

# Annotation of Functional Regulatory Elements in Livestock Species

Midwest ASAS

March 17, 2015

Huaijun Zhou<sup>1</sup>, Pablo Ross<sup>1</sup>, Ian Korf<sup>1</sup>, Mary Delany<sup>1</sup>, Hans Cheng<sup>2</sup>, Juan Medrano<sup>1</sup>, Alison Van Eenennaam<sup>1</sup>, Chris Tuggle<sup>3</sup>, Cathy Ernst<sup>4</sup>, Vicki Leesburg<sup>5</sup>

<sup>1</sup>Animal Science, University of California, Davis, CA

<sup>2</sup>ADOL, USDA-ARS, East Lansing, MI

<sup>3</sup>Animal Science, Iowa State University, Ames, IA

<sup>4</sup>Animal Science, Michigan State University, East Lansing, MI

<sup>5</sup> LRRL, USDA-ARS, Miles City, MT



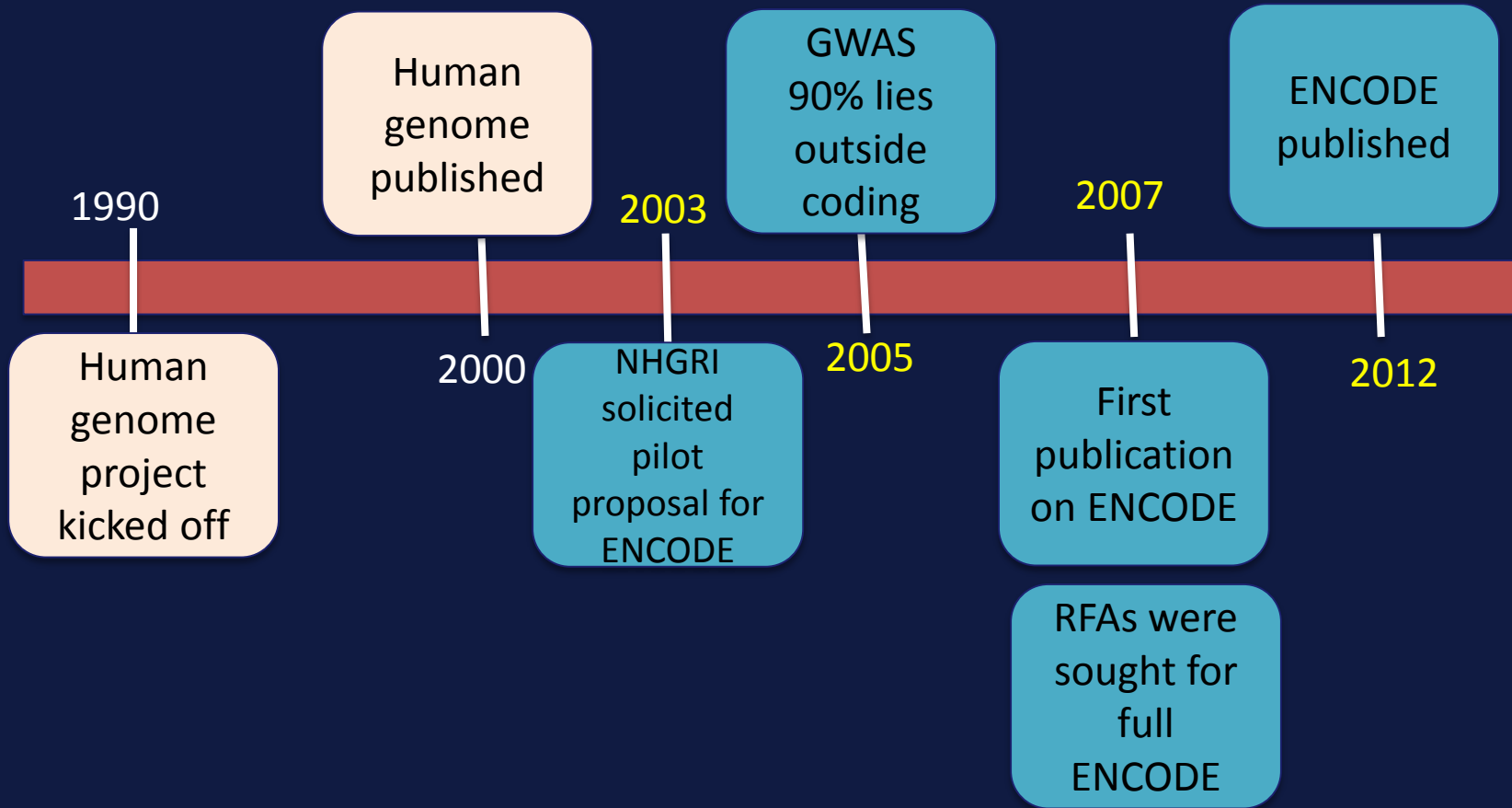
# Outline

- ENCODE (Encyclopedia of DNA Elements)
- Farm animal ENCODE pilot projects
- FAANG (Functional Annotation of Animal Genomes) initiative
- Future plan

# Human ENCODE Project

- 95% of genome is “junk” (<3% of genome is coding)
- cis regulatory elements occur within a limited genome distance
- **Objective:** Catalog functional elements in the human genome

# Human ENCODE Project Roadmap (200M \$)



# BY THE NUMBERS

The ENCODE project involved hundreds of people from around the world, and a lot of editing, disk space and phone calls.

**32** INSTITUTES



**442** CONSORTIUM MEMBERS

DATA



**1,649**  
EXPERIMENTS

**11,972**

FILES ANALYSED



**15 TB**

DISK SPACE USED



ENCODE Wiki

**741**  
WIKI  
CONTENT  
PAGES

**18,500**

PAGE EDITS SINCE 2008

**248,140**

VIEWS

TELECONFERENCING MAY 2008 TO JUNE 2012

**675**

CALLS MADE



**13**  
PARTICIPANTS  
PER CALL



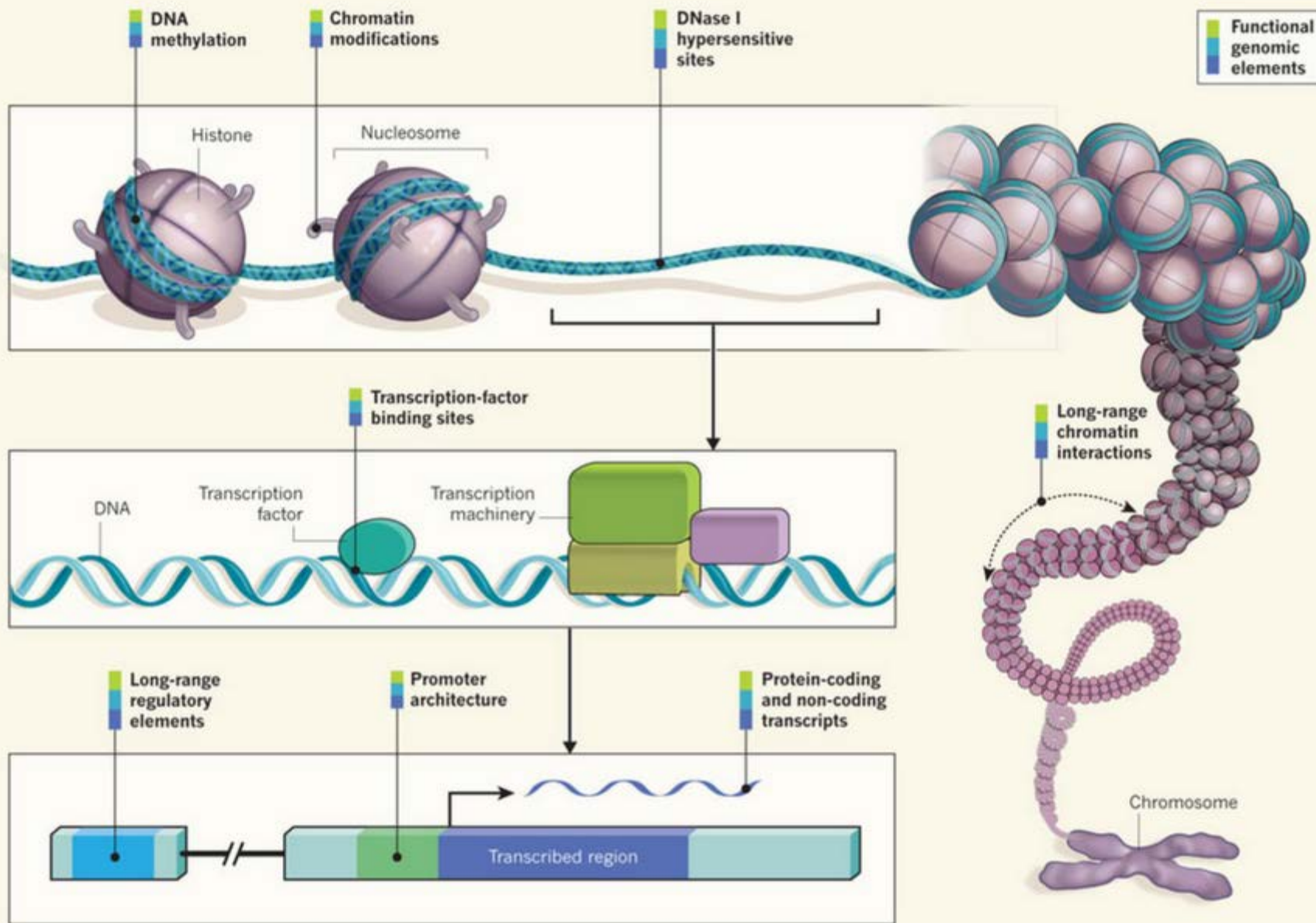
**45m**  
MINUTES PER CALL  
PER PARTICIPANT

**292**

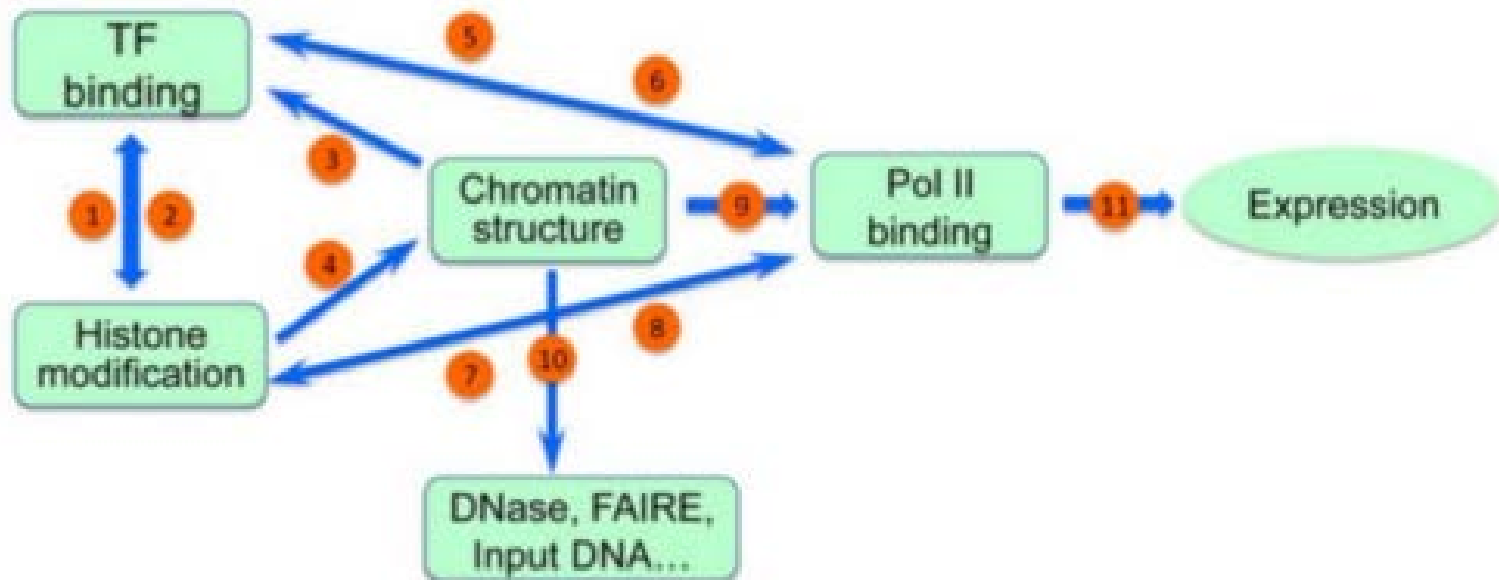
PERSON-DAYS SPENT  
ON CONFERENCE CALLS

TOTAL COST OF TELECONFERENCING = **£49,310.54**

Nature 489, 49–51 (06  
September 2012)  
doi:10.1038/489049a



## Transcriptional regulation – model

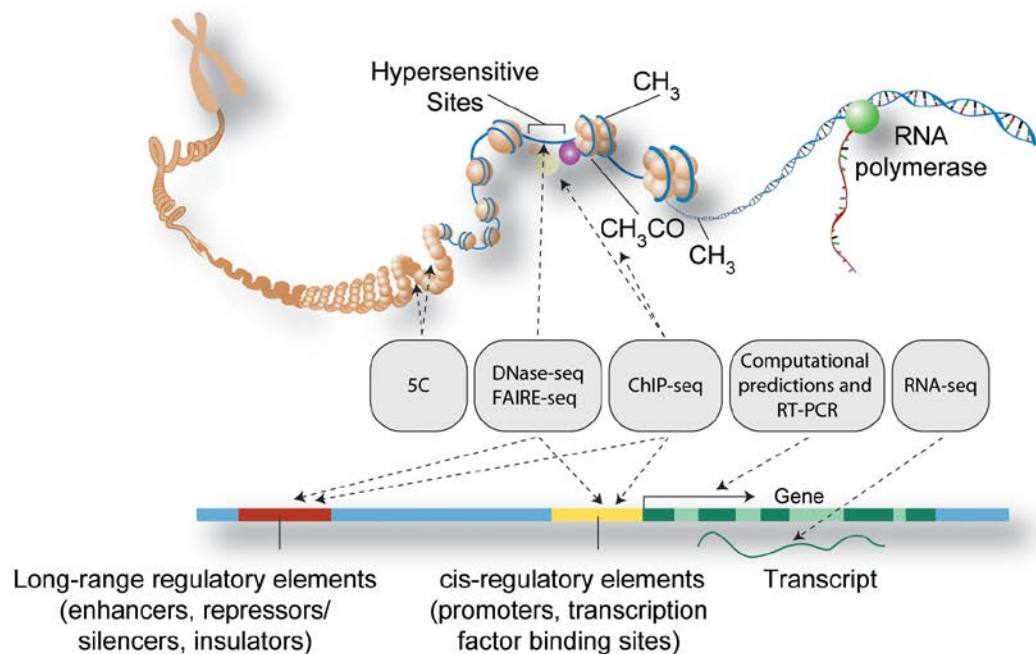


- (1) Recruiting histone modifiers
- (2) Recruiting TFs
- (3) Accessibility
- (4) Remodeling
- (5) Recruiting general TFs
- (6) Interacting with TFs

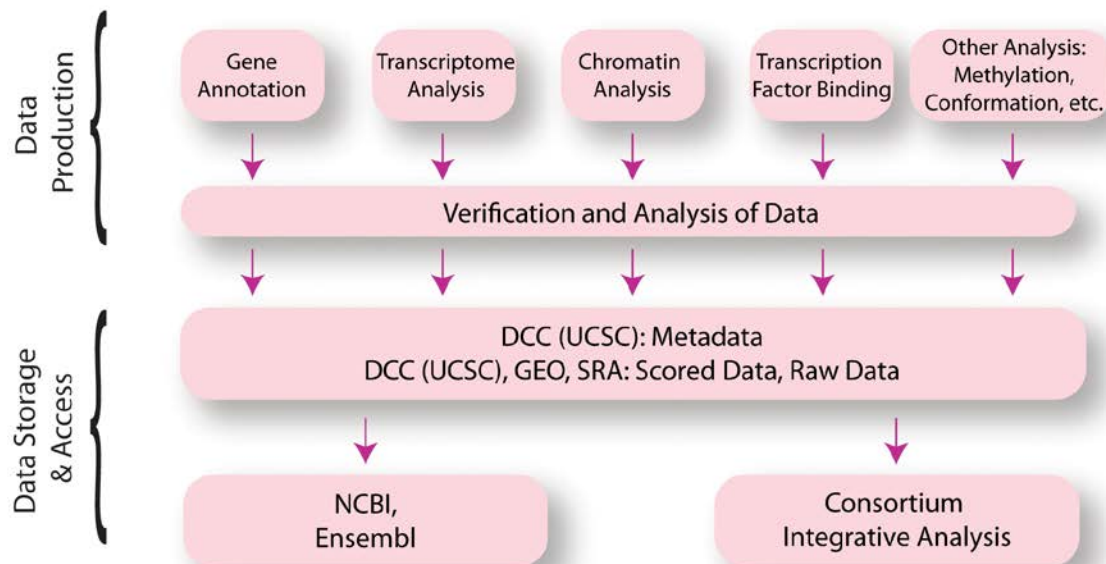
- (7) Recruit general TFs
- (8) Interacting with histone modifiers
- (9) Accessibility
- (10) Accessibility
- (11) Transcription



A.



B.



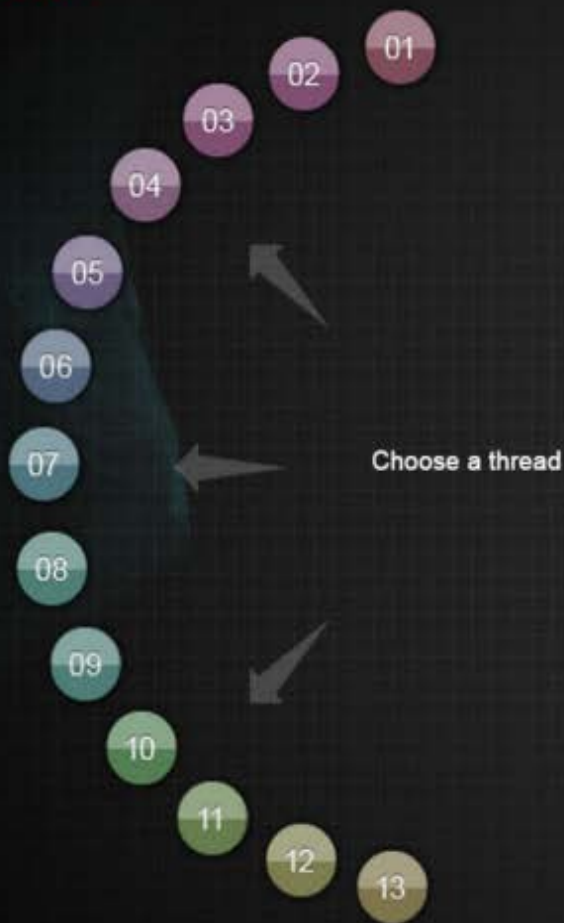
Plos Biology

DOI: 10.1371/journal.pbio.1001046



nature  
**ENCODE** explorer

## THREADS



30 papers published in 3  
journals at the same time

PRODUCED WITH  
SUPPORT FROM  
illumina

Welcome to the  
nature  
**ENCODE** explorer

Access the collected papers by exploring the thematic threads that run through them, with topics such as DNA methylation, RNA or machine learning.

Select a thread to start

# Key Findings

- 80% of the human genome participates in at least one functional activity
- Environment is critical in switch on or off for many genes
- 4 millions switches control ~21,000 genes
- “switches” are software and genes are “hardware”

# Key Findings

- 1% of the genome is distally regulated (Kbs to Mbs) using 5C
  - On an average 3.9 distal elements interacted with TSS
- 2.9 million DHS in 125 diverse cells and tissue types
- 10-12 expressed isoforms per gene per cell

# How about farm animals?

- Genome to Phenome
- Wealth genetic and phenotypic diversity
- Diverse evolution: Mammal, ruminant, avian species
- Extensive pedigree records for economically important complex traits

# Challenges

- Many species
- Many genetic lines
- Different development stages
- Limited resource
- Quality of genome assembly

# Genome-wide identification and annotation of functional regulatory regions in livestock species (New USDA grant)

- 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



# Specific Objectives

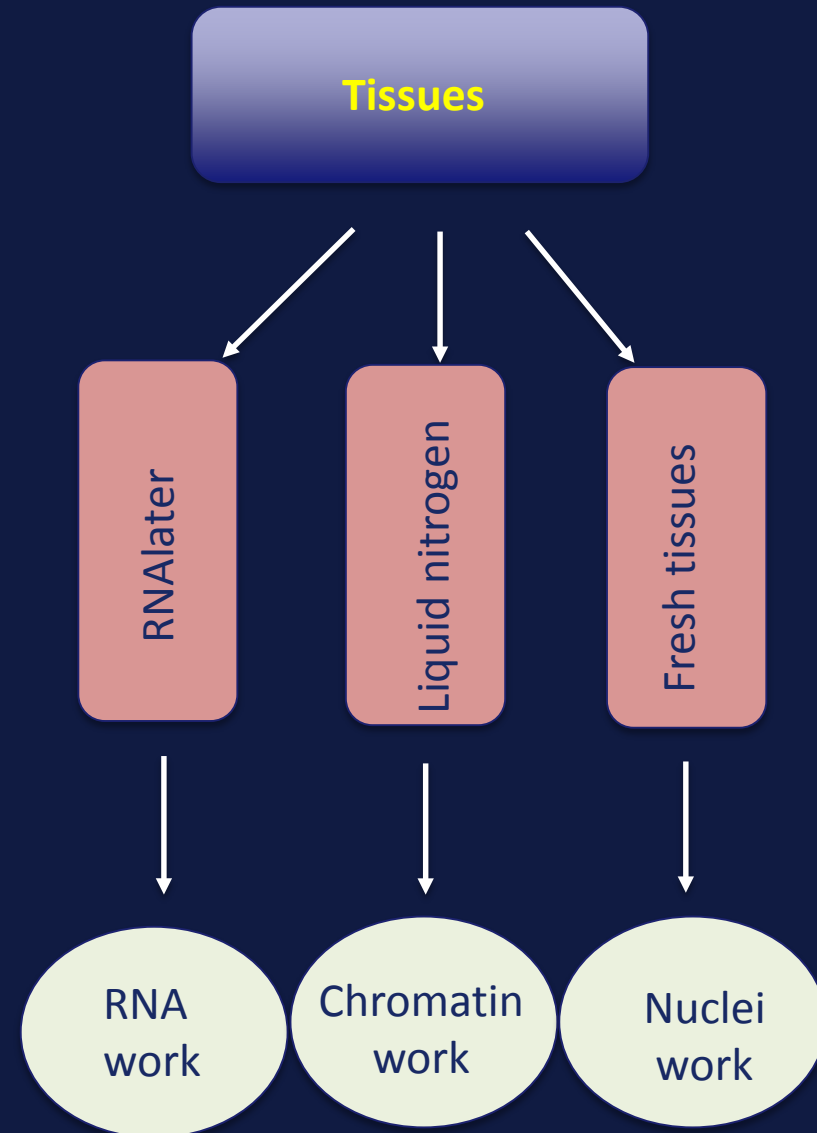
- Annotate chromatin states corresponding to DNase I hypersensitivity, four histone modifications, and one insulator element.
- Identify and annotate promoters, enhancers, and silencers by integrating information from RNA-seq, DNase I hypersensitivity, and ChIP-seq.
- Freely distribute all raw and annotated data via UCSC Genome Browser and Ensembl.

# Genetic lines

- Adult stages
- Chickens: F1 (**Line 6 X 7**) from ADOL, 2 males and 2 females at 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 at 14 months of age
- Pigs: **Yorkshire** from MSU, 2 littermates males at 6 months of age

## Tissues collection

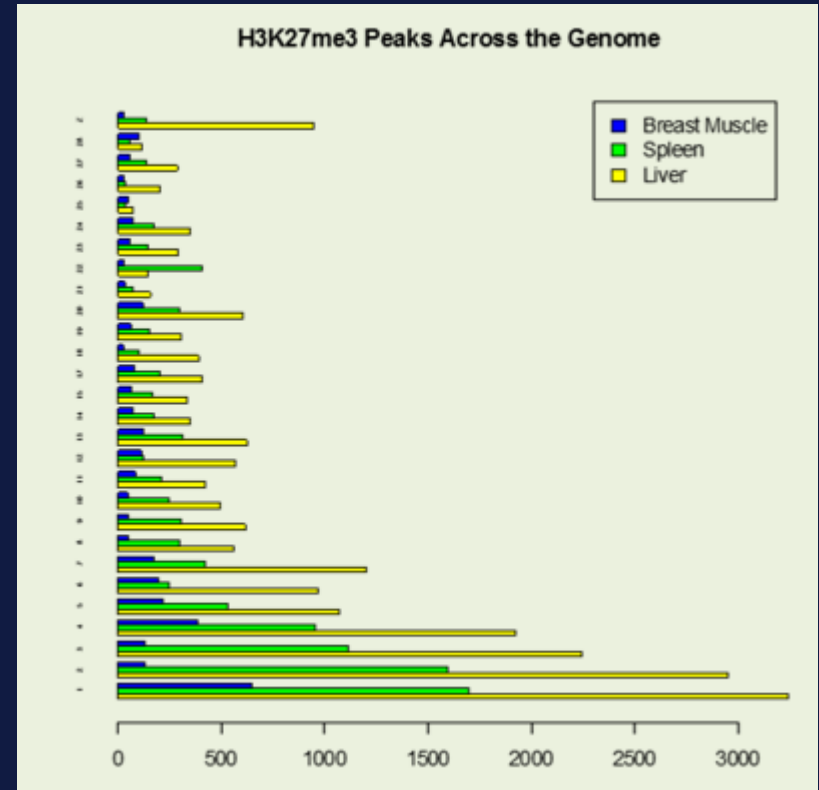
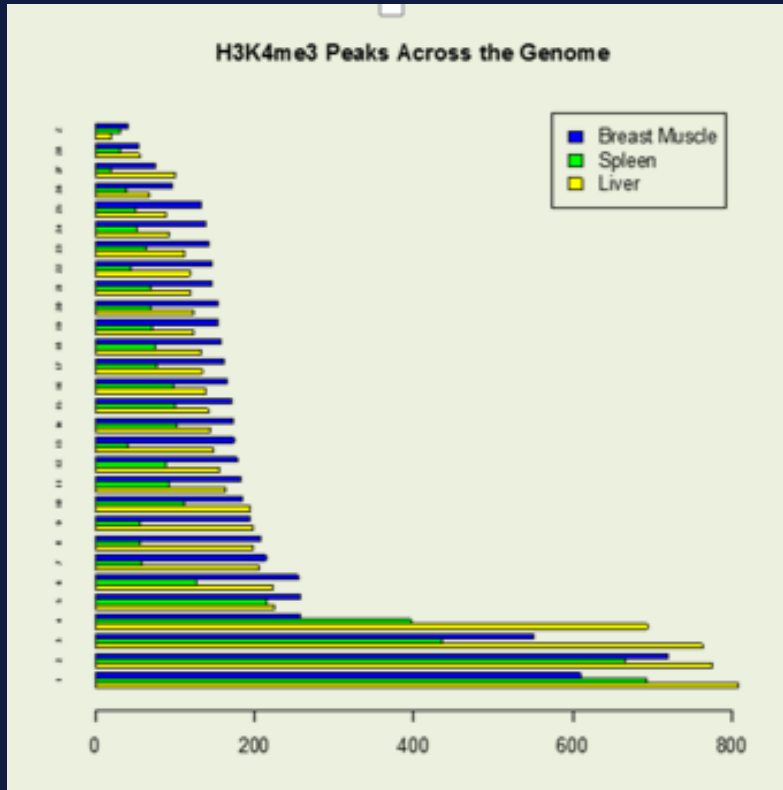
Tissues	Cow	Pig	chicken
Liver	X	X	X
Kidney	X	X	X
Skeletal muscle	X	X	X
Adipose	X	X	X
Spleen	X	X	X
Pancreas	X	X	X
Brain cortex/whole	X	X	X
Cerebellum	X	X	X
Hypothalamus		X	X
Heart	X	X	X
Lung	X	X	X
Trachea	X	X	X
Bladder	X	X	
Thyroid gland	X	X	X
Bursa			X
Esophagus	X	X	X
Thymus	X	X	X
Skin	X	X	X
Mammary gland	X	X	
Ovary	X	X	X
Oviduct	X	X	X
Uterus	X	X	
Testis	X	X	X
Bone marrow	X	X	X
Blood	X	X	X
Cartilage	X	X	X
Duodenum	X	X	X
Jejunum	X	X	X
Ileum	X	X	X
Cecum			X
Colon	X	X	X
Abomasum/Stomach/Gizzard	X	X	X



# Assays

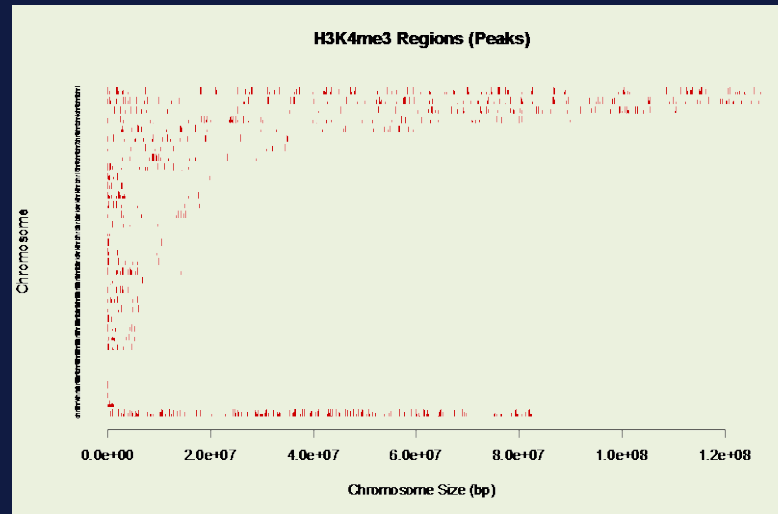
- Stranded RNA-seq
- Chromatin accessibility: DNase-seq (ATAC-seq)
- Histone modification marks
  - H3K4me3 (promoters of active genes and transcription starts)
  - H3K27me3 (silence)
  - H3K27AC (active regulatory elements)
  - H3K4Me1 (enhancers and other distal elements)
- Transcription factor CTCF (insulator)

# Distribution of H3K4me3 and H3K27me3 across chicken genome

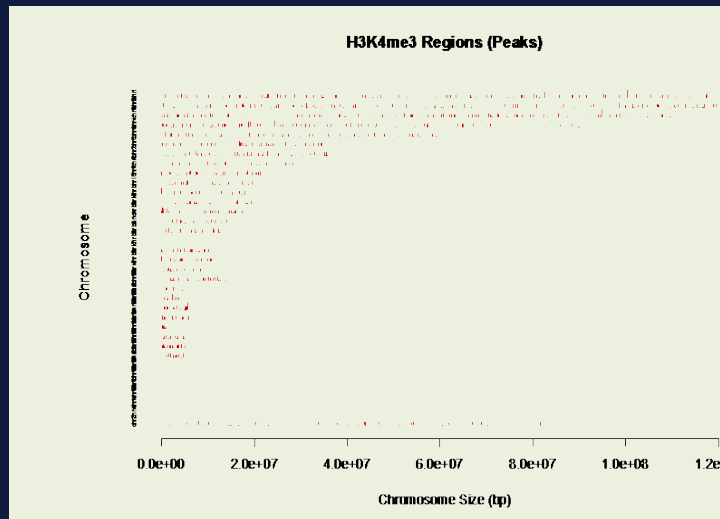


# H3K4me3 across chicken genome in 3 tissues

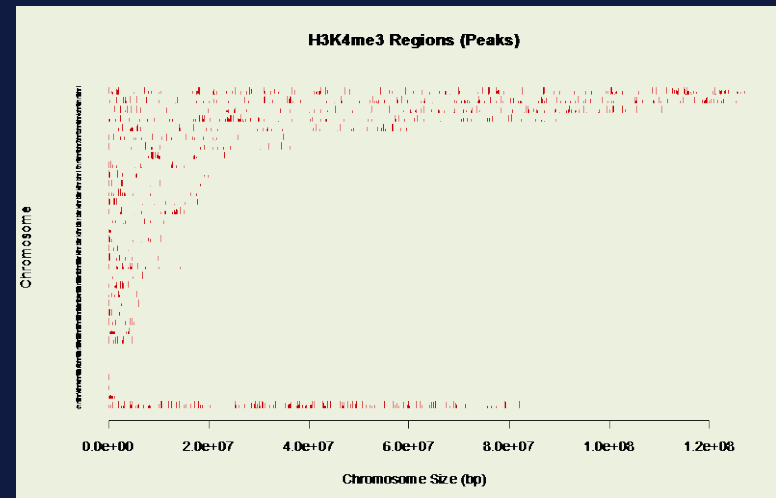
Liver



Spleen



Breast muscle



# Candidate genes demonstrating tissue-specific chromatin state

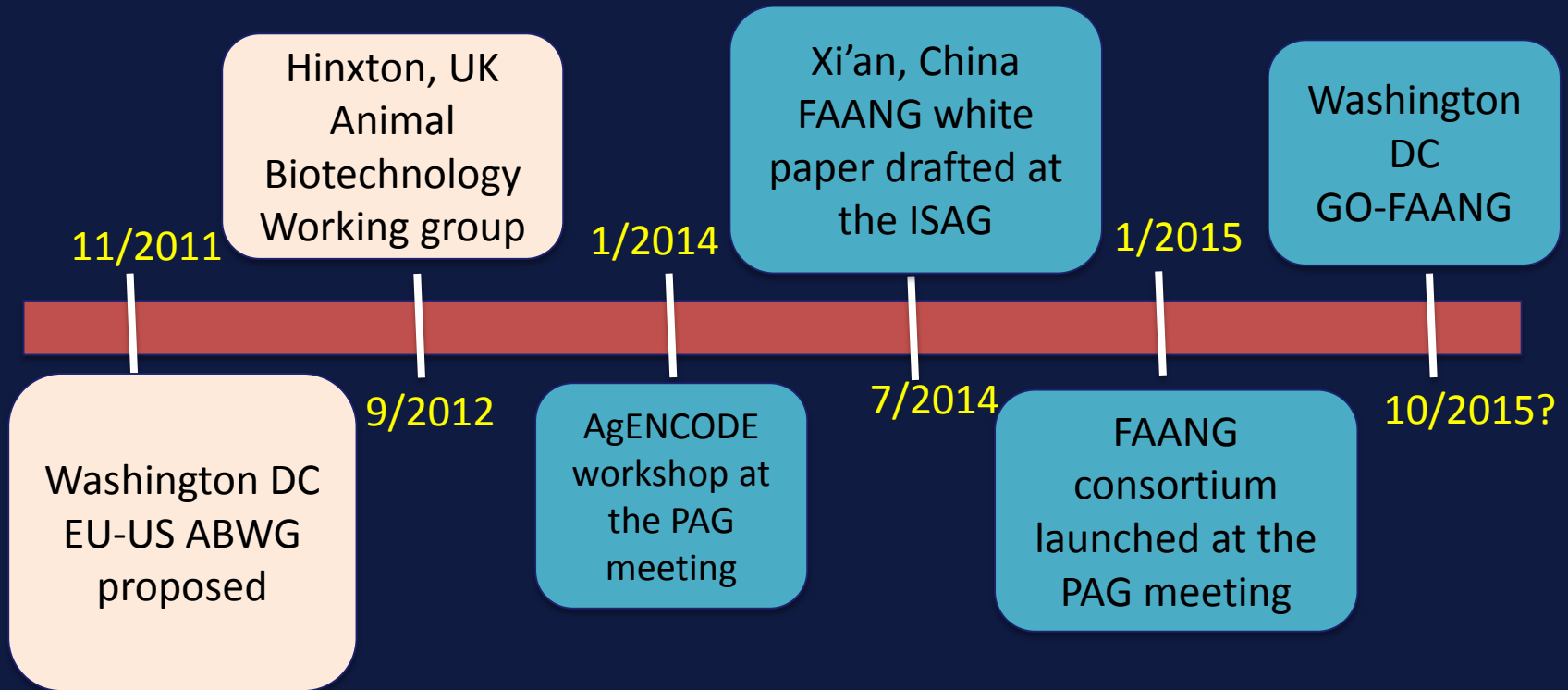




# Other pilot projects

- WUR-pigENCODE (Groenen et al.)
  - Pig
  - Duroc, Large White, Pietrain
  - RNA-seq, Small-RNA-seq, Methylation
- FR-AgENCODE (Giuffra et al.)
  - Pig: Large White; Chicken; White Leghorn; Cattle: Holstein; Goat: Alpine
  - RNA-seq, Small RNA-seq, and Hi-C
- Sheep ENCODE (Kijas et al)
  - USDA MARC , Baylor College of Medicine, Utah State University and CSIRO
  - Rambouillet ewe
  - Iso-seq, RNA-seq, Small-RNA-seq
- Pig ENCODE in China (PENCODE China, Huang et al)
  - 300 F6 mosaic pigs

# FAANG Consortium Roadmap (\$\$\$???)



[http:// www.FAANG.org](http://www.FAANG.org)



## Functional Annotation of ANimal Genomes (FAANG) Project

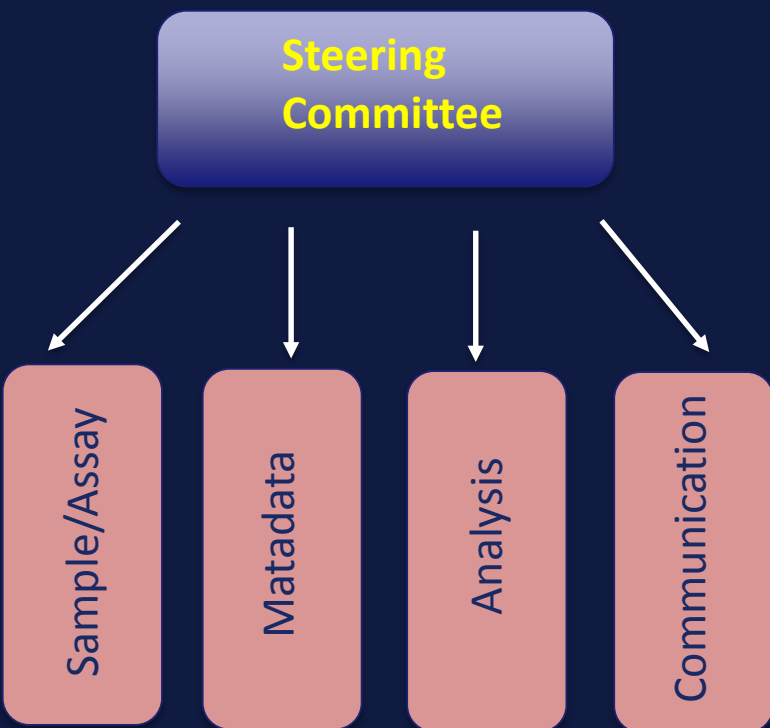
— A coordinated international action to accelerate Genome to Phenome

### Working Committees:

- Coordinators

	<b>Analysis</b>	<b>Communication</b>	<b>Metadata</b>	<b>Samples</b>	<b>Steering</b>
Clarke, Laura			✓		
Giuffra, Elisabetta				✓	
Reecy, James	✓				
Schmidt, Carl			✓		
Silverstein, Jeffrey		✓			
Tuggle, Christopher		✓			✓
Watson, Mick	✓				
Zhou, Huaijun				✓	

# FAANG Organization Chart



## Members

	Analysis	Communication	Metadata	Samples	Steering
Acloque, Herve				✓	
Amaral, Andreia			✓		
Archibald, Alan		✓	✓		✓
Brown, C. Titus	✓				
Burt, David	✓			✓	✓
Cheng, Hans		✓		✓	✓
Clarke, Laura	✓	✓	✓†	✓	✓
Crooijmans, Richard				✓	
Di Palma, Federica	✓				
Eory, Lel	✓				
Flicek, Paul	✓		✓		
Foissac, Sylvain	✓				
Giuffra, Elisabetta	✓	✓		✓†	✓
Groenen, Martien		✓			✓
HUE, Isabelle			✓		
Hayes, Ben				✓	✓
Huang, LuSheng					✓
Janacek, Sophie		✓			
Kijas, James				✓	✓
Kim, Heebal	✓		✓		✓
Kuo, Richard	✓				
MacLeod, James			✓		✓
Plastow, Graham		✓		✓	✓
Reecy, James	✓†				✓
Richardson, David	✓		✓		
Ross, Pablo	✓		✓		
Schmidt, Carl			✓†		✓
Silverstein, Jeffrey		✓†			✓
Song, Jiuzhou	✓				
Streeter, Ian	✓		✓		
Su, Ying				✓	
Swarbreck, David	✓				
Tosser-Klopp, Gwenola		✓	✓		✓
Tuggle, Christopher		✓†			✓†
Watson, Mick	✓†				✓
White, Stephen		✓		✓	✓
Yates, Andrew	✓		✓		
Zerbino, Daniel	✓		✓		
Zhou, Huaijun	✓	✓		✓†	✓
madsen, ole	✓				

# Future plan

- Sub-committee will work with animal genome community to coordinate FAANG activities
  - Sample protocol
  - Assay protocols
  - Data standard
  - Infrastructure
  - Foster collaborations
- Reach out more funding agencies worldwide
  - USDA, NSF, NIH, BBSRC, Europe Commission, Genome Canada, NSF China, Australia?, Korea?

# Future FAANG workshops/meetings

- July, 2015, Piacenza, Italy, ISAFG
- October, 2015, Washington DC, GO-FAANG
- January, 2016, FAANG workshop, PAG
- July, 2016, Salk city, UT, ISAG
- July, 2017, Dublin, ISAG

# Financial support



United States  
Department of  
Agriculture

National Institute  
of Food and  
Agriculture

USDA NIFA grant #2015-67015-22940

Poultry, Cattle, Swine  
Genome Coordination  
Funds

