Update: Identification of Regulatory Elements in Livestock Species

January 11, 2019

FAANG Workshop

H. Zhou[1], P. Ross[1], C. Kern[1], P. Saelao[1], Y. Wang[1], M. Halstead[1], K. Chanthavixay[1], X. Xu[1], I. Korf[1], M. Delany[1], H. Cheng[2], J. Medrano[1], A. Van Eenennaam[1], C. Tuggle[3], and C. Ernst[4], N. Trakooljul[5], K. Wimmers[5], Z. Jiang[6] [1]University of California, Davis, Davis, CA, USA; [2]USDA-ARS, Avian Disease and Oncology Laboratory, East Lansing, MI, USA; [3]Iowa State University, Ames, IA, USA; [4]Michigan 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

➢ 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

  ➢ 2 males and 2 females from same sire

  ➢ 14 months of age

➢ 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)
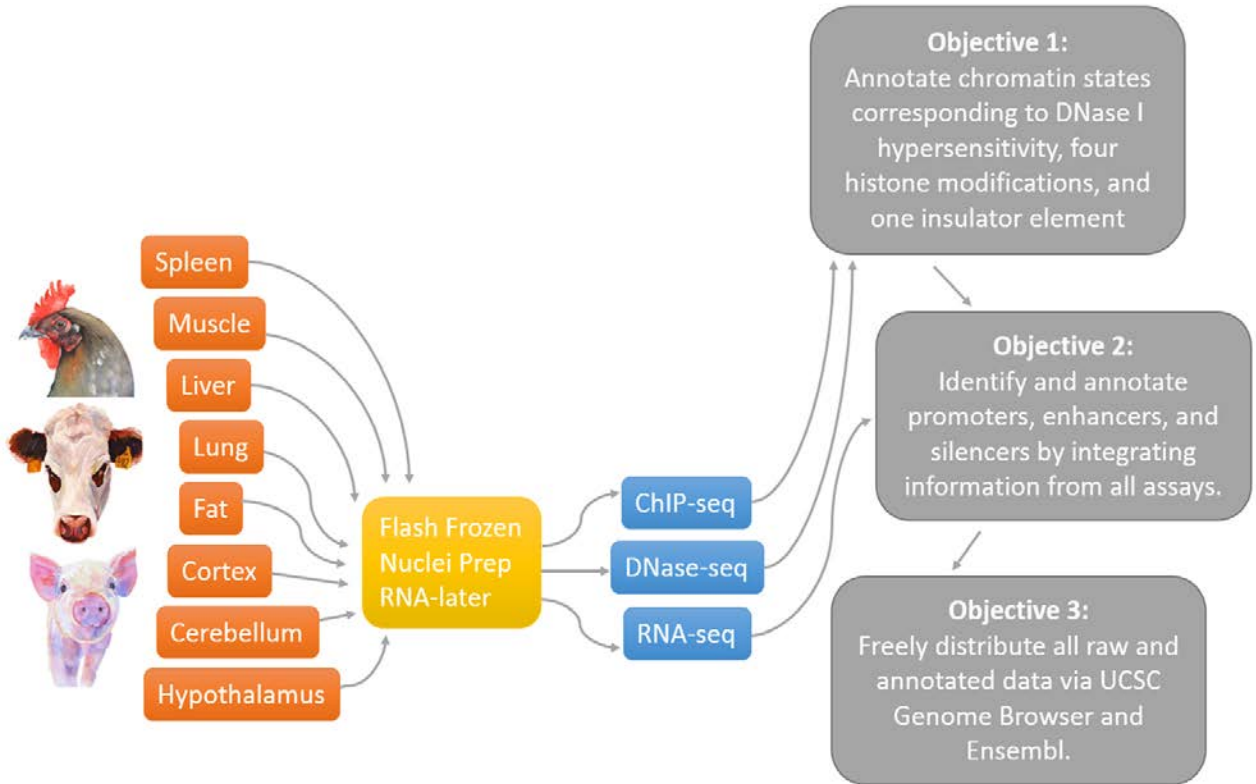
  ➢ Pig: 30 tissues (46 sampes:SAMEA4454570-4454614)

➢ 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)

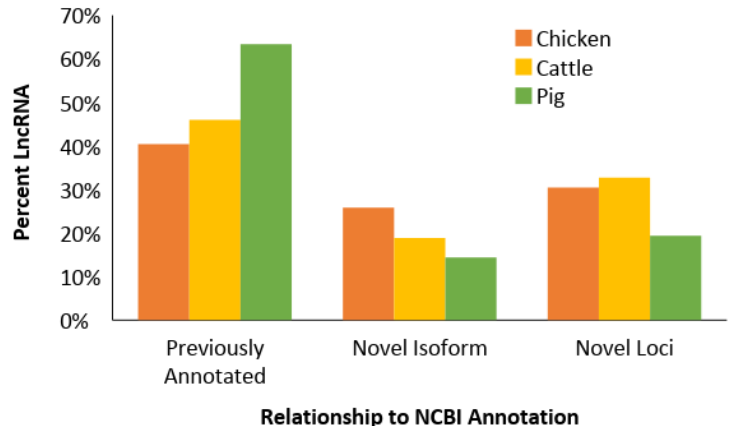➢ Transcription factor CTCF (insulator)

Current status

	Chicken		Pig		Cow
RNA-seq	Data Analysis		Data Analysis		Data Analysis
DNase/ATAC-seq	Data Analysis		Data Analysis		Data Analysis
H3K4me3	Data Analysis		Data Analysis	Chromatin Sheared	Data Analysis
H3K27me3	Data Analysis		Data Analysis	Chromatin Sheared	Data Analysis
H3K4me1	Data Analysis	Chromatin Sheared	Data Analysis	Chromatin Sheared	Data Analysis
H3K27ac	Data Analysis	Chromatin Sheared	Data Analysis	Chromatin Sheared	Data Analysis
CTCF	Data Analysis	Chromatin Sheared	Data Analysis	Chromatin Sheared	Data Analysis
RRBS-seq	Data Analysis		Data Analysis		Data Analysis
WTTS-seq	Data Analysis		Tissues Collected		Tissues Collected

LncRNA paper published

Kern et al. "Genome-wide identification of tissue-specific long non-coding RNA in three farm animal species." BMC Genomics (2018) 19:684

Species	LncRNA Identified
Chicken	9,393
Cattle	7,235
Pig	14,428

Peak calls on ChIP-seqs

H3K4me3	Chicken	Cattle	Pig
Adipose		21,534	
Cerebellum	24,980	30,708	
Cortex	22,966	25,201	
Hypothalamus		31,555	
Liver	20,881	30,554	31,614
Lung	16,099	34,923	27,696
Muscle	15,106	30,004	
Spleen	21,385	34,462	26,538

H3K27ac	Chicken	Cattle	Pig
Adipose		26,717	
Cerebellum	69,541	28,079	
Cortex	61,430	43,971	
Hypothalamus		29,781	
Liver	27,601	85,069	55,513
Lung	30,701	53,851	71,055
Muscle	15,909	61,325	
Spleen	33,174	52,154	68,111

H3K4me1	Chicken	Cattle	Pig
Adipose			
Cerebellum	56,046	23,960	
Cortex	53,459		
Hypothalamus		17,845	
Liver	37,289	63,211	47,684
Lung	14,477	58,986	68,057
Muscle	30,821	28,143	
Spleen	17,940	72,679	138,023

H3K27me3	Chicken	Cattle	Pig
Adipose		19,519	
Cerebellum	52,337	44,292	
Cortex	33,315	11,907	
Hypothalamus		53,328	
Liver	48,600	28,194	69,015
Lung	28,502	113,059	18,988
Muscle		31,277	
Spleen	26,701	50,579	36,578

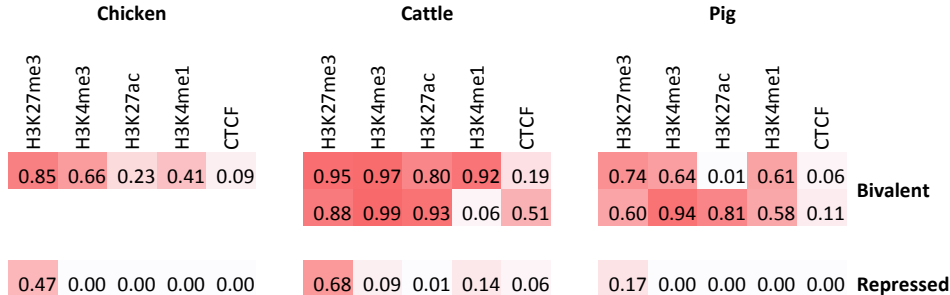| CTCF	Chicken	Cattle	Pig
Adipose			
Cerebellum		51,254	
Cortex	26,450	29,526	
Hypothalamus		13,187	
Liver	21,880	58,705	21,485
Lung	24,745	44,796	23,193
Muscle	12,203	30,909	
Spleen	42,969	11,580	35,082

ChromHMM models for chromatin state prediction

➢ Three models were created, one for each species, using all data passing quality metrics for that species.

➢ A common model was learned across tissues, then used to produce distinct chromatin state predictions for each tissue.

➢ All three models were created with 14 states for consistency between species.

Active promoter states

Chicken

H3K27me3	H3K4me3	H3K27ac	H3K4me1	CTCF
0.02	0.98	0.97	0.06	0.00
0.01	0.97	1.00	0.97	0.02
0.01	0.76	0.11	0.11	0.03
0.01	0.89	0.13	0.91	0.01
0.01	0.98	0.88	0.02	0.89
0.04	0.99	0.79	0.90	0.92

Cattle

H3K27me3	H3K4me3	H3K27ac	H3K4me1	CTCF
0.01	0.99	0.97	0.07	0.20
0.01	0.95	0.82	0.93	0.16
0.08	0.83	0.11	0.24	0.05

Pig

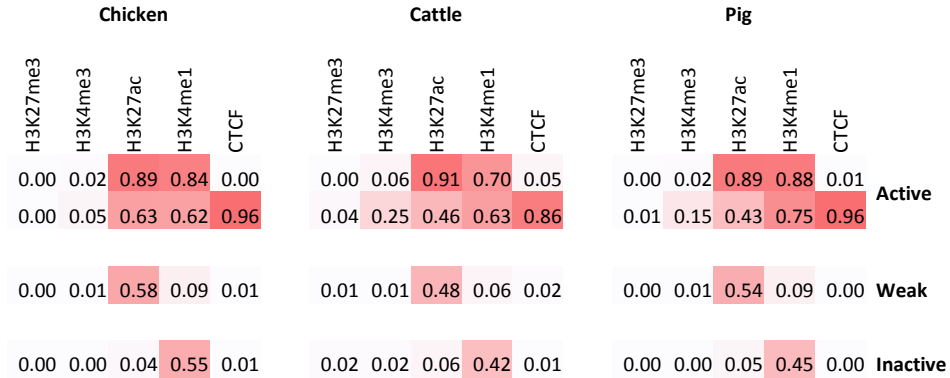| H3K27me3	H3K4me3	H3K27ac	H3K4me1	CTCF
0.00	0.98	0.96	0.02	0.02
0.01	0.98	1.00	0.97	0.03
0.02	0.80	0.09	0.18	0.03
0.01	0.80	0.12	0.88	0.02
0.00	0.99	0.93	0.28	0.94

➢ Active promoters are defined by presence of H3K4me3

➢ CTCF only appears in active promoters with H3K27ac

Bivalent promoter states

	Chicken						Cattle						Pig					
	H3K27me3	H3K4me3	H3K27ac	H3K4me1	CTCF		H3K27me3	H3K4me3	H3K27ac	H3K4me1	CTCF		H3K27me3	H3K4me3	H3K27ac	H3K4me1	CTCF	
	0.85	0.66	0.23	0.41	0.09		0.95	0.97	0.80	0.92	0.19		0.74	0.64	0.01	0.61	0.06	Bivalent
							0.88	0.99	0.93	0.06	0.51		0.60	0.94	0.81	0.58	0.11	
	0.47	0.00	0.00	0.00	0.00		0.68	0.09	0.01	0.14	0.06		0.17	0.00	0.00	0.00	0.00	Repressed

➢ Only one bivalent state in chicken

➢ Cattle bivalent states are distinguished by presence of H3K4me1 or CTCF

➢ Pig bivalent states are distinguished by presence of H3K27ac

➢ Pig repressed state has weak H3K27me3 probability

Enhancer states

	Chicken						Cattle						Pig					
	H3K27me3	H3K4me3	H3K27ac	H3K4me1	CTCF		H3K27me3	H3K4me3	H3K27ac	H3K4me1	CTCF		H3K27me3	H3K4me3	H3K27ac	H3K4me1	CTCF	
	0.00	0.02	0.89	0.84	0.00		0.00	0.06	0.91	0.70	0.05		0.00	0.02	0.89	0.88	0.01	Active
	0.00	0.05	0.63	0.62	0.96		0.04	0.25	0.46	0.63	0.86		0.01	0.15	0.43	0.75	0.96	
	0.00	0.01	0.58	0.09	0.01		0.01	0.01	0.48	0.06	0.02		0.00	0.01	0.54	0.09	0.00	Weak
	0.00	0.00	0.04	0.55	0.01		0.02	0.02	0.06	0.42	0.01		0.00	0.00	0.05	0.45	0.00	Inactive

➤ Active enhancers have H3K27ac and H3K4me1, may or may not have CTCF

➤ Weak (lower confidence) enhancers have H3K27ac only

➤ H3K4me1 without H3K27ac indicates inactive enhancers

Insulator and quiescent states

	Chicken						Cattle						Pig					
	H3K27me3	H3K4me3	H3K27ac	H3K4me1	CTCF		H3K27me3	H3K4me3	H3K27ac	H3K4me1	CTCF		H3K27me3	H3K4me3	H3K27ac	H3K4me1	CTCF	
	0.03	0.02	0.02	0.08	0.84		0.02	0.01	0.01	0.03	0.78		0.01	0.00	0.00	0.05	0.89	**Insulator**
	0.01	0.00	0.00	0.00	0.00		0.00	0.00	0.00	0.01	0.00		0.00	0.00	0.00	0.00	0.00	**Quiescent**
							0.00	0.00	0.00	0.00	0.00							
							0.06	0.00	0.00	0.00	0.00							

➤ Three quiescent states in cattle have distinct transition probabilities to other states

➤ Multiple quiescent states may improve prediction of other states, as ChromHMM predictions are influenced by nearby genomic regions.

Predicted promoter states at TSS

➢ In all three tissues, more TSS in pig are in an active state compared to chicken and cattle.

➢ Most TSS are active or quiescent.

		Chicken	Cattle	Pig
Liver	Active	38%	39%	49%
	Bivalent	7%	1%	6%
	Repressed	7%	3%	5%
	Quiescent	42%	48%	36%
	Other	6%	9%	4%
Lung	Active	34%	40%	45%
	Bivalent	5%	4%	9%
	Repressed	6%	4%	4%
	Quiescent	51%	47%	34%
	Other	4%	5%	8%
Spleen	Active	37%	38%	46%
	Bivalent	4%	4%	7%
	Repressed	2%	4%	3%
	Quiescent	52%	49%	35%
	Other	5%	5%	9%

Note: "Other" includes mainly enhancer states, which may indicate promoter activity without H3K4me3.

Active enhancers in chicken and cattle

	Chicken			Cattle		
	Total	Tissue-specific		Total	Tissue-specific	
Adipose				11,927	2,904	24.35%
Cerebellum	29,440	6,249	21.23%	572	335	58.57%
Cortex	23,373	6,378	27.29%	24,864	6,690	26.91%
Hypothalamus				15,196	2,651	17.45%
Liver	20,390	5,974	29.30%	55,617	29,796	53.57%
Lung	16,339	4,232	25.90%	28,826	7,554	26.21%
Muscle	15,815	6,839	43.24%	24,816	12,763	51.43%
Spleen	16,905	5,546	32.81%	33,660	12,947	38.46%
		Average	29.96%		Average	37.12%

➢ Enhancers are known to be highly tissue-specific

➢ Muscle has the highest percentage of tissue-specific enhancers in chickens, and is also very high in cattle.

Enriched GO terms of closest genes to muscle-specific enhancers

Chicken

Term	Fold Enrichment	FDR
muscle structure development	2.3	2.00E-02
regulation of nucleobase-containing compound metabolic process	1.4	2.10E-01
muscle tissue development	2.4	3.50E-01
regulation of nitrogen compound metabolic process	1.3	4.60E-01
regulation of cellular macromolecule biosynthetic process	1.4	7.00E-01
cellular macromolecule biosynthetic process	1.3	7.10E-01
regulation of macromolecule biosynthetic process	1.3	1.10E+00
STAT cascade	3	1.40E+00
striated muscle tissue development	2.3	1.70E+00
regulation of gene expression	1.3	2.20E+00
negative regulation of protein kinase activity	2.6	2.80E+00
negative regulation of protein modification process	2	2.80E+00
muscle organ development	2.3	3.00E+00
JAK-STAT cascade	2.9	3.10E+00
protein phosphorylation	1.5	3.20E+00

Cattle

Term	Fold Enrichment	FDR
organ morphogenesis	2.1	4.20E-24
embryonic morphogenesis	2.4	6.40E-24
cell fate commitment	3.4	7.10E-23
embryo development	2	3.30E-22
nervous system development	1.7	1.50E-21
generation of neurons	1.9	2.20E-20
neurogenesis	1.8	2.90E-20
regionalization	2.8	5.00E-20
neuron differentiation	1.9	6.70E-20
embryonic organ morphogenesis	3	1.10E-19
epithelium development	2	4.40E-19
embryonic organ development	2.6	6.50E-19
pattern specification process	2.5	7.60E-19
neuron fate commitment	5.2	1.10E-18
tissue morphogenesis	2.2	4.10E-17

Conservation of muscle-specific enhancers

➢ UCSC liftOver tool used to map enhancers between chicken and cattle. No mapping available directly between chicken and cattle, so both were mapped to human

➢ ~20% of chicken enhancers and ~83% of cattle enhancers could be mapped to human genome coordinates

➢ 71 muscle-specific enhancers overlapped between species, which could be associated with 9 chicken genes and 11 cattle genes

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 posters

P0425: Correlating Gene Expression with the Histone Modifications H3K4me3 and H3K27ac in High and Low CpG Content Promoters of Chickens, Cattle, and Pigs

P0527: Predicting Chromatin States to Identify Distinct Active Enhancers within Bursa Tissue of Two Inbred Chicken Lines Under NDV Infection and Heat Stress

P0287: Visualizing Tissue-Specific Regulation of the Equine Genome Using Histone Modification ChIP-Seq

Acknowledgements

UC Davis: Dr. Gerald Quon

Stam Lab (U. Washington)

V. Leesburg, USDA ARS
Jim Kent, UCSC
Laura Clarke, Paul Flicek, EBI
Bin Ren, UCSD

UC Davis FAANG Group
(chicken, pig, cattle and horse)