# Epigenetic Studies of Complex Human Disorders and Aging

by

Richie Hrachia Jeremian

A thesis submitted in conformity with the requirements
for the degree of Doctor of Philosophy

Institute of Medical Science
University of Toronto

# Epigenetic Studies of Complex Human Disorders and Aging

Richie Hrachia Jeremian

Doctor of Philosophy

Institute of Medical Science
University of Toronto

2018

## Abstract

Aging is a major risk factor for nearly all human illnesses. Several of these have a strong genetic and familial basis, but do not manifest until much later in life. Numerous studies have highlighted the potential of epigenetic studies to help quantify pathological and normal hallmarks of the aging process. In this dissertation, we investigated the role of DNA modification in bipolar disorder and found epigenetic deviations from biological age in individuals with a history of extreme suicidal behaviour. We then studied the simple genetic trait of lactase non-persistence (commonly known as lactose intolerance), the age-related reduction of lactase enzyme production that manifests after weaning in most but not all individuals, as a model to help understand how genotype influences age-related outcomes, and the role that DNA modification plays in mediating this process. Inherited genetic variants near the lactase gene are highly-associated with lactase persistence, the sustained activation of this gene into adulthood that is exhibited by about 30% of all humans. By performing high-resolution DNA modification profiling at the lactase gene, we found that both the age-related decline in levels of the lactase gene and its escape in persistent individuals occur as a function of genotype-dependent accumulation of epigenetic modifications; thus, possessing a given variant confers a certain "epigenetic aging trajectory". Further, we identified that the segment-specific expression of the lactase gene in a homogenous population of epithelial cells in mouse small intestine is dependent

upon DNA modification differences. The insights from these efforts have the potential to reshape the dogma which has to this point comprised solely of genetic factors, and to lay the foundation for standardized comprehensive modelling of the role of aging in human health and disease beyond simple genetic association.

# Acknowledgments

I am eternally grateful to the people in my life during this transformative ordeal – advisors, mentors, colleagues, peers, friends and family – together, we have created something special. None of this would have been possible without your guidance, love and support, and I am a better scientist and human being because of it. Until next time, keep it raw!

# Contributions

Richie Jeremian (author) solely prepared this thesis. All aspects of this body of work, including the planning, execution, analysis, and writing of all original research and publications was performed in whole or in part by the author. The following contributions by the author and by other individuals are formally and inclusively acknowledged.

Chapter 3

Mr. Richie Jeremian – experimental work [contribution to all figures], data analysis, manuscript authorship

Dr. John Strauss – mentorship & guidance, study design, manuscript review and senior authorship

Dr. Clement Zai, Mr. Yi-an Chen – assisted with SNP analysis and microarray data cleaning, respectively

Drs. Vincenzo de Luca, John Vincent & James Kennedy – study design and manuscript review

Chapter 4

Mr. Richie Jeremian – experimental work (microarray analysis, bisulfite-sequencing, animal work, DNA and RNA extraction, genotyping, CRISPR-Cas9 cell line work, CRISPR-Cas9 off-target analysis, qPCR) [contribution to figures 4-3, 4-4, 4-5, 4-6, 4-7, 4-8, 4-9, 4-12, 4-14, 4-15, 4-16, 4-23, 4-24, 4-25], assisted with manuscript review

Dr. Art Petronis & Dr. Michael Brudno – mentorship & guidance, study design, manuscript review and senior authorship

Dr. Viviane Labrie – study design and coordination of experiments, manuscript authorship

Dr. Orion Buske – assisted with study design, bioinformatic analysis, manuscript authorship

Mr. Edward Oh – Assisted with study design, statistical analysis, experimental work, assisted with manuscript review

Dr. Gabriel Oh, Mr. Daniel Groot – assisted with experimental design, performed bioinformatic analysis (bisulfite-sequencing)

Dr. Carolyn Ptak, Ms. Aiping Zhang, Dr. Sasha Ebrahimi, Mr. Akhil Nair – Assisted with peripheral experimental work

Dr. Juozas Goredivius and Mr. Karolis Koncevicius – assisted with bioinformatics analysis (microarrays)

Dr. Almantas Maleckas, Dr. Ruta Petereit, Dr. Aida Zvirbliene, Dr. Kestutis Adamonis, Dr. Limas Kupcinskas – assisted with obtaining and preparing human samples, experimental work and statistical analysis

Dr. Giedrius Gasiunas and Dr. Virginijus Siksnys – prepared the CRISPR-Cas9n cell line constructs

Dr. Edita Kriukiene – consulted on the mTAG approach

Chapter 5

Mr. Richie Jeremian – study design, experimental work (bisulfite-sequencing, animal work, qPCR, DNA and RNA extraction, ChIP) [contribution to all figures], manuscript co-authorship

Mr. Edward Oh – study design, bioinformatics analysis, manuscript co-authorship

Dr. Gabriel Oh, Mr. Daniel Groot – assisted with experimental design and bioinformatics analysis

Mr. Miki Susic – assisted with peripheral experiments

Dr. Art Petronis and Dr. Viviane Labrie – mentorship and guidance, study design, manuscript review and senior authorship

Dr. KwangHo Lee, Ms. Kelly Foy, Dr. Peter Laird – assisted with FACS and cell-type analysis

# Table of Contents

## Contents

# List of Tables

# List of Figures

# List of Appendices

# List of Abbreviations

| | |
|---|---|
| µl | Microlitre |
| 3' UTR / UTR3 | 3′ untranslated region |
| 5' UTR / UTR5 | 5′ untranslated region |
| 5caC | 5-carboxylcytosine |
| 5fC | 5-formylcytosine |
| 5hmC | 5-hydroxymethylcytosine |
| 5HTR1A | Serotonin 1A receptor |
| 5mC | 5-methylcytosine |
| AANAT | Aralkylamine N-acetyltransferase |
| ABCG1 | ATP binding cassette transporter G1 |
| ANK3 | Ankyrin-3 |
| APP | Amyloid beta precursor protein |
| ARGHDIA | Rho GDP dissociation inhibitor alpha |
| ATP | Adenosine triphosphate |
| AXL | Tyrosine-protein kinase receptor UFO |
| BD | Bipolar disorder |
| BDNF | Brain-derived neurotrophic factor |
| BH | Benjamini Hochberg |
| bp | Base pairs |
| BRCA1 | Breast cancer type 1 susceptibility |
| CACNA1C | Calcium channel, voltage-dependent, L type, alpha-1C subunit |
| CEBPA | CCAAT/enhancer-binding protein alpha |
| CH3 | Methyl group |
| ChIP | Chromatin immunoprecipitation |
| COG1 | Congenital disorder of glycosylation type IIg |
| COG2 | Component of oligomeric golgi complex 2 |
| CpG | Cytosine-guanine dinucleotide |
| CpH | Non-CpG site |
| CTCF | CCCTC-binding factor |
| CTNNA2 | Catenin alpha-2 |
| CYTH1 | Cytohesin-1 |
| DACOR1 | DNMT1-Associated Colon Cancer Repressed LncRNA 1 |
| DAPK | Death Associated Protein Kinase 1 |
| DNA | Deoxyribonucleic acid |
| DNMT | DNA methyltransferase |
| DOK7 | Docking protein 7 |
| DSM-IV | Diagnostic and Statistical Manual of Mental Disorders |
| DZ | Dizygotic |
| ecCEBP | Extra-coding CEBPA |
| EGFR | Epidermal growth factor receptor |

| | |
|---|---|
| ESR | Estrogen receptor alpha |
| ESRP2 | Epithelial splicing regulatory protein 2 |
| EWAS | Epigenome-wide association study |
| FACS | Fluorescence-activated cell sorting |
| FDR | False discovery rate |
| GAD1 | Glutamate decarboxylase 1 |
| GAD67 | Glutamate decarboxylase |
| GATA | GATA Binding Protein |
| GNATs | Gcn5-related N-acetyltransferases |
| GPR24 | Melanin-concentrating hormone 1 |
| GWAS | Genome-wide association study |
| H3K27Ac | Histone 3 lysine 27 acetylation |
| H3K27me3 | Histone 3 lysine 27 trimethylation |
| H3K36Ac | Histone 3 lysine 36 acetylation |
| H3K36me3 | Histone 3 lysine 36 trimethylation |
| H3K4me3 | Histone 3 lysine 4 trimethylation |
| H3K9Ac | Histone 3 lysine 9 acetylation |
| H3K9me3 | Histone 3 lysine 9 trimethylation |
| H4K12Ac | Histone 4 lysine 12 acetylation |
| HAT | Histone acetyltransferase |
| HCG9 | Human leukocyte antigen complex group 9 |
| HDAC | Histone deacetylase |
| HIRA | Histone cell cycle regulator |
| HNF1a | HNF1 homeobox A |
| HP1 | Heterochromatin protein 1 |
| IAP | Intracisternal A particle |
| ICD-10 | International Classification of Diseases 10th Revision |
| ICR | Imprinting control region |
| IDH | Isocitrate dehydrogenase |
| IGF2 | Insulin-like growth factor 2 |
| JMJD2 | Jumonji C domain containing protein family |
| KAP1 | KRAB-associated protein 1 |
| kb | Kilobase |
| KCNQ3 | Potassium voltage-gated channel subfamily Q member 3 |
| LCT | Lactase |
| lncRNA | Long non-coding ribonucleic acid |
| LSD1 | Lysine-specific demethylase 1 |
| m6A | N6-methyladenosine |
| MBD | Methyl-CpG binding domain |
| MCM6 | Minichromosome maintenance complex component 6 |
| MCT3 | Monocarboxylate transporter |
| MDD | Major depressive disorder |
| MeCP2 | Methyl-CpG binding protein complex 2 |

| | |
|---|---|
| METTL3 | N6-adenosine-methyltransferase subunit 3 |
| MGMT | O-6-Methylguanine-DNA Methyltransferase |
| miRNA | MicroRNA |
| MIXL1 | Mix paired-like homebox |
| MPP4 | Membrane palmitoylated protein 4 |
| mRNA | Messenger RNA |
| mTAG | Methyltransferase-directed transfer of activated groups |
| MYADML2 | Myeloid associated differentiation marker like 2 |
| MYST | (MOZ, Ybf2/Sas3, Sas2 and Tip60)-related HATs |
| MZ | Monozygotic |
| NDN | Necdin |
| ng | Nanogram |
| NSHELF | -4 to -2 kilobase region flanking the CpG island |
| NSHORE | -2 kilobase region flanking the CpG island |
| NUP133 | Nucleoporin 133 |
| ODZ4 | Teneurin 4 |
| OH | Hydroxyl group |
| p16 | Cyclin Dependent Kinase Inhibitor 2A |
| p300/CBP | p300-CREB-binding protein HATs |
| PAD4 | Peptidylarginine deiminase 4 |
| PDGFRA | Platelet derived growth factor receptor alpha |
| PDX-1 | Pancreatic and duodenal homeobox 1 |
| PGBD5 | piggyBac transposable element derived 5 |
| PHOSPHO1 | Phosphoethanolamine/phosphocholine phosphatase |
| POU2F1 | POU Class 2 Homeobox 1 |
| PRC2 | Polycomb repressor complex protein 2 |
| PTPRO | Receptor-type tyrosine-protein phosphatase O |
| PYCR1 | Pyrroline-5-carboxylate reductase 1 |
| qPCR | Quantitative polymerase chain reaction |
| Rab4a | Ras-related protein 4a |
| RELN | Reelin |
| RETN | Resistin |
| RHBDF2 | Rhomboid 5 homolog 2 |
| RNA | Ribonucleic acid |
| RNAi | RNA interference |
| rRNA | Ribosomal RNA |
| S.D. | Standard deviation |
| S.E.M. | Standard error of the mean |
| SAHA | Suberoylanilidehydroxamic acid |
| SAM | S-adenosyl methionine |
| SAS2 | Something about silencing 2 |
| Sat2 | Satellite 2 |
| SB | Suicidal behaviour |

| | |
|---|---|
| SCAN | Schedule for Clinical Assessments in Neuropsychiatry |
| SCZ | Schizophrenia |
| SETDB1 | Histone-lysine N-methyltransferase SETDB1 |
| sgRNA | Single guide RNA |
| SIR2 | NAD-dependent histone deacetylase SIR2 |
| SIRT1 | Sirtuin 1 |
| SKA2 | Spindle And Kinetochore Associated Complex Subunit 2 |
| SLC1A2 | Solute carrier family 1 member 2 |
| SLC26A11 | Solute carrier family 26 member 11 |
| SLC6A4 | Solute carrier family 6 member 4 |
| SNP | Single nucleotide polymorphism |
| snRNA | Small nuclear RNA |
| SNRPN | Small nuclear ribonucleoprotein-associated protein N |
| SOCS3 | Suppressor of cytokine signaling 3 |
| SREBF1 | Sterol regulatory element binding transcription factor 1 |
| SSHELF | +2 to +4 kilobase region flanking the CpG island |
| SSHORE | +2 kilobase region flanking the CpG island |
| SSU | SCAN Suicidality |
| ST6GALNAC1 | Alpha-N-acetylgalactosaminide alpha-2,6-sialyltransferase 1 |
| SUMO | Small ubiquitin-related modifier protein |
| SYNE1 | Spectrin repeat containing nuclear envelope protein 1 |
| TAFII250 | Transcription initiation factor TFIID subunit 1 |
| TBC1D16 | TRE2/BUB2/CDC16 domain family member 16 |
| Tet | Ten eleven translocation protein family |
| TRANK1 | Tetratricopeptide repeat and ankyrin repeat containing 1 |
| TRIM67 | Tripartite motif containing 67 |
| TrkB | Tropomyosin receptor kinase B |
| tRNA | Transfer RNA |
| TRPM7 | Transient receptor potential cation channel subfamily M member 7 |
| TSS1500 | 1500 nucleotides upstream of the transcription start site |
| TSS200 | 200 nucleotides upstream of the transcription start site |
| TXNIP | Thioredoxin interacting protein |
| UBE3A | Ubiquitin-protein ligase E3A |
| UHRF1 | Ubiquitin-like plant homeodomain and RING finger domain 1 |
| Xist | X-inactive specific transcript |
| μg | Microgram |

# 1 Literature Review

# 1.1 Epigenetics

## 1.1.1 Overview

Epigenetics is the study of heritable alterations to gene regulation and function that do not involve corresponding changes in the deoxyribonucleic acid (DNA) sequence. Although the term derives from 17th century usage (epigenesis, meaning “extra growth”), it was first coined by British embryologist Conrad Waddington in the 1940s, and has undergone several iterations of meaning to fit current paradigms of molecular biology (Felsenfeld, 2014). The transmission of epigenetic modifications was once thought to be limited to the mitotically dividing cells in a single organism, but a growing body of evidence suggests that this information can be transmitted from one generation to the next in simple organisms such as Drosophila and yeast, but also, to some extent, in mammals (Daxinger & Whitelaw, 2012).

Epigenetic factors are thought to function as a “second code” that, through a variety of mechanisms, have the capacity to regulate important local functions such as gene expression, DNA repair and imprinting, as well as global functions such as X chromosome inactivation and developmental programming. Epigenetic factors are highly specific to cells and tissues, and are important in driving and determining the identity and fate of genetically identical cells (Moosavi & Ardekani, 2016). In this way, epigenetics represents a bridge between genotype and phenotype, and can be a useful additional layer to account for outward manifestations of DNA-based factors. Epigenetic factors are common to all organisms, including invertebrates, plants and mammals. However, this review will focus on the field’s applicability to higher organisms, with a focus on complex human disorders and aging.

## 1.1.2 Epigenetic modifications of DNA

### 1.1.2.1 5-methylcytosine

The most widely studied category of DNA modification is known as DNA methylation, and involves covalent addition of a methyl group (-CH3) to the fifth carbon of cytosine from S-adenosyl methionine (SAM), producing 5-methylcytosine (5mC) (Robertson, 2005). In complex organisms, DNA modification occurs most commonly at a dinucleotide sequence of cytosine followed by guanine in the 5’ to 3’ direction, otherwise known as a CpG site (Lister et al., 2009).

CpG dinucleotides represent about 1% of the genome, and their distribution varies dramatically. Areas that are enriched in CpG dinucleotides, including about 70% of all gene promoter regions, are known as CpG islands. These regions, approximately 1 kilobase (kb) in length, are typically found directly upstream of gene promoters, but are also found within and in the absence of genes. Directly flanking CpG islands in promoter regions are CpG shores (<2 kilobases outward), CpG shelves (2-4 kilobases outward) and seas (>4 kilobases outward) (Sandoval et al., 2011).

The process of DNA methylation is orchestrated by the structurally and functionally distinct DNA methyltransferase (DNMT) family of enzymes. Two such enzymes, DNMT3A and DNMT3B are responsible for establishing the initial methylation profile (*de novo* DNA methylation) during epigenetic reprogramming of germ cells and in early embryogenesis (Cheng & Blumenthal, 2008; Okano, Bell, Haber, & Li, 1999). *De novo* DNA methylation is regulated with the aid of DNMT3-Like (DNMT3L), a protein that does not have a DNA-binding domain, but colocalizes with DNMT3A and DNMT3B. The tight-knit nature of these enzymes was confirmed in germ cells and retrotransposons, whereby mice carrying either a *Dnmt3L* knockout or *Dnmt3a* germ-cell conditional knockout, carried the same alterations in *de novo* methylation patterns (Bourc'his, Xu, Lin, Bollman, & Bestor, 2001; Kareta, Botello, Ennis, Chou, & Chedin, 2006).

By contrast, DNMT1 functions primarily in maintaining methylation marks following cell division (maintenance methylation). In order to accomplish this role, the DNMT1 protein has a 30-40-fold greater affinity for hemimethylated sites, that only carry methylation on one strand. The function of DNMT1 hinges upon its interactions with the accessory protein, ubiquitin-like plant homeodomain and RING finger domain 1 (UHRF1) (Bashtrykov, Jankevicius, Jurkowska, Ragozin, & Jeltsch, 2014).

There exists another DNMT homolog, DNMT2, the function of which is still poorly understood. Through several studies, it has been noted that DNMT2 is not required for either *de novo* or maintenance methylation, and its methylation activity was not detected *in vitro*. Thus, it has not been shown to play a role in embryogenesis or cell development. However, methylation activity of DNMT2 was demonstrated in mouse, *Arabidopsis thaliana* and *Drosophila*, in the

context of transfer ribonucleic acid (RNA) (Goll et al., 2006). These findings suggest that the role of DNMT2 may be limited to highly specialized functions.

DNA methylation has also been shown to occur at non-CpG sites, known as CpH sites, in cells undergoing development, embryonic stem cells and in the brain, as well as in plants, and non-mammals, though this phenomenon is less well understood (Guo et al., 2014).

#### 1.1.2.2 5-hydroxymethylcytosine and DNA demethylation

Following 5mC, the most prominent DNA modification is 5-hydroxymethylcytosine (5hmC), characterized by the addition of a hydroxyl group (-OH) to the methyl group of 5mC. This is accomplished through the oxidation of 5mC by the enzymatic action of the ten eleven translocation (Tet) family of proteins. 5hmC represents 5-10% of all cytosine modification in the genome, largely enriched in brain tissue, and is indistinguishable from 5mC using conventional methods (Tahiliani et al., 2009). However, this has recently been overcome through a variety of techniques that covalently modify the hydroxyl group for later recovery.

5hmC has garnered substantial interest in its connection to DNA demethylation. Given early discoveries of DNMTs, it was previously believed that DNA demethylation was a passive process. However, in the past decade, it has been established that 5hmC is the first intermediate in this pathway (Ito et al., 2011). 5hmC is enriched in neuronal tissue, and also in embryonic stem cells, where it has been associated with simultaneously permissive and repressive histone modifications (described in detail below), indicative of a developmentally poised state (Ficz et al., 2011; Ito et al., 2010; Khare et al., 2012). Moreover, 5hmC has been found to bind several methylation-specific enzymes, including UHRF1, to prevent methylation in specific regions (Frauer et al., 2011). Although the precise functions of 5hmC are poorly understood, these findings suggest that 5hmC plays a role in mediating gene expression and epigenomic maintenance, and is unlikely to simply be an intermediate in the DNA demethylation pathway.

#### 1.1.2.3 Other DNA modifications

Two additional modifications catalyzed from 5mC via Tet proteins are carboxylcytosine (5caC) and formylcytosine (5fC). Much like 5hmC, 5caC and 5fC have garnered interest as additional

intermediates in the DNA demethylation pathway. These marks are identified by thymine DNA glycosylase, which subsequently triggers the base excision repair pathway that replaces the recognized base with an unmodified cytosine (He et al., 2011; Ito et al., 2011; Kohli & Zhang, 2013). However, these modifications are indistinguishable from unmodified cytosines using existing methods. Moreover, given that 5caC and 5fC are found in substantially lower amounts even compared to 5hmC has added additional barriers in the understanding of their functional roles. In our studies, we did not distinguish between the modifications outlined above; thus, unless otherwise indicated, the term "DNA modification" will be used henceforth.

#### 1.1.2.4 Functions of DNA modification

DNA modification is a major contributor to genomic regulation, and is responsible for several functions, such as establishment of cell identity, disease development and mediation of gene expression (P. A. Jones, 2012). DNA modification in gene regions serves both repressive and permissive functions. Notably, modification of promoter regions is strongly associated with gene expression silencing, thought to be due to the blocking of positive transcription factors from binding to this region and initiating expression. Repression can also occur through recruitment of proteins that recognize modification marks. For instance, 5mC is recognized by methyl-CpG binding domain (MBD) proteins that in turn associate with inhibitory protein complexes. Three such proteins, Methyl-CpG binding protein complex 2 (MeCP2) and MBD2 can recruit histone deacetylase (HDAC) proteins, while MBD1 recruits Histone-lysine N-methyltransferase SETDB1 (SETDB1), a histone methyltransferase, to facilitate transcriptional repression (Klose & Bird, 2006). By contrast, DNA modification within the gene body has been positively associated with transcription, presumably through regulation of splicing. Although CpGs in these regions exist in low abundance, gene bodies containing CpG islands that likely served as promoters in early development. The modification of gene body CpG islands serves to promote preferential initiation of transcription (Illingworth et al., 2010).

CpGs found outside of islands are largely modified in somatic cells, particularly those contained in repetitive elements. Extensive modification of sequences such as retrotransposons, long interspersed nuclear elements, and satellites, serves to repress their action and promote genomic stability. CpH modification has also been shown to be present in abundance in both

cultured and induced pluripotent stem cells, as well as in embryonic stem cells (Guo et al., 2014). CpH modification is also enriched in mouse and human brain tissue, and accounts for about one quarter of all modification in adult mouse dentate neuron and adult mouse cortex. The bulk of this modification was not found near active genes or regions of high CpG density. Interestingly, inter-species conservation was noted in CpH modification patterns, whereby CpH levels were similar in human and mouse brain DNA at a majority (83%) of orthologous genes (Varley et al., 2013). CpH methylation was also shown to repress neuronal transcription in mice, likely through MeCP2 binding. Although it is not known how CpH modification is regulated, there is some evidence to suggest that it relies in part upon active lifetime maintenance by DNMT3A (Sharma, Klein, Barboza, Lodhi, & Toth, 2016). These findings underline the important but still poorly understood regulatory role of CpH methylation, particularly during developmental periods, and in the brain. Further investigation will help elucidate the regulatory role for CpH, particularly in genome regions devoid of CpGs, on neuronal function, plasticity and cognition (McGowan, 2016).

Further, numerous types of non-coding DNA sequences located near coding regions have been shown to play a substantial role in regulating transcription. Understanding the DNA modification profiles of these *cis*-regulatory elements (such as enhancers, insulators and silencers) has been important to uncovering the epigenetic contribution of gene regulation (Wittkopp & Kalay, 2011). First, enhancer sequences serve to promote transcription through direct or indirect binding of transcription factors; DNA modification of enhancers has been shown to inhibit this effect in numerous tissues (Ong & Corces, 2011). Moreover, differential modification of enhancers has been shown to be a successful predictor of cell lineage in breast cancer (Fleischer et al., 2017). Second, insulators are intergenic sequences that direct preferential gene expression by binding CCCTC-binding factor (CTCF), a protein that mediates looping of DNA and can promote or prevent crosstalk between enhancers and promoters. Modification of CTCF binding sites such as insulators can prevent CTCF recruitment, and contribute to transcriptional regulation; such modification is instrumental to genome-wide epigenetic process such as parental epigenetic imprinting (Bell, West, & Felsenfeld, 1999; Filippova et al., 1996; Hark et al., 2000; Lobanenkov et al., 1990).

### 1.1.2.5 Long non-coding RNA elements and DNA modification

Long non-coding RNA (lncRNA) elements are a group of transcribed RNA molecules that have a length of 200 or more nucleotides and are not translated into proteins. This diverse class of elements likely spans into the thousands in the human genome, and has been increasingly shown to play important regulatory functions in gene regulation, particularly when co-expressed with a nearby protein-coding gene (Cabili et al., 2011; Derrien et al., 2012). The action of lncRNAs has been found to be both repressive and permissive of nearby gene expression, depending on the nature of their interactions with other proteins. For instance, the interaction of an lncRNA element known as X-inactive specific transcript (*Xist*) with Polycomb repressor complex protein 2 (PRC2) is known to target loci for epigenetic silencing during X-inactivation (da Rocha et al., 2014). LncRNA elements can also interact with DNA modification machinery. For example, *Dali* is a lncRNA expressed in the central nervous system important for neural differentiation. *Dali* has the ability to inhibit distal promoter DNA modification through interaction with DNMT1, and its depletion is associated with enrichment of DNA modification. Expression of *Dali* was also associated with active histone modification marks, likely through histone interaction with the element (Chalei et al., 2014). Extra-coding CEBPA (*ecCEBP*) is an activating lncRNA that is co-expressed with the CCAAT/enhancer-binding protein alpha gene (*CEBPA*), the structure of which can interface with the active domain of DNMT1 and suppress DNA modification at gene regions (Di Ruscio et al., 2013).

Further, lncRNAs, like genes, are under epigenetic control. For example, a mouse model of breast cancer identified 69 differentially methylated loci found in the promoter regions of lncRNAs (Heilmann et al., 2017). One of these lncRNAs, *Esrp2-antisense* is hypomethylated in cancer tissue, and its co-expression with epithelial splicing regulatory protein 2 (ESRP2) is associated with over-proliferation, carcinogenesis and poor patient prognosis. Another differentially expressed lncRNA, DNMT1-Associated Colon Cancer Repressed LncRNA 1 (*DACOR1*) was found to be repressed in colon cancer cell lines (Merry et al., 2015). By contrast, overexpression of *DACOR1* was associated with accumulation of DNA modification at numerous loci. This is supported by findings that this element interacts with over 300 loci, several of which were differentially modified in colon cancer cells compared to control. Like other elements, *DACOR1* can interact with and target DNMT1 to methylate specific sequences.

The relative amounts and identity of expressed lncRNAs has important functional consequences, and regulation of lncRNA expression (epigenetic or otherwise) impacts how genes are regulated and may play a role in various disease states.

## 1.1.3 Histones and chromatin

### 1.1.3.1 Overview

The highly-condensed association of DNA and proteins is called chromatin, a complex connection subject to tremendous amounts of regulation that ultimately regulates and impacts genomic function. This tightly-condensed structure serves not only to protect DNA from damage, but also to dictate accessibility of DNA to transcription factors and other regulatory proteins in a fashion specific and appropriate to a given cell-type and stage of the cell cycle.

Chromatin thus exists as a dynamic entity that is subject to remodeling either condenses or loosens portions of the genome for silencing and expression. This can occur on a global scale through the action of adenosine triphosphate (ATP)-dependent chromatin remodeling complexes that have the ability to modify the position and density of DNA and its associated proteins. This process is crucial to key processes such as cell-cycle progression, development, DNA replication and chromosome segregation; and to cell-type and -stage appropriate gene expression. Chromatin can exist in one of two states: heterochromatin describes a tightly-bound configuration of histones and DNA, which prevents accessibility of transcription factors to the genome, and is associated with inhibited gene expression; by contrast, euchromatin is described by a loose association between chromatin and DNA, and is associated with transcriptional activity and gene expression (Bannister & Kouzarides, 2011).

### 1.1.3.2 Histones

Histones are a family of proteins that serve to condense and store DNA in the nucleus. The basic functional unit of chromatin is the nucleosome, made up of an octamer of two copies of histone proteins H2A, H2B, H3 and H4 condensed with 147 base pairs of DNA. Nucleosomes are joined together through the association of 80 base pairs (bp) of DNA and a linker histone protein called

H1. Histone proteins contain positively charged amino terminal chains that protrude outside the nucleosome. By nature of this conformation, these tails are subject to extensive modification in response to normal developmental and physiological factors, as well as from external factors, such as environment and lifestyle. Such modifications impact the physical association of DNA and histone proteins and can, in turn, impact the DNA modification profile and activity of genes and other sequence-based elements, and also play a significant role in DNA repair processes (Cutter & Hayes, 2015).

#### 1.1.3.3 Histone acetylation

Histone acetylation is a post-translational histone modification involving the addition of an acetyl group from Acetyl coenzyme A to amino groups of lysines (9, 14, 18, 23 and 46 on H3; 5, 8, 12 and 16 on H4) and arginines (2 on H3; 3 on H4) on N-terminal tails of histones. This addition is catalyzed through the action of one of several types of histone acetyltransferase (HAT) enzymes, and occurs on all four core histone proteins. HATs are sorted into five families of proteins based on presence of functional domains: Gcn5-related N-acetyltransferases (GNATs), (MOZ, Ybf2/Sas3, Sas2 and Tip60)-related HATs (MYST), p300-CREB-binding protein HATs (p300/CBP), transcription initiation factor TFIID subunit 1 (TAFII250) and nuclear receptor coactivators (Hodawadekar & Marmorstein, 2007). Histone acetylation can be removed via hydrolytic action of HDAC enzymes, grouped into four classes (Yang & Seto, 2007). Moreover, HDAC activity can be reversed through the action of HDAC inhibitors; these enzymes play an important role as regulators of synaptic plasticity, mood and memory formation, and are commonly used as treatments for neurodegenerative and psychiatric disorders. On a global scale, the interplay of HATs and HDACs has important implications for cell differentiation, cell cycle, apoptosis, DNA repair and transcriptional activation (Bannister & Kouzarides, 2011).

#### 1.1.3.4 Histone methylation

Histone methylation is another post-translational histone modification, and involves the covalent addition of a methyl group from S-adenosylmethionine to arginine and lysine residues of histone proteins. Arginine residues can be monomethylated or either symmetrically or asymmetrically

dimethylated, while lysine residues can carry one to three methylation marks. Histone methylation is mediated by three classes of enzymes: the SET domain protein lysine methyltransferases are involved in methylation of lysines 4, 9, 27 and 36 on H3 and lysine 20 on H4, while non-SET domain-containing methyltransferases govern methylation of lysine 79 on H3. Further, arginine methyltransferases modify arginines 2, 17 and 26 on H3 and arginine 3 of H4 (Lan & Shi, 2009; Lorenzo & Bedford, 2011).

Histones can also undergro demethylation through the action of various protein complexes. Lysine-specific demethylase 1 (LSD1) functions to remove methyl groups from lysines 4 and 9 on H3, while the Jumonji C domain containing protein family (JMJD2) demethylates lysines 9 and 36 on H3 (Klose & Zhang, 2007; Mosammaparast & Shi, 2010). Arginine demethylation occurs through the action of peptidylarginine deiminase 4 (PAD4), which can remove methyl groups from arginines 2, 8, 17 and 26 on H3, and arginine 3 on H4 (Wang et al., 2004).

Histone methylation can serve to either promote or repress gene expression, depending on the effect of the modification on the ability of DNA to dissociate from the nucleosome.

#### 1.1.3.5 Other histone modifications

In addition to histone acetylation and methylation, there exist several other known variants of histone modification with functional significance: proline isomerization, sumoylation, ubiquitination, ADP-ribosylation, phosphorylation; the latter four are, like acetylation and methylation, post-translational modifications (Bannister & Kouzarides, 2011).

First, proline isomerization involves the switching of cis-trans conformation in proline residues, and is mediated by spontaneous action of proline isomerase enzymes. This process is known to occur on prolines 30 and 38 on the tails of histone H3 proteins; proline 38 isomerization is thought to mediate the activity of Set2 in the methylation of H3 lysine 36 (Nelson, Santos-Rosa, & Kouzarides, 2006).

Second, sumoylation is the addition of a short polypeptide chain, known as a small ubiquitin-related modifier protein (SUMO), and is known to occur on lysines 6, 7, 16 and 17 of

H2B, lysine 126 of H2A and on all four lysines of the N-terminal tail of H4. This modification has a repressive effect on transcription, as it is known to oppose permissive marks such as acetylation and ubiquitination, likely through the recruitment of histone deacetylases and heterochromatin proteins (Shiio & Eisenman, 2003).

Third, ubiquitination involves covalent attachment of one or more strands of ubiquitin, a short polypeptide, to the ε-amino group of lysine. Addition of more than one ubiquitin strand tags proteins for degradation; by contrast, monoubiquitination of histone proteins is associated with transcriptional activity. For instance, ubiquitination of lysine 119 of H2A, and lysine 120 of H2B are marks of transcriptional activation found corresponding to active genes. Ubiquitination has also been shown to mediate other histone modifications; notably binding of histone deacetylase 6 to a ubiquitin strand on H2B lysine 123 was shown to be necessary for H3K4 and H3K79 methylation, and maintenance of a euchromatic state (Kim, Kim, Uchiki, Gygi, & Goldberg, 2009; M. G. Lee et al., 2007; Wang et al., 2004).

Fourth, ADP-ribosylation consists of the addition of one or greater ADP-ribose molecules, and can impact all four core histone proteins and the linker protein H1. This phenomenon generally underlies DNA damage, and addition of multiple such molecules serves to remove histones from the nucleosomes to allow for DNA repair to occur (Hassa, Haenni, Elser, & Hottiger, 2006).

Fifth, phosphorylation involves addition of a phosphate group through the action of kinases, and its removal via phosphatases. This occurs namely on serine 10, serine 139 and tyrosine 142 of an H2A variant called H2AX, and on serines 18, 173, 189 and tyrosine 11, 138 and 155 of H1 and H4. This modification is a component of chromatin condensation, regulation of expression and DNA repair, and is thought to be regulated through interaction between phosphate group and histone kinases/phosphatases (Oki, Aihara, & Ito, 2007).

#### 1.1.3.6 Histone modification, DNA modification and gene expression

DNA modification, and consequently gene expression are tightly-associated with modification state of corresponding histone proteins, and understanding this interplay underlies crucial aspects of chromatin dynamics and genomic function. As evidenced by the Roadmap Epigenomics

Project, enrichment of histone modification marks adds cell-, tissue- and species-specific functional significance to loci (ie. promoter activity, enhancer status), and provides a greater understanding of the genetic regulation of those regions (Kundaje et al., 2015).

Histone acetylation results in the conversion of a positive charge on nitrogen to a neutral one, resulting in dissociation of protein from DNA; this modification is thus typically associated with relaxed chromatin state, increased accessibility of transcription factors to DNA, and increased gene expression. Acetylation of H3 lysine 27 (H3K27Ac) is one such permissive mark, that is typically associated with active enhancer elements, and is found proximally and distally adjacent to transcription start sites (Tie et al., 2009). Similarly H3 lysine 9 acetylation (H3K9Ac) is a hallmark of active promoter regions (Tjeertes, Miller, & Jackson, 2009). By contrast, histone deacetylation is correlated with DNA methylation and transcriptional repression, through the recruitment of HDACs by MBD proteins. H3 lysine 36 acetylation (H3K36Ac) is also associated with transcriptional activation (Morris et al., 2007).

Histone methylation marks play varied roles and can be grouped on the basis of transcriptional activation or silencing. H3 lysine 9 trimethylation (H3K9me3) is positively correlated with DNA modification and transcriptional silencing, whereby MBD1 complexes with histone methyltransferases (Kondo, 2008). DNA modification may also be maintained through the interaction of DNMT1 and its associated protein UHRF1, with histone methyltransferases. By contrast, H3 lysine 4 trimethylation (H3K4me3) promotes gene activation through recruitment of transcription and chromatin remodeling factors (Zhou & Zhou, 2011). Further, H3 lysine 27 trimethylation (H3K27me3) is associated with gene repression by promoting remodeling into heterochromatin. Paradoxically, this type of modification does not coexist with DNA modification despite having a similar function (Ku et al., 2008). As described above, this mark can co-occupy a position with permissive H3K4me3, often to indicate a developmentally poised state, primed for future transcription. H3 lysine 36 trimethylation (H3K36me3) has numerous roles: it is found in exon regions and at the 3’ end of active genes, but is also found in abundance in heterochromatin (Barski et al., 2007; Guenther, Levine, Boyer, Jaenisch, & Young, 2007).

### 1.1.4 Epigenetic modifications of RNA

In addition to modifications to DNA and histone proteins, RNAs can also be subject to epigenetic modification. Though less attention is given to epigenetic modification of RNA compared to DNA and histones, the diverse roles of RNA underscore the need to consider RNA modification in understanding genomic regulation and cell function. Perhaps the most widely studied RNA modification is adenosine methylation, resulting in N6-methyladenosine (m6A). This primary target of adenosine methylation is messenger RNA (mRNA), but this phenomenon also occurs in transfer RNA (tRNA), ribosomal RNA (rRNA), small nuclear RNA (snRNA) and numerous species of lncRNA. m6A is created through the action of the N6-adenosine-methyltransferase subunit 3 (METTL3) protein, and is thought to be a target of recognition by downstream proteins. The presence of this mark is associated with a number of phenomena, including alternative splicing of mRNA, altered gene expression and cell development, and modification of cellular circadian period. Unsurprisingly, studies of m6A have implicated this modification in aberrant metabolism, embryonic development and cancer progression (Cui et al., 2017; Dai, Wang, Zhu, Jin, & Wang, 2018; Huisman, Manske, Carney, & Kalantry, 2017). Although it will not be considered further here, it is clear that RNA modification remains an active and ever-expanding topic of interest and importance.

### 1.1.5 Technologies for investigation of epigenetic modifications

#### 1.1.5.1 Overview

Several assays have been widely used to study DNA modification, such as methylation-sensitive restriction enzyme mapping, microarrays, as well as numerous methods involving bisulfite conversion, antibodies, enrichment and sequencing. In this section, I will discuss the methods primarily used in my studies, their function, and advantages and disadvantages of each.

#### 1.1.5.2 Microarray-based methods

Microarrays permit investigation of several hundred thousand predesigned loci in a cost-effective and high-throughput fashion. Each microarray consists of a chip containing two sets of probes

complementary to various genomic or methylomic regions, that fluoresce upon binding of fragmented input DNA sequences. The readout can provide information regarding sample genotype, DNA modification state or gene expression profile. The most popular microarray platform for methylome-wide investigation in the past several years has been the Illumina Infinium Human Methylation 450K BeadChip, which can assess DNA modification state at approximately 450,000 sites (Dedeurwaerder et al., 2011).

We have also made use of an enrichment-based technique known as methyltransferase-directed transfer of activated groups (mTAG), to interrogate DNA modification on a chromosome-wide level. mTAG uses a modified methyltransferase that tags and enriches for the unmethylated fraction of the genome, a much smaller proportion of the methylome. The product is then interrogated using tiling microarrays, containing contiguous sequences, and DNA modification state is inferred from the resulting data. mTAG permits profiling of a large proportion of the genome not covered by other methods, but is ultimately limited in its resolution compared to targeted bisulfite sequencing (Kriukiene et al., 2013).

Microarray-based methods allow for extensive screening, but are limited to loci contained on the chip; these loci are generally adjacent to or within coding regions, and do not represent the full spectrum of genomic complexity. Thus, microarrays are best used in exploratory studies to gain a "snapshot", rather than a comprehensive understanding of DNA modification state.

#### 1.1.5.3 Sequencing-based methods

The advent of cost-effective next generation sequencing has enabled methylomic profiling in a high-throughput and high-resolution manner. In our studies, investigation of DNA modification has relied heavily upon the use of bisulfite-sequencing using the padlock probe method. As in other bisulfite-based methods, input DNA is treated with sodium bisulfite, a compound that converts unmodified cytosines to uracil through deamination. The product subsequently anneals to and is amplified with up to several hundred thousand custom probes designed for each locus of interest, and prepared for sequencing. The sequencing data is later aligned to a reference genome to infer modified sites. This method is an improvement over others in that it allows for highly-targeted investigation of DNA modification (including non-CpG sites) and can

simultaneously accommodate a large number of samples using unique barcodes (Diep et al., 2012).

#### 1.1.5.4 Chromatin immunoprecipitation

Chromatin immunoprecipitation (ChIP) is a widely-used technique for investigating specific histone modifications associated with a locus of interest. In brief, tissue cross-linked with formaldehyde is treated with antibodies specific to one or more histone modifications. After protein capture, cross-links are removed and the sample is purified such that only DNA linked to the targeted histone modification remains. This DNA can be subjected to quantitative polymerase chain reaction (qPCR) or other sequencing methods for identification. ChIP is advantageous for identifying specific types of histone modification and has numerous applications depending on the loci of interest (ChIP-qPCR, ChIP-seq, ChIP-on-chip). However, in addition to requiring large amounts of input cells, ChIP is traditionally a qualitative technique, making precise comparisons of histone modifications between samples, especially rare cell-types, difficult (Milne, Zhao, & Hess, 2009).

### 1.1.6 Features of epigenetic processes

#### 1.1.6.1 Overview

A hallmark of epigenetic mechanisms is their dynamic and malleable nature; epigenetic profiles of cells can change dramatically over a lifetime, and also be altered over shorter periods of time (Kanherkar, Bhatia-Dey, & Csoka, 2014). Long-term epigenetic changes accompany aging and rely upon the complex interaction of several processes over time, and the effects of imprinting and developmental reprogramming. By contrast, short-term epigenetic changes can originate from environmental, dietary and pharmacologic exposure, and can be reversible. Thus, epigenetic modifications may serve as useful biosensors of external contribution, and have potential to be used as part of a treatment or intervention (Tammen, Friso, & Choi, 2013).

### 1.1.6.2 Diet and environmental exposure

The fields of environmental and nutritional epigenetics have been active in the past decade in the search for environmental, lifestyle and nutritional contributors to phenotype that are mediated by epigenetic factors.

Many nutritional and environmental epigenetic studies have described changes thought to stem from exposure to compounds that can reduce the available level of the SAM, a key methyl donor (Chiang et al., 1996). The most widely-studied dietary sources of methyl donors is folate (vitamin B9), enriched in many grains and legumes. Folate is known to maintain DNA modification, while vitamins B6 and B12 are essential for maintaining the availability of SAM (Kim, 2005; McKay & Mathers, 2011; Pufulete et al., 2005).

The yellow agouti mouse model is an earlier example of a diet-induced phenotypic change that is mediated by epigenetic factors. The mutant model contains an insertion of a retroelement-like intracisternal A particle (IAP) into the Agouti gene, that is transiently and ectopically expressed in numerous cell types. In contrast to its wild-type counterpart, the mutant agouti mouse has a yellow coat colour, and metabolic dysfunction that features hyperinsulinemia. This viable mutant gene, known as $A^{vy}$, is a metastable epiallele that is variably expressed on the basis of its epigenetic modification state, such that the degree of modification of the long terminal of the insertion mediates hair cycle-specific and ectopic expression of the mutant transcript, and determines coat colour (yellow to the wild-type-like pseudo-agouti colour), and metabolic symptoms. Interestingly, maternal supplementation with methyl donors and soy-derived isoflavone induced changes in offspring phenotypic distribution toward predominantly pseudo-agouti and facilitated escape from metabolic symptoms. These changes were shown to be associated with increased DNA methylation of IAP that persisted into adulthood. Moreover, this effect was conserved in three germ layers, suggesting early epigenetic effects at work (Dolinoy, 2008).

Far-reaching epigenetic alterations have reported as a product of lifestyle. First, in utero exposure to maternal smoking behaviour associated with reduced methylation of repeat sequences including Satellite 2 (Sat2) and Alu, and increased methylation in genes encoding Tyrosine-protein kinase receptor UFO (*AXL*) and Receptor-type tyrosine-protein phosphatase O (*PTPRO*) and Insulin-like growth factor 2 (*IGF2*) (Breton et al., 2009; Breton, Salam, &

Gilliland, 2011; Flom et al., 2011; Murphy et al., 2012). Moreover, adult smoking behaviour and second-hand smoke were found to be associated with increased methylation of three genes: Cyclin Dependent Kinase Inhibitor 2A (*p16*), O-6-Methylguanine-DNA Methyltransferase (*MGMT*) and Death Associated Protein Kinase 1 (*DAPK*) (Leng et al., 2012). Second, exposure of mouse embryos to alcohol was associated with epigenetic aberrations at dozens of genes, many of which have human homologs associated with neurodegenerative and neurodevelopmental disorders such as Alzheimer's (Amyloid beta precursor protein, *APP*), Parkinson's (Transient receptor potential cation channel subfamily M member 7, *TRPM7*), and Angelman syndrome (Ubiquitin-protein ligase E3A, *UBE3A*) (Liu, Balaraman, Wang, Nephew, & Zhou, 2009). Third, in utero exposure to bisphenol alpha, a compound commonly found in plastics, was shown to induce DNA methylation and histone modification alterations, and increased offspring susceptibility to prostate cancer (Bromer, Zhou, Taylor, Doherty, & Taylor, 2010); interestingly, these effects were reversible upon treatment with methylation donors (Dolinoy, Huang, & Jirtle, 2007). Taken together, these studies suggest that studying epigenetic factors are potentially useful biomarkers for measuring the impact of lifestyle and environment.

#### 1.1.6.3 Transgenerational epigenetic inheritance

Upon fertilization, the epigenome is subject to two major reprogramming effects that erase epigenetic marks established during gametogenesis. In fact, about 70% of all DNA modification marks are removed in early development, a process that promotes cellular totipotency and prepares the budding embryo for differentiation into many different cell types. This is gradually followed by genome-wide *de novo* modification during subsequent maturation. However, some epigenetic marks established in mature gametes are partially stable, have the potential to resist programming, and may be maintained across cell divisions. It is thus postulated that incomplete removal of epigenetic modification during reprogramming allows for potential to transfer molecular information to the next generation. This statement has broad implications because, if true, it suggests that parental lifestyle, environment, exposure and behaviour can impact the health and disease susceptibility of offspring (Daxinger & Whitelaw, 2012).

One of the most well-documented transgenerational effects pertains to the so called Dutch Hunger Winter study. Exposure to prolonged starvation over seven decades ago had a

profound effect on individuals, and also induced lasting metabolic consequences on their descendants generations later. Individuals who were prenatally exposed to famine had lower birth weight and increased intolerance to glucose in adulthood. This effect was, in part, accounted for by lower than expected DNA modification levels at five CpG sites at *IGF2* (Heijmans et al., 2008).

Animal studies have been instrumental to the precise quantification of transgenerational effects with potential underlying epigenetic effects. Much of the attention has fallen on parental diet as a modifier of offspring epigenetic state. For instance, consumption of a high fat diet in rodents is consistently associated with offspring predisposed to metabolic dysfunction (Carone et al., 2010). A study in rats showed that lifetime paternal high fat diet results in decreased beta cell mass in offspring, decreased body weight and impaired response to the glucose tolerance test (de Castro Barbosa et al., 2016). This effect was found to be most pronounced in female offspring, and persisted for up to two generations. In support of these findings, differentially-methylated regions were identified in both parental and offspring rat sperm cells on the basis of parental dietary regime. Many of these regions mapped to genes related to metabolism and housekeeping. Also, several microRNA (miRNA) genes were found to be differentially expressed in the parents with a high fat diet. One of these, *let-7c*, is a mediator of glucose insensitivity and was found to be overexpressed following high fat diet (Jiang, 2016).

These findings were supported by another study that used *in vitro* fertilization to account for any gestational epigenetic effects (Huypens et al., 2016). In this investigation, gametes from C57/BL6 mouse parents fed a high fat diet were collected, fused and implanted into a control surrogates fed a control diet. Despite this alteration, parental diet nevertheless influenced susceptibility to metabolic dysfunction. Females exhibited greater tendency toward obesity if derived from gametes of parents fed a high fat diet. This effect was also observed in males, but only if derived from oocytes of mothers with a high fat diet. The authors also noted that DNA modification marks found in these offspring were conserved in tissues from various germ types, suggesting that the alteration was made during early gamete development. A third study noted that parental exposure to a high fat diet influenced metabolism in offspring, and resulted in increased female offspring body weight and impaired glucose tolerance and sensitivity (Ng et al., 2010).

Epigenetic inheritance has been met with some criticism and resistance. It has been suggested that major findings from studies such as the agouti mutant mouse model (described in the previous section) are confounded by the fact that genomic regions containing transposable elements may be resistant to epigenetic reprogramming, and that changes induced in the germ-line are not consistently maintained in somatic cells of offspring (Daxinger & Whitelaw, 2012). Despite some limitations however, there is evidence to support many epidemiological and population studies that point to parental diet as a modifier of offspring phenotype and susceptibility to illness. The focus of future epigenetic inheritance studies should be on solidifying the connection between parental exposure and changes in offspring epigenetic modification state, and to identify specific loci that can escape intergenerational reprogramming.

#### 1.1.6.4 Epigenetic imprinting

Epigenetic factors are responsible for genomic imprinting, the silencing of loci depending on the basis of parental inheritance (i.e. maternal or paternal) that affects about 1% of all mammalian genes (Baran et al., 2015). Imprinting of genes results in a parent-of-origin-specific epigenetic silencing, that involves both DNA and histone modification. Moreover, imprinting is maintained over many cell divisions and passed on to various cell type. In many regions, imprinting occurs at loci known as the imprinting control region (ICR) and results in expression only of the non-imprinted allele (Lawson, Cheverud, & Wolf, 2013).

A classic example of epigenetic imprinting is the study of the *H19/IGF2* locus (Gong, Pan, & Chen, 2010). *IGF2* encodes for a ligand instrumental for activating the IGF1 receptor and triggering a signal transduction pathway important for early development with important growth promoting effects on muscle, bone and brain tissue. By contrast, *H19* encodes for a lncRNA element that has growth inhibiting effects. In humans, both are found on chromosome 11 and are regulated by the imprinting status of an ICR that divides these two loci and harbors a methylation-dependent CTCF binding site. On the maternal allele, the ICR remains unmethylated, permitting CTCF binding that blocks *IGF2* and promotes *H19* expression. By contrast, the ICR is methylated on the paternal allele, preventing CTCF binding and permitting *IGF2* expression. *H19/IGF2* imprinting has important developmental and metabolic consequences. For instance, decreased *IGF2* expression is associated with predisposition to

obesity, while aberrant *H19* expression is linked to developmental disorders such as Silver-Russell Syndrome and Beckwith-Wiedemann Syndrome (Huang et al., 2012). Moreover, loss of imprinting at this locus is known to be involved in the pathogenesis of colon cancer (Leick, Shoff, Wang, Congress, & Gallicano, 2012).

Regulation of X-inactivation is another function of epigenetic imprinting. In females, during embryonic development, one X chromosome is fully inactivated at random. This inactivated mark is stably maintained, and propagates clonally following cell division. The inactivated X forms a discrete bundle of heterochromatin called a Barr body. X-inactivation is initiated through the action of lncRNA *Xist*. Expression of *Xist* initiates chromosome-wide silencing via the recruitment of PRC1 and PRC2, exclusion of RNA polymerase II from the chromosome, and recruitment of repressive histone modification marks such as H3K27me3. DNMT1 contributes to the subsequent maintenance of this inactive state. X-inactivation function to maintain proper dosage of gene expression, given that females carry twice the copies of every X-linked allele as males. Nevertheless, dozens of genes have been shown to escape X-inactivation in a tissue- and cell-specific manner; moreover, the inactivation process can sometimes be skewed toward one X chromosome over another (Minks, Robinson, & Brown, 2008; Tukiainen et al., 2017). X-inactivation skewing was found to differ between individuals, and gradually increase with age (Gentilini et al., 2012). Dysfunction of the inactivation process is associated with aberrations and complex disorders; for instance, *Xist* is known to be upregulated in numerous cancers.

Improper imprinting is also associated with disease, most commonly Angelman and Prader-Willi syndromes. Both of these are developmental and neurologic disorders with distinct clinical presentation, and are linked to aberrant imprinting on chromosome 15. Angelman Syndrome is characterized by delayed intellectual development, neurological and cognitive impairment, and specific facial appearance; it is derived from inheriting a defective copy of *UBE3A*, a member of the ubiquitin pathway, or the inheritance of two paternal (imprinted) copies (Bird, 2014). Prader-Willi Syndrome is characterized by delayed physical development, metabolic problems, muscular dysfunction, behavioural issues. Prader-Willi Syndrome is caused by deletions or defects of the paternal copies of the Small nuclear ribonucleoprotein-associated protein N (*SNRPN*) and Necdin (*NDN*) genes, and nearby small RNA elements; given that this

region is maternally imprinted, lacking functional paternal copies leads to dysfunction (Angulo, Butler, & Cataletto, 2015).

#### 1.1.6.5 Epigenetic therapy

The plastic characteristic of epigenetic modifications is conducive to the development of pharmacologic therapies to modify or undo genomic changes that underlie disease. Although use of such therapies is not yet widespread, several types of drug candidates have shown promise, particularly in the treatment of various cancers.

A major target of intervention is DNA modification. For instance, 5-Azacytidine has been used to treat myelodysplastic syndrome under its trademark name Vidaza. The compound is a cytidine analogue that, when incorporated into DNA, binds irreversibly to DNMT1 and causes global reductions in DNA modification (Momparler, 2005). This effect is most pronounced in cancer cells, which divide more rapidly than surrounding cells.

A second target of intervention is histone modification. In particular, HDAC inhibitors, such as suberoylanilidehydroxamic acid (SAHA), have been widely-used for treatment of T cell lymphomas (Zhang, Richon, Ni, Talpur, & Duvic, 2005). Further, this class of drugs has been investigated preclinically to combat Duchenne muscular dystrophy, memory loss, ischemia, anxiety and post-traumatic stress disorder, and amyotrophic lateral sclerosis (Ganai, 2014).

Use of pharmacologic agents with epigenetic targets has many side-effects, given the widespread action of such compounds. However, the modest success of such treatments provides optimism for future studies.

#### 1.1.6.6 Epigenetic studies of twins

Twin studies of disease have been useful to overcome the limitations of many disease studies. Given that monozygotic twins (MZ) share a nearly identical genomic sequence, twin studies can avoid many common confounders such as age, sex, genetic factors and environment (Zwijnenburg, Meijers-Heijboer, & Boomsma, 2010). Over a lifetime, monozygotic twins

diverge in the context of phenotype and underlying physiology (Haque, Gottesman, & Wong, 2009; Machin, 1996). Studies of many complex disorders have cited discordance, whereby one twin is affected and the other disease-free (Boomsma, Busjahn, & Peltonen, 2002). Concordance can be low, as in cancer (0-16%) (Lichtenstein et al., 2000), or moderate-to-high, as in type 1 diabetes (61%), type 2 diabetes (41%), autism (60%), and schizophrenia (58%) (Beckmann & Franzek, 2000; Condon et al., 2008; Lehtovirta et al., 2010; Wong et al., 2014). This discrepancy is likely due to many interactions, for instance between genes and environment, and suggests that genotype alone does not account for full phenotypic variation.

Monozygotic twins are also more epigenetically similar to each other than dizygotic twins (DZ), presumably due to shared genetic sequence and derivation from the same fertilized oocyte, and thus being subjected to the same epigenetic reprogramming events. However, epigenetic discordance is nonetheless prevalent in monozygotic twins. Monozygotic twins were first found to epigenetically differ in study of peripheral lymphocytes; pairs were more epigenetically different with increasing age, dissimilarity of environment, medical history and health (Fraga et al., 2005). This was confirmed in a landmark study that showed epigenetic differences in three different tissues (white blood cells, buccal epithelial cells and intestine) of monozygotic twins matched for age and sex (Kaminsky et al., 2009). The same study noted greatest epigenetic similarity at CpG islands and promoter regions.

Epigenetic studies of discordant monozygotic twins can provide clues into the molecular processes underlying disease, and areas of genome most affected (Jaenisch & Bird, 2003). For example, studies of breast cancer-discordant monozygotic twins identified over 400 differentially methylated regions between affected and unaffected individuals in whole blood (Heyn, Carmona, et al., 2013). One of these differences was hypermethylation of the alternative promoter in the docking protein 7 gene (*DOK7*) in affected individuals. *DOK7* encodes a docking protein that activates a receptor tyrosine kinase; this difference was noted several years before diagnosis, suggesting this change occurs as a precursor in early stages of cancers, and points to its usefulness as a diagnostic biomarker. Another study looked at MZ discordance for childhood leukemia and secondary thyroid carcinoma; the authors found differential methylation of breast cancer type 1 susceptibility (*BRCA1*) in primary fibroblast DNA (Galetzka et al., 2012). *BRCA1* is a gene highly implicated in various cancers, as well as with epigenetic mosaicism and lower

levels of *BRCA1* expression in the affected twin (Bianco, Chenevix-Trench, Walsh, Cooper, & Dobrovic, 2000; Esteller et al., 2000).

## 1.2 Aging

### 1.2.1 Overview

Aging is a complex biological process that refers to the progressive decline of physiological and organismal function throughout the span of a lifetime. On a molecular level, aging can be described as the gradual loss of molecular and cellular fidelity after reaching sexual maturity, that results in eventual loss of normal cell function, leads to degradation of tissues and physiology and, eventually, death (Lopez-Otin, Blasco, Partridge, Serrano, & Kroemer, 2013). Aging is a distinct process from senescence, a state characterized by constraints to cellular replication after a certain period of time (known as the Hayflick limit); unlike aging, senescence serves a protective function that prevents unrestrained cell division, particularly of cells that have accrued damage (Hayflick & Moorhead, 1961).

Aging has traditionally been viewed as a naturally-occurring process (Callahan & Topinkova, 1998; Hayflick, 2007). Indeed, development and maturation (accompanied by hormonal changes, bone growth, brain development) are all beneficial consequences of aging. Moreover, the degree of aging is characterized as "normal" for a given age and sex. Many pathological byproducts of aging are also deemed inevitable and "normal". For instance, *post-mortem* brain samples of individuals aged 80 years and above contain an abundance of amyloid plaques; many older men are anecdotally known to "die with prostate cancer rather than from it"; and older individuals commonly experience hardened arteries and atherosclerosis (Blumenthal, 2003; Kirkegaard et al., 2018). Nevertheless, given that the majority of illnesses are progressive and present with severe outcomes over time, aging is a major risk factor for complex diseases, such as cardiovascular disease, cancer, neurodegenerative disorders, and many others (Bulterijs, Hull, Björk, & Roy, 2015). The study of aging is therefore instrumental to understanding the decline of bodily integrity and function that accompanies disease, and uncovering mechanisms that can potentially serve as targets for intervention and treatment.

#### 1.2.1.1 Model organisms

Model organisms have been instrumental for identifying pathways linked to aging. Studies using budding yeast, *Saccharomyces cerevisiae*, have investigated replicative capacity and lifespan on

the basis of the number of mitotic divisions cells can undergo, and their respective ability for post-mitotic survival, respectively (Kaeberlein, Burtner, & Kennedy, 2007). Similar studies have been performed in invertebrate organisms such as *Caenorhabditis elegans* and *Drosophila melanogaster*, but have not been able replicate the complexity of the aging process in higher organisms (Brandt & Vilcinskas, 2013; Tissenbaum, 2012). Investigations of more complex organisms, such as mice and the African turquoise killifish, as well as human cell lines, have also provided some insights into the molecular underpinnings of human aging (Harel et al., 2015; Quarrie & Riabowol, 2004). Although these studies show that aging is a process that can be intervened upon, use of model organisms has not adequately captured the genetic diversity and complexity of the human aging process, and many cell-models have limited applicability to *in vivo* aging tissue, due to factors such as passaging and immortalization of certain cell lines (Neumann et al., 2010; Sen, Shah, Nativio, & Berger, 2016).

#### 1.2.1.2 Molecular and environmental mechanisms of aging

Aging has been characterized as the molecularly-driven accumulation of harmful factors that contribute to the gradual undoing of normal physiology, and the inability of endogenous repair mechanisms to prevent this effect. Although there are likely hundreds if not thousands of contributing factors, several have been consistently linked to aging. For instance, the oxidative stress hypothesis posits that the build-up of reactive oxygen species contributes to systemic damage and imposes a limit to the number of times a cell can divide (Hayflick, 1991). Such stress has been shown to be strongly associated with telomere shortening, another age-related phenomenon. Telomeres are condensed sequences at chromosome ends that prevent the degradation of DNA and shorten with each progressive round of cell division (Proctor & Kirkwood, 2002). Possessing telomeres shorter than expected for a given age is thought to be a sign of accelerated aging; shorter than expected telomere length has been associated with poor lifestyle factors (smoking and obesity), early onset cardiovascular disease and certain cancers (Samani, Boultby, Butler, Thompson, & Goodall, 2001; Shammas, 2011).

It is thought that accumulation of somatic DNA mutations over the course of a lifetime can also impact aging. However, it was recently noted in yeast that this accumulation is not primarily responsible for the aging process. By contrast, genetic predisposition to age-related

disorders is likely tied to the gradual misregulation of genes of numerous genes over time. In support of this, greater age has been linked to stochastic changes in gene expression activity and increased transcriptional variability between cells in a given tissue. For instance, it was noted that cardiomyocytes derived from older mice varied in transcript level of heart-specific genes compared to younger mice (Warren et al., 2007). Further, older human hematopoietic stem cells had higher overall transcript levels. Although aging outcomes likely derive from the additive effect of numerous genetic factors, deviations from transcriptional activity in youth may also be due to the increase of expression from repetitive elements that are generally silenced from early age. Increased expression of such elements, including long terminal repeats, long and short interspersed repetitive elements, was noted in aged mouse somatic cells (De Cecco, Criscione, Peterson, et al., 2013). Further, human genetic association studies have revealed that variants in genes responsible for cell maintenance and metabolism may be key to longevity. Resilient mechanisms such as DNA repair, telomere conservation, heat shock response and improved clearance of free radicals are thought to be associated with an improved or persistent ability to regulate and maintain physiology (Raule et al., 2014; Rose et al., 2012; Singh et al., 2006; Soerensen et al., 2012).

Given that aging is a broad and dynamic process, it can also be influenced through factors such as diet, lifestyle and environment, although the exact interplay between these factors and inherited predisposition is not well understood. One of the most widely-cited lifestyle interventions on aging is caloric restriction (Cournil & Kirkwood, 2001). Studies in both lower organisms and mice have demonstrated the effect of restricted caloric intake, reduced metabolic rate and upregulated stress response to be positively correlated with longevity (Bordone & Guarente, 2005; Migliaccio et al., 1999; Ruggiero et al., 2008). For instance, mice fed *ad libitum* are prone to increased weight gain and metabolic disease, an effect that can be largely mitigated through caloric restriction (Fontana, Partridge, & Longo, 2010). The lifespan-extending effect is thought to be due to a reduction in reactive oxygen species that are created as byproduct of metabolism (Sohal & Forster, 2014).

#### 1.2.1.3 Aging, health and disease

Aging is strongly linked to a decline in health and increased risk of susceptibility to illness, and is thus instrumental in the disease process. Several genetic disorders are marked by accelerated aging and reduced lifespan, such as Hutchinson-Gilford Progeria Syndrome, Werner Syndrome and Dyskeratosis Congenita. These genetically-based diseases lead to disorganization and deterioration of chromatin structure due to impairment of DNA repair, and are marked by considerable genomic instability (Arancio, Pizzolanti, Genovese, Pitrone, & Giordano, 2014; Burtner & Kennedy, 2010). Investigation of the mechanisms driving these disorders has been useful to understanding the normal aging process.

A major question surrounding disease pertains to the temporal component by which diseases unravel. The vast majority of complex diseases have an inherited risk and a strong familial component, although pathology does not appear until (sometimes much) later in life, and age of onset may differ dramatically even among family members. Despite investigating specific mechanisms mediating each disease process, delayed onset of complex diseases is a phenomenon that remains poorly understood. For instance, neuropsychiatric disorders such as bipolar disorder (BD) and schizophrenia (SCZ) do not manifest until late adolescence or early adulthood, whereas neurodegenerative disorders such as Alzheimer's disease and Parkinson's disease, do not manifest until after age 60 in the majority of cases, or later (Huff, Growdon, Corkin, & Rosen, 1987). By contrast, several childhood-onset disorders such as certain forms of asthma, epilepsy and attention deficit and hyperactivity disorder, have been shown to disappear by virtue of children "outgrowing their illness" (Freeman, 1987; Panhuysen et al., 1997; Shaw et al., 2007). It has been demonstrated that genetic factors play a role in expediting or delaying age of onset of these illnesses, but the exact processes behind this are unclear (Rodríguez-Rodero et al., 2011).

#### 1.2.1.4 Summary

It is clear from the literature that aging is a dynamic and highly intricate process that relies upon the complex interactions of genetic substrates, proteins and the environment, in four-dimensional space. Although genetic variants are associated with disease, health and longevity, it is difficult to account for this complex interplay on the basis of unchanging factors such as genetic

sequence. Thus, adequately capturing the temporal dimension of these processes hinges upon access to factors that mediate the interplay between DNA and phenotype. In the subsequent sections, I will discuss the previously-demonstrated overlap between epigenetics and aging, and propose its usefulness as a tool to understand the temporal component of complex genetic disorders, with a focus on BD. Despite aging being considered a normal and inevitable process, gaining an improved understanding of the molecular mechanisms behind it allow for potential for intervention upon this process, and prevention of compromises to overall health and well-being.

## 1.2.2 Epigenetics and aging

### 1.2.2.1 Overview

Epigenetic modifications are promising to the study of complex phenotypes because they are malleable and can change in response to internal and external factors, including genetic sequence, environment and stochasticity. Moreover, epigenetic modifications account for temporal and spatial dynamics of the epigenome, which is crucial for understanding aging (Brunet & Berger, 2014). This is underscored by the fact that case-control studies of human disease are typically matched for age, a complex metric that can provide insight into the effects of developmental mechanisms, internal physiological changes, and factors external to the individual.

Several epigenetic mechanisms have been proposed to underlie the aging process. Collectively, these mechanisms point to substantial epigenomic changes over the span of a lifetime that have profound implications for chromosomal integrity, genomic stability, gene regulation, and consequently, function (Pal & Tyler, 2016).

### 1.2.2.2 Age-dependent DNA modification changes

DNA modification changes have been widely associated with the aging process. Mammalian aging is generally associated with hypomethylation of CpGs, an effect that has consequences on gene regulation and genome stability. Loss of CpG modification in repetitive regions erodes the

maintenance of these elements, and is accompanied by enrichment of permissive histone modification marks (H3K9Ac, H3K27Ac, H3K4me1, H3K4me and H3K4me3) (Jung & Pfeifer, 2015). These global reductions in CpG methylation can result in mobilization of transposable elements and their potential disruption of coding or regulatory sequences, and have also been observed in various cancer tissues. Changes in DNA modification patterns have also been shown to be driven by genetic variants such as single nucleotide polymorphisms (SNPs). This phenomenon may partially account for the variation in inherited disorders and, is useful for better understanding complex illnesses (Feinberg & Irizarry, 2010).

The DNA modification state of numerous loci follows a directional pattern over time. In addition to global CpG hypomethylation, several regions, particularly genes involved in early cell development, accumulate DNA modification; this is often accompanied by promoter region enrichment of repressive histone marks (H3K4me3, H3K27me3). The DNA modification state of 353 CpG sites has been found to be associated with biological age. Individual CpG modification was combined into a weighted score and observed to change measurably as a function of age. When this "DNA methylation age" score regressed is with chronological age, deviations are proposed to be a reflection of accelerated or decelerated aging (Horvath, 2013). Epigenetic age was found to be conserved in numerous tissues corresponding to their biological age, whereas pluripotent stem cells showed a lower epigenetic age. Moreover, this metric has been used to detect age acceleration in cancer tissue, post-traumatic stress disorder (Wolf et al., 2017), human immunodeficiency virus-related neurocognitive impairment (Levine et al., 2016), Alzheimer's Disease ( Levine, Lu, Bennett, & Horvath, 2015), and Parkinson's disease (Horvath & Ritz, 2015). Given that many of these CpG sites are located in or near glucocorticoid response elements and showed changes in DNA modification after dexamethasone challenge, suggests that epigenetic aging may be related, in part, to activation of the stress response (Zannas et al., 2015). This supports the notion that the epigenetic basis of many diseases may be a result of cumulative stress and gene misregulation.

A recent study found that DNA modification oscillates in a circadian fashion in liver and lung tissues, suggesting coordinated modification and de-modification as a product of time (Oh et al., 2018). Such oscillation of several CpG sites was also associated with long-term changes in DNA modification state and observed to dampen over time; this finding suggests that epigenetic oscillations may precede and serve to establish age-related accumulation (or loss) of

modification. The findings support the previously-established link between aging and disruption of circadian processes and implicate the role of DNA modification in this process (Froy, 2011; Kondratova & Kondratov, 2012).

#### 1.2.2.3 Heterochromatin loss

Aging has been found to be associated with decreased amounts of heterochromatin, that results in alterations to gene expression and nuclear architecture. This is thought to occur through the impairment of histone deacetylation in regions of heterochromatin (Tsurumi & Li, 2012). In yeast, HDAC inhibitor treatment and deletion of the NAD-dependent histone deacetylase SIR2 gene (*SIR2*) was found to decrease lifespan, whereas its overexpression had the opposite effect. Moreover, the sirtuin family of proteins has been found to regulate genomic stability, and was observed to redistribute from heterochromatin to unstable sites during the aging process (Guarente, 2011; Haigis & Sinclair, 2010). Sirtuin 1 (*Sirt1*), the mammalian homolog of Sir2, also functions to silence repetitive elements in mice, and is redistributed to sites of DNA damage, that results in changes to gene expression that are similar to those in the aging mouse brain (Oberdoerffer et al., 2008; Oberdoerffer & Sinclair, 2007). The heterochromatin loss model of aging has been supported through studies of two rare human genetic disorders, Hutchinson-Gilford progeria syndrome and Werner syndrome, that result in premature aging and shorter longevity. In these disorders, overall loss of heterochromatin is thought to result in genomic deregulation and accelerate the aging process (Goldman et al., 2004; W. Zhang et al., 2015).

#### 1.2.2.4 Histone protein degradation

In addition to loss of heterochromatin, degradation of histone proteins has also been found to influence the aging process. Yeast models have shown that aging is accompanied with a loss of nearly half of all core histone proteins, that results in de-repression of transcriptionally-silent regions and increased global expression of genes (Hu et al., 2014). The decrease of core histone proteins has been linked to their reduced synthesis; this was confirmed by studies in yeast, whereby increased supply of H3 and H4 histone proteins increased lifespan (Dang et al., 2009; Feser et al., 2010). Although it has not been shown *in vivo*, the effect of histone protein loss on

aging was also observed in worms, human diploid primary fibroblasts, and senescent human cells (Ivanov et al., 2013; Ni, Ebata, Alipanahiramandi, & Lee, 2012; O'Sullivan, Kubicek, Schreiber, & Karlseder, 2010). It is thought that decreased amounts of core histone proteins result in changes to chromatin landscape by increasing transcriptional access to DNA. This was demonstrated in budding yeast, whereby nucleosome occupancy decreased with age, and was reversible through overexpression of proteins H3 and H4 (Feser et al., 2010).

#### 1.2.2.5 Aging by retrotransposition

Reduced chromosomal integrity and regulation has also been found to underlie instability of the genome, and the derepression of transposable elements, which have frequently been associated with aging in several species. Although subject to substantial regulation and silencing, increased expression and activity of retrotransposable elements follows the loss of heterochromatin as cells age. This results in otherwise silent elements to become active and insert themselves into various parts of the genome, causing disruptions to cellular homeostasis (De Cecco, Criscione, Peckham, et al., 2013; Garcia et al., 2010; Hu et al., 2014; Wang et al., 2011). This has been demonstrated through studies of SIRT6, a deacetylase of H3K9Ac and H3K56Ac with adenosine 5'-diphosphate-ribosylase activity. SIRT6's ribosylation of KRAB-associated protein 1 (KAP1) increased its association with Heterochromatin protein 1 (HP1) and the packaging of transposable elements into heterochromatin. Throughout the aging process, SIRT6 is found to move away from these elements to tend to DNA damage in other genomic locations, an effect that inhibits the repression of silenced sequences (Haigis & Sinclair, 2010; Van Meter et al., 2014; Wood & Helfand, 2013). This is supported by findings that overexpression of histone proteins served to repress retrotransposon activity (Hu et al., 2014). The transposition-based aging model is upheld in several other studies: in adult stem cells, the onset of senescence was associated with increased transcription of retrotransposable elements and sustained DNA damage. Suppression of transcripts from retrotransposable genes counteracted this phenotype and ameliorated DNA damage foci. Increased transposition has also been noted in studies of the development of cancer and neurodegeneration, adding support to its role as an age-related process (Criscione, Zhang, Thompson, Sedivy, & Neretti, 2014; Lee et al., 2012; Reilly,

Faulkner, Dubnau, Ponomarev, & Gage, 2013; Sedivy, Banumathy, & Adams, 2008; Wang et al., 2011).

#### 1.2.2.6 Changes in histone variants

The aging process has also been associated with a change in the proportion of histone variants. Of particular importance are non-core proteins, that differ in their primary sequences and have several specialized roles in regulating the chromatin remodeling process (Kamakaka & Biggins, 2005). Notably, histone protein H3.3, a variant of H3 that is the most commonly expressed form during senescence, is important for maintenance of chromatin during this period (Filipescu, Muller, & Almouzni, 2014). This was confirmed by findings of excess incorporation of H3.3 being able to induce senescence (Duarte et al., 2014). Deletion of the yeast homolog of the histone cell cycle regulator protein (HIRA), the histone chaperone that accompanies H3.3 incorporation, extended lifespan (Feser et al., 2010). The positive correlation between H3.3 incorporation and aging was observed in mouse brain, where H3.3 levels increased with age, and levels of H3.1, which is mainly expressed during the S phase, decreased (Maze et al., 2015).

#### 1.2.2.7 Histone modification changes

There exist several histone modifications that have been associated with regulation of the aging process, most commonly lysine acetylation and methylation. Acetylation of lysine residue 56 on H3 (H3K56Ac) and of lysine residue 16 on H4 (H4K16Ac) are shown to mediate replicative aging. It is likely that aging occurs through a tight-knit balance of these two modifications that change in proportion over a lifetime. In yeast, H3K56Ac levels are reduced with age, while deletion of genes that mediate this removal has been shown to reduce lifespan and increase genomic instability. Moreover, deletion of a histone acetyltransferase that acetylates H3K56Ac also reduces lifespan. By contrast, H4K16Ac increases during aging, thought to be a result of decreases in a histone deacetylase Sir2. Sir2 overexpression, and deletion of Something about silencing 2 (*SAS2*), that codes for an H4K16 acetylase, increase lifespan. The age-associated increase of H4K16Ac is thought to occur via effects on chromatin structure on telomeres (Dang et al., 2009; Feser et al., 2010).

The interplay between histone acetyltransferases and histone deacetylases has also been important to the aging process. Mutation of NuA4, a histone acetyltransferase of H4 with longevity-implicated substrates, extends yeast lifespan (Lin et al., 2009; Rogina, Helfand, & Frankel, 2002). Further, deletion of genes that express portions of Rdp3, the histone H3 deacetylase complex, has also been shown to extend lifespan in yeast and flies (Frankel, Woods, Ziafazeli, & Rogina, 2015).

Histone acetylation has been shown to be important in the aging brain. Overall histone hypoacetylation has been observed in repetitive DNA elements in aged mice brain, likely due to erosion of chromatin integrity (Ryu, Kang, Yoo, Joe, & Chung, 2011). Age-related memory impairment in mice has is thought to be due to an inability to increase histone histone 4 lysine 12 acetylation (H4K12Ac), which is important in promotion transcription elongation, and maintenance of gene expression patterns. Restoration of H4K12Ac rescued gene expression and learning in aged mice, supporting this finding (Peleg et al., 2010).

#### 1.2.2.8 Nucleosome remodeling complex changes

The nucleosome remodeling complex is responsible for the dynamic nature of histones, and facilitates addition and removal of modifications that either condense or loosen it. This is mediated by ATP-dependent nucleosome remodeling complexes, such as Mi-2/NURD; deletions of this complex were shown to increase lifespan and resistance to environmental stress in *C. elegans*, fruit flies and plants (De Vaux et al., 2013; Pegoraro et al., 2009). Moreover, portions of chromatin remodeling factors, such as RB binding proteins 4 and 7, were found to be reduced in patients with Hutchinson-Gilford progeria syndrome and aged individuals, and their deletion was associated with increased endogenous DNA damage in cultured cells. Loss of these elements as a consequence of physiological aging was further related to alterations in chromatin structure (Pegoraro et al., 2009). These findings suggest that proteins relevant to nucleosome remodeling may serve a protective role from DNA damage, and their conservation in lower and higher species points to their candidacy as important targets to study human aging.

#### 1.2.2.9 lncRNA elements

As suggested above, long non-coding RNA elements are important regulators of epigenetic modifications through their association with proteins that are key to DNA and histone modification. lncRNAs have also been shown to associate with proteins to help maintain three-dimensional structure of packaged DNA and define borders between heterochromatin and euchromatin (Cohen & Jia, 2014). Alterations to these connections over a lifetime can impact disease processes, and suggest an important link between lncRNAs and aging. For example, *Gas5* is an element that is highly expressed in aged mouse brain, and is associated with impaired learning (Meier, Fellini, Jakovcevski, Schachner, & Morellini, 2010). Moreover, loss of imprinting at *IGF2/H19* locus was demonstrated during aging in mice and in human prostate cancer cells; this misregulation results in expression of previously silenced genes and disruption of CTCF binding. Given that prostate cancer risk typically increases with age, these findings suggest that lncRNAs may, in part, mediate this risk (Fu et al., 2008).

#### 1.2.2.10 Dietary restriction effects on epigenetic regulation of aging

Dietary restriction is one of the most widely-cited and easily-accessible means of lifespan extension, and has been cited in multiple species, from yeast, flies to mice and primates (Lee & Longo, 2016). The precise mechanisms of this effect are not fully understood, but thought to be mediated in part by epigenetic pathways. For example, dietary restriction regulates DNA methylation at several genes, possibly through increasing activity of DNMT enzymes (Li, Daniel, & Tollefsbol, 2011). Histone modifications also play a role, whereby activation of class III HDACs in several model organisms was associated with limited diet and greater longevity (Vaquero & Reinberg, 2009). Dietary restriction was also shown to induce certain gene expression patterns that may be protective, and to prevent differential expression characteristic of aging (Wood et al., 2015). For example, increased expression of mir-98-3p modulates histone remodeling and has a neuroprotective effect. Dietary restriction also impacts autophagy, DNA and cell repair, cell integrity, activation of tumour suppressor genes, likely through epigenetic changes (Pal & Tyler, 2016).

#### 1.2.2.11 Conclusion

These findings are in line with the observed degradation of mechanisms that silence transcription over a lifetime, and evidence that manipulation of these mechanisms has impact on lifespan. The above findings also highlight the importance of genomic maintenance in delaying the aging process, and point to numerous epigenetic mechanisms in regulating these processes. Susceptibility to age-related dysfunction may, in part, be due to loss of repressive marks and increase in permissive marks, resulting in increased activity of normally-silenced genes and dysregulation of genes linked to DNA repair, apoptosis, and normal cell function. As suggested by other findings, epigenetic marks are readily influenced by environmental stimuli and metabolism, and can be passed through cell division (Pal & Tyler, 2016). These findings give promise to the discovery of epigenetic targets for intervention in the aging process.

## 1.3 Epigenetics and complex disorders

### 1.3.1 Overview

Many complex disorders run in families, have inherited predisposition, but do not present themselves until much later in life. Despite having been the topic of large-scale genetic investigations, the etiopathogenesis of these disorders have only been modestly accounted for by genetic factors alone. Epigenetic studies add an additional layer by which to understand how these disorders unravel, and how numerous loci of small individual effect can synergistically precipitate symptoms. Epigenetic factors have also been shown to be consistent with several non-Mendelian features of complex disorders, including parent-of-origin effects, discordance among monozygotic twins, delayed age of onset, fluctuating disease course, major hormonal changes and effects of environment and lifestyle (Feil & Fraga, 2012; Oh & Petronis, 2008; Ollikainen et al., 2010; Petronis, 2001; Relton & Davey Smith, 2010). This notion is supported by studies of monozygotic twins that identified that genetic mechanisms explain about one-quarter of the differences observed in twins, positing that the remainder of difference likely occurred through stochastic epigenetic mechanisms over time. This understanding may also serve to bridge the discrepancy between heritability estimates of disease from family studies and those from genetic analysis, collectively termed the "missing heritability" (Labrie, Pai, & Petronis, 2012). Thus, investigating the role that these factors play may serve to improve understanding of the etiopathogenesis of complex disorders. Below is a brief summary of how epigenetics has been applied to better understand several key complex disorders.

### 1.3.2 Cancer

Cancer is a cluster of disorders characterized by abnormal and sometimes aggressive cell growth that has the ability to resist programmed cell death mechanisms and proliferate (metastasize) into other tissues. Development of cancer is strongly linked to abnormal gene expression, cell identity, response to internal and external cues, genetic mutations, abnormal cell division, loss of cell-specific repression of developmental genes. Given the breadth of alteration underlying many cancers, epigenetics lends itself well to studying this illness (Sharma, Kelly, & Jones, 2010).

Studies have identified epigenetic changes in cells that are indicative of a cancerous state. For instance, malignant cells exhibit methylome wide deviations from normal cells and altered chromatin structure indicative of aberrant epigenetic programming (Comet, Riising, Leblanc, & Helin, 2016; Suva, Riggi, & Bernstein, 2013). One study found that gain of function mutations in the enhancer of zeste homolog 2 enzyme, the catalytic subunit of PRC2, result in hyperactivity, increased histone methylation and the promotion of an early developmental state in B cells (Beguelin et al., 2013; McCabe et al., 2012). Epigenetic plasticity can also result in a more active state that allows cells to be highly proliferative, adapt to their environments, and resist apoptosis and response to treatment. For example, gain of function mutations in the isocitrate dehydrogenase gene (*IDH*) are commonly linked to tumorigenesis (Shlush et al., 2014). Overactivity of IDH protein inhibits function of DNA demethylases like Tet, and disrupts CTCF binding. This results in global hypermethylation patterns that are maintained in subsequent cell divisions, and disruption of normal DNA insulation and looping. Loss of insulator function has been linked to increased enhancer activity, and overexpression of genes that are normally kept silent. Moreover, insulator mutations have been found in colorectal, liver and esophageal cancers (Katainen et al., 2015). For instance, platelet derived growth factor receptor alpha (*PDGFRA*) is a glioma oncogene that is frequently aberrantly active (Hnisz et al., 2016). This effect was further confirmed by findings of overexpression of genes separated by CTCF in IDH-mutant tumours (Flavahan et al., 2016; Nora et al., 2012).

### 1.3.3 Cardiovascular disease

Cardiovascular disease is one of the leading causes of death in all countries, and the focus of many diagnostic and preventative campaigns in the past few decades (Nabel & Braunwald, 2012). It involves both genetic predisposition and environmental/lifestyle factors and has been the topic of several epigenetic studies. For instance, cardiac hypertrophy is associated with histone acetylation, whereby CBP and p300 overexpression was linked to increased risk of this phenotype in cardiac cells. This effect was not observed in mutant cells lacking HAT activity, and HDAC inhibitors were shown to have protective function again cardiac hypertrophy *in vivo* (Khalil, 2014). Studies of end-stage heart failure patients identified several epigenetic modifications involved in heart and vasculature remodeling pathways (Movassagh et al., 2010;

Movassagh et al., 2011). Moreover, atherosclerosis in humans was associated with hypermethylation of protective estrogen receptor alpha genes (*ESR1* and *ESR2*) in vascular smooth muscle cells (Kim et al., 2007; Post et al., 1999). Methylation of the monocarboxylate transporter gene (*MCT3*) was also associated with progression of atherosclerosis (Zhu, Goldschmidt-Clermont, & Dong, 2005). miRNAs have also been implicated in cardiovascular disease: changes in expression of genes involved in vascular integrity and health (inflammation, cholesterol transport, lipid uptake, cell differentiation and apoptosis) has been linked to cardiovascular aberrations (Kim et al., 2012; Leeper et al., 2011; Rayner et al., 2010). Though no significant therapeutic applications have been developed as of yet, investigation of epigenetic changes has improved understanding of cellular changes underlying pathophysiology of heart disease.

### 1.3.4 Type 2 diabetes

Type 2 diabetes is a metabolic disorder that is characterized by hyperglycemia from insulin resistance and insulin deficiency, and affects hundreds of millions of people worldwide. Impaired glucose metabolism leads to oxidative stress and drives damage to several tissues (retinopathy, kidney failure, neuropathy, macrovascular disease) (Robertson, Harmon, Tran, & Poitout, 2004). Type 2 diabetes has a strong heritable component, whereby having an affected parent increases risk by 30-40%. Genome-wide association studies (GWAS) in the past decade have identified dozens of loci associated with modest effect size (Dupuis et al., 2010; Voight et al., 2010). Epigenetic studies of individuals predisposed to famine (described briefly above) have found DNA and histone modifications in pancreatic beta cells responsible for insulin production, associated with sustained hyperglycemia. One of the identified changes was the silencing of Pdx1, a transcription factor necessary for pancreatic development and beta cell maturation, and ultimately necessary for normal insulin production (Park, Stoffers, Nicholls, & Simmons, 2008). An epigenome-wide association study (EWAS) of type 2 diabetes identified differential modification in GWAS loci, including several genes (thioredoxin interacting protein [*TXNIP*], ATP binding cassette transporter G1 [*ABCG1*], phosphoethanolamine/phosphocholine phosphatase [*PHOSPHO1*], suppressor of cytokine signaling 3 [*SOCS3*], sterol regulatory element binding transcription factor 1 [*SREBF1*]) relevant to glucose uptake, beta cell function,

metabolism and cell signalling (Chambers et al., 2015; Chen et al., 2016; Florath et al., 2016; Kulkarni et al., 2015; Toperoff et al., 2012). A study in mice investigated the impact of obesity on development of type 2 diabetes (A. Y. Kim et al., 2015). Obesity was associated with DNMT1 activation and hypermethylation and silencing of the adiponectin gene. This is relevant because adiponectin regulates insulin sensitivity. Moreover, plasma DNA modification at CpG sites in the promoter region of resistin (*RETN*), a gene that codes for a hormone involved in insulin resistance, plasma negatively correlated with its expression; having minor allele SNPs in the promoter negative correlated with DNA modification of this locus, suggesting a connection between SNPs and epigenetic factors (Nakatochi et al., 2015).

## 1.4 Bipolar disorder

### 1.4.1 Overview

The primary focus of our research has been on mood and psychiatric disorders, namely BD, a mood disorder that impacts about 1% of the population and is characterized by disorganized thought and behaviour (Merikangas et al., 2011). Other characteristics of BD include recurrent episodes of mania (in bipolar disorder type 1) or hypomania (bipolar disorder type 2), and depression. These episodes involve abnormally and persistently increased mood, irritability, energy, lack of sleep, racing thoughts and difficulty with impulse control and emotional regulation; hypomania is similar to mania, but has a shorter duration (at least 4 days vs. at least 1 week for mania) and is often not severe enough to cause tremendous reductions to functioning or hospitalization. Major depressive episodes are characterized by persistently reduced mood, decreased pleasure and interest in life, and changes in daily habits such as diet and sleeping patterns (Hilty, Leamon, Lim, Kelly, & Hales, 2006). Cyclothymia involves two or greater years of persistent hypomanic and depressive symptoms that are not severe enough to meet criteria for the above characterizations. Typically, symptoms of depression last longer than episodes of mania and hypomania. Collectively, these symptoms impair day-to-day function and quality of life (through ineligibility for employment, lack of productivity), and are associated with tremendous burden on individuals' health (through comorbidities and harmful behaviour), their families and caregivers (Ferrari et al., 2016).

### 1.4.2 Features of bipolar disorder

BD is characterized by fluctuating disease course, and symptoms recur and remit over time. Severity of illness can be impacted by proximal (socioeconomic environment, environment, use/abuse of psychoactive drugs and alcohol) and historical stressors (early life adversity). Both sexes are equally likely to be affected by BD type I, but women experience more frequent depressive episodes (Schneck et al., 2004). Manic episodes are characterized by increased neurotransmission of dopamine and glutamate in various regions of the brain (Chen, Henter, & Manji, 2010). This is followed by a compensatory reduction in dopaminergic signaling that precedes depressive symptoms (Malhi, Tanious, Das, Coulston, & Berk, 2013). BD is often

treated with mood-stabilizing compounds such as lithium, valproate and carbamazepine. Patients have also reported success with a combination of pharmacologic treatment and various forms of psychotherapy (Connolly & Thase, 2011).

Another feature of BD is delayed age of onset, whereby in the vast majority of cases, symptoms do not manifest themselves until late adolescence or early adulthood; the median age of onset is 25 years (Coryell et al., 2003; Schneck et al., 2004). Earlier age of onset (before 20 years) is associated with greater prevalence of rapid cycling and comorbid substance abuse, as well as overall reduction in function and quality of life (Ernst & Goldberg, 2004; Perlis et al., 2009).

Individuals who suffer from BD are at substantial risk of suicide, with nearly half of all affected persons attempting suicide at least once, and completed suicides being up to fifteen times and four times higher than in individuals suffering from unipolar depression and unaffected individuals, respectively (Brown, Beck, Steer, & Grisham, 2000; Harris & Barraclough, 1997). Suicidal behaviour is a complex phenotype that is hard to characterize, but like BD, has a strong familial component (Hawton, Malmberg, & Simkin, 2004). The greatest predictor of suicide is having a first-degree relative who committed suicide, with previous suicide attempt increasing the risk of a future event (Leverich et al., 2002). Additional risk factors for suicide include alcohol or substance use disorder, male sex, impulsive behaviour and persistent social isolation (Angst & Cassano, 2005; Balázs et al., 2006; Jamison, 2000).

Family and twin studies have demonstrated the highly heritable component of BD. Family studies in the past century have demonstrated that having a first degree relative with BD conferred a 7-fold increase of developing the illness (Gershon et al., 1982; Smoller & Finn, 2003). Collectively, these studies estimate BD heritability at about 70% (Craddock, O'Donovan, & Owen, 2005). Further, monozygotic twin studies have reported substantially higher concordance of BD as compared to that in dizygotic twins and other relatives (60% and 7%, respectively) (Bertelsen, Harvald, & Hauge, 1977; Craddock & Jones, 1999). However, progress in identifying causal molecular factors that predispose individuals to BD have faced limited success.

### 1.4.3 Genetic studies

Early genetic studies have focused on identifying candidate genes of BD risk. For example, BD is associated with a decreased expression of reelin (*RELN*) and glutamate decarboxylase (*GAD67*), genes that code for proteins involved in GABAergic neurotransmission (Guidotti et al., 2000). The brain-derived neurotrophic factor gene (*BDNF*) is another important candidate gene, and codes for a protein that has important functions in neurogenesis, learning, memory and mood. BDNF has been shown to be reduced in individuals with BD, particularly during manic and depressive episodes (de Oliveira et al., 2009; Machado-Vieira et al., 2007).

More recently, GWAS have profiled common genetic markers in thousands of population-, age- and sex-matched cases and controls, and have identified potential risk variants associated with BD (Shinozaki & Potash, 2014). SNPs found within several genes (calcium channel, voltage-dependent, L type, alpha-1C subunit [*CACNA1C*], ankyrin-3 [*ANK3*], spectrin repeat containing nuclear envelope protein 1 [*SYNE1*], tetratricopeptide repeat and ankyrin repeat containing 1 [*TRANK1*] and teneurin 4 [*ODZ4*]). The role of these genes includes maintaining voltage-gated channel function, transcriptional regulation, and cell signaling; however, direct connections to BD pathophysiology have yet to be made.

These findings illustrate that dysfunction in neuronal processes like neurotransmission, neuroplasticity and neuronal integrity are important in regulating mood, and may be dysfunctional in affected individuals. Most of the studies nonetheless suffer from low power owing to small effect sizes and small sample sizes, and thus do not fully account for the genetic contribution of the illness. The combination of affordable whole-genome sequencing and functional studies will likely identify further insights into common and rare variants mediate BD risk, and how these loci work synergistically in the disease process.

### 1.4.4 Epigenetic studies

Epigenetic studies of BD have provided context to understanding how genes are regulated in the disease state, and how their expression is changed as a function of pharmacologic treatment. First, a study of blood DNA from individuals with BD and SCZ found increased DNA modification in the promoter of serotonin 1A receptor gene (*5HTR1A*) in affected individuals,

potentially supporting the idea that this gene is downregulated in disease (Carrard, Salzmann, Malafosse, & Karege, 2011). A second study highlighted lower modification levels in an exon region of the potassium voltage-gated channel subfamily Q member 3 (*KCNQ3*), a gene that was previously associated with BD in several linkage and genome-wide association studies and is known to protect against neuronal excitotoxicity (Fidzinski et al., 2015; Kaminsky, Jones, et al., 2015). Another study used fine mapping techniques to investigate DNA modification, and found lower modification levels at the human leukocyte antigen complex group 9 (*HCG9*) gene in individuals with BD (Z. Kaminsky et al., 2012). Interestingly, this was consistently observed in blood, brain and germline samples. Moreover, DNA modification of *HCG9* was also dependent on rs1128306 genotype, supporting the idea that genetic variants can dictate epigenetic modification patterns.

Histone modification marks have also been associated with BD. For instance, one study reported increased H3K4me3 in the prefrontal cortex of individuals with BD and major depressive disorder (MDD), as well as increased expression of synapsin genes (involved in neurogenesis, synaptic transmission and plasticity)(Cruceanu et al., 2013).

Studies of monozygotic twins discordant for BD have observed epigenetic differences at several loci. One investigation used blood DNA and found differential modification between twins at two previously-implicated genes (melanin-concentrating hormone 1 [*GPR24*] and catenin alpha-2 [*CTNNA2*]), and hypomethylation of the alpha-N-acetylgalactosaminide alpha-2,6-sialyltransferase 1 gene (*ST6GALNAC1*) (Dempster et al., 2011). These genes mediate cell-cell interactions in the brain, and the latter overlaps with a rare duplication found in patients with SCZ (Xu et al., 2008). These findings may suggest the effect of epigenetically-measurable environmental risk factors in mediating BD phenotype.

Moreover, mood-stabilizing and antipsychotic medications have been demonstrated to exert some epigenetic mechanism of action. For instance, individuals in depressed or euthymic states had significantly higher *BDNF* modification compared to healthy controls; this effect was reversed by both lithium and valproate, although the change did not meet significance (Dell'Osso et al., 2014). Another study noted increased blood DNA modification levels in BD type I and II patients treated with valproic acid and quetiapine in several neurologically-relevant genes (*RELN*, solute carrier family 1 member 2 [*SLC1A2*], *IGF2*, *H19*, solute carrier family 6 member

4 [*SLC6A4*], glutamate decarboxylase 1 [*GAD1*]) (Houtepen, van Bergen, Vinkers, & Boks, 2016).

The findings above are varied, but support genetic studies that point to epigenetically-mediated changes in gene expression implicated in the disease process. Although these findings do not identify any distinct trend, the changes identified may be etiologically relevant and highlight several avenues by which epigenetic gene misregulation may play a role in BD. Moreover, although limited by sample size and direct clinical applicability, concordant epigenetic modification between various tissues may potentially be valuable diagnostic markers, and may represent specific avenues for intervention.

### 1.4.5 Accelerated aging in bipolar disorder

BD has typically been characterized as a cyclic disorder, with symptoms fluctuating over the course of the disease. However, several studies have identified progressive changes in phenotype of affected individuals, as well as many features that have also been observed in many late-onset disorders and in older individuals (Rizzo et al., 2014). For instance, individuals with BD have a higher prevalence and reduced age of onset of several age-related conditions, such as cancer, cardiovascular disease and metabolic syndrome (Crump, Sundquist, Winkleby, & Sundquist, 2013; Czepielewski, Daruy Filho, Brietzke, & Grassi-Oliveira, 2013; Soreca et al., 2008). Further, there is a demonstrated link between BD and dementia, underlied by shared alterations of inflammation networks, neurotrophic factors, and amyloid pathway (Aprahamian, Nunes, & Forlenza, 2013; Modabbernia, Taslimi, Brietzke, & Ashrafi, 2013). Collectively, these studies implicate cellular and molecular overlap between aging and the pathophysiology of BD.

First, some structural brain changes in BD resemble those found in the aging brain. This includes gray matter volume reduction in the orbital and medial prefrontal cortex, ventral striatum and mesotemporal cortex; increased amygdala size, and enlargement of third and lateral ventricles (Hallahan et al., 2011; Konarski et al., 2008; Lisy et al., 2011; Martinez-Aran et al., 2007). These changes are most pronounced in individuals who have experienced multiple episodes (Strakowski et al., 2002). Second, BD is associated with reductions of cognitive function that are proportional to the frequency of manic episodes, and commonly present during

euthymia (Martinez-Aran et al., 2005; Robinson et al., 2006). Impairments affect verbal memory, executive function and psychosocial function, and overlap with those found in early dementia and mild cognitive impairment (Albert et al., 1999; Martinez-Aran et al., 2007). Third, changes related to age-associated oxidative stress, which are common in numerous metabolic and neurodegenerative disorder, have also been detected in BD. Human and animal studies have observed mania-associated increases in mediators of oxidative stress and decreases in antioxidant proteins, effects that are reversible with lithium treatment (Gergerlioglu et al., 2007; Ghedim et al., 2012; Machado-Vieira et al., 2007). Fourth, deposition of amyloid beta plaques commonly associated with Alzheimer's Disease have been observed in patients suffering from both major depressive disorder and BD (Kessing & Andersen, 2004; Kita et al., 2009). In BD patients, levels of amyloid beta peptides are correlated with the number of lifetime affective episodes (Piccinni et al., 2012). Fifth, BD has been strongly linked to persistent inflammatory states, particularly during manic episodes (Brietzke et al., 2009); many of the observed effects, such as inflammation-related catabolism of neurotransmitter precursors, circulation of senescent CD8+CD28- cells, and reduction of regulatory T cells, are also common in older individuals (Capuron et al., 2011; do Prado et al., 2013). Sixth, a recent study noted that older individuals with BD exhibited epigenetic age acceleration compared to younger individuals and healthy controls; this effect was correlated with increased mitochondrial DNA copy number, a marker of aging and cumulative life stress (Fries et al., 2017). Collectively, these findings suggest that BD pathology may be characterized in the context of differential aging trajectory that may be due to a lifetime of altered neurobiology and behaviour, but may also have a genetic basis, be quantifiable epigenetically, and could reflect common aging pathways between BD and other disorders.

### 1.4.6 Conclusion

BD is a complex and highly heterogeneous disorder, with many factors contributing to psychopathology. The experience of suffering from BD can vary tremendously between individuals (some experience more manic episodes, others more depressive episodes), and is characterized by complexities such as fluctuating disease course, resistance to treatment, sexual dimorphism and peaks of susceptibility that overlap with hormonal changes. It is difficult, if not

impossible, to link a single factor or gene to the disorder. Nevertheless, the high heritability of BD and the wide applicability of epigenetics lends itself well to improving the characterization of the disorder.

## 1.5 Lactase persistence and non-persistence

### 1.5.1 Overview

Lactase phlorizin hydrolase (or simply, lactase) is a member of the beta-galactosidase enzymes that is required to hydrolyze the milk sugar lactose into its monosaccharide constituents, glucose and galactose. The digestion of lactose is crucial to survival in early life given mammals' primary subsistence on milk prior to weaning. Lactase is produced exclusively by brush border epithelial cells that line the villi in the small intestine, and is under the control of a gene of the same name (*LCT*) found on chromosome 2. Lactase is produced in abundance from birth into early childhood in humans and most other mammals, but decreases to near-negligible levels shortly thereafter in most but not all individuals. This later-onset reduction in lactase is responsible for the inability to break down lactose; undigested lactose is subsequently passaged into the large intestine, where it is processed by bacteria and is associated with intestinal discomfort such as bloating, gas, diarrhea and abdominal pain. This phenomenon is termed lactase non-persistence, but is more commonly known as lactose intolerance. Approximately 30% of humans retain varying degrees of the ability to produce lactase throughout adolescence and into adulthood, and are referred to as lactase persistent (Rasinpera, Forsblom, et al., 2005; Sahi, 1994). Lactase non-persistence can be measured clinically using the lactose hydrogen breath test and the lactose tolerance blood test following consumption of lactose (Ghoshal, Kumar, Chourasia, & Misra, 2009). Moreover, substantial reduction of *LCT* mRNA levels has been strongly associated with this phenotype (Mattar, de Campos Mazo, & Carrilho, 2012; Rasinpera, Kuokkanen, et al., 2005).

### 1.5.2 Population and genetic studies

Lactase persistence has been reported in numerous populations, and is thought to have undergone extensive directional selection in several pastoral groups that have historically relied upon dairy as a source of nutrition and water; evolutionary studies have traced this phenomenon back to between three and ten thousand years ago. Lactase persistence is most commonly found among Western and Northern European populations (over 90% in Scandinavian populations), but has also been observed in some Middle-Eastern, East African and South Asian populations (~30-

50%). By contrast, East Asian, Native American and Pacific Islander populations have a virtually non-existent prevalence of lactase persistence (less than 2%) (Mattar et al., 2012; Tishkoff et al., 2009).

Lactase persistence is associated with numerous population-specific genetic markers (namely, SNPs) that arose independently of each other through the process of convergent evolution. Many of these are found upstream of the *LCT* gene in various portions of the nearby housekeeping gene called minichromosome maintenance 6 (*MCM6*). The most widely-studied persistence-associated SNP, C/T-13910 (rs4988235), is located approximately 15 kilobases upstream of *LCT* gene within intron 13 of MCM6 (Poulter et al., 2003). In a Finnish study, possessing at least one copy of the T variant was nearly perfectly associated with persistence. This SNP is in linkage disequilibrium with another persistence-associated marker, G/A-22018 (rs182549), whereby the A allele confers persistence (Enattah et al., 2002). Despite these associations, the above two markers do not account for lactase persistence in all populations that readily consume dairy, particularly those outside Europe. More recently, studies in non-European populations have identified other variants associated with the trait. For instance, minor variants of several SNPs (G/C-14010 [rs145946881], T/G-13915 [rs41380347], and C/G-13907 [rs41525747]) are part of a conserved haplotype that confers persistence in individuals from the Arabian Peninsula and portions of East Africa that include present-day Ethiopia, Somalia, Sudan and Kenya (Liebert et al., 2017; Ranciaro et al., 2014).

Dietary tolerance of lactose is observed in all populations in early life, regardless of genotype and ancestral background. Moreover, studies in populations that commonly possess persistence-associated markers have shown that some individuals without these markers display a later-than-expected onset of lactose intolerance. These observations suggest that the regulatory effect conferred by persistence-associated SNPs does not play a role until later development (Rasinpera, Forsblom, et al., 2005; Seppo et al., 2008). Functional studies have shown that persistence-associated SNPs promote increased activity of the lactase promoter when cloned and transfected in rats (Olds & Sibley, 2003). This was confirmed through *in vitro* investigation of Caco-2, a human epithelial colorectal adenocarcinoma cell line derived from a Caucasian donor that carries the T-13910 genotype, expresses *LCT* and bears morphological similarity to intestinal epithelial cells (Fang, Ahn, Wodziak, & Sibley, 2012). A similar effect was noted for non-European variants G-14010, T-13915 and C-13907 (Tishkoff et al., 2007). The mechanisms

underlying this effect can be explained, in part, by the overlap between the loci and binding sites for transcriptional factors that are instrumental to intestinal cell development and function. Both T-13910 and C-14010 are associated with greater binding affinity of LCT-activating proteins, POU Class 2 Homeobox 1 (POU2F1) and HNF1 homeobox A (HNF1a) (Jensen et al., 2011; Mattar et al., 2012). The temporal interplay between several other factors with LCT-activating (GATA binding proteins, GATA-4, GATA-5, GATA-6; and homeobox protein Cdx2) and repressing function (pancreatic and duodenal homeobox 1, PDX-1) is known to be altered as a function of genotype, but the exact mechanisms behind this have yet to be elucidated (Mattar et al., 2012). Collectively, these findings point to the role of enhancers that remain active in persistence-associated haplotypes and consequently sustain *LCT* in an active state beyond early childhood.

### 1.5.3 Conclusion

The above studies have helped to delineate genetic differences between lactase persistent and non-persistent individuals, and have provided clues to understand processes underlying the differential regulation of the lactase gene. Despite the strong connection between evolutionarily-acquired genetic markers and escape from age-dependent silencing of *LCT*, the mechanisms driving this process remain poorly understood. Despite the limited clinical significance and consequence of lactase non-persistence on quality of life, this phenotype presents yet another example of delayed age of onset. Moreover, the simplicity of this phenotype (regulation of a single gene, expression from a single cell type) makes it a strong candidate for understanding how inherited factors may influence age-related changes to gene function and phenotype, potentially through epigenetic mechanisms.

## 1.6 Challenges in epigenetic studies

### 1.6.1 Overview

Aging is a multi-faceted biological phenomenon that is intimately linked with, and integral to understanding how complex human diseases manifest over the lifespan. Many age-related features (delayed age of onset, changes in symptom severity with time) are common across several disorders, and point to a potentially shared mechanism of pathogenesis.

Many complex disorders are highly heritable, suggesting the involvement of DNA risk factors. Genetic studies of such disorders have assumed a combination of inherited risk variants that interact with environment. However, genetic association studies have only uncovered a handful of markers with small effect sizes, and are limited in their applicability to function. Moreover, GWAS require hundreds of thousands of samples to investigate small effects of numerous markers. Further, environmental risk factors of complex disorders, particularly in major psychosis, have not led to the identification of specific causal risk factors. Collectively, these findings point to a sizable portion of "missing heritability" (the discrepancy between disease heritability measures derived from family studies and genetic markers), and point toward the usefulness of other models to understand complex disorders.

The studies outlined in this review demonstrate the broad applicability of epigenetics to studying nearly every facet of complex disease, including its temporal dimension. Epigenetic studies point to misregulation of genes that may accumulate over time; disease-related epigenetic modification may be inherited as well as acquired. The inclusion of epigenetic factors in the study of complex disorders has added an additional layer to understanding the complex interplay of factors in gene regulation. However, these studies have not always been successfully replicated, and are limited in their usefulness for a number of reasons. Thus, in order to effectively harness the usefulness of epigenetic tools, several important questions must be addressed in forthcoming studies.

### 1.6.2 Tissue and cell-type specificity

A major limitation of many human studies is lack of availability and use of clinically relevant tissues to investigate a disease of interest. Epigenetic modifications are highly-specific to cells and tissues, and even differ in the same cell type in a different tissue region. These differences have important functional consequences that may have relevance to the phenotype of interest. Thus, DNA modification of cells from more accessible tissues (such as peripheral blood, saliva and sperm) may not be concordant with those of the tissue of interest. Moreover, using bulk tissue that may consist of several different cell types corresponding to distinct functional regions may yield an averaged value of DNA modification, and not be sufficient to capture the molecular diversity of cell in a given tissue.

For instance, many studies of major psychosis used peripheral blood cell DNA for epigenetic analysis. Despite being readily available, this DNA traces its origins to dozens of distinct types of white blood cells. Indeed, many such studies have found modest (albeit significant) differences. However, at most loci, DNA modifications in blood DNA do not correlate well with that of the brain (Edgar, Jones, Meaney, Turecki, & Kobor, 2017; Hannon, Lunnon, Schalkwyk, & Mill, 2015; Jiang et al., 2015; Walton et al., 2016). For these reasons, blood-based analysis provides limited biological interpretation.

### 1.6.3 Phenotypic complexity

Many complex disorders are characterized by a spectrum of phenotypes and symptomatology, that makes close and accurate associations between individuals difficult, if not impossible, to make. For example, BD is a disorder that is characterized by clusters of symptoms that vary in their severity and do not impact all affected individuals equally. Moreover, clinical presentation can vary between patients on the basis of numerous factors including, age, sex, environment, medication status, as well as comorbid drug and alcohol abuse. Such complexity points to the synergistic effect of several interacting factors (genetics and epigenetics) that adds tremendous barriers to making simple comparisons between symptoms and genetic or epigenetic markers at specific regions.

### 1.6.4 Temporal dimension hard to capture

The majority of gene expression and epigenetic studies are cross-sectional, and thus capture epigenetic marks at a single time point, thereby neglecting the temporal dimension of the phenotype in question. Thus, using such studies to make inferences to the effects of aging is limited to studying individuals of a variety of ages to test for the heterogeneity of effects as a function of age. Epigenetic marks are only partially stable, and differ markedly between individuals, but also within individuals across the lifespan, and also at various points of a chronic illness. Thus, association of a phenotype with a molecular marker on the basis of a single measurement does not capture the progressive nature of age-related phenotypes, and may not be predictive of a true disease state. Future studies should incorporate a longitudinal component to better understand how epigenetic factors can change over time, and how these changes may be associated with phenotype.

### 1.6.5 Experimental depth versus breadth

The majority of conventional epigenetic profiling methods are limited to specific regions and provide low-resolution data, that may not be sufficient to capture phenotype-specific regions of interest. For example, many locus-specific studies of candidate genes have interrogated a small fraction of a gene, and based their conclusions on the epigenetic state of only a handful of sites. Similarly, microarray-based "epigenome-wide" data has breadth, but only encompasses specific sites (such as promoters, exons, or gene-body), often neglecting to capture introns or sites adjacent to genes. The nature of the data makes it difficult to detect "point epimutations" without *a priori* knowledge or a large sample size. This limitation can be improved upon by using high-resolution epigenomic profiling, or by targeting a larger proportion of CpG sites in a gene or region of interest.

### 1.6.6 Conclusion

These factors underscore the need for improved methodology in order to study the role of aging in complex disorders. In addition to more comprehensive experimental design, understanding the

impact of epigenetic factors in age-related diseases would benefit from a simple phenotype as a proof-of-concept (ideally limited to a single cell-type and genetic locus), the findings of which can be extrapolated. Addressing these limitations would make better use of clinically-relevant epigenetic markers, and allow for a much more nuanced understanding of health and disease.

# 2 Thesis Organization and Aims

## 2.1 Thesis Organization

This thesis is structured in a "paper format" to best reflect the sequential nature of the projects conducted during my doctoral studies. In Chapter 1, I provide an overview of the field of epigenetics and its usefulness in studying non-Mendelian traits and disorders. The primary aim of my work is to investigate the misregulation of genes that occurs as a product of aging, a process that is a strong candidate for understanding the basis of most complex disorders with a delayed age-of-onset; in this chapter, I also outline the aims of my work and the specific research questions that were tested. Chapters 3-5 represent original work conducted during this time with all supplemental material enclosed. Chapter 3 is a modified version of the paper published in Bipolar Disorders (Jeremian et al., 2017). Chapter 4 is a modified version of the paper published in Nature Structural & Molecular Biology (Labrie et al., 2016). Chapter 5 is a modified version of the paper published in Scientific Reports (Oh et al., 2017). I conclude my dissertation with Chapter 6, highlighting the key findings from each of these studies, provide a unifying discussion to merge the overlapping themes, and conclude with future directions.

## 2.2 Research Aims

The primary aim of this thesis is to investigate the contribution of genetic and epigenetic factors, namely DNA modification, to temporal phenotypes that arise both as a product of complex disorders, but also as part of the "normal" aging process. This was initially investigated in an exploratory study of BD and suicidal behaviour, and later expanded to a larger proof-of-concept study using lactase non-persistence as a simple model of DNA variation-dependent epigenetic aging.

The specific hypotheses tested are the following:

1) Individuals suffering from bipolar disorder who have a history of extreme suicidal behavior have a different "epigenetic age" profile compared to matched subjects with BD, without this history (Chapter 3).

2) DNA modification accounts for the haplotype-dependent regulation and aging of the lactase gene, which ultimately results in lactase non-persistence or lactase persistence (Chapter 4).

3) DNA modification of key regions in and near the lactase gene accounts for downregulation of the lactase gene in a small intestinal segment-specific manner (Chapter 5).

# 3 Investigation of correlations between DNA modification, suicidal behavior and aging

This chapter is modified from the following:

Jeremian, R., Chen, Y. A., De Luca, V., Vincent, J. B., Kennedy, J. L., Zai, C. C., & Strauss, J. (2017). Investigation of correlations between DNA methylation, suicidal behavior and aging. *Bipolar Disord*, 19(1), 32-40. doi:10.1111/bdi.12466

## 3.1 Abstract

Suicidal behavior (SB) is a major cause of mortality for patients diagnosed with bipolar disorder (BD). In this study, we investigated epigenetic differences in BD participants with and without a history of SB. We used suicidality scores constructed from Schedule for Clinical Assessments in Neuropsychiatry (SCAN) interview questions about suicidal thought and behavior to identify individuals from a BD cohort of n = 452; participants with the most extreme high (H-SB, n = 18) and most extreme low (L-SB, n = 22) scores were used as cases and controls, respectively. Epigenome-wide DNA modification patterns were compared between the two groups using the Illumina Infinium Human Methylation 450 BeadChip microarray and significant differentially-methylated regions were annotated for potential functional significance in brain and blood tissue. Further, DNA modification age was compared to chronological tissue age. We observed a greater correlation between DNA modification age and tissue age in controls ($R = 0.91$, $p < 0.0001$) than in the H-SB group ($R = 0.83$, $p < 0.0001$). We report significant findings at three loci based on a methylome scan of participants with BD for an SB phenotype, and potentially altered molecular aging in suicide attempters. Despite the small sample size, our proof-of-concept study highlights the potential for epigenetic factors to be useful in understanding the molecular underpinnings of suicide with the ultimate aim of its prevention.

## 3.2 Introduction

Individuals with bipolar disorder (BD) are at significant risk of suicidal behavior (SB), a spectrum which includes ideation, attempt and completion (Turecki, Ernst, Jollant, Labonté, & Mechawar, 2012). More than half of all individuals with BD have at least one lifetime suicide-related disturbance, and 8−20% are at risk of completed suicide (Hawton, Sutton, Haw, Sinclair, & Harriss, 2005; Jamison, 1986). A history of suicide attempt is associated with long-term disability and negative health outcomes (Balázs et al., 2006; Müller-Oerlinghausen, Berghöfer, & Bauer, 2002). Clinical and molecular findings support the idea that transmission of SB in mood and affective disorders has a familial and environmental basis whereby predictive factors are described as proximal, such as substance use, mood disorders and adherence to treatment, and distal, such as a family history of SB and adverse childhood experiences (Brent et al., 2004; Egeland & Sussex, 1985). Candidate genes that mediate stress response and neurotransmission have been associated with suicidal ideation and completion, but the small sample sizes of many genetic studies have failed to demonstrate genome-wide significant effects. Further, low-to-moderate concordance of SB in identical twins with BD suggests that SB follows a distinct pattern of heritability from that of the co-existing affective disorder (Bertelsen et al., 1977; Mathews et al., 2013). Thus, characterizing the clinically heterogeneous phenotype of suicide requires looking beyond DNA and investigating factors that are dynamic throughout the course of the disease state, can be monitored over time, and are mutable in response to pharmacologic treatment. Epigenetic phenomena, such as DNA modification and histone modification, have been shown to play a significant role in gene regulation and expression, and may serve, in part, to quantify the "missing heritability" of mood disorders and SB. For example, mouse models have demonstrated the effect of maladaptive maternal grooming behavior on hypothalamic-pituitary-adrenal axis dysregulation, which is modulated largely by heritable epigenetic changes in the brain (Turecki, 2014). Moreover, differential modification of genes involved in neurotransmission and mood is strongly associated with SB in peripheral blood and *post-mortem* tissue (Bani-Fatemi, Howe, & De Luca, 2015; Maussion et al., 2014). Notably, genetic and epigenetic effects at a single SNP in the Spindle And Kinetochore Associated Complex Subunit 2 gene (*SKA2*) were observed in *post-mortem* brain and peripheral blood tissue of at-risk participants; DNA modification state and expression of the gene, in conjunction with cortisol level and self-reports of depression, were highly reliable predictors of future suicide risk

(Kaminsky et al., 2015). Collectively, these findings demonstrate that epigenetic investigation of suicide provides opportunities to quantify the effects of dynamic social, biological and environmental stressors on mediating SB, and to identify accessible biomarkers of SB.

We performed an epigenetic analysis on DNA from individuals with BD who were grouped on the basis of a history of suicidal ideation and at least one lifetime suicide attempt. Using the Illumina BeadChip Human Methylation 450K array platform (Illumina Inc., San Diego, CA, USA), we observed highly significant differences in modification between cases and controls in three genomic regions enriched for epigenetic modifications corresponding to gene regulatory regions. BD participants with a history of SB showed less overall modification in the 5′ untranslated region of Membrane palmitoylated protein 4 (*MPP4*) ($p = 7.42\times10^{-7}$) and in intron 3 of TRE2/BUB2/CDC16 domain family member 16 (*TBC1D16*) ($p = 6.47\times10^{-7}$), while exon 1 of Nucleoporin 133 (*NUP133*) was less methylated in controls ($p = 1.17\text{x}10^{-6}$) (Figure 3-1). These regions were annotated with chromatin modification profiles of peripheral blood and *post-mortem* brain to investigate the concordance of epigenetic marks across these tissues. Our sample set was also subjected to DNA modification age analysis to link biological age to epigenetic age (Horvath, 2013). Based on previous epigenetic studies of SB, we expected that individuals with BD and a history of suicidality may be at a greater risk of trauma and heightened sensitivity to stressful events, which may be reflected in distinct DNA modification changes compared to BD patients without a history of suicidality.

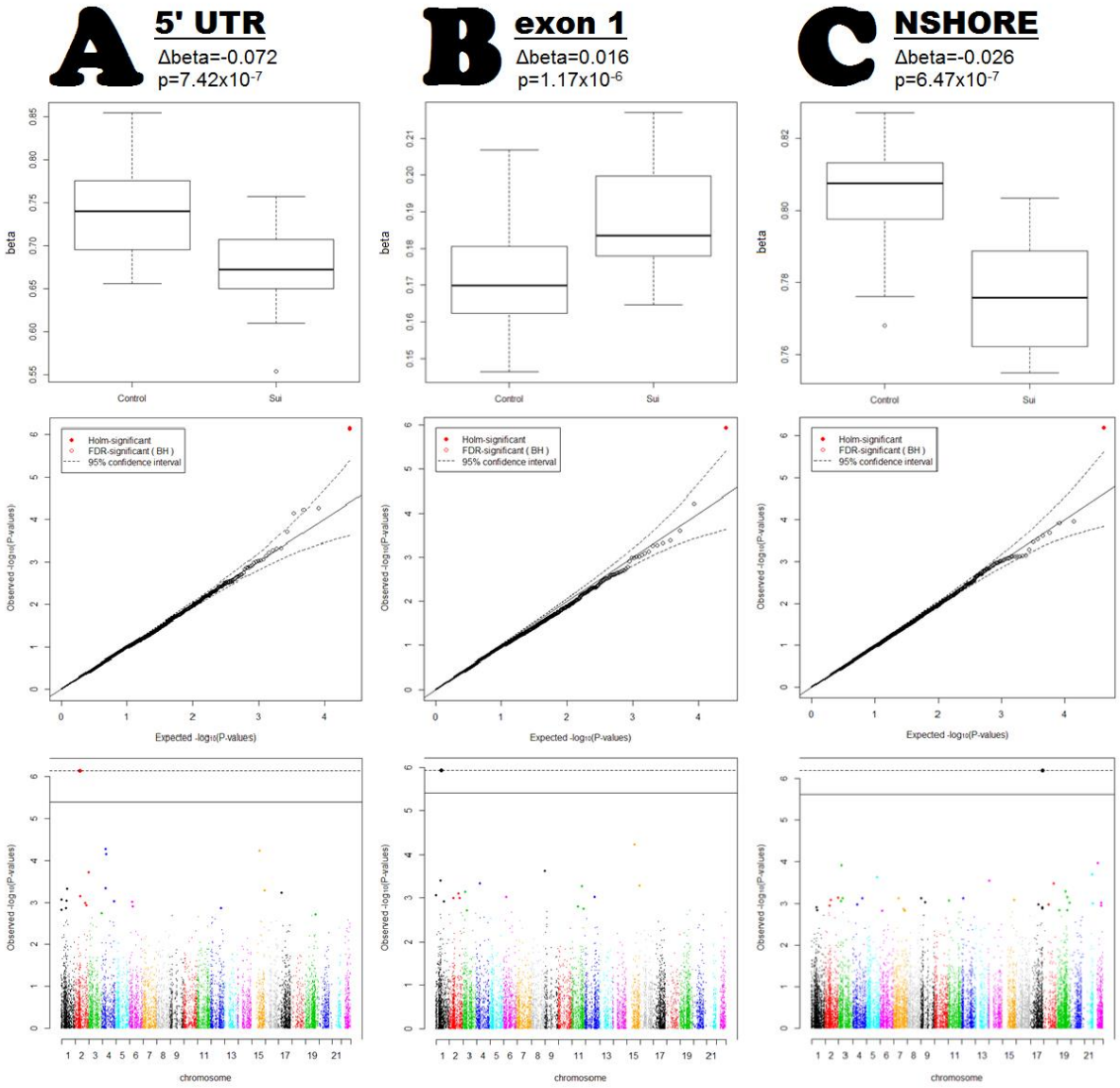



**Figure 3 - 1 . Summary of significantly differentially methylated regions between suicide (Sui) and control groups adjusted for false discovery rate (FDR) using Benjamini-Hochberg (BH) correction**

**[**Top] Age-adjusted box plots of modification difference; [Middle] Quantile- quantile plots of the association between DNA modification and group in the significant gene region; [Bottom] Manhattan plots of the association between DNA modification and group in the significant gene region. (A) The 5' untranslated region of Membrane palmitoylated protein 4 (*MPP4*) was less methylated in suicidal individuals compared to controls ($\Delta\beta = 0.072$, $p = 7.42\times10^{-7}$; FDR adjusted $q = 0.009$). (B) Exon 1 of Nucleoporin 133 (*NUP133*) was more methylated in suicidal individuals compared to controls ($\Delta\beta = 0.016$, $p = 1.17\times10^{-6}$; FDR adjusted $q = 0.015$). (C) Intron 3 of TRE2/BUB2/CDC16 domain family member 16 (*TBC1D16*; located in the -2 kilobase region flanking the CpG island [NSHORE] at chr17:77965805-77966296) was less methylated in suicidal individuals compared to controls ($\Delta\beta = 0.026$, $p = 6.47\times10^{-7}$; FDR adjusted $q = 0.013$). Coordinates are based on the GRCh37/hg19 genome build.

## 3.3 Materials and Methods

### 3.3.1 Patient Recruitment and sample collection

Individuals were chosen from a larger sample of BD patients from Toronto, Canada (n=452) who were previously recruited for a study at the Centre for Addiction and Mental Health (Scott et al., 2009). Inclusion criteria included a DSM-IV or International Classification of Diseases 10th Revision (ICD-10) diagnosis of BD subtype I or II, a minimum age of 18 years, and Northern and Western European ethnicity. We chose not to investigate SB in other psychiatric disorders due to limited sample availability, the small-scale study design and heterogeneity of the SB phenotype. Participants were chosen based on extreme measures, highest and lowest, of a two-question suicidality score, a modified version of the SCAN Suicidality (SSU) score derived from the Schedule for Clinical Assessments in Neuropsychiatry (SCAN) (Wing et al., 1990; Schosser et al., 2011). The modified SSU is the sum of the ratings for SCAN items 6.011 and 6.012. Exclusion criteria were a history of dependence on intravenous drugs, intellectual disability, relatedness to another study participant, manias which occurred only as a result of substance abuse or medication, medical illness, and the presence of mood-incongruent psychotic symptoms. BD diagnoses were made using a computerized algorithm (CATEGO) for the SCAN 2.1 interview, based on DSM-IV or ICD-10 criteria (Wing et al., 1990). Following diagnosis and ratings for suicidality, the extreme 24 most (H-SB) and 24 least (L-SB) suicidal patients were used as case and control groups, respectively. The above procedures were monitored and approved by the Centre for Addiction and Mental Health (Toronto, ON, Canada) Research Ethics Board. Written informed consent was obtained from all recruited participants prior to the study.

### 3.3.2 DNA and genotyping

Genomic DNA was obtained from white blood cells using a high salt extraction method (Lahiri & Nurnberger, 1991). DNA modification status was determined using the Illumina Human Methylation450 BeadChip platform, which can assess modification at over 450,000 CpG sites in 11 gene regions: 1500 nucleotides upstream of the transcription start site (TSS1500), 200 nucleotides upstream of the transcription start site (TSS200), the 5′ untranslated region (UTR5), the first exon (EXON1), the gene body (GENEBODY), the 3′ untranslated region (UTR3), the

CpG islands, -2 kilobase region flanking the CpG island (NSHORE), the -4 to -2 kilobase region flanking the CpG island (NSHELF), +2 kilobase region flanking the CpG island (SSHORE), and the +2 to +4 kilobase region flanking the CpG island (SSHELF) (Illumina). Out of 48 samples, eight were found to be partially degraded, and therefore, for these samples, 2 micrograms (μg) of DNA was used in bisulfite conversion. For the remaining 40 samples, 1 μg of DNA was used. All samples were processed in one batch using four chips. SNP data for 32 of the above individuals (13 cases and 19 controls) had been obtained as part of a previous study (Zai et al., 2015). SNP genotyping on the Illumina Sentrix Human Hap550 BeadChip was performed at Illumina Inc., with subjects being genotyped at the Genome Quebec facility (Montreal, QC, Canada).

### 3.3.3 Data analysis

Preliminary methylome analysis was performed using the Illumina GenomeStudio version 2011.1, which yields a β value, a measure of the density of methylated cytosines for each locus. Background intensities were used to normalize the data set. The β value was then calculated as the intensity of methylated probes divided by the sum of the intensity of methylated probes and the intensity of unmethylated probes; β values range between 0 (unmethylated) and 1 (fully methylated). Individual CpGs were analyzed for differences between case and control groups. Further, modification values were combined for CpG sites in the same genomic region. Subsequent analysis was performed using a modified version of the IMA R package version 3.1.2 (Wang et al., 2012). Peak correction was done to normalize modification levels detected by probes of the array. Eight samples (6 cases, 2 controls), containing more than 5% of probes with a detection P-value > 0.00001, were removed. The detection P-value is defined as the probability that the probe intensity is distinguishable from the background intensity. Forty samples (18 cases and 22 controls) were retained from the original set, and underwent quality control and data cleaning. As a quality control measure, any samples that showed outlier signals for any of the internal array control probes (which were used to assess the quality of bisulfite conversion, hybridization, single base extension, target removal, and non-specific probe extension) were removed from further analysis; no sample was removed from this step. Further data cleaning included the removal of 118,563 sites containing SNPs or targets of non-specific probes; 9,745

sites on sex chromosomes, and 1710 sites with a detection P-value > 0.05 in more than 75% of samples (Dedeurwaerder et al., 2011). Of the original 485,577 sites, 348,091 were retained. CpGassoc version 2.50, an R statistical package, was used to determine the significance of each gene region using a fixed effects model which used the chip ID as a covariate to adjust for batch effects (Barfield, Kilaru, Smith, & Conneely, 2012). A Wilcoxon rank-sum test was then performed for all CpGs in each gene region to compare case and control groups; the resulting P-values were adjusted using the false discovery rate method (Benjamini & Hochberg, 1995). DNA modification age and patient blood cell distribution were determined using the DNA methylation age calculator (Horvath, 2013). Pearson correlations, analysis of variance and figures were generated using Sigmaplot (Systat Software, Chicago, IL, USA). Significant regions were annotated with available Roadmap Epigenomics enrichment data (Release 9) for seven histone modifications (H3K4me1, H3K4me3, H3K9me3, H3K27me3, H3K36me3, H3K9ac and H3K27ac) in six brain regions (the anterior caudate, cyngulate gyrus, middle hippocampus, inferior temporal lobe, mid frontal lobe, and substantia nigra) and two blood cell types (naïve CD4 and CD8 T cells) (Kundaje et al., 2015). Regions of interest were mapped to enrichment data using the Genboree Workbench tool (Coarfa et al., 2014). Values corresponding to histone enrichment in peripheral blood and brain tissue were averaged for all CpGs of a given region, and graphed based on the magnitude of relative enrichment using Excel (Microsoft). Quality control of SNP genotyping data was performed using PLINK and R in a fashion similar to that described in Zai et al (Purcell et al., 2007; Zai et al., 2015). SNPs within and 10 kb upstream and downstream of the three significant regions were profiled. Individuals and SNPs with more than 1% missing genotypes were excluded. SNPs with minor allele frequencies of  < 0.01 were removed. We did not find related individuals (PI_HAT  > 0.05) or participants with extreme heterozygosity (< 4 S.D. from the mean). SNPs for which the genotypes deviate from Hardy−Weinberg equilibrium ($p < 0.000001$) were not included. Fifty two markers remained after additional quality control for missingness and minor allele frequency for the 32 individuals included in the case−control association analysis.

## 3.4 Results

We investigated peripheral blood DNA modification states in individuals diagnosed with BD who were grouped as controls and cases on the basis of extreme scores. Our sample of participants contained 22 controls (mean age 44.3, S.D. 16.3 years; 12 males, 10 females) and 18 cases (mean age 48.1, S.D. 10.9 years; 8 males, 10 females). We also used previously available SNP genotyping data for 32 of our participants in and near the three significant regions presented above and found five nominally significant alleles with a case−control association (Table 3-1). None of the surveyed markers survived multiple testing correction ($q > 0.05$).

Next, we compared tissue age and blood cell type distribution between case and control groups based on DNA modification in biological age-associated regions. Using the DNA methylation age calculator, we observed that the DNA modification age of tissue was more closely correlated with biological patient age in non-suicide control patients based on Pearson correlation ($R = 0.919$, $p < 0.0001$; Figure 3-2A) compared to cases ($R = 0.833$, $p < 0.0001$; Figure 3-2B). We did not observe differences in blood cell type distribution between groups (Figure 3-2C).

Further, we investigated tissue- and brain region-specific function, and the potential for inter-tissue overlap of our most significant differentially-modified regions. This was accomplished by annotating these regions with data obtained corresponding to enrichment profiles of seven histone modifications from the Roadmap Epigenomics Project (Kundaje et al., 2015); data were obtained from analysis of six brain regions and the two most abundant blood cell types in our cohort, naive CD4 and CD8 T lymphocytes (Figure 3-3). Of the three surveyed regions, exon 1 of *NUP133* exhibited the greatest degree of enrichment of H3K4 methylation, as well as H3K9 and H3K27 acetylation, across both blood and brain tissue. By contrast, we observed the lowest degree of chromatin modifications in the 5′ UTR of *MPP4*. A modest degree of chromatin modification enrichment was observed in intron 3 of *TBC1D16*, across four brain tissues and CD8 T cells, showing concordant enrichment in the middle hippocampus and CD8 T cells. No enrichment of H3K9 trimethylation was observed in any tissue.

| Chr | SNP | MinAllele | MinFreqCases | MajAllele | MinFreqControls | $\chi^2$ | p-value | Odds Ratio |
|---|---|---|---|---|---|---|---|---|
| 1 | rs6671385 | T | 0.3462 | C | 0.1316 | 4.159 | 0.04141 | 3.494 |
| 2 | rs2676325 | T | 0.3077 | C | 0.5789 | 4.561 | 0.0327 | 0.3232 |
| 17 | rs3751947 | A | 0.3077 | G | 0.07895 | 5.675 | 0.01721 | 5.185 |
| 17 | rs12450588 | A | 0.1923 | G | 0.05263 | 3.092 | 0.07869 | 4.286 |

**Table 3 - 1. Summary of nominally-significant SNPs with a case-control association**

Minor and major alleles are denoted as MinAllele and MajAllele, respectively. Minor allele frequency is displayed in cases (MinFreqCases) and controls (MinFreqControls) with corresponding chi-squared score and p-value. Odds ratio corresponds to the effect of the minor allele. All individuals in this analysis are of Northern and Western European descent.

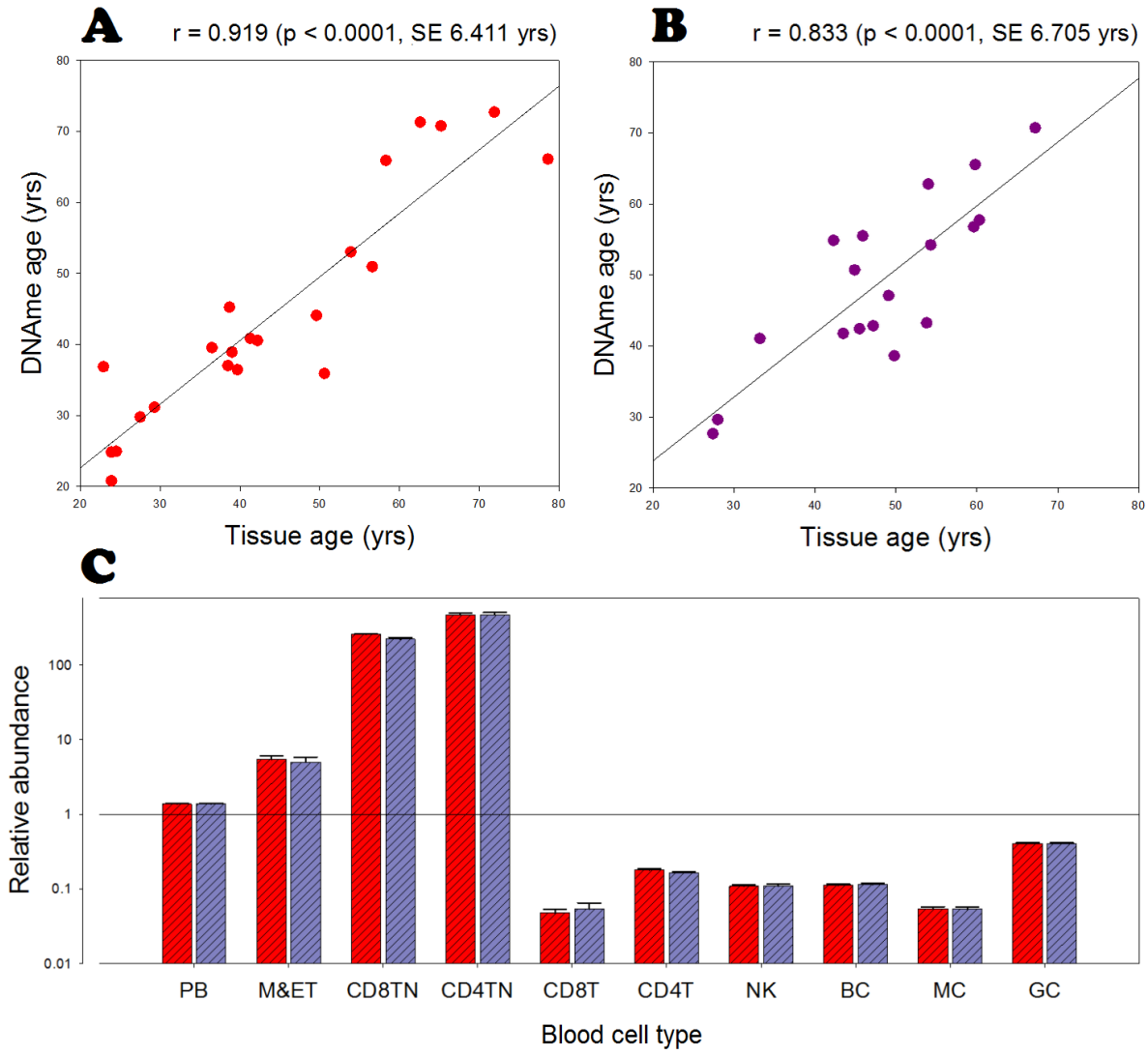



**Figure 3 - 2. Age analysis and white blood cell (WBC) composition by group**

(A) Correlation of tissue age and age- related DNA modification markers in the control group (R = 0.833, S.E.M. = 6.705 years, $p < 0.0001$). (B) Correlation of tissue age and age-related DNA modification markers in the suicidal group (R = 0.919, S.E.M. = 6.411 years, $p < 0.0001$). (C) Comparison of relative white blood cell distributions in patient samples indicated no significant differences between the control (red) and suicide (violet) groups (left to right: PB, plasmablasts; M&ET, memory and effector T cells; CD8TN, naïve CD8 T cells; CD4TN, naïve CD4 T cells; CD8T, CD8 T cells; CD4T, CD4 T cells; NK, natural killer cells; BC, B cells; MC, monocytes; GC, granulocytes). The first four counts represent the Horvath lymphocyte algorithm, while the latter six represent the standard Houseman blood count algorithm. Error bars represent S.E.M.

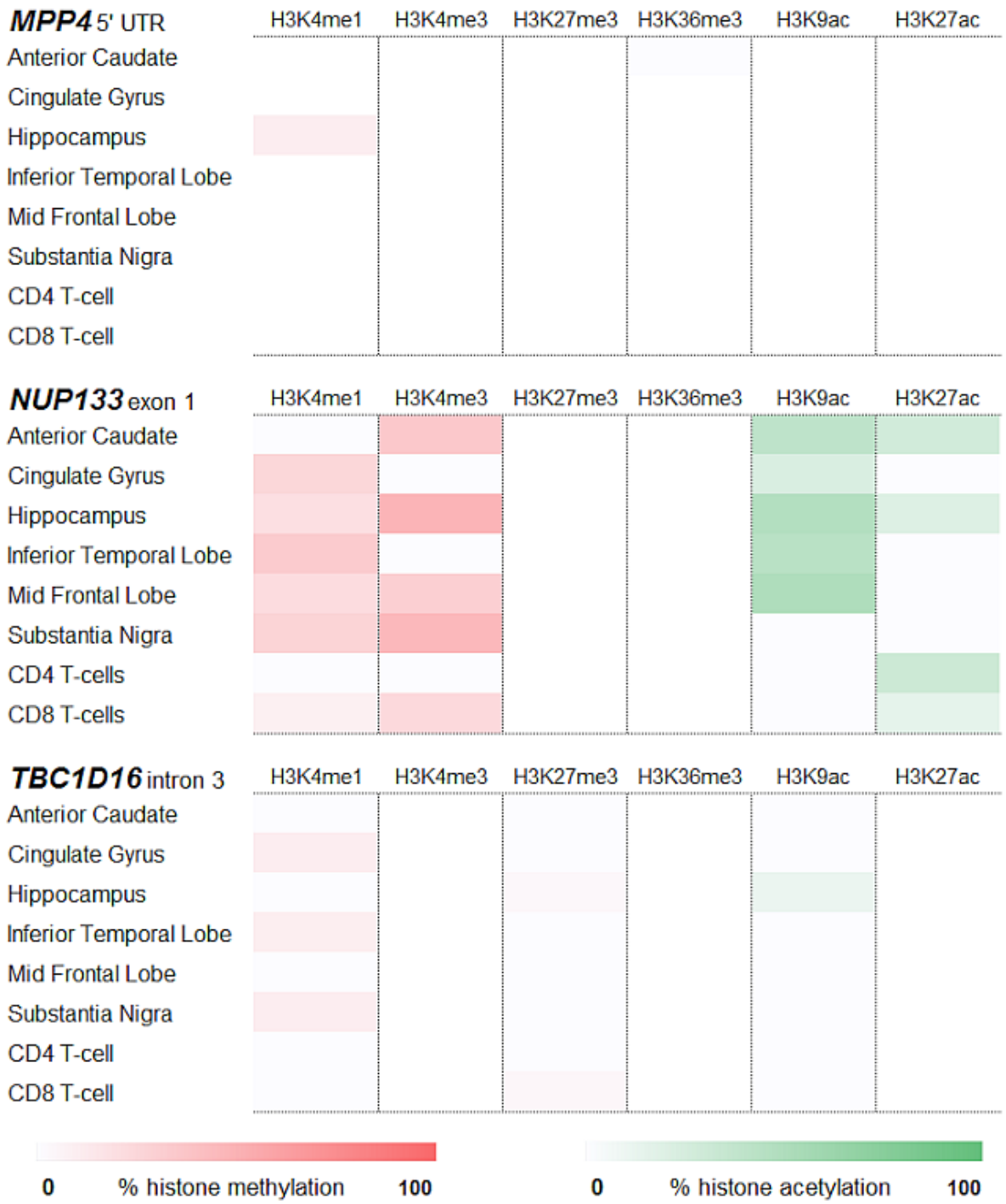



**Figure 3 - 3. Histone modification profiles of significantly differentially methylated regions in brain and peripheral blood tissue**

The mean relative enrichment of histone modifications in three significant regions (5' untranslated region [UTR5] of Membrane palmitoylated protein 4 [*MPP4*], exon 1 of Nucleoporin 133 [*NUP133*], and intron 3 of TRE2/BUB2/CDC16 domain family member 16 [*TBC1D16*]) across six brain regions and two peripheral blood cell types is shown.

## 3.5 Discussion

We interrogated the DNA modification state of nearly 350,000 CpGs, and found significant differential modification in three gene regions between individuals diagnosed with BD who were grouped on the basis of extremes of suicidality scores (Figure 3-1). We annotated the identified markers with histone modification enrichment data from brain and blood tissue to investigate the functional and regulatory potential of these loci (Figure 3-3). First, we observed greater DNA modification in cases at the 5′ UTR of *MPP4* which codes for a member of the family of membrane-associated guanylate kinases that is expressed exclusively in the retina, and regulates the activity of membrane calcium ATPases (Stohr et al., 2005; Stöhr, Stojic, & Weber, 2003; Stöhr & Weber, 2001; Yang et al., 2007). Bovine and porcine models have identified MPP4 protein localization in synaptic terminals of cone and rod photoreceptors, and highlighted its role in transduction of sensory processing (Stöhr et al., 2003). This finding is of interest given the evidence of an association between abnormal retinal response, affective disorders and the seasonality of suicidal behavior (Balogh, Benedek, & Kéri, 2008; Hébert, Beattie, Tam, Yatham, & Lam, 2004; Hébert, Dumont, & Lachapelle, 2002). The disruption of the visual system and other pathways involved in sensory transduction may be indicative of higher order cognitive deficits associated with mood disorders, including disruption of memory, executive function, and attention. In light of these findings, we posit that *MPP4* may be epigenetically mutable in response to circadian and seasonal rhythms, and a better understanding of its function in the context of BD could link the circadian and seasonal patterns of retinally-mediated precipitation of stress response and depressive symptoms to the risk of SB (Bunney & Bunney, 2000; Hasler, Drevets, Gould, Gottesman, & Manji, 2006; Lenox, Gould, & Manji, 2002).

Second, we observed differential case−control modification at exon 1 of *NUP133*. The NUP133 protein is a member of a complex involved in the cell cycle, and plays a role in spindle assembly, kinetochore function, cytokinesis, nuclear pore complex reassembly following mitotic exit and *de novo* nuclear pore complex assembly in interphase, and helps mediate nuclear mRNA export (Berke, Boehmer, Blobel, & Schwartz, 2004; Vasu et al., 2001; Wozniak, Burke, & Doye, 2010). The significant CpGs lie in a region that is enriched for regulatory histone modifications that are characteristic of active promoters (H3K4me3, H3K9ac and H3K27ac) and enhancers (H3K4me1) (Barski et al., 2007; Creyghton et al., 2010; Heintzman et al., 2007). These peaks are found throughout the gene, including introns 12 and 24. Collectively, these modifications span

nearly the entire spectrum of surveyed brain and blood tissue, which suggests an active function of *NUP133* in these tissues. *NUP133* has not previously been associated with mood disorders, but, given the abundance of regulatory activity of this gene in the brain, it is reasonable to suggest that epigenetic modifications may impact *NUP133* function and play a role in modifying BD pathology and, consequently, risk of SB. Moreover, the concordant epigenetic modifications we observed in brain and blood DNA in this gene may enable its use as a clinical marker of suicide in follow-up studies.

Third, we noted decreased modification in cases at a five-CpG shore in intron 3 of *TBC1D16*. Akin to the other significant regions, this locus is located in a DNase I hypersensitivity cluster, and is presumed to play a role in gene regulatory activity. This is supported by the profile of histone modifications in this region, as we observed H3K4me1 and H3K9ac enrichment in four brain regions, as well as a histone mark corresponding to an inactive gene promoter (H3K27me3) in the hippocampus and CD8 T cells (Barski et al., 2007). These findings may be of relevance to our study given the role that the TBC1D16 protein plays as a GTPase activator of the Ras-related protein 4a (Rab4a) signaling pathway (Goueli, Powell, Finger, & Pfeffer, 2012). This pathway modulates the trafficking and activity of the epidermal growth factor receptor (EGFR), a signaling receptor kinase that is involved in neural cell growth and survival, and was shown to exhibit crosstalk with tropomyosin receptor kinase B (TrkB). In the brain, TrkB is involved in neurotransmission, synaptic plasticity and neuronal integrity, and has been widely implicated in modulating the course and symptoms of BD along with its ligand brain-derived neurotrophic factor (Yoshii & Constantine-Paton, 2010). For instance, reduced levels of TrkB protein in frontal cortex astrocytes of suicide completers were attributed to hypermethylation of the 3′ UTR of a truncated TrkB transcript (Maussion et al., 2014). Moreover, a SNP in *EGFR* was found to be significant in a genome-wide association study of BD, but this finding was not supported by a subsequent study (Vizoso et al., 2015; Weber et al., 2011). Based on these data, we posit that DNA modification of *TBC1D16* may influence neuronal and astrocytic cellular stability through protein−protein interactions with TrkB, and could be a molecular marker that is reflective of a history of suicidal behavior. Our findings are corroborated by a recent study of panic-induced epigenetic modifications in mice. Notably, the latter two most significant loci described above are located near human homologs of genes implicated in this study (Mix paired-like homebox [*MIXL1*], Tripartite motif containing 67

[*TRIM67*], Component of oligomeric golgi complex 2 [*COG2*] and piggyBac transposable element derived 5 [*PGBD5*] near *NUP133*; and Congenital disorder of glycosylation type IIg [*COG1*], Cytohesin-1 [*CYTH1*], Pyrroline-5-carboxylate reductase 1 [*PYCR1*], Rhomboid 5 homolog 2 [*RHBDF2*], Aralkylamine N-acetyltransferase [*AANAT*], Solute carrier family 26 member 11 [*SLC26A11*], Rho GDP dissociation inhibitor alpha [*ARHGDIA*] and Myeloid associated differentiation marker like 2 [*MYADML2*] near *TBC1D16*) (Cittaro et al., 2016).

Although the statistically-significant SNPs we identified were not found to be investigated in prior research, our hope is that their presentation here may give these markers attention in future studies. It should also be noted that three nominally significant genes in our study are significantly differentially modified in panic mouse models (*PPTC7*, *PIWI1* and *GNAS*). Given the strong concordance between suicide attempt and panic disorder, these loci may be part of a common molecular pathway involved in both processes (Nam, Kim, & Roh, 2016; Nepon, Belik, Bolton, & Sareen, 2010).

We extended our analysis by investigating age-related signatures of DNA modification as a product of SB in our patient group. Using the freely available DNA methylation age calculator, we compared chronological age of sample tissue at collection date to its corresponding DNA modification age, and observed a weaker correlation between these two variables in H-SB individuals compared to controls. Based on the observed case−control differences, we hypothesize that blood tissue in individuals with a history of suicide attempt may exhibit greater age acceleration, a measure of the discrepancy between epigenetic and chronological age. This marker is posited to be a predictor of regulatory and genomic stability. For instance, aberrant age acceleration was observed in 6000 cancer cell lines, and is thought to impact cell survival (Horvath, 2013). In another study, increased epigenetic aging was associated with somatic mutations of growth and proliferation factors and increased progression of cancer (Lin & Wagner, 2015). Taking these findings into consideration, our observation of differential epigenetic age acceleration particularly in middle-aged patients may suggest that adverse physical and mental health outcomes stemming from a history of suicidal behavior may be reflected in epigenetically mediated changes which may have unexplored physiological effects. Use of the DNA methylation calculator also enabled us to perform epigenetic profiling of cell-specific modifications, and correct for potential differences in white blood cell fraction distribution between samples. The observed lack of cell-type differences, as generated using the

DNA methylation age calculator, validated our microarray analysis, and allowed for accurate case−control comparisons of the relative epigenetic contribution of each cell type to our phenotype of interest. We are the first to investigate altered epigenetic age acceleration as a potential marker of SB in BD patients. It is plausible that altered molecular aging may correlate to risk of suicide, or be a product of previous suicide attempt, or predisposing factors. To validate the inclusion of this marker in current molecular studies of suicide, our findings warrant an investigation of the molecular and cellular changes associated with age-related epigenetic signatures in suicide across multiple tissues.

This study should be viewed in the light of several limitations. Notably, correlations between peripheral blood DNA modification and disease-related molecular and brain region-specific processes have not been broadly demonstrated. Further, the low reproducibility of epigenetic findings across tissues remains a major impediment to the widespread implementation of epigenetic study of suicide in clinical settings. However, recent studies have demonstrated that epigenetic changes relevant to pathophysiology are not necessarily limited to brain tissue and that blood DNA may be a viable source for discovery of pathological epimutations (Cui et al., 2003; Rosa et al., 2008). Moreover, blood DNA has been successfully used for epigenome-wide profiling of mood and psychiatric disorders (Ewald et al., 2014; Walton et al., 2016). In light of positive findings and its relative availability in living individuals, blood DNA may be considered a potentially useful complement to brain tissue in the investigation of dynamic phenotypes, such as SB. Further, the modest statistical significance of SNPs adjacent to our differentially-methylated regions, coupled to the small sample size, adds a barrier to comprehensive functional investigation of the link between these variants and the suicide phenotype; however, our preliminary findings show promise and underscore the need for the inclusion of SNP analysis in subsequent studies to investigate whether suicide-associated genetic variants may also be associated with differential DNA modification. Another limitation of this study is the absence of a comprehensive analysis of patient history; our investigation was modestly powered, and additional analyses of potential interest were not feasible to pursue, including stratification by BD subtype, adverse life events, or comorbidity (such as anxiety disorder, non-intravenous substance abuse or personality disorder). Given the strong role of environmental factors in modifying the epigenome, our positive findings may be confounded by participants' histories, rather than representing their propensity for suicidality.

While of small individual effect size, this exploratory study nevertheless supports the hypothesis that individuals with BD and a history of SB may be epigenetically distinct from non-suicidal BD patients, and that these differences may suggest a differential aging trajectory that may have an impact of disease phenotype. Subsequent studies should focus on validation of our identified loci using targeted sequencing methods, and further investigation of their corresponding histone modifications in multiple tissues, including *post-mortem* brain, to identify their potential role in molecular networks which may predispose to or be reflective of SB.

## 3.6 Conclusion

We surveyed the epigenome by microarray to investigate the symptomatic complexity of suicide in BD, and observed significant epigenetic differences in several regions between BD patients with and without a history of high suicidality scores. This study is also the first to report on molecular aging acceleration effects in tissue from individuals with a history of SB, which may be of importance in improving understanding of molecular changes resulting from suicide attempt. Our findings may be relevant for identification of mutable substrates in pathways linked to suicidal- and stress-related behaviors in the context of mood and affective disorders. Further comparative epigenetic analysis on the basis of SB would improve understanding of the epigenetic interplay between inherited and environmental effects in BD, and provide an improved understanding of the epigenetic pathways that correspond to vulnerability to future SB, with the ultimate goal of preventing this fatal outcome in an already at-risk population.

# 4 Lactase non-persistence is directed by DNA-variation-dependent epigenetic aging

This chapter is modified from the following:

*Labrie, V., *Buske, O. J., Oh, E., Jeremian, R., Ptak, C., Gasiunas, G., Maleckas, A., Petereit, R., Zvirbliene, A., Adamonis, K., Kriukiene, E., Koncevicius, K., Gordevicius, J., Nair, A., Zhang, A., Ebrahimi, S., Oh, G., Siksnys, V., Kupcinskas, L., Brudno, M., Petronis, A. (2016). Lactase nonpersistence is directed by DNA-variation-dependent epigenetic aging. *Nat Struct Mol Biol*, 23(6), 566-573. doi:10.1038/nsmb.3227

* Equal contribution

## 4.1 Abstract

Inability to digest lactose due to lactase non-persistence is a common trait in adult mammals, with the exception of certain human populations that exhibit lactase persistence. It is not clear how the lactase gene (*LCT*) can be dramatically downregulated with age in most individuals, but remains active in some. We performed a comprehensive epigenetic study of the human and mouse intestine using chromosome-wide and locus-specific DNA modification profiling and targeted bisulfite sequencing. Epigenetically-controlled regulatory elements were found to account for the differences in lactase mRNA levels between individuals, intestinal cell types and species. The importance of these regulatory elements in modulating lactase mRNA levels was confirmed by CRISPR-Cas9-induced deletions. Genetic factors direct epigenetic changes occurring with age at the regulatory elements, as lactase persistence/non-persistence DNA haplotypes demonstrated markedly different epigenetic aging. Thus, genetic factors facilitate a gradual accumulation of epigenetic changes with age to affect phenotypic outcome.

## 4.2 Introduction

Lactose is the major carbohydrate component of milk and other dairy products. Digestion of lactose is mediated by lactase phlorizin hydrolase, an enzyme uniquely expressed in epithelial cells of the small intestine and encoded by the lactase (*LCT*) gene. In mammals, *LCT* expression initiates before birth, remains high during the nursing period and then progressively declines after weaning, resulting in lactase non-persistence. (Rasinpera, Kuokkanen, et al., 2005; Sahi, Launiala, & Laitinen, 1983). The age of onset of lactase non-persistence varies widely across populations, ranging between 1 and 20 years (mean 8.9 S.D. 5.5), and has even been reported as late as 42 years (Hegar & Widodo, 2015; Sahi, Isokoski, Jussila, & Launiala, 1972; Schirru et al., 2007; Seppo et al., 2008; Simoons, 1980; Tadesse, Yuen, & Leung, 1991; Wang et al., 1998). However, some humans are lactase persistent, for which the level of *LCT* mRNA, and consequently LCT activity, remains high in adulthood, especially in the jejunum of the small intestine. Human lactase non-persistence, the ancestral state of *LCT* regulation, segregates as an autosomal recessive Mendelian trait and causes lactose intolerance in over 65% of humans worldwide (Ingram, Mulcare, Itan, Thomas, & Swallow, 2009). Despite its prevalence, the molecular mechanisms mediating lactase non-persistence are not well understood.

To date, inter-individual differences in *LCT* expression in human adults have been attributed solely to DNA sequence variation upstream of *LCT*. In particular, the C/T-13910 single nucleotide polymorphism (rs4988235) found in intron 13 of the minichromosome maintance component complex 6 gene (*MCM6*) is associated with lactase persistence/non-persistence in European populations (Enattah et al., 2002). Lactase persistence in Europe is attributed to the T-13910 allele, part of an extended conserved >500 kb haplotype that has been under strong directional selection for the last 3-7,500 years (Bersaglieri et al., 2004; Enattah et al., 2007; Gamba et al., 2014; Itan, Powell, Beaumont, Burger, & Thomas, 2009; Poulter et al., 2003). This and other nearby genetic polymorphisms associated with *LCT* expression have been detected in other ethnic groups as well (Ingram et al., 2009; Jones et al., 2013; Ranciaro et al., 2014; Tishkoff et al., 2007), and across populations lactase persistence originated on several different haplotype backgrounds (B. L. Jones et al., 2013; Ranciaro et al., 2014; Tishkoff et al., 2007). In non-Europeans, these SNPs have been reported to explain 21-45% of lactase persistence, and certain individuals within African populations exhibit lactase persistence in the absence of known *LCT*-associated variants (Ingram et al., 2009; Ranciaro et al., 2014). Though

reporter assays examining *LCT* promoter induction and gel shifts (Jensen et al., 2011; Olds & Sibley, 2003; Troelsen, Olsen, Moller, & Sjostrom, 2003), and transgenic mouse experiments (Fang et al., 2012), support the importance of SNP C/T-13910 in *LCT* regulation in European populations, the reported differences between allelic variants are modest (T-13910 variant shows a ~1.5-fold higher activity versus ancestral variant) (Jensen et al., 2011). This suggests that, in addition to SNPs, there may be other factors contributing to *LCT* mRNA differences in lactase persistent and non-persistent individuals. It is also unclear what molecular mechanisms can account for the age-dependent changes of *LCT* expression; from very high in infancy to near complete inactivation in the majority of adults. Since DNA sequence is stable, more dynamic regulatory systems must be involved in the temporal dimension of lactase non-persistence.

Epigenetic modifications of DNA and histone proteins could contribute to lactase non-persistence as they effectively regulate gene transcription (Kundaje et al., 2015), differ markedly across tissues and cell types (Elliott et al., 2015; Kundaje et al., 2015), and also change in the same individual over time (Bell et al., 2012; Horvath, 2013). Indeed, evidence suggests that many manifestations of aging, including age-dependent diseases, have an epigenetic basis (De Jager et al., 2014; Gjoneska et al., 2015; Kundaje et al., 2015). In light of this, we conducted the first investigation examining the extent to which epigenetic processes mediate the age-dependent and cell type-specific downregulation of *LCT* that results in lactase non-persistence. In this study, we performed chromosome-wide profiling of DNA modification (consisting of methylation and other epigenetic cytosine modifications) using high density tiling microarrays, followed by targeted bisulfite sequencing-based interrogation of the human and mouse lactase genes in intestinal cells and other tissues. We detected differences in DNA modification densities at several distinct regulatory elements that direct the gradual decline in lactase gene expression following infancy in mammals. We then explored how genetic factors can impact age-specific changes in epigenetic marks, and found that SNP C/T-13910 containing haplotypes direct the epigenetic aging of *LCT/MCM6*. To validate the epigenetically-controlled regulatory elements for the lactase gene, we used RNA interference (RNAi) in human tissue culture and CRISPR-Cas9-induced genetic deletions in the mouse. Our study reveals that lactase non-persistence results from accumulation of transcriptionally suppressive epigenetic changes on haplotypes carrying the SNP C-13910 allele, while T-13910 containing haplotypes escape from epigenetic inactivation to facilitate lactase persistence.

## 4.3 Materials and Methods

### 4.3.1 Human samples

Jejunum specimens were collected from individuals of Northern European descent during laparoscopic Roux-En-Y gastric bypass surgery for obesity in the Departments of Gastroenterology and Surgery, Lithuanian University of Health Sciences (Kaunas, Lithuania). Two cohorts of jejunum surgical samples from unrelated individuals were collected: cohort 1 (n = 56 individuals, 21–72 years old) and cohort 2 (n = 59 individuals, 22–65 years old; Figure 4-1) Following removal, jejunum samples were immediately submerged in RNA*later* (Life Technologies), kept overnight at 4°C, followed by storage at −80°C. Blood (n = 58 individuals) and sperm (n = 18 individuals) samples were collected from unrelated individuals enrolled in laparoscopic surgery, and were from separate individuals than those in cohort 1 and 2. Blood samples were collected by venipuncture into heparin-coated tubes (BD), and then stored immediately at −80°C. Sperm samples provided by participants were stored immediately at −80°C. The study protocol was approved by the Bioethics Committee of Lithuanian University of Health Sciences (Protocol 2007-12-04 Nr.BE-2-55), and each patient signed informed consent to participate in the study. Human tissues were processed for mRNA analysis, DNA modification profiling and genotyping by researchers blind to any sample information.

### 4.3.2 Mouse jejunum samples

All animal procedures were approved by the Institutional Animal Care Committee of the Toronto Centre for Phenogenomics (TCP) and complied with the requirements of the Canadian Council on Animal Care and Province of Ontario Animals for Research Act. C57BL/6NCrl mice were bred at the TCP and housed in ventilated polycarbonate cages, and given ad libitum sterile food (Harlan 2918X) and water. Adult mice were housed by sex in groups of 2-5 littermates. The vivarium was maintained under controlled temperature (21°C [S.D. 1°C]) and humidity (50-60%), with a 12-h diurnal cycle (lights on: 0700-1900).

The jejunum of infant (postnatal day 6) and adult (postnatal day 60) mice was investigated. The small intestine was harvested and dissected into three equal parts; the duodenum, jejunum and ileum. Each of these anatomical segments was further dissected into three equal parts (proximal,

middle and distal) resulting in nine dissections of the small intestine. Only male mice were used in the chromosome-wide and targeted bisulfite sequencing study. For the CRISPR-Cas9 genetically-modified mice, approximately equal male and female mice ratios were used for each group, and no sex differences in steady-state *Lct* mRNA were observed. No statistical methods were used to predetermine sample size, but sample sizes were comparable to other studies involving mRNA analysis in mice with CRISPR-Cas9-induced mutation (Platt et al., 2014; Yin et al., 2014). No animals were excluded from the CRISPR-Cas9-based validation experiment. Mice were randomly allocated to the postnatal day 6 or 60 groups. Variance of group sample sizes was similar between groups in all analyses. Mouse intestinal samples were processed by researchers blind to genotype and experimental conditions.

### 4.3.3 mRNA analyses

Sections of small intestine (~30 mg) were homogenized with a hand-held homogenizer on ice (VWR; human jejunum samples) or a ceramic bead-based homogenizer (Precellys; for mouse jejunum samples) in 1 mL RLT-buffer (Qiagen) supplemented with 10 microlitres (µl) β-mercaptoethanol. Total RNA was subsequently extracted using RNeasy Mini Kit (Qiagen) according to the manufacturer's instructions. Purified RNA samples were treated with RNase-free DNase I (Qiagen) at room temperature for 30 minutes and then repurified with the RNeasy Mini Kit. RNA yield was quantified using UV spectrophotometry (NanoDrop ND-1000, Thermo Fisher Scientific), and RNA integrity was verified on an Agilent Bioanalyzer 2100 system (Agilent Technologies). Total RNA from human epithelial colorectal adenocarcinoma (Caco-2, ATCC) cells was extracted using TRIzol reagent (Life Technologies) according to the manufacturer's instructions. Extracted RNA samples were treated with DNase I (Qiagen), repurified using the RNeasy Mini Kit (Qiagen), and quantified and analyzed as described above. Total RNA (750 nanograms [ng]) was converted to cDNA using High-Capacity RNA-to-cDNA Kit (Life Technologies). For each sample, a negative control (without reverse transcriptase treatment) was prepared to confirm the absence of DNA contamination. Steady-state mRNA levels were assayed with TaqMan Gene Expression Master Mix (Life Technologies) using Applied Biosystems ViiA 7 Real-time PCR system. qPCR cycling conditions were: 50°C for 2 min, 95°C for 10 min and 40 cycles of (95°C for 15 sec, 60°C for 1 min). Samples were ran in

triplicate for each gene and a reference sample was used for inter-plate normalization. TaqMan gene expression probes (Life Technologies) for mouse and human are detailed in Appendix A-3. mRNA of villin 1 and sucrase-isomaltase were used as endogenous controls. ΔΔCt was used to calculate the relative steady-state mRNA levels of each sample. Detailed statistical analyses of the qPCR data were performed using Statistica (Statsoft). The data were analyzed using repeated-measure (RM) ANOVA with the appropriate between-subjects and within-subject factors. Significant effects and interactions were further analyzed by Tukey's honest significant difference (HSD) post hoc comparisons or by unpaired two-tailed Student's t-test to 95% confidence intervals.

### 4.3.4 Intestinal epithelial cell isolation and DNA extraction

A chelating method was used to isolate intestinal epithelial cells (enterocytes) from the rest of the jejunum (Weiser, 1973). Since lactase expression is high in enterocytes of the intestinal villi, while low in those of the crypts (Goda, Yasutake, Tanaka, & Takase, 1999; Yeh, Yeh, & Holt, 1991), we optimized the isolation protocol to yield enterocyte fractions principally from the villi (Figure 4-2). Jejunum samples were washed four times in 1 mL citrate buffer (27 mM Nacitrate, 5.6 mM Na2HPO4, 96 mM NaCl, 8 mM KH2PO4, 1.5 mM KCl, pH 7.4) containing 10 µl of DNase inactivation reagent (Life Technologies) with gentle agitation. Enterocytes were released by exposing jejunum samples to 1 mL Ca2+-chelating buffer (1.5 mM EDTA, 0.5 mM DTT, 10 mM NaH2PO4, 154 mM NaCl) at 37°C for 15 min with agitation (700 rpm). The remaining jejunum was separated from the released enterocytes and the previous step was repeated two additional times using fresh Ca2+-chelating buffer. For each individual, the released enterocytes were collected by centrifugation at 1000 rpm for 5 min at 4°C, and the pellet was used for DNA isolation.

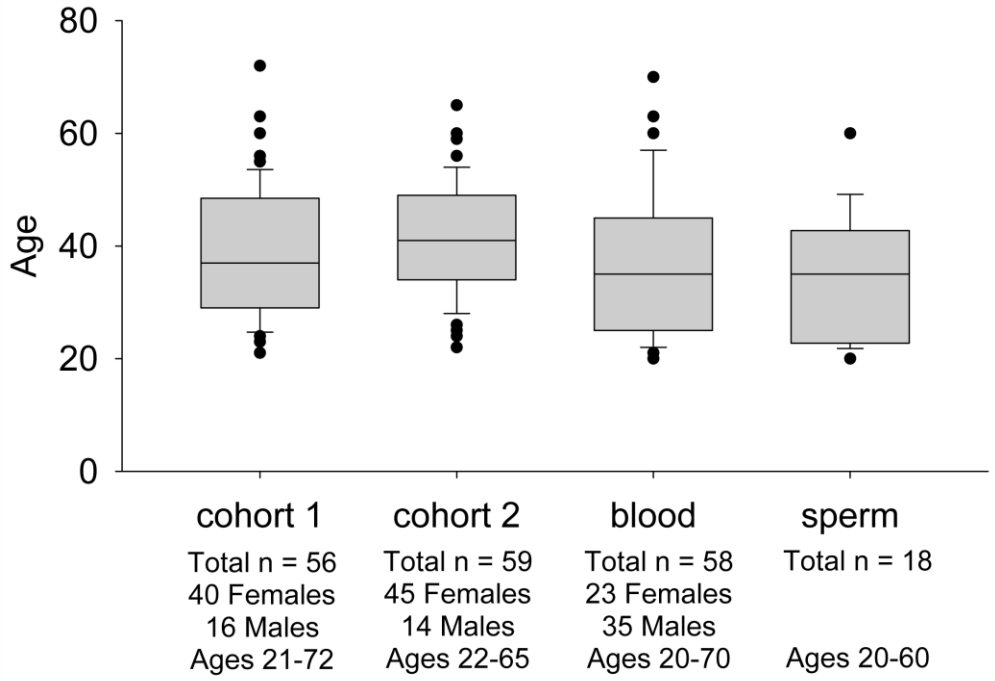



**Figure 4 - 1. Demographic information for human tissue samples**

Box and whisker plots showing the age and sex distribution of the human tissue samples. We attained jejunum surgical samples from 56 (cohort 1) and 59 (cohort 2) unrelated Caucasian individuals of Northern European descent. We also attained 58 blood samples and 18 sperm samples from separate, unrelated Caucasian individuals of Northern European descent.

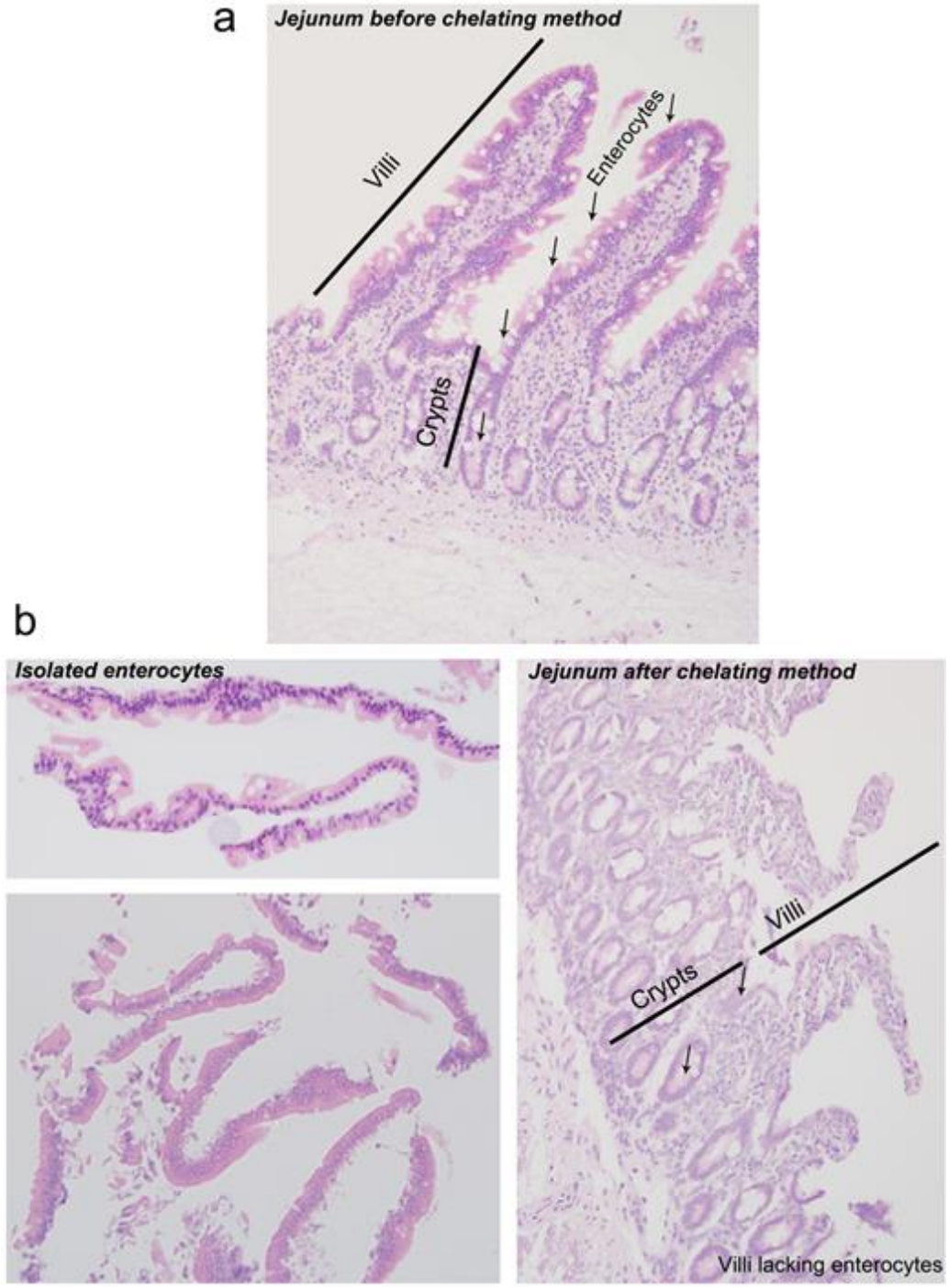



**Figure 4 - 2. Isolation of enterocytes from jejunum tissue samples**

A chelating method specifically removed enterocytes (epithelial cells) from intestinal samples. We optimized the chelating method to remove enterocytes specifically from the villi, where *LCT* expression is most abundant. (a) Jejunum tissue prior to the chelating method (H&E stain, 40X magnification). (b) Left panels: Enterocyte cell fraction isolated by chelating method (40X and 60X magnification). Right panel: Jejunum tissue after the chelating method shows that enterocytes primarily from villi have been isolated (40X magnification). Arrows indicate examples of enterocytes.

DNA was extracted separately for enterocytes and the jejunum lacking enterocytes for each individual using standard phenol-chloroform DNA extraction methods in combination with Phase-Lock tubes (5 Prime). DNA quantity was measured by UV spectrophotometry (NanoDrop ND-1000, Thermo Fisher Scientific), and DNA quality was verified on a 1% agarose gel. For DNA isolation from white blood cells, blood samples were first mixed with 2.5% Triton X100 and cells were pelleted by centrifugation at 400 x g for 15 min at 4°C. The pelleted white blood cells were washed in a buffer (10 mM Tris-HCl pH 7.6, 10 mM KCl, 10 mM MgCl2, 2 mM EDTA), centrifuged again and then DNA was extracted using standard phenol-chloroform methods. Sperm DNA isolation initially involved mixing sperm cells with 0.5% Triton X-100 and 10% SDS and centrifugation at 400 x g for 10 min at 4°C. Sperm cell pellets were washed once gently in PBS, pelleted again, and DNA was isolated with a standard phenol-chloroform approach. DNA quantity and quality was measured as described above.

### 4.3.5 Chromosome-wide DNA modification analysis

To perform our chromosome-wide scan we enriched the unmodified DNA fraction using the mTAG-click method, as described previously (Kriukiene et al., 2013). Briefly, genomic DNA was sonicated using a Covaris S220 instrument (200 cycles/burst, 10% duty factor, 3 min/sample) in 10 mM Tris-HCl (pH 8.5) to yield fragments with a peak size of 200 bp. Sheared DNA fragments were blunt-ended using the Fast DNA End Repair Kit (Thermo Scientific). DNA was then purified using the ZR-96 DNA Clean & Concentrator-5 kit (Zymo Research). For each sample, mTAG-click labeling was performed by mixing 300 ng of sheared and end repaired DNA with 25 μM Ado-6-azide cofactor and 90nM eM.SssI in Labelling buffer (10 mM Tris-HCl pH 7.6, 50 mM NaCl, 0.1 mg/ml BSA) in a 25 μl reaction. Reactions were incubated at 37°C for 30 min, followed by heating at 80°C for 10 min. Labelled unmodified cytosines were then biotinylated with 200 μM dibenzylcyclooctyne-S-S-PEG3-biotin (Click Chemistry Tools) in a 2 h incubation at 37˚C. Samples were again purified using the ZR-96 DNA Clean & Concentrator-5 kit (Zymo Research). Biotinylated DNA fragments were captured with 1.25 mg/ml Dynabeads MyOne Streptavidin T1 (Invitrogen) in 5X BW buffer (25 mM Tris-HCl (pH 7.5), 2.5 mM EDTA, and 5 M NaCl). Samples were mixed with beads at 900 rpm for 2 h at room temperature. The beads were washed three times with 3X BW buffer and twice with 1X BW buffer. A qPCR

reaction comparing an aliquot of each sample bound and unbound to beads confirmed the desired 25% labelling efficiency. DNA was then released from the beads and biotin tags by a 1 h incubation in 50 mM DTT and 60 mM Tris pH 7.8 with mixing at 900 rpm. The recovered DNA was ligated to adaptors (A-25 5'-AGTTACATCTTGTAGTCAGTCTCCA and A-19 5'-TGGAGACTGACTACAAGAT) in a reaction containing 0.5 mM ATP, 10 mM MgCl2, 5% PEG, 3.3 µM adaptors and 10 a.u. T4 Ligase (Thermo Scientific). Samples were ligated overnight at 22˚C followed by an enzyme inactivation step of 65˚C 15 min and gradual cooling to 25 ˚C -1˚C/10 sec. Immediately afterwards, samples were treated with 32 mM β-mercaptoethanol for 10 min at room temperature and PCR-amplified. PCR amplification was done in two rounds. The first PCR was performed by adding buffer (63 mM Tris-HCl pH 8.8, 17 mM (NH4)2SO4, 0.084% (v/v) Tween 20), 587 µM dNTPs, 1.5 µM adaptor-specific A-25 primer and 6.3U Taq polymerase (Thermo Scientific) for a 100 µl reaction. PCR cycling conditions were: 50°C for 1 min, 72°C for 2 min, 94°C for 5 min, 65°C for 1 min, 72°C for 5 min, 15 cycles of (94°C for 1 min, 65°C for 1 min, 72°C for 1 min), and 72°C for 2 min. The second round of PCR amplification contained 5 µl of the PCR product from the first round in Buffer (75 mM Tris-HCl (pH 8.8 at 25°C), 20 mM (NH4)2SO4, 0.01% (v/v) Tween 20), 2 mM MgCl2, 0.4 mM dNTPs with 0.02 mM dUTP, 1.5 µM A-25 primer and 6.3 U Taq polymerase (Thermo Scientific) in a 50 µl reaction. PCR cycling conditions: 95°C for 1 min, 10 cycles of (93°C for 30 sec, 65°C for 1 sec, 68°C for 90 sec + 5 sec/cycle), and 72°C for 5 min. PCR products were verified on a 2% agarose gel and purified with the QIAquick 96 PCR Purification kit (Qiagen).The unmodified DNA fraction was hybridized to GeneChip Human Tiling 2.0R Array (B array, covering chromosomes 2, 9 and 19) or GeneChip Mouse Tiling 2.0R Array (A array, chromosome 1, 9 and 19). Microarrays were processed as detailed in the Affymetrix ChIP Assay protocol and scanned using an Affymetrix GeneChip Scanner.

Tiling array data of isolated enterocytes and enterocyte-deficient jejunum were examined for human and mouse samples. For the microarray hybridization, 7 µg of purified uracil-containing amplicons were fragmented to 50–100 bp using GeneChip Double-stranded DNA Terminal Labeling kit (Affymetrix). Individual samples were hybridized to a separate GeneChip Human Tiling 2.0R Array (B array, covering chromosomes 2, 9 and 19) or GeneChip Mouse Tiling 2.0R Array (A array, chromosome 1, 9 and 19). The arrays were washed, stained and scanned using an Affymetrix GeneChip Scanner as described in the Affymetrix ChIP Assay protocol.

For the human samples, the tiling array data of isolated enterocytes and enterocyte-deficient jejunum from 56 individuals were analyzed, along with 10 input control samples (no mTAG-click enrichment). Replicates were included for 22 samples (7 sample pairs of enterocytes and 15 pairs of enterocyte-deficient jejunum). For the mouse samples, enterocytes from mice at postnatal day 6 (n = 14 mice) and 60 (n =15 mice) were examined. Tiling arrays intensities were extracted using bioconductor R package "AffyTiling". The mean intensities across input tracks were subtracted, and the intensities for the relevant chromosomes (human Hs-NCBIv36:chr2, mouse Mm-NCBIv33:chr1) were quantile normalized and log(2) transformed. Probe sequences were re-aligned to the genomes for each species (human GRCh37/hg19, mouse GRCm38/mm10) using mrfast-2.6.1.0 with no mismatches allowed. Reference genomic sequences were downloaded from the UCSC, with repetitive sequences masked ("random" and "chrUn" chromosomes were ignored). Only intensities for uniquely mapping probe sequences were considered. For each sample, normalized intensities were then smoothed using a running pseudomedian filter with span of 200 bp (Royce, Carriero, & Gerstein, 2007). The mean correlation of each sample to other samples of the same tissue was calculated, and samples with correlations less than $-2\sigma$ were excluded from further analysis (n = 4 human, n = 1 mouse). One additional human sample was excluded due to abnormally high *LCT* mRNA levels ($> \mu + 4\sigma$).Probe sequences were aligned to the genomes for each species (human GRCh37/hg19, mouse GRCm38/mm10) with no mismatches allowed, and only intensities for uniquely-mapping probe sequences were considered.

Probe filtering for the human samples involved a bootstrapping procedure that was used to identify the probes that were significantly more consistent between pairs of replicates than random samples of the same tissue. After filtering, 20 pairs of replicate samples remained (6 for enterocytes, 14 for enterocyte-deficient jejunum). First, the Pearson correlation was computed between each pair of replicates for every probe. Then the significance of this correlation was assessed against an empirical null distribution, which was calculated by measuring the correlation of 20 random pairs of samples, for each of 1000 iterations. Random pairs were chosen in a stratified fashion, such that they were always from the same class (both enterocyte or both enterocyte-deficient jejunum), while maintaining the same ratio of enterocyte and enterocyte-deficient jejunum samples as in the original replicates. Probes were kept for further analysis only if the average correlation between replicates was greater than or equal to that of

99% of random samplings, resulting in n = 57,583 probes. As expected, replicates clustered tightly together after this filtering step, but samples also clustered distinctly according to the tissue type. Remaining replicates were then averaged together, resulting in a single data point per sample (to avoid imbalance in the statistical analysis). Finally, probes were filtered to the 5,482 probes that were significantly different between enterocytes and rest of jejunum, based on a paired Mann-Whitney test at each probe, Benjamini-Hochberg multiple testing correction, and a false discovery rate (FDR) threshold of $q < 0.01$.

In the analysis of the human samples, the significance of correlation between DNA modification and *LCT* steady-state mRNA was calculated using Kendall's nonparametric correlation test at each probe, followed by Benjamini-Hochberg multiple testing correction. For the mouse samples, age-associated probes were identified by performing a Mann-Whitney test at each probe and then applying the Benjamini-Hochberg correction for multiple testing with a threshold of $q < 0.01$ (n = 2,680,563 sites). The Mann-Whitney and Kendall non-parametric tests were used to assess significance because the array intensity values were not normally distributed. Due to the limited sample size and correlation between nearby probes, the Benjamini-Hochberg method was used to correct for multiple testing. Chromosomes 9 and 19 were subsequently analyzed using the same method.

### 4.3.6 Bisulfite padlock probes for fine-mapping of DNA modification in humans and mice

DNA modification was examined at single nucleotide resolution in the *LCT* and *MCM6* loci of humans and mice using the bisulfite padlock probe technique (Diep et al., 2012). Padlock probes were generated to target the unique (non-repetitive) genome following bisulfite conversion using the ppDesigner software. Probes were designed for both the forward and reverse DNA strands using recent versions of the genome (human GRCh37/hg19, mouse GRCm38/mm10). Human probes (n = 466) surveyed a ~100 kb region surrounding the *LCT* and *MCM6* locus, while mouse probes (n = 404) were designed to cover an ~80 kb region centered over the *Lct* and *Mcm6*. Probe sequences are described in Appendix A-4. Padlock probes were synthesized using a programmable microfluidic microarray platform (LC Sciences). Each probe consisted of 20-25 bp probe segments (used to interrogate a ~180 bp bisulfite-converted region of interest) on either

end of a 30 bp common linker region (used for PCRamplification of captured regions). To prepare the probes, 1 nM of the mixed template oligonucleotides were PCR-amplified in the presence of 400 nM of each eMIP_CA1_F primer and eMIP_CA1_R primer (Appendix A-4), and 1X KAPA SYBR fast Universal qPCR Master Mix. The PCR conditions were: 95°C for 30 sec, 5 cycles of (95°C for 5 sec, 50°C for 1 min, 60°C for 30 sec), 12 cycles of (95°C for 5 sec, 60°C for 30 sec, 72°C for 30 sec), and 72°C for 10 min. PCR amplicons were purified with QIAquick PCR purification columns (Qiagen). The purified amplicons were then re-amplified in 15 PCR reactions each containing 0.02 nM of amplicon, 400 nM each of eMIP_CA1_F primer and eMIP_CA1_R primer, and 1X KAPA SYBR fast Universal qPCR Master Mix. The PCR conditions were: 95°C for 30 sec, 15 cycles of (95°C for 5 sec, 60°C for 30 sec, 72°C for 30 sec), and 72°C for 10 min. These PCR reactions were pooled and amplicons were purified with QIAquick PCR purification columns (Qiagen). Probes (4 µg) were then digested with wild type AlwI (10 U/µl, NEB) and Nb.BrsDI (10 U/µl, NEB). Both digestions were performed in cutSmart buffer (NEB) at 37°C for 1 h, 55°C for 20 min, 60°C for 20 min, 65°C for 1 h, followed by an enzyme inactivation step of 80°C for 20 min. The probes were purified with a QIAquick PCR purification column (Qiagen). Further purification was completed by running the probes on a 6% denaturing Urea-PAGE gel and electro-eluted using D-Tube Dialyzer 6-8KDa tubes (Novagen).

Genomic DNA was bisulfite-converted and purified using the EZ DNA Methylation-Lightning Kit (Zymo Research). In tandem, whole genome amplified DNA (WGA1 kit, Sigma), a genome devoid of DNA modifications, was also bisulfite-converted and used to verify the high efficiency of the bisulfite conversion reactions (i.e. ~98% conversion in human samples). For each sample, 200 ng of bisulfite-converted DNA was hybridized to 0.12 ng of probes in 1X Ampligase Buffer (EpiCentre), using the following reaction conditions: 98°C for 3 min, 85°C for 30 min, gradual cooling to 60°C (-1°C/min), 60°C for 60 min, gradual cooling to 56°C (-1°C/min), 56°C for 300 min, gradual cooling to 45°C , (-1°C/min), and 45°C for a minimum of 120 min. Target region extension and circularization was completed in a reaction containing 1X Ampligase Buffer (Epicentre), 2.5 U PfuTurbo Cx (Agilent Technologies), 0.4 mM dNTPs, 5 U Ampligase (Epicentre), 200 µM NAD (NEB) that was incubated at 56°C for 1 hr, followed by enzyme inactivation at 72°C for 20 min. Each sample was then digested with exonucleases, using 8U exonuclease I (USB), 40 U exonuclease III (USB), 6 U RecJf, 2.5 U Lambda

exonuclease, and 0.33X exonuclease III buffer (USB) in an incubation of 37°C for 30 min, 80°C for 20 min, 95°C for 5 min, and kept at 4°C. The samples were then PCR-amplified by mixing an 8 µl aliquot of DNA template with 1X PfuTurbo Cx Buffer, 0.25 mM dNTPs, 2.5 U PfuTurbo Cx DNA polymerase (Agilent Technologies), 0.45 µM AmpF6.4Sol Primer, 0.45 µM Reverse barcode primer (primers detailed Appendix A-4) in a 50 µl reaction. PCR conditions were: 98°C for 3 min, 35 cycles of (98°C for 30 sec, 50°C for 30 sec, 72°C for 1 min) and 72°C for 10 min. The PCR amplicons were visualized on a 2% agarose gel. Library pools were generated, containing ≤ 68 samples per pool. For each library pool, the desired ~330 bp band was excised from the agarose gel and purified using the QIAquick Gel Extraction kit (Qiagen). The purified libraries were then submitted for next-generation sequencing on an Illumina HiSeq 2500 machine in Rapid Run mode at the Donnelly Sequencing Centre in Toronto, Canada. Technical replicates contained in different sample pools confirmed a high reproducibility between the sequencing runs ($r > 0.99$).

### 4.3.7 Analysis of bisulfite padlock probes data

Using a custom pipeline (Buske, 2015) based on the Bismark tool (Krueger & Andrews, 2011), DNA modification density was interrogated at every cytosine covered by padlock probes. Reads were first trimmed and residual adapter sequences removed using Trimmomatic-0.32. To remove residual phiX DNA spiked-in as a sequencing control, reads were aligned to the phiX reference genome (NC_001422.1) using Bismark (560 bp from the beginning of the phiX genome was appended to the end to accommodate the circular nature of the genome for read alignment). Reads that did not align to phiX were aligned to the target reference genome (GRCh37/hg19 or GRCm38/mm10). Percent modification was estimated as the fraction of spanning reads that retained the reference "C", and were not converted to "T" from the bisulfite treatment. Modification estimates were only considered if 30 or more reads spanned the cytosine. For human intestine (enterocytes and enterocyte-deficient jejunum), sperm, and blood samples, DNA modification density was interrogated at 891 CpGs (14,830 cytosines) across 301 samples (plus 46 replicates). CpG sites were filtered for effect size and sparsity by excluding sites with an interquartile range < 0.05 (376 CpGs) or with coverage in fewer than 30 intestinal samples (233 CpGs). Exceptionally sparse samples with more than $\mu + 2\sigma$ missing data (18 samples) and outliers (14 samples with a mean pairwise Pearson correlation with other samples of the same

tissue < $\mu - 2\sigma$) were excluded. Remaining replicates (n = 35) were merged by taking the mean across replicates at each CpG site, resulting in 58 blood, 18 sperm, 100 enterocyte, and 104 enterocyte-deficient jejunum samples (280 samples overall). To avoid genomic variants confounding the bisulfite sequencing results, we extracted the positions of all common polymorphisms called as part of the 1000 Genomes Project (phase 3, v5a release; (Genomes Project et al., 2012)). Polymorphisms were filtered to those with allele frequencies ≥ 5% either across all populations or within just the European (EUR) population (most concordant with our human subjects). Cytosines within 1 bp of these polymorphisms were excluded from further analysis (17 CpGs). The correlation was measured between the DNA modification density at each remaining CpG (n = 282) and relative steady-state *LCT* mRNA using Kendall's nonparametric correlation test. Bonferroni multiple testing correction was then performed and CpGs were considered significant at a family-wise error rate threshold $\alpha < 0.01$. CpGs exhibiting cell-type specific expression were identified by performing a nonparametric Mann-Whitney test at each CpG between the sets of enterocyte and enterocyte-deficient jejunum samples, followed by Bonferroni multiple testing correction and a significance threshold of $\alpha < 0.01$. The Kendall and Mann-Whitney tests were selected because neither the DNA modification density nor the steady-state *LCT* mRNA was normally distributed.

For mouse samples, DNA modification density was interrogated at 859 CpGs (11,611 cytosines) in 29 enterocyte samples. CpG sites were filtered for effect size and sparsity by excluding sites with an interquartile range < 0.05 (236 CpGs) or with coverage in fewer than 15 enterocyte samples (498 CpGs). The significance of association between DNA modification and age was measured separately at each CpG (n = 298) by comparing the modification levels of mice at postnatal day 6 (n = 16 mice) and day 60 (n = 13 mice) with a Mann-Whitney nonparametric test. This test was selected because the DNA modification densities were not normally distributed. Multiple testing correction was performed using the Benjamini-Hochberg method with a FDR threshold $q < 0.05$.

### 4.3.8 Analysis of chromatin state

NarrowPeak peak calls were downloaded for 4 tracks (DNase, H3K4me1, H3K4me3, H3K27ac; see Appendix A-2) for human small intestine tissue from Roadmap Epigenomics uniform re-

processing unconsolidated epigenomes from Release 9 of the Epigenome Atlas (Roadmap Epigenomics et al., 2015) (http://egg2.wustl.edu/roadmap/data/byFileType/peaks/unconsolidated/narrowPeak/; Supplementary Table 2). Peak calls were filtered for significance (FDR $q < 0.01$) and intersected with the seven *LCT*-associated regions in Supplementary Table 1 (using bedtools (Quinlan & Hall, 2010)). *LCT*-associated regions smaller than 500 bp were symmetrically extended to 500 bp. A 100 kb region encompassing *LCT* and *MCM6* (chr2:136,540,000–136,640,000 [hg19]) was scanned for potential enhancer regions by identifying all overlapping H3K4me1 and H3K27ac peaks annotated with an FDR $q < 0.01$. For the mouse small intestine, we obtained processed data, including BroadPeak calls, for 4 tracks (POLR2A, H3K4me1, H3K4me3, H3K27ac) from ENCODE (www.encodeproject.org). These originated from the small intestine of 60-day old C57BL/6NCrl mice (same age and strain as used in the bisulfite padlock probe sequencing). Peak calls were lifted over from mm9 to mm10 using the UCSC LiftOver tool, and then intersected with the interrogated 80 kb region containing *Lct* and *Mcm6* (chr1:128,280,000–128,360,000 [mm10]).

### 4.3.9 Analysis of mammalian conservation

The significance of intronic conservation within mammals was assessed using 12-mer constraint predictions from a multiple sequence alignment of 29 mammals (Garber et al., 2009) (http://www.broadinstitute.org/ftp/pub/assemblies/mammals/29mammals/hg19/hg19_29way_pi_lods_elements_12mers.chr_specific.fdr_0.1_with_scores.txt.gz). Intronic regions were defined as the intervals of Refgene regions at least 100bp away from the nearest exon and within a 100 kb region centered on *LCT* and *MCM6* (chr2:136,540,000–136,640,000), resulting in 33 intronic intervals spanning 67,744 bp. Mammalian conserved elements were intersected with these intronic intervals using bedtools to identify 47 intronic conserved elements within *LCT* and *MCM6* (758 conserved bases, or 1.12% of intronic sequence). The expected number of conserved bases within a 500 bp window was modeled by randomly permuting the locations of the 47 conserved elements within the intronic intervals using bedtools *shuffle* and measuring the total number of conserved bases within a 500 bp region, for each of 10,000,000 iterations. A set of 25 overlapping 500 bp windows were sufficient to span all 47 conserved

elements, and the significance of each window was measured by comparing the number of conserved bases within the window to the permuted distribution.

### 4.3.10 RNAi technique

The relationship between lncRNA (*LOC100507600*) and *LCT* mRNA levels was examined by RNA interference in Caco-2 cells (ATCC). For this approach, Caco-2 cells were transfected with either an equimolar pool of 4 siRNAs (30 pmol) directed to *LOC100507600* (Qiagen FlexiTube GeneSolution; SI05737893, SI05737900, SI0737907 and SI0737914) or a scrambled-control siRNA (Qiagen AllStars Neg. siRNA AF488, SI03650318) using RNAiMAX (Life technologies) according the manufacturer's instructions. The siRNA or scrambled control group was randomly assigned to each cell well. After 48 h, transfection efficiency was confirmed for siRNAs conjugated to AlexaFluor-488 using an Olympus FV1200 confocal microscope and FluoView 1200 software (Olympus). Cells were then harvested and RNA was extracted using TRIzol reagent (Life Technologies). mRNA levels were analyzed by qPCR, as described above, to evaluate the consequence of *LOC100507600* knockdown on *LCT* expression. Sample sizes (n = 8 scrambled control-treated, 11 RNAi-treated cell cultures) were comparable to other studies involving RNAi (Gong, 2011; already in Endnote).

### 4.3.11 CRISPR-Cas9 genome-editing

Genetic deletions in the *Mcm6* intron 13, *Lct* intron 2 or *Lct* intron 1 of mice were generated using the CRISPR-Cas9 genome editing technique. Design of single guide RNAs (sgRNA) targeting these loci was completed with the MIT CRISPR Design tool (Ran, Hsu, Wright, et al., 2013). Guides were selected based on highest on-target specificity and avoided exons, intronic regions adjacent to exons (at least 100 bp), and repetitive regions.

For the Cas9 RNA-guided nuclease modified mice, sgRNA templates for *in vitro* transcription were amplified by high-fidelity PCR using a forward primer with the T7 promoter, the guide RNA sequence, and an anchor sequence for the pX330 plasmid (gift of Feng Zhang, Addgene 42230) and a reverse primer at 3′ terminus of the sgRNA sequence in pX330 (Primers detailed in Appendix A-3). The PCR templates were gel purified and used for *in vitro*

transcription using the MEGAshortscript T7 transcription kit (Life Technologies) followed by purification using the MEGAclear kit (Life Technologies). Cas9 mRNA was produced by *in vitro* transcription of a T7 promoter containing amplicon from pX330 using the mMESSAGE mMACHINE T7 Ultra Kit (Life Technologies). Cas9 mRNA (20 ng/μL) and sgRNA (10 ng/μL) were premixed in injection buffer (10 mM Tris, 0.1 mM EDTA, pH 7.5) and microinjected via continuous flow into the pronuclei of C57BL/6NCrl zygotes at the Toronto Centre for Phenogenomics, Canada. Mutation success rates were 18% for *Mcm6* (n = 2 mice), 11% *Lct* intron 1 (n = 2) and 38% for *Lct* intron 2 (n = 5). $F_0$ mice were backcrossed to wild-type C57BL/6NCrl mice and $F_1$ heterozygous mice were intercrossed, keeping each founder line separate. Of the 9 mutant lines, we investigated only one for each *Mcm6* and *Lct* intron 1, and three lines for *Lct* intron 2 (the three *Lct* intron 2 lines did not differ). $F_2$ offspring from the heterozygous mouse intercrosses were used in experiments.

For cell-based analysis, we used the Caco-2 cell line (colorectal adenocarcinoma; HTB-37; ATCC, Manssas, VA, USA), a model of human male Caucasian intestinal epithelial cells that is hyperdiploid (modal number 96), and exhibits chromosomal, as well as microsatellite instability (Gayet et al., 2001; Hermsen et al., 2002; Lengauer, Kinzler, & Vogelstein, 1997; Shibata, Peinado, Ionov, Malkhosyan, & Perucho, 1994). For genetic modifications in this line, we opted for the CRISPR-Cas9n system, which has been shown to increase genome editing specificity in cells (Ran, Hsu, Lin, et al., 2013). We generated deletions in *MCM6* intron 13, *LCT* intron 2 or *LCT* intron 1 in the human cells using the CRISPR-Cas9n protocol described previously (Ran, Hsu, Wright, et al., 2013). Caco-2 cells in the CRISPR-Cas9n and RNAi experiments were authenticated; showing a 100% match to the reference profile of Caco-2 (ATCC HTB-37) based on genetic analyses of 15 autosomal tandem repeat loci and the gender identity locus Amelogenin (performed by Genetica DNA Laboratories). A PCR-based test using the eMYCO Plus kit (iNtRON Biotechnology) verified the absence of mycoplasma contamination in the Caco-2 cells used (performed by Genetica DNA Laboratories).

CRISPR-induced deletions in both the mice and Caco-2 cells were confirmed by Sanger sequencing (ACGT Corporation in Toronto, Canada). *LCT* steady-state mRNA levels were measured in the CRISPR-Cas9-modified mice and CRISPR-Cas9n-modified Caco-2 cells using a qPCR assay (described above).

### 4.3.12 Genotyping and analysis of CRISPR-induced deletions

Human jejunum samples were genotyped for the genetic variants C/T-13910, G/A-22018 and the deletion in *LCT* intron 1. PCR amplicons were generated using the KAPA2G Fast kit (Kapa Biosystems) and 0.6 μM of each forward and reverse primer (Primers detailed in Appendix A-3) in a 25 μl reaction. PCR cycling conditions were: 95°C for 3min, 30 cycles of (98°C for 20 sec, 58.5°C (C/T-13910) or 61.5°C (G/A -22018, LCT intron 1 deletion) for 15 sec, 72°C for 15 sec) and 72°C for 1 min. Genetic variants were identified by Sanger sequencing (C/T-13910; ACGT Corporation in Toronto, Canada), restriction endonuclease digestion with BstUI (G/A -22018; NEB) or by PCR amplicon size[5] (deletion in *LCT* intron 1). Linkage disequilibrium and haplotype structure between these genetic variants was examined using the Haploview version 4.2 program (Barrett, Fry, Maller, & Daly, 2005).

For the CRISPR-Cas9 modified mice, genotyping was conducted with genomic DNA isolated from mice tail-clippings using a standard high salt method. Genomic DNA (100 ng) was PCR-amplified with 1X Herculase II buffer (Agilent Technologies), 1 mM dNTPs, 0.3 μM of each primer (Appendix A-3) and 1 U Herculase II Fusion DNA polymerase (Agilent Technologies) in a 50 μl reaction. PCR conditions were: 95°C for 7 min, 10 cycles of (95°C for 20 sec, 61°C (Mcm6, Lct intron 1) or 57°C (Lct intron 2) for 20 sec, 68°C for 5 min), 20 cycles of (95°C for 20 sec, 61°C (Mcm6, Lct intron 1) or 57°C (Lct intron 2) for 20 sec, 68°C for 5 min + 20 sec/cycle), and 68°C for 8 min. Mice carrying a CRISPR-Cas9 induced deletion in Mcm6 intron 13, Lct intron 1 or Lct intron 2 were genotyped by surveying PCR amplicon size (detailed in Appendix A-3) on a 1.5% agarose gel. For the CRISPR-Cas9n modifications in Caco-2 cells, genomic DNA was isolated for each cell colony with QuickExtract solution (15 μl/cell pellet, Epicentre) and incubated at 65°C for 6 min, and 98°C for 2 min. Screening for deletions in MCM6 intron 13 and LCT intron 2 involved generating PCR amplicons by combining 200 ng of genomic DNA with 1X Herculase II buffer (Agilent Technologies), 1 mM dNTPs, 0.5 μM of each primer (Appendix A-3), 1 U of Herculase II fusion DNA polymerase (Agilent Technologies) in a 50 μl reaction. PCR conditions were: 95°C for 2 min, 10 cycles of (95°C for 20 sec, 54.4°C (MCM6) or 58.8°C (LCT intron 2) for 20 sec, 68°C for 5 min), 20 cycles of (95°C for 20 sec, 54.4°C (MCM6) or 58.8°C (LCT intron 2) for 20 sec, 68°C for 5 min + 20 sec/cycle), 68°C for 8 min. Deletion in LCT intron 1 involved PCR containing 200 ng of genomic DNA mixed with 1X ThermoPol buffer (NEB), 2 mM MgCl2, 0.2 μM each primer

(Appendix A-3), 0.2 mM dNTPs, 5 U Taq (NEB) in a 100 µl reaction, and PCR cycling conditions were: 95°C for 10min, 40 cycles of (95°C for 30 sec, 58°C for 35 sec, 68°C for 75 sec), 68°C for 4min. Deletions were determined by examining PCR amplicon size (Appendix A-3) on a 1.5% agarose gel.

Off-target analysis for CRISPR-induced deletions was also performed for the mice and cell lines. Based on the ranked list of putative off-target sites provided by MIT CRISPR Design tool (Ran, Hsu, Lin, et al., 2013; Ran, Hsu, Wright, et al., 2013), we examined the most likely candidate regions for sgRNA off-target activity in both the mice with CRISPR-Cas9-induced deletions (2 sgRNA guides/deletion) and Caco-2 human cells with CRISPR-Cas9n-mediated deletions (4 sgRNA guides/deletion). A total of 46 potential off-target sites were tested in the genetically-modified mice and 172 sites were tested in the Caco-2 human cell lines (Appendix A-3 for regions tested and primers used). Candidate off-target regions were amplified via PCR in a 25 µl reaction containing a final concentration of 1X KAPA2G Fast HotStart ReadyMix (KAPA Biosystems), 0.5 µM forward and reverse primers (Appendix A-3) and 100 ng of input DNA. The following cycling conditions were implemented for all reactions: 95°C for 3 min, 30 cycles of (95°C for 15 sec, 56°C for 15 sec, 72°C for 6 sec), 72°C for 90 sec. PCR products were run on a 1.5% agarose gel to verify fragment size. An enzymatic purification was then performed by adding 30 U of exonuclease I (NEB) and 3 U of FastAP Thermosensitive Alkaline Phosphatase (Thermo Scientific), and incubating for 15 min at 37°C, followed by a 5 min inactivation at 75°C. PCR amplicons underwent Sanger sequencing (ACGT Corporation in Toronto, Canada) and sequences were analyzed by alignment with the mouse (GRCm38/mm10) or human (GRCh37/hg19) reference genome. To display the regions investigated we used the CIRCOS program version 0.67 (Krzywinski et al., 2009).

## 4.4 Results

### 4.4.1 Chromosome-wide investigation of lactase gene regulation

We surveyed DNA modification profiles in human enterocytes derived from Northern European individuals (n = 56; mean age 38.8, S.D. 12.0 years; 23 males and 33 females; BMI data not available) who underwent Roux-en-Y gastric bypass surgery at the Lithuanian University of Health Sciences (Kaunas, Lithuania). To explore epigenetic contributions to lactase non-persistence and persistence, we obtained surgical jejunum samples from unrelated human adults (Figure 4-1) and isolated villi enterocytes (intestinal epithelial cells) (Weiser, 1973) from the jejunum (Figure 4-2). Across individuals there was a wide range of steady-state *LCT* mRNA levels; differing up to 29-fold (Figure 4-3).

We then explored chromosome-wide DNA modification patterns in the human enterocyte samples (cohort 1, n = 56). The unmodified DNA fraction was enriched using the mTAG-click method (Kriukiene et al., 2013), followed by interrogation on Affymetrix high density human tiling arrays. The arrays used surveyed human chromosome 2, which contains *LCT* (and the same array also contained chromosomes 9 and 19). Examination across chromosome 2 identified 5 probes that exhibited significant correlation between steady-state *LCT* mRNA and the density of unmodified cytosines (FDR $q = 0.017$, $R^2 = 0.28–0.32$). All 5 probes were clustered in a ~170 bp region of *LCT* exon 1/intron 1, where higher DNA modification densities were associated with lower *LCT* mRNA levels (Figure 4-4 and Appendix A-1). As expected, no significant associations between DNA modification and *LCT* mRNA levels were detected on chromosomes 9 and 19 (Figure 4-5). DNA modifications in the jejunum lacking enterocytes also did not exhibit any significant correlations with *LCT* mRNA (Figure 4-6).

We next investigated whether DNA modification changes are involved in the age-related decline of the lactase gene. For this experiment, we compared chromosome-wide DNA modification profiles of jejunal enterocytes of infant and adult mice (postnatal day 6 *vs*. 60, respectively). Similar to lactase non-persistence in humans, mice show high intestinal lactase activity at birth, followed by a gradual decrease after infancy (Fang, Olds, & Sibley, 2006). Indeed, we found that *Lct* mRNA in the jejunum of inbred C57BL/6N adult mice at postnatal day 60 was reduced by 3-fold compared to infant mice at postnatal day 6 ($p < 10^{-4}$; Figure 4-7). For the epigenetic analysis, we used the mTAG-click method to enrich the unmodified DNA fraction

from infant and adult jejunal enterocytes, followed by interrogation on mouse tiling microarrays. We detected that 0.4% of probes in mouse chromosome 1, which contains *Lct*, exhibited significant differential DNA modification between the two age groups (11,206 probes, FDR $q <$ 0.01; Figure 4-8). Of these differentially-modified probes, 30 probes mapped to within 100 kb of *Lct*, clustering into 3 regions: *Lct* intron 2/exon 3, intron 8/exon 9, and exon 10 (~60-700 bp region size; 1.36- to 2.25-fold change; Figure 4-8 and Appendix A-1). At all these regions, adult animals demonstrated significantly greater densities of modified cytosines compared to infants, indicating an inverse relationship between DNA modification and *Lct* mRNA levels. Thus chromosome-wide analysis revealed epigenetically-modified regions in the lactase gene of both mice and humans, which may be important for its transcriptional regulation and the establishment of lactase non-persistence.

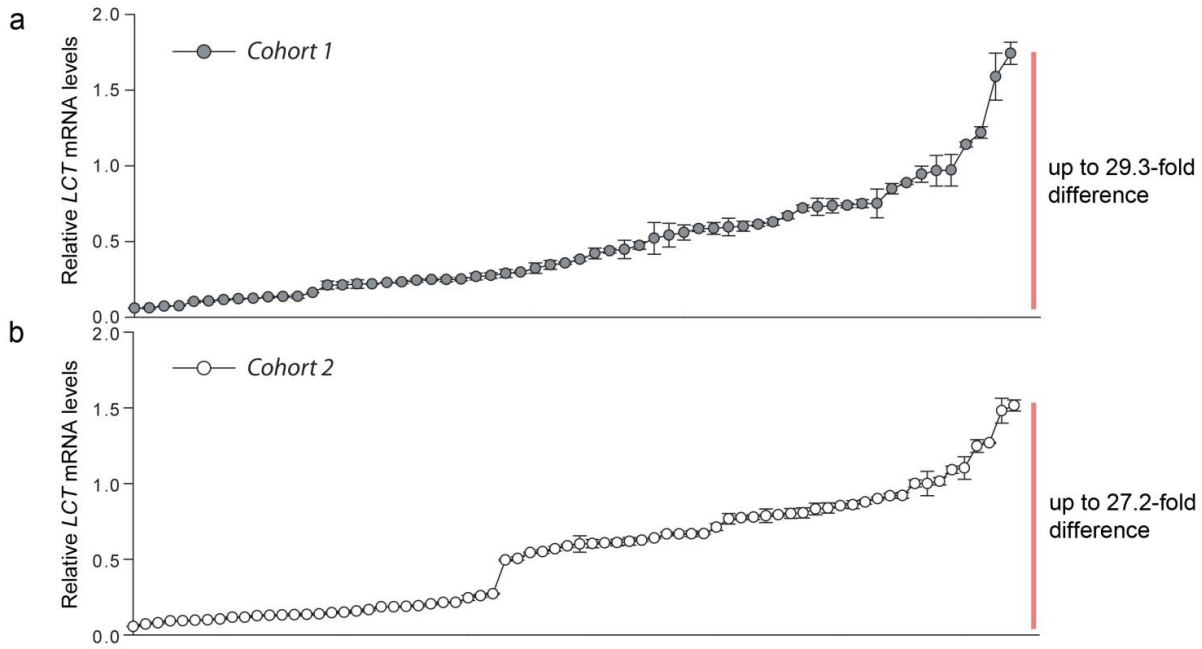



**Figure 4 - 3. Steady-state LCT mRNA levels in human jejunum samples**

Two cohorts of human samples were investigated. A wide range of *LCT* mRNA levels was represented in both cohorts. Cohort 1 n = 56 and cohort 2 n = 59. Data are means ± S.E.M.

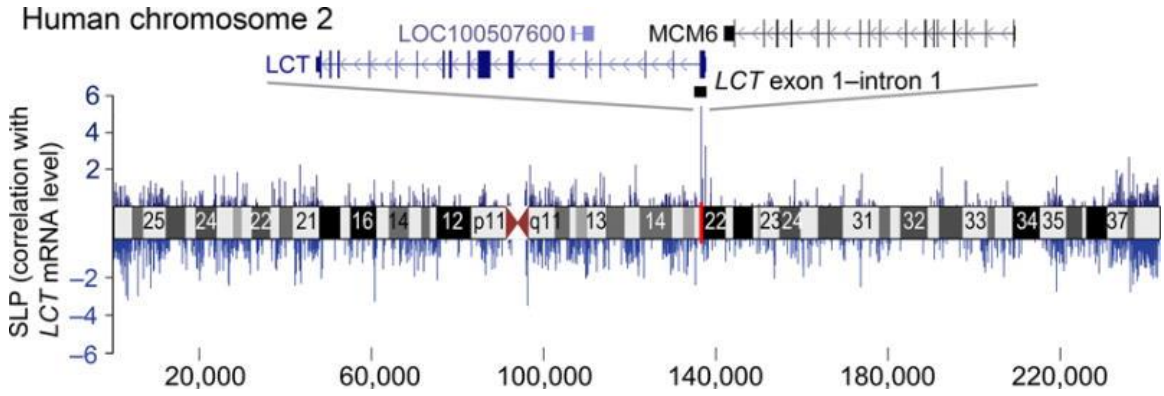



**Figure 4 - 4. Chromosome-wide scan to detect epigenetically-modified regions associated with LCT regulation in humans**

DNA modifications across the entire human chromosome containing the *LCT* gene investigated in jejunal enterocytes by a tiling array approach. Statistical significance of correlations (Y axis) between the density of unmodified cytosines across human chromosome 2 (X axis) and *LCT* steady-state mRNA levels in enterocytes (n = 56 individuals). The single region exhibiting significant correlation after chromosome-wide correction for multiple testing maps to *LCT* exon 1–intron 1. SLP (signed log10(*p*-value)) refers to the significance of correlation, with the sign corresponding to the direction of correlation.

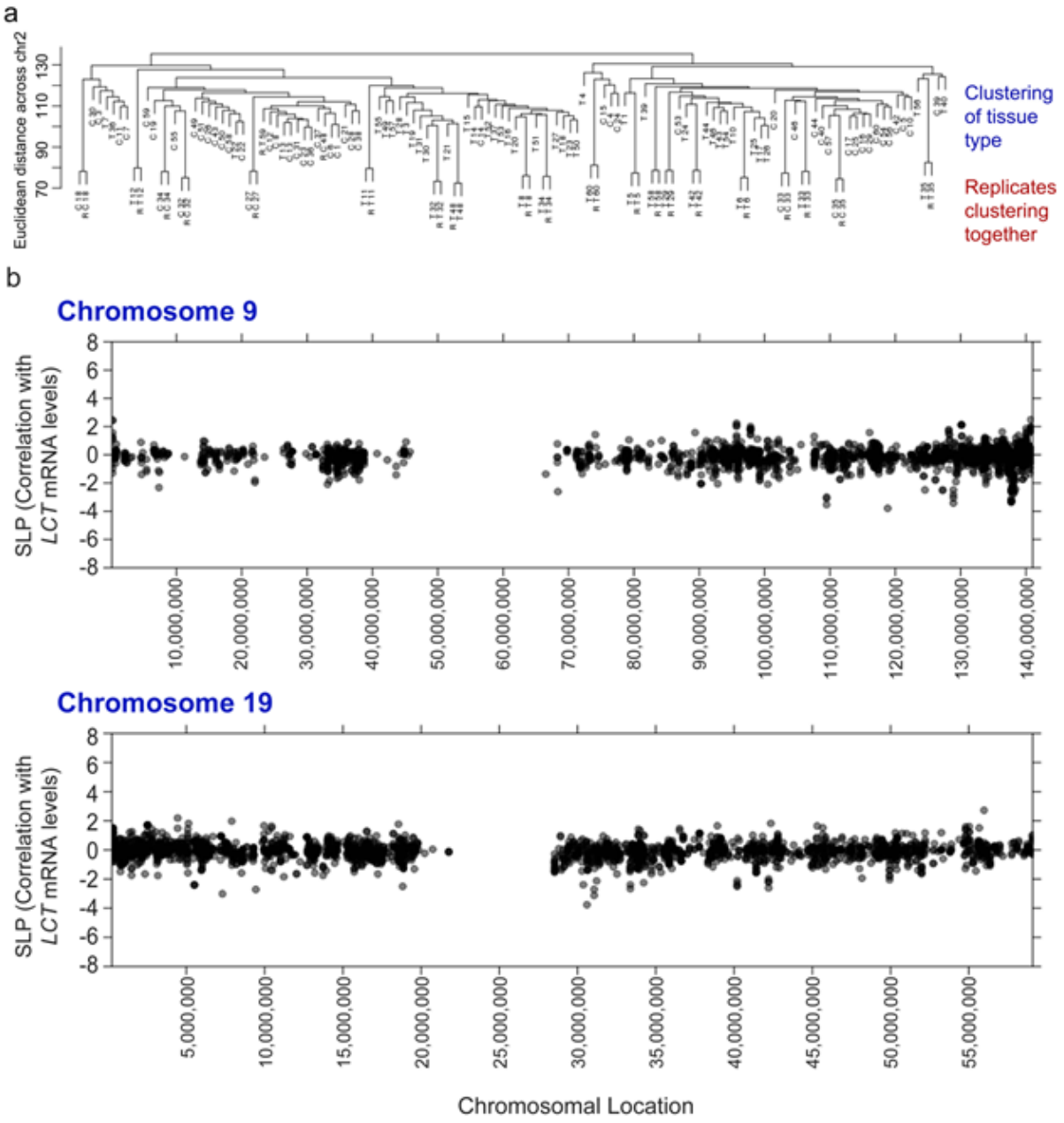



**Figure 4 - 5. Chromosome-wide profiling of DNA modifications associated with LCT mRNA levels in human enterocytes isolated from the jejunum (n = 56) using tiling microarrays**

A) Hierarchical clustering diagram showing strong similarity between replicates and tissue type after bootstrap filtering of probes. DNA modification analysis was completed for enterocyte cells ("C" labels) and the enterocyte-deficient jejunum ("T" labels), along with sample replicates ("R" labels). B) Manhattan plots demonstrating that DNA modification within chromosomes that do not contain the *LCT* gene (chromosome 9 and 19; GRCh37/hg19) is not significantly associated with *LCT* mRNA (all sites FDR $q \geq 0.05$). SLP refers to signed log10(p-value).

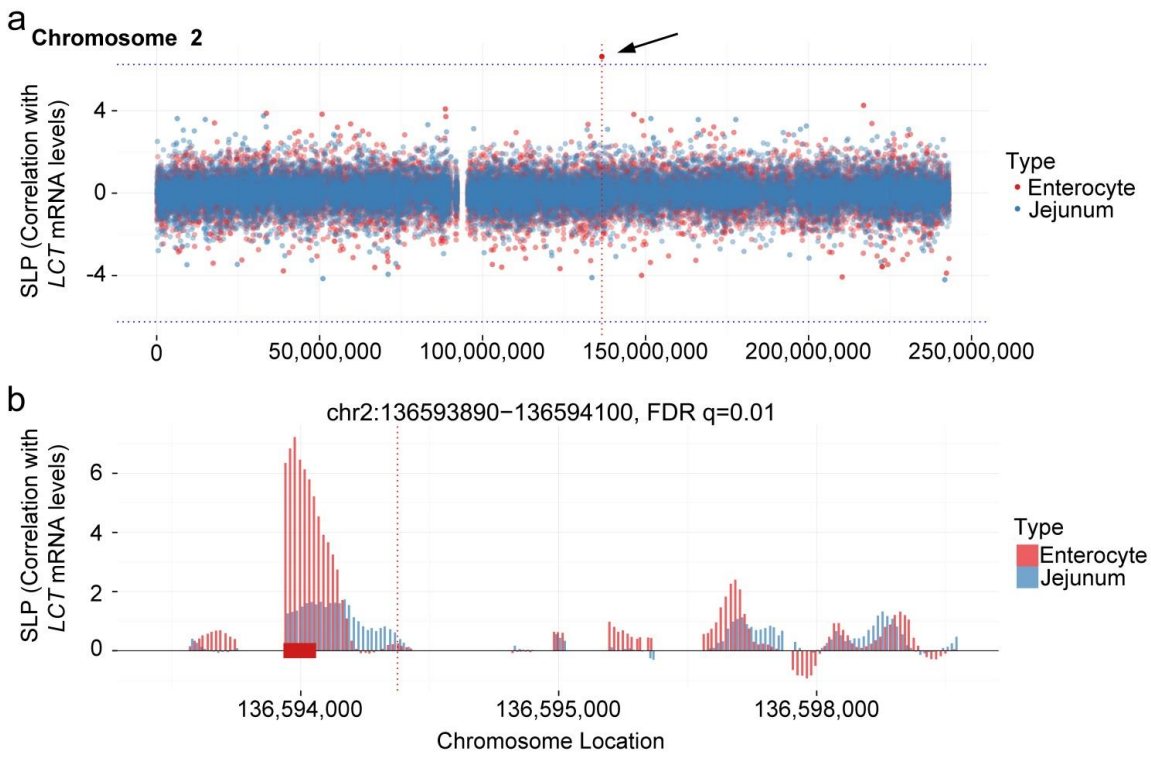



**Figure 4 - 6. Chromosome-wide analysis using human tiling microarrays found that DNA modification densities in enterocytes significantly correlated with LCT mRNA levels at LCT exon 1/intron 1**

The Parsimonious Temporal Aggregation (PTA) algorithm was used to merge similar adjacent probes and assayed resulting regions with the highest (10%) inter-sample variability on chromosome 2 (GRCh37/hg19). Both enterocytes (in red) and the enterocyte-deficient jejunum (in blue) were examined for each individual (n = 56). (a) On chromosome 2 a significant positive association was found between density of unmodified cytosines in *LCT* exon 1/intron 1 and *LCT* mRNA levels in enterocytes (indicated by arrow; R = 0.65, chr2:136,593,890-136,594,100). The horizontal blue line shows the significance threshold. DNA modifications in the enterocyte-deficient jejunum were not significantly associated with *LCT* mRNA levels on chromosome 2. (b) Close-up of the region where lower DNA modification densities were correlated with increased *LCT* mRNA levels in enterocytes (highlighted by horizontal red block). DNA modifications did not significantly correlate with *LCT* mRNA levels in the enterocyte-deficient jejunum. The vertical dashed red line shows the transcription start of *LCT*.

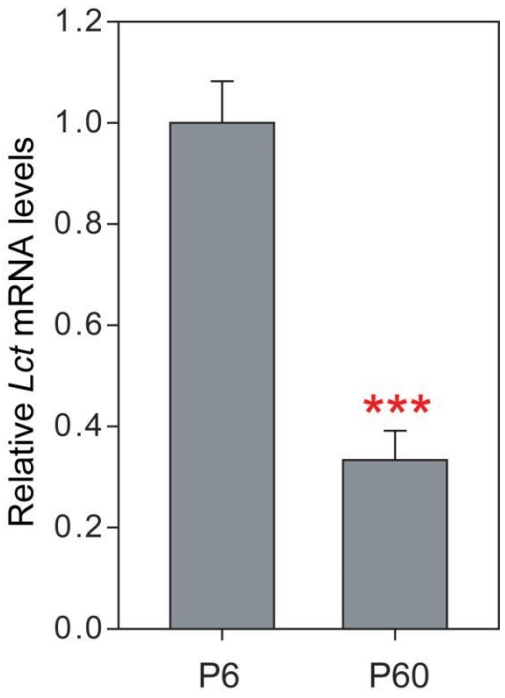



**Figure 4 - 7. Steady-state *Lct* mRNA levels in mouse jejunum samples**

*Lct* mRNA levels were examined in the jejunum of adult (post-natal day 60) and infant (post-natal day 6) mice. *Lct* mRNA levels are greater in infant mice (n = 11) than in adult mice (n = 7). *Lct* mRNA levels were determined by qPCR. ****p* < 0.0001; two-tailed Student's t-test. Data are means ± S.E.M.

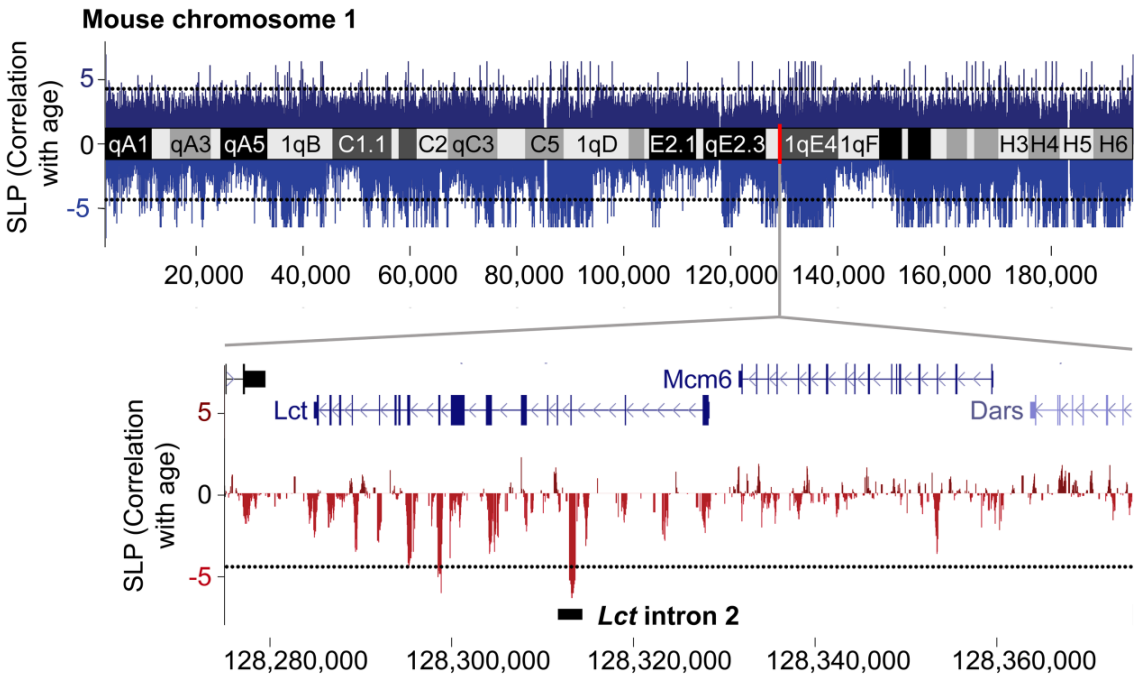



**Figure 4 - 8. Chromosome-wide scan of DNA modifications in aging mouse enterocytes identifies regions associated with the age-dependent downregulation of *Lct***

Numerous DNA modification differences were detected on chromosome 1 of jejunal enterocytes from postnatal day 6 mice (infants, n = 14) compared to day 60 (adults, n = 15) mice. Within the extended *Lct* and *Mcm6* region (100 kb), the largest DNA modification difference between the two mouse age groups were found at *Lct* intron 2. Smaller differences were observed at *Lct* intron 8/exon 9 and exon 10. In these regions, infant mice exhibited significantly higher density of unmodified cytosines compared to the adult mice. *Lct* downregulation with age was therefore correlated with a gain in DNA modifications at these regions. SLP indicates the significance of correlation with age, with negative values signifying higher DNA modification levels in adult mice. The dashed line indicates the significance threshold after Benjamini-Hochberg multiple testing correction.

## 4.4.2 Inter-individual and cell-type specific *LCT* regulatory regions

To complement the broad but relatively low resolution of the microarray-based approach, we fine-mapped the extended human *LCT* region using bisulfite padlock probe sequencing (Diep et al., 2012). For this, we designed 466 padlock probes covering a ~100 kb genomic region encompassing *LCT* (49 kb) and *MCM6* (37 kb). In addition to the first cohort used in the microarray experiment, a second cohort of jejunum samples (n = 59; mean age 42.6, S.D. 10.7 years; 21 males, 38 females) was added to this investigation, for a total of 115 samples (Figure 4-1). We first examined the regions mediating human inter-individual differences in *LCT*. In this

analysis, correlations between DNA modification densities and *LCT* mRNA levels were performed. In enterocytes, significant inverse correlations were detected between DNA modification density and steady-state *LCT* mRNA levels at 11 CpG sites ($p < 0.01–10^{-10}$ after Bonferroni correction). These 11 CpGs clustered into 2 regions: *MCM6* intron 13/exon 13 (~600 bp region size; $R^2 = 0.27–0.53$), which surrounds the lactase persistence/non-persistence SNP C/T-13910, and *MCM6* exon 16 (~30 bp region size; $R^2 = 0.20–0.25$; Figure 4-9 and Appendix A-1). This suggests that epigenetically-controlled regions within *MCM6* may mediate individual differences in *LCT* mRNA.

We then identified regions involved in the cell-type specific expression of *LCT* by comparing the DNA modification profiles of enterocytes with that of the enterocyte-deficient jejunum of the same individuals. Significant DNA modification differences were observed at 35 CpG sites ($p < 0.01–10^{-14}$ after Bonferroni correction), which clustered into seven distinct regions: *LCT* intron 5, intron 3, intron 2, exon 1, as well as *MCM6* exon 17, exon 16 and intron 13 (~25- 560 bp region size; 1.35- to 2.51-fold change between jejunum and enterocytes; Figure 4-9 and Appendix A-1). The *MCM6* intron 13 and exon 16 regions each overlapped the regions identified in the inter-individual analysis above. The *LCT* exon 1 and intron 2 regions we identified by targeted bisulfite sequencing were close to the respective regions detected in the tiling microarrays (Figure 4-10). Notably, the region in *LCT* intron 5 mapped to the promoter of a lncRNA, *LOC100507600*, while the exon 17 region mapped to 3' UTR of *MCM6.*

We also investigated whether the association of DNA modifications with *LCT* was unique to enterocytes or could be extended to tissue that does not express *LCT* mRNA. In white blood cells (n = 58), the seven *LCT*/*MCM6* regions were nearly fully modified and therefore vastly differed from those of intestinal tissues (averaged % DNA modification in blood, enterocyte-deficient jejunum and enterocytes: 73.3% ± 0.4%, 54.2% ± 0.8% and 37.1% ± 0.9%, respectively; Figure 4-11). This provides additional support that these regions facilitate the cell-type specificity of *LCT* modification.

To further explore the regulatory functions of the differentially modified *LCT* regions, we investigated chromatin reference maps of the human small intestine generated by the Epigenomics Roadmap (Bernstein et al., 2010; Consortium, 2012; ENCODE Project Consortium, 2012; A. Kundaje et al., 2015) (Figure 4-9). Our discovered regions in *LCT* exon 1,

introns 2 and 3, as well as *MCM6* intron 13, exon 16 and 3' UTR overlapped with DNase I hypersensitivity sites. Moreover, histone marks characteristic of active enhancers, H3K4me1 and H3K27ac, were detected at *LCT* introns 2 and 3, as well as *MCM6* intron 13 and exon 16 to 3'-UTR. No other enhancer sites were observed elsewhere in a 100 kb region centered over *LCT* and *MCM6*. Meanwhile, *LCT* exon 1 had an abundance of H3K4me3 and H3K27ac, indicative of an active promoter. Hence, functional chromatin signatures corroborate well with the sites we identified in our DNA modification study, further supporting the role of epigenetic mechanisms in the regulation of *LCT*.

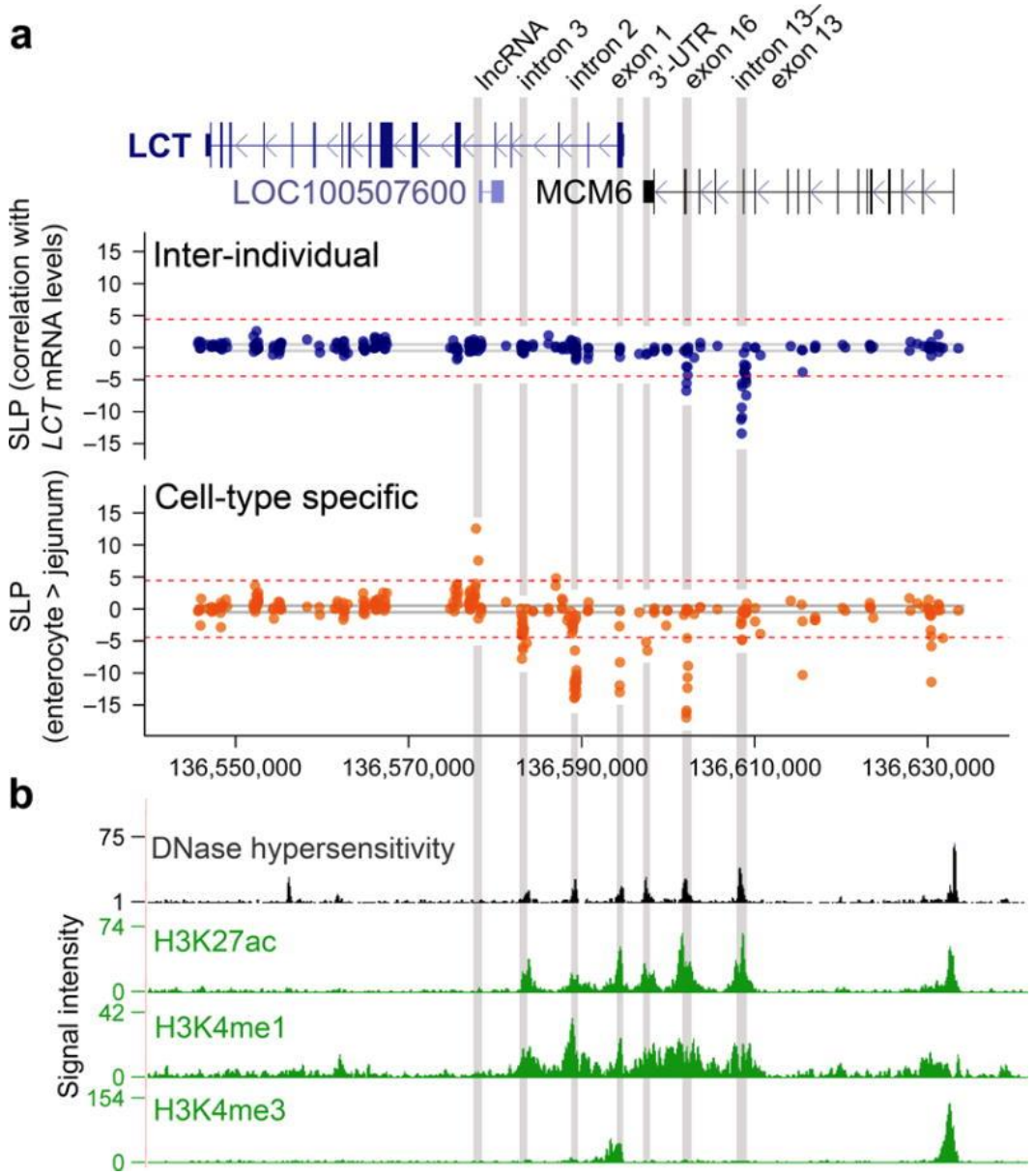
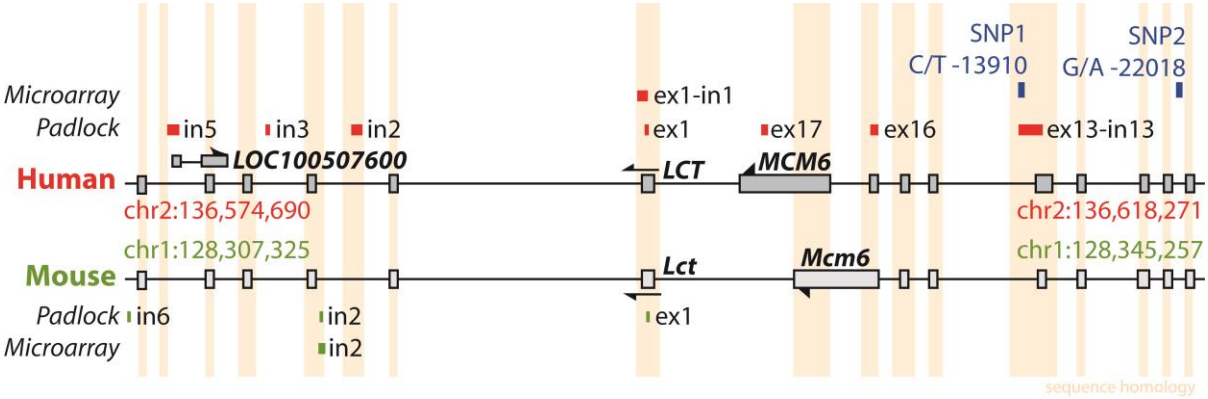



**Figure 4 - 9. Fine-mapping of DNA modifications to identify regions associated with inter-individual and cell-type specific LCT mRNA levels in human adults**

The bisulfite padlock probe approach profiled DNA modifications in human intestinal samples (n = 115 individuals). (**a, upper panel**) In enterocytes, DNA modifications significantly correlated with individual *LCT* mRNA levels at two regions in *MCM6*: exon 16 and intron 13–exon 13. SLP refers to the log10(*p*-value), with the sign corresponding to the direction of correlation. The dashed line indicates threshold significance after Bonferroni correction for multiple testing ($p < 0.01$). (**a, lower panel**) Cell-type specificity of *LCT* modification determined by comparison of enterocytes to the remainder of the jejunum. Seven regions within *LCT* and *MCM6* were associated with cell-type specificity of *LCT*. (**b**) Regions exhibiting inter-individual and cell-type specific *LCT–MCM6* modification differences (grey vertical bars) overlap with DNase I hypersensitivity sites and are located within enhancers (H3K4me1 and H3K27ac at *LCT* intron 3, intron 2 and *MCM6* 3′-UTR, exon 16 and intron 13–exon 13) and a promoter (H3K4me3 and

H3K27ac at *LCT* exon 1). The region at *LCT* intron 5 is located at the promoter of the lncRNA *LOC100507600*. Chromatin modification profiles from the fetal small intestine (n = 1 individual) mapped by the Epigenomics Roadmap.




**Alignment of *MCM6/Mcm6* intron 13 region in human and mouse**

```
homo_sapiens      ...............................................................................
mus_musculus      TGGTAGTCTGTGGCAGAGGGCTGGGAATTTAAGCCAGGTTACCAAAAGAGCACCTGTTTCAGAAACACAACAAACAAATGAACCTA
homo_sapiens      ...............................................................................
mus_musculus      GTCCTTTCACAGACAAAATTTGAATCCTAAATGAACTCTGTATATAAAAGTACACAAAAGCCTCTTAGTTTGAAGCAGATATGTGG
homo_sapiens      ..............................................................
mus_musculus      CTTCCAATATCACCCCTTCCCATAATCCGACATTTGCTTTGAGTCTAGGGCACCAAGAGCAATTAAAAATCTATGGAGCTAAGGTA
homo_sapiens      ...............................................................................
mus_musculus      AGAGTCATGAGTGTGCTTAGCATACACAAAGCCCTGAATTTTGCCCCAGCATATGCAAACCCAAAATAATGCCCACACCCGCCACA
homo_sapiens      .....ATTATACAAATGCAACCTAAGGAGGAGAGTTCCTTTGAGGCCAGGGGCTACATTATCTTATCTGTATTGCCAGCGCAGAGG
mus_musculus      CACTGATTATACAAC-----CCTAAGGAGGAAGAGTCCTCTGAGGC-AGGAGTCACA--GTCACGTTTGTATTCCCAGC-CAGAGG
                       *********      ***********    **** ****** *** *  ***   **   * ****** ***** ******
homo_sapiens      CCTACTAGTACATTGTAGGGTCTAAGTACATTTTTCCTGAATGAAAGGTATTAAATGGTAACTTACGTCTTTATGCACTCTATAAA
mus_musculus      CCAA---GGCCACTGTAGAGATTAAATAGGTTTCTCTTGCGTAAGT--------------ACTCACCTGTATCTGCACTCCATAAT
                  ** *   *  ** ***** *  *** **  *** ** **  * *               *** ** * * * ******* ****
homo_sapiens      CTAT.........................................
mus_musculus        TTACTGATTATAATAAACACTTTTATCCATCACTATTAGTCTAGACAGC                  **
```

**Figure 4 - 10. Regions exhibiting DNA modification densities that are significantly associated with lactase mRNA levels in humans and mice**

(Upper panel) Schematic drawing of the regions containing DNA modifications associated with *LCT* mRNA regulation, as identified in a chromosome-wide scan and fine mapping investigation. DNA modifications were investigated in jejunal enterocytes from human samples (cohorts 1 and 2) as well as infant and adult mice C57BL/6N inbred mice (post-natal day 6 vs 60). Microarrays were used in the chromosome-wide analysis and the bisulfite padlock probe technique was for the fine-mapping investigation. Epigenetically-controlled regions associated with lactase mRNA levels are indicated for humans (red) and mice (green). Peach bars indicate regions with high sequence homology between humans and mice. (Lower panel) Sequence alignment of *MCM6/Mcm6* intron 13 region in human (GRCh37/hg19; chr2: 136608600-136608770) and mouse (C57BL/6 strain, GRCm38/mm10; chr1:128337434-128337972). This region corresponds to the region deleted in the mouse CRISPR-Cas9 experiment. *MCM6* SNP C/T-13910 site is highlighted in yellow and * indicates DNA sequence homology between human and mouse.

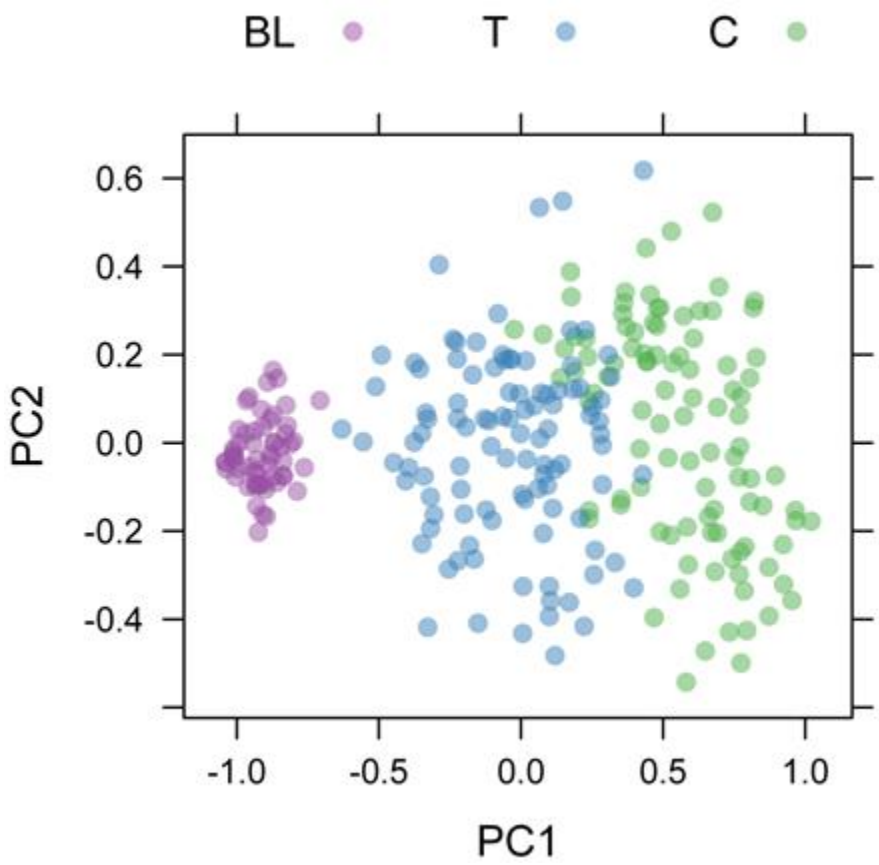



**Figure 4 - 11. Comparison of DNA modification densities at lactase gene regulatory regions in intestinal and non-intestinal tissue**

DNA modification profiles in enterocytes (C) differ from those of the enterocyte-deficient jejunum (T), and even more so from those of blood (BL). Principal component analysis was performed on the DNA modification levels for 262 samples across the seven key *LCT* regulatory regions identified in this study; each sample was considered independently from its donor. This study found 42 CpG sites associated with *LCT* regulation, which clustered into the following seven regions: *LCT* intron 5, intron 3, intron 2, and exon 1 as well as *MCM6* 3'-UTR, exon 16

and intron 13/exon 13. There were 31 samples without a measurement for 1 region (2 blood, 13 enterocyte, 19 enterocyte-deficient jejunum), and 3 samples without a measurement 2 regions (3 enterocyte-deficient jejunum); missing values were imputed using k-nearest-neighbors (k=10) using the "impute.knn" function within the R "impute" package. The first two principal components are plotted.

### 4.4.3 Genotype-dependent and temporal changes in *LCT* modification in humans

Since *LCT/MCM6* haplotypes exhibit a strong association with lactase non-persistence/persistence in adult Europeans (Bersaglieri et al., 2004; Gallego Romero et al., 2012; Poulter et al., 2003), we hypothesized that differential epigenetic modifications could occur in such haplotypes, contributing to the age-dependent regulation of *LCT*. To test this, we examined whether different genetic backgrounds at *LCT/MCM6* led to a divergence of DNA modifications in an age-dependent fashion. We first genotyped two *MCM6* SNPs (C/T-13910, rs4988235 and G/A-22018, rs182549) and an insertion-deletion in the 3' region of *LCT* intron 1 in the intestine samples (n = 115). All three genotypes showed association with steady-state *LCT* mRNA levels (Figure 4-12) and exhibited strong pairwise linkage disequilibrium, consistent with previous reports (Bersaglieri et al., 2004; Poulter et al., 2003) (D' > 0.9 and Figure 4-13). Further analysis focused on the SNP C/T-13910, for which C and T alleles are parts of the haplotypes associated with lactase non-persistence and lactase persistence, respectively (Bersaglieri et al., 2004; Enattah et al., 2002; Poulter et al., 2003).

We next investigated DNA modification changes at the seven regulatory regions (described in the above sections) in enterocyte samples stratified for genetic background. We observed that CC carriers (lactase non-persistent) displayed major DNA modification differences in comparison to TT carriers (lactase persistent) in the *MCM6* locus, particularly at the *MCM6* intron 13/exon 13 (82.1 ± 2.6% *vs*. 19.3 ± 4.1%; $p < 10^{-17}$) where the SNP C/T-13910 resides (Figure 4-14 and 4-15).

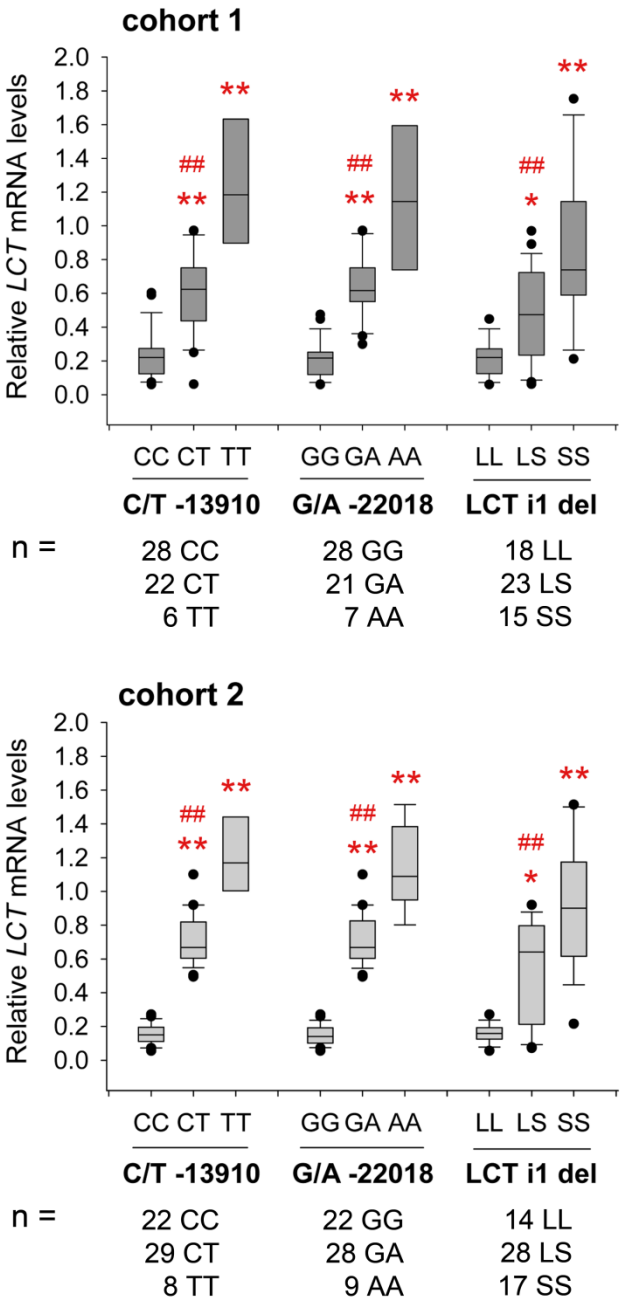



**Figure 4 - 12. *LCT* mRNA levels in the human jejunum relative to SNP genotypes for cohort 1 and 2.**

Individuals were genotyped for the SNP in *MCM6* intron 13 C/T-13910, the SNP in *MCM6* intron 9 G/A-22018 and the deletion (no deletion = L, deletion = S) in *LCT* intron 1. Total sample size was n = 56 individuals in cohort 1 and n =59 in cohort 2. The haplotype containing T-13910, A-22018, and deletion in *LCT* intron 1 was associated with higher *LCT* mRNA levels in both cohorts. Analysis of variance (ANOVA) found a SNP C/T-13910 haplotype effect in cohort 1 and 2 ($F_{2,53} = 58.5$, $p < 10^{-13}$ and $F_{2,56} = 206.2$, $p < 10^{-25}$, respectively). *$p < 0.05$ and

**$p < 10^{-3}$ compared to CC/GG/LL genotype; ##$p < 10^{-3}$ compared to TT/AA/SS genotype; Tukey's honest significant difference (HSD) post-hoc test.

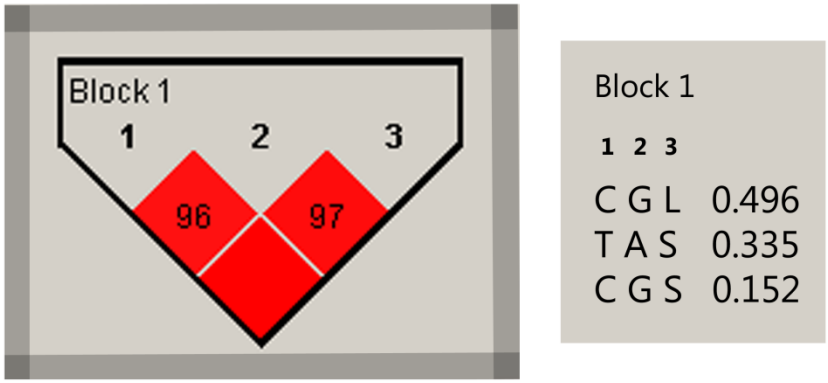



**Figure 4 - 13. Linkage disequilibrium in *LCT* and *MCM6* region in individuals with European ancestry**

For the intestinal samples used in this study (n =115 European individuals, cohort 1 and 2), three genetic variants were examined: *MCM6* intron 13 C/T-13910 (labelled 1), the SNP in *MCM6* intron 9 G/A-22018 (labelled 2) and the deletion in *LCT* intron 1 (labelled 3). (Left panel) Haplotype analysis revealed strong linkage disequilibrium (D' > 0.96) between these SNPs. D' values are shown in the boxes. (Right panel) Haplotypes and their population frequencies are shown. Haplotypes containing the -13910*C are lactase non-persistent and haplotype with -13910*T are lactase persistent. Haplotype analysis was performed using the Haploview 4.2 program.

On the basis of our data, our inference is that *MCM6* intron 13 completes most of its age-dependent epigenetic changes during the first two decades after birth (prior to the age of the youngest individuals in our sample set ranging from 21-72 years old). We posit that DNA modification changes at *MCM6* intron 13 for the haplotypes containing C-13910 likely occur after conception, as our investigation of sperm samples (n = 18) revealed no DNA modification differences between CC, CT and TT genotypes at this region (>94% modified in all genotypes; data not shown). The situation with *MCM6* exon 16 is likely similar, as it also shows significant differences between C-13910 and T-13910 carrying haplotypes (52.7 ± 3.1% *vs*. 32.8 ± 4.1%; $p < 0.01$), although it appears to amass more gradual DNA modification changes compared to *MCM6* intron 13.

Molecular events associated with lactase non-persistence or persistence do not stop at childhood or youth (Rasinpera, Kuokkanen, et al., 2005; Sahi et al., 1983). In our cohort of 21-72 year old individuals, we detected that *LCT* mRNA was significant negatively associated with age in the lactase non-persistent (CC) individuals ($R^2 = 0.2$, $p < 0.01$), but showed no significant association with age in lactase persistent (TT) individuals ($R^2 = 0.13$, $p = 0.2$). We then investigated age-related DNA modification changes within each of the seven regulatory regions in the enterocyte samples stratified for C/T-13910 genotypes. In lactase non-persistent individuals (CC), DNA modifications consistently increased with age across *LCT* and *MCM6* (Figure 4-16). Conversely, lactase persistent (TT) individuals showed mostly decreasing DNA modification densities in *MCM6* and *LCT* with age (Figure 4-16) (except at the lncRNA promoter, where greater DNA modification was associated with higher *LCT* mRNA). DNA modification changes observed in adulthood could be an extension of the *LCT/MCM6* epigenetic processes occurring in childhood. Overall, our findings suggest that lactase non-persistence may be mediated by an age-related increase in DNA modifications at regulatory elements in *MCM6* and *LCT*, while such elements are protected from epigenetic inactivation in lactase persistent individuals.

The age-related DNA modification changes we observed at *MCM6* were consistent with the trajectories of histone modifications across ages (Figure 4-17). Using Epigenomics Roadmap data, we found that in the fetal small intestine, *MCM6* exon 16 and intron 13 are marked by H3K4me1 and H3K27ac, which is indicative of active enhancers. However, at 3 years old, both modifications are greatly reduced at *MCM6* exon 16 and intron 13, indicating that these regions have lost their enhancer activity, an effect that persists into adulthood (no significant H3K27ac and H3K4me1 peaks in adults). The other *LCT/MCM6* enhancers (intron 2, intron 3 and 3'-UTR) and the *LCT* promoter (exon 1), which are all active in fetal life, are poised/inactive in 3 year old children and adults (lack of H3K27ac, but presence of H3K4me1 in enhancers and H3K4me3 in promoter). These findings suggest that inactivation of key regulatory elements in *LCT/MCM6* is involved in the age-related downregulation of *LCT* in humans.

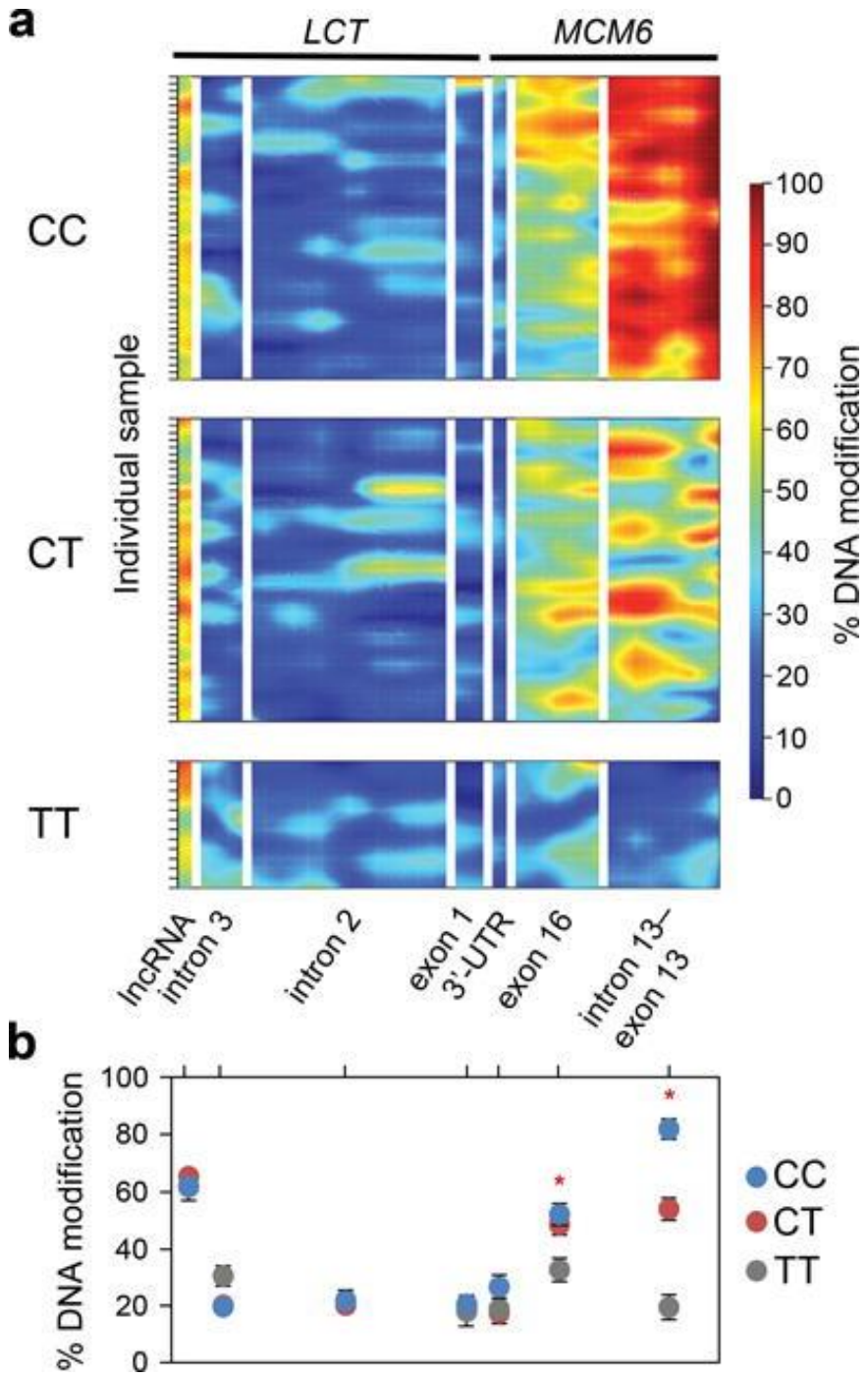



**Figure 4 - 14. The effect of genetic variation relevant to lactase non-persistence and persistence on DNA modification landscapes**

(**a**) Heatmaps show variation of DNA modifications in enterocytes at the seven regulatory sites associated with *LCT* regulation in the population stratified for SNP C/T-13910 genotype. *MCM6* contained the most dramatic modification differences between CC, CT and TT individuals (n = 43, 43, 14 individuals, respectively). Color scale indicates DNA modification densities; from absence (dark blue) to fully modified (dark red). Each bar on Y-axis represents a separate sample. (**b**) Average DNA modification density in human enterocyte samples at each of

the seven regulatory sites. CC individuals (lactase non-persistent) exhibited a 4-fold higher density of modified cytosines at *MCM6* intron 13–exon 13 compared to the TT individuals (lactase persistent) (82% in CC *vs*. 19% in TT; $F_{2,97} = 61.8$, $p < 10^{-17}$; one-way ANOVA). DNA modification density also significantly diverged between genotypes at *MCM6* exon 16 (53% in CC *vs.* 33% in TT; $F_{2,97} = 4.9$, $p < 0.01$; one-way ANOVA). Data are means ± S.E.M. (n = 43 CC, 43 CT and 14 TT individuals). * $p < 0.01$ comparison between homozygote carriers; Tukey's honest significant difference (HSD) post-hoc test.

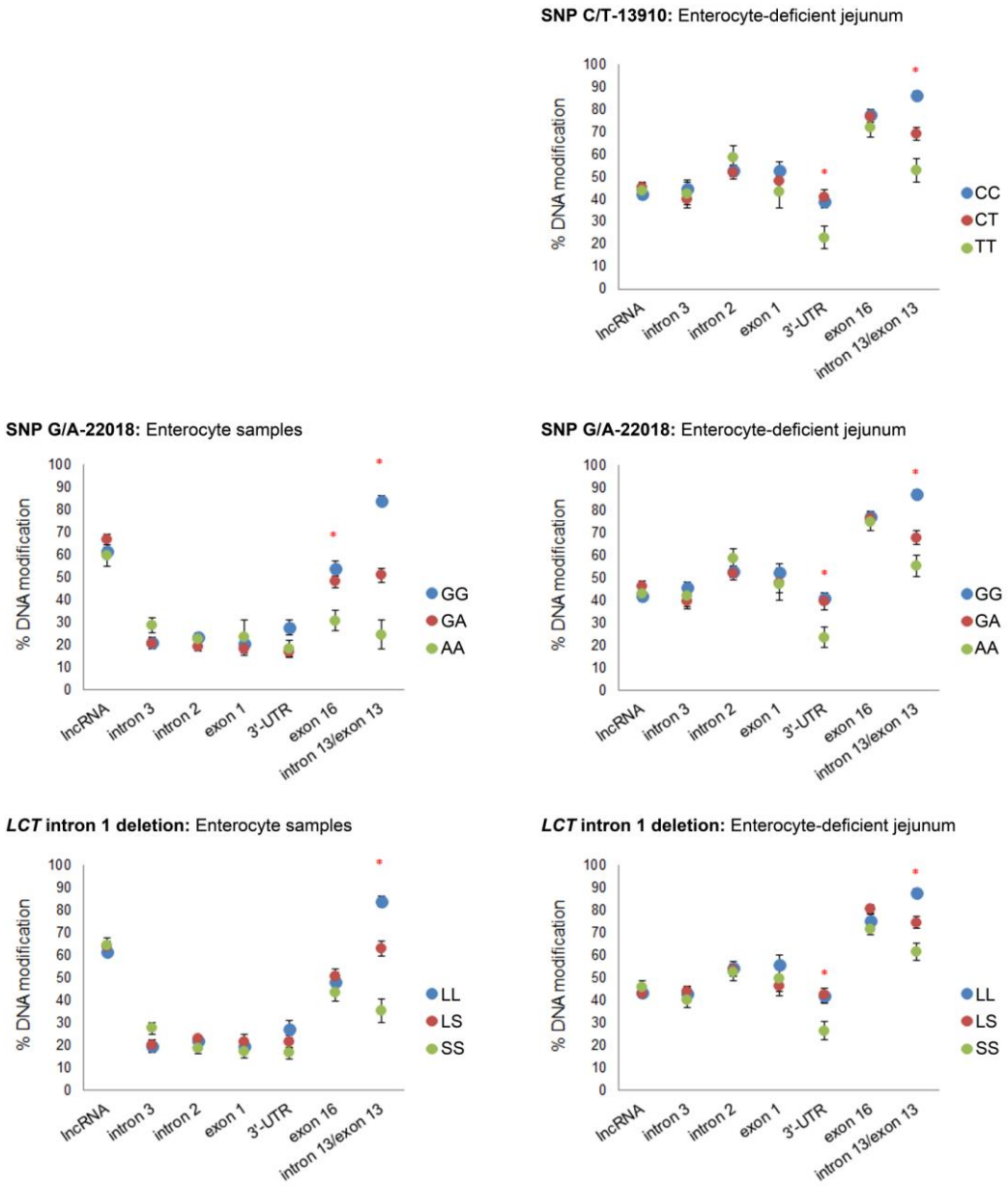



**Figure 4 - 15. DNA modification densities for LCT/MCM6 SNP genotypes**

DNA modifications in enterocyte samples and enterocyte-deficient jejunal samples were examined for each genotype of SNP C/T-13910, SNP G/A-22018 and the deletion in the 3'

region of *LCT* intron 1. The seven loci that demonstrated a significant correlation between DNA modifications and *LCT* mRNA levels were examined. In enterocytes, average DNA modification density significantly diverges between genotypes at *MCM6* exon 13/intron 13 ($p < 10^{-10}$) and at *MCM6* exon 16 (for G/A-22018 $p < 0.01$). In enterocyte-deficient jejunum, smaller DNA modification density differences are observed between genotypes at *MCM6* intron 13/exon 13 ($p < 10^{-5}$) as well as the 3'UTR region of *MCM6* ($p < 0.05$). *p*-values are one-way ANOVA. Sample sizes of SNP C/T-13910 were CC n = 43, CT n = 43 and TT n = 14; SNP G/A-22018 were GG n = 43, GA n = 42 and AA n = 15; *LCT* intron 1 deletion were LL n = 27, LS n = 46 and SS n = 27. Data are means ± S.E.M. DNA modification densities in enterocytes of SNP C/T-13910 for each of the seven regions is displayed in Figure 3 of main text.

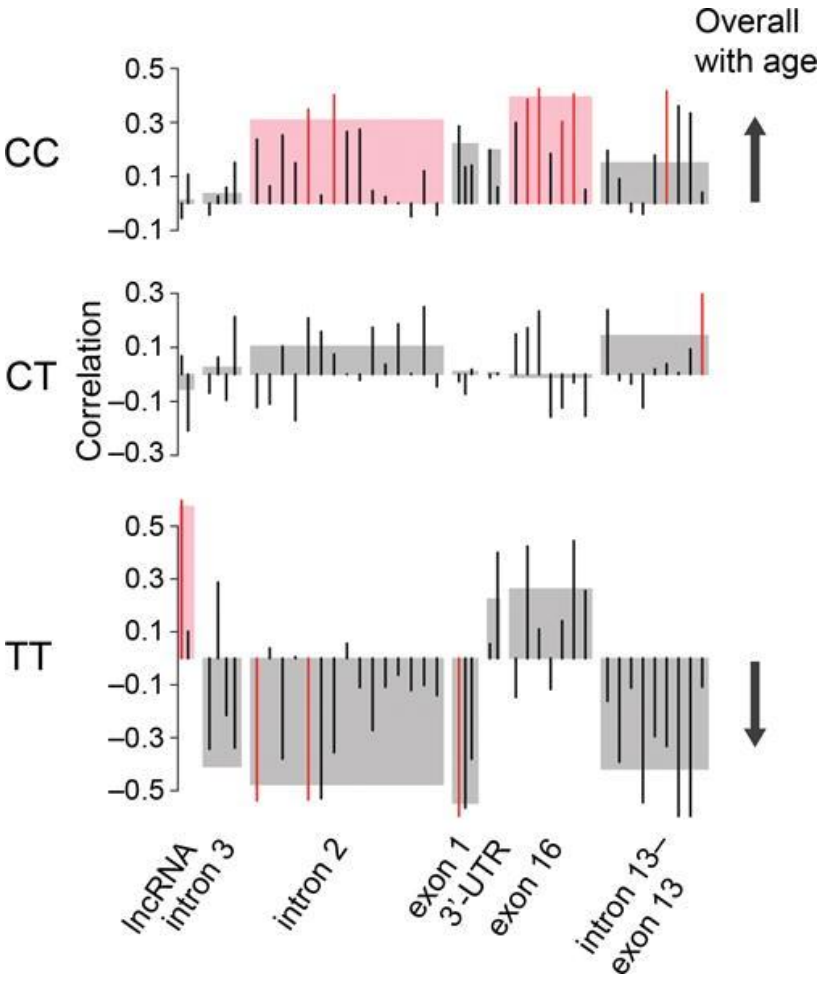



**Figure 4 - 16. Aging dynamics of DNA modification at *LCT–MCM6* regulatory sites between SNP -13910 CC, CT and TT carriers**

DNA modification changes occurring with age examined in enterocytes of CC, CT and TT carriers (n = 43, 43 and 14 individuals, respectively). To identify region-wide changes in response to age, average DNA modification densities at each of the seven epigenetically-controlled regions associated with *LCT* regulation were correlated with age. Pink and grey boxes

indicate significant ($p < 0.05$) and non-significant, respectively, region-wide correlations. To identify effects at individual cytosine sites, single CpGs within each of the seven regions were correlated with age. Red and black bars indicate significant ($p < 0.05$) and non-significant, respectively, single CpG correlations. In CC individuals, DNA modifications increased with age; an effect most apparent at *LCT* intron 2 (adjusted $R^2 = 0.07$) and *MCM6* exon 16 (adjusted $R^2 = 0.13$). TT individuals demonstrated DNA modifications that primarily decreased with age, except at the lncRNA promoter region which had an age-dependent increase in DNA modifications (adjusted $R^2 = 0.28$). Correlations between DNA modification and age, including those that did not reach statistical significance, were predominantly positive in CC individuals and predominantly negative in TT individuals (as indicated by arrow).

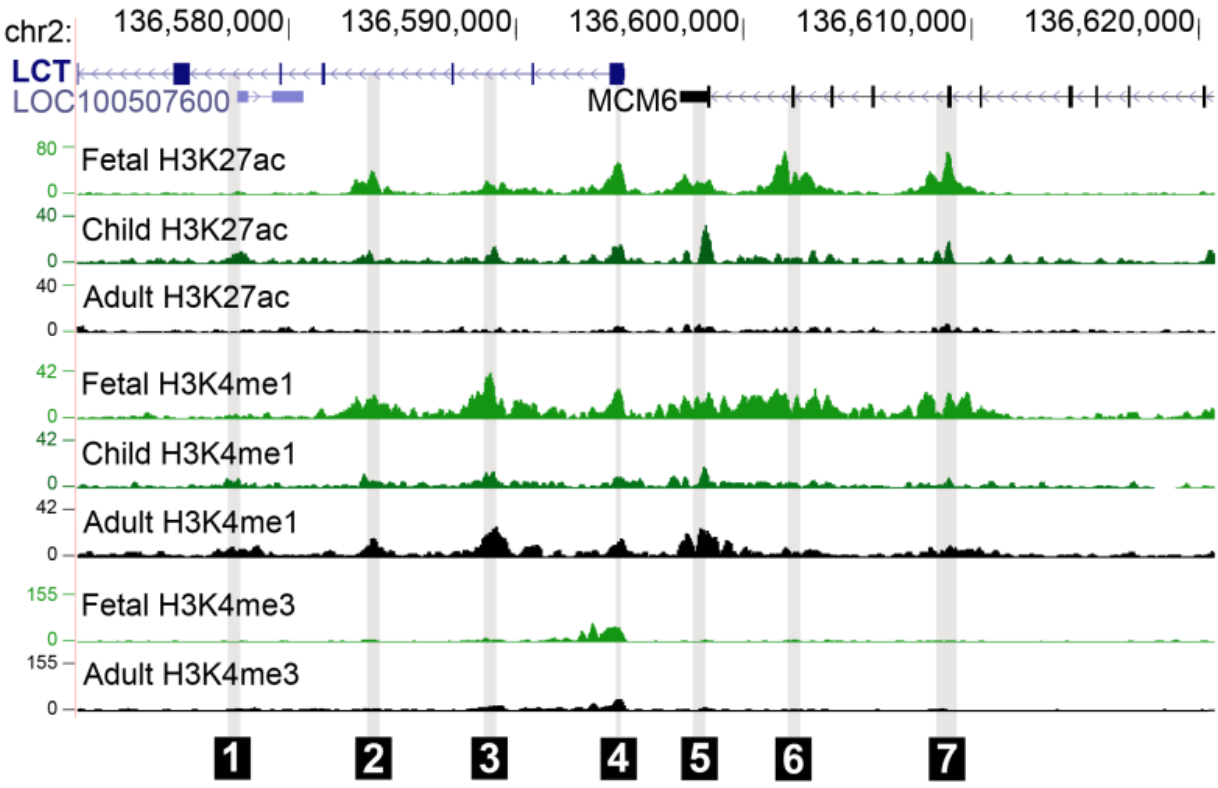



**Figure 4 - 17. Histone modification profiles change with age, especially at *MCM6* exon 16 and intron 13/exon 13**

Seven regions (denoted by black squares) within the *LCT* and *MCM6* were found to contain DNA modification densities significantly associated with steady-state *LCT* mRNA. These were: *LCT* intron 5 (lncRNA, label 1), intron 3 (label 2), intron 2 (label 3) and exon 1 (label 4), and *MCM6* exon 17 (3'-UTR, label 5), exon 16 (label 6) and intron 13/exon 13 (label 7). H3K27ac, H3K4me1 and H3K4me3 marks were examined in the *LCT* and *MCM6* of fetal (prenatal day 108), child (3 years old) and adult (59 years old) small intestine. At the fetal age, histone marks indicative of an enhancer (H3K4me1 and H3K27ac) are present at the *LCT* intron 2 and intron 3 as well as the *MCM6* intron 13/exon 13, exon 16 and 3'-UTR region. The *LCT* exon 1 region

contains histone marks that signal a promoter (H3K4me3, H3K27ac). *MCM6* exon 16 (label 6) and *MCM6* intron 13/exon 13 (label 7) show prominent reductions in enhancer marks in childhood, as demonstrate by a loss of both H3K4me1 and H3K27ac, and this effect persists into adulthood. The other *LCT* enhancers and the promoter at *LCT* exon 1 are present but in a poised/inactive state in adults, as indicated by the loss of H3K27ac and continued presence of H3K4me marks at these sites. Roadmap Epigenomics Project small intestine sample codes were H-24595 (fetal: 108 days old, male, n = 1), STL001 (child: 3 years old, male, n = 1) and 159 (adult: 59 years old, male, n = 1).

### 4.4.4 Evolutionary conservation of *LCT* regulation

We next explored the extent of evolutionary conservation of epigenetic factors contributing to the age-dependent downregulation of the lactase gene in mammals. For this, we used 404 padlock probes and bisulfite sequenced a ~80 kb region containing *Lct* and *Mcm6* in enterocytes of infant (6 days old) and adult (60 days old) mice. DNA modifications at *Lct* intron 2 and 8 (similar regions to those identified in the tiling arrays) were found to significantly increase with age in mice (FDR $q < 0.05$; fold changes: 1.60 and 1.25, respectively; Appendix A-1, Figure 4-10 and 4-18). In support, investigation of chromatin marks using ENCODE data of an adult (60 days old) C57BL/6N mouse identified an enhancer element (H3K27ac and H3K4me1 peaks) in *Lct* intron 2, along with an active promoter at *Lct* exon 1/intron 1 (H3K27ac, H3K4me3 and POLR2A peaks) and an inactive enhancer at *Mcm6* intron 13 (only H3K4me1; Figure 4-19).

Furthermore, the *LCT* intron 2 region is highly conserved across mammals, and contains 7 out of 47 mammalian intronic conserved elements in *LCT* and *MCM6* (139 conserved bases; permutation $p < 10^{-7}$; Figure 4-20 and 4-21). This large degree of mammalian sequence conservation was not found in any other intron of *LCT* or *MCM6*, with no significant conservation observed at the *MCM6* intron 13 region. Thus, converging evidence indicates that *LCT* intron 2 is an evolutionarily old regulatory element, common in mice and humans (and potentially across all mammals) for the development of lactase non-persistence. By contrast, the regulatory regions in *MCM6* appear to have emerged more recently, perhaps in response to evolution selection toward lactase persistence in human adults.

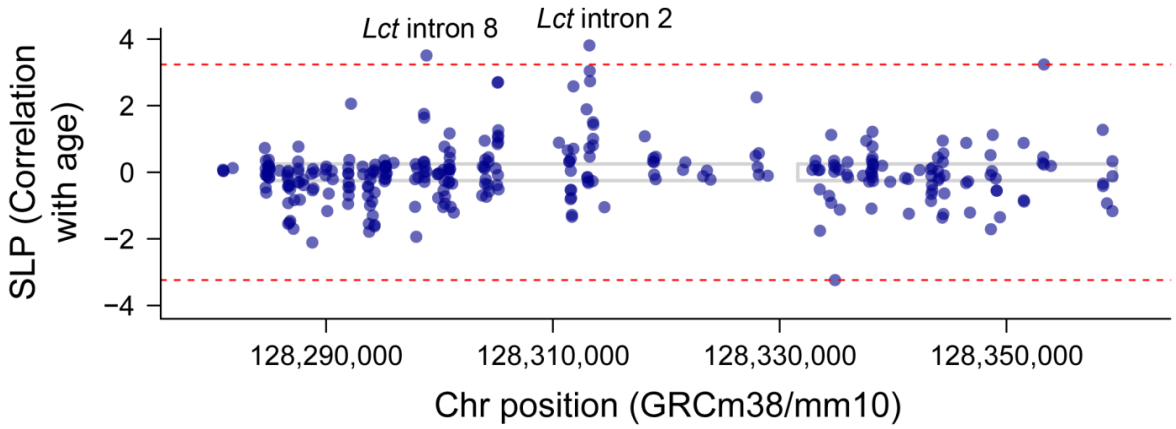



**Figure 4 - 18. Epigenetic modifications affecting lactase gene regulation in mice**

Mapping of DNA modifications in mice using the bisulfite padlock probe sequencing approach revealed regions involved in the age-dependent decline in *Lct*. Enterocytes of infant (post-natal day 6, n = 16) and adult mice (day 60, n = 13) were compared. DNA modification densities at *Lct* intron 2 and intron 8 significantly increased with age (FDR $q < 0.05$). Borderline significant associations between *Lct* mRNA levels and DNA modification density were also observed at *Mcm6* exon 15 and intron 3. Dashed line indicates the significance threshold.

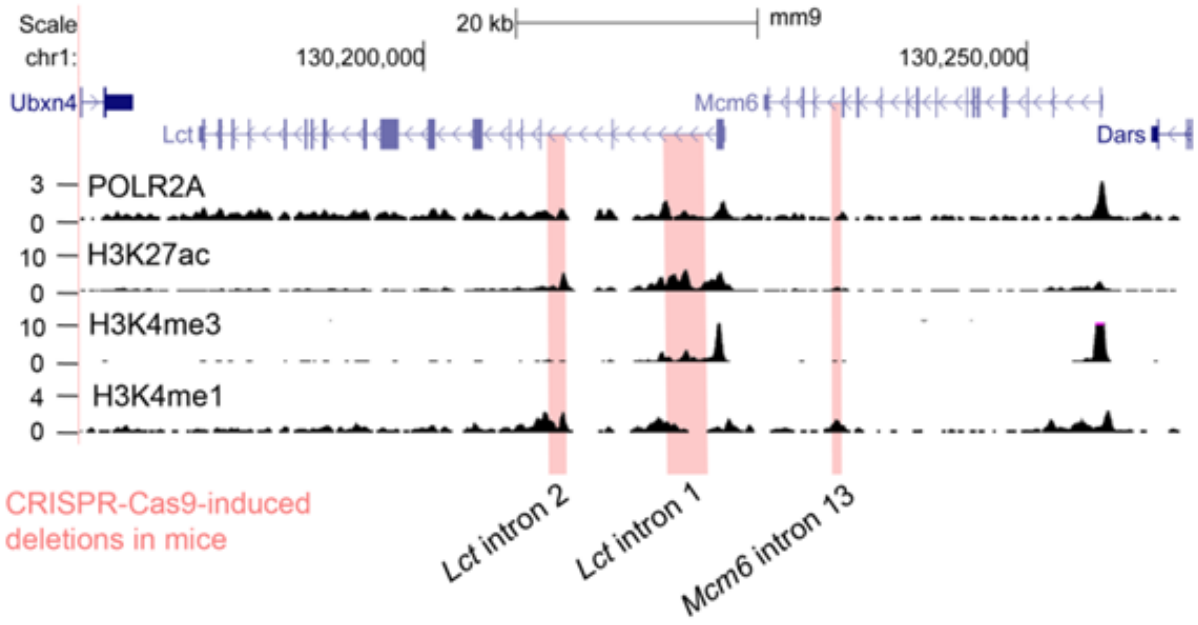



**Figure 4 - 19. Chromatin modification profiles in the small intestine of an adult (60 days old) C57BL/6N mouse**

Examination of ENCODE data revealed that *Lct* intron 2 contained an enhancer element (H3K4me1 and H3K27ac). There was also a promoter element (H3K4me3, H3K27ac, POLR2A)

in *Lct* exon 1/intron 1 and an inactive enhancer in *Mcm6* intron 13 (H3K4me1 only). The ENCODE sample code for the adult mouse small intestine (n = 1) was ENCBS157ENC. To investigate the capacity for epigenetically-controlled regulatory regions to alter *Lct* mRNA levels, we used the CRISPR-Cas9 genome editing technique to induce deletions in mice (deleted regions highlighted in pink).

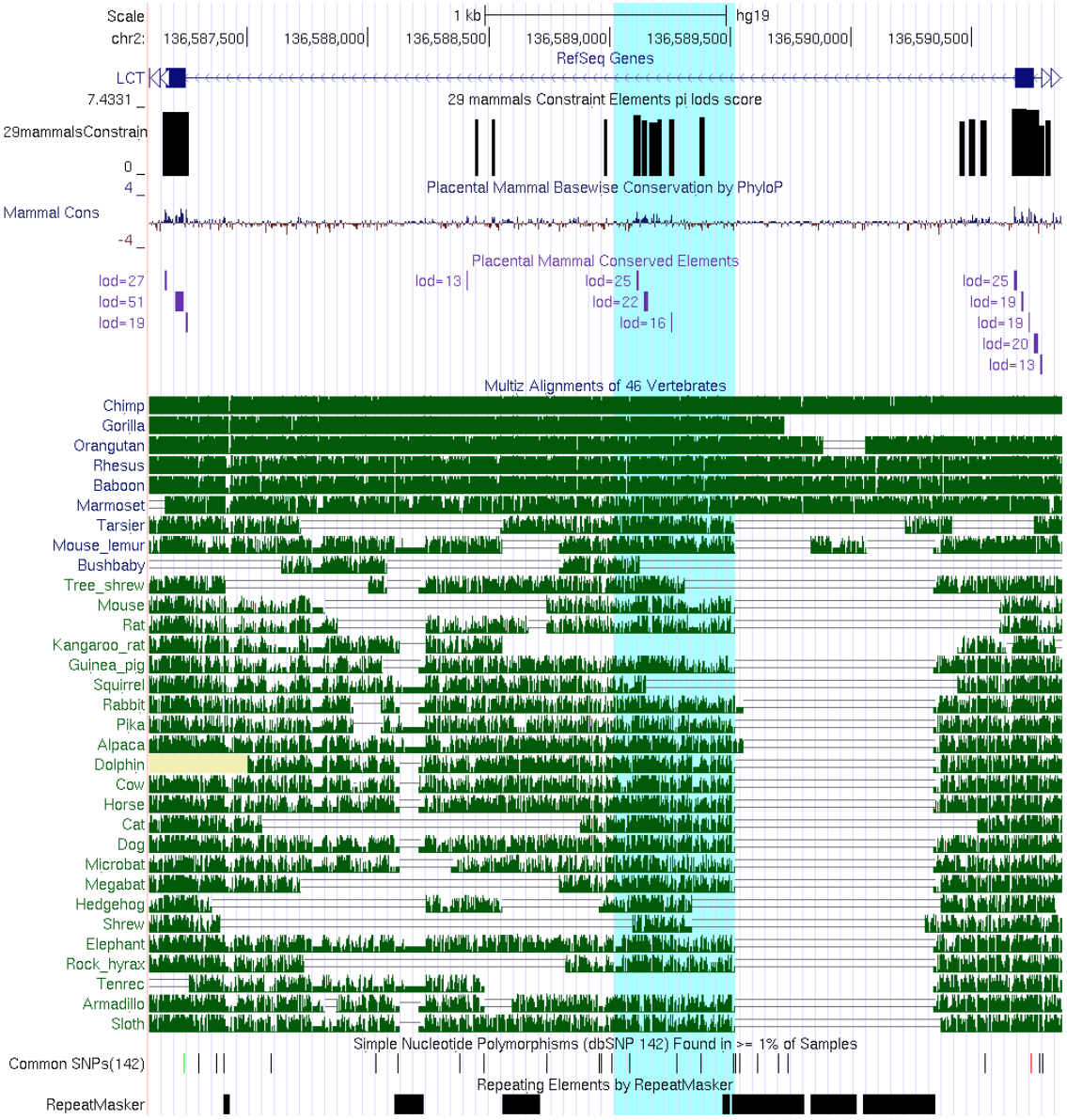



**Figure 4 - 20. Evolutionary conservation of the enhancer element at *LCT* intron 2**

Human (GRCh37/hg19) UCSC genome browser tracks showing high sequence conservation at the epigenetically-controlled region we identified in *LCT* intron 2 (highlighted in blue). Mammals show a sequence conservation peak and an enrichment in conserved elements at this *LCT* intron 2 region.

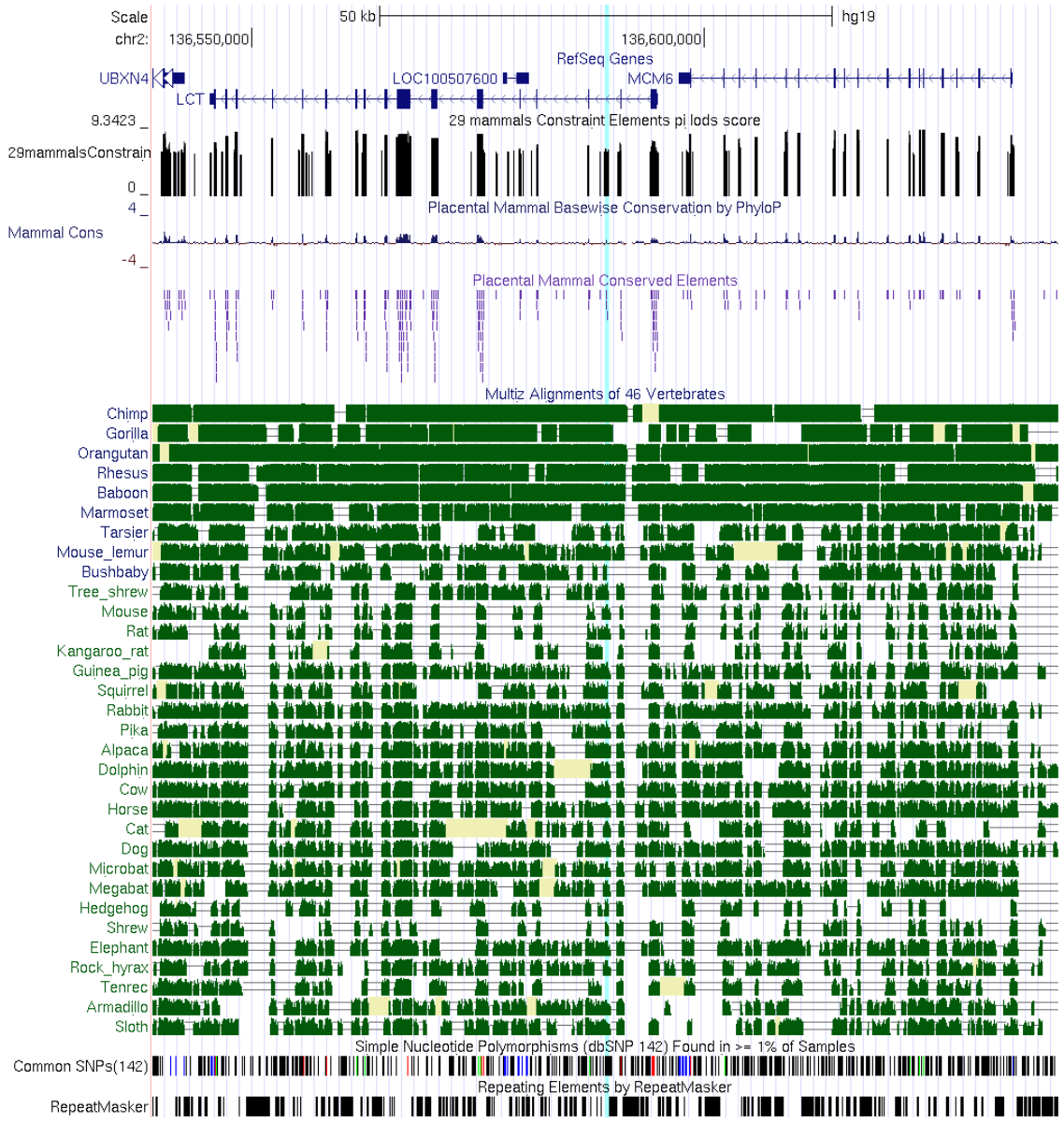



**Figure 4 - 21. *LCT* intron 2 shows substantial evolutionary conservation compared to other introns in *MCM6* and *LCT* region**

Human (GRCh37/hg19) UCSC genome browser tracks showing sequence conservation at *LCT* and *MCM6* locus. Mammals show high sequence conservation in the regulatory region we identified in *LCT* intron 2 (highlighted in blue). Outside of exons, sequence conservation at *LCT* and *MCM6* is uncommon.

## 4.4.5 LncRNA involvement in *LCT* expression

We further investigated the contribution of the previously uncharacterized lncRNA *LOC100507600* to the regulation of *LCT* mRNA. *LOC100507600* is transcribed from the

opposite DNA strand than *LCT*. Consequently, it follows that this lncRNA is a natural antisense transcript, a type of regulatory element that can be controlled by DNA modification (Li et al., 2015), and that is increasingly recognized to alter target gene expression (Di Ruscio et al., 2013; Magistri, Faghihi, St Laurent, & Wahlestedt, 2012). We examined the human jejunum cohort 2 (n = 59) and detected a strong positive correlation between *LOC100507600* lncRNA and *LCT* mRNA ($R^2 = 0.73$; $p < 10^{-9}$; Figure 4-22a).

Since DNA modification at this lncRNA increases with age in lactase persistent individuals (carriers of T-13910), we investigated the capacity of the lncRNA to modulate *LCT* mRNA levels. We performed an RNAi experiment in the human intestinal epithelial cell line Caco-2, which have the TT genotype at SNP -13910. Though Caco-2 cells do not fully recapitulate human enterocytes, they do exhibit measureable *LCT* mRNA levels. Transfection with small-interfering RNAs directed to *LOC100507600* reduced its expression by 20%, and resulted in concomitant 25% reduction in *LCT* mRNA ($p < 0.05$; Figure 4-22). Hence, *LCT* mRNA levels were in part regulated by this lncRNA in cells relevant to lactase persistence. The mRNA of the non-targeted *MCM6* remained unchanged by *LOC100507600* silencing (Figure 4-22b).

To understand how increased DNA modification at the lncRNA promoter could augment *LOC100507600* lncRNA and *LCT* mRNA levels, we surveyed ENCODE datasets (ENCODE Project, Consortium, 2012). There was a CTCF (insulator) binding site overlapping the lncRNA promoter that was nearly ubiquitous across human tissues (present in 51 out of 52 cell lines, including Caco-2). Increased DNA modification inhibits CTCF binding at such insulator sites, which has been shown to activate tissue-specific gene expression (Yu et al., 2013), and facilitate the interaction of neighboring enhancers with their target promoters (Guo et al., 2015; Ong & Corces, 2014). However, further investigation is required to understand this *LCT*-regulating lncRNA.

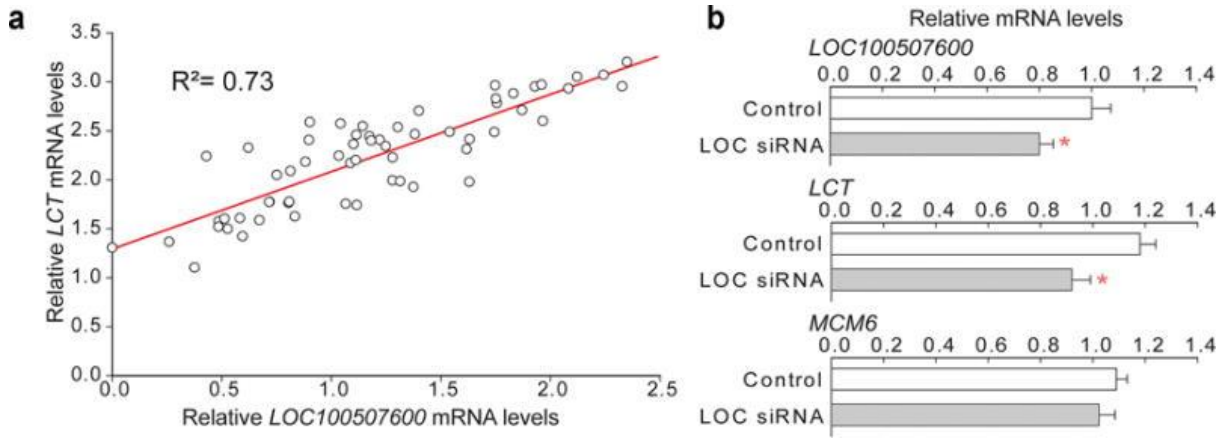



**Figure 4 - 22. A lncRNA LOC100507600 affects *LCT* mRNA levels in human intestinal cells**

(**a**) qPCR analysis of *LCT* and lncRNA levels in human enterocytes (n = 59 individuals). Regression plot showing a strong positive correlation between *LCT* mRNA and lncRNA levels. (**b**) RNAi-mediated reduction of lncRNA *LOC100507600* in human intestinal epithelial cells, Caco-2. Inhibition of lncRNA reduced steady-state *LCT* mRNA levels, but did not affect *MCM6* mRNA. Relative mRNA levels are shown for cells treated with a scrambled siRNA control (open bars) and a siRNA cocktail directed to lncRNA *LOC100507600* (grey bars). Data are means ± S.E.M. (n = 11 and 8 cell cultures). **p* < 0.05 compared to scrambled control; two-tailed Student's t-test.

### 4.4.6 Genetic manipulation of lactase regulatory elements

We used the CRISPR-Cas9 genome editing technology to determine whether the differentially modified DNA regions indeed contribute to the regulation of *Lct*. For this, we deleted intronic regions in *Lct* or *Mcm6* (Figure 4-23), and measured the effect on steady-state *Lct* mRNA (Figure 4-26). We deliberately targeted intronic regions only, avoiding splicing sites and exons that may compromise *Lct* mRNA processing and stability. In C57BL/6N mice, we deleted the *Lct* intron 2 region, which showed evidence of cross-species lactase gene regulation, and separately deleted a region in *Mcm6* intron 13, which was important to human *LCT* regulation. We also generated a deletion in *Lct* intron 1, serving as a positive control due to its overlap with a *Lct* promoter element. As a negative control, we measured *Mcm6* mRNA levels in mice carrying the *Mcm6* intron 13 deletion. We also surveyed and did not find any off-target activity of the CRISPR-Cas9 system in 46 candidate regions (Figure 4-24 and 4-25).

In adult and infant mice, deletion in *Lct* intron 1 as well as a deletion of *Lct* intron 2 caused widespread downregulation of *Lct* throughout the duodenum and jejunum (3- to 8-fold

reduction; $p < 0.001–10^{-12}$; Figure 4-26). The deletion in *Mcm6* intron 13 yielded less pronounced, segment-specific *Lct* mRNA reductions (less than 2-fold reduction; $p < 0.05–10^{-3}$; Figure 4-26). This Mcm6 intron 13 deletion did not alter *Mcm6* mRNA levels in adult or infant mice, supporting that our CRISPR-Cas9 deletions specifically affected *Lct* (Figure 4-26). Compared to the other deletions, the mutation in *Lct* intron 2 produced a significantly greater loss in *Lct* mRNA in adults than in infants (2-fold greater reduction in adult intestine, $p < 10^{-3}$; Figure 4-26). This suggests that the evolutionary conserved *Lct* intron 2 establishes a stronger control of *Lct* mRNA in adulthood, corroborating our above findings of aging mice.

We also examined the extent to which the epigenetically-regulated regions impact human *LCT* expression in cultured human intestinal epithelial cells, Caco-2. Caco-2 cells undergo spontaneous differentiation into enterocyte-like cells of the small intestine when grown past confluence (Troelsen et al., 2003). We assayed *LCT* expression in Caco-2 cells carrying a CRISPR-Cas9n-induced deletion in the regulatory elements in *MCM6* intron 13 or *LCT* intron 2. We also deleted a region not overlapping a regulatory element in *LCT* intron 1, serving as a negative control. Cells were examined at a time point prior-to-confluence (undifferentiated, cancer-like state, day 6) and post-confluence (differentiated, intestine epithelial-like state, day 15). The deletions did not affect *LCT* mRNA expression in the undifferentiated state (Figure 4-26). However, in the differentiated, epithelial-like cell state the deletion in *MCM6* intron 13 or *LCT* intron 2 resulted in a significant decrease in *LCT* mRNA levels ($p < 10^{-3}$; Figure 4-26), supporting the importance of these regions in regulation of *LCT* expression in humans.

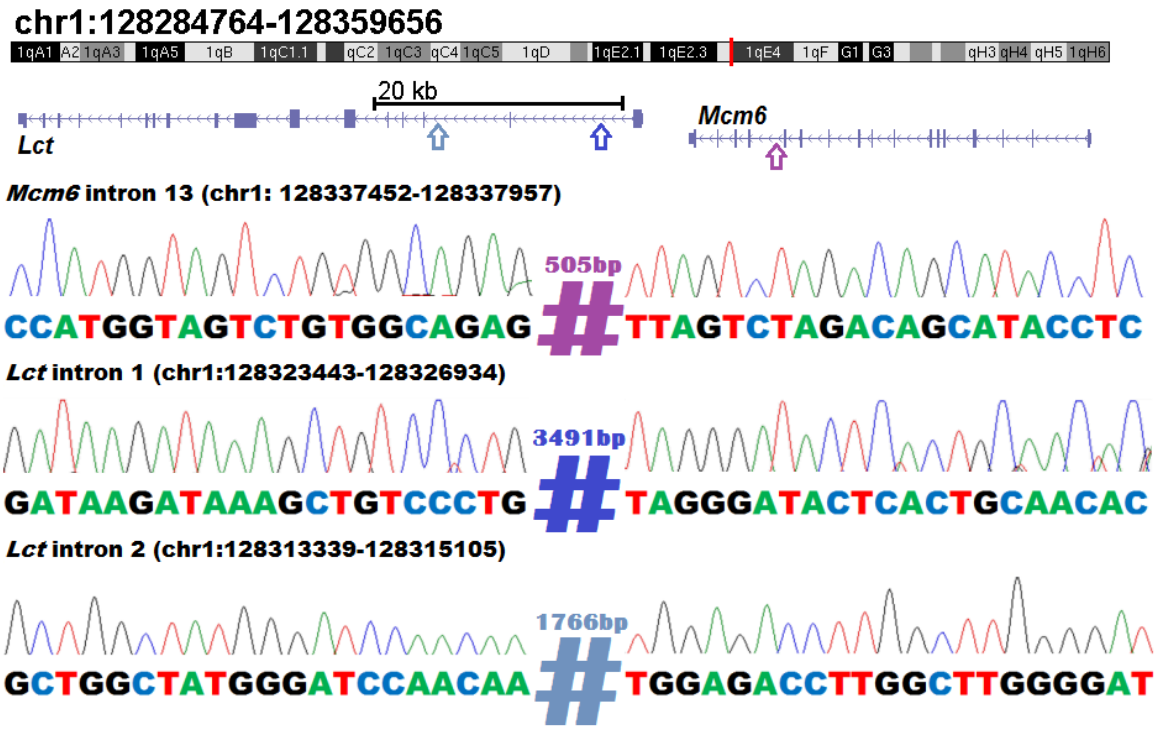



**Figure 4 - 23. CRISPR-Cas9-induced deletions in mice**

Chromatograms showing mutations in *Lct* intron 1, *Lct* intron 2 and *Mcm6* intron 13 generated by the CRISPR-Cas9 genome editing technique in C57BL/6N mice.

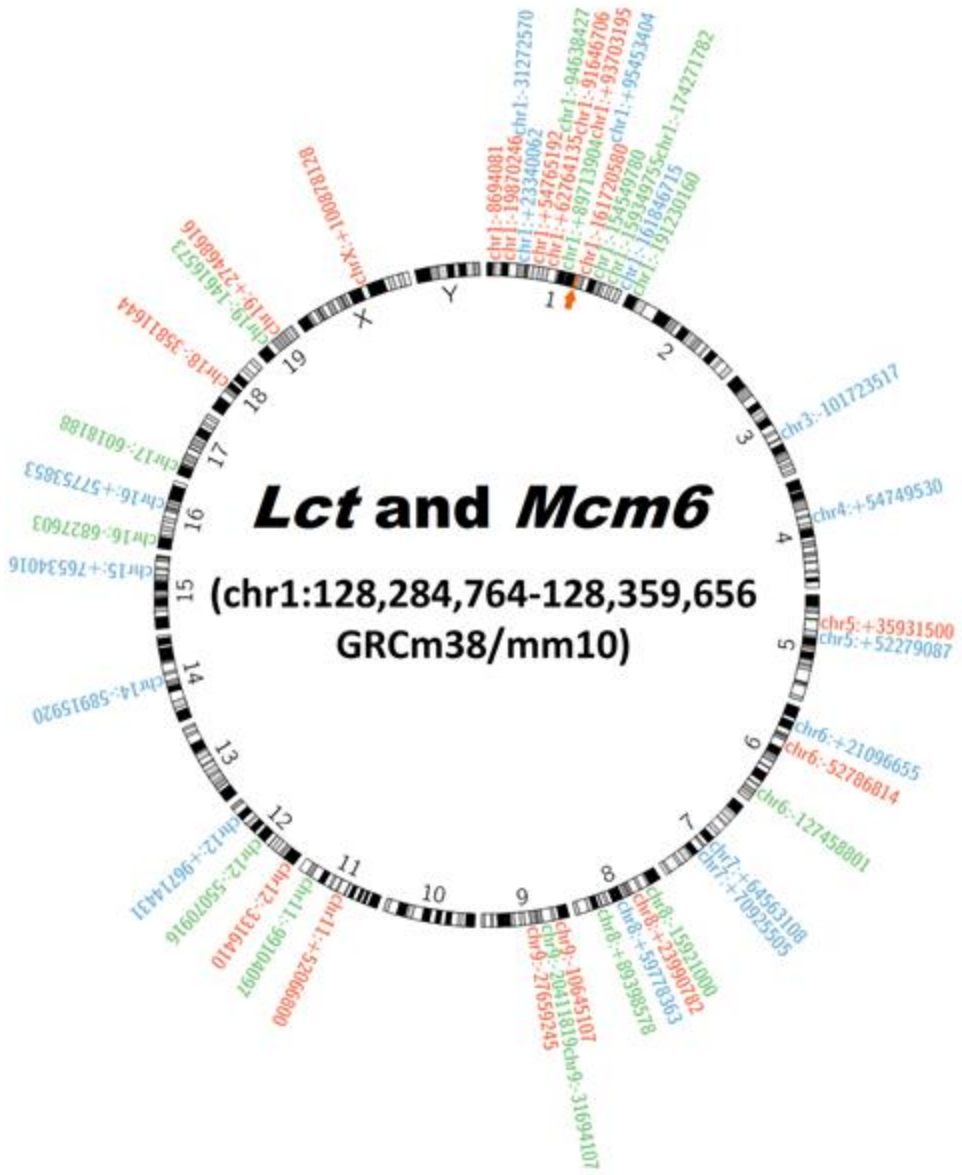



**Figure 4 - 24. Off-target analysis for the CRISPR-Cas9 modified mice**

The genomic regions most likely to be affected by non-specific activity of the sgRNAs used to guide the Cas9 nuclease for the deletions, as predicted by the CRISPR design tool, were examined by Sanger sequencing of PCR amplicons. A total of 46 sites were tested in the genetically modified mice and no off-target mutations were detected. Red, blue and green coordinates are the sites tested for the *Mcm6* intron 13, *Lct* intron 1 and *Lct* intron 2 sgRNA guides, respectively. The orange arrow indicates the *Lct* and *Mcm6* locus. Image created using CIRCOS version 0.67.

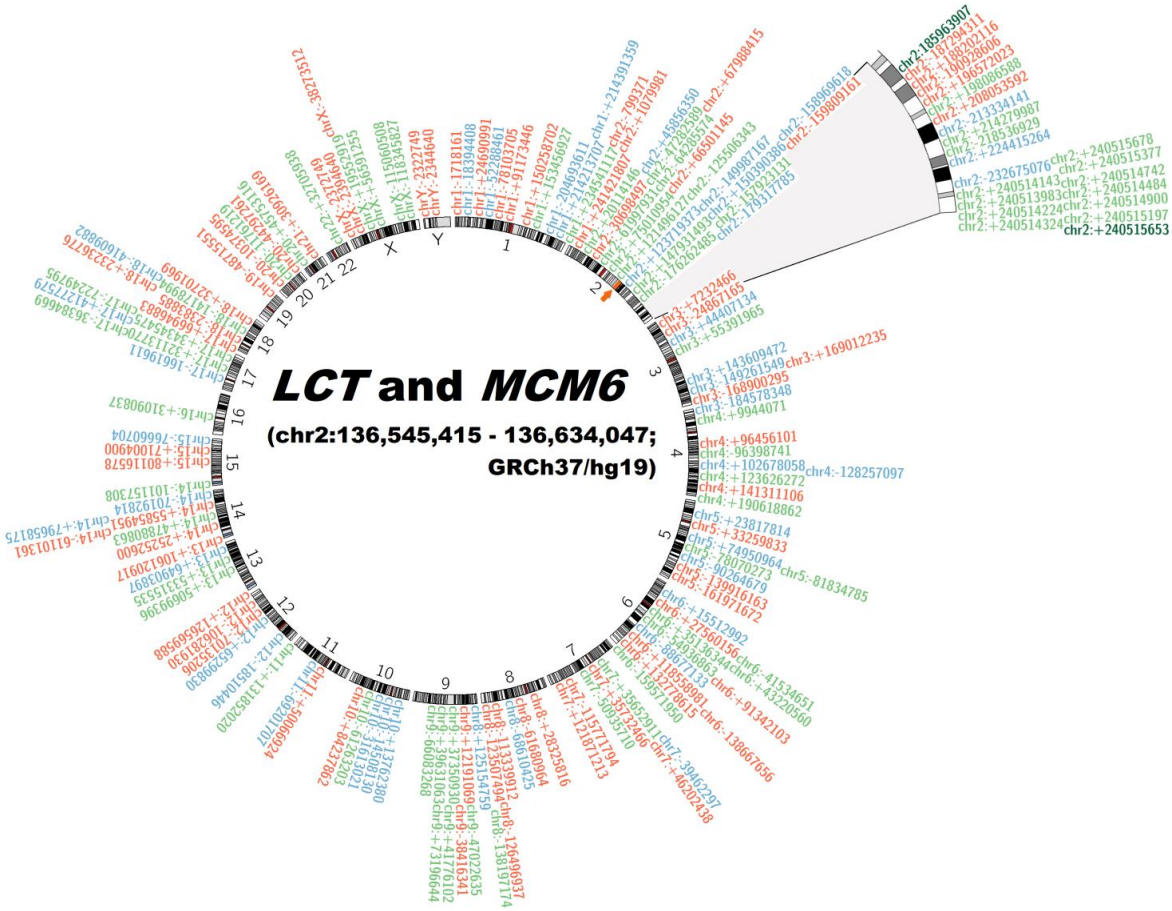



**Figure 4 - 25. Off-target analysis for the CRISPR-Cas9n modified human Caco-2 cell line**

The genomic regions most likely to be affected by non-specific activity of the sgRNAs used to guide the Cas9n deletions, as predicted by the CRISPR design tool, were examined by Sanger sequencing of PCR amplicons. A total of 172 sites were tested in the genetically modified cells and 2 off-target mutations were detected in the *LCT* intron 2 line (highlight in bold, top right insert). The off-target activity was at chr2: 185,963,907 (11 bp deletion) and chr2: 240,525,653 (21 bp insertion). These off-target deletions were both over 49 Mb from the *LCT* locus and unlikely to explain the differences in *LCT* mRNA in CRISPR-Cas9n-modified cells. Red, blue and green coordinates are the sites tested for the *Mcm6* intron 13 (n = 66 sites), *Lct* intron 1 (n = 41 sites) and *Lct* intron 2 (n = 65 sites) sgRNA guides, respectively. The orange arrow indicates the human *Lct* and *Mcm6* locus. Image created using CIRCOS version 0.67.

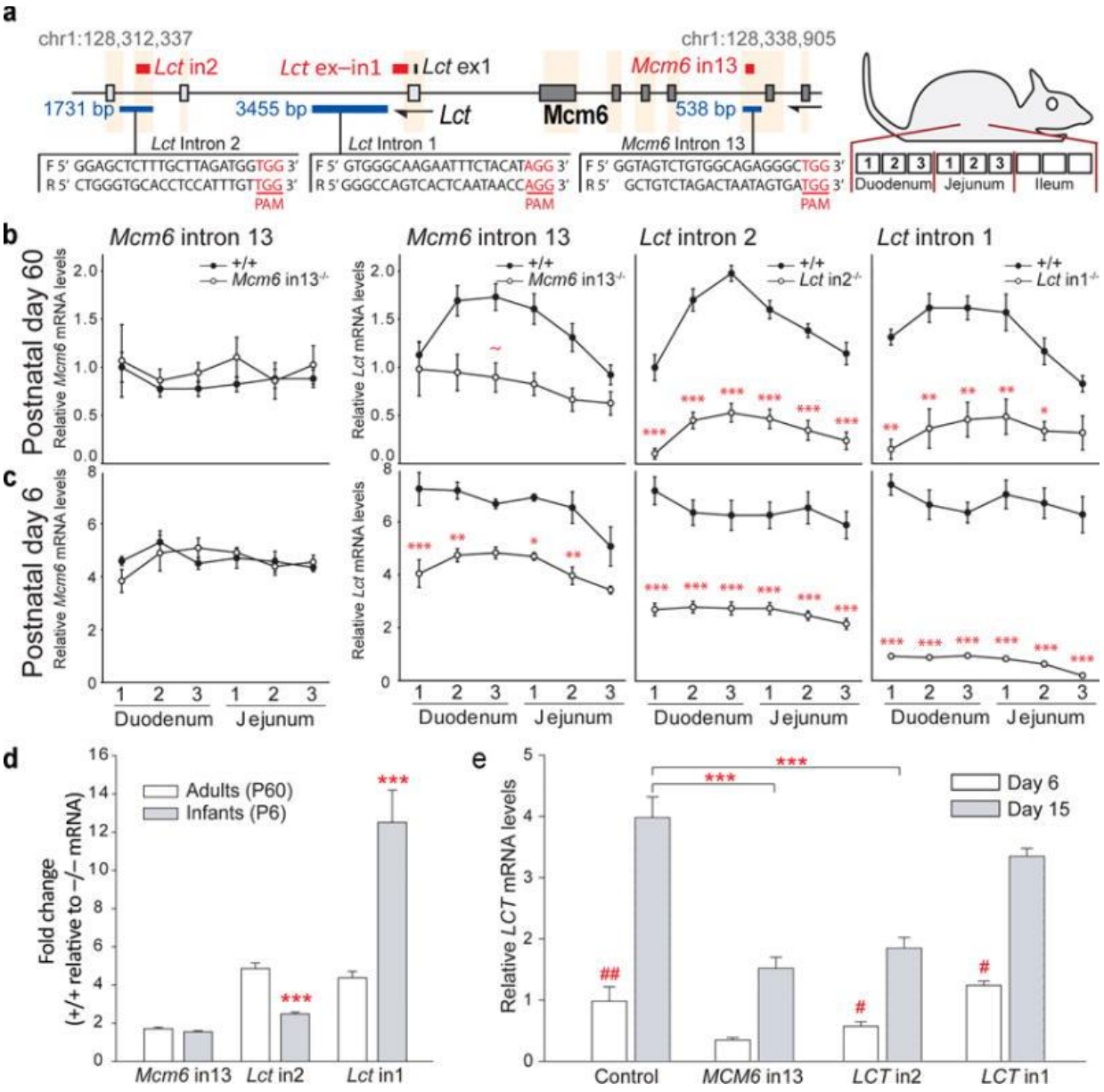



**Figure 4 - 26. CRISPR-Cas9-mediated deletions of lactase regulatory elements in mice and a human cell line**

(**a**) Schema showing guide RNA sequences and the resulting CRISPR-Cas9-mediated deletions in *Lct* intron 2, intron 1 and *Mcm6* intron 13 in mice. (**b, c**) *Lct* and *Mcm6* mRNA levels in the proximal, middle and distal sections of the duodenum and jejunum of postnatal day 60 (adult) and day 6 (infant) wild-type (+/+) and homozygote mutant (−/−) mice. One-way repeated measures ANOVAs showed significant loss in *Lct* mRNA due to regulatory element deletion in *Mcm6* intron 13 (adult: $F_{1,15} = 7.6$, $p < 0.05$; infant: $F_{1,7} = 43.7$, $p < 10^{-3}$), *Lct* intron 2 ($F_{1,30} = 162.3$, $p < 10^{-12}$; $F_{1,18} = 50.9$, $p < 10^{-5}$) and *Lct* intron 1 ($F_{1,10} = 27.8$, $p < 0.001$; $F_{1,8} = 137.2$, $p < 10^{-5}$). *$p < 0.05$, **$p < 0.01$, ***$p < 10^{-3}$ compared to wild-type mice. (**d**) Averaged fold change in *Lct* mRNA in adult and infant genetically-modified mice. Two-way ANOVA showed a significant mutation-age interaction ($F_{2,46} = 116.4$, $p < 10^{-21}$). ***$p < 10^{-3}$ compared adults with same mutation. (**e**) Deletions of regulatory elements in *MCM6*intron 13 and *LCT* intron 2, and a

negative control sequence in *LCT* intron 1 in human intestinal epithelial cells, Caco-2. Undifferentiated Caco-2 cells (day 6) and differentiated cells, which exhibit an enterocyte-like state (day 15) examined. *LCT* mRNA in differentiated, enterocyte-like Caco-2 cells was reduced by deletion in *MCM6* intron 13 and *LCT* intron 2 (genotype effect: $F_{3,176} = 11.7$, $p < 10^{-6}$ and genotype-day interaction: $F_{3,176} = 4.8$, $p < 0.01$; by two-way ANOVA). ***$p < 10^{-3}$ compared to Cas9n-transfected controls; #$p < 0.05$, ##$p < 10^{-4}$ within genotype. (**b, c, d, e**) *p*-values are Tukey's HSD post-hoc test. Data are means ± S.E.M. Mouse sample sizes (+/+,−/−): *Mcm6* intron 13 P60 n = 11, 6; P6 n = 4, 5), *Lct* intron 2 (P60 n = 13, 19; P6 n = 10, 10) and intron 1 (P60 n = 8, 4; P6 n = 6, 4). Cell culture sample sizes (left-to-right) n = 41, 46, 7, 8, 26, 32, 12 and 9.

## 4.5 Discussion and Conclusion

This study reveals that epigenetic factors are involved in the regulation of the human and mouse lactase genes, and by corollary lactase non-persistence and lactose intolerance. DNA modification studies, in combination with chromatin maps and our functional studies of mutant mice and cell lines, revealed quite a complex regulation of the lactase gene. Indeed, epigenetic modifications targeting several different regulatory elements account for species- and tissue-specific effects as well as the inter-individual variation of *LCT* expression. A key finding is that different *LCT/MCM6* variants confer differential accumulation of DNA modification over time. Our findings suggest that the lactase non-persistence C-13910 allele accumulate modified cytosines that silence the regulatory elements in *MCM6* and *LCT*, while the lactase persistence haplotype containing the T-13910 allele displays age-related modification changes to maintain *LCT* activity. Since infant *LCT* mRNA levels are high irrespective of DNA haplotype, it can be speculated that DNA variation is not directly involved in *LCT* transcription. Rather, genetic variation may mediate the accumulation or loss of epigenetic modifications at specific enhancers and promoters, which, in turn, directly orchestrate gene activity.

Our study is a first step in understanding the interactions of genetic and epigenetic mechanisms in the lactase gene regulation. Previous studies have reported that SNPs associated with lactase persistence/non-persistence show allelic differences in transcription factor binding (Jensen et al., 2011; Lewinsky et al., 2005). For example, compared to the C-13910 allele, the T allele shows greater binding to POU2F1 in combination with HNF1α (Jensen et al., 2011; Lewinsky et al., 2005). In addition, transcription factor binding differences co-occur with changes in DNA modifications chromosome-wide during the differentiation of enterocytes along the crypt-to-villi axis; a process necessary for *LCT* expression along this axis (Kaaij et al., 2013; Sheaffer et al., 2014). It has been shown that the DNA methylation machinery competes with the binding of transcription factors *in vivo* (Domcke et al., 2015). The age-dependent accumulation of modified cytosines at *LCT/MCM6* regulatory elements may be circumvented in the lactase persistent T-13910 allele due to its higher affinity for transcription factor binding. Notably, SNPs do not need to directly overlap *LCT/MCM6* enhancer/promoter elements, as local DNA modification profiles can be regulated by nearby changes to transcription factor motifs and DNA-transcription factor interactions (Domcke et al., 2015). However, epigenetic coordination between more distally located enhancers and promoters may require long-range looping

interactions of such elements. Our analysis of ENCODE data reveals an enrichment of CTCF binding sites at the lncRNA promoter and elsewhere within the *LCT/MCM6* locus, including the *MCM6* intron 13 region. Differential CTCF binding (as well as lncRNA recruitment) has been shown to affect chromatin conformation, which in turn alters accessibility and proximity of genes and regulatory elements (Guo et al., 2015; Ong & Corces, 2014; Xiang et al., 2014). DNA modification differences between *LCT/MCM6* haplotypes could influence the extent of CTCF binding, which in turn affects chromatin architecture and results in genetically- and epigenetically-driven alterations in enhancer-promoter interactions. Overall, genetic variation seems to be setting up the stage, either allowing or preventing DNA modification accumulation with age at key regulatory sites, which could impact regulatory site accessibility and interactions. In humans, DNA modifications within *MCM6* regulatory regions appears to be centrally involved in enabling lactase persistence (escape from the ancestral lactase non-persistence program), while epigenetic control of *LCT* regulatory regions seems to be highly relevant to cell-type specific expression of *LCT*. Our study also suggests an evolutionary conservation of epigenetic regulation, which appears to precede genetic polymorphisms for the evolutionarily favorable lactase persistence trait (Abi Khalil, 2014) (Heyn, Moran, et al., 2013; You & Jones, 2012). Future studies examining chromatin configuration (Ong & Corces, 2014) at the *LCT/MCM6* locus in aging individuals and targeted epigenetic editing with the CRISPR-Cas9 system (Hilton et al., 2015) will be necessary to further understand genetic-epigenetic contributions to lactase persistence/non-persistence.

Our study helps to further delineate the principles and guidelines for epigenomic studies of other traits and diseases. The main lessons learned are 1) the required investigation of the specific cell type centrally involved in the phenotype rather than bulk of the tissue or organ, 2) use of quantitative phenotypes and endophenotypes (in this case *LCT* mRNA levels that are the required template for production of lactase enzyme), 3) sites containing stretches of differentially modified CpGs were replicable 4) the benefit of cross-species comparisons which showed a partial conservation of regulatory elements between mice and humans, 5) examination of multiple epigenetic layers to gain insight into regulatory element functions, and 6) the potential for DNA sequence risk factors to closely associate with epigenetic abnormalities that together influence gene regulatory elements.

The significance of this study may extend far beyond the lactose-related phenotype and could benefit etiological investigations of complex aging diseases, such as cancer, type 2 diabetes, and Alzheimer's disease. Age-dependent DNA modification states have the potential to serve as an additional layer of context to genetic risk factors (e.g. GWAS hits). Our study suggests that such DNA variants become *bona fide* disease risk factors once they reach a critical mass of epigenetic misregulation in the aging cell. Genetic-epigenetic interactions may also be involved in the age-dependent loss of pathogenicity of risk genes. For example, children affected with asthma, epilepsy, atopic dermatitis, and attention deficit and hyperactivity disorder can partially or even fully recover upon reaching young adulthood (Belsky et al., 2013; Delaney & Smith, 2012). Thus, investigation of age-dependent epigenetic trajectories of disease-associated genetic variants should become an inseparable component of genome-wide association studies, and this effort may be pivotal to uncovering the origins of complex phenotypes.

# 5 Transcriptional heterogeneity in the lactase gene within cell-type is linked to the epigenome

This chapter is modified from the following:

*Oh, E., *Jeremian, R., Oh, G., Groot, D., Susic, M., Lee, K., Foy, K., Laird, P., Petronis, A., Labrie, V. (2017). Transcriptional heterogeneity in the lactase gene within cell-type is linked to the epigenome. *Sci Rep*, 7, 41843. doi:10.1038/srep41843

* Equal contribution

## 5.1 Abstract

Transcriptional variation in histologically- and genetically- identical cells is a widespread phenomenon in tissues, yet the processes conferring this heterogeneity are not well understood. To identify contributing factors, we analyzed epigenetic profiles associated with the *in vivo* transcriptional gradient of the mouse lactase gene (*Lct*), which occurs in enterocytes along the proximal-to-distal axis of the small intestine. We found that epigenetic signatures at enhancer and promoter elements align with transcriptional variation of *Lct* in enterocytes. Age and phenotype-specific environmental cues (lactose exposure after weaning) induced changes to epigenetic modifications and CTCF binding at select regulatory elements, which corresponded to the alterations in the intestinal *Lct* mRNA gradient. Thus, epigenetic modifications in combination with CTCF binding at regulatory elements account for the transcriptional gradient in *Lct* in cells of the same type. Epigenetic divergence within enterocytes may contribute to the functional specialization of intestinal subregions.

## 5.2 Introduction

Within a tissue, seemingly identical cells have been found to exhibit a large degree of variation in their transcriptomes (Dueck et al., 2015; Lovatt et al., 2014). Although this cell-to-cell variation can be partially stochastic (Munsky, Neuert, & van Oudenaarden, 2012), there are coordinated transcriptional gradients in cells of the same type across whole organs. For example, prominent transcriptional gradients have been detected in intestinal epithelial cells (enterocytes) along the length of the small intestine (Grun et al., 2015; Haber et al., 2017), in hepatocytes across liver zones (Halpern et al., 2017; Yang et al., 2017), in myocytes in different heart chambers (Liu et al., 2017; See et al., 2017; Skelly et al., 2018), and in adipocytes located in bone marrow (Acosta et al., 2017; Ehrlund et al., 2017; Spaethling et al., 2016). Gradients in transcriptional states enable cells that are histologically identical to perform biologically distinct roles, vary in their response to environmental stimuli, and exhibit differential disease vulnerability (Avraham et al., 2015; Dueck et al., 2015). Despite its evident biological significance, the molecular processes coordinating same cell-type transcriptional divergence across tissue subregions are not well understood.

Epigenetic mechanisms could contribute to transcriptional variation within the same cell type, as epigenetic modifications are important in regulating gene transcription (Kundaje et al., 2015), determining cell identity (Barrero, Boue, & Izpisua Belmonte, 2010; Sheaffer et al., 2014), and affecting genomic functions in response to aging and environmental cues (Benayoun, Pollina, & Brunet, 2015). Recently, DNA modification analysis of single cells from *in vitro* cultures found significant heterogeneity (Angermueller et al., 2016; Farlik et al., 2015; Stelzer, Shivalila, Soldner, Markoulaki, & Jaenisch, 2015), suggesting the potential for epigenetic marks to differ between cells of the same type in living organisms. In this study, we sought to examine the divergence of DNA modifications within a cell type or between subtypes of cells *in vivo* and explored its potential in regulating same cell type transcriptional gradients in tissue. To answer this question, we examined transcriptional and epigenetic variation of the lactase gene (*Lct*) in enterocytes along the proximal-to-distal axis of the mouse small intestine. *Lct*, which is responsible for lactose metabolism, represents an ideal model to identify DNA modifications that can contribute to transcriptional heterogeneity for tissue subregion specialization as it exhibits a distinguished transcriptional gradient in enterocytes; a gradual elevation from duodenum to jejunum, followed by a steady decline to the ileum (Fang et al., 2006; Lee, Russell, Montgomery,

& Krasinski, 1997; Rings et al., 1994). Recent studies have shown that *Lct* mRNA expression is dependent upon DNA modifications at key genomic regulatory elements (V. Labrie et al., 2016; Sheaffer et al., 2014), however, the extent to which DNA modifications regulate *Lct* transcription along the length of the intestine remains uninvestigated. In addition to this, we also examined how aging and the environment can modify *Lct* expression through DNA modifications at *Lct* genomic regulatory elements.

Here, we found that in enterocytes isolated from different intestinal regions, DNA modifications at specific genomic regulatory elements were concordant with the transcriptional variation of *Lct*. Aging and environmental exposures (lactose feeding after weaning) resulted in DNA modification changes at these regulatory sites and changes in *Lct* transcriptional patterns. Overall, DNA modification patterns support the aging- and environmentally-induced gradients of *Lct mRNA* and, more broadly, could affect phenotypic outcome by modifying the transcriptional program of a single gene within same cell types.

## 5.3 Materials and Methods

### 5.3.1 Mouse intestinal samples and milk treatments

All animal procedures were approved by the Institutional Animal Care Committee of the Toronto Centre for Phenogenomics (TCP) and compiled per the requirements of the Canadian Council on Animal Care and Province of Ontario Animals for Research Act. C57BL/6NCrl mice were bred at the TCP. Mice were housed in ventilated polycarbonate cages, and given *ad libitum* sterile food (Harlan 2918X) and water, unless stated otherwise. Adult mice were housed by sex in groups of 2-5 littermates. The vivarium was maintained under temperature (21°C ± 1°C) and humidity (50-60%), with a 12-hour diurnal cycle (lights on 0700-1900).

Infant mice at postnatal day 6 (P6) and adult mice at postnatal day 60 (P60) were used to investigate the segmental gradient and age-associated epigenetic changes. The small intestine was harvested and dissected into nine portions; the duodenum (segments 1-2), jejunum (segments 3-5) and ileum (segments 6-9). Each of these anatomical segments was further dissected into three equal parts resulting in nine dissections of the small intestine (segments 1-9). To investigate the diet/environmental associated epigenetic changes, postnatal day 30 mice were supplied *ad libitum* sterile food with either 2% lactose-containing organic milk (Grand Pré) or 2% lactose-free milk (Natrel) in the place of water for 60 days. Approximately 500 ml of lactose-containing milk (LAC+) was supplemented with 1 tablespoon (or ~12 g) of sucrose (Redpath) to promote a consumption level comparable to that of mice fed lactose-free milk every 1.5~2 days. No sex differences were observed in either DNA modification or *Lct* mRNA expression level. Mouse intestinal samples were processed by researchers blind to genotype and experimental conditions.

### 5.3.2 Enterocyte isolation and DNA extraction

Enterocytes were isolated specifically from the villi of each small intestine segment (approx. 100 mg) using a previously validated method (V. Labrie et al., 2016; Weiser, 1973). The segments were washed four times in 1 mL citrate buffer (27 mM Na-citrate, 5.6 mM $Na_2HPO_4$, 96 mM NaCl, 8 mM $KH_2PO_4$, 1.5 mM KCl, pH 7.4) containing 10 µl of DNase inactivation reagent (Life Technologies) with gentle agitation. Enterocytes were released with the addition of 1 mL $Ca^{2+}$-chelating buffer (1.5 mM EDTA, 0.5 mM DTT, 10 mM $NaH_2PO_4$, 154 mM NaCl) at 37°C

for 15 min with agitation (700 rpm). The intestinal segments and suspended enterocytes were separated, and the enterocytes were collected by centrifugation at 1000 rpm for 5 mins at 4°C. This was repeated twice using fresh $Ca^{2+}$-chelating buffer for maximal recovery of enterocytes. The purity and specificity of the villi enterocyte isolation was confirmed using histological investigation (Labrie et al., 2016) and fluorescence-activated cell sorting (see below). DNA was extracted from enterocytes and the small intestinal segment lacking enterocytes using standard phenol-chloroform DNA extraction methods in combination with Phase-Lock tubes (5 Prime) after an overnight treatment in Proteinase K. DNA quantity was measured by NanoDrop ND-1000, and DNA quality was verified on a 1% agarose gel. The purity and specificity of the villi enterocyte isolation was confirmed to be 84.8% ± 4.4 (median ± S.E.M.) across intestinal segments through histological analysis (V. Labrie et al., 2016) and fluorescence-activated cell sorting (FACS) using mice expressing fluorescent protein, mKate2, under the control of enterocyte-specific *Villin-1* promoter. The construct for the *Villin-mKate2* transgenic line was generated using the intestine-specific mouse *Villin-1* promoter excised from the pBSKSVillinMESSV40polyA vector (gift from Dr. Louvard) and a *mKate2* gene purchased from the Evrogen. The construct was then injected into B6D2F1 hybrid mouse male pronucleus at the USC transgenic core facility. Isolated enterocyte pellets were resuspended in 5ml Accutase (STEMCELL Technologies) and incubated at 37°C for 20 mins. Accutase treated cells were then centrifuged at 300xg for 5 minutes and the resulting cell pellets were resuspended in 500 µl PBS for flow sorting. Single-cell suspensions were stained with DAPI (4′,6′-Diamidino-2-phenylindole) and alive enterocytes were used for subsequent sorting. Purity of enterocytes following our chelating protocol was determine by comparing the same intestinal segments from a *Villin-mKate2* mouse to those of a mouse littermate lacking *mKate2* (Figure 5-1). The Van Andel Research Institute Flow Cytometry Facility examined cells on a CytoFlex S Flow Cytometer (Beckman Coulter Inc.) and data were analyzed using FlowJo (FlowJo LLC.) (Figure 5-1).

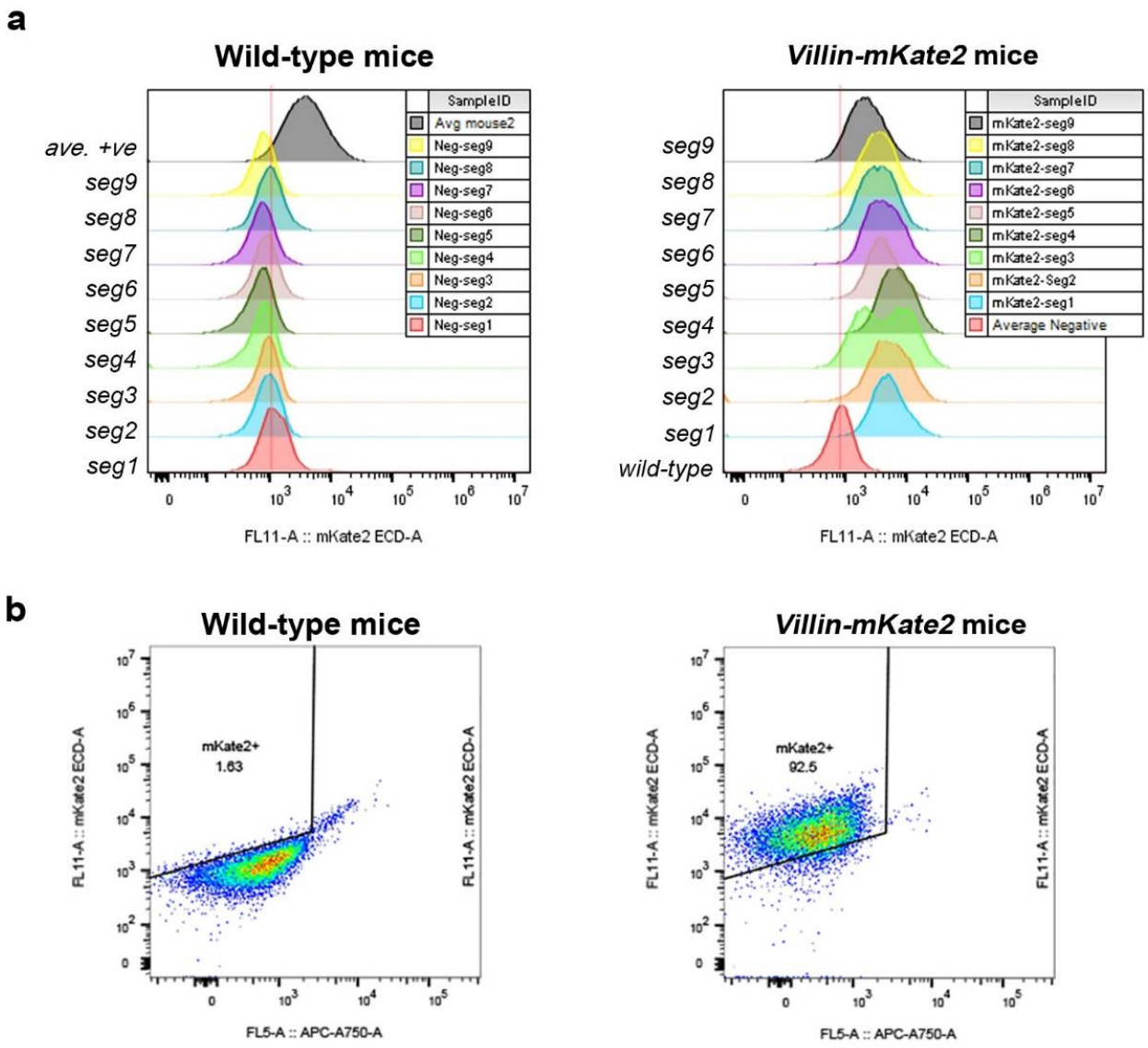



**Figure 5 - 1. Flow cytometry analysis of purity of enterocytes that were isolated using the chelating method**

Enterocytes were isolated from Villin-mKate2 expressing mice and wild-type littermates by intestinal segments. a) Histogram depicting isolated enterocytes from wild-type mice (left) and mice expressing mKate2 under the control of the enterocyte-specific villin promoter (right). Each of the nine intestinal segments were examined for mKate2 fluorescence (x-axis) in wild-type and Villin-mKate2 mice. Average positive and average negative refers to the average mKate2 signal across all intestinal segments of Villin-mKate2 and wild-type mice, respectively. b) Representative dot plot of isolated enterocytes. Using the chelating method, we were able to achieve ~84.8% purity in the intestinal segments. Y-axis: mKate2 fluorescence channel.

DNA modification state in the villi enterocytes was examined with single nucleotide resolution in the *Lct–Mcm6* locus of mice using the bisulfite padlock probe technique (Diep et al., 2012). Padlock probes (n = 314) were designed to target non-repetitive genomic sequences on both DNA strands (GRCm38/mm10; Appendix A-5). Enrichment of targeted bisulfite-converted DNA was done as described (Diep et al., 2012), and sequencing was performed on an Illumina HiSeq 2500. Preprocessed reads were mapped onto a bisulfite-converted mouse genome (GRCm38/mm10) using Bismark (Krueger & Andrews, 2011). Modification estimates were included only for cytosines with >30 reads. DNA modification density was interrogated at 7,580 modified cytosines (609 CpG sites) across 148 unique samples, and an additional 19 replicates. All data are available from the NCBI Gene Expression Omnibus (GEO) database under accession number GSE76373.

### 5.3.3 Fine-mapping of DNA modifications in enterocytes

DNA modification state was examined with single nucleotide resolution in the *Lct* and *Mcm6* loci of mice using the bisulfite padlock probe technique (Diep et al., 2012; V. Labrie et al., 2016). All investigations of DNA modifications were performed on isolated villi enterocytes. 314 padlock probes with 6 base pairs of degenerate sequences were targeted to unique (non-repetitive) genomic sequences (GRCm38/mm10). Probes covered 90kb of the region of interest for both forward and reverse DNA strands. Target DNA regions were designed for bisulfite converted DNA using ppDesigner software. CpG sites within probes were modified to capture both modified and unmodified cytosines. Probe sequences are found in Supplementary File 1. Using a custom pipeline based on the Bismark tool (Krueger & Andrews, 2011), DNA modification density was interrogated at every cytosine covered by padlock probes. Reads were first trimmed and residual adapter sequences removed using Trimmomatic-0.32. To remove phiX sequencing control DNA, reads were aligned to the phiX reference genome (NC_001422.1) using Bismark (560 bp from the beginning of the phiX genome was appended to the end to accommodate the circular nature of the genome for read alignment). Reads that did not align to phiX were aligned to the target reference genome (GRCm38/mm10). Percent modification was estimated as the fraction of spanning reads that retained the reference “C”, and were not converted to “T” from the bisulfite treatment. Modification estimates were included only in those

cytosines which received 30 or more reads. DNA modification density was interrogated at 7,580 cytosines (609 CpG sites inclusive) across 148 unique samples, and additional 19 replicates. CpG sites were analyzed for outliers (interquartile range <0.05) and sparsity issues (coverage in fewer than half of the samples), but none were found, and no sample-wise outliers were found by 1.5IQR limits. Biological variation exceeded technical noise, as the mean deviation of CpGs between randomly paired samples (average standard deviation, 0.0414 ± 0.0114) was higher than between were same sample replicates (average standard deviation, 0.0232 ± 0.0069).

### 5.3.4 Statistical analysis of DNA modifications

All statistical analyses were performed with R statistical software version 3.2.0. To investigate whether DNA modifications of isolated villi enterocytes were associated with transcriptional variation in *Lct* in enterocytes across the small intestine, we checked for normality at individual cytosines using a QQ plot, and performed for each individual cytosine a Pearson's correlation between modification density and steady-state *Lct* mRNA levels, using the 9 segments of the small intestine (n = 53, 5-6 mice per segment). Significance was set at $p < 0.01$ after Bonferroni correction for multiple testing. Clusters of significantly associated modified cytosines contained 3 or more cytosines within 500bp. P-values are expressed as the –log p-value of the correlation coefficient, with the sign (+/-) representing the direction of Pearson's correlation (SLP). Significant CpGs were further validated using segmental differences in effect size and % modification status along the proximal-distal axis of the intestine by one-way ANOVA. *Mcm6* mRNA and DNA modification of *Mcm6* regulatory regions (Figure 5-2) did not show significant differences between the intestinal segments. No significant sex differences were observed as determined by repeated measures (RM)-ANOVA.

For testing age-associated change in DNA modifications, segment 1, 3 and 7 of P6 infants were compared against corresponding segments in P60 adults (n = 3-6 per segment per age group). We determined which genomic sites showed significant age-associated differences by linear regression. Diet-associated DNA modification changes were identified by comparing LAC+ (n = 34; 6-7 per segment) and lac- (n = 31; 5-7 per segment) P90 mice. A group-wise variance matrix showed that heavily modified CpGs (>90%) lacked deviation and these were

removed from analysis in this experiment (Table 5-1). Significant age- and diet-induced DNA modification changes were identified after Bonferroni correction.

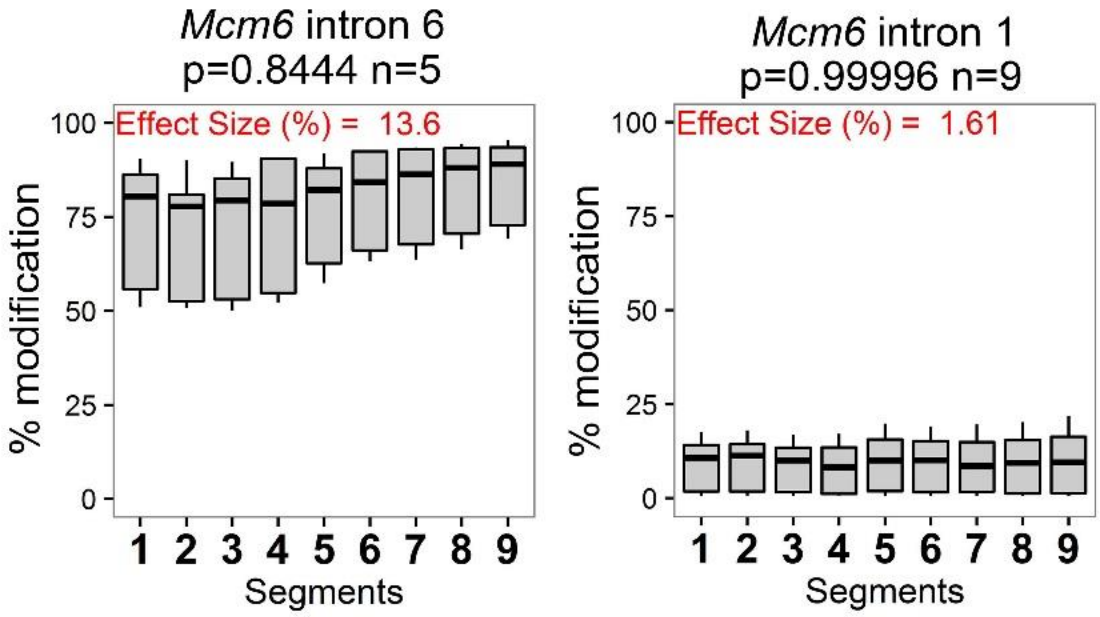



**Figure 5 - 2. Average% modification in *Lct* regulatory elements along the proximal-distal axis of the small intestine**

There is no segment-specific gradient in % modification status. $^{**}p < 0.01$, $^{***}p < 10^{-3}$ by Tukey's HSD post hoc test in comparison to segment 3.

### 5.3.5 Chromatin immunoprecipitation of CTCF

ChIP was performed to investigate CTCF binding in intestinal segments 1, 3, 5, 7 and 9 of both P6 and P60 mice (n = 3 per group), as well as segment 7 of LAC+ and lac- milk-fed mice (n = 3 per group). Tissue homogenization and ChIP were performed using the MAGnify ChIP kit (Life Technologies) per the manufacturer's instructions. Immunoprecipitation was performed overnight, using 3 µl monoclonal CTCF antibody (Pierce G.758.4) per ChIP reaction, and 1 µg of mouse IgG antibody (Life Technologies) for negative control reactions. A non-immunoprecipitated input control was also prepared for each sample. qPCR was performed in triplicate on all samples for five loci: *Mcm6* exon 13/intron 13 and exon 7, as well as *Lct* exon 1, intron 2 and intron 8. This was done using the SsoAdvanced Universal SYBR Green Supermix (Bio-Rad) using the Applied Biosystems ViiA 7 real-time PCR system. The reaction was carried

out per the following conditions in a 20 µl total reaction volume: 0.9X Universal SYBR Green Supermix, 10 µl recovered sample, 0.25 uM forward and reverse primer (primer sequences detailed in Appendix A-6); 98°C for 3 min and 50 cycles of (98°C for 15 sec, 55°C for 1 min). Percent input method with background normalization was calculated for further analysis.

| **Average $R^2$ (Pearson)** | **All CpGs** | **>90% mod** | **90-20% mod** | **<90% mod** | **<20% mod** |
|---|---|---|---|---|---|
| lactose-free milk vs itself | 0.9863 | 0.8703 | 0.9497 | 0.9824 | 0.8224 |
| Normal milk vs itself | 0.9899 | 0.8951 | 0.9563 | 0.9871 | 0.9130 |
| lactose-free vs Normal milk | 0.9893 | 0.8652 | 0.9595 | 0.9863 | 0.8502 |
| *Change in avg $R^2$* | *0.003586* | *0.024848* | *0.006649* | *0.004673* | *0.090554* |
| | | | | | |
| **Average Variance** | **All CpGs** | **>90% mod** | **90-20% mod** | **<90% mod** | **<20% mod** |
| lactose-free milk vs itself | 0.0705 | 0.0003 | 0.0313 | 0.0927 | 0.0016 |
| Normal milk vs itself | 0.0714 | 0.0004 | 0.0265 | 0.0925 | 0.0023 |
| lactose-free vs Normal milk | 0.0705 | 0.0003 | 0.0288 | 0.0919 | 0.0019 |
| *Change in avg variance* | *0.000837* | *0.000055* | *-0.004813* | *-0.000207* | *0.000708* |

**Table 5 - 1. DNA modification and diet at *Lct*/*Mcm6***

Diet-associated DNA modification between LAC+ (n = 39; 7-9 per segment) and lac- (n = 35; 5-9 per segment) investigated by Pearson's and variance matrices within each condition by modification status of individual CpGs.

## 5.4 Results

### 5.4.1 Within cell-type difference in *Lct* mRNA are correlated to epigenetic alterations

Enterocytes exhibit differing *Lct* transcriptional patterns along the proximal-distal axis of the small intestine, which enables intestinal subregion specialization in lactose metabolism (Figure 5-3a). We examined *Lct* mRNA levels in enterocytes from nine segments of the adult (postnatal day 60; P60) mouse small intestine; the proximal, middle and distal segment of the duodenum (segment 1-2), jejunum (segment 3-5), and ileum (segment 6-9), and observed a gradient in steady-state *Lct* mRNA levels (main effect of segment: $F_{8,46} = 23.4$, $p = 8.4\text{x}10^{-14}$; Figure 5-3b). *Lct* mRNA levels were highest in the proximal jejunum (segment 3), and then gradually declined toward the proximal duodenum ($p = 0.008$ compared to segment 3) and distal ileum ($p = 1.2\text{x}10^{-10}$). This indicates that enterocytes from the mid-duodenum to mid-jejunum are most specialized for the digestion of lactose.

Using a targeted bisulfite sequencing approach (Diep et al., 2012), we sought to determine whether DNA modifications (methylation and other cytosine modifications) could account for this within cell-type transcriptional gradient. In enterocytes selectively isolated from the villi of each of the nine intestinal segments, we characterized 7,580 cytosines (609 CpG sites & 6,971 CpH sites) along the *Lct* and its neighbouring gene, *Mcm6*. We included *Mcm6* in our investigation because DNA variation in this gene, particularly at *MCM6* intron 13, affects inter-individual differences in lactase levels in humans (Enattah et al., 2008; Enattah et al., 2002; Tishkoff et al., 2007). We identified several clusters of modified cytosines (3 or more in <500bp) which showed significant inverse correlations with *Lct* mRNA levels ($R^2 = 0.62\text{-}0.85$; $p < 0.01$ after Bonferroni correction; Figure 5-3c). Most notably, at *Mcm6* exon 13–intron 13, we observed a large cluster modified cytosines associated with *Lct* mRNA variation across the intestine (18 CpGs and 1 CpH site, $p = 3.3\text{x}10^{-3}$ to $1.4\text{x}10^{-11}$ after Bonferroni correction; $R^2 = 0.67\text{-}0.85$). DNA modification densities in this region were lowest in segment 3 (31%) and reached maximal levels in the distal ileum (CpH modifications in Figure 5-4, all DNA modification Figure 5-5). There were also CpG clusters strongly correlated with *Lct* mRNA at *Lct* exon 1 and *Lct* intron 2 (each genomic site containing 6 CpGs, $R^2 = 0.62\text{-}0.85$; $p = 1.8\text{x}10^{-2}$ to $4.0\text{x}10^{-12}$ after Bonferroni correction). Thus, DNA modifications at specific regions are related to the *Lct* transcriptional gradient within enterocytes along the small intestine.

We next examined ENCODE data of adult mouse small intestine (Consortium, 2012) to determine whether the sites we associated with transcriptional variation in *Lct* could function as genomic regulatory elements (Figure 5-3d). We found histone marks indicative of an enhancer element at *Mcm6* exon 13–intron 13 and *Lct* intron 2 (H3K4me1), and a promoter at *Lct* exon 1 (H3K4me3). *Mcm6* exon 13–intron 13, *Lct* exon 1 and intron 2 showed binding of CTCF, a protein that facilitates interactions between transcription regulatory sequences by affecting chromatin architecture (Guo et al., 2015; Ong & Corces, 2014). In addition, *Lct* exon 1 and intron 2 overlapped GATA1 binding sites. GATA1 is a member of GATA transcription factor family that modulates *Lct* expression (van Wering et al., 2004), and modifies epigenetic marks to promote subregion transcriptomic differences in the small intestine (Aronson et al., 2014). Thus, sites in which DNA modifications were significantly predictive of regional differences in *Lct* contained chromatin signatures of enhancers and a promoter, along with an enrichment in transcriptional regulatory proteins.

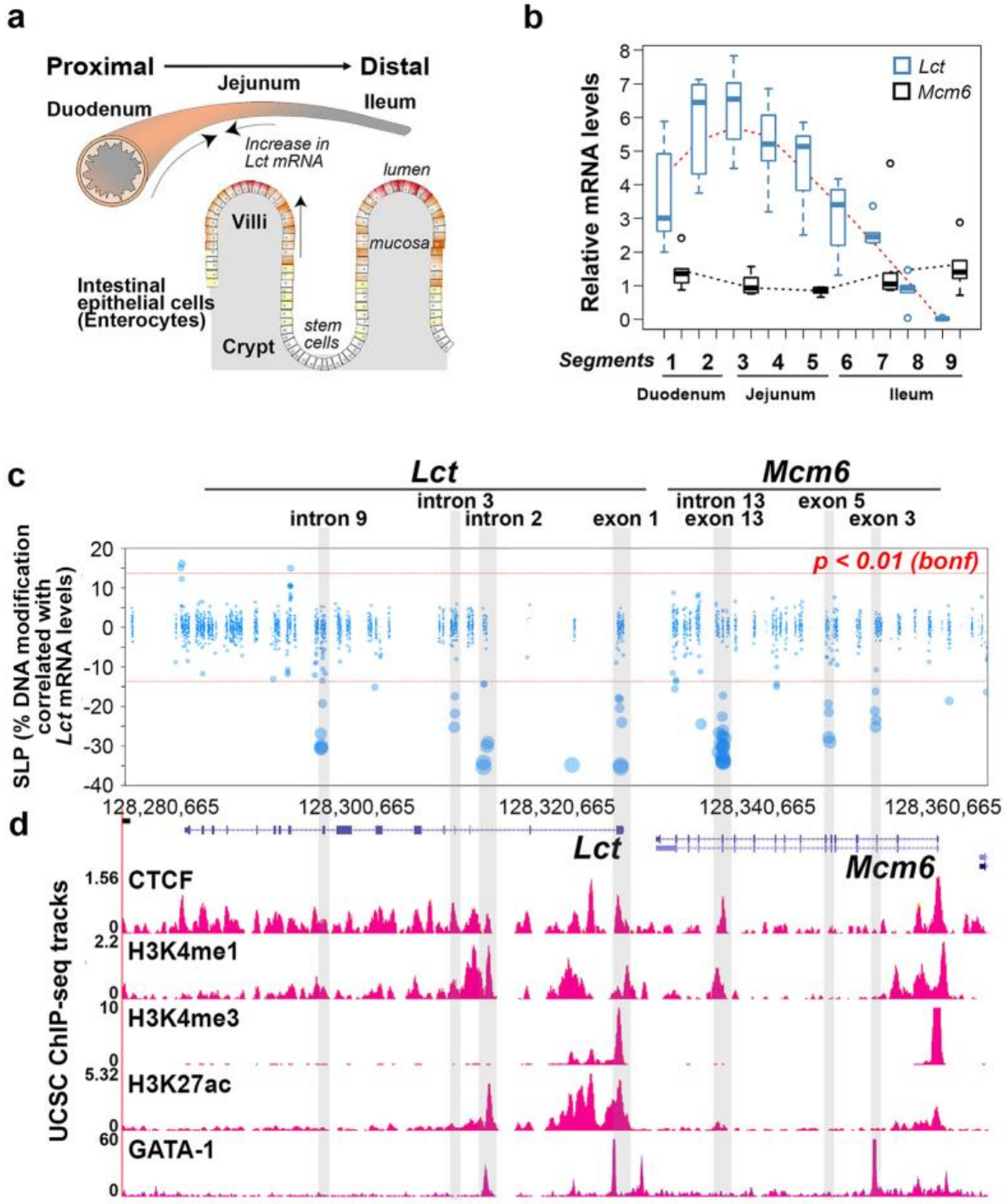



**Figure 5 - 3. Within cell-type variation in Lct mRNA along the proximal-to-distal axis of small intestine is correlated to DNA modification differences at genomic regulatory elements**

(**a**) Illustration of the proximal-to-distal axis of the small intestine and villi. (**b**) There is a segment-specific gradient in steady-state *Lct* mRNA levels. There is no transcriptional gradient in *Mcm6* along the intestine. Villin-1 was used as an endogenous control. Data are represented as

mean ± S.E.M. n = 5–7 mice. $^{**}p < 0.01$, $^{***}p < 10^{-3}$ by Tukey's HSD post hoc test in comparison to segment 3. (**c**) Enterocyte-specific DNA modification clusters (>3 cytosine modifications) within the *Lct–Mcm6* locus significantly correlated with *Lct* mRNA levels (n = 53; 9 segments × 5–6 mice per segment). Bonferroni correction for multiple testing ($p < 0.01$) is indicated by dash lines. SLP refers to the -log *p*-value of the Pearson's correlation $R^2$ (negative values signify an inverse correlation). (**d**) DNA modification clusters associated with the *Lct* transcriptional gradient (highlighted by grey vertical bars) overlap histone marks characteristic of promoters and enhancers, as well as the transcription factor GATA1 and the chromatin architectural protein CTCF (ChIP-seq data from ENCODE, n = 2).

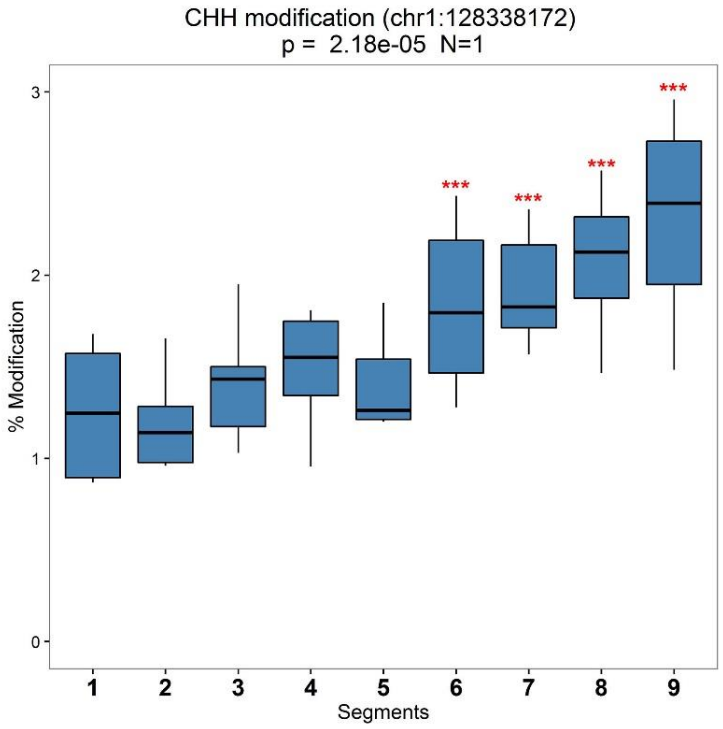



**Figure 5 - 4. CpH modification status along the mouse small intestine**

Box and whisker plots representing % CpH modification at chr1:128338172 along the length of the intestine (box represents the interquartile range; whiskers represent upper and lower extremes). Intestinal segmental effect (One-way ANOVA $p = 2.18 \times 10^{-5}$) was found to be significant. $^{***}p < 0.001$ by Tukey's HSD post hoc test in comparison to segment 1.

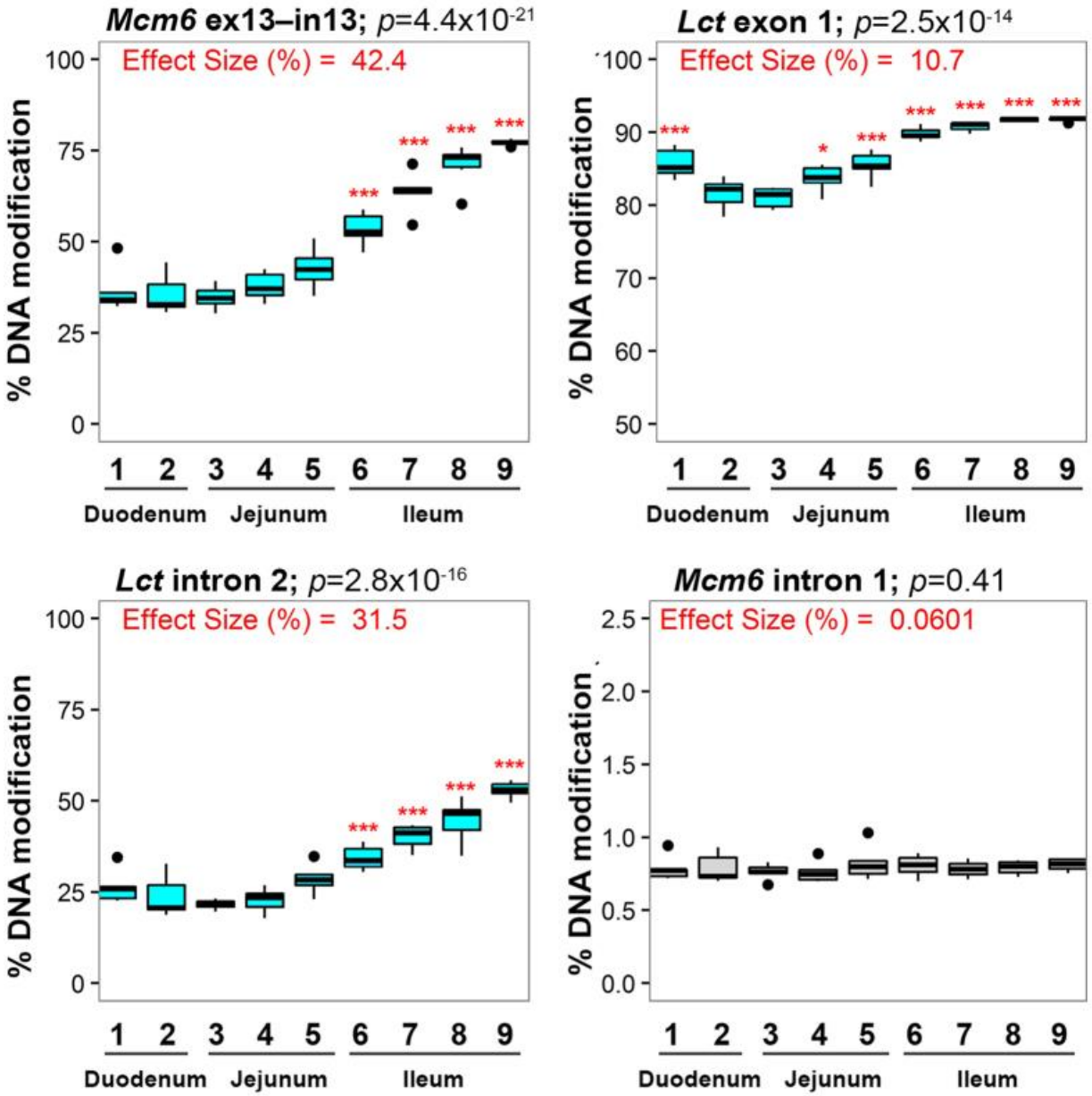



**Figure 5 - 5. Average % DNA modification (y-axis) at Mcm6 exon 13-intron 13, Lct intron 2 and Lct exon 1 is significantly associated with Lct transcriptional variation along the proximal-to-distal axis of the intestine (x-axis)**

*Mcm6* exon 13-intron 13, *Lct* intron 2 and *Lct* exon 1 regions selected based on findings of Figure 5-3c. Average % DNA modification of *Mcm6* intron 1 included as a negative control. *P*-value obtained by one-way ANOVA (n = 53; 9 segments × 5–6 mice per segment). **p* < 0.05, ****p* < $10^{-3}$ by Tukey's HSD post hoc test in comparison to segment 3.

### 5.4.2 The *Lct* transcriptional epigenetic gradients in enterocytes are age-dependent

We investigated whether transcriptional gradients could be established from age-associated epigenetic changes. To do this, we first examined *Lct* mRNA levels in enterocytes across the small intestine of infant mice (P6) and adult mice (P60). *Lct* mRNA levels in infants were mostly similar across intestinal segments, and were 5–15 fold higher than adults (main effect of age: $F_{8,85} = 13.23$, $p = 3.12\text{x}10^{-12}$; Figure 5-6a). This signifies that transcriptional differences in *Lct* along the small intestine manifests with age, where over time *Lct* becomes mainly suppressed in enterocytes of the proximal duodenum and ileum relative to medial segments.

Next, the extent to which DNA modifications diverge with age at the *Lct-Mcm6* locus was examined in villi enterocytes from different intestinal regions. In adults, segments 1, 3 and 7 (duodenal, jejunal and ileal segments) differed substantially in *Lct* abundance, and consistently, DNA modifications differed significantly between adult segments ($p < 0.01$ after Bonferroni correction; Figure 5-6b, x-axis). In infants, no significant DNA modification differences were observed between the intestinal segments (Figure 5-6b, y-axis), which reflects their lack of an *Lct* transcriptional gradient.

We then searched for epigenetically-controlled DNA sites that could contribute to the age-dependent establishment of the *Lct* transcriptional gradient. For this, we investigated *Lct–Mcm6* modifications in isolated villi enterocytes of adult and infant mice, comparing only matched segments between the age groups (Figure 5-6c). We found that *Mcm6* exon 13–intron 13 exhibited the highest localized gain of DNA modifications in adulthood (Figure 5-6c). The age-related DNA modification increase at *Mcm6* exon 13–intron 13 was most prevalent in the segment with the largest *Lct* mRNA loss (segment 7; age-segment interaction: $F_{7,78} = 5.78$, $p = 4.010^{-4}$; Figure 5-6d). DNA modification densities in *Mcm6* exon 13–intron 13 increased by 30% in segment 7 in P60 adults compared to infants ($p = 1.3\text{x}10^{-4}$), while there was an 18% increase segment 1 ($p = 1.3\text{x}10^{-4}$) and a 12% increase in segment 3 ($p = 0.0013$; Figure 5-6d). In older adults (P90), DNA modification levels continued to rise (by ~10%) in segment 1 and 3 ($p = 0.025$; Figure 5-6d), indicating continued epigenetic aging at this *Mcm6* site.

Thus, in addition to enabling lactase persistence in certain human populations (Enattah et al., 2002; Tishkoff et al., 2007), epigenetic regulation of *Mcm6* exon 13–intron 13 could be

central to the age-dependent decline of *Lct* in proximal and distal portions of the small intestine, leading to their inability to metabolize lactose. Thus, cells of the same type can exhibit a divergence in DNA modifications with age across tissue regions, which in turn may facilitate transcriptional gradients and the functional specialization of tissue subregions.

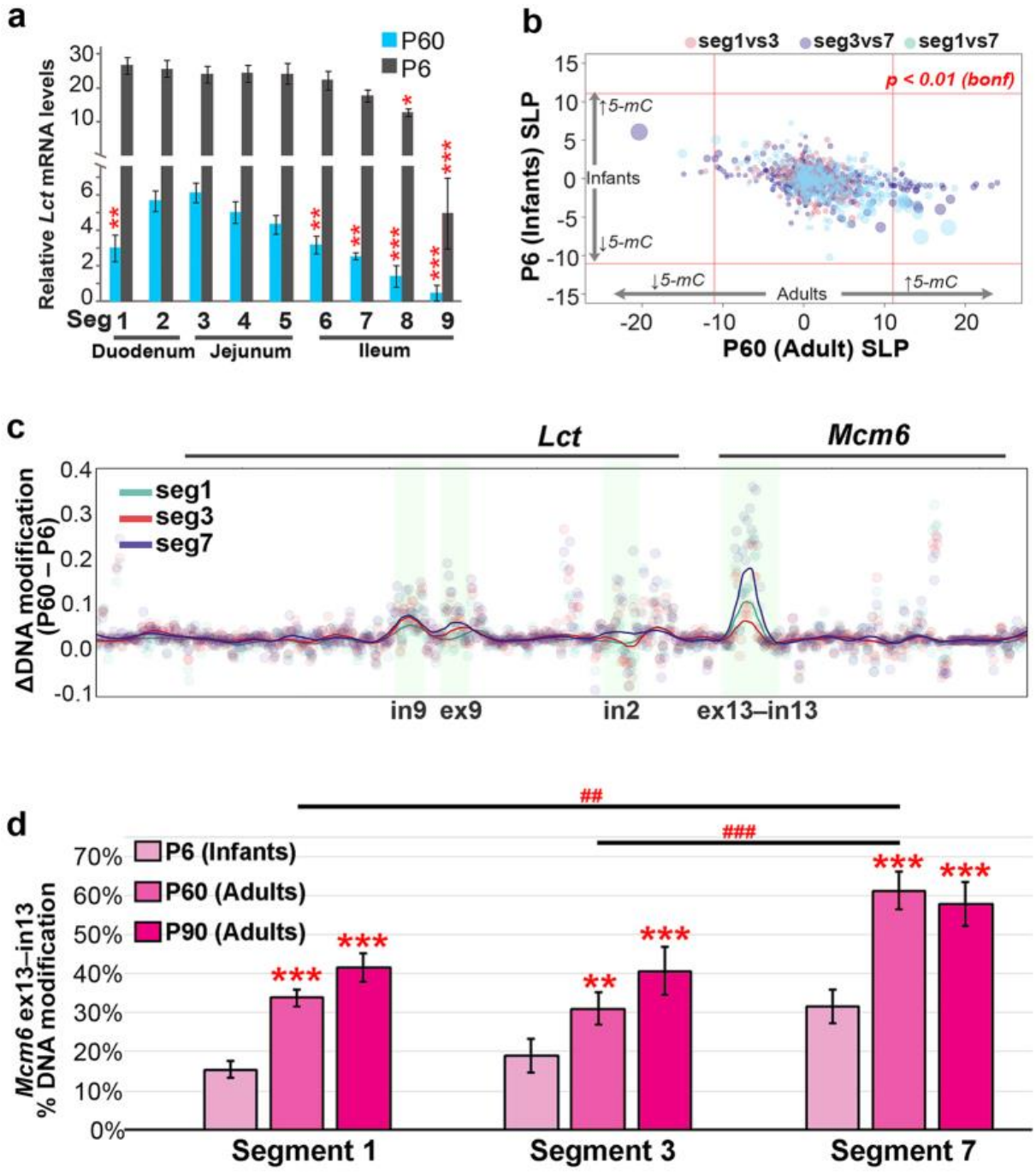



**Figure 5 - 6. Regional differences in Lct transcription arises with age and the divergence of DNA modification patterns within enterocytes**

(**a**) Infant mice (P6; n = 6) had higher *Lct* mRNA levels than adults (P60; n = 4–6), and had an attenuated transcriptional gradient. Data are represented as mean ± S.E.M. *$p < 0.05$, **$p < 0.01$, ***$p < 10^{-3}$ by Tukey's HSD post hoc test compared to age-matched segment 3. (**b**) Comparison of individual DNA modifications between intestinal segments (within age group) revealed a significant divergence in DNA modifications in villi enterocytes along the intestine of adults (x-axis), but not infants (y-axis). SLP refers to signed log *p*-value. Significant comparisons exceed the dashed red line ($p < 0.01$ after Bonferroni correction). (**c**) Identification of genomic sites exhibiting differential DNA modification with age (each dot represents a single CpG site). The *Mcm6* exon 13–intron 13 site in segment 7 (blue) displayed the largest increase in % DNA modification with age, with segment 1 (red) and segment 3 (turquoise) showing smaller DNA modification increases. There was a smaller gain in DNA modifications at *Lct* exon 9–intron 9 in adults. Modification density differences y-axis; average % DNA modification in adults minus in infants by individual CpGs (n = 632) fitted with LOESS curve for each segment. (**d**) Average % DNA modification densities at *Mcm6* exon 13–intron 13 site (Data are represented as mean ± S.E.M.; n = 14 CpGs) in enterocytes along the intestine of infants at P6 and adults at P60 and P90. DNA modification levels increased with age at *Mcm6* exon 13–intron 13, particularly in segment 7 enterocytes. Data are represented as mean ± S.E.M. **$p < 0.01$, ***$p < 10^{-3}$ by Tukey's HSD post hoc test compared to same segment of infant mice. $^{\#\#}p < 0.01$, $^{\#\#\#}p < 10^{-3}$ by Tukey's HSD post hoc test for between segments of P60 adults. (**b**–**d**) n = 3–7 per age group for each segment.

### 5.4.3 The environment modifies the *Lct* transcriptional gradient and induces epigenetic alterations

We determined whether a phenotype-related environmental signal (i.e. lactose exposure) could modify transcriptional patterns and induce epigenetic changes in enterocytes along the small intestine. For this experiment, we fed mice either lactose-containing milk (LAC+) or lactose-free milk (lac-) for 60 days. In response to lactose treatment, *Lct* mRNA levels increased in the enterocytes of distal intestinal segments (segment-diet interaction $F_{8,88} = 5.2$, $p = 2.0 \times 10^{-5}$; Figure 5-7a). Specifically, lactose exposure increased *Lct* mRNA levels in enterocytes of segments 7 ($p = 0.007$) and 8 by 1.3- to 2- fold ($p = 0.023$) compared to mice in the lactose-free group (Figure 5-7). To test whether the epigenome could be involved in the environmentally-induced changes in *Lct* mRNA across the intestine, we investigated DNA modifications in isolated villi enterocytes at the *Lct–Mcm6* locus in LAC+ and lac- mice. We evaluated intestinal segments in which *Lct* was induced (segments 7 and 8) or showed no statistically significant change (segments 1, 3 and 4). Distal intestinal exhibited a significant decrease in DNA modification densities at *Lct* exon 12 (segment 7 = 5.2%, $p = 4.7 \times 10^{-4}$; segment 8 = 4.7%, $p = 0.010$), *Lct* intron 8 (segment 8 =3.6%, $p = 3.6 \times 10^{-3}$), and *Lct* intron 2 (segment 8 = 6.4%; $p =$

0.012; Figure 5-7b and c). DNA modifications were not significantly different in segments 1, 3 and 4 Figure 5-7b and c). Although epigenetic modifications at *Mcm6* exon 13–intron 13 were found to be important for aging and regional differences in *Lct*, this genomic site was unchanged by milk-feeding, which could partly explain why *Lct* mRNA levels were not returned to that of pre-weaned infants after prolonged lactose exposure.

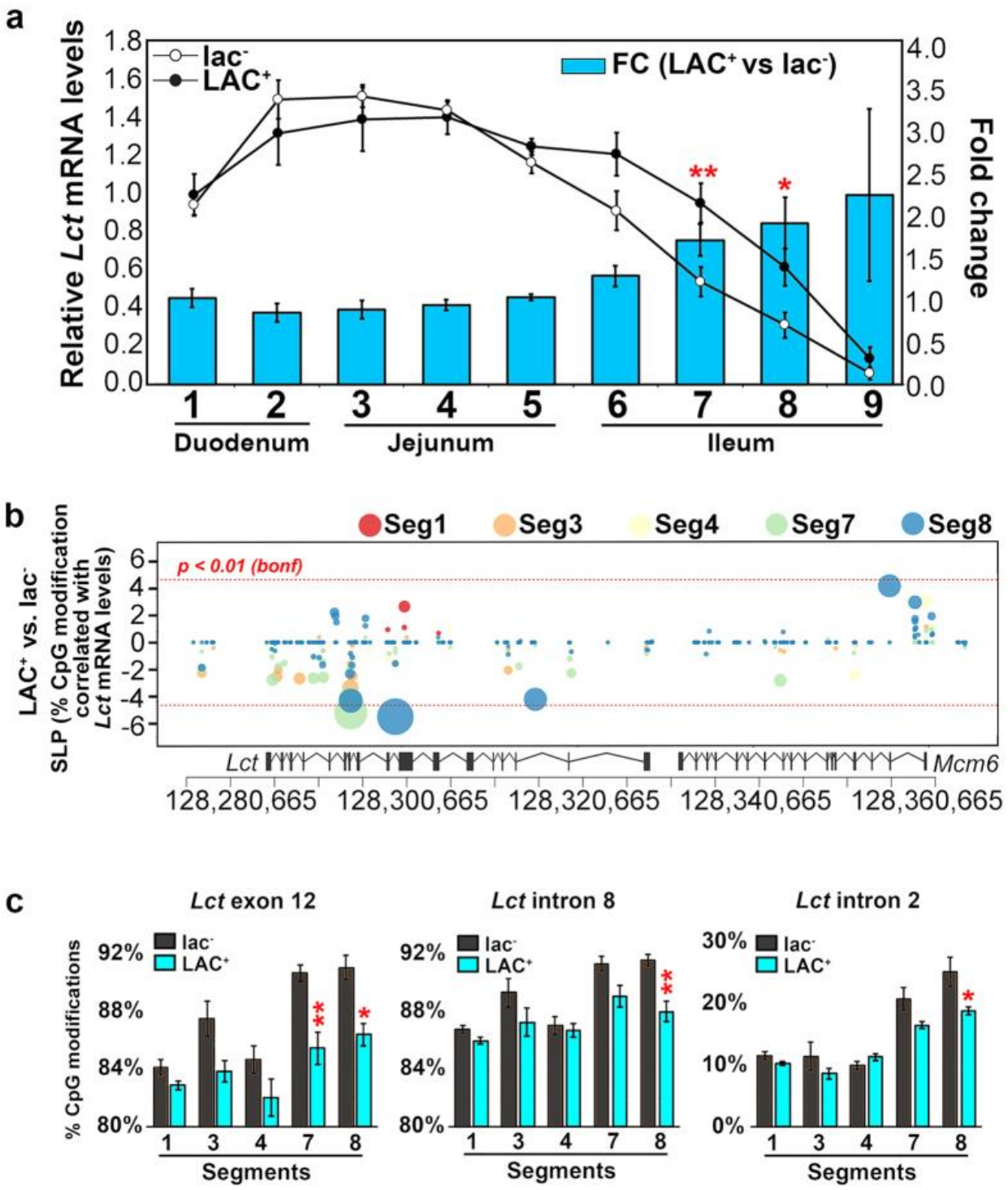



**Figure 5 - 7. Lactose-treatment induces DNA modification changes in intestinal subregions that correspond to changes in the *Lct* transcriptional gradient**

(**a**) Mice given lactose-containing milk ($LAC^+$, n = 6) for 60-days had segment-specific increases in *Lct* mRNA in comparison to mice given lactose-free milk ($lac^-$, n = 7). *Lct* mRNA following lactose exposure was increased in the ileum (seg7–8). $*p < 0.05$, $**p < 0.01$ by one-way ANOVA. (**b**) Changes in CpG modification were detected in intestinal segments that

displayed *Lct* mRNA changes in response to lactose treatment ($p < 0.01$ after Bonferroni correction; dashed red line; $LAC^+$ n = 6–7 mice per segment; $lac^-$ n = 5–7 mice per segment). (**c**) Reductions in CpG modifications occurred in response to lactose treatment at *Lct* exon 12 (Factorial ANOVA; $F_{4,55} = 2.3$, $p = 0.074$), intron 8 ($F_{4,55} = 2.3$, $p = 0.074$) and intron 2 ($F_{4,55} = 3.1$, $p = 0.023$) in distal intestinal segments. Data are represented as average % CpG modification density ± S.E.M. Effect of lactose treatment by intestinal segment was calculated by factorial ANOVA. *$p < 0.05$ **$p < 0.01$, and ***$p < 0.01$ by Tukey's HSD post hoc test for between lactose treatment within intestinal segment.

### 5.4.4 CTCF binding along the intestine targets epigenetically-controlled sites associated with *Lct* transcriptional variation

We wanted to examine the potential mechanism behind how DNA modification status of ~10kb upstream genomic region of the *Lct* gene was able to exert such strong effects on the *Lct* promoter. Based on our histone analysis, we chose to examine the contribution of the chromatin architectural protein CTCF because of its ability to mediate architectural DNA interactions that facilitate transcription (Ghirlando & Felsenfeld, 2016). Furthermore, sequence-dependent CTCF occupancy can be regulated by DNA modification status (Bell & Felsenfeld, 2000; Jones, 2012; Shukla et al., 2011; Stadler et al., 2011; Wang et al., 2012) which leads to differential splicing and gene transcription. The contribution of CTCF binding in the *Lct-Mcm6* locus in mediating transcriptional gradients in cells of the same type across the adult and infant mouse small intestine was investigated by ChIP-qPCR (Figure 5-8). In adults, CTCF binding was highest in the proximal jejunum (segment 3) and gradually declined in more proximal and distal intestinal regions (main effect segment: $F_{4,19} = 5.12$, $p = 0.0083$). Segment differences in CTCF binding were most apparent at *Mcm6* exon 13–intron 13, and to a lesser extent at *Lct* intron 2. In infant mice, on the other hand, we found that CTCF was absent from *Lct–Mcm6* regulatory elements (Figure 5-8).

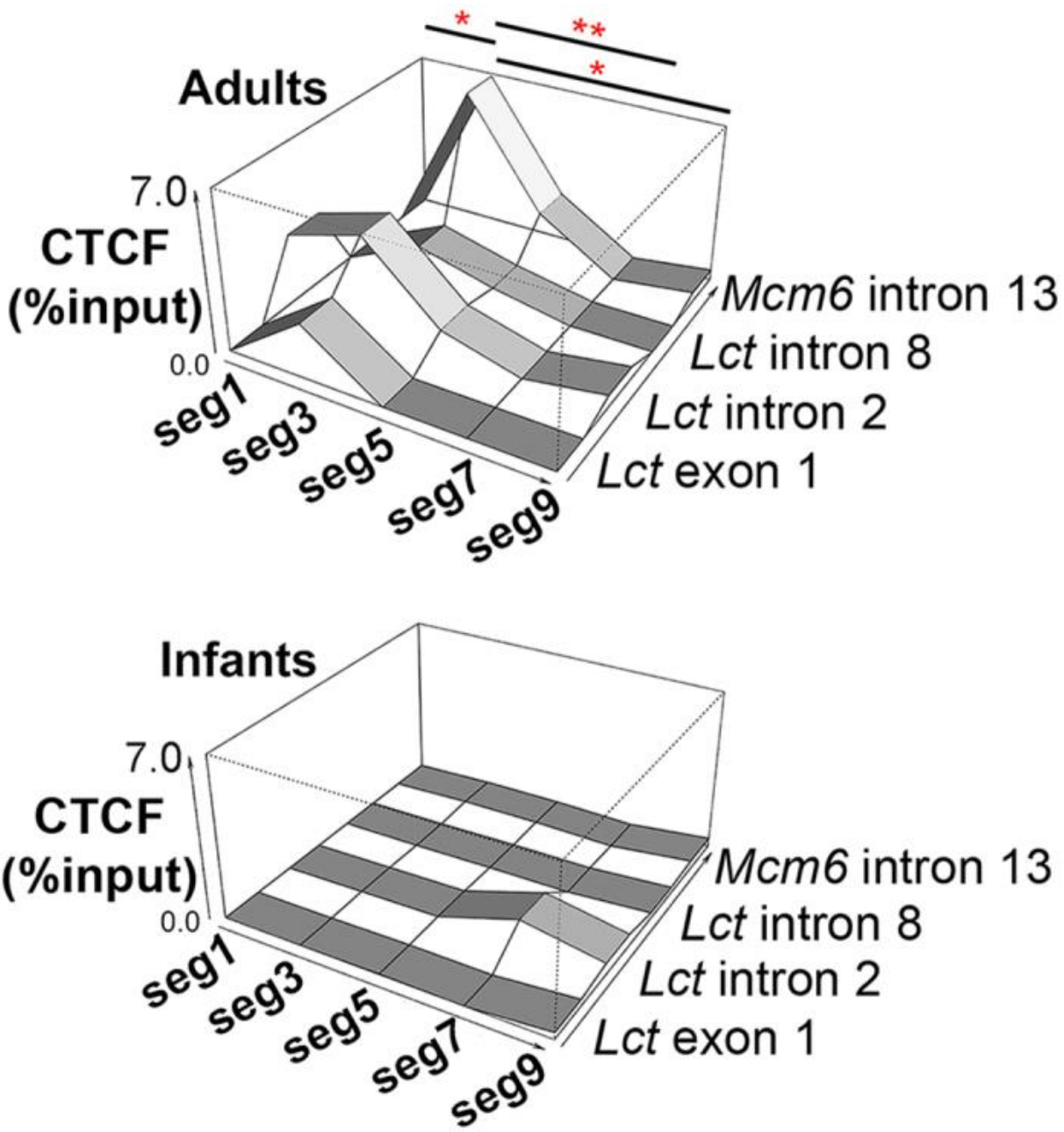



**Figure 5 - 8. CTCF binding at epigenetically-controlled sites in the Lct–Mcm6 locus corresponds to the *Lct* transcriptional gradient**

ChIP-qPCR analysis of CTCF binding at *Mcm6* intron 13, *Lct* intron 8, *Lct* intron 2 and *Lct* exon 1 in adults (P60; top) and infants (P6; bottom) by segment (Data are represented as mean; n = 3 per group). Data presented as percent input after background normalization. *$p < 0.05$ and **$p < 0.01$ by Tukey's HSD post hoc test for one-way ANOVA by effects of intestinal segment.

## 5.5 Discussion and Conclusions

Our results indicate that divergent epigenetic programming enables aging- and environmentally-induced changes in gene transcription occurring in intestinal epithelial cells of the villi (Figure 5-9). In particular, the *Mcm6* exon 13–intron 13 site was found to be a key modulator of the age-dependent establishment and maintenance of the *Lct* transcriptional gradient. These findings are consistent with reports that this upstream region is an evolutionarily conserved enhancer of *Lct* transcription based on genetic (lactase persistence SNPs in various populations) (Enattah et al., 2008; Tishkoff et al., 2007) and molecular evidence that demonstrate its ability to regulate *Lct* transcription *in vitro* (Jensen et al., 2011; Liebert et al., 2016; Troelsen et al., 2003). Our new observation that epigenetic influence of *Mcm6* exon 13–intron 13 is important to *Lct* regulation in both mice and humans (Enattah et al., 2002; Labrie et al., 2016; Tishkoff et al., 2007) adds an epigenetic layer to these previous findings. Furthermore, the *Mcm6* exon 13–intron 13 site was not sensitive to environmental cues (i.e. milk), indicative that age-dependent epigenetic programming of *Mcm6* exon 13–intron 13 is not malleable. By contrast, other regulatory elements affecting the *Lct* gradient (i.e. *Lct* intron 2 and intron 8) remained epigenetically adaptive to environmental signals, enabling a partial recovery of *Lct* expression in the adult intestine after weaning. Therefore, localized DNA modification changes accumulating with age may facilitate the intestinal *Lct* transcriptional gradient, while some remain partially dynamic to environmental signals.

CTCF occupancy was concordant with the *Lct* transcriptional gradient along the small intestine, and inversely correlated to DNA modification profiles at these genomic regulatory elements, particularly at *Mcm6* exon 13-intron 13. CTCF is a multifunctional protein that participates in many epigenetic regulatory functions, including insulation via enhancer blocking, imprinting, X chromosome inactivation, and both transcriptional activation and repression (Kim, Yu, & Kaang, 2015; Wallace & Felsenfeld, 2007).

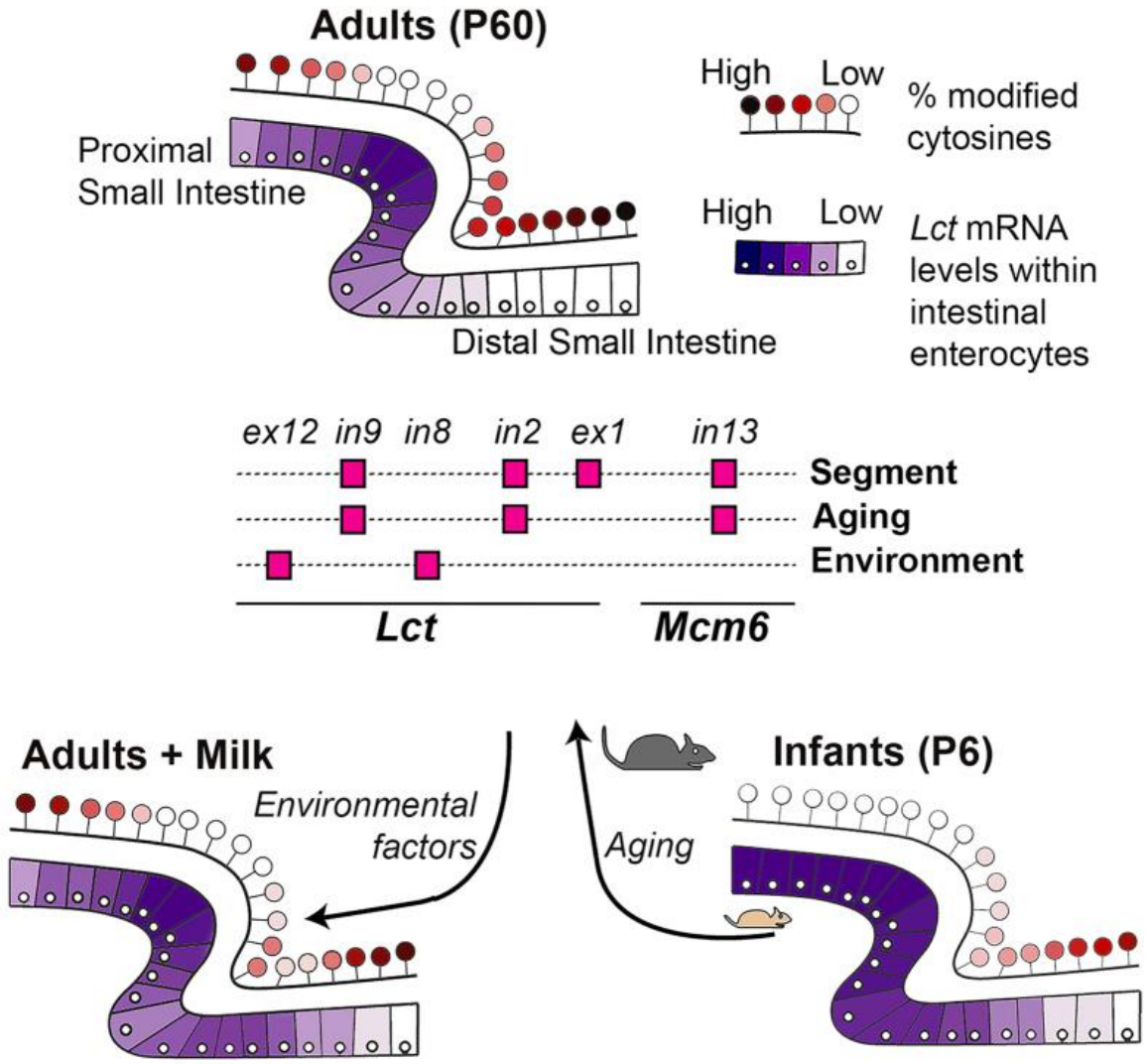



**Figure 5 - 9. Relationship between *Lct* transcriptional gradient and DNA modification status**

In enterocytes along the small intestine, transcriptional variation in the *Lct* gene corresponds to variation in DNA modification profiles at regulatory elements. The *Lct* transcriptional gradient is established with age and is modifiable by environmental signals; changes which co-occur with DNA modification density alterations in enterocytes at the same intestinal regions. The key genomic sites exhibiting DNA modification differences predictive of *Lct* transcriptional heterogeneity across intestinal segments and in response to aging and environmental signals are summarized.

CTCF can also influence DNA modification distribution both locally, through binding to chromatin boundaries, and distally, through effects on chromatin architecture (Hou, Zhao, Tanimoto, & Dean, 2008; Kemp et al., 2014; Wang et al., 2012). Here, our findings indicate that

CTCF binding at the *Lct–Mcm6* locus may function as a region-specific transcriptional activator in adult mice through DNA looping of distal enhancers, such as the *Mcm6* exon 13-intron 13 locus (Guo et al., 2015). Also, CTCF binding may limit the accumulation of cell-type specific DNA modifications with age at middle regions of the intestine relative to distal regions during development, whereas the absence of CTCF could promote downregulation of *Lct* through unobstructed accumulation of DNA modifications and epigenetic silencing. On the other hand, the lack of CTCF at *Lct-Mcm6* sites in infant mice signifies that CTCF binding is not required for the expression of high *Lct* mRNA per say. Rather, CTCF occupancy (and its putative effects on chromatin structure) occurs in tandem with DNA modification changes following weaning, to prevent uniform epigenetic silencing of *Lct* with age. The resulting effect in adulthood is that CTCF and the opposing DNA modification landscape work in concert to facilitate a transcriptional gradient in *Lct* across the intestine.

DNA modifications and CTCF may work in tandem with other transcription factors to create and maintain age-dependent transcriptional gradients in cells of the same type. Transcription factors have been shown to play a role in enterocyte differentiation along the crypt-to-vili axis (i.e., HNF-1a), intestinal subregion specialization (i.e., GATA4) (Aronson et al., 2014; Boudreau et al., 2002; Middendorp et al., 2014; Walker, Thompson, Kohlnhofer, Faber, & Battle, 2014), and interact with enhancer of genes (including *Lct*) which exhibit a transcriptional gradient (R. Fang et al., 2006; Jensen et al., 2011). For example, GATA4 has been shown to be upregulated in the proximal small intestine and downregulated in the distal small intestine to enable the development and divergence of duodenual and ileum portions of the intestine (Bosse et al., 2006; Kohlnhofer, Thompson, Walker, & Battle, 2016). Together, our findings suggest an interplay between molecular and epigenetic factors that facilitate the biological specialization of cell of the same type. The molecular mechanism behind this concerted effort of CTCF, DNA modifications, and transcription factors at enhancer region(s) should be investigated in future studies.

Profiling epigenetic modifications of individual cell types in a tissue region-specific manner could offer insights into tissue specialization. Our findings emphasize the importance for future studies examining epigenetic contributions to the transcriptional variances in multiple genes within a cell type, as these could help explain why tissue subregions can perform diverse biological functions (Anderle et al., 2005; Barth et al., 2005; Scheller et al., 2015) and vary

widely in disease susceptibility and treatment (Dalerba et al., 2011; Patel et al., 2014). In addition, our findings provide an important lesson for epigenetic studies of phenotypes, as failure to consider within cell-type transcriptional variation and epigenetic divergence limits the detection of biologically significant effects. Future epigenetic and gene regulation studies in health and disease will be greatly refined by not only isolating the cell type of interest, but by sampling a single cell type across tissue subregions, and across aging and environmental parameters.

# 6 General Discussion and Future Directions

## 6.1 General Discussion

The efforts outlined in this dissertation highlight the usefulness of incorporating DNA modification analysis into investigations of developmental phenotypes, with the ultimate aim of understanding the latency of complex diseases. In an exploratory study (Chapter 3), we identified differences in DNA modification patterns between individuals with BD when selected by the presence or absence of a history of extreme suicidal behaviour. We also identified SNPs that were nominally significantly-associated with such a history. The case group displayed epigenetic deviations from biological age, which suggested that suicidal behaviour could be linked to epigenetic modifications acting in tandem with genetic polymorphisms/predispositions. Given the limitations of this study (small sample size, small effect size, participant heterogeneity, lack of availability of relevant biological tissue), we chose to investigate these assertions further using the simple trait of lactase persistence/non-persistence (Chapter 4 and 5). Observing the age-related and epigenetically-mediated downregulation of the lactase gene in humans and mice was shown to be an effective preliminary model that minimized some of the variability associated with studying more complex phenotypes. In these studies, we identified subregions within the extended lactase gene locus that, over decades, accrued DNA modifications in a genotype- and diet- dependent fashion. These observations further highlight the potential of epigenetic analysis to annotate DNA polymorphisms with functional and regulatory significance, and underscore the importance of high resolution epigenetic profiling.

Our proof-of-concept investigations represent the first steps in understanding the role of epigenetic factors in lactase gene downregulation. However, further studies are necessary for a complete understanding of age-dependent epigenetic regulation of the lactase gene. First, the focus and design of our study centred around the -13910 C/T SNP which is predominantly associated with the trait in Northern European populations. Thus, our subjects were recruited from individuals of Northern European ancestry to encompass both C- and T- alleles. This approach proved to be advantageous in validating previous studies of this polymorphism, performed primarily using samples from individuals of Northern and Western European descent, and served as an effective preliminary step to investigating the variant-dependent epigenetic trajectory. Nevertheless, future research should involve incorporating polymorphisms from different ethnic populations, particularly those with a history of lactase persistence (Sub-Saharan African, Middle Eastern, South Asian) to allow us to quantify the impact of each marker on the

epigenetic aging process. Second, in these studies we used a chelating method for the enrichment of enterocytes from the villi of the intestinal epithelium (>85% purity as confirmed by flow cytometry). However, the enriched population potentially includes a multitude of other cell types (such as goblet cells, enteroendocrine cells, Paneth cells, adult stem cells, transit-amplifying cells, and other, more rare cell subtypes) which may contribute their own individual and distinct epigenetic factors to our analysis (Lee & Ayres, 2018). Thus, subsequent studies should ideally distinguish and isolate specific cell types and subtypes using FACS-based methods, or on the basis of transcriptional profiles to better parse out the effect of each cell type on observed DNA modification profiles. Third, we were only able to recruit adults, thereby missing the critical period during which lactase non-persistence develops. Our findings were recently replicated in a cross-sectional human study using subjects as young as age 8 (Leseva et al., 2018). Fourth, we used *LCT* mRNA expression level as an endophenotype for determining lactase persistence/non-persistence; a more clinically relevant assessment of this phenotype would involve the measurement of intestinal lactase phlorizin hydrolase enzyme levels (when possible), and lactase breath assay (Newcomer, McGill, Thomas, & Hofmann, 1975). Fifth, our methods did not distinguish between 5mC and other types of DNA modification with potential regulatory effects; interrogating 5hmC modification is feasible in subsequent studies, and may provide peripheral insights.

## 6.2 Future Directions

It is our hope that our experimental designs and resultant observations have laid ground for more comprehensive investigation of the genetic-epigenetic interplay that influences lactase persistence/non-persistence, and exploration of the relevance of such age-related changes in understanding more complex phenotypes, such as suicidality in BD. Toward these proximal and distal goals, we have planned several future experiments to validate the findings presented here.

Given the lack of feasibility of collecting intestinal samples from both children and adults from various non-European populations, we will screen and investigate CRISPR-Cas9 mutants and natural strains of mice that carry a variety of *Lct-Mcm6* polymorphisms. To this end, we have selected nearly a dozen mouse strains that exhibit genetic variation at the *Lct* enhancer, and plan to test these for the phenotype of lactase persistence (identified by *Lct* transcription levels). We will also expand the spectrum of genetic variation at this locus by using CRISPR-Cas9 mutagenesis to induce deletions; moreover, we will attempt to validate the sequence-specificity for lactase persistence phenotype by introducing new SNPs in several enhancer regions across *Lct/Mcm6*. We will select mouse strains with the most distinct age-dependent changes in *Lct* expression, and perform fine mapping of DNA modification in the *Lct-Mcm6* region. Next, we will investigate how DNA polymorphisms at regulatory regions in *Lct* contribute to epigenetic regulation of the gene in short-term (daily) and long-term (developmental/aging) dimensions. In addition to exploring DNA modification at this locus, we will interrogate the transcription factor binding and chromosome conformation profiles along the small intestine as a function of genotype and epigenotype. We hope to subsequently reproduce these findings in three-dimensional cell models of the human small intestine known as organoids, and use these to determine the epigenetic contribution of as many cell types and subtypes of the intestinal epithelium to the phenotypes of lactose intolerance and lactase persistence (Sugimoto & Sato, 2017). Using a variety of experiments, we hope to capture additional nuances of how this phenotype is regulated, and hopefully integrate these insights with accompanying studies of several human populations.

We also plan to validate our case-control genetic and epigenetic findings, and the epigenetic deviations from biological aging we observed in individuals suffering from BD and a history of suicidal behaviour. Our first step would be to attempt to reproduce these findings in a

larger cohort of individuals with a history of suicidal behaviour, controlling for the effects of ethnicity, age, sex, socioeconomic factors, early childhood adversity, medication, treatment; moreover, we hope to test a larger array of subjects, including those with other comorbid disorders (schizophrenia, substance abuse, major depressive disorder, panic disorder and generalized anxiety disorder). If successfully replicated in larger studies, these loci will be targets of interest for future high-resolution profiling of DNA modification in whole blood and *post-mortem* brain. We hope that the insights gained from these follow-up studies may provide additional molecular targets to understand not only how BD unfolds, but also how suicide risk can be quantified on an individual basis.

Looking ahead, the forthcoming decade holds promise for the field of epigenetics in improving the framework by which non-Mendelian traits and disorders are currently understood. A growing body of findings have the potential to begin uncovering epigenetic markers across the lifespan that could provide therapeutic (targeted drug design), diagnostic (identification of at-risk populations, early detection of disease) and prognostic (quantification of disease course, symptomatic load, response to treatment) opportunities for elucidation of many complex disorders, particularly those with a defined temporal component. The success of these studies hinges greatly upon improved technologies to permit more effective epigenetic profiling; these include more cost-effective and portable sequencing, improved separation of cell types, and less invasive biopsy methods. These advances have far-reaching translational potential to clinical outcomes, and could reshape the paradigms used to understand some of the most debilitating human illnesses.

# References

Abi Khalil, C. (2014). The emerging role of epigenetics in cardiovascular disease. *Ther Adv Chronic Dis, 5*(4), 178-187. doi:10.1177/2040622314529325

Acosta, J. R., Joost, S., Karlsson, K., Ehrlund, A., Li, X., Aouadi, M., . . . Laurencikiene, J. (2017). Single cell transcriptomics suggest that human adipocyte progenitor cells constitute a homogeneous cell population. *Stem Cell Res Ther, 8*(1), 250. doi:10.1186/s13287-017-0701-4

Albert, S. M., Michaels, K., Padilla, M., Pelton, G., Bell, K., Marder, K., . . . Devanand, D. P. (1999). Functional significance of mild cognitive impairment in elderly patients without a dementia diagnosis. *Am J Geriatr Psychiatry, 7*(3), 213-220.

Anderle, P., Sengstag, T., Mutch, D. M., Rumbo, M., Praz, V., Mansourian, R., . . . Roberts, M. A. (2005). Changes in the transcriptional profile of transporters in the intestine along the anterior-posterior and crypt-villus axes. *BMC Genomics, 6*, 69. doi:10.1186/1471-2164-6-69

Angermueller, C., Clark, S. J., Lee, H. J., Macaulay, I. C., Teng, M. J., Hu, T. X., . . . Reik, W. (2016). Parallel single-cell sequencing links transcriptional and epigenetic heterogeneity. *Nat Methods, 13*(3), 229-232. doi:10.1038/nmeth.3728

Angst, J., & Cassano, G. (2005). The mood spectrum: improving the diagnosis of bipolar disorder. *Bipolar Disord, 7 Suppl 4*, 4-12. doi:10.1111/j.1399-5618.2005.00210.x

Angulo, M. A., Butler, M. G., & Cataletto, M. E. (2015). Prader-Willi syndrome: a review of clinical, genetic, and endocrine findings. *J Endocrinol Invest, 38*(12), 1249-1263. doi:10.1007/s40618-015-0312-9

Aprahamian, I., Nunes, P. V., & Forlenza, O. V. (2013). Cognitive impairment and dementia in late-life bipolar disorder. *Curr Opin Psychiatry, 26*(1), 120-123. doi:10.1097/YCO.0b013e32835ac5f6

Arancio, W., Pizzolanti, G., Genovese, S. I., Pitrone, M., & Giordano, C. (2014). Epigenetic involvement in Hutchinson-Gilford progeria syndrome: a mini-review. *Gerontology, 60*(3), 197-203. doi:10.1159/000357206

Aronson, B. E., Rabello Aronson, S., Berkhout, R. P., Chavoushi, S. F., He, A., Pu, W. T., . . . Krasinski, S. D. (2014). GATA4 represses an ileal program of gene expression in the proximal small intestine by inhibiting the acetylation of histone H3, lysine 27. *Biochim Biophys Acta, 1839*(11), 1273-1282. doi:10.1016/j.bbagrm.2014.05.018

Avraham, R., Haseley, N., Brown, D., Penaranda, C., Jijon, H. B., Trombetta, J. J., . . . Hung, D. T. (2015). Pathogen Cell-to-Cell Variability Drives Heterogeneity in Host Immune Responses. *Cell, 162*(6), 1309-1321. doi:10.1016/j.cell.2015.08.027

Balázs, J., Benazzi, F., Rihmer, Z., Rihmer, A., Akiskal, K. K., & Akiskal, H. S. (2006). The close link between suicide attempts and mixed (bipolar) depression: Implications for suicide prevention. *Journal of Affective Disorders, 91*(2-3), 133-138. doi:10.1016/j.jad.2005.12.049

Balogh, Z., Benedek, G., & Kéri, S. (2008). Retinal dysfunctions in schizophrenia. *Progress in Neuro-Psychopharmacology and Biological Psychiatry, 32*(1), 297-300. doi:10.1016/j.pnpbp.2007.08.024

Bani-Fatemi, A., Howe, A. S., & De Luca, V. (2015). Epigenetic studies of suicidal behavior. *Neurocase, 21*(2), 134-143. doi:10.1080/13554794.2013.826679

Bannister, A. J., & Kouzarides, T. (2011). Regulation of chromatin by histone modifications. *Cell Res, 21*(3), 381-395. doi:10.1038/cr.2011.22

Baran, Y., Subramaniam, M., Biton, A., Tukiainen, T., Tsang, E. K., Rivas, M. A., . . . Lappalainen, T. (2015). The landscape of genomic imprinting across diverse adult human tissues. *Genome Res, 25*(7), 927-936. doi:10.1101/gr.192278.115

Barfield, R. T., Kilaru, V., Smith, A. K., & Conneely, K. N. (2012). CpGassoc: an R function for analysis of DNA methylation microarray data. *Bioinformatics, 28*(9), 1280-1281. doi:10.1093/bioinformatics/bts124

Barrero, M. J., Boue, S., & Izpisua Belmonte, J. C. (2010). Epigenetic mechanisms that regulate cell identity. *Cell Stem Cell, 7*(5), 565-570. doi:10.1016/j.stem.2010.10.009

Barrett, J. C., Fry, B., Maller, J., & Daly, M. J. (2005). Haploview: analysis and visualization of LD and haplotype maps. *Bioinformatics, 21*(2), 263-265. doi:10.1093/bioinformatics/bth457

Barski, A., Cuddapah, S., Cui, K., Roh, T. Y., Schones, D. E., Wang, Z., . . . Zhao, K. (2007). High-Resolution Profiling of Histone Methylations in the Human Genome. *Cell, 129*(4), 823-837. doi:10.1016/j.cell.2007.05.009

Barth, A. S., Merk, S., Arnoldi, E., Zwermann, L., Kloos, P., Gebauer, M., . . . Nabauer, M. (2005). Functional profiling of human atrial and ventricular gene expression. *Pflugers Arch, 450*(4), 201-208. doi:10.1007/s00424-005-1404-8

Bashtrykov, P., Jankevicius, G., Jurkowska, R. Z., Ragozin, S., & Jeltsch, A. (2014). The UHRF1 protein stimulates the activity and specificity of the maintenance DNA methyltransferase DNMT1 by an allosteric mechanism. *J Biol Chem, 289*(7), 4106-4115. doi:10.1074/jbc.M113.528893

Beckmann, H., & Franzek, E. (2000). The genetic heterogeneity of "schizophrenia". *World J Biol Psychiatry, 1*(1), 35-41.

Beguelin, W., Popovic, R., Teater, M., Jiang, Y., Bunting, K. L., Rosen, M., . . . Melnick, A. M. (2013). EZH2 is required for germinal center formation and somatic EZH2 mutations

promote lymphoid transformation. *Cancer Cell, 23*(5), 677-692. doi:10.1016/j.ccr.2013.04.011

Bell, A. C., & Felsenfeld, G. (2000). Methylation of a CTCF-dependent boundary controls imprinted expression of the Igf2 gene. *Nature, 405*(6785), 482-485. doi:10.1038/35013100

Bell, A. C., West, A. G., & Felsenfeld, G. (1999). The protein CTCF is required for the enhancer blocking activity of vertebrate insulators. *Cell, 98*(3), 387-396.

Bell, J. T., Tsai, P. C., Yang, T. P., Pidsley, R., Nisbet, J., Glass, D., . . . Deloukas, P. (2012). Epigenome-wide scans identify differentially methylated regions for age and age-related phenotypes in a healthy ageing population. *PLoS Genet, 8*(4), e1002629. doi:10.1371/journal.pgen.1002629

Belsky, D. W., Sears, M. R., Hancox, R. J., Harrington, H., Houts, R., Moffitt, T. E., . . . Caspi, A. (2013). Polygenic risk and the development and course of asthma: an analysis of data from a four-decade longitudinal study. *Lancet Respir Med, 1*(6), 453-461. doi:10.1016/S2213-2600(13)70101-2

Benayoun, B. A., Pollina, E. A., & Brunet, A. (2015). Epigenetic regulation of ageing: linking environmental inputs to genomic stability. *Nat Rev Mol Cell Biol, 16*(10), 593-610. doi:10.1038/nrm4048

Benjamini, Y., & Hochberg, Y. (1995). Benjamini Y, Hochberg Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society: Series B, 57*(1), 289-300. doi:10.2307/2346101

Berke, I. C., Boehmer, T., Blobel, G., & Schwartz, T. U. (2004). Structural and functional analysis of Nup133 domains reveals modular building blocks of the nuclear pore complex. *Journal of Cell Biology, 167*(4), 591-597. doi:10.1083/jcb.200408109

Bernstein, B. E., Stamatoyannopoulos, J. A., Costello, J. F., Ren, B., Milosavljevic, A., Meissner, A., . . . Thomson, J. A. (2010). The NIH Roadmap Epigenomics Mapping Consortium. *Nat Biotechnol, 28*(10), 1045-1048. doi:10.1038/nbt1010-1045

Bersaglieri, T., Sabeti, P. C., Patterson, N., Vanderploeg, T., Schaffner, S. F., Drake, J. A., . . . Hirschhorn, J. N. (2004). Genetic signatures of strong recent positive selection at the lactase gene. *Am J Hum Genet, 74*(6), 1111-1120. doi:10.1086/421051

Bertelsen, a., Harvald, B., & Hauge, M. (1977). A Danish twin study of manic-depressive disorders. *The British Journal of Psychiatry, 130*, 330-351. doi:10.1192/bjp.130.4.330

Bianco, T., Chenevix-Trench, G., Walsh, D. C., Cooper, J. E., & Dobrovic, A. (2000). Tumour-specific distribution of BRCA1 promoter region methylation supports a pathogenetic role in breast and ovarian cancer. *Carcinogenesis, 21*(2), 147-151.

Bird, L. M. (2014). Angelman syndrome: review of clinical and molecular aspects. *Appl Clin Genet, 7*, 93-104. doi:10.2147/tacg.s57386

Blumenthal, H. T. (2003). The aging-disease dichotomy: true or false? *J Gerontol A Biol Sci Med Sci, 58*(2), 138-145.

Boomsma, D., Busjahn, A., & Peltonen, L. (2002). Classical twin studies and beyond. *Nat Rev Genet, 3*(11), 872-882. doi:10.1038/nrg932

Bordone, L., & Guarente, L. (2005). Calorie restriction, SIRT1 and metabolism: understanding longevity. *Nat Rev Mol Cell Biol, 6*(4), 298-305. doi:10.1038/nrm1616

Bosse, T., Piaseckyj, C. M., Burghard, E., Fialkovich, J. J., Rajagopal, S., Pu, W. T., & Krasinski, S. D. (2006). Gata4 is essential for the maintenance of jejunal-ileal identities in the adult mouse small intestine. *Mol Cell Biol, 26*(23), 9060-9070. doi:10.1128/MCB.00124-06

Boudreau, F., Rings, E. H., van Wering, H. M., Kim, R. K., Swain, G. P., Krasinski, S. D., . . . Traber, P. G. (2002). Hepatocyte nuclear factor-1 alpha, GATA-4, and caudal related homeodomain protein Cdx2 interact functionally to modulate intestinal gene transcription. Implication for the developmental regulation of the sucrase-isomaltase gene. *J Biol Chem, 277*(35), 31909-31917. doi:10.1074/jbc.M204622200

Bourc'his, D., Xu, G. L., Lin, C. S., Bollman, B., & Bestor, T. H. (2001). Dnmt3L and the establishment of maternal genomic imprints. *Science, 294*(5551), 2536-2539. doi:10.1126/science.1065848

Brandt, A., & Vilcinskas, A. (2013). The Fruit Fly Drosophila melanogaster as a Model for Aging Research. *Adv Biochem Eng Biotechnol, 135*, 63-77. doi:10.1007/10_2013_193

Brent, D. A., Oquendo, M., Birmaher, B., Greenhill, L., Kolko, D., Stanley, B., . . . Mann, J. J. (2004). Familial transmission of mood disorders: convergence and divergence with transmission of suicidal behavior. *J Am Acad Child Adolesc Psychiatry, 43*(10), 1259-1266. doi:10.1097/01.chi.0000135619.38392.78

Breton, C. V., Byun, H. M., Wenten, M., Pan, F., Yang, A., & Gilliland, F. D. (2009). Prenatal tobacco smoke exposure affects global and gene-specific DNA methylation. *Am J Respir Crit Care Med, 180*(5), 462-467. doi:10.1164/rccm.200901-0135OC

Breton, C. V., Salam, M. T., & Gilliland, F. D. (2011). Heritability and role for the environment in DNA methylation in AXL receptor tyrosine kinase. *Epigenetics, 6*(7), 895-898. doi:10.4161/epi.6.7.15768

Brietzke, E., Stertz, L., Fernandes, B. S., Kauer-Sant'anna, M., Mascarenhas, M., Escosteguy Vargas, A., . . . Kapczinski, F. (2009). Comparison of cytokine levels in depressed, manic and euthymic patients with bipolar disorder. *J Affect Disord, 116*(3), 214-217. doi:10.1016/j.jad.2008.12.001

Bromer, J. G., Zhou, Y., Taylor, M. B., Doherty, L., & Taylor, H. S. (2010). Bisphenol-A exposure in utero leads to epigenetic alterations in the developmental programming of uterine estrogen response. *Faseb j, 24*(7), 2273-2280. doi:10.1096/fj.09-140533

Brown, G. K., Beck, A. T., Steer, R. A., & Grisham, J. R. (2000). Risk factors for suicide in psychiatric outpatients: a 20-year prospective study. *J Consult Clin Psychol, 68*(3), 371-377.

Brunet, A., & Berger, S. L. (2014). Epigenetics of aging and aging-related disease. *J Gerontol A Biol Sci Med Sci, 69 Suppl 1*, S17-20. doi:10.1093/gerona/glu042

Bulterijs, S., Hull, R. S., Björk, V. C. E., & Roy, A. G. (2015). It is time to classify biological aging as a disease. *Front Genet, 6*. doi:10.3389/fgene.2015.00205

Bunney, W. E., & Bunney, B. G. (2000). Molecular clock genes in man and lower animals: Possible implications for circadian abnormalities in depression (Vol. 22, pp. 335-345).

Burtner, C. R., & Kennedy, B. K. (2010). Progeria syndromes and ageing: what is the connection? *Nat Rev Mol Cell Biol, 11*(8), 567-578. doi:10.1038/nrm2944

Buske, O. J. (2015). Padlock-Pipeline *GitHub*.

Cabili, M. N., Trapnell, C., Goff, L., Koziol, M., Tazon-Vega, B., Regev, A., & Rinn, J. L. (2011). Integrative annotation of human large intergenic noncoding RNAs reveals global properties and specific subclasses. *Genes Dev, 25*(18), 1915-1927. doi:10.1101/gad.17446611

Callahan, D., & Topinkova, E. (1998). Is aging a preventable or curable disease? *Drugs Aging, 13*(2), 93-97.

Capuron, L., Schroecksnadel, S., Feart, C., Aubert, A., Higueret, D., Barberger-Gateau, P., . . . Fuchs, D. (2011). Chronic low-grade inflammation in elderly persons is associated with altered tryptophan and tyrosine metabolism: role in neuropsychiatric symptoms. *Biol Psychiatry, 70*(2), 175-182. doi:10.1016/j.biopsych.2010.12.006

Carone, B. R., Fauquier, L., Habib, N., Shea, J. M., Hart, C. E., Li, R., . . . Rando, O. J. (2010). Paternally-induced transgenerational environmental reprogramming of metabolic gene expression in mammals. *Cell, 143*(7), 1084-1096. doi:10.1016/j.cell.2010.12.008

Carrard, A., Salzmann, A., Malafosse, A., & Karege, F. (2011). Increased DNA methylation status of the serotonin receptor 5HTR1A gene promoter in schizophrenia and bipolar disorder. *Journal of Affective Disorders, 132*(3), 450-453. doi:10.1016/j.jad.2011.03.018

Chalei, V., Sansom, S. N., Kong, L., Lee, S., Montiel, J. F., Vance, K. W., & Ponting, C. P. (2014). The long non-coding RNA Dali is an epigenetic regulator of neural differentiation. *Elife, 3*, e04530. doi:10.7554/eLife.04530

Chambers, J. C., Loh, M., Lehne, B., Drong, A., Kriebel, J., Motta, V., . . . Kooner, J. S. (2015). Epigenome-wide association of DNA methylation markers in peripheral blood from Indian Asians and Europeans with incident type 2 diabetes: a nested case-control study. *Lancet Diabetes Endocrinol, 3*(7), 526-534. doi:10.1016/s2213-8587(15)00127-8

Chen, G., Henter, I. D., & Manji, H. K. (2010). Presynaptic glutamatergic dysfunction in bipolar disorder. *Biol Psychiatry, 67*(11), 1007-1009. doi:10.1016/j.biopsych.2010.03.027

Chen, Z., Miao, F., Paterson, A. D., Lachin, J. M., Zhang, L., Schones, D. E., . . . Natarajan, R. (2016). Epigenomic profiling reveals an association between persistence of DNA methylation and metabolic memory in the DCCT/EDIC type 1 diabetes cohort. *Proc Natl Acad Sci U S A, 113*(21), E3002-3011. doi:10.1073/pnas.1603712113

Cheng, X., & Blumenthal, R. M. (2008). Mammalian DNA methyltransferases: a structural perspective. *Structure, 16*(3), 341-350. doi:10.1016/j.str.2008.01.004

Chiang, P. K., Gordon, R. K., Tal, J., Zeng, G. C., Doctor, B. P., Pardhasaradhi, K., & McCann, P. P. (1996). S-Adenosylmethionine and methylation. *Faseb j, 10*(4), 471-480.

Cittaro, D., Lampis, V., Luchetti, A., Coccurello, R., Guffanti, A., Felsani, A., . . . Battaglia, M. (2016). Histone Modifications in a Mouse Model of Early Adversities and Panic Disorder : Role for Asic1 and Neurodevelopmental Genes. *Sci Rep, 6*(April), 1-10. doi:10.1038/srep25131

Coarfa, C., Pichot, C. S., Jackson, A., Tandon, A., Amin, V., Raghuraman, S., . . . Milosavljevic, A. (2014). Analysis of interactions between the epigenome and structural mutability of the genome using Genboree workbench tools. *BMC Bioinformatics, 15*(Suppl 7), 1-12. doi:10.1186/1471-2105-15-S7-S2

Cohen, A. L., & Jia, S. (2014). Noncoding RNAs and the borders of heterochromatin. *Wiley Interdiscip Rev RNA, 5*(6), 835-847. doi:10.1002/wrna.1249

Comet, I., Riising, E. M., Leblanc, B., & Helin, K. (2016). Maintaining cell identity: PRC2-mediated regulation of transcription and cancer. *Nat Rev Cancer, 16*(12), 803-810. doi:10.1038/nrc.2016.83

Condon, J., Shaw, J. E., Luciano, M., Kyvik, K. O., Martin, N. G., & Duffy, D. L. (2008). A study of diabetes mellitus within a large sample of Australian twins. *Twin Res Hum Genet, 11*(1), 28-40. doi:10.1375/twin.11.1.28

Connolly, K. R., & Thase, M. E. (2011). The clinical management of bipolar disorder: a review of evidence-based guidelines. *Prim Care Companion CNS Disord, 13*(4). doi:10.4088/PCC.10r01097

Consortium, E. P. (2012). An integrated encyclopedia of DNA elements in the human genome. *Nature, 489*(7414), 57-74. doi:10.1038/nature11247

Coryell, W., Solomon, D., Turvey, C., Keller, M., Leon, A. C., Endicott, J., . . . Mueller, T. (2003). The long-term course of rapid-cycling bipolar disorder. *Arch Gen Psychiatry, 60*(9), 914-920. doi:10.1001/archpsyc.60.9.914

Cournil, A., & Kirkwood, T. B. (2001). If you would live long, choose your parents well. *Trends Genet, 17*(5), 233-235.

Craddock, N., & Jones, I. (1999). Genetics of bipolar disorder. *J Med Genet, 36*(8), 585-594.

Craddock, N., O'Donovan, M. C., & Owen, M. J. (2005). The genetics of schizophrenia and bipolar disorder: dissecting psychosis. *J Med Genet, 42*(3), 193-204. doi:10.1136/jmg.2005.030718

Creyghton, M. P., Cheng, A. W., Welstead, G. G., Kooistra, T., Carey, B. W., Steine, E. J., . . . Jaenisch, R. (2010). Histone H3K27ac separates active from poised enhancers and predicts developmental state. *Proceedings of the National Academy of Sciences, 107*(50), 21931-21936. doi:10.1073/pnas.1016071107

Criscione, S. W., Zhang, Y., Thompson, W., Sedivy, J. M., & Neretti, N. (2014). Transcriptional landscape of repetitive elements in normal and cancer human cells. *BMC Genomics, 15*, 583. doi:10.1186/1471-2164-15-583

Cruceanu, C., Alda, M., Nagy, C., Freemantle, E., Rouleau, G. A., & Turecki, G. (2013). H3K4 tri-methylation in synapsin genes leads to different expression patterns in bipolar disorder and major depression. *Int J Neuropsychopharmacol, 16*(2), 289-299. doi:10.1017/s1461145712000363

Crump, C., Sundquist, K., Winkleby, M. A., & Sundquist, J. (2013). Comorbidities and mortality in bipolar disorder: a Swedish national cohort study. *JAMA Psychiatry, 70*(9), 931-939. doi:10.1001/jamapsychiatry.2013.1394

Cui, H., Cruz-Correa, M., Giardiello, F. M., Hutcheon, D. F., Kafonek, D. R., Brandenburg, S., . . . Feinberg, A. P. (2003). Loss of IGF2 imprinting: a potential marker of colorectal cancer risk. *Science, 299*(March), 1753-1755. doi:10.1126/science.1080902

Cui, Q., Shi, H., Ye, P., Li, L., Qu, Q., Sun, G., . . . Shi, Y. (2017). m(6)A RNA Methylation Regulates the Self-Renewal and Tumorigenesis of Glioblastoma Stem Cells. *Cell Rep, 18*(11), 2622-2634. doi:10.1016/j.celrep.2017.02.059

Cutter, A. R., & Hayes, J. J. (2015). A brief review of nucleosome structure. *FEBS Lett, 589*(20 Pt A), 2914-2922. doi:10.1016/j.febslet.2015.05.016

Czepielewski, L., Daruy Filho, L., Brietzke, E., & Grassi-Oliveira, R. (2013). Bipolar disorder and metabolic syndrome: a systematic review. *Rev Bras Psiquiatr, 35*(1), 88-93.

da Rocha, S. T., Boeva, V., Escamilla-Del-Arenal, M., Ancelin, K., Granier, C., Matias, N. R., . . . Heard, E. (2014). Jarid2 Is Implicated in the Initial Xist-Induced Targeting of PRC2 to the Inactive X Chromosome. *Mol Cell, 53*(2), 301-316. doi:10.1016/j.molcel.2014.01.002

Dai, D., Wang, H., Zhu, L., Jin, H., & Wang, X. (2018). N6-methyladenosine links RNA metabolism to cancer progression. *Cell Death Dis, 9*(2). doi:10.1038/s41419-017-0129-x

Dalerba, P., Kalisky, T., Sahoo, D., Rajendran, P. S., Rothenberg, M. E., Leyrat, A. A., . . . Quake, S. R. (2011). Single-cell dissection of transcriptional heterogeneity in human colon tumors. *Nat Biotechnol, 29*(12), 1120-1127. doi:10.1038/nbt.2038

Dang, W., Steffen, K. K., Perry, R., Dorsey, J. A., Johnson, F. B., Shilatifard, A., . . . Berger, S. L. (2009). Histone H4 lysine 16 acetylation regulates cellular lifespan. *Nature, 459*(7248), 802-807. doi:10.1038/nature08085

Daxinger, L., & Whitelaw, E. (2012). Understanding transgenerational epigenetic inheritance via the gametes in mammals. *Nat Rev Genet, 13*(3), 153-162. doi:10.1038/nrg3188

de Castro Barbosa, T., Ingerslev, L. R., Alm, P. S., Versteyhe, S., Massart, J., Rasmussen, M., . . . Barres, R. (2016). High-fat diet reprograms the epigenome of rat spermatozoa and transgenerationally affects metabolism of the offspring. *Mol Metab, 5*(3), 184-197. doi:10.1016/j.molmet.2015.12.002

De Cecco, M., Criscione, S. W., Peckham, E. J., Hillenmeyer, S., Hamm, E. A., Manivannan, J., . . . Sedivy, J. M. (2013). Genomes of replicatively senescent cells undergo global epigenetic changes leading to gene silencing and activation of transposable elements. *Aging Cell, 12*(2), 247-256. doi:10.1111/acel.12047

De Cecco, M., Criscione, S. W., Peterson, A. L., Neretti, N., Sedivy, J. M., & Kreiling, J. A. (2013). Transposable elements become active and mobile in the genomes of aging mammalian somatic tissues. *Aging (Albany NY), 5*(12), 867-883. doi:10.18632/aging.100621

De Jager, P. L., Srivastava, G., Lunnon, K., Burgess, J., Schalkwyk, L. C., Yu, L., . . . Bennett, D. A. (2014). Alzheimer's disease: early alterations in brain DNA methylation at ANK1, BIN1, RHBDF2 and other loci. *Nat Neurosci, 17*(9), 1156-1163. doi:10.1038/nn.3786

de Oliveira, G. S., Cereser, K. M., Fernandes, B. S., Kauer-Sant'Anna, M., Fries, G. R., Stertz, L., . . . Kapczinski, F. (2009). Decreased brain-derived neurotrophic factor in medicated and drug-free bipolar patients. *J Psychiatr Res, 43*(14), 1171-1174. doi:10.1016/j.jpsychires.2009.04.002

De Vaux, V., Pfefferli, C., Passannante, M., Belhaj, K., von Essen, A., Sprecher, S. G., . . . Wicky, C. (2013). The Caenorhabditis elegans LET-418/Mi2 plays a conserved role in lifespan regulation. *Aging Cell, 12*(6), 1012-1020. doi:10.1111/acel.12129

Dedeurwaerder, S., Defrance, M., Calonne, E., Denis, H., Sotiriou, C., & Fuks, F. (2011). Evaluation of the Infinium Methylation 450K technology. *Epigenomics, 3*(6), 771-784. doi:10.2217/epi.11.105

Delaney, L., & Smith, J. P. (2012). Childhood health: trends and consequences over the life course. *Future Child, 22*(1), 43-63.

Dell'Osso, B., D'Addario, C., Carlotta Palazzo, M., Benatti, B., Camuri, G., Galimberti, D., . . . Altamura, A. C. (2014). Epigenetic modulation of BDNF gene: differences in DNA methylation between unipolar and bipolar patients. *J Affect Disord, 166*, 330-333. doi:10.1016/j.jad.2014.05.020

Dempster, E. L., Pidsley, R., Schalkwyk, L. C., Owens, S., Georgiades, A., Kane, F., . . . Mill, J. (2011). Disease-associated epigenetic changes in monozygotic twins discordant for

schizophrenia and bipolar disorder. *Hum Mol Genet, 20*(24), 4786-4796. doi:10.1093/hmg/ddr416

Derrien, T., Johnson, R., Bussotti, G., Tanzer, A., Djebali, S., Tilgner, H., . . . Guigo, R. (2012). The GENCODE v7 catalog of human long noncoding RNAs: analysis of their gene structure, evolution, and expression. *Genome Res, 22*(9), 1775-1789. doi:10.1101/gr.132159.111

Di Ruscio, A., Ebralidze, A. K., Benoukraf, T., Amabile, G., Goff, L. A., Terragni, J., . . . Tenen, D. G. (2013). DNMT1-interacting RNAs block gene-specific DNA methylation. *Nature, 503*(7476), 371-376. doi:10.1038/nature12598

Diep, D., Plongthongkum, N., Gore, A., Fung, H. L., Shoemaker, R., & Zhang, K. (2012). Library-free methylation sequencing with bisulfite padlock probes. *Nat Methods, 9*(3), 270-272. doi:10.1038/nmeth.1871

do Prado, C. H., Rizzo, L. B., Wieck, A., Lopes, R. P., Teixeira, A. L., Grassi-Oliveira, R., & Bauer, M. E. (2013). Reduced regulatory T cells are associated with higher levels of Th1/TH17 cytokines and activated MAPK in type 1 bipolar disorder. *Psychoneuroendocrinology, 38*(5), 667-676. doi:10.1016/j.psyneuen.2012.08.005

Dolinoy, D. C. (2008). The agouti mouse model: an epigenetic biosensor for nutritional and environmental alterations on the fetal epigenome. *Nutr Rev, 66*(Suppl 1), S7-11. doi:10.1111/j.1753-4887.2008.00056.x

Dolinoy, D. C., Huang, D., & Jirtle, R. L. (2007). Maternal nutrient supplementation counteracts bisphenol A-induced DNA hypomethylation in early development. *Proc Natl Acad Sci U S A, 104*(32), 13056-13061. doi:10.1073/pnas.0703739104

Domcke, S., Bardet, A. F., Adrian Ginno, P., Hartl, D., Burger, L., & Schubeler, D. (2015). Competition between DNA methylation and transcription factors determines binding of NRF1. *Nature, 528*(7583), 575-579. doi:10.1038/nature16462

Duarte, L. F., Young, A. R., Wang, Z., Wu, H. A., Panda, T., Kou, Y., . . . Bernstein, E. (2014). Histone H3.3 and its proteolytically processed form drive a cellular senescence programme. *Nat Commun, 5*, 5210. doi:10.1038/ncomms6210

Dueck, H., Khaladkar, M., Kim, T. K., Spaethling, J. M., Francis, C., Suresh, S., . . . Kim, J. (2015). Deep sequencing reveals cell-type-specific patterns of single-cell transcriptome variation. *Genome Biol, 16*, 122. doi:10.1186/s13059-015-0683-4

Dupuis, J., Langenberg, C., Prokopenko, I., Saxena, R., Soranzo, N., Jackson, A. U., . . . Barroso, I. (2010). New genetic loci implicated in fasting glucose homeostasis and their impact on type 2 diabetes risk. *Nat Genet, 42*(2), 105-116. doi:10.1038/ng.520

Edgar, R. D., Jones, M. J., Meaney, M. J., Turecki, G., & Kobor, M. S. (2017). BECon: a tool for interpreting DNA methylation findings from blood in the context of brain. *Transl Psychiatry, 7*(8), e1187. doi:10.1038/tp.2017.171

Egeland, J. A., & Sussex, J. N. (1985). Suicide and family loading for affective disorders. *JAMA, 254*(7), 915-918.

Ehrlund, A., Acosta, J. R., Bjork, C., Heden, P., Douagi, I., Arner, P., & Laurencikiene, J. (2017). The cell-type specific transcriptome in human adipose tissue and influence of obesity on adipocyte progenitors. *Sci Data, 4*, 170164. doi:10.1038/sdata.2017.164

Elliott, G., Hong, C., Xing, X., Zhou, X., Li, D., Coarfa, C., . . . Costello, J. F. (2015). Intermediate DNA methylation is a conserved signature of genome regulation. *Nat Commun, 6*, 6363. doi:10.1038/ncomms7363

Enattah, N. S., Jensen, T. G., Nielsen, M., Lewinski, R., Kuokkanen, M., Rasinpera, H., . . . Peltonen, L. (2008). Independent introduction of two lactase-persistence alleles into human populations reflects different history of adaptation to milk culture. *Am J Hum Genet, 82*(1), 57-72. doi:10.1016/j.ajhg.2007.09.012

Enattah, N. S., Sahi, T., Savilahti, E., Terwilliger, J. D., Peltonen, L., & Jarvela, I. (2002). Identification of a variant associated with adult-type hypolactasia. *Nat Genet, 30*(2), 233-237. doi:10.1038/ng826

Enattah, N. S., Trudeau, A., Pimenoff, V., Maiuri, L., Auricchio, S., Greco, L., . . . Peltonen, L. (2007). Evidence of still-ongoing convergence evolution of the lactase persistence T-13910 alleles in humans. *Am J Hum Genet, 81*(3), 615-625. doi:10.1086/520705

Ernst, C. L., & Goldberg, J. F. (2004). Clinical features related to age at onset in bipolar disorder. *J Affect Disord, 82*(1), 21-27. doi:10.1016/j.jad.2003.10.002

Esteller, M., Silva, J. M., Dominguez, G., Bonilla, F., Matias-Guiu, X., Lerma, E., . . . Herman, J. G. (2000). Promoter hypermethylation and BRCA1 inactivation in sporadic breast and ovarian tumors. *J Natl Cancer Inst, 92*(7), 564-569.

Ewald, E. R., Wand, G. S., Seifuddin, F., Yang, X., Tamashiro, K. L., Potash, J. B., . . . Lee, R. S. (2014). Alterations in DNA methylation of Fkbp5 as a determinant of blood-brain correlation of glucocorticoid exposure. *Psychoneuroendocrinology, 44*, 112-122. doi:10.1016/j.psyneuen.2014.03.003

Fang, L., Ahn, J. K., Wodziak, D., & Sibley, E. (2012). The human lactase persistence-associated SNP -13910*T enables in vivo functional persistence of lactase promoter-reporter transgene expression. *Hum Genet, 131*(7), 1153-1159. doi:10.1007/s00439-012-1140-z

Fang, R., Olds, L. C., & Sibley, E. (2006). Spatio-temporal patterns of intestine-specific transcription factor expression during postnatal mouse gut development. *Gene Expr Patterns, 6*(4), 426-432. doi:10.1016/j.modgep.2005.09.003

Farlik, M., Sheffield, N. C., Nuzzo, A., Datlinger, P., Schonegger, A., Klughammer, J., & Bock, C. (2015). Single-cell DNA methylome sequencing and bioinformatic inference of epigenomic cell-state dynamics. *Cell Rep, 10*(8), 1386-1397. doi:10.1016/j.celrep.2015.02.001

Feil, R., & Fraga, M. F. (2012). Epigenetics and the environment: emerging patterns and implications. *Nat Rev Genet, 13*(2), 97-109. doi:10.1038/nrg3142

Feinberg, A. P., & Irizarry, R. A. (2010). Evolution in health and medicine Sackler colloquium: Stochastic epigenetic variation as a driving force of development, evolutionary adaptation, and disease. *Proc Natl Acad Sci U S A, 107 Suppl 1*, 1757-1764. doi:10.1073/pnas.0906183107

Felsenfeld, G. (2014). A Brief History of Epigenetics. *Cold Spring Harb Perspect Biol, 6*(1). doi:10.1101/cshperspect.a018200

Ferrari, A. J., Stockings, E., Khoo, J. P., Erskine, H. E., Degenhardt, L., Vos, T., & Whiteford, H. A. (2016). The prevalence and burden of bipolar disorder: findings from the Global Burden of Disease Study 2013. *Bipolar Disord, 18*(5), 440-450. doi:10.1111/bdi.12423

Feser, J., Truong, D., Das, C., Carson, J. J., Kieft, J., Harkness, T., & Tyler, J. K. (2010). Elevated histone expression promotes life span extension. *Mol Cell, 39*(5), 724-735. doi:10.1016/j.molcel.2010.08.015

Ficz, G., Branco, M. R., Seisenberger, S., Santos, F., Krueger, F., Hore, T. A., . . . Reik, W. (2011). Dynamic regulation of 5-hydroxymethylcytosine in mouse ES cells and during differentiation. *Nature, 473*(7347), 398-402. doi:10.1038/nature10008

Fidzinski, P., Korotkova, T., Heidenreich, M., Maier, N., Schuetze, S., Kobler, O., . . . Jentsch, T. J. (2015). KCNQ5 K(+) channels control hippocampal synaptic inhibition and fast network oscillations. *Nat Commun, 6*, 6254. doi:10.1038/ncomms7254

Filipescu, D., Muller, S., & Almouzni, G. (2014). Histone H3 variants and their chaperones during development and disease: contributing to epigenetic control. *Annu Rev Cell Dev Biol, 30*, 615-646. doi:10.1146/annurev-cellbio-100913-013311

Filippova, G. N., Fagerlie, S., Klenova, E. M., Myers, C., Dehner, Y., Goodwin, G., . . . Lobanenkov, V. V. (1996). An exceptionally conserved transcriptional repressor, CTCF, employs different combinations of zinc fingers to bind diverged promoter sequences of avian and mammalian c-myc oncogenes. *Mol Cell Biol, 16*(6), 2802-2813.

Flavahan, W. A., Drier, Y., Liau, B. B., Gillespie, S. M., Venteicher, A. S., Stemmer-Rachamimov, A. O., . . . Bernstein, B. E. (2016). Insulator dysfunction and oncogene activation in IDH mutant gliomas. *Nature, 529*(7584), 110-114. doi:10.1038/nature16490

Fleischer, T., Tekpli, X., Mathelier, A., Wang, S., Nebdal, D., Dhakal, H. P., . . . Kristensen, V. N. (2017). DNA methylation at enhancers identifies distinct breast cancer lineages. *Nat Commun, 8*(1), 1379. doi:10.1038/s41467-017-00510-x

Flom, J. D., Ferris, J. S., Liao, Y., Tehranifar, P., Richards, C. B., Cho, Y. H., . . . Terry, M. B. (2011). Prenatal Smoke Exposure and Genomic DNA Methylation in a Multiethnic Birth Cohort. *Cancer Epidemiol Biomarkers Prev, 20*(12), 2518-2523. doi:10.1158/1055-9965.epi-11-0553

Florath, I., Butterbach, K., Heiss, J., Bewerunge-Hudler, M., Zhang, Y., Schottker, B., & Brenner, H. (2016). Type 2 diabetes and leucocyte DNA methylation: an epigenome-wide association study in over 1,500 older adults. *Diabetologia, 59*(1), 130-138. doi:10.1007/s00125-015-3773-7

Fontana, L., Partridge, L., & Longo, V. D. (2010). Extending healthy life span--from yeast to humans. *Science, 328*(5976), 321-326. doi:10.1126/science.1172539

Fraga, M. F., Ballestar, E., Paz, M. F., Ropero, S., Setien, F., Ballestar, M. L., . . . Esteller, M. (2005). Epigenetic differences arise during the lifetime of monozygotic twins. *Proc Natl Acad Sci U S A, 102*(30), 10604-10609. doi:10.1073/pnas.0500398102

Frankel, S., Woods, J., Ziafazeli, T., & Rogina, B. (2015). RPD3 histone deacetylase and nutrition have distinct but interacting effects on Drosophila longevity. *Aging (Albany NY), 7*(12), 1112-1129. doi:10.18632/aging.100856

Frauer, C., Hoffmann, T., Bultmann, S., Casa, V., Cardoso, M. C., Antes, I., & Leonhardt, H. (2011). Recognition of 5-hydroxymethylcytosine by the Uhrf1 SRA domain. *PLoS ONE, 6*(6), e21306. doi:10.1371/journal.pone.0021306

Freeman, J. M. (1987). A clinical approach to the child with seizures and epilepsy. *Epilepsia, 28 Suppl 1*, S103-109.

Fries, G. R., Bauer, I. E., Scaini, G., Wu, M. J., Kazimi, I. F., Valvassori, S. S., . . . Quevedo, J. (2017). Accelerated epigenetic aging and mitochondrial DNA copy number in bipolar disorder. *Transl Psychiatry, 7*(12), 1283. doi:10.1038/s41398-017-0048-8

Froy, O. (2011). Circadian rhythms, aging, and life span in mammals. *Physiology (Bethesda), 26*(4), 225-235. doi:10.1152/physiol.00012.2011

Fu, V. X., Dobosy, J. R., Desotelle, J. A., Almassi, N., Ewald, J. A., Srinivasan, R., . . . Jarrard, D. F. (2008). Aging and cancer-related loss of insulin-like growth factor 2 imprinting in the mouse and human prostate. *Cancer Res, 68*(16), 6797-6802. doi:10.1158/0008-5472.can-08-1714

Galetzka, D., Hansmann, T., El Hajj, N., Weis, E., Irmscher, B., Ludwig, M., . . . Haaf, T. (2012). Monozygotic twins discordant for constitutive BRCA1 promoter methylation, childhood cancer and secondary cancer. *Epigenetics, 7*(1), 47-54. doi:10.4161/epi.7.1.18814

Gallego Romero, I., Basu Mallick, C., Liebert, A., Crivellaro, F., Chaubey, G., Itan, Y., . . . Kivisild, T. (2012). Herders of Indian and European cattle share their predominant allele for lactase persistence. *Mol Biol Evol, 29*(1), 249-260. doi:10.1093/molbev/msr190

Gamba, C., Jones, E. R., Teasdale, M. D., McLaughlin, R. L., Gonzalez-Fortes, G., Mattiangeli, V., . . . Pinhasi, R. (2014). Genome flux and stasis in a five millennium transect of European prehistory. *Nat Commun, 5*, 5257. doi:10.1038/ncomms6257

Ganai, S. A. (2014). HDAC Inhibitors Entinostat and Suberoylanilide Hydroxamic Acid (SAHA): The Ray of Hope for Cancer Therapy. In R. D. Wells, J. S. Bond, J. Klinman, B. S. S. Masters, & E. Bell (Eds.), *Molecular Life Sciences: An Encyclopedic Reference* (pp. 1-16). New York, NY: Springer New York.

Garber, M., Guttman, M., Clamp, M., Zody, M. C., Friedman, N., & Xie, X. (2009). Identifying novel constrained elements by exploiting biased substitution patterns. *Bioinformatics, 25*(12), i54-62. doi:10.1093/bioinformatics/btp190

Garcia, A. M., Calder, R. B., Dolle, M. E., Lundell, M., Kapahi, P., & Vijg, J. (2010). Age- and temperature-dependent somatic mutation accumulation in Drosophila melanogaster. *PLoS Genet, 6*(5), e1000950. doi:10.1371/journal.pgen.1000950

Gayet, J., Zhou, X. P., Duval, A., Rolland, S., Hoang, J. M., Cottu, P., & Hamelin, R. (2001). Extensive characterization of genetic alterations in a series of human colorectal cancer cell lines. *Oncogene, 20*(36), 5025-5032. doi:10.1038/sj.onc.1204611

Genomes Project, C., Abecasis, G. R., Auton, A., Brooks, L. D., DePristo, M. A., Durbin, R. M., . . . McVean, G. A. (2012). An integrated map of genetic variation from 1,092 human genomes. *Nature, 491*(7422), 56-65. doi:10.1038/nature11632

Gentilini, D., Castaldi, D., Mari, D., Monti, D., Franceschi, C., Di Blasio, A. M., & Vitale, G. (2012). Age-dependent skewing of X chromosome inactivation appears delayed in centenarians' offspring. Is there a role for allelic imbalance in healthy aging and longevity? *Aging Cell, 11*(2), 277-283. doi:10.1111/j.1474-9726.2012.00790.x

Gergerlioglu, H. S., Savas, H. A., Bulbul, F., Selek, S., Uz, E., & Yumru, M. (2007). Changes in nitric oxide level and superoxide dismutase activity during antimanic treatment. *Prog Neuropsychopharmacol Biol Psychiatry, 31*(3), 697-702. doi:10.1016/j.pnpbp.2006.12.020

Gershon, E. S., Hamovit, J., Guroff, J. J., Dibble, E., Leckman, J. F., Sceery, W., . . . Bunney, W. E., Jr. (1982). A family study of schizoaffective, bipolar I, bipolar II, unipolar, and normal control probands. *Arch Gen Psychiatry, 39*(10), 1157-1167.

Ghedim, F. V., Fraga Dde, B., Deroza, P. F., Oliveira, M. B., Valvassori, S. S., Steckert, A. V., . . . Zugno, A. I. (2012). Evaluation of behavioral and neurochemical changes induced by ketamine in rats: implications as an animal model of mania. *J Psychiatr Res, 46*(12), 1569-1575. doi:10.1016/j.jpsychires.2012.08.010

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

Ghoshal, U. C., Kumar, S., Chourasia, D., & Misra, A. (2009). Lactose hydrogen breath test versus lactose tolerance test in the tropics: does positive lactose tolerance test reflect more severe lactose malabsorption? *Trop Gastroenterol, 30*(2), 86-90.

Gjoneska, E., Pfenning, A. R., Mathys, H., Quon, G., Kundaje, A., Tsai, L. H., & Kellis, M. (2015). Conserved epigenomic signals in mice and humans reveal immune basis of Alzheimer's disease. *Nature, 518*(7539), 365-369. doi:10.1038/nature14252

Goda, T., Yasutake, H., Tanaka, T., & Takase, S. (1999). Lactase-phlorizin hydrolase and sucrase-isomaltase genes are expressed differently along the villus-crypt axis of rat jejunum. *J Nutr, 129*(6), 1107-1113.

Goldman, R. D., Shumaker, D. K., Erdos, M. R., Eriksson, M., Goldman, A. E., Gordon, L. B., . . . Collins, F. S. (2004). Accumulation of mutant lamin A causes progressive changes in nuclear architecture in Hutchinson-Gilford progeria syndrome. *Proc Natl Acad Sci U S A, 101*(24), 8963-8968. doi:10.1073/pnas.0402943101

Goll, M. G., Kirpekar, F., Maggert, K. A., Yoder, J. A., Hsieh, C. L., Zhang, X., . . . Bestor, T. H. (2006). Methylation of tRNAAsp by the DNA methyltransferase homolog Dnmt2. *Science, 311*(5759), 395-398. doi:10.1126/science.1120976

Gong, L., Pan, Y. X., & Chen, H. (2010). Gestational low protein diet in the rat mediates Igf2 gene expression in male offspring via altered hepatic DNA methylation. *Epigenetics, 5*(7), 619-626.

Goueli, B. S., Powell, M. B., Finger, E. C., & Pfeffer, S. R. (2012). TBC1D16 is a Rab4A GTPase activating protein that regulates receptor recycling and EGF receptor signaling. *Proceedings of the National Academy of Sciences, 109*(39), 15787-15792. doi:10.1073/pnas.1204540109

Grun, D., Lyubimova, A., Kester, L., Wiebrands, K., Basak, O., Sasaki, N., . . . van Oudenaarden, A. (2015). Single-cell messenger RNA sequencing reveals rare intestinal cell types. *Nature, 525*(7568), 251-255. doi:10.1038/nature14966

Guarente, L. (2011). Franklin H. Epstein Lecture: Sirtuins, aging, and medicine. *N Engl J Med, 364*(23), 2235-2244. doi:10.1056/NEJMra1100831

Guenther, M. G., Levine, S. S., Boyer, L. A., Jaenisch, R., & Young, R. A. (2007). A chromatin landmark and transcription initiation at most promoters in human cells. *Cell, 130*(1), 77-88. doi:10.1016/j.cell.2007.05.042

Guidotti, A., Auta, J., Davis, J. M., Di-Giorgi-Gerevini, V., Dwivedi, Y., Grayson, D. R., . . . Costa, E. (2000). Decrease in reelin and glutamic acid decarboxylase67 (GAD67) expression in schizophrenia and bipolar disorder: a postmortem brain study. *Arch Gen Psychiatry, 57*(11), 1061-1069.

Guo, J. U., Su, Y., Shin, J. H., Shin, J., Li, H., Xie, B., . . . Song, H. (2014). Distribution, recognition and regulation of non-CpG methylation in the adult mammalian brain. *Nat Neurosci, 17*(2), 215-222. doi:10.1038/nn.3607

Guo, Y., Xu, Q., Canzio, D., Shou, J., Li, J., Gorkin, D. U., . . . Wu, Q. (2015). CRISPR Inversion of CTCF Sites Alters Genome Topology and Enhancer/Promoter Function. *Cell, 162*(4), 900-910. doi:10.1016/j.cell.2015.07.038

Haber, A. L., Biton, M., Rogel, N., Herbst, R. H., Shekhar, K., Smillie, C., . . . Regev, A. (2017). A single-cell survey of the small intestinal epithelium. *Nature, 551*(7680), 333-339. doi:10.1038/nature24489

Haigis, M. C., & Sinclair, D. A. (2010). Mammalian sirtuins: biological insights and disease relevance. *Annu Rev Pathol, 5*, 253-295. doi:10.1146/annurev.pathol.4.110807.092250

Hallahan, B., Newell, J., Soares, J. C., Brambilla, P., Strakowski, S. M., Fleck, D. E., . . . McDonald, C. (2011). Structural magnetic resonance imaging in bipolar disorder: an international collaborative mega-analysis of individual adult patient data. *Biol Psychiatry, 69*(4), 326-335. doi:10.1016/j.biopsych.2010.08.029

Halpern, K. B., Shenhav, R., Matcovitch-Natan, O., Toth, B., Lemze, D., Golan, M., . . . Itzkovitz, S. (2017). Single-cell spatial reconstruction reveals global division of labour in the mammalian liver. *Nature, 542*(7641), 352-356. doi:10.1038/nature21065

Hannon, E., Lunnon, K., Schalkwyk, L., & Mill, J. (2015). Interindividual methylomic variation across blood, cortex, and cerebellum: implications for epigenetic studies of neurological and neuropsychiatric phenotypes. *Epigenetics, 10*(11), 1024-1032. doi:10.1080/15592294.2015.1100786

Haque, F. N., Gottesman, II, & Wong, A. H. (2009). Not really identical: epigenetic differences in monozygotic twins and implications for twin studies in psychiatry. *Am J Med Genet C Semin Med Genet, 151c*(2), 136-141. doi:10.1002/ajmg.c.30206

Harel, I., Benayoun, B. A., Machado, B., Singh, P. P., Hu, C. K., Pech, M. F., . . . Brunet, A. (2015). A platform for rapid exploration of aging and diseases in a naturally short-lived vertebrate. *Cell, 160*(5), 1013-1026. doi:10.1016/j.cell.2015.01.038

Hark, A. T., Schoenherr, C. J., Katz, D. J., Ingram, R. S., Levorse, J. M., & Tilghman, S. M. (2000). CTCF mediates methylation-sensitive enhancer-blocking activity at the H19/Igf2 locus. *Nature, 405*(6785), 486-489. doi:10.1038/35013106

Harris, E. C., & Barraclough, B. (1997). Suicide as an outcome for mental disorders. A meta-analysis. *Br J Psychiatry, 170*, 205-228.

Hasler, G., Drevets, W. C., Gould, T. D., Gottesman, I. I., & Manji, H. K. (2006). Toward constructing an endophenotype strategy for bipolar disorders (Vol. 60, pp. 93-105).

Hassa, P. O., Haenni, S. S., Elser, M., & Hottiger, M. O. (2006). Nuclear ADP-ribosylation reactions in mammalian cells: where are we today and where are we going? *Microbiol Mol Biol Rev, 70*(3), 789-829. doi:10.1128/mmbr.00040-05

Hawton, K., Malmberg, A., & Simkin, S. (2004). Suicide in doctors. A psychological autopsy study. *J Psychosom Res, 57*(1), 1-4. doi:10.1016/s0022-3999(03)00372-6

Hawton, K., Sutton, L., Haw, C., Sinclair, J., & Harriss, L. (2005). Suicide and attempted suicide in bipolar disorder: a systematic review of risk factors. *The Journal of clinical psychiatry, 66*(6), 693-704. doi:10.4088/JCP.v66n0604

Hayflick, L. (1991). Aging under glass. *Mutat Res, 256*(2-6), 69-80.

Hayflick, L. (2007). Biological aging is no longer an unsolved problem. *Ann N Y Acad Sci, 1100*, 1-13. doi:10.1196/annals.1395.001

Hayflick, L., & Moorhead, P. S. (1961). The serial cultivation of human diploid cell strains. *Exp Cell Res, 25*, 585-621.

He, Y. F., Li, B. Z., Li, Z., Liu, P., Wang, Y., Tang, Q., . . . Xu, G. L. (2011). Tet-mediated formation of 5-carboxylcytosine and its excision by TDG in mammalian DNA. *Science, 333*(6047), 1303-1307. doi:10.1126/science.1210944

Hébert, M., Beattie, C. W., Tam, E. M., Yatham, L. N., & Lam, R. W. (2004). Electroretinography in patients with winter seasonal affective disorder. *Psychiatry Research, 127*(1-2), 27-34. doi:10.1016/j.psychres.2004.03.006

Hébert, M., Dumont, M., & Lachapelle, P. (2002). Electrophysiological evidence suggesting a seasonal modulation of retinal sensitivity in subsyndromal winter depression. *Journal of Affective Disorders, 68*(2-3), 191-202. doi:10.1016/S0165-0327(00)00192-0

Hegar, B., & Widodo, A. (2015). Lactose intolerance in Indonesian children. *Asia Pac J Clin Nutr, 24 Suppl 1*, S31-40. doi:10.6133/apjcn.2015.24.s1.06

Heijmans, B. T., Tobi, E. W., Stein, A. D., Putter, H., Blauw, G. J., Susser, E. S., . . . Lumey, L. H. (2008). Persistent epigenetic differences associated with prenatal exposure to famine in humans. *Proc Natl Acad Sci U S A, 105*(44), 17046-17049. doi:10.1073/pnas.0806560105

Heilmann, K., Toth, R., Bossmann, C., Klimo, K., Plass, C., & Gerhauser, C. (2017). Genome-wide screen for differentially methylated long noncoding RNAs identifies Esrp2 and lncRNA Esrp2-as regulated by enhancer DNA methylation with prognostic relevance for human breast cancer. *Oncogene, 36*(46), 6446-6461. doi:10.1038/onc.2017.246

Heintzman, N. D., Stuart, R. K., Hon, G., Fu, Y., Ching, C. W., Hawkins, R. D., . . . Ren, B. (2007). Distinct and predictive chromatin signatures of transcriptional promoters and enhancers in the human genome. *Nat Genet, 39*(3), 311-318. doi:10.1038/ng1966

Hermsen, M., Postma, C., Baak, J., Weiss, M., Rapallo, A., Sciutto, A., . . . Meijer, G. (2002). Colorectal adenoma to carcinoma progression follows multiple pathways of chromosomal instability. *Gastroenterology, 123*(4), 1109-1119.

Heyn, H., Carmona, F. J., Gomez, A., Ferreira, H. J., Bell, J. T., Sayols, S., . . . Esteller, M. (2013). DNA methylation profiling in breast cancer discordant identical twins identifies DOK7 as novel epigenetic biomarker. *Carcinogenesis, 34*(1), 102-108. doi:10.1093/carcin/bgs321

Heyn, H., Moran, S., Hernando-Herraez, I., Sayols, S., Gomez, A., Sandoval, J., . . . Esteller, M. (2013). DNA methylation contributes to natural human variation. *Genome Res, 23*(9), 1363-1372. doi:10.1101/gr.154187.112

Hilton, I. B., D'Ippolito, A. M., Vockley, C. M., Thakore, P. I., Crawford, G. E., Reddy, T. E., & Gersbach, C. A. (2015). Epigenome editing by a CRISPR-Cas9-based acetyltransferase activates genes from promoters and enhancers. *Nat Biotechnol, 33*(5), 510-517. doi:10.1038/nbt.3199

Hilty, D. M., Leamon, M. H., Lim, R. F., Kelly, R. H., & Hales, R. E. (2006). A review of bipolar disorder in adults. *Psychiatry (Edgmont), 3*(9), 43-55.

Hnisz, D., Weintraub, A. S., Day, D. S., Valton, A. L., Bak, R. O., Li, C. H., . . . Young, R. A. (2016). Activation of proto-oncogenes by disruption of chromosome neighborhoods. *Science, 351*(6280), 1454-1458. doi:10.1126/science.aad9024

Hodawadekar, S. C., & Marmorstein, R. (2007). Chemistry of acetyl transfer by histone modifying enzymes: structure, mechanism and implications for effector design. *Oncogene, 26*(37), 5528-5540. doi:10.1038/sj.onc.1210619

Horvath, S. (2013). DNA methylation age of human tissues and cell types. *Genome Biol, 14*(10), R115. doi:10.1186/gb-2013-14-10-r115

Horvath, S., & Ritz, B. R. (2015). Increased epigenetic age and granulocyte counts in the blood of Parkinson's disease patients. *Aging (Albany NY), 7*(12), 1130-1142. doi:10.18632/aging.100859

Hou, C., Zhao, H., Tanimoto, K., & Dean, A. (2008). CTCF-dependent enhancer-blocking by alternative chromatin loop formation. *Proc Natl Acad Sci U S A, 105*(51), 20398-20403. doi:10.1073/pnas.0808506106

Houtepen, L. C., van Bergen, A. H., Vinkers, C. H., & Boks, M. P. (2016). DNA methylation signatures of mood stabilizers and antipsychotics in bipolar disorder. *Epigenomics, 8*(2), 197-208. doi:10.2217/epi.15.98

Hu, Z., Chen, K., Xia, Z., Chavez, M., Pal, S., Seol, J. H., . . . Tyler, J. K. (2014). Nucleosome loss leads to global transcriptional up-regulation and genomic instability during yeast aging. *Genes Dev, 28*(4), 396-408. doi:10.1101/gad.233221.113

Huang, R. C., Galati, J. C., Burrows, S., Beilin, L. J., Li, X., Pennell, C. E., . . . Craig, J. M. (2012). DNA methylation of the IGF2/H19 imprinting control region and adiposity distribution in young adults. *Clin Epigenetics, 4*(1), 21. doi:10.1186/1868-7083-4-21

Huff, F. J., Growdon, J. H., Corkin, S., & Rosen, T. J. (1987). Age at onset and rate of progression of Alzheimer's disease. *J Am Geriatr Soc, 35*(1), 27-30.

Huisman, B., Manske, G., Carney, S., & Kalantry, S. (2017). Functional Dissection of the m6A RNA Modification. *Trends Biochem Sci, 42*(2), 85-86. doi:10.1016/j.tibs.2016.12.004

Huypens, P., Sass, S., Wu, M., Dyckhoff, D., Tschop, M., Theis, F., . . . Beckers, J. (2016). Epigenetic germline inheritance of diet-induced obesity and insulin resistance. *Nat Genet, 48*(5), 497-499. doi:10.1038/ng.3527

Illingworth, R. S., Gruenewald-Schneider, U., Webb, S., Kerr, A. R., James, K. D., Turner, D. J., . . . Bird, A. P. (2010). Orphan CpG islands identify numerous conserved promoters in the mammalian genome. *PLoS Genet, 6*(9), e1001134. doi:10.1371/journal.pgen.1001134

Ingram, C. J., Mulcare, C. A., Itan, Y., Thomas, M. G., & Swallow, D. M. (2009). Lactose digestion and the evolutionary genetics of lactase persistence. *Hum Genet, 124*(6), 579-591. doi:10.1007/s00439-008-0593-6

Itan, Y., Powell, A., Beaumont, M. A., Burger, J., & Thomas, M. G. (2009). The origins of lactase persistence in Europe. *PLoS Comput Biol, 5*(8), e1000491. doi:10.1371/journal.pcbi.1000491

Ito, S., D'Alessio, A. C., Taranova, O. V., Hong, K., Sowers, L. C., & Zhang, Y. (2010). Role of Tet proteins in 5mC to 5hmC conversion, ES-cell self-renewal and inner cell mass specification. *Nature, 466*(7310), 1129-1133. doi:10.1038/nature09303

Ito, S., Shen, L., Dai, Q., Wu, S. C., Collins, L. B., Swenberg, J. A., . . . Zhang, Y. (2011). Tet proteins can convert 5-methylcytosine to 5-formylcytosine and 5-carboxylcytosine. *Science, 333*(6047), 1300-1303. doi:10.1126/science.1210597

Ivanov, A., Pawlikowski, J., Manoharan, I., van Tuyn, J., Nelson, D. M., Rai, T. S., . . . Adams, P. D. (2013). Lysosome-mediated processing of chromatin in senescence. *J Cell Biol, 202*(1), 129-143. doi:10.1083/jcb.201212110

J. K. Wing, M. D., T. Babor, M. D., T. Brugha, M. D., J. Burke, M. D., J. E. Cooper, M. D., R. Giel, M. D., . . . N. Sartorius, M. D. (1990). SCAN Schedules for Clinical Assessment in Neuropsychiatry. *Archives of General Psychiatry, 47*(6), 589-593.

Jaenisch, R., & Bird, A. (2003). Epigenetic regulation of gene expression: how the genome integrates intrinsic and environmental signals. *Nat Genet, 33 Suppl*, 245-254. doi:10.1038/ng1089

Jamison, K. R. (1986). Suicide and bipolar disorders. *Annals New York Academy of Sciences, 487*, 301-315.

Jamison, K. R. (2000). Suicide and bipolar disorder. *J Clin Psychiatry, 61 Suppl 9*, 47-51.

Jensen, T. G., Liebert, A., Lewinsky, R., Swallow, D. M., Olsen, J., & Troelsen, J. T. (2011). The -14010*C variant associated with lactase persistence is located between an Oct-1 and HNF1alpha binding site and increases lactase promoter activity. *Hum Genet, 130*(4), 483-493. doi:10.1007/s00439-011-0966-0

Jiang, R., Jones, M. J., Chen, E., Neumann, S. M., Fraser, H. B., Miller, G. E., & Kobor, M. S. (2015). Discordance of DNA methylation variance between two accessible human tissues. *Sci Rep, 5*, 8257. doi:10.1038/srep08257

Jiang, T. (2016). Let-7c miRNA Inhibits the Proliferation and Migration of Heat-Denatured Dermal Fibroblasts Through Down-Regulating HSP70. *39*(4), 345-351. doi:10.14348/molcells.2016.2336

Jones, B. L., Raga, T. O., Liebert, A., Zmarz, P., Bekele, E., Danielsen, E. T., . . . Swallow, D. M. (2013). Diversity of lactase persistence alleles in Ethiopia: signature of a soft selective sweep. *Am J Hum Genet, 93*(3), 538-544. doi:10.1016/j.ajhg.2013.07.008

Jones, P. A. (2012). Functions of DNA methylation: islands, start sites, gene bodies and beyond. *Nat Rev Genet, 13*(7), 484-492. doi:10.1038/nrg3230

Jung, M., & Pfeifer, G. P. (2015). Aging and DNA methylation. *BMC Biology, 13*(1), 7. doi:10.1186/s12915-015-0118-4

Kaaij, L. T., van de Wetering, M., Fang, F., Decato, B., Molaro, A., van de Werken, H. J., . . . Ketting, R. F. (2013). DNA methylation dynamics during intestinal stem cell differentiation reveals enhancers driving gene expression in the villus. *Genome Biol, 14*(5), R50. doi:10.1186/gb-2013-14-5-r50

Kaeberlein, M., Burtner, C. R., & Kennedy, B. K. (2007). Recent developments in yeast aging. *PLoS Genet, 3*(5), e84. doi:10.1371/journal.pgen.0030084

Kamakaka, R. T., & Biggins, S. (2005). Histone variants: deviants? *Genes Dev, 19*(3), 295-310. doi:10.1101/gad.1272805

Kaminsky, Z., Jones, I., Verma, R., Saleh, L., Trivedi, H., Guintivano, J., . . . Potash, J. B. (2015). DNA methylation and expression of KCNQ3 in bipolar disorder. *Bipolar Disord, 17*(2), 150-159. doi:10.1111/bdi.12230

Kaminsky, Z., Tochigi, M., Jia, P., Pal, M., Mill, J., Kwan, A., . . . Petronis, A. (2012). A multi-tissue analysis identifies HLA complex group 9 gene methylation differences in bipolar disorder. *Mol Psychiatry, 17*(7), 728-740. doi:10.1038/mp.2011.64

Kaminsky, Z., Wilcox, H. C., Eaton, W. W., Van Eck, K., Kilaru, V., Jovanovic, T., . . . Smith, A. K. (2015). Epigenetic and genetic variation at SKA2 predict suicidal behavior and post-traumatic stress disorder. *Translational Psychiatry, 5*(8), e627-e627. doi:10.1038/tp.2015.105

Kaminsky, Z. A., Tang, T., Wang, S. C., Ptak, C., Oh, G. H., Wong, A. H., . . . Petronis, A. (2009). DNA methylation profiles in monozygotic and dizygotic twins. *Nat Genet, 41*(2), 240-245. doi:10.1038/ng.286

Kanherkar, R. R., Bhatia-Dey, N., & Csoka, A. B. (2014). Epigenetics across the human lifespan. *Front Cell Dev Biol, 2*. doi:10.3389/fcell.2014.00049

Kareta, M. S., Botello, Z. M., Ennis, J. J., Chou, C., & Chedin, F. (2006). Reconstitution and mechanism of the stimulation of de novo methylation by human DNMT3L. *J Biol Chem, 281*(36), 25893-25902. doi:10.1074/jbc.M603140200

Katainen, R., Dave, K., Pitkanen, E., Palin, K., Kivioja, T., Valimaki, N., . . . Aaltonen, L. A. (2015). CTCF/cohesin-binding sites are frequently mutated in cancer. *Nat Genet, 47*(7), 818-821. doi:10.1038/ng.3335

Kemp, C. J., Moore, J. M., Moser, R., Bernard, B., Teater, M., Smith, L. E., . . . Filippova, G. N. (2014). CTCF haploinsufficiency destabilizes DNA methylation and predisposes to cancer. *Cell Rep, 7*(4), 1020-1029. doi:10.1016/j.celrep.2014.04.004

Kessing, L. V., & Andersen, P. K. (2004). Does the risk of developing dementia increase with the number of episodes in patients with depressive disorder and in patients with bipolar disorder? *J Neurol Neurosurg Psychiatry, 75*(12), 1662-1666. doi:10.1136/jnnp.2003.031773

Khare, T., Pai, S., Koncevicius, K., Pal, M., Kriukiene, E., Liutkeviciute, Z., . . . Petronis, A. (2012). 5-hmC in the brain is abundant in synaptic genes and shows differences at the exon-intron boundary. *Nat Struct Mol Biol, 19*(10), 1037-1043. doi:10.1038/nsmb.2372

Kim, A. Y., Park, Y. J., Pan, X., Shin, K. C., Kwak, S. H., Bassas, A. F., . . . Kim, J. B. (2015). Obesity-induced DNA hypermethylation of the adiponectin gene mediates insulin resistance. *Nat Commun, 6*, 7585. doi:10.1038/ncomms8585

Kim, H. T., Kim, K. P., Uchiki, T., Gygi, S. P., & Goldberg, A. L. (2009). S5a promotes protein degradation by blocking synthesis of nondegradable forked ubiquitin chains. *Embo j, 28*(13), 1867-1877. doi:10.1038/emboj.2009.115

Kim, J., Kim, J. Y., Song, K. S., Lee, Y. H., Seo, J. S., Jelinek, J., . . . Issa, J. P. (2007). Epigenetic changes in estrogen receptor beta gene in atherosclerotic cardiovascular tissues and in-vitro vascular senescence. *Biochim Biophys Acta, 1772*(1), 72-80. doi:10.1016/j.bbadis.2006.10.004

Kim, S., Yu, N. K., & Kaang, B. K. (2015). CTCF as a multifunctional protein in genome regulation and gene expression. *Exp Mol Med, 47*, e166. doi:10.1038/emm.2015.33

Kim, S. W., Ramasamy, K., Bouamar, H., Lin, A. P., Jiang, D., & Aguiar, R. C. (2012). MicroRNAs miR-125a and miR-125b constitutively activate the NF-kappaB pathway by targeting the tumor necrosis factor alpha-induced protein 3 (TNFAIP3, A20). *Proc Natl Acad Sci U S A, 109*(20), 7865-7870. doi:10.1073/pnas.1200081109

Kim, Y. I. (2005). Nutritional epigenetics: impact of folate deficiency on DNA methylation and colon cancer susceptibility. *J Nutr, 135*(11), 2703-2709. doi:10.1093/jn/135.11.2703

Kirkegaard, P., Edwards, A., Nielsen, T. L. O., Ørntoft, T. F., Sørensen, K. D., Borre, M., & Bro, F. (2018). Perceptions about screening for prostate cancer using genetic lifetime risk assessment: a qualitative study. *BMC Family Practice, 19*(1), 32. doi:10.1186/s12875-018-0717-6

Kita, Y., Baba, H., Maeshima, H., Nakano, Y., Suzuki, T., & Arai, H. (2009). Serum amyloid beta protein in young and elderly depression: a pilot study. *Psychogeriatrics, 9*(4), 180-185. doi:10.1111/j.1479-8301.2009.00293.x

Klose, R. J., & Bird, A. P. (2006). Genomic DNA methylation: the mark and its mediators. *Trends Biochem Sci, 31*(2), 89-97. doi:10.1016/j.tibs.2005.12.008

Klose, R. J., & Zhang, Y. (2007). Regulation of histone methylation by demethylimination and demethylation. *Nat Rev Mol Cell Biol, 8*(4), 307-318. doi:10.1038/nrm2143

Kohli, R. M., & Zhang, Y. (2013). TET enzymes, TDG and the dynamics of DNA demethylation. *Nature, 502*(7472), 472-479. doi:10.1038/nature12750

Kohlnhofer, B. M., Thompson, C. A., Walker, E. M., & Battle, M. A. (2016). GATA4 regulates epithelial cell proliferation to control intestinal growth and development in mice. *Cell Mol Gastroenterol Hepatol, 2*(2), 189-209. doi:10.1016/j.jcmgh.2015.11.010

Konarski, J. Z., McIntyre, R. S., Kennedy, S. H., Rafi-Tari, S., Soczynska, J. K., & Ketter, T. A. (2008). Volumetric neuroimaging investigations in mood disorders: bipolar disorder versus major depressive disorder. *Bipolar Disord, 10*(1), 1-37. doi:10.1111/j.1399-5618.2008.00435.x

Kondo, Y. (2008). Downregulation of Histone H3 Lysine 9 Methyltransferase G9a Induces Centrosome Disruption and Chromosome Instability in Cancer Cells. *3*(4). doi:10.1371/journal.pone.0002037

Kondratova, A. A., & Kondratov, R. V. (2012). The circadian clock and pathology of the ageing brain. *Nat Rev Neurosci, 13*(5), 325-335. doi:10.1038/nrn3208

Kriukiene, E., Labrie, V., Khare, T., Urbanaviciute, G., Lapinaite, A., Koncevicius, K., . . . Klimasauskas, S. (2013). DNA unmethylome profiling by covalent capture of CpG sites. *Nat Commun, 4*, 2190. doi:10.1038/ncomms3190

Krueger, F., & Andrews, S. R. (2011). Bismark: a flexible aligner and methylation caller for Bisulfite-Seq applications. *Bioinformatics, 27*(11), 1571-1572. doi:10.1093/bioinformatics/btr167

Krzywinski, M., Schein, J., Birol, I., Connors, J., Gascoyne, R., Horsman, D., . . . Marra, M. A. (2009). Circos: an information aesthetic for comparative genomics. *Genome Res, 19*(9), 1639-1645. doi:10.1101/gr.092759.109

Ku, M., Koche, R. P., Rheinbay, E., Mendenhall, E. M., Endoh, M., Mikkelsen, T. S., . . . Bernstein, B. E. (2008). Genomewide analysis of PRC1 and PRC2 occupancy identifies two classes of bivalent domains. *PLoS Genet, 4*(10), e1000242. doi:10.1371/journal.pgen.1000242

Kulkarni, H., Kos, M. Z., Neary, J., Dyer, T. D., Kent, J. W., Jr., Goring, H. H., . . . Carless, M. A. (2015). Novel epigenetic determinants of type 2 diabetes in Mexican-American families. *Hum Mol Genet, 24*(18), 5330-5344. doi:10.1093/hmg/ddv232

Kundaje, A., Meuleman, W., Ernst, J., Bilenky, M., Yen, A., Heravi-Moussavi, A., . . . Kellis, M. (2015). Integrative analysis of 111 reference human epigenomes. *Nature, 518*(7539), 317-330. doi:10.1038/nature14248

Kundaje, A., Meuleman, W., Ernst, J., Bilenky, M., Yen, A., Heravi-Moussavi, A., . . . Kellis, M. (2015). Integrative analysis of 111 reference human epigenomes. *Nature, 518*(7539), 317-330. doi:10.1038/nature14248

Labrie, V., Buske, O. J., Oh, E., Jeremian, R., Ptak, C., Gasiunas, G., . . . Petronis, A. (2016). Lactase nonpersistence is directed by DNA-variation-dependent epigenetic aging. *Nat Struct Mol Biol, 23*(6), 566-573. doi:10.1038/nsmb.3227

Labrie, V., Pai, S., & Petronis, A. (2012). Epigenetics of major psychosis: Progress, problems and perspectives. *Trends in Genetics, 28*(9), 427-435. doi:10.1016/j.tig.2012.04.002

Lahiri, D. K., & Nurnberger, J. I. (1991). A rapid non-enzymatic method for the preparation of HMW DNA from blood for RFLP studies. *Nucleic Acids Research, 19*(19), 5444-5444. doi:10.1093/nar/19.19.5444

Lan, F., & Shi, Y. (2009). Epigenetic regulation: methylation of histone and non-histone proteins. *Sci China C Life Sci, 52*(4), 311-322. doi:10.1007/s11427-009-0054-z

Lawson, H. A., Cheverud, J. M., & Wolf, J. B. (2013). Genomic imprinting and parent-of-origin effects on complex traits. *Nat Rev Genet, 14*(9), 609-617. doi:10.1038/nrg3543

Lee, C., & Longo, V. (2016). Dietary restriction with and without caloric restriction for healthy aging. *F1000Res, 5*. doi:10.12688/f1000research.7136.1

Lee, E., Iskow, R., Yang, L., Gokcumen, O., Haseley, P., Luquette, L. J., 3rd, . . . Park, P. J. (2012). Landscape of somatic retrotransposition in human cancers. *Science, 337*(6097), 967-971. doi:10.1126/science.1222077

Lee, M. F., Russell, R. M., Montgomery, R. K., & Krasinski, S. D. (1997). Total intestinal lactase and sucrase activities are reduced in aged rats. *J Nutr, 127*(7), 1382-1387.

Lee, M. G., Villa, R., Trojer, P., Norman, J., Yan, K. P., Reinberg, D., . . . Shiekhattar, R. (2007). Demethylation of H3K27 regulates polycomb recruitment and H2A ubiquitination. *Science, 318*(5849), 447-450. doi:10.1126/science.1149042

Lee, Y. M., & Ayres, J. S. (2018). Decoding the intestinal epithelium cell by cell. *Nat Immunol, 19*(1), 7-9. doi:10.1038/s41590-017-0011-0

Leeper, N. J., Raiesdana, A., Kojima, Y., Chun, H. J., Azuma, J., Maegdefessel, L., . . . Spin, J. M. (2011). MicroRNA-26a is a novel regulator of vascular smooth muscle cell function. *J Cell Physiol, 226*(4), 1035-1043. doi:10.1002/jcp.22422

Lehtovirta, M., Pietilainen, K. H., Levalahti, E., Heikkila, K., Groop, L., Silventoinen, K., . . . Kaprio, J. (2010). Evidence that BMI and type 2 diabetes share only a minor fraction of genetic variance: a follow-up study of 23,585 monozygotic and dizygotic twins from the Finnish Twin Cohort Study. *Diabetologia, 53*(7), 1314-1321. doi:10.1007/s00125-010-1746-4

Leick, M. B., Shoff, C. J., Wang, E. C., Congress, J. L., & Gallicano, G. I. (2012). Loss of imprinting of IGF2 and the epigenetic progenitor model of cancer. *Am J Stem Cells, 1*(1), 59-74.

Leng, S., Do, K., Yingling, C. M., Picchi, M. A., Wolf, H. J., Kennedy, T. C., . . . Belinsky, S. A. (2012). Defining a Gene Promoter Methylation Signature in Sputum for Lung Cancer Risk Assessment. *Clin Cancer Res, 18*(12), 3387-3395. doi:10.1158/1078-0432.ccr-11-3049

Lengauer, C., Kinzler, K. W., & Vogelstein, B. (1997). Genetic instability in colorectal cancers. *Nature, 386*(6625), 623-627. doi:10.1038/386623a0

Lenox, R. H., Gould, T. D., & Manji, H. K. (2002). Endophenotypes in bipolar disorder. *American Journal of Medical Genetics - Neuropsychiatric Genetics, 114*(4), 391-406. doi:10.1002/ajmg.10360

Leseva, M. N., Grand, R. J., Klett, H., Boerries, M., Busch, H., Binder, A. M., & Michels, K. B. (2018). Differences in DNA Methylation and Functional Expression in Lactase Persistent and Non-persistent Individuals. *Sci Rep, 8*(1), 5649. doi:10.1038/s41598-018-23957-4

Leverich, G. S., McElroy, S. L., Suppes, T., Keck, P. E., Jr., Denicoff, K. D., Nolen, W. A., . . . Post, R. M. (2002). Early physical and sexual abuse associated with an adverse course of bipolar illness. *Biol Psychiatry, 51*(4), 288-297.

Levine, A. J., Quach, A., Moore, D. J., Achim, C. L., Soontornniyomkij, V., Masliah, E., . . . Horvath, S. (2016). Accelerated epigenetic aging in brain is associated with pre-mortem HIV-associated neurocognitive disorders. *J Neurovirol, 22*(3), 366-375. doi:10.1007/s13365-015-0406-3

Levine, M. E., Lu, A. T., Bennett, D. A., & Horvath, S. (2015). Epigenetic age of the pre-frontal cortex is associated with neuritic plaques, amyloid load, and Alzheimer's disease related cognitive functioning. *Aging (Albany NY), 7*(12), 1198-1211. doi:10.18632/aging.100864

Lewinsky, R. H., Jensen, T. G., Moller, J., Stensballe, A., Olsen, J., & Troelsen, J. T. (2005). T-13910 DNA variant associated with lactase persistence interacts with Oct-1 and stimulates lactase promoter activity in vitro. *Hum Mol Genet, 14*(24), 3945-3953. doi:10.1093/hmg/ddi418

Li, Y., Daniel, M., & Tollefsbol, T. O. (2011). Epigenetic regulation of caloric restriction in aging. *BMC Med, 9*, 98. doi:10.1186/1741-7015-9-98

Li, Y., Zhang, Y., Li, S., Lu, J., Chen, J., Wang, Y., . . . Li, X. (2015). Genome-wide DNA methylome analysis reveals epigenetically dysregulated non-coding RNAs in human breast cancer. *Sci Rep, 5*, 8790. doi:10.1038/srep08790

Lichtenstein, P., Holm, N. V., Verkasalo, P. K., Iliadou, A., Kaprio, J., Koskenvuo, M., . . . Hemminki, K. (2000). Environmental and heritable factors in the causation of cancer--analyses of cohorts of twins from Sweden, Denmark, and Finland. *N Engl J Med, 343*(2), 78-85. doi:10.1056/nejm200007133430201

Liebert, A., Jones, B. L., Danielsen, E. T., Olsen, A. K., Swallow, D. M., & Troelsen, J. T. (2016). In Vitro Functional Analyses of Infrequent Nucleotide Variants in the Lactase Enhancer Reveal Different Molecular Routes to Increased Lactase Promoter Activity and Lactase Persistence. *Ann Hum Genet*. doi:10.1111/ahg.12167

Liebert, A., López, S., Jones, B. L., Montalva, N., Gerbault, P., Lau, W., . . . Swallow, D. M. (2017). World-wide distributions of lactase persistence alleles and the complex effects of recombination and selection. *Hum Genet, 136*(11), 1445-1453. doi:10.1007/s00439-017-1847-y

Lin, Q., & Wagner, W. (2015). Epigenetic aging signatures are coherently modified in cancer. *PLoS Genet, 11*(6), e1005334-e1005334. doi:10.1371/journal.pgen.1005334

Lin, Y. Y., Lu, J. Y., Zhang, J., Walter, W., Dang, W., Wan, J., . . . Zhu, H. (2009). Protein acetylation microarray reveals that NuA4 controls key metabolic target regulating gluconeogenesis. *Cell, 136*(6), 1073-1084. doi:10.1016/j.cell.2009.01.033

Lister, R., Pelizzola, M., Dowen, R. H., Hawkins, R. D., Hon, G., Tonti-Filippini, J., . . . Ecker, J. R. (2009). Human DNA methylomes at base resolution show widespread epigenomic differences. *Nature, 462*(7271), 315-322. doi:10.1038/nature08514

Lisy, M. E., Jarvis, K. B., DelBello, M. P., Mills, N. P., Weber, W. A., Fleck, D., . . . Adler, C. M. (2011). Progressive neurostructural changes in adolescent and adult patients with bipolar disorder. *Bipolar Disord, 13*(4), 396-405. doi:10.1111/j.1399-5618.2011.00927.x

Liu, Y., Balaraman, Y., Wang, G., Nephew, K. P., & Zhou, F. C. (2009). Alcohol exposure alters DNA methylation profiles in mouse embryos at early neurulation. *Epigenetics, 4*(7), 500-511.

Liu, Z., Wang, L., Welch, J. D., Ma, H., Zhou, Y., Vaseghi, H. R., . . . Qian, L. (2017). Single-cell transcriptomics reconstructs fate conversion from fibroblast to cardiomyocyte. *Nature, 551*(7678), 100-104. doi:10.1038/nature24454

Lobanenkov, V. V., Nicolas, R. H., Adler, V. V., Paterson, H., Klenova, E. M., Polotskaja, A. V., & Goodwin, G. H. (1990). A novel sequence-specific DNA binding protein which interacts with three regularly spaced direct repeats of the CCCTC-motif in the 5'-flanking sequence of the chicken c-myc gene. *Oncogene, 5*(12), 1743-1753.

Lopez-Otin, C., Blasco, M. A., Partridge, L., Serrano, M., & Kroemer, G. (2013). The hallmarks of aging. *Cell, 153*(6), 1194-1217. doi:10.1016/j.cell.2013.05.039

Lorenzo, A. D., & Bedford, M. T. (2011). Histone Arginine Methylation. *FEBS Lett, 585*(13), 2024-2031. doi:10.1016/j.febslet.2010.11.010

Lovatt, D., Ruble, B. K., Lee, J., Dueck, H., Kim, T. K., Fisher, S., . . . Eberwine, J. (2014). Transcriptome in vivo analysis (TIVA) of spatially defined single cells in live tissue. *Nat Methods, 11*(2), 190-196. doi:10.1038/nmeth.2804

Machado-Vieira, R., Dietrich, M. O., Leke, R., Cereser, V. H., Zanatto, V., Kapczinski, F., . . . Gentil, V. (2007). Decreased plasma brain derived neurotrophic factor levels in unmedicated bipolar patients during manic episode. *Biol Psychiatry, 61*(2), 142-144. doi:10.1016/j.biopsych.2006.03.070

Machin, G. A. (1996). Some causes of genotypic and phenotypic discordance in monozygotic twin pairs. *Am J Med Genet, 61*(3), 216-228. doi:10.1002/(sici)1096-8628(19960122)61:3<216::aid-ajmg5>3.0.co;2-s

Magistri, M., Faghihi, M. A., St Laurent, G., 3rd, & Wahlestedt, C. (2012). Regulation of chromatin structure by long noncoding RNAs: focus on natural antisense transcripts. *Trends Genet, 28*(8), 389-396. doi:10.1016/j.tig.2012.03.013

Malhi, G. S., Tanious, M., Das, P., Coulston, C. M., & Berk, M. (2013). Potential mechanisms of action of lithium in bipolar disorder. Current understanding. *CNS Drugs, 27*(2), 135-153. doi:10.1007/s40263-013-0039-0

Martinez-Aran, A., Vieta, E., Colom, F., Torrent, C., Reinares, M., Goikolea, J. M., . . . Sanchez-Moreno, J. (2005). Do cognitive complaints in euthymic bipolar patients reflect objective cognitive impairment? *Psychother Psychosom, 74*(5), 295-302. doi:10.1159/000086320

Martinez-Aran, A., Vieta, E., Torrent, C., Sanchez-Moreno, J., Goikolea, J. M., Salamero, M., . . . Ayuso-Mateos, J. L. (2007). Functional outcome in bipolar disorder: the role of clinical and cognitive factors. *Bipolar Disord, 9*(1-2), 103-113. doi:10.1111/j.1399-5618.2007.00327.x

Mathews, D. C., Richards, E. M., Niciu, M. J., Ionescu, D. F., Rasimas, J. J., & Zarate, C. A. (2013). Neurobiological Aspects of Suicide and Suicide Attempts in Bipolar Disorder. *Translational neuroscience, 4*(2), 1-24. doi:10.2478/s13380-013-0120-7

Mattar, R., de Campos Mazo, D. F., & Carrilho, F. J. (2012). Lactose intolerance: diagnosis, genetic, and clinical factors. *Clin Exp Gastroenterol, 5*, 113-121. doi:10.2147/ceg.s32368

Maussion, G., Yang, J., Suderman, M., Diallo, A., Nagy, C., Arnovitz, M., . . . Turecki, G. (2014). Functional DNA methylation in a transcript specific 3' UTR region of TrkB associates with suicide. *Epigenetics, 9*(8), 1061-1070. doi:10.4161/epi.29068

Maze, I., Wenderski, W., Noh, K. M., Bagot, R. C., Tzavaras, N., Purushothaman, I., . . . Allis, C. D. (2015). Critical Role of Histone Turnover in Neuronal Transcription and Plasticity. *Neuron, 87*(1), 77-94. doi:10.1016/j.neuron.2015.06.014

McCabe, M. T., Ott, H. M., Ganji, G., Korenchuk, S., Thompson, C., Van Aller, G. S., . . . Creasy, C. L. (2012). EZH2 inhibition as a therapeutic strategy for lymphoma with EZH2-activating mutations. *Nature, 492*(7427), 108-112. doi:10.1038/nature11606

McGowan, H. (2016). Bridging the Gap between DNA Methylation, DNA Methylation Readers, and Neurodevelopmental Disorders. *J Neurosci, 36*(26), 6851-6853. doi:10.1523/jneurosci.1130-16.2016

McKay, J. A., & Mathers, J. C. (2011). Diet induced epigenetic changes and their implications for health. *Acta Physiol (Oxf), 202*(2), 103-118. doi:10.1111/j.1748-1716.2011.02278.x

Meier, I., Fellini, L., Jakovcevski, M., Schachner, M., & Morellini, F. (2010). Expression of the snoRNA host gene gas5 in the hippocampus is upregulated by age and psychogenic stress and correlates with reduced novelty-induced behavior in C57BL/6 mice. *Hippocampus, 20*(9), 1027-1036. doi:10.1002/hipo.20701

Merikangas, K. R., Jin, R., He, J. P., Kessler, R. C., Lee, S., Sampson, N. A., . . . Zarkov, Z. (2011). Prevalence and correlates of bipolar spectrum disorder in the world mental health survey initiative. *Arch Gen Psychiatry, 68*(3), 241-251. doi:10.1001/archgenpsychiatry.2011.12

Merry, C. R., Forrest, M. E., Sabers, J. N., Beard, L., Gao, X. H., Hatzoglou, M., . . . Khalil, A. M. (2015). DNMT1-associated long non-coding RNAs regulate global gene expression and DNA methylation in colon cancer. *Hum Mol Genet, 24*(21), 6240-6253. doi:10.1093/hmg/ddv343

Middendorp, S., Schneeberger, K., Wiegerinck, C. L., Mokry, M., Akkerman, R. D., van Wijngaarden, S., . . . Nieuwenhuis, E. E. (2014). Adult stem cells in the small intestine are intrinsically programmed with their location-specific function. *Stem Cells, 32*(5), 1083-1091. doi:10.1002/stem.1655

Migliaccio, E., Giorgio, M., Mele, S., Pelicci, G., Reboldi, P., Pandolfi, P. P., . . . Pelicci, P. G. (1999). The p66shc adaptor protein controls oxidative stress response and life span in mammals. *Nature, 402*(6759), 309-313. doi:10.1038/46311

Milne, T. A., Zhao, K., & Hess, J. L. (2009). Chromatin immunoprecipitation (ChIP) for analysis of histone modifications and chromatin-associated proteins. *Methods Mol Biol, 538*, 409-423. doi:10.1007/978-1-59745-418-6_21

Minks, J., Robinson, W. P., & Brown, C. J. (2008). A skewed view of X chromosome inactivation. *J Clin Invest, 118*(1), 20-23. doi:10.1172/jci34470

Modabbernia, A., Taslimi, S., Brietzke, E., & Ashrafi, M. (2013). Cytokine alterations in bipolar disorder: a meta-analysis of 30 studies. *Biol Psychiatry, 74*(1), 15-25. doi:10.1016/j.biopsych.2013.01.007

Momparler, R. L. (2005). Epigenetic therapy of cancer with 5-aza-2'-deoxycytidine (decitabine). *Semin Oncol, 32*(5), 443-451. doi:10.1053/j.seminoncol.2005.07.008

Moosavi, A., & Motevalizadeh Ardekani, A. (2016). Role of Epigenetics in Biology and Human Diseases. *Iran Biomed J, 20*(5), 246-258.

Morris, S. A., Rao, B., Garcia, B. A., Hake, S. B., Diaz, R. L., Shabanowitz, J., . . . Strahl, B. D. (2007). Identification of histone H3 lysine 36 acetylation as a highly conserved histone modification. *J Biol Chem, 282*(10), 7632-7640. doi:10.1074/jbc.M607909200

Mosammaparast, N., & Shi, Y. (2010). Reversal of histone methylation: biochemical and molecular mechanisms of histone demethylases. *Annu Rev Biochem, 79*, 155-179. doi:10.1146/annurev.biochem.78.070907.103946

Movassagh, M., Choy, M. K., Goddard, M., Bennett, M. R., Down, T. A., & Foo, R. S. (2010). Differential DNA methylation correlates with differential expression of angiogenic factors in human heart failure. *PLoS ONE, 5*(1), e8564. doi:10.1371/journal.pone.0008564

Movassagh, M., Choy, M. K., Knowles, D. A., Cordeddu, L., Haider, S., Down, T., . . . Foo, R. S. (2011). Distinct epigenomic features in end-stage failing human hearts. *Circulation, 124*(22), 2411-2422. doi:10.1161/circulationaha.111.040071

Müller-Oerlinghausen, B., Berghöfer, A., & Bauer, M. (2002). Bipolar disorder. *Lancet, 359*(9302), 241-247. doi:10.1016/S0140-6736(02)07450-0

Munsky, B., Neuert, G., & van Oudenaarden, A. (2012). Using gene expression noise to understand gene regulation. *Science, 336*(6078), 183-187. doi:10.1126/science.1216379

Murphy, S. K., Adigun, A., Huang, Z., Overcash, F., Wang, F., Jirtle, R. L., . . . Hoyo, C. (2012). Gender-Specific Methylation Differences in Relation to Prenatal Exposure to Cigarette Smoke. *Gene, 494*(1), 36-43. doi:10.1016/j.gene.2011.11.062

Nabel, E. G., & Braunwald, E. (2012). A tale of coronary artery disease and myocardial infarction. *N Engl J Med, 366*(1), 54-63. doi:10.1056/NEJMra1112570

Nakatochi, M., Ichihara, S., Yamamoto, K., Ohnaka, K., Kato, Y., Yokota, S., . . . Yokota, M. (2015). Epigenome-wide association study suggests that SNPs in the promoter region of RETN influence plasma resistin level via effects on DNA methylation at neighbouring sites. *Diabetologia, 58*(12), 2781-2790. doi:10.1007/s00125-015-3763-9

Nam, Y.-y., Kim, C.-h., & Roh, D. (2016). ScienceDirect Comorbid panic disorder as an independent risk factor for suicide attempts in depressed outpatients. *Comprehensive Psychiatry*(67), 13-18. doi:10.1016/j.comppsych.2016.02.011

Nelson, C. J., Santos-Rosa, H., & Kouzarides, T. (2006). Proline isomerization of histone H3 regulates lysine methylation and gene expression. *Cell, 126*(5), 905-916. doi:10.1016/j.cell.2006.07.026

Nepon, J., Belik, S.-L., Bolton, J., & Sareen, J. (2010). The Relationship Between Anxiety Disorders and Suicide Attempts: Findings from the National Epidemiologic Survey on Alcohol and Related Conditions. *Depress Anxiety, 27*(9), 791-798. doi:10.1002/da.20674

Neumann, E., Riepl, B., Knedla, A., Lefèvre, S., Tarner, I. H., Grifka, J., . . . Müller-Ladner, U. (2010). Cell culture and passaging alters gene expression pattern and proliferation rate in rheumatoid arthritis synovial fibroblasts. *Arthritis Res Ther, 12*(3), R83. doi:10.1186/ar3010

Newcomer, A. D., McGill, D. B., Thomas, P. J., & Hofmann, A. F. (1975). Prospective comparison of indirect methods for detecting lactase deficiency. *N Engl J Med, 293*(24), 1232-1236. doi:10.1056/nejm197512112932405

Ng, S. F., Lin, R. C., Laybutt, D. R., Barres, R., Owens, J. A., & Morris, M. J. (2010). Chronic high-fat diet in fathers programs beta-cell dysfunction in female rat offspring. *Nature, 467*(7318), 963-966. doi:10.1038/nature09491

Ni, Z., Ebata, A., Alipanahiramandi, E., & Lee, S. S. (2012). Two SET domain containing genes link epigenetic changes and aging in Caenorhabditis elegans. *Aging Cell, 11*(2), 315-325. doi:10.1111/j.1474-9726.2011.00785.x

Nora, E. P., Lajoie, B. R., Schulz, E. G., Giorgetti, L., Okamoto, I., Servant, N., . . . Heard, E. (2012). Spatial partitioning of the regulatory landscape of the X-inactivation centre. *Nature, 485*(7398), 381-385. doi:10.1038/nature11049

O'Sullivan, R. J., Kubicek, S., Schreiber, S. L., & Karlseder, J. (2010). Reduced histone biosynthesis and chromatin changes arising from a damage signal at telomeres. *Nat Struct Mol Biol, 17*(10), 1218-1225. doi:10.1038/nsmb.1897

Oberdoerffer, P., Michan, S., McVay, M., Mostoslavsky, R., Vann, J., Park, S. K., . . . Sinclair, D. A. (2008). SIRT1 redistribution on chromatin promotes genomic stability but alters gene expression during aging. *Cell, 135*(5), 907-918. doi:10.1016/j.cell.2008.10.025

Oberdoerffer, P., & Sinclair, D. A. (2007). The role of nuclear architecture in genomic instability and ageing. *Nat Rev Mol Cell Biol, 8*(9), 692-702. doi:10.1038/nrm2238

Oh, G., Ebrahimi, S., Carlucci, M., Zhang, A., Nair, A., Groot, D. E., . . . Petronis, A. (2018). Cytosine modifications exhibit circadian oscillations that are involved in epigenetic diversity and aging. *Nat Commun, 9*(1), 644. doi:10.1038/s41467-018-03073-7

Oh, G., & Petronis, A. (2008). Environmental studies of schizophrenia through the prism of epigenetics. *Schizophr Bull, 34*(6), 1122-1129. doi:10.1093/schbul/sbn105

Okano, M., Bell, D. W., Haber, D. A., & Li, E. (1999). DNA methyltransferases Dnmt3a and Dnmt3b are essential for de novo methylation and mammalian development. *Cell, 99*(3), 247-257.

Oki, M., Aihara, H., & Ito, T. (2007). Role of histone phosphorylation in chromatin dynamics and its implications in diseases. *Subcell Biochem, 41*, 319-336.

Olds, L. C., & Sibley, E. (2003). Lactase persistence DNA variant enhances lactase promoter activity in vitro: functional role as a cis regulatory element. *Hum Mol Genet, 12*(18), 2333-2340. doi:10.1093/hmg/ddg244

Ollikainen, M., Smith, K. R., Joo, E. J., Ng, H. K., Andronikos, R., Novakovic, B., . . . Craig, J. M. (2010). DNA methylation analysis of multiple tissues from newborn twins reveals both genetic and intrauterine components to variation in the human neonatal epigenome. *Hum Mol Genet, 19*(21), 4176-4188. doi:10.1093/hmg/ddq336

Ong, C. T., & Corces, V. G. (2011). Enhancer function: new insights into the regulation of tissue-specific gene expression. *Nat Rev Genet, 12*(4), 283-293. doi:10.1038/nrg2957

Ong, C. T., & Corces, V. G. (2014). CTCF: an architectural protein bridging genome topology and function. *Nat Rev Genet, 15*(4), 234-246. doi:10.1038/nrg3663

Pal, S., & Tyler, J. K. (2016). Epigenetics and aging. *Sci Adv, 2*(7), e1600584. doi:10.1126/sciadv.1600584

Panhuysen, C. I., Vonk, J. M., Koeter, G. H., Schouten, J. P., van Altena, R., Bleecker, E. R., & Postma, D. S. (1997). Adult patients may outgrow their asthma: a 25-year follow-up study. *Am J Respir Crit Care Med, 155*(4), 1267-1272. doi:10.1164/ajrccm.155.4.9105065

Park, J. H., Stoffers, D. A., Nicholls, R. D., & Simmons, R. A. (2008). Development of type 2 diabetes following intrauterine growth retardation in rats is associated with progressive epigenetic silencing of Pdx1. *J Clin Invest, 118*(6), 2316-2324. doi:10.1172/jci33655

Patel, A. P., Tirosh, I., Trombetta, J. J., Shalek, A. K., Gillespie, S. M., Wakimoto, H., . . . Bernstein, B. E. (2014). Single-cell RNA-seq highlights intratumoral heterogeneity in primary glioblastoma. *Science, 344*(6190), 1396-1401. doi:10.1126/science.1254257

Pegoraro, G., Kubben, N., Wickert, U., Gohler, H., Hoffmann, K., & Misteli, T. (2009). Ageing-related chromatin defects through loss of the NURD complex. *Nat Cell Biol, 11*(10), 1261-1267. doi:10.1038/ncb1971

Peleg, S., Sananbenesi, F., Zovoilis, A., Burkhardt, S., Bahari-Javan, S., Agis-Balboa, R. C., . . . Fischer, A. (2010). Altered histone acetylation is associated with age-dependent memory impairment in mice. *Science, 328*(5979), 753-756. doi:10.1126/science.1186088

Perlis, R. H., Ostacher, M. J., Uher, R., Nierenberg, A. A., Casamassima, F., Kansky, C., . . . Sachs, G. S. (2009). Stability of symptoms across major depressive episodes in bipolar disorder. *Bipolar Disord, 11*(8), 867-875. doi:10.1111/j.1399-5618.2009.00764.x

Petronis, A. (2001). Human morbid genetics revisited: relevance of epigenetics. *Trends Genet, 17*(3), 142-146.

Piccinni, A., Origlia, N., Veltri, A., Vizzaccaro, C., Marazziti, D., Catena-Dell'osso, M., . . . Dell'osso, L. (2012). Plasma beta-amyloid peptides levels: a pilot study in bipolar depressed patients. *J Affect Disord, 138*(1-2), 160-164. doi:10.1016/j.jad.2011.12.042

Platt, R. J., Chen, S., Zhou, Y., Yim, M. J., Swiech, L., Kempton, H. R., . . . Zhang, F. (2014). CRISPR-Cas9 knockin mice for genome editing and cancer modeling. *Cell, 159*(2), 440-455. doi:10.1016/j.cell.2014.09.014

Post, W. S., Goldschmidt-Clermont, P. J., Wilhide, C. C., Heldman, A. W., Sussman, M. S., Ouyang, P., . . . Issa, J. P. (1999). Methylation of the estrogen receptor gene is associated with aging and atherosclerosis in the cardiovascular system. *Cardiovasc Res, 43*(4), 985-991.

Poulter, M., Hollox, E., Harvey, C. B., Mulcare, C., Peuhkuri, K., Kajander, K., . . . Swallow, D. M. (2003). The causal element for the lactase persistence/non-persistence polymorphism is located in a 1 Mb region of linkage disequilibrium in Europeans. *Ann Hum Genet, 67*(Pt 4), 298-311.

Proctor, C. J., & Kirkwood, T. B. (2002). Modelling telomere shortening and the role of oxidative stress. *Mech Ageing Dev, 123*(4), 351-363.

Pufulete, M., Al-Ghnaniem, R., Khushal, A., Appleby, P., Harris, N., Gout, S., . . . Sanders, T. A. (2005). Effect of folic acid supplementation on genomic DNA methylation in patients with colorectal adenoma. *Gut, 54*(5), 648-653. doi:10.1136/gut.2004.054718

Purcell, S., Neale, B., Todd-brown, K., Thomas, L., Ferreira, M. A. R., Bender, D., . . . Sham, P. C. (2007). PLINK : A Tool Set for Whole-Genome Association and Population-Based Linkage Analyses. *Am J Hum Genet., 81*(3), 559-575. doi:10.1086/519795

Quarrie, J. K., & Riabowol, K. T. (2004). Murine models of life span extension. *Sci Aging Knowledge Environ, 2004*(31), re5. doi:10.1126/sageke.2004.31.re5

Quinlan, A. R., & Hall, I. M. (2010). BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics, 26*(6), 841-842. doi:10.1093/bioinformatics/btq033

Ran, F. A., Hsu, P. D., Lin, C. Y., Gootenberg, J. S., Konermann, S., Trevino, A. E., . . . Zhang, F. (2013). Double nicking by RNA-guided CRISPR Cas9 for enhanced genome editing specificity. *Cell, 154*(6), 1380-1389. doi:10.1016/j.cell.2013.08.021

Ran, F. A., Hsu, P. D., Wright, J., Agarwala, V., Scott, D. A., & Zhang, F. (2013). Genome engineering using the CRISPR-Cas9 system. *Nat Protoc, 8*(11), 2281-2308. doi:10.1038/nprot.2013.143

Ranciaro, A., Campbell, M. C., Hirbo, J. B., Ko, W. Y., Froment, A., Anagnostou, P., . . . Tishkoff, S. A. (2014). Genetic origins of lactase persistence and the spread of pastoralism in Africa. *Am J Hum Genet, 94*(4), 496-510. doi:10.1016/j.ajhg.2014.02.009

Rasinpera, H., Forsblom, C., Enattah, N. S., Halonen, P., Salo, K., Victorzon, M., . . . Jarvela, I. (2005). The C/C-13910 genotype of adult-type hypolactasia is associated with an increased risk of colorectal cancer in the Finnish population. *Gut, 54*(5), 643-647. doi:10.1136/gut.2004.055939

Rasinpera, H., Kuokkanen, M., Kolho, K. L., Lindahl, H., Enattah, N. S., Savilahti, E., . . . Jarvela, I. (2005). Transcriptional downregulation of the lactase (LCT) gene during childhood. *Gut, 54*(11), 1660-1661. doi:10.1136/gut.2005.077404

Raule, N., Sevini, F., Li, S., Barbieri, A., Tallaro, F., Lomartire, L., . . . Franceschi, C. (2014). The co-occurrence of mtDNA mutations on different oxidative phosphorylation subunits, not detected by haplogroup analysis, affects human longevity and is population specific. *Aging Cell, 13*(3), 401-407. doi:10.1111/acel.12186

Rayner, K. J., Suarez, Y., Davalos, A., Parathath, S., Fitzgerald, M. L., Tamehiro, N., . . . Fernandez-Hernando, C. (2010). MiR-33 contributes to the regulation of cholesterol homeostasis. *Science, 328*(5985), 1570-1573. doi:10.1126/science.1189862

Reilly, M. T., Faulkner, G. J., Dubnau, J., Ponomarev, I., & Gage, F. H. (2013). The role of transposable elements in health and diseases of the central nervous system. *J Neurosci, 33*(45), 17577-17586. doi:10.1523/jneurosci.3369-13.2013

Relton, C. L., & Davey Smith, G. (2010). Epigenetic epidemiology of common complex disease: prospects for prediction, prevention, and treatment. *PLoS Med, 7*(10), e1000356. doi:10.1371/journal.pmed.1000356

Rings, E. H., Krasinski, S. D., van Beers, E. H., Moorman, A. F., Dekker, J., Montgomery, R. K., . . . Buller, H. A. (1994). Restriction of lactase gene expression along the proximal-to-distal axis of rat small intestine occurs during postnatal development. *Gastroenterology, 106*(5), 1223-1232.

Rizzo, L. B., Costa, L. G., Mansur, R. B., Swardfager, W., Belangero, S. I., Grassi-Oliveira, R., . . . Brietzke, E. (2014). The theory of bipolar disorder as an illness of accelerated aging: Implications for clinical care and research. *Neuroscience and Biobehavioral Reviews, 42*(Complete), 157-169. doi:10.1016/j.neubiorev.2014.02.004

Roadmap Epigenomics, C., Kundaje, A., Meuleman, W., Ernst, J., Bilenky, M., Yen, A., . . . Kellis, M. (2015). Integrative analysis of 111 reference human epigenomes. *Nature, 518*(7539), 317-330. doi:10.1038/nature14248

Robertson, K. D. (2005). DNA methylation and human disease. *Nat Rev Genet, 6*(8), 597-610. doi:10.1038/nrg1655

Robertson, R. P., Harmon, J., Tran, P. O., & Poitout, V. (2004). Beta-cell glucose toxicity, lipotoxicity, and chronic oxidative stress in type 2 diabetes. *Diabetes, 53 Suppl 1*, S119-124.

Robinson, L. J., Thompson, J. M., Gallagher, P., Goswami, U., Young, A. H., Ferrier, I. N., & Moore, P. B. (2006). A meta-analysis of cognitive deficits in euthymic patients with bipolar disorder. *J Affect Disord, 93*(1-3), 105-115. doi:10.1016/j.jad.2006.02.016

Rodríguez-Rodero, S., Fernández-Morera, J. L., Menéndez-Torre, E., Calvanese, V., Fernández, A. F., & Fraga, M. F. (2011). Aging Genetics and Aging. *Aging Dis, 2*(3), 186-195.

Rogina, B., Helfand, S. L., & Frankel, S. (2002). Longevity regulation by Drosophila Rpd3 deacetylase and caloric restriction. *Science, 298*(5599), 1745. doi:10.1126/science.1078986

Rosa, A., Picchioni, M. M., Kalidindi, S., Loat, C. S., Knight, J., Toulopoulou, T., . . . Craig, I. W. (2008). Differential methylation of the X-chromosome is a possible source of discordance for bipolar disorder female monozygotic twins. *American Journal of Medical Genetics, Part B: Neuropsychiatric Genetics, 147*(4), 459-462. doi:10.1002/ajmg.b.30616

Rose, M. R., Flatt, T., Graves, J. L., Greer, L. F., Martinez, D. E., Matos, M., . . . Shahrestani, P. (2012). What is Aging? *Front Genet, 3*. doi:10.3389/fgene.2012.00134

Royce, T. E., Carriero, N. J., & Gerstein, M. B. (2007). An efficient pseudomedian filter for tiling microrrays. *BMC Bioinformatics, 8*, 186. doi:10.1186/1471-2105-8-186

Ruggiero, C., Metter, E. J., Melenovsky, V., Cherubini, A., Najjar, S. S., Ble, A., . . . Ferrucci, L. (2008). High basal metabolic rate is a risk factor for mortality: the Baltimore Longitudinal Study of Aging. *J Gerontol A Biol Sci Med Sci, 63*(7), 698-706.

Ryu, S. H., Kang, K., Yoo, T., Joe, C. O., & Chung, J. H. (2011). Transcriptional repression of repeat-derived transcripts correlates with histone hypoacetylation at repetitive DNA elements in aged mice brain. *Exp Gerontol, 46*(10), 811-818. doi:10.1016/j.exger.2011.07.001

Sahi, T. (1994). Genetics and epidemiology of adult-type hypolactasia. *Scand J Gastroenterol Suppl, 202*, 7-20.

Sahi, T., Isokoski, M., Jussila, J., & Launiala, K. (1972). Lactose malabsorption in Finnish children of school age. *Acta Paediatr Scand, 61*(1), 11-16.

Sahi, T., Launiala, K., & Laitinen, H. (1983). Hypolactasia in a fixed cohort of young Finnish adults. A follow-up study. *Scand J Gastroenterol, 18*(7), 865-870.

Samani, N. J., Boultby, R., Butler, R., Thompson, J. R., & Goodall, A. H. (2001). Telomere shortening in atherosclerosis. *Lancet, 358*(9280), 472-473. doi:10.1016/s0140-6736(01)05633-1

Sandoval, J., Heyn, H., Moran, S., Serra-Musach, J., Pujana, M. A., Bibikova, M., & Esteller, M. (2011). Validation of a DNA methylation microarray for 450,000 CpG sites in the human genome. *Epigenetics, 6*(6), 692-702.

Scheller, E. L., Doucette, C. R., Learman, B. S., Cawthorn, W. P., Khandaker, S., Schell, B., . . . MacDougald, O. A. (2015). Region-specific variation in the properties of skeletal adipocytes reveals regulated and constitutive marrow adipose tissues. *Nat Commun, 6*, 7808. doi:10.1038/ncomms8808

Schirru, E., Corona, V., Usai-Satta, P., Scarpa, M., Cucca, F., De Virgiliis, S., . . . Congia, M. (2007). Decline of lactase activity and c/t-13910 variant in Sardinian childhood. *J Pediatr Gastroenterol Nutr, 45*(4), 503-506. doi:10.1097/MPG.0b013e31805b5899

Schneck, C. D., Miklowitz, D. J., Calabrese, J. R., Allen, M. H., Thomas, M. R., Wisniewski, S. R., . . . Sachs, G. S. (2004). Phenomenology of rapid-cycling bipolar disorder: data from the first 500 participants in the Systematic Treatment Enhancement Program. *Am J Psychiatry, 161*(10), 1902-1908. doi:10.1176/ajp.161.10.1902

Schosser, A., Butler, A. W., Ising, M., Perroud, N., Uher, R., Ng, M. Y., . . . Lewis, C. M. (2011). Genomewide association scan of suicidal thoughts and behaviour in major depression. *PLoS ONE, 6*(7). doi:10.1371/journal.pone.0020690

Scott, L. J., Muglia, P., Kong, X. Q., Guan, W., Flickinger, M., Upmanyu, R., . . . Boehnke, M. (2009). Genome-wide association and meta-analysis of bipolar disorder in individuals of European ancestry. *Proc Natl Acad Sci U S A, 106*(18), 7501-7506. doi:10.1073/pnas.0813386106

Sedivy, J. M., Banumathy, G., & Adams, P. D. (2008). Aging by epigenetics--a consequence of chromatin damage? *Exp Cell Res, 314*(9), 1909-1917. doi:10.1016/j.yexcr.2008.02.023

See, K., Tan, W. L. W., Lim, E. H., Tiang, Z., Lee, L. T., Li, P. Y. Q., . . . Foo, R. S. (2017). Single cardiomyocyte nuclear transcriptomes reveal a lincRNA-regulated de-differentiation and cell cycle stress-response in vivo. *Nat Commun, 8*(1), 225. doi:10.1038/s41467-017-00319-8

Sen, P., Shah, P. P., Nativio, R., & Berger, S. L. (2016). Epigenetic Mechanisms of Longevity and Aging. *Cell, 166*(4), 822-839. doi:10.1016/j.cell.2016.07.050

Seppo, L., Tuure, T., Korpela, R., Jarvela, I., Rasinpera, H., & Sahi, T. (2008). Can primary hypolactasia manifest itself after the age of 20 years? A two-decade follow-up study. *Scand J Gastroenterol, 43*(9), 1082-1087. doi:10.1080/00365520802095485

Shammas, M. A. (2011). Telomeres, lifestyle, cancer, and aging. *Curr Opin Clin Nutr Metab Care, 14*(1), 28-34. doi:10.1097/MCO.0b013e32834121b1

Sharma, A., Klein, S. L., Barboza, L., Lodhi, N., & Toth, M. (2016). Principles Governing DNA Methylation during Neuronal Lineage and Subtype Specification. *J Neurosci, 36*(5), 1711-1722. doi:10.1523/jneurosci.4037-15.2016

Sharma, S., Kelly, T. K., & Jones, P. A. (2010). Epigenetics in cancer. *Carcinogenesis, 31*(1), 27-36. doi:10.1093/carcin/bgp220

Shaw, P., Eckstrand, K., Sharp, W., Blumenthal, J., Lerch, J. P., Greenstein, D., . . . Rapoport, J. L. (2007). Attention-deficit/hyperactivity disorder is characterized by a delay in cortical maturation. *Proc Natl Acad Sci U S A, 104*(49), 19649-19654. doi:10.1073/pnas.0707741104

Sheaffer, K. L., Kim, R., Aoki, R., Elliott, E. N., Schug, J., Burger, L., . . . Kaestner, K. H. (2014). DNA methylation is required for the control of stem cell differentiation in the small intestine. *Genes Dev, 28*(6), 652-664. doi:10.1101/gad.230318.113

Shibata, D., Peinado, M. A., Ionov, Y., Malkhosyan, S., & Perucho, M. (1994). Genomic instability in repeated sequences is an early somatic event in colorectal tumorigenesis that persists after transformation. *Nat Genet, 6*(3), 273-281. doi:10.1038/ng0394-273

Shiio, Y., & Eisenman, R. N. (2003). Histone sumoylation is associated with transcriptional repression. *Proc Natl Acad Sci U S A, 100*(23), 13225-13230. doi:10.1073/pnas.1735528100

Shinozaki, G., & Potash, J. B. (2014). New developments in the genetics of bipolar disorder. *Curr Psychiatry Rep, 16*(11), 493. doi:10.1007/s11920-014-0493-5

Shlush, L. I., Zandi, S., Mitchell, A., Chen, W. C., Brandwein, J. M., Gupta, V., . . . Dick, J. E. (2014). Identification of pre-leukaemic haematopoietic stem cells in acute leukaemia. *Nature, 506*(7488), 328-333. doi:10.1038/nature13038

Shukla, S., Kavak, E., Gregory, M., Imashimizu, M., Shutinoski, B., Kashlev, M., . . . Oberdoerffer, S. (2011). CTCF-promoted RNA polymerase II pausing links DNA methylation to splicing. *Nature, 479*(7371), 74-79. doi:10.1038/nature10442

Simoons, F. J. (1980). Age of onset of lactose malabsorption. *Pediatrics, 66*(4), 646-648.

Singh, R., Kølvraa, S., Bross, P., Jensen, U. B., Gregersen, N., Tan, Q., . . . Rattan, S. I. S. (2006). Reduced heat shock response in human mononuclear cells during aging and its association with polymorphisms in HSP70 genes. *Cell Stress Chaperones, 11*(3), 208-215. doi:10.1379/csc-184r.1

Skelly, D. A., Squiers, G. T., McLellan, M. A., Bolisetty, M. T., Robson, P., Rosenthal, N. A., & Pinto, A. R. (2018). Single-Cell Transcriptional Profiling Reveals Cellular Diversity and Intercommunication in the Mouse Heart. *Cell Rep, 22*(3), 600-610. doi:10.1016/j.celrep.2017.12.072

Smoller, J. W., & Finn, C. T. (2003). Family, twin, and adoption studies of bipolar disorder. *Am J Med Genet C Semin Med Genet, 123c*(1), 48-58. doi:10.1002/ajmg.c.20013

Soerensen, M., Thinggaard, M., Nygaard, M., Dato, S., Tan, Q., Hjelmborg, J., . . . Christiansen, L. (2012). Genetic variation in TERT and TERC and human leukocyte telomere length and longevity: a cross-sectional and longitudinal analysis. *Aging Cell, 11*(2), 223-227. doi:10.1111/j.1474-9726.2011.00775.x

Sohal, R. S., & Forster, M. J. (2014). Caloric restriction and the aging process: a critique. *Free Radic Biol Med, 73*, 366-382. doi:10.1016/j.freeradbiomed.2014.05.015

Soreca, I., Fagiolini, A., Frank, E., Houck, P. R., Thompson, W. K., & Kupfer, D. J. (2008). Relationship of general medical burden, duration of illness and age in patients with bipolar I disorder. *J Psychiatr Res, 42*(11), 956-961. doi:10.1016/j.jpsychires.2007.10.009

Spaethling, J. M., Sanchez-Alavez, M., Lee, J., Xia, F. C., Dueck, H., Wang, W., . . . Eberwine, J. (2016). Single-cell transcriptomics and functional target validation of brown adipocytes show their complex roles in metabolic homeostasis. *Faseb j, 30*(1), 81-92. doi:10.1096/fj.15-273797

Stadler, M. B., Murr, R., Burger, L., Ivanek, R., Lienert, F., Scholer, A., . . . Schubeler, D. (2011). DNA-binding factors shape the mouse methylome at distal regulatory regions. *Nature, 480*(7378), 490-495. doi:10.1038/nature10716

Stelzer, Y., Shivalila, C. S., Soldner, F., Markoulaki, S., & Jaenisch, R. (2015). Tracing dynamic changes of DNA methylation at single-cell resolution. *Cell, 163*(1), 218-229. doi:10.1016/j.cell.2015.08.046

Stohr, H., Molday, L. L., Molday, R. S., Weber, B. H., Biedermann, B., Reichenbach, A., & Kramer, F. (2005). Membrane-associated guanylate kinase proteins MPP4 and MPP5 associate with Veli3 at distinct intercellular junctions of the neurosensory retina. *J Comp Neurol, 481*(1), 31-41. doi:10.1002/cne.20367

Stöhr, H., Stojic, J., & Weber, B. H. (2003). Cellular localization of the MPP4 protein in the mammalian retina. *Investigative Ophthalmology & Visual Science, 44*(12), 5067-5074. doi:10.1167/iovs.03-0702

Stöhr, H., & Weber, B. H. (2001). Cloning and characterization of the human retina-specific gene MPP4, a novel member of the p55 subfamily of MAGUK proteins. *Genomics, 74*(3), 377-384. doi:10.1006/geno.2001.6559

Strakowski, S. M., DelBello, M. P., Zimmerman, M. E., Getz, G. E., Mills, N. P., Ret, J., . . . Adler, C. M. (2002). Ventricular and periventricular structural volumes in first- versus multiple-episode bipolar disorder. *Am J Psychiatry, 159*(11), 1841-1847. doi:10.1176/appi.ajp.159.11.1841

Sugimoto, S., & Sato, T. (2017). Establishment of 3D Intestinal Organoid Cultures from Intestinal Stem Cells. *Methods Mol Biol, 1612*, 97-105. doi:10.1007/978-1-4939-7021-6_7

Suva, M. L., Riggi, N., & Bernstein, B. E. (2013). Epigenetic reprogramming in cancer. *Science, 339*(6127), 1567-1570. doi:10.1126/science.1230184

Tadesse, K., Yuen, R. C., & Leung, D. T. (1991). Late-onset hypolactasia in Hong Kong school children. *Ann Trop Paediatr, 11*(3), 289-292.

Tahiliani, M., Koh, K. P., Shen, Y., Pastor, W. A., Bandukwala, H., Brudno, Y., . . . Rao, A. (2009). Conversion of 5-Methylcytosine to 5-Hydroxymethylcytosine in Mammalian DNA by MLL Partner TET1. *Science, 324*(5929), 930-935. doi:10.1126/science.1170116

Tammen, S. A., Friso, S., & Choi, S. W. (2013). Epigenetics: the link between nature and nurture. *Mol Aspects Med, 34*(4), 753-764. doi:10.1016/j.mam.2012.07.018

Tie, F., Banerjee, R., Stratton, C. A., Prasad-Sinha, J., Stepanik, V., Zlobin, A., . . . Harte, P. J. (2009). CBP-mediated acetylation of histone H3 lysine 27 antagonizes Drosophila Polycomb silencing. *Development, 136*(18), 3131-3141. doi:10.1242/dev.037127

Tishkoff, S. A., Reed, F. A., Friedlaender, F. R., Ehret, C., Ranciaro, A., Froment, A., . . . Williams, S. M. (2009). The genetic structure and history of Africans and African Americans. *Science, 324*(5930), 1035-1044. doi:10.1126/science.1172257

Tishkoff, S. A., Reed, F. A., Ranciaro, A., Voight, B. F., Babbitt, C. C., Silverman, J. S., . . . Deloukas, P. (2007). Convergent adaptation of human lactase persistence in Africa and Europe. *Nat Genet, 39*(1), 31-40. doi:10.1038/ng1946

Tissenbaum, H. A. (2012). Genetics, life span, health span, and the aging process in Caenorhabditis elegans. *J Gerontol A Biol Sci Med Sci, 67*(5), 503-510. doi:10.1093/gerona/gls088

Tjeertes, J. V., Miller, K. M., & Jackson, S. P. (2009). Screen for DNA-damage-responsive histone modifications identifies H3K9Ac and H3K56Ac in human cells. *Embo j, 28*(13), 1878-1889. doi:10.1038/emboj.2009.119

Toperoff, G., Aran, D., Kark, J. D., Rosenberg, M., Dubnikov, T., Nissan, B., . . . Hellman, A. (2012). Genome-wide survey reveals predisposing diabetes type 2-related DNA methylation variations in human peripheral blood. *Hum Mol Genet, 21*(2), 371-383. doi:10.1093/hmg/ddr472

Troelsen, J. T., Olsen, J., Moller, J., & Sjostrom, H. (2003). An upstream polymorphism associated with lactase persistence has increased enhancer activity. *Gastroenterology, 125*(6), 1686-1694.

Tsurumi, A., & Li, W. X. (2012). Global heterochromatin loss: a unifying theory of aging? *Epigenetics, 7*(7), 680-688. doi:10.4161/epi.20540

Tukiainen, T., Villani, A. C., Yen, A., Rivas, M. A., Marshall, J. L., Satija, R., . . . MacArthur, D. G. (2017). Landscape of X chromosome inactivation across human tissues. *Nature, 550*(7675), 244-248. doi:10.1038/nature24265

Turecki, G. (2014). Epigenetics and suicidal behavior research pathways. *American Journal of Preventive Medicine, 47*(3S2), S144-151. doi:10.1016/j.amepre.2014.06.011

Turecki, G., Ernst, C., Jollant, F., Labonté, B., & Mechawar, N. (2012). The neurodevelopmental origins of suicidal behavior. *Trends in Neurosciences, 35*(1), 14-23. doi:10.1016/j.tins.2011.11.008

Van Meter, M., Kashyap, M., Rezazadeh, S., Geneva, A. J., Morello, T. D., Seluanov, A., & Gorbunova, V. (2014). SIRT6 represses LINE1 retrotransposons by ribosylating KAP1 but this repression fails with stress and age. *Nat Commun, 5*, 5011. doi:10.1038/ncomms6011

van Wering, H. M., Bosse, T., Musters, A., de Jong, E., de Jong, N., Hogen Esch, C. E., . . . Krasinski, S. D. (2004). Complex regulation of the lactase-phlorizin hydrolase promoter by GATA-4. *Am J Physiol Gastrointest Liver Physiol, 287*(4), G899-909. doi:10.1152/ajpgi.00150.2004

Vaquero, A., & Reinberg, D. (2009). Calorie restriction and the exercise of chromatin. *Genes Dev, 23*(16), 1849-1869. doi:10.1101/gad.1807009

Varley, K. E., Gertz, J., Bowling, K. M., Parker, S. L., Reddy, T. E., Pauli-Behn, F., . . . Myers, R. M. (2013). Dynamic DNA methylation across diverse human cell lines and tissues. *Genome Res, 23*(3), 555-567. doi:10.1101/gr.147942.112

Vasu, S., Shah, S., Orjalo, A., Park, M., Fischer, W. H., & Forbes, D. J. (2001). Novel vertebrate nucleoporins Nup133 and Nup160 play a role in mRNA export. *The Journal of cell biology, 155*(3), 339-354. doi:10.1083/jcb.200108007

Vizoso, M., Ferreira, H. J., Lopez-Serra, P., Carmona, F. J., Martínez-Cardús, A., Girotti, M. R., . . . Esteller, M. (2015). Epigenetic activation of a cryptic TBC1D16 transcript enhances melanoma progression by targeting EGFR. *Nature medicine, 21*(7), 741-750. doi:10.1038/nm.3863

Voight, B. F., Scott, L. J., Steinthorsdottir, V., Morris, A. P., Dina, C., Welch, R. P., . . . McCarthy, M. I. (2010). Twelve type 2 diabetes susceptibility loci identified through large-scale association analysis. *Nat Genet, 42*(7), 579-589. doi:10.1038/ng.609

Walker, E. M., Thompson, C. A., Kohlnhofer, B. M., Faber, M. L., & Battle, M. A. (2014). Characterization of the developing small intestine in the absence of either GATA4 or GATA6. *BMC Res Notes, 7*, 902. doi:10.1186/1756-0500-7-902

Wallace, J. A., & Felsenfeld, G. (2007). We gather together: insulators and genome organization. *Curr Opin Genet Dev, 17*(5), 400-407. doi:10.1016/j.gde.2007.08.005

Walton, E., Hass, J., Liu, J., Roffman, J. L., Bernardoni, F., Roessner, V., . . . Ehrlich, S. (2016). Correspondence of DNA methylation between blood and brain tissue and its application to schizophrenia research. *Schizophrenia Bulletin, 42*(2), 406-414. doi:10.1093/schbul/sbv074

Wang, D., Yan, L., Hu, Q., Sucheston, L. E., Higgins, M. J., Ambrosone, C. B., . . . Liu, S. (2012). IMA: An R package for high-throughput analysis of Illumina's 450K Infinium methylation data. *Bioinformatics, 28*(5), 729-730. doi:10.1093/bioinformatics/bts013

Wang, H., Maurano, M. T., Qu, H., Varley, K. E., Gertz, J., Pauli, F., . . . Stamatoyannopoulos, J. A. (2012). Widespread plasticity in CTCF occupancy linked to DNA methylation. *Genome Res, 22*(9), 1680-1688. doi:10.1101/gr.136101.111

Wang, J., Geesman, G. J., Hostikka, S. L., Atallah, M., Blackwell, B., Lee, E., . . . Lunyak, V. V. (2011). Inhibition of activated pericentromeric SINE/Alu repeat transcription in senescent human adult stem cells reinstates self-renewal. *Cell Cycle, 10*(17), 3016-3030. doi:10.4161/cc.10.17.17543

Wang, Y., Harvey, C. B., Hollox, E. J., Phillips, A. D., Poulter, M., Clay, P., . . . Swallow, D. M. (1998). The genetically programmed down-regulation of lactase in children. *Gastroenterology, 114*(6), 1230-1236.

Wang, Y., Wysocka, J., Sayegh, J., Lee, Y. H., Perlin, J. R., Leonelli, L., . . . Coonrod, S. A. (2004). Human PAD4 regulates histone arginine methylation levels via demethylimination. *Science, 306*(5694), 279-283. doi:10.1126/science.1101400

Warren, L. A., Rossi, D. J., Schiebinger, G. R., Weissman, I. L., Kim, S. K., & Quake, S. R. (2007). Transcriptional instability is not a universal attribute of aging. *Aging Cell, 6*(6), 775-782. doi:10.1111/j.1474-9726.2007.00337.x

Weber, H., Kittel-Schneider, S., Gessner, A., Domschke, K., Neuner, M., Jacob, C. P., . . . Reif, A. (2011). Cross-disorder analysis of bipolar risk genes: further evidence of DGKH as a risk gene for bipolar disorder, but also unipolar depression and adult ADHD. *Neuropsychopharmacology, 36*(10), 2076-2085. doi:10.1038/npp.2011.98

Weiser, M. M. (1973). Intestinal epithelial cell surface membrane glycoprotein synthesis. I. An indicator of cellular differentiation. *J Biol Chem, 248*(7), 2536-2541.

Wittkopp, P. J., & Kalay, G. (2011). Cis-regulatory elements: molecular mechanisms and evolutionary processes underlying divergence. *Nat Rev Genet, 13*(1), 59-69. doi:10.1038/nrg3095

Wolf, E. J., Logue, M. W., Stoop, T. B., Schichman, S. A., Stone, A., Sadeh, N., . . . Miller, M. W. (2017). Accelerated DNA Methylation Age: Associations with PTSD and Mortality. *Psychosom Med*. doi:10.1097/psy.0000000000000506

Wong, C. C. Y., Meaburn, E. L., Ronald, A., Price, T. S., Jeffries, A. R., Schalkwyk, L. C., . . . Mill, J. (2014). Methylomic analysis of monozygotic twins discordant for autism spectrum disorder and related behavioural traits. *Mol Psychiatry, 19*(4), 495-503. doi:10.1038/mp.2013.41

Wood, J. G., & Helfand, S. L. (2013). Chromatin structure and transposable elements in organismal aging. *Front Genet, 4*, 274. doi:10.3389/fgene.2013.00274

Wood, S. H., van Dam, S., Craig, T., Tacutu, R., O'Toole, A., Merry, B. J., & de Magalhaes, J. P. (2015). Transcriptome analysis in calorie-restricted rats implicates epigenetic and post-translational mechanisms in neuroprotection and aging. *Genome Biol, 16*, 285. doi:10.1186/s13059-015-0847-2

Wozniak, R., Burke, B., & Doye, V. (2010). Nuclear transport and the mitotic apparatus: An evolving relationship (Vol. 67, pp. 2215-2230).

Xiang, J. F., Yin, Q. F., Chen, T., Zhang, Y., Zhang, X. O., Wu, Z., . . . Chen, L. L. (2014). Human colorectal cancer-specific CCAT1-L lncRNA regulates long-range chromatin interactions at the MYC locus. *Cell Res, 24*(5), 513-531. doi:10.1038/cr.2014.35

Xu, B., Roos, J. L., Levy, S., van Rensburg, E. J., Gogos, J. A., & Karayiorgou, M. (2008). Strong association of de novo copy number mutations with sporadic schizophrenia. *Nat Genet, 40*(7), 880-885. doi:10.1038/ng.162

Yang, J., Pawlyk, B., Wen, X. H., Adamian, M., Soloviev, M., Michaud, N., . . . Li, T. (2007). Mpp4 is required for proper localization of plasma membrane calcium ATPases and maintenance of calcium homeostasis at the rod photoreceptor synaptic terminals. *Hum Mol Genet, 16*(9), 1017-1029. doi:10.1093/hmg/ddm047

Yang, L., Wang, W. H., Qiu, W. L., Guo, Z., Bi, E., & Xu, C. R. (2017). A single-cell transcriptomic analysis reveals precise pathways and regulatory mechanisms underlying hepatoblast differentiation. *Hepatology, 66*(5), 1387-1401. doi:10.1002/hep.29353

Yang, X. J., & Seto, E. (2007). HATs and HDACs: from structure, function and regulation to novel strategies for therapy and prevention. *Oncogene, 26*(37), 5310-5318. doi:10.1038/sj.onc.1210599

Yeh, K. Y., Yeh, M., & Holt, P. R. (1991). Intestinal lactase expression and epithelial cell transit in hormone-treated suckling rats. *Am J Physiol, 260*(3 Pt 1), G379-384. doi:10.1152/ajpgi.1991.260.3.G379

Yin, H., Xue, W., Chen, S., Bogorad, R. L., Benedetti, E., Grompe, M., . . . Anderson, D. G. (2014). Genome editing with Cas9 in adult mice corrects a disease mutation and phenotype. *Nat Biotechnol, 32*(6), 551-553. doi:10.1038/nbt.2884

Yoshii, A., & Constantine-Paton, M. (2010). Postsynaptic BDNF-TrkB signaling in synapse maturation, plasticity, and disease. *Developmental Neurobiology, 70*(5), 304-322. doi:10.1002/dneu.20765

You, J. S., & Jones, P. A. (2012). Cancer genetics and epigenetics: two sides of the same coin? *Cancer Cell, 22*(1), 9-20. doi:10.1016/j.ccr.2012.06.008

Yu, D. H., Ware, C., Waterland, R. A., Zhang, J., Chen, M. H., Gadkari, M., . . . Shen, L. (2013). Developmentally programmed 3' CpG island methylation confers tissue- and cell-type-specific transcriptional activation. *Mol Cell Biol, 33*(9), 1845-1858. doi:10.1128/MCB.01124-12

Zai, C. C., Gonçalves, V. F., Tiwari, A. K., Gagliano, S. A., Hosang, G., Luca, V. D., . . . Kennedy, J. L. (2015). A genome-wide association study of suicide severity scores in bipolar disorder. *J Psychiatr Res., 65*, 23-29. doi:10.1016/j.jpsychires.2014.11.002

Zannas, A. S., Arloth, J., Carrillo-Roa, T., Iurato, S., Roh, S., Ressler, K. J., . . . Mehta, D. (2015). Lifetime stress accelerates epigenetic aging in an urban, African American cohort: relevance of glucocorticoid signaling. *Genome Biol, 16*, 266. doi:10.1186/s13059-015-0828-5

Zhang, C., Richon, V., Ni, X., Talpur, R., & Duvic, M. (2005). Selective induction of apoptosis by histone deacetylase inhibitor SAHA in cutaneous T-cell lymphoma cells: relevance to mechanism of therapeutic action. *J Invest Dermatol, 125*(5), 1045-1052. doi:10.1111/j.0022-202X.2005.23925.x

Zhang, W., Li, J., Suzuki, K., Qu, J., Wang, P., Zhou, J., . . . Belmonte, J. C. (2015). Aging stem cells. A Werner syndrome stem cell model unveils heterochromatin alterations as a driver of human aging. *Science, 348*(6239), 1160-1163. doi:10.1126/science.aaa1356

Zhou, B. O., & Zhou, J. Q. (2011). Recent Transcription-induced Histone H3 Lysine 4 (H3K4) Methylation Inhibits Gene Reactivation. *J Biol Chem, 286*(40), 34770-34776. doi:10.1074/jbc.M111.273128

Zhu, S., Goldschmidt-Clermont, P. J., & Dong, C. (2005). Inactivation of monocarboxylate transporter MCT3 by DNA methylation in atherosclerosis. *Circulation, 112*(9), 1353-1361. doi:10.1161/circulationaha.104.519025

Zwijnenburg, P. J., Meijers-Heijboer, H., & Boomsma, D. I. (2010). Identical but not the same: the value of discordant monozygotic twins in genetic research. *Am J Med Genet B Neuropsychiatr Genet, 153b*(6), 1134-1149. doi:10.1002/ajmg.b.31091

# Appendices

**Appendix A - 1. Regions exhibiting DNA modification levels that are significantly associated with LCT mRNA levels in development and adulthood**

***Chromosome-wide investigation***

| **Locus** | **Region*** | **Species** | **–log10(*q*)†** | **Association** |
|---|---|---|---|---|
| *LCT* exon–intron 1 | chr2: 136,594,004–136,594,148 | Human | 1.78 | ↑5-mC = ↓*LCT* mRNA |
| *Lct exon 10* | chr1: 128,295,240–128,295,278 | Mouse | 2.14 | ↑5-mC with ↑age |
| *Lct* intron 8–exon 9 | chr1: 128,298,461–128,298,810 | Mouse | 3.05 | ↑5-mC with ↑age |
| *Lct* intron 2–exon 3 | chr1: 128,312,912–128,313,561 | Mouse | 3.20 | ↑5-mC with ↑age |

***Fine-mapping investigation: Human***

| **Locus** | **Region*** | **–log10(*p*)‡** | **Association** |
|---|---|---|---|
| *LCT* intron 5 (LncRNA promoter) | chr2: 136,577,792–136,578,051 | 10.10 | Cell-type specific; ↑5-mC = ↑*LCT* mRNA |
| *LCT* intron 3 | chr2: 136,583,097–136,583,652 | 5.33 | Cell-type specific; ↑5-mC = ↓*LCT* mRNA |
| *LCT* intron 2 | chr2: 136,589,129–136,589,405 | 11.48 | Cell-type specific; ↑5-mC = ↓*LCT* mRNA |
| *LCT* exon 1 | chr2: 136,594,396–136,594,421 | 10.56 | Cell-type specific ↑5-mC = ↓*LCT* mRNA |
| *MCM6* exon 17 (3'-UTR) | chr2: 136,597,453–136,597,614 | 4.07 | Cell-type specific ↑5-mC = ↓*LCT* mRNA |
| *MCM6* exon 16 | chr2: 136,602,097–136,602,294 | 4.27<br>14.51 | Inter-individual variation & Cell-type specific; ↑5-mC = ↓*LCT* mRNA |
| *MCM6* intron 13–exon 13 | chr2: 136,608,430–136,609,013 | 10.96<br>2.42 | Inter-individual variation & Cell-type specific; ↑5-mC = ↓*LCT* mRNA |

***Fine-mapping investigation: Mouse***

| Locus | Region* | –log10(*p*)§ | Association |
|---|---|---|---|
| *Lct intron 8* | chr1: 128,298,854 | 3.51 | ↑5-mC with ↑age |
| *Lct intron 2* | chr1: 128,313,221 | 3.81 | ↑5-mC with ↑age |

*Human regions are GRCh37/hg19, mouse regions are GRCm38/mm10; †, ‡, §Significance values correspond to the most significant site in each region; †FDR q < 0.05 after Benjamini-Hochberg multiple testing correction; ‡p < 0.01 after Bonferroni correction for multiple testing; §FDR q < 0.05.

**Appendix A - 2. Roadmap Epigenomics Project and ENCODE data sets used for chromatin accessibility and histone modification analysis**

| Sample code | Sources | Tissue | Age | Sex | Data sets |
|---|---|---|---|---|---|
| H-23769 | Roadmap Epigenomics Project | Fetal small intestine, human | 108 days | Male | DNase I HS |
| H-24595 | Roadmap Epigenomics Project | Fetal small intestine, human | 108 days | Male | H3K4me1<br>H3K4me3<br>H3K27ac |
| ENCBS157ENC | ENCODE | Small intestine, mouse<br>Strain: C57BL/6NCrl | 8 weeks (60 days) | Male | POLR2A<br>H3K4me1<br>H3K4me3 |

**Appendix A - 3. Genome editing guides and all primer sequences**
A summary of all sequences can be found here:
https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4899171/bin/NIHMS778525-supplement-supp_table3.xlsx

**Appendix A - 4. Padlock probes and PCR primers**
A summary of all sequences can be found here:
https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4899171/bin/NIHMS778525-supplement-supp_table4.xlsx

**Appendix A - 5. Additional padlock probe sequences**
A summary of all sequences can be found here:
https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5282553/bin/srep41843-s2.xls

**Appendix A - 6. CTCF ChIP-qPCR primer locations**

| Primer | Locus | Sequence | Location |
|---|---|---|---|
| x1-FWD | *Lct* exon 1 | TCTCAACAAGGTCCCCAAAG | chr1:128327796-128327815 |
| x1-REV | *Lct* exon 1 | TGCAGTGCTACCGACAACTC | chr1:128327950-128327969 |
| i2-FWD | *Lct* intron 2 | TCCTGCACAGAAGTGAGCTG | chr1:128313174-128313193 |
| i2-REV | *Lct* intron 2 | CCATAGCCAGCATACCCATC | chr1:128313309-128313328 |
| i8-FWD | *Lct* intron 8 | ATCCTTTCCGTGTCATCCAG | chr1:128299877-128299896 |
| i8-REV | *Lct* intron 8 | AGCCTTGATCTCCATGATGC | chr1:128300007-128300026 |
| *Mcm6*-FWD | *Mcm6* intron/exon 13 | ACTCATGCATTCCTGTTCCC | chr1:128338074-128338093 |
| *Mcm6*-REV | *Mcm6* intron/exon 13 | AGTACAAACGTCTCCGCCAG | chr1:128338224-128338243 |
| *Mcm6*-X7-FWD | *Mcm6* exon 7 | GAGGCTCGTGCAGAGATTGT | chr1:128348385-128348404 |
| *Mcm6*-X7-REV | *Mcm6* exon 7 | CCACTTCAGTTTGGAGGGAA | chr1:128348509-128348528 |
| Negative control A: noCTCF-up-F | | GTCATAGTTCCCTGGGCAAA | chr1:128490456-128490475 |
| Negative control A: noCTCF-up-R | | CAACAGGATCCAGGGTCAGT | chr1:128490568-128490587 |
| Negative control B: noCTCF-down-F | | CCCGAAGGAAAACAGGTGTA | chr1:127891693-127891712 |
| Negative control B: noCTCF-down-R | | GTTCAGAGCAAGCCCAAGAC | chr1:127891856-127891875 |