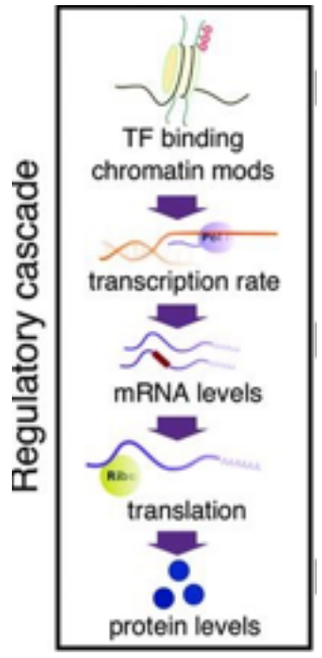
HUMAN GENETICS

RNA splicing is a primary link between genetic variation and disease

Yang I. Li,[1] Bryce van de Geijn,[2] Anil Raj,[1] David A. Knowles,[3,4] Allegra A. Petti,[5] David Golan,[1] Yoav Gilad,[2*] Jonathan K. Pritchard[1,6,7*]

JClub
Donghoon Lee
160725

What is the aim of the paper?



noncoding variants → *?* → **complex traits & diseases**

lack full understanding of how they act

What were known?

- eQTLs are highly enriched among the risk loci for complex diseases

- A large fraction of eQTLs are due to SNPs that affect TF binding or other aspects of chromatin function at enhancers or promoters

- Genetic variation might also affect gene regulation and function through pre-mRNA splicing (conflicting reports)

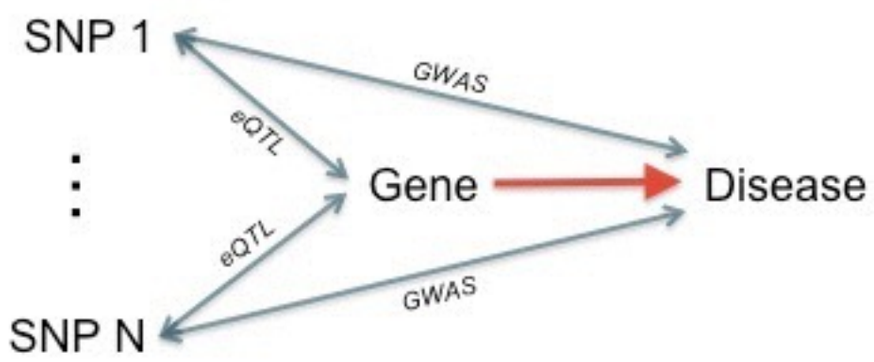
- QTL: locations of quantitative traits (traits or phenotypes that can be measured. i.e., height or skin pigmentation) in the genome

- eQTL: expression-QTL, how a given genotype (the DNA variants) at a particular QTL affects (increase or decrease) gene expression at that locus

- sQTL: splicing-QTL, how a given genotype affect alternative splicing pattern of mRNA precursors

- haQTL: histone-acetylation-QTL

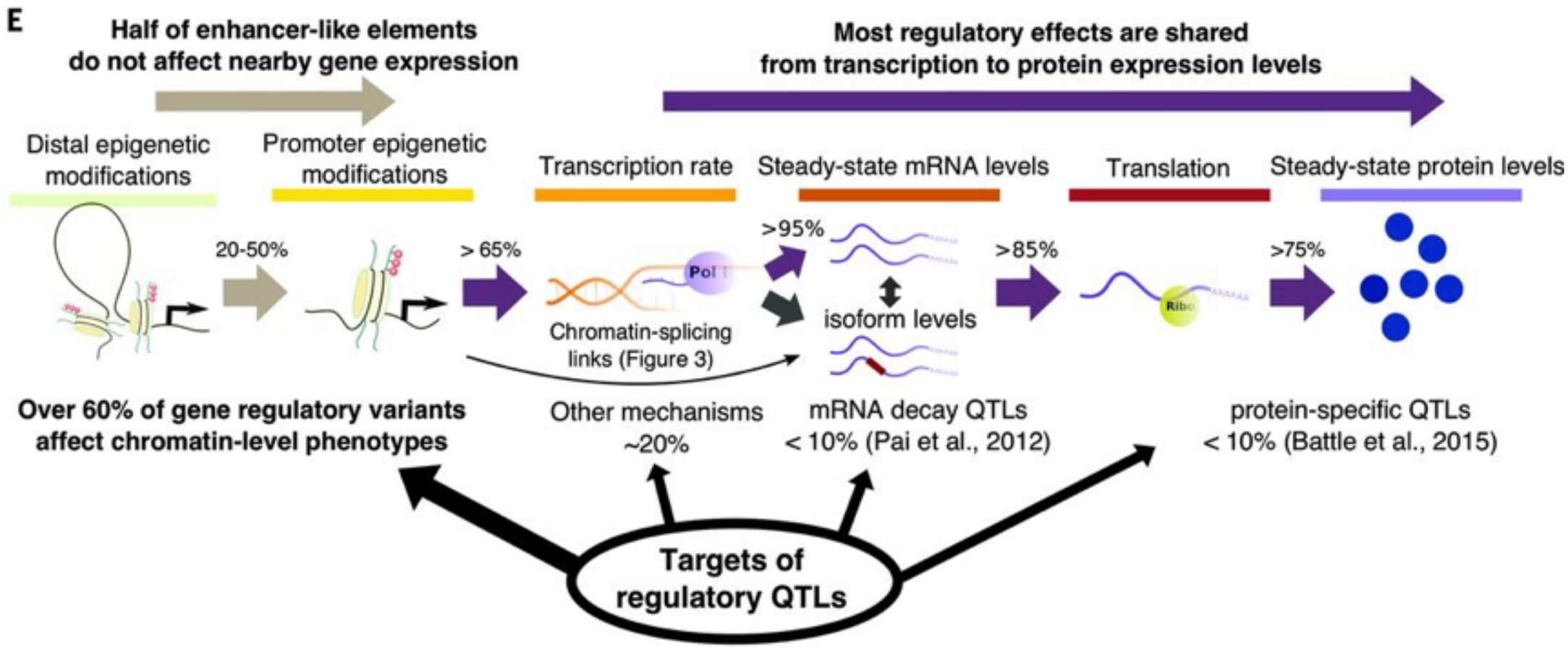
- xxxQTL: you-name-it



What did they find?

Quick Summary

noncoding variants → ***RNA Splicing*** → **complex traits & diseases**

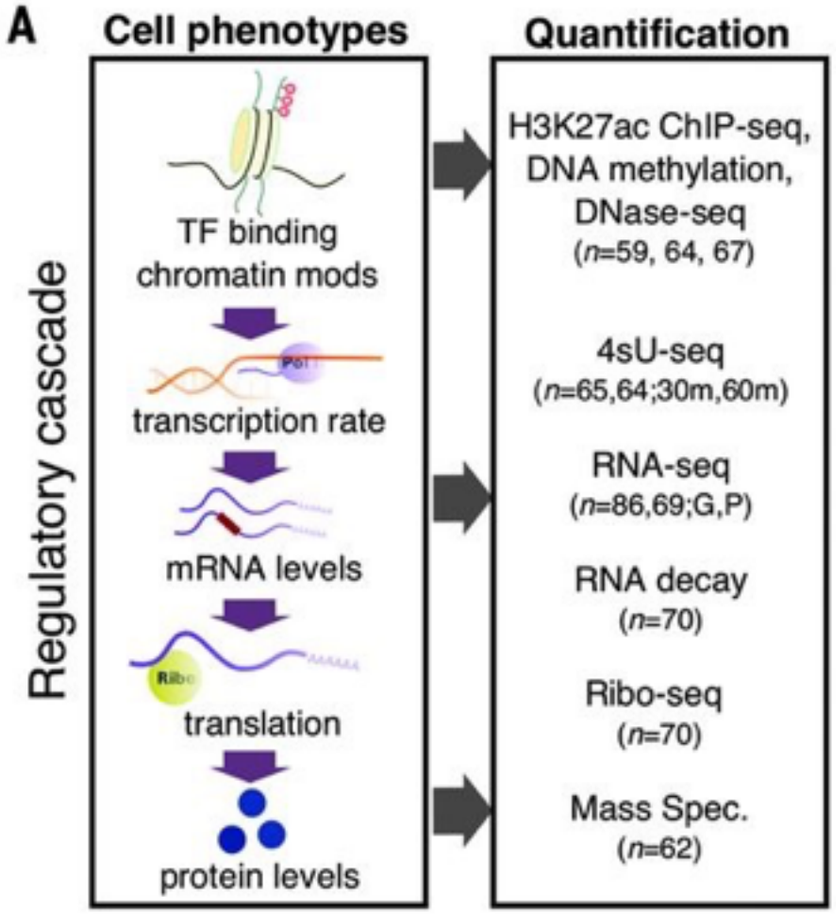


What did they do?

genetic variants

human lymphoblastoid cell lines (LCLs)

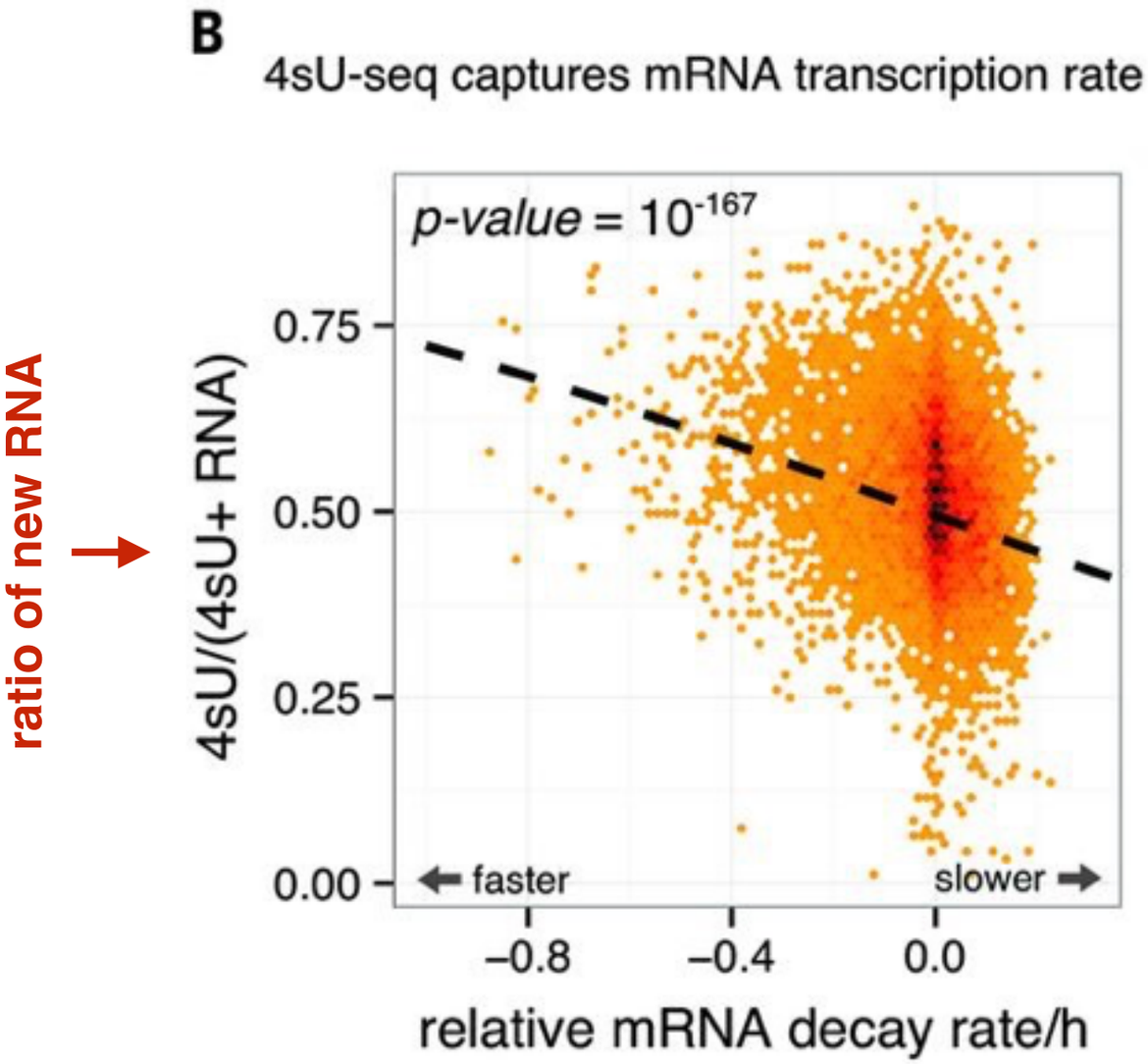
chromatin modifications
chromatin accessibility
DNA methylation
mRNA levels
transcription rate
RNA decay rate
translation levels
protein levels



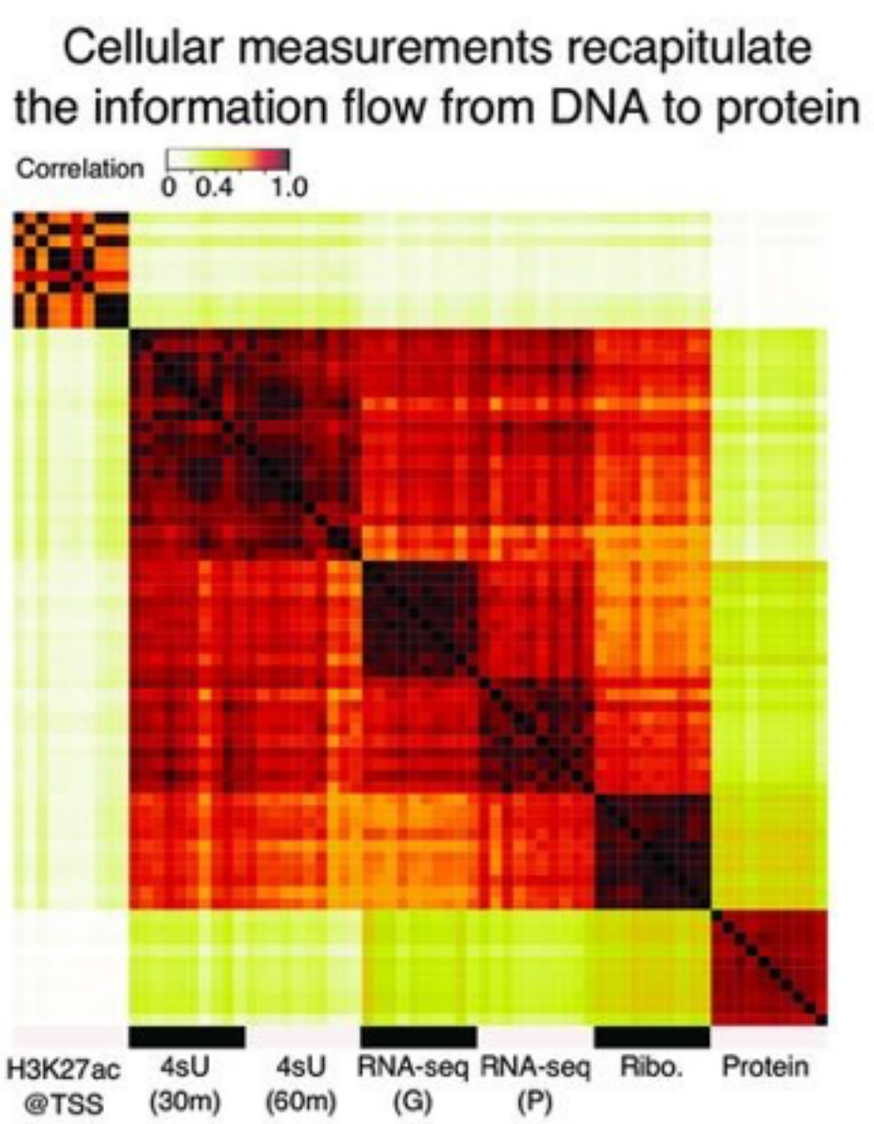
Data	Accession
H3K27ac	GSE58852 (GEO)
DNA methylation	GSE57483 (GEO)
DNase-seq	GSE31388 (GEO)
4sU-seq	GSE75220 (GEO) ← NEW
RNA-seq (Pickrell)	GSE19480 (GEO)
RNA-seq (GEUVADIS)	E-GEUV-3 (ArrayExpress)
RNA decay	GSE37451 (GEO)
ribo-seq	GSE61742 (GEO)
protein	PXD001406 (ProteomeXchange)

Table S8: Location of datasets used to call QTLs in this study.

all 8: 32 samples
≥ 6: 68 samples



steady-state = balance between transcription and decay



sequential ordered regulatory cascade:
promoter activity -> txn rates -> mRNA exp. lvls
-> translation lvls -> protein exp. lvls

All 1000G P1 SNPs with MAF ≥ 0.05 ±100kb genes

⇩

WASP: allele-specific software for robust molecular quantitative trait locus discovery

Bryce van de Geijn, Graham McVicker, Yoav Gilad & Jonathan K Pritchard

Affiliations | Contributions | Corresponding author

Nature Methods **12**, 1061–1063 (2015) | doi:10.1038/nmeth.3582
Received 23 June 2015 | Accepted 13 August 2015 | Published online 14 September 2015

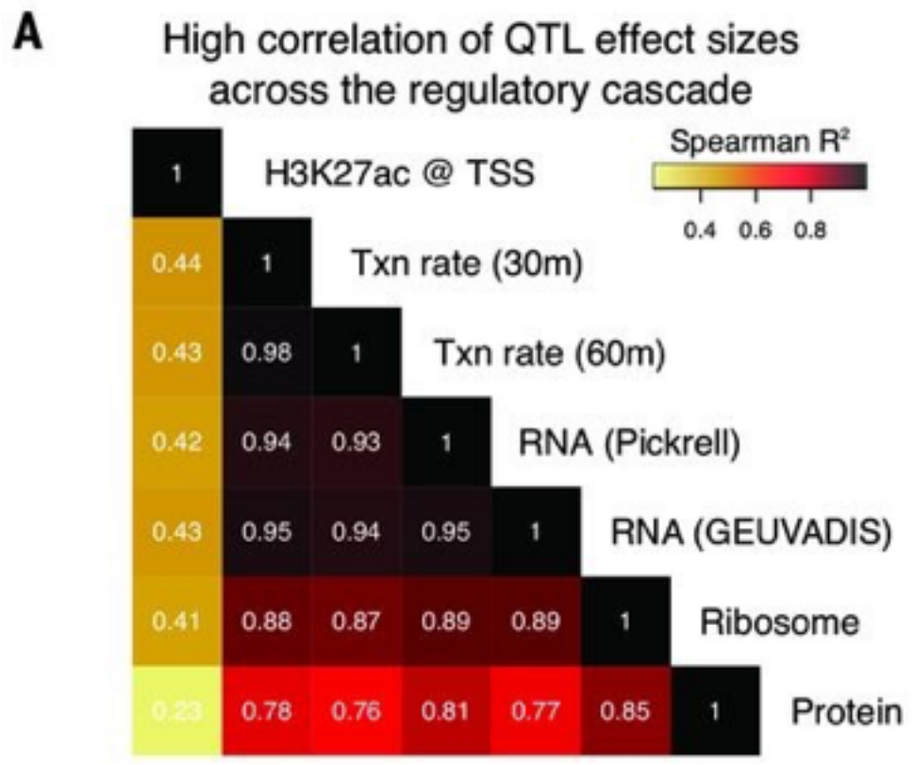
Allele-specific sequencing reads => QTLs

⇩

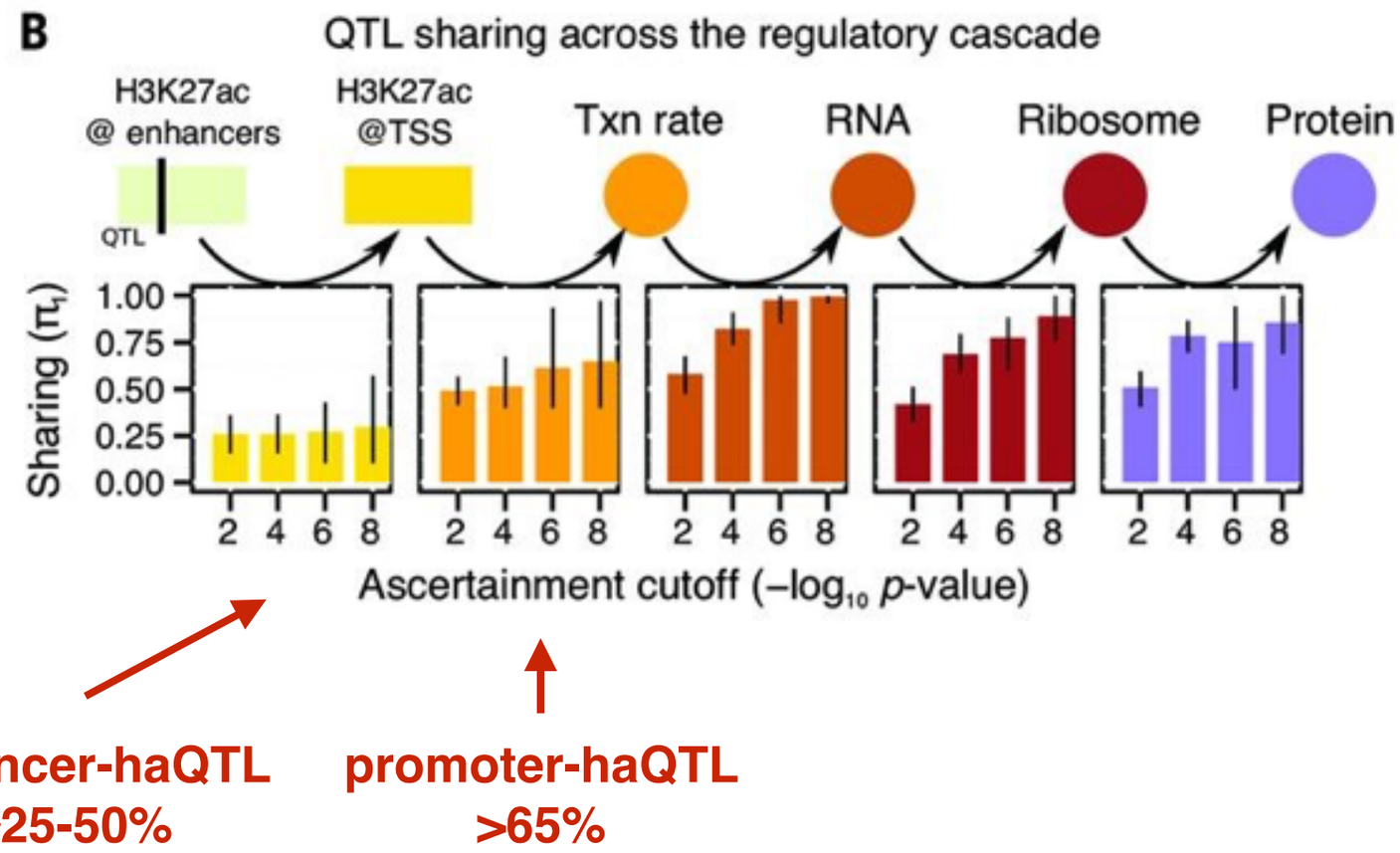
Mapping of QTLs across the 8 molecular phenotypes

https://www.encodeproject.org/software/wasp/

set of significant QTLs with different cutoff -> estimate sharing across the regulatory cascade



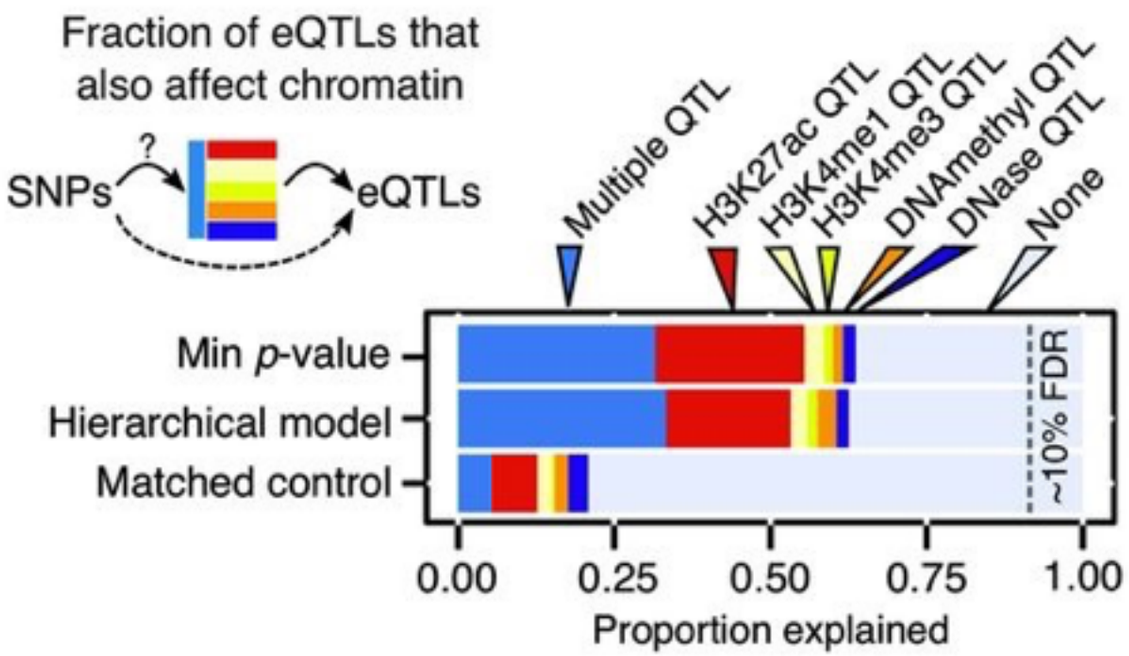
high correlation implies high proportion of eQTL sharing



enhancer-haQTL ~25-50%

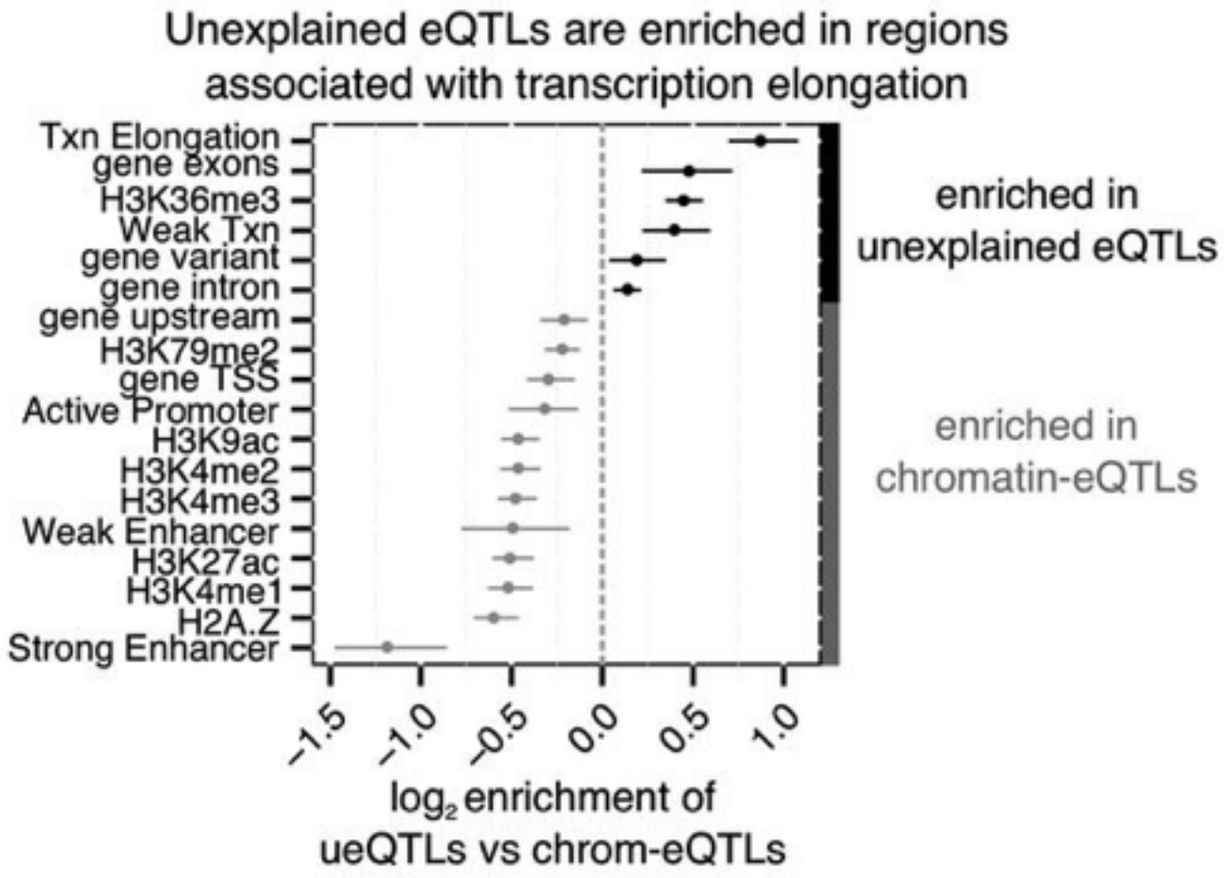
promoter-haQTL >65%

C

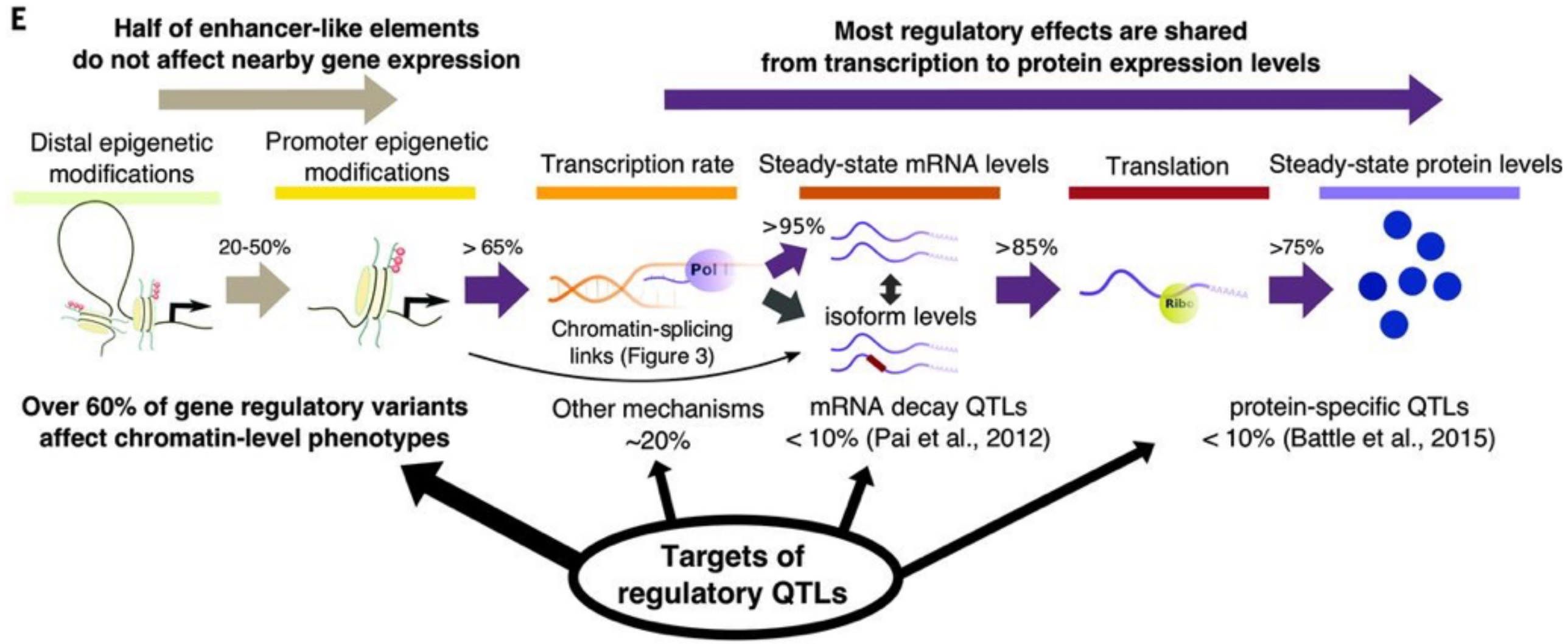


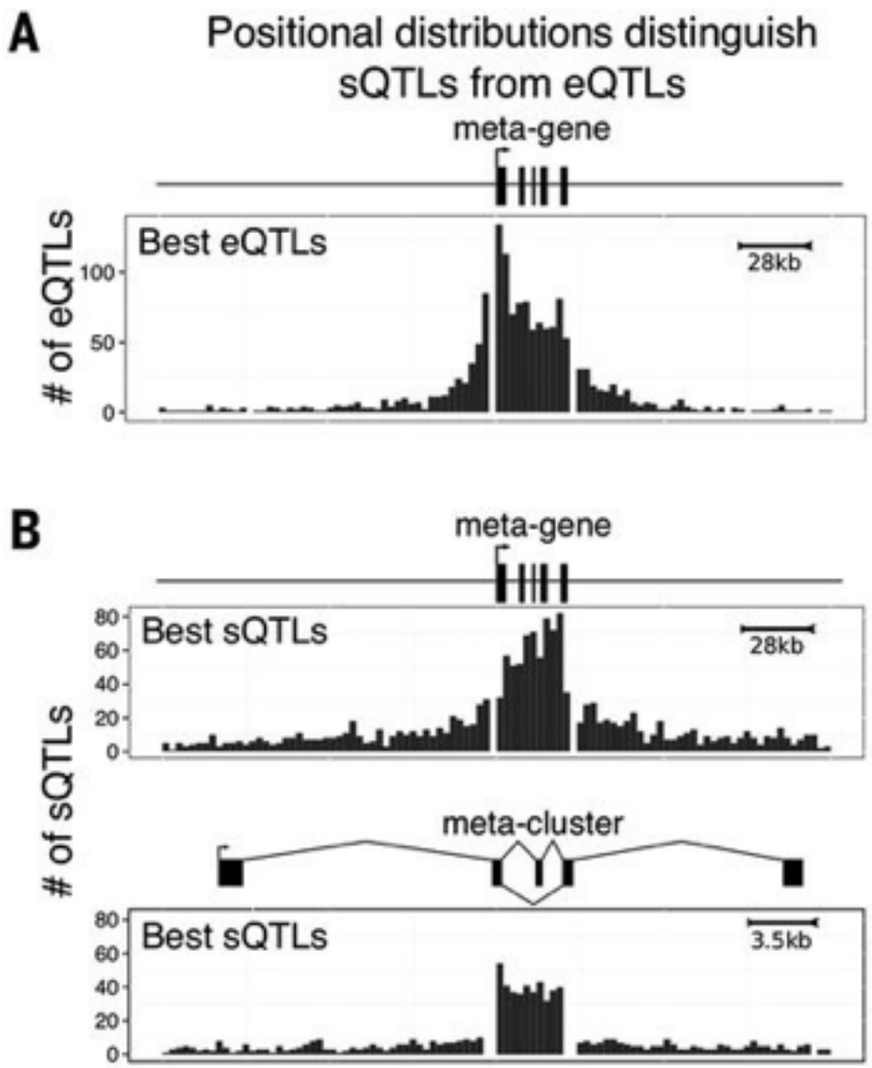
~ 65% eQTL also affect chromatin

D

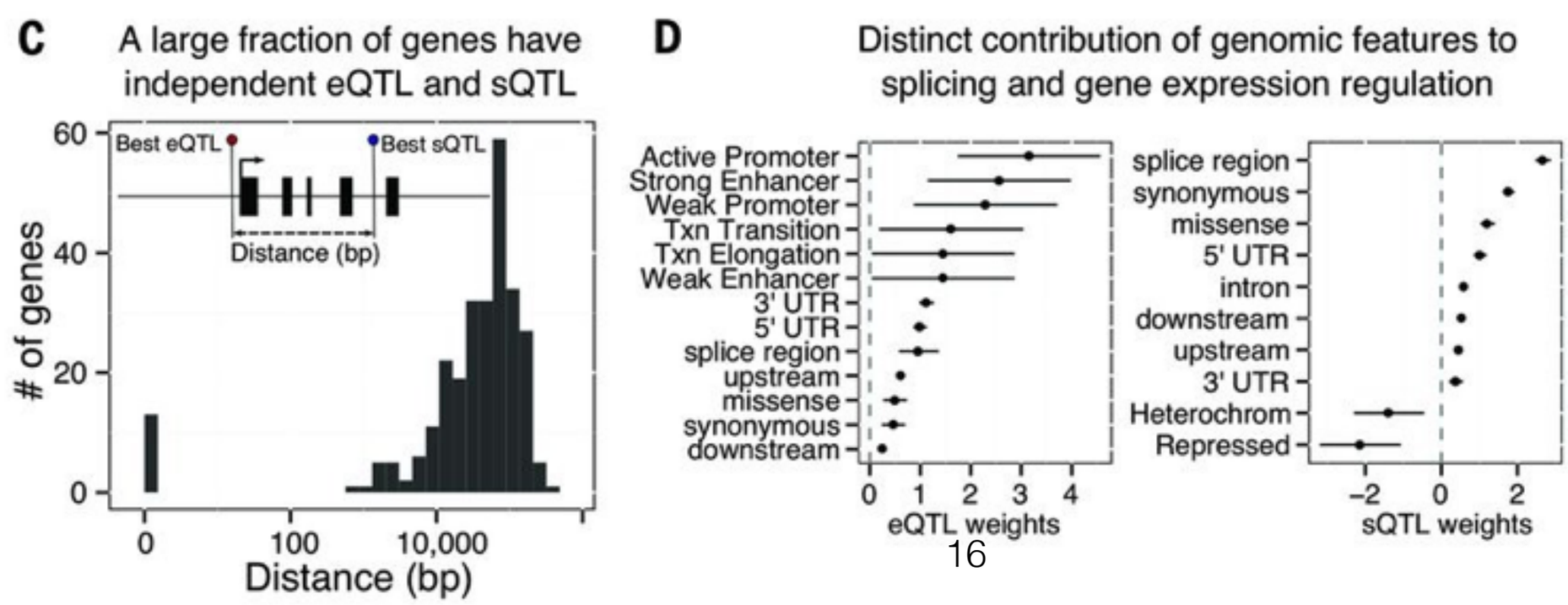


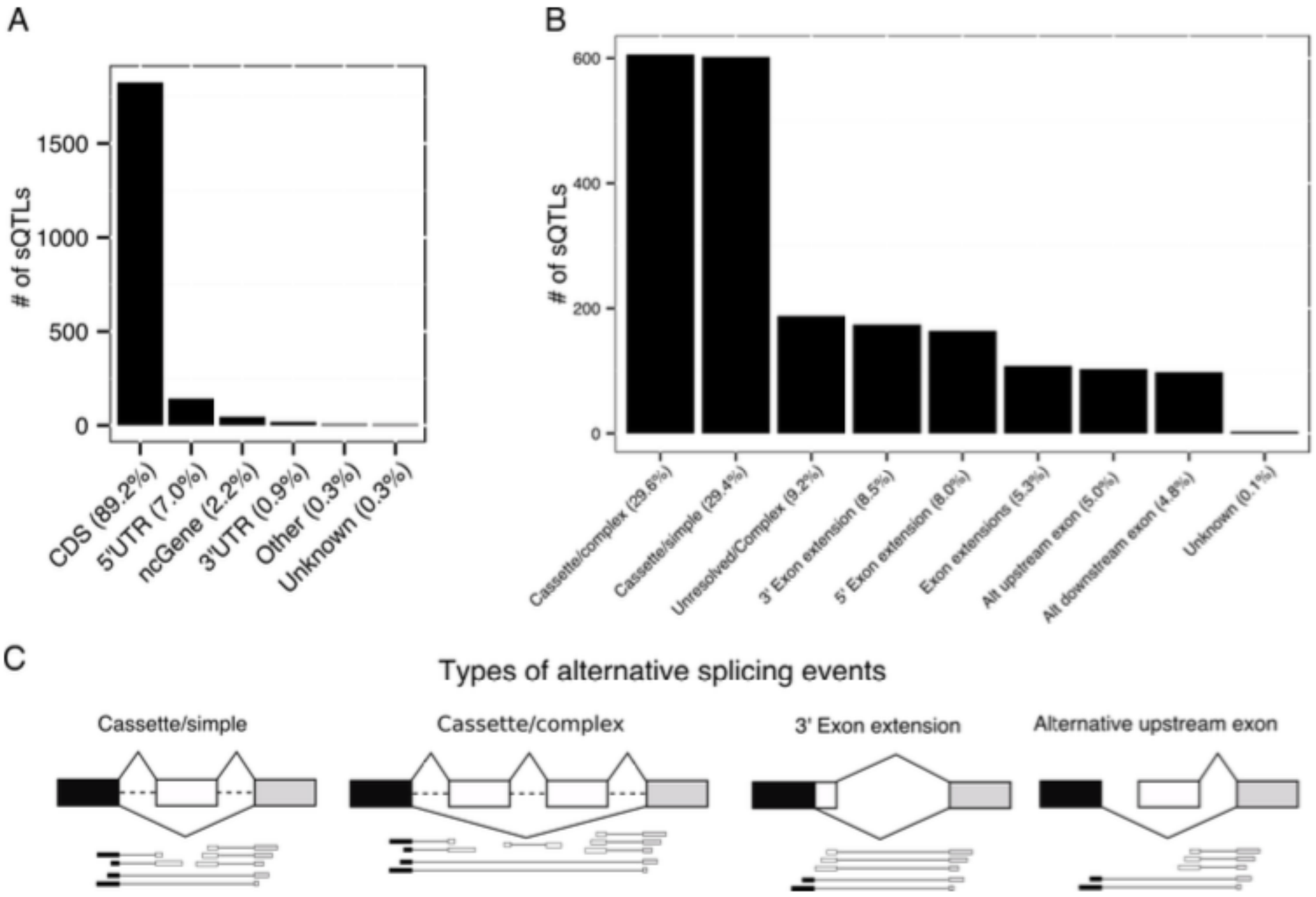
~ 35% (unexplained) eQTL “might be” associated with txn elongation


E
Half of enhancer-like elements
do not affect nearby gene expression
Most regulatory effects are shared
from transcription to protein expression levels
Distal epigenetic modifications
Promoter epigenetic modifications
Transcription rate
Steady-state mRNA levels
Translation
Steady-state protein levels
20-50%
> 65%
>95%
>85%
>75%
Pol II
Chromatin-splicing links (Figure 3)
isoform levels
Ribo
Over 60% of gene regulatory variants affect chromatin-level phenotypes
Other mechanisms ~20%
mRNA decay QTLs < 10% (Pai et al., 2012)
protein-specific QTLs < 10% (Battle et al., 2015)
Targets of regulatory QTLs

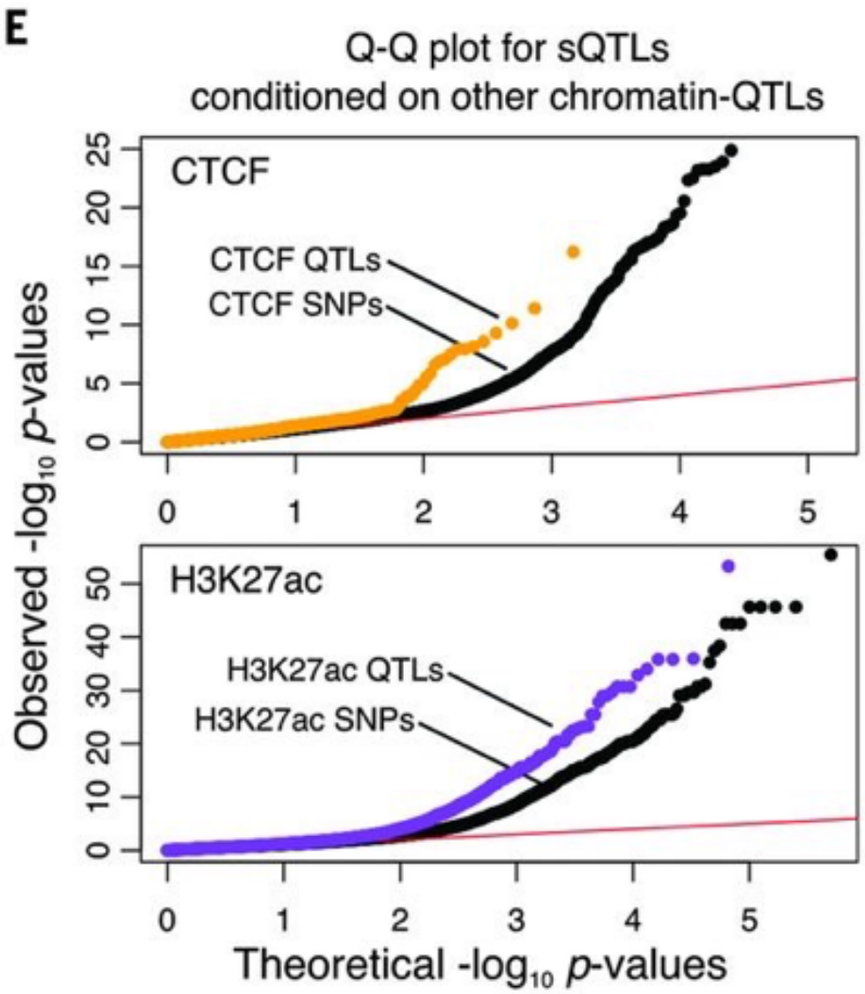


- **LeftCutter (deposited in biorxiv), which uses split reads, identified 2,983 sQTLs**
- **eQTLs are enriched near TSS, and sQTLs are enriched within gene bodies (esp. introns)**
- **sQTLs and eQTLs tend to be independent => then how do they affect traits?**





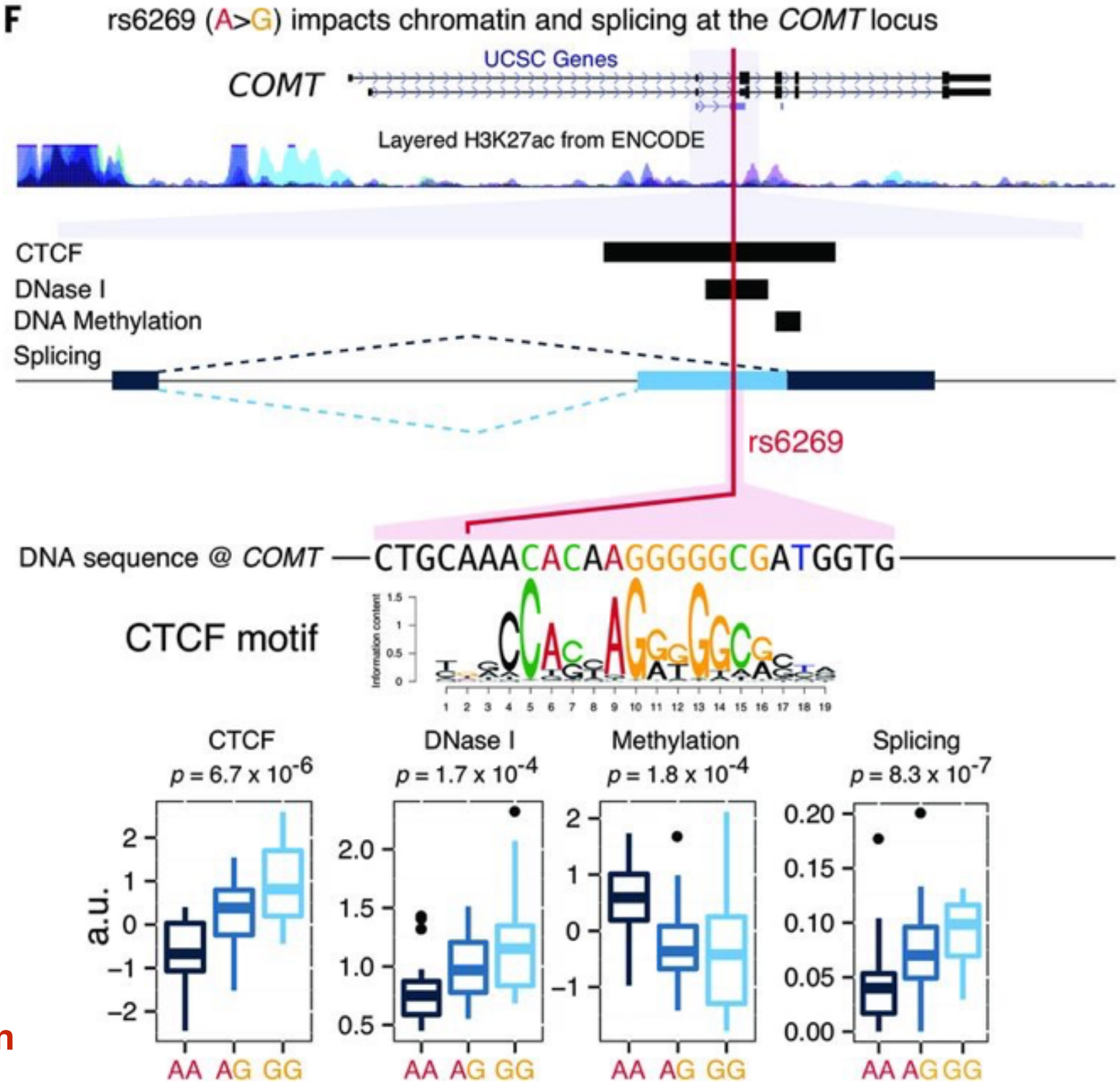
Although most sQTLs do not affect expression, ~90% affect CDS, meaning it could potentially affect protein function



chrom-QTLs (CTCF/H3K27ac) more likely to affect sQTLs than matched control SNPs

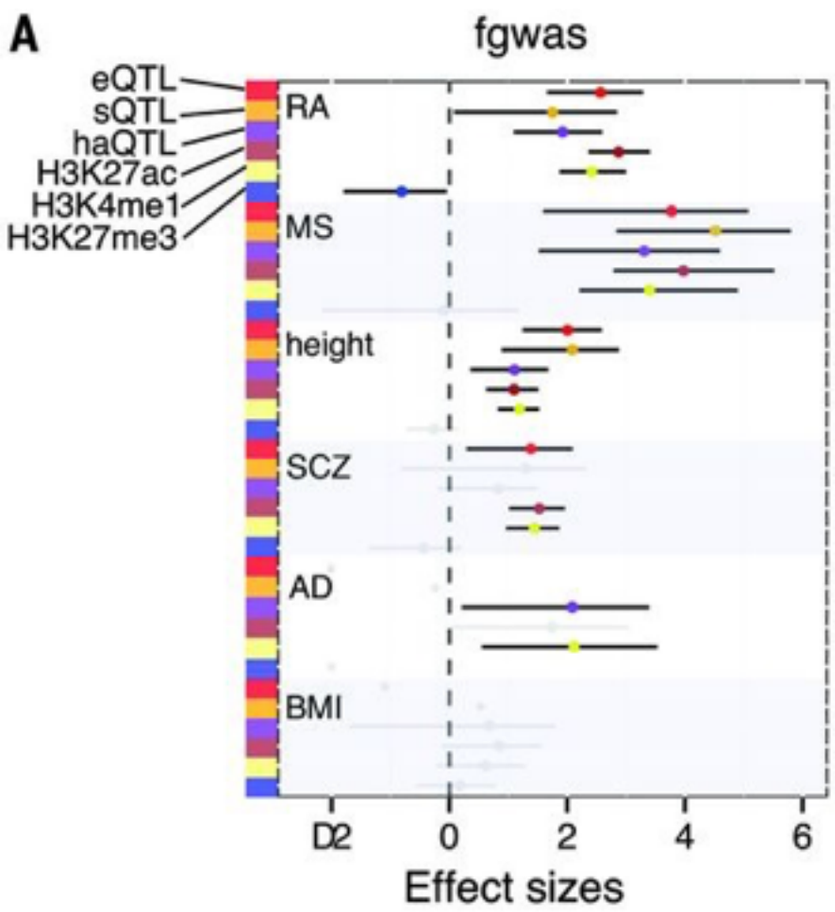
sQTLs were modestly enriched for association with chromatin-level phenotypes

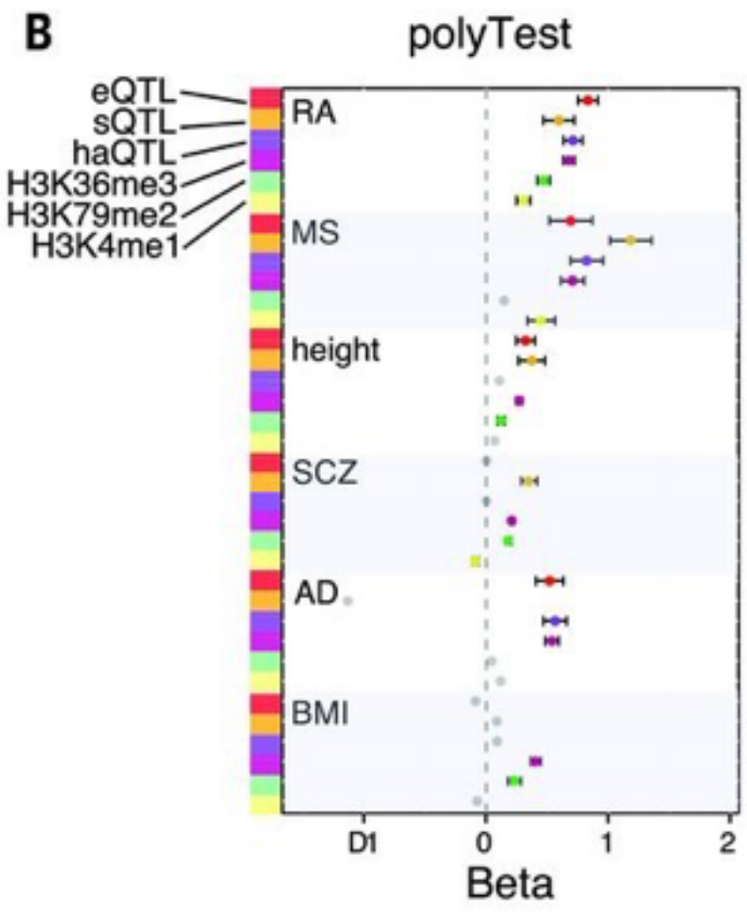
direct evidence that genetic variation can affect splicing by altering chromatin-level traits

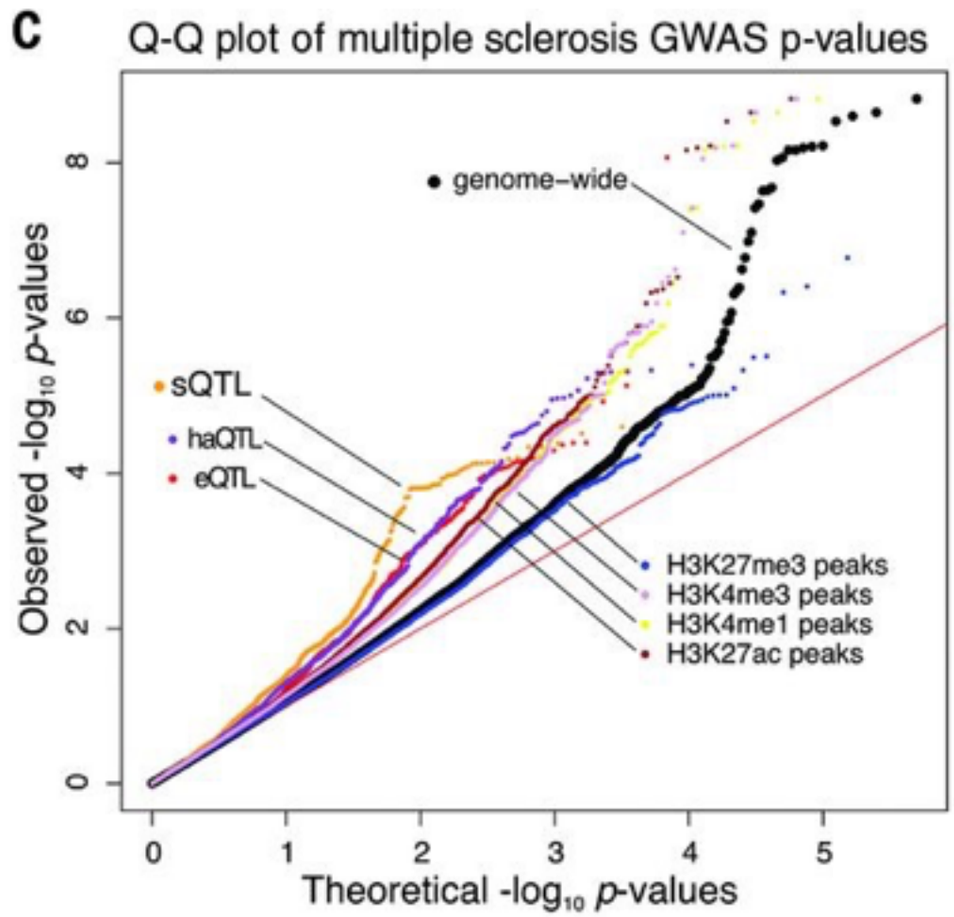


As expected, eQTLs associated with complex traits/diseases

sQTLs were enriched to a similar extent or in the case of multiple sclerosis to an even greater extent than eQTLs

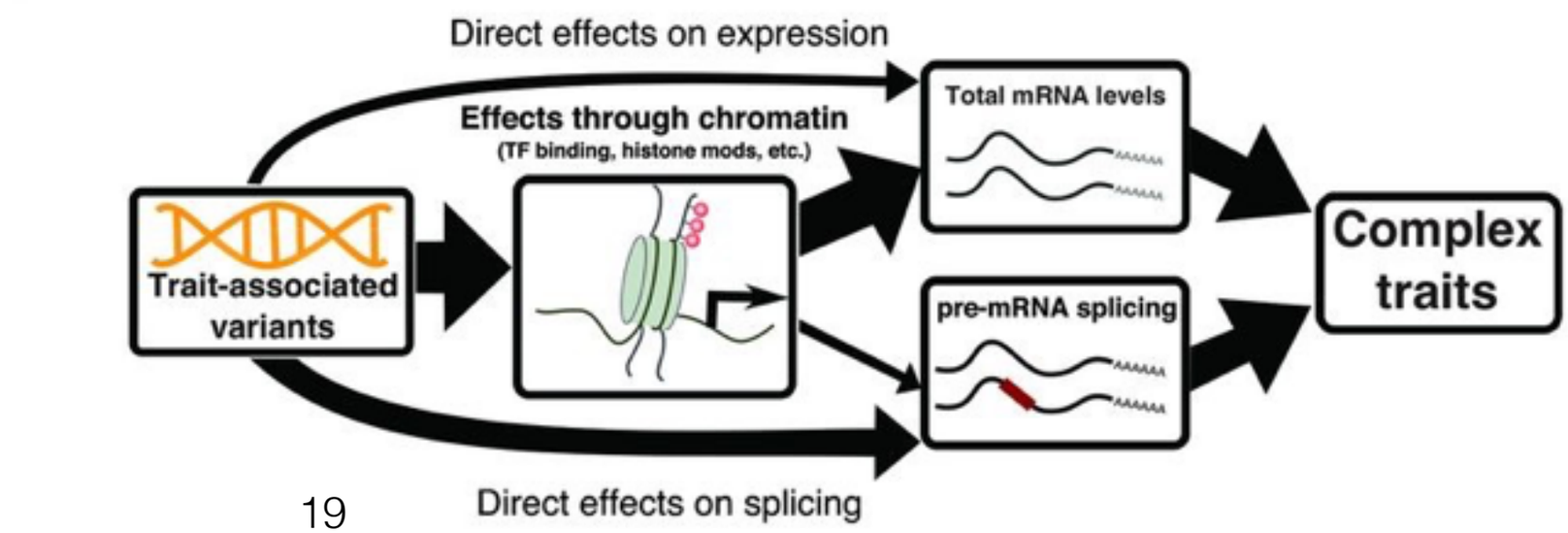




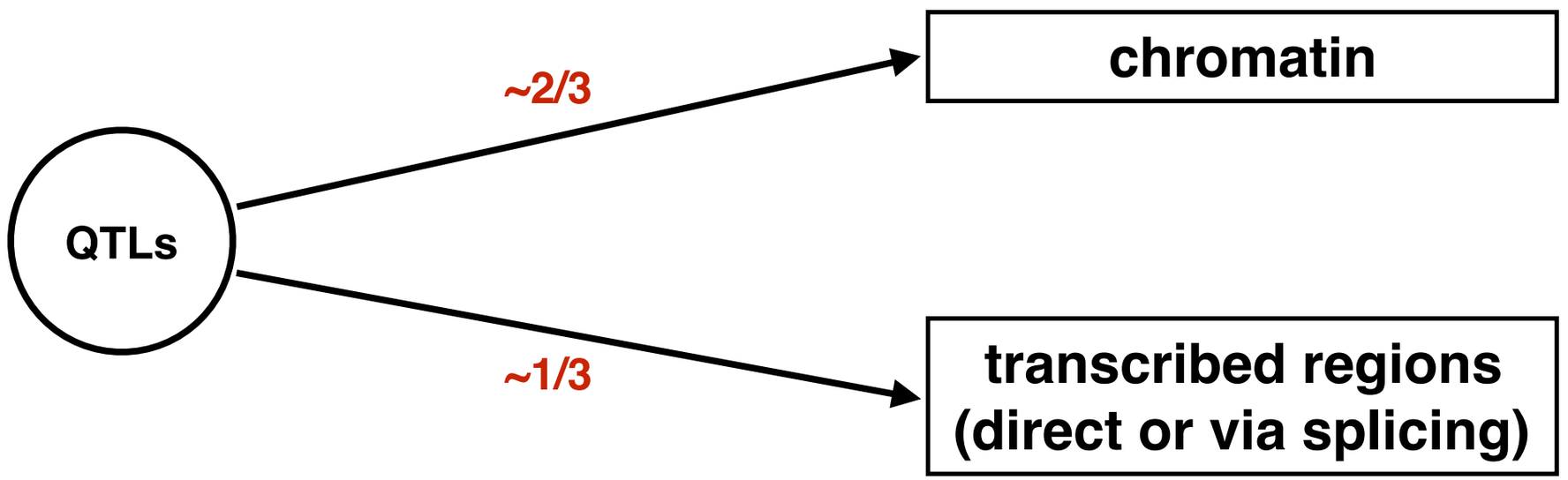


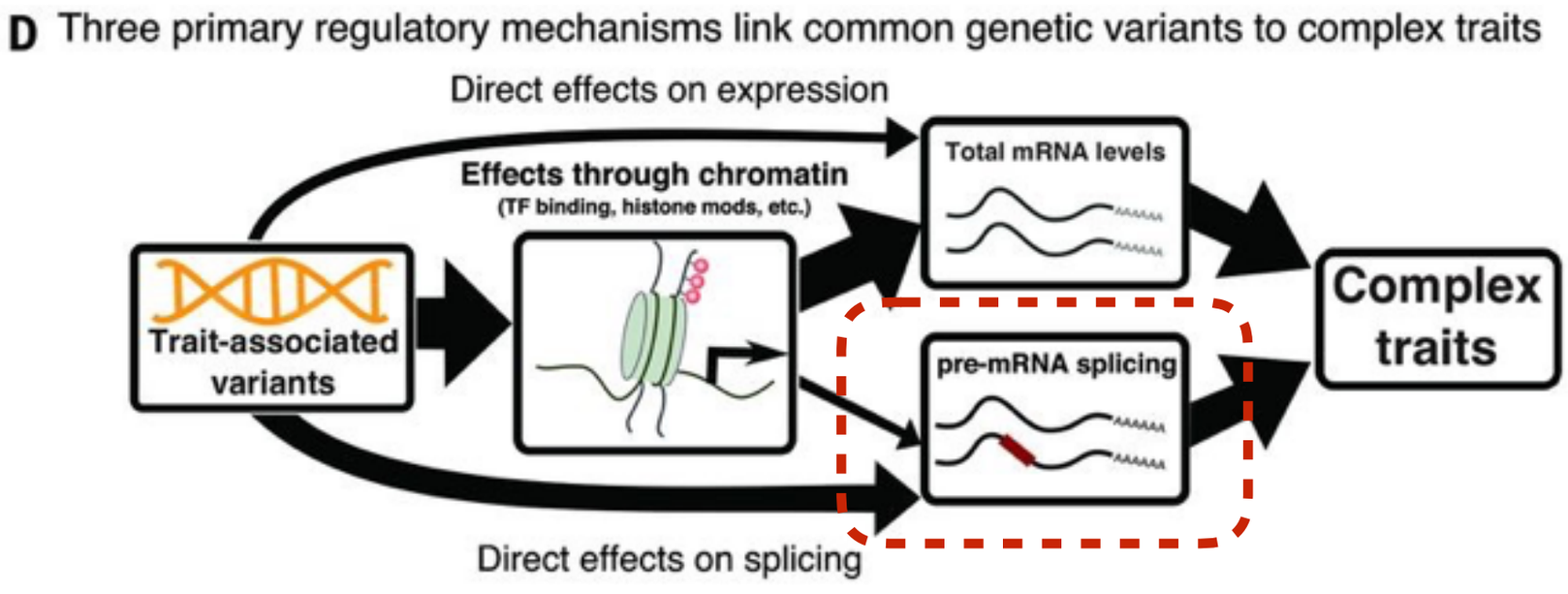
important role of RNA splicing in modulating phenotypic traits

variants that affect splicing make a major contribution to the genetics of complex diseases



Conclusion


~2/3
chromatin
QTLs
~1/3
transcribed regions
(direct or via splicing)


D Three primary regulatory mechanisms link common genetic variants to complex traits
Direct effects on expression
Effects through chromatin
(TF binding, histone mods, etc.)
Total mRNA levels
Trait-associated variants
pre-mRNA splicing
Complex traits
Direct effects on splicing

Critique

- Used 1 sample (YRI LCLs)
- Used 1 histone ChIP-seq (H3K27ac, active txn)
    - H3K36me3 has known to influence alt. splicing as well as nucleosome positioning, replication timing