

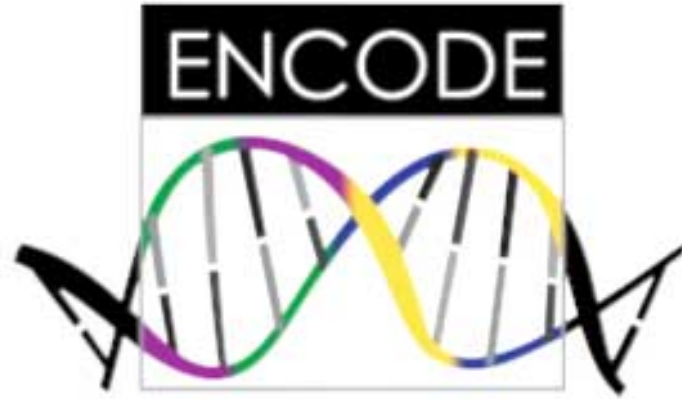
United States
Department of
Agriculture

National Institute
of Food and
Agriculture

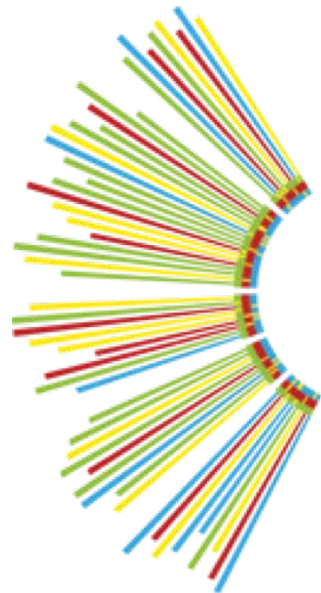
Functional Annotation of the Animal Genome FAANG

American Society of Animal Science, Canadian Society of Animal Science
Midwestern Section
March 11, 2019





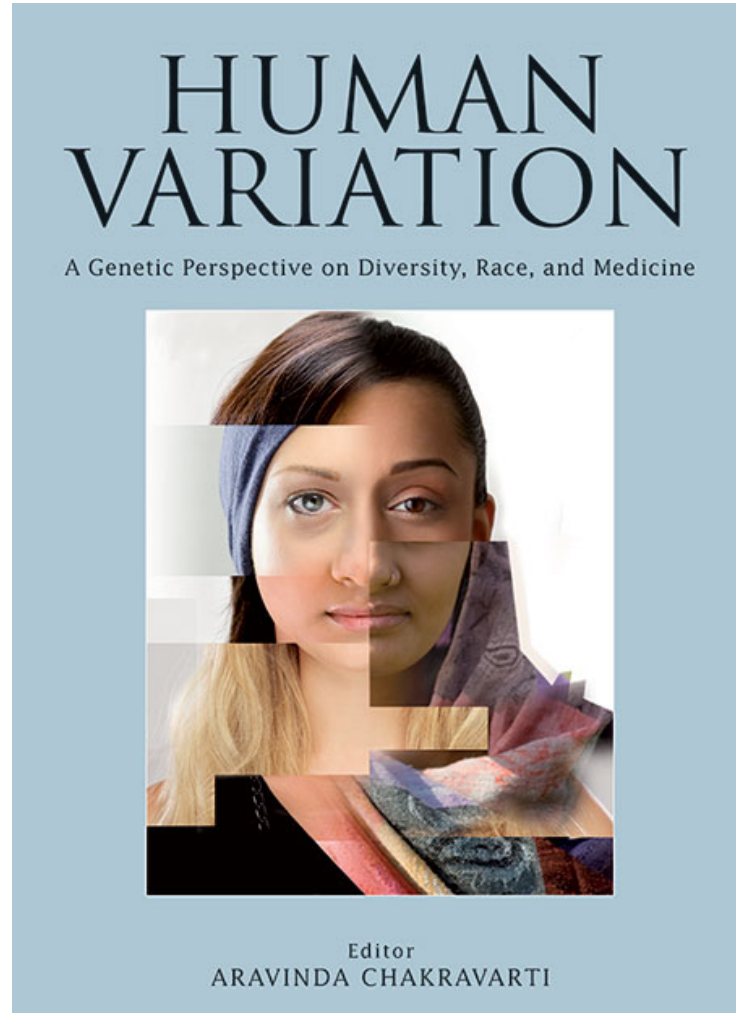
ENCODE: Encyclopedia of DNA Elements



FAANG
Functional Annotation of Animal Genomes

Biological Variation

- What are the sources of variation in biological traits?



Biological Variation

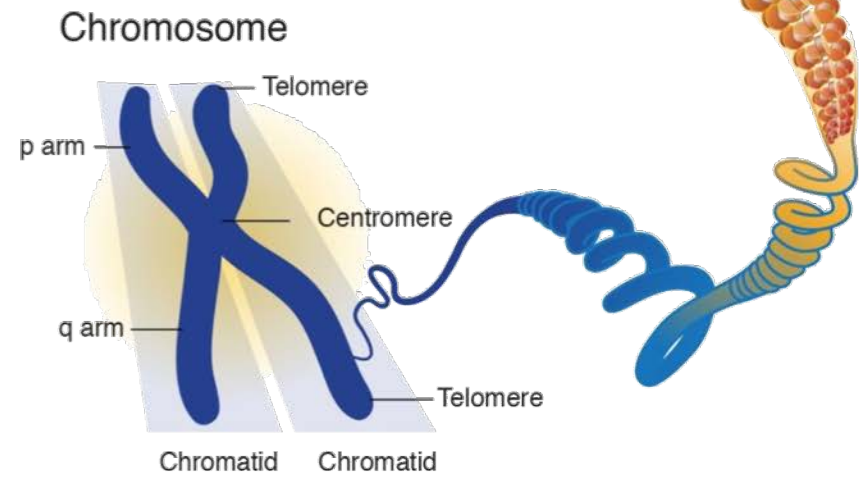
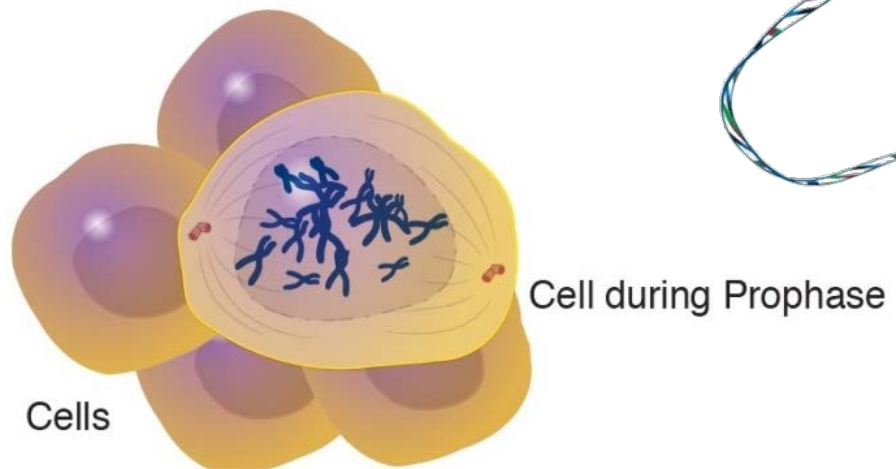
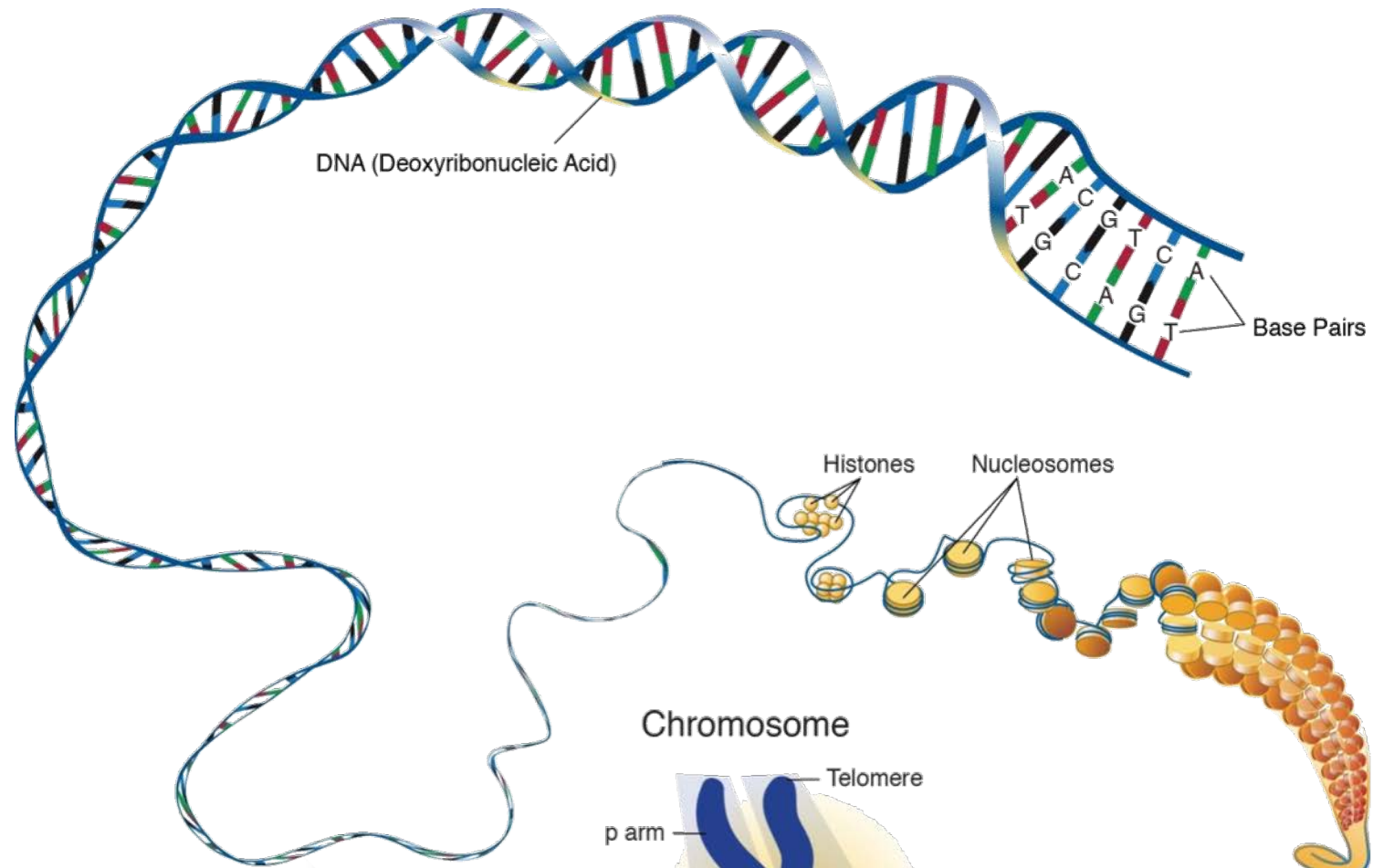
- What are the sources of variation in biological traits?
- Caused by
 - Genetics
 - Environment

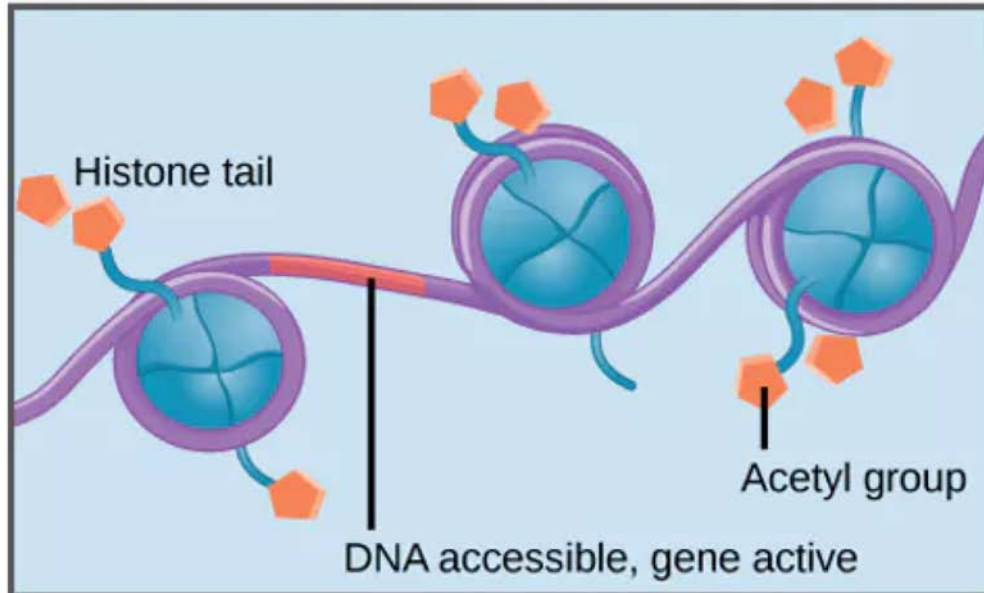
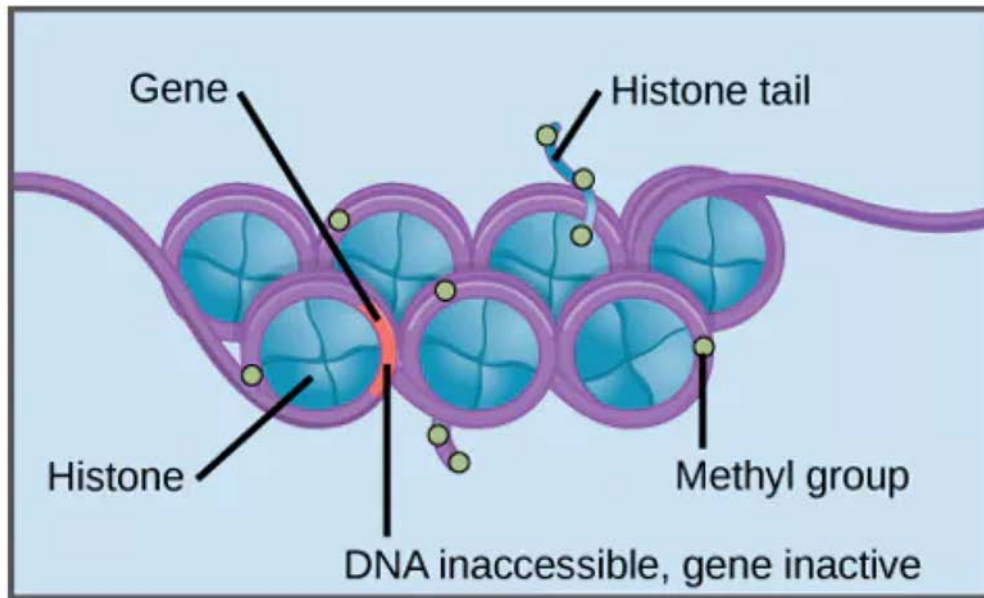


Biological Variation

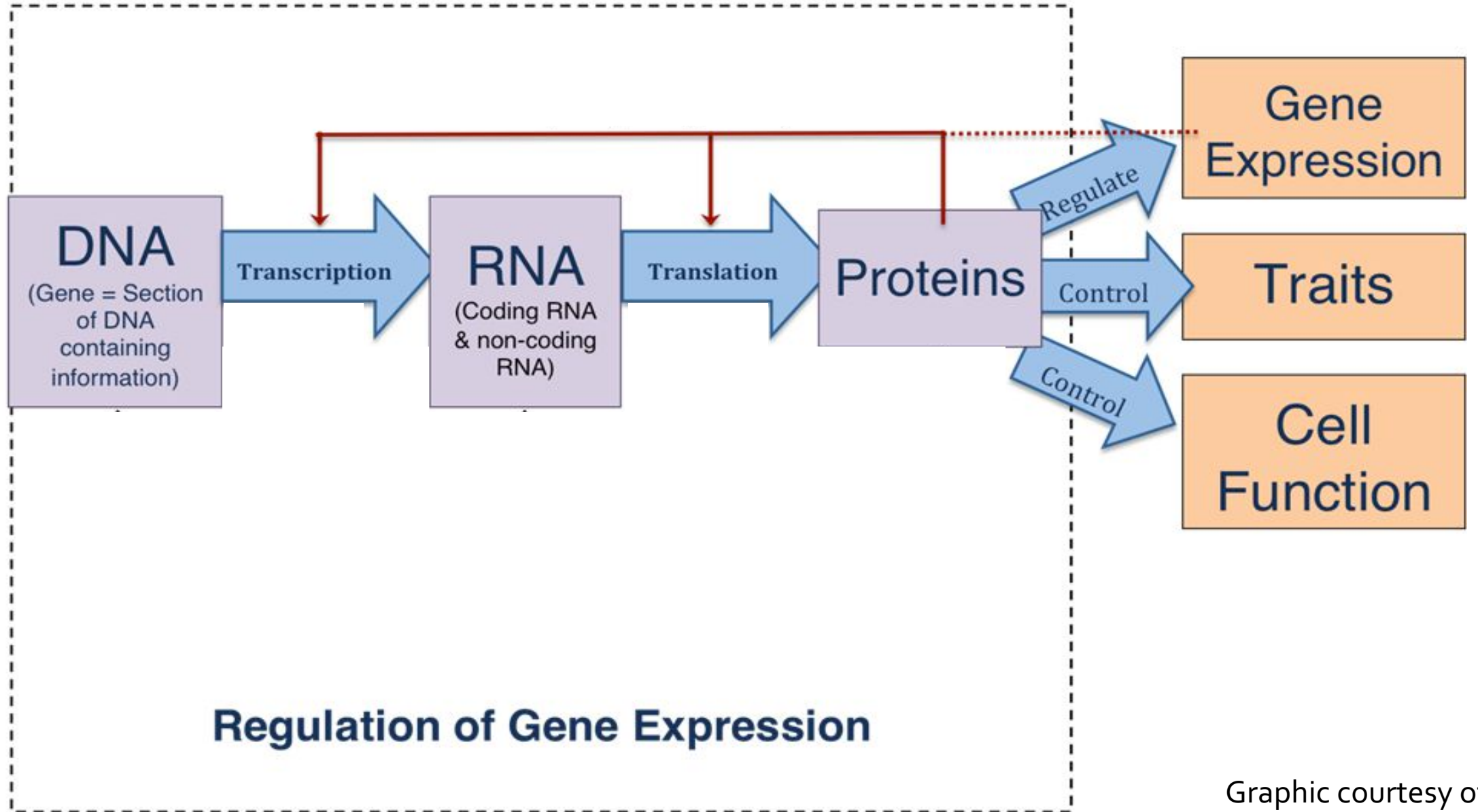
- What are the sources of variation in biological traits?
- Caused by
 - Genetics
 - Environment







- Methylation histone & DNA causes nucleosome to be tightly pack
- Transcription factors cannot bind DNA & genes are NOT expressed
- Histone acetylation results in loose packaging
- Transcription factor CAN bind DNA & genes can be expressed

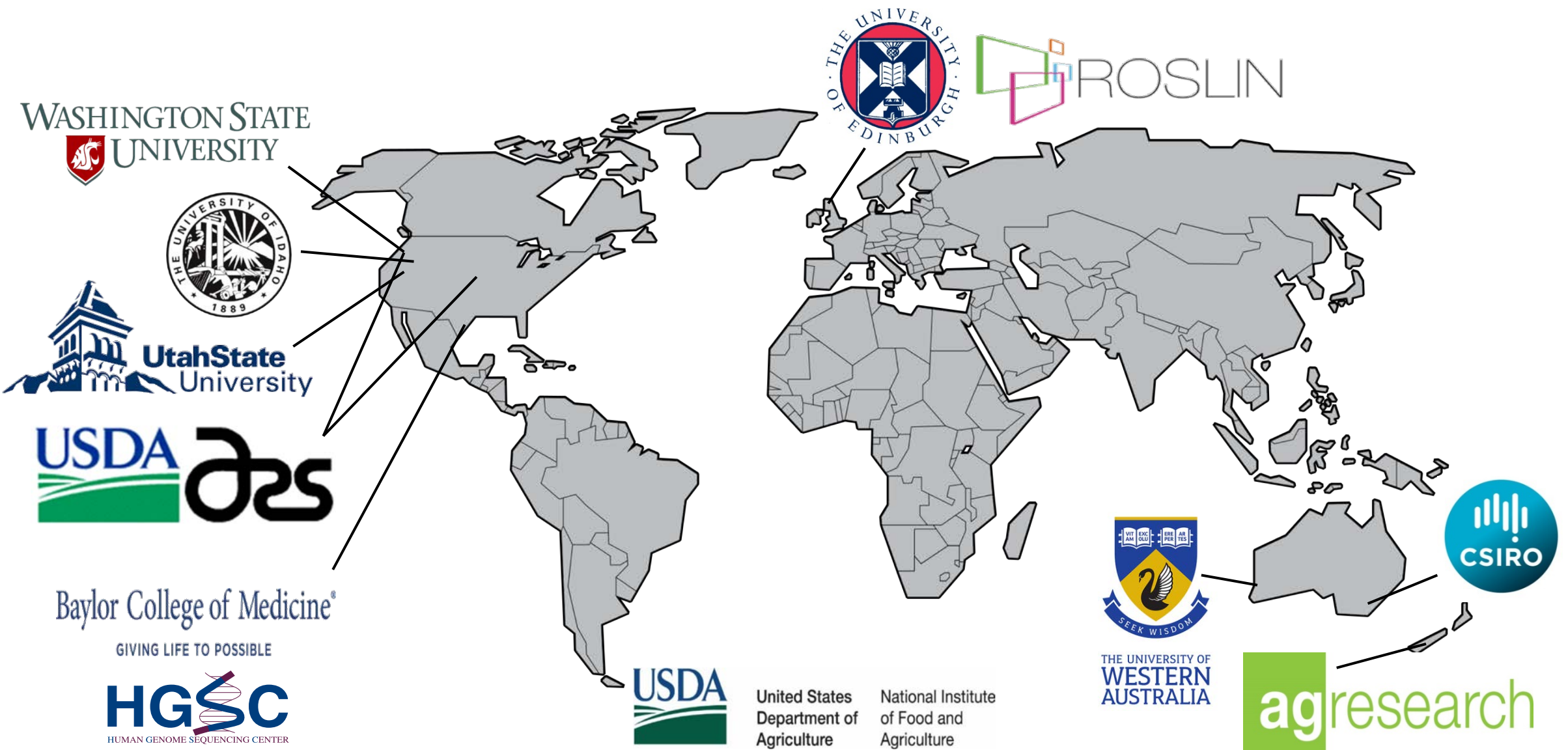


Graphic courtesy of
Marianne Dobrovoly

Ovine FAANG Project

Brenda Murdoch (University of Idaho), Stephen White (USDA, ARS, Washington State University), Michelle Mousel (USDA, ARS ADRC), Alisha Massa (Washington State University), Kim Worley (Baylor College of Medicine), Alan Archibald (Roslin Institute), Emily Clark (Roslin Institute), Brian Dalrymple (University of Western Australia), James Kijas (CSIRO), Shannon Clarke (AgResearch), Rudiger Brauning (AgResearch), Timothy Smith (USDA, ARS MARC), Tracey Hadfield (Utah State University), Noelle Cockett (Utah State University).

Acknowledgements



Objectives

1. Develop a deep and robust dataset of **transcribed** elements in the sheep genome.
2. Develop a deep and robust dataset of transcription **regulatory features** in the sheep genome.
3. Annotation of the ovine genome reference assembly.
4. Provide ovine FAANG data and support to the public.

Materials and Methods

- We generated a new high quality reference genome sequence *de novo* assembled using (PacBio) sequence and Hi-C data.
 - The 2.87 Gb assembly (Oar_rambouillet_v1.0, GCA_002742125.1)
- 100 tissues collected (snap & slow frozen) from Benz 2616 in April, 2016 at USU
 - BioProject PRJNA414087
- FAANG Annotations from the same animal - Oar_rambouillet_v1.0



Rambouillet ewes Benz 2616 (right)

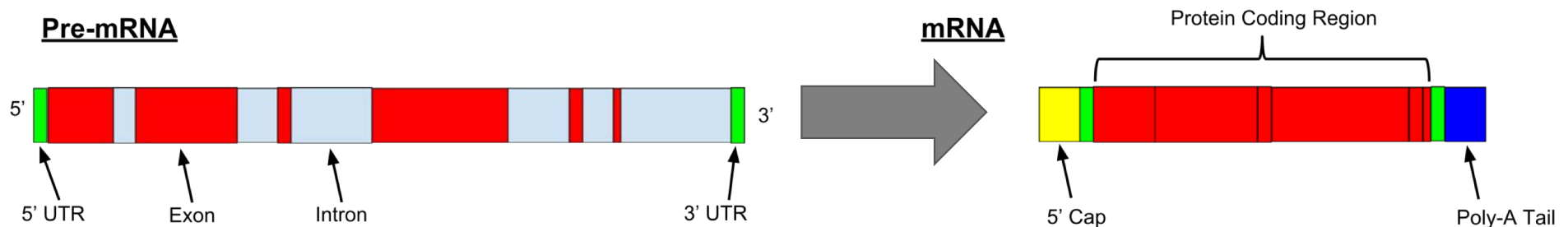
Baylor College of Medicine®

GIVING LIFE TO POSSIBLE

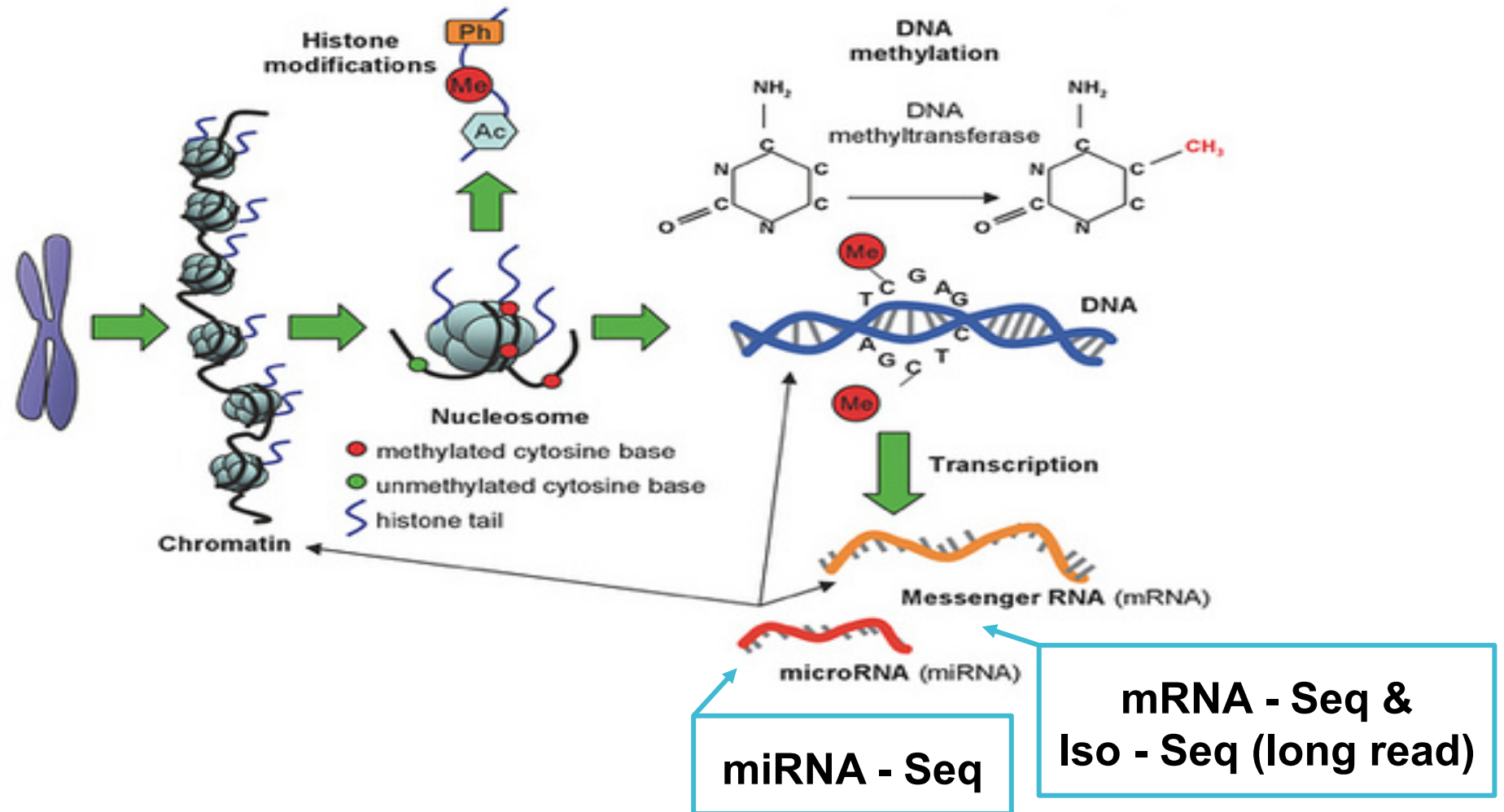
HGSC
HUMAN GENOME SEQUENCING CENTER

Objectives

1. Develop a deep and robust dataset of **transcribed** elements in the sheep genome.
 - Three types of transcriptome data on tissues
 - 1) Using poly-A+ mRNAs transcript sequence
 - 2) Small noncoding microRNAs sequence
 - 3) Full transcript (ISO) sequence PacBio



Sheep FAANG Assays



Objectives

1. Develop a deep and robust dataset of **transcribed** elements in the sheep genome.
 - RNA sequencing and transcriptome analyses is in progress.

| Transcript | Tissues run by Kim Worley | Tissues run by Tim Smith | Total Tissues |
|------------|---------------------------|--------------------------|---------------|
| mRNA | 29 | 31 | 60 |
| miRNA | 30 | - | 30 |
| Iso-Seq | 5 | 10 | 15 |

Objectives

1. Develop a deep and robust dataset of **transcribed elements** in the sheep genome.
2. Develop a deep and robust dataset of transcription **regulatory features** in the sheep genome.
 - 1) Cap Analysis of Gene Expression: CAGE – Sequencing
 - 2) Chromatin accessibility: ATAC – Sequencing
 - 3) Histone modification: Chromatin Immunoprecipitation – Sequencing
 - 4) DNA methylation: Reduced Representation Bisulfite and Whole Genome Bisulfite Sequencing

Sheep FAANG Assays

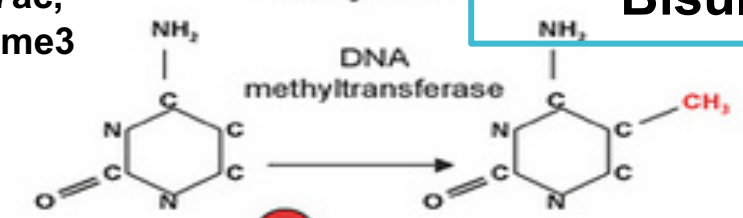
**Histone Modification: Chromatin Immunoprecipitation
ChIP- Seq**

Histone modifications

H3K4me3,
H3K4me1,
H3K27ac,
H3K27me3

**DNA Methylation:
Whole Genome &
Reduced Representation
Bisulfite - Seq**

DNA methylation

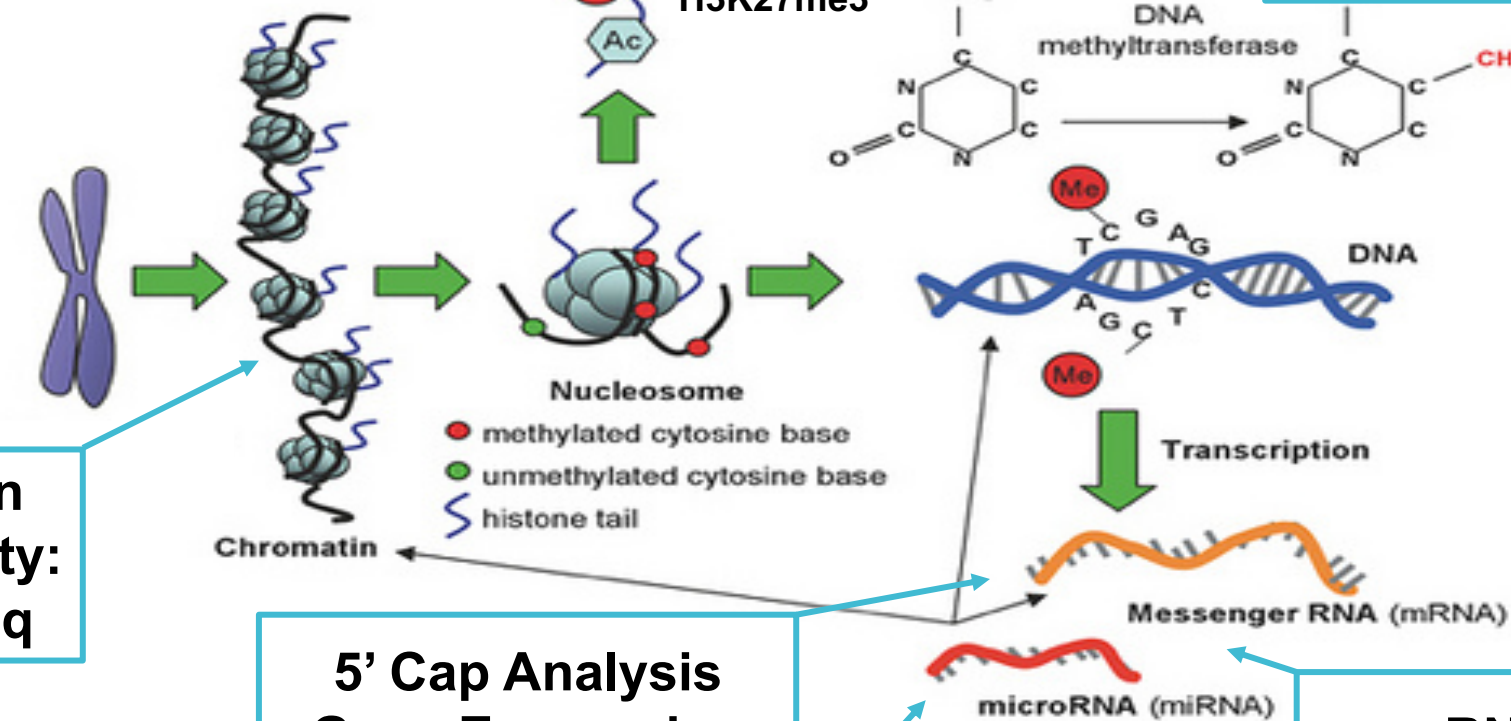


**Chromatin Accessibility:
ATAC - Seq**

**5' Cap Analysis
Gene Expression
CAGE - Seq**

miRNA - Seq

**mRNA - Seq &
Iso - Seq (long read)**

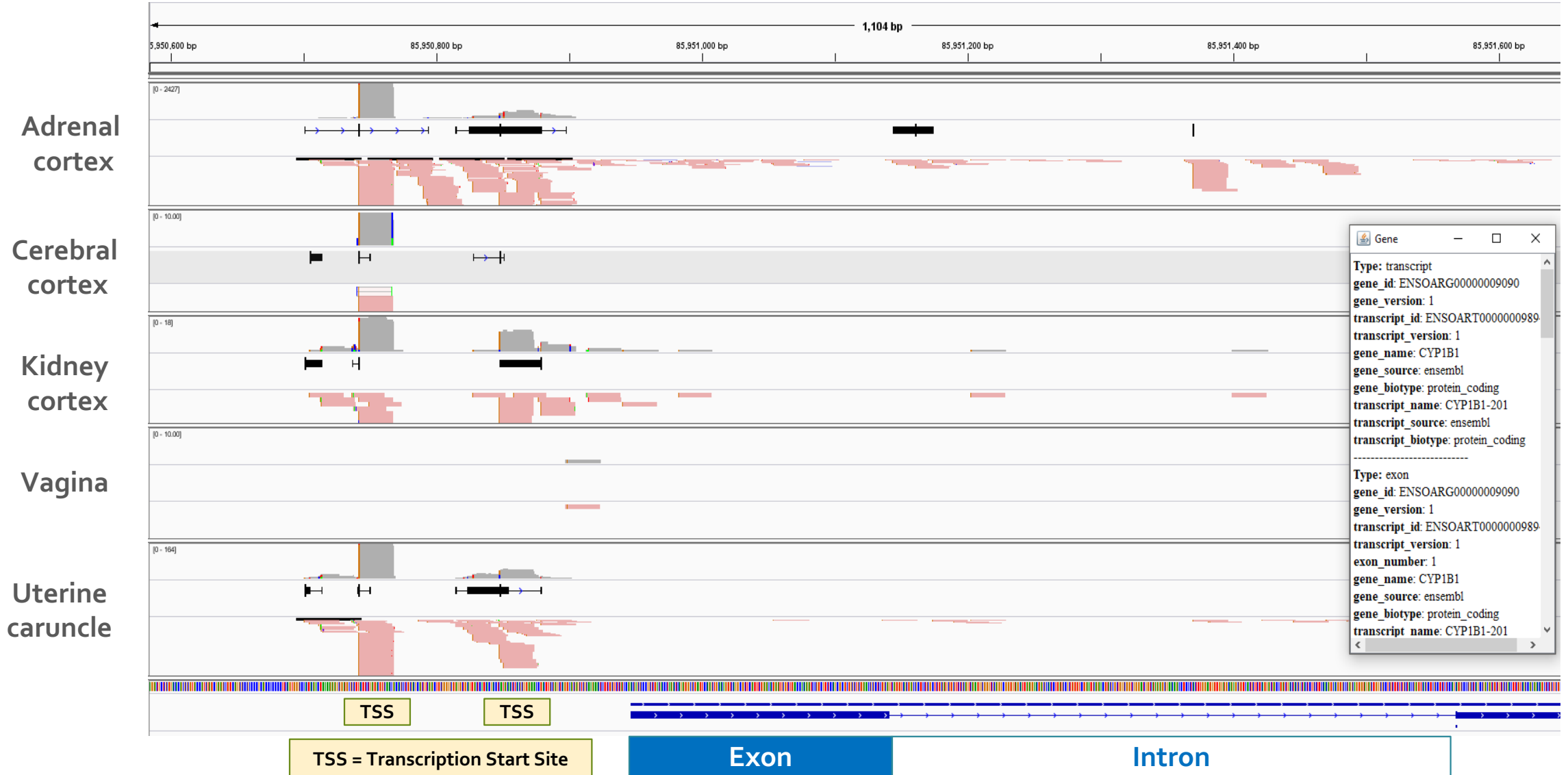


Regulatory Feature Assays

- CAGE assay & analyses, lead by Emily Clark and Alan Archibald is currently in progress.
 - In total 58 samples of mRNA were shipped to Scotland
 - CAGE assay sequenced on the HiSeq 2500 platform
 - To date 40 different tissues have been analyzed

CAGE – Sequence

CYP1B1- Cytochrome P450 family subfamily b1

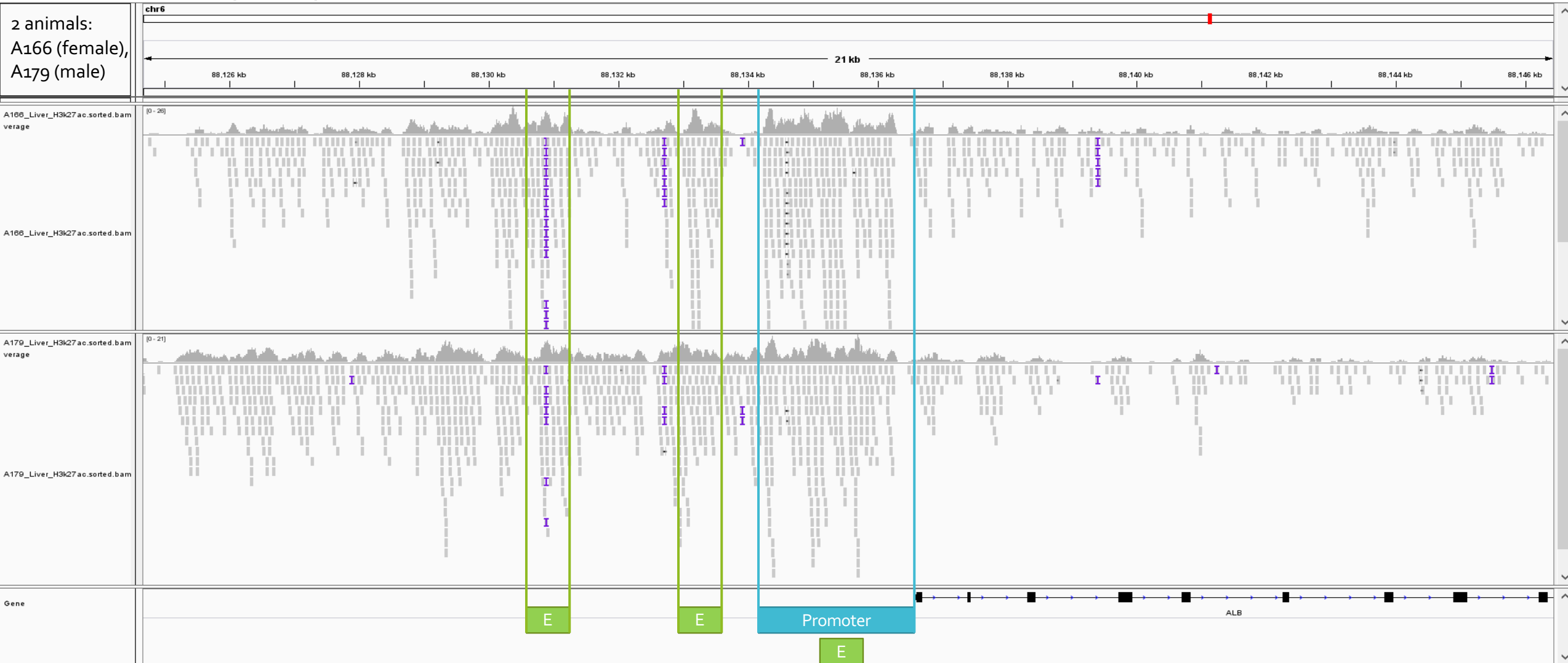


Regulatory Feature Assays

- ChIP-Seq & ATAC-Seq, are in progress in collaboration with Brenda Murdoch, Stephen White and Michelle Mousel.
- Separate pilot projects are underway to ensure protocols are working for sheep tissues.
 - Initial pilot experiment to develop protocols are being analyzed
 - 3 tissues from 2 female and 2 males

ChIP – Sequence for H3K27ac

Albumin (ALB): expected to be expressed in the liver and available for transcription

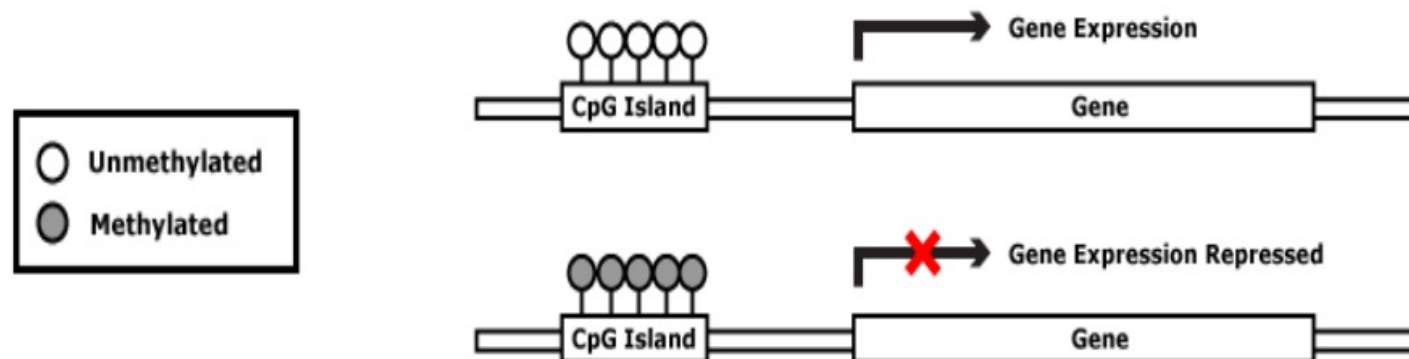


E = known enhancer in humans

Promoter = known promoter in humans

Regulatory Feature Assays

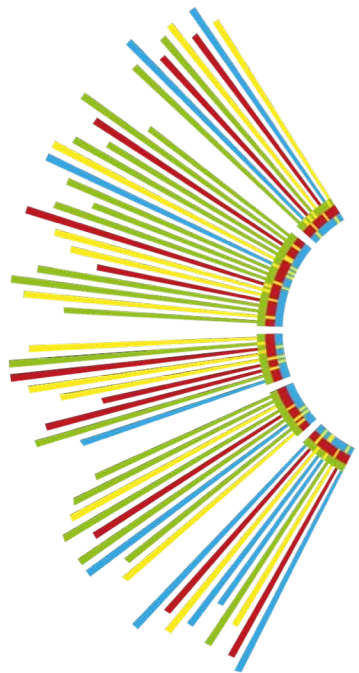
- DNA methylation, is being lead by Shannon Clarke and Rudiger Brauning.
 - Reduced Representation Bisulfite Sequencing
 - on the same 60 tissues as the RNA sequence
 - Whole Genome Bisulfite Sequencing
 - for 22 core FAANG tissues



Summary

1. Transcriptome data for 60 tissues.
2. Transcription **regulatory features** in the sheep genome.
 - CAGE for 40 tissues completed
3. Annotation of the ovine genome reference assembly.
4. Provide ovine FAANG data and support to the public.





United States
Department of
Agriculture

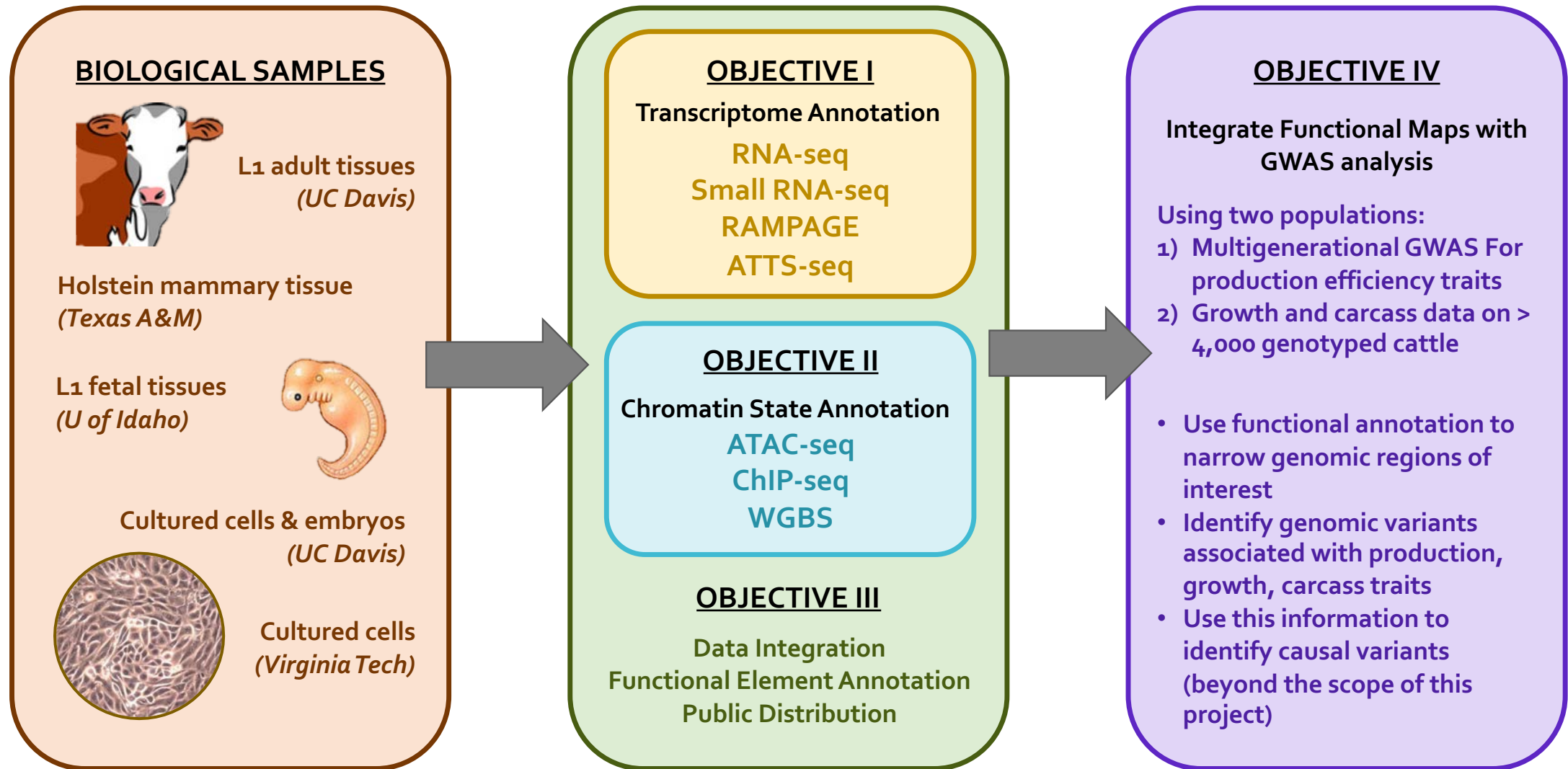
National Institute
of Food and
Agriculture

The Functional Annotation of the Bovine Genome

Pablo Ross ¹, Stephanie McKay ², Clare Gill ³, Monique Rijnkels ³, Brenda Murdoch ⁴, Tim Smith ⁵,
Huaijun Zhou ¹, Zhihua Jiang ⁶, James Reecy ⁷, Wansheng Liu ⁸, Honglin Jiang ⁹, Milton Thomas¹⁰

¹ University of California Davis, ² University of Vermont, ³ Texas A&M University, ⁴ University of Idaho, ⁵ USDA-ARS, ⁶ Washington State University, ⁷ Iowa State University, ⁸ Pennsylvania State University, ⁹ Virginia Tech, ¹⁰ Colorado State University

The Functional Annotation of the Bovine Genome



The Functional Annotation of the Bovine Genome

BIOLOGICAL SAMPLES



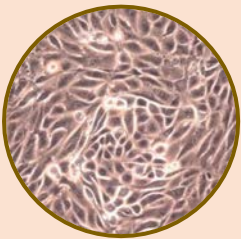
L1 adult tissues
(UC Davis)

Holstein mammary tissue
(Texas A&M)

L1 fetal tissues
(U of Idaho)



Cultured cells & embryos
(UC Davis)



Cultured cells
(Virginia Tech)

- 40 adult tissues from L1 Hereford line
- 5 stages of mammary gland development
- 8 fetal tissues from L1 Hereford line
- 8 primary cell lines

The Functional Annotation of the Bovine Genome

Expressed regions

| | |
|------------------|---------------------------------|
| RNA-seq | transcripts - variants |
| smRNA-seq | Small transcript |
| RAMPAGE | Transcription start sites |
| WTTS-seq | Transcription termination sites |
| Iso-seq | Large transcripts |

OBJECTIVE I

Transcriptome Annotation

RNA-seq
Small RNA-seq
RAMPAGE
ATTS-seq

OBJECTIVE II

Chromatin State Annotation

ATAC-seq
ChIP-seq
WGBS

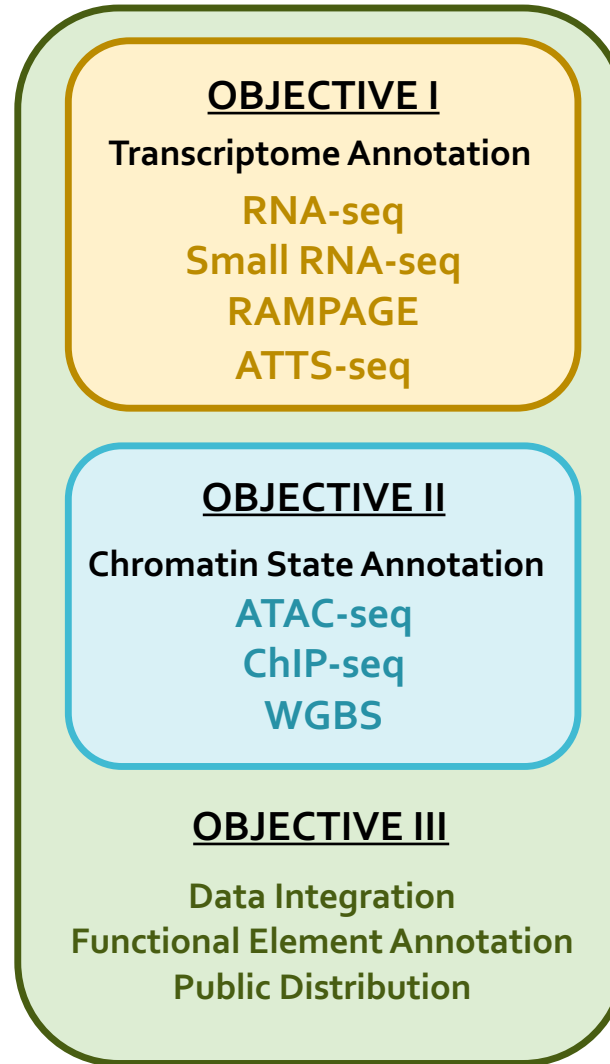
OBJECTIVE III

Data Integration
Functional Element Annotation
Public Distribution

The Functional Annotation of the Bovine Genome

Expressed regions

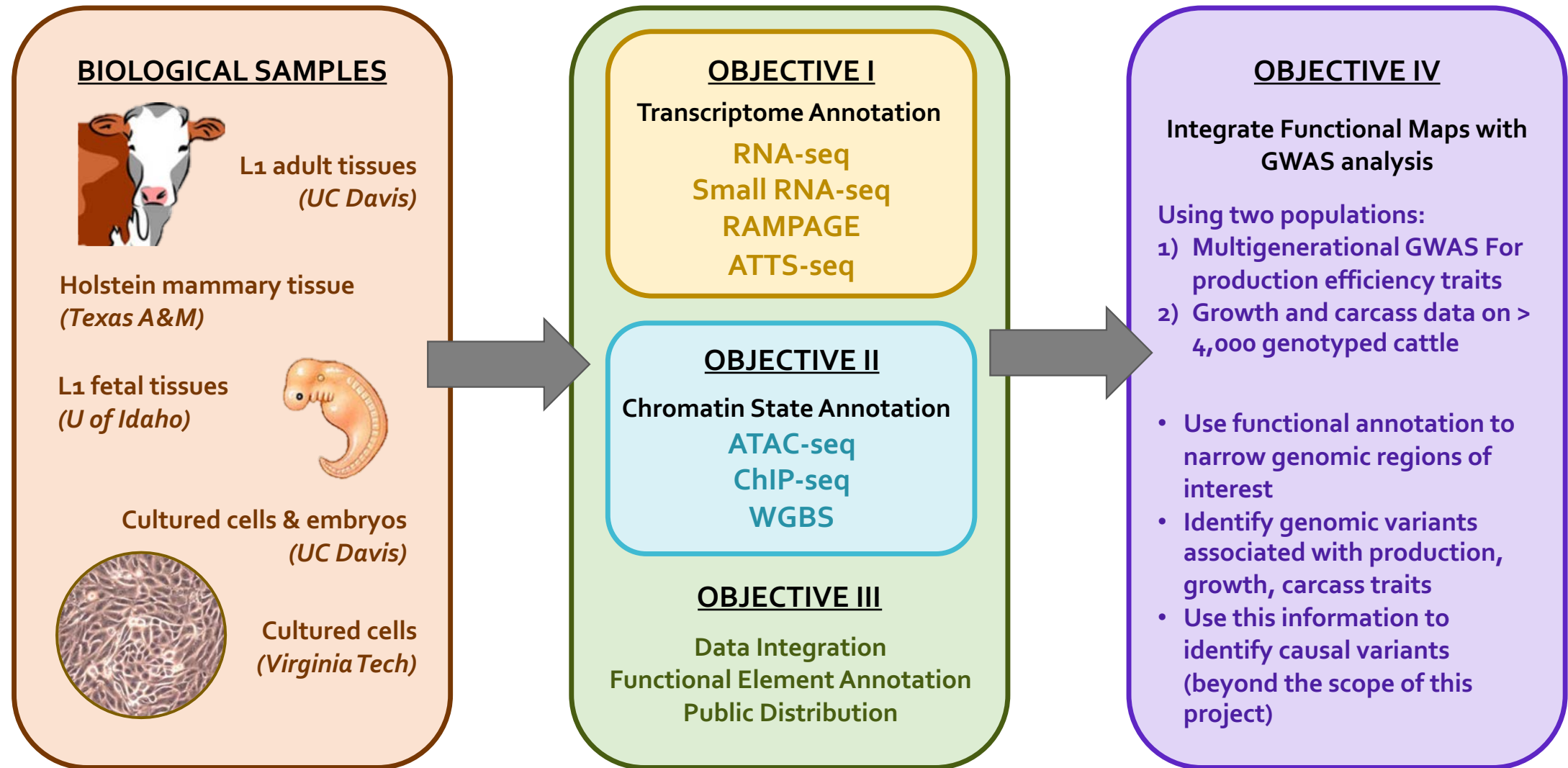
| | |
|------------------|---------------------------------|
| RNA-seq | transcripts - variants |
| smRNA-seq | Small transcript |
| RAMPAGE | Transcription start sites |
| WTTS-seq | Transcription termination sites |
| Iso-seq | Large transcripts |



Chromatin State

| | | |
|-----------------|-----------------|------------------------|
| | ATAC-seq | Open chromatin |
| CHIP-seq | H3K4me3 | Active promoters |
| | H3K27me3 | Repression |
| | H3K4me1 | Active enhancers |
| | H3K27ac | Enhancers & promoters |
| | CTCF | Insulators & promoters |
| | H3K9me3 | Heterochromatin |
| | H3K36me3 | Active gene bodies |
| | WGBS | DNA methylation |

The Functional Annotation of the Bovine Genome



Summary

- Transcriptome for 40 adult tissues from L1 Hereford line is close to completion
- Fetal tissues (3 male, 1 female) from L1 Hereford line have been collected
- Assay for primary cell lines are underway

Conclusions

1. Develop a deep and robust dataset of **transcribed** elements in the genome.
 - Sheep (60 tissues) & Cattle (40 tissues, mammary & fetal, cells)
2. Develop a deep and robust dataset of transcription **regulatory features** in the genome.
 - Sheep & Cattle
3. Annotation of the genome reference assembly.
4. Provide FAANG data and support to the public.

