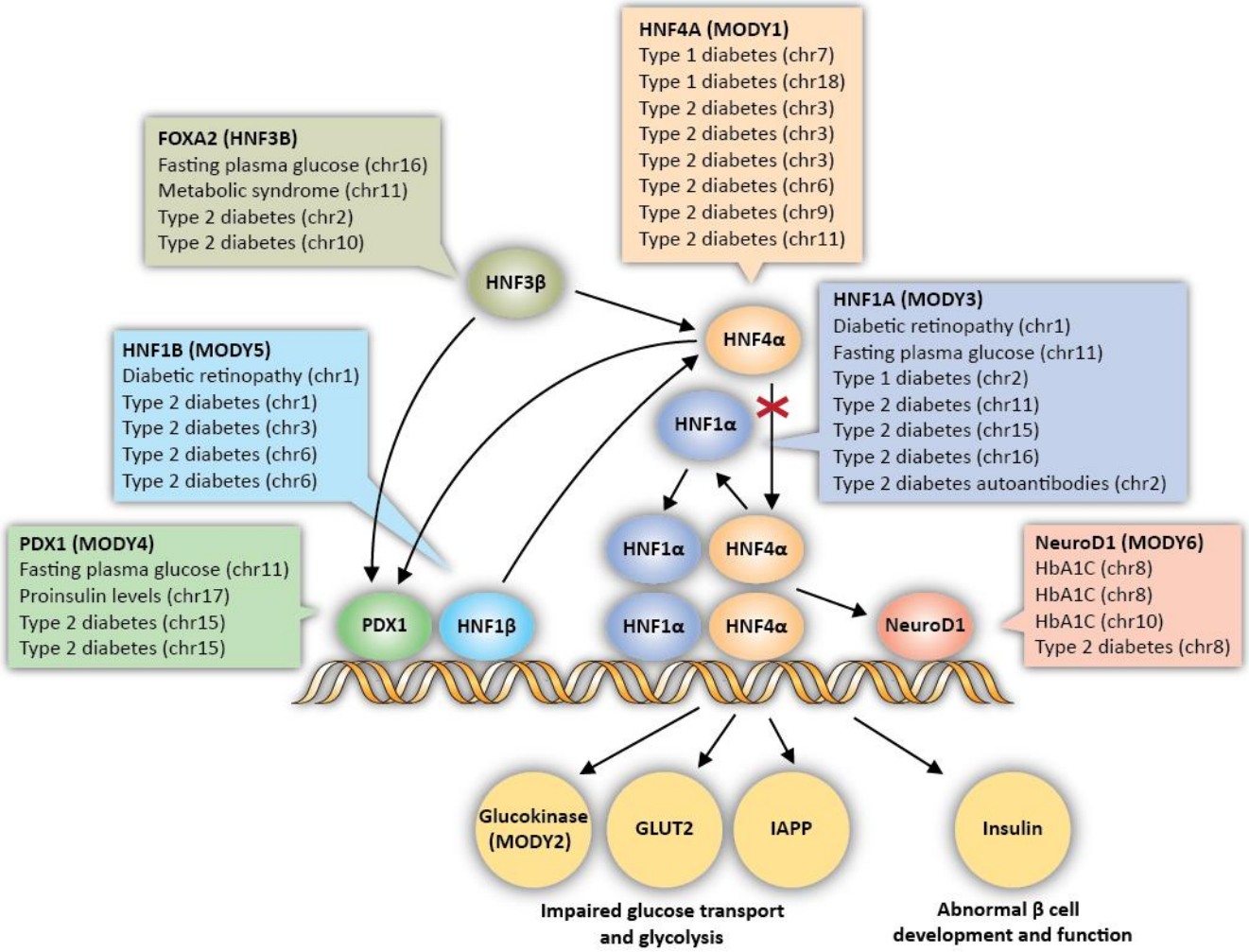
# Targets of transcription factor in promoters and enhancers







# Disease variants cluster in regulatory pathways



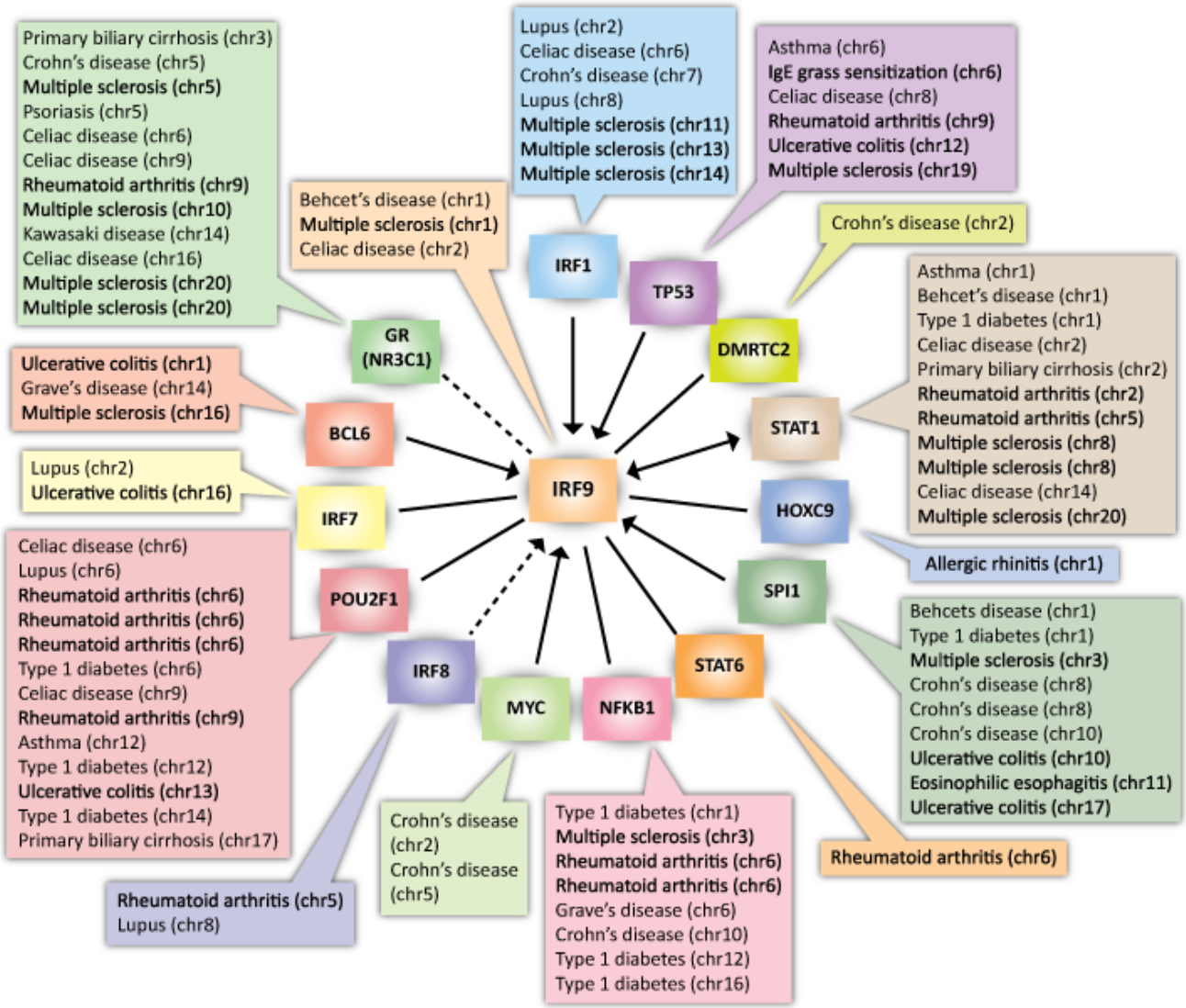
## Mendelian


## Common



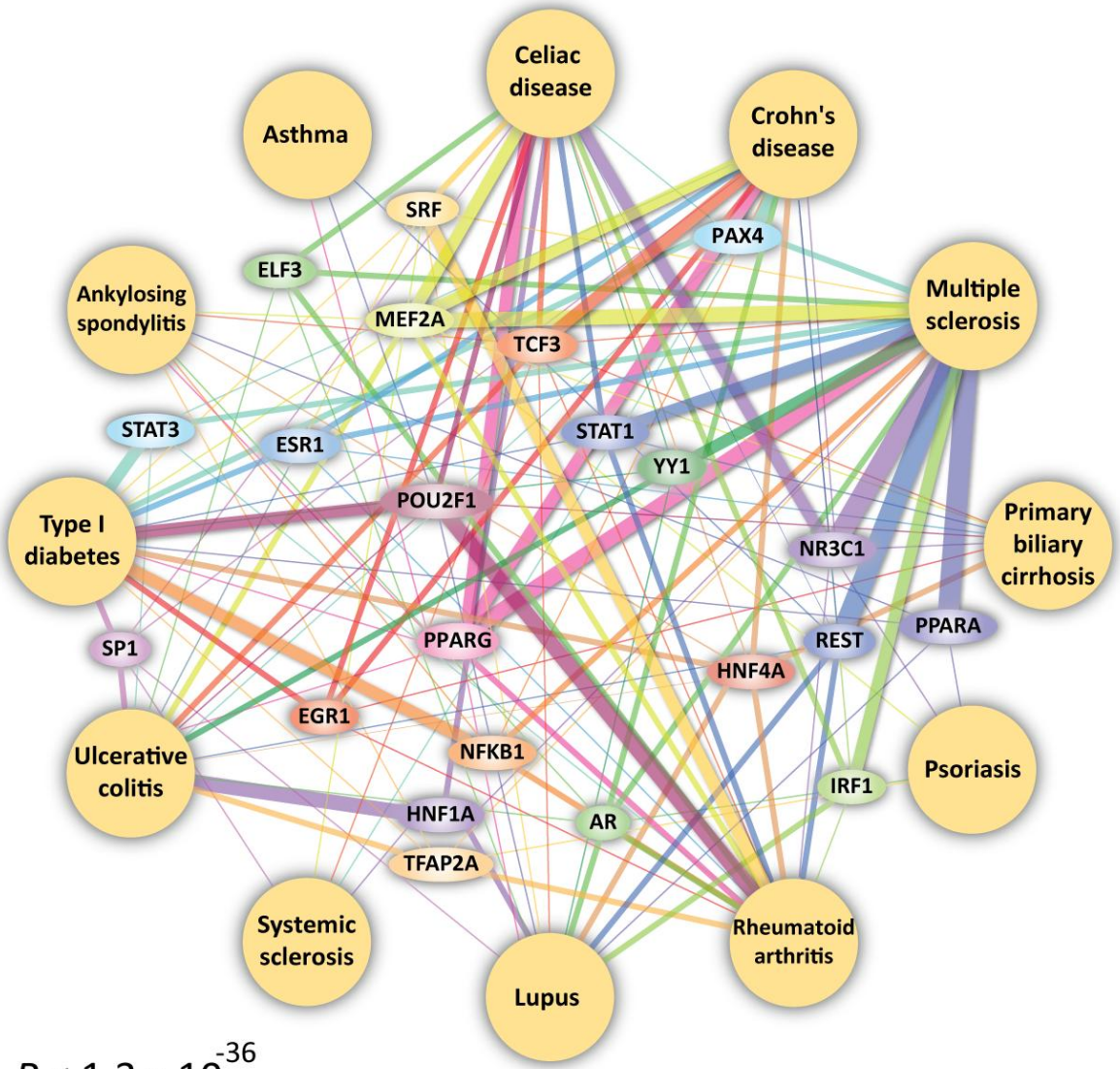



# ~25% of rGWAS inflammatory disease variants lie in IRF9 pathway




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

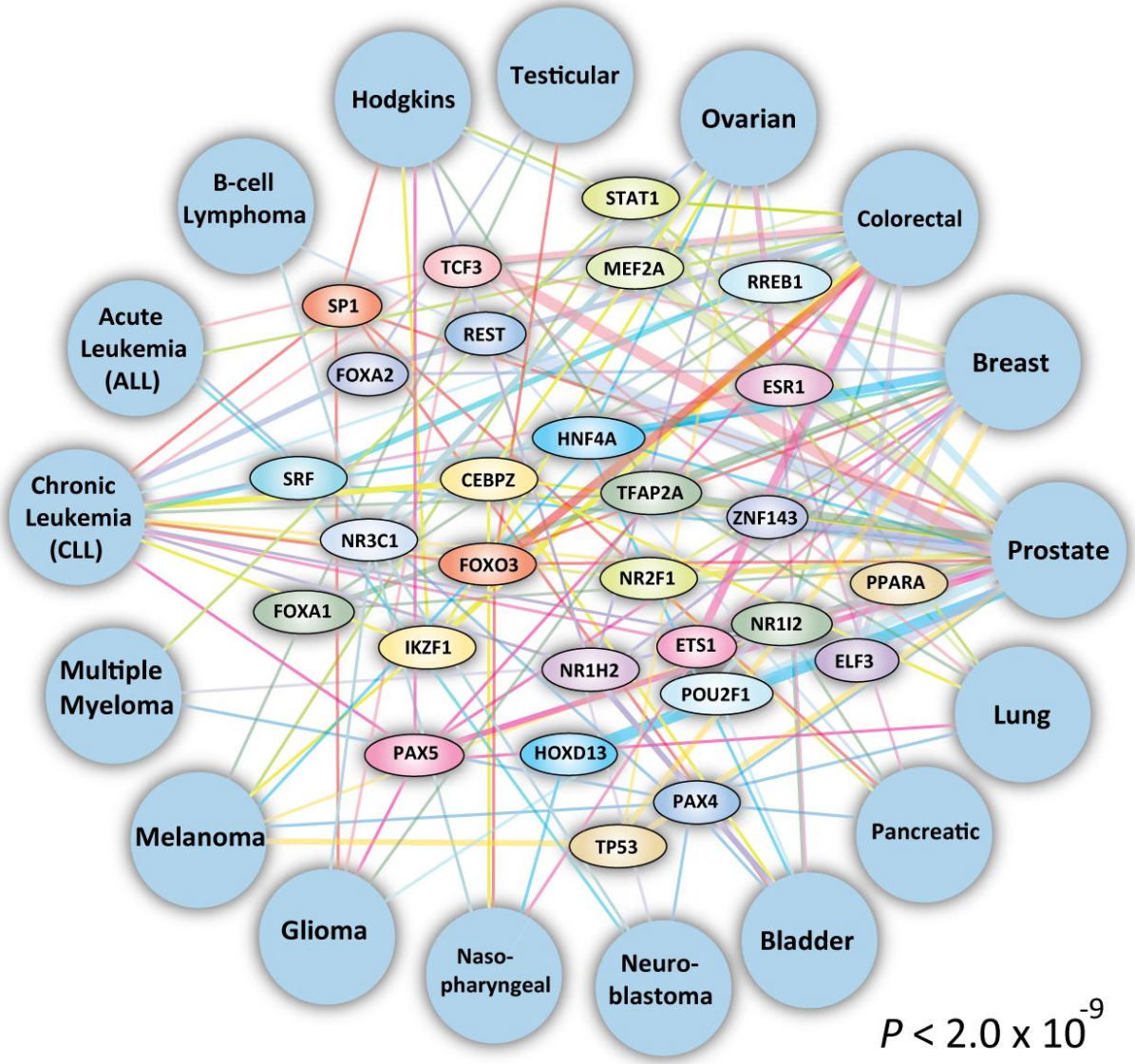
# A common regulatory network for autoimmune disease




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

*Maurano et al., Science 2012*

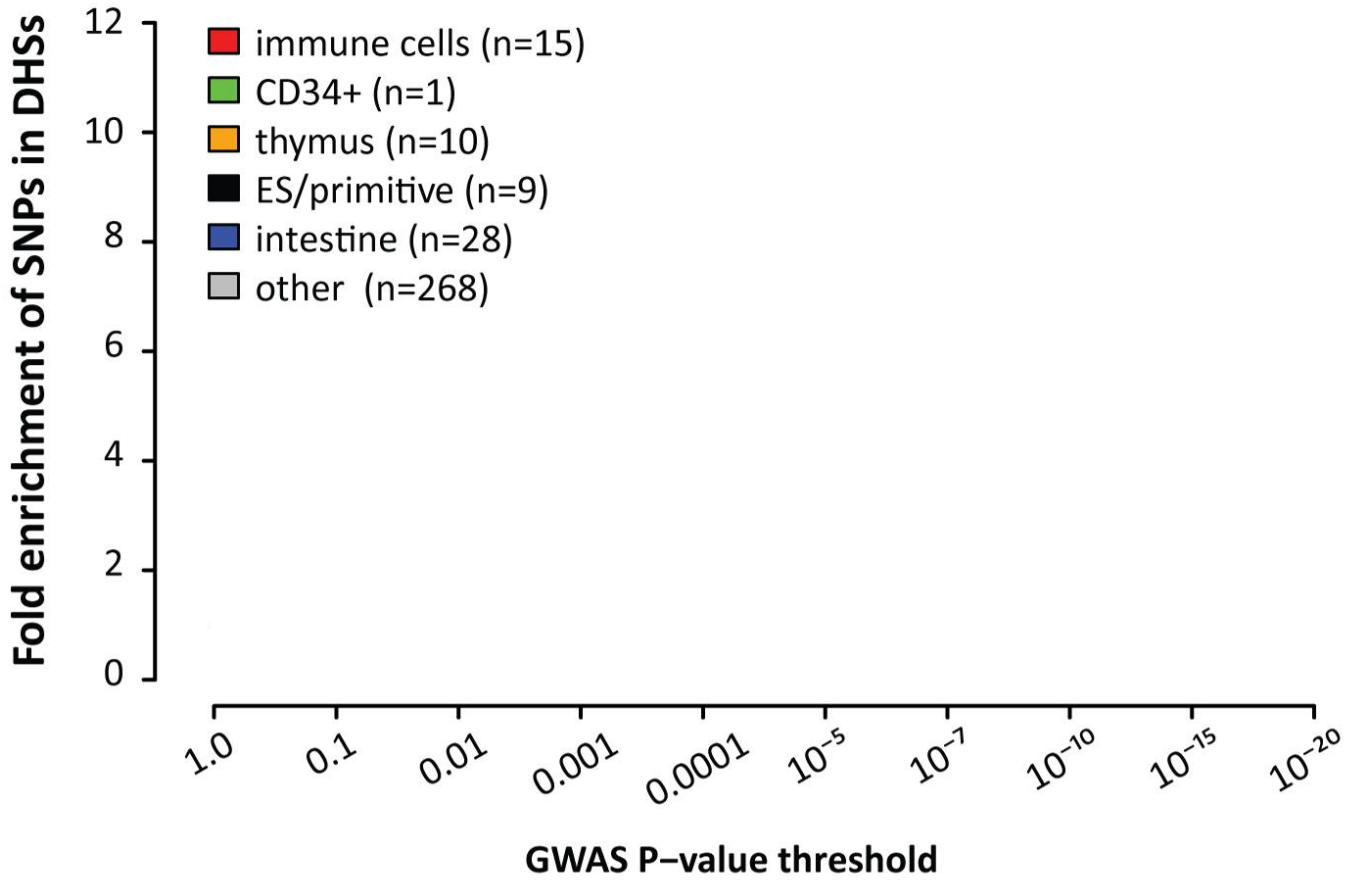
# A common regulatory network underlies diverse malignancies



#7

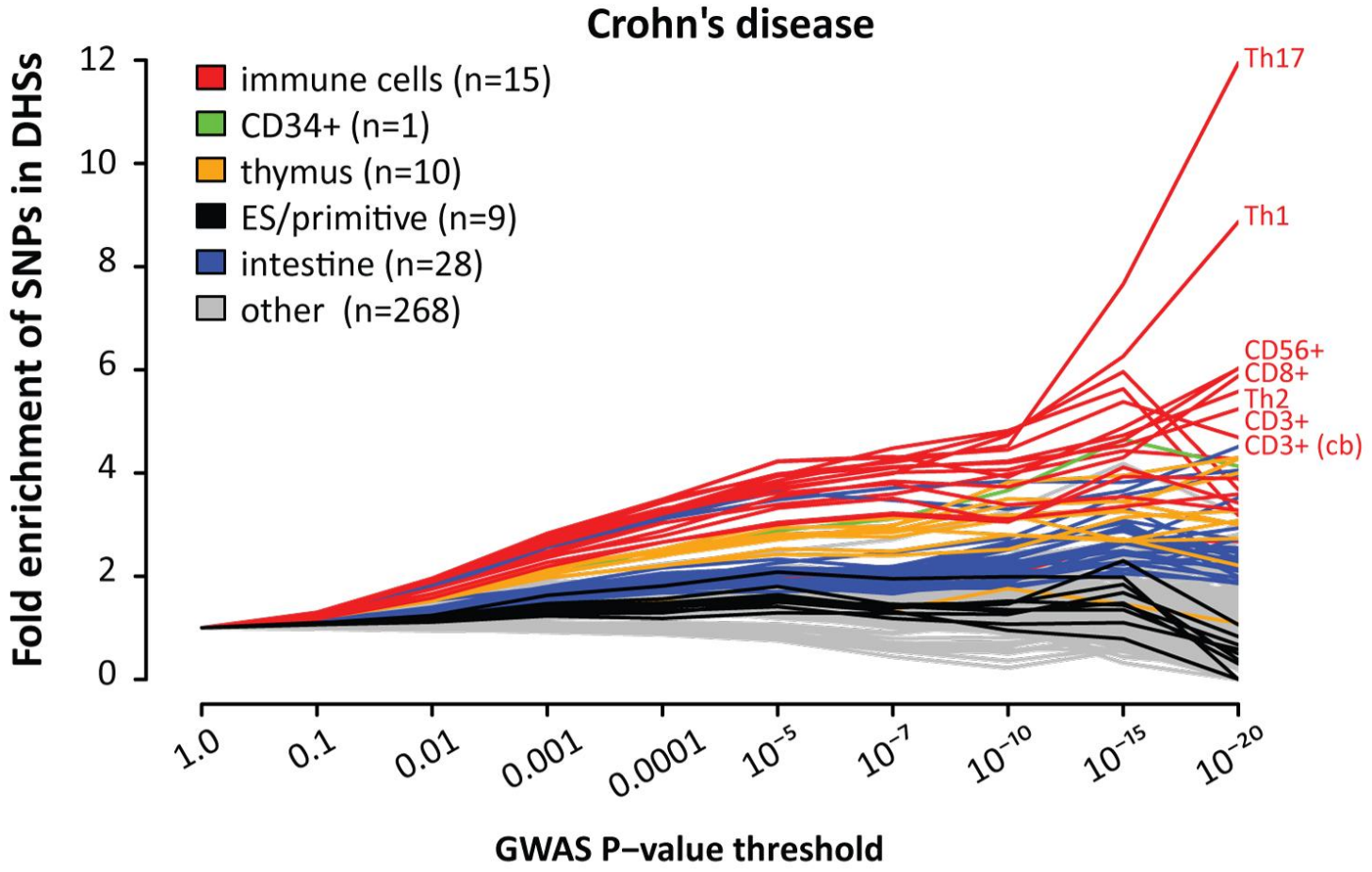
Many, many more variants show these effects than so-called genome-wide significant SNPs

# Selective enrichment of GWAS variants in pathogenic cell types



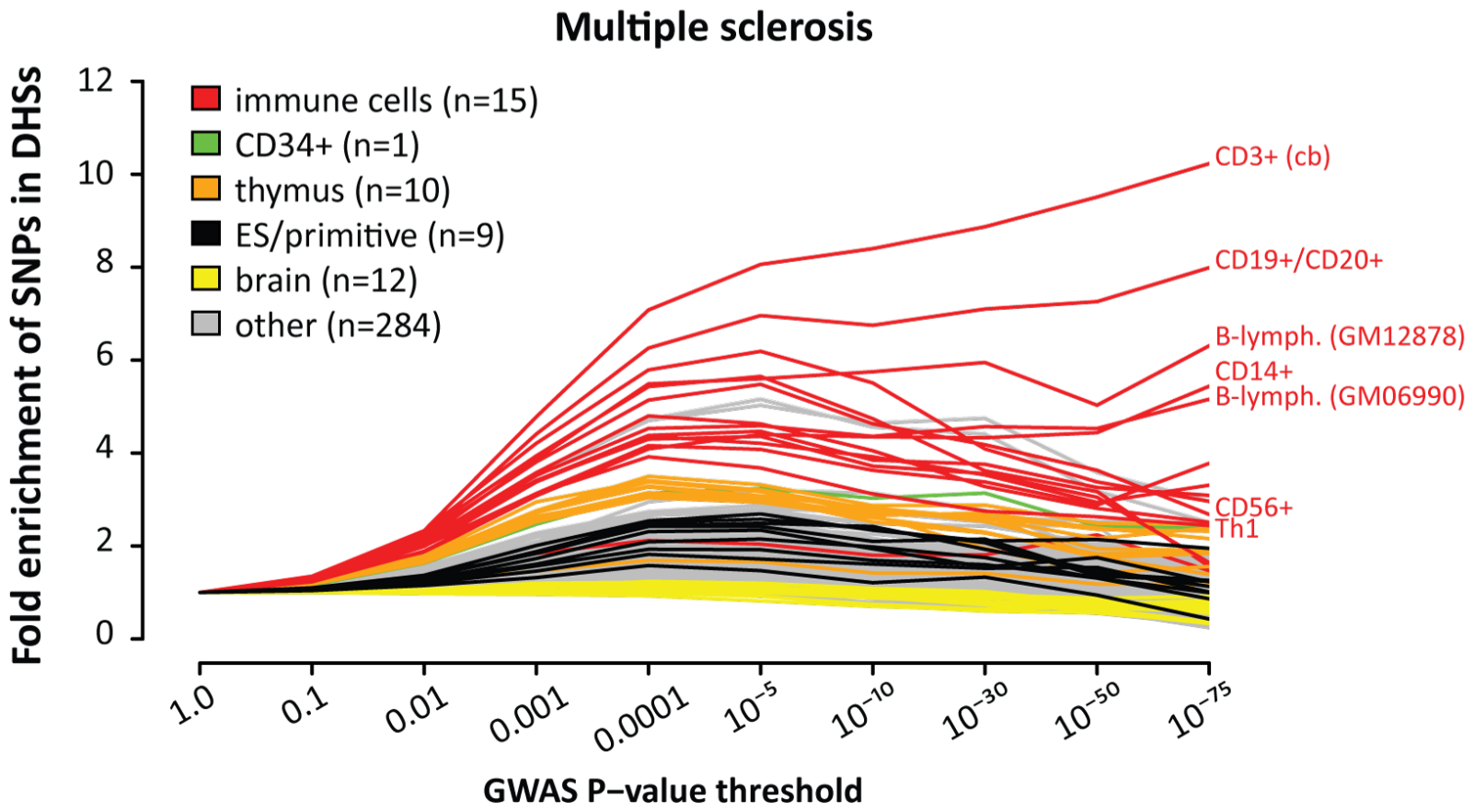

*Maurano et al., Science 2012*

# Selective enrichment of GWAS variants in pathogenic cell types



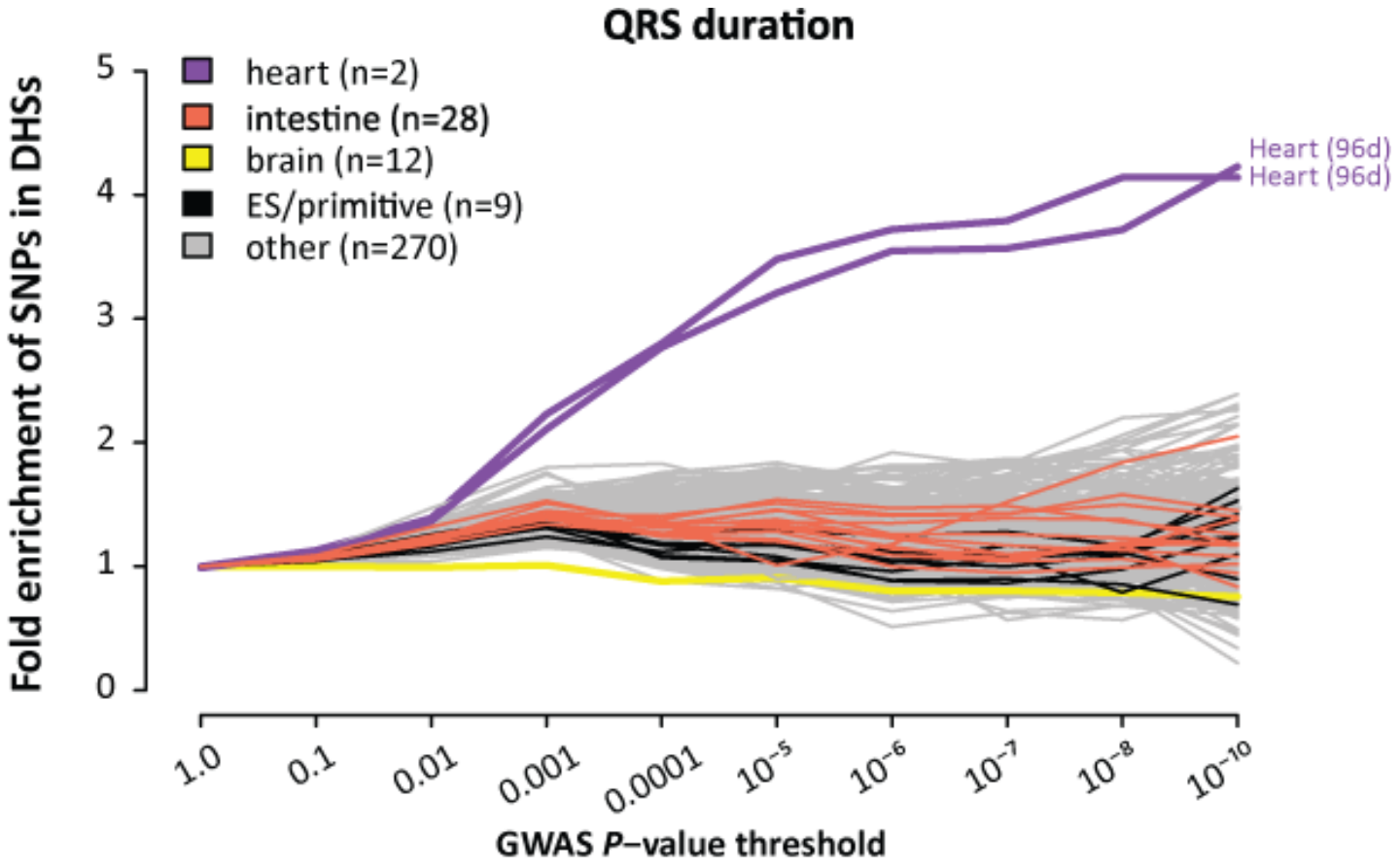

*Maurano et al., Science 2012*

# Selective enrichment of GWAS variants in pathogenic cell types




*Maurano et al., Science 2012*

# Selective enrichment of GWAS variants in pathogenic cell types




*Maurano et al., Science 2012*

# Summary & Implications

Systematic localization of disease-associated variation in regulatory DNA
→*Enables a coherent approach to understanding the role of non-coding variants*

Significant majority of variants localize in fetal-stage regulatory DNA
→ *Developmental contribution to common adult disorders and traits*

Common networks connect disorders within broader disease classes
→*Shared genetic liability at the level of transcription factor networks*

Many GWAS data are untapped
→*Many additional disease-associated variants that do not meet "genome-wide significance" also localize within regulatory DNA*

Deep maps of regulatory DNA in *normal* cells enable pathogenic insights
→*Pathogenic cell types for many disorders are not well-defined*

# Acknowledgements


Analysis

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

Data production

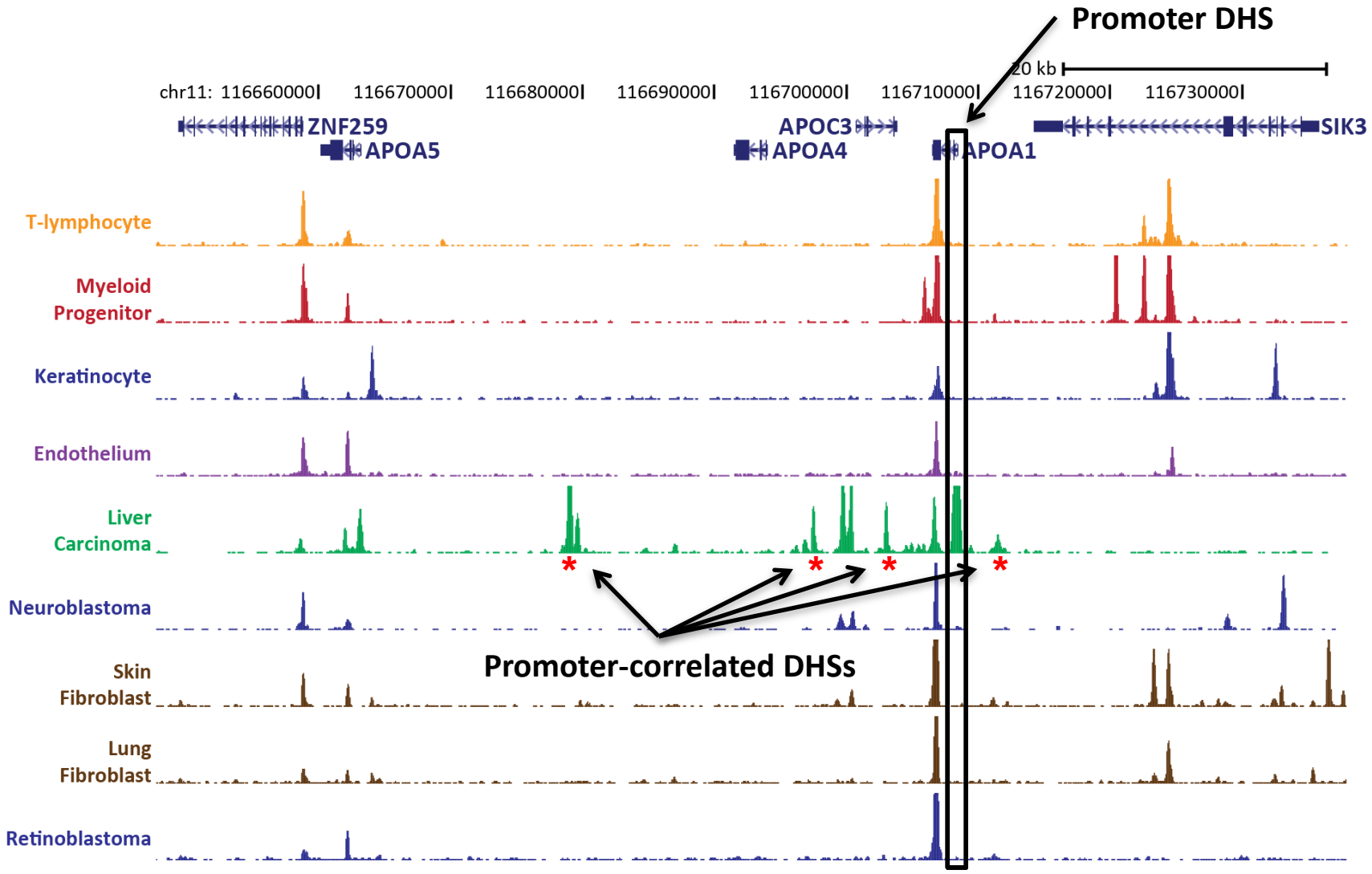
Raj Kaul, Scott Hansen, Peter Sabo, Molly Weaver, Theresa Cantwell, Kristin Lee, Shinny Vong, Vaughan Roach, Erica Gist, Sandra Stehling-Sun

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)**

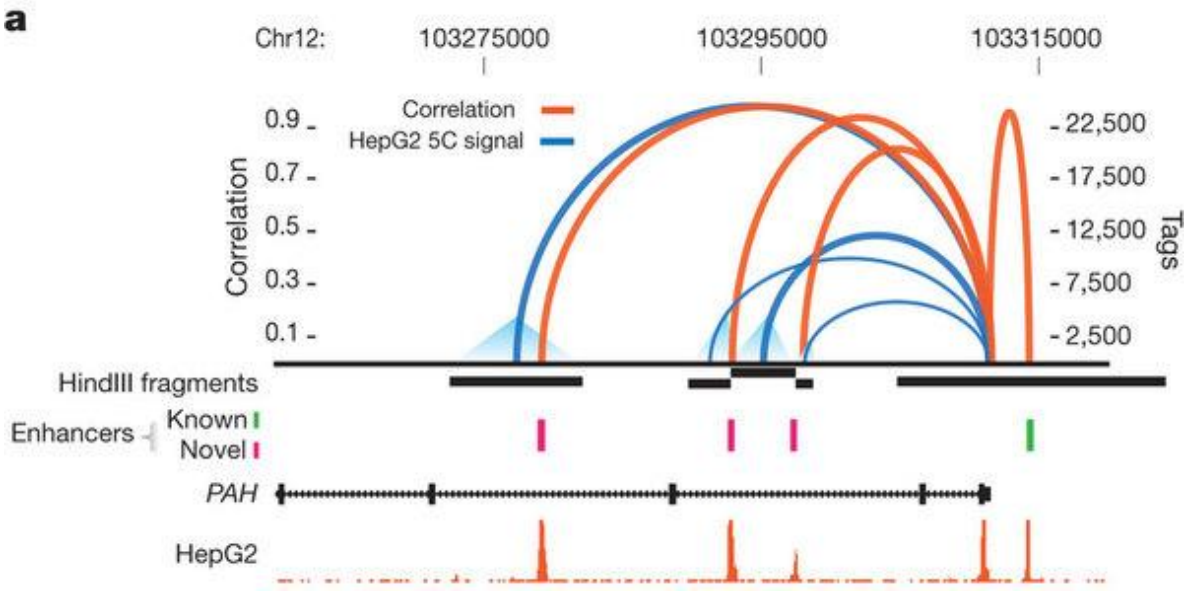
# Linking regulatory DNA with its cognate gene(s)

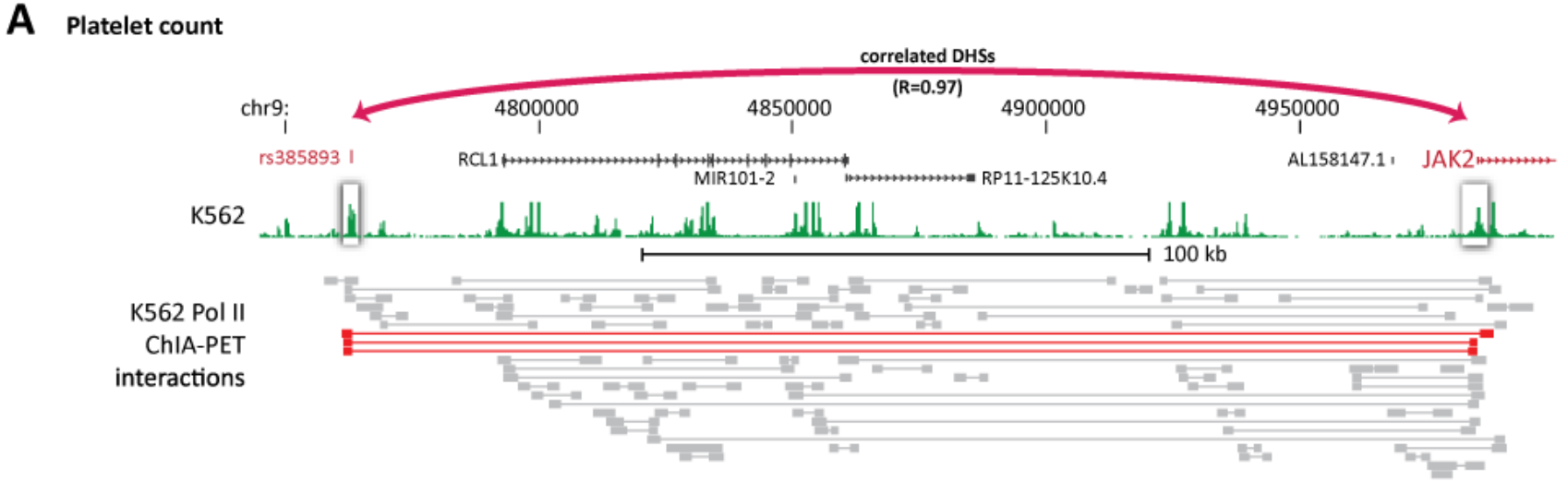



**Applied to 1,524,865 DHSs across 79 cell types (DHSs+/-500kb)**

**580,000 co-regulated DHSs**

# Correlated DHS-promoter pairs physically interact




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

# Disease/trait-associated variants cluster in physiologically relevant transcription factor binding sites