Distribution Agreement

In presenting this thesis or dissertation as a partial fulfillment of the requirements for an advanced degree from Emory University, I hereby grant to Emory University and its agents the non-exclusive license to archive, make accessible, and display my thesis or dissertation in whole or in part in all forms of media, now or hereafter known, including display on the world wide web. I understand that I may select some access restrictions as part of the online submission of this thesis or dissertation. I retain all ownership rights to the copyright of the thesis or dissertation. I also retain the right to use in future works (such as articles or books) all or part of this thesis or dissertation.

Signature: Date:

Roles of DNA methylation and hydroxymethylation in aging, immunosenescence, and gene regulation

By
Nicholas D. Johnson
Doctor of Philosophy

Graduate Division of Biological and biomedical Science
Population Biology, Ecology, and Evolution

_____
Karen N. Conneely, Ph.D
Advisor

_____
David Cutler, Ph.D
Committee Member

_____
Hao Wu, Ph.D
Committee Member

_____
Michael Epstein, Ph.D
Committee Member

_____
Peng Jin, Ph.D
Committee Member

Accepted:

_____
Lisa A. Tedesco
Dean of the James T. Laney School of Graduate Studies

_____
Date

Roles of DNA methylation and hydroxymethylation in aging, immunosenescence, and gene regulation

By

Nicholas D. Johnson
MPH, Georgia State University, 2016

Advisor: Karen N. Conneely, Ph.D.

An abstract of
A dissertation submitted to the Faculty of the
James T. Laney School of Graduate Studies of Emory University
in partial fulfillment of the requirements for the degree of

Doctor of Philosophy

Graduate Division of Biological and Biomedical Sciences
Population Biology, Ecology, and Evolution

2021

Abstract

Understanding types of DNA methylation and their effects on gene regulation
by Nicholas D. Johnson

DNA methylation (DNAm) refers to the binding of a methyl group to DNA, which typically occurs at a cytosine nucleotide that is directly followed by a guanine nucleotide from 5' to 3'. DNAm can hold genes in a stably repressed state. It is associated with age, many chronic diseases and biological processes. DNA hydroxymethylation (DNAhm) is a more recent discovery and less well-characterized than DNAm. Unlike DNAm, DNAhm does not bind to known gene repressors (MBD1, MBD2, MBD3) and is associated with cis-gene expression. I first review the role of DNAm and DNAhm in one such biological process known as immunosenescence. Immunosenescence is the general deterioration of the immune system with age, and it is characterized by functional changes in hematopoietic stem cells (HSCs) and specific blood cell types as well as changes in levels of numerous factors, particularly those involved in inflammation. DNAm and DNAhm is not only associated with many immunosenescence-related processes but both of these epigenetic modifications play a causal role in some of them. In addition to reviewing the role of DNAhm and DNAm in immunosenescence, this dissertation specifically investigates age-related DNAhm and its relationship to age-related gene expression. I observed an overrepresentation of directional consistency between age-related DNAhm and age-related gene expression, which contrasts the repressive gene regulatory role of DNAm. This section is followed by an investigation of the capacity of single-cell methylation sequencing to detect novel imprints. Allele-specific methylation (ASM) is the mechanism underlying the establishment of imprints, which regulate gene expression by holding one allele of a gene in a stably repressed state. In this section, I developed a criteria to detect ASM at the sample-level. I then compared ASM detected at the sample-level to the methylation state in individual cells at CpGs with heterozygous sites sufficiently close to the CpGs to distinguish the methylation state of each allele, which constituted a cellular-level gold-standard. In this comparison, ASM detected in ≥3 samples was also ASM in the cellular-level gold standard 95% of the time. These results demonstrate that single cell methylation sequencing is a powerful tool to detect novel imprints.

Understanding types of DNA methylation and their effects on gene regulation

By

Nicholas D. Johnson
MPH, Georgia State University, 2016

Advisor: Karen N. Conneely, Ph.D.

A dissertation submitted to the Faculty of the
James T. Laney School of Graduate Studies of Emory University
in partial fulfillment of the requirements for the degree of

Doctor of Philosophy

Graduate Division of Biological and Biomedical Sciences
Population Biology, Ecology, and Evolution

2021

ACKNOWLEDGEMENTS

I would like to acknowledge all of the support I have received throughout my doctoral work. I would first and foremost like to thank my advisor, Karen Conneely, for the invaluable and persistent support and counsel. Thank you to the members of my committee, David Cutler, Hao Wu, Michael Epstein, and Peng Jin for their encouragement and willingness to meet to clarify difficult concepts. I would also like to thank my former lab mates, Chloe Robins, Liz Kennedy, and Crystal Grant for the guidance, especially through some of the more stressful times of my doctoral work. Thank you to my PBEE cohort, Amanda Vicente, Brent Allman, Connor Morozumi, Loy Xingwen, Daniel Desautels, Donna McDermott for all their support and comradery. A special thanks to my brother Zack Johnson for his advice on single cell sequencing and guidance throughout my doctoral work.

TABLE OF CONTENTS

Chapter 1: Introduction

DNA methylation (DNAm) is the biological phenomenon whereby a DNA methyltransferase catalyzes the chemical reaction that binds a methyl group to a cytosine nucleotide (Okano et al. 1999). In mammals, these chemical reactions occur most commonly at CpG sites, a cytosine that is directly followed by a guanine nucleotide from 5' to 3' (Mendizabal et al. 2014).

The methylation of gene promoters is capable of holding genes in a stably repressed state (Illingworth and Bird 2009; P. A. Jones 2012). Proteins known as MBD1, MBD2, MBD4, and MecP2 (Boyes and Bird 1991; Nan et al. 1998; Wade 2001) are all capable of binding to methylated DNA, which results in transcriptional repression.

DNAm is robustly associated with human age. Many epigenome-wide association studies have observed this association in human whole blood (Bell et al. 2012; Bjornsson et al. 2008; Christensen et al. 2009; Teschendorff et al. 2010; Alisch et al. 2012) and other tissues (Teschendorff et al. 2010; Christensen et al. 2009). Thousands of CpG sites undergo rapid changes in DNAm with age during childhood, which become more gradual in adulthood (Alisch et al. 2012). The robust association between DNAm and age has allowed for the development of DNAm-based biomarkers of aging capable of accurately predicting chronological age across many human tissues (Horvath 2013). Further, some of these DNAm-based biomarkers can predict mortality better than chronological age alone, which suggests DNAm does not only predict age, but serves as a marker of senescence as well (Hannum et al. 2013; Horvath 2013; Levine et al. 2018; Marioni et al. 2015).

DNAm is associated with numerous biological phenomena, one of which is immunosenescence, the decline of the immune system with age. Immunosenescence is a complex biological process entangled in other processes, such as aging, shifts in cell composition, and numerous chronic diseases. It marks cell lineage skewing (H. Li et al. 2014) and signatures of DNAm across thousands of sites can be used to characterize cell type composition (Houseman et al. 2012; Reinius et al. 2012). Further, DNAm is associated with at least 30 cancerous (Weisenberger 2014) and 20 non-cancerous chronic diseases (De Jager et al. 2014; Chouliaras et al. 2012; Bernstein et al. 2016; Xu et al. 2018; Zaina et al. 2014; de Vega, Vernon, and McGowan 2014; Qiu et al. 2012; Sharma et al. 2014; Q. Zhao et al. 2013; Smolarek et al. 2010; Hessam et al. 2017; Ventham et al. 2016; Leoni et al. 2015; den Hollander and Meulenbelt 2015; Jowaed et al. 2010; Smith et al. 2011; de Andres et al. 2015; Wockner et al. 2014; Figueroa-Romero et al. 2012; Javierre et al. 2010; M. Zhao et al. 2018; Toperoff et al. 2012). While DNAm is associated with many immunosenescence-related processes, there is evidence that it also plays a causal role in at least some of these processes (Ko et al. 2010; Moran-Crusio et al. 2011; Challen et al. 2011; Pan et al. 2016; M. Zhao et al. 2018).

The extent to which DNA methylation serves as a mechanism regulating immunosenescence and other associated phenotypes vs. simply a biomarker of these processes is still an open question, which is further complicated by the fact that a closely related epigenetic modification to DNAm, known as DNA hydroxymethylation (DNAhm), is not as well studied due to its recent discovery. The importance of characterizing DNAhm apart from DNAm is that the two epigenetic modifications play distinct gene regulatory roles. DNAm is catalyzed to DNAhm via the ten-eleven translocation enzymes

(TET1, TET2, TET3) (Ito et al. 2010). Unlike DNAm, DNAhm is associated with cis-gene expression (Marco et al. 2016; J. Zhao et al. 2017; Colquitt et al. 2013) and is unable to bind to methyl-binding proteins (MBD1, MBD2, and MBD4), which are known to play a role in transcriptional repression (Jin, Kadam, and Pfeifer 2010; Boyes and Bird 1991; Nan et al. 1998; Wade 2001).

In chapter 2 of my dissertation, I review the literature to determine our current understanding of immunosenescence and the roles DNAm and DNAhm play in it. I first discuss our current understanding of shifts in cell composition that occur in both the adaptive and innate arms of the immune system. I subsequently overview the main changes observed during immunosenescence and whether these changes should be viewed as the immune system in decline or changes that are compensating for immune system decline. These sections are followed by sections presenting our current knowledge regarding the roles DNAm and DNAhm play in immunosenescence-related processes including aging, lineage skewing of immune cell types, inflammation, and chronic diseases. These sections also distinguish between evidence of associations of DNAm and DNAhm and evidence of causal roles for DNAm and DNAhm in these processes. To conclude chapter 2, I discuss future directions to better understand the relationship between these two epigenetic modifications and immunosenescence. Some of these future directions include, but are not limited to, a better characterization of DNAhm in immunosenescence-related processes considering its characterization lags behind that of DNA methylation, which constitutes a gap I attempt to close in part via the work presented in chapter 3.

Considering that aging is one of the biological processes related to immunosenescence, chapter 3 of my dissertation explores the relationship between age-related DNAhm and age-related gene expression in human immune cells, which complements previous work characterizing the relationship between age-related DNAm and age-related gene expression. I first use DNAhm-capture-sequencing data collected from 10 young (23-30 years) and 10 old (68-76 years) Caucasian females to test for age-differential DNAhm in 2000 bp across the genome. I subsequently investigate enrichment of age-related DNAhm for biological pathways with the hypothesis that it will be associated with immune-related processes. I further investigate DNAhm in distinct genomic features, including promoters, gene bodies, and enhancers to further elucidate its functional role. In addition to DNAhm, I use RNA-sequencing data collected from the same 20 individuals to test for age-differential gene expression transcriptome-wide. In regions where age-related DNAhm and age-related gene expression overlapped, I observe that most genes with age-related gene expression and most regions with age-related DNAhm decrease with age, and that age-related DNAhm and age-related gene expression show directional consistency in all but one region, which is consistent with previous work that found a positive correlation of DNAhm with cis-gene expression (Marco et al. 2016; J. Zhao et al. 2017; Colquitt et al. 2013). In addition to DNAhm, another biological process that plays a role in both gene regulation and immunosenescence is imprinting. Imprinting refers to the repression of a specific parental copy (maternal or paternal) of a gene or a gene cluster via allele-specific methylation (ASM). The consequence of imprinting-related ASM is allele-specific expression (Imbeault, Helleboid, and Trono 2017). Genomic imprints are established in the female and male

germlines and are transmitted and maintained into adult life (Edwards and Ferguson-Smith 2007; Reig and Concha 2012; Y. Li and Sasaki 2011). Thus, imprinted regions occur at the same genomic loci across healthy human individuals. Imprinting is involved in immunosenescence via *Dlk1-Gtl2*, a maternally imprinted gene that maintains the self-renewal capacity of hematopoietic stem cells (HSCs) (Serrano-Lopez and Cancelas 2016). A reduced capacity for self-renewal in the HSC cell lineage underlies skewing of the HSC lineage, which constitutes a hallmark of immunosenescence (Geiger, de Haan, and Carolina Florian 2013).

In chapter 4, I utilize single-cell methylation data collected from 610 B cells sampled from six healthy individuals to detect novel imprinted regions. I first quantify ASM at CpG sites within individual cells and used this information to call ASM at the sample-level. I hypothesize that there would be concordance of ASM across samples since the same regions are expected to be imprinted across human individuals. I then investigate whether there was an overrepresentation of ASM in previously reported imprinted genes and imprinting binding motifs. To further validate my results, I call heterozygous sites in each of the 6 healthy samples and differentiate reads at these sites by which allele they overlap, allowing me to directly observe cellular-level ASM, which I regard as the gold standard. I then compare the gold standard ASM to the detected sample-level ASM to assess the ability to detect ASM in the absence of allelic information provided by heterozygous sites.

I conclude my dissertation with chapter 5 where I summarize the findings of the foregoing chapters. In addition, I describe the broader implications these findings have on the field of epigenetics.

References

Alisch, Reid S., Benjamin G. Barwick, Pankaj Chopra, Leila K. Myrick, Glen A. Satten, Karen N. Conneely, and Stephen T. Warren. 2012. “Age-Associated DNA Methylation in Pediatric Populations.” *Genome Research* 22 (4): 623–32.

Andres, María C. de, Eva Perez-Pampin, Manuel Calaza, Francisco J. Santaclara, Ignacio Ortea, Juan J. Gomez-Reino, and Antonio Gonzalez. 2015. “Assessment of Global DNA Methylation in Peripheral Blood Cell Subpopulations of Early Rheumatoid Arthritis before and after Methotrexate.” *Arthritis Research & Therapy* 17 (August): 233.

Bell, Jordana T., Pei-Chien Tsai, Tsun-Po Yang, Ruth Pidsley, James Nisbet, Daniel Glass, Massimo Mangino, et al. 2012. “Epigenome-Wide Scans Identify Differentially Methylated Regions for Age and Age-Related Phenotypes in a Healthy Ageing Population.” *PLoS Genetics* 8 (4): e1002629.

Bernstein, Alison I., Yunting Lin, R. Craig Street, Li Lin, Qing Dai, Li Yu, Han Bao, et al. 2016. “5-Hydroxymethylation-Associated Epigenetic Modifiers of Alzheimer’s Disease Modulate Tau-Induced Neurotoxicity.” *Human Molecular Genetics* 25 (12): 2437–50.

Bjornsson, Hans T., Martin I. Sigurdsson, M. Daniele Fallin, Rafael A. Irizarry, Thor Aspelund, Hengmi Cui, Wenqiang Yu, et al. 2008. “Intra-Individual Change over Time in DNA Methylation with Familial Clustering.” *JAMA: The Journal of the American Medical Association* 299 (24): 2877–83.

Bollati, Valentina, Joel Schwartz, Robert Wright, Augusto Litonjua, Letizia Tarantini, Helen Suh, David Sparrow, Pantel Vokonas, and Andrea Baccarelli. 2009. “Decline in Genomic DNA Methylation through Aging in a Cohort of Elderly Subjects.” *Mechanisms of Ageing and Development* 130 (4): 234–39.

Boyes, Joan, and Adrian Bird. 1991. “DNA Methylation Inhibits Transcription Indirectly via a Methyl-CpG Binding Protein.” *Cell* 64 (6): 1123–34.

Challen, Grant A., Deqiang Sun, Mira Jeong, Min Luo, Jaroslav Jelinek, Jonathan S. Berg, Christoph Bock, et al. 2011. “Dnmt3a Is Essential for Hematopoietic Stem Cell Differentiation.” *Nature Genetics* 44 (1): 23–31.

Chouliaras, Leonidas, Daniel L. A. van den Hove, Gunter Kenis, Stella Keitel, Patrick R. Hof, Jim van Os, Harry W. M. Steinbusch, Christoph Schmitz, and Bart P. F. Rutten. 2012. “Age-Related Increase in Levels of 5-Hydroxymethylcytosine in Mouse Hippocampus Is Prevented by Caloric Restriction.” *Current Alzheimer Research* 9 (5): 536–44.

Christensen, Brock C., E. Andres Houseman, Carmen J. Marsit, Shichun Zheng, Margaret R. Wrensch, Joseph L. Wiemels, Heather H. Nelson, et al. 2009. “Aging and Environmental Exposures Alter Tissue-Specific DNA Methylation Dependent upon CpG Island Context.” *PLoS Genetics* 5 (8): e1000602.

Colquitt, Bradley M., William E. Allen, Gilad Barnea, and Stavros Lomvardas. 2013. “Alteration of Genic 5-Hydroxymethylcytosine Patterning in Olfactory Neurons Correlates with Changes in Gene Expression and Cell Identity.” *Proceedings of the National Academy of Sciences of the United States of America* 110 (36): 14682–87.

De Jager, Philip L., Gyan Srivastava, Katie Lunnon, Jeremy Burgess, Leonard C.

Schalkwyk, Lei Yu, Matthew L. Eaton, et al. 2014. “Alzheimer’s Disease: Early Alterations in Brain DNA Methylation at ANK1, BIN1, RHBDF2 and Other Loci.” *Nature Neuroscience* 17 (9): 1156–63.

Edwards, Carol A., and Anne C. Ferguson-Smith. 2007. “Mechanisms Regulating Imprinted Genes in Clusters.” *Current Opinion in Cell Biology* 19 (3): 281–89.

Figueroa-Romero, Claudia, Junguk Hur, Diane E. Bender, Colin E. Delaney, Michael D. Cataldo, Andrea L. Smith, Raymond Yung, Douglas M. Ruden, Brian C. Callaghan, and Eva L. Feldman. 2012. “Identification of Epigenetically Altered Genes in Sporadic Amyotrophic Lateral Sclerosis.” *PloS One* 7 (12): e52672.

Fraga, Mario F., Esteban Ballestar, Maria F. Paz, Santiago Ropero, Fernando Setien, Maria L. Ballestar, Damia Heine-Suñer, et al. 2005. “Epigenetic Differences Arise during the Lifetime of Monozygotic Twins.” *Proceedings of the National Academy of Sciences of the United States of America* 102 (30): 10604–9.

Fraga, Mario F., and Manel Esteller. 2007. “Epigenetics and Aging: The Targets and the Marks.” *Trends in Genetics: TIG* 23 (8): 413–18.

Geiger, Hartmut, Gerald de Haan, and M. Carolina Florian. 2013. “The Ageing Haematopoietic Stem Cell Compartment.” *Nature Reviews. Immunology* 13 (5): 376–89.

Hannum, Gregory, Justin Guinney, Ling Zhao, Li Zhang, Guy Hughes, Srinivas Sadda, Brandy Klotzle, et al. 2013. “Genome-Wide Methylation Profiles Reveal Quantitative Views of Human Aging Rates.” *Molecular Cell* 49 (2): 359–67.

Hessam, Schapoor, Michael Sand, Kerstin Lang, Heiko U. Käfferlein, Lisa Scholl, Thilo Gambichler, Marina Skrygan, Thomas Brüning, Eggert Stockfleth, and Falk G. Bechara. 2017. “Altered Global 5-Hydroxymethylation Status in Hidradenitis Suppurativa: Support for an Epigenetic Background.” *Dermatology* 233 (2-3): 129–35.

Hollander, Wouter den, and Ingrid Meulenbelt. 2015. “DNA Methylation in Osteoarthritis.” *Current Genomics* 16 (6): 419–26.

Horvath, Steve. 2013. “DNA Methylation Age of Human Tissues and Cell Types.” *Genome Biology* 14 (10): R115.

Houseman, Eugene Andres, William P. Accomando, Devin C. Koestler, Brock C. Christensen, Carmen J. Marsit, Heather H. Nelson, John K. Wiencke, and Karl T. Kelsey. 2012. “DNA Methylation Arrays as Surrogate Measures of Cell Mixture Distribution.” *BMC Bioinformatics* 13 (May): 86.

Illingworth, Robert S., and Adrian P. Bird. 2009. “CpG Islands - ‘A Rough Guide.’” *FEBS Letters* 583 (11): 1713–20.

Imbeault, Michaël, Pierre-Yves Helleboid, and Didier Trono. 2017. “KRAB Zinc-Finger Proteins Contribute to the Evolution of Gene Regulatory Networks.” *Nature* 543 (7646): 550–54.

Ito, Shinsuke, Ana C. D’Alessio, Olena V. Taranova, Kwonho Hong, Lawrence C. Sowers, and Yi Zhang. 2010. “Role of Tet Proteins in 5mC to 5hmC Conversion, ES-Cell Self-Renewal and Inner Cell Mass Specification.” *Nature* 466 (7310): 1129–33.

Javierre, Biola M., Agustin F. Fernandez, Julia Richter, Fatima Al-Shahrour, J. Ignacio Martin-Subero, Javier Rodriguez-Ubreva, Maria Berdasco, et al. 2010. “Changes in the Pattern of DNA Methylation Associate with Twin Discordance in Systemic

Lupus Erythematosus." *Genome Research* 20 (2): 170–79.
Jin, Seung-Gi, Swati Kadam, and Gerd P. Pfeifer. 2010. "Examination of the Specificity of DNA Methylation Profiling Techniques towards 5-Methylcytosine and 5-Hydroxymethylcytosine." *Nucleic Acids Research* 38 (11): e125.
Jones, Meaghan J., Sarah J. Goodman, and Michael S. Kobor. 2015. "DNA Methylation and Healthy Human Aging." *Aging Cell* 14 (6): 924–32.
Jones, Peter A. 2012. "Functions of DNA Methylation: Islands, Start Sites, Gene Bodies and beyond." *Nature Reviews. Genetics* 13 (7): 484–92.
Jowaed, Ahmad, Ina Schmitt, Oliver Kaut, and Ullrich Wüllner. 2010. "Methylation Regulates Alpha-Synuclein Expression and Is Decreased in Parkinson's Disease Patients' Brains." *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience* 30 (18): 6355–59.
Ko, Myunggon, Yun Huang, Anna M. Jankowska, Utz J. Pape, Mamta Tahiliani, Hozefa S. Bandukwala, Jungeun An, et al. 2010. "Impaired Hydroxylation of 5-Methylcytosine in Myeloid Cancers with Mutant TET2." *Nature* 468 (7325): 839–43.
Leoni, Cristina, Sara Montagner, Lorenzo Deho', Rocco D'Antuono, Giovanna De Matteis, Angelo V. Marzano, Serena Merante, Ester M. Orlandi, Roberta Zanotti, and Silvia Monticelli. 2015. "Reduced DNA Methylation and Hydroxymethylation in Patients with Systemic Mastocytosis." *European Journal of Haematology* 95 (6): 566–75.
Levine, Morgan E., Ake T. Lu, Austin Quach, Brian H. Chen, Themistocles L. Assimes, Stefania Bandinelli, Lifang Hou, et al. 2018. "An Epigenetic Biomarker of Aging for Lifespan and Healthspan." *Aging* 10 (4): 573–91.
Li, H., T. Zheng, B. Chen, G. Hong, W. Zhang, T. Shi, S. Li, L. Ao, C. Wang, and Z. Guo. 2014. "Similar Blood-Borne DNA Methylation Alterations in Cancer and Inflammatory Diseases Determined by Subpopulation Shifts in Peripheral Leukocytes." *British Journal of Cancer* 111 (3): 525–31.
Li, Yufeng, and Hiroyuki Sasaki. 2011. "Genomic Imprinting in Mammals: Its Life Cycle, Molecular Mechanisms and Reprogramming." *Cell Research* 21 (3): 466–73.
Marco, Asaf, Tatiana Kisliouk, Tzlil Tabachnik, Aron Weller, and Noam Meiri. 2016. "DNA CpG Methylation (5-Methylcytosine) and Its Derivative (5-Hydroxymethylcytosine) Alter Histone Posttranslational Modifications at the Pomc Promoter, Affecting the Impact of Perinatal Diet on Leanness and Obesity of the Offspring." *Diabetes* 65 (8): 2258–67.
Marioni, Riccardo E., Sonia Shah, Allan F. McRae, Stuart J. Ritchie, Graciela Muniz-Terrera, Sarah E. Harris, Jude Gibson, et al. 2015. "The Epigenetic Clock Is Correlated with Physical and Cognitive Fitness in the Lothian Birth Cohort 1936." *International Journal of Epidemiology* 44 (4): 1388–96.
Mendizabal, I., T. E. Keller, J. Zeng, and Soojin V. Yi. 2014. "Epigenetics and Evolution." *Integrative and Comparative Biology* 54 (1): 31–42.
Moran-Crusio, Kelly, Linsey Reavie, Alan Shih, Omar Abdel-Wahab, Delphine Ndiaye-Lobry, Camille Lobry, Maria E. Figueroa, et al. 2011. "Tet2 Loss Leads to Increased Hematopoietic Stem Cell Self-Renewal and Myeloid Transformation." *Cancer Cell* 20 (1): 11–24.
Nan, X., H. H. Ng, C. A. Johnson, C. D. Laherty, B. M. Turner, R. N. Eisenman, and A.

Bird. 1998. “Transcriptional Repression by the Methyl-CpG-Binding Protein MeCP2 Involves a Histone Deacetylase Complex.” *Nature* 393 (6683): 386–89.

Okano, M., D. W. Bell, D. A. Haber, and E. Li. 1999. “DNA Methyltransferases Dnmt3a and Dnmt3b Are Essential for de Novo Methylation and Mammalian Development.” *Cell* 99 (3): 247–57.

Pan, Zhiqiang, Ming Zhang, Tao Ma, Zhou-Ya Xue, Guo-Fang Li, Ling-Yun Hao, Li-Jiao Zhu, Yan-Qiang Li, Hai-Lei Ding, and Jun-Li Cao. 2016. “Hydroxymethylation of microRNA-365-3p Regulates Nociceptive Behaviors via Kcnh2.” *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience* 36 (9): 2769–81.

Qiu, Weiliang, Andrea Baccarelli, Vincent J. Carey, Nadia Boutaoui, Helene Bacherman, Barbara Klanderman, Stephen Rennard, et al. 2012. “Variable DNA Methylation Is Associated with Chronic Obstructive Pulmonary Disease and Lung Function.” *American Journal of Respiratory and Critical Care Medicine* 185 (4): 373–81.

Rakyan, Vardhman K., Thomas A. Down, Siarhei Maslau, Toby Andrew, Tsun-Po Yang, Huriya Beyan, Pamela Whittaker, et al. 2010. “Human Aging-Associated DNA Hypermethylation Occurs Preferentially at Bivalent Chromatin Domains.” *Genome Research* 20 (4): 434–39.

Reig, German, and Miguel L. Concha. 2012. “Genomic Imprinting and Embryonic Development.” *International Journal of Morphology = Revista Internacional de Morfologia* 30 (4): 1453–57.

Reinius, Lovisa E., Nathalie Acevedo, Maaike Joerink, Göran Pershagen, Sven-Erik Dahlén, Dario Greco, Cilla Söderhäll, Annika Scheynius, and Juha Kere. 2012. “Differential DNA Methylation in Purified Human Blood Cells: Implications for Cell Lineage and Studies on Disease Susceptibility.” *PloS One* 7 (7): e41361.

Serrano-Lopez, Juana, and Jose A. Cancelas. 2016. “Mom Knows Best: Imprinted Control of Hematopoietic Stem Cell Quiescence.” *Cell Stem Cell*.

Sharma, Priyanka, Gaurav Garg, Arun Kumar, Farhan Mohammad, Sudha Ramesh Kumar, Vinay Singh Tanwar, Satish Sati, et al. 2014. “Genome Wide DNA Methylation Profiling for Epigenetic Alteration in Coronary Artery Disease Patients.” *Gene* 541 (1): 31–40.

Smith, Alicia K., Karen N. Conneely, Varun Kilaru, Kristina B. Mercer, Tamara E. Weiss, Bekh Bradley, Yilang Tang, Charles F. Gillespie, Joseph F. Cubells, and Kerry J. Ressler. 2011. “Differential Immune System DNA Methylation and Cytokine Regulation in Post-Traumatic Stress Disorder.” *American Journal of Medical Genetics. Part B, Neuropsychiatric Genetics: The Official Publication of the International Society of Psychiatric Genetics* 156B (6): 700–708.

Smolarek, Iwona, Eliza Wyszko, Anna M. Barciszewska, Stanislaw Nowak, Iwona Gawronska, Anna Jablecka, and Miroslawa Z. Barciszewska. 2010. “Global DNA Methylation Changes in Blood of Patients with Essential Hypertension.” *Medical Science Monitor: International Medical Journal of Experimental and Clinical Research* 16 (3): CR149–55.

Teschendorff, Andrew E., Usha Menon, Aleksandra Gentry-Maharaj, Susan J. Ramus, Daniel J. Weisenberger, Hui Shen, Mihaela Campan, et al. 2010. “Age-Dependent DNA Methylation of Genes That Are Suppressed in Stem Cells Is a

Hallmark of Cancer." *Genome Research* 20 (4): 440–46.
Toperoff, Gidon, Dvir Aran, Jeremy D. Kark, Michael Rosenberg, Tatyana Dubnikov, Batel Nissan, Julio Wainstein, et al. 2012. "Genome-Wide Survey Reveals Predisposing Diabetes Type 2-Related DNA Methylation Variations in Human Peripheral Blood." *Human Molecular Genetics* 21 (2): 371–83.
Vega, Wilfred C. de, Suzanne D. Vernon, and Patrick O. McGowan. 2014. "DNA Methylation Modifications Associated with Chronic Fatigue Syndrome." *PloS One* 9 (8): e104757.
Ventham, N. T., N. A. Kennedy, A. T. Adams, R. Kalla, S. Heath, K. R. O'Leary, H. Drummond, et al. 2016. "Integrative Epigenome-Wide Analysis Demonstrates That DNA Methylation May Mediate Genetic Risk in Inflammatory Bowel Disease." *Nature Communications* 7 (November): 13507.
Wade, Paul A. 2001. "Methyl CpG-Binding Proteins and Transcriptional Repression." *BioEssays: News and Reviews in Molecular, Cellular and Developmental Biology* 23 (12): 1131–37.
Weisenberger, Daniel J. 2014. "Characterizing DNA Methylation Alterations from The Cancer Genome Atlas." *The Journal of Clinical Investigation* 124 (1): 17–23.
Wockner, L. F., E. P. Noble, B. R. Lawford, R. Mcd Young, C. P. Morris, V. L. J. Whitehall, and J. Voisey. 2014. "Genome-Wide DNA Methylation Analysis of Human Brain Tissue from Schizophrenia Patients." *Translational Psychiatry* 4 (January): e339.
Xu, Cheng-Jian, Cilla Söderhäll, Mariona Bustamante, Nour Baïz, Olena Gruzieva, Ulrike Gehring, Dan Mason, et al. 2018. "DNA Methylation in Childhood Asthma: An Epigenome-Wide Meta-Analysis." *The Lancet Respiratory Medicine* 6 (5): 379–88.
Zaina, Silvio, Holger Heyn, F. Javier Carmona, Nuray Varol, Sergi Sayols, Enric Condom, José Ramírez-Ruz, et al. 2014. "DNA Methylation Map of Human Atherosclerosis." *Circulation. Cardiovascular Genetics* 7 (5): 692–700.
Zhao, Jinying, Yun Zhu, Jingyun Yang, Lin Li, Hao Wu, Philip L. De Jager, Peng Jin, and David A. Bennett. 2017. "A Genome-Wide Profiling of Brain DNA Hydroxymethylation in Alzheimer's Disease." *Alzheimer's & Dementia: The Journal of the Alzheimer's Association* 13 (6): 674–88.
Zhao, Ming, Meng-Ying Li, Xiao-Fei Gao, Su-Jie Jia, Ke-Qin Gao, Yin Zhou, Hui-Hui Zhang, et al. 2018. "Downregulation of BDH2 Modulates Iron Homeostasis and Promotes DNA Demethylation in CD4 T Cells of Systemic Lupus Erythematosus." *Clinical Immunology* 187 (February): 113–21.
Zhao, Q., Y-C Fan, J. Zhao, S. Gao, Z-H Zhao, and K. Wang. 2013. "DNA Methylation Patterns of Peroxisome Proliferator-Activated Receptor Gamma Gene Associated with Liver Fibrosis and Inflammation in Chronic Hepatitis B." *Journal of Viral Hepatitis* 20 (6): 430–37.

Chapter 2: The role of DNA methylation and hydroxymethylation in immunosenescence.

This chapter was published as cited below:

Nicholas D. Johnson, and Karen N. Conneely. 2019. “The Role of DNA Methylation and Hydroxymethylation in Immunosenescence.” Ageing Research Reviews 51 (February): 11–23

1. Introduction

Age-related DNAm and DNAhm may contribute to immunosenescence by regulating or mediating the regulation of levels of immune-related factors and proportions of immune cell types throughout life. Investigating the role of these epigenetic modifications in immunosenescence may help answer fundamental questions about aging while simultaneously providing valuable information to the field of medicine. The well-established associations of DNAm with age and immune cell types in whole blood support a possible role of DNAm in immunosenescence. DNAhm has not been as well studied in this context, so its potential involvement in aging and immunosenescence represents an open area where a gap in knowledge can be closed. The goal of this review is to discuss current evidence on the possible roles of DNAm and DNAhm in immunosenescence and highlight gaps in our understanding.

In mammalian DNA, methylation is the process whereby an enzyme known as DNA methyltransferase binds a methyl group to a cytosine nucleotide at a CpG site (a cytosine that is directly followed by a guanine from 5’ to 3’) forming 5-methylcytosine (5mC) (Okano et al., 1999). Common DNA methyltransferases are DNMT3A and

DNMT3B, responsible for *de novo* methylation, and DNMT1, a maintenance methyltransferase that preserves the methylation state across mitotic divisions. In addition to methyl groups, hydroxymethyl groups have also been observed to be bound to cytosine nucleotides, forming 5-hydroxymethylcytosine (5hmC) (Fig. 1). Ten-eleven translocation (TET) enzymes are a group of three proteins (TET1, TET2, TET3), each of which is capable of catalyzing 5mC to 5hmC (Ito et al., 2010).

DNAm robustly associates with age, numerous chronic diseases, and has a well-studied role in gene regulation. DNAm signatures across thousands of sites can be used to characterize cell type composition (Houseman et al., 2012; Reinius et al., 2012), and have been shown to mark cell lineage skewing (Li et al., 2014), which is a well-known feature of immunosenescence. Shifts in lineage commitment within hematopoietic stem cells (HSCs), the precursor lineage to major immune cell types (Fig. 2), are a well-documented change that occurs with age (Geiger et al., 2013). Within HSCs, differential DNAm with age has been observed in genes expressed in cell lineages downstream of HSCs (Beerman et al., 2013). T cells, a lineage downstream to HSCs, also undergo age-related shifts in subpopulations (Tu and Rao, 2016). In T cells as well as peripheral leukocytes, association between age-related DNAm and gene expression of genes regulating T cell lineage has also been observed (Tserel et al., 2015). These studies, among others we will describe in this review, support the interpretation that 5mC is involved in immunosenescence.

An appreciation of the possible involvement of 5hmC in immunosenescence has lagged behind that of 5mC. One reason for this is that blood cells carry out numerous functions of the immune system, but 5hmC content is estimated to be 0.027% in human

whole blood, which is relatively low (Godderis et al., 2015). 5hmC content in the brain, for example, has been estimated at 13% (Wen et al., 2014). However, the low overall level of 5hmC in blood does not preclude it from playing a regulatory role at specific sites. A second reason is that the discovery of 5hmC in mammalian DNA is more recent and bisulfite-based methods used to detect DNAm are unable to distinguish between 5mC and 5hmC. More recently, however, methods have been developed that use chemical modification to differentiate 5mC from 5hmC prior to sequencing, allowing the specific detection and quantification of 5hmC (Booth et al., 2012; Höbartner, 2011; Song et al., 2012; Szwagierczak et al., 2010; Terragni et al., 2012).

Current evidence suggests that 5mC and 5hmC have distinct effects on gene expression. It is well understood that DNAm can hold a CpG island promoter in a stably repressed state (Illingworth and Bird, 2009; Jones, 2012). In contrast, recent work suggests that promoter DNAhm does not inhibit gene expression like DNAm and instead both promoter and gene body DNAhm associate with increased cis-gene expression (Colquitt et al., 2013; Marco et al., 2016; J. Zhao et al., 2017). This may be because proteins (MBD1, MBD2 and MBD4) known to bind to 5mC and contribute to transcriptional repression do not bind to 5hmC (Fig. 3) (Boyes and Bird, 1991; Jin et al., 2010; Nan et al., 1998; Wade, 2001). Instead, 5hmC has been suggested as an intermediate modification between DNAm and demethylation (He et al., 2011; Ito et al., 2011; Tahiliani et al., 2009; Wu and Zhang, 2010). Because they appear to have different regulatory functions, both 5mC and 5hmC warrant consideration when investigating gene regulatory mechanisms in immunosenescence and other biological processes.

The goal of this review is to explore a possible role for DNAm, DNAhm, and gene expression in immune system decline. We start by describing major features of immunosenescence, which is a difficult phenotype to characterize because the immune system is very complex, and the senescence of the immune system involves changes in the quantities of numerous factors carrying out multiple tasks within a variety of blood cell types and their precursor cell lineage, HSCs. We then describe the relationship between epigenetics and the immune system and discuss evidence suggesting that DNAm and DNAhm may play a role in immunosenescence. Next we discuss numerous studies reporting changes in DNAm and DNAhm with age and age-related chronic diseases, which, in many cases occur within immune-related genes or associate with their expression. In some cases, findings suggest some of these immune-related genes to play a role in the degeneration or deterioration characteristic of the disease being interrogated, but in most cases associations are observed without evidence of causality being established, leaving a gap in knowledge to be filled by future studies. We end by outlining potential future directions that could address this gap and advance the field.

2. Immunosenescence

Immunosenescence entails changes in both the innate and adaptive arms of the immune system with age, and is accompanied by immune lineage skewing and an increase in chronic, low-grade inflammation known as inflammaging (Agarwal and Busse, 2010; Franceschi et al., 2000). The cells of the innate and adaptive immune system differentiate from HSCs, which help replace roughly ten billion white blood cells every day of a person's life (Yoder, 2004). Leukocytes (white blood cells) in the human body

originate from the HSCs in the bone marrow that first differentiate into the myeloid and lymphoid lineages, thereafter differentiating into neutrophils, eosinophils, basophils, monocytes, and lymphocytes (Fig. 3). Monocytes have the potential to differentiate into macrophages or dendritic cells (DCs) and populate tissues (Geissmann et al., 2010). The lymphocytes of the lymphoid lineage then differentiate into natural killer (NK) cells, T cells, and B cells (Ye and Graf, 2007). Collectively, these cell types carry out various tasks of the immune system, such as phagocytosis, the recognition of “non-self” antigens, and antigen presentation (Janeway, 2001; Medzhitov and Janeway, 2002). In this section we first discuss major age-related changes to immune cell types followed by a discussion of the possible role these changes play to the health and aging of the organism.

2.1. Age-related changes in immune cell types

Innate immune system cells function as first responders to bacterial invaders or other microorganisms. These cells can phagocytose invaders and release cytokines and chemokines alerting other innate immune system cells of the invaders. The release of cytokines and chemokines, also known as the inflammatory response, alerts cells of the adaptive immune system, which are capable of “remembering” specific pathogens and targeted responses to these pathogens in the event of future invasions (Janeway, 2001). Cell lineages of both the innate and adaptive immune system are observed to undergo changes in immunosenescent individuals. HSCs, from which these cell lineages derive (Fig. 3), also undergo changes with age, including reduced function and skewing toward the myeloid lineage (Geiger et al., 2013).

Adaptive arm of the immune system:

T cells and B cells are major cell types of the adaptive immune system and work in a coordinated fashion. B cells are responsible for producing antibodies, which respond to specific antigens allowing the adaptive immune system to mount targeted responses to pathogens. T cells carry out many tasks of the adaptive immune system They provide immunity to intracellular and extracellular pathogens, mount attacks in response to infections, and retain “memory” of previous infections so that they can mount attacks in the event of future infections (Zhang and Bevan, 2011; Zhu and Paul, 2008).

One of the most prominent features of T cell aging relates to thymic involution (shrinkage of the thymus with age). Palmer (2013) reviews several studies indicating that thymic involution is responsible for a decreased output of naive T cells in animal models whereas in humans it remains a topic of debate. Regardless, naive T cell output has been observed to decrease with age in humans. In PBMCs of 39 human donors aged 6-90 years, T cell receptor diversity decreased with age, accompanied by a decrease in naive T cells (Britanova et al., 2014). Further, $CD8^+$ T cells are increasingly absent of CD28 with age (Weng et al., 2009).

Many age-associated changes in transcription are observed in T cells and their subsets (i.e. $CD4^+$, $CD8^+$, $CD8^+CD28^-$) (Chen et al., 2013). In $CD8^+$ T cells isolated from PBMCs of five young (23-27 years) and four old (65-80 years) individuals, Cao et al. (2010) observed 754 differentially expressed genes, 66% with decreasing expression and 34% with increasing expression, with overrepresentation of genes involved in immune response among genes with increasing expression. In $CD4^+$ T cells isolated from PBMCs of 423 participants in the Multi-Ethnic Study of Atherosclerosis (MESA) study, differential

expression with age was observed in 218 genes (Reynolds et al., 2015), with suggestive enrichment for immune response pathways among genes with increasing expression with age. In human PBMCs, T_h1 and T_h2 cell counts increase with age while the ratio of T_h2 to T_h1 cells decreases with age (Uciechowski et al., 2008).

B cells, which are responsible for producing antibodies, also undergo functional changes with age, which are reviewed by Cancro et al. (2009). B cells of aged individuals have reduced protein levels and expression of genes that contribute to developmental progression of B cells. Accompanying these changes, B cells have a diminished capacity to complete each stage of differentiation. Age-related alterations in the B1 and B2 cell subsets may reflect these changes: the B1 pathway predominates in prenatal and neonatal development whereas the B2 pathway predominates in young adult life, followed by a proportional increase in the B1 pathway in later life. In addition, B cells undergo a loss in receptor diversity with age (Cancro et al., 2009).

Innate arm of the immune system:

Cell lineages, including monocytes, macrophages, and dendritic cells (DCs) of the innate arm of the immune system undergo changes with age as well. Monocytes carry out various tasks in the immune system. They are the most numerous mononuclear phagocyte in the blood, and are capable of migrating from blood to tissues during inflammation, differentiating into macrophages, presenting antigens to T cells, and affecting T cell differentiation (Jakubzick et al., 2017). Despite their similarities, DCs and macrophages have distinct functions: while macrophages are primarily engaged in

maintenance of tissue immune integrity, such as bone homeostasis, DCs are efficient antigen presenters primarily involved in tissue immune response (Hashimoto et al., 2011).

In monocyte samples of 146 healthy adults (20-84 years), phagocytosis of monocytes was impaired with age and they exhibited altered expression of a number of CD molecules (Hearps et al., 2012). In 181 healthy adult subjects (18-88 years), the $CD14^{+}CD16^{+}$ subset of monocytes increased with age, accompanied by age-related changes in chemokine receptors and increases in serum monocyte chemoattractant protein-1, although total monocyte counts were comparable between young and old groups (Seidler et al., 2010). In $CD14^{+}$ monocytes isolated from PBMCs of 1,264 MESA participants (55–94 years), differential expression with age was observed in 2,704 genes, with nominally significant enrichment for immune response pathways (Reynolds et al., 2015).

While monocytes differentiate into macrophages, a recent review notes that the majority of macrophages derive from embryonic precursors and are self-maintained in tissues (Ginhoux and Jung, 2014). In a review of macrophage function and age-related functional decline, Linehan and Fitzgerald (2015) report a number of studies indicating that macrophages show age-related impairments in their ability to become activated in response to IFN-γ, to secrete cytokines in response to TLR stimulation, to present antigens, and to phagocytose and repair tissue. Agarwal and Busse (2010) review a number of murine and human studies suggesting mixed support for altered secretion of chemokines and other cytokines with age in macrophages and monocytes, including reduced levels of TNF-α, IL-1β, IL-2, IL-6, IL-12, CCL5, macrophage inflammatory protein 1α, and increased levels of IL-10. They note inconsistent findings that could be attributable to the fact that many of the murine studies were conducted using stimulated

spleen cell cultures of macrophages whereas human studies were carried out using monocytes isolated from peripheral blood (Agarwal and Busse, 2010).

With age DCs have impaired antigen processing and migration, but comparable cell count and phenotypes (Agrawal and Gupta, 2011). In addition, Wong and Goldstein (2013) review a number of studies that collectively suggests an age-related impaired ability of DCs to present antigens to T cells, although some conflicting reports exist. Gupta (2014) reviews mechanisms underlying age-related functional decline of dendritic cells (DCs), noting impaired phagocytosis and migration, and possible consequences of these age-related changes to other immune cell types.

Neutrophils and eosinophils, which are immune cells critical for combating microorganisms and parasites, respectively, also undergo age-related changes. Lord et al. (2001) describe a number of changes in neutrophils including a reduction in the number of microbes ingested per neutrophil, decreased Fc-mediated phagocytosis, and decreased expression of CD16 with age. They suggest low GM-CSF (due to the shift from T_h1 to T_h2 in T cells) and high TNF-α (due to increased secretion in monocytes) may be responsible for age-related reduction in neutrophil responsiveness and decreased survival at the site of infection (Lord et al., 2001). Eosinophils secrete granules containing antimicrobial substances (Faurschou and Borregaard, 2003), and Mathur et al. (2008) observed an age-related decrease in IL-5 stimulated degranulation (Mathur et al., 2008).

2.2. Immunosenescence: Dysfunction or Adaptation?

The specific age-related changes to cell types of the two arms of the immune system collectively engender immunosenescence and inflammaging. It is well established

that elevated inflammation is a risk factor for morbidity and mortality (Brüünsgaard and Pedersen, 2003). Nevertheless, a recent review by Fulop et al. (2017) argues that some immune system changes are beneficial and that the interplay between immunosenescence and inflammaging is more complex than conventionally described.

Fulop et al. (2017) propose that healthy aging entails an optimization of changes encompassing adaptive immune system remodeling, inflammaging, and anti-inflammaging. Changes to the adaptive immune system may serve to optimize resource allocation for changing needs of the organism. For example, thymic involution may decrease the TCR repertoire and the naive T cell compartment, but Fulop et al. (2017) note that it may benefit the organism by lowering energy consumption and that age-related increases of the memory T cell compartment may help the organism combat cognate pathogens and stave off infection. Furthermore, the increase in the Treg compartment may address a growing need for autoimmunity. Although inflammaging may contribute to disease, Fulop et al. (2017) suggest that when optimized, inflammaging may counterbalance the altered ability to fight off new infections brought about by adaptive remodeling, and that anti-inflammaging can serve to prevent inflammation from becoming excessive and destructive.

Other researchers have also argued for careful consideration of age-related changes to the immune system before they are deemed harmful, as many may be beneficial. For example, an increasing proportion of $CD8^+$ T cells are absent of CD28 with age, and Arosa (2002) notes that increased proportions of $CD8^+28^-$ T cells accompany more than 20 conditions and disorders. Although $CD8^+28^-$ T cells have been conventionally considered dysfunctional, terminally differentiated cells, a recent review

argues that these cells are involved in tissue repair and homeostasis (Arosa et al., 2016). Thus, the age-related increase in $CD8^{+}28^{-}$ T cells, rather than being harmful, may serve to address the increasing need for tissue maintenance of the senescing organism.

Lineage skewing of immune cell compartments is a major characteristic of immunosenescence, which helps give rise to the remodeling of the immune system described above. These changes are involved in complex outcomes including healthy vs. unhealthy aging as well as numerous disease states. These complex outcomes vary markedly between elderly individuals. DNAm and DNAhm can help build on our understanding of immunosenescence because numerous studies have observed associations of DNAm and DNAhm with lineage skewing, inflammation, age and many of the complex outcomes related to aging. We discuss these studies in the sections below and argue that future work on immunosenescence stands to benefit by investigating potential roles of DNAm and DNAhm.

3. Age-related DNA methylation and DNA hydroxymethylation

DNAm has been shown to robustly associate with human age, and this association has been observed across various tissues, with a large number of studies reporting epigenome-wide associations in human whole blood (Alisch et al., 2012; Bell et al., 2012; Bjornsson et al., 2008; Christensen et al., 2009; Teschendorff et al., 2010). During childhood, thousands of CpG sites undergo rapid DNAm changes with age, some of which become hypomethylated and others hypermethylated. The majority of these sites show significant, but less rapid DNAm changes in adulthood (Alisch et al., 2012). Gentilini et al. (2012) observed 13 hypermethylated and 15 hypomethylated sites in offspring of

centenarians compared to offspring of non-centenarians suggesting DNAm not only associates with age, but senescence. In addition to nuclear DNAm, mitochondrial DNAm also is associated with age and may be a biomarker of aging in both humans and murine models (D'Aquila et al., 2015; Iacobazzi et al., 2013).

To take advantage of the robust association with DNAm and age, several DNAm-based biomarkers of aging have been recently developed. Some of these epigenetic clocks can estimate age across numerous human tissues with high accuracy (Horvath, 2013) and predict outcomes such as mortality more accurately than chronological age alone (Hannum et al., 2013; Horvath, 2013; Levine et al., 2018; Marioni et al., 2015), suggesting it could detect healthy vs. unhealthy aging. Using Horvath's Clock, semi-supercentenarians (N=82) were observed to have an epigenetic age of 8.6 years younger than their chronological age, and offspring of semi-supercentenarians (N=63) to be 5.1 years younger than age-matched controls (N=47) (Horvath et al., 2015). This study demonstrates that DNAm can be used to study complex outcomes of immunosenescence, namely, healthy vs. unhealthy aging, and a recent review highlights the possibility of DNAm as a therapeutic target to increase longevity (Xiao et al., 2016). The accuracy of DNAm-based predictors as well as the stability of DNAm also allows for the estimation of age in anthropological and forensic contexts. Pedersen et al. (2014) were able to estimate age of a 4000-yr-old Paleo-Eskimo using DNAm from hair tissue, and Giulani et al. (2016) were able to predict age with high accuracy using DNAm levels extracted from the cementum and pulp of modern teeth based on CpG sites from the genes *ELOVL2, FHL2, and PENK*.

Many loci are consistently correlated with age across multiple studies. The best-known of these loci is DNAm in the CpG island of *ELOVL2*, which has been observed among the top significant hits from numerous epigenome-wide association studies of age (Florath et al., 2014; Gopalan et al., 2017; Karpf, 2012; Marttila et al., 2015; Reynolds et al., 2014; Rönn et al., 2015; Slieker et al., 2018; Steegenga et al., 2014). In a large cohort of 501 subjects aged 9 to 99 years, a Spearman's correlation of 0.92 between age and DNAm was observed in the CpG island of *ELOVL2* (Garagnani et al., 2012). In vitro cell replication experiments demonstrated that *ELOVL2* associated with cell replication rather than senescence and no association between longevity/mortality and whole blood DNAm was observed from participants of the Leiden Longitudinal Study (N=994; 89-104 years) (Bacalini et al., 2017). These observations suggest that DNAm of *ELOVL2* marks age, but not necessarily aging.

Among the epigenetic biomarkers of age discussed in this section, some are more sensitive to cell lineage skewing than others. Horvath's clock (Horvath, 2013) was designed to be tissue-independent and is thus largely unaffected by cell lineage skewing in blood samples and can predict DNAm across various cell types with similar accuracy. An earlier aging clock (Hannum et al., 2013) was constructed based on blood DNAm and may partially reflect cell lineage skewing, but its inclusion of extrinsic information beyond DNAm may result in more accurate predictions of mortality and health outcomes (Horvath and Raj, 2018). Based on this idea, a third biomarker, DNAm PhenoAge, was constructed to intentionally model extrinsic information such as cell count, inflammatory markers, and other clinical measures, that may inform health outcomes (Levine et al., 2018). The increased success of these biomarkers at predicting mortality is possibly indicative of the

importance to aging and mortality of immunosenescent-related processes, such as cell lineage skewing, as well as other extrinsic factors, such as disease state.

In contrast to DNAm, preliminary studies are just beginning to characterize associations between DNAhm and age. DNAhm has been observed to associate with age in the hippocampus and cerebellum of mice aged 7 days, 6 weeks, and 1 year (Szulwach et al., 2011). In human mesenchymal stem cells of 11 young (aged 2-29 years) and 6 old (63-89 years) donors, nominally significant hyper-DNAhm was reported at 785 CpG sites and hypo-DNAhm at 846 CpG sites (Toraño et al., 2016). However, this study did not observe genome-wide significant associations, so larger studies will be needed to assess the extent to which DNAhm associates with age in humans.

4. DNA methylation and hydroxymethylation in lineage skewing

Variation in levels of DNAm across cell types of whole blood at thousands of CpG sites embody cell type specific signatures sufficiently distinct that they are widely used to estimate proportions of blood cell types in human whole blood samples (Houseman et al., 2012; Reinius et al., 2012). A recent study of multiple cancers and inflammatory diseases used DNAm as a marker of lineage skewing, utilizing its ability to measure proportions of myeloid and lymphoid cells in human blood samples, and ultimately demonstrating a pattern of skewing toward the myeloid lineage in a number of diseases (Li et al., 2014). In an analysis of whole genome bisulfite sequencing from 112 samples from BLUEPRINT, DNAm was observed to distinguish between myeloid and lymphoid lineages and patterns of DNAm became more pronounced throughout B and T lymphocyte development

(Schuyler et al., 2016). DNAm is therefore useful to measure immunosenescence-related lineage skewing.

Since cell lineage skewing is a well known feature of immunosenescence, one could posit that the observation of age-related DNAm in whole blood is solely attributable to this skewing. This argument, however, would imply that associations between DNAm and age will not be observed within individual blood cell types, which is not the case. In the Multi-Ethnic Study of Atherosclerosis (MESA), age-associated DNAm was observed for 2,595 CpG sites in $CD4^+$ T cells (N=227) and 37,911 sites in monocytes (N=1,264), a subset of which associated with age-related expression of genes involved in antigen processing and presentation (Reynolds et al., 2014). A study in human $CD8^+$ T cells isolated from 50 young (22-34 years) and 50 old (73-84 years) individuals observed differential DNAm related to both age and skewing, further demonstrating that one does not confound the other (Tserel et al., 2015). Further, fewer age-differentially methylated CpG sites were observed in peripheral blood leukocytes (806 sites) compared to $CD4^+$ T cells (12,275 sites) and $CD8^+$ T cells (48,876 sites) from the same individuals.

Age-related DNAm changes also accompany shifts in T cell subsets, which are an important feature of immunosenescence. In Tserel et al. (2015), levels of DNAm at CpG sites within several genes involved in regulation of T cell lineage (*CD27, CD248, SATB1, TCF7, BCL11B, RUNX3*) inversely correlated with expression of these genes, and genes involved in immune response were observed to have decreased DNAm and increased gene expression with age. A CpG site in the promoter region of *IFN-γ* has been observed to become hypomethylated during T_h2 polarization in mice (Jones and Chen, 2006). Furthermore, *IL-4* undergoes demethylation within differentiating mouse T_h2 cells (Lee et

al., 2002). An age-related decrease in the ratio of T_h2 to T_h1 cells has been observed with age in humans (Uciechowski et al., 2008), for which the results of Lee et al. (2002) and Jones and Chen (2006) provide possible mechanisms involving DNAm.

DNAm changes are potentially involved in Treg dysfunction and may contribute to immunosenescence (Jasiulionis, 2018). Johnson et al. (2017) observed age-related DNAm with an accelerated rate of increase in DNAm in late life at 2 CpG sites in the *KLF14* promoter in human whole blood and replicated this finding in several other human tissues including monocytes and T cell subsets. *KLF14* has been suggested to affect proportions of naive and regulatory T cells via the regulation of *FOXP3* in mice (Sarmento et al., 2015). Rosenkranz et al. measured the frequency of regulatory T cells in young (23-40 years) and old (51–87 years) individuals and observed significantly higher frequency of regulatory T cells marked by FOXP3 in the old age group (2007). These findings support the interpretation that DNAm changes could influence relative proportions of Treg cell subsets in the elderly.

Although the number of HSCs undergoes a two- to ten-fold increase throughout life in both humans and mice, age-related functional defects in HSCs, including a reduction in self-renewal capacity and the skewing of lineage differentiation, are major factors in immunosenescence (Geiger et al., 2013). Several studies support the interpretation that DNAm and DNAhm are involved in cell lineage skewing of HSCs. In mice, Beerman et al. (2013) observed an age-related reduction in functional potential of transplanted HSCs, measured by the ability of the transplanted cells to reconstitute irradiated bone marrow, and transplants of aged HSCs exhibited skewing of reconstitution toward the myeloid lineage. In addition, age-related DNAm occurred in genes more highly

expressed or exclusively expressed in downstream cell lineages at a much higher frequency compared to genes more highly expressed or exclusively expressed in HSCs. This suggests regulatory consequences of DNAm changes in HSCs that do not occur until HSC differentiation. Comparing young vs. old mice, regions gaining DNAm with age were enriched for regions of open chromatin in lymphoid cells, while regions losing DNAm were enriched for regions of open chromatin in myeloid cells, which is consistent with a DNAm-driven skewing toward the myeloid lineage with age (Beerman et al., 2013). In transplantation experiments of DNMT3A-null and control HSCs in the bone marrow of mice, Challen et al. (2011) found that DNMT3A-null HSCs had higher self-proliferation capacity and lower differentiation capacity with a skewed contribution to peripheral blood of B cells compared to controls. Both hypo- and hyper-methylation were observed at sites in DNMT3A-null HSCs with hypermethylation accounting for 95% of the differential methylation occurring in CpG islands. Many genes involved in HSC differentiation were also downregulated in DNMT3a-null HSCs. Evidence also suggests a role for DNAhm in skewed differentiation of HSCs. In bone marrow genomic DNA of 88 patients with myeloid malignancies with *TET2* mutation and 17 healthy controls, 5hmC levels negatively associated with *TET2* mutation (Ko et al., 2010). This same study observed that in HSCs isolated from mice, downregulation of *TET2* via transduction with *TET2* short hairpin RNA skewed lineage differentiation toward the monocyte/macrophage lineage (Ko et al., 2010). Examination of a mouse model of *TET2* loss yielded similar results, showing that *TET2* loss-of-function mutation resulted in decreased 5hmC content, increased HSC self-renewal capacity, and skewing toward the myeloid lineage (Moran-Crusio et al., 2011).

Overall, these findings suggest that DNAm and DNAhm could play a role in immunosenescence, specifically, HSC aging.

5. DNAm in inflammation

A meta-analysis of 8863 individuals of European ancestry found strong evidence that DNAm associates with levels of an inflammatory factor, observing that DNAm at 218 CpG sites associates with CRP levels and replicating 58 of these CpG sites in an African-American panel (n=4111) (Ligthart et al., 2016). The most significant CpG site (cg10636246) is within 1500 bp of a gene called Absent in melanoma 2 (*AIM2*), which is expressed in adult peripheral blood $CD27^+$ B cells at steady state (Svensson et al., 2017). In macrophages, AIM2 is involved in an inflammasome response. In particular, in response to human cytomegalovirus (CMV) infection, macrophages made to be deficient of *AIM2* by use of small interference RNA (siRNA) had an impaired ability to induce an inflammasome response (indicated by IL- 1β and IL- 18) initiate cell death (indicated by Lactate Dehydrogenase), and curb the CMV life cycle (Huang et al., 2017).

Associations between DNAm and inflammatory factors have also been observed in studies of traits related to inflammation such as obesity. Studies investigating obesity and related conditions have reported dysregulation of DNAm associated with inflammatory factors. Yusuf et al. measured changes in plasma concentrations of patients with dyslipidemia pre- vs. post-intervention with a PPAR-α inhibitor, and observed concentrations of CRP, IL-2, and IL-6 to associate with DNAm at several CpG sites in $CD4^+$ T cells (2017). In PBMC samples of 186 overweight/obese subjects, PM10 (particulate matter 10 micrometers and smaller) exposure negatively associated with

DNAm of *CD14* and *TLR4* (Cantone et al., 2017). In a similar study of 165 obese subjects, DNAm of *TNF-α* negatively associated with nutrient intake of cholesterol, folic acid, β-carotene, carotenoids, and retinol in whole blood (Bollati et al., 2014).

Jasiulionis (2018) review a number of studies suggesting that DNAm mediates the effect of many environmental and lifestyle factors on aging. Some of these studies indicate that an enhanced inflammatory response results from promoter demethylation of *IL-6* in response to deficiency of zinc (Wong et al., 2015). A review by Haase and Rink (2009) suggest that the age-related decrease of zinc may play a role in inflammaging (Haase and Rink, 2009). DNAm may also be involved in inflammation associated with sociocultural and psychological factors. It is well known that living in disadvantaged neighborhoods is associated with poor health outcomes (Diez Roux and Mair, 2010). In monocytes purified from blood from 1,226 MESA participants, socioeconomic disadvantage and neighborhood social environment both associated with DNAm near a number of genes coding for inflammatory factors and in some cases associated changes in gene expression of these factors were also observed (Smith et al., 2017). In the same cohort, both low adult socioeconomic status and low social mobility associated with DNAm near several inflammation-related genes, which in turn associated with gene expression (Needham et al., 2015). In blood samples from participants of the EPIC Italy prospective cohort study (n=857), socioeconomic status associated with differential DNAm of probe sites within the gene body, the 5' untranslated region, or the region from the transcription start site to 1500 bp upstream of the transcription start site of six inflammation-related genes (Stringhini et al., 2015). In addition to sociocultural factors, a study examining the involvement of epigenetics in psychological phenotypes found that

anxiety, depression, and hostility were observed to associate with human whole blood DNAm in the promoter of inflammation-related factors Intercellular Adhesion Molecule-1 (*ICAM-1*) and coagulation factor III (*F3*) (Kim et al., 2016). In whole blood samples from participants of the Grady Trauma Project, post-traumatic stress disorder (PTSD) associated with DNAm of CpG sites within genes involved in inflammation. In the same study, PTSD also associated with plasma concentrations of several factors involved in immune system regulation including several interleukins and TNF-α (Smith et al., 2011).

DNAm changes also accompany changes to inflammatory factors that potentially relate to cardiovascular health in the elderly. In whole blood samples taken from 789 Normative Aging Study (NAS) participants aged 55-100 years, diastolic blood pressure negatively associated with DNAm of *IFN-γ* and positively associated with DNAm of toll-like receptor 2 (*TLR2*) and inducible nitric oxide synthase (*iNOR*) (Alexeeff et al., 2013). Age-related reductions in expression of *TLRs 1-9* as well as nitric oxide synthase has been observed in murine macrophages (Kissin et al., 1997; Renshaw et al., 2002). Another study of NAS participants found that decreased DNAm in LINE-1 repetitive elements associated with increased levels of VCAM-1 in serum samples (Baccarelli et al., 2010). VCAM-1 is an adhesion molecule on endothelial cells that binds lymphocytes (Osborn et al., 1989).

In breast cancer patients, chemotherapy was associated with decreased levels of DNAm at 8 CpG sites, all of which significantly or suggestively associated with levels of inflammatory factors sTNFR2 and IL-6 (Smith et al., 2014). The largest difference in DNAm was observed in 4 CpG sites in exon 11 of transmembrane protein 49, *TMEM49*. These four sites (though not the other four) also associated with CRP levels in the meta-

analysis discussed above (Ligthart et al., 2016), suggesting hypomethylation of these sites as a marker of general inflammation.

Changes to DNAm and DNAhm accompany many of the types of changes that characterize immunosenescence. Further study is needed to ascertain the extent to which these epigenetic changes may mediate immunosenescence. We can gain additional insight about possible roles of DNAm and DNAhm in immunosenescence by drawing upon the literature on disease epigenetics, considering that many diseases are marked by inflammation and share overlapping characteristics with immunosenescence.

6. DNA methylation and hydroxymethylation in chronic diseases

Chronic diseases constitute a complex array of outcomes resulting from aging and immune system dysfunction. In this section, we discuss associations of DNAm and DNAhm with various chronic diseases and how they may relate to immunosenescence.

6.1. DNAm and DNAhm in cognitive decline

Changes in 5mC and 5hmC have been reported in studies of age-related cognitive decline. Irier et al. (2014) observed environmental enrichment via the addition of plastic tubes and toys to reduce global 5hmC in the hippocampus and improve cognitive function in aged mice. In the Senescence Accelerated Mouse P8 (SAMP8), environmental enrichment also improved memory and cognition, which was accompanied by global decreases in 5hmC and increases in 5mC levels, along with decreases in expression of *IL-6* and *CXCL10* (Griñan-Ferré et al., 2016). CXCL10 is involved in the chemotaxis of mononuclear cells (Fife et al., 2001). Chemotaxis refers to the migration of cells toward

increasing or decreasing concentration of a chemical, which, in the case of the immune system, is often released at the site of an infection or injury. These results are consistent with the interpretation that 5mC and 5hmC and attenuated inflammaging are involved in the mitigation of neurodegenerative decline (Griñan-Ferré et al., 2016). Furthermore, mice whose cerebra are chronically hyperfused are used to model age-related cerebrovascular degeneration in humans. Compared to controls, the corpus callosum of hyperfused mice had increased levels of both 5hmC and Iba1-positive inflammatory microglia, but association between 5hmC and Iba1-positive inflammatory microglia was observed in both treatment groups (Tsenkina et al., 2014). Iba1 is specifically secreted by and involved in the activation of microglia, which are macrophages residing in the central nervous system (Greter and Merad, 2012; Ito et al., 1998). Involvement of DNAm and DNAhm in Alzheimer's disease (AD) has also been reported. In brain tissue of 460 individuals diagnosed with AD and 263 controls, differential DNAm between cases and controls was observed in 11 DMRs and associations were observed between DNAm and expression of 8 proximal genes, including the gene coding for RHBDF2 (De Jager et al., 2014), a protein observed to be necessary for the transport of TNF-α in mouse macrophages (Adrain et al., 2012). TNF-α is a molecule of considerable scientific and clinical interest that, to date, has been found to be involved in multiple signaling pathways, inflammation, immunity, and human diseases (Chen and Goeddel, 2002; Wajant et al., 2003). In mice used to model AD pathogenesis, a global decrease of 5hmC in hippocampus was observed (Shu et al., 2016). Differential DNAhm between post-mortem brain samples of individuals with Alzheimer's diseases (n=5) and controls (n=5) was observed in 325 genes (Bernstein et al., 2016).

6.2. DNAm and DNAhm in cancer

It has been long established that altered DNAm is observed in numerous cancers. Weisenberger (2014) notes several studies reporting that regions of repetitive elements, regions with low density of CpG sites, and lamin-associated domains are hypomethylated, while specific loci in CpG islands and shores are hypermethylated in human cancers, and tumor suppressor genes are often silenced due to hypermethylation of their promoter regions (Baylin, 2005). As of January 2, 2014, The Cancer Genome Atlas enumerated 30 cancers characterized by DNAm alterations in humans (Weisenberger, 2014). To date, DNAhm alterations have also been observed in at least 12 human cancer types, most of which are characterized by decreased 5hmC compared to controls as well as a decrease in 5hmC as the cancer progresses, and several studies have proposed that DNAhm could be used as a tool for cancer diagnosis and prognosis (Bhattacharyya et al., 2013; Chapman et al., 2015; Dong et al., 2015; Jäwert et al., 2013; Ko et al., 2010; Kroeze et al., 2014; Larson et al., 2014; Lian et al., 2012; Liao et al., 2016; Liu et al., 2013; Müller et al., 2012; Song et al., 2017; Thomson and Meehan, 2017; Yang et al., 2013; Ye and Li, 2014; Zhang et al., 2015).

DNAm and DNAhm may also be involved with inflammation in cancers. Using cell type specific DNAm microarray data to identify signatures of differential DNAm between myeloid and lymphoid cells, a study of multiple cancers noted that differential DNAm patterns observed in whole blood samples of cancer patients vs. controls were consistent with shifting cell populations: specifically, an increase of myeloid cells and a decrease of lymphoid cells within cancer patients. The study also noted high levels of overlap between

whole blood DNAm changes observed in cancer and those observed in inflammatory diseases (Li et al., 2014). In blood samples of participants from The Normative Aging Study (NAS; n=795), promoter hypermethylation of *IFN-γ*, *ICAM-1* and *IL-6* was observed to associate with prostate cancer incidence, and promoter hypermethylation of *IFN-γ* also associated with all-cancer incidence (Joyce et al., 2015). DNAhm is also suggested to play a role in the infiltration of immune cells into melanoma (Fu et al., 2017).

6.3. DNAm and DNAhm in non-cancerous diseases

Differential DNAm and DNAhm have been observed in non-cancerous chronic diseases as well. Levels of DNAm and DNAhm have been quantified between healthy individuals and those with chronic diseases across various tissues including peripheral blood mononuclear cells (PBMCs), brain, liver, heart, colon, and spinal cord, among others. Investigators have collected cell subtypes from these tissues, such as $CD4^+$ T cells from PBMCs and frontal cortex tissue from the brain. A representative selection of these studies is presented in Supplementary Table 2.1. Notably, many of the diseases showing robust associations with DNAm involve an inflammatory component and may share processes with immunosenescence, suggesting that the wealth of information generated by these studies could inform current knowledge on the relationship between DNAm, DNAhm, and immunosenescence.

Some inflammatory diseases are characterized by hundreds of disease-associated DNAm sites in blood: an example of this is inflammatory bowel disease (IBD). Karatzas et al. (2014) reviewed studies relating DNAm to IBD and its two principal subtypes Crohn's disease (CD) and ulcerative colitis (UC) in blood and other tissues, and

noted IBD-associated DNAm in 19 genes, CD-associated DNAm in 79 genes, and UC-associated DNAm in 91 genes. A systematic review focused on genes linked to inflammatory response identified 25 genes differentially methylated between UC cases and controls or UC inflamed and quiescent mucosa (Gould et al., 2016). More recently, a study comparing DNAm in PBMCs of 240 patients newly-diagnosed with IBD and 190 controls, found 439 sites to be differentially methylated, with nearby genes showing enrichment for immune function. This study also identified three differentially methylated regions which replicated in an independent cohort and covered the genes *TXK, ITGB2*, and *VMP1*. The authors observed IBD-associated hypermethylation of *TXK* promoter DNAm, which associated with reduced expression of *TXK* in whole blood and $CD8^+$ T cells (Ventham et al., 2016). *TXK* expression is involved in the development of human T helper cells (Kashiwakura et al., 1999). *ITGB2*, also known as *CD18*, is involved in leukocyte adhesion (Tan, 2012). As we discuss in section 5, CpG sites in *VMP1*, also known as *TMEM49*, have been reported to associate with chemotherapy and levels of inflammatory factors in humans (Smith et al., 2014), and with human CRP levels in a large meta-analysis (Ligthart et al., 2016). The association of *VMP1* DNAm with multiple conditions and inflammatory factors suggest its involvement in inflammation in general.

5mC and 5hmC has also been reported to be involved in the inflammation and immune system decline in systemic lupus erythematosus (SLE). SLE is an autoimmune disease that typically affects individuals in mid to late life, and shares many characteristics with immunosenescence, most notably in T cells (van den Hoogen et al., 2015). Wu et al. (2016) reviewed studies indicating that there is a global decrease of DNAm in PBMCs, B cells, and $CD4^+$ T cells in SLE, including demethylation of a number of genes encoding

CD molecules, cytokines, and pro-inflammatory markers. A recent study used RNA interference to experimentally alter global methylation levels by knocking down *BDH2* expression in $CD4^+$ T cells of SLE patients, and found that the induced hyper-DNAhm and hypo-DNAm resulted in increased expression of autoimmune-related genes such as *CD70*, *CD11a*, *CD40L*, and *PRF1* (M. Zhao et al., 2017). Overall, this suggests that, in the $CD4^+$ T cells of SLE patients, increased global DNAhm and decreased global DNAm can alter expression of the aforementioned autoimmune-related genes.

Studies investigating other chronic diseases involving an inflammatory response have also found evidence of a relationship between DNAm and inflammation. In white blood cells of patients with chronic obstructive pulmonary disease differential DNAm was observed at 349 CpG sites near genes enriched for immune and inflammatory system pathways (Qiu et al., 2012). In PBMCs of chronic hepatitis B patients, hyper-DNAm of *PPAR-γ* associated with inhibition of its transcription, and with liver inflammation and fibrosis, although the particular function of *PPAR-γ* in the liver is unclear (Zhao et al., 2013). In fibroblast-like synoviocytes (cells residing in joint cavities) of patients with rheumatoid arthritis compared to those with osteoarthritis, Karouzakis et al. (2014) observed promoter hypomethylation accompanied by high expression of T-box transcription factor (*TBX5*). In a human synovial sarcoma cell line, knocking down the expression of *TBX5* decreased and overexpression of *TBX5* increased production of pro-inflammatory cytokines (Hussain et al., 2018). Collectively, these studies suggest a potential interaction between DNAm and expression of inflammatory genes across a number of inflammatory chronic diseases, although the exact regulatory roles are not yet established. In fibroblast-like synoviocytes isolated from 14 patients with rheumatoid

arthritis and 12 patients with osteoarthritis that were stimulated with either IL-1β or TNF, *DNMT3A* expression decreased after stimulation of either cytokine and *DNMT1* expression decreased after IL-1β stimulation (Nakano et al., 2013). While other experimental work in humans has shown evidence of DNAm changes influencing levels of immune-related factors via altered expression (M. Zhao et al., 2017), this study raises the possibility that changes in immune-related factors may influence DNAm.

Changes in 5mC and 5hmC have also been observed in inflammation of other tissues, including skin, kidney, and spinal cord. Low global levels of 5hmC were observed in samples from lesional and perilesional skin cells of individuals with an inflammatory skin disease known as hidradenitis suppurativa (n=30) compared to healthy controls (n=30) suggesting DNAhm may play a role in skin cell inflammation as well (Hessam et al., 2017). In a study investigating the effect of diabetes on kidney function, decreased mitochondrial 5mC and 5hmC was observed in kidney tissue of streptozotocin-induced diabetic rats and accompanied diminished uric acid clearance (de Oliveira et al., 2017), the lack of which likely contributes to kidney inflammation (Wang et al., 2012). In the spinal cord of mice with formalin-induced acute inflammatory pain, an increase of TET1 and TET3 as well as 5hmC was observed. Moreover, injection of *TET1*-siRNA or *TET3*-siRNA in mice decreased 5hmC and alleviated formalin-induced nociceptive response compared to controls, suggesting that 5hmC may regulate nociceptive behavior (Pan et al., 2016).

Taken together, these findings highlight a potential role of DNAm and DNAhm in immune system decline with age in chronic disease patients, and demonstrate the utility of examining DNAm and DNAhm to better understand immunosenescence in numerous

chronic diseases. The majority of studies contributing to our current understanding of immunosenescence have not interrogated DNAm and DNAhm. Other work has observed DNAm and DNAhm to associate with immune-related factors, but has not directly investigated age. Moreover, the majority of results discussed above reflect associations rather than causal relationships. Thus, there are gaps in our understanding of the potential relationship between immunosenescence and these two epigenetic modifications. Below we outline the design of future studies suited to further our understanding of the role of DNAm and DNAhm in immunosenescence.

7. Future Directions

Given the DNAm, DNAhm, and gene expression changes reported to occur with aging, disease state, and features of immunosenescence such as lineage skewing, it is tempting to postulate potential causal mechanisms driving immunosenescence. For example, changes to DNAm and DNAhm in HSCs could drive changes to gene expression resulting in altered differentiation and other phenotypic changes to HSCs observed with age. It is possible that many of the immunosenescence-related changes observed in PBMCs are products of changes that begin in HSCs. Alternatively, changes observed in immune cells could originate subsequent to differentiation from HSCs. The mechanisms behind immunosenescence are difficult to disentangle given that the vast majority of studies reviewed here investigate associations, and the experimental studies have shown mixed evidence. The limited evidence establishing causal relationships constitutes a large gap for future studies to fill. Below, we suggest several promising strategies to investigate potential mechanisms underlying immunosenescence.

First, more controlled experimental studies are needed to directly assess potential causal relationships. For example, transplanting young, healthy bone marrow into immunosenescent murine models and observing changes such as a reversal of lineage skewing and a reversion to younger phenotypes within white blood cell lineages, as well as concurrent changes to DNAm, DNAhm, and gene expression within these cell types, could determine whether immunosenescent phenotypes are rooted in age-related phenotypic changes to HSCs. Observing DNAm, DNAhm, and gene expression changes in patients undergoing bone marrow transplants could also be similarly fruitful. Knockouts or siRNAs that target DNMT and TET genes known to manipulate DNAm and DNAhm (i.e. *DNMT1*, *DNMT3a*, *DNMT3b, TET1*, *TET2*, *TET3*), could help establish whether DNAm and DNAhm mediate phenotypic changes characteristic of immunosenescence, or are merely markers of such changes. While these experiments would alter global levels of DNAm and DNAhm, advancements such as CRISPR-Cas9 raise the possibility of experimentally manipulating DNAm and DNAhm at specific CpG sites (Hsu et al., 2014; Liu et al., 2016).

As a complement to experimental approaches, analytical approaches such as Mendelian Randomization (MR) have been proposed that are capable of making causal inferences from existing epidemiological data (Davey Smith and Hemani, 2014). While such methods are not substitutes for controlled experiments, they can be easily applied to human studies, can generate discoveries to be further validated through experimentation, and can take advantage of the wealth of large genomic datasets generated over the last decade. There are many publicly available datasets to which this approach could be applied. For example, the original Framingham Heart Study is a

longitudinal cohort of 5,209 participants of a wide age range with stored blood samples and an extremely rich set of adjudicated disease phenotype data (Mahmood et al., 2014). Application of MR to large datasets enables inference on whether locus-specific DNAm or DNAhm that associates with the expression of an immune-related factor is 1) causal or 2) consequential to changes in expression of the factor. Such inferences directly address the question of whether these epigenetic modifications regulate immunosenescence or simply mark it. MR relies on the assumption that alleles are passed to offspring independently of potential confounders such as environmental exposures. This assumption allows the use of genetic variants as instrumental variables that can be used to mimic randomization of levels of DNAm or immune factors. MR can thus be used to establish directional causal relationships between site-specific DNAm and immune-related factors, and to distinguish causal relationships from the situation where both DNAm and immune factors are influenced by a third confounding factor (Burgess et al., 2015).

Single cell sequencing is another approach that does not require an experimental setting that could further elucidate regulatory relationships in immunosenescence. For example, within a tissue sample, it could be the case that some cells have decreased locus-specific DNAm upregulating expression of an immune factor while other cells have increased DNAm at the same locus downregulating expression of the same factor. If DNAm and gene expression signals were measured across a cell population, these signals would cancel each other out. With single cell sequencing of both the methylome and transcriptome, however, such signals would be detectable (Clark et al., 2018;

Smallwood et al., 2014). Similar arguments can be applied to DNAhm of immune-related factors.

Each of these approaches has the potential to reinforce each other. The advantage of controlled experiments is that tissues and cell types can be experimentally manipulated to more directly interrogate immunosenescence-related processes, holding constant potential confounding factors. The advantage of the other approaches is that they may be applied to human studies, since they can use existing data from human tissue samples. MR has the potential to draw causal inferences at specific sites, while RNA interference approaches to manipulate levels of DNA methyltransferases and TET proteins allow inference of causal effects of changes in global levels of DNAm and DNAhm. However, if MR studies demonstrate evidence of epigenetic regulation of expression of particular genes involved in immunosenescence, RNA interference could be used to investigate the impact of knocking out those genes on the expression of other genes, identifying potential downstream pathways. Further, as the technology becomes more widely adapted, CRISPR-Cas9 can be used to validate causal inferences by experimentally manipulating DNAm and DNAhm at specific CpG sites. Therefore, a coordinated effort involving all of these strategies has the potential to be particularly fruitful.

8. Conclusion

Numerous studies support the interpretation that DNAm and DNAhm play a role in immune system decline, most notably in white blood cell types and HSCs. Here, we have highlighted findings of DNAm/DNAhm changes associated with many features of

immunosenescence, including lineage skewing, inflammatory factors, aging, and disease state (Table 2.1). However, for the most part the causal relationships underlying these associations and defining the role of DNAm/DNAhm in immunosenescence remain to be elucidated. To fill this gap, we have proposed a multi-pronged approach involving both experimental and observational studies to further our understanding of the specific roles DNAm and DNAhm may play in immunosenescence.

Characteristic	Description	Current understanding of connection with DNAm	Current understanding of connection with DNAhm
Aging (Section 3)	The general deterioration of the organism associated with an increased risk of morbidity and mortality.	Evidence of widespread site-specific changes in human blood cells and other tissues	Preliminary evidence of region- and site-specific changes in mouse and human brain tissue
Lineage skewing (Section 4)	Skewing toward the myeloid lineage in HSCs	Evidence of causal role for DNAm in HSC differentiation. (Challen et al. 2011)	Evidence for causal role of DNAhm in skewing toward myeloid lineage of HSCs (Ko et al. 2010; Moran-Crusio et al. 2011).
	Changes in T cell subsets	Site-specific changes (Lee et al. 2002; Jones and Chen 2006; Tserel 2015)	Uncharacterized
Inflammaging (Section 5)	Increase in stably low levels of inflammation	Numerous site-specific associations observed with inflammation and inflammatory diseases	Associations found in some inflammatory diseases Evidence of causal role inflammatory pain (Pan et al. 2016)
Chronic diseases (Section 6)	Chronic diseases can be viewed as complex outcomes of aging that vary markedly by severity and tissue affected.	Evidence of causal role in SLE (M. Zhao et al. 2017) Associations with numerous chronic diseases	

Table 2.1: Summary of characteristics of immunosenescence and associated epigenetic changes

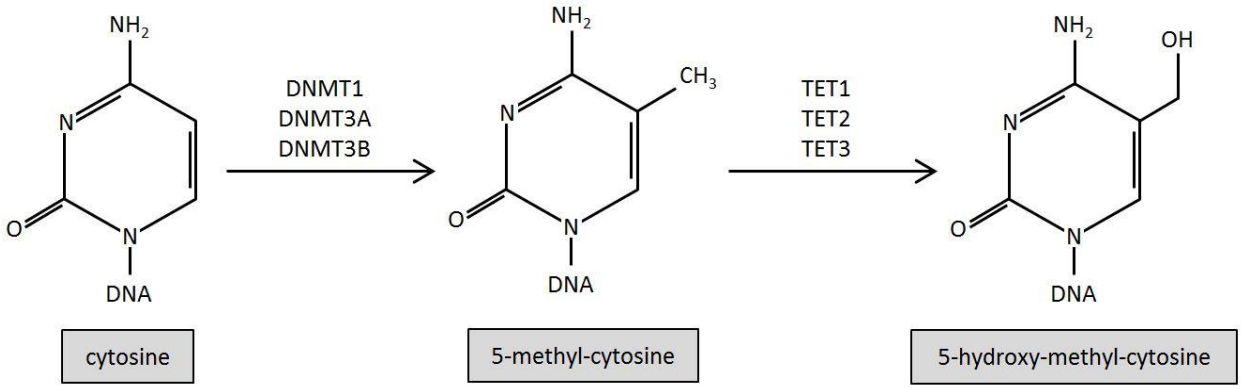


Figure 2.1. DNA methyltransferases (DNMT1, DNMT3A, DNMT3B) catalyze the methylation of cytosine to 5-methylcytosine (left and center), and the TET proteins catalyze the hydroxymethylation of 5-methylcytosine to 5-hydroxymethylcytosine.

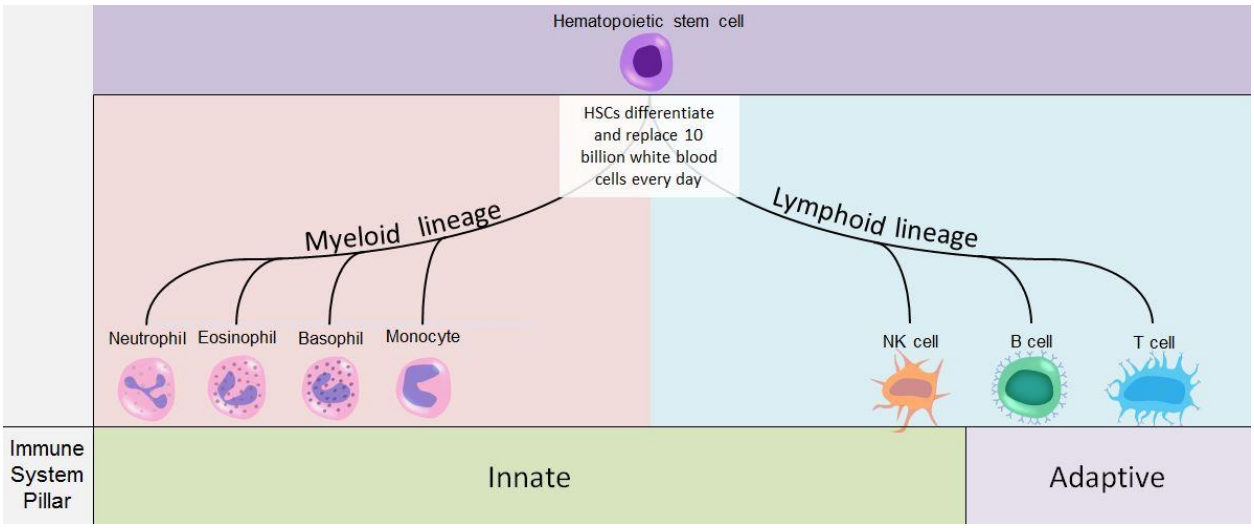


Figure 2.2. Lineage differentiation of hematopoietic stem cells into cell types of the myeloid and lymphoid lineage.

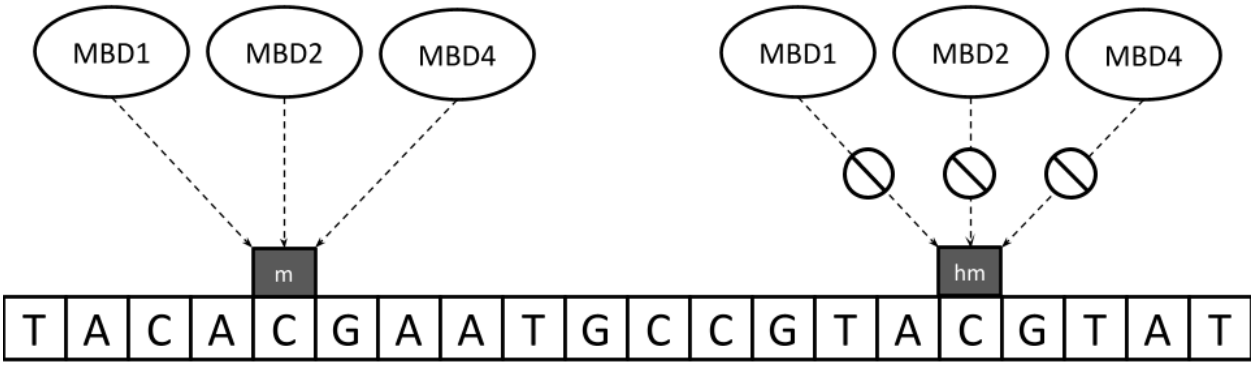


Figure 2.3. Methyl-binding proteins (MBD1, MBD2, MBD4), which repress transcription, can bind to methyl groups (m), but not hydroxymethyl groups (hm).

REFERENCES

Adrain, C., Zettl, M., Christova, Y., Taylor, N., Freeman, M., 2012. Tumor necrosis factor signaling requires iRhom2 to promote trafficking and activation of TACE. Science 335, 225–228.

Agarwal, S., Busse, P.J., 2010. Innate and adaptive immunosenescence. Ann. Allergy Asthma Immunol. 104, 183–190.

Agrawal, A., Gupta, S., 2011. Impact of aging on dendritic cell functions in humans. Ageing Res. Rev. 10, 336–345.

Alexeeff, S.E., Baccarelli, A.A., Halonen, J., Coull, B.A., Wright, R.O., Tarantini, L., Bollati, V., Sparrow, D., Vokonas, P., Schwartz, J., 2013. Association between blood pressure and DNA methylation of retrotransposons and pro-inflammatory genes. Int. J. Epidemiol. 42, 270–280.

Alisch, R.S., Barwick, B.G., Chopra, P., Myrick, L.K., Satten, G.A., Conneely, K.N., Warren, S.T., 2012. Age-associated DNA methylation in pediatric populations. Genome Res. 22, 623–632.

Arosa, F.A., 2002. CD8+CD28- T cells: certainties and uncertainties of a prevalent human T-cell subset. Immunol. Cell Biol. 80, 1–13.

Arosa, F.A., Esgalhado, A.J., Padrão, C.A., Cardoso, E.M., 2016. Divide, Conquer, and Sense: CD8CD28 T Cells in Perspective. Front. Immunol. 7, 665.

Bacalini, M.G., Deelen, J., Pirazzini, C., De Cecco, M., Giuliani, C., Lanzarini, C., Ravaioli, F., Marasco, E., van Heemst, D., Suchiman, H.E.D., Slieker, R., Giampieri, E., Recchioni, R., Mercheselli, F., Salvioli, S., Vitale, G., Olivieri, F., Spijkerman, A.M.W., Dollé, M.E.T., Sedivy, J.M., Castellani, G., Franceschi, C., Slagboom, P.E., Garagnani, P., 2017. Systemic Age-Associated DNA Hypermethylation of ELOVL2 Gene: In Vivo and In Vitro Evidences of a Cell Replication Process. J. Gerontol. A Biol. Sci. Med. Sci. 72, 1015–1023.

Baccarelli, A., Tarantini, L., Wright, R.O., Bollati, V., Litonjua, A.A., Zanobetti, A., Sparrow, D., Vokonas, P.S., Schwartz, J., 2010. Repetitive element DNA methylation and circulating endothelial and inflammation markers in the VA normative aging study. Epigenetics 5, 222–228.

Baylin, S.B., 2005. DNA methylation and gene silencing in cancer. Nat. Clin. Pract. Oncol. 2 Suppl 1, S4–11.

Beerman, I., Bock, C., Garrison, B.S., Smith, Z.D., Gu, H., Meissner, A., Rossi, D.J., 2013. Proliferation-dependent alterations of the DNA methylation landscape underlie hematopoietic stem cell aging. Cell Stem Cell 12, 413–425.

Bell, J.T., Tsai, P.-C., Yang, T.-P., Pidsley, R., Nisbet, J., Glass, D., Mangino, M., Zhai, G., Zhang, F., Valdes, A., Shin, S.-Y., Dempster, E.L., Murray, R.M., Grundberg, E., Hedman, A.K., Nica, A., Small, K.S., Dermitzakis, E.T., McCarthy, M.I., Mill, J., Spector, T.D., Deloukas, P., The MuTHER Consortium, 2012. Epigenome-Wide Scans Identify Differentially Methylated Regions for Age and Age-Related Phenotypes in a Healthy Ageing Population. PLoS Genet. 8, e1002629.

Bernstein, A.I., Lin, Y., Street, R.C., Lin, L., Dai, Q., Yu, L., Bao, H., Gearing, M., Lah, J.J., Nelson, P.T., He, C., Levey, A.I., Mullé, J.G., Duan, R., Jin, P., 2016. 5-Hydroxymethylation-associated epigenetic modifiers of Alzheimer's disease modulate Tau-induced neurotoxicity. Hum. Mol. Genet. 25, 2437–2450.

Bhattacharyya, S., Yu, Y., Suzuki, M., Campbell, N., Mazdo, J., Vasanthakumar, A.,

Bhagat, T.D., Nischal, S., Christopeit, M., Parekh, S., Steidl, U., Godley, L., Maitra, A., Greally, J.M., Verma, A., 2013. Genome-wide hydroxymethylation tested using the HELP-GT assay shows redistribution in cancer. Nucleic Acids Res. 41, e157.

Bjornsson, H.T., Sigurdsson, M.I., Fallin, M.D., Irizarry, R.A., Aspelund, T., Cui, H., Yu, W., Rongione, M.A., Ekström, T.J., Harris, T.B., Launer, L.J., Eiriksdottir, G., Leppert, M.F., Sapienza, C., Gudnason, V., Feinberg, A.P., 2008. Intra-individual change over time in DNA methylation with familial clustering. JAMA 299, 2877–2883.

Bollati, V., Favero, C., Albetti, B., Tarantini, L., Moroni, A., Byun, H.-M., Motta, V., Conti, D., Tirelli, A., Vigna, L., Bertazzi, P., Pesatori, A., 2014. Nutrients Intake is Associated with DNA Methylation of Candidate Inflammatory Genes in a Population of Obese Subjects. Nutrients 6, 4625–4639.

Booth, M.J., Branco, M.R., Ficz, G., Oxley, D., Krueger, F., Reik, W., Balasubramanian, S., 2012. Quantitative sequencing of 5-methylcytosine and 5-hydroxymethylcytosine at single-base resolution. Science 336, 934–937.

Boyes, J., Bird, A., 1991. DNA methylation inhibits transcription indirectly via a methyl-CpG binding protein. Cell 64, 1123–1134.

Britanova, O.V., Putintseva, E.V., Shugay, M., Merzlyak, E.M., Turchaninova, M.A., Staroverov, D.B., Bolotin, D.A., Lukyanov, S., Bogdanova, E.A., Mamedov, I.Z., Lebedev, Y.B., Chudakov, D.M., 2014. Age-related decrease in TCR repertoire diversity measured with deep and normalized sequence profiling. J. Immunol. 192, 2689–2698.

Brüünsgaard, H., Pedersen, B.K., 2003. Age-related inflammatory cytokines and disease. Immunol. Allergy Clin. North Am. 23, 15–39.

Burgess, S., Timpson, N.J., Ebrahim, S., Davey Smith, G., 2015. Mendelian randomization: where are we now and where are we going? Int. J. Epidemiol. 44, 379–388.

Cancro, M.P., Hao, Y., Scholz, J.L., Riley, R.L., Frasca, D., Dunn-Walters, D.K., Blomberg, B.B., 2009. B cells and aging: molecules and mechanisms. Trends Immunol. 30, 313–318.

Cantone, L., Iodice, S., Tarantini, L., Albetti, B., Restelli, I., Vigna, L., Bonzini, M., Pesatori, A.C., Bollati, V., 2017. Particulate matter exposure is associated with inflammatory gene methylation in obese subjects. Environ. Res. 152, 478–484.

Cao, J.-N., Gollapudi, S., Sharman, E.H., Jia, Z., Gupta, S., 2010. Age-related alterations of gene expression patterns in human CD8 T cells. Aging Cell 9, 453–453.

Challen, G.A., Sun, D., Jeong, M., Luo, M., Jelinek, J., Berg, J.S., Bock, C., Vasanthakumar, A., Gu, H., Xi, Y., Liang, S., Lu, Y., Darlington, G.J., Meissner, A., Issa, J.-P.J., Godley, L.A., Li, W., Goodell, M.A., 2011. Dnmt3a is essential for hematopoietic stem cell differentiation. Nat. Genet. 44, 23–31.

Chapman, C.G., Mariani, C.J., Wu, F., Meckel, K., Butun, F., Chuang, A., Madzo, J., Bissonette, M.B., Kwon, J.H., Godley, L.A., 2015. TET-catalyzed 5-hydroxymethylcytosine regulates gene expression in differentiating colonocytes and colon cancer. Sci. Rep. 5, 17568.

Chen, G., Goeddel, D.V., 2002. TNF-R1 signaling: a beautiful pathway. Science 296, 1634–1635.

Chen, G., Lustig, A., Weng, N.-P., 2013. T cell aging: a review of the transcriptional

changes determined from genome-wide analysis. Front. Immunol. 4, 121.
Christensen, B.C., Houseman, E.A., Marsit, C.J., Zheng, S., Wrensch, M.R., Wiemels, J.L., Nelson, H.H., Karagas, M.R., Padbury, J.F., Bueno, R., Sugarbaker, D.J., Yeh, R.-F., Wiencke, J.K., Kelsey, K.T., 2009. Aging and environmental exposures alter tissue-specific DNA methylation dependent upon CpG island context. PLoS Genet. 5, e1000602.
Clark, S.J., Argelaguet, R., Kapourani, C.-A., Stubbs, T.M., Lee, H.J., Alda-Catalinas, C., Krueger, F., Sanguinetti, G., Kelsey, G., Marioni, J.C., Stegle, O., Reik, W., 2018. scNMT-seq enables joint profiling of chromatin accessibility DNA methylation and transcription in single cells. Nat. Commun. 9, 781.
Colquitt, B.M., Allen, W.E., Barnea, G., Lomvardas, S., 2013. Alteration of genic 5-hydroxymethylcytosine patterning in olfactory neurons correlates with changes in gene expression and cell identity. Proc. Natl. Acad. Sci. U. S. A. 110, 14682–14687.
D'Aquila, P., Giordano, M., Montesanto, A., De Rango, F., Passarino, G., Bellizzi, D., 2015. Age-and gender-related pattern of methylation in the MT-RNR1 gene. Epigenomics 7, 707–716.
Davey Smith, G., Hemani, G., 2014. Mendelian randomization: genetic anchors for causal inference in epidemiological studies. Hum. Mol. Genet. 23, R89–98.
De Jager, P.L., Srivastava, G., Lunnon, K., Burgess, J., Schalkwyk, L.C., Yu, L., Eaton, M.L., Keenan, B.T., Ernst, J., McCabe, C., Tang, A., Raj, T., Replogle, J., Brodeur, W., Gabriel, S., Chai, H.S., Younkin, C., Younkin, S.G., Zou, F., Szyf, M., Epstein, C.B., Schneider, J.A., Bernstein, B.E., Meissner, A., Ertekin-Taner, N., Chibnik, L.B., Kellis, M., Mill, J., Bennett, D.A., 2014. Alzheimer's disease: early alterations in brain DNA methylation at ANK1, BIN1, RHBDF2 and other loci. Nat. Neurosci. 17, 1156–1163.
de Oliveira, A.A.F., de Oliveira, T.F., Bobadilla, L.L., Garcia, C.C.M., Berra, C.M., de Souza-Pinto, N.C., Medeiros, M.H.G., Di Mascio, P., Zatz, R., de M Loureiro, A.P., 2017. Sustained kidney biochemical derangement in treated experimental diabetes: a clue to metabolic memory. Sci. Rep. 7, 40544.
Diez Roux, A.V., Mair, C., 2010. Neighborhoods and health. Ann. N. Y. Acad. Sci. 1186, 125–145.
Dong, Z.-R., Zhang, C., Cai, J.-B., Zhang, P.-F., Shi, G.-M., Gao, D.-M., Sun, H.-C., Qiu, S.-J., Zhou, J., Ke, A.-W., Fan, J., 2015. Role of 5-hydroxymethylcytosine level in diagnosis and prognosis prediction of intrahepatic cholangiocarcinoma. Tumour Biol. 36, 2763–2771.
Faurschou, M., Borregaard, N., 2003. Neutrophil granules and secretory vesicles in inflammation. Microbes Infect. 5, 1317–1327.
Fife, B.T., Kennedy, K.J., Paniagua, M.C., Lukacs, N.W., Kunkel, S.L., Luster, A.D., Karpus, W.J., 2001. CXCL10 (IFN-gamma-inducible protein-10) control of encephalitogenic CD4+ T cell accumulation in the central nervous system during experimental autoimmune encephalomyelitis. J. Immunol. 166, 7617–7624.
Florath, I., Butterbach, K., Müller, H., Bewerunge-Hudler, M., Brenner, H., 2014. Cross-sectional and longitudinal changes in DNA methylation with age: an epigenome-wide analysis revealing over 60 novel age-associated CpG sites. Hum. Mol. Genet. 23, 1186–1201.

Franceschi, C., Bonafè, M., Valensin, S., Olivieri, F., De Luca, M., Ottaviani, E., De Benedictis, G., 2000. Inflamm-aging. An evolutionary perspective on immunosenescence. Ann. N. Y. Acad. Sci. 908, 244–254.

Fulop, T., Larbi, A., Dupuis, G., Le Page, A., Frost, E.H., Cohen, A.A., Witkowski, J.M., Franceschi, C., 2017. Immunosenescence and Inflamm-Aging As Two Sides of the Same Coin: Friends or Foes? Front. Immunol. 8, 1960.

Fu, S., Wu, H., Zhang, H., Lian, C.G., Lu, Q., 2017. DNA methylation/hydroxymethylation in melanoma. Oncotarget 8, 78163–78173.

Garagnani, P., Bacalini, M.G., Pirazzini, C., Gori, D., Giuliani, C., Mari, D., Di Blasio, A.M., Gentilini, D., Vitale, G., Collino, S., Rezzi, S., Castellani, G., Capri, M., Salvioli, S., Franceschi, C., 2012. Methylation of ELOVL2 gene as a new epigenetic marker of age. Aging Cell 11, 1132–1134.

Geiger, H., de Haan, G., Carolina Florian, M., 2013. The ageing haematopoietic stem cell compartment. Nat. Rev. Immunol. 13, 376–389.

Geissmann, F., Manz, M.G., Jung, S., Sieweke, M.H., Merad, M., Ley, K., 2010. Development of Monocytes, Macrophages, and Dendritic Cells. Science 327, 656–661.

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, 277–283.

Ginhoux, F., Jung, S., 2014. Monocytes and macrophages: developmental pathways and tissue homeostasis. Nat. Rev. Immunol. 14, 392–404.

Giuliani, C., Cilli, E., Bacalini, M.G., Pirazzini, C., Sazzini, M., Gruppioni, G., Franceschi, C., Garagnani, P., Luiselli, D., 2016. Inferring chronological age from DNA methylation patterns of human teeth. Am. J. Phys. Anthropol. 159, 585–595.

Godderis, L., Schouteden, C., Tabish, A., Poels, K., Hoet, P., Baccarelli, A.A., Van Landuyt, K., 2015. Global Methylation and Hydroxymethylation in DNA from Blood and Saliva in Healthy Volunteers. Biomed Res. Int. 2015, 845041.

Gopalan, S., Carja, O., Fagny, M., Patin, E., Myrick, J.W., McEwen, L.M., Mah, S.M., Kobor, M.S., Froment, A., Feldman, M.W., Quintana-Murci, L., Henn, B.M., 2017. Trends in DNA Methylation with Age Replicate Across Diverse Human Populations. Genetics 206, 1659–1674.

Gould, N.J., Davidson, K.L., Nwokolo, C.U., Arasaradnam, R.P., 2016. A systematic review of the role of DNA methylation on inflammatory genes in ulcerative colitis. Epigenomics 8, 667–684.

Greter, M., Merad, M., 2012. Regulation of microglia development and homeostasis. Glia 61, 121–127.

Griñan-Ferré, C., Puigoriol-Illamola, D., Palomera-Ávalos, V., Pérez-Cáceres, D., Companys-Alemany, J., Camins, A., Ortuño-Sahagún, D., Rodrigo, M.T., Pallàs, M., 2016. Environmental Enrichment Modified Epigenetic Mechanisms in SAMP8 Mouse Hippocampus by Reducing Oxidative Stress and Inflammaging and Achieving Neuroprotection. Front. Aging Neurosci. 8, 241.

Gupta, S., 2014. Role of dendritic cells in innate and adaptive immune response in human aging. Exp. Gerontol. 54, 47–52.

Haase, H., Rink, L., 2009. The immune system and the impact of zinc during aging.

Immun. Ageing 6, 9.
Hannum, G., Guinney, J., Zhao, L., Zhang, L., Hughes, G., Sadda, S., Klotzle, B., Bibikova, M., Fan, J.-B., Gao, Y., Deconde, R., Chen, M., Rajapakse, I., Friend, S., Ideker, T., Zhang, K., 2013. Genome-wide methylation profiles reveal quantitative views of human aging rates. Mol. Cell 49, 359–367.
Hashimoto, D., Miller, J., Merad, M., 2011. Dendritic cell and macrophage heterogeneity in vivo. Immunity 35, 323–335.
Hearps, A.C., Martin, G.E., Angelovich, T.A., Cheng, W.-J., Maisa, A., Landay, A.L., Jaworowski, A., Crowe, S.M., 2012. Aging is associated with chronic innate immune activation and dysregulation of monocyte phenotype and function. Aging Cell 11, 867–875.
Hessam, S., Sand, M., Lang, K., Käfferlein, H.U., Scholl, L., Gambichler, T., Skrygan, M., Brüning, T., Stockfleth, E., Bechara, F.G., 2017. Altered Global 5-Hydroxymethylation Status in Hidradenitis Suppurativa: Support for an Epigenetic Background. Dermatology 233, 129–135.
He, Y.-F., Li, B.-Z., Li, Z., Liu, P., Wang, Y., Tang, Q., Ding, J., Jia, Y., Chen, Z., Li, L., Sun, Y., Li, X., Dai, Q., Song, C.-X., Zhang, K., He, C., Xu, G.-L., 2011. Tet-mediated formation of 5-carboxylcytosine and its excision by TDG in mammalian DNA. Science 333, 1303–1307.
Höbartner, C., 2011. Enzymatic labeling of 5-hydroxymethylcytosine in DNA. Angew. Chem. Int. Ed Engl. 50, 4268–4270.
Horvath, S., 2013. DNA methylation age of human tissues and cell types. Genome Biol. 14, R115.
Horvath, S., Pirazzini, C., Bacalini, M.G., Gentilini, D., Di Blasio, A.M., Delledonne, M., Mari, D., Arosio, B., Monti, D., Passarino, G., De Rango, F., D'Aquila, P., Giuliani, C., Marasco, E., Collino, S., Descombes, P., Garagnani, P., Franceschi, C., 2015. Decreased epigenetic age of PBMCs from Italian semi-supercentenarians and their offspring. Aging 7, 1159–1170.
Horvath, S., Raj, K., 2018. DNA methylation-based biomarkers and the epigenetic clock theory of ageing. Nat. Rev. Genet. 19, 371–384.
Houseman, E.A., Accomando, W.P., Koestler, D.C., Christensen, B.C., Marsit, C.J., Nelson, H.H., Wiencke, J.K., Kelsey, K.T., 2012. DNA methylation arrays as surrogate measures of cell mixture distribution. BMC Bioinformatics 13, 86.
Hsu, P.D., Lander, E.S., Zhang, F., 2014. Development and Applications of CRISPR-Cas9 for Genome Engineering. Cell 157, 1262–1278.
Huang, Y., Liu, L., Ma, D., Liao, Y., Lu, Y., Huang, H., Qin, W., Liu, X., Fang, F., 2017. Human cytomegalovirus triggers the assembly of AIM2 inflammasome in THP-1-derived macrophages. J. Med. Virol. 89, 2188–2195.
Hussain, N., Zhu, W., Jiang, C., Xu, J., Wu, X., Geng, M., Hussain, S., Cai, Y., Xu, K., Xu, P., Han, Y., Sun, J., Meng, L., Lu, S., 2018. Down-regulation of miR-10a-5p in synoviocytes contributes to TBX5-controlled joint inflammation. J. Cell. Mol. Med. 22, 241–250.
Iacobazzi, V., Castegna, A., Infantino, V., Andria, G., 2013. Mitochondrial DNA methylation as a next-generation biomarker and diagnostic tool. Mol. Genet. Metab. 110, 25–34.
Illingworth, R.S., Bird, A.P., 2009. CpG islands - "A rough guide." FEBS Lett. 583, 1713–

1720.
Irier, H., Craig Street, R., Dave, R., Lin, L., Cai, C., Davis, T.H., Yao, B., Cheng, Y., Jin, P., 2014. Environmental enrichment modulates 5-hydroxymethylcytosine dynamics in hippocampus. Genomics 104, 376–382.
Ito, D., Imai, Y., Ohsawa, K., Nakajima, K., Fukuuchi, Y., Kohsaka, S., 1998. Microglia-specific localisation of a novel calcium binding protein, Iba1. Brain Res. Mol. Brain Res. 57, 1–9.
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, 1129–1133.
Ito, S., Shen, L., Dai, Q., Wu, S.C., Collins, L.B., Swenberg, J.A., He, C., Zhang, Y., 2011. Tet proteins can convert 5-methylcytosine to 5-formylcytosine and 5-carboxylcytosine. Science 333, 1300–1303.
Jakubzick, C.V., Randolph, G.J., Henson, P.M., 2017. Monocyte differentiation and antigen-presenting functions. Nat. Rev. Immunol. 17, 349–362.
Janeway, C.A., 2001. Immunobiology: The Immune System in Health and Disease. Churchill Livingstone.
Jasiulionis, M.G., 2018. Abnormal Epigenetic Regulation of Immune System during Aging. Front. Immunol. 9, 197.
Jäwert, F., Hasséus, B., Kjeller, G., Magnusson, B., Sand, L., Larsson, L., 2013. Loss of 5-hydroxymethylcytosine and TET2 in oral squamous cell carcinoma. Anticancer Res. 33, 4325–4328.
Jin, S.-G., Kadam, S., Pfeifer, G.P., 2010. Examination of the specificity of DNA methylation profiling techniques towards 5-methylcytosine and 5-hydroxymethylcytosine. Nucleic Acids Res. 38, e125.
Johnson, N.D., Wiener, H.W., Smith, A.K., Nishitani, S., Absher, D.M., Arnett, D.K., Aslibekyan, S., Conneely, K.N., 2017. Non-linear patterns in age-related DNA methylation may reflect CD4 T cell differentiation. Epigenetics 12, 492–503.
Jones, B., Chen, J., 2006. Inhibition of IFN-γ transcription by site-specific methylation during T helper cell development. EMBO J. 25, 2443–2452.
Jones, P.A., 2012. Functions of DNA methylation: islands, start sites, gene bodies and beyond. Nat. Rev. Genet. 13, 484–492.
Joyce, B.T., Gao, T., Liu, L., Zheng, Y., Liu, S., Zhang, W., Penedo, F., Dai, Q., Schwartz, J., Baccarelli, A.A., Hou, L., 2015. Longitudinal Study of DNA Methylation of Inflammatory Genes and Cancer Risk. Cancer Epidemiol. Biomarkers Prev. 24, 1531–1538.
Karatzas, P.S., Gazouli, M., Safioleas, M., Mantzaris, G.J., 2014. DNA methylation changes in inflammatory bowel disease. Ann. Gastroenterol. Hepatol. 27, 125–132.
Karouzakis, E., Trenkmann, M., Gay, R.E., Michel, B.A., Gay, S., Neidhart, M., 2014. Epigenome analysis reveals TBX5 as a novel transcription factor involved in the activation of rheumatoid arthritis synovial fibroblasts. J. Immunol. 193, 4945–4951.
Karpf, A.R., 2012. Epigenetic Alterations in Oncogenesis. Springer Science & Business Media.
Kashiwakura, J.-I., Suzuki, N., Nagafuchi, H., Takeno, M., Takeba, Y., Shimoyama, Y., Sakane, T., 1999. Txk, a Nonreceptor Tyrosine Kinase of the Tec Family, Is

Expressed in T Helper Type 1 Cells and Regulates Interferon γ Production in Human T Lymphocytes. J. Exp. Med. 190, 1147–1154.

Kim, D., Kubzansky, L.D., Baccarelli, A., Sparrow, D., Spiro, A., 3rd, Tarantini, L., Cantone, L., Vokonas, P., Schwartz, J., 2016. Psychological factors and DNA methylation of genes related to immune/inflammatory system markers: the VA Normative Aging Study. BMJ Open 6, e009790.

Kissin, E., Tomasi, M., McCartney-Francis, N., Gibbs, C.L., Smith, P.D., 1997. Age-related decline in murine macrophage production of nitric oxide. J. Infect. Dis. 175, 1004–1007.

Ko, M., Huang, Y., Jankowska, A.M., Pape, U.J., Tahiliani, M., Bandukwala, H.S., An, J., Lamperti, E.D., Koh, K.P., Ganetzky, R., Liu, X.S., Aravind, L., Agarwal, S., Maciejewski, J.P., Rao, A., 2010. Impaired hydroxylation of 5-methylcytosine in myeloid cancers with mutant TET2. Nature 468, 839–843.

Kroeze, L.I., Aslanyan, M.G., van Rooij, A., Koorenhof-Scheele, T.N., Massop, M., Carell, T., Boezeman, J.B., Marie, J.-P., Halkes, C.J.M., de Witte, T., Huls, G., Suciu, S., Wevers, R.A., van der Reijden, B.A., Jansen, J.H., EORTC Leukemia Group and GIMEMA, 2014. Characterization of acute myeloid leukemia based on levels of global hydroxymethylation. Blood 124, 1110–1118.

Larson, A.R., Dresser, K.A., Zhan, Q., Lezcano, C., Woda, B.A., Yosufi, B., Thompson, J.F., Scolyer, R.A., Mihm, M.C., Jr, Shi, Y.G., Murphy, G.F., Lian, C.G., 2014. Loss of 5-hydroxymethylcytosine correlates with increasing morphologic dysplasia in melanocytic tumors. Mod. Pathol. 27, 936–944.

Lee, D.U., Agarwal, S., Rao, A., 2002. Th2 lineage commitment and efficient IL-4 production involves extended demethylation of the IL-4 gene. Immunity 16, 649–660.

Levine, M.E., Lu, A.T., Quach, A., Chen, B.H., Assimes, T.L., Bandinelli, S., Hou, L., Baccarelli, A.A., Stewart, J.D., Li, Y., Whitsel, E.A., Wilson, J.G., Reiner, A.P., Aviv, A., Lohman, K., Liu, Y., Ferrucci, L., Horvath, S., 2018. An epigenetic biomarker of aging for lifespan and healthspan. Aging 10, 573–591.

Lian, C.G., Xu, Y., Ceol, C., Wu, F., Larson, A., Dresser, K., Xu, W., Tan, L., Hu, Y., Zhan, Q., Lee, C.-W., Hu, D., Lian, B.Q., Kleffel, S., Yang, Y., Neiswender, J., Khorasani, A.J., Fang, R., Lezcano, C., Duncan, L.M., Scolyer, R.A., Thompson, J.F., Kakavand, H., Houvras, Y., Zon, L.I., Mihm, M.C., Jr, Kaiser, U.B., Schatton, T., Woda, B.A., Murphy, G.F., Shi, Y.G., 2012. Loss of 5-hydroxymethylcytosine is an epigenetic hallmark of melanoma. Cell 150, 1135–1146.

Liao, Y., Gu, J., Wu, Y., Long, X., Ge, D.I., Xu, J., Ding, J., 2016. Low level of 5-Hydroxymethylcytosine predicts poor prognosis in non-small cell lung cancer. Oncol. Lett. 11, 3753–3760.

Ligthart, S., Marzi, C., Aslibekyan, S., Mendelson, M.M., Conneely, K.N., Tanaka, T., Colicino, E., Waite, L.L., Joehanes, R., Guan, W., Brody, J.A., Elks, C., Marioni, R., Jhun, M.A., Agha, G., Bressler, J., Ward-Caviness, C.K., Chen, B.H., Huan, T., Bakulski, K., Salfati, E.L., WHI-EMPC Investigators, Fiorito, G., CHARGE epigenetics of Coronary Heart Disease, Wahl, S., Schramm, K., Sha, J., Hernandez, D.G., Just, A.C., Smith, J.A., Sotoodehnia, N., Pilling, L.C., Pankow, J.S., Tsao, P.S., Liu, C., Zhao, W., Guarrera, S., Michopoulos, V.J., Smith, A.K., Peters, M.J., Melzer, D., Vokonas, P., Fornage, M., Prokisch, H., Bis, J.C., Chu,

A.Y., Herder, C., Grallert, H., Yao, C., Shah, S., McRae, A.F., Lin, H., Horvath, S., Fallin, D., Hofman, A., Wareham, N.J., Wiggins, K.L., Feinberg, A.P., Starr, J.M., Visscher, P.M., Murabito, J.M., Kardia, S.L.R., Absher, D.M., Binder, E.B., Singleton, A.B., Bandinelli, S., Peters, A., Waldenberger, M., Matullo, G., Schwartz, J.D., Demerath, E.W., Uitterlinden, A.G., van Meurs, J.B.J., Franco, O.H., Chen, Y.-D.I., Levy, D., Turner, S.T., Deary, I.J., Ressler, K.J., Dupuis, J., Ferrucci, L., Ong, K.K., Assimes, T.L., Boerwinkle, E., Koenig, W., Arnett, D.K., Baccarelli, A.A., Benjamin, E.J., Dehghan, A., 2016. DNA methylation signatures of chronic low-grade inflammation are associated with complex diseases. Genome Biol. 17, 255.

Li, H., Zheng, T., Chen, B., Hong, G., Zhang, W., Shi, T., Li, S., Ao, L., Wang, C., Guo, Z., 2014. Similar blood-borne DNA methylation alterations in cancer and inflammatory diseases determined by subpopulation shifts in peripheral leukocytes. Br. J. Cancer 111, 525–531.

Linehan, E., Fitzgerald, D.C., 2015. Ageing and the immune system: focus on macrophages. Eur. J. Microbiol. Immunol. 5, 14–24.

Liu, C., Liu, L., Chen, X., Shen, J., Shan, J., Xu, Y., Yang, Z., Wu, L., Xia, F., Bie, P., Cui, Y., Bian, X.-W., Qian, C., 2013. Decrease of 5-hydroxymethylcytosine is associated with progression of hepatocellular carcinoma through downregulation of TET1. PLoS One 8, e62828.

Liu, X.S., Wu, H., Ji, X., Stelzer, Y., Wu, X., Czauderna, S., Shu, J., Dadon, D., Young, R.A., Jaenisch, R., 2016. Editing DNA Methylation in the Mammalian Genome. Cell 167, 233–247.e17.

Lord, J.M., Butcher, S., Killampali, V., Lascelles, D., Salmon, M., 2001. Neutrophil ageing and immunesenescence. Mech. Ageing Dev. 122, 1521–1535.

Mahmood, S.S., Levy, D., Vasan, R.S., Wang, T.J., 2014. The Framingham Heart Study and the epidemiology of cardiovascular disease: a historical perspective. Lancet 383, 999–1008.

Marco, A., Kisliouk, T., Tabachnik, T., Weller, A., Meiri, N., 2016. DNA CpG Methylation (5-Methylcytosine) and Its Derivative (5-Hydroxymethylcytosine) Alter Histone Posttranslational Modifications at the Pomc Promoter, Affecting the Impact of Perinatal Diet on Leanness and Obesity of the Offspring. Diabetes 65, 2258–2267.

Marioni, R.E., Shah, S., McRae, A.F., Ritchie, S.J., Muniz-Terrera, G., Harris, S.E., Gibson, J., Redmond, P., Cox, S.R., Pattie, A., Corley, J., Taylor, A., Murphy, L., Starr, J.M., Horvath, S., Visscher, P.M., Wray, N.R., Deary, I.J., 2015. The epigenetic clock is correlated with physical and cognitive fitness in the Lothian Birth Cohort 1936. Int. J. Epidemiol. 44, 1388–1396.

Marttila, S., Kananen, L., Häyrynen, S., Jylhävä, J., Nevalainen, T., Hervonen, A., Jylhä, M., Nykter, M., Hurme, M., 2015. Ageing-associated changes in the human DNA methylome: genomic locations and effects on gene expression. BMC Genomics 16, 179.

Mathur, S.K., Schwantes, E.A., Jarjour, N.N., Busse, W.W., 2008. Age-related changes in eosinophil function in human subjects. Chest 133, 412–419.

Medzhitov, R., Janeway, C.A., Jr, 2002. Decoding the patterns of self and nonself by the innate immune system. Science 296, 298–300.

Moran-Crusio, K., Reavie, L., Shih, A., Abdel-Wahab, O., Ndiaye-Lobry, D., Lobry, C.,

Figueroa, M.E., Vasanthakumar, A., Patel, J., Zhao, X., Perna, F., Pandey, S., Madzo, J., Song, C., Dai, Q., He, C., Ibrahim, S., Beran, M., Zavadil, J., Nimer, S.D., Melnick, A., Godley, L.A., Aifantis, I., Levine, R.L., 2011. Tet2 loss leads to increased hematopoietic stem cell self-renewal and myeloid transformation. Cancer Cell 20, 11–24.

Müller, T., Gessi, M., Waha, A., Isselstein, L.J., Luxen, D., Freihoff, D., Freihoff, J., Becker, A., Simon, M., Hammes, J., Denkhaus, D., zur Mühlen, A., Pietsch, T., Waha, A., 2012. Nuclear exclusion of TET1 is associated with loss of 5-hydroxymethylcytosine in IDH1 wild-type gliomas. Am. J. Pathol. 181, 675–683.

Nakano, K., Boyle, D.L., Firestein, G.S., 2013. Regulation of DNA methylation in rheumatoid arthritis synoviocytes. J. Immunol. 190, 1297–1303.

Nan, X., Ng, H.H., Johnson, C.A., Laherty, C.D., Turner, B.M., Eisenman, R.N., Bird, A., 1998. Transcriptional repression by the methyl-CpG-binding protein MeCP2 involves a histone deacetylase complex. Nature 393, 386–389.

Needham, B.L., Smith, J.A., Zhao, W., Wang, X., Mukherjee, B., Kardia, S.L.R., Shively, C.A., Seeman, T.E., Liu, Y., Diez Roux, A.V., 2015. Life course socioeconomic status and DNA methylation in genes related to stress reactivity and inflammation: The multi-ethnic study of atherosclerosis. Epigenetics 10, 958–969.

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, 247–257.

Osborn, L., Hession, C., Tizard, R., Vassallo, C., Luhowskyj, S., Chi-Rosso, G., Lobb, R., 1989. Direct expression cloning of vascular cell adhesion molecule 1, a cytokine-induced endothelial protein that binds to lymphocytes. Cell 59, 1203–1211.

Palmer, D.B., 2013. The Effect of Age on Thymic Function. Front. Immunol. 4. https://doi.org/10.3389/fimmu.2013.00316

Pan, Z., Zhang, M., Ma, T., Xue, Z.-Y., Li, G.-F., Hao, L.-Y., Zhu, L.-J., Li, Y.-Q., Ding, H.-L., Cao, J.-L., 2016. Hydroxymethylation of microRNA-365-3p Regulates Nociceptive Behaviors via Kcnh2. J. Neurosci. 36, 2769–2781.

Pedersen, J.S., Valen, E., Velazquez, A.M.V., Parker, B.J., Rasmussen, M., Lindgreen, S., Lilje, B., Tobin, D.J., Kelly, T.K., Vang, S., Andersson, R., Jones, P.A., Hoover, C.A., Tikhonov, A., Prokhortchouk, E., Rubin, E.M., Sandelin, A., Gilbert, M.T.P., Krogh, A., Willerslev, E., Orlando, L., 2014. Genome-wide nucleosome map and cytosine methylation levels of an ancient human genome. Genome Res. 24, 454–466.

Qiu, W., Baccarelli, A., Carey, V.J., Boutaoui, N., Bacherman, H., Klanderman, B., Rennard, S., Agusti, A., Anderson, W., Lomas, D.A., DeMeo, D.L., 2012. Variable DNA methylation is associated with chronic obstructive pulmonary disease and lung function. Am. J. Respir. Crit. Care Med. 185, 373–381.

Reinius, L.E., Acevedo, N., Joerink, M., Pershagen, G., Dahlén, S.-E., Greco, D., Söderhäll, C., Scheynius, A., Kere, J., 2012. Differential DNA methylation in purified human blood cells: implications for cell lineage and studies on disease susceptibility. PLoS One 7, e41361.

Renshaw, M., Rockwell, J., Engleman, C., Gewirtz, A., Katz, J., Sambhara, S., 2002. Cutting edge: impaired Toll-like receptor expression and function in aging. J. Immunol. 169, 4697–4701.

Reynolds, L.M., Ding, J., Taylor, J.R., Lohman, K., Soranzo, N., de la Fuente, A., Liu, T.F., Johnson, C., Graham Barr, R., Register, T.C., Donohue, K.M., Talor, M.V., Cihakova, D., Gu, C., Divers, J., Siscovick, D., Burke, G., Post, W., Shea, S., Jacobs, D.R., Hoeschele, I., McCall, C.E., Kritchevsky, S.B., Herrington, D., Tracy, R.P., Liu, Y., 2015. Transcriptomic profiles of aging in purified human immune cells. BMC Genomics 16. https://doi.org/10.1186/s12864-015-1522-4

Reynolds, L.M., Taylor, J.R., Ding, J., Lohman, K., Johnson, C., Siscovick, D., Burke, G., Post, W., Shea, S., Jacobs, D.R., Jr., Stunnenberg, H., Kritchevsky, S.B., Hoeschele, I., McCall, C.E., Herrington, D.M., Tracy, R.P., Liu, Y., 2014. Age-related variations in the methylome associated with gene expression in human monocytes and T cells. Nat. Commun. 5, 5366.

Rönn, T., Volkov, P., Gillberg, L., Kokosar, M., Perfilyev, A., Jacobsen, A.L., Jørgensen, S.W., Brøns, C., Jansson, P.-A., Eriksson, K.-F., Pedersen, O., Hansen, T., Groop, L., Stener-Victorin, E., Vaag, A., Nilsson, E., Ling, C., 2015. Impact of age, BMI and HbA1c levels on the genome-wide DNA methylation and mRNA expression patterns in human adipose tissue and identification of epigenetic biomarkers in blood. Hum. Mol. Genet. 24, 3792–3813.

Rosenkranz, D., Weyer, S., Tolosa, E., Gaenslen, A., Berg, D., Leyhe, T., Gasser, T., Stoltze, L., 2007. Higher frequency of regulatory T cells in the elderly and increased suppressive activity in neurodegeneration. J. Neuroimmunol. 188, 117–127.

Sarmento, O.F., Svingen, P.A., Xiong, Y., Xavier, R.J., McGovern, D., Smyrk, T.C., Papadakis, K.A., Urrutia, R.A., Faubion, W.A., 2015. A Novel Role for Kruppel-like Factor 14 (KLF14) in T-Regulatory Cell Differentiation. Cellular and Molecular Gastroenterology and Hepatology 1, 188–202.e4.

Schuyler, R.P., Merkel, A., Raineri, E., Altucci, L., Vellenga, E., Martens, J.H.A., Pourfarzad, F., Kuijpers, T.W., Burden, F., Farrow, S., Downes, K., Ouwehand, W.H., Clarke, L., Datta, A., Lowy, E., Flicek, P., Frontini, M., Stunnenberg, H.G., Martín-Subero, J.I., Gut, I., Heath, S., 2016. Distinct Trends of DNA Methylation Patterning in the Innate and Adaptive Immune Systems. Cell Rep. 17, 2101–2111.

Seidler, S., Zimmermann, H.W., Bartneck, M., Trautwein, C., Tacke, F., 2010. Age-dependent alterations of monocyte subsets and monocyte-related chemokine pathways in healthy adults. BMC Immunol. 11, 30.

Shu, L., Sun, W., Li, L., Xu, Z., Lin, L., Xie, P., Shen, H., Huang, L., Xu, Q., Jin, P., Li, X., 2016. Genome-wide alteration of 5-hydroxymenthylcytosine in a mouse model of Alzheimer's disease. BMC Genomics 17. https://doi.org/10.1186/s12864-016-2731-1

Slieker, R.C., Relton, C.L., Gaunt, T.R., Slagboom, P.E., Heijmans, B.T., 2018. Age-related DNA methylation changes are tissue-specific with ELOVL2 promoter methylation as exception. Epigenetics Chromatin 11, 25.

Smallwood, S.A., Lee, H.J., Angermueller, C., Krueger, F., Saadeh, H., Peat, J., Andrews, S.R., Stegle, O., Reik, W., Kelsey, G., 2014. Single-cell genome-wide bisulfite sequencing for assessing epigenetic heterogeneity. Nat. Methods 11, 817–820.

Smith, A.K., Conneely, K.N., Kilaru, V., Mercer, K.B., Weiss, T.E., Bradley, B., Tang, Y., Gillespie, C.F., Cubells, J.F., Ressler, K.J., 2011. Differential immune system DNA methylation and cytokine regulation in post-traumatic stress disorder. Am. J. Med.

Genet. B Neuropsychiatr. Genet. 156B, 700–708.
Smith, A.K., Conneely, K.N., Pace, T.W.W., Mister, D., Felger, J.C., Kilaru, V., Akel, M.J., Vertino, P.M., Miller, A.H., Torres, M.A., 2014. Epigenetic changes associated with inflammation in breast cancer patients treated with chemotherapy. Brain Behav. Immun. 38, 227–236.
Smith, J.A., Zhao, W., Wang, X., Ratliff, S.M., Mukherjee, B., Kardia, S.L.R., Liu, Y., Roux, A.V.D., Needham, B.L., 2017. Neighborhood characteristics influence DNA methylation of genes involved in stress response and inflammation: The Multi-Ethnic Study of Atherosclerosis. Epigenetics 12, 662–673.
Song, C.-X., Yi, C., He, C., 2012. Mapping recently identified nucleotide variants in the genome and transcriptome. Nat. Biotechnol. 30, 1107–1116.
Song, C.-X., Yin, S., Ma, L., Wheeler, A., Chen, Y., Zhang, Y., Liu, B., Xiong, J., Zhang, W., Hu, J., Zhou, Z., Dong, B., Tian, Z., Jeffrey, S.S., Chua, M.-S., So, S., Li, W., Wei, Y., Diao, J., Xie, D., Quake, S.R., 2017. 5-Hydroxymethylcytosine signatures in cell-free DNA provide information about tumor types and stages. Cell Res. 27, 1231–1242.
Steegenga, W.T., Boekschoten, M.V., Lute, C., Hooiveld, G.J., de Groot, P.J., Morris, T.J., Teschendorff, A.E., Butcher, L.M., Beck, S., Müller, M., 2014. Genome-wide age-related changes in DNA methylation and gene expression in human PBMCs. Age 36, 9648.
Stringhini, S., Polidoro, S., Sacerdote, C., Kelly, R.S., van Veldhoven, K., Agnoli, C., Grioni, S., Tumino, R., Giurdanella, M.C., Panico, S., Mattiello, A., Palli, D., Masala, G., Gallo, V., Castagné, R., Paccaud, F., Campanella, G., Chadeau-Hyam, M., Vineis, P., 2015. Life-course socioeconomic status and DNA methylation of genes regulating inflammation. Int. J. Epidemiol. 44, 1320–1330.
Svensson, A., Patzi Churqui, M., Schlüter, K., Lind, L., Eriksson, K., 2017. Maturation-dependent expression of AIM2 in human B-cells. PLoS One 12, e0183268.
Szulwach, K.E., Li, X., Li, Y., Song, C.-X., Wu, H., Dai, Q., Irier, H., Upadhyay, A.K., Gearing, M., Levey, A.I., Vasanthakumar, A., Godley, L.A., Chang, Q., Cheng, X., He, C., Jin, P., 2011. 5-hmC-mediated epigenetic dynamics during postnatal neurodevelopment and aging. Nat. Neurosci. 14, 1607–1616.
Szwagierczak, A., Bultmann, S., Schmidt, C.S., Spada, F., Leonhardt, H., 2010. Sensitive enzymatic quantification of 5-hydroxymethylcytosine in genomic DNA. Nucleic Acids Res. 38, e181.
Tahiliani, M., Koh, K.P., Shen, Y., Pastor, W.A., Bandukwala, H., Brudno, Y., Agarwal, S., Iyer, L.M., Liu, D.R., Aravind, L., Rao, A., 2009. Conversion of 5-methylcytosine to 5-hydroxymethylcytosine in mammalian DNA by MLL partner TET1. Science 324, 930–935.
Tan, S., 2012. The leucocyte β2 (CD18) integrins: the structure, functional regulation and signalling properties. Biosci. Rep. 32, 241–269.
Terragni, J., Bitinaite, J., Zheng, Y., Pradhan, S., 2012. Biochemical characterization of recombinant β-glucosyltransferase and analysis of global 5-hydroxymethylcytosine in unique genomes. Biochemistry 51, 1009–1019.
Teschendorff, A.E., Menon, U., Gentry-Maharaj, A., Ramus, S.J., Weisenberger, D.J., Shen, H., Campan, M., Noushmehr, H., Bell, C.G., Maxwell, A.P., Savage, D.A., Mueller-Holzner, E., Marth, C., Kocjan, G., Gayther, S.A., Jones, A., Beck, S.,

Wagner, W., Laird, P.W., Jacobs, I.J., Widschwendter, M., 2010. Age-dependent DNA methylation of genes that are suppressed in stem cells is a hallmark of cancer. Genome Res. 20, 440–446.

Thomson, J.P., Meehan, R.R., 2017. The application of genome-wide 5-hydroxymethylcytosine studies in cancer research. Epigenomics 9, 77–91.

Toraño, E.G., Bayón, G.F., Del Real, Á., Sierra, M.I., García, M.G., Carella, A., Belmonte, T., Urdinguio, R.G., Cubillo, I., García-Castro, J., Delgado-Calle, J., Pérez-Campo, F.M., Riancho, J.A., Fraga, M.F., Fernández, A.F., 2016. Age-associated hydroxymethylation in human bone-marrow mesenchymal stem cells. J. Transl. Med. 14, 207.

Tsenkina, Y., Ruzov, A., Gliddon, C., Horsburgh, K., De Sousa, P.A., 2014. White matter tract and glial-associated changes in 5-hydroxymethylcytosine following chronic cerebral hypoperfusion. Brain Res. 1592, 82–100.

Tserel, L., Kolde, R., Limbach, M., Tretyakov, K., Kasela, S., Kisand, K., Saare, M., Vilo, J., Metspalu, A., Milani, L., Peterson, P., 2015. Age-related profiling of DNA methylation in CD8+ T cells reveals changes in immune response and transcriptional regulator genes. Sci. Rep. 5, 13107.

Tu, W., Rao, S., 2016. Mechanisms Underlying T Cell Immunosenescence: Aging and Cytomegalovirus Infection. Front. Microbiol. 7, 2111.

Uciechowski, P., Kahmann, L., Plümäkers, B., Malavolta, M., Mocchegiani, E., Dedoussis, G., Herbein, G., Jajte, J., Fulop, T., Rink, L., 2008. TH1 and TH2 cell polarization increases with aging and is modulated by zinc supplementation. Exp. Gerontol. 43, 493–498.

van den Hoogen, L.L., Sims, G.P., van Roon, J.A.G., Fritsch-Stork, R.D.E., 2015. Aging and Systemic Lupus Erythematosus - Immunosenescence and Beyond. Curr. Aging Sci. 8, 158–177.

Ventham, N.T., Kennedy, N.A., Adams, A.T., Kalla, R., Heath, S., O'Leary, K.R., Drummond, H., IBD BIOM consortium, IBD CHARACTER consortium, Wilson, D.C., Gut, I.G., Nimmo, E.R., Satsangi, J., 2016. Integrative epigenome-wide analysis demonstrates that DNA methylation may mediate genetic risk in inflammatory bowel disease. Nat. Commun. 7, 13507.

Wade, P.A., 2001. Methyl CpG-binding proteins and transcriptional repression. Bioessays 23, 1131–1137.

Wajant, H., Pfizenmaier, K., Scheurich, P., 2003. Tumor necrosis factor signaling. Cell Death Differ. 10, 45–65.

Wang, C., Pan, Y., Zhang, Q.-Y., Wang, F.-M., Kong, L.-D., 2012. Quercetin and allopurinol ameliorate kidney injury in STZ-treated rats with regulation of renal NLRP3 inflammasome activation and lipid accumulation. PLoS One 7, e38285.

Weisenberger, D.J., 2014. Characterizing DNA methylation alterations from The Cancer Genome Atlas. J. Clin. Invest. 124, 17–23.

Weng, N.-P., Akbar, A.N., Goronzy, J., 2009. CD28− T cells: their role in the age-associated decline of immune function. Trends Immunol. 30, 306–312.

Wen, L., Li, X., Yan, L., Tan, Y., Li, R., Zhao, Y., Wang, Y., Xie, J., Zhang, Y., Song, C., Yu, M., Liu, X., Zhu, P., Li, X., Hou, Y., Guo, H., Wu, X., He, C., Li, R., Tang, F., Qiao, J., 2014. Whole-genome analysis of 5-hydroxymethylcytosine and 5-methylcytosine at base resolution in the human brain. Genome Biol. 15, R49.

Wong, C., Goldstein, D.R., 2013. Impact of aging on antigen presentation cell function of dendritic cells. Curr. Opin. Immunol. 25, 535–541.

Wong, C.P., Rinaldi, N.A., Ho, E., 2015. Zinc deficiency enhanced inflammatory response by increasing immune cell activation and inducing IL6 promoter demethylation. Mol. Nutr. Food Res. 59, 991–999.

Wu, H., Zhao, M., Tan, L., Lu, Q., 2016. The key culprit in the pathogenesis of systemic lupus erythematosus: Aberrant DNA methylation. Autoimmun. Rev. 15, 684–689.

Wu, S.C., Zhang, Y., 2010. Active DNA demethylation: many roads lead to Rome. Nat. Rev. Mol. Cell Biol. 11, 607–620.

Xiao, F.-H., Kong, Q.-P., Perry, B., He, Y.-H., 2016. Progress on the role of DNA methylation in aging and longevity. Brief. Funct. Genomics 15, 454–459.

Yang, Q., Wu, K., Ji, M., Jin, W., He, N., Shi, B., Hou, P., 2013. Decreased 5-hydroxymethylcytosine (5-hmC) is an independent poor prognostic factor in gastric cancer patients. J. Biomed. Nanotechnol. 9, 1607–1616.

Ye, C., Li, L., 2014. 5-hydroxymethylcytosine: a new insight into epigenetics in cancer. Cancer Biol. Ther. 15, 10–15.

Ye, M., Graf, T., 2007. Early decisions in lymphoid development. Curr. Opin. Immunol. 19, 123–128.

Yoder, M.C., 2004. Blood cell progenitors: insights into the properties of stem cells. Anat. Rec. A Discov. Mol. Cell. Evol. Biol. 276, 66–74.

Yusuf, N., Hidalgo, B., Irvin, M.R., Sha, J., Zhi, D., Tiwari, H.K., Absher, D., Arnett, D.K., Aslibekyan, S.W., 2017. An epigenome-wide association study of inflammatory response to fenofibrate in the Genetics of Lipid Lowering Drugs and Diet Network. Pharmacogenomics 18, 1333–1341.

Zhang, L.-Y., Li, P.-L., Wang, T.-Z., Zhang, X.-C., 2015. Prognostic values of 5-hmC, 5-mC and TET2 in epithelial ovarian cancer. Arch. Gynecol. Obstet. 292, 891–897.

Zhang, N., Bevan, M.J., 2011. CD8(+) T cells: foot soldiers of the immune system. Immunity 35, 161–168.

Zhao, J., Zhu, Y., Yang, J., Li, L., Wu, H., De Jager, P.L., Jin, P., Bennett, D.A., 2017. A genome-wide profiling of brain DNA hydroxymethylation in Alzheimer's disease. Alzheimers. Dement. 13, 674–688.

Zhao, M., Li, M.-Y., Gao, X.-F., Jia, S.-J., Gao, K.-Q., Zhou, Y., Zhang, H.-H., Huang, Y., Wang, J., Wu, H.-J., Lu, Q.-J., 2017. Downregulation of BDH2 modulates iron homeostasis and promotes DNA demethylation in CD4+ T cells of systemic lupus erythematosus. Clin. Immunol. https://doi.org/10.1016/j.clim.2017.11.002

Zhao, Q., Fan, Y.-C., Zhao, J., Gao, S., Zhao, Z.-H., Wang, K., 2013. DNA methylation patterns of peroxisome proliferator-activated receptor gamma gene associated with liver fibrosis and inflammation in chronic hepatitis B. J. Viral Hepat. 20, 430–437.

Zhu, J., Paul, W.E., 2008. CD4 T cells: fates, functions, and faults. Blood 112, 1557–1569.

Chronic disease	Tissue studied	5mC/ 5hmC	Sites/Genes associated	Study Details	Publication
Alzheimer's disease (AD)	Brain tissue	5mC	71 DMLs associated with neuritic amyloid plaques in secondary analysis	491 AD subjects, 217 non-impaired, Bonferroni-corrected (α=0.05)	De Jager et al. (2014)
	Hippocampus	5mC/5hmC	Global decrease in 5mC and 5hmC	10 cases, 10 age-matched controls (α=0.05)	Chouliaras et al. (2012)
	Brain tissue	5hmC	325 genes	5 cases, 5 controls	Bernstein et al. (2016)
Asthma	Multiple tissues	5mC	27 DMLs discovered, 14 DMLs replicated	n=1548 in discovery cohort, n=3196 in replication cohorts Bonferroni-corrected in replication (α=0.05)	Xu et al. (2018)
Asthma (experimentally induced cell line)	CD4+ T cells	5mC	*IFNG*	4 to 6 animals per group	Brand et al. (2012)
Atherosclerosis	Aortic and carotid samples	5mC	1858 DMLs	15 cases, 15 controls, 19 carotid plaques Bonferroni-corrected (α=0.05)	Zaina et al. (2014)
Chronic fatigue syndrome	PBMCs	5mC	1192 DMLs in 826 genes	12 cases, 12 controls FDR<0.05	de Vega, Vernon, and McGowan (2014)
Chronic obstructive pulmonary disease (COPD)	White blood cells	5mC	349 DMLs	2 cohorts: 1. 620 cases, 325 controls 2. 181 cases 109 controls FDR<0.05	Qiu et al. (2012)
Coronary artery disease	Peripheral Lymphocytes	5mC	Increase in global 5mC	137 cases, 150 controls (α=0.05)	Sharma et al. (2008)

Hepatitis B	PBMCs	5mC	*PPAR-γ*	91 cases, 18 controls (α=0.05)	Q. Zhao et al. (2013)
Hypertension	Whole blood	5mC	Decrease in global 5mC	30 stage 1, 30 stage 2, 30 controls (α=0.05)	Smolarek et al. (2010)
Hidradenitis suppurativa	Skin	5hmC	Decrease in global 5hmC	30 cases, 30 controls ($p < 0.0001$)	Hessam et al. (2017)
Irritable Bowel Syndrome (IBD)	Whole blood	5mC	439 DMLs	240 cases, 190 controls Holm-corrected (α=0.05)	Ventham et al. (2016)
Cutaneous Mastocytosis	Peripheral blood (PB) Bone marrow (BM)	5mC/ 5hmC	Decrease in 5mC and 5hmC in BM and PB	24 cases, 31 controls (PB 5mC) 24 cases, 30 controls (PB 5hmC) 18 cases, 8 controls (BM 5mC and 5hmC)	Leoni et al. (2015)
Osteoarthritis (review)	Chondrocytes	5mC	*COL9, GDF5, DIO2, IL1β, MMP13, iNOS, SOX9, ADAMTS4, MMP13, MMP3, MMP9*	Review of 8 studies	Hollander and Meulenbelt (2015)
Osteoarthritis	Chondrocytes	5hmC	70591 DhMRs (44288 DhMRs increased, 26303 decreased)	4 cases; 4 controls FDR<0.2 applied	Taylor et al. (2015)
Parkinson's disease	pars compacta of substantia nigra, putamen, cortex	5mC	*SNCA intron 1*	12 cases, 14 controls (α=0.05)	Jowaed et al. (2010)
Post traumatic stress disorder	PBMCs	5mC	Global increase in 5mC, DMLs in TPR, CLEC9A, ACP5, ANXA2, TLR8	cases 25, controls 25 FDR<0.05	A. K. Smith et al. (2011)

Rheumatoid arthritis	T cells, B cells, NK cells, Monocytes, polymorphonuclear leukocytes	5mC	Global decrease in 5mC in T cells and monocytes	19 cases, 17 controls (α=0.05)	de Andres et al. (2015)
Schizophrenia	Frontal cortex	5mC	4641 probes in 2929 genes	24 cases, 24 controls FDR<0.05	Wockner et al. (2014)
Sporadic Amyotrophic lateral sclerosis	Spinal cord	5mC/5hmC	Global increase in 5mC and 5hmC; 4,261 DMLs in 3,574 genes	12 cases, 11 controls FDR<0.05	Figueroa-Romero et al. (2012)
Systemic lupus erythematosus	White blood cells	5mC	Differential DNAm in 49 genes	17 cases, 17 controls FDR < 0.25	Javierre et al. (2010)
	CD4+ T cells	5mC	86 DMLs	18 cases, 18 controls FDR<0.05	Coit et al. (2013)
	CD4+ T cells	5hmC	Global increase in 5hmC	10 cases, 10 controls (α=0.05)	M. Zhao et al. (2016)
Type II diabetes	Peripheral white blood cells	5mC	13 DMLs in 6 DMRs	710 cases, 459 controls FDR< 0.05	Toperoff et al. (2012)

Supplementary Table 2.1. Non-cancerous chronic diseases associated with 5mC and/or 5hmC. Abbreviations: DMLs: differentially methylated CpG sites; DhMRs: differentially hydroxymethylated regions

Chapter 3: Age-related DNA hydroxymethylation is enriched for gene expression and immune system processes in human peripheral blood.

This chapter was published as cited below:

Nicholas D. Johnson, Luoxiu Huang, Ronghua Li, Yun Li, Yuchen Yang, Hye Rim Kim, Crystal Grant, Hao Wu, Eric A Whitsel, Douglas P Kiel, Andrea A Baccarelli, Peng Jin, Joanne M Murabito, Karen N Conneely, 2019. “Age-Related DNA Hydroxymethylation Is Enriched for Gene Expression and Immune System Processes in Human Peripheral Blood.” Epigenetics, September. https://doi.org/10.1080/15592294.2019.1666651.

INTRODUCTION

DNA methylation (DNAm) is the process whereby enzymes known as DNA methyltransferases covalently bind methyl groups (-CH_3) to DNA. DNAm usually occurs at a CpG site, i.e. a 5’ cytosine nucleotide directly upstream of a 3’ guanine nucleotide (Jaenisch and Bird 2003). An epigenetic modification closely related to DNAm is DNA hydroxymethylation (DNAhm), distinguished by the hydroxylation of the methyl to a hydroxymethyl group (CH_2OH). The ten-eleven translocation (TET) enzymes are a family of three proteins (TET1, TET2, TET3) that catalyze the hydroxylation of 5-methylcytosine to 5-hydroxymethylcytosine (Ito et al. 2010).

DNAm undergoes changes in an orchestrated fashion throughout mammalian development, and coordinates with proteins to regulate gene expression states (Jaenisch

and Bird 2003). A large body of work has established that DNAm robustly associates with age in human whole blood (Bell et al. 2012; Bjornsson et al. 2008; Christensen et al. 2009; Teschendorff et al. 2010; Alisch et al. 2012), whereas age-related changes to DNAhm are less well characterized. This is partially because DNAhm was only discovered in mammals in 2009 (Tahiliani et al. 2009). Previously, bisulfite-based methods to detect DNAm were unable to differentiate between DNAm and DNAhm, but newer methods are able to detect and quantify DNAhm (Song, Yi, and He 2012; Terragni et al. 2012; Höbartner 2011; Szwagierczak et al. 2010; Booth et al. 2012). Initially, DNAhm was regarded as a transient epigenetic modification intermediate between DNAm and demethylation. However, recent work suggests it can be stably maintained and may serve as a blocker to proteins that would otherwise bind to methylated DNA (Hahn, Szabó, and Pfeifer 2014), which opens the possibility that DNAhm could play a more important role in gene regulation than its original understanding as a transient mark might suggest.

DNAhm has been most extensively studied in the brain, where 13% of all CpG sites have been reported to have high DNAhm levels (Wen et al. 2014). Studies have also found that DNAhm is enriched in active genes of human and mouse frontal cortex (Wen et al. 2014; Lister et al. 2013; Colquitt et al. 2013), and associates with active transcription and age throughout brain development (Szulwach et al. 2011; Hahn et al. 2013; Chouliaras et al. 2012; Song et al. 2011; Gross et al. 2015). In neural tissue of mice from embryonic day 7 to 6 weeks, Szulwach et al. (2011) observed an association between developmental stage and increased DNAhm in developmentally active genes (Szulwach et al. 2011). An association between gene activation and gene body DNAhm has also been observed in neural progenitor cells and neurons isolated from mice (Hahn

et al. 2013). These studies suggest that DNAhm may mediate the effect of gene activation on neurodevelopment.

Age-related gene body DNAhm in the brain is enriched for pathways associated with neurodegenerative disorders (Song et al. 2011). A genome-wide decrease in DNAhm was observed in a fragile-X mouse model (Yao et al. 2014). In post-mortem human prefrontal cortex, Bernstein et al. (2016) observed differential gene body DNAhm in Alzheimer's disease in 325 genes in both discovery and replication datasets (Bernstein et al. 2016). DNAhm has also been shown to associate with Alzheimer's-associated neuritic plaques and neurofibrillary tangles in postmortem human dorsolateral prefrontal cortex tissue (Zhao et al. 2017). Together, these studies suggest DNAhm may be involved in neurodegeneration.

Most previous studies have examined brain DNAhm, while few have investigated DNAhm in blood. Further, it is possible that whole blood DNAhm has escaped interest due to its low content, which is estimated to be 0.027% (Godderis et al. 2015). A longitudinal study of peripheral blood samples from mice observed 8,613 genomic regions differentially hydroxymethylated with age, and 28,196 CpG sites differentially methylated with age, with little overlap between the two (Kochmanski et al. 2018). Although regional changes with age in DNAhm have yet to be investigated in human blood, studies have observed global decreases in DNAhm with age (Truong et al. 2015; Xiong et al. 2015).

While previous work has established a potential role for DNAhm in active gene transcription as well as development and aging in brain tissue, the role it plays in blood remains unclear. The objective of our study was to investigate age-related DNAhm in human peripheral blood and its possible gene regulatory and biological roles. To do this,

we generated genome-wide DNAhm capture sequencing data and RNA-seq data in peripheral blood mononuclear cells (PBMCs) from 10 young and 10 old healthy Caucasian females. Our study is the first to investigate these changes regionally, and one of the first to investigate the interplay between genome-wide gene expression and DNAhm in peripheral blood.

RESULTS

To investigate associations between age, DNAhm, and gene expression, we performed DNAhm capture sequencing (Song, Yi, and He 2012) and RNA-sequencing on DNA samples and RNA samples, respectively, isolated from PBMCs of 10 young (23-30 years) and 10 old Caucasian females (68-76 years). The number of reads that uniquely mapped to the genome were 11,668,072 to 36,532,555 for DNAhm-capture sequencing and 13,354,493 to 36,314,069 for RNA-seq across samples. Among the samples from the 20 individuals, 48-81% of the DNAhm-capture reads (Supplementary Table 3.1) and 74-80% of the RNA-seq reads (Supplementary Table 3.2) uniquely mapped to the genome. DNAhm profiles clustered by both batch and age group in a principal components analysis (Supplementary Figure 3.1), indicating that both of these variables capture variation in the data. To account for the variation due to batch, batch was included as a covariate in all subsequent analyses of the DNAhm data. The sample gene expression profiles clustered into the young and old age groups (Supplementary Figure 3.2). These analyses show that both the gene expression profiles and DNAhm profiles of our sample data capture variation originating from sample age.

We first investigated DNAhm content across different genomic features. The mean DNAhm read density was highest in enhancers ($p<10^{-15}$) (Figure 3.1), followed by exons, which had a higher density than that observed in gene bodies ($p<10^{-15}$) or genome-wide ($p<10^{-15}$). 52% of promoters had no overlapping reads, although for reads overlapping promoter regions, mean density was comparable between exons and promoters.

We observed 6,650 differentially hydroxymethylated regions (DhMRs) (FDR<0.05): 4,664 associated with increased ((+) DhMRs) and 1,986 associated with decreased ((-) DhMRs) DNAhm with age (Figure 3.2; Supplementary Table 3.3). Log_2-fold-change ranged from 0.35 to 2.48 for (+) DhMRs and -0.37 to -2.14 for (-) DhMRs (Figure 3.3). The counts for the 20 samples for these 6,650 bins clustered by age group, with one exception (Figure 3.4). We observed 732 regions with multiple DhMRs directly adjacent to one another: 658 associated with lower DNAhm and 74 with higher DNAhm in the old age group. Most DhMRs overlapped a gene body, enhancer, or promoter with an overrepresentation of both (+) and (-) DhMRs overlapping gene bodies and enhancers (Table 3.1). The most significant DhMR had decreasing DNAhm with age and occurred on chromosome 20 in an intronic region of *TOX2* (Supplementary Figure 3.3), which belongs to a family of transcription factors that modify chromatin structure in T cell development (O'Flaherty and Kaye 2003), and was recently linked to $CD8^+$ T cell exhaustion (Seo et al. 2019). The second most significant DhMR was on chromosome 1 and also had decreasing DNAhm with age, and overlapped an exon and intronic regions of the gene *LRRN2,* which is overexpressed in malignant gliomas (Supplementary Figure 3.4) (Almeida et al. 1998). The most significant DhMR with increasing DNAhm was on chromosome 7 overlapping an intron of *STK17A* (Supplementary Figure 3.5), also known

as *DRAK1*, which is involved in apoptosis and is overexpressed in some cancers (Park et al. 2015).

To investigate whether our DhMRs mapped to regions showing differential DNAhm with age in brain, we compared our results to two studies of post-mortem cerebellar tissue samples. To maximize comparability, we re-analyzed raw data from the previous studies using similar methods of genome tiling, quality control, and testing for DhMRs between the studies. We observed a small but significant positive correlation between test statistics ($0.12 \leq r \leq 0.37$; Supplementary Figure 3.6) in a comparison of our DhMRs to age-DhMRs called from 19 publicly available cerebellar tissue DNAhm samples (age range: 4-55 years) (Cheng et al. 2018). However, we observed no correlation between the test statistics corresponding to our DhMRs and DhMRs called from DNAhm profiles of 6 young (<20 years) and 6 old (>69 years) cerebellar tissue samples from the NIH NeuroBioBank tissues repositories ($r<0.02$; Supplementary Figure 3.7) (Kim and Jin, n.d.) To investigate whether our DhMRs mapped to regions showing differential DNAm with age, we compared our DhMRs to age-related differential DNAm identified in human whole blood samples from two cohorts, one with similar demographics to our 20-sample cohort (European-American) and one African-American cohort (Johnson et al. 2017; Ali et al. 2015). Although differentially methylated regions (DMRs) were highly correlated between the two cohorts ($r=0.95$; Supplementary Figure 3.8), we did not observe a correlation

between our DhMR test statistics and DNAm test statistics for either cohort ($r<0.1$, Supplementary Figure 3.8).

To investigate the potential gene regulatory role of age-related DNAhm, we performed an age-related differential gene expression analysis. We observed 124 differentially expressed genes (DEGs) (FDR<0.05), 45 with higher ((+) DEGs) and 79 with lower ((-) DEGs) expression in old vs. young samples (Supplementary Table 3.5). For these 124 DEGs, the RNA-seq read counts clustered perfectly by age group (Supplementary Figure 3.9). We compared our DESeq2 test statistics to the published Z-scores of 1,497 genes significantly associated with age in a previous study of gene expression in whole blood (Peters et al. 2015) and observed a significant correlation between the two ($r=0.58$; $p<10^{-15}$; Figure 3.10), indicating that our results were consistent with those from a much larger study (N=14,983). We next investigated the overlap between DEGs and DhMRs in our study. 160 DhMRs overlapped 39 of the 124 DEGs. We observed directional consistency between age-related DNAhm and gene expression at 159 of these 160 DhMRs (Supplementary Table 3.5). 154 of these DhMRs exhibited age-related decreases in both DNAhm and gene expression.

Because DNAm marks cell lineage skewing (H. Li et al. 2014) and age-related cell lineage skewing is observed in both peripheral blood (Tu and Rao 2016) and hematopoietic stem cells (HSCs) (Geiger, de Haan, and Carolina Florian 2013), we investigated the possibility that age-related changes in cell type proportions could explain our age-related results. To do this, we estimated cell type proportions (Figure 3.11) from our RNA-seq data using the DeconRNASeq package (Gong and Szustakowski 2013). The DESeq2 test statistics generated with estimates of cell type proportions as covariates

strongly correlated with the test statistics obtained without cell type proportions included as covariates for both DEGs (Figure 3.12; r=0.64) and DhMRs (Figure 3.13; r=0.83). All DhMRs and 123 of the 124 DEGs from the original models showed directional consistency in the models with cell types as covariates, suggesting that our age-related results are robust to cell composition.

We also evaluated whether gene body, exonic, promoter, or enhancer DNAhm associated with gene expression of the same gene controlling for age and batch. We observed 16 genes with a significant positive association between gene expression and exonic DNAhm and none with a negative association (FDR<0.05; Supplementary Table 3.6). 25 genes had a significant association between gene expression and gene body DNAhm, 23 of which had a positive association (FDR<0.05; Supplementary Table 3.7). 3 of these 25 genes overlapped with the 16 genes with a significant association between exonic DNAhm and gene expression. The lack of overlapping results between exonic regions and gene bodies could be because exonic regions do not make up a large portion of the total gene length (~16%). Thus, 5hmC could feasibly overlap the gene body without overlapping an exon. We also observed 1 significant association between gene expression and enhancer DNAhm for an enhancer linked to the gene *TBX21* ($p=5.3x10^{-8}$, Supplementary Figure 3.14). However, this enhancer overlaps a substantial portion of the gene body (2274 bp) and we also observed a significant association between gene expression and 5hmC in both gene body and exonic regions for *TBX21;* thus, it is not clear which to attribute the association to. We did not observe a significant association between gene expression and promoter DNAhm for any gene. Among both significant and non-significant results, the majority of genes had a positive slope coefficient of gene

expression on DNAhm in gene bodies (62%), exonic regions (58%), promoters (56%), and enhancers (55%). Further, the average T-statistic corresponding to the slope coefficient of gene expression on DNAhm was significantly greater than zero for all four genomic features ($p<10^{-15}$; Supplementary Figure 3.15). The positive association between DNAhm and gene expression for most significant genes is consistent with previous work observing enrichment of DNAhm in active genes in the brain (Song et al. 2011; Szulwach et al. 2011; Mellén et al. 2012; Lister et al. 2013). To test whether age moderates the association between DNAhm and gene expression, we included an interaction term between these two variables in two additional analyses, one for gene body DNAhm and the other for exonic DNAhm. We only tested for an interaction among genes that were significant for the main effect analysis. We did not observe a significant interaction for exonic or enhancer regions. For gene body 5hmC, a tumor suppressor gene known as PRDM5 (Shu et al. 2011) had a significant interaction between 5hmC and age ($p=4.9x10^{-5}$; Supplementary Figure 3.17) suggesting a large effect of 5hmC on gene expression among older individuals that was little to non-existent among young individuals.

We performed GO analyses on 9 gene sets: the 1500 genes with the highest DNAhm content, genes showing age-related increased or decreased DNAhm in gene bodies, promoters, and enhancers, and (+) and (-) DEGs. After Bonferroni adjustment ($p<3.8x10^{-6}$) we observed enrichment for a total of 94 GO terms (Supplementary Table 3.8). We observed 7 GO terms enriched in positive gene body DhMRs as well as 1 enriched in negative enhancer DhMRs and 19 enriched in positive enhancer DhMRs (Table 3.2). Genes with high DNAhm were enriched for 66 GO terms and positive DEGs

were enriched for 1 GO term. Many of the significant GO terms observed are involved in immune system function. Genes with high DNAhm were enriched for cell activation involved in immune response, immune system development, immune effector process, immune response-activating cell surface receptor signaling pathway, immune response-regulating signaling pathway, phagocytosis, Fc-receptor signaling pathway, and Fc-gamma receptor signaling pathway involved in phagocytosis. Positive gene body DhMRs associated with leukocyte mediated immunity, neutrophil activation, neutrophil degranulation, and myeloid cell activation involved in immune response. The set of genes with positive enhancer DhMRs were enriched for neutrophil degranulation, immune response, immune system development, and positive regulation of leukocyte differentiation.

DISCUSSION

Our study is the first to discover region-specific changes in DNAhm with age in human blood. Among overlapping DhmRs and DEGs, the vast majority had the same direction of effect, i.e. both with associated increases with age or both with associated decreases with age. Further, we observed a positive correlation between gene expression and DNAhm, supporting a possible gene regulatory role for DNAhm in PBMCs. We observed no correlation between DMRs and DhMRs, which could suggest that DNAm and DNAhm regulate distinct genes. Based on these results, it is possible that DNAm and DNAhm work in tandem to shape the development and function of tissues throughout life.

Unlike CpG island promoter DNAm, which typically holds genes in a stably repressed state (Illingworth and Bird 2009; Jones 2012), the relationship between

promoter DNAhm and gene transcription may depend on the presence of certain transcription factors. Unlike DNAm, DNAhm does not bind to methyl-binding proteins (MBD1, MBD2, and MBD4), which are known repressors of transcription (Johnson and Conneely 2019; Jin, Kadam, and Pfeifer 2010). However, a transcriptional repressor known as methyl CpG binding protein-2 (MeCP2) binds to DNAm and DNAhm with similar affinity (Mellén et al. 2012). Further, work suggests that promoter DNAhm may mediate the binding of MeCP2 thereby blocking transcription (Zhubi et al. 2014). Among both significant and non-significant results, we observed 56% of genes to have a positive correlation between promoter DNAhm and gene expression. It is possible that the preponderance of positive associations between DNAhm and gene expression is a result of a lack of transcriptional repressors that bind to DNAhm.

The link between DNAhm and active transcription has been better established for gene bodies than promoter regions. Consistent with previous work (Szulwach et al. 2011; Hahn et al. 2013; Chouliaras et al. 2012; Song et al. 2011; Mellén et al. 2012; Gross et al. 2015), we observed mostly positive associations between DNAhm and gene expression: 16 of the 16 genes were positively associated among genes with an association between exonic DNAhm and gene expression, and 23 of 25 genes were positively associated among genes with an association between gene body DNAhm and gene expression. Further, for a slight but significant majority (62%) of all genes tested (including those that were not significant), the test statistic corresponding to the correlation between gene expression and DNAhm was positive. In contrast, previous work on human blood samples observed 60-70% of CpG sites with a significant association

between gene expression and DNAm to be negatively correlated in both exons and promoters, as well as other genomic contexts (Kennedy et al. 2018).

Among these 23 genes with an association between gene body DNAhm and gene expression, the gene *PRDM5* had a significant interaction between DNAhm and age (Supplementary Figure 3.16) indicative of a much stronger association among old subjects compared to young subjects. This could suggest that DNAhm regulates gene expression in an age-dependent fashion for *PRDM5*, whereby its effect on gene expression is strong among the elderly and negligible in younger individuals.

Previous work has observed the highest levels of DNAhm in exons and enhancers in prefrontal cortex (Gross et al. 2015). Consistent with this study, we observed the highest density of read counts overlapping enhancers followed by exons. The density in both enhancers and exons was significantly higher than gene bodies and genome-wide, and nominally significantly higher than promoters. It is unclear what role DNAhm might play in enhancers, but its similarly high abundance in both brain and blood may be indicative of a congruent role of this epigenetic modification across tissues.

In the brain, where DNAhm has been more extensively studied relative to other tissues, studies have observed enrichment of gene body DNAhm in highly transcribed genes (Song et al. 2011; Szulwach et al. 2011; Mellén et al. 2012; Lister et al. 2013). In our study, we observed directional consistency between age-related differences in peripheral blood DNAhm and gene expression. Further, in a comparison of our age-related DhMRs and previously published age-related differentially methylated loci, we observed no correlation (Supplementary Figure 3.6). These findings together with

previous work support the interpretation that DNAhm regulates distinct genes and may be involved in the upregulation of gene expression.

Gene ontology analyses support the interpretation that DNAhm is involved in immune system function, similar to DNAm (Johnson and Conneely 2019). In our study, genes showing increased DNAhm in gene bodies in older subjects were enriched for biological processes related to myeloid cell activation involved in immune response. One of the hallmarks of the aging immune system is skewing toward the myeloid lineage of HSCs (Geiger, de Haan, and Carolina Florian 2013). Experiments in mice have shown that both DNAm and DNAhm alter lineage skewing in HSCs (Challen et al. 2011; Ko et al. 2010; Moran-Crusio et al. 2011). Cell types of the myeloid lineage undergo major changes during immunosenescence such as an age-related impairment of phagocytosis among monocytes (Hearps et al. 2012), which are the most abundant mononuclear phagocytes found in blood (Jakubzick, Randolph, and Henson 2017). Our findings could indicate that age-related DNAhm is involved in these changes. Another myeloid cell type that incurs age-related changes is neutrophils, including an age-related impairment of degranulation (Lord et al. 2001). We observed genes showing increased DNAhm in older subjects to be enriched for biological processes related to neutrophil function including neutrophil activation and degranulation. Interestingly, neutrophils are not mononuclear and therefore should not be present in PBMCs. Although contamination of PBMCs with low-density granulocytes has been observed, this is more typically observed in diseases such as systemic lupus erythematosus whereas such contamination is usually negligible in healthy individuals (Zhang et al. 2017), such as the subjects in our study. The enrichment with neutrophil function could be an indication of cross-talk between cell

types, that is, age-related DNAhm could be regulating genes related to neutrophil function in PBMCs which is subsequently communicated to neutrophils.

In addition to age-related DNAhm, we observed enrichment for immune system function among genes with the highest levels of DNAhm, which suggests a possible role of DNAhm in the regulation of immune-related genes independent of age. One of the enriched biological processes was the Fc-gamma receptor signaling pathway involved in phagocytosis, which becomes activated in both macrophages (Huber et al. 2001) and neutrophils (Weisbart et al. 1988; Lord et al. 2001) to combat pathogens. We also observed enrichment for biological processes suggesting DNAhm is involved in immune response, including immune effector process, immune response-activating cell surface receptor signaling pathway, cell activation involved in immune response, and immune response-regulating signaling pathway.

One of the limitations of this study is the small sample size, which limits power and is likely why we only detected 124 DEGs, much fewer than a previous study that observed 1,497 DEGs in PBMCs of 14,983 individuals of European ancestry (Peters et al. 2015). In addition, we used a capture-based approach (hMe-Seal), which quantifies DNAhm regionally rather than at CpG site resolution. Thus, our age-related DNAhm results could potentially differ from those generated at CpG site resolution. Another limitation is the differences in sequencing lengths for batch 1 (151 bp) and batch 2 (51 bp), which could explain the lower alignment rate of batch 2 compared to batch 1 (Supplementary Table 3.1). Because batches were randomized, and batch was included as a covariate in all analyses, this difference will not bias or confound our analysis, but the lower coverage in batch 2 represents a further limit on power. Further, all 20 subjects were Caucasian and

female. While the homogeneity of our sample means it is unlikely that sex or ancestry could confound our results, it is possible that our results may not generalize to males or people of non-European ancestry. However, this was a pilot study intended to investigate 1) the possible presence of region-specific age-related DNAhm in PBMCs and 2) whether DNAhm was enriched for active transcription. We observed evidence of both of these phenomena. Further, our GO analyses suggest that age-related DNAhm may be involved in immune system function in PBMCs. Our findings motivate future investigations into age-related DNAhm in human PBMCs and its possible involvement in gene regulation and immunosenescence, ideally in cohorts that are larger and more heterogeneous.

CONCLUSION

Our study is the first to show that DNAhm associates with age at thousands of regions in human PBMCs. The directional consistency between age-related differences in DNAhm and gene expression supports previous work in brain reporting that DNAhm is enriched for active transcription. Our study adds to previous work suggesting a possible role of DNAm (Challen et al. 2011) and DNAhm (Ko et al. 2010; Moran-Crusio et al. 2011) in immunosenescence within HSCs, and supports the premise that age-related DNAhm plays a role in immune system function in PBMCs.

METHODS

Data Collection

We collected DNAhm and gene expression data from PBMCs of 10 young (23-30 years) and 10 old (68-76 years) Caucasian females, all of whom were disease-free,

unmedicated non-smokers. Ascertainment of participants and sample collection was performed by iSpecimen®, and samples were subsequently shipped to Emory University for analysis. We extracted DNA using the QIAGEN DNeasy Blood and Tissue Kit.

We extracted 1 µg of RNA from each subject using the QIAGEN RNeasy Kit, synthesized cDNA with Invitrogen Oligo(dT)20 primers, and amplified DNA with q-PCR analysis. We performed all laboratory analyses blinded from knowledge of age group of the samples collected.

DNAhm Data Generation

We performed DNAhm-enrichment using hMe-Seal, a previously described selective chemical labeling method (Song et al. 2011). To perform DNAhm labeling, we used a 100-µl solution containing 50 mM HEPES buffer (pH 7.9), 25 mM MgCl2, 300 ng µl− 1 sonicated genomic DNA (100–500 bp), 250 µM UDP-6-N3-Glu, and 2.25 µM wild-type β-glucosyltransferase. We incubated the reactions at 37 °C for 1 hour, after which we purified DNA substrates using either a Qiagen DNA purification kit or a phenol-chloroform precipitation followed by reconstitution in H_2O. With the addition of 150 µM dibenzocyclooctyne modified biotin in the DNA solution, we performed click chemistry followed by incubation of the solution for 2 hours at 37 °C. Using Pierce Monomeric Avidin Kit (Thermo), we purified samples according to the manufacturer's recommendations. Subsequent to elution, we concentrated DNA containing biotin-5-N3-gmC using 10 K

Amicon Ultra 0.5-mL Centrifugal Filters (Millipore) and then purified using a Qiagen DNA purification kit.

DNAhm-enriched DNA library preparation

We generated libraries following the Illumina protocol “Preparing Samples for ChIP Sequencing of DNA” (Part# 111257047 Rev. A). We initiated the protocol using 25 ng of DNAhm-captured DNA or input genomic DNA. We gel-purified fragments of approximately 150-300 bp subsequent to adapter ligation. We quantified PCR-amplified DNA libraries using an Agilent 2100 Bioanalyzer. We diluted these libraries to 6-8 pM for cluster generation and sequencing. We used version 4 Cluster Generation (Part #15002739), Sequencing Kits (#15005236) and Version 7.0 recipes to perform 38-cycle single-end sequencing. Due to a change in lab protocol between batches 1 and 2, the sequence read length was 151 bp for batch 1 and 51 bp for batch 2. We processed images and extracted sequences using the standard Illumina Pipeline.

DNAhm Data Processing

We used Bowtie 2 (Langmead and Salzberg 2012) to align samples to the hg38 build of the human genome using the pre-built index from Illumina’s iGenomes collection. Bowtie 2 outputs SAM-formatted files, which we then sorted, converted to BAM format, and removed reads with MAPQ values lower than 23 (Supplementary Figure 3.17; Supplementary Table 3.9) using samtools (Heng Li et al. 2009).

Creation of DNAhm Count Matrices

We used the Rsamtools (Morgan M, Pages H, Obenchain V, Hayden N 2018), GenomicRanges (Lawrence et al. 2013), and BSgenome.Hsapiens.UCSC.hg38 libraries in R Bioconductor to create a 1,544,146-row by 20-column matrix, each row corresponding to a 2000 bp non-overlapping region of the genome and each column corresponding to a sample. The value in the i^{th} row of the j^{th} column indicates the number of DNAhm reads overlapping the i^{th} 2000 bp segment of the j^{th} sample. To create count matrices for reads overlapping gene bodies and exonic regions, we used the same R Bioconductor packages and an Ensembl gene annotation file of 60,675 genes based on the hg38 build of the human genome. To determine DNAhm reads overlapping promoters, we used the NIH Roadmap Epigenomics Consortium's 18,692 PBMC-specific promoters, which were determined using ChromHMM (Roadmap Epigenomics Consortium et al. 2015; Ernst and Kellis 2012). In our analysis, we used 13,782 of the 18,692 promoters, which were within 2000 bp upstream of the transcription start site of the 60,675 Ensembl genes. We also created a count matrix of DNAhm reads overlapping enhancers using the GeneHancer annotation, which contains annotations of enhancers and their linked genes (Fishilevich et al. 2017).

RNA-seq library preparation

We generated RNA-seq libraries using 0.5 ug of total RNA samples from young or old PBMCs via the TruSeq RNA Sample Preparation Kit v2 (Illumina). We validated libraries by DNA Chips using Agilent 2100 Bioanalyzer. We then sequenced these libraries using 50-cycle single-end runs with Illumina HiSeq 2000.

RNA-seq Data Processing

We aligned RNA-seq reads to the hg38 build of the human genome and discarded reads shorter than 50 bps and with quality scores below 20 using the STAR (Dobin et al. 2013) RNA-seq aligner.

We used the Rsamtools, GenomicFeatures, and GenomicAlignments R Bioconductor packages to count RNA-seq reads overlapping each Ensembl-annotated gene of the hg38 build of the human genome. Using the same Ensembl-annotated genome, we counted RNA-seq reads that overlapped exons of these genes.

Quality Control of Count Matrices

For quality control of the RNA-seq and DNAhm read count matrices, we excluded genes/regions that had <10 reads for 5 or more samples. For the DNAhm and RNA-Seq Joint Data Analysis described below, we used genes that passed quality control for both the RNA-seq data using STAR and the DNAhm count matrix of reads overlapping gene exons.

To evaluate data clustering of the gene expression profiles using FPKM values and DNAhm profiles of the 2000 bp binned count matrix, we performed principal component analyses. We then plotted the second principal component against the first principal component for both the gene expression profiles and the DNAhm profiles.

Age-differential analyses of DNAhm and gene expression

To test for age-differentially expressed genes (DEGs) and age-differentially hydroxymethylated 2000 bp regions (DhMRs), we used the R Bioconductor package

DESeq2 which models read counts as a negative binomial distribution (Love, Huber, and Anders 2014). Read counts were included as the outcome variable and age group as the predictor variable. When testing for DhMRs we also included batch as a covariate. To correct for multiple testing, we used the Bioconductor qvalue package to apply an FDR cutoff of 0.05 (Dabney, Storey, and Warnes 2010).

Comparison of age-related DNAhm and DNAm

To compare our age-related DNAhm to age-related DNAm, we used CpG-specific results from a previous study investigating age-related DNAhm (N=336) (Johnson et al. 2017). We used Stouffer's method to combine site-specific T-statistics from the regression of DNAm on age into region-specific Z-scores for each DhMR.

DNAhm and RNA-Seq Joint Data Analysis

We used a regression model to test for an association between gene expression (in fragments per kilobase of mapped reads, or FPKM values) and quantile-normalized log-transformed DNAhm counts overlapping gene exons, gene bodies and gene promoters using the following model:

$$FPKM_{ij} = \beta_0 + \beta_{DNAhm} DNAhm_{ij} + \beta_{Age\ group} Age\ group_j + \beta_{Batch} Batch_j + \varepsilon_{ij}$$

where the response variable, $FPKM_{ij}$, corresponds to the FPKM value and the predictor variable, $DNAhm_{ij}$, corresponds to the normalized transformed DNAhm count for the i^{th} gene and the j^{th} individual. We also included two covariates: $Age\ group_j$ of the j^{th} individual and $Batch_j$, the batch run for the DNAhm reads of the j^{th} individual. The error term for the i^{th} gene and the j^{th} individual is denoted ε_{ij}. To test for an interaction between DNAhm and

age group, we performed additional regressions for gene bodies, gene exons, and enhancers by including the interaction term $DNAhm_{ij} \times Age\ group_j$ in the model. We only tested for an interaction among genes that had a significant main effect term β_{DNAhm}.

Cell Mixture Estimation

We used previously published cell-type-specific RNA-seq signatures from flow-sorted PBMCs (Monaco et al. 2019) and the R Bioconductor package DeconRNASeq (Gong and Szustakowski 2013) to estimate proportions of myeloid cells, T cells, and B cells in our samples of PBMCs. To investigate whether our models were robust to possible age-related lineage skewing, we compared the DESeq2 analyses run with and without cell type proportions included as covariates for both DEGs and DhMRs.

Gene Enrichment Analyses

We used the software package GOstats available in R Bioconductor to test for an association between gene sets and Gene Ontology (GO) terms (Falcon and Gentleman 2007). We first used DESeq2 to perform a feature-based age-differential analysis of DNAhm reads overlapping gene bodies, promoters, and enhancers, applying an FDR<0.05 on the combined set of genomic features. We then did a separate GO analysis on the following nine gene sets: DEGs with decreased and increased expression with age, genes with decreased DNAhm in gene bodies, promoters, enhancers and genes with increased DNAhm in gene bodies, promoters, and enhancers as well as the top 1,500 genes with the most overlapping DNAhm reads. For all GO analyses, we used a Bonferroni correction of α=0.05.

		Decreasing DhmRs	Increasing DhmRs	All 2000 bp bins
Overlap with:	Promoter	46 (0.02%)	5 (0.01%)	19481 (0.01%)
	Enhancer	852 (39%)	129 (34%)	321938 (21%)
	Gene body	1446 (65%)	265 (71%)	825172 (53%)
	Total	2205	374	1544146

Table 3.1. Total number (and percentages) of decreasing DhmRs, increasing DhmRs, and 2000 bp bins overlapping at least one promoter, enhancer, or gene body.

	Negative	Positive
Differential Gene Expression	1. atrioventricular valve development 2. mitral valve morphogenesis 3. kidney morphogenesis	1. oxygen transport 2. bicarbonate transport 3. hydrogen peroxide catabolic process 4. response to hydrogen peroxide 5. antibiotic metabolic process 6. peptidyl-glutamic acid carboxylation 7. response to oxidative stress
Differential DNAhm (gene body)	None	1. leukocyte mediated immunity 2. neutrophil activation 3. neutrophil degranulation 4. myeloid cell activation involved in immune response 5. cell activation 6. exocytosis 7. positive regulation of nitric-oxide synthase biosynthetic process
Differential DNAhm (promoter)	None	None
Differential DNAhm (enhancer)	1. positive regulation of RNA biosynthetic process 2. positive regulation of nucleobase-containing compound metabolic process	1. positive regulation of catalytic activity 2. neutrophil degranulation 3. cellular response to nitrogen compound 4. secretion 5. protein modification process 6. cell activation involved in immune response 7. regulation of protein localization 8. amide transport 9. organic substance transport 10. positive regulation of peptide secretion

Table 3.2. Significantly enriched GO terms by direction of differential DNAhm and genomic feature showing differential DNAhm.

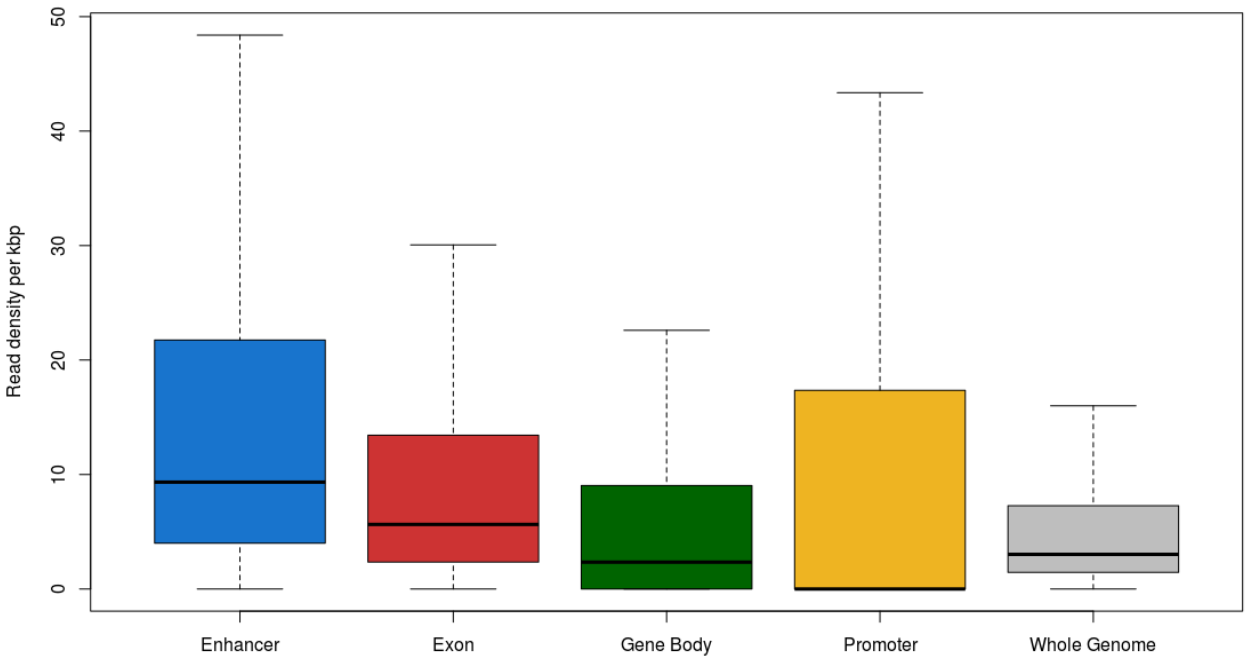


Figure 3.1. Overlapping reads per kbp across genomic features. To calculate read density, reads that overlapped a genomic feature were counted and averaged across individuals.

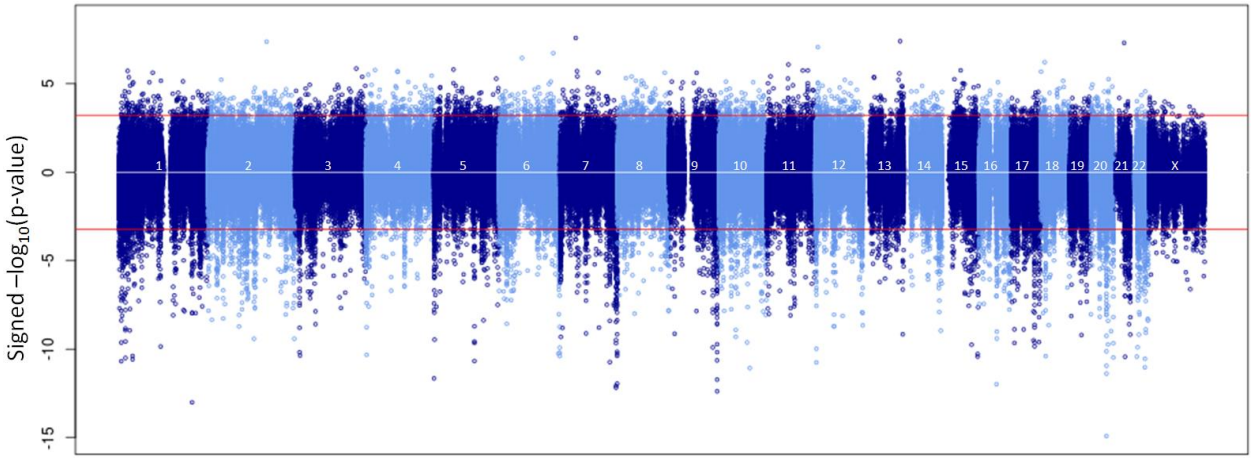


Figure 3.2. Manhattan plot of DhMR analysis. The y-axis represents the signed negative $\log_{10}$-p-value and the x-axis represents the chromosome number and position, with chromosome X appearing to the very right.

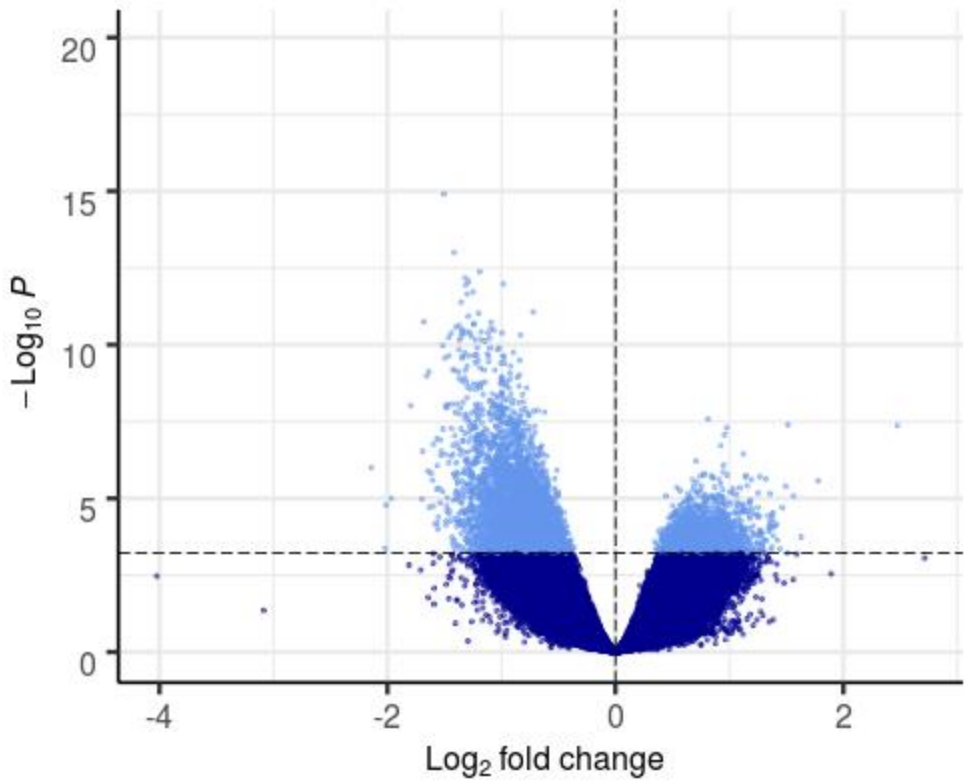


Figure 3.3. Volcano plot of DhMRs in 2000 bp bins. Bins in light blue are those with FDR<0.05.

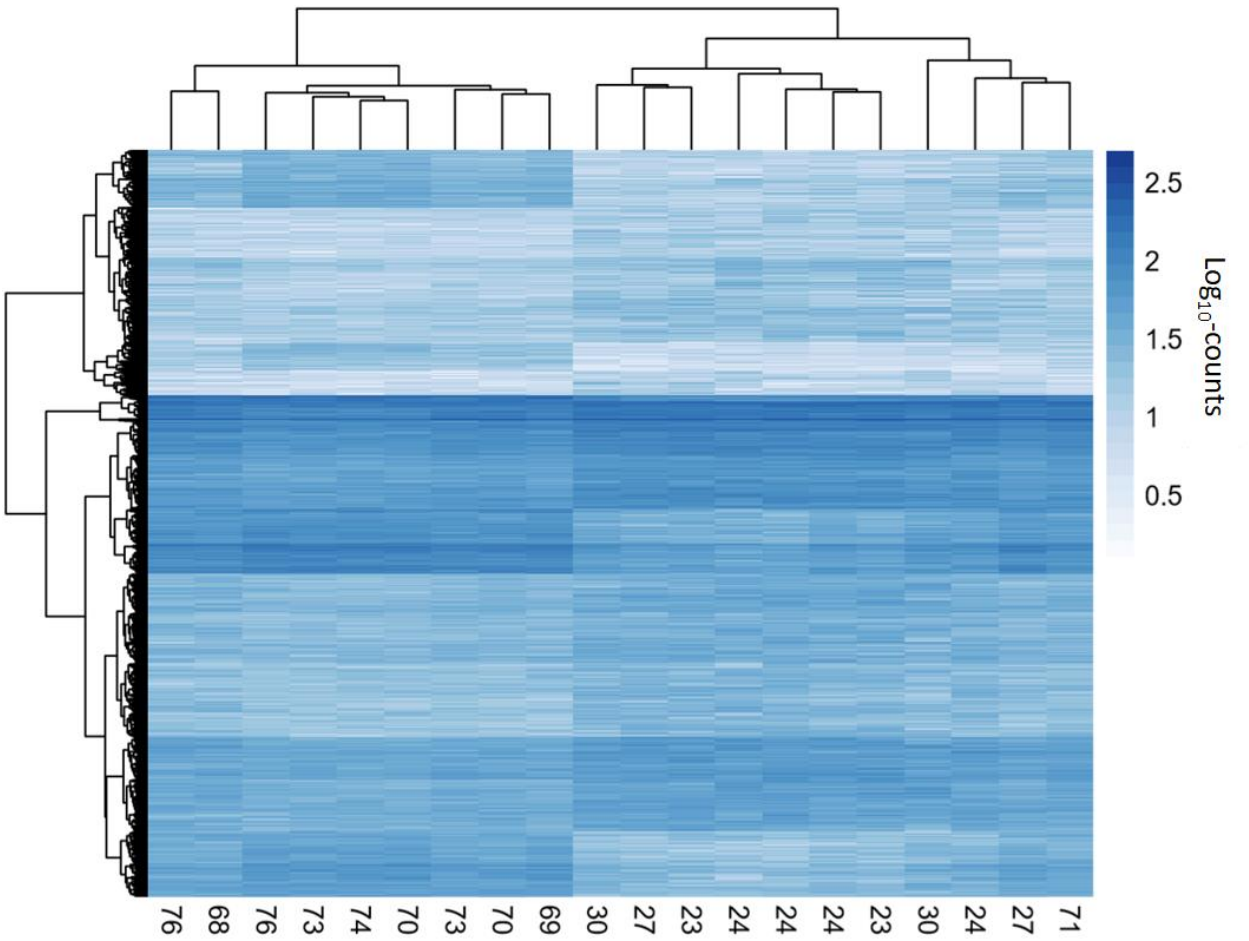


Figure 3.4. Heatmap of log_{10}-normalized counts of the 6,650 DhMRs. The numbers along the x-axis indicate ages.

REFERENCES

Ali, Omar, Diana Cerjak, Jack W. Kent Jr, Roland James, John Blangero, Melanie A. Carless, and Yi Zhang. 2015. “An Epigenetic Map of Age-Associated Autosomal Loci in Northern European Families at High Risk for the Metabolic Syndrome.” *Clinical Epigenetics* 7 (February): 12.

Alisch, Reid S., Benjamin G. Barwick, Pankaj Chopra, Leila K. Myrick, Glen A. Satten, Karen N. Conneely, and Stephen T. Warren. 2012. “Age-Associated DNA Methylation in Pediatric Populations.” *Genome Research* 22 (4): 623–32.

Almeida, A., X. X. Zhu, N. Vogt, R. Tyagi, M. Muleris, A. M. Dutrillaux, B. Dutrillaux, D. Ross, B. Malfoy, and S. Hanash. 1998. “GAC1, a New Member of the Leucine-Rich Repeat Superfamily on Chromosome Band 1q32.1, Is Amplified and Overexpressed in Malignant Gliomas.” *Oncogene* 16 (23): 2997–3002.

Bell, Jordana T., Pei-Chien Tsai, Tsun-Po Yang, Ruth Pidsley, James Nisbet, Daniel Glass, Massimo Mangino, et al. 2012. “Epigenome-Wide Scans Identify Differentially Methylated Regions for Age and Age-Related Phenotypes in a Healthy Ageing Population.” *PLoS Genetics* 8 (4): e1002629.

Bernstein, Alison I., Yunting Lin, R. Craig Street, Li Lin, Qing Dai, Li Yu, Han Bao, et al. 2016. “5-Hydroxymethylation-Associated Epigenetic Modifiers of Alzheimer’s Disease Modulate Tau-Induced Neurotoxicity.” *Human Molecular Genetics* 25 (12): 2437–50.

Bjornsson, Hans T., Martin I. Sigurdsson, M. Daniele Fallin, Rafael A. Irizarry, Thor Aspelund, Hengmi Cui, Wenqiang Yu, et al. 2008. “Intra-Individual Change over Time in DNA Methylation with Familial Clustering.” *JAMA: The Journal of the American Medical Association* 299 (24): 2877–83.

Booth, Michael J., Miguel R. Branco, Gabriella Ficz, David Oxley, Felix Krueger, Wolf Reik, and Shankar Balasubramanian. 2012. “Quantitative Sequencing of 5-Methylcytosine and 5-Hydroxymethylcytosine at Single-Base Resolution.” *Science* 336 (6083): 934–37.

Challen, Grant A., Deqiang Sun, Mira Jeong, Min Luo, Jaroslav Jelinek, Jonathan S. Berg, Christoph Bock, et al. 2011. “Dnmt3a Is Essential for Hematopoietic Stem Cell Differentiation.” *Nature Genetics* 44 (1): 23–31.

Cheng, Ying, Ziyi Li, Sasicha Manupipatpong, Li Lin, Xuekun Li, Tianlei Xu, Yong-Hui Jiang, Qiang Shu, Hao Wu, and Peng Jin. 2018. “5-Hydroxymethylcytosine Alterations in the Human Postmortem Brains of Autism Spectrum Disorder.” *Human Molecular Genetics* 27 (17): 2955–64.

Chouliaras, Leonidas, Daniel L. A. van den Hove, Gunter Kenis, Stella Keitel, Patrick R. Hof, Jim van Os, Harry W. M. Steinbusch, Christoph Schmitz, and Bart P. F. Rutten. 2012. “Age-Related Increase in Levels of 5-Hydroxymethylcytosine in Mouse Hippocampus Is Prevented by Caloric Restriction.” *Current Alzheimer Research* 9 (5): 536–44.

Christensen, Brock C., E. Andres Houseman, Carmen J. Marsit, Shichun Zheng, Margaret R. Wrensch, Joseph L. Wiemels, Heather H. Nelson, et al. 2009. “Aging and Environmental Exposures Alter Tissue-Specific DNA Methylation Dependent upon CpG Island Context.” *PLoS Genetics* 5 (8): e1000602.

Colquitt, Bradley M., William E. Allen, Gilad Barnea, and Stavros Lomvardas. 2013. “Alteration of Genic 5-Hydroxymethylcytosine Patterning in Olfactory Neurons

Correlates with Changes in Gene Expression and Cell Identity." *Proceedings of the National Academy of Sciences of the United States of America* 110 (36): 14682–87.

Dabney, Alan, John D. Storey, and G. R. Warnes. 2010. "Qvalue: Q-Value Estimation for False Discovery Rate Control." *R Package Version* 1 (0). ftp://ftp.uni-bayreuth.de/pub/math/statlib/R/CRAN/src/contrib/Descriptions/qvalue.html.

Dobin, Alexander, Carrie A. Davis, Felix Schlesinger, Jorg Drenkow, Chris Zaleski, Sonali Jha, Philippe Batut, Mark Chaisson, and Thomas R. Gingeras. 2013. "STAR: Ultrafast Universal RNA-Seq Aligner." *Bioinformatics* 29 (1): 15–21.

Ernst, Jason, and Manolis Kellis. 2012. "ChromHMM: Automating Chromatin-State Discovery and Characterization." *Nature Methods* 9 (3): 215–16.

Falcon, S., and R. Gentleman. 2007. "Using GOstats to Test Gene Lists for GO Term Association." *Bioinformatics* 23 (2): 257–58.

Fishilevich, Simon, Ron Nudel, Noa Rappaport, Rotem Hadar, Inbar Plaschkes, Tsippi Iny Stein, Naomi Rosen, et al. 2017. "GeneHancer: Genome-Wide Integration of Enhancers and Target Genes in GeneCards." *Database: The Journal of Biological Databases and Curation* 2017 (January). https://doi.org/10.1093/database/bax028.

Geiger, Hartmut, Gerald de Haan, and M. Carolina Florian. 2013. "The Ageing Haematopoietic Stem Cell Compartment." *Nature Reviews. Immunology* 13 (5): 376–89.

Godderis, Lode, Caroline Schouteden, Ali Tabish, Katrien Poels, Peter Hoet, Andrea A. Baccarelli, and Kirsten Van Landuyt. 2015. "Global Methylation and Hydroxymethylation in DNA from Blood and Saliva in Healthy Volunteers." *BioMed Research International* 2015 (May): 845041.

Gong, Ting, and Joseph D. Szustakowski. 2013. "DeconRNASeq: A Statistical Framework for Deconvolution of Heterogeneous Tissue Samples Based on mRNA-Seq Data." *Bioinformatics* 29 (8): 1083–85.

Gross, Jeffrey A., Alain Pacis, Gary G. Chen, Luis B. Barreiro, Carl Ernst, and Gustavo Turecki. 2015. "Characterizing 5-Hydroxymethylcytosine in Human Prefrontal Cortex at Single Base Resolution." *BMC Genomics* 16 (September): 672.

Hahn, Maria A., Runxiang Qiu, Xiwei Wu, Arthur X. Li, Heying Zhang, Jun Wang, Jonathan Jui, et al. 2013. "Dynamics of 5-Hydroxymethylcytosine and Chromatin Marks in Mammalian Neurogenesis." *Cell Reports* 3 (2): 291–300.

Hahn, Maria A., Piroska E. Szabó, and Gerd P. Pfeifer. 2014. "5-Hydroxymethylcytosine: A Stable or Transient DNA Modification?" *Genomics*. https://doi.org/10.1016/j.ygeno.2014.08.015.

Hearps, Anna C., Genevieve E. Martin, Thomas A. Angelovich, Wan-Jung Cheng, Anna Maisa, Alan L. Landay, Anthony Jaworowski, and Suzanne M. Crowe. 2012. "Aging Is Associated with Chronic Innate Immune Activation and Dysregulation of Monocyte Phenotype and Function." *Aging Cell* 11 (5): 867–75.

Höbartner, Claudia. 2011. "Enzymatic Labeling of 5-Hydroxymethylcytosine in DNA." *Angewandte Chemie* 50 (19): 4268–70.

Huber, V. C., J. M. Lynch, D. J. Bucher, J. Le, and D. W. Metzger. 2001. "Fc Receptor-Mediated Phagocytosis Makes a Significant Contribution to Clearance of Influenza Virus Infections." *Journal of Immunology* 166 (12): 7381–88.

Illingworth, Robert S., and Adrian P. Bird. 2009. “CpG Islands - ‘A Rough Guide.’” *FEBS Letters* 583 (11): 1713–20.

Ito, Shinsuke, Ana C. D’Alessio, Olena V. Taranova, Kwonho Hong, Lawrence C. Sowers, and Yi Zhang. 2010. “Role of Tet Proteins in 5mC to 5hmC Conversion, ES-Cell Self-Renewal and Inner Cell Mass Specification.” *Nature* 466 (7310): 1129–33.

Jaenisch, Rudolf, and Adrian Bird. 2003. “Epigenetic Regulation of Gene Expression: How the Genome Integrates Intrinsic and Environmental Signals.” *Nature Genetics* 33 Suppl (March): 245–54.

Jakubzick, Claudia V., Gwendalyn J. Randolph, and Peter M. Henson. 2017. “Monocyte Differentiation and Antigen-Presenting Functions.” *Nature Reviews. Immunology* 17 (6): 349–62.

Jin, Seung-Gi, Swati Kadam, and Gerd P. Pfeifer. 2010. “Examination of the Specificity of DNA Methylation Profiling Techniques towards 5-Methylcytosine and 5-Hydroxymethylcytosine.” *Nucleic Acids Research* 38 (11): e125.

Johnson, Nicholas D., and Karen N. Conneely. 2019. “The Role of DNA Methylation and Hydroxymethylation in Immunosenescence.” *Ageing Research Reviews* 51 (February): 11–23.

Johnson, Nicholas D., Howard W. Wiener, Alicia K. Smith, Shota Nishitani, Devin M. Absher, Donna K. Arnett, Stella Aslibekyan, and Karen N. Conneely. 2017. “Non-Linear Patterns in Age-Related DNA Methylation May Reflect CD4+ T Cell Differentiation.” *Epigenetics: Official Journal of the DNA Methylation Society* 12 (6): 492–503.

Jones, Peter A. 2012. “Functions of DNA Methylation: Islands, Start Sites, Gene Bodies and beyond.” *Nature Reviews. Genetics* 13 (7): 484–92.

Kennedy, Elizabeth M., George N. Goehring, Michael H. Nichols, Chloe Robins, Divya Mehta, Torsten Klengel, Eleazar Eskin, Alicia K. Smith, and Karen N. Conneely. 2018. “An Integrated -Omics Analysis of the Epigenetic Landscape of Gene Expression in Human Blood Cells.” *BMC Genomics* 19 (1): 476.

Kim, Hyerim, and Peng Jin. n.d. “Aging-Related Epigenetic Dynamics in Cerebellum.” *Unpublished Manuscript*.

Kochmanski, Joseph, Elizabeth H. Marchlewicz, Raymond G. Cavalcante, Maureen A. Sartor, and Dana C. Dolinoy. 2018. “Age-Related Epigenome-Wide DNA Methylation and Hydroxymethylation in Longitudinal Mouse Blood.” *Epigenetics: Official Journal of the DNA Methylation Society* 13 (7): 779–92.

Ko, Myunggon, Yun Huang, Anna M. Jankowska, Utz J. Pape, Mamta Tahiliani, Hozefa S. Bandukwala, Jungeun An, et al. 2010. “Impaired Hydroxylation of 5-Methylcytosine in Myeloid Cancers with Mutant TET2.” *Nature* 468 (7325): 839–43.

Langmead, Ben, and Steven L. Salzberg. 2012. “Fast Gapped-Read Alignment with Bowtie 2.” *Nature Methods* 9 (4): 357–59.

Lawrence, Michael, Wolfgang Huber, Hervé Pagès, Patrick Aboyoun, Marc Carlson, Robert Gentleman, Martin T. Morgan, and Vincent J. Carey. 2013. “Software for Computing and Annotating Genomic Ranges.” *PLoS Computational Biology* 9 (8): e1003118.

Li, Heng, Bob Handsaker, Alec Wysoker, Tim Fennell, Jue Ruan, Nils Homer, Gabor

Marth, Goncalo Abecasis, Richard Durbin, and 1000 Genome Project Data Processing Subgroup. 2009. “The Sequence Alignment/Map Format and SAMtools.” *Bioinformatics* 25 (16): 2078–79.

Li, H., T. Zheng, B. Chen, G. Hong, W. Zhang, T. Shi, S. Li, L. Ao, C. Wang, and Z. Guo. 2014. “Similar Blood-Borne DNA Methylation Alterations in Cancer and Inflammatory Diseases Determined by Subpopulation Shifts in Peripheral Leukocytes.” *British Journal of Cancer* 111 (3): 525–31.

Lister, Ryan, Eran A. Mukamel, Joseph R. Nery, Mark Urich, Clare A. Puddifoot, Nicholas D. Johnson, Jacinta Lucero, et al. 2013. “Global Epigenomic Reconfiguration during Mammalian Brain Development.” *Science* 341 (6146): 1237905.

Lord, J. M., S. Butcher, V. Killampali, D. Lascelles, and M. Salmon. 2001. “Neutrophil Ageing and Immunesenescence.” *Mechanisms of Ageing and Development* 122 (14): 1521–35.

Love, Michael I., Wolfgang Huber, and Simon Anders. 2014. “Moderated Estimation of Fold Change and Dispersion for RNA-Seq Data with DESeq2.” *Genome Biology* 15 (12): 550.

Mellén, Marian, Pinar Ayata, Scott Dewell, Skirmantas Kriaucionis, and Nathaniel Heintz. 2012. “MeCP2 Binds to 5hmC Enriched within Active Genes and Accessible Chromatin in the Nervous System.” *Cell* 151 (7): 1417–30.

Monaco, Gianni, Bernett Lee, Weili Xu, Seri Mustafah, You Yi Hwang, Christophe Carré, Nicolas Burdin, et al. 2019. “RNA-Seq Signatures Normalized by mRNA Abundance Allow Absolute Deconvolution of Human Immune Cell Types.” *Cell Reports* 26 (6): 1627–40.e7.

Moran-Crusio, Kelly, Linsey Reavie, Alan Shih, Omar Abdel-Wahab, Delphine Ndiaye-Lobry, Camille Lobry, Maria E. Figueroa, et al. 2011. “Tet2 Loss Leads to Increased Hematopoietic Stem Cell Self-Renewal and Myeloid Transformation.” *Cancer Cell* 20 (1): 11–24.

Morgan M, Pages H, Obenchain V, Hayden N. 2018. “Rsamtools: Binary Alignment (BAM), FASTA, Variant Call (BCF), and Tabix File Import.” https://doi.org/10.18129/B9.bioc.Rsamtools.

O’Flaherty, Emmett, and Jonathan Kaye. 2003. “TOX Defines a Conserved Subfamily of HMG-Box Proteins.” *BMC Genomics* 4 (1): 13.

Park, Y., W. Kim, J-M Lee, J. Park, J. K. Cho, K. Pang, J. Lee, et al. 2015. “Cytoplasmic DRAK1 Overexpressed in Head and Neck Cancers Inhibits TGF-β1 Tumor Suppressor Activity by Binding to Smad3 to Interrupt Its Complex Formation with Smad4.” *Oncogene*. https://doi.org/10.1038/onc.2014.423.

Peters, Marjolein J., Roby Joehanes, Luke C. Pilling, Claudia Schurmann, Karen N. Conneely, Joseph Powell, Eva Reinmaa, et al. 2015. “The Transcriptional Landscape of Age in Human Peripheral Blood.” *Nature Communications* 6 (October): 8570.

Roadmap Epigenomics Consortium, Anshul Kundaje, Wouter Meuleman, Jason Ernst, Misha Bilenky, Angela Yen, Alireza Heravi-Moussavi, et al. 2015. “Integrative Analysis of 111 Reference Human Epigenomes.” *Nature* 518 (7539): 317–30.

Seo, Hyungseok, Joyce Chen, Edahí González-Avalos, Daniela Samaniego-Castruita, Arundhoti Das, Yueqiang H. Wang, Isaac F. López-Moyado, et al. 2019. “TOX

and TOX2 Transcription Factors Cooperate with NR4A Transcription Factors to Impose CD8 T Cell Exhaustion." *Proceedings of the National Academy of Sciences of the United States of America* 116 (25): 12410–15.

Shu, Xing-Sheng, Hua Geng, Lili Li, Jianming Ying, Chunhong Ma, Yajun Wang, Fan Fong Poon, et al. 2011. "The Epigenetic Modifier PRDM5 Functions as a Tumor Suppressor through Modulating WNT/β-Catenin Signaling and Is Frequently Silenced in Multiple Tumors." *PloS One* 6 (11): e27346.

Song, Chun-Xiao, Keith E. Szulwach, Ye Fu, Qing Dai, Chengqi Yi, Xuekun Li, Yujing Li, et al. 2011. "Selective Chemical Labeling Reveals the Genome-Wide Distribution of 5-Hydroxymethylcytosine." *Nature Biotechnology* 29 (1): 68–72.

Song, Chun-Xiao, Chengqi Yi, and Chuan He. 2012. "Mapping Recently Identified Nucleotide Variants in the Genome and Transcriptome." *Nature Biotechnology* 30 (11): 1107–16.

Szulwach, Keith E., Xuekun Li, Yujing Li, Chun-Xiao Song, Hao Wu, Qing Dai, Hasan Irier, et al. 2011. "5-hmC-Mediated Epigenetic Dynamics during Postnatal Neurodevelopment and Aging." *Nature Neuroscience* 14 (12): 1607–16.

Szwagierczak, Aleksandra, Sebastian Bultmann, Christine S. Schmidt, Fabio Spada, and Heinrich Leonhardt. 2010. "Sensitive Enzymatic Quantification of 5-Hydroxymethylcytosine in Genomic DNA." *Nucleic Acids Research* 38 (19): e181.

Tahiliani, Mamta, Kian Peng Koh, Yinghua Shen, William A. Pastor, Hozefa Bandukwala, Yevgeny Brudno, Suneet Agarwal, et al. 2009. "Conversion of 5-Methylcytosine to 5-Hydroxymethylcytosine in Mammalian DNA by MLL Partner TET1." *Science* 324 (5929): 930–35.

Terragni, Jolyon, Jurate Bitinaite, Yu Zheng, and Sriharsa Pradhan. 2012. "Biochemical Characterization of Recombinant β-Glucosyltransferase and Analysis of Global 5-Hydroxymethylcytosine in Unique Genomes." *Biochemistry* 51 (5): 1009–19.

Teschendorff, Andrew E., Usha Menon, Aleksandra Gentry-Maharaj, Susan J. Ramus, Daniel J. Weisenberger, Hui Shen, Mihaela Campan, et al. 2010. "Age-Dependent DNA Methylation of Genes That Are Suppressed in Stem Cells Is a Hallmark of Cancer." *Genome Research* 20 (4): 440–46.

Truong, Thien Phu, Mamiko Sakata-Yanagimoto, Momoko Yamada, Genta Nagae, Terukazu Enami, Rie Nakamoto-Matsubara, Hiroyuki Aburatani, and Shigeru Chiba. 2015. "Age-Dependent Decrease of DNA Hydroxymethylation in Human T Cells." *Journal of Clinical and Experimental Hematopathology: JCEH* 55 (1): 1–6.

Tu, Wenjuan, and Sudha Rao. 2016. "Mechanisms Underlying T Cell Immunosenescence: Aging and Cytomegalovirus Infection." *Frontiers in Microbiology* 7 (December): 2111.

Weisbart, R. H., A. Kacena, A. Schuh, and D. W. Golde. 1988. "GM-CSF Induces Human Neutrophil IgA-Mediated Phagocytosis by an IgA Fc Receptor Activation Mechanism." *Nature* 332 (6165): 647–48.

Wen, Lu, Xianlong Li, Liying Yan, Yuexi Tan, Rong Li, Yangyu Zhao, Yan Wang, et al. 2014. "Whole-Genome Analysis of 5-Hydroxymethylcytosine and 5-Methylcytosine at Base Resolution in the Human Brain." *Genome Biology* 15 (3): R49.

Xiong, Jun, Han-Peng Jiang, Chun-Yan Peng, Qian-Yun Deng, Meng-Dan Lan, Huan Zeng, Fang Zheng, Yu-Qi Feng, and Bi-Feng Yuan. 2015. "DNA

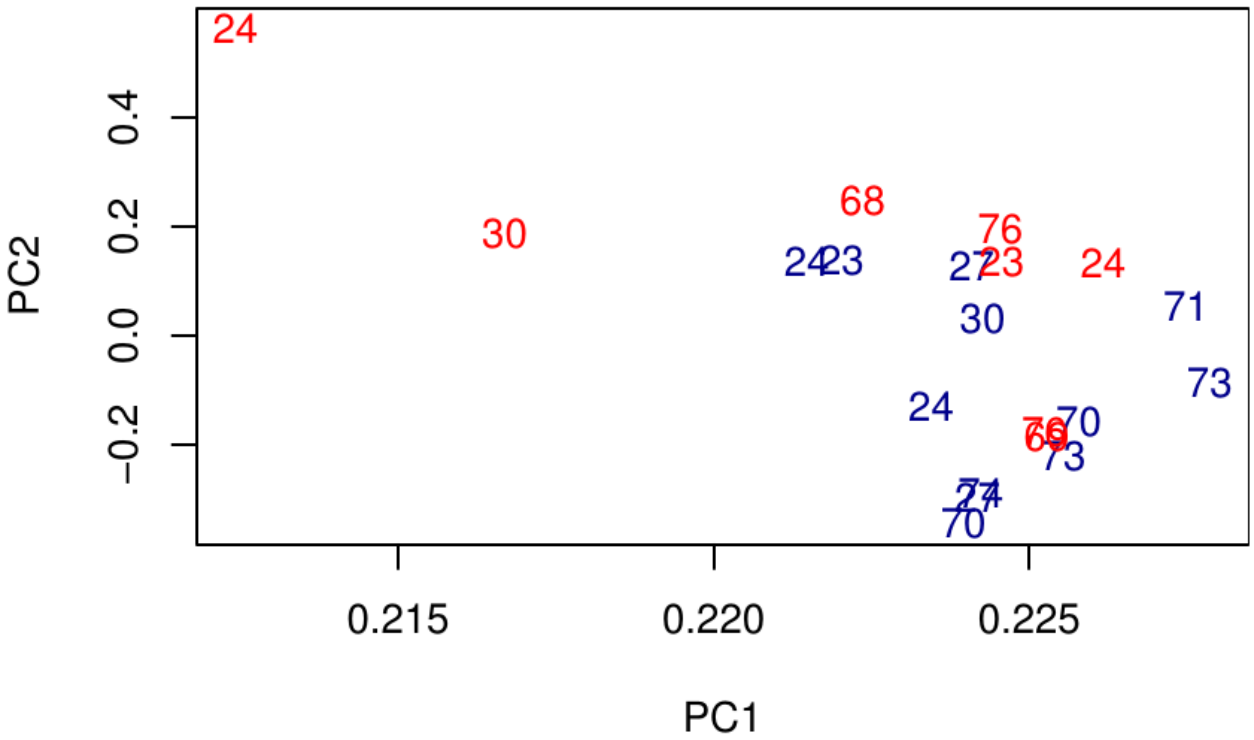
Hydroxymethylation Age of Human Blood Determined by Capillary Hydrophilic-Interaction Liquid Chromatography/mass Spectrometry." *Clinical Epigenetics*. https://doi.org/10.1186/s13148-015-0109-x.

Yao, Bing, Li Lin, R. Craig Street, Zachary A. Zalewski, Jocelyn N. Galloway, Hao Wu, David L. Nelson, and Peng Jin. 2014. "Genome-Wide Alteration of 5-Hydroxymethylcytosine in a Mouse Model of Fragile X-Associated Tremor/ataxia Syndrome." *Human Molecular Genetics* 23 (4): 1095–1107.

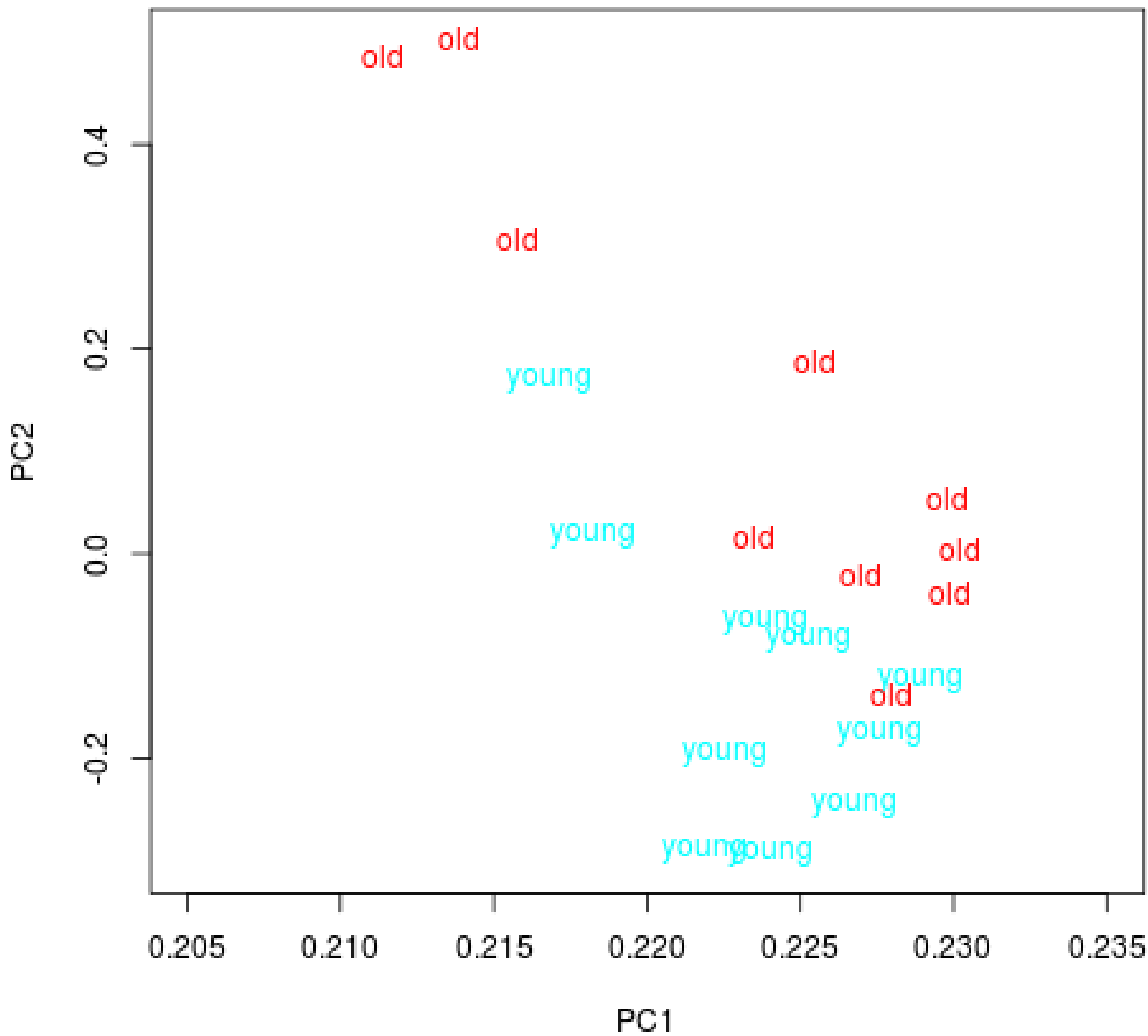
Zhang, Si-Gong, Yu-Xin Song, Xiao-Ming Shu, Hai-Li Shen, Han-Bo Yang, Rui-Xue Duo, and Guo-Chun Wang. 2017. "A Simple Method for Removing Low-Density Granulocytes to Purify T Lymphocytes from Peripheral Blood Mononuclear Cells." *Journal of Zhejiang University. Science. B* 18 (7): 605–14.

Zhao, Jinying, Yun Zhu, Jingyun Yang, Lin Li, Hao Wu, Philip L. De Jager, Peng Jin, and David A. Bennett. 2017. "A Genome-Wide Profiling of Brain DNA Hydroxymethylation in Alzheimer's Disease." *Alzheimer's & Dementia: The Journal of the Alzheimer's Association* 13 (6): 674–88.

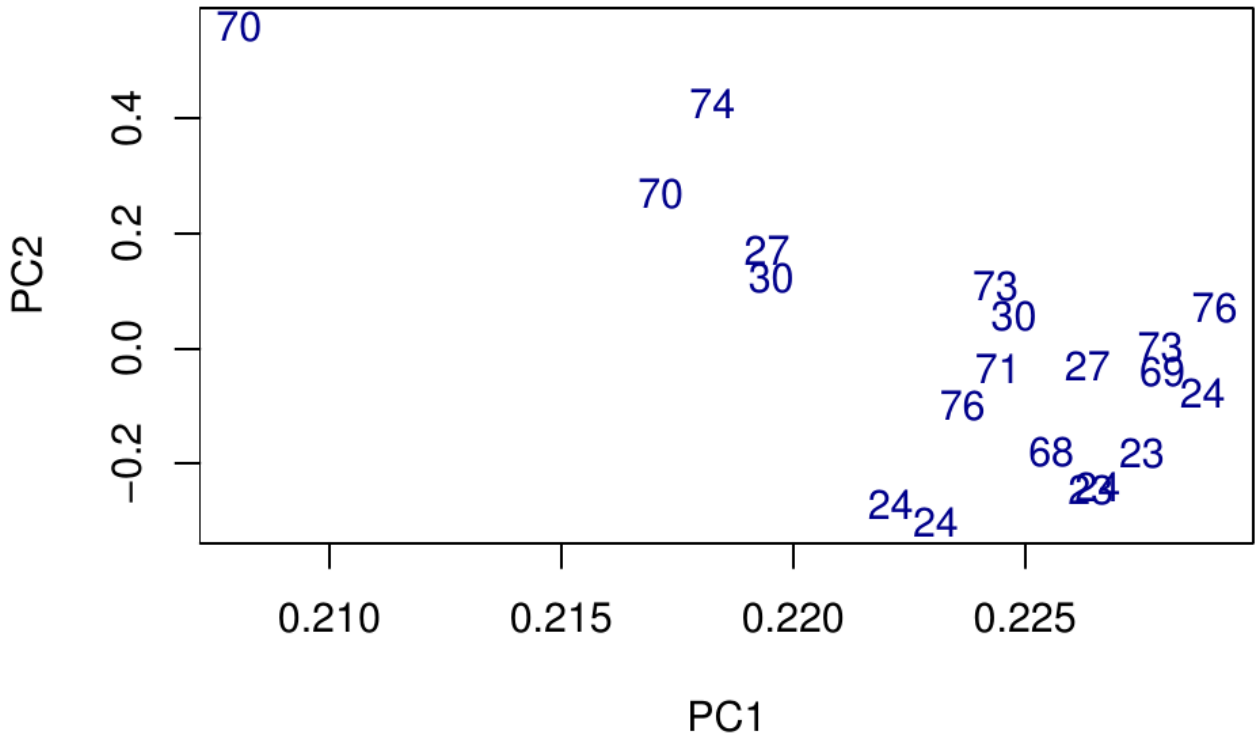
Zhubi, A., Y. Chen, E. Dong, E. H. Cook, A. Guidotti, and D. R. Grayson. 2014. "Increased Binding of MeCP2 to the GAD1 and RELN Promoters May Be Mediated by an Enrichment of 5-hmC in Autism Spectrum Disorder (ASD) Cerebellum." *Translational Psychiatry* 4 (January): e349.



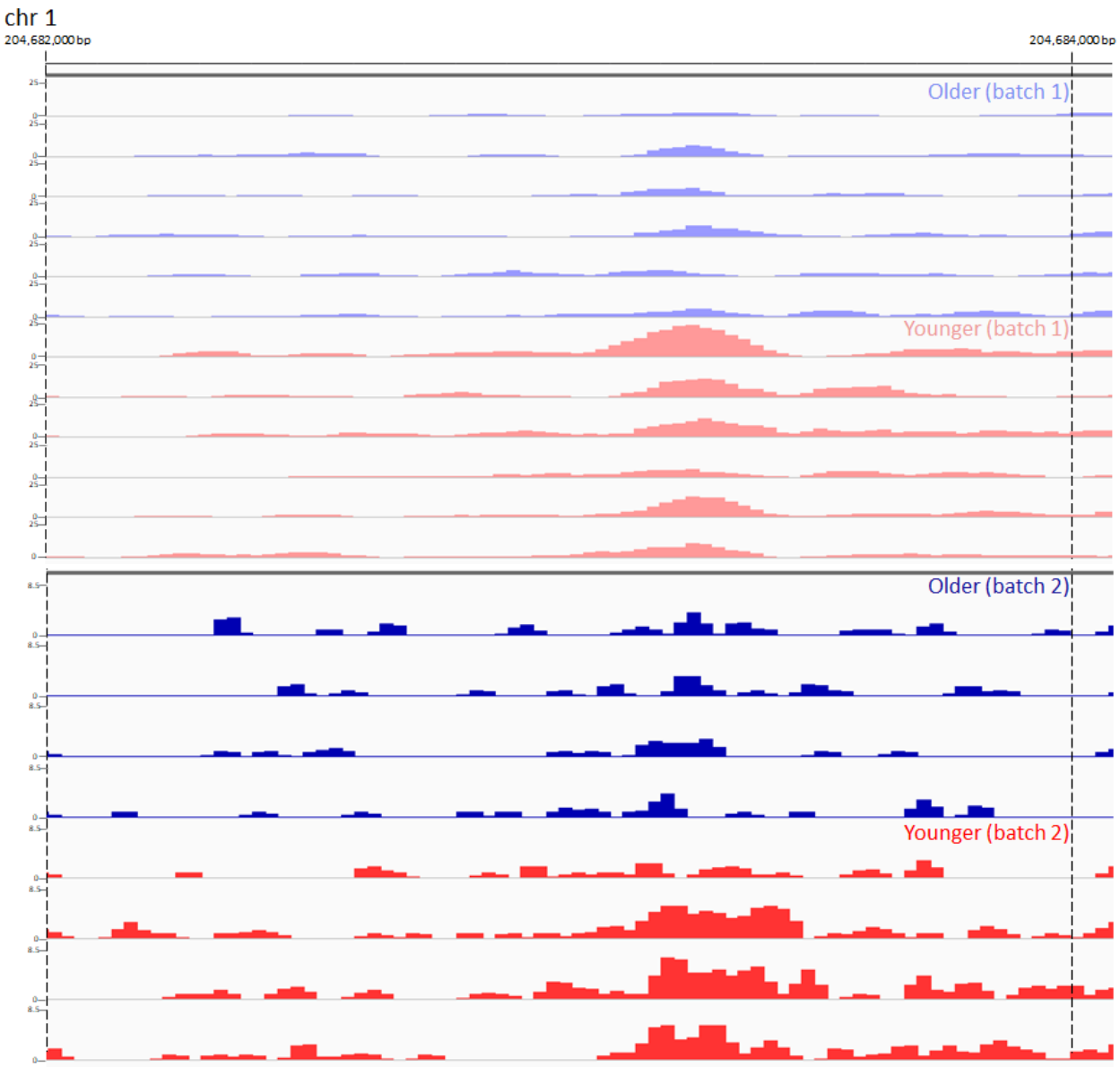
Supplementary Figure 3.1. Second principal component plotted against the first principal component for the sample data DNAhm profiles using the matrix of read counts overlapping 2000 bp bins. Batches 1 and 2 appear in red and blue, respectively. The numbers indicate age in years. There was a significant difference between PC1 scores (p=0.033) and PC2 scores (p=0.049) by age group, suggesting moderate clustering by age group. There was also a significant difference between PC2 scores by batch (p=0.028).



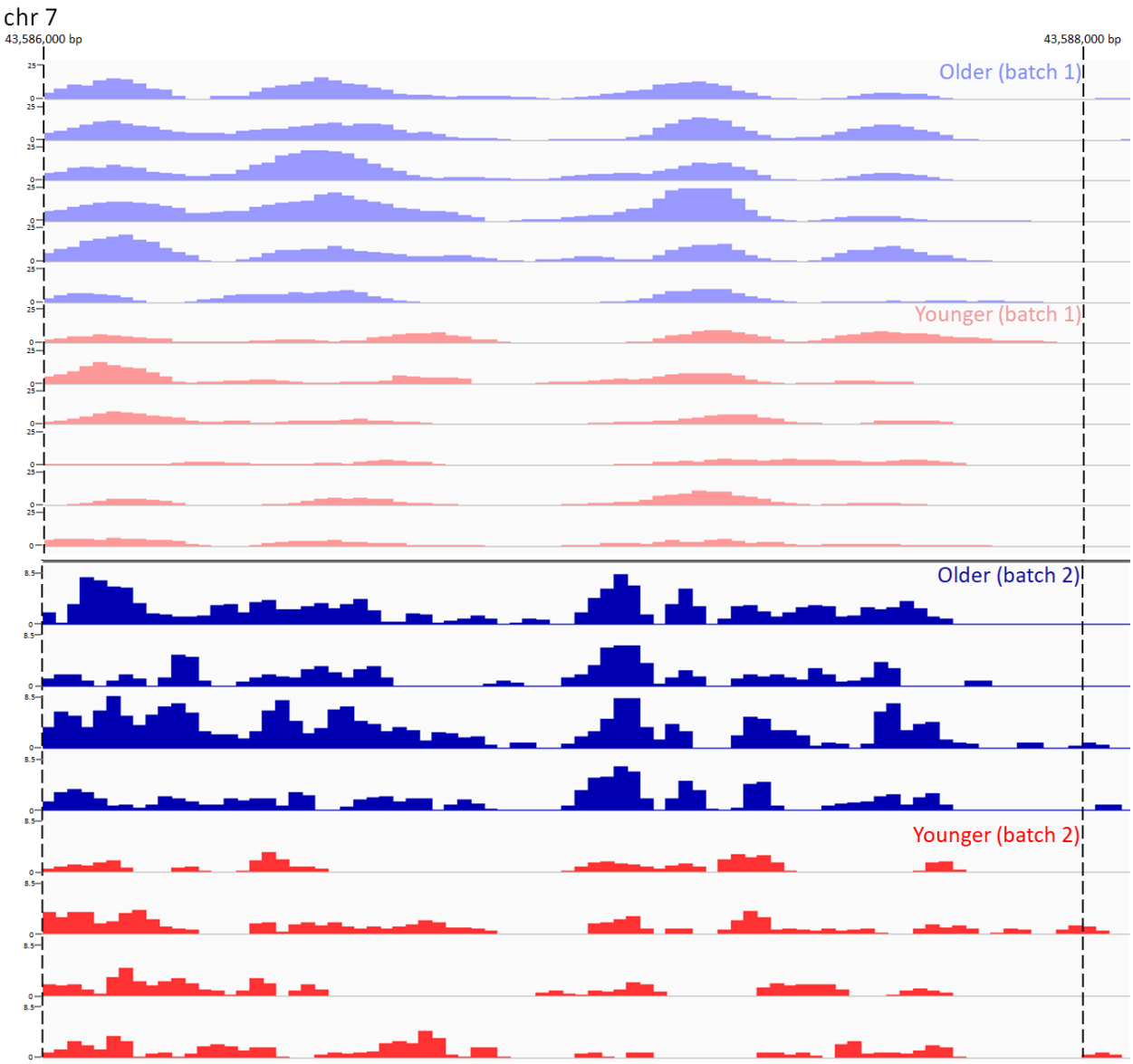
Supplementary Figure 3.2. Second principal component plotted against the first principal component for the sample RNA-seq profiles. The numbers indicate age in years. There was a significant difference between PC2 scores by age group (p=0.039).



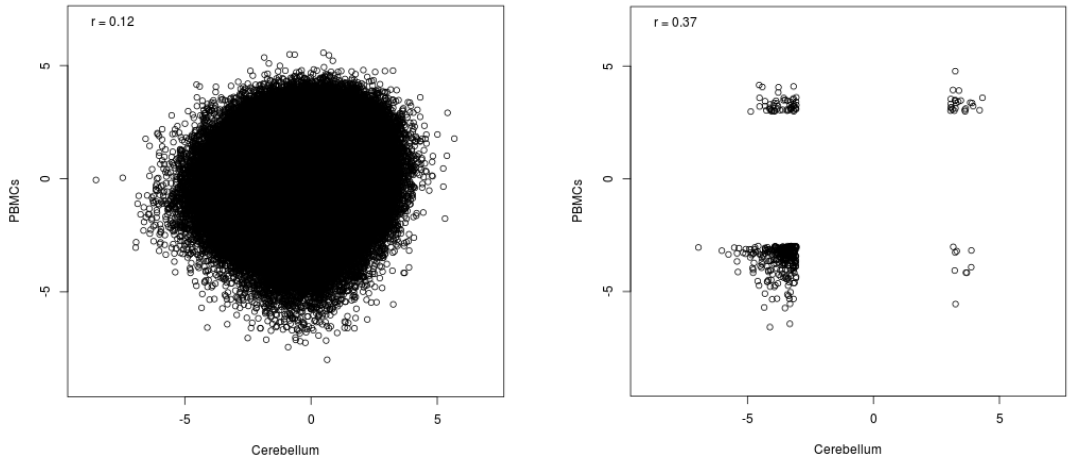
Supplementary Figure 3.3. Genome browser view of 5hmC reads for each sample overlapping the most significant dhMR (chr20:43,930,0001-43,930,2000), which is in the gene *TOX2*. The y-axis is the scale of read pileups (number of reads) and the x-axis is the genomic location. Samples are grouped by batch and age. Because the read length of batch 1 (151 bp) is three times the read length of batch 2 (51 bp), the read pileups are approximately three times as high in batch 1.



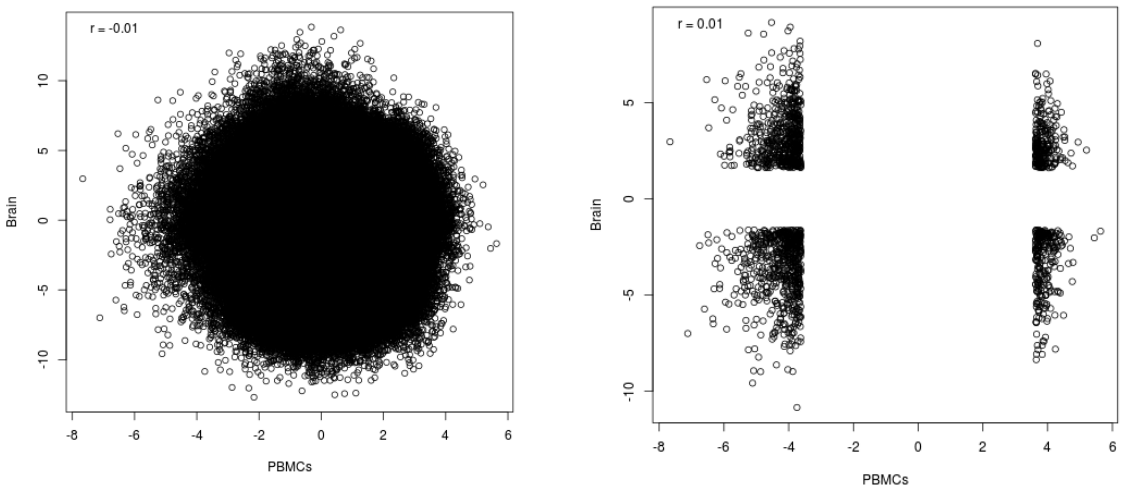
Supplementary Figure 3.4. Genome browser view of 5hmC reads for each sample overlapping the second most significant dhMR (chr1:204,682,001-204,684,000), which is in the gene *LRRN2*. The y-axis is the scale of read pileups (number of reads) and the x-axis is the genomic location. Samples are grouped by batch and age. Because the read length of batch 1 (151 bp) is three times the read length of batch 2 (51 bp), the read pileups are approximately three times as high in batch 1.



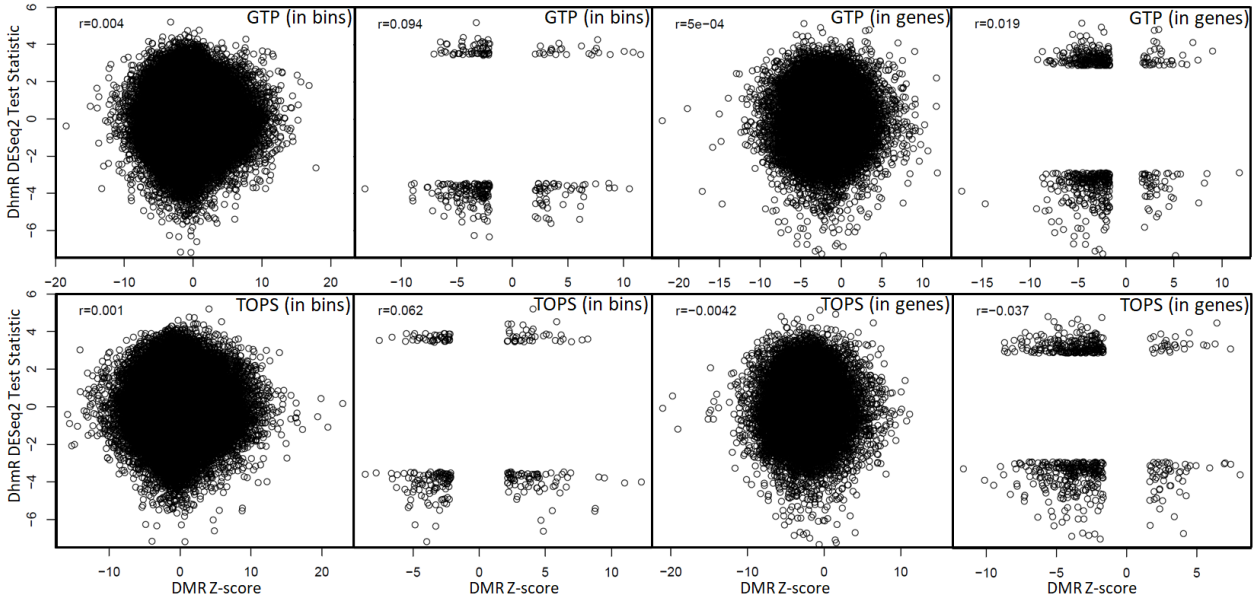
Supplementary Figure 3.5. Genome browser view of 5hmC reads for each sample overlapping the most significant dhMR with increasing hydroxymethylation (chr7:43,586,001-43,588,000), which is in the gene *STK17A*. The y-axis is the scale of read pileups (number of reads) and the x-axis is the genomic location. Samples are grouped by batch and age. Because the read length of batch 1 (151 bp) is three times the read length of batch 2 (51 bp), the read pileups are approximately three times as high in batch 1.



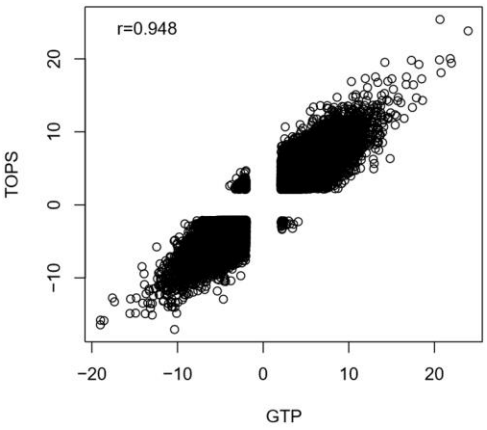
Supplementary Figure 3.6. DhmR test statistics from our study plotted against DhmRs from cerebellar brain tissue from a recently published study (GSE10712) (Cheng et al., 2018). The left panel includes all the overlapping statistics ($p<1x10^{-15}$) and the right panel includes DhmRs that passed on FDR correction of 0.1 ($p<1x10^{-13}$).

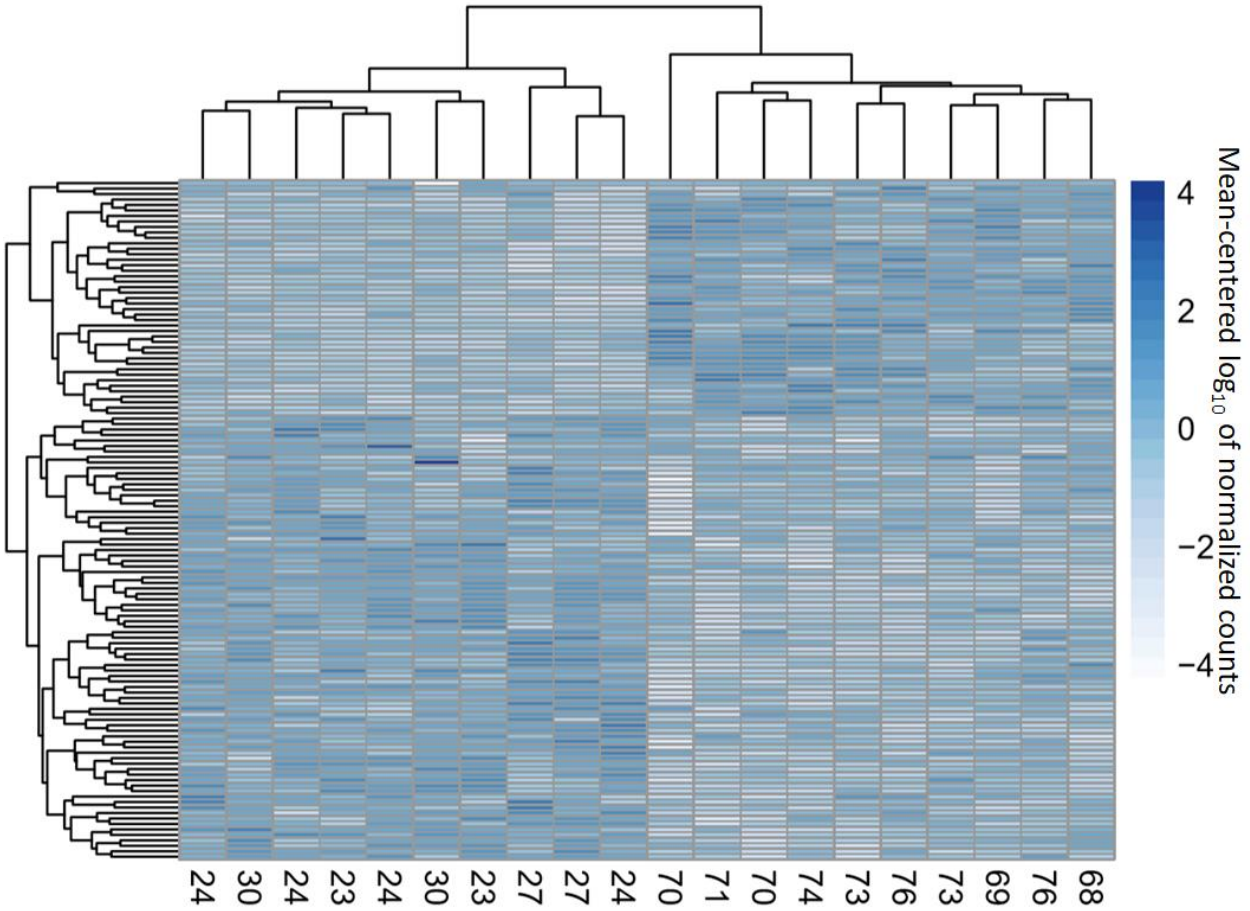


Supplementary Figure 3.7. DhmR test statistics from our study plotted against DhmRs from cerebellar brain tissue from the NIH NeuroBioBank tissues repositories (Kim & Jin, 2019). The left panel includes all the overlapping statistics and the right panel includes DhmRs that passed on FDR correction of 0.1.

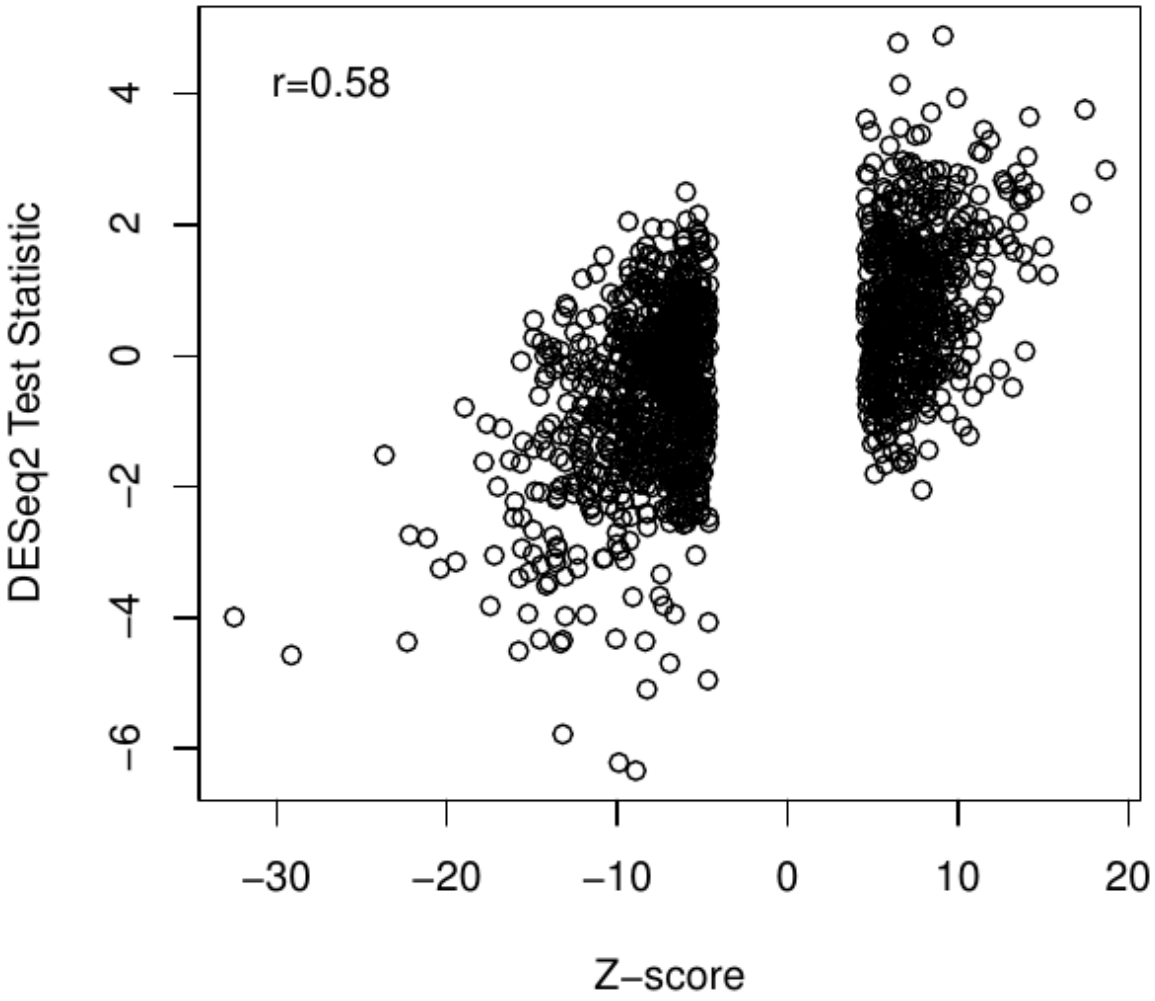


Supplementary Figure 3.8. DhmR test statistics from our study plotted against DMR test statistics from the Grady Trauma Project (top four panels) and the TOPS family study (lower four panels) in bins and genes, respectively. The first and third columns of panels include all the overlapping statistics and the second and fourth panels include test statistics that passed an FDR correction of 0.05. There was no correlation observed between DMRs and DhmRs statistics for either of the two cohorts. The panel to the right is a plot of test statistics in 2000 bp bins of TOPS and GTP plotted against each other.

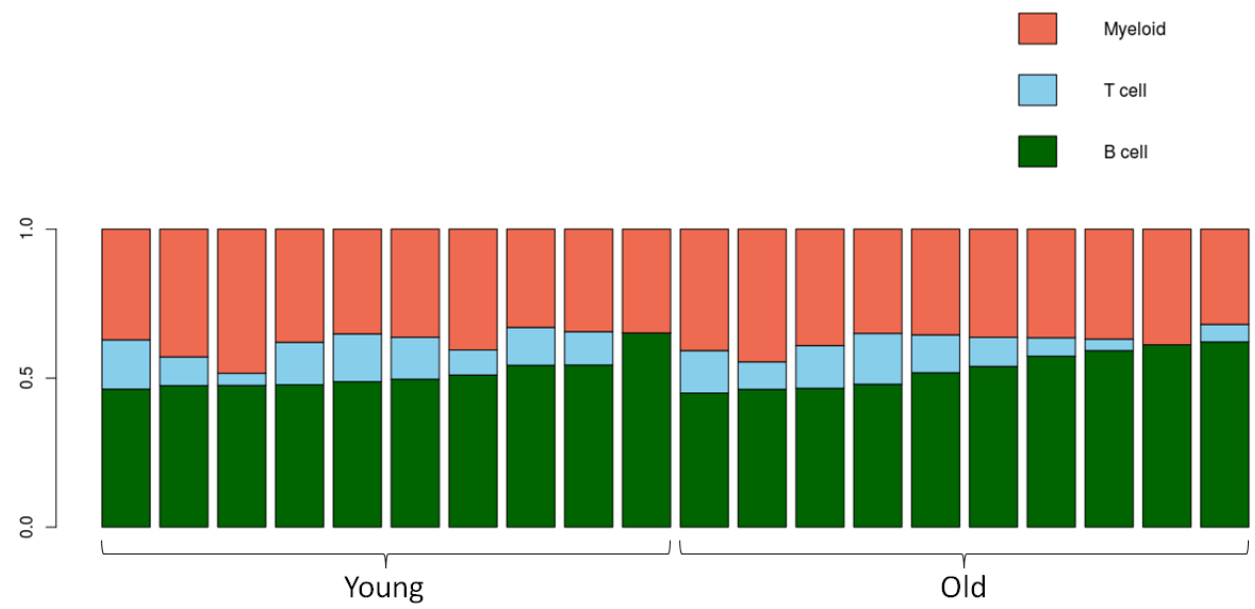




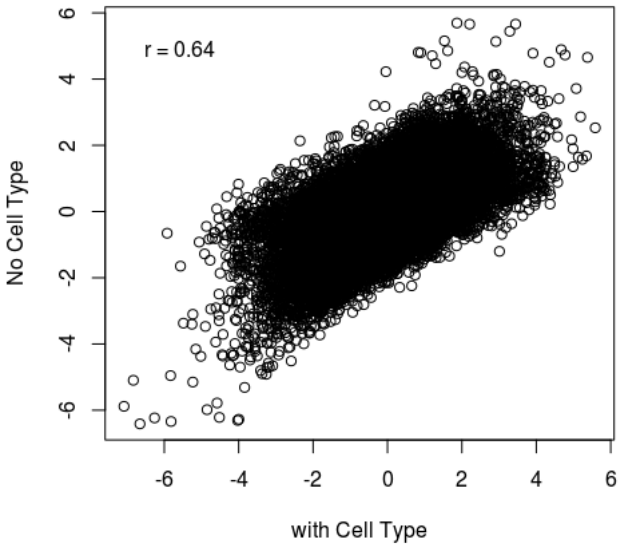
Supplementary Figure 3.9. Heatmap of $\log_{10}$ of normalized counts of the 6,650 124 DEGs. The numbers along the x-axis indicate ages.



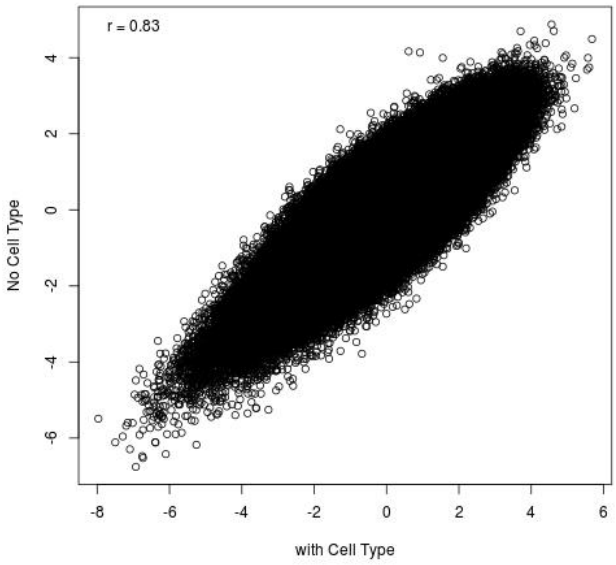
Supplementary Figure 3.10. DESeq2 Test Statistics for age-related gene expression plotted against the Z-scores for significant age-related genes from a previous study (Peters et al., 2015).



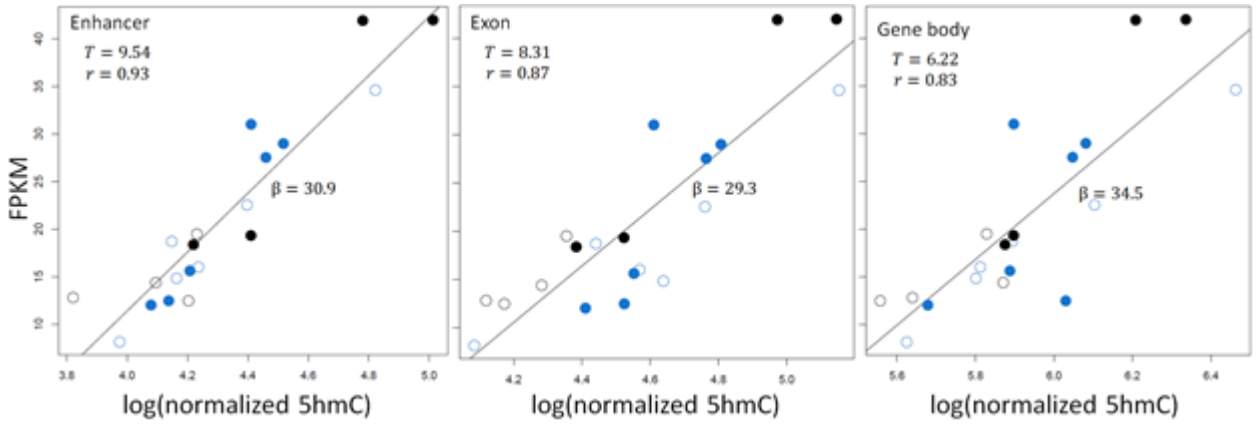
Supplementary Figure 3.11. Estimated cell type proportions using DeconRNASeq for each of the 20 individuals.



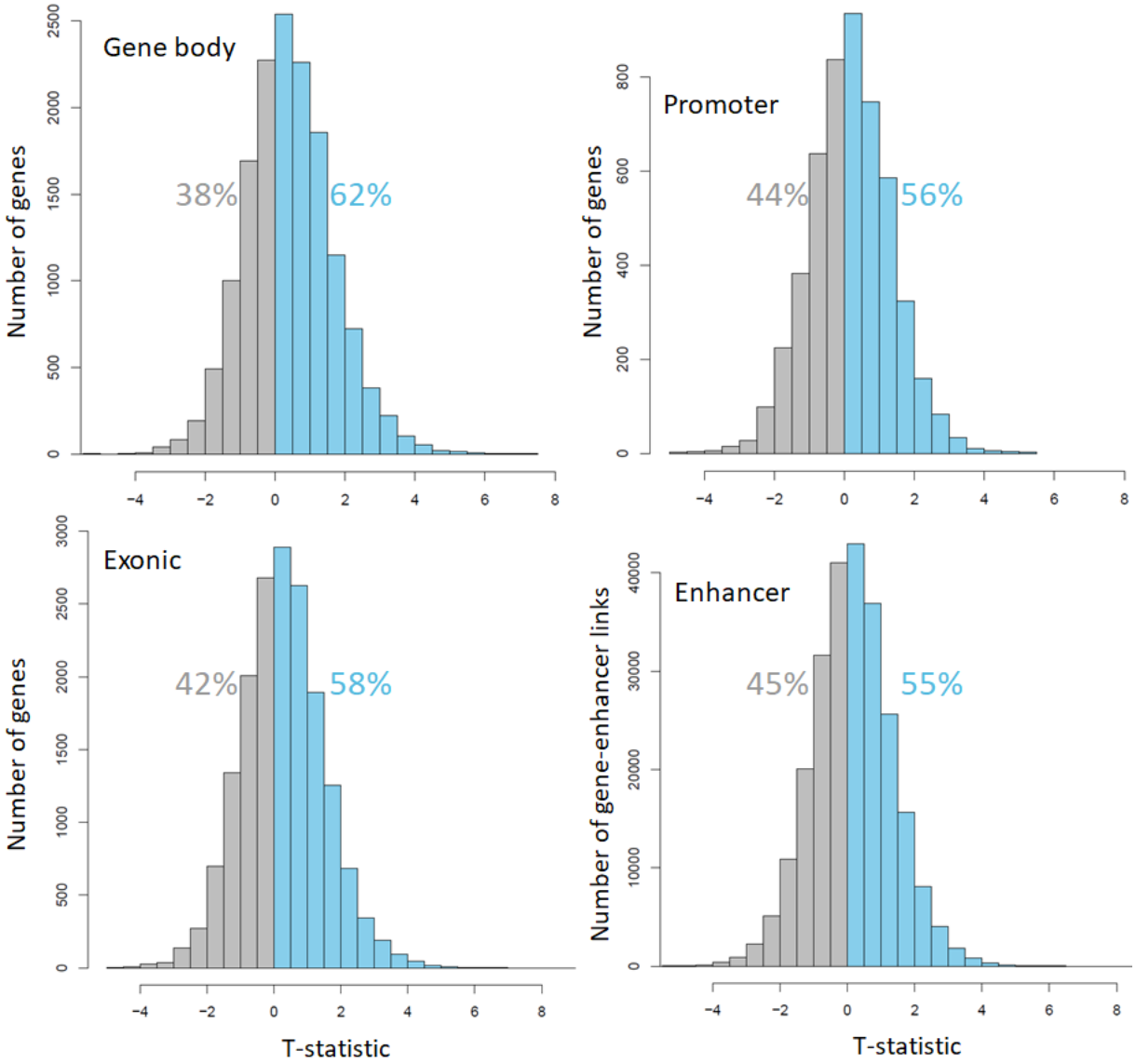
Supplementary Figure 3.12. DESeq2 test statistics for DEGs for the model without estimated cell types as covariates against test statistics for the model with cell types as covariates.



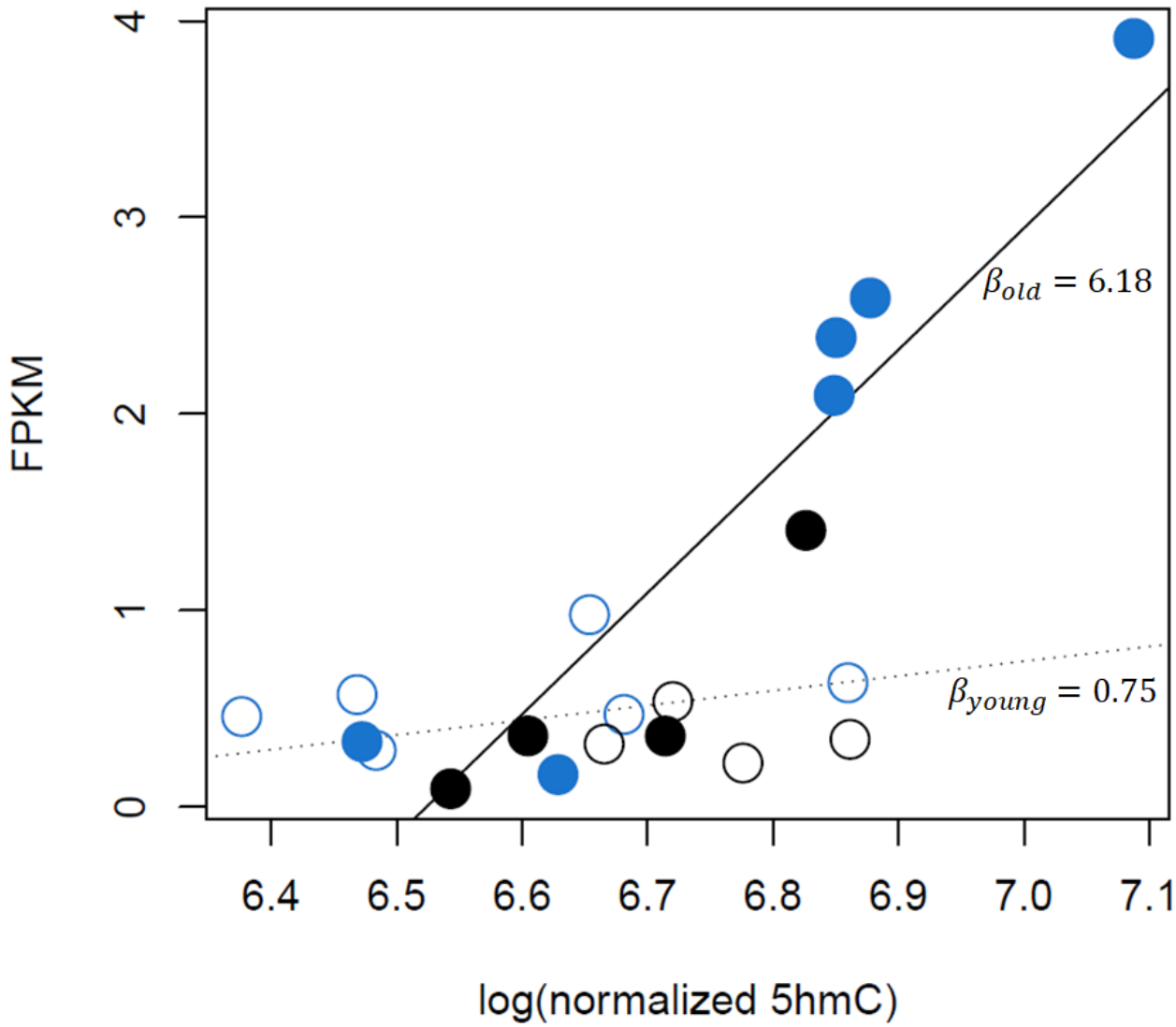
Supplementary Figure 3.13. DESeq2 test statistics for DhmRs for the model without estimated cell types as covariates against test statistics for the model with cell types as covariates.



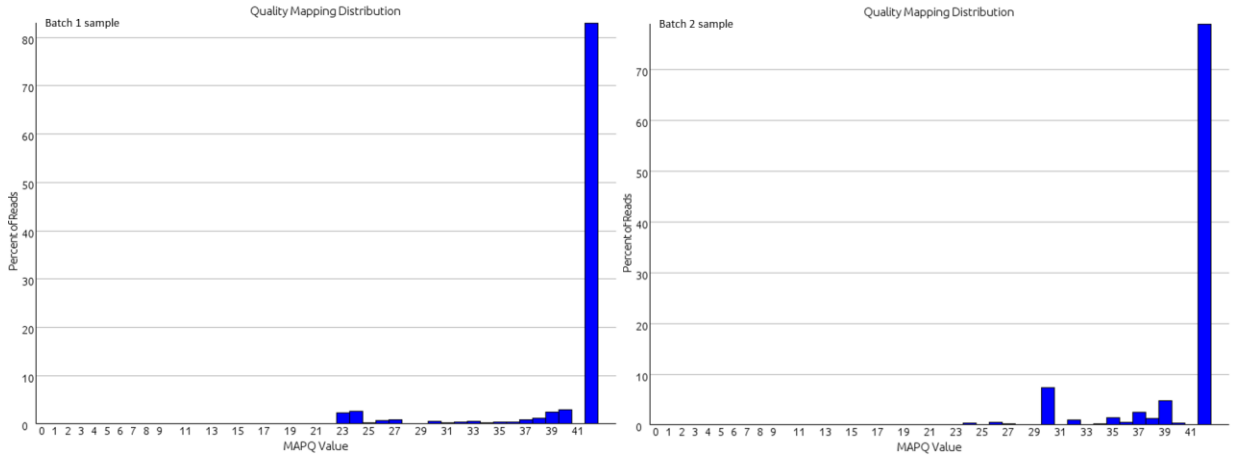
Supplementary Figure 3.14. FPKM scores plotted against the log normalized 5hmC reads in enhancer, exon and gene body for *TBX21*. Solid dots represent old and open dots represent young samples. Batch 1 samples are colored blue and batch 2 samples are colored black. The solid line is the regression line.



Supplementary Figure 3.15. Histograms of T-statistics corresponding the regression of gene expression FPKM values against 5hmC in gene bodies, promoters, exonic regions and enhancers. The average T-statistic was significantly greater than 0 for all four genomic features ($p<10^{-15}$). Genes with positive correlations are highlighted in blue. Percentages in gray and blue correspond to the percent of t-statistics above and below 0, respectively.



Supplementary Figure 3.16. FPKM scores plotted against the log normalized 5hmC reads for *PRDM5*, the one gene with a significant interaction between age and 5hmC ($p=4.9x10^{-5}$). Solid dots represent old and open dots represent young samples. Batch 1 samples are colored blue and batch 2 samples are colored black. The dotted line is the regression line for young and the solid line is the regression line for old.



Supplementary Figure 3.17. Distributions of bowtie
2 quality scores (MAPQ) from two samples, one from batch 1 (left), and the other from batch 2 (right).

Supplementary Table 3.1. 5hmC-capture sequencing data quality measures for each sample.

Batch	Age	Total Reads	% Duplicated	% Uniquely mapped
1	73	29156797	16.26%	79.07%
1	70	32431351	11.40%	78.96%
1	24	33769567	11.40%	80.34%
1	74	29024371	14.56%	77.47%
1	27	33814638	17.55%	77.98%
1	27	20868983	12.00%	76.67%
1	23	15756756	13.36%	74.05%
1	70	45002170	13.59%	81.18%
1	24	26246784	13.88%	77.00%
1	30	28292783	21.21%	72.41%
1	73	38342970	9.22%	72.42%
1	71	31511078	8.90%	61.58%
2	30	27478662	12.05%	52.57%
2	24	30663767	11.52%	60.34%
2	23	27297737	11.04%	60.42%
2	76	36305366	14.06%	53.47%
2	69	29001477	11.55%	65.72%
2	76	33908987	11.92%	62.84%
2	68	35300136	12.31%	51.69%
2	24	30672764	11.55%	47.72%

Supplementary Table 3.2. RNA-sequencing data quality measures for each sample.

Age	Total Reads	% Uniquely Mapped
23	22,794,274	78.67%
23	38,555,544	78.84%
24	23,802,391	76.45%
24	32,108,671	77.74%
24	31,812,722	79.69%
24	40,624,979	79.06%
27	30,060,644	74.16%
27	44,030,380	74.90%
30	28,915,303	78.38%
30	25,766,697	78.92%
68	44,622,123	80.21%
69	50,122,401	77.45%
70	19,311,004	75.97%
70	19,642,201	77.93%
71	33,423,319	76.27%
73	22,142,199	77.27%
73	32,757,101	79.50%
74	25,876,515	78.33%
76	40,590,448	77.36%
76	44,950,796	78.62%

Supplementary Table 3.3 is a csv file (SupplementaryTable3.3.csv)
Genomic information for the 6,650 DhmRs. Column 1 (binIndex) is a unique index for each of the 6,650 bins, column 2 (chr) indicates the chromosome of the bin, column 3 (start) and 4 (end) are the start and end positions of the bin according to the hg38 build of the human genome, column 5 (featureName) is the name of a genomic feature the bin overlaps, column 6 (featureType) indicates the type of overlapping feature (Gene Body, Promoter, or Enhancer), column 7 is the DESeq2 test statistic and column 8 is the corresponding p-value. **DhmRs overlapping multiple genomic features have multiple rows, one for each overlapping feature**.

Supplementary Table 3.4 is a csv file (SupplementaryTable3.4.csv)
Genomic information of the 124 DEGs including the gene name (column 1), the Ensembl name (column 2), chromosome (column 3), start (column 4) and stop position (column 5), DESeq2 test statistic (column 6), p-value (column 7), and q-value

Supplementary Table 3.5. Observed and expected counts of DhmRs overlapping DEGs with both associated increases (+) and decreases (-) with age. Directional consistency is greater than expected by chance. ($\chi_1^2 = 124.6$; $p = 2.2 \times 10^{-16}$).

	(-) DEG (-) DhmR	**(+) DEG (+) DhmR**	**(+) DEG (-) DhmR**	**(-) DEG (+) DhmR**
Observed	154	5	0	1
Expected	71.5	17.3	40.7	30.4

Supplementary Table 3.6. Genes with a significant association between gene expression and exonic DNAhm.

Gene Name	Gene Ensembl Name	T-statistic	p-value	q-value
SPON2	ENSG00000159674	8.66	1.96E-07	2.46E-03
TBX21	ENSG00000073861	8.31	3.36E-07	2.46E-03
CLUH	ENSG00000132361	6.82	4.13E-06	1.26E-02
SLC45A4	ENSG00000022567	6.78	4.45E-06	1.26E-02
STEAP4	ENSG00000127954	6.73	4.84E-06	1.26E-02
SEMA4C	ENSG00000168758	6.67	5.42E-06	1.26E-02
UBN1	ENSG00000118900	6.61	6.01E-06	1.26E-02
HLA-C	ENSG00000204525	6.33	9.94E-06	1.82E-02
DEF8	ENSG00000140995	6.20	1.26E-05	1.87E-02
STK40	ENSG00000196182	6.20	1.28E-05	1.87E-02
IGHM	ENSG00000211899	6.04	1.73E-05	2.30E-02
DENND3	ENSG00000105339	5.80	2.73E-05	3.34E-02
LRRC4	ENSG00000128594	5.70	3.32E-05	3.74E-02
CTA-833B7.2	ENSG00000183822	5.61	3.89E-05	4.07E-02
ZAP70	ENSG00000115085	5.57	4.21E-05	4.12E-02
TBC1D16	ENSG00000167291	5.52	4.66E-05	4.27E-02

Supplementary Table 3.7. Genes with a significant association between gene expression and gene body DNAhm.

Gene Name	Gene Ensembl Name	T-statistic	p-value
GLT1D1	ENSG00000151948	7.24	1.96E-06
HLA-DRB1	ENSG00000196126	6.92	3.46E-06
ATP10A	ENSG00000206190	6.54	6.87E-06
IGHM	ENSG00000211899	6.46	7.86E-06
IGHD	ENSG00000211898	6.29	1.08E-05
TBX21	ENSG00000073861	6.22	1.22E-05
CDA	ENSG00000158825	5.99	1.87E-05
YES1	ENSG00000176105	5.89	2.29E-05
NR1D2	ENSG00000174738	5.84	2.49E-05
JAKMIP1	ENSG00000152969	5.76	2.92E-05
SNRPD1	ENSG00000167088	5.74	3.06E-05
FAM169A	ENSG00000198780	5.68	3.38E-05
PPP3CC	ENSG00000120910	5.60	3.98E-05
BTNL8	ENSG00000113303	5.47	5.15E-05
APBA2	ENSG00000034053	5.34	6.59E-05
TRPM6	ENSG00000119121	5.34	6.65E-05
CYCS	ENSG00000172115	5.33	6.70E-05
RP11-164H13.1	ENSG00000257275	5.31	7.10E-05
KIAA0125	ENSG00000226777	5.27	7.62E-05
IRF2BP2	ENSG00000168264	5.26	7.75E-05
PRDM5	ENSG00000138738	5.25	7.97E-05
GTF2H2B	ENSG00000226259	5.19	9.03E-05
UBN1	ENSG00000118900	5.17	9.31E-05
KIAA1551	ENSG00000174718	-5.22	8.35E-05
PRTN3	ENSG00000196415	-6.85	3.92E-06

Supplementary Table 3.8. Significant GO terms for each GO analysis.

	Decreased with age	Increased with age
Differential Gene Expression	None	1. oxygen transport
Differential DNAhm (gene body)	None	1. leukocyte mediated immunity 2. neutrophil activation 3. neutrophil degranulation 4. myeloid cell activation involved in immune response 5. cell activation 6. regulated exocytosis 7. secretion
Differential DNAhm (promoter)	None	None
Differential DNAhm (enhancer)	1. nucleic acid metabolic process	1. neutrophil degranulation 2. positive regulation of catalytic activity 3. amide transport 4. organic substance transport 5. immune response 6. cellular response to organic substance 7. protein modification process 8. immune system development 9. RNA splicing, via transesterification reactions 10. positive regulation of leukocyte differentiation 11. secretion by cell 12. cellular response to nitrogen compound 13. regulation of peptide transport 14. vesicle-mediated transport 15. establishment of localization 16. response to organonitrogen compound 17. peptide secretion 18. positive regulation of secretion 19. positive regulation of metabolic process
Top 1500 genes with high DNAhm	1. protein modification process 2. cell projection morphogenesis 3. regulation of small GTPase mediated signal transduction 4. positive regulation of cellular metabolic process 5. positive regulation of RNA biosynthetic process 6. cellular response to stimulus 7. positive regulation of catalytic activity 8. positive regulation of biosynthetic process	

	9. positive regulation of nitrogen compound metabolic process 10. positive regulation of GTPase activity 11. anatomical structure morphogenesis 12. phosphorus metabolic process 13. regulation of Rho protein signal transduction 14. immune system development 15. negative regulation of nucleobase-containing compound metabolic process 16. regulation of cell differentiation 17. negative regulation of RNA biosynthetic process 18. regulation of cell morphogenesis 19. regulation of nucleic acid-templated transcription 20. vascular endothelial growth factor receptor signaling pathway 21. peptidyl-serine phosphorylation 22. macromolecule localization 23. negative regulation of transcription by RNA polymerase II 24. peptidyl-amino acid modification 25. positive regulation of transcription by RNA polymerase II 26. Fc receptor signaling pathway 27. positive regulation of cell adhesion 28. immune response-regulating signaling pathway 29. negative regulation of phosphorus metabolic process 30. immune response-activating cell surface receptor signaling pathway 31. multicellular organism development 32. positive regulation of substrate adhesion-dependent cell spreading 33. cell-substrate adhesion 34. negative regulation of cellular process 35. Fc-gamma receptor signaling pathway involved in phagocytosis 36. regulation of cell shape 37. nucleobase-containing compound biosynthetic process 38. chromatin organization 39. regulation of cellular macromolecule biosynthetic process 40. positive regulation of cell death 41. positive regulation of apoptotic process 42. adherens junction assembly 43. biological regulation 44. immune effector process 45. regulation of cell junction assembly 46. positive regulation of axonogenesis 47. negative regulation of biosynthetic process 48. platelet activation 49. peptidyl-threonine phosphorylation 50. actin filament bundle assembly 51. actin cytoskeleton reorganization 52. calcium ion transmembrane transport 53. regulation of cellular response to stress 54. regulation of cation transmembrane transport 55. cellular response to organonitrogen compound 56. positive regulation of developmental process 57. positive regulation of protein modification process 58. establishment of localization

	59. actin cytoskeleton organization 60. cell-substrate junction assembly 61. cell activation involved in immune response 62. phagocytosis 63. response to monoamine 64. establishment of cell polarity 65. cellular protein metabolic process 66. positive regulation of transcription, DNA-templated

Supplementary Table 3.9. Frequency of post-QC reads with each Bowtie 2 quality score for each sample. Each row indicates a sample and each column is a quality score. The majority of reads had the maximum quality score for all samples.

Batch	23	24	25	26	27	30	31	32	33	34	35	36	37	38	39	40	42
1	0.02	0.03	0.00	0.01	0.01	0.00	0.00	0.00	0.01	0.00	0.00	0.00	0.01	0.01	0.02	0.03	0.83
1	0.03	0.03	0.00	0.01	0.01	0.00	0.00	0.00	0.01	0.00	0.00	0.00	0.01	0.01	0.03	0.03	0.82
1	0.02	0.02	0.00	0.01	0.01	0.00	0.00	0.00	0.01	0.00	0.00	0.00	0.01	0.01	0.03	0.02	0.85
1	0.02	0.02	0.00	0.01	0.01	0.01	0.00	0.00	0.01	0.00	0.00	0.00	0.01	0.01	0.03	0.03	0.83
1	0.02	0.02	0.00	0.01	0.01	0.01	0.00	0.00	0.01	0.00	0.00	0.00	0.01	0.01	0.03	0.02	0.84
1	0.02	0.02	0.00	0.01	0.01	0.01	0.00	0.00	0.01	0.00	0.00	0.00	0.01	0.01	0.03	0.03	0.84
1	0.02	0.02	0.00	0.01	0.01	0.01	0.00	0.00	0.01	0.00	0.00	0.01	0.01	0.01	0.03	0.02	0.83
1	0.02	0.02	0.00	0.01	0.01	0.01	0.00	0.00	0.01	0.00	0.00	0.00	0.01	0.01	0.02	0.02	0.85
1	0.02	0.02	0.00	0.01	0.01	0.01	0.00	0.00	0.01	0.00	0.00	0.01	0.01	0.01	0.03	0.03	0.83
1	0.02	0.02	0.00	0.01	0.01	0.01	0.00	0.00	0.01	0.00	0.00	0.01	0.01	0.01	0.03	0.03	0.83
1	0.02	0.02	0.00	0.01	0.01	0.01	0.00	0.00	0.01	0.00	0.01	0.01	0.01	0.01	0.03	0.03	0.82
1	0.04	0.04	0.00	0.01	0.01	0.01	0.00	0.00	0.01	0.00	0.01	0.01	0.01	0.01	0.03	0.04	0.76
2	0.00	0.01	0.00	0.01	0.00	0.11	0.00	0.01	0.00	0.00	0.02	0.01	0.04	0.02	0.07	0.00	0.71
2	0.00	0.00	0.00	0.01	0.00	0.09	0.00	0.01	0.00	0.00	0.02	0.01	0.03	0.02	0.06	0.00	0.75
2	0.00	0.00	0.00	0.01	0.00	0.09	0.00	0.01	0.00	0.00	0.02	0.01	0.03	0.02	0.06	0.00	0.75
2	0.00	0.00	0.00	0.01	0.00	0.11	0.00	0.01	0.00	0.00	0.02	0.01	0.04	0.02	0.07	0.00	0.70
2	0.00	0.00	0.00	0.00	0.00	0.07	0.00	0.01	0.00	0.00	0.02	0.01	0.03	0.01	0.05	0.00	0.79
2	0.00	0.00	0.00	0.01	0.00	0.08	0.00	0.01	0.00	0.00	0.02	0.01	0.03	0.01	0.05	0.00	0.77
2	0.02	0.02	0.00	0.01	0.01	0.01	0.00	0.00	0.01	0.00	0.00	0.01	0.01	0.01	0.03	0.03	0.83
2	0.00	0.01	0.00	0.01	0.00	0.13	0.00	0.01	0.00	0.00	0.02	0.01	0.04	0.02	0.08	0.01	0.66

Chapter 4: Single cell methylation has the potential to detect novel imprints and distinguish between hemi-methylation and allele-specific methylation

INTRODUCTION

Imprinting is the parent-of-origin allele-specific repression of a single gene or a cluster of genes achieved via allele-specific methylation (ASM). In humans, imprinting-related proteins bind to ICRs resulting in allele-specific gene expression (ASE) (Imbeault, Helleboid, and Trono 2017). More specifically, zinc finger protein 57 (ZFP57) and zinc finger protein 445 (ZNF445/ZFP445) bind to the 6-mer imprinting binding motif (TGC^{m}CGC), which is present in all mouse and most human ICRs, and are responsible for maintaining imprints (Takahashi et al. 2019; Riso et al. 2016).

Genomic imprints are established in the female and male germlines, transmitted to the zygote stage (Reig and Concha 2012; Y. Li and Sasaki 2011). Post-fertilization imprints are resistant to epigenetic reprogramming allowing them to be faithfully maintained into adult life (Edwards and Ferguson-Smith 2007).

Imprinted genes are typically consistently imprinted across all tissues, but some genes are only imprinted in specific tissues or developmental stages. While the mechanisms underlying tissue-specific imprints are not fully characterized, the TET family of proteins may be involved in the erasure of imprints (Y. Li and Sasaki 2011). IGF2 is one example of a gene with tissue-specific imprinting: it is biallelically expressed in the human fetal brain but in the adult brain, it is maternally imprinted in the hypothalamus and globus pallidus but biallelically expressed in the pons1 (Elhamamsy 2017). In the GNAS locus, G protein alpha-subunit (G_s-alpha) is imprinted in neonatal brown adipose tissue

but not adult brown adipose tissue; it is biallelically expressed in most tissues, but imprinted in the paraventricular nucleus of the hypothalamus, gonads, renal proximal tubules, the pituitary glands, and thyroid (Weinstein et al. 2010). These examples highlight the importance of investigating imprinting within specific tissues, cell types, and developmental stages.

While it is possible for imprinting on one parental chromosome to repress a gene on that same parental copy, imprinting can also cause a gene on the same parental copy to be expressed. For example, the ICR corresponding to *Igf2r* is the promoter of *Airn*, a long non-coding RNA that, when expressed, represses genes in the *Igf2r* cluster (Sleutels, Zwart, and Barlow 2002). Ferguson-Smith (2011) notes that other maternally imprinted clusters also result in repression of protein-coding genes via the expression of non-coding transcripts, and further points out that, while imprints largely originate via germline differentially methylated regions, the mechanism by which they cause mono-allelic expression can vary.

Several lines of evidence suggest imprinting could play a major role in immunity and disease. A maternal-imprinted locus known as *Dlk1-Gtl2* maintains the self-renewal capacity of hematopoietic stem cells (HSCs) (Serrano-Lopez and Cancelas 2016), which is the precursor lineage of both myeloid and lymphoid immune cells. Further, skewing of the HSC lineage toward the myeloid lineage is a hallmark of immunosenescence, and a reduced self-renewal capacity underlies this skewing (Geiger, de Haan, and Carolina Florian 2013). In addition, neonatal maternal imprinting of the B cell repertoire in mice has been reported to have long-term consequences on functional immunity to infectious viruses (Fink et al. 2008). Imprinting plays a role in many diseases, most notably,

imprinting disorders such as Angleman syndrome and Prader-Willi syndrome as well as several other disorders that have been recently reviewed (Butler 2009; Elhamamsy 2017). Loss of imprinting results in tumorigenesis in mouse embryonic stem cells (Holm et al. 2005), and is associated with cancer risk in multiple cancer types (Feinberg, Ohlsson, and Henikoff 2006).

Here, we use single cell sequencing to investigate imprinting. Our study is the first to observe the allelic-methylation state in individual cells and compare it to the allelic-expression state in those same cells using single-cell reduced representation bisulfite sequencing (scRRBS) and single-cell RNA-sequencing (scRNAseq), respectively. We also develop criteria to detect CpG sites with allele-specific methylation (ASM) using scRRBS. We also called heterozygous variants, which we used to detect both methylation states (methylated and unmethylated) at each allele within individual cells, which allowed us to observe ASM at the cellular level. Among CpGs with ASM in ≥3 samples that also had a nearby heterozygous site that allowed us to observe the methylation state of both alleles, 95% of the CpGs also had cellular-level ASM.

METHODS

Data Processing

We processed single cell reduced representation bisulfite sequencing (scRRBS) data on 6 samples of B cells obtained from healthy individuals, which we obtained from a publicly available dataset of multiplexed fastq files deposited on the Gene Expression Omnibus (GSE109085) (Gaiti et al. 2019). scRRBS data for samples B01-B06 included data on 92, 108, 51, 169, 96, and 94 B cells, respectively. We used a custom script to demultiplex fastq files using cell-specific barcodes. To remove 6-bp barcodes and bases added during the end repair reaction, we used fastx_trimmer to trim off the first 7 bp from the 5' end of reads. To address the possibility that reads could contain adapters on the 3' end, we used TrimGalore, a trimming tool designed to address issues specific to RRBS sequencing (Krueger 2015). In particular, we used the following arguments to remove 3' end adapters along with experimentally added cytosines that do not necessarily correspond to the true methylation state:

--rrbs

--paired

--adapter

--three_prime_clip_R1 3

--three_prime_clip_R2 3

After trimming, we aligned the demultiplexed paired-end reads to the hg38 build of the human genome using the bismark aligner with bowtie 2. To account for bisulfite conversion, bismark aligns to both the human genome and an *in silico* bisulfite-converted genome for both the reference and complementary strands.

We used samtools to sort reads by genomic position and remove reads with MAPQ < 20. To quantify DNA methylation in each cell, we used the

bismark_methylation_extractor, which retains strand-specific information. That is, DNA methylation on one strand is measured separately from DNA methylation on the opposite strand. Importantly, strand-specific quantification of DNA methylation at the single cell resolution allows for the detection of hemi-methylation and can even distinguish it from allele-specific methylation. To remove low-information cells, we excluded any cells whose average read count was ≤5 across all samples.

Quantification of Potential Allele-Specific Methylation

To interrogate potential allele-specific methylation (ASM), we quantified the number of methylated reads and unmethylated reads at each CpG site at the cellular level. At sites with both unmethylated reads and methylated reads, we tested for allele-specific methylation within each cell by performing a likelihood ratio test. At a CpG site, we pooled methylated and unmethylated reads across cells with read types of both methylation states. If the CpG site corresponded to true allele-specific methylation (alternative hypothesis), we expected the pooled reads to follow a binomial distribution with p=0.5 (i.e. a 50/50 split between methylated and unmethylated reads). In contrast, our likelihood ratio test compared the alternative hypothesis to the null hypothesis that the pooled reads of two states, methylated and unmethylated, originated from a sequencing error modeled as a binomial distribution with p=0.01 (i.e. ~1% of reads correspond to one methylation state and ~99% correspond to the other methylation state;

Equation 1). Within each sample, we rejected the null hypothesis for the LRT if $p < \frac{0.05}{Number\ of\ CpGs\ tested}$.

$$\lambda_{LR} = -2ln\left(\frac{L(p_0=0.01)}{L(p_A=0.5)}\right) \sim \chi_1^2 \qquad (Equation\ 1)$$

$$H_A: X \sim Binom(N, p_A = 0.5)$$

$$H_0: X \sim Binom(N, p_0 = 0.01)$$

$$If\ the\ number\ of\ methylated\ reads\ <\ unmethylated\ reads, then$$

$$X\ is\ the\ number\ of\ methylated\ reads. Otherwise, X\ is\ the\ number\ of\ unmethylated\ reads$$

$$N\ is\ the\ total\ number\ of\ reads$$

Due to low coverage, most CpG sites did not have coverage of both alleles within a cell. Gaiti et al. (2019) observed biallelic coverage in ~5% of available germline SNPs within a cell among the CLL samples (See Extended Data Figure 2 of Gaiti et al. 2019), which is consistent with our observations in the healthy B cell samples. Therefore, a CpG site that is truly ASM typically would have reads corresponding to both methylation states in a minority of the cells with coverage, and among the remaining cells, some would have methylated reads and others would have unmethylated reads in an approximate 50/50 split. To address this, we complemented the likelihood ratio test above (which is applied only to cells with both methylated and unmethylated reads) with a binomial test performed on the remaining cells with data. That is, among these remaining cells, if there are X cells with unmethylated reads and N-X cells with methylated reads, we tested whether X deviated significantly from $X \sim Binomial(N, p = 0.5)$. As an additional criterion, at least 2 cells had to have reads indicative of both a methylated and unmethylated signal for the CpG. Thus, three measures were used to determine ASM within a sample: (1) at least 2 cells with both methylated and unmethylated reads, (2) a significant LR test on these cells

(Equation 1), (3) a non-significant binomial test on the number of unmethylated cells X and methylated cells N-X ($p > 0.05$). To correct for multiple testing, we used a common cut-off of $p < 10^{-8}$. Because it is slightly smaller than the Bonferroni threshold for B01 (1.1×10^{-8}), which had the most sites covered, this cutoff will conservatively control the family-wise error rate at <0.05.

Quantification of Potential Allele-Specific Methylation Using Heterozygous Variants

Quantification of Cellular-level ASM according to a gold standard

We called heterozygous variants in the B04 sample to allow creation of a gold standard measure of ASM for CpG sites on reads overlapping these variants. First, we used samtools to merge and sort the scRRBS aligned bam files into a single bam file. Next, we generated a pileup file using the samtools mpileup function with the --skip-indels setting to exclude insertions and deletions (H. Li et al. 2009). To exclude CpG sites, we also generated the pileup file with a genome masked to CpG sites. Subsequently, we called variants using bcftools. We then filtered out A/G and C/T variant calls because it is impossible to distinguish true heterozygous sites from a locus with both methylated and unmethylated cytosines that undergo bisulfite conversion.

Using a publicly available script (Lindenbaum 2015), we subsetted single-cell bam files so they only included reads overlapping both a heterozygous variant and a CpG site with both methylated and unmethylated read counts. For a CpG to qualify as having ASM according to our gold standard measure, one allele needed to have >3 unmethylated

reads and the other allele needed to have >3 methylated reads within a single cell among reads overlapping both the heterozygous site and CpG site. To be considered “consistent with ASM”, a CpG had to have one allele with 1-3 unmethylated reads and the other allele needed to have 1-3 methylated reads. We considered a CpG an artifact if >3 methylated reads and >3 unmethylated reads overlapped the same strand and the same allele. A CpG was considered hemi-methylated if >3 methylated reads overlapped one strand of an allele and >3 unmethylated reads overlapped the other strand of that same allele. CpGs that had both methylated and unmethylated reads, but did not meet any of the aforementioned criteria were considered inconclusive.

Finding Cells where Allelic Methylation and Allelic Expression were both Detectable

After finding CpGs near heterozygous SNPs whose unmethylated and methylated read counts could be subsetted by each allele, we searched the scRNAseq data for reads overlapping the same heterozygous SNP. We then matched the scRRBS data and scRNAseq data by cell index so that the cellular-level allelic methylation and allelic expression could be observed directly.

Overlap of Allele-specific Methylation with Imprinting Binding Motifs and Previously Reported Imprinted Genes

Imprinting Binding Motif Annotation

We used the matchPattern function within the R Bioconductor package Biostrings to find all instances of the imprinting binding motif in the hg38 build of the human genome. More specifically, we searched for the sequence TGCCGC as these are the first 6 nucleotides of the imprinting binding motif (the last nucleotide is variable).

Imprinted Genes Annotation

We obtained a list of imprinted genes from a publicly available catalogue of human imprinted genes (Jirtle 2012). We then used an Ensembl gtf file (Homo_sapiens.GRCh38.87.chr.gtf) to match each gene with its corresponding genomic coordinates in hg38 (Yates et al. 2019).

Distance from CpGs to Nearest Imprinted Genes and Binding Motifs

To investigate whether ASM sites were closer to imprinting binding motifs and imprinted genes compared to non-ASM CpGs, we used the distanceToNearest function in the R Bioconductor package GenomicRanges. The distribution of distances between CpGs and imprinted genes and the distribution of distances between CpGs and imprinting binding motifs were both approximately log-normally distributed. Thus, we log-transformed the distances and perform t-tests to compare the distances for all CpGs to the distances for ASM CpGs.

Association Analysis of Allele-specific Methylation and Imprinted Genes

Since ASM is often found in the promoter of imprinted genes, we included the region from the transcription start site to 2 kb upstream of the gene in our analysis. Using the Ensembl gene annotation file Homo_sapiens.GRCh38.87.chr.gtf , we used the subset of genes with coverage and further divided them into imprinted genes and those that have no known imprinting. From the genes with coverage we also made subsets of genes with allele-specific methylation and genes with no allele-specific methylation. We then made a 2x2 contingency table of imprinted/not imprinted genes vs. genes with ASM/noASM. To investigate whether ASM and imprinting were associated, we performed a Chi-square test of independence.

Association Analysis of Allele-specific Methylation and Imprinting Binding Motifs

Similar to the association analysis of allele-specific methylation and imprinted genes, we included the 2 kb region upstream of the transcription start site, and we used the same gene annotation file to create separate subsets of genes with binding motifs and those with no binding motif. Genes with no binding motif that overlapped genes with binding motifs were excluded from the analysis. We then made a 2x2 contingency table of genes with/without a binding motif vs. genes with ASM/noASM, and performed a Chi-square test of independence to investigate whether genes with ASM were associated with genes with an imprinting binding motif.

Comparison Across Healthy Samples

To investigate the concordance of ASM between healthy samples, we created contingency tables for each pairwise comparison between B01-B06 (15 pairwise

comparisons altogether) excluding sites that did not have coverage in both pairs. We then performed a Chi-square test of independence for each pairwise comparison and computed an odds ratio.

Comparison of Allele-specific Methylation with Allele-specific Expression

To investigate whether allele-specific methylation associated with allele-specific expression in the B04 sample, we counted the number of reads overlapping each allele of all genes overlapping a heterozygous variant called using the RRBS data as previously described. To compare DNA methylation and gene expression in gene bodies, we created a set of genes with ASM CpGs in the gene body and another set without ASM CpGs. Genes with no ASM that overlapped genes with ASM were excluded from the analysis. We further subdivided these two sets into genes with reads originating from one allele and genes with reads originating from two alleles. We then created a 2x2 contingency table and performed a Chi-square test of independence. We tested for an association between ASM in gene promoters and allelic expression in the same fashion except that ASM was quantified in the region from the transcription start site to 2000 bp upstream of the transcription start site.

Power Analysis

Detection of ASM requires the observation of both alleles. The probability of detecting both alleles increases with coverage. That is, as the number of reads N at a CpG site increases, the probability of detecting both alleles approaches 1:

$\lim_{N\to\infty} P(both\ alleles|N) = 1$.We observed that the log(mean number of reads) at a CpG site increases linearly with the number of cells with both methylated and unmethylated reads (i.e. cells with both alleles present). We were able to use this information to estimate the $P(1\ allele\ present)$ and $P(2\ alleles\ present)$. Assuming a log-normal distribution of reads, we estimated power for each CpG within each sample as follows.

$$P(both\ alleles\ present|N)$$

$$= \frac{P(2\ alleles\ present)(1-0.5^{N-1})\frac{1}{\sigma_2}e^{-\frac{(logN-\mu_2)^2}{2\sigma_2^2}}}{P(1\ allele\ present)\frac{1}{\sigma_1}e^{-\frac{(logN-\mu_1)^2}{2\sigma_1^2}} + P(2\ alleles\ present)\frac{1}{\sigma_2}e^{-\frac{(logN-\mu_2)^2}{2\sigma_2^2}}}$$

μ_1: mean of log-read counts in cells with 1 allele

μ_2: mean of log-read counts in cells with 2 alleles

σ_1: standard deviation of log-read counts in cells with 1 allele

σ_2: standard deviation of log-read counts in cells with 2 alleles

RESULTS

Sample Characteristics

After filtering out cells with low read counts during quality control, the final cell counts with DNA methylation data ranged from 51 to 169 across the 6 samples (B01-B06) (Table 4.1). B03 was the sample with the fewest reads (73,182,855 reads), and B01 was the sample with the most reads (354,878,229 reads).

Each cell typically had coverage in at least 100,000 CpGs with a median number of reads per CpG ranging from 8 to 97, after filtering out CpGs with <6 reads (see

Methods). B01 and B02 had an exceptionally large number of CpGs with coverage per cell compared to the other samples with a median number of CpGs of 522,614 and 247,196, respectively. At the sample level, the total number of CpGs with coverage ranged from 2,213,651 to 4,442,170. The total number of CpGs with coverage across all cells within a sample was 3,036,327 in B04, 2,679,381 in B05, and 2,222,858 in B06.

Allele-Specific Methylation in Healthy Samples

We observed 88,956 CpGs with allele-specific methylation in B01, 7,853 in B02, 12,050 in B03, 24,346 in the B04 (Figure 4.1), 3,348 in the B05 sample, and 2,461 in the B06 sample ($p < 10^{-8}$). We also observed a greater-than-expected overlap of ASM across samples. 139,014 CpG sites had ASM in at least one sample, with 7,430 in two samples, 618 in three samples, 41 in four samples, and 10 in five samples (Figure 4.2). We observed a significant association in every pairwise comparison of ASM between samples (Figure 4.3; Supplementary Table 4.1). No CpG site had ASM detected in all 6 samples, which is consistent with our estimates of power to detect ASM in all 6 samples being very small (Supplementary Figure 4.1). However, the power to detect ASM in at least one sample is much higher, which coincides with our finding that the ASM we detected in a single sample was usually not detected in other samples.

Allelic Methylation State at CpGs near Variant Sites

To assess the accuracy of our ASM calls, we compared these calls to a more traditional (gold standard) approach utilizing heterozygous variant calls, for a subset of CpGs that had neighboring heterozygous variants. We called heterozygous sites from the scRRBS data, and subsequently counted methylated and unmethylated reads at CpG sites that had reads overlapping both the heterozygous sites and the CpG. Table 4.2 shows the number of variants called within each sample, the number of overlapping CpGs with both

methylated and unmethylated reads overlapping these variants, and the number of those CpGs that were called as having cellular-level ASM. Within each cell, we classified a CpG as having cellular-level ASM according to the gold standard definition if it had >3 methylated reads at one allele with no unmethylated reads at that same allele as well as >3 unmethylated reads at the other allele with no methylated reads at that allele. We considered CpGs with one methylated allele and one unmethylated allele, but fewer reads than the aforementioned cutoff, as being "consistent with ASM" in that particular cell. These variant-based ASM calls were considered the gold standard because the presence of the variant allowed assignment of methylated and unmethylated reads to a specific strand and allele, B01 had the most CpGs with cellular-level ASM and B06 had the least (Table 4.3). At these CpG sites, the methylation state of the cytosines could be differentiated by both strand and allele, thereby allowing us to differentiate between hemi-methylation and allele-specific methylation at the cellular level. When comparing to the gold standard calls (ASM or consistent with ASM), the true positive rate of our sample-level ASM calls was 89% among CpGs with ASM called in ≥1 sample (Table 4.4). The true positive rate increased to 95% among CpGs with sample-level ASM called in ≥3 samples.

We also found that the power to detect ASM at a CpG was higher among CpGs with ASM detected in multiple samples (Supplementary Figure 4.2).

We used the B04 sample, which was the only sample with cell-matched scRNAseq and scRRBS data, to compare allele-specific DNAm to allele-specific gene expression. Among the 185 gold standard CpG sites in B04, we investigated whether gene expression reads overlapped any of the heterozygous variants, allowing assessment of allele-specific expression. Among the heterozygous variants corresponding to the 185 CpGs, we found that 4 variants had overlapping RNAseq reads originating from the same cells as the DNA methylation data (Table 4.5). We observed expression from a single allele at 3 of these sites. None of the CpG sites were called ASM at the sample-level because there was only 1 cell with both unmethylated and methylated reads. However, 2 of the sites had ASM based on the observed cellular data. Interestingly, it was the methylated allele that was expressed and the unmethylated allele that had no RNA-seq reads at these 2 sites. At the other 2 sites, it was also the methylated allele that was expressed, though in all cases it is difficult to distinguish between allelic dropout and allele-specific expression, especially given the low read counts.

Overrepresentation of Allele-Specific Methylation within and near Imprinting-related Genomic Features

Among genes with an intragenic or promoter CpG that showed evidence of ASM in at least 1, 2, or 3 samples, there was a significant overlap with previously-identified imprinted genes (Table 4.6; $\chi_1^2 \geq 42$; $p = 4.2 \times 10^{-8}$). In addition, we observed a greater-

than-expected overlap between genes with ASM and genes with an imprinting binding motif (Table 4.7; $\chi_1^2 \geq 85$; $p < 10^{-15}$). We also found that CpG sites with ASM were significantly closer to known imprinted genes in B01 ($t = 9.8, p = p = 5.9 \times 10^{-23}$), B04 ($t = 4.1, p = 4 \times 10^{-5}$) and B05 ($t = 3.1, p = 0.002$) whereas there was no significant difference between the distances to imprinting binding motifs between ASM and non-ASM CpGs for any of the 6 samples (Table 4.8).

No Association between Allele-specific Methylation and Allele-specific Expression

Using the heterozygous variants called in the scRRBS data, we were able to distinguish expression between the two alleles in 187 genes using scRNAseq. Among these genes, 150 were mono-allelic and 37 were bi-allelically expressed. We did not observe an association of allele-specific expression with either gene body or gene promoter allele-specific methylation (Supplementary Table 4.3).

DISCUSSION

The primary purpose of this study was to investigate allele-specific methylation (ASM) using scRRBS and explore the resolution at which it can be detected. This is the first study to investigate allele-specific methylation within individual cells and compare it to the allelic expression state in those same cells. We also developed a method that utilizes methylation information across cells to determine the sample-level allelic methylation state at CpG sites. The sample-level ASM we detected was overrepresented in imprinted genes and genes with imprinting binding motifs. We observed an overrepresentation of ASM in previously-identified imprinted genes (Table 4.7) and

imprinting binding motifs (Table 4.8). In addition to the 183 imprinted genes with ASM, 425 other genes also had ASM in 3 or more samples. This result supports the notion that there may be a few hundred more imprinted genes in addition to the ones previously reported. Using Illumina bisulfite sequencing to measure methylation near heterozygous variants, another study also found ASM to be relatively common with the percentage of variants with ASM ranging from 23 to 37% depending on the cell line (Shoemaker et al. 2010). While this could mean that imprinting is a more widespread phenomenon than we currently understand, it is possible that ASM plays a role distinct to imprinting or is simply a non-functional marker in some cases.

However, our ASM results are consistent with what we understand about imprinting in multiple ways. In addition to its overrepresentation in imprinted genes and imprinting binding motifs, there is also a greater-than-expected overlap of ASM across samples. There are some documented cases where the allelic methylation state differs by tissue and/or developmental stage (Weinstein et al. 2010; Elhamamsy 2017), but when investigating an imprinted locus in the same tissue, the allelic methylation state is expected to be the same across individuals. The high pairwise association we observed in our pairwise comparisons of ASM across B01-B06 are consistent with this expectation.

There were four CpGs for which the allelic methylation state and the allelic expression state were observable within the same cells. While we suspected the methylated alleles would not be expressed whereas unmethylated alleles would be expressed, we were not surprised that this was not the case considering the complex relationship between methylation and imprinting.

The mechanisms by which methylation causes imprinting vary from one imprinted cluster to the next, which renders it difficult to ascertain the expected genomic coordinates of allele specific methylation relative to the gene it regulates. According to the chromatin-loop model by Murrell (Murrell, Heeson, and Reik 2004), the unmethylated *H19* DMR and the unmethylated *Igf2* DMR on the maternal allele interact and correspond to *H19* being in an active domain while *Igf2* is in an inactive domain. Conversely, on the paternal allele, the two genes' DMRs are methylated with *Igf2* in the active domain and *H19* in the inactive domain. As mentioned earlier, the non-coding RNA transcript *Air* plays a repressive regulatory role in *Igr2r*. *Air* is anti-sense to and overlaps *Igf2r*, which is one of three genes it represses, along with *Slc22a2* and *Slc22a3* (Sleutels, Zwart, and Barlow 2002). The ICR is in the promoter of *Air*. Thus for genes whose imprinting is carried out similar to *Air* and the gene cluster it represses, it would be challenging to know where the allelic methylation should be located and how it should correspond the allelic expression state for each gene within the cluster without prior information about the specific imprinting dynamics of that cluster.

The complex mechanisms governing imprinting at the *Igf2/H19* imprinting cluster highlight the importance of determining markers that can identify features of imprinting. By mining publicly available data, McEwen and Ferguson-Smith (2010) showed that the histone marks H3K9me3 and H4K20me3 are both enriched at imprinted genes with differentially methylated promoters whereas imprinting control regions are marked by H3K4me3, H3K9me3 and H4K20me3.

We believe the use of single-cell methylation can provide additional information to further interrogate imprinting as we have done here. While bulk-sequencing studies are

often interested in the tissue-level DNA methylation state, it is important to note that at the cellular level, a CpG can take on many methylation states. A CpG has two cytosines on opposite strands, 1 base pair away from each other, each of which can take on a distinct methylation state (Figure 4.4). Since a typical cell has a paternal copy and maternal copy of DNA, there are at least 16 possible methylation states a CpG can have at the cellular-level. The number of possible methylation states goes well beyond 16 if we consider other forms of methylation (i.e. 5-hydroxymethylcytosine, 5-formylcytosine, 5-carboxycytosine) (Zhu et al. 2020). Furthermore, in a post-replication cell, where there is an additional copy of each parental copy, the number of possible DNA methylation states is even greater. In these cells, it can take many hours before the new DNA copy's methylation is restored (Petryk et al. 2020; Charlton et al. 2018) suggesting hemi-methylation is a common phenomenon.

Single-cell bisulfite sequencing in combination with computational tools like the bismark methylation extractor (Krueger and Andrews 2011) facilitates the measurement of a cell's methylation state in both a strand-specific and allele-specific fashion allowing for the differentiation between ASM and hemi-methylation. In our analysis of methylated and unmethylated reads overlapping a nearby heterozygous sites, Only 7 of the CpG sites where we observed possible cellular-level hemi-methylation overlapped the CpG sites with ASM called at the sample level. We believe that the ability for single-cell bisulfite sequencing to interrogate a CpG's cellular-level methylation state will make it an indispensable technology to understand the dynamics of imprinting in future work.

The primary limitations of this work is that both the scRRBS and the scRNAseq data were sparse. Due to low coverage, we were often only able to interrogate ASM in a

small subset of the total cells sequenced. This also renders the comparison of DNA methylation and gene expression challenging as evidenced by the fact that we were only able to compare cellular-level ASM and allelic gene expression in a few instances. This issue can be largely remedied by increasing the number of cells sequenced and we are encouraged to see that this is the trend in the field.

Our results indicate that the true positive rate to detect imprinted regions increases with sample size. When we compared sample-level ASM to the gold standard cellular-level ASM, we found that the true positive rate was higher among CpGs with detected sample-level ASM in a greater number of samples with CpGs. Among CpGs with ASM in ≥3 samples, the true positive rate was 95%. Future studies that collect single-cell methylation data in a greater number of samples have the potential to further increase the power to detect novel imprinted regions.

In this study we utilized cellular-level ASM to detect ASM at the sample-level. When comparing these results to the cellular-level ASM with nearby heterozygous SNPs, all the sample-level ASM also had cellular-level ASM. We also found an overrepresentation of ASM in previously reported imprinted genes and in genes with the 6-mer sequence corresponding to the imprinting binding motif. We demonstrate the capacity of single cell bisulfite sequencing to interrogate imprinting-related ASM, and we hope that future investigations will utilize this technology to further elucidate the complex dynamics of imprinting.

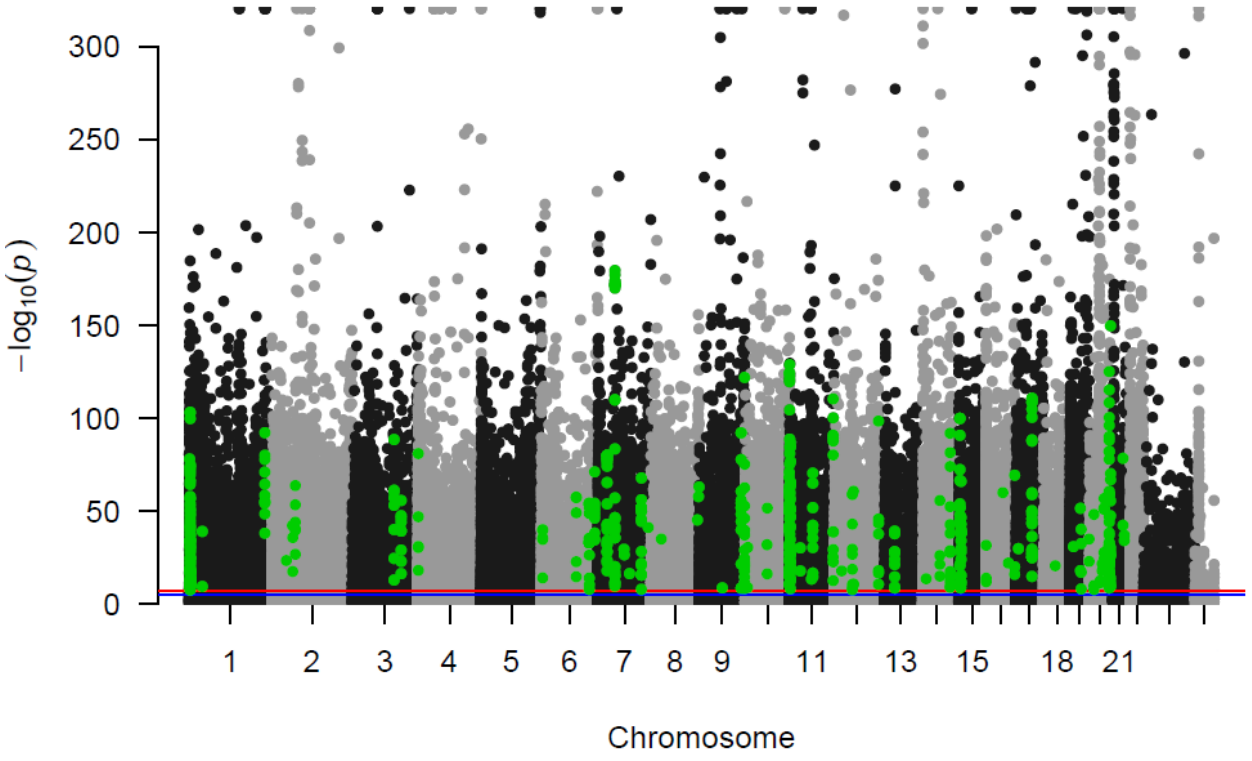


Figure 4.1. Manhattan plot of ASM results for the B04 sample. Green points represent 625 CpGs that overlapped either the gene body or promoter of 109 imprinted genes. The red horizontal line represents the Bonferroni cut-off and the blue horizontal line represents FDR=0.05.

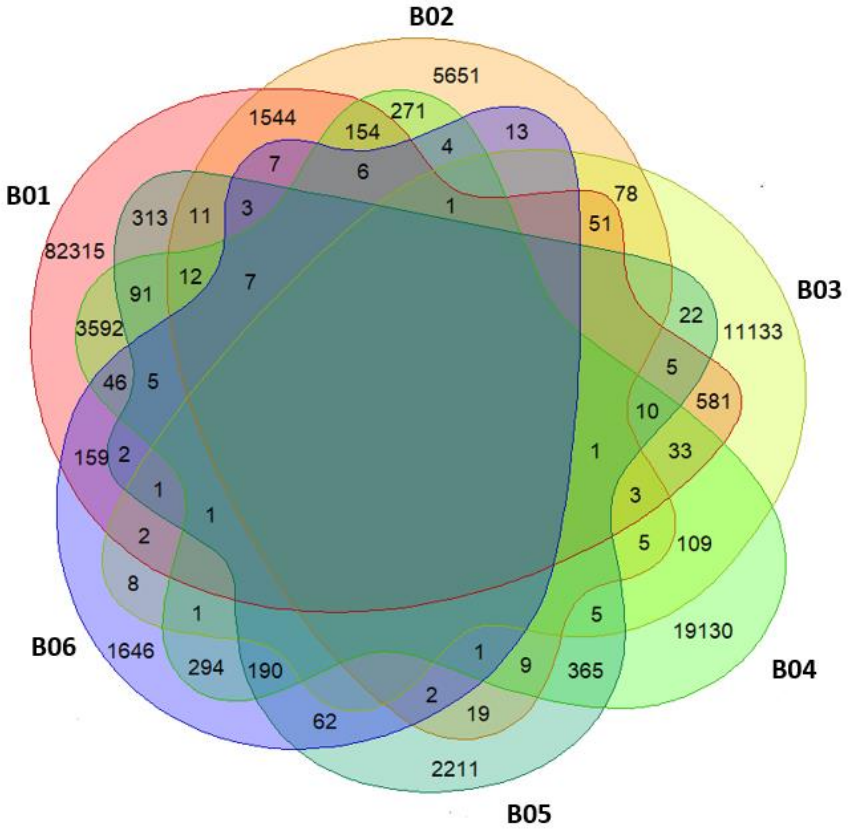


Figure 4.2. Overlap of ASM called in each sample. No CpG site had ASM detected in all 6 samples.

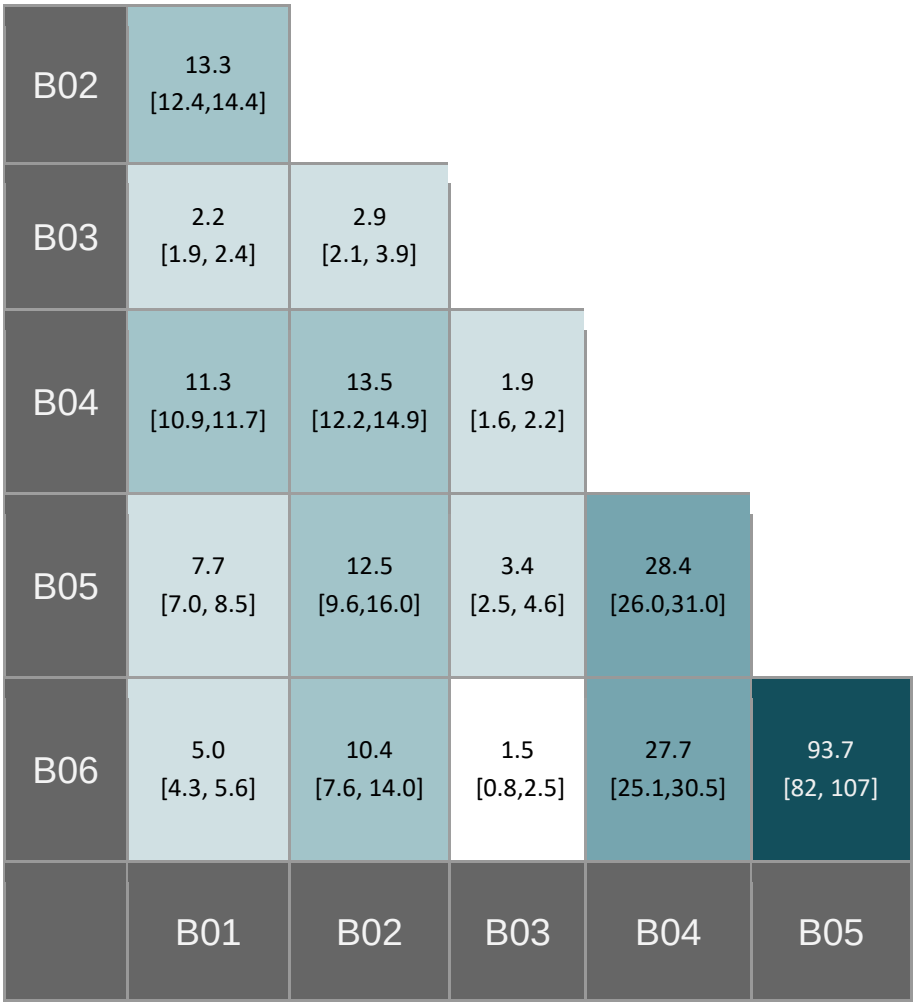


Figure 4.3. Pairwise comparison of ASM across the 6 samples. The strength of association is indicated by color-coding with darker colors corresponding to a stronger association.

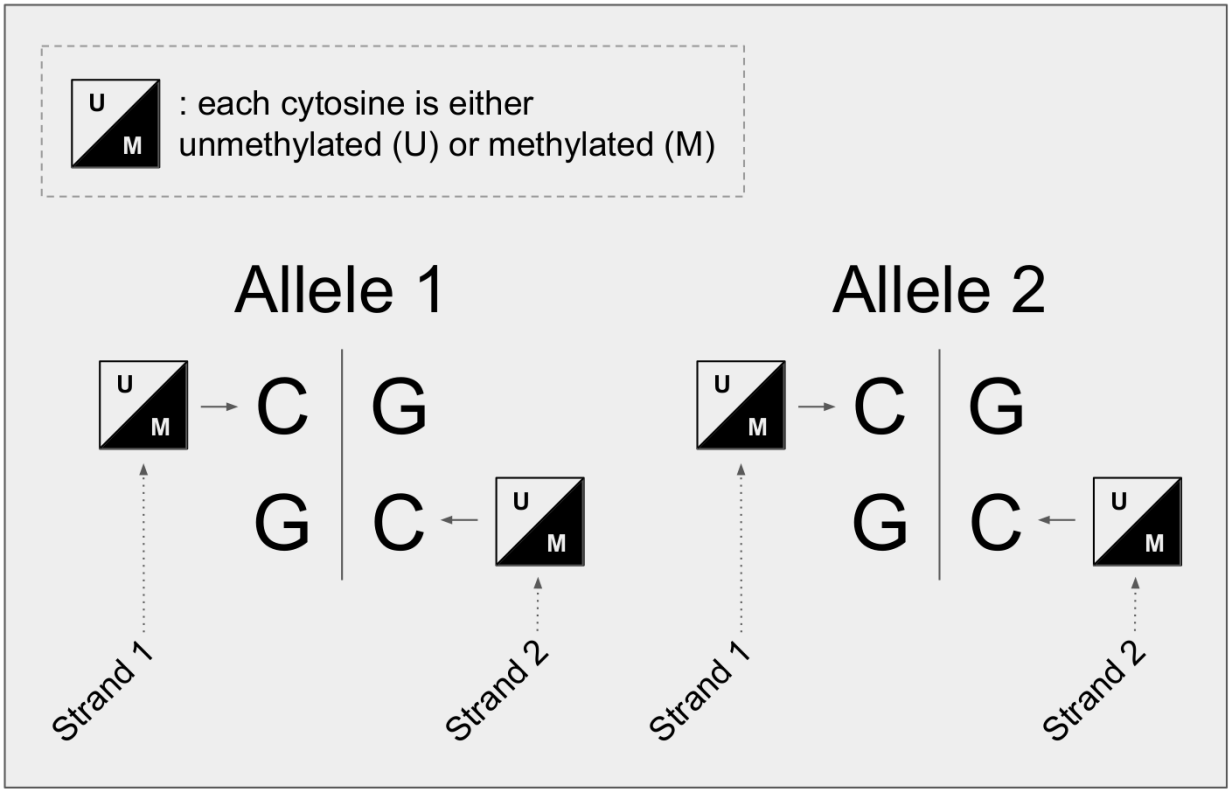


Figure 4.4. Diagram illustrating possible methylation states of a CpG site. For simplicity, we are ignoring intermediate methylation states (i.e. hydroxymethylation). Each allele (maternal and paternal) has a cytosine nucleotide on the forward strand (Strand 1) and the reverse strand (Strand 2), for a total of 4 cytosines. Each of the 4 cytosine has 2 possible methylation states (unmethylated or methylated), for a total of 16 states: 2^4=16.

Sample	# of cells after QC	Total Mapped Reads	Number of CpGs with coverage per cell - median (IQR)	CpGs with coverage in at least one cell	Range of median reads per CpG across cells
B01	92	354,878,229	522,614 (324,834, 743,488)	4,442,170	8 - 91
B02	108	100,863,075	247,196 (192,567, 340,879)	4,225,210	9 - 32
B03	51	73,182,855	163,221 (45,309, 242,833)	3,059,654	15 - 97
B04	169	186,089,626	127,059 (100,930, 149,319)	3,019,292	14 - 37
B05	96	101,886,925	146,594 (125,172, 145,121)	2,668,094	18 - 34
B06	94	85,752,832	98,148 (76,142, 102,974)	2,213,651	26 - 55

Table 4.1. Summary information for each sample

Sample	Variants Called	Overlapping CpGs	Overlapping CpGs w/ ASM	Consistent w/ ASM
B01	6237	991	412	246
B02	5030	200	121	38
B03	2336	80	23	11
B04	4775	185	52	46
B05	2645	26	7	10
B06	1620	12	4	5

Table 4.2. Total variants called, CpGs with reads overlapping the called variants of both unmethylated and methylated states, and the number of those overlapping CpGs with ASM.

B01		Sample-level ASM	
		No ASM	ASM
Observed cellular ASM (Gold Standard)	ASM	333	79
	Consistent w/ ASM	218	28
	Hemi-methylation	38	**6**
	Artifact	45	**7**
	Inconclusive	220	**17**

B02		Sample-level ASM	
		No ASM	ASM
Observed cellular ASM (Gold Standard)	ASM	108	13
	Consistent w/ ASM	37	1
	Hemi-methylation	2	**0**
	Artifact	4	**2**
	Inconclusive	32	**1**

B03		Sample-level ASM	
		No ASM	ASM
Observed cellular ASM (Gold Standard)	ASM	18	5
	Consistent w/ ASM	11	0
	Hemi-methylation	5	**1**
	Artifact	6	**0**
	Inconclusive	32	**2**

B04		Sample-level ASM	
		No ASM	ASM
Observed cellular ASM (Gold Standard)	ASM	40	12
	Consistent w/ ASM	46	0
	Hemi-methylation	5	**0**
	Artifact	9	**0**
	Inconclusive	73	**0**

B05		Sample-level ASM	
		No ASM	ASM
Observed cellular ASM (Gold Standard)	ASM	7	0
	Consistent w/ ASM	10	0
	Hemi-methylation	1	**0**
	Artifact	0	**0**
	Inconclusive	8	**0**

B06		Sample-level ASM	
		No ASM	ASM
Observed cellular ASM (Gold Standard)	ASM	0	1
	Consistent w/ ASM	5	0
	Hemi-methylation	0	**0**
	Artifact	0	**0**
	Inconclusive	3	**0**

Table 4.3. Sample-level ASM in each sample compared to the gold standard cellular-level ASM.

Number of samples with ASM	True positive rate (compared to gold standard)
≥1	89%
≥2	92%
≥3	95%

Table 4.4. True positive rates among CpGs with sample-level ASM in ≥1, ≥2, and ≥3 samples. A site was considered to be "consistent with ASM" based on the gold standard if it had at least one read overlapping one allele of the heterozygous site and at least one read overlapping the other allele. Sites considered gold standard "ASM" needed at least 4 reads overlapping each allele. For the comparison of sample-level ASM and gold standard ASM, we considered it a true positive if sample-level ASM was also "ASM" or "consistent with ASM", and a false positive (boldface numbers) if a CpG with sample-level ASM showed evidence of hemi-methylation or an artifact (such as doublets or non-specific alignment).

Cell Index	CpG	Variant Position		Genotype Call (from scRRBS)		Methylation State		RNAseq Allele Counts				Gene(s) overlapping variant
		Chr	Position	Allele 1	Allele 2	Allele 1	Allele 2	A	C	G	T	
B04.159	9:93084582-3	9	93084584	A	T	unmethylated	methylated	0	0	0	5	SUSD3
B04.114	10:129729433-4	10	129729441	T	A	methylated	methylated	7	1	0	0	MGMT
B04.114	19:49447586-7	19	49447581	A	C	unknown	methylated	0	5	0	0	PIH1D1
B04.159	22:47116957-8	22	47116970	A	T	unmethylated	methylated	0	0	0	4	TBC1D22A, FP325331.1

Table 4.5. Four variants with matching RNAseq data and RRBS data at the cellular level, which overlap the CpG sites in Table S2. In row two one of the RNA counts is a cytosine, which matches allele 1 from the scRRBS data if we assume that it was originally a cytosine that went through bisulfite conversion to a thymine. Row 2 corresponds to rs879280, 76% C, 24% A (1000 Genomes Project Consortium et al. 2015).

	Genes (Not Imprinted)	Genes (Imprinted)	Evidence for enrichment
Genes without ASM ≥1 Samples	17,335	37	*OR = 3.5 [2.4, 5.1]*
Genes with ASM ≥1 Samples	19,856	147	$\chi_1^2 = 50$ $p < 10^{-11}$
Genes without ASM ≥2 Samples	33,095	127	*OR = 3.7 [2.6, 5.0]*
Genes with ASM ≥2 Samples	4,061	57	$\chi_1^2 = 73$ $p < 10^{-15}$
Genes without ASM ≥3 Samples	36,724	171	*OR = 6.0 [3.0, 11]*
Genes with ASM ≥3 Samples	425	12	$\chi_1^2 = 42$ $p < 10^{-9}$

Table 4.6. Genes with ASM in ≥1 (top 2 rows), ≥2 (middle 2 rows), and ≥3 (bottom 2 rows) samples are overrepresented among previously-reported imprinted genes.

	Genes w/o binding motif	Genes w/ binding motif	Evidence for enrichment
Genes without ASM ≥1 Samples	9,486	7,921	*OR = 3.8 [3.6, 4.0]*
Genes with ASM ≥1 Samples	4,764	15,268	$\chi_1^2 = 3725$ $p < 10^{-15}$
Genes without ASM ≥2 Samples	13,560	19,703	*OR = 3.6 [3.3, 3.9]*
Genes with ASM ≥2 Samples	659	3,465	$\chi_1^2 = 955$ $p < 10^{-15}$
Genes without ASM ≥3 Samples	14,140	22,797	*OR = 3.1 [2.4, 4.0]*
Genes with ASM ≥3 Samples	73	365	$\chi_1^2 = 85$ $p < 10^{-15}$

Table 4.7. Genes with ASM in ≥1 (top 2 rows), ≥2 (middle 2 rows), and ≥3 (bottom 2 rows) samples are overrepresented among genes with imprinting binding motifs.

Sample	Difference in Mean Distance to nearest Imprinted Gene (ASM vs. All CpGs)				Difference in Mean Distance to nearest Imprinting Binding Motif (ASM vs. All CpGs)			
	Median Distance to ASM	Median Distance to non-ASM	T-statistic	P-value	Median Distance to ASM	Median Distance to non-ASM	T-statistic	P-value
B01	4.965 Mb	5.117 Mb	11.3	8.2E-30	4.483 kb	4.499 kb	1.28	0.20
B02	5.293 Mb	5.125 Mb	0.09	0.93	4.401 kb	4.499 kb	0.81	0.42
B03	5.207 Mb	5.157 Mb	-1.00	0.32	4.657 kb	4.495 kb	-1.46	0.14
B04	5.140 Mb	5.136 Mb	4.0	5.6E-05	4.545 kb	4.510 kb	0.86	0.39
B05	4,717 Mb	5.119 Mb	3.2	0.001	4.485 kb	4.501 kb	-0.63	0.53
B06	4.850 Mb	5.106 Mb	0.05	0.96	4,296 kb	4,494 kb	0.18	0.86

Table 4.8. Mean distances to the nearest imprinted gene of ASM CpGs compared to all CpGs for each sample (Column 2). Mean distances to the nearest imprinting binding motif of ASM CpGs compared to all CpGs for each sample (Column 3). T-statistics and corresponding p-values are based on log-transformed distances.

REFERENCES

1000 Genomes Project Consortium, Adam Auton, Lisa D. Brooks, Richard M. Durbin, Erik P. Garrison, Hyun Min Kang, Jan O. Korbel, et al. 2015. “A Global Reference for Human Genetic Variation.” *Nature* 526 (7571): 68–74.

Butler, Merlin G. 2009. “Genomic Imprinting Disorders in Humans: A Mini-Review.” *Journal of Assisted Reproduction and Genetics* 26 (9-10): 477–86.

Charlton, Jocelyn, Timothy L. Downing, Zachary D. Smith, Hongcang Gu, Kendell Clement, Ramona Pop, Veronika Akopian, et al. 2018. “Global Delay in Nascent Strand DNA Methylation.” *Nature Structural & Molecular Biology* 25 (4): 327–32.

Edwards, Carol A., and Anne C. Ferguson-Smith. 2007. “Mechanisms Regulating Imprinted Genes in Clusters.” *Current Opinion in Cell Biology* 19 (3): 281–89.

Elhamamsy, Amr Rafat. 2017. “Role of DNA Methylation in Imprinting Disorders: An Updated Review.” *Journal of Assisted Reproduction and Genetics* 34 (5): 549–62.

Feinberg, Andrew P., Rolf Ohlsson, and Steven Henikoff. 2006. “The Epigenetic Progenitor Origin of Human Cancer.” *Nature Reviews Genetics*. https://doi.org/10.1038/nrg1748.

Ferguson-Smith, Anne C. 2011. “Genomic Imprinting: The Emergence of an Epigenetic Paradigm.” *Nature Reviews. Genetics* 12 (8): 565–75.

Fink, Katja, Raphael Zellweger, Jacqueline Weber, Nataly Manjarrez-Orduno, Martin Holdener, Beatrice M. Senn, Hans Hengartner, Rolf M. Zinkernagel, and Andrew J. Macpherson. 2008. “Long-Term Maternal Imprinting of the Specific B Cell Repertoire by Maternal Antibodies.” *European Journal of Immunology* 38 (1): 90–101.

Gaiti, Federico, Ronan Chaligne, Hongcang Gu, Ryan M. Brand, Steven Kothen-Hill, Rafael C. Schulman, Kirill Grigorev, et al. 2019. “Epigenetic Evolution and Lineage Histories of Chronic Lymphocytic Leukaemia.” *Nature* 569 (7757): 576–80.

Geiger, Hartmut, Gerald de Haan, and M. Carolina Florian. 2013. “The Ageing Haematopoietic Stem Cell Compartment.” *Nature Reviews. Immunology* 13 (5): 376–89.

Holm, Teresa M., Laurie Jackson-Grusby, Tobias Brambrink, Yasuhiro Yamada, William M. Rideout 3rd, and Rudolf Jaenisch. 2005. “Global Loss of Imprinting Leads to Widespread Tumorigenesis in Adult Mice.” *Cancer Cell* 8 (4): 275–85.

Imbeault, Michaël, Pierre-Yves Helleboid, and Didier Trono. 2017. “KRAB Zinc-Finger Proteins Contribute to the Evolution of Gene Regulatory Networks.” *Nature* 543 (7646): 550–54.

Jirtle, James. 2012. “Imprinted Genes: Human.” Geneimprint. 2012. http://www.geneimprint.com/site/genes-by-species.Homo+sapiens.

Krueger, Felix. 2015. “Trim Galore.” *A Wrapper Tool around Cutadapt and FastQC to Consistently Apply Quality and Adapter Trimming to FastQ Files* 516: 517.

Krueger, Felix, and Simon R. Andrews. 2011. “Bismark: A Flexible Aligner and Methylation Caller for Bisulfite-Seq Applications.” *Bioinformatics* 27 (11): 1571–72.

Li, Heng, Bob Handsaker, Alec Wysoker, Tim Fennell, Jue Ruan, Nils Homer, Gabor

Marth, Goncalo Abecasis, Richard Durbin, and 1000 Genome Project Data Processing Subgroup. 2009. “The Sequence Alignment/Map Format and SAMtools.” *Bioinformatics* 25 (16): 2078–79.

Lindenbaum, Pierre. 2015. “JVarkit: Java-Based Utilities for Bioinformatics,” May. https://doi.org/10.6084/m9.figshare.1425030.v1.

Li, Yufeng, and Hiroyuki Sasaki. 2011. “Genomic Imprinting in Mammals: Its Life Cycle, Molecular Mechanisms and Reprogramming.” *Cell Research* 21 (3): 466–73.

McEwen, Kirsten R., and Anne C. Ferguson-Smith. 2010. “Distinguishing Epigenetic Marks of Developmental and Imprinting Regulation.” *Epigenetics & Chromatin* 3 (1): 2.

Murrell, Adele, Sarah Heeson, and Wolf Reik. 2004. “Interaction between Differentially Methylated Regions Partitions the Imprinted Genes Igf2 and H19 into Parent-Specific Chromatin Loops.” *Nature Genetics* 36 (8): 889–93.

Petryk, Nataliya, Sebastian Bultmann, Till Bartke, and Pierre-Antoine Defossez. 2020. “Staying True to Yourself: Mechanisms of DNA Methylation Maintenance in Mammals.” *Nucleic Acids Research*, December. https://doi.org/10.1093/nar/gkaa1154.

Reig, German, and Miguel L. Concha. 2012. “Genomic Imprinting and Embryonic Development.” *International Journal of Morphology = Revista Internacional de Morfologia* 30 (4): 1453–57.

Riso, Vincenzo, Marco Cammisa, Harpreet Kukreja, Zahra Anvar, Gaetano Verde, Angela Sparago, Basilia Acurzio, et al. 2016. “ZFP57 Maintains the Parent-of-Origin-Specific Expression of the Imprinted Genes and Differentially Affects Non-Imprinted Targets in Mouse Embryonic Stem Cells.” *Nucleic Acids Research* 44 (17): 8165–78.

Serrano-Lopez, Juana, and Jose A. Cancelas. 2016. “Mom Knows Best: Imprinted Control of Hematopoietic Stem Cell Quiescence.” *Cell Stem Cell*.

Shoemaker, Robert, Jie Deng, Wei Wang, and Kun Zhang. 2010. “Allele-Specific Methylation Is Prevalent and Is Contributed by CpG-SNPs in the Human Genome.” *Genome Research* 20 (7): 883–89.

Sleutels, Frank, Ronald Zwart, and Denise P. Barlow. 2002. “The Non-Coding Air RNA Is Required for Silencing Autosomal Imprinted Genes.” *Nature* 415 (6873): 810–13.

Takahashi, Nozomi, Andrea Coluccio, Christian W. Thorball, Evarist Planet, Hui Shi, Sandra Offner, Priscilla Turelli, Michael Imbeault, Anne C. Ferguson-Smith, and Didier Trono. 2019. “ZNF445 Is a Primary Regulator of Genomic Imprinting.” *Genes & Development* 33 (1-2): 49–54.

Weinstein, L. S., T. Xie, A. Qasem, J. Wang, and M. Chen. 2010. “The Role of GNAS and Other Imprinted Genes in the Development of Obesity.” *International Journal of Obesity* 34 (1): 6–17.

Yates, Andrew D., Premanand Achuthan, Wasiu Akanni, James Allen, Jamie Allen, Jorge Alvarez-Jarreta, M. Ridwan Amode, et al. 2019. “Ensembl 2020.” *Nucleic Acids Research* 48 (D1): D682–88.

Zhu, Hongming, He Zhu, Miao Tian, Dongying Wang, Jiaxing He, and Tianmin Xu. 2020. “DNA Methylation and Hydroxymethylation in Cervical Cancer: Diagnosis, Prognosis and Treatment.” *Frontiers in Genetics* 11 (April): 347.

SUPPLEMENT

Odds ratio: 13.3 [12.4, 14.4]		B02	
		No ASM	ASM
B01	No ASM	2,421,586	3,257
	ASM	50,380	904

$\chi_1^2 = 7{,}928;\ p < 10^{-15}$

Odds ratio: 2.2 [1.9, 2.4]		B03	
		No ASM	ASM
B01	No ASM	2,078,908	8,020
	ASM	45,349	378

$\chi_1^2 = 222;\ p < 10^{-15}$

Odds ratio: 11.3 [10.9,11.7]		B04	
		No ASM	ASM
B01	No ASM	2,944,332	20,384
	ASM	50,614	3,962

$\chi_1^2 = 28{,}930;\ p < 10^{-15}$

Odds ratio: 7.7 [7.0, 8.5]		B05	
		No ASM	ASM
B01	No ASM	2,421,997	2846
	ASM	50,822	462

$\chi_1^2 = 2{,}305;\ p < 10^{-15}$

Odds ratio: 5.0 [4.3, 5.6]		B06	
		No ASM	ASM
B01	No ASM	2,084,733	2195
	ASM	45489	238

$\chi_1^2 = 673;\ p < 10^{-15}$

Odds ratio: 2.9 [2.1, 3.9]		B03	
		No ASM	ASM
B02	No ASM	2,039,059	81903
	ASM	3,751	44

$\chi_1^2 = 52.7;\ p < 10^{-12}$

Odds ratio: 13.5 [12.2, 14.9]		B04	
		No ASM	ASM
B02	No ASM	2,448,744	23,222
	ASM	3,688	473

$\chi_1^2 = 4{,}755;\ p < 10^{-15}$

Odds ratio: 12.5 [9.6, 16.0]		B05	
		No ASM	ASM
B02	No ASM	2,660,544	3,283
	ASM	4,202	65

$\chi_1^2 = 655;\ p < 10^{-15}$

Odds ratio: 10.4 [7.6, 14.0]		B06	
		No ASM	ASM
B02	No ASM	2,044,952	2,297
	ASM	3,751	44

$\chi_1^2 = 355;\ p < 10^{-15}$

Odds ratio: 1.9 [1.6, 2.2]		B04	
		No ASM	ASM
B03	No ASM	2,101,310	22,947
	ASM	8,231	167

$\chi_1^2 = 64;\ p < 10^{-15}$

Odds ratio: 3.4 [2.5, 4.6]		B05	
		No ASM	ASM
B03	No ASM	2,039,689	3,121
	ASM	8,191	43

$\chi_1^2 = 70;\ p < 10^{-15}$

Odds ratio: 1.5 [0.8, 2.5]		B06	
		No ASM	ASM
B03	No ASM	2,202,699	2,447
	ASM	8,491	14

$\chi_1^2 = 1.7;\ p = 0.19$

Odds ratio: 28.4 [26.0, 31.0]		B05	
		No ASM	ASM
B04	No ASM	2,449,821	2,611
	ASM	22,998	697

$\chi_1^2 = 14{,}117;\ p < 10^{-15}$

Odds ratio: 27.7 [25.1, 30.5]		B06	
		No ASM	ASM
B04	No ASM	2,107,664	1,877
	ASM	22,558	556

$\chi_1^2 = 10{,}746;\ p < 10^{-15}$

Odds ratio: 93.7 [82, 107]		B06	
		No ASM	ASM
B05	No ASM	2,045,813	2,067
	ASM	2,890	274

$\chi_1^2 = 20{,}224;\ p < 10^{-15}$

Supplementary Table 4.1. Contingency Tables of pair-wise associations of allele-specific methylation across samples.

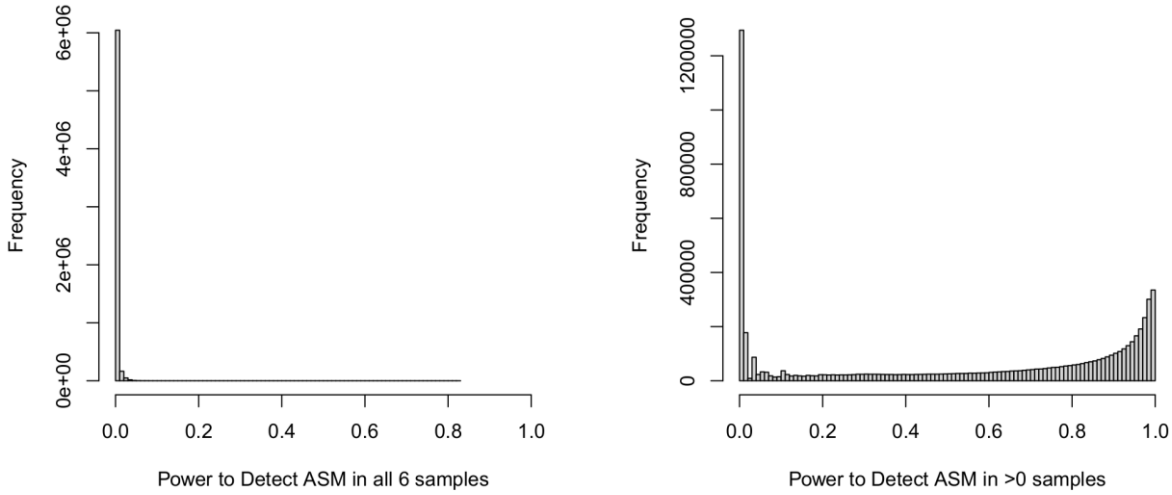
Supplementary Table 4.2.csv

Supplementary Table 4.2. The CpGs in B01-B06 sample with methylated and unmethylated reads overlapping both a CpG site and a variant.

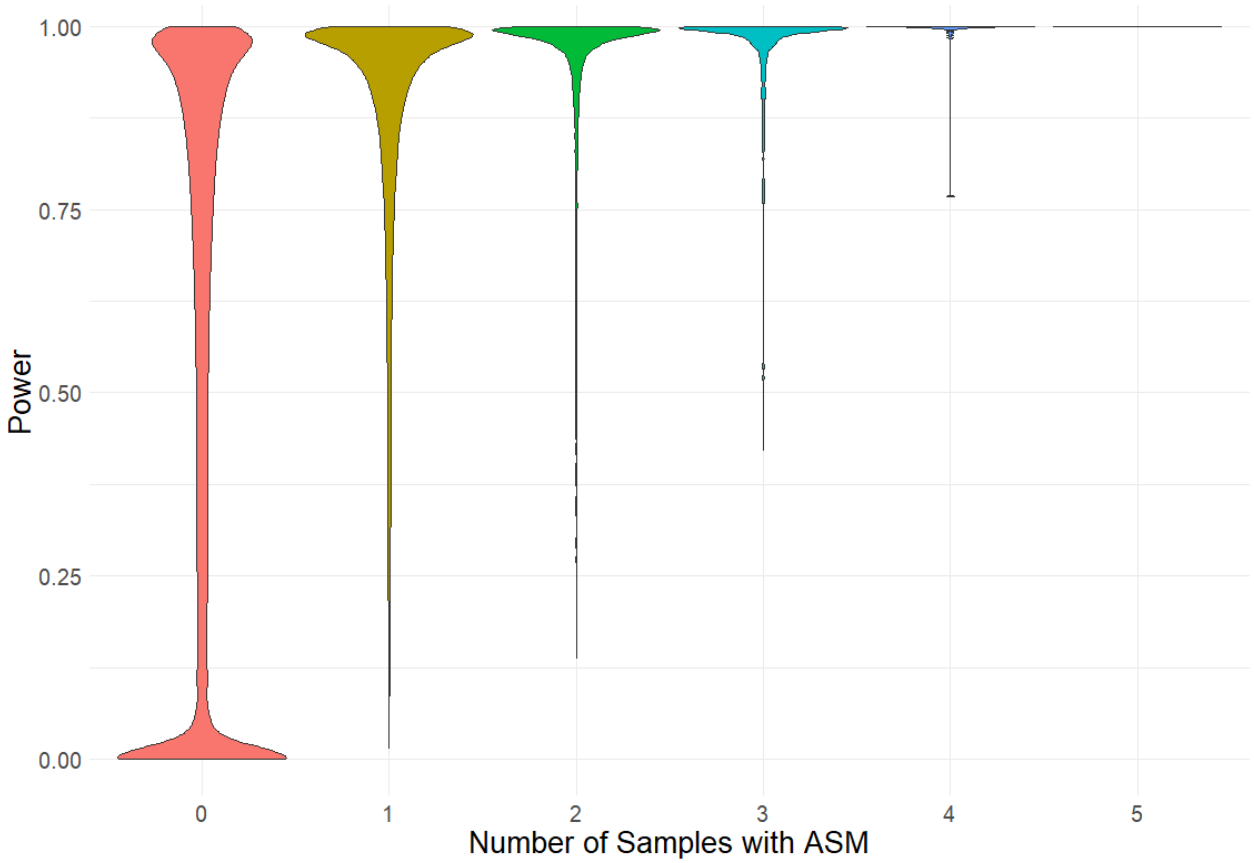
Gene body	**Mono-allelic expression**	**Bi-allelic expression**
No ASM	79	19
ASM	71	18

Promoter	**Mono-allelic expression**	**Bi-allelic expression**
No ASM	125	31
ASM	13	4

Supplementary Table 4.3. Among genes where mono- and bi-allelic expression was detected, there was no association of number of alleles expressed with ASM in either the gene body or promoter.



Supplementary Figure 4.1. Distribution of power to detect ASM in all 6 samples (left); Distribution of power to detect ASM in at least 1 sample (right), for all CpGs included in the study.



Supplementary Figure 4.2. Distribution of power to detect ASM for CpG sites with ASM detected in 0-5 samples.

Chapter 5: Conclusion

DNA methylation is an epigenetic modification that plays a gene regulatory role. In chapter 2, I reviewed the literature regarding our current understanding on the roles DNA methylation and DNA hydroxymethylation play in the decline of the immune system with age. I found that numerous studies have observed associations of DNA methylation and DNA hydroxymethylation with age and many processes involved in immune system decline. Although most of these studies investigated non-causal associations of these epigenetic modifications with immunosenescence, or related processes, some of these studies support a causal role of these epigenetic modifications.

Roles of these epigenetic modifications are not fully characterized in immunosenescence, aging, and other biological processes. The lack of characterization of these epigenetic modifications is partially because the discovery of DNA hydroxymethylation is recent (Song, Yi, and He 2012; Terragni et al. 2012; Höbartner 2011; Szwagierczak et al. 2010; Booth et al. 2012). In large part, this lack of characterization provides the motivation to further investigate age-related DNA hydroxymethylation and its relationship with gene expression in immune cells, which is the focus of chapter 3.

In chapter 3, I investigated age-related DNAhm and age-related gene expression in human peripheral blood mononuclear cells, which comprise both myeloid and lymphoid lineages of the immune system. I found that the abundance of DNAhm was higher in enhancers and exons compared to gene bodies and genome-wide. In spite of the fact that the abundance of DNAhm is relatively low in blood (Godderis et al., 2015), I observed

thousands of age-differentially hydroxymethylated regions (DhmRs) with an overrepresentation of these DhmRs in gene bodies and enhancers. When comparing the DhmRs to age-differentially expressed genes (DEGs), I observed directional consistency between DhmRs and DEGs in 159 of 160 DhmRs overlapping DEGs. This contrasts the repressive gene regulatory role that DNAm plays, and is suggestive of a distinct role for DNAhm. Unlike DNAm, DNAhm does not bind to known transcriptional repressors (MBD1, MBD2, and MBD4) (Boyes and Bird, 1991; Jin et al., 2010; Nan et al., 1998; Wade, 2001). Further, the directional consistency between age-related DNAhm and age-related DNAhm I observed is consistent with previous work that found an association between DNAhm and cis-gene expression (Marco et al. 2016; J. Zhao et al. 2017; Colquitt et al. 2013).

Gene enrichment analyses were also supportive of a role for DNAhm in immune system function. Gene bodies with age-related increases in DNAhm were enriched for myeloid cell activation involved in immune system response, neutrophil function, and neutrophil degranulation. Aside from age-related DNAhm, we also observed that the genes with the highest levels of DNAhm were enriched for Fc-gamma receptor signaling pathway involved in phagocytosis as well as several other GO terms indicative of involvement in immune response, which is consistent with the age-related impairment of phagocytosis observed in human monocytes (Hearps et al. 2012). Overall, these results are consistent with the role of age-related DNAhm in immune system processes within peripheral blood.

In addition to the fact that DNAhm has only recently been discovered, these epigenetic modifications have not been fully characterized for another reason, namely,

the lack of resolution of bulk sequencing technologies. The vast majority of the work characterizing these epigenetic modifications in immune system processes and immunosenescence have relied on targeted approaches and bulk sequencing technologies, but this technology lacks the resolution necessary to fully characterize the complexity of these epigenetic modifications, their effects on gene expression, and ability to shape biological processes. For example, a well-known hallmark of immunosenescence is lineage skewing of hematopoietic stem cells (HSCs) (Geiger, de Haan, and Carolina Florian 2013), which results from a reduced self-renewal capacity of this lineage. Both DNAm (Challen et al. 2011) and DNAhm (Moran-Crusio et al. 2011; Ko et al. 2010) play a causal role in HSC lineage skewing, which highlights the importance of these epigenetic modifications in immunosenescence. It is also known that HSC self-renewal capacity is maintained by an imprinted locus (Serrano-Lopez and Cancelas 2016). Allele-specific methylation (ASM) is the underlying mechanism resulting in imprinting. In chapter 4, I describe why single-cell methylation sequencing is better-suited than bulk-sequencing approaches to detect ASM, and subsequently investigate the ability of single-cell reduced representation bisulfite sequencing to detect novel imprinted loci.

I was able to detect measurements of methylation within individual cells and use this information to detect sample-level ASM in all six of our healthy samples of B cells. Many of these CpGs with ASM were detected in multiple samples. To further validate these results, I called heterozygous sites, which I used to quantify methylated and unmethylated reads overlapping each allele. One of the unprecedented utilities of single cell methylation

data is that, at these CpG sites with nearby heterozygous sites, it is possible to observe both methylation states (methylated and unmethylated) and distinguish which strand and which allele it originated from. Because strand, allele, and methylation information are observable at the cellular-level, it is possible to distinguish hemi-methylation from allele-specific methylation. Therefore, these CpGs with nearby heterozygous sites provide a gold standard, which can be used to validate ASM detected at the sample level. Remarkably, 95% of the CpG sites near heterozygous sites with sample-level ASM in ≥3 samples also had cellular-level ASM.

In addition to validating that our results were truly detecting ASM, we also investigated whether they met other expectations of imprinting. Firstly, imprinting is typically observable at the same regions across samples, and our results are consistent with this: we observed a greater-than-expected overlap of ASM across samples. We also observed an overrepresentation of sample-level ASM within previously reported imprinted genes and imprinting binding motifs. Overall, these results demonstrate that single cell methylation sequencing has the capacity to discover novel imprints.

DNAm and DNAhm are epigenetic modifications with a complex relationship with many biological processes. The extent to which these epigenetic modifications regulate gene expression and subsequently shape biological processes vs. simple marking these processes is an active area of research. In chapter 2, I reviewed the role of these

epigenetic modifications in one of these biological processes, namely, immunosenescence. One of the conclusions of chapter 2 is that our understanding of DNAhm lags behind that of DNAm, which provided motivation to pursue the research in chapter 3, where I investigated the relationship between age-related DNAhm and age-related gene expression in human peripheral blood. Despite the low levels of DNAhm I observed in blood, the discovery in chapter 3 suggests that DNAhm may play a role in aging and immunosenescence and is a worthy area of future investigations. In chapter 4, I explore the ability of the new technology to measure DNA methylation at the single cell level and its ability to detect imprinting. I show that single cell methylation sequencing is a powerful tool to detect novel imprints. The results presented here provide a basis to further understand the complex interplay among epigenetic modifications, such as DNAm and DNAhm, gene regulation, and the ultimate consequence on biological processes.

References

Booth, Michael J., Miguel R. Branco, Gabriella Ficz, David Oxley, Felix Krueger, Wolf Reik, and Shankar Balasubramanian. 2012. “Quantitative Sequencing of 5-Methylcytosine and 5-Hydroxymethylcytosine at Single-Base Resolution.” *Science* 336 (6083): 934–37.

Challen, Grant A., Deqiang Sun, Mira Jeong, Min Luo, Jaroslav Jelinek, Jonathan S. Berg, Christoph Bock, et al. 2011. “Dnmt3a Is Essential for Hematopoietic Stem Cell Differentiation.” *Nature Genetics* 44 (1): 23–31.

Geiger, Hartmut, Gerald de Haan, and M. Carolina Florian. 2013. “The Ageing Haematopoietic Stem Cell Compartment.” *Nature Reviews. Immunology* 13 (5): 376–89.

Höbartner, Claudia. 2011. “Enzymatic Labeling of 5-Hydroxymethylcytosine in DNA.” *Angewandte Chemie* 50 (19): 4268–70.

Ko, Myunggon, Yun Huang, Anna M. Jankowska, Utz J. Pape, Mamta Tahiliani, Hozefa S. Bandukwala, Jungeun An, et al. 2010. “Impaired Hydroxylation of 5-Methylcytosine in Myeloid Cancers with Mutant TET2.” *Nature* 468 (7325): 839–43.

Moran-Crusio, Kelly, Linsey Reavie, Alan Shih, Omar Abdel-Wahab, Delphine Ndiaye-Lobry, Camille Lobry, Maria E. Figueroa, et al. 2011. “Tet2 Loss Leads to Increased Hematopoietic Stem Cell Self-Renewal and Myeloid Transformation.” *Cancer Cell* 20 (1): 11–24.

Serrano-Lopez, Juana, and Jose A. Cancelas. 2016. “Mom Knows Best: Imprinted Control of Hematopoietic Stem Cell Quiescence.” *Cell Stem Cell*.

Song, Chun-Xiao, Chengqi Yi, and Chuan He. 2012. “Mapping Recently Identified Nucleotide Variants in the Genome and Transcriptome.” *Nature Biotechnology* 30 (11): 1107–16.

Szwagierczak, Aleksandra, Sebastian Bultmann, Christine S. Schmidt, Fabio Spada, and Heinrich Leonhardt. 2010. “Sensitive Enzymatic Quantification of 5-Hydroxymethylcytosine in Genomic DNA.” *Nucleic Acids Research* 38 (19): e181.

Terragni, Jolyon, Jurate Bitinaite, Yu Zheng, and Sriharsa Pradhan. 2012. “Biochemical Characterization of Recombinant β-Glucosyltransferase and Analysis of Global 5-Hydroxymethylcytosine in Unique Genomes.” *Biochemistry* 51 (5): 1009–19.