Advancing Sheep Genomics Research Through Population Genetics, Genome Assembly, and the Functional Annotation of Gene Regulatory Elements

A Dissertation
Presented in Partial Fulfillment of the Requirements for the
Degree of Doctor of Philosophy
with a
Major in Animal Physiology
in the
College of Graduate Studies
University of Idaho
by
Kimberly M. Davenport

Approved by:
Major Professor: Brenda M. Murdoch, Ph.D.
Committee Members: Mark M. McGuire, Ph.D.; Paul Hohenlohe, Ph.D.; Timothy P.L. Smith, Ph.D.
Department Administrator: Robert Collier, Ph.D.

December 2021

Abstract

Sheep are a globally important species raised for meat, milk, and fiber. Research in livestock genomics, including sheep, is important to increasing disease resilience and production of animal products while decreasing environmental impact of raising animals. Sheep have adapted to many different environments across the world, which has led to specialized traits including heat tolerance, disease resilience, and increased growth and meat quality. The first study presented in this dissertation found that sheep selected to fit production systems across the world are genetically different, despite originating from similar breed lineages. This can lead to further characterization of biological traits unique to populations of animals within a species. The assembly of high-quality reference genomes also leads to a better understanding of genetic diversity and identification of genetic variation. This includes mitochondrial genomes, which have historically contributed to phylogenetic studies in mammalian species. The assembly of the mitochondrial genome of the Rocky Mountain bighorn sheep and the entire genome of the Rambouillet sheep presented in this dissertation contribute valuable reference resources to the scientific community. Further studies using these assemblies will aid in better defining and understanding the relationships between wild and domestic sheep, as well as genetic variation and locations of genes and regulatory elements in domestic sheep. The functional annotation of the sheep genome presented in the last two chapters of this dissertation defines the locations of transcriptional regulatory elements across the genome from a large collection of tissues. Histone modification, open chromatin, and DNA methylation are all classified as transcriptional regulatory elements and were defined in these studies. These regulatory elements were annotated on the Rambouillet reference genome to provide further resources to the community to investigate the mechanisms of gene regulation in relation to traits important to the sheep industry. Overall, this research will advance sheep research and production of economically important sheep products including meat, milk, and wool across the world.

Acknowledgements

I would like to thank the following people for their professional assistance and support to this dissertation and my academic career:

Dr. Brenda Murdoch, my major professor, for outstanding mentorship, opportunities to learn and network, and continued investment into my research career.

Dr. Mark McGuire for serving on my committee and providing excellent advice, insight, and encouragement throughout my postgraduate education.

Dr. Paul Hohenlohe for serving on my committee and providing thought-provoking perspectives and insights into this research.

Dr. Timothy Smith for serving on my committee and providing writing and research support as well as mentorship for these projects.

Dr. Gordon Murdoch for research advice and education, guidance, and humor.

Dr. Stephanie McKay for networking advice, research insight, and support.

Dr. Benjamin Rosen for instruction on bioinformatics and genome assembly, research support, encouragement, and hosting my USDA AGIL visit.

Dr. Kara Thornton for the kind advice and inspiration to pursue a career in research.

Dr. Patricia Villamediana for research support and assistance.

Graduate students Gabrielle Becker, Morgan Stegemiller, and Katie Shira for research support and assistance.

Undergraduate students Parker Cendejas, Abby Davis, Brooklen Walker, Dana Kerner, Taylor Badigian, Hannah Jaeger, and Rebekka Job for research assistance.

The University of Idaho Animal, Veterinary, & Food Sciences department faculty, staff, and students for support throughout my postgraduate education.

This project was supported by Agriculture and Food Research Initiative grant no. 2017-67016-26301, Hatch grant no. IDA01566 from the USDA National Institute of Food and Agriculture, USDA-ARS Project 3040-31000-100-00-D for sequencing, and USDA-ARS Project 2090-32000-36-00-D for the care and management of sheep. Financial support for presenting this research in oral and poster format were obtained from the International Society of Animal Genomics US Graduate Students and Early Career Investigator Awards (Dublin, Ireland and Lleida, Spain), University of Idaho Graduate and Professional Student Association Travel Awards, Oxford Nanopore Technologies Travel Bursary, and the AVFS department.

Dedication

This dissertation is dedicated to my grandparents, parents, and sister, who have always supported me throughout my education and career pursuits.

Table of Contents

Abstract ii

Acknowledgements iii

Dedication iv

Table of Contents v

List of Tables xi

List of Figures xii

Chapter 1: Literature Review 1

- Introduction 1
  - The Sheep Industry in the United States 1
  - Brief Overview of Sheep Genetics Research 3
- Genome Assembly 5
  - Genome Assembly Process 5
  - Genome Assemblies in Model Species and Livestock 9
  - Mitochondrial Genome Assemblies 12
- Transcriptional Regulation in Mammals 13
  - Introduction to Transcriptional Regulation 13
  - Genome Organization 16
  - CCCTC-Binding Factor (CTCF) 20
  - Histone Modifications 21
  - DNA Methylation 23
  - Functional Annotation of Animal Genomes (FAANG) 25
- Conclusion 34
- References 35

Chapter 2: Genetic Structure and Admixture in Sheep from Terminal Breeds in the United States 88

- Summary 89
- Introduction 90

Materials and methods 91

Sample collection and DNA isolation 91

Genotyping and quality control 91

Observed heterozygosity, inbreeding coefficients, and F_{ST} calculations 91

Eigenvalue analyses 92

Hierarchical clustering 92

Admixture analysis 92

International Breed Comparisons 93

Results 93

Observed Heterozygosity and Inbreeding Coefficient 93

Wright's F_{ST} 93

Eigenvalue Analyses 94

Hierarchical Clustering Based on Identity by State 94

Admixture Analysis 95

Eigenvalue Plots of U.S. and International Comparisons 95

Discussion 96

Acknowledgements 98

Availability of data 98

References 99

Tables 105

Figures 107

Chapter 3: The Complete Mitochondrial Genome Sequence of Bighorn Sheep 113

Abstract 114

Manuscript 114

Acknowledgements 115

References 116

Chapter 4: An Improved Ovine Reference Genome Assembly to Facilitate In-Depth Functional Annotation of the Sheep Genome ... 118

Abstract ... 120

Background ... 120

Findings ... 120

Conclusions ... 120

Context ... 121

Methods ... 122

Sampling Strategy ... 122

Sequencing and Data Acquisition ... 122

Assembly ... 124

Scaffolding ... 124

Gap Filling and Polishing ... 125

RNA Sequencing ... 125

Annotation ... 125

Data Validation and Quality Control ... 127

Assembly Quality Statistics ... 127

RNA sequencing alignment ... 129

Annotation ... 129

Re-use potential ... 130

Availability of supporting data ... 130

Author contributions ... 131

Acknowledgements ... 131

Funding ... 131

References ... 132

Tables ... 138

Figures ... 140

Supplementary Material 144

Chapter 5: Characterizing Genetic Regulatory Elements in Ovine Tissues 150

Abstract 151

Introduction 152

Materials & Methods 155

Sample collection 155

Chromatin immunoprecipitation 155

ChIP-seq library preparation and sequencing 156

Whole genome bisulfite sequencing library preparation and sequencing 156

ChIP-seq data quality control, mapping, and peak calling 157

DNA methylation data quality control, mapping, and methylation level characterization 158

Chromatin state and CTCF motif analysis 158

Results 159

Mapping summary and statistics 160

ChIP-seq peak calling 160

Visual assessment of sequence pileup 162

Variability in histone marks between animals 163

Principal component analysis of DNA methylation 164

Methylation level at CG and non-CG sites 164

Chromatin state assignment and correlation with methylation status 164

Distribution of chromatin states in the genome and proximity to TSS 166

Similarities and differences of chromatin states between tissues 167

CTCF binding motifs 167

Discussion 168

Acknowledgements 173

Contribution to the field 174

Funding 174

Ethics statement....175

Data availability....175

Supplementary data....175

Author contributions....175

Conflict of interest statement....176

References....177

Tables....190

Figures....195

Supplementary Material....206

Chapter 6: Defining Genetic Regulatory Elements in the Ovine Genome Provides Insight into Transcriptional Regulation Across Tissues....221

Abstract....222

Introduction....223

Materials & Methods....224

Tissue Collection....224

Chromatin Immunoprecipitation with Sequencing....224

Assay for Transposase-Accessible Chromatin with Sequencing....226

Whole Genome and Reduced Representation Bisulfite Sequencing....227

RNA Sequencing....227

Chromatin State Characterization....228

Assay Integration....229

Results....229

Hierarchical Clustering of Sequence Signal....229

ATAC-seq and ChIP-seq Peak Annotation....230

Chromatin State Characterization....231

Chromatin State Overlap with Open Chromatin....232

Chromatin State and ATAC-seq Overlap with Hypermethylated and Hypomethylated Regions 233

Chromatin State and ATAC-seq Overlap with Transcript Expression 234

Tissue Comparisons of Chromatin States 235

Discussion 237

Funding 241

Acknowledgements 241

References 243

Figures 260

Supplementary Material 281

Chapter 7: Conclusion 292

List of Tables

Table 2.1: The mean observed heterozygosity and average estimated inbreeding coefficient including the 95% confidence interval for each group……………………………………………………………105

Table 2.2: Pairwise F_{ST}* between breeds of sheep……………………………………………….106

Table 4.1: Assembly quality statistics comparison………………………………………………138

Table 4.2: RNA-seq alignment statistics to ARS-UI_Ramb_v2.0 and Oar_rambouillet_v1.0 from five different tissues…………………………………………………………………………………139

Table 5.1: Average correlations of sequencing signal between all four animals……………………190

Table 5.2: Known CTCF motifs present in the top 10 most significant motifs across multiple samples………………………………………………………………………………………………191

Table 5.3: Top three *de novo* CTCF motifs present in each sample in liver………………………..192

Table 5.4: Top three *de novo* CTCF motifs present in each sample in spleen……………………...193

Table 5.5: Top three *de novo* CTCF motifs present in each sample in cerebellum…………………194

List of Figures

Figure 1.1: An outline of the genome assembly process 6

Figure 1.2: Example of manual curation and scaffold rearrangement using the Hi-C contact map into full length chromosomes.. 7

Figure 1.3: Example of DNA looping and enhancer regulatory element contact with promoter sequence and transcription initiation complex 14

Figure 1.4: Three-dimensional organization of the genome 16

Figure 1.5: Workflow of ATAC-seq protocol 27

Figure 1.6: A comparison of open chromatin characterization methods 28

Figure 1.7: Chromatin immunoprecipitation followed by sequencing (ChIP–seq) 29

Figure 1.8: Bisulfite treatment of DNA 31

Figure 1.9: Overview of Hi-C 32

Figure 2.1: Plot of calculated Eigenvalues for breeds of U.S. sheep 107

Figure 2.2: Rectangular cladogram of individuals 108

Figure 2.3: ADMIXTURE model clustering output with K-6 populations 109

Figure 2.4: Eigenvalue plots of U.S. sheep in this study compared to other breeds across the world as part of the Sheep HapMap study 111

Figure 4.1: Image of Benz 2616 Rambouillet ewe selected for the ovine reference genome assembly 140

Figure 4.2: Hi-C contact map comparison of ARS-UI_Ramb_v2.0 141

Figure 4.3: Assembly error comparison between ARS-UI_Ramb_v2.0, Oar_rambouillet_v1.0, and Oar_v4.0 142

Figure 4.4: Dotplot comparison of genome assemblies 143

Figure 5.1: The percent of the total number of peaks normalized per Mb on each chromosome. 196

Figure 5.2: Signal of H3K4me3 ChIP-seq peaks 2 kilobases upstream and downstream of transcription start sites (TSS) identified by CAGE assays 197

Figure 5.3: Integrative Genomics Viewer (IGV) screenshot of sequence pileup normalized with the input control for active and repressive histone marks and DNA methylation 199

Figure 5.4: Principal component analysis plot based on CG methylation 200

Figure 5.5: Methylation level at CG compared to non-CG sites in liver, spleen, and cerebellum. 201

Figure 5.6: Chromatin state description and ChromHMM heatmap with histone mark signal overlap 203

Figure 5.7: Percent of the genome in liver, spleen, and cerebellum assigned to each chromatin category 204

Figure 5.8: Percent of overlapping promoter, active enhancer, poised enhancer, and repressed enhancer chromatin state categories, and hypermethylated and hypomethylated regions 205

Figure 6.1: Cluster dendrograms of raw sequence signal displaying tissue relationships................. 265

Figure 6.2: Annotation of peaks as percent of peaks overlapping with each feature 270

Figure 6.3: Chromatin states across tissues incorporating signal from all four histone modifications 271

Figure 6.4: Chromatin state overlap with open chromatin regions.. 272

Figure 6.5: Comparison of hypermethylated and hypomethylated sites with chromatin states and open chromatin regions 273

Figure 6.6: Transcript per million (TPM) counts in chromatin states and open chromatin regions .. 274

Figure 6.7: Comparison of chromatin states across tissues 278

Figure 6.8: Comparison of promoter and active enhancer chromatin states in a subset of brain tissues, GI tissues, and tissues with immune related function 279

Figure 6.9: Screenshots from the Integrative Genomics Viewer (IGV) 280

Chapter 1: Literature Review

Introduction

The Sheep Industry in the United States

Sheep are an important agricultural species used for several purposes including meat, milk, and wool in the United States and across the world. The number of sheep in the United States reached 45 million in 1867 and peaked at 51 million in 1884 (Hahn, 2020). Sheep numbers declined to approximately 5 million and operations to approximately 80,000 by 2012; however, the number of sheep operations increased to over 101,000 in 2017 (National Academy of Sciences, 2008; Hahn, 2020). This increase in the number of operations is attributed to a rise in small-scale farms that produce less than 100 sheep for lamb and wool each year (Hahn, 2020).

A recent survey conducted by the American Sheep Industry (ASI) reported that 47% of producers across the United States plan to increase the number of breeding ewes on their operations in the next five years (Miller et al., 2016). The largest flocks reside in the western United States, with approximately 20% of all operations representing almost 80% of the national breeding ewe population (Miller et al., 2016). Conversely 73% of sheep operations raise 100 ewes or fewer, and these operations reported that they planned on expanding their flocks (Miller et al., 2016). Expansion of these smaller flocks could benefit the sheep industry by increasing consumption of locally raised lamb considering that approximately 50% of the lamb consumption in the United States is imported from Australia and New Zealand (National Academy of Sciences, 2008).

The continued advancement of the U.S. sheep industry will rely on the use of genetic selection to improve production. The ASI survey reported that producers, primarily seedstock and small-scale producers, ranked genetics as a high priority for improvement of their flocks (Miller et al., 2016). Genomic selection strategies are not currently widely applied in the sheep industry when compared to other livestock industries such as the dairy cattle industry. The sheep industry is beginning to implement estimated breeding values (EBVs) calculated from reported phenotypes for selection of desired sires and replacement ewes in their flocks. The National Sheep Improvement Program (NSIP) was established in 1987 and is an organization that encourages the use of EBVs for birth weight, weaning weight, carcass traits, and breed specific production traits (Wilson and Morrical, 1991; Lupton, 2008). This has allowed producers to access EBVs, however factors including cost and sufficient record keeping limit producer implementation of this program. Genetic selection moving forward will rely on precise phenotyping and record keeping along with the use of molecular genetic tools to more accurately predict the genetic materials that is inherited in the next generation and select desirable animals earlier in life.

One trait that the sheep industry is currently selecting against scrapie susceptibility. Scrapie is a prion disease that has previously devastated the sheep industry. Mutations at codon 171 of the prion gene *PRNP* that results in an “R” amino acid has been shown to be associated with less susceptibility compared with the “Q” which has been associated with greater susceptibility (Baylis and Goldman, 2004). Genetic variation in not only codon 171, but also 154 and 136 of ovine *PRNP* have been associated with susceptibility (Baylis and Goldman, 2004). Producers have capitalized on this knowledge to select against this harmful disease as part of the National Scrapie Eradication Program (Lynn et al., 2007; Melichior et al., 2010).

The sheep industry has several single marker genetic tests for causative mutations related to disease and production traits such as spider lamb, callipyge, and fecundity available for producer use (Cockett et al., 1999; Cockett and Beever, 2001; Wilson et al., 2001; McNatty et al., 2007; Freking et al., 1998). However, single marker genomic tests are not the only molecular genetic tool available to the sheep industry and researchers. The International Sheep Genomics Consortium (ISGC) has historically led the international effort to compile various single nucleotide polymorphisms (SNPs) identified throughout the sheep genome into panels for parentage testing and genomic evaluation (Heaton et al., 2014). The available SNP panels include the Ovine 15K, three different arrays consisting of approximately 50,000 SNPs (Illumina Ovine 50K BeadChip, Affymetrix Ovine 50K, and Neogen GGP Ovine 50K), and a high-density panel (Illumina Ovine HD) consisting of approximately 600,000 SNPs. These arrays include markers for parentage, some traits of interest such as scrapie, and SNPs that evenly span the entire genome. The recently released low-density and low-cost panel (Flock54) specifically targets parentage markers and markers related to disease and production traits (Job et al., 2019; Thorne et al., 2019). This panel gives producers the opportunity to use genomic testing at a lower cost to inform selection decisions in their flocks (Job et al., 2019; Thorne et al., 2019). In summary, the sheep industry in the United States will greatly benefit from implementation of genetic selection tools to improve meat, wool, and milk production.

Brief Overview of Sheep Genetics Research

Sheep populations across the world have diverged based on adaptation and selection for traits such as wool, growth, or milk production and this has led to greater breed specialization (Kijas et al., 2012; Zhang et al., 2013). The sheep HapMap project used the 50K array to characterize genetic relatedness of sheep across the world although both the high-density and 50K SNP arrays are commonly used for genomics research (Kijas et al., 2012). This study found that many breeds of sheep retained greater heterozygosity in comparison to other livestock species under selection such as cattle (McKay et al., 2008; Bovine HapMap Consortium et al., 2009; Kijas et al., 2012; Ciani et al., 2013; Gaouar et al., 2017).

Inbreeding coefficient and Wright's F_{ST} metrics have been used in addition to observed heterozygosity to assess relatedness in sheep across the world (Wright, 1965; Weir & Cockerham, 1984; Zhang et al., 2013; Al-Mamun et al., 2015; Michailidou et al., 2018). The genetic differences based on selection for specific purposes and production systems are apparent across different geographical locations revealed by clustering based on principle component analyses and identity by state matrices (Blackburn et al., 2011; Kijas et al., 2012). Regional differences have been identified even within sheep breeds including Suffolk in the United States (Kuehn et al., 2008). It is important to understand genetic differences of sheep based on breed and geographic location because this can improve the effectiveness of genomic selection and assess the applicability of information discovered in one breed to other breeds based on relatedness.

The 50K and HD SNP panels have also been used for genome-wide association studies to identify genetic relationships to traits of interest. This includes resistance for gastrointestinal nematodes and ovine lentivirus, which both substantially affect the sheep industry (Heaton et al., 2012; Heaton et al., 2013; Becker et al., 2020; Ahbara et al., 2021). Parasite and lentivirus infections are difficult to mitigate with treatment and often these pathogens develop resistance, therefore genetic selection for resistance provides a promising prevention strategy (Heaton et al., 2012; Heaton et al., 2013; Becker et al., 2020; Ahbara et al., 2021).

The production of lamb for consumption, which relies on litter size, growth, and meat quality traits, has also been studied in a wide variety of breeds including Icelandic, Finnsheep, Romanov, Wadi, Hu, Texel, Dorper, and Merino (Zhang et al., 2013; McRae et al., 2018; Mortimer et al., 2018; Xu et al., 2018). These studies identified quantitative trait loci (QTLs) associated with economically important, valuable, or desirable traits. Among these is callipyge, a trait that is characterized by increased muscling in the hindquarters (Cockett et al., 1996; Freking et al., 2002; Murphy et al., 2005; Murphy et al., 2006; Bidwell et al., 2014; Freking et al., 2018). The SNP associated with callipyge was found to disrupt an enhancer that was paternally imprinted (Cockett et al., 1996; Freking et al., 2002; Murphy et al., 2005; Murphy et al., 2006; Bidwell et al., 2014; Freking et al., 2018). Many other SNPs associated with traits of interest lie outside transcribed regions, and may reside in regulatory regions, however these have not yet been annotated in the sheep genome.

Many traits of interest to the sheep industry including fecundity have been examined with other technologies such as whole genome sequence (WGS) (Heaton et al., 2017). This is an important tool in sheep research and provides in-depth characterization of genetic variation throughout the genome. WGS has also been used to characterize and trace selection signatures across wild and domestic sheep as well as a tool to examine the accuracy of imputation from a high-density ovine SNP panel (700K) to sequence level variation (Bolormaa et al., 2019; Yang et al., 2020). The increased availability of WGS in sheep from research groups including the United States Department of Agriculture and the International Sheep Genomics Consortium is advantageous to researchers examining traits across a diverse collection of sheep breeds.

The sheep industry in the United States will greatly benefit from the increased use of genomic technologies and advancements in genetic research. The use of SNPs and WGS have enabled major advancements in associating genetic variation to traits of interest to the industry. These technologies rely heavily on a quality reference genome to precisely map these markers and sequence, as well as gene and regulatory element annotation, to deliver the most accurate results. Reference genomes in sheep and other species will be discussed in the subsequent section of this literature review.

Genome Assembly

Genome Assembly Process

Genome assembly algorithms have greatly advanced in the last several years. There are numerous genome assembly programs available for a range of uses including Velvet and HGAP which are optimized for small genomes (Zerbino and Birney, 2008; Chin et al., 2013), Falcon and Canu which are optimized for large genome assembly with long read PacBio and/or Oxford Nanopore data (Chin et al., 2016; Koren et al., 2017), and MaSuRCA which can be used for any size genome and both long and short read data (Zimin et al., 2013). An overview of the genome assembly process is outlined in Figure 1.1. One of the most commonly used assembly programs is Canu, an updated Celera assembler that was developed for single molecule sequencing such as PacBio or Oxford Nanopore (Myers et al., 2000; Miller et al., 2008; Koren et al., 2017). This program uses the MinHash Alignment Process (MHAP) to overlap and generate consensus sequence from single-molecule sequence (Berlin et al., 2015; Koren et al., 2017). Canu operates in three stages: correct, trim, and assemble (Koren et al., 2017). Raw sequence reads (PacBio, Oxford Nanopore, or a combination of the two) are overlapped and the highest quality overlap is selected and output as a consensus (Koren et al., 2017). The low-quality reads are identified in the overlapping sequence and trimmed, retaining the highest quality consensus sequence (Koren et al., 2017). Canu then makes a final pass to identify potential sequencing errors, constructs an overlap graph, generates consensus sequences, outputs assembled contigs, and calculates summary statistics (Koren et al., 2017).

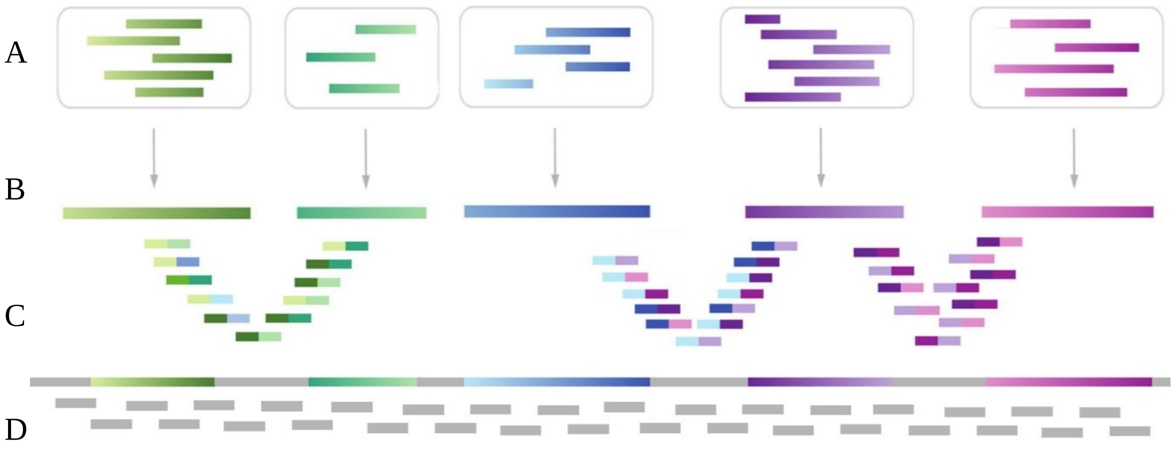


Figure 1.1: An outline of the genome assembly process. A) Whole genome sequence reads (short or long read). B) Assembly of sequence into contigs. C) Scaffolding of contigs into scaffolds. D) Gap filling and polishing of the assembly. Adapted from Giani et al., 2020.

The next step in the genome assembly process is scaffolding. There are several technologies that can aid in piecing together contigs into scaffolds including physical mapping, subcloning, and chromosome contact data (Ghurye & Pop, 2019). Physical mapping includes restriction mapping, which is implemented with enzymatic cleavage of DNA at specific recognition sites followed by sequencing to order and join contigs (Williams et al., 1992; Schwartz et al., 1993; Ghurye & Pop, 2019). Another type of physical mapping is optical mapping, which uses a restriction enzyme to digest DNA which is then stained with a fluorescent dye and joined together into a consensus optical map (Lawrence et al., 1991; Wu & Shi, 2002; Ghurye & Pop, 2019). Radiation hybrid mapping, another physical mapping technique, utilizes X-rays or radiation to randomly break DNA, which is then cloned into cell lines and sequenced (Lawrence et al., 1991; Wu & Shi, 2002; Ghurye & Pop, 2019). Subcloning is a process which involves fragmenting the genome into large pieces, transferring these DNA inserts into a vector such as a BAC, and sequencing of these fragments (Ghurye & Pop, 2019). This was first done by sequencing each end of BACs which were approximately 100kb in length (Rowen et al., 1997; Lander et al., 2001; Ghurye & Pop, 2019). The 10x Genomics technology uses this concept by partitioning large DNA fragments into droplets which are tagged with a barcode, sequenced, and divided into groups based on which DNA fragment they originated from (Mostovoy et al., 2016; Ghurye & Pop, 2019).

Additional scaffolding techniques include chromosomal contact mapping methods such as Hi-C and Chicago to utilize the three-dimensional structure of DNA to identify areas in close proximity (Lieberman-Aiden et al., 2009; Ghurye & Pop, 2019). The Hi-C protocol involves cross-linking of DNA, fragmenting with a restriction enzyme that leaves sticky ends, biotinylation and ligation of sticky ends, and sequencing of the chimeric DNA fragment (Lieberman-Aiden et al., 2009; Ghurye & Pop, 2019). The Chicago method is similar to Hi-C; however, this protocol starts with purified DNA and uses artificial nucleosomes to define contacts, which are then sequenced (Putnam et al., 2016; Ghurye & Pop, 2019). Contacts maps from both methods provide information about order and orientation of DNA sequence to inform the scaffolding process (Ghurye & Pop, 2019). A common and effective algorithm that scaffolds contigs using Hi-C information is the simple assembly scaffolder (SALSA) (Ghurye et al., 2017; Ghurye et al., 2019). The Hi-C sequence is first aligned to the genome, and then read coordinates are used to correct any contig mis-assemblies and construct scaffolds based on a scaffold graph and contact frequency between contigs (Ghurye et al., 2017; Ghurye et al., 2019).

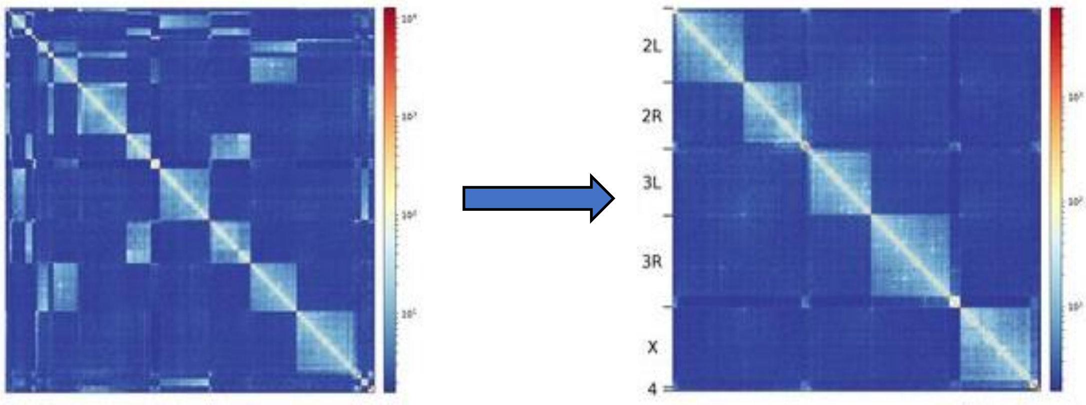


Figure 1.2: Example of manual curation and scaffold rearrangement using the Hi-C contact map into full length chromosomes. Adapted from Renschler et al., 2019.

Scaffolds are then visualized and manually curated after the scaffolding algorithm completes. The Hi-C data is mapped to the scaffolded assembly and visualized as a contact heatmap (Durand et al., 2019). The Hi-C contact information is then used to arrange and edit scaffolds (Figure 1.2) (Durand et al., 2018; Renschler et al., 2019). This step can be performed with program such as Juicer and Pretextmap (Durand et al., 2018). Scaffolded assemblies can also be compared with other genome assemblies, such as prior assemblies within the same species, by aligning the two assemblies to each other using minimap2 and visualizing similarities and differences using an interactive dotplot (Cabanettes & Klopp, 2018; Li, 2018).

The next step in the assembly process is gap filling. Gap filling may not be needed for some assemblies constructed from very long reads, such as Oxford nanopore (Heaton et al., 2021; Oppenheimer et al., 2021). Some scaffolded assemblies may have gaps that can become problematic when the assembly is released and used by the scientific community (English et al., 2012). A commonly used gap filling program is PBJelly which is designed specifically to use PacBio data (English et al., 2012). To accomplish gap filling, this program aligns PacBio WGS to the scaffolded genome and identifies reads that span gaps, and then fills these gaps in the assembly accordingly (English et al., 2012).

Polishing is the final step in the assembly process. This step aims to reduce error and deliver a more accurate assembly (Jain et al., 2018; Rhie et al., 2021). There are many polishing programs available including Nanopolish and freebayes (Jain et al., 2018; Rhie et al., 2021). Nanopolish aligns raw Oxford Nanopore fast5 sequence to the genome draft and searches for disagreements between the aligned reads and the assembly, which it then corrects based on the consensus of the raw Oxford Nanopore sequence (Jain et al., 2021). Freebayes is also used to correct errors by identifying disagreements between the draft genome and short read sequence, which generally has a greater accuracy, and subsequently corrects these errors based on the consensus of the short-read sequence (Rhie et al., 2021). The polished genome can then be evaluated for various quality metrics such as contig N50 (the length of the smallest contig where the sum of this contig and all larger contigs is equal to over 50% of the total assembly length) and L50 (the least number of contigs it takes to span half the genome), scaffold N50 (half of the scaffolds are of this size and larger) and L50 (the length of the smallest scaffold where the sum of this scaffold and all larger scaffolds is equal to over 50% of the total assembly length), per-base and kmer based quality scores, and other metrics such as mapping rate of sequence data such as RNA-seq data (Heaton et al., 2021; Oppenheimer et al., 2021).

Complete and accurate genome assemblies are important resources for the scientific community because they provide a baseline map that defines the location of genes, transcribed regions and genetic variation within the genome. The genomes assembled for model organisms as well as livestock, including reference genomes, will be discussed in the subsequent section.

Genome Assemblies in Model Species and Livestock

The precise annotation of a genome relies on an accurate reference assembly. The accuracy and completeness of a genome improves as technology rapidly advances. Genome assembly began with model organisms including *Caenorhabditis elegans* (Sulston et al., 1992), mouse (Mouse Genome Sequencing Consortium, 2002), and zebrafish (Howe et al., 2013). The large-scale project that involved sequencing the human genome for the first time began in 1990, with the first draft released in 2001 (International Human Genome Sequencing Consortium, 2001). The first human genome used Sanger technology to sequence bacterial artificial chromosome (BAC) clones generated from restriction enzyme fragmented DNA (International Human Genome Sequencing Consortium, 2001; Chial 2008). It was reported that this venture cost approximately $300 million (Etherington et al., 2020). Genome assembly technology has changed dramatically over two decades. The decrease in cost of next-generation sequencing including high throughput short read (Illumina) and long read (PacBio and Oxford Nanopore) as well as advancement of assembly algorithms has decreased cost of a simple, non-reference mammalian genome assembly to roughly $1,000 (Etherington et al., 2020).

Global efforts to sequence and assemble genomes of non-model species including livestock and wild species have followed the release of the human genome. The Genome 10K Community of Scientists aimed to collect tissue and DNA for sequencing of 10,000 vertebrate species, expecting a decrease in whole genome sequencing costs during the time the samples were collected (Genome 10K Community of Scientists, 2009). This community then evolved into the Vertebrate Genomes Project which aimed to assemble genomes from 71,657 vertebrate species (Rhie et al., 2021). A similar effort named the Earth Biogenomes Project has the goal to sequence, catalog, and characterize genomes from thousands of eukaryotes over the span of 10 years (Lewin et al., 2018). Within the agriculture sector, globally collaborative scientific communities studying agriculturally relevant livestock species including cattle, goats, and sheep assembled reference genomes for their respective species.

The sequencing and assembly of the first cattle genome set the stage for the sheep genome assembly. The first cattle genome, which was adopted as the first reference genome, was released in 2004, and later refined and published in 2009 (Liu et al, 2009). The animal that was selected to be used for this assembly due to her high level of inbreeding was a Hereford cow (L1 Dominette 01449) (Liu et al., 2009). The high level of inbreeding and reduced heterozygosity in this cow was intended to simplify genome assembly (Liu et al., 2009). The first cattle genome was assembled from BAC sequence combined with whole genome shotgun sequence (WGSS) (Liu et al., 2009). The BACs were pooled and sequenced to reduce cost, which was a similar approach to the rat and sea urchin genomes (Gibbs et al., 2004; Consortium SUGS, 2006; Liu et al., 2009). The cattle assembly was performed using the Atlas assembly system, which involves assembly of the BAC sequence alone and then in combination with the WGSS data (Liu et al., 2009). The assembled contigs and scaffolds were then placed onto chromosomes using the Integrated Bovine Map (Snelling et al., 2007; Liu et al., 2009). The Integrated Bovine Map included maps constructed from fragments of 290,797 BAC clones, linkage maps, and radiation hybrid maps (Ihara et al., 2004; Snelling et al., 2005; Everts-van der Wind et al., 2005; Itoh et al., 2005; McKay et al., 2007; Snelling et al., 2007). The reference genome assembly, Btau_3.1 was released as a resource to the scientific community. Soon after, another non-reference cattle genome, UMD2, was assembled from the same Hereford cow but used different assembly algorithms (Zimin et al., 2009). The UMD2 genome was assembled with a modified Celera assembler that was also used in the *Drosophila* genome (Zimin et al., 2009). Several years later, the reference cattle genome was updated to Btau_4.0 and Btau_5.0 as it was widely utilized by cattle researchers.

Genome assemblies have vastly improved in contiguity because of advances in sequencing technology and assembly algorithms. The most recent cattle reference genome, ARS-UCD1.2, was assembled using the same Hereford cow as the first reference (Rosen et al., 2020). This updated reference genome combined PacBio RSII long read WGS and

Illumina NextSeq500 short read WGS to achieve greater continuity with a contig N50 of over 25 Mb. T The ARS-UCD1.2 genome was assembled with the PacBio sequence using the Falcon assembler (Chin et al., 2016) and contigs were scaffolded using Dovetail Chicago (Putnam et al., 2016; Rosen et al., 2020), BtOM1.0 optical map (Zhou et al., 2015; Rosen et al., 2020) and a recombination map (Ma et al., 2015; Rosen et al., 2020). The Falcon assembler is diplotype-aware and assembles the genome by error correcting raw reads, overlapping reads together, phasing heterozygous SNPs by grouping these SNPs into haplotypes, and using these phased reads to generate contigs and haplotigs (Chin et al., 2016). The Dovetail Chicago method of scaffolding involves proximity ligation of DNA and linking these contigs into scaffolds based on three-dimensional contacts of DNA (Putnam et al., 2016). The optical map is another scaffolding technique created using a single molecule-based system that orders DNA fragments based on imaging (Zhou et al., 2015). The recombination map was created in Holstein cattle by examining linkage disequilibrium across the genome from genotypes from related animals (Ma et al., 2015).

The first cattle genome paved the way for the genome assemblies of other livestock, including sheep. The first sheep reference genome was assembled using sequence data generated from two Texel sheep (Jiang et al., 2014). This genome was assembled in multiple iterations and eventually Oar_v3.1 was released, which was a high-quality reference genome at the time (Jiang et al., 2014). The first sheep genome was assembled from WGSS (short read), BAC sequence, a radiation hybrid map, and a linkage map (Jiang et al., 2014). The contig N50 of Oar_v3.1 is approximately 40 kilobases (kb) and is 2.61 Gb in length (Jiang et al., 2014). This Texel genome was later updated using improved assembly algorithms and released as Oar_v4.0. The sheep reference was again updated in 2017 with a genome assembled from a Rambouillet ewe selected for the Functional Annotation of Animal Genomes project (Salavati et al., 2020). The Oar_rambouillet_v1.0 genome has a contig N50 of 2.57 Mb and was assembled using PacBio RSII and Illumina whole genome shotgun sequencing, followed by scaffolding with Hi-C (Salavati et al., 2020). The sheep reference genome Oar_rambouillet_v1.0 was a vast improvement from previous assemblies but has lower contiguity and quality when compared with other livestock genomes released at a similar time (Bickhart et al, 2017; Rosen et al., 2020).

There are current efforts in sheep and cattle to combine multiple genomes assembled from a several different breeds within a species to create a pan-genome. A pan-genome is essentially a reference genome that captures genetic variation across multiple individuals within a species (Sherman & Salzberg, 2020). Pan-genomes are currently being constructed and utilized in humans, crops, and bacteria (Sherman & Saslzberg, 2020; Siles et al., 2020; Coletta et al., 2021). A new strategy that is being used to create pan-genomes by assembling multiple livestock genomes involves crossing divergent breeds and collecting samples from the offspring to create haplotype-resolved genomes of each parent. This takes advantage of high heterozygosity and the ability to phase entire chromosomes back to the parental sequence, which is termed trio-binning (Koren et al., 2018; Low et al., 2020). The Angus (*Bos taurus*) and Brahman (*Bos indicus*) genome assemblies were derived from a 153-day F1 fetus and achieved contig N50s of 29.4 Mb for Angus and 23.4 Mb for Brahman (Low et al., 2020). Two additional assemblies were generated from a hybrid animal which was a cross between a yak female (*Bos grunniens*) and a Highland cattle male (*Bos taurus*)(Rice et al., 2020). These assemblies achieved impressive contig N50s of 79.8 Mb for yak and 72.8 Mb for Highland (Rice et al., 2020). Additional genomes were assembled from an American bison (*Bison bison*) and Simmental cattle (*Bos taurus*) cross, resulting in contig N50s of 70.8 Mb for the Simmental female and 68.5 Mb for the bison male (Heaton et al., 2021; Oppenheimer et al, 2021). These genomes strongly support the benefit of crossing divergent species to assemble haplotype-resolved genomes.

Mitochondrial Genome Assemblies

Several mitochondrial genomes have been completed for many different species in addition to full genome assemblies. Mitochondrial genomes are often included with the entire assembly; however, some mitochondrial genomes have been assembled separately since they can be informative in population genetics analyses and discerning phylogeny (Avise et al., 1987; Moritz, 1994; Moore, 1995). Several mitochondrial genomes exist for domestic sheep (*Ovis aries*) including the Texel and Rambouillet breeds (Hu & Gao, 2014; Salavati et al., 2020). These genomes are circular, over 16,600bp in length, and include 13 protein-coding genes, 22 transfer RNA genes, 2 ribosomal RNA genes, and a control or D-loop region (Hu

& Gao 2014). Other breeds of domestic sheep including Hamdani and Karadi breeds from Kurdistan, as well as Altay, Shandong large-tailed, small-tailed Hulun Buir, and Alpine Merono breeds from China have complete mitochondrial genomes (Fan et al., 2015; Mustafa et al., 2018; Qiao et al., 2020). Mitochondrial genomes of other *Ovis* species have also been assembled, which provides insight into phylogenetic differences between populations of wild sheep and their relationships to domestic sheep. This includes bighorn sheep (*Ovis canadensis*), snow sheep (*Ovis nivicola*), thinhorn sheep (*Ovis dalli*), and Turkish wild sheep (*Ovis gmelinii anatolica*) (Bunch et al., 2006; Miller et al., 2012; Demirci et al., 2013; Dotsev et al., 2019).

The genomes of livestock and other species have vastly improved in continuity and quality as sequencing technology and assembly algorithms improve. The cost of assembling these genomes has drastically decreased since the first human genome was released. The continued progress in technology will allow even more improvement in genome assembly in the future, as well as the construction of pan-genomes that incorporate genetic diversity from many different breeds and lineages.

Transcriptional Regulation in Mammals

Introduction to Transcriptional Regulation

The regulation of gene expression in mammals is essential for proper biological functionality, as well as tissue and cellular identity. Precise gene regulation is important for daily functionality and responses to environmental stimuli, which affects the expression of phenotypes such as growth and wool production important to the sheep industry. Transcriptional regulation is modulated by the binding of proteins to specific regulatory sequences that impact the activity of RNA polymerase. Regulatory elements including promoters and enhancers act in cis, or within proximity of neighboring elements they are acting upon. Promoter elements generally contain bindings sites for transcription factors, such as the TATA box and Inr sequence (Cooper, 2000; Calo & Wysocka, 2013; Andersson & Sandelin, 2020).

Enhancer regions are generally farther away from the transcription start site of a gene and outnumber protein coding genes, with over a million enhancers estimated to exist in the human genome (Cooper, 2000; Tippens et al., 2018). This demonstrates the complexity of gene regulation, with many enhancers having the capacity to regulate a single gene (Tippens et al., 2018). Enhancers act in cis and may be up to a megabase from their target gene in linear space (Tippens et al., 2018). The proximity in three-dimensional space is thought to facilitate enhancer and promoter interaction, and therefore gene regulation (Tippens et al., 2018). Enhancers act similarly to promoters because they often allow the binding of transcription factors which then interact with RNA polymerase (Cooper, 2000; Calo & Wysocka, 2013; Andersson & Sandelin, 2020). A depiction of this three-dimensional interaction is displayed in Figure 1.3. Genetic variation in these regulatory sequences can impact transcription factor binding and therefore transcriptional regulation (Cooper, 2000; Calo & Wysocka, 2013; Andersson & Sandelin, 2020).

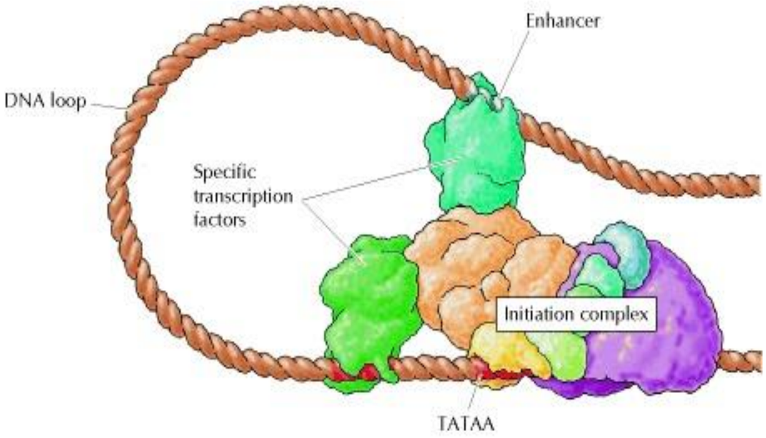


Figure 1.3: Example of DNA looping and enhancer regulatory element contact with promoter sequence and transcription initiation complex. Adapted from https://slideplayer.com/slide/7463417/.

The transcription factors that bind to regulatory sequences are an instrumental component of gene regulation. Transcriptional activators are the most well-studied group of transcription factors (Cooper, 2000). These proteins have two major domains, one that binds the target DNA sequence and one that interacts with the transcriptional machinery (Cooper, 2000; Calo & Wysocka, 2013; Andersson & Sandelin, 2020). There are four major families of DNA binding domains including zinc fingers, helix-turn-helix, leucine zipper, and helix-loop-helix (Cooper, 2000; Calo & Wysocka, 2013; Andersson & Sandelin, 2020). The zinc finger domains consist of zinc ions bound to alpha helix and beta sheet loops that directly interact with the DNA (Cooper, 2000; Calo & Wysocka, 2013; Andersson &

Sandelin, 2020). Helix-turn-helix domains consist of two or more coiled regions in which one region has direct contact with the DNA and the other regions stabilize the contact (Cooper, 2000; Calo & Wysocka, 2013; Andersson & Sandelin, 2020). Leucine zipper domains are comprised of polypeptide chains with hydrophobic areas that interact with each other and form a DNA binding helix (Cooper, 2000; Calo & Wysocka, 2013; Andersson & Sandelin, 2020). Helix-loop-helix domains are similar to leucine zippers except the protein dimerization has two helical regions that encompass the DNA instead of one (Cooper, 2000; Calo & Wysocka, 2013; Andersson & Sandelin, 2020).

Repressors also play a large role in transcriptional regulation in mammals. Repressors bind to promoter regions and throughout the genome and have domains that inhibit transcription by interacting with transcription factors (Calo & Wysocka, 2013; Andersson & Sandelin, 2020). Repressors play a key role in transcriptional regulation and are thought to be involved with tissue-specific expression, although they are not as historically well studied as activators (Cooper, 2000; Andersson & Sandelin, 2020).

Methylation of DNA at promoter sites can also repress transcription, although this is not always true (Andersson & Sandelin, 2020). Methylation of specific amino acids that are part of histone tails also can indicate repressed regions (Cooper, 2000; Calo & Wysocka, 2013; Andersson & Sandelin, 2020). Tight chromatin packaging is another method of repressing transcription, leaving the transcription start site inaccessible to RNA polymerase (Calo & Wysocka, 2013; Andersson & Sandelin, 2020). This packaging can be related to histone modifications that dictate histone compaction, and therefore silencing of transcription (Cooper, 2000; Andersson & Sandelin, 2020). Both DNA methylation and histone modification will be discussed in greater detail in a subsequent section of this review. Genome organization and packaging influences transcriptional repression as well as activation.

Genome Organization

Genomes are often visualized *in silico* in a linear manner, however within the nucleus, DNA has a deliberate three-dimensional organization. DNA has a hierarchical structure, with intentional folds and higher order organization that influences gene activity. The positioning of genes within this organization can influence transcription, with heterochromatic and repressed regions present at the outer portion of the nucleus, and more active genes present near the interior of the nucleus (Szabo et al., 2019; Bickmore, 2013). A visual depiction of higher order genome organization within the nucleus is displayed in Figure 1.4 (Szabo et al., 2019). Individual chromosomes are organized into chromosome territories, which represent distinct higher order packaging within the nucleus (Szabo et al., 2019; Cremer and Cremer, 2010). These territories contain A and B compartments that separate different types of chromatin (Lieberman-Aiden et al., 2009; Rao et al., 2014; Wang et al., 2016). A compartments contain active chromatin and regions that are gene rich, while B compartments contain mostly repressed chromatin (Szabo et al., 2019; Lieberman-Aiden et al., 2009; Rao et al., 2014).

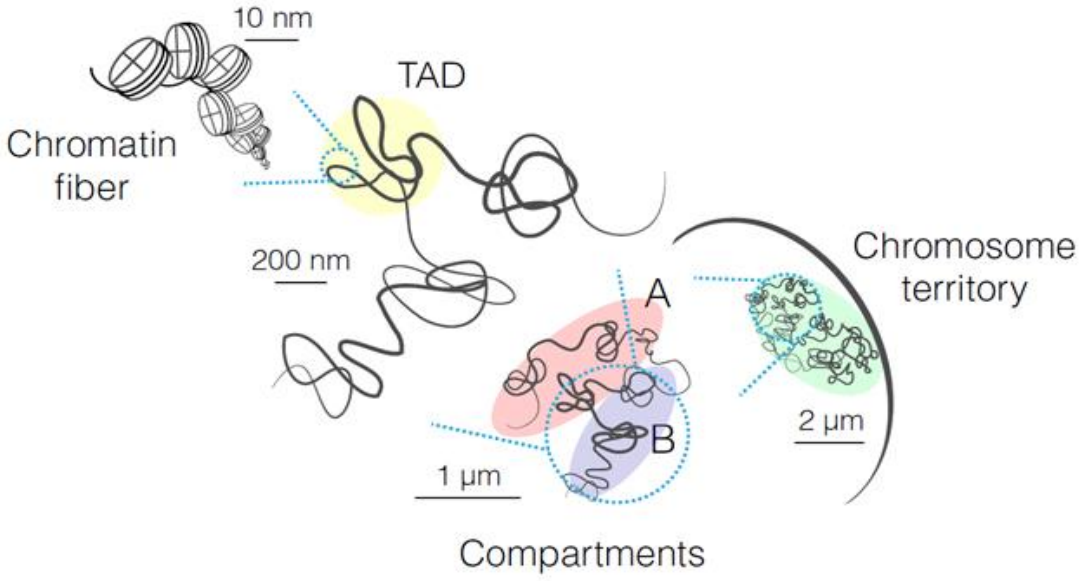


Figure 1.4: Three-dimensional organization of the genome within the nucleus into chromosome territories, A and B compartments, and topologically associated domains. This organization is instrumental in facilitating gene regulation. Adapted from Szabo et al., 2019.

Topologically associated domains (TADs) consist of genomic regions within chromatin compartments that are looped in close physical proximity in three-dimensional space, which facilitates intradomain interactions of genes and regulatory elements (Szabo et al., 2019; Nora et al., 2012). The size of these domains varies anywhere from 10 kilobases (kb) to hundreds of kb, with a median size in mice of 880 kb (Dixon et al., 2012; Hou et al., 2012; Nora et al., 2012; Sexton et al., 2012). Gene regulation and regulatory landscapes across the genome through three-dimensional organization can be influenced by TADs (Symmons et al., 2014; Bonev et al., 2017; Szabo et al., 2019) This facilitates contact between genes and regulatory elements such as promoters and enhancers located within the same TAD (Symmons et al., 2014; Bonev et al., 2017; Szabo et al., 2019). The influence of enhancers and promoters on gene expression is normally restricted within the same TAD (Bonev et al., 2017; Szabo et al., 2019). Smaller domains within TADs, called sub-TADs, are also believed to influence transcriptional regulation, and have a median size of 185 kb in the mouse genome (Rao et al., 2014; Rowley et al., 2017). Studies in mammalian species have shown that TADs are highly conserved (Dixon et al., 2012; Vietri Rudan et al., 2015; Szabo et al., 2019). Disruption of TADs can lead to mis-regulation of genes by facilitating non-intentional regulatory element contacts with genes, which may cause mis-expression of genes and lead to complications during development or even cancer (Lupianez et al., 2015; Flavahan et al., 2016; Franke et al., 2016; Hnisz et al., 2016; Lupianez et al., 2016; Weischenfeldt et al., 2017; Szabo et al., 2019).

The boundaries of TADs are characterized by the presence of CCCTC-binding factor (CTCF) and the structural maintenance of chromosomes (SMC) cohesin complex. CTCF plays an essential role in defining the boundaries of TADs. The removal or change in the CTCF binding site can shift TAD boundaries or even dismantle them completely (Sanborn et al., 2015; Lupianez et al., 2015; de Wit et al., 2015; Guo et al., 2015; Szabo et al., 2019). Approximately 75-95% of TAD boundaries in mice are associated with CTCF, depending on the cell type (Dixon et al., 2012; Bonev et al., 2017; Szabo et al., 2019). Approximately 5-25% of TAD boundaries are independent of CTCF (Nora et al., 2017; Szabo et al., 2019). These CTCF-absent boundaries have been found to separate A and B compartments within a chromosome territory (Dixon et al., 2012; Rao et al., 2014; Bonev et al., 2017; Rowley et al., 2017; Szabo et al., 2019).

The stability of TADs is also influenced by the SMC cohesin complex, which lies in the interior of the TAD (Sanborn et al., 2015; Fundenberg et al., 2016; Szabo et al., 2019). Most TAD domains were experimentally disrupted by the absence or depletion of CTCF, cohesin, or the cohesin loading factor Nipbl in mice (Nora et al., 2017; Rao et al., 2017; Schwarzer et al., 2017; Wutz et al., 2017; Szabo et al., 2019). The presence of cohesin on chromatin stabilizes TADs (Nora et al., 2017; Rao et al., 2017; Schwarzer et al., 2017; Wutz et al., 2017; Haarhuis et al., 2017; Nuebler et al., 2018; Szabo et al., 2019). Super resolution microscopy suggests that cohesin is required for cell type specific positioning of TADs (Szabo et al., 2019).

The formation of TADs is proposed to occur by the "loop extrusion method," which involves chromatin being extruded by the SMC complex until the chromatin either reaches two bound, convergent CTCF sites or the cohesin complex itself dissociates (Fundenberg et al., 2017; Szabo et al., 2019). The condensin II complex has also been observed to play a role in this loop extrusion process (Ganji et al., 2018; Szabo et al., 2019). The interaction between the cohesin and condensin complexes in this process has yet to be investigated. Interestingly, when cohesin is removed, chromatin still forms TADs through the loop extrusion method, however these loops are not positioned at CTCF boundaries (Bintu et al., 2018; Szabo et al., 2019).

CTCF is not only present at TAD boundaries, but it is also present within TADs. Enhancer-promoter pairs within TADs form smaller loop domains (sub-TADs) which are also defined by CTCF and cohesin, along with YYI (Rao et al., 2014; Phillips-Cremins et al., 2013; Szabo et al., 2019). The YYI protein is thought to contribute to cell type specific promoter and enhancer interactions with genes inside the sub-TAD (Weintraub et al., 2018; Szabo et al., 2019). CTCF acts as an insulator, which prevents ectopic regulatory element interaction both across TAD and sub-TAD boundaries. This process is discussed in further detail in a subsequent section of this review.

Domains that resemble TADs have been identified across other eukaryotes as well as prokaryotes. Four types of TADs exist according to the specific chromatin function and signature, including active domains, polycomb-repressed domains, domains devoid of chromatin and histone modifications, and heterochromatin domains in *Drosophila* (Sexton et al., 2012; Szabo et al., 2019). The *Caenorhabditis elegans* species exhibits defined domains with internal interactions were found only on the X chromosome (Crane et al., 2015; Szabo et al., 2019). Complete TAD formation and partitioning of the genome was not found in plants such as *Arabidopsis thaliana*, however boundary-like regions that contain regulatory compartments similar in role to those of *Drosophila* were identified in other species including maize, tomato, sorghum, foxtail millet, and rice (Grob et al., 2014; Wang et al., 2018; Dong et al., 2017; Szabo et al., 2019). Boundary-like regions termed chromosomal interaction domains were also found in bacteria, first in *Caulobacter crescentus* (Le et al., 2013; Szabo et al., 2019). Similar domains have also been identified in other species of bacteria including *Escherichia coli* and *Mycoplasma pneumoniae* (Espeli et al., 2008; Szabo et al., 2019). This demonstrates that higher order chromatin organization is not unique to mammals and is thought to influence gene regulation in many different species.

These higher order chromatin domains contain chromatin, which consists of DNA wrapped around nucleosomes. This basic level of genome organization was discovered in 1884 by Albrecht Kossel (Chen et al., 2014; Ramazi et al., 2020). Nucleosomes are histone octamers comprised of two of each H2A, H2B, H3, and H4 histone proteins (Cao and Yan 2012; Zhang et al., 2015; Maleszewska et al., 2016). Approximately 147 base pairs of DNA are wrapped around each nucleosome, and nucleosomes are joined by linker DNA and histone H1 to form chromatin (Crane-Robinson et al., 1997; Zhang et al., 2015; Ramazi et al., 2020). These histone proteins have N- and C- terminal tails which protrude from the nucleosome and can be post translationally modified (Zhang et al., 2015).

Post-translational modifications of histone tails can regulate chromatin structure, act to recruit proteins such as transcription factors to chromatin, and mark transcriptionally active or repressed chromatin (Zhang et al., 2015). Histone modifications can play an essential role in regulating the condensation of chromatin into accessible and inaccessible states (Peterson and Laniel, 2004; Sawan and Herceg 2010; Ramazi et al., 2020). The most abundant and studied histone modifications include methylation, acetylation, phosphorylation, and ubiquitylation, although others have been reported (Arnauda and Garcia, 2013; Zhang et al., 2015). Histone modifications and their function as regulators of gene expression will be discussed in greater detail in subsequent sections of this review.

CCCTC-Binding Factor (CTCF)

CTCF is known to have several different roles in addition to marking the boundaries of TADs. It was first discovered to be a transcriptional repressor of *c-MYC*, a zinc finger known to play a role in cell cycle progression and apoptosis (Lee & Iyer et al., 2012). The zinc finger, *c-MYC* is expressed in many different tissues and the sequence is conserved across vertebrate species (Klenova et al., 1993; Filippova et al., 1996; Lee & Iyer et al., 2012). A paralog of CTCF in testicular tissue, *BORIS*, is expressed and hypothesized to be involved in resetting epigenetic marks in the germline after erasure (Loukinov et al., 2002; Lee & Iyer et al., 2012). CTCF has also been shown to be involved with X chromosome inactivation in mammalian species (Chao et al., 2002; Lee & Iyer, 2012). Overexpression of CTCF in cell lines caused cell cycle arrest and apoptosis (Torrano et al., 2005; Qi et al., 2003; Lee & Iyer, 2012). A knockdown of CTCF led to cell proliferation and inhibited cell differentiation (Torrano et al., 2005; Qi et al., 2003; Lee & Iyer, 2012).

One function of CTCF is to block communication between promoters and enhancers to prevent inappropriate interactions, known as the “insulator function” (Lee & Iyer, 2012). This can occur at boundaries of TADs to prevent cross talk between loop domains, as well as at specific loci that are not meant to interact with each other. CTCF acts as an insulator and blocks enhancers upstream and downstream from interacting with the β-globin locus in mice and humans (Bell et al., 1999; Ristimaki et al., 1991; Lee & Iyer, 2012).

CTCF has also demonstrated involvement in genomic imprinting. CTCF acts as an insulator in mice by blocking communication between promoters and enhancers at the *Igf2* and *H19* locus (Bell et al., 2000; Lee & Iyer, 2012). Differences in DNA methylation between alleles exist in the imprinting control regions (ICR) of these two genes based on parental origin (Bell et al., 2000; Szabo et al., 2000; Kanduri et al., 2000; Lee & Iyer, 2012). CTCF binding is inhibited when the paternally inherited ICR is hypermethylated, which silences *H19* expression and indirectly activates transcription of *Igf2* (Bell et al., 2000; Szabo et al., 2000; Kanduri et al., 2000; Lee & Iyer, 2012). The hypomethylated ICR inherited maternally is permissive to CTCF binding, which activates *H19* and insulates the *Igf2* promoter from its distal enhancer, leading to repression of *Igf2* (Bell et al., 2000; Szabo et al., 2000; Kanduri et al., 2000; Lee & Iyer, 2012).

Histone Modifications

The chemical modifications of histone tails play an essential role in chromatin accessibility and gene regulation in eukaryotes. The first report of histone modifications having an inhibitory role in RNA synthesis was reported in 1951, and today we understand more about how this occurs (Allfrey et al.,1964; Ramazi et al., 2020). These modifications have been found to influence gene expression, DNA replication, DNA damage response and repair (Banerjee and Chakravarti, 2011; Khan et al., 2015; Kaimori et al., 2016; Ramazi et al., 2020), DNA condensation during mitosis and meiosis, chromatin packaging, cell cycle control, protein-protein interactions, and protein functions (Arnaudo and Garcia, 2013; Duan and Walther 2015; Shortreed et al., 2015; Ramazi et al., 2020). These modifications are added or removed by enzymes termed ‘writers’ and ‘erasers’ (Fan et al., 2015; Sadakierska-Chudy and Filip, 2015; Sabari et al., 2017; Ramazi et al., 2020).

The most abundant post-translational histone modifications are methylation, acetylation, phosphorylation, and ubiquitylation, however SUMOylation, ADP-ribosylation, deamination, arginine citrullination, N-formylation, crotonylation, propionylation, butyrylation, proline and aspartic acid isomerization, and biotinylation have also been observed (Hassan and Zempleni 2008; Wood et al., 2009; Sawan and Hercet 2010; Tan et al., 2011; Sadakierska-Chudy and Filip, 2015; Ramazi et al., 2020). Most histone modifications occur on the N-terminal of the tails; however, some modifications have been reported within the histone core (Sawan and Herceg, 2010; Ramazi et al., 2020). Post-translational modifications have been detected on over 60 residues of histones (Sawan and Herceg, 2010; Ramazi et al., 2020). Acetylation and methylation of histones H3 and H4 have been shown to be related to transcriptional activation or repression of genes (Kaimori et al., 2015), while phosphorylation of the same histones during mitosis and meiosis have been linked to chromosome condensation (Banerjee and Chakravarti, 2011; Ramazi et al., 2020).

The first histone modification to be studied in-depth after its discovery in 1964 was histone acetylation (Crane-Robinson et al., 1997; Ramazi et al., 2020). It was initially linked to transcriptional activation, but later found to also be involved with gene silencing, DNA repair, and cell-cycle progression (Vendone et al., 2005; Ramazi et al., 2020). The ‘writer’ enzymes that catalyze the transfer of an acetyl group from acetyl Co-A to the lysine on the N-terminal of the histone tail are called histone acetyltransferases (HATs) (Vendone et al., 2005; Ramazi et al., 2020). The addition of the acetyl group to the histone tail causes a shift in the charge of the histone protein which results in chromatin opening, therefore the chromatin is permissive to transcription (Han et al., 2016; Ramazi et al., 2020). The ‘eraser’ enzymes that remove acetyl groups from the N-terminal lysine are called histone deacetylases (HDAC) (Li et al., 2017; Ramazi et al., 2020). Removal of acetyl groups result in the DNA wrapping more tightly around nucleosomes, which is not permissive to gene transcription (Bannister and Kouzarides, 2011; Ramazi et al., 2020). Histone acetylation is associated with transcriptional activation, however the exact mechanism(s) underlying this association has yet to be fully uncovered (Ramazi et al., 2020).

Histone methylation occurs mainly on the arginine and lysine amino acids of histone tails (Bannister and Kouzarides, 2011). The addition of one, two, or three methyl groups does not alter the charge of the histone protein unlike acetylation or phosphorylation (Bannister and Kouzarides, 2011). Histone methyltransferase enzymes such as SUV39H1 methylate N-terminal tails of histones (Bannister and Kouzarides, 2011). Most histone methyltransferases have a SET domain that is relatively conserved across vertebrate species (Ng et al., 2009; Ramazi et al., 2020). The amino acids that differ in these SET domains play an important role in enzyme function and regulation in different species (Wood and Shilatifard, 2004; Ramazi et al., 2020). Erasers of methyl groups from histone tails occurs with histone demethylase enzymes (Greer and Shi, 2012; Sadakierska-Chudy and Filip, 2015; Wesche et al., 2017; Ramazi et al., 2020).

Active chromatin domains are characterized by histone methylation at lysine 4 in addition to histone acetylation (Zhang et al., 2015). Histone modifications are often denoted by nomenclature that includes the histone, such as H3, followed by the amino acid that is modified, such as K4, and the modification, such as one methyl group (me1). Enhancer regions are found to be enriched for H3K4me1 often with H3K27ac (Creyghton et al., 2010; Zhang et al., 2015), promoters are enriched H3K4me3 (Barrera et al., 2008; Zhang et al., 2015), and gene bodies are enriched with H3K36me3 and H3K79me3 (Ng et al., 2003; Pokholok et al., 2005; Zhang et al., 2015). H3K4me3 is often present near CpG islands found at 50-70% of unmethylated vertebrate promoters, and the addition of H3K27me3 indicates a bivalent promoter (Zhang et al., 2015). Methylation of histone tails does not always indicate active regulatory domains (Zhang et al., 2015). Methylation on H3K9, H3K20, and H3K27 indicate repressed regulatory domains and heterochromatin regions (Lohrum et al., 2007; Ng et al., 2009; Zhang et al., 2015). H3K27me3 is present at repressed enhancers and is identified by polycomb group proteins (LeRoy et al., 2013; Zhang et al., 2015), and H3K9me2 and H3K9me3 is involved in heterochromatin formation including centromere and telomere regions (Barski et al., 2007; Malezewska et al., 2016; Zhang et al., 2015).

DNA Methylation

The methylation of 5-methylcytosine (5mC) is another regulator of transcription in the mammalian genome (Zemach et al., 2010; Greenberg & Bourc’his, 2019). Mammalian genomes exhibit a high level of methylation with 70-80% of all CpGs methylated compared with other eukaryotes (Li & Zhang, 2014; Greenberg & Bourc’his, 2019). DNA methylation plays a large role in genomic imprinting and repressing transposons that may be harmful (Walsh et al., 1998; Borgel et al., 2010; Arand et al., 2012; Greenberg & Bourc’his, 2019). The inactivation of the X chromosome in females also involves DNA methylation in order to silence transcription from one X chromosome (Lock et al., 1987; Grant et al., 1992). Mis-regulation of DNA methylation during developmental and adult stages of life can lead to embryonic lethality and cancer (Li et al., 1992; Okano et al., 1999; Baylin et al., 2016). Two important stages of DNA methylation reprogramming occur during embryonic development, which influences CpG methylation patterns following fertilization and also during germ cell differentiation in both males and females (Monk et al., 1987; Sanford et al., 1987; Greenberg & Bourc’his, 2019).

DNA methylation involves three different phases: establishment (*de novo* methylation), maintenance, and demethylation. Different enzymes write, maintain, and erase these methylation patterns. The writers of DNA methylation are termed DNA methyltransferases (DNMT). The most prevalent DNMT enzymes in mammals are

DMNT3A and DMNT3B which are active in somatic cells and in the germline when partnered with DNMT3L (Okano et al., 1998; Okano et al., 1999; Bourc'his et al., 2001; Ooi et al., 2007). Maintenance of DNA methylation during cell division is also aided by DNMT, mainly DMNT1 which interacts with E3 ubiquitin-protein ligase (UHRF1) to methylate daughter DNA identically to the parental DNA strand (Nishiyama et al., 2013; Qin et al., 2015; Greenberg & Bourc'his, 2019). The demethylation of DNA occurs with TET methylcytosine dioxygenases by oxidizing 5mC to 5-hydroxymethylcytosine (5hmC), 5-formylcytosine (5fC), and 5-carboxylcytosine (5caC) (Tahiliani et al., 2009; He et al., 2011; Ito et al., 2011; Greenberg & Bourc'his, 2019).

The relationship of DNA methylation to open chromatin, histone modifications, and transcription has been studied in several species. Open chromatin regions that are accessible for transcription factor binding generally have very low levels or absence of DNA methylation (Stadler et al., 2011; Greenberg & Bourc'his, 2019). In general, DNA methylation can block transcription factors binding to their specific motifs, therefore inhibiting transcription (Yin et al., 2017; Greenberg & Bourc'his, 2019). The modification H3K4me3 repels the binding of a chromatin reading domain of DNMTs, therefore deterring the methylation of CpG sites at promoters of actively transcribed genes (Zhang et al., 2010; Piunti & Shilatifard, 2016; Greenberg & Bourc'his, 2019). DNA methylation is also present in repressive polycomb regions characterized by H3K27me3, which are inversely related to gene expression (Tanay et al., 2007; Brinkman et al., 2012; Jermann et al., 2014). Heterochromatin regions are also characterized by the presence of DNA methylation (Greenberg & Bourc'his, 2019). DNA methylation is often present within actively transcribed gene bodies which are often also marked by the H3K36me3 histone modification (Lister et al., 2009; Dhayalan et al., 2010; Zemach et al., 2010; Greenberg & Bourc'his, 2019). This suggests that DNA methylation is not always acting in a repressive way (Sun et al., 2005; Dhayalan et al., 2010; Greenberg & Bourc'his, 2019). DNA methylation is involved in many different biological pathways that influence the transcription of genes.

Functional Annotation of Animal Genomes (FAANG)

The functional annotation of animal genomes project (FAANG) aims to define genetic regulatory elements in domestic species. The encyclopedia of DNA elements (ENCODE) project has completed functional annotation of regulatory elements in human (ENCODE Project Consortium, 2012), mouse (Shen et al., 2012; Yue et al., 2014), *Drosophila* (Roy et al., 2010), *Caenorhabditis elegans* (Gerstein et al., 2010), and zebrafish (Sivasubbu et al., 2013) cells and tissues. Regulatory elements have yet to be defined across domestic and farmed animal species including sheep (Barbosa-Morais et al., 2012; Andersson et al., 2015). Many complex traits that are important in livestock industries are likely not controlled by variation in coding regions alone (Andersson, 2013; Schaub et al., 2012; Andersson et al., 2015). A variant in an enhancer region, which is also subject to polar overdominance, is the causal mutation for callipyge in sheep (Cockett et al., 1996; Freking et al., 2002; Bidwell et al., 2014). Myostatin protein expression is decreased by a new microRNA binding site caused by a single nucleotide change in the 3' untranslated region of the sheep myostatin gene (Clop et al., 2006; Andersson et al., 2015). Growth in pigs is influenced by a variant in an intron of *IGF2*, which also happens to be a regulatory region (Van Laere et al., 2003; Andersson et al., 2015). The functional annotation of domestic and farm animal genomes will provide further insight into the biological mechanisms behind complex, important traits.

Gene transcription and regulation can be characterized by defining transcribed loci, chromatin accessibility and architecture, histone modifications, DNA methylation, transcription factor binding sites, and genome conformation. To characterize transcribed loci, sequencing information is obtained from RNAs which include messenger RNA (mRNA), micro-RNA (miRNA), and long non-coding RNA (lncRNA). Sequencing different types of RNA molecules may help to identify novel transcripts, splice sites, coding regions, and allele-specific expression (Kukurba and Montgomery, 2015). There are many different types of non-coding RNA (ncRNA) that can regulate gene expression (Kukurba and Montgomery, 2015). The non-coding RNAs that were first discovered include ribosomal RNA (rRNA), transfer RNA (tRNA) involved with translation of mRNA, small nuclear RNA (snRNA) involved with splicing, and small nucleolar RNA (snoRNA) involved with modification of rRNA (Mattick and Makunin, 2006; Kukurba and Montgomery, 2015). Other types of ncRNA were discovered and linked to the regulation of gene expression including micro-RNA (miRNA) and piwi-interacting RNA (piRNA), and long non-coding RNA (lncRNA) (Okazaki et al., 2002; Stefani and Slack, 2008; Kukurba and Montgomery, 2015).

The FAANG community has begun to characterize the transcriptome across several species and a large collection of tissues. A gene expression atlas was created for the sheep using RNA-seq data (Clark et al., 2017). This dataset used network clustering to identify expression specific to tissues and prenatal, neonatal, juvenile, and adult developmental stages of Texel and Scottish Blackface sheep (Clark et al., 2017). A goat gene expression atlas was created from 17 tissues using RNA-seq which allowed for identification of expression unique to specific groups of tissues and identified unannotated genes in the current reference (Muriuki et al., 2019). A gene expression atlas created for the water buffalo included three different breeds (Mediterranean, Padharpuri, and Bhadawari) and clustered genes based on expression profiles (Young et al., 2019). The pig gene expression atlas took the approach of mining RNA-seq data from public repositories and attempted to normalize and compare the data to identify tissue and cell specific transcript expression (Summers et al., 2020). Allele-specific mRNA expression was also characterized in a Texel-crossed Scottish Blackface sheep, which sets the precedent for similar experiments in the future (Salavati et al., 2019). lncRNA has been characterized in cattle, chicken, and pigs across 8 tissues and compared between species (Kern et al., 2018). An experiment in cattle from divergent residual feed intake groups examined lncRNA differences related to this important phenotype (Nolte et al., 2020).

Additional methods to capture transcription information includes cap analysis of gene expression (CAGE) and RNA annotation and mapping of promoters for analysis of gene expression (RAMPAGE). The CAGE method captures the first 20 nucleotides from the 5' end of mRNA followed by sequencing to identify transcription start sites (TSS) (Shiraki et al., 2003). The RAMPAGE method uses reverse transcribed RNA to sequence the 5' complementary strand in order to identify TSS (Batut and Gingeras, 2013). Both methods have been utilized by the FAANG community to identify TSS throughout the genome in several tissues. TSS were identified in sheep with CAGE across 56 tissues and compared with both mRNA-seq and DNA methylation data (Salavati et al., 2020). This study revealed novel, un-annotated TSS across this large collection of tissues, and found that these sites were devoid of methylation (Salavati et al., 2020). A survey of RAMPAGE sequencing in 31 tissues in cattle also identified novel TSS throughout the genome and in a tissue specific capacity (Goszczynski and Halstead et al., 2021).

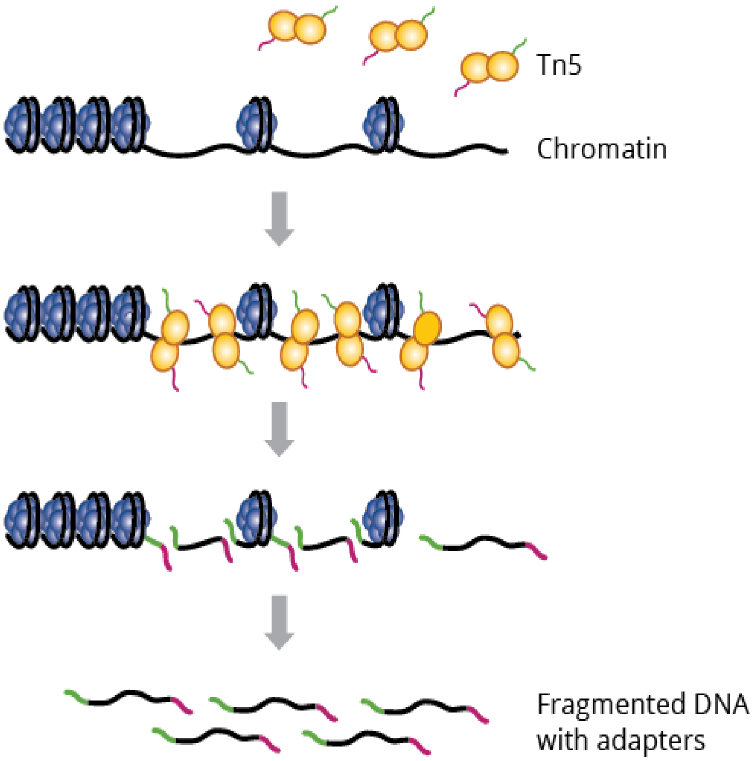


Figure 1.5: Workflow of ATAC-seq protocol involving the Tn5 enzyme cleaving accessible regions of chromatin, isolation of cleaved chromatin, and ligation of adapters prior to sequencing. Adapted from https://www.genewiz.com/Public/Services/Next-Generation-Sequencing/Epigenomics/ATAC-Seq/

Chromatin accessibility can be highly indicative of transcriptional activity throughout the genome (Yan et al., 2020). Several assays including assay of transposable accessible chromatin with sequencing (ATAC-seq) (Figure 1.6), DNase I hypersensitive sites sequencing (DNase-seq) Figure 1.6), and formaldehyde-assisted isolation of regulatory elements with sequencing (FAIRE-seq) (Figure 1.6) can all define chromatin accessibility (Buenrostro et al., 2013; Buenrostro et al., 2015; Song and Crawford, 2010; Thurman et al., 2012; Boyle et al., 2008; Giresi et al., 2007; Yan et al., 2020). The ATAC-seq assay incorporates a hyperactive Tn5 transposase to cut chromatin accessible to the enzyme and leaves a 9bp nick to ligate adaptors for sequencing (Yan et al., 2020). ATAC-seq is comparable in sensitivity and specificity to DNase-seq but generally outperforms FAIRE-seq as this method requires a much larger sample input (Buenrostro et al., 2013; Yan et al., 2020). The DNase-seq assay is similar to ATAC-seq in that it uses an enzyme, DNase I, that cleaves sites accessible and hypersensitive to the enzyme throughout the genome (Boyle et al., 2009). Both assays have been used prevalently to characterize open chromatin, but ATAC-seq has gained popularity due to the fact that it requires less sample input and is less labor intensive to perform (Li et al., 2019; Yan et al., 2020). Finally, FAIRE-seq involves crosslinking chromatin with formaldehyde, sonicating the chromatin, and then separating the "free" chromatin from the chromatin crosslinked to histones for sequencing (Giresi et al., 2007).

The FAANG community has performed some characterization of open chromatin in livestock using ATAC-seq and DNase-seq. The methodology for ATAC-seq with homogenized, cryopreserved tissue was recently published to show proof of concept in chicken lung by comparing ATAC-seq data to DNase-seq, ChIP-seq, and RNA-seq assays (Halstead et al., 2020). Both ATAC-seq and DNase-seq were integrated with different data types including ChIP-seq and RNA-seq across 8 tissues in cattle, chicken, and pig (Kern et al., 2021). This study displayed how open chromatin overlapped with other regulatory elements and transcription across the

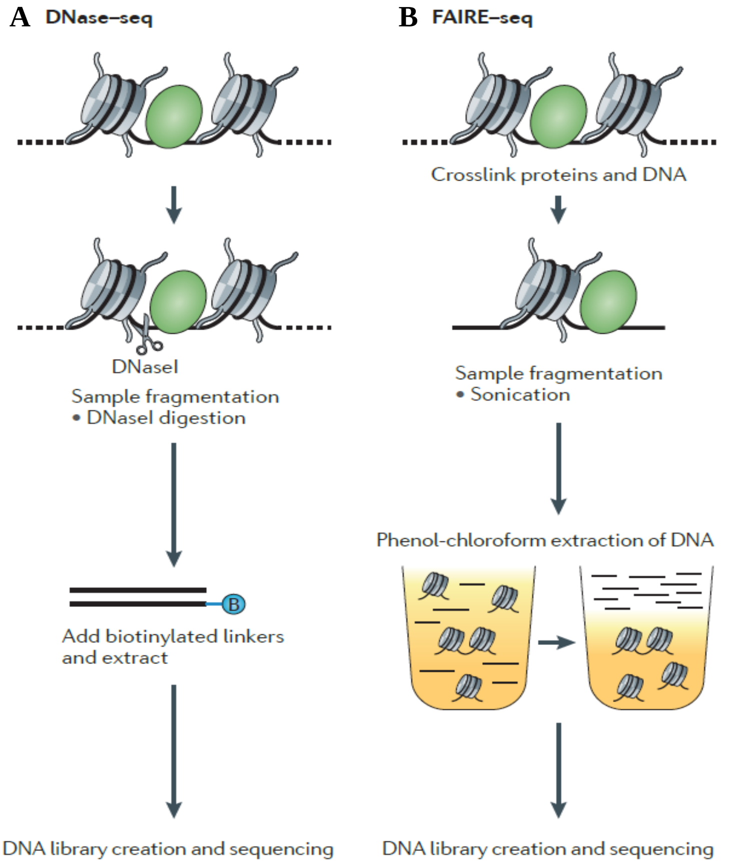

A DNase–seq
B FAIRE–seq
Crosslink proteins and DNA
DNaseI
Sample fragmentation
• DNaseI digestion
Sample fragmentation
• Sonication
Phenol-chloroform extraction of DNA
B
Add biotinylated linkers and extract
DNA library creation and sequencing
DNA library creation and sequencing

Figure 1.6: A comparison of open chromatin characterization methods. A) DNase–seq relies on digestion by the DNaseI nuclease to identify regions of nucleosome-depleted open chromatin where there are binding sites for all types of factors, but it cannot identify what specific factors are bound. B) Formaldehyde-assisted identification of regulatory elements (FAIRE–seq) similarly identifies nucleosome-depleted regions by extracting fragmented DNA that is not crosslinked to nucleosomes. Adapted from Furey, 2012.

genome and compared these regions both within and across cattle, chicken, and pig (Kern et al., 2021).

Histone modifications are commonly characterized by chromatin immunoprecipitation with sequencing (ChIP-seq) (Figure 1.7). The goal of ChIP-seq is to sequence the DNA wrapped around histones with specific modified tails or transcription factors (Landt et al., 2012). The ENCODE consortium has outlined a protocol for ChIP-seq that involves treating cells or tissues with formaldehyde to crosslink bound transcription factors or histones to DNA (Ren et al., 2000; Iyer et al., 2001; Landt et al., 2012). This step is followed by sonication or enzymatic digestion to shear chromatin into 100-300 bp fragments (Ren et al., 2000; Iyer et al., 2001; Landt et al., 2012). Chromatin is then enriched for the protein or histone modification of interest using targeted antibodies (Landt et al., 2012). Both monoclonal and polyclonal antibodies can be used depending on the goal of the project (Ma and Zhang, 2020). Monoclonal antibodies can generate the

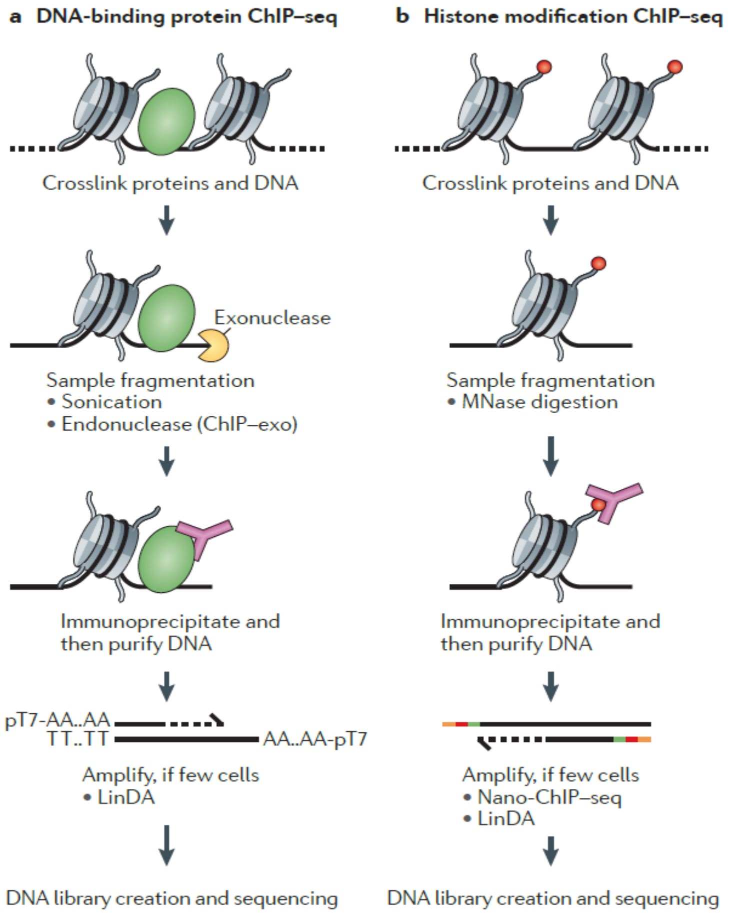

a DNA-binding protein ChIP–seq
b Histone modification ChIP–seq
Crosslink proteins and DNA
Crosslink proteins and DNA
Exonuclease
Sample fragmentation
• Sonication
• Endonuclease (ChIP–exo)
Sample fragmentation
• MNase digestion
Immunoprecipitate and then purify DNA
Immunoprecipitate and then purify DNA
pT7-AA..AA
TT..TT
AA..AA-pT7
Amplify, if few cells
• LinDA
Amplify, if few cells
• Nano-ChIP–seq
• LinDA
DNA library creation and sequencing
DNA library creation and sequencing

Figure 1.7: Chromatin immunoprecipitation followed by sequencing (ChIP–seq) for A) DNA-binding proteins such as transcription factors. Recent variations on the standard protocol include using endonuclease digestion instead of sonication (ChIP–exo) to increase the resolution of binding-site detection and to eliminate contaminating DNA, and DNA amplification after ChIP for samples with limited cells. B) ChIP–seq for histone modifications uses micrococcal nuclease (MNase) digestion to fragment DNA and can also now be run on low-quantity samples when combined with the additional post-ChIP amplification. Adapted from Furey, 2012.

most precise results from immunoprecipitation experiments if the antibody is produced for a specific epitope that exists in a particular species, while polyclonal antibodies capture a greater number of epitopes and can be used across different species (Kidder et al., 2011; Ma and Zhang, 2020). Following immunoprecipitation, the crosslinks in the chromatin are then reversed and DNA is isolated for high-throughput sequencing (Ren et al., 2000; Iyer et al., 2001; Landt et al., 2012). The ChIP-seq protocol can be performed without crosslinking the chromatin and rather relying on the innate binding strength of proteins and histones to DNA, which is termed native or natural ChIP-seq (O’neill, 2003; Kasinathan et al., 2014; Gilfillan et al., 2012; Ma and Zhang, 2020). This can result in some loss of enrichment for sites that are not bound well enough to withstand the immunoprecipitation protocol (Ma and Zhang, 2020).

Few studies have completed ChIP-seq for histone modifications and transcription factors such as CTCF in livestock species. A comparison of promoter and enhancer activity in liver defined by H3K4me3 and H3K27ac across mammals, including cattle and pigs, and found regions conserved across 20 species and regions that differed and related to evolutionary distance (Villar et al., 2015). A study examining differences in longissimus dorsi muscles of differing tenderness in cattle also utilized H3K4me3 peak enrichments as promoters to define potential variation in histone modifications related to a phenotype important to livestock industries (Zhao et al., 2015). Regulatory elements in sheep were first characterized by lifting over regions annotated by the ENCODE project in humans (Naval-Sanchez et al., 2018). The data used in this study included lifted over ChIP-seq and chromatin states for H3K4me3, H3K4me1, H3K36me3, H3K9me3, and H3K27me3 and one ChIP-seq experiment with H3K4me3, H3K27ac, and H3K27me3 in adipose tissue (Naval-Sanchez et al., 2018).

In cattle cell lines, the effect of butyrate on regulatory elements including H3K4me3, H3K4me1, H3K27ac, H3K27me3, H3K9ac, H3K9me3, and CTCF along with ATAC-seq, RNA-seq, and DNA methylation by whole genome bisulfite sequencing was examined (Fang et al., 2019; Kang et al., 2020). A study by Fang et al. found differences in these regulatory elements based on treatment with the histone deacetylase butyrate which is a prevalent volatile fatty acid produced in the stomach of ruminant species (Fang et al., 2019). In addition, a comparison of transcriptional regulation between chicken, pig, and cattle was performed using ChIP-seq of H3K4me3, H3K27ac, H3K4me1, H3K27me3, and CTCF as well as DNase-seq, ATAC-seq, and RNA-seq (Kern et al., 2021). A number of conserved and different regulatory regions were defined throughout the genome in these three species (Kern et al., 2021). Further, CTCF was used to predict TADs and define regulatory domains and target genes of promoter and enhancer regions (Kern et al., 2021). Integration of ChIP-seq with other datasets is important to defining regulatory elements and creating a better understanding of transcriptional regulation.

Characterizing DNA methylation is also important for defining regulatory elements throughout the genome. Methylation of CpG islands in promoter regions often indicate inactivity or silencing of that gene (Greenberg & Bourc’his, 2019). DNA methylation is also known to influence many phenotypes related to disease and development and is involved with genomic imprinting which was discussed in a previous section (Greenberg & Bourc’his, 2019).

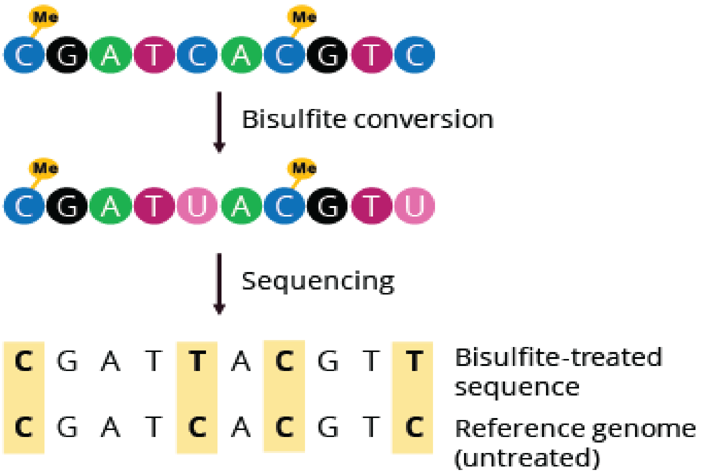


Figure 1.8: Bisulfite treatment of DNA involves conversion of unmethylated cytosine bases to uracil, followed by library preparation and sequencing. The unmethylated and methylated cytosine bases are then differentiated when the sequence is mapped to the reference genome (Adapted from https://www.genewiz.com/Public/Services/Next-Generation-Sequencing/Epigenomics/Whole-Genome-Bisulfite-Sequencing/).

Two common and effective methods to characterize DNA methylation include whole genome bisulfite sequencing (WGBS) and reduced representation bisulfite sequencing (RRBS) (Doherty and Couldrey, 2014). Bisulfite treatment is used to convert unmethylated cytosines into uracil in both WGBS and RRBS, while methylated cytosines are protected (Figure 1.8) (Frommer et al., 1992; Doherty and Couldrey, 2014). The entire genome is then sequenced, often with short-read sequencing, for WGBS while the RRBS method employs preferential selection of CpG-rich regions throughout the genome prior to sequencing (Meissner et al., 2005; Gu et al., 2011; Doherty and Couldrey, 2014). Methylation is informative independently and can complement other FAANG data from CAGE, ChIP-seq, ATAC-seq, and RNA-seq experiments (Fang et al., 2019; Salavati et al., 2020).

An additional assay that helps to define regulatory regions in the genome is Hi-C, which uses proximity ligation to characterize the three-dimensional organization and folding of the genome (Figure 1.9) (Oluwadare et al., 2019). This technology is used to not only identify topologically associated domains, but also assemble genomes by linking assembled contigs together during the scaffolding process (Oluwadare et al., 2019). This technology has been coupled with ChIP-seq to examine three-dimensional contacts of histone modifications and transcription factors (Mumbach et al., 2016; Bhattacharyya et al., 2019). This method is performed by using proximity ligation before chromatin is sheared, followed by biotinylation of sticky ends, immunoprecipitation of the biotin tagged DNA, and sequencing (Mumbach et

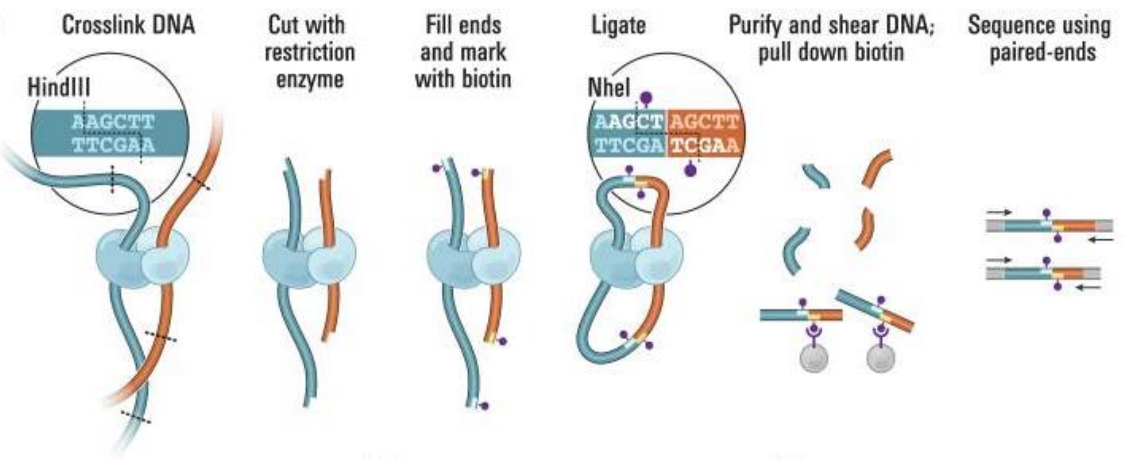

Crosslink DNA
Cut with restriction enzyme
Fill ends and mark with biotin
Ligate
Purify and shear DNA; pull down biotin
Sequence using paired-ends
HindIII
AAGCTT
TTCGAA
NheI
AAGCT AGCTT
TTCGA TCGAA

Figure 1.9: Overview of Hi-C. Cells are cross-linked with formaldehyde, resulting in covalent links between spatially adjacent chromatin segments (DNA fragments: dark blue, red; Proteins, which can mediate such interactions, are shown in light blue and cyan). Chromatin is digested with a restriction enzyme (here, HindIII; restriction site: dashed line, see inset) and the resulting sticky ends are filled in with nucleotides, one of which is biotinylated (purple dot). Ligation is performed under extremely dilute conditions to create chimeric molecules; the HindIII site is lost and a NheI site is created (inset). DNA is purified and sheared. Biotinylated junctions are isolated with streptavidin beads and identified by paired-end sequencing. Adapted from Lieberman-Aiden et al., 2009.

al., 2016; Bhattacharyya et al., 2019). This method is similar to chromatin interaction analysis by paired-end tag sequencing (ChIA-PET) except Hi-ChIP requires less input material (Mumbach et al., 2016; Bhattacharyya et al., 2019).

Other new and evolving technologies that require less input material include cleavage under targets and release using nuclease (CUT&RUN) and single cell sequencing. The CUT&RUN methodology is similar to ChIP-seq in that it requires an antibody specific to the target, such as a transcription factor or histone modification, to isolate and sequence regions of interest (Skene et al., 2018; Hainer & Fazzio, 2019). However, with CUT&RUN, the micrococcal nuclease is tethered to a recombinant protein A or G which recognizes the antibody, binds, and enzymatically cleaves the target region (Skene et al., 2018; Hainer & Fazzio, 2019). The small chromatin fragments are then separated and purified for sequencing (Skene et al., 2018; Hainer & Fazzio, 2019). This requires less input and can be performed at a single-cell level after sorting and separation of individual cells (Skene et al., 2018; Hainer & Fazzio, 2019). Other technologies that have been performed and sequenced at a single-cell level include RNA-seq and ATAC-seq (Buenrostro et al., 2015; Haque et al., 2017; Cusanovich et al., 2018; Stuart et al., 2019). This allows for precise separation of cell types which provides a snapshot of transcriptional and regulatory profiles in the many cell types that compose a tissue (Buenrostro et al., 2015; Haque et al., 2017; Cusanovich et al., 2018; Stuart et al., 2019).

Defining the location of regulatory elements will lead to a better understanding of the factors that influence transcriptional activity throughout the genome. Genetic variation within these regions is also important to characterize, as it can influence the regulation of transcription and affect a particular phenotype. Defining these regulatory regions and how variation influences regulatory mechanisms will lead to a better understanding of complex traits that are important to the sheep industry.

Conclusion

Sheep provide meat, milk, and wool to humans on a global scale. This important agricultural species has been selected for different production systems across the world and the genetic variation in sheep populations reflect this. Sheep have unique biological traits and features, and many research efforts aim to characterize the mechanisms that influence these traits. The use of genetics in sheep production is being integrated slowly into management practice. Sheep genetics research, however, is making vast progress as technology advances, with the end goal of helping the sheep industry in the United States and across the world select animals to best fit production goals. Advances such as more accurate genomes and functional annotation of regulatory regions will help uncover the biology behind many complex traits and gain a more precise understanding of the sheep species as well as mammals.

References

1. Ahbara AM, Rouatbi M, Gharbi M, Rekik M, Haile A, Rischkowsky B, Mwacharo JM. Genome-wide insights on gastrointestinal nematode resistance in autochthonous Tunisian sheep. Sci Rep. 2021 Apr 29;11(1):9250. doi: 10.1038/s41598-021-88501-3.

2. Al-Mamun HA, Clark SA, Kwan P, Gondro C. Genome-wide linkage disequilibrium and genetic diversity in five populations of Australian domestic sheep. Genet Sel Evol. 2015 Nov 24;47:90. doi: 10.1186/s12711-015-0169-6.

3. Andersson R, Sandelin A. Determinants of enhancer and promoter activities of regulatory elements. Nat Rev Genet. 2020 Feb;21(2):71-87. doi: 10.1038/s41576-019-0173-8.

4. Andersson L, Archibald AL, Bottema CD, Brauning R, Burgess SC, Burt DW, Casas E, Cheng HH, Clarke L, Couldrey C, Dalrymple BP, Elsik CG, Foissac S, Giuffra E, Groenen MA, Hayes BJ, Huang LS, Khatib H, Kijas JW, Kim H, Lunney JK, McCarthy FM, McEwan JC, Moore S, Nanduri B, Notredame C, Palti Y, Plastow GS, Reecy JM, Rohrer GA, Sarropoulou E, Schmidt CJ, Silverstein J, Tellam RL, Tixier-Boichard M, Tosser-Klopp G, Tuggle CK, Vilkki J, White SN, Zhao S, Zhou H; FAANG Consortium. Coordinated international action to accelerate genome-to-phenome with FAANG, the Functional Annotation of Animal Genomes project. Genome Biol. 2015 Mar 25;16(1):57.

5. Andersson L. Molecular consequences of animal breeding. Curr Opin Genet Dev. 2013 Jun;23(3):295-301. doi: 10.1016/j.gde.2013.02.014. Epub 2013 Apr 16. PMID: 23601626.

6. Arand J, Spieler D, Karius T, Branco MR, Meilinger D, Meissner A, Jenuwein T, Xu G, Leonhardt H, Wolf V, Walter J. In vivo control of CpG and non-CpG DNA

methylation by DNA methyltransferases. PLoS Genet. 2012 Jun;8(6):e1002750. doi: 10.1371/journal.pgen.1002750.

7. Arnaudo AM, Garcia BA. Proteomic characterization of novel histone post-translational modifications. Epigenetics Chromatin. 2013 Aug 1;6(1):24. doi: 10.1186/1756-8935-6-24.

8. Avise JC, Arnold J, Ball RM, Bermingham E, Lamb T, Neigel JE, Reeb CA, Saunders NC. Intraspecific phylogeography: the mitochondrial DNA bridge between population genetics and systematics. Annu Rev Ecol Syst. 1987;18:489–522. doi:10.1146/annurev.es.18.110187.002421.

9. Barbosa-Morais NL, Irimia M, Pan Q, Xiong HY, Gueroussov S, Lee LJ, Slobodeniuc V, Kutter C, Watt S, Colak R, Kim T, Misquitta-Ali CM, Wilson MD, Kim PM, Odom DT, Frey BJ, Blencowe BJ. The evolutionary landscape of alternative splicing in vertebrate species. Science. 2012 Dec 21;338(6114):1587-93. doi: 10.1126/science.1230612.

10. Barrera LO, Li Z, Smith AD, Arden KC, Cavenee WK, Zhang MQ, Green RD, Ren B. Genome-wide mapping and analysis of active promoters in mouse embryonic stem cells and adult organs. Genome Res. 2008 Jan;18(1):46-59. doi: 10.1101/gr.6654808. Epub 2007 Nov 27.

11. Bartolomei MS, Webber AL, Brunkow ME, Tilghman SM. Epigenetic mechanisms underlying the imprinting of the mouse H19 gene. Genes Dev. 1993 Sep;7(9):1663-73. doi: 10.1101/gad.7.9.1663. PMID: 7690336.

12. Batut P, Gingeras TR. RAMPAGE: promoter activity profiling by paired-end sequencing of 5'-complete cDNAs. Curr Protoc Mol Biol. 2013 Nov 11;104:Unit 25B.11. doi: 10.1002/0471142727.mb25b11s104.

13. Baylin SB, Jones PA. Epigenetic Determinants of Cancer. Cold Spring Harb Perspect Biol. 2016 Sep 1;8(9):a019505. doi: 10.1101/cshperspect.a019505.

14. Baylis M, Goldmann W. The genetics of scrapie in sheep and goats. Curr Mol Med. 2004 Jun;4(4):385-96. doi: 10.2174/1566524043360672. PMID: 15354869.

15. Becker GM, Davenport KM, Burke JM, Lewis RM, Miller JE, Morgan JLM, Notter DR, Murdoch BM. Genome-wide association study to identify genetic loci associated with gastrointestinal nematode resistance in Katahdin sheep. Anim Genet. 2020 Mar;51(2):330-335. doi: 10.1111/age.12895.

16. Bell AC, Felsenfeld G. Methylation of a CTCF-dependent boundary controls imprinted expression of the Igf2 gene. Nature. 2000 May 25;405(6785):482-5. doi: 10.1038/35013100.

17. Bell AC, West AG, Felsenfeld G. The protein CTCF is required for the enhancer blocking activity of vertebrate insulators. Cell. 1999 Aug 6;98(3):387-96. doi: 10.1016/s0092-8674(00)81967-4.

18. Berlin K, Koren S, Chin CS, Drake JP, Landolin JM, Phillippy AM. Assembling large genomes with single-molecule sequencing and locality-sensitive hashing. Nat Biotechnol. 2015 Jun;33(6):623-30. doi: 10.1038/nbt.3238. Epub 2015 May 25. Erratum in: Nat Biotechnol. 2015 Oct;33(10):1109.

19. Bhattacharyya S, Chandra V, Vijayanand P, Ay F. Identification of significant chromatin contacts from HiChIP data by FitHiChIP. Nat Commun. 2019 Sep 17;10(1):4221. doi: 10.1038/s41467-019-11950-y.

20. Bickhart DM, Rosen BD, Koren S, Sayre BL, Hastie AR, Chan S, Lee J, Lam ET, Liachko I, Sullivan ST, Burton JN, Huson HJ, Nystrom JC, Kelley CM, Hutchison JL, Zhou Y, Sun J, Crisà A, Ponce de León FA, Schwartz JC, Hammond JA, Waldbieser

GC, Schroeder SG, Liu GE, Dunham MJ, Shendure J, Sonstegard TS, Phillippy AM, Van Tassell CP, Smith TP. Single-molecule sequencing and chromatin conformation capture enable de novo reference assembly of the domestic goat genome. Nat Genet. 2017 Apr;49(4):643-650. doi: 10.1038/ng.3802. Epub 2017 Mar 6.

21. Bickmore WA. The spatial organization of the human genome. Annu Rev Genomics Hum Genet. 2013;14:67-84. doi: 10.1146/annurev-genom-091212-153515.

22. Bidwell CA, Waddell JN, Taxis TM, Yu H, Tellam RL, Neary MK, Cockett NE. New insights into polar overdominance in callipyge sheep. Anim Genet. 2014 Aug;45 Suppl 1:51-61. doi: 10.1111/age.12132. Epub 2014 Jul 2.

23. Bintu B, Mateo LJ, Su JH, Sinnott-Armstrong NA, Parker M, Kinrot S, Yamaya K, Boettiger AN, Zhuang X. Super-resolution chromatin tracing reveals domains and cooperative interactions in single cells. Science. 2018 Oct 26;362(6413):eaau1783. doi: 10.1126/science.aau1783.

24. Blackburn HD, Paiva SR, Wildeus S, Getz W, Waldron D, Stobart R, Bixby D, Purdy PH, Welsh C, Spiller S, Brown M. Genetic structure and diversity among sheep breeds in the United States: identification of the major gene pools. J Anim Sci. 2011 Aug;89(8):2336-48. doi: 10.2527/jas.2010-3354. Epub 2011 Mar 7.

25. Bolormaa S, Chamberlain AJ, Khansefid M, Stothard P, Swan AA, Mason B, Prowse-Wilkins CP, Duijvesteijn N, Moghaddar N, van der Werf JH, Daetwyler HD, MacLeod IM. Accuracy of imputation to whole-genome sequence in sheep. Genet Sel Evol. 2019 Jan 17;51(1):1. doi: 10.1186/s12711-018-0443-5.

26. Bonev B, Mendelson Cohen N, Szabo Q, Fritsch L, Papadopoulos GL, Lubling Y, Xu X, Lv X, Hugnot JP, Tanay A, Cavalli G. Multiscale 3D Genome Rewiring during Mouse Neural Development. Cell. 2017 Oct 19;171(3):557-572.e24. doi: 10.1016/j.cell.2017.09.043.

27. Borgel J, Guibert S, Li Y, Chiba H, Schübeler D, Sasaki H, Forné T, Weber M. Targets and dynamics of promoter DNA methylation during early mouse development. Nat Genet. 2010 Dec;42(12):1093-100. doi: 10.1038/ng.708.

28. Bourc'his D, Xu GL, Lin CS, Bollman B, Bestor TH. Dnmt3L and the establishment of maternal genomic imprints. Science. 2001 Dec 21;294(5551):2536-9. doi: 10.1126/science.1065848.

29. Bovine HapMap Consortium, Gibbs RA, Taylor JF, Van Tassell CP, Barendse W, Eversole KA, Gill CA, Green RD, Hamernik DL, Kappes SM, Lien S, Matukumalli LK, McEwan JC, Nazareth LV, Schnabel RD, Weinstock GM, Wheeler DA, Ajmone-Marsan P, Boettcher PJ, Caetano AR, Garcia JF, Hanotte O, Mariani P, Skow LC, Sonstegard TS, Williams JL, Diallo B, Hailemariam L, Martinez ML, Morris CA, Silva LO, Spelman RJ, Mulatu W, Zhao K, Abbey CA, Agaba M, Araujo FR, Bunch RJ, Burton J, Gorni C, Olivier H, Harrison BE, Luff B, Machado MA, Mwakaya J, Plastow G, Sim W, Smith T, Thomas MB, Valentini A, Williams P, Womack J, Woolliams JA, Liu Y, Qin X, Worley KC, Gao C, Jiang H, Moore SS, Ren Y, Song XZ, Bustamante CD, Hernandez RD, Muzny DM, Patil S, San Lucas A, Fu Q, Kent MP, Vega R, Matukumalli A, McWilliam S, Sclep G, Bryc K, Choi J, Gao H, Grefenstette JJ, Murdoch B, Stella A, Villa-Angulo R, Wright M, Aerts J, Jann O, Negrini R, Goddard ME, Hayes BJ, Bradley DG, Barbosa da Silva M, Lau LP, Liu GE, Lynn DJ, Panzitta F, Dodds KG. Genome-wide survey of SNP variation uncovers the genetic structure of cattle breeds. Science. 2009 Apr 24;324(5926):528-32. doi: 10.1126/science.1167936.

30. Boyle AP, Davis S, Shulha HP, Meltzer P, Margulies EH, Weng Z, Furey TS, Crawford GE. High-resolution mapping and characterization of open chromatin across the genome. Cell. 2008 Jan 25;132(2):311-22. doi: 10.1016/j.cell.2007.12.014.

31. Brinkman AB, Gu H, Bartels SJ, Zhang Y, Matarese F, Simmer F, Marks H, Bock C, Gnirke A, Meissner A, Stunnenberg HG. Sequential ChIP-bisulfite sequencing enables direct genome-scale investigation of chromatin and DNA methylation cross-talk. Genome Res. 2012 Jun;22(6):1128-38. doi: 10.1101/gr.133728.111.

32. Bunch TD, Wu C, Zhang YP, Wang S. Phylogenetic analysis of snow sheep (Ovis nivicola) and closely related taxa. J Hered. 2006 Jan-Feb;97(1):21-30. doi: 10.1093/jhered/esi127. Epub 2005 Nov 2. PMID: 16267166.

33. Buenrostro JD, Giresi PG, Zaba LC, Chang HY, Greenleaf WJ. Transposition of native chromatin for fast and sensitive epigenomic profiling of open chromatin, DNA-binding proteins and nucleosome position. Nat Methods. 2013 Dec;10(12):1213-8. doi: 10.1038/nmeth.2688. Epub 2013 Oct 6.

34. Buenrostro JD, Wu B, Chang HY, Greenleaf WJ. ATAC-seq: A Method for Assaying Chromatin Accessibility Genome-Wide. Curr Protoc Mol Biol. 2015 Jan 5;109:21.29.1-21.29.9. doi: 10.1002/0471142727.mb2129s109.

35. Buenrostro JD, Wu B, Litzenburger UM, Ruff D, Gonzales ML, Snyder MP, Chang HY, Greenleaf WJ. Single-cell chromatin accessibility reveals principles of regulatory variation. Nature. 2015 Jul 23;523(7561):486-90. doi: 10.1038/nature14590. Epub 2015 Jun 17.

36. Cabanettes F, Klopp C. D-GENIES: dot plot large genomes in an interactive, efficient and simple way. PeerJ. 2018 Jun 4;6:e4958. doi: 10.7717/peerj.4958.

37. Calo E, Wysocka J. Modification of enhancer chromatin: what, how, and why? Mol Cell. 2013 Mar 7;49(5):825-37. doi: 10.1016/j.molcel.2013.01.038.

38. Chial H. DNA sequencing technologies key to the Human Genome Project. Nat Ed. 2008;1(1):219.

39. Chin CS, Alexander DH, Marks P, Klammer AA, Drake J, Heiner C, Clum A, Copeland A, Huddleston J, Eichler EE, Turner SW, Korlach J. Nonhybrid, finished microbial genome assemblies from long-read SMRT sequencing data. Nat Methods. 2013 Jun;10(6):563-9. doi: 10.1038/nmeth.2474.

40. Chin CS, Peluso P, Sedlazeck FJ, Nattestad M, Concepcion GT, Clum A, Dunn C, O'Malley R, Figueroa-Balderas R, Morales-Cruz A, Cramer GR, Delledonne M, Luo C, Ecker JR, Cantu D, Rank DR, Schatz MC. Phased diploid genome assembly with single-molecule real-time sequencing. Nat Methods. 2016 Dec;13(12):1050-1054. doi: 10.1038/nmeth.4035.

41. Chao W, Huynh KD, Spencer RJ, Davidow LS, Lee JT. CTCF, a candidate trans-acting factor for X-inactivation choice. Science. 2002 Jan 11;295(5553):345-7. doi: 10.1126/science.1065982. Epub 2001 Dec 6. PMID: 11743158.

42. Ciani E, Crepaldi P, Nicoloso L, Lasagna E, Sarti FM, Moioli B, Napolitano F, Carta A, Usai G, D'Andrea M, Marletta D, Ciampolini R, Riggio V, Occidente M, Matassino D, Kompan D, Modesto P, Macciotta N, Ajmone-Marsan P, Pilla F. Genome-wide analysis of Italian sheep diversity reveals a strong geographic pattern and cryptic relationships between breeds. Anim Genet. 2014 Apr;45(2):256-66. doi: 10.1111/age.12106.

43. Clark EL, Archibald AL, Daetwyler HD, Groenen MAM, Harrison PW, Houston RD, Kühn C, Lien S, Macqueen DJ, Reecy JM, Robledo D, Watson M, Tuggle CK, Giuffra E. From FAANG to fork: application of highly annotated genomes to improve farmed animal production. Genome Biol. 2020 Nov 24;21(1):285. doi: 10.1186/s13059-020-02197-8.

44. Clark EL, Bush SJ, McCulloch MEB, Farquhar IL, Young R, Lefevre L, Pridans C, Tsang HG, Wu C, Afrasiabi C, Watson M, Whitelaw CB, Freeman TC, Summers KM, Archibald AL, Hume DA. A high resolution atlas of gene expression in the domestic sheep (Ovis aries). PLoS Genet. 2017 Sep 15;13(9):e1006997. doi: 10.1371/journal.pgen.1006997.

45. Clop A, Marcq F, Takeda H, Pirottin D, Tordoir X, Bibé B, Bouix J, Caiment F, Elsen JM, Eychenne F, Larzul C, Laville E, Meish F, Milenkovic D, Tobin J, Charlier C, Georges M. A mutation creating a potential illegitimate microRNA target site in the myostatin gene affects muscularity in sheep. Nat Genet. 2006 Jul;38(7):813-8. doi: 10.1038/ng1810.

46. Cockett NE, Beever J. Screening for the molecular defect causing spider lamb syndrome in sheep. 2001. US Patent No. 6,306,591 Bl.

47. Cockett NE, Jackson SP, Shay TL, Farnir F, Berghmans S, Snowder GD, Nielsen DM, Georges M. Polar overdominance at the ovine callipyge locus. Science. 1996 Jul 12;273(5272):236-8. doi: 10.1126/science.273.5272.236.

48. Cockett NE, Shay TL, Beever JE, Nielsen D, Albretsen J, Georges M, Peterson K, Stephens A, Vernon W, Timofeevskaia O, South S, Mork J, Maciulis A, Bunch TD. Localization of the locus causing Spider Lamb Syndrome to the distal end of ovine Chromosome 6. Mamm Genome. 1999 Jan;10(1):35-8. doi: 10.1007/s003359900938.

49. Cooper GM. The Cell: A Molecular Approach. 2nd edition. Sunderland (MA): Sinauer Associates; 2000. Regulation of Transcription in Eukaryotes. Available from: https://www.ncbi.nlm.nih.gov/books/NBK9904/.

50. Crane E, Bian Q, McCord RP, Lajoie BR, Wheeler BS, Ralston EJ, Uzawa S, Dekker J, Meyer BJ. Condensin-driven remodelling of X chromosome topology during dosage compensation. Nature. 2015 Jul 9;523(7559):240-4. doi: 10.1038/nature14450. Epub 2015 Jun 1.

51. Cremer T, Cremer M. Chromosome territories. Cold Spring Harb Perspect Biol. 2010 Mar;2(3):a003889. doi: 10.1101/cshperspect.a003889. PMID: 20300217; PMCID: PMC2829961.

52. Creyghton MP, Cheng AW, Welstead GG, Kooistra T, Carey BW, Steine EJ, Hanna J, Lodato MA, Frampton GM, Sharp PA, Boyer LA, Young RA, Jaenisch R. Histone H3K27ac separates active from poised enhancers and predicts developmental state. Proc Natl Acad Sci U S A. 2010 Dec 14;107(50):21931-6. doi: 10.1073/pnas.1016071107.

53. Cusanovich DA, Hill AJ, Aghamirzaie D, Daza RM, Pliner HA, Berletch JB, Filippova GN, Huang X, Christiansen L, DeWitt WS, Lee C, Regalado SG, Read DF, Steemers FJ, Disteche CM, Trapnell C, Shendure J. A Single-Cell Atlas of In Vivo Mammalian Chromatin Accessibility. Cell. 2018 Aug 23;174(5):1309-1324.e18. doi: 10.1016/j.cell.2018.06.052.

54. Della Coletta R, Qiu Y, Ou S, Hufford MB, Hirsch CN. How the pan-genome is changing crop genomics and improvement. Genome Biol. 2021 Jan 4;22(1):3. doi: 10.1186/s13059-020-02224-8.

55. Demirci S, Koban Baştanlar E, Dağtaş ND, Pişkin E, Engin A, Ozer F, Yüncü E, Doğan SA, Togan I. Mitochondrial DNA diversity of modern, ancient and wild sheep(Ovis gmelinii anatolica) from Turkey: new insights on the evolutionary history of sheep. PLoS One. 2013 Dec 11;8(12):e81952. doi: 10.1371/journal.pone.0081952.

56. Dhayalan A, Rajavelu A, Rathert P, Tamas R, Jurkowska RZ, Ragozin S, Jeltsch A. The Dnmt3a PWWP domain reads histone 3 lysine 36 trimethylation and guides DNA methylation. J Biol Chem. 2010 Aug 20;285(34):26114-20. doi: 10.1074/jbc.M109.089433.

57. Dixon JR, Selvaraj S, Yue F, Kim A, Li Y, Shen Y, Hu M, Liu JS, Ren B. Topological domains in mammalian genomes identified by analysis of chromatin interactions. Nature. 2012 Apr 11;485(7398):376-80. doi: 10.1038/nature11082.

58. Doherty R, Couldrey C. Exploring genome wide bisulfite sequencing for DNA methylation analysis in livestock: a technical assessment. Front Genet. 2014 May 13;5:126. doi: 10.3389/fgene.2014.00126.

59. Dong P, Tu X, Chu PY, Lü P, Zhu N, Grierson D, Du B, Li P, Zhong S. 3D Chromatin Architecture of Large Plant Genomes Determined by Local A/B Compartments. Mol Plant. 2017 Dec 4;10(12):1497-1509. doi: 10.1016/j.molp.2017.11.005. Epub 2017 Nov 22. PMID: 29175436.

60. DOSKOCIL J, SORM F. Distribution of 5-methylcytosine in pyrimidine sequences of deoxyribonucleic acids. Biochim Biophys Acta. 1962 Jun 11;55:953-9. doi: 10.1016/0006-3002(62)90909-5.

61. Dotsev AV, Kunz E, Shakhin AV, Petrov SN, Kostyunia OV, Okhlopkov IM, Deniskova TE, Barbato M, Vagirov VA, Medvedev DG Krebs S, Brem G, Medugorac I, Zinovieva NA. The first complete mitochondrial genomes of snow sheep (*Ovis nivicola*) and thinhorn sheep (*Ovis dalli*) and their phylogenetic implications for the genus *Ovis*. Mitochondrial DNA. 2019;4(1):1332-3. doi: 10.1080/23802359.2018.1535849.

62. ENCODE Project Consortium. An integrated encyclopedia of DNA elements in the human genome. Nature. 2012 Sep 6;489(7414):57-74. doi: 10.1038/nature11247.

63. English AC, Richards S, Han Y, Wang M, Vee V, Qu J, Qin X, Muzny DM, Reid JG, Worley KC, Gibbs RA. Mind the gap: upgrading genomes with Pacific Biosciences RS long-read sequencing technology. PLoS One. 2012;7(11):e47768. doi: 10.1371/journal.pone.0047768. Epub 2012 Nov 21.

64. Espeli O, Mercier R, Boccard F. DNA dynamics vary according to macrodomain topography in the E. coli chromosome. Mol Microbiol. 2008 Jun;68(6):1418-27. doi: 10.1111/j.1365-2958.2008.06239.x. Epub 2008 Apr 11.

65. Etherington GJ, Heavens D, Baker D, Lister A, McNelly R, Garcia G, Clavijo B, Macaulay I, Haerty W, Di Palma F. Sequencing smart: De novo sequencing and assembly approaches for a non-model mammal. Gigascience. 2020 May 1;9(5):giaa045. doi: 10.1093/gigascience/giaa045.

66. Everts-van der Wind A, Larkin DM, Green CA, Elliott JS, Olmstead CA, Chiu R, Schein JE, Marra MA, Womack JE, Lewin HA. A high-resolution whole-genome cattle-human comparative map reveals details of mammalian chromosome evolution. Proc Natl Acad Sci U S A. 2005 Dec 20;102(51):18526-31. doi: 10.1073/pnas.0509285102.

67. Fan H, Zhao F, Zhu C, Li F, Liu J, Zhang L, Wei C, Du L. Complete Mitochondrial Genome Sequences of Chinese Indigenous Sheep with Different Tail Types and an Analysis of Phylogenetic Evolution in Domestic Sheep. Asian-Australas J Anim Sci. 2016 May;29(5):631-9. doi: 10.5713/ajas.15.0473. Epub 2015 Sep 3.

68. Fang L, Liu S, Liu M, Kang X, Lin S, Li B, Connor EE, Baldwin RL 6th, Tenesa A, Ma L, Liu GE, Li CJ. Functional annotation of the cattle genome through systematic discovery and characterization of chromatin states and butyrate-induced variations. BMC Biol. 2019 Aug 16;17(1):68. doi: 10.1186/s12915-019-0687-8.

69. Feng S, Cokus SJ, Zhang X, Chen PY, Bostick M, Goll MG, Hetzel J, Jain J, Strauss SH, Halpern ME, Ukomadu C, Sadler KC, Pradhan S, Pellegrini M, Jacobsen SE. Conservation and divergence of methylation patterning in plants and animals. Proc Natl Acad Sci U S A. 2010 May 11;107(19):8689-94. doi: 10.1073/pnas.1002720107.

70. Filippova GN, Fagerlie S, Klenova EM, Myers C, Dehner Y, Goodwin G, Neiman PE, Collins SJ, Lobanenkov VV. An exceptionally conserved transcriptional repressor, CTCF, employs different combinations of zinc fingers to bind diverged promoter sequences of avian and mammalian c-myc oncogenes. Mol Cell Biol. 1996 Jun;16(6):2802-13. doi: 10.1128/MCB.16.6.2802.

71. Filippova GN, Thienes CP, Penn BH, Cho DH, Hu YJ, Moore JM, Klesert TR, Lobanenkov VV, Tapscott SJ. CTCF-binding sites flank CTG/CAG repeats and form a methylation-sensitive insulator at the DM1 locus. Nat Genet. 2001 Aug;28(4):335-43. doi: 10.1038/ng570.

72. Flavahan WA, Drier Y, Liau BB, Gillespie SM, Venteicher AS, Stemmer-Rachamimov AO, Suvà ML, Bernstein BE. Insulator dysfunction and oncogene activation in IDH mutant gliomas. Nature. 2016 Jan 7;529(7584):110-4. doi: 10.1038/nature16490.

73. Foissac S, Djebali S, Munyard K, Vialaneix N, Rau A, Muret K, Esquerré D, Zytnicki M, Derrien T, Bardou P, Blanc F, Cabau C, Crisci E, Dhorne-Pollet S, Drouet F, Faraut T, Gonzalez I, Goubil A, Lacroix-Lamandé S, Laurent F, Marthey S, Marti-Marimon M, Momal-Leisenring R, Mompart F, Quéré P, Robelin D, Cristobal MS, Tosser-Klopp

G, Vincent-Naulleau S, Fabre S, Pinard-Van der Laan MH, Klopp C, Tixier-Boichard M, Acloque H, Lagarrigue S, Giuffra E. Multi-species annotation of transcriptome and chromatin structure in domesticated animals. BMC Biol. 2019 Dec 30;17(1):108. doi: 10.1186/s12915-019-0726-5.

74. Franke M, Ibrahim DM, Andrey G, Schwarzer W, Heinrich V, Schöpflin R, Kraft K, Kempfer R, Jerković I, Chan WL, Spielmann M, Timmermann B, Wittler L, Kurth I, Cambiaso P, Zuffardi O, Houge G, Lambie L, Brancati F, Pombo A, Vingron M, Spitz F, Mundlos S. Formation of new chromatin domains determines pathogenicity of genomic duplications. Nature. 2016 Oct 13;538(7624):265-269. doi: 10.1038/nature19800.

75. Freking BA, Keele JW, Beattie CW, Kappes SM, Smith TP, Sonstegard TS, Nielsen MK, Leymaster KA. Evaluation of the ovine callipyge locus: I. Relative chromosomal position and gene action. J Anim Sci. 1998 Aug;76(8):2062-71. doi: 10.2527/1998.7682062x. PMID: 9734855.

76. Freking BA, King DA, Shackelford SD, Wheeler TL, Smith TPL. Effects and interactions of myostatin and callipyge mutations: I. Growth and carcass traits. J Anim Sci. 2018 Mar 6;96(2):454-461. doi: 10.1093/jas/skx055.

77. Freking BA, Murphy SK, Wylie AA, Rhodes SJ, Keele JW, Leymaster KA, Jirtle RL, Smith TP. Identification of the single base change causing the callipyge muscle hypertrophy phenotype, the only known example of polar overdominance in mammals. Genome Res. 2002 Oct;12(10):1496-506. doi: 10.1101/gr.571002.

78. Frommer M, McDonald LE, Millar DS, Collis CM, Watt F, Grigg GW, Molloy PL, Paul CL. A genomic sequencing protocol that yields a positive display of 5-

methylcytosine residues in individual DNA strands. Proc Natl Acad Sci U S A. 1992 Mar 1;89(5):1827-31. doi: 10.1073/pnas.89.5.1827.

79. Fudenberg G, Abdennur N, Imakaev M, Goloborodko A, Mirny LA. Emerging Evidence of Chromosome Folding by Loop Extrusion. Cold Spring Harb Symp Quant Biol. 2017;82:45-55. doi: 10.1101/sqb.2017.82.034710. Epub 2018 May 4.

80. Fudenberg G, Imakaev M, Lu C, Goloborodko A, Abdennur N, Mirny LA. Formation of Chromosomal Domains by Loop Extrusion. Cell Rep. 2016 May 31;15(9):2038-49. doi: 10.1016/j.celrep.2016.04.085.

81. Ganji M, Shaltiel IA, Bisht S, Kim E, Kalichava A, Haering CH, Dekker C. Real-time imaging of DNA loop extrusion by condensin. Science. 2018 Apr 6;360(6384):102-105. doi: 10.1126/science.aar7831.

82. Gaouar SB, Lafri M, Djaout A, El-Bouyahiaoui R, Bouri A, Bouchatal A, Maftah A, Ciani E, Da Silva AB. Genome-wide analysis highlights genetic dilution in Algerian sheep. Heredity (Edinb). 2017 Mar;118(3):293-301. doi: 10.1038/hdy.2016.86.

83. Genome 10K Community of Scientists. Genome 10K: a proposal to obtain whole-genome sequence for 10,000 vertebrate species. J Hered. 2009 Nov-Dec;100(6):659-74. doi: 10.1093/jhered/esp086.

84. Gerstein MB, Lu ZJ, Van Nostrand EL, Cheng C, Arshinoff BI, Liu T, Yip KY, Robilotto R, Rechtsteiner A, Ikegami K, Alves P, Chateigner A, Perry M, Morris M, Auerbach RK, Feng X, Leng J, Vielle A, Niu W, Rhrissorrakrai K, Agarwal A, Alexander RP, Barber G, Brdlik CM, Brennan J, Brouillet JJ, Carr A, Cheung MS, Clawson H, Contrino S, Dannenberg LO, Dernburg AF, Desai A, Dick L, Dosé AC, Du J, Egelhofer T, Ercan S, Euskirchen G, Ewing B, Feingold EA, Gassmann R, Good PJ,

Green P, Gullier F, Gutwein M, Guyer MS, Habegger L, Han T, Henikoff JG, Henz SR, Hinrichs A, Holster H, Hyman T, Iniguez AL, Janette J, Jensen M, Kato M, Kent WJ, Kephart E, Khivansara V, Khurana E, Kim JK, Kolasinska-Zwierz P, Lai EC, Latorre I, Leahey A, Lewis S, Lloyd P, Lochovsky L, Lowdon RF, Lubling Y, Lyne R, MacCoss M, Mackowiak SD, Mangone M, McKay S, Mecenas D, Merrihew G, Miller DM 3rd, Muroyama A, Murray JI, Ooi SL, Pham H, Phippen T, Preston EA, Rajewsky N, Rätsch G, Rosenbaum H, Rozowsky J, Rutherford K, Ruzanov P, Sarov M, Sasidharan R, Sboner A, Scheid P, Segal E, Shin H, Shou C, Slack FJ, Slightam C, Smith R, Spencer WC, Stinson EO, Taing S, Takasaki T, Vafeados D, Voronina K, Wang G, Washington NL, Whittle CM, Wu B, Yan KK, Zeller G, Zha Z, Zhong M, Zhou X; modENCODE Consortium, Ahringer J, Strome S, Gunsalus KC, Micklem G, Liu XS, Reinke V, Kim SK, Hillier LW, Henikoff S, Piano F, Snyder M, Stein L, Lieb JD, Waterston RH. Integrative analysis of the Caenorhabditis elegans genome by the modENCODE project. Science. 2010 Dec 24;330(6012):1775-87. doi: 10.1126/science.1196914. Epub 2010 Dec 22. Erratum in: Science. 2011 Jan 7;331(6013):30.

85. Ghurye J, Rhie A, Walenz BP, Schmitt A, Selvaraj S, Pop M, Phillippy AM, Koren S. Integrating Hi-C links with assembly graphs for chromosome-scale assembly. PLoS Comput Biol. 2019 Aug 21;15(8):e1007273. doi: 10.1371/journal.pcbi.1007273.

86. Ghurye J, Pop M. Modern technologies and algorithms for scaffolding assembled genomes. PLoS Comput Biol. 2019 Jun 5;15(6):e1006994. doi: 10.1371/journal.pcbi.1006994.

87. Ghurye J, Pop M, Koren S, Bickhart D, Chin CS. Scaffolding of long read assemblies using long range contact information. BMC Genomics. 2017 Jul 12;18(1):527. doi: 10.1186/s12864-017-3879-z.

88. Gibbs RA, Weinstock GM, Metzker ML, Muzny DM, Sodergren EJ, Scherer S, Scott G, Steffen D, Worley KC, Burch PE, Okwuonu G, Hines S, Lewis L, DeRamo C, Delgado O, Dugan-Rocha S, Miner G, Morgan M, Hawes A, Gill R, Celera, Holt RA, Adams MD, Amanatides PG, Baden-Tillson H, Barnstead M, Chin S, Evans CA, Ferriera S, Fosler C, Glodek A, Gu Z, Jennings D, Kraft CL, Nguyen T, Pfannkoch CM, Sitter C, Sutton GG, Venter JC, Woodage T, Smith D, Lee HM, Gustafson E, Cahill P, Kana A, Doucette-Stamm L, Weinstock K, Fechtel K, Weiss RB, Dunn DM, Green ED, Blakesley RW, Bouffard GG, De Jong PJ, Osoegawa K, Zhu B, Marra M, Schein J, Bosdet I, Fjell C, Jones S, Krzywinski M, Mathewson C, Siddiqui A, Wye N, McPherson J, Zhao S, Fraser CM, Shetty J, Shatsman S, Geer K, Chen Y, Abramzon S, Nierman WC, Havlak PH, Chen R, Durbin KJ, Egan A, Ren Y, Song XZ, Li B, Liu Y, Qin X, Cawley S, Worley KC, Cooney AJ, D'Souza LM, Martin K, Wu JQ, Gonzalez-Garay ML, Jackson AR, Kalafus KJ, McLeod MP, Milosavljevic A, Virk D, Volkov A, Wheeler DA, Zhang Z, Bailey JA, Eichler EE, Tuzun E, Birney E, Mongin E, Ureta-Vidal A, Woodwark C, Zdobnov E, Bork P, Suyama M, Torrents D, Alexandersson M, Trask BJ, Young JM, Huang H, Wang H, Xing H, Daniels S, Gietzen D, Schmidt J, Stevens K, Vitt U, Wingrove J, Camara F, Mar Albà M, Abril JF, Guigo R, Smit A, Dubchak I, Rubin EM, Couronne O, Poliakov A, Hübner N, Ganten D, Goesele C, Hummel O, Kreitler T, Lee YA, Monti J, Schulz H, Zimdahl H, Himmelbauer H, Lehrach H, Jacob HJ, Bromberg S, Gullings-Handley J, Jensen-Seaman MI, Kwitek AE, Lazar J, Pasko D, Tonellato PJ, Twigger S, Ponting CP, Duarte JM, Rice S, Goodstadt L, Beatson SA, Emes RD, Winter EE, Webber C, Brandt P, Nyakatura G, Adetobi M, Chiaromonte F, Elnitski L, Eswara P, Hardison RC, Hou M, Kolbe D, Makova K, Miller W, Nekrutenko A, Riemer C, Schwartz S, Taylor J, Yang S, Zhang Y, Lindpaintner K, Andrews TD, Caccamo M, Clamp M, Clarke L, Curwen V, Durbin R, Eyras E, Searle SM, Cooper GM, Batzoglou S, Brudno M, Sidow A, Stone EA, Venter JC, Payseur BA, Bourque G, López-Otín C, Puente XS, Chakrabarti K, Chatterji S, Dewey C, Pachter L, Bray N, Yap VB, Caspi A, Tesler G, Pevzner PA, Haussler D, Roskin KM, Baertsch R, Clawson H, Furey TS, Hinrichs AS, Karolchik D, Kent WJ, Rosenbloom KR, Trumbower H, Weirauch M, Cooper DN, Stenson PD, Ma B, Brent M, Arumugam M, Shteynberg D, Copley RR, Taylor MS, Riethman H, Mudunuri U,

Peterson J, Guyer M, Felsenfeld A, Old S, Mockrin S, Collins F; Rat Genome Sequencing Project Consortium. Genome sequence of the Brown Norway rat yields insights into mammalian evolution. Nature. 2004 Apr 1;428(6982):493-521. doi: 10.1038/nature02426.

89. Giani AM, Gallo GR, Gianfranceschi L, Formenti G. Long walk to genomics: History and current approaches to genome sequencing and assembly. Comput Struct Biotechnol J. 2019 Nov 17;18:9-19. doi: 10.1016/j.csbj.2019.11.002.

90. Gilfillan GD, Hughes T, Sheng Y, Hjorthaug HS, Straub T, Gervin K, Harris JR, Undlien DE, Lyle R. Limitations and possibilities of low cell number ChIP-seq. BMC Genomics. 2012 Nov 21;13:645. doi: 10.1186/1471-2164-13-645.

91. Giresi PG, Kim J, McDaniell RM, Iyer VR, Lieb JD. FAIRE (Formaldehyde-Assisted Isolation of Regulatory Elements) isolates active regulatory elements from human chromatin. Genome Res. 2007 Jun;17(6):877-85. doi: 10.1101/gr.5533506..

92. Giuffra E, Tuggle CK; FAANG Consortium. Functional Annotation of Animal Genomes (FAANG): Current Achievements and Roadmap. Annu Rev Anim Biosci. 2019 Feb 15;7:65-88. doi: 10.1146/annurev-animal-020518-114913.

93. Gonzales-Siles L, Karlsson R, Schmidt P, Salvà-Serra F, Jaén-Luchoro D, Skovbjerg S, Moore ERB, Gomila M. A Pangenome Approach for Discerning Species-Unique Gene Markers for Identifications of Streptococcus pneumoniae and Streptococcus pseudopneumoniae. Front Cell Infect Microbiol. 2020 May 19;10:222. doi: 10.3389/fcimb.2020.00222.

94. Goszczynski DE, Halstead MM, Islas-Trejo AD, Zhou H, Ross PJ. Transcription initiation mapping in 31 bovine tissues reveals complex promoter activity, pervasive

transcription, and tissue-specific promoter usage. Genome Res. 2021 Apr;31(4):732-744. doi: 10.1101/gr.267336.120.

95. Grant M, Zuccotti M, Monk M. Methylation of CpG sites of two X-linked genes coincides with X-inactivation in the female mouse embryo but not in the germ line. Nat Genet. 1992 Oct;2(2):161-6. doi: 10.1038/ng1092-161.

96. Greenberg MVC, Bourc'his D. The diverse roles of DNA methylation in mammalian development and disease. Nat Rev Mol Cell Biol. 2019 Oct;20(10):590-607. doi: 10.1038/s41580-019-0159-6.

97. Grob S, Schmid MW, Grossniklaus U. Hi-C analysis in Arabidopsis identifies the KNOT, a structure with similarities to the flamenco locus of Drosophila. Mol Cell. 2014 Sep 4;55(5):678-93. doi: 10.1016/j.molcel.2014.07.009.

98. Gu H, Smith ZD, Bock C, Boyle P, Gnirke A, Meissner A. Preparation of reduced representation bisulfite sequencing libraries for genome-scale DNA methylation profiling. Nat Protoc. 2011 Apr;6(4):468-81. doi: 10.1038/nprot.2010.190.

99. Guo Y, Xu Q, Canzio D, Shou J, Li J, Gorkin DU, Jung I, Wu H, Zhai Y, Tang Y, Lu Y, Wu Y, Jia Z, Li W, Zhang MQ, Ren B, Krainer AR, Maniatis T, Wu Q. CRISPR Inversion of CTCF Sites Alters Genome Topology and Enhancer/Promoter Function. Cell. 2015 Aug 13;162(4):900-10. doi: 10.1016/j.cell.2015.07.038.

100. Haarhuis JHI, van der Weide RH, Blomen VA, Yáñez-Cuna JO, Amendola M, van Ruiten MS, Krijger PHL, Teunissen H, Medema RH, van Steensel B, Brummelkamp TR, de Wit E, Rowland BD. The Cohesin Release Factor WAPL Restricts Chromatin Loop Extension. Cell. 2017 May 4;169(4):693-707.e14. doi: 10.1016/j.cell.2017.04.013.

101. Hahn W. Sheep, Lamb, & Mutton: Sector at a Glance. USDA Economic Research Service Report. June 24, 2020.

102. Hainer SJ, Fazzio TG. High-Resolution Chromatin Profiling Using CUT&RUN. Curr Protoc Mol Biol. 2019 Apr;126(1):e85. doi: 10.1002/cpmb.85.

103. Halstead MM, Kern C, Saelao P, Chanthavixay G, Wang Y, Delany ME, Zhou H, Ross PJ. Systematic alteration of ATAC-seq for profiling open chromatin in cryopreserved nuclei preparations from livestock tissues. Sci Rep. 2020 Mar 23;10(1):5230. doi: 10.1038/s41598-020-61678-9.

104. Haque A, Engel J, Teichmann SA, Lönnberg T. A practical guide to single-cell RNA-sequencing for biomedical research and clinical applications. Genome Med. 2017 Aug 18;9(1):75. doi: 10.1186/s13073-017-0467-4.

105. He YF, Li BZ, Li Z, Liu P, Wang Y, Tang Q, Ding J, Jia Y, Chen Z, Li L, Sun Y, Li X, Dai Q, Song CX, Zhang K, He C, Xu GL. Tet-mediated formation of 5-carboxylcytosine and its excision by TDG in mammalian DNA. Science. 2011 Sep 2;333(6047):1303-7. doi: 10.1126/science.1210944.

106. Heaton MP, Clawson ML, Chitko-Mckown CG, Leymaster KA, Smith TP, Harhay GP, White SN, Herrmann-Hoesing LM, Mousel MR, Lewis GS, Kalbfleisch TS, Keen JE, Laegreid WW. Reduced lentivirus susceptibility in sheep with TMEM154 mutations. PLoS Genet. 2012 Jan;8(1):e1002467. doi: 10.1371/journal.pgen.1002467.

107. Heaton MP, Kalbfleisch TS, Petrik DT, Simpson B, Kijas JW, Clawson ML, Chitko-McKown CG, Harhay GP, Leymaster KA; International Sheep Genomics Consortium.

Genetic testing for TMEM154 mutations associated with lentivirus susceptibility in sheep. PLoS One. 2013;8(2):e55490. doi: 10.1371/journal.pone.0055490.

108. Heaton MP, Smith TPL, Bickhart DM, Vander Ley BL, Kuehn LA, Oppenheimer J, Shafer WR, Schuetze FT, Stroud B, McClure JC, Barfield JP, Blackburn HD, Kalbfleisch TS, Davenport KM, Kuhn KL, Green RE, Shapiro B, Rosen BD. A Reference Genome Assembly of Simmental Cattle, Bos taurus taurus. J Hered. 2021 Mar 29;112(2):184-191. doi: 10.1093/jhered/esab002.

109. Heaton MP, Smith TPL, Freking BA, Workman AM, Bennett GL, Carnahan JK, Kalbfleisch TS. Using sheep genomes from diverse U.S. breeds to identify missense variants in genes affecting fecundity. F1000Res. 2017 Aug 2;6:1303. doi: 10.12688/f1000research.12216.1.

110. Hnisz D, Weintraub AS, Day DS, Valton AL, Bak RO, Li CH, Goldmann J, Lajoie BR, Fan ZP, Sigova AA, Reddy J, Borges-Rivera D, Lee TI, Jaenisch R, Porteus MH, Dekker J, Young RA. Activation of proto-oncogenes by disruption of chromosome neighborhoods. Science. 2016 Mar 25;351(6280):1454-1458. doi: 10.1126/science.aad9024.

111. Hou C, Li L, Qin ZS, Corces VG. Gene density, transcription, and insulators contribute to the partition of the Drosophila genome into physical domains. Mol Cell. 2012 Nov 9;48(3):471-84. doi: 10.1016/j.molcel.2012.08.031. Epub 2012 Oct 4.

112. Howe K, Clark MD, Torroja CF, Torrance J, Berthelot C, Muffato M, Collins JE, Humphray S, McLaren K, Matthews L, McLaren S, Sealy I, Caccamo M, Churcher C, Scott C, Barrett JC, Koch R, Rauch GJ, White S, Chow W, Kilian B, Quintais LT, Guerra-Assunção JA, Zhou Y, Gu Y, Yen J, Vogel JH, Eyre T, Redmond S, Banerjee R, Chi J, Fu B, Langley E, Maguire SF, Laird GK, Lloyd D, Kenyon E, Donaldson S,

Sehra H, Almeida-King J, Loveland J, Trevanion S, Jones M, Quail M, Willey D, Hunt A, Burton J, Sims S, McLay K, Plumb B, Davis J, Clee C, Oliver K, Clark R, Riddle C, Elliot D, Threadgold G, Harden G, Ware D, Begum S, Mortimore B, Kerry G, Heath P, Phillimore B, Tracey A, Corby N, Dunn M, Johnson C, Wood J, Clark S, Pelan S, Griffiths G, Smith M, Glithero R, Howden P, Barker N, Lloyd C, Stevens C, Harley J, Holt K, Panagiotidis G, Lovell J, Beasley H, Henderson C, Gordon D, Auger K, Wright D, Collins J, Raisen C, Dyer L, Leung K, Robertson L, Ambridge K, Leongamornlert D, McGuire S, Gilderthorp R, Griffiths C, Manthravadi D, Nichol S, Barker G, Whitehead S, Kay M, Brown J, Murnane C, Gray E, Humphries M, Sycamore N, Barker D, Saunders D, Wallis J, Babbage A, Hammond S, Mashreghi-Mohammadi M, Barr L, Martin S, Wray P, Ellington A, Matthews N, Ellwood M, Woodmansey R, Clark G, Cooper J, Tromans A, Grafham D, Skuce C, Pandian R, Andrews R, Harrison E, Kimberley A, Garnett J, Fosker N, Hall R, Garner P, Kelly D, Bird C, Palmer S, Gehring I, Berger A, Dooley CM, Ersan-Ürün Z, Eser C, Geiger H, Geisler M, Karotki L, Kirn A, Konantz J, Konantz M, Oberländer M, Rudolph-Geiger S, Teucke M, Lanz C, Raddatz G, Osoegawa K, Zhu B, Rapp A, Widaa S, Langford C, Yang F, Schuster SC, Carter NP, Harrow J, Ning Z, Herrero J, Searle SM, Enright A, Geisler R, Plasterk RH, Lee C, Westerfield M, de Jong PJ, Zon LI, Postlethwait JH, Nüsslein-Volhard C, Hubbard TJ, Roest Crollius H, Rogers J, Stemple DL. The zebrafish reference genome sequence and its relationship to the human genome. Nature. 2013 Apr 25;496(7446):498-503. doi: 10.1038/nature12111. Epub 2013 Apr 17. Erratum in: Nature. 2014 Jan 9;505(7482):248.

113. Hu XD, Gao LZ. The complete mitochondrial genome of domestic sheep, Ovis aries. Mitochondrial DNA A DNA Mapp Seq Anal. 2016;27(2):1425-7. doi: 10.3109/19401736.2014.953076.

114. Ihara N, Takasuga A, Mizoshita K, Takeda H, Sugimoto M, Mizoguchi Y, Hirano T, Itoh T, Watanabe T, Reed KM, Snelling WM, Kappes SM, Beattie CW, Bennett GL, Sugimoto Y. A comprehensive genetic map of the cattle genome based on 3802 microsatellites. Genome Res. 2004 Oct;14(10A):1987-98. doi: 10.1101/gr.2741704.

115. Ito S, Shen L, Dai Q, Wu SC, Collins LB, Swenberg JA, He C, Zhang Y. Tet proteins can convert 5-methylcytosine to 5-formylcytosine and 5-carboxylcytosine. Science. 2011 Sep 2;333(6047):1300-3. doi: 10.1126/science.1210597.

116. Itoh T, Watanabe T, Ihara N, Mariani P, Beattie CW, Sugimoto Y, Takasuga A. A comprehensive radiation hybrid map of the bovine genome comprising 5593 loci. Genomics. 2005 Apr;85(4):413-24. doi: 10.1016/j.ygeno.2004.12.007.

117. Iyer VR, Horak CE, Scafe CS, Botstein D, Snyder M, Brown PO. Genomic binding sites of the yeast cell-cycle transcription factors SBF and MBF. Nature. 2001 Jan 25;409(6819):533-8. doi: 10.1038/35054095.

118. Jain M, Koren S, Miga KH, Quick J, Rand AC, Sasani TA, Tyson JR, Beggs AD, Dilthey AT, Fiddes IT, Malla S, Marriott H, Nieto T, O'Grady J, Olsen HE, Pedersen BS, Rhie A, Richardson H, Quinlan AR, Snutch TP, Tee L, Paten B, Phillippy AM, Simpson JT, Loman NJ, Loose M. Nanopore sequencing and assembly of a human genome with ultra-long reads. Nat Biotechnol. 2018 Apr;36(4):338-345. doi: 10.1038/nbt.4060.

119. Jermann P, Hoerner L, Burger L, Schübeler D. Short sequences can efficiently recruit histone H3 lysine 27 trimethylation in the absence of enhancer activity and DNA methylation. Proc Natl Acad Sci U S A. 2014 Aug 19;111(33):E3415-21. doi: 10.1073/pnas.1400672111.

120. Jiang Y, Xie M, Chen W, Talbot R, Maddox JF, Faraut T, Wu C, Muzny DM, Li Y, Zhang W, Stanton JA, Brauning R, Barris WC, Hourlier T, Aken BL, Searle SMJ, Adelson DL, Bian C, Cam GR, Chen Y, Cheng S, DeSilva U, Dixen K, Dong Y, Fan G, Franklin IR, Fu S, Guan R, Highland MA, Holder ME, Huang G, Ingham AB,

Jhangiani SN, Kalra D, Kovar CL, Lee SL, Liu W, Liu X, Lu C, Lv T, Mathew T, McWilliam S, Menzies M, Pan S, Robelin D, Servin B, Townley D, Wang W, Wei B, White SN, Yang X, Ye C, Yue Y, Zeng P, Zhou Q, Hansen JB, Kristensen K, Gibbs RA, Flicek P, Warkup CC, Jones HE, Oddy VH, Nicholas FW, McEwan JC, Kijas J, Wang J, Worley KC, Archibald AL, Cockett N, Xu X, Wang W, Dalrymple BP. The sheep genome illuminates biology of the rumen and lipid metabolism. Science. 2014 Jun 6;344(6188):1168-1173. doi: 10.1126/science.1252806.

121. Job RJ, Duan M, Hunter SS, Rodriguez AM, Davenport KM, Eidman L, Murdoch B. Development of Flock54: A targeted genotyping panel for the sheep industry. Plant & Animal Genome Conference. 2019 Jan.

122. Kanduri C, Pant V, Loukinov D, Pugacheva E, Qi CF, Wolffe A, Ohlsson R, Lobanenkov VV. Functional association of CTCF with the insulator upstream of the H19 gene is parent of origin-specific and methylation-sensitive. Curr Biol. 2000 Jul 13;10(14):853-6. doi: 10.1016/s0960-9822(00)00597-2.

123. Kang X, Liu S, Fang L, Lin S, Liu M, Baldwin RL, Liu GE, Li CJ. Data of epigenomic profiling of histone marks and CTCF binding sites in bovine rumen epithelial primary cells before and after butyrate treatment. Data Brief. 2019 Dec 12;28:104983. doi: 10.1016/j.dib.2019.104983.

124. Kasinathan S, Orsi GA, Zentner GE, Ahmad K, Henikoff S. High-resolution mapping of transcription factor binding sites on native chromatin. Nat Methods. 2014 Feb;11(2):203-9. doi: 10.1038/nmeth.2766.

125. Kern C, Wang Y, Chitwood J, Korf I, Delany M, Cheng H, Medrano JF, Van Eenennaam AL, Ernst C, Ross P, Zhou H. Genome-wide identification of tissue-

specific long non-coding RNA in three farm animal species. BMC Genomics. 2018 Sep 18;19(1):684. doi: 10.1186/s12864-018-5037-7.

126. Kern C, Wang Y, Xu X, Pan Z, Halstead M, Chanthavixay G, Saelao P, Waters S, Xiang R, Chamberlain A, Korf I, Delany ME, Cheng HH, Medrano JF, Van Eenennaam AL, Tuggle CK, Ernst C, Flicek P, Quon G, Ross P, Zhou H. Functional annotations of three domestic animal genomes provide vital resources for comparative and agricultural research. Nat Commun. 2021 Mar 23;12(1):1821. doi: 10.1038/s41467-021-22100-8.

127. Kidder BL, Hu G, Zhao K. ChIP-Seq: technical considerations for obtaining high-quality data. Nat Immunol. 2011 Sep 20;12(10):918-22. doi: 10.1038/ni.2117.

128. Kijas JW, Lenstra JA, Hayes B, Boitard S, Porto Neto LR, San Cristobal M, Servin B, McCulloch R, Whan V, Gietzen K, Paiva S, Barendse W, Ciani E, Raadsma H, McEwan J, Dalrymple B; International Sheep Genomics Consortium Members. Genome-wide analysis of the world's sheep breeds reveals high levels of historic mixture and strong recent selection. PLoS Biol. 2012 Feb;10(2):e1001258. doi: 10.1371/journal.pbio.1001258.

129. Klenova EM, Nicolas RH, Paterson HF, Carne AF, Heath CM, Goodwin GH, Neiman PE, Lobanenkov VV. CTCF, a conserved nuclear factor required for optimal transcriptional activity of the chicken c-myc gene, is an 11-Zn-finger protein differentially expressed in multiple forms. Mol Cell Biol. 1993 Dec;13(12):7612-24. doi: 10.1128/mcb.13.12.7612-7624.1993.

130. Koren S, Rhie A, Walenz BP, Dilthey AT, Bickhart DM, Kingan SB, Hiendleder S, Williams JL, Smith TPL, Phillippy AM. De novo assembly of haplotype-resolved

genomes with trio binning. Nat Biotechnol. 2018 Oct 22:10.1038/nbt.4277. doi: 10.1038/nbt.4277.

131. Koren S, Walenz BP, Berlin K, Miller JR, Bergman NH, Phillippy AM. Canu: scalable and accurate long-read assembly via adaptive *k*-mer weighting and repeat separation. Genome Res. 2017 May;27(5):722-736. doi: 10.1101/gr.215087.116.

132. Kuehn LA, Lewis RM, Notter DR. Connectedness in Targhee and Suffolk flocks participating in the United States National Sheep Improvement Program. J Anim Sci. 2009 Feb;87(2):507-15. doi: 10.2527/jas.2008-1092.

133. Lander ES, Linton LM, Birren B, Nusbaum C, Zody MC, Baldwin J, Devon K, Dewar K, Doyle M, FitzHugh W, Funke R, Gage D, Harris K, Heaford A, Howland J, Kann L, Lehoczky J, LeVine R, McEwan P, McKernan K, Meldrim J, Mesirov JP, Miranda C, Morris W, Naylor J, Raymond C, Rosetti M, Santos R, Sheridan A, Sougnez C, Stange-Thomann Y, Stojanovic N, Subramanian A, Wyman D, Rogers J, Sulston J, Ainscough R, Beck S, Bentley D, Burton J, Clee C, Carter N, Coulson A, Deadman R, Deloukas P, Dunham A, Dunham I, Durbin R, French L, Grafham D, Gregory S, Hubbard T, Humphray S, Hunt A, Jones M, Lloyd C, McMurray A, Matthews L, Mercer S, Milne S, Mullikin JC, Mungall A, Plumb R, Ross M, Shownkeen R, Sims S, Waterston RH, Wilson RK, Hillier LW, McPherson JD, Marra MA, Mardis ER, Fulton LA, Chinwalla AT, Pepin KH, Gish WR, Chissoe SL, Wendl MC, Delehaunty KD, Miner TL, Delehaunty A, Kramer JB, Cook LL, Fulton RS, Johnson DL, Minx PJ, Clifton SW, Hawkins T, Branscomb E, Predki P, Richardson P, Wenning S, Slezak T, Doggett N, Cheng JF, Olsen A, Lucas S, Elkin C, Uberbacher E, Frazier M, Gibbs RA, Muzny DM, Scherer SE, Bouck JB, Sodergren EJ, Worley KC, Rives CM, Gorrell JH, Metzker ML, Naylor SL, Kucherlapati RS, Nelson DL, Weinstock GM, Sakaki Y, Fujiyama A, Hattori M, Yada T, Toyoda A, Itoh T, Kawagoe C, Watanabe H, Totoki Y, Taylor T, Weissenbach J, Heilig R, Saurin W, Artiguenave F, Brottier P, Bruls T, Pelletier E, Robert C, Wincker P, Smith DR, Doucette-Stamm L, Rubenfield M, Weinstock K, Lee

HM, Dubois J, Rosenthal A, Platzer M, Nyakatura G, Taudien S, Rump A, Yang H, Yu J, Wang J, Huang G, Gu J, Hood L, Rowen L, Madan A, Qin S, Davis RW, Federspiel NA, Abola AP, Proctor MJ, Myers RM, Schmutz J, Dickson M, Grimwood J, Cox DR, Olson MV, Kaul R, Raymond C, Shimizu N, Kawasaki K, Minoshima S, Evans GA, Athanasiou M, Schultz R, Roe BA, Chen F, Pan H, Ramser J, Lehrach H, Reinhardt R, McCombie WR, de la Bastide M, Dedhia N, Blöcker H, Hornischer K, Nordsiek G, Agarwala R, Aravind L, Bailey JA, Bateman A, Batzoglou S, Birney E, Bork P, Brown DG, Burge CB, Cerutti L, Chen HC, Church D, Clamp M, Copley RR, Doerks T, Eddy SR, Eichler EE, Furey TS, Galagan J, Gilbert JG, Harmon C, Hayashizaki Y, Haussler D, Hermjakob H, Hokamp K, Jang W, Johnson LS, Jones TA, Kasif S, Kaspryzk A, Kennedy S, Kent WJ, Kitts P, Koonin EV, Korf I, Kulp D, Lancet D, Lowe TM, McLysaght A, Mikkelsen T, Moran JV, Mulder N, Pollara VJ, Ponting CP, Schuler G, Schultz J, Slater G, Smit AF, Stupka E, Szustakowki J, Thierry-Mieg D, Thierry-Mieg J, Wagner L, Wallis J, Wheeler R, Williams A, Wolf YI, Wolfe KH, Yang SP, Yeh RF, Collins F, Guyer MS, Peterson J, Felsenfeld A, Wetterstrand KA, Patrinos A, Morgan MJ, de Jong P, Catanese JJ, Osoegawa K, Shizuya H, Choi S, Chen YJ, Szustakowki J; International Human Genome Sequencing Consortium. Initial sequencing and analysis of the human genome. Nature. 2001 Feb 15;409(6822):860-921. doi: 10.1038/35057062. Erratum in: Nature 2001 Aug 2;412(6846):565. Erratum in: Nature 2001 Jun 7;411(6838):720.

134. Landt SG, Marinov GK, Kundaje A, Kheradpour P, Pauli F, Batzoglou S, Bernstein BE, Bickel P, Brown JB, Cayting P, Chen Y, DeSalvo G, Epstein C, Fisher-Aylor KI, Euskirchen G, Gerstein M, Gertz J, Hartemink AJ, Hoffman MM, Iyer VR, Jung YL, Karmakar S, Kellis M, Kharchenko PV, Li Q, Liu T, Liu XS, Ma L, Milosavljevic A, Myers RM, Park PJ, Pazin MJ, Perry MD, Raha D, Reddy TE, Rozowsky J, Shoresh N, Sidow A, Slattery M, Stamatoyannopoulos JA, Tolstorukov MY, White KP, Xi S, Farnham PJ, Lieb JD, Wold BJ, Snyder M. ChIP-seq guidelines and practices of the ENCODE and modENCODE consortia. Genome Res. 2012 Sep;22(9):1813-31. doi: 10.1101/gr.136184.111.

135. Lawrence S, Morton NE, Cox DR. Radiation hybrid mapping. Proc Natl Acad Sci U S A. 1991 Sep 1;88(17):7477-80. doi: 10.1073/pnas.88.17.7477.

136. Le TB, Imakaev MV, Mirny LA, Laub MT. High-resolution mapping of the spatial organization of a bacterial chromosome. Science. 2013 Nov 8;342(6159):731-4. doi: 10.1126/science.1242059.

137. Lee BK, Iyer VR. Genome-wide studies of CCCTC-binding factor (CTCF) and cohesin provide insight into chromatin structure and regulation. J Biol Chem. 2012 Sep 7;287(37):30906-13. doi: 10.1074/jbc.R111.324962.

138. Lewin HA, Robinson GE, Kress WJ, Baker WJ, Coddington J, Crandall KA, Durbin R, Edwards SV, Forest F, Gilbert MTP, Goldstein MM, Grigoriev IV, Hackett KJ, Haussler D, Jarvis ED, Johnson WE, Patrinos A, Richards S, Castilla-Rubio JC, van Sluys MA, Soltis PS, Xu X, Yang H, Zhang G. Earth BioGenome Project: Sequencing life for the future of life. Proc Natl Acad Sci U S A. 2018 Apr 24;115(17):4325-4333. doi: 10.1073/pnas.1720115115.

139. Li E, Bestor TH, Jaenisch R. Targeted mutation of the DNA methyltransferase gene results in embryonic lethality. Cell. 1992 Jun 12;69(6):915-26. doi: 10.1016/0092-8674(92)90611-f.

140. Li E, Zhang Y. DNA methylation in mammals. Cold Spring Harb Perspect Biol. 2014 May 1;6(5):a019133. doi: 10.1101/cshperspect.a019133.

141. Li H. Minimap2: pairwise alignment for nucleotide sequences. Bioinformatics. 2018 Sep 15;34(18):3094-3100. doi: 10.1093/bioinformatics/bty191.

142. Li X, Yang J, Shen M, Xie XL, Liu GJ, Xu YX, Lv FH, Yang H, Yang YL, Liu CB, Zhou P, Wan PC, Zhang YS, Gao L, Yang JQ, Pi WH, Ren YL, Shen ZQ, Wang F, Deng J, Xu SS, Salehian-Dehkordi H, Hehua E, Esmailizadeh A, Dehghani-Qanatqestani M, Štěpánek O, Weimann C, Erhardt G, Amane A, Mwacharo JM, Han JL, Hanotte O, Lenstra JA, Kantanen J, Coltman DW, Kijas JW, Bruford MW, Periasamy K, Wang XH, Li MH. Whole-genome resequencing of wild and domestic sheep identifies genes associated with morphological and agronomic traits. Nat Commun. 2020 Jun 4;11(1):2815. doi: 10.1038/s41467-020-16485-1.

143. Li Z, Schulz MH, Look T, Begemann M, Zenke M, Costa IG. Identification of transcription factor binding sites using ATAC-seq. Genome Biol. 2019 Feb 26;20(1):45. doi: 10.1186/s13059-019-1642-2.

144. Lieberman-Aiden E, van Berkum NL, Williams L, Imakaev M, Ragoczy T, Telling A, Amit I, Lajoie BR, Sabo PJ, Dorschner MO, Sandstrom R, Bernstein B, Bender MA, Groudine M, Gnirke A, Stamatoyannopoulos J, Mirny LA, Lander ES, Dekker J. Comprehensive mapping of long-range interactions reveals folding principles of the human genome. Science. 2009 Oct 9;326(5950):289-93. doi: 10.1126/science.1181369.

145. Lister R, Pelizzola M, Dowen RH, Hawkins RD, Hon G, Tonti-Filippini J, Nery JR, Lee L, Ye Z, Ngo QM, Edsall L, Antosiewicz-Bourget J, Stewart R, Ruotti V, Millar AH, Thomson JA, Ren B, Ecker JR. Human DNA methylomes at base resolution show widespread epigenomic differences. Nature. 2009 Nov 19;462(7271):315-22. doi: 10.1038/nature08514..

146. Liu Y, Qin X, Song XZ, Jiang H, Shen Y, Durbin KJ, Lien S, Kent MP, Sodeland M, Ren Y, Zhang L, Sodergren E, Havlak P, Worley KC, Weinstock GM, Gibbs RA. Bos taurus genome assembly. BMC Genomics. 2009 Apr 24;10:180. doi: 10.1186/1471-2164-10-180.

147. Lock LF, Takagi N, Martin GR. Methylation of the Hprt gene on the inactive X occurs after chromosome inactivation. Cell. 1987 Jan 16;48(1):39-46. doi: 10.1016/0092-8674(87)90353-9.

148. Loukinov DI, Pugacheva E, Vatolin S, Pack SD, Moon H, Chernukhin I, Mannan P, Larsson E, Kanduri C, Vostrov AA, Cui H, Niemitz EL, Rasko JE, Docquier FM, Kistler M, Breen JJ, Zhuang Z, Quitschke WW, Renkawitz R, Klenova EM, Feinberg AP, Ohlsson R, Morse HC 3rd, Lobanenkov VV. BORIS, a novel male germ-line-specific protein associated with epigenetic reprogramming events, shares the same 11-zinc-finger domain with CTCF, the insulator protein involved in reading imprinting marks in the soma. Proc Natl Acad Sci U S A. 2002 May 14;99(10):6806-11. doi: 10.1073/pnas.092123699.

149. Low WY, Tearle R, Liu R, Koren S, Rhie A, Bickhart DM, Rosen BD, Kronenberg ZN, Kingan SB, Tseng E, Thibaud-Nissen F, Martin FJ, Billis K, Ghurye J, Hastie AR, Lee J, Pang AWC, Heaton MP, Phillippy AM, Hiendleder S, Smith TPL, Williams JL. Haplotype-resolved genomes provide insights into structural variation and gene content in Angus and Brahman cattle. Nat Commun. 2020 Apr 29;11(1):2071. doi: 10.1038/s41467-020-15848-y.

150. Lupiáñez DG, Kraft K, Heinrich V, Krawitz P, Brancati F, Klopocki E, Horn D, Kayserili H, Opitz JM, Laxova R, Santos-Simarro F, Gilbert-Dussardier B, Wittler L, Borschiwer M, Haas SA, Osterwalder M, Franke M, Timmermann B, Hecht J, Spielmann M, Visel A, Mundlos S. Disruptions of topological chromatin domains cause pathogenic rewiring of gene-enhancer interactions. Cell. 2015 May 21;161(5):1012-1025. doi: 10.1016/j.cell.2015.04.004.

151. Lupiáñez DG, Spielmann M, Mundlos S. Breaking TADs: How Alterations of Chromatin Domains Result in Disease. Trends Genet. 2016 Apr;32(4):225-237. doi: 10.1016/j.tig.2016.01.003.

152. Lupton CJ. Impacts of animal science research on United States sheep production and predictions for the future. J Anim Sci. 2008 Nov;86(11):3252-74. doi: 10.2527/jas.2008-1148.

153. Lynn T, Grannis J, Williams M, Marshall K, Miller R, Bush E, Bruntz S. An evaluation of scrapie surveillance in the United States. Prev Vet Med. 2007 Sep 14;81(1-3):70-9. doi: 10.1016/j.prevetmed.2007.04.001.

154. Ma L, O'Connell JR, VanRaden PM, Shen B, Padhi A, Sun C, Bickhart DM, Cole JB, Null DJ, Liu GE, Da Y, Wiggans GR. Cattle Sex-Specific Recombination and Genetic Control from a Large Pedigree Analysis. PLoS Genet. 2015 Nov 5;11(11):e1005387. doi: 10.1371/journal.pgen.1005387.

155. Ma S, Zhang Y. Profiling chromatin regulatory landscape: insights into the development of ChIP-seq and ATAC-seq. Mol Biomed. 2020;1:9. doi: 10.1186/s43556-020-00009-w.

156. McKay SD, Schnabel RD, Murdoch BM, Matukumalli LK, Aerts J, Coppieters W, Crews D, Dias Neto E, Gill CA, Gao C, Mannen H, Wang Z, Van Tassell CP, Williams JL, Taylor JF, Moore SS. An assessment of population structure in eight breeds of cattle using a whole genome SNP panel. BMC Genet. 2008 May 20;9:37. doi: 10.1186/1471-2156-9-37.

157. McNatty KP, Hudson NL, Whiting L, Reader KL, Lun S, Western A, Heath DA, Smith P, Moore LG, Juengel JL. The effects of immunizing sheep with different BMP15 or GDF9 peptide sequences on ovarian follicular activity and ovulation rate. Biol Reprod. 2007 Apr;76(4):552-60. doi: 10.1095/biolreprod.106.054361.

158. McKay SD, Schnabel RD, Murdoch BM, Aerts J, Gill CA, Gao C, Li C, Matukumalli LK, Stothard P, Wang Z, Van Tassell CP, Williams JL, Taylor JF, Moore SS. Construction of bovine whole-genome radiation hybrid and linkage maps using high-throughput genotyping. Anim Genet. 2007 Apr;38(2):120-5. doi: 10.1111/j.1365-2052.2006.01564.x.

159. McRae KM, Rowe SJ, Baird HJ, Bixley MJ, Clarke SM. Genome-wide association study of lung lesions and pleurisy in New Zealand lambs. J Anim Sci. 2018 Nov 21;96(11):4512-4520. doi: 10.1093/jas/sky323.

160. Melchior MB, Windig JJ, Hagenaars TJ, Bossers A, Davidse A, van Zijderveld FG. Eradication of scrapie with selective breeding: are we nearly there? BMC Vet Res. 2010 May 4;6:24. doi: 10.1186/1746-6148-6-24.

161. Meissner A, Gnirke A, Bell GW, Ramsahoye B, Lander ES, Jaenisch R. Reduced representation bisulfite sequencing for comparative high-resolution DNA methylation analysis. Nucleic Acids Res. 2005 Oct 13;33(18):5868-77. doi: 10.1093/nar/gki901.

162. Michailidou S, Tsangaris G, Fthenakis GC, Tzora A, Skoufos I, Karkabounas SC, Banos G, Argiriou A, Arsenos G. Genomic diversity and population structure of three autochthonous Greek sheep breeds assessed with genome-wide DNA arrays. Mol Genet Genomics. 2018 Jun;293(3):753-768. doi: 10.1007/s00438-018-1421-x.

163. Miller JR, Delcher AL, Koren S, Venter E, Walenz BP, Brownley A, Johnson J, Li K, Mobarry C, Sutton G. Aggressive assembly of pyrosequencing reads with mates. Bioinformatics. 2008 Dec 15;24(24):2818-24. doi: 10.1093/bioinformatics/btn548.

164. Miller JM, Malenfant RM, Moore SS, Coltman DW. Short reads, circular genome: skimming solid sequence to construct the bighorn sheep mitochondrial genome. J Hered. 2012 Jan-Feb;103(1):140-6. doi: 10.1093/jhered/esr104.

165. Miller LR, Stepanek Shiflett J, Marsh DJ, Rogers P. U.S. sheep industry research, development, and education priorities. American Sheep Industry Association, Inc. 2016 June.

166. modENCODE Consortium, Roy S, Ernst J, Kharchenko PV, Kheradpour P, Negre N, Eaton ML, Landolin JM, Bristow CA, Ma L, Lin MF, Washietl S, Arshinoff BI, Ay F, Meyer PE, Robine N, Washington NL, Di Stefano L, Berezikov E, Brown CD, Candeias R, Carlson JW, Carr A, Jungreis I, Marbach D, Sealfon R, Tolstorukov MY, Will S, Alekseyenko AA, Artieri C, Booth BW, Brooks AN, Dai Q, Davis CA, Duff MO, Feng X, Gorchakov AA, Gu T, Henikoff JG, Kapranov P, Li R, MacAlpine HK, Malone J, Minoda A, Nordman J, Okamura K, Perry M, Powell SK, Riddle NC, Sakai A, Samsonova A, Sandler JE, Schwartz YB, Sher N, Spokony R, Sturgill D, van Baren M, Wan KH, Yang L, Yu C, Feingold E, Good P, Guyer M, Lowdon R, Ahmad K, Andrews J, Berger B, Brenner SE, Brent MR, Cherbas L, Elgin SC, Gingeras TR, Grossman R, Hoskins RA, Kaufman TC, Kent W, Kuroda MI, Orr-Weaver T, Perrimon N, Pirrotta V, Posakony JW, Ren B, Russell S, Cherbas P, Graveley BR, Lewis S, Micklem G, Oliver B, Park PJ, Celniker SE, Henikoff S, Karpen GH, Lai EC, MacAlpine DM, Stein LD, White KP, Kellis M. Identification of functional elements and regulatory circuits by Drosophila modENCODE. Science. 2010 Dec 24;330(6012):1787-97. doi: 10.1126/science.1198374.

167. Monk M, Boubelik M, Lehnert S. Temporal and regional changes in DNA methylation in the embryonic, extraembryonic and germ cell lineages during mouse embryo development. Development. 1987 Mar;99(3):371-82.

168. Moore WS. INFERRING PHYLOGENIES FROM mtDNA VARIATION: MITOCHONDRIAL-GENE TREES VERSUS NUCLEAR-GENE TREES. Evolution. 1995 Aug;49(4):718-726. doi: 10.1111/j.1558-5646.1995.tb02308.x.

169. Moritz C. Defining evolutionary significant units for conservation. Trends Ecol Evol. 1994;9:373-5. doi:10.1016/0169-5347(94)90057-4.

170. Mortimer SI, Fogarty NM, van der Werf JHJ, Brown DJ, Swan AA, Jacob RH, Geesink GH, Hopkins DL, Hocking Edwards JE, Ponnampalam EN, Warner RD, Pearce KL, Pethick DW. Genetic correlations between meat quality traits and growth and carcass traits in Merino sheep1. J Anim Sci. 2018 Sep 7;96(9):3582-3598. doi: 10.1093/jas/sky232.

171. Mostovoy Y, Levy-Sakin M, Lam J, Lam ET, Hastie AR, Marks P, Lee J, Chu C, Lin C, Džakula Ž, Cao H, Schlebusch SA, Giorda K, Schnall-Levin M, Wall JD, Kwok PY. A hybrid approach for de novo human genome sequence assembly and phasing. Nat Methods. 2016 Jul;13(7):587-90. doi: 10.1038/nmeth.3865.

172. Mouse Genome Sequencing Consortium, Waterston RH, Lindblad-Toh K, Birney E, Rogers J, Abril JF, Agarwal P, Agarwala R, Ainscough R, Alexandersson M, An P, Antonarakis SE, Attwood J, Baertsch R, Bailey J, Barlow K, Beck S, Berry E, Birren B, Bloom T, Bork P, Botcherby M, Bray N, Brent MR, Brown DG, Brown SD, Bult C, Burton J, Butler J, Campbell RD, Carninci P, Cawley S, Chiaromonte F, Chinwalla AT, Church DM, Clamp M, Clee C, Collins FS, Cook LL, Copley RR, Coulson A, Couronne O, Cuff J, Curwen V, Cutts T, Daly M, David R, Davies J, Delehaunty KD, Deri J, Dermitzakis ET, Dewey C, Dickens NJ, Diekhans M, Dodge S, Dubchak I, Dunn DM, Eddy SR, Elnitski L, Emes RD, Eswara P, Eyras E, Felsenfeld A, Fewell GA, Flicek P, Foley K, Frankel WN, Fulton LA, Fulton RS, Furey TS, Gage D, Gibbs RA, Glusman G, Gnerre S, Goldman N, Goodstadt L, Grafham D, Graves TA, Green ED, Gregory S, Guigó R, Guyer M, Hardison RC, Haussler D, Hayashizaki Y, Hillier

LW, Hinrichs A, Hlavina W, Holzer T, Hsu F, Hua A, Hubbard T, Hunt A, Jackson I, Jaffe DB, Johnson LS, Jones M, Jones TA, Joy A, Kamal M, Karlsson EK, Karolchik D, Kasprzyk A, Kawai J, Keibler E, Kells C, Kent WJ, Kirby A, Kolbe DL, Korf I, Kucherlapati RS, Kulbokas EJ, Kulp D, Landers T, Leger JP, Leonard S, Letunic I, Levine R, Li J, Li M, Lloyd C, Lucas S, Ma B, Maglott DR, Mardis ER, Matthews L, Mauceli E, Mayer JH, McCarthy M, McCombie WR, McLaren S, McLay K, McPherson JD, Meldrim J, Meredith B, Mesirov JP, Miller W, Miner TL, Mongin E, Montgomery KT, Morgan M, Mott R, Mullikin JC, Muzny DM, Nash WE, Nelson JO, Nhan MN, Nicol R, Ning Z, Nusbaum C, O'Connor MJ, Okazaki Y, Oliver K, Overton-Larty E, Pachter L, Parra G, Pepin KH, Peterson J, Pevzner P, Plumb R, Pohl CS, Poliakov A, Ponce TC, Ponting CP, Potter S, Quail M, Reymond A, Roe BA, Roskin KM, Rubin EM, Rust AG, Santos R, Sapojnikov V, Schultz B, Schultz J, Schwartz MS, Schwartz S, Scott C, Seaman S, Searle S, Sharpe T, Sheridan A, Shownkeen R, Sims S, Singer JB, Slater G, Smit A, Smith DR, Spencer B, Stabenau A, Stange-Thomann N, Sugnet C, Suyama M, Tesler G, Thompson J, Torrents D, Trevaskis E, Tromp J, Ucla C, Ureta-Vidal A, Vinson JP, Von Niederhausern AC, Wade CM, Wall M, Weber RJ, Weiss RB, Wendl MC, West AP, Wetterstrand K, Wheeler R, Whelan S, Wierzbowski J, Willey D, Williams S, Wilson RK, Winter E, Worley KC, Wyman D, Yang S, Yang SP, Zdobnov EM, Zody MC, Lander ES. Initial sequencing and comparative analysis of the mouse genome. Nature. 2002 Dec 5;420(6915):520-62. doi: 10.1038/nature01262.

173. Mumbach MR, Rubin AJ, Flynn RA, Dai C, Khavari PA, Greenleaf WJ, Chang HY. HiChIP: efficient and sensitive analysis of protein-directed genome architecture. Nat Methods. 2016 Nov;13(11):919-922. doi: 10.1038/nmeth.3999.

174. Muriuki C, Bush SJ, Salavati M, McCulloch MEB, Lisowski ZM, Agaba M, Djikeng A, Hume DA, Clark EL. A Mini-Atlas of Gene Expression for the Domestic Goat (*Capra hircus*). Front Genet. 2019 Nov 4;10:1080. doi: 10.3389/fgene.2019.01080.

175. Murphy SK, Freking BA, Smith TP, Leymaster K, Nolan CM, Wylie AA, Evans HK, Jirtle RL. Abnormal postnatal maintenance of elevated DLK1 transcript levels in

callipyge sheep. Mamm Genome. 2005 Mar;16(3):171-83. doi: 10.1007/s00335-004-2421-1.

176. Murphy SK, Nolan CM, Huang Z, Kucera KS, Freking BA, Smith TP, Leymaster KA, Weidman JR, Jirtle RL. Callipyge mutation affects gene expression in cis: a potential role for chromatin structure. Genome Res. 2006 Mar;16(3):340-6. doi: 10.1101/gr.4389306.

177. Mustafa SI, Schwarzacher T, Heslop-Harrison JS. Complete mitogenomes from Kurdistani sheep: abundant centromeric nuclear copies representing diverse ancestors. Mitochondrial DNA A DNA Mapp Seq Anal. 2018 Dec;29(8):1180-1193. doi: 10.1080/24701394.2018.1431226.

178. Myers EW, Sutton GG, Delcher AL, Dew IM, Fasulo DP, Flanigan MJ, Kravitz SA, Mobarry CM, Reinert KH, Remington KA, Anson EL, Bolanos RA, Chou HH, Jordan CM, Halpern AL, Lonardi S, Beasley EM, Brandon RC, Chen L, Dunn PJ, Lai Z, Liang Y, Nusskern DR, Zhan M, Zhang Q, Zheng X, Rubin GM, Adams MD, Venter JC. A whole-genome assembly of Drosophila. Science. 2000 Mar 24;287(5461):2196-204. doi: 10.1126/science.287.5461.2196.

179. National Academy of Sciences. Changes in the Sheep Industry in the United States. 2008.

180. Naval-Sanchez M, Nguyen Q, McWilliam S, Porto-Neto LR, Tellam R, Vuocolo T, Reverter A, Perez-Enciso M, Brauning R, Clarke S, McCulloch A, Zamani W, Naderi S, Rezaei HR, Pompanon F, Taberlet P, Worley KC, Gibbs RA, Muzny DM, Jhangiani SN, Cockett N, Daetwyler H, Kijas J. Sheep genome functional annotation reveals proximal regulatory elements contributed to the evolution of modern breeds. Nat Commun. 2018 Feb 28;9(1):859. doi: 10.1038/s41467-017-02809-1.

181. Ng HH, Ciccone DN, Morshead KB, Oettinger MA, Struhl K. Lysine-79 of histone H3 is hypomethylated at silenced loci in yeast and mammalian cells: a potential mechanism for position-effect variegation. Proc Natl Acad Sci U S A. 2003 Feb 18;100(4):1820-5. doi: 10.1073/pnas.0437846100.

182. Nishiyama A, Yamaguchi L, Sharif J, Johmura Y, Kawamura T, Nakanishi K, Shimamura S, Arita K, Kodama T, Ishikawa F, Koseki H, Nakanishi M. Uhrf1-dependent H3K23 ubiquitylation couples maintenance DNA methylation and replication. Nature. 2013 Oct 10;502(7470):249-53. doi: 10.1038/nature12488.

183. Nolte W, Weikard R, Brunner RM, Albrecht E, Hammon HM, Reverter A, Küehn C. Identification and Annotation of Potential Function of Regulatory Antisense Long Non-Coding RNAs Related to Feed Efficiency in *Bos taurus* Bulls. Int J Mol Sci. 2020 May 6;21(9):3292. doi: 10.3390/ijms21093292.

184. Nora EP, Goloborodko A, Valton AL, Gibcus JH, Uebersohn A, Abdennur N, Dekker J, Mirny LA, Bruneau BG. Targeted Degradation of CTCF Decouples Local Insulation of Chromosome Domains from Genomic Compartmentalization. Cell. 2017 May 18;169(5):930-944.e22. doi: 10.1016/j.cell.2017.05.004.

185. Nora EP, Lajoie BR, Schulz EG, Giorgetti L, Okamoto I, Servant N, Piolot T, van Berkum NL, Meisig J, Sedat J, Gribnau J, Barillot E, Blüthgen N, Dekker J, Heard E. Spatial partitioning of the regulatory landscape of the X-inactivation centre. Nature. 2012 Apr 11;485(7398):381-5. doi: 10.1038/nature11049.

186. Nuebler J, Fudenberg G, Imakaev M, Abdennur N, Mirny LA. Chromatin organization by an interplay of loop extrusion and compartmental segregation. Proc Natl Acad Sci U S A. 2018 Jul 17;115(29):E6697-E6706. doi: 10.1073/pnas.1717730115.

187. Okano M, Bell DW, Haber DA, Li E. DNA methyltransferases Dnmt3a and Dnmt3b are essential for de novo methylation and mammalian development. Cell. 1999 Oct 29;99(3):247-57. doi: 10.1016/s0092-8674(00)81656-6.

188. Okano M, Xie S, Li E. Cloning and characterization of a family of novel mammalian DNA (cytosine-5) methyltransferases. Nat Genet. 1998 Jul;19(3):219-20. doi: 10.1038/890.

189. Oluwadare O, Highsmith M, Cheng J. An Overview of Methods for Reconstructing 3-D Chromosome and Genome Structures from Hi-C Data. Biol Proced Online. 2019 Apr 24;21:7. doi: 10.1186/s12575-019-0094-0.

190. O'Neill LP, Turner BM. Immunoprecipitation of native chromatin: NChIP. Methods. 2003 Sep;31(1):76-82. doi: 10.1016/s1046-2023(03)00090-2.

191. Ooi SK, Qiu C, Bernstein E, Li K, Jia D, Yang Z, Erdjument-Bromage H, Tempst P, Lin SP, Allis CD, Cheng X, Bestor TH. DNMT3L connects unmethylated lysine 4 of histone H3 to de novo methylation of DNA. Nature. 2007 Aug 9;448(7154):714-7. doi: 10.1038/nature05987.

192. Oppenheimer J, Rosen BD, Heaton MP, Vander Ley BL, Shafer WR, Schuetze FT, Stroud B, Kuehn LA, McClure JC, Barfield JP, Blackburn HD, Kalbfleisch TS, Bickhart DM, Davenport KM, Kuhn KL, Green RE, Shapiro B, Smith TPL. A Reference Genome Assembly of American Bison, Bison bison bison. J Hered. 2021 Mar 29;112(2):174-183. doi: 10.1093/jhered/esab003.

193. Phillips-Cremins JE, Sauria ME, Sanyal A, Gerasimova TI, Lajoie BR, Bell JS, Ong CT, Hookway TA, Guo C, Sun Y, Bland MJ, Wagstaff W, Dalton S, McDevitt TC, Sen R, Dekker J, Taylor J, Corces VG. Architectural protein subclasses shape 3D organization of genomes during lineage commitment. Cell. 2013 Jun 6;153(6):1281-95. doi: 10.1016/j.cell.2013.04.053.

194. Piunti A, Shilatifard A. Epigenetic balance of gene expression by Polycomb and COMPASS families. Science. 2016 Jun 3;352(6290):aad9780. doi: 10.1126/science.aad9780.

195. Pokholok DK, Harbison CT, Levine S, Cole M, Hannett NM, Lee TI, Bell GW, Walker K, Rolfe PA, Herbolsheimer E, Zeitlinger J, Lewitter F, Gifford DK, Young RA. Genome-wide map of nucleosome acetylation and methylation in yeast. Cell. 2005 Aug 26;122(4):517-27. doi: 10.1016/j.cell.2005.06.026.

196. Putnam NH, O'Connell BL, Stites JC, Rice BJ, Blanchette M, Calef R, Troll CJ, Fields A, Hartley PD, Sugnet CW, Haussler D, Rokhsar DS, Green RE. Chromosome-scale shotgun assembly using an in vitro method for long-range linkage. Genome Res. 2016 Mar;26(3):342-50. doi: 10.1101/gr.193474.115.

197. Qi CF, Martensson A, Mattioli M, Dalla-Favera R, Lobanenkov VV, Morse HC 3rd. CTCF functions as a critical regulator of cell-cycle arrest and death after ligation of the B cell receptor on immature B cells. Proc Natl Acad Sci U S A. 2003 Jan 21;100(2):633-8. doi: 10.1073/pnas.0237127100.

198. Qiao G, Zhang H, Zhu S, Yuan C, Zhao H, Han M, Yue Y, Yang B. The complete mitochondrial genome sequence and phylogenetic analysis of Alpine Merino sheep (*Ovis aries*). Mitochondrial DNA B Resour. 2020 Feb 3;5(1):990-991. doi: 10.1080/23802359.2020.1720536.

199. Qin W, Wolf P, Liu N, Link S, Smets M, La Mastra F, Forné I, Pichler G, Hörl D, Fellinger K, Spada F, Bonapace IM, Imhof A, Harz H, Leonhardt H. DNA methylation requires a DNMT1 ubiquitin interacting motif (UIM) and histone ubiquitination. Cell Res. 2015 Aug;25(8):911-29. doi: 10.1038/cr.2015.72.

200. Rao SSP, Huang SC, Glenn St Hilaire B, Engreitz JM, Perez EM, Kieffer-Kwon KR, Sanborn AL, Johnstone SE, Bascom GD, Bochkov ID, Huang X, Shamim MS, Shin J, Turner D, Ye Z, Omer AD, Robinson JT, Schlick T, Bernstein BE, Casellas R, Lander ES, Aiden EL. Cohesin Loss Eliminates All Loop Domains. Cell. 2017 Oct 5;171(2):305-320.e24. doi: 10.1016/j.cell.2017.09.026.

201. Rao SS, Huntley MH, Durand NC, Stamenova EK, Bochkov ID, Robinson JT, Sanborn AL, Machol I, Omer AD, Lander ES, Aiden EL. A 3D map of the human genome at kilobase resolution reveals principles of chromatin looping. Cell. 2014 Dec 18;159(7):1665-80. doi: 10.1016/j.cell.2014.11.021.

202. Ren B, Robert F, Wyrick JJ, Aparicio O, Jennings EG, Simon I, Zeitlinger J, Schreiber J, Hannett N, Kanin E, Volkert TL, Wilson CJ, Bell SP, Young RA. Genome-wide location and function of DNA binding proteins. Science. 2000 Dec 22;290(5500):2306-9. doi: 10.1126/science.290.5500.2306.

203. Renschler G, Richard G, Valsecchi CIK, Toscano S, Arrigoni L, Ramírez F, Akhtar A. Hi-C guided assemblies reveal conserved regulatory topologies on X and autosomes despite extensive genome shuffling. Genes Dev. 2019 Nov 1;33(21-22):1591-1612. doi: 10.1101/gad.328971.119.

204. Rhie A, McCarthy SA, Fedrigo O, Damas J, Formenti G, Koren S, Uliano-Silva M, Chow W, Fungtammasan A, Kim J, Lee C, Ko BJ, Chaisson M, Gedman GL, Cantin

LJ, Thibaud-Nissen F, Haggerty L, Bista I, Smith M, Haase B, Mountcastle J, Winkler S, Paez S, Howard J, Vernes SC, Lama TM, Grutzner F, Warren WC, Balakrishnan CN, Burt D, George JM, Biegler MT, Iorns D, Digby A, Eason D, Robertson B, Edwards T, Wilkinson M, Turner G, Meyer A, Kautt AF, Franchini P, Detrich HW 3rd, Svardal H, Wagner M, Naylor GJP, Pippel M, Malinsky M, Mooney M, Simbirsky M, Hannigan BT, Pesout T, Houck M, Misuraca A, Kingan SB, Hall R, Kronenberg Z, Sović I, Dunn C, Ning Z, Hastie A, Lee J, Selvaraj S, Green RE, Putnam NH, Gut I, Ghurye J, Garrison E, Sims Y, Collins J, Pelan S, Torrance J, Tracey A, Wood J, Dagnew RE, Guan D, London SE, Clayton DF, Mello CV, Friedrich SR, Lovell PV, Osipova E, Al-Ajli FO, Secomandi S, Kim H, Theofanopoulou C, Hiller M, Zhou Y, Harris RS, Makova KD, Medvedev P, Hoffman J, Masterson P, Clark K, Martin F, Howe K, Flicek P, Walenz BP, Kwak W, Clawson H, Diekhans M, Nassar L, Paten B, Kraus RHS, Crawford AJ, Gilbert MTP, Zhang G, Venkatesh B, Murphy RW, Koepfli KP, Shapiro B, Johnson WE, Di Palma F, Marques-Bonet T, Teeling EC, Warnow T, Graves JM, Ryder OA, Haussler D, O'Brien SJ, Korlach J, Lewin HA, Howe K, Myers EW, Durbin R, Phillippy AM, Jarvis ED. Towards complete and error-free genome assemblies of all vertebrate species. Nature. 2021 Apr;592(7856):737-746. doi: 10.1038/s41586-021-03451-0.

205. Rice ES, Koren S, Rhie A, Heaton MP, Kalbfleisch TS, Hardy T, Hackett PH, Bickhart DM, Rosen BD, Ley BV, Maurer NW, Green RE, Phillippy AM, Petersen JL, Smith TPL. Continuous chromosome-scale haplotypes assembled from a single interspecies F1 hybrid of yak and cattle. Gigascience. 2020 Apr 1;9(4):giaa029. doi: 10.1093/gigascience/giaa029.

206. Ristimäki A, Ylikorkala O, Pesonen K, Perheentupa J, Viinikka L. Human milk stimulates prostacyclin production by cultured human vascular endothelial cells. J Clin Endocrinol Metab. 1991 Mar;72(3):623-7. doi: 10.1210/jcem-72-3-623.

207. Rosen BD, Bickhart DM, Schnabel RD, Koren S, Elsik CG, Tseng E, Rowan TN, Low WY, Zimin A, Couldrey C, Hall R, Li W, Rhie A, Ghurye J, McKay SD, Thibaud-Nissen F, Hoffman J, Murdoch BM, Snelling WM, McDaneld TG, Hammond JA, Schwartz JC, Nandolo W, Hagen DE, Dreischer C, Schultheiss SJ, Schroeder SG, Phillippy AM, Cole JB, Van Tassell CP, Liu G, Smith TPL, Medrano JF. De novo assembly of the cattle reference genome with single-molecule sequencing. Gigascience. 2020 Mar 1;9(3):giaa021. doi: 10.1093/gigascience/giaa021.

208. Rowen L, Mahairas G, Hood L. Sequencing the human genome. Science. 1997 Oct 24;278(5338):605-7. doi: 10.1126/science.278.5338.605.

209. Rowley MJ, Nichols MH, Lyu X, Ando-Kuri M, Rivera ISM, Hermetz K, Wang P, Ruan Y, Corces VG. Evolutionarily Conserved Principles Predict 3D Chromatin Organization. Mol Cell. 2017 Sep 7;67(5):837-852.e7. doi: 10.1016/j.molcel.2017.07.022.

210. Sanborn AL, Rao SS, Huang SC, Durand NC, Huntley MH, Jewett AI, Bochkov ID, Chinnappan D, Cutkosky A, Li J, Geeting KP, Gnirke A, Melnikov A, McKenna D, Stamenova EK, Lander ES, Aiden EL. Chromatin extrusion explains key features of loop and domain formation in wild-type and engineered genomes. Proc Natl Acad Sci U S A. 2015 Nov 24;112(47):E6456-65. doi: 10.1073/pnas.1518552112.

211. Sanford JP, Clark HJ, Chapman VM, Rossant J. Differences in DNA methylation during oogenesis and spermatogenesis and their persistence during early embryogenesis in the mouse. Genes Dev. 1987 Dec;1(10):1039-46. doi: 10.1101/gad.1.10.1039.

212. Salavati M, Bush SJ, Palma-Vera S, McCulloch MEB, Hume DA, Clark EL. Elimination of Reference Mapping Bias Reveals Robust Immune Related Allele-Specific Expression in Crossbred Sheep. Front Genet. 2019 Sep 19;10:863. doi: 10.3389/fgene.2019.00863.

213. Sanborn AL, Rao SS, Huang SC, Durand NC, Huntley MH, Jewett AI, Bochkov ID, Chinnappan D, Cutkosky A, Li J, Geeting KP, Gnirke A, Melnikov A, McKenna D, Stamenova EK, Lander ES, Aiden EL. Chromatin extrusion explains key features of loop and domain formation in wild-type and engineered genomes. Proc Natl Acad Sci U S A. 2015 Nov 24;112(47):E6456-65. doi: 10.1073/pnas.1518552112.

214. Schwartz DC, Li X, Hernandez LI, Ramnarain SP, Huff EJ, Wang YK. Ordered restriction maps of Saccharomyces cerevisiae chromosomes constructed by optical mapping. Science. 1993 Oct 1;262(5130):110-4. doi: 10.1126/science.8211116.

215. Schwarzer W, Abdennur N, Goloborodko A, Pekowska A, Fudenberg G, Loe-Mie Y, Fonseca NA, Huber W, Haering CH, Mirny L, Spitz F. Two independent modes of chromatin organization revealed by cohesin removal. Nature. 2017 Nov 2;551(7678):51-56. doi: 10.1038/nature24281.

216. Sea Urchin Genome Sequencing Consortium, Sodergren E, Weinstock GM, Davidson EH, Cameron RA, Gibbs RA, Angerer RC, Angerer LM, Arnone MI, Burgess DR, Burke RD, Coffman JA, Dean M, Elphick MR, Ettensohn CA, Foltz KR, Hamdoun A, Hynes RO, Klein WH, Marzluff W, McClay DR, Morris RL, Mushegian A, Rast JP, Smith LC, Thorndyke MC, Vacquier VD, Wessel GM, Wray G, Zhang L, Elsik CG, Ermolaeva O, Hlavina W, Hofmann G, Kitts P, Landrum MJ, Mackey AJ, Maglott D, Panopoulou G, Poustka AJ, Pruitt K, Sapojnikov V, Song X, Souvorov A, Solovyev V, Wei Z, Whittaker CA, Worley K, Durbin KJ, Shen Y, Fedrigo O, Garfield D, Haygood R, Primus A, Satija R, Severson T, Gonzalez-Garay ML, Jackson AR, Milosavljevic A, Tong M, Killian CE, Livingston BT, Wilt FH, Adams N, Bellé R, Carbonneau S, Cheung R, Cormier P, Cosson B, Croce J, Fernandez-Guerra A, Genevière AM, Goel M, Kelkar H, Morales J, Mulner-Lorillon O, Robertson AJ, Goldstone JV, Cole B, Epel D, Gold B, Hahn ME, Howard-Ashby M, Scally M, Stegeman JJ, Allgood EL, Cool J, Judkins KM, McCafferty SS, Musante AM, Obar RA, Rawson AP, Rossetti BJ,

Gibbons IR, Hoffman MP, Leone A, Istrail S, Materna SC, Samanta MP, Stolc V, Tongprasit W, Tu Q, Bergeron KF, Brandhorst BP, Whittle J, Berney K, Bottjer DJ, Calestani C, Peterson K, Chow E, Yuan QA, Elhaik E, Graur D, Reese JT, Bosdet I, Heesun S, Marra MA, Schein J, Anderson MK, Brockton V, Buckley KM, Cohen AH, Fugmann SD, Hibino T, Loza-Coll M, Majeske AJ, Messier C, Nair SV, Pancer Z, Terwilliger DP, Agca C, Arboleda E, Chen N, Churcher AM, Hallböök F, Humphrey GW, Idris MM, Kiyama T, Liang S, Mellott D, Mu X, Murray G, Olinski RP, Raible F, Rowe M, Taylor JS, Tessmar-Raible K, Wang D, Wilson KH, Yaguchi S, Gaasterland T, Galindo BE, Gunaratne HJ, Juliano C, Kinukawa M, Moy GW, Neill AT, Nomura M, Raisch M, Reade A, Roux MM, Song JL, Su YH, Townley IK, Voronina E, Wong JL, Amore G, Branno M, Brown ER, Cavalieri V, Duboc V, Duloquin L, Flytzanis C, Gache C, Lapraz F, Lepage T, Locascio A, Martinez P, Matassi G, Matranga V, Range R, Rizzo F, Röttinger E, Beane W, Bradham C, Byrum C, Glenn T, Hussain S, Manning G, Miranda E, Thomason R, Walton K, Wikramanayke A, Wu SY, Xu R, Brown CT, Chen L, Gray RF, Lee PY, Nam J, Oliveri P, Smith J, Muzny D, Bell S, Chacko J, Cree A, Curry S, Davis C, Dinh H, Dugan-Rocha S, Fowler J, Gill R, Hamilton C, Hernandez J, Hines S, Hume J, Jackson L, Jolivet A, Kovar C, Lee S, Lewis L, Miner G, Morgan M, Nazareth LV, Okwuonu G, Parker D, Pu LL, Thorn R, Wright R. The genome of the sea urchin Strongylocentrotus purpuratus. Science. 2006 Nov 10;314(5801):941-52. doi: 10.1126/science.1133609. Erratum in: Science. 2007 Feb 9;315(5813):766.

217. Sexton T, Yaffe E, Kenigsberg E, Bantignies F, Leblanc B, Hoichman M, Parrinello H, Tanay A, Cavalli G. Three-dimensional folding and functional organization principles of the Drosophila genome. Cell. 2012 Feb 3;148(3):458-72. doi: 10.1016/j.cell.2012.01.010.

218. Shen Y, Yue F, McCleary DF, Ye Z, Edsall L, Kuan S, Wagner U, Dixon J, Lee L, Lobanenkov VV, Ren B. A map of the cis-regulatory sequences in the mouse genome. Nature. 2012 Aug 2;488(7409):116-20. doi: 10.1038/nature11243.

219. Sherman RM, Salzberg SL. Pan-genomics in the human genome era. Nat Rev Genet. 2020 Apr;21(4):243-254. doi: 10.1038/s41576-020-0210-7.

220. Shiraki T, Kondo S, Katayama S, Waki K, Kasukawa T, Kawaji H, Kodzius R, Watahiki A, Nakamura M, Arakawa T, Fukuda S, Sasaki D, Podhajska A, Harbers M, Kawai J, Carninci P, Hayashizaki Y. Cap analysis gene expression for high-throughput analysis of transcriptional starting point and identification of promoter usage. Proc Natl Acad Sci U S A. 2003 Dec 23;100(26):15776-81. doi: 10.1073/pnas.2136655100.

221. Sivasubbu S, Sachidanandan C, Scaria V. Time for the zebrafish ENCODE. J Genet. 2013 Dec;92(3):695-701. doi: 10.1007/s12041-013-0313-4.

222. Skene PJ, Henikoff JG, Henikoff S. Targeted in situ genome-wide profiling with high efficiency for low cell numbers. Nat Protoc. 2018 May;13(5):1006-1019. doi: 10.1038/nprot.2018.015.

223. Snelling WM, Casas E, Stone RT, Keele JW, Harhay GP, Bennett GL, Smith TP. Linkage mapping bovine EST-based SNP. BMC Genomics. 2005 May 19;6:74. doi: 10.1186/1471-2164-6-74.

224. Snelling WM, Chiu R, Schein JE, Hobbs M, Abbey CA, Adelson DL, Aerts J, Bennett GL, Bosdet IE, Boussaha M, Brauning R, Caetano AR, Costa MM, Crawford AM, Dalrymple BP, Eggen A, Everts-van der Wind A, Floriot S, Gautier M, Gill CA, Green RD, Holt R, Jann O, Jones SJ, Kappes SM, Keele JW, de Jong PJ, Larkin DM, Lewin HA, McEwan JC, McKay S, Marra MA, Mathewson CA, Matukumalli LK, Moore SS, Murdoch B, Nicholas FW, Osoegawa K, Roy A, Salih H, Schibler L, Schnabel RD, Silveri L, Skow LC, Smith TP, Sonstegard TS, Taylor JF, Tellam R, Van Tassell CP, Williams JL, Womack JE, Wye NH, Yang G, Zhao S; International Bovine BAC

Mapping Consortium. A physical map of the bovine genome. Genome Biol. 2007;8(8):R165. doi: 10.1186/gb-2007-8-8-r165.

225. Song L, Crawford GE. DNase-seq: a high-resolution technique for mapping active gene regulatory elements across the genome from mammalian cells. Cold Spring Harb Protoc. 2010 Feb;2010(2):pdb.prot5384. doi: 10.1101/pdb.prot5384.

226. Stadler MB, Murr R, Burger L, Ivanek R, Lienert F, Schöler A, van Nimwegen E, Wirbelauer C, Oakeley EJ, Gaidatzis D, Tiwari VK, Schübeler D. DNA-binding factors shape the mouse methylome at distal regulatory regions. Nature. 2011 Dec 14;480(7378):490-5. doi: 10.1038/nature10716..

227. Stuart T, Butler A, Hoffman P, Hafemeister C, Papalexi E, Mauck WM 3rd, Hao Y, Stoeckius M, Smibert P, Satija R. Comprehensive Integration of Single-Cell Data. Cell. 2019 Jun 13;177(7):1888-1902.e21. doi: 10.1016/j.cell.2019.05.031.

228. Sulston J, Du Z, Thomas K, Wilson R, Hillier L, Staden R, Halloran N, Green P, Thierry-Mieg J, Qiu L, et al. The C. elegans genome sequencing project: a beginning. Nature. 1992 Mar 5;356(6364):37-41. doi: 10.1038/356037a0.

229. Summers KM, Bush SJ, Wu C, Su AI, Muriuki C, Clark EL, Finlayson HA, Eory L, Waddell LA, Talbot R, Archibald AL, Hume DA. Functional Annotation of the Transcriptome of the Pig, *Sus scrofa*, Based Upon Network Analysis of an RNAseq Transcriptional Atlas. Front Genet. 2020 Feb 14;10:1355. doi: 10.3389/fgene.2019.01355.

230. Sun XJ, Wei J, Wu XY, Hu M, Wang L, Wang HH, Zhang QH, Chen SJ, Huang QH, Chen Z. Identification and characterization of a novel human histone H3 lysine 36-

specific methyltransferase. J Biol Chem. 2005 Oct 21;280(42):35261-71. doi: 10.1074/jbc.M504012200.

231. Szabo Q, Bantignies F, Cavalli G. Principles of genome folding into topologically associating domains. Sci Adv. 2019 Apr 10;5(4):eaaw1668. doi: 10.1126/sciadv.aaw1668.

232. Szabó P, Tang SH, Rentsendorj A, Pfeifer GP, Mann JR. Maternal-specific footprints at putative CTCF sites in the H19 imprinting control region give evidence for insulator function. Curr Biol. 2000 May 18;10(10):607-10. doi: 10.1016/s0960-9822(00)00489-9.

233. Symmons O, Uslu VV, Tsujimura T, Ruf S, Nassari S, Schwarzer W, Ettwiller L, Spitz F. Functional and topological characteristics of mammalian regulatory domains. Genome Res. 2014 Mar;24(3):390-400. doi: 10.1101/gr.163519.113.

234. Tahiliani M, Koh KP, Shen Y, Pastor WA, Bandukwala H, Brudno Y, Agarwal S, Iyer LM, Liu DR, Aravind L, Rao A. Conversion of 5-methylcytosine to 5-hydroxymethylcytosine in mammalian DNA by MLL partner TET1. Science. 2009 May 15;324(5929):930-5. doi: 10.1126/science.1170116.

235. Tanay A, O'Donnell AH, Damelin M, Bestor TH. Hyperconserved CpG domains underlie Polycomb-binding sites. Proc Natl Acad Sci U S A. 2007 Mar 27;104(13):5521-6. doi: 10.1073/pnas.0609746104.

236. Thorne JW, Eidman L, Duan M, Hunter SS, Davenport KM, Murdoch B. PSII-27 Determining genetic variation in sheep with Flock54: a genotyping by sequencing panel. J Anim Sci. 2019 Dec;97(Suppl 3):245. doi: 10.1093/jas/skz258.498.

237. Thurman RE, Rynes E, Humbert R, Vierstra J, Maurano MT, Haugen E, Sheffield NC, Stergachis AB, Wang H, Vernot B, Garg K, John S, Sandstrom R, Bates D, Boatman L, Canfield TK, Diegel M, Dunn D, Ebersol AK, Frum T, Giste E, Johnson AK, Johnson EM, Kutyavin T, Lajoie B, Lee BK, Lee K, London D, Lotakis D, Neph S, Neri F, Nguyen ED, Qu H, Reynolds AP, Roach V, Safi A, Sanchez ME, Sanyal A, Shafer A, Simon JM, Song L, Vong S, Weaver M, Yan Y, Zhang Z, Zhang Z, Lenhard B, Tewari M, Dorschner MO, Hansen RS, Navas PA, Stamatoyannopoulos G, Iyer VR, Lieb JD, Sunyaev SR, Akey JM, Sabo PJ, Kaul R, Furey TS, Dekker J, Crawford GE, Stamatoyannopoulos JA. The accessible chromatin landscape of the human genome. Nature. 2012 Sep 6;489(7414):75-82. doi: 10.1038/nature11232.

238. Tippens ND, Vihervaara A, Lis JT. Enhancer transcription: what, where, when, and why? Genes Dev. 2018 Jan 1;32(1):1-3. doi: 10.1101/gad.311605.118.

239. Torrano V, Chernukhin I, Docquier F, D'Arcy V, León J, Klenova E, Delgado MD. CTCF regulates growth and erythroid differentiation of human myeloid leukemia cells. J Biol Chem. 2005 Jul 29;280(30):28152-61. doi: 10.1074/jbc.M501481200.

240. Tuggle CK, Giuffra E, White SN, Clarke L, Zhou H, Ross PJ, Acloque H, Reecy JM, Archibald A, Bellone RR, Boichard M, Chamberlain A, Cheng H, Crooijmans RP, Delany ME, Finno CJ, Groenen MA, Hayes B, Lunney JK, Petersen JL, Plastow GS, Schmidt CJ, Song J, Watson M. GO-FAANG meeting: a Gathering On Functional Annotation of Animal Genomes. Anim Genet. 2016 Oct;47(5):528-33. doi: 10.1111/age.12466.

241. Van Laere AS, Nguyen M, Braunschweig M, Nezer C, Collette C, Moreau L, Archibald AL, Haley CS, Buys N, Tally M, Andersson G, Georges M, Andersson L. A regulatory mutation in IGF2 causes a major QTL effect on muscle growth in the pig. Nature. 2003 Oct 23;425(6960):832-6. doi: 10.1038/nature02064.

242. Vietri Rudan M, Barrington C, Henderson S, Ernst C, Odom DT, Tanay A, Hadjur S. Comparative Hi-C reveals that CTCF underlies evolution of chromosomal domain architecture. Cell Rep. 2015 Mar 3;10(8):1297-309. doi: 10.1016/j.celrep.2015.02.004.

243. Villar D, Berthelot C, Aldridge S, Rayner TF, Lukk M, Pignatelli M, Park TJ, Deaville R, Erichsen JT, Jasinska AJ, Turner JM, Bertelsen MF, Murchison EP, Flicek P, Odom DT. Enhancer evolution across 20 mammalian species. Cell. 2015 Jan 29;160(3):554-66. doi: 10.1016/j.cell.2015.01.006.

244. Walsh CP, Chaillet JR, Bestor TH. Transcription of IAP endogenous retroviruses is constrained by cytosine methylation. Nat Genet. 1998 Oct;20(2):116-7. doi: 10.1038/2413.

245. Wang C, Liu C, Roqueiro D, Grimm D, Schwab R, Becker C, Lanz C, Weigel D. Genome-wide analysis of local chromatin packing in Arabidopsis thaliana. Genome Res. 2015 Feb;25(2):246-56. doi: 10.1101/gr.170332.113.

246. Wang S, Su JH, Beliveau BJ, Bintu B, Moffitt JR, Wu CT, Zhuang X. Spatial organization of chromatin domains and compartments in single chromosomes. Science. 2016 Aug 5;353(6299):598-602. doi: 10.1126/science.aaf8084.

247. Weintraub AS, Li CH, Zamudio AV, Sigova AA, Hannett NM, Day DS, Abraham BJ, Cohen MA, Nabet B, Buckley DL, Guo YE, Hnisz D, Jaenisch R, Bradner JE, Gray NS, Young RA. YY1 Is a Structural Regulator of Enhancer-Promoter Loops. Cell. 2017 Dec 14;171(7):1573-1588.e28. doi: 10.1016/j.cell.2017.11.008.

248. Weir BS, Cockerham CC. Estimating F-statistics for the analysis of population structure. Evolution. 1984 Nov;38(6):1358-1370. doi: 10.1111/j.1558-5646.1984.tb05657.x.

249. Weischenfeldt J, Dubash T, Drainas AP, Mardin BR, Chen Y, Stütz AM, Waszak SM, Bosco G, Halvorsen AR, Raeder B, Efthymiopoulos T, Erkek S, Siegl C, Brenner H, Brustugun OT, Dieter SM, Northcott PA, Petersen I, Pfister SM, Schneider M, Solberg SK, Thunissen E, Weichert W, Zichner T, Thomas R, Peifer M, Helland A, Ball CR, Jechlinger M, Sotillo R, Glimm H, Korbel JO. Pan-cancer analysis of somatic copy-number alterations implicates IRS4 and IGF2 in enhancer hijacking. Nat Genet. 2017 Jan;49(1):65-74. doi: 10.1038/ng.3722.

250. Williams BD, Schrank B, Huynh C, Shownkeen R, Waterston RH. A genetic mapping system in Caenorhabditis elegans based on polymorphic sequence-tagged sites. Genetics. 1992 Jul;131(3):609-24.

251. Wilson DE, Morrical DG. The National Sheep Improvement Program: a review. J Anim Sci. 1991 Sep;69(9):3872-81. doi: 10.2527/1991.6993872x.

252. Wilson T, Wu XY, Juengel JL, Ross IK, Lumsden JM, Lord EA, Dodds KG, Walling GA, McEwan JC, O'Connell AR, McNatty KP, Montgomery GW. Highly prolific Booroola sheep have a mutation in the intracellular kinase domain of bone morphogenetic protein IB receptor (ALK-6) that is expressed in both oocytes and granulosa cells. Biol Reprod. 2001 Apr;64(4):1225-35. doi: 10.1095/biolreprod64.4.1225.

253. de Wit E, Vos ES, Holwerda SJ, Valdes-Quezada C, Verstegen MJ, Teunissen H, Splinter E, Wijchers PJ, Krijger PH, de Laat W. CTCF Binding Polarity Determines Chromatin Looping. Mol Cell. 2015 Nov 19;60(4):676-84. doi: 10.1016/j.molcel.2015.09.023.

254. Wright S. The interpretation of population structure by F-statistics with special regard to systems of mating. Evolution. 1965;19:395-420.

255. Wu R, Shi ZR. Comparison of chromogenic in situ hybridization, fluorescence in situ hybridization, and immunohistochemistry. Handbook of Immunohistochemistry and in situ hybridization of human carcinomas, Elsevier Academic Press. 2002;13-26.

256. Wutz G, Várnai C, Nagasaka K, Cisneros DA, Stocsits RR, Tang W, Schoenfelder S, Jessberger G, Muhar M, Hossain MJ, Walther N, Koch B, Kueblbeck M, Ellenberg J, Zuber J, Fraser P, Peters JM. Topologically associating domains and chromatin loops depend on cohesin and are regulated by CTCF, WAPL, and PDS5 proteins. EMBO J. 2017 Dec 15;36(24):3573-3599. doi: 10.15252/embj.201798004.

257. Xu SS, Gao L, Xie XL, Ren YL, Shen ZQ, Wang F, Shen M, Eyþórsdóttir E, Hallsson JH, Kiseleva T, Kantanen J, Li MH. Genome-Wide Association Analyses Highlight the Potential for Different Genetic Mechanisms for Litter Size Among Sheep Breeds. Front Genet. 2018 Apr 10;9:118. doi: 10.3389/fgene.2018.00118.

258. Yan F, Powell DR, Curtis DJ, Wong NC. From reads to insight: a hitchhiker's guide to ATAC-seq data analysis. Genome Biol. 2020 Feb 3;21(1):22. doi: 10.1186/s13059-020-1929-3.

259. Yin Y, Morgunova E, Jolma A, Kaasinen E, Sahu B, Khund-Sayeed S, Das PK, Kivioja T, Dave K, Zhong F, Nitta KR, Taipale M, Popov A, Ginno PA, Domcke S, Yan J, Schübeler D, Vinson C, Taipale J. Impact of cytosine methylation on DNA binding specificities of human transcription factors. Science. 2017 May 5;356(6337):eaaj2239. doi: 10.1126/science.aaj2239..

260. Young R, Lefevre L, Bush SJ, Joshi A, Singh SH, Jadhav SK, Dhanikachalam V, Lisowski ZM, Iamartino D, Summers KM, Williams JL, Archibald AL, Gokhale S, Kumar S, Hume DA. A Gene Expression Atlas of the Domestic Water Buffalo (*Bubalus bubalis*). Front Genet. 2019 Jul 24;10:668. doi: 10.3389/fgene.2019.00668.

261. Yue F, Cheng Y, Breschi A, Vierstra J, Wu W, Ryba T, Sandstrom R, Ma Z, Davis C, Pope BD, Shen Y, Pervouchine DD, Djebali S, Thurman RE, Kaul R, Rynes E, Kirilusha A, Marinov GK, Williams BA, Trout D, Amrhein H, Fisher-Aylor K, Antoshechkin I, DeSalvo G, See LH, Fastuca M, Drenkow J, Zaleski C, Dobin A, Prieto P, Lagarde J, Bussotti G, Tanzer A, Denas O, Li K, Bender MA, Zhang M, Byron R, Groudine MT, McCleary D, Pham L, Ye Z, Kuan S, Edsall L, Wu YC, Rasmussen MD, Bansal MS, Kellis M, Keller CA, Morrissey CS, Mishra T, Jain D, Dogan N, Harris RS, Cayting P, Kawli T, Boyle AP, Euskirchen G, Kundaje A, Lin S, Lin Y, Jansen C, Malladi VS, Cline MS, Erickson DT, Kirkup VM, Learned K, Sloan CA, Rosenbloom KR, Lacerda de Sousa B, Beal K, Pignatelli M, Flicek P, Lian J, Kahveci T, Lee D, Kent WJ, Ramalho Santos M, Herrero J, Notredame C, Johnson A, Vong S, Lee K, Bates D, Neri F, Diegel M, Canfield T, Sabo PJ, Wilken MS, Reh TA, Giste E, Shafer A, Kutyavin T, Haugen E, Dunn D, Reynolds AP, Neph S, Humbert R, Hansen RS, De Bruijn M, Selleri L, Rudensky A, Josefowicz S, Samstein R, Eichler EE, Orkin SH, Levasseur D, Papayannopoulou T, Chang KH, Skoultchi A, Gosh S, Disteche C, Treuting P, Wang Y, Weiss MJ, Blobel GA, Cao X, Zhong S, Wang T, Good PJ, Lowdon RF, Adams LB, Zhou XQ, Pazin MJ, Feingold EA, Wold B, Taylor J, Mortazavi A, Weissman SM, Stamatoyannopoulos JA, Snyder MP, Guigo R, Gingeras TR, Gilbert DM, Hardison RC, Beer MA, Ren B; Mouse ENCODE Consortium. A comparative encyclopedia of DNA elements in the mouse genome. Nature. 2014 Nov 20;515(7527):355-64. doi: 10.1038/nature13992.

262. Zemach A, McDaniel IE, Silva P, Zilberman D. Genome-wide evolutionary analysis of eukaryotic DNA methylation. Science. 2010 May 14;328(5980):916-9. doi: 10.1126/science.1186366.

263. Zerbino DR, Birney E. Velvet: algorithms for de novo short read assembly using de Bruijn graphs. Genome Res. 2008 May;18(5):821-9. doi: 10.1101/gr.074492.107.

264. Zhang L, Liu J, Zhao F, Ren H, Xu L, Lu J, Zhang S, Zhang X, Wei C, Lu G, Zheng Y, Du L. Genome-wide association studies for growth and meat production traits in sheep. PLoS One. 2013 Jun 25;8(6):e66569. doi: 10.1371/journal.pone.0066569.

265. Zhang L, Mousel MR, Wu X, Michal JJ, Zhou X, Ding B, Dodson MV, El-Halawany NK, Lewis GS, Jiang Z. Genome-wide genetic diversity and differentially selected regions among Suffolk, Rambouillet, Columbia, Polypay, and Targhee sheep. PLoS One. 2013 Jun 10;8(6):e65942. doi: 10.1371/journal.pone.0065942.

266. Zhang T, Cooper S, Brockdorff N. The interplay of histone modifications - writers that read. EMBO Rep. 2015 Nov;16(11):1467-81. doi: 10.15252/embr.201540945.

267. Zhang Y, Jurkowska R, Soeroes S, Rajavelu A, Dhayalan A, Bock I, Rathert P, Brandt O, Reinhardt R, Fischle W, Jeltsch A. Chromatin methylation activity of Dnmt3a and Dnmt3a/3L is guided by interaction of the ADD domain with the histone H3 tail. Nucleic Acids Res. 2010 Jul;38(13):4246-53. doi: 10.1093/nar/gkq147.

268. Zhao C, Carrillo JA, Tian F, Zan L, Updike SM, Zhao K, Zhan F, Song J. Genome-Wide H3K4me3 Analysis in Angus Cattle with Divergent Tenderness. PLoS One. 2015 Jun 18;10(6):e0115358. doi: 10.1371/journal.pone.0115358.

269. Zhou S, Goldstein S, Place M, Bechner M, Patino D, Potamousis K, Ravindran P, Pape L, Rincon G, Hernandez-Ortiz J, Medrano JF, Schwartz DC. A clone-free, single molecule map of the domestic cow (Bos taurus) genome. BMC Genomics. 2015 Aug 28;16(1):644. doi: 10.1186/s12864-015-1823-7.

270. Zimin AV, Delcher AL, Florea L, Kelley DR, Schatz MC, Puiu D, Hanrahan F, Pertea G, Van Tassell CP, Sonstegard TS, Marçais G, Roberts M, Subramanian P, Yorke JA, Salzberg SL. A whole-genome assembly of the domestic cow, Bos taurus. Genome Biol. 2009;10(4):R42. doi: 10.1186/gb-2009-10-4-r42.

271. Zimin AV, Marçais G, Puiu D, Roberts M, Salzberg SL, Yorke JA. The MaSuRCA genome assembler. Bioinformatics. 2013 Nov 1;29(21):2669-77. doi: 10.1093/bioinformatics/btt476.

Chapter 2: Genetic Structure and Admixture in Sheep from Terminal Breeds in the United States

Published in Animal Genetics.

Citation: Davenport KM, Hiemke C, McKay SD, Thorne JW, Lewis RM, Taylor T, Murdoch BM. Genetic structure and admixture in sheep from terminal breeds in the United States. Anim Genet. 2020 Mar;51(2):284-291. doi: 10.1111/age.12905.

KM Davenport[1], C Hiemke[2], SD McKay[3], JW Thorne[4,1], RM Lewis[5], T Taylor[6], BM Murdoch[1]

[1]Department of Animal and Veterinary Science, University of Idaho, Moscow, 83844, United States

[2]Niman Ranch and Mapleton Mynd Shropshires, Stoughton, 53589, United States

[3]Department of Animal and Veterinary Sciences, University of Vermont, Burlington, 05405, United States

[4]Texas A&M AgriLife Extension, San Angelo, 76901, United States

[5]Department of Animal Science, University of Nebraska-Lincoln, Lincoln, Nebraska 68583, United States

[6]Department of Animal Science, Arlington Research Station, University of Wisconsin-Madison, Arlington, Wisconsin 53911, United States

Corresponding Author
BM Murdoch
208-885-2088
bmurdoch@uidaho.edu

Summary

Selection for performance in diverse production settings has resulted in variation across sheep breeds worldwide. Although sheep are an important species to the United States (U.S.), the current genetic relationship among many terminal sire breeds is not well characterized. Suffolk, Hampshire, Shropshire, and Oxford (terminal) and Rambouillet (dual purpose) sheep (n=248) sampled from different flocks were genotyped using the Applied Biosystems Axiom Ovine Genotyping Array (50K), and additional Shropshire (n=26) using the Illumina Ovine SNP50 BeadChip. Relationships were investigated by calculating observed heterozygosity, inbreeding coefficients, eigenvalues, pairwise Wright's F_{ST} estimates, and an identity by state (IBS) matrix. The mean observed heterozygosity for each breed ranged from 0.30 to 0.35 and is consistent with data reported in other U.S. and Australian sheep. Suffolk from two different regions of the U.S. (Midwest and West) clustered separately in eigenvalue plots and the rectangular cladogram. Further, divergence was detected between Suffolk from different regions with Wright's F_{ST} estimate. Shropshire animals showed the greatest divergence from other terminal breeds in this study. Admixture between breeds was examined using ADMIXTURE, and based on cross validation estimates, the best fit number of populations (clusters) was K=6. The greatest admixture was observed within Hampshire, Suffolk, and Shropshire breeds. When plotting eigenvalues, U.S. terminal breeds clustered separately in comparison to sheep from other locations of the world. Understanding the genetic relationships between terminal sire breeds in sheep will inform us about the potential applicability of markers derived in one breed to other breeds based on relatedness.

Keywords genetic admixture, sheep, terminal sheep breeds, genetic relationships

Introduction

The production of lamb and wool is an important agricultural industry in the United States (U.S.), with approximately 5 million sheep and 80,000 operations (USDA ERS 2019). According to the American Sheep Industry National Animal Health Monitoring System's most recent study, 81.6% of operations raise sheep for meat purposes (American Sheep Industry 2011). The most popular breeds used for meat production include Suffolk, Hampshire, Shropshire, Oxford, and Southdown (American Sheep Industry 2011). To make progress in their own flocks, some U.S. lamb and wool producers have implemented quantitative genetic selection strategies using estimated breeding values (EBV) through the National Sheep Improvement Program (NSIP) to identify and select animals with desirable traits (Wilson & Morrical 1991; Notter 1998; Lupton 2008). As this program is more widely utilized, improvement of product quality and yield of lamb and wool products in the U.S. is anticipated to accelerate.

Previous research indicates that selection for various traits such as wool or growth within breeds of sheep has led to greater breed specialization across the world (Kijas *et al.* 2012; Zhang *et al.* 2013). However, many breeds of sheep have retained greater heterozygosity in comparison to other species, including cattle (Bovine HapMap Consortium *et al.* 2009; Kijas *et al.* 2012). Furthermore, sheep from similar locations have been reported to have high levels of admixture (Blackburn *et al.* 2011; Kijas *et al.* 2012).

The current genetic structure and level of admixture among terminal sire breeds in the U.S. has not been well characterized (Zhang *et al*. 2013). The objective of this study was to examine population structure and admixture in sheep from terminal breeds from U.S. sheep operations in collaboration with producers engaged with NSIP. Understanding the genetic relationships between terminal sire breeds in the U.S. will allow us to better understand the genetic relatedness of these breeds of sheep and assess potential applicability of information based on breed relatedness. Further, this study can help elucidate how biological differences segregate in different breeds, as well as between breeds of sheep.

Materials and methods

Sample collection and DNA isolation

A total of 248 sheep from terminal breeds of sheep including Hampshire (n = 45 from 6 flocks), Suffolk (n = 68 from 9 flocks in the Midwest and n = 37 from one flock, the University of Idaho Suffolk flock, in the West), Oxford (n = 11 from 2 flocks), and Shropshire (n = 44 from 5 flocks), as well as wool/dual purpose Rambouillet (n=43 from one flock) were genotyped for this study. Blood, semen, or tissue samples were collected by individual producers and shipped to the University of Idaho and DNA was isolated using the phenol chloroform method previously described (Sambrook *et al.* 1989).

Genotyping and quality control

Samples were genotyped using the Applied Biosystems^{TM} Axiom^{TM} Ovine Genotyping Array (50K) consisting of 51,572 single nucleotide polymorphisms (SNPs) (Thermo Fisher Scientific, catalog number 550898). A subset of Shropshire samples (n = 26) previously genotyped on the Ovine Illumina SNP50 Bead Chip consisting of 54,241 SNPs (Illumina catalogue number WG-420-1001) were also included in this dataset. The genotypic data for these samples, from each platform, were merged by SNP name and location in PLINK v1.90, with a total of 47,485 SNPs overlapping between the two panels. Quality control of genotype data was performed using PLINK v1.90 specifically excluding SNPs with a call rate of less than 0.90 and minor allele frequency less than 0.01, resulting in 45,864 SNPs remaining in the analyses (Purcell *et al.* 2007; Chang *et al.* 2015).

Observed heterozygosity, inbreeding coefficients, and F_{ST} calculations

The observed heterozygosity was estimated for each animal using PLINK v1.90 and averaged by breed (Purcell *et al*. 2007; Chang *et al.* 2015). Inbreeding coefficients were calculated for each animal based on the observed and expected homozygosity in PLINK v1.90, and the mean and 95% confidence interval were calculated with the R package ‘rcompanion’ in R version 3.6.1. To remove redundancy and provide a more accurate representation of variation, linkage disequilibrium (LD) pruning was performed using the --indep-pairwise function in PLINK v1.90 with an r^2 = 0.5, sliding window size of 50 SNPs, and shifts of 5 SNPs (Visser *et al.* 2016; Gilbert *et al*. 2017). After LD pruning, 40,121 SNPs remained for further analyses. Pairwise F_{ST} was estimated in PLINK v1.90 between breeds of sheep using the LD pruned dataset (Purcell *et al.* 2007; Chang *et al.* 2015).

Eigenvalue analyses

Eigenvalues were calculated using the filtered SNP dataset for terminal breeds only and then with Rambouillet in SNP and Variation Suite (SVS) version 8.7.2 (Golden Helix, Inc., Bozeman, MT, www.goldenhelix.com). The top two eigenvalues were plotted against each other in SVS.

Hierarchical clustering

An identity by state matrix (IBS) was calculated from the LD pruned dataset pairwise between all sheep using PLINK v1.90 --distance flag (Purcell *et al.* 2007; Chang *et al.* 2015). The matrix was read into R version 3.5.2 and hierarchical clustering based on the IBS matrix of Hamming distances between each animal using the 'hclust' function. The Bioconductor package 'ctc' was used in R version 3.5.1 to write a Newick file to import into Dendroscope 3 software (Huson & Scornavacca 2012). A rectangular cladogram was drawn from the Newick file in Dendroscope version 3.5.9 (Huson & Scornavacca 2012). Individual branch labels were coloured according to producer reported breed of sheep.

Admixture analysis

The program ADMIXTURE version 1.3.0 was implemented to examine admixture between all samples using the LD pruned genotypes in BED format (Alexander *et al*. 2009; Decker *et al*. 2014). The most probably number of K given populations was estimated using the lowest cross-validation error (Alexander *et al*. 2009; Akanno *et al*. 2018). Euclidean distances were calculated in R version 3.6.1 with the *adegenet* package and an analysis of molecular variance (AMOVA) was performed with the *pegas* package with 1000 permutations to statistically examine differences between populations (McKay *et al.* 2008; Paradis 2010; Jombart & Ahmed 2011).

International Breed Comparisons

Genotypes from 2,819 sheep from 74 breeds across the world were retrieved from the International Sheep Genome Consortium Sheep HapMap Database and used in comparison with U.S. terminal breeds including the addition of n=5 Dorset and n=7 Southdown from the U.S. The same set of 45,864 SNPs used with the U.S. terminal breeds were then merged with the same SNPs from the Sheep HapMap dataset. Eigenvalues were calculated between U.S. terminal breeds and the same breeds from other locations in the HapMap dataset, all U.S. breeds in this study and the same breeds present from other locations in the HapMap dataset, and all U.S. breeds in this study and the Sheep HapMap dataset.

Results

Observed Heterozygosity and Inbreeding Coefficient

To examine the relatedness of animals within each of the breeds, observed heterozygosity and average inbreeding coefficient were calculated. These statistics were calculated based on observed and expected homozygosity, estimated for each individual, and averaged for each breed (Table 2.1). The Oxford animals exhibited the greatest (0.35) observed heterozygosity and lowest inbreeding coefficients. Similar observed heterozygosity was exhibited by Shropshire (0.34), Western Suffolk (0.34), Suffolk (0.33), and Hampshire (0.33). Shropshire had the lowest inbreeding coefficient (0.09) in comparison to the Suffolk (0.13), Western Suffolk (0.14), and Hampshire (0.14). The group with the lowest observed heterozygosity (0.30) and highest inbreeding coefficient (0.16) was Rambouillet.

Wright's F_{ST}

Wright's F_{ST} was calculated pairwise between each group of animals to examine differentiation between breeds (Table 2.2) (Wright 1965; Weir & Cockerham 1984; Lenstra *et al.* 2012). In general, values between 0 and 0.05 are categorized as "little to no differentiation," values between 0.05 and 0.15 as "moderate differentiation", values between 0.15 and 0.25 as "great differentiation", and values above 0.25 as "very great differentiation" between populations tested (Weir & Cockerham 1984; Frankham *et al.* 2002). Rambouillet is considered greatly differentiated from all terminal breeds. Interestingly, Western Suffolk are considered moderately differentiated from other terminal breeds. Little to no difference was detected between Hampshire and Suffolk or Hampshire and Shropshire. Furthermore, although Western Suffolk and other Suffolk are not reported as different breeds, they too exhibit moderate differentiation.

Eigenvalue Analyses

To investigate how individuals from reported terminal breeds the U.S. group or cluster, eigenvalues were calculated and plotted for all samples (Figure 2.1). An eigenvalue plot for only terminal breeds of sheep (Figure 2.1A) as well as terminal breeds and Rambouillet sheep (Figure 2.1B) is displayed. In Figure 1A, the largest difference of eigenvalues is between Western Suffolk and Shropshire and can be observed on the x-axis of the plot shown. Further, the animals sampled for the Shropshire breed exhibited the largest spread of eigenvalue points. Interestingly, all Suffolk did not group together. The Suffolk animals sampled from most cluster closely with Hampshire animals, however, the Western Suffolk flock clusters separately from Hampshire and other Suffolk animals.

In Figure 2.1B, Rambouillet animals cluster together, and the entire breed clusters distinctly and away from the terminal sheep breeds on the largest eigenvalue axis. Similar to Figure 2.1A, sheep clustered primarily by breed with the exception of four Shropshire animals. The Suffolk samples do not all group together, with Western Suffolk clustering separately from other Suffolk animals. With these notable exceptions, animals within a breed clustered together.

Hierarchical Clustering Based on Identity by State

To examine how animals from breeds of sheep in the U.S. related to other breeds, hierarchical clustering was performed using an identity by state matrix. A rectangular cladogram was constructed to visualize the hierarchical clustering (Figure 2.2). All Western Suffolk, Oxford, and Rambouillet animals cluster together by breed. Rambouillet animals cluster in a distinct, separate branch from all other breeds, which is consistent with the eigenvalue plot. In general, most sheep are more identical by state to other animals within the same breed with a few notable exceptions.

Several reported Shropshire animals cluster with the Hampshire branches; these are the same animals that clustered with the Hampshire breed in the eigenvalue plots. A branch of Shropshire animals also clusters closely with a larger branch of Hampshire sheep. Additionally, Suffolk and Hampshire animals overlap and appear to cluster closely within the branches of the cladogram. Still, overall most breeds cluster independently with the few before mentioned exceptions.

Admixture Analysis

An admixture analysis was performed using the program ADMIXTURE to investigate the extent of admixture between different breeds of sheep in this study (Alexander *et al*. 2009; Decker *et al.* 2014; Getachew *et al.* 2017). The analysis was conducted using 2 through 10 given populations. The best fit of K given populations was determined as K=6 based on the cross-validation (CV) values calculated in ADMIXTURE (Supplemental Figure 2.1) (Akanno *et al*. 2018). Further, the AMOVA analyses showed significant (*P*<0.01) differences between the K=6 assigned populations.

In the best fit K=6 plot, admixture was detected within terminal breeds (Figure 2.3). Admixture between terminal breeds was observed in Hampshire, Oxford, Suffolk, and Shropshire, but the Western Suffolk population showed little admixture with other terminal breeds except Suffolk. Not surprisingly, the dual purpose Rambouillet sheep were different than the U.S. terminal breeds examined.

Eigenvalue Plots of U.S. and International Comparisons

To examine how U.S. sheep compare to other sheep across the world, genotyping data from this study was merged with data from the Sheep HapMap (Kijas *et al.* 2012; Kijas 2013). Eigenvalues were calculated and plotted with U.S. terminal breeds including additional Dorset and Southdown sheep from the U.S., and animals of the same breeds from the Sheep HapMap dataset (Figure 2.4A). Interestingly, the U.S. terminal breeds cluster closer to other breeds from the U.S. than the same reported breed, including Suffolk and Dorset, from other locations. When the genetic information for wool breeds of sheep are included, they cluster apart from the terminal breeds (Figure 2.4B). Figure 2.4B also shows the Irish Suffolk clustering closely with Suffolk from the U.S. Finally, when all samples are considered, the U.S. terminal breeds cluster with similar breeds from Australia and the United Kingdom (Figure 2.4C). In summary, animals cluster closest with those of similar geographic location in the eigenvalue plots.

Discussion

The observed heterozygosity results from this study are consistent with data reported in other breeds of sheep across the world (Kijas *et al*. 2012; Ciani *et al.* 2013; Gaouar *et al.* 2017). More specifically, the observed heterozygosity in most breeds was close to what was reported in Australian sheep (Kijas *et al.* 2012; Al-Mamun *et al.* 2015). In addition, the observed heterozygosity is consistent with other U.S. sheep including Suffolk, Rambouillet, Columbia, Polypay, and Targhee (Zhang *et al.* 2013). However, the breeds in this study had lower observed heterozygosity when compared to Boutsko, Karagouniko, and Chios breeds from Greece (Michailidou *et al.* 2018).

In our study, Oxford sheep exhibited the lowest average inbreeding coefficient and highest observed heterozygosity, similar to Finnsheep (Li *et al*. 2011). This is likely because these sheep were selected based on pedigree diversity from NSIP, whereas Western Suffolk had one of the highest inbreeding coefficients and is only represented by one flock. However, to our surprise the inbreeding coefficient for Western Suffolk was similar to Suffolk, which included animals from 10 separate flocks. Perhaps this is because these animals are the result and representative of the breeding strategies of purebred flocks. Other work in 97 sheep breeds across the world and Ethiopian sheep reported inbreeding coefficients between -0.07-0.16 and observed heterozygosity between 0.061-0.343, which was similar to our results (Edea *et al.* 2017; Zhang *et al.* 2018).

Despite similarity in inbreeding coefficient and heterozygosity estimates, Western Suffolk shows moderate differentiation from Suffolk whereas Hampshire, Oxford, Shropshire, and Suffolk show little to moderate differentiation from each other. The Western Suffolk consists of representatives from a “closed flock”, which may explain the divergence from the more broadly sampled Suffolk. The lack of differentiation observed between the Suffolk, Hampshire, and Shropshire is not surprising considering the prevalence of crossbreeding in many U.S. terminal breed flocks. It is worth noting that Southdown is thought to be a common ancestor for Hampshire, Shropshire, and Oxford breeds ancestor (Ryder 1964). These points are strongly supported by the results of the ADMIXTURE analysis. Furthermore, these results concur with previous research that reported a Wright’s F_{ST} = 0.1621 between Suffolk and Rambouillet, these breeds differ in origin as the Rambouillet breed was derived from Merino bloodlines (Dickinson & Lush 1933; Zhang *et al.* 2013).

Differences between breed groups can be visualized in the eigenvalue plots, where sheep cluster primarily by reported breed with the exception of a few animals. The separation of Suffolk from Western Suffolk is apparent, which is consistent with previous work that identified regional differences in Suffolk from the U.S. (Kuehn *et al.* 2008). The Shropshire breed has a large spread of eigenvalues and a few animals cluster with Oxford and Hampshire, suggesting the occurrence of crossbreeding. The distinct clustering of the Rambouillet away from other breeds clearly displays the genetic difference between terminal and wool/dual-purpose breeds in the U.S.

The K=6 plot, supported by the AMOVA analysis, shows sheep cluster primarily by breed with some level of admixture between all terminal breeds, with the exception of Western Suffolk which exhibits little admixture except with other Suffolk. The observed admixture within Hampshire, Suffolk, Oxford, and Shropshire is potentially due to the use of sires with composite influence from other breeds in U.S. commercial operations (Ercanbrack & Knight 1991; Norberg & Sørensen 2007). Rambouillet sheep showed little to no admixture with the U.S. terminal breeds examined in this study.

When U.S. sheep were compared with other populations across the world, sheep primarily cluster closest to other animals in similar geographic locations than to the same reported breeds in other parts of the world (Kijas *et al.* 2012). More specifically, Suffolk and Dorset animals cluster closer to other U.S. groups than to Suffolk from Australia and Ireland, or

Dorset from Australia or the United Kingdom. This observation may be partially attributed to the differences in selection and breeding strategies and in production systems across the world (Andersson 2012; Ćurković et al. 2014; Wang *et al.* 2015). In addition, the difference between terminal breeds and wool breeds is clearly observed, suggesting that there are clear genetic differences between breeds that have been selected for alternate production objectives and purposes (Blackburn *et al.* 2011; Zhang *et al.* 2013; Fariello *et al.* 2014).

In summary, we characterized relationships between sheep from terminal sire breed populations in the U.S. Internationally, there has been an increased emphasis on genetic selection of sheep for a variety of traits and purposes. Marker assisted selection is growing in popularity as new technology is rapidly developed, along with an increase in use of quantitative genetic programs that calculate EBVs. By better understanding the population structure and admixture between terminal breeds in the U.S. compared to breeds across the world, we can improve the effectiveness of this developing technology. Our research provides insight into current relatedness of the popular terminal breeds in the U.S. and the framework for future analyses on a larger scale.

Acknowledgements

This project was supported by Agriculture and Food Research Initiative Hatch grant #IDA01566 from the USDA National Institute of Food and Agriculture. We would like to thank Thermo Fisher Scientific for genotyping; and Curt Stanley, Shane Kirtchen, Virginia Tech/Scott Greiner, Hayden’s Hamps, North Dakota State University, Drewry, Reombke, Adams, University of Wisconsin, Kindred Crossing, Bingen, Mapleton Mynd, Knepp Shropshires, All Forage Farm, Dr. Fred Groverman, Formo, Bishop, Bar-Zel Suffolks, JMG Suffolks, Virginia Tech, Culham & Stevens, Reau Suffolks, Double L Livestock, and Dry Sandy for sample contribution.

Availability of data

Data (50K SNP) has been deposited in Open Science Framework (https://osf.io/d7s59/?view_only=9c85566d0ac542d89a62150524eaad0e).

References

1. Akanno E.C., Chen L., Abo-Ismail M.K., Crowley J.J., Wang Z., Li C., Basarab J.A., MacNeil M.D. & Plastow G.S. (2018) Genome-wide association scan for heterotic quantitative trait loci in multi-breed and crossbred beef cattle. Genetics, selection, and evolution: GSE 50, 48.

2. Al-Mamun H.A., Clark S.A., Kwan P. & Gondro C. (2015) Genome-wide linkage disequilibrium and genetic diversity in five populations of Australian domestic sheep. Genetics, selection, evolution: GSE 47, 90.

3. Alexander D.H., Novembre J. & Lange K. (2009) Fast model-based estimation of ancestry in unrelated individuals. Genome Research 19, 1655-64.

4. American Sheep Industry (2011) National Animal Health Monitoring System (NAHMS) Sheep 2011 Study. https://sheepusa.org/researcheducation-animalhealth-nahms.

5. Andersson L. (2012) How selective sweeps in domesticated animals provide new insight into biological mechanisms. Journal of Internal Medicine 271, 1-14.

6. Blackburn H.D., Paiva S.R., Wildeus S., Getz W., Waldron D., Stobart R., Bixby D., Purdy P.H., Welsh C., Spiller S. & Brown M. (2011) Genetic structure and diversity among sheep breeds in the United States: Identification of the major gene pools. Journal of Animal Science 89, 2336–48.

7. Bovine HapMap Consortium, Gibbs R.A., Taylor J.F. *et al.* (2009) Genome-wide survey of SNP variation uncovers the genetic structure of cattle breeds. Science 324, 528-32.

8. Chang C.C., Chow C.C., Tellier L.C., Vattikuti S., Purcell S.M. & Lee J.J. (2015) Second-generation PLINK: rising to the challenge of larger and richer datasets. Gigascience 4, 7.

9. Ciani E., Crepaldi P., Nicoloso L., Lasagna E., Sarti F.M., Moioli B., Napolitano F., Carta A., Usai G., D'Andrea M., Marletta D., Ciampolini R., Riggio V., Occidente M., Matassino D., Kompan D., Modesto P., Macciotta N., Ajmone-Marsan P. & Pilla F. (2014) Genome-wide analysis of Italian sheep diversity reveals a strong geographic pattern and cryptic relationship between breeds. Animal Genetics 45, 256-66.

10. Ćurković M., Ramljak J., Ivanković S., Mioč B., Ianković A., Pavić V., Brka M., Veit-Kensch C. & Medugorac I. (2015) The genetic diversity and structure of 18 sheep breeds exposed to isolation and selection. Journal of Animal Breeding and Genetics. 133, 71-80.

11. Decker J.E., McKay S.D., Rolf M.M., Kim J., Molina Alcalá A., Sonstegard T.S., Hanotte O., Götherström A., Seabury C.M., Praharani L., Babar M.E., Correia de Almeida Regitano L., Yildiz M.A., Heaton M.P., Liu W.S., Lei C.Z., Reecy J.M., Saif-Ur-Rehman M., Schnabel R.D. & Taylor J.F. (2014) Worldwide patterns of ancestry, divergence, and admixture in domesticated cattle. PLoS Genetics 10, e1004254.

12. Dickinson W.F. & Lush J.L. (1933) Inbreeding and genetic diversity of Rambouillet sheep in America. Journal of Heredity 24, 19-33.

13. Edea Z., Dessie T., Dadi H., Do K.T. & Kim K.S. (2017) Genetic diversity and population structure of Ethiopian sheep populations revealed by high-density SNP markers. Frontiers in Genetics 8, 218.

14. Fariello M.I., Bertrand S., Tosser-Klopp G., Rupp R., Moreno C., International Sheep Genomics Consortium, San Cristobal M. & Boitard S. (2014) Selection signatures in worldwide sheep popualations. PLoS ONE 9, e103813.

15. Frankham R., Ballou J.D. & Briscou D.A. (2002) Introduction to Conservation Genetics. Cambridge University Press, Cambridge, UK.

16. Gaouar S.B.S., Lafri M., Djaout A., El-Bouyahiaoui R., Bouri A., Bouchatal A., Maftah A., Ciani E. & Da Silva A.B. (2017) Genome-wide analysis highlights genetic dilution in Algerian sheep. Heredity 118, 293-301.

17. Getachew T., Huson H.J., Wurzinger M., Burgstaller J., Gizaw S., Haile A., Rischkowsky B., Brem G., Boison S.A., Mészáros G., Mwai A.O. & Sölkner J. (2017) Identifying highly informative genetic markers for quantification of ancestry proportions in crossbred sheep populations: implications for choosing optimum levels of admixture. BMC Genetics 18, 80.

18. Gilbert E., Carmi S., Ennis S., Wilson J.F. & Cavaller G.L. (2017) Genomic insights into the population structure and history of the Irish Travellers. Scientific Reports 7, 42187.

19. Hubisz M.J., Falush D., Stephens M. & Pritchard J.K. (2009) Inferring weak population structure with the assistance of sample group information. Molecular Ecology Resources 9, 1322-32.

20. Huson D.H. & Scornavacca C. (2012) Dendroscope 3: An interactive tool for rooted phylogenetic trees and networks. Systematic Biology 61, 1061-7.

21. Jombart T. & Ahmed I. (2011) adegenet 1.2-1: new tools for the analysis of genome-wide SNP data. Bioinformatics 27, 3070-1.

22. Kijas J.W., Lenstra J.A., Hayes B., Boitard S., Porto Neto L.R., San Cristobal M., Servin B., McCulloch R., Whan V., Gietzen K., Paiva S., Barendse W., Ciani E., Raadsma H., McEwan J., Dalrymple B. & International Sheep Genome Consortium Members (2012) Genome-wide analysis of the world's sheep breeds reveals high levels of historic mixture and strong recent selection. PLoS Biology 10, e1001258.

23. Kijas, J. (2013) ISGC SNP50 HapMap and sheep breed diversity genotypes v1. CSIRO. Data collection. https://doi.org/10.4225/08/51870B1E8EE56.

24. Kuehn L.A., Lewis R.M. & Notter D.R. (2008) Connectedness in Targhee and Suffolk flocks participating in the United States National Sheep Improvement Program. Journal of Animal Science 87, 507-15.

25. Lenstra J.A., Groeneveld L.F., Eding H., Kantanen J., Williams J.L., Taberlet P., Nicolazzi E.L., Sölkner J., Simianer H., Ciani E., Garcia J.F., Bruford M.W., Ajmone-Marsan P. & Weigend S. (2012) Molecular tools and analytical approaches for the characterization of farm animal genetic diversity. Animal Genetics 43, 483-502.

26. Li M.H., Strandén I., Tiirikka T., Sevón-Aimonen M.L. & Kantanen J. A comparison of approaches to estimate the inbreeding coefficient and pairwise relatedness using genomic and pedigree data in a sheep population. 6, e26256.

27. Lupton C.J. (2008) Impacts of animal science research on United States sheep production and predictions for the future. Journal of Animal Science 86, 3252-74.

28. McKay S.D., Schnabel R.D., Murdoch B.M., Matukumalli L.K., Aerts J., Coppieters W., Crews D., Dias Neto E., Gill C.A., Gao C., Mannen H., Wang Z., Van Tassell C.P., Williams J.L., Taylor J.F. & Moore S.S. (2008) An assessment of population structure in eight breeds of cattle using a whole genome SNP panel. BMC Genetics 9.

29. Michailidou S., Tsangaris G., Fthenakis G.C., Tzora A., Skoufos I., Karkabounas S.C., Banos G., Argiriou A. & Arsenos G. (2018) Genomic diversity and population structure of three autochthonous Greek sheep breeds assessed with genome-wide DNA arrays. Molecular Genetics Genomics 293, 753-68.

30. Notter, D. R. 1998. The U.S. National Sheep Improvement Program: across-flock genetic evaluations and new trait development. Journal of Animal Science 76, 2324-30.

31. Norberg E. & Sørensen A.C. (2007) Inbreeding trend and inbreeding depression in the Danish populations of Texel, Shropshire, and Oxford Down. *Journal of Animal Science* 85, 299-304.

32. Paradis E. (2010) pegas: an R package for population genetics with an integrated-modular approach. Bioinformatics 26, 419-20.

33. Purcell S., Neale B., Todd-Brown K., Thomas L., Ferreira M.A., Bender D., Maller J., Sklar P., de Bakker P.I., Daly M.J. & Sham P.C. (2007) PLINK: a toolset for whole-genome association and population-based linkage analysis. American Journal of Human Genetics 81, 559–75.

34. Ryder M.L. (1964) The history of sheep breeds in Britain. The Agricultural History Review 12, 1-12.

35. Sambrook, J., Fritsch, E. F. & Maniatis, T. (1989) Molecular cloning: a laboratory manual (No. Ed. 2). Cold spring harbor laboratory press.

36. United States Department of Agriculture Economic Research Service (USDA ERS) (2019) Sector at a glance. https://www.ers.usda.gov/topics/animal-products/sheep-lamb-mutton/sector-at-a-glance/.

37. Wang H., Zhang L., Cao J., Wu M., Ma Z., Liu Z., Liu R., Zhao F., Wei C. & Du L. (2015) Genome-wide specific selection in three domestic sheep breeds. PLoS ONE 10, e0128688.

38. Weir B.S. & Cockerham C.C. (1984) Estimating F-statistics for the analysis of population structure. Evolution 38, 1358-70.

39. Wilson D.E. & Morrical D.G. (1991) The National Sheep Improvement Program: a review. Journal of Animal Science 69, 3872-81.

40. Wright S. (1965) The interpretation of population structure by F-statistics with special regard to systems of mating. Evolution 19, 395–420.

41. Visser C., Lashmar S.F., Marle-Koster E.V., Poll M.A. & Allain D. (2016) Genetic diversity and population structure in South African, French and Argentinian Angora goats from genome-wide SNP data. PLoS ONE 11:e0154353

42. Zhang L., Mousel M.R., Wu X., Michal J.J., Zhou X., Ding B., Dodson M.V., El-Halawany N.K., Lewis G.S. & Jiang Z. (2013) Genome-wide genetic diversity and differentially selected regions among Suffolk, Rambouillet, Columbia, Polypay, and Targhee sheep. PLoS One 8, e65942.

43. Zhang M., Peng W.F., Hu X.J., Zhao Y.X., Lv F.H. & Yang J. (2018) Global genomic diversity and conservation priorities for domestic animals are associated with the economies of their regions of origin. Scientific Reports 8, 11677.

Tables

Table 2.1: The mean observed heterozygosity and average estimated inbreeding coefficient including the 95% confidence interval for each group

Breed	Observed Heterozygosity	Inbreeding Coefficient*	95% Confidence Interval for Inbreeding Coefficient
Hampshire	0.33	0.14	0.12-0.15
Suffolk	0.33	0.13	0.12-0.15
Western Suffolk	0.34	0.14	0.13-0.15
Oxford	0.35	0.05	0.01-0.09
Shropshire	0.34	0.09	0.04-0.11
Rambouillet	0.30	0.16	0.15-0.17

*Inbreeding coefficients are reported as Fhat2 and calculated by: (observed heterozygosity – expected) / (total – expected).

Table 2.2: Pairwise F_{ST}* between breeds of sheep.

	Hampshire	Suffolk	Western Suffolk	Oxford	Shropshire
Hampshire	0				
Suffolk	0.03	0			
Western Suffolk	0.09	0.07	0		
Oxford	0.06	0.06	0.13	0	
Shropshire	0.05	0.06	0.11	0.06	0
Rambouillet	0.17	0.17	0.23	0.18	0.16

*Wright's F_{ST} values between 0-0.05 are categorized as no differentiation, 0.06-0.15 as moderate differentiation, 0.16-0.25 as great differentiation, and >0.26 as very great differentiation.

Figures

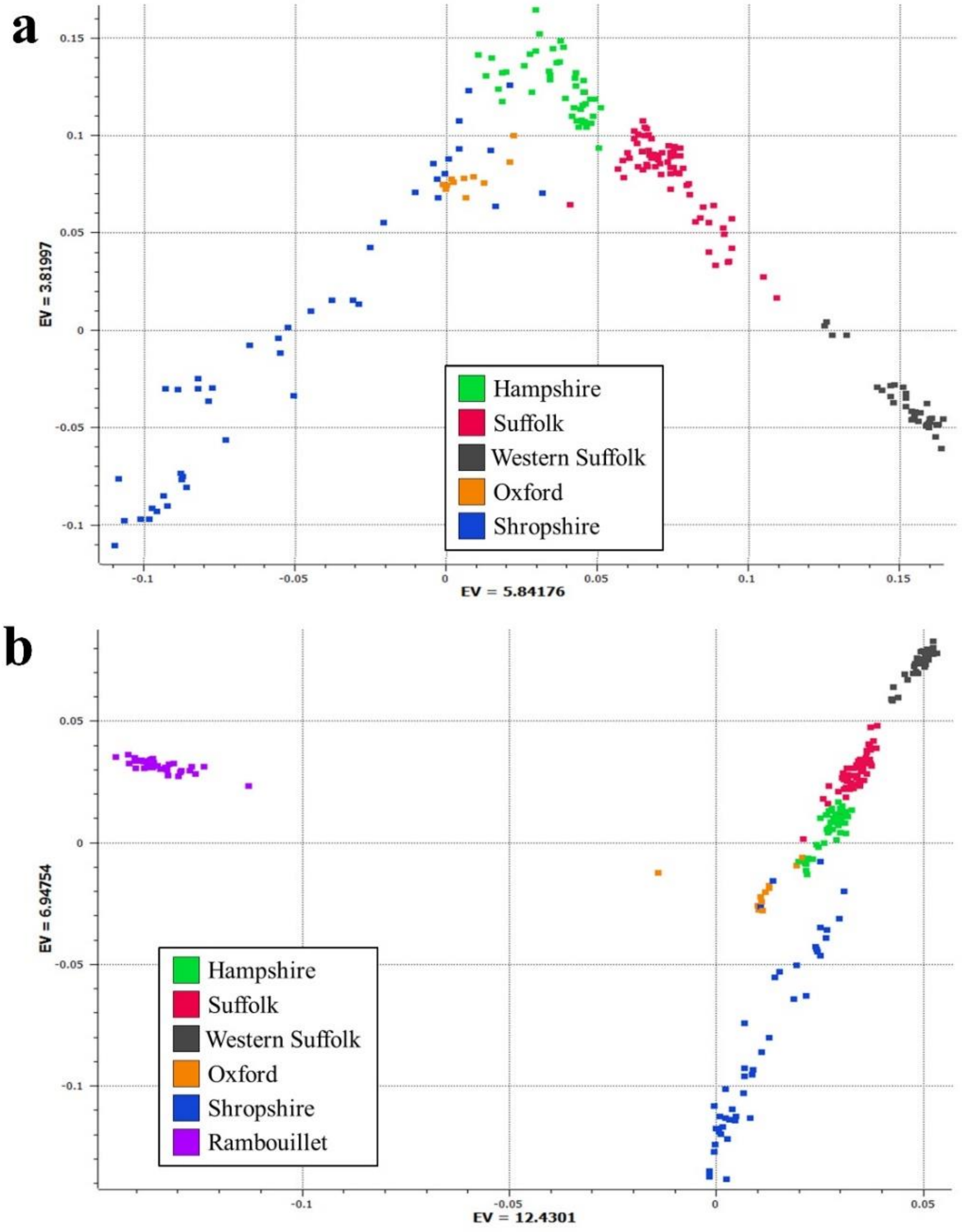


Figure 2.1: Plot of calculated Eigenvalues for breeds of U.S. sheep. (a) Eigenvalues plotted for U.S. terminal breeds of sheep. (b) Eigenvalues plotted for U.S. terminal breeds and Rambouillet sheep. Each point represents an individual animal and points are colored by reported breed.

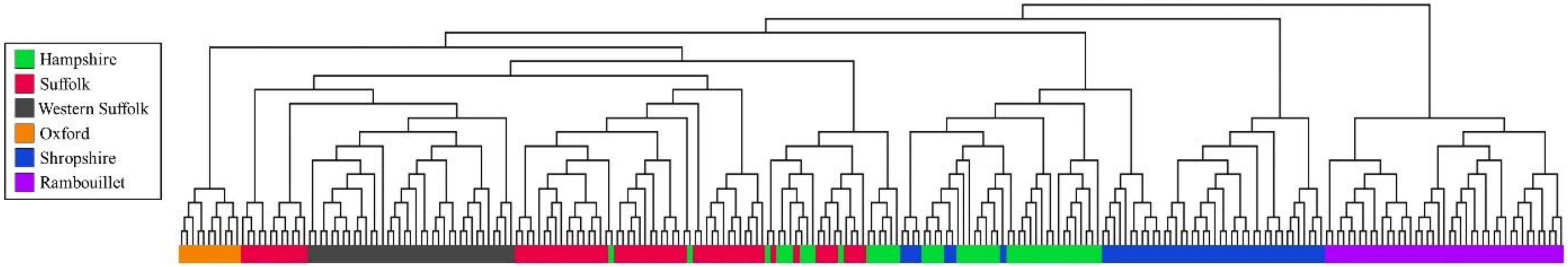


Figure 2.2: Rectangular cladogram of individuals clustered based on identity by state and coloured by reported breed.

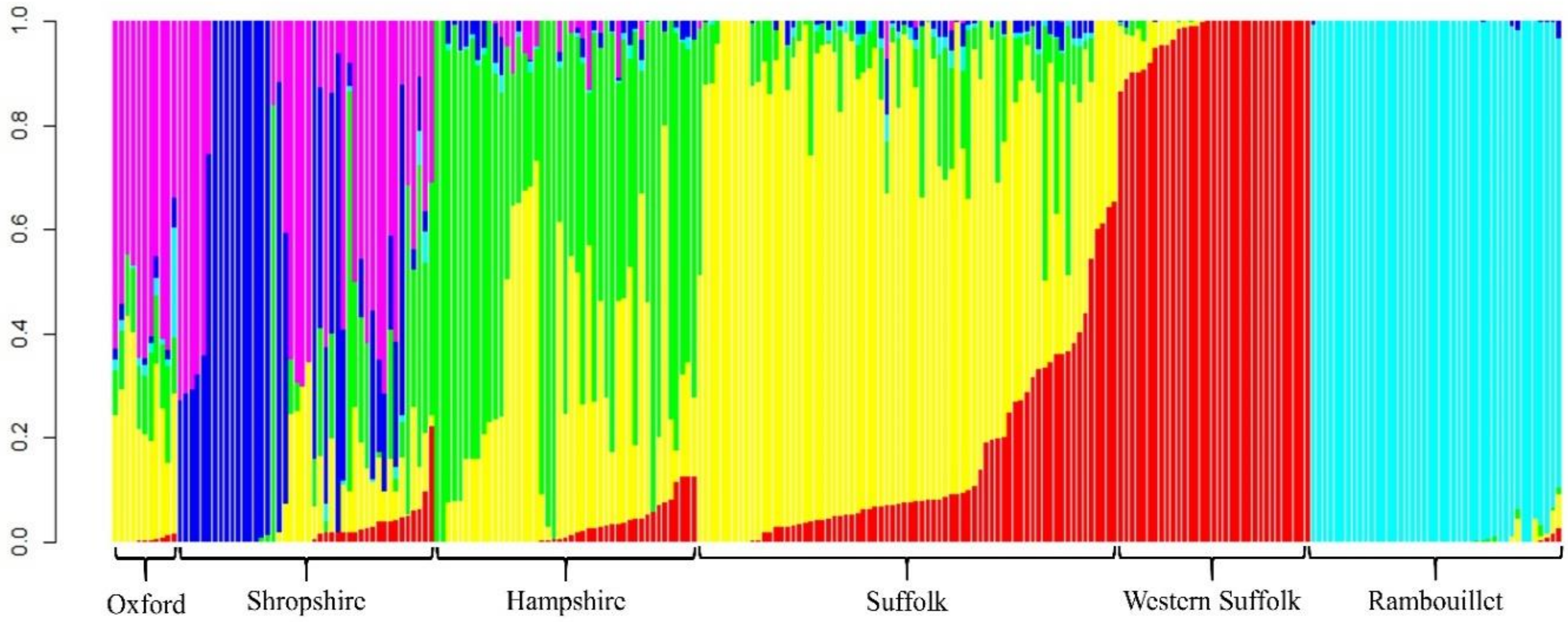


Figure 2.3: ADMIXTURE model clustering output with K-6 populations. Each bar represents an individual animal for each terminal breed and Rambouillet, and the six colours represent each K population cluster.

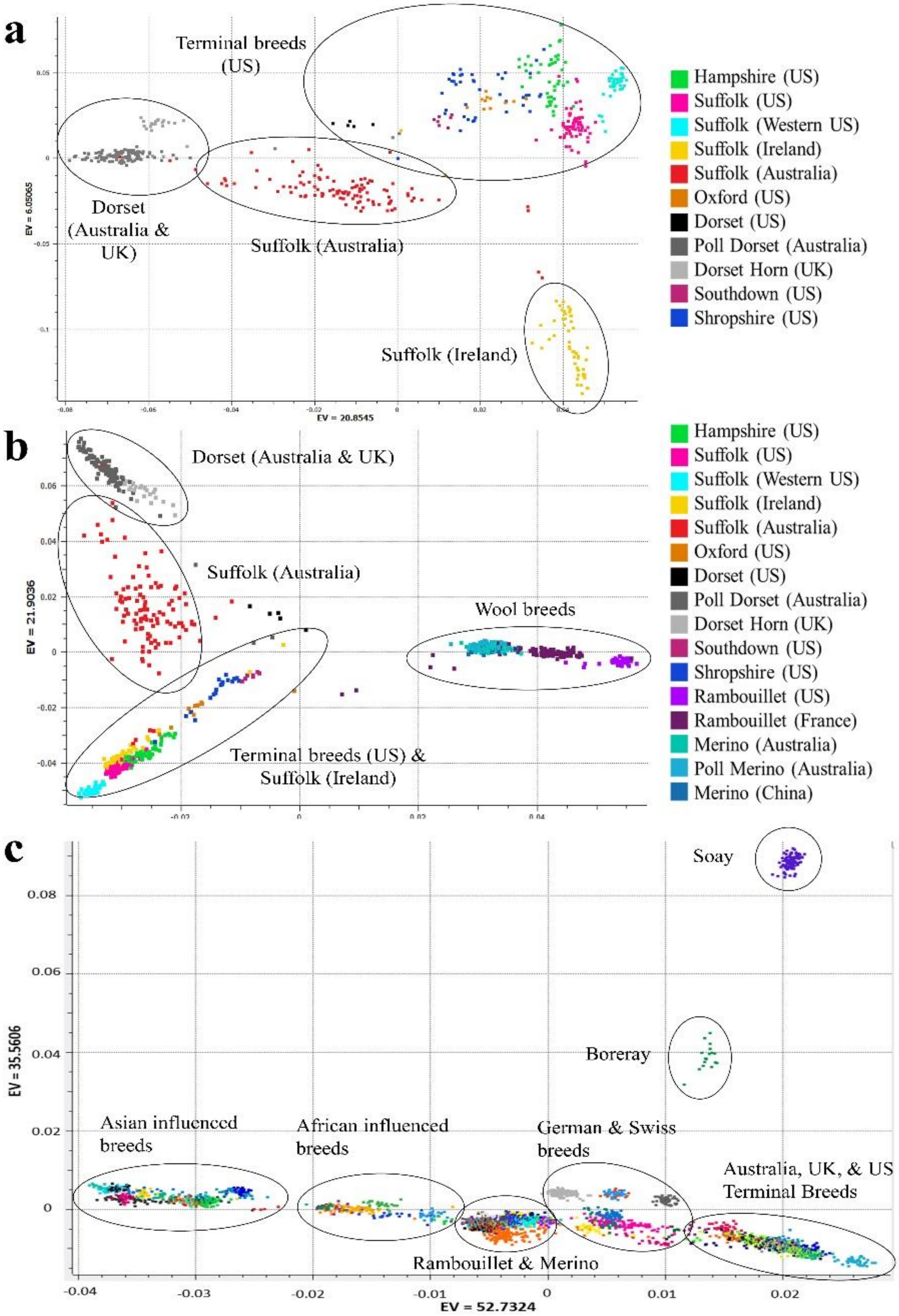
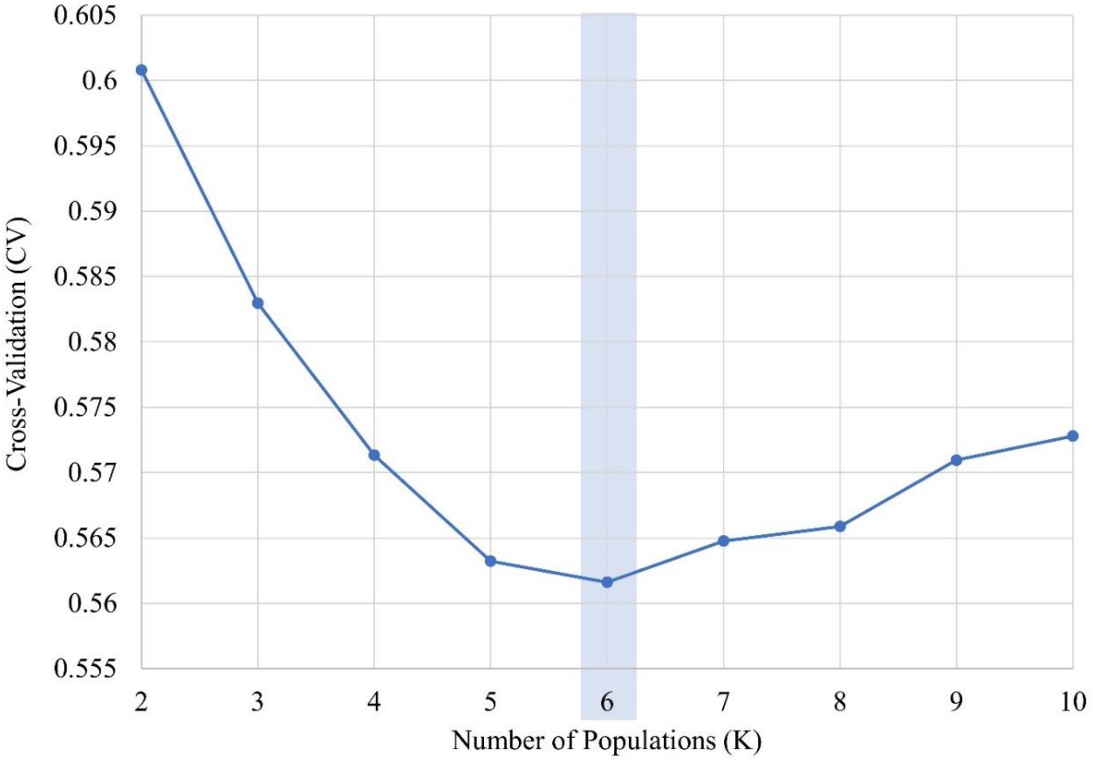

a
Terminal breeds (US)
Dorset (Australia & UK)
Suffolk (Australia)
Suffolk (Ireland)
0.05
0
-0.05
-0.1
EV = 6.05065
-0.08
-0.06
-0.04
-0.02
0
0.02
0.04
EV = 20.8545
Hampshire (US)
Suffolk (US)
Suffolk (Western US)
Suffolk (Ireland)
Suffolk (Australia)
Oxford (US)
Dorset (US)
Poll Dorset (Australia)
Dorset Horn (UK)
Southdown (US)
Shropshire (US)
b
Dorset (Australia & UK)
Suffolk (Australia)
Wool breeds
Terminal breeds (US) & Suffolk (Ireland)
0.06
0.04
0.02
0
-0.02
-0.04
EV = 21.9036
-0.02
0
0.02
0.04
Hampshire (US)
Suffolk (US)
Suffolk (Western US)
Suffolk (Ireland)
Suffolk (Australia)
Oxford (US)
Dorset (US)
Poll Dorset (Australia)
Dorset Horn (UK)
Southdown (US)
Shropshire (US)
Rambouillet (US)
Rambouillet (France)
Merino (Australia)
Poll Merino (Australia)
Merino (China)
c
Soay
Boreray
Asian influenced breeds
African influenced breeds
German & Swiss breeds
Australia, UK, & US Terminal Breeds
Rambouillet & Merino
0.08
0.06
0.04
0.02
0
EV = 35.5606
-0.04
-0.03
-0.02
-0.01
0
0.01
0.02
EV = 52.7324

Figure 2.4: Eigenvalue plots of U.S. sheep in this study compared to other breeds across the world as part of the Sheep HapMap study. (a) Eigenvalue plot of U.S. terminal breeds and Dorset and Suffolk HapMap breeds. (b) Eigenvalue plot of all U.S. sheep in this study compared to HapMap terminal and wool sheep. (c) Eigenvalue plot of U.S. sheep in this study compared to all breeds present in the Sheep HapMap study.



Supplementary Figure 2.1: ADMIXTURE cross-validation (CV) output plotted across K populations. The lowest CV value represents the most probable K number of populations for this dataset, K=6, which is highlighted in blue.

.

Chapter 3: The Complete Mitochondrial Genome Sequence of Bighorn Sheep

Published in Genome Announcements.

Citation: Davenport KM, Duan M, Hunter SS, New DD, Fagnan MW, Highland MA, Murdoch BM. Complete Mitochondrial Genome Sequence of Bighorn Sheep. Genome Announc. 2018 Jun 7;6(23):e00464-18. doi: 10.1128/genomeA.00464-18.

Kimberly M. Davenport[a], Mingrui Duan[a], Samuel S. Hunter[b], Daniel D. New[b], Matthew W. Fagnan[b], Margaret A. Highland[c], Brenda M. Murdoch[a]#

[a]Department of Animal and Veterinary Science, University of Idaho, Moscow, Idaho, USA

[b]Institute for Bioinformatics and Evolutionary Studies (IBEST), University of Idaho, Moscow, Idaho, USA

[c]Animal Disease Research Unit, Agricultural Research Services, USDA, Pullman, WA, USA

Running Head: The Complete Mitochondrial Genome of Bighorn Sheep

#Corresponding author: Brenda M. Murdoch
Tel: (208) 885-2088
Fax: (208) 885-6420
E-mail: bmurdoch@uidaho.edu

Abstract

We report here the complete mitochondrial genome sequence of a Rocky Mountain bighorn sheep (*Ovis canadensis*) in the United States. The circular genome has a size of 16,466 bp and contains 13 protein-coding genes, 22 tRNA genes, and two rRNA genes.

Manuscript

The bighorn sheep (*Ovis canadensis*) is an important ecological model for studying natural selection and evolution in Western North America (1,2,3,4). The population of bighorn sheep drastically declined in the early 20th century due to habitat loss, disease, and overharvest coinciding with European settlement, but has substantially rebounded because of conservation efforts and management strategies (5,6). However, this population decrease led to a bottleneck effect and reduced genetic diversity (4). Investigating genetic diversity and effective population sizes in bighorn sheep will aid in continued management and conservation of this species (4).

Mitochondrial genetic sequence has been used in many species for population genetics analyses and discerning phylogeny (7,8,9). Numerous mitochondrial genomes are available for different breeds of domestic sheep (*Ovis aries*), however only one has been released for bighorn sheep from Canada (10). Here, we report a complete mitochondrial genome of the Rocky Mountain bighorn sheep from an 8-month-old male in the United States. The animal was raised in a small cohort in captivity at Washington State University in Pullman, WA under the guidelines of the Institutional Animal Care and Use Committee and Association for Assessment and Accreditation of Laboratory Animal Care.

Mitochondrial DNA was extracted from liver with the Mitochondrial DNA Isolation Kit (Abcam, Cambridge, MA). Nextera shotgun libraries were produced and sequenced using a v3 600 cycle kit and Illumina MiSeq by the IBEST Genomics Resources Core at the University of Idaho. Adapter sequences were trimmed, low-quality ends were removed, and pair-end reads were overlapped by HTStream (https://github.com/ibest/HTStream). Cleaned data were assembled by the ARC software package v1.1.4-beta (https://github.com/ibest/ARC) using *Ovis aries* isolate GP092 mitochondrial genome (NCBI accession number KF302455) as seed reference to initialize the iterative assemblies. The assembly resulted in one circular contig, as confirmed by dot plot. The ends were overlapped and joined, and the resulting sequence was linearized such that the orientation started with tRNA-Phe to match other sheep mitochondrial genomes. The complete genome is 16,466bp with a GC content of 38.9%. The structural and functional annotation was performed with the mitochondrial genome annotation (MITOS) web server (11). Annotations of genes were checked using homology searches on GenBank and further improved by manual curation in Geneious version 9.1.8 (http://www.geneious.com) (12). The bighorn sheep mitochondrial genome is predicted to have 22 tRNA genes, 2 rRNA genes (12S and 16S) and 13 respiratory genes common to most animal mtDNA (*ATP6*, *ATP8*, *CYTB*, *COX1*, *COX2*, *COX3*, *ND1*, *ND2*, *ND3*, *ND4*, *ND4l*, *ND5*, and *ND6*). Alignment with other sheep mitochondrial genome sequences showed 99.6% identity with that of bighorn sheep, and 96% with sequences of domestic sheep, which is 3 million years divergent (2). This suggests that the mtDNA sequence we obtained is consistent with phylogenetic relationships for the studied populations of *Ovis* species. This complete mitochondrial genome provides an additional resource for phylogeographic and population genetic investigations in bighorn sheep, which contributes to future studies on sheep evolution and conservation efforts.

Accession Number. The mtDNA genome sequence has been deposited in GenBank under accession number MH094035.

Acknowledgements

This material is based upon work that is supported by the National Institute of Food and Agriculture, U.S. Department of Agriculture, Hatch grant IDA01566 and NIH COBRE grant P30GM103324.

References

1. Coltman D, Festa-Bianchet M, Jorgenson J, Strobeck C. 2002. Age-dependent sexual selection in bighorn rams. Proceedings of the Royal Society of London B: Biological Sciences 269:165–172.

2. Bunch TD, Wu C, Zhang YP, Wang S. 2006. Phylogenetic analysis of snow sheep (*Ovis nivicola*) and closely related taxa. J Hered. 97:21-30.

3. Miller JM, Poissant J, Hogg JT, Coltman DW. 2012. Genomic consequences of genetic rescue in an insular population of bighorn sheep (*Ovis canadensis*). Mol Ecol 21:1583–1596.

4. Kardos M, Luikart G, Bunch R, Dewey S, Edwards W, McWilliam S, Stephenson J, Allendorf FW, Hogg JT, Kijas J. 2015. Whole-genome resequencing uncovers molecular signatures of natural and sexual selection in wild bighorn sheep. Mol Ecol 24:5616-5632.

5. Buechner HK. 1960. The bighorn sheep in the United States, It's past, present, and future. Wildlife Monographs 4:3-174.

6. Rominger E. 2008. Ram harvest strategies for western states and provinces. Biennial Symposium of the Northern Wild Sheep and Goat Council 16:92-96.
7. Avise JC, Arnold J, Ball RM, Bermingham E, Lamb T, Neigel JE, Reeb CA, Saunders NC. 1987. Intraspecific phylogeography: the mitochondrial DNA bridge between population genetics and systematics. Annu Rev Ecol Evol Syst. 18:489-522.

8. Moritz C. 1994. Defining ‘evolutionary significant units’ for conservation. Trends Ecol Evol. 9:373-375.

9. Moore WS. 1995. Inferring phylogenies from mtDNA variation: mitochondrial-gene trees versus nuclear-gene trees. Evolution. 49:718-726.

10. Miller JM, Malenfant RM, Moore SS, Coltman DW. 2012. Short reads, circular genome: skimming solid sequence to construct the bighorn sheep mitochondrial genome. J Hered. 103:140-146.

11. Bernt M, Donath A, Jühling F, Externbrink F, Florentz C, Fritzsch G, Pütz J, Middendorf M, Stadler PF. 2013. MITOS: Improved *de novo* Metazoan Mitochondrial Genome Annotation. Mol Phylogenet Evol. 69:313-319.

12. Kearse M, Moir R, Wilson A, Stones-Havas S, Cheung M, Sturrock S, Buxton S, Cooper A, Markowitz S, Duran C, Thierer T, Ashton B, Meintjes P, Drummond A. 2012. Geneious basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data. Bioinformatics. 28:1647–1649.

Chapter 4: An Improved Ovine Reference Genome Assembly to Facilitate In-Depth Functional Annotation of the Sheep Genome

Forthcoming in GigaScience.

Kimberly M. Davenport[1], Derek M. Bickhart[2], Kim Worley[3], Shwetha C. Murali[4], Mazdak Salavati[5], Emily L. Clark[6], Noelle E. Cockett[7], Michael P. Heaton[8], Timothy P.L. Smith[9], Brenda M. Murdoch[10]*, and Benjamin D. Rosen[11]*

[1]Department of Animal, Veterinary, and Food Sciences, University of Idaho, 875 Perimeter Dr., Moscow, ID, United States 83843. Email: kmdavenport@uidaho.edu

[2]US Dairy Forage Research Center, USDA-ARS, 1925 Linden Drive, Madison, WI, United States 53706. Email: derek.bickhart@usda.gov

[3]Baylor College of Medicine, One Baylor Plaza, Houston, TX, United States 77030. Email: kworley@bcm.edu

[4]Baylor College of Medicine, One Baylor Plaza, Houston, TX, United States 77030. Email: shwethac@gmail.com

[5]The Roslin Institute, Royal (Dick) School of Veterinary Studies, The University of Edinburgh, Easter Bush Campus, Midlothian, United Kingdom, EH25 9RG, United Kingdom. Email: mazdak.salavati@roslin.ed.ac.uk

[6]The Roslin Institute, Royal (Dick) School of Veterinary Studies, The University of Edinburgh, Easter Bush Campus, Midlothian, United Kingdom, EH25 9RG. Email: emily.clark@roslin.ed.ac.uk

[7]Utah State University, Old Main Hill, Logan, UT 84322. Email: noelle.cockett@usu.edu

[8]US Meat Animal Research Center, USDA-ARS, State Spur 18D, Clay Center, NE 68933. Email: mike.heaton@usda.gov

[9]US Meat Animal Research Center, USDA-ARS, State Spur 18D, Clay Center, NE 68933. Email: tim.smith2@usda.gov

[10]Department of Animal, Veterinary, and Food Sciences, University of Idaho, 875 Perimeter Dr., Moscow, ID 83843. Email: bmurdoch@uidaho.edu

[11]Animal Genomics and Improvement Laboratory, USDA-ARS, 10300 Baltimore Avenue, Beltsville, MD 20705. Email: ben.rosen@usda.gov

Correspondence:

Brenda M. Murdoch
bmurdoch@uidaho.edu

Benjamin D. Rosen
ben.rosen@usda.gov

Abstract

Background

The domestic sheep (*Ovis aries*) is an important agricultural species raised for meat, wool, and milk across the world. A high-quality reference genome for this species enhances the ability to discover genetic mechanisms influencing biological traits. Further, a high-quality reference genome allows for precise functional annotation of gene regulatory elements. The rapid advances in genome assembly algorithms and emergence of increasingly long sequence read length provide the opportunity for an improved *de novo* assembly of the sheep reference genome.

Findings

Short-read Illumina (55x coverage), long-read PacBio (75x coverage), and Hi-C data from this ewe retrieved from public databases were combined with an additional 50x coverage of Oxford Nanopore data and assembled with canu v1.9. The assembled contigs were scaffolded using Hi-C data with Salsa v2.2, gaps filled with PBsuitev15.8.24, and polished with Nanopolish v0.12.5. After duplicate contig removal with PurgeDups v1.0.1, chromosomes were oriented and polished with two rounds of a pipeline which consisted of freebayes v1.3.1 to call variants, Merfin to validate them, and BCFtools to generate the consensus fasta. The ARS-UI_Ramb_v2.0 assembly has improved continuity (contig N50 of 43.18 Mb) with a 19-fold and 38-fold decrease in the number of scaffolds compared with Oar_rambouillet_v1.0 and Oar_v4.0. ARS-UI_Ramb_v2.0 has greater per-base accuracy and fewer insertions and deletions identified from mapped RNA sequence than previous assemblies.

Conclusions

The ARS-UI_Ramb_v2.0 assembly is a substantial improvement that will optimize the functional annotation of the sheep genome and facilitate improved mapping accuracy of genetic variant and expression data for traits in sheep.

Keywords: Rambouillet, genome assembly, reference genome, sheep

Context

The domestic sheep (*Ovis aries*) is a globally important livestock species raised for a variety of purposes including meat, wool, and milk. Domestication likely occurred in multiple events approximately 11,000 years ago [1-4]. Selection for desirable traits including meat, wool, and milk began approximately 4,000-5,000 years ago [2,4]. Modern sheep breeds exhibit a wide variety of phenotypes and adaptations to specific environments, for example the enhanced parasite tolerance evident in hair sheep [5,6]. As many as 1,400 breeds of sheep exist today [7-9] including the Rambouillet breed developed in France from a Merino fine wool lineage that is regarded for its ability to produce high quality wool as well as meat products in production systems across the world [10,11].

Genome research in sheep holds promise to improve efficiency and sustainability of production and reduce the environmental effects of animal agriculture [12]. The first sheep reference genome assembly was based on whole genome shotgun (WGS) short-read sequencing, scaffolded by genetic linkage and radiation hybrid maps. The sequence came from two unrelated Texel breed sheep, with the first assembly draft (Oar_v3.1; International Sheep Genomics Consortium, 2010) having a contig N50 of 40 kilobases (kb) and the update (Oar_v4.0) [13] boosting the N50 metric to 150 kb. More recently, the Ovine Functional Annotation of Animal Genomes (FAANG) project proposed to perform a variety of genome annotation assays for dozens of tissues from a single animal [14,15]. To maximize the success of assays that depend on mapping sequence data to a reference, the FAANG project assembled the genome of that animal, a female of the Rambouillet breed. The assembly, released in 2017 (Oar_rambouillet_v1.0, GenBank accession GCF_002742125; Worley et al., unpublished) is based on a combination of Pacific Biosciences RSII WGS long-read and Illumina short-read sequencing. It has an improved contig N50 of 2.6 megabases (Mb) and is generally regarded as the official reference assembly for global sheep research.

The continued maturation of long read sequencing technologies provided an opportunity to improve upon the sheep reference genome assembly. Since most of the proposed FAANG annotation assays had already been performed on the Rambouillet ewe, lung tissue from the same animal was chosen for DNA extraction. This allowed the use of existing long read data to supplement new, longer-read, Oxford Nanopore PromethION sequencing. We report a *de novo* assembly of the same Rambouillet ewe used for Oar_rambouillet_v1.0, based on approximately 50x coverage of nanopore reads (N50 47kb) and 75x coverage PacBio reads (N50 13kb). The new assembly, ARS-UI_Ramb_v2.0 offers a 20-fold improvement in contiguity and increased accuracy, providing a basis for regulatory element annotation in the FAANG project and facilitating the discovery of biological mechanisms that influence traits important in global sheep research and production.

Methods

Sampling Strategy

The fullblood Rambouillet ewe used for this genome assembly (Benz 2616, USMARC ID 200935900) (Figure 4.1) was selected by the Ovine Functional Annotation of Animal Genomes project and acquired from the USDA. Tissues were collected postmortem from the healthy six-year-old ewe as approved by the Utah State University Institutional Animal Care and Use Committee. A full description of the tissue collection strategy is available in the FAANG Data Coordination Center [15,16]. Details regarding the tissues collected from the animal are available under BioSample number SAMEG329607 [17].

Sequencing and Data Acquisition

DNA was extracted from approximately 50 mg of lung tissue using phenol:chloroform-based method as described (Logsdon 2019). Briefly, the frozen tissue was pulverized in a cryoPREP CP02 tissue disruption system (Covaris Inc., Woburn MA) as recommended by the manufacturer. The powdered tissue was transferred to a 50 mL conical tube and mixed in 200 µL of phosphate buffered saline (Sigma-Aldrich, St. Louis MO). The tissue was then diluted in 10 mL of buffer TLB (100mM NaCl, 10mM Tris-HCl pH 8.0, 25 mM EDTA, 0.5% SDS) and mixed by vortexing, then incubated with 20 µL 10 mg/mL RNase A at 37ºC for one hour with gentle shaking. Protein digestion was performed with 100 µL Proteinase K (20 mg/mL) at 50ºC for 2 hours, with slow rotation of the tube to mix every 30 minutes. The lysate was distributed equally into two 15 mL Phase Lock tubes (Quantabio, Beverly MA) and each tube received 5 mL of TE-saturated Phenol (Sigma-Aldrich, St. Louis MO) followed by mixing on a tube rotator at 20 RPM for 10 minutes at 22ºC. The aqueous layer was collected after separating at 2300xg for 10 minutes and transferred to another Phase Lock tube. A second extraction performed in the same way as the first was conducted using 2.5 mL phenol and 2.5 mL chloroform:isoamyl alcohol (Sigma). The final aqueous phase was transferred to a 50 mL conical tube and the DNA precipitated with 2 mL of 5M ammonium acetate and 15 mL of ice-cold 100% ethanol. The DNA was pulled from the alcohol using a Pasteur pipet "hook" and placed in 10 mL of cold 70% ethanol to wash the pellet. The ethanol was poured off and the DNA pellet dried for 20-30 minutes, then dissolved in a dark drawer at room temperature for 48 hours in 1 mL of 10mM Tris-Cl pH 8.5. Library preparation for Oxford Nanopore long read sequencing was performed with an LSK-109 template preparation kit as recommended by the manufacturer (Oxford Nanopore, Oxford U.K.) with modifications as described by Logsdon (https://www.protocols.io/view/hmw-gdna-purification-and-ont-ultra-long-read-data-bchhit36?comment_id=88927). The ligated template was sequenced with a PromethION instrument using four R9.4 flow cells. (Oxford Nanopore Technologies, Oxford, United Kingdom). Output as fast5 files were basecalled with Guppy v3.1 [18].

Sequence data used in the previous Oar_rambouillet_v1.0 assembly was retrieved from the Sequence Read Archive listed under project number PRJNA414087 [15]. PacBio RS II sequence generated from DNA extracted from whole blood was retrieved from SRX3445660, SRX3445661, SRX3445662, and SRX3445663. The Hi-C sequence data generated from liver using HindIII enzyme and sequenced at 150 bp paired end with an Illumina HiSeq X Ten was retrieved from SRX3399085 and SRX3399086. Short read whole genome sequencing from DNA extracted from whole blood collected from the Rambouillet ewe was performed with an Illumina HiSeq X Ten sequenced at 150 bp paired end and was retrieved from SRX3405602. Further details about these sequences can be found under the umbrella project number PRJNA414087. Short read 45 bp paired end whole genome sequence from an Illumina Genome Analyzer II generated from Texel sheep used in previous genome assemblies were retrieved from the Sequence Read Archive under accessions SRX511533-SRX511565 (BioProject PRJNA169880).

Assembly

Contigs were assembled with Oxford Nanopore and PacBio reads generated as described above using canu v1.8 through the trimmed reads stage of assembly. Parameters for contig construction were set as “batOptions=-dg 4 -db 4 -mo 1000” [19]. Canu v1.9 was used to complete the contig assembly because this update demonstrates better consensus generation of the overlapped contigs in the final step in the assembly process [20,21]. The corrected error rate option was set as “correctedErrorRate=0.105.”

Scaffolding

Two Hi-C datasets from liver tissue from two different library preparations were retrieved as described above. The Hi-C reads were first aligned to the polished contigs using the Arima Genomics mapping pipeline [22]. This pipeline first maps paired end reads individually with bwa-mem, then removes the 3’ end of reads identified as chimeric and span ligation junctions. Reads were then paired, filtered by mapping quality with samtools [23], and PCR duplicates removed with Picard [24]. The two Hi-C libraries were merged in the final step in the Arima pipeline to generate the merged BAM file. The BAM file was converted to a BED file for input into Salsa using the bedtools command bamToBed [25]. Salsa v2.2 was used for scaffolding by implementing “python run_pipeline.py -a contigs.fasta -l contigs.fasta.fai -b alignment.bed -e HindIII -o scaffolds -m yes” [26].

The Hi-C reads were aligned to the scaffolded assembly with the Arima Genomics mapping pipeline and then processed with PretextMap to visually evaluate the scaffolds as a contact map in PretextView [27]. The scaffolded assembly was also compared to *Oar_rambouillet_v1.0* by aligning the two genomes with “minimap2 -cx asm5 Oar_rambouillet_v1.0_genomic.fasta scaffolds.fasta > alignment.paf” [28]. A dotplot of the alignment was visualized with D-Genies [29]. Scaffolds were edited based on visual inspection of the contact map and dotplot, as well as the Hi-C alignment file. Scaffold joins and rearrangements were incorporated to the assembly using the *agp2fasta* mode of CombineFasta [30].

Gap Filling and Polishing

Gap filling was completed with pbsuite v15.8.24 using both the PacBio and Oxford Nanopore reads. Nanopolish v0.12.5 [31] with the NanoGrid parallel wrapper [32] was employed with the raw fast5 files generated from the PromethION sequencing to polish the assembly. Duplicates were removed using PurgeDups v1.0.1 [33]. The chromosome orientation was confirmed in the polished assembly by identifying telomeres and centromeres using RepeatMasker v4.1.1 [34]. The mitochondrial genome was identified by aligning the previously annotated mitochondrial sequence from Oar_rambouillet_v1.0 (RefSeq NC_001941.1) to the assembly contigs. Chromosomes were oriented centromere to telomere and placed in chromosome number order. The final polishing was performed with two rounds of freebayes v1.3.1 using the Illumina short read data after final chromosome orientations and mitochondrial genome were confirmed [35]. Variants used for polishing with both Nanopolish and freebayes were screened with Merfin [36] which evaluates the k-mer consequences of variant calls and filters unsupported variants.

RNA Sequencing

RNA sequencing data was generated from five tissues including skin, thalamus, pituitary, lymph node (mesenteric), and abomasum pylorus collected from the animal used to assemble the reference genome. Details regarding the RNA isolation protocol, library preparation, and sequencing as well as the raw data can be found in GenBank under BioProject PRJEB35292, specifically under SRA run numbers ERR3665717 (skin), ERR3728435 (thalamus), ERR3650379 (pituitary), ERR3665711 (lymph node mesenteric), and ERR3650373 (abomasum pylorus). Reads were trimmed with Trim Galore v0.6.4 [37] and alignment to both Rambouillet genomes was performed with STAR v2.7 using default parameters [38]. Indels were identified with bcftools mpileup, filtering allele depth (AD) at > 5 [39].

Annotation

The annotation for ARS-UI_Ramb_v2.0, NCBI Ovis aries Annotation Release 104, is available in RefSeq and other NCBI genome resources (https://ftp.ncbi.nlm.nih.gov/genomes/all/annotation_releases/9940/104).

Here we also provide a liftover of the annotation for Oar_rambouillet_v1.0 onto ARS-UI_Ramb_v2.0. The annotation used for the liftover was NCBI v103 GCF_002742125.1_Oar_rambouillet_v1.0_genomic.fna.gz. The GFF3 format gene annotation file was prepared for processing using liftOffv1.5.2 [40]. A set of matching chromosome names for Oar_rambouillet_v1.0 and ARS-UI_Ramb_v2.0 were generated according to the instructions for liftOff (*paste -d "," <(cut -d' ' -f1 ramb1.chr) <(cut -d' ' -f1 ramb2.chr) > chroms.txt*). The GFF file (annotation Ramb1LO2) generated by liftOff is included in Supplementary File 1 (Ramb_v1.0_NCBI103_lifted_over_ARS-UI_Ramb_v2.0.gff.gz).

To compare the breakdown of transcripts captured by the three annotations (Oar_Rambouillet_v1.0, Ramb1LO2 (liftover) and ARS-UI_Ramb_v2.0), we generated transcript expression estimates using Kallisto v0.44.0 [41]. For the lifted over gene annotation the GFF file (Ramb_v1.0_NCBI103_lifted_over_ARS-UI_Ramb_v2.0.gff.gz) was used to generate transcriptome sequence FASTA files, as a Kallisto index, for transcript expression estimation. Briefly, exonic blocks were extracted from the GFF3 file using the awk command (*awk '($3~/exon/)' input.gff)*. The getfasta and groupby plugins from bedtools v2.30.0 [42] were used to extract the exonic sequences and group them by transcript name. Exonic sequences for each transcript were appended in the correct order, to produce the complete sequence for each transcript. The FASTA format file for the whole transcriptome was created using all of the transcript level FASTA sequences for the liftover annotation Ramb1LO2 (liftover; Ramb1LO2_NCBI103_geneBank_rna.fa). The set of scripts used for this step are included in Supplementary File 1. The Kallisto indices for Oar_Rambouillet_v1.0 (GCF_002742125.1_Oar_rambouillet_v1.0_rna.fna.gz), Ramb1LO2 (liftover; Ramb1LO2_NCBI103_geneBank_rna.fa) and ARS-UI_Ramb_v2.0 (GCF_016772045.1_ARS-UI_Ramb_v2.0_rna.fna.gz) were then used with the RNA-Seq data from the 61 tissues from Benz2616 (GenBank BioProject PRJNA414087 and PRJEB35292) to estimate transcript level expression for every tissue as transcript per million mapped reads (TPM) and compared across the three annotations.

Data Validation and Quality Control

Assembly Quality Statistics

The four flow cells of PromethION data produced 136 gigabases (Gb) of WGS sequence (approximately 51x coverage) in reads having a read N50 of 47 kb. The initial generation of contigs used this data as well as 198.1 Gb of RSII data with a read N50 of 12.9 kb. The ARS-UI_Ramb_v2.0 assembly was submitted to NCBI GenBank under accession number GCA_016772045.1, and statistics of contigs and scaffolds following initial polishing, scaffolding with Hi-C data and manual editing, gap-filling, and final polishing, are shown in Table 1. The assembly improved on the Oar_v4.0/Oar_rambouillet_v1.0 sheep reference assemblies in all continuity measures (Table 4.1) including a 286/17-fold increase in contig N50 (the size of the shortest contig for which all larger contigs contain half of the total assembly), a 214/33-fold reduction in the number of contigs in the assembly and concomitant 209/13-fold reduction of contig L50 (the number of contigs making up half of the total assembly), and 38/19-fold reduction in total number of scaffolds. Manual curation of scaffolds using Hi-C data improved scaffold continuity and led to chromosome length scaffolds (Figure 4.2).

The Themis-ASM pipeline [43] was implemented to further assess assembly quality and compare sheep genome assemblies. Short read sequence from both the Rambouillet ewe used in this assembly and Texel sheep from previous sheep genome assemblies were used to compare ARS-UI_Ramb_v2.0 with Oar_rambouillet_v1.0 and Oar_v4.0 assemblies.

The k-mer based quality value and error rates improved with ARS-UI_Ramb_v2.0 compared with Oar_rambouillet_v1.0 and Oar_v4.0. This is also reflected in the proportion of complete assembly based on k-mers (merCompleteness), which is similar between ARS-UI_Ramb_v2.0 and Oar_rambouillet_v1.0 and both are higher than Oar_v4.0. Further, the SNP and indel quality value (baseQV) were greatest overall in ARS-UI_Ramb_v2.0 (41.84), followed by Oar_rambouillet_v1.0 (40.69) and Oar_v4.0 (32.40). The percentage of short reads not mapped to the genome was ≤1% in all three assemblies.

The completeness of ARS-UI_Ramb_v2.0 was evaluated by examining the presence or absence of evolutionarily conserved genes in each assembly using Benchmarking Universal Single-Copy Ortholog (BUSCO) scores generated as an output of the Themis-ASM pipeline. The percent of single copy complete BUSCOs were higher (90.7%) in ARS-UI_Ramb_v2.0 when compared with Oar_rambouillet_v1.0 (90.1%) and Oar_v4.0 (86.1%). Complete duplicated BUSCO percentage was highest in Oar_rambouillet_v1.0 (1.6%) compared with ARS-UI_Ramb_v2.0 (1.4%), and lowest in Oar_v4.0 (1.0%). Further, ARS-UI_Ramb_v2.0 had the lowest percent of fragmented and missing BUSCOs (2.0% and 5.9%, respectively) compared with Oar_rambouillet_v1.0 (2.1% and 6.2%, respectively) and Oar_v4.0 (3.7% and 9.2%, respectively).

The three sheep genome assemblies were also compared with a feature response curve in which the quality of the assembly is analyzed as a function of the features, or maximum number of possible errors, allowed in the contigs (Figure 4.3) [44]. Both the ARS-UI_Ramb_v2.0 and Oar_v4.0 feature response curves peak higher and to the left of Oar_rambouillet_v1.0, which indicate fewer errors in these assemblies (Figure 4.3A). The ARS-UI_Ramb_v2.0 genome also has fewer regions with either low or high coverage overall and for paired reads, suggesting fewer coverage issues, as well as fewer improperly paired or unmapped single reads when compared with other assemblies (Figure 4.3B). The number of high Comp/Expansion (CE) statistics in ARS-UI_Ramb_v2.0 was intermediate between Oar_rambouillet_v1.0 (higher) and Oar_v4.0 (lower), however this latest assembly had the lowest number of regions with low CE statistics.

Comparative alignment of ARS-UI_Ramb_v2.0 with previous assemblies Oar_rambouillet_v1.0 and Oar_v4.0 and visualization with a dotplot revealed a high amount of agreement between assemblies (Figure 4.4). Interestingly, chromosome 11 was improperly oriented in Oar_rambouillet_v1.0, and after confirming centromere and telomere locations on this chromosome, this was resolved in the ARS-UI_Ramb_v2.0 assembly. The percent identity between ARS-UI_Ramb_v2.0 is very high when compared with Oar_rambouillet_v1.0 which was expected considering the same animal was used in both assemblies. However, Oar_v4.0 was assembled from Texel sheep, which is apparent in the percent identity in the dotplot.

In summary, ARS-UI_Ramb_v2.0 offers greater contiguity, improved quality, more complete BUSCOs, and fewer assembly errors when compared with previous assemblies.

RNA sequencing alignment

Insertions and deletions (indels) in the ARS-UI_Ramb_v2.0 assembly were characterized and compared with Oar_rambouillet_v1.0 by mapping 150 bp paired-end RNA-seq data from skin, thalamus, pituitary, lymph node (mesenteric), and abomasum pylorus generated from the same animal used to assemble the reference genome. In all five tissues, ARS-UI_Ramb_v2.0 had nearly half of the number of indels compared with Oar_rambouillet_v1.0. Most indels identified in both assemblies were 1bp in length. The ARS-UI_Ramb_v2.0 had a greater number of uniquely mapped reads in each tissue when compared with Oar_rambouillet_v1.0, leading to an approximate 2% increase in the percent of uniquely mapped reads in most tissues except pituitary, which saw an almost 13% improvement. The number of reads that mapped to multiple loci decreased in the new assembly by 12.59% in pituitary, and 1-2% in other tissues. Further, ARS-UI_Ramb_v2.0 had fewer unmapped reads than Oar_rambouillet_v1.0 across all five tissues by an average of 0.15%.

Annotation

The ARS-UI_Ramb_v2.0 annotation represents a substantial improvement over the annotation on Oar_rambouillet_v1.0. For example, for ARS-UI_Ramb_v2.0 16,500 coding genes have an ortholog to human (compared to 16,319 for Oar_rambouillet_v1.0), and the BUSCO scores demonstrate that 99.1% of the gene models (cetartiodactyla_odb10) are complete in the new annotation versus 98.8% in the previous one. The annotation for ARS-UI_Ramb_v2.0 includes Iso-Sequencing for 8 tissues to improve contiguity of gene models, and CAGE sequencing for 56 tissues to define TSS, that were not used to annotate Oar_rambouillet_v1.0. The full report for the annotation release is available at: (https://www.ncbi.nlm.nih.gov/genome/annotation_euk/Ovis_aries/104).

Using Kallisto we compared the number of expressed transcripts, for the RNA-Seq dataset of 61 tissue samples from Benz2616, across the three annotations (Oar_Rambouillet_v1.0, Ramb1LO2 (liftover) and ARS-UI_Ramb_v2.0). There was a considerable increase in the number of transcripts captured by the annotation for ARS-UI_Ramb_v2.0 (60,064) relative to Oar_Rambouillet_v1.0 (42,058) and the liftover annotation (Ramb1LO2) (40,910) (Supplemental Figure x). This equates to approximately 20,000 new annotated gene models for ARS-UI_Ramb_v2.0 and further reflects the substantial improvement over the annotation for Oar_Rambouillet_v1.0.

The lifted over annotation we have generated will provide a resource for those who wish to compare their results for ARS-UI_Ramb_v2.0 to previous work using Oar_Rambouillet_v1.0. Only 2.7% of protein coding transcripts were lost (1148) lifting over the annotation for Oar_Rambouillet_v1.0 onto ARS-UI_Ramb_v2.0. According to the annotation report provided by NCBI (https://www.ncbi.nlm.nih.gov/genome/annotation_euk/Ovis_aries/104/), 70% of the annotations were identical or had only minor changes between and Oar_Rambouillet_v1.0 and ARS-UI_Ramb_v2.0.

Re-use potential

The ARS-UI_Ramb_v2.0 genome assembly serves as a reference for genetic investigation of traits important in sheep research and production across the world. This genome is assembled from the same animal used in the Ovine FAANG Project, which provides a high-quality basis for epigenetic annotation to serve the international sheep genomics community and scientific community at large.

Availability of supporting data

The data sets supporting the results of this article are available in the GenBank repository, GCA_016772045.1.

Author contributions

BMM, TPLS, DMB, and BDR conceptualized the study. BMM, NEC, MPH, and TPLS selected the animal and collected samples. KW and SCM facilitated the generation of RSII, short read, and Hi-C data. TPLS facilitated the nanopore long read data generation. KMD, DMB, TPLS, BMM, and BDR performed the genome assembly, scaffolding, RNA-sequencing alignment, polishing, and quality control. MS and ELC contributed the section describing the LiftOff annotation and comparative analysis of transcript expression estimates for the three annotations. KMD, DMB, MS, ELC, TPLS, BMM, and BDR generated tables and figures and drafted the manuscript. KMD, DMB, KW, SCM, MS, ELC, NEC, TPLS, BMM, and BDR edited the manuscript. All authors contributed to the article and approved the final version.

Acknowledgements

The authors thank Dr. Kristen Kuhn for technical support and Dr. Kreg Leymaster for overseeing the acquisition, animal care and housing, and interstate transportation of the Rambouillet ewe.

Funding

Funding was provided by Agriculture and Food Research Initiative Competitive grants from the USDA National Institute of Food and Agriculture supporting improvements of the sheep genomes (2013-67015-21228) and FAANG activities (2013-67015-21372, 2017-67016-26301). Additional funding was received from the International Sheep Genome Consortium (217201191442) and infrastructure support from a grant to R. Gibbs from the NIH NHGRI Large-Scale Sequencing Program (U54 HG003273).

DMB was supported by appropriated USDA CRIS project 5090-31000-026-00-D. TPLS was supported by appropriated USDA CRIS Project 3040-31000-100-00D. BDR was supported by appropriated USDA CRIS Project 8042-31000-001-00-D. The USDA does not endorse any products or services. Mentioning of trade names is for information purposes only. The USDA is an equal opportunity employer.

References

1. Pedrosa S, Uzun M, Arranz JJ, Gutiérrez-Gil B, San Primitivo F, Bayón Y. Evidence of three maternal lineages in Near Eastern sheep supporting multiple domestication events. Proc Biol Sci. 2005;272:2211-7.

2. Zeder MA. Domestication and early agriculture in the Mediterranean Basin: origins, diffusion, and impact. Proc Natl Acad Sci USA. 2008;105:11597-604.

3. Chessa B, Pereira F, Arnaud F, Amorim A, Goyache F, Mainland I, Kao RR, Pemberton JM, Beraldi D, Stear MJ, Alberti A, Pittau M, Iannuzzi L, Banabazi MH, Kazwala RR, Zhang YP, Arranz JJ, Ali BA, Wang Z, Uzun M, Dione MM, Olsaker I, Holm LE, Saarma U, Ahmad S, Marzanov N, Eythorsdottir E, Holland MJ, Ajmone-Marsan P, Bruford MW, Kantanen J, Spencer TE, Palmarini M. Revealing the history of sheep domestication using retrovirus integrations. Science. 2009;324:532-6.

4. Kijas JW, Lenstra JA, Hayes B, Boitard S, Porto Neto LR, San Cristobal M, Servin B, McCulloch R, Whan V, Gietzen K, Paiva S, Barendse W, Ciani E, Raadsma H, McEwan J, Dalrymple B, International Sheep Genomics Consortium Members. Genome-wide analysis of the world's sheep breeds reveals high levels of historic mixture and strong recent selection. PLoS Biol. 2012;doi:10.1371/journal.pbio.1001258.

5. Burke JM, Miller JE. Relative resistance of Dorper crossbred ewes to gastrointestinal nematode infection compared with St. Croix and Katahdin ewes in the southeastern United States. Vet Parasitol. 2002;109:265-75.

6. Bowdridge SA, Zajac AM, Notter DR. St. Croix sheep produce a rapid and greater cellular immune response contributing to reduced establishment of *Haemonchus contortus*. Vet Parasitol. 2015;208:204-10.

7. Scherf BD. World watch list for domestic animal diversity. 3rd ed. Rome: Food and Agriculture Organization of the United Nations; 2000.

8. Lv FH, Agha S, Kantanen J, Colli L, Stucki S, Kijas JW, Joost S, Li MH, Ajmone Marsan P. Adaptations to climate-mediated selective pressures in sheep. Mol Biol Evol. 2014;31:3324-43.

9. Cao YH, Xu SS, Shen M, Chen ZH, Gao L, Lv FH, Xie XL, Wang XH, Yang H, Liu CB, Zhou P, Wan PC, Zhang YS, Yang JQ, Pi WH, Hehua E, Berry DP, Barbato M, Esmailizadeh A, Nosrati M, Salehian-Dehkordi H, Dehghani-Qanatqestani M, Dotsev AV, Deniskova TE, Zinovieva NA, Brem G, Štěpánek O, Ciani E, Weimann C, Erhardt G, Mwacharo JM, Ahbara A, Han JL, Hanotte O, Miller JM, Sim Z, Coltman D, Kantanen J, Bruford MW, Lenstra JA, Kijas J, Li MH. Historical Introgression from Wild Relatives Enhanced Climatic Adaptation and Resistance to Pneumonia in Sheep. Mol Biol Evol. 2021;38:838-55.

10. Dickinson WF, Lush JL. Inbreeding and the genetic history of the Rambouillet sheep in America. J Hered. 1933;24:19-33.

11. Zhang L, Mousel MR, Wu X, Michal JJ, Zhou X, Ding B, Dodson MV, El-Halawany NK, Lewis GS, Jiang Z. Genome-wide genetic diversity and differentially selected regions among Suffolk, Rambouillet, Columbia, Polypay, and Targhee sheep. PLoS One. 2013;doi: 10.1371/journal.pone.0065942.

12. Rexroad C, Vallet J, Matukumalli LK, Reecy J, Bickhart D, Blackburn H, Boggess M, Cheng H, Clutter A, Cockett N, Ernst C, Fulton JE, Liu J, Lunney J, Neibergs H, Purcell C, Smith TPL, Sonstegard T, Taylor J, Telugu B, Eenennaam AV, Tassell CPV, Wells K. Genome to Phenome: Improving Animal Health, Production, and Well-Being - A New USDA Blueprint for Animal Genome Research 2018-2027. Front Genet. 2019;10:327.

13. Jiang Y, Xie M, Chen W, Talbot R, Maddox JF, Faraut T, Wu C, Muzny DM, Li Y, Zhang W, Stanton JA, Brauning R, Barris WC, Hourlier T, Aken BL, Searle SMJ,

Adelson DL, Bian C, Cam GR, Chen Y, Cheng S, DeSilva U, Dixen K, Dong Y, Fan G, Franklin IR, Fu S, Guan R, Highland MA, Holder ME, Huang G, Ingham AB, Jhangiani SN, Kalra D, Kovar CL, Lee SL, Liu W, Liu X, Lu C, Lv T, Mathew T, McWilliam S, Menzies M, Pan S, Robelin D, Servin B, Townley D, Wang W, Wei B, White SN, Yang X, Ye C, Yue Y, Zeng P, Zhou Q, Hansen JB, Kristensen K, Gibbs RA, Flicek P, Warkup CC, Jones HE, Oddy VH, Nicholas FW, McEwan JC, Kijas J, Wang J, Worley KC, Archibald AL, Cockett N, Xu X, Wang W, Dalrymple BP. The sheep genome illuminates biology of the rumen and lipid metabolism. Science. 2014;344:1168-1173.

14. Murdoch BM. The functional annotation of the sheep genome project. J Anim Sci. 2019;97:16.

15. Salavati M, Caulton A, Clark R, Gazova I, Smith TPL, Worley KC, Cockett NE, Archibald AL, Clarke SM, Murdoch BM, Clark EL. Global Analysis of Transcription Start Sites in the New Ovine Reference Genome (*Oar rambouillet v1.0*). Front Genet. 2020;11:580580.

16. FAANG Data Coordination Center. 2016. https://data.faang.org/api/fire_api/samples/USU_SOP_Ovine_Benz2616_Tissue_Collection_20160426.pdf.

17. European Bioinformatics Institute, BioSample SAMEG329607. 2016. https://www.ebi.ac.uk/biosamples/samples/SAMEG329607.

18. Guppy (2021). Guppy basecaller (Version 3.1) www.nanoporetech.com.

19. Koren S, Walenz BP, Berlin K, Miller JR, Bergman NH, Phillippy AM. Canu: scalable and accurate long-read assembly via adaptive *k*-mer weighting and repeat separation. Genome Res. 2017;27:722-36.

20. Heaton MP, Smith TPL, Bickhart DM, Vander Ley BL, Kuehn LA, Oppenheimer J, Shafer WR, Schuetze FT, Stroud B, McClure JC, Barfield JP, Blackburn HD, Kalbfleisch

TS, Davenport KM, Kuhn KL, Green RE, Shapiro B, Rosen BD. A Reference Genome Assembly of Simmental Cattle, Bos taurus taurus. J Hered. 2021;112:184-91.

21. Oppenheimer J, Rosen BD, Heaton MP, Vander Ley BL, Shafer WR, Schuetze FT, Stroud B, Kuehn LA, McClure JC, Barfield JP, Blackburn HD, Kalbfleisch TS, Bickhart DM, Davenport KM, Kuhn KL, Green RE, Shapiro B, Smith TPL. A Reference Genome Assembly of American Bison, Bison bison bison. J Hered. 2021;112:174-183.

22. Arima Genomics Mapping Pipeline (2019). ArimaGenomics https://github.com/ArimaGenomics/mapping_pipeline.

23. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R; 1000 Genome Project Data Processing Subgroup. The Sequence Alignment/Map format and SAMtools. Bioinformatics. 2009;25:2078-9.

24. PicardTools (2019). Picard Toolkit, Broad Institute (Version 2.9.2) http://broadinstitute.github.io/picard.

25. Quinlan AR. BEDTools: The Swiss-Army Tool for Genome Feature Analysis. Bioinformatics. 2014;47:1-34.

26. Ghurye J, Pop M, Koren S, Bickhart D, Chin CS. Scaffolding of long read assemblies using long range contact information. BMC Genomics. 2017;18:527.

27. Yardımcı GG, Noble W. Software tools for visualizing Hi-C data. Genome Biol. 2017;18:26.

28. Heng L. Minimap2: pairwise alignment for nucleotide sequences. Bioinformatics. 2018;34:3094-100.

29. D-Genies (2018). D-Genies (Version 1.2.0) https://github.com/genotoul-bioinfo/dgenies/releases/tag/v1.2.0.

30. CombineFasta agp2fasta (2020). CombineFasta (Version 0.0.17) https://github.com/njdbickhart/CombineFasta.

31. Loman N, Quick J Simpson J. A complete bacterial genome assembled de novo using only nanopore sequencing data. Nat Methods. 2015;12:733-735.

32. NanoGrid (2018). NanoGrid https://github.com/skoren/NanoGrid.

33. Guan D, McCarthy SA, Wood J, Howe K, Wang Y, Durbin R. Identifying and removing haplotypic duplication in primary genome assemblies. Bioinformatics. 2020;36:2896-8.

34. Smit AFA, Hubley R, Green P. RepeatMasker Open-4.0. 2013-2015. https://www.repeatmasker.org.

35. Garrison E, Marth G. Haplotype-based variant detection from short-read sequencing. arXiv preprint. 2012:1207.3907.

36. Merfin (2021). Merfin https://github.com/arangrhie/merfin.

37. Trim Galore (2020). TrimGalore (Version 0.6.6) https://www.bioinformatics.babraham.ac.uk/projects/trim_galore/.

38. Dobin A, Davis CA, Schlesinger F, Drenkow J, Zaleski C, Jha S, Batut P, Chaisson M, & Gingeras TR. STAR: ultrafast universal RNA-seq aligner. Bioinformatics. 2013;29:15-21.

39. Li H. A statistical framework for SNP calling, mutation discovery, association mapping and population genetical parameter estimation from sequencing data. Bioinformatics. 2011;27:2987-93.

40. Shumate A, Salzberg SL. Liftoff: accurate mapping of gene annotations. Bioinformatics. 2020;37:1639–43.

41. Bray NL, Pimentel H, Melsted P, Pachter L. Near-optimal probabilistic RNA-seq quantification. Nat Biotechnol. 2016;34:525-7.

42. Quinlan AR, Hall IM. BEDTools: a flexible suite of utilities for comparing genomic features. Bioinformatics. 2010;26:841-2.

43. Themis-ASM (2020). Themis-ASM pipeline https://github.com/njdbickhart/Themis-ASM.

44. Vezzi F, Narzisi G, Mishra B. Reevaluating Assembly Evaluations with Feature Response Curves: GAGE and Assemblathons. PLoS ONE. 2012;7:e52210.

45. Rhie A, Walenz BP, Koren S, Phillippy AM. Merqury: reference-free quality, completeness, and phasing assessment for genome assemblies. Genome Biol. 2020;21:245.

46. Koren S, Rhie A, Walenz BP, Dilthey AT, Bickhart DM, Kingan SB, Hiendleder S, Williams JL, Smith TPL, Phillippy AM. De novo assembly of haplotype-resolved genomes with trio binning. Nat Biotechnol. 2018;10.1038/nbt.4277.

Tables

Table 4.1: Assembly quality statistics comparison

Assembly Statistic	ARS-UI_Ramb_v2.0	Oar_rambouillet_v1.0	Oar_v4.0	Description
Total Length (Mb)	2628.15	2869.91	2615.52	Assembly length in Mbp
Contig Number	226	7,486	48,482	Total number of contigs
Contig N50 (bp)	43,178,051	2,572,683	150,472	Half the length of the assembly is in contigs of this size or greater
Contig L50 (number of contigs)	24	313	5,008	The smallest number of contigs whose length sum make up half of the assembly size
Scaffold Number	142	2,641	5,466	Total number of scaffolds and unplaced contigs in the assembly
merQV	44.7721*	32.1705*	31.9131**	Kmer based quality from Merqury, which estimates the frequency of consensus errors in the assembly [45]
merErrorRate	0.000033327*	0.00060662*	0.000643714**	Kmer based error rate from Merqury, which estimates error rate of the assembly based on errors in kmers [45]
merCompleteness	93.0479*	93.4711*	92.2182**	Proportion of complete assembly estimated by Merqury based on “reliable” kmers, or kmers unlikely to be caused by sequencing error [45]
baseQV	41.84*	40.69*	32.40**	SNP and INDEL quality value estimated from short read data mapped to the assembly [46]
Unmap%	0.96*	1.00*	0.73**	Percentage of short reads that are unmapped to each assembly [46]
COMPLETESC	90.7	90.1	86.1	Percent of complete, single copy BUSCOs
COMPLETEDUP	1.4	1.6	1.0	Percent of complete, duplicated BUSCOs
FRAGMENT	2.0	2.1	3.7	Percent of fragmented BUSCOs
MISSING	5.9	6.2	9.2	Percent of missing BUSCOs

*Short read sequencing from the Rambouillet ewe used to assemble both ARS-UI_Ramb_v2.0 and Oar_rambouillet_v1.0 was used in these quality values.

**Short read sequencing from the Texel animal used to assemble Oar_v4.0 was used in these quality values.

Tissue	Genome *	# input reads	# reads uniquely mapped	% of reads uniquely mapped	# reads multi-mapped	% reads multi-mapped	# reads unmapped	% reads unmapped	# indels
Skin	v2.0	62,630,134	53,990,480	86.20%	6,684,213	10.67%	1,955,441	3.12%	962
	v1.0		52,523,732	83.86%	8,114,599	12.96%	1,991,803	3.18%	2,512
	Δ	N/A	1,466,748	2.34%	-1,430,386	-2.29%	-36,362	-0.06%	-1,550
Thalamus	v2.0	54,655,873	45,721,452	83.65%	5,414,620	9.91%	3,519,801	6.44%	649
	v1.0		44,904,096	82.16%	6,126,363	11.21%	3,625,414	6.63%	1,054
	Δ	N/A	817,356	1.49%	-711,743	-1.30%	-105,613	-0.19%	-405
Pituitary	v2.0	43,368,663	39,710,031	91.56%	2,405,103	5.55%	1,253,529	2.89%	604
	v1.0		34,115,417	78.66%	7,866,251	18.14%	1,386,995	3.20%	960
	Δ	N/A	5,594,614	12.90%	-5,461,148	-12.59%	-133,466	-0.31%	-356
Lymph node – mesenteric	v2.0	43,673,576	38,819,419	88.88%	3,562,121	8.16%	1,292,036	2.96%	684
	v1.0		38,296,065	87.69%	4,057,915	9.29%	1,319,596	3.02%	999
	Δ	N/A	523,354	1.19%	-495,794	-1.13%	-27,560	-0.06%	-315
Abomasum pylorus	v2.0	45,977,534	41,018,529	89.21%	2,978,042	6.48%	1,980,963	4.31%	512
	v1.0		40,403,981	87.88%	3,533,015	7.68%	2,040,538	4.44%	846
	Δ	N/A	614,548	1.33%	-554,973	-1.20%	-59,575	-0.13%	-334

Table 4.2: RNA-seq alignment statistics to ARS-UI_Ramb_v2.0 and Oar_rambouillet_v1.0 from five different tissues.

* Genomes include v2.0 (ARS-UI_Ramb_v2.0) and v1.0 (Oar_rambouillet_v1.0) and the difference (∆).

Figures

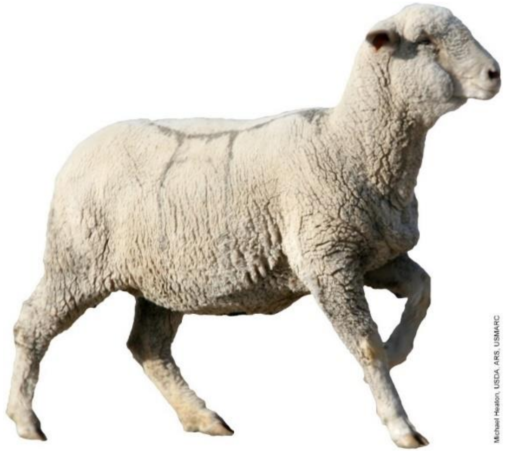


Figure 4.1: Image of Benz 2616 Rambouillet ewe selected for the ovine reference genome assembly. This image was shared by Dr. Michael P. Heaton, USDA ARS USMARC.

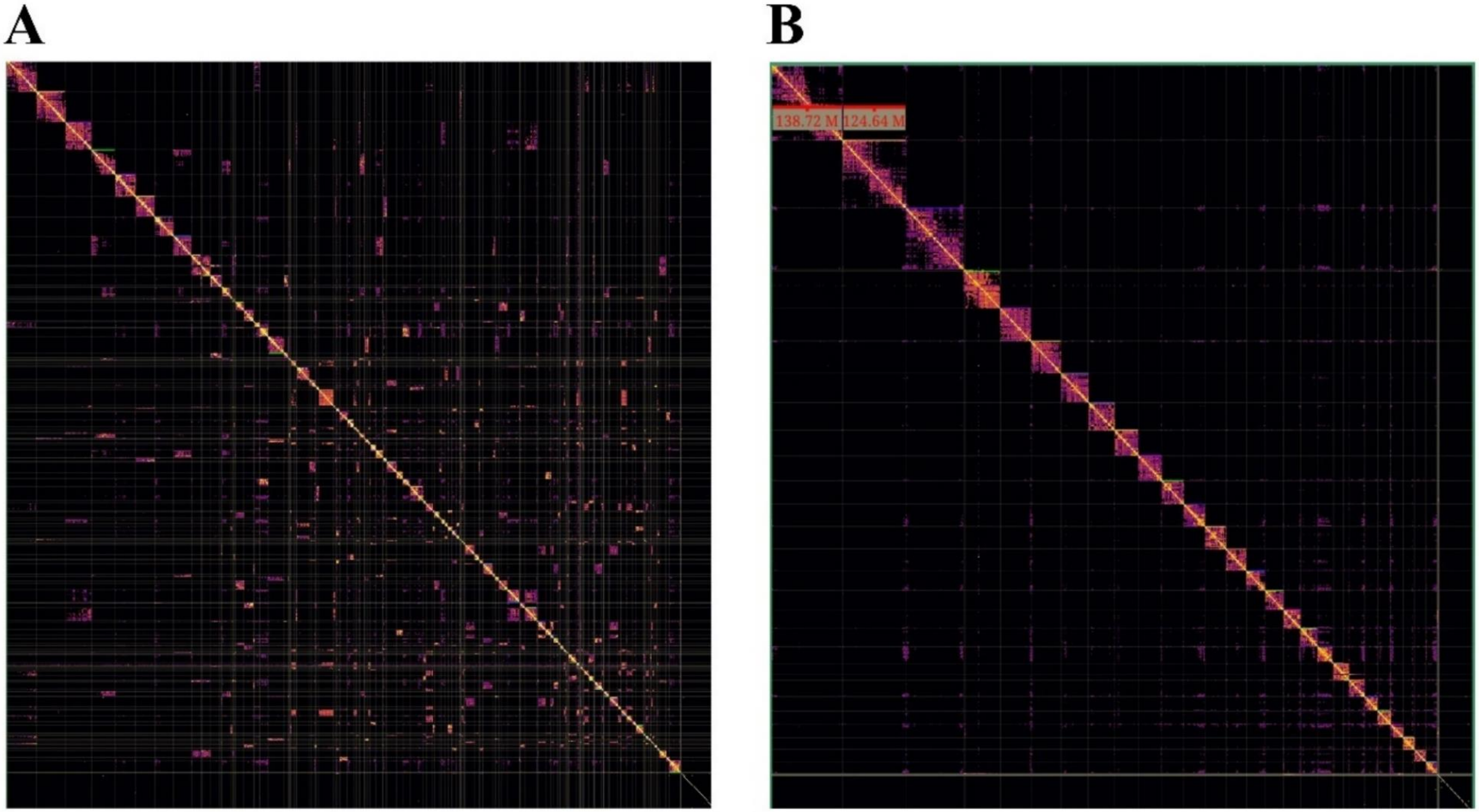
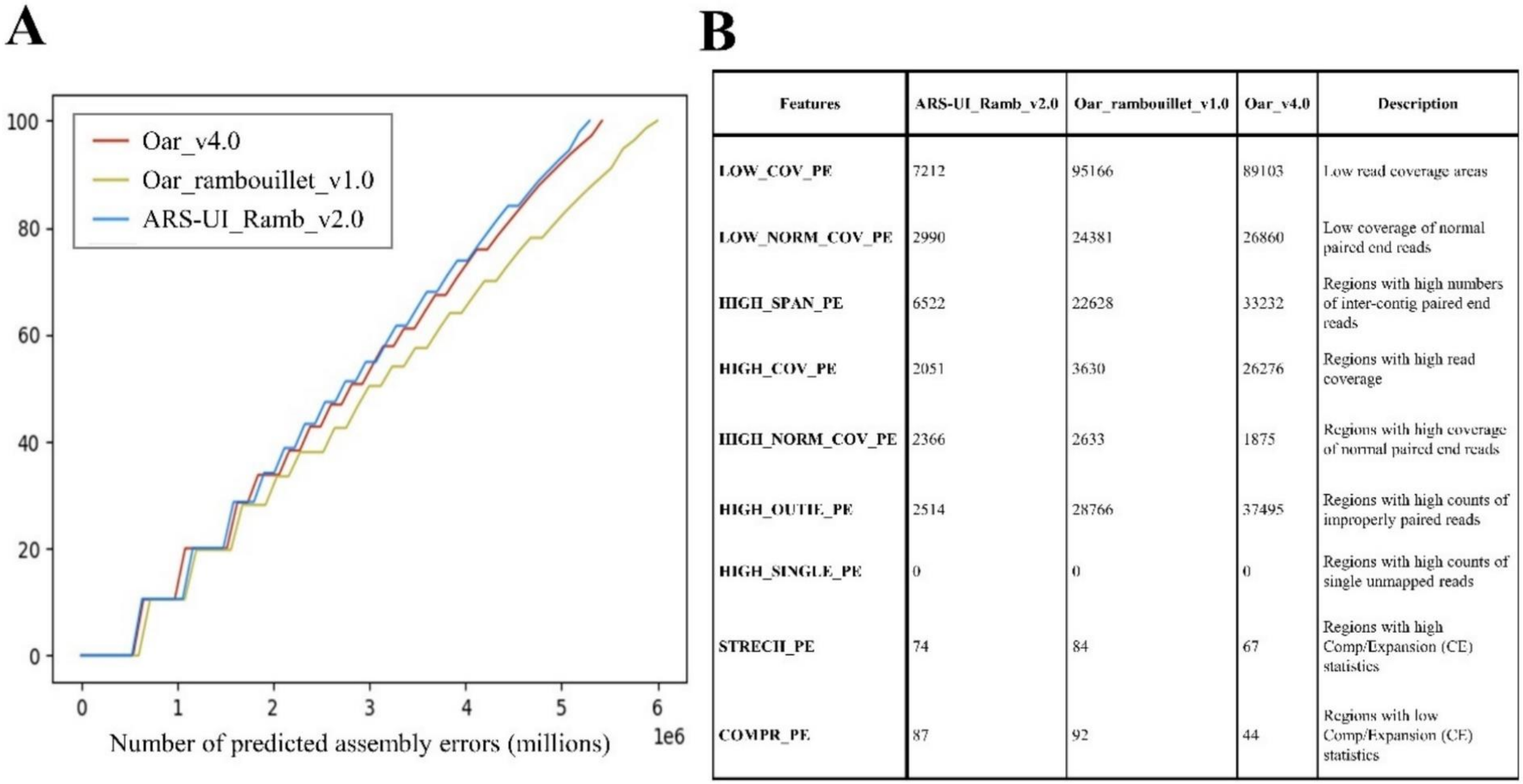


Figure 4.2: Hi-C contact map comparison of ARS-UI_Ramb_v2.0 A) directly after scaffolding and before manual curation and B) after manual curation with scaffold rearrangements and joins.



Features	ARS-UI_Ramb_v2.0	Oar_rambouillet_v1.0	Oar_v4.0	Description
LOW_COV_PE	7212	95166	89103	Low read coverage areas
LOW_NORM_COV_PE	2990	24381	26860	Low coverage of normal paired end reads
HIGH_SPAN_PE	6522	22628	33232	Regions with high numbers of inter-contig paired end reads
HIGH_COV_PE	2051	3630	26276	Regions with high read coverage
HIGH_NORM_COV_PE	2366	2633	1875	Regions with high coverage of normal paired end reads
HIGH_OUTIE_PE	2514	28766	37495	Regions with high counts of improperly paired reads
HIGH_SINGLE_PE	0	0	0	Regions with high counts of single unmapped reads
STRECH_PE	74	84	67	Regions with high Comp/Expansion (CE) statistics
COMPR_PE	87	92	44	Regions with low Comp/Expansion (CE) statistics

Figure 4.3: Assembly error comparison between ARS-UI_Ramb_v2.0, Oar_rambouillet_v1.0, and Oar_v4.0 in A) a feature response curve displaying sorted lengths of the assemblies with the fewest errors and B) specific feature counts for each genome and descriptions.

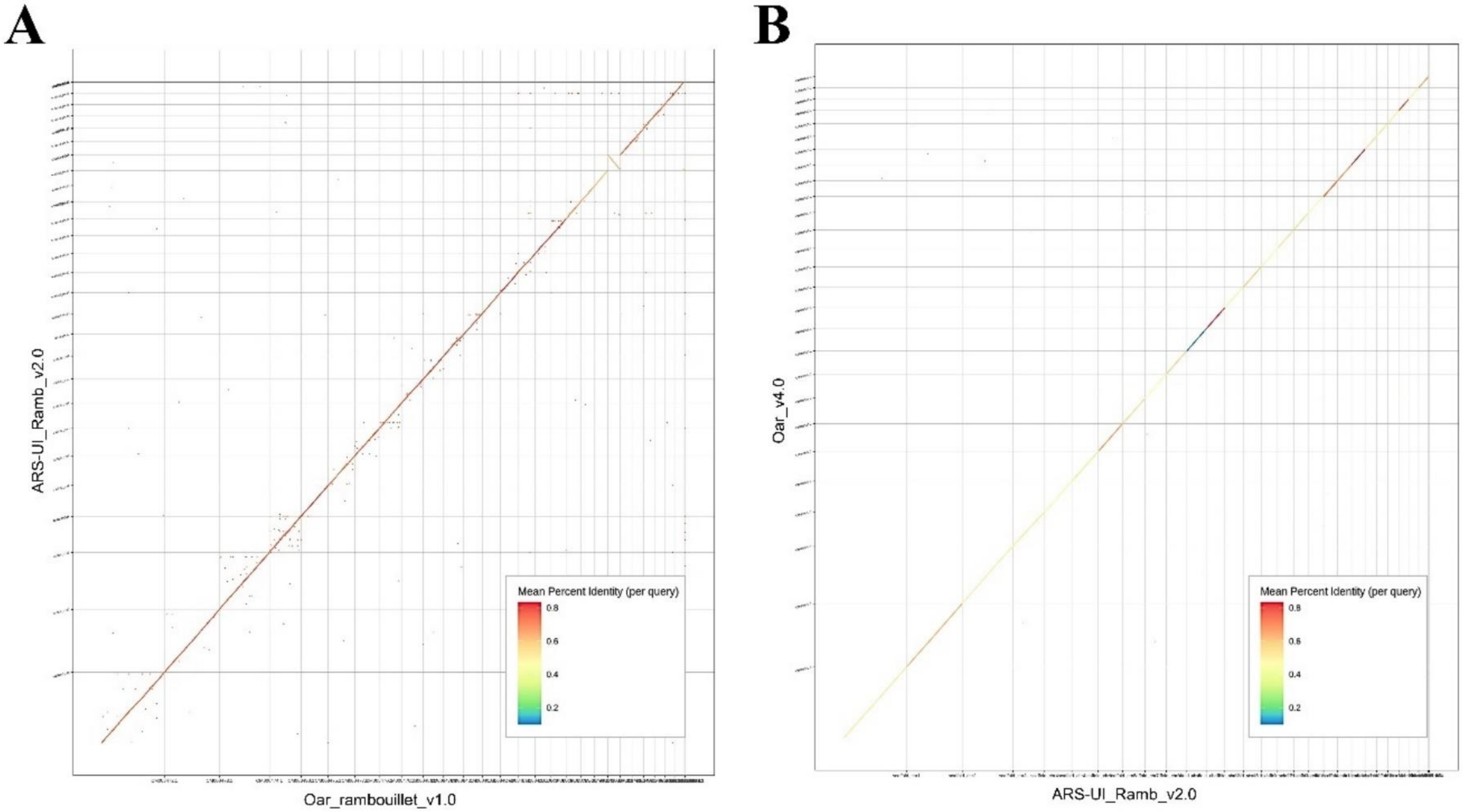
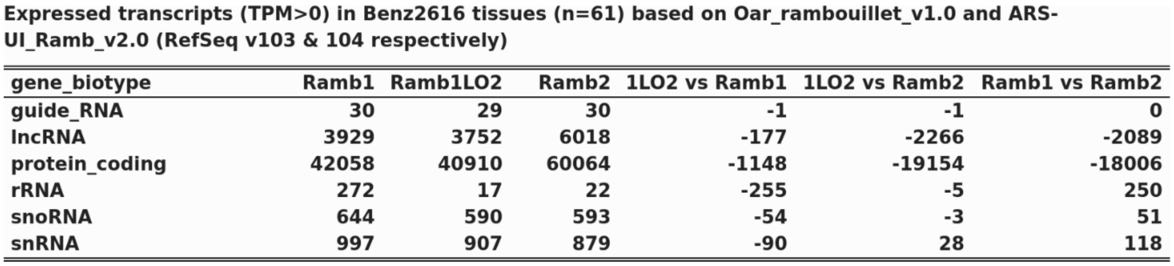


Figure 4.4: Dotplot comparison of genome assemblies between A) ARS-UI_Ramb_v2.0 and Oar_rambouillet_v1.0, and B) ARS-UI_Ramb_v2.0 and Oar_v4.0.

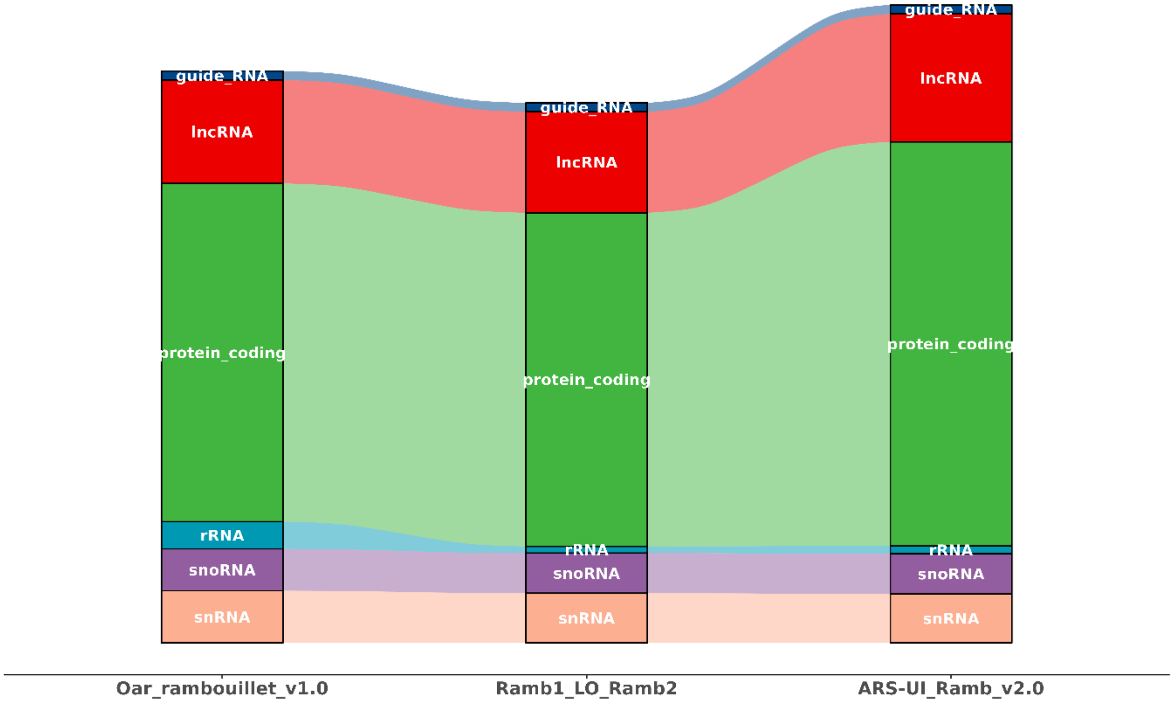
Supplementary Material

Supplementary Figure 4.1: Expressed transcripts (TPM > 0) in Benz2616 tissues (n=61) based on Oar_rambouillet_v1.0 and ARS-UI_Ramb_v2.0 (RefSeq v103 & 104 respectively).



Expressed transcripts (TPM>0) in Benz2616 tissues (n=61) based on Oar_rambouillet_v1.0 and ARS-UI_Ramb_v2.0 (RefSeq v103 & 104 respectively)

gene_biotype	Ramb1	Ramb1LO2	Ramb2	1LO2 vs Ramb1	1LO2 vs Ramb2	Ramb1 vs Ramb2
guide_RNA	30	29	30	-1	-1	0
lncRNA	3929	3752	6018	-177	-2266	-2089
protein_coding	42058	40910	60064	-1148	-19154	-18006
rRNA	272	17	22	-255	-5	250
snoRNA	644	590	593	-54	-3	51
snRNA	997	907	879	-90	28	118



Supplementary Document 4.1

```
#Reconstructing the transcriptome FASTA sequence for the lifted over gene annotation

#Tools used

module load pigz/2.3.3

module load BEDTools/2.30.0

module load samtools/1.9

module load kallisto/0.44.0


#Extracting the exonic block from the GFF file

zcat Ramb_v1.0_NCBI103_lifted_over_ARS-UI_Ramb_v2.0.gff.gz | \

awk '($3~/exon/)' | \

pigz > Ramb1LO2_NCBI103_exons.gff.gz


#Creating a modified GFF3 file format

paste <(zcat Ramb1LO2_NCBI103_exons.gff.gz | cut -f1-2) \

<(zcat Ramb1LO2_NCBI103_exons.gff.gz | cut -f9 | cut -d";" -f2| sed 's/Parent\=rna-//g') \

<(zcat Ramb1LO2_NCBI103_exons.gff.gz | cut -f4- ) > Ramb1LO2.gff


#Coversion to BED6 format

awk '{OFS="\t"; print $1,$4,$5,$3,0,$7}' Ramb1LO2.gff > Ramb1LO2.bed


#Example output

CM028704.1      42238   42395   XM_027962292.1  0       -
```

CM028704.1	42690	43127	XM_027962292.1	0	-
CM028704.1	43130	43378	XM_027962292.1	0	-
CM028704.1	43381	43588	XM_027962292.1	0	-
CM028704.1	43591	43756	XM_027962292.1	0	-
CM028704.1	43853	44085	XM_027962292.1	0	-
CM028704.1	45265	45335	XM_027962292.1	0	-
CM028704.1	46081	46232	XM_027962292.1	0	-
CM028704.1	46503	46709	XM_027962292.1	0	-
CM028704.1	74992	75652	XR_003588699.1	0	-
CM028704.1	76859	78063	XR_003588699.1	0	-
CM028704.1	78522	79261	XR_003588700.1	0	+
CM028704.1	79410	79494	XR_003588700.1	0	+
CM028704.1	147143	147427	XM_027962305.1	0	-
CM028704.1	147429	148122	XM_027962305.1	0	-
CM028704.1	148124	148170	XM_027962305.1	0	-
CM028704.1	148172	148597	XM_027962305.1	0	-
CM028704.1	150156	150304	XM_027962305.1	0	-
CM028704.1	158201	158296	XM_027964169.1	0	+
CM028704.1	164690	165052	XM_027964169.1	0	+
CM028704.1	165371	165532	XM_027964169.1	0	+
CM028704.1	166287	166321	XM_027964169.1	0	+
CM028704.1	166520	167538	XM_027964169.1	0	+

#Verifying the collapse of exon to transcript models and uniqueness

#Exons were grouped by transcript and counted per group (should be 1 == score column in the output bed)

#After sorting by the transcript id (-k 4,4) the grouping was based on the transcript name and strand (-g 4,6) and computation was done on chr,start,end and transcript id (-c 1,2,3,4). The distinct count of transcript ids for verification

```
sort -k4,4 Ramb1LO2.bed | \

bedtools groupby -g 4,6 -c 1,2,3,4 -o distinct,min,max,count_distinct | \

awk '{OFS="\t";print $3,$4,$5,$1,$6,$2}' | \

sort -V -k1,2 > Ramb1LO2_groupby.bed
```

#Example output

```
CM028704.1     42238   46709   XM_027962292.1  1      -
CM028704.1     74992   78063   XR_003588699.1  1      -
CM028704.1     78522   79494   XR_003588700.1  1      +
CM028704.1     147143  150304  XM_027962305.1  1      -
CM028704.1     158201  167538  XM_027964169.1  1      +
CM028704.1     176125  178445  XM_027964177.1  1      -
CM028704.1     183267  193065  XM_027962318.1  1      -
```

#Checking the total number of records in the final sorted BED file.

```
wc -l Ramb1LO2_groupby.bed
```

```
49899 Ramb1LO2_groupby.bed
awk '$5!=1' Ramb1LO2_groupby.bed | wc -l
0


#Extracting exonic level FASTA sequences
bedtools getfasta \
        -fi GCA_016772045.1_ARS-UI_Ramb_v2.0_genomic.fna \
        -bed Ramb1LO2.bed \
        -s -split -nameOnly > Ramb1LO2_NCBI103_geneBank_exons.fa


#Appending all exonic sequences from the same transcript id in the correct order


awk '/^>/ {if(prev!=$0) {prev=$0;printf("\n%s\n",$0);} next;} {printf("%s",$0);} END {printf("\n");}' \
        Ramb1LO2_NCBI103_geneBank_exons.fa > Ramb1LO2_NCBI103_geneBank_rna.fa


#Cleaning up the strand information from the fasta header
sed -i 's/(-)//g;s/(+)//g' Ramb1LO2_NCBI103_geneBank_rna.fa


#Buidling Kallisto index for the quantification step.
samtools faidx Ramb1LO2_NCBI103_geneBank_rna.fa
kallisto index -i Ramb1LO2_NCBI103.idx Ramb1LO2_NCBI103_geneBank_rna.fa
```

```
################################## TPM expression estimation

#!/bin/bash

#SGE flags

#$ -l h_rt=4:00:00

#$ -l h_vmem=8G

#$ -pe sharedmem 4

#$ -V

#$ -t 1-61


#Required modules

module load pigz/2.3.3

module load kallisto/0.44.0


#Kallisto runs

kallisto quant --bias -t ${vcpu} -i Ramb2_refseq104.idx -o ${sra_id}_kallisto_Ramb2 <(zcat ${infile}) <(zcat ${infile/_1P.fq.gz/_2P.fq.gz})

kallisto quant --bias -t ${vcpu} -i Ramb1_NCBI103.idx -o ${sra_id}_kallisto_Ramb1 <(zcat ${infile}) <(zcat ${infile/_1P.fq.gz/_2P.fq.gz})

kallisto quant --bias -t ${vcpu} -i Ramb1LO2_NCBI103.idx -o ${sra_id}_kallisto_Ramb1LO2 <(zcat ${infile}) <(zcat ${infile/_1P.fq.gz/_2P.fq.gz})
```

Chapter 5: Characterizing Genetic Regulatory Elements in Ovine Tissues

Published in Frontiers in Genetics.

Citation: Davenport KM, Massa AT, Bhattarai S, McKay SD, Mousel MR, Herndon MK, White SN, Cockett NE, Smith TPL, Murdoch BM; Ovine FAANG Project Consortium. Characterizing Genetic Regulatory Elements in Ovine Tissues. Front Genet. 2021 May 20;12:628849. doi: 10.3389/fgene.2021.628849.

Kimberly M. Davenport[1], Alisha T. Massa[2], Suraj Bhattarai[3], Stephanie D. McKay[3], Michelle R. Mousel[4,5], Maria K. Herndon[2], Stephen N. White[2,4,6], Noelle Cockett[7], Timothy P.L. Smith[8*], and Brenda M. Murdoch[1,6*] on behalf of The Ovine FAANG Project Consortium

[1]Department of Animal, Veterinary, and Food Science, University of Idaho, Moscow, ID
[2]Department of Veterinary Microbiology and Pathology, Washington State University, Pullman, WA
[3]University of Vermont, Burlington, VT
[4]USDA, ARS, Animal Disease Research Unit, Pullman, WA
[5]Paul G. Allen School for Global Animal Health, Washington State University, Pullman, WA
[6]Center for Reproductive Biology, Washington State University, Pullman, WA
[7]Utah State University, Logan, UT
[8]USDA, ARS, U.S. Meat Animal Research Center (USMARC), Clay Center, NE

* Correspondence:
Timothy P.L. Smith
tim.smith2@usda.gov

Brenda M. Murdoch
bmurdoch@uidaho.edu

Keywords: FAANG, epigenetics, ChIP-seq, WGBS, methylation, sheep, histone modifications, functional genomics

Abstract

The Ovine Functional Annotation of Animal Genomes (FAANG) project, part of the broader livestock species FAANG initiative, aims to identify and characterize gene regulatory elements in domestic sheep. Regulatory element annotation is essential for identifying genetic variants that affect health and production traits in this important agricultural species, as greater than 90% of variants underlying genetic effects are estimated to lie outside of transcribed regions. Histone modifications that distinguish active or repressed chromatin states, CTCF binding, and DNA methylation were used to characterize regulatory elements in liver, spleen, and cerebellum tissues from four yearling sheep. Chromatin immunoprecipitation with sequencing (ChIP-seq) was performed for H3K4me3, H3K27ac, H3K4me1, H3K27me3, and CTCF. Nine chromatin states including active promoters, active enhancers, poised enhancers, repressed enhancers, and insulators were characterized in each tissue using ChromHMM. Whole genome bisulfite sequencing (WGBS) was performed, to determine the complement of whole genome DNA methylation with the ChIP-seq data. Hypermethylated and hypomethylated regions were identified across tissues and these locations were compared with chromatin states to better distinguish and validate regulatory elements in these tissues. Interestingly, chromatin states with the poised enhancer mark H3K4me1 in spleen and cerebellum, and CTCF in liver displayed the greatest number of hypermethylated sites. Not surprisingly, active enhancers in liver and spleen, and promoters in cerebellum, displayed the greatest number of hypomethylated sites. Overall, chromatin states defined by histone marks and CTCF occupied approximately 22% of the genome in all three tissues. Further, liver and spleen displayed the greatest percent of active promoter (65%) active enhancer (81%) states in common, and liver and cerebellum displayed the greatest percent of poised enhancer (53%), repressed enhancer (68%), hypermethylated sites (75%), and hypomethylated sites (73%) in common. In addition, both known and *de novo* CTCF binding motifs were identified in all three tissues, with the highest number of unique motifs identified in cerebellum. In summary, this study has identified the regulatory regions of genes in three tissues that play key roles in defining health and economically important traits and has set the precedent for the characterization of regulatory elements in ovine tissues using the Rambouillet reference genome.

Introduction

Regulatory element characterization and chromatin state determination in relevant tissues was identified as a critical need for implementing precision breeding within the livestock industry by the Agricultural Animal Genomics Community (Rexroad et al., 2019). To this end, the Functional Annotation of Animal Genomes (FAANG) consortium and the Ovine FAANG project members seek to molecularly define the epigenome in food animals, including sheep (Andersson et al., 2015; Giuffra & Tuggle, 2019; Tuggle et al., 2016). Modelled upon the ENCODE project, (The ENCODE Project Consortium, 2012) FAANG aims to characterize the epigenome including chromatin histone modifications and DNA methylation (Andersson et al., 2015). The core objectives of the Ovine FAANG Project Consortium are to develop a deep and robust public database of transcriptional regulatory features in the sheep genome.

Sheep production for meat, milk, and wool is an important agricultural industry across the globe with more than 1 billion sheep suited to a diverse range of climates (Hegde 2019). This diversity is reflected in genetic differences between sheep breeds utilized for varied purposes (Al-Mamun et al., 2015; Meadows et al., 2008). Populations bred for different environments and for contrasting production traits provide the opportunity to study a range of phenotypes within the species. Analysis of elements that control gene expression in sheep tissues is needed as many complex traits such as rumen fatty acid metabolism, lanolin and wool production, growth, and carcass traits cannot be explained solely by variation in transcribed regions (Clark et al., 2017; Jiang et al., 2014; Villar et al., 2015; Kingsley et al., 2020). *In vivo* analysis of regulatory elements will allow researchers to test hypotheses of biological function of putative causal mutations in relevant production tissues. Understanding the phenotypic influences of genetic variance that lie in promoter and enhancer regions is important for trait prediction and the improvement of sheep production.

Functional variants that are causally implicated in phenotypic variation are increasingly found to lie outside of transcribed regions within DNA regulatory elements (Albert & Kruglyak, 2015; Xiang et al., 2019). These regulatory elements can be defined by epigenetic analyses that have not been systematically conducted in sheep. A library of putative regulatory elements in the sheep genome was recently predicted using inference from chromatin states defined in humans (Naval-Sanchez et al., 2018). However, direct experimental characterization of regulatory elements in individual ovine tissues is needed.

The work presented here represents the foundation in preparation for a deep survey using the same methodology across tissues of the index animal from which the new sheep reference genome was developed. Since the larger FAANG effort has N=2 for each tissue by design (i.e. a large array of tissues from the individual from which the genome was derived), the data collected here also provide a resource for evaluating the larger effort by permitting estimation of inter-individual variation in the appearance and tissue distribution of regulatory elements. Three tissues were selected for this study based on their prominence in defining production traits and to span tissues of endodermal, mesodermal, and ectodermal origin, and because each present unique procedural challenges for performing ChIP-seq assays. Liver is an endodermal-derived tissue that is a key metabolic component of the alimentary system (Villar et al., 2015) and contains a variety of complex carbohydrates that can inhibit various enzymatic reactions required in the ChIP-seq protocol. Spleen is a mesodermal-derived parenchymatous organ important for immune cell production and maturation and contains many natural deoxyribonucleases (DNase) which can present challenges to obtaining sufficient yield of high-quality DNA (Young & Sinsheimer, 1965). Cerebellum is an ectodermal-derived tissue representative of brain tissue and contains a high lipid content which can affect the efficiency of DNA extraction. With these three varied tissues, we developed workflows for assessing chromatin-associated histone modifications, CTCF binding sites, and DNA methylation to define regulatory elements.

The histone modifications characterized in this study include the trimethylation of histone 3 lysine 4 (H3K4me3) which denotes promoters and acetylation of histone 3 lysine 27 (H3K27ac) which denotes active enhancers (Barski et al., 2007; Wang et al., 2008). The monomethylation of histone 3 lysine 4 (H3K4me1) was characterized to explore poised enhancers, and the trimethylation of histone 3 lysine 27 (H3K27me3) was utilized to define repressed enhancers which silences gene expression in broad regions (Barski et al., 2007; Wang et al., 2008; Pauler et al., 2009). The CCTC-binding factor protein (CTCF) is a key component of the anchors at topologically associated domain boundaries (Ghirlando & Felsenfeld, 2016; Lee & Iyer, 2012). Determination of CTCF and multiple histone modifications, referred to as marks, allowed us to take advantage of the combinatorial nature of chromatin structure and gene expression regulation (Jenuwein & Allis, 2001; Wang et al., 2008) to categorize the sheep genome into chromatin states.

DNA methylation data derived from WGBS was incorporated to validate regulatory regions and chromatin states. In mammals, several groups have identified CpG islands that lack methylation are located at gene promoters (Deaton & Bird, 2011). Repressed promoters are marked by higher degrees of methylation associated with transcriptionally silenced gene expression (Weber et al., 2007). Histone methylation and DNA methylation are co-dependent epigenetic marks as enzymatic formation of one will guide the formation of the other and H3K4me3 may physically inhibit methylation of DNA during development (Meissner et al., 2008). Histone methylations and DNA methylation serve as templates for rebuilding one another during mitosis and meiosis and further reinforce segmentation of the genome into functional regions of active or repressed chromatin in adult somatic cells (Cedar & Bergman, 2009) justifying the utility of combined analysis in sheep.

Our objective for this study was to identify the locations of gene regulatory elements in sheep by characterizing histone modifications, CTCF binding, and DNA methylation for cerebellum, liver, and spleen. Defining regulatory elements in the sheep genome will provide the basis for a greater understanding of the mechanisms that underpin phenotypic variation in important health and production traits in sheep.

Materials & Methods

Sample collection

Tissue was collected postmortem from two pairs of healthy half siblings (one ewe and one wether per pair) totaling four yearling crossbred sheep (Columbia, Polypay, Rambouillet, Suffolk, Targhee) as approved by the Washington State University Institutional Animal Care and Use Committee. Small pieces of liver, spleen, and cerebellum tissues were collected within 30 minutes postmortem, briefly rinsed with ice cold 1 X PBS and promptly snap frozen in liquid nitrogen. Samples were transferred from liquid nitrogen directly into a -80 °C freezer for storage.

Chromatin immunoprecipitation

Chromatin Immunoprecipitation (ChIP) was performed using commercial antibodies for the histone modifications H3K4me3 (Abcam, cat. #ab8580), H3K4me1 (Abcam, cat. # ab8895), H3K27ac (Abcam, cat. #ab4729), H3K27me3 (Abcam, cat. #ab6002), and CTCF (Millipore, cat. #07-729) with SimpleChIP Plus Enzymatic Chromatin IP Kit according to manufacturer’s instructions (Cell Signaling Technologies cat. #9005, Danvers, MA, USA) (Johnson et al., 2007; Barski et al., 2007; Robertson et al., 2007; Mikkelsen et al., 2007; Park 2009). Briefly, tissue was cross-linked with 37% formaldehyde and disaggregated with a Dounce homogenizer. After cell membrane lysis, micrococcal nuclease (MNase) was added and incubated at 37°C and 200 rpm for 20 minutes to shear the chromatin. Next, the nuclear membrane was lysed, and the sheared chromatin isolated by centrifuging at 15,000 *x g* for 1 minute at 4°C. Chromatin was incubated with 1 µg of antibody overnight at 4 °C in a Hula mixer for 16 hours. The following morning, protein G-coated magnetic beads were added and incubated 2 hours at 4°C in a Hula mixer. The sample was washed twice with a low salt and once with a high salt buffer. Cross-links were reversed by incubating the sample at 65 °C for 30 minutes at 400 rpm in a thermomixer. Purification was performed with the DNA Purification Buffers and Spin Columns Kit following manufacturer’s instructions (Cell Signaling Technologies cat. #14209, Danvers, MA, USA).

ChIP-seq library preparation and sequencing

Purified DNA samples were quantified using the Qubit dsDNA HS Assay Kit (Thermo Fisher Scientific catalog number Q32854, Waltham, MA, USA). The DNA size and integrity were verified using a Fragment Analyzer (Agilent, Santa Clara, CA, USA). Libraries were prepared with the TruSeq ChIP Library Preparation Kit (Illumina, Inc. catalog number IP-202-1012, San Diego, CA, USA) for 75 base pair paired-end reads following manufacturer’s instructions and sequenced to at least 20 million mapped reads for “narrow” histone marks H3K4me3, H3K27ac, and CTCF libraries and at least 40 million mappable reads each for “broad” histone marks H3K4me1 and H3K27me3 libraries.

Whole genome bisulfite sequencing library preparation and sequencing

Whole genome bisulfite sequencing (WGBS) was performed as a service by Novogene (Novogene, Beijing, China) on liver, spleen, and cerebellum in all four animals. Briefly, DNA extracted from these tissues were subjected to agarose gel electrophoresis to test for DNA degradation and potential RNA contamination. The DNA was then quantified using Nanodrop spectrophotometer (NanoDrop Technologies, Rockland, DE), and Qubit2.0 fluorometer (Life Technologies, Carlsbad, CA, USA). Lambda phage DNA was spiked in as a negative control for DNA methylation. Since lambda phage DNA lacks DNA methylation, all the cytosines in its DNA should be converted to uracil during bisulfite conversion. Any unchanged cytosine in the lambda phage DNA can thus be used to determine the efficiency of bisulfite conversion. For library construction, DNA samples were fragmented into 200-400 bp using sonication (Covaris S220, Woburn, MA, USA). Next, end repair, A-ligation, and methylation sequencing adapter ligation was performed. The adapter sequences were 5’ adapter (5’-AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTCCGATCT – 3’) and 3’ adapter (5'-GATCGGAAGAGCACACGTCTGAACTCCAGTCACATCACGATCTCGTATGCCGTCTTCTGCTTG-3’). Following this, the DNA library was subjected to bisulfite treatment (EZ DNA Methylation Gold Kit, Zymo Research, Irvine, CA, USA). Library concentration was first quantified by Qubit2.0, was diluted to 1 ng/µl before checking insert size on Agilent

2100 (Agilent Technologies, Santa Clara, CA, USA), and was quantified with more accuracy by quantitative PCR (effective concentration of library > 2 nM). Libraries were then pooled per sample and sequenced paired end.

ChIP-seq data quality control, mapping, and peak calling

Quality control assessment of ChIP-seq reads were performed with FastQC, and Trim Galore was used to trim adapter sequences and low-quality bases. PCR duplicates were removed with Picard and the remaining read pair sequences were then mapped to the sheep reference genome *Oar_rambouillet_v1.0* with Bowtie2 (Langmead and Salzburg, 2012; Broad Institute, 2019). Cross-correlations were calculated using MACS2 *predictd* in Galaxy Version 2.1.1.20160309.1 (Supplementary Figure 5.1) (Afgan et al., 2018). Peaks for narrow histone marks H3K4me3 and H3K27ac as well as transcription factor CTCF were called using MACS2 with an input control and a false discovery rate of 0.05 (Zhang et al., 2008; Feng et al., 2012; Thomas et al., 2017). For broad peak histone modifications H3K4me1 and H3K27me3, SICER was implemented with the same input control and a false discovery rate of 0.05 to better account for broader sequence pileup distributions (Zang et al., 2009; Micsinai et al., 2012; Siska and Kechris, 2017). The number of uniquely mapped sequence, non-redundant fraction (NRF), and fraction of reads in peaks (FRiP) for each ChIP-seq sample was calculated using Picard (Afgan et al., 2018; Heinz et al., 2010; Landt et al., 2012; Friedman and Alm, 2012; Siska and Kechris, 2017) (Supplementary Table 5.1). Peak numbers were averaged across samples. Peaks common to multiple samples were determined with BEDTools intersect. The peaks common to three samples with the greatest NRF were determined for H3K4me3 (F1, M1, and M2 for liver, F2, M1, and M2 for spleen, and F1, M1, and M2 for cerebellum), H3K27ac (F1, M1, and M2 for liver, F2, M1, and M2 for spleen, and F1, M1, and M2 for cerebellum), H3K4me1 (F1, M1, and M2 for liver, F2, M1, and M2 for spleen, and F1, M1, and M2 for cerebellum), H3K27me3 (F1, M1, and M2 for liver, F2, M1, and M2 for spleen, and F2, M1, and M2 for cerebellum), and CTCF (F2, M1, and M2 for liver, F2, M1, and M2 for spleen, and F2, M1, and M2 for cerebellum). These consensus peaks were compared with transcription start site locations identified with CAGE assays from the ewe used to generate the reference genome using the deepTools *computeMatrix* function and heatmaps were plotted with the *plotHeatmap* function (Ramirez et al., 2016; Salavati et al., 2020). Further, peaks were annotated with the GTF file from reference genome *Oar_rambouillet_v1.0* and peaks were categorized as near a TSS (+2Kb to -2Kb), Exonic, Intronic, TTS (+1Kb to -1Kb), and Intergenic using the Homer *annotatePeaks.pl* function (Heinz et al., 2010). Further, normalized bigwig files depicting the sequence enrichment for each library were directly visualized with Integrative Genomics Viewer (IGV) for some gene regions which are known to be active and repressed in each tissue (Robinson et al., 2011). Spearman correlations were calculated between sample BAM signal files using deepTools in Galaxy Version 2.1.1.20160309.1 (Friedman and Alm, 2012; Ramírez et al., 2014; Siska and Kechris, 2017; Afgan et al., 2018).

DNA methylation data quality control, mapping, and methylation level characterization

The quality of raw sequences from whole genome bisulfite sequencing was assessed using FastQC v0.11.5. Adapters and low-quality bases (phred score < 20) were trimmed using Trimgalore v0.4.5 with default parameters. Cleaned data for each sample was aligned to the sheep reference genome *Oar_rambouillet_v1.0* using Bowtie2 aligner within BSseeker2 v2.1.8 with default parameters (Langmead et al., 2012; Guo et al., 2013). The X-chromosome was removed from the analysis to make male and female samples comparable. After mapping, bam files for the same individual sequenced on multiple lanes were merged, fixmated, sorted and PCR duplicates were removed using Samtools v1.6 (Li et al., 2009). The methylation level in each cytosine was determined using BSseeker2 with default parameters. Basic statistics on methylation were determined using the *mstat* fnction in CGmaptools v0.0.6 (Guo et al., 2018). Regions of the genome hypomethylated and hypermethylated for each sample were determined with methPipe v3.4.3 following the manual with default parameters (Song et al., 2013).

Chromatin state and CTCF motif analysis

Chromatin states were characterized by employing a Hidden Markov Model in ChromHMM, which assessed signal overlap between histone marks within a tissue and binned the genome into a given number of chromatin states (Ernst and Kellis 2010; Ernst and Kellis 2012; Ernst and Kellis, 2017; Gorkin et al., 2017; Gorkin et al., 2020). The two male samples (M1 and

M2) exhibited the greatest NRF and Spearman correlations and were therefore used in chromatin state analysis. The *LearnModel* function in ChromHMM was implemented with given chromatin states of 2 through 20 for each animal, and the model with the optimal number of chromatin states was examined using the *CompareModels* function in ChromHMM (Gorkin et al., 2017; Gorkin et al., 2020). The optimal number of chromatin states was determined as the model where the median Pearson correlation for all states plotted against each chromatin state model plateaued and were tightly correlated with the model with greatest number of states (Supplementary Figure 5.2) (Gorkin et al., 2017; Gorkin et al., 2020). The consensus of chromatin states between two animals (M1 and M2) was used to generate the heatmap and for further comparative analyses. Location similarities and differences between chromatin states, hypermethylated regions, and hypomethylated regions were assessed with BEDTools intersect within each tissue, and the consensus within each tissue was used to examine chromatin state and DNA methylation similarities and differences between liver, spleen, and cerebellum tissues (Quinlan, 2014). An UpsetR plot was generated to display chromatin state similarities and differences between tissues (Lex et al., 2014; Conway et al., 2017). Significantly enriched known and *de novo* CTCF motifs were identified and compared to other species by implementing the findMotifs.pl script in HOMER (Heinz et al., 2010). The proximity of annotated TSS generated from CAGE data to promoter chromatin states was examined with deepTools *computeMatrix* and *plotHeatmap* functions (Supplementary Figure 5.7) (Ramirez et al., 2016; Salavati et al., 2020).

Results

Genetic regulatory elements were characterized across the sheep genome in liver, spleen, and cerebellum using CTCF binding and chromatin immunoprecipitation with sequencing (ChIP-seq) of four histone marks, as well as DNA methylation status. Locating regulatory elements within and between tissues will provide the basis for identifying variation in these elements that may influence various phenotypic traits in sheep. Further, these results represent a resource for estimating inter-individual variation in the regulatory states of tissues to provide context for the FAANG project that aims to characterize these states in a broad array of tissues in a single individual from which the reference genome was produced.

Mapping summary and statistics

Mapping statistics were calculated to assess the assay quality, library preparation, and sequence coverage for each sample. Across animals, ChIP-seq reads had a consistent average mapping rate of 78.23%, 78.39%, and 76.82% to the *Oar_rambouillet_v1.0* genome for liver, spleen, and cerebellum, respectively. The number of uniquely mapped paired end reads averaged 40,757,252 for H3K4me3, 42,306,275 for H3K27ac, 53,171,657 for H3K4me1, 55,901,184 for H3K27me3, and 45,491,017 for CTCF across all three tissues. The number of uniquely mapped reads, NRF, and FRiP for each sample are displayed in Supplementary Table 5.1.

Whole genome bisulfite sequencing of cerebellum, liver, and spleen samples from the four sheep generated a total of 986, 1070, and 904 million paired end reads, respectively with a read length of 2 x 150 bp. The number of reads uniquely mapped to the reference genome was 84.24%, 78,86% and 82.48% for cerebellum, liver and spleen, respectively. The uniquely mapped bases covered the reference genome (*Oar_rambouillet_v1.0*; genome size ~2.87 Gb) at an average depth of 21x (range 18x to 26x). Bisulfite conversion rate was ~99.9% for all the samples. Mapping statistics for each tissue sample per sheep are displayed in Supplementary Table 5.2.

ChIP-seq peak calling

The locations of sequence signal enrichment were identified for all four histone marks and CTCF for each liver, spleen, and cerebellum sample by mapping the reads to the reference genome *Oar_rambouillet_v1.0*. The number of peaks normalized by chromosome length (in Mb; Figure 5.1) and the width of the peaks along the assembly were calculated from the mapped read depth. For each mark, the percent of the total number of peaks observed in the genome that lie on each chromosome is plotted in Figure 1 which shows an overall even distribution across chromosomes with some exceptions. The lowest number of peaks were called in narrow mark H3K4me3 (means of 10,458 in liver, 13,389 in spleen, and 16,911 in cerebellum), with the lowest number of peaks per Mb on chromosome 23 2.77 peaks/Mb), 26 (2.64 peaks/Mb), and 16 (2.47 peaks/Mb) in liver, spleen, and cerebellum, respectively. The greatest number of H3K4me3 peaks per Mb for liver, spleen, and cerebellum were on chromosomes 14 (6.16 peaks/Mb), 20 (5.17 peaks/Mb), and 11 (4.61 peaks/Mb), respectively. The average widths of H3K4me3 peaks were 168 bp, 178 bp, and 313 bp for liver, spleen, and cerebellum. The mean number of peaks called for the H3K27ac mark was 30,553 in liver, 35,327 in spleen, and 35,877 in cerebellum with the lowest number of peaks called on chromosomes 10 (2.54 peaks/Mb), 26 (2.25 peaks/Mb), and 6 (2.72 peaks/Mb) for the respective tissues. The greatest number of H3K27ac peaks were called on chromosome 11 for all three tissues and peak widths averaged 239 bp, 240 bp, and 238 bp in liver, spleen, and cerebellum for this narrow mark. The final narrow mark, CTCF, averaged 26,517 peaks in liver, 28,362 in spleen, and 26,244 in cerebellum. The lowest number of CTCF peaks were called on chromosome 24 (1.56 peaks/Mb for liver, 1.49 peaks/Mb for spleen, and 2.05 peaks/Mb in cerebellum), and the greatest number of peaks were called on chromosome 6 (5.50 peaks/Mb in liver, 5.73 peaks/Mb in spleen, and 5.07 peaks/Mb in cerebellum) for all three tissues. The width of CTCF peaks were similar to other narrow marks, with averages of 114 bp in liver, 265 bp in spleen, and 144 bp in cerebellum.

The greatest number of peaks were called in broad mark H3K4me1 (means of 47,828 in liver, 33,931 in spleen, and 51,766 in cerebellum), which is consistent with several tissues in cattle (Fang et al., 2019). Chromosomes with the lowest number of H3K4me1 peaks per Mb included 21 (2.34 peaks/Mb) for liver, 26 (2.90 peaks/Mb) for spleen, and 20 (3.12 peaks/Mb) for cerebellum, and the greatest number of peaks per Mb were on chromosome 7 (4.99 peaks/Mb for liver, 7.79 peaks/Mb for spleen, and 4.98 peaks/Mb in cerebellum) for all three tissues. The average width of broad peak H3K4me1 was greater than for the narrow peaks described above, as expected, at 948 bp for liver, 2,963 bp for spleen, and 1,909 bp for cerebellum. And lastly, the broad mark H3K27me3 had a lower average number of peaks called compared to H3K4me1 (mean of 39,162 in liver, 29,939 in spleen, and 26,244 in cerebellum). The lowest number of H3K27me3 peaks per Mb of chromosome length were on chromosomes 26 (3.04 peaks/Mb), 24 (2.58 peaks/Mb), and 11 (1.84 peaks/Mb) for liver, spleen, and cerebellum, respectively. The greatest number of peaks on chromosomes 13 (4.86 peaks/Mb) for liver, and 6 for both spleen (4.39 peaks/Mb) and cerebellum (4.94 peaks/Mb). The average width of broad H3K27me3 peaks was 440 bp in liver, 2,143 bp in spleen, and

653 bp in cerebellum. Peaks in common across the animals were calculated for all five ChIP-seq experiments and displayed for liver, spleen, and cerebellum (Supplementary Figure 5.2). Interestingly, half siblings (F1 and M1, F2 and M2) displayed a greater number of peaks in common with each other.

The proximity of H3K4me3 peaks to transcription start sites (TSS) was investigated by comparing consensus H3K4me3 peaks and CAGE data generated by Salavati et al., 2020. Not surprisingly, H3K4me3 peaks were detected on both sides of the TSS in liver, spleen, and cerebellum tissues. The signal distributions and heatmaps from 2 kilobases upstream and downstream of the TSS locations are displayed in Figure 2. In addition, the consensus peaks for H3K4me3, H3K27ac, H3K4me1, H3K27me3, and CTCF were annotated with the *Oar_rambouillet_v1.0* genome annotation file and these classifications are displayed in Supplementary Figures 3-5. The histone modification H3K4me3 had the greatest proportion of peaks annotated as near a TSS when compared with other histone modifications in all three tissues. H3K27ac and H3K4me1 histone modifications displayed intronic annotation most commonly, and H3K27me3 and CTCF displayed mostly intergenic peak annotation.

Visual assessment of sequence pileup

The peak predictions were directly examined in the Integrative Genomics Viewer (IGV; Robinson et al., 2011) for regions known to be active or repressed in the three tissues, to provide an evaluation of the success of the process in properly classifying chromatin states. One example of an expected active region for each liver, spleen, and cerebellum tissue as well as one region expected to be repressed in all tissues is displayed in Figure 5.3. Albumin (*ALB*), a gene that encodes a plasma protein synthesized in hepatocytes and expected to be active in liver, has one promoter and two enhancers annotated in humans that are within 2 kb upstream from the start of the gene (Frain et al., 1990; Hayashi et al., 1992; Bernardi et al., 2012; Fagerberg et al., 2014). Sequence pileup for active histone marks in liver were observed in all four sheep that overlap with approximate locations of regulatory elements of *ALB* in humans, and low levels of DNA methylation in these regions (Figure 5.3A). The region upstream of Solute carrier family 11 member 1 (*SLC11A1*), a gene expected to be active in spleen and encodes a membrane protein involved with macrophage development, displayed sequence pileup for active marks H3K4me3 and H3K27ac and low levels of DNA methylation directly upstream (Figure 5.3B) (Hedges et al., 2014). Paired box 6 (*PAX6*) is known to be involved in development of neural tissues and maturation of granule neurons in the cerebellum, and is known to have a promoter and multiple enhancers both upstream and downstream of the gene (Ha et al., 2015; Divya et al., 2016). Further, *PAX6* has greater expression in cerebellum than other tissues in sheep which is supported by the sequence pileup of active histone marks H3K4me3 and H3K27ac, with some activity of H3K4me1 and little DNA methylation (Figure 5.3C) (Jiang et al., 2014). In contrast, the REC8 meiotic recombination protein (*REC8*) is a gene that encodes a meiosis specific protein involved in synapsis of sister chromatids that is not expected to be active in liver, spleen, or cerebellum (Xu et al., 2005). This gene location shows no sequence pileup in all four sheep in liver, spleen, or cerebellum and several methylated regions (Figure 5.3D).

Variability in histone marks between animals

Correlations were calculated for histone marks and for DNA methylation between samples to evaluate inter-animal variation in sequence pileup signal for liver, spleen, and cerebellum (Friedman and Alm, 2012; Siska and Kechris, 2017). Correlations of ChIP-seq data (Spearman) and DNA methylation data (Pearson) averages for all four animals and males only (in parentheses) are provided in Table 1. The narrow mark H3K4me3 was moderately correlated between all four animals in liver (0.66) and spleen (0.54), and highly correlated in cerebellum (0.85). In males, H3K4me3 was highly correlated in liver (0.86), spleen (0.71), and cerebellum (0.88). The narrow mark H3K27ac was highly correlated between samples across all three tissues in liver (0.89 overall and 0.95 in males), spleen (0.78 overall and 0.84 in males), and cerebellum (0.70 overall and 0.91 in males).

The broad mark H3K4me1 also showed high correlation in two tissues including liver (0.71 overall and 0.93 in males) and cerebellum (0.82 overall and 0.91 in males) but the correlation in spleen was markedly lower (0.47 overall and 0.56 in males) and overall the correlations between spleen samples were lower than liver and cerebellum for all four histone marks. This is evident in H3K27me3 in spleen (0.37 overall and 0.44 in males) compared to liver (0.58 overall and 0.74 in males) and cerebellum (0.72 overall and 0.83 in males). The correlations of DNA methylation signal between samples ranged from 0.70-0.76, with liver and cerebellum displaying the greatest correlation between the two males (0.76). However, sex differences in correlations were not observed, as each female has a moderate to high correlation with both the other female (0.54-0.84) and both males (0.44-0.92) for each mark within all three tissues.

Principal component analysis of DNA methylation

A principal component analysis was performed with the DNA methylation data to investigate similarity and differences between samples and tissues. Eigenvalues were calculated based on position of CG methylation signal in all animals for all three tissues, and the first two eigenvalues (PC1 and PC2) were plotted (Figure 5.4). Samples cluster distinctly by tissue type rather than by sex or individual animal. The greatest spread of points within a tissue was observed in liver. The first eigenvalue (PC1, 27.56%) shows separation of liver, spleen, and cerebellum. The second eigenvalue (PC2, 12.16%) shows another dimension of separation of cerebellum and liver from spleen.

Methylation level at CG and non-CG sites

Average methylation levels were calculated and compared in each of the three tissues in both the CG and non-CG sites (Figure 5.5A). Non-CG sites are defined as CHG and CHH where H is either A/T/C. CG sites have an average methylation level ranging between 70-81% across different tissues. Specifically, cerebellum samples have an average methylation level of 81.4% whereas liver and spleen samples have an average methylation level of 70.3% and 76.9%, respectively. The average methylation level of cytosines at non-CG contexts (CHG and CHH) is nine-fold higher in cerebellum (1.7-2.1%) compared to spleen and liver samples (0.2%) (Figure 5.5B).

Chromatin state assignment and correlation with methylation status

The relative positions of the combination of specific histone marks provide a more complete definition of the overall regulatory chromatin state than individual peak calling. Regulatory elements were defined for two animals (M1 and M2) using a hidden Markov model employed by ChromHMM which assigns 200bp bins across the genome to a given number of chromatin states based on combined histone modification signal profiles (Ernst and Kellis 2010; Ernst and Kellis, 2017). The genome was categorized into 2 through 20 chromatin states using ChromHMM. The optimal number of states was determined to be 9, as it was the lowest number of states that had greater than 0.95 correlation of all samples to 20 states, which captures the complexity of the data with fewer states (see Supplementary Figure 5.2) (Gorkin et al., 2017; Gorkin et al., 2020). These 9 chromatin states are categorized as: promoter, active enhancer, poised enhancer, repressed enhancer, CTCF, and three or four states of quiescent/low signal. The consensus of chromatin states assigned to both M1 and M2 was used for further analyses.

The signal of all the histone marks and the 9 chromatin states for each tissue is displayed as heatmaps in Figure 6. Regions with primarily H3K4me3 signal often overlapping with H3K27ac are considered promoters, regions with strong H3K27ac signal are considered active enhancers, regions with H3K4me1 often paired with weak H3K27me3 signal are considered poised enhancers, and regions with strong H3K27me3 signal are considered repressed enhancers (Wang et al., 2008; Creyghton et al., 2010; Core et al., 2011; Carelli et al., 2018). All four of these categories of regulatory elements were observed and displayed in the heatmaps, with the addition of a weak poised enhancer state in spleen and weak repressed enhancer state in cerebellum which both displayed lower but still distinguishable signal. In addition, regions with CTCF signal which overlap with other marks including H3K4me1 and H3K27me3 were observed in liver and cerebellum. Lastly, quiescent/low states had very little signal in any of the five marks.

The correlation of DNA methylation status with predicted chromatin state was examined by estimating the number of hyper- and hypomethylated regions per Mb within the boundaries of the regulatory elements in the 9 defined chromatin states. The greatest number of hypomethylated regions were observed in active enhancer regions in liver and spleen, and in active promoter regions in cerebellum, as expected if our process was correctly identifying regulatory elements and classifying them as actively transcribed genes. The greatest number of hypermethylated regions were observed in poised enhancers and CTCF in liver, weak poised and poised enhancer regions in spleen, and poised enhancer regions in cerebellum, also consistent with the process correctly classifying regulatory elements.

Distribution of chromatin states in the genome and proximity to TSS

The chromosomal segments spanned by regulatory elements, as defined by the histone mark peaks, were combined and summarized to estimate the overall extent and percent of the genome representing regulatory elements and their chromatin state among the three tissues examined. Chromatin states from the ChromHMM analyses were categorized and combined into promoter, active enhancer, poised enhancer including weak poised enhancers, repressed enhancer including weak repressed enhancers, and quiescent or low signal categories and averaged for each tissue (Figure 5.7). Promoters comprise 2.95% of the genome in liver, 3.35% in spleen, and 1.85% in cerebellum, and active enhancers occupy 5.04% of the genome in liver, 4.30% in spleen, and 3.74% in cerebellum. In addition, 4.38% of the genome in liver, 4.63% in spleen, and 2.68% in cerebellum are categorized as poised enhancers while 7.78% of the genome in liver, 4.96% in spleen, and 9.89% are cerebellum are considered repressed enhancers. The percent of the genome that had primarily CTCF signal was 2.92% in liver, 3.19% in spleen, and 2.94% in cerebellum. Cumulatively, states considered as enriched with histone mark and CTCF signal intensity by ChromHMM, which includes promoter, enhancer, and CTCF functional elements, comprises approximately 23.08% of the genome in liver, 20.44% in spleen, and 21.10% in cerebellum. Not surprisingly, the largest percent of the genome, 76.91% in liver, 79.56% in spleen, and 78.90% in cerebellum was categorized as quiescent or low signal.

The locations of assigned promoter chromatin states were compared with TSS generated from CAGE data for liver, spleen, and cerebellum. Both the signal distribution and heatmap plots display a strong signal before and after the TSS in all three tissues (Supplementary Figure 5.7). This signal is similar to the distribution of the H3K4me3 peak signal before and after TSS, which is not surprising as the ChromHMM model assigns promoter states based on the presence of H3K4me3 signal. It is worth noting that the CAGE data used in this study was generated from the reference genome animal, a Rambouillet, which is different from the crossbred animals used in this study and may explain some of the signal noise.

Similarities and differences of chromatin states between tissues

Similarities and differences of promoters, enhancers, and methylated regions within and between tissues were examined and percentages of overlap are displayed in Figure 8. Active promoters were 64.76% similar between liver and spleen, 25.39% between liver and cerebellum, and 35.69% between spleen and cerebellum. Liver had 81.09% and 51.10% of active enhancers in common with spleen and cerebellum, respectively. Spleen and cerebellum had 53.85% similarity of active enhancers. Poised enhancers were shared 51.90% between liver and spleen, 52.72% between liver and cerebellum, and 38.27% between spleen and cerebellum. The percent of repressed enhancers that overlapped between liver and spleen was 56.05%. Liver and cerebellum repressed enhancers overlapped 67.90%, and spleen and cerebellum repressed enhancers overlapped 41.66%. Hypermethylated genomic locations overlapped 4.42% and hypomethylated regions overlapped 56.05% between liver and spleen. Liver and cerebellum displayed more similar hypermethylated and hypomethylated regions, 75.42% and 72.89% respectively, than spleen and cerebellum, 19.44% and 32.51% respectively.

CTCF binding motifs

The insulator CTCF is often present at the boundaries of topologically associated domains (TADs), compartments of chromatin interactions, across the genome (Beagan & Phillips-Cremins, 2020). The location of significant (*P*<0.00001) CTCF binding motifs both known from previous research and *de novo* were identified across the genome in liver, spleen, and cerebellum (Heinz et al., 2010). Of these, thirteen were present in at least three animals (Table 2). Three motifs, MYB3R4, MYB3R1, and Pdx1, were significantly enriched in liver, spleen, and cerebellum tissues. Liver and spleen exhibited the most significantly enriched CTCF motifs in common (TAGL, Six2, RRTF1, Sox6, SVP, TGA2). One motif, ZBTB19, was enriched in spleen and cerebellum. Cerebellum had three enriched motifs (Elk4, Pho2, BZR1) not present in liver or spleen. In addition, *de novo* motifs were identified in all three tissues. The top three most significant *de novo* motifs per sample in liver, spleen, and cerebellum are reported in Table 3, Table 4, and Table 5, respectively. Of the total number of *de novo* motifs, sixteen, thirteen, and twenty-one were identified as unique to liver, spleen, and cerebellum, respectively. Sixteen *de novo* motifs were identified in both liver and spleen, while cerebellum had only three *de novo* motifs in common with the other tissues.

Discussion

The goal of this study was to characterize regulatory elements in ovine liver, spleen, and cerebellum using ChIP-seq and WGBS. The three selected tissues, liver, spleen, and cerebellum each represent a different developmental origin and are important to metabolism, immune response, and motor control, respectively. We have demonstrated the successful application of the micrococcal nuclease ChIP protocol across these tissues and the bioinformatic pipeline for the analysis of ChIP-seq in sheep. Furthermore, this study has incorporated the value of coupled histone modification and DNA methylation data towards a better understanding of regulatory regions in the sheep genome.

Micrococcal nuclease was used to shear the chromatin because it provided a consistent and reproducible shearing across samples and tissue types. A limitation of the micrococcal nuclease may be increased likelihood of the appearance of duplicated reads due to similarity of cut sites in the chromatin; however, several studies have not found substantial bias when duplicates were removed (Allan et al., 2012; David et al., 2017; Gutiérrez et al., 2017; Chereji et al., 2019). Further, shearing with micrococcal nuclease to approximately 1-2 nucleosome lengths may contribute to slightly different characteristics, including width, of peaks called from these experiments.

Sequence read pileups were examined in IGV near genes known to be active and inactive in humans and expected to be conserved across species. This provided a means of examining genes with of known promoters and expression patterns as positive and negative controls for both ChIP-seq experiments and WGBS and provided insight into potential similarity of regulatory elements across species. Several genes known to be active across different mammalian species in liver, spleen, and cerebellum showed sequence pileup of active histone marks which likely indicated the presence of active regulatory elements. Inversely, genes known to be active during meiotic processes and quiescent during adult stages in several mammalian species showed no sequence pileup of histone marks and presence of DNA methylation, which likely indicates inactivity of regulatory elements.

Consistency of regulatory element identification by ChIP-seq and DNA methylation for each tissue between the four individual animals was evaluated by calculating Spearman and Pearson correlations, respectively. Correlations between samples for both ChIP-seq and DNA methylation were within the ranges previously reported with sequence data (Peng et al., 2010; Siska and Kechris, 2017). Further, correlations between ChIP-seq biological replicates have been reported as low as 0.3-0.4, with technical replicates reported as high as 0.9 (Friedman and Alm, 2012; Siska and Kechris, 2017). The results for these sheep tissues therefore achieve equivalent or improved results compared with previously reported pipelines for regulatory element identification and characterization and demonstrate a tissue-specific moderate variation across biological replicates. Spleen displayed the highest variation between biological replicates, with correlations between 0.44 and 0.84 among histone marks, although DNA methylation was consistent across replicates including spleen. Given that splenic tissue is an acutely responsive immunological tissue, perhaps it is not surprising that we observed greater variation in the biological replicates.

The CG methylation signal for all four samples clustered distinctly by tissue in a principle component analysis, indicating the clear differences in DNA methylation between tissues (Figure 3). This finding is supported by others that have shown the greatest differences in methylation occur between tissue types rather than between individuals (Pai, 2011; Zhang, 2013) and consistent with the requirement for a particular set of genes to be active and therefore de-methylated depending on tissue function. Cerebellum samples demonstrated a higher level of both CG and non-CG methylation compared to liver and spleen. Brain tissues are known to differ from other tissues in methylation patterns in other species, and further cerebellum has been shown to be different than other brain tissues (Gibbs et al., 2010; Cantrell et al., 2019)

The enrichment of individual histone marks was examined by identifying peaks in each sample. The number of peaks identified in these sheep liver, spleen, and cerebellum samples were consistent with other studies in sheep adipose, cattle liver, cattle muscle, cattle rumen epithelium, human liver, and mouse liver (Supplementary Table 3) (Naval-Sanchez et al., 2018; Villar et al., 2015; Zhao et al., 2015; Fang et al., 2019). Many chromosomes had differences in peak numbers normalized by chromosome length between tissues, indicating potential tissue specificity of some peaks. Narrow marks H3K3me3, H3K27ac, and CTCF had a shorter average width than broad marks H3K4me1 and H3K27me3, which may be influenced by the program and statistical model used to call peaks as well as the shearing method (Zhang et al., 2088; Zang et al., 2009). Because micrococcal nuclease as used for shearing, the length of the narrow peaks more closely resembles the size of a single nucleosome.

H3K4me3 peaks were enriched annotated TSS inn all three tissues. The peaks and heatmap signature signals are similar to several other ChIP-seq experiments in human PBMCs and CD14+ cells, as well as mouse liver (Schones et al., 2008; Quinodoz et al., 2014; Uchiyama et al., 2018). Peaks from all histone modifications and CTCF were also annotated with regions defined in the *Oar_rambouillet_v1.0* genome. In liver, spleen, and cerebellum, the most TSS were identified near (within 2 kilobases of distance on either side) to H3K4me3 peaks, which is not surprising. Many H3K27ac and H3K4me1 peaks, which indicate the presence active or poised enhancers, were located in intronic regions. Repressed enhancers marked by H3K27me3 were located mostly in intergenic regions, along with CTCF, which may be indicative of insulated TAD boundaries not in close proximity of genes. Further work with additional animals in combination with RNA expression and TSS analyses are needed to examine regulatory element activity outside of previously annotated regions of the sheep genome.

The genomic segments identified by histone mark peaks were evaluated for overlap between marks and CTCF binding. This broader view of the regulatory landscape lends a better understanding of gene regulation at each location than individual marks (Park 2018). Active promoters have been shown to exhibit greater enrichment of H3K4me3 than other histone marks in addition to the often present H3K27ac (Wang et al., 2008; Creyghton et al., 2010; Carelli et al., 2018). However, if lysine 4 is monomethylated (H3K4me1), this indicates the presence of a poised enhancer, in which enrichment of lysine 27 can be acetylated or trimethylated depending on the state and activity of the enhancer (Heintzman et al., 2007; Wang et al., 2008; Creyghton et al., 2010; Carelli et al., 2018). Low H3K4me3 coincident with high H3K27ac signal has been reported to be common at enhancers near genes undergoing highly active transcription (Core et al., 2011; Carelli et al., 2018). Repressed enhancers are generally characterized by H3K27me3 signal (Carelli et al., 2018). However, H3K27me3 has also been shown to be enriched near the promoter or gene body in genes being expressed at a relatively low rate (Young et al., 2011; Flensburg et al., 2014). The chromatin states characterized in this study are similar to what others have previously described in cattle (Fang et al., 2019). Further, the weak poised enhancer category detected in spleen and weak repressed enhancer category detected in cerebellum demonstrate that different tissues may have varying chromatin states, which supports the importance of characterizing chromatin states across tissues within a species.

Hypermethylated and hypomethylated regions of the sheep genome were defined across liver, spleen, and cerebellum tissues. The number of hypermethylated and hypomethylated regions per Mb in each of the nine chromatin states were quantified. The data presented in this study demonstrates an enrichment of hypermethylated regions in chromatin states with prominent H3K4me1 (primarily poised enhancers), and hypomethylated regions in active enhancers and promoters enriched with H3K27ac and H3K4me3. These results agree with previous research in humans and mice which indicate that active enhancer activity is inversely correlated with DNA methylation (Aran and Hellman, 2013; Barwick et al., 2016; Bell and Vertino, 2017). Interestingly, the presence of H3K4me1 was found to be positively correlated with DNA methylation, specifically intermediate methylation (25-75%), in mice (Zhang et al., 2009; Teng and Tan, 2012; Sharifi-Zarchi et al., 2017). Further, enhancers enriched with H3K27ac and promoters enriched with H3K4me3 had less DNA methylation than other regions (Sharifi-Zarchi et al., 2017).

Approximately 20% of the sheep genome was assigned to a chromatin state category including promoters, active, poised, and repressed enhancers, and CTCF in liver, spleen, and cerebellum. In cattle, a previous study similarly assigned approximately 30% of the genome to a either chromatin state or areas with open chromatin in rumen epithelium (Fang et al., 2019). The locations of many regulatory elements were similar between liver and spleen in this study; however, a greater difference was observed in active enhancers and promoters between cerebellum compared with liver and spleen. Since distinct differences in gene expression and regulation have been observed between cerebellum and other tissues in sheep, this difference is not surprising (Jiang et al., 2014).

The CCCTC-binding factor (CTCF) along with cohesins were shown to be present at the boundaries of TADs in humans and mice (Dixon et al., 2012; Phillips-Cremins et al., 2013; Rao et al., 2014; Vietri Rudan et al., 2015; Szabo et al., 2019). Depending on cell type, 75-95% of TAD boundaries defined by Hi-C chromatin capture have shown CTCF signal in mice (Bonev et al., 2017; Szabo et al., 2019). The chromatin states in this study that display primarily CTCF could be representative of these domain boundaries, however Hi-C data is required to confirm which will be possible for the data produced in the FAANG study of the reference ewe, where Hi-C data is also available. In addition to helping define TAD boundaries, CTCF has also been identified near enhancers and promoters within TADs in humans and mice, which then form smaller loop domains with cohesins and the protein YY1 (Phillips-Cremins et al., 2013; Weintraub et al., 2017; Szabo et al., 2019). The chromatin state analysis may be detecting some of these within-TAD loop interactions, with overlap between CTCF and H3K27me3 as well as H3K4me1 signal shown in the chromatin state heatmaps in liver and cerebellum. Signal from CTCF, H3K27me3, and H3K4me1 marks within one chromatin state was also observed in another study in cattle rumen epithelial tissue and Madin-Darby bovine kidney epithelial cells (Fang et al., 2019).

Motif analysis of CTCF resulted in both known and *de novo* motifs identified in more than one tissue. A large number CTCF binding motifs are similar in sequence across mammalian species including cattle (Filippova et al., 1996; Schmidt et al., 2012; 25; Wang et al., 2018). Wang and associates identified putative CTCF binding motifs in the bovine genome with 82 CTCF motif profiles with similar sequence in human, mouse, dog, and macaque (Schmidt et al., 2012; Wang et al., 2018). In this study, significant motifs identified in ovine liver, spleen, and cerebellum were also identified in human, mouse, fly (*Drosophila melanogaster*), and yeast (*Saccharomyces cerevisiae*) within the HOMER motif database (Heinz et al., 2010).

This experiment examines regulatory elements in multiple sheep tissues and individuals with ChIP-seq and WGBS methylation assays. These data provide putative categories of biological functions for regulatory DNA and will facilitate identification of epigenetic variation that control phenotypic traits in sheep. Epigenetic annotation is especially important for revealing the biology behind interesting complex traits since genetic variation does not always reveal the entire story. Epigenetic variation may play a larger role in traits uniquely expressed in a specific tissue or recently evolved rare traits. Identification of causal regulatory variants will allow more rapid genetic improvement for health and production traits in the meat, milk, and wool industries across sheep populations. Causal variants have the highest utility across breeds and allow more efficient assimilation of genetic markers into marker-assisted selection and genomic selection algorithms. The protocols and analysis pipeline optimized here for validation and the eventual annotation of DNA regulatory elements are valuable resources for The Ovine FAANG Project consortium and the International Sheep Genomics Consortium.

Acknowledgements

The authors would like to thank Codie Durfee and Kaneesha Hemmerling for assistance with tissue collection, Jacky Carnahan, Kristen Kuhn, and Kelsey McClure for assistance with sequencing, and Dr. Mingrui Duan for assistance with the bioinformatics analysis pipeline.

Members of the Ovine FAANG Project Consortium (listed by institution)

Brenda Murdoch (University of Idaho)
Kimberly Davenport (University of Idaho)
Stephen White (USDA, ARS, ADRU, Washington State University)
Michelle Mousel (USDA, ARS, ADRU, Washington State University)
Alisha Massa (Washington State University)

Kim Worley (Baylor College of Medicine)
Alan Archibald (The Roslin Institute, University of Edinburgh)
Emily Clark (The Roslin Institute, University of Edinburgh)
Brian Dalrymple (University of Western Australia)
James Kijas (CSIRO)
Shannon Clarke (AgResearch)
Rudiger Brauning (AgReseach)
Timothy Smith (USDA, ARS, MARC)
Tracey Hadfield (Utah State University)
Noelle Cockett (Utah State University)

Contribution to the field

Functional annotation of regulatory elements in sheep is vital for understanding complex phenotypic traits related to health as well as food and fiber production in this globally important species. Greater than 90% of variation underlying genetic effects on phenotypic traits are estimated to lie outside of transcribed regions. Therefore, it is important to define regions that regulate gene transcription across the genome to gain a greater understanding of the mechanisms that influence economically important traits. This study defines four histone modifications H3K4me3 (promoters), H3K27ac (active enhancers), H3K4me1 (poised enhancers), and H3K27me3 (repressed enhancers) and CTCF as well as global DNA methylation across three tissues that play key roles in health and production traits. This study provides novel regulatory element annotation from histone modifications, CTCF, and DNA methylation in liver, spleen, and cerebellum tissues in sheep. This will set the precedent for the characterization of regulatory elements in ovine tissues.

Funding

This project was supported by Agriculture and Food Research Initiative grant no. 2017-67016-26301, Hatch grant no. IDA01566 from the USDA National Institute of Food and

Agriculture, USDA-NIFA-2016-67016-24766 for DNA methylation, USDA-ARS Project 3040-31000-100-00-D for sequencing, and USDA-ARS Project 2090-32000-36-00-D for the care and management of sheep.

Ethics statement

Sheep were housed and cared for under Animal Subject Approval Form (ASAF) 04618 titled "Maintenance of MCF and OPPV free sheep flocks" approved by the Washington State University Institutional Animal Care and Use Committee (IACUC) and euthanized for tissue collection under ASAF 6003.

Data availability

Ewe F1 tissues were submitted to EBI BioSamples as SAMEA7423843, ewe F2 as SAMEA7423844, wether M1 as SAMEA7423846, and wether M2 as SAMEA7423847. The data is hosted by the European Nucleotide Archive under BioProject PRJEB41457 (https://www.ncbi.nlm.nih.gov/bioproject/?term=PRJEB41457) and is publicly available.

Supplementary data

Supplementary data is provided in Supplementary Table 1, Supplementary Table 2, Supplementary Table 3, Supplementary Figure 1, Supplementary Figure 2, Supplementary Figure 3, Supplementary Figure 4, Supplementary Figure 5, Supplementary Figure 6 and Supplementary Figure 7.

Author contributions

MRM, MKH, SNW, SDM, TPLS, and BMM designed the study. KMD, ATM, MRM, MKH, and BMM collected samples. BMM, KMD, MRM, SNW, ATM, SDM and SB facilitated the ChIP-Seq and methylation experiments, data analyses, and drafted the manuscript. TPLS facilitated ChIP-seq library preparation and sequencing. KMD, ATM, SB, SDM, MRM, MKH, SNW, TPLS, and BMM discussed and interpreted results. SDM, MRM, SNW, NC, TPLS, and BMM acquired funding. All authors contributed to the article and approved the final version.

Conflict of interest statement

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

References

1. Afgan, E., Baker, D., Batut, B., van den Beek, M., Bouvier, D., Čech, M., et al. (2018). The Galaxy platform for accessible, reproducible and collaborative biomedical analyses: 2018 update. Nuc. Acids Research. 46, 537–544. doi:10.1093/nar/gky379

2. Al-Mamun, H.A., Clark, S.A., Kwan, P., and Gondro, C. (2015). Genome-wide linkage disequilibrium and genetic diversity in five populations of Australian domestic sheep. Genet. Sel. Evol. 47, 90. doi:10.1186/s12711-015-0169-6

3. Albert, F.W., and Kruglyak, L. (2015). The role of regulatory variation in complex traits and disease. Nat. Rev. Genet. 16(4), 197–212. doi:10.1038/nrg3891

4. Andersson, L., Archibald, A.L., Bottema, C.D., Brauning, R., Burgess, S.C., Burt, D.W., et al. (2015). Coordinated international action to accelerate genome-to-phenome with FAANG, the Functional Annotation of Animal Genomes project. Gen. Biol. 16(1), 57. doi:10.1186/s13059-015-0622-4

5. Aran, D., and Hellman, A. (2013). DNA methylation of transcriptional enhancers and cancer predisposition. Cell. 154(1), 11-13. doi:10.1016/j.cell.2013.06.018

6. Barski, A., Cuddapah, S., Cui, K., Roh, T.Y., Schones, D.E., Wang, Z., et al. (2007). High-resolution profiling of histone methylations in the human genome. Cell. 129(4), 823–837. doi:10.1016/j.cell.2007.05.009

7. Bell, J.S.K., and Vertino, P.M. (2017). Orphan CpG islands define a novel class of highly active enhancers. Epigenetics. 12:6, 449-464. doi:10.1080/15592294.2017.1297910

8. Bernardi, M., Maggioli, C., and Zaccherini, G. (2012). Human albumin in the management of complications of liver cirrhosis. Crit. Care. 16(2), 211. doi: 10.1186/cc11218

9. Bonev B., Cohen N.M., Szabo Q., Fritsch L., Papadopoulos G.L., Lubling Y., et al. (2017). Multiscale 3D genome rewiring during mouse neural development. Cell. 171, 557–572. doi:10.1016/j.cell.2017.09.043

10. Cantrell, B., Lachance, H., Murdoch, B., Sjoquist, J., Funston, R., Weaber, R., and McKay, S. (2019) Global DNA methylation in the limbic system of cattle. Epigenomes. 3, 8. doi:10.3390/epigenomes3020008

11. Carelli, F.N., Liechti, A., Halbert, J., Warnefors, M., and Kaessmann, H. (2018). Repurposing of promoters and enhancers during mammalian evolution. Nat. Commun. 9(1), 4066. doi: 10.1038/s41467-018-06544-z

12. Cedar, H., and Bergman, Y. (2009). Linking DNA methylation and histone modification: patterns and paradigms. Nat. Rev. Genet. 10(5), 295–304. doi:10.1038/nrg2540

13. Chereji, R.V., Bryson, T.D., and Henikoff, S. (2019). Quantitative MNase-seq accurately maps nucleosome occupancy levels. Genome Biol. 20, 198. doi:10.1186/s13059-019-1815-z

14. Clark, E.L., Bush, S.J., McCulloch, M., Farquhar, I.L., Young, R., Lefevre, L., et al. (2017). A high resolution atlas of gene expression in the domestic sheep (Ovis aries). PLoS Genet. 13(9), e1006997. doi:10.1371/journal.pgen.1006997

15. Core, L.J., Martins, A.L., Danko, C.G., Waters, C.T., Siepel, A., and Lis, J.T. (2014). Analysis of nascent RNA identifies a unified architecture of initiation regions at

mammalian promoters and enhancers. Nat. Genet. 46(12), 1311-1320. doi: 10.1038/ng.3142

16. Creyghton, M.P., Cheng, A.W., Welstead, G.G., Kooistra, T., Carey, B.W., Steine, E.J., et al. (2010). Histone H3K27ac separates active from poised enhancers and predicts developmental state. Proc. Natl. Acad. Sci. USA. 107(50), 21931-21936. doi: 10.1073/pnas.1016071107

17. David, S.A., Piégu, B., Hennequet-Antier, C., Pannetier, M., Aguirre-Lavin, T., Crochet, S., et al. (2017). An Assessment of Fixed and Native Chromatin Preparation Methods to Study Histone Post-Translational Modifications at a Whole Genome Scale in Skeletal Muscle Tissue. Biol. Proc. Online. 19, 10. doi:10.1186/s12575-017-0059-0

18. Divya, T., Lalitha, S., Parvathy, S., Subashini, C., Sanalkumar, R., Bindu Dhanesh S., et al. (2016). Regulation of Tlx3 by Pax6 is required for the restricted expression of Chrnα3 in Cerebellar Granule Neuron progenitors during development. Sci. Rep. 6, 30337. doi:10.1038/srep30337

19. Dixon J.R., Selvaraj S., Yue F., Kim A., Li Y., Shen Y., et al. (2012). Topological domains in mammalian genomes identified by analysis of chromatin interactions. Nature. 485, 376–380. doi:10.1038/nature11082

20. ENCODE Project Consortium (2012). An integrated encyclopedia of DNA elements in the human genome. Nature. 489(7414), 57–74. doi:10.1038/nature11247

21. Ernst, J., and Kellis, M. (2010). Discovery and characterization of chromatin states for systematic annotation of the human genome. Nat Biotechnol. 28(8), 817-825. doi: 10.1038/nbt.1662

22. Ernst, J., and Kellis, M. (2012). ChromHMM: automating chromatin-state discovery and characterization. Nat Methods. 9(3), 215-216. doi: 10.1038/nmeth.1906

23. Ernst, J., and Kellis, M. (2017). Chromatin-state discovery and genome annotation with ChromHMM. Nat Protoc. 12(12), 2478–2492. doi:10.1038/nprot.2017.124

24. Fagerberg. L., Hallström, B.M., Oksvold, P., Kampf, C., Djureinovic, D., Odeberg, J., et al. (2014). Analysis of the human tissue-specific expression by genome-wide integration of transcriptomics and antibody-based proteomics. Mol. Cell. Proteomics. 13(2), 397-406. doi: 10.1074/mcp.M113.035600

25. Fang, L., Liu, S., Liu, M., Kang, X., Lin, S., Li, B., et al. (2019). Functional annotation of the cattle genome through systematic discovery and characterization of chromatin states and butyrate-induced variations. BMC Biol. 17(1), 68. doi:10.1186/s12915-019-0687-8

26. Feng, J., Liu, T., Qin, B., Zhang, Y., and Liu, X.S. (2012). Identifying ChIP-seq enrichment using MACS. Nat Protoc. 7(9), 1728-1740. doi: 10.1038/nprot.2012.101

27. Filippova, G.N., Fagerlie, S., Klenova, E.M., Myers, C., Dehner, Y., Goodwin, G., et al. (1996). An exceptionally conserved transcriptional repressor, CTCF, employs different combinations of zinc fingers to bind diverged promoter sequences of avian and mammalian c-myc oncogenes. Mol. Cell. Biol. 16(6), 2802–2813. doi: 10.1128/MCB.16.6.2802

28. Flensburg, C., Kinkel, S.A., Keniry, A., Blewitt, M.E., and Oshlack, A. (2014). A comparison of control samples for ChIP-seq of histone modifications. Front. Genet. 5, 329. doi:10.3389/fgene.2014.00329

29. Frain, M., Hardon, E., Ciliberto, G., and Sala-Trepat, J.M. (1990). Binding of a liver-specific factor to the human albumin gene promoter and enhancer. Mol. Cell. Biol. 10(3), 991-999. doi: 10.1128/mcb.10.3.991

30. Friedman, J., and Alm, E.J. (2012). Inferring correlation networks from genomic survey data. PLoS Comput. Biol. 8(9), e1002687. doi: 10.1371/journal.pcbi.1002687

31. Ghirlando, R., and Felsenfeld, G. (2016). CTCF: making the right connections. Genes Dev. 30(8), 881–891. doi:10.1101/gad.277863.116

32. Giuffra, E., Tuggle, C. K., and FAANG Consortium (2019). Functional Annotation of Animal Genomes (FAANG): Current Achievements and Roadmap. Ann. Rev. Anim. Biosci. 7, 65–88. doi:10.1146/annurev-animal-020518-114913

33. Gorkin, D.U., Barozzi, I., Zhang, Y., Lee, A.Y., Zhao, Y., Wildberg A., et al. (2017). Systematic mapping of chromatin state landscapes during mouse development. bioRxiv. 166652. doi: doi:10.1101/166652

34. Gorkin, D.U., Barozzi, I., Zhao, Y., Zhang, Y., Huang, H., Lee, A.Y., et al. (2020). An atlas of dynamic chromatin landscapes in mouse fetal development. Nature. 583, 744-751. doi:10.1038/s41586-020-2093-3

35. Guo, W., Fiziev, P., Yan, W., Cokus, S., Sun, X., Zhang, M.Q., et al. (2013). BS-Seeker2: a versatile aligning pipeline for bisulfite sequencing data. BMC Genomics. 14, 774. doi:10.1186/1471-2164-14-774

36. Guo, W., Zhu, P., Pellegrini, M., Zhang, M.Q., Wang, X., and Ni, Z. (2018). CGmapTools improves the precision of heterozygous SNV calls and supports allele-specific methylation detection and visualization in bisulfite-sequencing data. Bioinformatics. 34(3), 381–387. doi:10.1093/bioinformatics/btx595

37. Gutiérrez, G., Millán-Zambrano, G., Medina, D.A., Jordán-Pla, A., Pérez-Ortín, J.E., Peñate, X., and Chávez, S. (2017). Subtracting the sequence bias from partially digested MNase-seq data reveals a general contribution of TFIIS to nucleosome positioning. Epigenetics Chromatin. 10(1), 58. doi: 10.1186/s13072-017-0165-x

38. Ha, T., Swanson, D., Larouche, M., Glenn, R., Weeden, D., Zhang, P., et al. (2015). CbGRiTs: Cerebellar gene regulation in time and space. Dev. Biol. 397(1), 18-30. doi:10.1016/j.ydbio.2014.09.032

39. Hayashi, Y., Chan, J., Nakabayashi, H., Hashimoto, T., and Tamaoki, T. (1992). Identification and characterization of two enhancers of the human albumin gene. J Biol. Chem. 267(21), 14580-14585.

40. Hedges, J.F., Kimmel, E., Snyder, D.T., Jerome, M., and Jutila, M.A. (2013). Solute carrier 11A1 is expressed by innate lymphocytes and augments their activation. J. Immunol. 190(8), 4263-4273. doi: 10.4049/jimmunol.1200732

41. Hegde, N.G. (2019) Livestock development for sustainable livelihood of small farmers. Asian J Res Anim Vet Sci. 3(2), 1-17.

42. Heintzman, N.D., Stuart, R.K., Hon, G., Fu, Y., Ching, C.W., Hawkins, R.D., et al. (2007). Distinct and predictive chromatin signatures of transcriptional promoters and enhancers in the human genome. Nat. Genet. 39(3), 311-318. doi: 10.1038/ng1966

43. Heinz, S., Benner, C., Spann, N., Bertolino, E., Lin, Y.C., Laslo, P., et al. (2010). Simple combinations of lineage-determining transcription factors prime *cis*-regulatory elements required for macrophage and B cell identities. Mol. Cell. 38: 576–589. doi:10.1016/j.molcel.2010.05.004

44. Jenuwein, T., and Allis, C.D. (2001). Translating the histone code. Science. 293(5532), 1074–1080. doi:10.1126/science.1063127

45. Jiang, Y., Xie, M., Chen, W., Talbot, R., Maddox, J.F., Faraut, T., et al. (2014).The sheep genome illuminates biology of the rumen and lipid metabolism. Science. 344(6188), 1168-1173. doi: 10.1126/science.1252806

46. Johnson, D.S., Mortazavi, A., Myers, R.M., and Wold, B. (2007). Genome-wide mapping of in vivo protein-DNA interactions. Science. 316(5830), 1497–1502. doi:10.1126/science.1141319

47. Kingsley, N.B., Kern, C., Creppe, C., Hales, E.N., Zhou, H., Kalbfleisch, T.S., et al. (2019). Functionally Annotating Regulatory Elements in the Equine Genome Using Histone Mark ChIP-Seq. Genes. 11(1), 3. doi:10.3390/genes11010003

48. Landt, S.G., Marinov, G.K., Kundaje, A., Kheradpour, P., Pauli, F., Batzoglou, S., et al. (2012). ChIP-seq guidelines and practices of the ENCODE and modENCODE consortia. Genome Res. 22(9), 1813-1831. doi: 10.1101/gr.136184.111

49. Langmead, B., and Salzberg, S. (2012). Fast gapped-read alignment with Bowtie 2. Nature Methods. 9, 357-359.

50. Lee, B.K., and Iyer, V.R. (2012). Genome-wide studies of CCCTC-binding factor (CTCF) and cohesin provide insight into chromatin structure and regulation. J. Biol. Chem. 287(37), 30906–30913. doi:10.1074/jbc.R111.324962

51. Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., et al. (2009). The Sequence Alignment/Map format and SAMtools. Bioinformatics. 25(16), 2078–2079. doi:10.1093/bioinformatics/btp352

52. Meadows, J.R., Chan, E.K., and Kijas, J.W. (2008). Linkage disequilibrium compared between five populations of domestic sheep. BMC Genet. 9, 61. doi:10.1186/1471-2156-9-61

53. Meissner, A., Mikkelsen, T.S., Gu, H., Wernig, M., Hanna, J., Sivachenko, A., et al. (2008). Genome-scale DNA methylation maps of pluripotent and differentiated cells. Nature. 454(7205), 766–770. doi:10.1038/nature07107

54. Micsinai, M., Parisi, F., Strino, F., Asp, P., Dynlacht, B.D., and Kluger, Y. (2012). Picking ChIP-seq peak detectors for analyzing chromatin modification experiments. Nucleic Acids Res. 40(9), e70. doi: 10.1093/nar/gks048

55. Mikkelsen, T.S., Ku, M., Jaffe, D.B., Issac, B., Lieberman, E., Giannoukos, G., et al. (2007). Genome-wide maps of chromatin state in pluripotent and lineage-committed cells. Nature. 448(7153), 553-560. doi: 10.1038/nature06008

56. Naval-Sanchez, M., Nguyen, Q., McWilliam, S., Porto-Neto, L.R., Tellam, R., Vuocolo T., et al. (2018). Sheep genome functional annotation reveals proximal regulatory elements contributed to the evolution of modern breeds. Nat. Commun. 9(1), 859. doi:10.1038/s41467-017-02809-1

57. Pai, A.A., Bell, J.T., Marioni, J.C., Pritchard, J.K., and Gilad, Y. (2011). A Genome-Wide Study of DNA Methylation Patterns and Gene Expression Levels in Multiple Human and Chimpanzee Tissues. PLOS Genet. 7(2), e1001316. doi:10.1371/journal.pgen.1001316

58. Park, H.S. (2018). A Short Report on the Markov Property of DNA Sequences on 200-bp Genomic Units of Roadmap Genomics ChromHMM Annotations: A Computational Perspective. Genomics Inform. 16(4), e27. doi: 10.5808/GI.2018.16.4.e27

59. Pauler, F.M., Sloane, M.A., Huang, R., Regha, K., Koerner, M.V., Tamir, I., et al. (2009). H3K27me3 forms BLOCs over silent genes and intergenic regions and

specifies a histone banding pattern on a mouse autosomal chromosome. Genome Res. 19(2), 221–233. doi:10.1101/gr.080861.108

60. Peng, S., Kuroda, M.I. and Park, P.J. (2010). Quantized correlation coefficient for measuring reproducibility of ChIP-chip data. BMC Bioinformatics. 11, 399. doi:10.1186/1471-2105-11-399

61. Phillips-Cremins J.E., Sauria M.E.G., Sanyal A., Gerasimova T.I., Lajoie B.R., Bell J.S.K., et al. (2013). Architectural protein subclasses shape 3D organization of genomes during lineage commitment. Cell. 153, 1281–1295. doi:10.1016/j.cell.2013.04.053

62. "Picard Toolkit." 2019. Broad Institute, GitHub Repository. http://broadinstitute.github.io/picard/; Broad Institute

63. Quinlan, A.R. (2014). BEDTools: The Swiss-Army Tool for Genome Feature Analysis. Curr. Protoc. Bioinformatics. 47, 11.12.1-34. doi: 10.1002/0471250953.bi1112s47

64. Quinodoz, M., Naef F., Gobet C., and Gustafson K. (2014). Characteristic bimodal profiles of RNA polymerase II at thousands of active mammalian promoters. Genome Biol. 15:R85. doi: 10.1186/gb-2014-15-6-r85.

65. Ramírez, F., Dündar, F., Diehl, S., Grüning, B.A., and Manke, T. (2014). deepTools: a flexible platform for exploring deep-sequencing data. Nucleic Acids Res. 42, 187-191. doi: 10.1093/nar/gku365

66. Rao S.S.P., Huntley M.H., Durand N.C., Stamenova E.K., Bochkov I.D., Robinson J.T., et al. (2014). A 3D map of the human genome at kilobase resolution reveals principles of chromatin looping. Cell. 159, 1665–1680. doi:10.1016/j.cell.2014.11.021

67. Rexroad, C., Vallet, J., Matukumalli, L.K., Reecy, J., Bickhart, D., Blackburn, H., et al. (2019). Genome to Phenome: Improving Animal Health, Production, and Well-Being - A New USDA Blueprint for Animal Genome Research 2018-2027. Front Genet. 10, 327. doi:10.3389/fgene.2019.00327

68. Robertson, G., Hirst, M., Bainbridge, M., Bilenky, M., Zhao, Y., Zeng, T., et al. (2007). Genome-wide profiles of STAT1 DNA association using chromatin immunoprecipitation and massively parallel sequencing. Nat. Methods. 4, 651–657. doi:10.1038/nmeth1068

69. Robinson, J.T., Thorvaldsdóttir, H., Winckler, W., Guttman, M., Lander, E.S., Getz, G., and Mesirov, J.P. (2011). Integrative genomics viewer. Nat. Biotechnol. 29(1), 24-26. doi: 10.1038/nbt.1754

70. Salavati, M., Caulton, A., Clark, R., Gazova, I., Smith, T.P.L., Worley, K.C., Cockett, N.E., Archibald, A., Clarke, S.M., Murdoch, B.M., Clark, E.L. (2020). Global analysis of transcription start sites in the new ovine reference genome (*Oar rambouillet v1.0*). Front. Genet. 11, 580580. doi: 10.3389/fgene.2020.580580.

71. Schmidt, D., Schwalie Petra, C., Wilson Michael, D., Ballester, B., Gonçalves, Â., Kutter, C., et al. (2012). Waves of retrotransposon expansion remodel genome organization and CTCF binding in multiple mammalian lineages. Cell. 148(1), 335–348. doi: 10.1016/j.cell.2011.11.058

72. Schones, D.E., Cui, K., Cuddapah, S., Roh, T.Y., Barski, A., Wang, Z., Wei, G., and Zhao, K. (2008). Dynamic regulation of nucleosome positioning in the human genome. Cell. 132(5), 887-98. doi: 10.1016/j.cell.2008.02.022.

73. Siska, C., and Kechris, K. (2017). Differential correlation for sequencing data. BMC Res. Notes. 10(1), 54. doi: 10.1186/s13104-016-2331-9

74. Song, Q., Decato, B., Hong, E.E., Zhou, M., Fang, F., Qu, J., et al. (2013). A reference methylome database and analysis pipeline to facilitate integrative and comparative epigenomics. PloS One. 8(12), e81148. doi:10.1371/journal.pone.0081148

75. Szabo, Q., Bantignies, F., and Cavalli, G. (2019). Principles of genome folding into topologically associating domains. Sci. Adv. 5(4), 1668. doi:10.1126/sciadv.aaw1668

76. Teng, L., and Tan, K. (2012). Finding combinatorial histone code by semi-supervised biclustering. BMC Genomics. 13, 301. doi:10.1186/1471-2164-13-301

77. Thomas, R., Thomas, S., Holloway, A.K., and Pollard, K.S. (2017). Features that define the best ChIP-seq peak calling algorithms. Brief Bioinform. 18(3), 441-450. doi: 10.1093/bib/bbw035

78. Tuggle, C.K., Giuffra, E., White, S.N., Clarke, L., Zhou, H., Ross, P.J., et al. (2016). GO-FAANG meeting: a Gathering On Functional Annotation of Animal Genomes. Anim. Genet. 47(5), 528–533. doi:10.1111/age.12466

79. Uchiyama, R., Kupkova, K., Shetty, S.J., Linford, A.S., Pray-Grant, M.G., Wagar, L.E., Davis, M.M., Haque, R., Gaultier, A., Mayo, M.W., Grant, P.A., Petri Jr., W.A., Bekiranov, S., and Auble, D.T. (2018). Histone H3 lysine 4 methylation signature associated with human undernutrition. PNAS. 115(48), E11264-E11273. doi: 10.1073/pnas.1722125115.

80. Vietri Rudan M., Barrington C., Henderson S., Ernst C., Odom D.T., Tanay A., et al. (2015). Comparative Hi-C reveals that CTCF underlies evolution of chromosomal domain architecture. Cell Rep. 10, 1297–1309. doi:10.1016/j.celrep.2015.02.004

81. Villar, D., Berthelot, C., Aldridge, S., Rayner, T.F., Lukk, M., Pignatelli, M., et al. (2015). Enhancer evolution across 20 mammalian species. Cell. 160(3), 554–566. doi:10.1016/j.cell.2015.01.006

82. Wang, M., Hancock, T.P., Chamberlain, A.J., Vander Jagt, C.J., Pryce, J.E., Cocks, B.G., et al. (2018). Putative bovine topological association domains and CTCF binding motifs can reduce the search space for causative regulatory variants of complex traits. BMC Genomics. 19(1), 395. doi: 10.1186/s12864-018-4800-0

83. Wang, Z., Zang, C., Rosenfeld, J.A., Schones, D.E., Barski, A., Cuddapah, S., et al. (2008). Combinatorial patterns of histone acetylations and methylations in the human genome. Nat Genet. 40(7), 897-903. doi: 10.1038/ng.154

84. Weber, M., Hellmann, I., Stadler, M.B., Ramos, L., Pääbo, S., Rebhan, M., and Schübeler, D. (2007). Distribution, silencing potential and evolutionary impact of promoter DNA methylation in the human genome. Nat. Genet. 39(4), 457–466. doi:10.1038/ng1990

85. Weintraub A.S., Li C.H., Zamudio A.V., Sigova A.A., Hannett N.M., Day D.S., et al. (2017). YY1 is a structural regulator of enhancer-promoter loops. Cell. 171, 1573–1588.e28. doi:10.1016/j.cell.2017.11.008

86. Xiang, R., Berg, I., MacLeod, I.M., Hayes, B.J., Prowse-Wilkins, C.P., Wang, M., et al. (2019). Quantifying the contribution of sequence variants with regulatory and evolutionary significance to 34 bovine complex traits. Proc. Nat. Acad. Sci. USA. 116(39), 19398–19408. doi:10.1073/pnas.1904159116

87. Xu, H., Beasley, M.D., Warren, W.D., van der Horst, G.T., and McKay, M.J. (2005). Absence of mouse REC8 cohesin promots synapsis of sister chromatids in meiosis. Dev. Cell. 8(6), 949-961.

88. Young, E.T., and Sinsheimer, R.L. (1965). A comparison of the initial actions of spleen deoxyribonuclease and pancreas deoxyribonuclease. J. Biol. Chem. 240, 1274-1280

89. Young, M.D., Willson, T.A., Wakefield, M.J., Trounson, E., Hilton, D.J., Blewitt, M.E., et al. (2011). ChIP-seq analysis reveals distinct H3K27me3 profiles that correlate with transcriptional activity. Nucleic Acids Res. 39, 7415–7427. doi: 10.1093/nar/gkr416

90. Zang, C., Schones, D.E., Zeng, C., Cui, K., Zhao, K., and Peng, W. (2009). A clustering approach for identification of enriched domains from histone modification ChIP-Seq data. Bioinformatics.25(15), 1952-1958. doi: 10.1093/bioinformatics/btp340

91. Zhang, B., Zhou, Y., Lin, N., Lowdon, R. F., Hong, C., Nagarajan, R. P., et al. (2013). Functional DNA methylation differences between tissues, cell types, and across individuals discovered using the M&M algorithm. Genome Res. 23(9), 1522–1540. doi:10.1101/gr.156539.113

92. Zhang, X., Bernatavichute, Y.V., Cokus, S., Pellegrini, M., and Jacobsen, S.E. (2009). Genomewide analysis of mono-, di- and trimethylation of histone H3 lysine 4 in Arabidopsis Thaliana. Genome Biol. 10(6), R62. doi:10.1186/gb-2009-10-6-r62

93. Zhang, Y., Liu, T., Meyer, C.A., Eeckhoute, J., Johnson, D.S., Bernstein, B.E., et al. (2008). Model-based analysis of ChIP-Seq (MACS). Genome Biol. 9(9), R137. doi: 10.1186/gb-2008-9-9-r137

94. Zhao, C., Carrillo, J.A., Tian, F., Zan, L., Updike S.M., Zhao, K., et al. (2015). Genome-Wide H3K4me3 Analysis in Angus Cattle with Divergent Tenderness. PLoS One. 10(6), e0115358. doi:10.1371/journal.pone.0115358

Tables

Table 5.1: Average correlations of sequencing signal between all four animals. Spearman correlations were used for ChIP-seq data and Pearson correlations were used for DNA methylation data. Parentheses indicate correlations between the replicates used in the ChromHMM chromatin state analysis.

Tissue	H3K4me3	H3K27ac	H3K4me1	H3K27me3	DNA Methylation
Liver	0.66 (0.86)	0.89 (0.95)	0.71 (0.93)	0.58 (0.74)	0.72 (0.76)
Spleen	0.54 (0.71)	0.78 (0.84)	0.47 (0.56)	0.37 (0.44)	0.70 (0.74)
Cerebellum	0.85 (0.88)	0.70 (0.91)	0.82 (0.91)	0.72 (0.83)	0.73 (0.76)

Table 5.2: Known CTCF motifs present in the top 10 most significant motifs across multiple samples.

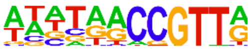
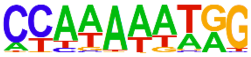
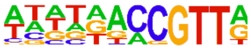
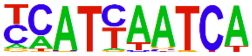
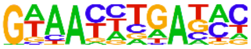
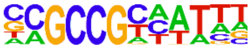
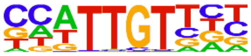
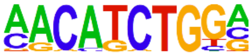
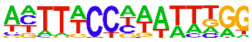
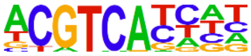
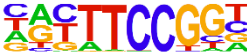
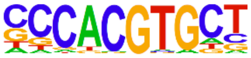
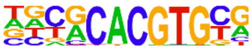
Known Motif Name	Known Motif	Tissue	Number of Samples	Mean P-value	Mean Percent of Target Sequences with Motif	Mean Percent of Background Sequences with Motif
MYB3R4 (MYB)		Liver, Spleen, Cerebellum	7	1E-2612	13.54%	1.27%
TAGL1 (MADS)		Liver, Spleen	6	1E-2167	44.33%	18.98%
MYB3R1 (MYB)		Liver, Spleen, Cerebellum	6	1E-1632	12.85%	2.42%
Pdx1 (Homeobox)		Liver, Spleen, Cerebellum	6	1E-1475	37.21%	17.82%
Six2 (Homeobox)		Liver, Spleen	5	1E-1486	30.93%	13.64%
RRTF1 (APTEREBP)		Liver, Spleen	4	1E-1655	7.16%	0.55%
Sox6 (HMG)		Liver, Spleen	4	1E-931	40.23%	23.32%
ZBTB19 (Zf)		Spleen, Cerebellum	4	1E-418	8.27%	3.07%
SVP (MADS)		Liver, Spleen	3	1E-1897	28.39%	9.82%
TGA2 (bZIP)		Liver, Spleen	3	1E-1792	16.14%	3.27%
Elk4 (ETS)		Cerebellum	3	1E-61	3.69%	2.07%
Pho2 (bHLH)		Cerebellum	3	1E-32	1.72%	0.92%
BZR1 (BZR)		Cerebellum	3	1E-29	0.68%	0.25%

Table 5.3: Top three *de novo* CTCF motifs present in each sample in liver.

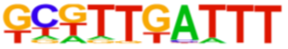
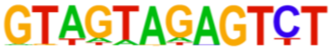
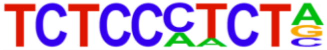
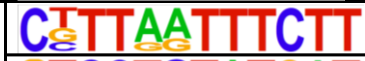
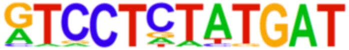
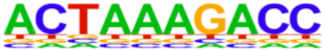
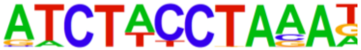
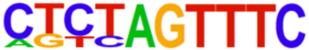
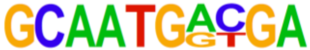
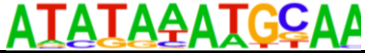
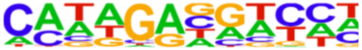
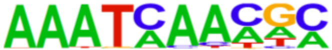
Animal	***De novo* Motif**	**P-value**	**Percent of Target Sequences with Motif**	**Percent of Background Sequences with Motif**
F1		1E-3278	31.57%	3.37%
		1E-3012	23.91%	1.74%
		1E-2873	23.31%	1.76%
F2		1E-1388	7.03%	0.62%
		1E-1349	6.55%	0.54%
		1E-1345	6.11%	0.45%
M1		1E-8604	21.40%	0.41%
		1E-7739	26.14%	1.17%
		1E-7299	21.19%	0.63%
M2		1E-10234	44.68%	4.07%
		1E-8614	34.87%	2.59%
		1E-8422	42.53%	4.89%

Table 5.4: Top three *de novo* CTCF motifs present in each sample in spleen.

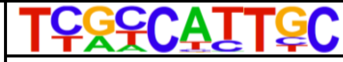
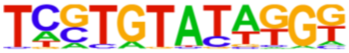
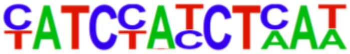
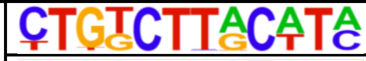
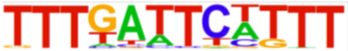
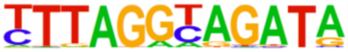
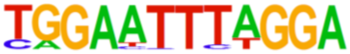
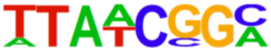
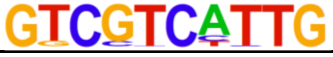
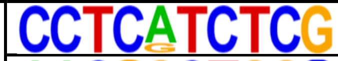
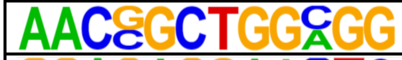
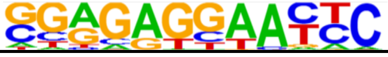
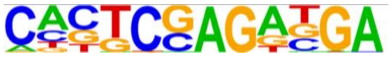
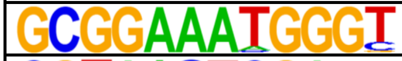
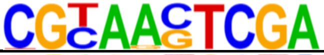
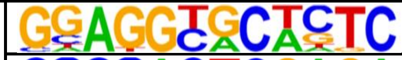
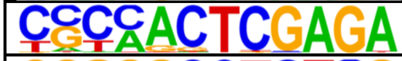
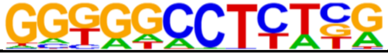
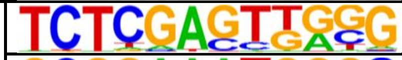
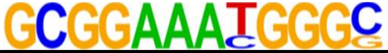
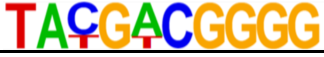
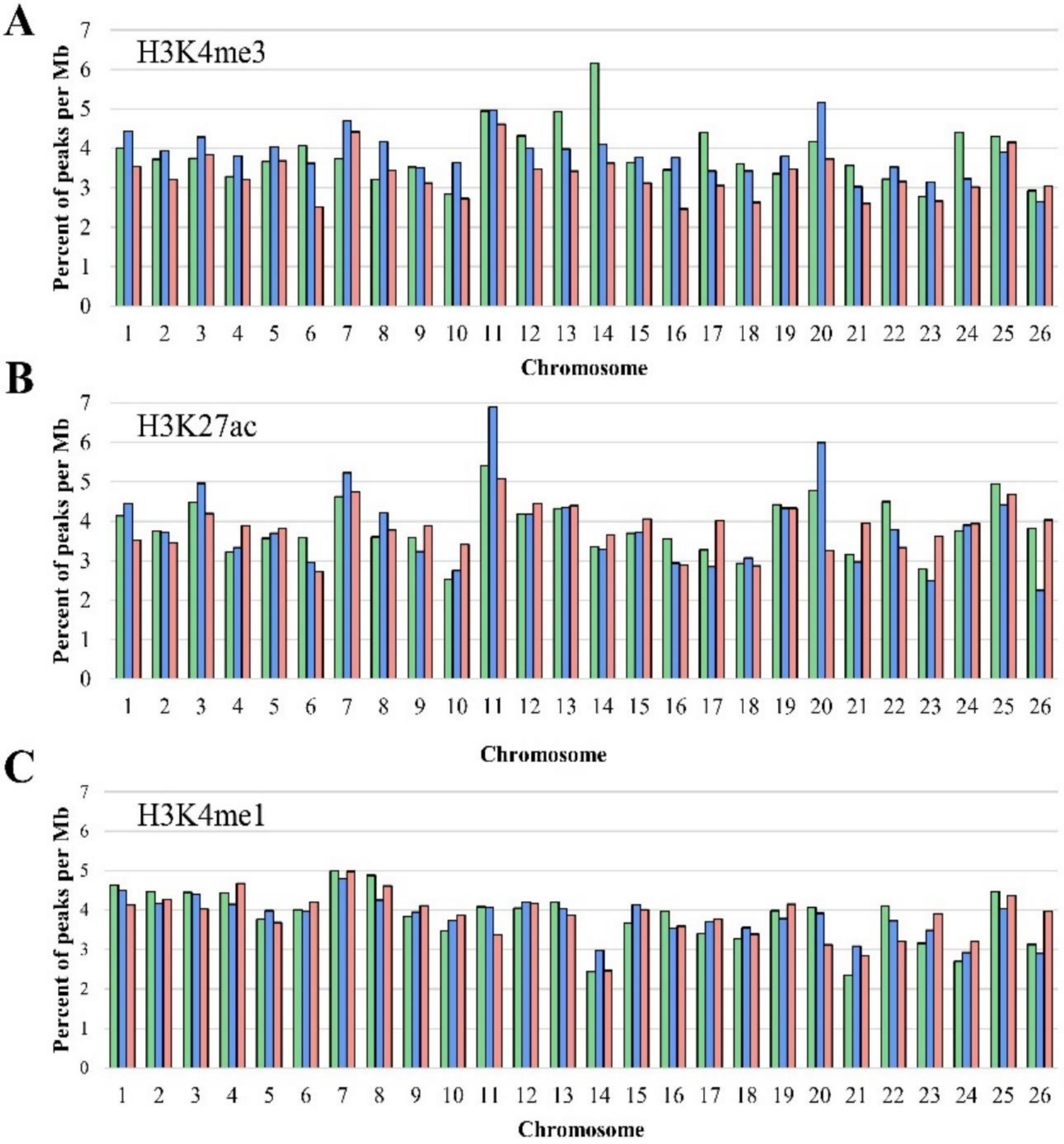
Animal	*De novo* Motif	P-value	Percent of Target Sequences with Motif	Percent of Background Sequences with Motif
F2		1E-12441	29.41%	0.72%
		1E-12221	38.23%	2.03%
		1E-12174	38.03%	2.01%
M1		1E-7022	24.88%	1.16%
		1E-6916	30.15%	2.24%
		1E-6704	25.65%	1.42%
M2		1E-5921	24.42%	0.88%
		1E-5710	24.12%	0.93%
		1E-5440	20.87%	0.61%

Table 5.5: Top three *de novo* CTCF motifs present in each sample in cerebellum.

Animal	*De novo* Motif	P-value	Percent of Target Sequences with Motif	Percent of Background Sequences with Motif
F1		1E-910	2.92%	0.01%
		1E-756	2.50%	0.01%
		1E-735	2.44%	0.01%
F2		1E-1078	3.49%	0.04%
		1E-875	1.85%	0.00%
		1E-842	2.01%	0.01%
M1		1E-946	1.42%	0.00%
		1E-800	1.24%	0.01%
		1E-697	1.25%	0.01%
M2		1E-677	1.02%	0.00%
		1E-565	0.88%	0.01%
		1E-509	0.80%	0.01%

Figures


A
H3K4me3
B
H3K27ac
C
H3K4me1
Percent of peaks per Mb
0
1
2
3
4
5
6
7
Chromosome
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26

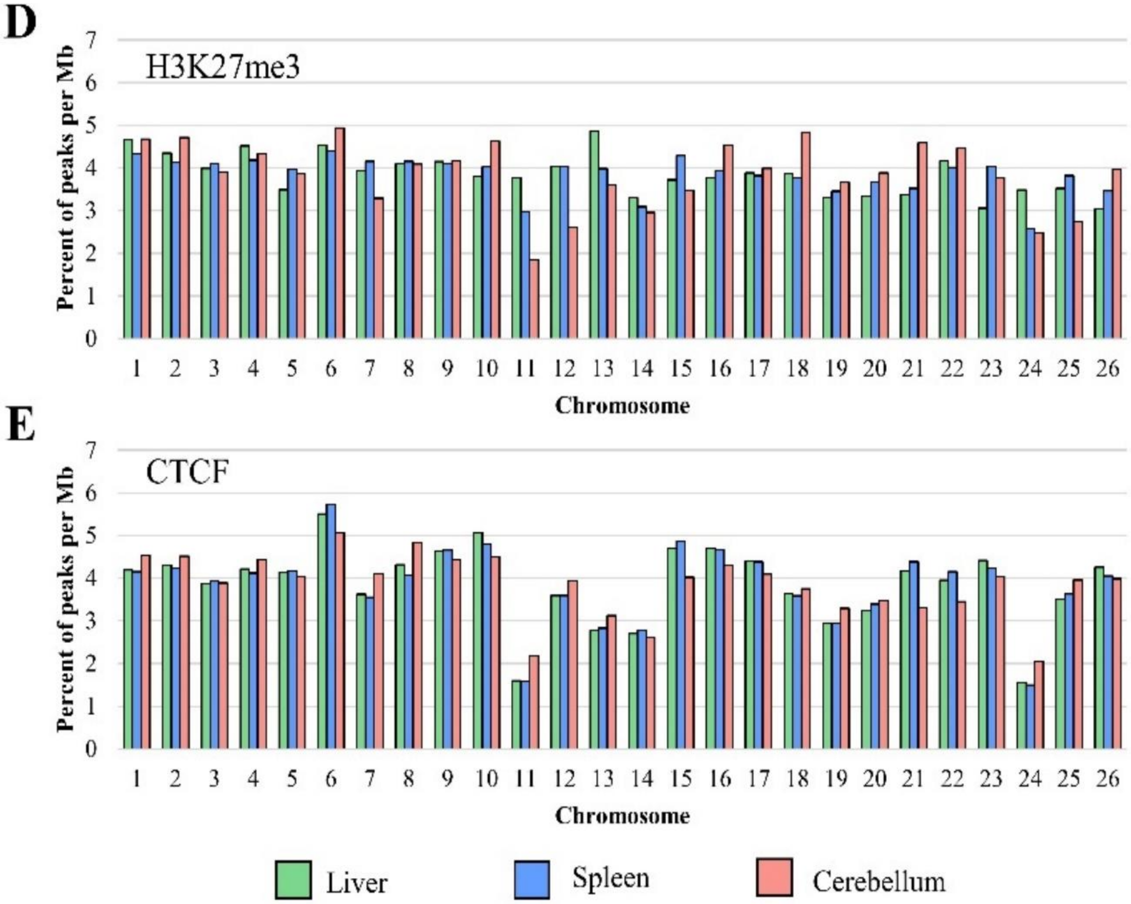


Figure 5.1: The percent of the total number of peaks normalized per Mb on each chromosome for (A) H3K4me3, (B) H3K27ac, (C) H3K4me1, (D) H3K27me3, and (E) CTCF averaged from all four animals (F1, F2, M1, M2).

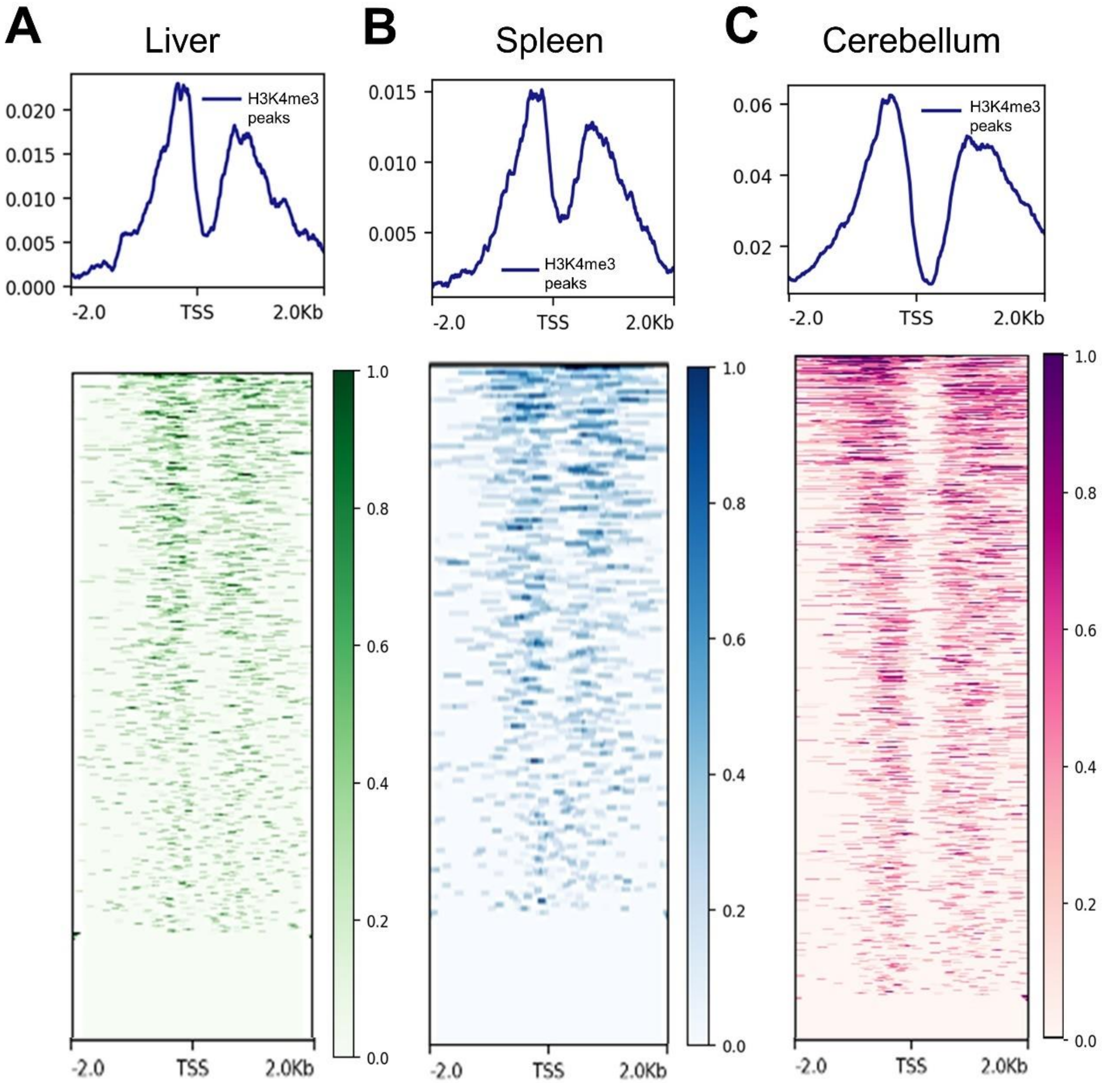
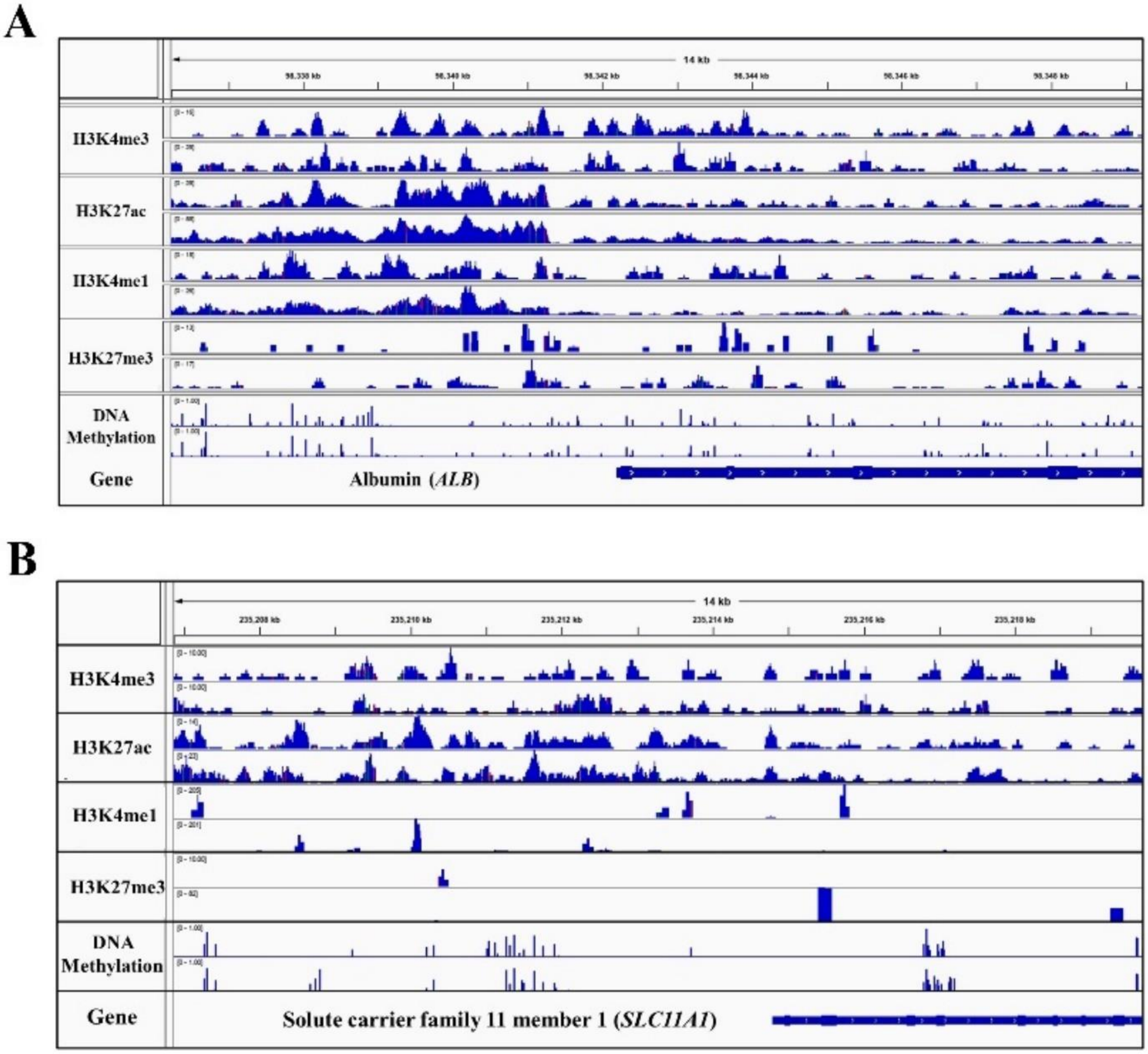


Figure 5.2: Signal of H3K4me3 ChIP-seq peaks 2 kilobases upstream and downstream of transcription start sites (TSS) identified by CAGE assays. A) Liver H3K4me3 signal (from F1, M1, and M2 consensus peaks) near TSS annotated in the reference genome, B) spleen H3K4me3 signal (from F2, M1, and M2 consensus peaks) near annotated TSS, C) cerebellum H3K4me3 signal (from F1, M1, and M2 consensus peaks) near annotated TSS.


A
14 kb
98,338 kb
98,340 kb
98,342 kb
98,344 kb
98,346 kb
98,348 kb
H3K4me3
H3K27ac
H3K4me1
H3K27me3
DNA Methylation
Gene
Albumin (*ALB*)
B
14 kb
235,208 kb
235,210 kb
235,212 kb
235,214 kb
235,216 kb
235,218 kb
H3K4me3
H3K27ac
H3K4me1
H3K27me3
DNA Methylation
Gene
Solute carrier family 11 member 1 (*SLC11A1*)

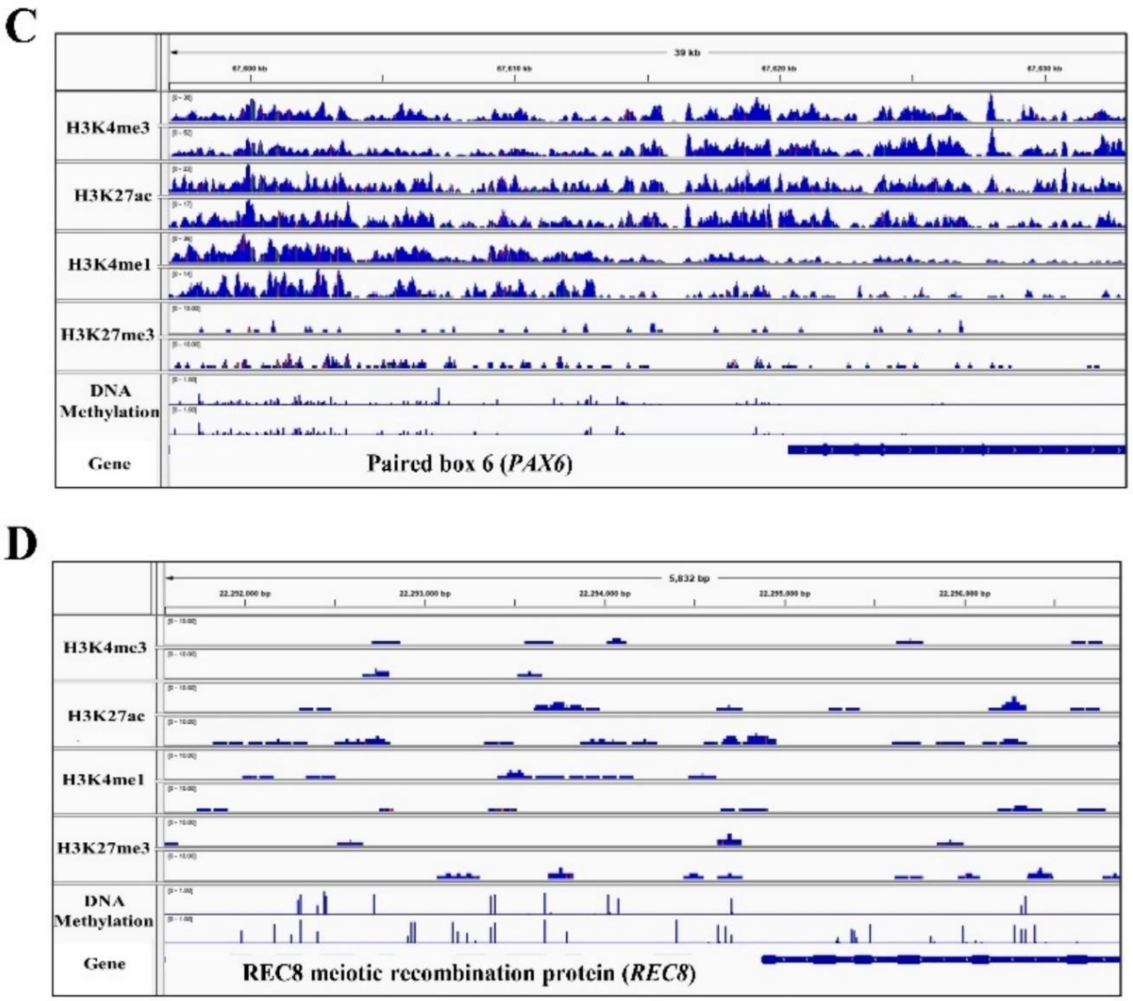


Figure 5.3: Integrative Genomics Viewer (IGV) screenshot of sequence pileup normalized with the input control for active and repressive histone marks and DNA methylation in two representative samples (M1 and M2) for (A) positive control Albumin (*ALB*) gene in liver, (B) positive control Solute carrier family 11 member 1 (*SLC11A1*) in spleen, (C) positive control Paired box 6 (*PAX6*) in cerebellum, (D) negative control REC8 gene (*REC8*) in all three tissues.

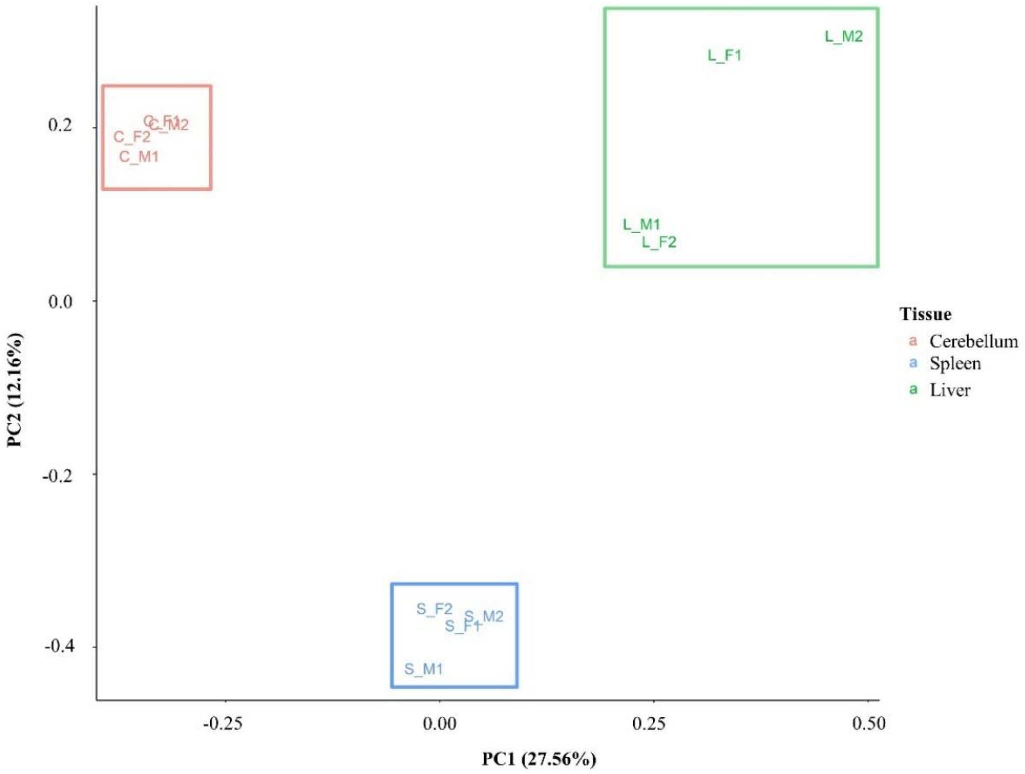


Figure 5.4: Principal component analysis plot based on CG methylation. Four animals are labelled as F1, F2, M1 and M2. Cerebellum, liver, and spleen samples are labelled as C, L and S, respectively.

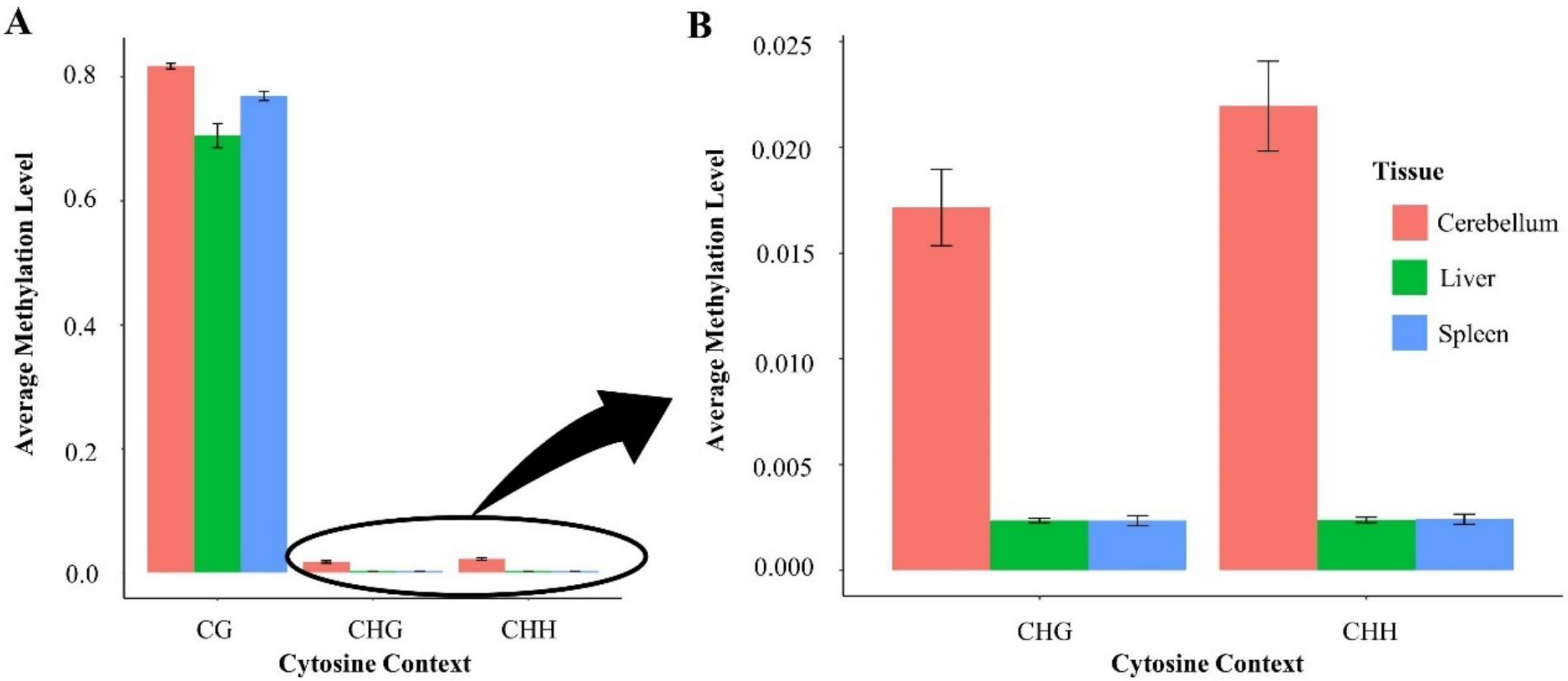
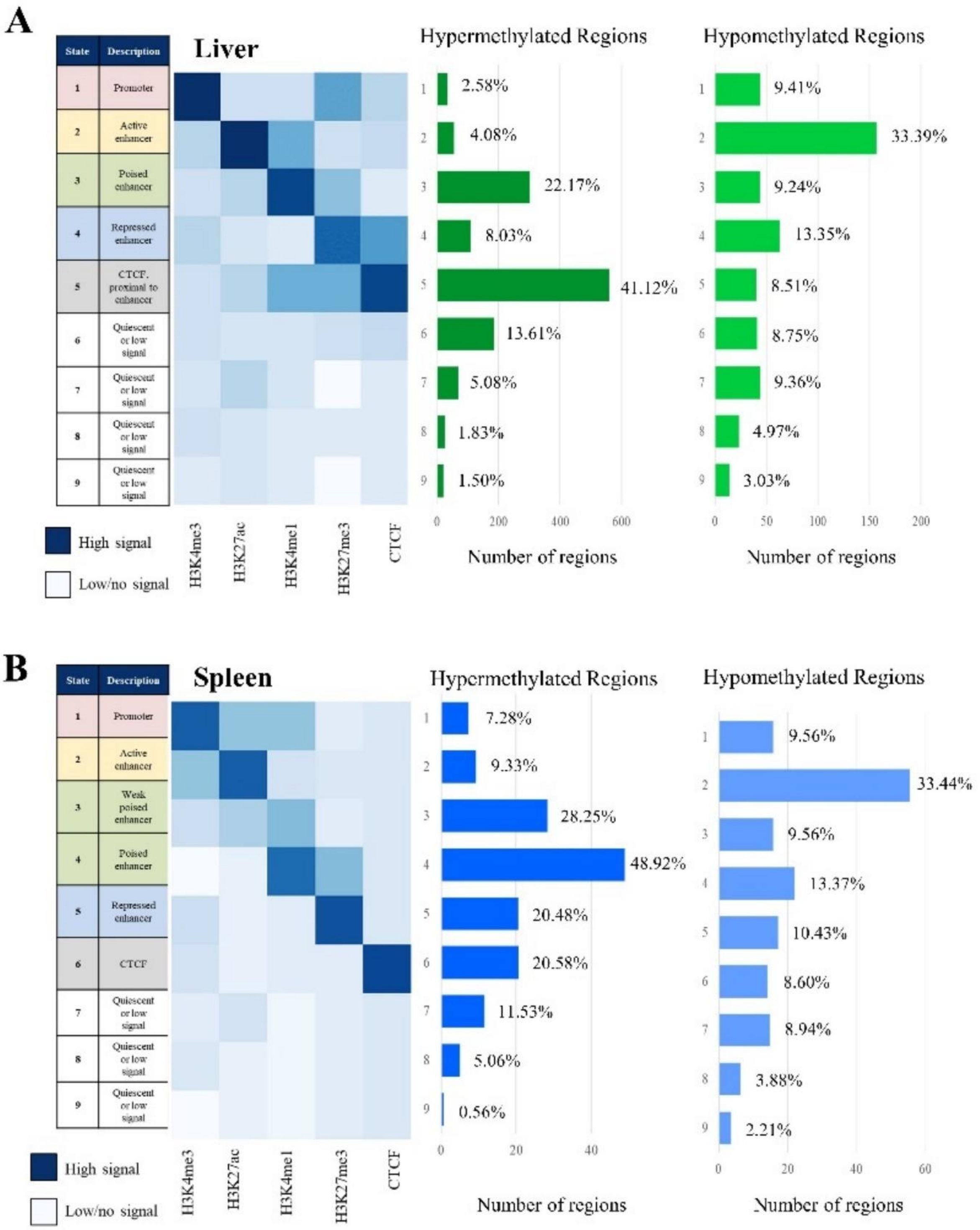
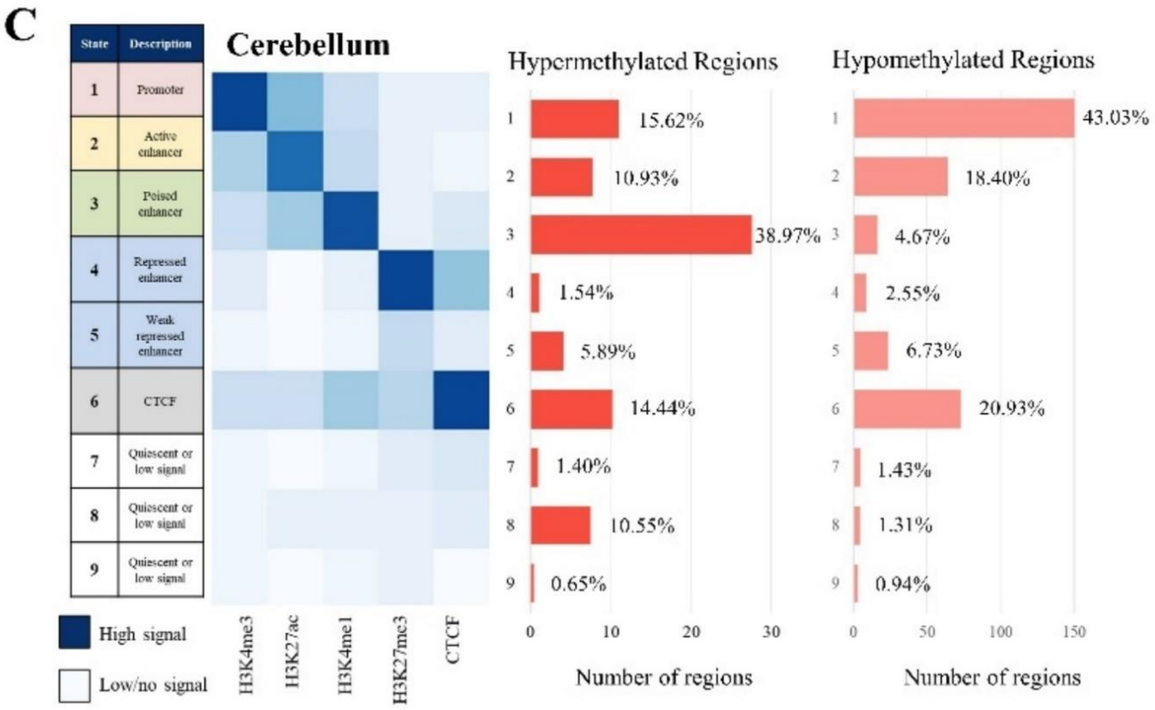


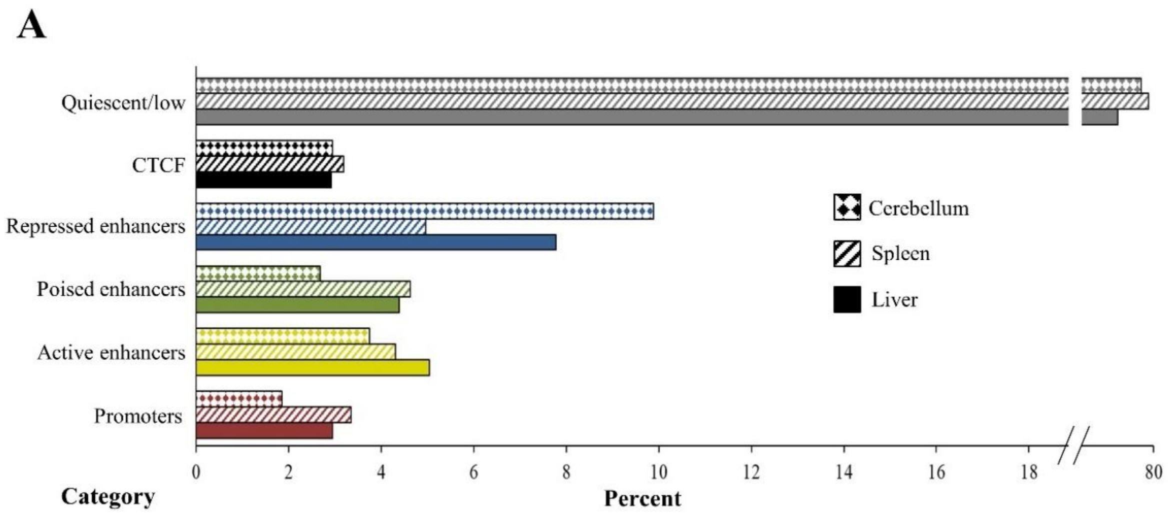
Figure 5.5: (A) Methylation level at CG compared to non-CG sites in liver, spleen, and cerebellum, and (B) Methylation level at non-CG sites (CHG and CHH) sites in each tissue enlarged.


A
State
Description
1 Promoter
2 Active enhancer
3 Poised enhancer
4 Repressed enhancer
5 CTCF, proximal to enhancer
6 Quiescent or low signal
7 Quiescent or low signal
8 Quiescent or low signal
9 Quiescent or low signal
Liver
High signal
Low/no signal
H3K4me3
H3K27ac
H3K4me1
H3K27me3
CTCF
Hypermethylated Regions
1 2.58%
2 4.08%
3 22.17%
4 8.03%
5 41.12%
6 13.61%
7 5.08%
8 1.83%
9 1.50%
0
200
400
600
Number of regions
Hypomethylated Regions
1 9.41%
2 33.39%
3 9.24%
4 13.35%
5 8.51%
6 8.75%
7 9.36%
8 4.97%
9 3.03%
0
50
100
150
200
Number of regions
B
State
Description
1 Promoter
2 Active enhancer
3 Weak poised enhancer
4 Poised enhancer
5 Repressed enhancer
6 CTCF
7 Quiescent or low signal
8 Quiescent or low signal
9 Quiescent or low signal
Spleen
High signal
Low/no signal
H3K4me3
H3K27ac
H3K4me1
H3K27me3
CTCF
Hypermethylated Regions
1 7.28%
2 9.33%
3 28.25%
4 48.92%
5 20.48%
6 20.58%
7 11.53%
8 5.06%
9 0.56%
0
20
40
Number of regions
Hypomethylated Regions
1 9.56%
2 33.44%
3 9.56%
4 13.37%
5 10.43%
6 8.60%
7 8.94%
8 3.88%
9 2.21%
0
20
40
60
Number of regions

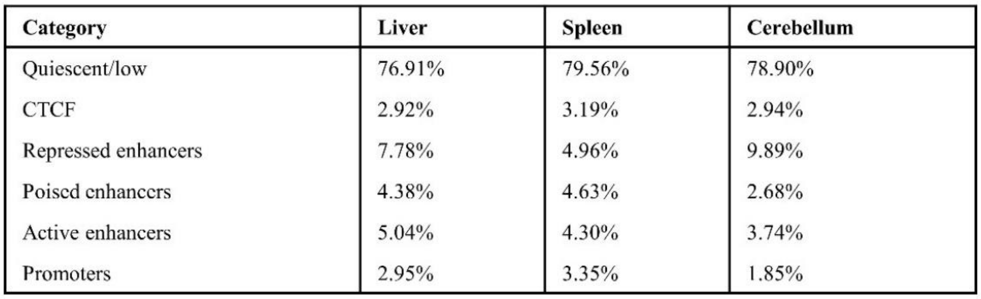


State	Description
1	Promoter
2	Active enhancer
3	Poised enhancer
4	Repressed enhancer
5	Weak repressed enhancer
6	CTCF
7	Quiescent or low signal
8	Quiescent or low signal
9	Quiescent or low signal

Figure 5.6: Chromatin state description and ChromHMM heatmap with histone mark signal overlap consensus from M1 and M2 compared with the number of hypermethylated regions and hypomethylated region consensus per Mb for M1 and M2 for (A) liver, (B) spleen, and (C) cerebellum.



B



Category	Liver	Spleen	Cerebellum
Quiescent/low	76.91%	79.56%	78.90%
CTCF	2.92%	3.19%	2.94%
Repressed enhancers	7.78%	4.96%	9.89%
Poised enhancers	4.38%	4.63%	2.68%
Active enhancers	5.04%	4.30%	3.74%
Promoters	2.95%	3.35%	1.85%

Figure 5.7: Percent of the genome in liver, spleen, and cerebellum (from M1 and M2) assigned to each category of quiescent/low (grey), CTCF (black), repressed enhancer (blue), poised enhancer (green), active enhancer (gold), and promoter (red) depicted visually in (A) the bar graph, and numerically in (B) the table.

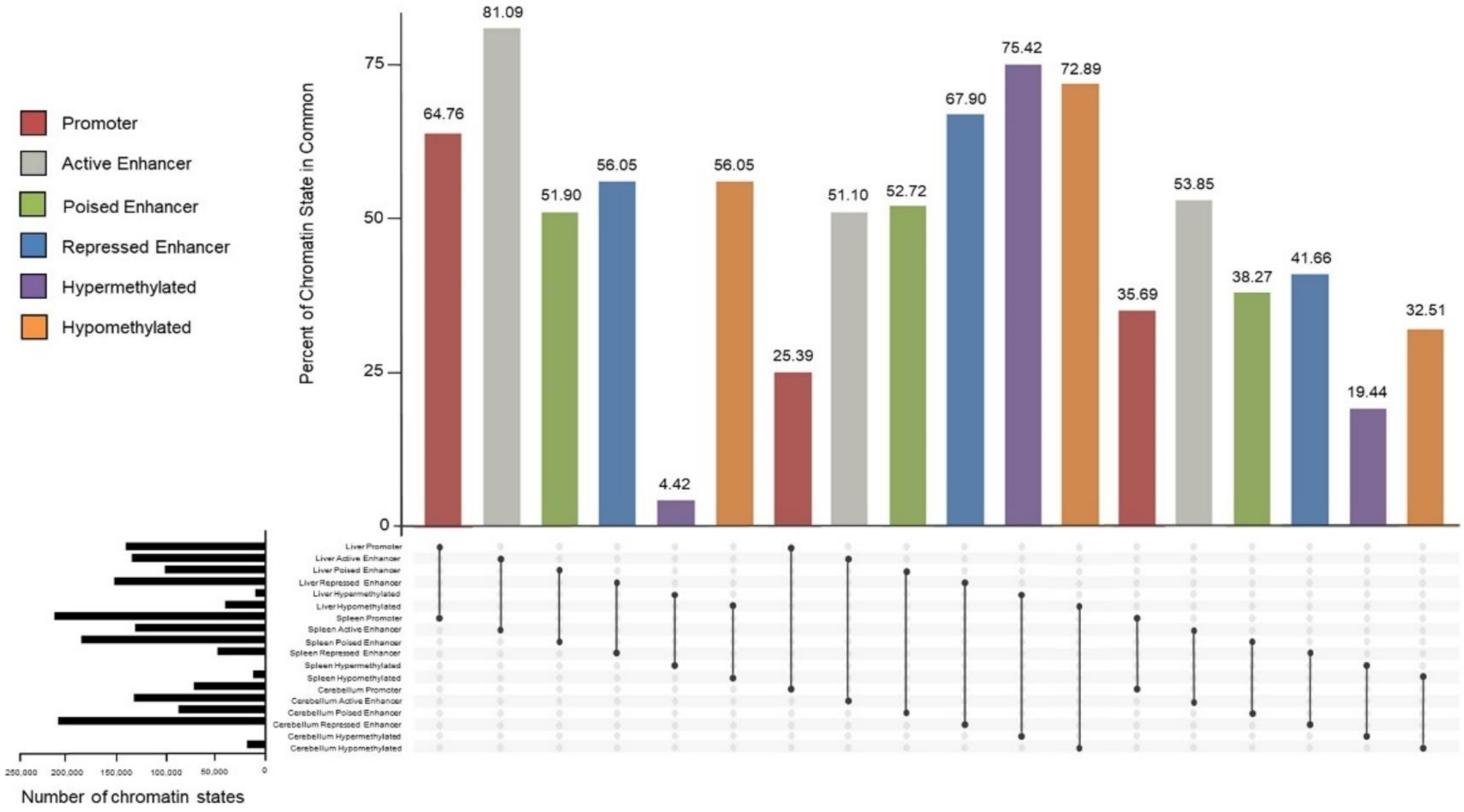


Figure 5.8: Percent of overlapping promoter (red), active enhancer (grey), poised enhancer (green), and repressed enhancer (blue) chromatin state categories, and hypermethylated (purple) and hypomethylated (orange) regions between liver, spleen, and cerebellum tissues the consensus categories from M1 and M2. The total number of chromatin states for each tissue is displayed in black horizontal bars

Supplementary Material

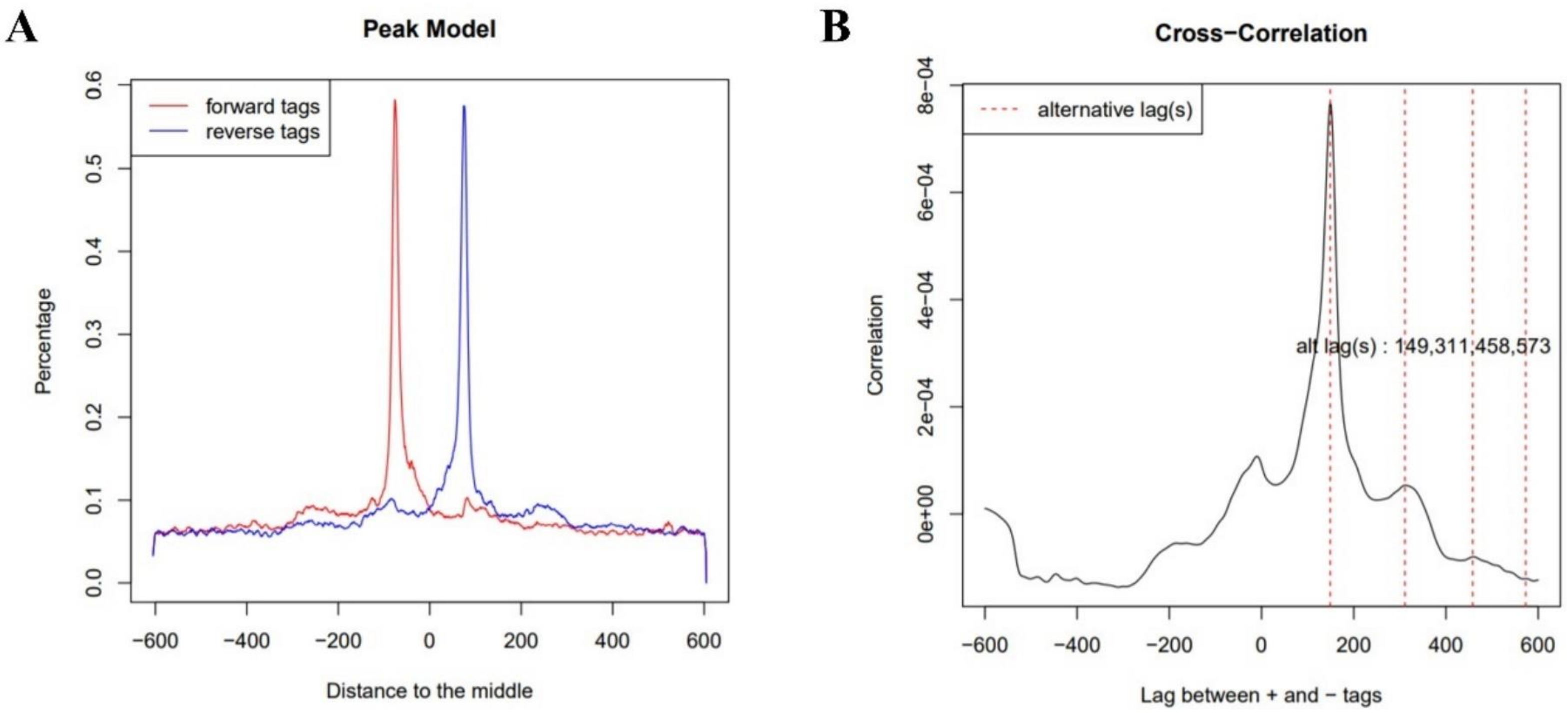
Supplementary Table 5.1: ChIP-seq quality metrics for each library. In the sample label column, L, S, and C represent liver, spleen, and cerebellum, respectively. F and M followed by a number represent female and male animal numbers, respectively. NRF represents the non-redundant fraction of the library and FRiP represents the fraction of reads in peaks.

Tissue	Histone Mark	Sample	Unique Mapping %	Number of Uniquely Mapped Reads	NRF	Number of Peaks	FRiP
Liver	**H3K4me3**	L_F1	76.26%	91,892,045	0.43	10,648	0.01
		L_F2	75.41%	51,369,218	0.27	9,062	0.01
		L_M1	74.35%	27,240,631	0.82	10,745	0.01
		L_M2	74.04%	37,575,540	0.83	11,376	0.01
	H3K27ac	L_F1	83.57%	21,882,178	0.69	25,464	0.02
		L_F2	84.38%	55,814,359	0.04	29,661	0.07
		L_M1	83.66%	59,037,311	0.81	27,123	0.01
		L_M2	82.24%	28,313,543	0.91	39,965	0.02
	H3K4me1	L_F1	80.26%	31,870,135	0.65	40,632	0.14
		L_F2	84.06%	74,783,676	0.62	30,153	0.15
		L_M1	82.98%	33,532,918	0.90	60,874	0.25
		L_M2	80.56%	54,530,540	0.91	59,655	0.26
	H3K27me3	L_F1	78.98%	46,772,724	0.31	33,340	0.01
		L_M1	79.98%	42,684,307	0.59	44,710	0.16
		L_M2	76.54%	46,898,657	0.84	33,583	0.08
	CTCF	L_F1	62.41%	69,727,193	0.01	29,893	0.07
		L_F2	70.58%	40,958,273	0.03	30,762	0.03
		L_M1	78.55%	41,219,481	0.07	16,036	0.14
		L_M2	81.50%	39,388,145	0.19	29,378	0.13
Spleen	**H3K4me3**	S_F1	73.77%	25,536,400	0.27	16,936	0.01
		S_F2	77.72%	25,364,432	0.78	10,381	0.01
		S_M1	78.84%	34,624,081	0.79	12,640	0.01
		S_M2	77.63%	28,406,550	0.80	13,601	0.01
	H3K27ac	S_F1	83.45%	38,909,684	0.47	43,423	0.02
		S_F2	82.32%	73,276,800	0.49	37,966	0.07
		S_M1	84.98%	41,271,154	0.84	38,294	0.02
		S_M2	79.74%	53,722,712	0.60	21,626	0.01
	H3K4me1	S_F1	75.54%	46,268,769	0.04	24,923	0.28
		S_F2	82.21%	91,786,917	0.06	39,769	0.15
		S_M1	77.83%	49,634,682	0.42	22,742	0.09

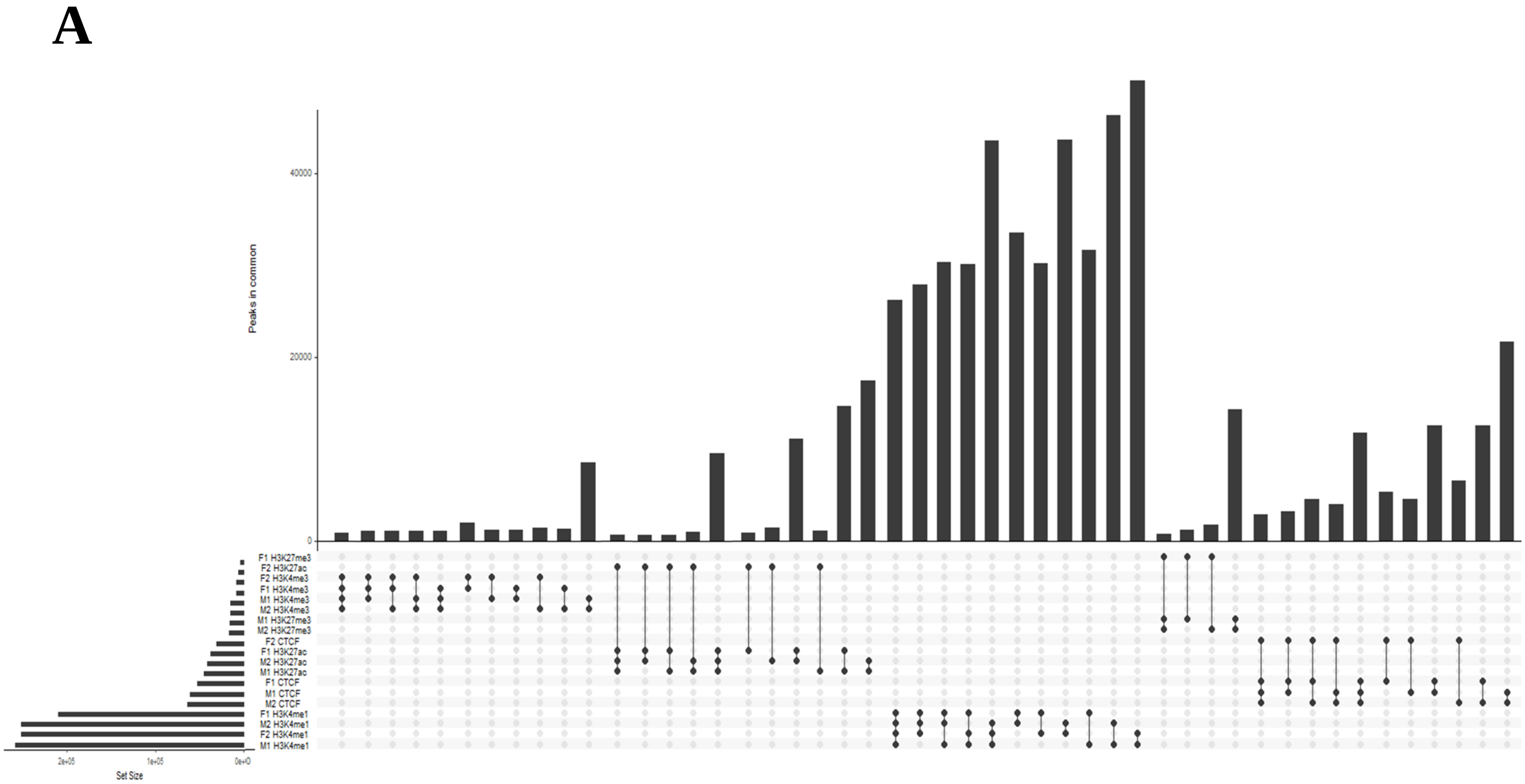
		S_M2	77.08%	37,868,373	0.06	30,762	0.25
	H3K27me3	S_F2	79.22%	56,617,636	0.04	22,051	0.76
		S_M1	79.03%	59,598,255	0.14	36,974	0.12
		S_M2	77.21%	40,734,132	0.06	22,482	0.39
	CTCF	S_F2	77.92%	73,893,274	0.02	33,599	0.16
		S_M1	76.93%	43,394,622	0.18	29,006	0.03
		S_M2	74.08%	41,908,970	0.06	22,482	0.06
Cerebellum	**H3K4me3**	C_F1	80.47%	38,211,509	0.86	16,542	0.01
		C_F2	80.14%	57,349,641	0.81	16,116	0.01
		C_M1	78.49%	24,280,387	0.95	21,463	0.01
		C_M2	79.31%	37,236,584	0.89	13,524	0.01
	H3K27ac	C_F1	75.77%	36,868,004	0.83	19,850	0.01
		C_F2	82.61%	40,111,785	0.71	40,069	0.01
		C_M1	83.27%	32,440,683	0.81	37,388	0.01
		C_M2	83.53%	26,027,091	0.93	30,174	0.01
	H3K4me1	C_F1	77.99%	31,415,709	0.94	47,600	0.10
		C_F2	82.35%	90,253,571	0.89	58,760	0.26
		C_M1	80.24%	38,566,786	0.92	51,334	0.36
		C_M2	79.97%	57,547,807	0.95	49,369	0.26
	H3K27me3	C_F1	75.56%	53,713,322	0.84	59,107	0.16
		C_F2	71.92%	54,875,161	0.94	31,614	0.14
		C_M1	74.11%	107,690,841	0.89	29,032	0.07
		C_M2	74.88%	98,190,441	0.98	27,818	0.09
	CTCF	C_F1	75.56%	53,713,322	0.25	22,405	0.01
		C_F2	75.51%	39,423,803	0.59	20,455	0.01
		C_M1	74.71%	28,256,482	0.74	30,170	0.01
		C_M2	75.00%	34,533,188	0.95	31,945	0.01

Supplementary Table 5.2: DNA methylation quality metrics for each library. In the sample label column, L, S, and C represent liver, spleen, and cerebellum, respectively. F and M followed by a number represent female and male animal numbers, respectively.

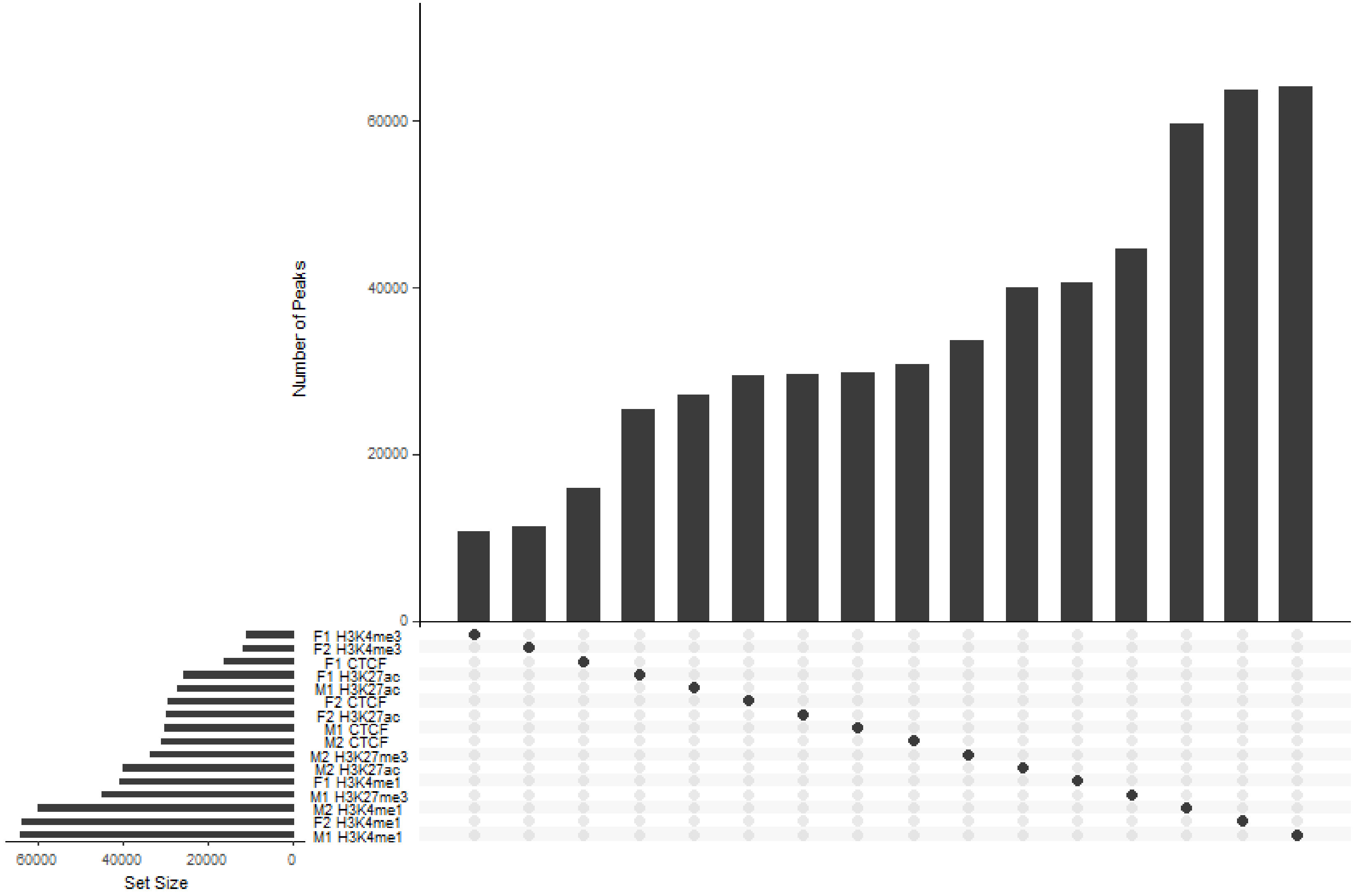
Tissue	Sample	Total PE Reads	Validated PE Reads	Uniquely mapped PE Reads	Unmapped read pairs	Bases used for mapping	Bases uniquely mapped	mCG (%)	mCHG (%)	mCHH (%)	Mappability (%)
Liver	L_F1	250802363	250773326	197117104	53656222	74105547825	58284149715	69.36	0.24	0.25	78.60
	L_F2	305118183	305106713	234081202	71025511	90938436967	69802293190	69.75	0.25	0.25	76.72
	L_M1	272377520	272367391	225870709	46496682	81189499413	67357790499	71.40	0.24	0.24	82.93
	L_M2	251382724	251374337	194073342	57300995	74923337158	57872016595	66.72	0.24	0.23	77.20
Spleen	S_F1	218460634	218451691	177966878	40484813	65111181250	53071678722	75.68	0.25	0.26	81.47
	S_F2	201358535	201273605	168451070	32822535	58730553503	49198143343	77.69	0.21	0.29	83.69
	S_M1	271369128	271357982	222096734	49261248	80874164415	66226685025	76.44	0.25	0.25	81.85
	S_M2	213583077	213573474	177063138	36510336	63652634020	52796851913	75.80	0.24	0.24	82.91
Cerebellum	C_F1	217567096	217558899	183491333	34067566	64855584915	54720375734	79.50	1.60	2.01	84.34
	C_F2	301162248	301152811	254518260	46634551	89774041124	75900622947	81.06	1.60	2.03	84.51
	C_M1	210546324	210540044	178412725	32127319	62755392823	53201279166	80.33	1.88	2.38	84.74
	C_M2	256873526	256864142	214165602	42698540	76567651870	63867562503	80.39	1.54	1.99	83.38
Average per tissue	Liver	1,079,680,790	1,079,621,767	851,142,357	228,479,410	321,156,821,363	253,316,249,999	69.31	0.24	0.24	78.86
	Spleen	904,771,374	904,656,752	745,577,820	159,078,932	268,368,533,188	221,293,359,003	76.40	0.24	0.26	82.48
	Cerebellum	986,149,194	986,115,896	830,587,920	155,527,976	293,952,670,732	247,689,840,350	80.32	1.65	2.10	84.24



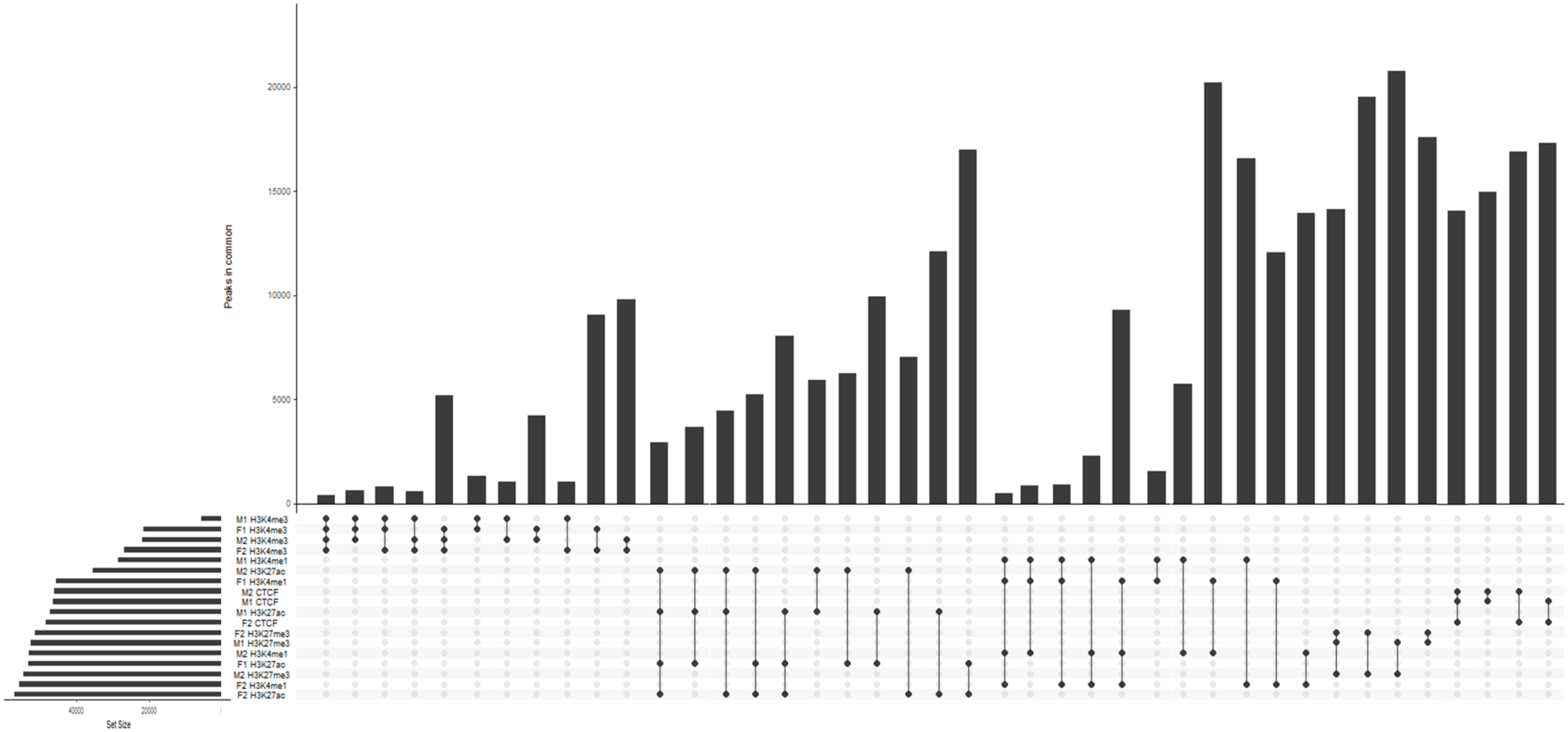
Supplementary Figure 5.1: (A) Peak model and (B) cross-correlation graph for narrow marks (MACS2).

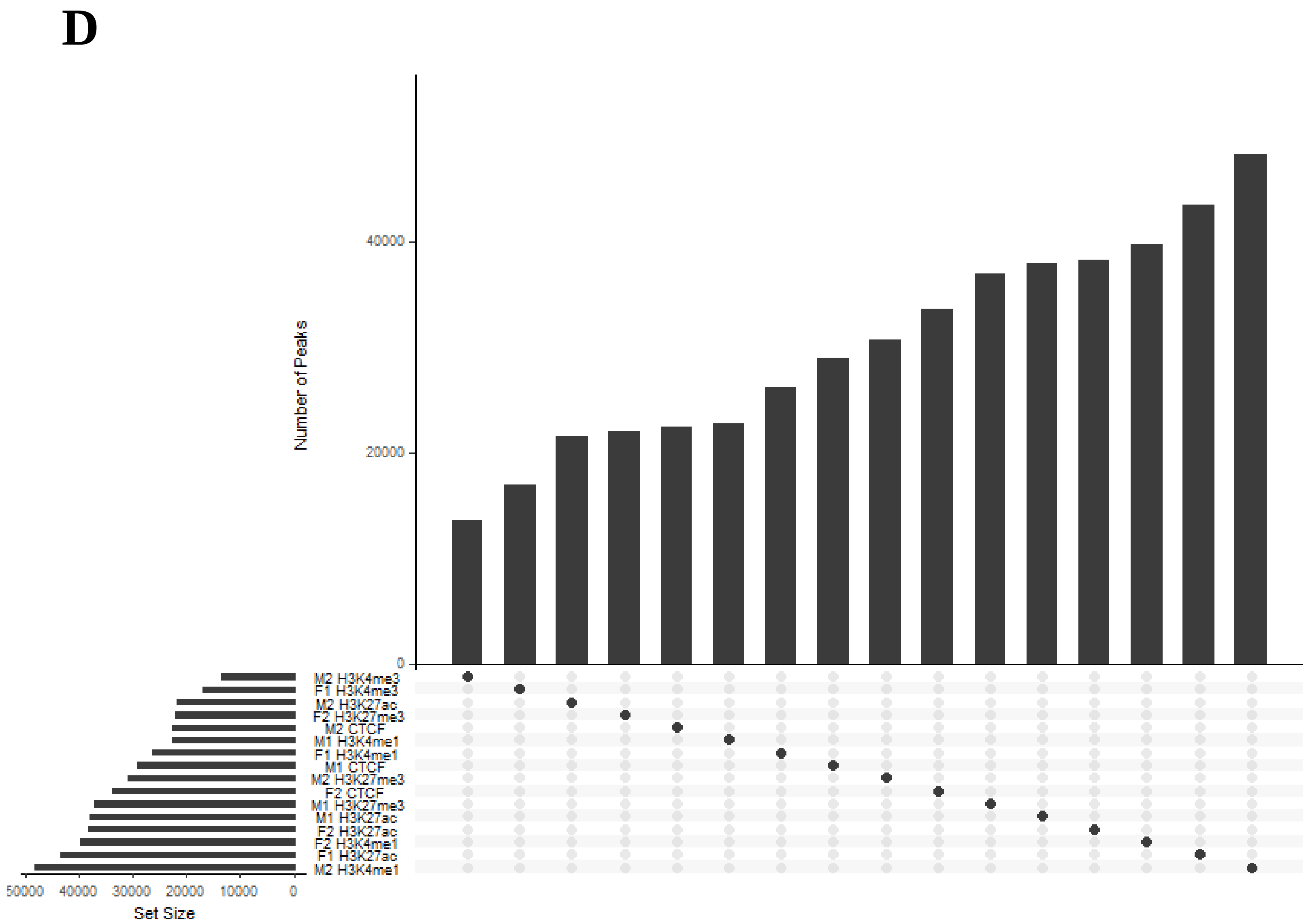

A
Peaks in common
40000
20000
0
F1 H3K27me3
F2 H3K27ac
F2 H3K4me3
F1 H3K4me3
M1 H3K4me3
M2 H3K4me3
M1 H3K27me3
M2 H3K27me3
F2 CTCF
F1 H3K27ac
M2 H3K27ac
M1 H3K27ac
F1 CTCF
M1 CTCF
M2 CTCF
F1 H3K4me1
M2 H3K4me1
F2 H3K4me1
M1 H3K4me1
2e+05
1e+05
0e+0
Set Size

B

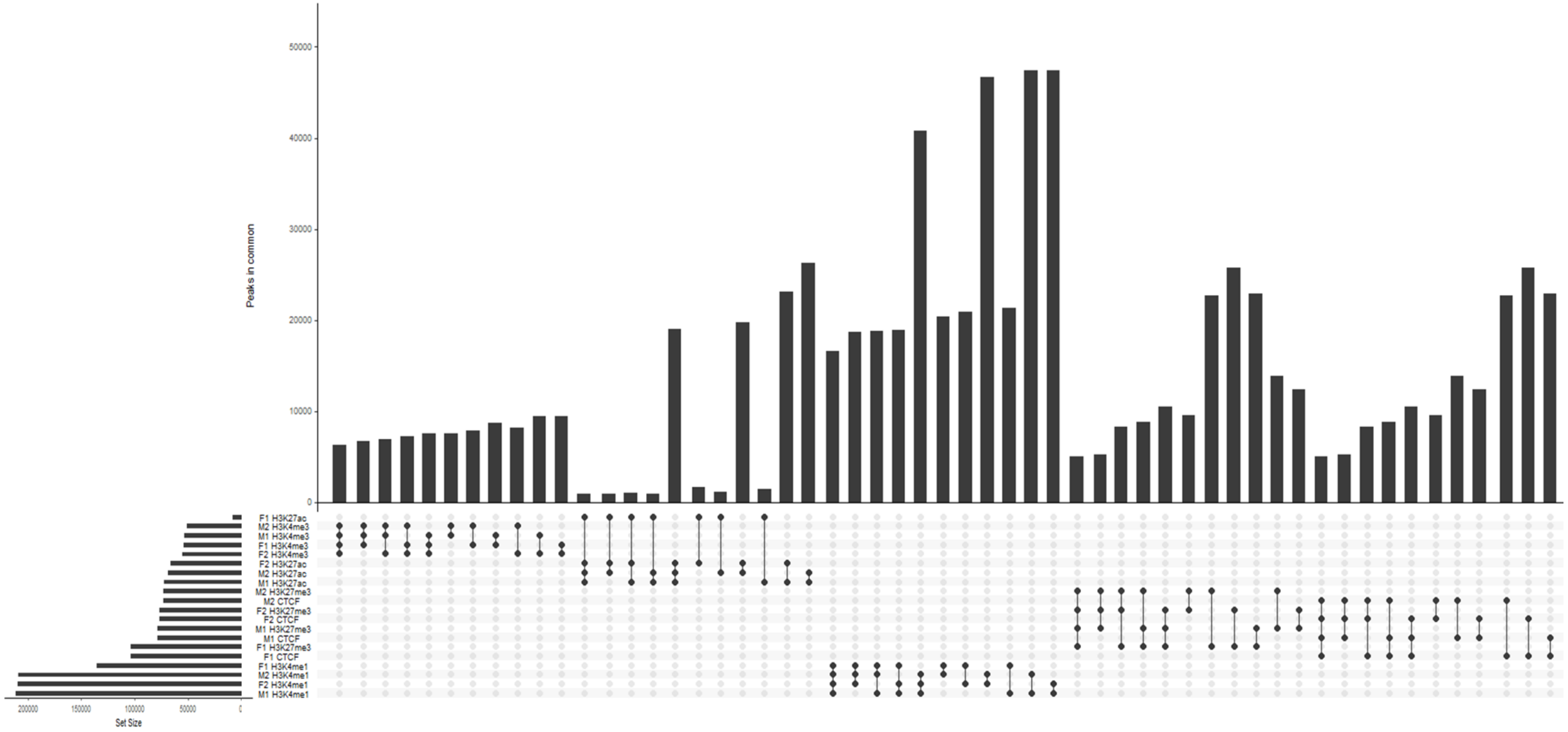

Number of Peaks
60000
40000
20000
0
F1 H3K4me3
F2 H3K4me3
F1 CTCF
F1 H3K27ac
M1 H3K27ac
F2 CTCF
F2 H3K27ac
M1 CTCF
M2 CTCF
M2 H3K27me3
M2 H3K27ac
F1 H3K4me1
M1 H3K27me3
M2 H3K4me1
F2 H3K4me1
M1 H3K4me1
60000
40000
20000
0
Set Size

C

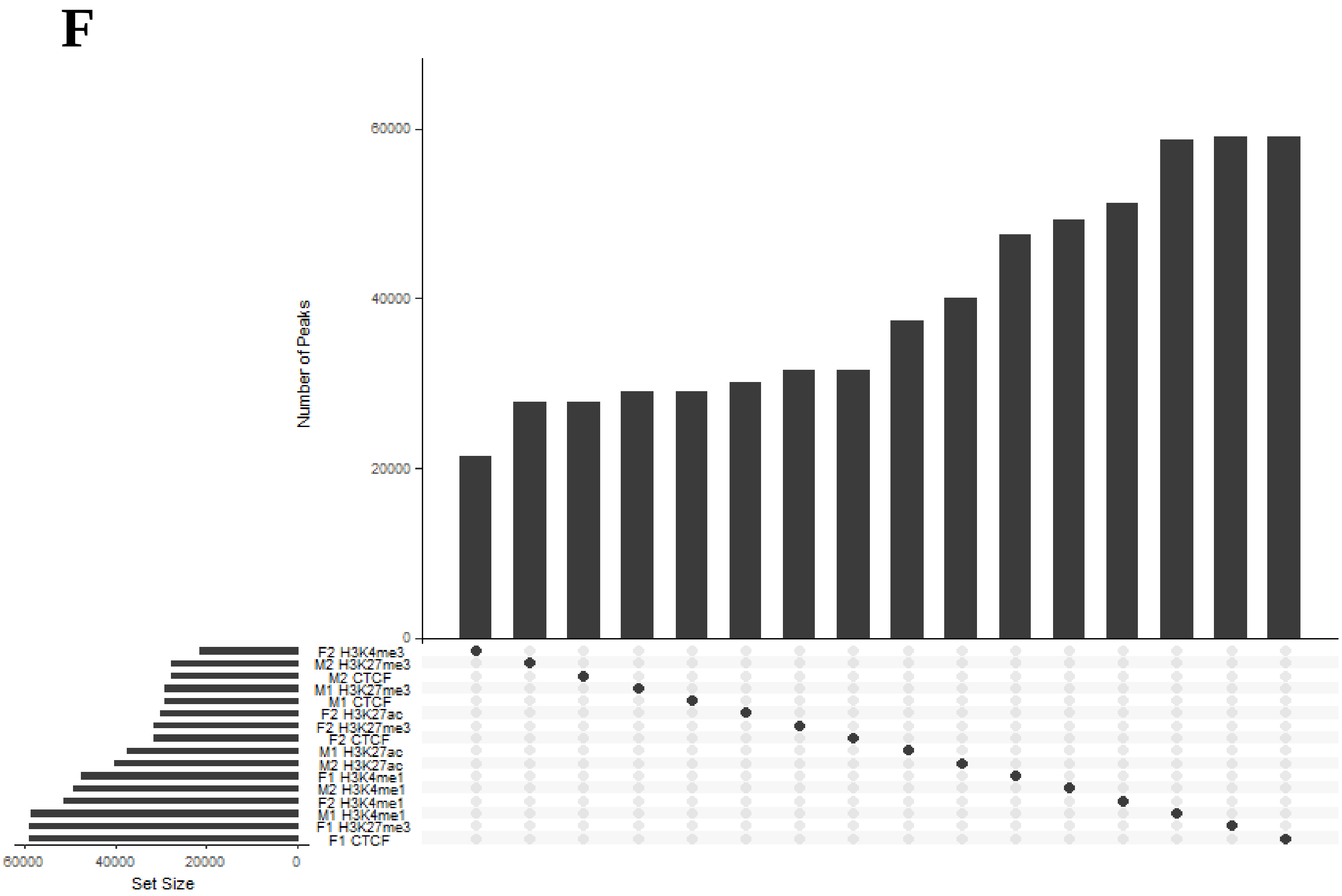

Peaks in common
20000
15000
10000
5000
0
M1 H3K4me3
F1 H3K4me3
M2 H3K4me3
F2 H3K4me3
M1 H3K4me1
M2 H3K27ac
F1 H3K4me1
M2 CTCF
M1 CTCF
M1 H3K27ac
F2 CTCF
F2 H3K27me3
M1 H3K27me3
M2 H3K4me1
F1 H3K27ac
M2 H3K27me3
F2 H3K4me1
F2 H3K27ac
40000
20000
Set Size


D
Number of Peaks
40000
20000
0
M2 H3K4me3
F1 H3K4me3
M2 H3K27ac
F2 H3K27me3
M2 CTCF
M1 H3K4me1
F1 H3K4me1
M1 CTCF
M2 H3K27me3
F2 CTCF
M1 H3K27me3
M1 H3K27ac
F2 H3K27ac
F2 H3K4me1
F1 H3K27ac
M2 H3K4me1
50000
40000
30000
20000
10000
0
Set Size

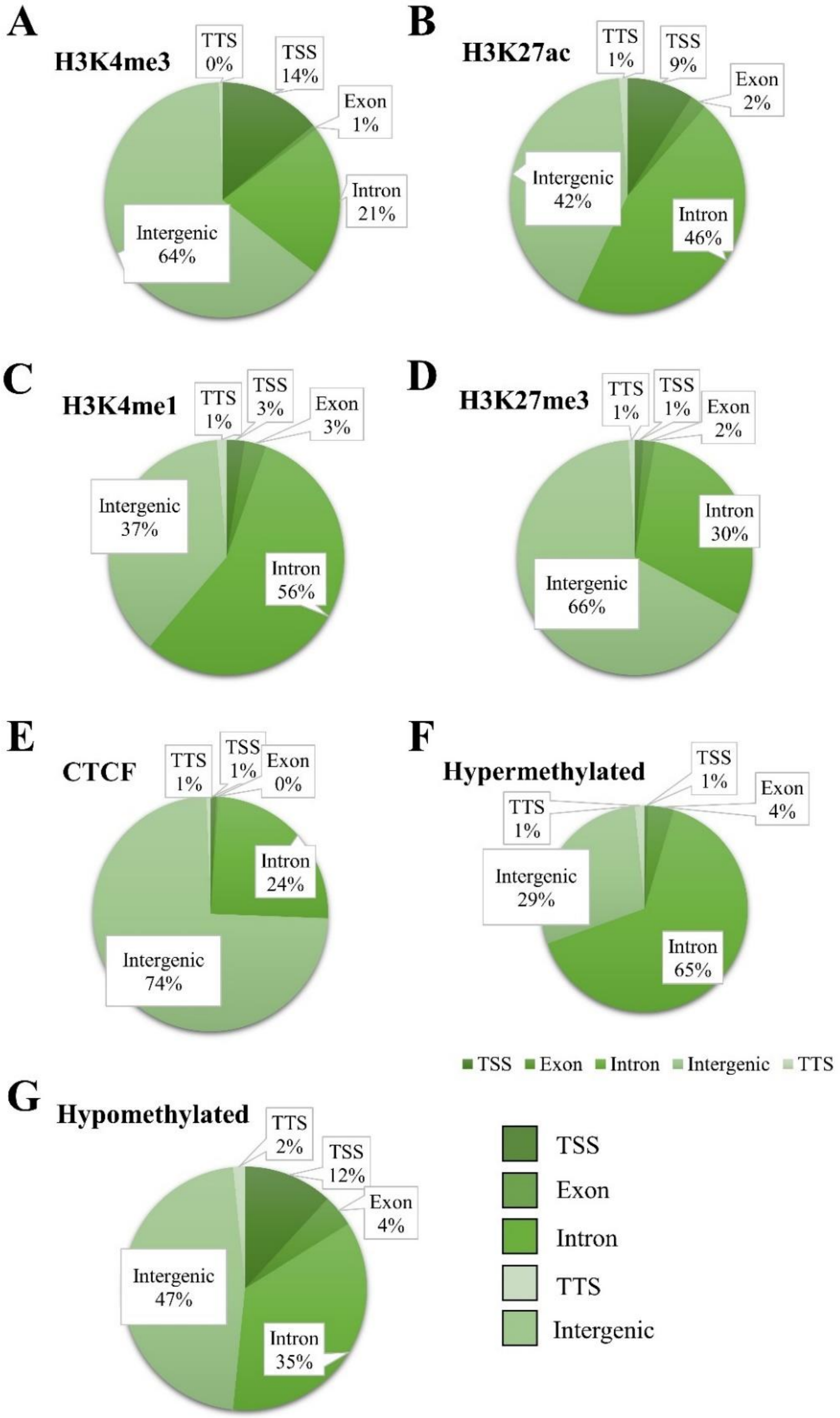
E



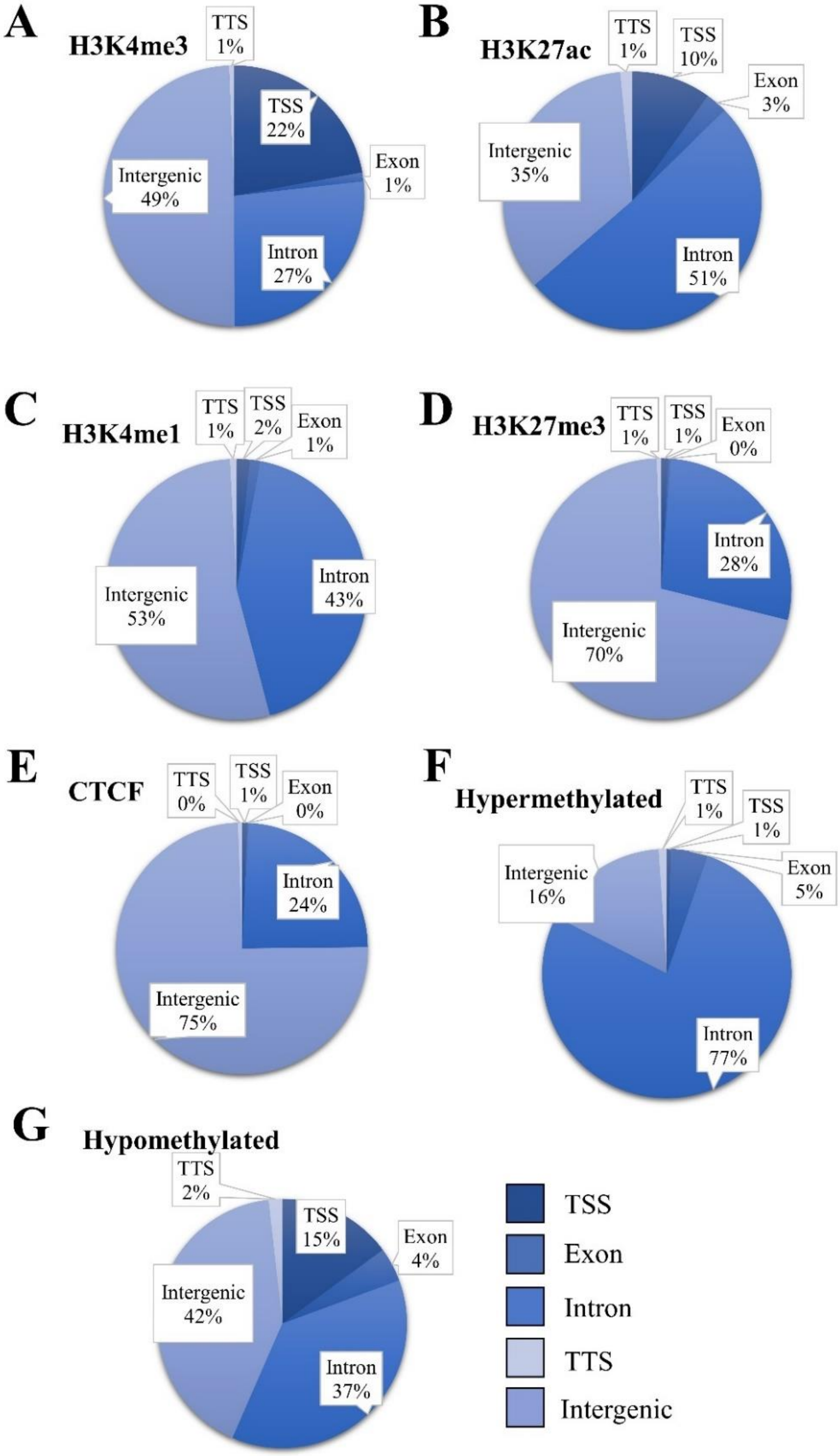
Peaks in common
0
10000
20000
30000
40000
50000
F1 H3K27ac
M2 H3K4me3
M1 H3K4me3
F1 H3K4me3
F2 H3K4me3
F2 H3K27ac
M2 H3K27ac
M1 H3K27ac
M2 H3K27me3
M2 CTCF
F2 H3K27me3
F2 CTCF
M1 H3K27me3
M1 CTCF
F1 H3K27me3
F1 CTCF
F1 H3K4me1
M2 H3K4me1
F2 H3K4me1
M1 H3K4me1
200000
150000
100000
50000
0
Set Size



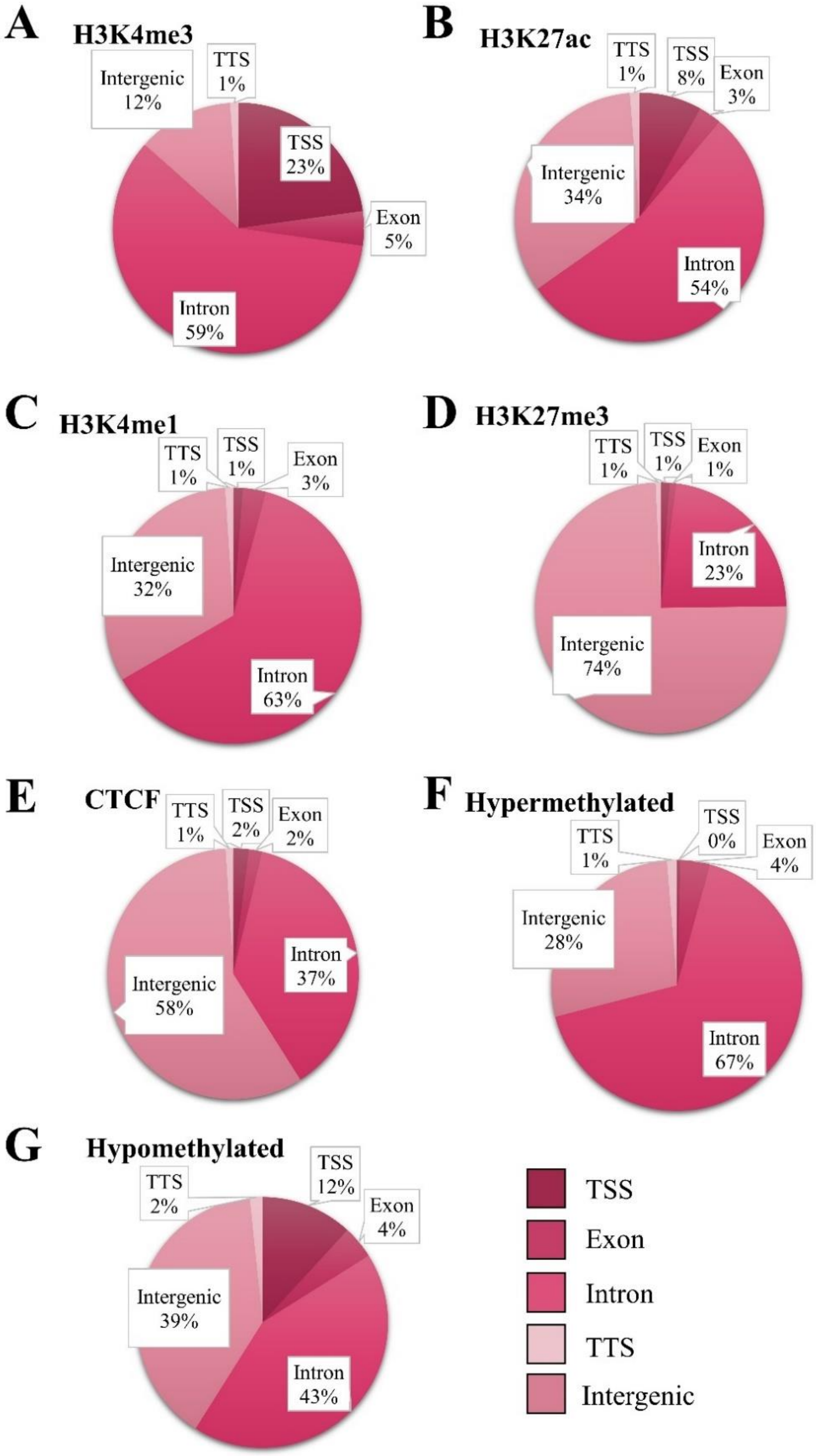
Supplementary Figure 5.2: Histone and CTCF peaks in common between all samples (A, C, and E) and total peak numbers (B, D, and F) per sample in (A) and (B) liver samples, (C) and (D) spleen samples, and (E) and (F) cerebellum samples.



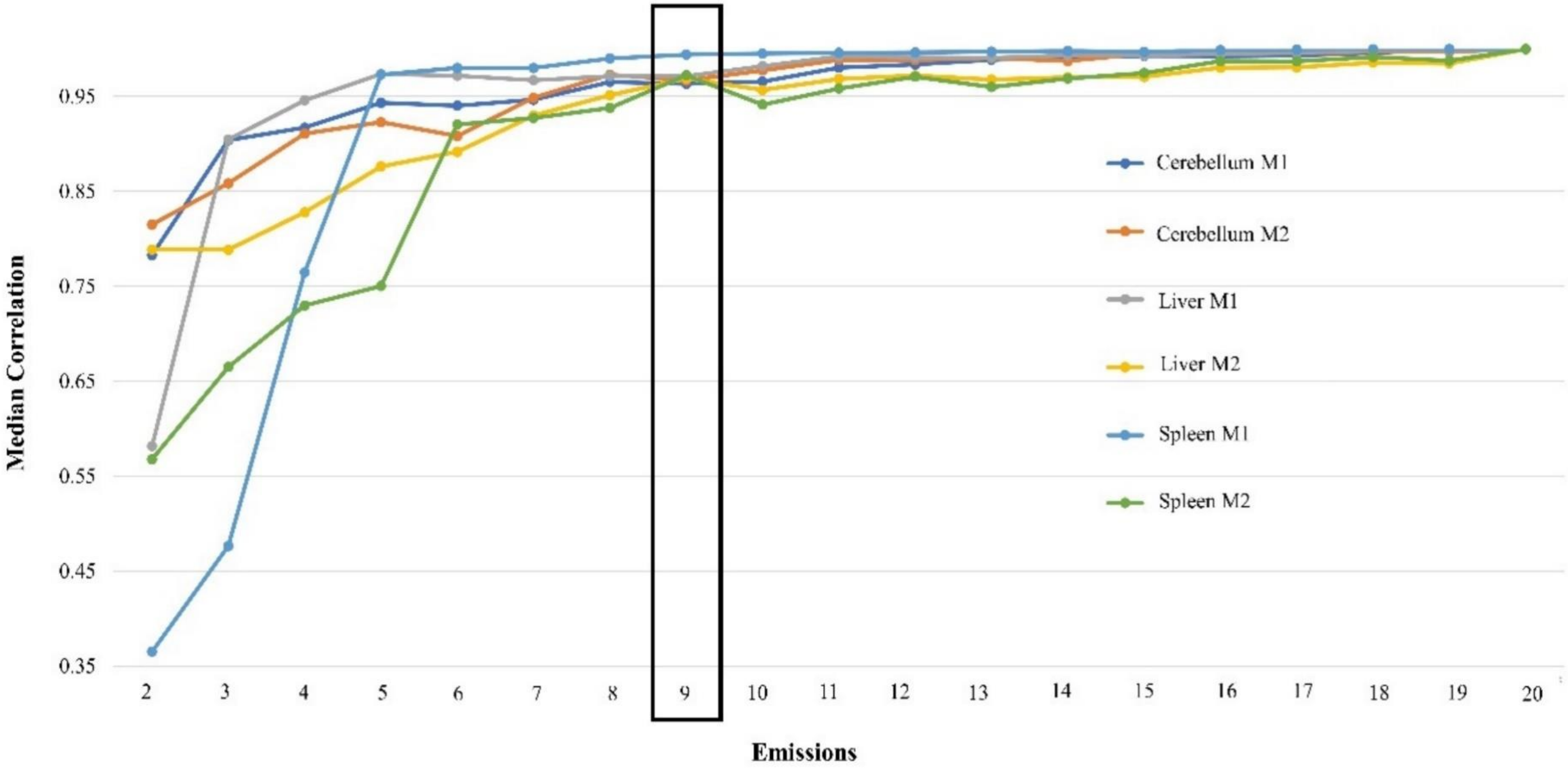
Supplementary Figure 5.3: Liver annotation (*Oar_rambouillet_v1.0*) for A) H3K4me3, B) H3K27ac, C) H3K4me1, D) H3K27me3, E) CTCF, F) hypermethylated regions, and G) hypomethylated regions.



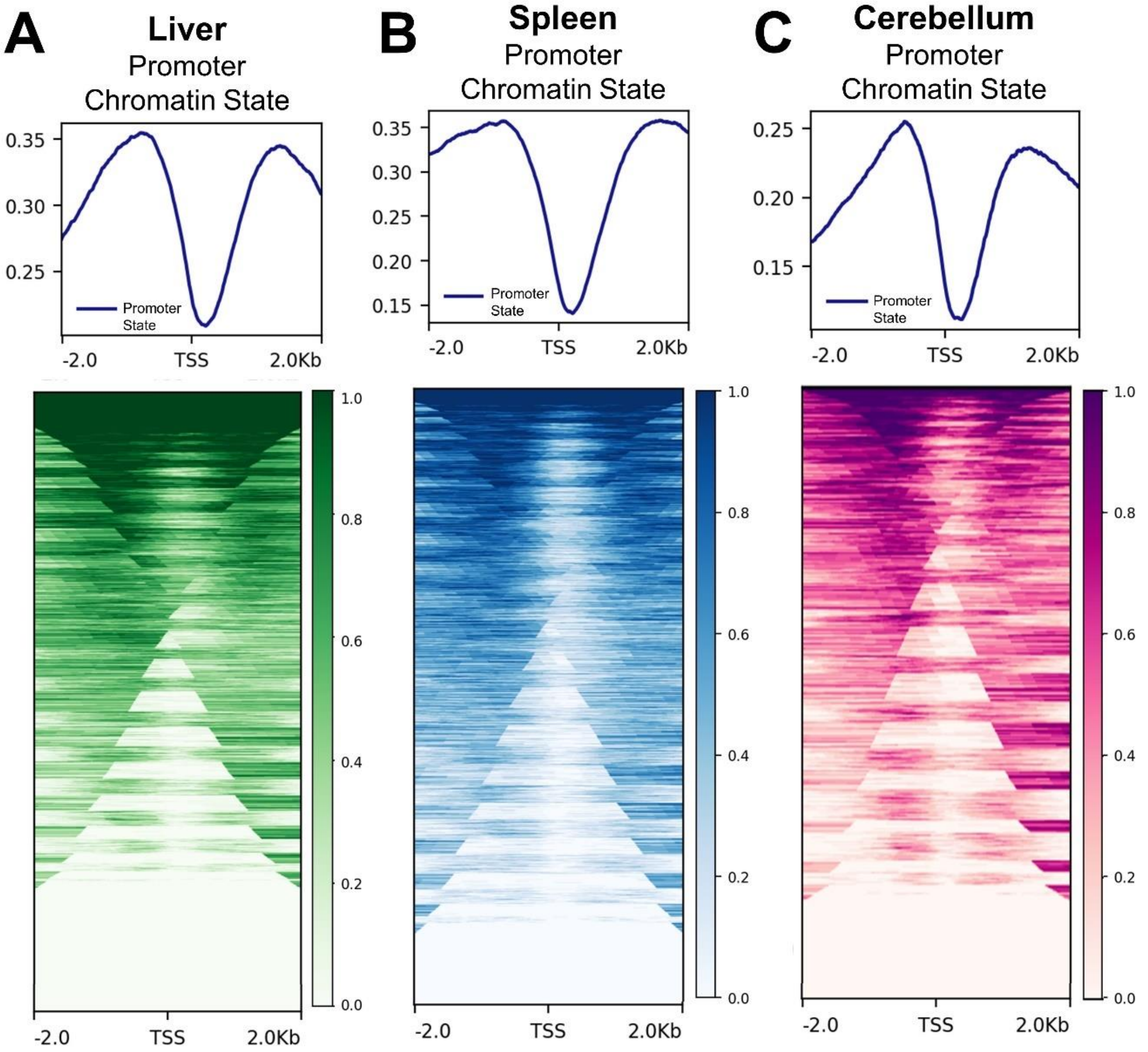
Supplementary Figure 5.4: Spleen annotation (*Oar_rambouillet_v1.0*) for A) H3K4me3, B) H3K27ac, C) H3K4me1, D) H3K27me3, E) CTCF, F) hypermethylated regions, and G) hypomethylated regions.



Supplementary Figure 5.5: Cerebellum annotation (*Oar_rambouillet_v1.0*) for A) H3K4me3, B) H3K27ac, C) H3K4me1, D) H3K27me3, E) CTCF, F) hypermethylated regions, and G) hypomethylated regions.



Supplementary Figure 5.6: The median correlation of each chromatin state model in male 1 (M1) and male 2 (M2) with 2-20 given emissions compared with 20 emission states. The optimal number of states was selected as the lowest number of states with above 0.95 correlation (9 states).



Supplementary Figure 5.7: Promoter chromatin state proximity to annotated transcription start sites identified with CAGE assays in A) liver, B) spleen, and C) cerebellum.

Chapter 6: Defining Genetic Regulatory Elements in the Ovine Genome Provides Insight into Transcriptional Regulation Across Tissues

Kimberly M. Davenport[1], Mazdak Salavati[2], Alex Caulton[3], Emily L. Clark[2], Alan Archibald[2], Shannon Clarke[3], Rudiger Brauning[3], Alisha T. Massa[4], Michelle R. Mousel[5], Stephen N. White[5,6,] Kim C. Worley[7], Brian Dalrymple[8], James Kijas[9], Tracy Hadfield[10], Benjamin D. Rosen[11], Timothy P.L. Smith[12], Noelle E. Cockett[10], and Brenda M. Murdoch[1,6]

[1]Department of Animal, Veterinary, & Food Sciences, University of Idaho, USA

[2]The Roslin Institute and Royal (Dick) School of Veterinary Studies, University of Edinburgh, UK

[3]AgResearch, New Zealand

[4]Department of Veterinary Microbiology and Pathology, Washington State University, USA

[5]USDA, ARS, Animal Disease Research Unit, USA

[6]Center for Reproductive Biology, Washington State University, USA

[7]Baylor College of Medicine-Human Genome Sequencing Center, USA

[8]University of Western Australia

[9]CSIRO Agricultural Flagship, Australia

[10]Utah State University, USA

[11]USDA ARS Beltsville Agricultural Research Center (BARC), USA

[12]USDA, ARS, U.S. Meat Animal Research Center (USMARC), USA

Abstract

Defining the locations of genetic regulatory elements is critical for understanding the regulatory mechanisms of complex phenotypic traits related to production traits and health in all species. The Ovine Functional Annotation of Animal Genomes (FAANG) Project aims to characterize transcriptional regulatory elements across the sheep genome in a large collection of tissues to facilitate a better understanding of the biological mechanisms influencing phenotypic traits in sheep. Approximately 100 tissues were collected from the Rambouillet ewe, Benz 2616, used to assemble the ovine reference genome ARS-UI_Ramb v2.0. Assays including sequencing of messenger RNA (mRNA-seq), microRNA (miRNA-seq), and long non-coding RNA (Iso-seq), cap analysis of gene expression (CAGE), chromatin immunoprecipitation with sequencing (ChIP-seq), assay for transposase-accessible chromatin with sequencing (ATAC-seq), whole genome bisulfite sequencing (WGBS) and reduced representation bisulfite sequencing (RRBS) were performed on a subset of these tissues. This manuscript details the ChIP-seq (H3K4me3, H3K27ac, H3K4me1, and H3K27me3), ATAC-seq, DNA methylation, and RNA-seq overlay between tissues. Nine chromatin states depicting promoters and enhancers (active, poised, and repressed) across the genome were defined using ChromHMM with histone modifications and compared across tissues. These chromatin states in combination with ATAC-seq, DNA methylation, and RNA-seq provide the basis of functional annotation in the ovine genome. These data suggest that active promoter and enhancer states reside in open chromatin regions with a greater number of both transcriptional activity and hypomethylated regions than other states. Further, poised and repressed enhancers did not primarily reside in open chromatin and had less transcriptional activity and more hypermethylated sites compared with active states. Together these data support each other to define transcriptional regulatory regions throughout the ovine genome. Characterizing regulatory elements in sheep will provide a valuable resource to facilitate a deeper understanding of how gene-regulation control influences complex traits in this globally important livestock species.

Keywords: FAANG, Sheep, ChIP-seq, ATAC-seq, RNA-seq, WGBS, RRBS, gene regulation, transcription, regulatory elements, functional annotation

Introduction

Defining the locations of genetic regulatory elements in the genome is extremely important for understanding transcriptional regulation and how it relates to phenotypes of interest. The encyclopedia of DNA elements (ENCODE) project has performed functional annotation in human cells and tissue types (ENCODE Project Consortium, 2012; Andersson et al., 2015). This project was closely followed by other model species including mouse (Shen et al., 2012; Yue et al., 2014), *Caenorhabditis elegans* (Gerstein et al., 2010), and zebrafish (Sivasubbu et al., 2013). These studies have already revealed numerous differences in regulatory elements across species (Gerstein et al., 2010; Barbosa-Morais et al., 2012; Sivasubbu et al., 2013; Yue et al., 2014). The functional annotation of animal genomes initiative (FAANG), that aims to identify genetic regulatory elements across the genomes of in domesticated species, will not only contribute to comparative and evolutionary perspectives of the control of transcription but also to enabling studies of the genetic control of economically important complex traits (Andersson et al., 2015; Tuggle et al., 2016; Giuffra et al., 2019; Clark et al., 2020). Transcribed loci, transcription start site locations, chromatin accessibility, histone modifications, methylation, and transcription factor binding sites are all important to complete the FAANG initiative in agricultural species. It is critical to define these regulatory regions in sheep to better understand phenotypes related to important traits such as meat, milk, and wool.

The FAANG community has begun to characterize the transcriptome across several species and a large collection of tissues. A gene expression atlas was created for the sheep using RNA-seq data (Clark et al., 2017). This study identified tissue specific expression at four developmental stages in Texel and Scottish Blackface sheep (Clark et al., 2017). Gebe expression atlases have also been established for other farmed animals including cattle (Harbay et al., 2010; Fang et al., 2020), water buffalo (Young et al., 2019), pigs (Summers et al., 2020), and chickens (Bush et al., 2018). Several studies have now advanced to the identification of regulatory sequences in the genomes of farmed animal species including cattle (Fang et al., 2019; Kang et al., 2020; Kern et al., 2021), pig, (Kern et al., 2021), chicken (Kern et al., 2021), and sheep (Massa et al., 2021; Davenport et al., 2021). Regulatory regions were first characterized in sheep by lifting over annotation from the human genome and using this enhanced annotation of the sheep genome to examine selection and domestication (Naval-Sanchez et al., 2018). Other studies have characterized regulatory elements with ChIP-seq in alveolar macrophages (Massa et al., 2021), and ChIP-seq and WGBS in liver, spleen, and brain tissue (Davenport et al., 2021). Here we report the first definition of transcriptional regulatory elements in a large collection of tissues with integration of data from the FAANG core assays to define gene regulatory regions in the reference sheep genome.

Materials & Methods

Tissue Collection

Tissues were collected from a healthy six-year-old Rambouillet ewe (Benz 2616, USMARC ID 200935900) as approved by the Utah State University Institutional Animal Care and Use Committee on April 29, 2016, at the Utah Veterinary Diagnostic Laboratory. This animal was selected by the Ovine Functional Annotation of Animal Genomes project and acquired from the USDA. Tissue samples collected postmortem were immediately transferred to cryogenic vials, followed by snap freezing in liquid nitrogen and subsequent transfer to -80°C freezers for storage. Lungs were subjected to a bronchiolar lavage to collect alveolar macrophages as described in Cordier et al., 1990. Further details regarding the tissue collection protocol are available through the FAANG Data Coordination Center’s data portal (data.faang.org), and metadata information logged according to FAANG guidelines can be found under BioSample accession SAMEG329607 (Harrison et al., 2018; Salavati et al., 2020).

Chromatin Immunoprecipitation with Sequencing

Chromatin immunoprecipitation (ChIP) was performed for 47 tissues (Supplementary Table 6.1) as described in Davenport et al., 2021. The SimpleChIP Plus Enzymatic Chromatin IP Kit (Cell Signaling Technologies, cat. #9005, Danvers, MA, USA) was used for immunoprecipitation according to the manufacturer’s instructions (Johnson et al., 2007; Barski et al., 2007; Robertson et al., 2007; Mikkelsen et al., 2007). Antibodies specific for histone modifications H3K4me3 (Abcam, cat. #ab8580), H3K27ac (Abcam, cat. #ab4729), H3K4me1 (Abcam, cat. #ab8895), and H3K27me3 (Abcam, cat. #ab6002) were used for immunoprecipitation. Tissues were crosslinked with 37% formaldehyde, disaggregated with Dounce homogenization, and chromatin was sheared with micrococcal nuclease (MNase) by incubating on a thermomixer for 20 minutes at 200 rpm and 37°C., Chromatin was incubated with 1 µg of antibody overnight (16 hours) at 4°C on a Hula mixer following nuclear membrane lysis and isolation of sheared chromatin by centrifugation at 15,000 *x g* for 1 minute. An input control (20 µl from each immunoprecipitation) was removed from the sheared chromatin for each tissue prior to incubation with antibodies and stored at -20°C until purification. Protein-G coated magnetic beads were used to isolate chromatin bound to specific antibodies and purification of enriched chromatin was performed with the DNA Purification Buffers and Spin Columns Kit following manufacturer's instructions (Cell Signaling Technologies, cat. #14209, Danvers, MA, USA).

Purified DNA was then quantified using a Qubit dsDNA HS Assay Kit (Thermo Fisher Scientific cat. #Q32854, Waltham, MA, USA), and size and integrity of DNA was verified with a Fragment Analyzer (Agilent, Santa Clara, CA, USA). Libraries were prepared as 75 base pair paired end reads with the TruSeq ChIP Library Preparation Kit (Illumina, Inc., cat. #IP-202-1012, San Diego, CA, USA) following manufacturer's instructions. Sequencing libraries for H3K4me3, H3K27ac, and input controls were sequenced to achieve at least 20 million uniquely mapped reads, and H3K4me1 and H3K27me3 libraries were sequenced to achieve at least 45 million uniquely mapped reads to comply with ENCODE standards for ChIP-seq (ENCODE Project Consortium, 2012).

Quality control of sequence data was performed with FastQC and MultiQC, followed by trimming of low-quality bases and adaptors with Trim Galore v0.64.1. Sequences were then mapped to the *ARS-UI_Ramb_v2.0* sheep genome using Bowtie2 (Langmead & Salzburg, 2012). Duplicates were removed with Picard (Broad Institute, 2019). Peaks were called for H3K4me3 and H3K27ac with MACS2 using a tissue specific input control and a false discovery rate of 0.05 (Zhang et al., 2008; Feng et al., 2012). Peaks for H3K4me1 and H3K27me3 were called with SICER with an input control and false discovery rate of 0.05 to better account for broad sequence pileup distributions (Zang et al., 2009; Micsiani et al., 2012; Siska and Kechris, 2017; Davenport et al., 2021). Quality statistics including the number of mapped reads, number of uniquely mapped reads, and non-redundant fraction

(NRF) were calculated for each sample using Samtools and Picard (Afgan et al., 2018; Heinz et al., 2010; Landt et al., 2012; Friedman and Alm, 2012; Siska and Kechris, 2017).

The average overall mapping percentages were 98.69%, 98.59%, 98.72%, and 97.79% for H3K4me3, H3K27ac, H3K4me1, and H3K27me3, respectively (Supplementary Table 6.2). The unique mapping percentages were 78.95%, 77.86%, 79.90%, and 77.78% for H3K4me3, H3K27ac, H3K4me1, and H3K27me3, respectively. All H3K4me3 and H3K27ac libraries reached above 20 million uniquely mapped reads, while all H3K4me1 and H3K27me3 libraries reached above 45 million uniquely mapped reads (Supplementary Table 6.3). The overall average duplication rate across sequence libraries was 0.16 (Supplementary Table 6.3).

Peaks were annotated using Homer with the *annotatepeaks.pl* function and classified as near a TSS (+2kb to -2kb), near a TTS site (within 2kb), within an exon, and intergenic. Dendrograms were created from ChIP-seq sequence signal from each of the histone modifications using a Spearman correlation calculated with deepTools and displayed as a cluster dendrogram in the factoextra package of R v3.6.2 (Salavati et al., 2020).

Assay for Transposase-Accessible Chromatin with Sequencing

Assay for Transposase Accessible Chromatin (ATAC) was performed for 33 tissues (Supplementary Table 6.1) by the University of California San Diego Center for Epigenomics. Libraries were sequenced to greater than 45 million reads. Quality control was performed with FastQC and MultiQC, and low-quality bases and adapters were trimmed with Trim Galore v0.64.1. Sequence reads were mapped to *ARS-UI_Ramb_v2.0* with Bowtie2 and subsequently, reads that mapped to the mitochondrial genome were removed with Samtools. Duplicates were then removed with Picard. Mapped files were indexed for visualization in IGV using Samtools. Peaks were called with MACS2 following ENCODE guidelines with -B -p 0.01 --nomodel --shift -75 --extsize 150 --SPMR flags (Gorkin et al., 2020; ENCODE Project Consortium, 2012). Peaks were annotated with the GTF file from the *ARS-UI_Ramb_v2.0* genome using the Homer *annotatePeaks.pl* function. ATAC-seq peaks were classified as near a TSS (+2kb to -2kb), Exonic, near a TTS (+1kb to -1kb), and Intergenic. Tissue similarities and differences between peaks were examined with BEDTools intersect. The dendrogram was created from raw ATAC-seq sequence signal using a Spearman correlation calculated with deepTools and displayed as a cluster dendrogram in the factoextra package of R v3.6.2 (Salavati et al., 2020).

The mapping and quality statistics of ATAC-seq is displayed in Supplementary Table 6.2. The overall average mapping percentage for ATAC-seq data was 95.78%, and the unique mapping percentage was 64.95% on average. The average proportion of duplicated reads across ATAC-seq sequence libraries was 0.19.

Whole Genome and Reduced Representation Bisulfite Sequencing

The DNA was extracted from flash frozen tissue (n=59 tissues, Supplementary Table 6.1) using phenol:chloroform:isoamyl alcohol following the protocol outlined in Salavati et al., 2020. The DNA was quantified with a Qubit dsDNA HS Assay Kit (Qiagen) and libraries were prepared by the Garvan Institute of Medical Research, Darlinghurst, Sydney, NSW, Australia. Bisulfite conversion was performed using the EZ DNA Methylation-Gold Kit (Zymo Research, CA, United States) following manufacturer's instructions, and then libraries were indexed with the Accel-NGS Methyl-seq DNA kit (Swift Biosciences, MI, United States) following manufacturer's instructions (Salavati et al., 2020). Libraries were sequenced 150bp paired end on an Illumina HiSeq X to achieve a minimum of 10x genome coverage for WGBS. Both WGBS and RRBS data were mapped to *ARS-UI_Ramb_v2.0*. Hypermethylated and hypomethylated regions were identified using the same methodology described in Salavati et al., 2020. Further details regarding these protocols are available at the FAANG Data Coordination Center (Salavati et al., 2020). The data processing, hierarchical clustering and dendrogram creation, and analyses were performed as described in Salavati et al., 2020 except the reads were mapped to the *ARS-UI_Ramb_v2.0* genome.

RNA Sequencing

RNA was isolated from snap frozen tissues by cryopulverizing tissue and placing in 2 mL of Trizol. The tissues were then homogenized in the Trizol and split into 1mL aliquots in 1.5 mL microcentrifuge. Samples were then centrifuged for 5 minutes at 14,000 x g and 4°C. The aqueous (clear) layer for each tissue was transferred to a new tube, avoiding any fat that may have risen to the top of the liquid. Exactly 400 µL of chloroform was added to each tube, shaken vigorously, and incubated at room temperature for 15 minutes. The samples were centrifuged for 15 minutes and 14,000 x g and 4°C, followed by transfer of the aqueous layer into another 1.5 mL microcentrifuge tube. 500 µL of isopropanol was then added to each tube and incubated for 10 minutes at room temperature, followed by centrifugation at 14,000 x g and 4°C for 30 minutes. The supernatant from each tube as discarded, and the pellet washed with 75% ethanol. The purified RNA from each tissue was resuspended in nuclease free water and quantified with a Nanodrop, as well as quality evaluated with a Fragment Analyzer. RNA was then treated with DNase I (1 µL of 2000U/mL per 10µL RNA) and cleaned up with the RNeasy MinElute spin column kit (Qiagen).

Sequencing libraries were prepared using the Illumina TruSeq Stranded mRNA Library Preparation Kit (Illumina, Inc.) according to manufacturer’s instructions. Libraries were sequenced on an Illumina NextSeq 500 to achieve at least 65 million reads. Further details and metadata for these samples are available in GenBank under BioProject PRJEB35292. The quality of RNA-seq libraries for each tissue was examined with FastQC and MultiQC, followed by trimming of adapters and low-quality sequence with Trim Galore v0.6.4. The RNA-seq data had an overall average percent of reads mapped of 96.06%, and 87.90% uniquely mapped reads. These libraries on average reached approximately 40 million reads each. Transcripts per million (TPM) counts were calculated with Kallisto using the *ARS-UI_Ramb_v2.0* genome and annotation.

Chromatin State Characterization

Chromatin states were defined using ChromHMM, which employs a Hidden Markov Model to assess histone modification signal enrichment overlap within a tissue (Ernst and Kellis 2010; Ernst and Kellis 2012; Ernst and Kellis, 2017; Gorkin et al., 2020). The optimal number of chromatin states was determined following Gorkin et al., 2020 by implementing ChromHMM with 2 through 16 chromatin states in the *LearnModel* function, followed by calculating the median Pearson correlation with the *CompareModels* function. The optimal number of states was determined at the point where each tissue plateaued and was tightly correlated with the model with the greatest number of states (Supplementary Figure 1) (Gorkin et al., 2020). The percent of the genome occupied by each chromatin state was averaged across tissues. Chromatin state location similarities and differences between tissues was examined with BEDTools intersect and displayed with an UpsetR plot (Quinlan et al., 2014; Conway et al., 2017).

Assay Integration

Overlay of ChIP-seq, ATAC-seq, DNA methylation, and RNA-seq was performed to define regulatory regions in the sheep genome. The overlap between ATAC-seq peaks and chromatin states for each tissue was determined with BEDTools intersect and averaged across tissues. The number of hypermethylated and hypomethylated regions in transcription start site and enhancer regions was evaluated with BEDTools intersect and the number of regions were quantified and averaged across tissues. The TPM counts from RNA-seq were overlayed with chromatin states to determine transcriptional activity across tissues.

Results

Defining the locations of genetic regulatory elements in the sheep genome was achieved by characterizing open chromatin, histone modifications, DNA methylation, and RNA expression across a large collection of tissues. These experiments were performed on tissues collected from the same animal used to assemble the *ARS-UI_Ramb_v2.0* genome and were accordingly mapped to this genome to provide a resource for annotation. The relationships between the locations of defined regulatory elements as well as regulatory regions across tissues were examined to explore gene regulation in sheep.

Hierarchical Clustering of Sequence Signal

Similarities between tissues were examined for RRBS, ATAC-seq, and ChIP-seq experiments with sequence signal from mapped reads using Spearman correlations and hierarchical clustering into dendrograms (Figure 6.1). The dendrogram displaying clustering of RRBS data (Figure 6.1A, generated by Alex Caulton) displays clear clustering by tissue type, including several branches with the ruminant stomach complex, intestinal tissues, immune-related tissues, muscle tissues, and nervous tissues clustering together. The ATAC-seq dendrogram (Figure 6.1B) also displays distinct clustering by tissue type including adrenal, nervous, GI, and muscle tissues. The ATAC-seq dendrogram also suggests similarity in open chromatin regions between GI tissues including ileum and Peyer's patch with immune related tissues including spleen and lung, with tonsil and spiral colon in a separate branch. The close clustering of GI and immune tissues is also present in the ChIP-seq dendrograms for H3K4me3, H3K27ac, H3K4me1, and H3K27me3 (Figure 6.1 C-F, respectively).

Muscle tissue and tissues that have a muscular layer also cluster together, including the left ventricle with skeletal muscle in the ATAC-seq dendrogram, abomasum and cecum with the biceps femoris in the H3K4me3 dendrogram, the rumen (ventral location) with the longissimus dorsi and diaphragm in the H3K27ac dendrogram, the biceps femoris and esophagus in the H3K4me1 dendrogram, and the rumen (ventral location) with the biceps femoris and rectum in the H3K27me3 dendrogram. This clustering suggests similarity of muscle tissue even across smooth, skeletal, and cardiac.

ATAC-seq and ChIP-seq Peak Annotation

The annotation file from ARS-UI_Ramb_v2.0 (https://www.ncbi.nlm.nih.gov/genome/annotation_euk/Ovis_aries/104/) retrieved from NCBI was used to determine the proximity of ATAC-seq and ChIP-seq peaks to genes and gene features including promoters/transcription start sites (TSS), exons, and transcription termination sites (TTS) (Figure 6.2). A mean of 28% of ATAC-seq peaks were identified within 2 kb from a promoter/TSS, 4% of peaks were in an exon, 9% of peaks were within 2kb from a TTS, and 59% were not located within 2 kb these features and considered intergenic on average across tissues (Figure 6.2A). The number of ATAC-seq peaks in close proximity to genes was similar across tissues. The greatest percent of H3K4me3 peaks were near promoters/TSS locations (mean of 41%), followed by intergenic areas (mean of 36%), TTS (mean of 20%), and within exons (mean of 3%) (Figure 6.2B). The number of H3K4me3 peaks near genes varied across tissues with many of the GI tissues having a higher percent of peaks in intergenic regions. The greatest percent of H3K27ac peaks were located in intergenic regions (mean of 61%), followed by promoter/TSS (mean of 27%), TTS (mean of 10%), and exonic (mean of 2%). The greatest percent of H3K4me1 peaks were also located in intergenic regions (mean of 72%), followed by promoter/TSS (mean of 19%), TTS (mean of 7%), and exon (mean of 2%). The greatest percent of the H3K27me3 peaks was in intergenic regions (mean of 88%), followed by promoter/TSS (mean of 8%), TTS (mean of 3%), and exons (mean of 1%). The percent of peaks in close proximity to genes was consistent across tissues for H3K4me1 and H3K27me3.

Chromatin State Characterization

Chromatin states were defined with a Hidden Markov Model in ChromHMM to characterize overlap between histone modifications which indicate the presence of promoter, enhancer, and repressor states throughout the genome (Figure 6.3). The optimal number of states was determined to be 9 based on the median correlation of each model of 2 through 16 states to the largest number of states (Roadmap Epigenomics Consortium et al., 2015; Gorkin et al., 2020). The 9 states define specific regulatory element locations throughout the genome with minimal redundancy.

The regulatory elements represented in each chromatin state were defined previously by the Roadmap Epigenomics Consortium based on the presence of sequence signal from each histone modification (Roadmap Epigenomics Consortium et al., 2015). The active TSS states were defined by H3K4me3 in combination with the presence of signal from H3K27ac, H3K4me1, and H3K27me3 histone modifications, active enhancer states were defined with primarily H3K27ac signal, poised enhancer states were defined with primarily H3K4me1 signal, and repressed enhancer states were defined with H3K27me3 signal (Roadmap Epigenomics Consortium et al., 2015; Gorkin et al., 2020). States with very low or absent signal were considered quiescent (Roadmap Epigenomics Consortium et al., 2015; Gorkin et al., 2020). The chromatin states defined as active, poised, and repressed in this study represent approximately 13% of the genome (Figure 6.3C). Almost 1% of the genome was occupied by active TSS and flanking active TSS, while only approximately 0.25% of the genome was occupied by bivalent flanking TSS. Active enhancers occupied approximately 1.75% of the genome, poised enhancers occupied approximately 2.5% of the genome, and repressed enhancers occupied approximately 2.25% of the genome. The weak repressor state had the greatest genome occupancy of the regulatory states (just over 3% of the genome).

Chromatin State Overlap with Open Chromatin

The overlap of open chromatin regions defined by ATAC-seq peaks and chromatin states was performed by examining common peak locations and calculating the percent of each chromatin state that resides in an ATAC-seq peak across tissues (Figure 6.4). This analysis provided further insight into the activity and characteristics of each chromatin state. The ATAC-seq data resulted in an average of 176,864 peaks across tissues with an average width of 536 bp. The greatest number of ATAC-seq peaks (n=239,409) was identified in Peyer’s patch tissue, while the lowest number of peaks (n=105,845) was identified in biceps femoris tissue. The widest peak (705 bp) was identified in heart left ventricle tissue, while the narrowest peak (364 bp) was identified in mesenteric lymph node tissue.

The chromatin states exhibiting the highest percent overlap with ATAC-seq include TssA (mean of 86.24% across tissues) and TssAFlnk (mean of 77.86% across tissues), which denotes active transcription start sites, as well as EnhA (mean of 69.17% across tissues) which denotes active enhancer states. Cerebellum tissue had the greatest overlap of TssA states with open chromatin (99.66%), whereas reticulum tissue had the least overlap (71.17%). A greater percent of the TssAFlnk chromatin state from gallbladder tissue resided in open chromatin regions (89.25%), while jejunum had the least percent of TssAFlnk states in open chromatin (44.08%). The EnhA state had the greatest percent of that state in open chromatin in biceps femoris muscle tissue (92.17%), while the lowest was in cerebellum tissue (40.51%). The BivFlnk state, which represents a transcription start site that exists in a bivalent state and includes the presence of the repressor mark H3K27me3 along with the active promoter mark of H3K4me3, has less overlap (mean of 60.37% across tissues) with ATAC-seq than other TSS-related states. The tissue with the least BivFlnk chromatin state overlay with open chromatin was kidney medulla (38.93%), while the greatest overlap was diaphragm tissue (81.03%).

Poised enhancers also have fewer states that overlap with ATAC-seq (mean of 40.98% across tissues). The greatest percent of EnhP chromatin states that overlapped with open chromatin were from ileum tissue (67.24%) while the lowest percent of EnhP chromatin state overlap were from spleen tissue (16.39%). Repressed polycomb states had the least overlap with ATAC-seq (mean of 17.54% across tissues for ReprPC and mean of 19.26% across tissues for ReprPCWk). The lowest percent of ReprPC chromatin state overlap with open chromatin was in biceps femoris muscle tissue (0.69%), while the greatest overlap was with duodenum tissue (36.56%). The ReprPCWk chromatin state had the greatest overlap with open chromatin in adrenal medulla tissue (37.87%) and the least overlap in longissimus dorsi muscle tissue (12.93%). An average of 71.96% of the quiescent (Quies) states overlapped with ATAC-seq peaks. The greatest overlap of the Quies state with open chromatin was in lung tissue (87.57%) and the lowest overlap was in cerebral cortex tissue (48.68%). The active chromatin states that displayed histone modifications H3K4me3 and H3K27ac signal resided in open chromatin, while repressed chromatin states that displayed H3K27me3 signal did not reside in open chromatin regions.

Chromatin State and ATAC-seq Overlap with Hypermethylated and Hypomethylated Regions

The number of hypermethylated and hypomethylated sites identified in the genome from the RRBS data was similar across tissues. Hypermethylated sites ranged from 18,744 sites in jejunum tissue to 27,327 sites in kidney medulla and averaged 21,565 sites across tissues. Hypomethylated sites averaged 20,451 across tissues and ranged from 18,679 sites in omasum tissue to 22,705 sites in ovary tissue.

Chromatin state locations were compared with hypermethylated (Figure 6.5A) and hypomethylated (Figure 6.5B) sites throughout the genome (Figure 6.5). The chromatin states with the greatest number of hypermethylated regions include the repressed polycomb states (ReprPC, 545 sites and ReprPCWk, 860 sites) as well as the poised enhancer state (EnhP, 634 sites, Figure 6.5A). Fewer hypermethylated regions (240 sites) resided in active enhancer states (EnhA) when compared with other enhancer states, and the least number of hypermethylated regions were identified in the active promoter/TSS states (TssA, 132 sites and TssAFlnk, 116 sites) (Figure 6.5A). The bivalent/flanking transcription start site state (BivFlnk) displayed 448 hypermethylated sites, which was greater than other promoter states. The greatest number of hypomethylated sites was observed in active TSSs (TssA, 6,239 sites), followed by active enhancer states (EnhA, 4,367 sites) and flanking active TSSs (TssAFlnk, 4,164 sites). The bivalent flanking TSS state exhibited the fewest number of hypomethylated regions that resided in that state (663 sites). The poised enhancer (EnhP) state as well as the repressed enhancer states (ReprPC and ReprPCWk) displayed fewer hypomethylated sites (2,017, 1,178, and 2,133 sites, respectively) when compared with other chromatin states.

The number of hypermethylated and hypomethylated sites that resided within open chromatin were also quantified across tissues (Figure 6.5C). The number of hypomethylated sites that resided in ATAC-seq peaks was greater than the number of hypermethylated sites that overlapped by almost four-fold. The average number of overlapping hypermethylated sites was 5,297, which was approximately 24.74% of the total number of hypermethylated sites identified in the genome across tissues. The number of hypomethylated sites averaged 16,461 which was approximately 80.60% of the total number of hypomethylated sites identified across tissues. Some tissues had a greater number of both hypermethylated and hypomethylated sites that resided in ATAC-seq peaks, such as the rectum (16,729 hypermethylated and 12,088 hypomethylated sites). Ovary had very few hypermethylated sites (18 hypermethylated sites) that resided in open chromatin compared to hypomethylated sites (19,500 hypomethylated sites) in open chromatin. This information suggested that some tissue specificity was observed, and highly methylated regions existed outside of open chromatin, and lowly methylated regions resided within open chromatin regions.

Chromatin State and ATAC-seq Overlap with Transcript Expression

The expression of RNA transcripts that occurred in each promoter and enhancer state as well as within ATAC-seq peaks was determined by quantifying TPM in each tissue (Figure 6.6). The active promoter states (TssA and TssAFlnk) had the most TPM counts within these states across tissues with averages of 472,377 and 366,054, respectively (Figure 6.6A). Bladder and oviduct tissues had the greatest numbers of TPM (665,421 and 634,382 TPM) in TssA and TssAFlnk states, respectively. The lowest numbers of TPM in TssA and TssAFlnk were 230,479 and 136,790 TPM in cerebellum and tongue tissues, respectively. The active enhancer state (EnhA) had the next largest TPM counts within this state with an average of 300,932, followed by the poised enhancer state (EnhP) with an average of 170,884 TPM (Figure 6.6A). The EnhA state had the greatest TPM overlay of 452,994 in lung tissue, and the least TPM overlay of 131,507 in jejunum tissue. The greatest number of TPM in the EnhP state (348,113) was in skin tissue, while the lowest number of TPM (17,925) was in descending colon tissue. The repressed enhancer states (ReprPC and

ReprPCWk) had the lowest TPM counts and were over ten-fold less than active promoter states with averages of 20,092 and 30,229 TPM, respectively (Figure 6.6A). The greatest number of TPM in ReprPC and ReprPCWk chromatin states was 54,710 TPM in descending colon and 110,532 TPM in rumen atrium, respectively, while the lowest TPM in these states was 554 TPM in omasum and 682 TPM in semimembranosus muscle, respectively.

The open chromatin regions designated by ATAC-seq peaks had greater transcript expression across tissues than regions with no ATAC-seq peaks (Figure 6.6B). Regions with ATAC-seq peaks displayed 683,771 TPM on average across tissues, while regions without ATAC-seq peaks displayed 49,271 TPM tissue average. The percent of the total number of TPM that resided in open chromatin regions was 92.92% on average across tissues, whereas a tissue average of 7.08% of the total TPM resided in regions outside of ATAC-seq peaks. The greatest TPM overlay within ATAC-seq peaks was 848,414 TPM in spleen tissue, and the lowest TPM overlay was 320,053 TPM in gallbladder tissue. The regions without ATAC-seq peaks displayed the least overlay of 19,804 TPM in heart left ventricle tissue and the greatest overlay of 100,556 TPM in ileum tissue. Regions defined as active promoter and enhancer states, as well as open chromatin, had greater overall transcript expression than regions defined as repressed and devoid of ATAC-seq peaks.

Tissue Comparisons of Chromatin States

The chromatin states were then compared across tissues to examine potential similarities and differences in the location of these regulatory regions (Figure 6.7). Active promoter states (Tss and TssA) demonstrated the greatest similarity between muscular tissues such as the diaphragm, biceps femoris, and GI tract with muscular layers such as the spiral colon (Figure 6.7A). Brain tissues cerebellum and cerebral cortex showed similar active promoter states as well. Tissues that had the least number of active promoter states in common were cerebral cortex and diaphragm. The active enhancer states (EnhA) displayed the greatest similarity between abomasum and abomasum pylorus, ileum and Peyer's patch, and longissimus dorsi and semimembranosus (Figure 6.7B). The tissues that displayed the least similarity of active enhancer states with other tissues were skin and ovary.

Poised enhancer states showed the greatest similarity between tissues with muscular layers (omasum and longissimus dorsi), skeletal muscle (longissimus dorsi, semimembranosus, and biceps femoris), and GI tissues (ileum, Peyer’s patch, and jejunum) (Figure 6.7C). Brain tissues also showed great similarity of poised enhancer states. The tissues that had the least number of poised enhancer states in common were alveolar macrophages, ovary, and bladder. Repressed enhancers showed a similar pattern to poised enhancers regarding chromatin state similarity (Figure 6.7D). Abomasum pylorus and abomasum, all three skeletal muscles, ileum and Peyer’s patch, and brain tissues showed the greatest similarity in repressed enhancer states. The tissues with the least repressed enhancer states in common with other tissues were vagina, parathyroid, and uterus.

A pattern observed most prominently in the promoter and active enhancer chromatin state tissue comparison was the overlap between GI, brain, and immune tissues (Figure 6.8). The ileum and Peyer’s patch tissues showed 84% and 88% similarity of promoter and active enhancer states to each other, respectively. These two tissues also demonstrated over 60% similarity to spleen, mesenteric lymph node, and tonsil in promoters and active enhancers. The jejunum and descending colon also showed similarity to immune-related tissues in promoter and active enhancer states, along with spiral colon in active enhancer states. The cerebral cortex and cerebellum tissues showed great similarity to each other (95% in promoters and 81% in active enhancers). Cerebral cortex has 67% and 63% of active enhancers in common with descending colon and lymph node mesenteric, respectively. The relationship between the gut, brain, and immune system was further examined by visualization of sequence signal tracks in the Integrative Genomics Viewer (IGV). The genes encoding IL-10 and one of its receptors were chosen as an example because these genes are involved in gut/brain/immune signaling. An example of *IL-10* and *IL-10RA* signal tracks is displayed in Figure 6.9. Promoter H3K4me3 signal was apparent for ileum and lymph node, as well as ATAC-seq signal in lymph node (Figure 6.9A). The *IL-10* receptor alpha was expressed in cerebellum as shown with RNA-seq tracks and H3K4me3 signal directly upstream of the 5’ end and within this gene (6.9B).

Discussion

This study identified key regulatory elements in the sheep genome, specifically promoters and enhancers across almost 50 tissues that represent all three developmental tissue types. The identification of enhancers is more challenging and requires additional data as described here although promoters might be readily identified as sequences within 2 kb of transcription start sites identified by RNA-seq data alone. We have been able to identify regulatory sequences that are active across many tissues, active in some tissues and not others, and those in a poised state.

The data were processed through quality control metrics, which met standards previously set by the ENCODE consortium (ENCODE Project Consortium et al., 2012). The peaks called for ATAC-seq open chromatin regions and histone modifications H3K4me3, H3K27ac, H3K4me1, and H3K27me3 were compared with gene locations as annotated by NCBI on the *ARS-UI_Ramb_v2.0* genome assembly. The ATAC-seq peaks resided in mostly intergenic locations, and the percent of total peaks near (+/- 2kb) promoter/TSS, within exons, and near (+/- 2kb) TTS was similar across tissue. Histone modification H3K4me3 annotations had a greater percent of peaks near promoter/TSS and TTS regions, however this value was not consistent across tissues. Fewer peaks are called for H3K4me3 (approximately 15,000) compared with ATAC-seq (over 100,000) and H3K4me3 is more closely related to tissue specific gene expression (Tian et al., 2011; Zhang & Zhang, 2011; Benayoun et al., 2015; Ishibashi et al., 2021), this may contribute to the differences in annotation. The differences between tissues may be indicative of tissue specific expression and regulation, in which some tissues may have a greater number of active regulatory elements near genes than others. The H3K27ac histone modification also displayed variation across tissues in the percent of peaks in proximity to or within genic elements (TSS, TTS, and exons). Fewer H3K4me1 and H3K27me3 peaks were identified near TSS and TTS and within genes compared to H3K4me3 and H3K27ac, however the percent of peaks in proximity to or that directly overlay these elements were more consistent across tissues. This suggested that histone modifications that mark the presence of active regulatory elements may have been more related to tissue specific gene activation and expression than poised or repressed regulatory elements.

The number of chromatin states assessed using H3K4me3, H3K27ac, H3K4me1, and H3K27me3 revealed 9 states which were consistent with the states reported in previous studies in sheep (Davenport et al., 2021), cattle (Fang et al., 2019; Kern et al., 2021), and humans (Gorkin et al., 2020). The active chromatin states, including TssA, TssAFlnk, and EnhA, resided in open chromatin regions which indicated that these regions were available for transcription. Hypomethylated regions were also found in active promoter and enhancer states, whereas hypermethylated regions resided in poised and repressed enhancer states. This study found that most of the hypomethylated sites resided in open chromatin regions as defined by ATAC-seq peaks. These active, poised, and repressed regulatory locations across different data types agreed with previous research in humans, mice, and other mammals (Aran and Hellman, 2013; Barwick et al., 2016; Bell and Vertino, 2017; Sharifi-Zarchi et al., 2017; Davenport et al., 2021).

The hypermethylated sites were more frequently located in the repressor polycomb and poised enhancer states, which agreed with previous studies in humans, mice, and sheep (Zhang et al., 2009; Teng and Tan, 2012; Sharifi-Zarchi et al., 2017; Davenport et al., 2021). DNA methylation has been found to be inversely correlated with active enhancer activity (Aran and Hellman, 2013; Barwock et al., 2016; Bell and Vertino, 2017). The H3K4me1 histone modification which is indicative of poised enhancer regions has been found to be positively correlated with DNA methylation in mice (Zhang et al., 2009; Teng and Tan, 2012; Sharifi-Zarchi et al., 2017). DNA methylation in poised enhancer regions is hypothesized to play a role in enhancer switching between active and repressed states (Teng and Tan, 2012; Sharifi-Zarchi et al., 2017).

This study identified a greater number of hypermethylated sites in open chromatin regions in the rectum when compared with other tissues. Mis-regulation of DNA methylation in the rectum and colon have been linked to colorectal cancers in humans, specifically increased methylation in tumors compared with non-cancerous rectal tissues (Molinari et al., 2013; Kaz et al., 2014; Exner et al., 2015). The ovary conversely had very few hypermethylated sites identified in open chromatin regions. Previous studies in mice and pigs have found that global DNA methylation decreased in ovarian tissue with advancing maternal age which is also associated with decreased fertility (Xi et al., 2019; Uysal & Ozturk, 2020). There may have been less methylation in the ovary because of the age of the ewe, especially in locations that overlapped with open chromatin. Previous studies in humans, mice, and cattle have also shown a difference in methylation across tissues and during the aging process (Bjornsson et al., 2008; Zhang et al., 2013; Bell et al., 2019; Zhou et al., 2020). These tissue specific differences have likely influenced the variation in hypermethylated sites in open chromatin observed in this study.

Greater gene transcription was identified in active and open chromatin states across tissues. The highest TPM counts were found in active chromatin states and open chromatin, which indicated that the histone modifications and open chromatin regions were indicative of transcript expression in tissues. The repressor regions and regions without ATAC-seq peaks had very few TPM counts, which indicated that transcript expression is indeed repressed in these regions. The bivalent flanking promoter state (BivFlnk) also had very little transcript expression. This lack of transcript expression in bivalent promoters was also seen in studies with tissues and cell lines from mice and humans (Stergachis et al., 2014; Kinkley et al., 2016; Yan et al., 2016; Mas et al., 2018; van der Velde et al., 2021). Promoter bivalency was hypothesized to be important during both development and adult stages in mammals as some genes were only actively transcribed during specific developmental states and in response to environmental signaling transcription of some genes were activated or repressed based on developmental and environmental signaling (Stergachis et al., 2014; Kinkley et al., 2016; Yan et al., 2016; Mas et al., 2018; van der Velde et al., 2021).

Data from complimentary assays and across tissues in this study provided a comprehensive collection of regulatory element locations across tissues to inform the functional annotation of the reference Rambouillet sheep genome. The overall mapped reads and uniquely mapped reads percentage were both improved in this study which used the *ARS-UI_Ramb_v2.0* genome (Massa et al., 2021; Davenport et al., 2021). The clustering of sequence signal displayed resemblance between tissues with similar functions and developmental lineages, such as the gastrointestinal tract. This hierarchical clustering of histone modification and open chromatin signal also implied a close relationship between the GI and immune tissues, especially the ileum and Peyer’s patch with lung, spleen, and mesenteric lymph node.

Shared regulatory elements in the GI and immune tissues are also implied in chromatin state similarities and differences across tissues. These relationships are especially apparent in the promoter and active enhancer states. Chromatin states in the brain also had shared regulatory elements with GI and immune tissues in the active enhancer state. The similarity of active regulatory elements across GI, brain, and immune related tissues alludes to the important physiological feedback mechanism of the gut/brain/immune axis (Rutsch et al., 2020; Gwak & Chang, 2021). The gut and the brain communicate in multiple ways including the autonomic and enteric nervous system, endocrine system, hypothalamic-pituitary-adrenal axis, microbiota metabolites, and the immune system (Rutsch et al., 2020). Much of the communication between the gut, brain, and immune system has not been well understood. Feedback between the central nervous system and the gut has been shown in previous studies with microbial metabolites and immune cells and their products that have passed the blood brain barrier in both normal and disease states (Kipnis et al., 2012; Ellwardt et al., 2016; Engelhardt et al., 2016; Kipnis, 2016; Inserra et al., 2018 Negi & Das, 2018; Rutsch et al., 2020).

Communication between the gut, brain, and immune systems has been required for homeostasis and signaling of a disease state in the body, in which pro-inflammatory and anti-inflammatory cytokines have played a major role (Iyer & Cheng, 2012; Rutsch et al., 2020; Wei et al., 2020; Jacobson et al., 2021). An example of an anti-inflammatory cytokine involved with gut mucosal homeostasis and maintenance of neuronal cells including microglia is interleukin 10 (*IL-10*). The *IL-10* gene is known to be involved with innate and adaptive immune response and signaling in the GI tract (Franke et al., 2008; Jostins et al., 2012; Shouval et al., 2014). This gene has been studied in relation to inflammatory bowel disease (IBD), in which genetic variation in both *IL-10* and *IL-10* receptor alpha was associated with IBD risk in humans (Franke et al., 2008; Glocker et al., 2009; Jostins et al., 2012; Kotlarz et al., 2012; Moran et al., 2013). Tissue of the GI tract has been shown to express *IL-10* in addition to immune cells such as macrophages (Autschbach et al., 1998; Rutsch et al., 2020; Wei et al., 2020; Jacobson et al., 2021). The gene *IL-10* and its receptor were also expressed in neurons such as microglia in the brain as part of maintenance of homeostasis in the central nervous system (Lobo-Silva et al., 2016; Burmeister & Marriott, 2018; Kathrin Uhde et al., 2018). The Th2 immune response involved *IL-10* and was previously shown to be involved with parasite resistance in humans and sheep (Schopf et al., 2002; Gautam et al., 2011; Shouval et al., 2014; Becker et al., 2020). Parasite resistance has been a trait selected for in sheep by producers across the world. The *IL-10* gene and its receptor were therefore chosen as an example to demonstrate expression and regulatory element presence in a gene known to be involved with the communication between the gut, brain, and immune system.

Defining genetic regulatory element locations in the sheep genome across a large collection of tissues provided an important resource for the scientific community. This study provided an important first step in providing the annotation of regulatory elements in ovine tissues. Sheep have been an important species raised for food, fiber, and milk across the world, and defining these regulatory elements will better equip researchers to understand biological mechanisms that influence economically important traits such as growth, wool quality, milk production, and disease resistance.

Funding

This material is based upon work that is supported by the National Institute of Food and Agriculture, U.S. Department of Agriculture, award number USDA-NIFA-2017-67016-26301. ALA and ELC were supported by BBSRC grant BB/S02008X/1 and ALA, ELC and MS by BBSRC Institute Strategic Programme grant BBS/E/D/10002070.

Acknowledgements

The authors would like to thank Codie Durfee and Kaneesha Hemmerling for assistance with tissue collection, Jacky Carnahan, Kristen Kuhn, and Kelsey McClure for assistance with sequencing, and Dr. Mingrui Duan for assistance with the bioinformatics analysis pipeline.

Members of the Ovine FAANG Project Consortium (listed by institution)

Brenda Murdoch (University of Idaho)

Kimberly Davenport (University of Idaho)

Stephen White (USDA, ARS, ADRU, Washington State University)

Michelle Mousel (USDA, ARS, ADRU, Washington State University)

Alisha Massa (Washington State University)

Kim Worley (Baylor College of Medicine)

Alan Archibald (The Roslin Institute, University of Edinburgh)

Emily Clark (The Roslin Institute, University of Edinburgh)

Mazdak Salavati (The Roslin Institute, University of Edinburgh)

Brian Dalrymple (University of Western Australia)

James Kijas (CSIRO)

Shannon Clarke (AgResearch)

Alex Caulton (AgResearch)

Rudiger Brauning (AgReseach)

Timothy Smith (USDA, ARS, MARC)

Tracey Hadfield (Utah State University)

Noelle Cockett (Utah State University)

References

1. Afgan E, Baker D, Batut B, van den Beek M, Bouvier D, Cech M, Chilton J, Clements D, Coraor N, Grüning BA, Guerler A, Hillman-Jackson J, Hiltemann S, Jalili V, Rasche H, Soranzo N, Goecks J, Taylor J, Nekrutenko A, Blankenberg D. The Galaxy platform for accessible, reproducible and collaborative biomedical analyses: 2018 update. Nucleic Acids Res. 2018 Jul 2;46(W1):W537-W544. doi: 10.1093/nar/gky379.

2. Andersson L, Archibald AL, Bottema CD, Brauning R, Burgess SC, Burt DW, Casas E, Cheng HH, Clarke L, Couldrey C, Dalrymple BP, Elsik CG, Foissac S, Giuffra E, Groenen MA, Hayes BJ, Huang LS, Khatib H, Kijas JW, Kim H, Lunney JK, McCarthy FM, McEwan JC, Moore S, Nanduri B, Notredame C, Palti Y, Plastow GS, Reecy JM, Rohrer GA, Sarropoulou E, Schmidt CJ, Silverstein J, Tellam RL, Tixier-Boichard M, Tosser-Klopp G, Tuggle CK, Vilkki J, White SN, Zhao S, Zhou H; FAANG Consortium. Coordinated international action to accelerate genome-to-phenome with FAANG, the Functional Annotation of Animal Genomes project. Genome Biol. 2015 Mar 25;16(1):57. doi: 10.1186/s13059-015-0622-4.

3. Aran D, Hellman A. DNA methylation of transcriptional enhancers and cancer predisposition. Cell. 2013 Jul 3;154(1):11-3. doi: 10.1016/j.cell.2013.06.018.

4. Autschbach F, Braunstein J, Helmke B, Zuna I, Schürmann G, Niemir ZI, Wallich R, Otto HF, Meuer SC. In situ expression of interleukin-10 in noninflamed human gut and in inflammatory bowel disease. Am J Pathol. 1998 Jul;153(1):121-30. doi: 10.1016/S0002-9440(10)65552-6.

5. Barbosa-Morais NL, Irimia M, Pan Q, Xiong HY, Gueroussov S, Lee LJ, Slobodeniuc V, Kutter C, Watt S, Colak R, Kim T, Misquitta-Ali CM, Wilson MD, Kim PM, Odom DT, Frey BJ, Blencowe BJ. The evolutionary landscape of alternative splicing in vertebrate species. Science. 2012 Dec 21;338(6114):1587-93. doi: 10.1126/science.1230612.

6. Barski A, Cuddapah S, Cui K, Roh TY, Schones DE, Wang Z, Wei G, Chepelev I, Zhao K. High-resolution profiling of histone methylations in the human genome. Cell. 2007 May 18;129(4):823-37. doi: 10.1016/j.cell.2007.05.009.

7. Barwick BG, Scharer CD, Bally APR, Boss JM. Plasma cell differentiation is coupled to division-dependent DNA hypomethylation and gene regulation. Nat Immunol. 2016 Oct;17(10):1216-1225. doi: 10.1038/ni.3519.

8. Becker GM, Davenport KM, Burke JM, Lewis RM, Miller JE, Morgan JLM, Notter DR, Murdoch BM. Genome-wide association study to identify genetic loci associated with gastrointestinal nematode resistance in Katahdin sheep. Anim Genet. 2020 Mar;51(2):330-335. doi: 10.1111/age.12895.

9. Bell CG, Lowe R, Adams PD, Baccarelli AA, Beck S, Bell JT, Christensen BC, Gladyshev VN, Heijmans BT, Horvath S, Ideker T, Issa JJ, Kelsey KT, Marioni RE, Reik W, Relton CL, Schalkwyk LC, Teschendorff AE, Wagner W, Zhang K, Rakyan VK. DNA methylation aging clocks: challenges and recommendations. Genome Biol. 2019 Nov 25;20(1):249. doi: 10.1186/s13059-019-1824-y.

10. Bell JSK, Vertino PM. Orphan CpG islands define a novel class of highly active enhancers. Epigenetics. 2017 Jun 3;12(6):449-464. doi: 10.1080/15592294.2017.1297910.

11. Benayoun BA, Pollina EA, Ucar D, Mahmoudi S, Karra K, Wong ED, Devarajan K, Daugherty AC, Kundaje AB, Mancini E, Hitz BC, Gupta R, Rando TA, Baker JC, Snyder MP, Cherry JM, Brunet A. H3K4me3 breadth is linked to cell identity and transcriptional consistency. Cell. 2014 Jul 31;158(3):673-88. doi: 10.1016/j.cell.2014.06.027.

12. Bjornsson HT, Sigurdsson MI, Fallin MD, Irizarry RA, Aspelund T, Cui H, Yu W, Rongione MA, Ekström TJ, Harris TB, Launer LJ, Eiriksdottir G, Leppert MF, Sapienza C, Gudnason V, Feinberg AP. Intra-individual change over time in DNA methylation with familial clustering. JAMA. 2008 Jun 25;299(24):2877-83. doi: 10.1001/jama.299.24.2877.

13. Burmeister AR, Marriott I. The Interleukin-10 Family of Cytokines and Their Role in the CNS. Front Cell Neurosci. 2018 Nov 27;12:458. doi: 10.3389/fncel.2018.00458.

14. Clark EL, Bush SJ, McCulloch MEB, Farquhar IL, Young R, Lefevre L, Pridans C, Tsang HG, Wu C, Afrasiabi C, Watson M, Whitelaw CB, Freeman TC, Summers KM, Archibald AL, Hume DA. A high resolution atlas of gene expression in the domestic sheep (Ovis aries). PLoS Genet. 2017 Sep 15;13(9):e1006997. doi: 10.1371/journal.pgen.1006997.

15. Conway JR, Lex A, Gehlenborg N. UpSetR: an R package for the visualization of intersecting sets and their properties. Bioinformatics. 2017 Sep 15;33(18):2938-2940. doi: 10.1093/bioinformatics/btx364.

16. Cordier G, Cozon G, Greenland T, Rocher F, Guiguen F, Guerret S, Brune J, Mornex JF. In vivo activation of alveolar macrophages in ovine lentivirus infection. Clin Immunol Immunopathol. 1990 Jun;55(3):355-67. doi: 10.1016/0090-1229(90)90124-9.

17. Davenport KM, Massa AT, Bhattarai S, McKay SD, Mousel MR, Herndon MK, White SN, Cockett NE, Smith TPL, Murdoch BM; Ovine FAANG Project Consortium. Characterizing Genetic Regulatory Elements in Ovine Tissues. Front Genet. 2021 May 20;12:628849. doi: 10.3389/fgene.2021.628849.

18. Ellwardt E, Walsh JT, Kipnis J, Zipp F. Understanding the Role of T Cells in CNS Homeostasis. Trends Immunol. 2016 Feb;37(2):154-165. doi: 10.1016/j.it.2015.12.008.

19. Engelhardt B, Carare RO, Bechmann I, Flügel A, Laman JD, Weller RO. Vascular, glial, and lymphatic immune gateways of the central nervous system. Acta Neuropathol. 2016 Sep;132(3):317-38. doi: 10.1007/s00401-016-1606-5.

20. ENCODE Project Consortium. An integrated encyclopedia of DNA elements in the human genome. Nature. 2012 Sep 6;489(7414):57-74. doi: 10.1038/nature11247.

21. Ernst J, Kellis M. Discovery and characterization of chromatin states for systematic annotation of the human genome. Nat Biotechnol. 2010 Aug;28(8):817-25. doi: 10.1038/nbt.1662.

22. Ernst J, Kellis M. ChromHMM: automating chromatin-state discovery and characterization. Nat Methods. 2012 Feb 28;9(3):215-6. doi: 10.1038/nmeth.1906.

23. Ernst J, Kellis M. Chromatin-state discovery and genome annotation with ChromHMM. Nat Protoc. 2017 Dec;12(12):2478-2492. doi: 10.1038/nprot.2017.124.

24. Exner R, Pulverer W, Diem M, Spaller L, Woltering L, Schreiber M, Wolf B, Sonntagbauer M, Schröder F, Stift J, Wrba F, Bergmann M, Weinhäusel A, Egger G. Potential of DNA methylation in rectal cancer as diagnostic and prognostic biomarkers. Br J Cancer. 2015 Sep 29;113(7):1035-45. doi: 10.1038/bjc.2015.303.

25. Fang L, Liu S, Liu M, Kang X, Lin S, Li B, Connor EE, Baldwin RL 6th, Tenesa A, Ma L, Liu GE, Li CJ. Functional annotation of the cattle genome through systematic discovery and characterization of chromatin states and butyrate-induced variations. BMC Biol. 2019 Aug 16;17(1):68. doi: 10.1186/s12915-019-0687-8.

26. Feng J, Liu T, Qin B, Zhang Y, Liu XS. Identifying ChIP-seq enrichment using MACS. Nat Protoc. 2012 Sep;7(9):1728-40. doi: 10.1038/nprot.2012.101.

27. Franke A, Balschun T, Karlsen TH, Sventoraityte J, Nikolaus S, Mayr G, Domingues FS, Albrecht M, Nothnagel M, Ellinghaus D, Sina C, Onnie CM, Weersma RK, Stokkers PC, Wijmenga C, Gazouli M, Strachan D, McArdle WL, Vermeire S, Rutgeerts P, Rosenstiel P, Krawczak M, Vatn MH; IBSEN study group, Mathew CG, Schreiber S. Sequence variants in IL10, ARPC2 and multiple other loci contribute to ulcerative colitis susceptibility. Nat Genet. 2008 Nov;40(11):1319-23. doi: 10.1038/ng.221.

28. Friedman J, Alm EJ. Inferring correlation networks from genomic survey data. PLoS Comput Biol. 2012;8(9):e1002687. doi: 10.1371/journal.pcbi.1002687.

29. Gautam S, Kumar R, Maurya R, Nylén S, Ansari N, Rai M, Sundar S, Sacks D. IL-10 neutralization promotes parasite clearance in splenic aspirate cells from patients with visceral leishmaniasis. J Infect Dis. 2011 Oct 1;204(7):1134-7. doi: 10.1093/infdis/jir461.

30. Gerstein MB, Lu ZJ, Van Nostrand EL, Cheng C, Arshinoff BI, Liu T, Yip KY, Robilotto R, Rechtsteiner A, Ikegami K, Alves P, Chateigner A, Perry M, Morris M, Auerbach RK, Feng X, Leng J, Vielle A, Niu W, Rhrissorrakrai K, Agarwal A, Alexander RP, Barber G, Brdlik CM, Brennan J, Brouillet JJ, Carr A, Cheung MS, Clawson H, Contrino S, Dannenberg LO, Dernburg AF, Desai A, Dick L, Dosé AC, Du J, Egelhofer T, Ercan S, Euskirchen G, Ewing B, Feingold EA, Gassmann R, Good PJ, Green P, Gullier F, Gutwein M, Guyer MS, Habegger L, Han T, Henikoff JG, Henz SR, Hinrichs A, Holster H, Hyman T, Iniguez AL, Janette J, Jensen M, Kato M, Kent WJ, Kephart E, Khivansara V, Khurana E, Kim JK, Kolasinska-Zwierz P, Lai EC, Latorre I, Leahey A, Lewis S, Lloyd P, Lochovsky L, Lowdon RF,

Lubling Y, Lyne R, MacCoss M, Mackowiak SD, Mangone M, McKay S, Mecenas D, Merrihew G, Miller DM 3rd, Muroyama A, Murray JI, Ooi SL, Pham H, Phippen T, Preston EA, Rajewsky N, Rätsch G, Rosenbaum H, Rozowsky J, Rutherford K, Ruzanov P, Sarov M, Sasidharan R, Sboner A, Scheid P, Segal E, Shin H, Shou C, Slack FJ, Slightam C, Smith R, Spencer WC, Stinson EO, Taing S, Takasaki T, Vafeados D, Voronina K, Wang G, Washington NL, Whittle CM, Wu B, Yan KK, Zeller G, Zha Z, Zhong M, Zhou X; modENCODE Consortium, Ahringer J, Strome S, Gunsalus KC, Micklem G, Liu XS, Reinke V, Kim SK, Hillier LW, Henikoff S, Piano F, Snyder M, Stein L, Lieb JD, Waterston RH. Integrative analysis of the Caenorhabditis elegans genome by the modENCODE project. Science. 2010 Dec 24;330(6012):1775-87. doi: 10.1126/science.1196914.

31. Giuffra E, Tuggle CK; FAANG Consortium. Functional Annotation of Animal Genomes (FAANG): Current Achievements and Roadmap. Annu Rev Anim Biosci. 2019 Feb 15;7:65-88. doi: 10.1146/annurev-animal-020518-114913.

32. Glocker EO, Kotlarz D, Boztug K, Gertz EM, Schäffer AA, Noyan F, Perro M, Diestelhorst J, Allroth A, Murugan D, Hätscher N, Pfeifer D, Sykora KW, Sauer M, Kreipe H, Lacher M, Nustede R, Woellner C, Baumann U, Salzer U, Koletzko S, Shah N, Segal AW, Sauerbrey A, Buderus S, Snapper SB, Grimbacher B, Klein C. Inflammatory bowel disease and mutations affecting the interleukin-10 receptor. N Engl J Med. 2009 Nov 19;361(21):2033-45. doi: 10.1056/NEJMoa0907206.

33. Gorkin DU, Barozzi I, Zhao Y, Zhang Y, Huang H, Lee AY, Li B, Chiou J, Wildberg A, Ding B, Zhang B, Wang M, Strattan JS, Davidson JM, Qiu Y, Afzal V, Akiyama JA, Plajzer-Frick I, Novak CS, Kato M, Garvin TH, Pham QT, Harrington AN, Mannion BJ, Lee EA, Fukuda-Yuzawa Y, He Y, Preissl S, Chee S, Han JY, Williams BA, Trout D, Amrhein H, Yang H, Cherry JM, Wang W, Gaulton K, Ecker JR, Shen Y, Dickel DE, Visel A, Pennacchio LA, Ren B. An atlas of dynamic chromatin landscapes in mouse fetal development. Nature. 2020 Jul;583(7818):744-751. doi: 10.1038/s41586-020-2093-3.

34. Gwak MG, Chang SY. Gut-Brain Connection: Microbiome, Gut Barrier, and Environmental Sensors. Immune Netw. 2021 Jun 16;21(3):e20. doi: 10.4110/in.2021.21.e20.

35. Halstead MM, Ma X, Zhou C, Schultz RM, Ross PJ. Chromatin remodeling in bovine embryos indicates species-specific regulation of genome activation. Nat Commun. 2020 Sep 17;11(1):4654. doi: 10.1038/s41467-020-18508-3.

36. Harrison PW, Fan J, Richardson D, Clarke L, Zerbino D, Cochrane G, Archibald AL, Schmidt CJ, Flicek P. FAANG, establishing metadata standards, validation and best practices for the farmed and companion animal community. Anim Genet. 2018 Dec;49(6):520-526. doi: 10.1111/age.12736.

37. Heinz S, Benner C, Spann N, Bertolino E, Lin YC, Laslo P, Cheng JX, Murre C, Singh H, Glass CK. Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and B cell identities. Mol Cell. 2010 May 28;38(4):576-89. doi: 10.1016/j.molcel.2010.05.004

38. Ishibashi M, Ikeda S, Minami N. Comparative analysis of histone H3K4me3 modifications between blastocysts and somatic tissues in cattle. Sci Rep. 2021 Apr 15;11(1):8253. doi: 10.1038/s41598-021-87683-0.

39. Inserra A, Rogers GB, Licinio J, Wong ML. The Microbiota-Inflammasome Hypothesis of Major Depression. Bioessays. 2018 Sep;40(9):e1800027. doi: 10.1002/bies.201800027.

40. Iyer SS, Cheng G. Role of interleukin 10 transcriptional regulation in inflammation and autoimmune disease. Crit Rev Immunol. 2012;32(1):23-63. doi: 10.1615/critrevimmunol.v32.i1.30.

41. Jacobson A, Yang D, Vella M, Chiu IM. The intestinal neuro-immune axis: crosstalk between neurons, immune cells, and microbes. Mucosal Immunol. 2021 May;14(3):555-565. doi: 10.1038/s41385-020-00368-1.

42. Johnson DS, Mortazavi A, Myers RM, Wold B. Genome-wide mapping of in vivo protein-DNA interactions. Science. 2007 Jun 8;316(5830):1497-502. doi: 10.1126/science.1141319.

43. Jostins L, Ripke S, Weersma RK, Duerr RH, McGovern DP, Hui KY, Lee JC, Schumm LP, Sharma Y, Anderson CA, Essers J, Mitrovic M, Ning K, Cleynen I, Theatre E, Spain SL, Raychaudhuri S, Goyette P, Wei Z, Abraham C, Achkar JP, Ahmad T, Amininejad L, Ananthakrishnan AN, Andersen V, Andrews JM, Baidoo L, Balschun T, Bampton PA, Bitton A, Boucher G, Brand S, Büning C, Cohain A, Cichon S, D'Amato M, De Jong D, Devaney KL, Dubinsky M, Edwards C, Ellinghaus D, Ferguson LR, Franchimont D, Fransen K, Gearry R, Georges M, Gieger C, Glas J, Haritunians T, Hart A, Hawkey C, Hedl M, Hu X, Karlsen TH, Kupcinskas L, Kugathasan S, Latiano A, Laukens D, Lawrance IC, Lees CW, Louis E, Mahy G, Mansfield J, Morgan AR, Mowat C, Newman W, Palmieri O, Ponsioen CY, Potocnik U, Prescott NJ, Regueiro M, Rotter JI, Russell RK, Sanderson JD, Sans M, Satsangi J, Schreiber S, Simms LA, Sventoraityte J, Targan SR, Taylor KD, Tremelling M, Verspaget HW, De Vos M, Wijmenga C, Wilson DC, Winkelmann J, Xavier RJ, Zeissig S, Zhang B, Zhang CK, Zhao H; International IBD Genetics Consortium (IIBDGC), Silverberg MS, Annese V, Hakonarson H, Brant SR, Radford-Smith G, Mathew CG, Rioux JD, Schadt EE, Daly MJ, Franke A, Parkes M, Vermeire S, Barrett JC, Cho JH. Host-microbe interactions have shaped the genetic architecture of inflammatory bowel disease. Nature. 2012 Nov 1;491(7422):119-24. doi: 10.1038/nature11582.

44. Kang X, Liu S, Fang L, Lin S, Liu M, Baldwin RL, Liu GE, Li CJ. Data of epigenomic profiling of histone marks and CTCF binding sites in bovine rumen epithelial primary cells before and after butyrate treatment. Data Brief. 2019 Dec 12;28:104983. doi: 10.1016/j.dib.2019.104983.

45. Kaz AM, Wong CJ, Dzieciatkowski S, Luo Y, Schoen RE, Grady WM. Patterns of DNA methylation in the normal colon vary by anatomical location, gender, and age. Epigenetics. 2014 Apr;9(4):492-502. doi: 10.4161/epi.27650.

46. Kern C, Wang Y, Xu X, Pan Z, Halstead M, Chanthavixay G, Saelao P, Waters S, Xiang R, Chamberlain A, Korf I, Delany ME, Cheng HH, Medrano JF, Van Eenennaam AL, Tuggle CK, Ernst C, Flicek P, Quon G, Ross P, Zhou H. Functional annotations of three domestic animal genomes provide vital resources for comparative and agricultural research. Nat Commun. 2021 Mar 23;12(1):1821. doi: 10.1038/s41467-021-22100-8.

47. Kinkley S, Helmuth J, Polansky JK, Dunkel I, Gasparoni G, Fröhler S, Chen W, Walter J, Hamann A, Chung HR. reChIP-seq reveals widespread bivalency of H3K4me3 and H3K27me3 in CD4(+) memory T cells. Nat Commun. 2016 Aug 17;7:12514. doi: 10.1038/ncomms12514.

48. Kipnis J. Multifaceted interactions between adaptive immunity and the central nervous system. Science. 2016 Aug 19;353(6301):766-71. doi: 10.1126/science.aag2638.

49. Kipnis J, Gadani S, Derecki NC. Pro-cognitive properties of T cells. Nat Rev Immunol. 2012 Sep;12(9):663-9. doi: 10.1038/nri3280.

50. Kotlarz D, Beier R, Murugan D, Diestelhorst J, Jensen O, Boztug K, Pfeifer D, Kreipe H, Pfister ED, Baumann U, Puchalka J, Bohne J, Egritas O, Dalgic B, Kolho KL, Sauerbrey A, Buderus S, Güngör T, Enninger A, Koda YK, Guariso G, Weiss B, Corbacioglu S, Socha P, Uslu N, Metin A, Wahbeh GT, Husain K, Ramadan D, Al-Herz W, Grimbacher B, Sauer M, Sykora KW, Koletzko S, Klein C. Loss of interleukin-10 signaling and infantile inflammatory bowel disease: implications for diagnosis and therapy. Gastroenterology. 2012 Aug;143(2):347-55. doi: 10.1053/j.gastro.2012.04.045..

51. Landt SG, Marinov GK, Kundaje A, Kheradpour P, Pauli F, Batzoglou S, Bernstein BE, Bickel P, Brown JB, Cayting P, Chen Y, DeSalvo G, Epstein C, Fisher-Aylor KI, Euskirchen G, Gerstein M, Gertz J, Hartemink AJ, Hoffman MM, Iyer VR, Jung YL, Karmakar S, Kellis M, Kharchenko PV, Li Q, Liu T, Liu XS, Ma L, Milosavljevic A, Myers RM, Park PJ, Pazin MJ, Perry MD, Raha D, Reddy TE, Rozowsky J, Shoresh N, Sidow A, Slattery M, Stamatoyannopoulos JA, Tolstorukov MY, White KP, Xi S, Farnham PJ, Lieb JD, Wold BJ, Snyder M. ChIP-seq guidelines and practices of the ENCODE and modENCODE consortia. Genome Res. 2012 Sep;22(9):1813-31. doi: 10.1101/gr.136184.111..

52. Langmead B, Salzberg SL. Fast gapped-read alignment with Bowtie 2. Nat Methods. 2012 Mar 4;9(4):357-9. doi: 10.1038/nmeth.1923.

53. Li Z, Schulz MH, Look T, Begemann M, Zenke M, Costa IG. Identification of transcription factor binding sites using ATAC-seq. Genome Biol. 2019 Feb 26;20(1):45. doi: 10.1186/s13059-019-1642-2.

54. Lobo-Silva D, Carriche GM, Castro AG, Roque S, Saraiva M. Balancing the immune response in the brain: IL-10 and its regulation. J Neuroinflammation. 2016 Nov 24;13(1):297. doi: 10.1186/s12974-016-0763-8.

55. Massa AT, Mousel MR, Herndon MK, Herndon DR, Murdoch BM, White SN. Genome-Wide Histone Modifications and CTCF Enrichment Predict Gene Expression in Sheep Macrophages. Front Genet. 2021 Jan 7;11:612031. doi: 10.3389/fgene.2020.612031.

56. Mas G, Blanco E, Ballaré C, Sansó M, Spill YG, Hu D, Aoi Y, Le Dily F, Shilatifard A, Marti-Renom MA, Di Croce L. Promoter bivalency favors an open chromatin architecture in embryonic stem cells. Nat Genet. 2018 Oct;50(10):1452-1462. doi: 10.1038/s41588-018-0218-5..

57. Micsinai M, Parisi F, Strino F, Asp P, Dynlacht BD, Kluger Y. Picking ChIP-seq peak detectors for analyzing chromatin modification experiments. Nucleic Acids Res. 2012 May;40(9):e70. doi: 10.1093/nar/gks048.

58. Mikkelsen TS, Ku M, Jaffe DB, Issac B, Lieberman E, Giannoukos G, Alvarez P, Brockman W, Kim TK, Koche RP, Lee W, Mendenhall E, O'Donovan A, Presser A, Russ C, Xie X, Meissner A, Wernig M, Jaenisch R, Nusbaum C, Lander ES, Bernstein BE. Genome-wide maps of chromatin state in pluripotent and lineage-committed cells. Nature. 2007 Aug 2;448(7153):553-60. doi: 10.1038/nature06008.

59. Molinari C, Casadio V, Foca F, Zingaretti C, Giannini M, Avanzolini A, Lucci E, Saragoni L, Passardi A, Amadori D, Calistri D, Zoli W. Gene methylation in rectal cancer: predictive marker of response to chemoradiotherapy? J Cell Physiol. 2013 Dec;228(12):2343-9. doi: 10.1002/jcp.24405.

60. Moran CJ, Walters TD, Guo CH, Kugathasan S, Klein C, Turner D, Wolters VM, Bandsma RH, Mouzaki M, Zachos M, Langer JC, Cutz E, Benseler SM, Roifman CM, Silverberg MS, Griffiths AM, Snapper SB, Muise AM. IL-10R polymorphisms are associated with very-early-onset ulcerative colitis. Inflamm Bowel Dis. 2013 Jan;19(1):115-23. doi: 10.1002/ibd.22974.

61. Naval-Sanchez M, Nguyen Q, McWilliam S, Porto-Neto LR, Tellam R, Vuocolo T, Reverter A, Perez-Enciso M, Brauning R, Clarke S, McCulloch A, Zamani W, Naderi S, Rezaei HR, Pompanon F, Taberlet P, Worley KC, Gibbs RA, Muzny DM, Jhangiani SN, Cockett N, Daetwyler H, Kijas J. Sheep genome functional annotation reveals proximal regulatory elements contributed to the evolution of modern breeds. Nat Commun. 2018 Feb 28;9(1):859. doi: 10.1038/s41467-017-02809-1.

62. Negi N, Das BK. CNS: Not an immunoprivilaged site anymore but a virtual secondary lymphoid organ. Int Rev Immunol. 2018 Jan 2;37(1):57-68. doi: 10.1080/08830185.2017.1357719.

63. Quinlan AR. BEDTools: The Swiss-Army Tool for Genome Feature Analysis. Curr Protoc Bioinformatics. 2014 Sep 8;47:11.12.1-34. doi: 10.1002/0471250953.bi1112s47.

64. Ramírez F, Dündar F, Diehl S, Grüning BA, Manke T. deepTools: a flexible platform for exploring deep-sequencing data. Nucleic Acids Res. 2014 Jul;42(Web Server issue):W187-91. doi: 10.1093/nar/gku365.

65. Roadmap Epigenomics Consortium, Kundaje A, Meuleman W, Ernst J, Bilenky M, Yen A, Heravi-Moussavi A, Kheradpour P, Zhang Z, Wang J, Ziller MJ, Amin V, Whitaker JW, Schultz MD, Ward LD, Sarkar A, Quon G, Sandstrom RS, Eaton ML, Wu YC, Pfenning AR, Wang X, Claussnitzer M, Liu Y, Coarfa C, Harris RA, Shoresh N, Epstein CB, Gjoneska E, Leung D, Xie W, Hawkins RD, Lister R, Hong C, Gascard P, Mungall AJ, Moore R, Chuah E, Tam A, Canfield TK, Hansen RS, Kaul R, Sabo PJ, Bansal MS, Carles A, Dixon JR, Farh KH, Feizi S, Karlic R, Kim AR, Kulkarni A, Li D, Lowdon R, Elliott G, Mercer TR, Neph SJ, Onuchic V, Polak P, Rajagopal N, Ray P, Sallari RC, Siebenthall KT, Sinnott-Armstrong NA, Stevens M, Thurman RE, Wu J, Zhang B, Zhou X, Beaudet AE, Boyer LA, De Jager PL, Farnham PJ, Fisher SJ, Haussler D, Jones SJ, Li W, Marra MA, McManus MT,

Sunyaev S, Thomson JA, Tlsty TD, Tsai LH, Wang W, Waterland RA, Zhang MQ, Chadwick LH, Bernstein BE, Costello JF, Ecker JR, Hirst M, Meissner A, Milosavljevic A, Ren B, Stamatoyannopoulos JA, Wang T, Kellis M. Integrative analysis of 111 reference human epigenomes. Nature. 2015 Feb 19;518(7539):317-30. doi: 10.1038/nature14248.

66. Robertson G, Hirst M, Bainbridge M, Bilenky M, Zhao Y, Zeng T, Euskirchen G, Bernier B, Varhol R, Delaney A, Thiessen N, Griffith OL, He A, Marra M, Snyder M, Jones S. Genome-wide profiles of STAT1 DNA association using chromatin immunoprecipitation and massively parallel sequencing. Nat Methods. 2007 Aug;4(8):651-7. doi: 10.1038/nmeth1068.

67. Rutsch A, Kantsjö JB, Ronchi F. The Gut-Brain Axis: How Microbiota and Host Inflammasome Influence Brain Physiology and Pathology. Front Immunol. 2020 Dec 10;11:604179. doi: 10.3389/fimmu.2020.604179.

68. Salavati M, Caulton A, Clark R, Gazova I, Smith TPL, Worley KC, Cockett NE, Archibald AL, Clarke SM, Murdoch BM, Clark EL. Global Analysis of Transcription Start Sites in the New Ovine Reference Genome (*Oar rambouillet v1.0*). Front Genet. 2020 Oct 23;11:580580. doi: 10.3389/fgene.2020.580580.

69. Schopf LR, Hoffmann KF, Cheever AW, Urban JF Jr, Wynn TA. IL-10 is critical for host resistance and survival during gastrointestinal helminth infection. J Immunol. 2002 Mar 1;168(5):2383-92. doi: 10.4049/jimmunol.168.5.2383.

70. Sharifi-Zarchi A, Gerovska D, Adachi K, Totonchi M, Pezeshk H, Taft RJ, Schöler HR, Chitsaz H, Sadeghi M, Baharvand H, Araúzo-Bravo MJ. DNA methylation regulates discrimination of enhancers from promoters through a H3K4me1-H3K4me3 seesaw mechanism. BMC Genomics. 2017 Dec 12;18(1):964. doi: 10.1186/s12864-017-4353-7.

71. Shen Y, Yue F, McCleary DF, Ye Z, Edsall L, Kuan S, Wagner U, Dixon J, Lee L, Lobanenkov VV, Ren B. A map of the cis-regulatory sequences in the mouse genome. Nature. 2012 Aug 2;488(7409):116-20. doi: 10.1038/nature11243.

72. Shouval DS, Ouahed J, Biswas A, Goettel JA, Horwitz BH, Klein C, Muise AM, Snapper SB. Interleukin 10 receptor signaling: master regulator of intestinal mucosal homeostasis in mice and humans. Adv Immunol. 2014;122:177-210. doi: 10.1016/B978-0-12-800267-4.00005-5.

73. Siska C, Kechris K. Differential correlation for sequencing data. BMC Res Notes. 2017 Jan 19;10(1):54. doi: 10.1186/s13104-016-2331-9.

74. Sivasubbu S, Sachidanandan C, Scaria V. Time for the zebrafish ENCODE. J Genet. 2013 Dec;92(3):695-701. doi: 10.1007/s12041-013-0313-4.

75. Stergachis AB, Neph S, Sandstrom R, Haugen E, Reynolds AP, Zhang M, Byron R, Canfield T, Stelhing-Sun S, Lee K, Thurman RE, Vong S, Bates D, Neri F, Diegel M, Giste E, Dunn D, Vierstra J, Hansen RS, Johnson AK, Sabo PJ, Wilken MS, Reh TA, Treuting PM, Kaul R, Groudine M, Bender MA, Borenstein E, Stamatoyannopoulos JA. Conservation of trans-acting circuitry during mammalian regulatory evolution. Nature. 2014 Nov 20;515(7527):365-70. doi: 10.1038/nature13972.

76. Teng L, Tan K. Finding combinatorial histone code by semi-supervised biclustering. BMC Genomics. 2012 Jul 3;13:301. doi: 10.1186/1471-2164-13-301.

77. Tian Y, Jia Z, Wang J, Huang Z, Tang J, Zheng Y, Tang Y, Wang Q, Tian Z, Yang D, Zhang Y, Fu X, Song J, Liu S, van Velkinburgh JC, Wu Y, Ni B. Global mapping of H3K4me1 and H3K4me3 reveals the chromatin state-based cell type-specific gene regulation in human Treg cells. PLoS One. 2011;6(11):e27770. doi: 10.1371/journal.pone.0027770.

78. Tuggle CK, Giuffra E, White SN, Clarke L, Zhou H, Ross PJ, Acloque H, Reecy JM, Archibald A, Bellone RR, Boichard M, Chamberlain A, Cheng H, Crooijmans RP, Delany ME, Finno CJ, Groenen MA, Hayes B, Lunney JK, Petersen JL, Plastow GS, Schmidt CJ, Song J, Watson M. GO-FAANG meeting: a Gathering On Functional Annotation of Animal Genomes. Anim Genet. 2016 Oct;47(5):528-33. doi: 10.1111/age.12466.

79. Uhde AK, Ciurkiewicz M, Herder V, Khan MA, Hensel N, Claus P, Beckstette M, Teich R, Floess S, Baumgärtner W, Jung K, Huehn J, Beineke A. Intact interleukin-10 receptor signaling protects from hippocampal damage elicited by experimental neurotropic virus infection of SJL mice. Sci Rep. 2018 Apr 17;8(1):6106. doi: 10.1038/s41598-018-24378-z..

80. Uysal F, Ozturk S. The loss of global DNA methylation due to decreased DNMT expression in the postnatal mouse ovaries may associate with infertility emerging during ovarian aging. Histochem Cell Biol. 2020 Sep;154(3):301-314. doi: 10.1007/s00418-020-01890-w.

81. van der Velde A, Fan K, Tsuji J, Moore JE, Purcaro MJ, Pratt HE, Weng Z. Annotation of chromatin states in 66 complete mouse epigenomes during development. Commun Biol. 2021 Feb 22;4(1):239. doi: 10.1038/s42003-021-01756-4.

82. Wei HX, Wang B, Li B. IL-10 and IL-22 in Mucosal Immunity: Driving Protection and Pathology. Front Immunol. 2020 Jun 26;11:1315. doi: 10.3389/fimmu.2020.01315.

83. Xi X, Zou Q, Wei Y, Chen Y, Wang X, Lv D, Li P, Wen A, Zhu L, Tang G, Ma J, Li M, Li X, Jiang Y. Dynamic Changes of DNA Methylation and Transcriptome Expression in Porcine Ovaries during Aging. Biomed Res Int. 2019 Oct 30;2019:8732023. doi: 10.1155/2019/8732023.

84. Yan L, Guo H, Hu B, Li R, Yong J, Zhao Y, Zhi X, Fan X, Guo F, Wang X, Wang W, Wei Y, Wang Y, Wen L, Qiao J, Tang F. Epigenomic Landscape of Human Fetal Brain, Heart, and Liver. J Biol Chem. 2016 Feb 26;291(9):4386-98. doi: 10.1074/jbc.M115.672931..

85. Yue F, Cheng Y, Breschi A, Vierstra J, Wu W, Ryba T, Sandstrom R, Ma Z, Davis C, Pope BD, Shen Y, Pervouchine DD, Djebali S, Thurman RE, Kaul R, Rynes E, Kirilusha A, Marinov GK, Williams BA, Trout D, Amrhein H, Fisher-Aylor K, Antoshechkin I, DeSalvo G, See LH, Fastuca M, Drenkow J, Zaleski C, Dobin A, Prieto P, Lagarde J, Bussotti G, Tanzer A, Denas O, Li K, Bender MA, Zhang M, Byron R, Groudine MT, McCleary D, Pham L, Ye Z, Kuan S, Edsall L, Wu YC, Rasmussen MD, Bansal MS, Kellis M, Keller CA, Morrissey CS, Mishra T, Jain D, Dogan N, Harris RS, Cayting P, Kawli T, Boyle AP, Euskirchen G, Kundaje A, Lin S, Lin Y, Jansen C, Malladi VS, Cline MS, Erickson DT, Kirkup VM, Learned K, Sloan CA, Rosenbloom KR, Lacerda de Sousa B, Beal K, Pignatelli M, Flicek P, Lian J, Kahveci T, Lee D, Kent WJ, Ramalho Santos M, Herrero J, Notredame C, Johnson A, Vong S, Lee K, Bates D, Neri F, Diegel M, Canfield T, Sabo PJ, Wilken MS, Reh TA, Giste E, Shafer A, Kutyavin T, Haugen E, Dunn D, Reynolds AP, Neph S, Humbert R, Hansen RS, De Bruijn M, Selleri L, Rudensky A, Josefowicz S, Samstein R, Eichler EE, Orkin SH, Levasseur D, Papayannopoulou T, Chang KH, Skoultchi A, Gosh S, Disteche C, Treuting P, Wang Y, Weiss MJ, Blobel GA, Cao X, Zhong S, Wang T, Good PJ, Lowdon RF, Adams LB, Zhou XQ, Pazin MJ, Feingold EA, Wold B, Taylor J, Mortazavi A, Weissman SM, Stamatoyannopoulos JA, Snyder MP, Guigo R, Gingeras TR, Gilbert DM, Hardison RC, Beer MA, Ren B; Mouse ENCODE Consortium. A comparative encyclopedia of DNA elements in the mouse genome. Nature. 2014 Nov 20;515(7527):355-64. doi: 10.1038/nature13992.

86. Zang C, Schones DE, Zeng C, Cui K, Zhao K, Peng W. A clustering approach for identification of enriched domains from histone modification ChIP-Seq data. Bioinformatics. 2009 Aug 1;25(15):1952-8. doi: 10.1093/bioinformatics/btp340.

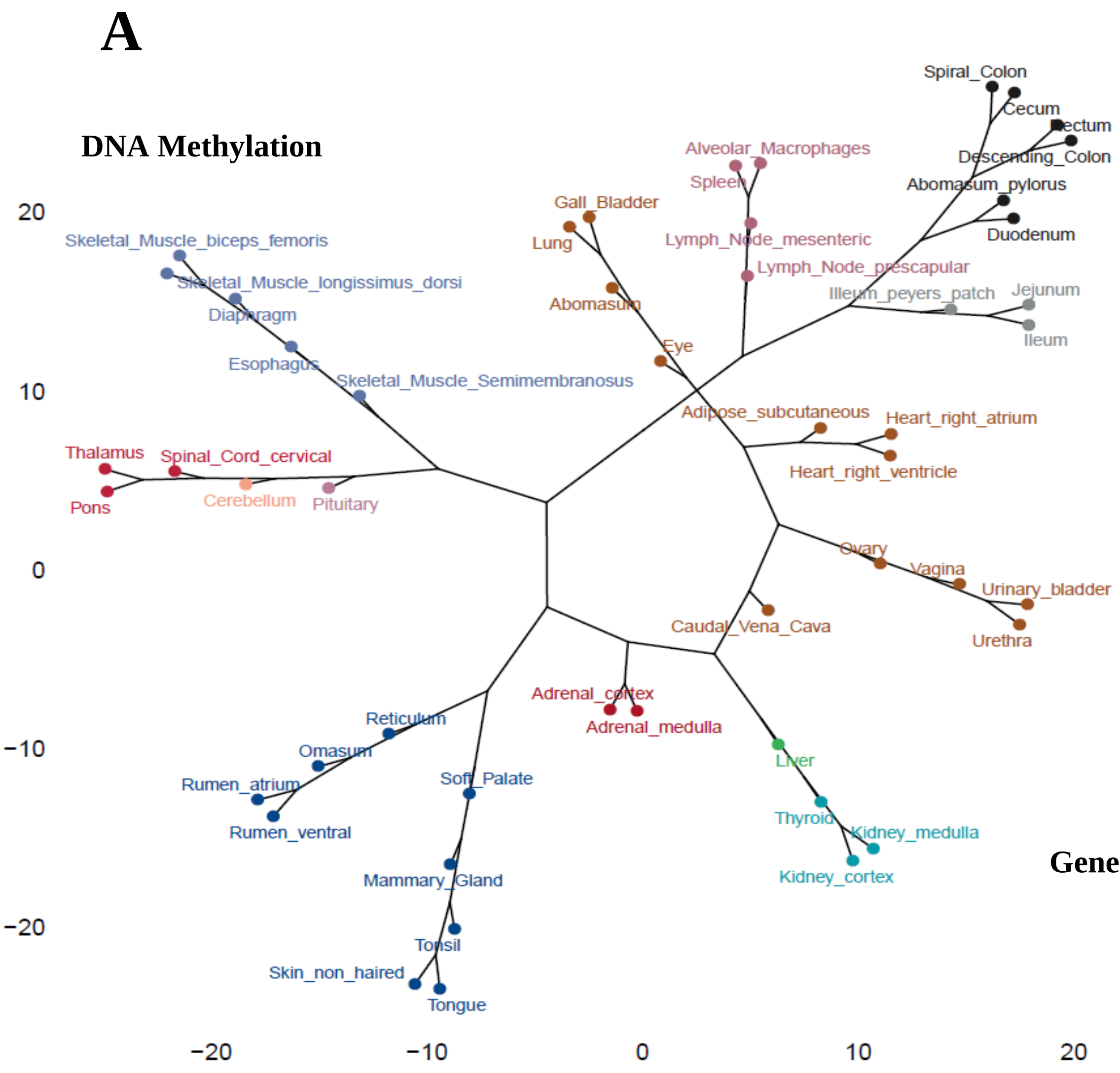
87. Zhang B, Zhou Y, Lin N, Lowdon RF, Hong C, Nagarajan RP, Cheng JB, Li D, Stevens M, Lee HJ, Xing X, Zhou J, Sundaram V, Elliott G, Gu J, Shi T, Gascard P, Sigaroudinia M, Tlsty TD, Kadlecek T, Weiss A, O'Geen H, Farnham PJ, Maire CL, Ligon KL, Madden PA, Tam A, Moore R, Hirst M, Marra MA, Zhang B, Costello JF, Wang T. Functional DNA methylation differences between tissues, cell types, and across individuals discovered using the M&M algorithm. Genome Res. 2013 Sep;23(9):1522-40. doi: 10.1101/gr.156539.113.

88. Zhang X, Bernatavichute YV, Cokus S, Pellegrini M, Jacobsen SE. Genome-wide analysis of mono-, di- and trimethylation of histone H3 lysine 4 in Arabidopsis thaliana. Genome Biol. 2009;10(6):R62. doi: 10.1186/gb-2009-10-6-r62.

89. Zhang Z, Zhang MQ. Histone modification profiles are predictive for tissue/cell-type specific expression of both protein-coding and microRNA genes. BMC Bioinformatics. 2011 May 14;12:155. doi: 10.1186/1471-2105-12-155.

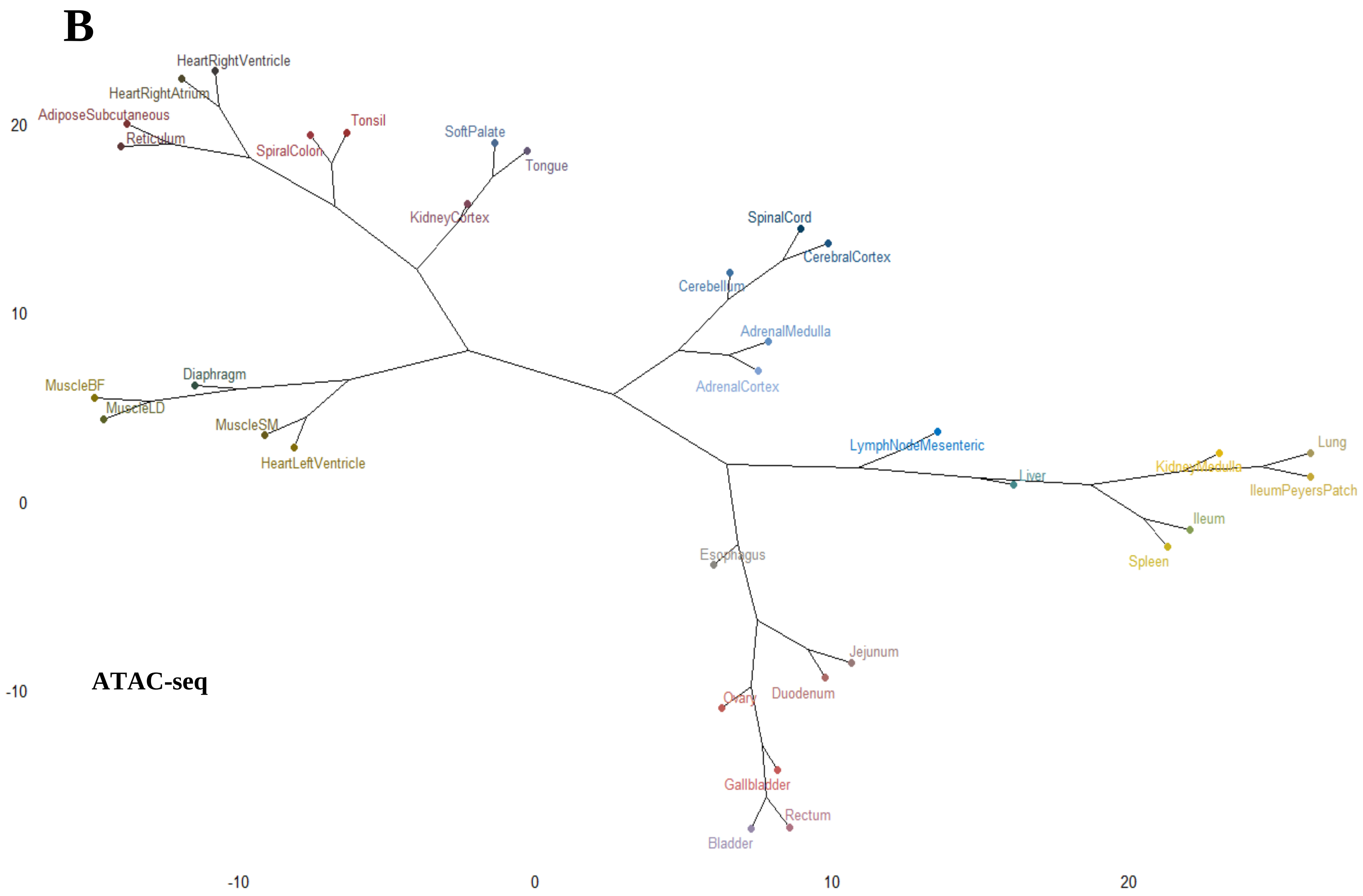
90. Zhou Y, Liu S, Hu Y, Fang L, Gao Y, Xia H, Schroeder SG, Rosen BD, Connor EE, Li CJ, Baldwin RL, Cole JB, Van Tassell CP, Yang L, Ma L, Liu GE. Comparative whole genome DNA methylation profiling across cattle tissues reveals global and tissue-specific methylation patterns. BMC Biol. 2020 Jul 6;18(1):85. doi: 10.1186/s12915-020-00793-5.

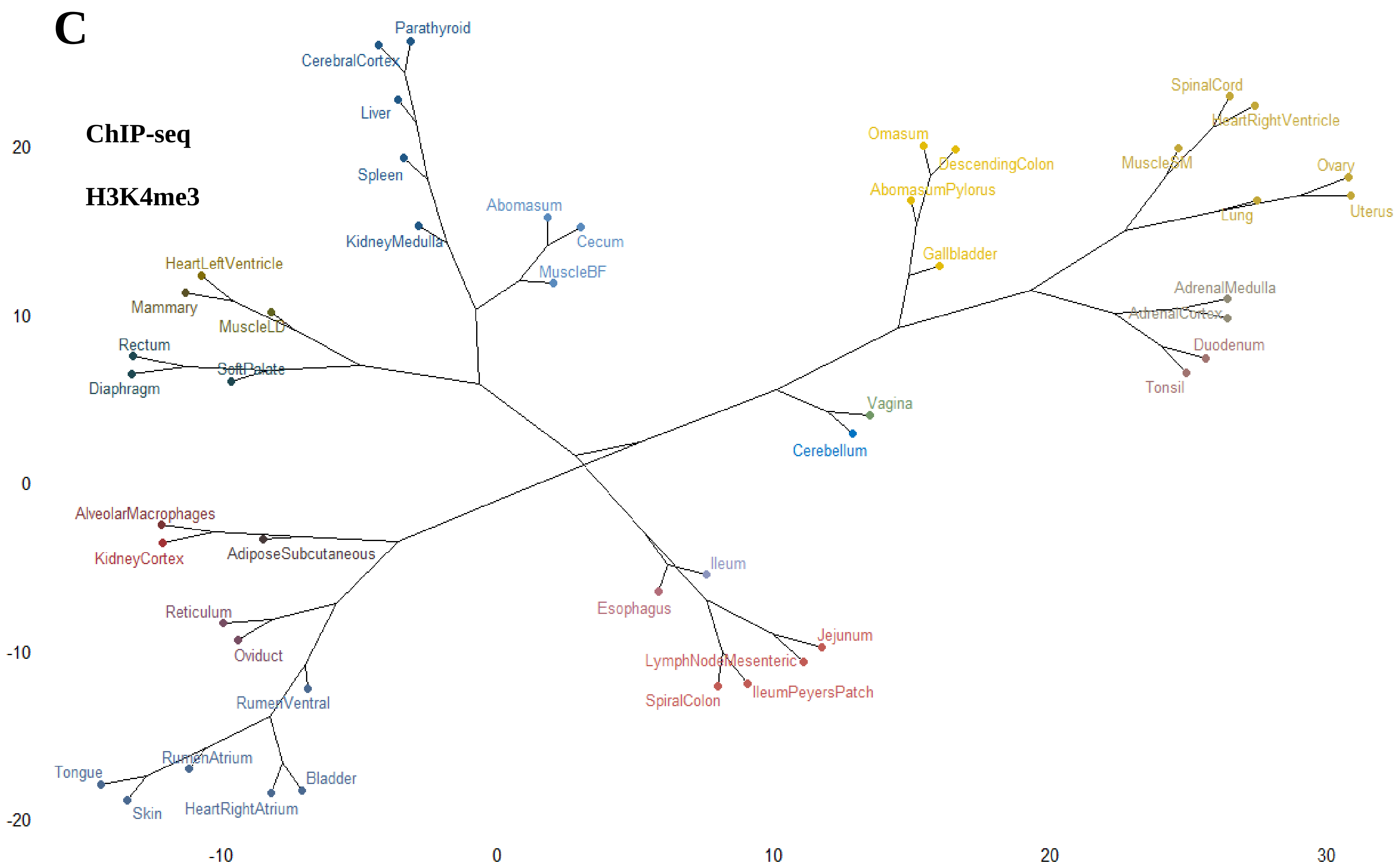
Figures

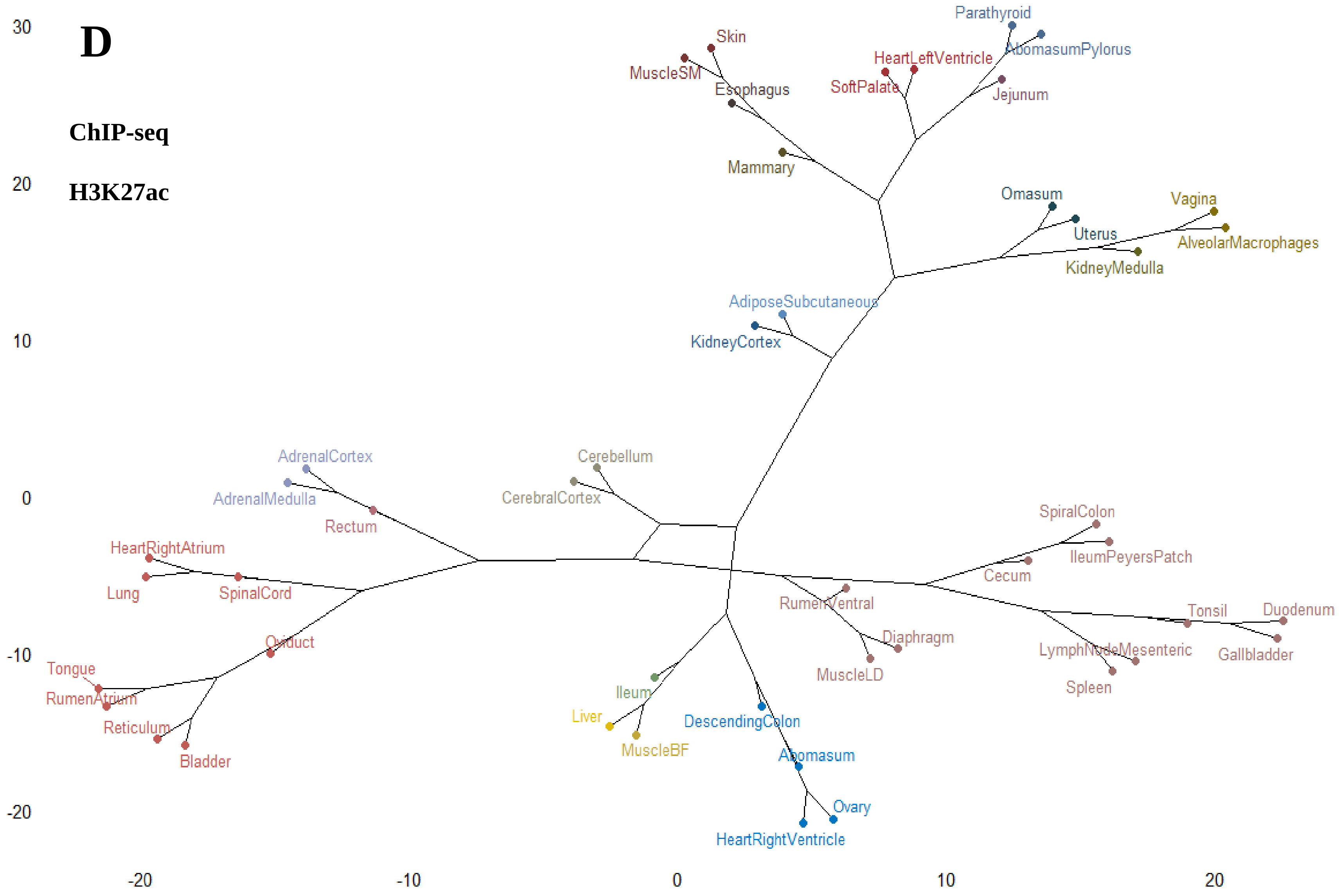


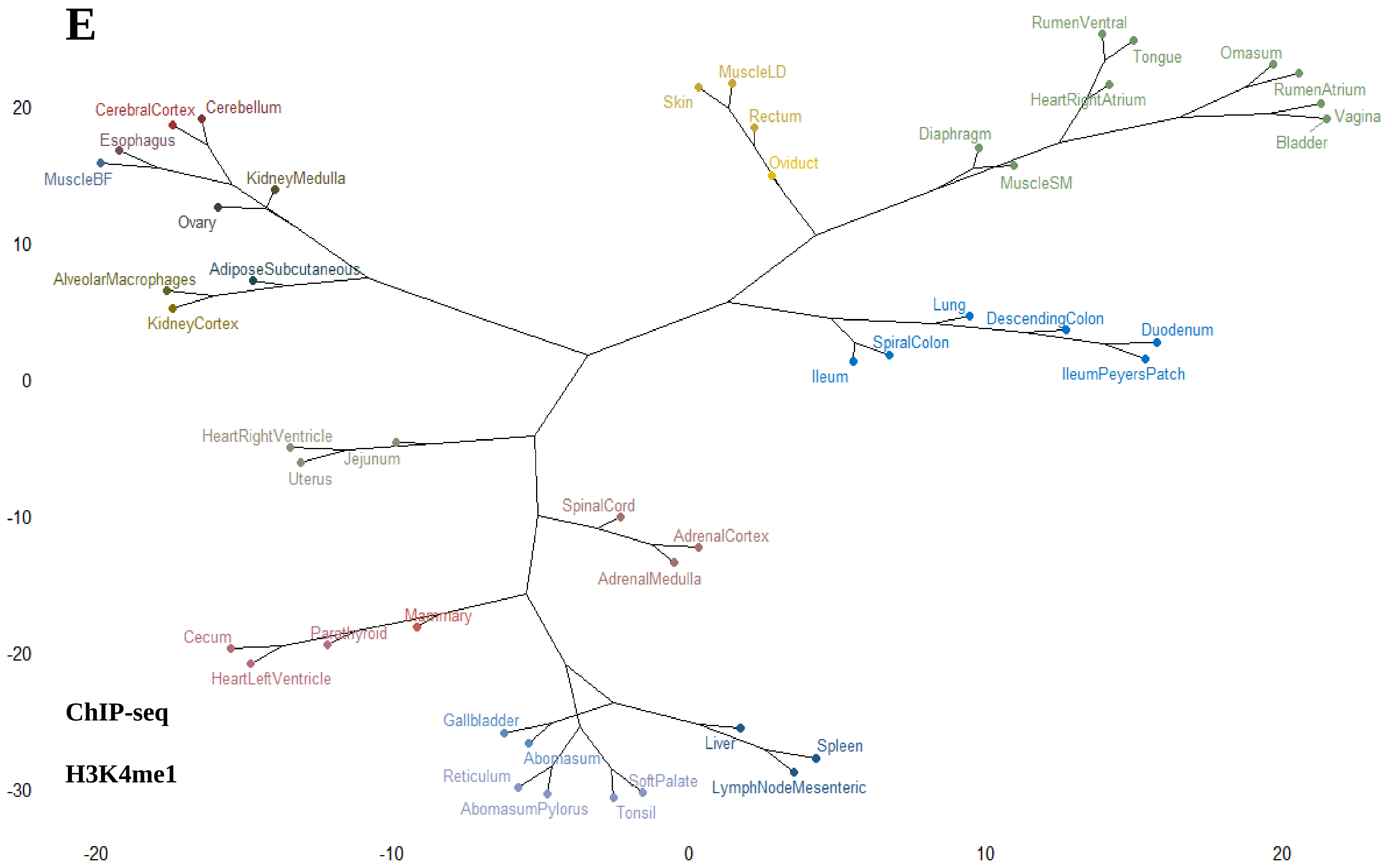
DNA Methylation

Generated by Alex Caulton


B
HeartRightVentricle
HeartRightAtrium
AdiposeSubcutaneous
Reticulum
SpiralColon
Tonsil
SoftPalate
Tongue
KidneyCortex
SpinalCord
CerebralCortex
Cerebellum
AdrenalMedulla
AdrenalCortex
Diaphragm
MuscleBF
MuscleLD
MuscleSM
HeartLeftVentricle
LymphNodeMesenteric
Liver
Lung
KidneyMedulla
IleumPeyersPatch
Ileum
Spleen
Esophagus
Jejunum
Duodenum
Ovary
Gallbladder
Rectum
Bladder
ATAC-seq
20
10
0
-10
-10
0
10
20


C
ChIP-seq
H3K4me3
Parathyroid
CerebralCortex
Liver
Spleen
KidneyMedulla
Abomasum
Cecum
MuscleBF
HeartLeftVentricle
Mammary
MuscleLD
Rectum
SoftPalate
Diaphragm
Omasum
DescendingColon
AbomasumPylorus
Gallbladder
SpinalCord
HeartRightVentricle
MuscleSM
Ovary
Lung
Uterus
AdrenalMedulla
AdrenalCortex
Duodenum
Tonsil
Vagina
Cerebellum
AlveolarMacrophages
KidneyCortex
AdiposeSubcutaneous
Ileum
Esophagus
Jejunum
LymphNodeMesenteric
IleumPeyersPatch
SpiralColon
Reticulum
Oviduct
RumenVentral
RumenAtrium
Tongue
Skin
HeartRightAtrium
Bladder
20
10
0
-10
-20
-10
0
10
20
30


D
ChIP-seq
H3K27ac
30
20
10
0
-10
-20
-20
-10
0
10
20
Parathyroid
AbomasumPylorus
Skin
MuscleSM
HeartLeftVentricle
SoftPalate
Esophagus
Jejunum
Mammary
Omasum
Vagina
Uterus
AlveolarMacrophages
KidneyMedulla
AdiposeSubcutaneous
KidneyCortex
AdrenalCortex
Cerebellum
AdrenalMedulla
CerebralCortex
Rectum
SpiralColon
HeartRightAtrium
IleumPeyersPatch
Cecum
Lung
SpinalCord
RumenVentral
Tonsil
Duodenum
Diaphragm
Oviduct
LymphNodeMesenteric
Gallbladder
Tongue
MuscleLD
Spleen
RumenAtrium
Ileum
Reticulum
DescendingColon
Liver
Bladder
MuscleBF
Abomasum
Ovary
HeartRightVentricle


E
ChIP-seq
H3K4me1
20
10
0
-10
-20
-30
-20
-10
0
10
20
CerebralCortex
Cerebellum
Esophagus
MuscleBF
KidneyMedulla
Ovary
AdiposeSubcutaneous
AlveolarMacrophages
KidneyCortex
MuscleLD
Skin
Rectum
Oviduct
RumenVentral
Tongue
HeartRightAtrium
Omasum
RumenAtrium
Vagina
Bladder
Diaphragm
MuscleSM
Lung
DescendingColon
Duodenum
SpiralColon
Ileum
IleumPeyersPatch
HeartRightVentricle
Jejunum
Uterus
SpinalCord
AdrenalCortex
AdrenalMedulla
Mammary
Parathyroid
Cecum
HeartLeftVentricle
Gallbladder
Abomasum
Reticulum
AbomasumPylorus
SoftPalate
Tonsil
Liver
Spleen
LymphNodeMesenteric

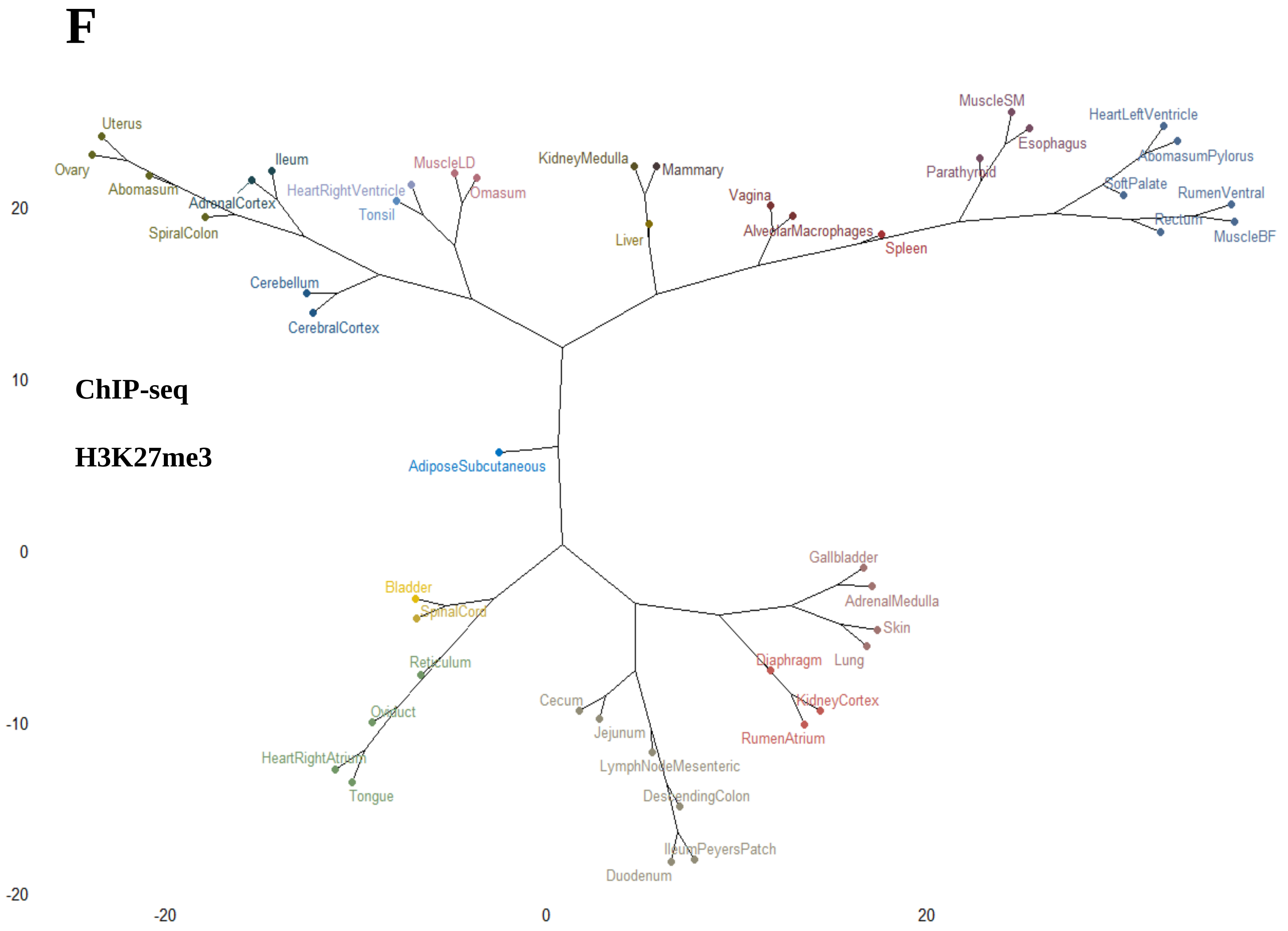
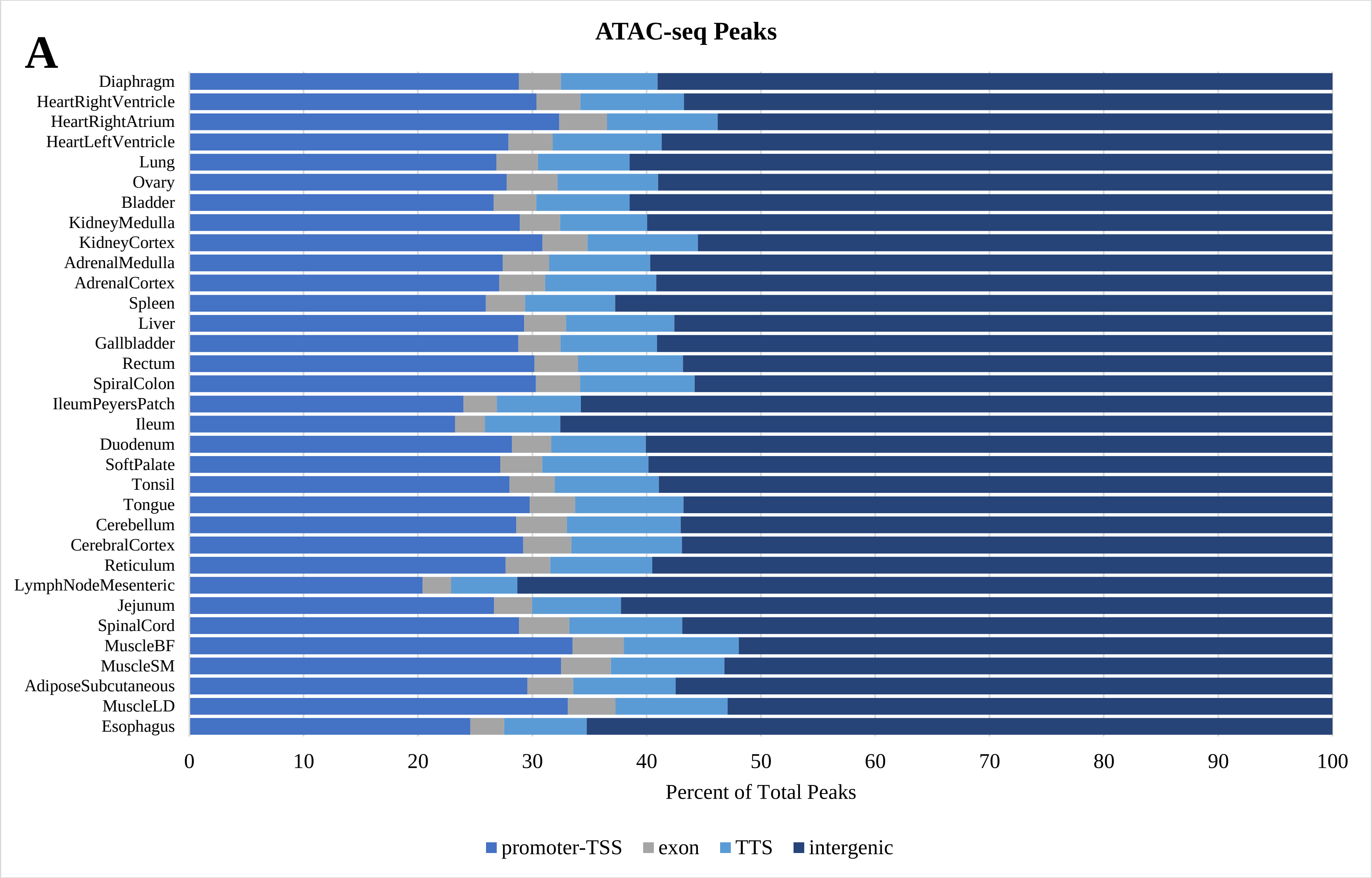
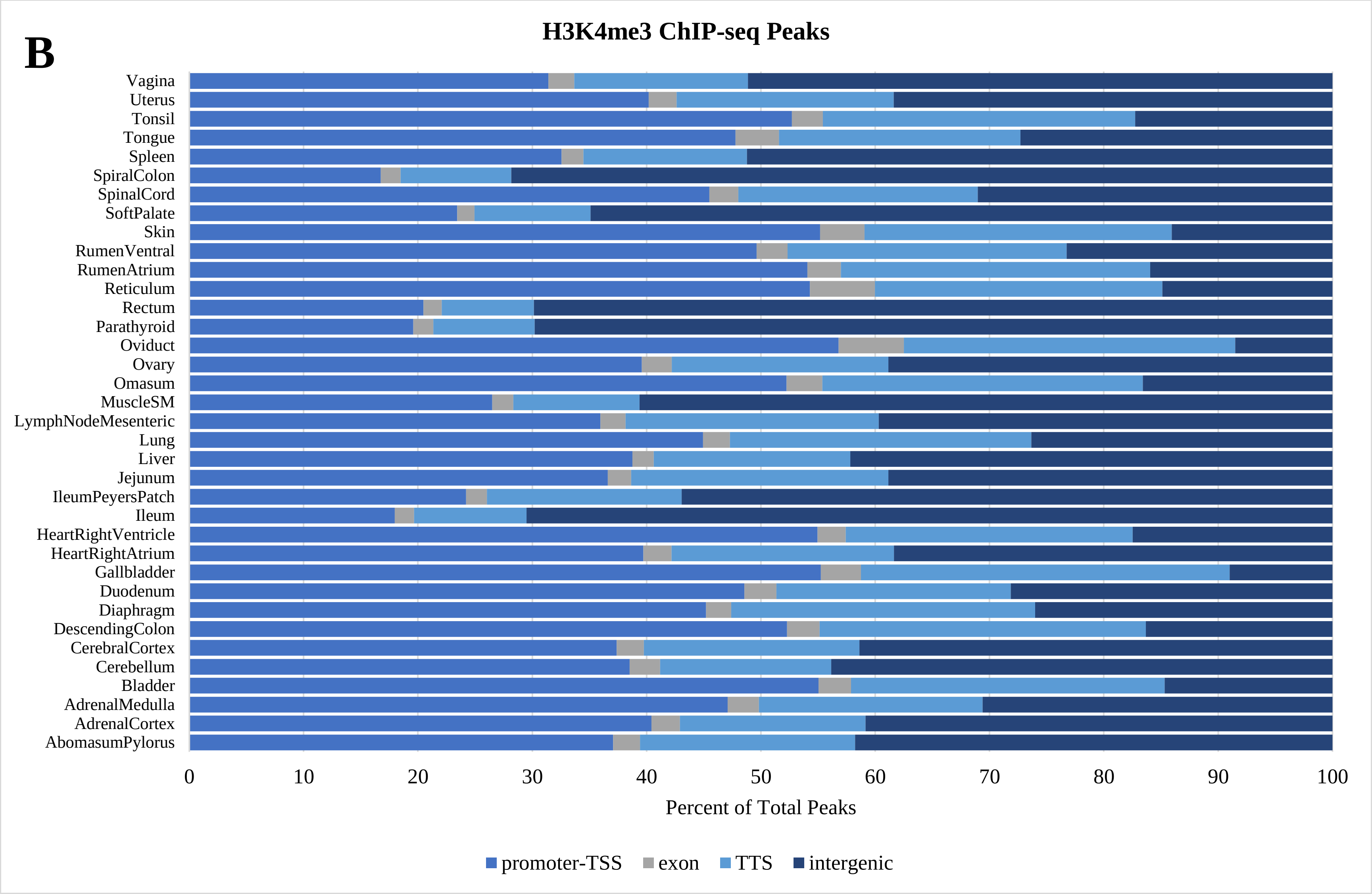
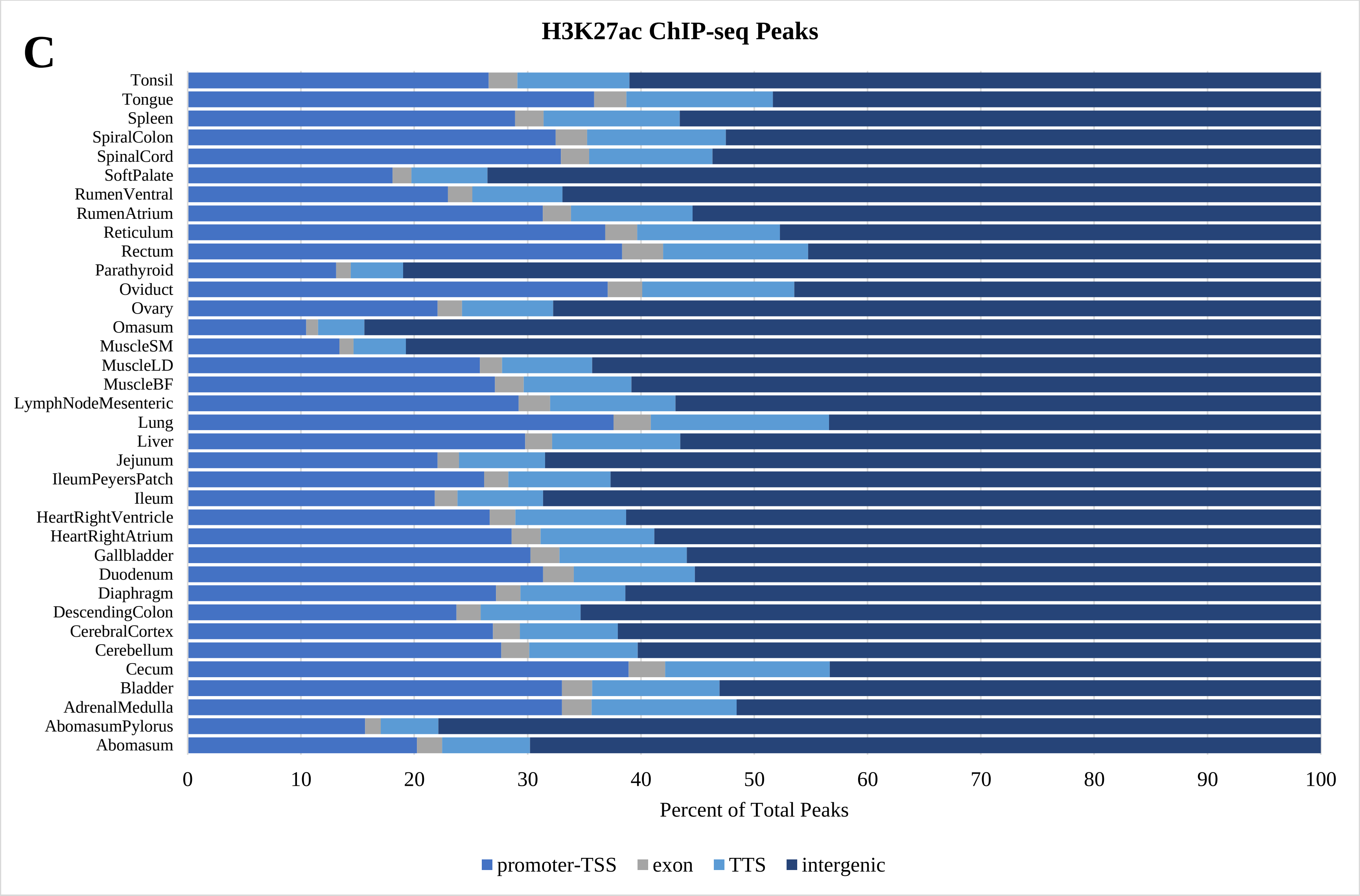
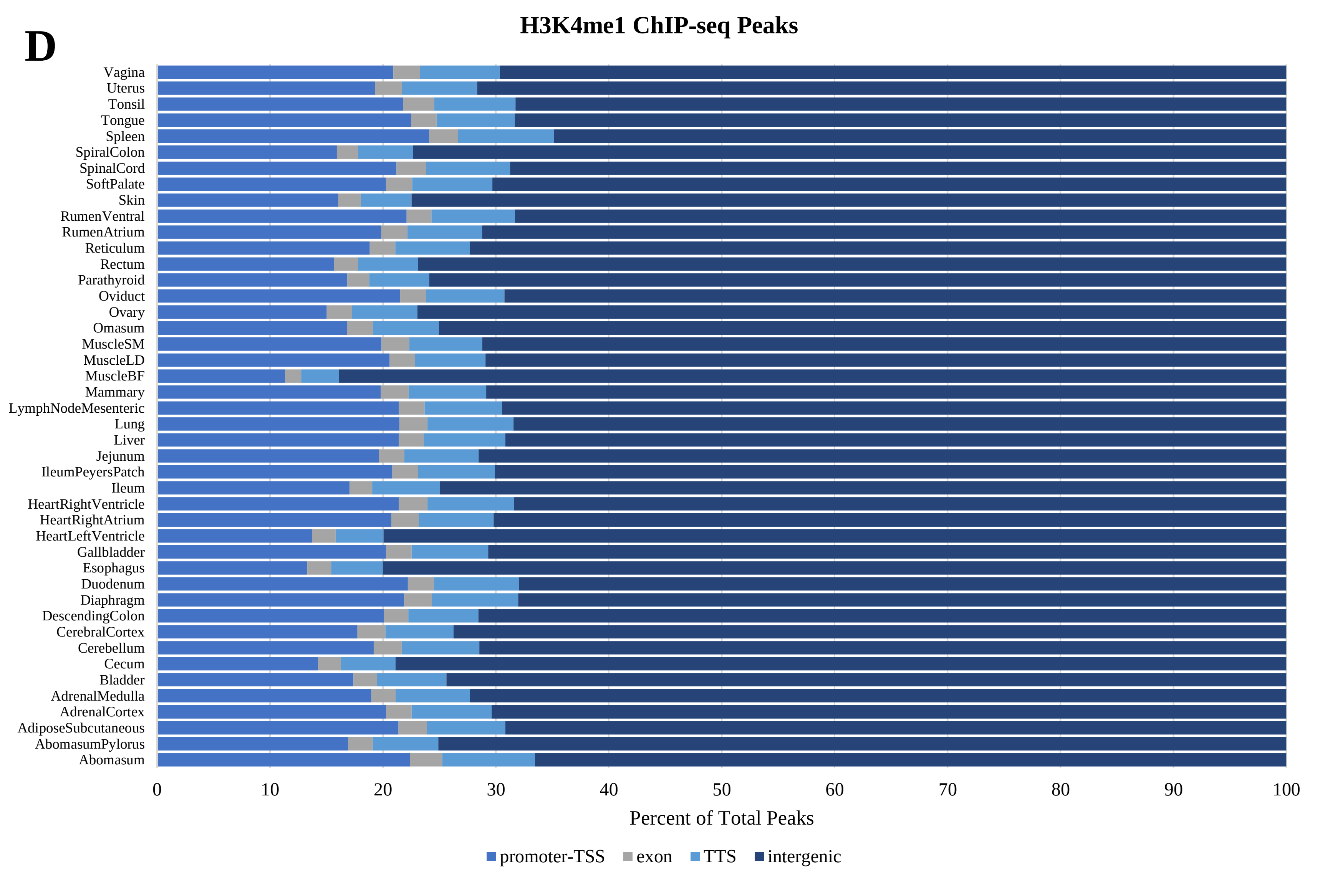


Figure 6.1: Cluster dendrograms of raw sequence signal displaying tissue relationships in A) DNA methylation, B) ATAC-seq, C) ChIP-seq of H3K4me3, D) ChIP-seq of H3K27ac, E) ChIP-seq of H3K4me1, and F) ChIP-seq of H3K27me3 data.


A
ATAC-seq Peaks
Diaphragm
HeartRightVentricle
HeartRightAtrium
HeartLeftVentricle
Lung
Ovary
Bladder
KidneyMedulla
KidneyCortex
AdrenalMedulla
AdrenalCortex
Spleen
Liver
Gallbladder
Rectum
SpiralColon
IleumPeyersPatch
Ileum
Duodenum
SoftPalate
Tonsil
Tongue
Cerebellum
CerebralCortex
Reticulum
LymphNodeMesenteric
Jejunum
SpinalCord
MuscleBF
MuscleSM
AdiposeSubcutaneous
MuscleLD
Esophagus
0
10
20
30
40
50
60
70
80
90
100
Percent of Total Peaks
promoter-TSS
exon
TTS
intergenic


B
H3K4me3 ChIP-seq Peaks
Vagina
Uterus
Tonsil
Tongue
Spleen
SpiralColon
SpinalCord
SoftPalate
Skin
RumenVentral
RumenAtrium
Reticulum
Rectum
Parathyroid
Oviduct
Ovary
Omasum
MuscleSM
LymphNodeMesenteric
Lung
Liver
Jejunum
IleumPeyersPatch
Ileum
HeartRightVentricle
HeartRightAtrium
Gallbladder
Duodenum
Diaphragm
DescendingColon
CerebralCortex
Cerebellum
Bladder
AdrenalMedulla
AdrenalCortex
AbomasumPylorus
0
10
20
30
40
50
60
70
80
90
100
Percent of Total Peaks
promoter-TSS
exon
TTS
intergenic


C
H3K27ac ChIP-seq Peaks
Tonsil
Tongue
Spleen
SpiralColon
SpinalCord
SoftPalate
RumenVentral
RumenAtrium
Reticulum
Rectum
Parathyroid
Oviduct
Ovary
Omasum
MuscleSM
MuscleLD
MuscleBF
LymphNodeMesenteric
Lung
Liver
Jejunum
IleumPeyersPatch
Ileum
HeartRightVentricle
HeartRightAtrium
Gallbladder
Duodenum
Diaphragm
DescendingColon
CerebralCortex
Cerebellum
Cecum
Bladder
AdrenalMedulla
AbomasumPylorus
Abomasum
0
10
20
30
40
50
60
70
80
90
100
Percent of Total Peaks
promoter-TSS
exon
TTS
intergenic


D
H3K4me1 ChIP-seq Peaks
Vagina
Uterus
Tonsil
Tongue
Spleen
SpiralColon
SpinalCord
SoftPalate
Skin
RumenVentral
RumenAtrium
Reticulum
Rectum
Parathyroid
Oviduct
Ovary
Omasum
MuscleSM
MuscleLD
MuscleBF
Mammary
LymphNodeMesenteric
Lung
Liver
Jejunum
IleumPeyersPatch
Ileum
HeartRightVentricle
HeartRightAtrium
HeartLeftVentricle
Gallbladder
Esophagus
Duodenum
Diaphragm
DescendingColon
CerebralCortex
Cerebellum
Cecum
Bladder
AdrenalMedulla
AdrenalCortex
AdiposeSubcutaneous
AbomasumPylorus
Abomasum
0
10
20
30
40
50
60
70
80
90
100
Percent of Total Peaks
promoter-TSS
exon
TTS
intergenic

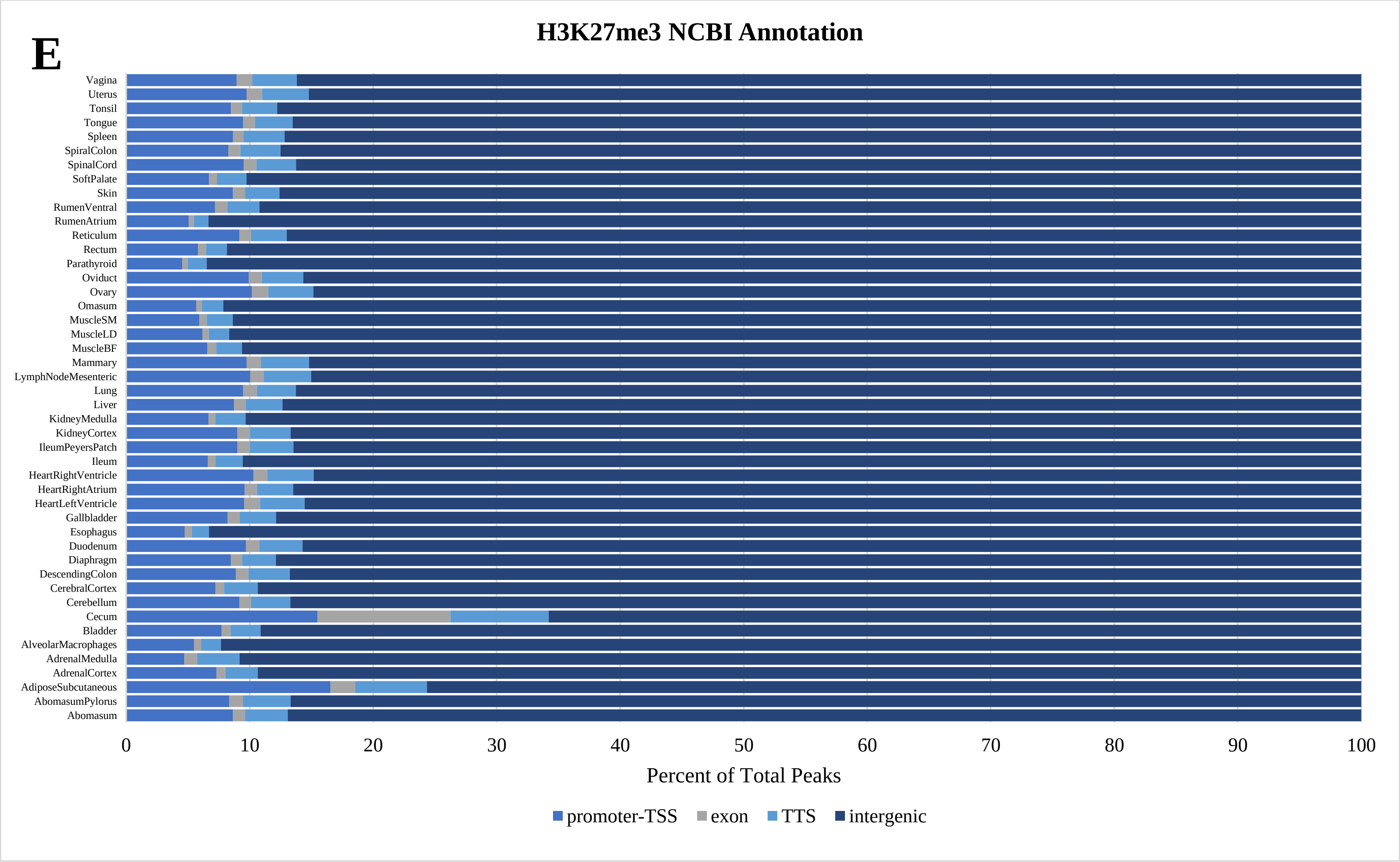
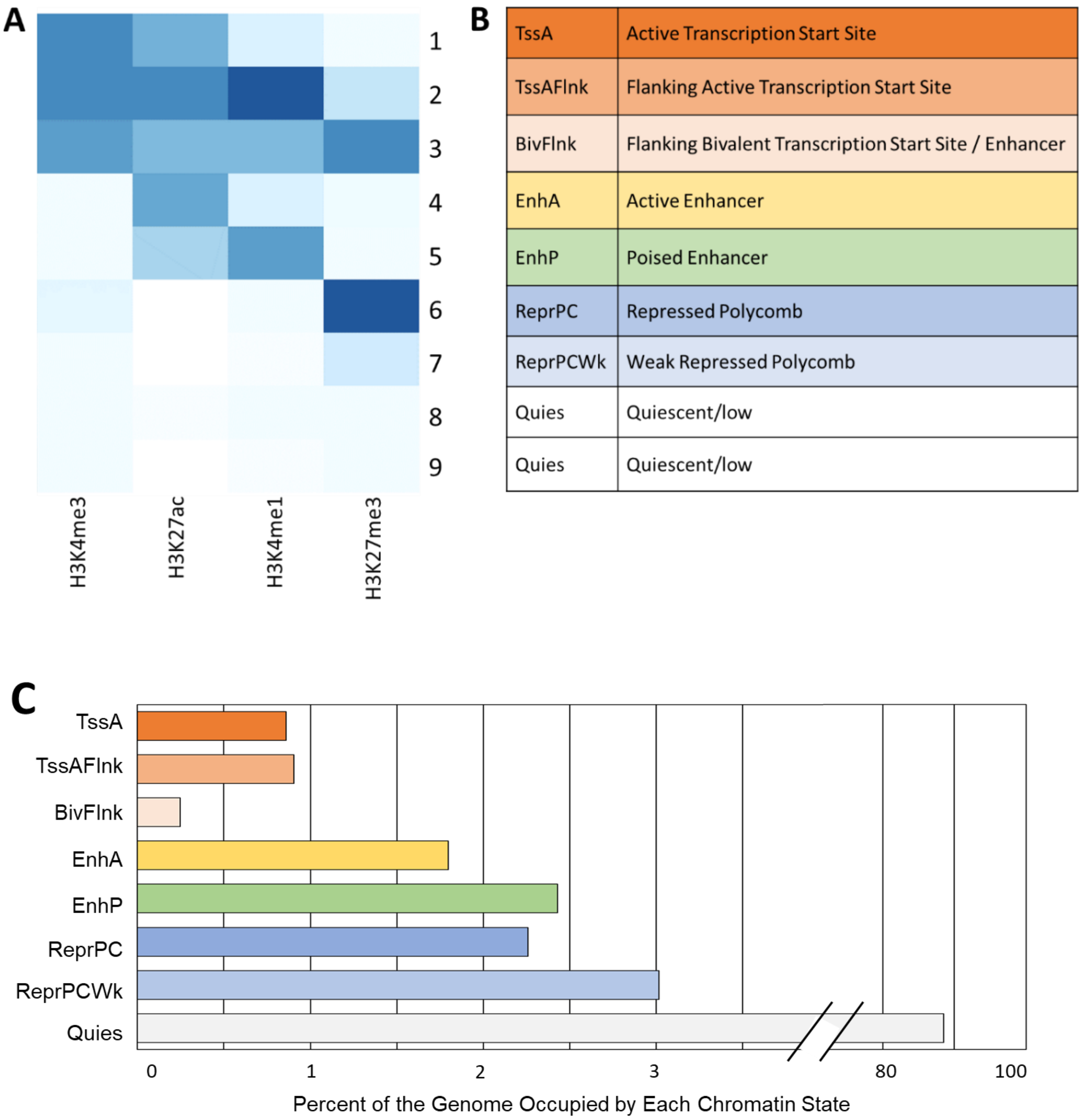


Figure 6.2: Annotation of peaks as percent of peaks overlapping with each feature from A) ATAC-seq and ChIP-seq B)H3K4me3, C) H3K27ac, D) H3K4me1, and E) H3K27me3 using the NCBI annotation release for the *ARS-UI_Ramb_v2.0 genome.*



TssA	Active Transcription Start Site
TssAFlnk	Flanking Active Transcription Start Site
BivFlnk	Flanking Bivalent Transcription Start Site / Enhancer
EnhA	Active Enhancer
EnhP	Poised Enhancer
ReprPC	Repressed Polycomb
ReprPCWk	Weak Repressed Polycomb
Quies	Quiescent/low
Quies	Quiescent/low

Figure 6.3: Chromatin states across tissues incorporating signal from all four histone modifications (H3K4me3, H3K27ac, H3K4me1, and H3K27me3). A) Heatmap of histone modification sequence signal in all nine chromatin states. B) Short name and description of each chromatin state based on the characteristic ChIP-seq signal (adapted from Roadmap Epigenomics Consortium and Gorkin et al., 2020). C) Genome occupancy of each chromatin state displayed as a percent of the genome.

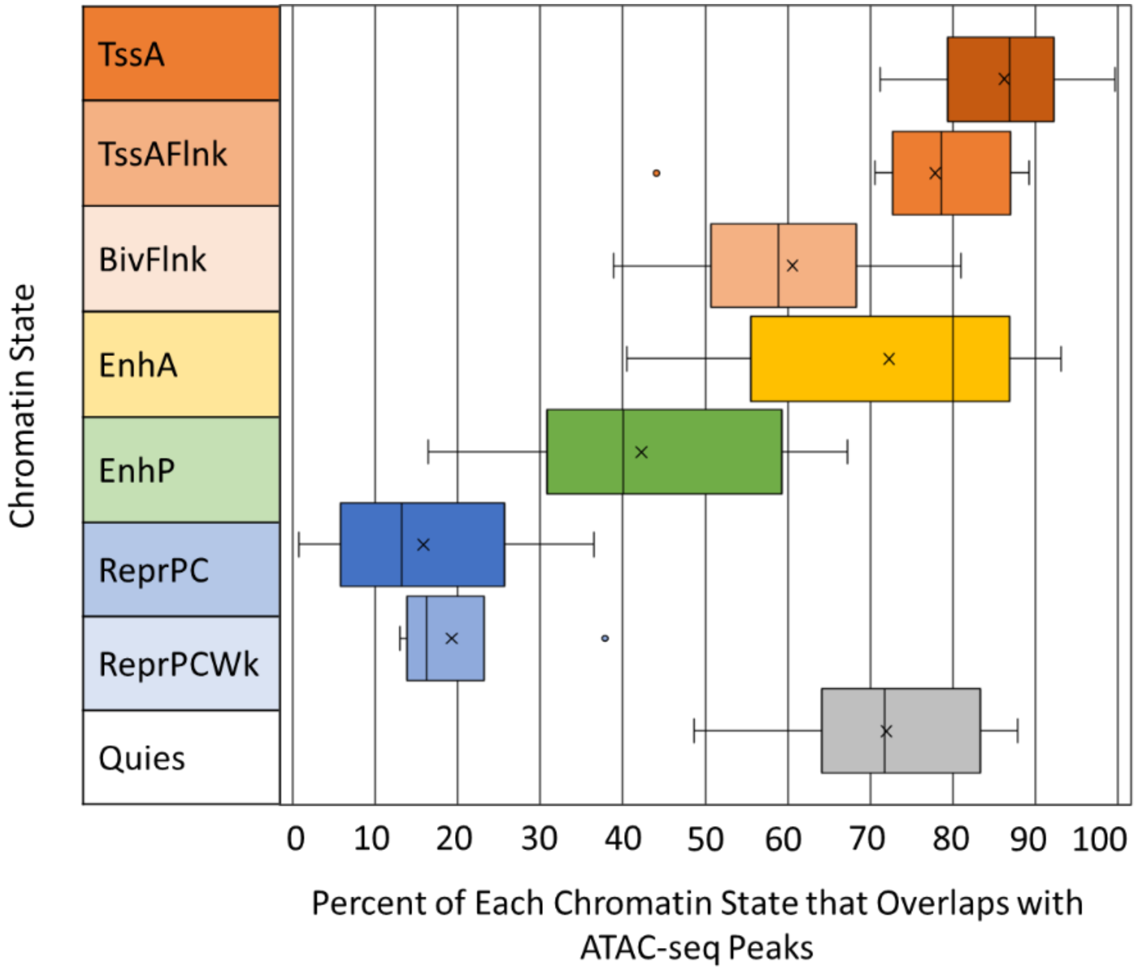


Figure 6.4: Chromatin state overlap with open chromatin regions. A) Chromatin state heatmap with states 1-9 defined by ChromHMM. B) Chromatin state name abbreviation. C) Percent of each chromatin state that overlaps with open chromatin as defined by ATAC-seq peaks as a boxplot with defined mean (solid line) and median (x) of the 33 tissues with ATAC-seq information.

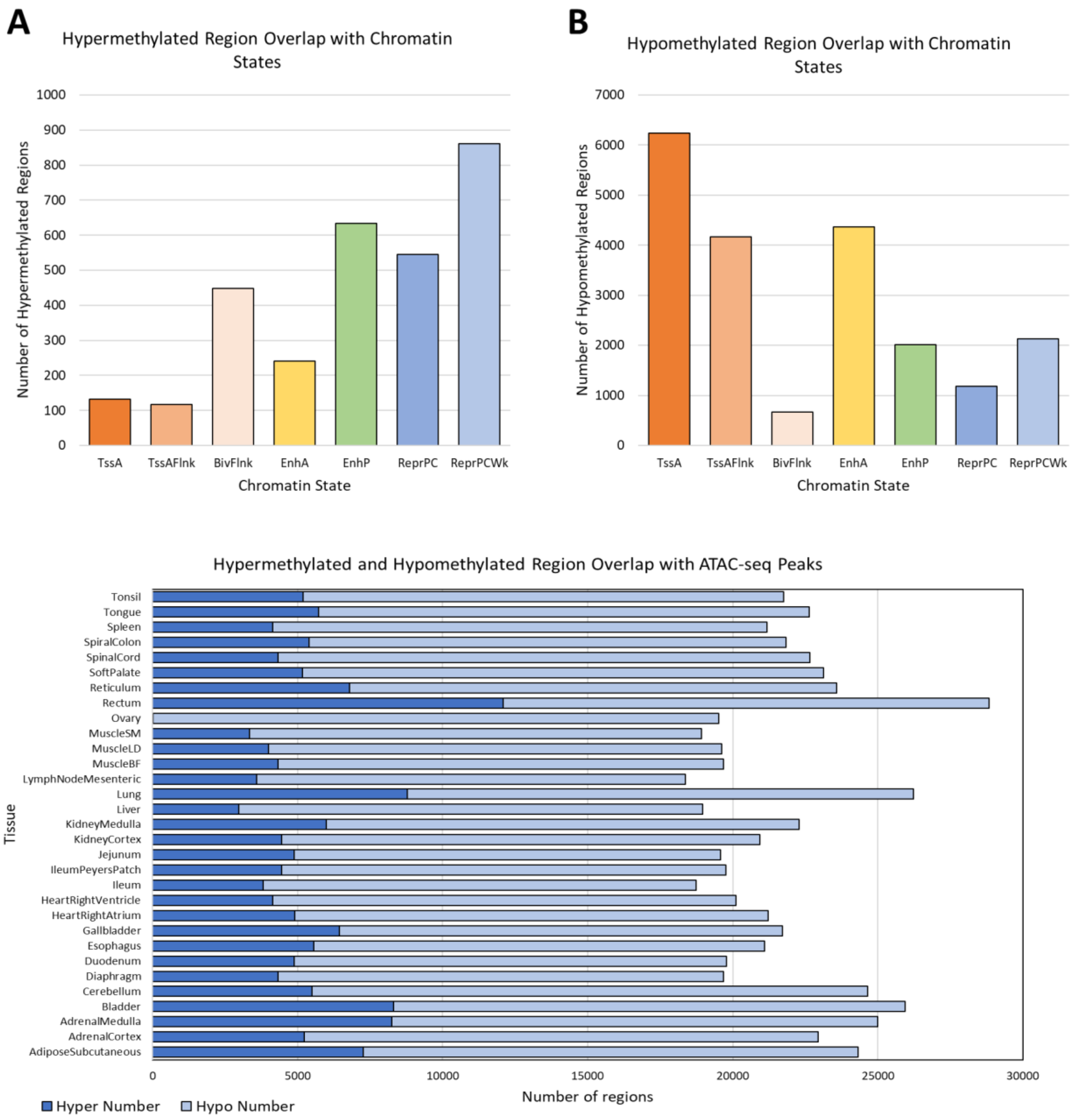


Figure 6.5: Comparison of hypermethylated and hypomethylated sites with chromatin states and open chromatin regions. A) The average number of hypermethylated and B) hypomethylated regions across tissues that overlap with chromatin states containing regulatory elements. C) The number of hypermethylated and hypomethylated regions that overlap with open chromatin denoted by ATAC-seq peaks across tissues.

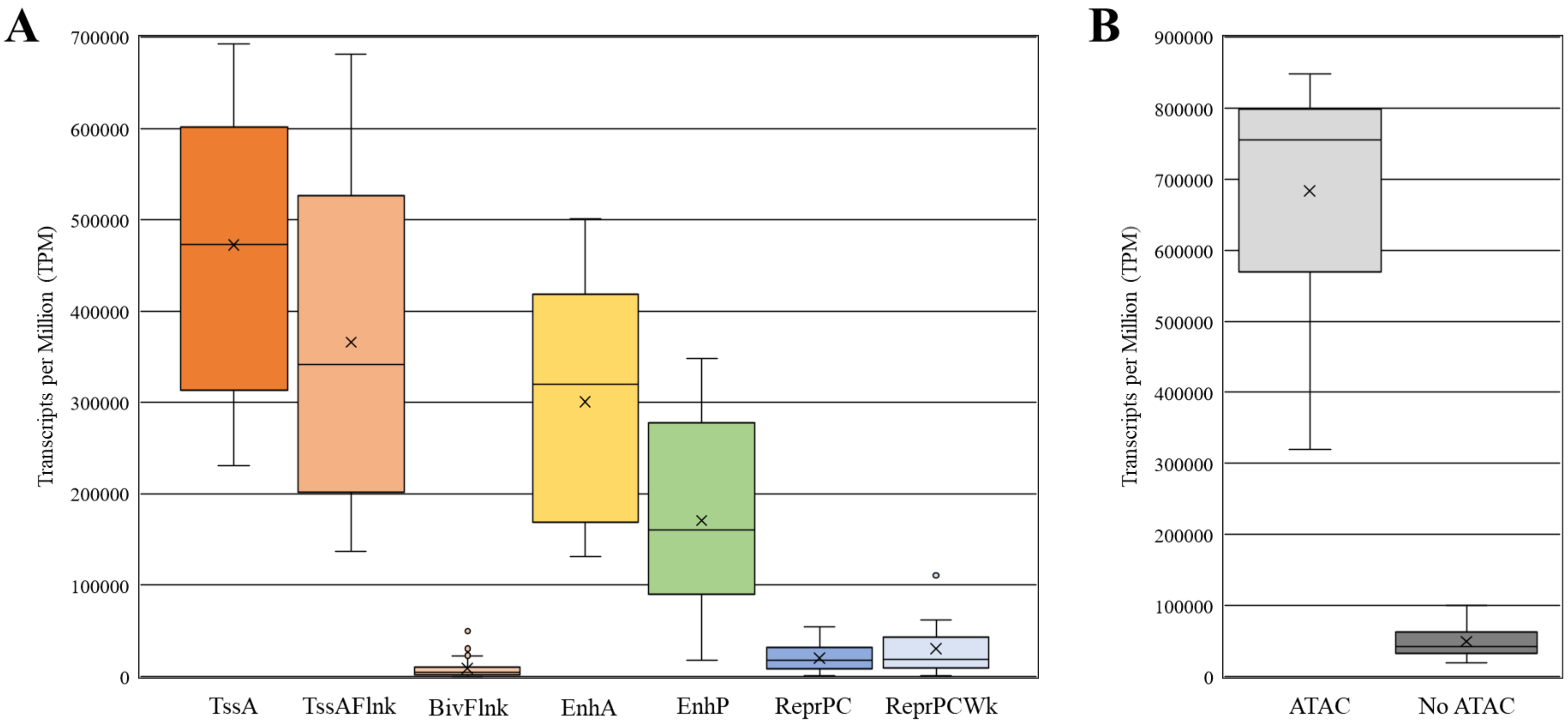
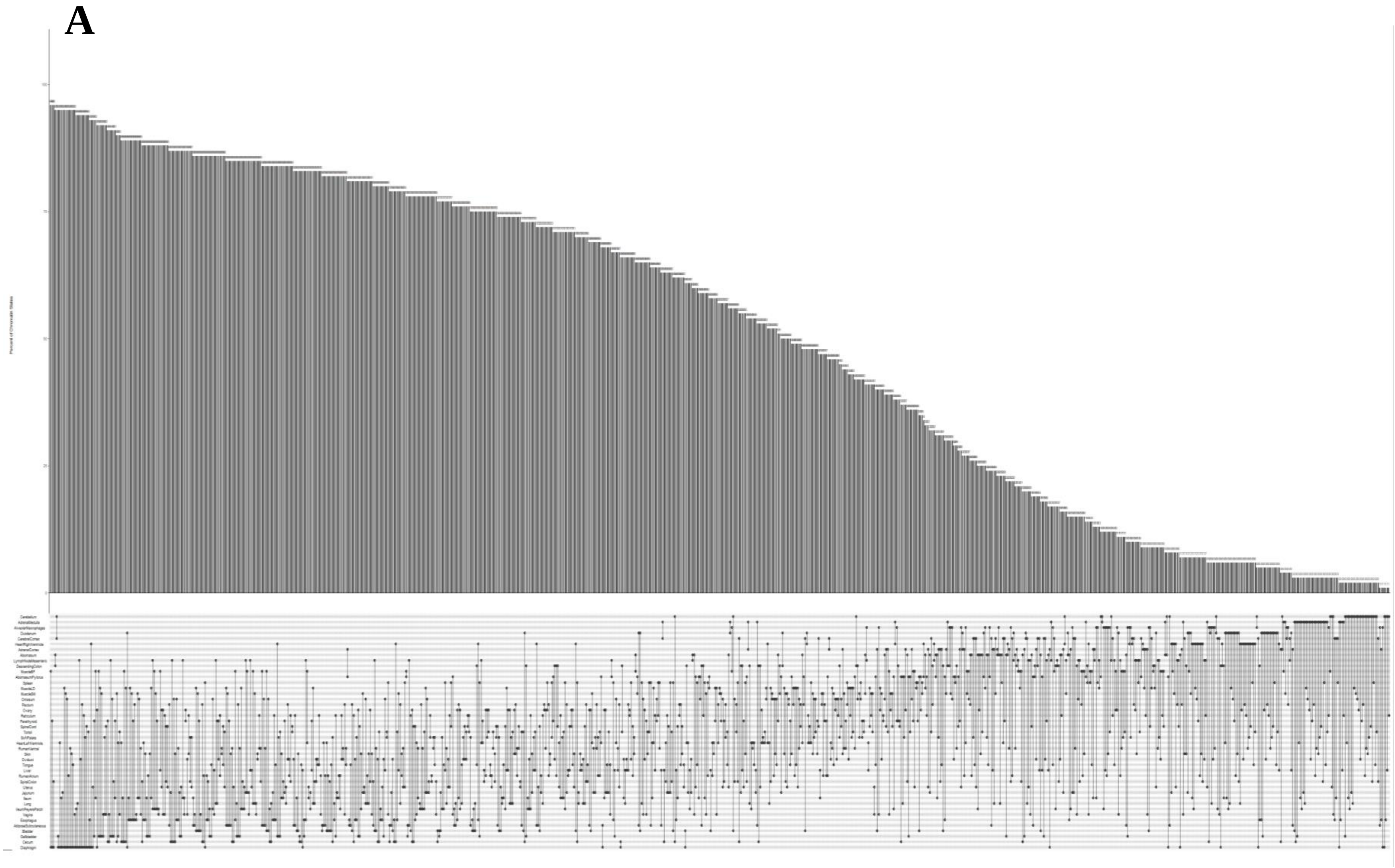
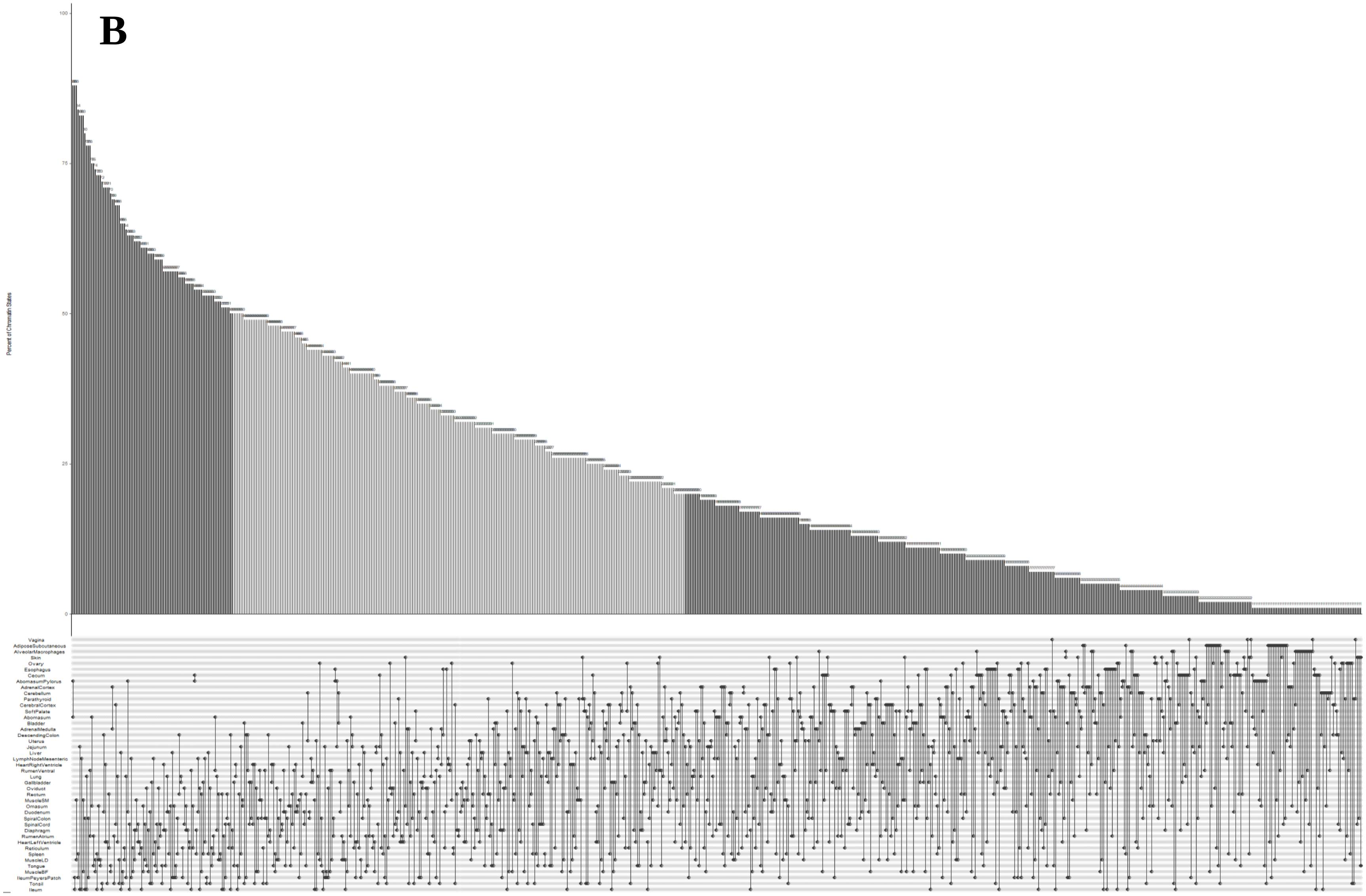
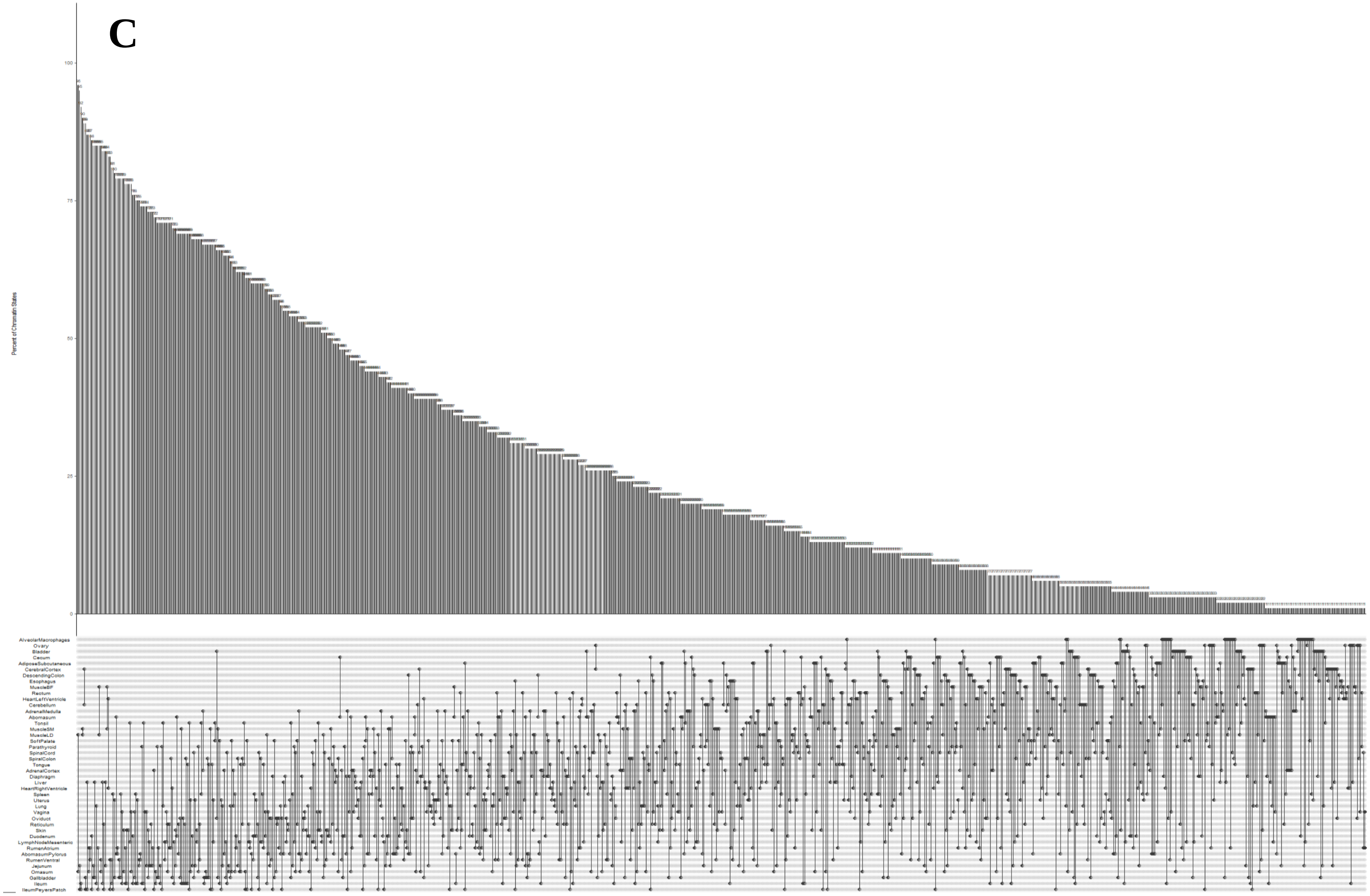


Figure 6.6: Transcript per million (TPM) counts in A) different chromatin states and B) in open chromatin regions denoted by ATAC-seq peaks (ATAC) and regions without ATAC-seq peaks (No ATAC).


A
Percent of Chromatin States
100
75
50
25
0
Cerebellum
AdrenalMedulla
AlveolarMacrophages
Duodenum
CerebralCortex
HeartRightVentricle
AdrenalCortex
Abomasum
LymphNodeMesenteric
DescendingColon
MuscleBF
AbomasumPylorus
Spleen
MuscleLD
MuscleSM
Omasum
Rectum
Ovary
Reticulum
Parathyroid
SpinalCord
Tonsil
SoftPalate
HeartLeftVentricle
RumenVentral
Skin
Oviduct
Tongue
Liver
RumenAtrium
SpiralColon
Uterus
Jejunum
Ileum
Lung
IleumPeyersPatch
Vagina
Esophagus
AdiposeSubcutaneous
Bladder
Gallbladder
Cecum
Diaphragm

B
100
75
50
25
0
Percent of Chromatin States
Vagina
AdiposeSuboutaneous
AlveolarMacrophages
Skin
Ovary
Esophagus
Ceoum
AbomasumPylorus
AdrenalCortex
Cerebellum
Parathyroid
CerebralCortex
SoftPalate
Abomasum
Bladder
AdrenalMedulla
DescendingColon
Uterus
Jejunum
Liver
LymphNodeMesenteric
HeartRightVentricle
RumenVentral
Lung
Gallbladder
Oviduct
Rectum
MuscleSM
Omasum
Duodenum
SpiralColon
SpinalCord
Diaphragm
RumenAtrium
HeartLeftVentricle
Reticulum
Spleen
MuscleLD
Tongue
MuscleBF
IleumPeyersPatch
Tonsil
Ileum


C
Percent of Chromatin States
100
75
50
25
0
AlveolarMacrophages
Ovary
Bladder
Cecum
AdiposeSubcutaneous
CerebralCortex
DescendingColon
Esophagus
MuscleBF
Rectum
HeartLeftVentricle
Cerebellum
AdrenalMedulla
Abomasum
Tonsil
MuscleSM
MuscleLD
SoftPalate
Parathyroid
SpinalCord
SpiralColon
Tongue
AdrenalCortex
Diaphragm
Liver
HeartRightVentricle
Spleen
Uterus
Lung
Vagina
Oviduct
Reticulum
Skin
Duodenum
LymphNodeMesenteric
RumenAtrium
AbomasumPylorus
RumenVentral
Jejunum
Omasum
Gallbladder
Ileum
IleumPeyersPatch

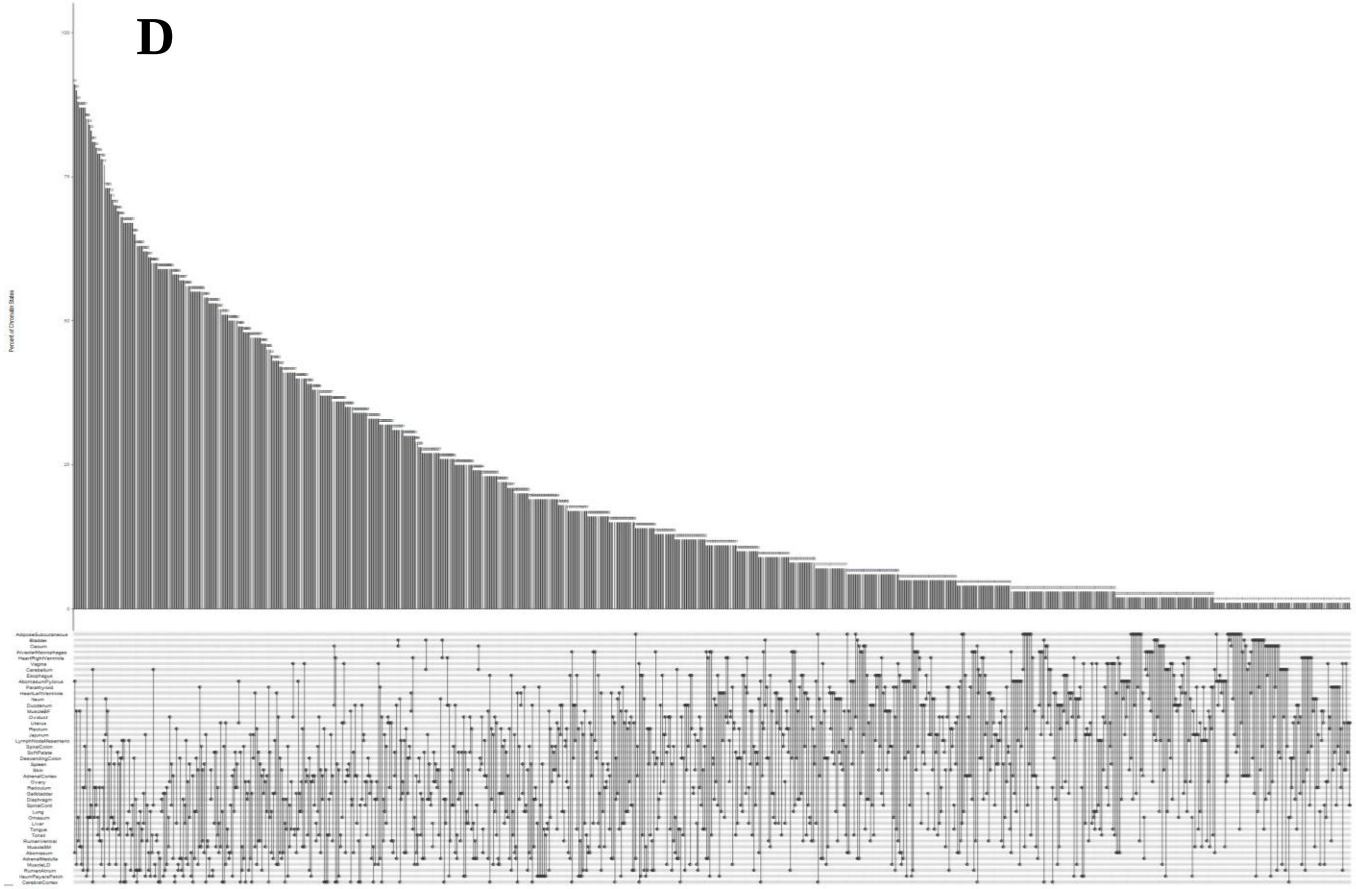


Figure 6.7: Comparison of A) promoter, B) active enhancer, C) poised enhancer, and D) repressed enhancer chromatin states across tissues. The tissues are listed and each pairwise overlap is represented by a dot and line. The percent of chromatin states in common is represented by the bar chart value.

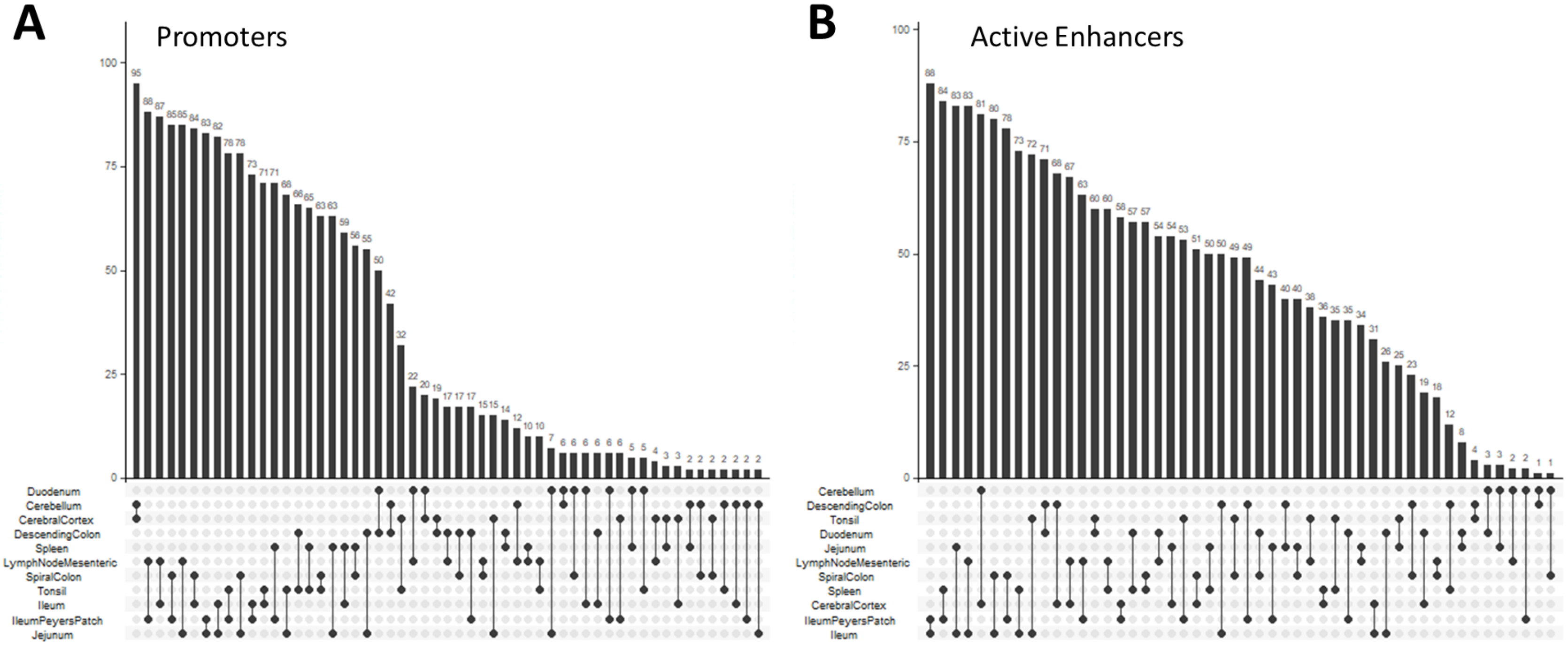


Figure 6.8: Comparison of A) promoter and B) active enhancer chromatin states in a subset of brain tissues, GI tissues, and tissues with immune related function. The bars indicate the percent similarity between tissues and the dots with lines indicate which tissues are being compared. Tissue comparisons are sorted to display the tissues with the most in common to the least in common for each chromatin state.

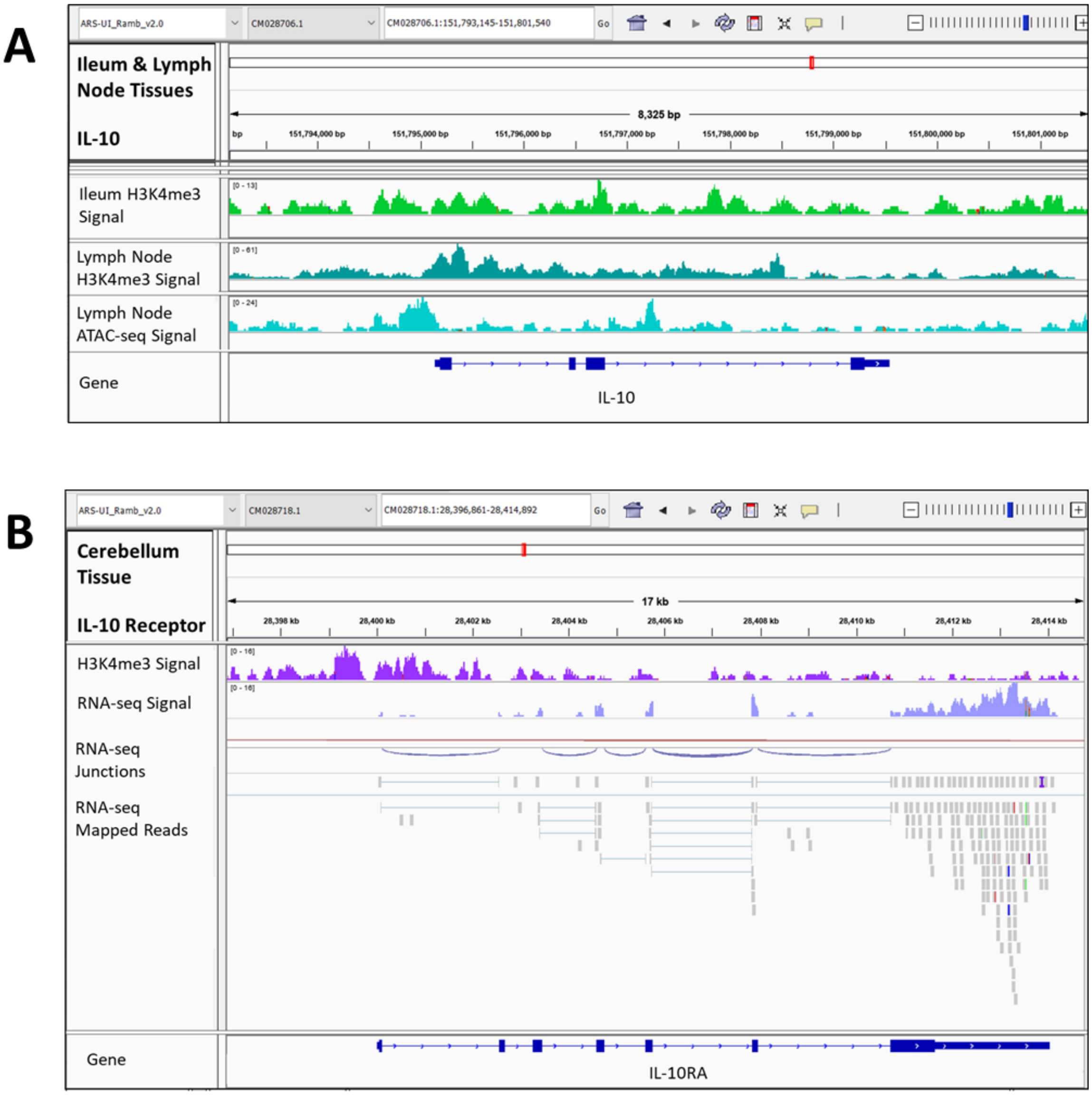


Figure 6.9: Screenshots from the Integrative Genomics Viewer (IGV) in the A) IL-10 gene for H3K4me3 signal in ileum and lymph node mesenteric tissues, and ATAC-seq signal in lymph node tissue, and B) the IL-10 receptor alpha in cerebellum tissue H3K4me3 signal and RNA-seq signal. This cytokine is known to be involved with signaling between the GI, immune system, and brain.

Supplementary Material

Supplementary Table 6.1: List of tissues used in each assay.

Assay	*Tissues*	*Total number of tissues*
Chromatin immunoprecipitation with sequencing (ChIP-seq)	Abomasum	47
	Abomasum Pylorus	
	Adipose Subcutaneous	
	Adrenal Cortex	
	Adrenal Medulla	
	Alveolar Macrophages	
	Bladder	
	Cecum	
	Cerebellum	
	Cerebral Cortex	
	Descending Colon	
	Diaphragm	
	Duodenum	
	Esophagus	
	Gallbladder	
	Heart Left Ventricle	
	Heart Right Atrium	
	Heart Right Ventricle	
	Ileum	
	Ileum Peyers Patch	
	Jejunum	
	Kidney Cortex	
	Kidney Medulla	
	Liver	
	Lung	
	Lymph Node Mesenteric	
	Mammary	
	Muscle BF (biceps femoris)	
	Muscle LD (longissimus dorsi)	
	Muscle SM (semimembranosus)	
	Omasum	
	Ovary	
	Oviduct	
	Parathyroid	
	Rectum	
	Reticulum	
	Rumen Atrium	
	Rumen Ventral	
	Skin	
	Soft Palate	
	Spinal Cord	
	Spiral Colon	
	Spleen	
	Tongue	
	Tonsil	
	Uterus	
	Vagina	

Assay for transposase accessible chromatin with sequencing (ATAC-seq)	Adipose Subcutaneous	33
	Adrenal Cortex	
	Adrenal Medulla	
	Bladder	
	Cerebellum	
	Cerebral Cortex	
	Diaphragm	
	Duodenum	
	Esophagus	
	Gallbladder	
	Heart Left Ventricle	
	Heart Right Atrium	
	Heart Right Ventricle	
	Ileum	
	Ileum Peyer’s Patch	
	Jejunum	
	Kidney Cortex	
	Kidney Medulla	
	Liver	
	Lung	
	Lymph Node Mesenteric	
	Muscle BF (biceps femoris)	
	Muscle LD (longissimus dorsi)	
	Muscle SM (semimembranosus)	
	Ovary	
	Rectum	
	Reticulum	
	Soft Palate	
	Spinal Cord	
	Spiral Colon	
	Spleen	
	Tongue	
	Tonsil	
Reduced representation bisulfite sequencing (RRBS)	Abomasum	51
	Abomasum Pylorus	
	Adipose Subcutaneous	
	Adrenal Cortex	
	Adrenal Medulla	
	Alveolar Macrophages	
	Atrioventricular valve, left	
	Bladder	
	Cecum	
	Cerebellum	
	Descending Colon	
	Diaphragm	
	Duodenum	
	Esophagus	
	Eye (retina)	
	Gallbladder	
	Heart Right Atrium	
	Heart Right Ventricle	
	Ileum	
	Ileum Peyer’s Patch	
	Jejunum	
	Kidney Cortex	
	Kidney Medulla	

	Liver	
	Lung	
	Lymph Node Mesenteric	
	Mammary	
	Muscle BF (biceps femoris)	
	Muscle LD (longissimus dorsi)	
	Muscle SM (semimembranosus)	
	Omasum	
	Ovary	
	Pituitary	
	Pons	
	Rectum	
	Reticulum	
	Rumen Atrium	
	Rumen Ventral	
	Skin	
	Soft Palate	
	Spinal Cord	
	Spiral Colon	
	Spleen	
	Thalamus	
	Thyroid	
	Tongue	
	Tonsil	
	Uterus	
	Vagina	
	Vena Cava (heart)	
Whole genome bisulfite sequencing (WGBS)	Alveolar Macrophages	8
	Cerebral Cortex	
	Cerebellum	
	Lung	
	Muscle BF (biceps femoris)	
	Muscle LD (longissimus dorsi)	
	Ovary	
	Rumen Atrium	
RNA sequencing (RNA-seq)	Abomasum	60
	Abomasum Pylorus	
	Adrenal Cortex	
	Adrenal Medulla	
	Alveolar Macrophages	
	Bladder	
	Caudal Vena Cava	
	Cecum	
	Cerebellum	
	Cerebral Cortex	
	Descending Colon	
	Diaphragm	
	Duodenum	
	Esophagus	
	Gallbladder	
	Heart Left Ventricle	
	Heart Right Atrium	
	Heart Right Ventricle	
	Hippocampus	
	Ileum	
	Ileum Peyer’s Patch	

Jejunum
Kidney Cortex
Kidney Medulla
Left Atrioventricular Valve
Liver
Lung
Lymph Node Mandibular
Lymph Node Mesenteric
Lymph Node Prescapular
Mammary
Muscle BF (biceps femoris)
Muscle LD (longissimus dorsi)
Muscle SM (semimembranosus)
Muscle SS (supraspinatus)
Omasum
Ovary
Oviduct
Pituitary
Pons
Rectum
Reticulum
Retina (eye)
Rumen Atrium
Rumen Ventral
Skin
Soft Palate
Spinal Cord
Spiral Colon
Spleen
Thalamus
Thyroid
Tongue
Tonsil
Ureter
Urethra
Uterus
Vagina

Supplementary Table 6.2: Mapping statistics for ATAC-seq data of the 33 tissues collected from Benz 2616.

Sample	***Number of reads sequenced***	***Number of reads mapped***	***% of total reads mapped***	***% uniquely mapped***	***Proportion of duplication***
Jejunum	58,923,671	57,273,449	97.20	64.21	0.12
LymphNodeMesenteric	57,974,995	57,132,608	98.55	64.14	0.11
Reticulum	60,334,457	20,356,093	33.74	44.14	0.15
CerebralCortex	60,81,0682	60,065,604	98.77	59.58	0.25
Cerebellum	75,8116,39	74,770,240	98.63	64.69	0.23
Tongue	61,639,525	58,490,545	94.89	68.25	0.15
Tonsil	64,654,489	62,090,200	96.03	66.62	0.17
SoftPalate	77,743,714	75,142,117	96.65	71.25	0.20
Duodenum	77,394,800	76,075,603	98.30	65.76	0.15
Ileum	70,784,691	68,005,758	96.07	65.73	0.14
IleumPeyersPatch	66,399,065	64,356,660	96.92	65.70	0.14
SpiralColon	74,221,101	71,816,346	96.76	70.21	0.15
Rectum	72,777,399	69,601,828	95.64	65.56	0.17
Gallbladder	60,464,950	59,717,929	98.76	65.10	0.15
Liver	71,684,845	70,834,466	98.81	55.45	0.33
Spleen	62,928,779	62,067,418	98.63	68.05	0.15
AdrenalCortex	59,283,162	58,499,859	98.68	66.92	0.18
AdrenalMedulla	67,598,623	66,834,629	98.87	61.81	0.21
KidneyCortex	57,444,492	56,769,574	98.83	63.43	0.25
KidneyMedulla	58,136,547	57,469,910	98.85	63.06	0.17
Bladder	71,670,850	70,879,177	98.90	68.37	0.16
Ovary	60,900,338	60,009,863	98.54	66.59	0.15
Lung	84,617,326	83,508,578	98.69	66.86	0.14
HeartLeftVentricle	53,483,251	52,144,748	97.50	61.32	0.27
HeartRightAtrium	46,108,019	45,436,348	98.54	70.32	0.13
HeartRightVentricle	58,664,953	57,538,982	98.08	50.80	0.34
Diaphragm	53,887,012	52,943,272	98.25	67.93	0.22
Esophagus	52,795,396	49,668,705	94.08	63.51	0.18
MuscleLD	51,359,597	49,879,210	97.12	69.29	0.19
AdiposeSubcutaneous	61,223,902	58,979,513	96.33	69.40	0.20
MuscleSM	55,760,963	54,750,250	98.19	75.83	0.20
MuscleBF	49,700,991	48,982,274	98.55	70.18	0.21
SpinalCord	57,003,960	56,175,834	98.55	63.33	0.22

Supplementary Table 6.3: Mapping statistics for ChIP-seq data from the 47 tissues collected from Benz 2616.

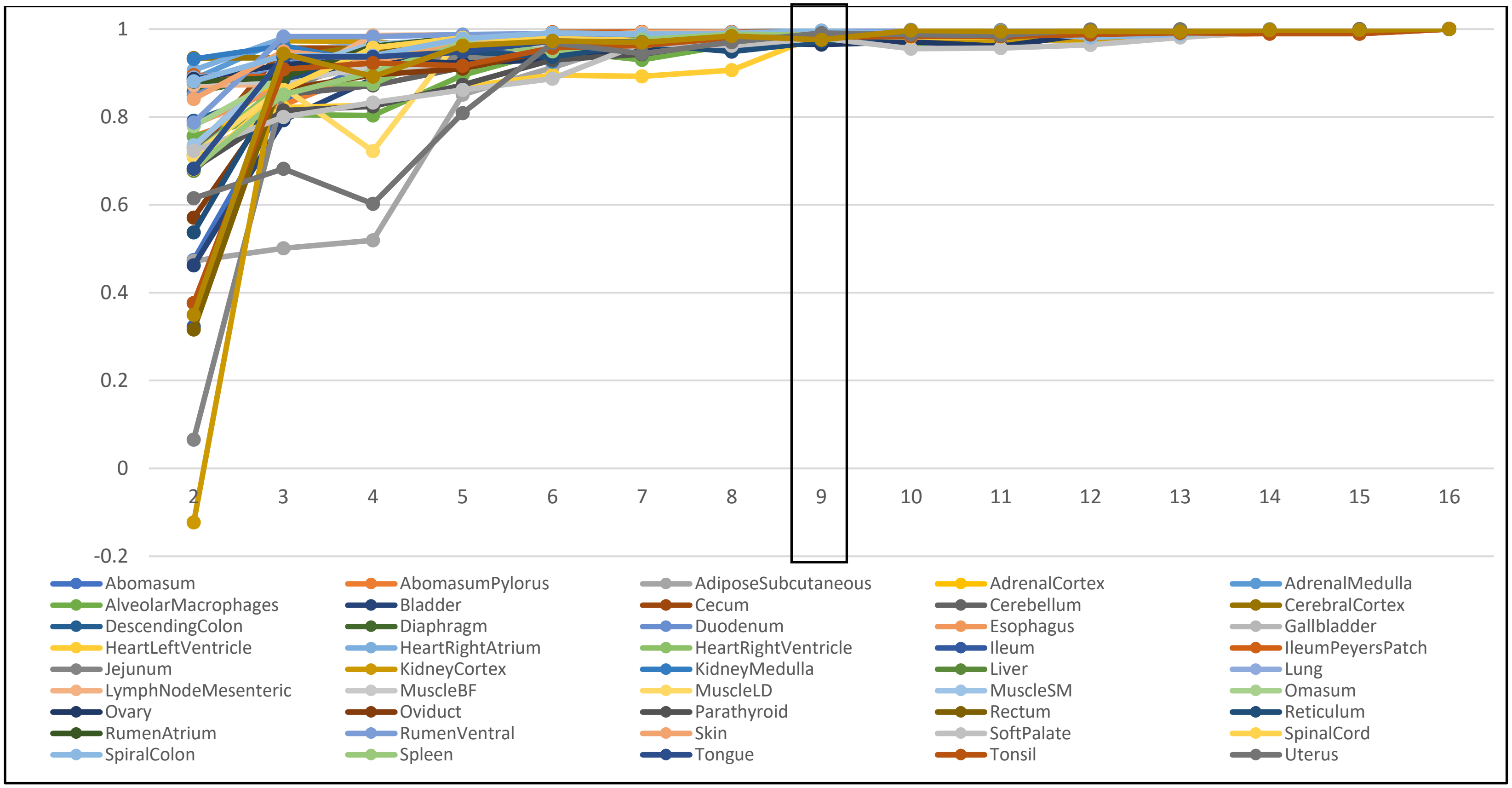
Tissue	**Number of reads sequenced**	**Number of reads mapped**	**% of reads mapped**	**Number of reads uniquely mapped**	**Proportion of duplication**
H3K4me3					
Abomasum	59533622	58932445	98.99	47001795	0.28
AbomasumPylorus	37434682	37087853	99.07	29554682	0.04
AdiposeSubcutaneous	56656894	55826911	98.54	44730618	0.42
AdrenalCortex	72064162	70514810	97.85	56894656	0.31
AdrenalMedulla	46689098	46262245	99.09	36861043	0.13
AlveolarMacrophages	30951270	30521775	98.61	24436028	0.09
Bladder	70067972	69191745	98.75	55318664	0.19
Cecum	50837904	50388306	99.12	40136526	0.27
Cerebellum	77098054	76112385	98.72	60868914	0.35
CerebralCortex	61032508	60451086	99.05	48185166	0.03
DescendingColon	73286752	72728850	99.24	57859891	0.03
Diaphragm	46844792	45862285	97.9	36983964	0.03
Duodenum	40957636	40520871	98.93	32336054	0.03
Esophagus	58783200	58171482	98.96	46409337	0.03
Gallbladder	51243926	50728754	98.99	40457080	0.03
HeartLeftVentricle	40401084	39921713	98.81	31896656	0.43
HeartRightAtrium	51325062	50742736	98.87	40521137	0.31
HeartRightVentricle	50805426	49915396	98.25	40110884	0.13
Ileum	51676892	51255811	99.19	40798907	0.02
IleumPeyersPatch	39783460	39455614	99.18	31409042	0.03
Jejunum	56069742	55571484	99.11	44267062	0.04
KidneyCortex	60322406	59927416	99.35	47624540	0.97
KidneyMedulla	66534074	65955572	99.13	52528652	0.42
Liver	46909948	46555377	99.24	37035404	0.05
Lung	59648666	58912259	98.77	47092622	0.22
LymphNodeMesenteric	77173236	75831918	98.26	60928270	0.05
Mammary	89105700	87332202	98.01	70348951	0.72
MuscleBF	60039886	59431645	98.99	47401490	0.06
MuscleLD	45753170	45300644	99.01	36122128	0.04
MuscleSM	38027324	37576259	98.81	30022573	0.04
Omasum	57752794	57194470	99.03	45595831	0.04
Ovary	92050280	90648492	98.48	72673697	0.42
Oviduct	86746398	85192426	98.21	68486282	0.52
Parathyroid	57395074	56216784	97.95	45313411	0.06
Rectum	54282262	53690878	98.91	42855846	0.15
Reticulum	42557644	41865807	98.37	33599260	0.41

RumenAtrium	64859256	64079416	98.8	51206383	0.06
RumenVentral	41381154	40805783	98.61	32670422	0.05
Skin	67256388	66340975	98.64	53098919	0.15
SoftPalate	36778256	36412893	99.01	29036434	0.08
SpinalCord	85787896	83900354	97.8	67729544	0.47
SpiralColon	56129794	55042234	98.06	44314473	0.04
Spleen	38116320	37868834	99.35	30092835	0.05
Tongue	77948428	76024847	97.53	61540284	0.08
Tonsil	68655640	67526875	98.36	54203628	0.12
Uterus	103177318	101827096	98.69	81458493	0.34
Vagina	62771856	61525806	98.01	49558381	0.77
		H3K27ac			
Abomasum	61525992	60873901	98.94	47904138	0.25
AbomasumPylorus	52023452	51494086	98.98	40505460	0.03
AdiposeSubcutaneous	52673024	51892855	98.52	41011217	0.13
AdrenalCortex	105187854	103128275	98.04	81899264	0.3
AdrenalMedulla	31259430	30959973	99.04	24338593	0.04
AlveolarMacrophages	74455406	73465952	98.67	57970980	0.15
Bladder	64037774	63349746	98.93	49859811	0.11
Cecum	49640298	49119125	98.95	38649937	0.02
Cerebellum	51075184	50614197	99.1	39767139	0.04
CerebralCortex	41582644	41150821	98.96	32376247	0.03
DescendingColon	32581578	32267865	99.04	25368017	0.02
Diaphragm	48419952	47263481	97.61	37699775	0.03
Duodenum	45165746	44765291	99.11	35166050	0.02
Esophagus	51521138	50989724	98.97	40114359	0.02
Gallbladder	50678474	50087840	98.83	39458260	0.05
HeartLeftVentricle	31946058	31394126	98.27	24873201	0.06
HeartRightAtrium	70360882	69658927	99	54782983	0.14
HeartRightVentricle	47110534	46544770	98.8	36680262	0.05
Ileum	60755496	60231268	99.14	47304230	0.03
IleumPeyersPatch	45490366	45138860	99.23	35418799	0.03
Jejunum	56285096	55836090	99.2	43823576	0.04
KidneyCortex	47826524	47542241	99.41	37237732	0.97
KidneyMedulla	52481548	52071654	99.22	40862134	0.47
Liver	46800868	46469245	99.29	36439156	0.79
Lung	55309844	54767893	99.02	43064245	0.04
LymphNodeMesenteric	86197870	84564990	98.11	67113662	0.3
MuscleBF	49722910	49114925	98.78	38714258	0.34
MuscleLD	55047960	54457680	98.93	42860342	0.07
MuscleSM	57891660	57043165	98.53	45074447	0.03
Omasum	52753408	52204768	98.96	41073804	0.05
Ovary	61581182	60773006	98.69	47947109	0.19
Oviduct	49538666	48983692	98.88	38570806	0.13

Parathyroid	47636396	46759286	98.16	37089698	0.06
Rectum	34995448	34560773	98.76	27247456	0.06
Reticulum	53353414	52531719	98.46	41540969	0.09
RumenAtrium	60438186	59680848	98.75	47057172	0.02
RumenVentral	65001844	63678260	97.96	50610436	0.04
Skin	42063106	41673000	99.07	32750335	0.06
SoftPalate	75474156	74735450	99.02	58764178	0.13
SpinalCord	95475094	93377408	97.8	74336909	0.19
SpiralColon	77075588	75546231	98.02	60011053	0.03
Spleen	54814518	54474283	99.38	42678584	0.32
Tongue	76198678	74142192	97.3	59328291	0.06
Tonsil	37376302	36746566	98.32	29101189	0.11
Uterus	56064144	55539477	99.06	43651543	0.21
Vagina	47372536	46886259	98.97	36884257	0.1
			H3K4me1		
Abomasum	97536006	95601830	98.02	75941535	0.32
AbomasumPylorus	81100114	80059174	98.72	63144549	0.03
AdiposeSubcutaneous	63216732	62227020	98.43	49220548	0.23
AdrenalCortex	108708406	106908934	98.34	84640365	0.34
AdrenalMedulla	73514844	72755785	98.97	57238658	0.04
AlveolarMacrophages	118087904	116806862	98.92	91943243	0.15
Bladder	64475714	63824692	98.99	50200791	0.06
Cecum	51963396	51545463	99.2	40458701	0.02
Cerebellum	65057816	64406075	99	50654016	0.04
CerebralCortex	69534344	68889583	99.07	54139441	0.03
DescendingColon	65655046	65138908	99.21	51119019	0.02
Diaphragm	52173962	50766598	97.3	40622647	0.03
Duodenum	75719558	74765287	98.74	58955248	0.02
Esophagus	60258028	59578130	98.87	46916901	0.02
Gallbladder	67452112	66716401	98.91	52518215	0.03
HeartLeftVentricle	128732094	126263033	98.08	100230809	0.04
HeartRightAtrium	52573384	51993355	98.9	40933637	0.08
HeartRightVentricle	55993524	54965823	98.16	43596558	0.05
Ileum	60576812	59967180	98.99	47165106	0.02
IleumPeyersPatch	60301648	59695419	98.99	46950864	0.02
Jejunum	59001800	58507839	99.16	45938802	0.04
KidneyCortex	48635658	48308581	99.33	37867724	0.97
KidneyMedulla	88441658	87630674	99.08	68860675	0.24
Liver	47296152	46913756	99.19	36824784	0.16
Lung	64530968	63793887	98.86	50243812	0.06
LymphNodeMesenteric	68676656	67047026	97.63	53471645	0.05
Mammary	63370290	62818257	99.13	49340108	0.27
MuscleBF	55970974	55310927	98.82	43579001	0.04
MuscleLD	72990418	71959538	98.59	56830340	0.08

MuscleSM	69918370	68649963	98.19	54438443	0.03
Omasum	63801906	63262176	99.15	49676165	0.03
Ovary	58315174	57288890	98.24	45404195	0.1
Oviduct	51037720	50516940	98.98	39737969	0.09
Parathyroid	101598294	99930545	98.36	79104432	0.06
Rectum	64257610	63369112	98.62	50030976	0.05
Reticulum	74595696	73468848	98.49	58080209	0.1
RumenAtrium	51989214	51423328	98.91	40478803	0.03
RumenVentral	82330202	81275849	98.72	64102296	0.05
Skin	68816764	68109441	98.97	53580733	0.04
SoftPalate	77709902	76939784	99.01	60504930	0.23
SpinalCord	48173990	47594536	98.8	37508269	0.14
SpiralColon	76521306	75262605	98.36	59579489	0.04
Spleen	50754512	50299523	99.1	39517464	0.1
Tongue	76999506	75157273	97.61	59951816	0.04
Tonsil	52064832	51367494	98.66	40537679	0.12
Uterus	65614474	65037724	99.12	51087430	0.1
Vagina	97304122	96129915	98.79	75760990	0.51
	H3K27me3				
Abomasum	76046116	75144060	98.81	59148670	0.48
AbomasumPylorus	82484288	81352012	98.63	64156280	0.05
AdiposeSubcutaneous	86178814	84117851	97.61	67029882	0.33
AdrenalCortex	65023764	63698597	97.96	50575484	0.09
AdrenalMedulla	66530654	65838443	98.96	51747543	0.13
AlveolarMacrophages	72610458	71649415	98.68	56476415	0.1
Bladder	58157152	57492336	98.86	45234633	0.09
Cecum	75462276	64340305	85.26	58694559	0.04
Cerebellum	92154006	91090298	98.85	71677386	0.14
CerebralCortex	72437098	71643242	98.9	56341575	0.07
DescendingColon	81747298	80512941	98.49	63583049	0.05
Diaphragm	76106572	74723187	98.18	59195692	0.03
Duodenum	56424992	55875539	99.03	43887359	0.02
Esophagus	77236326	76448704	98.98	60074415	0.02
Gallbladder	53792356	53351957	99.18	41839695	0.03
HeartLeftVentricle	107180636	105908364	98.81	83365099	0.11
HeartRightAtrium	61201878	60673183	99.14	47602821	0.06
HeartRightVentricle	57090060	56588332	99.12	44404649	0.32
Ileum	59207088	58680045	99.11	46051274	0.03
IleumPeyersPatch	63005144	62260977	98.82	49005402	0.04
Jejunum	87742980	62000554	70.66	68246490	0.06
KidneyCortex	60692712	60035327	98.92	47206792	0.08
KidneyMedulla	64335354	63859899	99.26	50040039	0.77
Liver	66836674	66207847	99.06	51985566	0.93
Lung	81070052	80297142	99.05	63056287	0.06

LymphNodeMesenteric	55783150	54543848	97.78	43388135	0.15
Mammary	59981452	58889130	98.18	46653574	0.78
MuscleBF	75574726	74725266	98.88	58782022	0.1
MuscleLD	60136916	59533949	99	46774494	0.04
MuscleSM	75474572	74245908	98.37	58704123	0.03
Omasum	68665934	67490789	98.29	53408364	0.03
Ovary	55876452	55116628	98.64	43460705	0.15
Oviduct	68661448	67615291	98.48	53404875	0.17
Parathyroid	96798568	95277580	98.43	75289927	0.15
Rectum	133364284	131378982	98.51	103730741	0.42
Reticulum	61975694	61207986	98.76	48204695	0.08
RumenAtrium	344114132	337553106	98.09	267651972	0.08
RumenVentral	87861134	86269784	98.19	68338391	0.04
Skin	86717954	85690942	98.82	67449225	0.09
SoftPalate	76995168	76129394	98.88	59886842	0.29
SpinalCord	61375616	60288328	98.23	47737955	0.42
SpiralColon	76570534	75239920	98.26	59556562	0.14
Spleen	61361852	60796383	99.08	47727249	0.72
Tongue	42811994	41934050	97.95	33299169	0.05
Tonsil	82072404	81185674	98.92	63835916	0.69
Uterus	66946702	66354629	99.12	52071145	0.29
Vagina	57770606	57196729	99.01	44933978	0.15



Supplementary Figure 6.1: Median correlation of each chromatin state to 16 states across tissues. At 9 states, the average median correlation from all tissues reached above 0.98 and was chosen as the optimal number of states.

Chapter 7: Conclusion

Sheep are a globally important species raised for meat, milk, and wool. The continued improvement of sheep production will rely on genomic technologies for selection of animals that exhibit desirable traits earlier in life. A brief overview of sheep production in the United States and research in genetics and genomics is presented in the first chapter of this dissertation. The selection and adaptation of sheep to diverse production systems and environments result in breed specialization and differences in phenotypic traits. Global studies across breeds reveal that sheep from similar lineages, such as those selected for meat or wool, exhibit greater genetic relatedness when compared with sheep selected for a very different trait. Sheep within breeds also show genetic divergence based on location and regional selection. The second chapter of this dissertation characterizes the relationships between sheep breeds in the United States compared with similar breed lineages from across the world as part of the Sheep HapMap Project. This study observes genetic differentiation between meat and wool breeds as well as regional genetic differences in the Suffolk breed. Sheep within the same breed, such as Suffolk and Rambouillet, are genetically distinct depending on geographic locations. This study aids sheep researchers and the sheep industry in understanding genetic differences across breeds of sheep in the United States and assessing applicability of genomic technologies developed in one breed to another breed based on genetic relatedness.

Research in sheep genetics to improve sheep production is reliant on an accurate reference genome to assess genetic variation and define the locations of genes and regulatory elements. The rapid progress in sequencing technologies, including the advent of next-generation sequencing, paired with enhanced computational resources allow for updated and improved genome assemblies. Genome assemblies, including mitochondrial assemblies, also aid in discerning phylogeny and relatedness between wild and domestic *Ovis* species, such as bighorn sheep (*Ovis canadensis*) and domestic sheep (*Ovis aries*). The third chapter of this dissertation describes the first mitochondrial genome assembly of a Rocky Mountain Bighorn Sheep from the United States. This study provides a valuable resource for phylogenetic and comparative studies between the bighorn sheep and other species. Mitochondrial genome assemblies from domestic sheep are often included with the whole genome assembly releases. The fourth chapter of this dissertation introduces a genome assembled from a Rambouillet ewe selected as part of the functional annotation of animal genomes project. This updated genome features vast improvements in contiguity and quality when compared with previous sheep genome assemblies and is comparable in quality with other livestock reference genomes. The updated Rambouillet genome, *ARS-UI_Ramb_v2.0,* will serve as the reference for the sheep species and provide a valuable resource for the scientific community to investigate genetic relationships to traits of interest using a more accurate reference.

The improved quality and contiguity of the most recent sheep reference genome offers more accurate locations of genes and the opportunity to more accurately define the locations of genetic regulatory elements across different tissues. Regulatory elements are known to influence gene transcription contain genetic variation that can influence phenotypes. Defining regulatory elements throughout the genome will provide the resources for the functional annotation of the sheep genome and facilitate further research in the influence of genetic regulatory elements on traits of interest in sheep. The fifth chapter of this dissertation describes the locations of histone modifications and DNA methylation in sheep liver, spleen, and cerebellum tissues. The study identifies tissue specific chromatin states and depicts the overlay of active promoter and enhancer states with hypomethylated regions, and conversely repressed and poised states with hypermethylated regions. The protocols and analyses pipelines developed in this study are used in the larger study of genetic regulatory elements in the sixth chapter of this dissertation. This study characterizes histone modifications, open chromatin, DNA methylation, transcription start sites, and transcript expression across almost 50 tissues collected from the same Rambouillet sheep used in the reference genome assembly. Active promoter and enhancer regions across tissues overlay with transcript expression, open chromatin, and hypomethylated sites. The repressed and poised enhancer states contain hypermethylated regions and are not present in open chromatin with greater transcript expression. Similarities and differences in genetic regulatory elements, particularly active regions, also allude to the physiological and regulatory relationship between the gut, brain, and immune tissues.

The studies in this dissertation describe improvements to efforts in sheep genomics research to further understand relationships between genetics and phenotypes of interest to the sheep industry. Examining population genetics and genetic relatedness of sheep, assembling the sheep reference genome, and defining genetic regulatory elements in the sheep genome provide valuable information and resources for the improvement of sheep research and production. The sheep industry will rely on genomic information for greater improvement of meat, wool, and milk production to feed and clothe a growing global population. This research contributes to this effort in several different aspects of sheep genetics and genomics research.