Aggregating hundreds of ENCODE epigenomic annotations to predict cell-state-specific regulatory elements

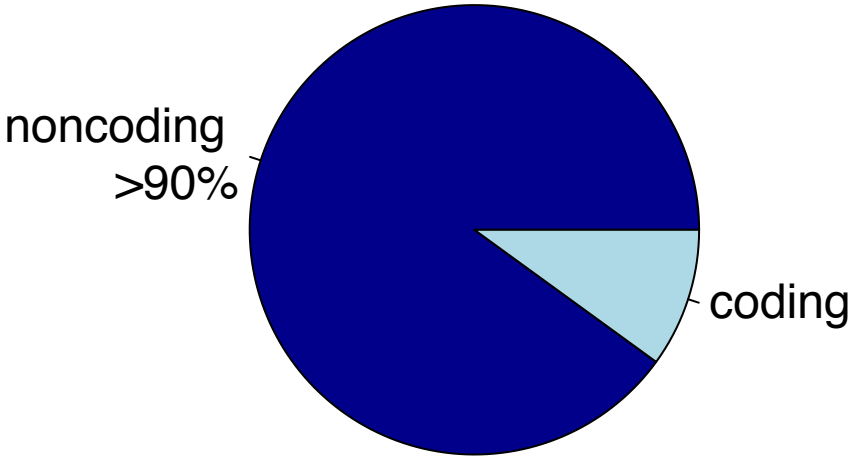
Tiffany Amariuta

Ph.D. Candidate, Raychaudhuri Lab

Harvard University, Boston, MA

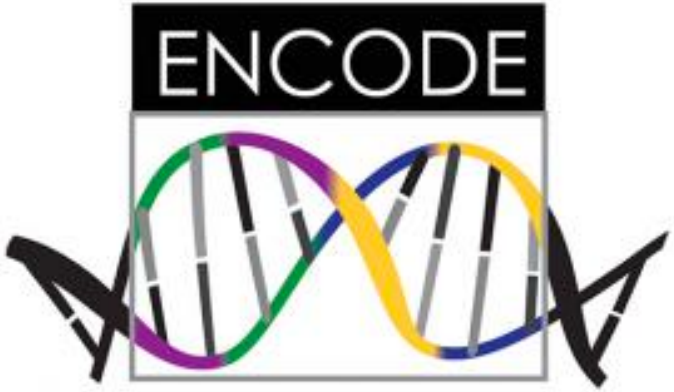
pre-ASHG ENCODE Workshop 2018

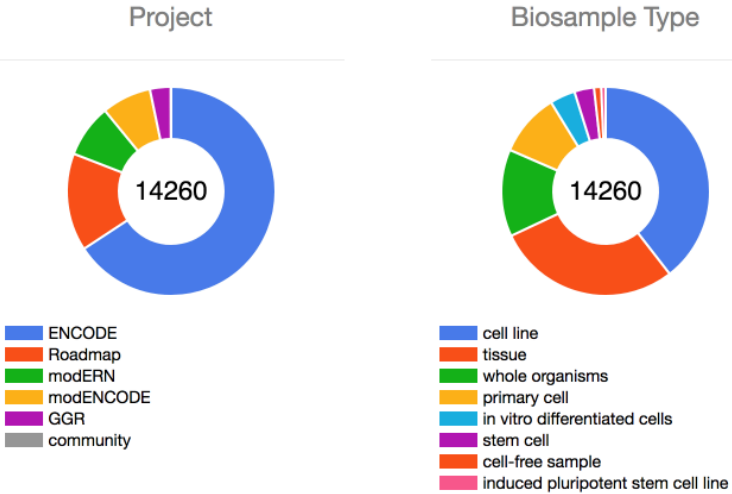
Majority of GWAS variants are noncoding



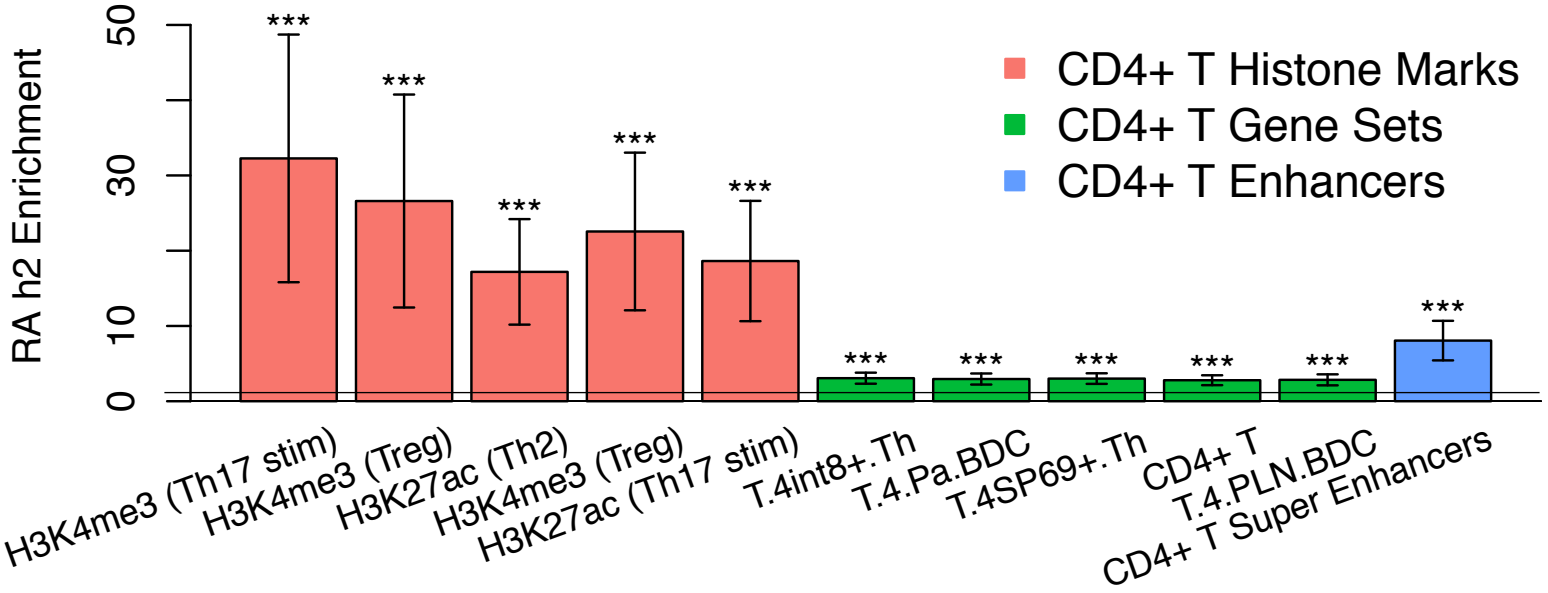
(Hrdlickova et al 2014 Review, BBA-Molecular Basis of Disease)

Big data: genomic functional annotation





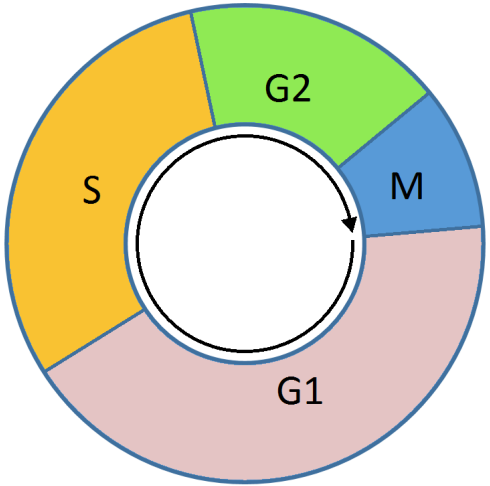
Gene regulatory elements enriched for disease heritability



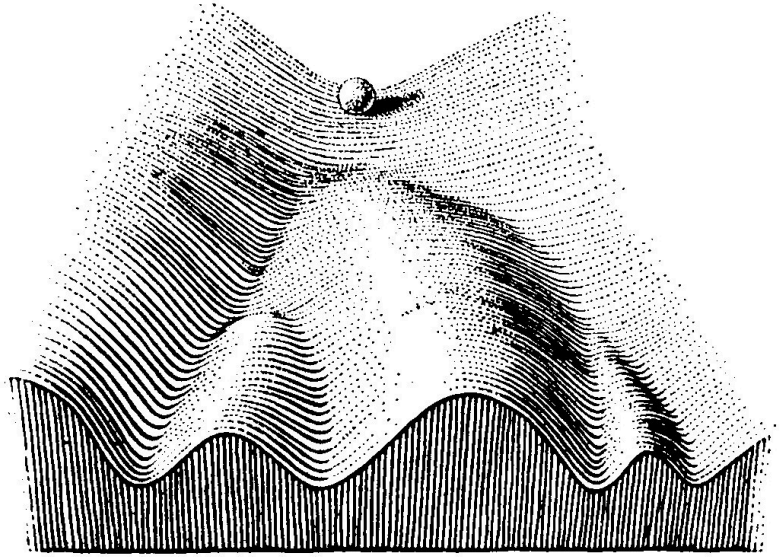
data from Finucane et al 2015 Nat Genet, Finucane et al 2018 Nat Genet

Distinguishing specific from nonspecific gene regulatory elements

NON-SPECIFIC

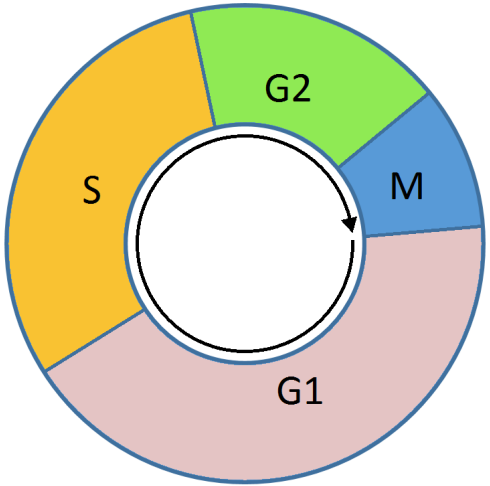


SPECIFIC

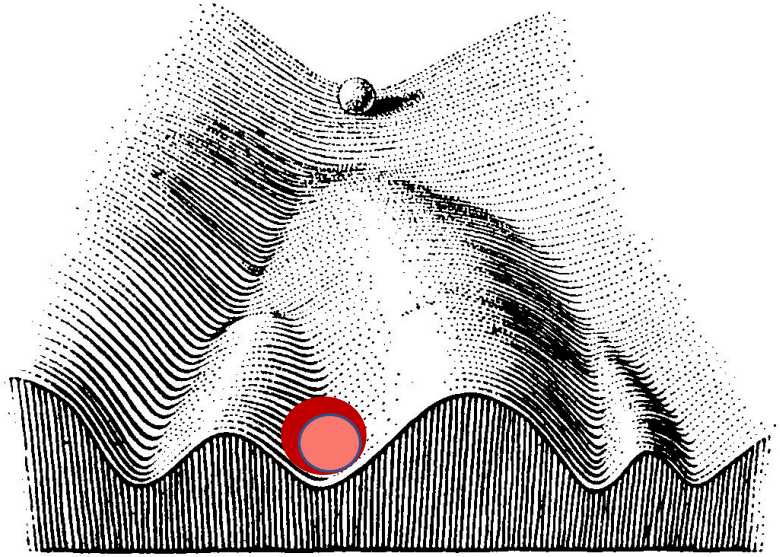


Distinguishing specific from nonspecific gene regulatory elements

NON-SPECIFIC

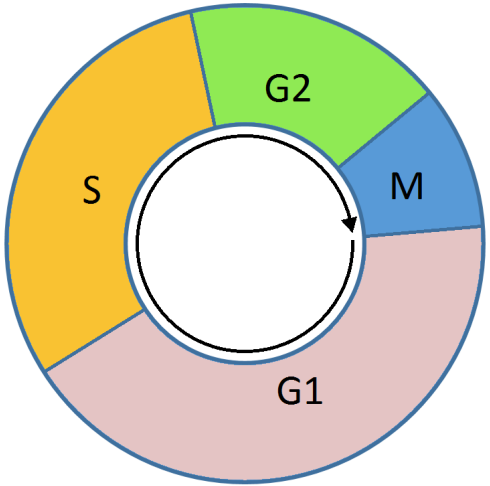


SPECIFIC

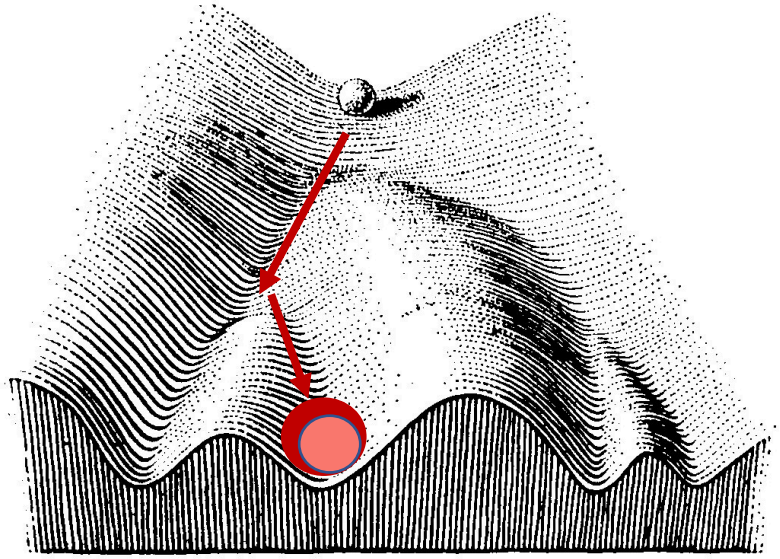


Distinguishing specific from nonspecific gene regulatory elements

NON-SPECIFIC

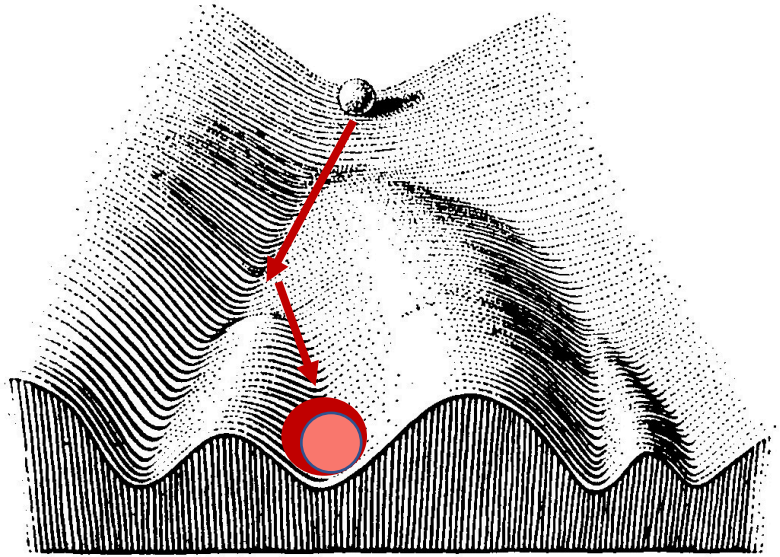


SPECIFIC

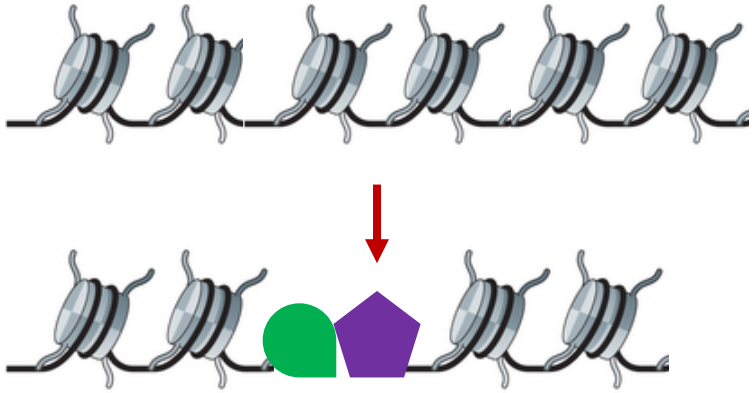


hypothesis:
MORE ENRICHED FOR DISEASE HERITABILITY

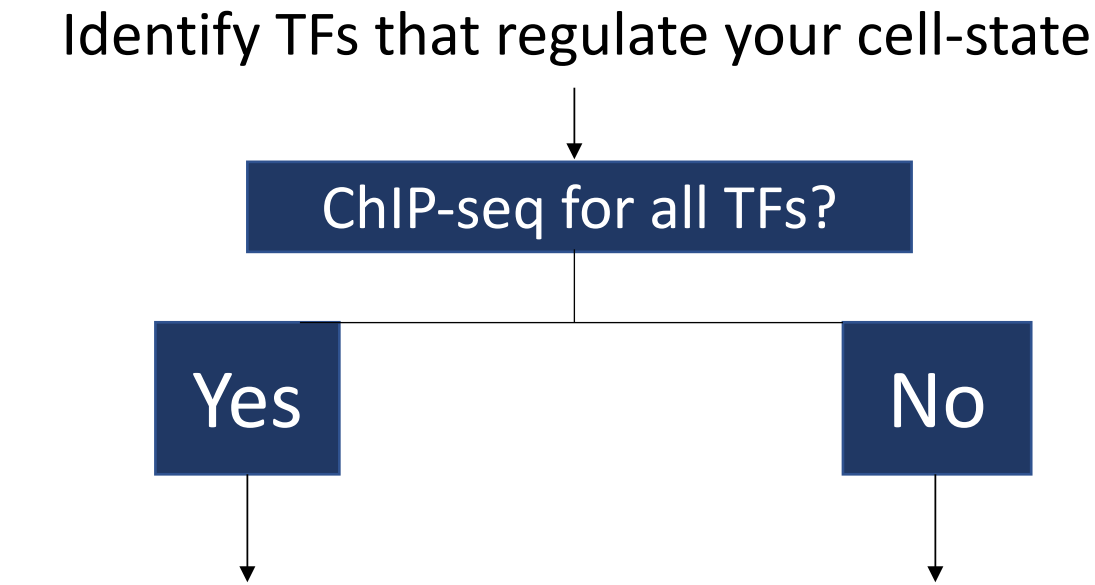
Regulatory elements may be identified by bound TFs and local epigenome



specific genome-wide regulatory changes



Defining cell-state-specific gene regulatory elements



Regulatory element:
any spot where TF binds

1. Identify TF motifs
2. Grow cells
3. Open chromatin assay

Regulatory element: peaks containing TF motifs

Defining cell-state-specific gene regulatory elements

Identify TFs that regulate your cell-state

ChIP-seq for all TFs?

May not know the TFs, have ChIP-seq for all TFs, have enough cells, or have a means to assay cell-state

Regulatory element:
any spot where TF binds

1. Identify TF motifs
2. Grow cells
3. Open chromatin assay

Regulatory element: peaks containing TF motifs

Defining cell-state-specific gene regulatory elements

Identify TFs that regulate your cell-state

↓

ChIP-seq for all TFs?

would include generic regulatory elements as well

↓

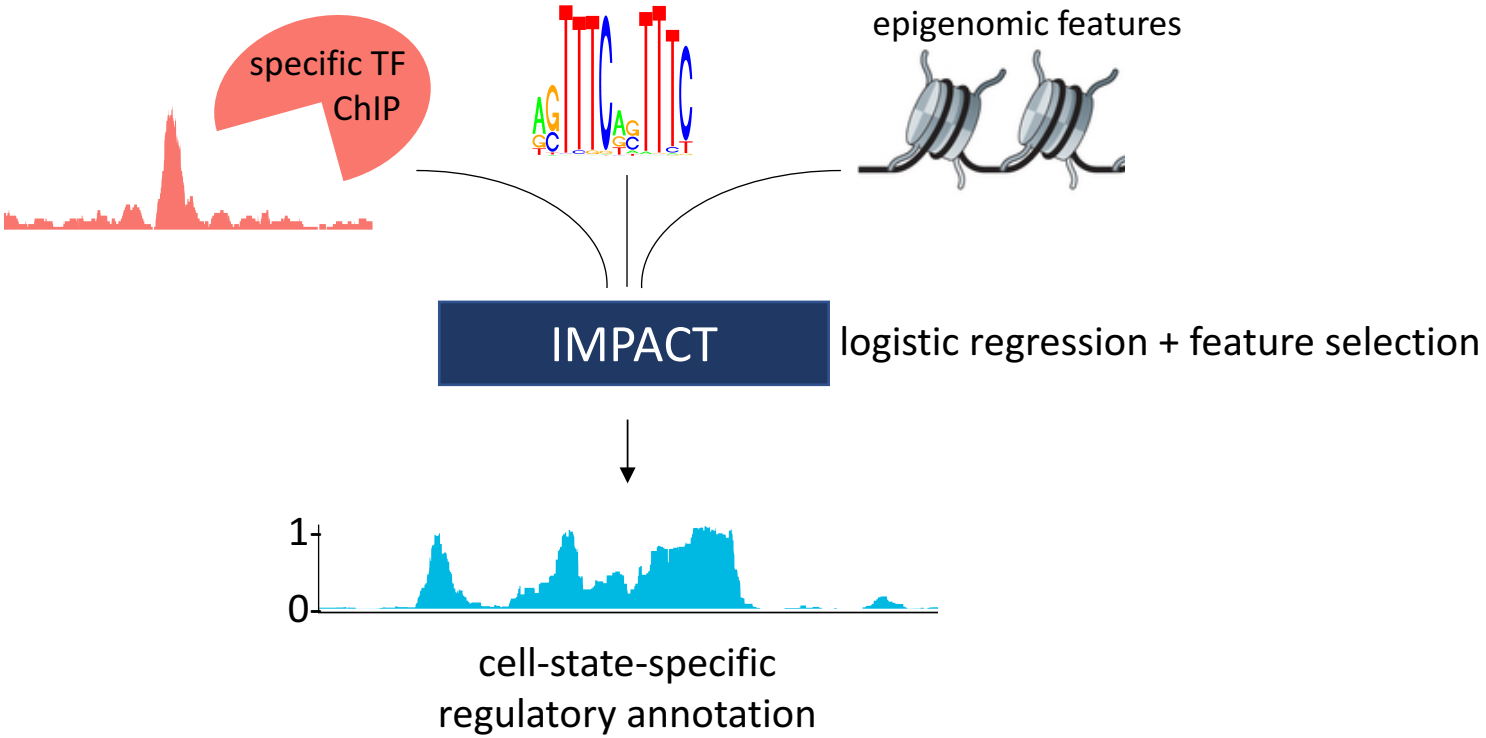
Regulatory element:
any spot where TF binds

↓

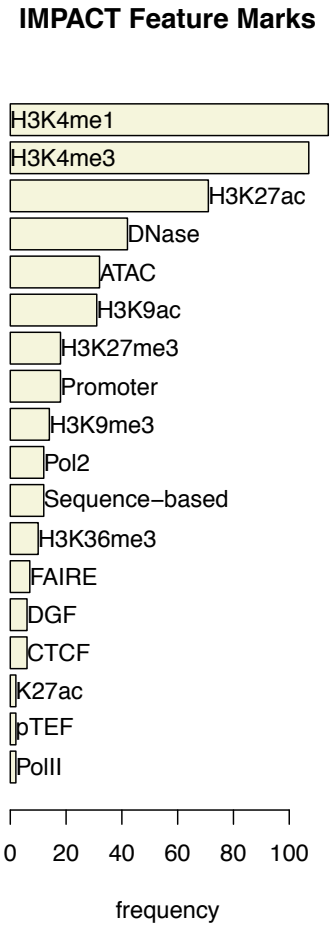
1. Identify TF motifs
2. Grow cells
3. Open chromatin assay

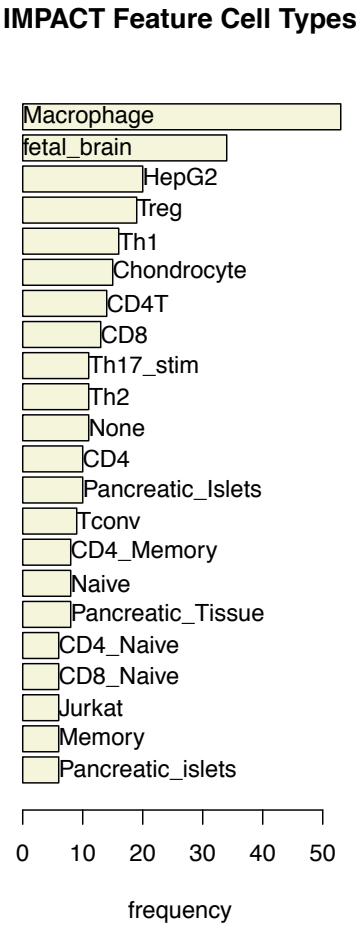
Regulatory element: peaks containing TF motifs

IMPACT
(Inference and Modeling of Phenotype-related ACtive Transcription)



IMPACT Epigenomic Features

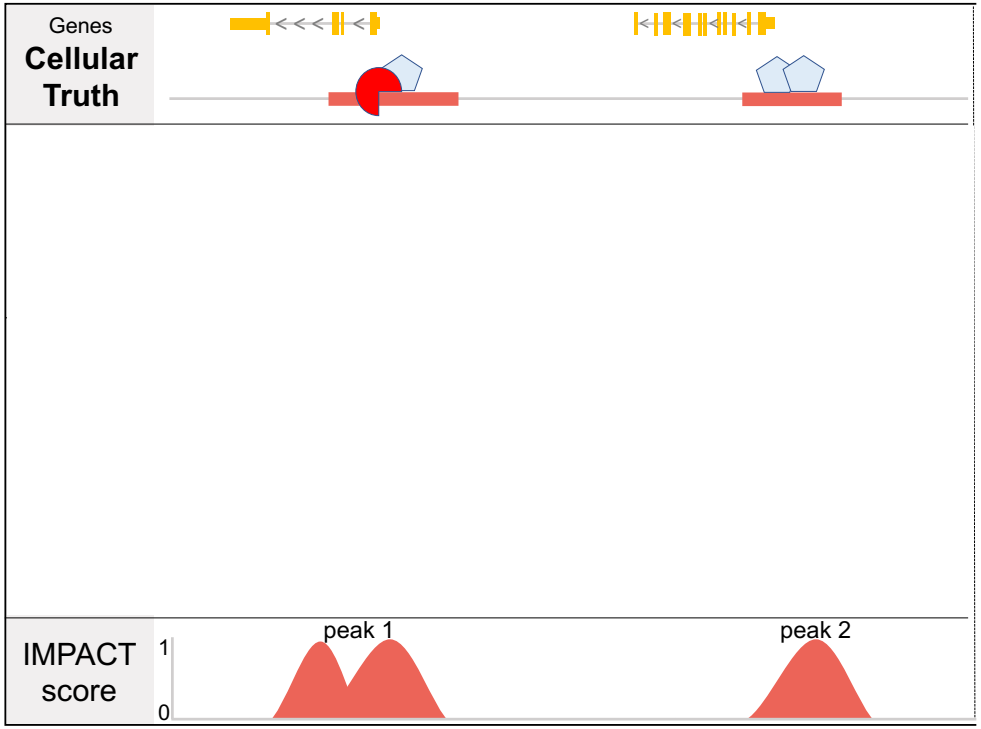




IMPACT model

key TF

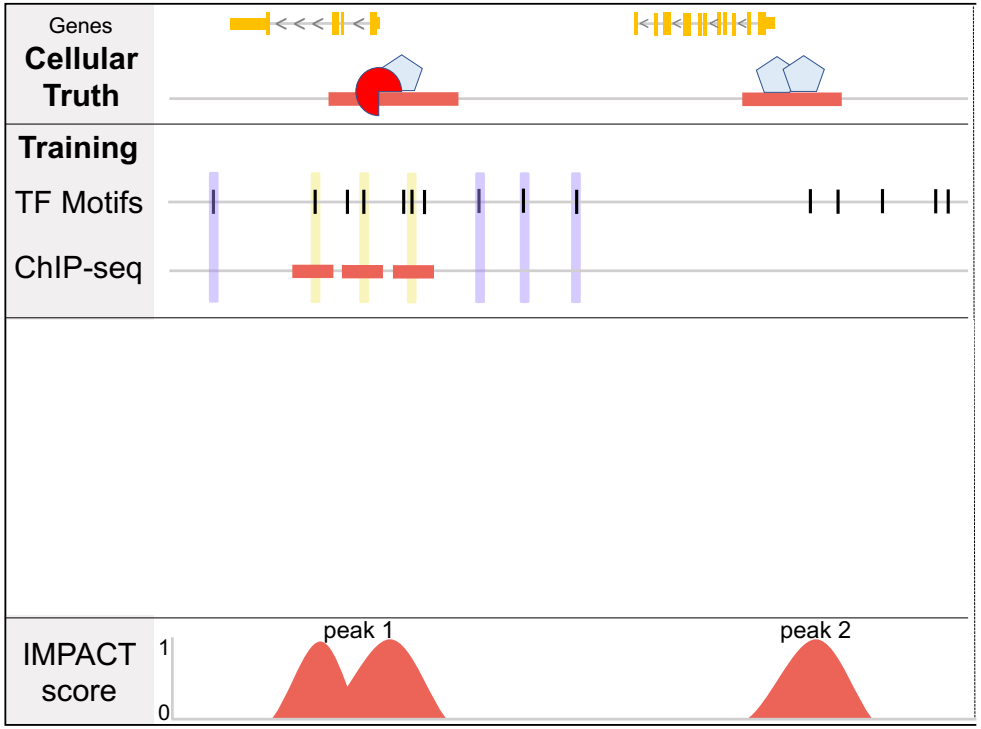
key TF coregulators



IMPACT model

key TF

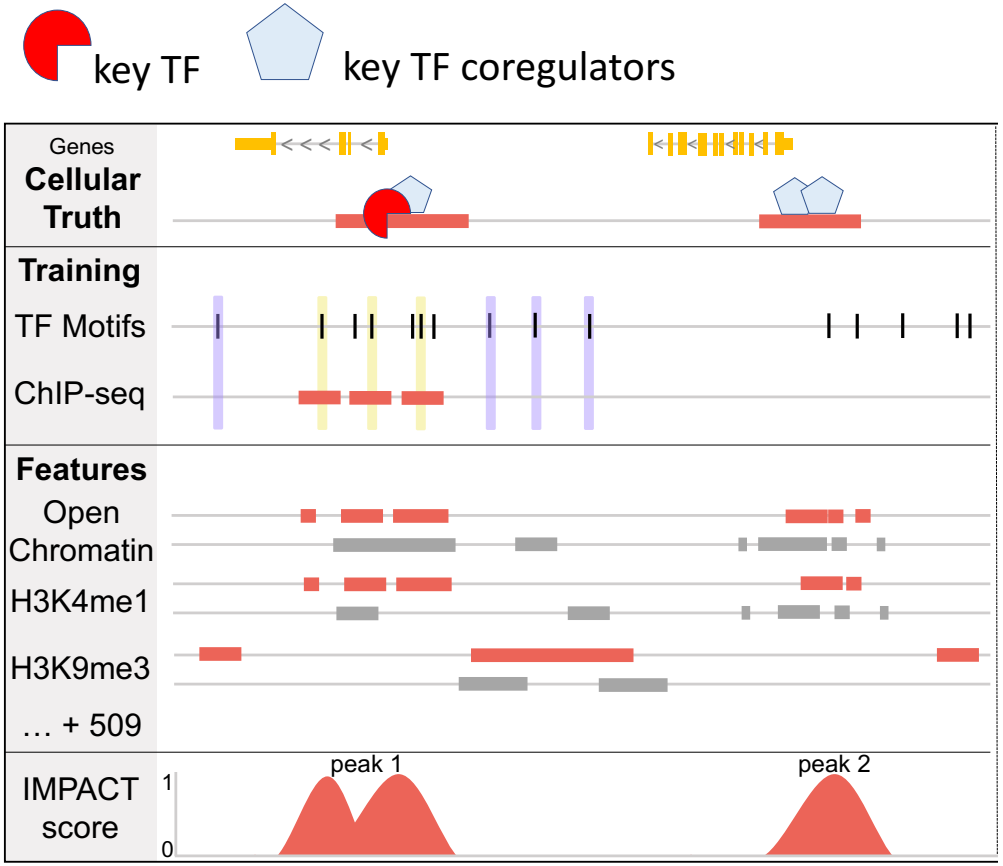
key TF coregulators



training regulatory regions

training non-regulatory regions

IMPACT learns a cell-state-specific epigenomic signature

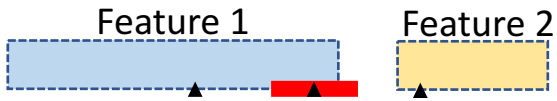


training regulatory regions

training non-regulatory regions

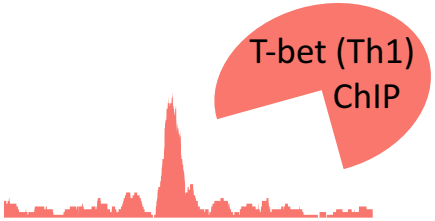
cell-type-specific regulatory elements

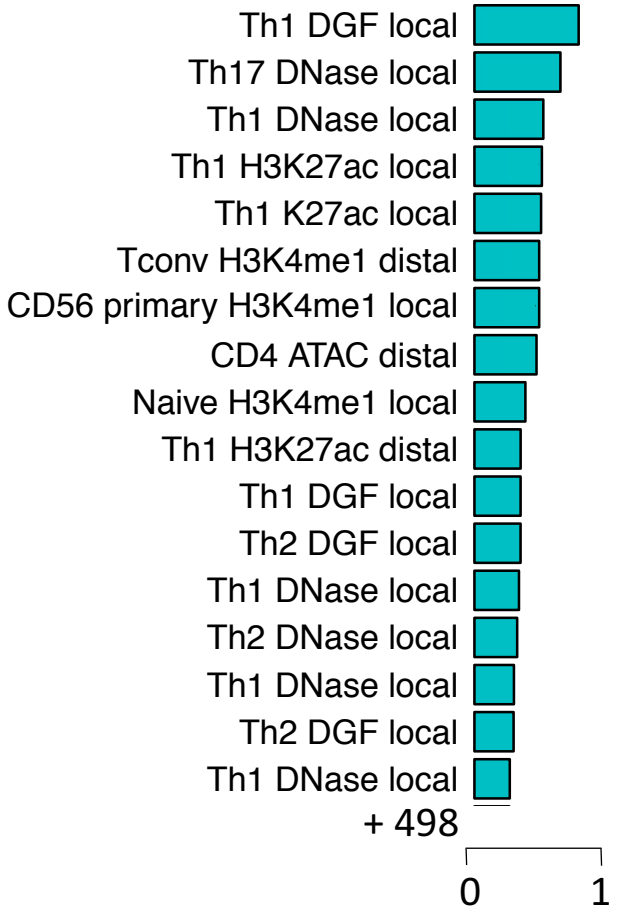
cell-type-nonspecific regulatory elements



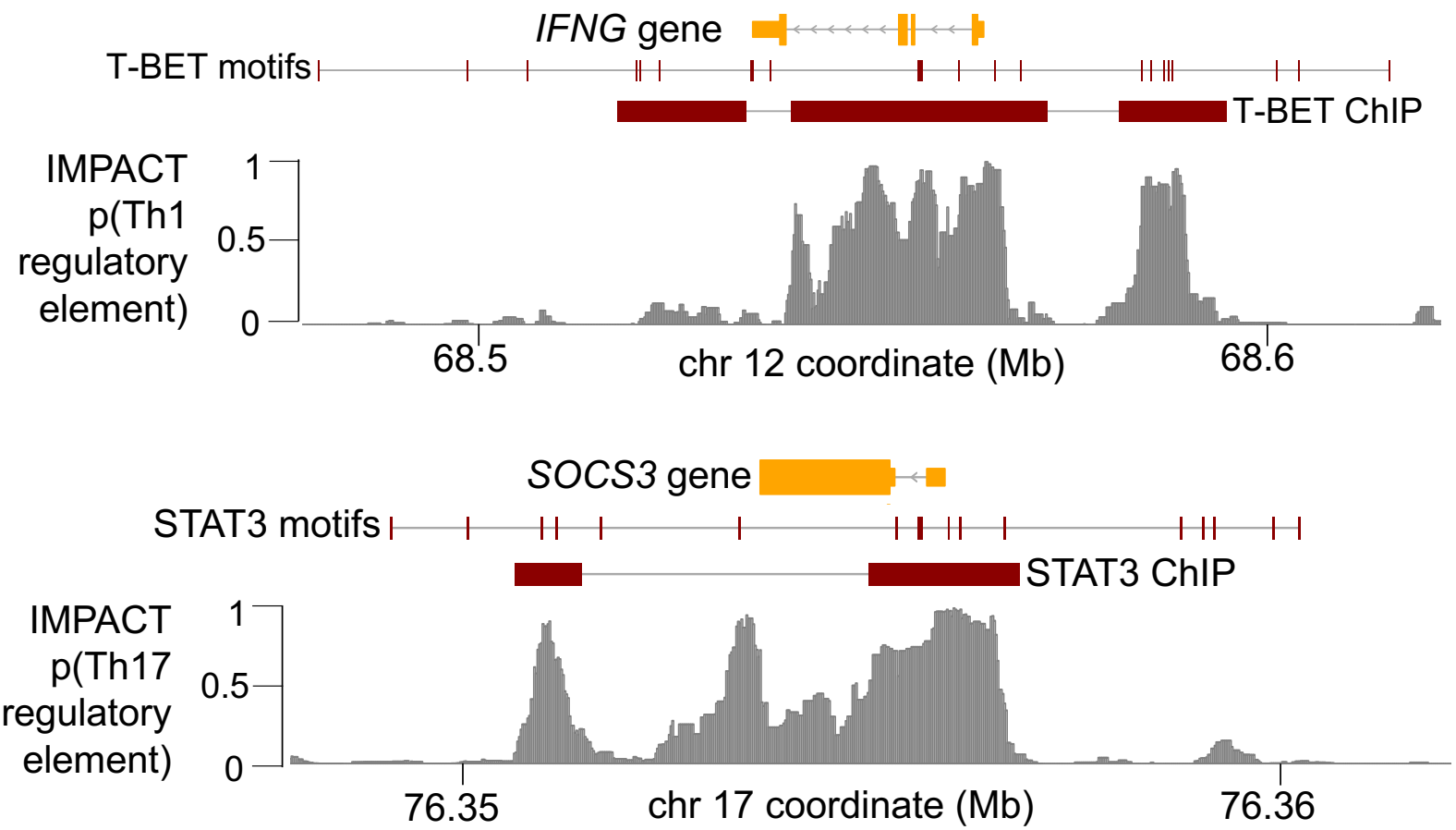
Region	Feat.1	Feat.2	Feat.1	Feat.2
motif 1	1	0	1	1
	local	local	distal	distal

IMPACT epigenomic signature





Epigenomic resolution is finer than ChIP-seq



IMPACT Validation

- IMPACT model assumes that binding sites of each TF will constitute a global epigenomic signature. *(Cowper Sal-Iari et al Nat Genet 2012)*

IMPACT Validation

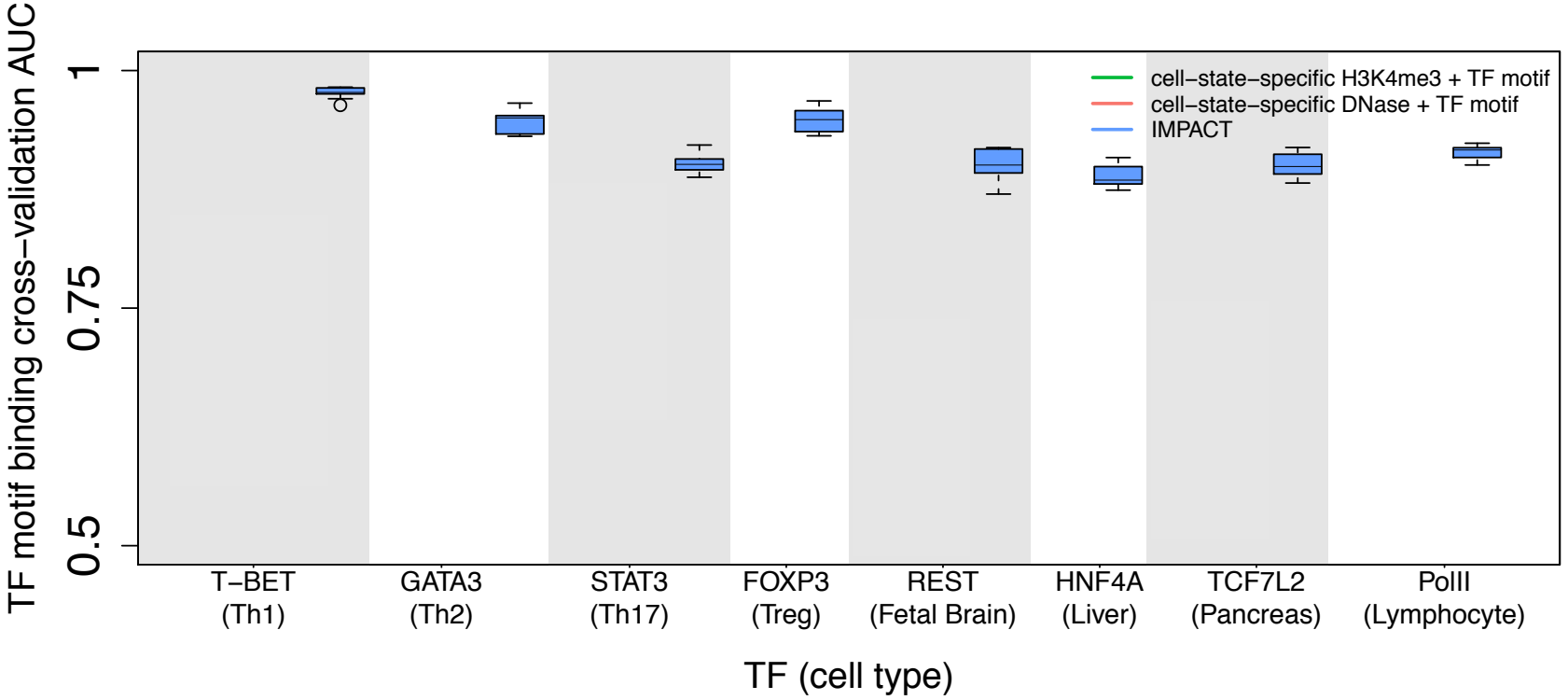
- IMPACT model assumes that binding sites of each TF will constitute a global epigenomic signature. *(Cowper Sal-Iari et al Nat Genet 2012)*

- If the signature is robust, each binding site genome-wide should resemble this signature.

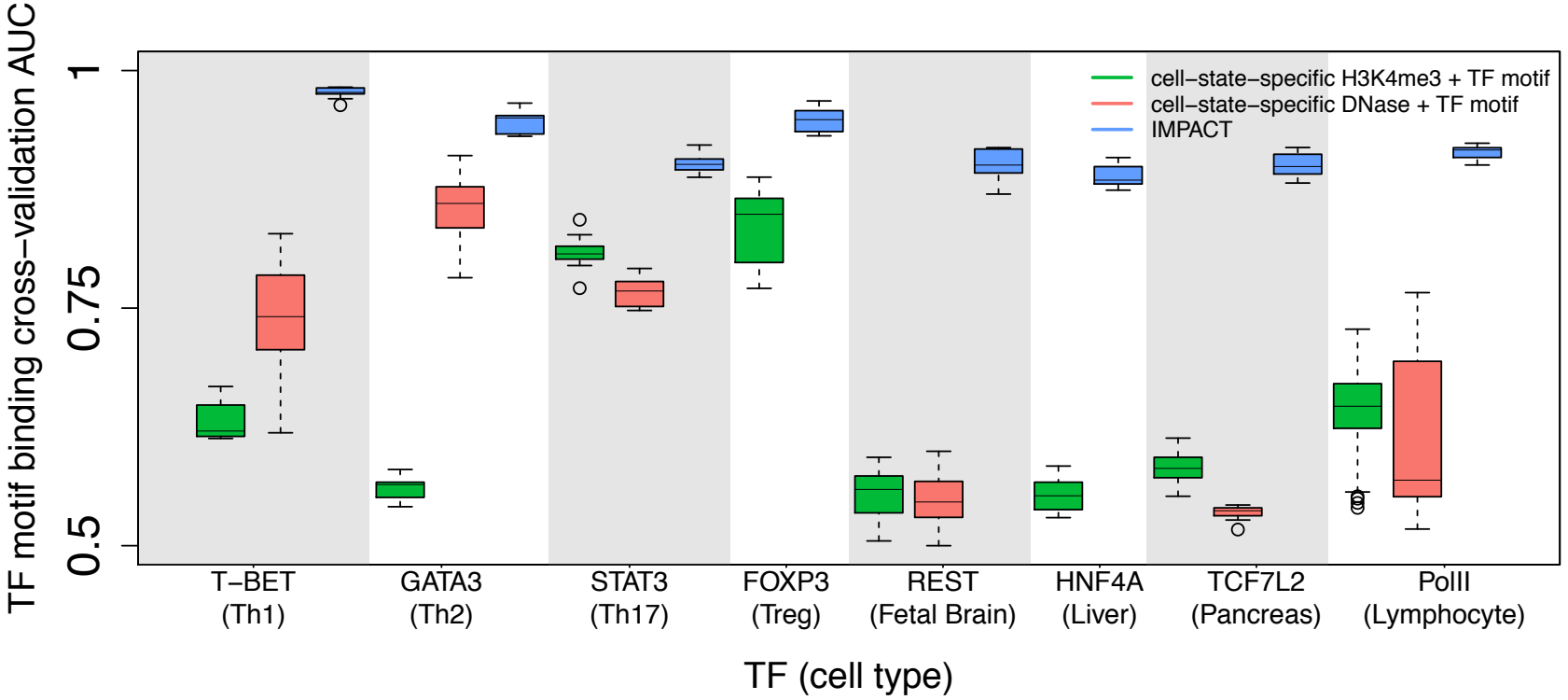
IMPACT Validation

- IMPACT model assumes that binding sites of each TF will constitute a global epigenomic signature. *(Cowper Sal-Iari et al Nat Genet 2012)*
- If the signature is robust, each binding site genome-wide should resemble this signature.
- Therefore, IMPACT should be able to accurately predict TF binding.
- **TF binding prediction is notoriously difficult**.

IMPACT reidentifies TF binding



IMPACT is more accurate than open chromatin + motif



average increase in AUC from 0.66 to 0.92

IMPACT Application: Expression Causal Variation

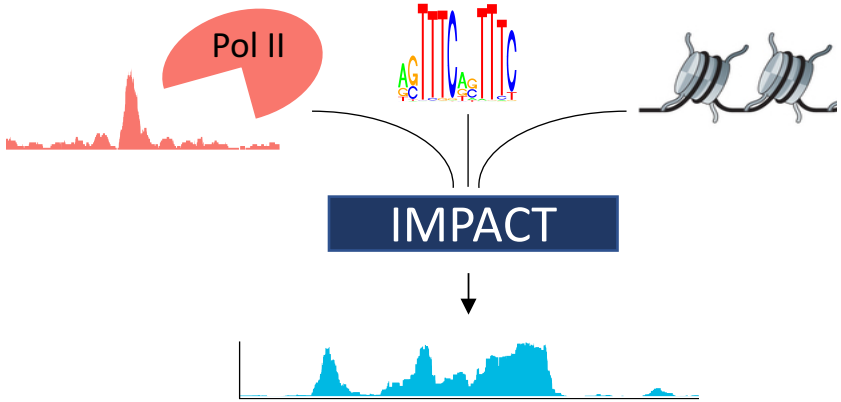
- **cis eQTL** causal variation is known to be concentrated around promoters and the TSS. *(Liu et al AJHG 2017)*

IMPACT Application: Expression Causal Variation

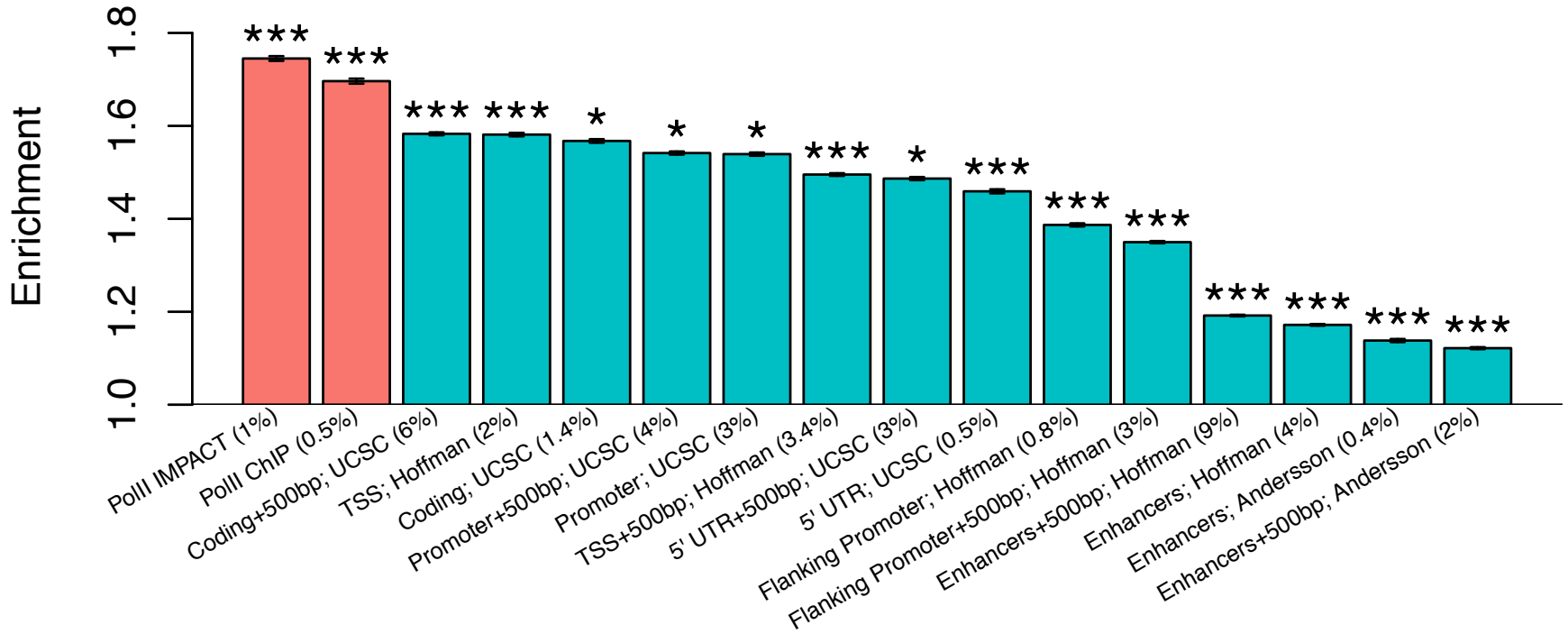
- **cis eQTL** causal variation is known to be concentrated around promoters and the TSS. *(Liu et al AJHG 2017)*

- *Hypothesis*

IMPACT **RNA polymerase II** annotation would better capture this variation.



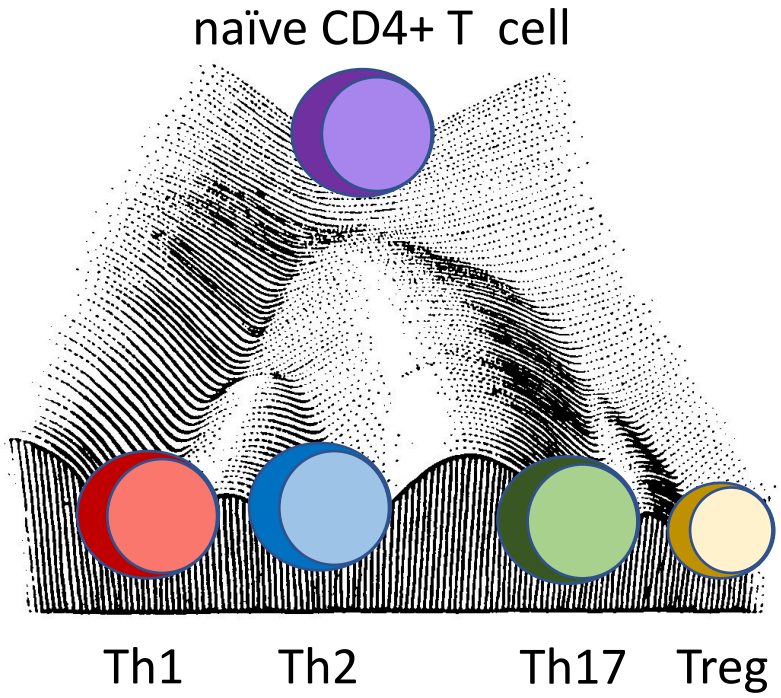
IMPACT Application: Expression Causal Variation



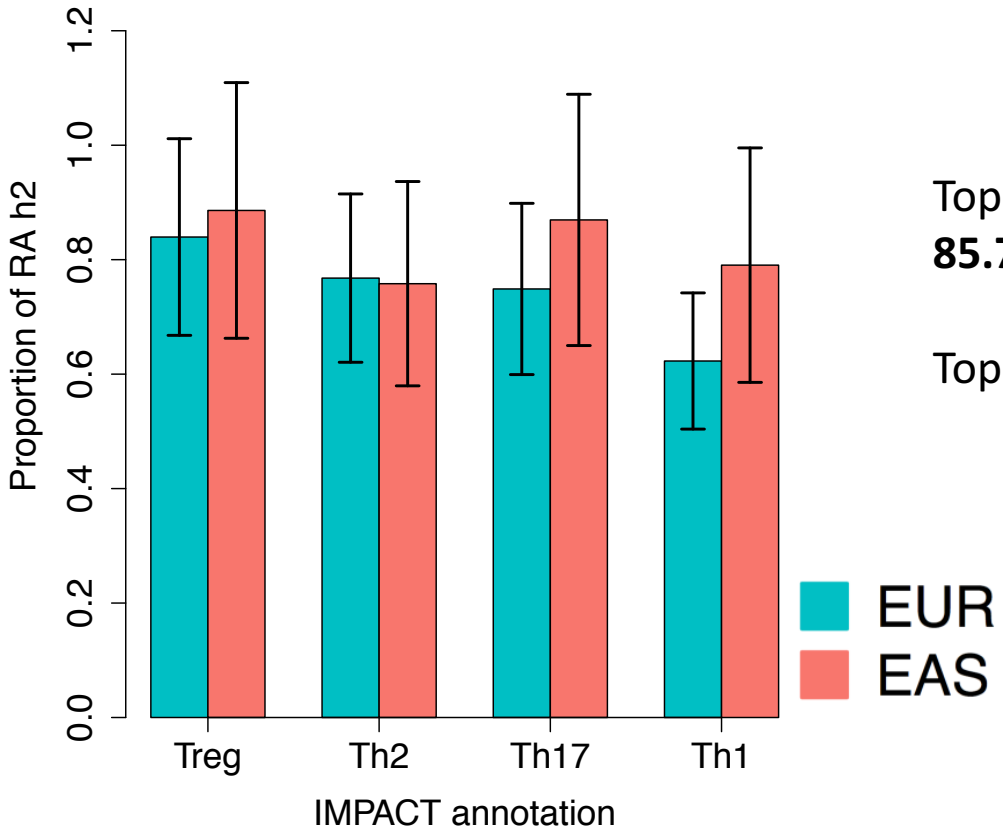
25% average increase in enrichment

IMPACT Application: Polygenic Causal Variation

- CD4+ T cells are strongly implicated in RA pathogenesis
- CD4+ T cell-states: Th1, Th2, Th17, Tregs, etc…



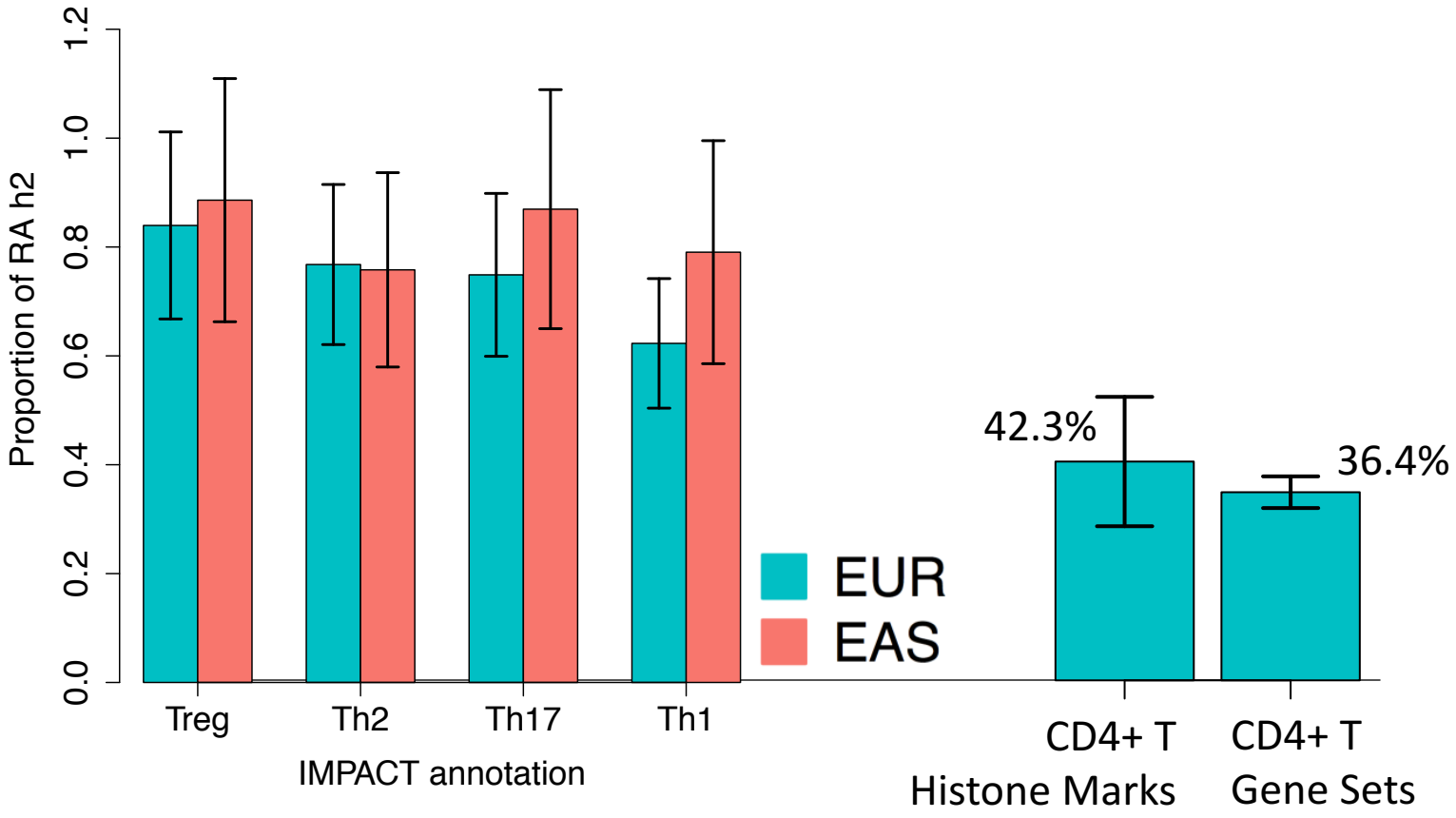
IMPACT explains majority of RA h2



Top 5% of IMPACT Treg SNPs explain **85.7%** of RA h2.

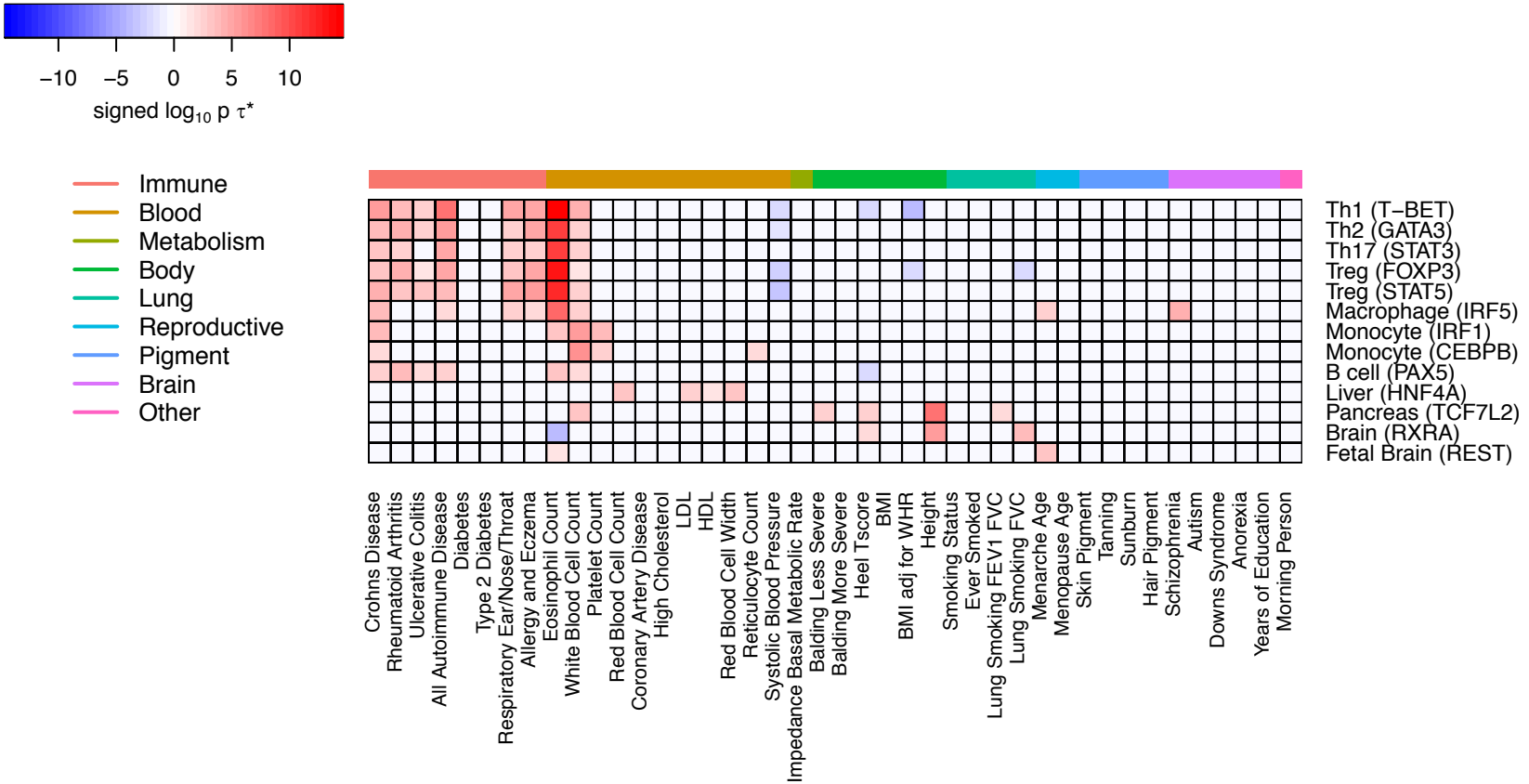
Top 9.8% explain **97.3%** of RA h2.

Most comprehensive explanation for RA h2



IMPACT applied to 42 polygenic traits

IMPACT applied to 42 polygenic traits



IMPACT applied to 42 polygenic traits: highlights

- CD4+ T IMPACT (T-BET, STAT3, GATA3, FOXP3) → explains heritability of (auto)immune traits
- Liver IMPACT (HNF4A) → explains heritability of LDL, HDL, + some blood traits
- Macrophage IMPACT (IRF5) → explains heritability of Schizophrenia, + immune traits
- Monocyte IMPACT (CEBPB, IRF1) → explains heritability of blood traits, not immune traits
- CD4+ Treg IMPACT (STAT5) replicates CD4+ Treg IMPACT (FOXP3)

IMPACT Application: A priori identification of functional variants

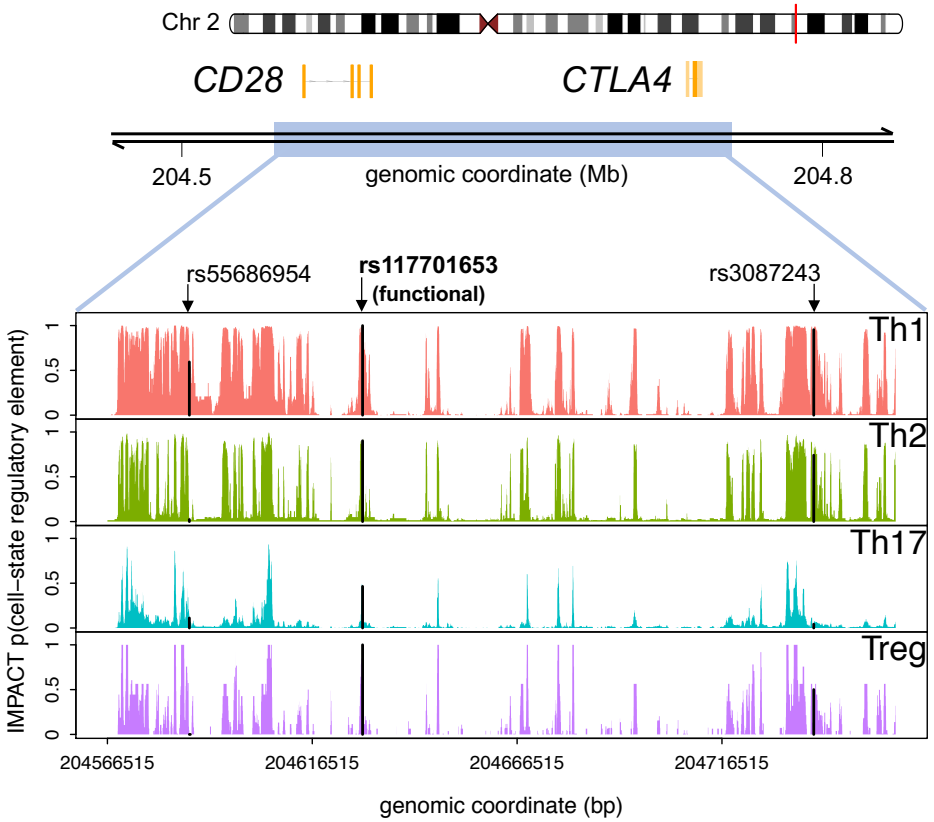
- Fine-mapping 20 RA risk loci using GWAS of 11,475 European RA cases and 15,870 controls *(Westra et al 2018 Nature Genetics)*

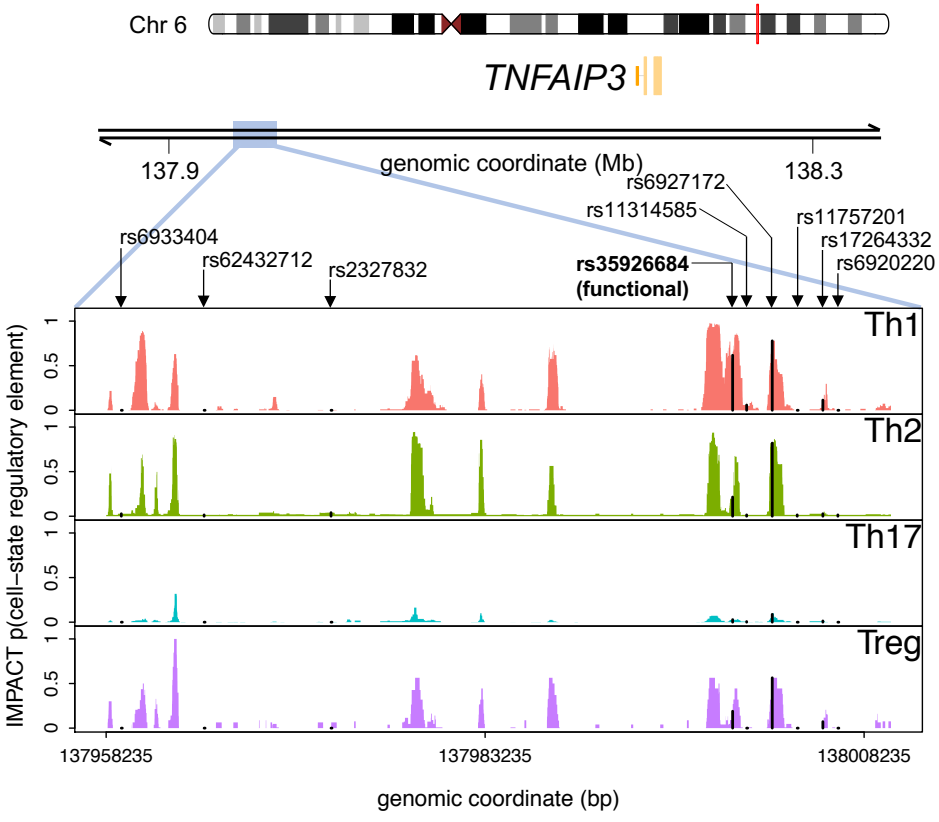
- Identification of putatively causal variants:

    1. Differential binding of CD4+ T nuclear extract (EMSA)

    2. Differential enhancer activity (luciferase assays)

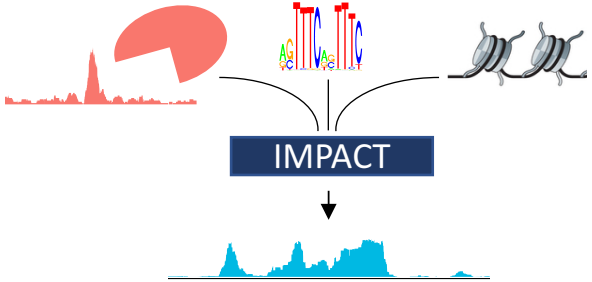
IMPACT Application: A priori identification of functional variants



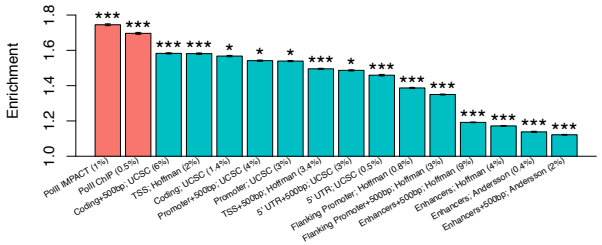


Applications of IMPACT

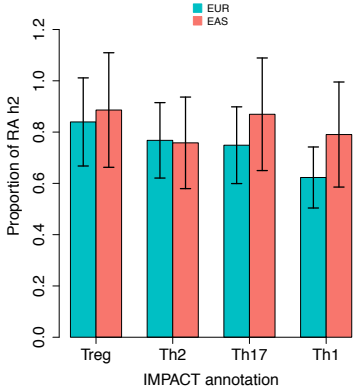
1. Cell-state-specific regulation



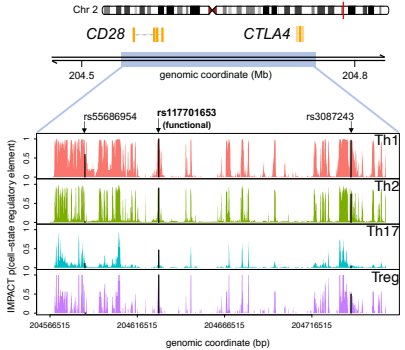
2. Causal variation of gene expression



3. Causal variation of polygenic traits



4. Functional fine-mapping



Thank You

Raychaudhuri Lab
Soumya Raychaudhuri
Yang Luo
Emma Davenport
Harm-Jan Westra

Price Lab
Alkes Price
Steven Gazal
Bryce van de Geijn

email: tamariuta@gmail.com
twitter: @TAmariuta

IMPACT manuscript on *bioRxiv*!