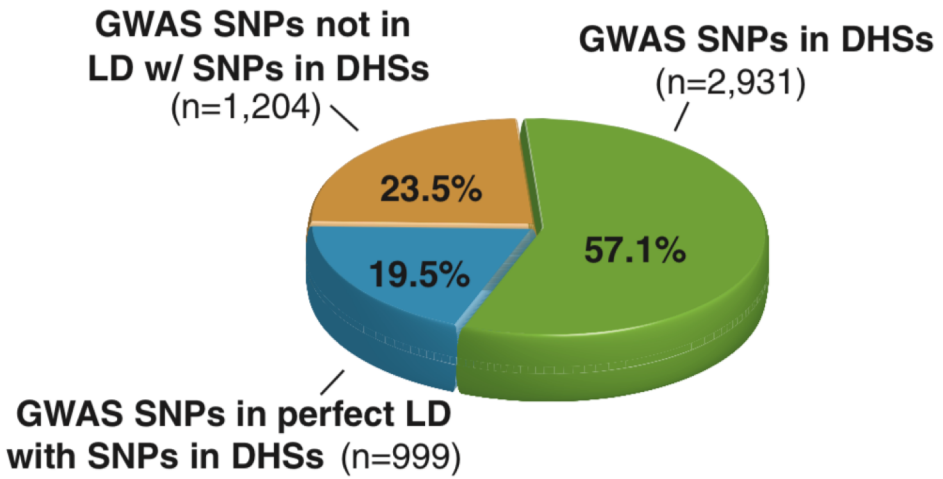
Heritability enrichment of regulatory elements is concentrated in elements with ancient sequence age and conserved function

Margaux Hujoel
Doctoral Student
Harvard T.H. Chan School of Public Health
Department of Biostatistics
ASHG 2018

Enhancers and promoters impact disease

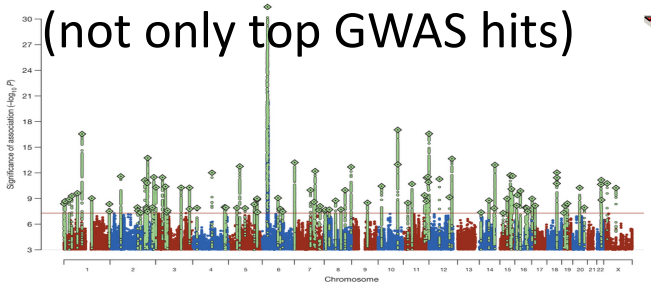
- Disease-associated variants and disease heritability are concentrated in regulatory elements, such as enhancer and promoters
- **Goal: to determine *which* enhancers and promoters are most important**



(Maurano et al. 2012 *Science*; Trynka et al. 2013 *Nat Genet*; Pickrell 2014 *AJHG*; Finucane et al. 2015 *Nat Genet*)

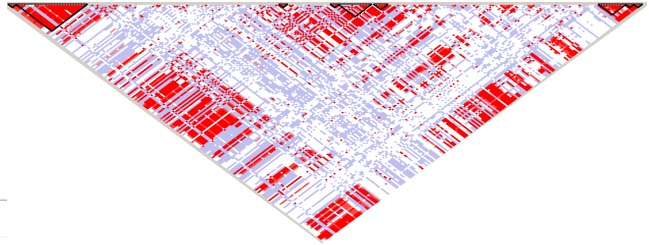
Stratified LD Score Regression (S-LDSC) is a method to partition disease heritability

GWAS summary statistics (not only top GWAS hits)

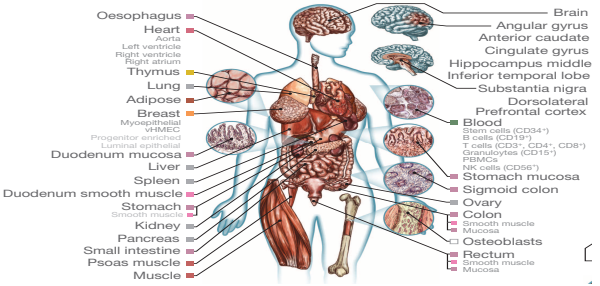


PGC-SCZ 2014 *Nature*

Reference panel (1000G)



Functional annotations



Roadmap Epigenomics Consortium 2015 *Nature*

S-LDSC with baseline-LD model (75 functional annotations)

Annotation **Enrichment**

$$\text{Enrichment} = \frac{\text{Prop. h}^2}{\text{Prop. SNPs}}$$

Annotation **Effect Size** (τ_c^*)

Conditional on other annotations

(Finucane et al. 2015 *Nat Genet*, Gazal et al. 2017 *Nat Genet*)

Summary of datasets analyzed

We utilize summary statistics for 41 independent traits and diseases ($\overline{N}$=320K).

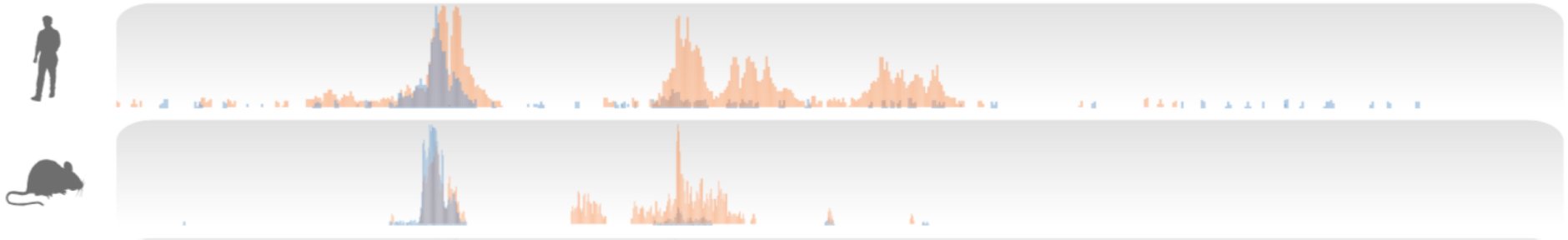
- UK Biobank summary statistics are publicly available
- Enrichment and τ* is meta-analyzed over 41 traits

Results are obtained by conditioning on the baseline-LD model (75 functional annotations) and enhancer and promoter annotations

(Gazal et al. 2017 *Nat Genet*; Loh et al. 2018 *Nat Genet*; Hormozdiari et al. 2018 *Nat Genet*)

Enhancer and promoter annotations

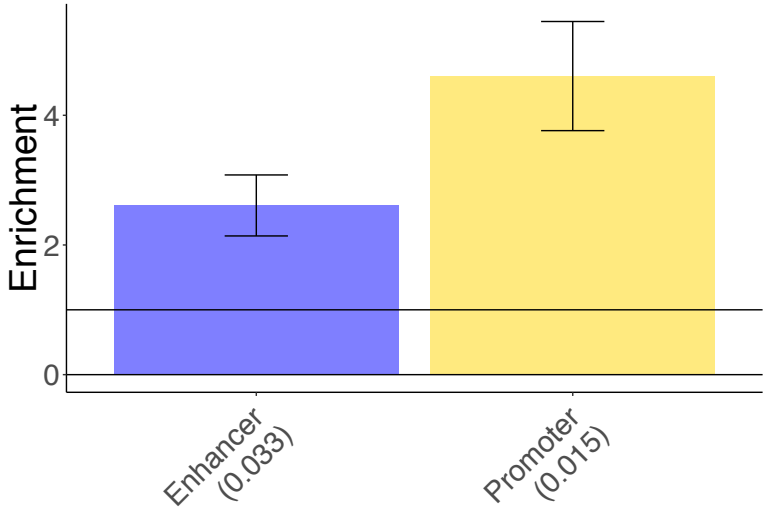
- Annotated enhancers and promoters using histone marks H3K27ac and H3K4me3
- 20 mammalian liver genomes
- Used only biologically reproducible peaks present in two or more replicates



(Villar et al. 2015 *Cell*)

Human enhancers and promoters are enriched for disease heritability

- Enhancers were 2.6x enriched (p=2.5e-12)
- Promoters were 4.6x enriched (p=3.2e-17)



(Villar et al. 2015 *Cell*)

Results meta-analyzed across 41 traits

Outline

① Enhancers and promoters with ancient sequence age

② Functionally conserved enhancers and promoters

③ Promoters of loss-of-function intolerant genes

④ All of the above are informative in a joint model

Outline

① Enhancers and promoters with ancient sequence age

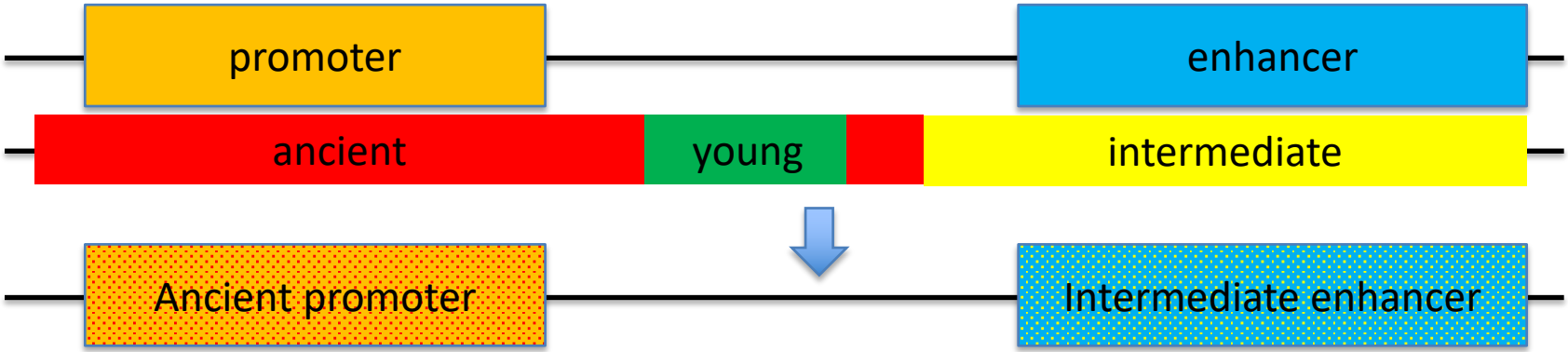
② Functionally conserved enhancers and promoters

③ Promoters of loss-of-function intolerant genes

④ All of the above are informative in a joint model

Underlying sequence age

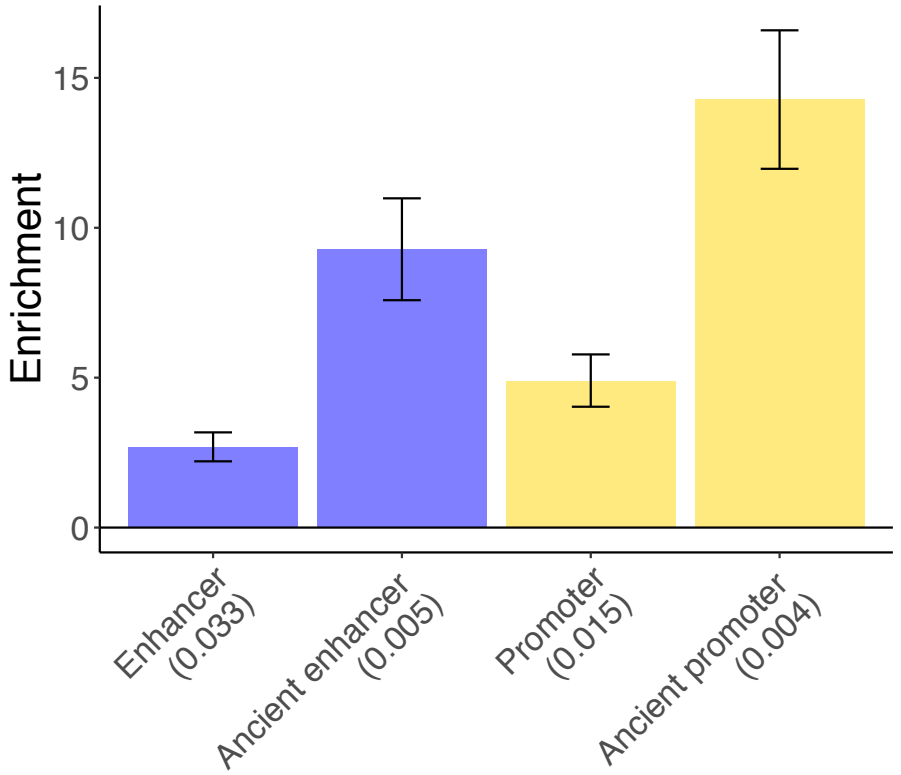
- Date the underlying sequence age of human enhancers and promoters
- Young (< 160 MYA), intermediate (~160 MYA), ancient (> 160 MYA)
- 16% of enhancers are ancient, 28% of promoters are ancient



(Marnetto et al. 2018 *AJHG*)

Heritability enrichment is concentrated in human elements with ancient sequence age

- Ancient human enhancers were 9.3x enriched, compared to 2.7x for all human enhancers (p=4e-15 for difference)
- Ancient human promoters were 14.3x enriched, compared to 4.9x for all human promoters (p=2e-18 for difference)



Results meta-analyzed across 41 traits (Hujoel et al. 2018 *bioRxiv*)

Outline

① Enhancers and promoters with ancient sequence age

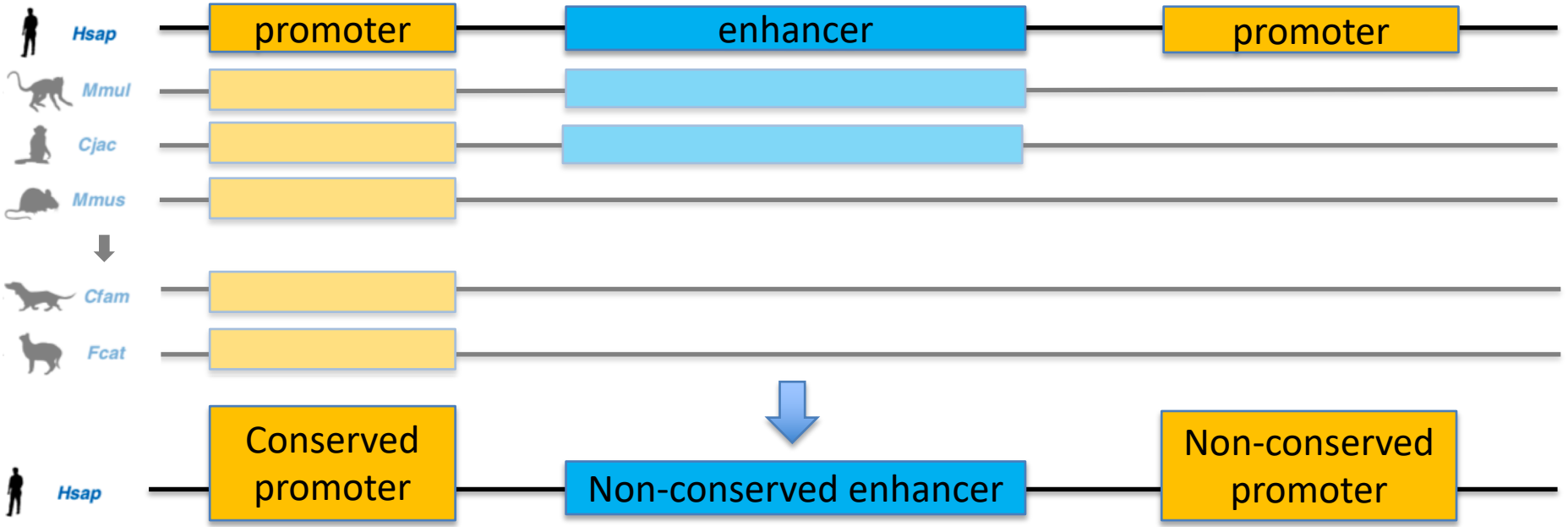
② Functionally conserved enhancers and promoters

③ Promoters of loss-of-function intolerant genes

④ All of the above are informative in a joint model

Conserved function across species

- Determine the number of mammalian species in which the human enhancer and promoter were also functional in
- conserved = conserved function in ≥ 5 of 9 other mammalian species
- Promoters tend to be more conserved (53% vs. 16%)

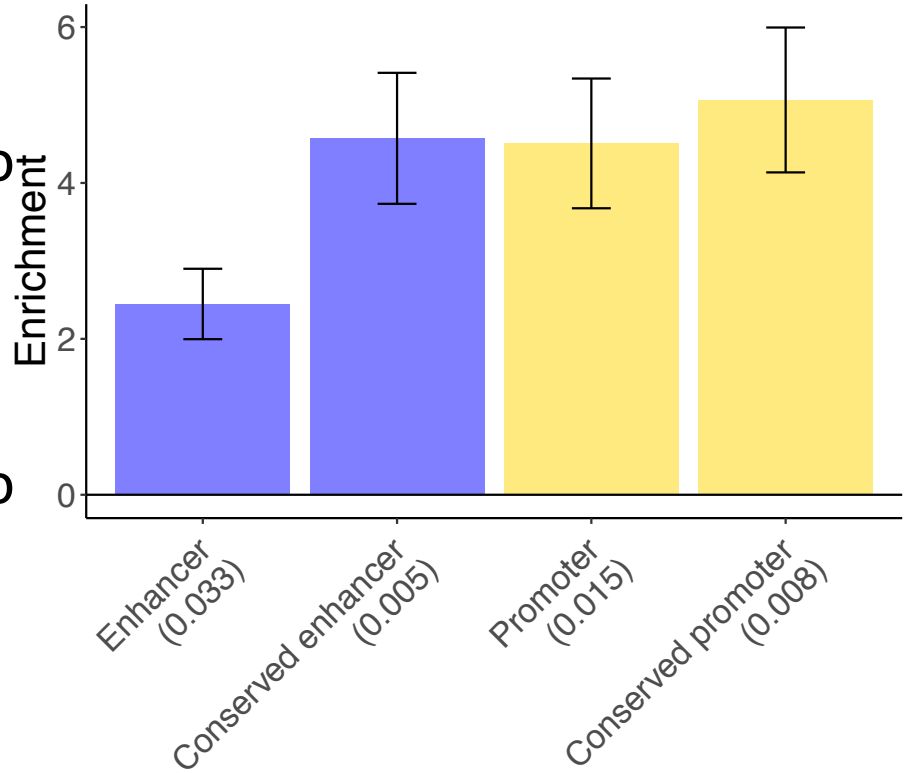


(Villar et al. 2015 *Cell;* Vermunt et al. 2016 *Nat Neuro*; Trizzino et al. 2017 *Genome Res*)

Heritability enrichment is concentrated in functionally conserved enhancers and promoters

- Conserved human enhancers were 4.6x enriched, compared to 2.4x for all human enhancers (p=3e-12 for difference)
- Conserved human promoters were 5.1x enriched, compared to 4.5x for all human promoters (p=0.022 for difference)

Results meta-analyzed across 41 traits



Outline

① Enhancers and promoters with ancient sequence age

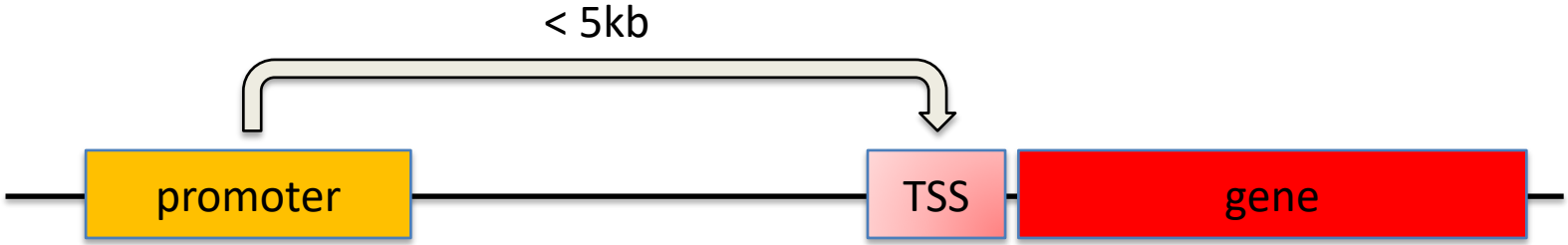
② Functionally conserved enhancers and promoters

③ Promoters of loss-of-function intolerant genes

④ All of the above are informative in a joint model

Target gene

- Identify the target gene of human promoters

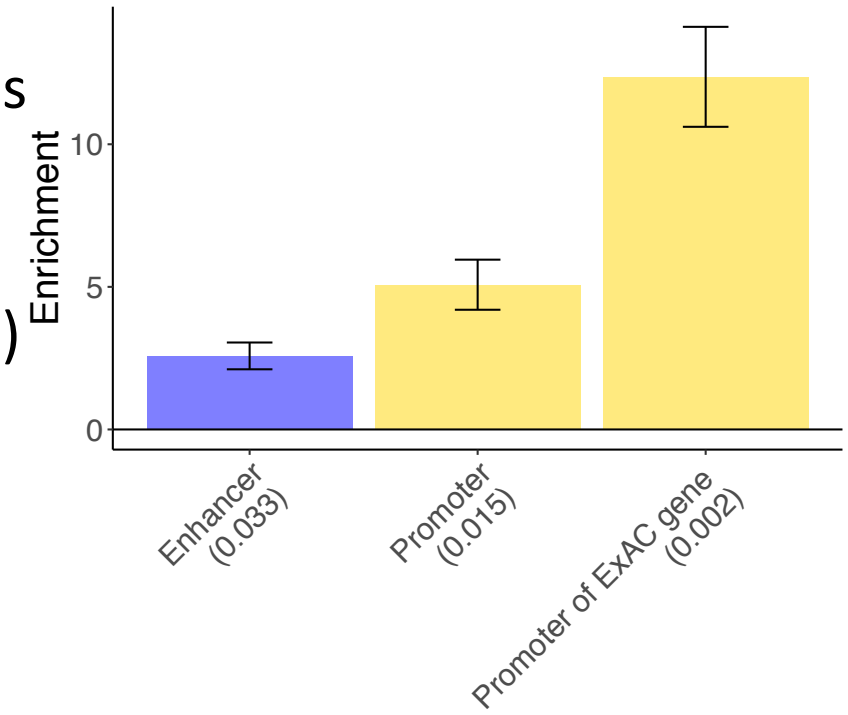


- Genes looked at:
  - ExAC loss-of-function (LoF) intolerant genes
  - Ancient genes (emerged before the vertebrates split; ~500 MYA)
  - Genes with a mouse ortholog

(Domazet-Loso and Tautz 2008 *Mol. Biol. Evol*;
Neme and Tautz 2013 *BMC Genomics*; Gao et al. 2018 *Cell*; Lek et al. 2016 *Nature*)

Heritability enrichment is concentrated in human promoters of ExAC LoF intolerant genes

- 16% of human promoters are promoters of ExAC Lof intolerant genes
- Promoters of ExAC genes were 12.4x enriched, compared to 5.1x for all promoters (p=9e-16 for the difference)
- Annotations for promoters of ancient genes or genes with a mouse ortholog were not conditionally significant



Results meta-analyzed across 41 traits

Outline

① Enhancers and promoters with ancient sequence age

② Functionally conserved enhancers and promoters

③ Promoters of loss-of-function intolerant genes

④ All of the above are informative in a joint model

Obtaining final joint model

Included annotations relating to:

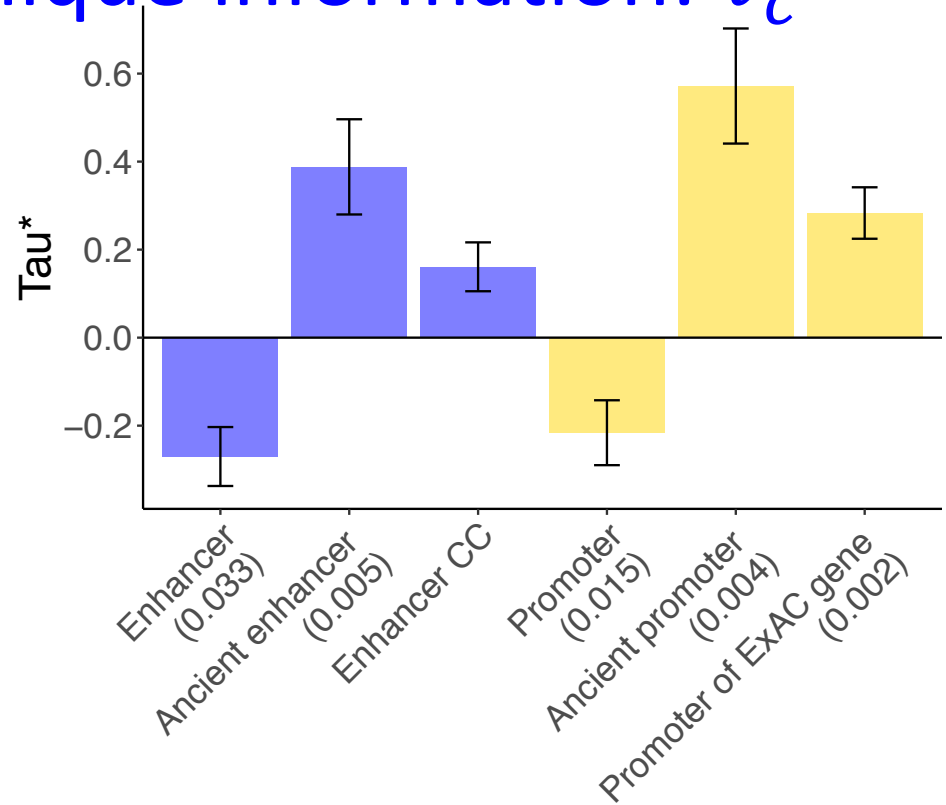
- Sequence age: enhancer/promoter w/ ancient sequence age
- Conserved function: enhancer/promoter conservation count
- Target gene: promoter of ExAC LoF intolerant genes

Iteratively removed annotations that were not conditionally significant (adjusting for multiple testing; based on τ_c^* p-value)

Sequence age, conserved function, and target gene each provide unique information: τ_c^*

Final model included

- Ancient enhancer
- Enhancer conservation count
- Ancient promoter
- Promoter of ExAC LoF intolerant gene

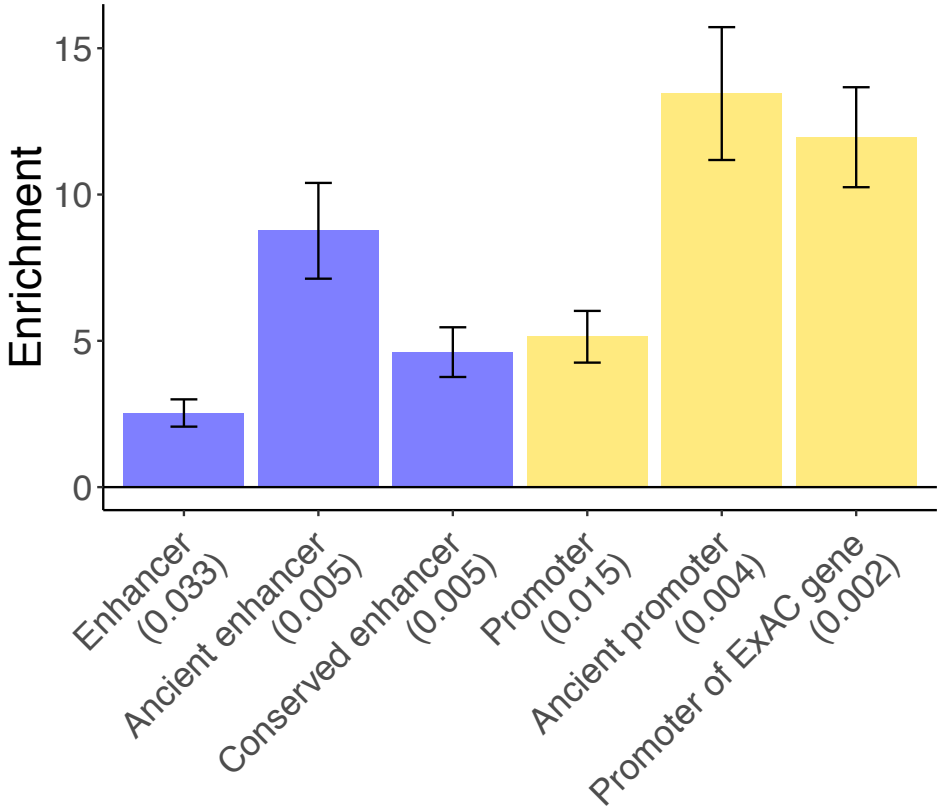


Results meta-analyzed across 41 traits

τ_c^*: the proportionate change in per-SNP h^2 per one s.d. increase in the value of the annotation

Sequence age, conserved function, and target gene provide unique information: enrichment

- Enrichment remains consistent with previous models
- Within both enhancers and promoters, heritability is particularly concentrated within elements with ancient sequence age

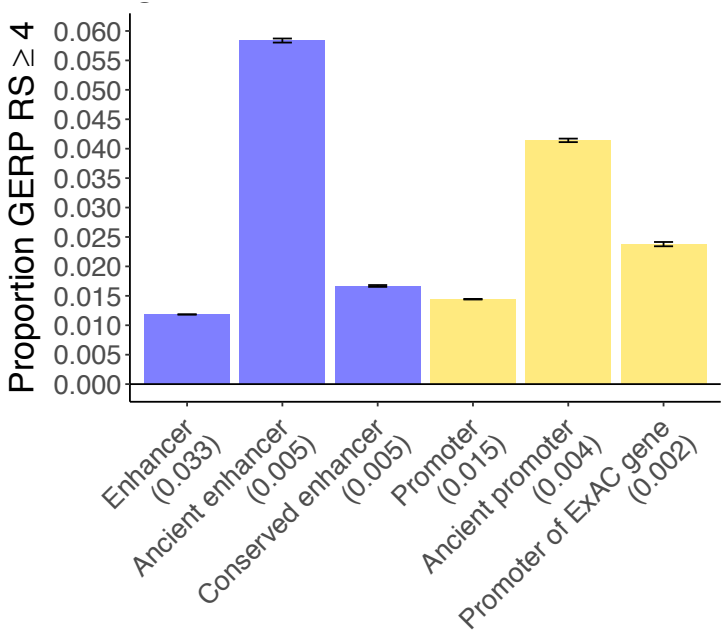


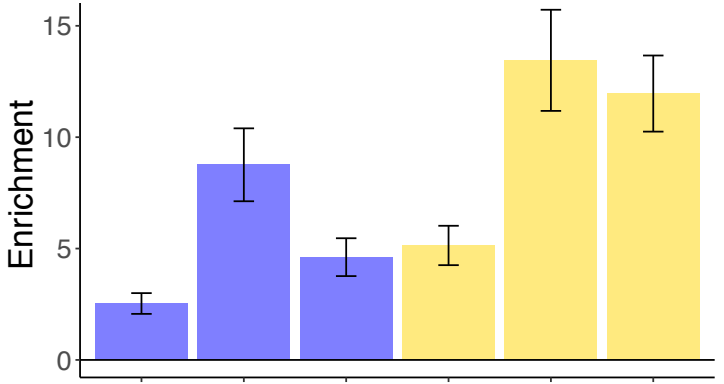
Results meta-analyzed across 41 traits

Negative selection metrics mirror these findings

- GERP RS score: a larger score is indicative of stronger negative selection
- Larger proportion of SNPs under stronger negative selection within elements that are more enriched
- Similar results for other measures of negative selection

(Gazal et al. 2017 *Nat Genet;*
Davydov et al. 2010 *PLoS Comput Biol*)





Conclusions

- Heritability enrichment within human enhancers and promoters is concentrated:
  - In elements with ancient sequence age
  - In elements with conserved function across many species
  - In promoters of ExAC LoF intolerant genes

- The mean value of several measures of negative selection within these genomic annotations mirrored all of these findings.

Acknowledgments

Steven Gazal
Farhad Hormozdiari
Bryce van de Geijn
Alkes Price

P. Flicek, P. Provero, D. Marnetto, H. Finucane

Additional thanks to UK Biobank

Disease heritability enrichment of regulatory elements is concentrated in elements with ancient sequence age and conserved function across species
Hujoel *et al.* 2018 **bioRxiv:420166**