Functional Annotation of Regulatory Elements in Livestock Species

H. Zhou[1], P. Ross[1], C. Kern[1], P. Saelao[1], Y. Wang[1], J. Chitwood[1], M. Halstead[1], I. Korf[2], M. Delany[1], H. Cheng[3], J. Medrano[1], A. Van Eenennaam[1], C. Tuggle[4], C. Ernst[5]

[1]Department of Animal Science, University of California, Davis, CA, USA
[2]Genome Center, University of California, Davis, CA, USA
[3]USDA-ARS, Avian Disease and Oncology Laboratory, East Lansing, MI, USA
[4]Department of Animal Science, Iowa State University, Ames, IA, USA
[5]Department of Animal Science, Michigan State University, East Lansing, MI, USA

USDA NIFA grant No. 2015-67015-22940

- H. Zhou (PD), P. Ross, I. Korf (co-PDs), UC Davis
- Collaborators:
  - Poultry Genome Coordinators: M. Delany, H. Cheng
  - Cattle Genome Coordinators: J. Medrano, A. Van Eenennaam
  - Swine Genome Coordinators: C. Tuggle, C. Ernst
  - V. Leesburg, USDA ARS
  - Jim Kent, UCSC
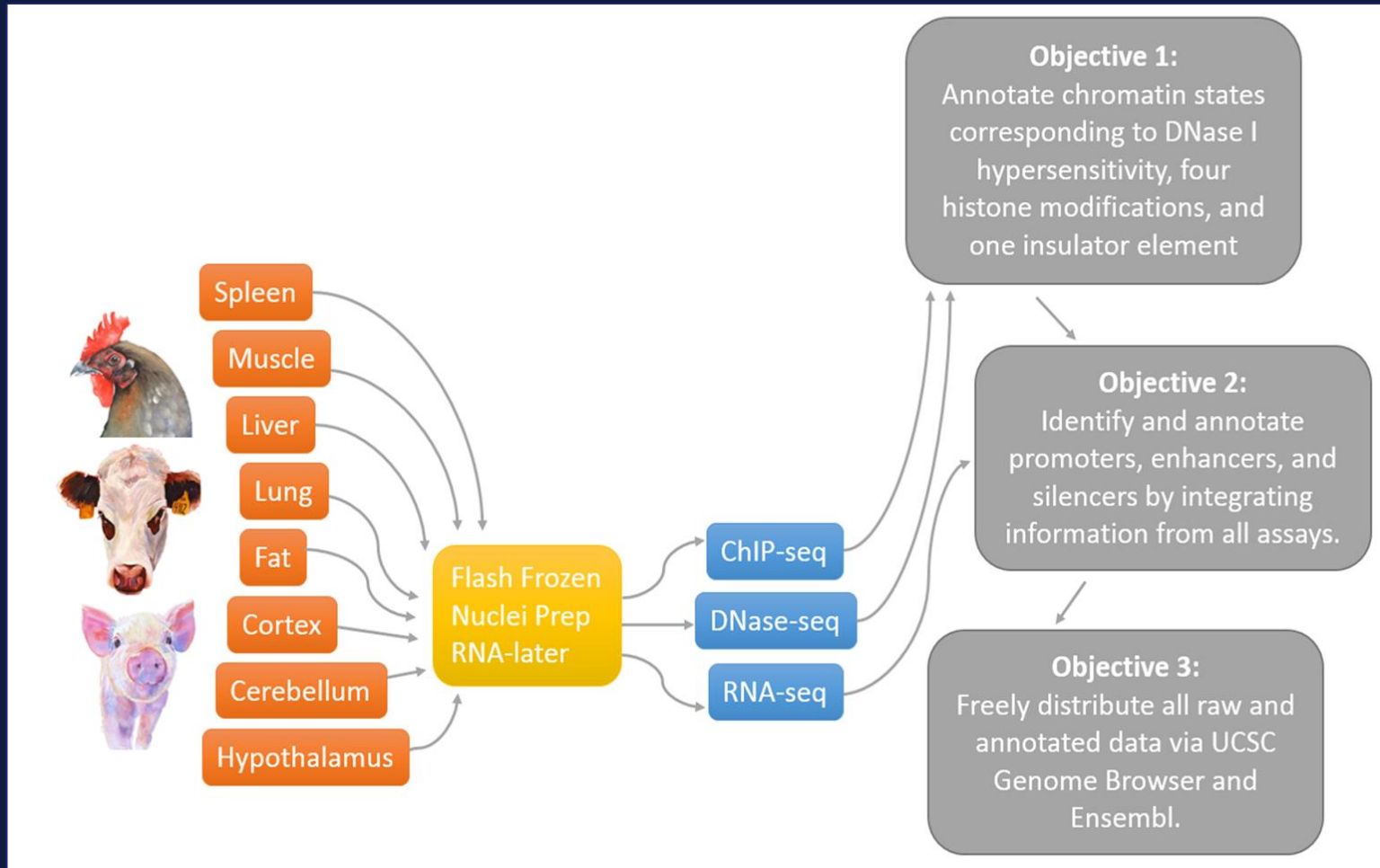 - Laura Clarke, Paul Flicek, EBI
  - Bin Ren, UCSD

Overall Goal

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

Project Overview

Genetic lines

- Adult stages
- Chickens:
  - 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,
  - 2 males and 2 females from the same sire
  - 14 months of age
- Pigs:
  - Yorkshire from MSU
  - 2 littermates males
  - 6 months of age

Tissues Collected

- Wide range of tissues collected
- Scope of current project:
  - 2 male biological replicates from each species
  - 8 tissues: Adipose, Cerebellum, Cortex, Hypothalamus, Liver, Lung, Muscle, Spleen.
- Total tissues collected:
  - Chicken: 29 tissues
  - Cattle: 102 tissues
  - Pig: 30 tissues

Assays

- Stranded RNA-seq
- Chromatin accessibility: DNase-seq (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)

Current Progress

	Chicken	Cow	Pig
RNA-Seq	Data Analysis	Data Analysis	Data Analysis
DNase-Seq	Data Analysis	Tissues Collected	Tissues Collected
H3K4me3	Data Analysis 6 / 8 Tissues — IP Completed 2 / 8	Tissues Collected	IP Completed 2 / 8 — Chromatin Sheared 6 / 8 Tissues
H3K27me3	Data Analysis 6 / 8 Tissues — IP Completed 2 / 8	Tissues Collected	IP Completed 2 / 8 — Chromatin Sheared 6 / 8 Tissues
H3K4me1	IP Completed	Tissues Collected	Tissues Collected
H3K27ac	IP Completed	Tissues Collected	Tissues Collected
CTCF	IP Completed	Tissues Collected	Tissues Collected

Aligned and Filtered Read Counts for RNA-seq

	Chicken	Cattle	Pig
Adipose	198,929,564	156,656,620	119,721,691
Cerebellum	242,807,223	246,658,282	152,762,359
Cortex	236,147,593	119,721,576	126,240,107
Hypothalamus	244,215,661	142,709,163	132,786,659
Liver	244,674,805	119,617,850	104,210,750
Lung	205,055,604	138,746,254	198,053,139
Muscle	238,435,618	140,106,635	155,724,909
Spleen	201,084,991	150,804,156	125,682,422

Aligned and Filtered ChIP-Seq Reads from Chicken

Tissues	H3K4me3	H3K27me3	Control
Cerebellum	24,331,529	26,132,801	27,935,330
Cortex	35,745,439	20,781,566	56,091,853
Hypothalamus	42,025,119	48,644,786	60,546,445
Liver	29,485,908	35,850,744	51,835,760
Lung	33,971,248	30,219,806	51,907,342
Spleen	28,553,718	37,290,417	44,115,980

DNase-seq Status for Chicken

	Adipose	Cerebellum	Cortex	Hypothalamus	Liver	Lung	Muscle	Spleen
Bird A								
Bird B								
Reads	29,236,452	107,520,255	42,946,606	59,644,084	93,681,598	79,033,732	26,730,117	237,834,942

# Aligned and Filtered DNase-Seq Reads from Chicken

	Galgal4	Galgal5
Adipose	29,236,452	27,464,894
Cerebellum	107,520,255	105,329,616
Cortex	42,946,606	39,999,573
Hypothalamus	59,644,084	56,589,035
Liver	93,681,598	88,742,141
Lung	79,033,732	77,450,516
Muscle	26,730,117	24,635,982
Spleen	237,834,942	221,475,988

Alignment Rates to Chicken Genome

	RNA-Seq	DNase-Seq	H3K4me3	H3K27me3
Adipose	92.92%	64.85%		
Cerebellum	92.60%	72.55%	79.00%	81.00%
Cortex	91.98%	43.26%	89.00%	70.00%
Hypothalamus	92.79%	69.42%	83.00%	88.00%
Liver	96.50%	78.97%	78.00%	70.00%
Lung	94.62%	78.92%	90.00%	89.00%
Muscle	91.49%	59.46%		
Spleen	93.38%	76.54%	89.00%	90.00%

Distribution of DHS peaks identified in 8 tissues in the chicken
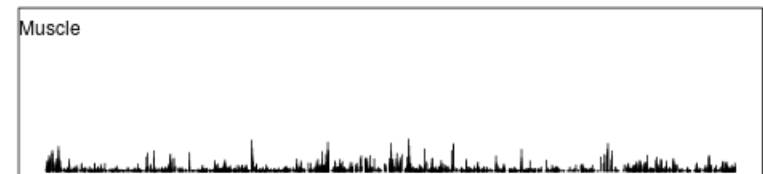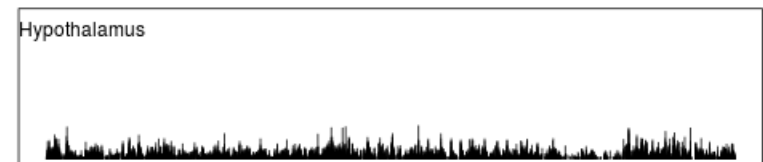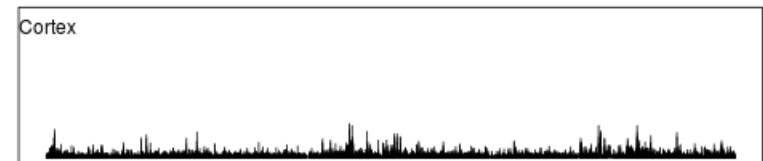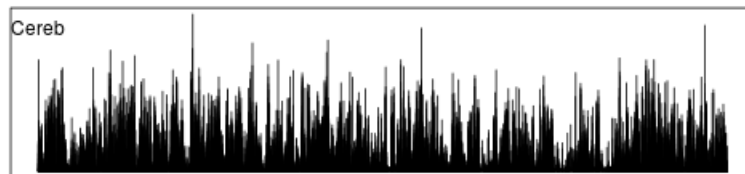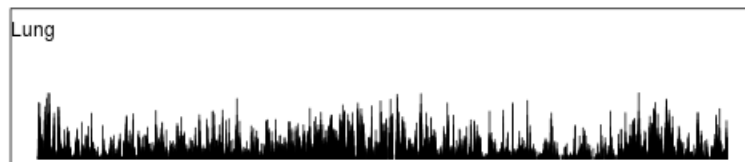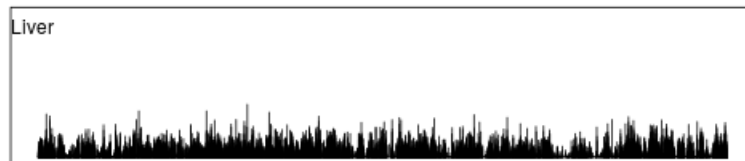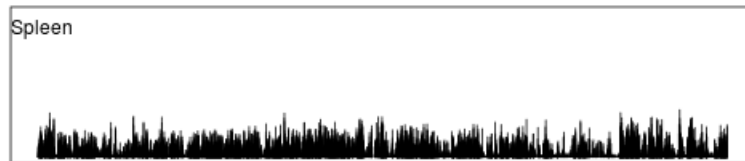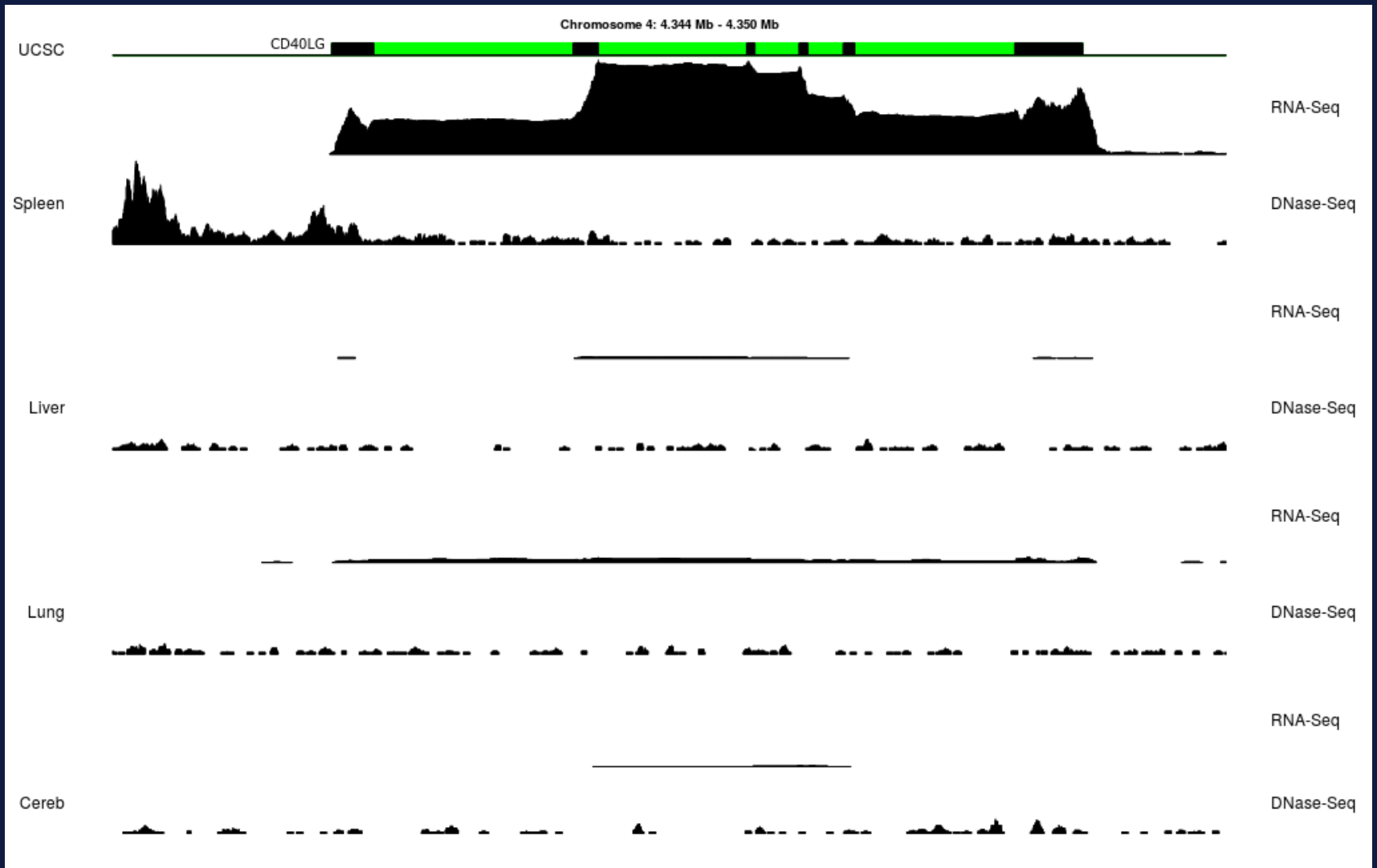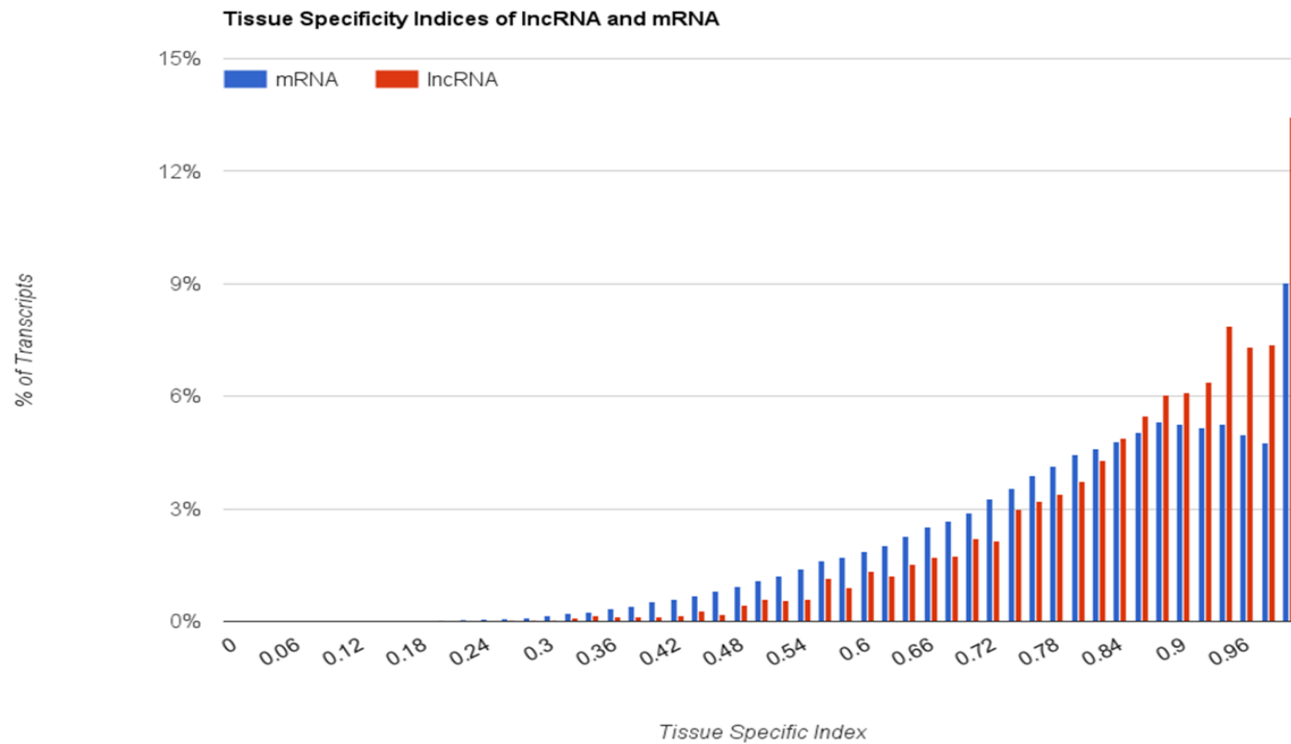
Tissue-Specific DHS Peaks

- For each tissue
  - Create list of genes with a DHS peak in promoter or gene region.
  - Remove genes that appear in more than one tissue
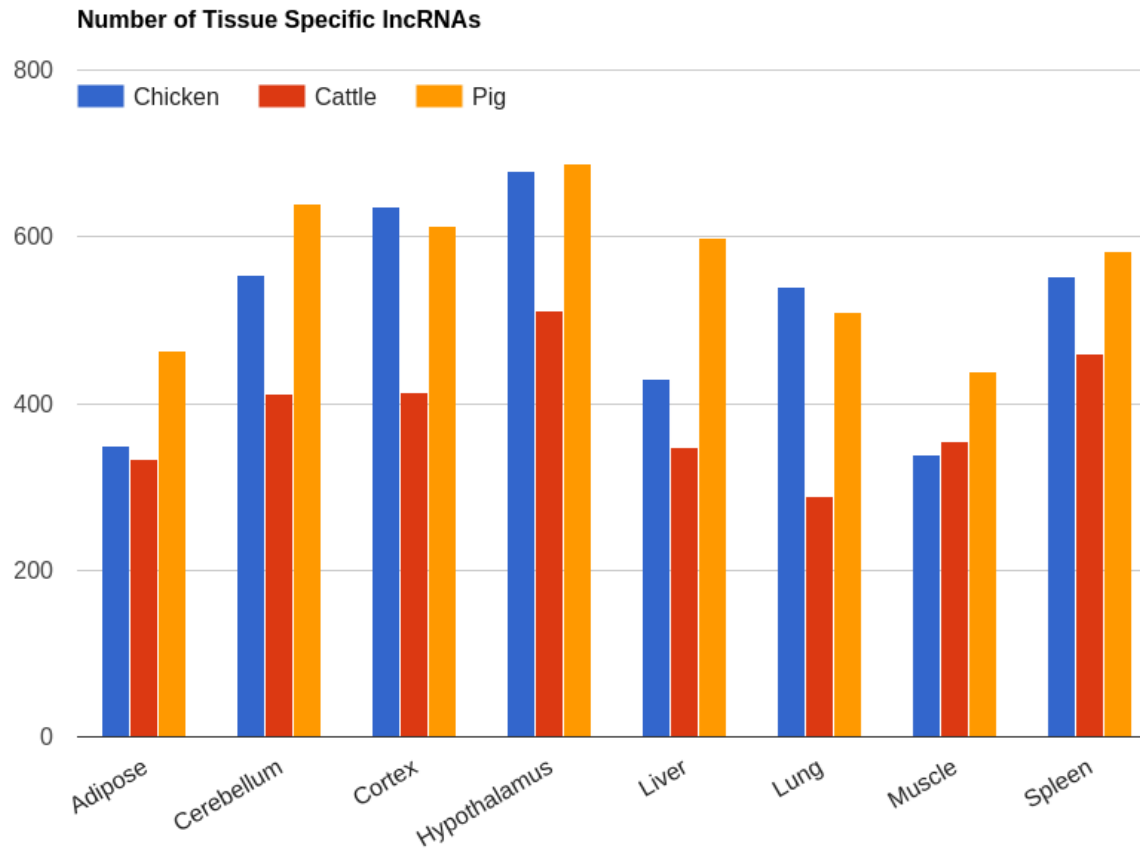 - Combine with RNA-Seq expression

Spleen-Specific DHS Peaks

Long ncRNAs Are More Tissue-Specific than mRNA

# Number of Tissue-specific lncRNAs in chicken, cattle and pig

Number of Orthologous Long ncRNAs

	Chicken	Cattle	Pig	Human	Mouse
Chicken	7364				
Cattle	10	5918			
Pig	7	193	8439		
Human	18	175	181	27817	
Mouse	7	52	51	348	12169

Identification of Promoter, Enhancer, and Insulator Elements

- Every DNase-seq and ChIP-seq peak is categorized based on genomic location: Promoter (2kb upstream of TSS), 5' UTR, 3' UTR, Exon, Intron, Intergenic.
- The following elements are then defined based on the previous categorization:
  - Active Promoter: DNase-seq site and H3K4me3 peak in promoter or 5' region.
  - Inactive Promoter: DNase-seq site and H3K27me3 peak in promoter or 5' region.
  - Enhancer: Intergenic DNase-seq sites co-localizing with H3K4me1 and H3K27ac.
  - Potential Insulator: CTCF and DNase-seq sites in promoter or 5' region.

Integrated RNA-seq, DNase-seq, and ChIP-seq assays

Assay Read Depth in Chicken Lung
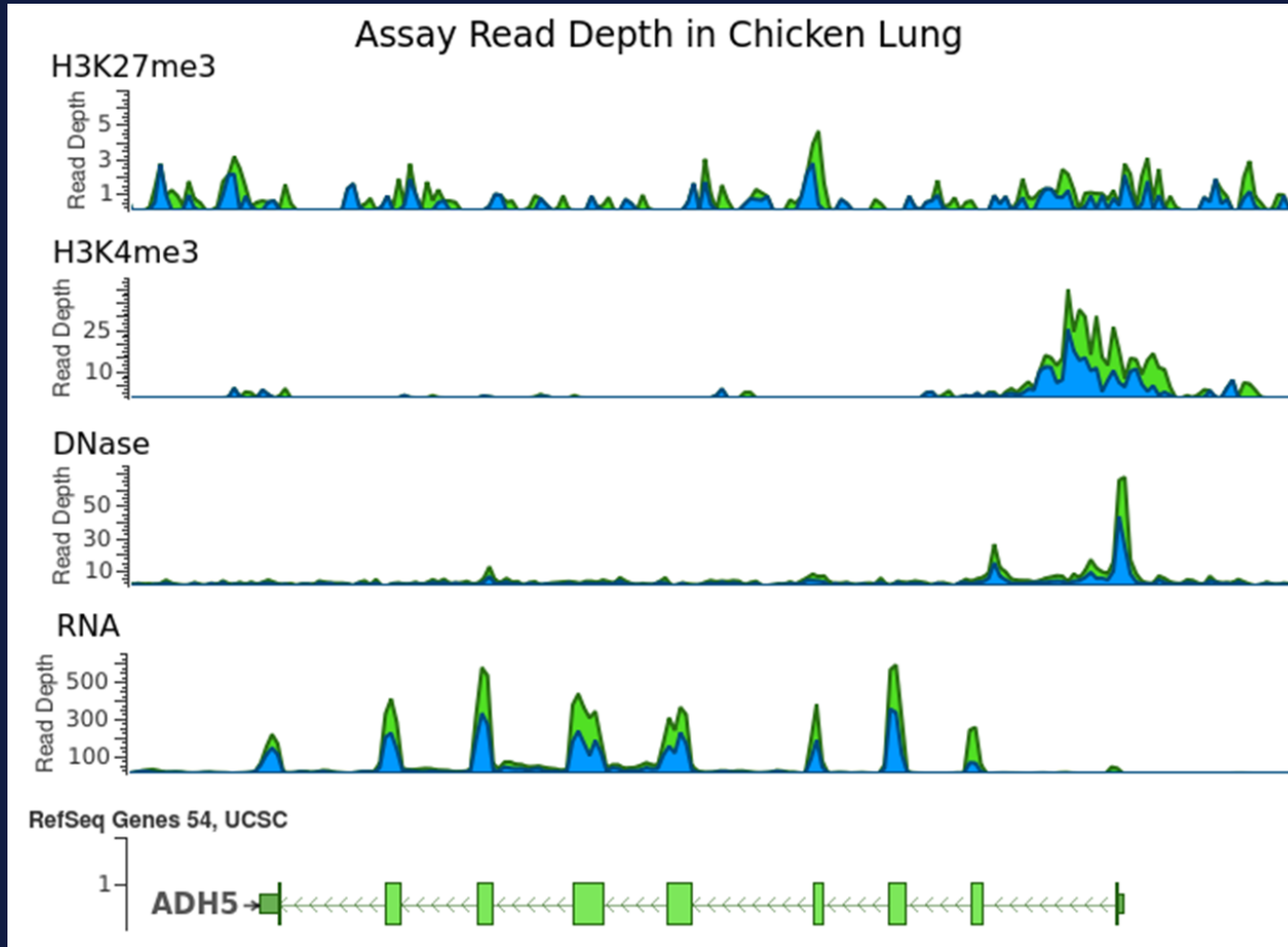
Future plan

- Continue DNase-seq and ChIP-seq assays

- Integrate Assays to Determine Functional Roles

- Develop an integrative HMM (Hidden Markov Model) for identifying regulatory regions.

- Look for functional relationships using a Gene Set Enrichment Analysis (GSEA).

Financial support

USDA NIFA grant #2015-67015-22940;
2016-67015-24579

Poultry, Cattle, Swine Genome Coordination Funds