Perspectives and Discoveries from the US Bovine FAANG project

Dr. James Koltes

Iowa State University

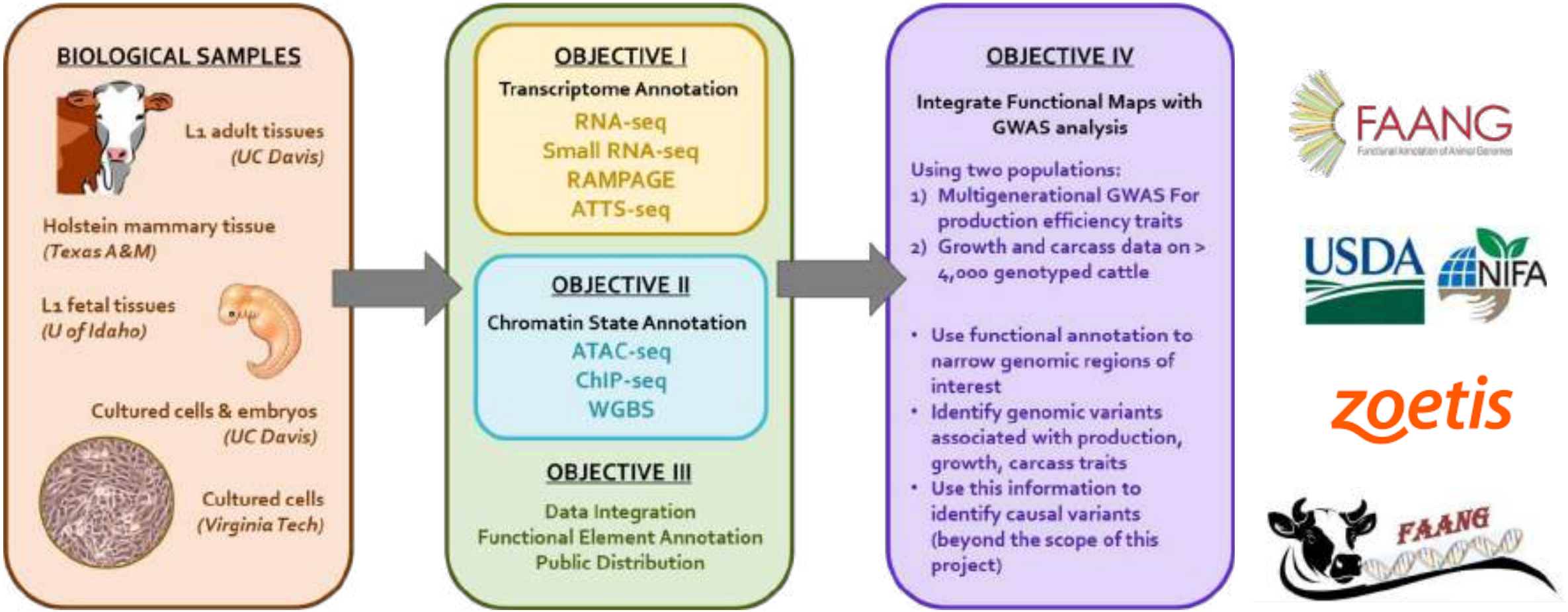
jekoltes@iastate.edu

Functional Annotation of the Bovine Genome

USDA-NIFA-AFRI GRANT NO: 2018-67015-27500

Huaijun Zhou, Pablo Ross, Stephanie McKay, Monique Rijnkels, Brenda Murdoch, Tim Smith, Clare Gill, Zhihua Jiang, James Reecy, James Koltes, Wansheng Liu, Honglin Jiang, Juan Mendrano, Angela Canovas, Graham Plastow

University of California Davis, University of Vermont, Texas A&M University, University of Idaho, USDA-ARS, Washington State University, Iowa State University, Pennsylvania State University, Virginia Tech, University of Guelph, University of Alberta.



Additional Contributors

Sarah Corum, Gonzalo Rincon, Avinash Baktula, Donald Nkrumah

Shally Xu, Zhangyuan Pan, Dailu Guan, Ying Wang, Claire Prowse-Wilkins

Kim Davenport, Gabrielle Becker, Morgan Stegemiller, Denise Konetchy

Jannifer Michal, Xiaohui Zhang, Yunqi Zhang, Hongyang Wang

Pengcheng Lyu

Kristen Kuhn

Chandlar Kern, Wenzhi Ma

Hamid Beiki
Bruna Petry

Guosong Wang

Tissues and cells analyzed for RNA-seq and small RNA-seq (47 total)

Tissue	Tissue	Tissue	Tissue
abomasum	esophagus	tongue	ovary
adipose	fetal brain	jejunum	pituitary
mammary gland (adult)	fetal gonad	kidney cortex	placental caruncle
bladder	fetal heart	kidney	placental cotyledon
bone marrow	fetal kidney	mammary gland (latepregnant)	reticulum
brain (frontal cortex)	fetal liver	longissimus dorsi muscle	rumen
cecum	fetal muscle	lung	skin
cerebellum	fetal spleen	lymph nodes	spleen
colon	fetal umbilicalcord	mammary gland (virgin)	testis
duodenum	follicular cells	mammary gland (midpregnant)	thymus
mammary gland (earlylactating)	gall bladder	myoblast	uterine endometrium
epididimus caput	Ileum	omasum	

Adult and fetal tissues from cattle closely related to Dominette (Hereford; UCD_ARS1.2)

Mammary gland tissue from Holstein

Samples collected and epigenomic data generated

- 40 adult tissues from L1 Hereford line
- 8 fetal tissues from L1 Hereford line
- 4 primary cell lines (Pre-Myocytes, Myocytes, Pre-Adipocytes, Adipocytes)
- 5 stages of Holstein mammary gland development

ASSAYS-BY-SEQUENCE

Expressed regions **(47 cell/tissues)**

RNA-seq	Large transcripts expression - variants
smRNA-seq	Small transcript expression
RAMPAGE	Transcription start sites – promoter activity
WTTS-seq	Transcription termination sites

Chromatin states **(40 cells/ tissues)**

WGBS		DNA methylation
ATAC-seq		Open Chromatin profiling
ChIP-seq	H3K4me3	Active promoters
	H3K27me3	Polycomb repression
	H3K4me1	Active enhancers
	H3K27ac	Enhancers and promoters
	CTCF	Insulators and promoters
	H3K9me3	Heterochromatin
	H3K36me3	Active gene bodies

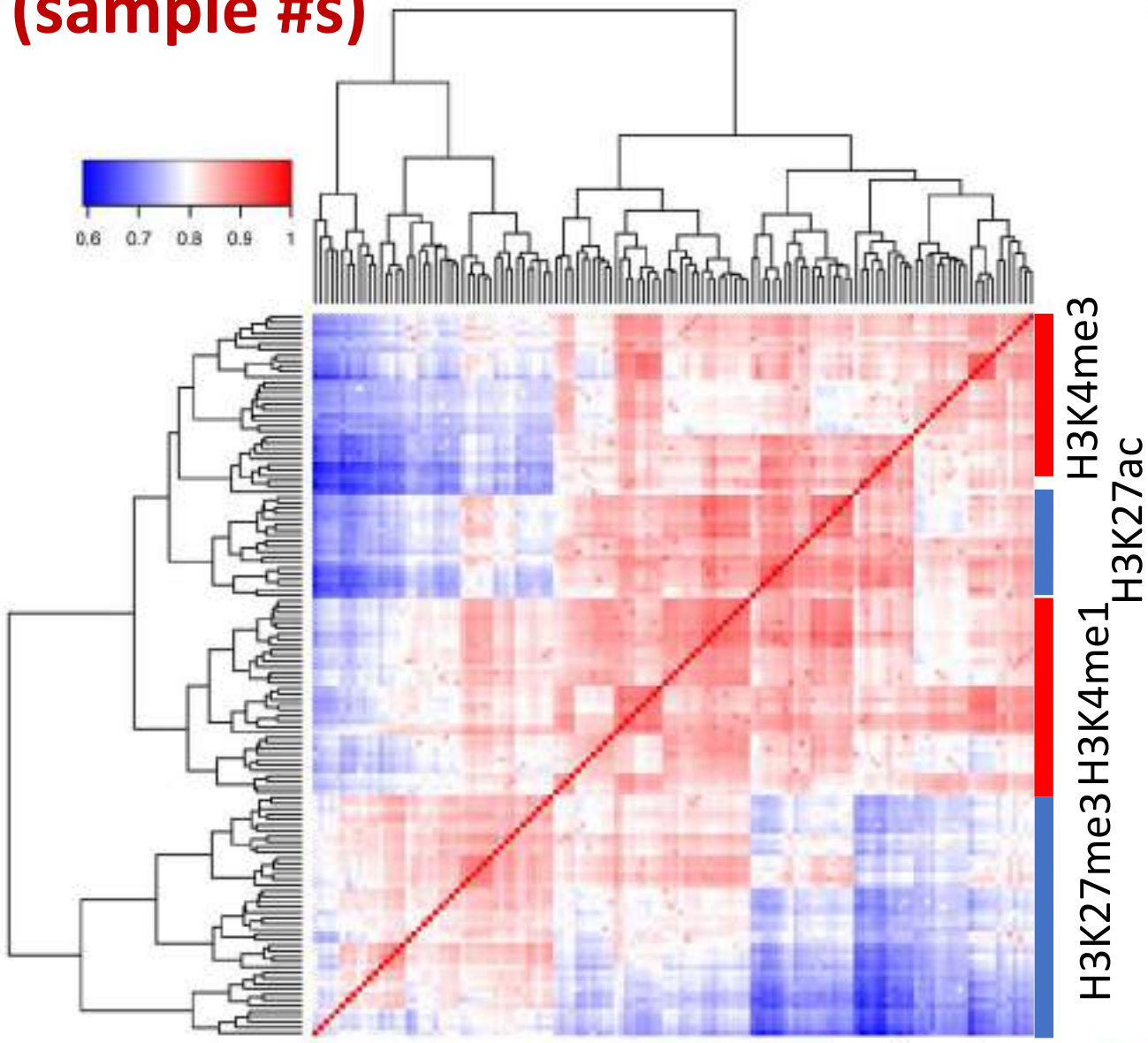
Generation of epigenomic data (sample #s)

- Histone marks (ChIP-seq)

	Adult tissue	Fetal tissue	Cell line	Total
CTCF	52	20	8	80
H3K27ac	52	20	8	80
H3K27me3	46	20	8	74
H3K4me1	52	20	8	80
H3K4me3	52	20	8	80
H3K36me3	29	20	8	57
H3K9me3	38	20	8	66

- WGBS: 95
- ATAC-seq: 114
- WTTS:76

Replicates per tissue = 2



Functionally annotate epigenomic elements of the bovine genome

(Develop annotations of epigenetic regulatory elements)

Build a map of of regulatory elements by integrating:

a) Project data
   - 27 Adult tissues (Herefords)
   - 5 Fetal tissues (Herefords)
   - Mammary gland (Holstein 5 stages)
   - 4 Cell lines

b) Public data (Holstein) ←
   - 23 Adult tissues
   - 4 Fetal tissues

Australian Data Set

	Adult tissue	Fetal tissue	Cell line	Total
CTCF	47	8	1	56
H3K27ac	47	8	1	56
H3K27me3	47	6	1	56
H3K4me1	51	8	1	60
H3K4me3	51	8	1	60

→ **A total of 46 adult tissues, 6 fetal tissues, and 5 cell lines**

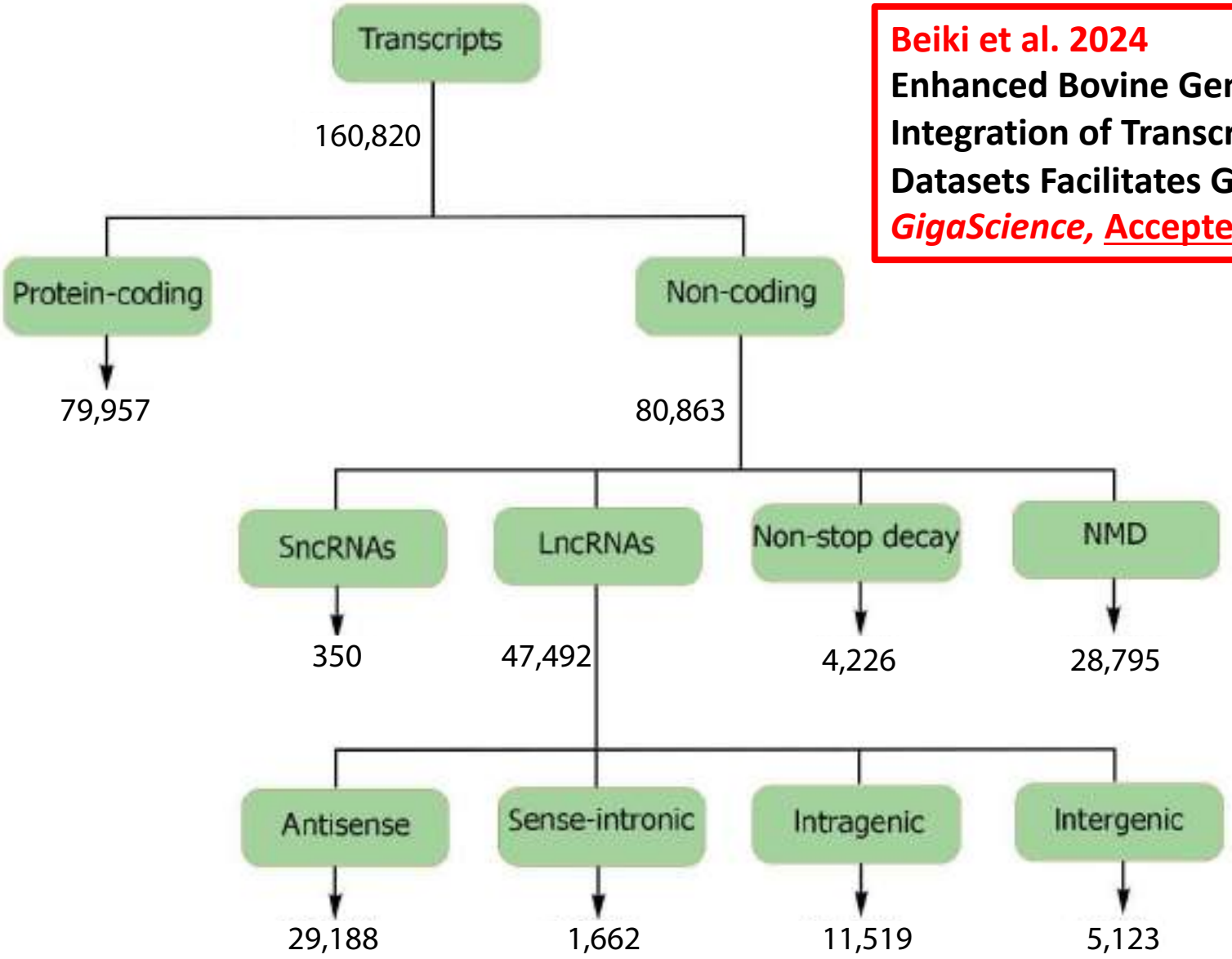
Results: Summary of expressed transcripts/genes

Feature	Annotation[1]		
	Current project	Ensembl (Release 2021-03)	NCBI (Release 106)
Number of genes	35,150 (21,193)	27,607 (21,880)	35,143 (21,355)
Number of transcripts	160,820 (79,957)	43,984 (37,538)	83,195 (47,280)
Number of spliced transcripts	130,531	37,299	73,423
Number of transcripts per gene	4.9	1.5	2.3
Median number of 5' UTRs per gene	2	1	1
Median number of 3' UTRs per gene	1	1	1

[1]Numbers in parentheses indicate the number of protein-coding genes/transcripts.

Beiki et al., 2024. *GigaScience* accepted

Classification of the predicted transcripts into different biotypes and discovery of 110,965 new transcripts



Beiki et al. 2024
Enhanced Bovine Genome Annotation Through Integration of Transcriptomic and Epi-Transcriptomics Datasets Facilitates Genomic Biology
***GigaScience*, Accepted**

- **118,563 transcripts (73% of the total) were structurally validated by independent assays**
- **69% of transcripts were previously unannotated**
  - **86% derived from annotated genes**
  - **14% completed novel**

Genes/Transcripts expressions across tissues between adult and fetal stages

- Fetal testis showed the highest number of expressed genes while fetal brain and muscle tissues showed the highest number and percentage of non-coding genes
- Fetal tissues had a significantly higher rate of alternative splicing events than adult tissues
- Fetal tissues had significantly higher proportions of unique non-coding transcripts compared to protein-coding transcripts than adult tissues
- 106 non-coding genes identified in fetal tissues that switched to protein-coding genes in their matched adult tissues

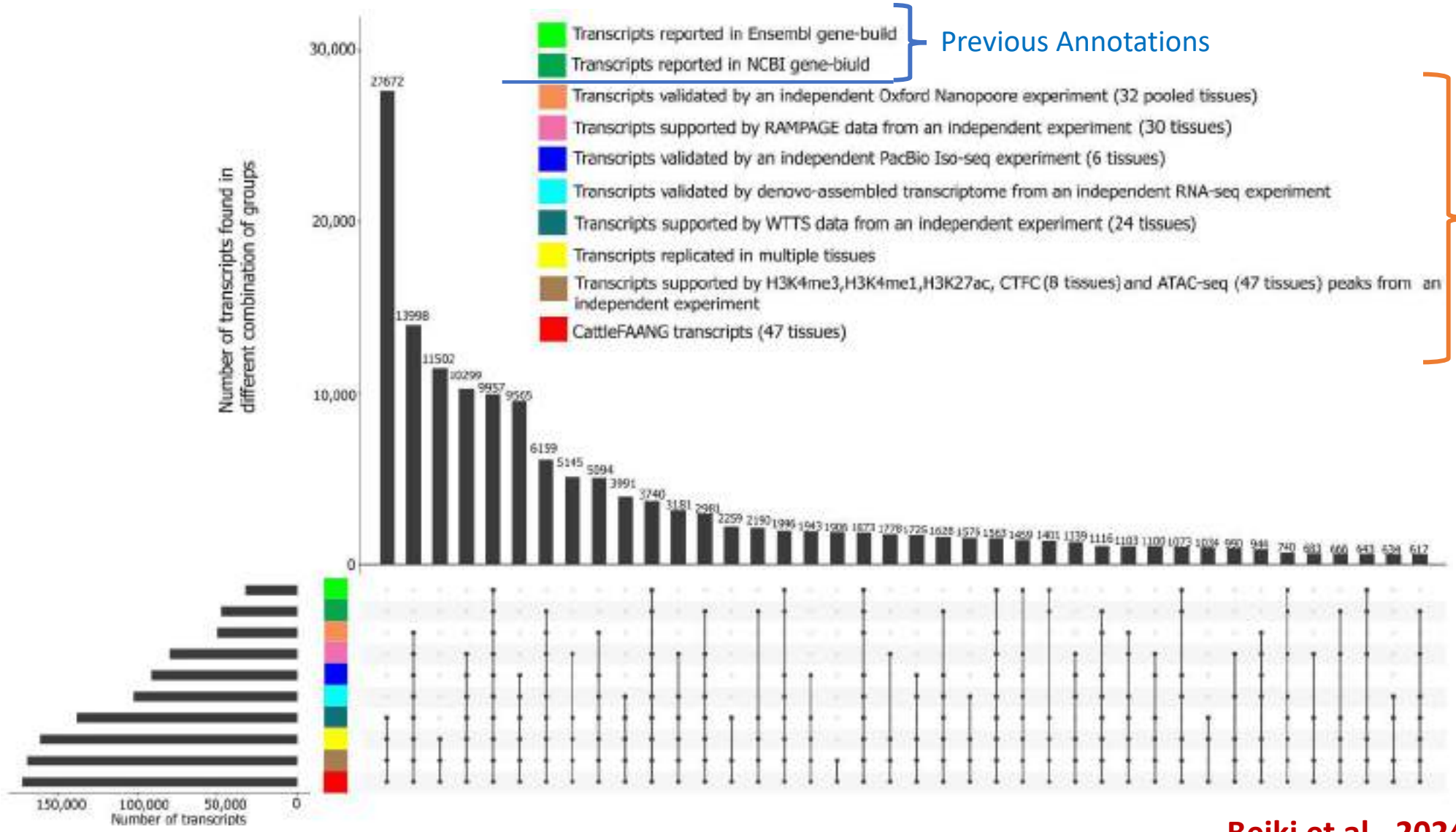
Beiki et al., 2024. *GigaScience* accepted

Tissue specificity

- Nine percent of all genes (3,174) and transcripts (15,562) were only expressed in a single tissue
- The majority of tissue-specific genes (75%) and transcripts (84%) were un-annotated
- Testis and thymus had the highest number of tissue-specific genes and transcripts
- As we expected, the expression level of tissue-specific genes and transcripts was significantly lower than that of their non-tissue-specific counterparts

Beiki et al., 2024. *GigaScience* accepted

Validation of predicted *transcripts* using independent data from different technologies



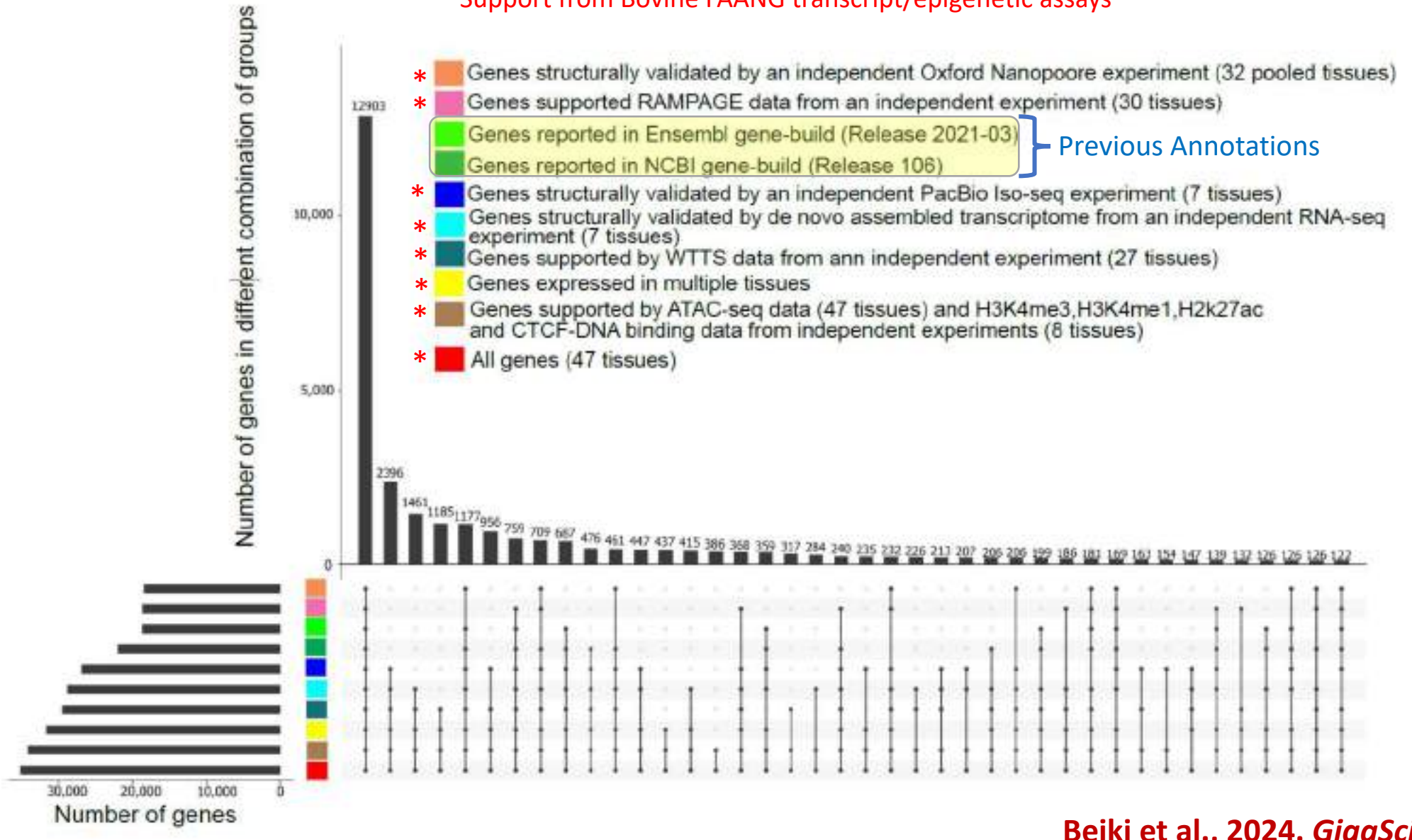
Previous Annotations

Supported by *different* transcript/ epigenetic Bovine FAANG assays

Beiki et al., 2024. *GigaScience* accepted

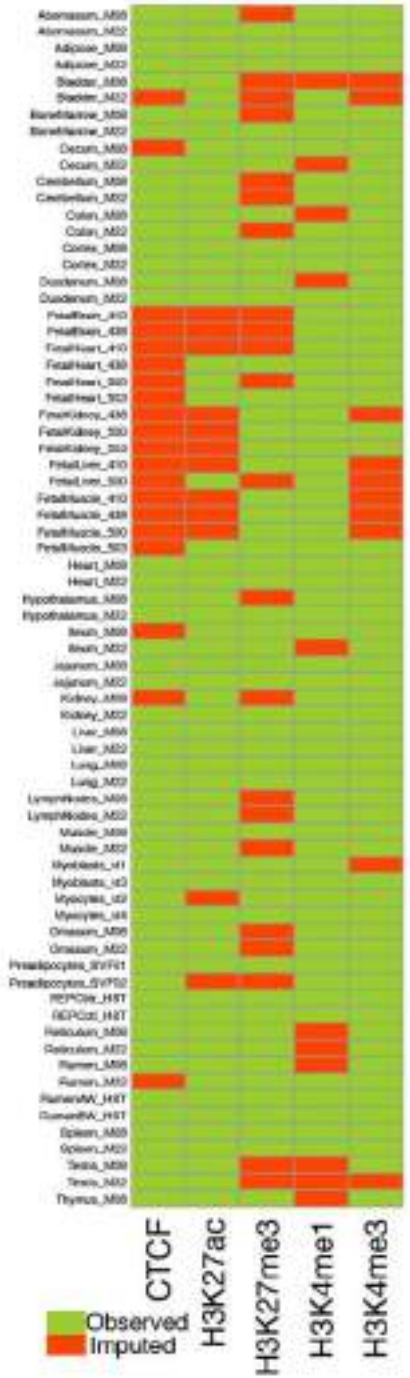
Validation of predicted *genes* using independent data from different technologies

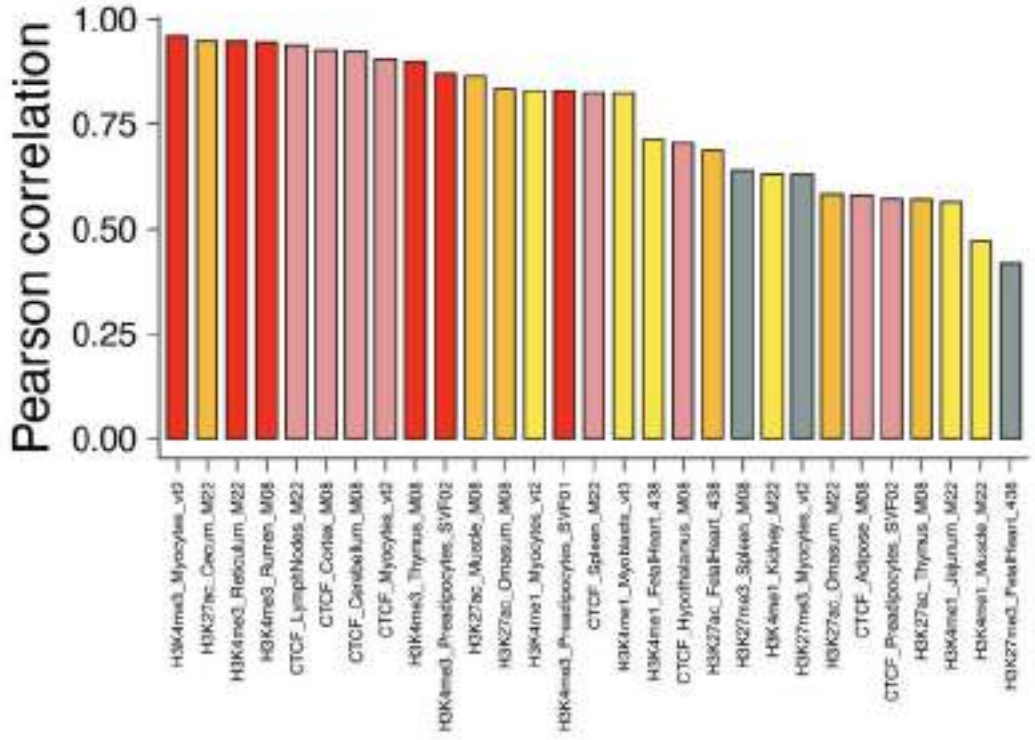
*Support from Bovine FAANG transcript/epigenetic assays

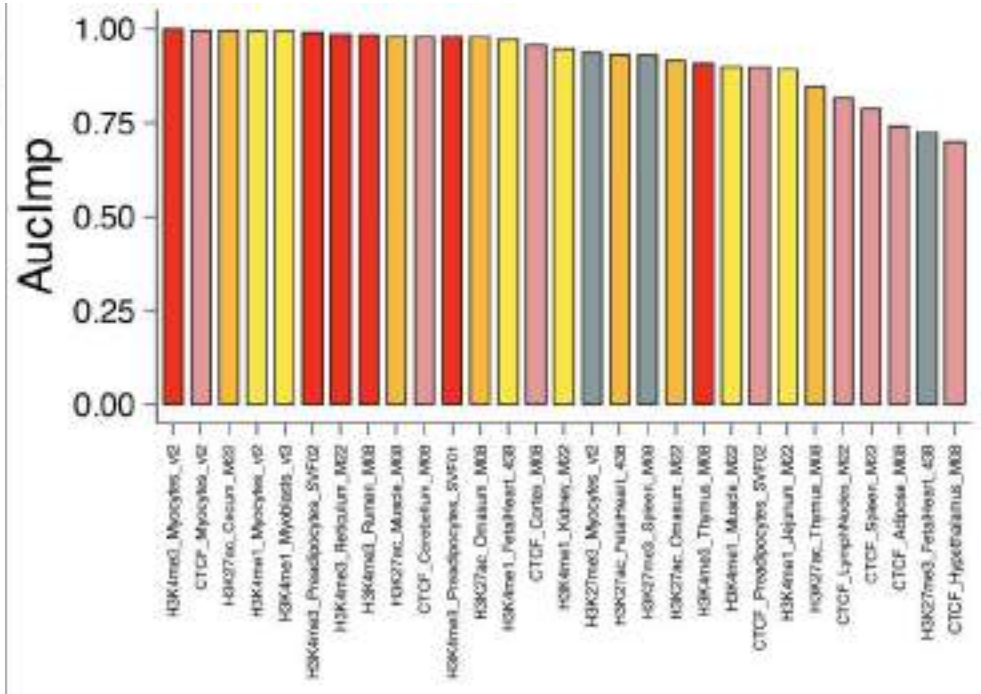


Beiki et al., 2024. *GigaScience* accepted

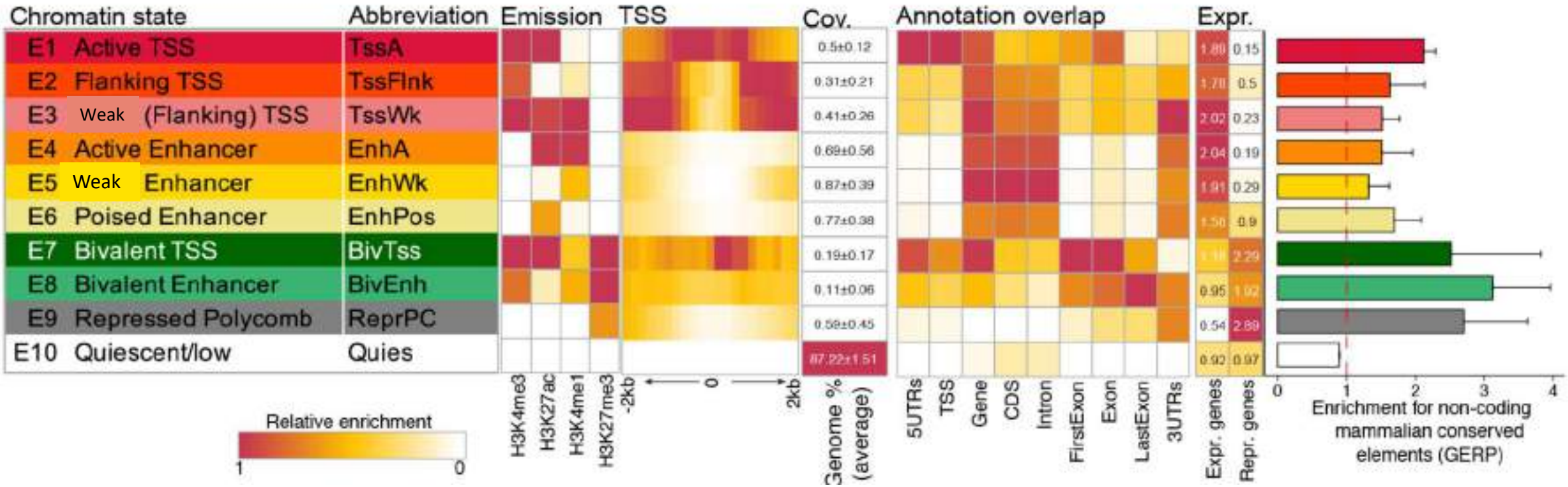
Imputation of missing ChIP-seq marks





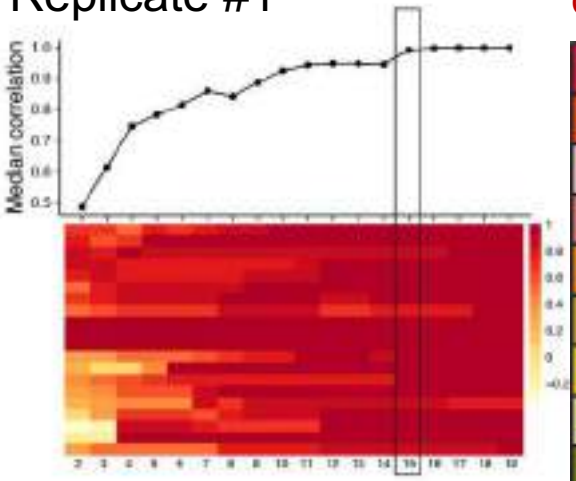


Initial chromatin state models: ‘Core’ 10-state model

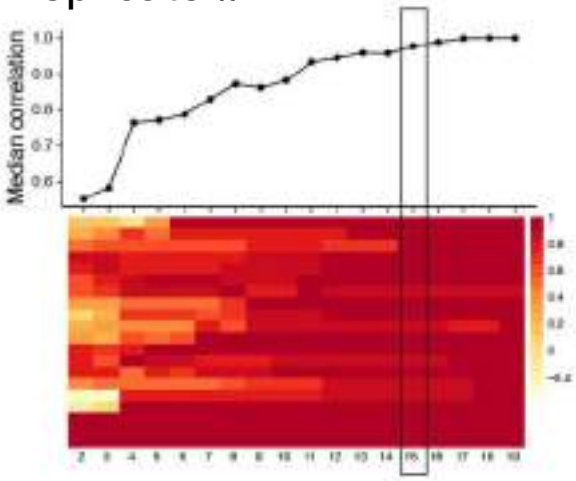


‘Expanded’ 15-state model

Replicate #1



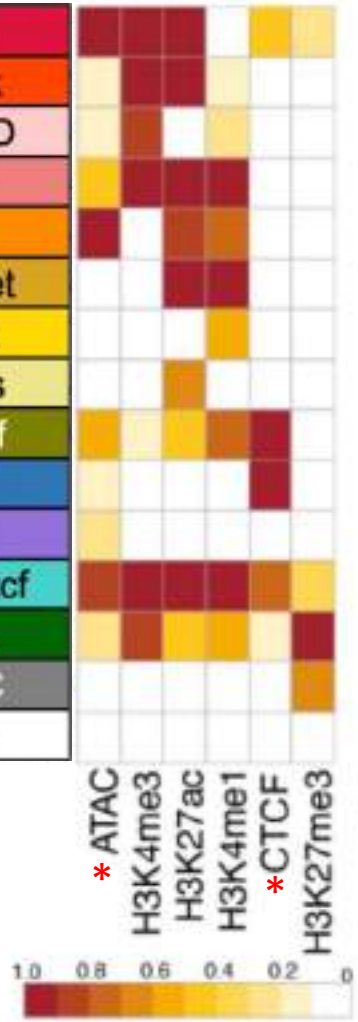
Replicate #2



Chromatin States

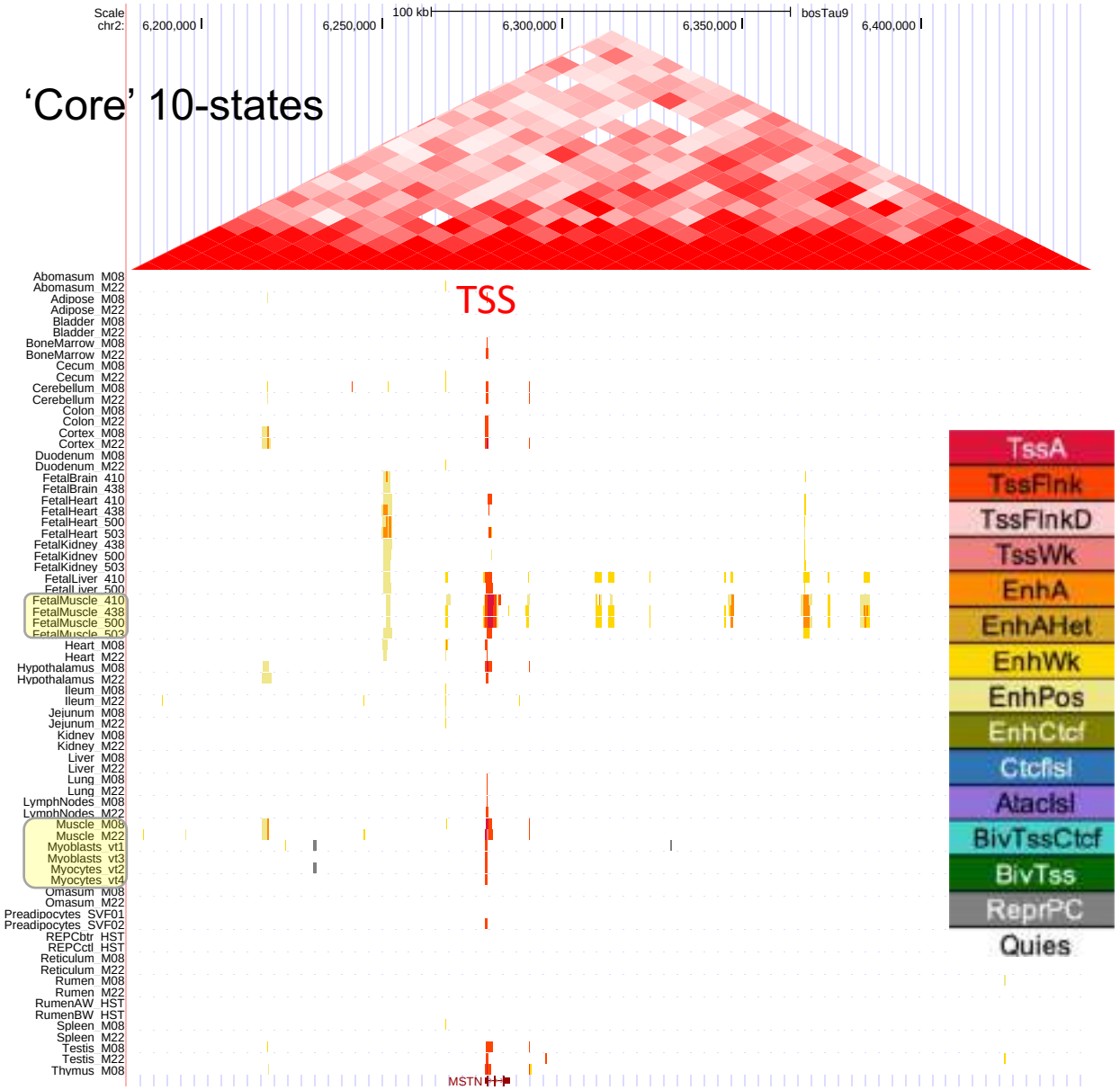
E1	Active TSS	TssA
E2	Flanking TSS	TssFlnk
E3	Flanking TSS downstream	TssFlnkD
E4	Weak Flanking) TSS	TssWk
E5	Active Enhancer	EnhA
E6	Active Enhancer without ATAC (hetero)	EnhAHet
E7	Weak Enhancer	EnhWk
E8	Poised Enhancer	EnhPos
E9	Enhancer with CTCF	EnhCtcf
E10	CTCF island	CtcfIsl
E11	ATAC island	AtacIsl
E12	Bivalent TSS with CTCF	BivTssCtcf
E13	Bivalent TSS	BivTss
E14	Repressed Polycomb	ReprPC
E15	Quiescent/low	Quies

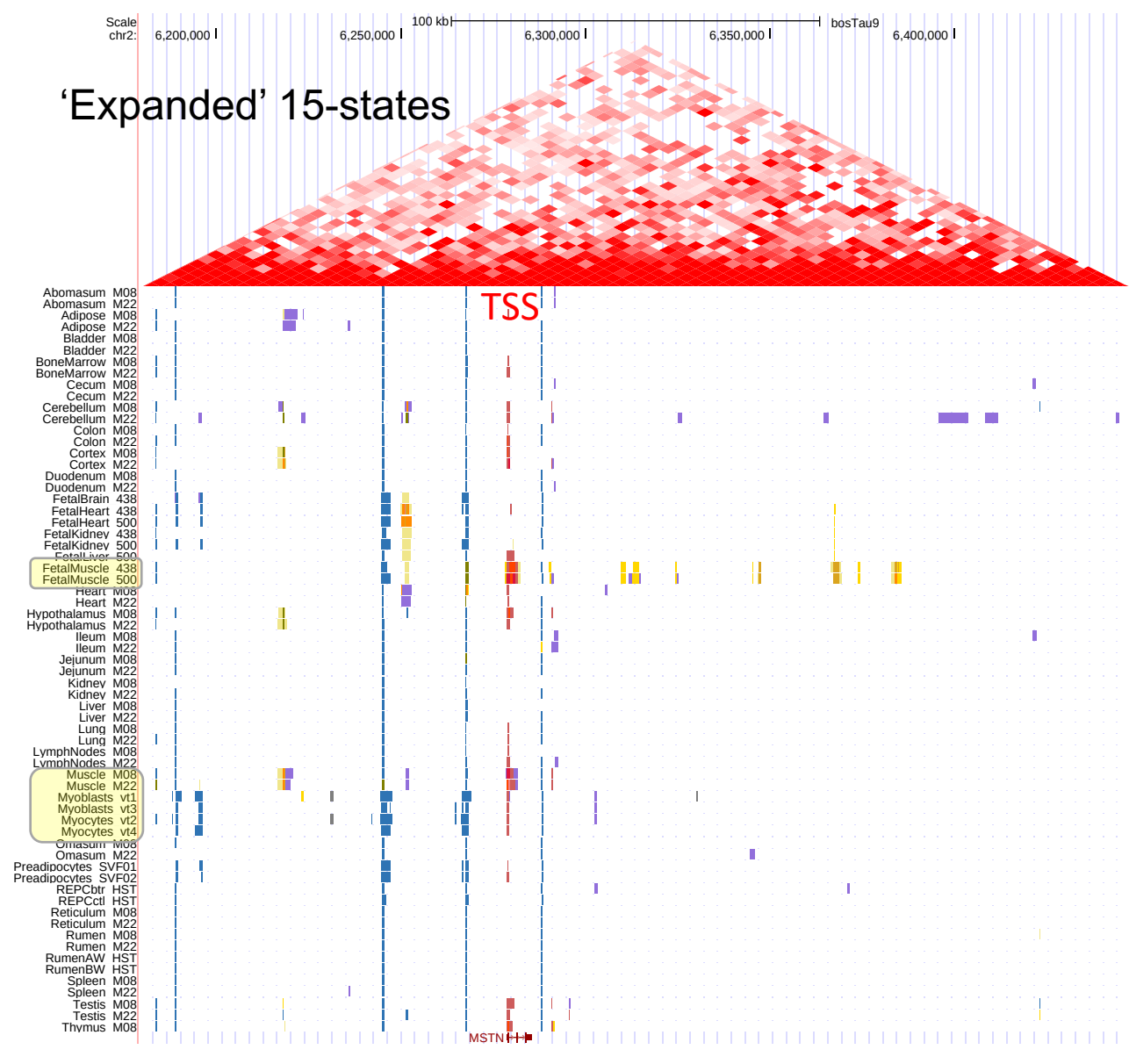
Two additional Histone marks



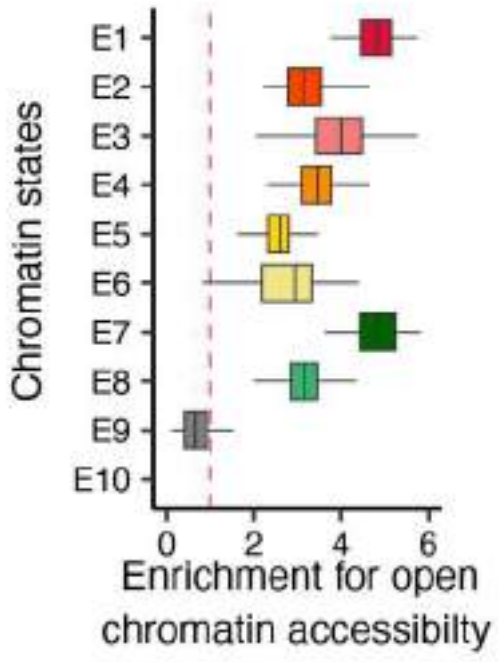
Location relative to TSS

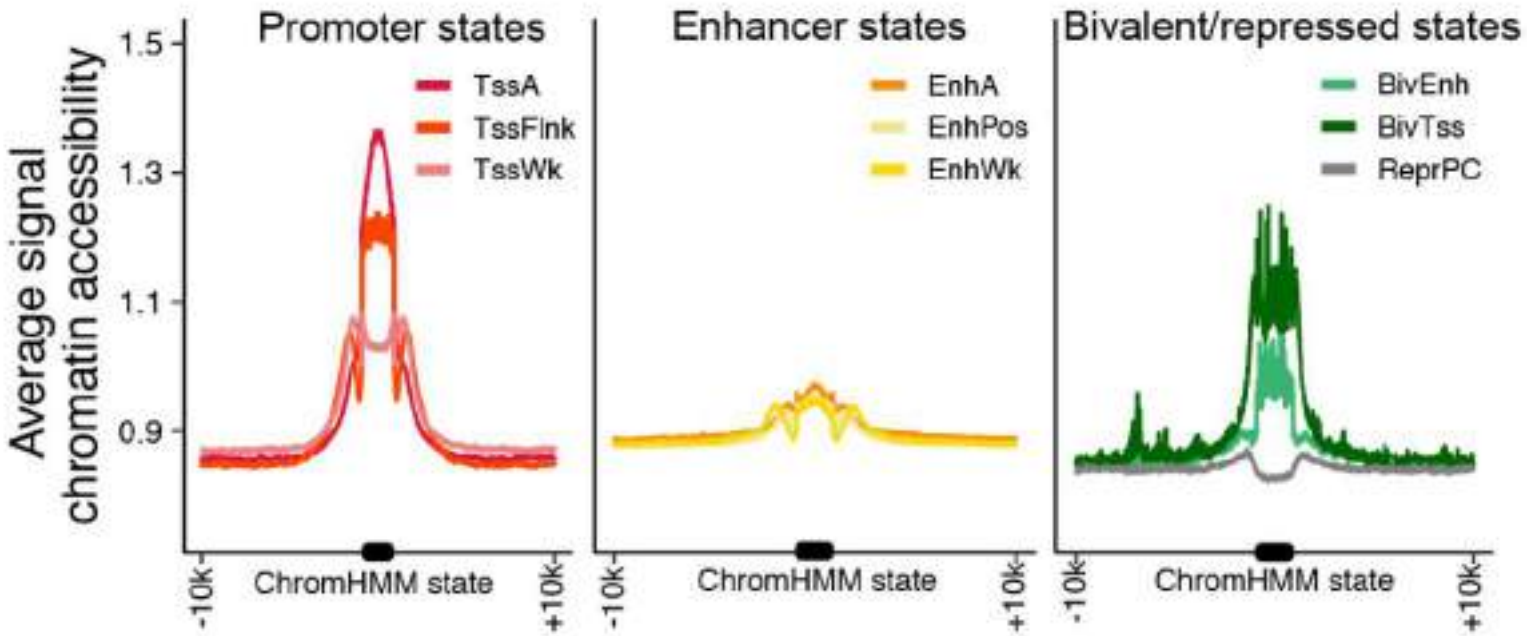
Chromatin states at the *MSTN* locus





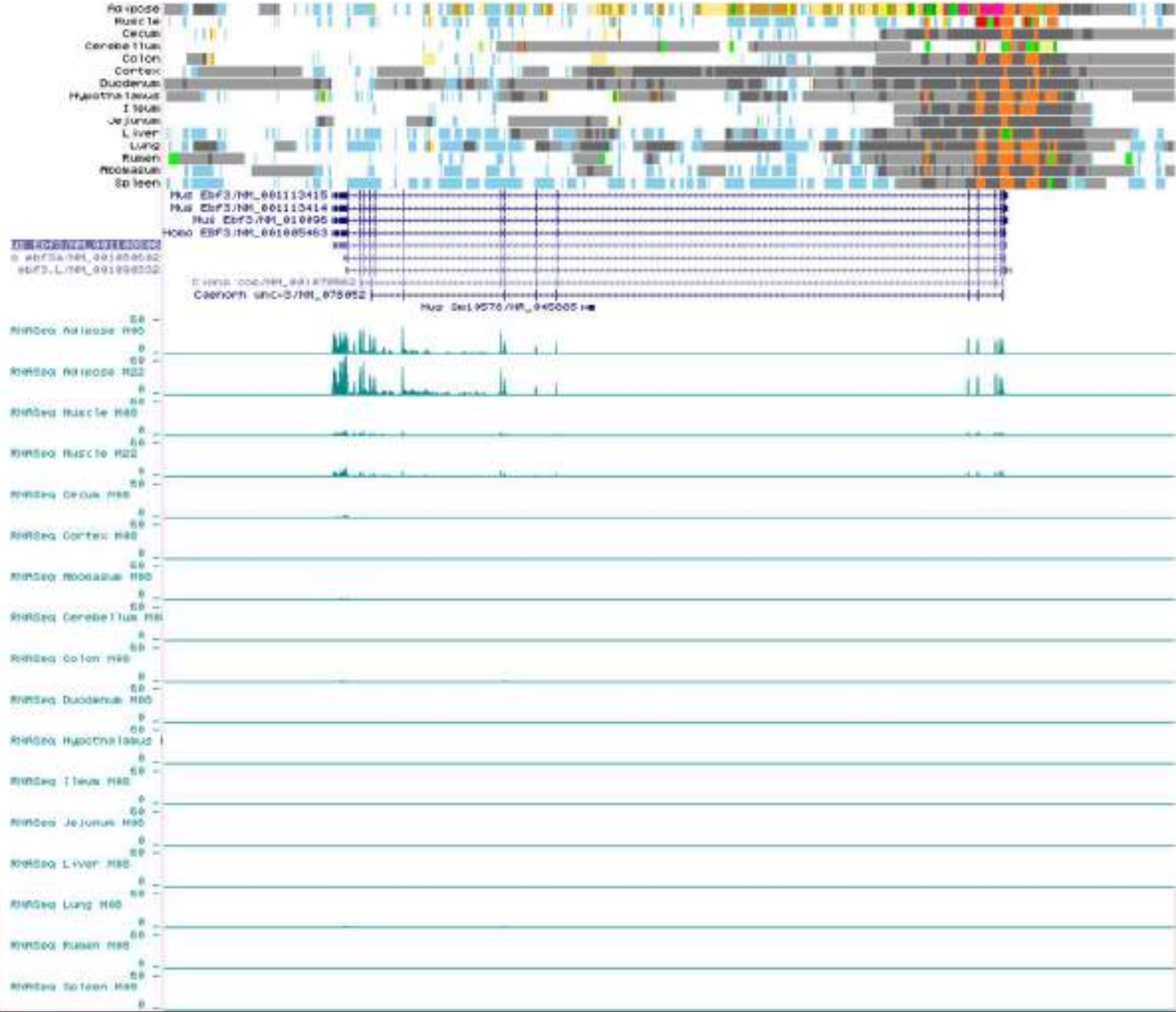
Open chromatin accessibility regions are indicative of active regulatory elements



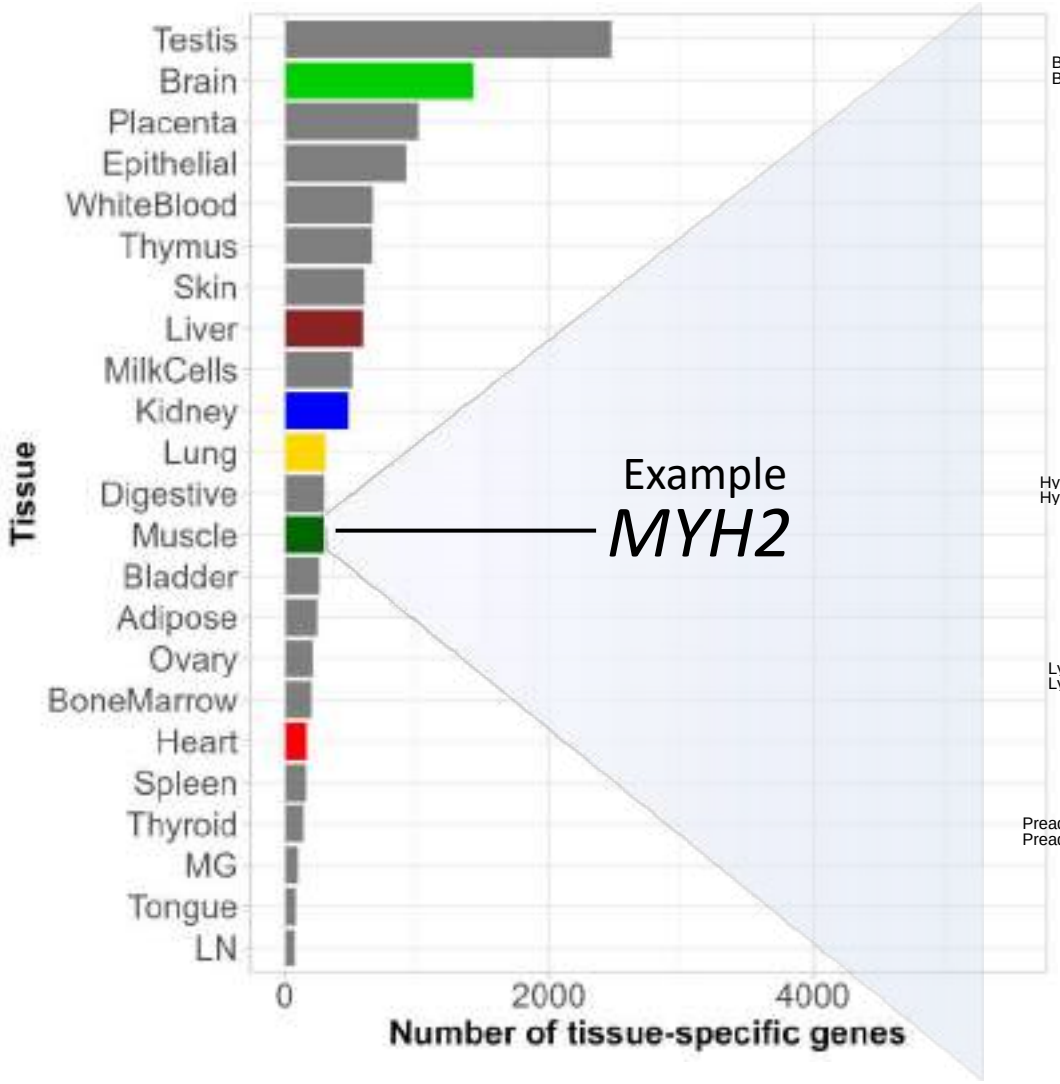


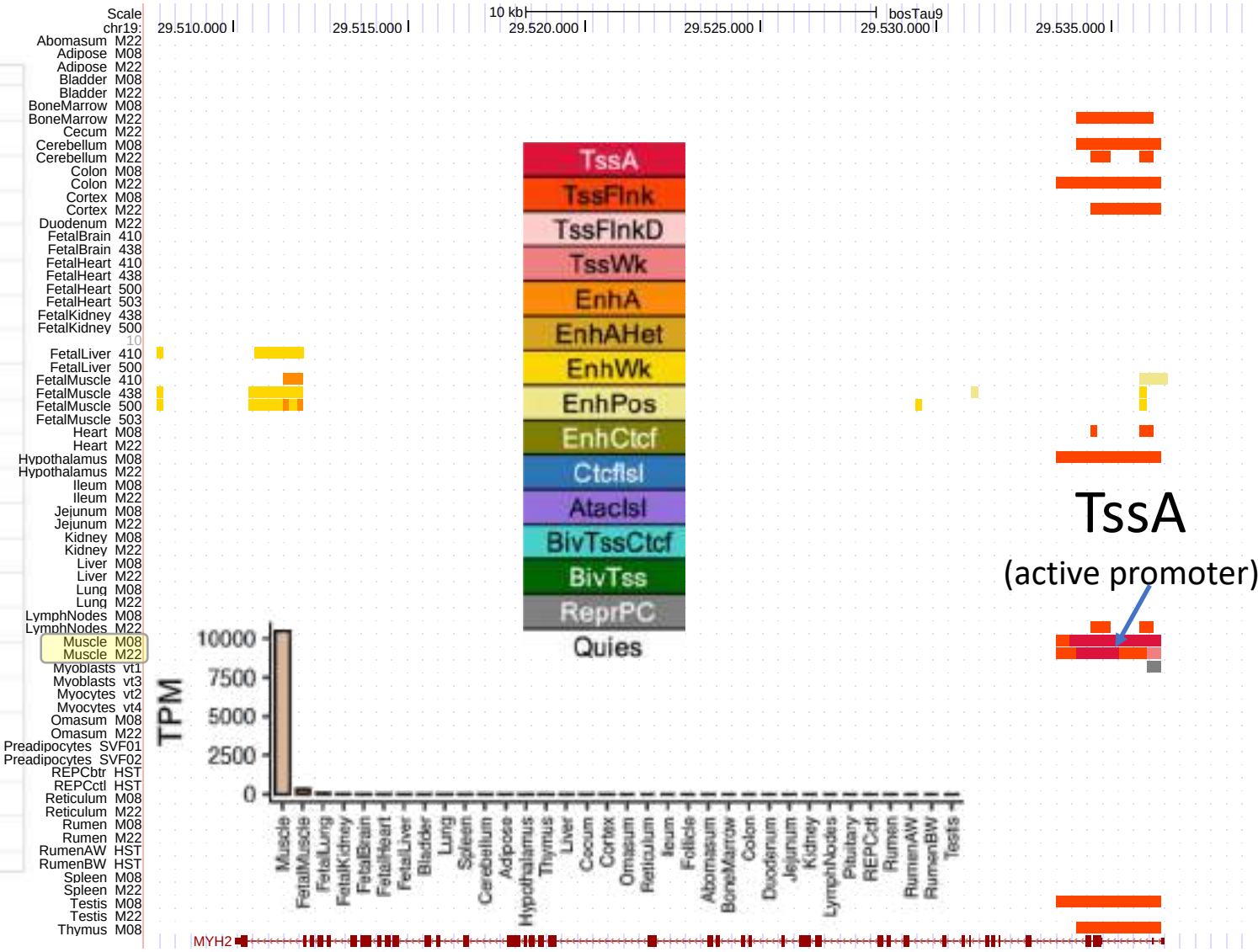
Example of chromatin state tissue specificity

- *EBF3* specifically expressed in adipose tissue.
- Adipose has many tissue-specific enhancers

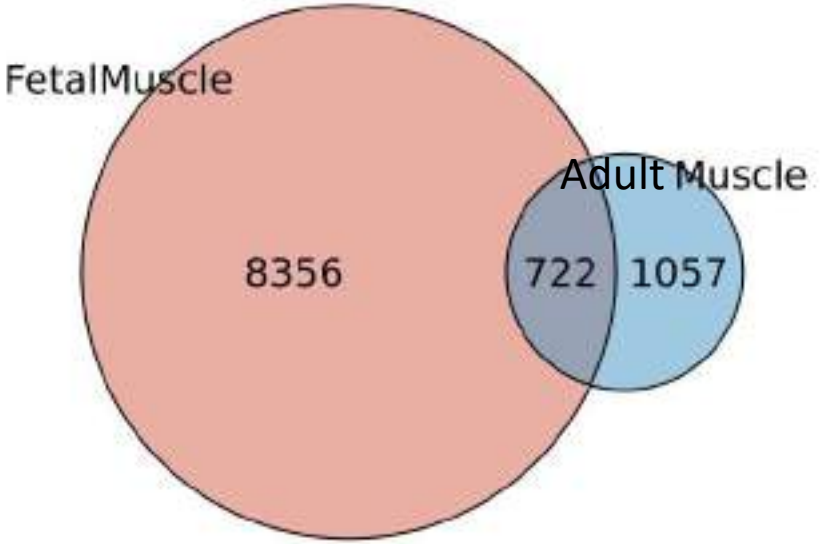


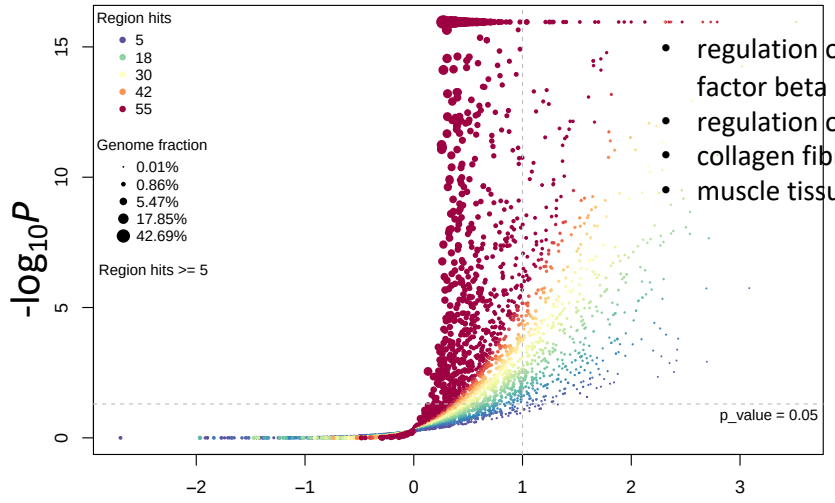
Tissue-specific gene expression contributed by tissue-specific regulators- muscle MYH2 example



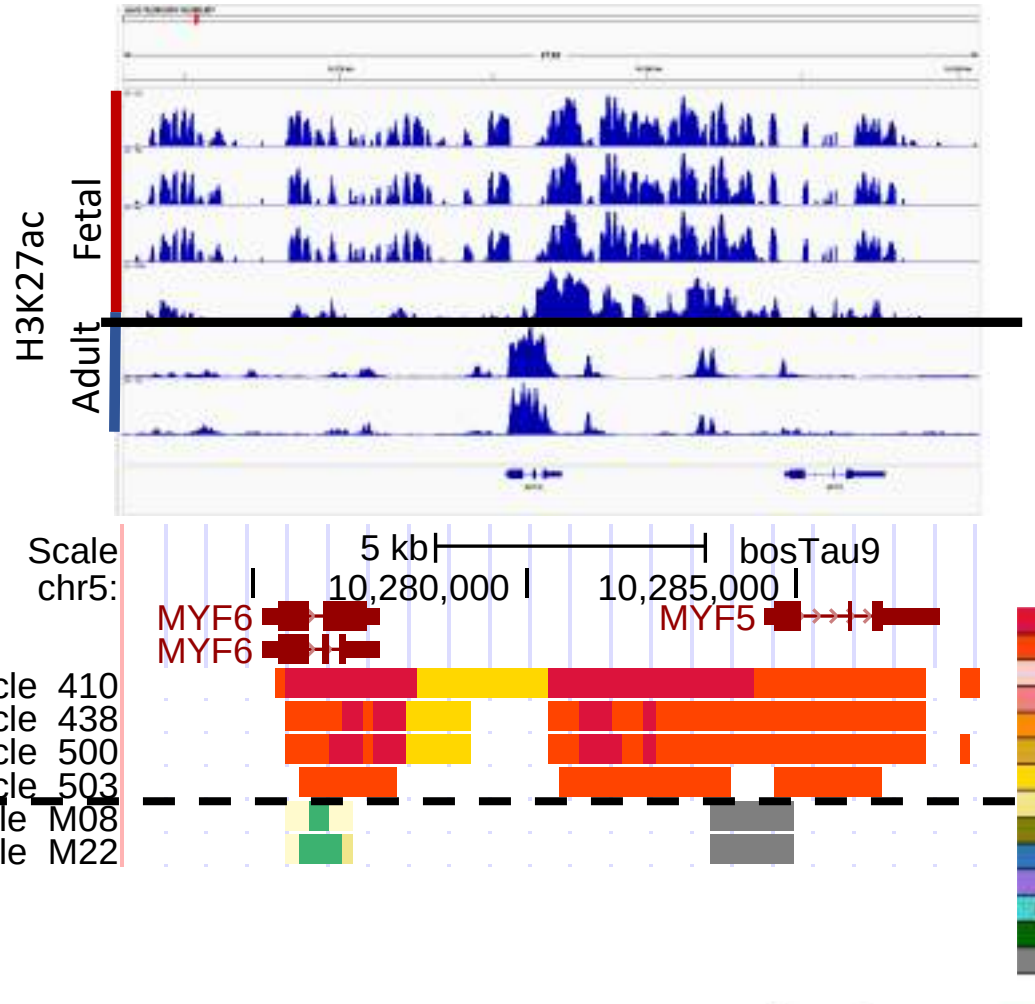


Differential active enhancer (E4) between fetal and adult muscle: changes in chromatin state over development- muscle MYF 5 & 6 examples

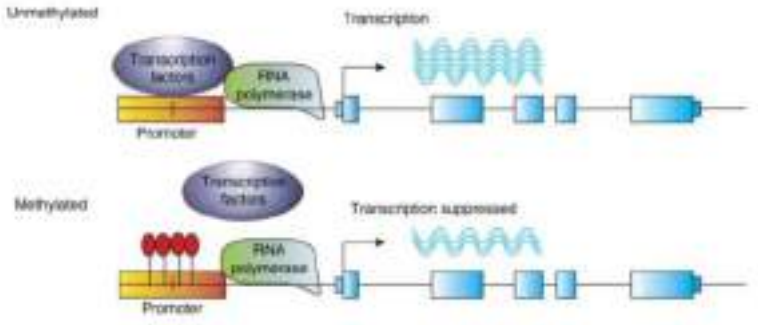




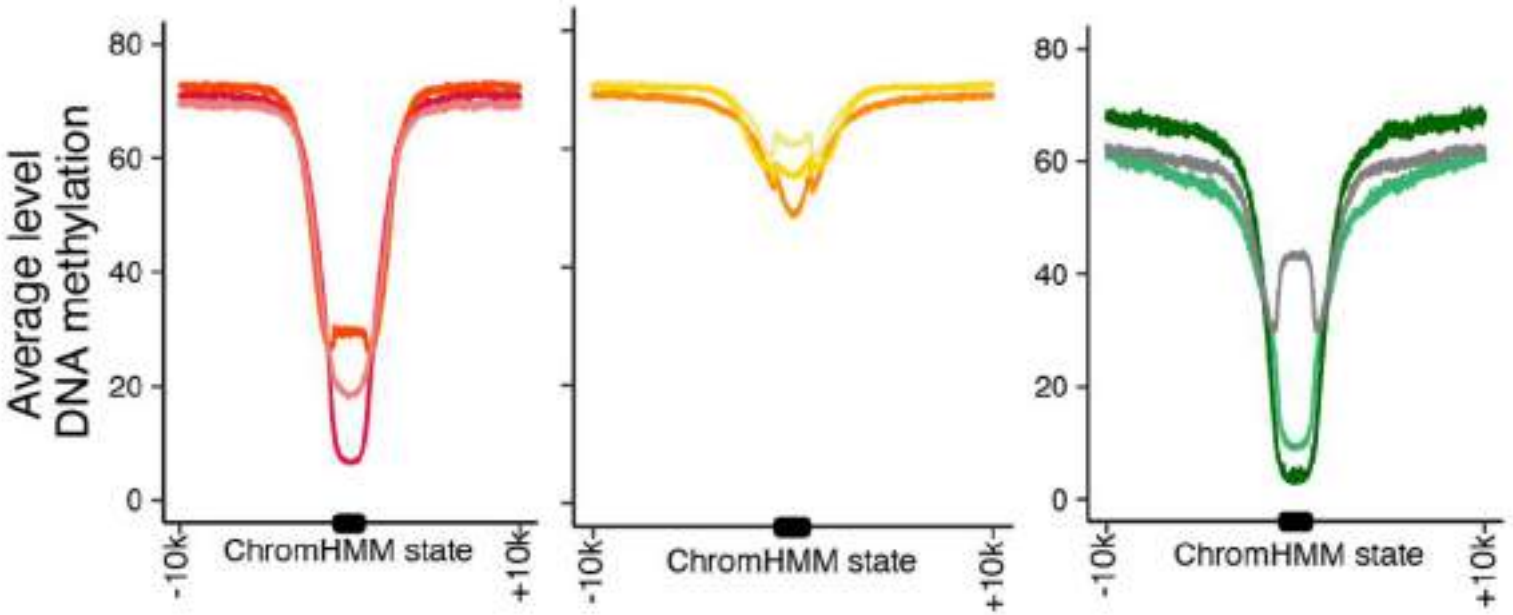
- regulation of transforming growth factor beta production
- regulation of myelination
- collagen fibril organization
- muscle tissue development

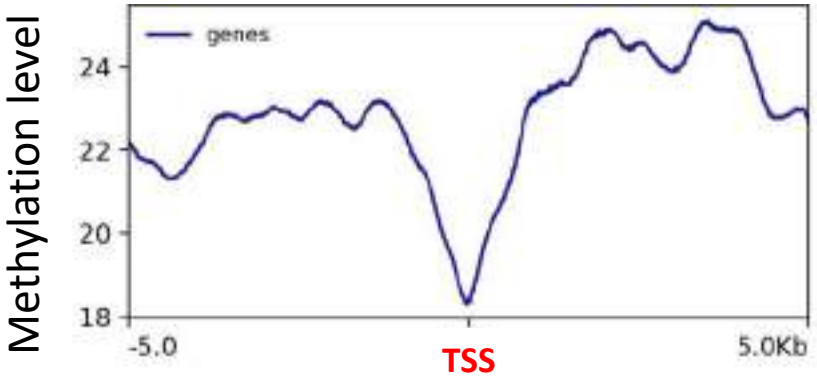


DNA methylation and chromatin states



Wu et al. 2018



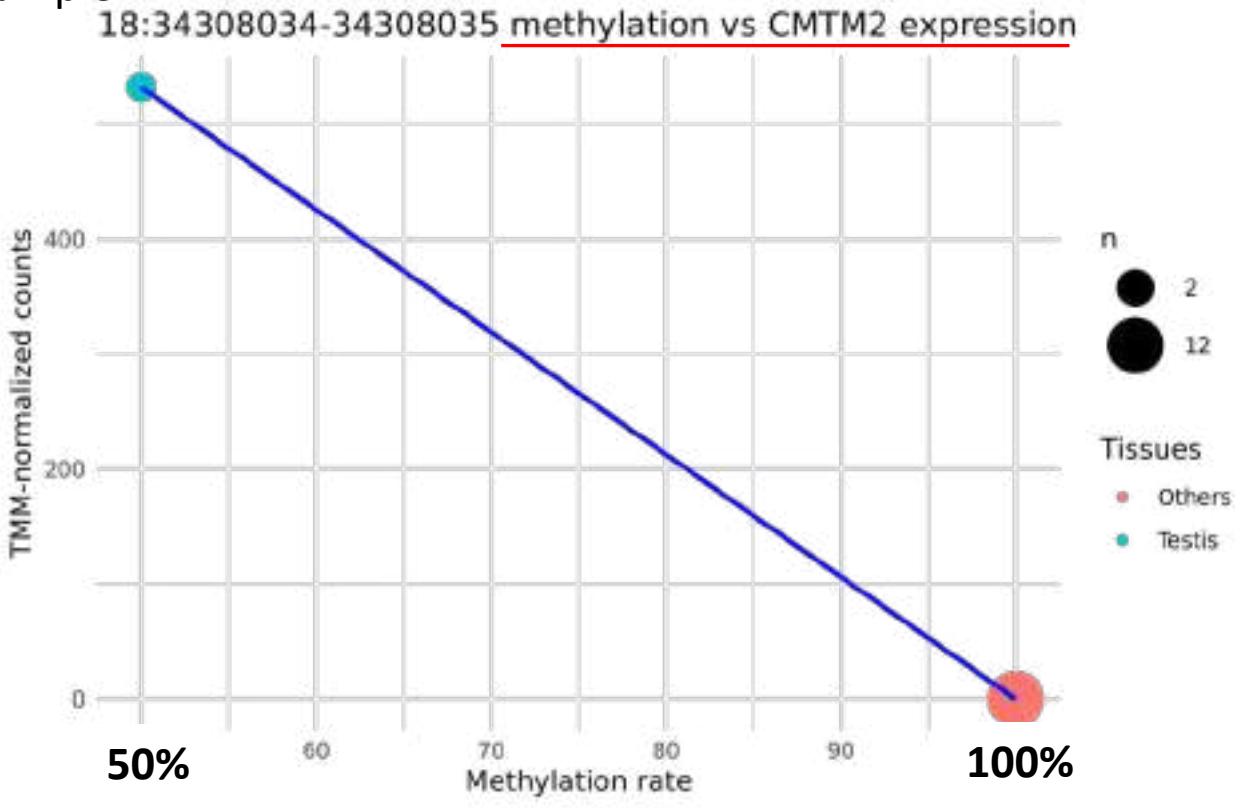


- Promoter-like states show lower methylation levels, confirming the well-known negative correlation between promoter methylation and gene expression

Differentially methylated regions (DMRs) across tissues- CMTM2 example across tissues

- Identified 208,665 differentially methylated regions (DMRs) across tissues, predicted to affect 1,080,550 motifs

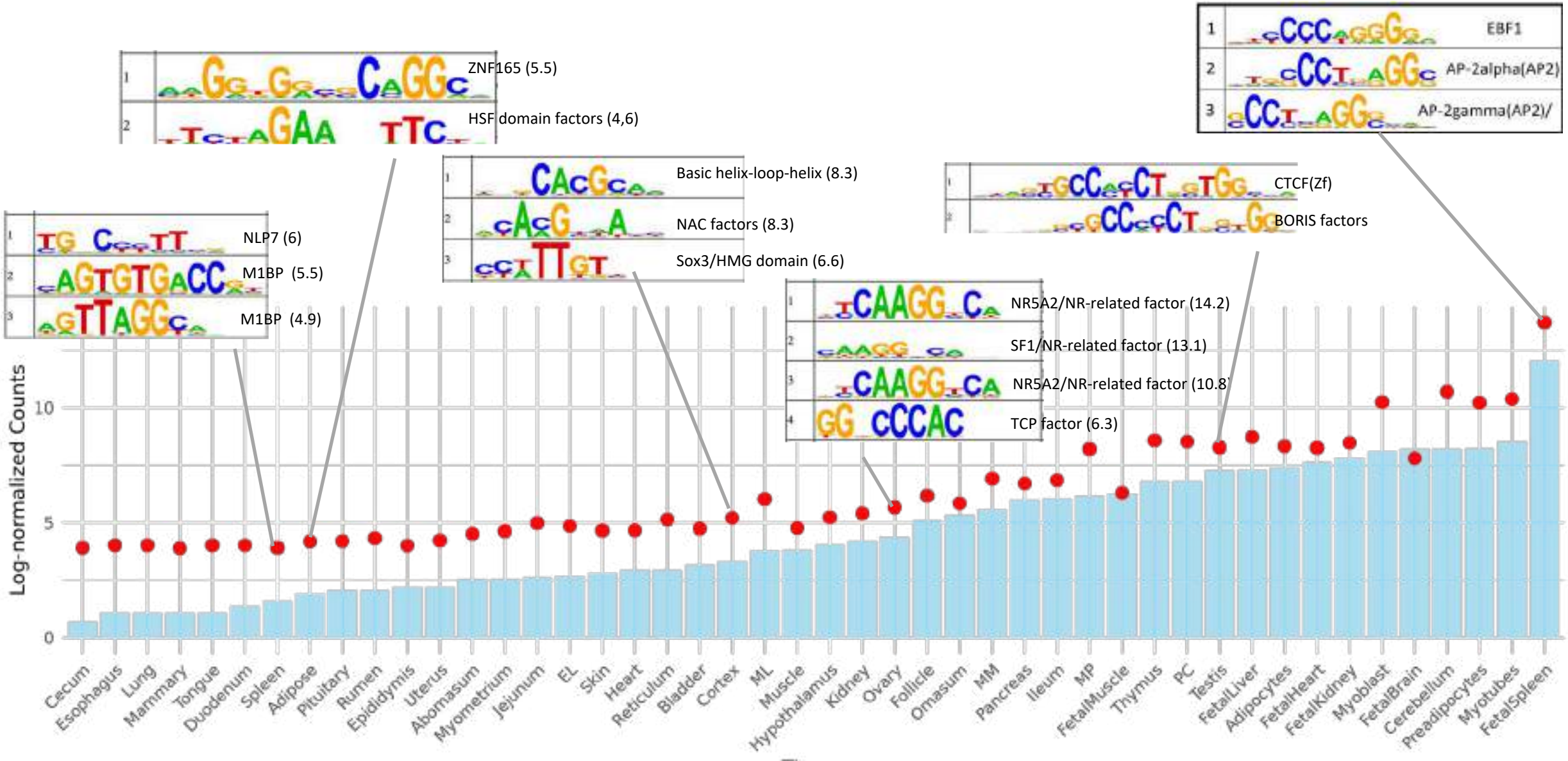
Example:



Two cytosines within a DMR for Testis correlated (**rho=-1**) with *CMTM2* → spermatogenesis and reproduction[1]

[1]10.1016/j.theriogenology.2019.07.026

Ongoing: evaluating the number of DMRs that may lead to motif disruption by promoter methylation



On-Going Research

- Allele-specific epigenomic regulation across tissues
- Characterization of regulatory elements (tissue specificity, switch/repurpose, super-enhancers/enhancer modules)
- Functional mechanism of regulatory elements in gene regulation (integrating with CattleGTEx)
- Application on complex traits/adaptive evolution (GWAS)
- Building open-access portals for data sharing

Future Considerations- What “needs” to be done next?

Needs:

- Regulatory Element Build for cattle
- Across breed characterization of regulatory elements
- Single Cell characterization/ annotation of regulatory markings
  - Integrated with bulk data
- More functional investigation: CRISPR screens with cell lines (other approached?)
- More investigation of chromatin conformation (impact of genotype and on molecular phenotypes)
- Outreach/ promoting increased use in the broader Animal Science disciplines
- Linking genotype to phenotype- application of regulatory/epigenetic data

Challenges:

- Need more information to link G2P (see above)
- Need support for more bioinformatics students and post-docs

Questions?

Thanks for your attention!

Contact:

jekoltes@iastate.edu

IOWA STATE UNIVERSITY
Department of Animal Science
Enriching lives through animals

Animal Breeding & Genetics

Project #s:
2018-67015-27500
2015-67015-22940