

ASSESSMENT OF STATE-SPECIFIC DNA METHYLATION PATTERNS TO IMPROVE
FUNCTIONAL ANNOTATION OF FARM ANIMAL GENOMES

By

Ryan James Corbett

A DISSERTATION

Submitted to
Michigan State University
in partial fulfillment of the requirements
for the degree of

Genetics—Doctor of Philosophy

2021

ABSTRACT

ASSESSMENT OF STATE-SPECIFIC DNA METHYLATION PATTERNS TO IMPROVE FUNCTIONAL ANNOTATION OF FARM ANIMAL GENOMES

By

Ryan James Corbett

Over the past several decades, genetic advancements in the domestic pig (*Sus scrofa*) and other farm animal species have resulted in increased economic output and expanded use of these organisms as biomedical models to study human disease. However, limited functional annotation of the porcine genome—particularly in non-coding regulatory regions—hinders both identification of causal genes for complex traits and translational research capabilities. The Functional Annotation of Animal Genomes consortium seeks to map functional elements in domesticated animal genomes in part by performing sequencing assays to characterize the animal epigenome, as specific chromatin modifications have been shown to be predictive of regulatory regions. DNA methylation is the most ubiquitous epigenetic modification made to the DNA molecule, and in mammals occurs almost exclusively at cytosines in CpG dinucleotides. DNA methylation exerts regulatory effects through numerous mechanisms, including the occlusion of transcription factors at activating regulatory regions, and as such has been shown to play important roles in establishing spatiotemporal gene expression. Furthermore, differential methylation has been associated with genomic imprinting and stress-induced physiological changes in mammals. Assessment of DNA methylation in the pig and other farm animal species has thus far been limited in scope. In this dissertation, I have characterized state-specific DNA methylation patterns in farm animal genomes across a diverse collection of cell types, developmental stages, and environmental conditions, to enhance understanding of epigenetic gene regulation in livestock and poultry. **First**, I demonstrate that sorted porcine immune cells exhibit unique DNA methylation landscapes that are strongly

correlated with local and distal gene expression as well as binding sites for transcription factors regulating immune cell-specific functions. The co-localization of immune cell differentially methylated regions with GWAS SNPs for immune-related traits supports the use of epigenomics assays to increase functional annotation of economically relevant genomic regions. **Second**, I show that development of four porcine fetal tissues (whole brain, liver, loin muscle, and placenta) is associated with increased differentiation of DNA methylation profiles that likely contributes to tissue-specific transcriptomes and transcription factor regulatory potential. I also report widespread allele-biased methylation in fetal tissues associated with breed-specific gene regulation as well as putative regions of genomic imprinting events. **Third**, I characterize associations between environmental stimuli and DNA methylation patterns in two studies. I show that piglet weaning correlates with changes in peripheral blood mononuclear cell DNA methylation, and that increased weaning stress is associated with increased methylation and decreased expression of T cell-enriched genes, suggesting a diminished adaptive immune response. Lastly, I assess the impact of broiler chick incubation parameters on cardiac DNA methylation and observe significant temperature-associated differential methylation of genes involved in heart morphogenesis. I identified differentially methylated and expressed genes between temperature treatments that may influence environment-driven differences in cardiovascular development. In conclusion, I have performed the most expansive survey of whole-genome DNA methylation in farm animal species to date and have identified thousands of putative regulatory elements influencing state-specific gene and phenotype expression. These data will be a valuable resource for future functional annotation efforts seeking to identify mechanistic links between genetic and phenotypic variation in animal species.

This work is dedicated to my parents, Paul and Debbie, and my grandparents, Rodger and Joanne. They filled my childhood with endless zoo trips, library visits, and toy animal figurines, all of which sparked my interest in science and resulted in the production of this dissertation.

ACKNOWLEDGMENTS

I would first and foremost like to thank my advisor, Dr. Cathy Ernst, for her years of support and mentorship throughout my PhD. It is because of Cathy's tireless efforts that I have been afforded so many research and professional development opportunities over these last six years. I will always be grateful for her willingness to take me on as a graduate student, and for giving me the independence to develop new skills and grow as a scientist. I would also like to thank Nancy Raney for performing the bulk of the lab work necessary to complete my projects. I had the privilege of working alongside several graduate and undergraduate students while in the Ernst lab—Scott Funkhouser, Kaitlyn Daza, Deborah Velez-Irrizarry, Andrea Luttmann, Laura Ford, and Jenna Grabowski—and want to thank all of them for their technical and emotional support during my time in the group.

My research would not have been possible without the support of numerous collaborators. I would like to thank Ole Madsen for his mentorship during my time as a visiting scholar at Wageningen University and Research. I would also like to acknowledge several collaborators on the USDA Pig FAANG project, including Chris Tuggle, Juber Herrera-Urbe, Haibo Liu, Crystal Loving, and Dan Nonneman. As a geneticist applying my skillset to immunology and developmental biology research, I was incredibly appreciative of your expertise in these areas to help guide my projects.

Lastly, I would like to thank my family and friends who have supported me throughout this journey, including (but not limited to) Jeff Plegaria, Katerina-Lay Pruitt, Amanda Koenig, and Agustin González-Reymúndez. I couldn't have asked for a better academic and emotional support group these past few years.

TABLE OF CONTENTS

LIST OF TABLES	ix
LIST OF FIGURES	xi
CHAPTER 1 INTRODUCTION	1
1.1 Increasing power of swine genomics research through functional annotation.....	1
1.2 DNA methylation and its role in gene regulation and expression.....	3
1.3 DNA methylation patterns govern tissue- and stage-specific gene expression	7
1.4 DNA methylation as a mediator of environment-induced physiological changes	9
1.5 Project Goals & Significance.....	11
CHAPTER 2 ASSESSMENT OF DNA METHYLATION IN PORCINE IMMUNE CELLS REVEALS NOVEL REGULATORY ELEMENTS ASSOCIATED WITH CELL-SPECIFIC GENE EXPRESSION AND IMMUNE TRAITS	13
2.1 Abstract.....	13
2.2 Introduction.....	14
2.3 Materials & Methods	17
2.3.1 Blood Collection and Separation	17
2.3.2 Immune Cell Sorting from PBMCs	18
2.3.3 DNA Isolation and Bisulfite Sequencing.....	18
2.3.4 WGBS Bioinformatics Analyses	19
2.3.5 cDMR and cLMR Identification.....	20
2.3.6 Integration of Gene Methylation and Expression Data.....	20
2.3.7 Transcription Factor Binding Motif Enrichment Analyses	21
2.3.8 GWAS SNP Enrichment Analysis.....	21
2.3.9 RT-qPCR Analyses.....	22
2.4 Results.....	23
2.4.1 Porcine immune cells exhibit unique DNA methylation patterns associated with cell-specific co-receptor activation and biological processes	23
2.4.2 Immune cell differential methylation is strongly associated with differential gene expression	27
2.4.3 Cell lowly methylated regions are enriched for cell-specific transcription factor binding motifs	30
2.4.4 Immune cell differential methylation co-localizes with candidate loci for immune capacity and disease traits	32
2.5 Discussion.....	34
2.6 Acknowledgements.....	40
CHAPTER 3 PIG FETAL SKELETAL MUSCLE DEVELOPMENT IS ASSOCIATED WITH GENOME-WIDE DNA HYPOMETHYLATION AND CORRESPONDING ALTERATIONS IN TRANSCRIPT AND MICRO-RNA EXPRESSION	41
3.1 Abstract.....	41

3.2	Introduction.....	42
3.3	Materials & Methods	44
3.3.1	Sample Collection and Nucleic Acid Isolation.....	44
3.3.2	Bioinformatics Analyses.....	45
3.3.3	Methylation Analyses	45
3.3.4	Differential Expression Analyses	46
3.3.5	Motif Enrichment Analyses	47
3.3.6	GWAS SNP Enrichment Analyses	47
3.4	Results & Discussion.....	47
3.4.1	Fetal pig skeletal muscle development is associated with genome-wide and regional hypomethylation	47
3.4.2	Skeletal muscle differential methylation is enriched in gene features and micro-RNAs	50
3.4.3	Differential methylation is strongly associated with differential transcript and miRNA expression.....	52
3.4.4	Hypomethylated regions are enriched for myogenic regulatory factors binding sites	55
3.4.5	Differentially methylated regions co-localize with GWAS SNPs for muscle and meat quality traits.....	57
3.5	Conclusions.....	57
3.6	Acknowledgements.....	59

CHAPTER 4 DNA METHYLATION PATTERNS IN PIG FETAL TISSUES ARE ASSOCIATED WITH UNIQUE DEVELOPMENTAL TRAJECTORIES AND ALLELE-BIASED GENE REGULATION.....		60
4.1	Abstract.....	60
4.2	Introduction.....	61
4.3	Materials and Methods	64
4.3.1	Sample Collection.....	64
4.3.2	DNA Isolation and Bisulfite Sequencing.....	65
4.3.3	WGBS Bioinformatics	65
4.3.4	DNA Methylation Analyses.....	66
4.3.5	RNA-seq Bioinformatics and Gene Expression Analyses.....	66
4.3.6	Motif Enrichment analyses	67
4.3.7	Allele-Specific Methylation Analyses	67
4.4	Results.....	68
4.4.1	Global DNA methylation patterns are associated with tissue type and developmental stage.....	68
4.4.2	Developmental differential DNA methylation is associated with tissue- and stage-specific differential gene expression.....	72
4.4.3	Differentially methylated regions co-localize with tissue-specific transcription factor binding motifs.....	75
4.4.4	Genotype-dependent allele-biased methylation is widespread in pig fetal tissues ..	76
4.4.5	Genotype-independent allele biases in pig fetal tissues reveal putative novel imprinted loci	81
4.5	Discussion.....	84
4.6	Acknowledgements.....	89

CHAPTER 5 WEANING INDUCES STRESS-DEPENDENT DNA METHYLATION AND TRANSCRIPTIONAL CHANGES IN PIGLET PBMCS	90
5.1 Abstract.....	90
5.2 Introduction.....	91
5.3 Materials & Methods	93
5.3.1 Sample Collection.....	93
5.3.2 Nucleic Acid Isolation and Sequencing.....	93
5.3.3 WGBS Bioinformatics	94
5.3.4 RNA-sequencing Bioinformatics.....	95
5.3.5 Assessment of <i>NR3C1</i> Methylation and Expression	95
5.4 Results.....	96
5.4.1 Serum cortisol and lesion count measurements identify low- and high-stress animals following weaning	96
5.4.2 Weaning differential methylation is associated with lymphocyte immune response genes	97
5.4.3 Weaning differential gene expression is associated with differential gene methylation	100
5.4.4 <i>NR3C1</i> differential methylation and expression is indicative of altered HPA axis response.....	103
5.5 Discussion.....	104
5.6 Acknowledgements.....	107
CHAPTER 6 GENOME-WIDE ASSESSMENT OF DNA METHYLATION IN CHICKEN CARDIAC TISSUE EXPOSED TO DIFFERENT INCUBATION TEMPERATURES AND CO ₂ LEVELS	108
6.1 Abstract.....	108
6.2 Introduction.....	109
6.3 Material and Methods	111
6.3.1 Experimental Design and Sample Collection	111
6.3.2 Statistical Analyses of Heart Weight	112
6.3.3 Sample Processing and Bisulfite Sequencing.....	112
6.3.4 Bioinformatics Analyses.....	113
6.3.5 Differential Methylation Analyses.....	113
6.3.6 CpG Annotation	114
6.3.7 Motif Enrichment Analysis.....	115
6.3.8 Epigenome-Wide Association Study	115
6.3.9 Quantitative reverse-transcription PCR	115
6.4 Results.....	116
6.4.1 Broiler chicks differ significantly in HW between ESTs	116
6.4.2 RRBS reads capture predominantly lowly methylated regions of the chicken genome	117
6.4.3 EST impacts the methylation state of genes involved in general and heart-specific developmental processes.....	118
6.4.4 EST differential methylation impacts binding sites for TFs involved in cell proliferation.....	120
6.4.5 CO ₂ differential methylation impacts genes involved in heart development and response to hypoxia.....	122

6.4.6	EWAS identifies methylation signatures significantly associated with HW	124
6.4.7	RT-qPCR identifies genes with coordinated differences in gene methylation and expression between EST treatments	125
6.5	Discussion.....	126
6.6	Acknowledgments	131
CHAPTER 7	CONCLUSIONS & FUTURE DIRECTIONS.....	132
APPENDICES	138
APPENDIX A	CHAPTER 2 SUPPLEMENTARY MATERIAL.....	139
APPENDIX B	CHAPTER 3 SUPPLEMENTARY MATERIAL.....	142
APPENDIX C	CHAPTER 4 SUPPLEMENTARY MATERIAL.....	144
APPENDIX D	CHAPTER 5 SUPPLEMENTARY MATERIAL.....	147
APPENDIX E	CHAPTER 6 SUPPLEMENTARY MATERIAL	148
REFERENCES	149

LIST OF TABLES

Table 2.1: Summary of Immune Cell WGBS Libraries and Global DNA Methylation.....	23
Table 2.2: Enriched GO terms among cell lowly methylated genes.....	27
Table 2.3: SNPs within cLMRs associated with transcript abundance	33
Table 3.1: Summary of differentially methylated regions (DMRs) and genes (DMGs)	49
Table 3.2: Enriched GO terms among promoter differentially methylated genes.....	51
Table 4.1: Number of tissue-lowly methylated regions (LMRs) and enriched genes	69
Table 4.2: Enriched GO terms among promoter differentially methylated genes (DMGs)	73
Table 4.3: Top enriched GO terms among tissue breed-ABM genes	79
Table 4.4: Novel genes exhibiting imprinting-like allele methylation and expression biases.....	84
Table 5.1: Enriched GO terms among promoter- and gene body-DMGs.....	99
Table 5.2: GO Terms enriched among DMGs in low and high stress (LS & HS) groups	100
Table 5.3: Weaning differential methylation and expression of T cell co-receptor genes	101
Table 5.4: GO terms enriched among upregulated genes in low and high stress groups	102
Table 6.1: Enriched GO terms among DMGs between EST treatments	120
Table 6.2: Uniquely enriched GO terms among TFs of hypomethylated motifs.....	121
Table 6.3: Summary of CO ₂ differential methylation analyses	122
Table 6.4: Enriched GO terms among CO ₂ intragenic-DMGs related to heart/muscle development and hypoxia	123
Table A.1: Summary of identified cell lowly methylated regions and overlapping genes.....	139
Table A.2: Number of expression-enriched genes for each immune cell type.....	140
Table B.1: Fetal <i>longissimus dorsi</i> whole-genome bisulfite sequencing summary statistics	142
Table B.2: Fetal <i>longissimus dorsi</i> RNA-seq and smRNA-seq summary statistics	143
Table C.1: Fetal tissue whole-genome bisulfite sequencing library summary statistics	144

Table C.2: Summary of allele-specific sorting of whole-genome bisulfite sequencing reads.....	146
Table D.1. PBMC whole-genome bisulfite sequencing summary statistics	147
Table D.2. PBMC RNA-sequencing read summary	147
Table E.1. Chick heart reduced representation bisulfite sequencing summary statistics	148

LIST OF FIGURES

Figure 1.1: DNA methylation in mammals and its associations with gene expression.....	6
Figure 1.2: Overview of project goals and rationale.....	12
Figure 2.1: Evidence for lineage-specific immune cell differential methylation	25
Figure 2.2: Porcine immune cell differential methylation is associated with differential gene expression	29
Figure 2.3 Cell lowly methylated regions (cLMRs) are enriched within expression-enriched genes	30
Figure 2.4 Porcine immune cell lowly methylated regions (cLMRs) co-localized with unique transcription factor binding motifs.....	31
Figure 2.5 Immune cell differentially methylated regions (cDMRs) harbor immune capacity GWAS SNPs.....	34
Figure 3.1: Fetal skeletal muscle development is associated with genome-wide and site-specific differential DNA methylation	49
Figure 3.2: Fetal skeletal muscle differential methylation exhibits feature-specific enrichment.....	51
Figure 3.3: Fetal skeletal muscle differential methylation is associated with differential transcript and miRNA abundance	53
Figure 3.4: DNA hypomethylation is enriched for myogenic regulatory factor (MRF) binding motifs	56
Figure 3.5: Differentially methylated regions are enriched for muscle and meat quality trait GWAS SNPs.....	58
Figure 4.1: Landscape of porcine fetal DNA methylation.....	69
Figure 4.2: Tissue-specific lowly methylated regions (LMRs) in the developing pig fetus.....	71
Figure 4.3: Developmental differential DNA methylation in porcine fetal tissues	74
Figure 4.4: Pig fetal tissue differentially methylated regions (DMRs) are associated with unique transcription factor (TF) binding motifs	77
Figure 4.5: Breed allele-biased methylation (ABM) is widespread in pig fetal tissues	78
Figure 4.6: Evidence for imprinting-like allele-biased methylation in pig fetal tissues	82

Figure 5.1: Total post-weaning lesion score plotted against percent change in cortisol concentration (post- vs. pre-weaning) in nine piglets	96
Figure 5.2: Differential methylation in post- versus pre-weaning PBMCs	97
Figure 5.3: Glucocorticoid receptor (<i>NR3C1</i>) gene methylation and expression in response to weaning	103
Figure 6.1: Increased egg shell temperature (EST) is associated with significantly lower heart weight.....	117
Figure 6.2: Epigenome-wide association study for heart weight.....	124
Figure 6.3: Relative fold changes in transcript abundance (38.9°C vs. 37.8°C) for select differentially methylated genes	126
Figure A.1: Associations between immune cell global methylation and DNA methyltransferase (<i>DNMT</i>) expression.....	139
Figure A.2: Methylation rates across immune cell marker genes.....	140
Figure A.3: Correlation dot plots of transcript abundance versus promoter expression of immune cell marker genes	141
Figure A.4: Transcript abundance of five transcription factors with cLMR-enriched binding motifs	141
Figure B.1: Principal component analysis plot fetal <i>longissimus dorsi</i> muscle samples by global CpG methylation rates	142
Figure B.2: Differential methylation of miRNAs in developing fetal skeletal muscle	143
Figure C.1: Correlation between first principal component eigenvalues and global methylation rates in pig fetal tissues.	144
Figure C.2: Normalized heatmap of DMR enrichment in gene and CpG features.....	145
Figure C.3: Principal component analysis (PCA) plots of allele methylation rates by tissue	145
Figure E.1: Egg shell temperature (EST) differential methylation in genes influencing heart development.....	148

CHAPTER 1

INTRODUCTION

1.1 Increasing power of swine genomics research through functional annotation

Over the past several decades, advancements in genetics research in the domestic pig (*Sus scrofa*) have led to improvements in selection for agriculturally relevant traits as well as its use as a biomedical model. The development and application of porcine SNP genotyping arrays in genome-wide association studies (GWAS) has led to the identification of variants and quantitative trait loci (QTL) associated with hundreds of growth, reproductive, and disease traits [1, 2]; to date 33,143 such associations have been submitted to the Pig QTL database (<https://www.animalgenome.org/cgi-bin/QTLdb/SS/index>). Furthermore, high-quality genome assemblies are publicly available to advance swine research: a reference genome assembly created from a combination of long-read and short-read sequencing strategies greatly improved *S. scrofa* gene annotation over the previous assembly, and short-read assemblies for 12 additional pig breeds have been generated to better represent the diversity of the porcine genome sequence [3]. In addition to common anatomical and physiological characteristics, the observed similarities in genome sequence and gene content between pig and human have resulted in increased use of the pig as a biomedical model in cardiovascular, brain, and gut physiology research, among other areas [4]. For example, comparative genomics studies have revealed greater immune gene conservation between pigs and humans than between mice and humans, strengthening the argument for continued use of the pig as an infectious disease model [5].

Despite increased knowledge of genetic variation and sequence in the pig, a limited understanding of genome function hinders continued progress in genetic selection and translational

research. A significant proportion of identified pig QTL are megabases in length due to high linkage disequilibrium in tested populations, and many of these lie within intergenic regions of the genome for which no functional role has been assigned [6]. It is being increasingly appreciated that such non-coding regions of the genome harbor *cis*-regulatory elements that play important roles in influencing gene expression. These include but are not limited to: enhancers that act distally to increase rates of transcription through chromatin looping [7]; insulator elements that delineate chromatin into topologically-associated domains [8]; and silencers that inhibit gene expression via the recruitment of transcriptional repressors [9]. In addition to their effects on transcript diversity, intergenic regions of the genome have been shown to disproportionately harbor variants that influence complex trait variation [10]. Recognizing the wealth of knowledge to be gained by identifying such regulatory regions, the Encyclopedia of DNA Elements (ENCODE) project was initiated in 2011 with the goal of increasing functional annotation in the human genome. To date, ENCODE has assigned a functional role to 80% of the human genome, and has identified hundreds of thousands of putative enhancer, silencer, and insulator elements [11]. Similar large-scale annotation projects have been undertaken in mice (the mouseENCODE project) and other model organisms such as *Drosophila* and *C. elegans* (modENCODE) and have produced similar findings [12–14]. ENCODE project data has been subsequently utilized to refine previously identified GWAS regions as well as predict causal variants for disease traits and their target genes [15, 16]. Pigs and other domesticated animals still lag far behind humans and model organisms with respect to their own genome annotation, despite the potential use of such data in improving agricultural output and biomedical research.

In light of this gap in knowledge, the Functional Annotation of Animal Genomes (FAANG) consortium was founded in 2015 to enhance the discovery of functional elements in domesticated

animal genomes. The findings of this consortium will not only aid in the identification of mechanistic links between genotypic and phenotypic variation, but increase the efficacy of animal species as models for human disease [17]. To achieve such ends, the FAANG consortium seeks to perform next-generation sequencing assays to identify functional genetic loci across tissues and cell types, developmental stages, and environmental conditions. Many proposed assays will characterize transcribed portions of the genome, including transcription start sites and novel transcript isoforms. However, a large proportion of FAANG efforts is dedicated to describing proximal and distal *cis*-regulatory regions that have been shown in other species to influence gene and phenotype expression. To date, FAANG projects have begun to elucidate conserved and species-specific regulatory elements through the assessment of chromatin states, including within a small subset of tissues in the pig [18, 19]. Of note, these studies report that, despite relative lack of genomic conservation across mammalian species within intergenic enhancers, these elements were nevertheless functionally conserved to regulate similar genes and signaling pathways.

1.2 DNA methylation and its role in gene regulation and expression

While regulatory elements can bear characteristic DNA sequences that may allow for computational prediction of their locations, knowledge of their activation status and tissue-specificity cannot be discerned from genome sequence data alone [20]. It is now recognized that such state-specific regulatory information can be obtained through assessment of animal epigenomes. Broadly defined as the repertoire of chemical modifications made to chromatin that do not involve changes in DNA sequence, the epigenome is primarily established by two classes of mechanisms: those that act on the DNA molecule itself (e.g., DNA methylation and hydroxymethylation) and those resulting in histone tail modifications. Previous research has shed light on the prevalence of characteristic epigenomic modifications at various regulatory elements,

leading to their increased use as molecular signposts to predict function. For example, trimethylation of lysine 4 on the histone 3 tail (H3K4me3) is highly prevalent at gene promoters, while monomethylation of the same residue (H3K4me1) co-localizes with enhancer elements [21]. Other modifications are predictive of repressive elements, such as trimethylation of H3 lysine 27 (H3K27me3) that contributes to chromatin compaction [22]. Sequencing strategies to assess genome-wide modifications have been implemented by the FAANG consortium and have been successful in the early prediction of functional elements in animal genomes. Studies have shown that not only do epigenomic modifications associate with tissue-specific gene regulation [19], but that the deposition of histone marks is highly dynamic throughout such processes as the innate immune response in the pig [23], strengthening the argument for continued spatiotemporal assessment of farm animal epigenomes.

Among epigenetic mechanisms, DNA methylation is the most widespread modification made to the DNA molecule and refers broadly to the enzymatic addition of a methyl group to DNA bases. In mammals, DNA methylation occurs almost exclusively at cytosine bases, and is further restricted to cytosine-guanine dinucleotides (CpGs) in most tissues with the exception of the nervous system and embryonic stem cells [24, 25]. Mammalian genomes generally exhibit CpG methylation rates between 60% and 80%, with local methylation being strongly inversely correlated with CpG density [26]. While the majority of mammalian genomes are CpG-sparse and exhibit high methylation rates, regions of high CpG density, referred to as CpG islands, exhibit low methylation rates on average, due in part to local histone modifications that inhibit DNA methylation [27]. Cellular DNA methylation levels are carefully regulated by the activity of DNA methylation and demethylation enzymes. Cytosine methylation is performed by the DNA methyltransferase family of enzymes, of which three members—DNMT1, DNMT3A, and

DNMT3B—actively deposit methyl groups (Figure 1.1A). DNMT1 acts as a maintenance methyltransferase, and functions primarily at newly replicated DNA to preserve methylation patterns at hemi-methylated bases [28]. DNMT3A and DNMT3B are both responsible for *de novo* DNA methylation, with *DNMT3B* expression more restricted to early embryonic development [29]. Loss of methylation can occur either passively following successive rounds of DNA replication, or actively via TET methylcytosine deoxygenase (TET) activity that initiates the DNA demethylation pathway [30].

Studies over the past several decades have revealed context-specific relationships between DNA methylation and local and distal gene expression. At activating regulatory elements, such as gene promoters and intergenic and intronic enhancers, hypermethylation has generally been associated with suppression of corresponding gene transcription [25, 31, 32], and several mechanisms governing this inverse correlation have been proposed (Figure 1.1B). In many cases, DNA methylation has been shown to occlude transcription factor binding through the alteration of *cis*-acting elements [33]. Conversely, methyl-binding proteins recognize methylated DNA sequences and act as transcriptional repressors by recruiting other chromatin remodeling enzymes that promote a heterochromatic state and reduce RNA polymerase II elongation efficiency [34, 35]. While once thought to exert solely repressive effects, it is now acknowledged that, in several contexts, DNA methylation levels are positively correlated with rates of gene transcription (Figure 1.1C). DNA hypermethylation is often observed in highly expressed genes [36]; while this has been attributed to the prevention of spurious transcription from alternative transcription start sites [37], much of the DNA methylation in this context has been hypothesized

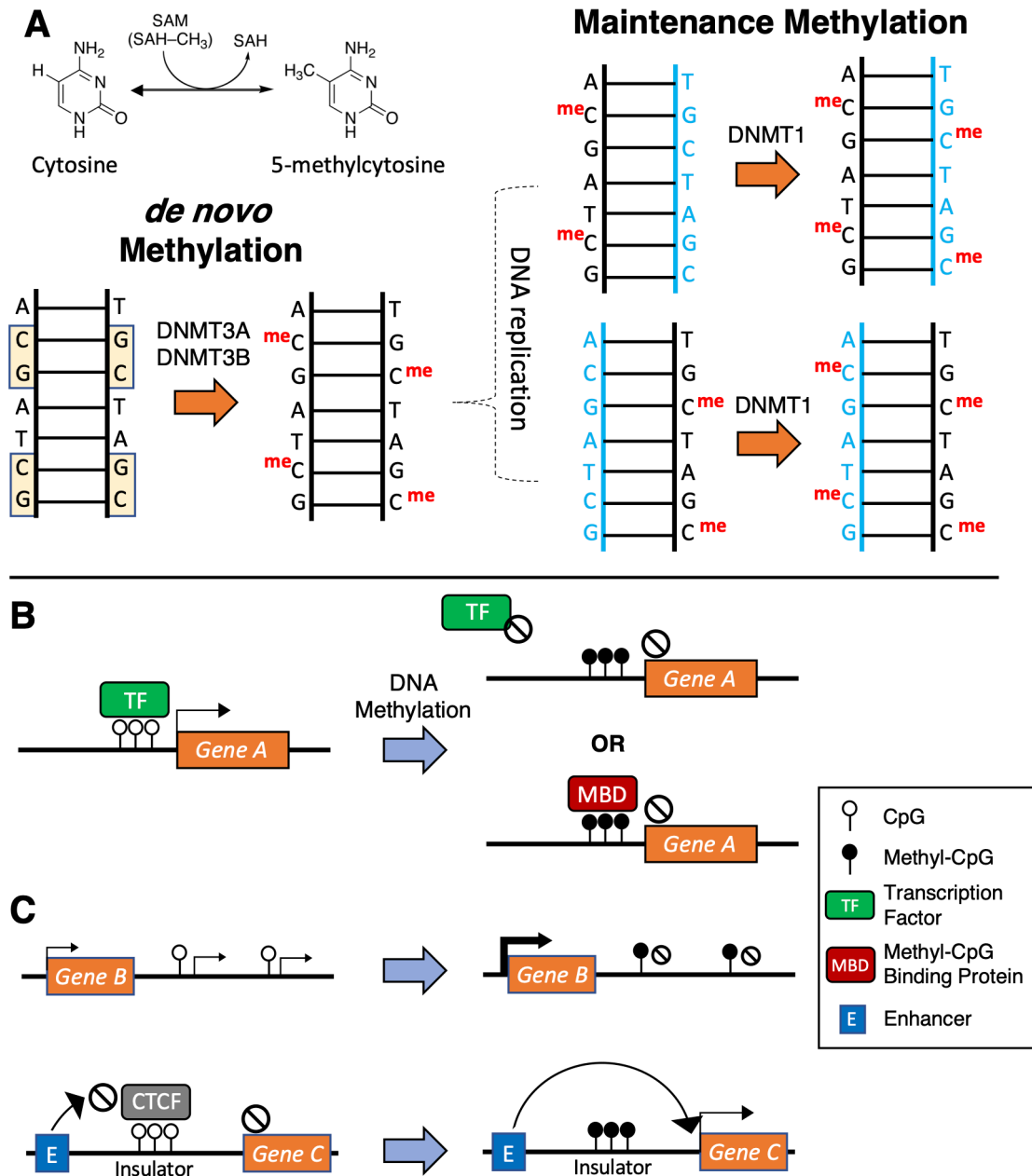


Figure 1.2. DNA methylation in mammals and its associations with gene expression.

(A) Methylation of cytosine-guanine (CpG) dinucleotides can occur *de novo* via DNA methyltransferase 3A (DNMT3A) and DNMT3B, or at hemi-methylated DNA following DNA replication via DNMT1-directed maintenance methylation. (B) Methylation of activating regulatory elements (e.g., promoters, enhancers) is associated with gene repression via the exclusion of transcription factor (TF) binding or the recruitment of methyl-CpG binding proteins (MBDs) and other transcriptional repressors. (C) Methylation in some contexts can have positive associations with gene expression. In the case of *Gene B*, intragenic methylation inhibits spurious transcription from alternative start sites. In the case of *Gene C*, methylation of an insulator element prevents CTCF binding and the creation of a chromatin boundary, allowing for an upstream enhancer to activate transcription.

to be passively deposited as a result of increased accessibility of DNA by RNA polymerase activity [38]. Lastly, suppressive effects of DNA methylation may indirectly activate expression of another gene. For example, hypermethylation of an insulator element in the *H19/IGF2* gene cluster inhibits a nearby enhancer from acting on *H19*, but promotes its acting on *IGF2*, while the opposite pattern is observed under insulator hypomethylation [39]. In the *IGF2R* locus, hypermethylation of an intragenic element promotes *IGF2R* expression by inhibiting transcription of an *IGF2R* antisense RNA [40]. In conclusion, while feature-specific correlations between DNA methylation and gene expression have been identified across mammalian species, the mechanisms governing this relationship appear to be highly gene-specific and are thus worthy of further study.

1.3 DNA methylation patterns govern tissue- and stage-specific gene expression

Since the advent of bisulfite-sequencing strategies to assess genome-wide DNA methylation, functional annotation projects in humans and other model organisms have provided unique insights into methylation variation across diverse tissues and cell types. Initial DNA methylation data generated from the ENCODE project reported that over 90% of CpGs exhibited variable methylation levels across assayed samples, and that such variation was strongly correlated with local gene expression as well as enhancer activity [25]. Furthermore, assessment of methylation states across neighboring CpGs has led to the discovery of human tissue- and cell-specific differentially methylated regions (DMRs) that represent putative regulatory elements governing state-specific gene expression [32, 41]. Tissue-specific DMRs were not only found to be enriched for binding motifs of relevant transcription factors, but disproportionately harbored GWAS SNPs associated with tissue-related traits, highlighting the potential role of these regions in regulating phenotype expression. Initial surveys of DNA methylation patterns in porcine tissues have revealed similar genome-wide methylation patterns and context-specific associations with gene

expression as has been reported in other mammals [42]. However, knowledge of such patterns across a wide set of porcine states is still limited, particularly in cell types relevant to industry performance and biomedical research use. For example, little is known about the contribution of DNA methylation to regulatory variation in porcine immune cells, despite its potential usefulness in selection for disease resistance phenotypes and in translational research applications.

In contrast to the often-stable patterns observed in differentiated animal tissues, DNA methylation is highly variable throughout early mammalian development. While the early embryo is subject to global demethylation followed by remethylation to erase most parentally-inherited epigenomic marks, DNA methylation patterns continue to be highly dynamic throughout cell lineage commitment, differentiation, and maturation in coordination with stage-specific gene expression demands [43]. Studies of human fetal development have shown that tissue-specific DNA methylation patterns are observed by the first trimester, with subsequent differential methylation being associated with suppression of early development genes and activation of tissue-specific genes [44]. Despite the fact that many economically important phenotypes in the pig are influenced by prenatal physiology and gene expression [45, 46], assessment of epigenomic variation and discovery in pig fetal tissues has been limited. Additionally, epigenomic variation likely contributes to observed developmental differences in anatomy, physiology, and gene expression between divergent pig breeds, although the exact mechanistic links between genetic background and developmental variation have been largely unexplored.

In addition to exhibiting temporal variation, DNA methylation plays important roles in conferring parent-of-origin-specific regulatory effects. The phenomenon of genomic imprinting results in biased or exclusive gene expression from the paternal or maternal allele, and primarily affects genes involved in fetal growth and development [47–49]. As imprinting has been shown to

occur even in diploid organisms with genetically identical alleles, epigenetic mechanisms were hypothesized as primary regulators of this process. Indeed, it is now recognized that imprinted alleles possess imprinting control regions (ICRs) at which differential DNA modifications—in particular DNA methylation—result in biased or monoallelic gene expression [49]. While over 100 genes in humans and mice have been identified as imprinted, and many corresponding ICRs have been identified, the number of experimentally-verified imprinted genes and ICRs in the pig and other livestock species is comparatively few [50]. However, the identification of genetic variation in imprinted alleles in livestock species has led to the recognition that such QTL are inheritance-specific. In the pig, two alleles have been identified in the paternally-expressed *IGF2* gene that differ in a single G-to-A point mutation; the ‘A’ allele results not only in increased *IGF2* expression, but increased muscle mass and lower body fat, although this effect is only observed on the *IGF2*-expressing paternal allele [51, 52]. Such knowledge of parent-of-origin-specific genetic effects across imprinted loci in farm animal genomes has the potential to improve quantitative genetics models and genomic selection strategies.

1.4 DNA methylation as a mediator of environment-induced physiological changes

It has long been appreciated that early life stress can result in epigenomic reprogramming that has physiological and behavioral outcomes in animal species [53]. Due to the dynamic nature of DNA methylation during pre- and perinatal development, the methylome is hypothesized to be more susceptible to environmental insults during this period. Thus, research in humans and rodents has been focused on the impact of early life stimuli on epigenetic regulation in stress response systems, particularly the hypothalamic-pituitary-adrenal (HPA) axis. Studies in rats have demonstrated that levels of perinatal maternal care are significantly associated with neuronal DNA methylation patterns; of note, decreased maternal grooming has been associated with increased gene

methylation and decreased expression of the glucocorticoid receptor gene (*NR3C1*), a critical negative feedback regulator of the HPA axis response [54, 55]. Studies assessing the effects of stress on DNA methylation have also sought to characterize patterns in peripheral blood, as this can be more easily sampled and has been shown to serve as a suitable ‘surrogate’ for methylation and expression patterns in other tissues [56]. Human studies have reported similar effects of early life stress on *NR3C1* methylation in peripheral blood [57, 58], and, in cattle, transportation stress has been associated with HPA axis gene methylation changes, including within genes involved in corticotropin releasing hormone signaling [59, 60]. Recent studies in the pig have begun to elucidate DNA methylation changes in the brain associated with early life nutritional stress and viral infection [61]. However, pigs experience numerous stressors in production systems that likely elicit specific molecular and physiological responses, necessitating additional studies of stress-induced epigenomic alterations and their impact on gene expression.

Research over the past several decades have also proposed epigenetic mechanisms as mediators of abiotic stress-induced physiological changes in animals [62, 63]. Temperature is among the most consequential abiotic factor influencing animal growth, metabolism, and behavior; as such, numerous studies in livestock and poultry species have examined epigenomic patterns associated with variable environmental temperature. Studies of heat stress in livestock species have reported significant DNA methylation differences in skeletal muscle and whole blood, impacting genes involved in lipid metabolism and immune and stress response, respectively [64, 65]. Poultry species are known to be particularly sensitive to changes in incubation temperature, with profound effects on early and later-life metabolism and thermoregulation [66–68]. Thermal manipulation studies during broiler chick embryonic development have reported alterations in histone modifications patterns in the hypothalamus associated with brain development and metabolism

[69]. Furthermore, adult differential DNA methylation patterns in heat shock protein genes have been observed in brain tissue of broiler chickens exposed to early life heat stress versus control birds, providing evidence for epigenetic ‘memory’ of environmental insults [70]. The extent to which such changes are present in the most temperature-responsive organ systems, such as the cardiovascular system, remain unexplored in poultry species.

1.5 Project Goals & Significance

In summary, there is a need for increased characterization of DNA methylation patterns across farm animal cell types, developmental stages, and environmental conditions. Knowledge of such state-specific epigenomic modifications will aid in the identification of regulatory elements associated with gene expression and provide additional layers of regulatory information in poorly-annotated regions of animal genomes. This dissertation presents findings from whole-genome DNA methylation analyses of porcine and avian tissues and cell types of economic relevance. First, I report that sorted porcine immune cells exhibit unique DNA methylation patterns that significantly co-localize with cell-enriched gene expression, binding motifs for transcription factors regulating immune cell function, and GWAS SNPs for immune capacity traits (Figure 1.2A, Chapter 2). Second, I demonstrate that fetal tissue differentiation is associated with genome-wide DNA methylation changes that aid in establishment of unique transcriptomic profiles, and that allele-specific methylation is widespread in pig fetal tissues in both a genotype-dependent and independent manner (Figure 1.2B, Chapters 3 & 4). Lastly, I provide evidence that farm animal methylomes are responsive to stresses experienced in production settings, including weaning stress in piglets (Chapter 5) and incubation temperature and CO₂ concentration in broiler chicks (Figure 1.2C, Chapter 6). The results presented in this dissertation will not only serve as a valuable resource in genome annotation-driven searches for causative variants influencing economically

important traits, but have also revealed novel genomic loci subject to allele-specific and environment-dependent epigenetic gene regulation.

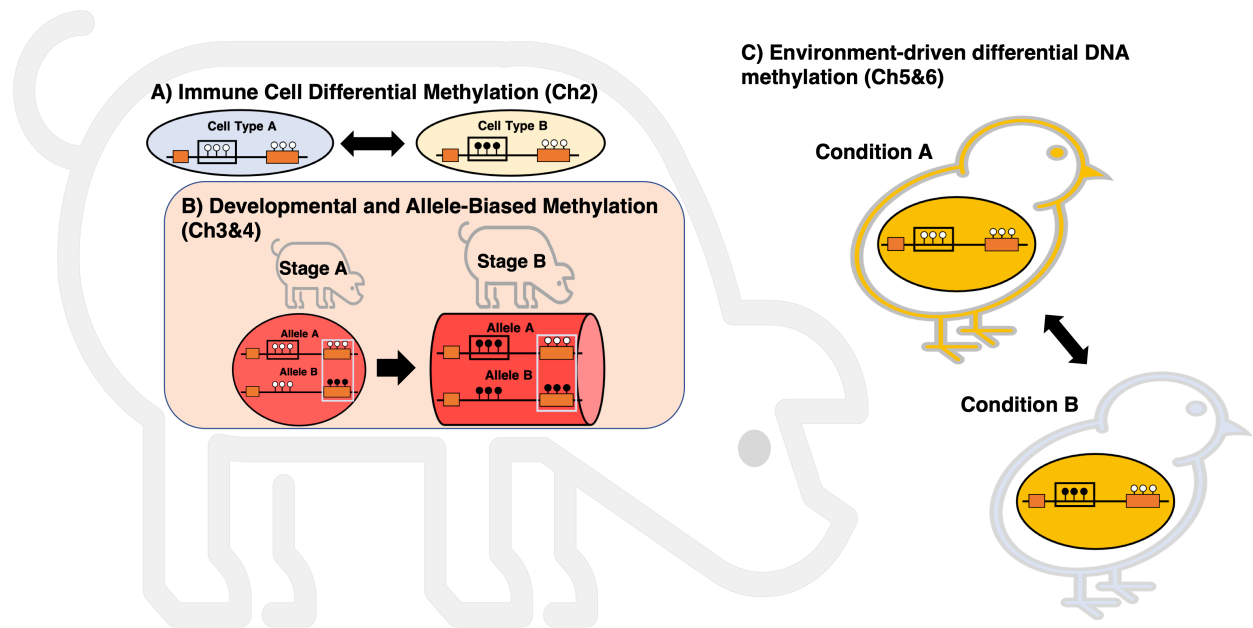


Figure 1.2. Overview of project goals and rationale. (A) Assessment of DNA methylation patterns in adult porcine immune cells will reveal cell-type specific differentially methylated regions (DMRs, black boxes) associated with immune cell gene regulation and improve functional annotation of the porcine immunome. (B) Identification of stage-specific DMRs in differentiating pig fetal tissues (e.g., muscle, shown) will increase understanding of epigenetic gene regulation during development of economically-relevant organ systems. Furthermore, analysis of allele-biased methylation (white boxes) will uncover novel genetic loci under breed-specific and imprinting regulation. (C) Differential DNA methylation patterns between animals in different environmental conditions is indicative of putative long-term changes in gene regulation as a result of external stimuli, and will reveal candidate genes driving environment-induced physiological changes. White circles = unmethylated CpGs, black circles = methylated CpGs.

CHAPTER 2

ASSESSMENT OF DNA METHYLATION IN PORCINE IMMUNE CELLS REVEALS NOVEL REGULATORY ELEMENTS ASSOCIATED WITH CELL-SPECIFIC GENE EXPRESSION AND IMMUNE TRAITS

This chapter is currently being prepared for publication. It was prepared alongside co-authors Andrea M. Luttman, Juber Herrera-Urbe, Haibo Liu, Nancy E. Raney, Jenna M. Grabowski, Crystal L. Loving, Christopher K. Tuggle, and Catherine W. Ernst.

2.1 Abstract

The porcine immune system possesses a vast repertoire of broad-mammalian and species-enriched cell types that are critical in combatting infection. Genetics studies have enhanced pig selection practices for disease resistance phenotypes as well as increased the efficacy of the porcine model in biomedical research; however limited functional annotation of the porcine immunome has hindered progress on both fronts. Among various epigenetic mechanisms that regulate mammalian gene expression, DNA methylation is the most ubiquitous modification made to the DNA molecule and has been shown to influence transcription factor binding as well as gene and phenotype expression. Human and mouse DNA methylation studies have improved mapping of regulatory elements in these species, but comparable studies in the pig have been limited in scope. We performed whole-genome bisulfite sequencing in nine porcine immune cell populations to assess cell-specific DNA methylation patterns and their associations with: 1) cell-enriched functions and gene expression, 2) transcription factor binding motifs, and 3) GWAS SNPs for immune capacity and disease traits. Whole blood was collected from two crossbred barrows and subjected to density gradient centrifugation to remove red blood cells and separate out neutrophils, and peripheral blood mononuclear cells underwent magnetic- and fluorescence-activated cell sorting into myeloid

cells, natural killer (NK) cells, two B cell fractions (CD21⁺ and CD21⁻) and four T cell fractions (CD4⁺, CD8⁺, CD4+CD8⁺, and SWC6γδ⁺). We identified 54,391 immune cell differentially methylated regions (cDMRs), and clustering by cDMR methylation rate grouped samples by cell lineage. 32,737 cDMRs were classified as cell lowly methylated regions (cLMRs) in at least one cell type (methylation<75% and z-score<-1), and cLMRs were broadly enriched in genic regions as well as regions of intermediate CpG density. Immune cDMRs exhibited methylation rates that were significantly correlated with local transcript abundance across cell types, with the majority of these correlations being negative. Furthermore, cell lowly methylated genes were overrepresented among expression-enriched genes for the same cell type, suggesting that low methylation is strongly associated with cell-specific gene activation. Motif analysis of cLMR sequences revealed cell type-specific enrichment of transcription factor binding motifs among B, T, myeloid, and NK cells, indicating that cell-specific methylation patterns may influence accessibility by trans-acting factors. Lastly, cDMRs were specifically enriched for immune capacity GWAS SNPs; many such overlaps occurred within genes known to influence immune cell development and function and have previously been associated with immune phenotypes, including *CD8B* and *NDRG1*. Overall, our DNA methylation data improve functional annotation of the porcine genome through characterization of epigenomic regulatory patterns that contribute to immune cell identity and function, and increase the potential for identifying mechanistic links between genotype and immune and disease trait variation.

2.2 Introduction

The porcine immune system plays critical roles in combatting infectious diseases, including those prevalent in production systems [71]. As in other mammals, pig immunity is conferred by two defense systems – innate and adaptive immunity. Innate immunity involves many barrier systems

and immune cells including myeloid-derived macrophages, dendritic cells and granulocytes as well as lymphoid-derived natural killer (NK) cells. Adaptive immunity refers to acquired immunity and is conferred by a system of lymphoid-derived B and T cells. In addition to adaptive immune cells phenotyped in humans and mice, pigs exhibit an overrepresentation of specific T cell subsets in peripheral blood. These include CD4⁺CD8⁺ double positive (DP) T cells, which express CD8αα as opposed to CD8αβ on conventional cytotoxic T cells and are functionally characterized as a memory T cell population with MHC-II restriction [72, 73]. In addition, pigs are considered a gamma-delta (γδ) high species with frequencies of circulating γδ T cells of up to 30% [74]. γδ T cells are defined by the expression of T cell receptors (TCR) composed of gamma and delta subunits, as opposed to the alpha and beta chain TCR of more conventional αβ T cells. γδ T cells play roles in both the innate and adaptive immune response [75] and differential expression of CD2 and CD8α on γδ T cells is indicative of function [76]. Increased functional characterization of pig leukocytes—and in particular pig-enriched T cell populations—will inform understanding of their unique roles in immune response pathways.

Genetics studies in the pig have enhanced the discovery of DNA variants associated with immune-related traits: to date 3,231 and 610 genetic associations with immune capacity and disease susceptibility traits, respectively, have been submitted to the Pig QTL Database (<https://www.animalgenome.org/cgi-bin/QTLdb/SS/index>). Furthermore, comparative genomics studies revealed greater preservation of human immune genes and gene sequences in pigs relative to mice [5], making the pig a promising biomedical model to study genetic contributions to human diseases [77, 78]. However limited functional annotation of the porcine genome—particularly within regulatory regions—has limited both the search for causative variants influencing disease and production traits in pigs as well as the translational capabilities of the pig as a model organism

[79]. To this end, the Functional Annotation of Animal Genomes (FAANG) consortium was initiated to map functional elements in domesticated animal species through the use of high-throughput sequencing data generated from tissues and cell types of relevance, including those of the immune system [17, 18]. Among the core FAANG assays are many that assess epigenomic modifications that regulate chromatin accessibility and gene expression, including DNA methylation, histone modifications, and long non-coding RNAs [63]. FAANG projects in the pig have begun characterizing epigenomic patterns in immune cell populations, including in LPS- and Poly-I:C-stimulated cells [23, 80, 81].

DNA methylation involves the enzymatic addition of a methyl group to the 5-position of cytosine rings, producing 5-methylcytosine. Methylation occurs almost exclusively at CpG dinucleotides in mammals and has context-specific associations with gene expression. At gene promoters and enhancers, methylation generally functions to decrease levels of transcription through the alteration of transcription factor (TF) binding sites or the recruitment of transcriptional repressors and chromatin-modifying enzymes [25, 31, 82]. Generation of DNA methylomes from diverse tissues and cell types across animal species has led to the identification of differentially methylated regions (DMRs) that have vastly improved understanding of tissue- and cell-specific epigenetic gene regulation. Many such studies have identified strong associations between cell-specific lowly methylated regions (LMRs) and TF binding motifs as well as GWAS SNPs for relevant traits, highlighting the potential significance of these regions and of DNA methylation in regulating gene and phenotype expression [32, 41, 83–85]. DNA methylation patterns have previously been characterized in heterogeneous porcine immune cell populations [80, 86]. However, assessment of methylation patterns in specific cell types has not been performed, but could provide insight on transcriptional regulation and hence function relevant to pig health.

Here we report the first genome-wide DNA methylation study in porcine granulocytes (primarily neutrophils) and eight sorted immune cell populations: myeloid cells, NK cells, two B cell fractions (CD21⁺ and CD21⁻) and four T cell fractions (CD8⁺, CD4⁺, CD8⁺CD4⁺, and SWC6 γ δ ⁺). Using whole-genome bisulfite sequencing (WGBS), we identified cell-differential DNA methylation patterns strongly associated with enriched gene expression and TF binding sites governing cell specificity. Furthermore, we report DMRs overlapping previously-identified GWAS SNPs for immune-related traits, suggesting they may play important roles in regulating gene expression that impacts phenotypic variation. Our data massively improve functional annotation of the porcine immunome and provide unique insight into epigenetic gene regulation in understudied immune cell populations.

2.3 Materials & Methods

2.3.1 Blood Collection and Separation

Blood was collected from two crossbred (predominantly Large White and Landrace heritage) male pigs between 4 and 6 months of age. Pigs were housed in BSL2 rooms at the National Animal Disease Center (Ames, IA) and all animal procedures were performed in compliance with and with approval by the Institutional Animal Care and Use Committee. Blood was drawn and peripheral blood mononuclear cells (PBMCs) were isolated as described in Herrera-Urbe et al. 2021 [81]. For granulocyte isolation, blood was collected in BD Vacutainer ACD solution A tubes and subjected to dextran sedimentation using 6% Dextran / 0.9% NaCl solution at room temperature for 45-60 minutes. The supernatant was transferred to a conical tube and centrifuged for 12 minutes at 300 RCF and 4° C, after which pelleted erythrocytes were lysed with ACK Lysing buffer per manufacturer's instructions (Thermo Fisher). The pellet was resuspended and overlaid onto

Ficoll-Histopaque-1077 and centrifuged for 30 min at 450 RCF at room temperature. The mononuclear cell layer was discarded and the pellet containing granulocytes (primarily neutrophils) was resuspended in phosphate buffered saline (PBS) and centrifuged at 450 RCF for 5 minutes. The resulting pellet was resuspended in 2 mL HBSS, and viable cells were enumerated using the Count and Viability Assay Kit on the MUSE® detection system (Merck Millipore).

2.3.2 Immune Cell Sorting from PBMCs

PBMCs underwent magnetic- and fluorescence-activated cell sorting (MACS/FACS) as previously described [81]. Briefly, PBMCs were incubated with biotin labeled anti-porcine CD3 ϵ (PPT3, Washington State University Monoclonal Antibody Center) and separated into CD3 ϵ - positive and CD3 ϵ -negative fractions on LS columns. Both fractions were further separated by FACS into four subpopulations based on extracellular markers. Isolated subpopulations in the CD3 ϵ -positive fraction included SWC6 $\gamma\delta$ ⁺ T cells (**SWC6gdT**; CD3 ϵ ⁺SWC6⁺), CD4⁺ T cells (**CD4T**; CD3 ϵ ⁺SWC6⁻CD8 α ⁻CD4⁺), CD8⁺ T cells (**CD8T**; CD3 ϵ ⁺SWC6⁻CD8 α ⁺CD4⁻), and CD8⁺CD4⁺ T cells (**CD4CD8T**; CD3 ϵ ⁺SWC6⁻CD8 α ⁺CD4⁺). Isolated subpopulations from the CD ϵ -negative fraction included myeloid leukocytes (**myeloid**; CD3 ϵ ⁻CD172 α ⁺CD8 α ⁻), NK cells (**NK**; CD3 ϵ ⁻CD172 α ⁻CD8 α ⁺), CD21⁺ B cells (**CD21pB**; CD3 ϵ ⁻CD172 α ⁻CD8 α ⁻CD21⁺), and CD21⁻ B cells (**CD21nB**; CD3 ϵ ⁻CD172 α ⁻CD8 α ⁻CD21⁻). A fraction of each sorted population and isolated granulocytes were flash-frozen and stored at -20°C.

2.3.3 DNA Isolation and Bisulfite Sequencing

DNA was isolated from cells using the Qiagen AllPrep DNA/RNA Minikit and quantified using a Qubit fluorometer. Prior to library preparation, sample DNA was spiked with unmethylated lambda phage DNA (Promega) at a concentration of 5 ng lambda DNA/1 μ g sample DNA. DNA

was fragmented to approximately 350bp using a Covaris M220 Sonicator, and bisulfite-converted using the Zymo EZ DNA Methylation-Gold Kit according to manufacturer's instructions (Zymo Research). Bisulfite sequencing libraries were prepared using the Accel-NGS Methyl-Seq DNA Library Kit and Methyl-Seq Combinatorial Dual Indexing Kit (Swift Biosciences). Completed libraries were quantified and QC'ed using Qubit dsDNA HS and Agilent 4200 TapeStation HS DNA1000 assays, respectively.

Sequencing libraries were divided into three pools, and WGBS was performed on each pool across three flow cell lanes on an Illumina HiSeq 4000 instrument in 2x150PE format using HiSeq 4000 reagents. A PhiX control DNA library was spiked into each lane at 10% of the total to account for the unbalanced base composition inherent in Methyl-Seq libraries. Base calling was done by Illumina Real Time Analysis (RTA) v2.7.7 and output of RTA was demultiplexed and converted to FastQ format with Illumina Bcl2fastq v2.19.1.

2.3.4 WGBS Bioinformatics Analyses

WGBS libraries were trimmed of technical sequences and low-quality bases using Trimmomatic v.0.39 [87]. Forward and reverse reads were subjected to removal of the first 10 and 15 bases, respectively, according to the Swift Biosciences Library Kit instructions. Reads were further trimmed using the following parameters: ILLUMINACLIP:<adapter sequence>:2:30:10 LEADING:25 TRAILING:25 AVGQUAL:20 MINLEN:30. Trimmed reads were aligned to the *Sus scrofa* reference genome (v11.1) using Bismark [88]. The bismark_genome_preparation command was used to prepare an *in-silico* bisulfite-converted bowtie2 index using default parameters. WGBS paired-end alignment was performed using the parameters: -X 1000 --score_min L,0,-0.6. Unmapped forwards reads were merged with forwards reads left unpaired following trimming and aligned using the same parameters. All libraries were also aligned to the

lambda phage genome using default parameters, and the percentage of methylated cytosines among lambda-derived reads was subtracted from 100 to calculate bisulfite conversion efficiency. Aligned WGBS reads were deduplicated using the `bismark_deduplicate` command, and CpG methylation reports were generated using the `bismark_methylation_extractor` command using default parameters.

2.3.5 cDMR and cLMR Identification

Genome regional methylation rates were calculated using the *methyKit* R package [89]. Briefly, CpG reports from all samples were merged, and genome tiling was performed to calculate average methylation rates for 1kb non-overlapping regions in the pig genome. Regions with coverage >25 in at least 7 cell types per animal were retained for further analysis. We generated linear mixed models for each region with average methylation rate as a response and included the fixed effect of cell type and random effect of animal. We assessed the significance of the cell effect using two-way analysis of variance (ANOVA), and regions with a multiple test-corrected cell effect false discovery rate (FDR) <0.01 were defined as cell differentially methylated regions (cDMRs). Standardized scores (z-scores) were calculated from methylation rates at each cDMR, and those regions with z-score <-1 and mean methylation rate <75% in a given cell type were further classified as cell lowly methylated regions (cLMRs) for that cell type. cLMRs were annotated using the *genomation* R package [90], and cLMR genes were submitted to the Panther database for gene set enrichment analysis [91, 92].

2.3.6 Integration of Gene Methylation and Expression Data

RNA-sequencing data from the same sorted immune cell populations was previously reported and was used in the current analysis [81]. The transcript data from bulk sorted PBMCs was used in the current analysis, and this is the first report on the neutrophil RNA-seq data. Briefly, transcript

abundance was quantified as transcripts per kilobase million (TPM) for all samples, and genes exhibiting cell-enriched expression were identified using the *DESeq2* R package [93] and defined as genes with log2-fold change >1 and FDR <0.05 in a cell type relative to all other cell types. We calculated Pearson correlation coefficients between cDMR methylation rates and transcript abundance of overlapping Ensembl-annotated genes, further separating cDMRs into those overlapping promoters ($<2\text{kb}$ from transcription start site), gene bodies (intragenic), and transcription termination sites (TTSs). We compared correlation distributions to those derived from a random sampling of 1kb regions for each feature and corresponding transcript abundances. To determine the degree of association between low gene methylation and enriched expression, we calculated enrichment p-values between lowly methylated genes and expression-enriched genes for each cell type using hypergeometric tests performed using R software.

2.3.7 Transcription Factor Binding Motif Enrichment Analyses

We extracted cLMR sequences as well as sequences from an equal number of random regions for each cell type. cLMR sequences were submitted for Analysis of Motif Enrichment by the MEME Suite [94, 95], using random sequences as controls. Motifs from the JASPAR CORE vertebrates NON-REDUNDANT (*in-vivo* and *in-silico*), UniProbe Mouse, and Jolma2013 Human and Mouse databases [96–98] were scored using the average odds score method, and motif enrichment was calculated using Fisher’s exact test. Motif enrichment clustering was performed using the *pheatmap* R package v.1.0.12 [99].

2.3.8 GWAS SNP Enrichment Analysis

GWAS SNPs associated with immune-related and nonrelated traits were downloaded from the Pig QTL Database [100]. We performed hypergeometric tests using R software to calculate the enrichment of peak SNPs for select trait classes within cDMRs.

2.3.9 RT-qPCR Analyses

We quantified transcript abundances of genes overlapping cLMRs and previously-reported GWAS SNPs using quantitative reverse-transcriptase polymerase chain reaction (RT-qPCR). Buffy coat samples collected from an equal number of male and female 7-week-old Yorkshire pigs representing selected SNP genotypes underwent RNA extraction with TRIzol. 2 ug RNA was reverse transcribed using the High Capacity cDNA Reverse Transcription Kit (Applied Biosystems) and cDNA was quantified on a Nanodrop spectrophotometer. *B2M* and *GAPDH* were selected as reference genes due to their reported consistent expression across PBMC isolations [101, 102]. qPCR assays were performed in duplicate on a StepOnePlus Real-time PCR Instrument (Applied Biosystems) using 5 µl cDNA (500 ng total), 1 µl TaqMan Gene Expression Assay (Applied Biosystems Assay Nos. Ss03391669 (*CD8A*), Ss06890240 (*NDRG1*), Ss03374854_g1 (*B2M*), Ss03391154_m1 (*GAPDH*)), 10 µl TaqMan Fast Advanced Mastermix (Applied Biosystems), and 4 µl water. Reaction conditions were 50°C for 2 min and 95°C for 2 min, followed by 40 cycles of 95°C for 1 s and 60°C for 20 s. Delta Cts (dCts) were obtained for each sample by subtracting the geometric mean of reference gene Cts from the test gene Ct, and ANOVA and post-hoc pairwise comparisons were performed to identify significant differences between genotypes.

Table 2.1. Summary of Immune Cell WGBS Libraries and Global DNA Methylation

Cell Type	Symbol	No. Raw Reads (M)	Mapping Rate (%)	Conversion Efficiency (%)	Global Methylation Rate (%)*	Rep Corr.
CD21 ⁻ B cells	CD21nB	152.4	88.5	99.4	82.2 ^{abc}	0.80
CD21 ⁺ B cells	CD21pB	150.6	90.3	99.4	80.3 ^{bc}	0.83
Myeloid cells	Myeloid	146.0	88.8	99.4	83.5 ^a	0.81
Neutrophils	Neut	162.6	90.5	99.4	82.3 ^{ab}	0.80
NK cells	NK	172.8	88.9	99.4	80.3 ^{bc}	0.80
CD4 ⁺ T cells	CD4T	174.1	88.3	99.4	84.3 ^a	0.84
CD8 ⁺ T cells	CD8T	183.8	88.6	99.4	83.1 ^a	0.81
CD4 ⁺ CD8 ⁺ T cells	CD4CD8T	170.1	91.0	99.4	80.1 ^c	0.83
SWC6 $\gamma\delta$ ⁺ T cells	SWC6gdT	168.1	88.6	99.4	84.1 ^a	0.81

*Letters indicate statistically significant differences between cell types

2.4 Results

2.4.1 Porcine immune cells exhibit unique DNA methylation patterns associated with cell-specific co-receptor activation and biological processes

We generated 146-184M WGBS reads for each immune cell subpopulation, of which 88.3-91.0% aligned to the *S. scrofa* reference genome (Table 2.1). We observed bisulfite conversion rates >99% and Pearson correlation coefficients between replicates >0.8, meeting the standards for WGBS libraries previously set by ENCODE (<https://www.encodeproject.org/data-standards/wgbs/>). Average global CpG methylation rates ranged from 80.1% to 84.3%, with significant differences observed between cell types. Methylation rates of myeloid, CD4T, CD8T, and SWC6gdT cells were significantly higher than those of CD21pB, NK, and CD4CD8T cells. We assessed whether global methylation rates were associated with corresponding expression of DNA methyltransferases (*DNMTs*) and observed a moderate positive correlation with *DNMT3A* ($r=0.469$, $p=0.057$), a marginal negative correlation with *DNMT1* ($r=-0.249$, $p=0.336$) and a

significant negative correlation with *DNMT3B* ($r=-0.517$, $p=0.034$; Figure A.1). When taking into account the positive association with *DNMT3A* and the negative associations with *DNMT1* and *DNMT3B*, overall *DNMT* abundance was most significantly correlated with CpG methylation ($r=0.649$, $p=5.00E-3$), demonstrating that combined *DNMT* expression explained a significant proportion of global methylation variation. DNMT1 and DNMT3B play important roles in methylation of replicating DNA by methylating newly synthesized DNA and centromeric regions, respectively [103]. Indeed, relative centromeric methylation in porcine immune cells was positively correlated with *DNMT3B* expression ($r=0.613$, $p=5.13E-03$), suggesting that DNMT3B activity may be localized to these regions. Furthermore, we observed that the genes most significantly correlated with global CpG methylation were enriched for DNA replication GO terms (data not shown). We have thus identified methyltransferase and DNA replication genes as significant correlates with DNA methylation levels, suggesting that these processes have a significant impact on methylation signatures in porcine immune cells.

We identified 54,391 regions at which methylation rate was significantly associated with cell type, hereby classified as cell differentially methylated regions (cDMRs). A principal component analysis (PCA) revealed that PC1 and PC2 explained 41.1% and 29.2% of the variance in cDMR methylation, respectively, and clearly separated cell types into those of the B cell, T cell, and myeloid lineages (Figure 2.1A). To determine if differential immune cell methylation was occurring within expected genes, we scanned for cDMRs within genes encoding proteins used for cell sorting. We identified multiple cDMRs within the *CD3* multi-gene locus at which T cell methylation was significantly lower than in other cell types (Figure 2.1B). Furthermore, NK cells, known to express *CD3D* and *CD3G* but not *CD3E*, exhibited relatively low methylation at *CD3D*

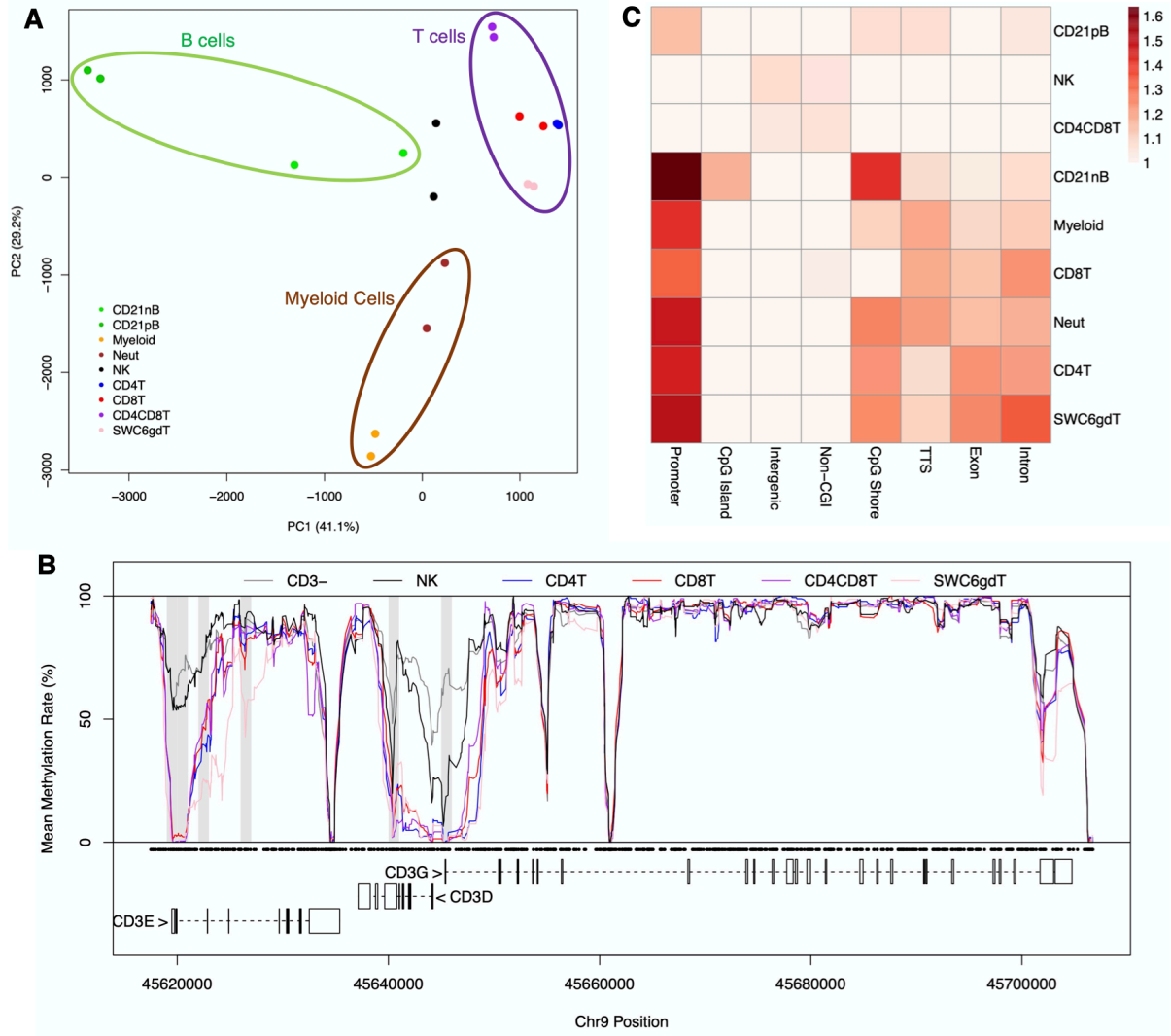


Figure 2.1. Evidence for lineage-specific immune cell differential methylation. (A) Principal component analysis plot of porcine immune cells based on cell differentially methylated region (cDMR) methylation rate. (B) Methylation rates across the *CD3* gene locus for expressing and non-expressing cell types. Gray boxes indicate cDMRs, and black dots indicate CpG coordinates. (C) Heatmap of immune cell lowly methylated region (cLMR) enrichment scores in genomic features.

and *CD3G* promoters but high methylation comparable to other CD3ε- cell types at the *CD3E* promoter, demonstrating the high resolution of WGBS libraries. We observed additional cDMRs exhibiting hypomethylation at the immune cell marker genes *CD4* (in CD4T cells and CD4CD8T cells), *CD8A* (in NK cells, CD8T cells, and CD4CD8T cells), *SIRPA* (in myeloid cells and

neutrophils) and *CD19* (in B cells; Figure A.2), revealing putative novel sites of gene regulation associated with receptor expression.

To further classify cDMRs based on cell types in which low methylation was observed, we designated all cDMRs with methylation rate <75% and z-score <-1 in a given cell type as cell lowly methylated regions (cLMRs). A total of 32,737 cLMRs were identified, ranging from 1,196 in CD21nB cells to 13,701 in CD21pB cells (Table A.1). We mapped cLMRs to gene features as well as CpG Islands (CGIs) and identified feature- and cell-specific enrichment (Figure 2.1C). cLMRs in six of the nine immune cell types were most significantly enriched in gene promoters, and to a lesser extent in other gene features (exons, introns, and TTSs) as well as CpG shores. Conversely, little to no enrichment was observed in intergenic regions, CGIs, and non-CGIs across cLMRs, suggesting that cell-specific low methylation indicative of gene regulation is more prominent in genomic regions proximal to genes as well as loci of intermediate CpG density. LMRs in three cell types (CD21pB, NK and CD4CD8T cells) were not enriched for any genomic features, which is in agreement with the observed global hypomethylation of these immune cell types relative to other cell types. We submitted cLMR-overlapping genes of each cell type for GSEA and identified uniquely enriched GO terms associated with respective cell functions (Table 2.2). Notably, we identified significant enrichment of CD4CD8T cell LMR genes for processes related to interferon gamma production and interleukin-15 signaling, in agreement with known function of CD4CD8T cells in the porcine periphery [104]. Overall, these results indicate a strong association between immune cell differential methylation and marker gene expression as well as genes involved in cell-specific biological processes.

Table 2.2. Enriched GO terms among cell lowly methylated genes

GO Term	No. Genes	Enrichment	FDR
CD21nB cell LMR Genes			
Regulation of B cell receptor signaling pathway	9	5.54	1.06E-02
B cell activation	32	3.14	2.17E-05
B cell differentiation	21	2.95	3.62E-03
Neut LMR Genes			
Positive regulation of cell motility	107	1.81	1.25E-05
Neutrophil activation involved in immune response	94	1.80	7.59E-05
Neutrophil degranulation	93	1.80	9.29E-05
CD4T cell LMR Genes			
Alpha-beta T cell differentiation	16	6.9	1.85E-06
CD4-positive, alpha-beta T cell differentiation	11	6.81	2.24E-04
T-helper cell differentiation	8	6.67	3.95E-03
CD4CD8T cell LMR Genes			
Interleukin-15 mediated signaling pathway	8	5.32	1.65E-02
Positive regulation of interferon-gamma production	21	2.71	6.90E-03
Regulation of interferon-gamma production	27	2.22	1.14E-02

2.4.2 Immune cell differential methylation is strongly associated with differential gene expression

To determine if immune cell differential methylation was associated with differential expression of overlapping genes, we utilized bulk RNA-sequencing data generated from the same porcine immune cell populations [81] and tested for significant associations between cDMRs and transcript abundance of overlapping genes. Methylation rates of promoter and intragenic cDMRs were heavily biased towards being significantly correlated with local transcript abundance compared to a random sampling of regions from each respective feature (Figure 2.2A-B). For cDMRs in

intergenic regions, we identified the gene with the most proximal transcription start site (TSS) and saw a similar overrepresentation of significant correlations between intergenic methylation and gene expression as those for cDMRs and overlapping genes (Figure 2.2C). cDMRs across gene features were more enriched for negative associations with transcript abundance, although a small but significant enrichment for positive associations was also evident (Figure 2.2D). We observed that genes in which cDMR methylation was positively correlated with transcript abundance were more lowly expressed on average compared to genes with negative methylation-transcript abundance associations, and furthermore that these positively correlated cDMRs exhibited higher methylation rates on average (data not shown). Within immune cell marker genes, cDMRs exhibited methylation rates that were significantly negatively correlated with abundances of corresponding transcripts (Figure A.3). Furthermore, by identifying the most proximal TSSs to intergenic cDMRs, we identified a region approximately 19kb upstream of the *SIRPA* TSS that was lowly methylated in myeloid cells and neutrophils (Figure 2.2E). While there was no evidence of TSS-proximal differential methylation between *SIRPA*-expressing and non-expressing cell types, methylation at this upstream locus was significantly negatively correlated with *SIRPA* abundance ($r = -0.94$), suggesting putative enhancer-like function. A region of open chromatin as measured by ATAC-seq has been identified in the same region in myeloid cells (Juber Herrera-Urbe, personal communication), providing additional evidence for a myeloid-specific enhancer element at this locus. Collectively, these results elucidate context-specific associations between DNA methylation and both proximal and distal gene expression, and highlight putative sites of transcriptional epigenetic gene regulation.

We calculated the degree to which cLMRs associated with cell-enriched gene expression using the same RNA-seq data sets. We identified 2,895 genes exhibiting enriched gene expression

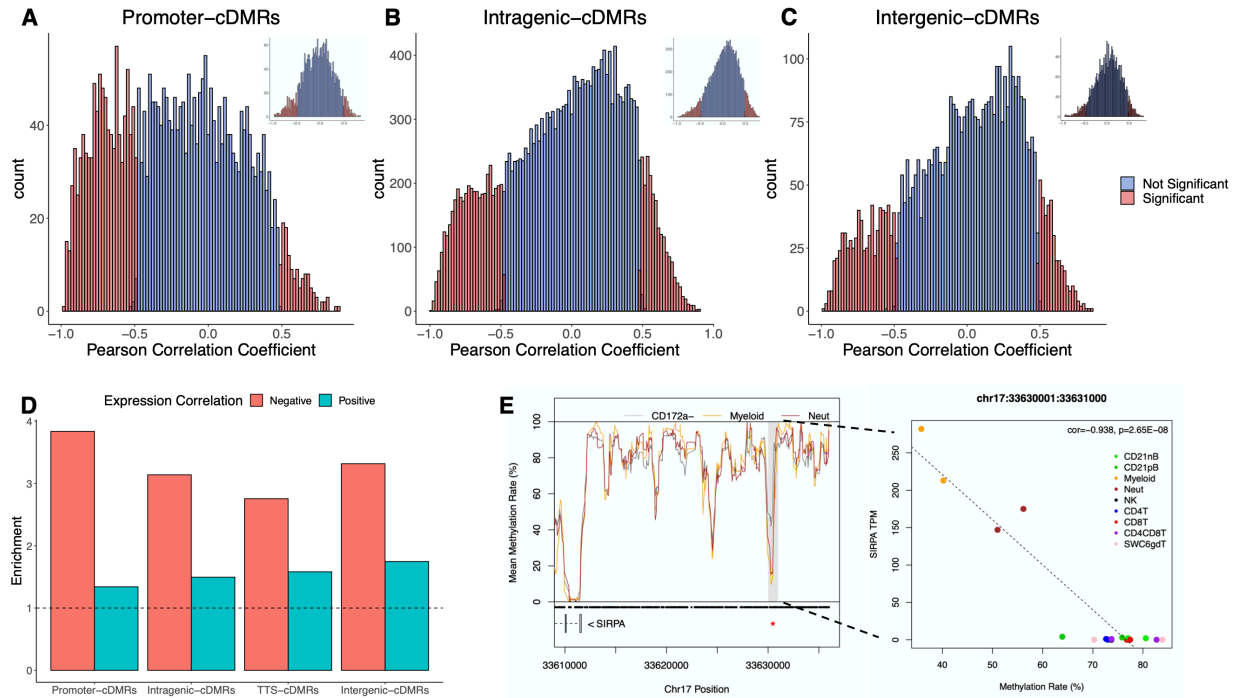


Figure 2.2. Porcine immune cell differential methylation is associated with differential gene expression. (A-C) Histograms of cDMR methylation-gene expression Pearson correlation coefficients in promoter, intragenic, and intergenic regions, inset by distributions at a random sampling of regions overlapping each feature. (D) Enrichment of cDMRs for regions positively and negatively correlated with gene expression, separated by gene feature. (E) A myeloid cell and neutrophil lowly methylated region located ~19kb upstream of the *SIRPA* gene is significantly negatively correlated with *SIRPA* abundance.

in one or more cell types, ranging from 244 genes in CD4CD8T cells to 1,261 in myeloid cells (Table A.2). Overall, cLMRs were highly overrepresented among expression-enriched genes of the same or related cell types (Figure 2.3A-C). Promoter and TTS cLMRs tended to have stronger and exclusive associations with expression-enriched genes of the same cell type; however, this pattern was also observed to a lesser extent among intragenic cLMRs, demonstrating that low methylation outside of promoter regions is associated with enriched gene expression. We provide here ample evidence that cell-specific differential methylation in porcine immune cells is highly correlated with transcript abundance and co-localizes with genes associated with immune cell state.

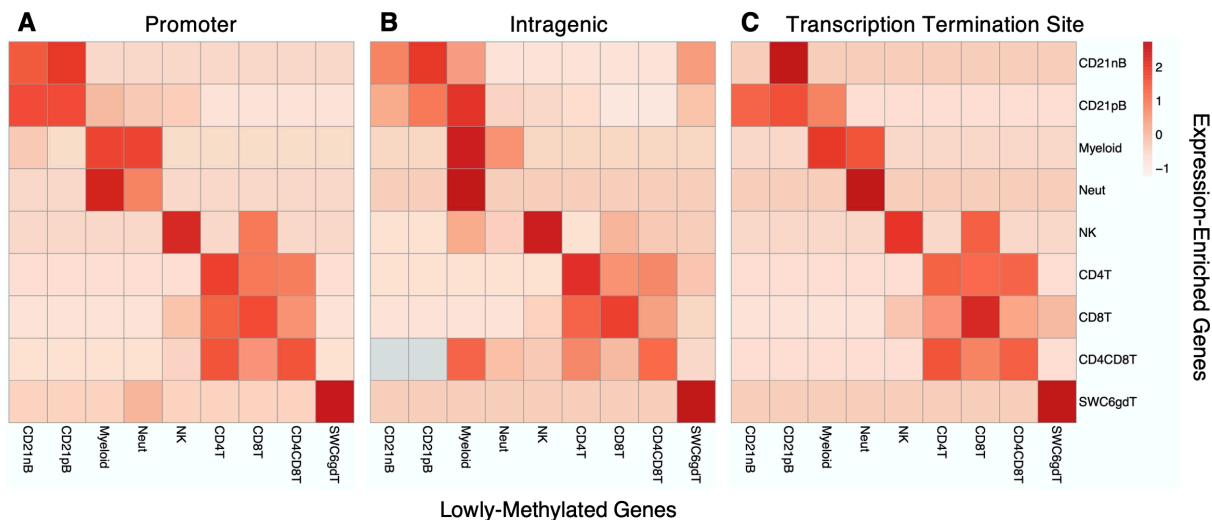


Figure 2.3. Cell lowly methylated regions (cLMRs) are enriched within expression-enriched genes. (A-C) Heatmap of normalized enrichment p-values between expression-enriched genes and genes overlapping promoter, intragenic, and transcription termination site cLMRs for each cell type.

2.4.3 Cell lowly methylated regions are enriched for cell-specific transcription factor binding motifs

Because DNA methylation is known to inhibit TF activity, and lowly methylated regions are thus more likely to be permissive to TF binding, we submitted cLMR and random control sequences for analysis of motif enrichment to identify enriched TF binding motifs. Clustering of cells based on cLMR enrichment for 1808 human motifs grouped cell types into B cells, myeloid cells, NK and CD4CD8T cells, and the remaining T cell subpopulations (data not shown). The most enriched binding motifs among cLMRs were highly cell-type specific, and many play an important role in the respective cell type's development, maturation, and function (Figure 2.4A). These included motifs for transcription factor E2-alpha (TCF3) in lymphocytes and the CCAAT enhancer binding protein (CEBP) family in myeloid cells. Other binding motifs were enriched among lymphocyte subtypes: motifs of early B cell factor 1 (EBF1) and paired box 5 (PAX5) were enriched among B cell LMRs, while transcription factor 7 (TCF7) and lymphoid enhancer binding factor 1 (LEF1) motifs were enriched in $\alpha\beta$ TCR cell (CD4T, CD8T, CD4CD8T) LMRs. Several TFs possessed

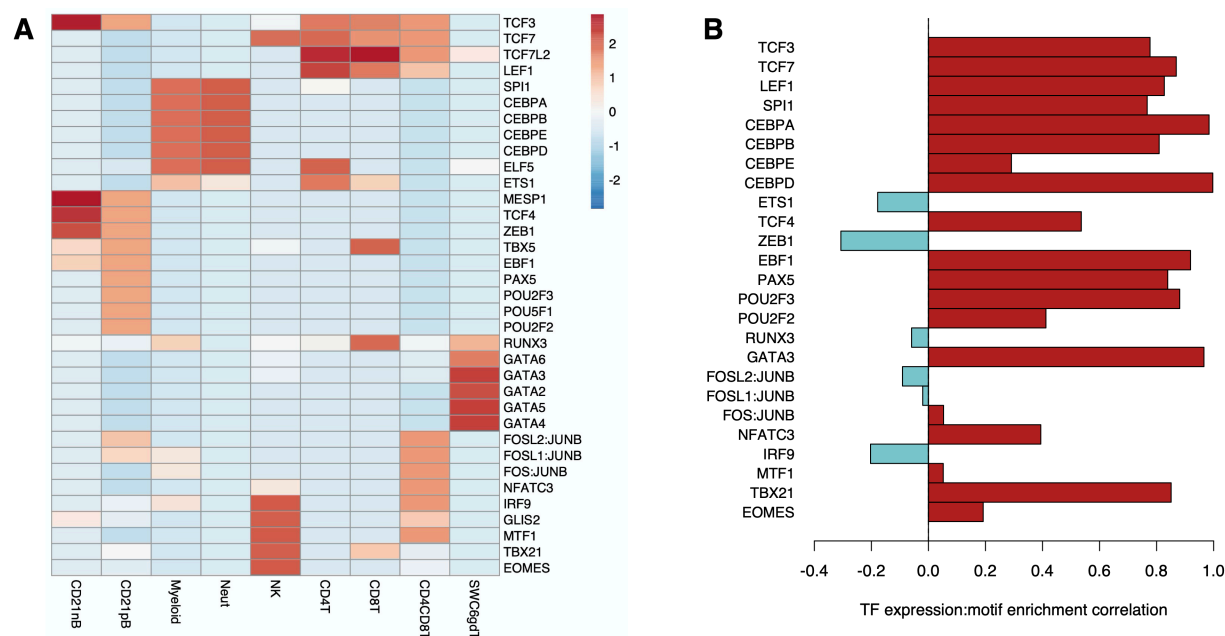


Figure 2.4. Porcine immune cell lowly methylated regions (cLMRs) co-localize with unique transcription factor (TF) binding motifs. (A) Normalized heatmap of enrichment scores for most significantly enriched TF binding motifs among cLMRs. **(B)** Pearson correlation coefficients between TF expression and $-\log_{10}$ p-value of motif enrichment among cLMRs. Red and blue bars indicate positive and negative correlations, respectively.

enriched motifs among LMRs for a single cell type, such as T-bet (TBX21) and Eomesodermin (EOMES) among NK cell LMRs, Fos:JunB heterodimers among CD4CD8T cell LMRs, and GATA binding proteins (GATAs) among SWC6gdT cell LMRs.

To determine if binding motif enrichment among cLMRs was associated with increased expression of corresponding TFs, we assessed transcript abundances for all TFs in Figure 2.4A. Of the 25 expressed TFs, 13 exhibited transcript abundances that were significantly positively correlated ($r > 0.5$) with cLMR enrichment for the corresponding binding motif, and all but 6 correlations were positive (Figure 2.4B). Furthermore, many of the TFs exhibited significantly enriched expression in the same cell types for which their binding motifs were enriched among cLMRs: *EBF1*, *POU2F3*, *PAX5*, *TCF3*, and *TCF4* (B cells), *TCF7* and *LEF1* (T cells), *TBX21* (NK cells), *CEBPA/B/D* and *SPI1* (myeloid cells), and *GATA3* (SWC6gdT cells) (Figure A.4).

These data support the conclusion that regions of cell-specific low methylation are associated with regulatory potential of biologically-relevant trans-acting factors regulating immune cell development and function.

2.4.4 Immune cell differential methylation co-localizes with candidate loci for immune capacity and disease traits

To determine if observed immune cell differential methylation was associated with genomic regions influencing economically important traits, we identified cDMRs harboring previously-reported GWAS SNPs in the Pig QTL Database. Among various trait classes, we identified cDMR enrichment exclusively for SNPs associated with immune capacity traits, and no significant enrichment for SNPs associated with growth, reproductive, or behavioral traits (Figure 2.5A). Fifty-three immune capacity GWAS SNPs overlapped cDMRs, and several were within biologically-relevant genes exhibiting transcript abundance that was significantly correlated with regional methylation rate (Table 2.3). We identified a SNP associated with the traits ‘CD8⁺ T cell percentage’ and ‘CD8⁺ T cell percentage’ that co-localized with a CD8T cell LMR within *CD8B* (Figure 2.5B) [105]. This gene encodes the beta subunit of the CD8 T cell co-receptor and exhibited the highest transcript abundance in CD8T cells (log2FoldChange=6.32). The GWAS SNP (rs81371115, 4:g.57971247C>T) lies in an intronic CpG upstream of exon 2. To determine if this SNP was also associated with local gene expression, we quantified transcript abundances of *CD8B* and *CD8A* via RT-qPCR in PBMCs across rs81371115 genotypes in an unrelated MSU pig resource population. While *CD8B* abundance was too low to detect significant differences, SNP genotype was significantly associated with *CD8A* expression (p=0.027), with the TT genotype exhibiting significantly lower abundance than CC and CT genotypes (Figure 2.5C). These results corroborate previous findings that rs81371115 genotype is associated with CD8⁺ T cell

composition phenotypes, and demonstrate the regulatory potential of the LMR overlapping this SNP in influencing CD8 co-receptor expression.

Lastly, we also queried disease-associated GWAS SNPs for co-localization with cDMRs, and identified a myeloid and NK cell LMR in *N-Myc Downstream Regulated 1 (NDRG1)* harboring a SNP associated with number of mummified piglets (rs327164077, 4:g.8034723T>C). The phenotype of mummified piglets is often caused by porcine reproductive and respiratory syndrome virus (PRRSv) infection in the sow [106], and *NDRG1* has previously been implicated in PRRSv response [107]. In adult porcine PBMCs, we observed significantly lower *NDRG1* abundance in the CT genotype relative to CC (p=0.001) while the TT genotype exhibited intermediate expression, providing evidence of a shared genetic association for reproductive performance and *NDRG1* expression (Figure 2.5D). In summary, our results have enhanced

Table 2.3. SNPs within cLMRs associated with transcript abundance

SNP Pos	Trait	cLMR Cell	Gene	Meth-TPM cor	Meth-TPM pvalue	Expression Enrichment
3:30799911	CD8+T%, CD8-T%	CD4T, SWC6gdT	<i>SNX29</i>	0.673	0.003	CD21pB
3:56530168	CD8+T%, CD8-T%, CD3-CD8-T%	CD21nB, SWC6gdT	<i>ZAP70</i>	-0.752	0.0005	CD21nB, SWC6gdT, Neut
3:57971274		CD8T	<i>CD8B</i>	-0.839	2.56E-05	CD8T, NK
3:58968002		NK	<i>ATOH8</i>	-0.715	0.0012	CD8T, NK
8:30368512	Lymphocyte percentage, 20 d	CD21pB, Myeloid	<i>KLHL5</i>	-0.446	0.073	Myeloid
4:8034723	No. of mummified pigs	Myeloid, NK	<i>NDRG1</i>	-0.614	0.0087	Myeloid

functional annotation in genomic regions at which genetic variation is associated with immune traits, and have further linked such genetic variation to local gene expression.

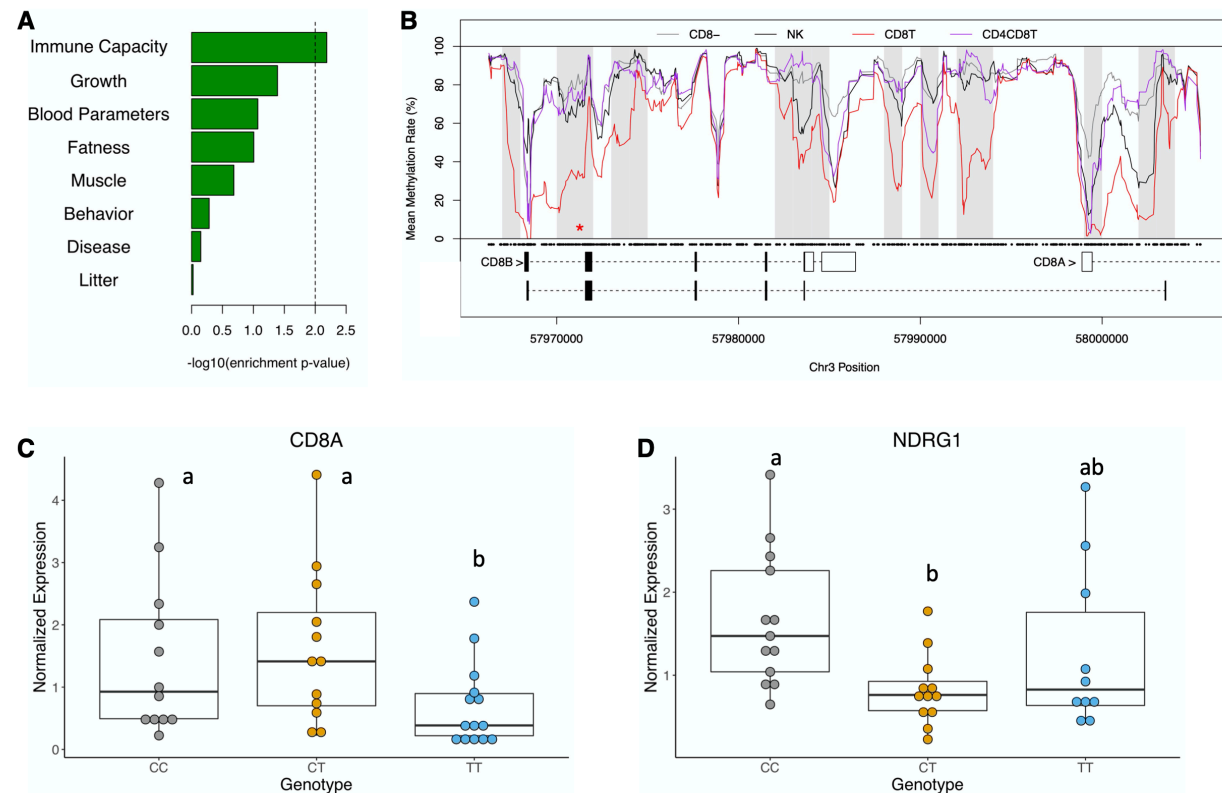


Figure 2.5. Immune cell differentially methylated region (cDMRs) harbor immune capacity GWAS SNPs. (A) Enrichment barplot of GWAS SNPs for various trait classes among cDMRs. Dashed line indicates threshold for significance. (B) *CD8B/CD8A* locus methylation plot including a CD8T cell LMR overlapping a CD8⁺ T cell %/CD8⁺ T cell % GWAS SNP (red asterisk). Gray boxes indicate cDMRs, and black dots indicate CpG coordinates. (C&D) Normalized *CD8A* and *NDRG1* abundance in porcine peripheral blood mononuclear cells by *CD8B* and *NDRG1* SNP genotype, respectively. Different letters indicate statistically significant differences (FDR<0.05).

2.5 Discussion

We report here the first genome-wide assessment of DNA methylation in different porcine circulating immune cells, and correlate methylation with gene expression of the same sorted cells. Using these data, we have identified thousands of genomic regions where methylation rate is associated with immune cell type, which suggests that these regions play important roles in cell-

specific gene regulation. Previous ENCODE studies have surveyed DNA methylation in B and T cell subpopulations from healthy individuals and leukemia patients, and have reported similar trends of cell-specific DNA methylation patterns between cell types and states [25, 41]. We have surveyed a greater diversity of circulating immune cell populations in the pig derived from both myeloid and lymphoid lineages, including T cell populations that are uniquely overrepresented in the pig relative to human and mouse. Our DNA methylation analyses thus provide unique insights into epigenetic gene regulation in understudied immune cells.

We observed global DNA methylation rates between 80-84% that are comparable to those reported across mammalian species [27]. Among cell types, CD21pB cells, NK cells, and CD4CD8T cells possessed significantly lower methylation rates relative to other immune cells. Previous studies have shown that B and NK cell development and activation are associated with global DNA hypomethylation [108–111]. We identified significant correlates with global methylation rates, most notably abundance of DNA methyltransferase transcripts. While DNMT3A, DNMT1 and DNMT3B are all responsible for catalyzing cytosine methylation, transcript abundances of the latter two were negatively correlated with global CpG methylation rates in our study, such that their abundances subtracted from *DNMT3A* abundance explained the greatest proportion of variance in immune cell methylation rate. DNMT3A is responsible for *de novo* DNA methylation, thus higher expression would be expected to result in increased methylation rates genome-wide [112]. The negative correlations between global methylation and *DNMT1* and *DNMT3B* abundance may be due to their associations with replicating DNA and dividing cells. Newly synthesized DNA is inherently hemi-methylated following replication, and requires the activity of maintenance methyltransferase *DNMT1* to conserve CpG methylation [113]. Furthermore, *DNMT3B*, a *de novo* methyltransferase, can promote mitotic division by

methyating centromeric regions and establishing chromosome stability [103]. We provide evidence that global hypomethylation of cell types may be associated with increased rates of cell division due to their higher rates of relative centromeric methylation as well as increased expression of genes involved in DNA replication and cell division. Overall, these data suggest that variation in genome-wide DNA methylation levels can be explained in part by differences in *DNMT* expression and, potentially more directly, mitotic capacity of cell types.

We identified tens of thousands of genomic loci exhibiting differential methylation associated with porcine immune cell type. As regions of epigenetic variation, cDMRs represent likely regions of gene regulation that are associated with cell identity and function. In support of this claim, clustering by cDMR methylation rate broadly grouped samples by cell lineage, and numerous cDMRs were present within immune cell-specific co-receptors. The majority of cDMRs were further classified as cLMRs for the cell types known to express the encoded marker, suggesting that these loci may promote expression of the resulting transcript. While cLMRs of some marker genes were limited to promoter regions (e.g. *CD4*, *CD19*), others exhibited cell-specific low methylation in intragenic regions (e.g. *CD8A*, *SIRPA*). Lowly methylated intragenic regions associated with gene activation have been shown to be predictive of intronic enhancer elements in human tissues and cell lines [25], suggesting similar regions in porcine immune cells are promising candidates for distal regulatory elements. Among cell lowly methylated genes, we identified enrichment for cell-specific processes for the majority of cell types, demonstrating that low methylation genome-wide occurs disproportionately in functionally-relevant loci.

By integrating methylation data with transcript abundance data from the same porcine immune cell samples, we identified an overrepresentation of regions at which methylation rate was significantly correlated with expression of overlapping genes. Intergenic cDMRs were also

disproportionately correlated with abundance of the most proximal gene's transcript, signifying putative roles for these regions as distal regulators of gene expression. The majority of significant cDMR-transcript abundance correlations—particularly in promoter regions—were negative, which is in agreement with the predominantly repressive role of DNA methylation at promoter and enhancer regions previously reported in other mammals [25, 32]. However, enrichment for positive correlations between cDMR methylation and expression was also observed, and was more common in genes exhibiting lower expression and higher methylation. This finding is in agreement with previous work reporting expression-dependent relationships between differential gene methylation and transcript abundance: for genes of lower basal expression levels, methylation increases with increasing transcription rates due to increased chromatin opening by RNA Polymerase II (Pol II), but at a certain expression threshold methylation disrupts RNA Pol II and vice versa [38]. We demonstrate that integrating immune cell methylation data with gene expression data can identify putative distal-acting regulatory elements, exemplified by a putative myeloid enhancer upstream of *SIRPA* at which methylation was negatively correlated with *SIRPA* abundance. Cell-enriched gene expression was most strongly correlated with lowly methylated regions, suggesting that any active role DNA methylation plays in regulating porcine immune gene expression is largely suppressive. The colocalization of enriched genes with LMRs across all gene features signify the importance of intragenic epigenetic gene regulation in maintaining cell-type specific gene expression.

We have shown that immune cell differential methylation is associated with regions harboring TF binding motifs, many of which govern cell-specific functions. It has long been understood that differential methylation acts in coordination with changes in trans-acting factor binding to regulate gene expression, although the exact mechanisms governing this relationship

remain debated and are likely context-specific [114]. In general, DNA hypomethylation in promoters and enhancers is associated with increased binding of activating TFs, while hypermethylation is associated with exclusion of activating TFs and recruitment of transcriptional repressors [31, 82]. Human and mouse DNA methylation studies report enrichment for binding sites of functionally-relevant TFs within LMRs unique to tissues and cell types [32, 41, 85]. Similarly, we identified enriched TF motifs among cLMRs that play known roles in regulating development and function, including: the CEBP family of TFs in myeloid cells [115–117]; EBF1 and PAX5 in B cells [118, 119], TCF7 in T cells [120], and EOMES and T-bet/TBX21 in NK cells [121]. Furthermore, we identified overrepresented motifs within porcine-enriched T cell LMRs that provide insight into gene regulation in these subpopulations. Multiple GATA TF motifs were enriched solely among SWC6gdT cell LMRs in our analysis. Among these, *GATA3* has previously been shown to be highly expressed in porcine $\gamma\delta$ T cells, and also exhibited enriched expression in our study ($\log_2\text{FoldChange}=4.89$) [122]. It is thus likely that *GATA3* plays a critical role in mature $\gamma\delta$ T cells that is unique from its role in other T cell subtypes. A recent single-cell RNA-sequencing analysis in PBMCs from the same individuals reported greater *GATA3* expression in $\text{CD2}^- \gamma\delta$ T cell clusters relative to $\text{CD2}^+ \gamma\delta$ T cell clusters, suggesting even further restriction of *GATA3* activity to certain $\gamma\delta$ T cell subpopulations [81]. In CD4CD8T cell LMRs, we identified enrichment of multiple AP-1 TF complex motifs, with c-Fos:JunB heterodimers being the most enriched. *c-fos* deficiency in mice has previously been linked to decreased AB T cell production in cultured thymocytes [123]. However, the presence of lowly methylated AP-1 binding sites in mature DP T cells indicates a potential outsized role for these TFs beyond thymic T cell development. AP-1 TF binding has been shown to activate IFN γ , which DP T cells are known to

express at high levels [124]. Future studies should seek to assess the genome-wide consequences of AP-1 motif hypomethylation and its association with AP-1 activity in porcine DP T cells.

Lastly, we provide evidence that differential methylation is associated with reported QTL for pig immune traits. Previous studies have identified tissue- and cell-specific enrichment of LMRs for human and livestock GWAS SNPs of relevant traits [41, 83, 84]. We similarly identified immune cell DMR enrichment exclusively for pig immune capacity traits, and not for traits related to growth, reproduction, and behavior that are less likely to be directly impacted by immune system gene regulation. In addition to observed overall enrichment, cDMRs overlapped GWAS SNPs within strong candidate genes for associated traits. These included a SNP associated with various T cell subpopulation percentages in a CD8T cell LMR overlapping the *CD8B/CD8A* locus. We show here that the same SNP is associated with PBMC *CD8A* expression in a separate pig population, demonstrating that CD8 co-receptor transcript abundance may serve as an intermediate molecular phenotype linking genetic variation with previously observed variation in T cell composition. Furthermore we also show that a SNP within a myeloid and NK cell LMR overlapping *NDRG1* is significantly associated with *NDRG1* expression, while also having been previously associated with number of mummified piglets [125]. Both reproductive performance and *NDRG1* expression have been shown to be negatively impacted by PRRSV infection, with downregulation of *NDRG1* contributing to increased PRRSV replication rate [106, 107]. It is thus intriguing to hypothesize that *NDRG1* expression may act as an intermediate molecular phenotype linking *NDRG1* genotype with observed reproductive variation, and that this relationship may be further affected by viral infection.

In summary, we have vastly improved the understanding of epigenetic gene regulation in porcine immune cells through the generation of DNA methylation profiles across diverse cell

types. These data sets will contribute to the efforts of the Functional Annotation of Animal Genomes (FAANG) project by providing additional regulatory information in important porcine cell types. Additionally, we believe the results presented here will prove valuable in further understanding how cell-specific epigenomic modifications contribute to pig immune and disease phenotypes, and inspire future studies seeking to integrate functional annotation data to enhance the search for causative variants for complex traits.

2.6 Acknowledgements

Sequencing was performed at the Michigan State University Research Technology and Support Facility. Computing resources were provided by Michigan State University's High Performance Computing Cluster. The authors would like to thank Kristen Byrne for her assistance with lab work.

CHAPTER 3

PIG FETAL SKELETAL MUSCLE DEVELOPMENT IS ASSOCIATED WITH GENOME-WIDE DNA HYPOMETHYLATION AND CORRESPONDING ALTERATIONS IN TRANSCRIPT AND MICRO-RNA EXPRESSION

This chapter is currently being prepared for publication. It was prepared alongside co-authors Laura M. Ford, Nancy E. Raney, Jenna M. Grabowski, and Catherine W. Ernst.

3.1 Abstract

Fetal myogenesis represents a critical period of porcine skeletal muscle development and requires the coordinated expression of thousands of genes. While numerous studies have transcriptionally profiled prenatal skeletal muscle, the *cis*-regulatory dynamics associated with porcine myogenesis remain poorly characterized. Epigenetic mechanisms play important roles in regulating development and differentiation; among such mechanisms, DNA methylation exhibits context-specific associations with gene expression and has been shown to be highly dynamic during developmental processes. We performed whole-genome bisulfite sequencing to assess DNA methylation in pig fetal *longissimus dorsi* (LD) muscle at 41 and 70 days of gestation (dg), as well as RNA- and small RNA-sequencing in the same samples to identify coordinated changes in methylation and expression between myogenic stages. We observed global loss of DNA methylation in LD muscle at 70 dg: among the 45,739 differentially methylated regions (DMRs) identified, the majority (N=34,232) were hypomethylated at 70 vs. 41 dg. Developmental DMRs exhibited feature-specific enrichment in gene regulatory regions, as well as in regions proximal to micro-RNAs (miRNAs) that play known roles in myogenesis. Integration of DNA methylation and gene expression data revealed strong associations between differential transcript abundance and gene methylation. Of particular note was the finding that differential miRNA methylation was

significantly negatively correlated with abundance, and that differential expression patterns continued for assayed miRNAs well into postnatal life. Motif analysis revealed significant enrichment of myogenic regulatory factor binding motifs among hypomethylated regions, suggesting that loss of methylation may function in part to increase accessibility of muscle-specific transcription factors. Lastly, we show that developmental DMRs are enriched for GWAS SNPs associated with muscle physiology and meat quality traits, demonstrating the potential for epigenetic processes in these regions to influence phenotypic diversity. Our results enhance understanding of DNA methylation dynamics in pig fetal skeletal muscle, and reveal putative *cis*-regulatory elements under epigenetic control during porcine myogenesis.

3.2 Introduction

Skeletal muscle accounts for the majority of carcass weight in pigs and other livestock species, and its physiology influences numerous meat quality phenotypes [126]. Fetal myogenesis represents a critical window in skeletal muscle growth and development: while muscle continues to grow postnatally via hypertrophy, the fiber numbers in most muscle types are established prenatally in two waves of myoblast fusion and differentiation [127]. In the pig, primary myogenesis occurs between 30 and 60 days gestation (dg), at which time myoblasts fuse *de novo* to form primary myotubes [45]. During secondary myogenesis (54-90 dg) myoblasts fuse using primary myotubes as a scaffold to form secondary myotubes [45]. It has long been recognized that the migration, proliferation, and differentiation of skeletal muscle precursors requires the coordinated expression of thousands of genes, and the family of myogenic regulatory factors (MRFs)—including MYOD1, MYF5, myogenin (MYOG), and MRF4/MYF6—have emerged as critical regulators of these processes [128]. Over the past two decades transcriptome-wide profiling of fetal skeletal muscle has been performed in multiple pig breeds [129–134]. Furthermore,

expression of small RNAs (smRNAs) in developing muscle has gained increased appreciation in recent years. In particular micro-RNAs (miRNAs)—a class of smRNAs known to inhibit gene expression via binding of target transcripts—have been shown to play important roles throughout muscle development [135], and a number of studies have assessed miRNA expression in prenatal and postnatal porcine skeletal muscle [136–139].

Despite insights into gene expression dynamics throughout muscle development, the *cis*-regulatory elements driving expression of RNAs and miRNAs during pig skeletal muscle development have remained understudied, particularly those regulated by epigenetic processes. Among epigenetic modifications, DNA methylation is the most prevalent modification made to the DNA molecule, and in mammals primarily involves the enzymatic addition of a methyl group to the 5-position of cytosines in CpG dinucleotides. DNA methylation has been shown to exhibit context-specific associations with gene expression: methylation in promoter and enhancer regions is generally inversely correlated with transcript abundance, either through the alteration of transcription factor (TF) binding sites or the recruitment of transcriptional repressors and chromatin-modifying enzymes [25, 31, 82]. Intragenic DNA methylation has been shown to be highly gene- and site-specific, due in part to the high frequency of intronic regulatory elements [25]. Several studies have recently reported DNA methylation patterns in adult porcine skeletal muscle as well as across developmental stages [140–144]. However, associations between DNA methylation and RNA and smRNA expression have not been extensively studied, limiting understanding of epigenetic gene regulation by DNA methylation during myogenesis.

We have performed whole-genome bisulfite sequencing (WGBS) in tandem with RNA-sequencing (RNA-seq) and smRNA-seq in porcine fetal *longissimus dorsi* (LD) muscle at 41 dg and 70 dg, representing stages of primary and secondary myogenesis, respectively. The goals of

the current study were to identify differentially methylated regions (DMRs) between myogenic stages and characterize their relationships with differential RNA and miRNA abundance. We demonstrate that LD muscle DNA is extensively hypomethylated during developmental progression, and is strongly associated with gene expression dynamics as well as muscle-specific *cis*-regulatory elements. Our results greatly increase understanding of epigenetic gene regulation during pig fetal skeletal muscle development and reveal putative regulatory regions that may contribute to variation in skeletal muscle phenotypes.

3.3 Materials & Methods

3.3.1 Sample Collection and Nucleic Acid Isolation

LD muscle was sampled from fetuses of three Yorkshire x Landrace gilts each at 41 dg and 70 dg. 41 dg littermate samples were pooled irrespective of sex, while pooled 70 dg samples contained tissue from only female fetuses. DNA was isolated using the PureLink Genomic DNA Kit (Invitrogen), and quality and quantity were assessed on the Qubit fluorometer. DNA samples were spiked with unmethylated lambda phage DNA (5 ng lambda DNA/ 1ug sample DNA) to assess bisulfite conversion rates. Sodium bisulfite conversion was performed using the Zymo EZ DNA Methylation-Gold Kit, and sequencing libraries were prepared using the Kapa Hyper Prep DNA Kit (Roche). WGBS was performed as described in Section 2.3.3.

RNA was isolated from five of the six LD samples using the TRIzol extraction method, and isolated from one sample (41dg_3) using the miRNAeasy mini kit (Qiagen). RNA quality and quantity were assessed on the Agilent Bioanalyzer. RNA-and smRNA-seq libraries were prepared using the Illumina TruSeq Stranded Total RNA Ribozero Library Preparation Kit and the Illumina TruSeq Small RNA Library Preparation Kit, respectively, following manufacturer's instructions.

Completed libraries were QC'ed and quantified using a combination of Qubit dsDNA High Sensitivity (HS) and Agilent Bioanalyzer HS DNA assays. RNA-seq (2x150bp PE) and smRNA-seq (1x50bp single-end) were performed on an Illumina HiSeq 4000 instrument.

3.3.2 Bioinformatics Analyses

WGBS read trimming, alignment, and methylation extraction were performed as described in Section 2.3.4. Due to the suboptimal conversion rates obtained across samples, we used the `filter_non_conversion` command from Bismark to remove aligned reads with ≥ 3 consecutive methylated non-CpG cytosines. This filtering resulted in homogeneous bisulfite conversion rates $> 99\%$ across samples (Table B.1).

RNA-seq and smRNA-seq reads were trimmed of adapters and low-quality bases using Trimmomatic [87]. Trimmed RNA-seq reads were aligned to the *Sus scrofa* reference genome using TopHat2 [145], and gene counts were obtained from uniquely aligned reads using HTSeq-count [146]. smRNA-seq reads were aligned to the *S. scrofa* reference genome and counts of mature miRNAs were obtained using the miRDeep2 suite [147].

3.3.3 Methylation Analyses

Regional differential methylation analysis between stages was performed using the *methylKit* R package [89]. Briefly, average CpG methylation rates were calculated for 1kb non-overlapping regions of the *S. scrofa* genome, and regions covered by at least 20 reads were retained for further analysis. Logistic regression models were fitted for each region and an analysis-of-deviance was performed to determine the significance of the stage parameter's inclusion in the model. Regions with a mean difference $> 10\%$ between stages and $FDR < 0.05$ were considered DMRs. Annotation of DMRs and gene set enrichment analysis (GSEA) of differentially methylated genes was

performed as described in Section 2.3.5. We classified any gene with a promoter-DMR as a promoter-DMG, and any gene with an intragenic-DMR an intragenic-DMG.

3.3.4 Differential Expression Analyses

Differential expression analysis of transcripts and miRNAs was performed using the *DESeq2* R package [93]. Briefly, sample-specific size factors were determined to account for differences in sequencing depth, and gene-specific dispersion parameters were estimated to define mean-variance relationships. Lastly, a negative binomial generalized linear model was fitted and a Wald test for the significance of stage coefficient (β_i) for gene counts was performed. Genes and miRNAs with an absolute \log_2 -fold-change > 1 between stages and a Benjamini-Hochberg adjusted p-value < 0.05 were considered differentially expressed. GSEA of DEGs was assessed using the Panther Database.

DNMT transcripts and miRNAs were selected for expanded expression profiling using reverse-transcriptase quantitative PCR (RT-qPCR). Briefly, total RNA at three additional ages (105 dg and 1 wk and 5 wk post-natal) was isolated using the TRIzol extraction method, and RNA and miRNA cDNA was reverse transcribed using the High-Capacity cDNA Reverse Transcription Kit and Taqman Advanced miRNA cDNA Synthesis Kit, respectively (Applied Biosystems). *PPIA* and *HPRT1* were selected as reference genes for *DNMT* RT-qPCR and let-7a and miR-30e were selected as reference assays for miRNA RT-qPCR based on their observed stable expression across assayed developmental stages (data not shown). qPCR reactions were performed in triplicate as described in Section 2.3.9, with the modification of using 50 ng of cDNA at 5 μ l total volume per reaction well. Statistical analyses of qPCR data were performed as described in Section 2.3.9 to test for significant differences between developmental stages.

3.3.5 Motif Enrichment Analyses

Hyper- and hypomethylated DMR sequences were submitted for analysis of motif enrichment using the MEME suite [94]. Analyses were performed as described in Section 2.3.7.

3.3.6 GWAS SNP Enrichment Analyses

We extracted genomic coordinates of peak GWAS SNPs for all published traits from the Pig QTL database [100]. For each trait, enrichment scores were calculated corresponding to the associated GWAS SNPs' enrichment within hyper- and hypomethylated DMRs.

3.4 Results & Discussion

3.4.1 Fetal pig skeletal muscle development is associated with genome-wide and regional hypomethylation

We obtained 146-167M WGBS PE reads per sample (Table B.1), from which 402,438 1kb genomic regions with sufficient coverage were retained for differential methylation analyses. 43,481 regions overlapped gene promoters, while 192,777 and 166,180 overlapped intragenic and intergenic regions, respectively. In total, 73.0% of annotated genes in the pig genome were covered by at least one region.

The greatest proportion of variance in LD methylation could be explained by gestational age (Figure B.1). Furthermore, we observed significant global DNA hypomethylation at 70 dg relative to 41 dg (Figure 3.1A; $t=15.72$, $p=6.24E-4$). To determine if global methylation differences were associated with differential expression of DNA methylation and demethylation enzymes, we assessed transcript abundances of DNA methyltransferases (*DNMTs*) and TET methylcytosine dioxygenases (*TETs*) between stages using RNA-seq data from the same samples. We observed significant decreases in abundance at 70 dg for *DNMT1* ($\log_2FC=-0.54$, $FDR=5.65E-5$), *DNMT3A* ($\log_2FC=-0.68$, $FDR=7.78E-5$), and *DNMT3B* ($\log_2FC=-0.62$, $FDR=9.63E-3$).

Conversely, members of the *TET* family, known to participate in active DNA demethylation [30], were significantly upregulated at 70 dg, including *TET1* (log2FC=0.61, FDR=1.48E-5) and *TET2* (log2FC=0.65, FDR=5.10E-5). Yang *et al.* observed similar decreases in *DNMT* expression and moderate increases in *TET* abundance during prenatal LD muscle development [141]; however we observe equal changes in magnitude among both gene families. We quantified expression of *DNMT* transcripts at additional stages of prenatal LD muscle development via RT-qPCR; while within-stage variances were too large to detect significant differences in *DNMT1* and *DNMT3A* abundance, we observed a significant decrease in *DNMT3B* abundance after 41 dg (Figure 3.1B; $p < 0.001$) that validates patterns observed in RNA-seq data and demonstrates that *DNMT3B* is primarily active in early muscle development. These data suggest that global hypomethylation of developing pig fetal skeletal muscle may be due in part to decreased DNA methyltransferase activity and increased activity of demethylating enzymes.

We identified 45,739 DMRs between gestational ages, of which 11,507 were hypermethylated at 70 dg relative to 41 dg and 34,232 hypomethylated (Table 3.1). We surveyed MRF genes to determine if differential methylation was present in expected genomic regions. Within the *MYF5* and *MYF6* locus, *MYF5* was significantly promoter-hypermethylated at 70 dg, whereas *MYF6* was significantly hypomethylated upstream of its transcription start site (TSS; Figure 3.1C). *MYF5* is the earliest expressed MRF and primarily functions in myoblast proliferation and determination [148], while *MYF6* functions in muscle cell differentiation [149, 150]. These patterns are thus in agreement with expected downregulation of *MYF5* and upregulation of *MYF6* as muscle development progresses, and demonstrate that differential methylation is evident at myogenic TFs.

Table 3.1. Summary of differentially methylated regions (DMRs) and genes (DMGs)

	DMRs	Promoter-DMGs	Intragenic-DMGs
Total (70dg vs 41dg)	45,739	3,635	7,269
Hypermethylated	11,507	1,414	2,438
Hypomethylated	34,232	2,221	4,831

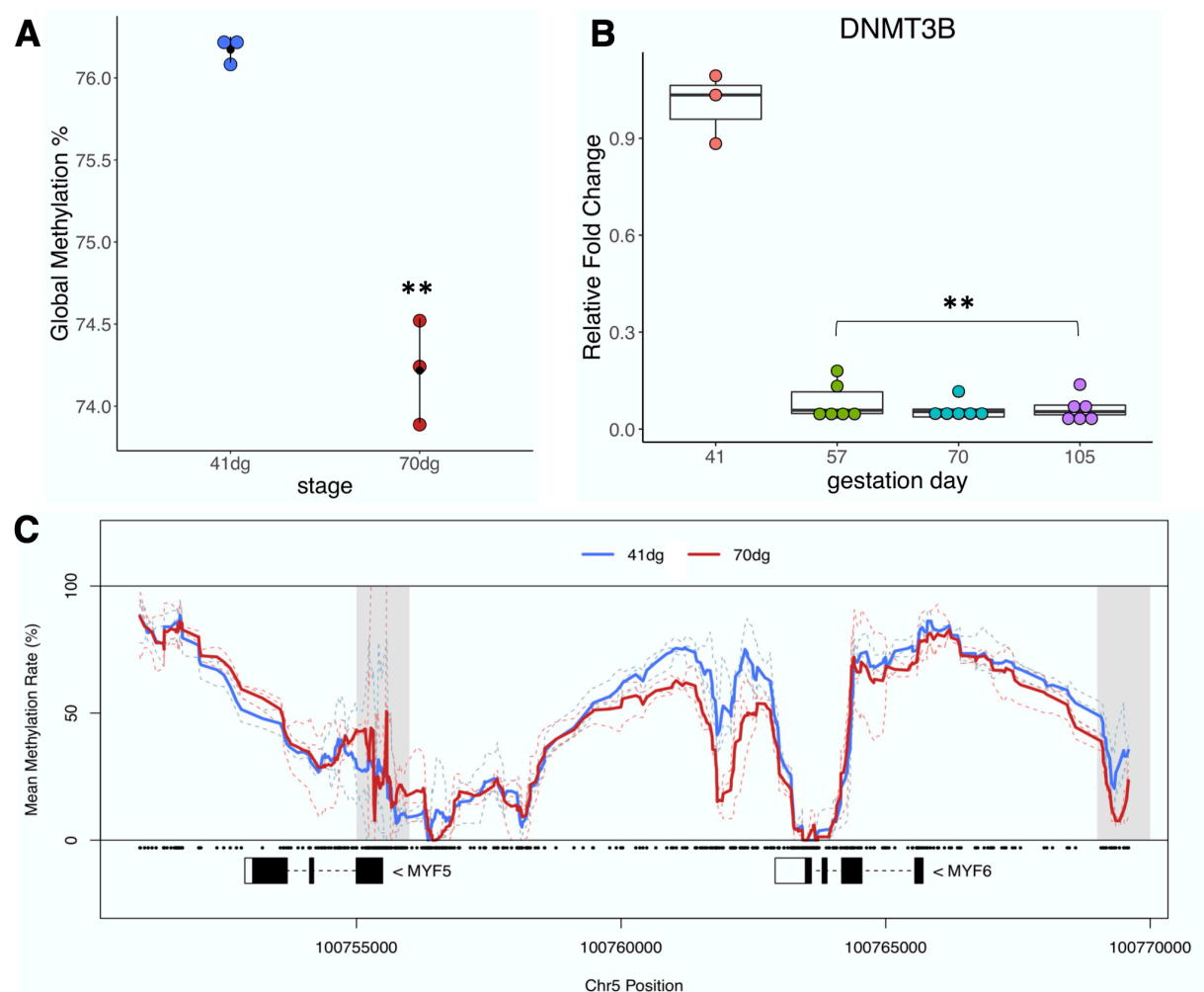


Figure 3.1. Fetal skeletal muscle development is associated with genome-wide and site-specific differential DNA methylation. (A) Dot plot of global DNA methylation of fetal *longissimus dorsi* (LD) muscle by developmental stage. **(B)** Fold change in *DNMT3B* abundance in fetal LD muscle relative to 41 days gestation across multiple developmental stages. **(C)** Gene methylation plot for muscle-specific transcription factors *MYF5* and *MYF6*. Gray boxes indicate differentially methylated regions, and dots indicate CpG coordinates. ** $p < 0.05$

3.4.2 Skeletal muscle differential methylation is enriched in gene features and micro-RNAs

To determine if differential methylation was overrepresented in specific genomic regions, we annotated DMRs with respect to gene features, CpG islands (CGIs) and shores, and miRNA genes. Enrichment analyses revealed feature-specific enrichment of hyper- and hypomethylated regions (Figure 3.2A). Hypermethylated regions were significantly enriched in gene promoters and miRNAs, as well as within CGIs and CpG shores. The magnitude of DNA hypomethylation significantly impacted feature enrichment—potentially due to the global hypomethylation observed at 70 dg—however enrichment for miRNAs and CpG shores was consistently observed. CpG islands and shores are strongly associated with gene promoters and other gene regulatory features [151]. While generally non-methylated, studies have shown that a small proportion of CGIs become methylated during normal tissue differentiation [152–154]. Our data showing dynamic CGI methylation in pig fetal LD muscle is in agreement with these previous findings, and also highlight a potential role for DNA methylation in regulating miRNAs.

Due to the enrichment of DNA hyper- and hypomethylation in the vicinity of miRNAs, we looked more closely at patterns of miRNA methylation with particular focus on a class of miRNAs termed ‘myomiRs’ that play important roles in regulating myogenesis [135]. We identified hypomethylated DMRs within 2kb of four intronic myomiR genes—*ssc-mir-1*, *ssc-mir-133a-1*, *ssc-mir-206*, and *ssc-mir-208a*—suggesting that miRNAs may be epigenetically regulated independently of their host genes to confer activation (Figure B.2A-D). Furthermore we also observed significant hypomethylation flanking *ssc-mir-378-2* and *ssc-mir-338* (Figure 3.2B-C), which encode miRNAs promoting muscle differentiation and mitochondrial respiration, respectively [155, 156]. Among the most significantly hypermethylated miRNAs were *ssc-mir-196a* and *ssc-mir-615*, both of which lie in the embryonic homeobox (*HOX*) gene cluster and have

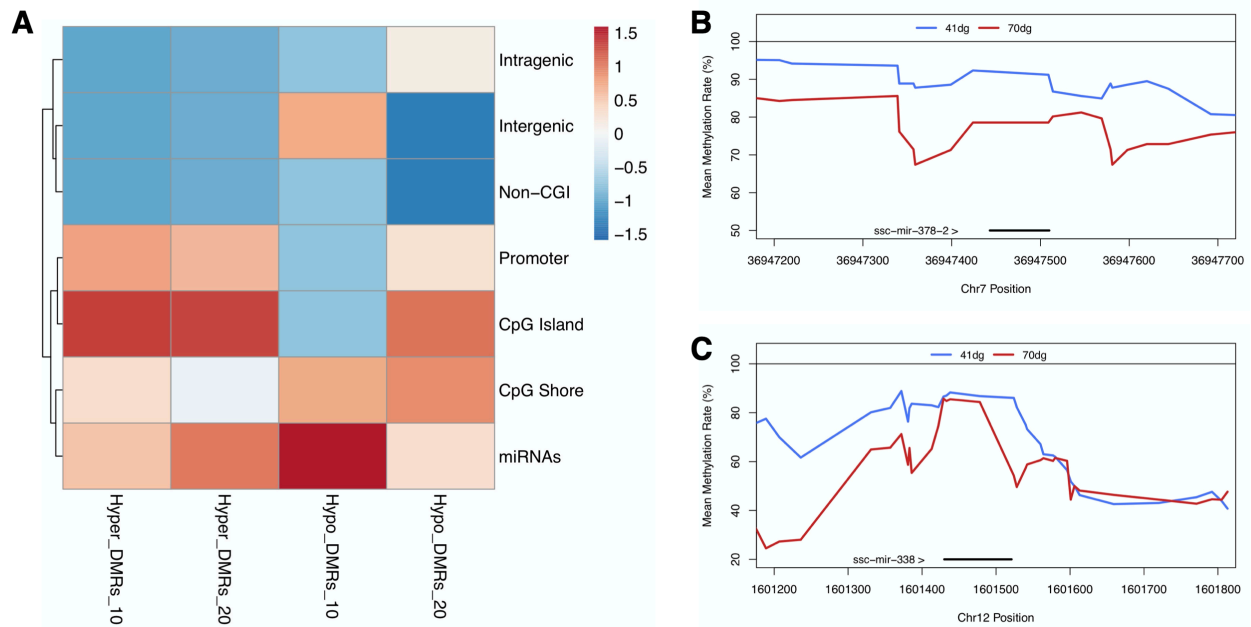


Figure 3.2. Fetal skeletal muscle differential methylation exhibits feature-specific enrichment. (A) Enrichment of hyper- and hypomethylated regions in gene features, CpG islands, and miRNAs. (B&C) Stage-specific methylation plots for regions overlapping ssc-mir-378-2 and ssc-mir-338, respectively. Black horizontal lines indicate miRNA gene coordinates.

Table 3.2. Enriched GO terms among promoter differentially methylated genes

<i>Promoter Hypermethylated Genes</i>			
GO Term	No. Genes	Enrichment	FDR
Muscle cell fate commitment	6	6.36	4.48E-02
Regulation of animal organ formation	11	6.28	8.50E-04
Muscle tissue morphogenesis	13	2.92	4.93E-02
Muscle organ development	40	2.12	2.44E-03
DNA-binding transcription factor activity	179	1.83	5.51E-11
<i>Promoter Hypomethylated Genes</i>			
GO Term	No. Genes	Enrichment	FDR
Regulation of release of sequestered calcium ion into cytosol	21	2.5	4.04E-02
Regulation of cation transmembrane transport	39	2.2	5.34E-03
Actin binding	78	1.63	2.09E-02
Sarcomere	40	1.83	3.40E-02
Actin cytoskeleton	84	1.55	1.49E-02

previously been shown to be downregulated during mammalian muscle development [137, 157] (Figure B.2E-F). These data demonstrate that DNA hyper- and hypomethylation are associated with embryonic and muscle-enriched miRNAs, respectively.

Annotation of DMRs resulted in identification of 3,635 promoter DMGs and 7,269 intragenic DMGs (Table 3.1), and we identified uniquely enriched gene ontology (GO) terms between hyper- and hypomethylated gene sets (Table 3.2). Promoter-hypermethylated genes were enriched for terms related to organ and muscle development, as well as for terms associated with transcriptional regulation. Conversely, promoter-hypomethylated genes were broadly enriched for numerous GO terms, reflecting the global hypomethylation at 70 dg. However, the most enriched GO terms were related to calcium transport as well as structural components of muscle (e.g., sarcomere and actin cytoskeleton). Intragenic DMGs followed a similar trend in which the most enriched terms among hypermethylated genes were associated with early developmental processes, and the most enriched among hypomethylated genes were related to calcium and ion transport. In summary, the enriched terms associated with DMGs are in agreement with expected downregulation (via hypermethylation) of genes involved in early muscle cell development, and activation (via hypomethylation) of genes involved in muscle hypertrophy, including those encoding components of muscle fibers and calcium transport pathways.

3.4.3 Differential methylation is strongly associated with differential transcript and miRNA expression

In order to assess the putative regulatory effects of identified DMRs, we performed both RNA-seq and smRNA-seq in the same fetal LD samples to characterize differential expression patterns and their association with differential methylation. We obtained 37-58M RNA-seq reads per sample and 17,414 genes exhibited non-zero gene counts (Table B.2). Among these, 1,546 genes were

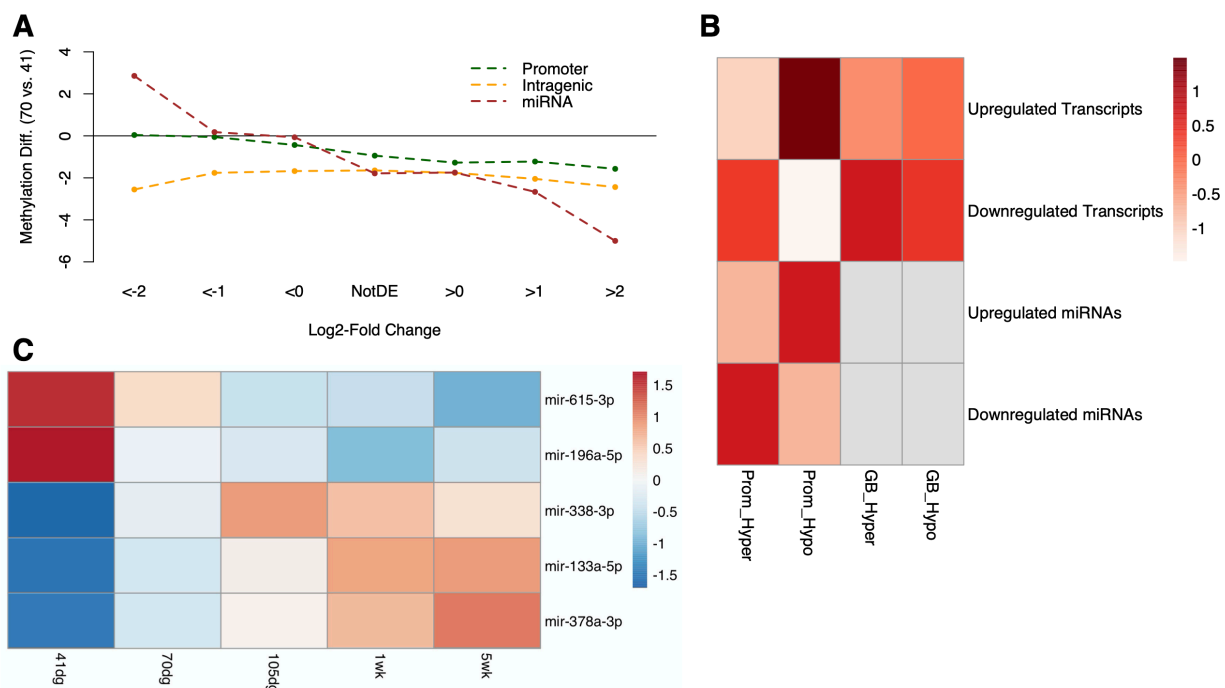


Figure 3.3. Fetal skeletal muscle differential methylation is associated with differential transcript and miRNA abundance. (A) Line plot of differential methylation by level of transcript abundance fold change between stages. NotDE = Not differentially expressed. (B) Enrichment heatmap for promoter and intragenic (GB) differentially methylated genes among differentially expressed transcripts and miRNAs. (C) Normalized transcript abundances for five differentially methylated and DE miRNAs across pre- and postnatal *longissimus dorsi* developmental stages.

differentially expressed between stages (867 upregulated and 679 downregulated at 70 vs. 41 dg). Upregulated genes were significantly enriched for muscle-specific GO terms, including muscle contraction and muscle filament sliding, and downregulated genes were significantly enriched for terms related to early development and synaptic signaling (data not shown). Additionally, we generated 36-69M smRNA-seq reads across samples (Table B.2), and among smRNA species the majority of reads mapped to miRNAs (data not shown). 356 miRNAs were expressed across all samples, and 35 and 42 exhibited significantly increased and decreased abundance at 70 vs. 41 dg, respectively.

To determine the degree to which differential methylation was associated with changes in transcript and miRNA abundance, we categorized genes based on their direction and magnitude of

differential expression between stages and calculated average methylation differences among groups. Increases in magnitude and direction of differential expression at 70 dg were associated with greater corresponding hypomethylation of promoters and miRNAs (Figure 3.3A). These findings are in agreement with reported inverse correlations between promoter methylation and gene expression [82], and suggest similar mechanisms are regulating miRNAs. Hypomethylation of intragenic regions was associated only with magnitude of differential expression, such that the most up- and downregulated genes exhibited the greatest degree of intragenic hypomethylation on average. Conflicting data has previously been reported on the relationship between intragenic methylation and gene expression, suggesting more context-specific associations than those observed at promoters [25].

We further characterized associations between differential methylation and expression by calculating the level of enrichment for DMGs among DEGs (Figure 3.3B). Among hyper- and hypomethylated regions located within 2kb of gene and miRNA TSSs, enrichment was observed for differential expression in the opposite direction, such that promoter-hypermethylated genes were enriched among downregulated transcripts and miRNAs, and promoter-hypomethylated genes were enriched among upregulated transcripts and miRNAs. Broader enrichment was observed for intragenic DMGs among DEGs, although, as in promoters, up- and downregulated genes were most enriched among hypo- and hypermethylated genes, respectively.

We quantified abundance of select differentially methylated and expressed miRNAs at additional stages of fetal and post-natal skeletal muscle development via RT-qPCR (Figure 3.3C). Abundance of hypomethylated and upregulated miRNAs miR-338, miR-133a, and miR-378a continued to increase throughout prenatal stages, and, in the case of the latter two, peaked in abundance at 5 weeks of age. miR-378a and miR-133a promote muscle differentiation through the

downregulation of *myogenic repressor* (myoR) and MAPK signaling, respectively, and have been reported as among the most abundant miRNAs in adult porcine skeletal muscle [136]. Expression of miR-338, once thought to be highly brain-specific, has recently been identified as expressed in porcine neonate skeletal muscle, as well as upregulated in bovine skeletal muscle exhibiting increased shear force [158, 159]; our data provide additional evidence that miR-338 regulation may be important in developing skeletal muscle, and that DNA methylation may in turn regulate its expression. Among the hypermethylated and downregulated miRNAs miR-196a and miR-615, expression continued to decrease or remain low with increasing age. These results indicate that differential methylation in fetal muscle precedes continued differential expression of overlapping miRNAs throughout skeletal muscle development.

3.4.4 Hypomethylated regions are enriched for myogenic regulatory factor binding sites

Differential methylation is known to exert regulatory effects in part through the alteration of TF binding sites [31, 82]. We therefore queried LD muscle developmental DMR sequences for enrichment of TF binding motifs, with particular interest in determining the degree to which MRF binding motifs were enriched in DMRs. Hypomethylated DMR sequences were enriched for MYOD1, MYF6, and MYOG motifs, while no such enrichment was observed among hypermethylated DMR sequences (Figure 3.4A). Enrichment of MRF motifs among hypomethylated regions correlates with expression patterns between stages, with *MYOD1*, *MYOG*, and *MYF6* exhibiting moderate to significant upregulation at 70 dg relative to 41 dg (Figure 3.4B). As MRFs are known to regulate one another's transcription [128], we assessed whether DMRs within MRF genes were associated with other MRF binding motifs. Indeed, we identified a MYOD1 motif within a hypomethylated region 5kb upstream of the *MYOG* TSS, and a MYOG motif within a hypomethylated region 4kb upstream of *MYF6*, which is in agreement with expected

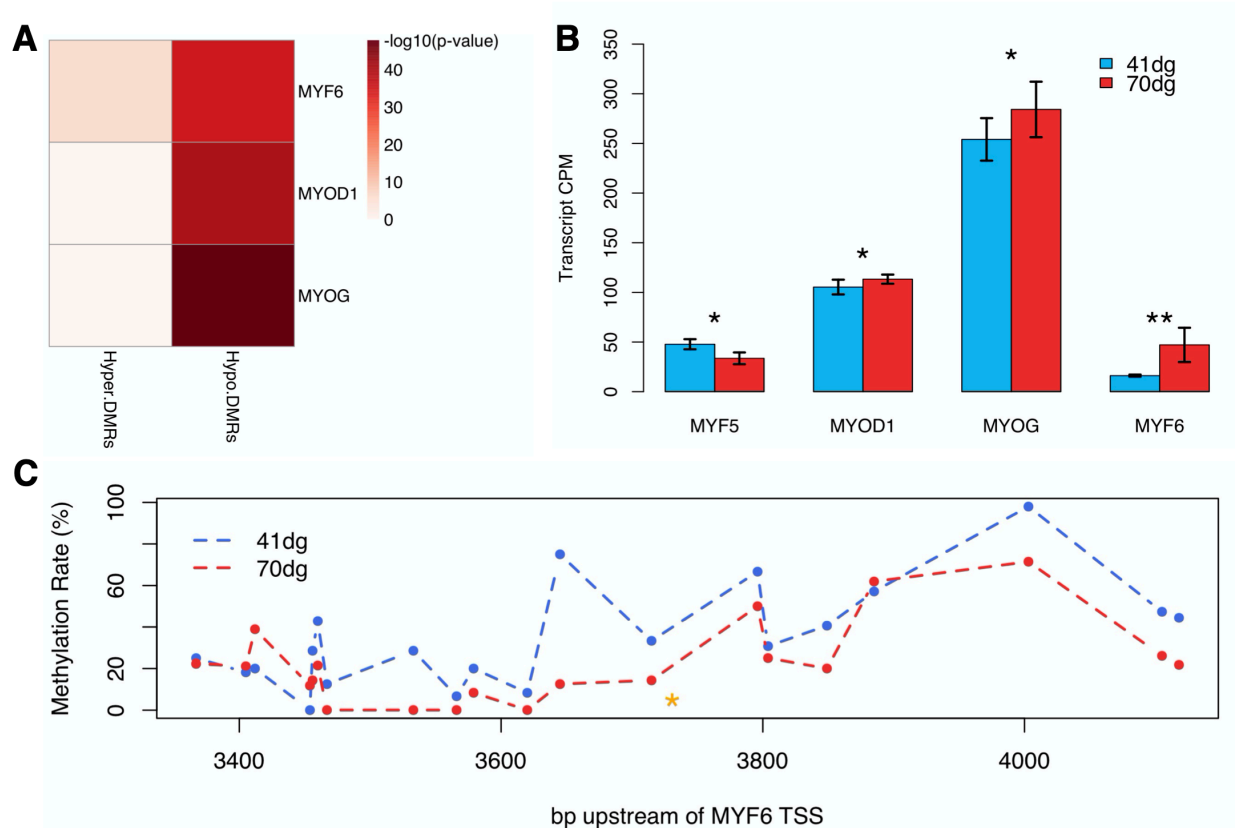


Figure 3.4. DNA hypomethylation is enriched for myogenic regulatory factor (MRF) binding motifs. (A) Heatmap of $-\log_{10}$ enrichment p-values of differentially methylated regions (DMRs) for MRF binding motifs. (B) MRF transcript abundances between developmental stages. (C) CpG methylation rates in a DMR upstream of *MYF6* overlapping a MYOG binding motif (orange asterisk). * $p<0.05$, **FDR<0.05.

MRF regulatory interactions. Additionally, a site-specific differential methylation analysis revealed that the MYOG motif upstream of *MYF6* co-localized with the most hypomethylated CpGs (Figure 3.4C), providing further evidence for an association between differential methylation and MYOG regulatory potential at this locus. Our data demonstrate that a large proportion of hypomethylated regions may serve regulatory functions via increased binding potential of MRFs involved in skeletal muscle differentiation.

3.4.5 Differentially methylated regions co-localize with GWAS SNPs for muscle and meat quality traits

Regions of differential methylation represents putative sites of gene regulation with potential consequences on phenotype expression. We therefore curated GWAS SNPs from the Pig QTL Database and searched for trait classes for which associated SNPs were enriched among fetal LD muscle DMRs. We identified significant enrichment among DMRs for GWAS SNPs associated with muscle, meat quality, and fatty acid traits (Figure 3.5A). The majority of trait SNPs were enriched among hypomethylated regions as opposed to hypermethylated regions, suggesting that regions that lose methylation—and can be hypothesized to be more active—with increasing age may be more influential on phenotypes related to skeletal muscle growth and physiology.

Of note, we identified a GWAS SNP for muscle pH at 24 hr post-mortem that overlapped a hypomethylated region in the first intron of *SCN4A* (Figure 3.5B). This gene encodes a subunit of a voltage-gated sodium channel responsible for generation of action potentials in skeletal muscle [160]. We observed significant upregulation of *SCN4A* at 70 dg ($\log_2\text{FC}=3.18$, $\text{FDR}=1.39\text{E-}43$), suggesting that alterations in DNA methylation at this region may be required for changes in gene expression. These results have elucidated developmental DMRs at which genetic variation has been associated with muscle phenotypes, and provide additional functional data that may aid in the search for causative variants within these regions.

3.5 Conclusions

We have identified thousands of DMRs that represent putative sites of gene regulation in pig fetal skeletal muscle. While observed global methylation patterns were strongly associated with differential expression of methylating and demethylating enzymes, we provide ample evidence that a large proportion of differential DNA methylation correlates with differential gene expression

and predicted *cis*-regulatory elements. Furthermore, dynamic methylation in intronic, muscle-enriched miRNAs suggests an important role for DNA methylation in regulating miRNA expression independently from their host genes.

We demonstrate that hypomethylated regions in late-myogenic fetal LD muscle are enriched for numerous muscle- and meat-related trait GWAS SNPs. Current functional annotation efforts in domesticated animal species seek to provide layers of regulatory information in genomic regions previously identified as associated with complex and disease traits [17]. LD muscle developmental DMRs identified in our study may represent stage-specific regulatory elements that contribute to prenatally-influenced phenotypes, or else signify permanent epigenomic changes in differentiated muscle tissue that contribute to gene and phenotype expression in postnatal and adult life. Knowledge of such stage-specific epigenomic patterns will enhance understanding of spatiotemporal specificity of functional elements in the porcine genome and improve the search for causative variants influencing traits of economic relevance.

3.6 Acknowledgements

Computing resources were provided by Michigan State University's High Performance Computing Cluster. Sequencing was performed at the Michigan State University Research Technology Support Facility. The authors wish to thank Mingang Lei and Valencia Rilington for their technical assistance in the laboratory and Kaitlyn Daza for her intellectual support.

CHAPTER 4

DNA METHYLATION PATTERNS IN PIG FETAL TISSUES ARE ASSOCIATED WITH UNIQUE DEVELOPMENTAL TRAJECTORIES AND ALLELE-BIASED GENE REGULATION

This chapter is currently being drafted for publication. It was prepared alongside co-authors Haibo Liu, Tim P.L Smith, Nancy E. Raney, Christopher K. Tuggle, Dan J. Nonneman, and Catherine W. Ernst

4.1 Abstract

Early-life gene regulation has diverse consequences on economically important phenotypes in the pig, yet annotation of regulatory elements in developing porcine tissues has thus far been limited. It is now recognized that gene regulatory regions bear characteristic epigenetic signatures, and DNA methylation has been extensively characterized in mammals as having context-specific effects on local and distal gene expression. Furthermore, allele-biased methylation (ABM) has been reported in a genotype-dependent manner associated with sequence variation between parental alleles, as well as a genotype-independent manner indicative of genomic imprinting; both classes of ABM have been shown to influence mammalian development. Research into developmental DNA methylation patterns in the pig has thus far been restricted to single-tissue analyses, and the extent of ABM in developing tissues is currently unknown. We have performed whole-genome bisulfite sequencing (WGBS) to assess DNA methylation patterns in four pig fetal tissues (brain, liver, loin muscle, and placenta) at 30 and 70 days gestation (dg) in females derived from Meishan (MS) and White Composite (WC) reciprocal crosses. We identified unique DNA methylation landscapes associated with tissue and developmental stage, with increasing fetal age resulting in an increase in the number of lowly methylated regions across tissues. 14.2% of the

genome was significantly differentially methylated between stages in at least one tissue. Differentially methylated regions (DMRs) disproportionately overlapped tissue- or stage-differentially expressed genes, with hypomethylated regions at 70 dg being more associated with upregulated and tissue-enriched genes at 70 dg in most tissues. Hypermethylated DMRs across tissues were enriched for similar transcription factor (TF) motifs, while hypomethylated regions exhibited tissue-specific motif enrichment of TFs associated with the respective tissue's differentiation. We also performed allele-specific mapping of WGBS reads and identified global ABM between MS and WC alleles that affected genes involved in processes such as dopamine transport and placenta labyrinthine layer development that may confer breed-specific behavioral and reproductive differences. *Methionine synthase reductase (MTRR)* was consistently promoter-hypomethylated in MS alleles, and this was associated with corresponding breed allele-biased expression (ABE) in an isoform-dependent manner. Lastly, we identified thousands of imprinting-like ABM regions that overlapped known imprinted gene clusters and regions homologous to human imprinting control regions. Among novel genes exhibiting ABM and ABE, protocadherin GA4 (*PCDHGA4*) exhibited the strongest imprinting-like ABM in the fetal brain and was associated with isoform-specific parental ABE. These results have provided unique insight into developmental and allele-specific DNA methylation patterns indicative of novel regulatory regions in porcine fetal tissues of economic relevance.

4.2 Introduction

Early life development has profound effects on future livestock performance, with numerous economically important phenotypes in the pig having been attributed to prenatal determinants. For example, the number of skeletal muscle fibers in most muscles is determined prenatally, and decreased pig placental weight has been associated with decreases in growth potential [45, 46].

Organ development requires appropriate spatiotemporal expression of thousands of genes involved in cellular commitment, differentiation, and maturation. Thus numerous transcriptome profiling studies have been performed in the pig across fetal ages, and have elucidated candidate genes regulating development of neuronal, skeletal muscle, and placental tissues, among others [129, 161–164]. Much less understood are the non-transcribed regulatory regions governing dynamic fetal gene expression in pigs and other livestock species. Epigenomic modifications result in heritable changes in gene expression without altering DNA sequence, and various subtypes of these modifications have been shown to be predictive of regulatory elements including enhancers, silencers, and insulators [11]. While epigenomic patterns are known to be highly variable throughout early life in coordination with changing transcriptional demands, their limited annotation in pig fetal tissues hinders the search for gene regulatory regions influencing gene and phenotype expression.

Among epigenetic mechanisms, DNA methylation is the most prevalent modification made to the DNA molecule and in mammals generally involves the addition of a methyl group to the 5-position of cytosines in CpG dinucleotides [27]. DNA methylation exhibits context-specific associations with gene expression: while DNA methylation levels at promoters and enhancers are generally negatively correlated with transcription levels through the inhibition of transcription factor (TF) binding or recruitment of repressors, the transcriptional consequences of methylation levels in other regions, including gene bodies, has been shown to be positive or negative in a gene-specific manner [25, 33, 34]. The mammalian DNA methylation landscape is highly dynamic during early development. Following genome-wide demethylation and subsequent remethylation to erase parentally-inherited epigenetic marks during pre-implantation, the DNA methylome continues to undergo rapid changes in coordination with cell-specific developmental trajectories

[43]. Assessment of DNA methylation patterns in human fetal tissues of diverse cell lineages has revealed putative regulatory elements associated with activation of early development and tissue-specific genes, and these were found to gain and lose methylation marks with increasing age, respectively [44]. Characterization of DNA methylation patterns has recently been performed in the pig fetus [141, 165]; however these studies have focused on single tissues, thereby limiting our understanding of tissue-specific methylation dynamics during prenatal development.

Allele biases in gene regulation and expression have been widely reported in mammalian species as having diverse developmental consequences. Genotype-dependent allele-biased methylation (ABM) arises from DNA sequence variation between parental alleles and has been extensively characterized from human bisulfite sequencing data [166–169]. Assessment of this phenomenon in the pig can provide unique insights into breed-specific gene regulation, particularly in crossbred individuals of highly divergent parentage. Chinese pig breeds have gained commercial interest due to their increased prolificacy and litter survival relative to European breeds, and have also been shown to exhibit greater docility and higher rates of intramuscular fat [170, 171]. Mammalian species also exhibit numerous instances of genotype-independent allelic biases in expression of autosomal genes. Genomic imprinting results in monoallelic or allele-biased expression (ABE) from either the maternal or paternal allele, and is known to effect dozens of genes involved in embryonic and fetal development [48]. In many cases, imprinting regulation has been shown to be governed by an imprinting control region (ICR) at which differential methylation between parental alleles is causative of observed ABE [172]. Knowledge of inheritance patterns of imprinted genes has proven useful in the livestock industry, as genotypes in imprinted loci have been shown to influence complex trait variation in a parent-of-origin-specific manner [51].

However validation of imprinted genes and their corresponding ICRs in the pig is currently limited in scope [173].

In the current study, we have performed whole-genome bisulfite sequencing (WGBS) to characterize DNA methylation patterns and their developmental variation in pig fetal brain, liver, loin muscle, and placenta tissue at two ages (30 and 70 days gestation (dg)). As fetuses were derived from reciprocal crosses of divergent pig breeds—Meishan (MS) and White Composite (WC)—we leveraged the genetic variation between alleles to identify allelic biases in methylation associated with breed-of-origin and with maternal or paternal alleles in an imprinting-like manner. We identified diverging patterns of DNA methylation with increasing fetal age across porcine tissues, and show that developmental differential DNA methylation is associated with differential and tissue-enriched gene expression. Furthermore, we report extensive ABM between breed alleles impacting biological pathways associated with phenotypic differences in parental pig breeds, and identify imprinting-like ABM regions overlapping known and putative-novel imprinted genes. These results have elucidated novel regulatory elements associated with prenatal porcine tissue development, as well as regions governing allele-specific gene regulation and expression.

4.3 Materials and Methods

4.3.1 Sample Collection

Fetal tissues were collected from litters of WC x MS reciprocal crosses at the USDA ARS Meat Animal Research Center (USDA MARC), and collection was done in compliance with the US MARC Animal Care Guidelines. Sampling was performed at 30 dg and 70 dg in one litter per cross type, with litters of the same cross type sampled at different stages having the same sire but different unrelated dams. Whole brain, liver, loin muscle, and placenta tissue were collected after

caesarean on slaughtered gilts immediately after electrocution, and tissue samples were immediately flash frozen in liquid nitrogen and stored at -80° C. Fetal sex was determined by genotyping an X-linked breed-specific SNP in *SERPINA7* (rs45431492) using an *HphI* PCR-RFLP assay, and heterozygous genotypes were classified as female [174]. Tissues from two females per litter were randomly selected for sequencing analyses.

4.3.2 DNA Isolation and Bisulfite Sequencing

Tissue DNA was isolated using either: 1) overnight digestion with digestion buffer, 20% SDS, RNase A and proteinase K followed by ethanol precipitation (30 dg liver and all 70 dg samples), or 2) the Qiagen AllPrep DNA/RNA MiniKit (remaining 30 dg samples). Bisulfite sequencing libraries were prepared and QC'ed as described in Section 2.3.3. Libraries were divided into four pools by tissue type, and each pool was sequenced on an Illumina HiSeq 4000 instrument as described in Section 2.3.3.

4.3.3 WGBS Bioinformatics

WGBS libraries were trimmed of technical sequences and low-quality bases using Trimmomatic v.0.39 as described in Section 2.3.4 [87]. To increase efficiency of read mapping and accuracy of methylation estimates, we created animal-specific genome fasta files using whole-genome sequencing (WGS) data from the same fetuses (Liu et al., personal communication). Briefly, DNA variant calling was performed on aligned WGS reads using GATK3 [175], and identified homozygous and heterozygous variants were replaced and masked, respectively, in the corresponding animal-specific fasta file. Trimmed WGBS reads were aligned to animal-specific genomes using Bismark as described in Section 2.3.4 [88]. Aligned WGBS reads were deduplicated and calculation of CpG methylation rates was performed as described in Section 2.3.4.

4.3.4 DNA Methylation Analyses

Genome regional methylation rates were calculated using the *methyKit* R package [89]. Briefly, CpG reports from all samples were merged, and genome tiling was performed to calculate average methylation rates for 1kb non-overlapping regions in the *Sus scrofa* genome. Within a stage, tissue-specific lowly methylated regions (LMRs) were classified as those regions with z-score <-1 and methylation rate $<75\%$ in only samples derived from a single tissue type. LMRs were annotated with respect to gene and CpG features using the *genomation* R package [90], and gene set enrichment analysis (GSEA) of tissue LMR genes was performed using the PANTHER database [91, 92].

Differential methylation analyses between fetal ages for each tissue were performed using the *methyKit* R package as described in Section 3.3.3, with stage and cross type included as fixed effects in logistic regression models. Genomic regions with a mean methylation difference $>10\%$ between stages and $FDR < 1E-5$ were considered differentially methylated regions (DMRs). DMR annotation and GSEA of differentially methylated genes (DMGs) was performed as described previously.

4.3.5 RNA-seq Bioinformatics and Gene Expression Analyses

RNA-sequencing (RNA-seq) was performed on the same fetal tissue samples on an Illumina NextSeq 500 in PE format at USDA MARC. Reads were trimmed and aligned, and gene counts were obtained as described in Section 3.3.2. Differential expression analyses were performed using the *DESeq2* R package [93]. For identification of tissue-specific gene expression, eight separate analyses were performed contrasting expression of one tissue versus expression of all other tissues within each stage. Analyses were performed as described in Section 3.3.4, with generalized linear models including the fixed effects of tissue, cross type and sequencing batch. Genes with \log_2 -

fold change >1 in a single tissue relative to other tissues and FDR <0.05 were considered tissue-enriched. Developmental differential expression analyses were performed for each tissue and included cross type and sequencing batch as a fixed effect. Genes with log₂-Fold Change >1 and FDR <0.05 between stages were considered differentially expressed. The degree of overlap between DMGs and both tissue-enriched genes and differentially expressed genes was assessed using hypergeometric tests (R software).

4.3.6 Motif Enrichment analyses

Developmental DMR sequences were submitted for analysis of motif enrichment using the MEME suite as described in Section 2.3.7 [94, 95]. Clustering of the top ten most enriched motifs for each DMR category (hyper- or -hypomethylated in each tissue) was performed using the *pheatmap* R package [99].

4.3.7 Allele-Specific Methylation Analyses

We performed allele-specific sorting of WGBS reads using SNPsplit v.0.3.4 [176]. Briefly, heterozygous SNPs in fetal genomes were N-masked prior to WGBS read alignment. Using WGS data from the parents of each reciprocal cross, genotype calls were made at identified fetal heterozygous SNPs using GATK3 [175] (Liu et al., personal communication). For each fetus, alleles were assigned to one of two parental genomes if 1) parents were homozygous for opposite alleles, or 2) One parent was homozygous for one allele and the other parent was heterozygous. Aligned WGBS reads overlapping assigned SNPs were sorted using SNPsplit into two separate allele-specific alignment files, after which methylation calling was performed using *bismark_methylation_extractor* with default parameters.

Allele-biased methylation (ABM) analyses were performed using the *methyKit* R package. Merging of methylation data and genome tiling were performed as described previously. To

identify ABM patterns associated with breed of origin, logistic regression models were fitted separately for each tissue and stage, and corrected for the effect of parental sex of derived alleles, and tested for the effect of breed. For identification of ABM patterns associated with paternal vs. maternal alleles, tissue- and stage-specific models were fitted, corrected for the effect of breed of origin, and tested for the effect of parental sex. Regions exhibiting a methylation difference >10% and FDR <0.01 were classified as exhibiting significant ABM. Allele-biased expression (ABE) analyses have previously been performed in the same tissue samples (Liu et al., personal communication), and were utilized here to identify genes exhibiting both ABE and ABM.

4.4 Results

4.4.1 Global DNA methylation patterns are associated with tissue type and developmental stage

We obtained 162-254M PE WGBS reads per fetal tissue sample, of which an average of 85.9% uniquely mapped to their respective animal-specific genome assemblies (Table C.1). Library bisulfite conversion rates were $\geq 99\%$ for all samples, meeting the standards for WGBS data sets proposed by ENCODE (<https://www.encodeproject.org/data-standards/wgbs/>). Global CpG methylation rates varied significantly between tissue types ($F = 789.3$, $p < 2.2e-16$), with brain and muscle samples across both fetal ages being hypermethylated relative to liver and placenta (Figure 1A). Furthermore, stage-specific differences in global CpG methylation were also observed: 70 dg brain, muscle and placenta were hypomethylated relative to 30 dg samples, while 70 dg liver was significantly hypermethylated relative to 30 dg. To determine if the observed variance in global methylation was associated with differences in DNA methylation or demethylation enzyme expression, we calculated correlation coefficients between sample methylation and DNA methyltransferase (*DNMT*) and Tet methylcytosine dioxygenase (*TET*) transcript abundance from

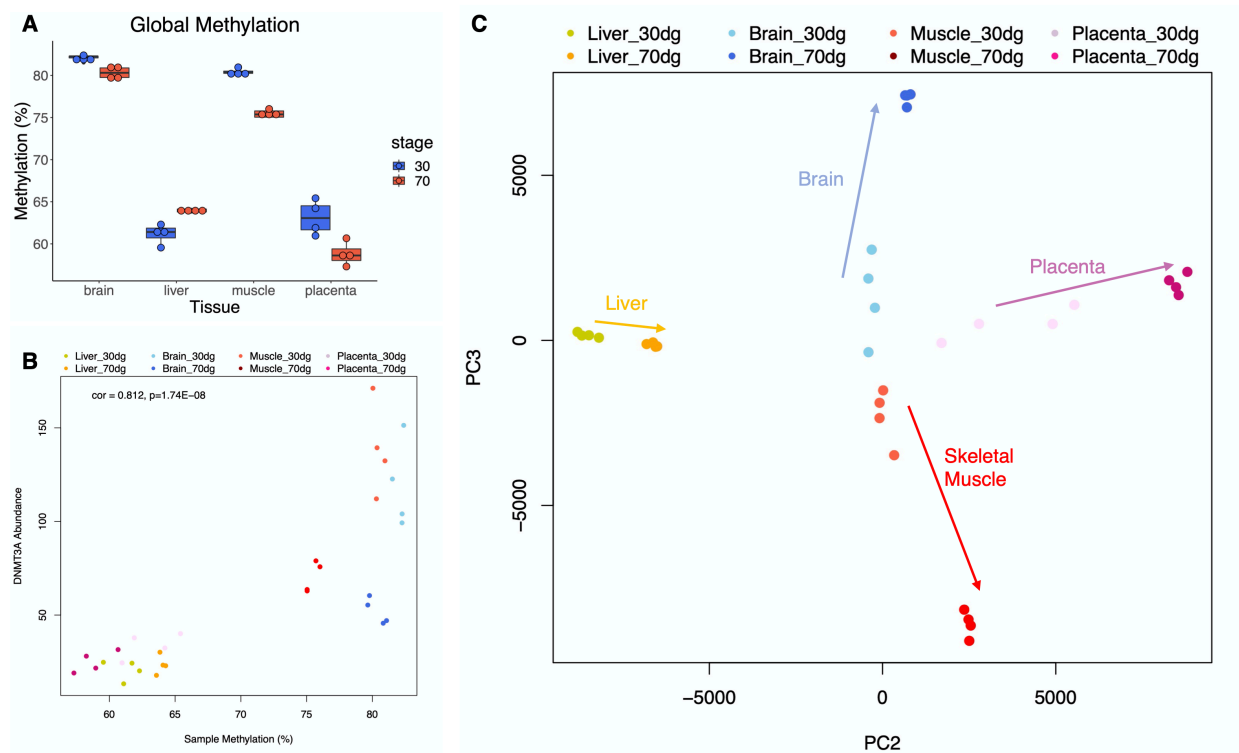


Figure 4.1. Landscape of porcine fetal DNA methylation. (A) Boxplot of global CpG methylation rates by tissue and stage. (B) Scatter plot of sample *DNMT3A* abundance against global CpG methylation rate. (C) Principal component analysis (PCA) plot of PC2 (6.2%) and PC3 (4.1%). Arrows denote tissue-specific direction of 30 dg to 70 dg samples.

Table 4.1. Number of tissue-lowly methylated regions (LMRs) and enriched genes

Tissue	No. LMRs, 30 dg	No. LMRs, 70 dg	No. Enriched Genes, 30 dg	No. Enriched Genes, 70 dg
Brain	2554	24880	1495	3926
Liver	316209	121328	3681	2566
Muscle	1857	17311	443	1344
Placenta	147135	504276	2261	2451

RNA-seq data. Global methylation level was most significantly positively correlated with *DNMT3A* abundance (Figure 1B), while all other correlations were not significant.

A principal component analysis of fetal tissue DNA methylation revealed that the greatest proportion of variance (PC1=63.2%) could be attributed to global methylation rate (Figure C.1). The second and third principal components clearly separated samples according to tissue and fetal

age (Figure 1C). Among brain, muscle, and placenta samples, 30 dg samples exhibited weak clustering by tissue type, while the 70 dg samples exhibited better clustering and were more separated by tissue. These results indicate that global DNA methylation patterns are associated with differentiation of porcine tissues during fetal development.

To determine how DNA methylation may contribute to state-specific gene regulation during fetal development, we identified tissue-specific lowly methylation regions (LMRs) at each stage (Table 4.1). There was an overall greater number of LMRs at 70 dg versus 30 dg, consistent with the notion that tissue differentiation is associated with unique regions of demethylation [44]. In agreement with observed global methylation differences, brain, muscle, and placenta possessed more LMRs at 70 dg relative to 30 dg, while the liver possessed fewer. Lowly methylated genes (LMGs) were broadly associated with tissue-enriched GO terms, and, in many cases, enrichment was greater in 70 dg LMGs for each tissue type while, in muscle, enrichment for muscle-specific terms was only observed at 70 dg (Figure 4.2A). Additionally, emergence of tissue-specific low methylation was observed at 70 dg for tissue-enriched genes: *NEUROD6* (brain), *APOH* (liver), *MYH7* (muscle) and *CGA* (placenta) (Figure 4.2B-E). This was achieved either through tissue-specific DNA hypomethylation at 70 dg (*NEUROD6*, *APOH*, *MYH7*) or via hypermethylation of non-expressing tissues (*CGA*). These data demonstrate that tissue-specific hypomethylation during fetal development is highly associated with markers of tissue differentiation and enriched biological processes.

We identified thousands of genes exhibiting tissue-specific expression at both stages (Table 4.1) and determined the degree to which enriched gene expression was associated with low methylation. While enrichment between expression-enriched genes and LMGs was non-specific

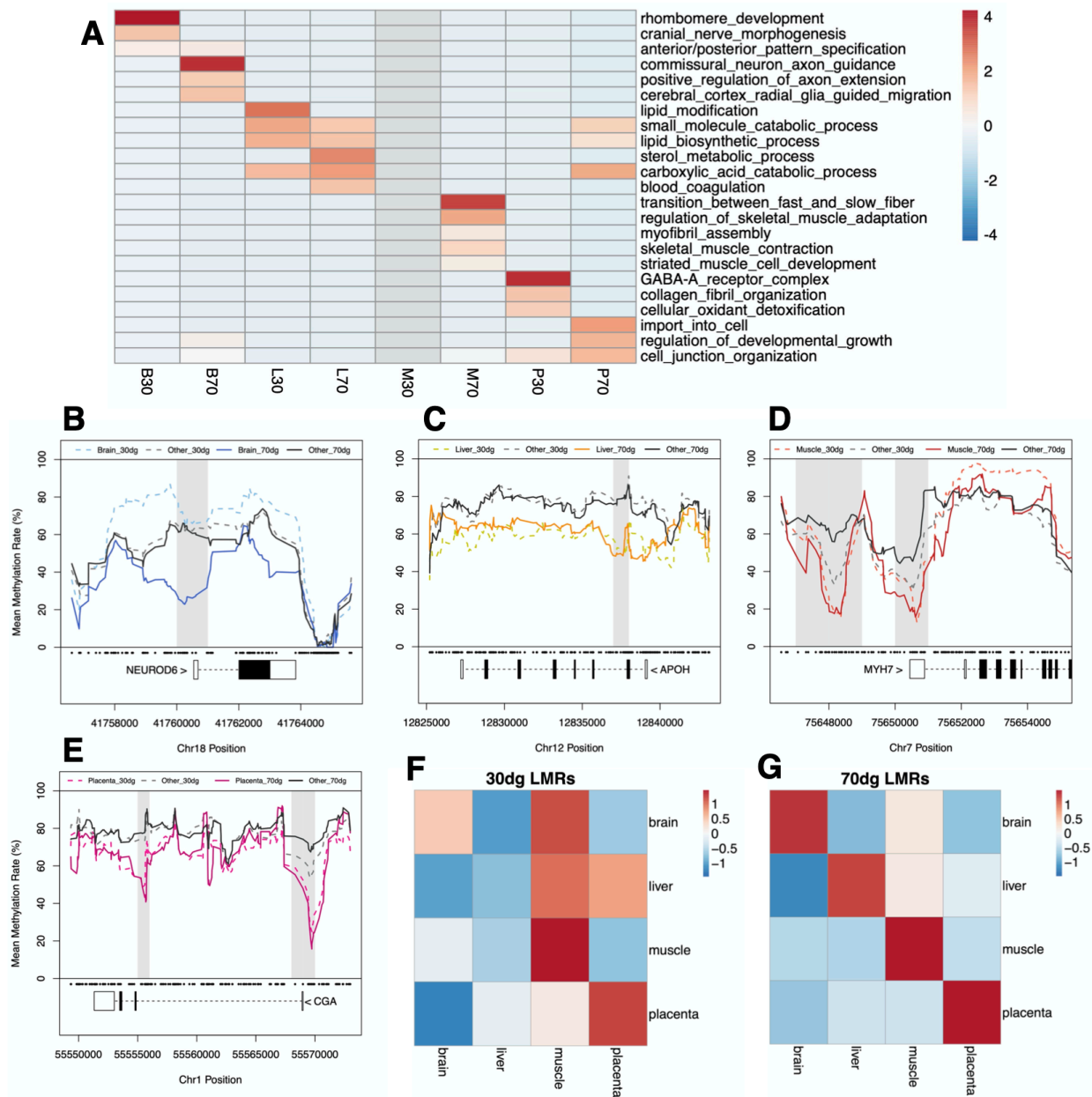


Figure 4.2. Tissue-specific lowly methylated regions in the developing pig fetus. (A) Heatmap of enriched gene ontology (GO) terms among tissue lowly methylated genes (LMGs). (B-E) Methylation levels of *NEUROD6*, *APOH*, *MYH7*, and *CGA* in fetal brain, liver, skeletal muscle, and placenta, respectively, plotted alongside average methylation rates in other tissues. (F-G) Normalized heatmaps of enrichment scores between tissue LMGs and expression-enriched genes at 30 and 70 dg.

at 30 dg, we observed exclusive enrichment for expression-enriched genes among LMGs of the same tissue type at 70 dg (Figure 4.2F-G). These results indicate that tissue-specific DNA methylation is more strongly associated with differential gene expression with increasing fetal age, and that other factors besides transcriptional demands may influence early fetal DNA methylation patterns.

4.4.2 Developmental differential DNA methylation is associated with tissue- and stage-specific differential gene expression

To characterize epigenomic variation associated with porcine tissue development, we performed differential methylation analyses between fetal ages. We identified a total of 338,523 DMRs across all tissues, comprising 14.2% of the *S. scrofa* genome (Figure 4.3A). Consistent with changes in global methylation between stages, the liver exhibited a greater proportion of hyper-DMRs while the remaining tissues had a greater proportion of hypo-DMRs at 70 dg. Across all tissues, DMRs were significantly enriched in CpG islands (CGIs) and shores in the porcine genome (Figure C.2). GO enrichment analysis of genes overlapping CGIs revealed an enrichment for early developmental processes (data not shown), suggesting that dynamic CGI methylation, while generally unobserved in mammalian systems, may be necessary during prenatal organ development and differentiation. Differentially methylated genes (DMGs) in each tissue were enriched for unique biological processes that are generally in agreement with expected gene activation and suppression throughout development (Table 4.2). Within promoters, hypomethylated genes were associated with differentiating tissue-specific processes (e.g., ‘Neuron recognition’ in brain and ‘myofibril assembly’ in muscle), while hypermethylated genes were associated with general and tissue-specific early developmental processes (e.g. ‘hemopoiesis’ in liver and ‘muscle tissue morphogenesis’ in muscle).

Table 4.2. Enriched GO terms among promoter differentially methylated genes (DMGs)

DMG Class	Go Term	No. Genes	Enrichment	FDR
Brain Hyper- DMGs	embryonic neurocranium	4	12.01	3.33E-02
	morphogenesis			
	dentate gyrus development	5	9.01	2.23E-02
	forebrain regionalization	6	7.05	2.02E-02
Brain Hypo- DMGs	Neuron recognition	19	2.87	2.51E-02
	Oligodendrocyte differentiation	23	2.42	3.17E-02
	Regulation of neurotransmitter	30	2.07	4.19E-02
	transport			
Liver Hyper- DMGs	anterior/posterior pattern specification	34	3.45	2.61E-06
	embryonic organ development	51	2.41	2.94E-05
	hemopoiesis	52	1.83	8.90E-03
Liver Hypo- DMGs	regulation of lipid metabolic process	20	3.03	1.70E-02
	protein modification process	84	1.61	1.12E-02
	metabolic process	198	1.37	1.37E-04
Muscle Hyper- DMGs	proximal/distal pattern formation	15	7.45	2.51E-06
	neuromuscular junction development	9	3.78	3.05E-02
	muscle tissue morphogenesis	13	3.18	1.48E-02
Muscle Hypo- DMGs	myofibril assembly	24	2.4	3.98E-02
	muscle contraction	85	2.01	4.38E-05
	striated muscle cell differentiation	64	1.91	2.06E-03
Placenta Hyper- DMGs	Trophoblast giant cell differentiation	4	15.96	1.82E-02
	dorsal/ventral pattern formation	11	7.51	9.65E-05
	anterior/posterior pattern specification	29	7.41	7.83E-13
Placenta Hypo- DMGs	regulation of plasma membrane-	49	1.87	1.75E-02
	bounded cell projection assembly			
	myeloid leukocyte activation	125	1.51	5.36E-03
	organic substance transport	417	1.43	7.96E-09

To determine the degree to which developmental differential DNA methylation is associated with tissue- and developmental differential gene expression, we performed hypergeometric tests between tissue DMGs and 1) tissue-enriched genes and 2) Stage DEGs for each tissue. Overall differential gene methylation was strongly associated with tissue-enriched genes at 70 dg: brain-, muscle- and placenta-enriched genes were disproportionately hypomethylated across genomic features, although hypermethylation in a subset of intragenic and CpG-sparse regions of the genome was also observed in brain- and muscle-enriched genes (Figure 4.3B). Similar DMG enrichment was not observed among 30 dg tissue-enriched genes, indicating

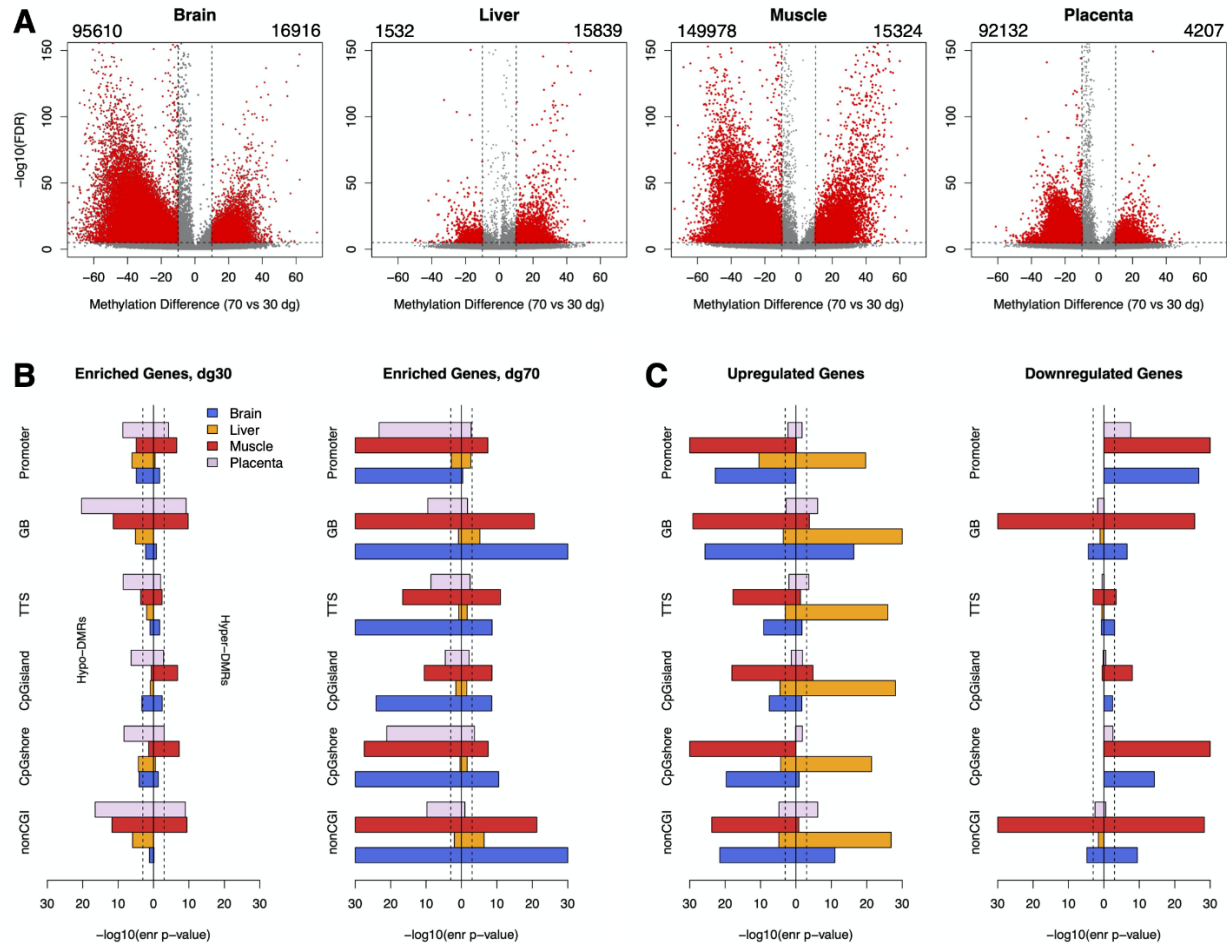


Figure 4.3. Developmental differential DNA methylation in porcine fetal tissues. (A) Volcano plots of regional differential methylation between fetal ages in each tissue. Left and right numbers on each plot denote number of hypo- and hypermethylated regions, respectively. (B) Bar plots of enrichment p-values between tissue-enriched genes and developmental DMRs within gene and CpG features. (C) Bar plots of enrichment p-values between differentially expressed genes and developmental DMRs within gene and CpG features. Vertical dashed lines indicate thresholds for significance.

many genes gaining tissue-specificity in expression throughout development also undergo dynamic methylation—and in particular hypomethylation—between developmental ages.

We identified 1,471-4,558 DEGs between fetal ages for each tissue type, and these were significantly enriched for tissue DMGs (Figure 4.3C). Brain and muscle generally exhibited canonical associations between differential methylation and expression: upregulated genes were enriched for hypomethylated genes across gene features, while downregulated genes were

enriched for gene hypermethylation, particularly in promoters. Conversely, upregulated genes at 70 dg in liver were broadly enriched for hypermethylated genes, and liver DMRs were not associated with gene downregulation. Given the unique associations observed between methylation and gene expression dynamics in the liver, we explored other potential factors influencing methylation patterns. Dividing cells are inherently hypomethylated due to the formation of hemi-methylated DNA following replication. We found that cell division and DNA replication processes were uniquely enriched among liver-specific genes at both stages, and furthermore that downregulated genes in 70 dg liver were enriched for similar processes (data not shown). Most notably, 70 dg liver was enriched for expression of *DNMT1*, the methyltransferase responsible for methylating newly-synthesized DNA strands [28]. These data provide evidence that global hypermethylation of fetal liver from 30 dg to 70 dg may be driven primarily by decreases in DNA replication and cell division, and may explain the lack of association between differential methylation and expression that is observed in other tissues.

4.4.3 Differentially methylated regions co-localize with tissue-specific transcription factor binding motifs

To assess whether regions of developmental differential methylation co-localized with TF binding motifs, we performed analysis of motif enrichment among DMRs. Clustering DMRs by motif enrichment scores grouped all hypermethylated regions together, suggesting that similar motifs gain methylation during fetal development regardless of tissue (Figure 4.4A). Conversely hypomethylated regions in fetal brain, muscle, and placenta exhibited distinct motif enrichment profiles, demonstrating loss of methylation of TF motifs in a tissue-dependent manner. These clustering patterns were preserved when considering only the most significantly enriched motifs in each set of DMRs, and revealed broad and tissue-specific enrichment for TF families (Figure 4.4B). Hypermethylated regions across tissue types were enriched for binding motifs of

transcription factor AP-2 (TFAP2) and SP TF family members that are known to play important roles in early animal development and cellular differentiation [177, 178]. Hypomethylated regions exhibited unique motif enrichment for TFs known to play important roles in tissue-specific development: nuclear factor I X (NFIX) and notochord homeobox (NOTO) in brain, myogenic differentiation 1 (MYOD1) and myogenin (MYOG) in skeletal muscle, and aryl hydrocarbon receptor (AHR), aryl hydrocarbon receptor nuclear translocator (ARNT), and hypoxia-inducible factor 1A (HIF1A) in placenta [179].

To determine if TFs of DM motifs also exhibited coordinated changes in expression, we assessed patterns of differential transcript abundance of TFs in Figure 4.4B. Among hypermethylated motifs across 70 dg fetal tissues, we observed corresponding trends towards downregulation for most TF transcripts (Figure 4.4C). Conversely TFs with hypomethylated motifs in specific tissues exhibited increases in transcript abundance in that same tissue, including *NFIX* and *NOTO* (brain) and *MYOD1* and *MYOG* (skeletal muscle) (Figure 4.4D). These data provide additional molecular evidence that developmental differential methylation is associated with binding sites for tissue- and stage-specific transcription factors.

4.4.4 Genotype-dependent allele-biased methylation is widespread in pig fetal tissues

Twenty-five to 33% of WGBS reads across sample libraries overlapped heterozygous variants and could be confidently assigned to one of two parental alleles (Table C.2). Consistent with variation observed in the diploid genome, we observed global differences in DNA methylation associated with fetal age (Figure C.3). However, across all tissues significant global differences were also observed between breed alleles, with those derived from the MS parents being significantly hypomethylated relative to the WC allele from the same individual (data not shown).

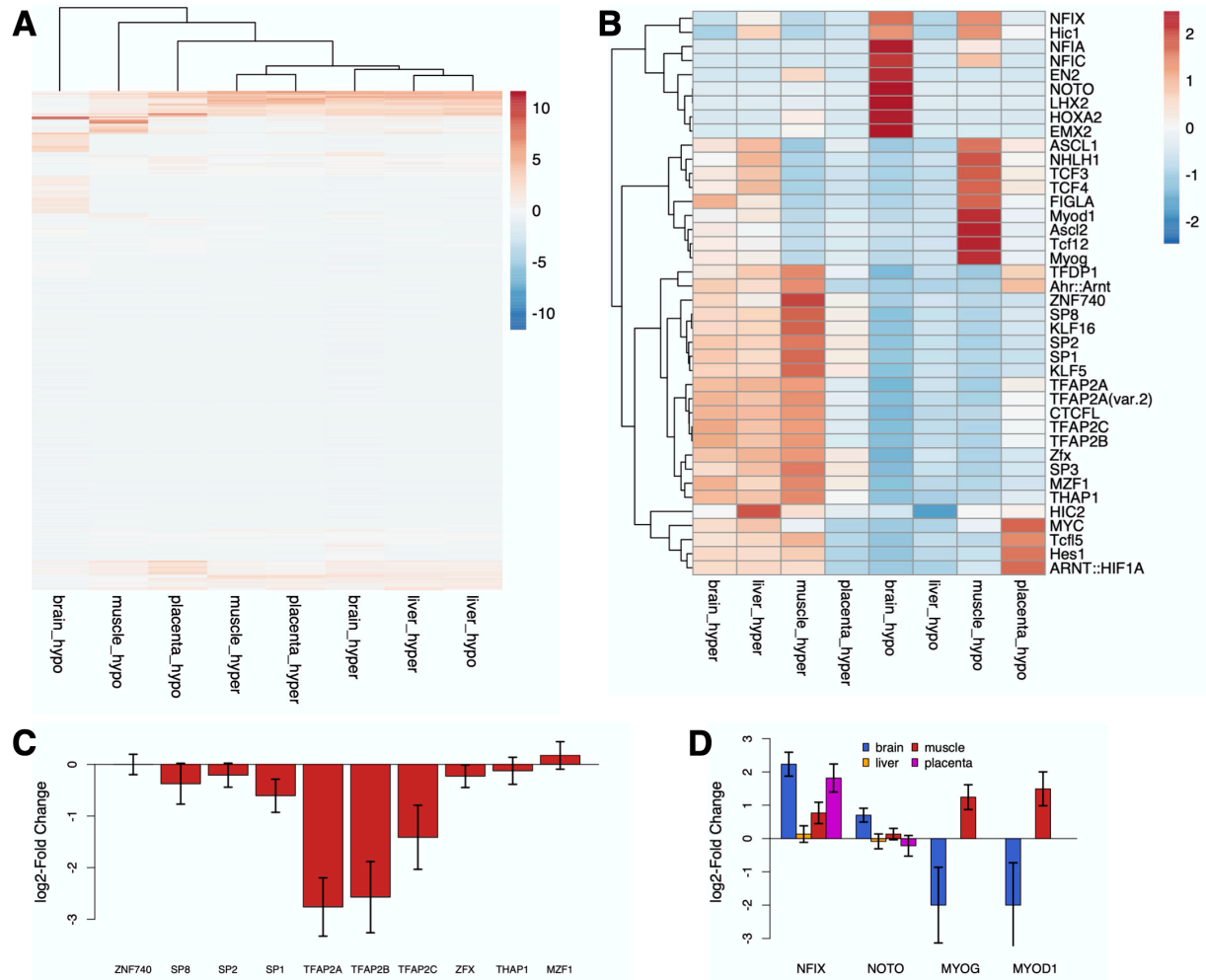


Figure 4.4. Pig fetal tissue differentially methylated regions (DMRs) are associated with unique transcription factor (TF) binding motifs. (A) Normalized clustered heatmap of DMR enrichment p-values for TF motifs in the JASPAR human motif database. **(B)** Normalized heatmap of the most enriched TF motifs among fetal DMRs. **(C)** Tissue-averaged log₂-fold change in abundance of TFs with commonly-enriched motifs across hypermethylated regions. **(D)** log₂-fold change in abundance of TFs with enriched motifs among hypomethylated-DMRs in fetal brain (*NFIX*, *NOTO*) and skeletal muscle (*MYOG*, *MYOD1*).

We identified 116,467 regions exhibiting breed ABM in at least one comparison (Figure 4.5A). In all tissues a greater number of breed-ABM regions were identified at 70 dg relative to 30 dg, suggesting that breed-specific gene regulation may be more prevalent as development progresses. Gene-set enrichment analysis of breed-ABM genes revealed uniquely enriched GO terms related to organ morphology and physiology (Table 4.3). 30 dg brain breed-ABM genes

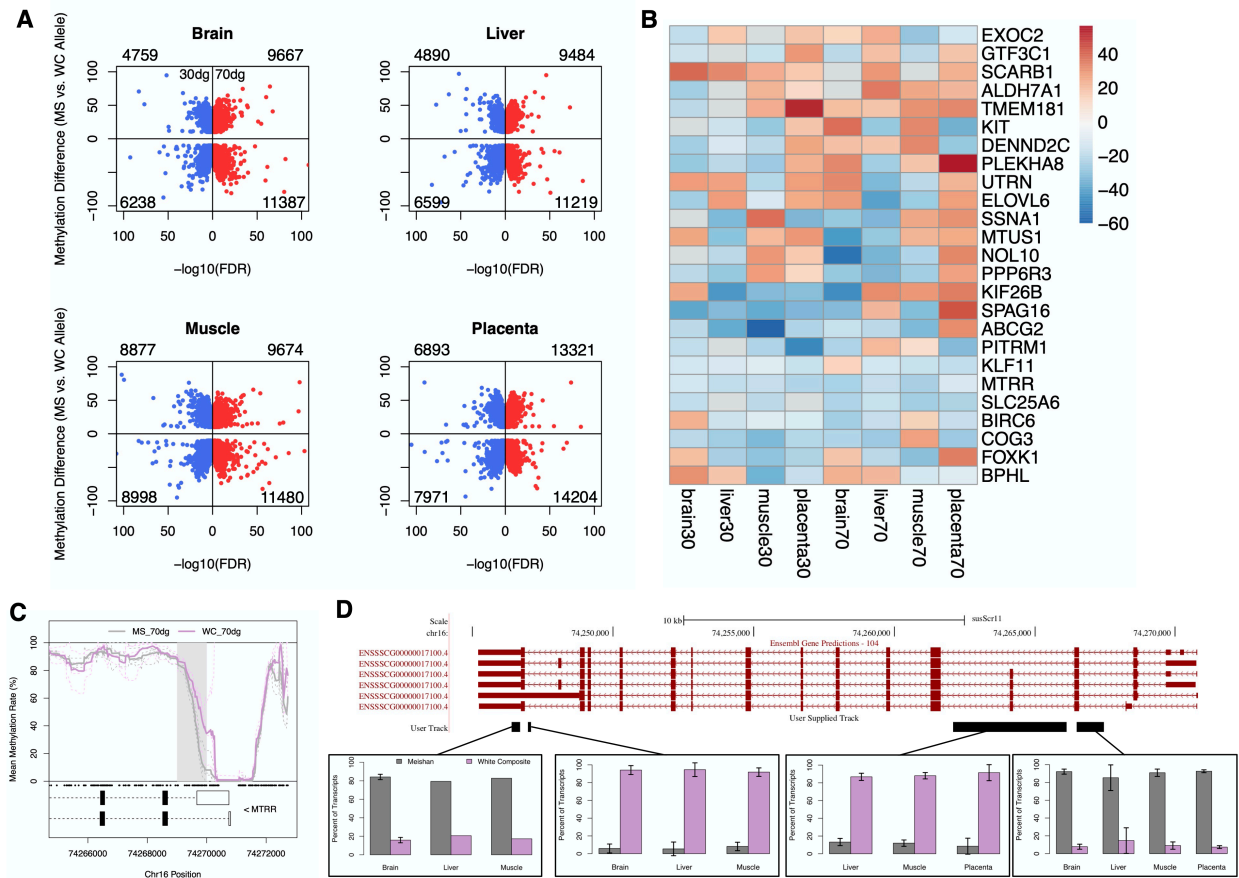


Figure 4.5. Breed allele-biased methylation (ABM) is widespread in pig fetal tissues. (A) Volcano plots of breed allele differential methylation within each tissue at 30 dg (blue) and 70 dg (red). Upper and lower numbers in each plot represent number of WC and MS-hypomethylated regions, respectively. **(B)** Heatmap of methylation differences (Meishan vs. White Composite) in genes exhibiting allele-biased methylation and expression in multiple tissue. **(C)** *MTRR* promoter methylation plot in 70 dg skeletal muscle. Gray box denotes region of significant ABM. **(D)** Gene track of *MTRR* transcript isoforms. Black bars indicate regions of detected allele-biased expression, with corresponding bar plots showing percentage of transcripts derived from each breed allele in pig fetal tissues. Error bars indicate standard deviation.

were significantly enriched for terms associated with dopamine transport. This group of genes included multiple solute carrier family members: the dopamine transporter *SLC6A3* as well as *SLC22A1*, *SLC22A2*, and *SLC22A3*, all of which are responsible for translocation of dopamine and other neurotransmitters [180]. The majority of dopamine-associated genes were MS-hypomethylated, suggesting common epigenetic regulation of this process. In muscle, breed-ABM genes were most enriched for the term ‘negative regulation of muscle hypertrophy’ and contained

an equal number of MS- and WC-hypomethylated genes. Placenta breed-ABM genes were enriched for the term ‘labyrinthine layer morphogenesis’, with related genes being primarily MS hypomethylated and, in the case of *FGFR2*, also exhibiting a significant MS-bias in expression. ABM genes in the placenta were also enriched for the term ‘Tolerance induction to self-antigen’, with all related genes being hypomethylated in WC alleles. These results indicate potential implications of breed- and tissue-specific gene regulation on processes influencing behavior, growth, and *in utero* development.

Table 4.3. Top enriched GO terms among tissue breed-ABM genes

Tissue-Stage	GO Term	Enrichment	FDR	Genes
Brain-30 dg	Dopamine Uptake	4.61	3.43E-02	<u>DBH</u> , <u>NET1</u> , <u>PRKN</u> , <u>SNCA</u> , <u>SLC22A2</u> , <u>SLC22A1</u> , <u>SLC22A3</u> , <u>SLC6A3</u>
Muscle-70 dg	Negative regulation of muscle hypertrophy	2.37	4.20E-02	<u>ATP2B2</u> , <u>FOXO1</u> , <u>GATA5</u> , <u>GLRX3</u> , <u>G6PD</u> , <u>JARID2</u> , <u>NOTCH1</u> , <u>LMNA</u> , <u>MLIP</u> , <u>P2RX4</u> , <u>PAK1</u> , <u>PPARA</u> , <u>SMAD3</u> , <u>SMAD4</u> , <u>TRIM63</u> , <u>YY1</u>
Placenta-30 dg	Labyrinthine layer Development	2.66	4.17E-02	<u>BIRC6</u> , <u>BMP7</u> , <u>CITED1</u> , <u>DNAJB2</u> , <u>FGFR2*</u> , <u>GRB2</u> , <u>GRHL2</u> , <u>IL10</u> , <u>MAP2K1</u> , <u>NCOA1</u> , <u>ST14</u> , <u>VASH1</u> , <u>WNT2</u> , <u>WNT7B</u>
Placenta-30 dg	Tolerance induction to self-antigen	5.31	3.82E-02	<u>AIRE</u> , <u>BLK</u> , <u>FOXP3</u> , <u>LYN</u> , <u>XKR8</u>

Bold: hypomethylated in Meishan (MS); Underline: hypomethylated in White Composite (WC); **Bold & Underline**: regions of MS and WC hypomethylation

*Gene exhibits MS ABE

Allele-biased expression (ABE) within the same pig fetal tissues has previously been reported (Liu et al., personal communication), and we sought to identify genes exhibiting coordinated allele biases in expression and methylation. A total of 2,348 genes exhibited ABE in a breed-specific manner, and these were significantly enriched among breed-ABM genes ($p=7.2E-246$). Interestingly, we did not observe tissue- or stage-specific enrichment between ABM and ABE genes, due in part to the high degree of overlap between ABM and ABE genes across multiple tissues and stages (data not shown). We therefore identified the most common ABM and ABE genes observed across tissues and fetal ages. Twenty-five such genes were common to at least six tissue-stage combinations (Figure 4.5B). Among these were genes previously shown to exhibit differential genetic selection in European vs. Chinese pig breeds, including two genes known to influence coat color (*EXOC2* and *KIT*) as well as *BPHL*, *FOXK1*, and *UTRN* [181–183].

In all tissues and stages, methionine synthase reductase (*MTRR*) exhibited MS hypomethylation in a region overlapping the promoter and first exon (Figure 4.5C). This gene encodes an enzyme responsible for reactivation of methionine synthase, and *MTRR* deficiency has been associated with multiple morphological and physiological phenotypes in the mammalian brain, liver, and placenta. Due to its reported associations with organ systems relevant to this study, we further assessed how breed-ABM in *MTRR* correlated with expression patterns. We identified *MTRR* transcript regions with conflicting breed-ABE (Figure 4.5D). In two regions overlapping exons common to all annotated *MTRR* isoforms, MS alleles represented a significantly higher proportion of transcripts in all tissues. However, in regions overlapping more isoform-specific exons, WC alleles were the predominant transcript detected. These findings suggest that MS hypomethylation in the *MTRR* promoter is associated with overall MS bias in *MTRR* expression, but that WC alleles may express specific isoforms at higher frequencies.

4.4.5 Genotype-independent allele biases in pig fetal tissues reveal putative novel imprinted loci

We assessed patterns of ABM between paternal and maternal alleles as putative regions of imprinting regulation. We identified 40,030 regions exhibiting imprinting-like ABM across tissues and stages (diff >20, FDR <0.01) and, as with breed-ABM, a greater number of imprinting-like ABM regions were observed at 70 dg in all tissues. To validate our results, we queried the regions with the greatest methylation biases between parental alleles, and these primarily overlapped known imprinted genes (Figure 4.6A). Imprinting-like ABM ‘hotspots’ were also identified containing multiple ABM regions, and these were within known imprinting clusters such as *DLK1*, *GNAS*, *KCNQ1*, and *Pws*. When testing for overall enrichment of imprinting-like ABM genes among known human imprinted genes, the degree of overlap was not significant ($p=0.32$). However, when considering only the genes exhibiting ABM >50% between parental alleles, the observed enrichment was statistically significant ($p=2.47E-03$), indicating that greater ABM biases are associated with previously-validated human imprinted genes.

Differential parental methylation of ICRs is known to act as a primary regulator of imprinting monoallelic expression. We therefore sought to determine whether imprinting-like ABM regions overlapped putative ICRs that have been extensively studied in other mammalian species. The human and mouse *H19/insulin-like growth factor 2 (H19/IGF2)* ICR contains multiple CTCF binding sites, and paternal hypermethylation of this locus results in CTCF inhibition and a nearby enhancer activating *IGF2* expression solely from the paternal allele. We identified a paternally-hypermethylated region in multiple tissues located in an intergenic region between *IGF2* and *H19*, and this region shared significant sequence identity with a segment of the human *H19/IGF2* ICR (Figure 4.6B). The putative CTCF binding site within this locus also overlapped the most significantly DM CpGs within the ABM region, providing additional

evidence that observed ABM is associated with differential insulator potential and putative ICR activity. We also queried for putative ICRs in the paternally-imprinted *IGF2 receptor* (*IGF2R*); in humans and mice, an intragenic region of *IGF2R* regulates expression of an *IGF2R* antisense RNA (*AIRN*), and maternal *IGF2R* expression has been shown to arise from DNA hypermethylation of this locus that suppresses *AIRN* expression, while this locus is hypomethylated on the paternal allele. We identified a maternally-hypermethylated region within *IGF2R* in 70 dg skeletal muscle,

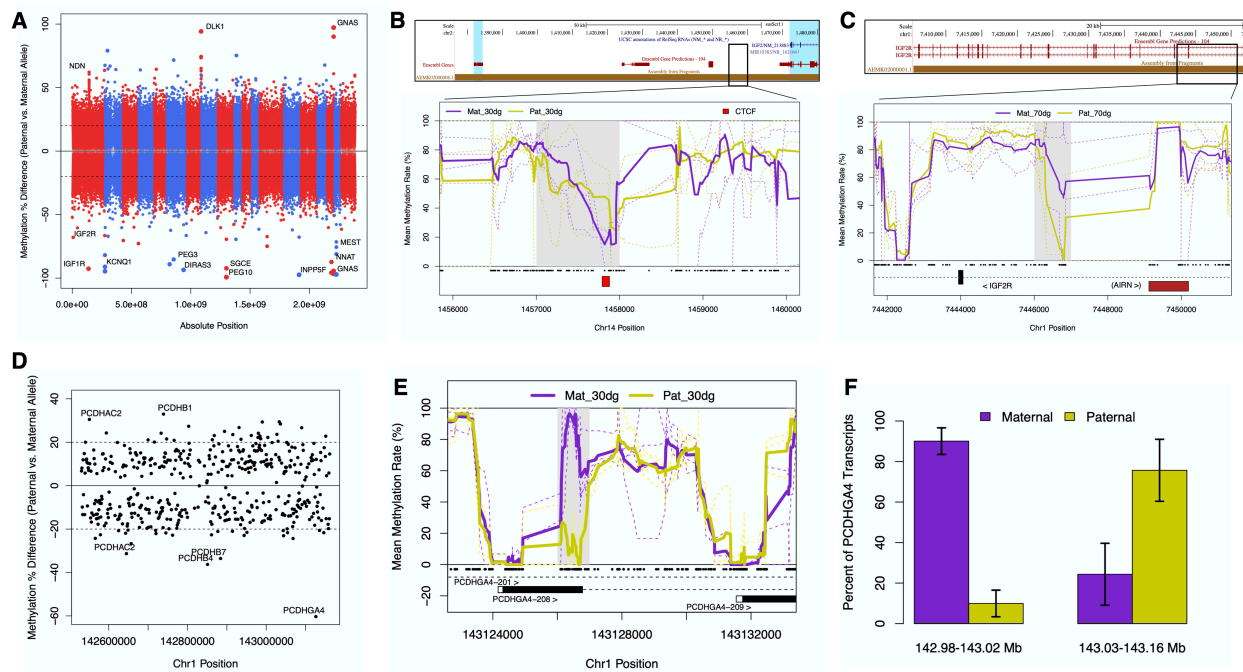


Figure 4.6. Evidence for imprinting-like allele-biased methylation in pig fetal tissues. (A) Miami plot of maximum methylation difference between parental alleles at each tested genomic region across tissues and fetal ages. Labels indicate regions overlapping known mammalian imprinted genes, and alternating colors denote different chromosomes. (B) Methylation plot of a region downstream of *IGF2* in 30 dg liver at which a paternally-hypermethylated region shares homology with a segment of the human *H19/IGF2* imprinting control region and a CTCF binding site. (C) Methylation plot of an *IGF2R* intragenic region at which a maternally-hypermethylated region lies upstream of a region with homology to the human *IGF2R* antisense RNA (*AIRN*). Bold lines indicate parental averages across individual alleles (dashed lines), and black dots denote CpG coordinates. (D) Dot plot of maximum differential methylation between parental alleles among regions in the *Sus scrofa* protocadherin locus. (E) Methylation plot of a *PCDHGA4* promoter region exhibiting extreme differential methylation between parental alleles in 30 dg brain. Bold lines indicate parental averages of individual alleles (dashed lines). (F) Bar plot showing percentage of *PCDHGA4* transcripts derived from parental alleles within two separate regions in the gene locus. Error bars indicate standard deviation.

and a BLAST search of the region downstream of this locus revealed significant sequence identity to the human *AIRN* lncRNA gene (Figure 4.6C). Thus, our assessment of paternal and maternal ABM has revealed not only significant overlap with known imprinted genes, but also with regions known to regulate genomic imprinting in other species, providing evidence for their similar function in the pig.

We identified 453 genes exhibiting ABE between maternal and paternal alleles and these were enriched among imprinting-like ABM genes ($p=1.33E-37$). Twenty-three ABE genes exhibited high imprinting-like ABM (methylation difference $>30\%$) and have not been characterized as imprinted in other mammals (Table 4.4). Protocadherin gamma-A4 (*PCDHGA4*) exhibited the strongest imprinting-like ABM, and was among multiple genes in the protocadherin domain on *S. scrofa* chromosome 1 exhibiting significant ABM (Figure 4.6D). Porcine *PCDHGA4* possesses 10 known transcript isoforms, and we observed significant paternal hypomethylation in a downstream promoter region in 30 dg brain (Figure 4.6E). Assessment of *PCDHGA4* ABE revealed isoform-dependent parental biases in expression: transcripts derived from the most upstream regions of *PCDHGA4* exhibited significant maternal biases in abundance, while those transcripts derived from regions in the vicinity of and downstream of the paternally-hypomethylated promoter exhibited paternal ABE (Figure 4.6E). These results suggest a putative imprinting-like mechanism whereby differential allelic methylation is associated with differential *PCDHGA4* isoform usage between paternal and maternal alleles in the pig fetal brain.

Table 4.4. Novel genes exhibiting imprinting-like allele methylation and expression biases

Tissue	30 dg	70 dg
Brain	<i>P3H1</i> , <i>PCDHGA4</i>	<i>AMOTL1</i> , CACNB2 , SEMA5A , TOR1AIP1
Liver	OXR1	<i>LASPI</i> , PARD3B
Muscle	SEMA5A , TGFBR3	<i>COMMD2</i> , PARD3B , SEMA5A
Placenta	FSCN1 , TGFBR3	ATP6V1B1 , <i>CYP20A1</i> , DNER , FBLN5 , <i>HGS</i> , IFFO2 , PARD3B , PTPRQ , RGS6 , SLC1A2 , <i>SOX5</i> , TRPV5

Bold: Maternal allele-biased expression (ABE); *Italics:* paternal ABE

Bold and Italics: ABE from both alleles at different locations in transcript

4.5 Discussion

We present here the first multi-tissue survey of genome-wide DNA methylation in the developing pig fetus. By profiling organs derived from unique germ layers—the endoderm (liver), mesoderm (skeletal muscle), and ectoderm (brain)—as well as the trophoctoderm (placenta), our data capture temporal DNA methylation dynamics across distinct cell lineages. Furthermore, the morphology and physiology of these organs influence a myriad of economically important pig phenotypes related to growth, behavior, and meat quality, thereby making them excellent specimens in which to better understand epigenetic gene regulation during their respective development.

We observed tissue and stage variation in global DNA methylation that is consistent with tissue differentiation in the developing fetus, with methylomes of 70 dg samples being more distinct from one another than their 30 dg counterparts. This is consistent with findings in Sliker et al. which assessed DNA methylation patterns in four human fetal tissues, including whole brain and placenta [44]. Our results demonstrate that differentiation of organ tissues is concurrent with differentiation of DNA methylation profiles that likely reflect changing gene regulatory demands throughout development. Beyond observed global differences in methylation, we identified distinct regional methylation signatures in fetal tissues indicative of unique regulatory landscapes. Tissue-specific regions of low methylation have previously been shown to act as regulatory regions

through the permissive binding of unique transcription factors [184]; we therefore sought to identify LMRs in pig fetal tissues to be classified as putative regulatory elements. The overall gain in tissue LMRs from 30 dg to 70 dg demonstrates that unique hypomethylation of genes and regulatory regions is associated with distinct differentiation trajectories. We not only observed that markers of differentiation for each tissue display low methylation specificity at 70 dg, but that there is a greater enrichment for tissue-specific GO terms among LMGs at 70 dg vs. 30 dg. Interestingly, we also observed tissue-specific enrichment for LMGs among expression-enriched genes exclusively at 70 dg, while this enrichment was not specific at 30 dg. Early development is known to involve large-scale changes in genome architecture in response to cellular identity commitment and differentiation [185]. Tissue-specific low methylation may therefore more accurately reflect gene expression demands with increasing fetal age.

Assessment of distinct regions of hyper- and hypomethylation between fetal ages revealed both common and tissue-specific methylation dynamics associated with gene regulation and expression. The enrichment of tissue DMRs in CpG islands and shores is consistent with previous findings that a small subset of CGIs is dynamically methylated during mammalian development [152, 154]. Moreover, genes overlapping DM CGIs have been shown to be involved in pattern specification and other early embryonic processes, and we show that hypermethylated CGIs across tissues are associated with *homeobox* and *t-box* gene families, among others. Differential methylation patterns were strongly associated with tissue-enriched and differential gene expression, with most tissues exhibiting relationships that are consistent with known associations between methylation and expression. These developmental DMRs overlapping DEGs represent likely *cis*-acting regulatory elements governing tissue-specific developmental gene expression.

In contrast to the associations between differential methylation and expression seen in other tissues, the fetal liver exhibited distinct correlations that emphasize its unique methylation dynamics reported in this study. In addition to being the only globally-hypermethylated tissue at 70 dg, fetal liver DMRs were not associated with liver-enriched genes, and upregulated genes were disproportionately hypermethylated as opposed to hypomethylated as was observed in other tissues. We provide evidence here that liver methylation patterns may be more influenced by changes in proliferative capacity than are patterns in other tissues by demonstrating that 1) downregulated genes in the 70 dg fetal liver are uniquely enriched for DNA replication and cell division GO terms, and 2) the 70 dg fetal liver exhibits enriched expression of genes involved in these processes. Previous research has similarly observed global DNA hypermethylation in human adult vs. fetal liver that is not associated with gene downregulation, and reported that genes involved in DNA replication were enriched in the fetal liver [186]. These results suggest that the developing liver exhibits stronger replication-dependent methylation levels than those observed in other fetal tissues. As fully-differentiated hepatocytes can re-enter the cell cycle well into adult life, this likely has unique impacts on the epigenome not observed in other mature tissues.

The reciprocal crosses of two divergent pig breeds used in this study allowed for extensive profiling of ABM in a tissue- and stage-specific manner. Allele-specific methylation patterns have previously been reported in human populations and can be attributed to differences in genetic sequence [167, 187, 188]. We similarly identified global and region-specific differences between fetal alleles derived from MS and WC parents. We demonstrate that tissue-specific ABM is associated with biological pathways previously linked to observed phenotypic differences between Chinese and European pig breeds. Breed-ABM genes in the fetal brain were most enriched for terms related to dopamine signaling, and dopamine transport and metabolism gene polymorphisms

have previously been associated with variation in aggressive behavior in MS vs. Large White breeds [171]. ABM genes in the fetal placenta were disproportionately associated with development of the placenta labyrinthine layer, the region of extensive vascularization and nutrient exchange between mother and fetus [189]. Increased vascularization of MS placentas has been extensively characterized as the primary mechanism to meet fetal nutritional demands in the absence of placental growth [170]. Among labyrinth-associated ABM genes, *FGFR2* also exhibited MS-biased expression. A human case-control study demonstrated that *FGFR2* methylation was negatively correlated with fetal birth weight and placental surface area, implicating epigenetic regulation of *FGFR2* in adverse developmental outcomes [190]. These results have thus identified strong tissue-specific candidate genes and pathways influencing behavioral and reproductive traits in pig breeds.

We identified consistent *MTRR* promoter hypomethylation in MS alleles across all tissues. *MTRR*-encoded methionine synthase reductase is required for optimal synthesis of methionine, an amino acid that is essential for diverse biological processes including protein and phospholipid biosynthesis as well as DNA methylation [191]. Studies in *MTRR*-deficient mouse models have reported numerous morphological and physiological consequences, including decreased embryonic and placental weight, decreased liver glycogen content, and decreased acetylcholine and DNA methylation levels in the brain [192–194]. *MTRR* hypomethylation in MS alleles was associated with overall MS-biased *MTRR* expression, as evidenced by significantly higher proportions of MS-derived transcripts at common exons. *MTRR* expression has previously been shown to be upregulated in the endometrium of MS vs. WC sows, as has gene expression of other methionine synthesis genes in the placenta when comparing the same pig breeds [195, 196]. However, we observed significant WC bias in abundance of isoform-specific exons, suggesting

the WC genetic background may express unique *MTRR* isoforms at higher frequencies. Due to their diverse physiological roles as well as observed differential expression in porcine reproductive tissues, *MTRR* and other methionine synthesis genes represent intriguing candidates regulating complex trait variation between Chinese and European pig breeds.

Leveraging genetic variation to assess allelic methylation also resulted in the identification of widespread genotype-independent ABM between parental alleles. We found that the majority of regions exhibiting extreme ABM between maternal and paternal alleles in our analysis overlapped known mammalian imprinted genes, and that clusters of imprinting-like ABM regions co-localized with known imprinted gene clusters. Numerous imprinted gene clusters have been shown to possess a *cis*-regulatory region at which differential methylation between maternal and paternal alleles governs monoallelic gene expression [49]. We identified imprinting-like ABM regions in both *H19/IGF2* and *IGF2R*, and their co-localization with homologous ICR-proximal elements in humans—a CTCF binding site and *AIRN* at *IGF2* and *IGF2R*, respectively—provide evidence for a similar regulatory role for these regions in the pig. Both *IGF2* and *IGF2R* are known to be mono-allelically expressed [197, 198], and while an extreme difference in methylation was observed at the putative *IGF2R* ICR, the difference at the *IGF2* ABM locus was only moderate (~20%). As the human *H19/IGF2* ICR is known to contain seven CTCF binding sites [199], it is possible that ABM at each element contributes cumulatively to exclusive *IGF2* expression from the paternal allele.

We identified dozens of genes exhibiting novel imprinting-like ABM and ABE in pig fetal tissues. Among these, *PCDHGA4* exhibited extreme paternal hypomethylation in the 30 dg brain and was among multiple regions of imprinting-like ABM in the *S. scrofa* protocadherin gene cluster. Protocadherin genes have previously been shown to exhibit allele biases in expression in

a stochastic manner similar to X chromosome inactivation, such that, at a given locus, monoallelic expression may occur from the maternal or paternal allele [200]. We provide evidence here for an imprinting-like bias in both protocadherin methylation and expression that is distinct from the random allelic variation that has been reported in humans and mice, particularly in *PCDHGA4*. While the observed paternal hypomethylation of *PCDHGA4* in 30 dg brain was associated with paternal ABE of nearby transcript isoforms, more distal and upstream isoforms exhibited maternal ABE. These results provide evidence for non-random *PCDHGA4* activation from distinct TSSs between parental alleles in the developing pig brain. Future research should seek to further assess the extent of imprinting-like biases in porcine protocadherin gene expression, and whether differential methylation of a *cis*-acting element governs these biases as has been observed in other imprinted genes.

In conclusion, we have dramatically improved the functional annotation of fetal porcine tissues through the assessment of DNA methylation patterns at two critical stages of prenatal development. Identified DMRs represent putative novel regulatory regions that may govern temporal gene expression. Furthermore, we report thousands of genes exhibiting allele biases in methylation that represent candidate loci governing breed-specific and imprinting-like gene regulation. These data will serve as a valuable resource for future endeavors seeking to define enhancers, silencers, and insulator elements important for pig fetal tissue development.

4.6 Acknowledgements

Computing resources were provided by Michigan State University's High Performance Computing Cluster. Sequencing was performed at the Michigan State University Research Technology Support Facility.

CHAPTER 5

WEANING INDUCES STRESS-DEPENDENT DNA METHYLATION AND TRANSCRIPTIONAL CHANGES IN PIGLET PBMCs

This chapter has been published previously [86]. It was prepared alongside co-authors Andrea M. Luttman, Kaitlin E. Wurtz, Janice M. Siegford, Nancy E. Raney, Laura M. Ford, and Catherine W. Ernst.

5.1 Abstract

Changes to the epigenome, including those to DNA methylation, have been proposed as mechanisms by which stress can induce long-term physiological changes in livestock species. Pig weaning is associated with dietary and social stress, both of which elicit an immune response and changes to the hypothalamic-pituitary-adrenal (HPA) axis. While differential methylation following stress has been assessed in model organisms, it remains poorly understood how the pig methylome is altered by stressors in production settings. We quantified changes in CpG methylation and transcript abundance in piglet peripheral blood mononuclear cells (PBMCs) following weaning, and also assessed differential patterns in pigs exhibiting high and low stress response as measured by cortisol concentration and lesion scores. Blood was collected from nine gilt piglets 24h before and after weaning, and whole-genome bisulfite sequencing (WGBS) and RNA-sequencing were performed on six and nine animals, respectively, at both time points. We identified 2,674 differentially methylated regions (DMRs) which were enriched within promoters of genes associated with lymphocyte stimulation and transcriptional regulation. Stress groups displayed unique differential methylation and expression patterns associated with activation and suppression of T cell immunity in low and high stress animals, respectively. Differential

methylation was strongly associated with differential expression; specifically, upregulated genes were enriched among hypomethylated genes. We observed post-weaning hypermethylation of the glucocorticoid receptor (*NR3C1*) promoter, and a significant decrease in *NR3C1* expression (n=9, p=6.1E-3). Our results indicate weaning-associated stress elicits genome-wide methylation changes associated with differential gene expression, reduced T cell activation, and an altered HPA axis response.

5.2 Introduction

Livestock animals experience numerous stressors throughout their lifetime that result in short- and long-term effects on physiology and performance [201–205]. Weaning represents a period of dietary and social stress in pig production systems with acute effects on digestive physiology and immune response [206]. In addition to direct effects of weaning, piglets are exposed to unfamiliar individuals in nursery pens, which results in aggressive encounters and skin lesion development [207]. Overall blood lymphocyte concentrations have been shown to be significantly reduced in weaned pigs [208], and additional psychosocial stress experienced during weaning has been associated with lower concentrations of T cell subpopulations [209]. One reliable indicator of stress response following pig weaning and mixing is lesion counts, as this measure has been shown to be associated with aggressive behavior, hypothalamic-pituitary-adrenal axis (HPA) activity, and risk of infection, as well as negatively associated with immunocompetence [210–213]. Weaning stress can have long-term consequences on gut health and disease susceptibility [214], yet the molecular mechanisms underlying these relationships remain poorly characterized. It has been proposed that epigenetic mechanisms may link stress from environmental stimuli to lasting changes in gene expression and physiology [62, 215].

DNA methylation is an epigenetic modification involving the enzymatic addition of a methyl group to the 5-carbon of cytosine rings, producing 5-methylcytosine. Methylation occurs almost exclusively at CpG dinucleotides in mammals and has context-specific associations with gene expression. In gene promoters and intronic enhancers, methylation generally functions to decrease levels of transcription through the alteration of transcription factor binding sites or the recruitment of transcriptional repressors [25, 31, 82]. Numerous studies have assessed DNA methylation patterns in peripheral blood mononuclear cells (PBMCs), as these can be repeatedly obtained from the same subjects and provide indications of alteration in stress response pathways; for instance, human studies have found increased levels of stress exposure to be associated with increased glucocorticoid receptor (*NR3C1*) methylation and decreased *NR3C1* expression in blood [56, 216, 217]. DNA methylation studies have been performed in pig tissues in response to different stressors [64, 80]. However, assessment of DNA methylation in the pig in response to natural production stressors has not been extensively studied, nor has the association between differential methylation and gene expression.

In the current study, we sought to identify differentially methylated regions (DMRs) and differentially expressed genes (DEGs) associated with weaning. We also assessed differential methylation and expression separately in pigs exhibiting high and low levels of weaning stress as determined by changes in serum cortisol concentration and post-weaning lesion counts. Lastly, we looked specifically at the effect of weaning on *NR3C1* methylation and expression as an indicator of alterations in the HPA axis response.

5.3 Materials & Methods

5.3.1 Sample Collection

Blood was sampled from nine crossbred gilts (four and five from two litters) weaned at 28-29 d of age. At weaning, littermates were separated into nursery pens containing 6 gilts of various litters. Samples were collected 24 h before and 24 h after weaning, and body lesions were counted immediately before and 24 h after weaning by the same trained counter [207]. Blood was collected from each animal per time point in Vacutainer whole blood collection tubes with no additive (Becton Dickinson) for serum isolation and Vacutainer tubes containing freeze-dried sodium heparin (VWR) for PBMC isolation. PBMCs were isolated from each animal at both time points following centrifugation in ACCUSPIN™ System-Histopaque-1077® tubes (Sigma Aldrich) according to manufacturer's instructions. Serum cortisol concentrations were measured using the DetectX Cortisol Enzyme Immunoassay from Serum and Plasma Kit (Arbor Assays). Optical density was measured at 450nm using a microplate reader, and readings were converted to ng/ml concentrations using Arbor Assays software. We assessed changes in serum cortisol concentration and post-weaning lesion counts as phenotypic measurements of stress response. High stress (HS) and low stress (LS) animals were visually selected as those at the extremes of the two-dimensional plot of the two phenotypes, and further subjected to DNA methylation analyses (Fig S1).

5.3.2 Nucleic Acid Isolation and Sequencing

DNA was isolated from six pre- and six post-weaning PBMC samples using the PureLink Genomic DNA Kit. Prior to library preparation, DNA samples were spiked with unmethylated lambda phage DNA (5 ng lambda DNA/1 µg sample DNA) to assess sodium bisulfite conversion. DNA was sodium bisulfite converted using the Zymo EZ DNA Methylation-Gold Kit. Eight of the 12 libraries were prepared using the Kapa Hyper Prep DNA Kit (Roche), and the remaining four

libraries were prepared using the Swift Biosciences Accel-NGS Methyl-Seq Library Kit. WGBS was performed as described in Section 2.3.3.

PBMC RNA from all animals at both time points was isolated using the miRNeasy Mini Kit (Qiagen). For eight of the 18 RNA samples, ribosomal RNA (rRNA) was depleted using the Illumina Ribo-Zero Gold Kit; the remaining RNA samples were subject to rRNA and globin RNA depletion using Illumina Ribo-Zero Plus Kit. All sequencing libraries were prepared using the Illumina TruSeq Total RNA Library Preparation Kit. RNA-sequencing (RNA-seq) was performed as described in Section 3.3.1.

5.3.3 WGBS Bioinformatics

WGBS read trimming, alignment, deduplication, and methylation extraction was performed as described in Section 2.3.4. Due to sub-optimal conversion in some samples, we ran Bismark's `filter_non_conversion` command with default parameters to remove reads that contained ≥ 3 consecutive methylated non-CpG cytosines. Conversion rates were $\geq 99.5\%$ in all libraries following this filtering.

Differential methylation analysis between stages was performed using the *methylKit* R package v.1.6.3 as described in Section 3.3.3. [89]. Logistic regression models were fitted for each region accounting for the fixed effects of library prep and animal within library prep, and random effect of litter to test if stage (post- vs. pre-weaning) had a significant effect on the log odds ratio of regional CpG methylation rate. DMRs were classified as those with methylation difference $>10\%$ and $FDR < 0.05$. Separate post-weaning differential methylation analyses were also run for LS and HS animals using the same methods.

DMRs were annotated with respect to their overlap with gene features to identify differentially methylated genes (DMGs). Genes were classified as promoter- or gene body-DMGs

if they contained a DMR in their promoter (2kb up- or downstream of transcription start site (TSS)) or gene body, respectively. DMRs not overlapping a gene were classified as intergenic. DMG lists were submitted for gene set enrichment analysis (GSEA) using GOrilla software [218].

5.3.4 RNA-sequencing Bioinformatics

RNA-seq reads were trimmed of adapters and low-quality bases using Trimmomatic (HEADCROP:10 LEADING:25 TRAILING:25 AVGQUAL:20 MINLEN:30). Trimmed reads were aligned to the *S. scrofa* reference genome using TopHat2 [145] (parameter: --library-type 'fr-firststrand'), and gene counts were obtained using HTSeq-count [146] (parameters: -m intersection-nonempty -i gene_id -t exon -s reverse).

Differential expression analyses were performed using the *DESeq2* R package [93]. A negative binomial model was fitted that tested for the effect of stage, and corrected for the effects of litter, library preparation and animal within library preparation. DEGs were classified as those with FDR <0.05, regardless of log2-Fold Change. DEGs were submitted for GSEA using GOrilla to identify enriched GO terms. Four separate *DESeq2* analyses were performed for: 1) all animals (n=9/stage), 2) HS animals and 3) LS animals (n=3/stage), and 4) animals with WGBS data (n=6/stage).

5.3.5 Assessment of *NR3C1* Methylation and Expression

We assessed *NR3C1* promoters for regional and site-specific methylation differences, and PBMC *NR3C1* transcript abundance was measured by RNA-seq and RT-qPCR. Total RNA from PBMCs was reverse transcribed as described in Section 2.3.9. *B2M* and *GAPDH* were used as reference genes due to their reported stable expression in PBMCs [101, 102]. qPCR assays were performed in triplicate as described in Section 2.3.9, with the modification of adding 50 ng cDNA at 5 µl volume to each reaction well. Delta Cts (dCts) were obtained for each sample by subtracting the

geometric mean of the reference gene Cts from the *NR3C1* Ct, and a paired t-test was performed to assess the significance of stage on dCts. Fold change in abundance at post- versus pre-weaning was calculated using the $2^{-\Delta\Delta Ct}$ method.

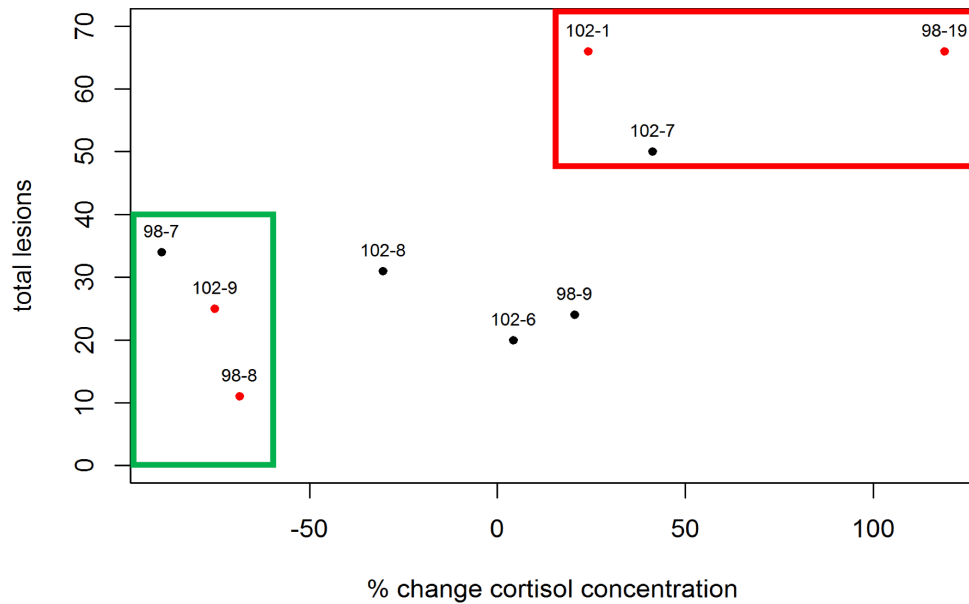


Figure 5.1. Total post-weaning lesion score plotted against percent change in cortisol concentration (post- vs. pre-weaning) in nine piglets. Samples in green and red boxes were designated as low stress and high stress, respectively.

5.4 Results

5.4.1 Serum cortisol and lesion count measurements identify low- and high-stress animals following weaning

We measured serum cortisol concentrations and skin lesion counts before and after weaning in nine gilts in order to assess differential stress response. Percent change in serum cortisol concentration was significantly positively correlated with post-weaning lesion counts ($r=0.70$, $p=0.036$; Figure 5.1). We designated the three animals exhibiting the lowest and highest values of these parameters as LS (two gilts from litter 98 and one from litter 102) and HS (one gilt from litter 98 and two from litter 102), respectively.

5.4.2 Weaning differential methylation is associated with lymphocyte immune response genes

We obtained 128-194M WGBS reads across samples, of which 86-89% uniquely aligned to the *S. scrofa* reference genome (Table D.1). We achieved sufficient coverage to assess differential methylation post- versus pre-weaning at 972,067 1kb regions. We observed clustering of samples based on litter (data not shown), and thus corrected for this as a random effect in our differential methylation analysis. We identified 2,674 DMRs between stages when considering all animals, of which 1,363 were hypermethylated and 1,311 hypomethylated post-weaning (Figure 5.2A). We annotated DMRs along with all tested regions to gene features, and observed a 3.4-fold and 2.1-fold overrepresentation of hyper- and hypo-DMRs in gene promoters, respectively (Figure 5.2B). Conversely, there was a depletion of DMRs in gene bodies and intergenic regions.

We also assessed post- versus pre-weaning differential methylation separately in LS and HS animals, and identified unique patterns between groups. HS animals possessed a greater number of DMRs overall (9,945 vs 6,141); however, while HS animals had roughly the same number of hyper- and hypo-DMRs (4,938 and 5,007, respectively), LS animals had a greater proportion of hypo-DMRs (n=3,473) relative to hyper-DMRs (n=2,668). Similar to DMRs among all animals, HS hyper-DMRs were more overrepresented in promoters than were HS hypo-DMRs (Figure 5.2B), while LS hyper and hypo-DMRs were overrepresented in promoters at similar magnitudes. These results indicate that differential methylation at weaning is strongly associated with gene promoters, and suggest that the proportions of hyper- and hypo-DMRs are dependent on stress level.

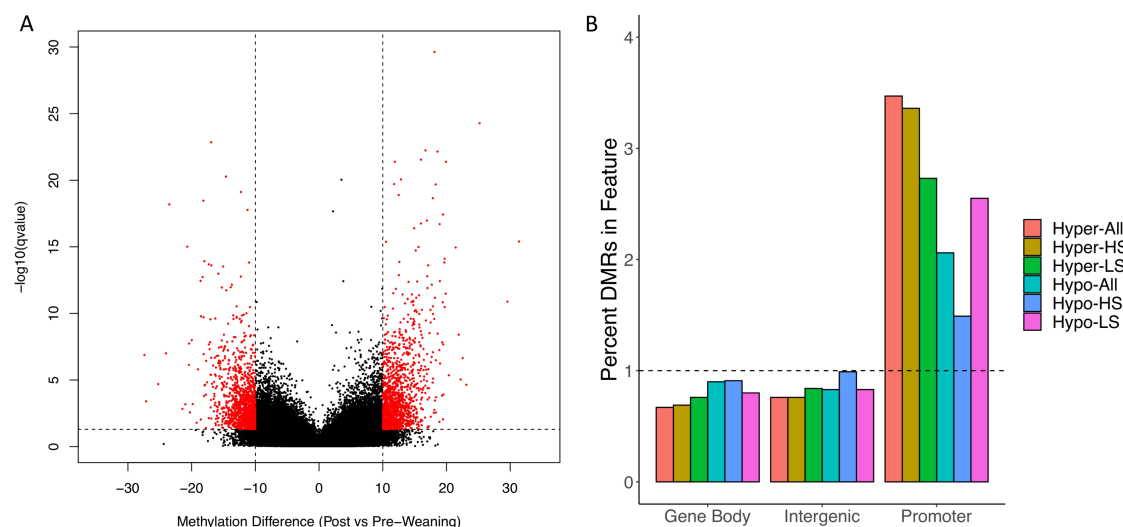


Figure 5.2. Differential methylation in post- versus pre-weaning PBMCs. (A) Volcano plot of methylation difference against $-\log_{10}(qvalue)$. Red dots indicate significant differentially methylated regions (DMRs). **(B)** Enrichment of hypomethylated and hypermethylated DMRs for All, HS, and LS pigs in gene features. Horizontal line indicates expected relative proportion of DMRs in feature if no enrichment (i.e. 1). HS = high stress; LS = low stress.

To determine if differential methylation influences genes involved in similar biological processes, we submitted DMGs for gene set enrichment analysis (Table 5.1). Among all animals, promoter-DMGs were enriched for T cell- and plasma cell-specific processes. Genes hypermethylated in their promoters included interleukin 2 receptor alpha (*IL2RA*) and LIM domain only 1 (*LMO1*), whereas hypomethylated genes included several involved in T cell proliferation (*CD3E* and *TNFSF14*) and apoptosis (*LGALS1*). While hypermethylated gene body-DMGs were not enriched for specific processes related to immunity, we identified several such processes enriched among hypomethylated DMGs including ‘Positive regulation of lymphocyte differentiation’ and ‘B cell receptor signaling pathway’.

Table 5.1. Enriched GO Terms among promoter- and gene body-DMGs

Hypermethylated Promoter-DMGs			
GO Term	Enrichment	P-value	Genes
Catecholamine secretion	34.05	8.60E-04	<i>MECP2, NISCH</i>
Regulation of T cell homeostatic proliferation	34.05	8.60E-04	<i>IL2RA, LMO1</i>
Hypomethylated Promoter-DMGs			
GO Term	Enrichment	P-value	Genes
T cell costimulation	7.44	5.30E-04	<i>PTPN6, TNFSF14, LGALS1, CD3E,</i>
Lymphocyte costimulation	7.28	5.8E-04	<i>PIK3R1</i>
Plasma cell differentiation	43.66	6.98E-04	<i>LGALS1, ITM2A</i>
Hypomethylated Gene Body-DMGs			
GO Term	Enrichment	P-value	Genes
Positive regulation of lymphocyte differentiation	3.62	2.07E-04	<i>PIK3R6, RASGRP1, IL4R, TOX, ZBTB7B, IL2RA, IL12RB1, PTPRC, INPP5D, ZMIZ1, RUNX1</i>
Regulation of response to stimulus	1.28	2.68E-04	157 genes
Positive regulation of leukocyte differentiation	2.92	3.06E-04	<i>PIK3R6, IL4R, TOX, IL2RA, TRIB1, IL12RB1, GNAS, INPP5D, RUNX1, RASGRP1, ZBTB46, ZBTB7B, PTPRC, ZMIZ1</i>
Immune response-activating cells surface receptor signaling pathway	2.31	4.34E-04	<i>VAV3, ELMO2, ICAM2, MAPK1, BLK, INPP5D, PSMB7, MYO1G, CYFIP2, ACTB, LCP2, PRKCB, BTRC, EIF2B2, MYO1C, NFATC2, PTPRC, ARPC1B, CD247, PDE4B</i>
B cell receptor signaling pathway	5.50	6.17E-04	<i>VAV3, ELMO2, ICAM2, MAPK1, BLK, PTPRC, NFATC2</i>

We also identified shared and unique enriched GO terms among DMGs in LS and HS animals (Table 5.2). In both groups, post-weaning promoter hypomethylation was enriched in genes involved in leukocyte differentiation and activation. Post-weaning hypomethylated genes in the LS group were also enriched for T cell-related processes (‘alpha-beta T cell receptor complex’, ‘regulation of gamma-delta T cell activation’). Conversely, terms related to T cells (‘alpha-beta T cell activation’, T cell differentiation’) were enriched among hypermethylated genes in the HS group. As gene hypomethylation is generally associated with gene activation, these data suggest that weaning stress level is negatively associated with T cell activity in part through differences in

Table 5.2. GO Terms enriched among DMGs in low and high stress (LS & HS) groups

GO Term	No. Genes	Enrichment	P-value
<i>LS Hypermethylated Genes</i>			
Platelet activation	14	2.74	4.98E-04
<i>LS Hypomethylated Genes</i>			
Positive regulation of myeloid cell differentiation	13	3.15	2.12E-04
Alpha-beta T cell receptor complex	4	15.7	3.21E-05
Regulation of gamma-delta T cell activation	4	8.29	5.47E-04
<i>HS Hypermethylated Genes</i>			
T cell differentiation	21	2.57	4.18E-04
Alpha-beta T cell activation	11	3.11	5.07E-04
B cell receptor signaling pathway	10	3.93	1.04E-04
<i>HS Hypomethylated Genes</i>			
Regulation of leukocyte degranulation	8	5.65	6.23E-05
Leukocyte activation	47	1.74	1.47E-04

DNA methylation. To further test this hypothesis, we assessed T cell co-receptor genes for post-weaning differential methylation in both stress groups. We identified opposing differential methylation in these genes between stress groups (Table 5.3). Regions in *CD3E*, *CD3D*, *CD3G*, and *CD4* were hypomethylated post-weaning in the LS group but not DM in the HS group, while a *CD8B* region was hypermethylated post-weaning in the HS group and not DM in the LS group.

5.4.3 Weaning differential gene expression is associated with differential gene methylation

We assessed PBMC gene expression in nine gilts via RNA-seq and obtained 62-108M reads per sample (Table D.2). We identified 13,580 expressed genes and 1,480 DEGs between stages (721 upregulated and 759 downregulated post-weaning). Numerous GO terms were enriched among upregulated genes, the most significant of which were related to immune, inflammatory, and stress response (data not shown). Downregulated genes were enriched for terms related to transcriptional and post-transcriptional regulation (data not shown).

Table 5.3. Weaning differential methylation and expression of T cell co-receptor genes.

Gene	DMR location	Gene Feature	Methylation Difference (Post vs Pre)			Log2 Fold Change (post vs. pre-weaning)		
			All	LS	HS	All	LS	HS
<i>CD3E</i>	9:45620001	Promoter	-9.13	-13.17	-6.02	0.076	0.064	0.024
	9:45621001	Promoter	-16.43	-19.64	-17.50			
	9:45624001	Gene Body	-6.63	-13.63	2.98			
<i>CD3D</i>	9:45645001	Promoter	-5.08	-18.71	2.95	0.273	0.165	0.081
<i>CD3G</i>						-0.460	-0.528	-0.588
<i>CD4</i>	5:63916001	Promoter	-5.39	-17.70	6.06	-0.122	-0.143	-0.387
<i>CD8A</i>	---	---	---	---	---	-0.311	-0.079	-0.59
<i>CD8B</i>	3:57970001	Promoter	16.50	15.17	17.49	-0.80	-0.251	-1.67
	3:57988001	Gene Body	11.93	-0.81	20.57			

Differential expression analyses were also performed on LS and HS animals separately. A larger number of weaning DEGs were identified in the LS group (134 upregulated and 42 downregulated post-weaning) compared to the HS group (49 genes upregulated and 13 downregulated). DEGs were submitted for GSEA, and only upregulated genes contained enriched GO terms. There were numerous terms enriched among both sets of upregulated genes (e.g. ‘Inflammatory response’, ‘Positive regulation of cytokine production’), but unique GO terms were also enriched (Table 5.4). LS upregulated genes were enriched for terms related to viral response, type I interferon signaling, and NIK/NF-kappaB signaling. HS upregulated genes were enriched for terms related to apoptosis and negative regulation of CD4-positive, alpha-beta T cell proliferation and activation.

To determine the degree of overlap between DEGs and DMGs, we identified DEGs between stages in the same six animals for which methylation data were generated. There was a total of 387 DEGs (275 upregulated and 112 downregulated), and these were significantly enriched among DMGs. Twenty-eight DEGs were also promoter-DMGs, and there was particular enrichment for upregulated genes among hypomethylated promoter-DMGs ($p=3.56E-3$).

Additionally, 29 DEGs were also gene body-DM, and there was again enrichment for upregulated genes among hypomethylated gene body-DMGs ($p=0.021$). There is thus evidence that differential methylation is strongly associated with differential expression in post-weaning piglet PBMCs, and specifically that hypomethylation is associated with increased gene expression.

Table 5.4. GO terms enriched among upregulated genes in low and high stress groups

GO Term	No. Genes	Enrichment	P-value
<i>LS Upregulated Genes</i>			
Inflammatory response	23	5.95	4.35E-12
Positive regulation of cytokine production	27	5.99	3.96E-14
Type I interferon signaling pathway	17	29.38	1.12E-21
Interferon-gamma-mediated signaling pathway	13	19.4	4.88E-14
Positive regulation of NIK/NF-kappaB signaling	6	7.43	1.44E-04
Response to virus	36	12.06	2.02E-29
<i>HS Upregulated Genes</i>			
Inflammatory response	13	8.51	2.46E-09
Positive regulation of cytokine production	10	5.62	9.16E-06
Positive regulation of T cell apoptotic process	3	55.46	1.69E-05
Positive regulation of lymphocyte apoptotic process	3	41.59	4.37E-05
Negative regulation of alpha-beta T cell activation	3	19.2	4.87E-04
Negative regulation of alpha-beta T cell proliferation	2	66.55	3.51E-04

Because T cell co-receptor genes exhibited unique differential methylation patterns between stress groups, we determined the extent to which such patterns associated with differential expression between HS and LS animals. *CD8B*, which was gene body-hypermethylated only in the HS group post-weaning, was also significantly downregulated in the HS group (Table 5.3; $\log_2FC=-1.67$) but not in the LS group ($\log_2FC=-0.25$). *CD4* exhibited post-weaning promoter hypomethylation in the LS group but not in the HS group, and exhibited a greater decrease in expression in the HS group ($\log_2FC= -0.387$) although this difference was not statistically significant. Changes in expression of genes with associated differences in regional methylation

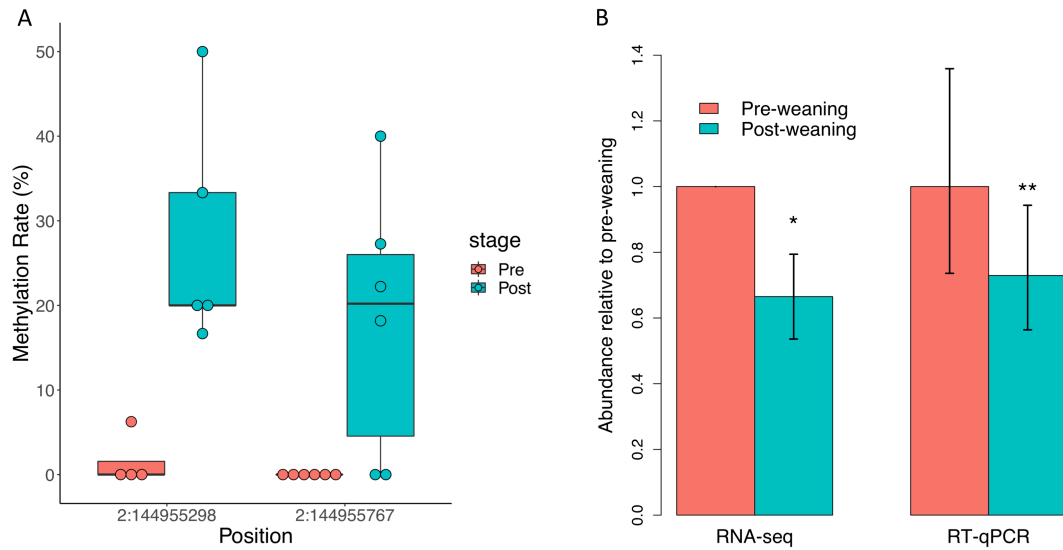


Figure 5.3. Glucocorticoid receptor (*NR3C1*) gene methylation and expression in response to weaning. (A) Two CpG sites in the *NR3C1* promoter (797 and 328 bp downstream of TSS) are significantly hypermethylated post-weaning. **(B)** *NR3C1* transcript abundance is significantly reduced post-weaning, as measured by RNA-sequencing and RT-qPCR.

suggest that such regions may harbor regulatory elements that dictate stress-dependent T cell gene expression.

5.4.4 *NR3C1* differential methylation and expression is indicative of altered HPA axis response

We did not observe significant regional differences in CpG methylation associated with weaning in the two *NR3C1* promoters in the pig genome. We thus assessed individual CpG methylation differences and identified two differentially methylated CpGs (DMCs), both of which were hypermethylated post-weaning (Figure 5.3A). These DMCs lie 797 and 328 bp downstream of the first *NR3C1* TSS, and exhibited 17% and 21% increases in methylation post-weaning, respectively. We observed a corresponding decrease in *NR3C1* transcript abundance via RNA-seq (Figure 5.3B; $\log_2FC = -0.588$, $p = 0.026$), which was also validated via RT-qPCR ($\log_2FC = -0.455$, $p = 6.1E-3$). These results recapitulate findings in human and mouse studies that stress exposure is associated with *NR3C1* promoter hypermethylation and decreased expression in peripheral tissues.

5.5 Discussion

This study has identified epigenetic alterations in pigs as a response to weaning stress by assessing DNA methylation patterns in piglet PBMCs prior to and after weaning and mixing with unfamiliar individuals. By selecting animals displaying extremes in post-weaning serum cortisol change and skin lesion counts, we were able to assess not only the overall effect of weaning on DNA methylation in PBMCs, but also how responses varied depending on stress level. PBMCs are a valuable peripheral cell type to study in the context of weaning for numerous reasons; first, they consist primarily of monocytes and lymphocytes whose activity, proliferation, and differentiation has been shown to be significantly altered by weaning-associated stress [208, 219, 220]. Furthermore, PBMC expression and methylation of genes involved in the HPA axis—namely *NR3C1*—have been shown to be suitable ‘surrogates’ for measurement of gene activity in neuronal tissues [56].

We observed global CpG methylation rates in piglet PBMCs between 79.1% and 82.9%. A high proportion of variation in DNA methylation could be attributed to litter, which emphasized the need to correct for this variable when assessing for weaning-specific differential methylation patterns. The presence of 2,674 DMRs between stages indicates that weaning-associated changes in DNA methylation were present at many genomic loci. Additionally, we observed unique differential methylation patterns between LS and HS animals, with HS animals possessing more DMRs but LS animals having a greater proportion of hypomethylated regions. Previous studies in livestock species have identified differential methylation patterns in stressed versus non-stressed animals. Hao et al. 2016 identified thousands of DMRs in *longissimus dorsi* muscle of heat-stressed vs. control pigs within genes involved in energy metabolism, stress response, and calcium signaling [64]. Multiple studies have identified differential lymphocyte DNA methylation

associated with prenatal transportation stress in Brahman bulls and heifers [59, 221, 222], and identified thousands of DM loci enriched in stress and immune response genes. Our results indicate that weaning stress also alters DNA methylation patterns in pig immune cells—particularly within gene promoters—and that the magnitude and direction of such alterations may be dependent on the level of stress experienced.

When considering all animals, post-weaning differential methylation impacted genes involved in immune cell processes. *IL2RA* exhibited hypermethylation in post-weaning PBMCs; this gene has previously been shown to possess extensive promoter hypomethylation in activated CD4⁺ T cells [223], suggesting that the opposite state observed here may suppress such activation. Clear differences were observed when assessing GO enrichment among LS and HS DMGs. Particularly, LS animals exhibited hypomethylation of genes involved in T cell activation and differentiation, while HS animals exhibited hypermethylation of genes involved in these processes. Overall, expression patterns were consistent with the differential methylation observed in LS and HS animals in terms of impacted biological processes. Upregulated HS genes were enriched for GO terms related to T cell apoptosis and negative regulation of CD4⁺ T cell proliferation, and these were not observed among the LS upregulated genes. Previous studies have shown CD4⁺ T cell concentrations to be the most significantly reduced following periods of psychosocial and weaning stress [209]. Our data suggest that differential gene regulation by DNA methylation may play a role in a reduced T cell response with increasing levels of stress. This was particularly evident when assessing differential methylation and expression by stress group among T cell co-receptor genes. Many of these genes exhibited post-weaning differential methylation patterns indicative of lower gene activation in the HS group compared to the LS group, and *CD8B* also exhibited significantly lower expression post-weaning only in the HS group. *CD8B* encodes a subunit of the

CD8 co-receptor in cytotoxic T cells. Studies have shown that corticosterone injections decrease CD8⁺ T cell concentrations in humans [224, 225], and that in pigs cytotoxic T cell concentrations decrease following weaning [208]. Our methylation analyses have identified a DMR in the gene body of *CD8B* that may act in regulating *CD8B* expression, particularly in response to weaning stress and cortisol levels.

Lastly, we observed significant post-weaning hypermethylation in CpGs in the *NR3C1* promoter, and a corresponding decrease in expression. The glucocorticoid receptor not only functions as an inducer of cortisol-mediated transcription in peripheral tissues, but also regulates the HPA axis response in a negative feedback loop [226]. *NR3C1* hypermethylation and downregulation in the hypothalamus has often been an indicator of stress vulnerability, and recent studies have shown comparable patterns in PBMCs [216, 217]. Our data show that pigs exhibit similar patterns of *NR3C1* hypermethylation and downregulation in response to weaning stress. However, we did not observe significant differences in post-weaning *NR3C1* methylation and expression between LS and HS pigs, potentially due to our limited sample size in this study.

In summary, we have elucidated epigenetic patterns of acute weaning-associated stress response in pigs. Future studies should seek to assess patterns of methylation and expression in PBMCs at later periods following weaning to assess long-term effects of weaning stress on pig immunity and performance. Additionally, assessment of other tissues involved in the HPA axis would provide a more direct measurement of stress response that could be compared to regulation and expression of genes in peripheral tissues. Continued analysis of genes undergoing stress-dependent gene regulation and expression may reveal biomarkers that are predictive of pig stress resilience.

5.6 Acknowledgements

The authors wish to thank Kevin Turner and the MSU Swine Teaching and Research Center for rearing and caring for the animals, and for their assistance with blood sample collection, as well as the MSU Genomics Core for their technical and sequencing support.

CHAPTER 6

GENOME-WIDE ASSESSMENT OF DNA METHYLATION IN CHICKEN CARDIAC TISSUE EXPOSED TO DIFFERENT INCUBATION TEMPERATURES AND CO₂ LEVELS

This chapter has been published previously [227]. The manuscript was prepared alongside co-authors Marinus F.W. te Pas, Henry van den Brand, Martien A.M. Groenen, Richard P.M.A. Crooijmans, Catherine W. Ernst, and Ole Madsen.

6.1 Abstract

Temperature and CO₂ concentration during incubation have profound effects on broiler chick development, and numerous studies have identified significant effects on hatch heart weight (HW) as a result of differences in these parameters. Early-life environment has also been shown to affect broiler performance later in life; it has thus been suggested that epigenetic mechanisms may mediate long-term physiological changes induced by environmental stimuli. DNA methylation is an epigenetic modification that can confer heritable changes in gene expression. Using reduced-representation bisulfite sequencing (RRBS), we assessed DNA methylation patterns in cardiac tissue of 84 broiler hatchlings incubated at two egg shell temperatures (EST; 37.8°C and 38.9°C) and three CO₂ concentrations (0.1%, 0.4%, and 0.8%) from day 8 of incubation onward. We assessed differential methylation between EST treatments and identified 2,175 differentially methylated (DM) CpGs (1,121 hypermethylated, 1,054 hypomethylated at 38.9° vs. 37.8°) in 269 gene promoters and 949 intragenic regions. DM genes (DMGs) were associated with heart developmental processes, including cardiomyocyte proliferation and differentiation. We identified enriched binding motifs among DM loci, including those for transcription factors associated with cell proliferation and heart development among hypomethylated CpGs that suggest increased

binding ability at higher EST. We identified 9,823 DM CpGs between at least two CO₂ treatments, with the greatest difference observed between 0.8% and 0.1% CO₂ that disproportionately impacted genes involved in cardiac muscle development and response to low oxygen levels. Using HW measurements from the same chicks, we performed an epigenome-wide association study (EWAS) for HW, and identified 23 significantly associated CpGs, nine of which were also DM between ESTs. We found corresponding differences in transcript abundance between ESTs in three DMGs (*ABLIM2*, *PITX2*, and *THRSP*). Hypomethylation of an exonic CpG in *PITX2* at 38.9°C was associated with increased expression, and suggests increased cell proliferation in broiler hatchlings incubated at higher temperatures. Overall, these results identified numerous epigenetic associations between chick incubation factors and heart development that may manifest in long-term differences in animal performance.

6.2 Introduction

Early-life environmental parameters have profound effects on broiler chick development and post-hatch performance. Incubation temperature is one of the most important factors influencing embryonic growth, development, and physiology [228–231]. In studies assessing the effects of incubation egg shell temperature (EST) on organ growth, heart weight (HW) is consistently observed to be negatively correlated with EST [67, 202, 232–234]. Mechanisms linking incubation EST and heart development have been proposed; studies have shown that increased incubation temperature decreases the mitotic index of cardiomyocytes [235], as well as the concentration of circulating T3, the metabolically active form of thyroid hormone [236, 237]. Additionally, phenotypic differences in adult broilers associated with differences in early life temperature have been observed. Chicks incubated at a higher EST from embryonic day 7 (E7) to hatch were found to have a higher incidence of ascites mortality at six weeks of age [238, 239]. However, broilers

exposed to increased temperature early in life exhibited decreased mortality when exposed to heat stress again at 6 weeks of age relative to unexposed animals, suggesting that early heat stress confers thermotolerance that can improve performance upon subsequent exposures [201].

In addition to temperature, CO₂ concentration during the incubation period has been suggested to be important for regulating chick growth and physiology. Embryos incubated at increased CO₂ concentration have been found to exhibit numerous phenotypes: in addition to a significantly greater average HW at hatch [234], lower blood oxygen levels relative to control chicks have also been observed [240]. Increased incubator CO₂ concentration was associated with decreased ascites mortality in broilers later in life [241], suggesting long-term effects of early life CO₂ exposure. Alterations made to the epigenome have been proposed as one mechanism by which early-life environmental differences can manifest in phenotypic differences in adult precocial birds [62, 242], and it has recently been shown that embryonic thermal manipulation induces epigenetic modifications in the hypothalamus of 35-day old broilers [69].

Numerous epigenetic molecular mechanisms including DNA methylation, histone tail modifications, and non-coding RNA interactions have been observed to induce mitotically heritable changes in gene expression without altering DNA sequence [63, 243]. DNA methylation is an epigenetic modification involving the enzymatic addition of a methyl group to the 5-carbon of cytosine rings, producing 5-methylcytosine. Methylation occurs almost exclusively at CpG dinucleotides in vertebrates and has context-specific associations with gene expression. In gene promoters, methylation generally functions to decrease levels of transcription through the alteration of transcription factor (TF) binding sites, the recruitment of transcriptional repressors, or changes in chromatin conformation [31, 82]. Conversely, gene body methylation is generally associated with increased levels of transcription [35, 37, 244], although negative associations have

been identified in the context of genic enhancers [25]. Epigenetic processes are known to be altered in response to environmental perturbations, and this phenomenon has been studied in livestock species in response to heat stress [64, 70, 245]. Additionally, numerous epigenome-wide association studies (EWAS) have been performed in humans and mice to identify loci at which methylation level is significantly associated with complex and disease traits [246, 247]. Bisulfite-sequencing approaches allow for genome-wide assessment of DNA methylation patterns, yet this approach has been underutilized in the chicken thus far [248–251].

In the current study, we assessed DNA methylation patterns in cardiac tissue of broiler chicks incubated at normal or high EST (37.8 or 38.9°C, respectively) and low, medium, or high CO₂ concentration (0.1%, 0.4%, 0.8%, respectively) from embryonic day 8 (E8) until hatch. Using reduced representation bisulfite sequencing (RRBS), we tested for differences in CpG methylation rates between EST and CO₂ treatments and identified thousands of loci exhibiting differential methylation that are associated with heart development. Additionally, by integrating phenotypic records from the same samples, we performed an EWAS and identified loci at which methylation rate was significantly associated with HW, and many of these sites were within genes with known roles in heart development. Our results provide evidence that EST and CO₂ may be influencing heart growth and physiology through changes in cardiac DNA methylation patterns.

6.3 Material and Methods

6.3.1 Experimental Design and Sample Collection

Heart tissue samples were collected from a 2x3 factorial study. Briefly, Cobb 500 broiler eggs were obtained from commercial broiler breeder farms and incubated in six consecutive batches. From day 0 to 8 of incubation (E8), all eggs were incubated at an EST of 37.8°C and 0.1% CO₂.

At E8 eggs were divided into two ESTs (37.8°C and 38.9°C) and three CO₂ concentrations (0.1%, 0.4%, and 0.8% CO₂) until hatch. Both EST treatments were applied in all batches, but CO₂ treatment application varied between batches. Time until hatch was recorded for each chick, and animals were removed from the incubator six hours post-hatch. Chicks were then killed by cervical dislocation followed by decapitation. Hearts were removed and stored at -20°C until further analysis. Heart samples derived from the left ventricle of 84 chicks (13 to 15 samples per EST-CO₂ combination) were utilized for DNA isolation and RRBS.

6.3.2 Statistical Analyses of Heart Weight

We tested for the significance of EST, CO₂ concentration, and their interaction on HW using analysis of variance (ANOVA). To account for the significantly shorter incubation time of 38.9°C chicks, we constructed a linear model with HW as the response variable and incubation time as a fixed effect, and subsequently used the model residuals as a response variable to test the significance of the environmental parameter main effects and their interaction.

6.3.3 Sample Processing and Bisulfite Sequencing

DNA and RNA from heart samples were isolated using the AllPrep DNA/RNA Mini Kit (Qiagen) following manufacturer's instructions. The RRBS libraries were prepared with the Ovation RRBS Library construction kit from Nugen following manufacturer's instructions. Isolated DNA was cut with MspI and a fragment size selection of 20-250 bases was applied. Samples were pooled across 9 flow cell lanes with each pool containing no more than 10 samples, and all pools were sequenced for 161 cycles on a HiSeq 2500 using the TruSeq SBS sequencing kit version 4. Fastq files were generated and demultiplexed with the bcl2fastq v2.17.1.14 Conversion Software (Illumina).

6.3.4 Bioinformatics Analyses

Raw RRBS FASTQ files were trimmed of adapter sequences using Trim Galore v0.5.0 with Cutadapter v1.8.1 (Babraham Bioinformatics) and the parameters: -a AGATCGGAAGAGC. Reads were subsequently filtered for only those beginning with the expected YGG trinucleotide sequence using a python script provided by NuGEN Technologies (<https://github.com/nugentechnologies/NuMetRRBS>).

Read alignment was performed using BS-seeker2 v2.1.3 [252]. An RRBS bowtie2 index was generated from the *Gallus gallus* GRCg6a reference genome assembly using `bs_seeker_build.py` with the following parameters: `--aligner bowtie2 --rrbs --low 10 --up 280`. Trimmed reads were subsequently aligned to the RRBS index using `bs_seeker2-align.py` with the following parameters: `--rbbs --low 10 --up 280 --mismatches 0.05 --aligner=bowtie2 --bt2-p 10 --bt--local --bt2-N 1`. Binary alignment map (BAM) files were converted to CGmap files using CGmaptools v0.1.2 [253], which reports methylation rates for all covered CpGs.

CGmap files were used to calculate sample global methylation rates, and to assess their associations with EST, CO₂, and individual CpG methylation rates. We identified an overrepresentation of CpG sites where methylation was significantly correlated with sample global CpG methylation rate (data not shown). We therefore corrected for this factor in subsequent site-specific differential methylation analyses.

6.3.5 Differential Methylation Analyses

We filtered CGmap files to only include CpGs with coverage of at least 10 reads within a sample. Filtered CGmap files were subsequently converted to CpG report text files for *methylKit* analysis using a custom python script. Differential methylation analyses were performed using the *methylKit* R package v1.8.1 [89]. Briefly, we removed CpGs in the 99.5th percentile of coverage

for each sample to remove potential PCR duplicates. For the EST differential methylation analysis (38.9°C vs. 37.8°C), CpGs were retained if they were covered in at least 25 samples per EST treatment over all CO₂ treatments. A logistic regression model was fitted for each CpG accounting for the covariates of CO₂ level and sample mean CpG methylation rate to test if EST had a significant effect on the log odds ratio of the CpG methylation rate. For CO₂ differential methylation analyses, CpGs were retained if they were covered in at least 16 samples per CO₂ treatment over both EST treatments. Three different analyses were performed to contrast the three different CO₂ treatment pairs (0.4% vs. 0.1%, 0.8% vs. 0.4%, and 0.8% vs. 0.1% CO₂). Again, a logistic regression model was fitted for each CpG, accounting for covariates of EST and sample mean CpG methylation rate, and testing for the effect of CO₂ level. A CpG site was considered significantly differentially methylated (DM) if the difference in methylation rate between treatments was greater than 10% and the corresponding q-value was less than 0.05 in both EST and CO₂ comparisons.

6.3.6 CpG Annotation

All analyzed CpGs were annotated with respect to overlap with genes using the genomation R package v1.14.0 [90]. We defined promoter CpGs as those within 2kb upstream or 200bp downstream of a gene transcription start site (TSS). Intragenic CpGs were defined as CpGs within any other region of a gene, while all remaining CpGs were classified as intergenic. Promoter- and intragenic-DM genes (DMGs) were defined as those with a DM promoter and intragenic CpG, respectively. To determine if DMGs were disproportionately involved in certain biological processes, gene lists were submitted for gene set enrichment analysis (GSEA) using the Panther database [91, 92]. We submitted promoter- and intragenic-DMG lists separately, and a background

list composed of all genes with CpGs covered by our RRBS data was applied. We considered GO terms significantly enriched at $FDR < 0.05$.

6.3.7 Motif Enrichment Analysis

We extracted 100bp genomic sequences centered around EST hyper- and hypomethylated CpGs, as well as 100bp sequences at 10k random CpGs included in the DM analysis to use as control sequences. Analysis of motif enrichment was performed using the MEME suite as described in Section 2.3.7. We identified motifs that were uniquely enriched among hyper- or hypomethylated sequences, as well as their corresponding TFs. The genes encoding these TFs were submitted to GSEA using the PANTHER database to identify enriched GO terms.

6.3.8 Epigenome-Wide Association Study

We performed an EWAS to identify significant associations between CpG methylation rates and HW. CpG sites with a methylation rate variance ≥ 1 across samples with coverage ≥ 10 were retained for further analysis. For each CpG site, a linear model was fitted with methylation rate as a response variable and including the fixed effects of sample mean methylation rate and CO₂ level. Residual methylation rates were extracted and used for subsequent association analyses. For each CpG site, a linear model was fitted with log₂-transformed HW as a response variable, and CO₂ concentration, sample mean methylation rate, and CpG methylation rate as fixed effects. Analysis of variance was performed to assess the effect of the CpG methylation rate on HW, and this process was repeated for all sites. Associations were considered significant at $p < 1E-5$.

6.3.9 Quantitative reverse-transcription PCR

We performed quantitative reverse-transcription PCR (RT-qPCR) on eight DMGs (*TBX1*, *TBX4*, *THRSP*, *PITX2*, *ABLIM2*, *NGB*, *GSC*, and *DAG1*). Total RNA from six samples per EST treatment

was reverse transcribed using Superscript II Reverse Transcriptase (Invitrogen). Custom primers were designed for the eight DMGs and two stably methylated genes (*MEAF6* and *SRRMI*), as well as for *YWHAZ* and *RPL13* which have been cited as suitable reference genes in chicken heart [254]. All assays were performed in singleton on a QuantStudio 5 Real-Time PCR System (Thermo Fisher) using 3.75 µl cDNA template (10ng total), 1.25 µl of forward and reverse primers, and 6.25 µl MESA BLUE qPCR MasterMix (Eurogentec). Cycling conditions were 50°C for 2 min. and 95°C for 10 min., followed by 40 cycles of 95°C for 15 s and 60°C for 1 min., followed by a melt curve stage of 95°C for 15 s and 60°C for 1 min. and a dissociation step of 95°C for 15 s. Delta Cts (dCts) were obtained for each gene per sample by subtracting the test gene Ct from the geometric mean of the reference gene Cts, and analyses of variance were performed to assess the significance of EST on dCts correcting for sample CO₂ concentrations. Fold changes in abundance at 38.9°C relative to 37.8°C were calculated using the 2^{-ddCt} method.

6.4 Results

6.4.1 Broiler chicks differ significantly in HW between ESTs

We measured HW in all broiler chicks six hours post-hatch. Chicks incubated at 38.9°C exhibited significantly lower HW relative to 37.8°C chicks (Figure 6.1; $F=37.53$, $p=3.19E-08$). There was no significant effect of CO₂ concentration ($F=0.254$, $p=0.62$) or EST-CO₂ interaction ($F=0.257$, $p=0.61$) on HW. When correcting for the significantly shorter incubation time of 38.9°C chicks ($F=52.79$, $p=2.18E-10$), the difference in HW between ESTs remained significant ($F=12.75$, $p=6.05E-04$). These results recapitulate the well-established negative association between

incubation EST and HW that has been found in numerous studies [67, 202, 232–234], and provide strong support for the assessment of associated differences in cardiac methylation between ESTs.

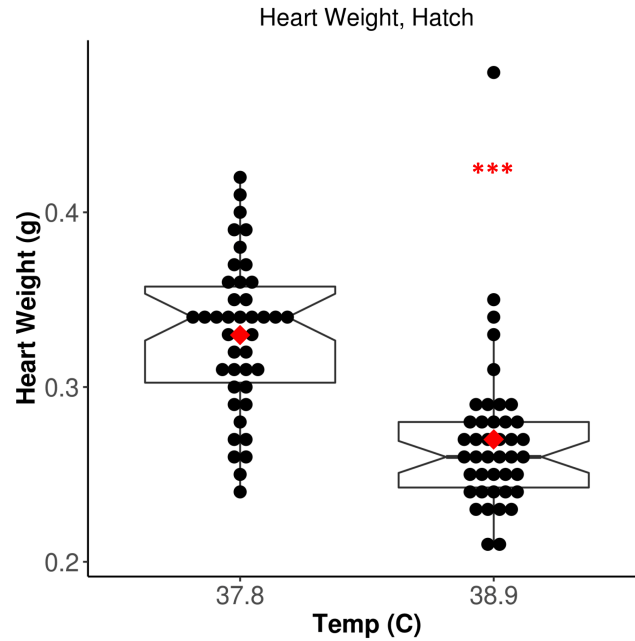


Figure 6.1. Increased egg shell temperature (EST) is associated with significantly lower heart weight. Notched horizontal lines and red dots indicate median and mean values, respectively. *** $p < 0.001$.

6.4.2 RRBS reads capture predominantly lowly methylated regions of the chicken genome

An average of 18.9 million RRBS reads were generated per sample, of which 18.3 million remained after trimming (Table E.1). Mapping of RRBS reads to the *G. gallus* reference genome resulted in an average unique alignment rate of 73.8%, and an average of 3.1 million CpGs covered at a read depth $\geq 10X$. Global CpG methylation rates across all samples were low on average (17.7%), likely due in part to the fact that RRBS targets CpG-rich regions of the genome which are lowly methylated [255]. Average CHG and CHH methylation rates were 0.55% and 0.56%, respectively; given the inherently low prevalence of methylation in non-CpG contexts in vertebrates [24], this suggests high bisulfite conversion efficiency of RRBS reads.

A total of 1,874,588 CpG sites were covered by reads in at least 25 samples per EST treatment, and the majority of these (40.56%) were within gene promoters. 35.18% of CpGs were within gene bodies (25.67% in introns and 12.51% in exons), while 21.26% were intergenic. Promoter CpGs exhibited the lowest methylation rate on average (7.76%), with 85% of sites having an average methylation rate less than 10% (data not shown). The percentage of lowly methylated (<10%) CpGs in exons, introns, and intergenic regions was significantly lower—42.74%, 54.52%, and 51.77%, respectively (data not shown). The low methylation observed at a majority of CpGs indicates that many loci covered by our RRBS data are not dynamically methylated in response to the applied environmental conditions, but remain relatively unmethylated regardless of EST or CO₂ treatment.

6.4.3 EST impacts the methylation state of genes involved in general and heart-specific developmental processes

We identified a total of 2,175 DM CpGs between EST treatments, of which 1,121 were hypermethylated and 1,054 hypomethylated at 38.9°C relative to 37.8°C. Differential methylation was more likely to occur outside of promoter regions; of the 760,332 promoter CpGs tested, only 285 (0.05%) were DM, while 0.14%, 0.18%, and 0.15% of exonic, intronic, and intergenic CpGs, respectively, were DM. In total 269 genes were promoter-DM (promoter-DMGs; 123 hyper- and 146 hypomethylated), while 949 genes were intragenic-DM (intragenic-DMGs; 526 hyper- and 423 hypomethylated).

DMG lists were submitted to Panther for GSEA, and we identified uniquely enriched biological processes among different gene sets (Table 6.2). We did not identify enriched GO terms when considering hyper- and hypomethylated promoter-DMGs separately, however when combining the two groups we identified enrichment of the terms ‘Multicellular Organism Development’, ‘Animal Organ Development’, and ‘Anatomical Structure Development’. Among

promoter-DMGs, several are known to play roles in heart development, including T-box transcription factors 1 and 4 (*TBX1* and *TBX4*) that were hypomethylated at 38.9°C, and two genes involved in thyroid hormone signaling: thyroid hormone receptor alpha (*THRA*) and thyroid hormone responsive (*THRSP*), which were hypo- and hypermethylated, respectively (Figure E.1). As promoter methylation is generally inversely correlated with gene expression, hypomethylation of these gene promoters at 38.9°C is indicative of increased potential for gene activation, while hypermethylation suggests decreased activation.

Intragenic-DMGs were enriched for processes related to cardiac muscle development (Table 6.2). Specifically, hypermethylated intragenic-DMGs were enriched for GO terms such as ‘Heart Development’, ‘Regulation of Myoblast Differentiation’, and ‘Cardiovascular System Development’, while hypomethylated intragenic-DMGs were enriched for terms including ‘Cell Differentiation’, ‘Cell Adhesion’, and ‘Glycosaminoglycan Metabolic Process’. Among genes associated with these GO terms are many involved in cardiomyocyte proliferation and differentiation, including: dystroglycan 1 (*DAG1*, hyper- and hypomethylated; Figure E.1), paired like homeodomain 2 (*PITX2*, hypomethylated; Figure E.1), and Erb-B2 receptor tyrosine kinase 2 (*ERBB2*, hypomethylated) and 4 (*ERBB4*, hyper- and hypomethylated). Associations between intragenic methylation and gene expression are context-specific [25]. Nevertheless, temperature-associated differential methylation within genes involved in heart development demonstrates the potential for corresponding differences in gene expression.

Table 6.1. Enriched GO Terms among DMGs between EST treatments

Promoter-DMGs			
GO Term	No. Genes	Enrichment	P-value
anatomical structure development	86	1.53	1.19E-05
animal organ development	60	1.81	4.22E-06
multicellular organism development	84	1.61	2.06E-06
Intragenic Hypermethylated Genes			
GO Term	No. Genes	Enrichment	P-value
anatomical structure development	179	1.82	1.93E-18
cell differentiation	131	1.95	3.15E-11
cardiovascular system development	29	3.04	2.71E-07
negative regulation of myoblast differentiation	5	10.21	2.37E-04
heart development	22	2.30	4.33E-04
regulation of cardiac muscle tissue development	7	4.82	9.30E-04
Intragenic Hypomethylated Genes			
GO Term	No. Genes	Enrichment	P-value
anatomical structure development	165	1.97	2.32E-21
cell differentiation	111	1.93	1.65E-12
cell adhesion	36	2.53	8.86E-07
positive regulation of transcription	50	2.08	1.20E-06
glycosaminoglycan metabolic process	11	4.64	4.52E-05

6.4.4 EST differential methylation impacts binding sites for TFs involved in cell proliferation

As methylation is known to affect gene regulation through the alteration of TF binding sites, we sought to identify enriched motifs in the vicinity of DM CpGs between ESTs. We submitted 100-bp sequences centered on the 2,175 DM CpGs for motif enrichment analysis, and were specifically interested in identifying motifs that were uniquely enriched among hyper- or hypomethylated sequences. We identified 71 motifs for 49 TFs uniquely enriched among hypermethylated regions, and 120 motifs for 81 TFs enriched among hypomethylated regions (data not shown). Among the most uniquely enriched binding motifs for hypermethylated regions were numerous members of the zinc finger and BTB domain containing (ZBTB) family (ZBTB7C, ZBTB7A, ZBTB7B,

ZBTB49), many of which have been shown to negatively regulate cell proliferation [256, 257]. Among hypermethylated regions, the most uniquely enriched binding motif was for musculin (MSC), which functions as a repressor of the myogenic factor *MYOD* [258].

We submitted the gene IDs for TFs of hyper- and hypomethylated motifs to the PANTHER database for GO enrichment analysis. Uniquely enriched terms among TFs of hypermethylated motifs were not directly related to heart development or function; however, TFs of hypomethylated motifs were uniquely enriched for the GO terms ‘cardiovascular system development’ and ‘positive regulation of cell population proliferation’ (Table 6.2). Genes associated with these terms included several with known roles in positively regulating cardiomyocyte proliferation, including retinoic acid receptor alpha (*RARA*) [259], T-box transcription factor 20 (*TBX20*) [260], lymphoid enhancer binding factor 1 (*LEF1*) [261] and *PITX2*, which was also found to be hypomethylated at 38.9°C. Overall, these results reveal the potential for decreased binding ability (via hypermethylation) of TFs involved in negatively regulating cell proliferation, and increased binding ability (via hypomethylation) of TFs involved in positively regulating cardiomyocyte proliferation at 38.9°C versus 37.8°C.

Table 6.2. Uniquely enriched GO terms among TFs of hypomethylated motifs

GO Term	No. Genes	Enrichment	p-value
embryo development ending in birth or egg hatching	13	5.42	8.08E-07
vasculature development	9	4.77	1.20E-04
cardiovascular system development	9	4.68	1.38E-04
circulatory system development	12	3.79	7.37E-05
positive regulation of cell population proliferation	12	3.58	1.25E-04
cellular response to stress	17	2.67	1.59E-04

Table 6.3. Summary of CO₂ differential methylation analyses

Contrast	DM CpGs	Hyper-methylated	Hypo-methylated	Promoter Hyper/Hypo DMGs	Intragenic Hyper/Hypo DMGs
0.4 vs 0.1%	3652	1951	1701	124/106	524/517
0.8% vs 0.4%	4695	2760	1935	165/156	744/514
0.8% vs 0.1%	4482	2782	1670	176/119	671/457

6.4.5 CO₂ differential methylation impacts genes involved in heart development and response to hypoxia

We identified a total of 9,823 DM CpG between at least one pair of CO₂ treatments. Among these, 3,652 were DM between 0.4% and 0.1% CO₂ chicks, 4,695 between 0.8% and 0.4% CO₂, and 4,482 between 0.8% and 0.1% CO₂ (Table 6.3). In each contrast, the higher CO₂ treatment was associated with a higher proportion of hypermethylated CpGs, especially when comparing 0.8% CO₂ to the other treatments.

DMGs were enriched for more heart- and muscle-specific GO terms when contrasting 0.8% CO₂ to either of the other two CO₂ treatments (Table 6.4). Among hypermethylated intragenic-DMGs for each of the three contrasts, 18 such GO terms were enriched among the 0.8% vs. 0.1% contrast, while only five and six terms were enriched at 0.4% vs. 0.1% and 0.8% vs. 0.4%, respectively. Among hypomethylated intragenic-DMGs, the 0.8% vs. 0.4% CO₂ contrast yielded the greatest number of enriched heart/muscle-related terms (n=23), while the 0.4% vs. 0.1% and 0.8% vs. 0.1% yielded only three and seven, respectively. Intragenic-DMGs at 0.8% CO₂ versus 0.4% and 0.1% included: *PDLIM7*, whose resulting protein regulates valve annulus size and hemostasis [262]; *TGFBI*, which regulates postnatal cardiomyocyte differentiation [263]; and *ILK*, which has been shown to exhibit protective effects against cardiomyopathy [264]. Additionally, we also identified an enrichment of DMGs associated with response to low oxygen only when comparing 0.8% to 0.1% CO₂ (Table 6.4). Hypomethylated intragenic-DMGs were enriched for

Table 6.4. Enriched GO terms among CO₂ DMGs related to heart/muscle development and hypoxia

Intragenic-hypermethylated Genes			
GO Term	0.4 vs. 0.1 CO₂	0.8 vs. 0.4 CO₂	0.8 vs. 0.1 CO₂
positive regulation of striated muscle cell differentiation	2.3E-04		2.5E-04
positive regulation of muscle organ development	6.5E-04		7.6E-04
cardiac ventricle development	5.1E-04		3.5E-04
cardiac chamber development	4.7E-04	5.9E-04	5.2E-04
muscle organ development		3.4E-05	2.9E-05
muscle structure development		1.6E-05	5.2E-06
muscle tissue development		9.7E-04	
heart development		1.8E-04	1.3E-04
cardiovascular system development		5.3E-09	1.7E-07
positive regulation of myotube differentiation			1.2E-03
positive regulation of striated muscle cell differentiation			2.5E-04
positive regulation of muscle cell differentiation			5.7E-05
positive regulation of muscle tissue development			8.3E-04
cardiac chamber morphogenesis			1.3E-03
cardiac muscle tissue development			1.2E-03
Intragenic-hypomethylated Genes			
GO Term	0.4 vs. 0.1 CO₂	0.8 vs. 0.4 CO₂	0.8 vs. 0.1 CO₂
positive regulation of striated muscle tissue development	5.1E-04		
positive regulation of muscle organ development	5.1E-04		
mitral valve development		2.8E-05	
heart valve formation		4.8E-04	
pulmonary valve morphogenesis		1.2E-03	
atrioventricular valve development		5.8E-04	
aortic valve morphogenesis		6.9E-04	
heart valve morphogenesis		1.1E-05	2.0E-04
aortic valve development		1.1E-03	
cardiac atrium development		3.6E-04	
heart valve development		2.5E-05	3.7E-04
ventricular septum development		1.1E-04	
cardiac septum development		1.2E-04	
cardiac ventricle development		4.3E-04	
cardiac muscle tissue development		3.2E-04	
heart morphogenesis		6.0E-04	1.5E-04
heart development		1.5E-04	6.1E-04
cardiac chamber morphogenesis			6.6E-04
cardiovascular system development			2.5E-06
cellular response to oxidative stress			1.0E-03
response to hypoxia			1.1E-04
response to decreased oxygen levels			1.7E-04

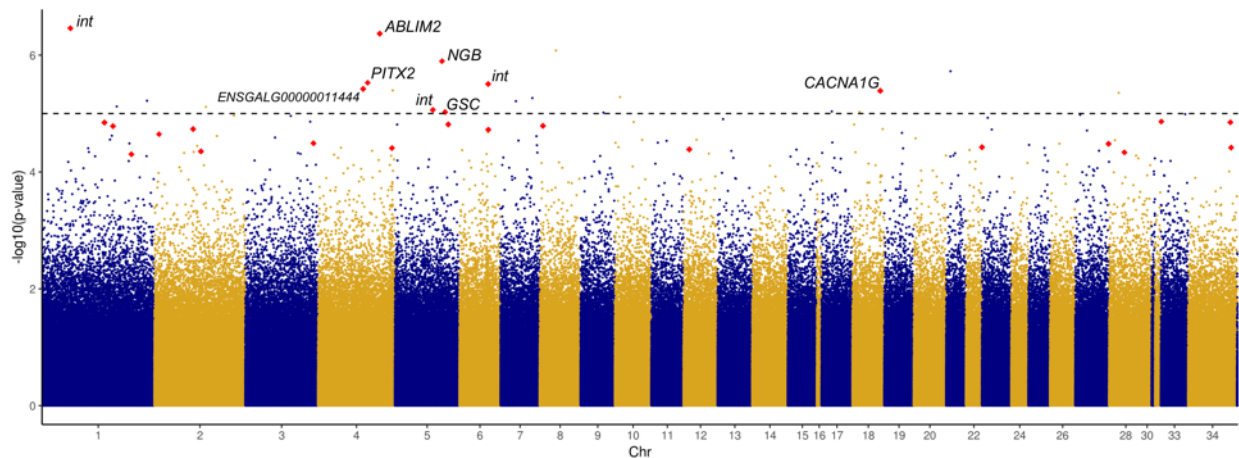


Figure 6.2. Epigenome-wide association study for heart weight. Manhattan plot of $-\log_{10}$ p-values of association between CpG methylation rate and \log_2 -transformed relative heart weight, ordered by chromosome. Red points indicate DM CpGs; significantly associated and DM CpGs are labeled with their corresponding gene ID (int = intergenic).

the GO terms ‘response to decreased oxygen levels’ and ‘response to hypoxia’, and these were not enriched among the DMGs in the other two contrasts.

6.4.6 EWAS identifies methylation signatures significantly associated with HW

Given the significant difference in HW between chicks of different incubation ESTs, we sought to identify CpG sites at which methylation rate was significantly associated with HW. We tested 1,521,346 CpG sites and identified 23 significant associations (Figure 6.2), which represented a modest enrichment in significant p-values compared to a random distribution (data not shown). Three of these CpGs were within gene promoter regions (in *GFII*, *SP9*, and *ST6GALNAC4*), 13 were within gene bodies (exons or introns), and seven were in intergenic regions.

Among the 23 significantly associated CpG sites, nine were also significantly DM between EST treatments, and only one site—the hypomethylated CpG in *PITX2*—had a methylation difference greater than 10%. This site was found to be significantly positively correlated with HW ($r=0.60$, $p=2.97E-06$), along with four other sites: two exonic CpGs in neuroglobin (*NGB*) and goosecoid homeobox (*GSC*), an intronic CpG in calcium voltage-gated channel subunit alpha1 G

(*CACNA1G*), and an intergenic CpG. Four hypermethylated CpGs were negatively associated with HW: an exonic CpG in *ENSGALG00000011444*, an intronic CpG in actin binding LIM protein family member 2 (*ABLIM2*), and two intergenic CpGs.

6.4.7 RT-qPCR identifies genes with coordinated differences in gene methylation and expression between EST treatments

We assessed transcript abundance of eight genes exhibiting differential methylation and/or significant associations with HW via RT-qPCR, to determine if differential methylation was associated with differences in gene expression. We detected quantifiable transcript abundance for all genes except *NGB*, which was undetectable in multiple samples in both EST treatments. Three genes—*ABLIM2*, *PITX2*, and *THRSP*—exhibited significant differences in transcript abundance between EST treatments (Figure 6.3). In all three cases, abundance was higher in the treatment with lower methylation. We observed significantly lower *THRSP* abundance in 38.9°C chicks ($p=0.046$), which is consistent with the promoter hypermethylation observed. Abundance of *PITX2*, which contained a hypomethylated CpG in its last exon, was significantly higher in 38.9°C chicks ($p=7.59E-03$). *ABLIM2*, containing a hypermethylated intronic CpG, exhibited lower abundance at 38.9°C ($p=1.37E-03$). Among the remaining five DMGs, the differences in expression were negatively associated with the differences in methylation for *DAG1*, while the differences were positively associated for *TBX1*, *TBX4*, and *GSC*. We also assessed transcript abundance in two genes that did not exhibit differential methylation (*MEAF6* and *SRRM1*), and did not detect differences in abundance between ESTs. These results have identified candidate genes exhibiting coordinated differences in methylation and expression in the developing chick heart that are associated with incubation EST.

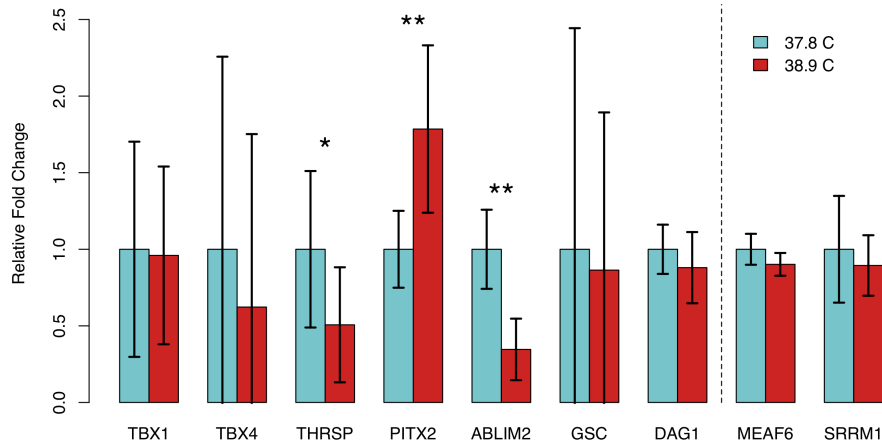


Figure 6.3. Relative fold changes in transcript abundance (38.9°C relative to 37.8°C) for seven expressed genes exhibiting differential methylation and 2 stably methylated genes between temperature treatments. Error bars indicate standard deviation. * $p < 0.05$, ** $p < 0.01$.

6.5 Discussion

To the best of our knowledge, this study represents the largest epigenome-wide analysis in livestock species, the first EWAS in broiler chickens, and the first study assessing genome-wide DNA methylation between chickens incubated in different environments. RRBS was performed in cardiac tissue collected from broiler chicks in a study that recapitulated the established negative correlation between incubation EST and HW. We observed significantly lower HW in chicks incubated at 38.9°C versus 37.8°C, which was independent of differences in incubation length. The heart appears to be uniquely impacted by incubation EST, a finding that has been observed in numerous studies showing up to 30% lower HW in 38.9°C chicks with no significant differences in organs such as the liver, intestines or stomach [67, 234, 238, 265].

Using RRBS, we assessed methylation at over 3 million CpGs across 84 heart samples. Global methylation rates in these samples were low (13.8-20.4%), due to the fact that a majority of sites exhibited methylation rates less than 10%. RRBS reads are disproportionately derived from CpG dense regions of the genome, which are known to be less methylated than the genome at large

[255]. However, RRBS data from eight pig tissues revealed an average methylation rate of 41% [42]. Thus the low methylation observed in our data may be specific to the species, tissue, or stage of development. To date, whole-genome bisulfite sequencing studies in *G. gallus* and other avian species have reported global CpG methylation rates between 50% and 65% [248–251, 266], which is lower than the 60% to 80% that has been reported in mammalian studies. It is possible that the low levels of CpG methylation detected in our data can be attributed to lower CpG methylation rates in avian genomes.

We identified thousands of DM CpGs between ESTs, and annotation of DMGs revealed an enrichment for general and heart-specific developmental processes. While promoter-DMGs were only significantly enriched for processes related to general development, many of these genes are known to be involved in heart development. Among these were two T-box transcription factors, *TBX1* and *TBX4*, that were promoter-hypomethylated at 38.9°C. *TBX1* has been shown to promote cell proliferation and inhibit differentiation of heart cells [267, 268]. *TBX4* has been shown to be expressed in the developing heart, and has been hypothesized to play a role in regulating tissue-specific gene expression [269]. Also among promoter-DMGs were two genes involved in response to thyroid hormone (*THRA*, hypomethylated and *THRSP*, hypermethylated at 38.9°C), which has previously been hypothesized to mediate temperature-driven differences in heart growth. *THRA* is the major thyroid hormone receptor expressed in the heart [270], and *THRA* mutations in zebrafish have been linked to defects in heart development [271]. *THRSP* is expressed in chicken cardiac tissue [272], and research has shown that overexpression of *THRSP* results in decreased levels of genes involved in heart development [273]. Intragenic-DMGs were enriched for GO terms related to heart development. Among these, *DAG1* possessed hypo- and hypermethylated CpGs, and encodes a glycoprotein that has been shown to play an important role in the inhibition of

cardiomyocyte proliferation in mice [274]. *PITX2* contained an exonic CpG that was significantly hypomethylated at 38.9°C. This gene encodes a homeobox TF with roles in heart development that have been reviewed extensively [275]; of note, it has been shown to promote cardiomyocyte proliferation in mice [276]. *ERBB2* and *ERBB4* encode receptor kinases that have been shown to promote cardiomyocyte proliferation and survival [277]. The observed differential epigenetic regulation of genes involved in cardiomyocyte proliferation and differentiation may be contributing to observed differences in HW between chicks of different incubation ESTs.

Furthermore, assessment of enriched binding motifs among DM loci provides evidence that temperature-differential methylation is impacting the developmental trajectories of the chick heart. Methylation is known to reduce the binding potential of TFs by altering binding motif sequences [33]. At 38.9°C, hypomethylated binding motifs—which would be predicted to be more accessible due to lower methylation levels—were enriched for TFs involved in positive regulation of cardiomyocyte proliferation. Conversely, the binding sites for TFs involved in negative regulation of cell proliferation—namely TFs in the ZBTB family—were the most enriched motifs among hypermethylated regions, which would be expected to be less accessible. The more favorable binding of pro-proliferation TFs in the cardiac methylome of chicks at higher temperatures suggests that observed differences in HW may be associated with differences in rates of differentiation and maturation.

We also identified thousands of DM CpGs between CO₂ treatments. Increasing CO₂ concentrations above 0.8% during late stages of incubation have benefits in commercial settings such as reduced hatching time variability [278]. However, continuous incubation in a hypercapnic environment has been shown to increase ascites incidence later in life [240]. Not only did differential methylation at 0.8% CO₂ relative to the lower treatments disproportionately affect

genes involved in heart development, but contrasting 0.8% with 0.1% CO₂ revealed differential methylation of genes involved in hypoxia response. These included: *ETSI*, a TF that has been shown to activate hypoxia-inducible genes [279]; *SRC*, encoding a tyrosine kinase that is upregulated in rat cardiomyocytes by hypoxia to activate MAPK signaling pathways [280]; and *DNM2*, encoding a GTPase and actin-binding protein that has been shown to be downregulated in cardiomyocytes subject to hypoxic conditions [281]. Hypoxia-inducible genes being specifically overrepresented when comparing extremes for CO₂ indicate a cardiac response to differences in gas exchange via differences in gene regulation. Incubation CO₂ level was not associated with differences in HW in this study, which has been shown previously [234]. However, CO₂ concentration may have observable effects on physiological processes that were not assessed here—including angiogenesis—that may contribute to adverse conditions later in life including increased ascites risk.

Our EWAS revealed a small but significant enrichment of association p-values between CpG methylation and HW, suggesting a relationship between gene regulation in the heart and its size at hatch. A DM CpG in an exon of *PITX2* was also the most significantly positively correlated with HW. We identified additional significant CpGs in our EWAS that were also DM, albeit at a difference less than 10%. An intragenic CpG in *ABLIM2* was significantly hypermethylated at 38.9°C and negatively correlated with HW; this gene encodes a actin binding protein that has been shown to be highly expressed in striated muscle and is thought to regulate cytoskeletal signaling pathways [282]. CpGs in these genes, along with those identified in other genes and in intergenic regions, are strong candidates for identifying epigenetic-mediated associations between temperature and heart development.

We identified coordinated changes in methylation and expression for three of the eight tested genes: *ABLIM2*, *PITX2*, and *THRSP*. In all three cases, direction of the difference in gene expression was opposite that of the difference in methylation between ESTs. This inverse relationship was expected at *THRSP*, since the DM CpG was immediately upstream of the TSS in the likely promoter region. Intragenic methylation in *ABLIM2* and *PITX2* were also negatively associated with expression, and this is in contrast to findings in other animal studies that the majority of CpGs in such regions are positively correlated with gene expression. Cases of intragenic methylation being negatively associated with gene expression have been reported by ENCODE [25], and they found these CpGs to be enriched in intragenic enhancers. The presence of CpGs in *ABLIM2* and *PITX2* at which methylation rate is inversely correlated with gene expression suggests that these CpGs may be located within gene regulatory elements. To explore this idea, we searched for binding motifs in the genomic regions flanking in DM CpGs in *ABLIM2* and *PITX2*. The *PITX2* DM regions had strong sequence similarity to the Krueppel Like Factor 5 (*KLF5*) binding motif, and this TF has been shown to be upregulated during heat stress in chickens previously [283]. Additionally, the *ABLIM2* DM regions showed strong sequence similarity to the binding motif of Achaete-Scute Family BHLH Transcription Factor 2 (*ASCL2*), which is a known inhibitor of myogenic differentiation [284]. While differential expression was not observed for five of the eight tested DMGs, this could be due in part to the limited number of samples that were assayed via qPCR (n=6/trt) relative to the number assayed via RRBS (n=42/trt).

Among these three DM and differentially expressed genes, *PITX2* has the most well-established role in heart development, specifically in promoting cardiomyocyte proliferation. In this study, increased EST was associated with both decreased methylation at an exonic CpG in *PITX2* as well as increased *PITX2* transcript abundance. As *PITX2* is known to promote

cardiomyocyte proliferation, this suggests an increased proliferative capacity in the hearts of 38.9°C chicks. This conflicts with results reported in Romanoff 1960, in which increased incubation temperature was associated with a reduced mitotic index from E8 to hatch [235]. However, our results indicate that proliferation in high EST chicks may be increased relative to lower EST chicks as a partial compensatory mechanism immediately following hatch, and that developmental trajectories of hearts at different ESTs continue to be affected even after incubation.

In conclusion, this study has identified differential methylation patterns in the post-hatch chick heart associated with differences in incubation EST and CO₂ level, and such differences may be impacting heart growth and development through associated changes in TF binding and gene expression. Future studies should seek to further assess differences in methylation between incubation treatments at later stages of post-hatch life. Additionally, characterizing epigenetic patterns in additional organs responsible for regulating body temperature, including the hypothalamus and thyroid gland, may help in characterizing the mechanisms regulating temperature-driven differences in heart development. Knowledge of such epigenetic signatures influenced by early-life environment may benefit animal breeding by serving as predictors of future animal performance.

6.6 Acknowledgments

This study was financially supported by the Dutch Ministry of Economic Affairs (TKI-Toeslag) and the Breed4Food partners Cobb Europe, CRV, Hendrix Genetics and Topigs Norsvin. We thank the Breed4Food EWAS committee for useful discussion during the project and we thank Bert Dibbits for technical laboratory support. RJC was funded under a USDA NIFA NNF Training Grant (#2015-38420-23697).

CHAPTER 7

CONCLUSIONS & FUTURE DIRECTIONS

The results presented in chapters 2 through 6 of this dissertation represent the most expansive survey of whole-genome DNA methylation in farm animal genomes to date. DNA methylation is the most prevalent epigenetic modification made to DNA molecule, and acts in concert with other chromatin modifiers to establish cell-, stage-, and environment-specific gene expression. I have demonstrated that state-specific DNA methylation patterns in farm animal species exhibit similar characteristics to those reported in other mammals in terms of global prevalence and distribution. Furthermore, we demonstrate that variation in DNA methylation between cell types and time points is strongly associated with differential gene expression as well as transcription factor binding motifs, revealing putative regulatory and transcriptional consequences of differentially methylated regions (DMRs). These data sets will prove valuable to future Functional Annotation of Animal Genomes (FAANG) consortium efforts to utilize epigenomics data to enhance discovery of regulatory elements influencing complex trait variation [17]. To achieve this end, raw bisulfite sequencing data will be publicly available for use by other scientists, and DNA methylation data in different tissues and cell type will be available for viewing in appropriate databases such as the University of California Santa Cruz (UCSC) Genome Browser.

In Chapter 2, I have defined differentially methylated regions in sorted porcine immune cells that represent likely regulatory regions governing immune cell state and function. We associated a large proportion of these regions with either 1) local gene expression or 2) regulatory potential of immune-related transcription factors, including in pig-enriched T cell subpopulations for which similar data in other mammals is currently unavailable. DMRs were also enriched among genetic loci associated with immune capacity traits, demonstrating their potential use in

characterizing mechanistic links between genetic variation and immune system phenotypes. Fluorescence activated cell-sorting of peripheral blood mononuclear cells allowed for assessment of DNA methylation in distinct leukocyte subpopulations. However, it is known that there is extensive heterogeneity even within the sorted cell types sequenced in this study, as evidenced by single-cell RNA-sequencing (scRNA-seq) performed in PBMCs from the same animals [81]. Therefore, results still reflect composite epigenomic profiles that may be fine-tuned in the future by additional single-cell sequencing assays.

Many current functional annotation projects in farm animal species seek to define distinct chromatin states across a diverse set of tissues and cell types, with particular focus on those systems known to contribute to phenotypes of economic interest [18]. Thus, similar analyses as performed in Chapter 2 to define DNA methylation patterns across the peripheral immune system should be performed in other pig organ systems, including the digestive and reproductive systems. DNA methylation is also known to interact with other epigenetic modifications to modify locus-specific regulatory activity, emphasizing the need to integrate epigenomics assays to better define local and distal-acting *cis* elements [27]. For example, preliminary data in adult porcine skeletal muscle and digestive tissues demonstrated that tissue-specific lowly methylated regions (LMRs) disproportionately overlapped regions displaying activating histone modifications (e.g., H3K27ac, H3K4me3) and were mutually exclusive from repressive histone modifications (H3K27me3) (Pan et al., personal communication). Generation of similar histone modification data in porcine immune cells would lead to increased efficacy of regulatory element prediction through the use of evidence from multiple epigenomics data sets. All functional assays performed in porcine immune cells and other organ systems has the potential to enhance the search for causative variants influencing complex traits. The use of functional biological data has proven useful in the

refinement of previously identified QTL in livestock species [285, 286]. Furthermore, functional data generated in the pig has been utilized in GWAS studies to weight SNPs in annotated regions more highly than unannotated SNPs, and resulted in the identification of novel loci associated with feed efficiency traits [287, 288]. DNA methylation data generated in this chapter may prove useful in performing similar analyses of immune system traits.

In Chapters 3 and 4, we identified hundreds of thousands of DMRs between different developmental stages, representing putative sites of stage-specific gene regulation in pig fetal organs. We chose to assess DNA methylation at fetal ages representing critical periods of organ development; for example, 41 and 70 days gestation represent midpoints of primary and secondary pig skeletal muscle myogenesis, respectively. However, it is likely that there is extensive variation in gene regulation—and thus DNA methylation—within the interval between surveyed stages. It would be beneficial to collect epigenomics data at intermediate time points to better describe temporal specificity of regulatory element activation. Pig prenatal physiology is known to have impacts on growth and meat quality traits [45]. I show in Chapter 3 that areas that lose methylation during skeletal muscle development—and thus may be more active in gene regulation—are enriched for GWAS SNPs for muscle and meat-related traits measured in adult pigs, suggesting that these regions may play important roles in regulating phenotype expression early in life. The use of epigenomics data from fetal timepoints may thus prove useful in determining any associations between observed chromatin state and later-life animal performance.

In Chapter 4, I also report extensive allele-biased methylation (ABM) in pig fetal tissues, representing the first report of allele-specific chromatin events in the pig. Fetuses were derived from crosses of highly divergent pig breeds—European White Composite and Chinese Meishan—that allowed for identification of extensive breed-specific ABM that is indicative of distinct gene

regulatory landscapes. We identified candidate genes exhibiting both ABM and allele-biased expression between breed alleles, including genes such as methionine synthase reductase (*MTRR*) that has been shown to regulate diverse developmental phenotypes in other mammals. Future research characterizing the impacts of breed-specific expression of genes identified in this study may elucidate regulatory pathways driving behavioral, growth, and reproductive differences between European and Chinese pig breeds. Furthermore, assessment of molecular biases between breed alleles in other pig populations may provide additional insight into regulatory differences associated with pig genetic background. I have also reported thousands of regions exhibiting imprinting-like biases in methylation, representing known and novel imprinted genes and control regions. These candidate genes will also require extensive validation (e.g., by pyrosequencing), particularly those genes that have not been reported as imprinted in other species. Additionally, our analyses likely only captured a fraction of imprinting-like ABM events due to our limited SNP coverage of the porcine genome (~15%). The use of different pig populations or different methodologies may be beneficial in further discovery of candidate imprinted loci.

In Chapters 5 and 6, I have defined DMRs associated with pig weaning and broiler chick incubation parameters, respectively, thus representing sites of epigenetic alteration in response to environmental condition. In chapter 5, I observed extensive differential methylation of piglet peripheral blood mononuclear cells associated with acute weaning stress, particularly among T cell receptor genes that is in agreement with previously-observed diminished response of T cells following weaning [208]. In chapter 6, I demonstrate that broiler chick incubation temperature and CO₂ concentration is associated with differential methylation of genes involved in heart development and hypoxia response, respectively, and furthermore identified candidate genes that may mediate temperature-driven differences in heart morphology. The data presented in chapter 6

represent the largest epigenome-wide associated study to date in farm animal species, and underlines the potential for population-wide assessment of DNA methylation to identify epigenomic signatures associated with complex trait variation.

Chapters 5 and 6 both identified DNA methylation differences associated with short-term response to environmental stimuli—piglet PBMCs were collected 24 hours post-weaning, and broiler chick heart tissue was collected 6 hours after hatch from different incubation settings. We therefore did not measure any potential long-term consequences of these stimuli by quantifying DNA methylation at later time points. However, weaning stress and incubation parameters have been shown to impact adult pig and broiler performance, respectively [289, 290], and short-term differences in DNA methylation observed here indicate the potential for persistent changes in response to environment via this mechanism. Additionally, organ systems beyond those surveyed in chapters 5 and 6 are likely similarly impacted by applied treatments and may even drive observed physiological and molecular differences in peripheral tissues. For example, psychosocial stress in pigs and incubation temperature in broiler chicks have been shown to alter gene expression and histone modification patterns in neuronal and endocrine tissues [69, 209]. Assessment of environment-driven epigenetic changes in a larger set of tissues will provide a more comprehensive understanding of physiological pathways impacted by various external stimuli. Knowledge of the mechanisms mediating long-term impacts of early life-environment may aid in future selection practices for resilience phenotypes.

In conclusion, the results presented in this dissertation have provided new insights into spatiotemporal DNA methylation patterns associated with cell identity, fetal development, and environment in farm animal species. The data generated by these projects will serve as a resource for the FAANG consortium to combine with other transcriptomic and epigenomics data sets to

map functional elements in the porcine genome. Enhanced functional annotation of porcine and avian genomes using epigenomics data will provide novel mechanistic frameworks that may be leveraged for continued genetic improvement of these species for agricultural and biomedical applications.

APPENDICES

APPENDIX A

CHAPTER 2 SUPPLEMENTARY MATERIAL

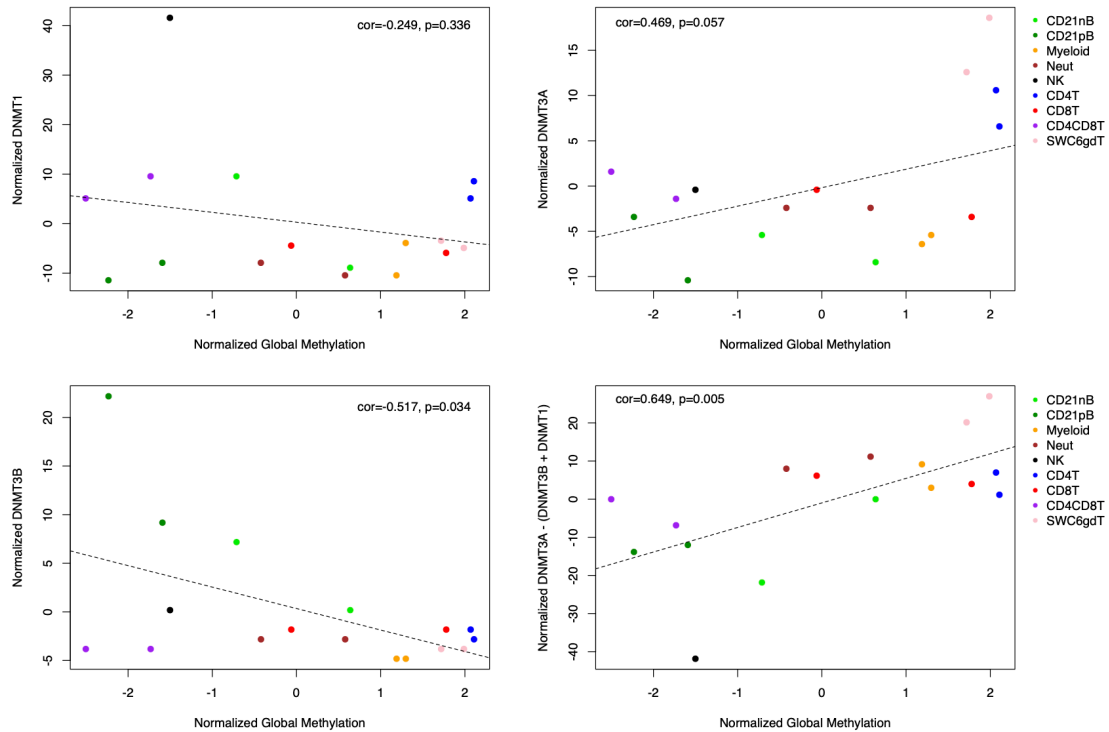


Figure A.1. Associations between immune cell global methylation and DNA methyltransferase (DNMT) expression. Moderate negative and positive correlations were observed between global methylation and normalized transcript abundance of *DNMT1* (A) and *DNMT3A* (B), respectively. Normalized *DNMT3B* abundance was significantly negatively correlated with global DNA methylation (C). Considering abundance of all DNMTs, the normalized abundances of *DNMT1* and *DNMT3B* subtracted from *DNMT3A* abundance was significantly positively correlated with global methylation (D).

Table A.1. Summary of identified cell lowly methylated regions and overlapping genes

Cell Type	No. cLMRs	No. Unique cLMRs	No. Genes	No. Unique Genes
CD21nB	1196	174 (14.5%)	860	42 (4.9%)
CD21pB	13701	9398 (68.6%)	5327	2201 (41.3%)
Myeloid	7959	4640 (58.3%)	3975	1243 (31.8%)
Neut	2837	666 (23.5%)	1880	194 (10.3%)
NK	4894	2410 (49.2%)	2268	476 (21.0%)
CD4T	1785	450 (25.2%)	1057	96 (9.1%)
CD8T	1493	375 (25.1%)	901	64 (7.1%)
CD4CD8T	7873	3598 (45.7%)	3412	815 (23.9%)
SWC6gdT	1655	907 (54.8%)	1061	183 (17.24%)

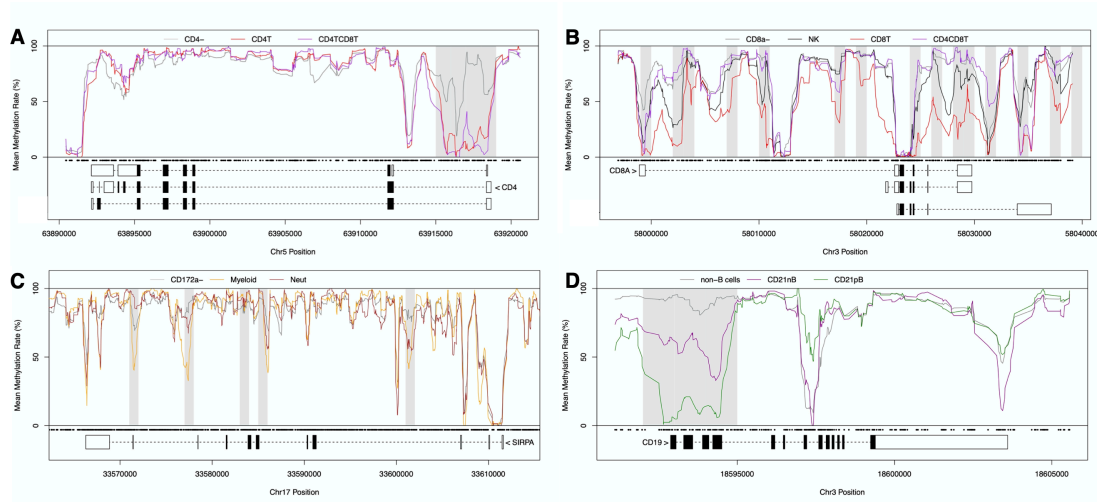


Figure A.2. Methylation rates across immune cell marker genes (A) *CD4*, (B) *CD8A*, (C) *SIRPA*, and (D) *CD19* for cell types positive and negative for each co-receptor. Gray boxes indicate cell differentially methylated regions (cDMRs) within each gene locus, and black dots indicate CpG coordinates.

Table A.2. Number of expression-enriched genes for each immune cell type

Cell Type	No. cLMRs
CD21nB	1196
CD21pB	13701
Myeloid	7959
Neut	2837
NK	4894
CD4T	1785
CD8T	1493
CD4CD8T	7873
SWC6gdT	1655

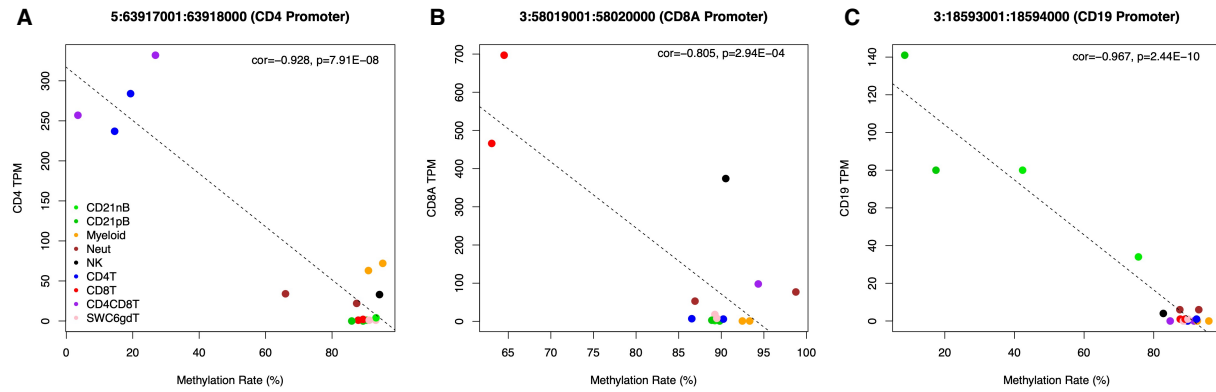


Figure A.3. Correlation dot plots of transcript abundance versus promoter expression of marker genes. (A) *CD4*, (B) *CD8A*, and (C) *CD19*.

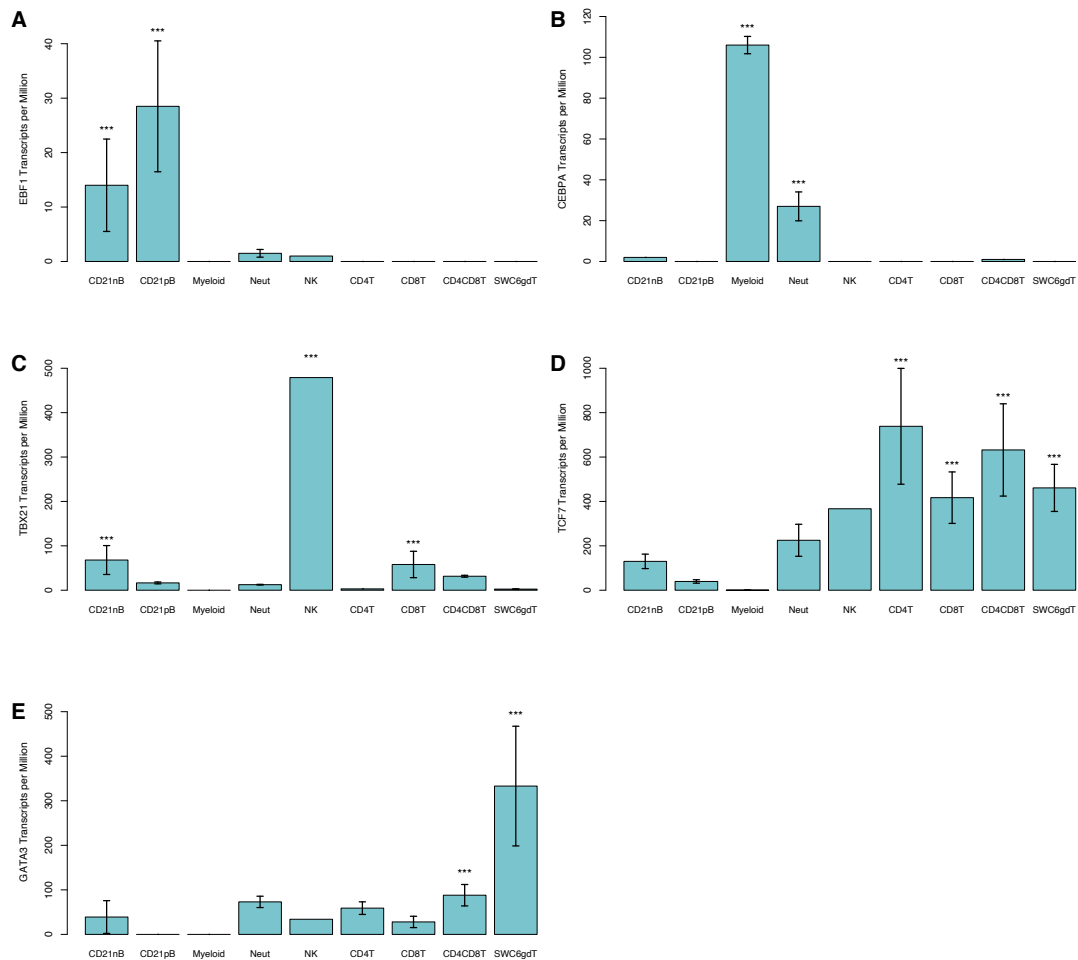


Figure A.4. Transcript abundance of five transcription factors with cLMR-enriched binding motifs. *EBF1* (A), *CEBPA* (B), *TBX21* (C), *TCF7* (D), and *GATA3* (E).

*** = Significantly enriched gene expression.

APPENDIX B

CHAPTER 3 SUPPLEMENTARY MATERIAL

Table B.1. Fetal *longissimus dorsi* WGBS summary statistics

Sample	No. Raw Reads	Unique Alignment Rate (%)	Bisulfite Conversion Rate (%)	Global CpG Methylation Rate (%)
41dg_1	167,861,798	89.1	99.2	76.2
41dg_2	146,709,279	87.7	99.3	76.23
41dg_3	163,976,885	89.4	99.2	76.08
70dg_1	163,646,699	83.6	99.3	73.89
70dg_2	161,096,923	87.5	99.3	74.24
70dg_3	160,140,859	88.0	99.2	74.52

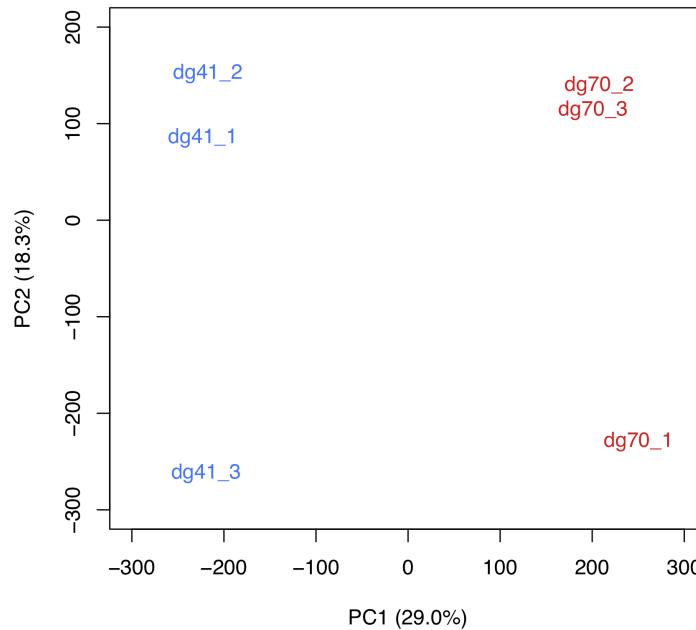


Figure B.1. Principal component analysis plot of fetal *longissimus dorsi* muscle samples by global CpG methylation rates.

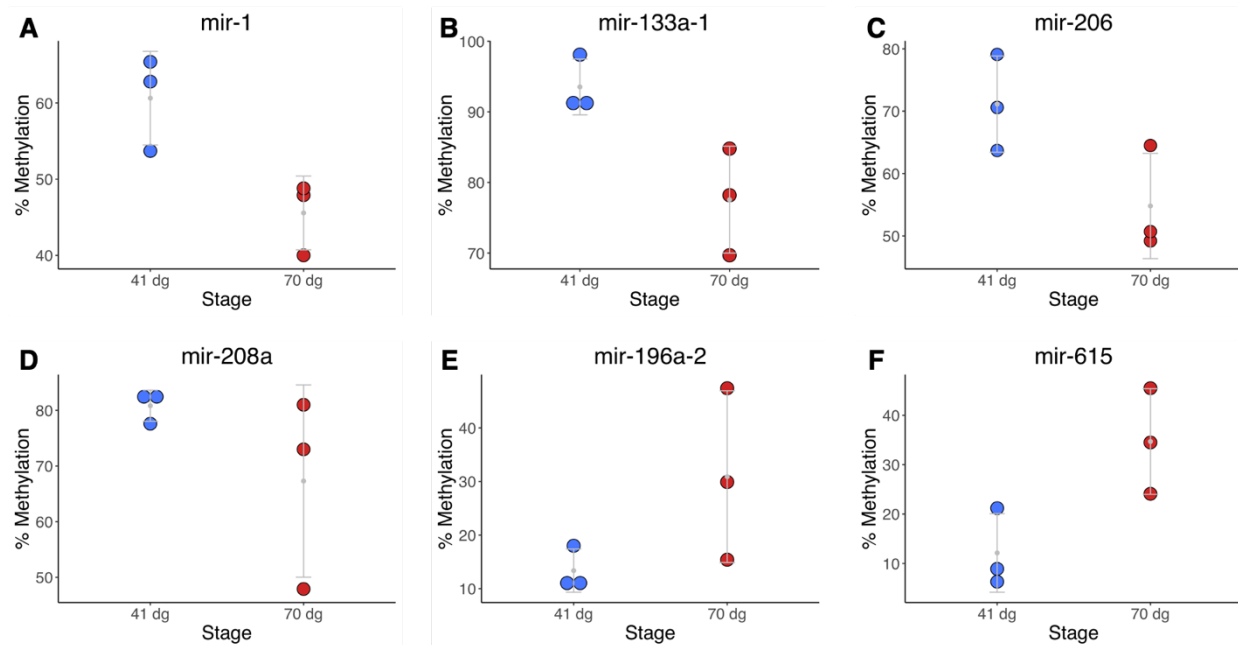


Figure B.2. Differential methylation of miRNAs in developing fetal skeletal muscle. (A-D) Methylation dot plots for hypomethylated regions proximal to mir-1, mir-133a-1, mir-206, and mir-208a. **(E-F)** Methylation dot plots for hypermethylated regions proximal to mir-196a-2 and mir-615. Gray dots and lines indicate mean and standard deviation, respectively.

Table B.2. RNA-seq and smRNA-seq summary statistics

Sample	RNA-seq		smRNA-seq	
	No. Reads	Unique Alignment Rate (%)	No. Reads	Unique Alignment Rate (%)
41dg_1	41,793,915	92.1	63,063,427	90.6
41dg_2	41,239,214	93.1	69,077,421	90.7
41dg_3	52,383,096	93.2	40,699,087	81.4
70dg_1	57,796,034	93.0	53,306,905	90.6
70dg_2	58,798,551	94.0	66,421,476	90.7
70dg_3	37,142,338	93.1	36,303,827	91.4

APPENDIX C

CHAPTER 4 SUPPLEMENTARY MATERIAL

Table C.1. Fetal tissue whole-genome bisulfite sequencing library summary statistics

Tissue, Stage	No. Raw Reads (Range)	Unique Mapping Rate (%)	Bisulfite Conversion Efficiency (Avg, %)	Average Methylation Rate (%)		
				CpG	CHG	CHH
Brain, 30 dg	181-254M	84.4-84.8	99.1	79.9	1.4	1.5
Brain, 70 dg	185-221M	84.2-86.5	99.1	76.9	1.5	1.6
Liver, 30 dg	214-246M	82.8-83.4	99.5	62.7	1.0	0.9
Liver, 70 dg	189-221M	81.8-82.4	99.5	65.2	0.9	0.9
Muscle, 30 dg	198-223M	85.0-85.9	99.1	78.0	1.5	1.7
Muscle, 70 dg	176-216M	84.8-85.7	99.1	73.9	1.5	1.6
Placenta, 30 dg	186-244M	86.6-87.4	99.1	62.8	1.4	1.5
Placenta, 70 dg	162-212M	85.1-86.5	99.1	60.1	1.6	1.7

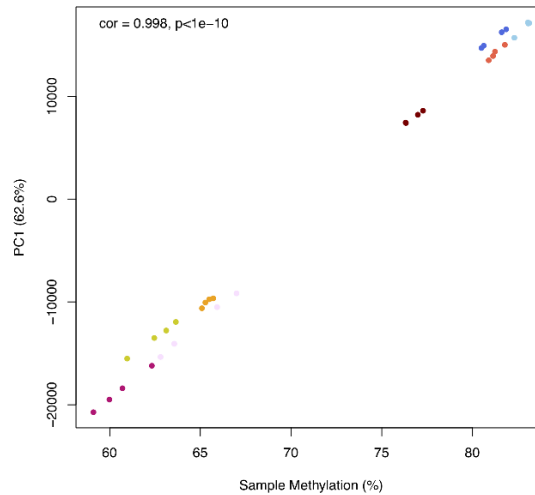


Figure C.1 Correlation between first principal component eigenvalues and global methylation rates in pig fetal tissues.

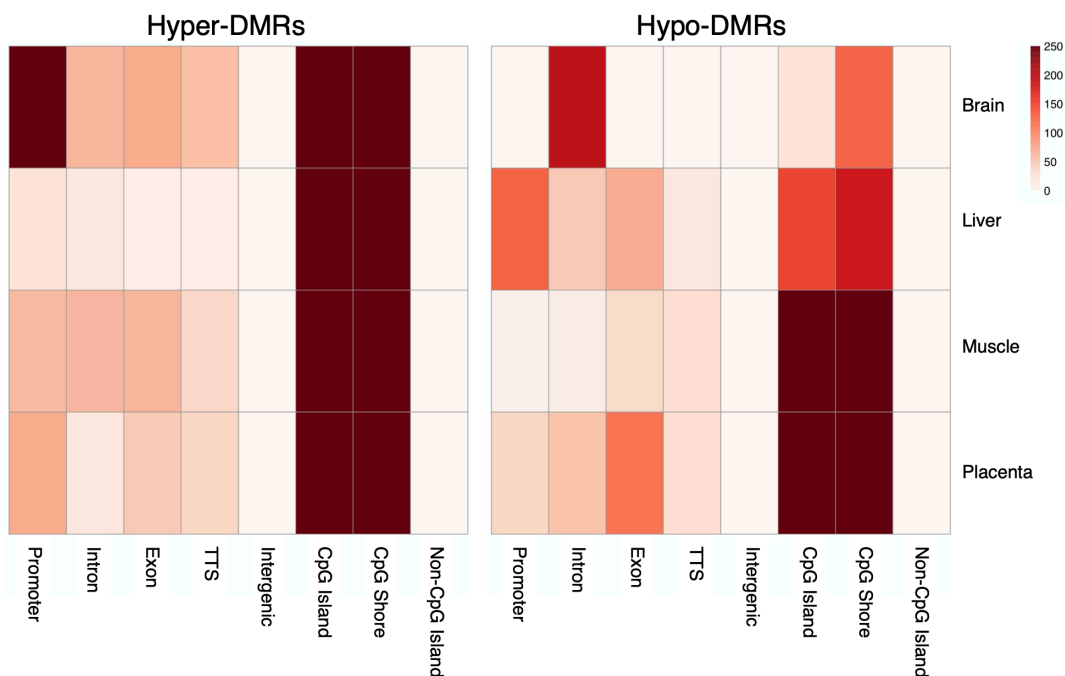


Figure C.2. Normalized heatmap of DMR enrichment in gene and CpG features.

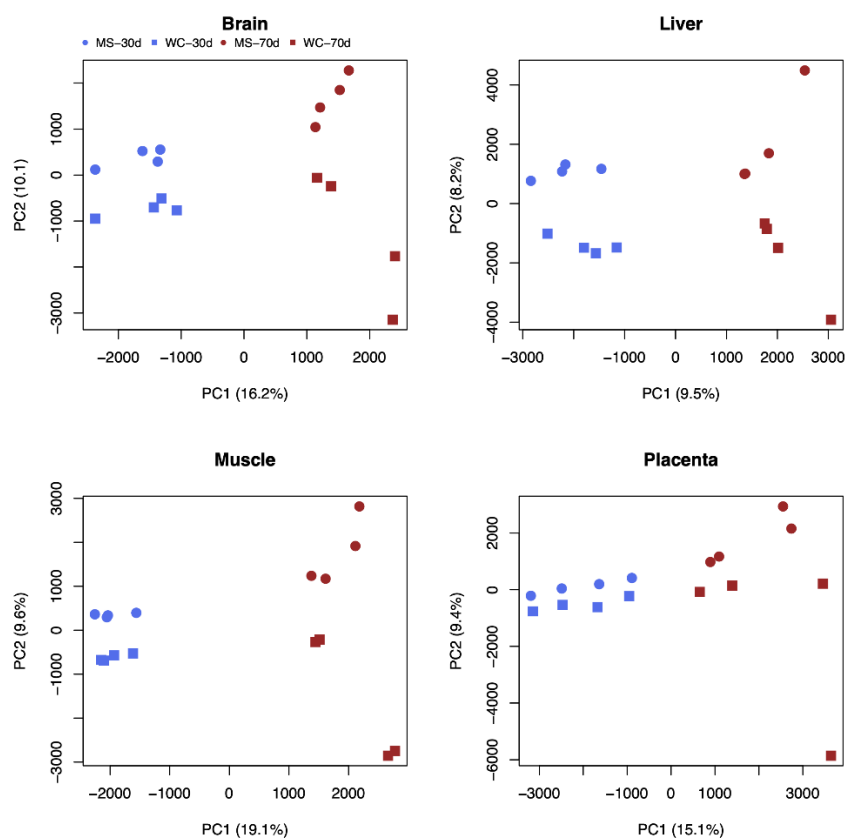


Figure C.3. Principal component analysis (PCA) plots of allele methylation rates by tissue.

Table C.2. Summary of allele-specific sorting of WGBS reads

Sample	% Reads assigned to Meishan allele	% Reads assigned to White allele	% Reads unassigned	% Reads conflicting
Brain_30dg_1	14.29	14.5	70.4	0.82
Liver_30dg_1	16.13	16.39	66.54	0.93
Muscle_30dg_1	14.59	14.79	69.82	0.8
Placenta_30dg_1	14.4	14.74	70.02	0.83
Brain_30dg_2	14.66	14.82	69.67	0.86
Liver_30dg_2	15.92	16.11	67.03	0.94
Muscle_30dg_2	14.62	14.76	69.77	0.85
Placenta_30dg_2	14.65	14.92	69.55	0.88
Brain_30dg_3	12.49	12.6	74.19	0.72
Liver_30dg_3	13.98	14.14	71.01	0.87
Muscle_30dg_3	12.45	12.57	74.22	0.76
Placenta_30dg_3	12.8	12.58	73.88	0.74
Brain_30dg_4	12.79	12.91	73.47	0.83
Liver_30dg_4	14.45	14.59	69.96	0.99
Muscle_30dg_4	12.81	12.92	73.44	0.83
Placenta_30dg_4	13.27	13.1	72.74	0.89
Brain_70dg_1	13.31	13.5	72.25	0.94
Liver_70dg_1	14.61	14.79	69.55	1.04
Muscle_70dg_1	13.39	13.49	72.17	0.95
Placenta_70dg_1	13.95	13.41	71.56	1.08
Brain_70dg_2	13.95	11.17	72.95	1.92
Liver_70dg_2	14.48	14.64	69.75	1.13
Muscle_70dg_2	14.19	11.36	72.5	1.95
Placenta_70dg_2	14.34	11.49	72.24	1.93
Brain_70dg_3	14.58	14.72	69.74	0.95
Liver_70dg_3	16.18	16.36	66.36	1.11
Muscle_70dg_3	14.76	14.88	69.46	0.9
Placenta_70dg_3	14.11	14.42	70.59	0.89
Brain_70dg_4	14.88	13.66	68.98	2.49
Liver_70dg_4	16.26	16.52	66.12	1.1
Muscle_70dg_4	14.76	13.51	69.37	2.36
Placenta_70dg_4	14.55	13.23	69.84	2.38

APPENDIX D

CHAPTER 5 SUPPLEMENTARY MATERIAL

Table D.1. PBMC whole-genome bisulfite sequencing summary statistics

Sample ID	No. Raw Reads	% Uniquely Mapped	Bisulfite Conversion Efficiency (%)	Global CpG Methylation Rate (%)
98-7_Pre	171,913,620	88.08	99.5	80.80
98-7_Post	148,621,253	88.25	99.5	81.10
98-8_Pre	129,793,223	85.95	99.6	79.12
98-8_Post	146,807,971	89.19	99.6	80.13
98-19_Pre	160,226,014	89.16	99.6	80.60
98-19_Post	140,287,192	88.75	99.6	80.97
102-1_Pre	136,042,171	88.20	99.7	80.10
102-1_Post	128,217,018	88.77	99.6	81.02
102-7_Pre	173,164,750	87.64	99.4	80.60
102-7_Post	154,119,196	86.76	99.4	80.30
102-9_Pre	146,145,767	88.97	99.6	82.85
102-9_Post	194,528,096	89.28	99.6	79.93

Table D.2. PBMC RNA-sequencing read summary

Sample ID	No. Raw Reads	% Trimmed	% Uniquely Aligned
98-8_Pre	108,119,016	99.98	51.97
98-8_Post	69,009,261	99.99	52.72
98-19_Pre	85,203,141	99.99	52.86
98-19_Post	95,056,244	99.99	49.27
102-1_Pre	75,870,039	99.99	55.75
102-1_Post	95,187,993	99.98	52.33
102-9_Pre	95,458,301	99.99	51.13
102-9_Post	75,575,119	99.99	52.21
98-7_Pre	64,258,319	99.97	64.90
98-7_Post	72,477,865	99.99	67.38
98-9_Pre	75,632,223	99.98	63.60
98-9_Post	61,798,005	99.98	62.72
102-6_Pre	62,488,241	99.98	66.59
102-6_Post	65,092,215	99.97	60.10
102-7_Pre	72,732,496	99.99	54.00
102-7_Post	64,944,140	99.98	64.06
102-8_Pre	107,578,353	99.92	69.27
102-8_Post	62,706,578	99.99	64.57

APPENDIX E

CHAPTER 6 SUPPLEMENTARY MATERIAL

Table E.1. Chick heart reduced representation bisulfite sequencing summary statistics

	No. Raw Reads	No. Trimmed Reads	Unique Mapping Rate (%)	CpG Meth Rate (%)	CHG Meth Rate (%)	CHH Meth Rate (%)	CpGs Covered
Mean	18.9M	18.3	73.8	17.7	0.55	0.56	3.1M
Range	10.1-34.1M	9.6-32.8M	77.1-69.9	13.8-20.4	0.48-0.78	0.49-0.77	2.3-3.5M

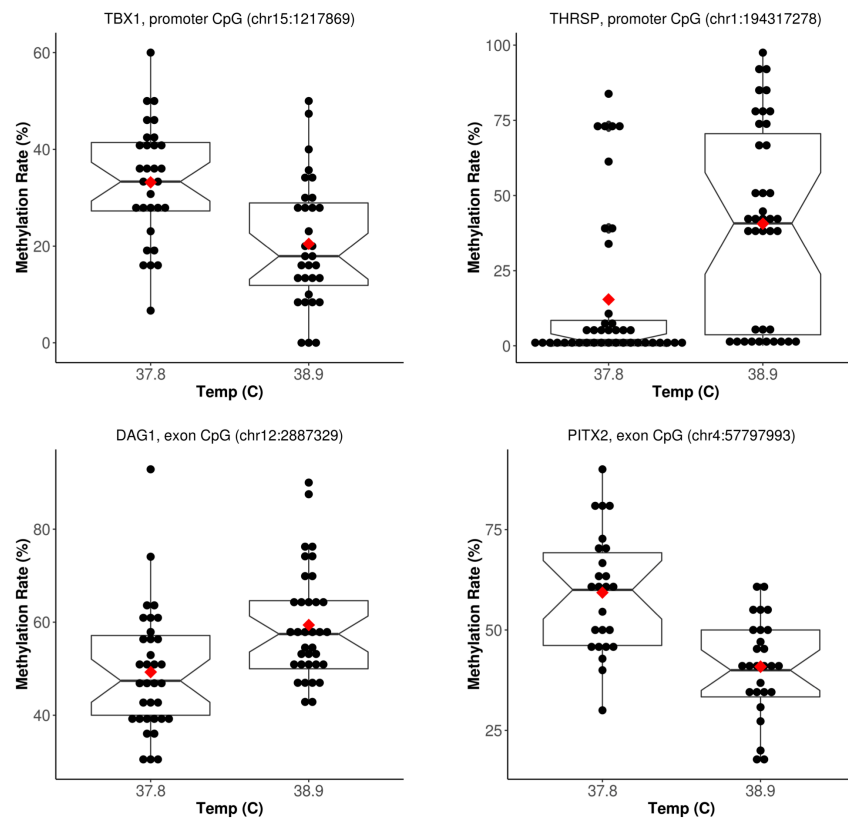


Figure E.1. Egg shell temperature (EST) differential methylation in genes influencing heart development. Dot plots of methylation rates by EST for CpGs in the promoters of *TBX1* and *THRSP* and exons of *DAG1* and *PITX2*. Notched horizontal lines and red dots indicate median and men methylation rate, respectively.

REFERENCES

REFERENCES

1. Rothschild MF, Hu ZL, Jiang Z. Advances in QTL mapping in pigs. *Int J Biol Sci.* 2007;3:192–7. doi:10.7150/ijbs.3.192.
2. Ernst CW, Steibel JP. Molecular advances in QTL discovery and application in pig breeding. *Trends Genet.* 2013;29:215–24. doi:10.1016/j.tig.2013.02.002.
3. Warr A, Affara N, Aken B, Beiki H, Bickhart DM, Billis K, et al. An improved pig reference genome sequence to enable pig genetics and genomics research. 2021;9:1–14. doi:10.1093/gigascience/giaa051.
4. Lunney JK. Advances in swine biomedical model genomics. *International Journal of Biological Sciences.* 2007;3:179–84. doi:10.7150/ijbs.3.179.
5. Dawson HD, Loveland JE, Pascal G, Gilbert JGR, Uenishi H, Mann KM, et al. Structural and functional annotation of the porcine immunome. *BMC Genomics.* 2013;14:332. doi:10.1186/1471-2164-14-332.
6. Badke YM, Bates RO, Ernst CW, Schwab C, Steibel JP. Estimation of linkage disequilibrium in four US pig breeds. *BMC Genomics.* 2012;13:24. doi:10.1186/1471-2164-13-24.
7. Visel A, Rubin EM, Pennacchio LA. Genomic views of distant-acting enhancers. *Nature.* 2009;461:199–205. doi:10.1038/nature08451.
8. Udvardy A, Maine E, Schedl P. The 87A7 chromomere. Identification of novel chromatin structures flanking the heat shock locus that may define the boundaries of higher order domains. *J Mol Biol.* 1985;185:341–58.
9. Ogbourne S, Antalis TM. Transcriptional control and the role of silencers in transcriptional regulation in eukaryotes. *Biochem J.* 1998;331:1–14.
10. Kindt ASD, Navarro P, Semple CAM, Haley CS. The genomic signature of trait-associated variants. *BMC Genomics.* 2013;14:1. doi:10.1186/1471-2164-14-108.
11. ENCODE Project et al. An integrated encyclopedia of DNA elements in the human genome. 2012. doi:10.1038/nature11247.
12. Yue F, Cheng Y, Breschi A, Vierstra J, Wu W, Ryba T, et al. A comparative encyclopedia of DNA elements in the mouse genome. *Nature.* 2014;515:355–64. doi:10.1038/nature13992.
13. Roy S, Ernst J, Kharchenko P V., Kheradpour P, Negre N, Eaton ML, et al. Identification of functional elements and regulatory circuits by *Drosophila* modENCODE. *Science.* 2010;330:1787–97. doi:10.1126/science.1198374.
14. Gerstein MB, Lu ZJ, Van Nostrand EL, Cheng C, Arshinoff BI, Liu T, et al. Integrative analysis of the *Caenorhabditis elegans* genome by the modENCODE project. *Science.*

2010;330:1775–87. doi:10.1126/science.1196914.

15. Schaub MA, Boyle AP, Kundaje A, Batzoglou S, Snyder M. Linking disease associations with regulatory information in the human genome. *Genome Res.* 2012;22:1748–59. doi:10.1101/gr.136127.111.

16. Corces MR, Granja JM, Shams S, Louie BH, Seoane JA, Zhou W, et al. The chromatin accessibility landscape of primary human cancers. *Science.* 2018;362. doi:10.1126/science.aav1898.

17. Andersson L, Archibald AL, Bottema CD, Brauning R, Burgess SC, Burt DW, et al. Coordinated international action to accelerate genome-to-phenome with FAANG, the Functional Annotation of Animal Genomes project. *Genome Biol.* 2015;16. doi:10.1186/s13059-015-0622-4.

18. Giuffra E, Tuggle CK. Functional Annotation of Animal Genomes (FAANG): Current Achievements and Roadmap. *Annual Review of Animal Biosciences.* 2019;7:65–88. doi:10.1146/annurev-animal-020518-114913.

19. Kern C, Wang Y, Xu X, Pan Z, Halstead M, Chanthavixay G, et al. Functional annotations of three domestic animal genomes provide vital resources for comparative and agricultural research. *Nat Commun.* 2021;12. doi:10.1038/s41467-021-22100-8.

20. Schmidt D, Wilson MD, Ballester B, Schwalie PC, Brown GD, Marshall A, et al. Five-vertebrate ChIP-seq reveals the evolutionary dynamics of transcription factor binding. *Science.* 2010;328:1036–40. doi:10.1126/science.1186176.

21. Heintzman ND, Stuart RK, Hon G, Fu Y, Ching CW, Hawkins RD, et al. Distinct and predictive chromatin signatures of transcriptional promoters and enhancers in the human genome. *Nat Genet.* 2007;39:311–8. doi:10.1038/ng1966.

22. Morey L, Helin K. Polycomb group protein-mediated repression of transcription. *Trends Biochem Sci.* 2010;35:323–32.

23. Herrera-Urbe J, Liu H, Byrne KA, Bond ZF, Loving CL, Tuggle CK. Changes in H3K27ac at Gene Regulatory Regions in Porcine Alveolar Macrophages Following LPS or PolyIC Exposure. *Front Genet.* 2020;11. doi:10.3389/fgene.2020.00817.

24. Ziller MJ, Müller F, Liao J, Zhang Y, Gu H. Genomic Distribution and Inter-Sample Variation of Non-CpG Methylation across Human Cell Types. *PLoS Genet.* 2011;7:1002389. doi:10.1371/journal.pgen.1002389.

25. Varley KE, Gertz J, Bowling KM, Parker SL, Reddy TE, Pauli-Behn F, et al. Dynamic DNA methylation across diverse human cell lines and tissues. *Genome Res.* 2013;23:555–67. doi:10.1101/gr.147942.112.

26. Ehrlich M, Gama-Sosa MA, Huang LH, Midgett RM, Kuo KC, Mccune RA, et al. Amount and distribution of 5-methylcytosine in human DNA from different types of tissues or cells.

- Nucleic Acids Res. 1982;10:2709–21. doi:10.1093/nar/10.8.2709.
27. Bird A. DNA methylation patterns and epigenetic memory. *Genes and Development*. 2002;16:6–21. doi:10.1101/gad.947102.
 28. Bestor TH. Activation of mammalian DNA methyltransferase by cleavage of a Zn binding regulatory domain. *EMBO J*. 1992;11:2611–7. doi:10.1002/j.1460-2075.1992.tb05326.x.
 29. Okano M, Bell DW, Haber DA, Li E. DNA methyltransferases Dnmt3a and Dnmt3b are essential for de novo methylation and mammalian development. *Cell*. 1999;99:247–57.
 30. Wu X, Zhang Y. TET-mediated active DNA demethylation: Mechanism, function and beyond. *Nat Rev Genet*. 2017;18:517–34. doi:10.1038/nrg.2017.33.
 31. Bird AP, Wolffe AP. Methylation-induced repression-belts, braces, and chromatin. *Cell*. 1999;99:451–4.
 32. Rakyan VK, Down TA, Thorne NP, Flicek P, Kulesha E, Gräf S, et al. An integrated resource for genome-wide identification and analysis of human tissue-specific differentially methylated regions (tDMRs). *Genome Res*. 2008;18:1518–29. doi:10.1101/gr.077479.108.
 33. Yin Y, Morgunova E, Jolma A, Kaasinen E, Sahu B, Khund-Sayeed S, et al. Impact of cytosine methylation on DNA binding specificities of human transcription factors. *Science*. 2017;356. doi:10.1126/science.aaj2239.
 34. Nan X, Ng HH, Johnson CA, Laherty CD, Turner BM, Eisenman RN, et al. Transcriptional repression by the methyl-CpG-binding protein MeCP2 involves a histone deacetylase complex. *Nature*. 1998;393:386–9. doi:10.1038/30764.
 35. Lorincz MC, Dickerson DR, Schmitt M, Groudine M. Intragenic DNA methylation alters chromatin structure and elongation efficiency in mammalian cells. *Nat Struct Mol Biol*. 2004;11:1068–75.
 36. Aran D, Toperoff G, Rosenberg M, Hellman A. Replication timing-related and gene body-specific methylation of active human genes. *Hum Mol Genet*. 2011;20:670–80. doi:10.1093/hmg/ddq513.
 37. Maunakea AK, Nagarajan RP, Bilenky M, Ballinger TJ, Dsouza C, Fouse SD, et al. Conserved role of intragenic DNA methylation in regulating alternative promoters. *Nature*. 2010;466:253–7.
 38. Jjingo D, Conley AB, Yi S V., Lunyak V V., King Jordan I. On the presence and role of human gene-body DNA methylation. *Oncotarget*. 2012;3:462–74. doi:10.18632/oncotarget.497.
 39. Thorvaldsen JL, Duran KL, Bartolomei MS. Deletion of the H19 differentially methylated domain results in loss of imprinted expression of H19 and Igf2. *Genes Dev*. 1998;12:3693–702. doi:10.1101/gad.12.23.3693.

40. Zwart R, Sleutels F, Wutz A, Schinkel AH, Barlow DP. Bidirectional action of the Igf2r imprint control element on upstream and downstream imprinted genes. *Genes Dev.* 2001;15:2361–6. doi:10.1101/gad.206201.
41. Ziller MJ, Gu H, Müller F, Donaghey J, Tsai LTY, Kohlbacher O, et al. Charting a dynamic DNA methylation landscape of the human genome. *Nature.* 2013;500:477–81. doi:10.1038/nature12433.
42. Schachtschneider KM, Madsen O, Park C, Rund LA, Groenen MAM, Schook LB. Adult porcine genome-wide DNA methylation patterns support pigs as a biomedical model. *BMC Genomics.* 2011;16. doi:10.1186/s12864-015-1938-x.
43. Smith ZD, Meissner A. DNA methylation: Roles in mammalian development. *Nat Rev Genet.* 2013;14:204–20. doi:10.1038/nrg3354.
44. Sliker RC, Roost MS, van Iperen L, Suchiman HED, Tobi EW, Carlotti F, et al. DNA Methylation Landscapes of Human Fetal Development. *PLOS Genet.* 2015;11:e1005583. doi:10.1371/journal.pgen.1005583.
45. Wigmore PMC, Strickland NC. Muscle development in large and small pig fetuses. *J Anat.* 1983;137:235–45. <https://www.ncbi.nlm.nih.gov.proxy2.cl.msu.edu/pmc/articles/PMC1171817/pdf/janat00206-0014.pdf>. Accessed 23 Jul 2017.
46. Foxcroft GR, Dixon WT, Dyck MK, Novak S, Harding JC, Almeida FC. Prenatal programming of postnatal development in the pig. *Soc Reprod Fertil Suppl.* 2009;66:213–31.
47. Bartolomei MS, Zemel S, Tilghman SM. Parental imprinting of the mouse H19 gene. *Nature.* 1991;351:153–5. doi:10.1038/351153a0.
48. Bartolomei MS. Genomic imprinting: Employing and avoiding epigenetic processes. *Genes and Development.* 2009;23:2124–33. doi:10.1101/gad.1841409.
49. Barlow DP, Bartolomei MS. Genomic imprinting in mammals. *Cold Spring Harb Perspect Biol.* 2014;6. doi:10.1101/cshperspect.a018382.
50. Magee DA, Spillane C, Berkowicz EW, Sikora KM, MacHugh DE. Imprinted loci in domestic livestock species as epigenomic targets for artificial selection of complex traits. *Anim Genet.* 2014;45:25–39. doi:10.1111/age.12168.
51. Van Laere AS, Nguyen M, Braunschweig M, Nezer C, Collette C, Moreau L, et al. A regulatory mutation in IGF2 causes a major QTL effect on muscle growth in the pig. *Nature.* 2003;425:832–6.
52. Markljung E, Jiang L, Jaffe JD, Mikkelsen TS, Wallerman O, Larhammar M, et al. ZBED6, a novel transcription factor derived from a domesticated DNA transposon regulates IGF2 expression and muscle growth. *PLoS Biol.* 2009;7. doi:10.1371/journal.pbio.1000256.

53. Mazzio EA, Soliman KFA. Basic concepts of epigenetics impact of environmental signals on gene expression. *Epigenetics*. 2012;7:119–30. doi:10.4161/epi.7.2.18764.
54. Weaver ICG, Cervoni N, Champagne FA, D’alessio AC, Sharma S, Seckl JR, et al. Epigenetic programming by maternal behavior. *Nat Neurosci*. 2004;7. doi:10.1038/nn1276.
55. McGowan PO, Suderman M, Sasaki A, Huang TCT, Hallett M, Meaney MJ, et al. Broad epigenetic signature of maternal care in the brain of adult rats. *PLoS One*. 2011;6. doi:10.1371/journal.pone.0014739.
56. González Ramírez C, Villavicencio Queijeiro A, Jiménez Morales S, Bárcenas López D, Hidalgo Miranda A, Ruiz Chow A, et al. The NR3C1 gene expression is a potential surrogate biomarker for risk and diagnosis of posttraumatic stress disorder. *Psychiatry Res*. 2020;284:112797.
57. Hompes T, Izzi B, Gellens E, Morreels M, Fieuws S, Pexsters A, et al. Investigating the influence of maternal cortisol and emotional state during pregnancy on the DNA methylation status of the glucocorticoid receptor gene (NR3C1) promoter region in cord blood. *J Psychiatr Res*. 2013;47:880–91. doi:10.1016/j.jpsychires.2013.03.009.
58. Oberlander TF, Weinberg J, Papsdorf M, Grunau R, Misri S, Devlin AM. Prenatal exposure to maternal depression, neonatal methylation of human glucocorticoid receptor gene (NR3C1) and infant cortisol stress responses. *Epigenetics*. 2008;3:97–106.
59. Baker EC, Cilkiz KZ, Riggs PK, Littlejohn BP, Long CR, Welsh TH, et al. Effect of prenatal transportation stress on DNA methylation in Brahman heifers. *Livest Sci*. 2020;240:104116.
60. Littlejohn BP, Price DM, Neuendorff DA, Carroll JA, Vann RC, Riggs PK, et al. Prenatal transportation stress alters genome-wide DNA methylation in suckling Brahman bull calves. *J Anim Sci*. 2018;96:5075–99. doi:10.1093/jas/sky350.
61. Schachtschneider KM, Welge ME, Auvil LS, Chaki S, Rund LA, Madsen O, et al. Altered hippocampal epigenetic regulation underlying reduced cognitive development in response to early life environmental insults. *Genes (Basel)*. 2020;11. doi:10.3390/genes11020162.
62. Tzschentke B, Basta D. Early development of neuronal hypothalamic thermosensitivity in birds: influence of epigenetic temperature adaptation. *Comp Biochem Physiol Part A Mol Integr Physiol*. 2002;131:825–32. doi:10.1016/S1095-6433(02)00020-X.
63. Ibeagha-Awemu EM, Zhao X. Epigenetic marks: regulators of livestock phenotypes and conceivable sources of missing variation in livestock improvement programs. *Front Genet*. 2015;6:302. doi:10.3389/fgene.2015.00302.
64. Hao Y, Cui Y, Gu X. Genome-wide DNA methylation profiles changes associated with constant heat stress in pigs as measured by bisulfite sequencing. *Sci Rep*. 2016;6:27507. doi:10.1038/srep27507.
65. Del Corvo M, Lazzari B, Capra E, Zavarez L, Milanesi M, Utsunomiya YT, et al. Methylome

Patterns of Cattle Adaptation to Heat Stress. *Front Genet.* 2021;12:864.
doi:10.3389/fgene.2021.633132.

66. Yahav S, Rath RS, Shinder D. The effect of thermal manipulations during embryogenesis of broiler chicks (*Gallus domesticus*) on hatchability, body weight and thermoregulation after hatch. *J Therm Biol.* 2004;29:245–50. doi:10.1016/j.jtherbio.2004.03.002.

67. Maatjens CM, Van Roovert-Reijrink IAM, Engel B, Van Der Pol CW, Kemp B, Van Den Brand H. Temperature during the last week of incubation. I. Effects on hatching pattern and broiler chicken embryonic organ development. *Poult Sci.* 2016;95:956–65.
doi:10.3382/ps/pev447.

68. Maatjens CM, Van Roovert-Reijrink IAM, Van Den Anker I, Engel B, Van Der Pol CW, Kemp B, et al. Temperature during the last week of incubation. II. Effects on first week broiler development and performance. *Poult Sci.* 2016;95:2136–44. doi:10.3382/ps/pew145.

69. David S-A, Vitorino Carvalho A, Gimmonnet C, Brionne A, Hennequet-Antier C, Piégu B, et al. Thermal Manipulation During Embryogenesis Impacts H3K4me3 and H3K27me3 Histone Marks in Chicken Hypothalamus. *Front Genet.* 2019;10:1207. doi:10.3389/fgene.2019.01207.

70. Vinoth A, Thirunalasundari T, Shanmugam M, Uthrakumar A, Suji S, Rajkumar U. Evaluation of DNA methylation and mRNA expression of heat shock proteins in thermal manipulated chicken. *Cell Stress Chaperones.* 2018;23:235–52.

71. Summerfield A. Special issue on porcine immunology: An introduction from the guest editor. *Dev Comp Immunol.* 2009;33:265–6.

72. Zuckermann FA, Gaskins HR. Distribution of porcine CD4/CD8 double-positive T lymphocytes in mucosa-associated lymphoid tissues. *Immunology.* 1996;87:493–9.
doi:10.1046/j.1365-2567.1996.494570.x.

73. Zuckermann FA, Husmann RJ. Functional and phenotypic analysis of porcine peripheral blood CD4/CD8 double-positive T cells. *Immunology.* 1996;87:500–12.
<http://www.ncbi.nlm.nih.gov/pubmed/8778040><http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=PMC1384123>. Accessed 23 Sep 2020.

74. Yang H, Parkhouse RME. Differential expression of CD8 epitopes amongst porcine CD8-positive functional lymphocyte subsets. *Immunology.* 1997;92:45–52.

75. Takamatsu HH, Denyer MS, Stirling C, Cox S, Aggarwal N, Dash P, et al. Porcine $\gamma\delta$ T cells: Possible roles on the innate and adaptive immune responses following virus infection. *Vet Immunol Immunopathol.* 2006;112:49–61.

76. Stepanova K, Sinkora M. Porcine $\gamma\delta$ T Lymphocytes Can Be Categorized into Two Functionally and Developmentally Distinct Subsets according to Expression of CD2 and Level of TCR. *J Immunol.* 2013;190:2111–20. doi:10.4049/jimmunol.1202890.

77. Meurens F, Summerfield A, Nauwynck H, Saif L, Gerds V. The pig: A model for human

- infectious diseases. *Trends in Microbiology*. 2012;20:50–7. doi:10.1016/j.tim.2011.11.002.
78. Roth JA, Tuggle CK. Livestock models in translational medicine. *ILAR J*. 2015;56:1–6. doi:10.1093/ilar/ilv011.
79. Wang M, Ibeagha-Awemu EM. Impacts of Epigenetic Processes on the Health and Productivity of Livestock. *Frontiers in Genetics*. 2021;11:613636. doi:10.3389/fgene.2020.613636.
80. Wang H, Wang J, Ning C, Zheng X, Fu J, Wang A, et al. Genome-wide DNA methylation and transcriptome analyses reveal genes involved in immune responses of pig peripheral blood mononuclear cells to poly I:C. *Sci Rep*. 2017;7:9709. doi:10.1038/s41598-017-10648-9.
81. Herrera-Urbe J, Wiarda JE, Sivasankaran SK, Daharsh L, Liu H, Byrne KA, et al. Reference Transcriptomes of Porcine Peripheral Immune Cells Created Through Bulk and Single-Cell RNA Sequencing. *Front Genet*. 2021;12:689406. doi:10.3389/fgene.2021.689406.
82. Greenberg MVC, Bourc'his D. The diverse roles of DNA methylation in mammalian development and disease. *Nat Rev Mol Cell Biol*. 2019;20:590–607. doi:10.1038/s41580-019-0159-6.
83. Li M, Wu H, Luo Z, Xia Y, Guan J, Wang T, et al. An atlas of DnA methylomes in porcine adipose and muscle tissues. *Nat Commun*. 2012;3. doi:10.1038/ncomms1854.
84. Fang L, Zhou Y, Liu S, Jiang J, Bickhart DM, Null DJ, et al. Integrating Signals from Sperm Methylome Analysis and Genome-Wide Association Study for a Better Understanding of Male Fertility in Cattle. *Epigenomes*. 2019;3:10. doi:10.3390/epigenomes3020010.
85. King AD, Huang K, Rubbi L, Liu S, Wang CY, Wang Y, et al. Reversible Regulation of Promoter and Enhancer Histone Landscape by DNA Methylation in Mouse Embryonic Stem Cells. *Cell Rep*. 2016;17:289–302. doi:10.1016/j.celrep.2016.08.083.
86. Corbett RJ, Luttman AM, Wurtz KE, Siegford JM, Raney NE, Ford LM, et al. Weaning Induces Stress-Dependent DNA Methylation and Transcriptional Changes in Piglet PBMCs. *Front Genet*. 2021;12:92. doi:10.3389/fgene.2021.633564.
87. Bolger AM, Lohse M, Usadel B. Trimmomatic: A flexible trimmer for Illumina sequence data. *Bioinformatics*. 2014;30:2114–20. doi:10.1093/bioinformatics/btu170.
88. Krueger F, Andrews SR. Bismark: a flexible aligner and methylation caller for Bisulfite-Seq applications. *Bioinformatics*. 2011;27:1571–2. doi:10.1093/bioinformatics/btr167.
89. Akalin A, Kormaksson M, Li S, Garrett-Bakelman FE, Figueroa ME, Melnick A, et al. MethylKit: a comprehensive R package for the analysis of genome-wide DNA methylation profiles. *Genome Biol*. 2012;13. doi:10.1186/gb-2012-13-10-R87.
90. Akalin A, Franke V, Vlahoviček K, Mason CE, Schübeler D. Genomation: A toolkit to summarize, annotate and visualize genomic intervals. *Bioinformatics*. 2015;31:1127–9.

doi:10.1093/bioinformatics/btu775.

91. Thomas PD, Kejariwal A, Guo N, Mi H, Campbell MJ, Muruganujan A, et al. Applications for protein sequence-function evolution data: mRNA/protein expression analysis and coding SNP scoring tools. *Nucleic Acids Res.* 2006;34. doi:10.1093/nar/gkl229.
92. Thomas PD, Campbell MJ, Kejariwal A, Mi H, Karlak B, Daverman R, et al. PANTHER: A Library of Protein Families and Subfamilies Indexed by Function. *Genome Res.* 2003;13:2129–41. doi:10.1101/GR.772403.
93. Love MI, Huber W, Anders S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* 2011;15. doi:10.1186/s13059-014-0550-8.
94. Bailey TL, Boden M, Buske FA, Frith M, Grant CE, Clementi L, et al. MEME SUITE: tools for motif discovery and searching. *Nucleic Acids Res.* 2009;37. doi:10.1093/nar/gkp335.
95. McLeay RC, Bailey TL. Motif Enrichment Analysis: A unified framework and an evaluation on ChIP data. 2010. doi:10.1186/1471-2105-11-165.
96. Hume MA, Barrera LA, Gisselbrecht SS, Bulyk ML. UniPROBE, update 2015: New tools and content for the online database of protein-binding microarray data on protein-DNA interactions. *Nucleic Acids Res.* 2015;43:D117–22. doi:10.1093/nar/gku1045.
97. Jolma A, Yan J, Whittington T, Toivonen J, Nitta KR, Rastas P, et al. DNA-binding specificities of human transcription factors. *Cell.* 2013;152:327–39. doi:10.1016/j.cell.2012.12.009.
98. Fornes O, Castro-Mondragon JA, Khan A, Van Der Lee R, Zhang X, Richmond PA, et al. JASPAR 2020: Update of the open-Access database of transcription factor binding profiles. *Nucleic Acids Res.* 2020;48:D87–92. doi:10.1093/nar/gkz1001.
99. Kolde R. pheatmap : Pretty Heatmaps. R package version 1.0.8. 2015;:1–7. <https://cran.r-project.org/web/packages/pheatmap/pheatmap.pdf>.
100. Hu Z-L, Park CA, Reecy JM. Building a livestock genetic and genomic information knowledgebase through integrative developments of Animal QTLdb and CorrDB. *Nucleic Acids Res.* 2018;47:701–10. doi:10.1093/nar/gky1084.
101. Cinar MU, Islam MA, Pröll M, Kocamis H, Tholen E, Tesfaye D, et al. Evaluation of suitable reference genes for gene expression studies in porcine PBMCs in response to LPS and LTA. *BMC Res Notes.* 2013;6.
102. Wang J, Wang Y, Wang H, Hao X, Wu Y, Guo J. Selection of reference genes for gene expression studies in porcine whole blood and peripheral blood mononuclear cells under polyinosinic:Polycytidylic acid stimulation. *Asian-Australasian J Anim Sci.* 2014;27:471–8.
103. Gopalakrishnan S, Sullivan BA, Trazzi S, Della Valle G, Robertson KD. DNMT3B interacts with constitutive centromere protein CENP-C to modulate DNA methylation and the

histone code at centromeric regions. *Hum Mol Genet.* 2009;18:3178–93. doi:10.1093/hmg/ddp256.

104. Mucida D, Husain MM, Muroi S, Van Wijk F, Shinnakasu R, Naoe Y, et al. Transcriptional reprogramming of mature CD4 + helper T cells generates distinct MHC class II-restricted cytotoxic T lymphocytes. *Nat Immunol.* 2013;14:281–9. doi:10.1038/ni.2523.

105. Zhang J, Chen JH, Liu XD, Wang HY, Liu XL, Li XY, et al. Genomewide association studies for hematological traits and T lymphocyte subpopulations in a Duroc × Erhualian F2 resource population. *J Anim Sci.* 2016;94:5028–41. doi:10.2527/jas.2016-0924.

106. Zimmerman J., Osorio F., Benfield D., Murtaugh M. SG and TM. PRRS virus (porcine arterivirus). In: *Diseases of Swine* (9th edition). 9th edition. Blackwell Publishing Company; 2006. p. 387–417.

107. Wang J, Liu J-Y, Shao K-Y, Han Y-Q, Li G-L, Ming S-L, et al. Porcine Reproductive and Respiratory Syndrome Virus Activates Lipophagy To Facilitate Viral Replication through Downregulation of NDRG1 Expression. *J Virol.* 2019;93. doi:10.1128/jvi.00526-19.

108. Kulis M, Merkel A, Heath S, Queirós AC, Schuyler RP, Castellano G, et al. Whole-genome fingerprint of the DNA methylome during human B cell differentiation. *Nat Genet.* 2015;47:746–56. doi:10.1038/ng.3291.

109. Barwick BG, Scharer CD, Bally APR, Boss JM. Plasma cell differentiation is coupled to division-dependent DNA hypomethylation and gene regulation. *Nat Immunol.* 2016;17:1216–25. doi:10.1038/ni.3519.

110. Lee ST, Xiao Y, Muench MO, Xiao J, Fomin ME, Wiencke JK, et al. A global DNA methylation and gene expression analysis of early human B-cell development reveals a demethylation signature and transcription factor network. *Nucleic Acids Res.* 2012;40:11339–51. doi:10.1093/nar/gks957.

111. Wiencke JK, Butler R, Hsuang G, Eliot M, Kim S, Sepulveda MA, et al. The DNA methylation profile of activated human natural killer cells. *Epigenetics.* 2016;11:363–80. doi:10.1080/15592294.2016.1163454.

112. Okano M, Bell DW, Haber DA, Li E. DNA methyltransferases Dnmt3a and Dnmt3b are essential for de novo methylation and mammalian development. *Cell.* 1999;99:247–57. doi:10.1016/S0092-8674(00)81656-6.

113. Hermann A, Goyal R, Jeltsch A. The Dnmt1 DNA-(cytosine-C5)-methyltransferase methylates DNA processively with high preference for hemimethylated target sites. *J Biol Chem.* 2004;279:48350–9. doi:10.1074/jbc.M403427200.

114. Zhu H, Wang G, Qian J. Transcription factors as readers and effectors of DNA methylation. *Nat Rev Genet.* 2016;17:551–65. doi:10.1038/nrg.2016.83.

115. Hasemann MS, Lauridsen FKB, Waage J, Jakobsen JS, Frank AK, Schuster MB, et al.

C/EBP α Is Required for Long-Term Self-Renewal and Lineage Priming of Hematopoietic Stem Cells and for the Maintenance of Epigenetic Configurations in Multipotent Progenitors. *PLoS Genet.* 2014;10:1004079. doi:10.1371/journal.pgen.1004079.

116. Williams SC, Du Y, Schwartz RC, Weiler SR, Ortiz M, Keller JR, et al. C/EBP ϵ is a myeloid-specific activator of cytokine, chemokine, and macrophage-colony-stimulating factor receptor genes. 1998. doi:10.1074/jbc.273.22.13493.

117. Lekstrom-Himes J, Xanthopoulos KG. Biological role of the CCAAT/enhancer-binding protein family of transcription factors. *Journal of Biological Chemistry.* 1998;273:28545–8. doi:10.1074/jbc.273.44.28545.

118. Nechanitzky R, Akbas D, Scherer S, Györy I, Hoyler T, Ramamoorthy S, et al. Transcription factor EBF1 is essential for the maintenance of B cell identity and prevention of alternative fates in committed cells. *Nat Immunol.* 2013;14:867–75. doi:10.1038/ni.2641.

119. Urbánek P, Wang ZQ, Fetka I, Wagner EF, Busslinger M. Complete block of early B cell differentiation and altered patterning of the posterior midbrain in mice lacking Pax5/BSAP. *Cell.* 1994;79:901–12.

120. Yu Q, Sharma A, Sen JM. TCF1 and β -catenin regulate T cell development and function. *Immunologic Research.* 2010;47:45–55. doi:10.1007/s12026-009-8137-2.

121. Gordon SM, Chaix J, Rupp LJ, Wu J, Madera S, Sun JC, et al. The Transcription Factors T-bet and Eomes Control Key Checkpoints of Natural Killer Cell Maturation. *Immunity.* 2012;36:55–67.

122. Rodríguez-Gómez IM, Talker SC, Käser T, Stadler M, Reiter L, Ladinig A, et al. Expression of T-Bet, Eomesodermin, and GATA-3 Correlates With Distinct Phenotypes and Functional Properties in Porcine $\gamma\delta$ T Cells. *Front Immunol.* 2019;10 MAR:396. doi:10.3389/fimmu.2019.00396.

123. Riera-Sans L, Behrens A. Regulation of $\alpha\beta/\gamma\delta$ T Cell Development by the Activator Protein 1 Transcription Factor c-Jun. *J Immunol.* 2007;178:5690–700. doi:10.4049/jimmunol.178.9.5690.

124. Samten B, Townsend JC, Weis SE, Bhounik A, Klucar P, Shams H, et al. CREB, ATF, and AP-1 Transcription Factors Regulate IFN- γ Secretion by Human T Cells in Response to Mycobacterial Antigen. *J Immunol.* 2008;181:2056–64. doi:10.4049/jimmunol.181.3.2056.

125. Chen Z, Ye S, Teng J, Diao S, Yuan X, Chen Z, et al. Genome-wide association studies for the number of animals born alive and dead in duroc pigs. *Theriogenology.* 2019;139:36–42.

126. England EM, Scheffler TL, Kasten SC, Matarneh SK, Gerrard DE. Exploring the unknowns involved in the transformation of muscle to meat. *Meat Sci.* 2013;95:837–43.

127. Wigmore PM, Evans DJR. Molecular and cellular mechanisms involved in the generation of fiber diversity during myogenesis. *Int Rev Cytol.* 2002;216:175–232. doi:10.1016/S0074-

7696(02)16006-2.

128. Asfour HA, Allouh MZ, Said RS. Myogenic regulatory factors: The orchestrators of myogenesis after 30 years of discovery. *Exp Biol Med*. 2018;243:118–28. doi:10.1177/1535370217749494.

129. Sollero BP, Guimarães SEF, Rilmington VD, Tempelman RJ, Raney NE, Steibel JP, et al. Transcriptional profiling during foetal skeletal muscle development of Piau and Yorkshire-Landrace cross-bred pigs. *Anim Genet*. 2011;42:600–12. doi:10.1111/j.1365-2052.2011.02186.x.

130. Te Pas MFW, De Wit AAW, Priem J, Cagnazzo M, Davoli R, Russo V, et al. Transcriptome expression profiles in prenatal pigs in relation to myogenesis. *J Muscle Res Cell Motil*. 2005;26:157–65.

131. Zhao S-H, Nettleton D, Liu W, Fitzsimmons C, Ernst CW, Raney NE, et al. Complementary DNA macroarray analyses of differential gene expression in porcine fetal and postnatal muscle1. *J Anim Sci*. 2003;81:2179–88. doi:10.2527/2003.8192179x.

132. Cagnazzo M, te Pas MFW, Priem J, de Wit AAC, Pool MH, Davoli R, et al. Comparison of prenatal muscle tissue expression profiles of two pig breeds differing in muscle characteristics1. *J Anim Sci*. 2006;84:1–10. doi:10.2527/2006.8411.

133. Muráni E, Murániová M, Ponsuksili S, Schellander K, Wimmers K. Identification of genes differentially expressed during prenatal development of skeletal muscle in two pig breeds differing in muscularity. *BMC Dev Biol*. 2007;7:109. doi:10.1186/1471-213X-7-109.

134. Tang Z, Li Y, Wan P, Li X, Zhao S, Liu B, et al. LongSAGE analysis of skeletal muscle at three prenatal stages in Tongcheng and Landrace pigs. *Genome Biol*. 2007;8:R115. doi:10.1186/gb-2007-8-6-r115.

135. Horak M, Novak J, Bienertova-Vasku J. Muscle-specific microRNAs in skeletal muscle development. *Dev Biol*. 2016;410:1–13. doi:10.1016/j.ydbio.2015.12.013.

136. Daza KR, Steibel JP, Velez-Irizarry D, Raney NE, Bates RO, Ernst CW. Profiling and characterization of a longissimus dorsi muscle microRNA dataset from an F2 Duroc × Pietrain pig resource population. *Genomics Data*. 2017;13:50–3. doi:10.1016/j.gdata.2017.07.006.

137. Li M, Xia Y, Gu Y, Zhang K, Lang Q, Chen L, et al. MicroRNAome of porcine pre- and postnatal development. *PLoS One*. 2010;5. doi:10.1371/journal.pone.0011541.

138. Liu X, Trakooljul N, Hadlich F, Muráni E, Wimmers K, Ponsuksili S. MicroRNA-mRNA regulatory networking fine-tunes the porcine muscle fiber type, muscular mitochondrial respiratory and metabolic enzyme activities. *BMC Genomics*. 2016;17:531. doi:10.1186/s12864-016-2850-8.

139. Verardo L, Nascimento C, Silva F, Gasparino E, Toriyama E, Barbosa A, et al. Identification and expression levels of pig miRNAs in skeletal muscle. *Livest Sci*. 2013;154:45–54. doi:10.1016/j.livsci.2013.02.019.

140. Hao Y, Cui Y, Gu X. Genome-wide DNA methylation profiles changes associated with constant heat stress in pigs as measured by bisulfite sequencing. *Sci Rep*. 2016;6. doi:10.1038/srep27507.
141. Yang Y, Fan X, Yan J, Chen M, Zhu M, Tang Y, et al. A comprehensive epigenome atlas reveals DNA methylation regulating skeletal muscle development. *Nucleic Acids Res*. 2021;49:1313–29. doi:10.1093/nar/gkaa1203.
142. Ponsuksili S, Trakooljul N, Basavaraj S, Hadlich F, Murani E, Wimmers K. Epigenome-wide skeletal muscle DNA methylation profiles at the background of distinct metabolic types and ryanodine receptor variation in pigs. *BMC Genomics*. 2019;20. doi:10.1186/s12864-019-5880-1.
143. Jin L, Jiang Z, Xia Y, Lou P, Chen L, Wang H, et al. Genome-wide DNA methylation changes in skeletal muscle between young and middle-aged pigs. *BMC Genomics*. 2014;15. doi:10.1186/1471-2164-15-653.
144. Zhang W, Zhang S, Xu Y, Ma Y, Zhang D, Li X, et al. The DNA methylation status of wnt and Tgfb signals is a key factor on functional regulation of skeletal muscle satellite cell development. *Front Genet*. 2019;10 MAR. doi:10.3389/fgene.2019.00220.
145. Kim D, Pertea G, Trapnell C, Pimentel H, Kelley R, Salzberg SL. TopHat2: accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions. *Genome Biol*. 2013;14:R36. doi:10.1186/gb-2013-14-4-r36.
146. Anders S, Pyl PT, Huber W. HTSeq--a Python framework to work with high-throughput sequencing data. *Bioinformatics*. 2015;31:166–9. doi:10.1093/bioinformatics/btu638.
147. Friedländer MR, MacKowiak SD, Li N, Chen W, Rajewsky N. MiRDeep2 accurately identifies known and hundreds of novel microRNA genes in seven animal clades. *Nucleic Acids Res*. 2012;40:37–52.
148. Ott MO, Bober E, Lyons G, Arnold H, Buckingham M. Early expression of the myogenic regulatory gene, myf-5, in precursor cells of skeletal muscle in the mouse embryo. *Development*. 1991;111:1097–107.
149. Kassam-Duchossoy L, Gayraud-Morel B, Gomès D, Rocancourt D, Buckingham M, Shinin V, et al. Mrf4 determines skeletal muscle identity in Myf5:Myod double-mutant mice. *Nature*. 2004;431:466–71. doi:10.1038/nature02876.
150. Summerbell D, Halai C, Rigby PWJ. Expression of the myogenic regulatory factor Mrf4 precedes or is contemporaneous with that of Myf5 in the somitic bud. *Mech Dev*. 2002;117:331–5.
151. Deaton AM, Bird A. CpG islands and the regulation of transcription. *Genes Dev*. 2011;25:1010–22. doi:10.1101/gad.2037511.
152. Schilling E, Rehli M. Global, comparative analysis of tissue-specific promoter CpG

methylation. *Genomics*. 2007;90:314–23.

153. Shen L, Kondo Y, Guo Y, Zhang J, Zhang L, Ahmed S, et al. Genome-wide profiling of DNA methylation reveals a class of normally methylated CpG island promoters. *PLoS Genet*. 2007;3:2023–6. doi:10.1371/journal.pgen.0030181.

154. Mohn F, Weber M, Rebhan M, Roloff TC, Richter J, Stadler MB, et al. Lineage-Specific Polycomb Targets and De Novo DNA Methylation Define Restriction and Potential of Neuronal Progenitors. *Mol Cell*. 2008;30:755–66. doi:10.1016/j.molcel.2008.05.007.

155. Aschrafi A, Schwechter AD, Mameza MG, Natera-Naranjo O, Gioio AE, Kaplan BB. MicroRNA-338 regulates local cytochrome c oxidase IV mRNA levels and oxidative phosphorylation in the axons of sympathetic neurons. *J Neurosci*. 2008;28:12581–90. doi:10.1523/JNEUROSCI.3338-08.2008.

156. Gagan J, Dey BK, Layer R, Yan Z, Dutta A. MicroRNA-378 Targets the Myogenic Repressor MyoR during Myoblast Differentiation. *J Biol Chem*. 2011;286:19431. doi:10.1074/JBC.M111.219006.

157. Siengdee P, Trakooljul N, Murani E, Schwerin M, Wimmers K, Ponsuksili S. MicroRNAs Regulate Cellular ATP Levels by Targeting Mitochondrial Energy Metabolism Genes during C2C12 Myoblast Differentiation. *PLoS One*. 2015;10:e0127850. doi:10.1371/journal.pone.0127850.

158. Kappeler BIG, Regitano LCA, Poleti MD, Cesar ASM, Moreira GCM, Gasparin G, et al. MiRNAs differentially expressed in skeletal muscle of animals with divergent estimated breeding values for beef tenderness. *BMC Mol Biol*. 2019;20:1. doi:10.1186/s12867-018-0118-3.

159. McDanel TG, Smith TP, Doumit ME, Miles JR, Coutinho LL, Sonstegard TS, et al. MicroRNA transcriptome profiles during swine skeletal muscle development. *BMC Genomics*. 2009;10:77. doi:10.1186/1471-2164-10-77.

160. Tsujino A, Maertenst C, Ohno K, Shen XM, Fukuda T, Harper CM, et al. Myasthenic syndrome caused by mutation of the SCN4A sodium channel. *Proc Natl Acad Sci U S A*. 2003;100:7377–82. doi:10.1073/pnas.1230273100.

161. Muráni E, Murániová M, Ponsuksili S, Schellander K, Wimmers K. Identification of genes differentially expressed during prenatal development of skeletal muscle in two pig breeds differing in muscularity. *BMC Dev Biol*. 2007;7:109. doi:10.1186/1471-213X-7-109.

162. Podolska A, Kaczkowski B, Busk PK, Søkilde R, Litman T, Fredholm M, et al. Microrna expression profiling of the porcine developing brain. *PLoS One*. 2011;6. doi:10.1371/journal.pone.0014494.

163. Krombeen SK, Shankar V, Noorai RE, Saski CA, Sharp JL, Wilson ME, et al. The identification of differentially expressed genes between extremes of placental efficiency in maternal line gilts on day 95 of gestation. *BMC Genomics*. 2019;20:1–15. doi:10.1186/s12864-

019-5626-0.

164. Kwon SG, Hwang JH, Park DH, Kim TW, Kang DG, Kang KH, et al. Identification of differentially expressed genes associated with litter size in Berkshire pig placenta. *PLoS One*. 2016;11. doi:10.1371/journal.pone.0153311.
165. Hwang JH, An SM, Kwon S, Park DH, Kim TW, Kang DG, et al. DNA methylation patterns and gene expression associated with litter size in Berkshire pig placenta. *PLoS One*. 2017;12. doi:10.1371/journal.pone.0184539.
166. Pastinen T. Genome-wide allele-specific analysis: Insights into regulatory variation. *Nat Rev Genet*. 2010;11:533–8. doi:10.1038/nrg2815.
167. Kerkel K, Spadola A, Yuan E, Kosek J, Jiang L, Hod E, et al. Genomic surveys by methylation-sensitive SNP analysis identify sequence-dependent allele-specific DNA methylation. *Nat Genet*. 2008;40:904–8. doi:10.1038/ng.174.
168. Schalkwyk LC, Meaburn EL, Smith R, Dempster EL, Jeffries AR, Davies MN, et al. Allelic Skewing of DNA Methylation Is Widespread across the Genome. *Am J Hum Genet*. 2010;86:196–212. doi:10.1016/j.ajhg.2010.01.014.
169. Abante J, Fang Y, Feinberg AP, Goutsias J. Detection of haplotype-dependent allele-specific DNA methylation in WGBS data. *Nat Commun*. 2020;11:1–13. doi:10.1038/s41467-020-19077-1.
170. Wilson ME, Biensen NJ, Youngs CR, Ford SP. Development of Meishan and Yorkshire littermate conceptuses in either a Meishan or Yorkshire uterine environment to day 90 of gestation and to term. *Biol Reprod*. 1998;58:905–10. doi:10.1095/biolreprod58.4.905.
171. Chu Q, Liang T, Fu L, Li H, Zhou BO. Behavioural genetic differences between Chinese and European pigs. *J Genet*. 2011;96:707–15. doi:10.1007/s12041-017-0826-3.
172. Barlow DP, Bartolomei MS. Genomic imprinting in mammals. *Cold Spring Harb Perspect Biol*. 2014;6. doi:10.1101/cshperspect.a018382.
173. Magee DA, Spillane C, Berkowicz EW, Sikora KM, MacHugh DE. Imprinted loci in domestic livestock species as epigenomic targets for artificial selection of complex traits. *Anim Genet*. 2014;45:25–39. doi:10.1111/age.12168.
174. Nonneman D, Rohrer GA, Wise TH, Lunstra DD, Ford JJ. A Variant of Porcine Thyroxine-Binding Globulin Has Reduced Affinity for Thyroxine and Is Associated with Testis Size. *Biol Reprod*. 2005;72:214–20. doi:10.1095/BIOLREPROD.104.031922.
175. McKenna A, Hanna M, Banks E, Sivachenko A, Cibulskis K, Kernytsky A, et al. The genome analysis toolkit: A MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res*. 2010;20:1297–303. doi:10.1101/gr.107524.110.
176. Krueger F, Andrews SR. SNPsplit: Allele-specific splitting of alignments between genomes

with known SNP genotypes. *F1000Research*. 2016;5:1479. doi:10.12688/f1000research.9037.1.

177. Marin M, Karis A, Visser P, Grosveld F, Philipsen S. Transcription factor Sp1 is essential for early embryonic development but dispensable for cell growth and differentiation. *Cell*. 1997;89:619–28. doi:10.1016/S0092-8674(00)80243-3.

178. Eckert D, Buhl S, Weber S, Jäger R, Schorle H. The AP-2 family of transcription factors. *Genome Biology*. 2005;6:1–8. doi:10.1186/gb-2005-6-13-246.

179. Hernández-Hernández JM, García-González EG, Brun CE, Rudnicki MA. The myogenic regulatory factors, determinants of muscle development, cell identity and regeneration. *Seminars in Cell and Developmental Biology*. 2017;72:10–8.

180. Ciliax BJ, Drash GW, Staley JK, Haber S, Mobley CJ, Miller GW, et al. Immunocytochemical localization of the dopamine transporter in human brain. *J Comp Neurol*. 1999;409:38–56.

181. Sun G, Liang X, Qin K, Qin Y, Shi X, Cong P, et al. Functional Analysis of KIT Gene Structural Mutations Causing the Porcine Dominant White Phenotype Using Genome Edited Mouse Models. *Front Genet*. 2020;11. doi:10.3389/fgene.2020.00138.

182. Li Z, Wei S, Li H, Wu K, Cai Z, Li D, et al. Genome-wide genetic structure and differentially selected regions among Landrace, Erhualian, and Meishan pigs using specific-locus amplified fragment sequencing. *Sci Rep*. 2017;7. doi:10.1038/s41598-017-09969-6.

183. Zhao Q bo, Oyelami FO, Qadri QR, Sun H, Xu Z, Wang Q shan, et al. Identifying the unique characteristics of the Chinese indigenous pig breeds in the Yangtze River Delta region for precise conservation. *BMC Genomics*. 2021;22. doi:10.1186/s12864-021-07476-7.

184. Stadler MB, Murr R, Burger L, Ivanek R, Lienert F, Schöler A, et al. DNA-binding factors shape the mouse methylome at distal regulatory regions. *Nature*. 2011;480:490–5. doi:10.1038/nature10716.

185. Suelves M, Carrió E, Núñez-Álvarez Y, Peinado MA. DNA methylation dynamics in cellular commitment and differentiation. *Brief Funct Genomics*. 2016;15:443–53. doi:10.1093/bfpg/elw017.

186. Huse SM, Gruppuso PA, Boekelheide K, Sanders JA. Patterns of gene expression and DNA methylation in human fetal and adult liver. *BMC Genomics*. 2015;16. doi:10.1186/s12864-015-2066-3.

187. Fang F, Hodges E, Molaro A, Dean M, Hannon GJ, Smith AD. Genomic landscape of human allele-specific DNA methylation. *Proc Natl Acad Sci U S A*. 2012;109:7332–7. doi:10.1073/pnas.1201310109.

188. Zhang Y, Rohde C, Reinhardt R, Voelcker-Rehage C, Jeltsch A. Non-imprinted allele-specific DNA methylation on human autosomes. *Genome Biol*. 2009;10:R138. doi:10.1186/gb-2009-10-12-r138.

189. Rinkenberger J, Werb Z. The labyrinthine placenta. *Nat Genet.* 2000;25:248–50.
190. Tian FY, Wang XM, Xie C, Zhao B, Niu Z, Fan L, et al. Placental surface area mediates the association between FGFR2 methylation in placenta and full-term low birth weight in girls. *Clin Epigenetics.* 2018;10. doi:10.1186/s13148-018-0472-5.
191. Elmore CL, Wu X, Leclerc D, Watson ED, Bottiglieri T, Krupenko NI, et al. Metabolic derangement of methionine and folate metabolism in mice deficient in methionine synthase reductase. *Mol Genet Metab.* 2007;91:85–97. doi:10.1016/j.ymgme.2007.02.001.
192. Jadavji NM, Bahous RH, Deng L, Malysheva O, Grand'maison M, Bedell BJ, et al. Mouse model for deficiency of methionine synthase reductase exhibits short-term memory impairment and disturbances in brain choline metabolism. *Biochem J.* 2014;461:205–12. doi:10.1042/BJ20131568.
193. Deng L, Elmore CL, Lawrance AK, Matthews RG, Rozen R. Methionine synthase reductase deficiency results in adverse reproductive outcomes and congenital heart defects in mice. *Mol Genet Metab.* 2008;94:336–42.
194. Sowton AP, Padmanabhan N, Tunster SJ, McNally BD, Murgia A, Yusuf A, et al. Mtrr hypomorphic mutation alters liver morphology, metabolism and fuel storage in mice. *Mol Genet Metab Reports.* 2020;23. doi:10.1016/j.ymgmr.2020.100580.
195. Bischoff SR, Tsai SQ, Hardison NE, Motsinger-Reif AA, Freking BA. Differences in X-Chromosome Transcriptional Activity and Cholesterol Metabolism between Placentae from Swine Breeds from Asian and Western Origins. *PLoS One.* 2013;8:55345. doi:10.1371/journal.pone.0055345.
196. Gu T, Zhu M jin, Schroyen M, Qu L, Nettleton D, Kuhar D, et al. Endometrial gene expression profiling in pregnant Meishan and Yorkshire pigs on day 12 of gestation. *BMC Genomics.* 2014;15:156. doi:10.1186/1471-2164-15-156.
197. DeChiara TM, Robertson EJ, Efstratiadis A. Parental imprinting of the mouse insulin-like growth factor II gene. *Cell.* 1991;64:849–59.
198. Barlow DP, Stöger R, Herrmann BG, Saito K, Schweifer N. The mouse insulin-like growth factor type-2 receptor is imprinted and closely linked to the Tme locus. *Nature.* 1991;349:84–7. doi:10.1038/349084a0.
199. Higashimoto K, Jozaki K, Kosho T, Matsubara K, Fuke T, Yamada D, et al. A novel de novo point mutation of the OCT-binding site in the IGF2 /H19-imprinting control region in a Beckwith-Wiedemann syndrome patient. *Clin Genet.* 2014;86:539–44.
200. Chess A. Monoallelic expression of protocadherin genes. *Nat Genet.* 2005;37:120–1. doi:10.1038/ng0205-120.
201. Yahav S, Hurwitz S. Induction of thermotolerance in male broiler chickens by temperature conditioning at an early age. *Poult Sci.* 1996;75:402–6. doi:10.3382/ps.0750402.

202. Molenaar R, Hulet R, Meijerhof R, Maatjens CM, Kemp B, Van den Brand H. High eggshell temperatures during incubation decrease growth performance and increase the incidence of ascites in broiler chickens. *Poult Sci.* 2011;90:624–32. doi:10.3382/ps.2010-00970.
203. Geverink NA, Kappers A, Van De Burgwal JA, Lambooij E, Blokhuis HJ, Wiegant VM. Effects of Regular Moving and Handling on the Behavioral and Physiological Responses of Pigs to Preslaughter Treatment and Consequences for Subsequent Meat Quality. *J Anim Sci.* 1998;76:2080–5. doi:10.2527/1998.7682080x.
204. Johnson JS, Aardsma MA, Duttlinger AW, Kpodo KR. Early life thermal stress: Impact on future thermotolerance, stress response, behavior, and intestinal morphology in piglets exposed to a heat stress challenge during simulated transport. *J Anim Sci.* 2018;96:1640–53.
205. Li Y, Song Z, Kerr KA, Moeser AJ. Chronic social stress in pigs impairs intestinal barrier and nutrient transporter function, and alters neuro-immune mediator and receptor expression. *PLoS One.* 2017;12.
206. Lallès JP, Boudry G, Favier C, Le Floc’h N, Luron I, Montagne L, et al. Gut function and dysfunction in young pigs: Physiology. *Anim Res.* 2004;53:301–16.
207. Wurtz KE, Siegford JM, Bates RO, Ernst CW, Steibel JP. Estimation of genetic parameters for lesion scores and growth traits in group-housed pigs. *J Anim Sci.* 2017;95:4310–7. doi:10.2527/jas2017.1757.
208. Kick AR, Tompkins MB, Flowers WL, Whisnant CS, Almond GW. Effects of stress associated with weaning on the adaptive immune system in pigs. *J Anim Sci.* 2012;90:649–56. doi:10.2527/jas.2010-3470.
209. Tuchscherer M, Kanitz E, Puppe B, Tuchscherer A, Viergutz T. Changes in endocrine and immune responses of neonatal pigs exposed to a psychosocial stressor. *Res Vet Sci.* 2009;87:380–8.
210. Turner SP, Roehe R, D’Eath RB, Ison SH, Farish M, Jack MC, et al. Genetic validation of postmixing skin injuries in pigs as an indicator of aggressiveness and the relationship with injuries under more stable social conditions. *J Anim Sci.* 2009;87:3076–82.
211. Turner SP, White IMS, Brotherstone S, Farnworth MJ, Knap PW, Penny P, et al. Heritability of post-mixing aggressiveness in grower-stage pigs and its relationship with production traits. *Anim Sci.* 2006;82:615–20.
212. Fernandez X, Meunier-Salaün MC, Mormede P. Agonistic behavior, plasma stress hormones, and metabolites in response to dyadic encounters in domestic pigs: Interrelationships and effect of dominance status. *Physiol Behav.* 1994;56:841–7.
213. Morrow-Tesch JL, McGlone JJ, Salak-Johnson JL. Heat and social stress effects on pig immune measures. *J Anim Sci.* 1994;72:2599–609. doi:10.2527/1994.72102599x.
214. Moeser AJ, Pohl CS, Rajput M. Weaning stress and gastrointestinal barrier development:

Implications for lifelong gut health in pigs. *Animal Nutrition*. 2017;3:313–21.

215. González-Recio O, Toro MA, Bach A. Past, present, and future of epigenetics applied to livestock breeding. *Front Genet*. 2015;6:305. doi:10.3389/fgene.2015.00305.

216. Oberlander TF, Weinberg J, Papsdorf M, Grunau R, Misri S, Devlin AM. Prenatal exposure to maternal depression, neonatal methylation of human glucocorticoid receptor gene (NR3C1) and infant cortisol stress responses. *Epigenetics*. 2008;3:97–106.

217. Perroud N, Paoloni-Giacobino A, Prada P, Olié E, Salzmann A, Nicastro R, et al. Increased methylation of glucocorticoid receptor gene (NR3C1) in adults with a history of childhood maltreatment: A link with the severity and type of trauma. *Transl Psychiatry*. 2011;1. doi:10.1038/tp.2011.60.

218. Eden E, Navon R, Steinfeld I, Lipson D, Yakhini Z. GOrilla: A tool for discovery and visualization of enriched GO terms in ranked gene lists. *BMC Bioinformatics*. 2009;10. doi:10.1186/1471-2105-10-48.

219. Blecha F, Pollmann DS, Nichols DA. Weaning pigs at an early age decreases cellular immunity. 1983. doi:10.2527/jas1983.562396x.

220. Niekamp SR, Sutherland MA, Dahl GE, Salak-Johnson JL. Immune responses of piglets to weaning stress: Impacts of photoperiod. *J Anim Sci*. 2007;85:93–100. doi:10.2527/jas.2006-153.

221. Littlejohn BP, Price DM, Neuendorff DA, Carroll JA, Vann RC, Riggs PK, et al. Prenatal transportation stress alters genome-wide DNA methylation in suckling Brahman bull calves. *J Anim Sci*. 2018;96:5075–99. doi:10.1093/jas/sky350.

222. Cilkiz KZ, Baker EC, Riggs PK, Littlejohn BP, Long CR, Welsh TH, et al. Genome-wide DNA methylation alteration in prenatally stressed Brahman heifer calves with the advancement of age. *Epigenetics*. 2020;:1–18. doi:10.1080/15592294.2020.1805694.

223. Belot MP, Castell AL, Le Fur S, Bougnères P. Dynamic demethylation of the IL2RA promoter during in vitro CD4⁺ T cell activation in association with IL2RA expression. *Epigenetics*. 2018;13:459–72.

224. Nair A, Hunzeker J, Bonneau RH. Modulation of microglia and CD8⁺ T cell activation during the development of stress-induced herpes simplex virus type-1 encephalitis. *Brain Behav Immun*. 2007;21:791–806.

225. Ashcraft KA, Hunzeker J, Bonneau RH. Psychological stress impairs the local CD8⁺ T cell response to mucosal HSV-1 infection and allows for increased pathogenicity via a glucocorticoid receptor-mediated mechanism. *Psychoneuroendocrinology*. 2008;33:951–63. doi:10.1016/j.psyneuen.2008.04.010.

226. de Kloet ER, Vreugdenhil E, Oitzl MS, Joëls M. Brain Corticosteroid Receptor Balance in Health and Disease 1. *Endocr Rev*. 1998;19:269–301.

227. Corbett RJ, te Pas MFW, van den Brand H, Groenen MAM, Crooijmans RPMA, Ernst CW, et al. Genome-Wide Assessment of DNA Methylation in Chicken Cardiac Tissue Exposed to Different Incubation Temperatures and CO₂ Levels. *Front Genet.* 2020;11. doi:10.3389/fgene.2020.558189.
228. Ricklefs RE. Comparative analysis of avian embryonic growth. *J Exp Zool.* 1987;244 SUPPL. 1:309–23. <http://www.ncbi.nlm.nih.gov/pubmed/3598499>. Accessed 21 Jan 2020.
229. French NA. Effect of incubation temperature on the gross pathology of turkey embryos. *Br Poult Sci.* 1994;35:363–71.
230. Christensen VL, Donaldson WE, Nestor KE. Length of the plateau and pipping stages of incubation affects the physiology and survival of turkeys. *Br Poult Sci.* 1999;40:297–303.
231. Lourens A, van den Brand H, Meijerhof R, Kemp B. Effect of eggshell temperature during incubation on embryo development, hatchability, and posthatch development. *Poult Sci.* 2005;84:914–20. doi:10.1093/ps/84.6.914.
232. Leksrisompong N, Romero-Sanchez H, Plumstead PW, Brannan KE, Brake J. Broiler Incubation. 1. Effect of Elevated Temperature During Late Incubation on Body Weight and Organs of Chicks. *Poult Sci.* 2007;86:2685–91. doi:10.3382/ps.2007-00170.
233. Molenaar R, Meijerhof R, van den Anker I, Heetkamp MJW, van den Borne JJGC, Kemp B, et al. Effect of eggshell temperature and oxygen concentration on survival rate and nutrient utilization in chicken embryos. *Poult Sci.* 2010;89:2010–21. doi:10.3382/ps.2010-00787.
234. Maatjens CM, Reijrink IAM, Molenaar R, van der Pol CW, Kemp B, van den Brand H. Temperature and CO₂ during the hatching phase. I. Effects on chick quality and organ development. *Poult Sci.* 2014;93:645–54. doi:10.3382/ps.2013-03490.
235. Romanoff AL. *The Avian Embryo: Structural and Functional Development.* Macmillan, New York; 1960.
236. Kühn ER, Decuypere E, Colen LM, Chadwick A, Heyns W, Michels H. 59 Orcadian Rhythm of Corticosterone, Prolactin and Thyroid Hormones in Serum and Endocrine Glands of Post-hatch Chicks, Incubated at Different Temperatures. In: *Chronobiology 1982-1983.* S. Karger AG; 2015. p. 336–44. doi:10.1159/000410999.
237. Yahav S, Straschnow A, Plavnik I, Hurwitz S. Effects of diurnally cycling versus constant temperatures on chicken growth and food intake. *Br Poult Sci.* 1996;37:43–54.
238. Molenaar R, Hulet R, Meijerhof R, Maatjens CM, Kemp B, van den Brand H. High eggshell temperatures during incubation decrease growth performance and increase the incidence of ascites in broiler chickens. *Poult Sci.* 2011;90:624–32. doi:10.3382/ps.2010-00970.
239. Sozcu A, Ipek A. Acute and chronic eggshell temperature manipulations during hatching term influence hatchability, broiler performance, and ascites incidence. *Poultry Science.* 2015;94:319–27. doi:10.3382/ps/peu080.

240. Hassanzadeh M, Buyse J, Decuypere E. Further evidence for the involvement of cardiac β -adrenergic receptors in right ventricle hypertrophy and ascites in broiler chickens. *Avian Pathol.* 2002;31:177–81. doi:10.1080/03079450120118676.
241. Buys N, Dewil E, Gonzales E, Decuypere E. Avian Pathology Different CO₂ levels during incubation interact with hatching time and ascites susceptibility in two broiler lines selected for different growth rate Different CO₂ levels during incubation interact with hatching time and ascites susceptibility. *Avian Pathol.* 1998;27:605–12. doi:10.1080/03079459808419391.
242. Nichelmann M. Perinatal epigenetic temperature adaptation in avian species: Comparison of turkey and Muscovy duck. *J Therm Biol.* 2004;29:613–9. doi:10.1016/j.jtherbio.2004.08.032.
243. Jaenisch R, Bird A. Epigenetic regulation of gene expression: How the genome integrates intrinsic and environmental signals. *Nat Genet.* 2003;33:245–54. doi:10.1038/ng1089.
244. Ball MP, Li JB, Gao Y, Lee J-H, Leproust E, Park I-H, et al. Targeted and genome-scale methylomics reveals gene body signatures in human cell lines HHS Public Access Author manuscript. *Nat Biotechnol.* 2009;27:361–8. doi:10.1038/nbt.1533.
245. Skibieli AL, Peñagaricano F, Amorín R, Ahmed BM, Dahl GE, Laporta J. In Utero Heat Stress Alters the Offspring Epigenome. *Sci Rep.* 2018;8.
246. Orozco LD, Morselli M, Rubbi L, Guo W, Go J, Shi H, et al. Epigenome-wide association of liver methylation patterns and complex metabolic traits in mice. *Cell Metab.* 2015;21:905–17.
247. Flanagan JM. Epigenome-wide association studies (EWAS): past, present, and future. In: *Methods in molecular biology* (Clifton, N.J.). Humana Press, New York, NY; 2015. p. 51–63. doi:10.1007/978-1-4939-1804-1_3.
248. Lee I, Rasoul BA, Holub AS, Lejeune A, Enke RA, Timp W. Whole genome DNA methylation sequencing of the chicken retina, cornea and brain. *Sci Data.* 2017;4:170148. doi:10.1038/sdata.2017.148.
249. Li J, Li R, Wang Y, Hu X, Zhao Y, Li L, et al. Genome-wide DNA methylome variation in two genetically distinct chicken lines using MethylC-seq. *BMC Genomics.* 2015;16:851. doi:10.1186/s12864-015-2098-8.
250. Zhang Z, Du H, Bai L, Yang C, Li Q, Li X, et al. Whole genome bisulfite sequencing reveals unique adaptations to high-altitude environments in Tibetan chickens. *PLoS One.* 2018;13:e0193597. doi:10.1371/journal.pone.0193597.
251. Zhang M, Yan F-B, Li F, Jiang K-R, Li D-H, Han R-L, et al. Genome-wide DNA methylation profiles reveal novel candidate genes associated with meat quality at different age stages in hens. *Sci Rep.* 2017;7:45564. doi:10.1038/srep45564.
252. Guo W, Fiziev P, Yan W, Cokus S, Sun X, Zhang MQ, et al. BS-Seeker2: a versatile aligning pipeline for bisulfite sequencing data. 2013. doi:10.1186/1471-2164-14-774.

253. Guo W, Zhu P, Pellegrini M, Zhang MQ, Wang X, Ni Z. CGmapTools improves the precision of heterozygous SNV calls and supports allele-specific methylation detection and visualization in bisulfite-sequencing data. *Bioinformatics*. 2018;34:381–7. doi:10.1093/bioinformatics/btx595.
254. Hassanpour H, Bahadoran S, Farhadfar F, Fallahi Chamali Z, Nazari H, Kaewduangta W. Identification of reliable reference genes for quantitative real-time PCR in lung and heart of pulmonary hypertensive chickens. *Poult Sci*. 2018;97:4048–56. doi:10.3382/ps/pey258.
255. Meissner A, Mikkelsen TS, Gu H, Wernig M, Hanna J, Sivachenko A, et al. Genome-scale DNA methylation maps of pluripotent and differentiated cells. *Nature*. 2008;454:766–70. doi:10.1038/nature07107.
256. Jeon BN, Kim MK, Yoon JH, Kim MY, An H, Noh HJ, et al. Two ZNF509 (ZBTB49) isoforms induce cell-cycle arrest by activating transcription of p21/CDKN1A and RB upon exposure to genotoxic stress. *Nucleic Acids Res*. 2014;42:11447–61.
257. Liu XS, Haines JE, Mehanna EK, Genet MD, Ben-Sahra I, Asara JM, et al. ZBTB7A acts as a tumor suppressor through the transcriptional repression of glycolysis. *Genes Dev*. 2014;28:1917–28.
258. Zhao P, Hoffman EP. Myosin isoforms and repression of MyoD in muscle regeneration. *Biochem Biophys Res Commun*. 2006;342:835–42.
259. Xavier-Neto J, Sousa Costa ÂM, Figueira ACM, Caiaffa CD, Amaral FN do, Peres LMC, et al. Signaling through retinoic acid receptors in cardiac development: Doing the right things at the right times. *Biochim Biophys Acta - Gene Regul Mech*. 2015;1849:94–111.
260. Lu F, Langenbacher A, Chen JN. Tbx20 drives cardiac progenitor formation and cardiomyocyte proliferation in zebrafish. *Dev Biol*. 2017;421:139–48.
261. Ye B, Li L, Xu H, Chen Y, Li F. Opposing roles of TCF7/LEF1 and TCF7L2 in cyclin D2 and Bmp4 expression and cardiomyocyte cell cycle control during late heart development. *Lab Invest*. 2019;99:807–18.
262. Krcmery J, Gupta R, Sadleir RW, Ahrens MJ, Misener S. Loss of the Cytoskeletal Protein Pdlim7 Predisposes Mice to Heart Defects and Hemostatic Dysfunction. *PLoS One*. 2013;8:80809. doi:10.1371/journal.pone.0080809.
263. Engelmann GL, Boehm KD, Birchenall-Roberts MC, Ruscetti FW. Transforming growth factor-beta1 in heart development. *Mech Dev*. 1992;38:85–97.
264. Hannigan GE, Coles JG, Dedhar S. Integrin-linked kinase at the heart of cardiac contractility, repair, and disease. *Circ Res*. 2007;100:1408–14. doi:10.1161/01.RES.0000265233.40455.62.
265. Nangsuay A, Meijerhof R, Van Den Anker I, Heetkamp MJW, Morita VDS, Kemp B, et al. Effects of breeder age, broiler strain, and eggshell temperature on development and physiological

- status of embryos and hatchlings. *Poultry Science*. 2016;95:1666–79. doi:10.3382/ps/pew080.
266. Derks MFL, Schachtschneider KM, Madsen O, Schijlen E, Verhoeven KJF, van Oers K. Gene and transposable element methylation in great tit (*Parus major*) brain and blood. *BMC Genomics*. 2016;17:332. doi:10.1186/s12864-016-2653-y.
267. Chen L, Fulcoli FG, Tang S, Baldini A. *Tbx1* regulates proliferation and differentiation of multipotent heart progenitors. *Circ Res*. 2009;105:842–51. doi:10.1161/CIRCRESAHA.109.200295.
268. Xu H, Morishima M, Wylie JN, Schwartz RJ, Bruneau BG, Lindsay EA, et al. *Tbx1* has a dual role in the morphogenesis of the cardiac outflow tract. *Development*. 2004;131:3217–27. doi:10.1242/dev.01174.
269. Krause A, Zacharias W, Camarata T, Linkhart B, Law E, Lischke A, et al. *Tbx5* and *Tbx4* transcription factors interact with a new chicken PDZ-LIM protein in limb and heart development. *Dev Biol*. 2004;273:106–20. doi:10.1016/J.YDBIO.2004.05.024.
270. Gloss B, Trost SU, Bluhm WF, Swanson EA, Clark R, Winkfein R, et al. Cardiac ion channel expression and contractile function in mice with deletion of thyroid hormone receptor α or β . *Endocrinology*. 2001;142:544–50. doi:10.1210/endo.142.2.7935.
271. Marelli F, Carra S, Rurale G, Cotelli F, Persani L. *In vivo* Functional Consequences of Human *THRA* Variants Expressed in the Zebrafish. *Thyroid*. 2017;27:279–91. doi:10.1089/thy.2016.0373.
272. Ren J, Xu N, Zheng H, Tian W, Li H, Li Z, et al. Expression of Thyroid Hormone Responsive SPOT 14 Gene Is Regulated by Estrogen in Chicken (*Gallus gallus*). *Sci Rep*. 2017;7:10243. doi:10.1038/s41598-017-08452-6.
273. Wellberg EA, Rudolph MC, Lewis AS, Padilla-Just N, Jedlicka P, Anderson SM. Modulation of tumor fatty acids, through overexpression or loss of thyroid hormone responsive protein spot 14 is associated with altered growth and metastasis. *Breast Cancer Res*. 2014;16.
274. Morikawa Y, Heallen T, Leach J, Xiao Y, Martin JF. Dystrophin Glycoprotein Complex Sequesters Yap to Inhibit Cardiomyocyte Proliferation HHS Public Access. *Nature*. 2017;547:227–31. doi:10.1038/nature22979.
275. Franco D, Sedmera D, Lozano-Velasco E. Multiple Roles of *Pitx2* in Cardiac Development and Disease. *J Cardiovasc Dev Dis*. 2017;4. doi:10.3390/jcdd4040016.
276. Tao G, Kahr PC, Morikawa Y, Zhang M, Rahmani M, Heallen TR, et al. *Pitx2* promotes heart repair by activating the antioxidant response after cardiac injury. *Nature*. 2016;534:119–23.
277. Zhao YY, Sawyer DR, Baliga RR, Opel DJ, Han X, Marchionni MA, et al. Neuregulins promote survival and growth of cardiac myocytes: Persistence of ErbB2 and ErbB4 expression in neonatal and adult ventricular myocytes. 1998. doi:10.1074/jbc.273.17.10261.

278. Van Rooy-Reijrink IAM. Incubation affects chick quality. 2016. <https://www.poultryworld.net/Genetics/Articles/2013/5/Incubation-affects-chick-quality-1183725W/>. Accessed 10 Aug 2020.
279. Salnikow K, Aprelikova O, Ivanov S, Tackett S, Kaczmarek M, Karaczyn A, et al. Regulation of hypoxia-inducible genes by ETS1 transcription factor. *Carcinogenesis*. 2008;29:1493–9.
280. Seko Y, Tobe K, Takahashi N, Kaburagi Y, Kadowaki T, Yazaki Y. Hypoxia and hypoxia/reoxygenation activate Src family tyrosine kinases and p21(ras) in cultured rat cardiac myocytes. *Biochem Biophys Res Commun*. 1996;226:530–5.
281. Shi D, Xie D, Zhang H, Zhao H, Huang J, Li C, et al. Reduction in dynamin-2 is implicated in ischaemic cardiac arrhythmias. *J Cell Mol Med*. 2014;18:1992–9.
282. Barrientos T, Frank D, Kuwahara K, Bezprozvannaya S, Pipes GCT, Bassel-Duby R, et al. Two novel members of the ABLIM protein family, ABLIM-2 and -3, associate with STARS and directly bind F-actin. *J Biol Chem*. 2007;282:8393–403. doi:10.1074/jbc.M607549200.
283. Sun L, Lamont SJ, Cooksey AM, McCarthy F, Tudor CO, Vijay-Shanker K, et al. Transcriptome response to heat stress in a chicken hepatocellular carcinoma cell line. *Cell Stress Chaperones*. 2015;20:939–50.
284. Wang C, Wang M, Arrington J, Shan T, Yue F, Nie Y, et al. Ascl2 inhibits myogenesis by antagonizing the transcriptional activity of myogenic regulatory factors. *Dev*. 2017;144:235–47.
285. Cai Z, Guldbrandtsen B, Lund MS, Sahana G. Prioritizing candidate genes post-GWAS using multiple sources of data for mastitis resistance in dairy cattle. *BMC Genomics*. 2018;19. doi:10.1186/S12864-018-5050-X.
286. Ballester M, Ramayo-Caldas Y, Revilla M, Corominas J, Castelló A, Estellé J, et al. Integration of liver gene co-expression networks and eGWAs analyses highlighted candidate regulators implicated in lipid metabolism in pigs. *Sci Rep*. 2017;7:46539. doi:10.1038/srep46539.
287. MacLeod IM, Bowman PJ, Jagt CJ Vander, Haile-Mariam M, Kemper KE, Chamberlain AJ, et al. Exploiting biological priors and sequence variants enhances QTL discovery and genomic prediction of complex traits. *BMC Genomics*. 2016;17. doi:10.1186/S12864-016-2443-6.
288. Keel BN, Snelling WM, Lindholm-Perry AK, Oliver WT, Kuehn LA, Rohrer GA. Using SNP Weights Derived From Gene Expression Modules to Improve GWAS Power for Feed Efficiency in Pigs. *Front Genet*. 2019;10:1. doi:10.3389/FGENE.2019.01339.
289. Boudry G, Péron V, Le Huërou-Luron I, Lallès JP, Sève B. Weaning Induces Both Transient and Long-Lasting Modifications of Absorptive, Secretory, and Barrier Properties of Piglet Intestine. *J Nutr*. 2004;134:2256–62.

290. Baarendse PJJ, Kemp B, Van Den Brand H. Early-age housing temperature affects subsequent broiler chicken performance. *Br Poult Sci.* 2006;47:125–30.
doi:10.1080/00071660600610575.