

Characterization of Nucleosome-Depleted Regions of Transcriptionally Active Chromatin in Chicken and Human Genomes

by

Tasnim H. Beacon

A Thesis submitted to the Faculty of Graduate Studies
of The University of Manitoba
in partial fulfilment of the requirements of the degree of

MASTER OF SCIENCE
2020

Department of Biochemistry and Medical Genetics
Max Rady College of Medicine, Rady Faculty of Health Sciences
University of Manitoba
Winnipeg

Copyright © 2020 by Tasnim H. Beacon

Abstract

The mammalian genome is organized into spatial domains at multiple scales such as compartments, topologically associating domains and chromatin loops. Such folding of the genome is essential to modulate gene expression at the structural and functional level. Histone post-translational modifications (acetylated histones, H3K4me3 and H3K27ac) associated with transcriptionally active chromatin regions are referred to as active histone marks and are often associated with the active compartment (compartment A). Current evidence suggests that protein arginine-N-methyl transferases and their catalyzed histone post-translational modifications, H3R2me2s and H4R3me2a, are also associated with compartment A and are involved in establishing and maintaining the transcriptionally active chromatin state in vertebrate genomes.

DNase I sensitivity and dynamic histone acetylation are features of transcribed chromatin. Within the DNase I sensitive region are DNase I hypersensitive sites which are devoid of nucleosomes and mark the presence of regulatory elements (e.g. promoters and enhancers) in the genome. Nucleosomes in the regulatory regions are destabilized by histone post-translational modifications and chromatin remodelling enzymes. Therefore, to understand the organization of transcriptionally active chromatin architecture, it is crucial to characterize the nucleosome-depleted regions of the genome in relation to nucleosome dynamics. For my project, I am investigating the features of nucleosome depleted regions in the chicken and human genomes. I have applied the technique of formaldehyde-assisted isolation of regulatory elements (FAIRE) coupled with next-generation sequencing to map out the position of the regulatory regions in relation to the active marks. FAIRE-sequencing was compared to chromatin immunoprecipitation-sequencing of H3K4me3, H3K27ac, H3R2me2s and H4R3me2a. Based on the sequenced data,

FAIRE-sequencing readily identified all of the promoter regions in chicken polychromatic erythrocytes. Further I identified regions that had an atypical chromatin structure in avian and human cells, and in these regions were genes involved in cell identity. Chromatin signatures characterized in active genes, which were involved in different biological processes, were comparable to the chromatin features observed in human genome.

Acknowledgement

When I came to write this section of my thesis the only proverb that came to my mind was that “it takes a village to raise a child” which in my case translates to a large group of individuals without whom my project would have never been accomplished. Although there were obstacles (e.g. firebrands and covid-19), I am grateful to the kind souls who helped me through this.

First and foremost, I would like to express my appreciation and gratitude to my supervisor Dr. Jim Davie who was not only a great mentor but has always treated me like his colleague to help me excel in my project and direct me to great ideas. With his helpful suggestions and comments, I have expanded my knowledge of epigenetics as well as participated in scientific curriculums that helped me build my CV. I would also like to thank my committee members Dr. Ted Lakowski and Dr. Etienne Leygue for their valuable questions and comments. Dr. Jeff Wiggle for guiding me through the BMG timeline. Philip Dufresne and Chloe Lepage for their administrative support. I am also thankful to my previous lab members Aleksandar Ilic, Veronica Lau, Debbie Tsuyuki and Ms. Cheryl Peltier for their expert guidance with my sample preparation and experimental technicalities. Research Manitoba for funding my research. I also want to thank Central animal care facility for their technical help with my chicken blood collection and the BMG family for their immense support in the span of 2 years that felt like it went by in no time.

Last but not the least I would like to thank my friends and family, especially my brother, without his support I could not excel as much as I did now.

TABLE OF CONTENTS

| | |
|--|-------------|
| Abstract | i |
| Acknowledgements | iii |
| List of Abbreviations | xi |
| List of Tables | xvi |
| List of Figures..... | xvii |
| List of Permitted Copyrighted Materials | xxi |
| CHAPTER 1: INTRODUCTION | 1 |
| 1.1 What is Epigenetics?..... | 2 |
| 1.2 Chromatin Composition..... | 3 |
| 1.3 Regulation of Chromatin Structure..... | 4 |
| 1.3.1 Open Chromatin/ Nucleosome Depleted Region (NDR)..... | 8 |
| 1.4 Active Histone PTMs of Promoters and Enhancers..... | 10 |
| 1.4.1 Lysine Methylation..... | 10 |
| 1.4.2 H3K4me3..... | 12 |
| 1.4.3 Histone Acetylation..... | 13 |
| 1.4.4 H3K27ac..... | 14 |
| 1.5 PRMT Enzymes: Role in Transcriptionally Active Chromatin..... | 15 |
| 1.5.1 H3R2me2s..... | 16 |
| 1.5.2 H4R3me2a..... | 17 |
| 1.6 Epigenetic Methods..... | 17 |
| 1.7 Innate Immunity..... | 20 |
| 1.7.1 Epigenetic Regulators of Innate Immunity..... | 22 |

| | |
|---|-----------|
| 1.7.2 Avian Innate Immunity..... | 24 |
| 1.8 Chicken Model System..... | 26 |
| 1.8.1 Genetic Makeup..... | 26 |
| 1.8.2 Erythropoiesis..... | 29 |
| 1.8.3 Chicken Erythrocyte Chromatin Structure..... | 30 |
| 1.8.4 β-Globin Domain..... | 32 |
| 1.8.5 Three-Dimensional Organization of the Chicken Genome..... | 33 |
| 1.8.6 Chicken DNA Methylome..... | 35 |
| 1.8.7 Poultry Epigenetics..... | 37 |
| 1.9.1 Hypotheses..... | 38 |
| 1.9.2 Aims..... | 38 |
| 1.9.3 Research Strategy..... | 39 |
| 2.0 References..... | 40 |
| CHAPTER 2: Methods & Materials..... | 55 |
| 2.1 Chicken Erythrocyte Processing..... | 56 |
| 2.1.1 Ethics statement..... | 56 |
| 2.1.2 Treatment of chickens to induce anemia..... | 56 |
| 2.1.3 Blood collection and polychromatic erythrocyte collection..... | 57 |
| 2.2 Erythrocyte Treatment and Assays..... | 58 |
| 2.2.1 Adenosine dialdehyde treatment..... | 58 |
| 2.2.2 Sodium butyrate treatment..... | 59 |
| 2.3 Protein quantification and standardization..... | 59 |
| 2.4 Protein Analysis and Electrophoretic Methods..... | 60 |

| | |
|--|-----------|
| 2.4.1 Immunoblotting antibody list..... | 60 |
| 2.4.2 SDS-PAGE electrophoresis and immunoblotting..... | 60 |
| 2.4.3 Acetic acid-urea-Triton-X100 (AUT) polyacrylamide gel electrophoresis..... | 62 |
| 2.4.4 1% DNA agarose gel..... | 64 |
| 2.4.5 Histone extraction..... | 65 |
| 2.5 Staining Methods..... | 66 |
| 2.5.1 Ponceau staining..... | 66 |
| 2.5.2 Coomassie blue staining..... | 66 |
| 2.5.3 8-Anilino-1-naphthalenesulfonate (ANS) staining..... | 67 |
| 2.6 FAIRE (Formaldehyde assisted isolation of regulatory region)..... | 67 |
| 2.6.1 FAIRE DNA Sample Purification using Agencourt AMPure XP Beads..... | 69 |
| 2.7 Quantitative Polymerase Chain Reaction (qPCR)..... | 71 |
| 2.8 FAIRE-seq DNA Library Preparation and Sequencing..... | 72 |
| 2.9 Bioinformatic Analysis..... | 72 |
| 2.10 References..... | 76 |
| CHAPTER 3: Dynamic histone acetylation and impact on histone PTMs and variants..... | 77 |
| 3.1 Introduction..... | 78 |
| 3.1.1 Active Histone Post-Translational Modifications..... | 78 |
| 3.1.2 Lysine Acetylation..... | 80 |
| 3.1.3 Dynamic Histone Acetylation..... | 81 |
| 3.1.4 Histone Variants..... | 82 |
| 3.1.5 Study Design and Objective..... | 83 |
| 3.2 Methods and Materials..... | 83 |

| | |
|--|-----------|
| 3.2.1. Ethics statement..... | 83 |
| 3.2.2. Preparation of chicken polychromatic erythrocytes and sodium butyrate treatment..... | 84 |
| 3.2.3 Protein analysis and immunoblotting..... | 84 |
| 3.3 Results..... | 85 |
| 3.3.1 H3 Tail Modifications..... | 85 |
| 3.3.2 H4 Tail Modifications..... | 89 |
| 3.4 Discussion..... | 91 |
| 3.5 References..... | 95 |
| CHAPTER 4: Chromatin organization of transcribed genes in chicken polychromatic erythrocytes..... | 99 |
| 4.1 Abstract..... | 101 |
| 4.2 Introduction..... | 102 |
| 4.3. Materials and methods..... | 104 |
| 4.3.1 Ethics statement..... | 104 |
| 4.3.2 Isolation of chicken erythrocytes..... | 105 |
| 4.3.3 Chromatin fractionation..... | 105 |
| 4.3.4 ChIP-sequencing..... | 105 |
| 4.3.5 FAIRE..... | 106 |
| 4.3.6 Quantitative PCR analysis..... | 106 |
| 4.3.7 Sequencing and data analyses..... | 107 |
| 4.3.8 Statistical analysis..... | 107 |
| 4.3.9 Heatmap analysis..... | 107 |

| | |
|---|------------|
| 4.4 Results and Discussion..... | 108 |
| 4.4.1 Carbonic anhydrase (CA2) gene on macrochromosome 2..... | 108 |
| 4.4.2 Ferritin, heavy polypeptide 1 (FTH1) locus on macrochromosome 5..... | 110 |
| 4.4.3 β-Globin locus on macrochromosome 1..... | 112 |
| 4.4.4 Histone H5 (H1F0) locus on macrochromosome 1..... | 114 |
| 4.4.5 α-Globin (HBA) locus on microchromosome 14..... | 116 |
| 4.4.6 ARIHI and NCOA4 genes..... | 118 |
| 4.4.7 Open chromatin and modified nucleosomes..... | 120 |
| 4.4.8 FAIRE and CpG Islands (CGIs)..... | 120 |
| 4.5 Conclusion..... | 122 |
| 4.6 Conflict of interest..... | 123 |
| 4.7 Funding..... | 123 |
| 4.8 References..... | 124 |
| 4.9 Supplementary Data..... | 129 |
| CHAPTER 5: Atypical chromatin structure of immune-related genes expressed in chicken erythrocytes..... | 130 |
| 5.1 Abstract..... | 132 |
| 5.2 Introduction..... | 133 |
| 5.3 Materials and methods..... | 135 |
| 5.3.1 Ethics statement..... | 135 |
| 5.3.2 Isolation of chicken erythrocytes and treatments..... | 135 |
| 5.3.3 RNA extraction and RT-PCR..... | 136 |
| 5.3.4 ChIP-sequencing..... | 136 |

| | |
|---|------------|
| 5.3.5 Chromatin fractionation, RNA sequencing, and FAIRE sequencing..... | 137 |
| 5.3.6 Bioinformatics data analyses..... | 137 |
| 5.3.7 Statistical analysis..... | 138 |
| 5.4 Results..... | 138 |
| 5.4.1 Polychromatic erythrocytes from chickens express genes involved in the immune response..... | 138 |
| 5.4.2 Genomic features of genes involved in the immune response..... | 140 |
| 5.4.3 Atypical chromatin organization of genes involved in the immune response... | 141 |
| 5.4.4 Nucleosome instability in human gene bodies associated with H3K27ac..... | 145 |
| 5.5 Discussion..... | 147 |
| 5.6 Acknowledgements..... | 150 |
| 5.7 References..... | 151 |
| 5.8 Supplementary Data..... | 156 |
| CHAPTER 6: Genomic landscape of transcriptionally active histone arginine methylation marks, H3R2me2s and H4R3me2a, relative to nucleosome depleted regions..... | 168 |
| 6.1 Abstract..... | 170 |
| 6.2 Introduction..... | 171 |
| 6.3 Materials and methods..... | 173 |
| 6.3.1 Ethics statement..... | 173 |
| 6.3.2 Preparation of chicken polychromatic erythrocytes and adenosine dialdehyde treatment..... | 174 |
| 6.3.3 Chromatin immunoprecipitation assay and sequencing..... | 174 |
| 6.3.4 Bioinformatics analyses..... | 175 |

| | |
|---|------------|
| 6.4 Results..... | 176 |
| 6.4.1 Association of PRMT1, PRMT5, H4R3me2a and H3R2me2s with transcribed chromatin..... | 176 |
| 6.4.2 Combinatory patterns of histone PTMs..... | 185 |
| 6.4.3 Dynamic methylation of histones..... | 187 |
| 6.5 Discussion..... | 187 |
| 6.6 Declaration of Competing Interest..... | 191 |
| 6.7 Acknowledgements..... | 191 |
| 6.8 References..... | 192 |
| 6.9 Supplementary Data..... | 199 |
| CHAPTER 7: Investigation of H3K4me3 broad domains in different cell types reveal importance of long chromatin signatures in human and chicken genomes..... | 202 |
| 7.1 Introduction..... | 203 |
| 7.2 Methods..... | 204 |
| 7.3 Results..... | 205 |
| 7.4 Discussion..... | 211 |
| 7.5 References..... | 214 |
| CHAPTER 8: Discussion and Future direction..... | 216 |
| 8.1 Study insight and future direction..... | 217 |
| 8.2 Study limitation and challenges..... | 228 |
| 8.3 Significance of the project..... | 229 |
| 8.4 References..... | 231 |

LIST OF ABBREVIATIONS

| | |
|----------|---|
| AdOx | Adenosine dialdehyde |
| AEV | Avian erythroblastosis virus |
| ATAC | Assay for Transposase-Accessible Chromatin using sequencing |
| ATP | Adenosine triphosphate |
| AUT | Acetic acid-urea-triton-x |
| bp | Base pair |
| BRG1 | Brahma-related gene-1 |
| BSA | Bovine serum albumin |
| CBP | CREB-binding protein |
| cDNA | Complementary deoxyribonucleic acid |
| CFU | Colony-Forming Unit (-E) |
| CHD1 | Chromodomain helicase DNA binding protein 1 |
| ChIP | Chromatin immunoprecipitation |
| ChIP-seq | ChIP followed by high-throughput sequencing |
| CGIs | CpG islands |
| 3C | Chromosome Conformation Capture Analysis |
| DNA | Deoxyribonucleic acid |
| DNMTs | DNA methyltransferase enzymes |
| EDTA | (Ethylenedinitrilo) tetraacetic acid |
| ENCODE | Encyclopedia of DNA Elements |

| | |
|----------|---|
| eRNA | Enhancer RNA |
| EZH2 | Enhancer of Zeste 2 Polycomb Repressive Complex 2 Subunit |
| FAIRE | Formaldehyde assisted isolation of regulatory elements |
| FBS | Fetal bovine serum |
| GEO | Gene Expression Omnibus |
| GRCg6a | Gallus gallus (chicken) genome assembly 6 |
| H | Hours |
| HDAC | Histone deacetylase |
| HRP | Horseradish peroxidase |
| H3K9ac | Histone H3 acetylation on lysine 9 |
| H3K9me3 | Histone H3 trimethylation on lysine 9 |
| H3K4me3 | Histone H3 trimethylation on lysine 4 |
| H3K36me3 | Histone H3 trimethylation on lysine 36 |
| H3Ac | Histone H3 acetylation on lysine 9 and 14 |
| H3R2me2s | Histone H3 dimethyl symmetric |
| H4R3me2a | Histone H4 dimethyl asymmetric |
| H4K20me2 | Histone H4 dimethylation on lysine 20 |
| H4Ac | H4 penta hyperacetylated |
| IB | Immunoblot |
| IEG | Immediate-early gene |
| IFNs | Interferones |
| IGV | Integrated Genomics Viewer |
| IL | Interleukin |

| | |
|----------------|---|
| IP | Immunoprecipitated fraction |
| ISGs | Interferon stimulated genes |
| JARID | Jumonji domain ARID-containing protein |
| JMJD | Jumonji domain |
| JNK | c-Jun N-terminal kinase |
| K562 | Chronic myeloid leukemia cell line |
| KAT | Lysine acetyltransferases |
| KDa | Kilodalton |
| KMT | Lysine methyltransferase |
| lncRNA | long non-coding RNAs |
| LSD1 | Lysine-specific histone demethylase-1 |
| MAPK | Mitogen-activated protein kinase |
| MBD | Methyl-CpG-binding domain-containing protein |
| MCF7 | Michigan Cancer Foundation-7 |
| MEM | Minimum Essential Medium |
| Min | Minutes |
| miRNA | MicroRNA |
| MLL | Mixed-lineage leukemia gene |
| MNase | Micrococcal nuclease |
| ncRNA | Non-coding RNA |
| NDRs | Nucleosome-depleted regions |
| NF- κ B | Nuclear factor-kappa B |
| NuRD | Nucleosome-remodeling and deacetylase repressor |

| | |
|------------------|---|
| Ocean-C | Open chromatin enrichment and network Hi-C |
| PAGE | Polyacrylamide gel electrophoresis |
| PBS | Phosphate buffered saline |
| PCR | Polymerase chain reaction |
| PCAF | Protein CBP-associated factor |
| P _E | EDTA insoluble chromatin fraction |
| PHD | Plant homeodomain |
| p53 | protein 53 |
| P ₁₅₀ | 150 mM NaCl insoluble chromatin fraction |
| PolyI:C | Polyinosinic:polycytidylic acid |
| PRGs | Primary response genes |
| PRMTs | Protein arginine methyltransferases |
| PTM | Post-translational modification |
| qPCR | quantitative PCR |
| RLR | RIG-I-like receptors |
| RNA | Ribonucleic acid |
| RNA-seq | Next-generation RNA-sequencing |
| RNAPII | RNA polymerase II |
| RT-PCR | Reverse transcription-polymerase chain reaction |
| SAH | S-adenosylhomocysteine |
| SDC | Sodium deoxycholate |
| SDS | Sodium dodecyl sulfate |
| SDS-PAGE | SDS-Polyacrylamide gel electrophoresis |

| | |
|------------------|--|
| Sec | Seconds |
| S _E | EDTA insoluble chromatin fraction |
| S ₁₅₀ | 150 mM NaCl insoluble chromatin fraction |
| SETD | SET Domain |
| siRNA | Small interfering RNA |
| SUV | Suppressor of Variegation |
| SWI/SNF | Switch/sucrose non-fermentable |
| TADs | Topologically associating domains |
| TBS | Tris buffered saline |
| TBST | Tris buffered saline with Tween-20 |
| TLR | TOLL LIKE RECEPTOR |
| Tris | Tris (hydroxymethyl)aminomethane |
| TNF | Tumor necrosis factor |
| UCSC | University of California Santa Cruz |
| UV | Ultraviolet |
| WDR | WD Repeat Domain |

LIST OF TABLES

| | |
|--|-----|
| Supplementary Table 4.1. Primers used in FAIRE-qPCR experiments..... | 129 |
| Table 5.1. Genes participating in the immune response and expressed in chicken polychromatic erythrocytes..... | 139 |
| Supplementary Table 5.S1: Primers used for RT-PCR..... | 156 |
| Supplementary Table 5.S2: Primers used in FAIRE-qPCR experiments..... | 157 |
| Supplementary Table 5.S3. Ranking of genes involved in the immune response according to expression levels in chicken polychromatic erythrocytes..... | 158 |
| Table 7.1: Prediction outcome of the <i>de novo</i> motif investigated using GOMO..... | 211 |

LIST OF FIGURES

| | |
|---|-----|
| Figure 1.1: Chromatin regulating enzymes (reader/ writer/ eraser) modulating N-terminal tail PTMs of the core histones..... | 5 |
| Figure 1.2: List of common histone post-translational modifications and DNA-methylation associated with annotated genes..... | 9 |
| Figure 1.3: Phylogenetic tree showing evolutionary relation and divergence time (in million of years) among vertebrate human, chicken and rodent..... | 28 |
| Figure 1.4: Folding and organization of chromatin at multiple scale..... | 35 |
| Figure 2.1: Key steps summary of FAIRE..... | 70 |
| Figure 2.2: Pipeline generated in Linux to process FASTQ files to produce intermediate files (bam/sam/bed) to bedgraphs or bigwigs that can be visualized using IGV, Partek Flow or UCSC browser..... | 74 |
| Figure 2.3: The bioinformatic pipeline designed to process ChIP-seq and RNA-seq data..... | 75 |
| Figure 3.1: PTMs analyzed on H3-tail..... | 88 |
| Figure 3.2: PTMs analyzed on H4-tail..... | 90 |
| Figure 4.1: Chromatin profile and transcriptional activity of the <i>CA2</i> genomic region..... | 109 |
| Figure 4.2: Chromatin profile and transcriptional activity of the <i>FTH1</i> genomic region..... | 111 |
| Figure 4.3. Chromatin profile of the β -globin genomic region..... | 113 |

| | |
|--|-----|
| Figure 4.4: Chromatin profile and transcriptional activity of the <i>HIF0</i> (H5) genomic region..... | 115 |
| Figure 4.5: Chromatin profile and transcriptional activity of the α -globin..... | 117 |
| Figure 4.6: Chromatin profile and transcriptional activity of the <i>ARIH1</i> and <i>NCOA4</i> genomic regions..... | 119 |
| Figure 4.7: Heatmaps depicting density coverage of FAIRE peaks over the TSS (a) and CGI (b) spanning 5 kb on each side of TSS or CGI from galGal3 database of UCSC Genome Browser..... | 121 |
| Figure 5.1: Chromatin profile of the <i>IRF8</i> gene region..... | 141 |
| Figure 5.2: Chromatin profile of the <i>NFKB2</i> gene region..... | 143 |
| Figure 5.3: Chromatin profile of the <i>TLR21</i> gene region..... | 144 |
| Figure 5.4: Chromatin profile of the human <i>IKZF1</i> tumor suppressor gene region in CD4+ T cells..... | 146 |
| Figure 5.5: Model for nucleosome instability over the gene body of chicken erythrocyte immune genes and human genes involved in cell identity..... | 149 |
| Supplementary Figure 5.S1: Fractionation of avian erythrocyte chromatin..... | 160 |
| Supplementary Figure 5.S2: Poly I:C mediated induction of immune genes in chicken polychromatic erythrocytes..... | 161 |
| Supplementary Figure 5.S3: Chromatin profile of the <i>IRF1</i> gene region..... | 162 |
| Supplementary Figure 5.S4: Chromatin profile of the <i>TLR3</i> gene region..... | 163 |

| | |
|---|-----|
| Supplementary Figure 5.S5. Chromatin profile of the MYD88 gene region..... | 164 |
| Supplementary Figure 5.S6: Chromatin profile of the IRF7 gene region..... | 165 |
| Supplementary Figure 5.S7: Chromatin profile of the <i>IL1B</i> gene region..... | 166 |
| Supplementary Figure 5.S8: Chromatin profile of the human GATA2 gene region in erythroleukemic K562 cells..... | 167 |
| Figure 6.1: Genomic distribution of histone arginine methylated marks..... | 177 |
| Figure 6.2: Categorization of expressed genes associated with histone arginine methylated marks..... | 178 |
| Figure 6.3: Chromatin profile along a 100 kb region of macrochromosome 5 covering highly expressed <i>FTH1</i> and <i>RAB3IL1</i> genes..... | 180 |
| Figure 6.4: Chromatin profile along a 70 kb region of macrochromosome 1 covering highly expressed β -globin (<i>HBBA</i>) gene and β -globin domain..... | 181 |
| Figure 6.5: Chromatin profile along a 145 kb region of microchromosome 11 covering the weakly expressed <i>PRMT7</i> gene..... | 182 |
| Figure 6.6. Profile of H3K4me3, H3K27ac, H4R3me2a and H3R2me2s around CpG islands.. | 183 |
| Figure 6.7: Location of histone arginine methylated marks relative to the TSS and other histone PTMs..... | 184 |
| Figure 6.8: Arginine methylation of histone H3 variants and turnover of methylated histones.. | 186 |
| Figure 6.9: Models for recruitment of PRMT1 to active genes..... | 190 |

| | |
|--|-----|
| Supplementary File 6.1. H3R2me2s and H4R3me2a profiles as a function of gene expression..... | 199 |
| Supplementary Figure 6.2: Chromatin profile along a 35 kb region of macrochromosome 6 covering the weakly expressed NFKB2 gene..... | 200 |
| Supplementary File 6.3: Correlation between histone modifications..... | 201 |
| Figure 7.1: Breadth distribution of H3K4me3 ChIP-seq seq in MDA-MB-231 cells..... | 206 |
| Figure 7.2: Breadth distribution of H3K4me3 ChIP-seq seq in chicken polychromatic erythrocytes..... | 207 |
| Figure 7.3: Immune gene <i>NFKB2</i> of polychromatic erythrocytes depicting H3K4me3 and H3K27ac broad domain in alignment with open chromatin region. Positions of CpG islands are shown..... | 208 |
| Figure 7.4: Comparison of human <i>GATA2</i> gene in K562 cells with polychromatic erythrocytes <i>GATA2</i> depicting similar broad domain features in additional active marks H3R2me2s and H4R3me2a..... | 209 |
| Figure 7.5: H3K27ac ChIA-PET interactome in K562 cell for <i>GATA2</i> aligned with broad domain tracks of H3K4me3 and H3K27ac..... | 210 |
| Figure 8.1: Proposed model for the function of CHD1 in remodelling nucleosomes..... | 222 |
| Figure 8.2: Analysis of different PTMs and open chromatin broad domain..... | 224 |
| Figure 8.3: Snapshot of <i>FTH1</i> gene depicting the tracks of F1-seq, FAIRE-seq and ChIP-seq of H3K4me3, H3K27ac, H3R2me2s and H4R3me2a to show their genomic distribution and their relation to dynamic acetylation..... | 228 |

LIST OF COPYRIGHTED MATERIAL FOR WHICH PERMISSION WAS OBTAINED

Jahan, S., Beacon, T. H., Xu, W., and Davie, J. R. (2020) Atypical chromatin structure of immune-related genes expressed in chicken erythrocytes. *Biochemistry and Cell Biology* 98, 171–177.

Beacon, T. H., Xu, W., and Davie, J. R. (2020) Genomic landscape of transcriptionally active histone arginine methylation marks, H3R2me2s and H4R3me2a, relative to nucleosome depleted regions. *Gene* 742.

Jahan, S., Beacon, T. H., He, S., Gonzalez, C., Xu, W., et al. (2019) Chromatin organization of transcribed genes in chicken polychromatic erythrocytes. *Gene* 699, 80–87.

Sepehri, Z., Beacon, T.H., Osman, F.D.S., Jahan, S. and Davie, J.R. (2019). *DNA Methylation and Chromatin Modifications*. 1st ed. Elsevier Inc., Chapter-2. *Nutritional Epigenomics*.

Chapter-1

Introduction

1.1 What is Epigenetics?

The term epigenetics was first coined by Waddington in 1942 with the intention of understanding the complex interaction of genes, their products and the modifying enzymes and how it resulted into a phenotype (1). With the advance in genomics and molecular techniques, the field is gaining an understanding of the mechanism(s) by which genes are positioned in chromosomes in specific genomes and how the genes function at molecular level. In the era blessed with the advancement of sequencing and bioinformatics, platforms such as Ensemble, Gene Expression Omnibus (GEO), National Center for Biotechnology Information, University of California Santa Cruz (UCSC) Genome Browser and Encyclopedia of DNA Elements (ENCODE) have made it possible to look at different genomes and their gene composition to investigate the base sequence to translated protein and their underlying function and pathway. This has expanded our knowledge to how these genes are controlled and how chromatin is modified. This led to the modern definition of epigenetics as study of changes in gene function that are heritable without inducing a change in the base sequence of DNA. Epigenetic control of gene expression is the key to understanding the underlying chromatin structure in a given state of an epigenome as well as understanding the mechanism of diseases investigated in different fields such as developmental biology, oncology, environmental, nutritional and microbiology.

1.2 Chromatin Composition

Chromatin typically exists in two states: euchromatin and heterochromatin. The less tightly packed or accessible chromatin is the euchromatin form associated with histone post-translational modifications (PTMs)/ active marks (histone H3 trimethylated at lysine 4, H3K4me3, histone H3 acetylated at lysine 27, H3K27ac and histone H3 trimethylated at lysine 1, H3K4me1) whereas heterochromatin is the repressed form mostly associated with repressed marks such as histone H3 tri-/di- methylated at lysine 27 and 9, H3K27me3 and H3K9me2/3. The basic unit that forms the foundation of chromatin is the nucleosome repeat. Each mononucleosome is composed of a histone octamer (two each of the core histones H2A, H2B, H3 and H4) around which is wrapped 147 base pairs of genomic DNAs. The nucleosomes are joined by a stretch of DNA called linker DNA, forming what is sometimes called the beads on a string structure. All histones have variants, products of different genes, classified as replication-independent and replication-dependent. H3.1 and H3.2 are common examples of replication-dependent, and H3.3 as replication-independent. H3.3 is therefore almost always available in the histone pool to rebuild displaced nucleosomes. Nucleosomes are typically displaced by histone chaperones and chromatin remodelers which cause the nucleosome to be evicted. The reformed nucleosome is often modified by replacing canonical histones with variants, especially outside of S phase of the cell cycle. The histones and their N-termini undergo modification which drives the on/off state of the chromatin. Such modifications are referred as histone PTMs, and the most common ones are acetylation, methylation, phosphorylation, ubiquitination and sumoylation. These histone PTMs are involved in cross-regulatory mechanisms which involves one PTM influencing the modification/function of another PTM usually on the same histone tail, commonly referred as histone modification crosstalk. Each

of these PTMs and their crosstalk are driven by a given set of enzymes which orchestrates its function individually or as a part of a complex of proteins to drive the function of the specific cell involved (2). The recruited proteins or enzymes serve one of three functions: writer, reader or eraser. Writers catalyze the addition of the modification, erasers catalyze their removal, and readers are proteins that specifically bind to various PTMs which recruit other proteins, complexes or signal other events. (Figure 1.1). Such complex control of histone PTMs helps to maintain and establish transcriptionally active chromatin by regulating chromatin accessibility and recruitment of transcription factors, chromatin remodeling enzymes and RNA polymerase II (RNAPII) (3).

1.3 Regulation of Chromatin Structure

The transcriptionally active chromatin state is maintained by the effort of different active histone PTMs and DNA modifications (Figure 1.1). Histone PTM associated with an active gene are H3K4me3 and H3K27ac at the promoter region, H3K4me1/ H3K27ac marking its enhancer region, and histone 3 trimethylated at lysine 36, H3K36me3 at the gene body. While the active state mostly requires presence of active histone PTMs and overall acetylation of the chromatin, the condensation of the chromatin is regulated mostly by the repressed marks H3K9me3 and H3K27me3 (Figure 1.1, 1.2). The context of modifications is also important. For example, the presence of the typically permissive modification H3K4me3 with the typically repressive H3K27me3 marks a ‘poised promoter’ or a gene whose active or repressed expression state has yet to be decided.

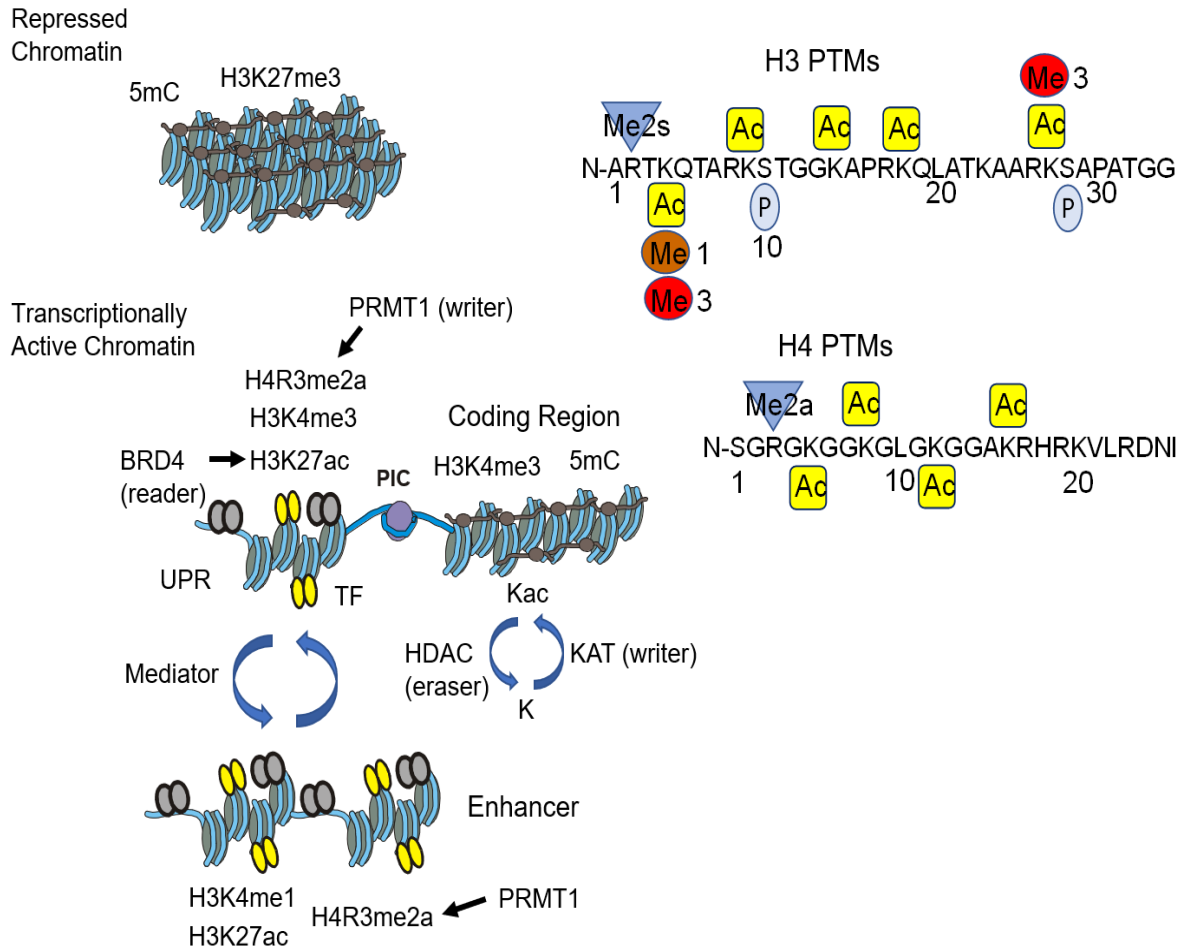


Figure 1.1: Chromatin regulating enzymes (reader/ writer/ eraser) modulating N-terminal tail PTMs of the core histones. Repressive marks, H3K27me3 and 5mC, play roles in the condensed structure of repressed chromatin. Active marks of transcriptionally active chromatin are: upstream promoter region (UPR), H4R3me2a, H3K4me3, H3K27ac which binds BRD4 (a “reader”); coding region, H3K4me3 after first exon, 5mC, dynamic acetylated histones (Kac) which is catalyzed by KATs (a “writer”) and HDACs (an “eraser”); enhancer, H4R3me2a, H3K4me1, H3K27ac, mediator. Enhancer reversibly interacts with UPR. PRMT1 and mediator are involved in this interaction. Transcription factors (TF) bind to UPR and enhancer.

Histone PTMs are mostly found on the N-terminal of the nucleosomal histones but can also occur in the globular domain. Mass spectroscopy analysis of the N-termini of H3 and H4 revealed H3 N-terminus to be the most modified of all the histones and can accommodate up to seven PTM combinations in a single N-terminal fragment (e.g. co-occurrence of lysine methylation and acetylation on the same tail) (4). Modifications, other than on the histone tail, mainly involve changes in biophysical properties of the nucleosome. These include acetylation of histone 3 lysine 56, 155 and 122 (H3K56, H3K155 and H3K122), and phosphorylation of threonine 118 on (H3T118), and result in exposing and unwrapping of DNA and increased nucleosome mobility (5). Acetylation of histone 4 on lysine 16, H4K16 prevents formation of the 30 nm chromatin fibre (6). One of the main functions of histone PTMs is to form the 'histone code', where a PTM might work as individual or in combination (one or more histone tails) to mediate their designated function. They often function through antagonistic or synergistic crosstalk that is decoded by specific effector proteins (reader) to regulate the downstream chromatin events (7). As an example, H4R3me2a, formed by PRMT1, recruits the lysine acetyltransferases p300 and p300/CREB-binding protein (CBP)-associated factor (PCAF), which then acetylates H4 at K8 and K12 and H3 at K9 and K14 (8). Lysine and arginine methylation of histones can either act as a binding site or prevent the binding of other modifiers to the site. An example of negative cross-talk is the function of H3 asymmetric di-methylation (H3R2me2a) in preventing mixed lineage leukemia (MLL) driven formation of di- and tri-methylation of H3 lysine 4 (H3K4me2/H3K4me3) whereas, presence of H3K4me3 prevents protein arginine methyltransferase (PRMT) 6-mediated H3R2me2a (9). Histone PTMs also regulate signal transduction pathways. Methylation of H3K4 at the phosphatidylinositol glycan anchor biosynthesis class P (*PIGP*) (promoter upregulates Toll-like receptor 4 (TLR4) -mediated innate immune signalling (10). Taken together, the histone

PTMs regulate DNA interaction, reader recruitment and chromatin structure to create an epigenetic signature/ mark read by the RNAPII transcription machinery. Some of the less common, newly discovered histone marks include lysine propionylation, lysine butyrylation, lysine crotonylation, lysine 2-hydroxyisobutyrylation, lysine malonylation, and lysine succinylation (11).

Another level of control of the chromatin state is offered through the methylation of the DNA by DNA methyltransferase enzymes (DNMTs). The de novo methyltransferases (DNMT3a and DNMT3b) catalyze methylation by addition of a methyl group onto the fifth carbon of cytosine residues in DNA (5^{mC}) position whereas DNMT1 functions to maintain DNA methylation through addition of methyl groups during replication. The major function of these enzymes is to establish and maintain heterochromatinization, gene silencing and promote transcriptional repression. Contrast to the previous knowledge of DNA methylation, current investigation of methylome profile suggest that presence of highly conserved DNA methylation within the gene bodies correlates positively with the upregulation of gene expression. DNA methylation in the gene body is conserved and maintained across various species by both DNMT1 and co-factor ubiquitin-like containing PHD and ring finger domains (UHRF1) which binds to the DNA methylation sites (12). The enzymes achieve their repressive function by modifying chromatin through the binding of histone deacetylase (HDAC) and lysine methyltransferases (Suppressor Of Variegation 3-9 Homolog 1/2, SUV39h 1/2 and G9) (13). DNA methylation (5^{mC}) at the promoter prevents binding of transcription factors. Further, DNA methylation at CpG islands at a promoter prevents the binding of protein/enzyme complexes, and in doing so help maintain transcriptional repression.

The DNA methylation events are also regulated by non-coding RNAs which have recently been shown to be critical regulators of chromatin compaction events in mammalian cells (14). Non-coding RNAs are divided into two groups based on their length in bp: long non-coding RNAs (lncRNAs) are typically greater than or equal to 200 bp whereas the small non-coding RNAs (sncRNAs) are typically shorter than 200 bp and include small interfering RNA (siRNAs), microRNAs (miRNAs), and PIWI-interacting RNAs. siRNA mediates transcriptional gene silencing by controlling EZH2 (Enhancer Of Zeste 2 Polycomb Repressive Complex 2 Subunit) activity and eventually DNA methylation (15). miRNAs are known to have similar mechanism as siRNAs, but mostly regulate post transcriptional events. Several histone and DNA modifications regulate the expression of the miRNA, which in turn regulate the chromatin structure by regulating the modifiers (e.g. KATs) and genes involved in epigenetic regulatory mechanism forming a regulatory circuit (16).

1.3.1 Open Chromatin/ Nucleosome Depleted Region (NDR)

The topological organization of nucleosomes across the genome is dynamic and irregular. Only about ~2–3% of the total DNA sequence is accessible/nucleosome-free. A nucleosome depleted/free region (NDR) is an open chromatin region in the genome that is devoid of a nucleosome. An NDR is characterized by the absence of nucleosomes, sensitivity to the DNaseI (deoxyribonuclease I) enzyme compared to the rest of the genome, and is marked by presence of an hypersensitive site and regulatory elements. Open chromatin is often found at transcription start site (TSS), transcription termination site (TTS), CCCTC-binding factor (CTCF) binding sites,

which are typically found in regulatory elements such as promoter, enhancer, insulator and transcription factor binding sites (17).

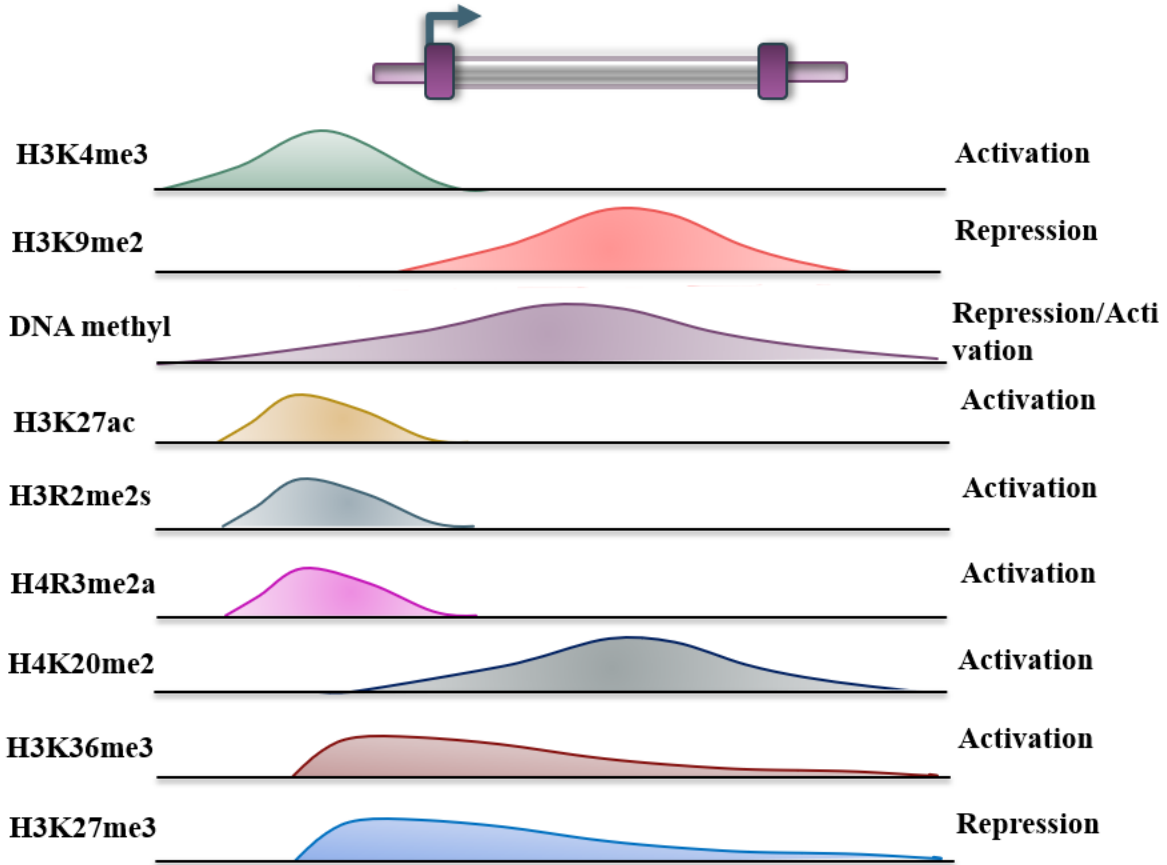


Figure 1.2: List of common histone post-translational modifications and DNA-methylation associated with annotated genes. The arrow shows the transcription start site. The positions of the histone PTMs and DNA modification (5^{me}C) near the transcription start site, 5' end of the gene and along the gene body in a transcribed or repressed state are shown.

The accessible DNA can also be found overlapping with regions at repetitive elements which is linked to chromatin remodelling at the differentiation stage (18). NDR plays a key role in

establishing and maintaining cell identity by modulating the regulatory element interaction network. Altering the length of nucleosome-free DNA regions can positively or negatively affect the rate of enhancer-promoter communication overall affecting the 3D organization of the genome. Certain NDRs are conserved across species (e.g. promoter regions) other NDRs are established in response to environmental and developmental stimuli. The accessibility to the naked DNA is itself regulated dynamically by chromatin remodellers, transcription factors (TFs) such as pioneer transcription factors and histone PTMs. While NDRs regulate TF binding depending on cell type, the TFs compete with histones and other DNA-binding proteins alternating between nucleosome occupancy and nucleosome-depleted DNA (19). The nucleosomes occupying the NDR are usually fragile and have a faster rate of turnover. Expressed genes are often associated with larger NDRs. Recent studies have attempted to understand the mechanism and establishment of NDR to gain insight into clusters of open regulatory elements/super-enhancer.

1.4 Active Histone PTMs of Promoters and Enhancers

1.4.1 Lysine Methylation and Demethylation

The lysine methylation writers, named lysine methyltransferases (KMTs), are grouped into eight families (KMT1-8). These include KMT1 A-F, KMT2 A-H, KMT3 A-G, KMT4, KMT5 A-C, KMT6, KMT7 and KMT8 A-F. The lysine methylation erasers, named lysine demethylases (KDMs), are grouped into eight families as well: KDM1 A/B, KDM2 A/B, KDM3 A-C, KDM4 A-D, KDM5 A-D, KDM6 A/B, KDM7 A-C and KDM8 (21). KMTs use S-adenosyl methionine (SAM) as a methyl donor. Typically, KDMs demethylate the methylated lysine but S-adenosyl

homocysteine (SAH) is also known to act as a competitive inhibitor to KMTs preventing methylation of lysine. KMT often methylate specific lysines on their respective substrates. Presence of previous methylation on the lysine residue and interacting partner of the enzyme poses direct affect on its degree of activity. For example, MLL can catalyze methylation of H3K4 to H3K4me2 but can trimethylate H3K4 only in the presence of KMT2A. Lysine methylation reader domains specifically bind to methylation sites and are capable for differentiation among mono- di and trimethyl-lysine sites. These reader domains include ADD (ATRX-DNMT3-DNMT3L), ankyrin, bromo-adjacent homology (BAH), chromobarrel, chromodomain, double chromodomain (DCD), MBT (malignant brain tumor), PHD (plant homeodomain), PWWP (Pro-Trp-Trp-Pro), tandem Tudor domain (TTD), Tudor, WD40 and the zinc finger CW (zf-CW) (22) (5). For example, the Tudor domain of SGF29 is essential for the recruitment of the Spt-Ada-Gcn5 acetyltransferase complex to sites of H3K4me3, which further enables acetylation of H3 tails. The dynamics of histone lysine methylation and its impact on the chromatin state are often understood through study of genome-wide methylation patterns. Such studies include chromatin profiling of regulatory regions such as promoter states, transcribed states, active intergenic regions, repressed intergenic regions, and repetitive elements. Histone lysine methylation is also known as the key dynamic modification that controls cell cycle events (23). Therefore, proteins involved in cell cycle dependent gene expression and chromatin structure maintenance are regulatory targets of KMTs and KDMs.

1.4.2 H3K4me3

H3K4me3 is an active promoter mark of expressed genes. H3K4me3-modified nucleosomes across a specific genome are located by chromatin immunoprecipitation-DNA sequencing (ChIP-seq). This active mark is preferentially located with CpG islands (CGIs). In the genome, majority of CpG dinucleotides are methylated except the ones that are within the CGIs. CGIs have an average GC frequency of $\geq 50\%$ and CpG observed/expected ratio of ≥ 0.6 in regions of ≥ 400 – 500 bps. Their distribution across the genome exhibits two distinct patterns, which are conserved in yeast, plants, worms, flies, birds and mammals. The first, found in most active genes, has H3K4me3 restricted to around the first exon and at the promoter. In the second rarer distribution, H3K4me3 covers a substantial part of the gene body and regions upstream and downstream of the gene, called the H3K4me3 broad domain signature. Genes with this signature are involved in the regulation of transcription, development, memory formation and disease (24). Enzymes (writers) that catalyze the methylation of H3K4 include MLL 1-4, SETD 1A/1B (25, 26). Recruiters of these enzymes include CXXC-type zinc finger protein 1, CXXC1 (a CpG binding protein), H2B ubiquitination, histone H3R2 dimethyl symmetric, H3R2me2s that directs the enzymes (e.g. SETD1A/B) to trimethylate H3K4 at specific sites of the gene body. Regulation of methylation of H3K4 also includes the activities of demethylases (erasers) that are in the lysine demethylase 5 (KDM5) family (5A/JARID1A/RBP2, 5B/JARID1B/PLU1, 5C/JARID1C/SMCX and 5D/JARID1D/SMCY). H3K4me3-modified nucleosomes also attract other protein complexes/readers including KATs, Chromodomain Helicase DNA Binding Protein 1 (CHD1), facilitates chromatin transcription complex, PAF complex, U2 small nuclear ribonucleoprotein complex (U2 snRNP), Splicing factor U2AF 65, SAGA-associated factor 29, and inhibitor of

growth proteins. The H3K4me3 levels are modulated by both the DNA and RNA associated multiprotein complexes machinery that forms an interconnected hub. For example, knocking down *ASH2L*, a component of the SETD1A/B enzyme complex, resulted in reduced H3K4me3 at the promoter, lower association of U2 snRNP with chromatin as well as lessened pre-mRNA splicing efficiency of an interferon inducible gene (20).

1.4.3 Histone Acetylation

One of the most common modification of N-terminal tail of the core histones is lysine acetylation, and it is mostly associated with transcriptionally active chromatin. The histone acetylation writers are KATs, deacetylation easers are HDACs, and the acetylation mark readers are bromodomain containing proteins (for example, BRD4 in Figure 1.1). KATs are categorized into four different groups; GCN5, MYST, (SAS/MOZ), P300/CBP and SRC/p160 nuclear receptor coactivator family (27). HDACs are subdivided into 4 families. The rate of histone acetylation can vary across genomic regions with some regions having a faster rate of dynamic acetylation while some have slower rates or none. Histone acetylation leads to a more relaxed structure of the chromatin by preventing H1-mediated chromatin compaction. Sodium butyrate, a weak HDAC inhibitor, increases histone acetylation and facilitates the solubility of the transcriptionally active chromatin at physiological salt concentrations. Transcription factors such as Smads, p53, and Nuclear factor-kappa B (NF- κ B) is a substrate for KATs, the acetylation of which regulates their DNA-binding ability, nuclear localization and interactions with other transcriptional regulators

(28). Imbalances in histone acetylation and deacetylation can lead to various pathological conditions which led researchers investigate KATs, HDACs, and bromodomain-containing proteins as targets for therapeutic intervention. HDACs have gained attention as potential targets to design drugs (HDAC inhibitors) against cancer, inflammation, and assorted neurological disorders.

1.4.4 H3K27ac

First identified in yeast (29), the active histone PTMs H3K27ac and H3K4me1 typically marks active enhancers, while H3K27ac and H3K4me3 are found at the promoter region of transcriptionally active genes. The formation of H3K27 is catalyzed by the KAT CBP/p300. Current literature suggests that the mechanism of acetylation of H3K27 involves bromodomain-containing protein 4's (BRD4) bromodomains interacting with p300 and CBP to modulate the KAT activity. RPD3 is the only known deacetylase specific for H3K27ac (30). The acetylation on H3K27 is read by the chromatin remodeler Brahma-related gene-1/BRG1 which further helps to enhance transcriptional activity (31). It is also reported that H3K27ac might also prevent trimethylation of H3K27 by polycomb repressive complex 2/PRC2 complex in Polycomb target genes (30). Enhancers are typically known to control cell-type specific gene expression through its interaction with the promoter within the topological domain. Recently, H3K27ac broad chromatin domains or the presence of a cluster of enhancers have been defined as super-enhancers. An enhancer hub found in close proximity of genes, some of which are involved in cell identity, are often associated with coactivators, mediators and transcription factors, resulting in an expanded H3K27ac domain (32). Given the poor understanding in defining super enhancers from regular

enhancers, further investigation will shed light on function and implication of H3K27ac broad domains.

1.5 PRMT Enzymes: Role in Transcriptionally Active Chromatin

PRMTs catalyze the transfer of a methyl group from a methyl donor, SAM, to the terminal guanidine-nitrogen atoms which generates SAH and methylated arginine residues within proteins (33). These enzymes are classified into 3 groups. Type I PRMTs catalyze the formation of monomethyl arginine and asymmetric dimethylarginine. Type II PRMTs catalyze monomethyl arginine and symmetric dimethylarginine, and Type III form only monomethyl arginine. This thesis will focus on the Type I enzyme PRMT1 and its formation of the active mark H4R3me2a, and the Type II enzyme PRMT5 and its formation of the active mark H3R2me2s and the repressive mark H4R3me2s. Previous knockdown experiments studies have shown the involvement of the PRMTs in transcription. PRMT1 is found in a protein complex with transcription factors along with KATs PCAF and SRC-1 (34). PRMT1 knockdown in avian erythroblastosis virus (AEV)-transformed cells arrested at the colony-forming unit (CFU-E) stage (known as 6C2 cells) resulted in the loss of H4R3methylation and H3/H4 acetylation in the chicken β -globin locus (35). Another group showed that knockdown of PRMT1 in a mouse erythroleukemia cell line resulted in loss of overall acetylation as well as promoter-enhancer communication in the β -globin locus, suggesting its role in erythroid transcription (36). Similar knockdown studies of PRMT5 in mammalian cells resulted in loss of H3R2me2s and H3K4me3 marks (37, 38). In the chronic myeloid leukemia cell line K562, knockdown of PRMT5 and associated complex components led to a reduction of

H4R3me2s (39). Together, these studies demonstrate the importance of PRMTs in establishing/maintaining transcriptionally active chromatin structure and active histone marks. Current literature suggests that arginine methylation of H3 and H4, catalyzed by PRMT, is required to form other active marks such as acetylated H3/H4. Due to the role of PRMTs in normal physiology and several diseases, understanding the function of PRMTs, which are modify histones and non-histones and are involved in a number of cellular processes, may make way to discover novel therapeutic targets.

1.5.1 H3R2me2s

Symmetric dimethylation of H3R2 is another active mark involved in maintaining the transcriptionally active chromatin. It is catalyzed by PRMT5. H3R2me2 is usually localized at the promoter with the active mark H3K4me3 and is known to guide binding factors to recognize the promoter mark H3K4me3 (40). WDR5 binding to H3 distinguishes asymmetric from symmetric methylation of H3R2 as WDR5 (subunit of several co-activator complexes such as MLL, SET1A, SET1B, NLS1, and ATAC) only reads H3R2me2s. H3R2me2s in turn recruits WDR5 to promote the methylation of H3K4 and produce H3K4me3. H3R2me2s functions to prevent heterochromatinization by blocking RBBP7 (retinoblastoma binding protein 7), a central component of co-repressor complexes Sin3a, NURD and PRC2 (41). There is currently no identified eraser for this active mark; however, JMJD6 is hypothesized to function as the demethylase of this histone PTM. H3R2me2s is also known to mark regions enriched with

acetylated H3/H4 and H3K4me1 at poised promoters. The function of H3R2me2s is known to keep genes poised until transcriptional activation or differentiation (41).

1.5.2 H4R3me2a

Asymmetric methylation of H4R3 is catalyzed by PRMT1. The reported functions of H4R3me2a include stimulating histone acetylation by CBP/p300 (35) and the establishment of an active chromatin structure (42). So far only a handful of protein complexes have been listed that are known to interact with methylated arginine and these include Tudor domain-containing protein 3 (TDRD3), DNA methyltransferase 3a (DNMT3a), RNA polymerase-associated protein 1 (PAF1) complex, and p300/CBP-associated factor (PCAF) (36, 43, 44). While most of these interactions have not yet been characterized, binding of Tudor Domain Containing 3/TDRD3 to H4R3me2a via a Tudor domain has been recognized as a structural motif for binding of arginine-methylated non-histone proteins. Although known as an active mark, H4R3me2a is still understudied and further investigation is required to characterize its localization.

1.6 Epigenetic Methods

With the advances in high throughput sequencing and bioinformatic analysis, several genome projects such as ENCODE and Human Genome Project have made it possible to study genomic hallmarks and diseases by mapping DNA methylation, histone modifications, regulatory elements and transcription factors. Genome-wide mapping is key to understanding epigenetic

regulation as well as targeting genes and base sequences for drug discovery and gene therapy. In order to gain insight into disease mechanisms, researchers often study the genome specific gene expression, differential expression and enrichment profile of a normal cell versus a diseased cell (which may be a cell challenged by a pathogen). Therefore, one key approach is to combine the epigenomic, transcriptomic and proteomic data to map out the interaction of the target epigenetic regulators. One of the most commonly used technique to look at histone modification/transcription factor and their chromatin signature is chromatin immunoprecipitation combined with next-generation sequencing (ChIP-seq). The method utilizes an antibody specific to the protein of interest to capture the DNA associated with the protein. The bound DNA is then identified by sequencing and validated by qPCR using primers to the region of enrichment. It is often crucial to look at open chromatin where the regulatory elements are found. This involves investigation of the promoter, enhancer or insulator region. The most commonly used techniques are DNase I-seq, (Formaldehyde assisted isolation of regulatory elements using sequencing) FAIRE-seq and (Assay for Transposase-Accessible Chromatin using sequencing) ATAC-seq. While these techniques serve the same purpose, each of them has their individual biases and are often mapped against ChIP-seq or micrococcal nuclease-seq (sequencing of isolated mononucleosomes). DNase I and ATAC-seq are enzyme dependent techniques. The DNase I enzyme preferentially cleaves the hypersensitive sites whereas Tn5 transposase enzyme opts for the cut and paste mechanism leading to cleaving of exposed DNA and insertion of sequencing adapters. FAIRE-seq, on the other hand, employs the phenol-chloroform DNA extraction method to separate the protein-free DNA from DNA that is bound to proteins (45).

Currently there are four widely used techniques to perform genome-wide DNA methylation analysis, which are: Methylated DNA immunoprecipitation sequencing/MeDIP, Methyl-CpG Binding Domain (MBD)-Based Capture and Sequencing/MBD Cap-seq, whole genome bisulfite sequencing and Illumina's Infinium Methylation assay. MeDIP method involves the immunoprecipitation of methylated genomic DNA with a 5-methylcytidine specific antibody. MBD Cap-seq (Methylminer) is proven to provide higher genomic coverage and uses the methyl-CpG binding domain (46) of the methyl-CpG-binding domain protein 2, a member of the MBD protein family, to capture methylated DNA (47). This technique is useful to distinguish 5mC from 5hmC. Illumina's Infinium Methylation assay, similar to bisulfite sequencing and pyrosequencing, focuses on the analysis of bisulfite-converted DNA to quantify cytosine methylation (48).

Three-dimensional structure of chromatin, arrangement of compartments and topologically associating domains (TADs) are also part of epigenetic studies to understand chromatin conformation and investigate interactions of regulatory elements. High-resolution techniques commonly used to study the three dimensional structure and interaction include Chromosome Conformation Capture Analysis/3C and 3C-based strategies such as 3C, 4C, 5C, Chromatin Interaction Analysis with Paired-End Tag Sequencing (ChIA-PET) and Hi-C analysis (49). Three-dimensional genome profiling is commonly used to identify epigenetic changes associated with development, aging and memory. Although ongoing epigenetic research typically uses one of the above-mentioned methods, the combination of different epigenetic techniques has often been useful to resolve new challenges. The technique Ocean-C (open chromatin enrichment and network Hi-C) combines the key steps of FAIRE-seq and Hi-C and has been proven to be an

excellent way to isolate enhancers and map out their long and short interactions overcoming the shortcomings of using either FAIRE-seq or Hi-C alone (50). Therefore, combining and exploring different epigenetic approaches to normal and disease states will help us unravel the functions of understudied histone modifications, chromatin-associated complexes, chromatin remodelers, transcription factors, and the effect of changes in acetylation and methylation states of the chromatin.

1.7 Innate Immunity

Innate immunity, developed early in animal evolution, is the non-specific branch of the immune system that responds naturally depending on the physiological state and genetic factors of a living system. As the first line of defense, its primary role is to restrict the spread of the foreign pathogen throughout the body. The innate response is typically short-term and is responsible for both the cell- and stimulus-specific transcriptional response upon foreign microbe exposure. The innate immune response involves specific cells (e.g., macrophages, dendritic and natural killer cells), secreted proteins (e.g interleukins, chemokines and cytokines), receptor-mediated signaling, cell-to-cell communication and conditioning of the adaptive immune response. The innate response is initiated by recognizing the specific pathogen-associated molecular patterns (PAMPs) on the pathogen's surface via cell specific pattern recognition receptors (PRRs) such as Toll-like receptors (TLRs), RIG-I-like receptors (RIG-I), NOD-like receptors (NLRs), C-type lectin receptors (CLRs) and cyclic GMP-AMP synthase (cGAS)- stimulator of interferon genes (51). The binding of PAMPs with PRR triggers the release of cytokines, which are chemical

messengers that regulate cell differentiation, activation/regulation of immune response and gene expression. Interleukins, a subtype of cytokines, mediate leukocyte interactions by inducing nearby cells to release cytokines, leading to a cytokine cascade/storm. A second class of cytokine, interferons (INF), function by signaling and promoting infected cell apoptosis, viral replication inhibition and downstream immune cell activation. Interferons signal uninfected cells to reduce protein synthesis by altering their gene expression which increases the cell's defense against viral invasion (52). The first line of pro-inflammatory cytokines induces inflammation and phagocytosis by several leukocytes and eventually destruction of infected cells by natural killer (NK) cells. Cytokines also signal the nervous system to alleviate symptoms such as sickness, fatigue, increased body temperature to encourage the individual to rest and prevent them from spreading the infection to others (53).

PAMPs regulate activation of inflammatory and antimicrobial response in the host through PRR-induced signal transduction pathways ultimately leading to activation of gene expression and adaptation to environmental change. With the advancement in high-throughput sequencing techniques, chromatin modifications and epigenetic regulators were easily identified that equally contribute to strengthen the host's antimicrobial defense. The key epigenetic processes and regulators that nurture the innate immunity during an infection are DNA methylation, histone PTMs, chromatin remodelers and non-coding RNAs (ncRNAs) (54). Modulation of gene expression at the chromatin level by innate signalling pathways after pathogen sensing ensures proper development and differentiation of distinct innate immune cell (monocyte and lymphocytes), recognition and transduction of infection signals, secretion of inflammatory cytokines and effector molecules, and memory-like innate immune responses. Conversely, delay

in pathogen sensing could result in viral latency leading to relapse of the infection or a more complicated form of disease state (55). Therefore, investigating key epigenetic players that modulate host–pathogen epigenetic interactions will shed light on the mechanisms of innate immune response and potential therapeutic targets for infectious disease.

1.7.1 Epigenetic Regulators of Innate Immunity

In order to initiate the innate immune response, the signal transduction is regulated by several chromatin modifiers and protein factors positively (e.g. KDM5A, EZH1, p300, HDAC9, SETD2, KMT2B and DNMT3A) and negatively (e.g. several different lncRNA and lincRNA) by regulating transcription of key signal transducers (JAKs, TBK1, IFI16, cGAS, RIG-I, CD14, MYD88, TRAF3/6) and transcription factors (STAT1/2, p65, p50, IRF3, NF- κ B) controlling the expression of antimicrobial proteins, variety of Interferones/IFNs and cytokine (51)(56)(57). Expression of key genes are regulated by both active and repressive marks. A recent report suggests demethylation of H3K27me₃ of *IFN β* promoter by KDM6A promotes transcription of *IL-6* and enrichment of H3K27ac during virus infection in macrophages (58). KDM6 is also involved in *IFN- β* expression but by a mechanism that does not require the demethylating activity of KDM6. In this mechanism KDM6 recruited MLL4 resulting in the production of H3K4me₂ at the S-IRE1 region of *IFNB1* gene and increased expression of the IFN- β -specific enhancer-derived RNA (eRNA) *S-IRE1*, which reflects full activation of the IFN- β enhancer (58). While SMYD5 (writer for H4K20me₃) represses TLR4-mediated promoter (59), lysine methyltransferase 2B (KMT2B) upregulates TLR4-mediated signalling, by increasing the levels of H3K4me₃ at the *PIGP* gene promoter (60). Increases in H3K27me₃ within the promoter of mice *Tollip*, negative

regulator of TLR signalling, by EZH1 enhances the TLR signalling further (61). DNMT3A positively regulates TBK1–IRF3 signalling by upregulating HDAC9 activity leading to *IRF3* phosphorylation (62). p300-mediated acetylation and methylation by NSD3 of *IRF3* promotes antiviral immunity (63)(64). Activation of interferon-induced and *IRF3*-dependent primary response genes (PRGs) with promoters without CGIs requires the chromatin remodeler, SWI/SNF-remodelling complex (65)(66). Pro-inflammatory stimulation in macrophages upregulates PCAF/GCN5 mediated acetylation of H4K5, H4K8 and/or H4K12 of PRGs. The acetylation is then read by the reader BRD4 to initiate RNA polymerase II transcriptional elongation (67). PRGs like IFN and tumor necrosis factor (TNF) genes remain poised and are first in line to get induced upon microbial exposure. Interferon stimulated genes (ISGs) are known to be a critical regulator of viral infections since they can target almost every stage of a virus life cycle. The highly inducible ISGs includes IFIT family members. Another important group of ISGs are *IFITM1/2/3* which are upregulated at the earliest to prevent the entry of the virus or combating the early stage of viral replication in the host. These genes are known to prevent entry of several viruses such as SARS-CoV, HIV-1, filovirus, FLUAV, VSV, WNV, YFV and DENV (68). PRMTs regulate pro-inflammatory cytokine expression both dependent and independent of their enzymatic activity and catalyzed histone PTMs (69). Little research has been done on the function and expression level of histone PTMs, their readers, writers and erasers in the regulation of genes involved in the innate immune response pathway. Inadequate epigenetic research on gene activation and gene enrichment profiles has led to poor database resources to identify histone PTMs and their epigenetic regulators of innate immune genes. Further investigation of mechanisms of how histone PTMs and their modifiers, and transcription factors regulate the signalling pathway and effector molecules of the innate immune response will provide more insight on epigenetic regulation of innate immunity.

1.7.2 Avian Innate Immunity

The avian innate response components are similar to the ones produced in mammals. In chicken, the innate response utilizes the same cells as mammals (e.g. macrophages, NK cells, and dendritic cells) including heterophils (counterparts of mammalian neutrophils). Chicken heterophils are found to express several different TLR mRNAs resulting in upregulation of *TLR1La*, *TLR1Lb*, *TLR2a*, *TLR2b*, *TLR4*, and *TLR5* on bacterial exposure, *TLR3* and *TLR7* for viral exposure suggesting a broad range of pathogen recognition (70). Heterophils responds to various TLRs stimuli by upregulating the expression of cytokines and chemokines to contribute to the resistance against pathogen invasion. Avian TLRs include *TLR1La*, *TLR1Lb*, *TLR2a*, *TLR2b*, *TLR3*, *TLR4*, *TLR5*, *TLR7*, *TLR15*, and *TLR21*. Other than gas exchange and transport, the immunological role of the erythrocyte includes contribution to the host immunity by upregulating cytokine transcripts in response to the TLR ligands. One study reported upregulation of *TLR 2,3,4,5* and *21*, immunological genes such as type I IFNs and interleukin (*IL*)-8 in response to TLR ligand treatment. Although no structural similarity was observed, overexpressing chicken *TLR21* leads to activation of NF- κ B, which is typically observed in the activation of the mammalian *TLR9* pathway. One assumption about the TLR-mediated response in unaffected erythrocytes, is that upregulating such transcripts could provide protection from and prevent the spread of pathogens (71). Mature chicken macrophages are of immunological research interest and are often collected after induced maturation when exposed to inflammatory cytokines. DCs in avian system function in a similar way as in mammals. After processing foreign antigens, they migrate to local draining lymph nodes where they upregulate costimulatory molecules and stimulate further naive T cells. Interestingly, chickens lack lymph nodes and therefore the site of antigen presentation and DC

maturation in chickens remains to be characterized (72). Unlike mammals, chickens do not have RIG-I but rather express two other members of the chicken RIG-I-like receptors (RLR) family, MDA5 and LGP-2. Like mammalian *MDA5*, stimulation by dsRNA and type I IFN results in upregulation of *MDA5* which is a major sensor of avian virus infection and leads to upregulation of the expression of innate immune genes, which induces the adaptive immune response in immune cells. The multiple levels of protection offered by *MDA5* is known to compensate for the missing *RIG-I* and is a major target for vaccine development. Chicken *LGP-2* contrasts mammalian *LGP-2*, where its knockdown lead to reduced avian influenza virus-induced chicken *IFN- β* promoter activity (73). Chicken *IRF1* and *IRF7* are also known to modulate the regulation of the *IFN- β* promoter and anti-viral immunity. Therefore, the chicken TLR and RLR signaling pathways appear to be similar to that of mammalian RLR signaling pathway and similarly leads to the production of type I IFNs and inflammatory cytokines through the activation of IRF and NF- κ B (74).

The chicken has been a crucial model system when it comes to vaccine development. Some key milestones in using the chicken as a disease model include: the first lab developed vaccine by Louise Pasteur after his work of immunization of chicken with *Pasteurella* (75, 76), and B cells were first identified in chicken as antibody producing cells and natural model for Marek's disease and human T cell lymphomas (77). Chicken also plays a big role in the history of cell biology and in the study of innate immunity. The IFN activity was first discovered in chicken and has been characterized for interfering with the viral replication (78). To date the annotated IFN genes in the *Gallus gallus* (chicken) genome assembly 6 (GRCg6a) assembly include *IFNA3*, *IFNARI*, *IFNAR2*, *IFNG*, *IFNGR1*, *IFNGR2*, *IFNL3A* and *IFNWI*. Like mammals, chickens use the JAK-

STAT signalling pathway to drive the transcription of ISGs. Analysis of chicken chemokines revealed that chicken expresses many orthologs of their mammalian chemokines although the DNA base sequence similarity was found to be less identical (79). To date the annotated chicken chemokine genes in the GRCg6a assembly include *IL-1*, *IL-2*, *IL-3*, *IL-4*, *IL-5*, *IL-6*, *IL-7*, *IL-8*, *IL-9*, *IL-10*, *IL-11*, *IL-12*, *IL-13*, *IL-15*, *IL-16*, *IL-17*, *IL-18*, *IL-21*, *IL-22* and other variants. To date, limited epigenetic research has been reported to identify the epigenetic players in the avian innate immune system. Current literature shows that human erythrocytes are also involved in innate immunity other than their obvious function as oxygen transporters. Although the mammalian erythrocyte has lost the ability to carry out transcription and translation, they are still able to interact with inflammatory molecules such as pathogens, chemokines and nucleic acids which makes the human red blood cell comparable to the chicken erythrocyte (80). Avian innate immunity is considered similar to mammalian both structurally and functionally. Characterizing the immunity genes would open the door to target crucial regulators in both avian and mammal infections.

1.8 Chicken Model System

1.8.1 Genetic Makeup

The first complete avian genome sequence was obtained from the undomesticated chicken, red jungle fowl (*Gallus gallus*). The karyotyping revealed 78 chicken chromosomes where 1-10, Z and W are classified as macrochromosomes and 11-38 are microchromosomes (81). Approximately 88% of the sequence has been anchored to chromosomes, which include autosomes

1–24, 26–28, and 32, and sex chromosomes W and Z. The remaining unanchored contigs have been concatenated into the chromosome chrUn (82). Chickens are known to share a common ancestor with the mammals and long blocks of conserved synteny although diverged approximately 300 million years ago (Figure 1.3). Regarding the chromosomal organization of genes, the human genome is closer to the chicken than to rodents. Chickens are located between mammals and fish in the tree of evolution. The common ancestry has made chicken an interesting model in the scientific community to study the commonalities among mammals and other vertebrates shared through evolution (83). Through genome comparative study, the International Chicken Genome Sequencing Consortium reported that although chicken and human shared approximately same number of genes, the chicken genome was about one-third of the human genome. The Chicken genome has about 20,000-23,000 genes in its 1 billion DNA base pairs, compared with the human count of 20,000-25,000 genes in 2.8 billion DNA base pairs (84). The difference suggested the less DNA in chicken genome reflected less amount (only ~15% repetitive DNA) of repeat and duplicate DNA elements and reduced number of pseudogenes. Since the chicken genome is devoid of unnecessary sequences yet has all the useful information, it is considered a good model for the exploration of transcriptionally active chromatin structure, nucleosome-depleted (regulatory) regions and evolution.

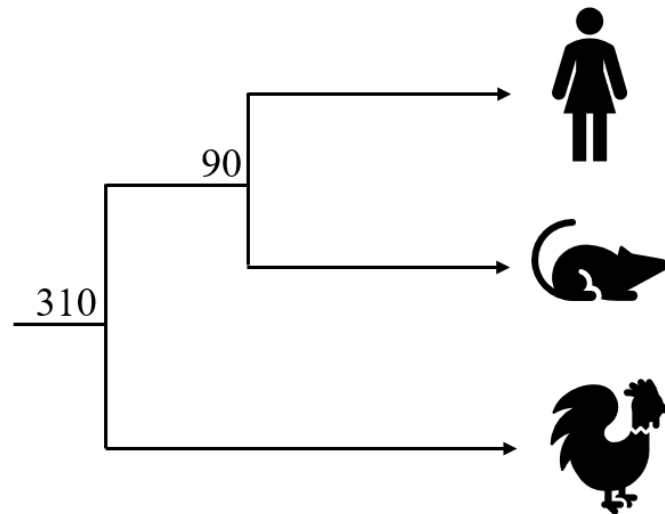


Figure 1.3: Phylogenetic tree showing evolutionary relation and divergence time (in million of years) among vertebrate human, chicken and rodent.

By comparing the chicken genes to that of human, investigators found sequence and functional similarity of genes involved in cellular functions, immune response, environment adaptation and reproduction. Scientists have also investigated the non-coding transcriptome of the red jungle fowl *Gallus gallus* genome and have reported 9681 lncRNAs according to the ncRNA database NONCODE database (85). Previously considered as junk RNA, their investigation is opening the door to a whole new perspective of their possible role in regulation and conservation among vertebrates. Gene mapping studies have shown that on aligning the human genes with that of chicken, around 2000 human genes were found to have a different start site than what was first

reported (84). Such discrepancies in evaluating the "true" start sites of human genes play a vital role in understanding gene boundaries and to target them for gene therapy.

1.8.2 Erythropoiesis

The process of erythropoiesis has always been an interest of study to understand cell differentiation and development, morphology of transitioning cells, and the changes in DNA, RNA and protein synthesis associated with each stage. The non-mammalian erythropoiesis has achieved more attention due to the retaining of the nucleus in the fully mature erythrocyte. Although the mature cell is transcriptionally inert, the polychromatic cells are transcriptionally active and have gained much attention in the study of erythropoiesis. Avian erythropoiesis studies have shown that upon inducing anemia through phenyl hydrazine treatment, it is possible to isolate large quantities of polychromatic from the circulating blood of the recovering birds (86). The differentiation changes can be visually understood by observing the change in nuclear, cytoplasmic and cell shape changes. Due to the success of techniques that isolate avian erythrocyte nuclei, the avian erythrocyte has been valuable in the study of chromatin structure, transcription and epigenetic modifications (87). These nucleated erythrocytes have also gained attention in the studies of evolutionary and lineage diversity. While the stages in avian erythropoiesis are often comparable to that of mammals, scientists have now backed up the process of expulsion of the nuclei in mammalian red blood cell as an evolutionary pressure. The early mammals appeared in the Triassic period; During that period, the oxygen content in the atmosphere naturally favored the loss of nuclei to contain more oxygen-carrying hemoglobin, thus more circulation of oxygenated blood to

survive (88). Unlike mammals, birds evolved in the Jurassic period when the atmosphere posed no selective pressure to lose their red blood cell nucleus (89).

1.8.3 Chicken Erythrocyte Chromatin Structure

The chicken nucleosome is composed of the same core histones (H2A, H2B, H3 and H4) as found in humans. The distinction lies in the presence of a variant of the linker histone H1, known as H5. Chicken mature erythrocytes have the linker histones H1 and H5 present in the ratio of 0.4:0.9 per nucleosome which contributes to the structured folding of the nucleosomal units. Both mature and polychromatic erythrocytes have limited regions of DNase I sensitive chromatin, which is less condensed than the bulk of chromatin. The main histone PTM that drives this unfolding is acetylation. Studies on histone acetylation dynamics revealed that approximately 3.7% of the histones were dynamically undergoing acetylation and deacetylation (90). Fractionation of 0.15 M NaCl soluble chromatin coupled with acetic acid-urea-triton-X100 (AUT) electrophoresis and western blotting have successfully been used in the past to isolate and characterize active gene enrich chromatin fractions (91). The 0.15 M salt-soluble polynucleosomes were enriched in acetylated H2B, H2A.Z and H4, poly and monoubiquitinated H2A and H2B and variants H3.3 and H2A.Z (often found next to NDRs of active promoters, enhancers and insulator regions) (92). The low-salt insoluble fraction was reported to be enriched in active genes which is speculated to be due to the presence of large multiprotein preinitiation transcription complexes (93). These experiments were crucial in making the statement that dynamically acetylated histones maintain

transcriptionally active chromatin as well as alter linker histone and DNA interaction; contributing to the unfolded higher structure of the chromatin (94).

The nuclear content of avian erythrocytes (early/mid/late polychromatic erythrocytes) undergo changes as they approach the mature erythrocyte (terminally differentiated) stage. H5 is highly phosphorylated in the early stages of erythropoiesis and plays a key role in compaction of chromatin in the nucleus. Therefore, the H5 plays a key role in the mature cells to maintain the repressed stage and heterochromatinization. Although the genome is considered inactive, the mature cell is still able to synthesize RNA (95). Histones are typically acetylated during the polychromatic stages which declines as they approach maturation (3.7 versus 2.1% of total potential acetylation sites) (96). The presence of H5 in chicken cells is an early marker of the erythroid lineage and therefore is already present in 6C2 cells. AEV (differentiates at 42°C) readily transforms bone marrow cells *in vitro*, resulting in leukemic erythroblast which has long been used to study hematopoietic cell differentiation in leukemic cell to gain insight into the pathway of chicken erythroid differentiation (97). Developmental studies have shown that as the cells transition from CFU-E to maturation, H5 increases 4-fold. This increase was concluded to be mainly due to independent regulation of H5 genes, compared to other histones and also due to selective deposition of H5 in the mature cells (98). One study has shown that the erythroid specific β -globin and H5 genes are transcriptionally active in the chicken polychromatic erythrocytes but not in the mature cells due to decline/lack of availability of transcription factor Sp1 and GATA-1 (99). In chicken adult reticulocytes the H5 enhancer interacting trans-acting factor GATA-1 is highly expressed, whereas GATA-2 and GATA-3 are transcribed at lower levels (100)(101). Such features made chicken a great model organism to study the events of chromatin condensation and

decondensation through isolation of the erythrocyte nuclei and subjecting it to varying ionic conditions.

1.8.4 β -Globin Domain

The chicken model is often used to study the function of the globin genes coding for embryonic and adult globin. The chicken α -globin gene cluster consists of three functional genes: π (*HBZ*), α^D (*HBM*), and α^A (*HBA1*) (102), while the β -globin gene cluster has four: ρ (*HBG1*), β^H (*HBE1*), β^A (*HBG2*), and *E* (*HBE*) (103). The stages of development are modulated by gene activity specific to that stage, histone modification and DNA methylation. For example, the DNA methylation pattern of the promoter and gene expression shows an inverse correlation in β globin genes (ρ and β^A). Garry Felsenfeld and his team have done extensive research to characterize the chicken folate receptor/ β -globin locus to study different histone PTMs that mark each developmental stage and correlate with transcriptional activity. Characterization of the β -globin domain includes identifying cis- and trans-regulatory elements key to transcriptional and developmental regulation. DNase-I hypersensitivity assays revealed that the 5' end of the chicken adult β -globin gene contains a hypersensitive region of 115 bp (104). The regions of hypersensitivity (about 100-fold more sensitive than bulk chromatin) are nucleosome-depleted regions associated with regulatory elements such as enhancers, locus control regions (LCR) and promoters. Identifying all the regulatory element is the crucial step to understanding embryogenesis, development and differentiation. Polychromatic erythrocytes express the adult β^A -globin gene as do 15-day old chicken embryo erythrocytes, but do not express the β^H globin gene

as cells of late embryos and newly hatched chickens do. Several investigators have used this nucleated, non-replicating G₀-phase cells to investigate the globin genes. The studies reported that in transcriptionally active chromatin, the acetylated histones are typically located around the TSS of expressed genes but for α - and β -globin genes in mammalian and chicken erythroid cells, the dynamically acetylated histones are broadly distributed covering transcriptionally competent and expressed globin genes. Davie and his team recently characterized all the transcriptionally active domains along with the globin domain (80). The key finding was that at the domain's edges, histone acetylation drops defining acetylation domains similar to the DNase I sensitive β globin domain. Felsenfeld has also looked at transcription and replication independent variant H3.3 distribution over locus control/regulatory elements in active domains. The H3.3 enrichment over regulatory elements at the folate receptor was also observed at developmental stages when the gene itself was not expressed implying the presence of H3.3 correlates with the presence of the accessible transcriptionally active domain. The presence of replication-independent H3.3 over coding regions was reported to be due to a gene specific mechanism such as chromatin remodeling or histone PTMs (105). Although α - and β -globin genes are located on different chromosomes, the chicken β globin gene domain organization is similar to its mammalian counterpart (95).

1.8.5 Three-Dimensional Organization of the Chicken Genome

Approximately 1–2% of polychromatic and mature erythrocyte chromatin has dynamically acetylated histones allowing only a small portion of the genome to remain in a decondensed state (80). Organization of chromatin into defined domains helps to facilitate and maintain gene

expression. The three-dimensional chromatin structure of the chicken erythrocyte genome is still understudied. Typically, the genome is organized into compartments and TADs at an intermediate scale (10 kb – few Mb). TADs are mainly domains that have boundaries enriched with CTCF, the cohesin complex, housekeeping and tRNA genes (106). Compartments are divided into type A and B. Compartment A is associated with active marks whereas compartment B is composed of repressed marks. Within each TAD domain, the interaction of the promoter and enhancer is regulated such that it does not interfere with the regulation of the nearby TADs (Figure 1.4). The promoter-enhancer interaction is also influenced by their three-dimensional arrangement e.g. on circular template, insulator near the enhancer reduced the expression, but in linear template decreased expression is only observed when insulator placed between the enhancer and promoter (107). High-resolution genome-wide Hi-C analysis of chromatin interactions revealed a different structural arrangement in chicken embryonic fibroblast vs chicken erythrocytes (108). The chicken embryonic fibroblast genome is partitioned into compartments and TADs with CTCF binding sites encapsulating transcriptional activity within the particular domain. However, the polychromatic erythrocyte genome has well organized A and B compartment but does not have a TAD organization indicating only long-range interaction of promoter and enhancer which differs from what is observed in the mammal genome. Although the sizes of the chicken TADs are considered similar to mammals in size, from an evolutionary perspective, this information is not enough to draw conclusions about their common ancestral TAD arrangement and at what point of the avian evolution the TADs were lost.

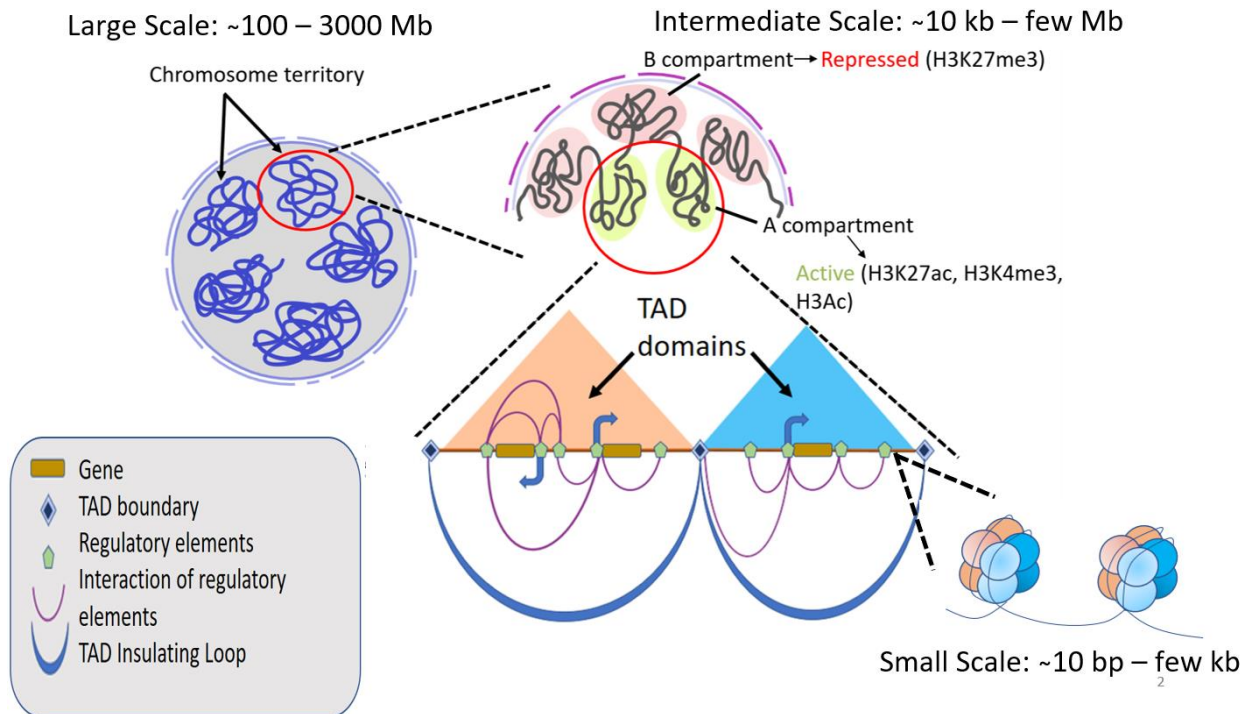


Figure 1.4: Folding and organization of chromatin at multiple scale.

1.8.6 Chicken DNA Methylome

The methylation of genomic DNA is another important regulator of embryogenesis and cell differentiation. The gene expression profile of the methylated DNA not only provides information about the epigenomic landscape of the genome under study but also provides information about certain mutations, behavioural changes and environmental adaptability of the species being investigated. The red jungle fowl is believed to be the ancestor of the domestic chicken breed (e.g. white leghorn) which has adapted to the poultry environment over years. Comparative studies of the DNA methylation pattern in different bred populations has gained

interest among epigenetic researchers to study the evolutionary methylome alterations in chicken breeds. The genome wide study of chicken DNA methylation revealed that, despite their breed type, the methylation pattern *Gallus gallus* to be similar to that of mammals and plants. Such methylation pattern typically includes enriched methylation in the gene body and repetitive element regions and depletion at the TSS and TTS. Similar to the human genome CGIs remained unmethylated (109). CGIs predate the evolution of vertebrates and are known as the hallmark of transcription initiation. These evolutionary footprints are found at annotated TSS, intergenic and intragenic regions of annotated genes. CGIs with no known function are termed as orphan CGIs and are assumed to mark promoter of novel transcripts not yet annotated (110). Chicken promoters are often categorized based on their associated CGI: promoters with long CGIs (>800 bp, LCGI) and promoters with no CGIs (NCGI). LCGIs are mostly associated with transcriptional, developmental and biosynthetic processes, whereas NCGI mostly occupy promoters of genes involved in defense response, behaviour and neurological processes. To further understand the evolution of promoter regions in vertebrates, researchers have compared the abundance and distribution patterns of sequence motifs in these regions. Although very few studies about the methylation pattern have been reported, the chicken has been used to determine the function of the DNA demethylase TET. In a recent study, one group cloned the chicken TET genes and found their regulatory function over the adult β globin genes. TET was seen to cause demethylation of the promoters of the β^A globin gene in erythroid cells. The same group observed decreased expression of the β^A globin gene during TET1 knockdown experiment in chicken erythroid progenitor cells (111). Although the chicken genome is thought to contain a much smaller number of repetitive elements in comparison to mammals, the short tandem repeat motif frequency in chicken promoter was higher when compared to mammalian promoters. Scientists interpret these

findings to be linked to conservation through promoter evolution and insightful regarding the structure of avian promoter (112).

1.8.7 Poultry Epigenetics

The chicken model was initially important in the field of embryology and developmental studies but soon expanded its horizon in the interdisciplinary fields of immunology and epigenetics. Apart from erythrocytes and immune cells, other tissues have also been used in biomedical research. The presence of histone H5 in chick embryo fibroblasts has also been exploited in inhibitory studies to test the effect of interferon treatment on the expression of immune and histone H1 and H5 genes (113). With the advancement in genome analysis, it is now possible to study all the histone PTMs and epigenetic regulators in chicken to characterize the ortholog genes in chicken crucial in human diseases and viral infections with no up to date cure. Agricultural researchers have also made great effort in deciphering the chicken genome to combat avian flu. The genomic component of their study is to understand how genetic variation and environment influence the susceptibility of different strains to get sick. Domesticated from their wild bred, red jungle fowl, the broiler chicken became a popular source of protein in our day to day diet due to its better success rate in growth, feeding efficiency, disease resistance and survival (114). Due to great success in comparative studies of domestication effects, broiler chicken became popular model organism in the field of poultry epigenetics for researchers interested in enhancing the nutritional value of poultry produces and egg products. Scientists have made extensive progress in studying the epigenetics patterns of the breeds to perfect domestication traits and generate broiler

generations that intensified the farm production system. Although chicken has been extensively used for a diverse array of research, very little focus has been shed on the transgenerational epigenetics of breeder and broiler chicken population. It has been reported that unwanted innate immunity effects might be the main reason for a mismatch in environment between breeder and broiler generation (115). In an effort to improve the health of the broiler/breeder generation and understand the epigenetic programming that modulates the environmental and inherited cues it is very important to shed light on the transgenerational effects of poultry breeding on breeders and broiler generations.

1.9.1 Hypotheses

1. Mapping of nucleosome depleted regions using FAIRE-seq will identify all regulatory regions
2. Specific chromatin signatures (e.g. H3K4me3 broad domains) align with gene functions

1.9.2 Aims

1. To map out nucleosome-depleted regions using FAIRE to identify regulatory regions in chicken polychromatic erythrocytes
2. To characterize the histone PTMs, variants and chromatin surrounding the nucleosome-depleted regions
3. To integrate the chromatin and transcriptome features towards understanding the organization of transcriptionally active chromatin in chicken polychromatic erythrocytes

1.9.3 Research Strategy

Histone H3 and H4 acetylation dynamics were studied in polychromatic erythrocytes incubated for one hour in the presence of the HDAC inhibitor, sodium butyrate. It is known that the histone undergoing dynamic acetylation are associated with transcriptionally active chromatin. Thus, this analysis explored the acetylation dynamics of H3 variants (H3.2 and H3.3) and H4 and the response of histone methylation to this short term HDAC (Chapter 3). Further understanding of the functional significance of the dynamics of the H3 and H4 PTMs was made possible in the studies of the location of the active histone PTMs (H3K4me3, H3K27ac, H3R2me2s, H4R3me2a) relative to nucleosome free regions which were present at promoters of transcribed genes (Chapter 4 and 6). Further characterization of the nucleosome depleted regions in the polychromatic erythrocyte was done to identify regions in the genome other than promoters that had this open chromatin feature (Chapter 5 and 7).

2.0 References

- [1] Waddington, C. H. (1942) Canalization of development and the inheritance of acquired characters. *Nature* 150, 563–565.
- [2] Suganuma, T., and Workman, J. L. (2008) Crosstalk among Histone Modifications. *Cell* 135, 604–607.
- [3] Temple, I. K. (2014) Lyle Armstrong: Epigenetics. *Human Genetics* 133, 813–814.
- [4] Lothrop, A. P., Torres, M. P., and Fuchs, S. M. (2013) Deciphering post-translational modification codes. *FEBS Letters* 587, 1247–1257.
- [5] Rothbart, S. B., and Strahl, B. D. (2014) Interpreting the language of histone and DNA modifications. *Biochimica et Biophysica Acta - Gene Regulatory Mechanisms* 1839, 627–643.
- [6] Delcuve, G. P., Rastegar, M., and Davie, J. R. (2009) Epigenetic control. *Journal of Cellular Physiology* 219, 243–250.
- [7] Lee, B. M., and Mahadevan, L. C. (2009) Stability of histone modifications across mammalian genomes: Implications for “epigenetic” marking. *Journal of Cellular Biochemistry* 108, 22–34.
- [8] Huang, S., Litt, M., and Felsenfeld, G. (2005) Methylation of histone H4 by arginine methyltransferase PRMT1 is essential in vivo for many subsequent histone modifications. *Genes and Development* 19, 1885–1893.

- [9] Guccione, E., Bassi, C., Casadio, F., Martinato, F., Cesaroni, M., et al. (2007) Methylation of histone H3R2 by PRMT6 and H3K4 by an MLL complex are mutually exclusive. *Nature* 449, 933–937.
- [10] Zhang, Q., and Cao, X. (2019) Epigenetic regulation of the innate immune response to infection. *Nature Reviews Immunology* 19, 417–432.
- [11] Barnes, C. E., English, D. M., and Cowley, S. M. (2019) Acetylation and Co: An expanding repertoire of histone acylations regulates chromatin and transcription. *Essays in Biochemistry* 63, 97–107.
- [12] Ndlovu, 'Matladi N., Denis, H., and Fuks, F. (2011) Exposing the DNA methylome iceberg. *Trends in Biochemical Sciences* 36, 381–387.
- [13] Cheung, P., and Lau, P. (2005) Epigenetic regulation by histone methylation and histone variants. *Molecular Endocrinology* 19, 563–573.
- [14] Francia, S. (2015) Non-coding RNA: Sequence-specific guide for chromatin modification and DNA damage signaling. *Frontiers in Genetics* 6, 320.
- [15] Wei, J. W., Huang, K., Yang, C., and Kang, C. S. (2017) Non-coding RNAs as regulators in epigenetics (Review). *Oncology Reports* 37, 3–9.
- [16] Carpenter, S., Ricci, E. P., Mercier, B. C., Moore, M. J., and Fitzgerald, K. A. (2014) Post-transcriptional regulation of gene expression in innate immunity. *Nature Reviews Immunology* 14, 361–376.
- [17] Qu, J., Yi, G., and Zhou, H. (2019) P63 cooperates with CTCF to modulate chromatin architecture in skin keratinocytes. *Epigenetics and Chromatin* 12, 31.

- [18] Gomez, N. C., Hepperla, A. J., Dumitru, R., Simon, J. M., Fang, F., et al. (2016) Widespread Chromatin Accessibility at Repetitive Elements Links Stem Cells with Human Cancer. *Cell Reports* 17, 1607–1620.
- [19] Klemm, S. L., Shipony, Z., and Greenleaf, W. J. (2019) Chromatin accessibility and the regulatory epigenome. *Nature Reviews Genetics* 20, 207–220.
- [20] Davie, J. R., Xu, W., and Delcuve, G. P. (2015) Histone H3K4 trimethylation: Dynamic interplay with pre-mRNA splicing¹. *Biochemistry and Cell Biology* 94, 1–11.
- [21] Black, J. C., Van Rechem, C., and Whetstine, J. R. (2012) Histone Lysine Methylation Dynamics: Establishment, Regulation, and Biological Impact. *Molecular Cell* 48, 491–507.
- [22] Kutateladze, T. G. (2011) Snapshot: Histone readers. *Cell* 146, :842-842.e1.
- [23] Ernst, J., Kheradpour, P., Mikkelsen, T. S., Shores, N., Ward, L. D., et al. (2011) Mapping and analysis of chromatin state dynamics in nine human cell types. *Nature* 473, 43–49.
- [24] Liu, X., Wang, C., Liu, W., Li, J., Li, C., et al. (2016) Distinct features of H3K4me3 and H3K27me3 chromatin domains in pre-implantation embryos. *Nature* 537, 558–562.
- [25] Shinsky, S. A., Monteith, K. E., Viggiano, S., and Cosgrove, M. S. (2015) Biochemical reconstitution and phylogenetic comparison of human SET1 family core complexes involved in histone methylation. *Journal of Biological Chemistry* 290, 6361–6375.
- [26] Shilatifard, A. (2012) The COMPASS Family of Histone H3K4 Methylases: Mechanisms of Regulation in Development and Disease Pathogenesis. *Annual Review of Biochemistry* 81, 65–95.

- [27] Davie, J. R., and Spencer, V. A. (1999) Control of histone modifications. *Journal of Cellular Biochemistry* 75, 141–148.
- [28] MH, K., and CD, A. (1998) Roles of Histone Acetyltransferases and Deacetylases in Gene Regulation. *BioEssays : news and reviews in molecular, cellular and developmental biology* 20, 615–626.
- [29] Suka, N., Suka, Y., Carmen, A. A., Wu, J., and Grunstein, M. (2001) Highly specific antibodies determine histone acetylation site usage in yeast heterochromatin and euchromatin. *Molecular Cell* 8, 473–479.
- [30] Tie, F., Banerjee, R., Stratton, C. A., Prasad-Sinha, J., Stepanik, V., et al. (2009) CBP-mediated acetylation of histone H3 lysine 27 antagonizes *Drosophila* Polycomb silencing. *Development* 136, 3131–3141.
- [31] Wu, T., Kamikawa, Y. F., and Donohoe, M. E. (2018) Brd4's Bromodomains Mediate Histone H3 Acetylation and Chromatin Remodeling in Pluripotent Cells through P300 and Brg1. *Cell Reports* 25, 1756–1771.
- [32] Huang, J., Li, K., Cai, W., Liu, X., Zhang, Y., et al. (2018) Dissecting super-enhancer hierarchy based on chromatin interactions. *Nature Communications* 9, 943.
- [33] Tewary, S. K., Zheng, Y. G., and Ho, M. C. (2019) Protein arginine methyltransferases: insights into the enzyme structure and mechanism at the atomic level. *Cellular and Molecular Life Sciences* 76, 2917–2932.
- [34] Huang, S., Li, X., Yusufzai, T. M., Qiu, Y., and Felsenfeld, G. (2007) USF1 recruits histone modification complexes and is critical for maintenance of a chromatin barrier. *Mol Cell*

- Biol 27, 7991–8002.
- [35] Wang, H., Huang, Z. Q., Xia, L., Feng, Q., Erdjument-Bromage, H., et al. (2001) Methylation of histone H4 at arginine 3 facilitating transcriptional activation by nuclear hormone receptor. *Science* 293, 853–857.
- [36] Li, X., Hu, X., Patel, B., Zhou, Z., Liang, S., et al. (2010) H4R3 methylation facilitates beta-globin transcription by regulating histone acetyltransferase binding and H3 acetylation. *Blood* 115, 2028–2037.
- [37] Chiang, K., Zielinska, A. E., Shaaban, A. M., Sanchez-Bailon, M. P., Jarrold, J., et al. (2017) PRMT5 Is a Critical Regulator of Breast Cancer Stem Cell Function via Histone Methylation and FOXP1 Expression. *Cell Reports* 21, 3498–3513.
- [38] Scaglione, A., Patzig, J., Liang, J., Frawley, R., Bok, J., et al. (2018) PRMT5-mediated regulation of developmental myelination. *Nature Communications* 9, 2840.
- [39] Rank, G., Cerruti, L., Simpson, R. J., Moritz, R. L., Jane, S. M., et al. (2010) Identification of a PRMT5-dependent repressor complex linked to silencing of human fetal globin gene expression. *Blood* 16, 1585–1592.
- [40] Yuan, C. C., Matthews, A. G., Jin, Y., Chen, C. F., Chapman, B. A., et al. (2012) Histone H3R2 symmetric dimethylation and histone H3K4 trimethylation are tightly correlated in eukaryotic genomes. *Cell Rep* 1, 83–90.
- [41] Migliori, V., Muller, J., Phalke, S., Low, D., Bezzi, M., et al. (2012) Symmetric dimethylation of H3R2 is a newly identified histone mark that supports euchromatin maintenance. *Nat Struct Mol Biol* 19, 136–144.

- [42] Li, J. J., Zhou, F., Zhan, D., Gao, Q., Cui, N., et al. (2012) A novel histone H4 arginine 3 methylation-sensitive histone H4 binding activity and transcriptional regulatory function for signal recognition particle subunits SRP68 and SRP72. *Journal of Biological Chemistry* 287, 40641–40651.
- [43] Zhao, Q., Rank, G., Tan, Y. T., Li, H., Moritz, R. L., et al. (2009) PRMT5-mediated methylation of histone H4R3 recruits DNMT3A, coupling histone and DNA methylation in gene silencing. *Nature Structural and Molecular Biology* 16, 304–311.
- [44] Yang, Y., Lu, Y., Espejo, A., Wu, J., Xu, W., et al. (2010) TDRD3 Is an Effector Molecule for Arginine-Methylated Histone Marks. *Molecular Cell* 40, 1016–1023.
- [45] Tsompana, M., and Buck, M. J. (2014) Chromatin accessibility: a window into the genome. *Epigenetics Chromatin* 7, 33.
- [46] Aberg, K. A., Xie, L., Chan, R. F., Zhao, M., Pandey, A. K., et al. (2015) Evaluation of methyl-binding domain based enrichment approaches revisited. *PLoS ONE* 10, e0132205–e0132205.
- [47] Nair, S. S., Coolen, M. W., Stirzaker, C., Song, J. Z., Statham, A. L., et al. (2011) Comparison of methyl-DNA immunoprecipitation (MeDIP) and methyl-CpG binding domain (MBD) protein capture for genome-wide DNA methylation analysis reveal CpG sequence coverage bias. *Epigenetics* 6, 34–44.
- [48] Yong, W. S., Hsu, F. M., and Chen, P. Y. (2016) Profiling genome-wide DNA methylation. *Epigenetics and Chromatin* 9, 26.
- [49] Han, J., Zhang, Z., and Wang, K. (2018) 3C and 3C-based techniques: the powerful tools

for spatial genome organization deciphering. *Mol Cytogenet* 11, 21.

- [50] Li, T., Jia, L., Cao, Y., Chen, Q., and Li, C. (2018) OCEAN-C: mapping hubs of open chromatin interactions across the genome reveals gene regulatory networks. *Genome Biol* 19, 54.
- [51] Zhang, Q., and Cao, X. (2019) Epigenetic regulation of the innate immune response to infection. *Nature Reviews Immunology* 19, 417–432.
- [52] Mogensen, T. H. (2009) Pathogen recognition and inflammatory signaling in innate immune defenses. *Clinical Microbiology Reviews* 22, 240–273.
- [53] Zhang, J. M., and An, J. (2007) Cytokines, inflammation, and pain. *International Anesthesiology Clinics* 45, 27–37.
- [54] Allis, C. D., and Jenuwein, T. (2016) The molecular hallmarks of epigenetic control. *Nature Reviews Genetics* 17, 487–500.
- [55] Lieberman, P. M. (2016) Epigenetics and Genetics of Viral Latency. *Cell Host and Microbe* 19, 619–628.
- [56] Schäfer, A., and Baric, R. S. (2017) Epigenetic landscape during coronavirus infection. *Pathogens* 6, 8.
- [57] Zhao, D., Zhang, Q., Liu, Y., Li, X., Zhao, K., et al. (2016) H3K4me3 Demethylase Kdm5a Is Required for NK Cell Activation by Associating with p50 to Suppress SOCS1. *Cell Reports* 15, 288–299.
- [58] Li, X., Zhang, Q., Shi, Q., Liu, Y., Zhao, K., et al. (2017) Demethylase Kdm6a

- epigenetically promotes IL-6 and IFN- β production in macrophages. *Journal of Autoimmunity* 80, 85–94.
- [59] Stender, J. D., Pascual, G., Liu, W., Kaikkonen, M. U., Do, K., et al. (2012) Control of Proinflammatory Gene Programs by Regulated Trimethylation and Demethylation of Histone H4K20. *Molecular Cell* 48, 28–38.
- [60] Austenaa, L., Barozzi, I., Chronowska, A., Termanini, A., Ostuni, R., et al. (2012) The Histone Methyltransferase Wbp7 Controls Macrophage Function through GPI Glycolipid Anchor Synthesis. *Immunity* 36, 572–585.
- [61] Liu, Y., Zhang, Q., Ding, Y., Li, X., Zhao, D., et al. (2015) Histone Lysine Methyltransferase Ezh1 Promotes TLR-Triggered Inflammatory Cytokine Production by Suppressing Tollip. *The Journal of Immunology* 194, 2838–2846.
- [62] Li, X., Zhang, Q., Ding, Y., Liu, Y., Zhao, D., et al. (2016) Methyltransferase Dnmt3a upregulates HDAC9 to deacetylate the kinase TBK1 for activation of antiviral innate immunity. *Nature Immunology* 17, 806–815.
- [63] Suhara, W., Yoneyama, M., Kitabayashi, I., and Fujita, T. (2002) Direct involvement of CREB-binding protein/p300 in sequence-specific DNA binding of virus-activated interferon regulatory factor-3 holocomplex. *Journal of Biological Chemistry* 277, 22304–22313.
- [64] Wang, C., Wang, Q., Xu, X., Xie, B., Zhao, Y., et al. (2017) The methyltransferase NSD3 promotes antiviral innate immunity via direct lysine methylation of IRF3. *Journal of Experimental Medicine* 214, 3597–3610.

- [65] Smale, S. T., Tarakhovsky, A., and Natoli, G. (2014) Chromatin Contributions to the Regulation of Innate Immunity. *Annual Review of Immunology* 32, 489–511.
- [66] Ramirez-Carrozzi, V. R., Braas, D., Bhatt, D. M., Cheng, C. S., Hong, C., et al. (2009) A Unifying Model for the Selective Regulation of Inducible Transcription by CpG Islands and Nucleosome Remodeling. *Cell* 138, 114–128.
- [67] Hargreaves, D. C., Horng, T., and Medzhitov, R. (2009) Control of inducible gene expression by signal-dependent transcriptional elongation. *Cell* 138, 129–145.
- [68] Schoggins, J. W., and Rice, C. M. (2011) Interferon-stimulated genes and their antiviral effector functions. *Current Opinion in Virology* 1, 519–525.
- [69] Jayne, S., Rothgiesser, K. M., and Hottiger, M. O. (2009) CARM1 but not Its Enzymatic Activity Is Required for Transcriptional Coactivation of NF- κ B-Dependent Gene Expression. *Journal of Molecular Biology* 394, 485–495.
- [70] Kogut, M. H., Iqbal, M., He, H., Philbin, V., Kaiser, P., et al. (2005) Expression and function of Toll-like receptors in chicken heterophils. *Developmental and Comparative Immunology* 29, 791–807.
- [71] St Paul, M., Paolucci, S., Barjesteh, N., Wood, R. D., and Sharif, S. (2013) Chicken erythrocytes respond to Toll-like receptor ligands by up-regulating cytokine transcripts. *Res Vet Sci* 95, 87–91.
- [72] Kaiser, P. (2010) Advances in avian immunology-prospects for disease control: A review. *Avian Pathology* 39, 309–324.
- [73] Liniger, M., Summerfield, A., Zimmer, G., McCullough, K. C., and Ruggli, N. (2012)

- Chicken Cells Sense Influenza A Virus Infection through MDA5 and CARDIF Signaling Involving LGP2. *Journal of Virology* 86, 705–717.
- [74] Lee, C.-C., Tung, C.-Y., Wu, C. C., and Lin, T. L. (2019) AVIAN INNATE IMMUNITY WITH AN EMPHASIS ON CHICKEN MELANOMA DIFFERENTIATION-ASSOCIATED GENE 5 (MDA5). *Taiwan Veterinary Journal* 45, 43–55.
- [75] Pasteur (1881) ON THE GERM THEORY. *Science* os-2, 420–422.
- [76] Smith, K. A. (2012) Louis Pasteur, the father of immunology? *Frontiers in Immunology* 3, 1–10.
- [77] Witter, R. L. (2000) Protective efficacy of Marek’s disease vaccines. *Current Topics in Microbiology and Immunology* 255, 57–90.
- [78] Isaacs, A., Lindenmann, J., A, I., and J, L. (1987) Virus interference. I. The interferon. *Journal of Interferon Research* 147, 258–267.
- [79] Kaiser, P., Poh, T. Y., Rothwell, L., Avery, S., Balu, S., et al. (2005) A genomic analysis of chicken cytokines and chemokines. *Journal of Interferon and Cytokine Research* 25, 467–484.
- [80] Jahan, S., Xu, W., He, S., Gonzalez, C., Delcuve, G. P., et al. (2016) The chicken erythrocyte epigenome. *Epigenetics and Chromatin* 9, 1–11.
- [81] Masabanda, J. S., Burt, D. W., O’Brien, P. C., Vignal, A., Fillon, V., et al. (2004) Molecular cytogenetic definition of the chicken genome: the first complete avian karyotype. *Genetics* 166, 1367–1373.

- [82] Wicker, T., Robertson, J. S., Schulze, S. R., Feltus, F. A., Magrini, V., et al. (2005) The repetitive landscape of the chicken genome. *Genome Research* 15, 126–136.
- [83] Burt, D., and Pourquie, O. (2003) Chicken genome - Science nuggets to come soon. *Science* 300, 1669.
- [84] International Chicken Genome Sequencing, C. (2004) Sequence and comparative analysis of the chicken genome provide unique perspectives on vertebrate evolution. *Nature* 432, 695–716.
- [85] Hong, H., Chai, H. H., Nam, K., Lim, D., Lee, K. T., et al. (2018) Non-coding transcriptome maps across twenty tissues of the Korean Black Chicken, Yeonsan Ogye. *International Journal of Molecular Sciences* 19, 2359.
- [86] Williams, A. F. (1972) DNA synthesis in purified populations of avian erythroid cells. *Journal of Cell Science* 10, 27–46.
- [87] Sinclair, G. D., and Brasch, K. (1975) The nucleated erythrocyte: a model of cell differentiation. *Revue Canadienne de Biologie*, 34,287-303.
- [88] Soslau, G. (2020) The role of the red blood cell and platelet in the evolution of mammalian and avian endothermy. *Journal of Experimental Zoology Part B: Molecular and Developmental Evolution* 334, 113–127.
- [89] Gavrilov, V. M. (2013) Origin and development of homiothermy: A case study of avian energetics. *Advances in Bioscience and Biotechnology* 04, 1–17.
- [90] Zhang, D. E., and Nelson, D. A. (1988) Histone acetylation in chicken erythrocytes. Rates of deacetylation in immature and mature red blood cells. *Biochem J* 250, 241–245.

- [91] Delcuve, G. P., and Davie, J. R. (1992) Western blotting and immunochemical detection of histones electrophoretically resolved on acid-urea-Triton- and sodium dodecyl sulfate-polyacrylamide gels. *Analytical Biochemistry* 200, 339–341.
- [92] Jin, C., Zang, C., Wei, G., Cui, K., Peng, W., et al. (2009) H3.3/H2A.Z double variant-containing nucleosomes mark “nucleosome-free regions” of active promoters and other regulatory regions. *Nat Genet* 41, 941–945.
- [93] Delcuve, G. P., and Davie, J. R. (1989) Chromatin structure of erythroid-specific genes of immature and mature chicken erythrocytes. *Biochemical Journal* 263, 179–186.
- [94] Ridsdale, J. A., Hendzel, M. J., Delcuve, G. P., and Davie, J. R. (1990) Histone acetylation alters the capacity of the H1 histones to condense transcriptionally active/competent chromatin. *Journal of Biological Chemistry* 265, 5150–5156.
- [95] Gasaryan, K. G. (1982) Genome Activity and Gene Expression in Avian Erythroid Cells. *International Review of Cytology* 74, 95–126.
- [96] Zhang, D., and Nelson, D. A. (1986) Histone acetylation in chicken erythrocytes. Estimation of the percentage of sites actively modified. *Biochem J* 240, 857–862.
- [97] Beug, H., Palmieri, S., Freudenstein, C., Zentgraf, H., and Graf, T. (1982) Hormone-dependent terminal differentiation in vitro of chicken erythroleukemia cells transformed by ts mutants of avian erythroblastosis virus. *Cell* 28, 907–919.
- [98] Affolter, M., Cote, J., Renaud, J., and Ruiz-Carrillo, A. (1987) Regulation of histone and beta A-globin gene expression during differentiation of chicken erythroid cells. *Mol Cell Biol* 7, 3663–3672.

- [99] Sun, J. M., Penner, C. G., and Davie, J. R. (1993) Repression of histone H5 gene expression in chicken mature erythrocytes is correlated with reduced DNA-binding activities of transcription factors SP1 and GATA-1. *FEBS Letters* 331, 141–144.
- [100] Yamamoto, M., Ko, L. J., Leonard, M. W., Beug, H., Orkin, S. H., et al. (1990) Activity and tissue-specific expression of the transcription factor NF-E1 multigene family. *Genes and Development* 4, 1650–1662.
- [101] Penner, C. G., and Davie, J. R. (1992) Multisubunit erythroid complexes binding to the enhancer element of the chicken histone H5 gene. *Biochemical Journal* 283, 905–911.
- [102] Alev, C., Shinmyozu, K., McIntyre, B. A. S., and Sheng, G. (2009) Genomic organization of zebra finch alpha and beta globin genes and their expression in primitive and definitive blood in comparison with globins in chicken. *Development Genes and Evolution* 219, 353–360.
- [103] Ulianov, S. V., Gavrilov, A. A., and Razin, S. V. (2012) Spatial organization of the chicken beta-globin gene domain in erythroid cells of embryonic and adult lineages. *Epigenetics and Chromatin* 5, 16.
- [104] McGhee, J. D., Wood, W. I., Dolan, M., Engel, J. D., and Felsenfeld, G. (1981) A 200 base pair region at the 5' end of the chicken adult β -globin gene is accessible to nuclease digestion. *Cell* 27, 45–55.
- [105] Jin, C., and Felsenfeld, G. (2006) Distribution of histone H3.3 in hematopoietic cell lineages. *Proc Natl Acad Sci U S A* 103, 574–579.
- [106] Rocha, P. P., Raviram, R., Bonneau, R., and Skok, J. A. (2015) Breaking TADs: Insights

- into hierarchical genome organization. *Epigenomics* 7, 523–526.
- [107] Recillas-Targa, F., Bell, A. C., and Felsenfeld, G. (1999) Positional enhancer-blocking activity of the chicken β -globin insulator in transiently transfected cells. *Proceedings of the National Academy of Sciences of the United States of America* 96, 14354–14359.
- [108] Fishman, V., Battulin, N., Nuriddinov, M., Maslova, A., Zlotina, A., et al. (2019) 3D organization of chicken genome demonstrates evolutionary conservation of topologically associated domains and highlights unique architecture of erythrocytes' chromatin. *Nucleic Acids Res* 47, 648–665.
- [109] Li, Q., Li, N., Hu, X., Li, J., Du, Z., et al. (2011) Genome-wide mapping of dna methylation in chicken. *PLoS ONE* 6, e19428.
- [110] Deaton, A. M., and Bird, A. (2011) CpG islands and the regulation of transcription. *Genes and Development* 25, 1010–1022.
- [111] Okuzaki, Y., Kaneoka, H., Nishijima, K. I., Murakami, S., Ozawa, Y., et al. (2017) Molecular cloning of chicken TET family genes and role of chicken TET1 in erythropoiesis. *Biochemical and Biophysical Research Communications* 490, 753–759.
- [112] Abe, H., and Gemmell, N. J. (2014) Abundance, arrangement, and function of sequence motifs in the chicken promoters. *BMC Genomics* 15, 1–12.
- [113] Grün, J., Redmann-Müller, I., Blum, D., Degen, H. J., Doenecke, D., et al. (1991) Regulation of histone H5 and H1^o gene expression under the control of vaccinia virus-specific sequences in interferon-treated chick embryo fibroblasts. *Virology* 180, 535–542.
- [114] Rubin, C. J., Zody, M. C., Eriksson, J., Meadows, J. R. S., Sherwood, E., et al. (2010)

Whole-genome resequencing reveals loci under selection during chicken domestication.
Nature 464, 587–591.

- [115] Berghof, T. V. L., Parmentier, H. K., and Lammers, A. (2013) Transgenerational epigenetic effects on innate immunity in broilers: An underestimated field to be explored? Poultry Science 92, 2904–2913.

Chapter-2

Methods & Materials

2.1 Chicken Erythrocyte Processing

2.1.1 Ethics statement

All methods involving the use of chickens were approved by the Bannatyne Campus Animal Care Committee (BC ACC) and carried out in accordance with Research Ethics and Compliance guidelines and regulations of the University of Manitoba. The animal use protocol is renewed every year. Technical services regarding animal treatment, blood collection and emergency euthanization were provided by Animal Care & Veterinary Services at the Bannatyne Campus. Batch of 12 birds were purchased per treatment. Controlled drug Ketamine and Rompun/Xylazine was purchased from Central Animal Care Services, University of Manitoba. The University is currently in full compliance with the Canadian Council on Animal Care who has certified that the animal care and use program at the University of Manitoba is in accordance with the standards of Good Animal Practice. The birds were purchased through Central Animal Care Services, University of Manitoba and were housed at the Bannatyne Animal Care facility under standard conditions.

2.1.2 Treatment of chickens to induce anemia

Adult White Leghorn chickens were injected with phenylhydrazine (0.15g phenylhydrazine, 2.5 ml ddH₂O and 3.5 ml 95% EtOH) to make them anemic. The intramuscular injections were given over the course of 6 consecutive days in the following schedule:

| Day | 1 | 2 | 3 | 4 | 5 | 6 |
|-------------|-----|-----|-----|-----|-----|-----|
| Volume (ml) | 0.7 | 0.7 | 0.4 | 0.6 | 0.7 | 0.8 |

The birds were weighed everyday and weight loss was recorded as they went through the process of becoming anemic. The birds were monitored 3 times a day for appearance, weight, behaviour, horn color and body condition. Sickness symptoms include red eyes, drastic comb color change, reduced movement, stopped laying eggs, reduced appetite, not making noise and physical weakness.

2.1.3 Blood collection and polychromatic erythrocyte collection

The chickens were anesthetized on the 7th day by injecting 0.5 ml anesthetic solution (3 parts ketamine: 1 part rompun: 2 parts phosphate-buffered saline) in the breast. The birds were made to bleed by an incision of the jugular vein with a sterilized surgical blade. The blood was collected in 500 ml plastic beakers containing collection buffer (pH 7.4: 75 mM NaCl, 25 mM EDTA, 25 mM Tris-HCl) (1). The final composition of each beaker was blood and collecting buffer in the ratio of 1:1. The blood is then filtered for contaminants by passing it through 4-5 layers of cheesecloth secured over a buchner funnel. The filtered blood was then collected in a 2 L flask. The beakers were kept on ice throughout the blood collection and purification procedure.

Aliquots of 12 ml blood samples were pipetted into cold polycarbonated 15 ml test tubes with lid. The tubes were then centrifuged at 4000 rpm for 5-10 minutes at 4°C in Thermo Scientific™ Sorvall™ ST 8R benchtop centrifuge (SS-34 rotor). Subjecting the samples to centrifugation causes the white blood cells to emerge on the surface of the red blood cell pellet and form a white buffy coat. The white cell layer was then discarded with a Pasteur pipette by vacuum aspiration. The pellet, containing the erythrocytes, was again suspended in collection buffer and washed at least 3-4 times. The packed purified polychromatic erythrocytes were then stored at -80° C for long term storage.

2.2 Erythrocyte Treatment and Assays

2.2.1 Adenosine dialdehyde treatment

Packed polychromatic erythrocytes were resuspended at one-third volume packed erythrocytes and two-thirds volume media minimum essential medium (MEM) Alpha (1X) (Life Technologies, cat#12571-063). For methylation studies, the cells were then treated with and without 10 µM adenosine dialdehyde (Sigma), (a SAH hydrolase inhibitor), for 1 and 12 h in a water bath set to 37°C and rotation speed at 2. Cells were then collected by centrifugation and stored at -80°C (2).

2.2.2 Sodium butyrate treatment

For histone acetylation studies, packed polychromatic erythrocytes were resuspended at one-third volume packed erythrocytes and two-thirds volume MEM Alpha (1X) (Life Technologies, cat#12571-063). The cells were then treated with and without 10 mM sodium butyrate for 1 hour in a water bath set to 37°C, with gentle rotation. Cells were then collected by centrifugation and stored at -80° C.

2.3 Protein quantification and standardization

Protein quantification of histones was performed using Pierce BSA Protein Assay Kit (Thermo Scientific) and bovine serum albumin (BSA) as a standard. BSA stock of 2 mg/mL was serially diluted using nuclease-free water to prepare standard concentrations of 0, 0.2, 0.4, 0.6, 0.8, and 1 µg/mL. Ten µL standards or samples were mixed with 190 µL of Reagent A/B mixture (50:1) in a 96 well clear plate. Each mixture was assayed in triplicate. Plate with the samples and standards were then incubated at 37°C for 30 min. Plate was read using Biotek Epoch spectrophotometer, and protein concentration was determined using a standard curve generated by the Gen5 program.

2.4 Protein Analysis and Electrophoretic Methods

2.4.1 Immunoblotting antibody list

Acid extracted histones from polychromatic erythrocytes were resolved in 15% SDS PAGE gels. Histones (PTMs and variants) were resolved by AUT 15% polyacrylamide gel electrophoresis. Immunoblotting analyses were done with the following antibodies to the following histone PTMs and variants: H4R3me2a (Cat No: 39705, Active motif), H3R2me2s (Cat No: ABE-460, Millipore), H3.3 (Cat No: 09-838, Millipore), H3K4me3 (Cat No: 07-473, Millipore), H3K27ac (Cat No: ab4729, Abcam), H3K4me3 (Cat No: ab8580, Abcam), H4K20me2 (Cat No: ab9052, Abcam), H4K5ac (Cat No: Millipore 07-327), H4 penta hyperacetylated (AcH4) (Cat No: 06-946, Millipore), H3K9K14 (AcH3) (Cat No: ab232952, Abcam) and H3K79me2 (Cat No: 39143, Active motif), H3 (Cat No: ab1791, Abcam).

2.4.2 SDS-PAGE electrophoresis and immunoblotting

Ten µg of histone was denatured by boiling in 2x Laemmli's [0.125 M Tris-HCl, pH 6.8, 4% SDS, 20% glycerol, 10% v/v β-mercaptoethanol (BME), and 0.004% (w/v) bromophenol blue] buffer for 5 mins at 100°C. SDS-polyacrylamide gel electrophoresis (SDS-PAGE) was applied to separate proteins based on their molecular weights according to Laemmli's protocol (3). Proteins

were typically resolved on 10% or 15% polyacrylamide gels using Mini-Protean® 3 Cell apparatus (BioRad). Gels were run at a constant voltage of 120V for approximately 1.5-2.0 hour till dye front migrates out of the gel or to obtain the desired resolution. The gels were then transferred to 0.45 µm nitrocellulose membrane (Bio-Rad) using the Trans-Blot Turbo Transfer System (Bio-Rad) following manufacturer's protocol. Upon completion of the transfer, the nitrocellulose membranes were baked for 20-30 minutes at 65°C in a conventional drying chamber. The membrane was then blocked in the appropriate percentage BSA or milk prepared with 0.1% TBST buffer (0.1% v/v tween-20 in Tris-buffered saline) for one hour at room temperature with shaking. Following the blocking step, the membrane was incubated with appropriate dilution of antibody with fresh blocking buffer for one hour at room temperature with rotation (or overnight at 4°C) on a rotator (Boekel Scientific, Model 260200). The membrane was then washed 3x (10 mins per wash) in 0.1% TSBT. Horseradish peroxidase (HRP) conjugated secondary antibody was then added to the membrane and incubated for one hour at room temperature with rotation followed by 3x 10-minute wash with 0.1% TBST on rocking platform (VWR, model# 200). Developing solution was prepared by mixing ECL substrate and oxidizing agent (Perkin Elmer ECL) in a 1:1 ratio; 1 mL of the cocktail/membrane. Incubate the membranes for 1 minute. The membrane was then exposed for different times and bands were recorded using an AlphaInnotech.

2.4.3 Acetic acid-urea-Triton-X100 (AUT) polyacrylamide gel electrophoresis

Fifteen % polyacrylamide AUT (acid-6.6M/ urea-0.375%/ Triton-X-100-0.3M) mini slab gel was prepared using the following recipe for 0.8 mm plate gel:

| | |
|---------------------------------|-------------|
| Urea | 3.2 g |
| Acrylamide/BIS (30:0.8) | 4 ml |
| 4% (v/v) TEMED/43.1 Acetic acid | 1 ml |
| 0.004% (w/v) riboflavin | 0.8 ml |
| 0.3 M Triton-X-100 | 160 μ l |
| Thiodiglycol | 80 μ l |

Recipe for 0.8 mm plate stacking gel:

| | |
|-------------------------|------------|
| Urea | 0.8 g |
| Acrylamide/BIS (30:0.8) | 0.5 ml |
| 3 M KAc (pH 4) | 0.25 ml |
| 0.004% (w/v) riboflavin | 0.2 ml |
| 0.3 M Triton-X-100 | 40 μ l |
| Thiodiglycol | 20 μ l |

The gel was poured and photopolymerized by placing it in front of a light box for about 5 hours. The samples were prepared for loading by mixing with an equal amount of sample buffer (8 M urea, 0.75 M KAc, pH 4.0, 30% (w/v) sucrose, 0.1% (w/v) pyronin Y) (4). The gel was then set to run at 150 V for 4 hours at 4°C using Mini-Protean® 3 Cell apparatus (BioRad). The gel can be stained with Coomassie blue solution (0.25% Coomassie blue G-250 in 45% methanol and 9% acetic acid) for visualization. For transferring, the unstained gels were first washed 2 X (30 min per wash) in 100 ml of 50 mM acetic acid, 0.5% SDS. The gel was then incubated for 30 min (2X) in 100 ml of buffer 0 (5% β -mercaptoethanol, 2.3% (w/v) SDS, and 62.5 mM Tris-HCl, pH 6.8)

(5). AUT minislab gel was then transferred to 0.45 μm nitrocellulose membrane (positioned at the anode side of the gel) with a Bio-Rad TransBlot apparatus. The protein transfer was carried out in the presence of transfer buffer [25 mM Caps, pH 10, and 20% (v/v) methanol] at 70 V for 2 h or at 30 V overnight with cooling to 4°C. Following transfer, the histones that were transferred onto the nitrocellulose filter were visualized by staining the nitrocellulose filter with ponceau stain. Similar to SDS-PAGE electrophoresis, the membrane was dried and blocked in appropriate percentage BSA or milk prepared with 0.1% TBST for one hour at room temperature with shaking. The membrane was then incubated with appropriate dilution of antibody with fresh blocking buffer for one hour at room temperature with rotation (or overnight at 4°C) on a rotator (Boekel Scientific, Model 260200). The membrane was then washed 3x (10 mins per wash) in 0.1% TBST. HRP-conjugated secondary antibody was then added to the membrane and incubated for one hour at room temperature with rotation followed by 3x 10-minute wash with 0.1% TBST. Developing solution was prepared by mixing ECL substrate and oxidizing agent (Perkin Elmer ECL) in a 1:1 ratio; 1 mL of the cocktail/membrane. Incubate the membranes for 1 minutes. The membrane was then exposed for different times, and band intensities were recorded using Chemidoc.

2.4.4 1% DNA agarose gel

One % agarose gel electrophoresis was used to separate DNA according to size (bp) and was used to determine shearing size and for PCR product analyses. The agarose gel was prepared by dissolving 0.5 g of agarose (Invitrogen) in 50 mL of 1X TBE buffer from 10X stock (1 M Tris base, 1 M boric acid and 0.02 M EDTA pH 8.0). The mixture was heated in the microwave for 1-

2 minutes. 1xGelStar nucleic acid stain (Lonza) was added to the cooled liquid agarose and was solidified. The samples were mixed with DNA loading buffer before loading them onto the gel. The gel was run in the presence of 1xTBE buffer at 150V for 1 hour. The 1xTBE buffer was used as a running buffer for the agarose gel electrophoresis. The gel was then visualized using the FluorChem HD2 software (Alpha-Innotech)/ Chemidoc.

2.4.5 Histone extraction

A frozen pellet of polychromatic erythrocytes (approx. 7 million cells) was resuspended in 1-5 mL of NP-40 buffer (10 mM TRIS-Cl, pH 7.6, 150 mM NaCl, 1.5 mM MgCl₂, 0.1% NP-40, 1 mM PMSF (add just before using), 1 mM sodium orthovanadate (added just before using), 10 mM sodium fluoride (added just before using), 25 mM β-glycerophosphate (added just before using)) and homogenized 10X (size of the homogenizer depends on sample volume). The suspension was then passed through a syringe with 22-gauge needle 3X and centrifuged (SS-34 rotor) at 4000 rpm for 4 min (4°C). The supernatant was then discarded, and the pellet was resuspend in 1-5 ml RSB buffer (10 mM TRIS-Cl pH 7.4, 3mM MgCl₂, 10 mM NaCl, 1mM PMSF (added just before using), 1 mM sodium orthovanadate (add before using), 10 mM sodium fluoride (add before using), 25 mM β-glycerophosphate (add before using). Four N H₂SO₄ was added to a 0.4 N final concentration and left on ice for 20 mins. The sample was then centrifuged, and the supernatant was subjected to TCA extraction to a final concentration of 18% (w/v) to precipitate the histones. The precipitate was collected by discarding the supernatant and washed with 1x 1mL acetone/HCl (5μl acid/ml

acetone) and 2x 1mL acetone. The final pellet was dried and resuspended in ddH₂O (100-500µl). The concentration was quantified using bicinchoninic acid assay (absorbance = 562 nm).

2.5 Staining Methods

2.5.1 Ponceau staining

The blot transfer membrane was placed in a plastic box and stained with Ponceau S stain, which had (0.5% (w/v) Ponceau S dissolved in 1% (v/v) acetic acid, for 30 seconds to 1 minute. The membrane was then destained by several changes of water for 30 seconds to 1 minute each and dried.

2.5.2 Coomassie blue staining

The gel was first soaked in fixing solution (50% (v/v) methanol, 10% (v/v) acetic acid, 40% dH₂O) for 30 mins. The gel was stained with Coomassie Blue R-250 solution (0.25% (w/v) Coomassie blue G-250 in 45% (v/v) methanol and 9% (v/v) acetic acid) for 1-2 hour. The gel was then destained for 2-24 hours with destaining solution (5% (v/v) methanol, 7.5% (v/v) acetic acid, 87.5% dH₂O) with kim wipes on the edge of the box to soak away the dye faster.

2.5.3 8-Anilino-1-naphthalenesulfonate (ANS) staining

The gel was immersed in 100 ml of ANS solution (90 $\mu\text{g}/\text{mL}$ ANS powder in 10 mM Tris-HCl, pH 7.4) and stained for 30 mins at room temperature. The gel was then destained with several changes of ddH₂O water in a rotatory shaker for 30 min -1 hour. The gel was then visualized with UV transilluminator box or chemidoc (6).

2.6 FAIRE (Formaldehyde assisted isolation of regulatory region)

The following steps were used in this protocol:

1. One mL polychromatic erythrocytes were resuspended in 10 ml X1 PBS. The cells were then cross-linked with 0.6% formaldehyde for 10 mins. The reaction was stopped with 125 mM glycine. The cells were washed 3 times with X1 PBS.
2. Resuspend the cell pellet in lysis buffer (10 mM Tris-HCl (pH 8.0), 2% (vol/vol) Triton X-100, 1% SDS, 100 mM NaCl and 1 mM EDTA) and homogenize X10 and leave it on ice for 5 mins. The lysate was briefly centrifuged for 5 mins, 4000 rpm (SS-34 rotor) at 4°C. This homogenization step was repeated twice, and the supernatant was stored (7).
3. The cell lysate was sonicated using focused-ultrasonicator (Covaris S220) to obtain an average DNA fragment size of approximately 300-500 bp (10 min sonication, average incident power (watt) = 7.0, peak incident power (watt) = 140.0, duty factor (percentage) = 5.0, cycle/burst(count) 200, duration (seconds) = 60).

4. The shearing of the DNA was checked for optimization by either running the sample in Bioanalyzer or using 1% agarose DNA gel. Aliquot the remaining cell lysate into 1.5-ml tubes and store at -80°C indefinitely.
5. Centrifuge aliquots of lysate at 15,000–20,000g for 5 min at 4°C to pellet cell debris. The supernatants were transferred to fresh 1.5-ml tubes and 1 volume of phenol/chloroform/isoamyl alcohol (Invitrogen) was added (Figure 2.1). The contents of the tube were then mixed on a vortex for 10 s and then centrifuged at 12,000g for 5 min in a tabletop centrifuge, and the aqueous (top) layer was transferred to another fresh 1.5-ml tubes.
6. One-tenth volume of 3 M sodium acetate (pH 5.2), two volumes of 95% (vol/vol) ethanol and 1 volume of glycogen was added to each tube containing the aqueous layer. Incubate the tubes at -80°C for 30 min or longer.
7. Centrifuge at 12,000g for 15 min at 4°C to precipitate the DNA. The supernatant was carefully aspirated without disturbing the DNA pellets. Wash the pellets with 500 μl ice-cold 70% (vol/vol) ethanol. Centrifuge at 12,000g for 5 min at 4°C and store the DNA pellet.
8. The pellet was then dried with a SpeedVac or by leaving tubes open for 10–20 min, then resuspend the pellets in 50 μL dH₂O.
9. Add 1 μL of DNase-free RNaseA and incubate for 30 min at 37°C . Add 1 μL of proteinase K, incubate at 55°C for 1 h, then incubate overnight at 65°C to reverse DNA-DNA cross-links.

10. The DNA was then purified with Zymo-I spin columns (Qiagen) using 5 volumes of DNA binding buffer and 750 μL of wash buffer for each washing. Elute twice with 10 μL of 10 mM Tris-HCl (pH 7.4).
11. Quantify FAIRE DNA with Qubit or NanoDrop. I recommend fluorometry-based quantification.
12. The ratio of FAIRE DNA isolated with respect to input control DNA isolated should not exceed 5% and will ideally fall in the 1–3% range.

2.6.1 FAIRE DNA Sample Purification using Agencourt AMPure XP Beads

The following steps were used in this protocol:

1. Mix the Agencourt AMPure XP beads to resuspend.
2. Add 45 μL (0.9X) of resuspended Agencourt AMPure XP Beads to the reaction (~50 μL). Mix thoroughly on a vortex mixer or by pipetting up and down at least 10 times.
3. Incubate for 5 minutes at room temperature.
4. Quickly spin the tube in a microcentrifuge and place the tube on an appropriate magnetic stand to separate beads from supernatant. After the solution is clear (about 5 minutes), carefully remove and discard the supernatant. Be careful not to disturb the beads that contain the DNA targets.
5. Add 200 μL of 80% freshly prepared ethanol to the tube while in the magnetic stand. Incubate at room temperature for 30 seconds, and then carefully remove and discard the supernatant.

6. Repeat Step 5 once for a total of 2 washing steps.
7. Air dry beads for 5 minutes while the tube/PCR plate is on the magnetic stand with the lid open. **Caution: Do not over dry the beads. This may result in lower recovery of DNA target.**
8. Remove the tube/plate from the magnet. Elute the DNA target from the beads by adding 23 μL of 10 mM Tris-HCl or 0.1X TE.
9. Mix well on a vortex mixer or by pipetting up and down. Quickly spin the tube in a microcentrifuge and incubate for 2 minutes at room temperature.
10. Put the tube in the magnetic stand until the solution is clear. Without disturbing the bead pellet, carefully transfer 20 μL of the supernatant to a fresh, sterile microfuge tube and store at -20°C .

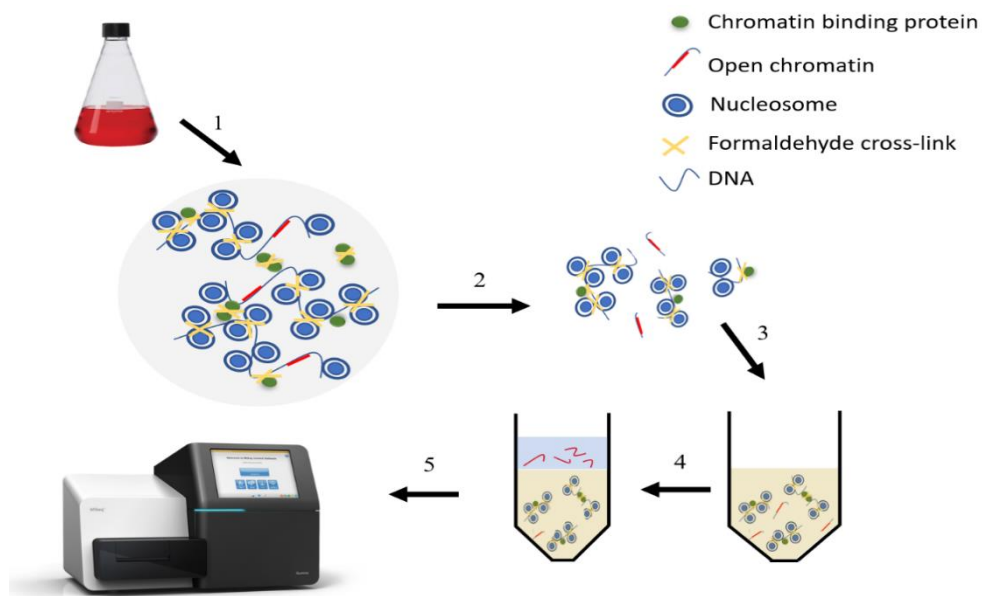


Figure 2.1: Key steps summary of FAIRE. 1. Formaldehyde crosslinking of cells. 2. Sonication to generate DNA fragments. 3. Phenol/chloroform DNA extraction. 4. Isolation of protein free DNA. 5. DNA sequencing.

2.7 Quantitative Polymerase Chain Reaction (qPCR)

In a MicroAmp optical 96-well reaction plate, add 10 μL SYBR Green master mix, 9.8 μL of DNA (0.1 $\text{ng}/\mu\text{L}$), 0.1 μL of Forward and Reverse primers (working concentration as per IDT) for a total of 20 μL reaction per well. Each primer was optimized using the temperature gradient. The template was checked for primer-dimer using 1% agarose gel. The verified templates were sequenced by the RIOH sequencing facility. The following program was used:


98.0°C for 3:00 minutes

98.0°C for 0:15 minutes

62.0°C for 0:30 minutes*

72.0°C for 0:15 minutes

GOTO 2 (39X)



95.0°C for 0:10 minutes

62.0°C for 0:05 minutes*

95.0°C for 0.5 minutes

* Melting curve °C / or the temperature specific for the primer

2.8 FAIRE-seq DNA Library Preparation and Sequencing

The FAIRE-seq DNA library was prepared by the NGS facility of the University of Manitoba using NEBNext Ultra II DNA Library Prep Kit for Illumina (NEB#E7645). The final library was size selected for 300-400 bp. PCR Enrichment of Adaptor-ligated DNA was performed using NEBNext Multiplex Oligos for Illumina (Set 2, NEB #E7500). All DNA measurements were performed using Qubit. DNA library was quantified using KAPA Library Quantification Kit Illumina platform. The quantified libraries were loaded on the flowchip of the Illumina MiSeq in the Manitoba NGS facility for sequencing.

2.9 Bioinformatic Analysis

To map out the regions of FAIRE-seq data, active histone PTMs (H3K4me3, H3K27ac, H3R2me2s and H4R3me2a) ChIP-seq data and active domain F1-seq data were analyzed simultaneously. FAIRE-seq FASTQ file was analyzed using a pipeline in Linux to align it with the latest genome galGal6 to generate a bam file comparable to the ChIP-seq bam files for H3K4me3, H3K27ac, H3R2me2s and H4R3me2a in galGal6 (generated by a bioinformatic centre) (Figure 2.2). A similar pipeline was used to analyze and realign F1-seq data. The FASTQ files were initially monitored for quality of raw sequence data coming from high throughput sequencing pipelines using FASTQC. The FASTQ file was trimmed and adjusted for sequence length. The

file was then mapped against the reference genome using bowtie2. The bam file was further indexed and sorted using SAMtools and Bedtools. Filtered bam files can then be used for peak calling (narrow and broad) against input files using Macs2. Bedfiles or bedgraphs can be visualized with Integrative Genome Browser (IGV) or UCSC browser. The peak annotation was done using Homer. R and deepTools were used to do the downstream integrative analysis of FAIRE peak enrichment around TSS and CGIs. A compute matrix was generated for computing signal distribution relative to scale-region/reference-point. The sliding window and bin size were adjusted as per genomic distribution resolution. FAIRE-seq, F1-seq, CHIP-seq (H3K4me3, H3K27ac, H3R2me2s and H4R3me2a) and RNA sequencing (RNA-seq) files were also analyzed using Partek Flow software. The pipeline designed was used for peak calling, genome annotation, TSS profiling, Venn diagram, visualization and downstream integrative analysis was uploaded in Partek Flow cloud (Figure 2.3). Gene Ontology was performed using the Panther classification system.

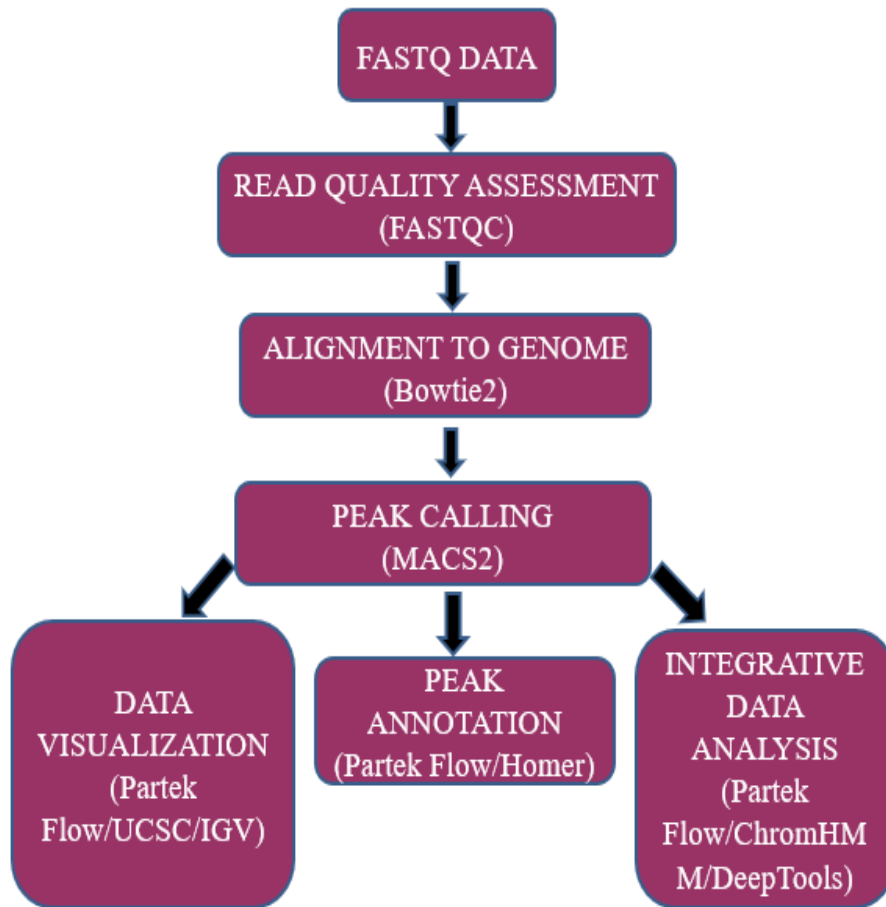


Figure 2.2: Pipeline generated in Linux to process FASTQ files to produce intermediate files (bam/sam/bed) to bedgraphs or bigwigs that can be visualized using IGV, Partek Flow or UCSC browser.

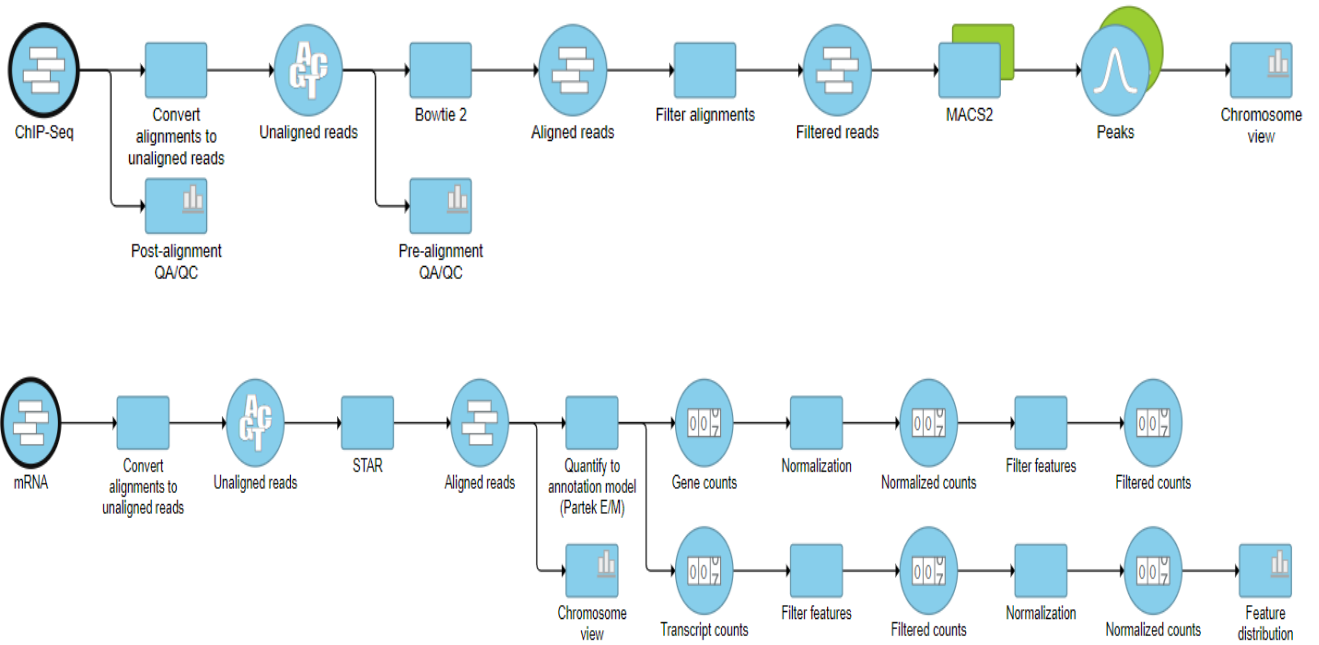


Figure 2.3: The bioinformatic pipeline designed to process ChIP-seq and RNA-seq data. The flow is composed of tools required for peak calling, alignment, annotation, integrative analysis and visualization.

2.10 References

- [1] Delcuve, G. P., and Davie, J. R. (1989) Chromatin structure of erythroid-specific genes of immature and mature chicken erythrocytes. *Biochemical Journal* 263, 179–186.
- [2] Hendzel, M. J., and Davie, J. R. (1992) Acetylation and methylation of histones H3 and H4 in chicken immature erythrocytes are not directly coupled. *Biochemical and Biophysical Research Communications* 185, 414–419.
- [3] Laemmli, U. K. (1970) Cleavage of structural proteins during the assembly of the head of bacteriophage T4. *Nature* 227, 680–685.
- [4] Walker, J. M., and Waterborg, J. H. (2002) Acid-Urea-Triton Polyacrylamide Gel Electrophoresis of Histones. *Protein Protocols Handbook*, 113–123.
- [5] Delcuve, G. P., and Davie, J. R. (1992) Western blotting and immunochemical detection of histones electrophoretically resolved on acid-urea-Triton- and sodium dodecyl sulfate-polyacrylamide gels. *Analytical Biochemistry* 200, 339–341.
- [6] Ma, W., Zhu, G., Zhao, T., Cong, W., Ye, W., et al. (2013) An improved protocol for better detection of protein using 8-anilino-1-naphthalenesulfonate. *Analytical Sciences* 29, 255–262.
- [7] Simon, J. M., Giresi, P. G., Davis, I. J., and Lieb, J. D. (2012) Using formaldehyde-assisted isolation of regulatory elements (FAIRE) to isolate active regulatory DNA. *Nature Protocols* 7, 256–267.

Chapter-3

Dynamic histone acetylation and impact on histone PTMs and variants

3.1 Introduction

3.1.1 Active Histone Post-Translational Modifications

Epigenetic mechanisms are involved in the regulation of all the biological processes in the living body. The regulatory circuit involves interplay among the epigenetic regulators and their modulated events. Some of the key players involved in regulating such events include: histone post-translational modifications, histone variants, DNA methylation, chromatin regulatory proteins and regulatory RNA. The basic unit of chromatin is the nucleosome. It assumes a “beads on a string” structure where the DNA (147 bp) assumes a 1.65 turn around a histone octamer core, consisting of two H2A/H2B and H3/H4 heterodimers. H1, a linker histone, joins the nucleosomes together and enhances chromatin fibre folding to fit the DNA and interacting proteins within the nucleus. Apart from H1 mediated compaction, chromatin organization is modulated through regulation of histone variant (substitutes for canonical histones), chromatin remodeler (histone slide and eviction), and histone chaperones (histone deposition and eviction). Chromatin organization impacts gene regulation and DNA-dependent processes (1). The events of nucleosome positioning within the genome are both evolutionary and modified through regulatory processes to expose the DNA sequences accessible to proteins governing transcription, DNA repair, replication and other regulators involved in maintaining the chromatin landscape.

Interphase chromosomes are composed of two different compartments (10 kb – few Mb): compartment A (euchromatin) is typically associated with active histone PTMs and compartment B is associated with repressed chromatin (heterochromatin). The active state is defined by transcriptional activity, associated histone PTMs, and accessible DNA. Highly expressed genes harbour more asymmetric distributions of different PTMs on both sides of the TSS. Histone acetylation usually marks active genes as does H3K4me1/2/3 whereas H3K9me2/3 and H3K27me3 (written by the EZH2 subunit of the PRC2 complex) constitute repressive marks (2). H3K4me3 is a modification associated with promoters; H3K4me2 is associated with promoters and enhancers, H3K4me1 is associated with enhancers. H3K27me3, H3K4me3 and H3K4me2 jointly mark poised promoters. H3K9ac (promoter) and H3K27ac (enhancer) are associated with active regulatory regions. Arginine marks H4R3me2a and the PRMT5-catalyzed H3R2me2s locate largely to introns and intergenic regions of expressed genes (3). H3K36me3 and H4K20me1 associate with transcribed regions. While methylated lysine 20 on H4, H4K20me1 is an active mark, H4K20me2 remains controversial. H4K20me2 is one of the cell-cycle dependent PTMs that regulate transition from compact mitotic chromosomes to decondensed interphase structure of the chromatin. During G1 phase, about 90% of the genome contains H4K20me2 (4). While it is not exactly clear if it is associated with active genes, it is involved in the SET8 mediated compaction. While most acetylation marks are associated with gene activation, H4K20ac is involved in transcriptional repression and is reported to be enriched at promoters of lowly expressed genes (5). Given the multitude functionality of histone PTMs in transcriptional initiation, elongation, enhancer activity, regulator binding and repression, chromatin profiling provides a powerful means to understand their combinatory activity better.

3.1.2 Lysine Acetylation

Each of the core histones contains a globular core domain and a flexible N-terminal tail with an array of highly conserved lysine residues. Estimated number of lysine residues present in each N-terminal tail includes 4 of the first 15 residues in histone H2A (27%), 8 of 24 in H2B (33%), 8 of 36 in H3 (22%) and 5 of 20 in H4 (25%) (6). The main sites of acetylation for the histones are: H2A, K5; H2B, K12, K15, K20, K24; H3, K9, K14, K18, K23; H4, K5, K8, K12, K16 (7). Histone acetylation was first reported by Vincent Allfrey and his group in 1964. The enzymes KAT and HDAC function by catalysing the addition and removal of acetylation, respectively. Till date, three of the best studied KAT families include: MYST family, GNAT family and p300 family. Allfrey's group also identified HDAC activity in the nuclei. HDACs exist in 4 different classes. Class I- IV are referred as "classic" HDACs and class III members are referred as sirtuins (8). In 1978, both Davie and Allfrey reported that n- butyrate is an HDAC inhibitor (9). Acetylated lysine on the N-terminal is read by the bromodomain module and Tandem-PHD containing at least two PHD domains. Bromodomain can also recognize non-histone acetylation, for example, BRD3 recognizes the CBP acetylated transcription factor GATA1(10). Acetylation mediates multiple signaling cascades regulating embryonic development, stress response , metabolism and diseases such as neurological disorders (5).

3.1.3 Dynamic Histone Acetylation

Active chromatin domains are distinguished by their increased sensitivity to DNase I and association with highly acetylated histones. Transcriptionally active chromatin is associated with histones that are rapidly acetylated and deacetylated (1-2%), whereas bulk of the genome are either unacetylated or statically mono- or di-acetylated. Dynamic studies of acetylation in chicken erythrocytes reported only a small percentage (3.7% in chicken polychromatic and 2.1% in chicken mature erythrocytes) of the modified lysine residues of the N-terminal tails of core histones became dynamically acetylated and deacetylated (11). Upon sodium butyrate treatment for one hour, hyperacetylation of H2B, H3 and H4 and to a lesser extent H2A were observed. Subsequently when incubated in the absence of n-butyrate, these hyperacetylated histones were rapidly deacetylated ($t_{1/2} \sim 5$ min for tetra-acetylated H4) but deacetylated slowly when mono- or di-acetylated ($t_{1/2} \sim 90$ min) (12). Histone hyperacetylation causes chromatin decondensation accessibility of DNA to transcription factors, whereas histone hypoacetylation promotes chromatin condensation and transcriptional repression. Salt solubility of transcriptionally active chromatin corresponds with the level of histone acetylation. Therefore, the histone acetylation state is the primary hinderance against H1-mediated chromatin folding. The estimation of acetylation or methylation rate are often affected by heterogeneity of their associated PTMs due to continuous turnover, transient deposition and removal and heterogeneous cell population.

3.1.4 Histone Variants

All histones have variants that are expressed during DNA synthesis (e.g. H3.1, H3.2 and H2A.1; replication dependent) and those that are expressed throughout the cell cycle (H3.3, H2A.Z; replication independent). Although variants/replacement histone of the same canonical histone, H3.1 and H3.3 are components of different histone assembly complexes and function differently. H3.3 protein differs by only four and five unique residues from histones H3.2 and H3.1. H3.1 and H3.2 have repressive and active histone marks; in contrast, H3.3 is enriched in active histone marks. Mutation studies of variants and their associated chaperone proteins provide evidence suggesting their role in regulation of diverse processes such as differentiation and proliferation, meiosis and nuclear reprogramming. Deposition of H3.3 and H2A.Z are observed in both active and repressed regions of the chromatin. Nucleosomes containing both variants exhibit an unusual degree of sensitivity to restriction enzymes and instability in the cell setting possibly because the dimer containing nucleosome is heavily post-translationally modified. Accumulation of H3.3 with age in vertebrate non-dividing tissues can represent up to 60% or more of the total H3 contents of chromatin, reducing its destabilizing role. In mouse, H2B.1, H2A.2 and H2A.X increase as H3.3 increases with mouse age, and levels of H2B.2 and H2A.1 decrease (13). Both H2A.Z and H3.3 variants are targeted to the nucleosomes near TSS, in a histone acetylation dependent way, by chromatin remodelers RSC, SWR, and HIRA. H3.3 is not always associated with actively transcribed genes; the death associated protein, DAXX, and the α -thalassemia X-linked mental retardation protein, ATRX, are reported to deposit H3.3 into telomeric and pericentric heterochromatin regions. This suggests that H3.3 functions as a boundary factor for

chromatin regulation rather than being directly involved in promoting transcription (14). A recent study in mammalian sperm reported involvement of variant H3.3 in transgenerational epigenetic inheritance. The study showed direct association of H3.3 and H3K4me3-enriched imprinted loci involved in conveying instructive epigenetic information to the zygote (15). Taken together H3.3 plays crucial role in the establishment of the chromatin landscape.

3.1.5 Study Design and Objective

Chicken polychromatic erythrocyte histones participating in dynamic acetylation become hyperacetylated after one-hour treatment of the cells with sodium butyrate. Applying this well documented protocol, I determined which of the H3 variants and H3/H4 PTMs were participating in dynamic histone acetylation.

3.2 Methods and Materials

3.2.1. Ethics statement

All methods involving the use of chicken were approved by the committee and carried out in accordance with its guidelines and regulations. The University is currently in full compliance with the Canadian Council on Animal Care who has certified that the animal care and use program at the University of Manitoba is in accordance with the standards of Good Animal Practice. The

birds were purchased through Central Animal Care Services, University of Manitoba and were housed under standard conditions.

3.2.2. Preparation of chicken polychromatic erythrocytes and sodium butyrate treatment

Anemia was induced in female Adult White Leghorn chickens by phenylhydrazine treatment, and pooled cells were purified as described in Risdale et al, 1990 (16). Polychromatic erythrocytes were incubated with 10 mM sodium butyrate for 60 min in (MEM) Alpha (1X) (Life Technologies, cat#12571-063) to induce a state of histone hyperacetylation.

3.2.3 Protein analysis and immunoblotting

Histones were acid extracted from the nuclei of cells collected from both treated and untreated polychromatic erythrocytes and subjected to 15% acid/urea/Triton X-100 gel electrophoresis (17) and immunoblotting (as mentioned in chapter 2) with antibodies (18) against H4R3me2a (Cat No: 39705, Active motif), H3R2me2s (Cat No: ABE-460, Millipore), H3.3 (Cat No: 09-838, Millipore), H3K27ac (Cat No: ab4729, Abcam), H3K4me3 (Cat No: ab8580, Abcam), H4K20me2 (Cat No: ab9052, Abcam), H4K5ac (Cat No: Millipore 07-327), H4 penta hyperacetylated (H4Ac) (Cat No: 06-946, Millipore), H3K9K14 (H3Ac) (Cat No: ab232952, Abcam) and H3K79me2 (Cat No: 39143, Active motif), H3 (Cat No: ab1791, Abcam), and H3K9ac (Cat No: ab4441, Abcam).

3.3 Results

3.3.1 H3 Tail Modifications

One of the most common methods of separating histone variants from their canonical counterparts is using AUT polyacrylamide gel electrophoresis which resolves histones according to size, charge and hydrophobicity (17). Treatment of chicken polychromatic erythrocytes with sodium butyrate induces hyperacetylation of histones in the small population of chromatin that has dynamically acetylated histones (11). Acetylation shifts the net positive charge of the histone/variant involved in dynamic acetylation to a less positively charged state causing a shift up in the acetylated histone compared to unmodified histone. On the other hand, methylation preserves the positive charge on the nitrogen atom of the ϵ -amino group, and therefore there is no change in the mobility of the methylated histone. However, a histone may also participate in dynamic acetylation but not show a change in mobility in the gel after histone deacetylase inhibition. In such cases, the histone acetylated at mainly one lysine site will exhibit an increase in intensity when HDAC activity is inhibited by butyrate..

The Commassie Blue stained AUT gel shows that H3.2 is more abundant than H3.3 in polychromatic erythrocytes (Figure 3.1). The relative staining intensity of H3.2 and H3.3 did not change after butyrate treatment. Further, there was no evidence for the reduced mobility of H3.2 or H3.3 in butyrate treated cells, consistent with a small population of histones participating in dynamic acetylation.

Next I determined whether specific H3 PTMs and variants had a reduced mobility after butyrate treatment, which would indicate that these histone forms had multiple acetylation sites participating in dynamic acetylation and had multiple acetylation sites, or a change in intensity (Figure 3.1). The histones were resolved on separate AUT gels and then immunostained with the indicated antibody. Going from left to right in Figure 3.1, I will present the results of each immunoblot. For total H3, I found no indication for reduced mobility of H3.2 and H3.3 after butyrate treatment. As resolution of H3.2 and H3.3 was poor in this AUT gel, I used an antibody specific against H3.3. As with the stained gel, reduced mobility of H3.3 after butyrate treatment was not observed. H3.2 and H3.3 were both acetylated, and H3.2Ac and H3.3Ac showed a shift expected after butyrate incubation and an increase in intensity. With H3.3, a major shift after butyrate incubation was not observed. Although some literature reports H3.3 incorporation in heterochromatin regions, further experimentation such as ChIP-seq would provide an answer. H3.2 and H3.3 were equally modified by R2me2s. When compared to the stained gel, which showed H3.2 being more abundant than H3.3, it is evident that H3.3 was preferentially modified by R2me2s. Butyrate treatment did not result in a reduced mobility of H3.2 or H3.3 modified by R2me2s. K4me3 was seen with both H3.2 and H3.3. However, as with R2me2s, H3.3 was preferentially modified by K4me3. The major shift in the gel after butyrate treatment suggests H3.3 K4me3, and perhaps H3.2, are dynamically acetylated and acetylated at multiple sites which is consistent with Mahadevan's finding (19). Both H3.2 and H3.3 were acetylated at K9, but there appeared to be a preferential acetylation of K9 of H3.3. An increase in intensity and perhaps in mobility was observed for H3.2K9ac and H3.3K9ac. K27ac, an enhancer mark, was seen with both H3.2 and H3.3 fairly equally, suggesting that H3.3 was enriched in this PTM. Although a change

in mobility of H3.2K27ac and H3.3K27ac was not observed after butyrate treatment, the intensity of H3.2K27ac and H3.3K27ac increased. Both H3.2 and H3.3 had K79me2, which is a PTM associated with transcriptional elongation (20). There may be a change in intensity, but not in mobility of this modified H3.

In summary, H3.3 was enriched in active marks K9ac, K27ac, K4me3 and R2me2s, consistent with previous reports (21). H3.3K4me3 had multiple acetylation sites and became hyperacetylated after butyrate treatment.

(H3 PTMs)

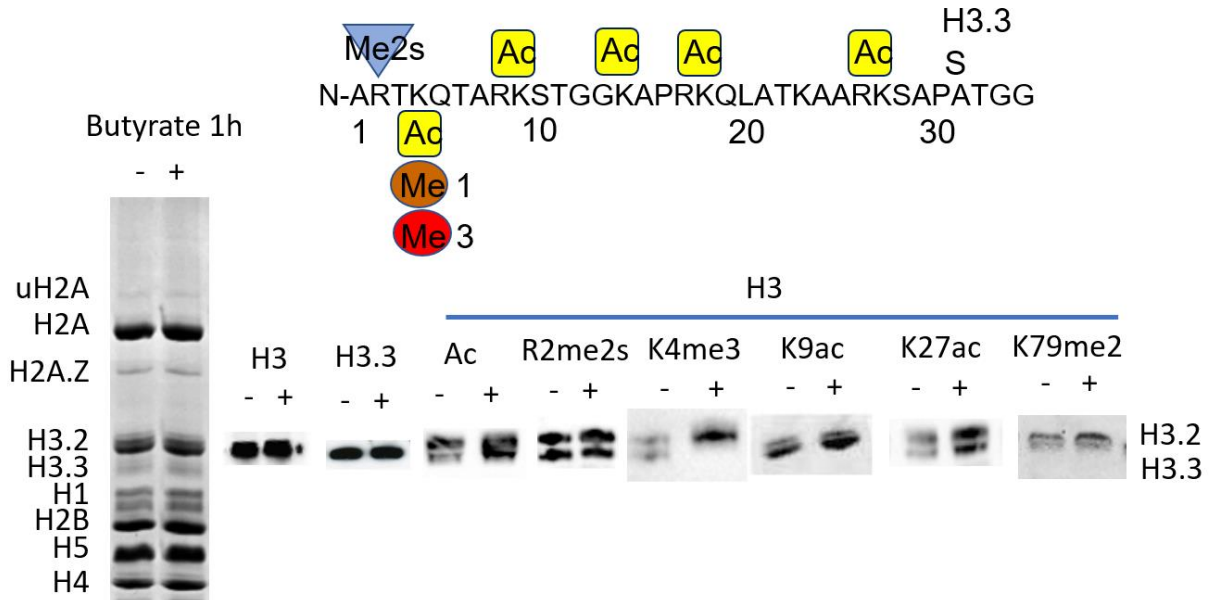


Figure 3.1: PTMs analyzed on H3-tail. Acid extracted histones from n-butyrate treated and non-treated polychromatic erythrocytes were electrophoretically separated on 15% AUT gel. The histones were then transferred on to the nitrocellulose membrane and blotted with antibodies against the indicated histone variant or PTM. The Coomassie Blue stained AUT gel show the histone patterns.

3.3.2 H4 Tail Modifications

The acetylation state of H4 was investigated in a similar way by comparing sodium butyrate treated cells versus non-treated cells run side by side. The Coomassie Blue stained gel shows equal loading of histones (Figure 3.2). Further, there was no evidence for the reduced mobility of H4 in buyrate treated cells, consistent with a small population of histones participating in dynamic acetylation.

Going from left to right in Figure 3.2, I will present the results of each immunoblot. The blot imunoastained for H4Ac serves as a control to show the different acetylated forms before and after butyrate treatment. In the absence of butyrate, H4 was mainly monoacetylated. After buyrate treatment, there was an increase in intensity of the four acetylated forms of H4. H4 was modified with more R3me2a after butyrat treatment. Although the band intensity increased, there was no major shift up in the gel. H4K5ac intensity increased and exhibited a reduced mobility after butyrate incubation. A major change was seen with H4K20me2, which although considered repressed, significantly increased in intensity after butyrate treatment. However, a reduced mobility of H4K20me2 was not observed.

In summary, the intensities of H4 modified at R3me2a, K5ac and K20me2 increased after butyrate treatment.

(H4 PTMs)

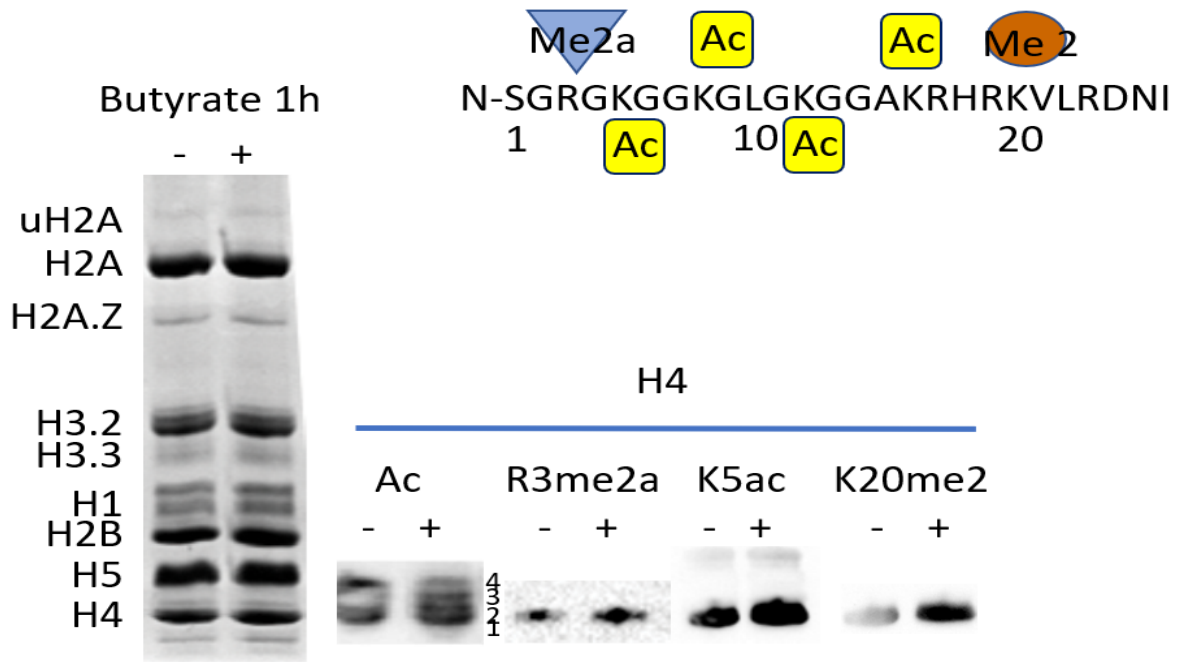


Figure 3.2: PTMs analyzed on H4-tail. Acid extracted histones from n-butyrate treated and non-treated polychromatic erythrocytes were electrophoretically separated on 15% AUT gel. The histones were then transferred on to the nitrocellulose membrane and blotted with antibodies against the indicated H4 PTMs. The Coomassie Blue stained AUT gels show the histone patterns.

3.4 Discussion

On global level, dynamic histone acetylation is observed using immunoblot analyses of histones resolved by AUT gel electrophoresis. The butyrate treatment causes dynamically acetylated H3 and H4 histones to undergo rapid hyperacetylation, suggesting that the dynamic acetyl groups are continuously turned over by the combined effort of KATs and HDACs. Among the H3 variants, H3.3 is preferentially modified with active marks and incorporated to rebuild nucleosomes displaced at regulatory regions of the genome. The PTMs investigated on the H3 tail suggest that not all of the H3.3-associated PTMs are altered when HDACs are inhibited (for example, H3.3R2me2s). In a previous study, we had shown that genes involved in innate immunity and had low expression were associated with unstable nucleosomes. One example is the *TLR21* gene in which the gene body was not enriched in F1 (indicates overall acetylation state). In this example, H3.3 could rebuild nucleosomes that are re-associating with the gene body after CHD1 displacement. In such a case, my results suggest that H3.3 would not be acetylated.

H3.3K4me3 and H3.3K9ac were hyperacetylated after the butyrate incubation. An increase in intensity and a decrease in mobility in the AUT gel is consistent with H3.3K4me3 and H3.3H9ac being further acetylated at multiple lysine sites after HDAC inhibition. Early studies by Vincent Allfrey and colleagues on gene induction, correlated hyperacetylation with gene induction and showed H3K9ac and H3K4me3 had comparable distributions across promoters of inducible genes and map specifically to first exon-intron boundaries (22). My result is consistent with this finding. In contrast to H3.3K4me3, H3.3K27ac did not show a reduced mobility but had an increase in

intensity after butyrate treatment. The increase in intensity could be due to CBP/p300 increasing the acetylation at this site in the absence of HDAC activity. This observation also shows that the other acetylation sites (e.g. K9, K14) of an H3.3 tail acetylated at K27 are not becoming more acetylated.

Our previous finding along with the literature suggested H3R2me2s is located together with H3K4me3 in intergenic regions and introns. This result is consistent with the finding that R2me2s and K79ac are found in H3.3 and H3.2. One study reported that the rate of nucleosome turnover is negatively correlated with H3K27me3 at regulatory regions and with H3K36me3 at gene bodies (23). Taken together, my data suggest open chromatin H3 PTMs (that is, H3 in nucleosomes located next to promoters) are more dynamically acetylated and preferentially decorate H3.3 whereas H3 PTMs in the gene body do not participate in dynamic acetylation.

Dynamic acetylation of H4 was also consistent with what has been reported in the literature. Like R2me2s on the H3 tail after butyrate treatment, R3me2a on the H4 tail did not show a drastic shift in band mobility. One speculation for this observation is that increased histone acetylation alters the activities of demethylating enzymes; PRMT and/or demethylating enzyme and/or remodelers. It has been reported that increase of H4 acetylation during early response to butyrate on K5ac, K8ac, K12ac and K16ac takes place primarily on histones containing K20me2 (24). Combinatory PTM analysis by mass spectrometry would provide further explanation. Another study reports elevated H4 acetylation (H4ac) was linked to four-fold increase in BRCA1 occupancy coupled with diminished 53BP1 association to DSB chromatin occupancy to direct DNA repair mechanism concluding acetylation prevents 53BP1 from binding to H4K20me2 (25).

Therefore, the striking change in K20me2 could be due to its acetylation dependent function in DSB repair. Co-immunoprecipitation and interactome analysis will further help us understand how acetylation of histone H4 is proteoform specific and its relation with methylation state. Transcription factors recruit coactivators with KAT activity (e.g., p300/CBP) to promote acetylation, while transcriptional repressors recruit corepressors with HDAC activity. Regulatory motif enrichment analysis will help us target transcription factors that modulate the mechanism of hyperacetylation.

Histone variants play a crucial role in the genome which are often tracked by studying their deposition profile and biophysical properties. These aspects of variants differ from their canonical histones in terms of regulating chromatin accessibility, turnover of nucleosomes and buffering nucleosomal density to maintain transcriptional and heterochromatic regions as they undergo modification by different histone PTMs. Therefore, it is safe to speculate from our results that histone variants and histone acetylation work synergistically to maintain an open chromatin structure. Histone variants are known to control cell specific expression and therefore they might have special deposition around cell-identity genes and their associated PTM domains. Understanding the consequences of increasing or decreasing combinatorial variant deposition and their direct impact on human health is still infancy. Mutations in a single allele of the p300 and CBP genes have been identified in multiple cancers. Translocation of MLL1, KMT family, accounts for ~80% of infant leukemias and 5%–10% of adult leukemias (26). Mis-regulation of HDACs has been linked to multiple cancers as they control hallmarks of cancer cellular processes (8). Current challenges in designing HDAC inhibitors include sufficient potency, selectivity and

cell-permeability (27). Taken together, improving screening methodologies and understanding the mechanism of acetylation regulators is beneficial to human health and disease.

3.5 References

- [1] Rothbart, S. B., and Strahl, B. D. (2014) Interpreting the language of histone and DNA modifications. *Biochimica et Biophysica Acta - Gene Regulatory Mechanisms* 1839, 627–643.
- [2] Hyun, K., Jeon, J., Park, K., and Kim, J. (2017) Writing, erasing and reading histone lysine methylations. *Experimental and Molecular Medicine* 49, e324.
- [3] Beacon, T. H., Xu, W., and Davie, J. R. (2020) Genomic landscape of transcriptionally active histone arginine methylation marks, H3R2me2s and H4R3me2a, relative to nucleosome depleted regions. *Gene* 742.
- [4] Tsang, L. W. K., Hu, N., and Alan Underhill, D. (2010) Comparative analyses of SUV420H1 isoforms and SUV420H2 reveal differences in their cellular localization and effects on myogenic differentiation. *PLoS ONE* 5.
- [5] Sheikh, B. N., and Akhtar, A. (2019) The many lives of KATs — detectors, integrators and modulators of the cellular environment. *Nature Reviews Genetics* 20, 7–23.
- [6] Barnes, C. E., English, D. M., and Cowley, S. M. (2019) Acetylation and Co: An expanding repertoire of histone acylations regulates chromatin and transcription. *Essays in Biochemistry* 63, 97–107.
- [7] Zhang, D., and Nelson, D. A. (1986) Histone acetylation in chicken erythrocytes. Estimation of the percentage of sites actively modified. *Biochem J* 240, 857–862.
- [8] Witt, O., Deubzer, H. E., Milde, T., and Oehme, I. (2009) HDAC family: What are the cancer relevant targets? *Cancer Letters* 277, 8–21.

- [9] Candido, E. P. M., Reeves, R., and Davie, J. R. (1978) Sodium butyrate inhibits histone deacetylation in cultured cells. *Cell* 14, 105–113.
- [10] Davie, J. R. (2003) Inhibition of Histone Deacetylase Activity by Butyrate. *J Nutr* 133, 2485S-2493S.
- [11] Zhang, D. E., and Nelson, D. A. (1988) Histone acetylation in chicken erythrocytes. Rates of acetylation and evidence that histones in both active and potentially active chromatin are rapidly modified. *The Biochemical journal* 250, 233–240.
- [12] Zhang, D. E., and Nelson, D. A. (1988) Histone acetylation in chicken erythrocytes. Rates of deacetylation in immature and mature red blood cells. *The Biochemical journal* 250, 241–245.
- [13] Cheema, M. S., and Ausió, J. (2015) The structural determinants behind the epigenetic role of histone variants. *Genes* 6, 685–713.
- [14] Bush, K. M., Yuen, B. T. K., Barrilleaux, B. L., Riggs, J. W., O'Geen, H., et al. (2013) Endogenous mammalian histone H3.3 exhibits chromatin-related functions during development. *Epigenetics and Chromatin* 6, 7.
- [15] Maze, I., Noh, K. M., Soshnev, A. A., and Allis, C. D. (2014) Every amino acid matters: Essential contributions of histone variants to mammalian development and disease. *Nature Reviews Genetics* 15, 259–271.
- [16] Ridsdale, J. A., Hendzel, M. J., Delcuve, G. P., and Davie, J. R. (1990) Histone acetylation alters the capacity of the H1 histones to condense transcriptionally active/competent chromatin. *Journal of Biological Chemistry* 265, 5150–5156.

- [17] Davie, J. R. (1982) Two-dimensional gel systems for rapid histone analysis for use in minislab polyacrylamide gel electrophoresis. *Analytical Biochemistry* 120, 276–281.
- [18] Delcuve, G. P., and Davie, J. R. (1992) Western blotting and immunochemical detection of histones electrophoretically resolved on acid-urea-Triton- and sodium dodecyl sulfate-polyacrylamide gels. *Analytical Biochemistry* 200, 339–341.
- [19] Crump, N. T., Hazzalin, C. A., Bowers, E. M., Alani, R. M., Cole, P. A., et al. (2011) Dynamic acetylation of all lysine-4 trimethylated histone H3 is evolutionarily conserved and mediated by p300/CBP. *Proceedings of the National Academy of Sciences of the United States of America* 108, 7814–7819.
- [20] Godfrey, L., Crump, N. T., Thorne, R., Lau, I. J., Repapi, E., et al. (2019) DOT1L inhibition reveals a distinct subset of enhancers dependent on H3K79 methylation. *Nature Communications* 10, 1–15.
- [21] Mckittrick, E., Gafken, P. R., Ahmad, K., and Henikoff, S. (2004) Histone H3.3 is enriched in covalent modifications associated with active chromatin.
- [22] Bieberstein, N. I., Oesterreich, F. C., Straube, K., and Neugebauer, K. M. (2012) First exon length controls active chromatin signatures and transcription. *Cell Reports* 2, 62–68.
- [23] Kraushaar, D. C., Jin, W., Maunakea, A., Abraham, B., Ha, M., et al. (2013) Genome-wide incorporation dynamics reveal distinct categories of turnover for the histone variant H3.3. *Genome Biology* 14, R121.
- [24] Wang, T., Holt, M. V., and Young, N. L. (2018) Early butyrate induced acetylation of histone H4 is proteoform specific and linked to methylation state. *Epigenetics* 13, 519–535.

- [25] Tang, J., Cho, N. W., Cui, G., Manion, E. M., Shanbhag, N. M., et al. (2013) Acetylation limits 53BP1 association with damaged chromatin to promote homologous recombination. *Nature Structural and Molecular Biology* 20, 317–325.
- [26] Audia, J. E., and Campbell, R. M. (2016) Histone modifications and cancer. *Cold Spring Harbor Perspectives in Biology* 8, a019521.
- [27] Delcuve, G. P., Khan, D. H., and Davie, J. R. (2013) Targeting class i histone deacetylases in cancer therapy. *Expert Opinion on Therapeutic Targets* 17, 29–41.

Chapter-4

Chromatin organization of transcribed genes in chicken polychromatic erythrocytes

Chromatin organization of transcribed genes in chicken polychromatic erythrocytes

Sanzida Jahan¹, **Tasnim H. Beacon**¹, Shihua He, Carolina Gonzalez, Wayne Xu, Geneviève P. Delcuve, Shuo Jia, Pingzhao Hu, James R. Davie

¹Co-first authors

Source of the published article:

Gene, Volume 699, 30 May 2019, Pages 80-87

<https://doi.org/10.1016/j.gene.2019.03.001>

Author contribution statement

SJ carried out the chromatin fractionation procedures, the ChIP-seq, and RNA-seq assays, prepared the first draft of the manuscript, participated in data interpretation, data analyses, and figure preparation. **THB** did the FAIRE sequencing and assays, data analyses and prepared figures. WX, PH, SJ and **THB** performed the bioinformatics analyses. SH participated in ChIP-seq assays. CG prepared the DNA and RNA libraries. GPD participated in data interpretation, drafted and wrote the manuscript. JRD conceived of the study, participated in its design and coordination and reviewed the manuscript. All authors read and approved the final manuscript.

4.1 Abstract

Transcriptional regulation is impacted by the organization of the genome into chromatin compartments and domains. We previously reported the application of a biochemical fractionation protocol to isolate highly enriched transcribed DNA from chicken polychromatic erythrocytes. In conjunction with next-generation DNA and RNA sequencing as well as chromatin immunoprecipitation-DNA sequencing, we identified all the active chromosomal compartments and determined their structural signatures in relation to expression levels. Highly expressed genes were found in broad dynamically highly acetylated, salt-soluble chromatin compartments, while poorly or moderately expressed genes exhibited a narrow stretch of salt-soluble chromatin limited to their 5' or body region. Here, we present the detailed characteristics, including the location of NDR and CGI, of several transcriptionally active chromatin compartments. These chromatin patterns illustrate how the salt solubility profile of a genomic region aids in the annotation of genes expressed in erythroid cells and contributes to the identification of functional features such as regulatory regions.

Keywords: Nucleosome free chromatin regions, Chromatin fractionation, Histone acetylation, H3K4me3, H3K27ac

4.2 Introduction

In the interphase nucleus, chromatin is organized into compartments and TADs (1). Compartment A contains active genes, while compartment B has repressed genes; these two compartments can be distinguished by their histone and DNA modifications (2, 3). A compartment may have several TADs (4). However, in contrast to other vertebrate tissues, chicken polychromatic erythrocytes lack TADs (5).

The study of chicken erythrocyte chromatin has been important in detailing features of transcribed chromatin and the organization of the genome. DNase I sensitivity and dynamic histone acetylation are features of transcribed chromatin that were first discovered in the chicken erythrocyte model system (6–9). For example, the α - and β -globin gene clusters and their flanking regions were preferentially sensitive to DNase I (a feature of compartment A), compared to the inactive ovalbumin gene (6). The extent of the DNase sensitivity defined the chromatin DNase I-sensitive domain (6, 10). Later studies showed that these preferentially DNase I-sensitive chromatin regions were characterized by the overall presence of highly acetylated histones in chicken, mouse and human erythroid cells (11–14).

Chicken polychromatic erythrocytes isolated from anemic chicken are nucleated G0 phase cells that are non-replicating. Thus, histone post-translational modifications related to cell cycle do not confound the analyses of transcribed chromatin. In our experience, the chicken polychromatic erythrocyte is the only eukaryotic cell source in which a biochemical fractionation

protocol is capable of isolating polynucleosomes (fraction F1 chromatin) that are soluble at physiological ionic strength and are highly enriched in transcribed DNA (that is, isolation of chromatin compartment A) (15). In contrast, chromatin fragments extracted from micrococcal nuclease digested *Drosophila* nuclei by 80 or 150 mM NaCl were enriched in transcribed DNA but were predominantly mononucleosome in length (16). Thus, the much longer 150 mM salt soluble chromatin fragments from chicken polychromatic erythrocytes are more informative about chromatin features of the DNase I-sensitive chromatin domains (compartment A).

Due to a particularly high density of H1/H5 linker histones (1.3-1.4 molecules per nucleosome compared to the typical 0.7 to 0.8 per nucleosome) (17), the bulk of chicken polychromatic erythrocyte chromatin is extremely condensed and insoluble at physiological ionic strength. However, dynamically acetylated histones associated with transcriptionally active/poised chromatin prevent H1/H5 from rendering active/poised gene polynucleosomes insoluble at physiological ionic strength (18). Chicken polychromatic erythrocyte F1 chromatin is enriched in active histone marks (highly acetylated four core histones, H3K4me3, uH2B) and in atypical nucleosomes (U-shaped) (15, 19, 20).

We exploited this powerful chromatin fractionation procedure to isolate transcriptionally active/poised, genome-wide, native chromatin (fraction F1), which was analyzed by next-generation DNA sequencing (F1 DNA-seq) (21). The F1 DNA-seq reads corresponded to DNase I-sensitive chromatin domains in chicken polychromatic erythrocytes. In conjunction with RNA-seq and ChIP-seq, we identified all the active chromosomal compartments and determined their structural signatures in relation to expression levels (21). In combination, these approaches

illustrated the features of the chromatin compartment A containing expressed genes.

The analyses of the F1 DNA-seq patterns containing expressed genes suggested that there was a wealth of information within the F1 DNA-seq data about the organization of transcribed chromatin (21). The F1 DNA-seq reads in a cluster varied in intensity and breadth and often there was a loss of reads in a cluster. We speculated that this loss of reads corresponded to a nucleosome free stretch of DNA containing regulatory regions. To help decipher the F1 DNA-seq information, we applied FAIRE-sequencing to map NDR in the erythrocyte chromatin. Herein we highlight the salient features of several transcriptionally active chromatin regions (compartment A).

4.3. Materials and methods

4.3.1 Ethics statement

All methods involving the use of chicken were approved by the committee and carried out in accordance with its guidelines and regulations. The University is currently in full compliance with the Canadian Council on Animal Care who has certified that the animal care and use program at the University of Manitoba is in accordance with the standards of Good Animal Practice. The birds were purchased through Central Animal Care Services, University of Manitoba and were housed under standard conditions.

4.3.2 Isolation of chicken erythrocytes

Polychromatic erythrocytes were isolated from anemic female Adult White Leghorn chickens as described (15, 21).

4.3.3 Chromatin fractionation

Chicken polychromatic erythrocyte nuclei, prepared as described (15, 22) were incubated with micrococcal nuclease, and chromatin fragments soluble in a low ionic strength solution containing 10 mM EDTA were recovered in fraction S_E . Chromatin fraction S_E was made 150 mM in NaCl, and chromatin fragments from the salt-soluble fraction (S_{150}) were size-resolved on a Bio-Gel A-1.5m column to isolate the F1 fraction containing polynucleosomes (20).

4.3.4 ChIP-sequencing

ChIP-seq assays, using antibodies against H3K27ac or H3K4me3 from Abcam, were done as previously described (21). In brief, chicken polychromatic erythrocytes were treated with 0.5% formaldehyde for 10 min, washed with RSB buffer (10 mM Tris-HCl pH 7.5, 10 mM NaCl, 3 mM MgCl) and pelleted. Following cell lysis in RSB plus 0.5% NP-40 and phosphatase/protease inhibitors, the pellet was resuspended with an appropriate volume (approximately 2 mL) of MNase

digestion buffer (10 mM Tris-HCl pH 7.5, 0.25M sucrose, 75 mM NaCl). Nuclei were lysed using 0.5% SDS. Chromatin was sheared using a probe sonicator (Fisher scientific, sonic dismembrator, model#100) at setting 3. Sonication time was optimized until the average fragment size of 200-300 bp was achieved. Immunoprecipitation with specific antibodies and subsequent procedures were performed as described previously.

4.3.5 FAIRE

FAIRE was optimized according to a previous protocol (23). Polychromatic erythrocytes were cross-linked with 0.6% formaldehyde, and the reaction was stopped with 125 mM glycine. The cell pellet was resuspended in lysis buffer and sonicated using focused-ultrasonicator (Covaris S220) to obtain an average DNA fragment size of approximately 300-500 bp (10 min sonication). The cell lysate was subjected to phenol/chloroform/isoamyl alcohol (Invitrogen) extraction, and the aqueous layer containing DNA (open chromatin) was recovered. The samples were then de-crosslinked and isolated DNA was purified using PCR purification kit (Qiagen). FAIRE was done with three biological repeats either by sequencing or PCR analyses.

4.3.6 Quantitative PCR analysis

FAIRE enrichment was determined using quantitative PCR method. Sequenced amplicons (Supplementary Table 4.1) were used to compare signals from the FAIRE sample relative to the

input. The relative enrichment was calculated using the comparative C_t method for the *FTH1* and *CA2* genes shown in Figure 4.2.

4.3.7 Sequencing and data analyses

DNA libraries and strand-specific RNA libraries preparations, sequencing and mapping of data as well as active chromatin detection (sequences enriched in fraction F1 chromatin) were previously described (21). The raw reads in BAM files were aligned to galGal3 as reference, converted to WIG files (default setting, scale=1) and visualized using IGV, UCSC Genome Browser. The genes were annotated using UCSC RefSeq chicken genes. The sequencing data are available from GEO under accession numbers GSE75955 and GSE117595.

4.3.8 Statistical analysis

Generation of all graphs and statistical analyses (student t-test) were done using GraphPad Prism 8.0 version.

4.3.9 Heatmap analysis

Peak density was plotted in reference to the annotated TSS and CGI sites downloaded from the UCSC Genome Browser for galGal3. A sliding window of 5 kb upstream and 5 kb downstream of the reference centre point was used to display the peaks per 50-base bin by a R script.

4.4 Results and Discussion

4.4.1 Carbonic anhydrase (*CA2*) gene on macrochromosome 2

The *CA2* (carbonic anhydrase II) gene, on macrochromosome 2, is highly expressed in polychromatic erythrocytes (21, 24). The entire gene and flanking regions were enriched in the F1 chromatin fraction (Figure 4.1a), which contains chromatin fragments that were soluble in 150 mM NaCl. An intense FAIRE-seq peak was found at the promoter region of *CA2* and was positioned 5' to the transcript (+) reads. Anti-sense reads (transcript -) were seen on both sides of the FAIRE-seq peak. The FAIRE-seq peak aligned in the split in the H3K4me3 and H3K27ac signals (see red line in Figure 4.1a). Figure 4.1b shows an expanded view of the *CA2* gene region. The FAIRE peak located 5' to the *CA2* gene aligns with a break in the F1, H3K4me3 and H3K27ac peaks, which is consistent with a NDR in the *CA2* promoter region. This observation shows that the loss of reads in a F1-DNA seq cluster may correspond to a nucleosome-free regulatory region. However, several breaks (that is, loss of reads) in F1- DNA seq track did not align with a peak in FAIRE-seq (Figure 4.1b). The NDR also positioned with a CGI that was 1810 bp in length. Long CGIs with lengths >800 bp are found in promoters of chicken genes involved in development, regulation of transcription and regulation of biosynthetic processes (25).

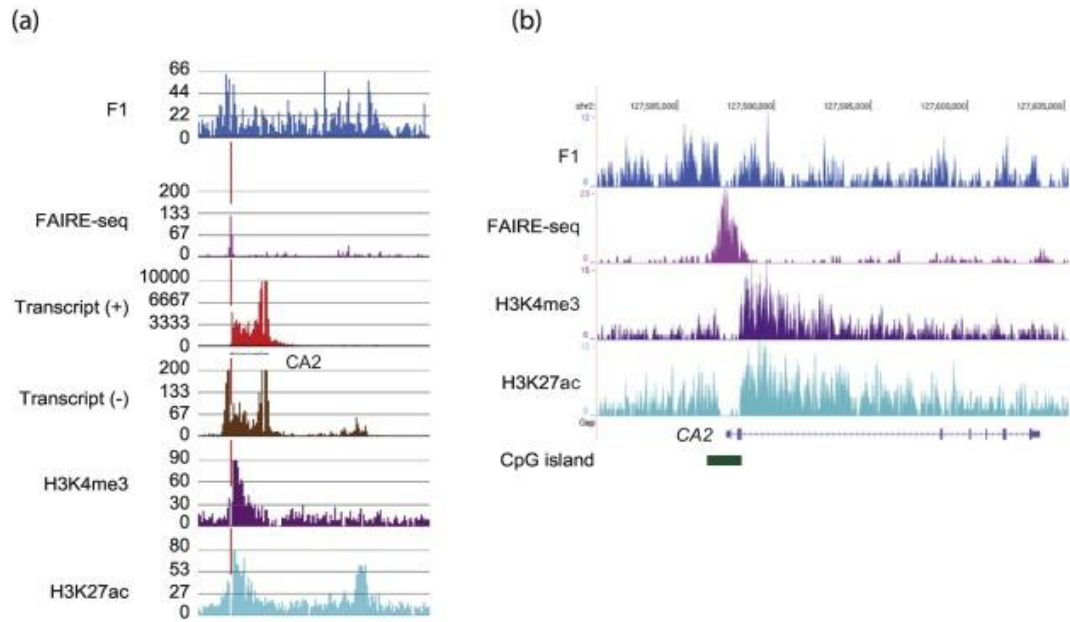


Figure 4.1: Chromatin profile and transcriptional activity of the *CA2* genomic region. (a) Signal tracks showing DNA enrichment in F1 fraction, FAIRE-seq track, transcripts on (+) and (-) strands and H3 modifications (H3K4me3 and H3K27ac). Transcripts (with exons as black boxes) are shown below their template strand. Vertical red lines illustrate the position within the genomic region of FAIRE peak relative to the drop in F1 DNA-seq reads, H3K27ac peak and H3K4me3 peak. The mapped BAM or WIG files were visualized using tools from the Partek Genomic Suite v6 (21)(b) An expanded view of the region containing the FAIRE peak is shown. The signal tracks for F1, FAIRE, H3K4me3 and H3K27ac were visualized using the Integrative Genomics Viewer. The position of the CGI is shown.

4.4.2 Ferritin, heavy polypeptide 1 (*FTH1*) locus on macrochromosome 5

The transcribed *FTH1* gene was in a 46 kb compartment, along with two other annotated genes, one of which (*RAB3IL1*) was expressed at a much lower level than *FTH1* (Figure 4.2a). The entire gene and flanking regions were enriched in the F1 chromatin fraction (Figure 4.2a). The 5' end of the *FTH1* gene was enriched in H3K27ac and H3K4me3, and antisense transcription (– strand) occurred along the coding and upstream regions of the gene. FAIRE-seq analyses showed a peak at the promoter region of the *FTH1* gene and next to the transcript (–) reads. Anti-sense transcript (+) reads were observed on both sides of the FAIRE peak. This peak aligned with a loss of reads in the F1 DNA-seq track (see red line in Figure 4.2a). This FAIRE peak was located next to the H3K4me3 peak. The FAIRE peak also separated two regions with H3K27ac. Figure 4.2b shows an enlargement of the *FTH1* gene region. The FAIRE peak located 5' to the *FTH1* gene aligns with a break in the F1, H3K4me3 and H3K27ac peaks, which is consistent with a NDR at the *FTH1* promoter. Also, in this region is a sequencing gap (a gap in the galGal3 assembly) which splits the FAIRE peak. Positioned with the FAIRE peak were two CGIs, 582 and 375 bp in length, located on either side of the sequencing gap. These results are consistent with a NDR that aligned with a break (trough) in the F1, H3K4me3 and H3K27ac peaks. To validate the FAIRE peaks found in the 5' regions of the *CA2* and *FTH1* genes, FAIRE-PCR was done with two biological repeats. Figure 4.2c shows the enrichment of the *CA2* promoter region and *FTH1* exon 1 sequences in the FAIRE fraction. An amplicon covering the *FTH1* exon 3 was not enriched in the FAIRE fraction. Thus, FAIRE-PCR results were consistent with the FAIRE-seq data.

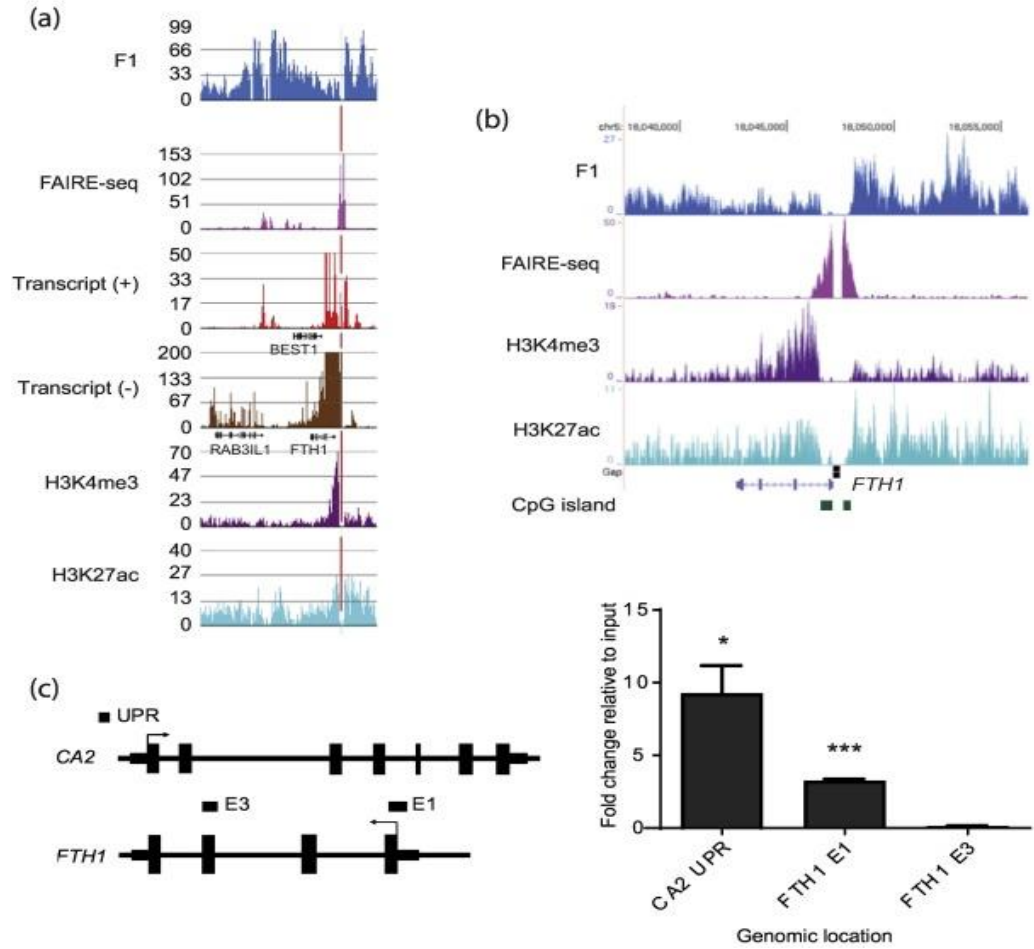


Figure 4.2: Chromatin profile and transcriptional activity of the *FTH1* genomic region. As described in Figure 4.1(a and b). (b) The black square below the H3K27ac ChIP-seq track shows the position of a gap in the *galGal3* assembly. (c) Map shows the *CA2* and *FTH1* genes and positions of the amplicons to analyze the enrichment of DNA regions in the FAIRE fractions shown in a bar graph. *CA2* PR, *CA2* promoter region; *FTH1* exon 1, *FTH1* E1; *FTH1* exon 3, *FTH1* E3. The error bars represent the standard deviation of the mean from the two biological replicates. Statistical significance was calculated by comparing *CA2* PR or *FTH1* E1 to the control (*FTH1* E3). The P values were calculated with the Student T-test (*<0.05, ***<0.001).

4.4.3 β -Globin locus on macrochromosome 1

The β -globin locus, as defined by the DNA-seq reads of the low-salt soluble F1 chromatin, co-mapped with the β -globin domain as defined by DNase I sensitivity and histone acetylation (21). Furthermore, within the domain, F1 enrichment reads paralleled the high acetylation profile and allowed the detection of DNase I hypersensitive sites identifying the β -globin LCR. Within the LCR we noted drops in the F1 reads. The location of the troughs mapped with the hypersensitive sites, suggesting that the troughs were NDRs likely associated with transcription factors (26, 27). Figure 4.3 shows that several pronounced NDR as detected by FAIRE-seq were located throughout the LCR at the location of HS4 and HS1. These FAIRE-seq peaks aligned with the valleys in the F1 DNA-seq reads. The HS4 site binds to several transcription factors including CTCF, GATA1 and NF-E2 and insulator activity (26, 28). There are several CGIs within the β -globin domain with a 352 bp long CGI positioning with HS4. A prominent FAIRE-seq peak and CGI (225 bp) may correspond to the promoter region of the adult β -globin (*HBG2*) gene. In general, the FAIRE-seq peaks aligned with the valleys in the F1 DNA-seq tracks; however, the intensity of the FAIRE-seq peaks did not correlate with the depth of the valley in the F1 DNA-seq profile.

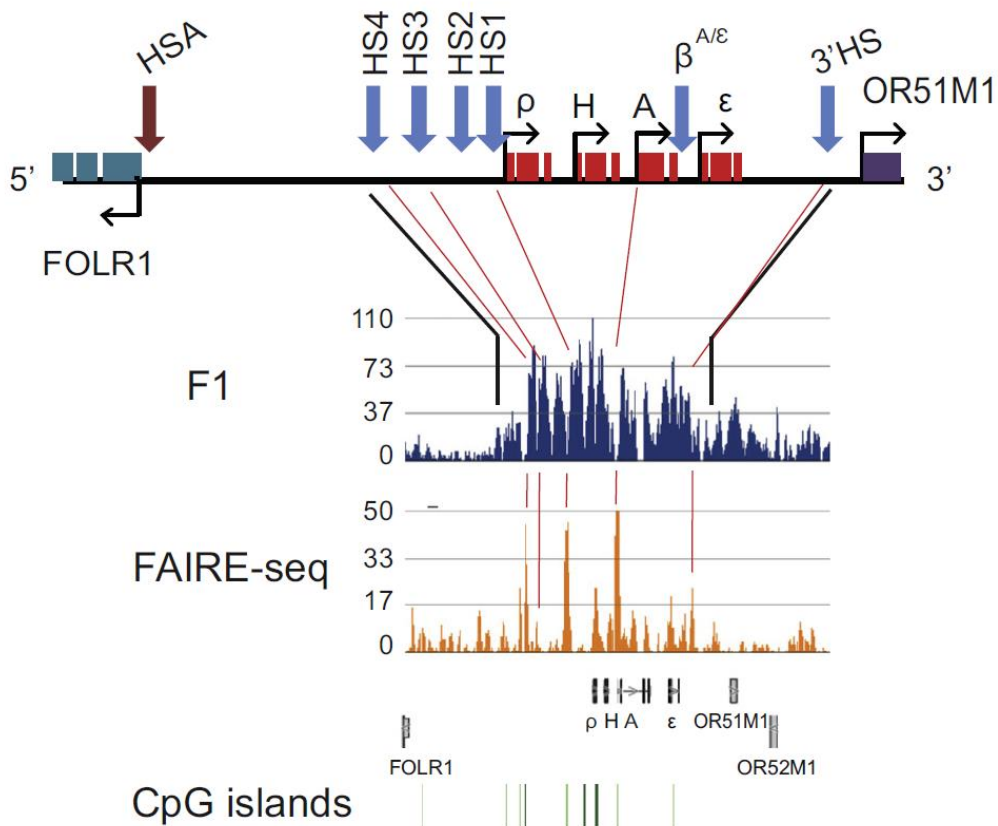


Figure 4.3. Chromatin profile of the β -globin genomic region. Schematic of the β -globin domain, detailing the developmentally regulated β -globin genes and DNase I hypersensitive sites, signal tracks showing DNA enrichment in F1 fraction and FAIRE-seq track. Vertical red lines illustrate the position within the genomic region of FAIRE peak relative to the drop in F1 DNA-seq reads. The position of the CGIs is shown.

4.4.4 Histone H5 (H1F0) locus on macrochromosome 1

The chromatin compartment shown in Figure 4.4a has the highly expressed intronless H5 gene as well as other genes that were either poorly expressed (*ANKRD54*, *GCAT*) or silent (*GALR3*, *NOGGIN4*) in chicken polychromatic erythrocytes (21). Expression of the H5 gene, coding for an H1 isoform only found in avian erythrocytes, is regulated by 5' and 3' enhancers (29) and in accordance, H3K27ac peaks coincided with these sites. Moreover, eRNA was evident in the antisense direction (+ strand). H5, a short gene (< 1 kb) without intron, was enriched in H3K4me3 and transcribed at a low levels in the antisense direction throughout the gene body.

As with the β -globin domain, the FAIRE-seq peaks aligned with the breaks in the F1 DNA-seq reads (see red lines in Figure 4.4a). The FAIRE-seq peaks (NDR) were located at the promoters and enhancers of expressed genes in this locus (Figure 4.4b). The FAIRE peaks often aligned with the CGIs. The FAIRE-seq peak aligned with the break in F1 DNA-seq reads located at the promoter region of the expressed *MICALL1* gene. Our interpretation of this result is that there is a NDR at the promoter region which is flanked on either side by acetylated chromatin that is soluble at physiological ionic strength. Following alignment of the FAIRE-seq track with that of the F1 DNA-seq track, we concluded that often the break in the F1 DNA-seq pattern aligned with a NDR; however, exceptions to this were observed.

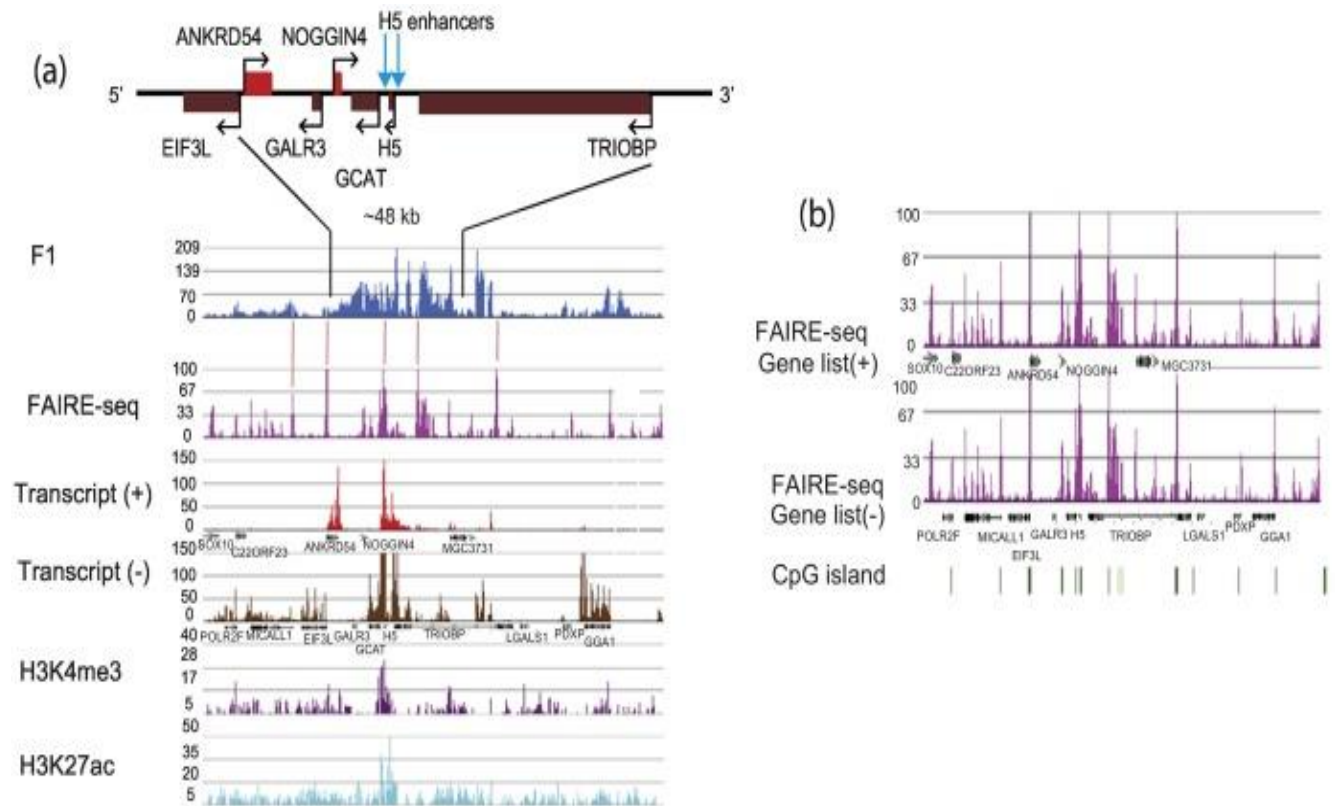


Figure 4.4: Chromatin profile and transcriptional activity of the *HIF0* (H5) genomic region. (a) Schematic of the genomic region containing the *HIF0* gene, signal tracks showing DNA enrichment in F1 fraction, FAIRE-seq track, transcripts on (+) and (-) strands and H3 modifications. Vertical red lines illustrate the position within the genomic region of FAIRE peak relative to the drop in F1 DNA-seq reads. Transcripts (with exons as black boxes) are shown below their template strand. (b) FAIRE-seq track aligned with transcripts across *HIF0* genomic region. The position of the CGIs is shown.

4.4.5 α -Globin (HBA) locus on microchromosome 14

Our data outlined a 60 kb stretch of chromatin, containing the α - globin genes, that was soluble at physiological ionic strength and coincided with a highly acetylated chromatin region in erythroid cells (Figure 4.5) (12). In contrast to the β -globin locus, the α -globin compartment, containing three genes, was not bordered by specific elements and partly overlaps the housekeeping gene *NPRL3* (10, 30). The α -globin locus includes a putative LCR (α MRE) and an insulator (HS-14.9 also known as α EHS-1.4), both located in the housekeeping gene *NPRL3* (also known as *CGTHBA* or *C16orf35*), as well as an enhancer (3'enh) 1 kb downstream from the α A gene (Figure 4.5) (31). These three regulatory regions displayed the H3K27ac mark and eRNA transcription on the (–) strand, although RNA originating from 3'enh was not certain, as it was masked by α A transcription (Figure 4.5). The FAIRE-seq peaks aligned with the α MRE and HS-14.9 but had a lower intensity than the NDR at β -globin HS4 site.

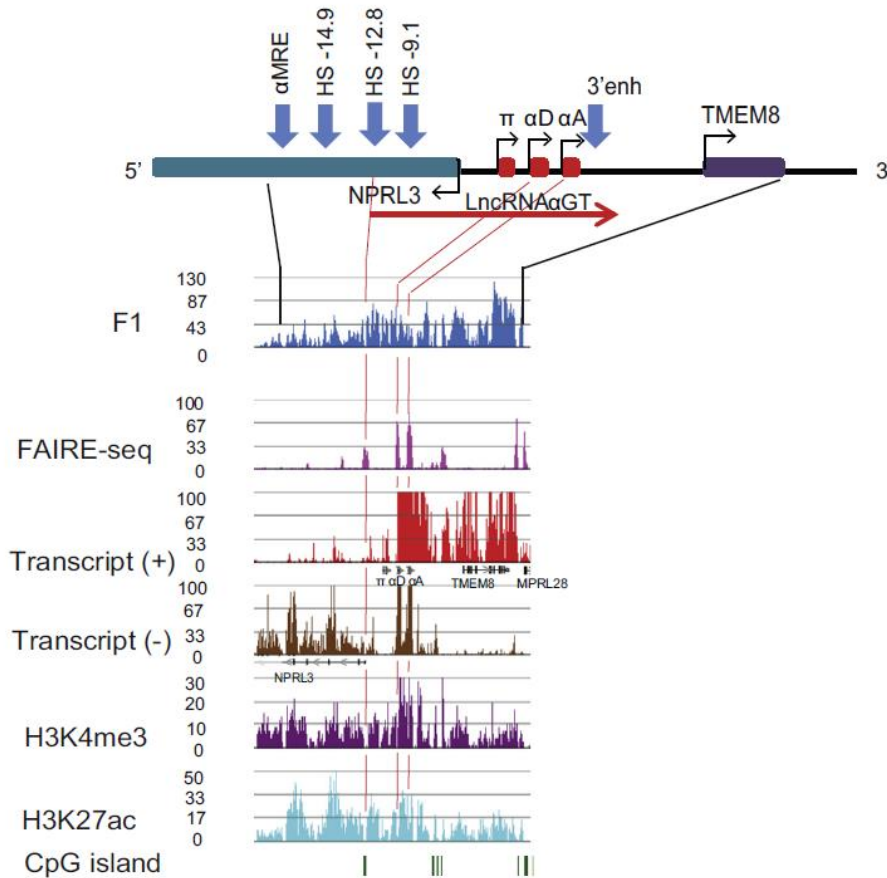


Figure 4.5: Chromatin profile and transcriptional activity of the α -globin. As described in Figure 4.4.

The main transcripts in the α -globin domain were from the adult αA and αD -globin genes, which also displayed low levels of antisense transcription through the gene bodies. Also evident were lower expression of the *TMEM8* and *NPRL3* genes. There might be a low level of the non-coding 23 kb RNA *LncRNA αGt* , an epigenetic participant in adult gene expression (Figure 4.5) (32). H3K27ac and H3K4me3 showed enrichment on the adult α -globin genes. FAIRE-seq peaks (NDR) were observed over the promoter regions of the *NPRL3*, and adult αA - and αD -globin genes as were CGIs.

4.4.6 *ARIH1* and *NCOA4* genes

ARIH1 (ariadne RBR E3 ubiquitin protein ligase 1) shared a 154 kb chromatin compartment on microchromosome 10, with two other genes expressed at a low level (*HEXA* and *PKM2*) (Figure 4.6a). The FAIRE peaks (NDR) were found between F1 DNA-seq peaks and at promoter regions of genes in this region. A FAIRE peak and two CGIs (629, 593 bp) were located at promoter region of the *ARIH1* gene and next to this FAIRE peak were intense H3K4me3 and H3K27ac peaks. Another FAIRE peak and CGI (692 bp) was located at the *BBS4* promoter, but this weakly expressed gene lacked pronounced H3K4me3 and H3K27ac peaks.

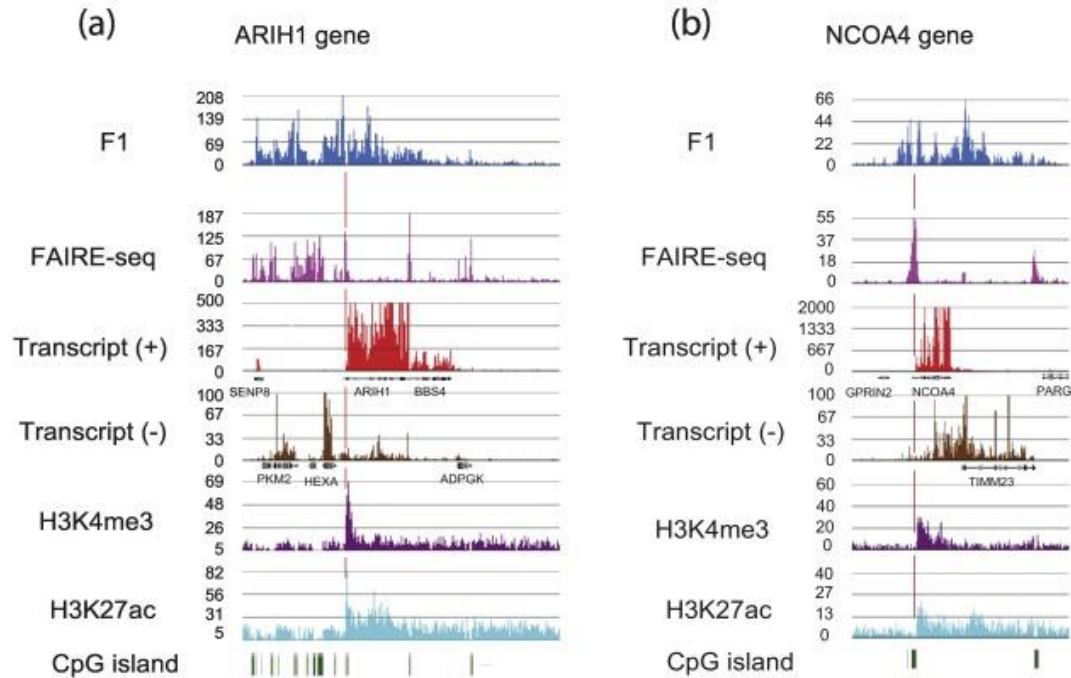


Figure 4.6: Chromatin profile and transcriptional activity of the *ARIH1* and *NCOA4* genomic regions. As described in Figure 4.4. (a) *ARIH1*, (b) *NCOA4*. The position of the CGIs is shown.

For the locus containing the *NCOA4* (nuclear receptor coactivator 4) on macrochromosome 6 (Figure 4.6b) gene, the strongest FAIRE-seq peak was located at a region lacking F1 DNA-seq reads and aligned with the *NCOA4* promoter region and a CGI (1114 bp). Pronounced H3K4me3 and H3K27ac peaks were found next to this FAIRE-seq peak. Among the less intense FAIRE-seq peaks, one maps to the promoter region of *TIMM23* gene and a CGI (952 bp).

4.4.7 Open chromatin and modified nucleosomes

Analyses of several transcribed genes showed that H3K4me3 and H3K27ac peaks were highest at the boundary of the NDR and then gradually losing intensity further away from the nucleosome region. It is possible that the transcription factor(s) binding to the NDR recruit lysine acetyltransferases and lysine methyltransferases to modify the nearby nucleosomes. Assuming that the enzymes are “fixed” to the transcription factor binding site, we determined what is the average maximal reach of the enzymes. From analyses of ten genes, we calculated a maximal distance of 2300 bp for H3K4me3 and 1930 for H3K27ac. Using a nucleosome repeat length of 212 bp for chicken erythrocyte, we calculate that the enzymes could contact 9 to 10 nucleosomes.

4.4.8 FAIRE and CpG Islands (CGIs)

To understand the distribution of open chromatin regions, the FAIRE peaks were plotted over the annotated TSS and CGIs (Figure 4.7) and were seen to overlap, showing a strong association of the FAIRE peaks with chicken promoters. A similar correlation of FAIRE peaks over TSS was reported for human promoters (23, 33). We observed several FAIRE peaks (NDR) that did not map to a promoter of an annotated gene. Such FAIRE peaks present in intergenic, intragenic and distal regions are known as orphan peaks (33, 34). These orphan peaks may identify regulatory regions to genes not yet annotated or may represent regions of the genome involved in higher order organization of the chicken erythrocyte genome(34, 35). Further characterization of

the orphan FAIRE peaks will add to our knowledge of the chromatin structure and regulation of gene expression of the chicken erythrocyte.

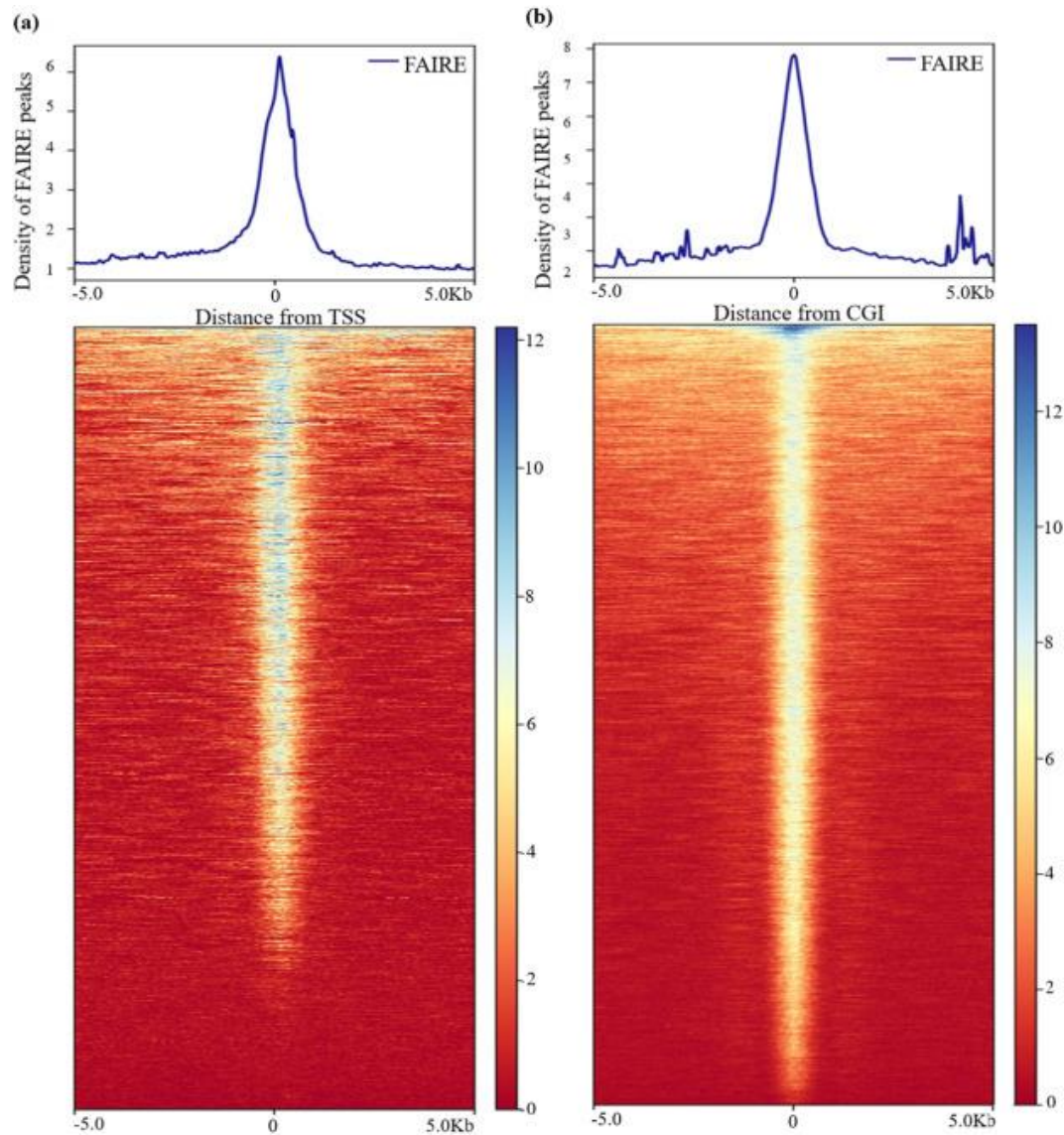


Figure 4.7: Heatmaps depicting density coverage of FAIRE peaks over the TSS (a) and CGI (b) spanning 5 kb on each side of TSS or CGI from galGal3 database of UCSC Genome Browser.

4.5 Conclusion

One to two percent of the chicken polychromatic erythrocyte epigenome is organized into DNase I sensitive, dynamically acetylated chromatin regions (compartment A). These regions are soluble at physiological ionic strength as shown in our F1-sequencing results. Breaks in the F1 DNA-seq peaks often aligned with NDR located at the regulatory (promoter, LCR) regions of genes. Thus, in combination with FAIRE-seq to map the NDR, we further appreciated the information in the F1-seq data and the breadth of the compartment. Further, our studies clearly demonstrated that FAIRE-seq readily identified promoter regions of transcribed genes, and such information will be useful in white Leghorn chicken genome annotation to map expressed genes.

Repressed chromatin (compartment B) is not soluble at physiological ionic strength, suggesting that solubility of the chromatin is an important factor in the phase separation of the A/B compartments. The solubility at physiological ionic strength is dependent on the highly acetylated state of the core histones which prevent H1/H5-mediated chromatin condensation (18). Incubating polychromatic erythrocytes in the absence of an histone deacetylase inhibitor, results in rapid deacetylation of the core histones bound to transcriptionally active chromatin, rendering the chromatin insoluble at physiological ionic strength (18).

Despite 310 million years of separate evolution, there are long blocks of conserved synteny between the chicken and human genomes. Regarding the chromosomal organization of genes, the

human genome is closer to the chicken than to rodents (36). Our studies may provide insights into the human erythrocyte genome organization.

4.6 Conflict of interest

Authors declare no conflicts of interest.

4.7 Funding

This work was supported by a grant from the Natural Sciences and Engineering Research Council of Canada (RGPIN-2017-05927) to J.R.D.

4.8 References

- [1] Serizay, J., and Ahringer, J. (2018) Genome organization at different scales: nature, formation and function. *Current Opinion in Cell Biology* 52, 145–153.
- [2] Lieberman-Aiden, E., Van Berkum, N. L., Williams, L., Imakaev, M., Ragozy, T., et al. (2009) Comprehensive mapping of long-range interactions reveals folding principles of the human genome. *Science* 326, 289–293.
- [3] Fortin, J. P., and Hansen, K. D. (2015) Reconstructing A/B compartments as revealed by Hi-C using long-range correlations in epigenetic data. *Genome Biology* 16, 180.
- [4] Gorkin, D. U., Leung, D., and Ren, B. (2014) The 3D genome in transcriptional regulation and pluripotency. *Cell Stem Cell* 14, 762–775.
- [5] Fishman, V., Battulin, N., Nuriddinov, M., Maslova, A., Zlotina, A., et al. (2019) 3D organization of chicken genome demonstrates evolutionary conservation of topologically associated domains and highlights unique architecture of erythrocytes' chromatin. *Nucleic Acids Research* 47, 648–665.
- [6] Stalder, J., Larsen, A., Engel, J. D., Dolan, M., Groudine, M., et al. (1980) Tissue-specific DNA cleavages in the globin chromatin domain introduced by DNAase I. *Cell* 20, 451–460.
- [7] Zhang, D. E., and Nelson, D. A. (1988) Histone acetylation in chicken erythrocytes. Rates of acetylation and evidence that histones in both active and potentially active chromatin are rapidly modified. *The Biochemical journal* 250, 233–240.
- [8] Hebbes, T. R., Clayton, A. L., Thorne, A. W., and Crane-Robinson, C. (1994) Core histone

- hyperacetylation co-maps with generalized DNase I sensitivity in the chicken β -globin chromosomal domain. *EMBO Journal* 13, 1823–1830.
- [9] Ghirlando, R., Giles, K., Gowher, H., Xiao, T., Xu, Z., et al. (2012) Chromatin domains, insulators, and the regulation of gene expression. *Biochimica et Biophysica Acta - Gene Regulatory Mechanisms* 1819, 644–651.
- [10] Furlan-Magaril, M., Rebollar, E., Guerrero, G., Fernández, A., Moltó, E., et al. (2011) An insulator embedded in the chicken α -globin locus regulates chromatin domain configuration and differential gene expression. *Nucleic Acids Research* 39, 89–103.
- [11] Hebbes, T. R., Thorne, A. W., and Crane-Robinson, C. (1988) A direct link between core histone acetylation and transcriptionally active chromatin. *The EMBO Journal* 7, 1395–1402.
- [12] Anguita, E., Johnson, C. A., Wood, W. G., Turner, B. M., and Higgs, D. R. (2001) Identification of a conserved erythroid specific domain of histone acetylation across the α -globin gene cluster. *Proceedings of the National Academy of Sciences of the United States of America* 98, 12114–12119.
- [13] Litt, M. D., Simpson, M., Lix Recillas-Targa, F. Â., Prioleau, M.-N. È. L., and Felsenfeld, G. (2001) Transitions in histone acetylation reveal boundaries of three separately regulated neighboring loci. *The EMBO Journal* 20, 2224–2235.
- [14] Razin, S. V., Ulianov, S. V., Ioudinkova, E. S., Gushchanskaya, E. S., Gavrilov, A. A., et al. (2012) Domains of α - and β -globin genes in the context of the structural-functional organization of the eukaryotic genome. *Biochemistry (mosc)* 77, 1409–1423.
- [15] Delcuve, G. P., and Davie, J. R. (1989) Chromatin structure of erythroid-specific genes of

- immature and mature chicken erythrocytes. *Biochemical Journal* 263, 179–186.
- [16] Henikoff, S., Henikoff, J. G., Sakai, A., Loeb, G. B., and Ahmad, K. (2009) Genome-wide profiling of salt fractions maps physical properties of chromatin. *Genome Research* 19, 460–469.
- [17] Bates, D. L., and Thomas, J. O. (1981) H1 and H5: one or two molecules per nucleosome? *Nucleic Acids Research* 9, 5883–5894.
- [18] Ridsdale, J. A., Hendzel, M. J., Delcuve, G. P., and Davie, J. R. (1990) Histone acetylation alters the capacity of the H1 histones to condense transcriptionally active/competent chromatin. *Journal of Biological Chemistry* 265, 5150–5156.
- [19] Locklear, L., Ridsdale, J. A., Bazett-Jones, D. P., and Davie, J. R. (1990) Ultrastructure of transcriptionally competent chromatin.
- [20] Sun, J.-M., Chen, H. Y., Espino, P. S., and Davie, J. R. (2007) Phosphorylated serine 28 of histone H3 is associated with destabilized nucleosomes in transcribed chromatin. *Nucleic Acids Research* 35, 6640–6647.
- [21] Jahan, S., Xu, W., He, S., Gonzalez, C., Delcuve, G. P., et al. (2016) The chicken erythrocyte epigenome. *Epigenetics and Chromatin* 9, 1–11.
- [22] Ridsdale, J. A., Rattner, J. B., and Davie, J. R. (1988) Erythroid-specific gene chromatin has an altered association with linker histones. *Nucleic Acids Research* 16, 5915–5926.
- [23] Simon, J. M., Giresi, P. G., Davis, I. J., and Lieb, J. D. (2012) Using formaldehyde-assisted isolation of regulatory elements (FAIRE) to isolate active regulatory DNA. *Nature Protocols* 7, 256–267.
- [24] Yoshihara, C. M., Lee, J.-D., and Dodgson, J. B. (1987) The chicken carbonic anhydrase II

gene: evidence for a recent shift in intron position.

- [25] Abe, H., and Gemmell, N. J. (2014) Abundance, arrangement, and function of sequence motifs in the chicken promoters. *BMC Genomics* 15, 1–12.
- [26] Stamatoyannopoulos, J. A., Goodwin, A., Joyce, T., and Lowrey, C. H. (1995) NF-E2 and GATA binding motifs are required for the formation of DNase I hypersensitive site 4 of the human beta-globin locus control region. *The EMBO journal* 14, 106–16.
- [27] Boyes, J., and Felsenfeld, G. (1996) Tissue-specific factors additively increase the probability of the all-or-none formation of a hypersensitive site. *The EMBO Journal* 15, 2496–2507.
- [28] Yusufzai, T. M., and Felsenfeld, G. (2004) The 5'-HS4 chicken β -globin insulator is a CTCF-dependent nuclear matrix-associated element. *Proceedings of the National Academy of Sciences of the United States of America* 101, 8620–8624.
- [29] Rousseau, S., Asselin, M., Renaud, J., and Ruiz-Carrillo, A. (1993) Transcription of the histone H5 gene is regulated by three differentiation-specific enhancers. *Molecular and Cellular Biology* 13, 4904–4917.
- [30] Klochkov, D., Rincón-Arano, H., Ioudinkova, E. S., Valadez-Graham, V., Gavrillov, A., et al. (2006) A CTCF-Dependent Silencer Located in the Differentially Methylated Area May Regulate Expression of a Housekeeping Gene Overlapping a Tissue-Specific Gene Domain. *Molecular and Cellular Biology* 26, 1589–1597.
- [31] Gavrillov, A. A., and Razin, S. V. (2008) Study of spatial organization of chicken α -globin gene domain by 3c technique. *Biochemistry (Moscow)* 73, 1192–1199.
- [32] Arriaga-Canon, C., Fonseca-Guzmán, Y., Valdes-Quezada, C., Arzate-Mejía, R., Guerrero,

- G., et al. (2014) A long non-coding RNA promotes full activation of adult gene expression in the chicken α -globin domain. *Epigenetics* 9, 173–181.
- [33] Giresi, P. G., Kim, J., McDaniell, R. M., Iyer, V. R., and Lieb, J. D. (2007) FAIRE (Formaldehyde-Assisted Isolation of Regulatory Elements) isolates active regulatory elements from human chromatin. *Genome Research* 17, 877–885.
- [34] Sarda, S., and Hannenhalli, S. (2018) Orphan CpG islands as alternative promoters. *Transcription* 9, 171–176.
- [35] Gushchanskaya, E. S., Artemov, A. V., Ulyanov, S. V., Logacheva, M. D., Penin, A. A., et al. (2014) The clustering of CpG islands may constitute an important determinant of the 3D organization of interphase chromosomes. *Epigenetics* 9, 951–963.
- [36] Hillier, L. W., Miller, W., Birney, E., Warren, W., Hardison, R. C., et al. (2004) Sequence and comparative analysis of the chicken genome provide unique perspectives on vertebrate evolution. *Nature* 432, 695–716.

4.9 Supplementary Data

Supplementary Table 4.1. Primers used in FAIRE-qPCR experiments

| Gene | Region | Primer sequence |
|---|---------------|--|
| Chicken <i>FTH1</i> | Exon 1 | FP: CCACCGCATCTCTCTCTTTC RP: GCGTACAGCTCCAGGTTGAT |
| | Exon 3 | FP: TCGTGATGACTGGGAGAATG RP: TGCCAATTTGTGCAGCTCTA |
| Chicken <i>CA2</i> | PR | FP: CGCGTTTCCTACAAGGTGAG RP: GACGCCCTGGTTCTTACTT |
| PR: promoter region; FP: Forward primer; RP: Reverse primer; 5' to 3' shown | | |

Chapter-5

Atypical chromatin structure of immune-related genes expressed in chicken erythrocytes

Atypical chromatin structure of immune-related genes expressed in chicken erythrocytes

Sanzida Jahan, **Tasnim H. Beacon**, Wayne Xu, James R. Davie

Source of the published article:

Biochemistry and Cell Biology, 2020, 98(2): 171-177

<https://doi-org.uml.idm.oclc.org/10.1139/bcb-2019-0107>

Author contribution statement

S.J. carried out F1 DNA-seq, RNA-seq for chicken polychromatic erythrocytes and 6C2, RT-PCR and induction study, performed bioinformatics analyses, designed experiments, wrote the first draft of the manuscript, analyzed the data, and prepared figures; **T.H.B.** performed the FAIRE sequencing and assays, data analyses, bioinformatic analyses; and prepared figures; W.X. performed bioinformatics analyses; J.R.D. conceived the study, participated in its design and coordination, and reviewed and wrote the final version of the manuscript. All of the authors read and approved the final manuscript.

5.1 Abstract

The major biological role of red blood cells is to carry oxygen to the tissues in the body. However, another role of the erythroid cell is to participate in the immune response. Mature erythrocytes from chickens express Toll-like receptors and several cytokines in response to stimulation of the immune system. We previously reported the application of a biochemical fractionation protocol to isolate highly enriched transcribed DNA from polychromatic erythrocytes from chickens. In conjunction with next-generation DNA, RNA sequencing, chromatin immunoprecipitation-DNA sequencing, and FAIRE sequencing, we identified the active chromosomal compartments and determined their structural signatures in relation to expression levels. Here, we present the detailed chromatin characteristics of erythroid genes participating in the innate immune response. Our studies revealed an atypical chromatin structure for several genes coding for Toll-like receptors, interleukins, and interferon regulatory factors. The body of these genes had NDR intermingled with nucleosomes modified with H3K4me3 and H3K27ac, suggesting a dynamic unstable chromatin structure. We further show that human genes involved in cell identity have gene bodies with the same chromatin-instability features as the chicken polychromatic erythrocyte genes participating in the innate immune response.

Keywords: chicken erythrocyte, innate immune response, NDR, active histone marks, salt soluble chromatin

5.2 Introduction

The major function of erythrocytes is to carry oxygen through the circulation from lungs to tissues of the body. However, nucleated erythrocytes of salmon, trout, and chicken, and human nucleated erythroid cells also participate in the immune response (1–5). Non-mammalian vertebrates such as birds, fish, amphibians, and reptiles contain a nucleus at the mature erythrocyte stage. Human erythroblasts are also nucleated but become enucleated at the erythrocyte terminal stage.

Genes involved in innate immunity are induced by TLR ligands in mature erythrocytes from chickens (2, 4, 6). TLRs, which are type I transmembrane proteins, serve as the major component of pattern recognition receptors (4). The TLR-mediated signaling event leads to release of cytokines and interferon molecules from the infected cells (7). Stimulation of mature chicken erythrocytes with polyinosinic–polycytidylic acid (poly I:C), which mimics viral dsRNA, reportedly induces TLR3 and IFN α transcripts (2). Findings from these studies provided evidence that chicken erythrocytes respond to various ligand-mediated immune responses, and thus, contribute to defense against invading microorganisms (6). However, the chromatin structure of genes participating in the immune response has not been characterized.

An analysis of the chicken gene promoter structure revealed that chicken promoter sequences varied in their GC contents and the presence of a long CpG island (CGI; >800 bp)(8). Genes with promoters with long CGIs were associated with morphogenesis, development, transcription processes, and biosynthetic processes, while genes involved in defense response (response to bacteria and leukocyte activation) lacked long CGIs. This study showed that the structural features of promoters of genes involved in the immune response differ from those of genes involved in the gas exchange function.

Our studies use chicken polychromatic erythrocytes, which are isolated from anemic birds(9), to characterize the structure and organization of transcriptionally active chromatin(10–13). Polychromatic erythrocytes from chickens are transcriptionally active nucleated cells that are arrested in the G0 phase of the cell cycle. We have taken advantage of the features of polychromatic erythrocyte chromatin to develop a fractionation protocol that isolates transcriptionally active chromatin (chromatin fraction F1) (12, 14) (Supplementary data, Figure 5.S1). Transcriptionally active chromatin domains with acetylated and methylated histones (H3K4me3; an active mark) that are soluble in 150 mmol/L NaCl were identified by next-generation sequencing of the F1 chromatin fraction (F1 DNA-seq). Together with transcriptome analyses by RNA-seq, H3K4me3 and H3K27ac ChIP-seq, FAIRE-seq, and CGI profiling, we have characterized the features of transcribed genes involved in oxygen transport activity and gene regulation (12, 13). In this study, we present the atypical chromatin structure of genes involved in the innate immune signaling system in polychromatic erythrocytes from chickens.

5.3 Materials and methods

5.3.1 Ethics statement

Ethical approval was obtained from the University of Manitoba Animal Care Committee. All methods involving the use of chickens were approved by the committee and carried out in accordance with its guidelines and regulations. The University is currently in full compliance with the Canadian Council on Animal Care who has certified that the animal care and use program at the University of Manitoba is in accordance with the standards of Good Animal Practice. The birds were purchased through Central Animal Care Services, University of Manitoba, and were housed under standard conditions.

5.3.2 Isolation of chicken erythrocytes and treatments

Polychromatic erythrocytes were isolated from anemic female Adult White Leghorn chickens as described previously (12, 14). Briefly, the erythrocytes were separated from white blood cells by centrifugation, followed by several washes of the erythrocyte pellet (15). Microscopic examination of the erythrocyte preparation confirmed the lack of white blood cells in the preparation. Polychromatic erythrocytes were treated with 50 µg/mL of poly I:C for either 1, 3, 6, 12, or 24 h in MEM Alpha (1×) (Life Technologies, cat#12571-063).

5.3.3 RNA extraction and RT-PCR

RNA from polychromatic erythrocytes was isolated using RNeasy Plus mini kit (QIAGEN) following the manufacturer's instructions. DNase (Promega) digestion was performed to remove any genomic DNA in the purified RNA. For qPCR analysis, complementary DNA (cDNA) was generated from purified total RNA (400 ng) using M-MLV reverse transcriptase and Oligo dT primers (Invitrogen). Quantitative real-time PCR was performed with cDNA (2.0 ng) using SYBR Green real-time PCR on the iCycler IQ5 (BioRad). Primers for the regions analyzed are described in the Supplementary data, Table 5. S1.

5.3.4 ChIP-sequencing

ChIP-seq was performed as previously described (12). Briefly, the nuclei were lysed after cross-linking of cells with 0.5% formaldehyde for 10 min, and the chromatin was sheared to 250 bp using an ultrasonic dismembrator (Fisher). The ChIP assays were performed with anti-H3K4me3 (Abcam, ab8589) and anti-H3K27ac (Abcam, ab4729) antibodies. Isotype-specific nonrelated IgG was used as a negative control for each ChIP assay. ChIP and input DNA was further processed, purified and quantitated using Qubit® 2.0 fluorometer (Life Technologies). Input and ChIP DNA quality was analyzed using 2000 Bioanalyzer (Agilent). ChIP-seq was performed in two biological replicates using the same protocol.

5.3.5 Chromatin fractionation, RNA sequencing, and FAIRE sequencing

Chromatin fractionation, isolation and sequencing of chromatin fraction F1, RNA sequencing, and FAIRE sequencing were done as described in Jahan et al. 2016, 2019 (12, 13). FAIRE was done with four biological repeats, either by sequencing or PCR analyses. Primers for the regions analyzed are described in the Supplementary data, Table 5. S2.

5.3.6 Bioinformatics data analyses

The bioinformatics data analyses were conducted as previously described (12). Briefly, the SOLiD sequencing data of F1 DNA-seq, RNA-seq, and ChIP-seq were mapped on chicken reference genome galGal3 using Lifescope v2.5.1 software (Life Technologies) with 2-mismatch settings. FAIRE-seq was done using an Illumina MiSeq. The genes were annotated using UCSC RefSeq chicken genes. The sequencing data are available from the GEO; <https://www.ncbi.nlm.nih.gov/submit/ncbi/geo/> under the accession number GSE141015. For active chromatin detection and genomic distribution, we applied a clustering approach (SICER)(16) to identify islands of DNA-seq enrichment using F1 DNA-seq-mapped BAM files as inputs. The window and gap sizes were chosen to be 1 kb each. The SE DNA-seq data were used for background subtraction. The binding site candidates were identified by comparing the ChIP sample with the input sample using the model-based analysis of ChIP-seq (MACS) peak caller (17) with a *P* value $1e-5$ and a high limit *mfold* cutoff of 30, and a low limit of 10. Transcriptional levels were detected

using the LifeScope whole-transcriptome mapping module. We used the reads per kilobase per million (RPKM) to assign gene transcription levels. Browser views of gene tracks and ChIP-seq data and peaks were visualized using IGV after data normalization using igvtools (18) or Partek Genomics Suite v6.6.

5.3.7 Statistical analysis

All graphs and statistical analyses were performed using GraphPad Prism (version 8.0). Data presented are the mean \pm SEM. Statistical analyses were performed using 1-way ANOVA, followed by Dunnett's multiple comparisons test where appropriate differences were considered significant for values of $P \leq 0.05$. For analyses of FAIRE-PCR, statistical significance was calculated by comparing the FAIRE DNA to the input DNA. The P values were calculated using Student's t test.

5.4 Results

5.4.1 Polychromatic erythrocytes from chickens express genes involved in the immune response

Unstimulated mature erythrocytes from chickens express TLR, chemokine, and *IFNA3* genes (2, 4). We searched our previously published RNA-seq data to determine the steady state expression level of genes involved in the immune response in polychromatic erythrocytes from chickens

(12). Genes coding for TLRs, interleukins, chemokines, and interferon regulatory factors were expressed to differing extents in these erythroid cells (Table 5.1; Supplementary data, Table 5.S3). Thus, as with mature erythrocytes, polychromatic erythrocytes express genes involved in the immune response.

Table 5.1. Genes participating in the immune response and expressed in chicken polychromatic erythrocytes.

Table 5.1. Genes participating in the immune response and expressed in chicken polychromatic erythrocytes.

| Toll-like receptors | | Adaptor proteins | | Interleukins | | Chemokines | |
|---------------------|-----|-------------------------------|------|--------------|------|------------|----|
| TLR2 | + | MYD88 | ++ | IL1B | + | CCL4 | + |
| TLR3 | +++ | TIRAP | ++++ | IL2 | + | CCL20 | ++ |
| TLR4 | + | TRAM1 | ++++ | IL3 | + | CXCL12 | ++ |
| TLR5 | + | TICAM1 | ++ | IL5 | + | CXCL14 | + |
| TLR6 | ++ | | | IL7 | ++ | | |
| TLR7 | ++ | | | IL8 | + | | |
| TLR15 | + | | | IL9 | ++ | | |
| TLR21 | + | | | IL15 | +++ | | |
| | | | | IL16 | + | | |
| | | | | IL18 | + | | |
| | | | | IL22 | + | | |
| Interferons | | Interferon regulatory factors | | Others | | | |
| Type I: | | IRF1 | ++++ | NFKB2 | ++ | | |
| IFNA3 | + | IRF2 | ++++ | IRAK2 | ++ | | |
| IFNW1 | ++ | IRF4 | + | CIAPIN1 | ++++ | | |
| | | IRF7 | ++++ | | | | |
| Type II: | | IRF8 | ++++ | | | | |
| IFNG | + | IRF9 | +++ | | | | |

Note: Steady-state expression levels (in uninduced chicken polychromatic erythrocytes): +, <51 RPKM; ++, 1-5 RPKM; +++, 5-50 RPKM; +++++, >50 RPKM; MYD88, myeloid differentiation primary response 88; TIRAP, TIR domain containing adaptor protein; TRAM1, translocation associated membrane protein 1-like 1; TICAM1, Toll like receptor adaptor molecule 1; CCL4 and 20, C-C motif chemokine ligand 4 and 20; CXCL12 and 14, C-X-C motif chemokine ligand 12 and 14; IFNA3, interferon; IFNW1, interferon omega 1; IFNG, interferon gamma; NFKB2, nuclear factor kappa B subunit 2; IRAK2, interleukin 1 receptor associated kinase 2; CIAPIN1, cytokine induced apoptosis inhibitor 1.

Chicken chromosomes are organized as macrochromosomes (chromosomes 1-8) and microchromosomes (9-38 and W). The expressed genes involved in the immune response were located on both macro- and microchromosomes (data not shown).

To determine whether the genes involved in the immune response were inducible in polychromatic erythrocytes, we incubated the cells with poly I:C, which is a synthetic analog for a double-stranded RNA virus and stimulates innate immunity via the TLR3-mediated pathway. In agreement with the studies using mature erythrocytes from chickens, *TLR3*, *IRF1*, *IL15*, *TLR21*, *CIAPIN1* and *TIRAP* were induced with poly I:C (Supplementary data, Figure 5.S2).

5.4.2 Genomic features of genes involved in the immune response

Common chromatin features of chicken genes expressed in polychromatic erythrocytes included a CGI, a H3K4me3, a H3K27ac, and a NDR over the promoter region (12, 13). Further, the promoter region of the gene was enriched in the F1 chromatin fraction (Supplementary data, Figure 5.S1), which contains chromatin fragments that are soluble at physiological ionic strength. Figure 5.1 shows the *IRF8* gene, which had a CGI over the promoter region and high steady state transcript levels. H3K27ac, H3K4me3, and a NDR as detected by FAIRE-seq were present at the upstream promoter region. The upstream promoter region and 5' end of the gene were enriched in the F1 salt-soluble chromatin fraction. Several other genes involved in the immune response including *IRF1*, *CAPIN1*, *TIRAP*, *TLR3*, *MYD88*, *IRAK2*, and *IL15* had most of these chromatin features (Figure 5.1; Supplementary data, Figures 5.S3–5.S5). These genes had moderate-to-high steady state levels of transcripts (Table 5.1).

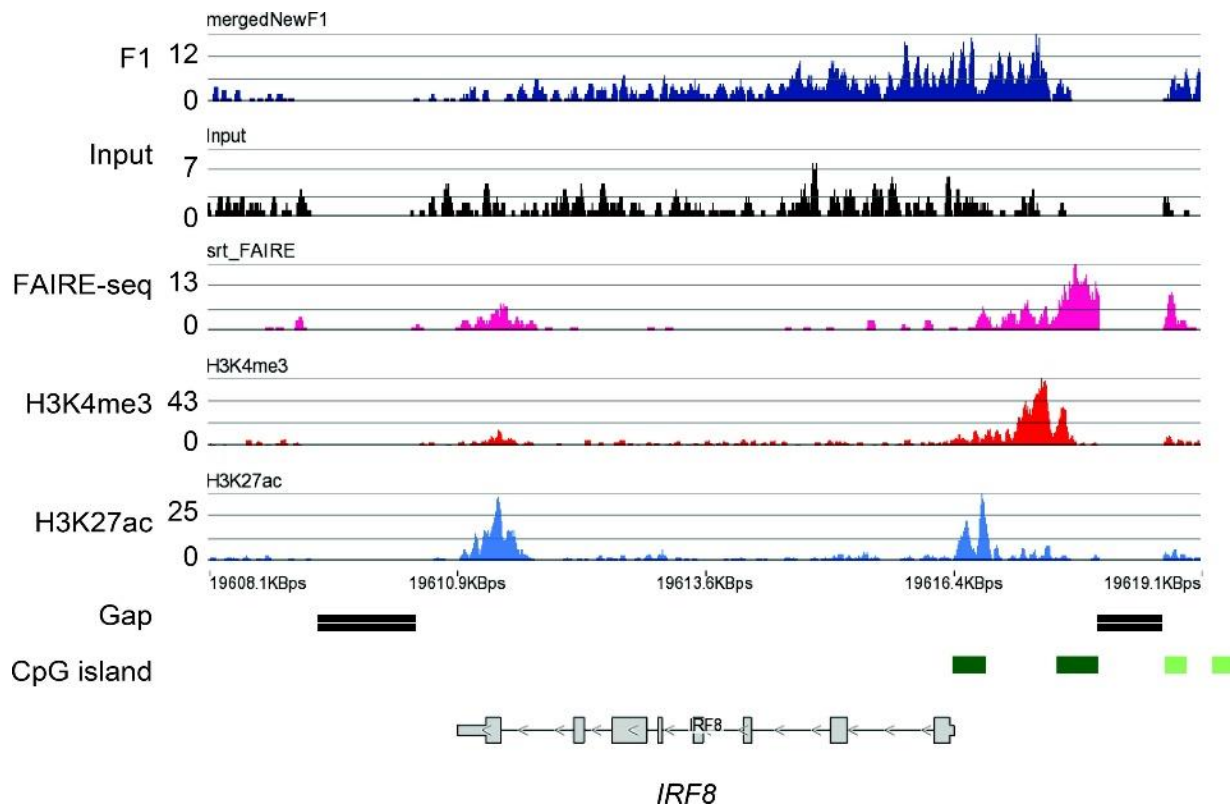


Figure 5.1: Chromatin profile of the *IRF8* gene region. The signal tracks for F1, FAIRE, H3K4me3, H3K27ac, and input DNA for the ChIP-seq studies are shown. The positions of the CpG islands and sequencing gaps are shown.

5.4.3 Atypical chromatin organization of genes involved in the immune response

In contrast to the typical chromatin organization of transcribed genes, several genes (*IRF7*, *NFKB2*, *TLR21*, *IL1B*, *IFNA3*) involved in the immune response had an unusual chromatin structure, with H3K4me3, H3K27ac, and nucleosome-free features throughout the gene body (Figure 5.2 and 5.3; Supplementary data, Figures 5.S6 and 5.S7). Some of these genes had a CGI

at the promoter (*IRF7*), and others did not (*TLR21*). The steady state level of the transcripts from these genes varied from high to low. The *NFKB2* gene, which had moderate transcript levels, had CGIs over the gene body (Figure 5.2). There was a sequencing gap at the promoter region, preventing further analyses of the promoter region. However, there was H3K27ac, H3K4me3, and a NDR at the beginning of the gene. There was little enrichment of the gene and flanking regions in the F1 chromatin fraction. Like the *IRF7* gene, the gene body was extensively associated with H3K27ac, H3K4me3, and the NDR that continued downstream of the gene. The *TLR21* gene had very low transcript levels (Table 5.1). There was a CGI positioned well-upstream of the gene and another CGI at the 3' end of the gene (Figure 5.3A). H3K4me3, H3K27ac, and the nucleosome-free chromatin regions were over the entire gene body. These features were not found at the promoter region of the gene. Little enrichment of the gene and flanking regions was observed in chromatin fraction F1. To validate the FAIRE peaks, FAIRE-PCR was done with three biological repeats. Figure 5.3B shows the enrichment of the *TLR21* exon 2, but not intron 1, in the FAIRE fraction, consistent with what was observed in the FAIRE-seq.

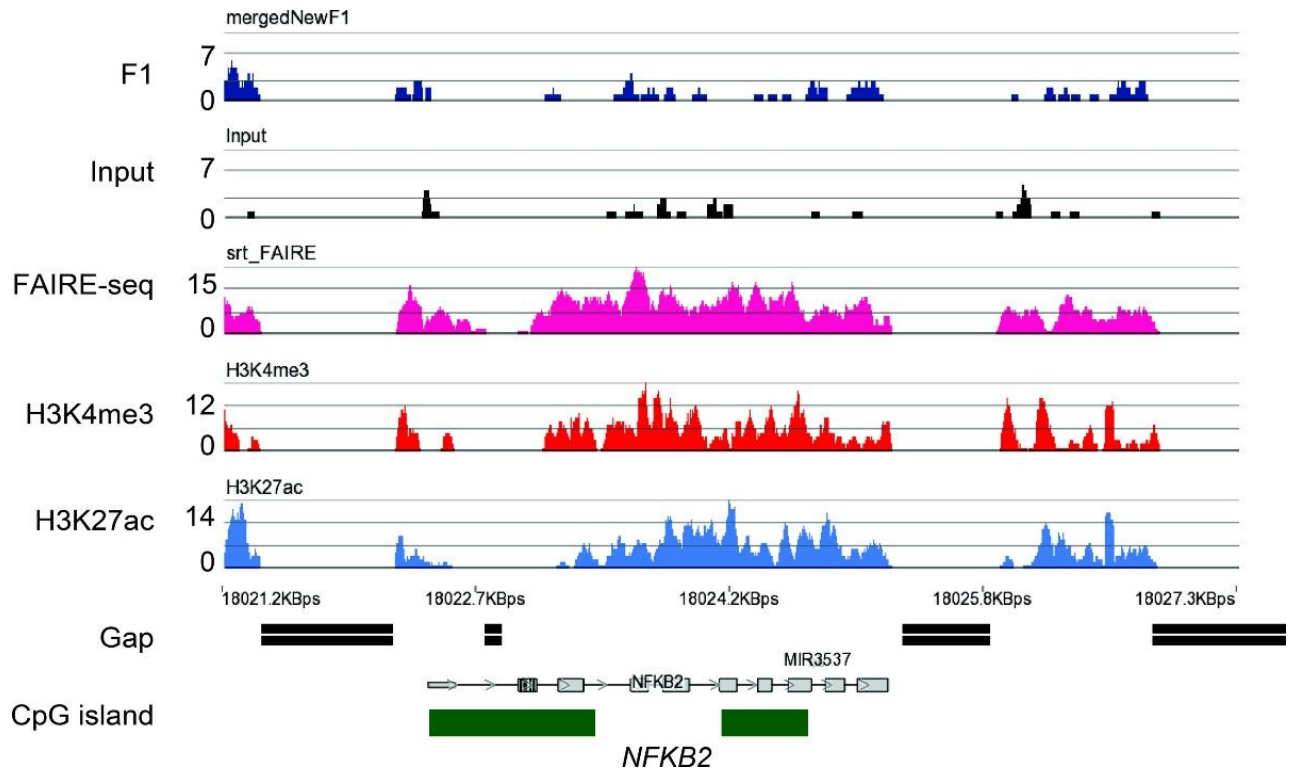


Figure 5.2: Chromatin profile of the *NFKB2* gene region. The signal tracks for F1, FAIRE, H3K4me3, H3K27ac, and input DNA for the ChIP-seq studies are shown. The positions of the CpG islands and sequencing gaps are shown.

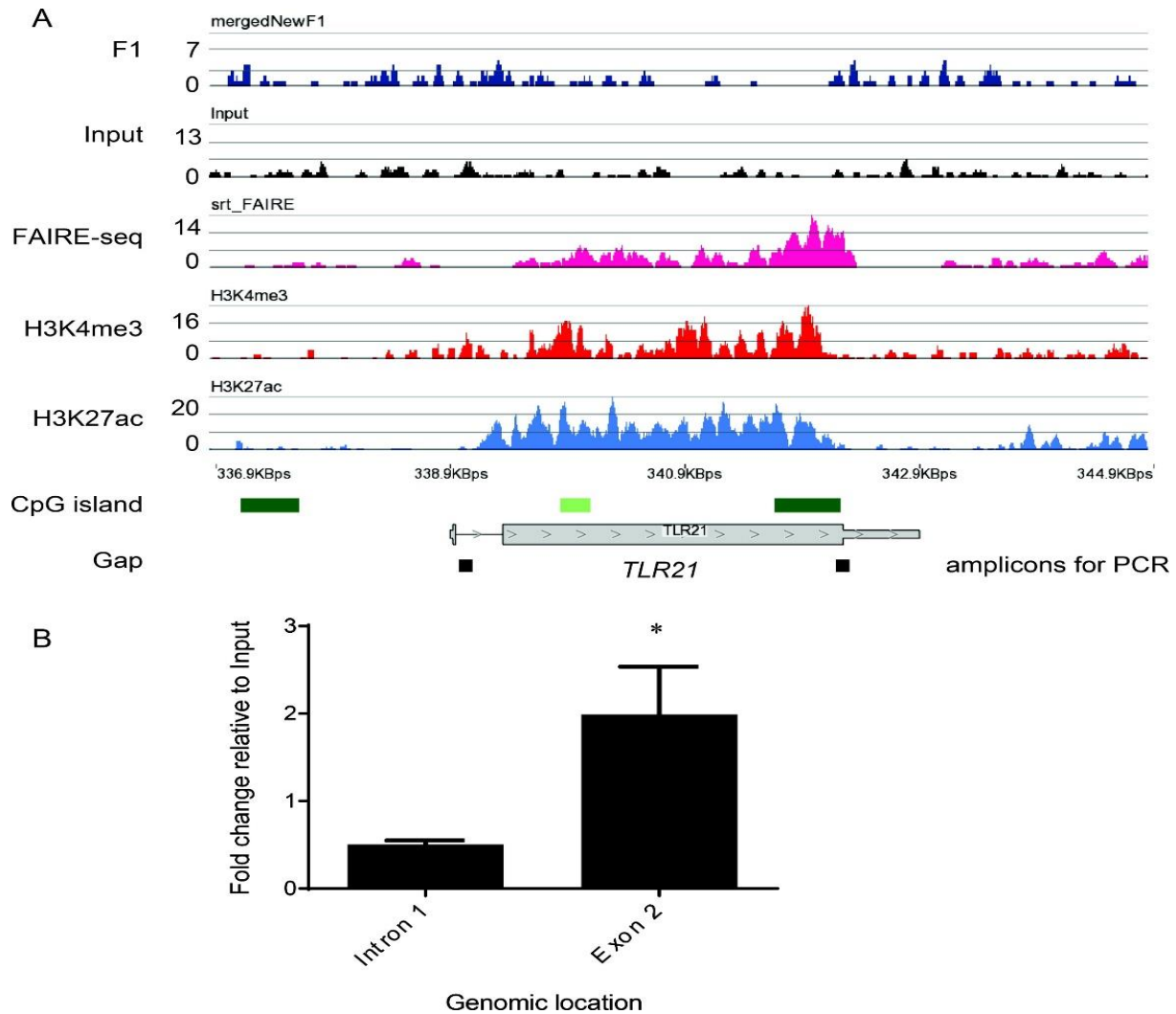


Figure 5.3: Chromatin profile of the *TLR21* gene region. (A) The signal tracks for F1, FAIRE, H3K4me3, H3K27ac, and input DNA for the ChIP-seq studies are shown. The positions of the CpG islands and sequencing gaps are shown. The position of the amplicons to analyze the enrichment of DNA regions in the FAIRE fraction are shown in (A) and the bar graph is shown in (B). The error bars represent the standard deviation of the mean from the three biological replicates. Statistical significance was calculated by comparing the FAIRE DNA with the input DNA. The P values were calculated using Student's t test (*, $P < 0.05$).

Another feature of the *IRF7*, *IL1B*, *NFKB2*, and *TLR21* genes was a repeat pattern of H3K27ac over the gene body, which was reminiscent of positioned nucleosomes (see Figures 5.2 and 5.3). We determined the average length of repeat H3K27ac peaks along the bodies of these genes. The average distance of the peaks was 158 bp (*IL1B*, 160; *IRF7*, 146; *NFKB2*, 148; *TLR21*, 176). This value is consistent with positioned H3K27 modified nucleosomes along these gene bodies.

5.4.4 Nucleosome instability in human gene bodies associated with H3K27ac

Genes involved in cell identity are marked by broad domains of H3K4me3 and, in some cases, H3K27ac also. These broad H3K4me3 domains may cover an entire gene body (19–21). Prompted by our observations that gene bodies with extensive H3K27ac/H3K4me3 modifications exhibited nucleosome instability, we determined whether the nucleosome-free feature, as measured by DNase-seq (DNase I hypersensitive sites sequencing) (22), was associated with the bodies of genes associated with the broad domains of H3K4me3 and H3K27ac. Using data available in ENCODE, we examined the IKAROS family zinc finger 1 (*IKZF1*) tumor suppressor gene, which has H3K4me3/H3K27ac over much of the gene body in CD4+ T cells (20). Figure 5.4 shows that the gene body had an accessible NDR that tracked with H3K27ac. In contrast, the fidgetin like 1 (*FIGNL1*) gene had H3K27ac, H3K4me3, and a NDR at the promoter/5' end of the gene body, which is typical of the organization of most transcribed genes. The GATA binding protein 2 (*GATA2*) gene expressed in K562 cells has a broad domain of H3K4me3 and H3K27ac. As with the *IKZF1* gene, the *GATA2* gene body had a nucleosome-free feature (Supplementary data, Figure 5.S8). These results provide evidence that nucleosomes modified by H3K4me3 and

H3K27ac and associated with transcribed gene bodies in chicken or human cells, have an unstable character.

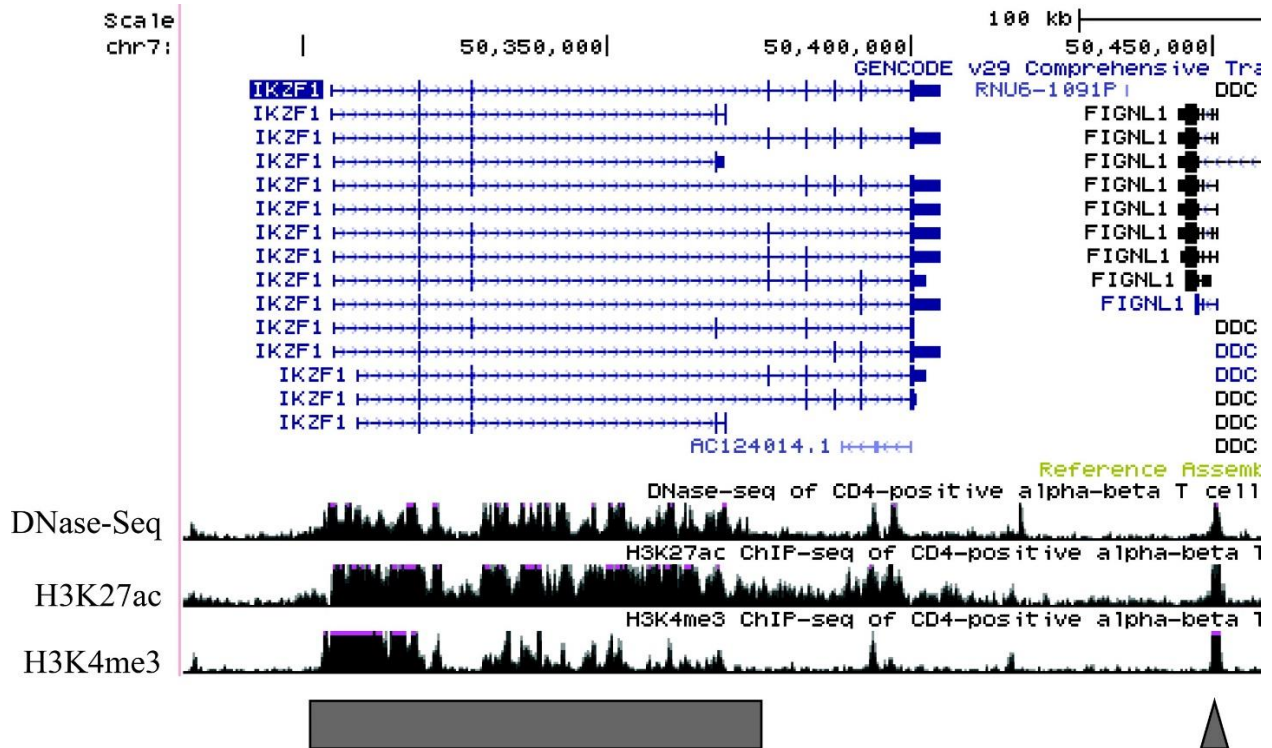


Figure 5.4: Chromatin profile of the human *IKZF1* tumor suppressor gene region in CD4+ T cells. The signal tracks for DNase-seq, H3K27ac, and H3K4me3 are shown. The rectangle shows the position of the broad H3K4me3 region over the *IKZF1* gene, while the triangle points to the location of the H3K4me3/H3K27ac/DNase seq peak at the promoter/5' end of the *FIGNL1* gene. Data taken from the ENCODE portal (23).

5.5 Discussion

The genes involved in innate immunity are expressed in mature and polychromatic erythrocytes from chickens. The investigators studying the induction of the innate immunity genes in chicken mature erythrocytes applied rigorous methods, such as consecutive density gradient centrifugations(2), to ensure that the erythrocytes were not contaminated with white blood cells. In addition to visually checking that our preparations of polychromatic erythrocytes did not contain white blood cells, we analyzed the transcriptome of 6C2 cells, and found that genes involved in innate immunity were expressed (data not shown). Taken together, these studies demonstrate that the red blood cells of chickens participate in protecting the host from microbial pathogens. This physiological role of the chicken and fish erythrocytes participating in the immune response is not unique to birds and fish, as mammalian (human and mouse) erythroblasts are also involved in the immune response (3, 5).

What is causing the basal expression of the genes participating in the immune response in polychromatic erythrocytes from chickens is unknown; possible players are microbial pathogens and stress (24). Stress has been shown to cause altered DNA methylation of genes involved in the immune system in the red blood cells of adult hens (25). It is of interest to understand the relationship between DNA methylation and chromatin organization of the innate immunity genes.

Several genes characterized in this study had H3K4me3, H3K27ac, and NDR over the gene body. The combination of a NDR and modified nucleosomes suggests that the nucleosomes along these gene bodies are undergoing dissolution and reassembly (Figure 5.5). Analyses of the

H3K27ac peaks along the *IRF7*, *IL1B*, *NFKB2*, and *TLR21* genes revealed a periodicity of the nucleosomes along the gene body, suggesting that there may be a chromatin-remodeling activity positioning the nucleosomes. The current literature provides evidence that the chromatin remodeler is chromodomain helicase DNA binding protein 1 (CHD1)(26–28). Because CHD1 is involved in nucleosome disassembly and nucleosome spacing (26), this ATP-driven chromatin remodeler may be responsible for instability and the positioning of H3K27ac nucleosomes along these genes. In *Drosophila*, CHD1 plays a role in the regulation of the immune response, stress response, and detoxification processes (29). Studies are underway to explore the role of CHD1 in the dynamics of nucleosome positioning and stability in chicken erythrocytes.

Histone modifications (acetylation, ubiquitination) and variants (H3.3) contribute to nucleosome instability (10, 31–33). These variants (H3.3) and modifications may confer a transcriptional memory onto the genes involved in the innate immune response (34). CHD1 may have a role in introducing H3.3 into nucleosomes following reassembly (26, 35, 36). We have reported that nucleosomal histones of transcribed genes exchanged with newly synthesized histones, independent of transcription (10). Exchange was most evident for the newly synthesized histones H3.3, H2A, H2A.Z and H2B, with H2A and H2B being prominently ubiquitinated and acetylated. This feature of nucleosome instability over the gene body was not unique to the erythroid genes of chickens, as shown with analyses of human genes that have broad H3K4me3 and H3K27ac covering the gene body (Figure 5.4). Further, we have observed this nucleosome instability over the body of breast cancer genes that have broad H3K4me3 domains (C. Lopez, G. Nardocci, I. Kovalchuk, G.P. Delcuve, T.H. Beacon, and J.R. Davie, unpublished data). The dynamic nucleosome structure and composition of genes involved in the immune response and cell identity may facilitate transcriptional elongation. Taken together, the results from this study

provide an excellent example of how the analysis of chromatin organization in chicken erythrocytes informs us about the human context.

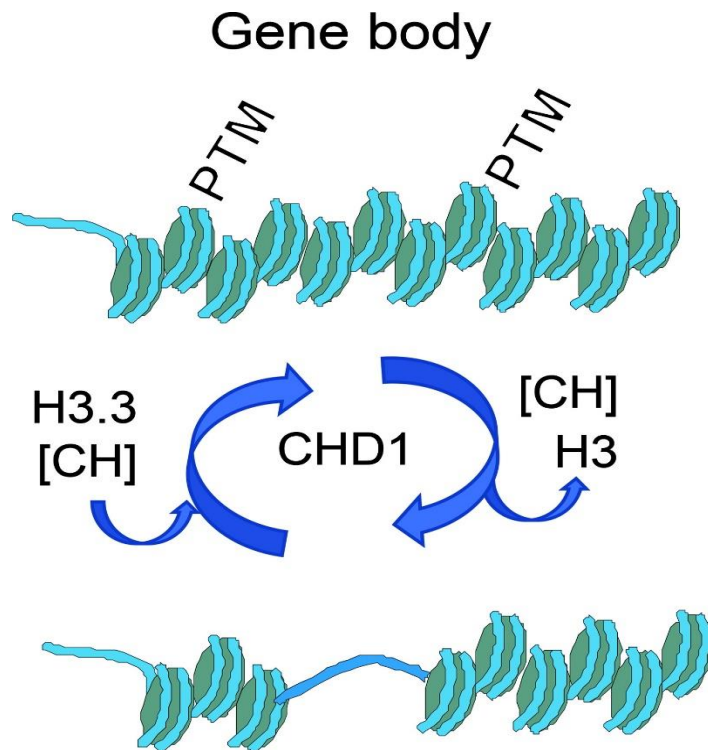


Figure 5.5: Model for nucleosome instability over the gene body of chicken erythrocyte immune genes and human genes involved in cell identity. Nucleosomes are being removed and reassembled due to the action of chromodomain helicase DNA binding protein 1 (CHD1). CHD1 binds to nucleosomes with H3K4me3 (30). Core histone (CH) post-translational modifications (PTMs) would be removed with the displaced nucleosome. Nucleosome reassembly would allow for incorporation of histone variants such as H3.3 and a new set of core histone PTMs. The dynamic nucleosome organization of the coding region would facilitate transcription elongation.

5.6 Acknowledgements

This work was supported by a grant from the Natural Sciences and Engineering Research Council of Canada (RGPIN-2016-5911) to J.R.D., a Research Manitoba award funded by Children's Hospital Research Institute of Manitoba and the Research Institute in Oncology and Hematology to S.J., and a Graduate Enhancement of Tri-Council Stipends award to T.H.B. We thank Zahra Sepehri for reviewing the literature of stress and immune response in chickens. We thank Aleksandar Ilic for preparation of the sequencing libraries.

5.7 References

- [1] Morera, D., and MacKenzie, S. A. (2011) Is there a direct role for erythrocytes in the immune response? *Veterinary Research* 42, 89.
- [2] Morera, D., Roher, N., Ribas, L., Balasch, J. C., Doñate, C., et al. (2011) RNA-Seq Reveals an Integrated Immune Response in Nucleated Erythrocytes. *PLoS ONE* 6, e26998.
- [3] Sennikov, S. V., Injelevskaya, T. V., Krysov, S. V., Silkov, A. N., Kovinev, I. B., et al. (2004) Production of hemo- and immunoregulatory cytokines by erythroblast antigen+ and glycophorin A+ cells from human bone marrow. *BMC Cell Biology* 5, 39.
- [4] St. Paul, M., Paolucci, S., Barjesteh, N., Wood, R. D., and Sharif, S. (2013) Chicken erythrocytes respond to Toll-like receptor ligands by up-regulating cytokine transcripts. *Research in Veterinary Science* 95, 87–91.
- [5] Elahi, S. (2014) New insight into an old concept: Role of immature erythroid cells in immune pathogenesis of neonatal infection. *Frontiers in Immunology* 5, 376.
- [6] Arighi, C., Shamovsky, V., Masci, A. M., Ruttenberg, A., Smith, B., et al. (2015) Correction: Toll-Like Receptor Signaling in Vertebrates: Testing the Integration of Protein, Complex, and Pathway Data in the Protein Ontology Framework. *PLOS ONE* 10, e0131148.
- [7] Medzhitov, R. (2007) Recognition of microorganisms and activation of the immune response. *Nature* 449, 819–826.
- [8] Abe, H., and Gemmell, N. J. (2014) Abundance, arrangement, and function of sequence

- motifs in the chicken promoters. *BMC Genomics* 15, 1–12.
- [9] Williams, A. F. (1972) DNA synthesis in purified populations of avian erythroid cells. *Journal of Cell Science* 10, 27–46.
- [10] Hendzel, M. J., and Davie, J. R. (1990) Nucleosomal histones of transcriptionally active/competent chromatin preferentially exchange with newly synthesized histones in quiescent chicken erythrocytes. *Biochemical Journal* 271, 67–73.
- [11] Locklear, L., Ridsdale¹, J. A., Bazett-Jones, D. P., and Davie¹, J. R. (1990) Ultrastructure of transcriptionally competent chromatin.
- [12] Jahan, S., Xu, W., He, S., Gonzalez, C., Delcuve, G. P., et al. (2016) The chicken erythrocyte epigenome. *Epigenetics and Chromatin* 9, 1–11.
- [13] Jahan, S., Beacon, T. H., He, S., Gonzalez, C., Xu, W., et al. (2019) Chromatin organization of transcribed genes in chicken polychromatic erythrocytes. *Gene* 699, 80–87.
- [14] Delcuve, G. P., and Davie, J. R. (1989) Chromatin structure of erythroid-specific genes of immature and mature chicken erythrocytes. *Biochemical Journal* 263, 179–186.
- [15] Rocha[@], E., Davies¹, J. R., Van Holde^{\$ll}, K. E., and Weintraub, H. (1984) Differential Salt Fractionation of Active and Inactive Genomic Domains in Chicken Erythrocyte". *The Journal of biological chemistry* 259, 8558–8563.
- [16] Zang, C., Schones, D. E., Zeng, C., Cui, K., Zhao, K., et al. (2009) Data and text mining A clustering approach for identification of enriched domains from histone modification ChIP-Seq data. *Bioinformatics* 25, 1952–1958.
- [17] Zhang, Y., Liu, T., Meyer, C. A., Eeckhoutte, J., Johnson, D. S., et al. (2008) Model-based analysis of ChIP-Seq (MACS). *Genome Biology* 9, R137.

- [18] Robinson, J. T., Thorvaldsdóttir, H., Winckler, W., Guttman, M., Lander, E. S., et al. (2011) Integrative genomics viewer. *Nature Biotechnology* 29, 24–26.
- [19] Benayoun, B. A., Pollina, E. A., Ucar, D., Mahmoudi, S., Karra, K., et al. (2014) H3K4me3 breadth is linked to cell identity and transcriptional consistency. *Cell* 158, 673–688.
- [20] Chen, K., Chen, Z., Wu, D., Zhang, L., Lin, X., et al. (2015) Broad H3K4me3 is associated with increased transcription elongation and enhancer activity at tumor-suppressor genes. *Nature Genetics* 47, 1149–1157.
- [21] Cao, F., Fang, Y., Tan, H. K., Goh, Y., Choy, J. Y. H., et al. (2017) Super-enhancers and broad h3k4me3 domains form complex gene regulatory circuits involving chromatin interactions. *Scientific Reports* 7, 2186.
- [22] Song, L., and Crawford, G. E. (2010) DNase-seq: A high-resolution technique for mapping active gene regulatory elements across the genome from mammalian cells. *Cold Spring Harbor Protocols* 2010, pdb.prot5384.
- [23] Davis, C. A., Hitz, B. C., Sloan, C. A., Chan, E. T., Davidson, J. M., et al. (2018) The Encyclopedia of DNA elements (ENCODE): data portal update. *Nucleic Acids Research* 46, D794–D801.
- [24] BN, H., MG, N., G, D. V. R., J, B., H, V. D. B., et al. (2005) Severe Feed Restriction Enhances Innate Immunity but Suppresses Cellular Immunity in Chicken Lines Divergently Selected for Antibody Responses. *Poultry science* 84, 1520–1529.
- [25] Pértille, F., Brantsæter, M., Nordgreen, J., Coutinho, L. L., Janczak, A. M., et al. (2017) DNA methylation profiles in red blood cells of adult hens correlate with their rearing conditions. *Journal of Experimental Biology* 220, 3579–3587.

- [26] Siggens, L., Cordeddu, L., Rönnerblad, M., Lennartsson, A., and Ekwall, K. (2015) Transcription-coupled recruitment of human CHD1 and CHD2 influences chromatin accessibility and histone H3 and H3.3 occupancy at active chromatin regions. *Epigenetics and Chromatin* 8, 4.
- [27] Zubek, J., Stitzel, M. L., Ucar, D., and Plewczynski, D. M. (2016) Computational inference of H3K4me3 and H3K27ac domain length. *PeerJ* 4, e1750.
- [28] Marfella, C. G. A., and Imbalzano, A. N. (2007) The Chd family of chromatin remodelers. *Mutation Research - Fundamental and Molecular Mechanisms of Mutagenesis* 618, 30–40.
- [29] Sebald, J., Morettini, S., Podhraski, V., Lass-Flörl, C., and Lusser, A. (2012) CHD1 Contributes to Intestinal Resistance against Infection by *P. aeruginosa* in *Drosophila melanogaster*. *PLoS ONE* 7, e43144.
- [30] Davie, J. R., Xu, W., and Delcuve, G. P. (2015) Histone H3K4 trimethylation: Dynamic interplay with pre-mRNA splicing. *Biochemistry and Cell Biology* 94, 1–11.
- [31] Li, W., Nagaraja, S., Delcuve, G. P., Hendzel, M. J., and Davie, J. R. (1993) Effects of histone acetylation, ubiquitination and variants on nucleosome stability. *Biochemical Journal* 296, 737–744.
- [32] Jin, C., and Felsenfeld, G. (2006) Distribution of histone H3.3 in hematopoietic cell lineages. *Proceedings of the National Academy of Sciences of the United States of America* 103, 574–579.
- [33] Jin, C., and Felsenfeld, G. (2007) Nucleosome stability mediated by histone variants H3.3 and H2A.Z. *Genes and Development* 21, 1519–1529.
- [34] Kamada, R., Yang, W., Zhang, Y., Patel, M. C., Yang, Y., et al. (2018) Interferon

stimulation creates chromatin marks and establishes transcriptional memory. *Proceedings of the National Academy of Sciences of the United States of America* 115, E9162–E9171.

[35] Konev, A. Y., Tribus, M., Sung, Y. P., Podhraski, V., Chin, Y. L., et al. (2007) CHD1 motor protein is required for deposition of histone variant H3.3 into chromatin in vivo. *Science* 317, 1087–1090.

[36] Zhang, K., Rajput, S. K., Wang, S., Folger, J. K., Knott, J. G., et al. (2016) CHD1 Regulates Deposition of Histone Variant H3.3 During Bovine Early Embryonic Development1. *Biology of Reproduction* 94, 140,1–8.

5.8 Supplementary Data

Supplementary Table 5.S1: Primers used for RT-PCR

| Primer | Sequences |
|-----------|-------------------------------|
| TLR3-F | 5'-CCCTGATGGAGTGTGTTGCTT-3' |
| TLR3-R | 5'-CCAGGGTTTTGAAAGGATCA-3' |
| TLR21-F | 5'-AAAGGAGAAAGCGGCTGAG-3' |
| TLR21-R | 5'-GACAAGGACAGGGACAGAGC-3' |
| IL1B-F | 5'-CTGAGCACACCACAGTGG-3' |
| IL1B-R | 5'-GCAGCAGTTTGGTCATGG-3' |
| IRF-1-F | 5'-TCATCTCATCTCGTCTCATCTCA-3' |
| IRF-1-R | 5'-CTGTGCTGTGCTGTGTTGTG-3' |
| TIRAP-1-F | 5'-CAGCCCCACCTCAGACAC-3' |
| TIRAP-1-R | 5'-GGTGGAAAGGCTGGAATC-3' |
| CIAPIN1-F | 5'-CTGTGAGATTGGCGTGGAC-3' |
| CIAPIN1-R | 5'-GAGCGGGATAGAGGTGAGAG-3' |
| IL-15-F | 5'-GCAATGTATTTCCCGATCCA-3' |
| IL-15-R | 5'-CTCCGGCAGAGTTTTGTGTT-3' |
| 18S-F | 5'-GTAACCCGTTGAACCCATT-3' |
| 18S-R | 5'-CCATCCAATCGGTAGTAGCG-3' |

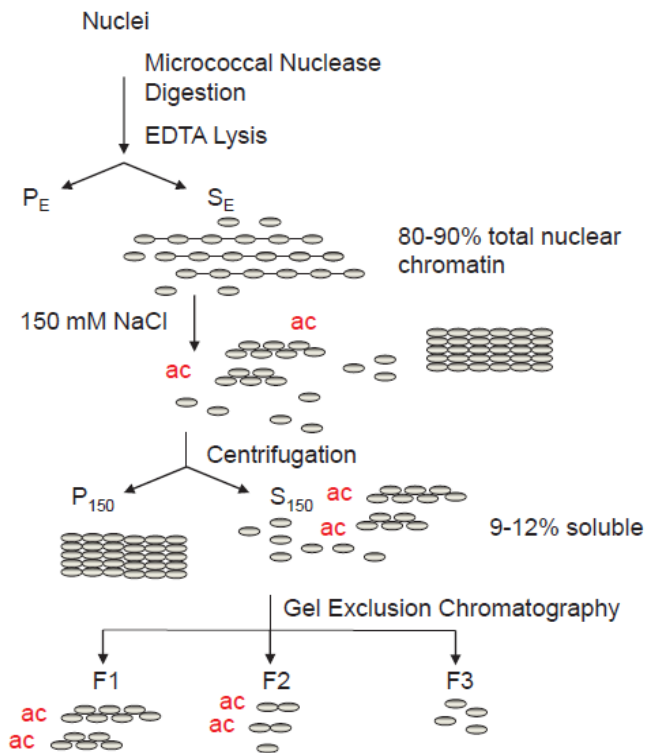
Supplementary Table 5.S2. Primers used in FAIRE-qPCR experiments

| Gene | Region | Primer sequence |
|--|---------------|--|
| <i>TLR21</i> | Intron 1 | FP: CCAGTCCTACATTGGTCCTATC RP: GGCATATGGAGTTGAGGATTTG |
| | Exon 2 | FP: ATGTAGTGGAGGTGTTAG RP: CATATGGATAGTTGCTGTC |
| <i>IRF7</i> | Intron 7 | FP: CATTCCCTGGCACCCATATC RP: CCTCTCACCTTATGTCACCCT |
| | Exon 3 | FP: GCCAAATGGAAGACCAACTTC RP: AACCGCGTAGACCTTGTG |
| | Exon 8 | FP: GCCTGAAGAAGTGCAAGGT RP: AACACCCTGAAGTCGAAGATG |
| FP: Forward primer; RP: Reverse primer; 5' to 3' shown | | |

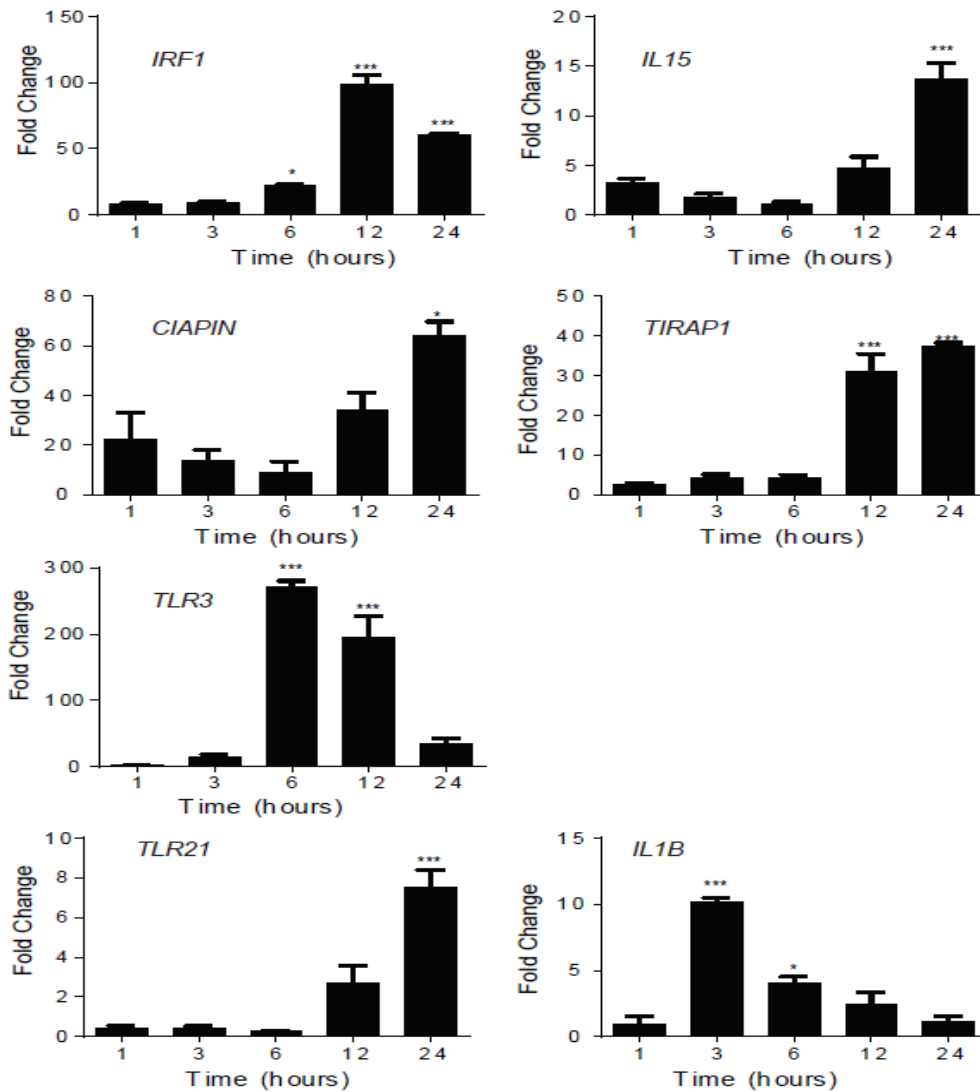
Supplementary Table 5.S3. Ranking of genes involved in the immune response according to expression levels in chicken polychromatic erythrocytes. All 5479 genes from the galGal3 RefSeq database were placed in order of their level of expression (mean RPKM from two biological samples) and were then divided into five 20th-percentile classes in relation to expression level. First 20 %-ile (red filled), second 20 %-ile (green filled), third 20 %-ile (blue filled), fourth 20 %-ile (dark blue filled) and fifth 20 %-ile (violet filled) (Jahan et al. 2016). Out of these 5479 genes, only the expressed genes involved in the immune response are listed. PR CGI indicates whether the gene had a CGI over the promoter region.

| Gene | Chr | Strand | Start | End | Transcripts | RPKM | PR CGI |
|---------|-------|--------|------------|------------|--------------|---------|--------|
| IRF1 | chr13 | - | 17453125 | high | NM_205415 | 1107.42 | Y |
| IL15 | chr4 | + | 31129925 | 31162068 | NM_204571 | 378.92 | N |
| IRF2 | chr4 | + | 40,941,772 | 40,965,995 | NM_205196 | 215 | N |
| IRF7 | chr5 | + | 16,950,071 | 16,954,585 | NM_205372 | 158 | Y |
| TRAM1 | chr2 | - | 121362691 | 121380275 | NM_204400 | 123.57 | Y |
| IRF8 | chr11 | - | 19,610,851 | 19,616,382 | NM_205416 | 87.1 | Y |
| CIAPIN1 | chr11 | + | 566342 | 572320 | NM_001005834 | 79.45 | Y |
| TIRAP | chr24 | + | 426283 | 429814 | NM_001024829 | 60.95 | Y |
| TLR3 | chr4 | + | 63155888 | 63160902 | NM_001011691 | 27.29 | N |
| IRF9 | chr20 | - | 10,018,363 | 10,021,470 | NM_204558 | 26.5 | Y |
| MYD88 | chr2 | + | 4,730,082 | 4,742,683 | NM_001030962 | 5.02 | Y |
| NFKB2 | chr6 | + | 18,022,427 | 18,025,184 | NM_204413 | 4.77 | Y |
| IRAK2 | chr12 | + | 20,078,601 | 20,092,031 | NM_001030605 | 4.62 | N |
| CCL20 | chr9 | + | 10,616,596 | 10,618,706 | NM_204438 | 4.37 | N |
| IL9 | chr13 | - | 15,573,175 | 15,575,874 | NM_001037825 | 3.94 | N |
| TLR7 | chr1 | - | 126823955 | 126830698 | NM_001011688 | 3.29 | N |
| TICAM1 | chr28 | + | 4,419,724 | 4,422,751 | NM_001081506 | 2.64 | Y |
| IFNW1 | chrZ | + | 6,888,979 | 6,889,590 | NM_001024836 | 1.90 | N |
| IL7 | chr2 | - | 125117020 | 125123266 | NM_001037833 | 1.365 | N |
| TLR6 | chr4 | + | 71,549,952 | 71,555,092 | NM_001081709 | 1.33 | N |
| CXCL12 | chr6 | - | 20,112,185 | 20,120,138 | NM_204510 | 1.28 | Y |
| TLR21 | chr11 | + | 338863 | 342853 | NM_001030558 | 0.96 | N |
| TLR15 | chr3 | - | 2,945,856 | 2,948,462 | NM_001037835 | 0.84 | N |

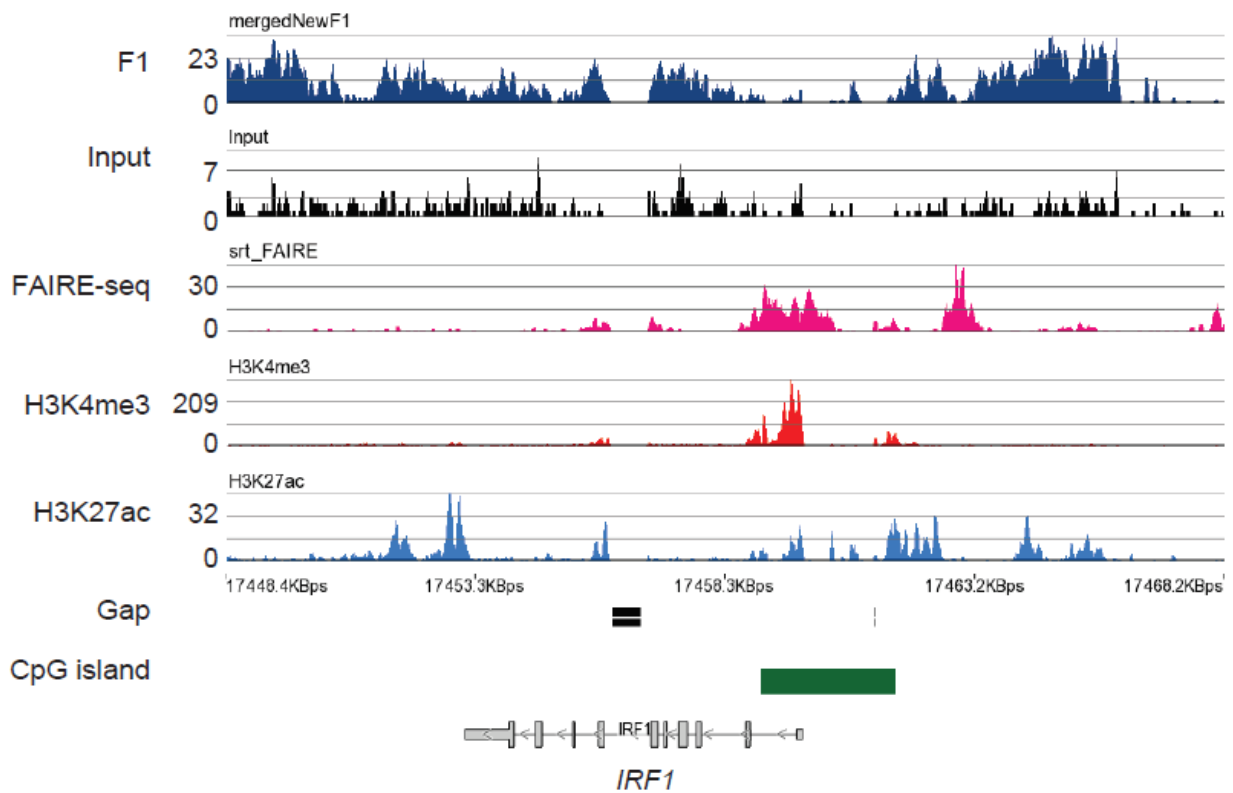
| | | | | | | | |
|--------|-------|---|------------|------------|--------------|------|--------|
| IFNG | chr1 | + | 87,329,807 | 87,333,781 | NM_205046 | 0.71 | N |
| CCL4 | chr19 | + | 373,753 | 375,322 | NM_204720 | 0.66 | Y weak |
| IFNA3 | chrZ | + | 6,896,104 | 6,896,866 | NM_205427 | 0.65 | N |
| TLR2 | chr4 | + | 21,101,227 | 21,108,624 | NM_204278 | 0.53 | N |
| IL18 | chr24 | + | 6,291,980 | 6,295,249 | NM_204608 | 0.53 | Y |
| IL16 | chr10 | - | 13,802,549 | 13,810,903 | NM_204352 | 0.5 | N |
| TLR4 | chr17 | + | 4,062,994 | 4,068,447 | NM_001030693 | 0.46 | N |
| TLR5 | chr3 | + | 18,975,945 | 18,978,530 | NM_001024586 | 0.39 | N |
| IL1B | chr22 | - | 3,876,886 | 3,878,491 | NM_204524 | 0.35 | Y |
| IL2 | chr4 | + | 55,255,551 | 55,258,596 | NM_204153 | 0.34 | N |
| IL3 | chr13 | + | 17,245,232 | 17,250,126 | NM_001007083 | 0.32 | N |
| IL8 | chr4 | + | 52,446,739 | 52,449,903 | NM_205498 | 0.32 | N |
| IL5 | chr13 | - | 17,472,321 | 17,483,539 | NM_001007084 | 0.30 | N |
| IL22 | chr1 | - | 36,974,029 | 36,976,482 | NM_001199614 | 0.29 | N |
| CXCL14 | chr13 | + | 15,684,189 | 15,691,454 | NM_204712 | 0.16 | Y |



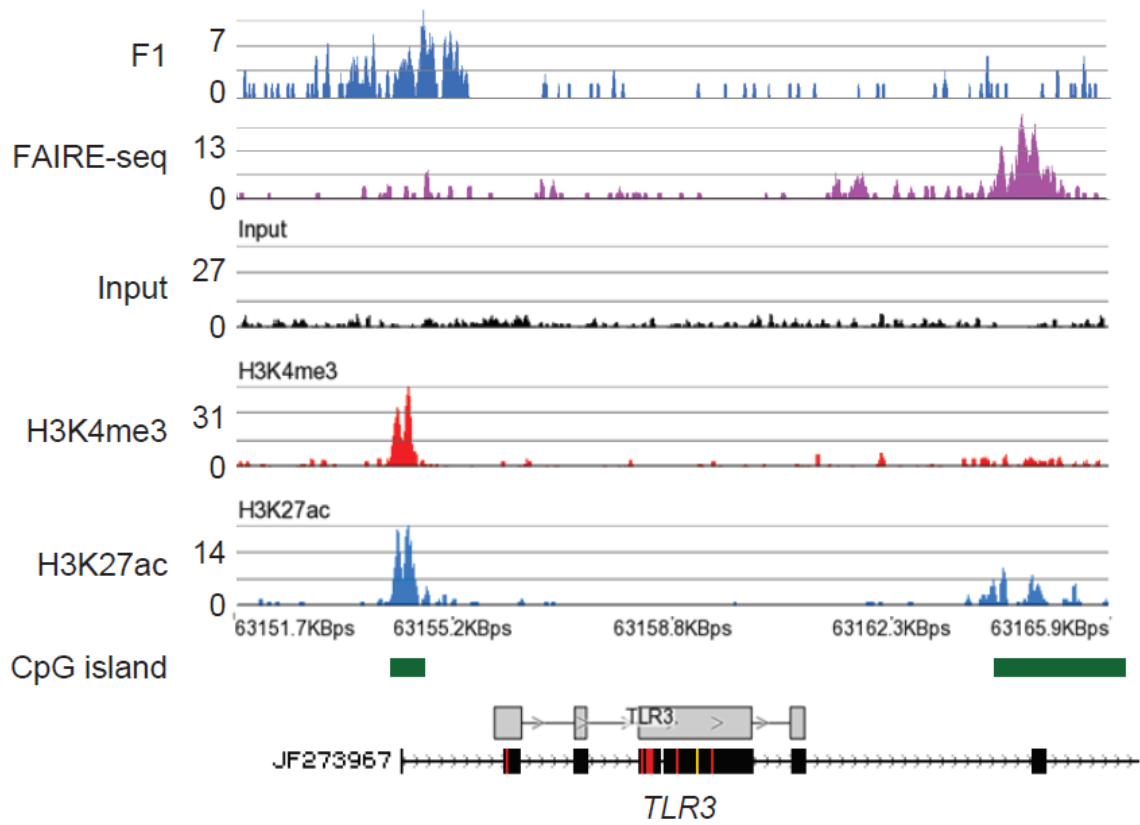
Supplementary Figure 5.S1: Fractionation of avian erythrocyte chromatin. Chicken polychromatic erythrocyte nuclei were incubated with micrococcal nuclease, and chromatin fragments soluble in a low ionic strength solution containing 10 mM EDTA were recovered in fraction S_E. Chromatin fraction S_E was made 150 mM in NaCl, and chromatin fragments from the salt-soluble fraction (S₁₅₀) were size-resolved on a Bio-Gel A-1.5 m column to isolate the F1 fraction containing polynucleosomes.



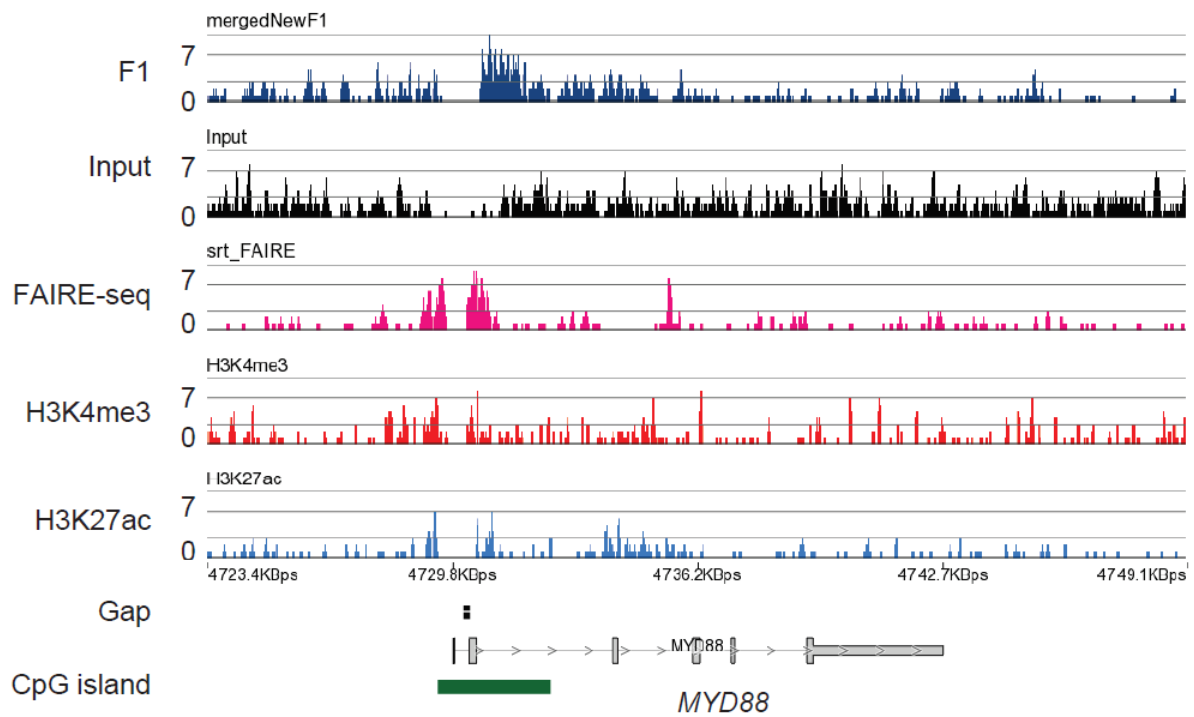
Supplementary Figure 5.S2: Poly I:C mediated induction of immune genes in chicken polychromatic erythrocytes. Fold changes were normalized to 18S rRNA levels and were calculated by comparing with untreated samples for each time point. The genes are arranged according to the gene's basal transcript levels (Table 1), with *IRF1*, *IL15*, *CIAPIN* and *TIRAP1* being +, *TLR3* being +, and *TLR21* and *IL1B* being +. Error bars represent the standard error of the mean from three independent biological experiments. Statistical significance was calculated in reference to the 1-hour treatment using one-way ANOVA * $P < 0.05$, ** $P < 0.01$ or *** $P < 0.001$.



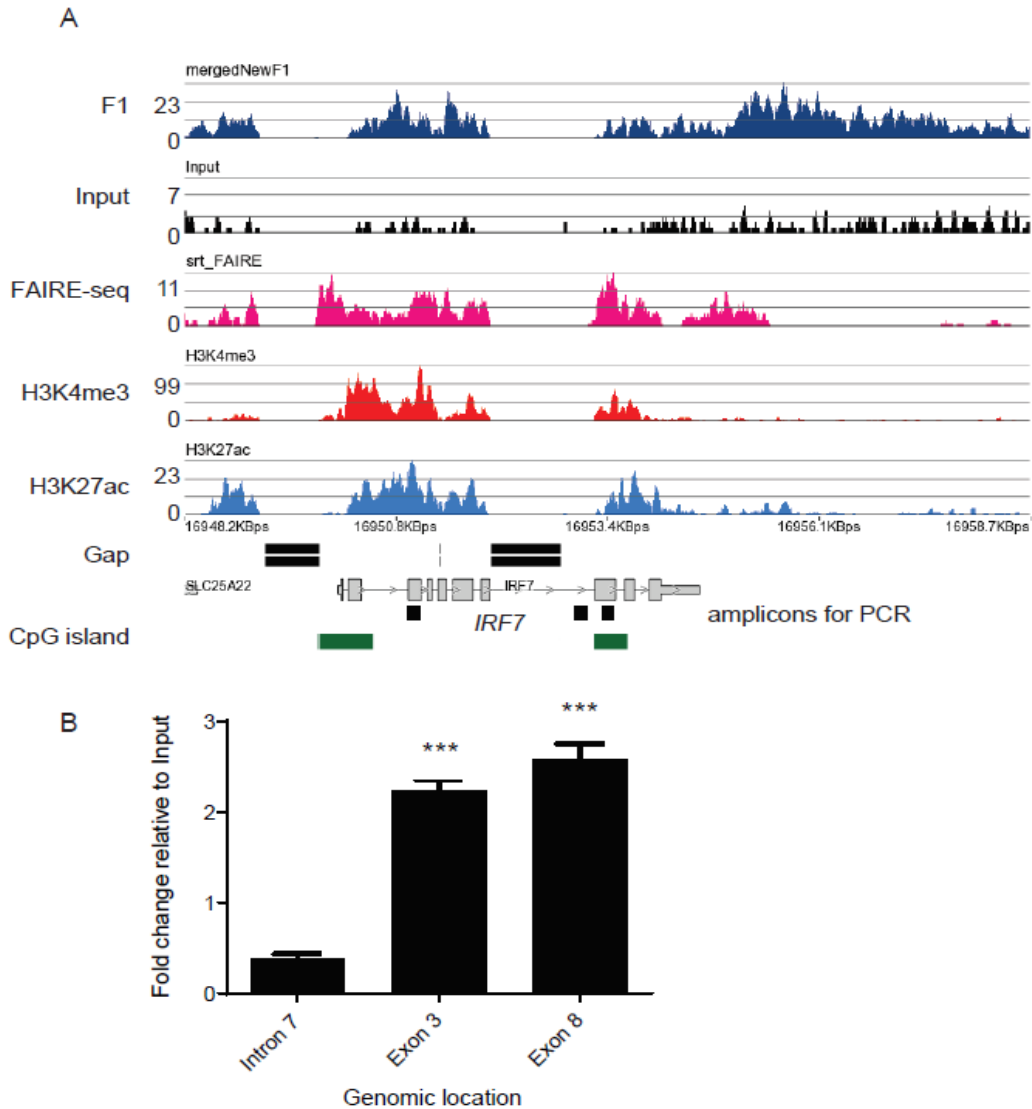
Supplementary Figure 5.S3: Chromatin profile of the *IRF1* gene region. The signal tracks for F1, FAIRE, H3K4me3, H3K27ac and input DNA for the ChIP-seq studies are shown. The positions of the CGIs and sequencing gaps are shown.



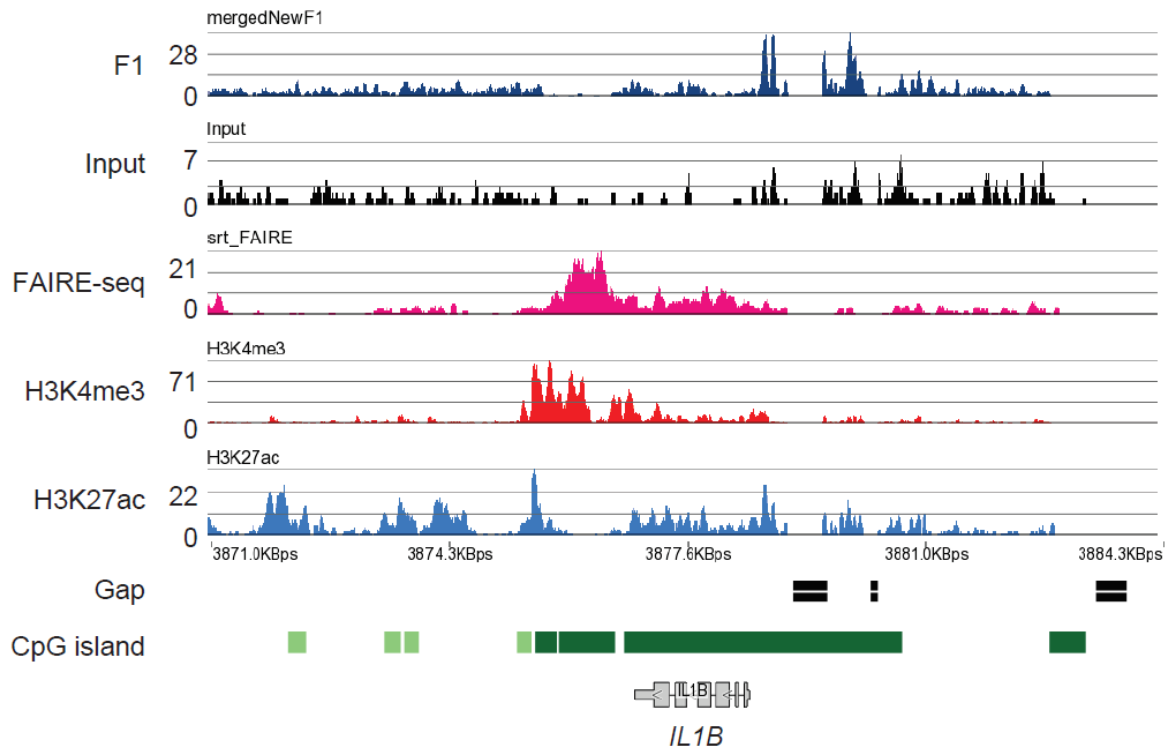
Supplementary Figure 5.S4: Chromatin profile of the TLR3 gene region. As in Supplementary Figure 5.S3.



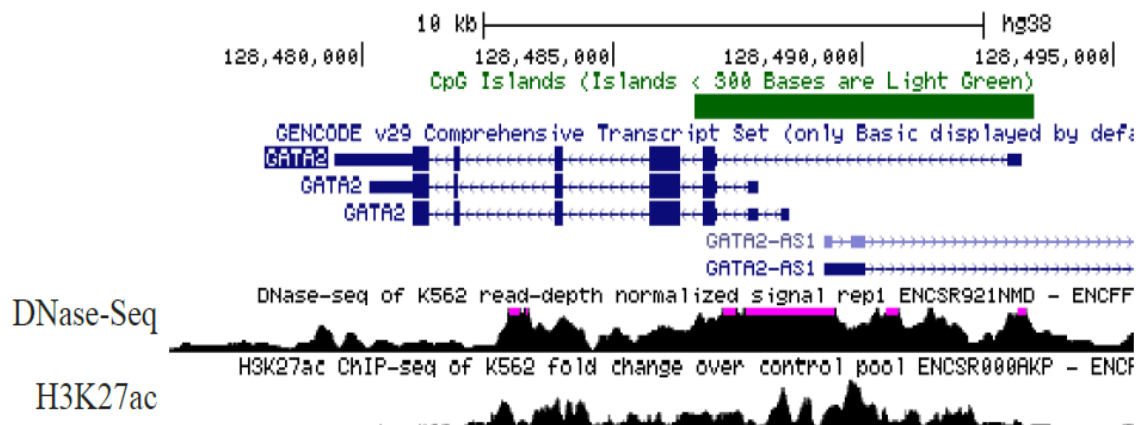
Supplementary Figure 5.S5. Chromatin profile of the MYD88 gene region. As in Supplementary Figure 5.S3.



Supplementary Figure 5.S6: Chromatin profile of the *IRF7* gene region. (A) The signal tracks for F1, FAIRE, H3K4me3, H3K27ac and input DNA for the ChIP-seq studies are shown. The positions of the CGIs and sequencing gaps are shown. The position of the amplicons to analyze the enrichment of DNA regions in the FAIRE fraction are shown in (A) and the bar graph is shown in (B). The error bars represent the standard deviation of the mean from the three biological replicates. Statistical significance was calculated by comparing FAIRE DNA to the input DNA. The P values were calculated with the Student t-test (***)(<math>P < 0.001</math>).



Supplementary Figure 5.S7: Chromatin profile of the *IL1B* gene region. As in Supplementary Figure 5.S3.



Supplementary Figure 5.S8: Chromatin profile of the human *GATA2* gene region in erythroleukemic K562 cells. Data taken from ENCODE portal (PMID: 29126249) (<http://www.genome.ucsc.edu/ENCODE/>).

Chapter-6

Genomic landscape of transcriptionally active histone arginine methylation marks, H3R2me2s and H4R3me2a, relative to nucleosome depleted regions

**Genomic landscape of transcriptionally active histone arginine
methylation marks, H3R2me2s and H4R3me2a, relative to
nucleosome depleted regions**

Tasnim H. Beacon, Wayne Xu James R. Davie

Source of the published article:

Gene, Volume 742, 5 June 2020, 144593

<https://doi.org/10.1016/j.gene.2020.144593>

Author contributions

J.R.D designed the studies. **T.H.B** conducted the experiments; **T.H.B** carried out the illustrated bioinformatic analyses using the galGal 6 assembly; X.W. carried out the initial bioinformatic analyses using the galGal 3 assembly. J.R.D. provided insight for analysis and with **T.H.B.** prepared the manuscript.

6.1 Abstract

PRMT1 and the product of this enzyme H4R3me2a are important in the establishment and maintenance of chicken and murine erythrocyte transcriptionally active chromatin. Silencing the expression of PRMT1 results in loss of acetylated histones H3 and H4 and methylated H3K4 and prevents erythropoiesis. Here, we show that H4R3me2a and the PRMT5-catalyzed H3R2me2s locate largely to introns of expressed genes and intergenic regions, with both marks co-localizing in the chicken polychromatic erythrocyte genome. H4R3me2a and H3R2me2s were associated with histone marks of active promoters and enhancers, as well as with the body of genes that have an atypical chromatin structure, with nucleosome depleted regions. H4R3me2a co-localized with acetylated H3K27. Previous studies have shown that PRMT1 was bound to CBP/p300, suggesting a role of PRMT1-mediated H4R3me2a in CBP/p300 recruitment and H3K27 acetylation. Moreover, PRMT1 might be a key enzyme affected when S-adenosyl methionine levels are reduced in metabolic disorders.

Keywords: Protein arginine methyl transferase, Nucleosome, Chromatin, Histone arginine methylation, Histone acetylation, Hypomethylation

6.2 Introduction

Arginine methylation of proteins and histones catalyzed by PRMTs can be either symmetrical or asymmetrical (1). Histone arginine methylation is involved in the activation and repression of gene expression (2). PRMTs catalyze arginine methylation by using SAM to form asymmetric (ω -NG, NG-dimethyl- arginine) (type I) or symmetric (ω -NG, N'G-dimethylarginine) (type II) or monomethyl arginine. PRMT1 and 6 are type I methyltransferases, while PRMT5 is a type II enzyme (1, 3) . PRMT1 catalyzes the formation of H4R3me2a (active mark), while PRMT5 catalyzes the formation of H3R2me2s (active mark) and H4R3me2s (repressive mark) (4, 5). There are reports on the genomic mapping of H3R2me2s and H4R3me2a in human B cells and human mesenchymal stem cells, respectively (4, 6). However, neither report studied both marks together.

In vertebrates, chromatin is organized into compartments and TADs (7). However, chicken erythrocytes apparently do not have TADs (8). Compartment A contains active genes, while compartment B has repressed genes; these two compartments can be distinguished by their histone and DNA modifications (9–11).

PRMT1 and the enzyme's product H4R3me2a are important in the establishment and maintenance of compartment A chromatin. Silencing the expression of PRMT1, which catalyzes H4R3me2a, results in loss of acetylated H3 and H4 and methylated H3K4, disruption of the interaction between the LCR and the β -globin promoter, and erythropoiesis does not take place (12–14).

The chicken karyotype consists of 38 autosomes and a pair of sex chromosomes (ZW female, ZZ male), and is made up of macrochromosomes (1 – 8, Z) and microchromosomes (9–38 and W) (15–19) . The chicken genome is three times smaller than the human genome, but has about the same number of genes, with 60% of them having a single human orthologue. Moreover, there are long blocks of conserved synteny between the chicken and human genomes (18). In terms of chromosomal organization of genes, the human genome is closer to the chicken than to rodents. It is possible that genomic location of elements (enhancers, LCRs) regulating chicken and human erythroid expressed genes are similar.

Chicken polychromatic erythrocytes are transcriptionally active nucleated cells that are arrested in G0 phase of the cell cycle. As with other vertebrates, the properties of compartment A chromatin of chicken erythrocytes include an enrichment in modified histones [acetylated histones, H3 methylated at K4, H3 phosphorylated at S28, ubiquitinated H2B, H3R2me2s, H4R3me2a], PRMT1, PRMT5, destabilized U shaped nucleosomes that actively exchange their histones with newly synthesized histones (replication-independent variants, for example H3.3), NDR, and solubility at physiological ionic strength (20–28). Transcriptome analyses of chicken polychromatic erythrocytes, showed expression of genes involved in oxygen transport and the innate immune response as in human erythrocytes (28, 29). Several genes involved in innate immunity and erythropoiesis had an atypical chromatin structure in which H3K4me3, H3K27ac and destabilized nucleosomes were present throughout the gene body (28).

Here, we show for the first time the location of H3R2me2s and H4R3me2a in a vertebrate genome. H3 variants and modifications co-localizing with R2me2s and H4R3m2a, and the chromatin-associated proteins bound to PRMT1 and PRMT5 were identified. This study adds to our knowledge of the role of the PRMT1 in maintaining compartment A chromatin.

6.3 Materials and methods

6.3.1 Ethics statement

All methods involving the use of chicken were approved by the committee and carried out in accordance with its guidelines and regulations. The University is currently in full compliance with the Canadian Council on Animal Care who has certified that the animal care and use program at the University of Manitoba is in accordance with the standards of Good Animal Practice. The birds were purchased through Central Animal Care Services, University of Manitoba and were housed under standard conditions.

6.3.2 Preparation of chicken polychromatic erythrocytes and adenosine dialdehyde treatment

Anemia was induced in female Adult White Leghorn chickens by phenylhydrazine treatment, and pooled cells were purified as described previously (20). Polychromatic erythrocytes were treated with and without 25 mM adenosine dialdehyde (Sigma), SAH hydrolase inhibitor, for either 0, 1 or 12 h in MEM Alpha (1X) (Life Technologies, cat#12571–063) in two biological experiments. Histones were acid-extracted at the indicated time points and resolved by 15% SDS PAGE gels and the resulting blots immunostained with antibodies against H3K4me3, H3R2me2s and H4R3me2a.

6.3.3 Chromatin immunoprecipitation assay and sequencing

ChIP-seq (H3R2me2s and H4R3me2a) library preparations were performed according to the NEB Next ChIP-seq library preparation protocol and were sequenced on the Illumina HiSeq 4000 System at Genome Quebec and on the 5500 SOLiD™ System in the Manitoba Next Generation Sequencing facility as described (26, 29). Sequencing data for ChIP-seq for H3R2me2s, H4R3me2a, H3K4me3 and H3K27ac (26, 28) and FAIRE-seq (27) are available from GEO under accession numbers GSE141015, GSE117595 and GSE145398. Previous ChIP-seq and FAIRE-seq were validated by ChIP and FAIRE assays in triplicate biological assays.

6.3.4 Bioinformatics analyses

ChIP-seq reads were mapped to version 6 of the *Gallus gallus* genome assembly (galGal6; November 2019 release) using bowtie2 v2.2.4 using default parameters for end-to-end mapping. Peak calling was performed using MACS2 v2.1.1. Peaks were annotated with the Ensembl GRCg6a (release 99) chicken gene annotation GTF file to define intron, intergenic, CDS exon, TSS, TTS and 5'/3' UTR exon positions. FAIRE-seq reads were similarly mapped against the reference genome (galGal6). The bam files were indexed and sorted using bedtools and samtools. All the sequenced bam files were processed with Partek Flow pipeline to generate the chromosome view tracks. Pie chart of genomic distribution and overlapping TSS plots and Venn diagrams were plotted with Partek flow peak analysis task with the galGal6 annotation model. Plots of CGI were generated using deepTools (v3.3.0) (same parameters as in (27)). CGI locations for galGal6 were downloaded from the UCSC Table Browser. Gene Ontology for the arginine modified expressed genes were performed using Panther classification system (30), v15 (released 2020-02-14).

6.4 Results

6.4.1 Association of PRMT1, PRMT5, H4R3me2a and H3R2me2s with transcribed chromatin

H3R2me2s and H4R3me2a located to the transcriptionally active chromatin fraction (compartment A) in chicken polychromatic erythrocytes (26). To further delineate the distribution of H3R2me2s and H4R3me2a, ChIP-seq was done on two sequencing platforms (5500 SOLiD™ System and Illumina HiSeq 4000 System) (26). Although the location of the H3R2me2s or H4R3me2a peaks were comparable for both sequencing runs, the ChIP-seq peak definition was better for the SOLiD™ System. Hence, the data analyses shown next are from the H3R2me2s and H4R3me2a ChIP-seq from the SOLiD™ System.

H3R2me2s and H4R3me2a distributions were similar among the genomic elements, being greatest at introns and intergenic regions (Figure 6.1). The transcriptome data for the galGal6 polychromatic RNA-seq were compared against the ChIP-seq data for H3R2me2s and H4R3me2a to identify the expressed genes enriched in both arginine marks (5247). The Venn diagram in Figure 6.2A categorized the genes into expressed genes enriched in both arginine marks (5247), expressed genes but not with any arginine marks (865), expressed and only marked by H3R2me2s (381), and expressed and only marked by H4R3me2a (339). The 5247 genes with both the arginine marks were further analyzed to outline the top biological processes and molecular functions

associated as shown in the Gene Ontology (Figure 6.2B, C). The bar plot represents the significantly enriched biological processes indicating the importance of arginine marks in regulation of mostly cellular processes in the *Gallus gallus* genome. This finding confirmed the previously well established involvement of PRMT1 and PRMT5 in regulating diverse cellular processes (31). The molecular function analysis also showed the most enrichment in binding and catalytic activity which is consistent with the role of the PRMT enzymes (32).

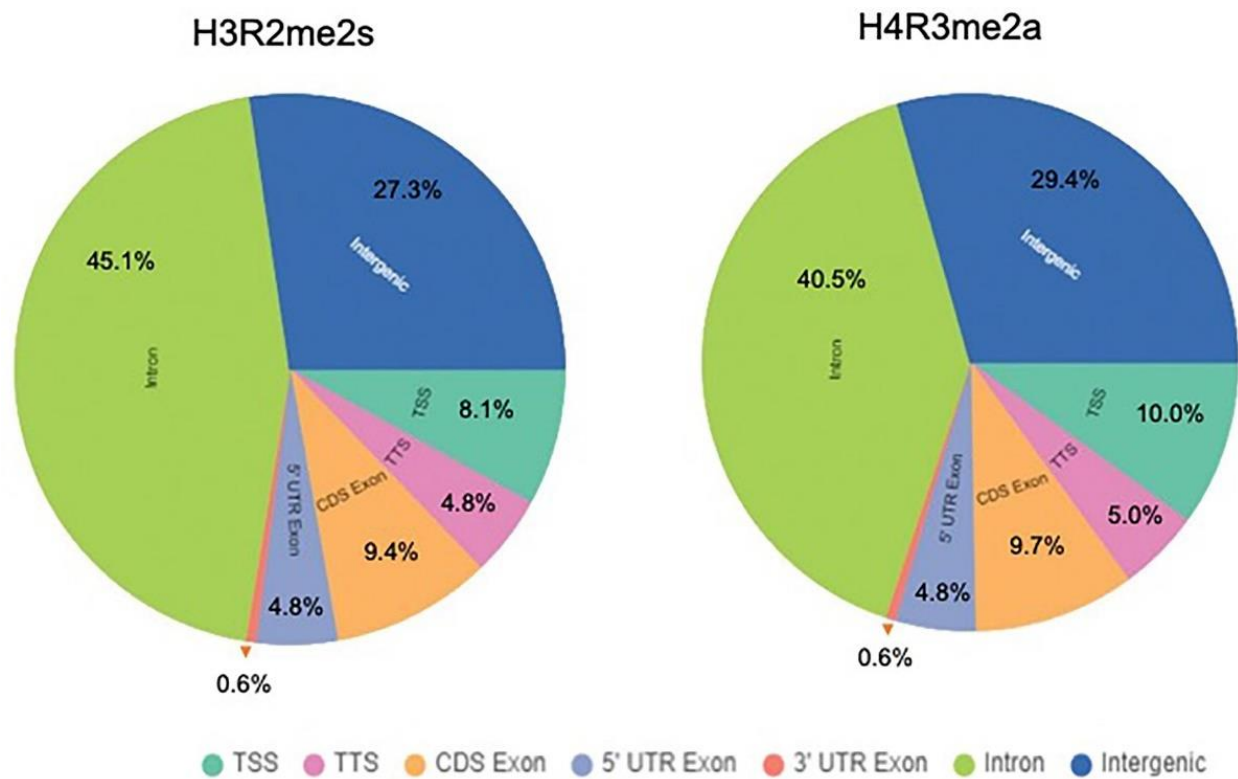


Figure 6.1: Genomic distribution of histone arginine methylated marks. Pie chart depicting genomic distribution of histone modifications H4R3me2a and H3R2me2s. Genomic enrichment for intergenic, exonic, intronic, TTS, TSS regions are color coded.

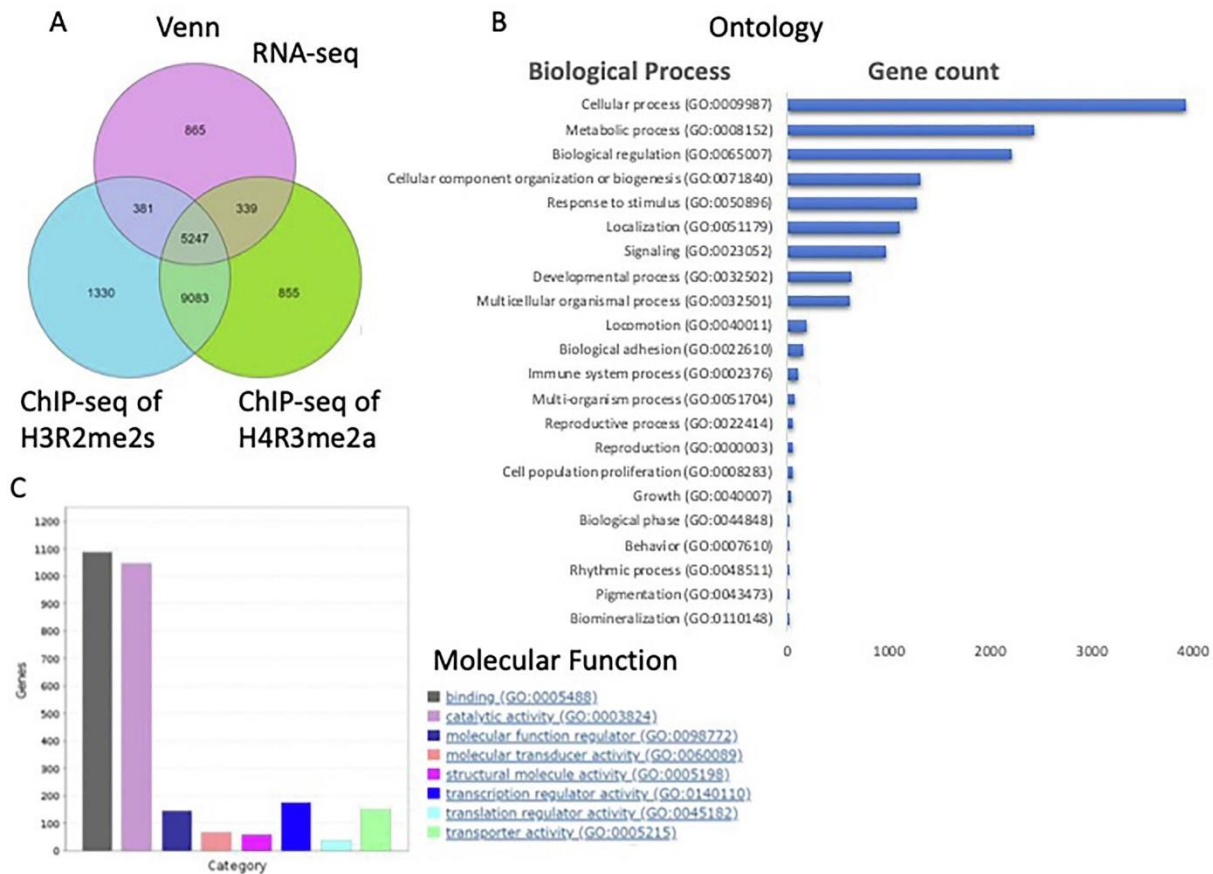


Figure 6.2: Categorization of expressed genes associated with histone arginine methylated marks. A) Venn diagram showing the overlap of expressed genes among separately processed RNA-seq, ChIP-seq peaks of H4R3me2a and H3R2me2s. The ChIP-seq peaks were annotated to the galGal6 annotation model build using Partek. B) Gene-ontology analysis of the 5247-arginine mark-regulated expressed genes identified in (A) are represented in a clustered bar plot. The Y-axis represents broad gene ontology (GO) categories for Biological process. C) GO terms of the molecular function of the overlapping genes are shown in the clustered column plot.

The genomic locations of these H3R2me2s and H4R3me2a were compared to that of H3K27ac and H3K4me3. Sample tracks of the distribution of these histone post-translational modifications (PTMs) on macrochromosomes and microchromosomes are shown.

Within the 100 kb region on macrochromosome 5 shown in Figure 6.3, expression of *FTH1* (ferritin heavy chain 1) and *RAB3IL1* (RAB3A interacting protein like 1) was greatest as suggested by the steady state transcript levels in chicken polychromatic erythrocytes (29). *FADS2* (fatty acid desaturase 2) was also expressed but at a low level. The *FTH1* TSS had a NDR positioned with a CGI surrounded on either side by intense peaks of H3K4me3, H3K27ac, H3R2me2s and H4R3me2a. The *RAB3IL1* TSS had a broader NDR positioned with the CGI and was surrounded by H3K4me3, which was more acute on the 5' end of the gene body, H3K27ac, H3R2me2s and H4R3me2a. The H3K27ac, H3R2me2s and H4R3me2a peaks were more pronounced upstream of the TSS. The *FADS2* TSS had a distinct NDR positioned with a CGI and H3K27ac, H3R2me2s and H4R3me2a peaks downstream of the TSS. An H3K4me3 peak was not observed. In addition to the TSS, H3R2me2s and H4R3me2a located in introns and intergenic regions. At chr5: 17,010,000–17,024,116, there was a region with nucleosome-free character, with multiple peaks of H3K27ac, H3R2me2s, H4R3me2a and less intense peaks for H3K4me3, which would be consistent with a regulatory element(s), possibly an enhancer or LCR. A similar chromatin organization was observed with the LCR of the β -globin (*HBBA*) gene (Figure 6.4) which had NDR at the DNase I hypersensitive sites (27) interspersed with nucleosomes modified by H3K27ac, H3R2me2s and H4R3me2a. The four PTMs were present throughout the 35-kb β -

globin domain. H4R3me2a and H3K27ac had intense peaks at the *HBBA* promoter, while an intense H3K4me3 peak was observed immediately downstream of the *HBBA* TSS. The *HBBA* enhancer had intense peaks of H4R3me2a and H3K27ac.

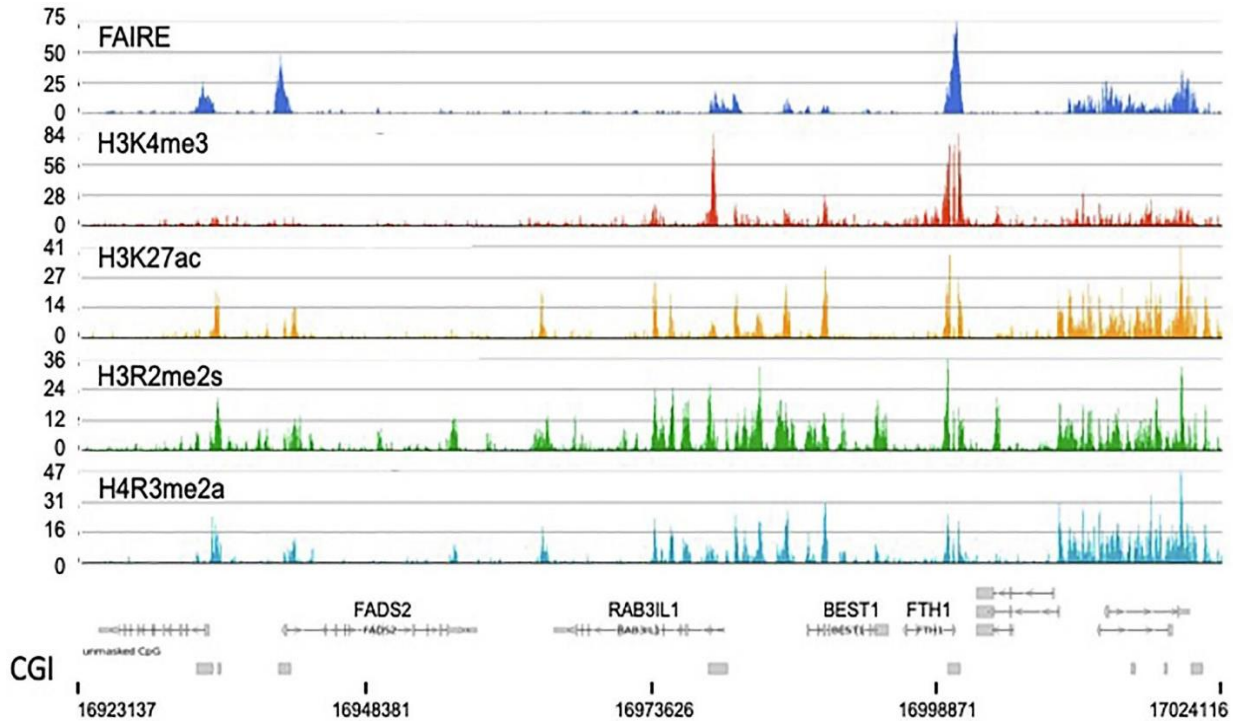


Figure 6.3: Chromatin profile along a 100 kb region of macrochromosome 5 covering highly expressed *FTH1* and *RAB3IL1* genes. Partek chromosome view snapshot detailing the position of the FAIRE-seq peaks and H3K4me3, H3K27ac, H4R3me2a and H3R2me2s ChIP-seq peaks. Positions of CpG islands (CGI) are indicated. Transcripts (with exons as boxes) and direction of transcription are depicted.

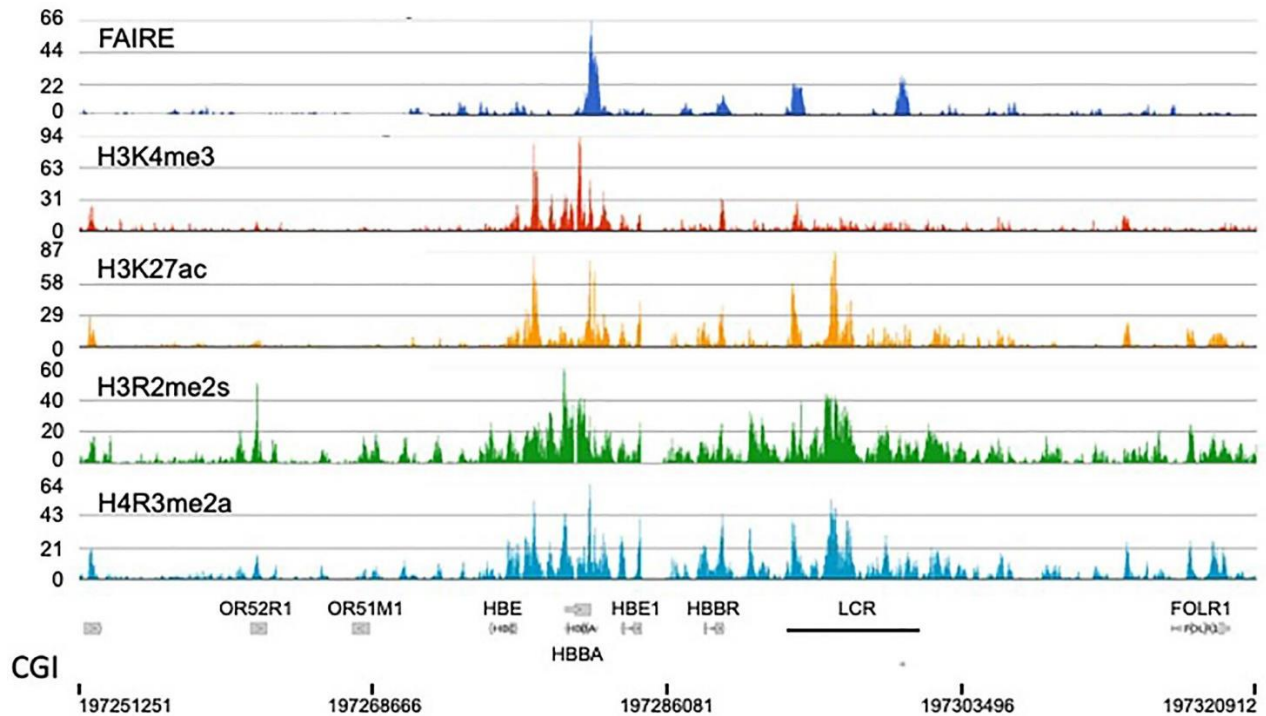


Figure 6.4: Chromatin profile along a 70 kb region of macrochromosome 1 covering highly expressed β -globin (*HBBA*) gene and β -globin domain. As described in Figure 6.3.

A 145 kb region on microchromosome 11 is shown in Figure 6.5. Within this region, the transcript levels from the *PRMT7* (protein arginine methyltransferase 7) and *TLR21* (toll-like receptor 21) genes were low, with expression of the other genes in this region not being observed (29). The NDR, identified by FAIRE-seq peaks, generally marked the position of the TSS and colocalized with a CGI (27). Also localizing with the CGIs were the four histone PTMs. The *PRMT7* gene, for example, had the four PTMs around the TSS with a CGI. Figure 6.6 shows that there was a strong association of the four marks with the CGIs. However, the distribution of H4R3me2a at the CGI differs from that of H3R2me2s. With the H4R3me2a peak centered at the

CGI, while H3R2me2s peaks were on both sides of the CGI. We had previously reported that CGIs were a common feature of active promoters in chicken erythroid cells (27). Further, chicken genes often have promoters with a CGI and such CGI marked promoters are with genes involved in transcriptional regulation and development (33).

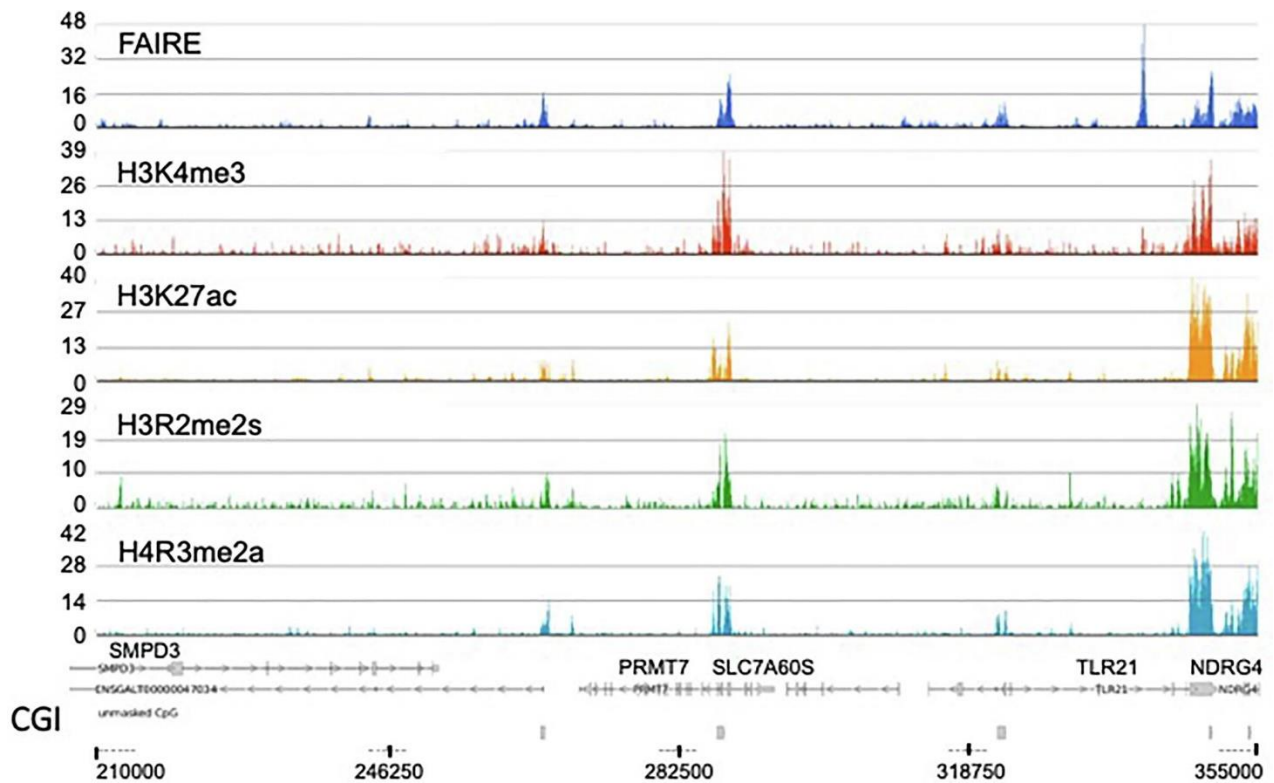


Figure 6.5: Chromatin profile along a 145 kb region of microchromosome 11 covering the weakly expressed *PRMT7* gene. As described in Figure 6.3.

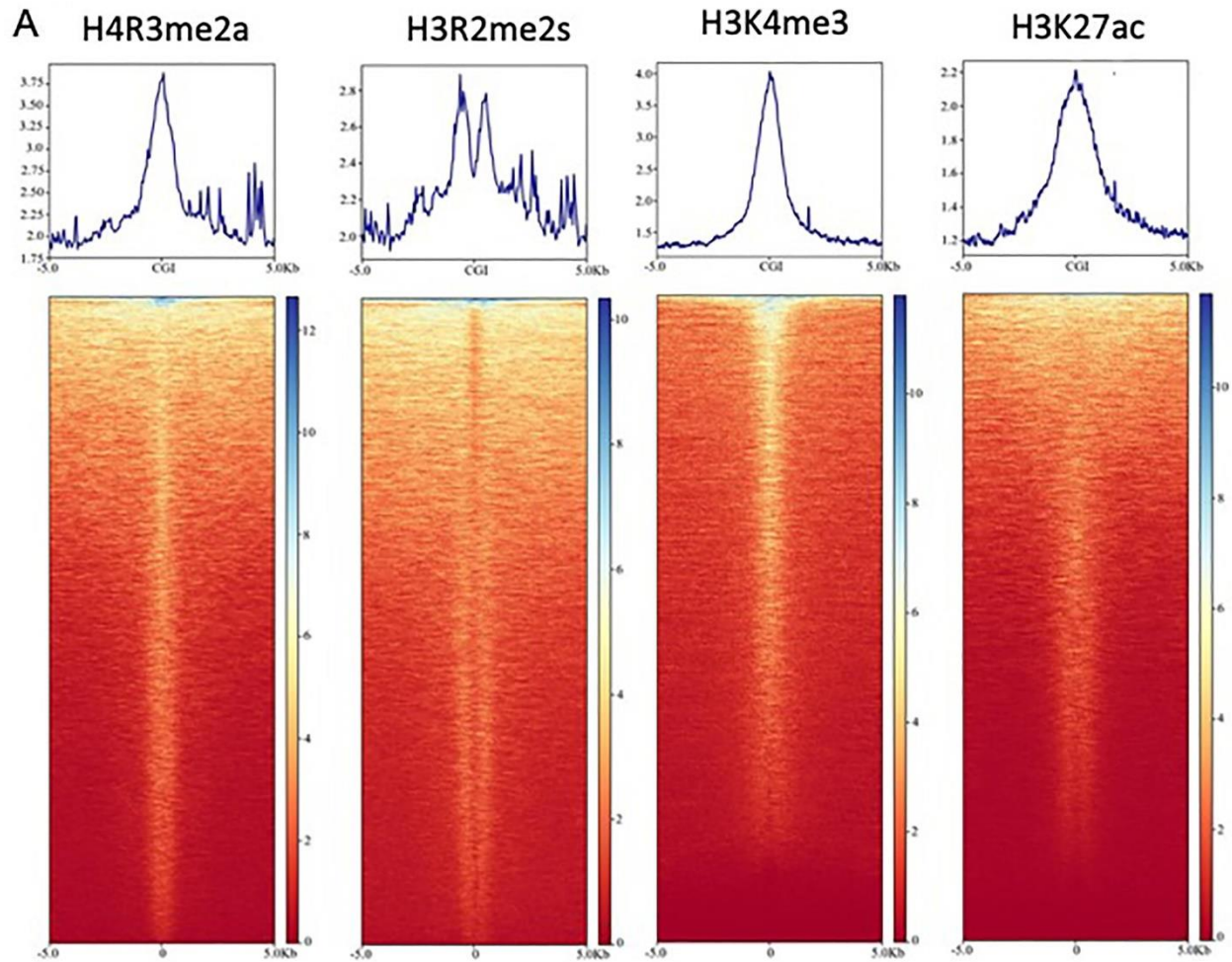


Figure 6.6. Profile of H3K4me3, H3K27ac, H4R3me2a and H3R2me2s around CpG islands. Histone mark enrichment distribution near CpG islands (± 5 kb from center of islands).

Analyses of the distribution of H3K4me3, H3R2me2s and H4R3me2a around the TSS showed H3K4me3 had the commonly observed distribution of being more abundant downstream of the TSS than upstream of the TSS (34–36) (Figure 6.7A). In contrast, H3R2me2s and H4R3me2a were more pronounced upstream of the TSS. H3K27ac was present equally on either side of the TSS of expressed genes similar to that reported for human cells (37) (analyses not

shown). The distribution of H3R2me2s and H4R3me2a relative to the TSS was further analyzed relative to the expression of polychromatic erythrocyte genes which were divided into 5 groups (29) (Supplementary Figure 6.1). H3R2me2s was strong on either side of the TSS but declining at the TSS for highly expressed genes (top 20%). H4R3me2a, in contrast, was most intense upstream of the TSS and was intense at the TSS for highly expressed genes.

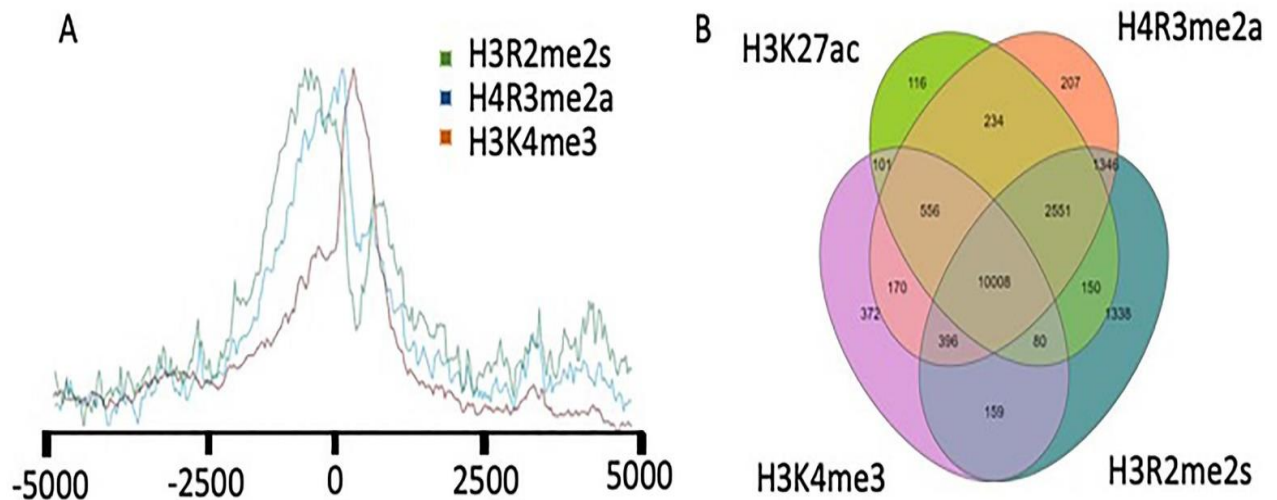


Figure 6.7. Location of histone arginine methylated marks relative to the TSS and other histone PTMs. A) Overlapped TSS plots of histone PTMs spanning 5 kb on each side. B) Correlation between histone modifications. Overlaps between H3K4me3, H3K27ac, H3R3me2a and H3R2me2s peaks are shown as a Venn diagram.

We recently reported that polychromatic erythrocytes express genes involved in innate immunity and that several of these genes had an atypical chromatin structure with NDR and

modified nucleosomes being intermingled throughout the gene body. One of these genes was *NFKB2*, nuclear factor kappa B subunit 2, located on macrochromosome 6 (28) . In addition to H3K4me3 and H3K27ac, the *NFKB2* gene body had nucleosomes with H3R2me2s and H4R3me2a (Supplementary Figure 6.2).

In summary, H3R2me2 and H4R3me2a are associated with transcriptionally active genes at the TSS and introns. Both marks are present in the intergenic space often with H3K27ac and are associated with regulatory regions such as LCRs.

6.4.2 Combinatory patterns of histone PTMs

Co-occurring histone PTM (H3K4me3, H3K27ac, H3R2me2s and H4R3me2a) peaks were identified by counting peak overlaps using bedtools. The overlap between peaks as presented in a Venn diagram is shown in Figure 6.7B. For the four PTMs, 10,008 peaks overlapped. A greater number of ChIP-seq peaks overlapped for H3R2me2s and H4R3me2a (14,301). Scatterplots further show the strong correlation between these histone PTMs (Supplementary Figure 6.3A-C). Correlation coefficients for H3R2me2s versus H4R3me2a, H3R2me2s versus H3K27ac, and H4R3me2a versus H3K27ac were all strong ($r = 0.83, 0.74, 0.93$, respectively).

In chicken erythrocytes, there are two variants for H3, H3.2 and H3.3. H3.2 is a replication-dependent histone variant, while H3.3 is a replication-independent, replacement histone variant (38) . Histones were resolved on AUT polyacrylamide gels, which separate histones according to

size, charge and hydrophobicity (39), and immunoblotted with antibodies against H3K4me3, H3R2me2s and H3.3 (Figure 6.8A). H3.3 migrates faster than H3.2 on the AUT gels. Both H3.2 and H3.3 had the K4me3 and R2me2s PTMs.

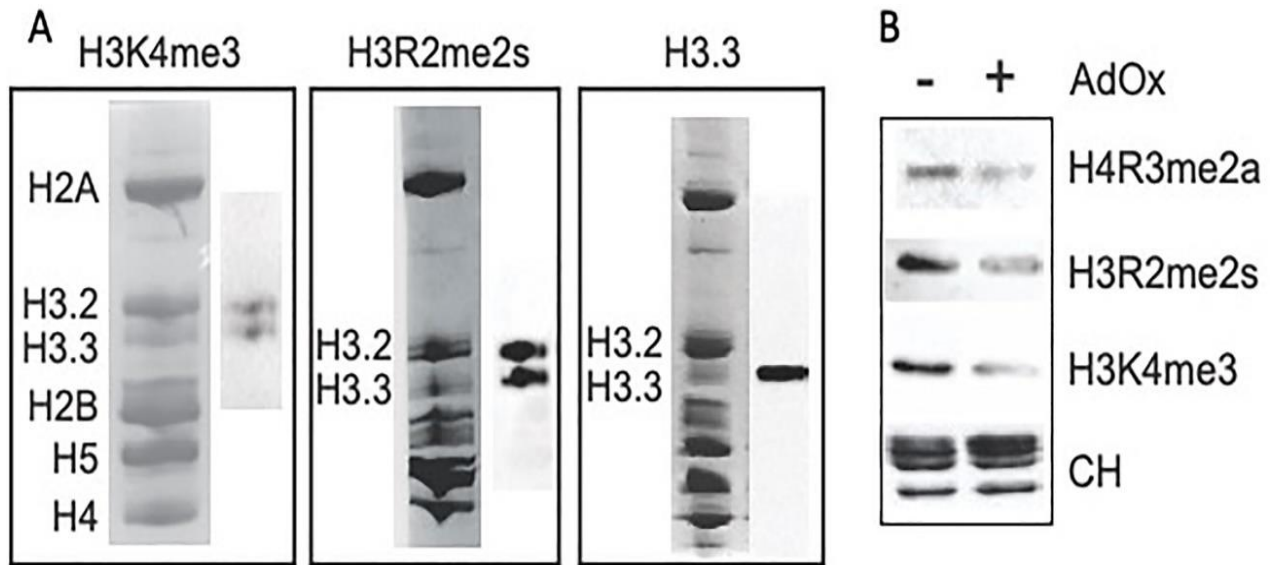


Figure 6.8: Arginine methylation of histone H3 variants and turnover of methylated histones. A) Acid-extracted histones were electrophoresed on a 15% acid-urea-Triton X-100 polyacrylamide gel, stained with Coomassie Blue (left lanes) or immunostained with the indicated antibodies (right lanes). B) Turnover of methylated histone. Acid-extracted histones isolated from polychromatic erythrocytes treated or not with adenosine dialdehyde (AdOx) were resolved by 15% SDS PAGE and the blots immunostained with the indicated antibodies.

In summary, both histone H3 variants are modified with R2me2s, and we find a strong correlation between H4R3me2a and H3K27ac.

6.4.3 Dynamic methylation of histones

We had previously reported that inhibition of histone methyltransferase activity was achieved by using adenosine dialdehyde, an inhibitor of SAH, leading to a decrease in the concentration of SAM required for methyltransferase activity (40, 41) . To determine whether the levels of H3R2me2s and H4R3me2a would decrease once SAM concentrations were lowered, we treated freshly isolated chicken polychromatic erythrocytes with adenosine dialdehyde for 12 h and detected the methylated histones by immunoblotting. After 12 h, levels of H4R3me2a, H3R2me2s and H3K4me3 were decreased compared to control levels (Figure 6.8B). There was little change in the levels of these three histone PTMs after 1-hour treatment (data not shown). These observations show that there are mechanisms in place to reduce the levels of these histone PTMs once the methyltransferase activity is diminished.

6.5 Discussion

In this work, we present the genomic landscape of H3R2me2s and H4R3me2a in chicken polychromatic erythrocytes. H3R2me2s and H4R3me2a located to compartment A and were often found together at introns, intergenic regions, exons and TSS in chicken polychromatic erythrocytes. At the TSS, there was a distinction between H3R2me2s and H4R3me2a, with

H3R2me2s being pronounced more upstream of the TSS than was H4R3me2a. These two histone PTMs also differed in their distribution relative to CGIs, with H4R3me2a locating at the CGI and with H3R2me2s being pronounced on either side of the CGI.

Previous studies have demonstrated a close link between H3R2me2s and H3K4me3 (42). K4me3, K4me2 and K27ac are on the same H3 tail as R2me2s (4, 26, 42, 43). Yuan et al observed H3R2me2s on both sides of the TSS, being more intense downstream of the TSS of highly expressed murine genes (42). In human B cells, H3R2me2s was located primarily upstream of the TSS(4). Our study with chicken polychromatic erythrocytes found H3R2me2s to be on both sides of the TSS with a tendency to be more prominent upstream of the TSS of highly expressed genes. A similar distribution of H3R2me2s was found at the CGIs.

H3R2me2s and H4R3me2a differed in their association with H3K27ac, with H4R3me2a having the stronger association. In our preliminary mass spectrometry analysis of the PRMT1 associated proteins, we observed an interaction between PRMT1 with p300 in chicken erythrocytes consistent with PRMT1/CBP/p300 interactions in chicken erythrocytic cells and human cells (13, 44). The interaction between PRMT1 and CBP/p300 explains the observed strong correlation between H4R3me2a and H3K27ac. These observations are consistent with PRMT1 recruiting p300/CBP which in turn acetylates H3 at K27, a mark of active promoters, active enhancers and LCRs. Further, nucleosomes with H4R3me2s potentiate CBP/p300 acetylation activity (13).

PRMT1 and PRMT5 were present in the active DNA enriched chromatin fraction S₁₅₀ and low ionic strength insoluble fraction P_E (26) . We had previously reported that histone methyltransferase activity was greatest in these fractions (45) . The P_E fraction, which contains the nuclear substructure, is also rich in lysine acetyltransferase and histone deacetylase activities (46).

PRMT1 is required to establish and maintain the transcriptionally active chromatin state (12, 47) . Interaction between the LCR and β -globin promoter in the β -globin gene domain was prevented when PRMT1 was knocked down. This observation suggests that PRMT1 has roles in activation of enhancers via the enzyme's recruitment of CBP/p300 and H3K27 acetylation and in interaction of the enhancer/super enhancer/LCR with the target promoter, a step that can take place before enhancer activation (13) . Important to initiating enhancer activation would be the recruitment of PRMT1. Among the mechanisms recruiting PRMT1 to regulatory sites are transcription factors like USF1 which bind to PRMT1 (48) (Figure 6.9). Our preliminary mass spectrometry results showed that transcription factors Sp1 and Sp3 were associated with PRMT1, but these interactions need to be validated. Further, we have preliminary evidence that PRMT1 interacts weakly with PRMT5 and CXXC1 (CFP1). CXXC1 is a component of the SETD1A (SET domain containing 1A, histone lysine methyltransferase). CXXC1 may aid in the recruitment of PRMT1 and PRMT5 to CGIs found at many of the active erythroid promoters (27) ; however, these interactions require further study. A recent study reports that CXXC1 can bind to genomic regions that do not have CGIs which increases the range by which CXXC1 distributes PRMT1 in the genome (49) . Lastly, PRMT1 (and 5) association with the elongating RNAPII may account for location of H4R3me2a and H3R2me2s along transcribed gene bodies (26).

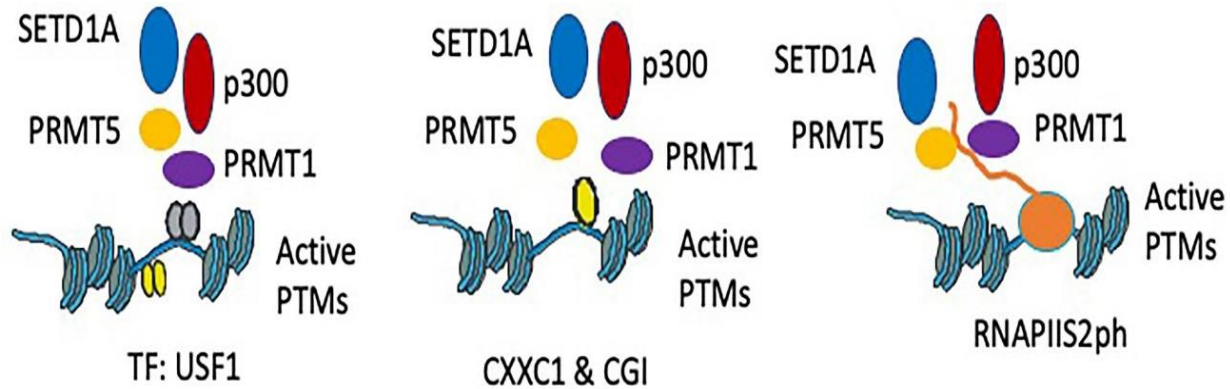


Figure 6.9: Models for recruitment of PRMT1 to active genes. A) Transcription factors such as USF1 recruit PRMT1 complex to regulatory regions of genes. The PRMT1 complex consists of PRMT5, SETD1A and CBP/p300. B) CXXC1 (CFP1) recruits the PRMT1 complex to CGIs and non-CGI regions of the genome. C) PRMT1 and PRMT5 associate with RNAPIIS2ph to modify nucleosomes along the coding region of genes.

In summary, our results show that H3R2me2s and H4R3me2a located to regulatory regions of the genome and marked genes that have an atypical chromatin structure. PRMT1-mediated H4R3me2a is involved in CBP/p300 recruitment and H3K27 acetylation. Multiple mechanisms are involved in the recruitment of PRMT1 to specific genomic sites (Figure 6.9). These activities will contribute to PRMT1 role in gene expression and cellular differentiation. Further, we propose that PRMT1 would be a critical enzyme impacted when SAM levels are reduced in metabolic disorders (50).

6.6 Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

6.7 Acknowledgements

This work was supported by a grant from the Natural Sciences and Engineering Research Council of Canada (RGPIN-2017-05927) to J.R.D., by a Canada Research Chair (to J.R.D.) and a Research Manitoba/Cancer Care Manitoba Master's Studentship, 2019 (to T.H.B.). We thank Geneviève Delcuve for reviewing the manuscript.

6.8 References

- [1] Jahan, S., and Davie, J. R. (2015) Protein arginine methyltransferases (PRMTs): Role in chromatin organization. *Advances in Biological Regulation* 57, 173–184.
- [2] Chen, D., Ma, M., Hong, H., Koh, S. S., Huang, S. M., et al. (1999) Regulation of transcription by a protein methyltransferase. *Science* 284, 2174–2177.
- [3] Lorton, B. M., and Shechter, D. (2019) Cellular consequences of arginine methylation. *Cellular and Molecular Life Sciences* 76, 2933–2956.
- [4] Migliori, V., Müller, J., Phalke, S., Low, D., Bezzi, M., et al. (2012) Symmetric dimethylation of H3R2 is a newly identified histone mark that supports euchromatin maintenance. *Nature Structural and Molecular Biology* 19, 136–145.
- [5] Girardot, M., Hirasawa, R., Kacem, S., Fritsch, L., Pontis, J., et al. (2014) PRMT5-mediated histone H4 arginine-3 symmetrical dimethylation marks chromatin at G+C-rich regions of the mouse genome. *Nucleic Acids Research* 42, 235–248.
- [6] Min, Z., Xiaomeng, L., Zheng, L., Yangge, D., Xuejiao, L., et al. (2019) Asymmetrical methyltransferase PRMT3 regulates human mesenchymal stem cell osteogenesis via miR-3648. *Cell Death and Disease* 10, 581.
- [7] Serizay, J., and Ahringer, J. (2018) Genome organization at different scales: nature, formation and function. *Current Opinion in Cell Biology* 52, 145–153.
- [8] Fishman, V., Battulin, N., Nuriddinov, M., Maslova, A., Zlotina, A., et al. (2019) 3D organization of chicken genome demonstrates evolutionary conservation of topologically

- associated domains and highlights unique architecture of erythrocytes' chromatin. *Nucleic Acids Research* 47, 648–665.
- [9] Lieberman-Aiden, E., Van Berkum, N. L., Williams, L., Imakaev, M., Ragozy, T., et al. (2009) Comprehensive mapping of long-range interactions reveals folding principles of the human genome. *Science* 326, 289–293.
- [10] Fortin, J. P., and Hansen, K. D. (2015) Reconstructing A/B compartments as revealed by Hi-C using long-range correlations in epigenetic data. *Genome Biology* 16, 180.
- [11] Bascom, G. D., and Schlick, T. (2018) Chromatin Fiber Folding Directed by Cooperative Histone Tail Acetylation and Linker Histone Binding. *Biophysical Journal* 114, 2376–2385.
- [12] Huang, S., Litt, M., and Felsenfeld, G. (2005) Methylation of histone H4 by arginine methyltransferase PRMT1 is essential in vivo for many subsequent histone modifications. *Genes and Development* 19, 1885–1893.
- [13] Li, X., Hu, X., Patel, B., Zhou, Z., Liang, S., et al. (2010) H4R3 methylation facilitates β -globin transcription by regulating histone acetyltransferase binding and H3 acetylation. *Blood* 115, 2028–2037.
- [14] Zhu, L., He, X., Dong, H., Sun, J., Wang, H., et al. (2019) Protein arginine methyltransferase 1 is required for maintenance of normal adult hematopoiesis. *International Journal of Biological Sciences* 15, 2763–2773.
- [15] McQueen, H. A., Siriaco, G., and Bird, A. P. (1998) Chicken microchromosomes are hyperacetylated, early replicating, and gene rich. *Genome Research* 8, 621–630.
- [16] Smith, J., and Burt, D. W. (1998) Parameters of the chicken genome (*Gallus gallus*). *Animal Genetics* 29, 290–294.

- [17] Habermann, F. A., Cremer, M., Walter, J., Kreth, G., Von Hase, J., et al. (2001) Arrangements of macro- and microchromosomes in chicken cells. *Chromosome Research* 9, 569–584.
- [18] Hillier, L. W., Miller, W., Birney, E., Warren, W., Hardison, R. C., et al. (2004) Sequence and comparative analysis of the chicken genome provide unique perspectives on vertebrate evolution. *Nature* 432, 695–716.
- [19] Masabanda, J. S., Burt, D. W., O'Brien, P. C. M., Vignal, A., Fillon, V., et al. (2004) Molecular Cytogenetic Definition of the Chicken Genome: The First Complete Avian Karyotype. *Genetics* 166, 1367–1373.
- [20] Delcuve, G. P., and Davie, J. R. (1989) Chromatin structure of erythroid-specific genes of immature and mature chicken erythrocytes. *Biochemical Journal* 263, 179–186.
- [21] Hendzel, M. J., and Davie, J. R. (1990) Nucleosomal histones of transcriptionally active/competent chromatin preferentially exchange with newly synthesized histones in quiescent chicken erythrocytes. *Biochemical Journal* 271, 67–73.
- [22] Locklear, L., Ridsdale, J. A., Bazett-Jones, D. P., and Davie, J. R. (1990) Ultrastructure of transcriptionally competent chromatin.
- [23] Ridsdale, J. A., Hendzel, M. J., Delcuve, G. P., and Davie, J. R. (1990) Histone acetylation alters the capacity of the H1 histones to condense transcriptionally active/competent chromatin. *Journal of Biological Chemistry* 265, 5150–5156.
- [24] Li, W., Nagaraja, S., Delcuve, G. P., Hendzel, M. J., and Davie, J. R. (1993) Effects of histone acetylation, ubiquitination and variants on nucleosome stability. *Biochemical Journal* 296, 737–744.

- [25] Sun, J.-M., Chen, H. Y., Espino, P. S., and Davie, J. R. (2007) Phosphorylated serine 28 of histone H3 is associated with destabilized nucleosomes in transcribed chromatin. *Nucleic Acids Research* 35, 6640–6647.
- [26] Jahan, S. (2017) Characterization of transcriptionally active chicken erythrocyte chromatin. University of manitoba, 478.
- [27] Jahan, S., Beacon, T. H., He, S., Gonzalez, C., Xu, W., et al. (2019) Chromatin organization of transcribed genes in chicken polychromatic erythrocytes. *Gene* 699, 80–87.
- [28] Jahan, S., Beacon, T. H., Xu, W., and Davie, J. R. (2020) Atypical chromatin structure of immune-related genes expressed in chicken erythrocytes. *Biochemistry and Cell Biology* 98, 171–177.
- [29] Jahan, S., Xu, W., He, S., Gonzalez, C., Delcuve, G. P., et al. (2016) The chicken erythrocyte epigenome. *Epigenetics and Chromatin* 9, 1–11.
- [30] Mi, H., Muruganujan, A., Ebert, D., Huang, X., and Thomas, P. D. (2019) PANTHER version 14: more genomes, a new PANTHER GO-slim and improvements in enrichment analysis tools. *Nucleic Acids Research* 47, 419–426.
- [31] Guccione, E., and Richard, S. (2019) The regulation, functions and clinical relevance of arginine methylation. *Nature Reviews Molecular Cell Biology* 20, 642–657.
- [32] Bedford, M. T., and Clarke, S. G. (2009) Protein Arginine Methylation in Mammals: Who, What, and Why. *Molecular Cell* 33, 1–13.
- [33] Abe, H., and Gemmell, N. J. (2014) Abundance, arrangement, and function of sequence motifs in the chicken promoters. *BMC Genomics* 15, 1–12.
- [34] Barski, A., Cuddapah, S., Cui, K., Roh, T. Y., Schones, D. E., et al. (2007) High-Resolution

- Profiling of Histone Methylations in the Human Genome. *Cell* 129, 823–837.
- [35] Bieberstein, N. I., Oesterreich, F. C., Straube, K., and Neugebauer, K. M. (2012) First exon length controls active chromatin signatures and transcription. *Cell Reports* 2, 62–68.
- [36] Davie, J. R., Xu, W., and Delcuve, G. P. (2015) Histone H3K4 trimethylation: Dynamic interplay with pre-mRNA splicing¹. *Biochemistry and Cell Biology* 94, 1–11.
- [37] Wang, Z., Zang, C., Rosenfeld, J. A., Schones, D. E., Barski, A., et al. (2008) Combinatorial patterns of histone acetylations and methylations in the human genome. *Nature Genetics* 40, 897–903.
- [38] Tvardovskiy, A., Schwämmle, V., Kempf, S. J., Rogowska-Wrzesinska, A., and Jensen, O. N. (2017) Accumulation of histone variant H3.3 with age is associated with profound changes in the histone methylation landscape. *Nucleic Acids Research* 45, 9272–9289.
- [39] Delcuve, G. P., and Davie, J. R. (1992) Western blotting and immunochemical detection of histones electrophoretically resolved on acid-urea-Triton- and sodium dodecyl sulfate-polyacrylamide gels. *Analytical Biochemistry* 200, 339–341.
- [40] Bartel, R. L., and Borchardt, R. T. (1984) Effects of adenosine dialdehyde on S-adenosylhomocysteine hydrolase and S-adenosylmethionine-dependent transmethylation in mouse L929 cells. *Molecular Pharmacology* 25, 418–424.
- [41] Chuang, C. Y., Chang, C. P., Lee, Y. J., Lin, W. L., Chang, W. W., et al. (2017) PRMT1 expression is elevated in head and neck cancer and inhibition of protein arginine methylation by adenosine dialdehyde or PRMT1 knockdown downregulates proliferation and migration of oral cancer cells. *Oncology Reports* 38, 1115–1123.
- [42] Yuan, C. C., Matthews, A. G. W., Jin, Y., Chen, C. F., Chapman, B. A., et al. (2012) Histone

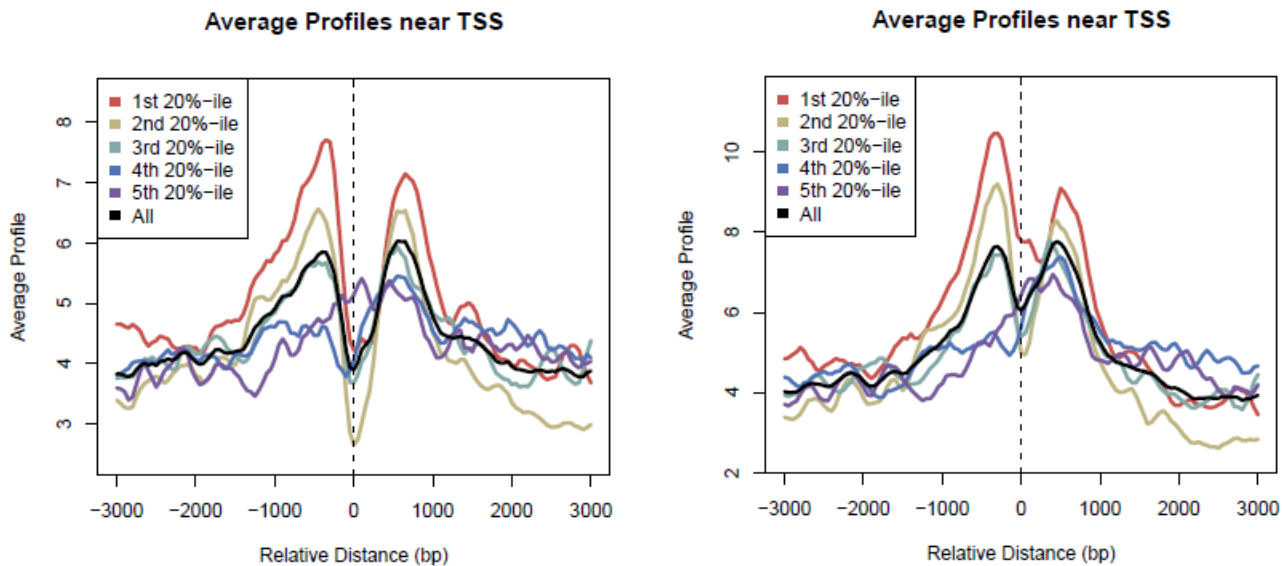
- H3R2 Symmetric Dimethylation and Histone H3K4 Trimethylation Are Tightly Correlated in Eukaryotic Genomes. *Cell Reports* 1, 83–90.
- [43] Binda, O. (2013) On your histone mark, SET, methylate! *Epigenetics* 8, 457–463.
- [44] An, W., Kim, J., and Roeder, R. G. (2004) Ordered cooperative functions of PRMT1, p300, and CARM1 in transcriptional activation by p53. *Cell* 117, 735–748.
- [45] Hendzel, M. J., and Davies, J. R. (1989) Distribution of Methylated Histones and Histone Methyltransferases in Chicken Erythrocyte Chromatin*. *Journal of Biological Chemistry* 264, 19208–19214.
- [46] Hendzels, M. J., Sun, J.-M., Chen, H. Y., Rattners, J. B., and Davie, J. R. (1994) Histone Acetyltransferase Is Associated with the Nuclear Matrix". *The Journal of biological chemistry* 269, 22894–22901.
- [47] Zhao, Y., Lu, Q., Li, C., Wang, X., Jiang, L., et al. (2019) PRMT1 regulates the tumour-initiating properties of esophageal squamous cell carcinoma through histone H4 arginine methylation coupled with transcriptional activation. *Cell Death and Disease* 10, 359.
- [48] Huang, S., Li, X., Yusufzai, T. M., Qiu, Y., and Felsenfeld, G. (2007) USF1 Recruits Histone Modification Complexes and Is Critical for Maintenance of a Chromatin Barrier. *Molecular and Cellular Biology* 27, 7991–8002.
- [49] Van De Lagemaat, L. N., Flenley, M., Lynch, M. D., Garrick, D., Tomlinson, S. R., et al. (2018) CpG binding protein (CFP1) occupies open chromatin regions of active genes, including enhancers and non-CpG islands *06 Biological Sciences 0601 Biochemistry and Cell Biology 06 Biological Sciences 0604 Genetics. Epigenetics and Chromatin* 11, 59.
- [50] Shen, W., Gao, C., Cueto, R., Liu, L., Fu, H., et al. (2020) Homocysteine-methionine cycle

is a metabolic sensor system controlling methylation-regulated pathological signaling.
Redox Biology 28, 101322.

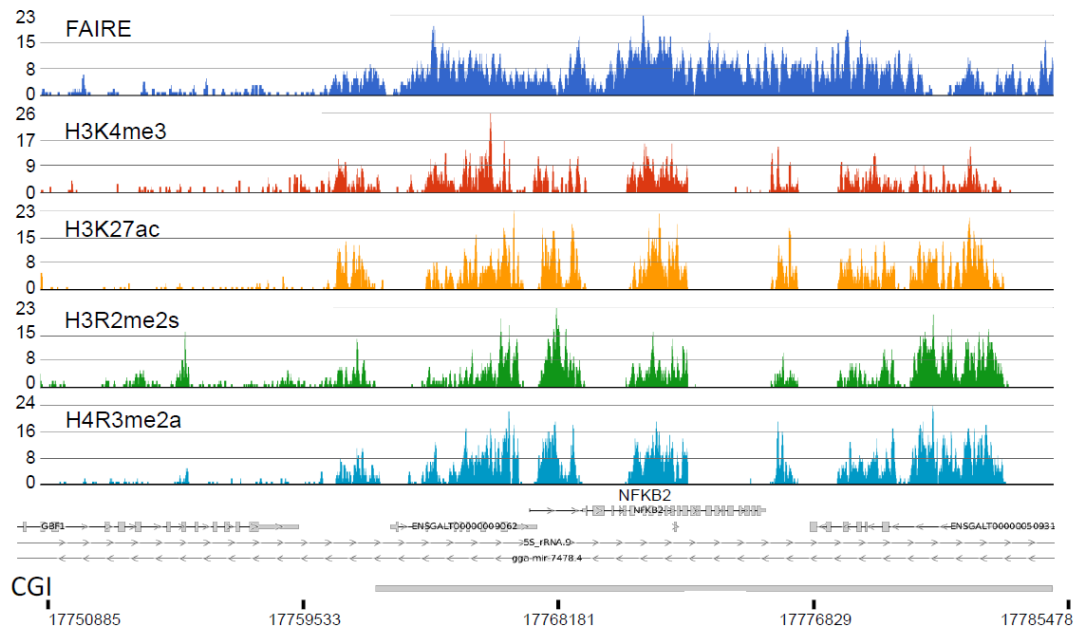
6.9 Supplementary Data

H3R2me2s

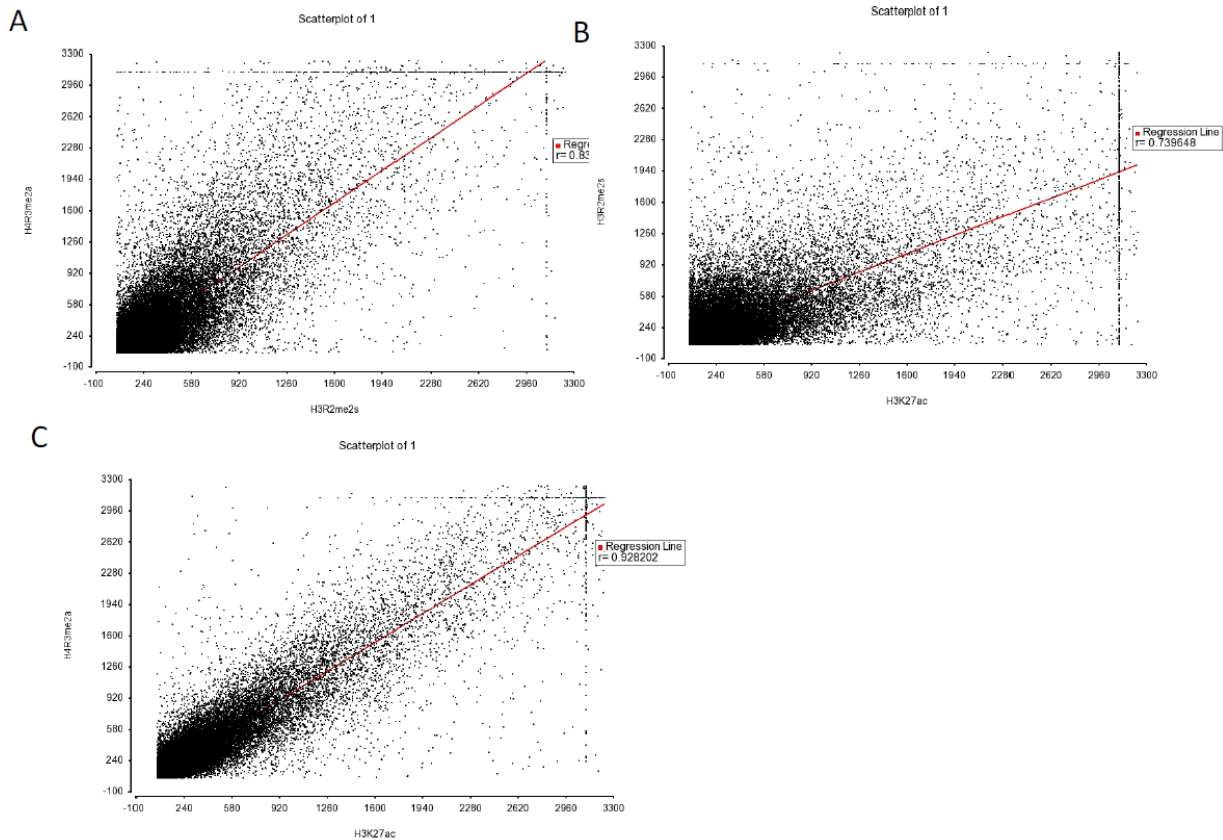
H4R3me2a



Supplementary file 6.1. H3R2me2s and H4R3me2a profiles as a function of gene expression. The ChIP-seq peak profiling around TSS were generated and displayed using the cis-regulatory element annotation system (CEAS) program. H3R2me2s and H4R3me2a TSS-centered profiles were divided into quintile classes based on gene expression levels (Jahan et al., 2016). All 5479 genes from the galGal3 RefSeq database were ranked from top to bottom, according to their level of expression. These genes were profiled for H3R2me2s and H4R3me2a spanning 3 kb on each side of TSS.



Supplementary Figure 6.2: Chromatin profile along a 35 kb region of macrochromosome 6 covering the weakly expressed NFKB2 gene. As described in Figure 6.3.



Supplementary File 6.3: Correlation between histone modifications. A - C) The common peak identifiers were generated by detecting overlapping peaks among H4R3me2a and H3R2me2s (A), H3R2me2s and H3K27ac (B) and H4R3me2a and H3K27ac (C) ChIP-seq experiments, using galGal3 as reference genome. Overlapping of peaks within 100 bp were merged by Homer software package. The Pearson correlations were calculated and plotted by Partek software. The calculated regression line is shown.

Chapter-7

Investigation of H3K4me3 broad domains in different cell types reveal importance of long chromatin signatures in human and chicken genomes

7.1 Introduction

Histone H3 trimethylated at lysine 4 (H3K4me3) is an active chromatin mark of expressed genes. The enrichment and location of H3K4me3 across the genome exhibits two distinct distribution patterns, which are conserved in yeast, plants, worms, flies, birds and mammals. The first, found for most active genes, have H3K4me3 restricted to around the first exon and promoter. In the second rarer distribution, H3K4me3 covers a substantial part of the gene body and regions upstream and downstream of the gene; the H3K4me3 broad domain signature. The first study on the broad H3K4me3 domain was reported in 2014 (1). It is known as the hallmark for cell-identity genes. H3K4me3 and its regulator (KMT and KDM) are known as the key players in neurodevelopmental disease (e.g. schizophrenia and autism) and memory formation (2). Analysis of H3K4me3 broad domains in human neural signalling genes revealed this as a therapeutic target for neurodevelopmental disease conserved in chimpanzee, macaque and mouse (3). In mouse oocytes, broad H3K4me3 mark development associated genes and their removal by KDM5A and KDM5B are required for normal zygotic genome activation and early embryo development (4). KDM4A-mediated H3K9me3 demethylation regulates normal epigenetic reprogramming and pre-implantation development by preserving the H3K4me3 broad domain (5). Taken together, this signature is a regulator of transcription of cell-type specific and disease associated genes.

Super-enhancers are cluster of enhancers that function together as proximal or distal regulators of gene interaction and activation. They are known to modulate key oncogenic factors in multiple myeloma, lung cancer, and glioblastoma multiforme (6). Several studies have linked

H3K4me3 broad domain and super-enhancer to co-regulate tumor suppressor genes and oncogenes (6). In this study, we explore the co-localization of H3K4me3 and H3K27 broad domains and show that normal and cancer cell types have unique sets of genes marked with H3K4me3 broad chromatin domains, important in normal and disease functions in vertebrates.

7.2 Methods

To investigate the H3K4me3 broad domain features, I analyzed publicly available ChIP-seq data in the GEO database for a human triple negative breast cancer cell line (MDA-MB-231), human erythroleukemia cell line (K562) and chicken polychromatic erythrocytes. The bioinformatic pipeline to analyze the massive parallel sequencing data was built in Partek flow (details in chapter 2). Peak detection for both distributions (narrow and broad) were identified using MACS2 (Model-based Analysis of ChIP-Seq) peak caller and annotated with the reference genome annotation model to identify the expressed genes. The genes with the broad H3K4me3 domain signature chromatin have been mapped against other active histone marks and regulatory regions e.g. ChIP-seq of H3K27ac, H3R2me2s, H4R3me2a and FAIRE-seq and DNase I-seq data. Motif enrichment analysis of the de novo motif was done using MEME Suite. Plots were prepared using R.

7.3 Results

For broad domain investigation, we analyzed publicly available H3K4me3 ChIP-seq data for K562 cell, MDA-MB-231 cell and chicken polychromatic erythrocyte from GEO and ENCODE. Consistent with previous findings, most H3K4me3 peaks were stretched up to 1-2 kb around the TSS with the breadth limiting within exon 1 of most active genes. A small percentage of the H3K4me3 marks had a broader distribution around TSS extending on both 5' and 3' but mostly in the 5' to 3' direction going into the gene body. Some H3K4me3 domains were seen to cover the whole gene irrespective of gene length (Figure 7.1). The genes marked by such broad domain differed across cell types. Therefore, the genes marked with the H3K4me3 broad domain in MDA-MB-231 cells were not found in K562 cells showing cell type specificity of the genes marked with broad domains (results not shown). We then compared the human H3K4me3 signature to that of chicken (polychromatic erythrocyte) for vertebrate comparative analysis and saw the same distribution of H3K4me3, that is mostly narrow peaks and rarer broad regions (Figure 7.2). We also discovered that chicken erythroid genes involved in innate immunity had the H3K4me3 broad chromatin signature (7). Another consistent observation of the broad H3K4me3 domains distribution was that breadth of the broad domain had no correlation to gene expression level, ChIP-peak intensity and annotated promoter region position. In the three cell types, promoter free H3K4me3 domains were observed spanning across the gene body and intergenic regions.

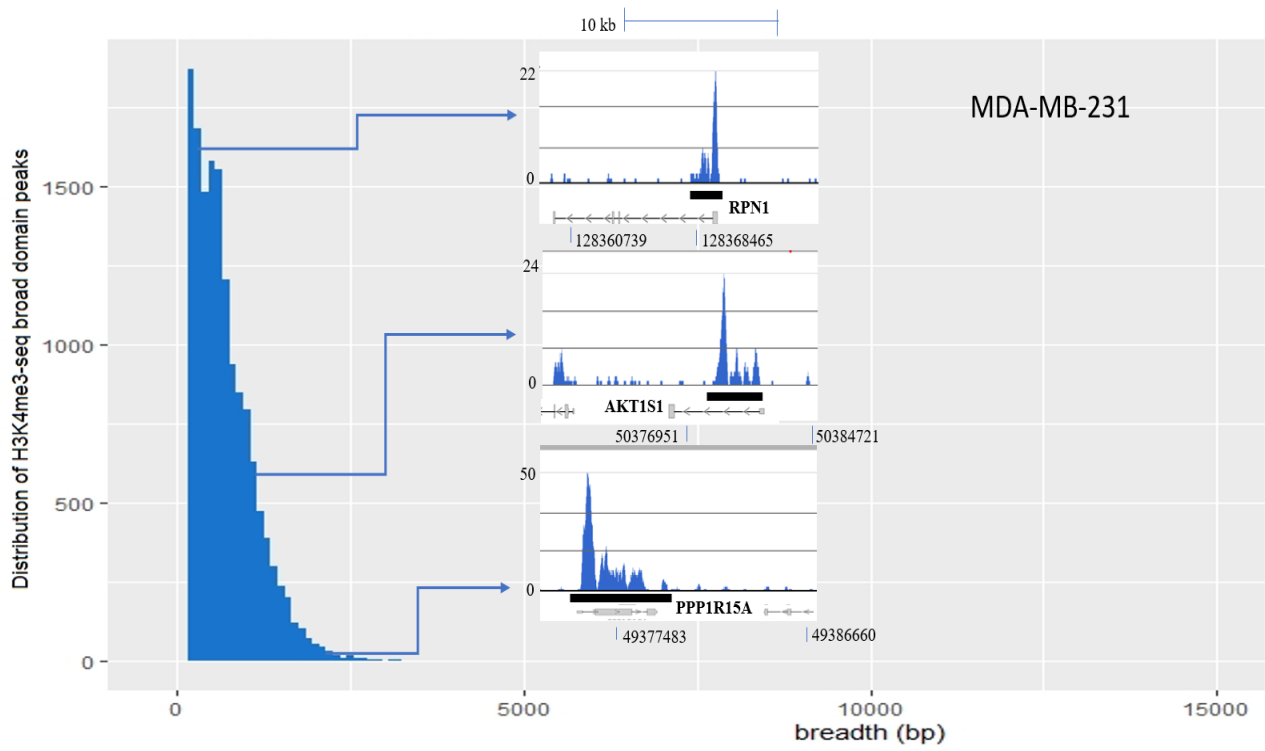


Figure 7.1: Breadth distribution of H3K4me3 ChIP-seq seq in MDA-MB-231 cells. The illustration shows the distinction between sharp and broad peak around TSS of active genes.

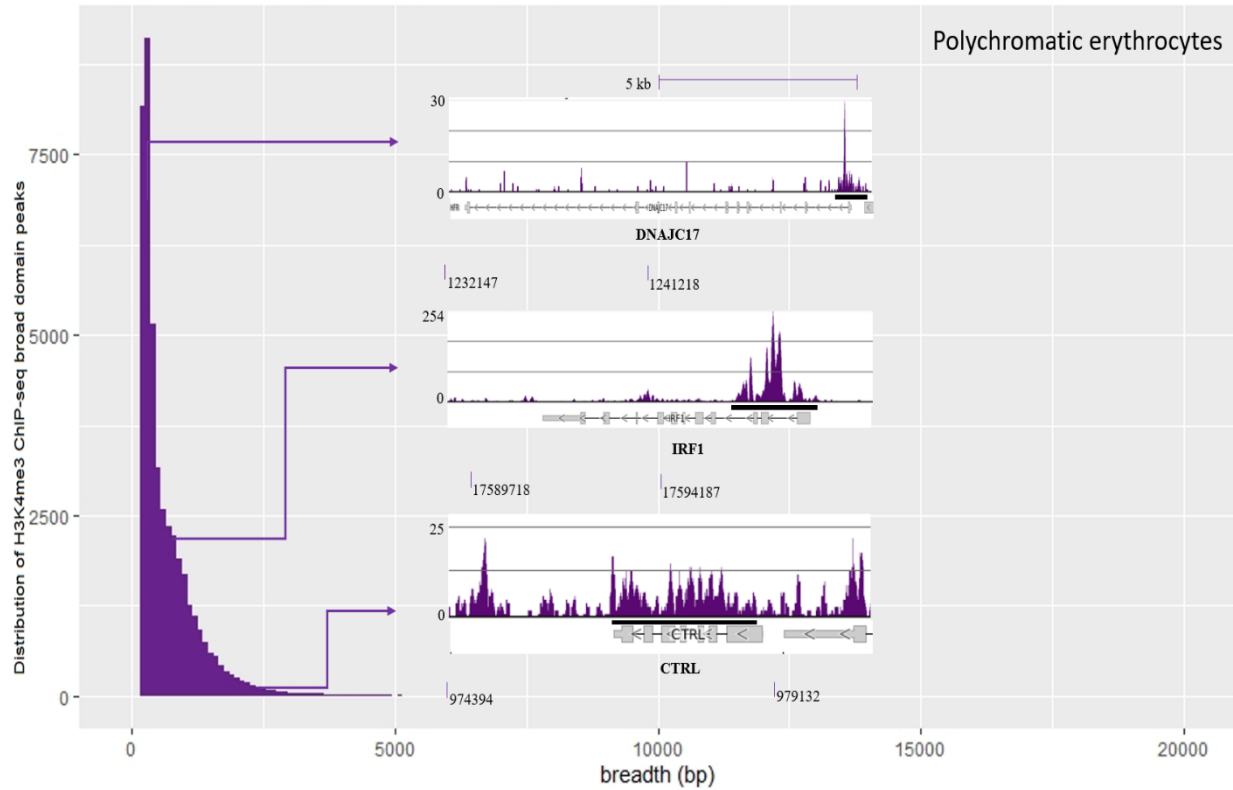


Figure 7.2: Breadth distribution of H3K4me3 ChIP-seq seq in chicken polychromatic erythrocytes. The illustration shows the distinction between sharp and broad peak around TSS of active genes.

To further characterize the H3K4me3 broad domains, we compared the H3K4me3 ChIP-seq data with H3K27ac ChIP-seq data and nucleosome-free DNA regions (sometimes referred to as open chromatin) data from DNase I-seq and FAIRE-seq. Apart from α and β globin genes, chicken polychromatic erythrocyte immune genes such as *IRF7*, *IRF1*, *IFIMT1/5*, *NF-kB2* (Figure 7.3), *IL6*, *IL10*, *TP53I11*, *TNFAIP6*, *TNFAIP8L1* and regulator/proto-oncogenes such as *MYC*, *ROS1*, wnt signalling genes *Wnt10A* *Wnt2B*, rho family gene *RHOC*, major vault protein gene *MVP*, apoptotic gene *BCL2*, *BCL2L10*, MAPK signaling gene *MAPK4*, notch signalling pathway gene

NOTCH1, transmembrane protein *TMEM*, sialidase gene *NEU3*, neuronal differentiation gene *NEUROD4* were observed to have both H3K4me3 and H3K27ac broad domains and nucleosome depletion along the gene body. This reconfirmed that H3K4me3 and H3K27ac broad domains overlap with open chromatin regions, consistent with the previous findings. Next we compared *GATA2*, a gene involved in cell identity in erythroid cells, across human and chicken genome (Figure 7.4). We observed similar broad domain distribution. For the chicken *GATA2* gene we also looked at the active marks H3R2me2s and H4R3me2a which similarly aligned with the distribution of H3K4me3 and H3K27ac and open chromatin profiles in chicken polychromatic erythrocytes.

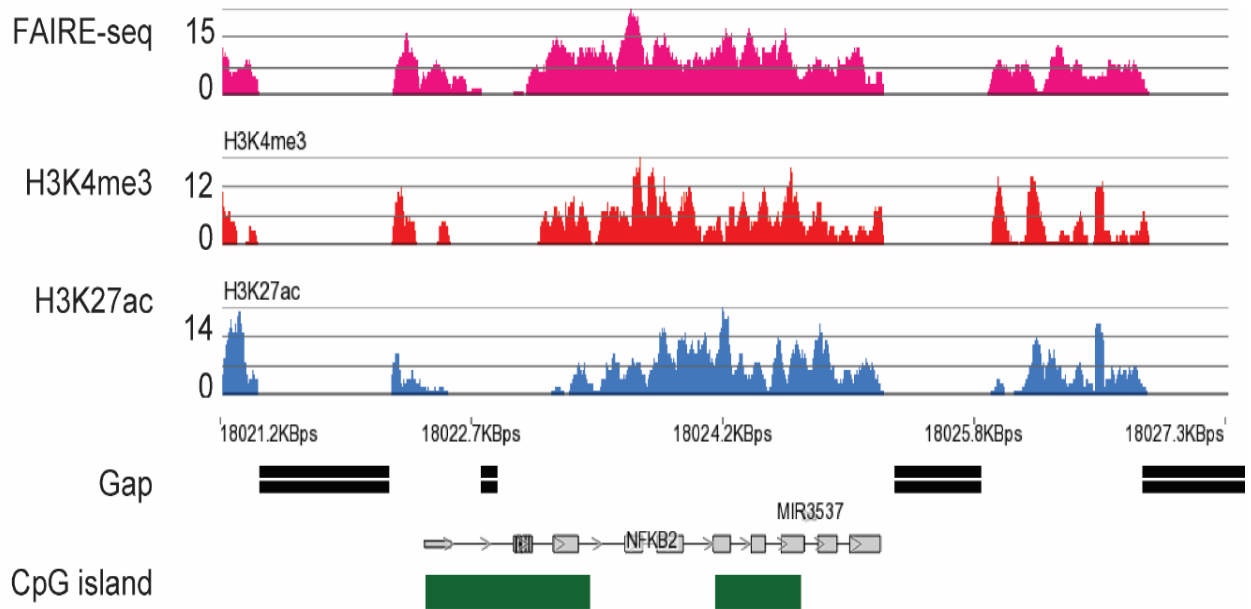
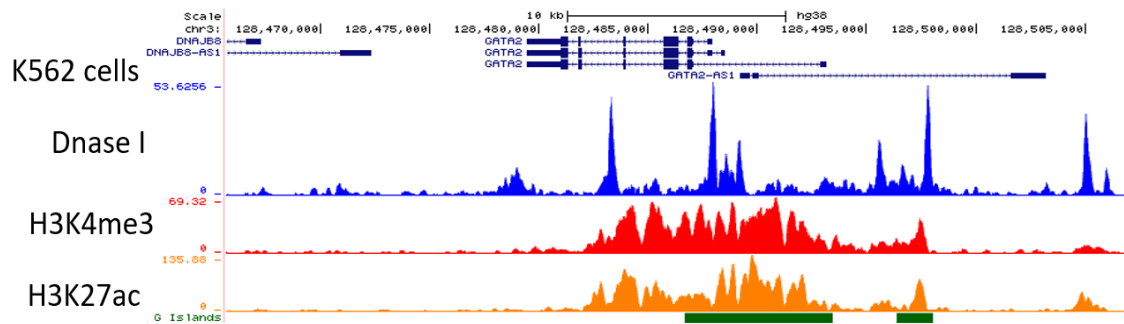


Figure 7.3: Immune gene *NFKB2* of polychromatic erythrocytes depicting H3K4me3 and H3K27ac broad domain in alignment with open chromatin region. Positions of CpG islands are shown.



Chicken polychromatic erythrocyte

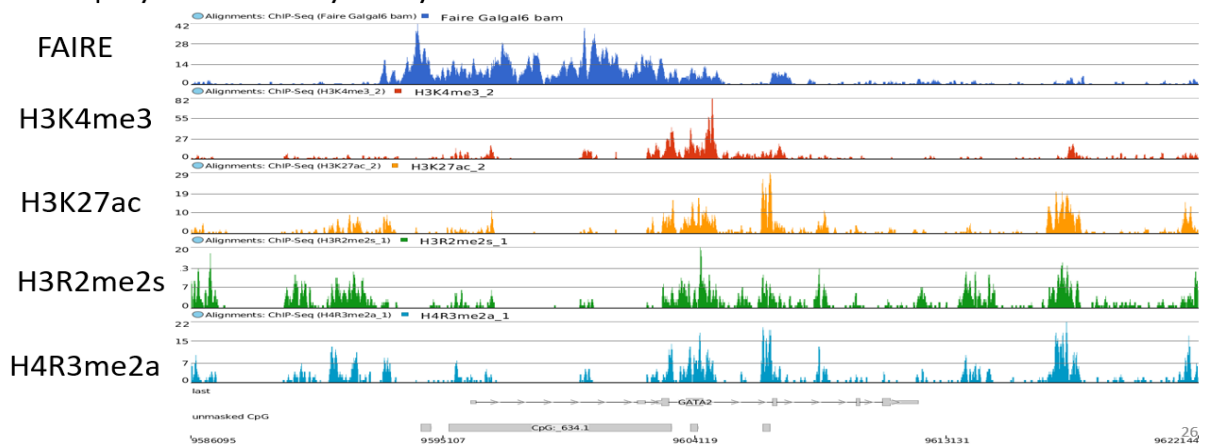


Figure 7.4: Comparison of human *GATA2* gene in K562 cells with polychromatic erythrocytes *GATA2* depicting similar broad domain features in additional active marks H3R2me2s and H4R3me2a.

To gain perspective on the biological relevance of these broad domain, we looked at the publicly available ChIA-PET data of K562 cells to understand H3K27ac mediated chromatin interactions around the *GATA2* gene (Figure 7.5). These data suggested that the H3K27ac broad domain was functioning as a super enhancer. The looping of this domain to the near by genes was supportive of the proximal super-enhancer characteristic (6). I next did a motif enrichment analysis

which further confirmed the intron in the broad domain was acting as a regulatory element (Table 7.1).

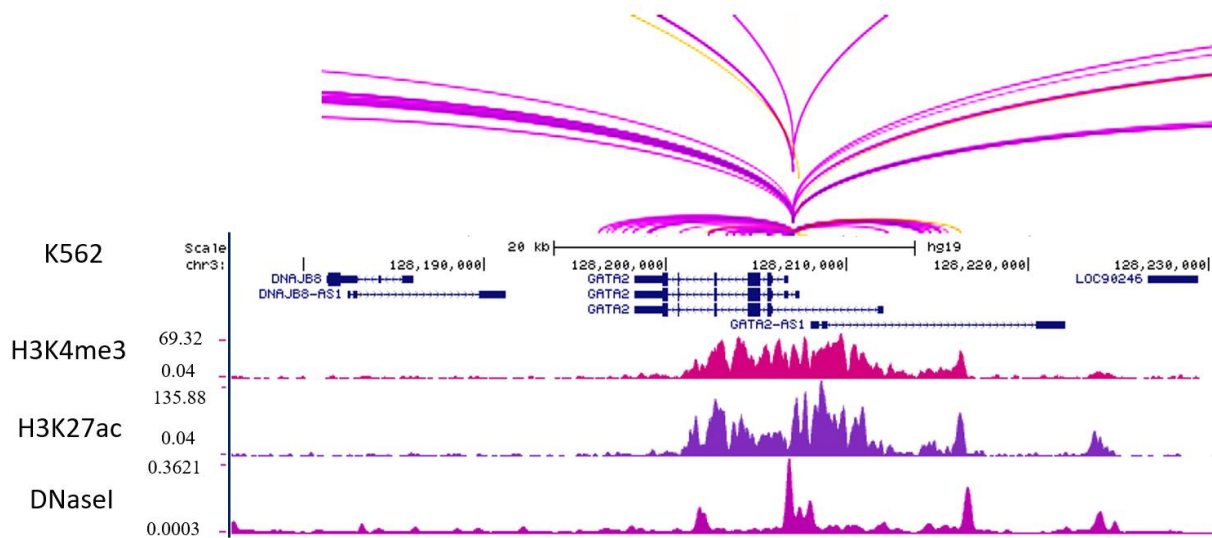



Figure 7.5: H3K27ac ChIA-PET interactome in K562 cell for *GATA2* aligned with broad domain tracks of H3K4me3 and H3K27ac. This level of interactions is typically observed with enhancers and super-enhancers suggesting that the intron in the broad domain is acting as a regulatory element (8). (Access at <http://promoter.bx.psu.edu/hi-c/chiapet.php?method=chiapet&species=human&assembly=hg19&tissue=K562&target=H3K27ac&lab=Snyder&gene=GATA2&window=&sessionID=&browser=ucsc>)

Table 7.1: Prediction outcome of the *de novo* motif investigated using GOMO

| Motif | score | p-value | q-value | Specificity | GO name (Molecular function) | Log likelihood ratio |
|---|-----------|-----------|-----------|-------------|-------------------------------|----------------------|
|  | 6.624e-05 | 2.169e-05 | 3.047e-02 | ~83% | transcription factor activity | 1,376,350.00 |

7.4 Discussion

To this date, it has been reported that genes with H3K4me3 broad signature and super-enhancer/broad H3K27ac domain are involved in cell identity and cancer (9). My result was consistent with this finding. Interestingly my analysis of stretched broad domains of H3K4me3 and H3K27ac also included immune genes involved in innate immunity. One study reported that genes with broad H3K4me3 domains were highly expressed in both peripheral-blood-derived and skin-derived human mast cells (10), but here I report for the first time that the overlapping of broad domains of H3K4me3 and H3K27ac broad domain in genes involved in immunity and several different signalling pathways. The most interesting finding was the association of open chromatin along the stretched histone domains, all in alignment. Such alignment could signify the association of unstable nucleosomes. One interpretation could be open chromatin and broad domains work in harmony for pre-initiation complex assembly (11). Similarly, the ATP-dependent chromatin-remodeling factor CHD1 interacts with H3K4me3, modifies nucleosome, mediators (12, 13) and

recruits' factors for mRNA maturation (14). Taken together, the H3K4me3 and H3K27ac broad domain and open chromatin domains facilitate transcription initiation and elongation.

We further exploited the ChIA-PET data in K562 chronic myelogenous leukemia cells to understand the association of H3K4me3 and H3K27ac broad domain with chromatin interaction. Our result suggested the broad domains were more involved in chromatin interaction compared to typical ones. This is also consistent with the findings of Cao et al; they compared overlapping super-enhancers and broad H3K4me3 domains in the K562 chronic myelogenous leukemia and MCF-7 breast cancer cell line with chromatin interactions and reported super-enhancers and broad H3K4me3 domains showed higher association with chromatin interactions than their typical counterparts (6). The looping interaction and motif enrichment analysis were consistent with the presence of intergenic super enhancers as seen in *GATA2* gene. The de novo motif identified was confirmed to have transcription factor binding which again ties to the function of these domains in transcriptional consistency. Further conclusions can be made through meta-analysis of high-throughput genomics data, designing algorithms and use of machine-learning models, and validating the prediction through experimental analysis (qPCR or genome-wide track investigation).

To date very few studies have explored the H3K4me3 broad domain conservation of genes involved in cell identity in specific cell types across different species. Therefore, my study will provide more insight about the conservation and function of genes marked with H3K4me3 broad domain in human and chicken erythroid cells. From a pathology perspective, this work shows the

significance of identifying genes with H3K4me3/H3K27ac broad domains as a tool to discover genes critical to normal and cancer cells.

7.5 References

- [1] Benayoun, B. A., Pollina, E. A., Ucar, D., Mahmoudi, S., Karra, K., et al. (2014) H3K4me3 breadth is linked to cell identity and transcriptional consistency. *Cell* 158, 673–688.
- [2] Park, S., Kim, G. W., Kwon, S. H., and Lee, J. S. (2020) Broad domains of histone H3 lysine 4 trimethylation in transcriptional regulation and disease. *FEBS Journal*.
- [3] Dincer, A., Gavin, D. P., Xu, K., Zhang, B., Dudley, J. T., et al. (2015) Deciphering H3K4me3 broad domains associated with gene-regulatory networks and conserved epigenomic landscapes in the human brain. *Translational Psychiatry* 5, e679–e679.
- [4] Dahl, J. A., Jung, I., Aanes, H., Greggains, G. D., Manaf, A., et al. (2016) Broad histone H3K4me3 domains in mouse oocytes modulate maternal-to-zygotic transition. *Nature* 537, 548–552.
- [5] Sankar, A., Lerdrup, M., Manaf, A., Johansen, J. V., Gonzalez, J. M., et al. (2020) KDM4A regulates the maternal-to-zygotic transition by protecting broad H3K4me3 domains from H3K9me3 invasion in oocytes. *Nature Cell Biology* 22, 380–388.
- [6] Cao, F., Fang, Y., Tan, H. K., Goh, Y., Choy, J. Y. H., et al. (2017) Super-enhancers and broad h3k4me3 domains form complex gene regulatory circuits involving chromatin interactions. *Scientific Reports* 7, 2186.
- [7] Jahan, S., Beacon, T. H., Xu, W., and Davie, J. R. (2020) Atypical chromatin structure of immune-related genes expressed in chicken erythrocytes. *Biochemistry and Cell Biology* 98, 171–177.

- [8] Kent, W. J., Sugnet, C. W., Furey, T. S., Roskin, K. M., Pringle, T. H., et al. (2002) The Human Genome Browser at UCSC. *Genome Research* 12, 996–1006.
- [9] Chen, K., Chen, Z., Wu, D., Zhang, L., Lin, X., et al. (2015) Broad H3K4me3 is associated with increased transcription elongation and enhancer activity at tumor-suppressor genes. *Nature Genetics* 47, 1149–1157.
- [10] Cildir, G., Toubia, J., Yip, K. H., Zhou, M., Pant, H., et al. (2019) Genome-wide Analyses of Chromatin State in Human Mast Cells Reveal Molecular Drivers and Mediators of Allergic and Inflammatory Diseases. *Immunity* 51, 949-965.e6.
- [11] Lauberth, S. M., Nakayama, T., Wu, X., Ferris, A. L., Tang, Z., et al. (2013) H3K4me3 interactions with TAF3 regulate preinitiation complex assembly and selective gene activation. *Cell* 152, 1021–1036.
- [12] Lee, Y., Park, D., and Iyer, V. R. (2017) The ATP-dependent chromatin remodeler Chd1 is recruited by transcription elongation factors and maintains H3K4me3/H3K36me3 domains at actively transcribed and spliced genes. *Nucleic acids research* 45, 8646.
- [13] Lin, J. J., Lehmann, L. W., Bonora, G., Sridharan, R., Vashisht, A. A., et al. (2011) Mediator coordinates PIC assembly with recruitment of CHD1. *Genes and Development* 25, 2198–2209.
- [14] Davie, J. R., Xu, W., and Delcuve, G. P. (2015) Histone H3K4 trimethylation: Dynamic interplay with pre-mRNA splicing¹. *Biochemistry and Cell Biology* 94, 1–11.

Chapter-8

Discussion and Future direction

8.1 Study insight and future direction

In Chapter 4 “Chromatin organization of transcribed genes in chicken polychromatic erythrocyte”, we combined F1-seq, Chip-seq (H3K4me3 and H3K27ac), RNA-seq and FAIRE-seq to characterize features of chromatin compartment A containing expressed genes. Our analysis showed that FAIRE-seq readily identified promoter regions but failed to map out the enhancers as efficiently. While we could detect the enhancer regions through PCR, sequencing results did not show the enhancer peaks. We assumed this discrepancy in both results could be due to fragment size selection during library preparation. Typically, the peaks identified by FAIRE-seq were localized to the 5' regions of the gene body. H3K4me3 and H3K27ac peaks were intense at the boundaries of NDRs. We speculated that the NDRs harboured the writers for the PTMs investigated. The chromatin modifying enzymes (the writers) are usually recruited by transcription factors which would bind to the NDRs within upstream promoter regions. For example, *CA2* gene from chapter 4 has NDR covering the transcription start site and upstream promoter element. The breadth of the NDR would be more than what is covered by the general transcription factors (TFIID, TFIIB etc.) and the preinitiation complex which has to be nucleosome free (1). At the upstream promoter region would be sites for multiple transcription factors that would recruit the chromatin modifying/remodeling factors. Our analysis of ten genes showed that the average reach of the writers was a maximal distance of 2300 bp for H3K4me3 and 1930 bp for H3K27ac, overall showing the length of modification was about 9-10 nucleosomes starting from TSS. We also found the TSS and CGI profile of FAIRE-seq comparable to the human promoter finding (2, 3). This

study was a setting step to further investigate the comparable and conserved features in chicken and human genome.

We expanded our study of compartment A genes by investigating two more active marks, H3R2me2s and H4R3me2a, in Chapter 6 “Genomic landscape of transcriptionally active histone arginine methylation marks, H3R2me2s and H4R3me2a, relative to NDRs.” The distribution of arginine modifications H3R2me2s and H4R3me2a genome wide was determined by analysis of the ChIP-seq data, and the chromatin structure of regions with these histone PTMs was characterized by FAIRE-seq in chicken polychromatic erythrocytes. The sequencing data showed that on either side of the NDR were nucleosomes marked with H3K4me3, H3K27ac, H3R2me2s and H4R3me2a. Genome distribution analysis showed that H4R3me2a marked active promoters, gene body and enhancer regions, whereas H3R2me2s associated with active/poised enhancers and the 5' end of gene body. Correlation studies showed that mononucleosomes modified by H3R2me2s were also modified by H4R3me2a, indicating their co-dependent functioning to maintain active chromatin. By comparing the transcriptome data (RNA-seq data aligned with the galGal6 genome assembly) with the ChIP-seq data of H3R2me2s and H4R3me2a, I showed that that expressed genes were enriched in both methylated arginine marks. This study supported the importance of PRMT1 in recruiting CBP/p300 which acetylates H3K27 (a mark of active enhancers and promoters). Adenosine dialdehyde inhibition of the enzymatic activity of histone methyltransferases, including PRMT1 and PRMT5, showed a significant decrease in the three active histone marks, H3K4me3, H3R2me2s and H4R3me2a after 12 h treatment suggesting mechanisms are in place to reduce the levels of these histone PTMs. Possible mechanisms to

decrease the levels of H3R2me2s and H4R3me2a involve the activity of a methyl arginine demethylase (to date not identified) and/or nucleosome turnover. PRMT1 and PRMT5 knock down (or knock out) or inhibition studies of chicken polychromatic erythrocytes using specific inhibitors will further reveal whether loss of PRMT1 or PRMT5 activity reduces acetylated histone, decreases H3K4me3 levels and alters gene expression. The relationship between PRMT1, CBP/p300 and H3K27ac is of particular interest as PRMT1 may play a critical role in enhancer chromatin structure and function. It has been reported that knockdown of PRMT1 in mouse erythroleukemic cells resulted in the loss of the interaction of the β -globin LCR with the β^{maj} promoter and failure of β -globin gene transcription (4). It is currently not known how many erythroid enhancers require PRMT1 activity. To address could be investigated by using Ocean-C in cells in which PRMT1 activity has been inhibited. The technique Ocean-C combines the key steps of FAIRE-seq and Hi-C and has been proven to be an excellent way to isolate enhancers and map out long and short distance interactions overcoming the shortcomings on the individual use of FAIRE-seq and Hi-C techniques (41).

Correlations between GO terms and CGI length revealed, the chicken promoters could be segregated into two categories: chicken promoters with long CGIs (>800 bp, LCGI) promoters, which occupied the top 10% of the genes in terms of CGI length and biological processes such as morphogenesis, development, transcription processes, and biosynthetic processes (explored in Chapter 4). “No CGI” (NCGI) promoter (5) was used to refer to promoters without CGI and typically harboured biological processes such as immune system and defense mechanisms. In Chapter 5 “Atypical chromatin structure of immune-related genes expressed in chicken erythrocytes”, we explored the function of chicken polychromatic erythrocytes in the innate

immune signaling and characterized genes coding for Toll-like receptors, interleukins, and interferon regulatory factors. While most genes resonated with the categories defined above, some genes had an atypical chromatin structure with NDRs in the gene body. We characterized the atypical chromatin feature of the immune genes as having the presence of broad H3K4me3 domains, nucleosome depletion over the gene body, and a stretched domain of H3K27ac nucleosome repeat units. While the exact mechanism is still unknown, it is quite clear from the expression of the immune genes that these genes might contribute to the identity and function of the erythrocyte in combating microbial pathogens. We also analyzed the transcriptome of the chicken early erythroid CFU-E stage cell line 6C2 and found that genes involved in innate immunity had a comparable expression profile (Chapter 5; data not shown). Therefore, to further gain insight into the chicken immune gene response and their open chromatin signature, Faire-seq of 6C2 will be a crucial experiment. In mammals, interferon stimulating genes such as the *IFITM* genes are upregulated early in the innate response to prevent the cellular entry of several viruses such as HIV-1, filovirus, FLUAV, VSV, WNV, YFV, DENV and SARS-CoV (6). On investigating the *IFITM* genes in human (*IFITM 1/2/3/5*) and chicken (*IFITM 1/3/5*), I saw a similar broad H3K4me3 domain signature in both genomes. In chicken polychromatic erythrocytes, the genes having this atypical chromatin structure and involved in innate immunity (e.g. the *NFKB2* gene) were poorly expressed. I speculate that these genes have the atypical chromatin structure to allow rapid transcription elongation by RNA polymerase II in response to an invading virus. Therefore, further investigation into the formation and maintenance of the atypical broad domain could be crucial in addressing not only avian flu and diseases but also human health and current pandemics COVID-19.

Interestingly, the function of the erythrocyte in the innate immune response is not unique to birds. Participation of erythrocytes in immune response was observed in fish, mouse (erythroblast) as well as human (erythroblasts and erythrocytes) (7–9). To identify the key epigenetic modifier that lead to the presence of NDR and modified nucleosomes at the same time, we proposed a model (Figure 8.1) where we propose that CHD1 remodels nucleosomes and introduces H3.3 into the broad domains. Given the role of CHD1 in nucleosome disassembly, nucleosome spacing (10) and regulation of the immune response (11), CHD1 mediated atypical chromatin structure might be essential for the biological event of transcriptional memory ingrained into the immune genes for their function in innate response. Therefore, ChIP-seq of CHD1 of chicken erythrocyte will be helpful to further address the investigation.

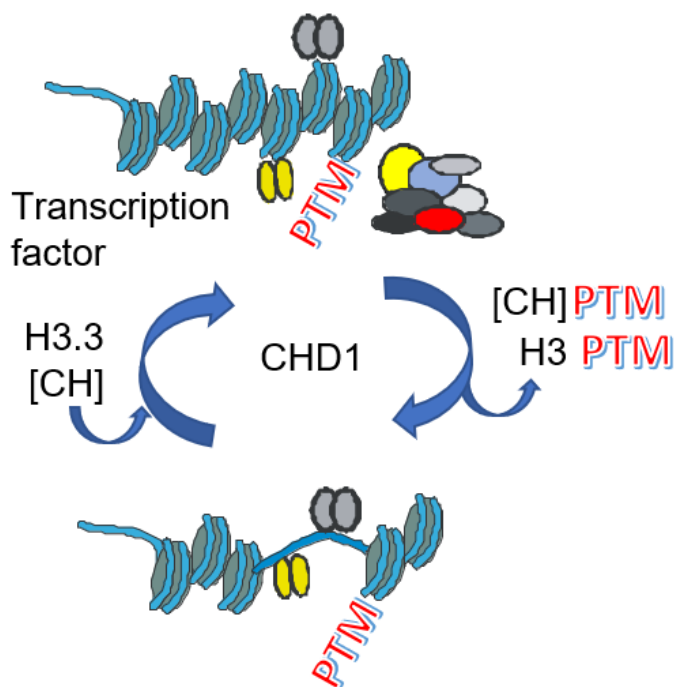
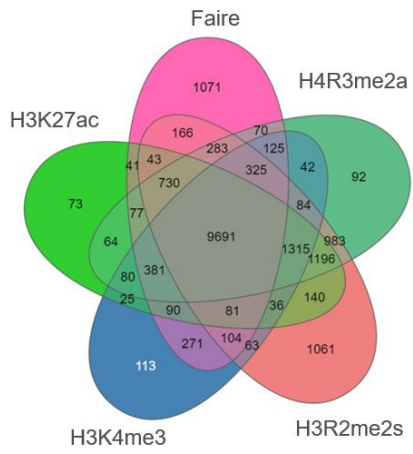


Figure 8.1: Proposed model for the function of CHD1 in remodelling nucleosomes. The remodelled nucleosomes have the canonical H3 histone replaced by the variant histone H3.3.

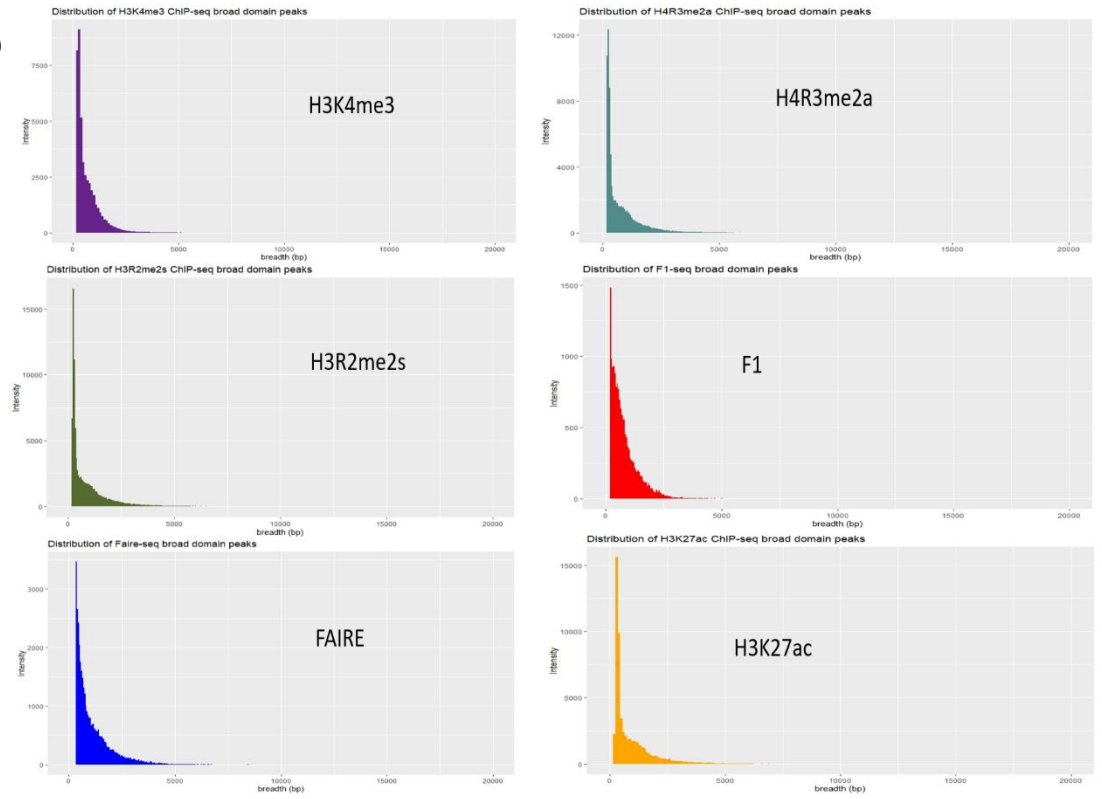
In this study we also reported that broad domain is not unique to chicken but is an important hallmark in human genome as well. Zubek et al reported in human derived cell lines, extended regions of CHD1 binding is associated with longer H3K27ac and H3K4me3 domains (12). This finding led to the design of the study in Chapter 7 “Investigation of H3K4me3 broad domains in different cell types reveal importance of long chromatin signatures in human and chicken genomes”. Consistent with the literature, we found genes involved in cell identity were marked with the H3K4me3 broad domain. Further analyses of H3K4me3 broad domain genes in different cell types (chicken and human) demonstrated that typically H3K27ac marks the same region as H3K4me3 and that H3K4me3/H3K27ac broad domains were associated with H3R2me2s,

H4R3me2a and unstable nucleosomes as made evident using the FAIRE and DNase I hypersensitivity assays, e.g. *GATA2* in K562 cells. To further investigate the genes with such signature, I mapped out all the broad domains (Figure 8.2a) of H3K4me3, H3K27ac, H3R2me2s, H4R3me2a, FAIRE, and identified 9691 genes (Figure 8.2b) which require further investigation. I also assumed that the broad domains would be associated with highly acetylated histones. This will be addressed through further investigation of F1 broad domains (the physiological ionic strength solubility of the chromatin in F1 is directly related to the histone acetylation state). Preliminary analyses of F1-seq data showed that F1-enriched sequences are in narrow and broad peaks (Figure 8.2b). I also investigated the pathways associated with broad open chromatin (Figure 8.2c). Further investigation of the genes enriched in such pathways will identify genes that have the atypical chromatin structure associated with H3K4me3/H3K27ac broad domains.

(a)



(b)



(c)

GO ENRICHMENT ANALYSIS

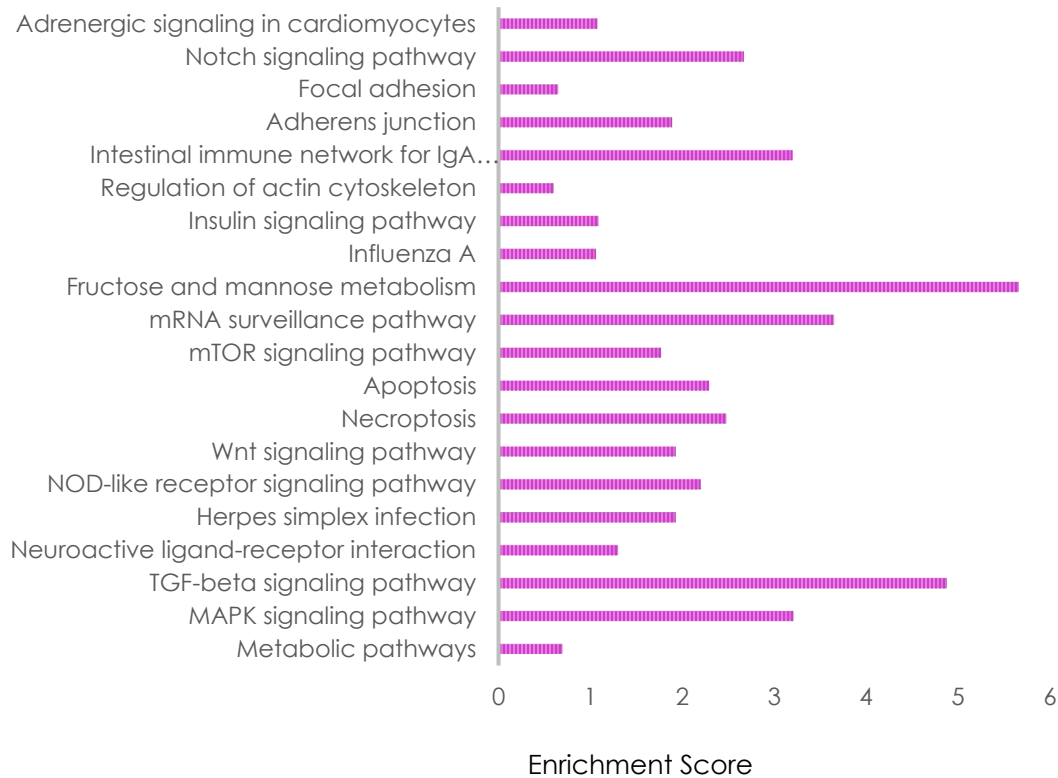


Figure 8.2: Analysis of different PTMs and open chromatin broad domain. (a) Venn diagram depicting the genes present within the set of overlapping broad domains of H3K4me3, H3K27ac, H3R2me2s, H4R3me2a, FAIRE. (b) Breadth distribution of ChIP-seq peaks of broad domains of H3K4me3, H3K27ac, H3R2me2s, H4R3me2a, FAIRE-seq and F1-seq. (c) Gene ontology depicting molecular function enrichment of genes with broad open chromatin regions (broad FAIRE-seq regions).

Study of interactome data of *GATA2* broad domains in K562 cells suggest that these domains interact with super enhancers and may themselves have enhancer activity. While more evidence is required to prove their enhancer activity, studies such as CRISPR Cas9 can be designed

to delete the region and record the gene expression. Chromatin state detection, differential expression analysis of associated genes, gene ontology, motif analysis and further downstream analysis is ongoing using Partek Flow, chromHMM and chromstaR tools.

In Chapter 3, the study “Investigation of histone variants H3.3 and H3.2 and their associated active histone post-translational modification using acid-urea-triton-x electrophoresis”, I identified chicken polychromatic erythrocyte histones participating in or being impacted when dynamically acetylated histones became hyperacetylated after one-hour treatment of cells with sodium butyrate. Applying this well documented protocol, I determined which of the H3 variants and H3/H4 PTMs were participating in dynamic histone acetylation. The most significant finding was with H3K4me3 which shows that H3.3 is preferentially trimethylated at K4 and that H3.3K4me3 at promoters and 5’ end of genes is dynamically acetylated. This suggests that HDAC inhibition results in multiple site acetylation in the H3.3K4me3 tail. This makes sense since H3K4me3 recruits a variety of KATs (readers of the H3K4me3 mark) which would acetylate the histones at multiple sites. Looking at the tracks for *FTH1* gene it appears that F1 locates with H3K4me3 which participates in dynamic multisite acetylation (Figure 8.3). Correlation studies will further help us to further explore this relationship. H3.3 is also preferentially modified by K27ac. H3.3K27ac has a much broader distribution than H3K4me3 in being associated with promoters, intergenic regions and introns (particularly with genes with an atypical chromatin structure). Genomic analysis suggests that H3.3K27ac is often in regions lacking K4me3 as such seen in Figure 4.1a for example. I propose that H3.3K27ac is mostly acetylated at one site (K27) and that this site is dynamically acetylated (i.e., CBP/p300 is acetylating K27 and HDACs deacetylating K27ac). However, for nucleosomes that have H3.3 modified at K27ac and K4me3,

I would expect that this H3.3 would participate in multisite dynamic acetylation. This can be further investigated by immunoprecipitation and mass spectrometry analysis of H3.3K27ac and H3.3K4me3/K27ac nucleosomes. I did not observe striking changes in H3.3R2me2s after butyrate treatment. This observation suggests that inhibition of HDAC activity by butyrate does not alter the activity of PRMT5 or the accessibility of H3.3R2 to the enzyme. As with H3K27ac, H3R2me2s has a much broader genomic distribution than H3K4me3. However, there is a population of nucleosomes that have H3 with R2me2s and K4me3 on the same tail (see Discussion in Chapter 6). These nucleosomes modified at R2me2s and K4me3 at upstream promoter regions and at the 5' end of expressed genes likely participate in multisite dynamic acetylation. On the other hand, H4R3me2a maps strongly with H3K27ac. Thus, when comes to inhibiting HDACs, no major shift in the AUT gel after butyrate treatment was seen for H4R3me2a as this nucleosome is often associated with H3K27ac which is acetylated mainly at one site. The lack of shift also suggests that the KATs do not acetylate H4R3me2a at multiple K sites. An increase in the intensity of H4R3me2a after butyrate treatment could possibly be due to a change in the activity of PRMT1. PRMT1 has multiple acetylation sites (<https://www.phosphosite.org/proteinAction.action?id=6954&showAllSites=true>), some of which may be acetylated by CBP/p300 which interacts with PRMT1. This idea would require further investigation.

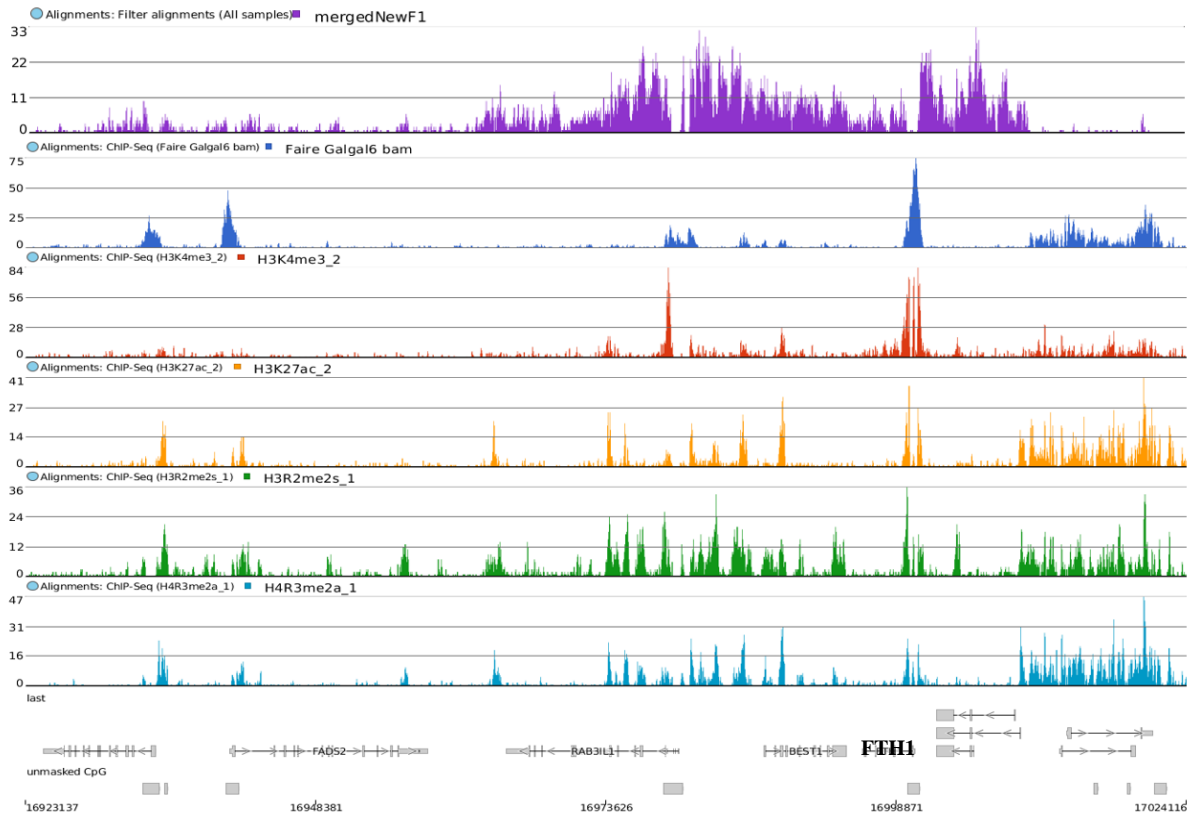


Figure 8.3: Snapshot of *FTH1* gene depicting the tracks of F1-seq, FAIRE-seq and ChIP-seq of H3K4me3, H3K27ac, H3R2me2s and H4R3me2a to show their genomic distribution and their relation to dynamic acetylation.

8.2 Study limitation and challenges

The major bottleneck for this project was finding tools for downstream computational analysis that works for the most recent chicken genome (*Gallus gallus* 6). Given most of our sequencing data were aligned with the latest chicken genome as reference genome, integrative

analysis using the same reference genome was extremely important for downstream biological interpretations and designing follow-up studies. The other struggle was analysis of broad domain. While most algorithms are designed to detect well-defined sharp peaks, usage of that to call for broad domain can generate many false positive or false negative calls (13). While tools such as Macs2, histoneHMM and Chromhmm have a broad domain identifying feature, they all come with their algorithmic or technical limitations such as overpredicting algorithm specific region, miscalling of pileups and blacklisted region (14).

8.3 Significance of the project

Chicken has approximately 20,000-23,000 genes in its 1 billion DNA base pairs, compared to human count of 20,000-25,000 genes in 2.8 billion (15). While both share approximately the same number of genes, the absence of other DNA elements (repetitive) makes chicken a simplified model to study and target the gene of interest, orthologous to their human counterpart. In this study I have shown the characterization of several genes, the signature of which is comparable to their distant relative human. Therefore, in pharmacological perspective, preclinical drug trials or gene therapy studies can be designed using the cost-effective chicken model.

FAIRE-seq readily identified promoter regions of transcribed genes contributing to gene annotation, bridging sequence gaps, identifying novel gene locations and atypical chromatin features in genes involved in innate immunity and cell identity in chicken and human genome. Loss of PRMT1 and PRMT5 has been linked to embryonic lethality confirming their crucial role

in the erythroid development of the organism (16). PRMTs are also known to regulate pro-inflammatory cytokine expression both dependent and independent of their enzymatic activity and catalyzed PTMs (17). Frequent H3R2/R8 and H4R3 hypomethylation leads to metabolic disorders, autoimmune disease and cardiovascular disease (18). Therefore, my study of PRMTs and their catalyzed PTMs will provide insight in their role in modulating active chromatin.

The broad H3K4me3 domain chromatin structure found in some genes plays a crucial role in transcription, development, memory formation, and several diseases, including cancer and autoimmune diseases (19). Implementation of computational model and chromatin interaction analysis on patient samples will help us to understand how the broad H3K4me3 domain activates cell-type-specific or disease-specific genes. Similarly, analysis of broad domains in other active marks will reveal a whole new perspective to study histone modifications.

8.4 References

- [1] Yoshihara, C. M., Lee, J.-D., and Dodgson, J. B. (1987) The chicken carbonic anhydrase II gene: evidence for a recent shift in intron position.
- [2] Giresi, P. G., Kim, J., McDaniel, R. M., Iyer, V. R., and Lieb, J. D. (2007) FAIRE (Formaldehyde-Assisted Isolation of Regulatory Elements) isolates active regulatory elements from human chromatin. *Genome Research* 17, 877–885.
- [3] Simon, J. M., Giresi, P. G., Davis, I. J., and Lieb, J. D. (2012) Using formaldehyde-assisted isolation of regulatory elements (FAIRE) to isolate active regulatory DNA. *Nature Protocols* 7, 256–267.
- [4] Li, X., Hu, X., Patel, B., Zhou, Z., Liang, S., et al. (2010) H4R3 methylation facilitates β -globin transcription by regulating histone acetyltransferase binding and H3 acetylation. *Blood* 115, 2028–2037.
- [5] Abe, H., and Gemmell, N. J. (2014) Abundance, arrangement, and function of sequence motifs in the chicken promoters. *BMC Genomics* 15, 1–12.
- [6] Schoggins, J. W., and Rice, C. M. (2011) Interferon-stimulated genes and their antiviral effector functions. *Current Opinion in Virology* 1, 519–525.
- [7] Sennikov, S. V., Injelevskaya, T. V., Krysov, S. V., Silkov, A. N., Kovinev, I. B., et al. (2004) Production of hemo- and immunoregulatory cytokines by erythroblast antigen+ and glycophorin A+ cells from human bone marrow. *BMC Cell Biology* 5, 39.
- [8] Elahi, S. (2014) New insight into an old concept: Role of immature erythroid cells in

- immune pathogenesis of neonatal infection. *Frontiers in Immunology* 5, 376.
- [9] Anderson, H. L., Brodsky, I. E., and Mangalmurti, N. S. (2018) The Evolving Erythrocyte: Red Blood Cells as Modulators of Innate Immunity. *The Journal of Immunology* 201, 1343–1351.
- [10] Siggins, L., Cordeddu, L., Rönnerblad, M., Lennartsson, A., and Ekwall, K. (2015) Transcription-coupled recruitment of human CHD1 and CHD2 influences chromatin accessibility and histone H3 and H3.3 occupancy at active chromatin regions. *Epigenetics and Chromatin* 8, 4.
- [11] Sebald, J., Morettini, S., Podhraski, V., Lass-Flörl, C., and Lusser, A. (2012) CHD1 Contributes to Intestinal Resistance against Infection by *P. aeruginosa* in *Drosophila melanogaster*. *PLoS ONE* 7, e43144.
- [12] Zubek, J., Stitzel, M. L., Ucar, D., and Plewczynski, D. M. (2016) Computational inference of H3K4me3 and H3K27ac domain length. *PeerJ* 4, e1750.
- [13] Heinig, M., Colomé-Tatché, M., Taudt, A., Rintisch, C., Schafer, S., et al. (2015) histoneHMM: Differential analysis of histone modifications with broad genomic footprints. *BMC Bioinformatics* 16, 60.
- [14] Carroll, T. S., Liang, Z., Salama, R., Stark, R., and de Santiago, I. (2014) Impact of artifact removal on ChIP quality metrics in ChIP-seq and ChIP-exo data. *Frontiers in Genetics* 5, 75.
- [15] Hillier, L. W., Miller, W., Birney, E., Warren, W., Hardison, R. C., et al. (2004) Sequence and comparative analysis of the chicken genome provide unique perspectives on vertebrate

evolution. *Nature* 432, 695–716.

- [16] Poulard, C., Corbo, L., and Le Romancer, M. (2016) Protein arginine methylation/demethylation and cancer. *Oncotarget* 7, 67532–67550.
- [17] Jayne, S., Rothgiesser, K. M., and Hottiger, M. O. (2009) CARM1 but not Its Enzymatic Activity Is Required for Transcriptional Coactivation of NF- κ B-Dependent Gene Expression. *Journal of Molecular Biology* 394, 485–495.
- [18] Shen, W., Gao, C., Cueto, R., Liu, L., Fu, H., et al. (2020) Homocysteine-methionine cycle is a metabolic sensor system controlling methylation-regulated pathological signaling. *Redox Biology* 28, 101322.
- [19] Park, S., Kim, G. W., Kwon, S. H., and Lee, J. S. (2020) Broad domains of histone H3 lysine 4 trimethylation in transcriptional regulation and disease. *FEBS Journal*.