The Ovine Functional Annotation of Animal Genome Project

Brenda M. Murdoch, Shangqian Xie, Kimberly M. Davenport, Mazdak Salavati, Emily Clark, Alan Archibald, Stephen N. White, Michelle R. Mousel, Kim C. Worley, Stephanie D. McKay, Alex Caulton, Shannon Clarke, Rudiger Brauning, Tracy Hadfield, Benjamin D. Rosen, Timothy P.L. Smith, and Noelle E. Cockett

Ovine Reference Genome

ARS-UI_Ramb_v2.0

- Go-FAANG meeting (2015)

- FAANG piolet project funded (chicken, cattle & pig)

- The ovine FAANG project was the first single specie US FAANG project funded

- New reference assembly & over 100 tissues were collected from the animal used for the reference assembly (NRSP8 – sheep)

Ovine Reference Genome

ARS-UI_Ramb_v2.0

- Updated sheep reference genome for more accurate functional annotation
  - Technology updates
  - Improved contiguity
  - Recently annotated by NCBI (GCA_016772045.1)

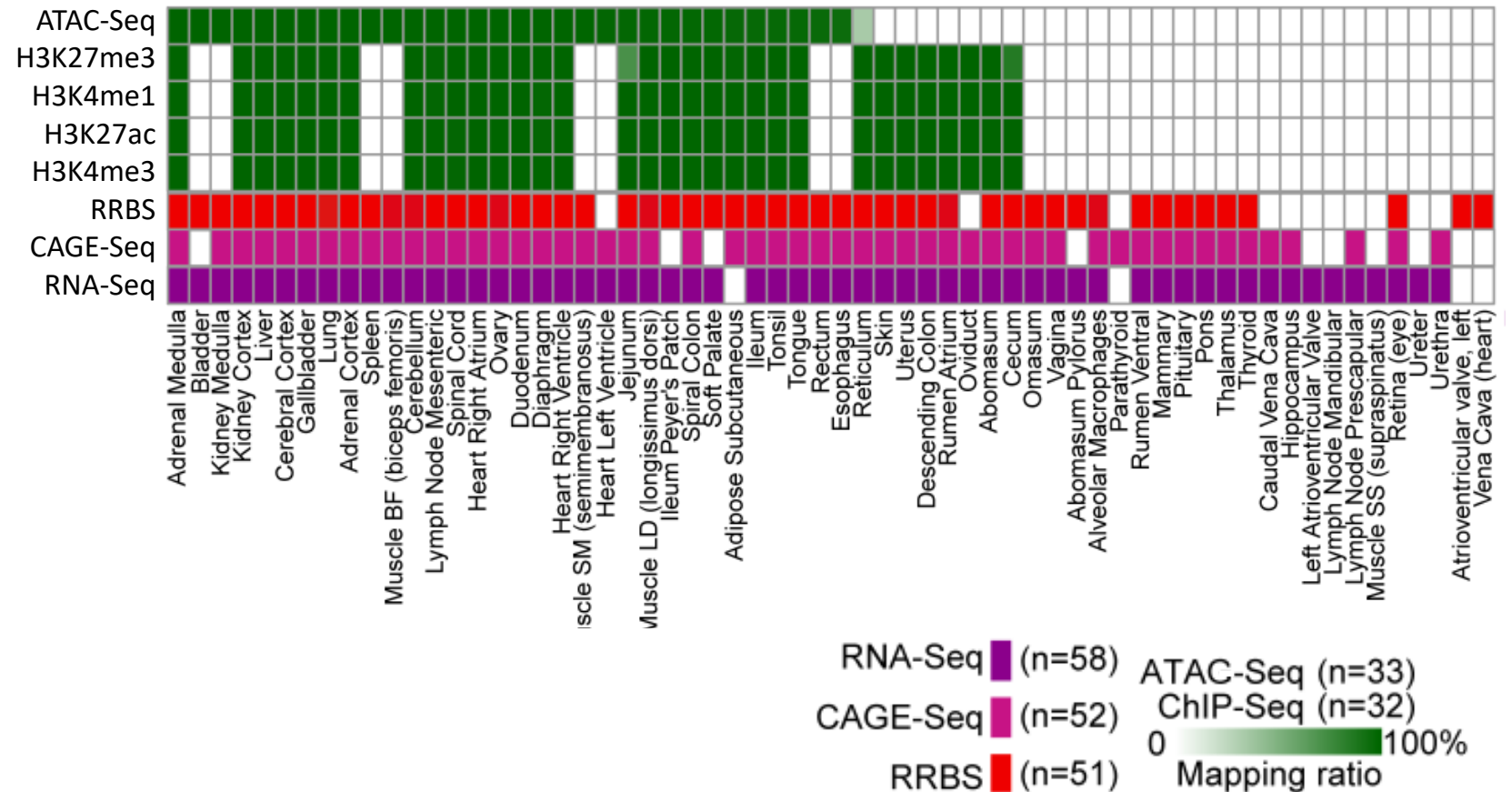|	Contig N50 (Mb)	LG50 (contigs)	Number of contigs	Assembly size (Gb)	Release date
Oar_v4.0	0.15	5,008	48,482	2.62	2012
Oar_rambouillet_v1	2.57	313	7,486	2.87	2017
ARS-UI_Ramb_v2.0*	43.18	24	226	2.63	2021
ARS-UI_Ramb_v3.0	43.18	24	226	2.63	2023

*Davenport KM, Bickhart DM, Worley K, Murali SC, Cockett NE, Heaton MP, Smith TPL, Murdoch BM, Rosen BD. An improved ovine reference genome assembly to facilitate in depth functional annotation of the sheep genome. (2022) GigaScience.
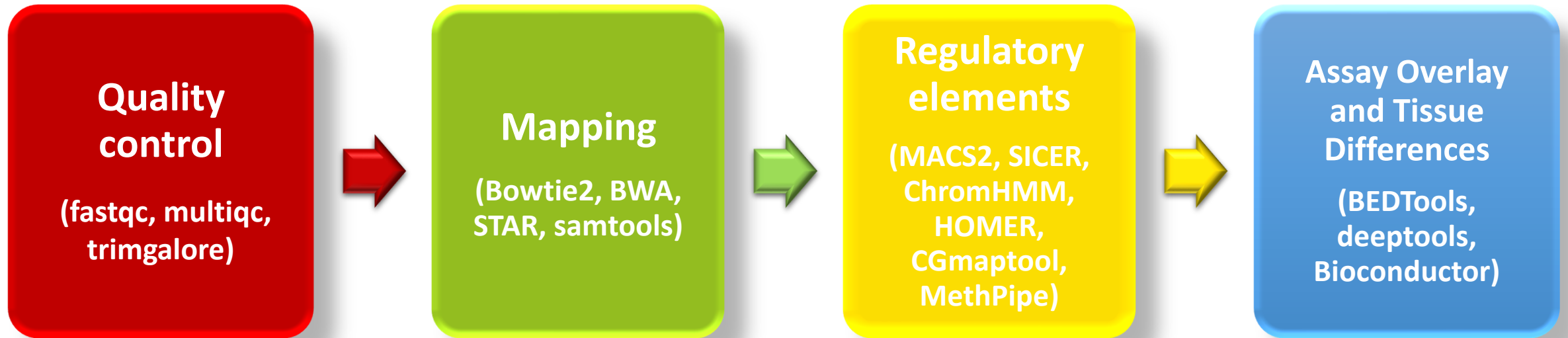
Ovine Reference Genome

ARS-UI_Ramb_v2.0

- Tissues from reference genome for FAANG
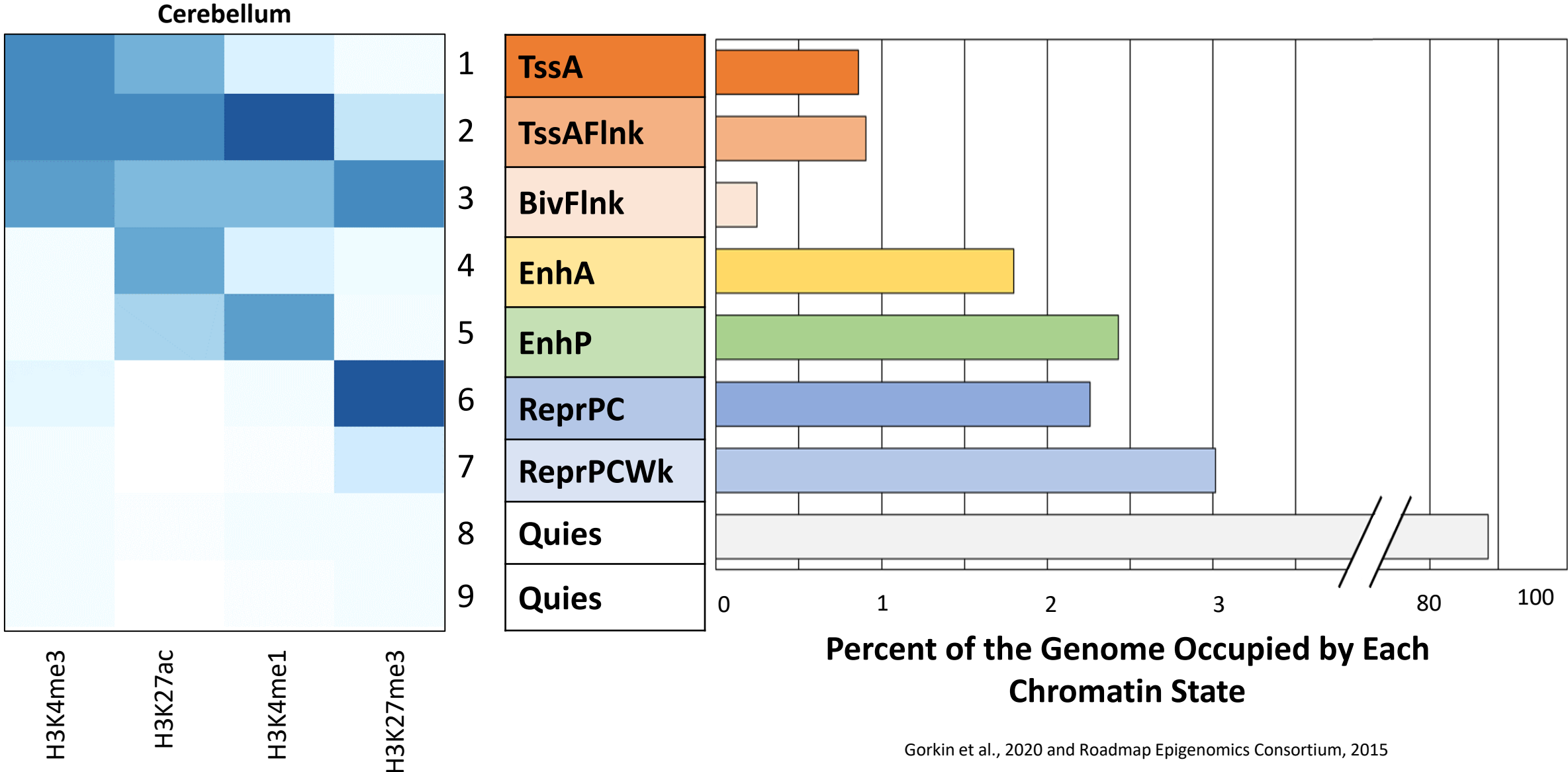- Remap to updated sheep reference genome for more accurate functional annotation

Bioinformatic Pipelines
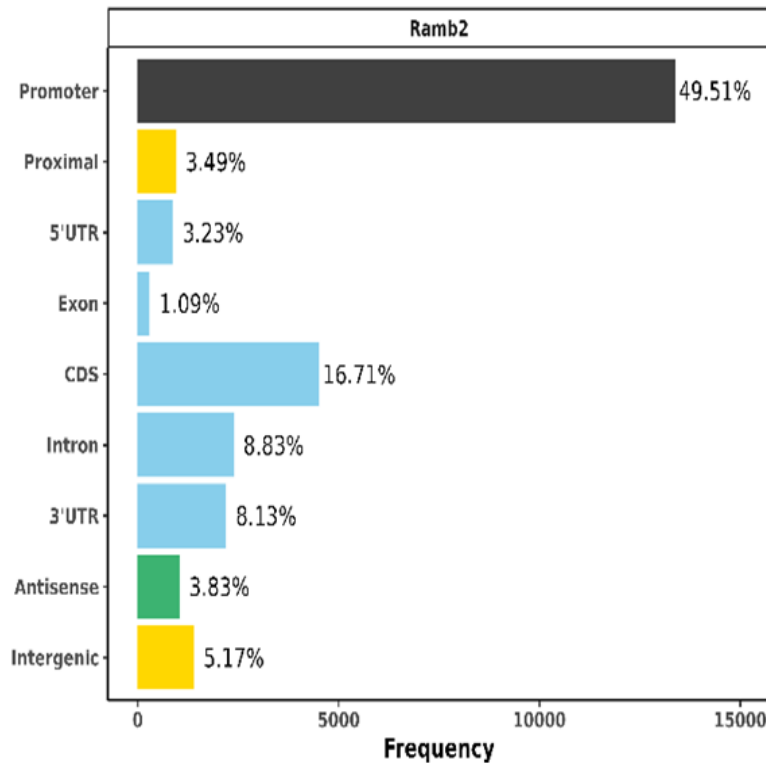
- ChIP-seq, ATAC-seq, RRBS data

Quality control

(fastqc, multiqc, trimgalore)

Mapping

(Bowtie2, BWA, STAR, samtools)

Regulatory elements

(MACS2, SICER, ChromHMM, HOMER, CGmaptool, MethPipe)

Assay Overlay and Tissue Differences

(BEDTools, deeptools, Bioconductor)

Chromatin States & Genome Occupancy

Cerebellum

1 TssA
2 TssAFlnk
3 BivFlnk
4 EnhA
5 EnhP
6 ReprPC
7 ReprPCWk
8 Quies
9 Quies

H3K4me3 H3K27ac H3K4me1 H3K27me3

0   1   2   3   80   100

Percent of the Genome Occupied by Each Chromatin State
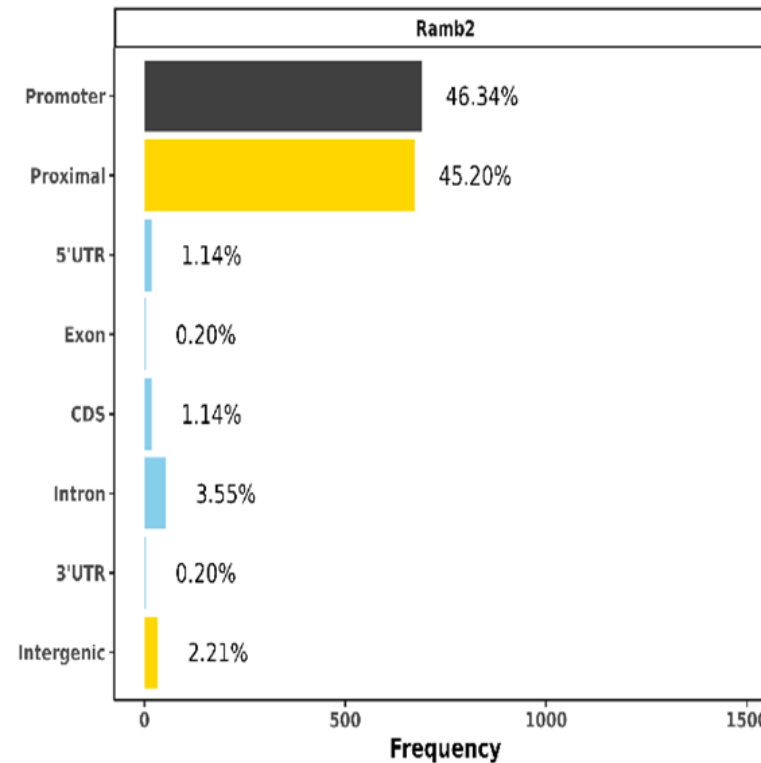
Gorkin et al., 2020 and Roadmap Epigenomics Consortium, 2015

CAGE Signal with Chromatin States

Uni-directional TSS clusters
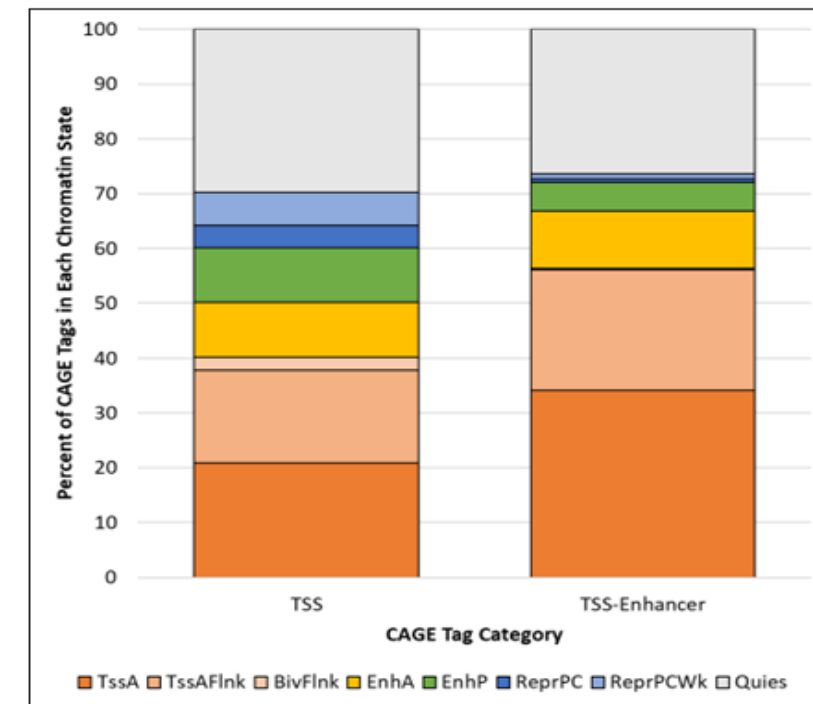

TSS clusters the highest in promoter
(± 100 bp of the 5 UTR)

Bi-directional TSS clusters

TSS-Enhancer clusters the highest
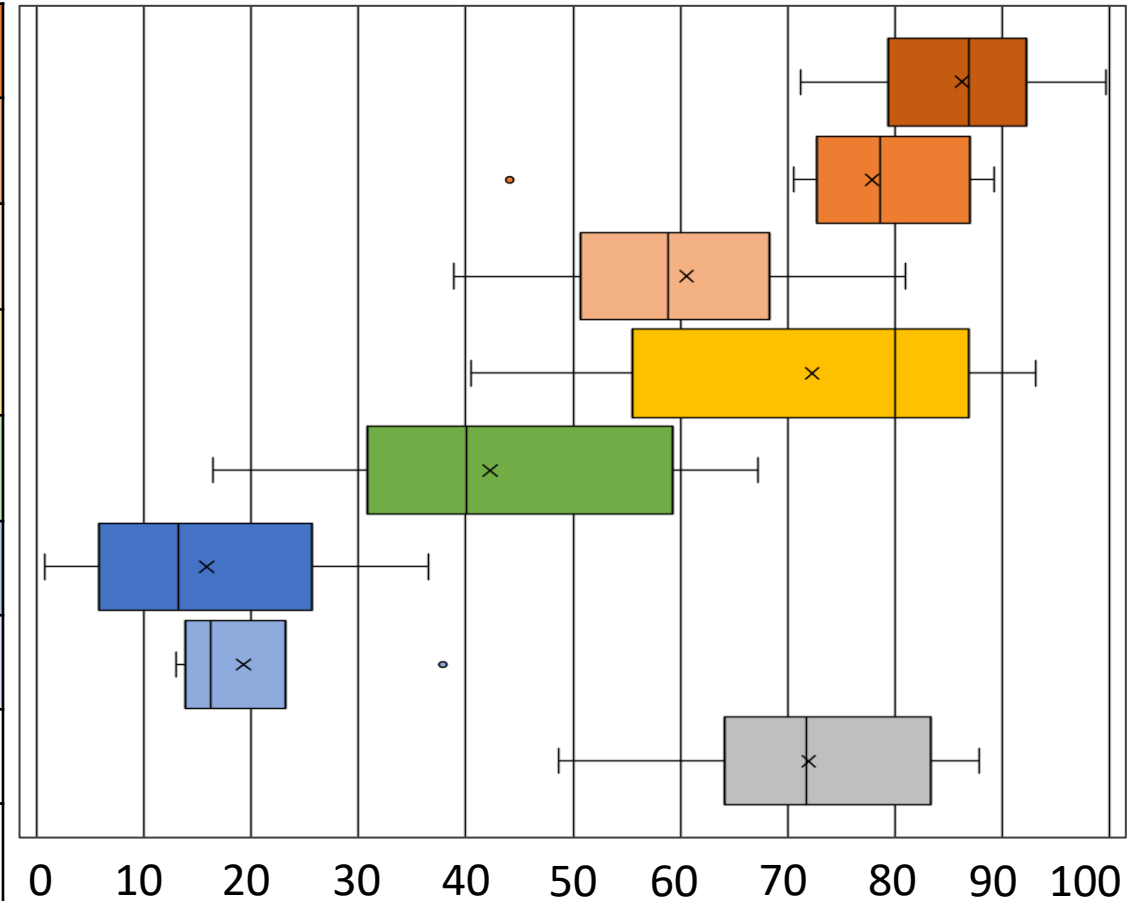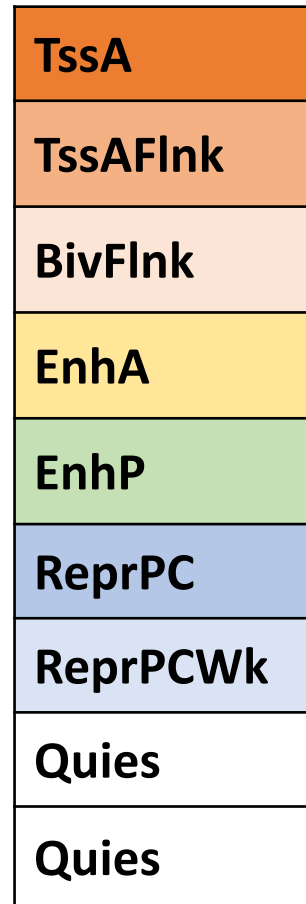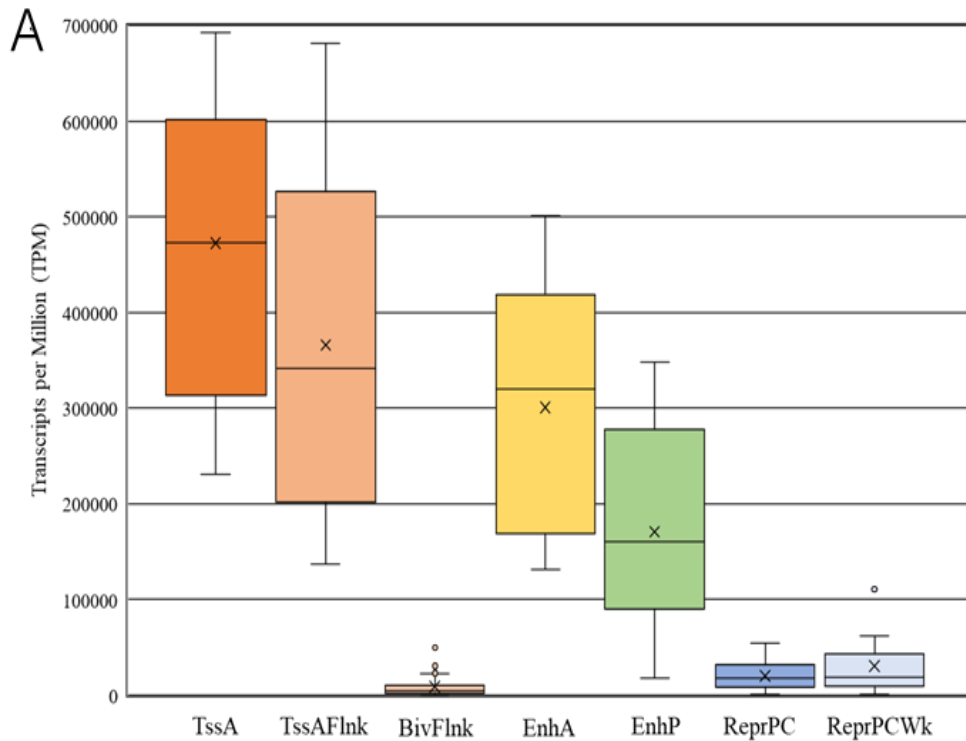upstream of the TSS
(1,000 bp upstream of the 5'UTR)

Overlap with chromatin states

~70% of TSS & TSS-Enhancer tags
overlap with chromatin states 1-7

Chromatin States with RNA & ATAC-seq

RNA Transcript that Overlaps with Each Chromatin State

Percent of Each Chromatin State that Overlaps with Open Chromatin Regions (ATAC-seq peaks)
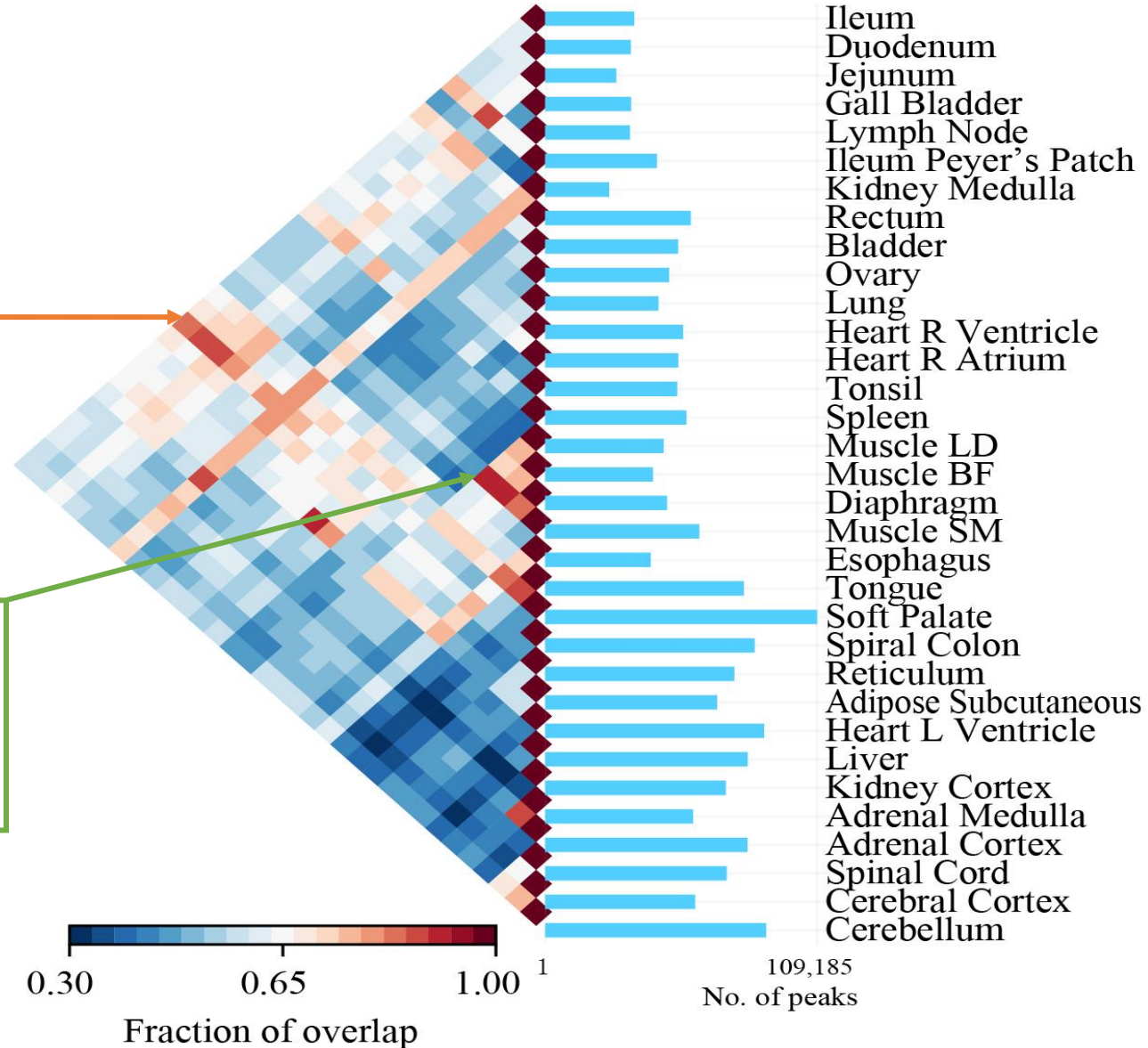
ATAC-seq Signals

Pairwise Intersection

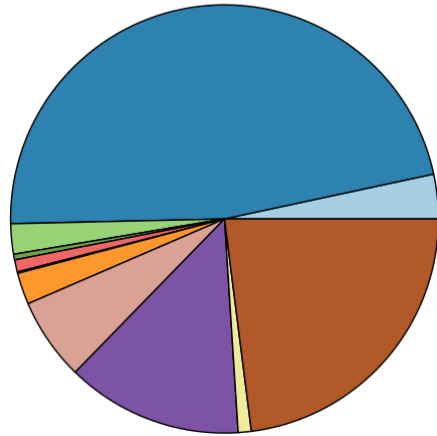**GIT Tissues**
- Ileum, Duodenum, Jejunum with Spiral Colon

Muscle Tissues
- Muscle LD, Muscle BF, Diaphragm, with Muscle SM

Ileum
Duodenum
Jejunum
Gall Bladder
Lymph Node
Ileum Peyer's Patch
Kidney Medulla
Rectum
Bladder
Ovary
Lung
Heart R Ventricle
Heart R Atrium
Tonsil
Spleen
Muscle LD
Muscle BF
Diaphragm
Muscle SM
Esophagus
Tongue
Soft Palate
Spiral Colon
Reticulum
Adipose Subcutaneous
Heart L Ventricle
Liver
Kidney Cortex
Adrenal Medulla
Adrenal Cortex
Spinal Cord
Cerebral Cortex
Cerebellum

0.30   0.65   1.00
Fraction of overlap

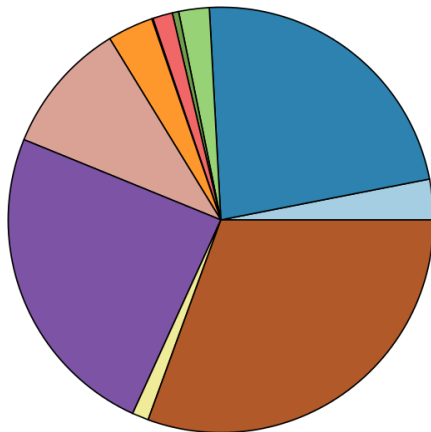1   109,185
                     No. of peaks

ATAC-seq Signals
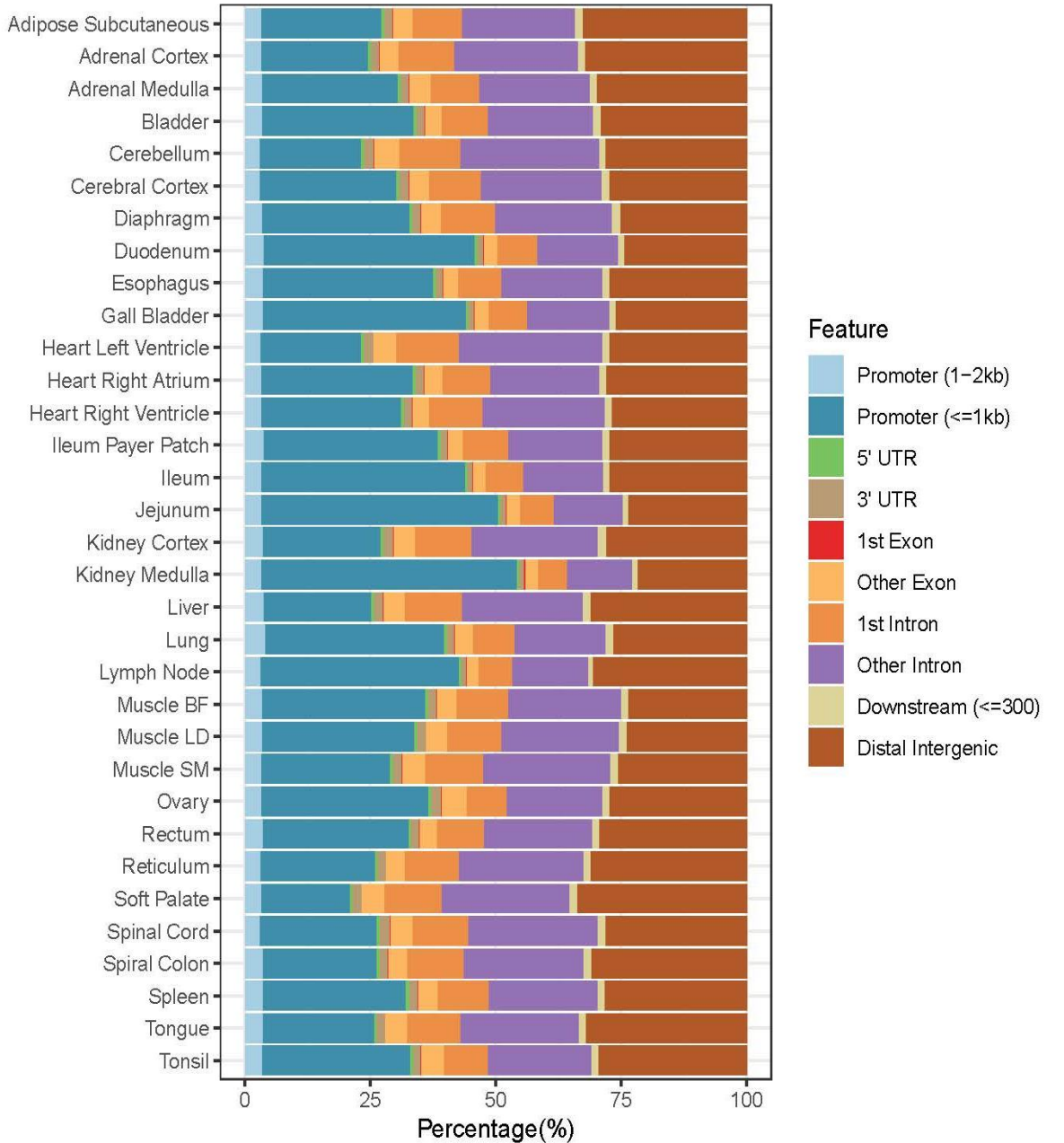
**Jejunum
Distribution of Peaks**

- Promoter (1–2kb) (3.36%)
- Promoter (<=1kb) (47%)
- Promoter (2–3kb) (2.29%)
- 5' UTR (0.42%)
- 3' UTR (0.92%)
- 1st Exon (0.11%)
- Other Exon (2.38%)
- 1st Intron (6.23%)
- Other Intron (13.31%)
- Downstream (<=300) (0.99%)
- Distal Intergenic (22.99%)

**Reticulum
Distribution of Peaks**

- Promoter (1–2kb) (3.08%)
- Promoter (<=1kb) (22.79%)
- Promoter (2–3kb) (2.32%)
- 5' UTR (0.48%)
- 3' UTR (1.51%)
- 1st Exon (0.09%)
- Other Exon (3.48%)
- 1st Intron (10.09%)
- Other Intron (24.35%)
- Downstream (<=300) (1.25%)
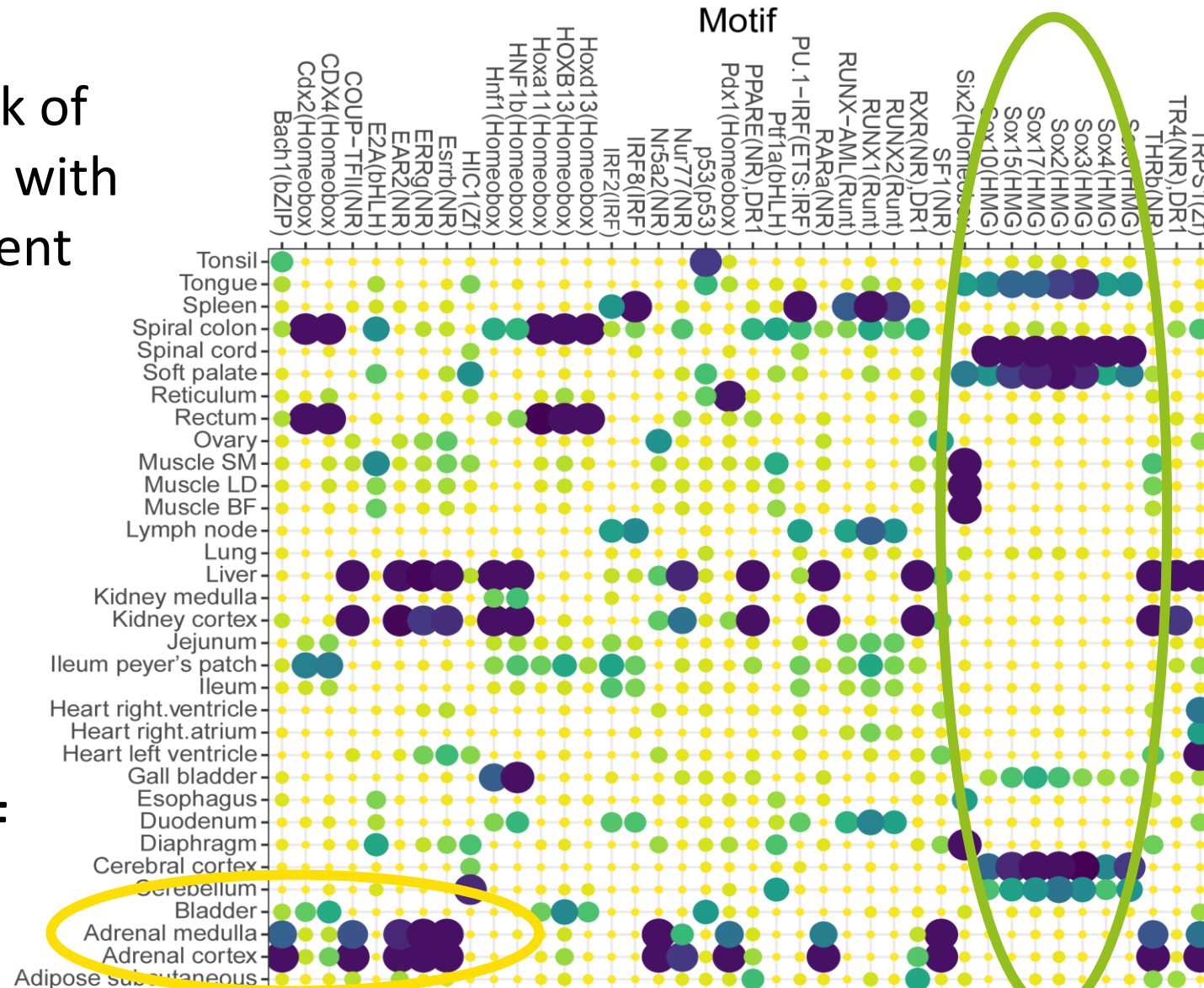- Distal Intergenic (30.56%)

Feature Distribution

Adipose Subcutaneous, Adrenal Cortex, Adrenal Medulla, Bladder, Cerebellum, Cerebral Cortex, Diaphragm, Duodenum, Esophagus, Gall Bladder, Heart Left Ventricle, Heart Right Atrium, Heart Right Ventricle, Ileum Payer Patch, Ileum, Jejunum, Kidney Cortex, Kidney Medulla, Liver, Lung, Lymph Node, Muscle BF, Muscle LD, Muscle SM, Ovary, Rectum, Reticulum, Soft Palate, Spinal Cord, Spiral Colon, Spleen, Tongue, Tonsil

Percentage(%)

Feature
- Promoter (1–2kb)
- Promoter (<=1kb)
- 5' UTR
- 3' UTR
- 1st Exon
- Other Exon
- 1st Intron
- Other Intron
- Downstream (<=300)
- Distal Intergenic

ATAC-seq TFBS across tissues

Regulatory network of the 40 TFBS motifs with the largest coefficient of variation values

SOX TF family
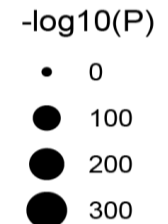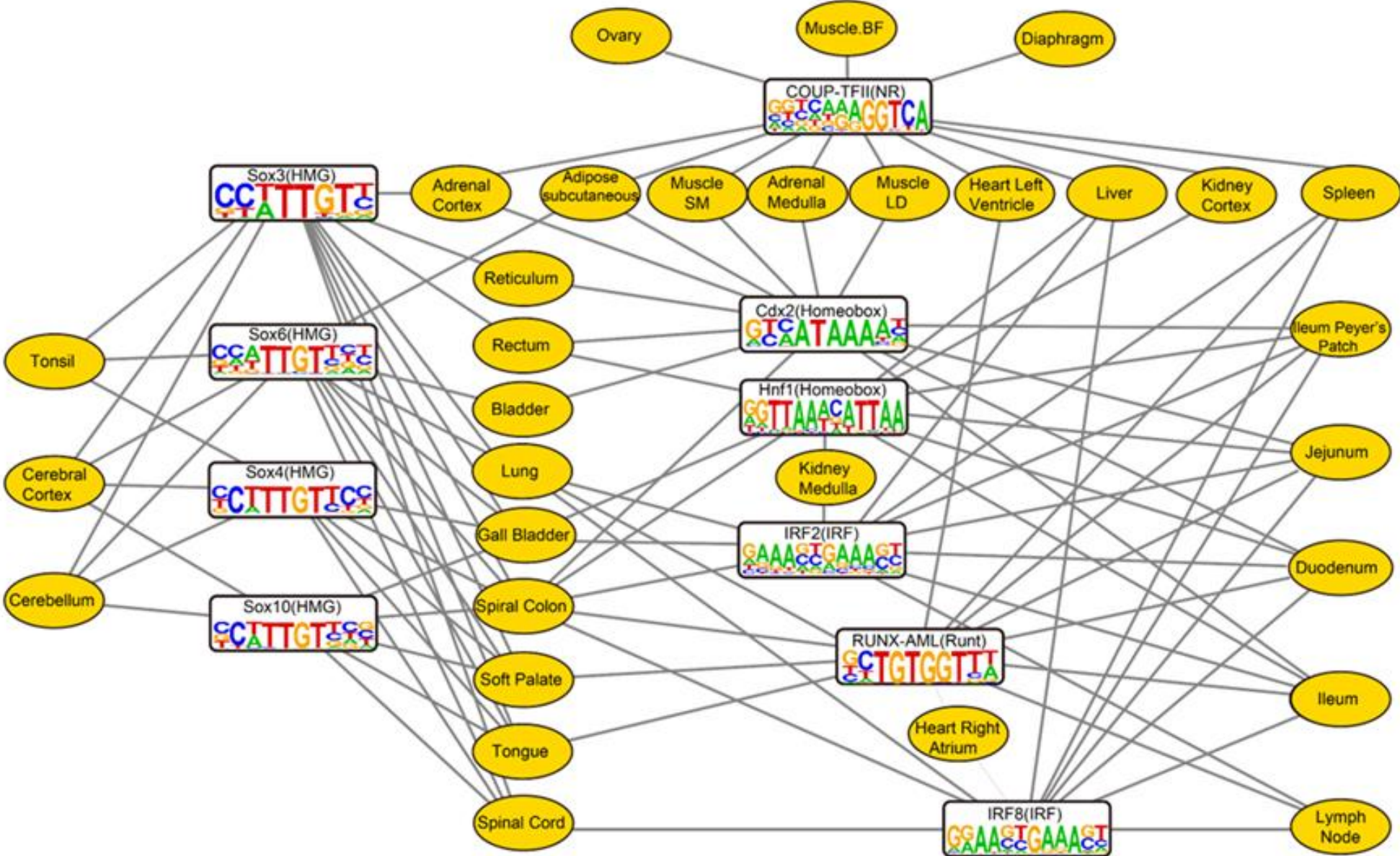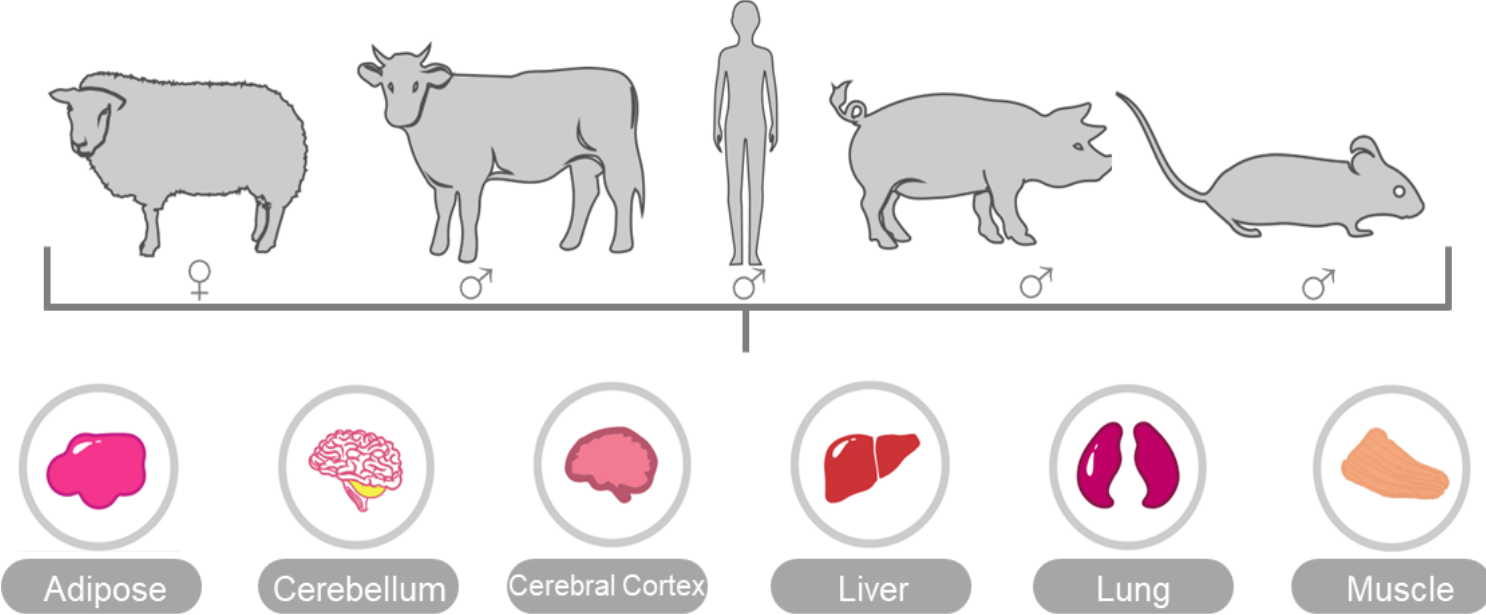- Tongue,
- soft palate,
- spinal cord,
- cerebellum,
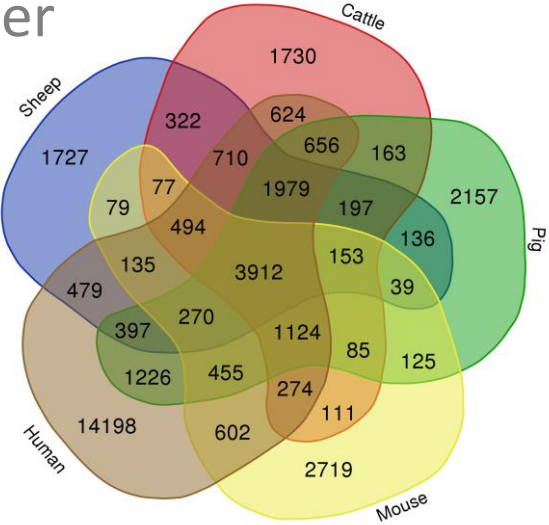- cerebral cortex

Nuclear Receptor TF
- Adrenal

ATAC-seq TFBS across tissues

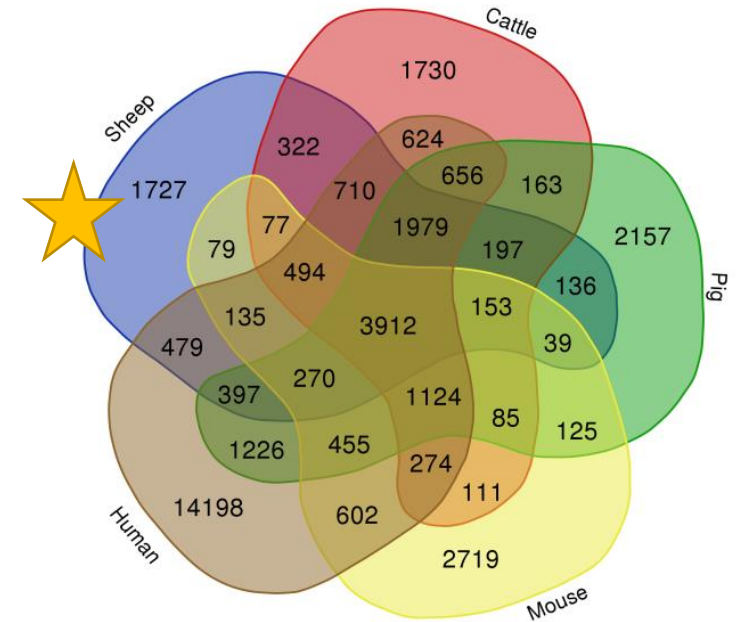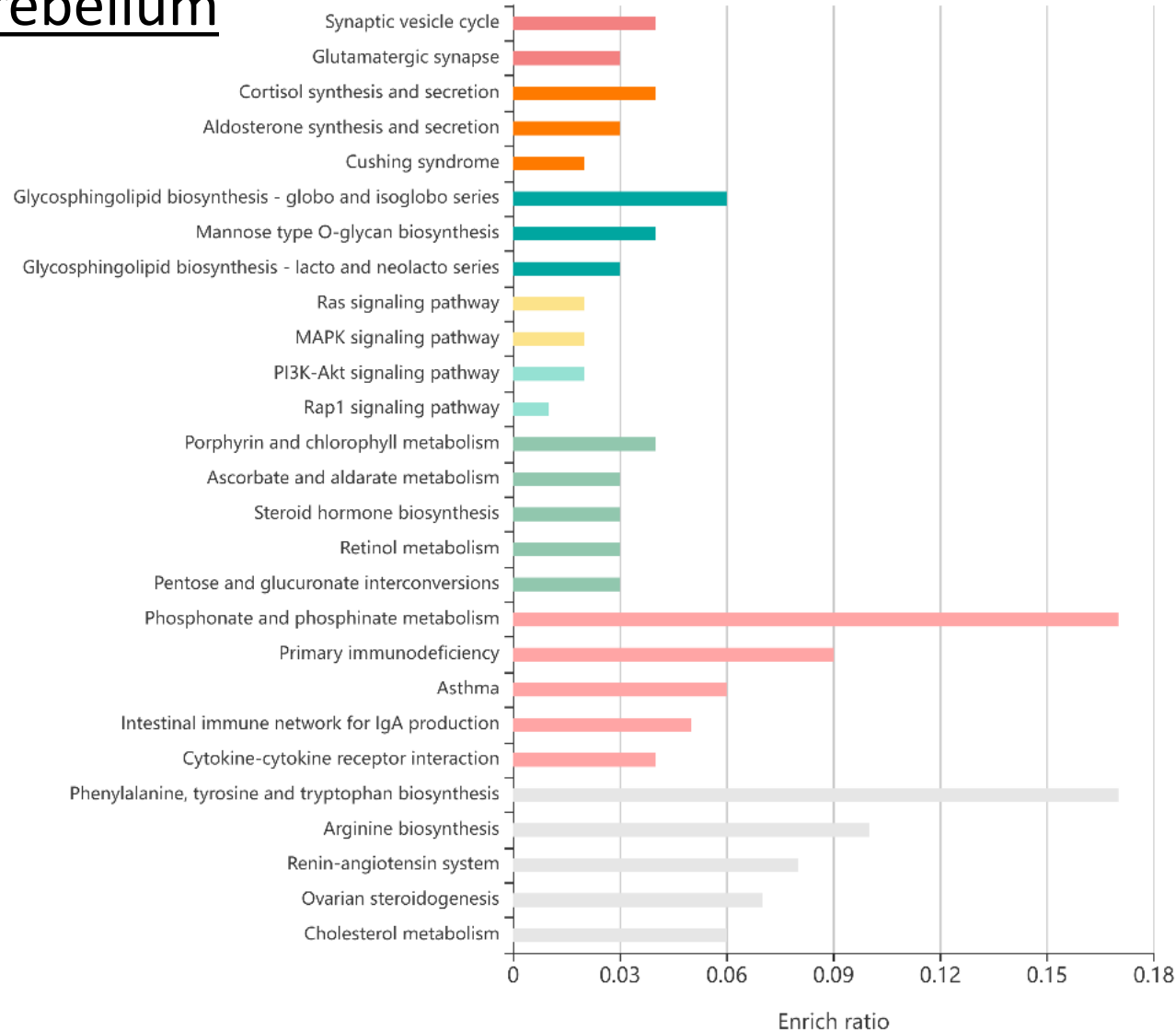# Comparison of Promoter & Enhancer Regions
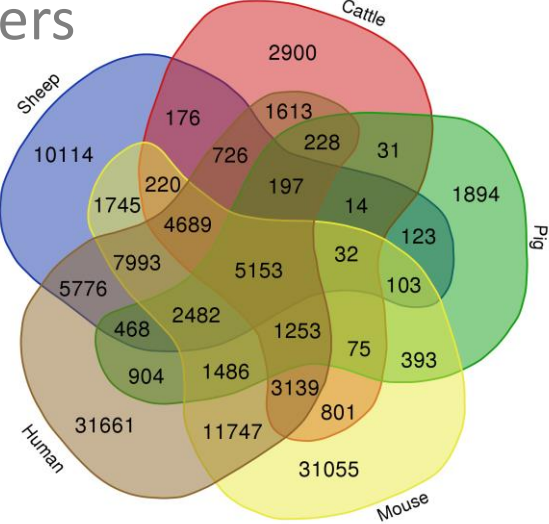
Comparison of Promoters (KEGG)
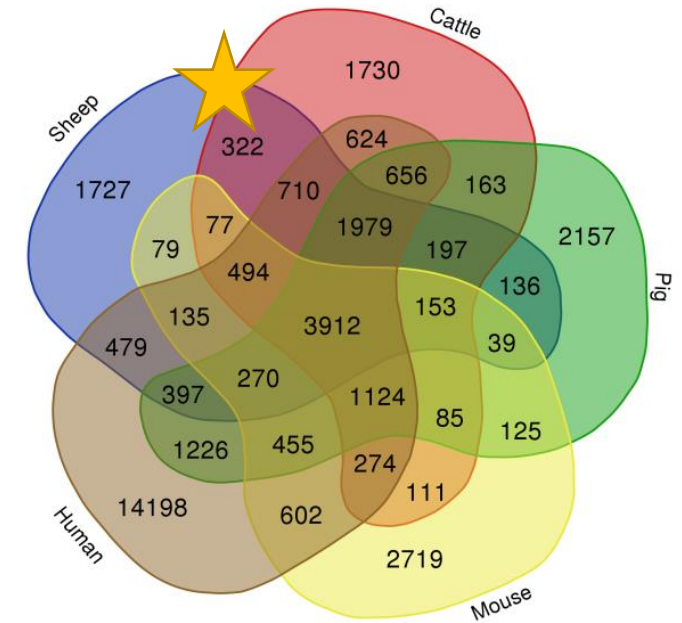
Cerebellum

Adipose

- Phenylalanine, tyrosine & tryptophan biosynthesis (ER=0.17)

Cortex

- Riboflavin metabolism (ER=0.13)
- Renin-angiotensin system (ER=0.12)

Comparison of Promoters

- 322 promoters were in common in sheep and cattle not in non-ruminants
  - suggested that their function might be related to ruminant biology

- KEGG pathways enriched of the ruminant-specific promoters included butanoate metabolism
  - significant role in energy metabolism known to influence rumen nutrition utilization and function
  - Three genes (*BDH1*, *ACSM3* and *ALDH5A1*) identified in sheep and cattle contributed to the butanoate metabolism

- Confirmed increase transcript of *ACSM3* in the liver of sheep and cattle in comparison to pig

Summary

- Defined regulatory elements across tissues in sheep

- Active promoter and enhancer regions reside in open chromatin, have less methylation and greater transcription

- Poised and repressed enhancers are not in open chromatin regions, have greater methylation and less transcription

- Basis for understanding tissues specific and conserved gene regulation of ruminant species

Acknowledgements

We would like to thank all those that contributed to the FAANG tissue collection at USU and the International Sheep Genomics Consortium.

Questions?