

# The regulatory landscape of genes, traits & disease or: *using DNase-seq data as a scaffold for integrative analyses*

Wouter Meuleman, PhD

Investigator

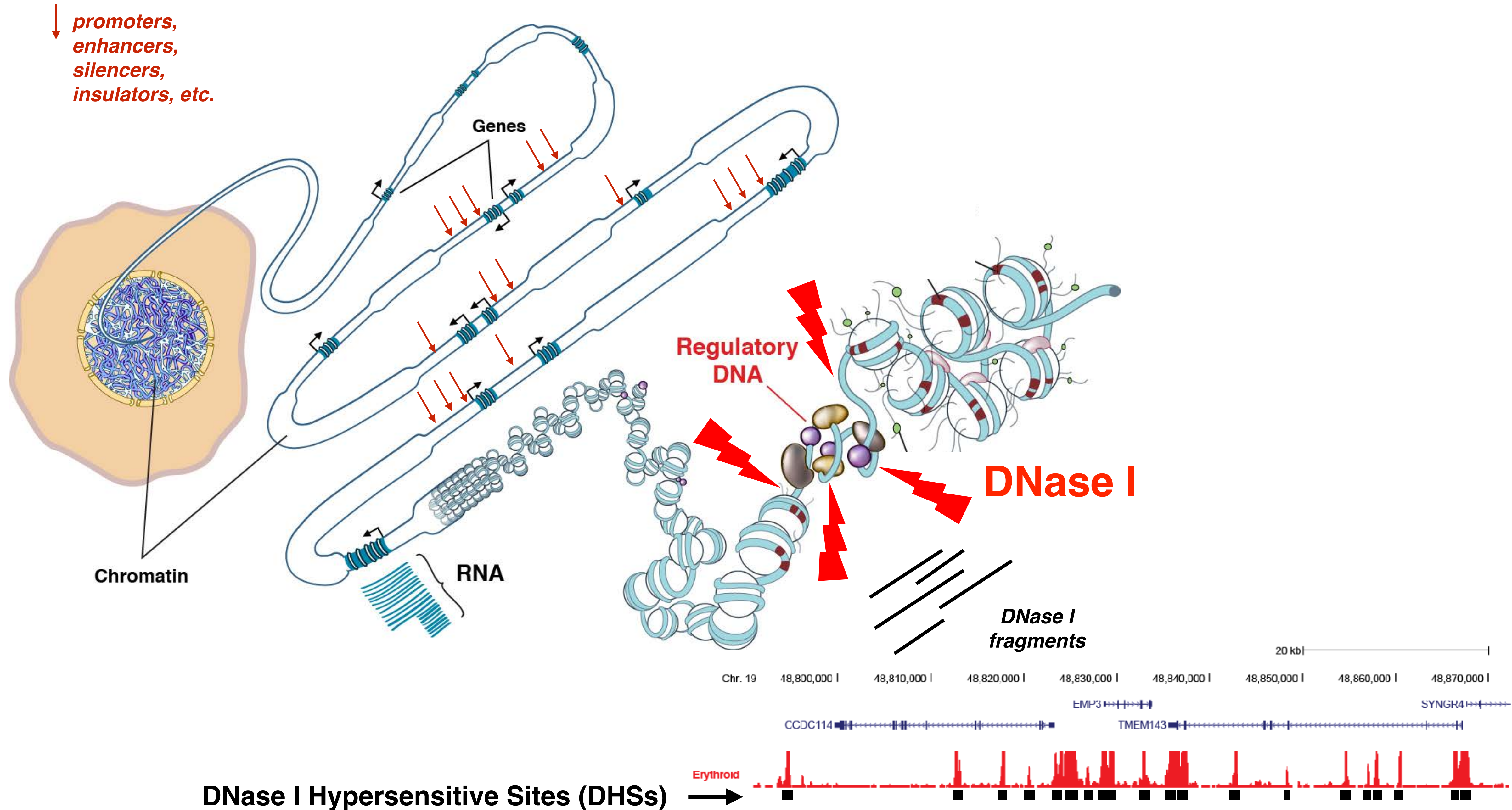
Altius Institute for Biomedical Sciences  
Seattle, WA USA



@nameluem

[www.meuleman.org](http://www.meuleman.org)

# The regulatory genome can be mapped using DNase I digestion



# Challenges and opportunities for integrative data analysis

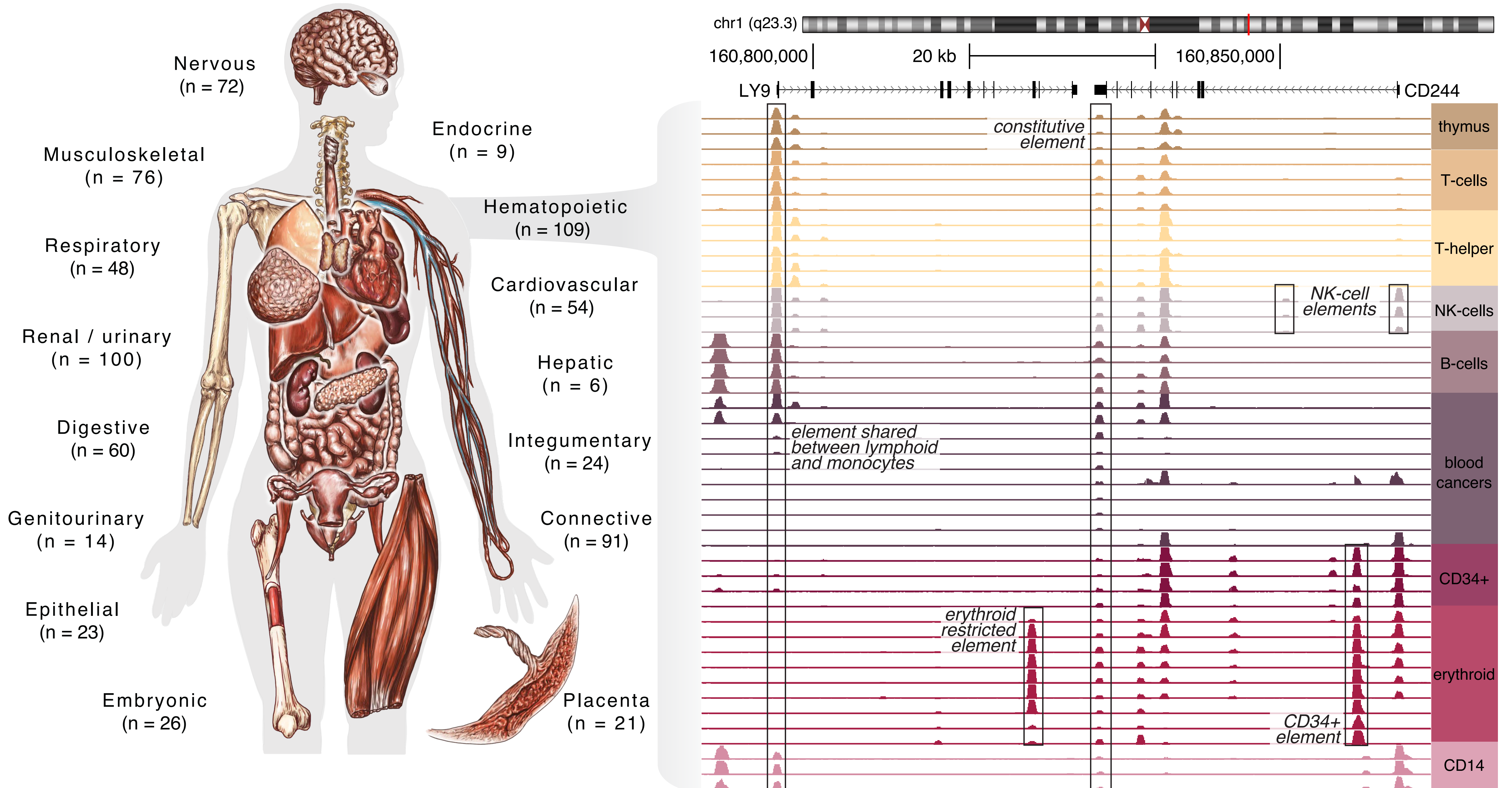


A challenge in integrative analyses is the lack of ***common coordinate systems*** and ***pragmatic annotations***

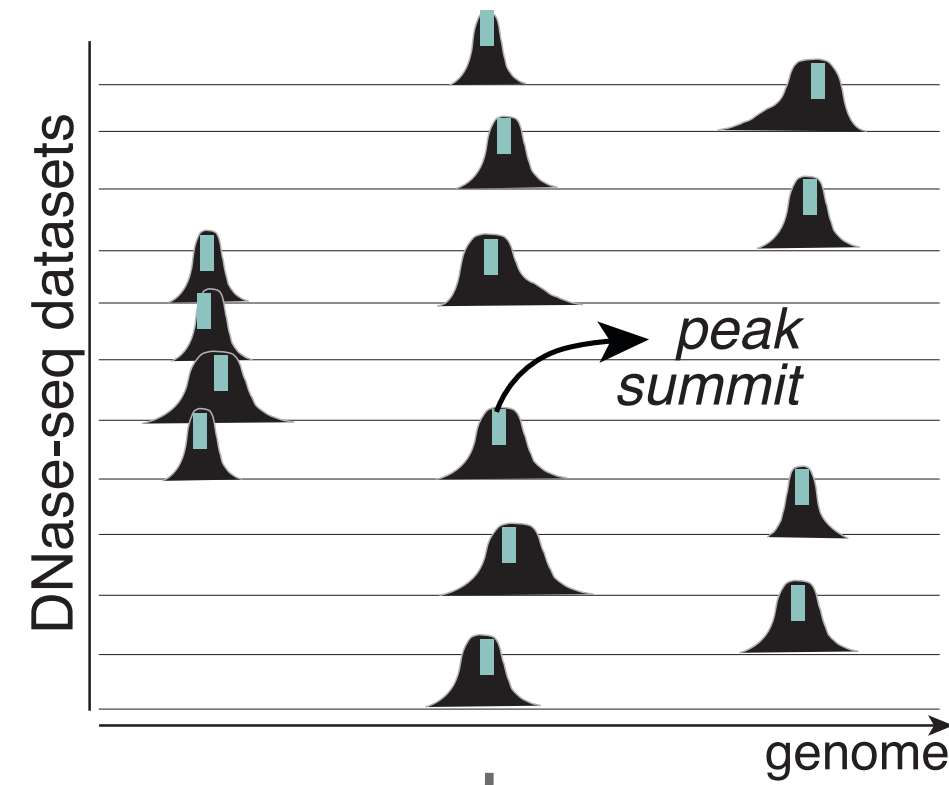
High quality bulk reference maps may anchor downstream analyses and provide scaffolds for complementary assay data

Towards comprehensively annotated maps

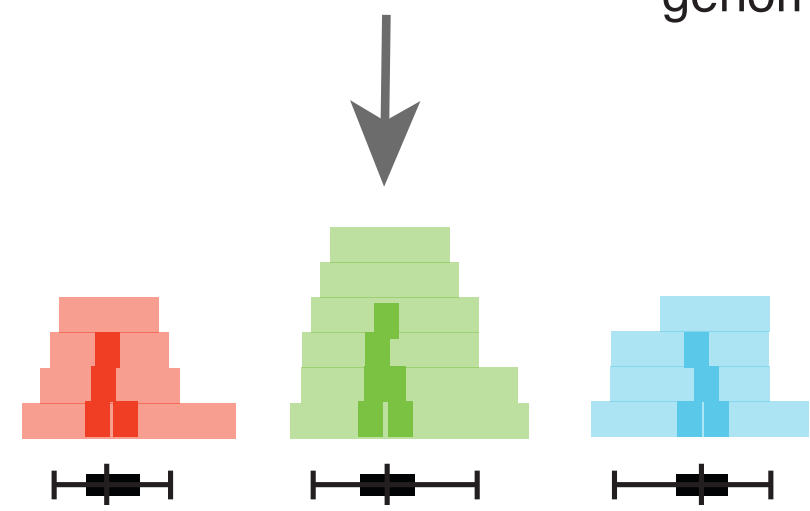
# A survey of chromatin accessibility across 400+ cell types and states



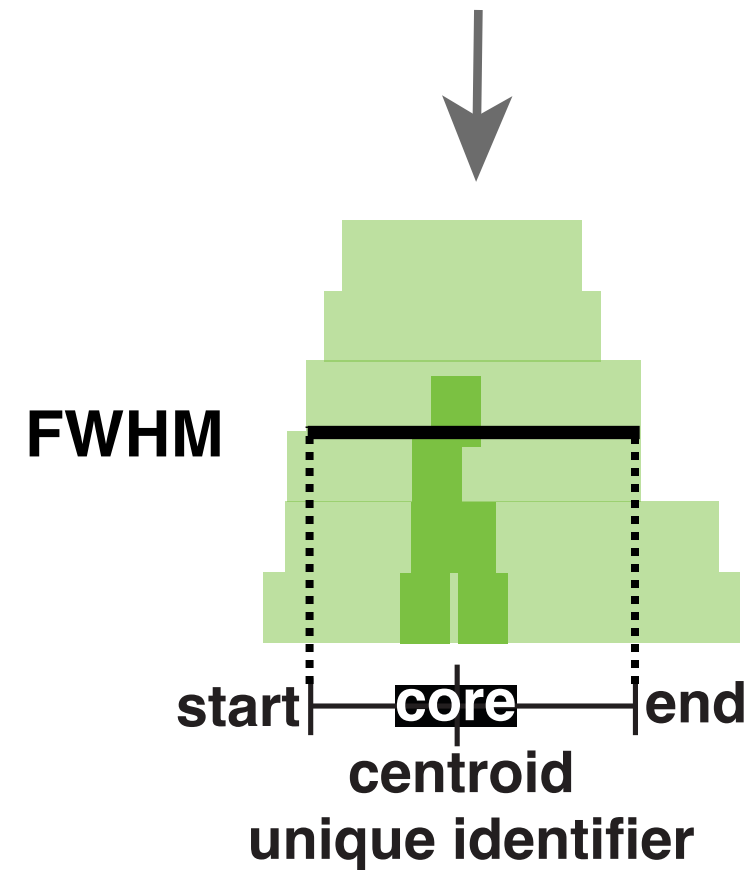
# A common coordinate system of DHSs



76.5 million  
DHS locations  
and summits



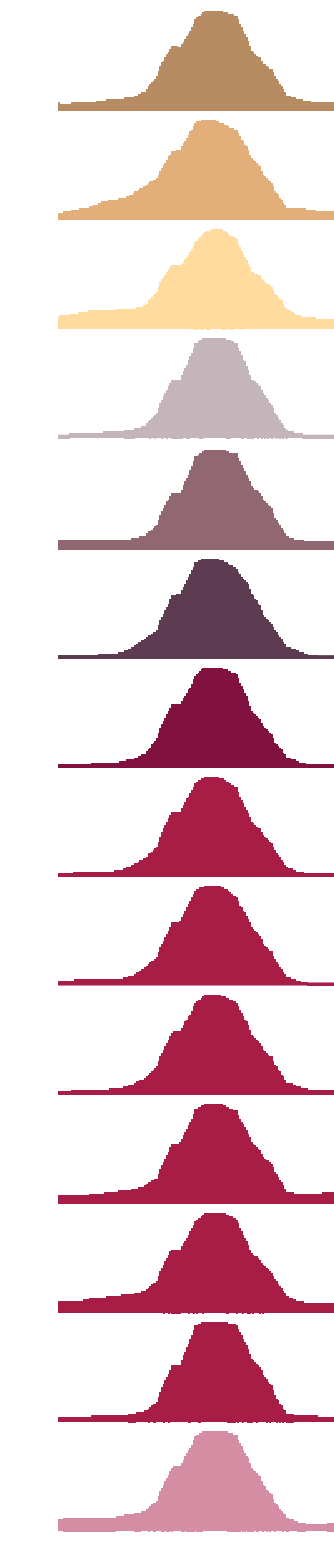
3.59 million  
DHS archetypes



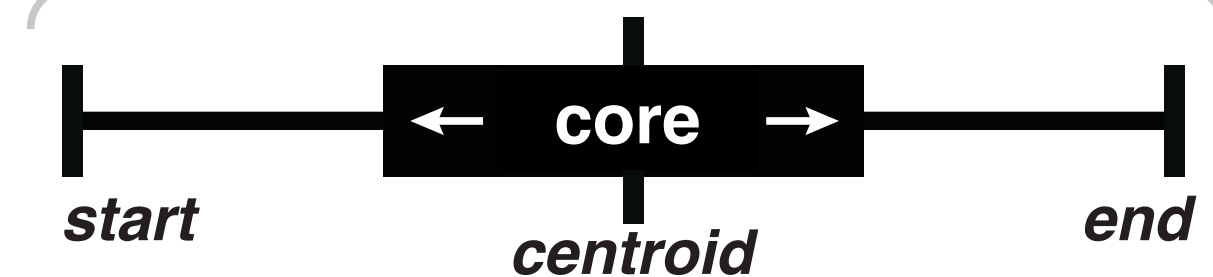
Annotated Index  
of DHS archetypes

- Thymus
- CD3+
- T2 helper cells, induced
- CD56+
- CD20+
- OCI Ly7
- CD34+
- CD34 (day 4)
- CD34 (day 6)
- CD34 (day 8)
- CD34 (day 15)
- CD34 (day 17)
- CD34 (day 18)
- CD14+

constitutive,  
well-positioned



14.882878

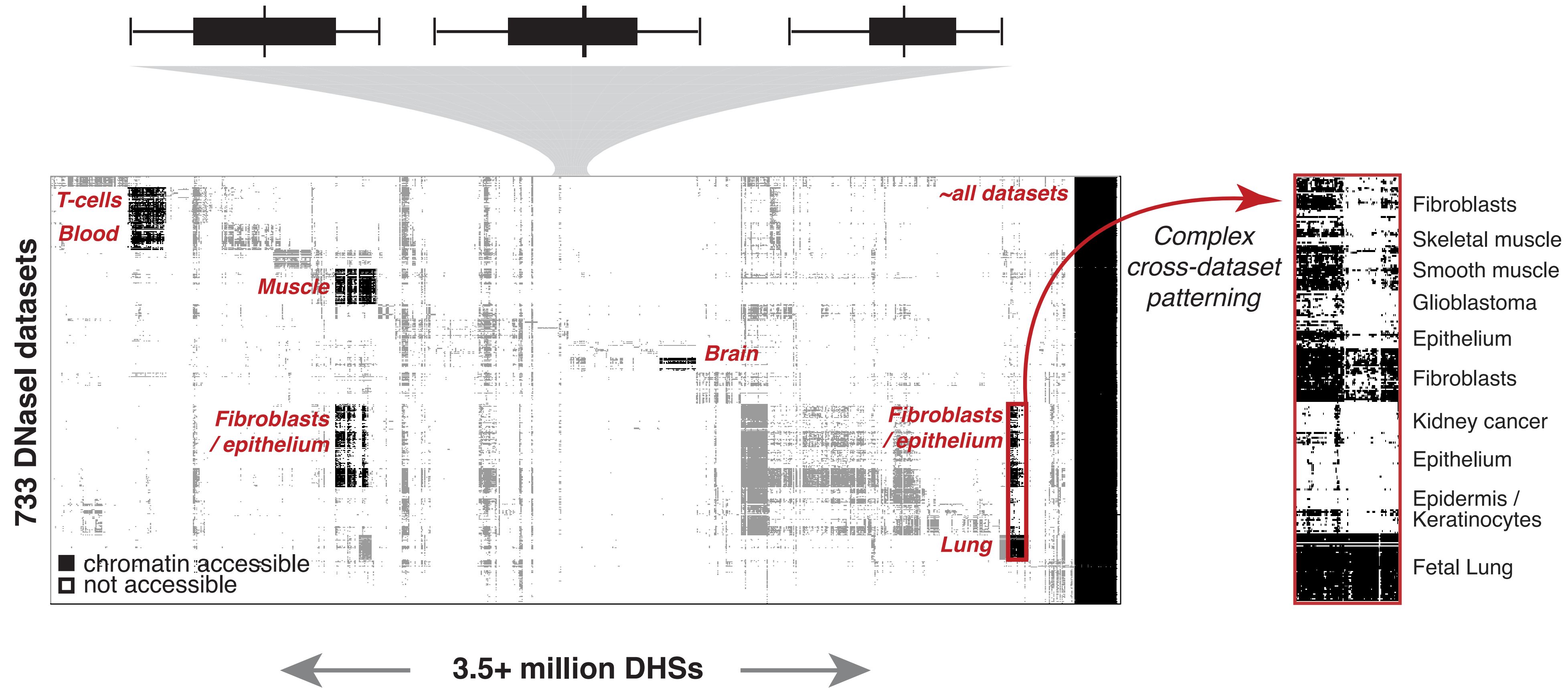


14.882878

CD34+  
element

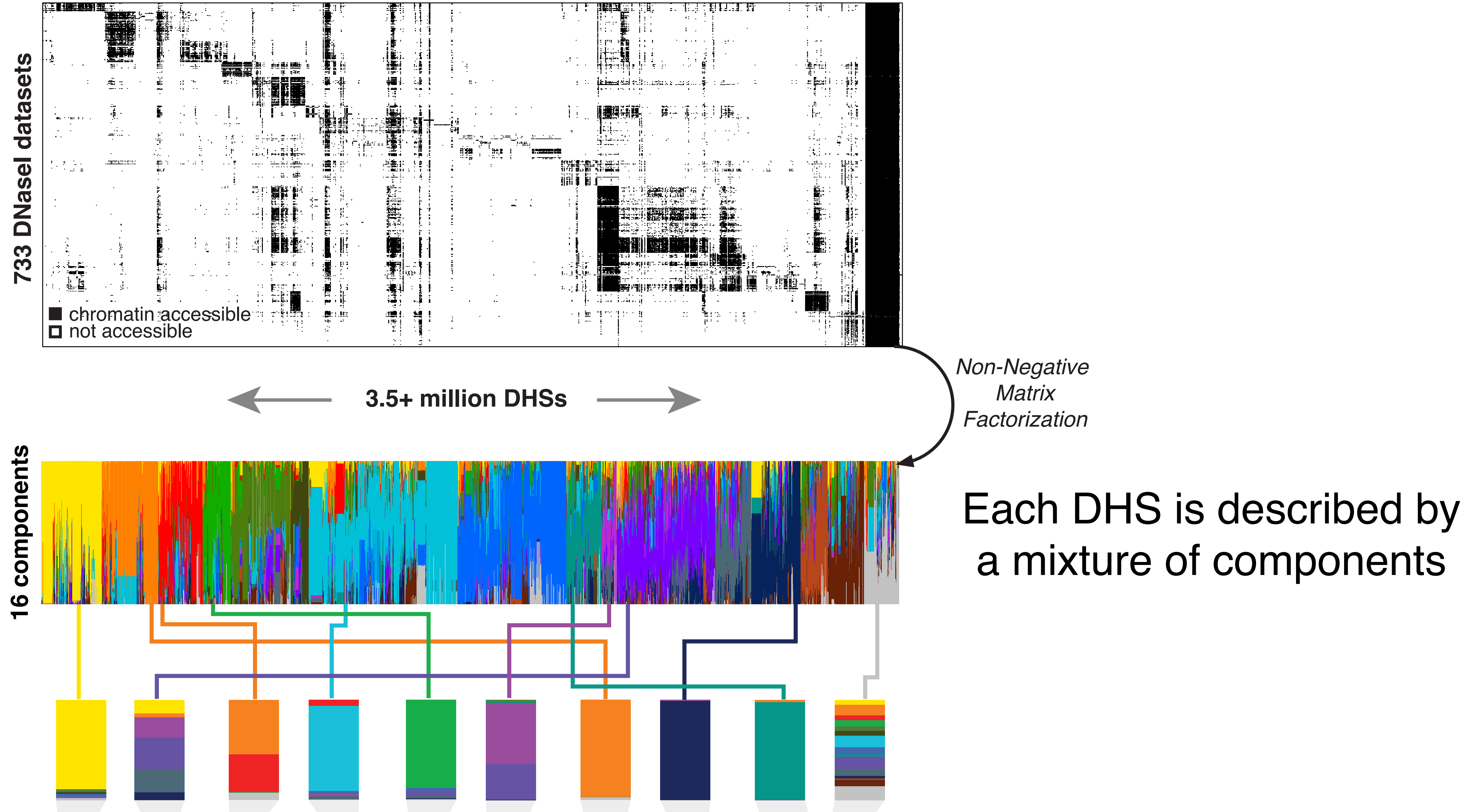
CD14

# Complex DHS patterning across cell types and states



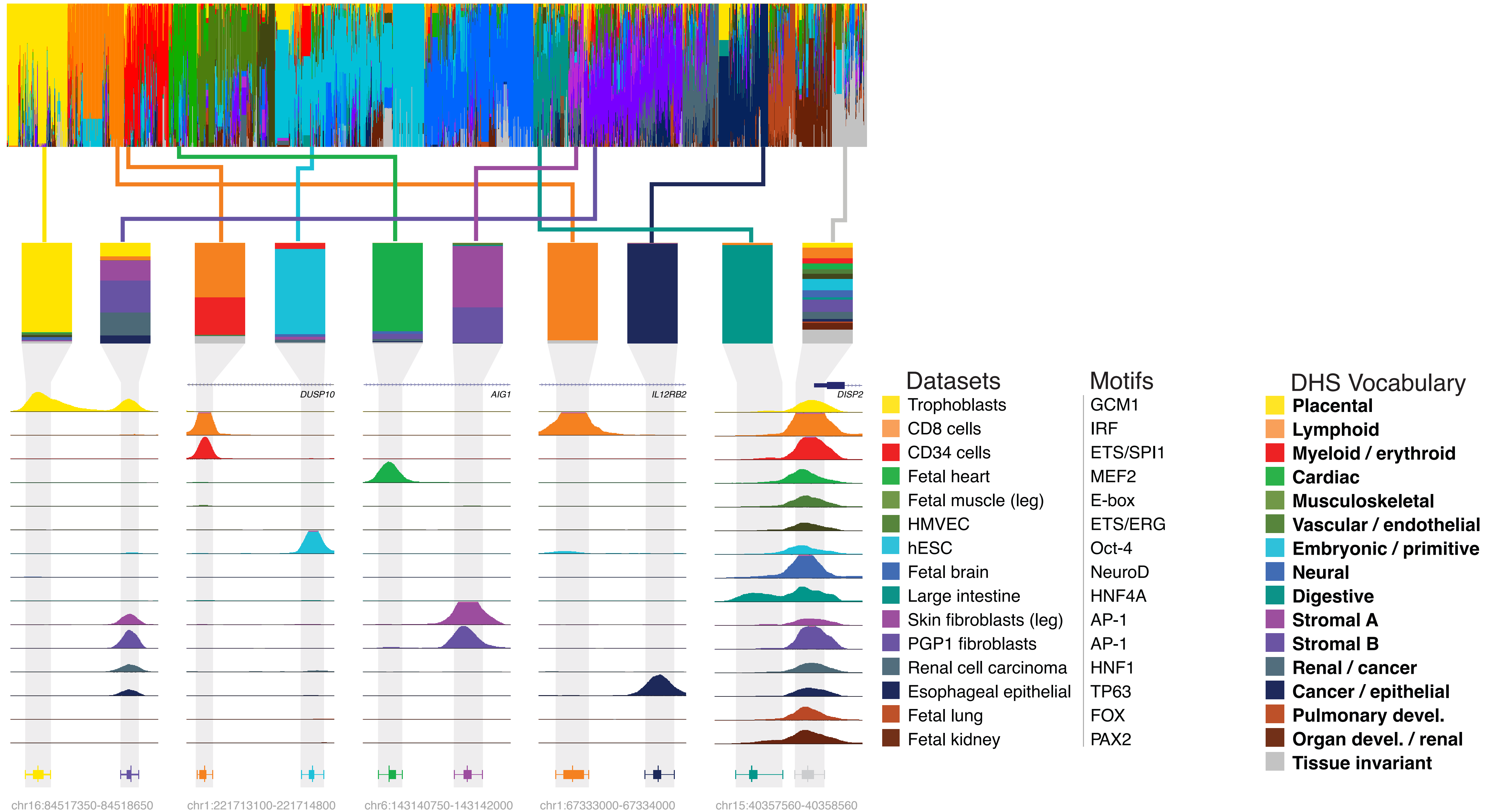
Although generally quite cell type/state selective, DHSs are often shared between broader cellular contexts

# DHS patterns can be decomposed into *components*



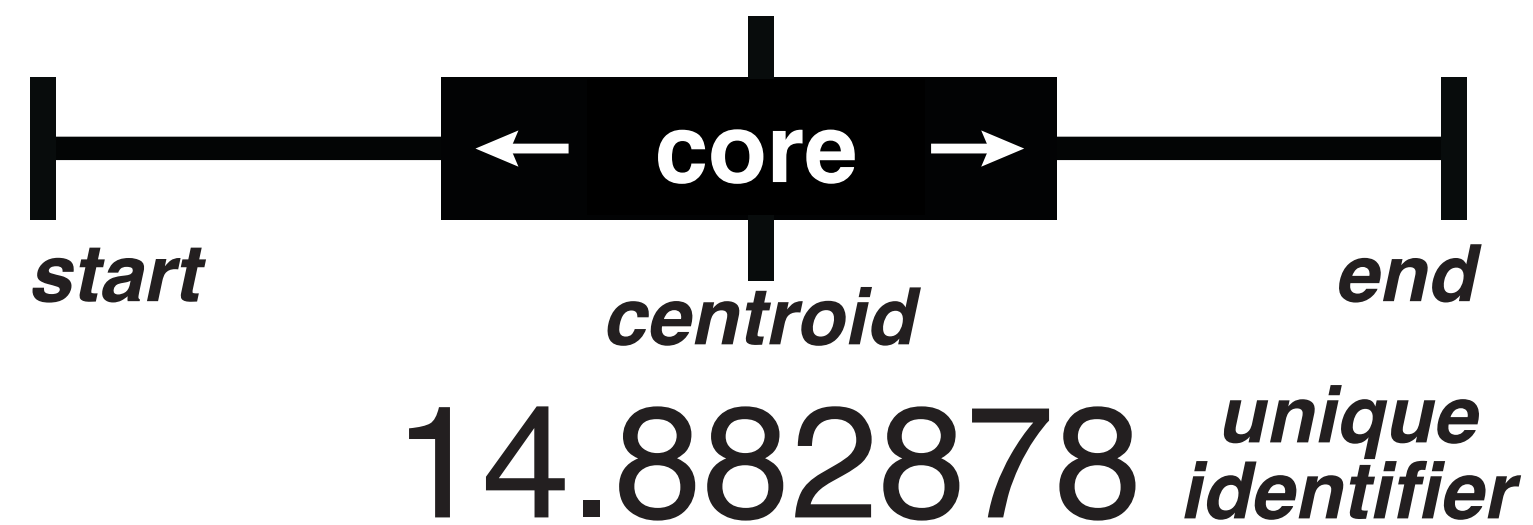


# Components reflect distinct biological contexts and regulatory signal



# DHS Index and Vocabulary: a novel annotation of regulatory DNA

## DHS Index



3.5M+ DHSs

Richly annotated and  
indexed across cell types

## DHS Vocabulary

- Placental
- Lymphoid
- Myeloid / erythroid
- Cardiac
- Musculoskeletal
- Vascular / endothelial
- Embryonic / primitive
- Neural
- Digestive
- Stromal A
- Stromal B
- Renal / cancer
- Cancer / epithelial
- Pulmonary devel.
- Organ devel. / renal
- Tissue invariant

# index.altius.org

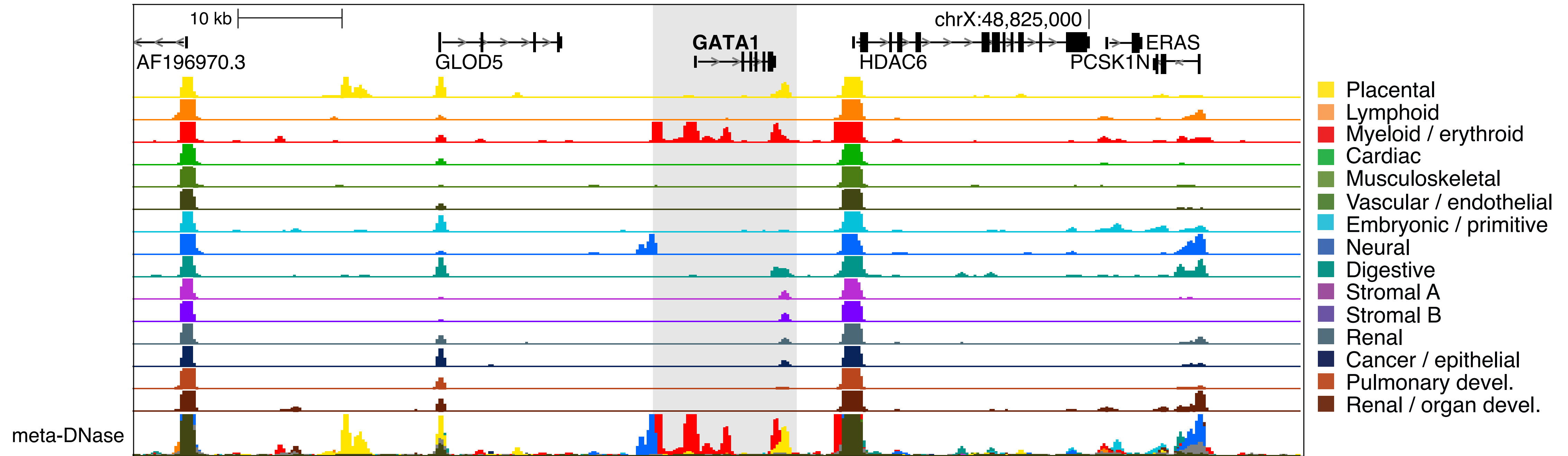
Go

I'm Feeling Lucky

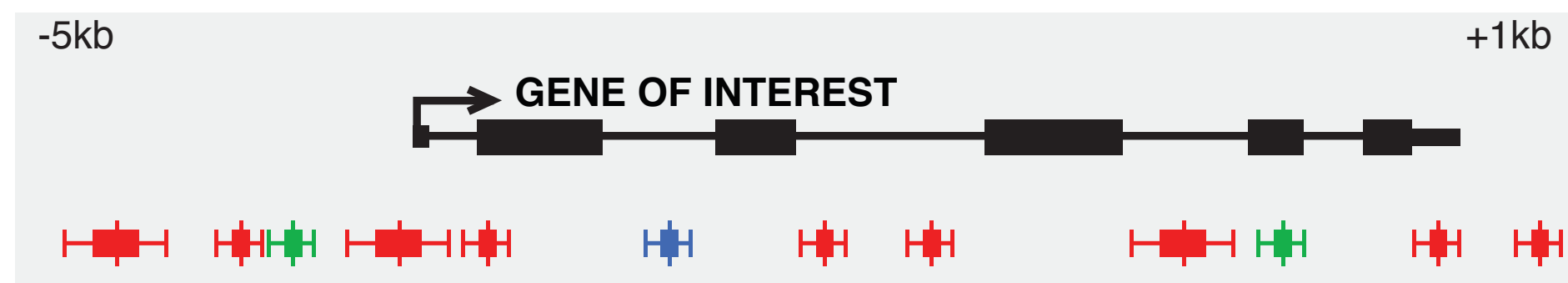
e.g., use query terms like HGNC symbols (HOXA1, NFKB1, etc.) or genomic regions (chr17:41155790-41317987, etc.)

# The regulatory landscape of genes

# Regions around genes show component-specific patterning



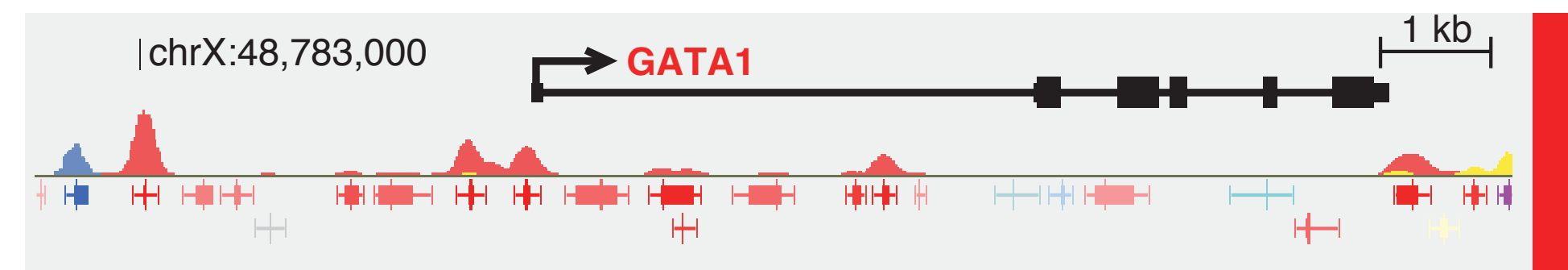
# Genes can be labeled based on their regulatory landscape



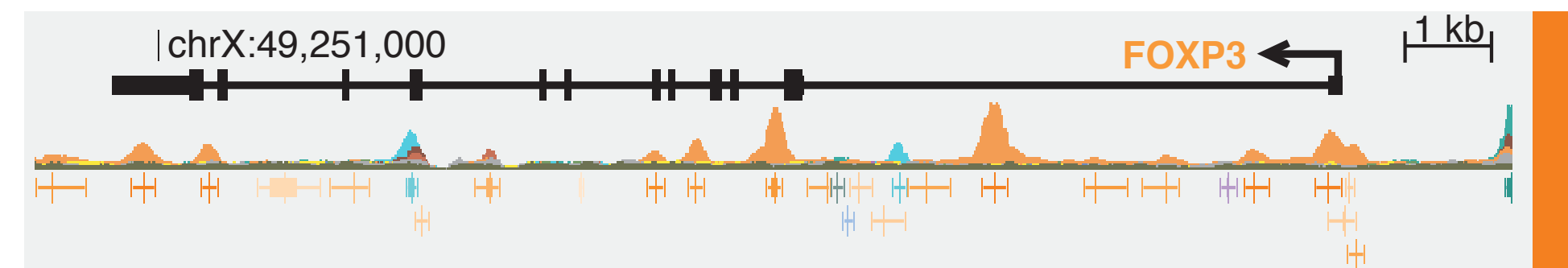
1 Obtain DHSs proximal to gene of interest

2 Test for overrepresentation of Vocabulary components (Fisher's Exact test, FDR 5%)

3 Assign gene to Vocabulary component(s)



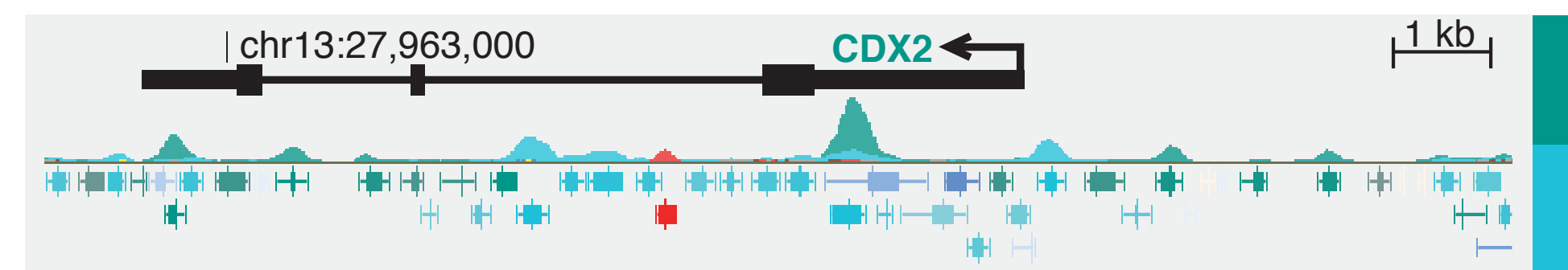
Myeloid / erythroid



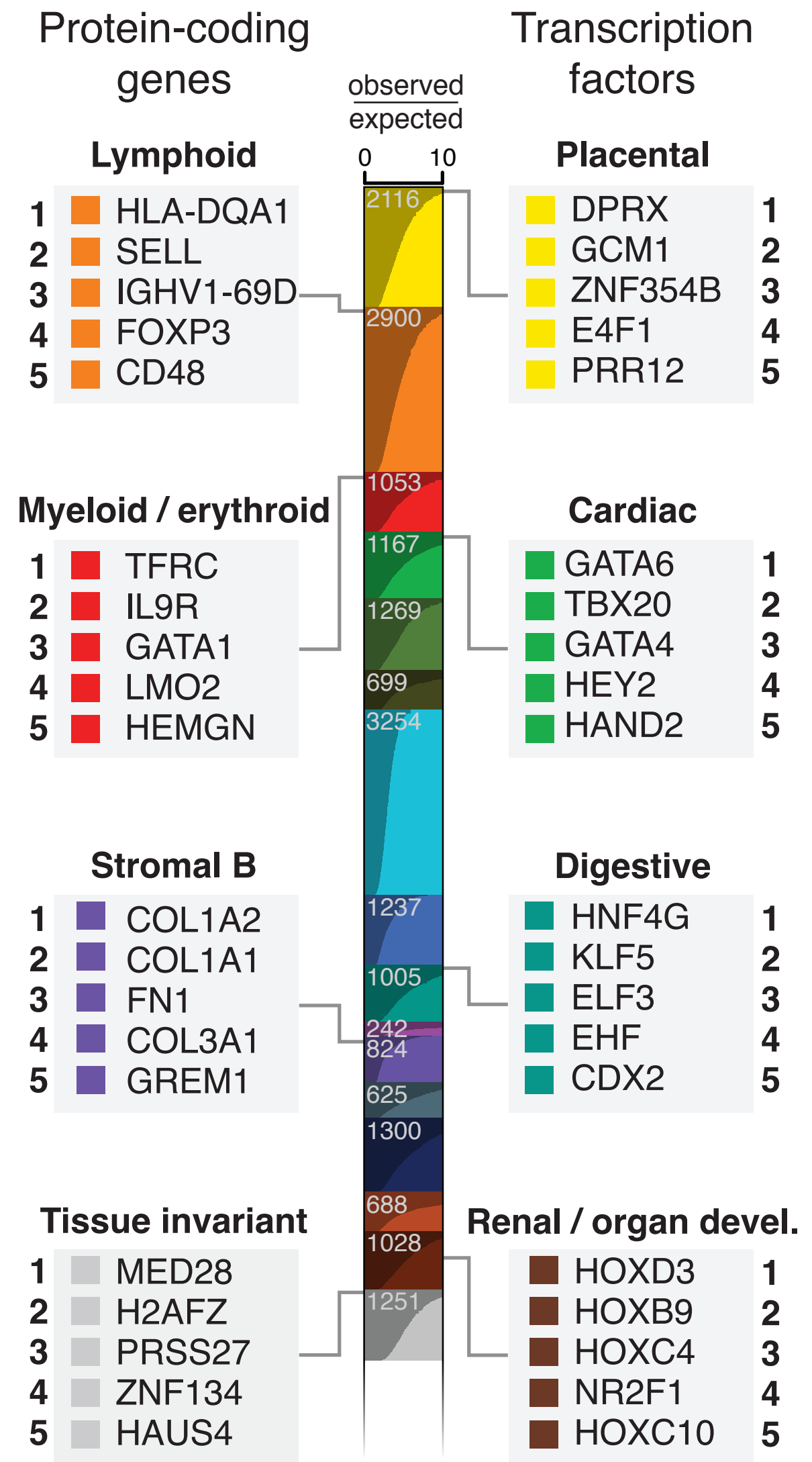
Lymphoid



Organ devel. / renal

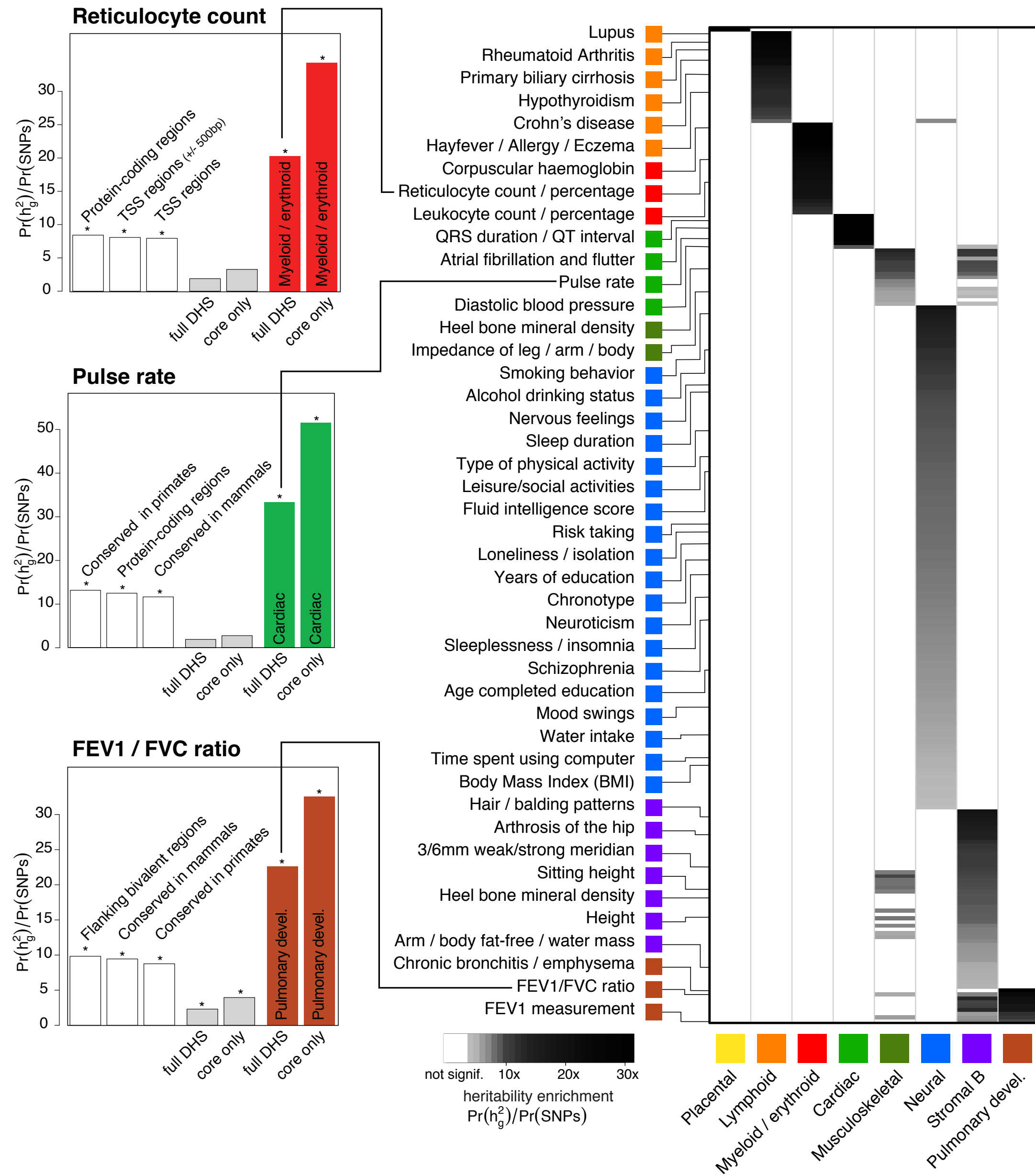


Digestive & Embryonic / primitive



Concentration of trait heritability

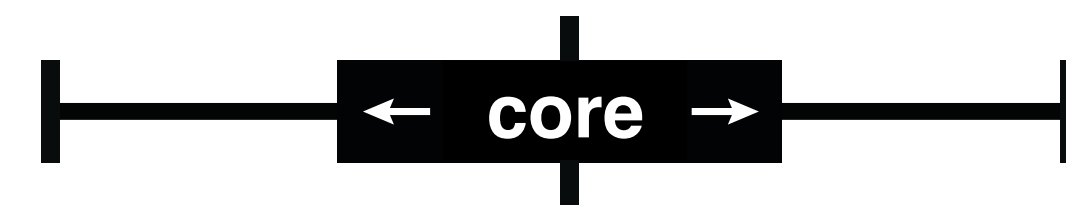
# GWAS signal is strongly enriched in DHS components



Estimate heritability enrichment per genome-wide annotation

DHS components show strong enrichment for heritability signal

particularly in the 'core' regions

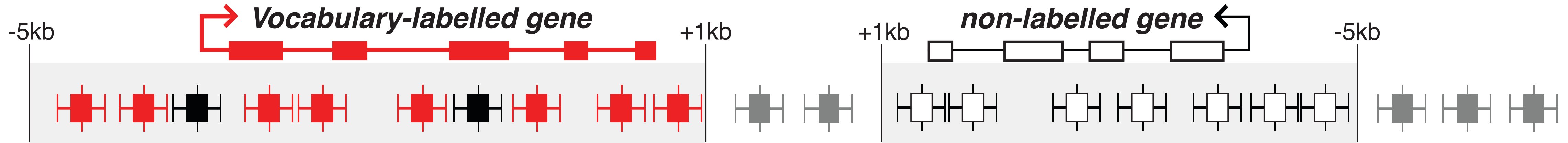


UK Biobank traits

LD score regression with partitioned heritability



# GWAS signal is concentrated in component concordant genic DHSs



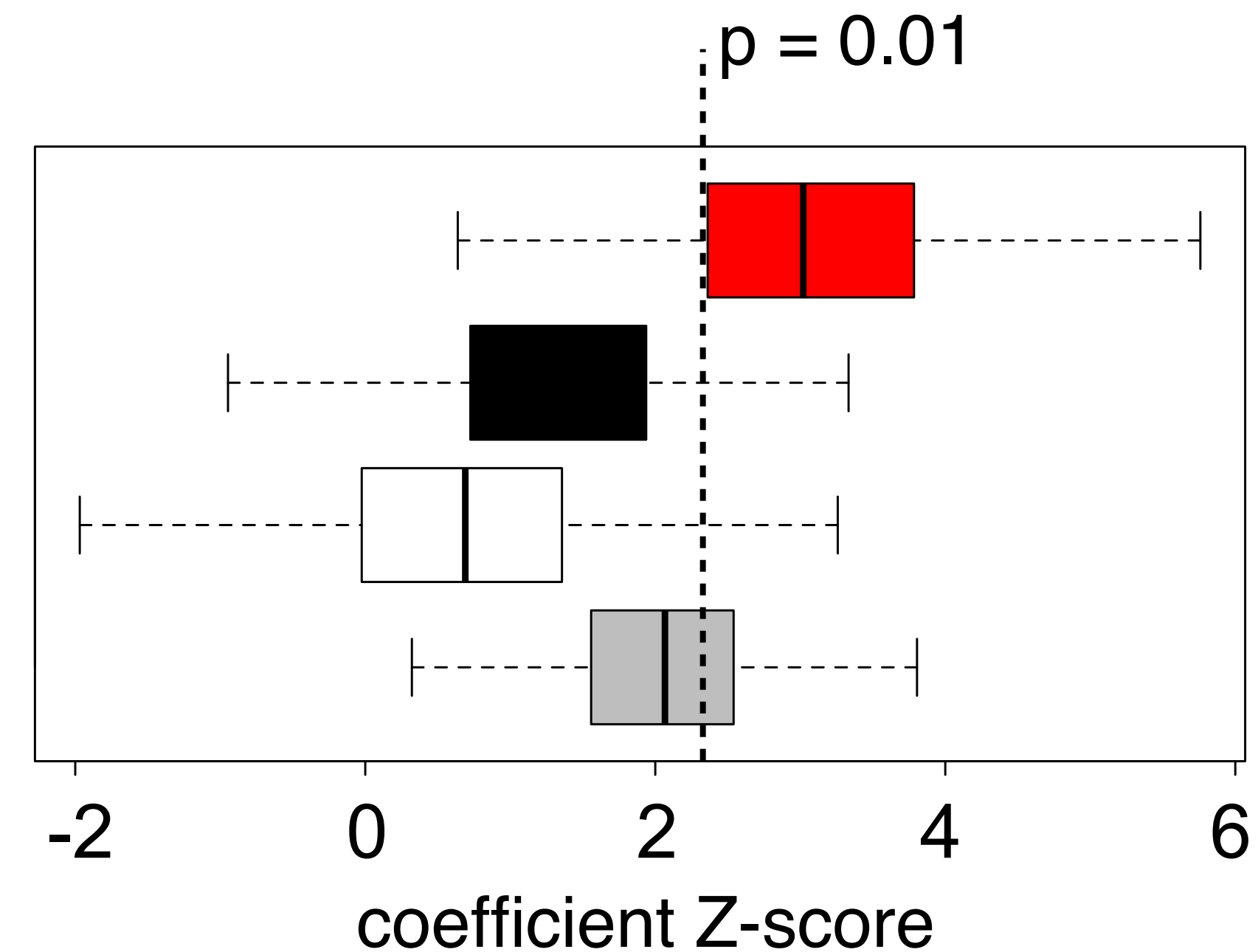
## Gene landscape DHS types:

Component concordant (17%) 

Component discordant (34%) 

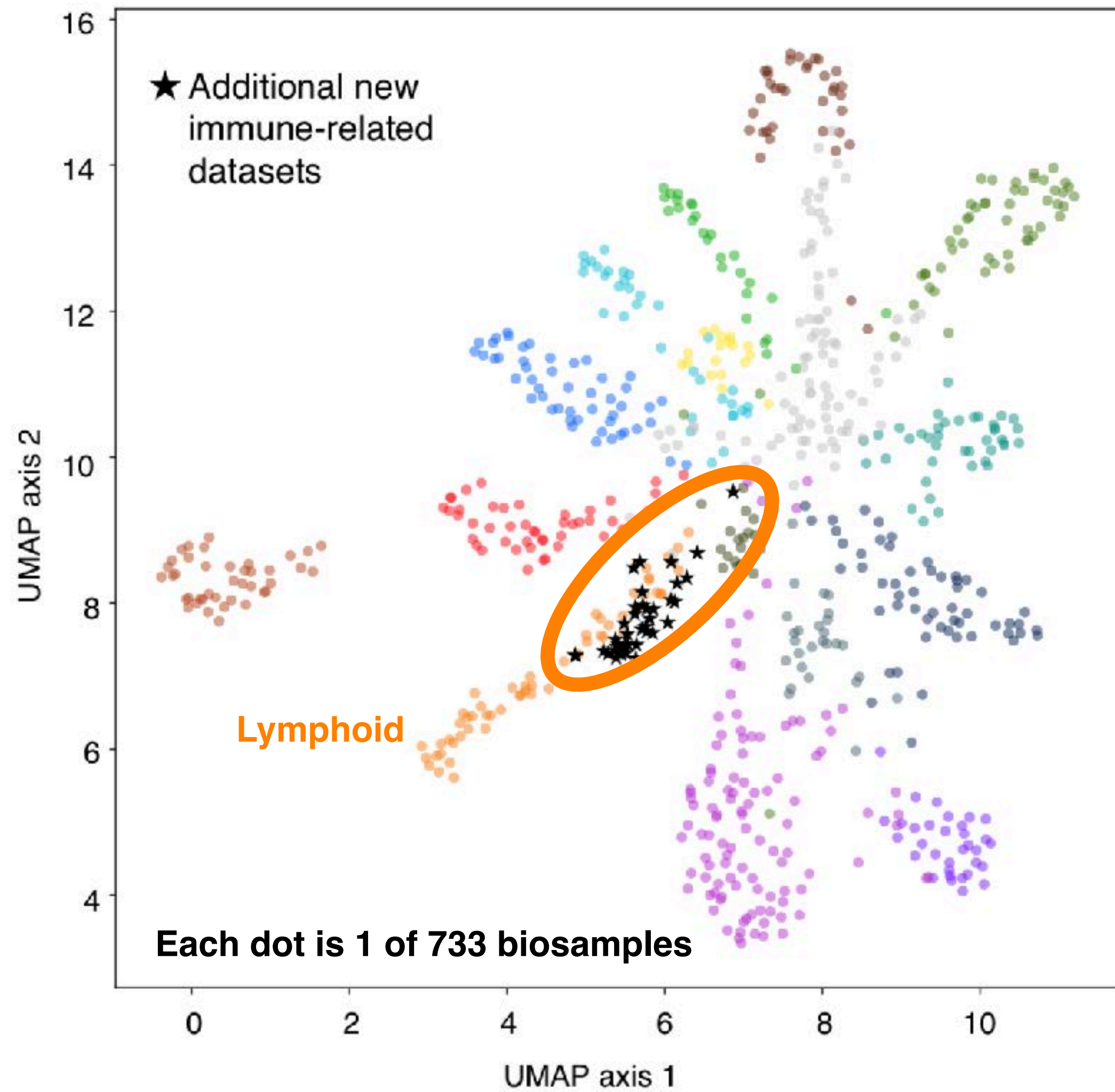
Genic control (18%) 

Inter-genic control (35%) 

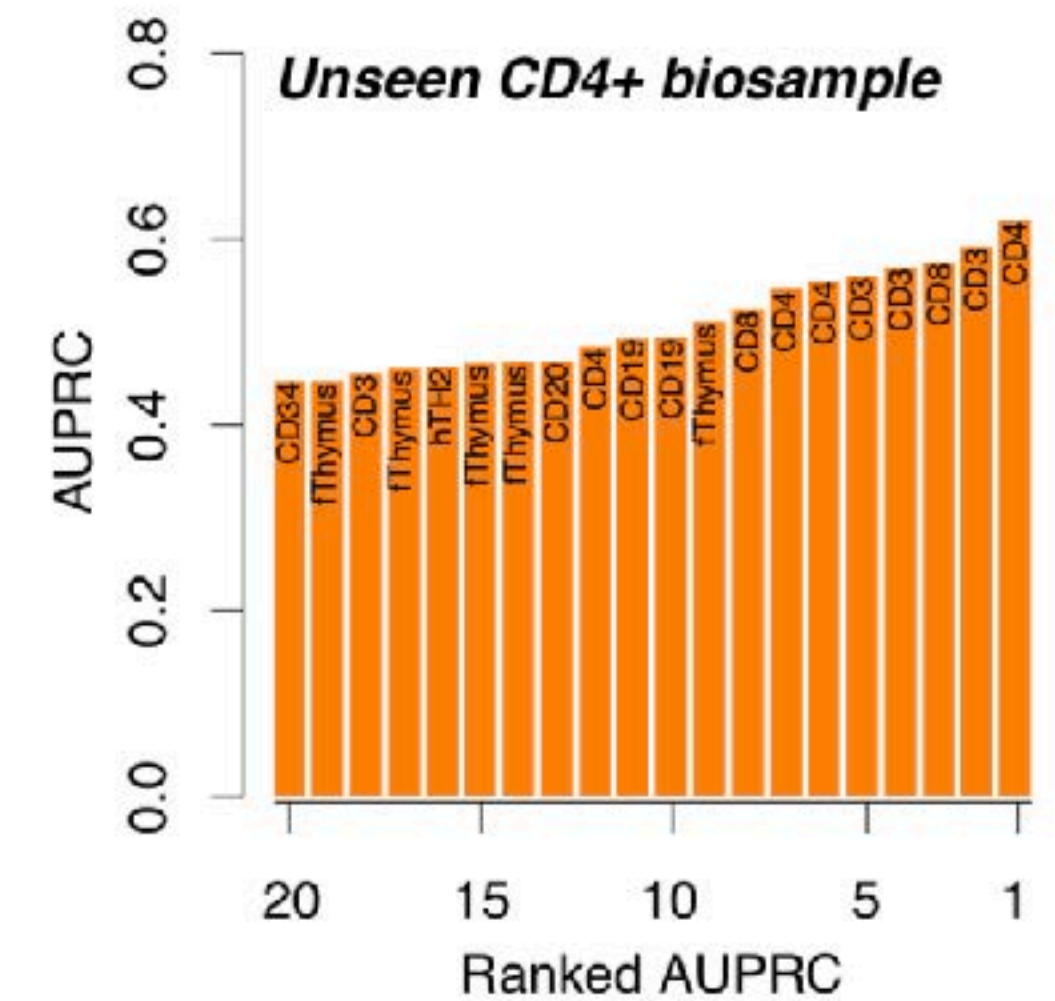
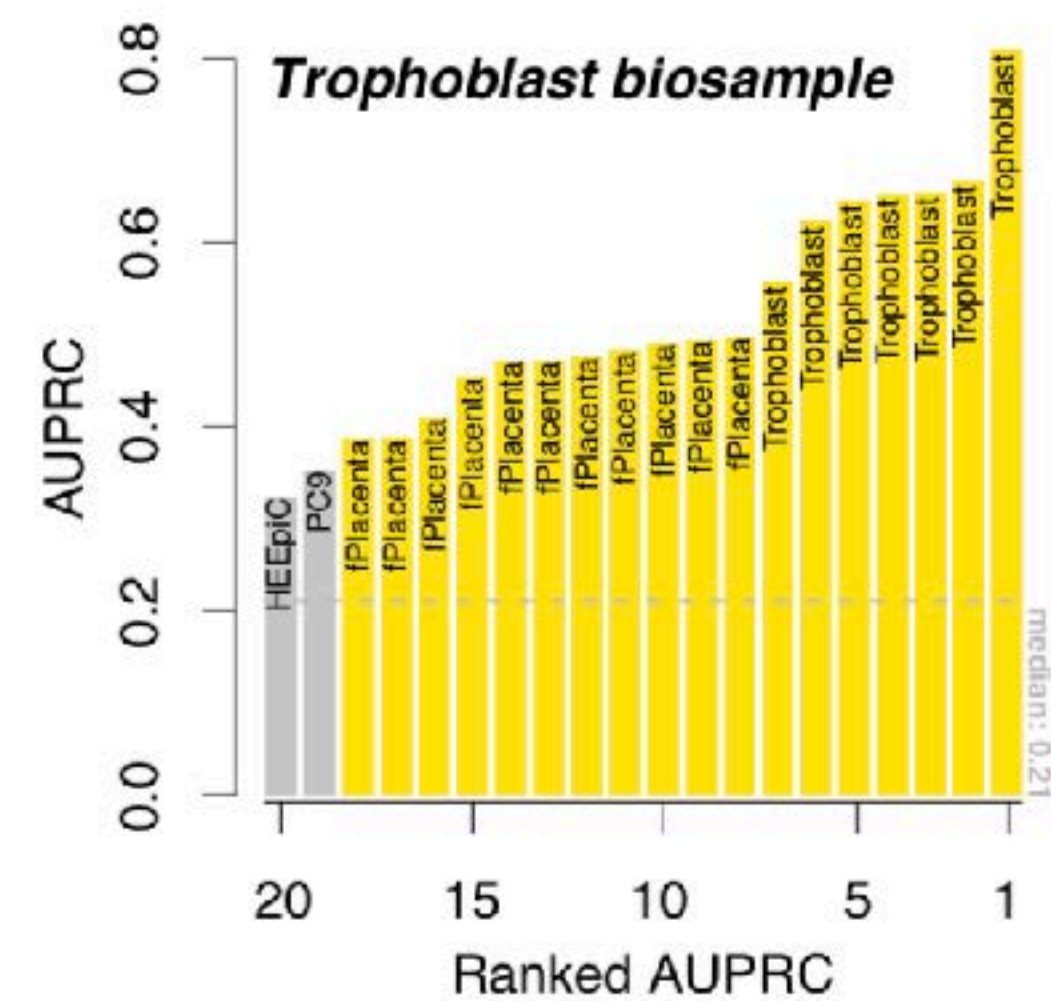
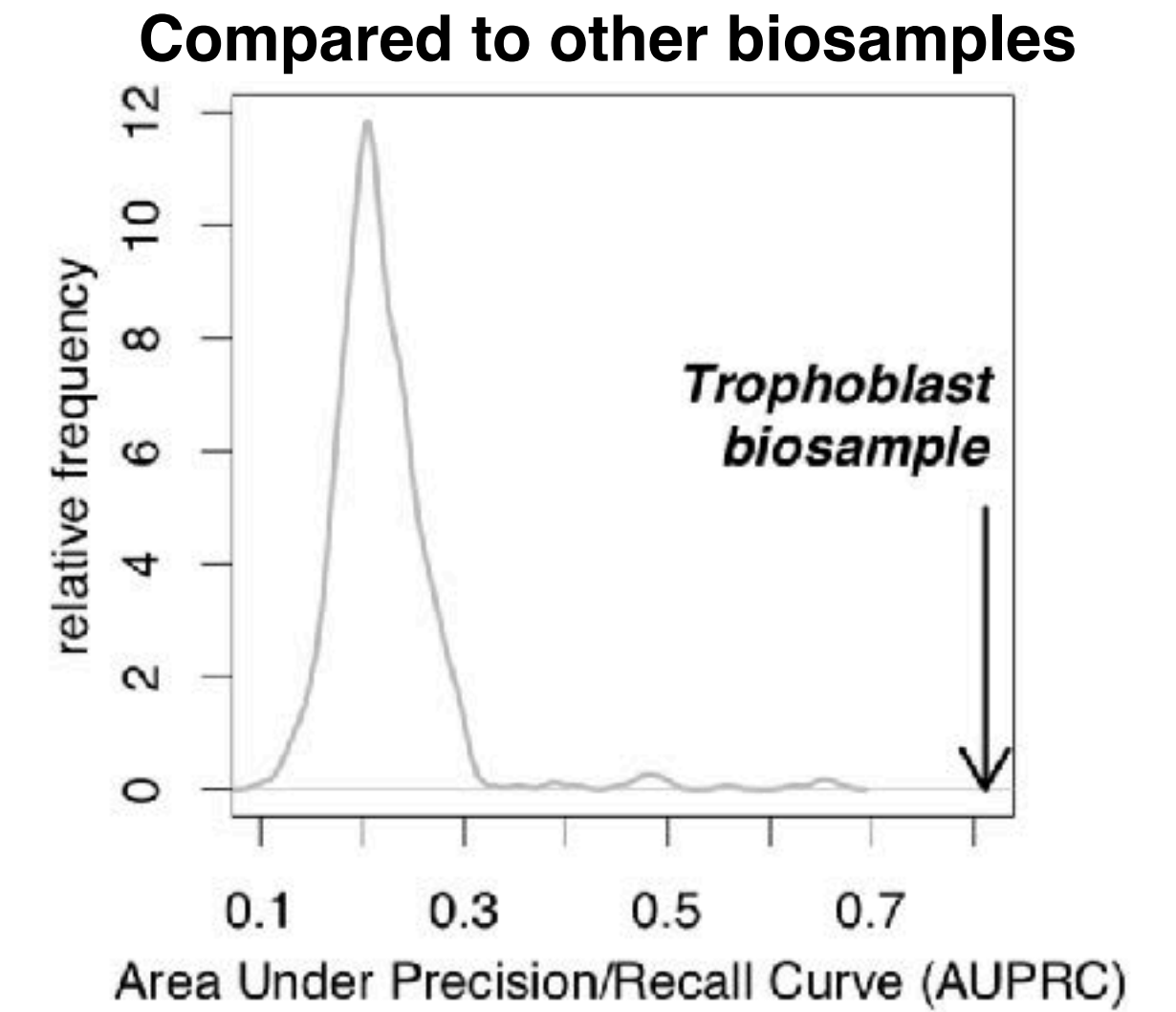
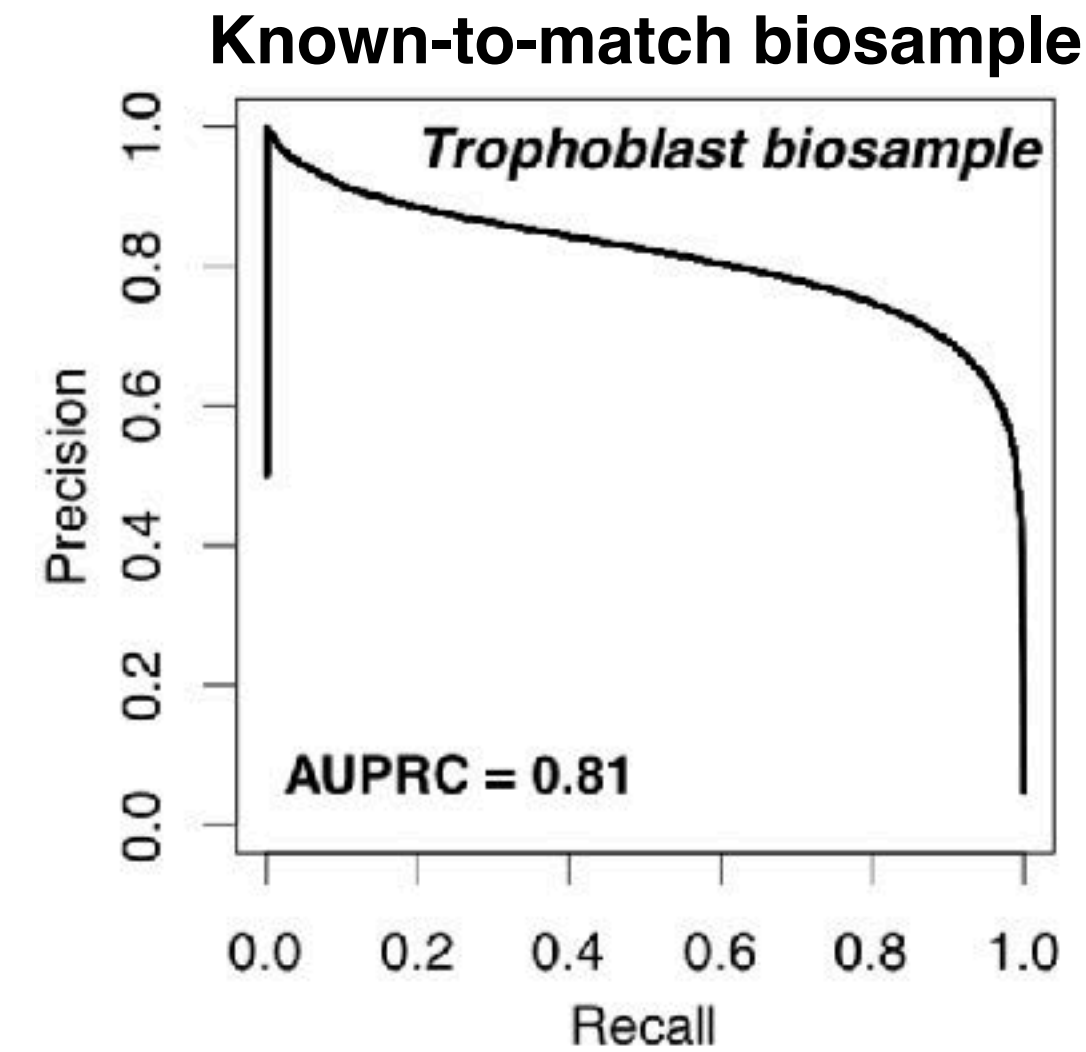


Transition from a discovery era to a detection era

# Annotation of unseen datasets by embedding into existing space




Using DHS peak calls



Peak call agnostic

# Take-home messages



We provide a high-resolution common coordinate system for regulatory DNA, along with a component-based annotation

These annotations enable:

- characterization of gene regulatory landscapes
- prioritization of disease-associated genetic signals
- integration of unseen datasets, agnostic of explicit peak calls

# Acknowledgements

## Data analysis

Sasha Muratov  
Eric Rynes  
Alex Reynolds  
Eric Haugen  
Nasi Teodosiadis

## Data coordination

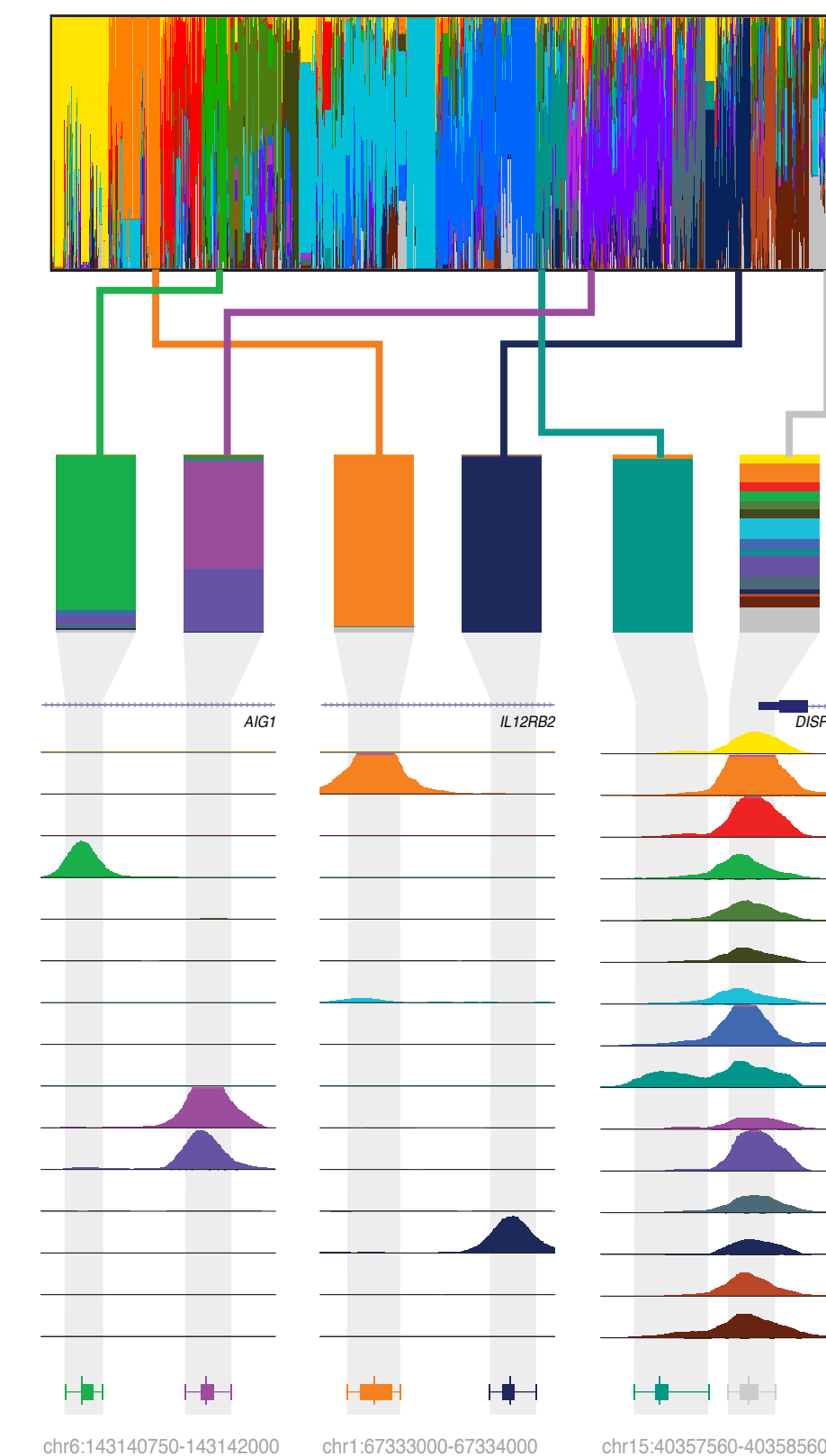
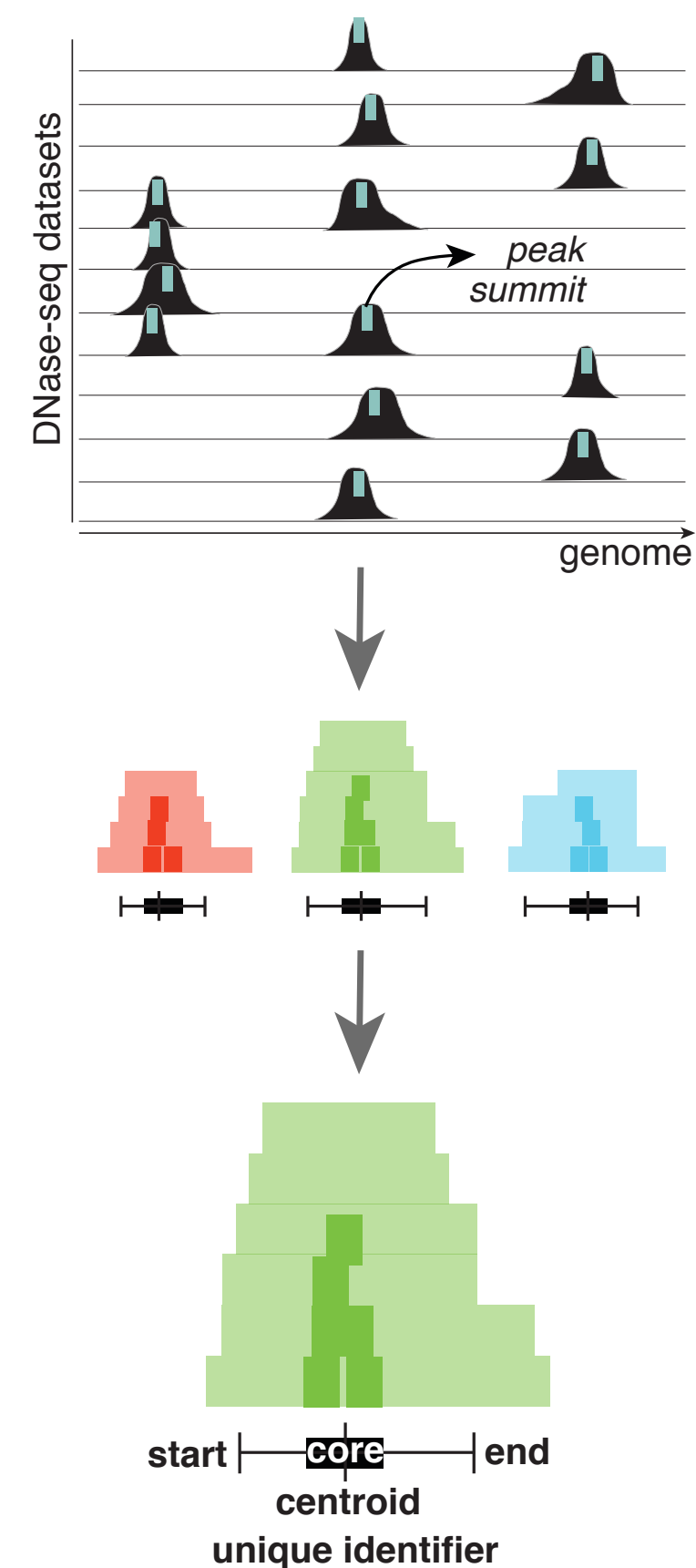
Richard Sandstrom  
Audra Johnson  
Jemma Nelson  
Mark Frerker  
Michael Buckley  
Rajinder Kaul

## Data generation

Jessica Halow  
Kristen Lee  
Daniel Bates  
Morgan Diegel  
Douglass Dunn  
Fidencio Neri

## Miscellaneous

Chad Lundberg  
Rae Senarighi  
Tim Mercer  
Jeff Vierstra  
John Stam.



*paper*



bioRxiv

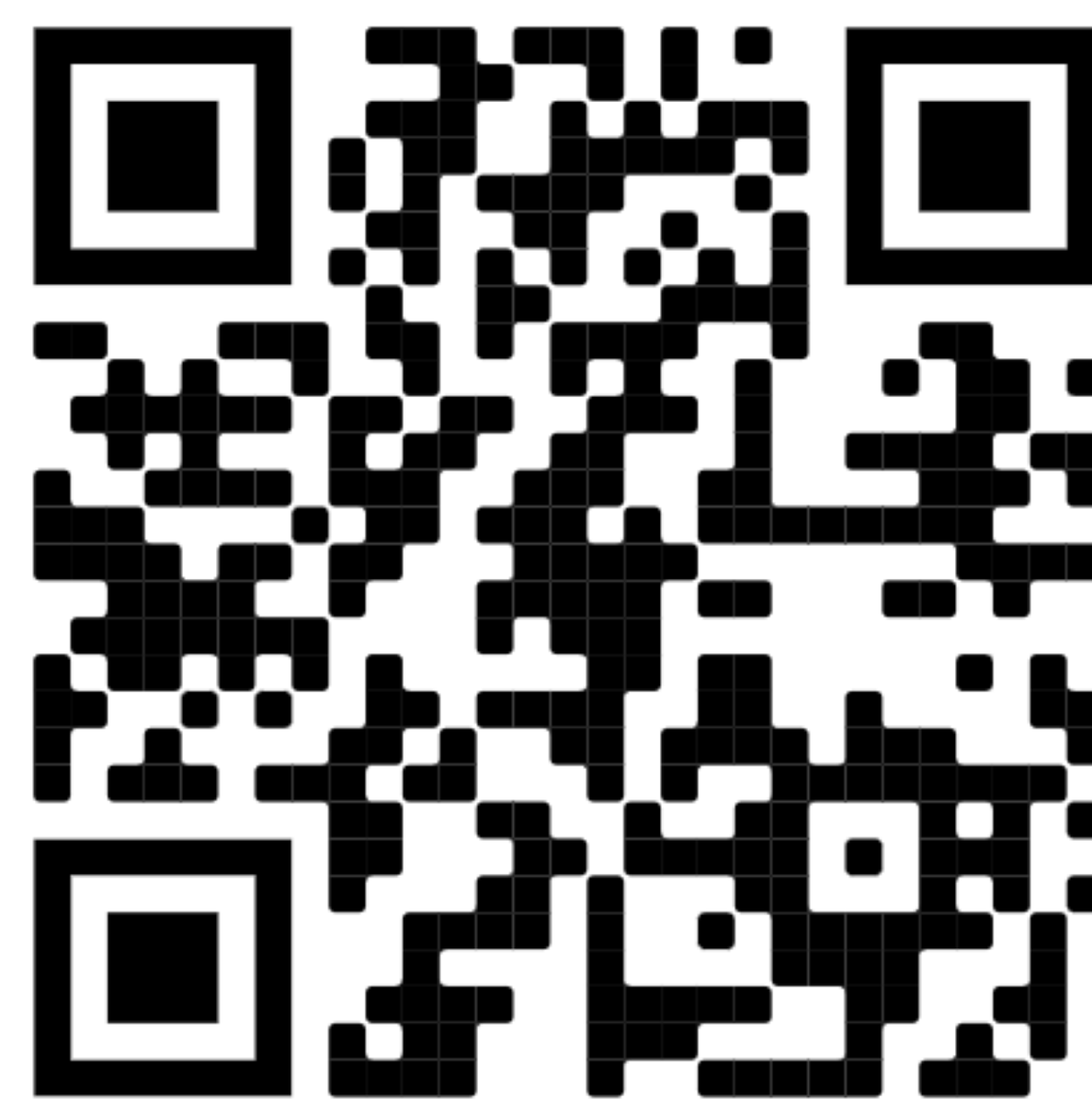
[index.altius.org](https://index.altius.org)

*browser*



[meuleman.org](https://meuleman.org)

*data*



zenodo

