

Update: Identification of Regulatory Elements in Livestock Species

January 12, 2018

FAANG Workshop

H. Zhou¹, P. Ross¹, C. Kern¹, P. Saelao¹, Y. Wang¹, M. Halstead¹, K. Chanthavixay¹, I. Korf¹, M. Delany¹, H. Cheng², J. Medrano¹, A. Van Eenennaam¹, C. Tuggle³, and C. Ernst⁴, N. Trakooljul⁵, K. Wimmers⁵, Z. Jiang⁶ ¹University of California, Davis, Davis, CA, USA; ²USDA-ARS, Avian Disease and Oncology Laboratory, East Lansing, MI, USA; ³Iowa State University, Ames, IA, USA; ⁴Michigan State University, East Lansing, MI, USA; FBN, Institute for Genome Biology, Dummerstorf, Germany; Washington State University, Pullman, WA, USA



Overall Goal

Generate a comprehensive resource of functional regulatory elements for the chicken, cattle, and pig genomes

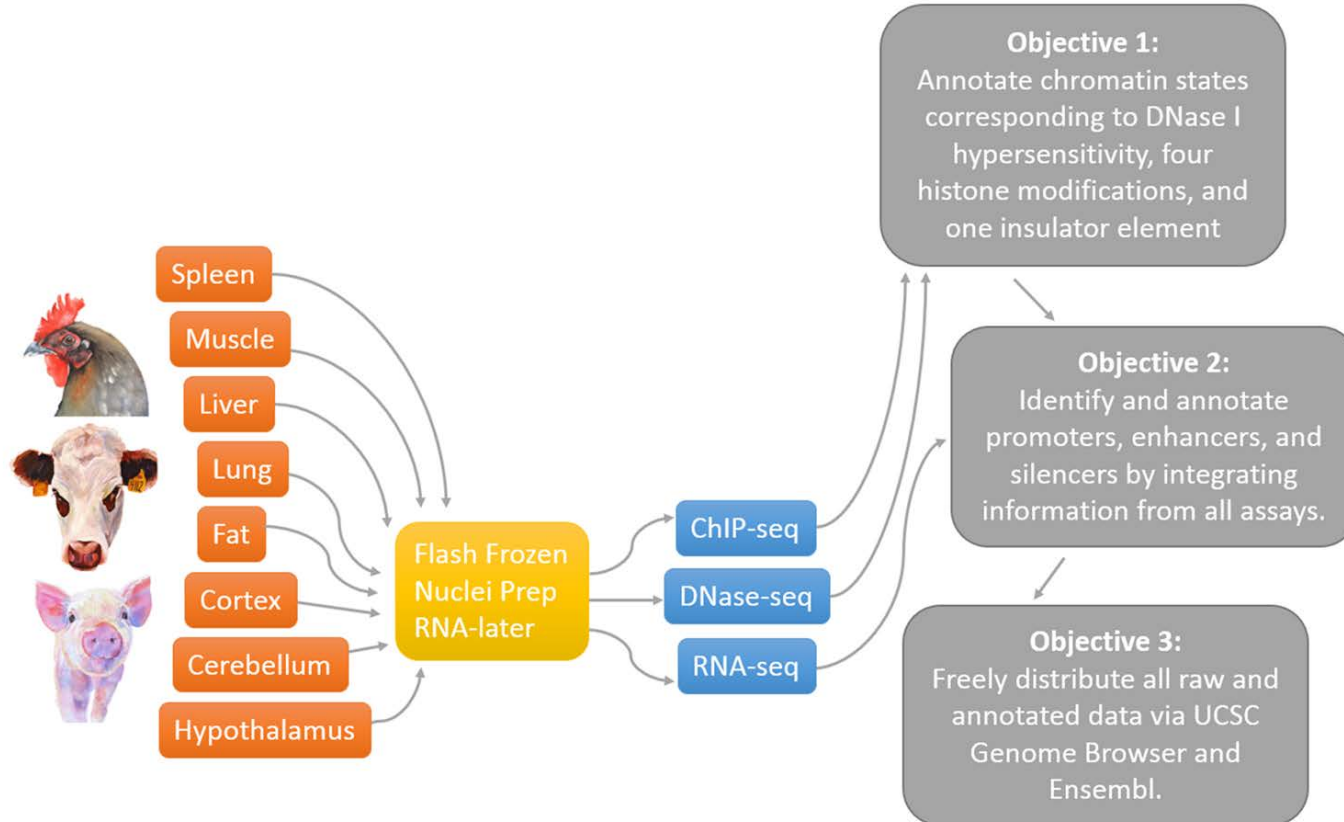
Livestock breeds

- Adult stages
 - 2 males and 2 females from same sire
 - 14 months of age
- Chicken
 - F1 (Line 6 x 7) from ADOL
 - 2 males and 2 females
 - 20 weeks of age
- Cattle
 - Line 1 Hereford from Fort Keogh Livestock and Range Research Lab
- Pig
 - Yorkshire from MSU
 - 2 male littermates (castrated)
 - 6 months of age

Tissues collected

- Wide range of tissues collected
- Total tissues collected:
 - Chicken: 29 tissues (92 samples: SAMEA4454482-4455404)
 - Cattle: 102 tissues (326 samples: SAMEA4454615-4455481)
 - Pig: 30 tissues (46 samples: SAMEA4454570-4454614)
- Scope of current project:
 - 2 male biological replicates from each species
 - 8 tissues: Adipose, Cerebellum, Cortex, Hypothalamus, Liver, Lung, Muscle, Spleen

Project overview



Assays

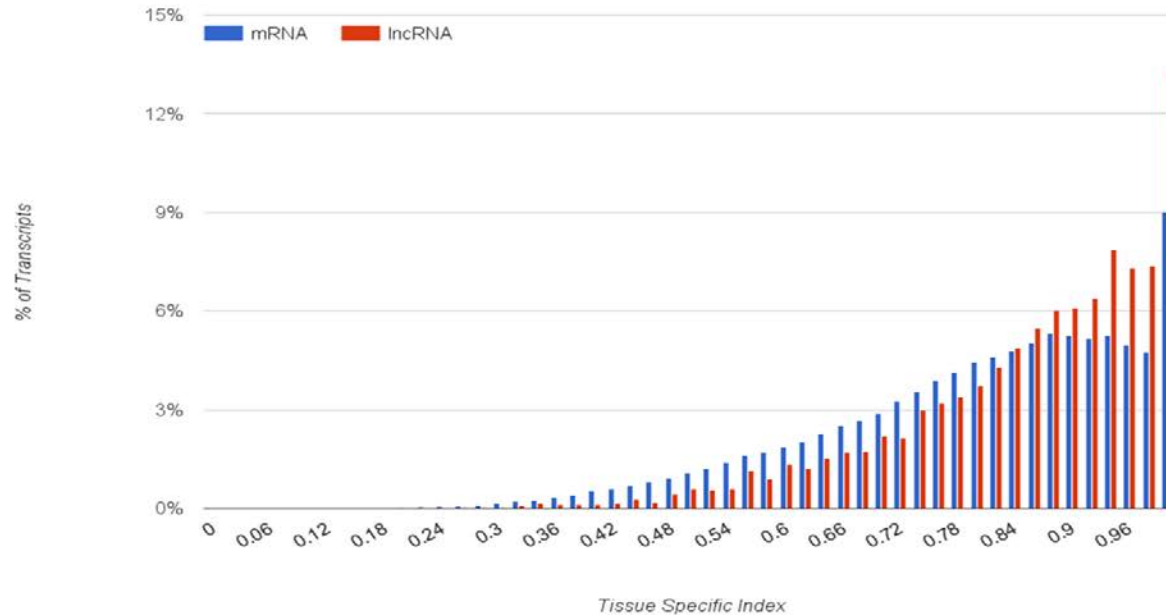
- Stranded RNA-seq
- Chromatin accessibility: DNase-seq or ATAC-seq
- Histone modification marks
 - H3K4me3 (promoters of active genes and transcription starts)
 - H3K27me3 (silence)
 - H3K27ac (active regulatory elements)
 - H3K4me1 (enhancers and other distal elements)
- Transcription factor CTCF (insulator)

Progress of UC Davis FAANG



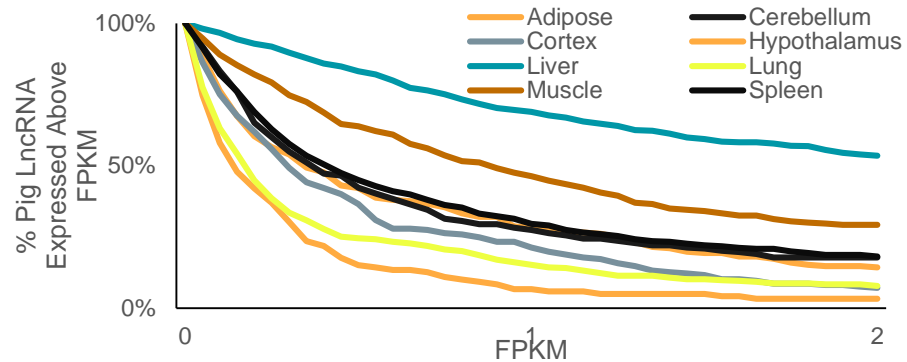
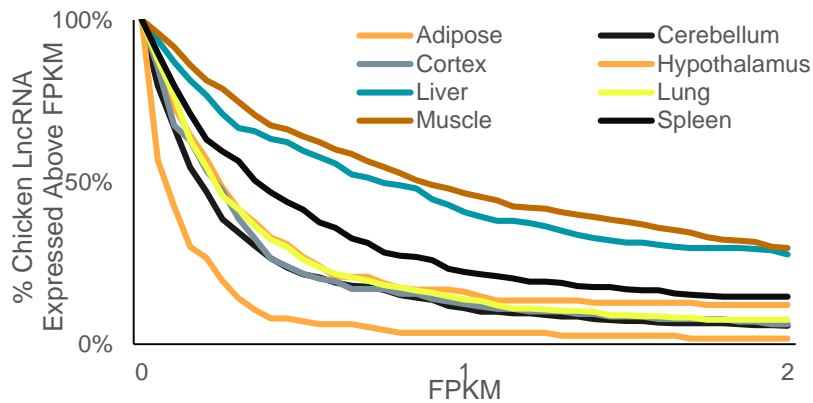
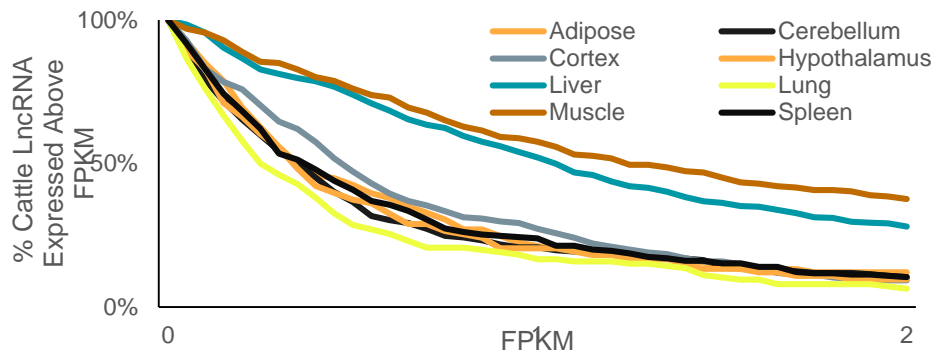
	RNA-seq	Data Analysis		Data Analysis		Data Analysis	
	DNase/ATAC-seq	Data Analysis		Data Analysis	Chromatin Sheared	Data Analysis	
	H3K4me3	Data Analysis		Data Analysis	Chromatin Sheared	Data Analysis	Chromatin Sheared
	H3K27me3	Data Analysis		Data Analysis	Chromatin Sheared	Data Analysis	Chromatin Sheared
	H3K4me1	Data Analysis	Chromatin Sheared	Data Analysis	Chromatin Sheared	Data Analysis	Chromatin Sheared
	H3K27ac	Data Analysis	Chromatin Sheared	Data Analysis	Chromatin Sheared	Data Analysis	Chromatin Sheared
	CTCF	Data Analysis	Chromatin Sheared	Data Analysis	Chromatin Sheared	Data Analysis	Chromatin Sheared
(FBN)	RRBS-seq	Data Analysis		Data Analysis		Tissues Collected	
(WSU)	WTTS-seq	Data Analysis		Tissues Collected		Tissues Collected	

Long ncRNAs are more tissue-specific than mRNA



Expression of tissue-specific lncRNAs

	Chicken	Cattle	Pig
Adipose	149	174	237
Cerebellum	389	202	258
Cortex	129	195	197
Hypothalamus	113	83	119
Liver	300	361	517
Lung	291	126	334
Muscle	273	226	246
Spleen	301	230	331



Enriched GO terms for genes associated with conserved lncRNAs across 3 species

Chicken	Pig	Cattle
chromatin silencing	chromatin assembly	chromatin assembly
negative regulation of gene expression, epigenetic	nucleosome assembly	chromatin assembly or disassembly
nucleosome assembly	chromatin assembly or disassembly	nucleosome assembly
chromatin assembly	DNA packaging	DNA packaging
chromatin assembly or disassembly	nucleosome organization	nucleosome organization
gene silencing	chromatin organization	chromatin organization
nucleosome organization	DNA conformation change	DNA conformation change
cell fate commitment	regulation of gene expression	protein-DNA complex assembly
embryonic skeletal system morphogenesis	chromatin silencing	protein-DNA complex subunit organization
regulation of production of molecular mediator of immune response	negative regulation of gene expression, epigenetic	innate immune response in mucosa

Enriched GO terms for genes associated with tissue-specific lncRNAs in chicken

Adipose	Cerebellum	Cortex	Hypothalamus	Liver	Lung	Muscle	Spleen
skeletal system development	nervous system development	nervous system development	cellular macromolecule biosynthetic process	chemotaxis	Notch signaling pathway	growth cellular	regulation of transcription, DNA-templated
ossification	regulation of developmental growth	regulation of cell differentiation	negative regulation of transcription from RNA polymerase II promoter	taxis	negative regulation of growth	macromolecule biosynthetic process	regulation of RNA metabolic process
gene expression	modulation of synaptic transmission	regulation of multicellular organismal development	negative regulation of cellular biosynthetic process	eye morphogenesis	regulation of glycoprotein biosynthetic process	developmental growth	regulation of nucleobase-containing compound metabolic process
regulation of gene expression	memory	neuron differentiation	organ morphogenesis	camera-type eye morphogenesis	response to metal ion	reproductive structure development	nucleobase-containing compound biosynthetic process
transcription, DNA-templated	neuron projection development	neuron projection development	negative regulation of biosynthetic process	regulation of leukocyte chemotaxis	regulation of glycoprotein metabolic process	reproductive system development	regulation of nitrogen compound metabolic process

Enriched GO terms for genes associated with tissue-specific lncRNAs in cattle

Adipose	Cerebellum	Cortex	Hypothalamu s	Liver	Lung	Muscle	Spleen
skeletal system morphogenesis	regionalization	response to endogenous stimulus	central nervous system development	monocarboxylic acid transport	pattern specification process	muscle tissue development	lymphocyte activation
skeletal system development	pattern specification process	response to organic substance	brain development	carboxylic acid metabolic process	embryonic morphogenesis	organ morphogenesis	regulation of immune system process
positive regulation of cell proliferation	spinal cord development	positive regulation of transferase activity	head development	organic acid metabolic process	embryonic organ morphogenesis	positive regulation of myoblast differentiation	immune system development
protein modification process	modulation of excitatory postsynaptic potential	regulation of signal transduction	nervous system development	oxoacid metabolic process	organ morphogenesis	striated muscle tissue development	T cell differentiation
cellular protein modification process	regulation of membrane potential	response to organonitrogen compound	cell development	monocarboxylic acid metabolic process	anatomical structure formation	skeletal muscle tissue development	single organism cell adhesion

Peaks called in chicken, pig and cattle

Chicken	H3K4me3	H3K27me3	H3K4me1	H3K27ac	CTCF	DHS
Liver	12,987	29,265	36,098	26,106	19,517	56,443
Lung	14,279	2,401	8,848	23,867	7,831	73,486
Spleen	18,089	965	12,639	21,961	12,379	51,615
Cereb.	23,057	28,379				88,401

Pig	H3K4me3	H3K27me3	H3K4me1	H3K27ac	CTCF	ATAC-seq
Liver	28,082	50,147	20,169	110,963	17,414	93,700
Lung	24,467	14,009	20,755	28,727	18,914	169,438
Spleen	23,504	25,770	29,163	57,422	28,545	117,246

Cattle	H3K4me3	H3K27me3	H3K4me1	H3K27ac	ATAC-seq
Liver		9,356	48,320	59,947	214,035
Lung	27,500	28,975	1,760	16,758	368,757
Spleen	24,057	4	6,090	37,694	358,115

Conservation of H3K4me3 and H3K4me1 peaks with human ENCODE

H3K4me3		<u>Peaks in conserved genomic regions</u>		<u>Matches ENCODE peak</u>	
Chicken	Liver	856	665	77.69%	
	Lung	547	289	52.83%	
	Spleen	378	321	84.92%	
Pig	Liver	8,533	5,902	69.17%	
	Lung	8,130	4,707	57.90%	
	Spleen	7,236	6,134	84.77%	

H3K4me1		<u>Peaks in conserved genomic regions</u>		<u>Matches ENCODE peak</u>	
Chicken	Lung	375	37	9.87%	
	Spleen	89	15	16.85%	
Pig	Lung	4,047	1,368	33.80%	
	Spleen	4,423	626	14.15%	

Identification of promoter, enhancer, and insulator elements

DNase-seq and ChIP-seq peak is categorized based on genomic location:
Promoter (2kb upstream of TSS), 5' UTR, 3' UTR, Exon, Intron, Intergenic.

Regulatory Element	Activation Marks					
	ATAC-seq	H3K4me3	H3K27me3	H3K27ac	H3K4me1	CTCF
Active Promoter	+	+	-	+	-	+/-
Inactive Promoter	+	-	+	-	+	+/-
Active Enhancer	+	+/-	+/-	+	+	+/-
Inactive Enhancer	+	+/-	+/-	-	+	+/-
Insulator	+	-	-	-	-	+

Chicken chromatin model

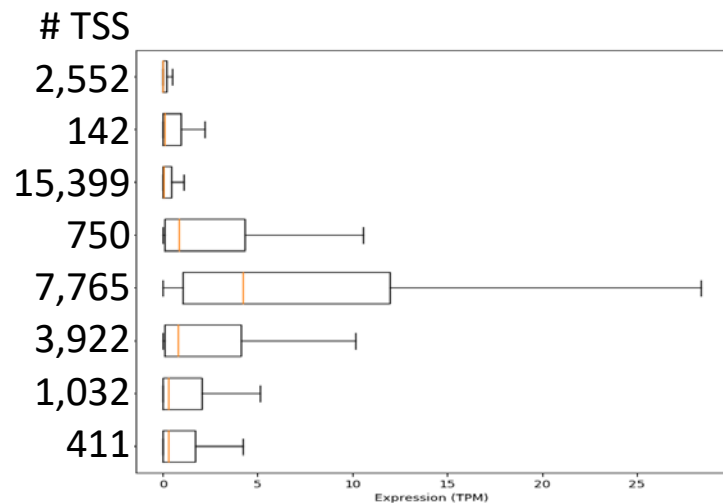
State	H3K4me3	H3K27ac	H3K4me1	H3K27me3	CTCF	Genome %	Exon	Gene	Promoter	TSS
1	1	1	11	41	4	2	2	1	5	4
2	1	0	0	3	0	0	1	1	1	1
3	0	0	0	0	0	83	1	1	1	1
4	2	46	8	0	4	1	3	1	3	3
5	95	65	40	7	20	1	4	1	10	18
6	0	32	89	0	6	9	1	1	3	2
7	0	0	0	0	1	3	1	1	2	1
8	0	0	0	0	1	1	1	1	2	2

Observation frequency
(%)

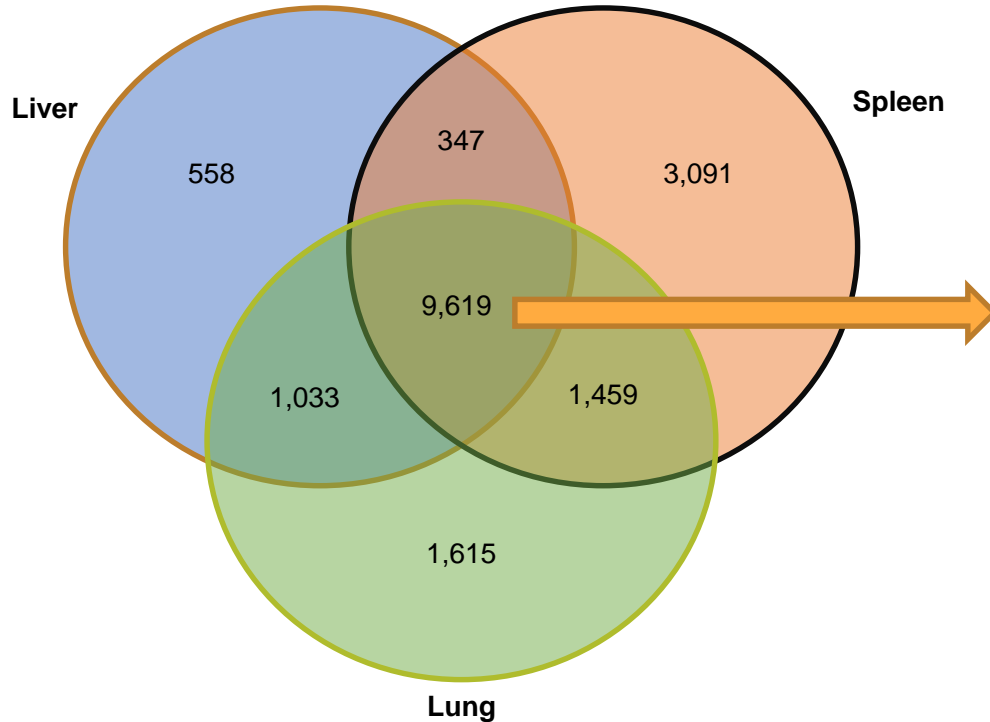
Feature Enrichment

Chicken chromatin model

State	Median Length (kb)	Candidate state annotation
1	1.6	Polycomb repressed
2	0.2	Heterochrom; low signal
3	6.8	Heterochrom; low signal
4	1.2	Enhancer
5	1.2	Active promoter
6	0.6	Enhancer
7	0.4	Heterochrom; low signal
8	0.2	Heterochrom; low signal



Promoters (State 5) in chickens

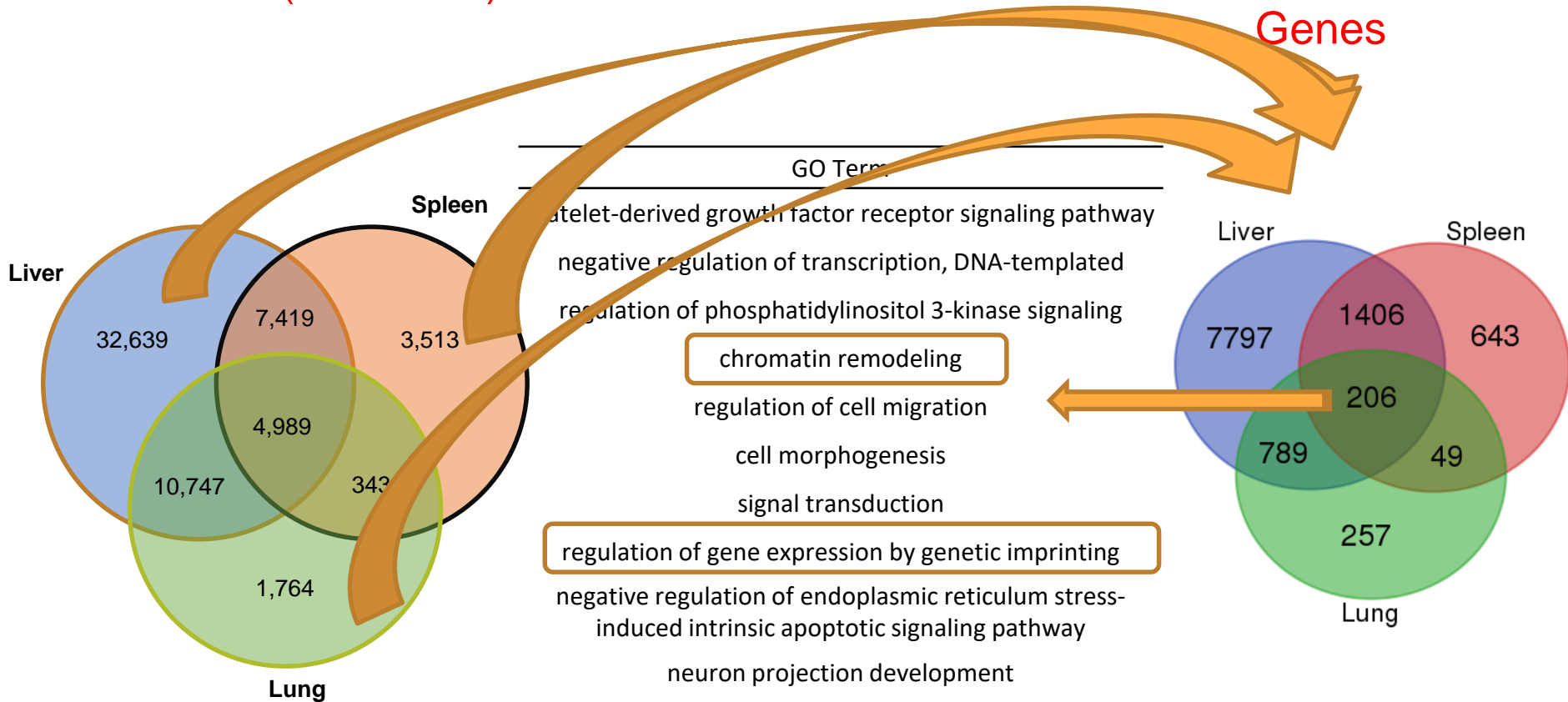


GO Term
translation
protein folding
mRNA processing
DNA replication
ER to Golgi vesicle-mediated transport
cell division
ER-associated ubiquitin-dependent protein catabolic process
mRNA splicing, via spliceosome
retrograde transport, endosome to Golgi
protein stabilization

Enriched GO terms for tissue-specific promoters

Liver	Lung	Spleen
biosynthetic process	protein ubiquitination involved in ubiquitin-dependent protein catabolic process	intracellular signal transduction
glutathione metabolic process	signal transduction	T cell receptor signaling pathway
response to glucose	protein deubiquitination	transmembrane receptor protein tyrosine kinase signaling pathway
insulin receptor signaling pathway	ubiquitin-dependent protein catabolic process	extrinsic apoptotic signaling pathway via death domain receptors
tryptophan catabolic process to kynurenine	transcription, DNA-templated	regulation of GTPase activity
anthranilate metabolic process	erythrocyte differentiation	multicellular organism development
quinolinate biosynthetic process	mRNA splicing, via spliceosome	positive regulation of filopodium assembly
reproductive structure development	protein K63-linked ubiquitination	negative thymic T cell selection
retinoid metabolic process	positive regulation of transcription from RNA polymerase II promoter	negative regulation of gene expression
'de novo' NAD biosynthetic process from tryptophan	positive regulation of nuclear-transcribed mRNA poly(A) tail shortening	positive regulation of transcription from RNA polymerase II promoter

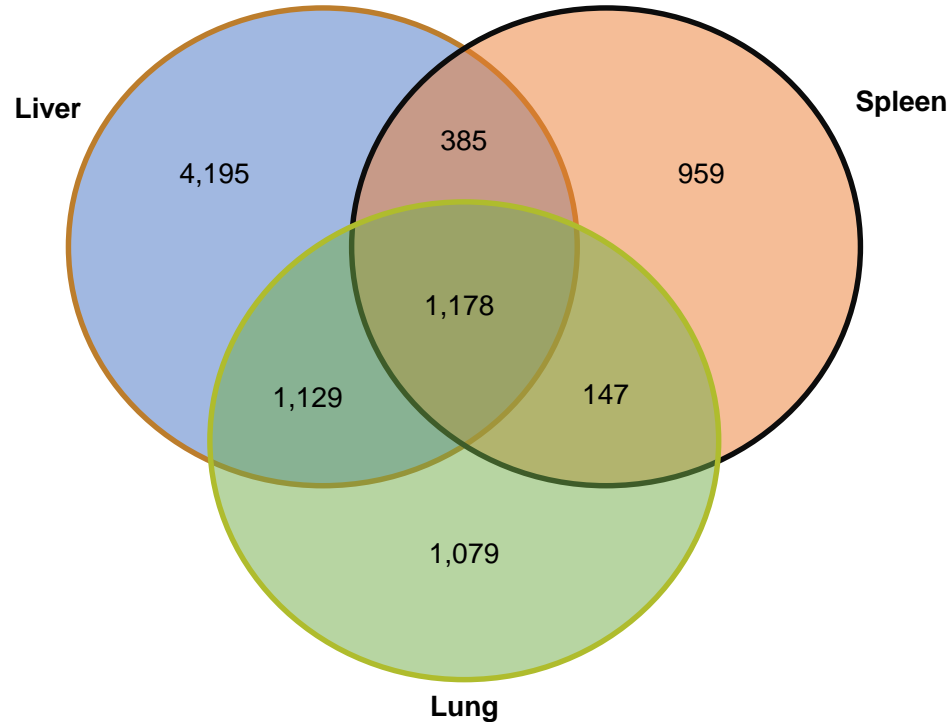
Enhancers (State 4+6) in chickens

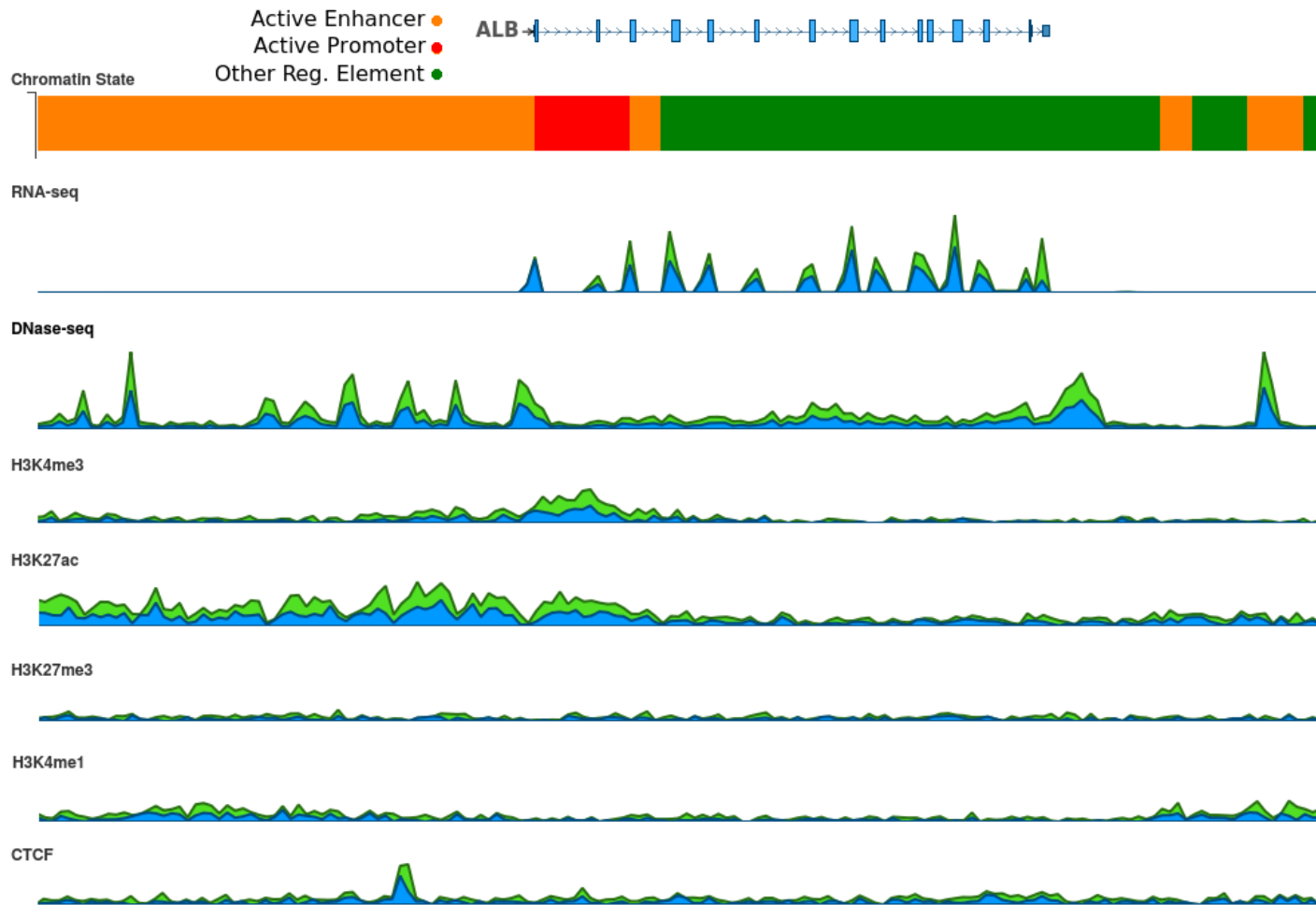


Enriched GO terms genes associated with tissue-specific enhancers from a single tissue

Liver	Lung	Spleen
intracellular protein transport	negative regulation of smooth muscle cell proliferation	immune response
metabolic process	positive regulation of transcription of nuclear large rRNA transcript from RNA polymerase I promoter	axonogenesis
protein folding	SRP-dependent cotranslational protein targeting to membrane	positive regulation of MAPK cascade
cellular response to amino acid starvation	anterior/posterior pattern specification	B cell receptor signaling pathway
histone acetylation	regulation of transcription from RNA polymerase II promoter	chemotaxis
social behavior	inner ear receptor stereocilium organization	positive regulation of Rho protein signal transduction
positive regulation of fibroblast proliferation	androgen receptor signaling pathway	MyD88-dependent toll-like receptor signaling pathway
regulation of cardiac muscle contraction	heart morphogenesis	cellular response to alkaloid
apoptotic cell clearance	negative regulation of protein catabolic process	heterophilic cell-cell adhesion via plasma membrane cell adhesion molecules
double-strand break repair via homologous recombination	protein folding	T cell differentiation

Polycomb repressed (State 1) in chickens





Whole transcriptome termini site sequencing in chickens

- Sequences 3' ends of transcripts
- Identifies alternative polyadenylation sites
- Expressed APS defined as ≥ 10 reads
- Total of 42,094 APSs (alternative polyadenylation sites) found across tissues

<u>Tissue</u>	<u>Expressed APSs</u>
Liver A	5,845
Liver B	3,986
Lung A	10,228
Lung B	4,853
Spleen A	8,024
Spleen B	5,557
Muscle A	6,905
Muscle B	4,420
Adipose A	5,779
Adipose B	9,982
Cortex A	15,483
Cortex B	11,186
Cerebellum A	13,511
Cerebellum B	12,124

WTTS-seq correlation with chromatin marks in chickens

	<u>Liver</u>	<u>Lung</u>	<u>Spleen</u>
DNase-seq	20.07%	36.89%	29.09%
H3K4me3	21.46%	44.13%	20.57%
H3K27me3	2.01%	17.51%	3.78%
H3K27ac	24.32%	44.69%	31.25%
H3K4me1	20.47%	39.07%	30.40%
CTCF	23.12%	49.15%	34.40%

Future work

- Finish sequencing for the rest of assays
- Integrative analysis with all assays by ChromHMM
- Distribute raw and annotated data via Ensembl and UCSC

Additional talks and posters

Talks:

Saturday 8:45am W166: Identification of regulatory regions in the swine genome

Tuesday afternoon: Attacking ATAC-seq in cryopreserved nuclei from livestock tissues

Posters:

P0509: Modification of ATAC-Seq Permits Profiling of Open Chromatin in Cryopreserved Chicken Lung

P0527: Genome-Wide Identification and Analysis of CTCF Binding Sites in chickens and pigs

P0460: Tissue Specific ChIP-Seq Analysis of Four Histone Modifications and an Insulator Element in Bovine Adult Male Tissue Samples

P0340: Optimization of Equine ChIP-Seq for the Functional Annotation of Animal Genomes

P0342: Update on the Equine FAANG Initiative: How the Community Is Using the Biobank

Acknowledgements

UC Davis: Dr. Gerald Quon

Stam Lab (U. Washington)



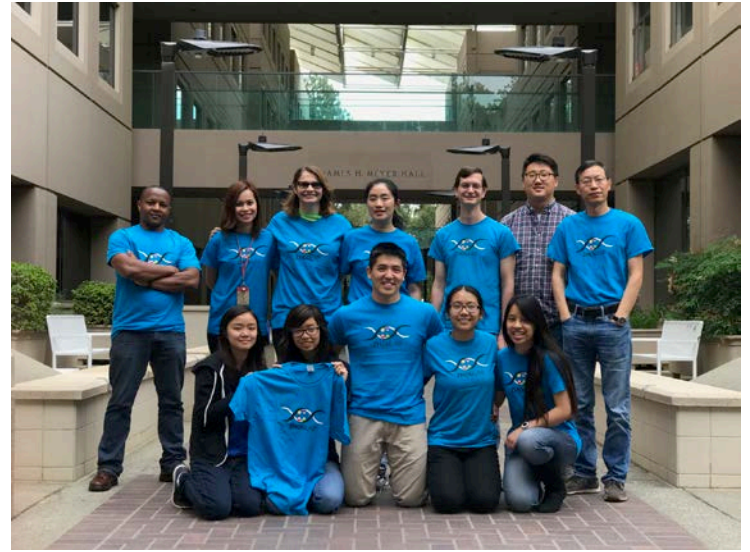
United States
Department of
Agriculture

National Institute
of Food and
Agriculture



This project is funded by Agriculture and Food Research Initiative Competitive Grant no. 2015-67015-22940 and National Need Fellowship Award #20143842021796 from the USDA National Institute of Food and Agriculture.

Further funding provided by the U.S. Poultry, Cattle, and Swine Genomes Coordination Funds



Zhou Lab 2017

V. Leesburg, USDA ARS

Jim Kent, UCSC

Laura Clarke, Paul Flicek, EBI

Bin Ren, UCSD

UC Davis FAANG Group

(chicken, pig, cattle and horse)

