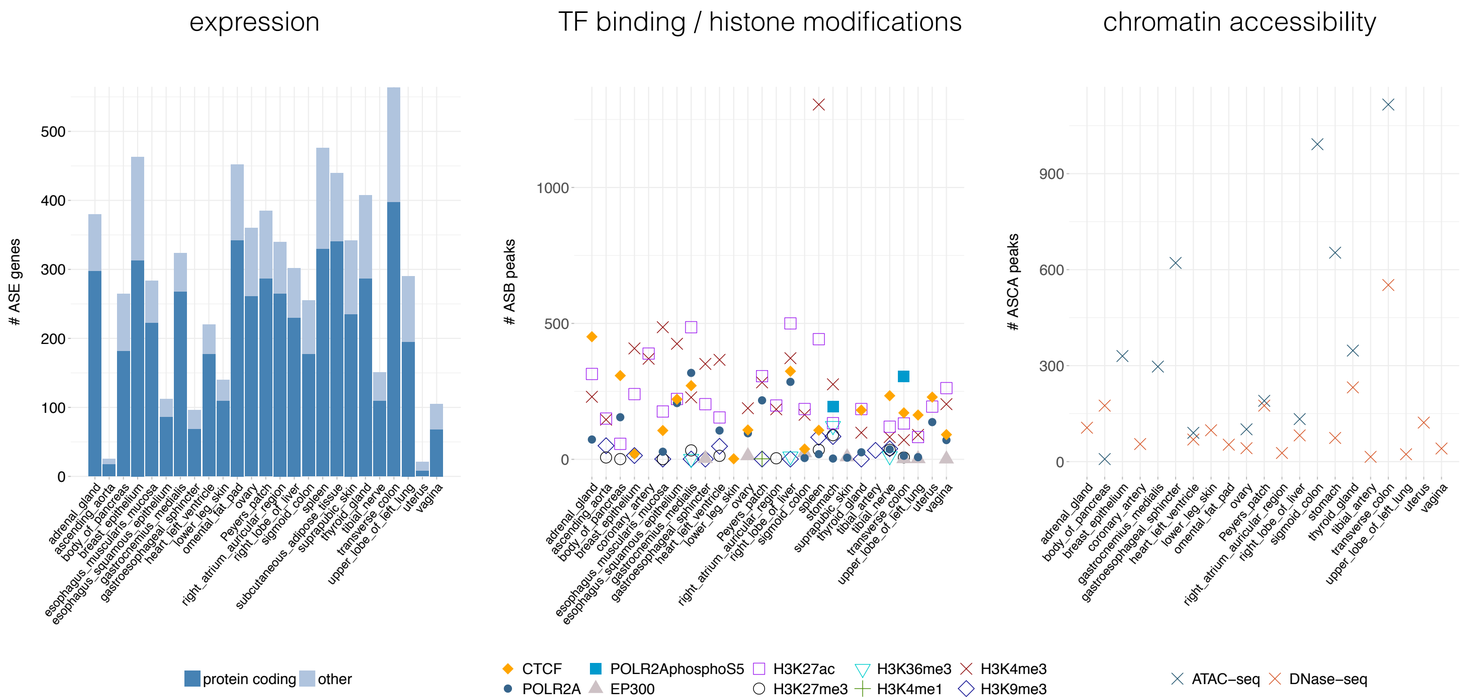
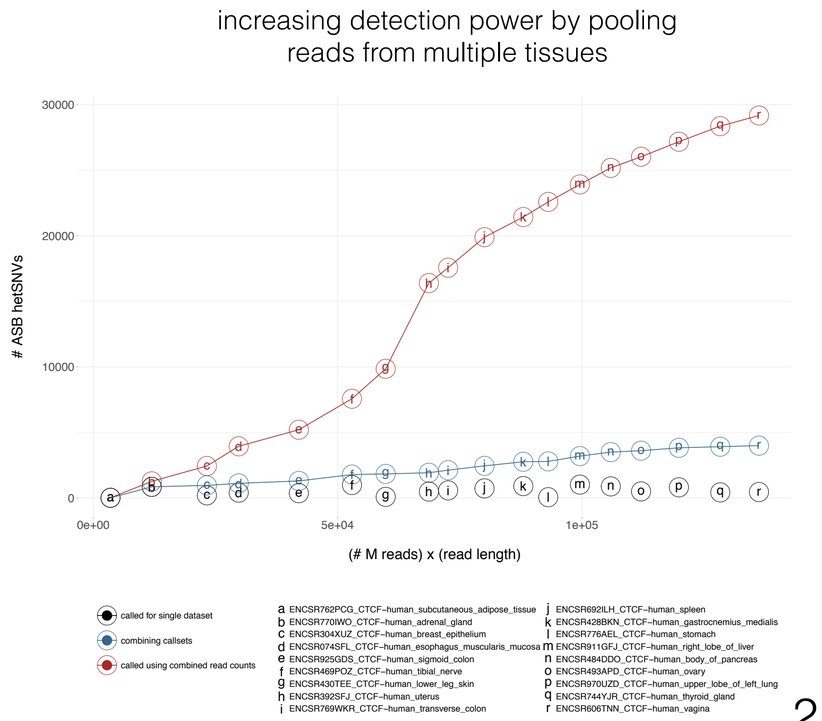
Summary of allele-specific analyses and current plans

Timur Galeev, Joel Rozowsky, Mark Gerstein

2018-05-09

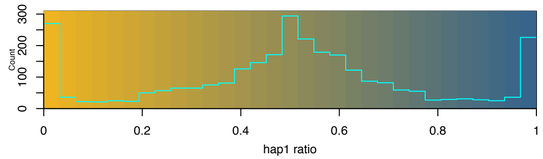
Calling allele-specific hetSNVs, genes, and peaks (e.g. shown for ENC-003):



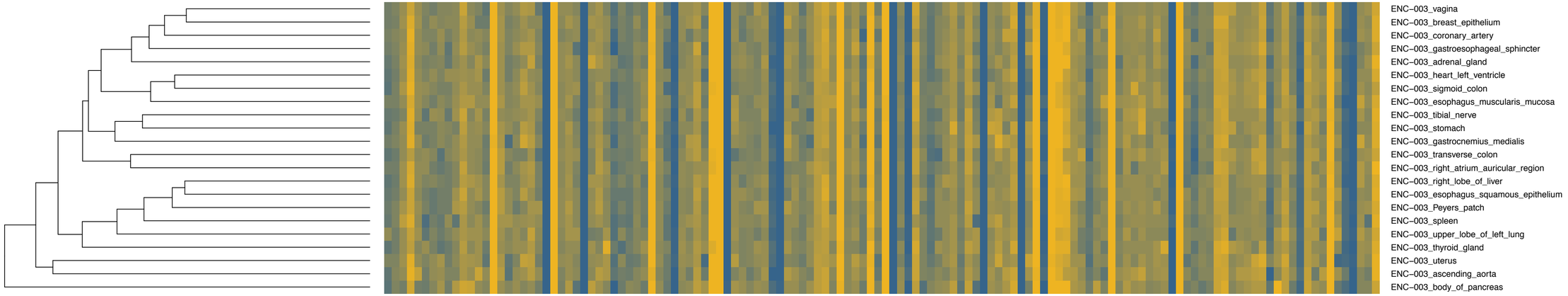


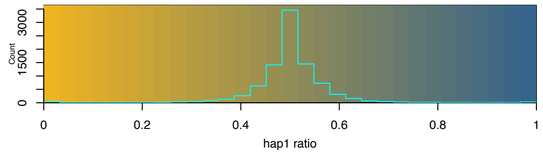
- Most analyses have been focused on ENC-002, ENC-003, will extend when high-quality genomes will be released for the two other individuals;
- add new datasets that became available since last processing;
- will also generate ‘combined call-sets’ using pooling samples from different tissues.

ASB and ASE across tissues

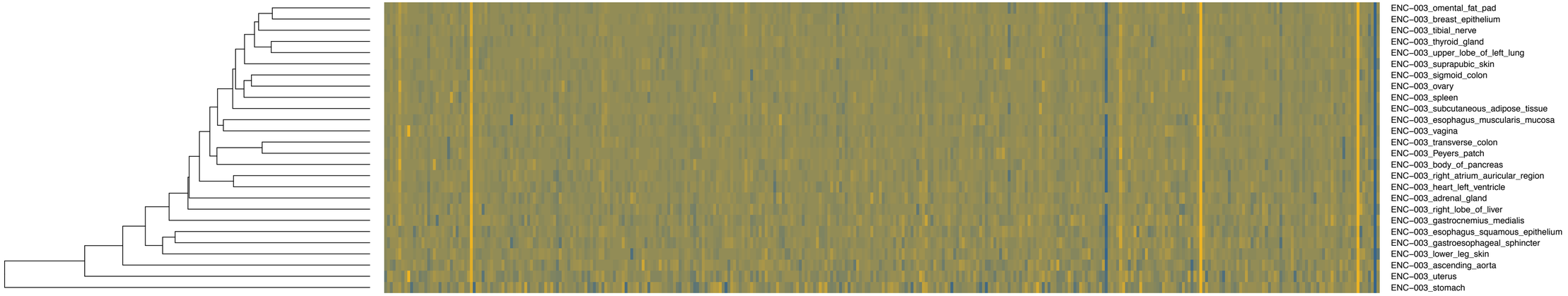


ASB (H3K27ac) in ‘cell type-agnostic’ candidate *cis*-Regulatory Elements



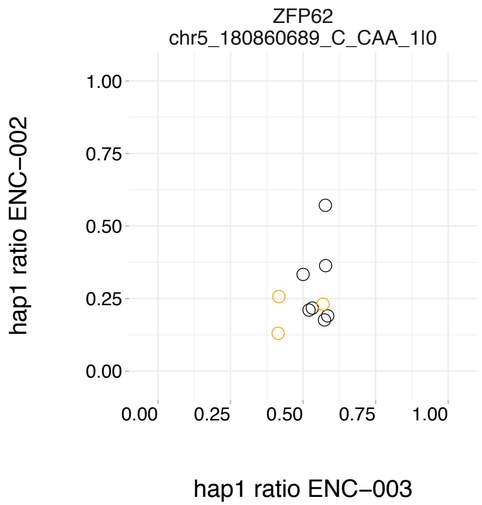


ASE, protein-coding genes



Relating allele-specific behavior with variation

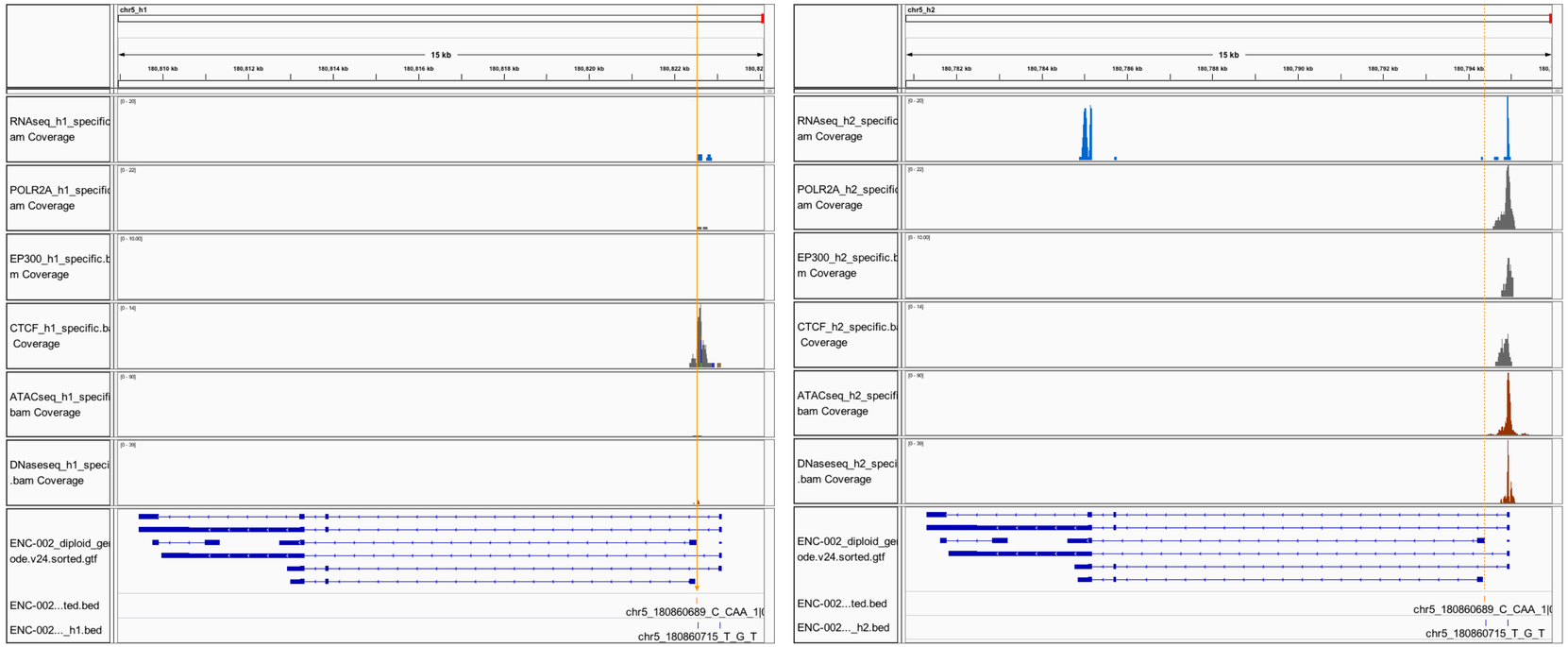
- Variants in regulatory elements



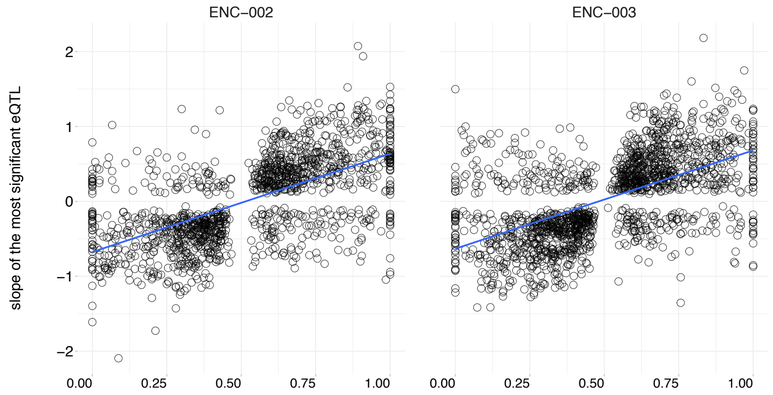
hap1

2bp INS in hap1

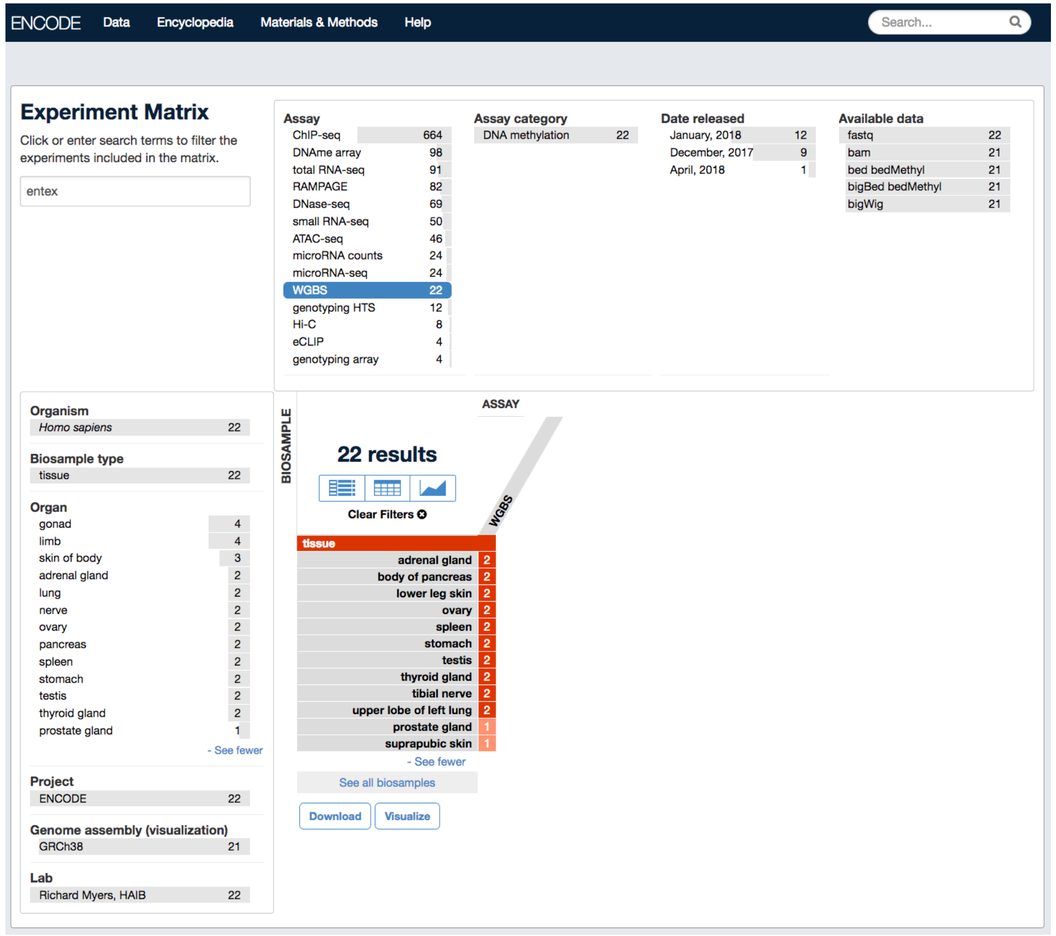
hap2

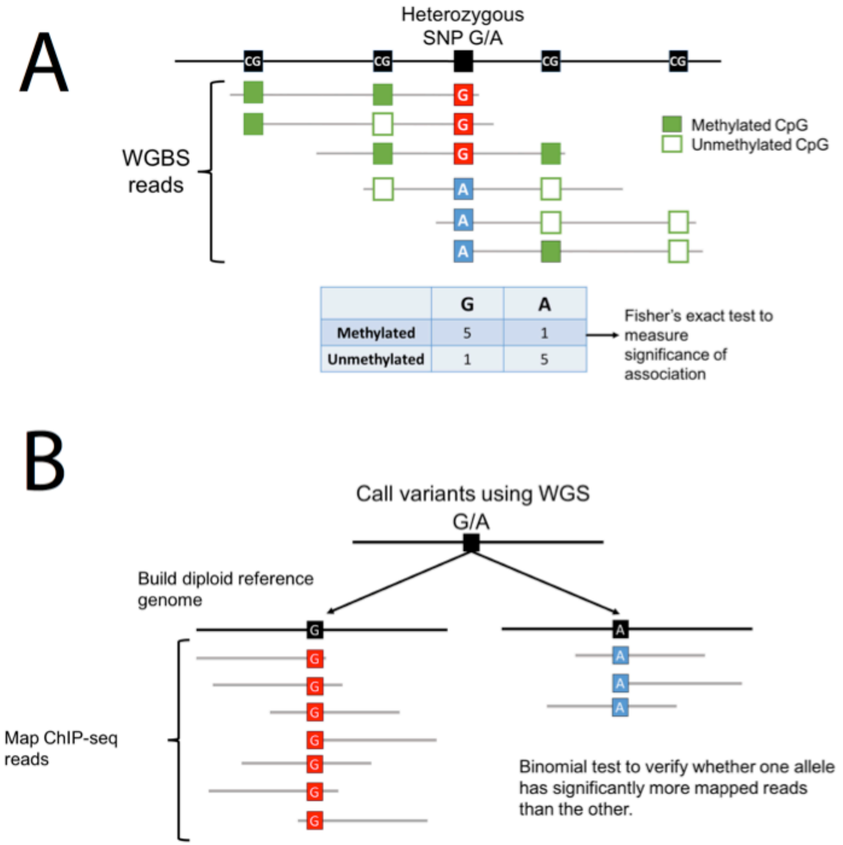


- Relationship with known eQTLs



Plans for calling allele-specific methylation (ASM) events





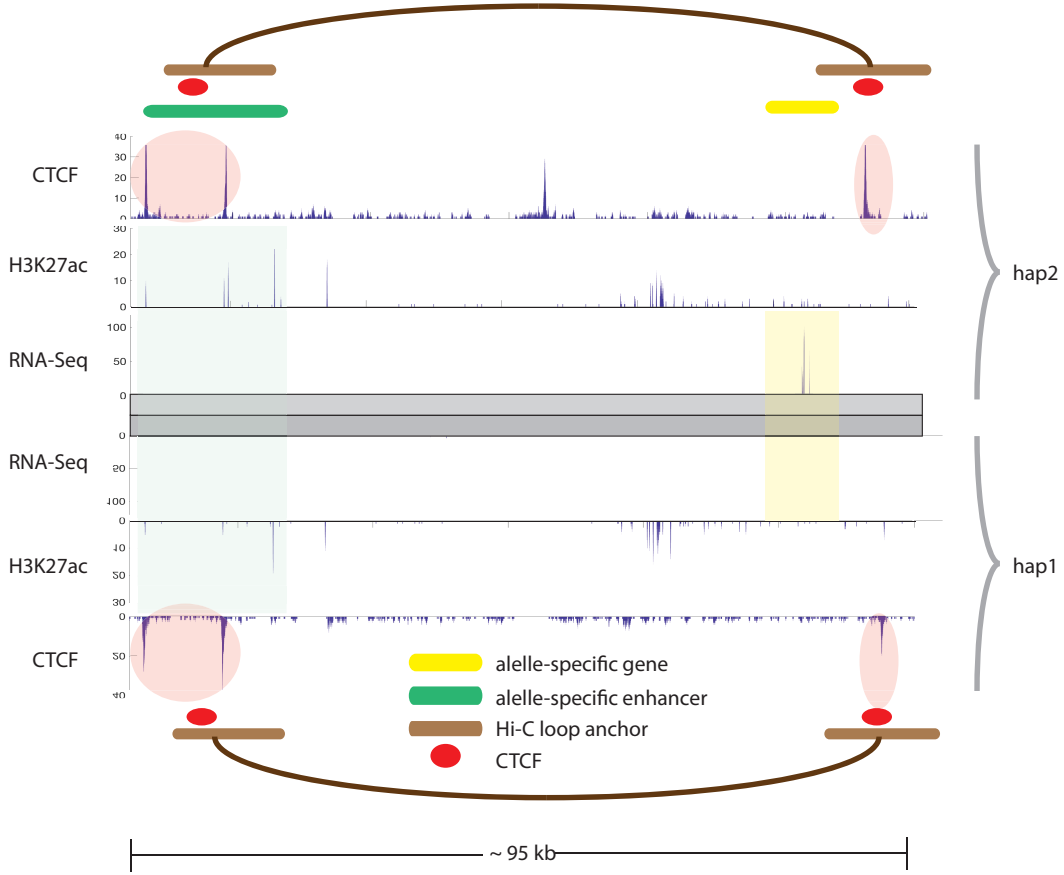
- and other assays, e.g. smallRNA-seq, RAMPAGE, Hi-C

Integration of multiple chromatin assays to understand allele-specific gene expression mechanism

GG, TG and Gerstein Lab

Allele – specific regulation of MYBPC2 gene

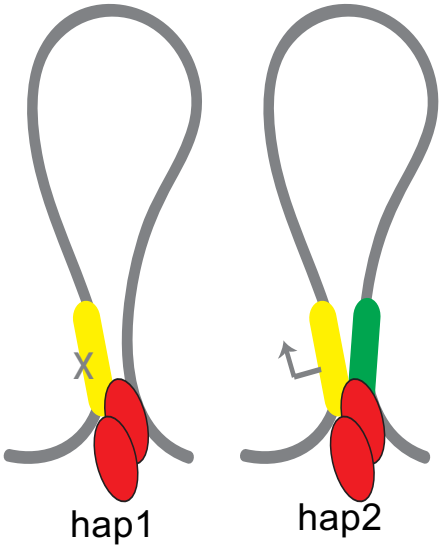
example from ENC-002 muscle tissue



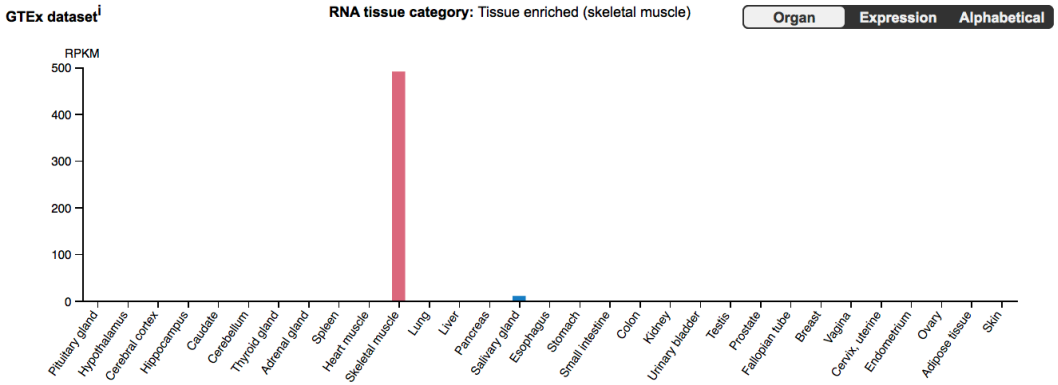
ASE gene: MYBPC2
Chr19: 50,432,903 – 50,466,321

ASB H3K27ac peaks:
~Chr19: 50,360,000 – 50,370,000

Hi-C loop anchor 1:
Chr19: 50,365,000 – 50,370,000
Hi-C loop anchor 2:
Chr19: 50,460,000 – 50,465,000
(slightly more loops in hap2 but no significant allelelic imbalance)

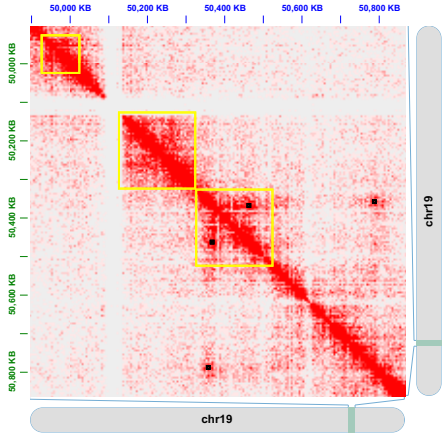


Myosin binding protein C : a highly tissue-specific gene

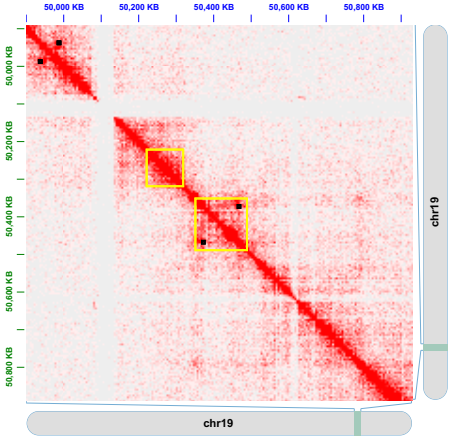


TAD and loop structures are different in different tissues

G. medialis 54 male



T colon 54 male



Goals:

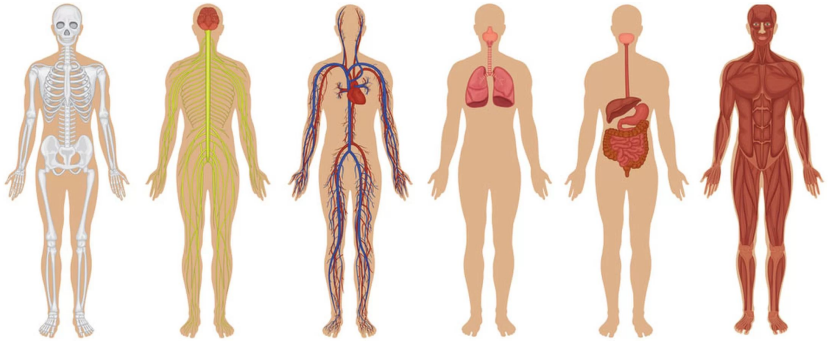
- Multi-assay AS calls on haploblocks: many transcription factors and histone marks simultaneously
- Annotation of AS candidate regulatory regions
- Incorporation of AS Hi-C calls with AS CTCF calls
- 3D genome comparison of each haploblock with inference on regulatory activity

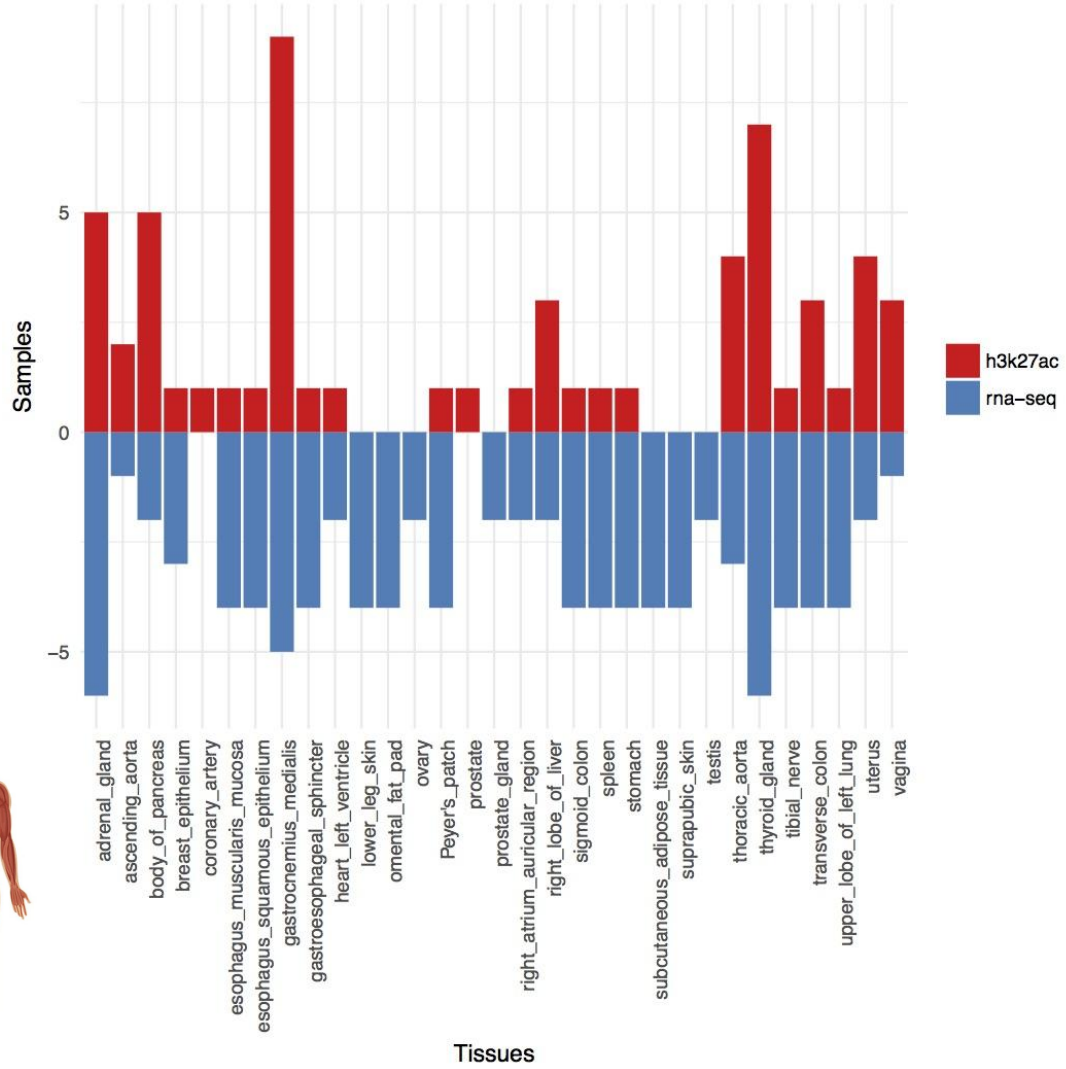
Global Inter-tissue spectral analysis

Fabio Navarro

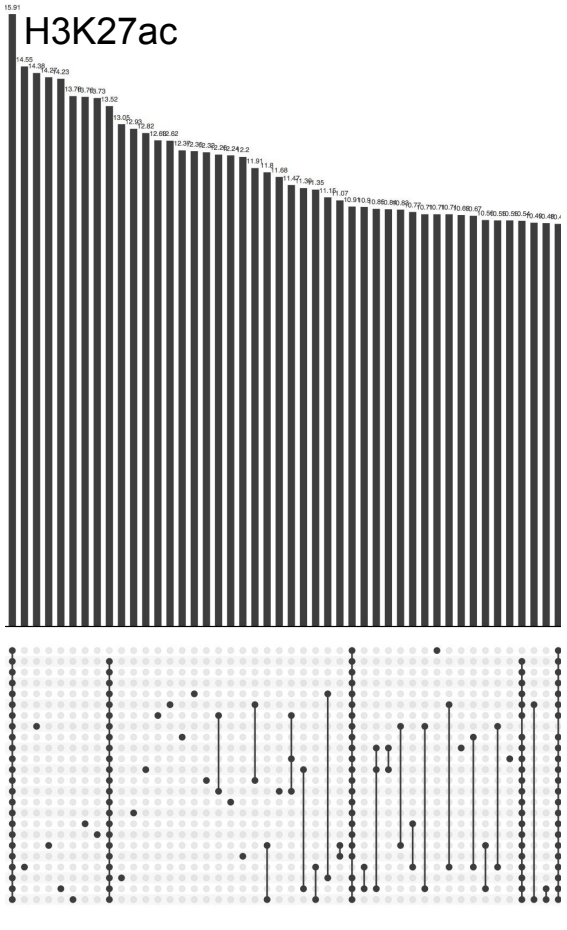
Datasets

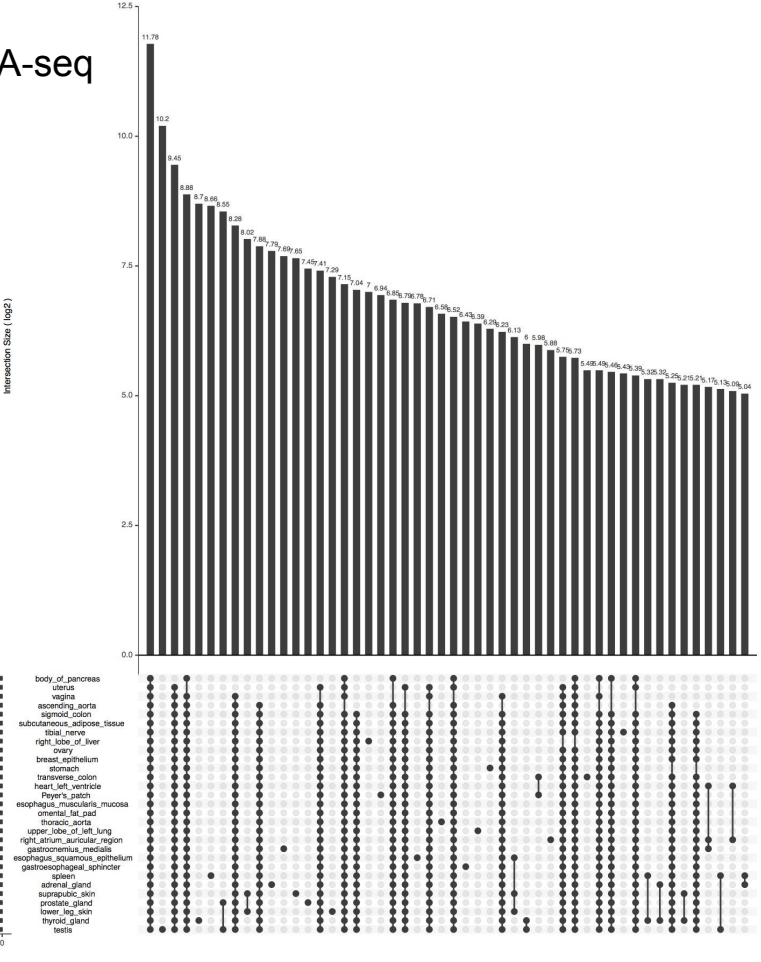
- Future analysis
  - RAMPAGE
  - ATAC-Seq/DNAse
  - Other Hist mod/RNA-Pol2



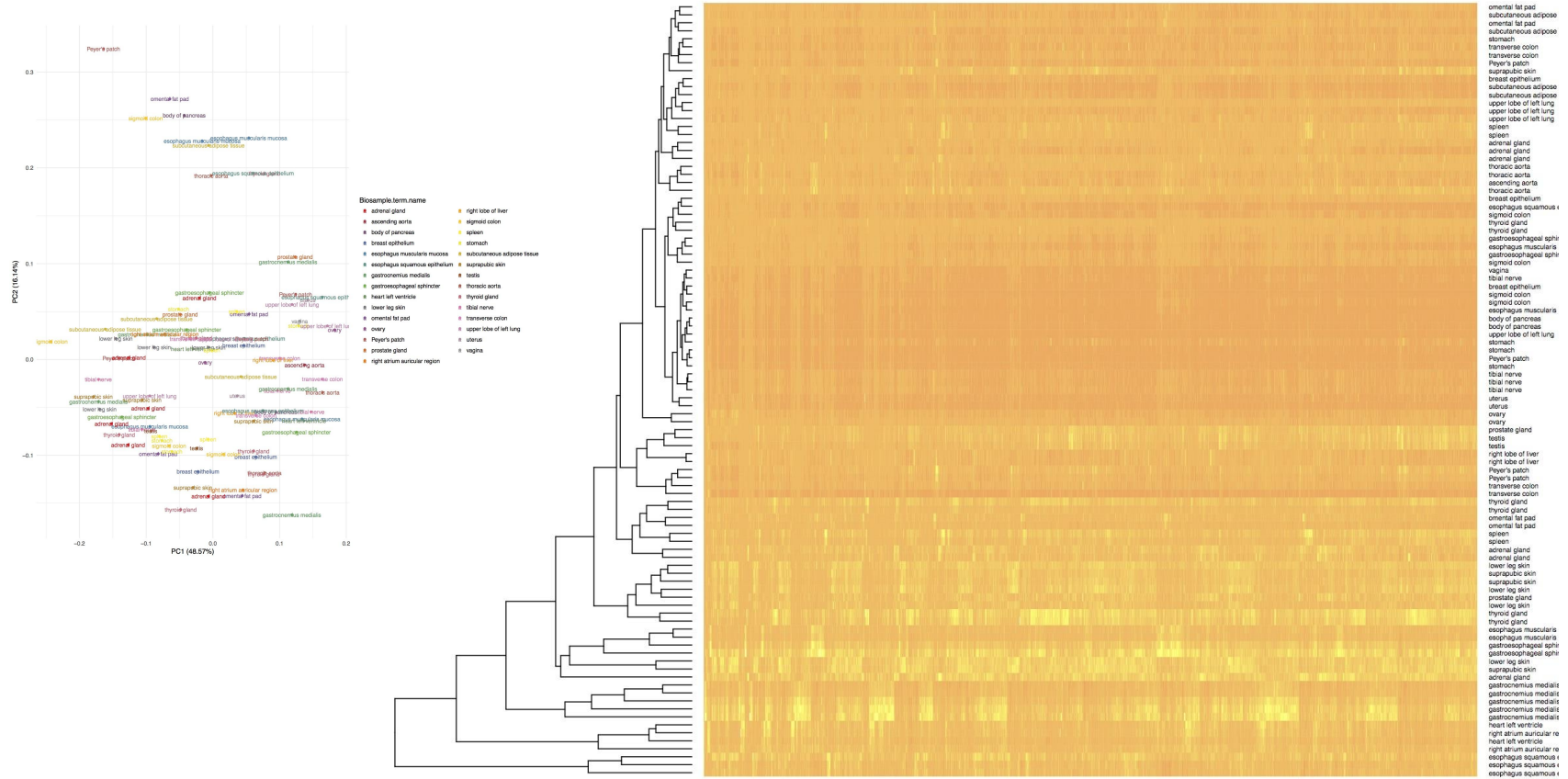


Cross tissue regulatory/transcriptome similarity

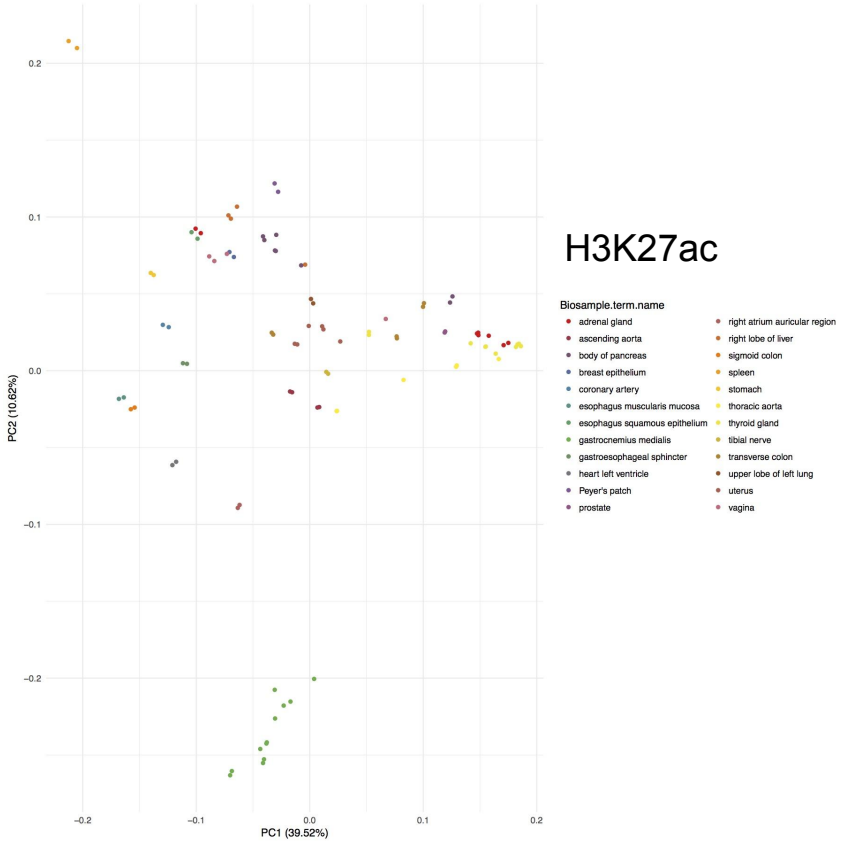




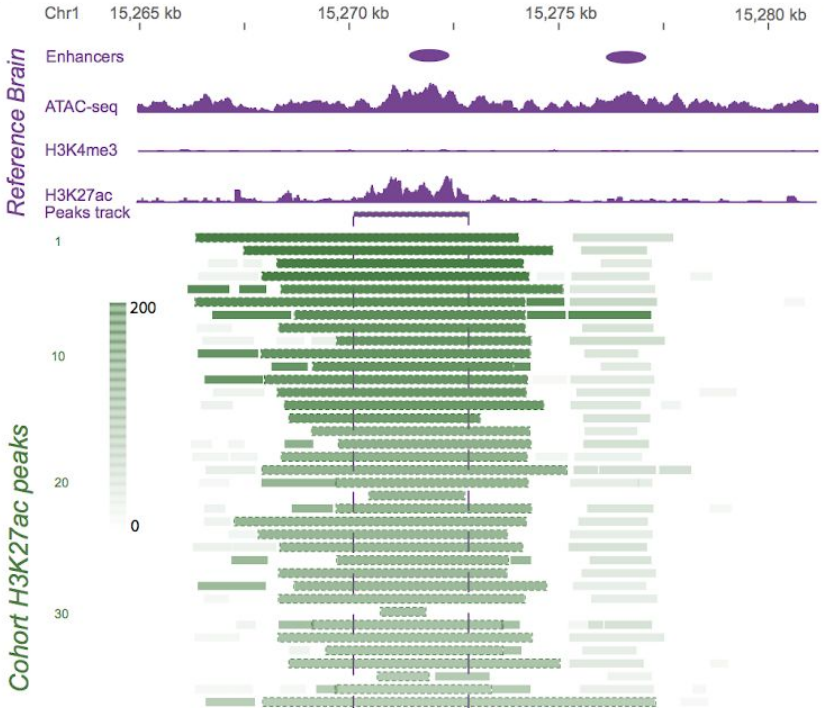
RNA-seq tissue clustering

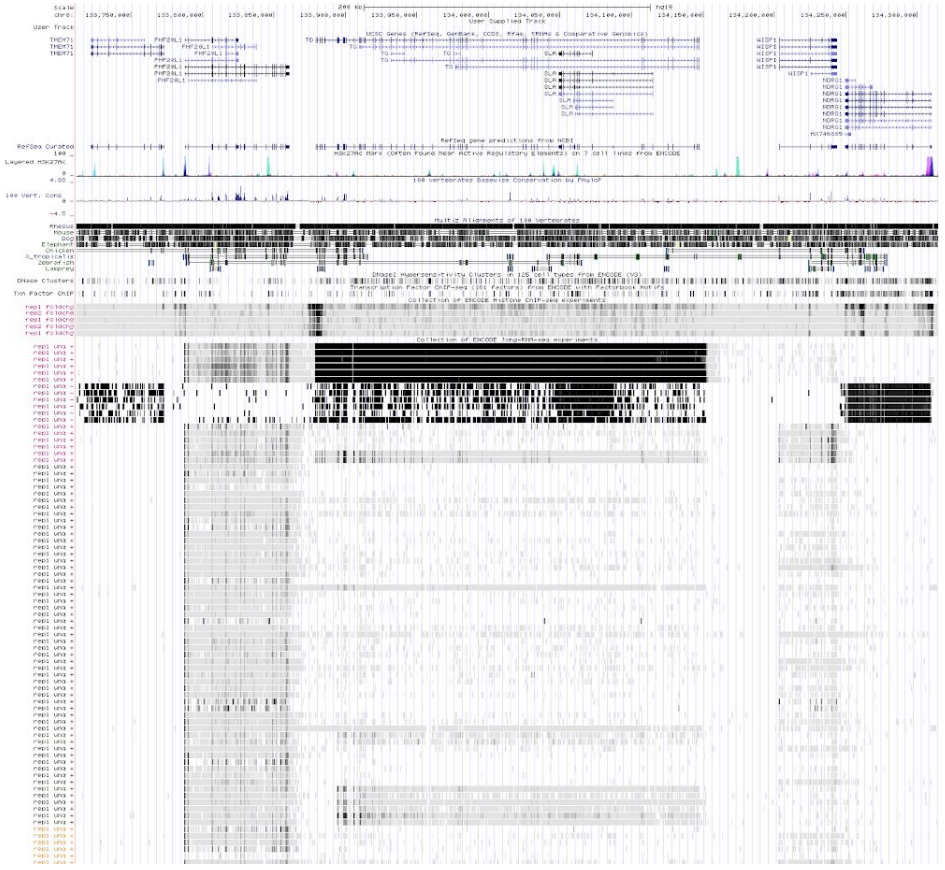


Dimensionality Reduction (PCA)

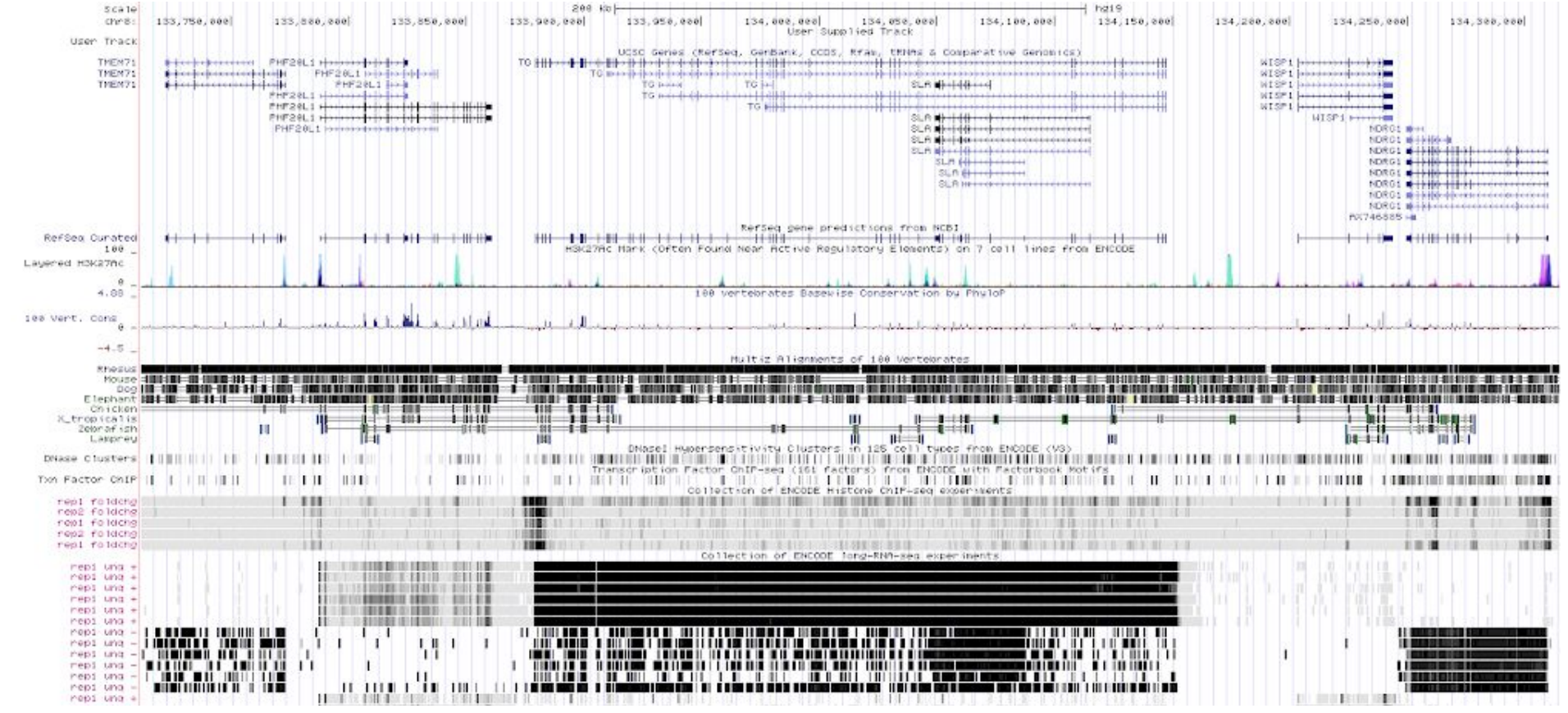


Inter individual/tissue variation





Inter individual/tissue variation



eQTL enrichment analyses

Mengting Gu and Yucheng T. Yang
Gerstein Lab

Enrichment of eQTL/GWAS SNPs in functional regulatory regions in the genome

- Identify consistent potential regulatory regions in the genome
- Test for enrichment of eQTL/GWAS SNPs in active regions for each tissue

ENC-003

Enrichment of eQTL/GWAS SNPs in functional regulatory regions in the genome

- Identify consistent potential regulatory regions in the genome
- Test for enrichment of eQTL/GWAS SNPs in active regions for each tissue

- We are developing a list of enhancers for EN-TEX tissues using DNA accessibility, histone modification and transcription factor binding data in each tissue
- Consistent candidate cis regulatory elements (ENCODE annotation)

Enrichment of eQTL/GWAS SNPs in functional regulatory regions in the genome

- Identify consistent potential regulatory regions in the genome
- Test for enrichment of eQTL/GWAS SNPs in active regions for each tissue

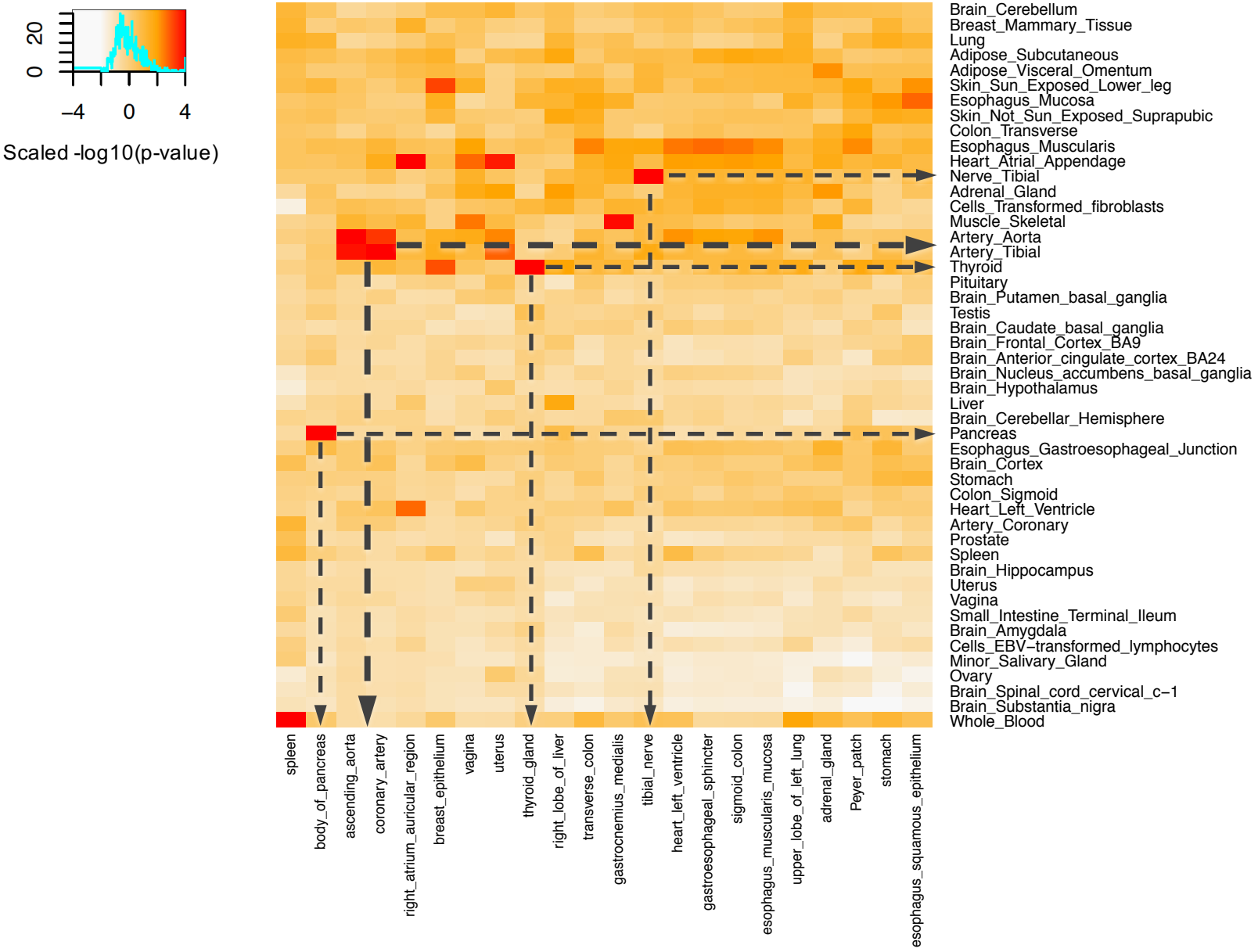
- We are developing a list of enhancers for EN-TEX tissues using DNA accessibility, histone modification and transcription factor binding data in each tissue
- Consistent candidate cis regulatory elements (ENCODE annotation)

- Proof of concept analysis:

  Enrichment test based on merged H3K27ac peaks

  Use top ranked H3K27ac peaks in each tissue (excluding promoters) as putative enhancers

Preliminary results of eQTL enrichment for H3K27ac peaks



Possible artifacts in GWAS/eQTL enrichment analysis

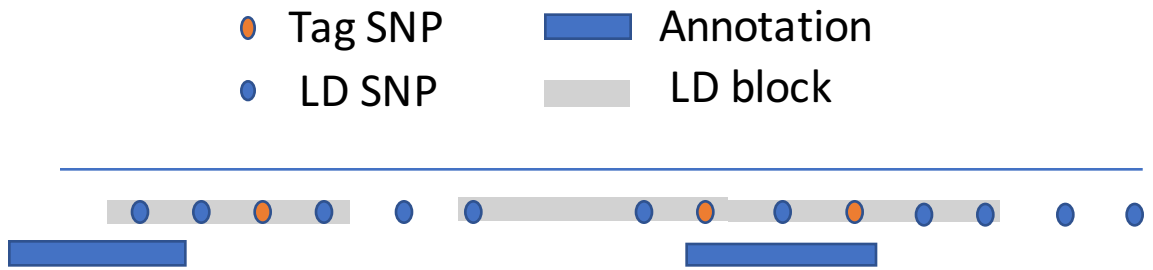
Jing Zhang

Gerstein Lab

Possible Artifacts in GWAS/eQTL enrichment analysis

Artifacts from missing LD SNPs

- Especially a problem in GWAS
- LD pruning before the test
- Tag SNP may sit outside of annotation while causal SNP is its LD SNP
- True for eQTL analysis if 1kg SNPs are used as background



Artifacts from duplicating LD SNPs

- Two LD SNPs are linked to two Tag SNPs

Artifacts from size of LD blocks

- LD block changes in size up to several order

Artifacts from size of SNP allele frequencies blocks

- GWAS and eQTL SNPs are relatively common SNPs

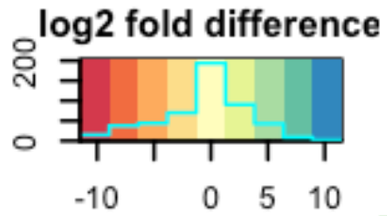
Artifacts from distance to annotations

- Gene proximal and distal

SV impact on allele-specific expression

Xiangmeng Kong, Mark Gerstein
Yale University

SV overlapped exons in ENC002



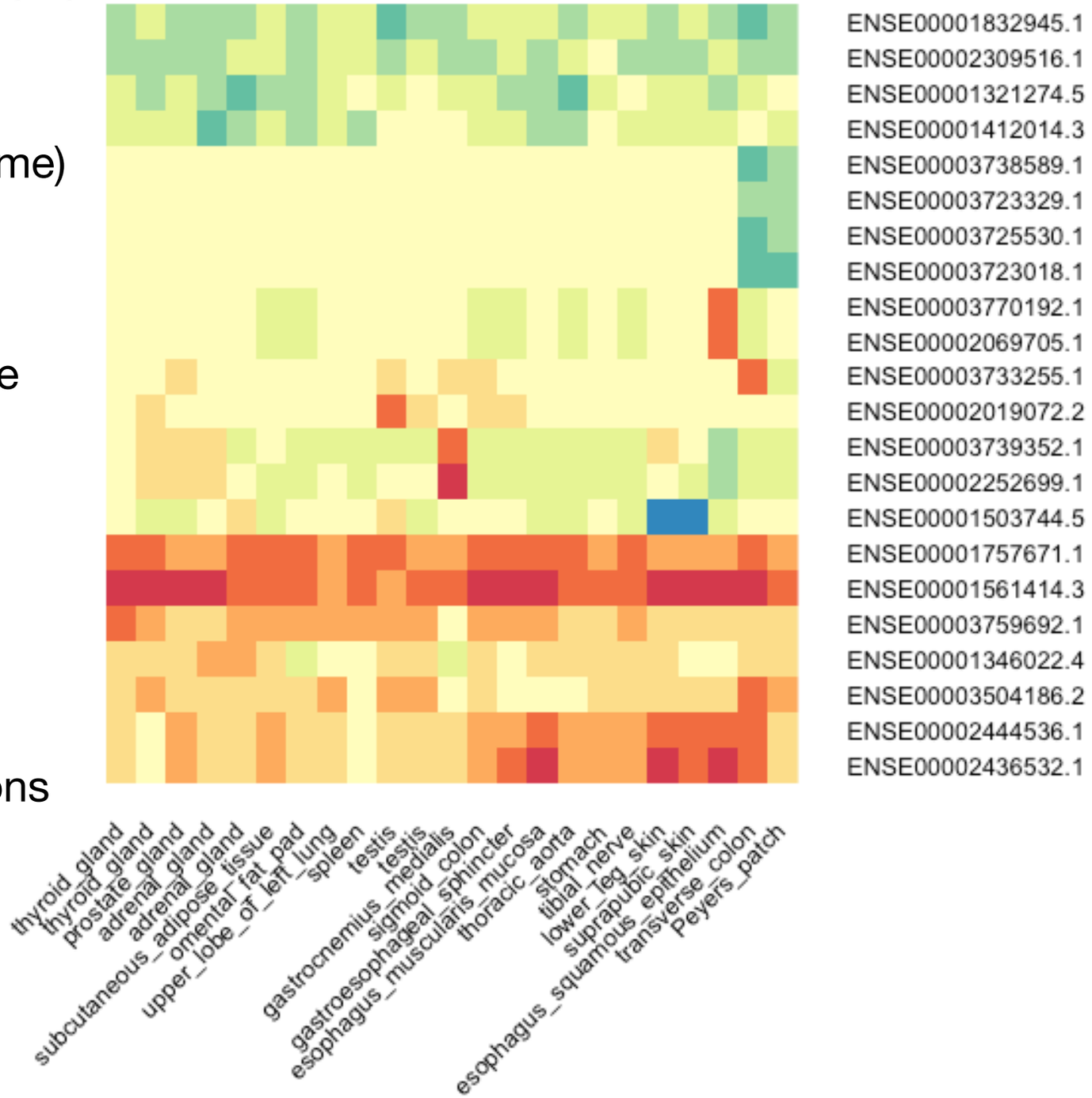
Total SVs: 21672

Overlapped exons: 926 (906 in Personal Genome)

Only reads uniquely mapped to one haplotype

$$log_2\left(\frac{h1\ counts+1}{h2\ counts+1}\right)$$

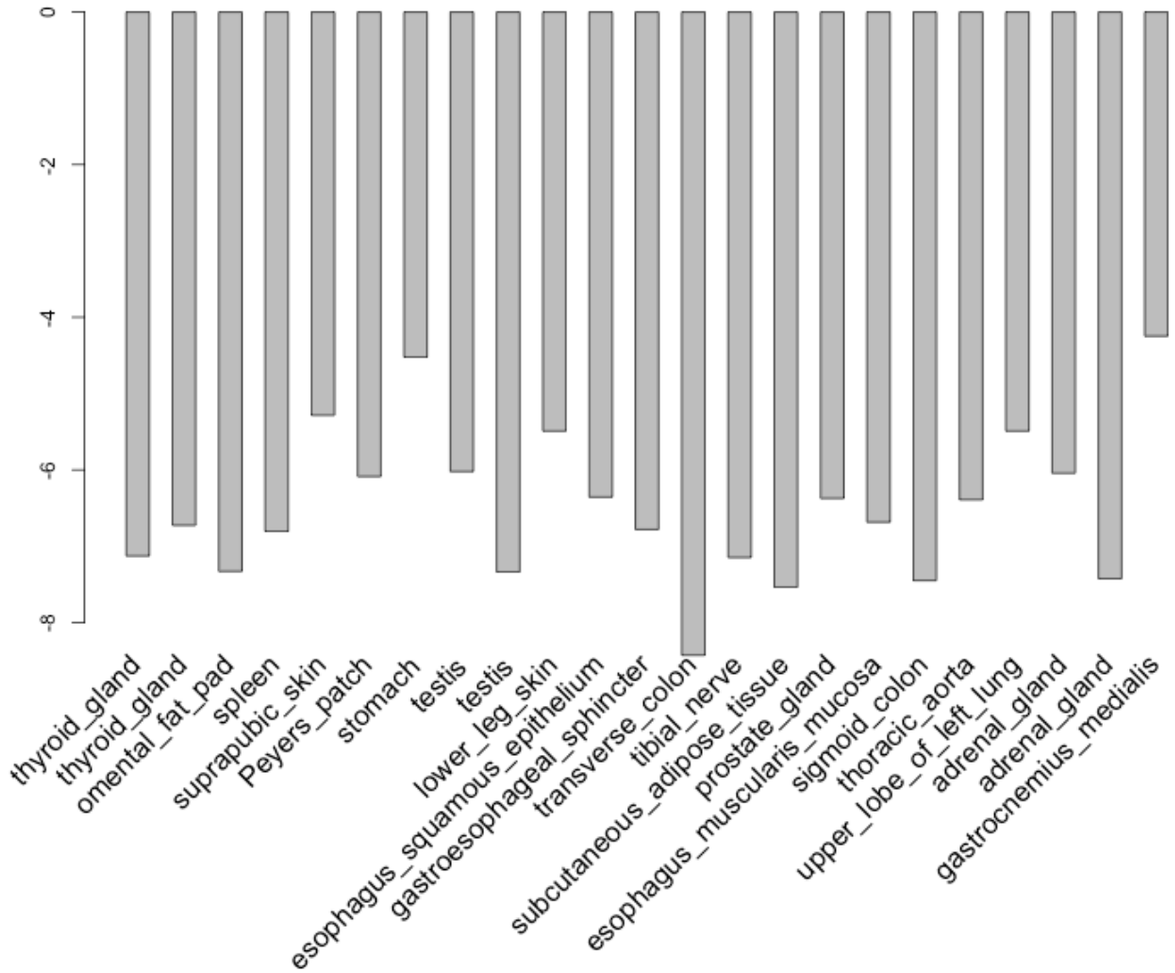
Significant fold change in >=1 tissue: 265 exons



SV impact gene cases

p< 0.05 in all samples: 2

```
chr9  133019485  133020874   ENSE00001561414.3
chr9  133019413  133021194   DEL  GT:DR:DV1|0:35:17
```




```
chr13 110936758  110939784   ENSE00001757671.1
chr13 110938283  110938283   INS  GT:DR:DV1|0:29:35
```