Primate Skeletal Epigenetics:

Evolutionary Implications of DNA Methylation Patterns

in the Skeletal Tissues of Human and Nonhuman Primates

by

Genevieve Housman

A Dissertation Presented in Partial Fulfillment
of the Requirements for the Degree
Doctor of Philosophy

Approved April 2017 by the
Graduate Supervisory Committee:

Anne Stone, Chair
Ellen Quillen
Kenro Kusumi
Christopher Stojanowski

ARIZONA STATE UNIVERSITY

May 2017

ABSTRACT

Within the primate lineage, skeletal traits that contribute to inter-specific anatomical variation and enable varied niche occupations and forms of locomotion are often described as the result of environmental adaptations. However, skeletal phenotypes are more accurately defined as complex traits, and environmental, genetic, and epigenetic mechanisms, such as DNA methylation which regulates gene expression, all contribute to these phenotypes. Nevertheless, skeletal complexity in relation to epigenetic variation has not been assessed across the primate order. In order to gain a complete understanding of the evolution of skeletal phenotypes across primates, it is necessary to study skeletal epigenetics in primates. This study attempts to fill this gap by identifying intra- and inter-specific variation in primate skeletal tissue methylation in order to test whether specific features of skeletal form are related to specific variations in methylation. Specifically, methylation arrays and gene-specific methylation sequencing are used to identify DNA methylation patterns in femoral trabecular bone and cartilage of several nonhuman primate species. Samples include baboons (*Papio spp.*), macaques (*Macaca mulatta*), vervets (*Chlorocebus aethiops*), chimpanzees (*Pan troglodytes*), and marmosets (*Callithrix jacchus*), and the efficiencies of these methods are validated in each taxon. Within one nonhuman primate species (baboons), intra-specific variations in methylation patterns are identified across a range of comparative levels, including skeletal tissue differences (bone vs. cartilage), age cohort differences (adults vs. juveniles), and skeletal disease state differences (osteoarthritic vs. healthy), and some of the identified patterns are evolutionarily conserved with those known in humans. Additionally, in all nonhuman primate species, intra-specific methylation variation in association with nonpathological

femur morphologies is assessed. Lastly, inter-specific changes in methylation are evaluated among all nonhuman primate taxa and used to provide a phylogenetic framework for methylation changes previously identified in the hominin lineage. Overall, findings from this work reveal how skeletal DNA methylation patterns vary within and among primate species and relate to skeletal phenotypes, and together they inform our understanding of epigenetic regulation and complex skeletal trait evolution in primates.

ACKNOWLEDGMENTS

I would like to thank my dissertation committee for their assistance during the development of my research, support during my data collection, and feedback during the final synthesis and writing of my dissertation. I want to specifically thank Lorena Havill for allowing me to initiate the baboon skeletal DNA methylation pilot project that started it all. Additionally, I am especially grateful to Ellen Quillen who granted me access to her lab and samples at the Texas Biomedical Research Institute and provided extremely helpful guidance during my data collection and analyses. I am also appreciative of Christopher Stojanowski and Kenro Kusumi for their helpful feedback and conversations. Lastly, I would like to express my immense gratitude to Anne C. Stone, my dissertation chair, for her constant support and guidance during my graduate student career.

I would also like to express my gratitude to the many researchers at the Texas Biomedical Research Institute that granted me access to their facilities, resources, and samples. Additionally, this research would not have been possible without the skeletal remains of several nonhuman primates housed at the Texas Biomedical Research Institute. Finally, I would like to thank all the funding agencies that enabled this research – the Max and Minnie Tomerlin Voelcker Foundation, the William and Ella Owens Foundation for Biomedical Research, the Southwest National Primate Research Center, Sigma Xi, the Leakey Foundation, the Wenner-Gren Foundation [Gr. 9310], the Nacey Maggioncalda Foundation, the International Primatological Society, the Center for Evolution and Medicine at ASU, the Graduate Research and Support Program at ASU, and the School of Human Evolution and Social Change at ASU.

I would also like to thank my fellow labmates in the Molecular Anthropology Laboratory and specifically María A. Nieves-Colón and Tanvi Honap for their assistance, encouragement, and friendship while we navigated through the highs and lows of academia together, taught ourselves to be bioinformaticians, and pushed each other to be amazing scientists. I am also grateful to my other fellow doctoral students for their willingness to help and support me regardless of my divergent epigenetic interests. Specifically, to John Rowan and Kathleen Paul, thank you for your friendship and for making sure I stayed somewhat involved in social activities, and to Halszka Glowacka, thank you for being an amazing roommate and helping me to find nonacademic activities that kept me healthy and strong throughout my dissertation.

Last but certainly not least, I am eternally grateful to my closest friends and family for their never-ending support of me and my academic pursuits. Specifically, to my parents, Jeanne Bowen and David Housman, thank you for your love and encouragement through all my research endeavors and my life in general. Finally, to Eric D. Johnson, thank you for everything that you have done to help me during this process, from bringing me dinner on the many occasions I worked late into the night on campus to guiding me in the development of a massive relational database, teaching me more efficient research workflow pipelines, and helping me understand complicated statistical analyses. You pushed me to think critically about my research, and our discussions have helped me to become a better scientist, educator, and human. I cannot thank you enough.

Beyond this, I am indebted to so many other people that have helped me to get to this point in my career and life. This includes my primary and secondary educators, my first research supervisor during my undergraduate education, Sibaji Sarkar, and numerous

other faculty, staff members, colleagues, and associates that have helped me along my journey. I apologize if I forgot to mention anyone, but please know, I am eternally grateful for your support.

TABLE OF CONTENTS

Page

APPENDIX Page

APPENDIX Page

APPENDIX Page

APPENDIX Page

LIST OF TABLES

Table Page

Table Page

LIST OF FIGURES

Figure Page

Figure Page

Figure Page

Figure Page

LIST OF EQUATIONS

Equation Page

CHAPTER 1

INTRODUCTION

Across the primate order, different species are characterized by divergent skeletal traits that contribute to inter-specific anatomical variation and enable varied niche occupations and forms of locomotion. These phenotypic distinctions are often described as the result of environmental adaptations. However, skeletal phenotypes are more accurately defined as complex traits, and environmental, genetic, and epigenetic mechanisms all contribute to these phenotypes. Nevertheless, skeletal complexity in relation to epigenetic variation has not been assessed across the primate order. In order to gain a complete understanding of the evolution of skeletal phenotypes across primates, it is necessary to study skeletal epigenetics in primates.

The importance of gene regulation for primate phenotypic diversity was originally noted by King and Wilson in 1975 and has gained credibility as the extent of genetic similarity among phenotypically distinct primate taxa has been clarified. The epigenome comprises a level of gene regulation that can change in response to environmental factors, and within the epigenome, DNA methylation serves as one form of gene regulation. Although general changes to mammalian epigenomes have been examined (Sharif et al. 2010), work on nonhuman primates has been limited to whole-genome methylation patterns of a small number of tissue-specific cells from a small number of species. Very few studies have tried to relate DNA methylation to variation in specific phenotypes. The research presented here begins to remedy this by examining the association between variation in DNA methylation and skeletal phenotypes among several nonhuman primate species.

Focusing on skeletal phenotypes is valuable for several reasons. First, skeletal anatomy varies across primates, and these underlying skeletal differences impact the overall anatomy of an animal, which in turn influences the range of niches it can occupy and forms of locomotion and movement it can perform. Secondly, skeletal morphology is readily used to reconstruct extinct species within the primate lineage. Inferences made from preserved skeletal remains within the fossil record inform our understanding of primate evolution. Therefore, understanding how extant primate skeletal traits vary and how underlying genetic and epigenetic components contribute to this morphological variation is crucial for proper evaluation of both ancient and modern primate skeletal systems. Third, while skeletal phenotypes are influenced by genetic (Goldring and Marcu 2012) and environmental (Henriksen et al. 2014; Macrini et al. 2013) forces, epigenetic factors also play an important role in bone development and maintenance (Delgado-Calle et al. 2013; García-Ibarbia et al. 2013; Iliopoulos et al. 2008; Loughlin and Reynard 2015; Ramos et al. 2014; Reynard et al. 2014). Thus, it is logical to hypothesize that these epigenetic mechanisms may also be involved in the evolution of diverse skeletal phenotypes across the primate order. Lastly, the emerging field of ancient epigenetics, which assesses DNA methylation patterns in ancient hominin skeletal remains (Smith et al. 2015; Gokhman et al. 2014), is lacking information on DNA methylation patterns in skeletal tissues from nonhuman primates. As such, these ancient hominin skeletal epigenetic patterns cannot currently be put into a broader phylogenetic or evolutionary context.

Researchers are currently exploring primate DNA methylation variation in relation to several behavioral, soft tissue, and disease-related phenotypes (Farcas et al.

2009; Hernando-Herraez et al. 2013; Martin et al. 2011; Molaro et al. 2011; Pai et al. 2011; Zeng et al. 2012). Additionally, medical fields are examining the relationship between DNA methylation variation and the manifestation of several bone pathologies in humans and model organisms (Bovée et al. 2010; Dimitriou et al. 2011; Goldring and Marcu 2012; Iliopoulos et al. 2008; Kasaai et al. 2013; Y. Liu et al. 2013; Ralston and Uitterlinden 2010; Rivadeneira et al. 2009). However, scientists have not yet studied the impact of this mechanism on nonhuman primate hard tissues or skeletal phenotypes. This study attempts to fill this knowledge gap by assessing how epigenetic patterns vary within and among primate taxa and in relation to skeletal phenotypes. Specifically, the overarching goals of this research are to identify genome-wide and gene-specific DNA methylation patterns in nonhuman primate skeletal tissues and assess variation both intra-specifically by determining how patterns differ between tissue types, between age ranges, between skeletal disease states, and between nonpathological skeletal morphologies and inter-specifically across several nonhuman primate species with a wide phylogenetic distribution.

Samples include baboons (*Papio spp.*, n=74), macaques (*Macaca mulatta*, n=10), vervets (*Chlorocebus aethiops*, n=10), chimpanzees (*Pan troglodytes*, n=4), and marmosets (*Callithrix jacchus*, n=6). Within baboons, samples include skeletally healthy adults (n=28), osteoarthritic adults (n=28), and skeletally healthy juveniles (n=18). Skeletal tissues collected from nonhuman primates include trabecular bone and cartilage from the medial condyles of right distal femora. Methylation arrays and gene-specific methylation sequencing were then used to assess how DNA methylation patterns in these tissues varied intra- and inter-specifically.

Specifically, in Chapter 2, I use the Illumina Infinium Human Methylation450 BeadChip (450K array) to identify DNA methylation patterns in bone and cartilage of age-matched, adult female baboons, five with and five without knee osteoarthritis (OA), in order to validate that this methylation array can be used for nonhuman primate skeletal tissue DNA extracts and to assess whether DNA methylation variation is associated with OA in baboons and in a manner similar to that observed in humans. Similarly, in Chapter 3, I use the Illumina Infinium MethylationEPIC BeadChip (EPIC array) to explore the evolution of OA epigenetics further by identifying DNA methylation patterns in bone and cartilage of 56 pedigreed, adult baboons, 28 with and 28 without knee OA, and by assessing whether DNA methylation variation is associated with OA in baboons and in a manner similar to that observed in humans. In Chapter 4, I use the EPIC array to examine the evolution of aging epigenetics by identifying DNA methylation patterns in bone from 46 pedigreed baboons, 28 that were adults and 18 that were juveniles, and assessing whether DNA methylation variation is associated with aging in baboons and in a manner similar to that observed in humans. Lastly, in Chapter 5, I validate that the EPIC array can be used for skeletal tissue DNA extracts from several nonhuman primate species and assess how genome-wide and gene-specific DNA methylation in bone varies intra-specifically in relation to nonpathological femur bone morphologies and inter-specifically for 28 baboons, 10 macaques, 10 vervets, 4 chimpanzees, and 6 marmosets. Overall, the findings from this research reveal how skeletal DNA methylation patterns vary within and among primate species and relate to skeletal phenotypes, and together they inform our understanding of epigenetic regulation and complex skeletal trait evolution in primates.

CHAPTER 2

ASSESSMENT OF DNA METHYLATION PATTERNS IN THE BONE AND CARTILAGE OF A NONHUMAN PRIMATE MODEL OF OSTEOARTHRITIS

Abstract

The degenerative joint disease of osteoarthritis (OA) impacts humans and several other animals. Thus, the mechanisms underlying this disorder may be evolutionary conserved. In particular, variation in skeletal tissue DNA methylation patterns are thought to be a critical mechanism in the development of OA. However, the associations between DNA methylation and OA development have not been optimized or readily studied in nonhuman primates. The Illumina Infinium Human Methylation450 BeadChip (450K array) is a cost-efficient application for assessing genome-wide DNA methylation patterns. Although it was designed for human DNA, the 450K array has also been successfully used for nonhuman primates because of the relative conservation between these organisms' genomes. Baboons (*Papio spp.*) serve as important models of disease and develop OA at rates similar to those observed in humans, so further investigation of the associations between DNA methylation patterns and OA development in this organism will advance the evolutionary understanding of this disease. Here, I used the 450K array to identify DNA methylation patterns in femur bone and cartilage of adult female baboons, five with and five without knee OA. I validated that the hybridization efficiency of 450K array probes is related to the degree of sequence similarity between the probes and the baboon genome. Additionally, approximately 44% of the 450K array probes reliably align to the baboon genome, contain a CpG site of interest, and maintain a

wide distribution throughout the genome. I also found that filtering probes using alignment similarity criteria retains more efficiently hybridized probes than filtering probes using gene symbol similarity criteria. Both filtering methods identified significantly differentially methylated positions (DMPs) between healthy and OA individuals in cartilage tissues, and some of these patterns overlap with those previously identified in humans. Conversely, in bone tissues, no DMPs were found between disease states, and no DMPs were found between tissue types. Overall, I conclude that the 450K array can be used to measure genome-wide DNA methylation in baboon tissues and identify significant associations with complex traits. The results of this study indicate that some DNA methylation patterns associated with OA are evolutionarily conserved while others are not, and this warrants further investigation in a larger and more phylogenetically diverse sample set.

Key Words

DNA methylation, baboon, osteoarthritis, bone, cartilage, nonhuman primate

Introduction

Osteoarthritis (OA) is a complex degenerative joint disease, and OA of the knee is one of the leading causes of disability across the globe (Cross et al. 2014). Thus, research endeavors to describe the molecular mechanisms that contribute to this disorder are underway. Both genetic and environmental factors have some effect (Blagojevic et al. 2010; Cooper et al. 2000; D. T. Felson and Zhang 1998; David T. Felson 2004; Henriksen et al. 2014; Jordan et al. 2007; Macrini et al. 2013; Rossignol et al. 2005).

However, epigenetic factors, such as DNA methylation which regulates gene expression, are now thought to play a more influential role in the development of degenerative skeletal disorders like OA (Delgado-Calle et al. 2013; den Hollander et al. 2014; Fernández-Tajes et al. 2014; García-Ibarbia et al. 2013; Goldring and Marcu 2012; Iliopoulos et al. 2008; Jeffries et al. 2016; Y. Liu et al. 2013; Loughlin and Reynard 2015; Moazedi-Fuerst et al. 2014; Ramos et al. 2014; Reynard et al. 2014; Rushton et al. 2014a).

Animal models, such as mice, rats, rabbits, guinea pigs, dogs, sheep, goats, and horses, have been essential in discerning some of the processes inherent to OA development (Bendele 2001; Kuyinu et al. 2016; Cucchiarini et al. 2016). Nevertheless, all of these animals are limited in their ability to fully inform our understanding of human OA, so the search to find a gold standard animal model for OA is still ongoing (Ameye and Young 2006; Lampropoulou-Adamidou et al. 2014). Additionally, while the conservation of this disorder in several species implies that the mechanisms contributing to OA may be evolutionarily conserved, few studies of OA have taken an evolutionary perspective (Ostrer et al. 2006; Rugg-Gunn et al. 2005). Lastly, although variation in skeletal tissue DNA methylation patterns are thought to be involved in the development and progression of OA, this epigenetic mechanism has not been readily studied in animal models because assays to assess variation in this regulatory level have not been optimized.

Nonhuman primates can serve as important models of disease for humans because they are phylogenetically close to humans. Baboons (*Papio spp.*) are a particularly good model of disease, especially OA (Cox et al. 2013), as they naturally develop OA at rates

similar to those observed in humans (Macrini et al. 2013; Cox et al. 2013; O'Connor 2006). Additionally, because of their evolutionary proximity to humans, further investigation of the molecular processes innate to OA development and progression in baboons as compared to these mechanisms in humans will advance the evolutionary understanding of this disease. Finally, the relative genetic conservation between baboons and humans makes the optimization and use of standardized DNA methylation assays possible. Specifically, the Infinium Human Methylation450 BeadChip (450K array), which is a cost-efficient application for assessing genome-wide DNA methylation patterns in humans, has been successfully used for some nonhuman primate species. These and other nonhuman primate DNA methylation studies have primarily used DNA extracted from blood or other soft tissues (Enard et al. 2004; Farcas et al. 2009; Fukuda et al. 2013; Hernando-Herraez et al. 2013; Kothapalli et al. 2007; Lindskog et al. 2014; Martin et al. 2011; Molaro et al. 2011; Ong et al. 2014; Pai et al. 2011; Provencal et al. 2012; Zeng et al. 2012). However, this technique has not yet been used to study DNA methylation variation in baboon skeletal tissues or how it relates to the development of OA in a nonhuman primate species.

For this study, I used the 450K array to identify DNA methylation patterns in femur bone and cartilage of age-matched female baboons, five with and five without knee OA, in order to validate that this technique can be used for nonhuman primate skeletal tissue DNA extracts and to assess whether DNA methylation variation is associated with OA in baboons and in a manner similar to that observed in humans.

Methods

Ethics Statement

Nonhuman primate tissue samples included were opportunistically collected at routine necropsy of these animals. No animals were sacrificed for this study, and no living animals were used in this study.

Samples

Baboon (*Papio spp.*) samples come from captive colonies at the Southwest National Primate Research Center in the Texas Biomedical Research Institute. These samples are ideal because many environmental factors that influence skeletal development and maintenance (e.g., diet and exposure to sunlight, which influences vitamin D production) are controlled and consistent across individuals.

Femora were opportunistically collected at routine necropsy of these animals and stored in -20°C freezers at the Texas Biomedical Research Institute after dissection. These preparation and storage conditions ensured the preservation of skeletal DNA methylation patterns.

Samples include skeletally healthy adult baboons (n=5) and adult baboons with severe osteoarthritis (OA, n=5). Age ranges are comparable between each group (Table 1), and only females were included in this study.

Assessment of Osteoarthritis

Classification of adult baboons as having healthy or OA knees was determined through visual examination of the distal femora and macroscopic inspection of the distal

articular surface cartilage. Each specimen was assigned an OA severity score. Briefly, Grade 1 is unaffected, Grade 2 is mild OA as indicated by cartilage fibrillation, Grade 3 is moderate OA as indicated by cartilage lesions, and Grade 4 is advanced OA as indicated by eburnation (Macrini et al. 2013). From this, binary classifications were made such that all healthy adult baboons have 100% Grade 1 on one or both distal femora, and all OA adult baboons have a variable percentage of Grades 3 or 4 on one or both distal femora (Figure 1).

Table 1. Baboon Samples for 450K Array Osteoarthritis Study.

Comparative Group	No.	Age
Healthy bone	5	19.30±1.70
OA bone	5	19.24±1.73
Healthy Cartilage	5	19.30±1.70
OA Cartilage	5	19.24±1.73

Table outlines the number (No.) and the average age in years plus or minus one standard deviation (Age) of individuals in each comparative group.

DNA Extraction

From the distal femoral condyles, cartilage scrapings were collected using scalpels and processed with a homogenizer, and trabecular bone samples were obtained using a Dremel and pulverized into bone dust using a BioPulverizer. Both tissues are included in this project because human skeletal epigenetic studies are based on trabecular bone and cartilage, and it is important to standardize tissue type for comparative purposes. Additionally, these tissues are both clinically relevant in terms of disease progression. DNA was extracted from these processed tissues using a phenol-chloroform

protocol optimized for skeletal tissues (Barnett and Larson 2012) and quantified using both Nanodrop and Qubit machines (APPENDIX C).

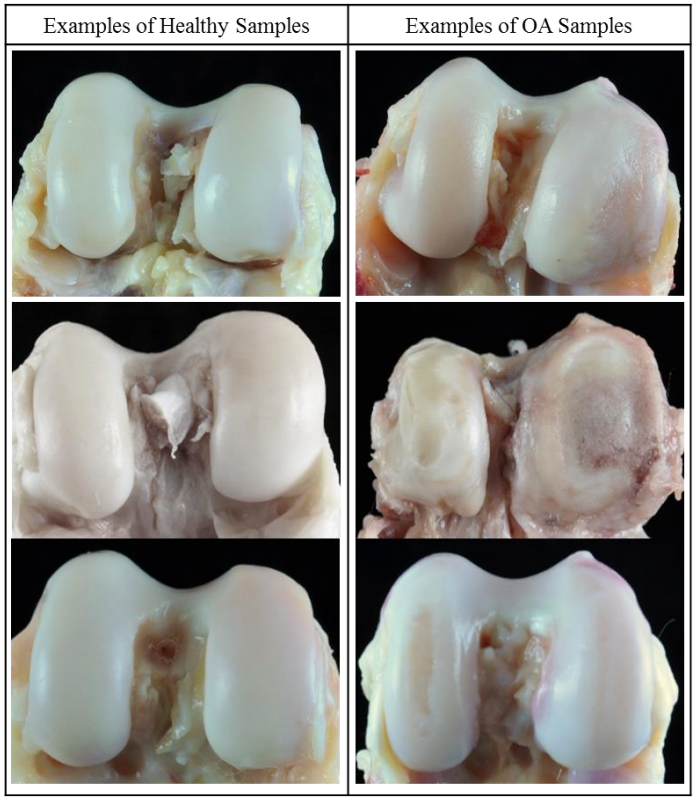

Figure 1. Examples of Healthy and Osteoarthritic Baboon Knee Joints.

Representative examples of baboon knees (distal femora) that are healthy or have OA.

Genome-Wide DNA Methylation Profiling

Genome-wide DNA methylation was assessed using the Infinium HumanMethylation450 BeadChips (450K array). These arrays analyze the methylation status of over 485,000 sites throughout the genome, covering 99% of RefSeq genes and 96% of the UCSC-defined CpG islands and their flanking regions. For each sample, approximately 500ng of genomic DNA (APPENDIX C) was bisulfite converted using the

EZ DNA Methylation™ Gold Kit according to the manufacturer's instructions (Zymo Research), with modifications described in the Infinium Methylation Assay Protocol. Following manufacturer guidelines (Illumina), this processed DNA was then whole-genome amplified, enzymatically fragmented, hybridized to the arrays, and imaged using the Illumina iScan system. The array data discussed here are available in APPENDIX B.

Methylation Data Processing

Raw fluorescent data were normalized to account for the noise inherent within and between the arrays themselves. Specifically, I performed a normal-exponential out-of-band (Noob) background correction method with dye-bias normalization (Triche et al. 2013) to adjust for background fluorescence and dye-based biases and followed this with a between-array normalization method (functional normalization) (Fortin et al. 2014) which removes unwanted variation by regressing out variability explained by the control probes present on the array as implemented in the minfi package in R (Aryee et al. 2014; Fortin et al. 2016), which is part of the Bioconductor project (Huber et al. 2015). This method has been found to outperform other existing approaches for studies that compare conditions with known large-scale differences (Fortin et al. 2014), such as those assessed in this study.

After normalization, methylation values (β values) for each site were calculated as the ratio of methylated probe signal intensity to the sum of both methylated and unmethylated probe signal intensities (Equation 1). These β values range from 0 to 1 and represent the average methylation levels at each site across the entire population of cells

from which DNA was extracted (0 = completely unmethylated sites, 1 = fully methylated sites).

Equation 1: β Value= $\frac{\text{Methylated Signal}}{(\text{Methylated Signal}+\text{Unmethylated Signal})}$

Every β value in the Infinium platform is accompanied by a detection p-value, and those with failed detection levels (p-value > 0.05) in greater than 10% of samples were removed from downstream analyses.

The probes on the arrays were designed to specifically hybridize with human DNA, so my use of nonhuman primate DNA required that probes non-specific to the baboon genome, which could produce biased methylation measurements, be computationally filtered out and excluded from downstream analyses. This was accomplished using two different methods modified from (Hernando-Herraez et al. 2013; Ong et al. 2014).

For both methods, I used blastn (Altschul et al. 1997) to map the 485,512 50bp probes onto the *Papio anubis* genome (Assembly: Panu_2.0, Accession: GCF_000264685.2) using an e-value threshold of e^{-10}. I retained probes that successfully mapped to the baboon genome, had only 1 unique BLAST hit, and targeted CpG sites (APPENDIX A). Then, for the first method, which used criteria based on sequence alignment, I only retained probes that had 0 mismatches in 5bp closest to and including the CpG site, and had 0-2 mismatches in 45bp not including the CpG site. For the second method, which used criteria based on gene symbol similarities, I identified the closest baboon gene to each probe site and checked for corresponding gene name matches

between humans and baboons. For baboons, this information was obtained from GFF and Ensembl BioMart data. Only those probes with partial or complete gene matches were retained. Additionally, β values associated with cross-reactive probes (Y. Chen et al. 2013), probes containing SNPs at the CpG site, probes detecting SNP information, probes detecting methylation at non-CpG sites, and probes targeting sites within the sex chromosomes were removed using the minfi package in R (Aryee et al. 2014; Fortin et al. 2016) (Figure 2).

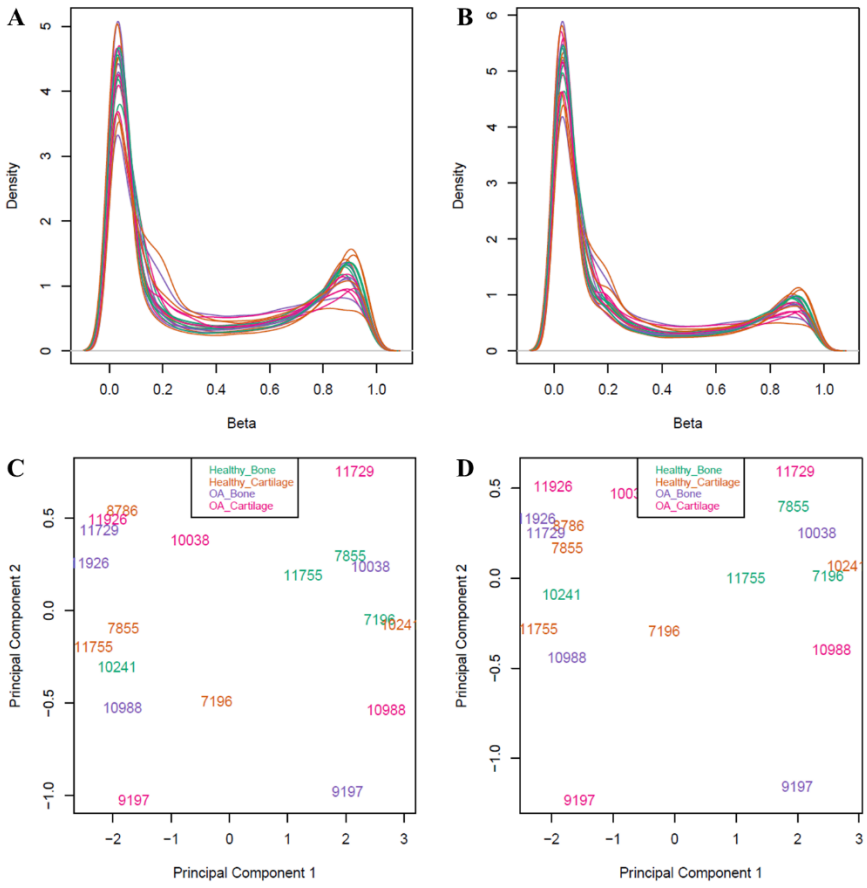

Figure 2. Normalized and Filtered Methylation Data for 450K Array Baboon Osteoarthritis Study.

Density plots of β values after normalization and probe filtering using the alignment filter criteria (A) or the gene symbol filter criteria (B). Multidimensional scaling plots showing the first two principle components that describe genome-wide methylation variation after normalization and filtering using the alignment filter criteria (C) or the gene symbol filter criteria (D). Each point represents one sample that is either from healthy bone, healthy cartilage, OA bone, or OA cartilage. In the multidimensional scaling plots, these categories do not form distinct clusters.

Differential Methylation Analyses

Because β values have high heteroscedasticity, they are not statistically valid for use in differential methylation analyses (Du et al. 2010). Thus, M values were calculated and used in these analyses instead (Equation 2).

Equation 2: $\text{M Value} = \log\left(\frac{\text{Methylated Signal}}{\text{Unmethylated Signal}}\right)$

In order to identify sites that were significantly differentially methylated across comparative groups, I designed and tested generalized linear mixed models (GLMMs) which related the variables of interest to the DNA methylation patterns for each site, while accounting for latent variables (Maksimovic et al. 2016). Sites found to have significant associations were classified as significantly differentially methylated positions (DMPs).

Specifically, a GLMM was used to estimate differences in methylation levels for each of the following contrasts:

1. between bone and cartilage in OA baboons
2. between bone and cartilage in healthy baboons
3. between OA and healthy baboon bone
4. between OA and healthy baboon cartilage
5. among all 4 combinations of tissue type and disease state (healthy bone vs. healthy cartilage vs. OA bone vs. OA cartilage)

Additional variables included in this GLMM were unknown latent variables calculated using the iteratively re-weighted least squares approach in the sva package in R (Jaffe and Irizarry 2014; Jeffrey T. Leek et al. 2012; J. T. Leek and Storey 2008; Jeffrey T. Leek and Storey 2007). The 4 latent variables estimated were included to help mitigate any unknown batch and cell heterogeneity effects on methylation variation at each site. No predefined batch effects for the arrays were included because these did not appear to have large effects on the data (Figure 3).

Alternative methods to account for cell heterogeneity exist, but they are specific to whole blood (Jaffe and Irizarry 2014; Morris and Beck 2015), require reference epigenetic data, or are reference free methods (Houseman et al. 2014) that are comparable to the sva method (Kaushal et al. 2015). Out of the known cell types in skeletal tissues (Horvath, Mah, et al. 2015), only chondrocytes and osteoblasts have reference epigenomes available on the International Human Epigenomics Consortium, and these are only for humans, not nonhuman primates. Thus, because no standard

method is available to correct for the heterogeneous cell structure in nonhuman primate skeletal tissue, I chose the described sva method.

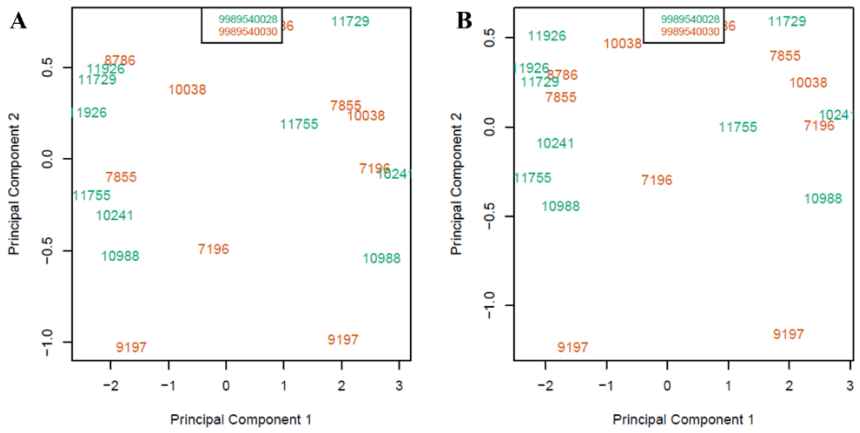

Figure 3. Batch Effects on Normalized and Filtered Methylation Data for 450K Array Baboon Osteoarthritis Study.

Multidimensional scaling plots showing the first two principle components that describe genome-wide methylation variation after normalization and filtering using the alignment filter criteria (A) or the gene symbol filter criteria (B). Each point represents one sample that is either from healthy bone, healthy cartilage, OA bone, or OA cartilage. The colors indicate which array (9989540028 or 9989540030) the samples were run on. These batch effects do not appear to cluster samples into distinct groups. Furthermore, they do not cluster samples based on their healthy bone, healthy cartilage, OA bone, or OA cartilage groupings (see Figure 2).

This GLMM design matrix (Equation 3) was fit to the M value array data by generalized least squares using the limma package in R (Ritchie et al. 2015; Phipson et al.

2016; Huber et al. 2015), and the estimated coefficients and standard errors for the defined tissue type and disease status contrasts were computed. Lastly, for each coefficient, an empirical Bayes approach (McCarthy and Smyth 2009; Lönnstedt and Speed 2002; Phipson et al. 2016; Smyth 2004) was used to compute moderated t-statistics, log-odds ratios of differential methylation, and associated p-values adjusted for multiple testing (Benjamini and Hochberg 1995). Significant DMPs for the effect of tissue type and disease status contrasts were defined as those having log fold changes in M values corresponding to an adjusted p-value of less than 0.05.

Equation 3: methylation ~ tissue type and disease status contrasts + latent variables

Results

The aim of this study was to use the 450K array to identify DNA methylation patterns in femur bone and cartilage of age-matched female baboons, five with and five without knee OA. In order to do this, I first assessed the effectiveness of the 450K array in identifying DNA methylation patterns in baboon DNA and of different probe filtering methods.

Alignment of 450K Array Probes with the Baboon Genome

Probes from the 450K array were aligned to the baboon genome using methods modified from (Hernando-Herraez et al. 2013; Ong et al. 2014) (APPENDIX A). Out of the 485,512 50bp probes on the array, 213,858 probes (44%) map to the baboon genome with e-values less than e^{-10}, have only unique BLAST hits, and target a CpG site (Figure

4). Out of these reliably mapped probes, 133,264 probes (62%) were retained after the alignment filter criteria, while 130,307 probes (61%) were retained after the gene symbol filter criteria (Figure 4). 83,142 probes overlapped between both filtering methods (62% for the alignment filter criteria and 64% for the gene symbol filter criteria, Figure 5).

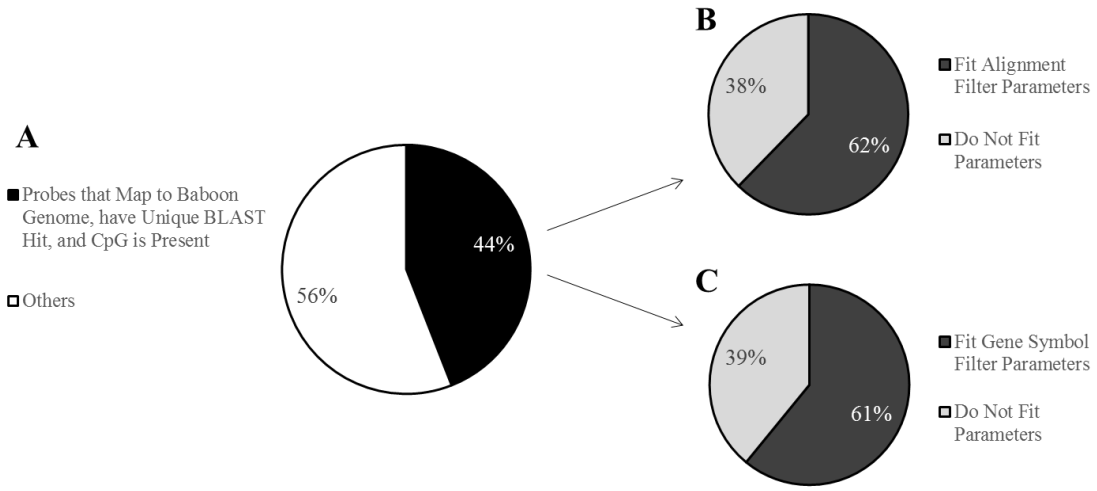

Figure 4. Filtering Effects on 450K Array Probes for Baboons.

(A) Pie chart showing the percent of 450K array probes that map to the baboon (*Papio anubis*) genome with e-values less than e^{-10}, have only unique BLAST hits, and target a CpG site. Out of 485,512 probes total, 213,858 probes (44%) meet these criteria. (B) Pie chart showing the percent of probes, out of those that successfully mapped to the baboon genome, that contain 0 mismatches in 5bp of the probe by and including the targeted CpG site and 0-2 mismatches in 45bp of the probe not including the CpG site. Out of the 213,858 mapped probes, 133,264 probes (62%) meet these criteria. (C) Pie chart showing the percent of probes, out of those that successfully mapped to the baboon genome, with

gene symbol matches to humans. Out of the 213,858 mapped probes, 130,307 probes (61%) meet these criteria.

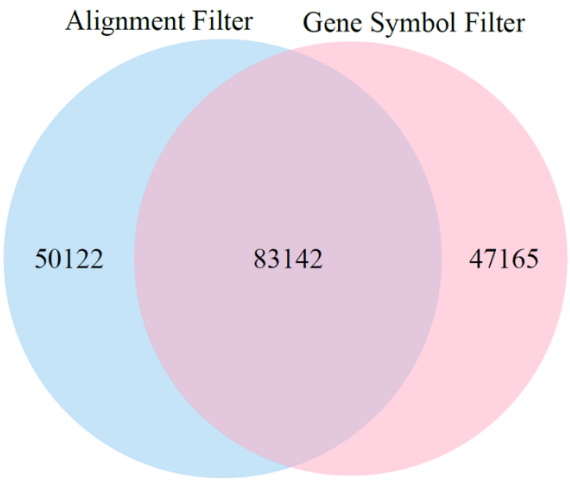

Figure 5. Overlap of 450K Array Probes for Baboons Using Different Filtering Methods. Venn diagram showing the number of probes that overlap between the alignment filter criteria and the gene symbol filter criteria. Out of the 133,264 probes that meet the alignment filter criteria and the 130,307 probes that meet the gene symbol criteria, 83,142 probes (62% and 64% respectively) overlap in both filters.

Probes that reliably mapped to the baboon genome, that met the alignment filter criteria, or that met the gene symbol criteria covered approximately 18,800 genes with an average coverage of 9, 6, or 8 probes per gene, respectively (APPENDIX D). Additionally, the retained probes covered a range of locations with respect to genes and CpG islands (APPENDIX D), indicating that these filtered probes maintain a wide distribution throughout the genome.

After filtering out cross-reactive probes (Y. Chen et al. 2013), probes containing SNPs at the CpG site, probes detecting SNP information, probes detecting methylation at non-CpG sites, and probes targeting sites within the sex chromosomes, a final set of 120,305 probes were retained for the alignment filter criteria, and a final set of 112,760 probes were retained for the gene symbol criteria (Figure 2).

Effectiveness of 450K Array Probes using Baboon DNA

To determine how effectively the 450K array probes measured DNA methylation in baboon DNA, I performed Spearman correlation tests between the hybridization efficiency of each probe and parameters defining the alignment quality of each probe to the baboon genome. Specifically, both probe alignment bitscores and percent identity were significantly negatively correlated with probe hybridization efficiency, and probe alignment e-values were significantly positively correlated with probe hybridization efficiency, regardless of filtering criteria (APPENDIX E). However, filtering probes using the alignment filter criteria retained proportionally more successfully hybridized probes than filtering probes using the gene symbol filter criteria (Figure 6). Thus, filtering probes using the alignment filter criteria likely produces more reliable results.

Differential Methylation and Osteoarthritis

Significant DMPs were only identified between healthy and OA individuals in cartilage tissues (Table 2). All of these DMPs displayed decreased methylation in OA cartilage samples as compared to healthy cartilage samples, and some of these patterns

overlapped with those previously identified in humans. Conversely, no DMPs were found between tissue types or between disease states in bone tissues.

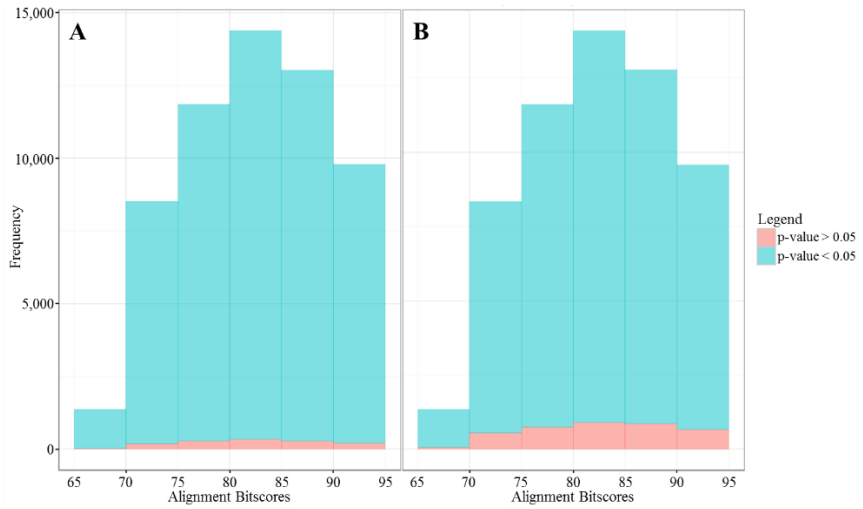

Figure 6. Hybridization Efficiencies of 450K Array Probes Retained for Baboon Osteoarthritis Study.

Histogram of alignment bitscores for 450K array probes with detection p-values > 0.05 (red) and < 0.05 (blue). These p-values were averaged across all samples, and probes included meet the alignment filter criteria (A) or the gene symbol filter criteria (B). For probes meeting the alignment filter criteria (A), 3,123 had detection p-values > 0.05, and 130,141 had detection p-values < 0.05. For probes meeting the gene symbol filter criteria (B), 8,695 had detection p-values > 0.05, and 121,612 had detection p-values < 0.05. For all probes that successfully mapped to the baboon genome with e-values less than e^{-10}, had only unique BLAST hits, and targeted a CpG site, 16,715 had detection p-values > 0.05, and 197,143 had detection p-values < 0.05.

Table 2. Number of Significant DMPs Identified in the 450K Array Baboon Osteoarthritis Study.

	Differential Methylation	OA Bone vs. OA Cartilage	Healthy Bone vs. Healthy Cartilage	OA Bone vs. Healthy Bone	OA Cartilage vs. Healthy Cartilage	OA Bone vs. Healthy Bone vs. Healthy Cartilage vs. OA Cartilage
Alignment Filter Probes	Significant (negative)	0	0	0	6	0
	Not Significant	120,305	120,305	120,305	120,299	120,305
	Significant (positive)	0	0	0	0	0
Gene Symbol Filter Probes	Significant (negative)	0	0	0	2	0
	Not Significant	112,760	112,760	112,760	112,758	112,760
	Significant (positive)	0	0	0	0	0

Table showing the number of significant DMPs between comparative groups. Results are shown for probes filtered using the alignment filter criteria and probes filtered using the gene symbol filter criteria, and for both sets, significant DMPs were only identified between OA cartilage and healthy cartilage. For all of these DMPs, OA cartilage samples have decreased methylation as compared to healthy cartilage samples.

When filtering probes using the alignment filer criteria, 6 significant DMPs were identified between OA cartilage samples and healthy cartilage samples, while only 2 DMPs were identified when filtering probes using the gene symbol criteria (Table 3). One locus matched between these filtering methods. *RUNX1* has previously been found to be differentially methylated in OA and healthy cartilage in humans, with OA cartilage having lower methylation as compared to healthy cartilage (Fernández-Tajes et al. 2014). The other genes associated with these probes have not previously been associated with OA in humans (Alvarez-Garcia et al. 2016; Aref-Eshghi et al. 2015; Delgado-Calle et al. 2013; Fernández-Tajes et al. 2014; García-Ibarbia et al. 2013; Goldring and Marcu 2012;

Iliopoulos et al. 2008; Jeffries et al. 2016; Moazedi-Fuerst et al. 2014; Ramos et al. 2014; Reynard et al. 2014; Rushton et al. 2014a; Saito et al. 2010).

Table 3. Gene Details of Significant DMPs Identified in the 450K Array Baboon Osteoarthritis Study.

	450K Array Probe ID	Log Fold Change in M Values	Adjusted P-Value	Human Gene Symbol	Baboon Gene Symbol	Baboon Chromosome	Baboon CpG Position
Alignment Filter Probes	cg05295841	-3.61359	0.024654	*KLHL26*	*CRTC1*	19	17216922
	cg17890983	-1.96830	0.024654	*RFXAP*	*RFXAP*	17	15599705
	cg02329670	-1.49812	0.041104	*MIR497; MIR195*	*LOC103878622*	16	6672282
	cg18456803	-2.04780	0.041104	*ELF1*	*WBP4; LOC103879193*	17	19630529
	cg13030790	-3.54370	0.041104	*RUNX1*	*RUNX1*	3	11500594
	cg24721647	-2.18375	0.041104	*ACSL1*	*ACSL1*	5	173909497
Gene Symbol Filter Probes	cg17890983	-1.96789	0.037417	*RFXAP*	*RFXAP*	17	15599705
	cg04759112	-1.63196	0.038075	*CMIP*	*CMIP*	20	63369656

Table showing the details on the significant DMPs between OA cartilage and healthy cartilage. Results are shown for probes filtered using the alignment criteria and probes filtered using the gene symbol criteria. For all of these DMPs, OA cartilage samples are hypomethylated as compared to healthy cartilage samples.

Discussion

Here, I used the 450K array to identify DNA methylation variation in bone and cartilage tissues from a baboon model of OA. This was done both to determine the effectiveness of this application for baboon DNA and to assess the evolutionary conservation of epigenetic-OA associations in the primate lineage.

I show that using the 450K array is feasible in baboon tissues. *In silico* probe filtering methods (Hernando-Herraez et al. 2013; Ong et al. 2014) indicated that 44% of

all human probes could be reliably mapped to the baboon genome and contained a CpG locus. This number was lower than expected since previous researchers were able to use these same methods to reliably map 61% of the human probes to the Cynomologus macaque genome (Ong et al. 2014), another Old World monkey that is a close phylogenetic relative to baboons. This discrepancy in number may be due to the quality of each nonhuman primates' genome assembly. While both are well annotated, the average scaffold length (88,649,475) and contig length (86,040) of the macaque genome (Assembly: Macaca_fascicularis_5.0, Accession: GCF_000264685.2) are higher than those (528,927 and 40,262) of the baboon genome.

Subsequent *in silico* analyses based on sequence alignment criteria (Hernando-Herraez et al. 2013) and based on gene symbol criteria (Ong et al. 2014) retained similar numbers of probes (Figure 4) that maintained wide and comparable distributions throughout the genome (APPENDIX D). However, only a little more than half of the resulting probes for each filtering technique overlapped with one another (Figure 5). This discrepancy is likely due to the incomplete nature of the baboon genome annotation. More than half of the probes that fit the alignment filter criteria but not the gene symbol criteria (28,699 out of 50,117) are associated with generic gene symbol identifiers (LOC) to indicate the as of yet unknown functions of these regions. Conversely, all of the probes that fit the gene symbol criteria but not the alignment filter criteria have over 3 mismatches with the baboon genome on average and have a maximum of 9 mismatches with the baboon genome. These high mismatch numbers are a potential concern for proper and accurate probe and baboon DNA hybridization.

Fittingly, after applying the 450K array to measure DNA methylation patterns of genomic material extracted from baboon skeletal tissues, I found that the hybridization efficiency of probes was significantly correlated with the alignment quality of each probe to the baboon genome, and thus, the degree of sequence conservation. The majority of filtered probes for both *in silico* methods passed quality controls and produced robust signals on the array, indicating that either filtering technique may be appropriate for future research. However, because the filtering method based on the alignment filter criteria retained a larger proportion of successfully hybridized probes than the method based on the gene symbol criteria (Figure 6) and because this method is less influenced by the degree of genome assembly annotation, I recommend that this alignment filter criteria method be preferentially used in subsequent nonhuman primate studies.

This work is an extension of previous work and uses the 450K array to study DNA methylation in baboons. The 450K array is advantageous because it is cost efficient per sample and simultaneously measures a large number of CpG loci with a broad genomic representation (Michels et al. 2013). Similar to this study, previous researchers have used the 450K array to measure DNA methylation patterns in great apes (Hernando-Herraez et al. 2013), which are closer to humans evolutionarily than baboons, and in macaques (Ong et al. 2014), which are comparable in proximity to humans evolutionarily as compared to baboons. All together these studies open new areas of research that incorporate animal models of disease or an evolutionary perspective of diseases across phylogenies, and the work presented here begins to advance such areas of research.

Specifically, I used a baboon model of OA to assess the evolutionary conservation of epigenetic-OA associations in the primate lineage. To do this, I identified significant

DMPs between healthy and OA individuals in cartilage and bone tissues. I also looked for DMPs between tissue types and between all four combinations of disease state and tissue type (healthy cartilage vs. OA cartilage vs. healthy bone vs. OA bone). However, DMPs were only found between healthy and OA individuals in cartilage tissues (Table 2), and all of these loci showed hypomethylation in OA cartilage samples as compared to healthy cartilage samples. This corresponds to the general global hypomethylation that is also observed in OA cartilage as compared to healthy cartilage. Six DMPs were identified when using the alignment filter criteria, and two DMPs were identified when using the gene symbol filter criteria (Table 3). All together these loci are associated with 8 genes – *KLHL26*, *RFXAP*, *MIR497*, *MIR195*, *ELF1*, *RUNX1*, *ACSL1*, and *CMIP* – that have a variety of functions.

Some of these genes have functions directly related to skeletal development and maintenance. For instance, *RUNX1* (Gene ID: 861), also known as runt related transcription factor 1, is involved in the regulation of bone and cartilage cell development and differentiation (Stein et al. 2004). Additionally, *MIR497* (Gene ID: 574456) and *MIR195* (Gene ID: 406971) are non-coding microRNAs that are involved in post-transcriptional regulation (Wei et al. 2015). While both of these microRNAs have roles in the development of cancer (Li et al. 2011; L. Liu et al. 2010), they also play important regulatory roles in the differentiation of mesenchymal stromal/stem cells into bone related cells (Almeida et al. 2016).

Other genes have functions associated with the immune system, which may have proximal roles in the development of OA. In particular, *RFXAP* (Gene ID: 5994), also known as regulatory factor X associated protein, codes for a protein that assists in the

transcriptional activation of major histocompatibility class II genes which are critical for the development and control of the immune system (Garvie and Boss 2008). Additionally, *CMIP* (Gene ID: 80790), also known as c-Maf inducing protein, codes for a protein that is involved in the T-cell signaling pathway, and SNPs within this gene have been associated with chronic diseases like diabetes (Dastani et al. 2012).

The remaining genes do not have functions related to skeletal phenotypes, which makes their involvement in OA less clear. For example, *KLHL26* (Gene ID: 55295), also known as kelch like family member 26, is part of a family of proteins that may be involved in protein ubiquitination (Dhanoa et al. 2013). Additionally, *ACSL1* (Gene ID: 2180), also known as acyl-CoA synthetase long-chain family member 1, codes for a protein that assists in the biosynthesis of lipids and degradation of fatty acids, and SNPs within this gene have been associated with chronic diseases like diabetes (Manichaikul et al. 2016). Lastly, *ELF1* (Gene ID: 1997), also known as E74 like E26 transformation-specific related transcription factor 1, is an important positive regulator of the Hox cofactor Myeloid ectropic viral integration site 1 (*MEIS1*) which is involved in developmental processes (Xiang et al. 2010).

Out of all of these DMPs and their associated genes, *RUNX1* is the only gene that has previously been associated with OA in humans. Specifically, *RUNX1* was found to be differentially methylated in OA and healthy cartilage in humans, with OA cartilage displaying hypomethylation as compared to healthy cartilage (Fernández-Tajes et al. 2014). As of yet, none of the remaining DMPs and their associated genes have been identified as candidate loci in human OA studies (Alvarez-Garcia et al. 2016; Aref-Eshghi et al. 2015; Delgado-Calle et al. 2013; Fernández-Tajes et al. 2014; García-

Ibarbia et al. 2013; Goldring and Marcu 2012; Iliopoulos et al. 2008; Jeffries et al. 2016; Moazedi-Fuerst et al. 2014; Ramos et al. 2014; Reynard et al. 2014; Rushton et al. 2014a; Saito et al. 2010).

Overall, these findings indicate that some DNA methylation patterns associated with OA are evolutionarily conserved between humans and baboons while others are not. Differences may exist between these two species simply because human studies have not identified all OA related changes in methylation. Alternatively, they may be due to general speciation events that took place during the evolution of these taxonomic groups, to slight differences in the development or manifestation of OA in these species, or artifacts of the experimental design itself. For instance, the sample size of this study (n=10) is rather small, and all individuals included were female. The small number of individuals likely reduced my power to detect potentially important OA related variants, and the inclusion of only one sex may have biased my results such that identified OA variants are actually female specific variants. Thus, in order to better identify candidate epigenetic alterations that underlie variation in knee OA, a larger sample set that includes both sexes should be considered. Nevertheless, using baboons as a model of OA in this study has begun to clarify the evolutionary conservation of this disorder, and future research in this animal model will help provide insight into the development and progression of OA in order to begin designing preventative and therapeutic agents (Cox et al. 2013).

In conclusion, I determined that the 450K array can be used to measure genome-wide DNA methylation in baboon tissues and identify significant associations with complex traits. This is the first study to specifically assess DNA methylation in skeletal

tissues from a nonhuman primate using this method. From an evolutionary perspective, the results of this study reveal DNA methylation variation in one species and in two skeletal tissues, as well as the degree to which the common skeletal condition of OA affects that variation. Some methylation variation is associated with genes that impact skeletal development and maintenance, and this may have direct downstream regulatory and phenotypic effects. Additionally, while some DNA methylation patterns associated with OA in baboons appear to be evolutionarily conserved with humans, others do not. These findings warrant further investigation in a larger and more phylogenetically diverse sample set.

Acknowledgements

Lorena Havill, Erin Sybout, Sophia Johnson, Mel Carless, Kara Peterson, Laura Cox, Kenneth Lange, Jerry Glenn, & Clint Christensen

Department of Genetics, Texas Biomedical Research Institute, San Antonio, TX, USA

Funding

Max and Minnie Tomerlin Voelcker Foundation to L.M. Havill

William and Ella Owens Foundation for Biomedical Research to L.M. Havill

SNPRC Internship Funds to G. Housman

ASU Chapter of Sigma Xi Grant-in-Aid of Research to G. Housman

CHAPTER 3

AN EVOLUTIONARY PERSPECTIVE OF DNA METHYLATION PATTERNS IN SKELETAL TISSUES USING A NONHUMAN PRIMATE MODEL OF OSTEOARTHRITIS

Abstract

Epigenetic factors, such as DNA methylation, play an influential role in the development of the degenerative join disease osteoarthritis (OA). These molecular mechanisms have been heavily studied in humans, and although OA impacts several other animals in addition to humans, few efforts have taken an evolutionary perspective. Here, I explore the evolution of OA epigenetics by assessing how DNA methylation variation relates to knee OA development in a baboon primate model (*Papio spp.*) and by comparing these findings to what is known in humans. Genome-wide DNA methylation patterns were identified in trabecular bone and cartilage of the right distal femora from 56 pedigreed, adult baboons (28 with and 28 without knee OA) using the Illumina Infinium MethylationEPIC BeadChip (EPIC array). Several significantly differentially methylated positions (DMPs) were found between tissue types. Within cartilage tissue, many DMPs were also identified between healthy and OA individuals. Conversely, very few DMPs were identified between disease states in bone tissue. Overall, these finding provide some insight into the etiology of OA. Furthermore, some genes containing DMPs overlap with and display methylation patterns similar to those previously identified in human OA studies, while others genes do not. These results provide insight into the evolutionary conservation of epigenetic mechanisms associated with OA. From an evolutionary

perspective, these results provide evidence for DNA methylation variation in skeletal tissue from one primate species and two skeletal tissues. They also reveal the degree to which the common skeletal condition OA affects this variation.

Key Words

Osteoarthritis, DNA methylation, evolution, epigenome, bone, cartilage, baboon

Introduction

Osteoarthritis (OA) is a chronic and degenerative joint disease. It is characterized by a progressive degradation of cartilage and underlying subchondral bone within a joint (Glyn-Jones et al. 2015) which leads to significant pain and functional limitations of the affected joint. According to the WHO, OA is present in 9.6% of men and 18.0% of women ages 60 or older world-wide. Of those affected, 80% have movement limitations and 25% are unable to perform major daily activities of life (WHO | Chronic Rheumatic Conditions 2016). The CDC further notes that OA of the knee joint is especially prevalent in the USA (Osteoarthritis (OA) | Arthritis | CDC 2016), and it is also one of the leading causes of disability across the globe (Cross et al. 2014). The burden of OA on society demands that researchers identify the factors contributing to and aiding in the development and progression of this disease.

Although significant work has been done in this area, the complete etiology of OA is still unclear. This is because OA pathogenesis appears to be multifactorial, with both genetic and environmental influences (Blagojevic et al. 2010; Cooper et al. 2000; D. T. Felson and Zhang 1998; David T. Felson 2004; Henriksen et al. 2014; Johnson and

Hunter 2014; Jordan et al. 2007; Macrini et al. 2013; Rossignol et al. 2005). Additionally, epigenetic factors, such as DNA methylation which regulates gene expression, are now thought to play a more influential role in the development of degenerative skeletal disorders like OA (Delgado-Calle et al. 2013; den Hollander et al. 2014; Fernández-Tajes et al. 2014; García-Ibarbia et al. 2013; Goldring and Marcu 2012; Iliopoulos et al. 2008; Jeffries et al. 2016; Y. Liu et al. 2013; Loughlin and Reynard 2015; Moazedi-Fuerst et al. 2014; Ramos et al. 2014; Reynard et al. 2014; Rushton et al. 2014a). The investigation of human OA epigenetics in both bone and cartilage tissues has revealed thousands of differentially methylated candidate genes, but whether this epigenetic variation truly contributes to the development of OA and by which pathways remains unknown. Accomplishing such research in humans is limited due to experimentation ethics. Thus, finding a suitable model organism in which tissue collection and direct OA progression assessment are possible is necessary for discovering the mechanisms involved in OA pathogenesis.

Current animal models of OA include mice, rats, rabbits, guinea pigs, dogs, sheep, goats, and horses (Bendele 2001; Kuyinu et al. 2016; Cucchiarini et al. 2016). Because the majority of these animal models do not naturally develop OA, they are limited in their ability to fully inform our understanding of human OA. Most animal models require transgenics, surgical procedures, drug injections, or non-invasive damage to a joint to induce OA, and even then, the physical manifestation of OA in these models only replicates certain stages of human OA (Bendele 2001; Kuyinu et al. 2016). Additionally, in those models that do naturally develop OA, such as guinea pigs, the occurrence of this disease across individuals differs from that in humans. Specifically in guinea pigs, males

have more consistent pathological alterations than females (Bendele 2001), while in humans, females have a higher occurrence of OA than males (Cross et al. 2014).

Conversely, among nonhuman primates, baboons develop knee OA naturally and at rates similar to those observed in humans. Like humans, the prevalence of severe OA in baboons is higher in females than in males (Macrini et al. 2013). Additionally, in both baboons and humans, the occurrence of OA is not an inevitable consequence of aging. For instance, at the Southwest National Primate Research Center (SNPRC), approximately 66% of older baboons develop OA, and the remaining show no distal femur articular cartilage degradation (Cox et al. 2013). This is comparable to the almost two-thirds of Americans (≥65 years old) that develop OA (O'Connor 2006) and the almost one-third of human tissue donors (70-90 years old) that show no manifestations of knee OA (Loeser and Shakoor 2003).

In general, nonhuman primates can serve as important models of disease for humans because they are phylogenetically close to humans. Because baboons also develop and present OA in a manner similar to that observed in humans, baboons may be a more suitable model of OA than those currently used. Furthermore, in captive colonies of baboons, environmental factors can be regulated and controlled, thus enabling more detailed investigations of the molecular mechanisms contributing to OA pathogenesis than can be achieved in humans (Macrini et al. 2013; Cox et al. 2013).

Lastly, because of their evolutionary proximity to humans, using baboons as an animal model of OA will advance the evolutionary understanding of this disease, a perspective that has not been readily explored (Ostrer et al. 2006; Rugg-Gunn et al. 2005). The comparable manifestations of OA between humans and phylogenetically

close primate relatives (baboons) as compared to less similar manifestations of OA between humans and more distantly related animals (Bendele 2001), implies that the potential to develop this disease is somewhat evolutionary conserved across species while also susceptible to change over evolutionary time. Thus, the molecular processes innate to OA development and progression may also be influenced by evolutionary forces. Overall, investigating the molecular processes associated with OA in baboons and comparing how these findings relate to those known in humans, particularly given the fact that the pathogenesis of this disease is similar between both species, will both provide greater insight into the etiology of OA and the evolution of this disease.

For this study, I explored the evolution of OA epigenetics by identifying DNA methylation patterns in femur trabecular bone and cartilage of 56 pedigreed, adult baboons, 28 with and 28 without knee OA, and assessing whether DNA methylation variation is associated with OA in baboons and in a manner similar to that observed in humans.

Methods

Ethics Statement

Nonhuman primate tissue samples included were opportunistically collected at routine necropsy of these animals. No animals were sacrificed for this study, and no living animals were used in this study.

Samples

Baboon (*Papio spp.*) samples come from captive colonies at the SNPRC in the Texas Biomedical Research Institute. These samples are ideal because many environmental factors that influence skeletal development and maintenance (e.g., diet and exposure to sunlight, which influences vitamin D production) are controlled and consistent across individuals. Additionally, these animals have a tracked pedigree, which denotes the genetic relationships among all individuals.

Femora were opportunistically collected at routine necropsy of these animals and stored in -20°C freezers at the Texas Biomedical Research Institute after dissection. These preparation and storage conditions ensured the preservation of skeletal DNA methylation patterns.

Samples include skeletally healthy adult baboons (n=28) and adult baboons with severe osteoarthritis (OA, n=28). Age ranges are comparable between each group, and both sexes are represented (Figure 7, APPENDIX G). This is important as many skeletal features, such as overall bone shape and susceptibility to diseases of skeletal maintenance are sex and age dependent (O'Connor 2006).

Assessment of Osteoarthritis

Classification of adult baboons as having healthy or OA knees was determined through visual examination of the distal femora and macroscopic inspection of the distal articular surface cartilage. Each specimen was assigned an OA severity score. Briefly, Grade 1 is unaffected, Grade 2 is mild OA as indicated by cartilage fibrillation, Grade 3 is moderate OA as indicated by cartilage lesions, and Grade 4 is advanced OA as

indicated by eburnation (Macrini et al. 2013). From this, binary classifications were made such that all healthy adult baboons have 100% Grade 1 on one or both distal femora, and all OA adult baboons have a variable percentage of Grades 3 or 4 on one or both distal femora (Figure 1).

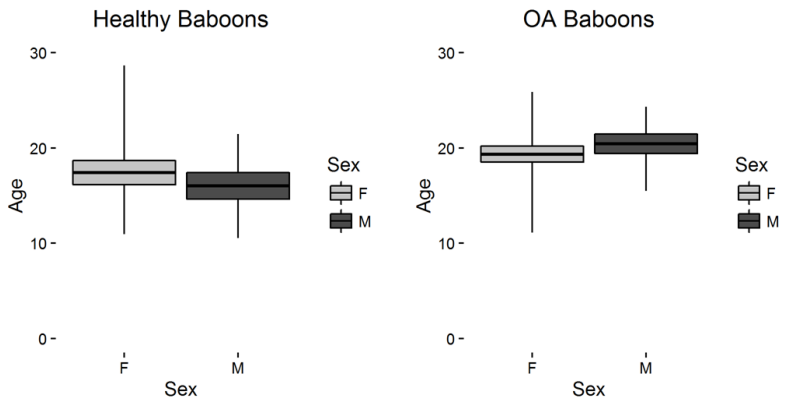

Figure 7. Baboon Sample Set Ages for EPIC Array Osteoarthritis Study.
Box plots depict average ages plus or minus one standard deviation (box), as well as full range of ages (whiskers), for male (M) and female (F) baboons that are skeletally healthy or have OA. For males and females combined, healthy adult baboons (n=28) are 16.90±5.02 years, and OA adult baboons (n=28) are 19.73±3.41 years.

DNA Extraction

DNA was extracted from femoral trabecular bone and cartilage using a phenol-chloroform protocol optimized for skeletal tissues (Barnett and Larson 2012). From the distal femoral condyles, cartilage scrapings were collected using scalpels and processed with a homogenizer, and trabecular bone was collected using coring devices and pulverized into bone dust using a SPEX SamplePrep Freezer/Mill. This region of the joint

was selected because this location is the common site of OA development in baboons and humans. Specifically, cartilage was obtained from the inferior aspect of the medial condyle on the right distal femur. Additionally, bone cores were obtained from a transverse plane through the center of the medial condyle on the right distal femur, such that the articular surface remained preserved. Cortical bone was removed from these cores using a Dremel.

Both tissues are included in this project because they are clinically relevant with respect to disease progression. As such, human skeletal epigenetic studies are based on both trabecular bone and cartilage, so for comparative purposes, it is also important to standardize tissue type. These tissues have distinct functions and occupy distinct portions of the femur. Trabecular bone comprises the internal spongy osseous tissue that contributes to femoral shape morphology, while cartilage comprises the external joint-associated tissue at the proximal and distal ends of femora. Trabecular bone and cortical bone remodeling, which begin before birth and continue throughout life, contribute the development and maintenance of femoral shape (Clarke 2008). However, trabecular bone in growing individuals influences both trabecular and cortical morphology in adulthood (Q. Wang et al. 2011), and this suggests that the epigenetics of trabecular bone may be of more interest initially than that of cortical bone. Lastly, although trabecular bone is not ideal for epigenetic analyses because it contains several cell types (Horvath, Mah, et al. 2015), statistical methods can correct for this heterogeneity.

Cartilage methylation patterns are known to vary between joints and between different sites within a joint (den Hollander et al. 2014; Jeffries et al. 2016; Loughlin and Reynard 2015; Moazedi-Fuerst et al. 2014; Rushton et al. 2014b). Although similar

studies of bone methylation patterns have not been conducted yet, the number and types of cells, and therefore epigenetic signatures, are expected to vary across different portions of the femur. Thus, tissues were collected from the same portion of the femur in order to minimize this variation between samples and comparative groups.

Genome-Wide DNA Methylation Profiling

Genome-wide DNA methylation was assessed using Illumina Infinium MethylationEPIC microarrays (EPIC array). These arrays analyze the methylation status of over 850,000 sites throughout the genome, covering over 90% of the sites on the Infinium HumanMethylation450 BeadChip as well as an additional 350,000 sites within enhancer regions. For each sample, 400ng of genomic DNA was bisulfite converted using the EZ DNA Methylation™ Gold Kit according to the manufacturer's instructions (Zymo Research), with modifications described in the Infinium Methylation Assay Protocol. Following manufacturer guidelines (Illumina), this processed DNA was then whole-genome amplified, enzymatically fragmented, hybridized to the arrays, and imaged using the Illumina iScan system. The array data discussed here are available in APPENDIX F.

Methylation Data Processing

Raw fluorescent data were normalized to account for the noise inherent within and between the arrays themselves. Specifically, I performed a normal-exponential out-of-band (Noob) background correction method with dye-bias normalization (Triche et al. 2013) to adjust for background fluorescence and dye-based biases. This was followed with a between-array normalization method (functional normalization) (Fortin et al.

2014), which removes unwanted variation by regressing out variability explained by the control probes present on the array as implemented in the minfi package in R (Aryee et al. 2014; Fortin et al. 2016) which is part of the Bioconductor project (Huber et al. 2015). This method has been found to outperform other existing approaches for studies that compare conditions with known large-scale differences (Fortin et al. 2014), such as those assessed in this study.

After normalization, methylation values (β values) for each site were calculated as the ratio of methylated probe signal intensity to the sum of both methylated and unmethylated probe signal intensities (Equation 1). These β values range from 0 to 1 and represent the average methylation levels at each site across the entire population of cells from which DNA was extracted (0 =completely unmethylated sites, 1 = fully methylated sites).

Every β value in the Infinium platform is accompanied by a detection p-value, and those with failed detection levels (p-value > 0.05) in greater than 10% of samples were removed from downstream analyses. Additionally, samples in which more than 30% of the β value had a detection p-value > 0.05 were removed from downstream analyses.

The probes on the arrays were designed to hybridize specifically with human DNA, so my use of nonhuman primate DNA required that probes non-specific to the baboon genome, which could produce biased methylation measurements, be computationally filtered out and excluded from downstream analyses. This was accomplished using methods modified from (Hernando-Herraez et al. 2013; Ong et al. 2014).

Briefly, I used blastn (Altschul et al. 1997) to map the 866,837 50bp probes onto the *Papio anubis* genome (Assembly: Panu_2.0, Accession: GCF_000264685.2) using an e-value threshold of e^{-10}. I only retained probes that successfully mapped to the baboon genome, had only 1 unique BLAST hit, targeted CpG sites, had 0 mismatches in 5bp closest to and including the CpG site, and had 0-2 mismatches in 45bp not including the CpG site (APPENDIX A). This filtering retained 209,802 probes.

Additionally, β values associated with cross-reactive probes (McCartney et al. 2016), probes containing SNPs at the CpG site, probes detecting SNP information, probes detecting methylation at non-CpG sites, and probes targeting sites within the sex chromosomes were removed using the minfi package in R (Aryee et al. 2014; Fortin et al. 2016) (Figure 8). This filtering retained a final set of 191,954 probes.

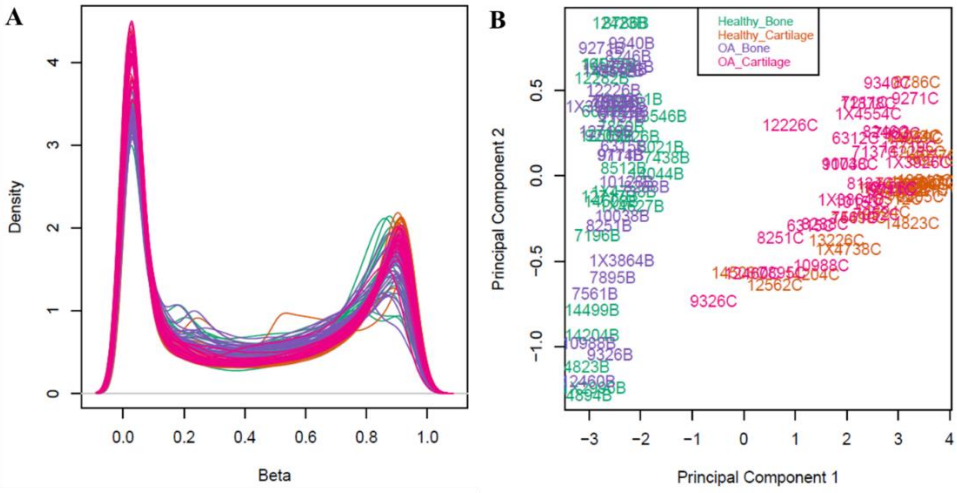

Figure 8. Normalized and Filtered Methylation Data for EPIC Array Baboon Osteoarthritis Study.

(A) Density plots of β values after normalization and probe filtering. (B) Multidimensional scaling plot showing the first two principle components that describe

genome-wide methylation variation after normalization and probe filtering. Each point represents one sample that is either from healthy bone, healthy cartilage, OA bone, or OA cartilage.

Differential Methylation Analyses

Because β values have high heteroscedasticity, they are not statistically valid for use in differential methylation analyses (Du et al. 2010). Thus, M values were calculated and used in these analyses instead (Equation 2).

In order to identify sites that were significantly differentially methylated across comparative groups, I designed and tested generalized linear mixed models (GLMMs) which related the variables of interest to the DNA methylation patterns for each site, while accounting for the effects of additional variables, batch effects, and latent variables (Maksimovic et al. 2016). Sites found to have significant associations were classified as significantly differentially methylated positions (DMPs).

Specifically, a GLMM was used to estimate differences in methylation levels for each of the following contrasts:

1. between bone and cartilage in OA baboons
2. between bone and cartilage in healthy baboons
3. between OA and healthy baboon bone
4. between OA and healthy baboon cartilage
5. among all 4 combinations of tissue type and disease state (healthy bone vs. healthy cartilage vs. OA bone vs. OA cartilage)

Additional variables included in this GLMM were sex, age (years), steady state weight (kg), known batch effects (e.g., array number and position), and unknown latent variables calculated using the iteratively re-weighted least squares approach in the sva package in R (Jaffe and Irizarry 2014; Jeffrey T. Leek et al. 2012; J. T. Leek and Storey 2008; Jeffrey T. Leek and Storey 2007). The 14 latent variables estimated were included to help mitigate any unknown batch and cell heterogeneity effects on methylation variation at each site.

Alternative methods to account for cell heterogeneity exist, but they are specific to whole blood (Jaffe and Irizarry 2014; Morris and Beck 2015), require reference epigenetic data, or are reference free methods (Houseman et al. 2014) that are comparable to the sva method (Kaushal et al. 2015). Out of the known cell types in skeletal tissues (Horvath, Mah, et al. 2015), only chondrocytes and osteoblasts have reference epigenomes available on the International Human Epigenomics Consortium, and these are only for humans, not nonhuman primates. Thus, because no standard method is available to correct for the heterogeneous cell structure in nonhuman primate skeletal tissue, I chose the described sva method.

This GLMM design matrix (*Equation 4*) was fit to the M value array data by generalized least squares using the limma package in R (Ritchie et al. 2015; Phipson et al. 2016; Huber et al. 2015), and the estimated coefficients and standard errors for the defined tissue type and disease status contrasts were computed. Because each baboon contributed both a bone sample and a cartilage sample, an inter-subject correlation was performed to account for these repeat measures (Smyth et al. 2005) and included in the GLMM. Lastly, for each coefficient, an empirical Bayes approach (McCarthy and Smyth

2009; Lönnstedt and Speed 2002; Phipson et al. 2016; Smyth 2004) was used to compute moderated t-statistics, log-odds ratios of differential methylation, and associated p-values adjusted for multiple testing (Benjamini and Hochberg 1995). Significant DMPs for the effect of tissue type, disease status, or both were defined as those having log fold changes in M values corresponding to an adjusted p-value of less than 0.05.

In order to account for genetic relatedness, the coefficients of relatedness (phi2 = 2 x kinship coefficients), or the expected proportions of alleles that are identical by descent between 2 individuals, were computed from a known pedigree using the kinship2 package in R (Therneau et al. 2015). Following this, two new GLMMs were designed and tested using the lmekin function of the coxme package in R (Therneau 2015). The first GLMM regressed methylation levels (M values) against the tissue type and disease status contrast effects while adjusting for other variables (sex, batch effects, latent variables) as fixed effects and kinship (phi2) as a random effect (*Equation 5*) (Zaghlool et al. 2015), and the second performed the same regression with the tissue type and disease status contrast effects removed (*Equation 6*). The log likelihoods of each model were then compared using a chi-square test to determine which model better explained the variation in methylation. For this test, the degrees of freedom were calculated as the absolute difference in the Akaike's information criteria for each model (Mazerolle 2016). When the model containing the tissue type and disease status contrast effects performed significantly better than the alternative model (p-value < 0.05), this confirmed that the site remained a significant DMP for the effects of tissue type, disease status, or both when adjusting for the added effects of kinship. Conversely, when the model containing the tissue type and disease status contrast effects did not perform better than the

alternative model (p-value ≥ 0.05), this indicated that the site was not a significant DMP for the effect of tissue type, disease status, or both when adjusting for the added effects of kinship. In this instance, this site was no longer considered a significant DMP.

Equation 4: methylation ~ tissue type and disease status contrasts + sex + age + weight + batch effects + latent variables

Equation 5: methylation ~ tissue type and disease status contrasts + sex + age + weight + batch effects + latent variables + kinship

Equation 6: methylation ~ sex + age + weight + batch effects + latent variables + kinship

Lastly, I further examined significant DMPs that had at least a 10% change in mean methylation between comparative groups ($\Delta\beta \geq 0.1$), as these may have greater biological impact than others (Hernando-Herraez et al. 2013). The gene ontology (GO) and KEGG pathway enrichment for significant CpGs while taking into account the differing number of probes per gene present on the array was determined using the missMethyl package in R (Geeleher et al. 2013; Young et al. 2010; Ritchie et al. 2015; Benjamini and Hochberg 1995). Significantly enriched (FDR < 0.05) GO biological processes were subsequently summarized using REViGO which removed redundant GO terms (retained only 50% of the full list of significant terms) and visualized the remaining terms in a semantic similarity-based scatterplot (Supek et al. 2011). Semantic similarity was calculated using the simRel score, which is a functional similarity measure that ranges from 0 for terms that have no similarity to 1 for terms with maximum similarity (Schlicker et al. 2006).

In addition to DMPs, differentially methylated regions (DMRs) were also identified between each comparative group using the DMRcate package in R (Peters et al. 2015; Wand and Jones 1994; Duong 2013). This method is only concerned with the spatial proximity of loci examined and is not biased by any annotations associated with these loci. For these analyses, the individual DMP t-statistics, which were derived by fitting the M value array data to a GLMM design matrix (Equation 4) by generalized least squares using the limma package in R (Ritchie et al. 2015; Phipson et al. 2016; Huber et al. 2015), were smoothed across each chromosome using a recommended Gaussian kernel bandwidth of 1000 base pairs with a scaling factor of 2. An expected value of this smoothed estimate with no experimental effects was also modelled using a Satterthwaite approximation (Satterthwaite 1946) in order to calculate a subsequent significance test for each DMP. A default threshold was then applied to p-values adjusted for multiple testing (Benjamini and Hochberg 1995) to identify FDR-corrected significant DMPs. Finally, these significant DMPs were agglomerated together into DMRs based on chromosomal location and such that each DMR contained at least 2 CpG sites that were less than 1000 base pairs apart.

Results

The aim of this study was to use the EPIC array to identify DNA methylation patterns in femur bone and cartilage of baboons, 28 with and 28 without knee OA. In order to do this, I first assessed the effectiveness of the EPIC array in identifying DNA methylation patterns in baboon DNA and of different probe filtering methods.

Alignment of EPIC Array Probes with the Baboon Genome

Probes from the EPIC array were aligned to the baboon genome using methods modified from (Hernando-Herraez et al. 2013; Ong et al. 2014) (APPENDIX A). Out of the 866,837 50bp probes on the array, 209,802 probes map to the baboon genome with e-values less than e^{-10}, have only unique BLAST hits, target a CpG site, and meet the described alignment filter criteria (Figure 9). These probes covered approximately 23,446 genes with an average coverage of 8 probes per gene. Additionally, the retained probes covered a range of locations with respect to genes and CpG islands (APPENDIX H), indicating that these filtered probes maintain a wide distribution throughout the genome. After filtering out cross-reactive probes (Y. Chen et al. 2013), probes containing SNPs at the CpG site, probes detecting SNP information, probes detecting methylation at non-CpG sites, and probes targeting sites within the sex chromosomes a final set of 191,954 probes were retained for downstream analyses.

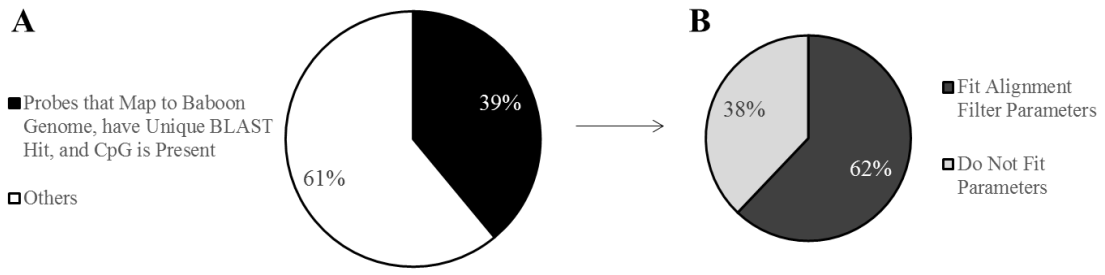

Figure 9. Filtering Effects on EPIC Array Probes for Baboons.

(A) Pie chart showing the percent of EPIC array probes that map to the baboon (*Papio anubis*) genome with e-values < e^{-10}, have only unique BLAST hits, and target a CpG site. Out of 866,837 probes total, 337,818 probes (39%) meet these criteria. (B) Pie chart

showing the percent of probes, out of those that successfully mapped to the baboon genome, that contain 0 mismatches in 5bp of the probe by and including the targeted CpG site and 0-2 mismatches in 45bp of the probe not including the CpG site. Out of the 337,818 mapped probes, 209,802 probes (62%) meet these criteria.

Effectiveness of EPIC Array Probes using Baboon DNA

To determine how effectively the EPIC array probes measured DNA methylation in baboon DNA, I performed Spearman correlation tests between the hybridization efficiency of each probe and parameters defining the alignment quality of each probe to the baboon genome. Specifically, both probe alignment bitscores and percent identity were significantly negatively correlated with probe hybridization efficiency, and probe alignment e-values were significantly positively correlated with probe hybridization efficiency, regardless of filtering criteria (APPENDIX I). Additionally, filtered probes retained a large proportion of successfully hybridized probes (Figure 10).

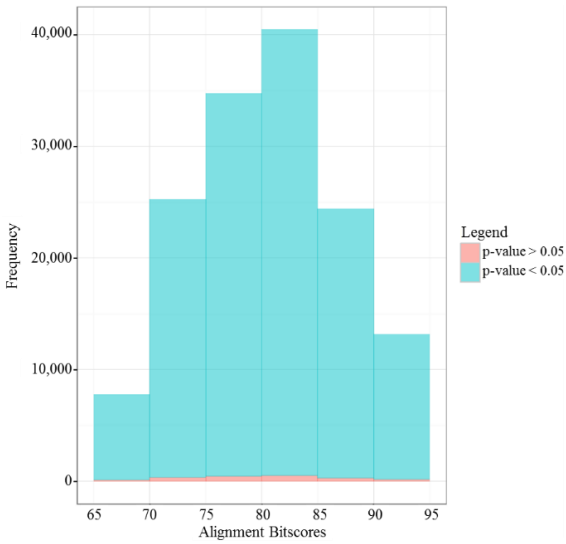

Figure 10. Hybridization Efficiencies of EPIC Array Probes Retained for Baboon Osteoarthritis Study.

Histogram of alignment bitscores for EPIC array probes with detection p-values > 0.05 (red) and < 0.05 (blue). These p-values were averaged across all samples, and probes included meet the alignment filter criteria. For these probes 2,815 had detection p-values > 0.05, and 206,987 had detection p-values < 0.05.

Differential Methylation and Osteoarthritis

Significant DMPs were interrogated from 191,954 sites and identified between disease statuses (OA vs. healthy) and tissue types (bone vs. cartilage), as well as among these variables in combination (Table 4). Accounting for kinship slightly reduces these DMP counts, but does not diminish their distribution across a variety of functional genomic regions and proximities to CpG islands (APPENDIX J). Using a $\Delta\beta \geq 0.1$ threshold substantially decreases the final number of significant DMPs per comparative group (Figure 11, *Table 5*, APPENDIX K). Overall, more DMPs were found between

tissue types than between disease states, and cartilage samples revealed more DMPs between disease states than did bone samples. This pattern holds true for significant DMRs, as well (Table 6, APPENDIX L).

More than half of all DMPs and those DMPs with $\Delta\beta \geq 0.1$ between tissue types in both healthy and OA individuals are hypermethylated in bone as compared to cartilage (Table 4, Table 5), while the global methylation patterns between disease statuses are more complicated. For all DMPs between disease statuses in bone tissues, most are hypermethylated in OA individuals as compared to healthy individuals. Conversely, for all DMPs between disease statuses in cartilage tissues, most are hypomethylated in OA individuals as compared to healthy individuals (Table 4). When just examining those DMPs with $\Delta\beta \geq 0.1$, OA individuals show increased hypomethylation as compared to healthy individuals when examining both bone and cartilage (Table 5).

Table 4. Number of Significant DMPs Identified in the EPIC Array Baboon Osteoarthritis Study.

Differential Methylation	Healthy Bone vs. Healthy Cartilage	OA Bone vs. OA Cartilage	OA Bone vs. Healthy Bone	OA Cartilage vs. Healthy Cartilage	OA Bone vs. Healthy Bone vs. Healthy Cartilage vs. OA Cartilage
Significant (negative)	49,990	43,936	98	11,698	2,143
Not Significant	64,435	71,890	191,570	170,582	186,818
Significant (positive)	77,529	76,128	286	9,674	2,993

Table showing the number of significant DMPs between comparative groups. Results are shown for probes filtered using the alignment criteria, and for these, significant DMPs were identified in all comparative groups.

The variation in methylation patterns at DMPs with $\Delta\beta \geq 0.1$ clusters bone and cartilage tissue types into distinct and separate groups, but does not cluster OA and healthy individuals as effectively (Figure 12). While OA and healthy samples within cartilage differentiate relatively well, except in a couple instances, OA and healthy samples within bone do not clearly differentiate. Additionally, significant DMPs with $\Delta\beta \geq 0.1$ for all comparative groups are associated with several genes that have distinct GO biological processes (Figure 15, APPENDIX M) and KEGG pathway functions (APPENDIX N).

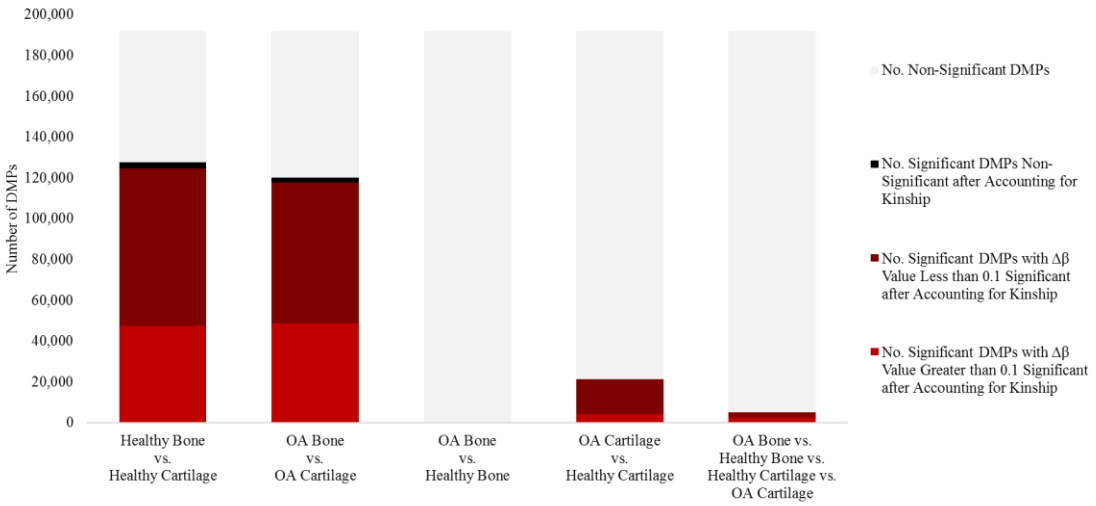

Figure 11. Number of Significant DMPs Identified in the EPIC Array Baboon Osteoarthritis Study.

Bar chart showing the number of significant DMPs between comparative groups. Results include the number of significant DMPs that remained statistically significant after accounting for kinship, the number of significant DMPs that did not remain statistically significant after accounting for kinship, and the number of loci that were not found to be

statistically significant. Those significant DMPs that remained statistically significant after accounting for kinship were additionally split into those that had $\Delta\beta < 0.1$ and those that had $\Delta\beta \geq 0.1$. Approximately 66.4% (127,519) of probes were differentially methylated between bone and cartilage in healthy baboons, 62.6% (120,064) between bone and cartilage in OA baboons, 0.2% (384) between OA and healthy baboons in bone, 11.1% (21,372) between OA and healthy baboons in cartilage, and 2.7% (5,136) between all four combinations of disease state and tissue type. When accounting for genetic relatedness, 2.2% (2,807), 2.0% (2,419), 0.5% (2), 0.8% (172), and 0.5% (24) of the originally identified significant DMPs do not maintain significant methylation associations, respectively. Of those significant DMPs that remained statistically significant after accounting for kinship, only 38.0% (47,386), 41.3% (48,562), 10.2% (39), 20.3% (4,298), and 52.4% (2,676) had $\Delta\beta \geq 0.1$, implying the difference may have had regulatory and biological effects.

Table 5. Number of Significant DMPs with $\Delta\beta \geq 0.1$ Identified in the EPIC Array Baboon Osteoarthritis Study.

Differential Methylation	Healthy Bone vs. Healthy Cartilage	OA Bone vs. OA Cartilage	OA Bone vs. Healthy Bone	OA Cartilage vs. Healthy Cartilage	OA Bone vs. Healthy Bone vs. Healthy Cartilage vs. OA Cartilage
Significant (negative)	20,742	20,562	27	2,998	1,106
Not Significant	144,444	143,165	191,915	187,653	189,275
Significant (positive)	26,768	28,227	12	1,303	1,573

Table showing the number of significant DMPs with $\Delta\beta \geq 0.1$ between comparative groups. Results are shown for probes filtered using the alignment criteria, and for these, significant DMPs were identified in all comparative groups.

Table 6. Number of Significant DMRs Identified in the EPIC Array Baboon Osteoarthritis Study.

	No. Significant DMRs	No. Associated Genes	Average CpGs per DMR (Min-Max)	Average Length of DMR (Min-Max)	Associated Gene Symbols
Healthy Bone vs. Healthy Cartilage	21,538	13,083	4 (2-107)	777bp (3bp-15,989bp)	*COL11A2, ESR1, FBXL8, FEZF2, HOXA11, HOXA11, HOXB3, HOXB5, HOXB6, HOXB-AS3, HOXB-AS3, HOXC4, HSD17B8, HSD17B8, HSF4, MIR219, NRM, PPP1R18, PTPRG-AS1, RING1, RNY4P10, RP11, RUNX3, RXRB, RXRB, SLC39A7, ZNF70P1*
OA Bone vs. OA Cartilage	20,272	12,496	4 (2-107)	774bp (3bp-15,989bp)	*COL11A2, DDAH2, ESR1, FEZF2, HOXA10, HOXA10-AS, HOXA9, HOXB3, HOXB5, HOXB6, HOXB-AS3, HSD17B8, MIR196B, MIR219, PPT2, PPT2-EGFL8, PRRT1, PTPRG-AS1, RING1, RNY4P10, RP1, RUNX3, RXRB, SLC39A7, ZNF70P1*
OA Bone vs. Healthy Bone	84	112	6 (2-17)	489bp (41bp-1,716bp)	*ANKRD13D, C1orf145, DOCK1, GLDN, HOXD8, LIPE, LPPR3, RNF180, SH3RF3, SH3RF3-AS1, SSH3*
OA Cartilage vs. Healthy Cartilage	4,154	2,768	3 (2-40)	507bp (3bp-6,159bp)	*DCN, FBXL8, HOXC6, HOXC9, HOXC-AS1, HSF4, MIR4740, PPP1R18, RP11, SIX1, SMG6, VMP1, ZBTB18*
OA Bone vs. Healthy Bone vs. Healthy Cartilage vs. OA Cartilage	802	597	4 (2-29)	486bp (4bp-5,003bp)	*DCN, DUSP1, ELF1, HOXC6, HOXC9, HOXC-AS1, HOXC-AS2, SIX1, SMG6, ZBTB18*

Table showing the number of significant DMRs identified between comparative groups, the number of unique gene names that overlapped with these regions, the average number of CpGs per DMR along with the minimum and maximum number, average length of DMRs along with the minimum and maximum length, and gene symbol names associated with the top 10 DMRs for each comparative group. For additional details see the Appendix L.

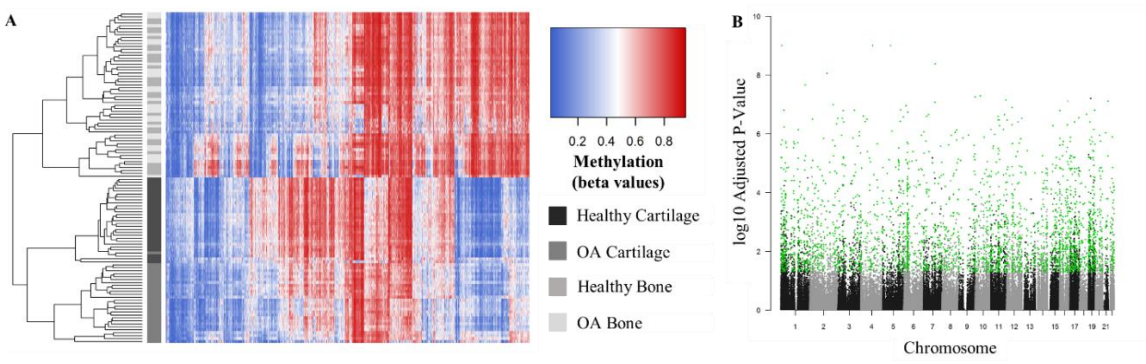

Figure 12. Methylation Levels at Significant DMPs Identified in the EPIC Array Baboon Osteoarthritis Study for All Four Combinations of Disease State and Tissue Type. Heatmaps depicting the DNA methylation levels (β values) of all DMPs with $\Delta\beta \geq 0.1$ between all four combinations of disease state and tissue type (x-axis) in all baboon samples (y-axis, n=112). Red indicates higher methylation at a DMP, while blue indicates lower methylation at a DMP. The dendrogram of all samples (y-axis) clusters individuals based on the similarity of their methylation patterns. Bone and cartilage tissues form distinct clusters, and OA and healthy samples within cartilage separate relatively well. However, these methylation patterns do not cluster OA and healthy samples within bone into distinct groups. See Figure 13 for additional information. (B) Manhattan plot showing the log10 adjusted p-values of all positions examined between all four combinations of disease state and tissue type in all baboon samples (n=112). Samples highlighted in green are the significant DMPs with $\Delta\beta \geq 0.1$ between all four combinations of disease state and tissue type that are displayed in the heatmap. See Figure 14 for additional information.

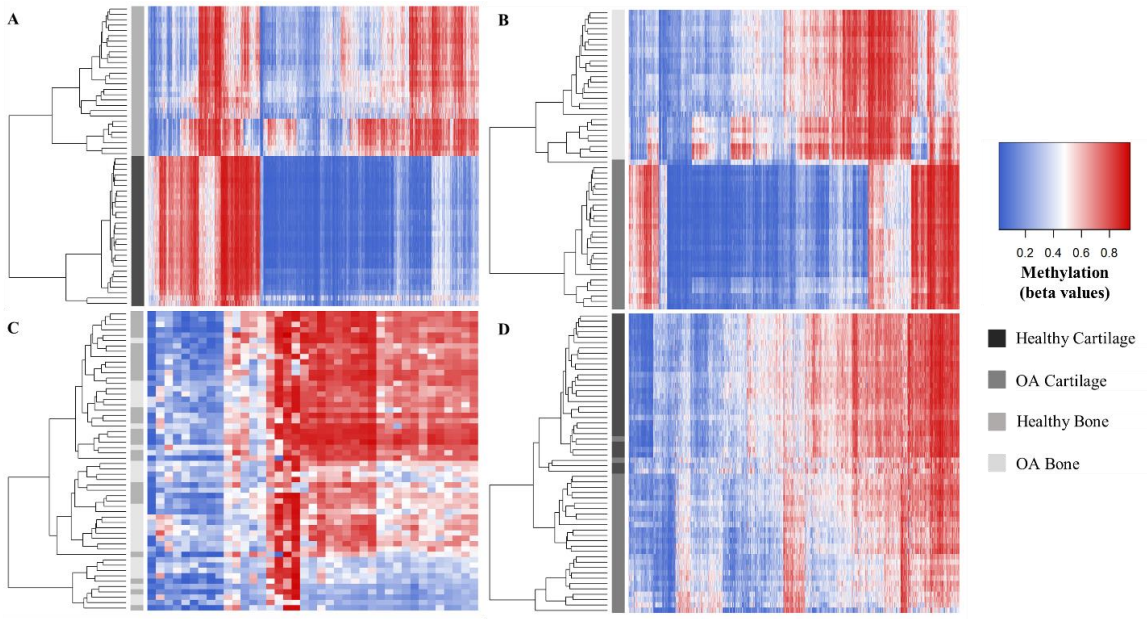

Figure 13. Heatmaps of Methylation Levels at Significant DMPs Identified in the EPIC Array Baboon Osteoarthritis Study for Several Disease State and Tissue Type Comparisons.

Heatmaps depicting the DNA methylation levels (β values) of (A) the top 15,000 DMPs with $\Delta\beta \geq 0.1$ between bone and cartilage (x-axis) in healthy baboons (y-axis, n=56), (B) the top 15,000 DMPs with $\Delta\beta \geq 0.1$ between bone and cartilage (x-axis) in OA baboons (y-axis, n=56), (C) all DMPs with $\Delta\beta \geq 0.1$ between OA and healthy baboons (x-axis) in bone tissues (y-axis, n=56), and (D) all DMPs with $\Delta\beta \geq 0.1$ between OA and healthy baboons (x-axis) in cartilage tissues (y-axis, n=56). Red indicates higher methylation at a DMP, while blue indicates lower methylation at a DMP. The dendrogram of all samples (y-axis) clusters individuals based on the similarity of their methylation patterns. Bone and cartilage tissues form distinct clusters, and OA and healthy samples within cartilage separate relatively well. However, these methylation patterns do not cluster OA and healthy samples within bone into distinct groups.

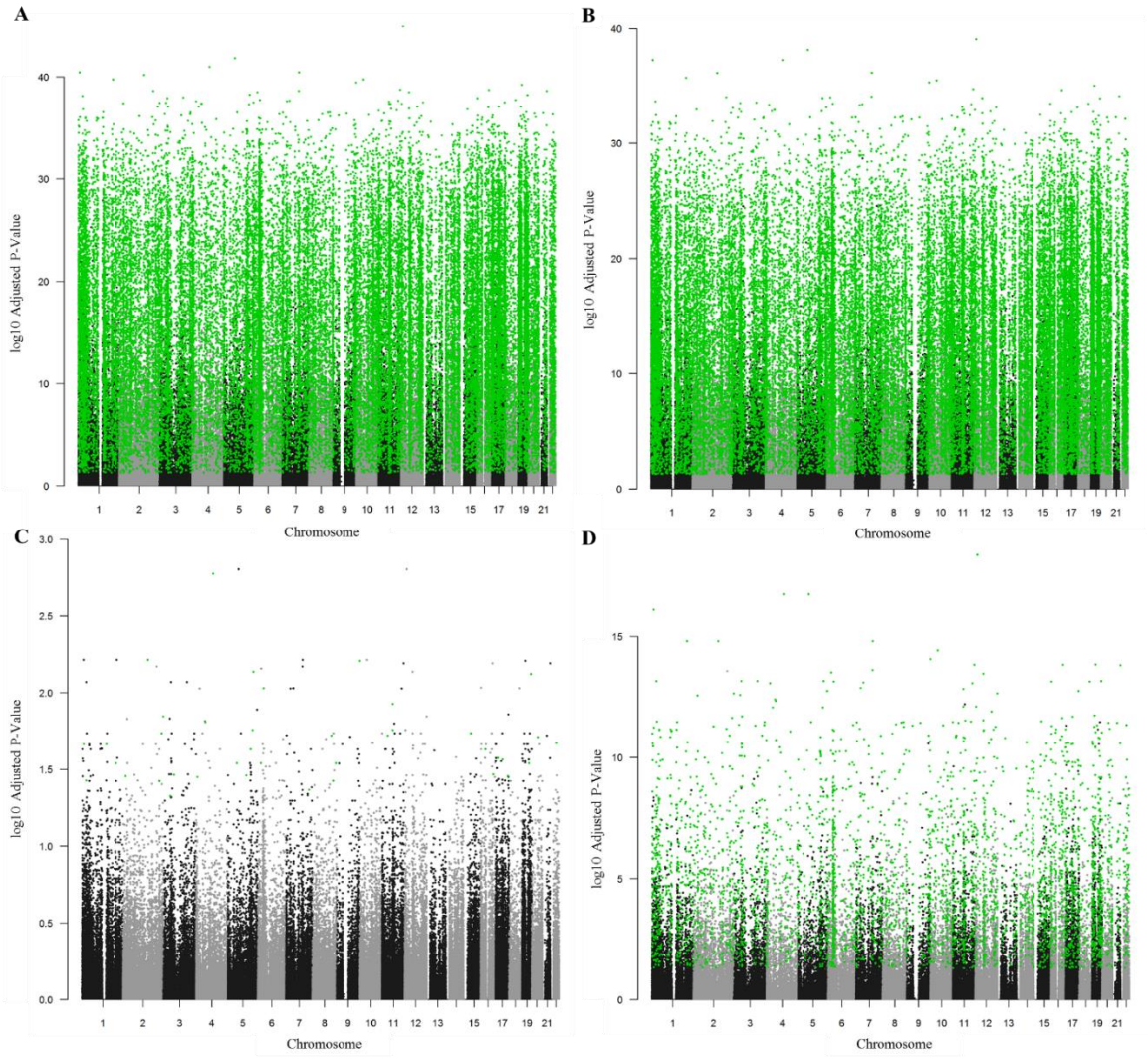

Figure 14. Manhattan plots of Significant DMPs Identified in the EPIC Array Baboon Osteoarthritis Study for Several Disease State and Tissue Type Comparisons.

Manhattan plot showing the log10 adjusted p-values of all positions examined between (A) bone and cartilage in healthy baboons (n=56), (B) bone and cartilage in OA baboons (n=56), (C) OA and healthy baboons in bone tissues (n=56), (D) OA and healthy baboons in cartilage tissues (n=56). Samples highlighted in green are the significant DMPs with $\Delta\beta \geq 0.1$ between each comparative group that are displayed in the heatmaps in Figure 13.

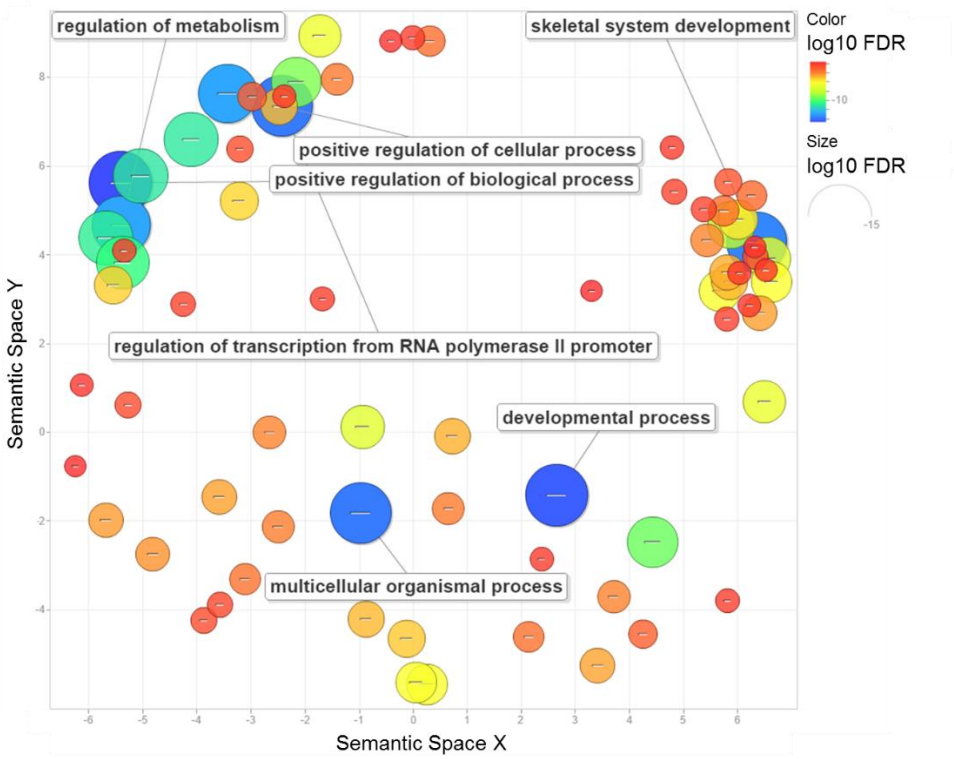

Figure 15. GO Biological Processes Enriched for Significant DMPs All Four Combinations of Disease State and Tissue Type in Baboons.

Multidimensional scaling plot summarizing the GO biological process terms that are significantly enriched (FDR < 0.05) for significant DMPs with $\Delta\beta \geq 0.1$, taking into account the differing number of probes per gene present on the EPIC array. DMPs were identified between all four combinations of disease state and tissue type. REViGO was used to remove redundant GO terms (retained only 50% of the full list of significant terms, see Appendix M) and to visualize the remaining terms in a semantic similarity-based scatterplot (Supek et al. 2011). Semantic similarity was calculated using the simRel score, which is a functional similarity measure that ranges from 0 for terms that have no similarity to 1 for terms with maximum similarity (Schlicker et al. 2006). These pairwise semantic similarity scores are then plotted in multidimensional scaling space such that

similar GO terms are located close to one another in the plot. The color and size of each GO term are based on the log10 FDR value, and some of the most significant GO terms have their descriptions provided in the plot. See Figure 16 for additional information.

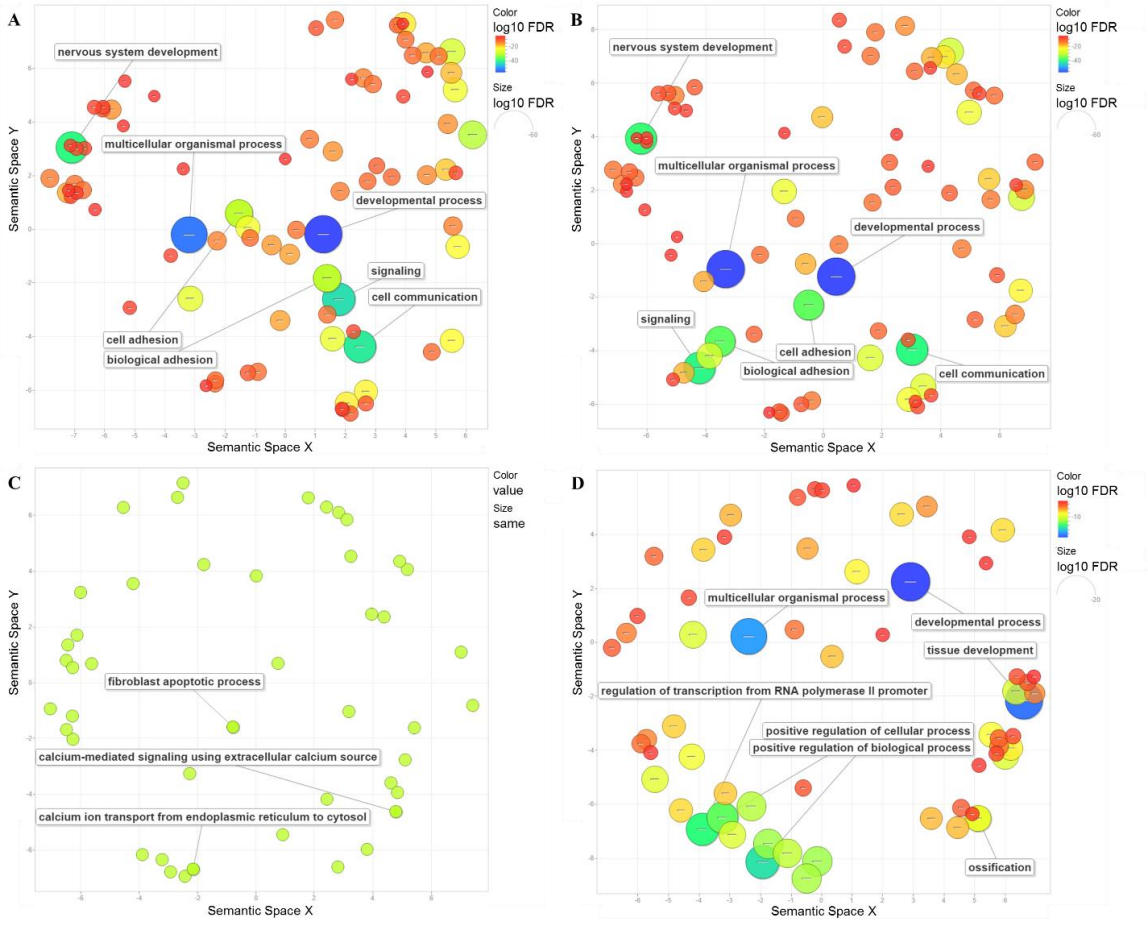

Figure 16. GO Biological Processes Enriched for Significant DMPs Identified in Several Disease State and Tissue Type Comparisons in Baboons.

Multidimensional scaling plot summarizing the GO biological process terms that are significantly enriched (FDR < 0.05) for significant DMPs with $\Delta\beta \geq 0.1$, taking into account the differing number of probes per gene present on the EPIC array. DMPs were identified between bone and cartilage in OA baboons (A), between bone and cartilage in

healthy baboons (B), between healthy and OA baboon bone (C), and between healthy and OA baboon cartilage (D). For OA versus healthy bone, no GO categories significant at 5% FDR were identified, so GO functions with p-values < 0.05 were included in plot. See Figure 8 for information on REViGO methods for making figures.

When comparing the DMPs with $\Delta\beta \geq 0.1$, DMRs, GO functions, and KEGG pathways that were identified between each comparative group, distinct patterns were identified. Out of the 54,302 unique DMPs with $\Delta\beta \geq 0.1$ identified between bone and cartilage in healthy and OA individuals, 77% (41,646) overlap. Similarly, out of the 24,079 unique DMRs identified between bone and cartilage in healthy and OA individuals, 71% (17,085) overlap. Again, this pattern holds up for GO functions, in which 78% (1,118 out of 1,427) overlap, and KEGG pathways, which are almost completely coincidental (289 out of 296 identical) (Figure 17). Thus, the locations and functional associations of differential methylation between bone and cartilage tissues are very similar in both healthy and OA individuals.

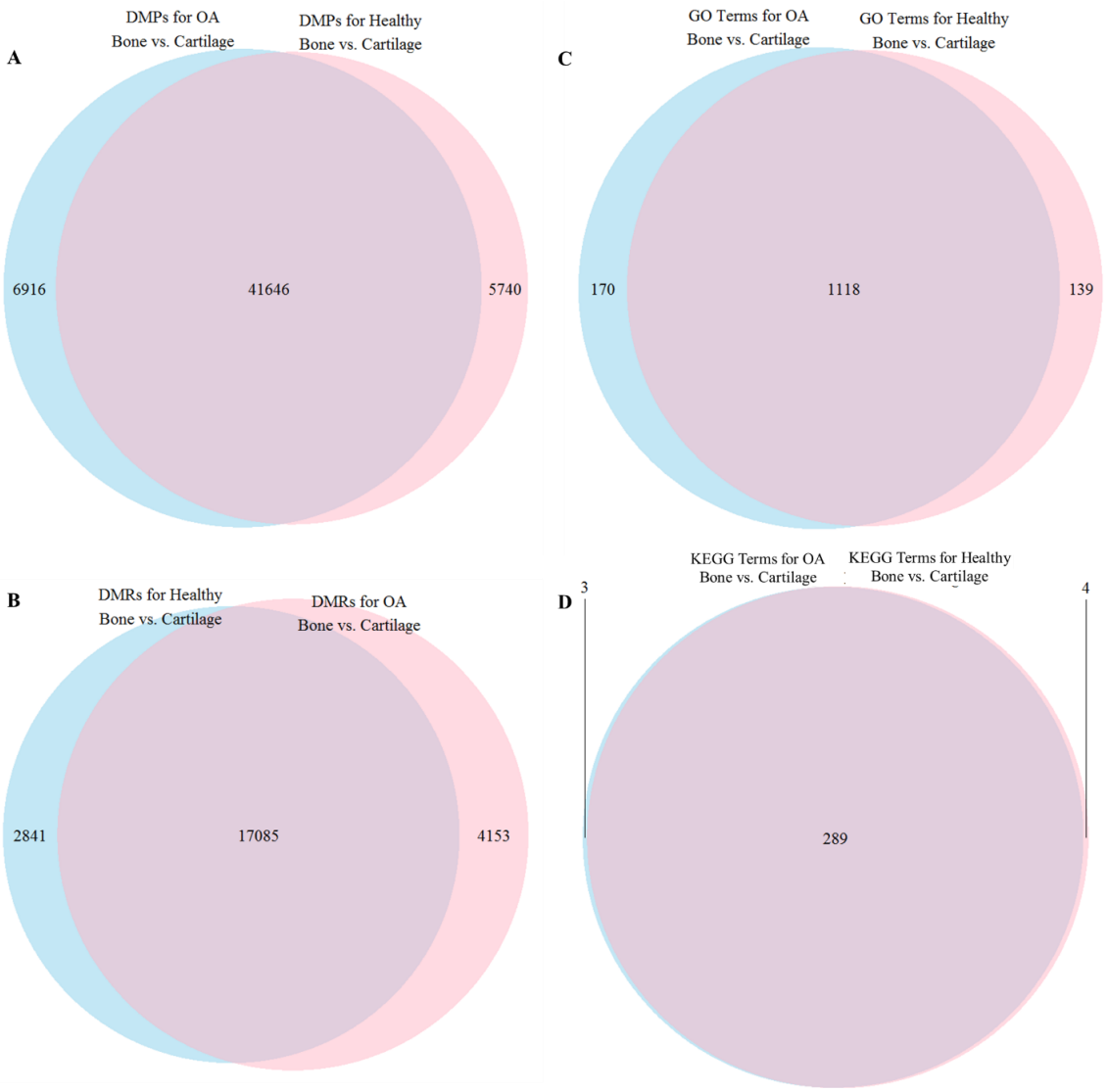

Figure 17. Overlap in Differential Methylation Findings between Bone and Cartilage of Healthy versus OA Baboons.

Venn diagrams showing the overlap in significant DMPs with $\Delta\beta \geq 0.1$ between bone and cartilage in healthy versus OA baboons (A), the overlap in significant DMRs between bone and cartilage in healthy versus OA baboons (B), the overlap in GO biological process terms that are significantly enriched for DMPs with $\Delta\beta \geq$ than 0.1 between bone and cartilage in healthy versus OA baboons (C), and the overlap in KEGG pathways that

are significantly enriched for DMPs with $\Delta\beta \geq 0.1$ between bone and cartilage in healthy versus OA baboons (D).

Conversely, out of the 4,332 unique DMPs with $\Delta\beta \geq 0.1$ identified between OA and healthy individuals in bone and cartilage tissues, only 0.1% (5) are identical. Similarly, out of the 3,547 unique DMRs identified between OA and healthy individuals in bone and cartilage tissues, 0.1% (4) are identical. This pattern continues for GO functions, in which 1.9% (13 out of 692) are identical, but not for KEGG pathways, in which all pathways identified in bone are encompassed within those identified in cartilage (Figure 18). Thus, the locations of differential methylation between healthy and OA individuals are very different in bone versus cartilage tissues. This corresponds with different biological functional associations. However, in spite of these differences, the molecular pathways through which the differential methylation between healthy and OA individuals impacts functional differences may be similar, with those identified in bone completely subsumed by those identified in cartilage. Overall, this suggests while the proximal changes in methylation and their downstream functional impacts associated with the development of OA are different in each tissue, the pathways connecting these are somewhat similar across tissues.

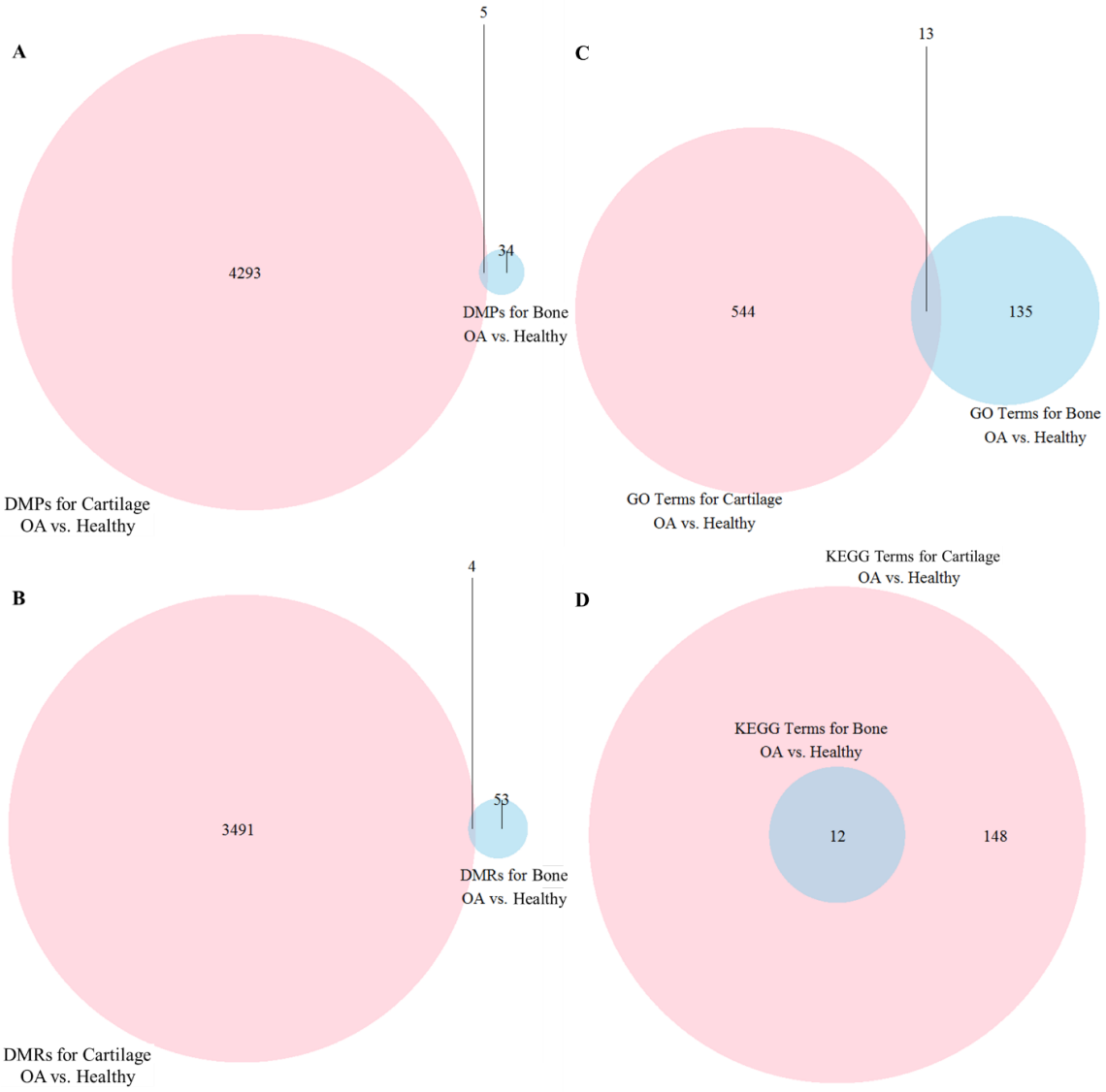

Figure 18. Overlap in Differential Methylation Findings between Healthy and OA Baboons when Using Bone versus Cartilage.

Venn diagrams showing the overlap in significant DMPs with $\Delta\beta \geq 0.1$ between OA and healthy baboons when assessing bone versus cartilage tissues (A), the overlap in significant DMRs between OA and healthy baboons when assessing bone versus cartilage tissues (B), the overlap in GO biological process terms that are enriched for DMPs with $\Delta\beta \geq 0.1$ between OA and healthy baboons when assessing bone versus cartilage tissues

(C), and the overlap in KEGG pathways that are significantly enriched for DMPs with $\Delta\beta \geq 0.1$ between OA and healthy baboons when assessing bone versus cartilage tissues (D). No significantly enriched GO terms were found for DMPs with $\Delta\beta \geq 0.1$ between OA and healthy baboons when assessing bone, so those with p-values < 0.05 are displayed.

A subset of the DMPs identified between OA and healthy baboons were associated with genes that overlap with those previously identified as being differentially methylated in human OA studies, and some genes identified as being differentially methylated in human OA studies were not identified in the current baboon OA study (Table 7, Table 8, APPENDIX O, APPENDIX P). Of those genes that do overlap between humans and baboons, some show methylation pattern that are identical between species, while others have opposing patterns (Alvarez-Garcia et al. 2016; Aref-Eshghi et al. 2015; Delgado-Calle et al. 2013; Fernández-Tajes et al. 2014; García-Ibarbia et al. 2013; Goldring and Marcu 2012; Iliopoulos et al. 2008; Jeffries et al. 2016; Moazedi-Fuerst et al. 2014; Ramos et al. 2014; Reynard et al. 2014; Rushton et al. 2014a; Saito et al. 2010). Specific examples of this are as follows.

Table 7. Overlap of Differential Methylation Associations from Human and Baboon Osteoarthritis Studies.

Previous Human OA Methylation Findings	Current Study Baboon OA Methylation Findings	Genes
Hypermethylated	Hypermethylated	*ERBB2, HOXA13*, IL11RA, IRX3, IRX5, KRT5, MARVELD1, MMP16, PRDM8, PRTFDC1, PTHLH, SPON1, TBX4*, TGFBR2, VAX2, ZADH2*
	Hypomethylated	*BMP2, BMP7, CD28, COL13A1, COL14A1, COL15A1, DLX5, FLJ42709, HOXC4, HOXD8, HS3ST3A1, INSR, NOL4, PPFIA1, SLITRK5, SUSD1, WNT11*
	Conflicting	*ACVR1B, ADAMTS2, ALX4, APCDD1, ATOH8, BCL11B, C10orf11, CCBE1, CDH11, CDH13, CDKN2B, COL5A1, CSNK1E, CYR61, DISP1, DLX6AS*, DYSF*, EDNRB, EGFR, FERD3L, FGD4, FGFR2, FGFRL1, FOXF2, FST, FZD10, GRB10, HDAC11, HMGA2, ID4, KCNQ1, KDM4B, KLF4, LSP1, MAFB, MAFF, MN1, MSI2, MSX1, MSX2, NCOR2, NID2, NPAS2, NPAS3, NRG1, NRP2, PRKAG2, PTPN14, RBP1, RORC, SIX2, SLC15A1, SMAD3, SMAD7, SORBS2, SSBP3, THSD4, TIMP3, TMEM200A, TNS1, TRHDE, VASN*
Hypomethylated	Hypomethylated	*ACVR1, AGPAT9, ALOX5AP, ARRDC2, BCAT1, BCL6, C10orf90, CHD1L, COL7A1, CP, CTSZ, DUSP1, ENG, FLJ43663*, LOC148696*, MLLT10, MMP13, MT2A, PGS1, S100A10, SEMA4D, SIT1, SLC39A7*, SRGN, WIPF1, NA*, NA**
	Hypermethylated	*CNGB1, DUSP5, GLI3, LEPR, MSH3, ODZ4, RARRES1, ROBO2, SYNJ2, UACA*
	Conflicting	*ACTRT2, ADAMTS4, ADAMTS5, ARAP1, ARHGAP9, ATXN1*, ATXN7L1, BEST3, BMP6, CD82, CHST11, CHSY1*, COL6A3, CPA1, CPNE2, CRISPLD2, ECM1, ERGIC1*, ETV6, FHAD1, FILIP1, FOXP1, FTO, GLIS1, GLIS3, H2AFY, HDAC4, HRNBP3, KDM2B, LAMB3, LDLRAD3, LMO7, LPP, LRRFIP1*, MAP3K4, MIR548H4, MTHFD1, MYO18A, NAV2*, NEK7, NFIL3*, NIN, OSBPL10, PMEPA1, RUNX1*, RUNX2, RUNX3, SHANK2, SLC7A5, SMOC2, TGFBR3, THRB*, TIMP2, TNXB, TTLL5*, WWTR1*
Conflicting	Hypermethylated	*ABL1*
	Hypomethylated	*NFATC1, SLC2A1*
	Conflicting	*CREB5, NPR3*, RORA, SPRY4, TGFB2, TRIO*

Details on the previous findings of differential methylation associations with human OA that do overlap with the current findings of differential methylation association with baboon OA. Table outlines several genes that were found to be differentially methylated in human OA studies, whether significant differential methylation with a $\Delta\beta \geq 0.1$ was also identified in the current baboon OA study, and whether the baboon methylation pattern (hypermethylation vs. hypomethylation) was similar to that in humans. Conflicting methylation for previous human studies means that different studies reported both hypermethylation and hypomethylation of a gene and for the current baboon study

means that some sites associated with a particular gene are significantly hypermethylated while others are significantly hypomethylated. All of the genes listed had at least 1 site associated with it that was tested on the EPIC array. Additional details can be found in Appendix O and Appendix P. Normal Font = genes significantly differentially methylated in OA vs. healthy in only cartilage, Italicized Font = gene significantly differentially methylated in OA vs. healthy in only bone, * = gene contains significant DMP that is identical to one identified in previous study.

Table 8. Non-Overlapping Differential Methylation Associations between Human and Baboon Osteoarthritis Studies.

Previous Human OA Methylation Level Findings	Current Study Baboon OA Methylation Level Findings	Genes
Hypermethylated	No significant differential methylation	*ADORA3, BIN2, CD5, COL18A1, FOXF1, IGF2AS, PLCXD3, TUBAL3, YPEL1, ZNF549, ADAMTS14, ARRDC4, C10orf116, C1orf229, C3orf35, C5orf4, CACNA2D1, CLDN5, CNTN1, COL9A3, CSPG5, CTSA, CYYR1, DPYS, EPO, FAM163A, FBXO39, FGF4, FLT4, FOXI2, GALR1, GLRB, GNMT, GRIN3A, HBA1, HCN4, HIST1H4H, HLA-G, HTRA4, IRX4, KDR, LAMA1, LTBR, NKX2-2, NTSR1, PAK7, PCDHB1, PDE10A, PDE1C, PHYHIPL, PKD2L1, PRRT2, PRSS22, PTPN6, PTPRO, RBL1, RBP4, RPS6KB1, SALL3, SLC27A4, SLC30A2, SNX31, SOX11, SOX17, SPEF2, STEAP4, THAP10, THBD, TMEM150A, TRIM58, TYSND1, WNT2, ZFP28, ZMYND19, ZNF454, ZNF667, ZNF678*
Hypomethylated	No significant differential methylation	*A1CF, AIP, ARFRP1, BPIL1, BST2, C13orf16, C14orf38, C8orf34, CD59, CLCN7, CLPP, COL20A1, DEFB129, DNTT, GTF2H3, GUCA1A, HINT1, IL18, IL32, JPH2, MTRR, NAA25, NTRK3, NUDCD3, OR11A1, OR51S1, PAX8, PIP5KL1, PTPN11, RB1CC1, SAA3P, SSU72, STK35, SYNPO2L, TMEM156, ANKRA2, ARHGEF4, BLMH, BMP1, C10orf81, COL22A1, CSNK1A1L, CSRP3, CXCR6, DNASE1L1, EDARADD, EFCAB6, EIF2B1, EMR2, FAM5C, FCGR2A, IL8, LCN2, LTA, LY9, MS4A1, NLRP10, NR1I3, ORMDL2, RBP2, SCAMP1, SMCR8, SP140, TMEM86B, TMIGD2, WNT8A, ZNF180*

Details on the previous findings of differential methylation associations with human OA that do not overlap with the current findings of differential methylation association with baboon OA. Table outlines several genes that were found to be differentially methylated

in human OA studies but were not found to have any significant differential methylation in the current baboon OA study, and if so whether the methylation pattern (hypermethylation vs. hypomethylation) was similar to that in humans. Additional details can be found in Appendix P.

TBX4 displays similar OA hypermethylation patterns in human (Alvarez-Garcia et al. 2016; Fernández-Tajes et al. 2014) and baboon cartilage (Figure 19). In baboon cartilage, out of the 24 CpG sites associated with *TBX4* that were examined, 8 DMPs were identified. All of these were hypermethylated in OA, and 1 had a $\Delta\beta \geq 0.1$ (APPENDIX Q). Additionally, 2 DMRs were identified in close proximity to this gene – one that is 1880bp long and contains 5 CpG sites (hg19 chr17:59533844-59535723) and one that is 259bp long and contains 2 CpGs (hg19 chr17:59532306-59532564). Conversely, in bone, no DMPs or DMRs were found between healthy and OA individuals. (APPENDIX Q).

HOXD8 displays similar OA hypermethylation patterns in humans (Delgado-Calle et al. 2013) and baboon bone but opposing OA methylation patterns in baboon cartilage (Figure 20). In baboon bone, out of the 26 CpG sites associated with *HOXD8* that were examined, 4 DMPs were identified. All of these were hypermethylated in OA, but none had $\Delta\beta \geq 0.1$ (APPENDIX Q). Additionally, 1 DMR that is 1716bp long and contains 12 CpGs (hg19 chr2:176993841-176995556) was identified in close proximity to this gene, and 1 CpG site (cg15520279) was identical between this study and human OA studies. Conversely, in baboon cartilage, 6 DMPs were identified. All of these were hypomethylated in OA, and all had $\Delta\beta \geq 0.1$ (APPENDIX Q). Additionally, 2 DMRs

were identified in this gene – one that is 753bp long and contains 3 CpGs (hg19 chr2:176993089-176993841) and one that is 730bp long and contains 2 CpGs (hg19 chr2:176995556-176996285) (APPENDIX Q).

LEPR displays opposing OA methylation patterns between humans with hypomethylation (Aref-Eshghi et al. 2015) and baboons with hypermethylation (Figure 21). In baboon cartilage, out of the 20 CpG sites associated with *LEPR* that were examined, 3 DMPs were identified. All of these were hypermethylated, and 1 had $\Delta\beta \geq 0.1$ (APPENDIX Q). Similarly, in baboon bone, 1 DMP was identified. This was hypermethylated but did not have a $\Delta\beta \geq 0.1$ (APPENDIX Q). No DMRs were identified in close proximity to this gene for baboon cartilage or bone.

Lastly, *RUNX1* displays opposing OA methylation patterns between humans with hypomethylation (Alvarez-Garcia et al. 2016; Fernández-Tajes et al. 2014) and baboon cartilage with a mixture of hypo- and hyper-methylated sites (Figure 22). In baboon cartilage, out of the 60 CpG sites associated with *RUNX1* that were examined, 16 DMPs were identified of which 9 were hypomethylated and 7 were hypermethylated. Additionally, 6 of the hypomethylated DMPs had a $\Delta\beta \geq 0.1$, while only 1 of the hypermethylated DMPs had a $\Delta\beta \geq 0.1$ (APPENDIX Q). Regarding DMRs, one that is 1198bp long and contains 10 CpGs (hg19 chr21:36258497-36259694) was found to be hypomethylated in close proximity to this gene, and one that is 489bp long and contains 6 CpGs (hg19 chr21:36421467-36421955) was found to be hypermethylated in close proximity to this gene. Finally, 1 CpG site (cg13030790), which was hypomethylated in OA baboon cartilage, was identical between this study and human OA studies.

Conversely, in bone, no DMPs or DMRs were found between healthy and OA individuals (APPENDIX Q).

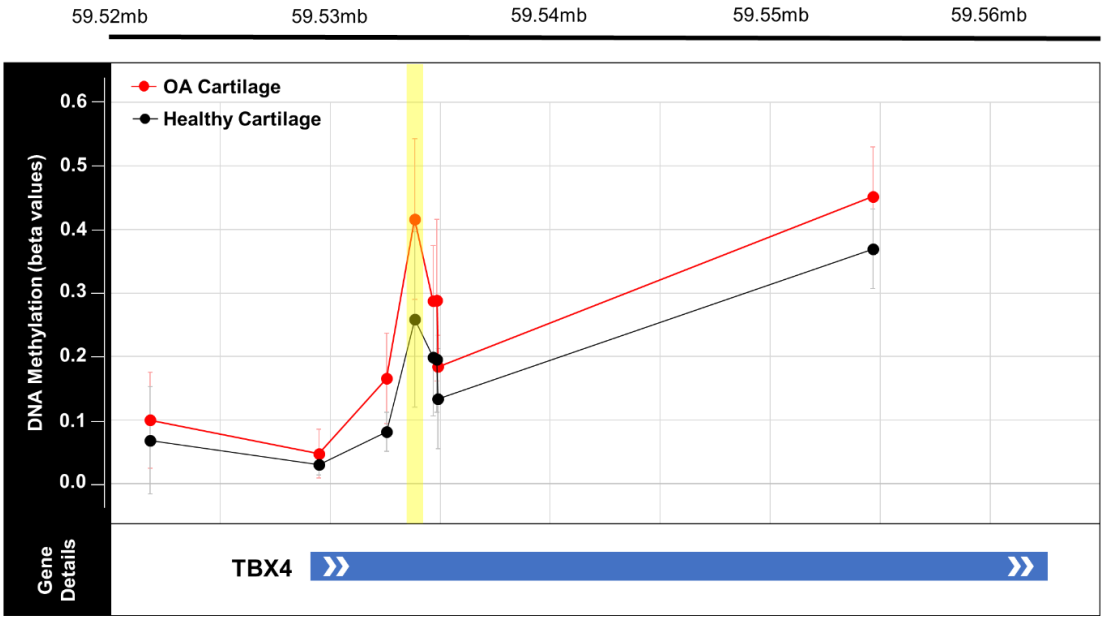

Figure 19. Methylation Levels Across *TBX4* in Healthy and OA Baboon Cartilage. Plot of the methylation levels of significant DMPs across the *TBX4* gene (hg19 chr17:59529134-59562471). Plot shows the average β values for each DMP with error bars indicating 1 standard deviation in each direction for each comparative group (red = OA baboon cartilage, black = healthy baboon cartilage). DMP chromosomal position in relation to the *TBX4* gene is also depicted. *TBX4* in baboon cartilage displays similar OA hypermethylation patterns as those observed in humans (Alvarez-Garcia et al. 2016; Fernández-Tajes et al. 2014). In baboon cartilage, out of the 24 CpG sites associated with *TBX4* that were examined, 8 DMPs were identified. All of these were hypermethylated in OA, and 1 had a $\Delta\beta \geq 0.1$ (highlighted in yellow). Additionally, 2 DMRs were identified in close proximity to this gene – one that is 1880bp long and contains 5 CpG sites (hg19

chr17:59533844-59535723) and one that is 259bp long and contains 2 CpGs (hg19 chr17:59532306-59532564). Conversely, in bone, no DMPs or DMRs were found between healthy and OA individuals. See Appendix Q for additional information.

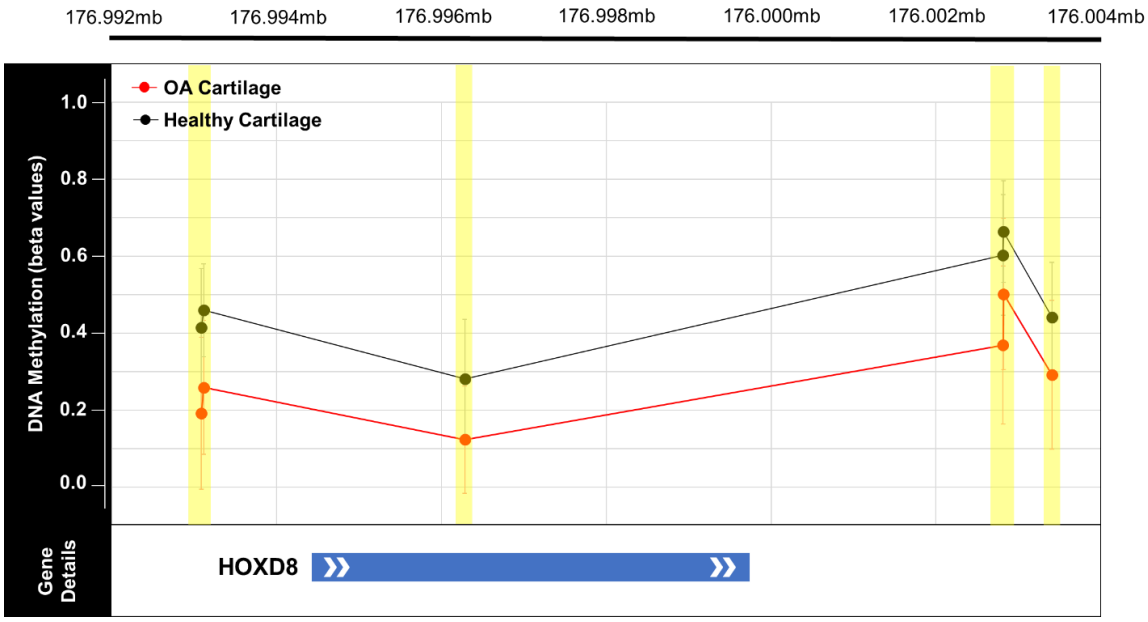

Figure 20. Methylation Levels Across *HOXD8* in Healthy and OA Baboon Cartilage. Plot of the methylation levels of significant DMPs across the *HOXD8* gene (hg19 chr2: 176994422-176997423). Plot shows the average β values for each DMP with error bars indicating 1 standard deviation in each direction for each comparative group (red = OA baboon cartilage, black = healthy baboon cartilage). DMP chromosomal position in relation to the *HOXD8* gene is also depicted. *HOXD8* in baboon cartilage is hypomethylated in OA as opposed to the OA hypermethylation patterns which have been observed in humans (Alvarez-Garcia et al. 2016; Fernández-Tajes et al. 2014). In baboon cartilage, out of the 26 CpG sites associated with *HOXD8* that were examined, 6 DMPs were identified. All of these were hypomethylated in OA, and all had $\Delta\beta \geq 0.1$.

Additionally, 2 DMRs were identified in close proximity to this gene – one that is 753bp long and contains 3 CpGs (hg19 chr2:176993089-176993841) and one that is 730bp long and contains 2 CpGs (hg19 chr2:176995556-176996285). Conversely, OA methylation patterns in baboon bone match those observed in humans. See Appendix Q for additional information.

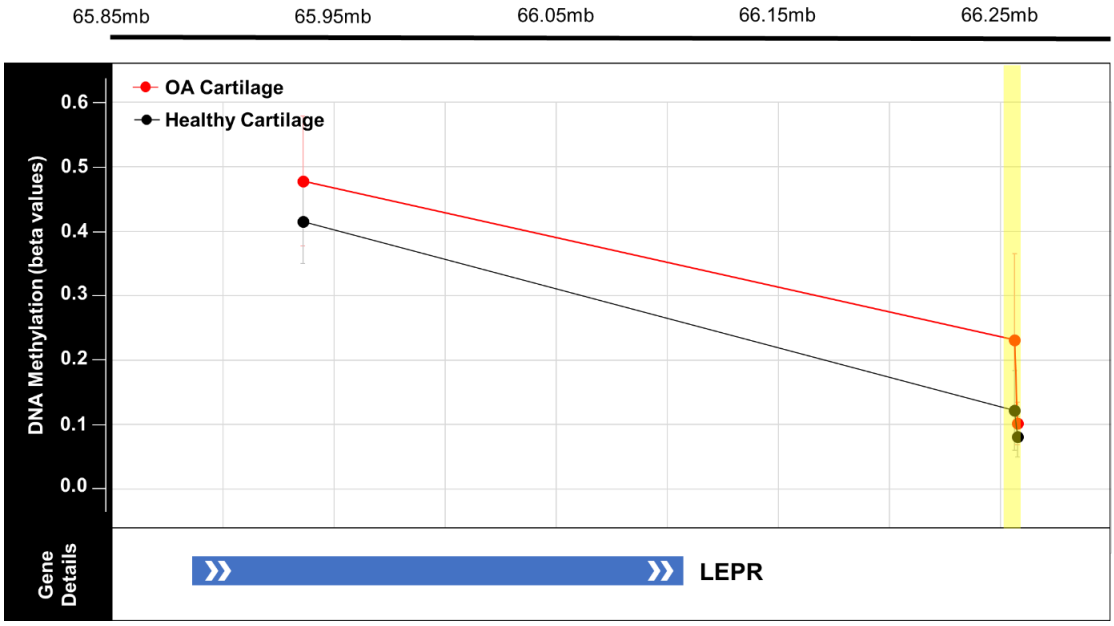

Figure 21. Methylation Levels Across *LEPR* in Healthy and OA Baboon Cartilage. Plot of the methylation levels of significant DMPs across the *LEPR* gene (hg19 chr1:65886335-66103176). Plot shows the average β values for each DMP with error bars indicating 1 standard deviation in each direction for each comparative group (red = OA baboon cartilage, black = healthy baboon cartilage). DMP chromosomal position in relation to the *LEPR* gene is also depicted. *LEPR* in baboon cartilage displays OA hypermethylation which opposes the OA hypomethylation patterns observed in humans (Aref-Eshghi et al. 2015). In baboon cartilage, out of the 20 CpG sites associated with

LEPR that were examined, 3 DMPs were identified. All of these were hypermethylated, and 1 had $\Delta\beta \geq 0.1$. Similarly, in baboon bone, 1 DMP was identified which is also hypermethylated in OA. No DMRs were identified in close proximity to this gene for baboon cartilage or bone. See Appendix Q for additional information.

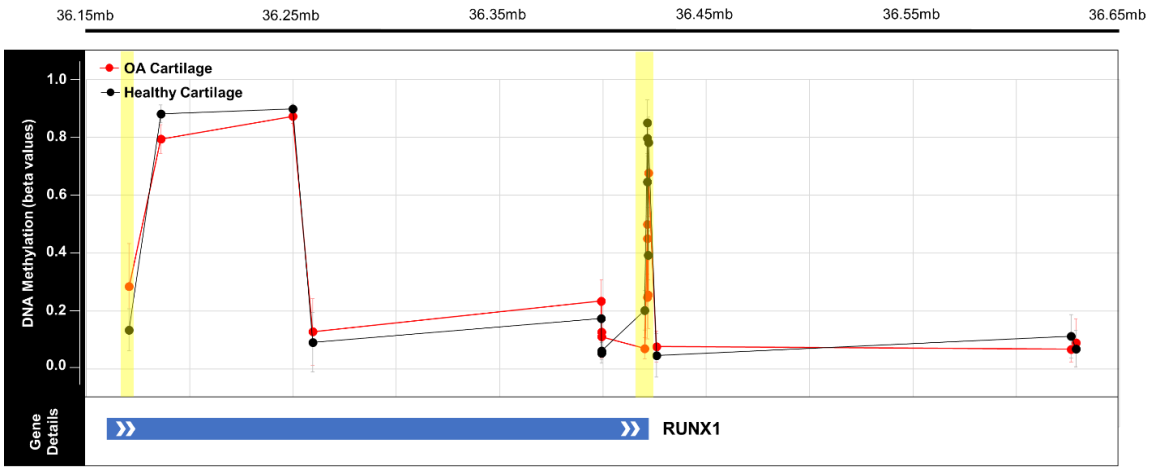

Figure 22. Methylation Levels Across *RUNX1* in Healthy and OA Baboon Cartilage. Plot of the methylation levels of significant DMPs across the *RUNX1* gene (hg19 chr21:36160098-36421595). Plot shows the average β values for each DMP with error bars indicating 1 standard deviation in each direction for each comparative group (red = OA baboon cartilage, black = healthy baboon cartilage). DMP chromosomal position in relation to the *RUNX1* gene is also depicted. *RUNX1* in baboon cartilage displays a mixture of OA hypo- and hyper-methylation patterns which opposes the OA hypomethylation patterns observed in humans (Alvarez-Garcia et al. 2016; Fernández-Tajes et al. 2014). In baboon cartilage, out of the 60 CpG sites associated with *RUNX1* that were examined, 16 DMPs were identified of which 9 were hypomethylated and 7 were hypermethylated. Additionally, 6 of the hypomethylated DMPs had a $\Delta\beta \geq 0.1$,

while only 1 of the hypermethylated DMPs had a $\Delta\beta \geq 0.1$. Regarding DMRs, one that is 1198bp long and contains 10 CpGs (hg19 chr21:36258497-36259694) was found to be hypomethylated in close proximity to this gene, and one that is 489bp long and contains 6 CpGs (hg19 chr21:36421467-36421955) was found to be hypermethylated in close proximity to this gene. Finally, 1 CpG site (cg13030790), which was hypomethylated in OA baboon cartilage, was identical between this study and human OA studies. Conversely, in bone, no DMPs or DMRs were found between healthy and OA individuals. See Appendix Q for additional information.

Discussion

Here, I identified DNA methylation variation in femoral bone and cartilage tissues from a baboon model of OA. This was done in order to assess the evolutionary conservation of epigenetic-OA associations in the primate lineage.

Using *in silico* probe filtering methods (Hernando-Herraez et al. 2013; Ong et al. 2014), I found that 24% of all human probes on the EPIC array reliably mapped to the baboon genome, contained a CpG locus, and met specific alignment filter criteria (Figure 9). This proportion was slightly lower than expected based on the previous findings of 27% retention for the 450K array (see Chapter 2). This discrepancy may be due to the altered design of the EPIC array as compared to the 450K array and to the quality of the baboon genome assembly.

Specifically, the EPIC array assays 866,837 probes, while the 450K array assays 485,512 probes. Approximately 90% of the probes on the 450K array are identical to those on the EPIC array, and the remaining EPIC array probes target CpGs in enhancer

regions. Fittingly, 60% of the retained probes in this EPIC array study are identical to those retained in the 450K array study, and the remaining 40% of retained probes only exist on the EPIC array. Additionally, those probes retained in the 450K array study but not in this EPIC array study were not retained because they are not included on the EPIC array. Finally, the relatively small number of retained probes targeting CpGs in enhancer regions may be due to the baboon genome assembly quality. While the baboon genome is well annotated, the average scaffold length (528,927) and contig length (40,262) are relatively low compared to other primates. Thus, sequence information for enhancer regions in baboons may not be as deeply covered as that for humans.

Despite this proportional discrepancy, the retained EPIC array probes maintained a wide distribution throughout the genome (APPENDIX H), had hybridization efficiencies significantly correlated with the alignment quality of each probe to the baboon genome (APPENDIX I), and predominantly passed quality controls to produce robust signals on the array (Figure 10). Thus, the quality of the retained EPIC array probes is comparable to previous findings for the 450K array (see Chapter 2).

The final set of filtered EPIC array probes for the baboon model of OA was then used to assess the evolutionary conservation of epigenetic-OA associations in the primate lineage. To do this, I identified significant DMPs between healthy and OA individuals in bone and cartilage tissues. I also looked for DMPs between tissue types and between all four combinations of disease state and tissue type (healthy cartilage vs. OA cartilage vs. healthy bone vs. OA bone).

DMPs were identified in all of these comparative groups, with more DMPs found between tissue types than between disease states and more disease state DMPs found in

cartilage samples than in bone samples (65% of sites significant in healthy bone vs. healthy cartilage, 61% of sites significant in OA bone vs. OA cartilage, 0.2% of sites significant in OA bone vs. healthy bone, and 11% of sites significant in OA cartilage vs. healthy cartilage). Many of these significant DMPs had biologically insignificant changes in mean methylation between comparative groups, so only those with at least a 10% change ($\Delta\beta \geq 0.1$) were considered further. This greatly reduced the overall number of DMPs considered but did not substantially affect the proportional trends observed between tissue types and disease states in relation to one another (25% of sites significant in healthy bone vs. healthy cartilage, 25% of sites significant in OA bone vs. OA cartilage, 0.02% of sites significant in OA bone vs. healthy bone, and 2% of sites significant in OA cartilage vs. healthy cartilage).

Regardless of whether the $\Delta\beta$ cutoff was used or not, the current baboon OA study identified substantially more DMPs than previous investigations of baboon OA using the 450K array (see Chapter 2). This is likely due to the increased sample size of the current study, as the number of actual sites examined was comparable between studies. Because of this increase in DMPs, subsequent analyses of significant DMRs were able to be performed. The number of DMRs identified between tissue types and disease states showed trends similar to those for DMPs, and the genes associated with these DMRs overlapped with those associated with DMPs (Table 6). These findings further confirm the results of the DMP analyses and their functional implications.

The epigenetic profiles of different tissue types were split such that a little more than half of the significant DMPs with $\Delta\beta \geq 0.1$ were hypermethylated in bone as compared to cartilage and the remaining DMPs were hypomethylated in bone as

compared to cartilage (Table 5). This pattern clearly distinguishes bone from cartilage (Figure 13, Figure 14) and was identical for baboons with healthy skeletal tissues and baboons with OA (Figure 17), which implies that the regulatory functions of this differential methylation between tissues were similar regardless of disease state. The predominant functions of genes associated with these DMPs included multicellular organismal process, developmental process, biological adhesion, cell adhesion, cell communication, signaling, and nervous system development (Figure 16). Excluding nervous system development, these biological processes seem appropriate given the roles of bone and cartilage in the development and maintenance of the skeleton. Nervous system development may simply be due to the methods used in this study, which reduced functional categories into large overarching groups. This interpretation is supported by the fact that this biological process encapsulates functions such as skeletal system development and regulation of ossification, which seem appropriate given the roles of bone and cartilage in the development and maintenance of the skeleton. Nevertheless, this function may also be a byproduct of the collected tissues' cellular heterogeneity. Regardless, the functional associations with tissue type differential methylation are almost identical between healthy baboons and OA baboons (Figure 17).

Comparable numbers and associated functions of significant DMPs between healthy bone vs. healthy cartilage and between OA bone vs. OA cartilage indicates that these tissues retain distinct epigenetic profiles regardless of their disease state. Maintaining a distinct epigenetic profile seems likely in healthy bone and cartilage as these tissues have distinct functions within the skeletal system. However, looking at the physical manifestation of OA, it may appear that degraded cartilage and bone blend to

become almost one pathological tissue type (Figure 1). If this was indeed the case, then the epigenetic profiles of OA bone and OA cartilage should look more similar than those between healthy bone and healthy cartilage. However, this was not found. Thus, although there appear to be distinct epigenetic alterations between disease states within bone and cartilage, these alterations are tissue-specific and do not make degraded OA bone and cartilage appear more like each other than unlike each other.

As stated, the epigenetic profiles of different disease statuses were more complicated than those of different tissues. In both bone and cartilage tissues a little more than half of the significant DMPs with $\Delta\beta \geq 0.1$ for each were hypomethylated in OA baboons as compared to healthy baboon and the remaining DMPs were hypermethylated OA baboons as compared to healthy baboons (Table 5). However, while this pattern readily distinguishes OA baboons from healthy baboons in cartilage, this pattern does not successfully differentiation OA baboons from healthy baboons in bone (Figure 13, Figure 14). Superficially, this difference is due to the substantially more disease related DMPs identified in cartilage than in bone. However, this difference in numbers of DMPs may have etiological implication. For instance, it suggests that cartilage epigenetics may have a more influential role in the development of OA than does bone epigenetics. This finding is reinforced by several human OA studies (Alvarez-Garcia et al. 2016; Aref-Eshghi et al. 2015; Fernández-Tajes et al. 2014; Jeffries et al. 2016; Moazedi-Fuerst et al. 2014; Ramos et al. 2014; Rushton et al. 2014a). On the other hand, it may simply be an artifact of bone tissue being more heterogeneous than cartilage tissue, and thus, less able for OA epigenetic signals to be isolated using the methods performed in this study.

Moreover, the disease related DMPs identified in bone and cartilage barely overlap (Figure 18), which indicates that the regulatory functions of the differential methylation between disease states is distinct between tissues. Too few DMPs were found between OA and healthy bone tissues to identify significantly enriched biological processes, but some general functions identified included fibroblast apoptotic function, calcium mediated signaling, and calcium ion transport. Conversely, several gene functions were enriched between OA and healthy cartilage, the predominant being multicellular organismal process, developmental process, tissue development, positive regulation of biological process, positive regulation of cellular process, regulation of transcription from RNA polymerase II promotor, and ossification (Figure 16). These functional differences are consistent given that the locations and amounts of differential methylation between healthy and OA individuals are very different in bone versus cartilage. However, in spite of these functional differences, the molecular pathways associated with differential methylation are similar, with those identified in bone completely subsumed by those identified in cartilage (Figure 18). This suggests that while the proximal changes in methylation and their downstream functional impacts associated with the development of OA are different in each tissue, the pathways connecting these are somewhat similar across tissues.

Overall, these disease-related differential methylated findings indicate that cartilage epigenetics may have a more influential role in the pathogenesis of OA than does bone epigenetics, which is reinforced by previous findings in humans (Alvarez-Garcia et al. 2016; Aref-Eshghi et al. 2015; Fernández-Tajes et al. 2014; Jeffries et al. 2016; Moazedi-Fuerst et al. 2014; Ramos et al. 2014; Rushton et al. 2014a). If the

epigenetic profiles of both tissues do directly contribute to the development and progression of OA, they do so through differential methylation of different regions and the use of overlapping molecular pathways which together contribute to differential regulation of different biological processes. Overall, while these finding provide some additional insight into the etiology of OA, they do not completely resolve this issues.

These disease related DMPs do provide further insight into the evolution of OA epigenetics, though. When compared with genes found to be differentially methylated in human OA studies (Alvarez-Garcia et al. 2016; Aref-Eshghi et al. 2015; Delgado-Calle et al. 2013; Fernández-Tajes et al. 2014; García-Ibarbia et al. 2013; Goldring and Marcu 2012; Iliopoulos et al. 2008; Jeffries et al. 2016; Moazedi-Fuerst et al. 2014; Ramos et al. 2014; Reynard et al. 2014; Rushton et al. 2014a; Saito et al. 2010), 197 genes found to be differentially methylated in healthy and OA baboons overlapped (Table 8). Additionally, 144 genes previously identified as having differential methylation in human OA studies were not identified as being significantly differentially methylated with or without a $\Delta\beta \geq 0.1$ cutoff in the current baboon OA study (APPENDIX Q). Of those genes that do overlap between humans and baboons, 43 have similar methylation patterns and 154 have alternative patterns (*Table 7. Overlap of Differential Methylation Associations from Human and Baboon Osteoarthritis Studies.Table 7*). Thus, some OA epigenetic patterns appear to be evolutionarily conserved between baboons and humans, while many show no evidence of conservation or display reversed signals. Specific examples of this variation in conservation are as follows.

In some instances, the OA methylation patterns observed in baboons matched those observed in humans. For instance, *TBX4* (Gene ID: 9496) displays similar OA

hypermethylation in human cartilage (Alvarez-Garcia et al. 2016; Fernández-Tajes et al. 2014) and baboon cartilage (Figure 19). *TBX4*, also known as T-box-4, codes for a transcription factor that regulates developmental processes of the lower limbs (Simon 1999). Mutations of this gene in humans have resulted in skeletal pathologies in the joints associated with the femoral bone (Bongers et al. 2004; K. Wang et al. 2010). These same areas of the leg, specifically the knee joint and hip joint, are also the predominant areas of OA development in humans (Cross et al. 2014; Cushnaghan and Dieppe 1991). Therefore, less severe alterations of the epigenetic profile of *TBX4* which may affect the regulation of this gene's expression may be mechanistically involved in the chronic pathogenesis of OA as evidenced by the increase in methylation of this gene in both OA humans and OA baboons. Additionally, since *TBX4* is involved in hindlimb specification and development, changes in methylation across this gene early during life may impact these functions and ultimately predispose an individual to OA. Furthermore, the findings of this study suggest that this OA methylation candidate gene and its pattern of susceptibility appear to be evolutionarily conserved within the primate lineage.

In other cases, the OA methylation patterns observed in one baboon tissue matched those observed in humans, but the OA methylation patterns observed in the other baboon tissue were opposite to those observed in humans. An example of this is *HOXD8* (Gene ID: 3234), which displays similar OA hypermethylation patterns in humans (Delgado-Calle et al. 2013) and baboon bone but opposing OA methylation patterns in baboon cartilage (Figure 20). *HOXD8*, also known as homeobox D8, codes for a transcription factors that regulates morphogenesis primarily of the lower limb, as evidenced by the lower limb malformations that result from mutations in this gene

(Goodman 2002; Del Campo et al. 1999). Therefore, similar to *TBX4*, less severe alterations of the epigenetic profile of *HOXD8* which may affect the regulation of this gene's expression may be mechanistically involved in the chronic pathogenesis of OA. This is supported by the differential methylation of this gene in both OA humans and OA baboons. In baboons, OA bone was found to be hypermethylated which is comparable to previous findings in humans (Delgado-Calle et al. 2013), while OA cartilage was found to be hypomethylated. This discrepancy suggests that the evolution of these OA epigenetic profiles may not be conserved. However, the differential methylation observed in humans was identified in bone tissues. Therefore, the susceptibility patterns in bone tissues may be evolutionarily conserved. Conversely, differential methylation of *HOXD8* has not been identified in human cartilage. This absence of findings in human cartilage may be due to the fact that many studies of human cartilage compare degraded and non-degraded cartilage within the same OA joint (Jeffries et al. 2016; Moazedi-Fuerst et al. 2014) as oppose to cartilage from OA versus healthy joint or evaluate healthy cartilage of the hip joint (Aref-Eshghi et al. 2015; Rushton et al. 2014b) as opposed to healthy cartilage of the knee joint. Thus, the susceptibility patterns in baboon cartilage tissues may also exist in human cartilage but have not been identified yet, or they might be isolated to baboons and not evolutionarily conserved.

Additionally, for some genes, the OA methylation patterns observed in both baboon tissues were opposite to those observed in humans. This is true for the gene *LEPR* (Gene ID: 3953), which displays opposing OA methylation patterns between humans with hypomethylation (Aref-Eshghi et al. 2015) and baboons with hypermethylated (Figure 21). *LEPR*, also known as leptin receptor, codes for a protein receptor that binds

and mediates the functions of leptin, a hormone that regulates signaling pathways involved in fat metabolisms, immune and inflammatory responses, and wound healing. The upregulation of *LEPR* is associated with osteogenic cell differentiation (Niu et al. 2015), and the release of leptin from human osteoblasts promotes bone mineralization via its interaction with *LEPR* (Enjuanes et al. 2002; Iwamoto et al. 2004). This gene's role in skeletal development and maintenance is further evidenced by studies that found alteration in the leptin pathway associated with low bone density skeletal disorders like osteoporosis (Crabbe et al. 2006; Fairbrother et al. 2007; H. J. Lee et al. 2014; Kim et al. 2008; Tam et al. 2014; Ye and Lu 2013). Moreover, altered regulation of the leptin signaling pathway has been observed in rheumatoid arthritis (Gómez-Bañuelos et al. 2015) and osteoarthritis (Simopoulou et al. 2007; Aref-Eshghi et al. 2015). Thus, differential regulation of *LEPR* as mediated by alterations in this gene's epigenetic profile may be mechanistically involved in the chronic pathogenesis of OA. Furthermore, mutations in *LEPR* have been associated with obesity (Becer et al. 2013; de Luis et al. 2013; Kimber et al. 2008; Répásy et al. 2014), and the risk of developing OA is higher in obese individuals than non-obese individuals (Cooper et al. 2000). While this alternative pathway through which *LEPR* regulation can influences OA susceptibility may be important in humans, it does not seem as relevant for the baboons included in this study, as the steady state weights of OA and healthy individuals were comparable (APPENDIX G). Regardless, the multifactorial pathways through which *LEPR* may impact OA development and progression may explain why the epigenetic profiles of this gene are not evolutionary conserved, and actually show opposite hypomethylation and hypermethylation signals, between humans and baboons.

Lastly, differentially methylated genes that overlapped between humans and baboons could sometimes show a specific pattern in humans but a mixture of signals in baboons. For example, *RUNX1* (Gene ID: 861) displays opposing OA methylation patterns between humans with hypomethylation (Alvarez-Garcia et al. 2016; Fernández-Tajes et al. 2014) and baboon cartilage with a mixture of hypo- and hyper-methylated sites (Figure 22). *RUNX1*, also known as runt related transcription factor 1, is involved in the regulation of bone and cartilage cell development and differentiation (Stein et al. 2004). In addition to its association with osteoarthritis, polymorphisms in *RUNX1* show variable effects on the development of rheumatoid arthritis (Y. H. Lee et al. 2015, 22; Martínez et al. 2006; Okada et al. 2014; Orozco et al. 2006, 22; Takata et al. 2008). Given this gene's known functions and its previous associations with skeletal disorders, it seems reasonable that changes to *RUNX1*'s epigenetic profile may affect the regulation of this gene and contribute to the chronic pathogenesis of OA in both humans and baboons. This idea is further supported by previous findings that one site associated with this gene was identified as being significantly hypomethylated in baboon OA cartilage as compared to healthy OA cartilage (see Chapter 2). Moreover, these comparable results provide evidence for the consistency of using methylation arrays to detect differential methylation levels in baboons. When comparing these baboon methylation patterns to those previously identified in humans, though, this OA methylation candidate gene and its pattern of susceptibility do not appear to be evolutionarily conserved within the primate lineage.

Overall, these findings indicate that some DNA methylation patterns associated with OA are evolutionarily conserved between humans and baboons, while others are not.

Species differences in OA epigenetics may be due to general speciation events that took place during the evolution of these taxonomic groups, to slight differences in the development or manifestation of OA in these species, or artifacts of the experimental design of this study or human OA studies. For instance, although the present baboon study increased the sample size and representation of both sexes as compared to a previous investigation of baboon OA epigenetics (see Chapter 2), the sample size was still lower than some human OA studies (n=117) (Rushton et al. 2014a). This smaller number of individuals may have reduced my power to detect potentially important OA related variants, which may explain why fewer OA candidate methylation genes were identified in baboons than humans. However, many human OA studies have comparable sample sizes to the current baboon study (n=21, n=18, n=53, n=45, n=30, n=12, n=15, n=33, respectively) (Alvarez-Garcia et al. 2016; Aref-Eshghi et al. 2015; Delgado-Calle et al. 2013; Fernández-Tajes et al. 2014; García-Ibarbia et al. 2013; Jeffries et al. 2016; Moazedi-Fuerst et al. 2014; Ramos et al. 2014). Thus, in order to identify candidate epigenetic alterations that underlie variation in knee OA better, even larger sample sets of both baboons and humans that possibly only focus on the epigenetics of cartilage tissues should be considered.

Additionally, the current study only evaluated one population of baboons; whereas, the list of OA candidate methylation genes considered came from several studies of OA in different populations of humans (Alvarez-Garcia et al. 2016; Aref-Eshghi et al. 2015; Delgado-Calle et al. 2013; Fernández-Tajes et al. 2014; García-Ibarbia et al. 2013; Goldring and Marcu 2012; Iliopoulos et al. 2008; Jeffries et al. 2016; Moazedi-Fuerst et al. 2014; Ramos et al. 2014; Reynard et al. 2014; Rushton et al. 2014a;

Saito et al. 2010). This may have preferentially biased my human gene list to have more than could be identified in any one population. In order to account for this, a replicate study of OA epigenetics should be done in baboons to determine if this additional replicate alters the results.

Lastly, this baboon study used a stringent DMP cutoff threshold that limited the resulting genes to only those that were associated with DMPs having an average change in mean methylation between comparative groups greater than or equal to 10%. This was as a precaution against including sites that likely had little biological relevance (Hernando-Herraez et al. 2013). However, while some human OA studies enforce comparable thresholds (Alvarez-Garcia et al. 2016; Aref-Eshghi et al. 2015; Jeffries et al. 2016; Rushton et al. 2014b), some do not (Delgado-Calle et al. 2013; Ramos et al. 2014; Reynard et al. 2014). Therefore, many genes previously classified as being differentially methylated in human OA may be false positives. Until further work to identify the mechanisms through which OA develops and progresses is done in humans or other model systems, the validity of currently known OA candidate methylation genes will remain unknown.

Regardless of these potential confounding factors, using baboons as a model of OA in this study has begun to clarify the etiology of this disorder and the evolutionary conservation of epigenetic mechanisms associated with this disorder. This is the first study to specifically assess DNA methylation in skeletal tissues from a nonhuman primate using the EPIC array and serves as a follow up to a previous study using the 450K array (see Chapter 2). In conclusion, from an evolutionary perspective, the findings of this study inform our understanding of DNA methylation variation in one species and

in two skeletal tissues, as well as the degree to which the common skeletal condition of OA affects this variation. These findings warrant further investigation in a larger and more phylogenetically diverse sample set, and future research in this baboon model of OA will help provide insight into the pathogenesis of OA.

Acknowledgements

Anne Sheldrake, Jaydee Turner, Kara Peterson, Mel Carless, & Laura Cox

Department of Genetics, Texas Biomedical Research Institute, San Antonio, TX, USA

Funding

Leakey Foundation Research Grant for Doctoral Students to G. Housman

Wenner-Gren Foundation Dissertation Fieldwork Grant (Gr. 9310) to G. Housman

James F. Nacey Fellowship from the Nacey Maggioncalda Foundation to G. Housman

International Primatological Society Research Grant to G. Housman

Sigma Xi Grant-in-Aid of Research to G. Housman

Center for Evolution and Medicine Venture Fund (ASU) to G. Housman

Graduate Research and Support Program Grant (GPSA, ASU) to G. Housman

Graduate Student Research Grant (SHESC, ASU) to G. Housman

CHAPTER 4

AN EVOLUTIONARY PERSPECTIVE OF DNA METHYLATION ASSOCIATED WITH AGE WITHING THE PRIMATE LINEAGE

Abstract

Aging is thought to be a developmentally regulated process that is controlled by epigenetic mechanisms, such as DNA methylation. Methylation patterns throughout the genome show distinct changes with age in humans. However, the evolution of age epigenetics and the degree to which epigenetic signatures of aging are conserved between species has not readily been explored. Here, I examine the evolution of aging epigenetics specifically within the primate lineage by identifying age-related DNA methylation pattern in baboons (*Papio spp.*) and by comparing these findings to what is known in humans. Genome-wide DNA methylation patterns were identified in femoral trabecular bone from 46 pedigreed baboons, 28 that were adults and 18 that were juveniles using the Illumina Infinium MethylationEPIC BeadChip (EPIC array). Several significantly differentially methylated positions (DMPs) were found between these age cohorts, and similar to other animals, adult baboons display global hypomethylation as compared to juvenile baboons. The significant age-related DMPs identified are associated with genes involved in developmental processes and pathways related to the progression of diseases of aging. Additionally, while some of these age-related DMPs overlap with and display methylation patterns similar to those previously identified in human aging studies, the majority of previously identified age-related methylation loci in humans were not significantly differentially methylated in baboons. Nevertheless, methylation levels at

these human loci are still able to differentiate baboon age cohorts. Overall, these results reveal how DNA methylation varies with respect to age in skeletal tissues from one primate species and provide insight into the evolutionary conservation of aging epigenetics within the primate lineage.

Key Words

Aging, DNA methylation, evolution, epigenome, bone, baboon

Introduction

The physical effects of aging are apparent across a wide range of tissues. Over the course of aging, the dermis loses its elasticity, hair loses its pigmentation and thins, bones become brittle, muscles weaken, blood vessels stiffen, injured tissue regeneration slows, and the capacity of the immune system to protect against infections and cancers deteriorates (Ryu et al. 2008; Van Neste and Tobin 2004; Ho et al. 2005). This general senescence, which is present across several different tissues, may be the result of similar underlying molecular mechanisms (Koch and Wagner 2011).

While the accumulation of DNA damage and the shortening of telomeres have been associated with senescence, epigenetic modifications such as DNA methylation also have a role in this developmentally regulated process (Campisi and Vijg 2009; Christensen et al. 2009; Fraga and Esteller 2007; Gonzalo 2010; Koch et al. 2011; Koch and Wagner 2011; Maegawa et al. 2010; Marciniak-Czochra et al. 2009; Martino et al. 2011; Mugatroyd et al. 2010; Oberdoerffer and Sinclair 2007; Rakyan et al. 2010; Schellenberg et al. 2011; D. L. Thompson et al. 1998). The involvement of DNA

methylation changes within germline cells, during infancy, and throughout adolescence further supports this idea (McCartney et al. 2016; Massart et al. 2016).

Among several species, genome-wide DNA hypomethylation has been associated with aging (Berdyshev et al. 1967; Vanyushin et al. 1973; Wilson et al. 1987). Simultaneously, at a finer resolution, certain regions of the genome become hypermethylated with age while others become hypomethylated with age (J. T. Bell et al. 2012; Bollati et al. 2009; Christensen et al. 2009; Fraga et al. 2007; Fraga and Esteller 2007; Horvath et al. 2012; Mugatroyd et al. 2010; Rakyan and Beck 2006; Rodríguez-Rodero et al. 2010; Teschendorff et al. 2010). Initial discoveries of age-related DNA methylation changes showed tissue-specific patterns (Christensen et al. 2009; R. F. Thompson et al. 2010). Specifically, age-associated changes were identified in mesenchymal stem cells and fibroblasts (Koch et al. 2011; Bork et al. 2010; Wagner et al. 2010), tissues of the dermis and epidermis (Grönniger et al. 2010), blood (Rakyan et al. 2010; Y. Chen et al. 2011), cord blood (Bocker et al. 2011; Adkins et al. 2011), cervical smears (Teschendorff et al. 2010), and saliva (Bocklandt et al. 2011), and these tissue-specific methylation profiles were highly reproducible (Maegawa et al. 2010; R. F. Thompson et al. 2010).

When the results of tissue-specific studies were examined together, some differential methylation patterns overlapped between tissue types and were found to be independent of sex, disease state, and methylation array platform (J. T. Bell et al. 2012; Bocklandt et al. 2011; Hernandez et al. 2011; Horvath et al. 2012; Koch and Wagner 2011; Numata et al. 2012; Rakyan et al. 2010; Teschendorff et al. 2010). These findings prompted the curation of CpG sites with tissue-independent, age-related DNA

methylation changes for use in predicting individuals' ages (Koch and Wagner 2011; Horvath 2013). An early study determined that 5 CpG sites could successfully estimate donor age (Koch and Wagner 2011). Four of these sites are hypermethylated with age (*TRIM58* - cg07533148, *KCNQ1DN* - cg01530101, *NPTX2* - cg1279989, *GRIA2* - cg25148589), and one is hypomethylated with age (*BIRC4BP* - cg23571857). A more recent study expanded on this set to include over 300 CpG sites that have a mixture of age-related hypomethylation and hypermethylation (Horvath 2013). The cumulative effect of methylation levels across these loci serves as a robust estimator of age, regardless of tissue type, in both young and old individuals (Bocklandt et al. 2011; Hannum et al. 2013; Horvath 2013; Lin et al. 2016; Spiers et al. 2015; Walker et al. 2015). Additionally, methylation at these more than 300 CpG sites can predict all-cause mortality (Marioni, Shah, McRae, Chen, et al. 2015; B. H. Chen et al. 2016), correlate with mental and physical fitness levels of older individuals (Marioni, Shah, McRae, Ritchie, et al. 2015; Levine et al. 2015), and associate with several diseases and conditions of aging (Horvath 2013; Horvath et al. 2014; Horvath, Garagnani, et al. 2015; Horvath, Mah, et al. 2015; Horvath and Levine 2015; Horvath and Ritz 2015; Horvath et al. 2016; Levine et al. 2015; Lowe et al. 2016).

Overall, although these methylation profiles aim to predict chronological age, they actually only define an epigenetic age. While epigenetic age often approximates chronological age, exceptions to this relationship have been observed. Thus, chronological age-matched individuals can have different epigenetic ages. Older epigenetic ages than expected have been associated with an increased risk for death from all causes (Marioni, Shah, McRae, Chen, et al. 2015; Christiansen et al. 2016; Perna et al.

2016), and younger epigenetic ages than expected have been associated with semi-supercentenarians, or individuals who lived to be at least 105-109 years of age (Marioni, Shah, McRae, Chen, et al. 2015). This discrepancy between chronological age and epigenetic age may provide insights into the susceptibilities of certain individuals to early senescence and diseases of aging. Indeed, the epigenetics of aging has already begun to inform our understanding of chronic disease epigenetics (C. G. Bell et al. 2016), and as such, understanding aging epigenetics more fully is crucial for these continued efforts.

One aspect of this field that has not readily been explored yet is how evolution has impacted aging processes and mechanisms. Most aging epigenetics research has been conducted in humans, which hinders our ability to take an evolutionary perspective. While global hypomethylation with aging appears to be evolutionarily conserved across several species (Berdyshev et al. 1967; Vanyushin et al. 1973; Wilson et al. 1987), it is unknown how other epigenetic signatures of aging have changed throughout animal evolution. Here, I begin to explore the evolution of aging epigenetics within the primate lineage by examining these patterns in one nonhuman primate and relating them to what is known for humans.

Specifically, for this study, I explored the evolution of aging epigenetics by identifying DNA methylation patterns in femoral trabecular bone from 46 pedigreed baboons, 28 that were adults and 18 that were juveniles, and assessing whether DNA methylation variation is associated with aging in baboons and in a manner similar to that observed in humans.

Methods

Ethics Statement

Nonhuman primate tissue samples included were opportunistically collected at routine necropsy of these animals. No animals were sacrificed for this study, and no living animals were used in this study.

Nonhuman Primate Samples

Baboon (*Papio spp.*) samples come from captive colonies at the Southwest National Primate Research Center in the Texas Biomedical Research Institute. These animals have a tracked pedigree, which denotes the genetic relationships among all individuals.

Femora were opportunistically collected at routine necropsy of these animals and stored in -20°C freezers at the Texas Biomedical Research Institute after dissection. These preparation and storage conditions ensured the preservation of skeletal DNA methylation patterns.

Samples include adult (n=28) and juvenile (n=18) baboons with no observable pathologies of the distal femur articular surface, and both sexes are represented (Figure 23, APPENDIX S). Adults are classified as sexually mature individuals, and juveniles are classified as sexually immature individuals. Thus, the ages of adults and juveniles included in this study fall outside the range of ages at which baboons go through puberty and reach sexual maturity (4-9 years) (Cawthon Lang, 2006).

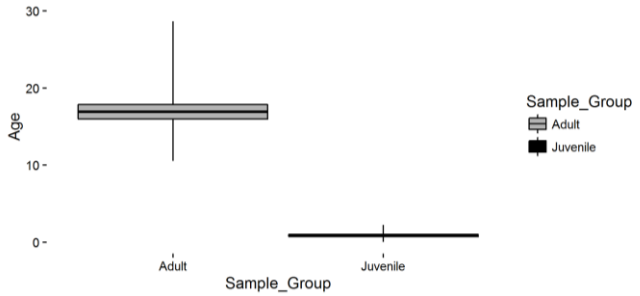

Figure 23. Baboon Sample Set Ages for EPIC Array Aging Study.

Box depicts the average age and one standard deviation, and whiskers depict the full range of ages. Healthy adult baboons (n=28) are 16.90±5.02 years, and healthy juvenile (n=18) are 0.85±0.63 years.

DNA Extraction

DNA was extracted from femoral trabecular bone using a phenol-chloroform protocol optimized for skeletal tissues (Barnett and Larson 2012). From the distal femoral condyles, trabecular bone was collected using coring devices and pulverized into bone dust using a SPEX SamplePrep Freezer/Mill. Specifically, bone cores were obtained from a transverse plane through the center of the medial condyle on the right distal femur, such that the articular surface remained preserved. Cortical bone was removed from these cores using a Dremel.

Cartilage methylation patterns are known to vary between joints and between different sites within a joint (den Hollander et al. 2014; Jeffries et al. 2016; Loughlin and Reynard 2015; Moazedi-Fuerst et al. 2014; Rushton et al. 2014b). Although similar studies of bone methylation patterns have not been conducted yet, the number and types of cells, and therefore epigenetic signatures, are expected to vary across different portions

of the femur. Thus, tissues were collected from the same portion of the femur in order to minimize this variation between samples and comparative groups.

Genome-Wide DNA Methylation Profiling

Genome-wide DNA methylation was assessed using Illumina Infinium MethylationEPIC microarrays (EPIC array). As previously described (see Chapter 3), these arrays analyze the methylation status of over 850,000 sites throughout the genome, covering over 90% of the sites on the Infinium HumanMethylation450 BeadChip as well as an additional 350,000 sites within enhancer regions. For each sample, 400ng of genomic DNA was bisulfite converted using the EZ DNA Methylation™ Gold Kit according to the manufacturer's instructions (Zymo Research), with modifications described in the Infinium Methylation Assay Protocol. Following manufacturer guidelines (Illumina), this processed DNA was then whole-genome amplified, enzymatically fragmented, hybridized to the arrays, and imaged using the Illumina iScan system. The array data discussed here are available in APPENDIX R.

Methylation Data Processing

Raw fluorescent data were normalized to account for the noise inherent within and between the arrays themselves. Specifically, I performed a normal-exponential out-of-band (Noob) background correction method with dye-bias normalization (Triche et al. 2013) to adjust for background fluorescence and dye-based biases and followed this with a between-array normalization method (functional normalization) (Fortin et al. 2014), which removes unwanted variation by regressing out variability explained by the control

probes present on the array as implemented in the minfi package in R (Aryee et al. 2014; Fortin et al. 2016) that is part of the Bioconductor project (Huber et al. 2015). This method has been found to outperform other existing approaches for studies that compare conditions with known large-scale differences (Fortin et al. 2014), such as those assessed in this study.

After normalization, methylation values (β values) for each site were calculated as the ratio of methylated probe signal intensity to the sum of both methylated and unmethylated probe signal intensities (Equation 1). These β values range from 0 to 1 and represent the average methylation levels at each site across the entire population of cells from which DNA was extracted (0 =completely unmethylated sites, 1 = fully methylated sites).

Every β value in the Infinium platform is accompanied by a detection p-value, and those with failed detection levels (p-value > 0.05) in greater than 10% of samples were removed from downstream analyses. Additionally, samples in which more than 30% of the β value had a detection p-value > 0.05 were removed from downstream analyses.

The probes on the arrays were designed to specifically hybridize with human DNA, so my use of nonhuman primate DNA required that probes non-specific to the baboon genome, which could produce biased methylation measurements, be computationally filtered out and excluded from downstream analyses. This was accomplished using methods modified from (Hernando-Herraez et al. 2013; Ong et al. 2014). Briefly, I used blastn (Altschul et al. 1997) to map the 866,837 50bp probes onto the *Papio anubis* genome (Assembly: Panu_2.0, Accession: GCF_000264685.2) using an e-value threshold of e^{-10}. I only retained probes that successfully mapped to the baboon

genome, had only 1 unique BLAST hit, targeted CpG sites, had 0 mismatches in 5bp closest to and including the CpG site, and had 0-2 mismatches in 45bp not including the CpG site. This filtering retained 209,802 probes.

Additionally, β values associated with cross-reactive probes (McCartney et al. 2016), probes containing SNPs at the CpG site, probes detecting SNP information, probes detecting methylation at non-CpG sites, and probes targeting sites within the sex chromosomes were removed using the minfi package in R (Aryee et al. 2014; Fortin et al. 2016) (Figure 24). This filtering retained a final set of 191,631 probes.

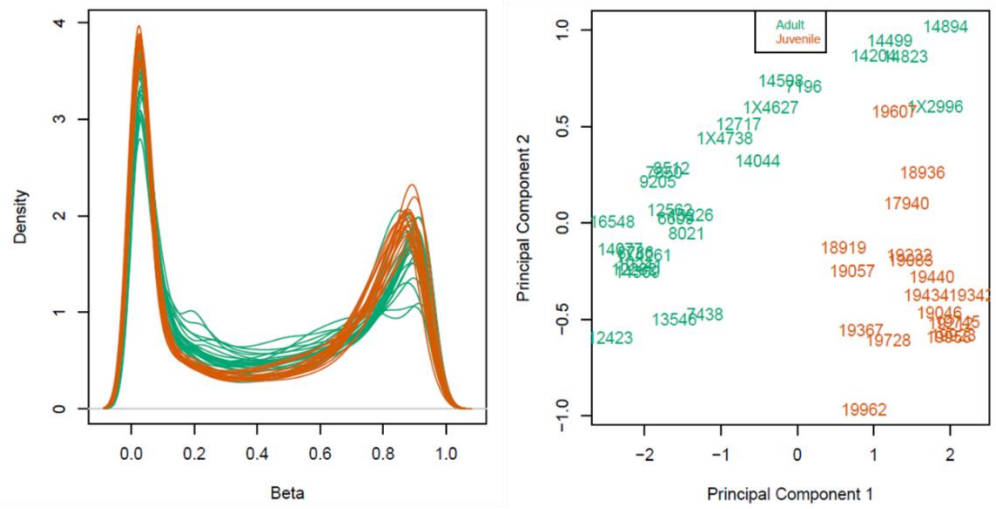

Figure 24. Normalized and Filtered Methylation Data for EPIC Array Baboon Aging Study.

Density plots of β-values after normalization and probe filtering using the alignment criteria (A). Multidimensional scaling plots showing the first two principle components that describe genome-wide methylation variation after normalization and filtering using

the alignment criteria (B). Each point represents one sample that is either from an adult or a juvenile.

Differential Methylation Analyses

Because β values have high heteroscedasticity, they are not statistically valid for use in differential methylation analyses (Du et al. 2010). Thus, M values were calculated and used in these analyses instead (Equation 2).

In order to identify sites that were significantly differentially methylated across comparative groups, I designed and tested generalized linear mixed models (GLMMs) which related the variables of interest to the DNA methylation patterns for each site, while accounting for the effects of additional variables, batch effects, and latent variables (Maksimovic et al. 2016). Sites found to have significant associations were classified as significantly differentially methylated positions (DMPs).

Specifically, a GLMM was used to estimate differences in methylation levels between adult and juvenile baboons. Additional variables included in this GLMM were sex, known batch effects (e.g., array number and position), and unknown latent variables calculated using the iteratively re-weighted least squares approach in the sva package in R (Jaffe and Irizarry 2014; Jeffrey T. Leek et al. 2012; J. T. Leek and Storey 2008; Jeffrey T. Leek and Storey 2007). The 3 latent variables estimated were included to help mitigate any unknown batch and cell heterogeneity effects on methylation variation at each site.

Alternative methods to account for cell heterogeneity exist, but they are specific to whole blood (Jaffe and Irizarry 2014; Morris and Beck 2015), require reference

epigenetic data, or are reference free methods (Houseman et al. 2014) that are comparable to the sva method (Kaushal et al. 2015). Out of the known cell types in skeletal tissues (Horvath, Mah, et al. 2015), only chondrocytes and osteoblasts have reference epigenomes available on the International Human Epigenomics Consortium, and these are only for humans, not nonhuman primates. Thus, because no standard method is available to correct for the heterogeneous cell structure in nonhuman primate skeletal tissue, I chose the described sva method.

This GLMM design matrix (*Equation 7*) was fit to the M value array data by generalized least squares using the limma package in R (Ritchie et al. 2015; Phipson et al. 2016; Huber et al. 2015), and the estimated coefficients and standard errors for the defined age cohort contrast were computed. Lastly, for each coefficient, an empirical Bayes approach (McCarthy and Smyth 2009; Lönnstedt and Speed 2002; Phipson et al. 2016; Smyth 2004) was used to compute moderated t-statistics, log-odds ratios of differential methylation, and associated p-values adjusted for multiple testing (Benjamini and Hochberg 1995). Significant DMPs for the effect of age cohort contrasts were defined as those having log fold changes in M values corresponding to an adjusted p-value of less than 0.05.

In order to account for genetic relatedness, the coefficients of relatedness (phi2 = 2 x kinship coefficients), or the expected proportions of alleles that are identical by descent between 2 individuals, were computed from a known pedigree using the kinship2 package in R (Therneau et al. 2015). Following this, two new GLMMs were designed and tested using the lmekin function of the coxme package in R (Therneau 2015). The first GLMM regressed methylation levels (M values) against the age cohort contrast

effect while adjusting for other variables (sex, batch effects, latent variables) as fixed effects and kinship (phi2) as a random effect (*Equation 8*) (Zaghlool et al. 2015), and the second performed the same regression with the age cohort contrast effect removed (*Equation 9*). The log likelihoods of each model were then compared using a chi-square test to determine which model better explained the variation in methylation. For this test, the degrees of freedom were calculated as the absolute difference in the Akaike's information criteria for each model (Mazerolle 2016). When the model containing the age cohort contrast effect performed significantly better than the alternative model (p-value < 0.05), this confirmed that the site remained a significant DMP for the effect of age cohort contrasts when adjusting for the added effects of kinship. Conversely, when the model containing the age cohort contrast effect did not perform better than the alternative model (p-value ≥ 0.05), this indicated that the site was not a significant DMP for the effect of age cohort contrasts when adjusting for the added effects of kinship. In this instance, this site was no longer considered a significant DMP.

Equation 7: methylation ~ age cohort contrasts + sex + batch effects + latent variables

Equation 8: methylation ~ age cohort contrasts + sex + batch effects + latent variables + kinship

Equation 9: methylation ~ sex + batch effects + latent variables + kinship

Lastly, I further examined significant DMPs that had at least a 10% change in mean methylation between comparative groups ($\Delta\beta \geq 0.1$), as these may have greater biological impact than others (Hernando-Herraez et al. 2013). The gene ontology and KEGG pathway enrichment for significant CpGs while taking into account the differing

number of probes per gene present on the array was determined using the missMethyl package in R (Geeleher et al. 2013; Young et al. 2010; Ritchie et al. 2015; Benjamini and Hochberg 1995). Significantly enriched (FDR < 0.05) GO biological processes were subsequently summarized using REViGO which removed redundant GO terms (retained only 50% of the full list of significant terms) and visualized the remaining terms in a semantic similarity-based scatterplot (Supek et al. 2011). Semantic similarity was calculated using the simRel score, which is a functional similarity measure that ranges from 0 for terms that have no similarity to 1 for terms with maximum similarity (Schlicker et al. 2006).

In addition to DMPs, differentially methylated regions (DMRs) were also identified between each comparative group using the DMRcate package in R (Peters et al. 2015; Wand and Jones 1994; Duong 2013). This method is only concerned with the spatial proximity of loci examined and is not biased by any annotations associated with these loci. For these analyses, the individual DMP t-statistics, which were derived by fitting the M value array data to a GLMM design matrix (*Equation 7*) by generalized least squares using the limma package in R (Ritchie et al. 2015; Phipson et al. 2016; Huber et al. 2015), were smoothed across each chromosome using a recommended Gaussian kernel bandwidth of 1000 base pairs with a scaling factor of 2. An expected value of this smoothed estimate with no experimental effects was also modelled using a Satterthwaite approximation (Satterthwaite 1946) in order to calculate a subsequent significance test for each DMP. A default threshold was then applied to p-values adjusted for multiple testing (Benjamini and Hochberg 1995) to identify FDR-corrected significant DMPs. Finally, these significant DMPs were agglomerated together into DMRs based on

chromosomal location and such that each DMR contained at least 2 CpG sites that were less than 1000 base pairs apart.

Results

The aim of this study was to use the EPIC array to identify DNA methylation patterns in femoral trabecular bone of baboons with varied ages, 28 adults and 18 juveniles. In order to do this, I first assessed the effectiveness of the EPIC array in identifying DNA methylation patterns in baboon DNA.

Alignment of EPIC Array Probes with the Baboon Genome

Similar to the methods in Chapter 3, probes from the EPIC array were aligned to the baboon genome using methods modified from (Hernando-Herraez et al. 2013; Ong et al. 2014) (APPENDIX A). Out of the 866,837 50bp probes on the array, 209,802 probes map to the baboon genome with e-values less than e^{-10}, have only unique BLAST hits, target a CpG site, and meet the described alignment filter criteria (Figure 9). These probes covered approximately 23,446 genes with an average coverage of 8 probes per gene. Additionally, the retained probes covered a range of locations with respect to genes and CpG islands (APPENDIX H), indicating that these filtered probes maintain a wide distribution throughout the genome. After filtering out cross-reactive probes (Y. Chen et al. 2013), probes containing SNPs at the CpG site, probes detecting SNP information, probes detecting methylation at non-CpG sites, and probes targeting sites within the sex chromosomes, a final set of 191,631 probes were retained for downstream analyses.

Effectiveness of EPIC Array Probes using Baboon DNA

To determine how effectively the EPIC array probes measured DNA methylation in baboon DNA, I performed Spearman correlation tests between the hybridization efficiency of each probe and parameters defining the alignment quality of each probe to the baboon genome. Specifically, both probe alignment bitscores and percent identity were significantly negatively correlated with probe hybridization efficiency, and probe alignment e-values were significantly positively correlated with probe hybridization efficiency, regardless of filtering criteria (APPENDIX T). Additionally, filtered probes retained a large proportion of successfully hybridized probes (Figure 25).

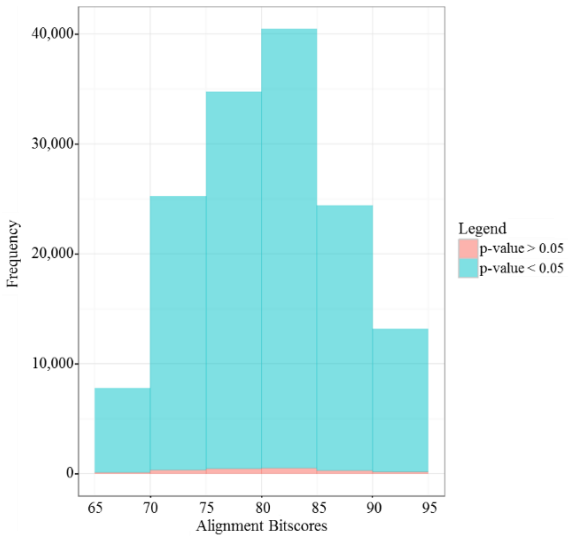

Figure 25. Hybridization Efficiencies of EPIC Array Probes Retained for Baboon Aging Study.

Histogram of alignment bitscores for EPIC array probes with detection p-values > 0.05 (red) and < 0.05 (blue). These p-values were averaged across all samples, and probes

included meet the alignment filter criteria. For these probes 2,815 had detection p-values > 0.05, and 206,987 had detection p-values < 0.05.

Differential Methylation and Aging

Significant DMPs were interrogated from 191,631 sites and identified between age cohorts (adult vs. juvenile) (Figure 26, APPENDIX U). Accounting for kinship slightly reduced these DMP counts, and using a $\Delta\beta \geq 0.1$ threshold substantially decreased the final number of significant DMPs between adults and juveniles. However, these filters did not diminish the overall distribution of DMPs across a variety of functional genomic regions and proximities to CpG islands (APPENDIX V). Several significant DMRs were also identified between adults and juveniles, with patterns similar to those identified among DMPs (Table 9, APPENDIX W).

The methylation patterns at significant DMPs with $\Delta\beta \geq 0.1$ generally distinguish adult baboons from juvenile baboons. For instance, more than half of all these DMPs are hypomethylated in adults as compared to juveniles (Table 10). Additionally, these DMPs are associated with several genes that have gene ontology (GO) biological processes predominantly involved in developmental functions (Figure 27, APPENDIX X) and KEGG pathway functions predominantly involved in development, maintenance, and the progression of diseases of aging (APPENDIX Y). Regardless of these unifying feature, the specific quantitative methylation level at each DMP do not separate adults and juveniles into distinct clusters (Figure 28). Rather, five female adult baboons (ages 11.19 years, 28.66 years, 12.83 years, 10.96 years, and 11.64 years) cluster more closely with juvenile baboons than with other adults.

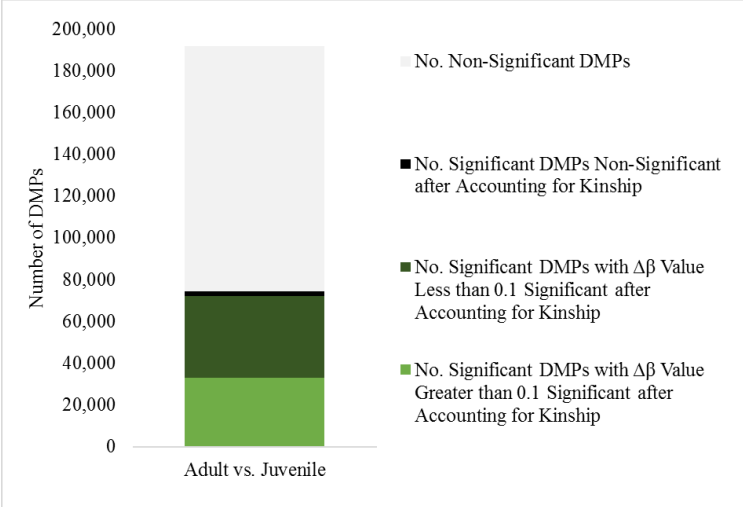

Figure 26. Number of Significant DMPs Identified in the EPIC Array Baboon Aging Study.

Bar chart showing the number of significant DMPs between adult and juvenile baboons. Results include the number of significant DMPs that remained statistically significant after accounting for kinship, the number of significant DMPs that did not remain statistically significant after accounting for kinship, and the number of loci that were not found to be statistically significant. Those significant DMPs that remained statistically significant after accounting for kinship were additionally split into those that had $\Delta\beta < 0.1$ and those that had $\Delta\beta \geq 0.1$. Approximately 38.9% (74,545) of probes were differentially methylated between adults and juveniles. When accounting for genetic relatedness, 3.1% (2,332) of the originally identified significant DMPs do not maintain significant methylation associations. Of those significant DMPs that remained statistically significant after accounting for kinship, only 45.8% (33,057) had $\Delta\beta \geq 0.1$, implying the difference may have had regulatory and biological effects.

Table 9. Number of Significant DMRs Identified in the EPIC Array Baboon Aging Study.

	No. Significant DMRs	No. Associated Genes	Average CpGs per DMR (Min-Max)	Average Length of DMR (Min-Max)	Associated Gene Symbols
Adult vs. Juvenile	12,962	9,023	5 (2-93)	791bp (3bp-11,695bp)	*EN1, FOXG1, HOTAIRM1, HOXA2, KCNQ1DN, PAX6, PPT2, PPT2-EGFL8, RCN1, RP11, SP9, TBR1, ZIC1, ZIC4*

Table showing the number of significant DMRs identified between adult and juvenile baboons, the number of unique gene names that overlapped with these regions, the average number of CpGs per DMR along with the minimum and maximum number, average length of DMRs along with the minimum and maximum length, and gene symbol names associated with the top 10 DMRs for each comparative group. For additional details see Appendix W.

Table 10. Number of Significant DMPs Identified in the EPIC Array Baboon Aging Study.

Differential Methylation	Adult vs. Juvenile
Significant (negative)	20,665
Not Significant	158,574
Significant (positive)	12,392

Table showing the number of significant DMPs with $\Delta\beta \geq 0.1$ between comparative groups when accounting for kinship. Results are shown for probes filtered using the alignment criteria, and for these, significant DMPs were identified in all comparative groups.

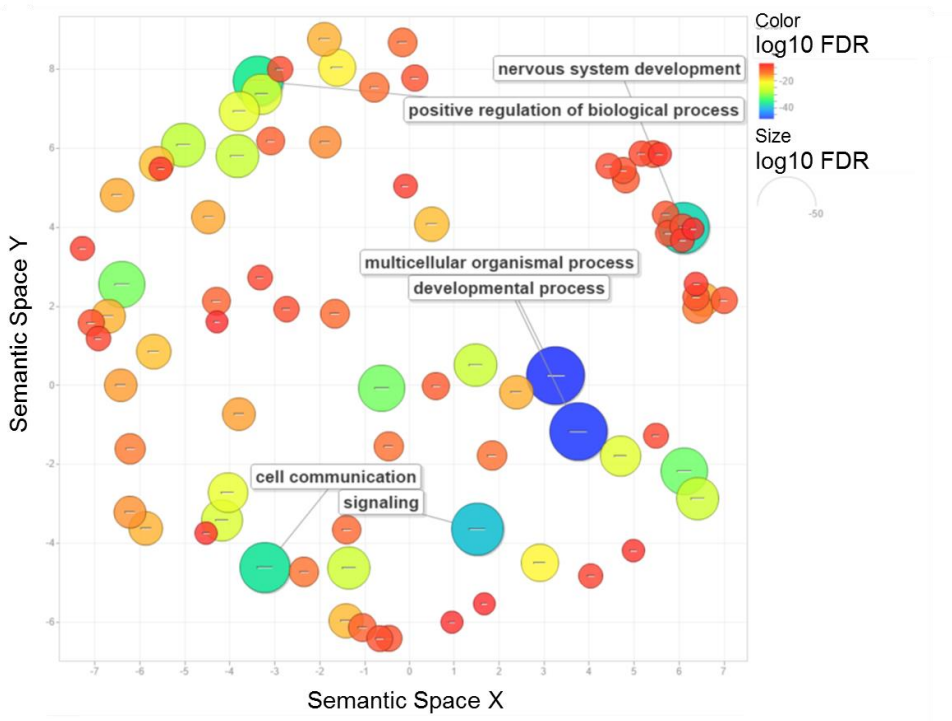

Figure 27. GO Biological Processes Enriched for Significant DMPs Associated with Aging in Baboons.

Multidimensional scaling plot summarizing the GO biological process terms that are significantly enriched (FDR < 0.05) for significant DMPs with $\Delta\beta \geq 0.1$, taking into account the differing number of probes per gene present on the EPIC array. DMPs were identified between adult and juvenile baboons. REViGO was used to remove redundant GO terms (retained only 50% of the full list of significant terms) and to visualize the remaining terms in a semantic similarity-based scatterplot (Supek et al. 2011). Semantic similarity was calculated using the simRel score, which is a functional similarity measure that ranges from 0 for terms that have no similarity to 1 for terms with maximum similarity (Schlicker et al. 2006). These pairwise semantic similarity scores are then plotted in multidimensional scaling space such that similar GO terms are located close to

one another in the plot. The color and size of each GO term are based on the log10 FDR value, and some of the most significant GO terms have their descriptions provided in the plot. See APPENDIX X for additional information.

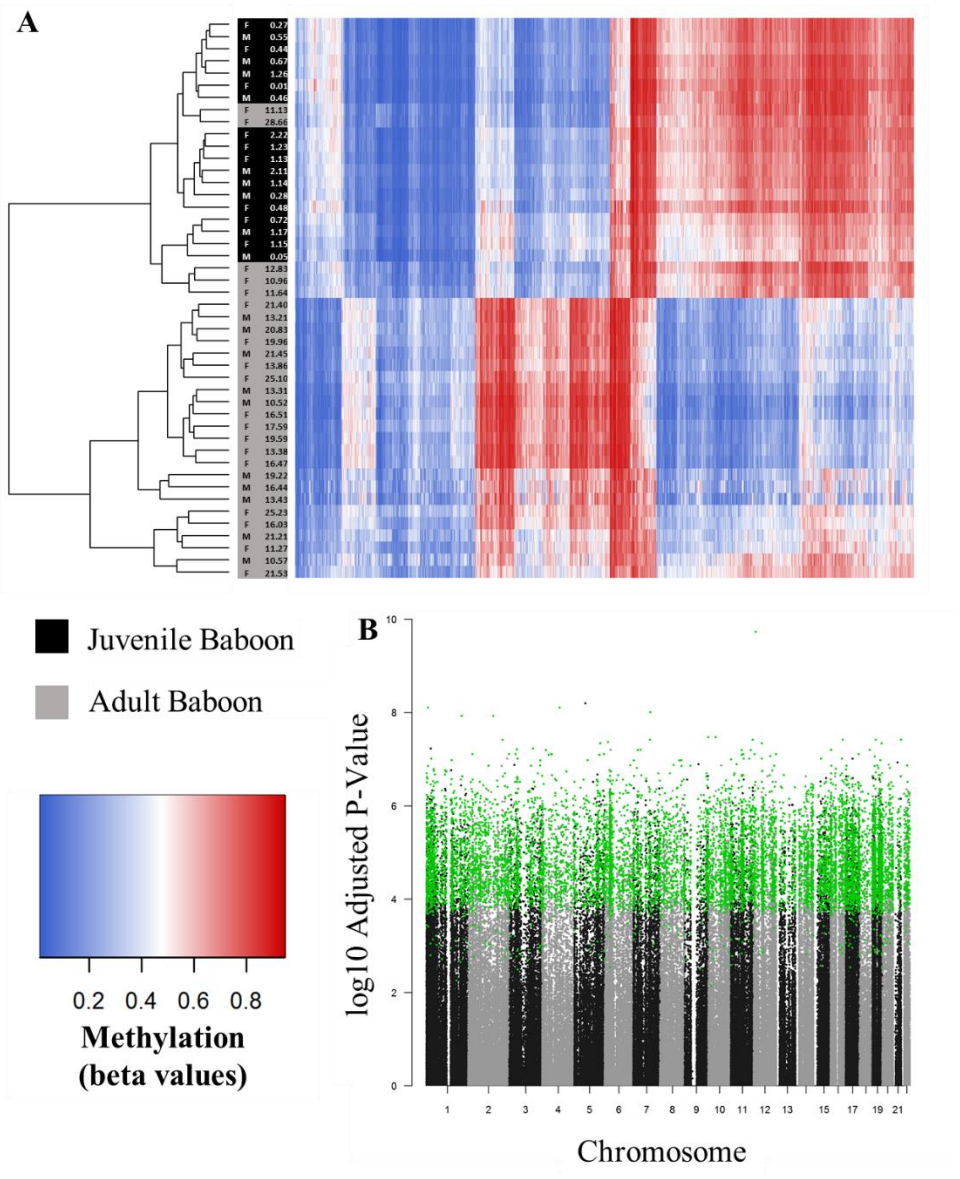

Figure 28. Methylation Levels at the Top 10,000 Significant DMPs Identified in the EPIC Array Baboon Aging Study.

(A) Heatmap depicting the DNA methylation levels (β values) of the top 10,000 DMPs with $\Delta\beta \geq 0.1$ between adults and juveniles (x-axis) in all baboon samples (y-axis, n=46). The sex and age of each baboon sample are also provided (y-axis). Red indicates higher methylation at a DMP, while blue indicates lower methylation at a DMP. The dendrogram of all samples (y-axis) clusters individuals based on the similarity of their methylation patterns. Adults and juveniles mostly form distinct clusters, with a few female adults (ages 11.19 years, 28.66 years, 12.83 years, 10.96 years, and 11.64 years) clustering with juveniles. (B) Manhattan plot showing the log10 adjusted p-values of all positions examined between adults and juveniles in all baboon samples (n=46). Samples highlighted in green are the top 10,000 significant DMPs with $\Delta\beta \geq 0.1$ between age cohorts that are displayed in the heatmap.

A subset of significant DMPs identified between adult and juvenile baboons overlap with CpGs previously identified as being differentially methylated in human aging studies. Human aging studies across a large range of tissues have identified over 300 CpGs that become hypomethylated or hypermethylated with age (Koch and Wagner 2011; Horvath 2013), but only 137 of these CpGs were able to be evaluated in baboons. Of these, 14 were found to be significantly differentially methylated across baboon ages (Table 11), and the rest were either not found to be statistically significant (n=75) or were statistically significant but had average changes in methylation levels less than 0.1 (n=48) (APPENDIX Z). Comparisons at the gene level as opposed to the CpG site level produce similar results (APPENDIX Z).

Table 11. Overlap of Differential Methylation Associations from Human and Baboon Aging Studies.

Current Study Baboon Aging Methylation Level Findings	Previous Human Findings	
	Hypermethylation with Age	Hypomethylation with Age
Same Methylation Pattern	*ALOX12* (cg03760483) *DIRAS3* (cg22901840) *FOXG1B* (cg02681442) *KCNQ1DN* (cg01530101) *LTBP3* (cg08965235) *NHLRC1* (cg22736354) *TBX5* (cg21907579) *VGF* (cg04084157)	*ACOT11* (cg10266490) *BCMO1* (cg22947000) *Bles03* (cg13547237)
Opposite Methylation Pattern	*GEFT* (cg02364642) *PLK1* (cg26003813)	*MBNL1* (cg14423778)

Details on the previous findings of differential methylation associations with human aging that overlap with the current findings of differential methylation associations with baboon aging. Table outlines several CpG sites associated with genes that were found to be differentially methylated in humans of different ages and baboons of different ages ($\Delta\beta \geq 0.1$) and whether the previously identified human methylation patterns (hypermethylation vs. hypomethylation) of each site were similar to that in baboons. Additional details on this CpG comparison or a more general gene comparison between humans and baboons can be found in APPENDIX Z.

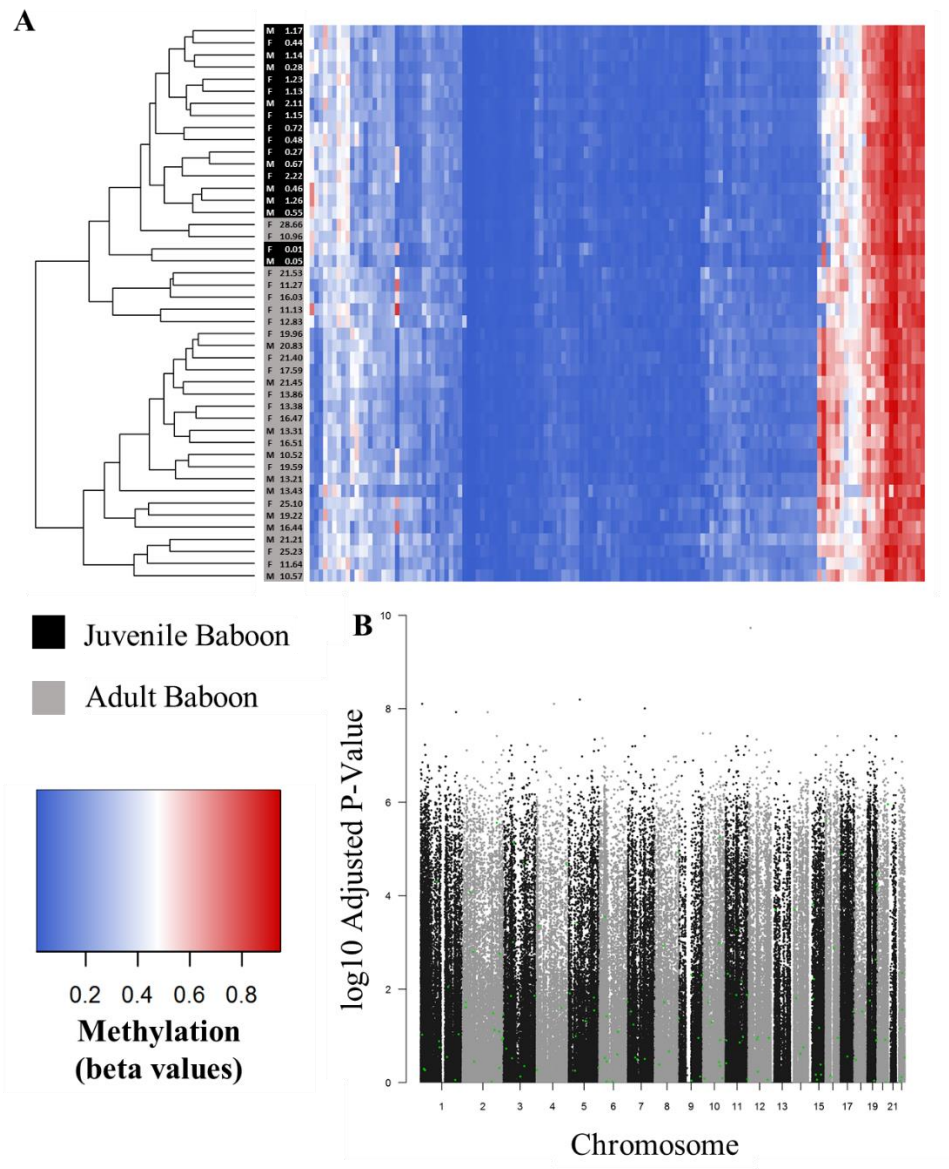

Figure 29. Methylation Levels at DMPs in the EPIC Array Baboon Aging Study that are also Associated with Aging in Humans.

(A) Heatmap depicting the DNA methylation levels (β values) of all CpG sites (n=137, x-axis) that are differentially methylated in humans with respect to age that were also tested in the current baboon study. The sex and age of each baboon sample tested are also provided (n=46, y-axis). Red indicates higher methylation at a DMP, while blue indicates

lower methylation at a DMP. The dendrogram of all samples (y-axis) clusters individuals based on the similarity of their methylation patterns. Although the majority of these sites were not significantly differentially methylated between adult and juvenile baboons, they do generally separate juveniles from adults, with the exception of some female adults (ages 28.66 years, 10.96 years, 21.53 years, 11.27 years, 16.03 years, 11.13 years, and 12.83 years) which cluster with juveniles as opposed to adults. (B) Manhattan plot showing the log10 adjusted p-values of all positions examined between adults and juveniles in all baboon samples (n=46). Samples highlighted in green are the DMPs that are differentially methylated in humans with respect to age, that were also tested in the current baboon study, and that are displayed in the heatmap.

Although the majority of age-related methylation loci in humans that were tested in the current baboon study (n=137) were not significantly differentially methylated between adult and juvenile baboons, they do generally separate these age cohorts (Figure 29). However, when assessing the quantitative methylation levels at each of these CpGs, seven female adults (ages 28.66 years, 10.96 years, 21.53 years, 11.27 years, 16.03 years, 11.13 years, and 12.83 years) appear to cluster with juveniles as opposed to other adult baboons. Three of these adult baboons are among the five that diverged from the expected when considering all significant DMPs (Figure 28). Of those CpG sites that are significant age-related DMPs in baboon and match those known in humans (n=14), some show methylation patterns that are identical between species, while others have opposing patterns (Table 11) (Koch and Wagner 2011; Horvath 2013).

For example, *KCNQ1DN* displays general hypermethylation with age in both humans (Koch and Wagner 2011) and baboons (Figure 30). In baboons, out of the 50 CpG sites associated with *KCNQ1DN* that were examined, 36 DMPs were identified with $\Delta\beta \geq 0.1$. Of these, 30 DMPs were hypermethylated in adults, and 6 were hypomethylated in adults (APPENDIX AA). Additionally, 1 DMR was identified in close proximity to this gene that is 1,517bp long and contains 29 CpGs (hg19 chr11:2889602-2891118), and this DMR had the highest level of significance among all identified DMRs (APPENDIX W). Lastly, the CpG site that is hypermethylated with age in humans (cg01530101) (Koch and Wagner 2011) is also hypermethylated with age in baboons.

Conversely, *MBNL1* displays opposing methylation patterns in humans and baboons, with age-associated hypomethylation in humans (Horvath 2013) and predominantly age-associated hypermethylation in baboons (Figure 31). In baboon bone, out of the 40 CpG sites associated with *MBNL1* that were examined, 4 DMPs were identified with $\Delta\beta \geq 0.1$. Of these, 3 DMPs were hypermethylated in adults, and 1 was hypomethylated in adults (APPENDIX AA). Additionally, 1 DMR that is 2,723bp long and contains 16 CpGs (hg19 chr3:151984822-151987544) (APPENDIX W) was identified in close proximity to this gene. Lastly, the CpG site that is hypomethylated with age in humans (cg14423778) (Horvath 2013) is hypermethylated with age in baboons.

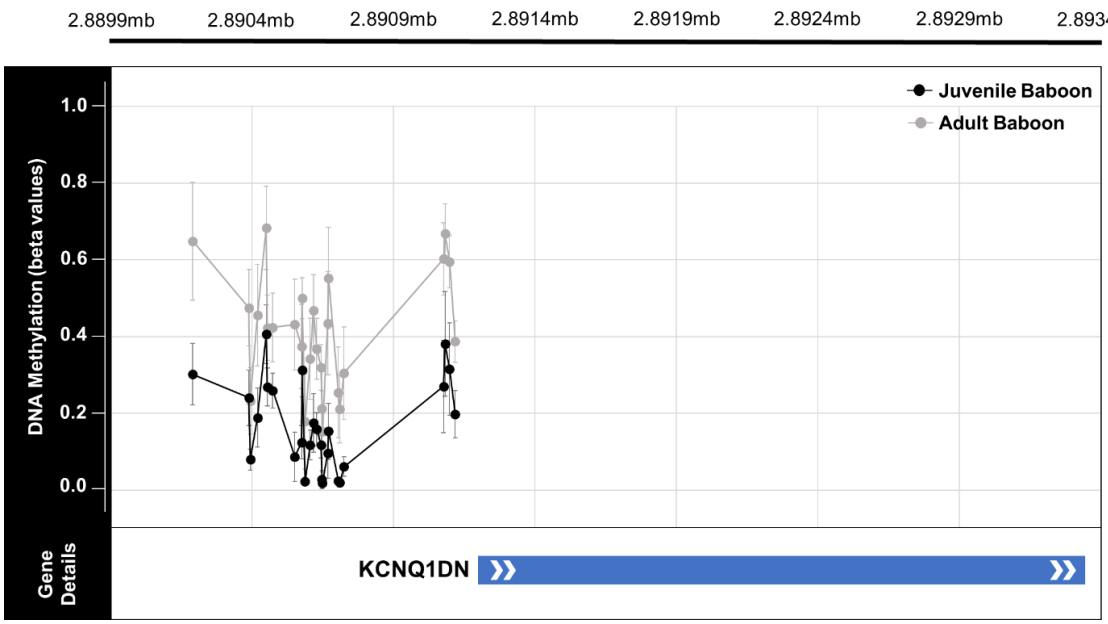

Figure 30. Methylation Levels Across *KCNQ1DN* in Adult and Juvenile Baboons. Plot of the methylation levels of significant DMPs across the *KCNQ1DN* gene (hg19 chr11:2891263-2893335). Plot shows the average β values for each DMP with error bars indicating 1 standard deviation in each direction for each comparative group (light grey = adult baboon bone, black = juvenile baboon bone). DMP chromosomal position in relation to the *KCNQ1DN* gene is also depicted. Similar to humans, *KCNQ1DN* is overwhelmingly hypermethylated in older baboons as compared to younger baboons (Koch and Wagner 2011). In baboons, 36 DMPs were identified with $\Delta\beta \geq 0.1$ (26 of which are depicted here). Of these, 30 DMPs were hypermethylated in adults, and 6 were hypomethylated in adults. Additionally, 1 DMR was identified in close proximity to this gene that is 1517bp long and contains 29 CpGs (hg19 chr11:2889602-2891118), and this DMR had the highest level of significance among all identified DMRs (Appendix W). Lastly, the CpG site that is hypermethylated with age in humans (cg01530101) (Koch and

Wagner 2011) is also hypermethylated with age in baboons. See Appendix AA for additional information.

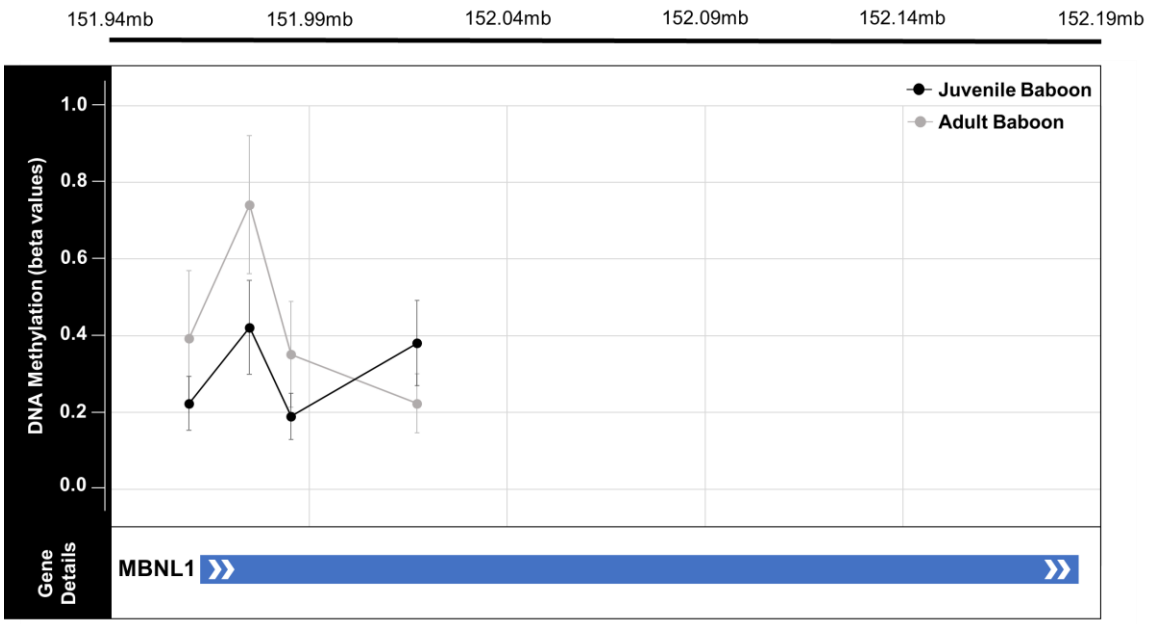

Figure 31. Methylation Levels Across *MBNL1* in Adult and Juvenile Baboons.

Plot of the methylation levels of significant DMPs across the *MBNL1* gene (hg19 chr3:151962120-152183569). Plot shows the average β values for each DMP with error bars indicating 1 standard deviation in each direction for each comparative group (light grey = adult baboon bone, black = juvenile baboon bone). DMP chromosomal position in relation to the *MBNL1* gene is also depicted. In contrast to humans in which *MBNL1* becomes hypomethylated with age (Horvath 2013), *MBNL1* is primarily hypermethylated in older baboons as compared to younger baboons. In baboons, 4 DMPs were identified with $\Delta\beta \geq 0.1$ (all of which are depicted here). Of these, 3 DMPs were hypermethylated in adults, and 1 was hypomethylated in adults. Additionally, 1 DMR was identified in close proximity to this gene that is 2723bp long and contains 16 CpGs (hg19

chr3:151984822-151987544) (Appendix W). Lastly, the CpG site that is hypomethylated with age in humans (cg14423778) (Horvath 2013) is hypermethylated with age in baboons. See Appendix AA for additional information.

Discussion

Here, I identified DNA methylation patterns in femoral trabecular bone from adult and juvenile baboons. This was done in order to identify age-related changes in DNA methylation and assess the evolutionary conservation of these changes in the primate lineage.

Using *in silico* probe filtering methods (Hernando-Herraez et al. 2013; Ong et al. 2014) and additional filtering criteria, I retained 191,631 EPIC array probes that reliably mapped to the baboon genome, contained a CpG locus, met specific alignment filter criteria (Figure 9), maintained a wide distribution throughout the genome (APPENDIX H), had hybridization efficiencies significantly correlated with the alignment quality of each probe to the baboon genome (APPENDIX T), and predominantly passed quality controls to produce robust signals on the array (Figure 25). The size and efficiencies of this final set of probes was comparable to previous findings for the 450K array (see Chapter 2) and EPIC array (see Chapter 3).

From these filtered probes, several significant DMPs were detected between different baboon age cohorts (Figure 26, APPENDIX U). Many of these DMPs had biologically insignificant changes in mean methylation, so only those with at least a 10% change ($\Delta\beta \geq 0.1$) were considered further. This reduced the overall number of DMPs considered to 17.25% of the originally interrogated sites. The number of DMRs identified

between age cohorts showed trends similar to those for DMPs, and the genes associated with these DMRs overlapped with those associated with DMPs (Table 9, APPENDIX W). These findings further confirm the results of the DMP analyses and their functional implications.

Overall, within baboons, the number of DMPs and DMRs found between adults and juveniles is greater than that found between individuals with different skeletal disease states (osteoarthritic vs. healthy) but smaller than that found between different skeletal tissue types (bone vs. cartilage) (see Chapter 3). This finding is to be expected since the amount of DNA methylation which influences some aspects of gene regulation and expression (Suzuki and Bird 2008; Singer et al. 2015) should fluctuate depending on how different cellular functions are between comparative groups (Zhang et al. 2013). In the case of different tissues, substantial DNA methylation differences may enable some of the distinct gene regulation and expression that is necessary for the cells in these tissues to promote different tissue functions. In the case of different age cohorts, within the same skeletal tissue, there is an emphasis on growth and development in juveniles as compared to an emphasis on maintenance in adults. These differences likely require some DNA methylation and gene regulatory differences, but not so much as to alter the general function of this tissue to anything other than bone-related functions. Lastly, in the case of osteoarthritic disease states, fewer regulatory changes are likely needed to initiate the dysregulation of tissue function than to promote the substantial functional differences between age ranges and tissues types.

The epigenetic profiles of each age cohort, as defined by the significant DMPs with $\Delta\beta \geq 0.1$, generally distinguished adult baboons from juvenile baboons. For

instance, global methylation levels in adult baboons were decreased as compared to juveniles (Table 10). This genome-wide reduction of methylation with aging has also been observed in a phylogenetically diverse set of organisms (Berdyshev et al. 1967; Vanyushin et al. 1973; Wilson et al. 1987). Additionally, because juvenile bone tissue is undergoing growth and development while adult bone tissue is focused on maintenance, it is fitting that the significant DMPs identified are associated with several genes predominantly involved in developmental processes and pathways (Figure 27, APPENDIX X, APPENDIX Y). Also of note is the abundance of genes associated with pathways involved in the development and progression of diseases of aging, such as cancer (APPENDIX Y).

Despite these overarching epigenetic features that distinguish adult baboons from juvenile baboons, the specific quantitative methylation levels at each DMP do not completely separate adults and juveniles into distinct clusters (Figure 28). Rather, five female adult baboons appear to cluster more closely with juvenile baboons than with other adult baboons. This might seem reasonable if the adults clustering with juveniles were the youngest individuals in the adult cohort. However, their ages range from 10.96 years to 28.66 years, which encompasses the maximum age spread among the adult female baboons included in this study. Some adult male baboons were slightly younger than this age range (10.52 and 10.57 years), but the 28.66-year-old female baboon is the oldest individual in the sample set by almost 3.5 years.

These unexpected findings may be due to potential sex differences in epigenetic signatures or to general discrepancies between chronological ages and epigenetic ages. In females, bone tissues are susceptible to variations in hormone regulation with age

(Väänänen and Härkönen 1996) and the nutritional demands required during pregnancy (Kovacs 2000), and these effects may influence the epigenetic age of females in ways that have not been identified yet. Additionally, studies in humans have found that not all individuals with identical chronological ages have identical epigenetic ages. For instance, people with epigenetic ages older than their chronological ages have overall increased risks for death from all causes (Marioni, Shah, McRae, Chen, et al. 2015; Christiansen et al. 2016; Perna et al. 2016), while people with epigenetic ages younger than their chronological ages have been associated with the capacity to become semi-supercentenarians (Marioni, Shah, McRae, Chen, et al. 2015). The discrepancy between epigenetic age and chronological age in some baboons in this study may simply be an example of the same discrepancy that has been observed in humans. Alternatively, epigenetic signatures of age may not be as clearly demarcated in bone tissues or may not be as robust at separating baboons age cohorts. If the former is true, though, researchers should be wary about extending these methods to estimate the ages of skeletal remains preserved in the archaeological record. Instead, evaluating DNA methylation patterns in teeth may serve as a more suitable method for aging efforts in these conditions (Giuliani et al. 2016). Conversely, the sample size utilized in this study may simply not have enough power to fully differentiate adults and juveniles.

These age-related DMPs, however, do provide further insight into the evolution of aging epigenetics. Out of the 137 CpG that have been found to be differentially methylated with age in humans (Koch and Wagner 2011; Horvath 2013) and which were also tested in baboons in the current study, 14 were found to be significantly differentially methylated across baboon ages (Table 11). Of these 14 CpG sites, some

show methylation pattern that are identical between humans and baboons, while others have reversed patterns between species (Table 11). Despite the majority of these human age-related DMPs not being significantly differentially methylated between baboon age cohorts, the methylation patterns across these sites generally separated adult baboons from juvenile baboons, except for seven female adults with ages ranging from 10.96 years to 28.66 years that cluster more closely with juveniles than other adult baboons (Figure 29). Three of these adult female baboons are identical to those that clustered with juvenile baboons when considering all significant DMPs (Figure 28). As described above, this incomplete sorting of different age cohorts may be due to subtle sex effects on the epigenome, the possible disconnect between chronological and epigenetic ages, or an aspect of the study design. Alternatively, since many of these human age-related DMPs are not significantly differentially methylated in baboons, there may not be enough power to distinctly cluster each age group using these particular loci. Thus, while some age-related epigenetic patterns appear to be evolutionarily conserved between baboons and humans, many show no evidence of conservation, and some display opposite signals. Specific examples of this variation in conservation are as follows.

In some instances, the age-related methylation patterns observed in baboons matched those observed in humans. For instance, *KCNQ1DN* (Gene ID: 55539) displays similar age associated hypermethylation in humans (Koch and Wagner 2011) and baboons (Figure 30). *KCNQ1DN*, also known as KCNQ1 downstream neighbor, is a non-protein coding gene that is imprinted and expressed by the maternal allele but otherwise has a generally unknown function. It has been associated with aging (Koch and Wagner 2011) and may possibly be involved in Wilms' tumorigenesis, which is characterized by

maternal-specific loss of heterozygosity of the chromosomal region surrounding the *KCNQ1DN* imprinted gene, as well as several other imprinted genes (Xin et al. 2000). Because of its limited functional annotations, it is unclear whether there is a biological function for an increase in methylation at this gene with age. Nevertheless, the findings of this study suggest that this age-related methylation candidate gene and its pattern of change over time appear to be evolutionarily conserved within the primate lineage.

Conversely, for some genes, the age-related methylation patterns observed in baboons were opposite to those observed in humans. This is true for the gene *MBNL1* (Gene ID: 4154), which displays hypomethylation with age in humans (Horvath 2013) and hypermethylation with age in baboons (Figure 31). *MBNL1*, also known as muscleblind like splicing regulator 1, codes for a splicing factor protein that regulates alternative splicing of pre-mRNAs (Gates et al. 2011; Purcell et al. 2012; Edge et al. 2013). It also appears to be involved in the progression of myotonic dystrophy which is a disorder characterized by muscle weakness and wasting (Gates et al. 2011; Purcell et al. 2012; Edge et al. 2013). Although its function in gene regulation via splicing is important, its role in aging is unclear. Nevertheless, the findings of this study suggest that this age-related methylation candidate gene and its pattern of change over time are not evolutionarily conserved within the primate lineage.

Overall, these age-related differential methylated findings indicate that global methylation patterns between age cohorts are comparable between humans and baboon and that the genes associated with these epigenetic changes are involved in developmental processes and pathways, which thus, may play crucial roles in distinguishing juvenile and adult phenotypes. However, when examined more closely,

these methylation patterns do not definitively distinguish all adults from all juveniles. Rather, there appears to be a mixing of age-related epigenetic signatures that may be due to subtle sex effects, the possible disconnect between chronological and epigenetic ages, or an aspect of the study design. Finally, the findings of this study indicate that some DNA methylation patterns associated with aging are evolutionarily conserved between humans and baboons, while others are not. Species differences in aging epigenetics may be due to general speciation events that took place during the evolution of these taxonomic groups, to slight differences in the aging processes of these species, or artifacts of the experimental design of the current baboon study or previous human aging studies.

For instance, the present baboon study includes only 46 animals, while the study which identified over 300 age-related methylation sites in humans (Horvath 2013) was a meta-analysis which included 81 human studies with an average sample size of 96 individuals. This smaller number of individuals may have reduced my power to detect significant differential methylation in baboons at several age-related methylation sites. Additionally, the CpG sites found to have age-related differential methylation across tissues (Koch and Wagner 2011; Horvath 2013) used older versions of the EPIC array, which only examined 27,000 or 450,000 sites throughout the human genome. This difference in methods may account for some differences observed between humans and baboons. Alternatively, the differences observed between humans and baboons may be a result of the differences in tissues examined. While previous human aging meta-analyses have included methylation data from skeletal tissues such as cartilage (Horvath 2013), most samples are from blood or other soft tissues (Horvath 2013; Koch and Wagner

2011), not bone. Thus, the absence of bone methylation patterns in human aging studies, may contribute to the lack of methylation pattern overlap between humans and baboon found in the present study.

Lastly, this baboon study used a stringent DMP cutoff threshold that limited the resulting genes to only those that were associated with DMPs having an average change in mean methylation between comparative groups greater than or equal to 10%. This was a precaution against including sites that likely had little biological relevance (Hernando-Herraez et al. 2013). However, age-related epigenetic changes may not have any explicit regulatory effects on these genes. Some CpG sites found to have age-related differential methylation across tissues (Koch and Wagner 2011) were selected without consideration of gene function because previous studies have shown that individual site-specific methylation changes are not readily associated with differential gene expression (Koch et al. 2011; Bork et al. 2010; Y. Chen et al. 2011). Rather, differential gene expression is made possible through the accumulation of several methylation changes within promotor regions (Suzuki and Bird 2008) or across the gene body (Singer et al. 2015). Thus, the increased filtering used in the current baboon study may have increased the number of false negatives, and further work to identify epigenetic aging profiles in nonhuman primate tissues should be done to validate the current findings of this research.

Regardless of these potential confounding factors, using baboons in this study has begun to clarify the evolutionary conservation of epigenetic aging processes within the primate lineage. This is the first study to specifically assess aging effects on the DNA methylation profiles of skeletal tissues from a nonhuman primate. In summary, from an evolutionary perspective, the findings of this study inform our understanding of DNA

methylation variation in one skeletal tissue from another primate, as well as the degree to which chronological age affects this variation. These findings warrant further investigation in a larger and more phylogenetically diverse sample set, and future research in this area will provide insight into the evolution of aging and senescence.

Acknowledgements

Anne Sheldrake, Jaydee Turner, Kara Peterson, Mel Carless, & Laura Cox

Department of Genetics, Texas Biomedical Research Institute, San Antonio, TX, USA

Funding

Leakey Foundation Research Grant for Doctoral Students to G. Housman

Wenner-Gren Foundation Dissertation Fieldwork Grant (Gr. 9310) to G. Housman

James F. Nacey Fellowship from the Nacey Maggioncalda Foundation to G. Housman

International Primatological Society Research Grant to G. Housman

Sigma Xi Grant-in-Aid of Research to G. Housman

Center for Evolution and Medicine Venture Fund (ASU) to G. Housman

Graduate Research and Support Program Grant (GPSA, ASU) to G. Housman

Graduate Student Research Grant (SHESC, ASU) to G. Housman

CHAPTER 5

INTRA- AND INTER-SPECIFIC INVESTIGATIONS OF SKELETAL DNA METHYLATION PATTERNS AND FEMUR MORPHOLOGY IN NONHUMAN PRIMATES

Abstract

Complex skeletal traits are the product of genetic, environmental, and epigenetic mechanisms. DNA methylation is one such epigenetic mechanism which regulates gene expression. Because epigenetic modifications are influential in the development and maintenance of skeletal traits, they may also contribute to the evolution of primate skeletal anatomy. Further, they may also contribute to general differences between species. Skeletal morphology in relation to DNA methylation variation has not, however, been assessed in nonhuman primates. Additionally, epigenetic variation in a phylogenetically diverse set of nonhuman primates has not been evaluated. This study addresses this knowledge gap by identifying intra-specific methylation variation in primate skeletal tissue to test the hypothesis that specific features of femur morphology are associated with specific variations in methylation and by identifying inter-specific methylation variation between nonhuman primate species to determine if lineage specific patterns have evolved and may contribute to species-specific morphologies. Here, I used the Illumina Infinium MethylationEPIC BeadChip (EPIC array) to identify DNA methylation patterns in femur trabecular bone from baboons (n=28), macaques (n=10), vervets (n=10), chimpanzees (n=4), and marmosets (n=6). I validated that the hybridization efficiency of EPIC array probes is related to the degree of sequence

similarity between the probes and each nonhuman primate genome, and I determined that approximately 39%, 39%, 39%, 76%, and 17% of the EPIC array probes reliably align to the baboon, macaque, vervet, chimpanzee, and marmoset genomes, respectively, contain a CpG site of interest, and maintain a wide distribution throughout the genome. I also found that filtering probes using alignment similarity criteria retains more efficiently hybridized probes than filtering probes using gene symbol similarity criteria. Additionally, significant differential methylation was identified in a subset of morphological variants within species. However, these significantly differentially methylated positions (DMPs) likely do not have large biological effects and may be confounded by other variables associated with morphological variation. Furthermore, I found several sites that show species-specific methylation patterns. Higher resolution of methylation variation across a subset of these regions confirms these patterns and provides more insight into their evolution history. Finally, genome-wide DNA methylation patterns across all 39,802 sites examined produce a topology that reflects known phylogenetic relationships between taxa. From an evolutionary perspective, these findings give us an appreciation of DNA methylation variation in skeletal tissue within and among five nonhuman primate species. They also provide insight into the degree to which this epigenetic variation relates to variation in skeletal morphology and taxonomic differences. Expansion of this sample set and more focused testing of additional genes and the resulting functions of DNA methylation changes will further inform our understanding of epigenetic regulation and complex trait evolution in primates.

Key Words

DNA methylation, nonhuman primates, evolution, epigenome, bone

Introduction

Primates distinguish themselves from other mammals with their unique suite of anatomical features that initially enabled arboreal niche occupation and subsequently evolved to fit a myriad of habitats and forms of locomotion, the most unique being hominin bipedalism. Skeletal features related to varied body forms are often described as the result of environmental adaptations. However, skeletal morphology is more accurately defined as the result of complex processes, and environmental (Henriksen et al. 2014; Macrini et al. 2013), genetic (Goldring and Marcu 2012), and epigenetic mechanisms (Delgado-Calle et al. 2013; García-Ibarbia et al. 2013; Y. Liu et al. 2013; Reynard et al. 2014) all contribute to these phenotypes. DNA methylation is one such epigenetic mechanism that regulates gene expression, and given its involvement in the development and maintenance of skeletal traits, it may also be involved in the evolution of diverse skeletal anatomies and speciation divergences in general.

Epigenetic contributions to primate phenotypic variation were first considered by (King and Wilson 1975). They proposed that anatomical and behavioral differences between humans and chimpanzees were more likely "based on changes in the mechanisms controlling the expression of genes than on sequence changes in proteins" (King and Wilson 1975, p.107), and studies to understand methylation variation across species began soon afterwards (Gama-Sosa et al. 1983). Although general changes to mammalian epigenomes have been examined (Sharif et al. 2010), primate-specific DNA

methylation studies have been limited. Most work has focused on epigenetic variation in humans – how it varies across distinct tissues within individuals (Lister et al. 2009; Slieker et al. 2013), across different individuals (Oates et al. 2006; Petronis et al. 2003; Weksberg et al. 2002), across populations (Heyn et al. 2013; Rakyan et al. 2004), in relation to aging processes (Fraga et al. 2005), in relation to diet (Jacob et al. 1998; Rampersaud et al. 2000; Shelnutt et al. 2004), and how it is inherited across generations (Flanagan et al. 2006; Gibbs et al. 2010; Suter et al. 2004; van Dongen et al. 2014). Important within species methylation variants have been identified in these studies.

Similarly, epigenetic variation has been identified between primate species. Support was initially inferred from underlying genomic sequences (Prendergast et al. 2007; Haygood et al. 2007; C. G. Bell et al. 2012). For instance, several promoter CpG densities vary across primates. These likely relate to regulatory methylation differences across species as primate CpG densities correlate with methylation levels (Weber et al. 2007). Additionally, gene expression studies, which primarily focus on brain tissues (Babbitt et al. 2010; Cáceres et al. 2003; Warner et al. 2009) and a small set of other soft tissues (Blekhman et al. 2008; Karere et al. 2010; Karere et al. 2012, 2; Karere et al. 2013; Tung et al. 2015), have also noted regulatory differences across species. Brain tissues tend to be the focus of DNA methylation studies, as well, and methylation differences in these tissues have evolved across primates and contributed to resultant brain phenotypes and disease vulnerabilities (Enard et al. 2004; Farcas et al. 2009; Zeng et al. 2012; Kothapalli et al. 2007; Provencal et al. 2012). Thus, methylation-phenotype relationships can be identified in primates. Primate methylation patterns in blood cells and other soft tissues have also been studied, but not to the same degree (Pai et al. 2011;

Molaro et al. 2011; Martin et al. 2011; Fukuda et al. 2013; Lindskog et al. 2014; Hernando-Herraez et al. 2013). Interestingly, two studies using soft tissues and blood identified differential methylation and expression of genes essential for skeletal development (*RUNX1*, *RUNX3*, and *COL2A1*) between some primates (Lindskog et al. 2014; Hernando-Herraez et al. 2013). Additionally, ancient DNA methylation studies have found other skeletal developmental genes (*HOXD* complex) differentially methylated among modern humans and ancient hominins (Gokhman et al. 2014). These findings suggest that primates do exhibit distinct epigenetic patterns and that the epigenetics of skeletal development may be an important direction of future research.

The skeletal system comprises an important set of complex phenotypes, and the underlying molecular contributions to this system have been heavily studied in humans and model organism (Zamli et al. 2014), but little of this work has taken an evolutionary perspective (Ostrer et al. 2006; Rugg-Gunn et al. 2005). As with most phenotypes, skeletal phenotypes are influenced by genetic and environmental factors, but epigenetic factors such as DNA methylation are now thought to play an important role in skeletal tissue development and maintenance (Delgado-Calle et al. 2013; García-Ibarbia et al. 2013; Goldring and Marcu 2012; Iliopoulos et al. 2008; Loughlin and Reynard 2015; Ramos et al. 2014; Reynard et al. 2014). For instance, epigenetic processes are influential in regulating skeletal muscle development (Brand-Saberi 2005; Ling et al. 2012; Palacios and Puri 2006; Pandorf et al. 2009; Zwetsloot et al. 2009) which can impact the adjacent skeletal scaffolding system. Several genes involved in human skeletal development appear to be differentially methylated across fetal and adult developmental stages (de Andrés et al. 2013). Lastly, methylation variation in humans and model organisms has

been implicated in several skeletal pathologies and disorders, such as rheumatoid arthritis, osteoporosis, and osteoarthritis (Bovée et al. 2010; Delgado-Calle et al. 2013; den Hollander et al. 2014; Dimitriou et al. 2011; Fernández-Tajes et al. 2014; García-Ibarbia et al. 2013; Goldring and Marcu 2012; Iliopoulos et al. 2008; Jeffries et al. 2014; Jeffries et al. 2016; Kasaai et al. 2013; Y. Liu et al. 2013; Loughlin and Reynard 2015; Moazedi-Fuerst et al. 2014; Ralston and Uitterlinden 2010; Ramos et al. 2014; Reynard et al. 2014; Rivadeneira et al. 2009; Rushton et al. 2014b). Some of these studies are the first to assess methylation patterns in human skeletal tissues (Delgado-Calle et al. 2013; den Hollander et al. 2014; Fernández-Tajes et al. 2014; García-Ibarbia et al. 2013; Jeffries et al. 2014; Moazedi-Fuerst et al. 2014; Rushton et al. 2014b), which is important for truly identifying the relationship between epigenetic variation and skeletal phenotypic variation.

The skeletal system is an important complex phenotype to examine further in primates because it shows varied morphologies across taxa, because underlying skeletal differences contribute to overall morphological diversity, and because skeletal morphology is readily used to reconstruct the anatomy and locomotor capabilities of extinct species within the primate lineage (Ankel-Simons 2007; Fleagle 1999; Leigh and Shea 1995; Schultz 1930; Schultz 1937). With respect to human evolution, femur morphology is of particular interest because of its role in bipedalism. Specific human features, such as the bicondylar angle and lateral lipping, biomechanically enable bipedalism. However, the range of mechanisms that enable the development of such skeletal features are not entirely understood. Specifically, the involvement of epigenetic variation in nonpathological skeletal phenotypes has not been studied. Nevertheless, it is

important to consider these mechanisms when interpreting skeletal features, such as those seen in the fossil record, and when making broad conclusions about primate evolution from such elements. Lastly, the emerging field of ancient epigenetics is beginning to assess DNA methylation patterns in ancient hominin skeletal remains. Many of these come from anatomically modern humans, but additional work is being done in Neandertal and Denisovan remains (Smith et al. 2015; Gokhman et al. 2014). However, the field does not have a clear understanding of DNA methylation patterns in skeletal tissues from nonhuman primates, so it's difficult to put these ancient hominin skeletal epigenetic patterns into a broader phylogenetic or evolutionary context.

Overall, there are clear knowledge gaps in our understanding of nonhuman primate skeletal complexity in relation to epigenetic variation and epigenetic differences between phylogenetically diverse nonhuman primate species. The present study begins to remedy this by assessing how genome-wide and gene-specific DNA methylation in primate skeletal tissues varies intra- and inter-specifically and in relation to femur form. Specifically, for this study, I explored the evolution of the epigenome and its relation to nonpathological skeletal traits within the primate lineage by identifying DNA methylation patterns in femur trabecular bone from 28 baboons, 10 macaques, 10 vervets, 4 chimpanzees, and 6 marmosets and assessed how DNA methylation variation is associated intra-specifically with respect to femur morphology variation and inter-specifically.

Methods

Ethics Statement

Nonhuman primate tissue samples included were opportunistically collected at routine necropsy of these animals. No animals were sacrificed for this study, and no living animals were used in this study.

Nonhuman Primate Samples

Nonhuman primate samples come from captive colonies of chimpanzees (*Pan troglodytes*), baboons (*Papio spp.*), rhesus macaques (*Macaca mulatta*), and marmosets (*Callithrix jacchus*) from the Southwest National Primate Research Center in the Texas Biomedical Research Institute, as well as vervets (*Chlorocebus aethiops*) from the Wake Forest/UCLA Vervet Research Colony in North Carolina.

Femora were opportunistically collected at routine necropsy of these animals and stored in -20°C freezers at the Texas Biomedical Research Institute after dissection. These preparation and storage conditions ensured the preservation of skeletal DNA methylation patterns.

Samples include baboons (n=28), macaques (n=10), vervets (n=10), chimpanzees (n=4), and marmosets (n=6). Age ranges are comparable between each group, and both sexes are represented (Figure 32, APPENDIX BB). This is important as many skeletal features, such as overall bone shape and susceptibility to diseases of skeletal maintenance (e.g., OA), are sex and age dependent.

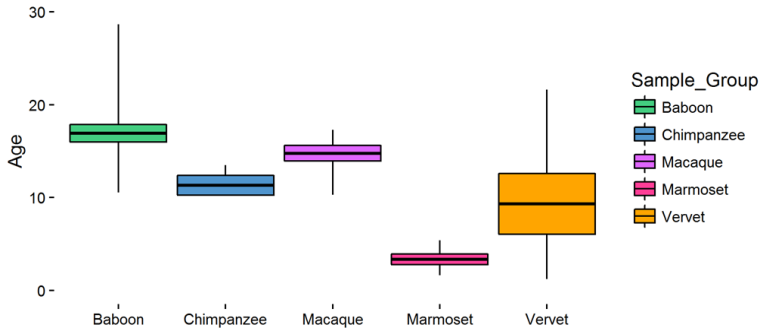

Figure 32. Nonhuman Primate Sample Set Ages for EPIC Array Intra- and Inter-Specific Study.

Box plot of sample ages. Box depicts the average age and one standard deviation, and whiskers depict the full range of ages. Baboons (n=28) are 16.90±5.02 years, chimpanzees (n=4) are 11.31±1.87 years, macaques (n=10) are 14.75±2.65 years, marmosets (n=6) are 3.34±1.41 years, and vervets (n=10) are 9.31±10.30 years.

Assessment of Nonpathological Morphologies

On nonhuman primate femora, 29 linear morphology traits (Figure 33, Table 12) were measured using calipers. These measurements characterize overall femur shape. Error for each linear measurement was determined by performing triplicate measurements on approximately 10% of the samples in each comparative group. These measurements were spaced throughout the entire data collection period. Error was calculated as the mean absolute difference divided by the mean (Corner et al. 1992; White and Folkens 2000). Only measurements with less than 5% error were retained for downstream analyses (APPENDIX CC).

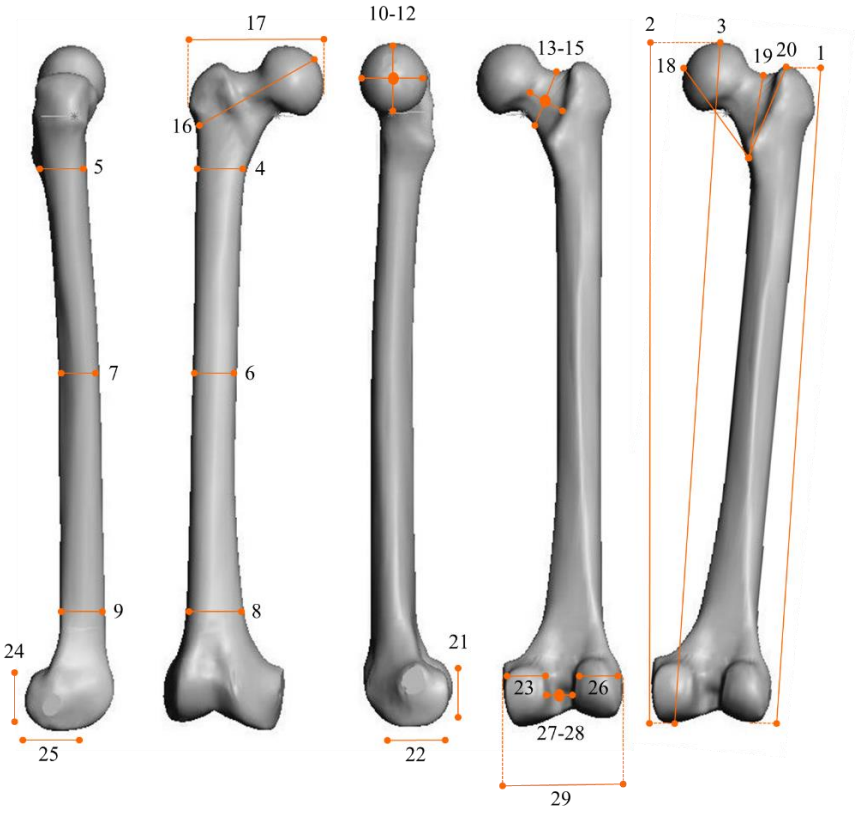

Figure 33. Nonhuman Primate Morphological Measurements.

Linear morphological measurements collected from the right femur of each nonhuman primate. Measurements selected were based on (McHenry and Corruccini 1978; Terzidis et al. 2012). See Table 12 for a detailed description of these measurements.

Table 12. Nonhuman Primate Morphological Measurements.

No.	Measurement	Definition
1	femur length	greatest distance between greater trochanter and lateral condyle parallel with long axis of bone
2	bicondylar femur length	distance between most superior point on head and plane joining most inferior points on lateral and medial condyles
3	maximum femur length	distance between most superior point on head and most inferior point on medial condyle
4	superior shaft width	medial-lateral width of shaft inferior to lesser trochanter
5	superior shaft depth	anterior-posterior depth of shaft inferior to lesser trochanter
6	middle shaft width	medial-lateral width of shaft at midpoint
7	middle shaft depth	anterior-posterior depth of shaft at midpoint
8	inferior shaft width	medial-lateral width of inferior most portion of shaft
9	inferior shaft depth	anterior-posterior depth of inferior most portion of shaft
10	head height	greatest superior-inferior distance along junction between the head and neck
11	head length	greatest anterior-posterior distance along junction between the head and neck
12	head width	distance between the articular surface of head and junction between the head and neck
13	anatomical neck length	length from intertrochanteric crest to junction between head and neck
14	anatomical neck height	superior-inferior width at midpoint of anatomical neck
15	anatomical neck depth	anterior-posterior depth at midpoint of anatomical neck
16	biomechanical neck length	greatest distance between articular surface of head and lateral surface of greater trochanter parallel with axis of neck
17	proximal width	projected distance between most medial point on head and lateral point on greater trochanter
18	lesser trochanter to head	maximum distance between inferior border of lesser trochanter and medial surface of head
19	lesser trochanter to neck	minimum distance between inferior border of lesser trochanter and superior border of neck
20	lesser to greater trochanter	distance between inferior border of lesser trochanter and center of superior border of greater trochanter
21	medial condyle height	greatest superior-inferior distance along medial condyle
22	medial condyle depth	greatest anterior-posterior distance along articular surface of medial condyle
23	medial condyle width	greatest medial-lateral distance along medial condyle
24	lateral condyle height	greatest superior-inferior distance along lateral condyle
25	lateral condyle depth	greatest anterior-posterior distance along articular surface of lateral condyle
26	lateral condyle width	greatest medial-lateral distance along lateral condyle
27	intercondylar notch width	distance between the medial and lateral condyles
28	intercondylar notch depth	anterior-posterior depth between posterior aspect of distal shaft and posterior aspect of condyles
29	bicondylar width	maximum transverse diameter of distal end

Linear morphological measurements collected from the right femur of each nonhuman primate. Measurements selected were based on (McHenry and Corruccini 1978; Terzidis et al. 2012). See Figure 33 for a visual representation of these measurements.

DNA Extraction

DNA was extracted from femoral trabecular bone using a phenol-chloroform protocol optimized for skeletal tissues (Barnett and Larson 2012). From the distal femoral condyles, trabecular bone was collected using coring devices and pulverized into bone dust using a SPEX SamplePrep Freezer/Mill. Specifically, bone cores were obtained from a transverse plane through the center of the medial condyle on the right distal femur, such that the articular surface remained preserved. Cortical bone was removed from these cores using a Dremel.

Trabecular bone is used in this study because several human skeletal epigenetic studies are based on trabecular bone, and it is important to standardize tissue type for comparative purposes. Trabecular bone comprises the internal spongy osseous tissue that contributes to femoral shape morphology and remodeling, which begins before birth and continues throughout life (Clarke 2008). However, trabecular bone in growing individuals influences both trabecular and cortical morphology in adulthood (Q. Wang et al. 2011), and this suggests that the epigenetics of trabecular bone may be of more interest initially than that of cortical bone. Lastly, although trabecular bone is not ideal for epigenetic analyses because it contains several cell types (Horvath, Mah, et al. 2015), statistical methods can correct for this heterogeneity.

Cartilage methylation patterns are known to vary between joints and between different sites within a joint (den Hollander et al. 2014; Jeffries et al. 2016; Loughlin and Reynard 2015; Moazedi-Fuerst et al. 2014; Rushton et al. 2014b). Although similar studies of bone methylation patterns have not been conducted yet, the number and types of cells, and therefore epigenetic signatures, are expected to vary across different portions

of the femur. Thus, tissues were collected from the same portion of the femur in order to minimize this variation between samples and comparative groups.

Genome-Wide DNA Methylation Profiling

Genome-wide DNA methylation was assessed using Illumina Infinium MethylationEPIC microarrays (EPIC array). These arrays analyze the methylation status of over 850,000 sites throughout the genome, covering over 90% of the sites on the Infinium HumanMethylation450 BeadChip as well as an additional 350,000 sites within enhancer regions. For each sample, 400ng of genomic DNA was bisulfite converted using the EZ DNA Methylation™ Gold Kit according to the manufacturer's instructions (Zymo Research), with modifications described in the Infinium Methylation Assay Protocol. Following manufacturer guidelines (Illumina), this processed DNA was then whole-genome amplified, enzymatically fragmented, hybridized to the arrays, and imaged using the Illumina iScan system. The array data discussed here are available in APPENDIX DD.

Methylation Data Processing

Raw fluorescent data were normalized to account for the noise inherent within and between the arrays themselves. Specifically, I performed a normal-exponential out-of-band (Noob) background correction method with dye-bias normalization (Triche et al. 2013) to adjust for background fluorescence and dye-based biases and followed this with a between-array normalization method (functional normalization) (Fortin et al. 2014) which removes unwanted variation by regressing out variability explained by the control

probes present on the array as implemented in the minfi package in R (Aryee et al. 2014; Fortin et al. 2016) which is part of the Bioconductor project (Huber et al. 2015). This method has been found to outperform other existing approaches for studies that compare conditions with known large-scale differences (Fortin et al. 2014), such as those assessed in this study.

After normalization, methylation values (β values) for each site were calculated as the ratio of methylated probe signal intensity to the sum of both methylated and unmethylated probe signal intensities (Equation 1). These β values range from 0 to 1 and represent the average methylation levels at each site across the entire population of cells from which DNA was extracted (0 =completely unmethylated sites, 1 = fully methylated sites).

Every β value in the Infinium platform is accompanied by a detection p-value, and those with failed detection levels (p-value > 0.05) in greater than 10% of samples were removed from downstream analyses. Additionally, samples in which more than 30% of the β value had a detection p-value > 0.05 were removed from downstream analyses.

The probes on the arrays were designed to hybridize specifically with human DNA, so my use of nonhuman primate DNA required that probes non-specific to any of the included nonhuman primate genomes, which could produce biased methylation measurements, be computationally filtered out and excluded from downstream analyses. This was accomplished using two different methods modified from (Hernando-Herraez et al. 2013; Ong et al. 2014).

For both methods, I used blastn (Altschul et al. 1997) to map the 866,837 50bp probes onto the baboon, macaque, vervet, chimpanzee, and marmoset genomes (Table

13) using an e-value threshold of e^{-10}. I retained probes that successfully mapped to each genome, had only 1 unique BLAST hit, targeted CpG sites, had 0 mismatches in 5bp closest to and including the CpG site, and had 0-2 mismatches in 45bp not including the CpG site. For the second method, which used criteria based on gene symbol similarities, I identified the closest nonhuman primate gene to each probe site and checked for corresponding gene name matches between humans and each nonhuman primate. This information was obtained from different sources for each taxon (Table 13). Only those probes with partial or complete gene matches were retained.

Table 13. Nonhuman Primate Genomes Used for Probe Filtering Methods.

Species	Assembly	Accession	Average Scaffold Length	Average Contig Length	Gene Information
Baboon (*Papio anubis*)	Panu_2.0	GCF_000264685.2	528,927	40,262	GFF, Ensembl BioMart
Macaque (*Macaca mulatta*)	Mmul_8.0.1	GCF_000772875.2	4,193,270	107,156	GFF
Vervet (*Chlorocebus aethiops*)	Chlorocebus_sabeus_1.1	GCF_000409795.2	81,825,804	162,724	GFF, Ensembl BioMart
Chimpanzee (*Pan troglodytes*)	Pan_tro_3.0	GCF_000001515.7	26,972,556	72,226	GFF
Marmoset (*Callithrix jacchus*)	Callithrix_jacchus-3.2	GCF_000004665.1	5,167,444	29,273	GFF, Ensembl BioMart

Baboon, macaque, vervet, chimpanzee, and marmoset genome assemblies and accession numbers used for probe filtering methods. Average scaffold lengths and average contig lengths for each genome also provided, as well as the gene information data sources for the gene symbol probe filtering method.

Additionally, β values associated with cross-reactive probes (McCartney et al. 2016), probes containing SNPs at the CpG site, probes detecting SNP information, probes detecting methylation at non-CpG sites, and probes targeting sites within the sex

chromosomes were removed using the minfi package in R (Aryee et al. 2014; Fortin et al. 2016) (Figure 34).

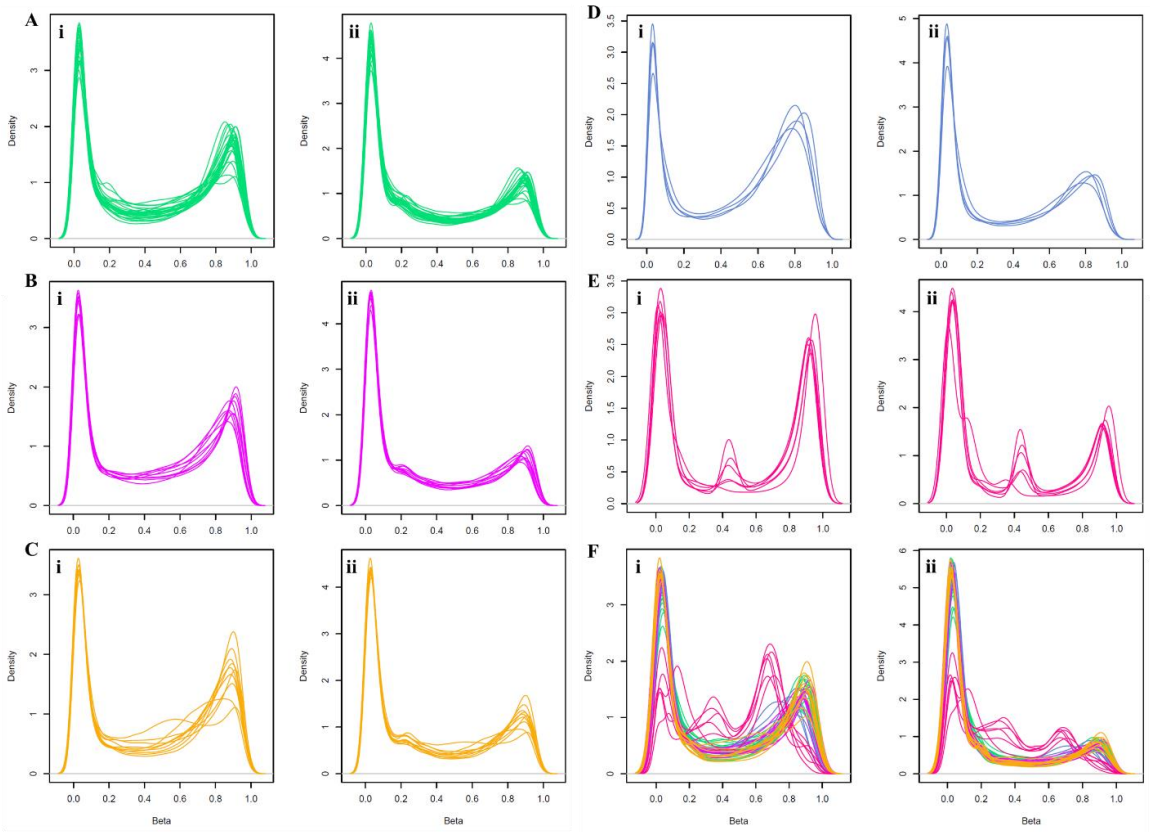

Figure 34. Normalized and Filtered Methylation Data for EPIC Array Nonhuman Primate Intra- and Inter-Specific Study.

Density plots of β-values after normalization and probe filtering using the alignment criteria (i) or gene symbol criteria (ii) for baboons (A), macaques (B), vervets (C), chimpanzees (D), marmosets (E), and all the combination of all taxa together (F), respectively.

Differential Methylation Analyses

Because β values have high heteroscedasticity, they are not statistically valid for use in differential methylation analyses (Du et al. 2010). Thus, M values were calculated and used in these analyses instead (Equation 2).

In order to identify sites that were significantly differentially methylated across comparative groups, I designed and tested generalized linear mixed models (GLMMs) which related the variables of interest to the DNA methylation patterns for each site, while accounting for the effects of additional variables, batch effects, and latent variables (Maksimovic et al. 2016). Sites found to have significant associations were classified as significantly differentially methylated positions (DMPs). Specifically, a GLMM was used to estimate differences in methylation levels associated with the femur morphology within each taxonomic group (intra-specific) and between each taxonomic group (inter-specific).

Intra-Specific Analyses

For the intra-specific analyses, variables included in each GLMM, in addition to the femur morphologies within each taxonomic group, were sex, age (years), and steady state weight (kg) when available, as well as unknown latent variables calculated using the iteratively re-weighted least squares approach in the sva package in R (Jaffe and Irizarry 2014; Jeffrey T. Leek et al. 2012; J. T. Leek and Storey 2008; Jeffrey T. Leek and Storey 2007). Variable numbers of latent variables estimated for each morphology were included to help mitigate any unknown batch and cell heterogeneity effects on methylation variation at each site (Table 14).

Table 14. Samples and Latent Variables Included in Nonhuman Primate Intra-Specific Morphology Study.

Measurement	Baboon		Macaque		Vervet		Chimpanzee		Marmoset	
	Samples	Latent Variables	Samples	Latent Variables	Samples	Latent Variables	Samples	Latent Variables	Samples	Latent Variables
femur length	21	3	10	1	10	2	4	0	5	1
bicondylar femur length	21	3	10	1	10	2	4	1	6	1
maximum femur length	21	3	10	1	10	2	4	1	6	1
superior shaft width	26	5	10	1	10	2	4	1	6	0
superior shaft depth	26	5	10	1	10	2	4	0	6	0
middle shaft width	21	3	10	1	10	2	4	0	6	1
middle shaft depth	21	3	10	1	10	2	4	0	6	0
inferior shaft width	28	4	10	1	10	2	4	0	6	0
inferior shaft depth	28	5	10	1	10	2	4	0	6	1
head height	26	4	10	1	10	2	4	1	6	0
head length	26	3	10	1	10	2	4	0	6	1
head width	26	3	10	1	10	2	4	0	6	0
anatomical neck length	26	4	10	1	10	2	4	1	6	0
anatomical neck height	26	4	10	1	10	2	4	0	6	1
anatomical neck depth	26	4	10	1	10	2	4	0	6	1
biomechanical neck length	25	5	10	1	10	2	4	0	6	1
proximal width	25	5	10	2	10	2	4	0	6	0
lesser trochanter to head	26	4	10	1	10	2	4	1	6	1
lesser trochanter to neck	26	4	10	1	10	2	4	1	6	1
lesser to greater trochanter	26	3	10	1	10	2	4	1	5	1
medial condyle height	28	4	10	1	10	2	4	1	6	0
medial condyle depth	26	4	10	1	10	2	4	1	5	1
medial condyle width	28	4	10	1	10	2	4	1	6	1
lateral condyle height	27	5	10	2	10	2	4	1	6	0
lateral condyle depth	28	5	10	1	10	2	4	0	5	1
lateral condyle width	28	5	10	1	10	2	4	0	6	0
intercondylar notch width	28	4	10	1	10	2	4	1	6	1
intercondylar notch depth	28	5	10	1	10	2	4	1	6	0
bicondylar width	28	4	10	1	10	2	4	0	6	1

Table showing the number of samples and latent variables included in each GLMM testing for associations between DNA methylation levels and femur morphology within each species.

Alternative methods to account for cell heterogeneity exist, but they are specific to whole blood (Jaffe and Irizarry 2014; Morris and Beck 2015), require reference epigenetic data, or are reference free methods (Houseman et al. 2014) that are comparable to the sva method (Kaushal et al. 2015). Out of the known cell types in skeletal tissues (Horvath, Mah, et al. 2015), only chondrocytes and osteoblasts have reference epigenomes available on the International Human Epigenomics Consortium, and these are only for humans, not nonhuman primates. Thus, because no standard method is available to correct for the heterogeneous cell structure in nonhuman primate skeletal tissue, I chose the described sva method.

Each GLMM design matrix (*Equation 10* for baboons, *Equation 11* for macaques, vervets, and marmosets when $n > 5$, and *Equation 12* for chimpanzees and marmosets when $n \leq 5$) was fit to corresponding M value array data by generalized least squares using the limma package in R (Ritchie et al. 2015; Phipson et al. 2016; Huber et al. 2015), and the estimated coefficients and standard errors for each morphology were computed. Lastly, for each coefficient, an empirical Bayes approach (McCarthy and Smyth 2009; Lönnstedt and Speed 2002; Phipson et al. 2016; Smyth 2004) was used to compute moderated t-statistics, log-odds ratios of differential methylation, and associated p-values adjusted for multiple testing (Benjamini and Hochberg 1995). Significant DMPs for the effect of each morphology were defined as those having log fold changes in M values corresponding to an adjusted p-value of less than 0.05.

Equation 10: methylation ~ femur morphology + sex + age + weight + latent variables
Equation 11: methylation ~ femur morphology + sex + age + latent variables

Equation 12: methylation ~ femur morphology + latent variables

Lastly, the gene ontology and KEGG pathway enrichment for significant CpGs while taking into account the differing number of probes per gene present on the array was determined using the missMethyl package in R (Geeleher et al. 2013; Young et al. 2010; Ritchie et al. 2015; Benjamini and Hochberg 1995).

Inter-Specific Analyses

For the inter-specific analyses, variables in addition to taxonomic grouping included in the GLMM were sex, age (years), known batch effects (e.g., array number and position), and unknown latent variables calculated using the iteratively re-weighted least squares approach in the sva package in R (Jaffe and Irizarry 2014; Jeffrey T. Leek et al. 2012; J. T. Leek and Storey 2008; Jeffrey T. Leek and Storey 2007). The four latent variables estimated for each morphology were included to help mitigate any unknown batch and cell heterogeneity effects on methylation variation at each site. As in the intra-specific analyses, I used sva to correct for heterogeneous cell structure.

The GLMM design matrix (*Equation 13*) was fit to the M value array data by generalized least squares using the limma package in R (Ritchie et al. 2015; Phipson et al. 2016; Huber et al. 2015), and the estimated coefficients and standard errors for taxonomic group effects were computed. Lastly, for each coefficient, an empirical Bayes approach (McCarthy and Smyth 2009; Lönnstedt and Speed 2002; Phipson et al. 2016; Smyth 2004) was used to compute moderated t-statistics, log-odds ratios of differential methylation, and associated p-values adjusted for multiple testing (Benjamini and

Hochberg 1995). Significant DMPs for the effect of taxonomy were defined as those having log fold changes in M values corresponding to an adjusted p-value of less than 0.05.

Equation 13: methylation ~ taxonomic group + sex + age + batch effects + latent variables

To determine only those methylation differences that represent fixed changes between genera, I used methods similar to those described in (Hernando-Herraez et al. 2013). Briefly, I identified significant DMPs between all possible pairwise comparisons between taxa as described (n=10: baboon-macaque, baboon-vervet, baboon-chimpanzee, baboon-marmoset, macaque-vervet, macaque-chimpanzee, macaque-marmoset, vervet-chimpanzee, vervet-marmoset, chimpanzee-marmoset). I then defined a significant DMP as taxon-specific if it was found to be significant in all 4 pairwise comparisons containing the taxon of interest but not found in any of the remaining pairwise comparisons. The gene ontology and KEGG pathway enrichment of these DMPs was then determined as described above.

Additionally, global changes in methylation were calculated using distance matrices, as described in (Hernando-Herraez et al. 2013), of the methylation levels for all finalized 39,802 filtered probes. For species-level changes, I averaged the β values per probe within a species and then used Euclidean distances to calculate the difference between every two species. For individual-level changes, I used Euclidean distances to calculate the difference between every two individuals. I used the ape package in R (Paradis et al. 2004) to estimate a neighbor joining tree using these distance matrices. For

each resulting tree, 1000 bootstraps were performed to determine confidence values for each branch.

Gene-Specific DNA Methylation Profiling and Analyses

Based on the inter-specific DNA methylation patterns identified in this study and those identified in other evolutionary anthropological studies (Gokhman et al. 2014), the *HOXD10* gene was selected for subsequent DNA methylation profiling and analysis at a higher resolution using gene-specific sequencing techniques. Specifically, primers were designed and optimized to PCR amplify regions spanning across the entire *HOXD10* gene, as well as upstream and downstream several hundred bases (hg19 chr2:176980532-176985117), in each nonhuman primate species for regular and bisulfite treated DNA (APPENDIX EE, APPENDIX FF).

These gene-specific assays were performed in a subset of the samples tested using the EPIC array and included chimpanzees (n=3), baboons (n=3), macaques (n=3), vervets (n=3), and marmosets (n=3) (APPENDIX BB). As described above, DNA was extracted from femoral trabecular bone using a phenol-chloroform protocol optimized for skeletal tissues (Barnett and Larson 2012). Bisulfite treated DNA was bisulfite converted using the EZ DNA Methylation^{TM} Gold Kit according to the manufacturer's instructions (Zymo Research). Successful PCR amplification was confirmed using standard gel electrophoresis. Gene-specific PCR products were then purified using a standard exonuclease I and shrimp alkaline phosphatase protocol and sequenced on the Applied Biosystems 3730 capillary sequencer at the DNA Laboratory at Arizona State University.

Regular and bisulfite sequences were aligned to the appropriate nonhuman primate references within the Enredo-Pecan-Orthus (EPO) whole-genome multiple alignments of several primate genomes [Ensembl Compara.8_primates_EPO] (Paten, Herrero, Beal, et al. 2008; Paten, Herrero, Fitzgerald, et al. 2008) using MEGA7 (Kumar et al. 2016) and Geneious version 9.1.2 (Kearse et al. 2012). Manual annotation of these sequences within each sample confirmed that the gene sequences belong to the appropriate primate species and that the regular and bisulfite treated sequences only differ in cytosine composition. The number and distribution of methylated loci throughout the *HOXD10* gene were then identified. These were compared to the sample specific findings from the EPIC array in order to validate the genome-wide methylation results. Lastly, methylated loci were compared within and among species to provide a higher resolution of methylation variation within this targeted gene.

Results

The aim of this study was to use the EPIC array to identify DNA methylation patterns in femoral bone from five nonhuman primate species – 28 baboons, 10 macaques, 10 vervets, 4 chimpanzees, and 6 marmosets – in order to determine how methylation varies intra-specifically with respect to femur morphology and how it varies inter-specifically. In order to do this, I first assessed the effectiveness of the EPIC array in identifying DNA methylation patterns in baboon DNA and of different probe filtering methods.

Alignment of EPIC Array Probes with Nonhuman Primate Genomes

Probes from the EPIC array were aligned to the baboon, macaque, vervet, chimpanzee, and marmoset genomes using methods modified from (Hernando-Herraez et al. 2013; Ong et al. 2014) (APPENDIX A). Out of the 866,837 50bp probes on the array, 39% map to the baboon genome, 39% map to the macaque genome, 39% map to the vervet genome, 76% map to the chimpanzee genome, and 17% map to the marmoset genome with e-values less than e^{-10}, with only unique BLAST hits, and that target a CpG site (Table 15). Out of these reliably mapped probes, 24% in baboons, 24% in macaques, 24% in vervets, 71% in chimpanzees, and 9% in marmosets were retained after the alignment filter criteria (Figure 35, Table 15). Conversely, 22% in baboons, 19% in macaques, 23% in vervets, 32% in chimpanzees, and 10% in marmosets were retained after the gene symbol filter criteria (Figure 35, Table 15). Overall, 36,248 probes are shared among all taxa for the alignment filter criteria, while 36,248 probes are shared among all taxa for the gene symbol filter criteria.

Table 15. Number of EPIC Array Probes Retained for Nonhuman Primates.

	Total Mapped Probes	Alignment Filter Probes	Gene Symbol Filter Probes
Baboon	337,818	209,802	190,703
Macaque	335,046	207,703	164,754
Vervet	336,786	207,650	195,555
Chimpanzee	657,913	622,819	273,306
Marmoset	143,407	74,599	85,770

Details on the numbers of probes that successfully mapped to each nonhuman primate genome with e-values less than e-10, had only unique BLAST hits, and targeted a CpG

site (Total Mapped Probes), probes that fit the alignment filter criteria (Alignment Filter Probes), and probes that fit the gene symbol filter criteria (Gene Symbol Filter Probes).

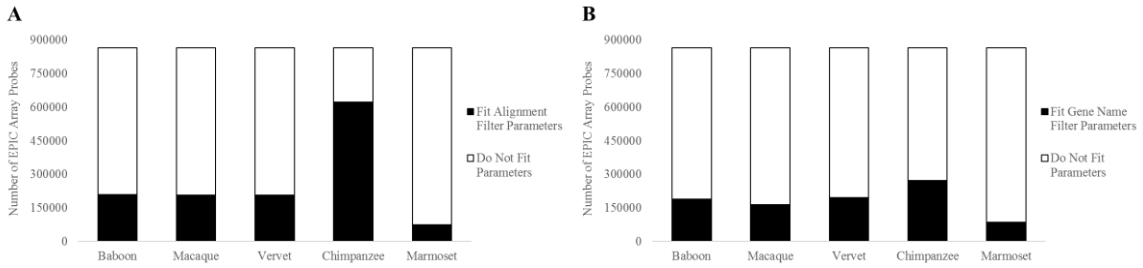

Figure 35. Filtering Effects on EPIC Array Probes for Nonhuman Primates.

(A) Bar chart showing the percent of EPIC array probes that map to the baboon (*Papio anubis*), macaque (*Macaca mulatta*), vervet (*Chlorocebus aethiops*), chimpanzee (*Pan troglodytes*), or marmoset (*Callithrix jacchus*) genomes with e-values < e^{-10}, have only unique BLAST hits, and target a CpG site, as well as contain 0 mismatches in 5bp of the probe by and including the targeted CpG site and 0-2 mismatches in 45bp of the probe not including the CpG site. Out of 866,837 probes total, 209,802 (24%) meet these criteria for baboons, 207,703 (24%) for macaques, 207,650 (24%) for vervets, 622,819 probes (72%) for chimpanzees, and 74,599 (9%) for marmosets. (B) Bar chart showing the percent of EPIC array probes that map to each nonhuman primate genome with e-values < e^{-10}, have only unique BLAST hits, and target a CpG site, as well as are found in a gene that matches the human gene. Out of 866,837 probes total, 190,703 (22%) meet these criteria for baboons, 164,754 (19%) for macaques, 195,555 (23%) for vervets, 273,306 probes (32%) for chimpanzees, and 85,770 (10%) for marmosets.

When comparing how probes overlap between the alignment filter criteria and the gene symbol criteria for each species, 121,308 probes overlapped between both filtering methods in baboons (58% and 64% respectively), 104,616 probes overlapped in macaques (50% and 63% respectively), 123,500 probes overlapped in vervets (59% and 63% respectively), 260,263 probes overlapped in chimpanzees (42% and 95% respectively), 44,989 probes overlapped in marmosets (60% and 52% respectively), and 16,916 probes overlapped in the combination of all taxa (40% and 47% respectively) (Figure 36).

Probes that reliably mapped to each nonhuman primate genome, that met the alignment filter criteria, or that met the gene symbol criteria covered 9,779 to 24,279 genes with average coverages ranging from 4 to 20 probes per gene (APPENDIX GG). Additionally, the retained probes covered a range of locations with respect to genes and CpG islands (APPENDIX GG), indicating that these filtered probes maintain a wide distribution throughout the genome.

After filtering out cross-reactive probes (Y. Chen et al. 2013), probes containing SNPs at the CpG site, probes detecting SNP information, probes detecting methylation at non-CpG sites, and probes targeting sites within the sex chromosomes a final set of 189,858 probes were retained for the alignment filter criteria in baboons, 190,898 probes in macaques, 191,639 probes in vervets, 576,804 probes in chimpanzees, 68,709 probes in marmosets, and 39,802 probes shared among species. Conversely, a final set of 165,529 probes were retained for the gene symbol filter criteria in baboons, 146,585 probes in macaques, 175,592 probes in vervets, 254,231 probes in chimpanzees, 75,002 probes in marmosets, and 33,254 probes shared among species (Figure 34).

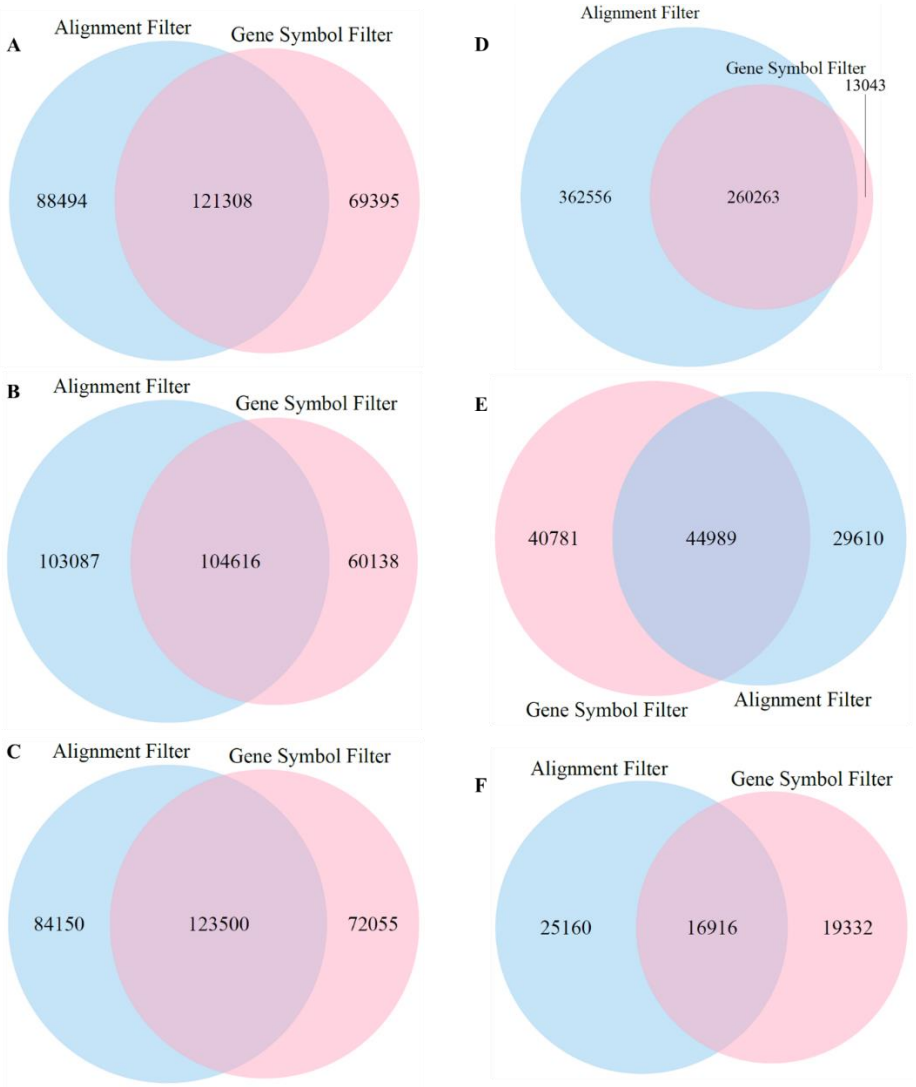

Figure 36. Overlap of 450K Array Probes for Nonhuman Primates Using Different Filtering Methods.

Venn diagrams showing the number of probes that overlap between the alignment filter criteria and the gene symbol criteria for each species. (A) For baboons, out of the 209,802 probes that meet the alignment filter criteria and the 190,703 probes that meet the gene symbol criteria, 121,308 probes (58% and 64% respectively) overlap in both filters. (B) For macaques, out of the 207,703 probes that meet the alignment filter criteria and the 164,754 probes that meet the gene symbol criteria, 104,616 probes (50% and 63%

respectively) overlap in both filters. (C) For vervets, out of the 207,650 probes that meet the alignment filter criteria and the 195,555 probes that meet the gene symbol criteria, 123,500 probes (59% and 63% respectively) overlap in both filters. (D) For chimpanzees, out of the 622,819 probes that meet the alignment filter criteria and the 273,306 probes that meet the gene symbol criteria, 260,263 probes (42% and 95% respectively) overlap in both filters. (E) For marmosets, out of the 74,599 probes that meet the alignment filter criteria and the 85,770 probes that meet the gene symbol criteria, 44,989 probes (60% and 52% respectively) overlap in both filters. (F) For probes that align to all nonhuman primate genomes, out of the 42,076 probes that meet the alignment filter criteria and the 36,248 probes that meet the gene symbol criteria, 16,916 probes (40% and 47% respectively) overlap in both filters.

Effectiveness of EPIC Array Probes using Nonhuman Primate DNA

To determine how effectively the EPIC array probes measured DNA methylation in baboon DNA, I performed Spearman correlation tests between the hybridization efficiency of each probe and parameters defining the alignment quality of each probe to each nonhuman primate genome. Specifically, both probe alignment bitscores and percent identity were significantly negatively correlated with probe hybridization efficiency, and probe alignment e-values were significantly positively correlated with probe hybridization efficiency, regardless of filtering criteria (APPENDIX HH). However, filtering probes using the alignment filter criteria retained proportionally more successfully hybridized probes than filtering probes using the gene symbol filter criteria (Figure 37). Thus, filtering probes using the alignment filter criteria likely produces more

reliable results, so only these data were used for downstream differential methylation analyses.

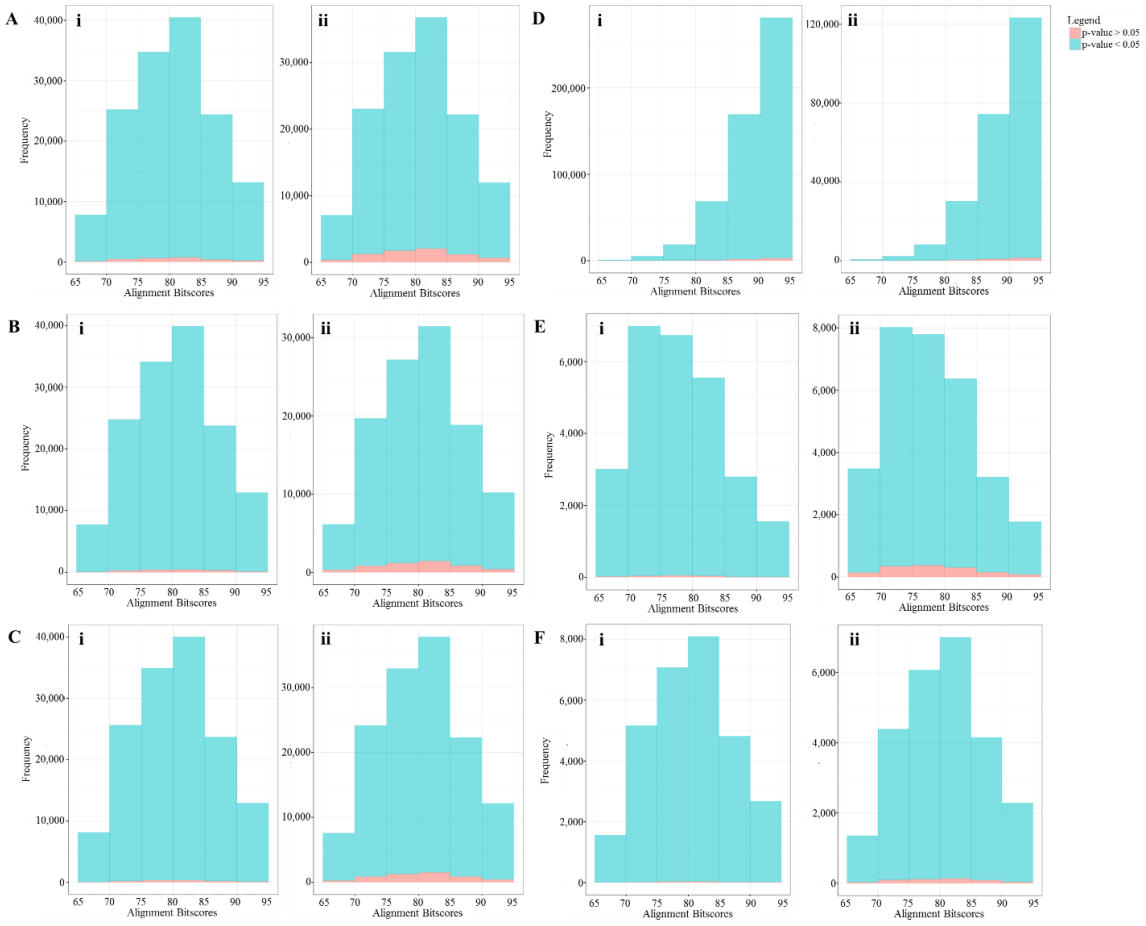

Figure 37. Hybridization Efficiencies of EPIC Array Probes Retained for Nonhuman Primate Intra- and Inter-Specific Study.

Histogram of alignment bitscores for EPIC array probes with detection p-values > 0.05 (red) and < 0.05 (blue). These p-values were averaged across all samples within each species, and probes included meet (i) the alignment filter criteria or (ii) the gene symbol filter criteria. (A) Baboons: For probes meeting the alignment filter criteria (i), 3,880 had detection p-values > 0.05, and 205,922 had detection p-values < 0.05. For probes meeting

the gene symbol filter criteria (ii), 10,571 had detection p-values > 0.05, and 180,132 had detection p-values < 0.05. For all probes that successfully mapped to the baboon genome with e-values < e^{-10}, had only unique BLAST hits, and targeted a CpG site, 21,977 had detection p-values > 0.05, and 315,841 had detection p-values < 0.05. (B) Macaques: For probes meeting the alignment filter criteria (i), 2,586 had detection p-values > 0.05, and 205,117 had detection p-values < 0.05. For probes meeting the gene symbol filter criteria (ii), 7,442 had detection p-values > 0.05, and 157,312 had detection p-values < 0.05. For all probes that successfully mapped to the baboon genome with e-values < e^{-10}, had only unique BLAST hits, and targeted a CpG site, 17,821 had detection p-values > 0.05, and 317,225 had detection p-values < 0.05. (C) Vervets: For probes meeting the alignment filter criteria (i) 2,007 had detection p-values > 0.05, and 205,643 had detection p-values < 0.05. For probes meeting the gene symbol filter criteria (ii), 7,732 had detection p-values > 0.05, and 187,823 had detection p-values < 0.05. For all probes that successfully mapped to the baboon genome with e-values < e^{-10}, had only unique BLAST hits, and targeted a CpG site, 15,405 had detection p-values > 0.05, and 321,381 had detection p-values < 0.05. (D) Chimpanzees: For probes meeting the alignment filter criteria (i), 6,120 had detection p-values > 0.05, and 616,699 had detection p-values < 0.05. For probes meeting the gene symbol filter criteria (ii), 3,241 had detection p-values > 0.05, and 270,065 had detection p-values < 0.05. For all probes that successfully mapped to the baboon genome with e-values < e^{-10}, had only unique BLAST hits, and targeted a CpG site, 9,982 had detection p-values > 0.05, and 647,931 had detection p-values < 0.05. (E) Marmosets: For probes meeting the alignment filter criteria (i), 595 had detection p-values > 0.05, and 74,004 had detection p-values < 0.05. For probes meeting the gene

symbol filter criteria (ii), 3,993 had detection p-values > 0.05, and 81,777 had detection p-values < 0.05. For all probes that successfully mapped to the baboon genome with e-values < e^{-10}, had only unique BLAST hits, and targeted a CpG site, 7,481 had detection p-values > 0.05, and 135,926 had detection p-values < 0.05. (F) All Nonhuman Primate Species Combined: For probes meeting the alignment filter criteria (i), 201 had detection p-values > 0.05, and 41,875 had detection p-values < 0.05. For probes meeting the gene symbol filter criteria (ii), 770 had detection p-values > 0.05, and 35,478 had detection p-values < 0.05.

Differential Methylation and Intra-Specific Morphological Variation

Measurements of 29 linear morphology traits (Figure 33, Table 12) were collected from each nonhuman primate right femur (Figure 38), and all of these measurements had less than 5% error, except those for intercondylar notch depth in macaques (APPENDIX CC). Significant DMPs associated with each intra-specific linear morphology were interrogated from 189,858 sites in in baboons, 190,898 sites in macaques, 191,639 sites in vervets, 576,804 sites in chimpanzees, and 68,709 sites in marmosets (Table 16).

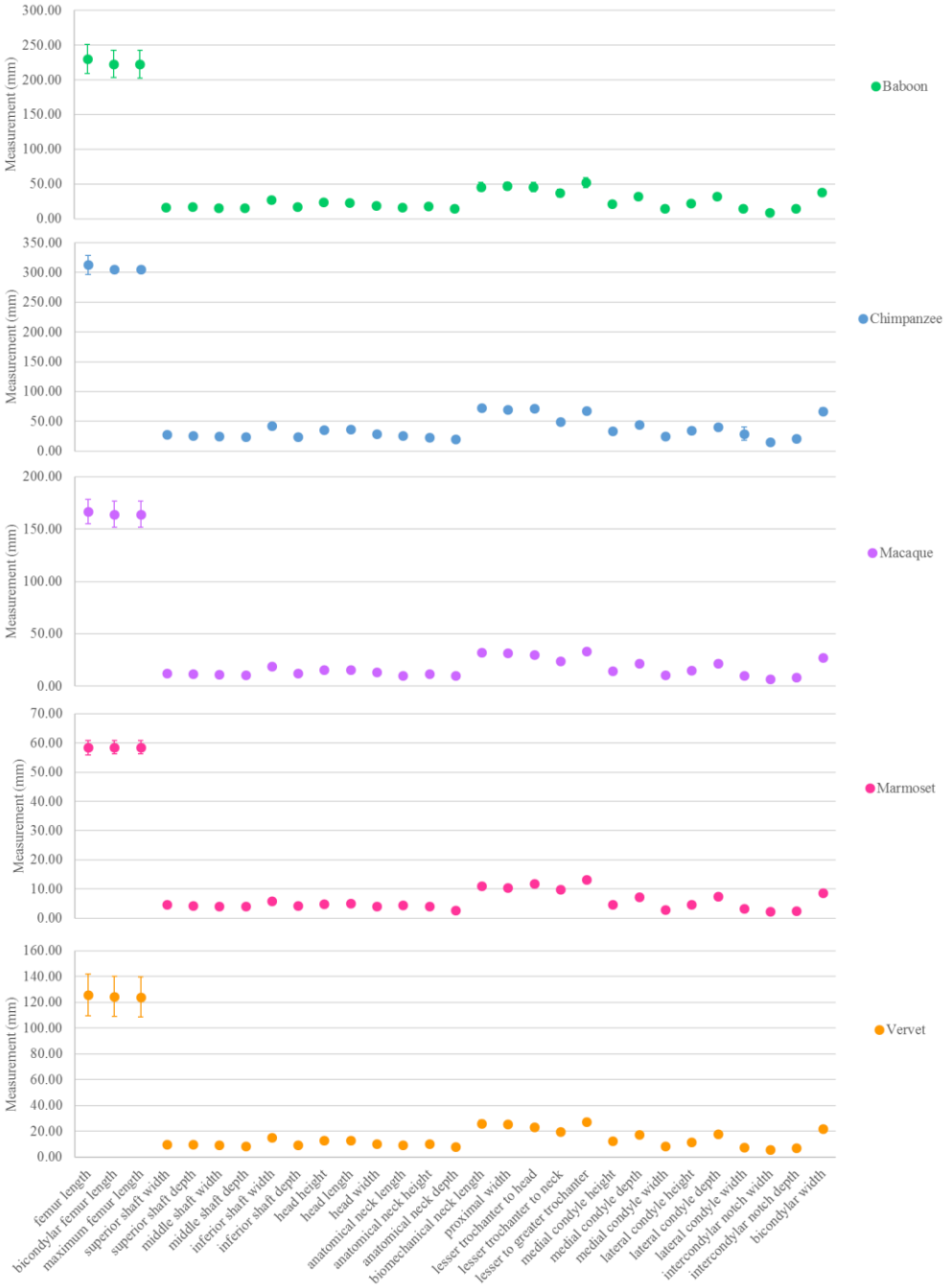

Figure 38. Results of Nonhuman Primate Morphological Measurements.

Plot of linear morphological measurements in each species. Plot depicts the average measurement in millimeters with error bars displaying one standard deviation in each direction. See Appendix CC for additional details.

Table 16. Number of Significant DMPs Identified in the EPIC Array Nonhuman Primate Intra-Specific Study.

Measurement	Significant DMPs				
	Baboon	Macaque	Vervet	Chimpanzee	Marmoset
femur length	0	0	0	0	0
bicondylar femur length	1	0	0	0	0
maximum femur length	1	0	0	0	0
superior shaft width	0	0	1	0	0
superior shaft depth	0	0	0	0	0
middle shaft width	0	0	0	0	0
middle shaft depth	0	0	0	0	0
inferior shaft width	0	0	2	0	0
inferior shaft depth	0	0	0	0	0
head height	0	0	0	0	0
head length	0	0	0	0	0
head width	0	0	0	0	0
anatomical neck length	0	0	0	273	0
anatomical neck height	0	0	1	0	0
anatomical neck depth	0	0	0	0	0
biomechanical neck length	0	0	0	0	0
proximal width	0	1	0	0	0
lesser trochanter to head	0	0	0	0	0
lesser trochanter to neck	0	0	0	0	0
lesser to greater trochanter	0	0	0	0	0
medial condyle height	0	0	0	0	0
medial condyle depth	0	0	0	0	0
medial condyle width	0	7	0	0	0
lateral condyle height	0	0	0	0	0
lateral condyle depth	0	0	0	0	0
lateral condyle width	0	0	0	0	0
intercondylar notch width	0	0	0	0	0
intercondylar notch depth	0	0	0	0	0
bicondylar width	0	0	0	0	0

Table showing the number of significant DMPs associated with each linear morphology in each species. Results are shown for probes filtered using the alignment criteria. The number of total sites tested for baboons was 189,858, for macaques was 190,898, for

vervets was 191,639, for chimpanzees was 576,804, and for marmosets was 68,709. The number of samples and latent variables included in each GLMM are provided in *Table 14*.

Specifically, in baboons, 1 site was found to be hypomethylated with increasing bicondylar femur length and increasing maximum femur length. In macaques, 1 site was found to be hypermethylated with increasing proximal femur width, 1 site was found to be hypermethylated with increasing medial condyle width, and 6 sites were found to be hypomethylated with increasing medial condyle width. In vervets, 1 site was found to be hypomethylated with increasing superior shaft width, 2 sites were found to be hypomethylated with increasing inferior shaft width, and 1 site was found to be hypermethylated with increasing anatomical neck height. In chimpanzees, 216 sites were found to be hypomethylated and 57 sites were found to be hypermethylated with increasing anatomical neck length. Lastly, in marmosets, no sites were found to have differential methylation with variation in morphology (APPENDIX II, APPENDIX JJ). While the maximum change in β values (Δβ) for most of these DMPs is greater than 0.1, the actual Δβ between individuals with the largest morphology measurements and individuals with the smallest morphology measurements is less than 0.1 for several DMPs (APPENDIX JJ).

Tests for enrichment of gene ontology (GO) and KEGG pathway functions were done for those intra-specific morphologies that had more than 2 significant DMPs associated with them. However, no GO biological processes were found to be enriched in DMPs associated with macaque medial condyle width or chimpanzee anatomical neck

length. Additionally, KEGG pathway functions were only found to be enriched in DMPs associated with chimpanzee anatomical neck length (Table 17).

Table 17. KEGG Pathways Enriched for Significant DMPs Associated with Anatomical Neck Lengths in Chimpanzees.

KEGG ID	KEGG Pathway	No. Genes Total	No. Genes with DMPs	P-Value	FDR
path:hsa04662	B cell receptor signaling pathway	65	3	0.0000922857	0.020544140
path:hsa04658	Th1 and Th2 cell differentiation	81	3	0.0001905770	0.020544140
path:hsa04650	Natural killer cell mediated cytotoxicity	92	3	0.0002224360	0.020544140
path:hsa04660	T cell receptor signaling pathway	97	3	0.0002617090	0.020544140
path:hsa04140	Autophagy - animal	119	3	0.0005084010	0.031927559
path:hsa01100	Metabolic pathways	1,111	6	0.0009562600	0.047626289
path:hsa04310	Wnt signaling pathway	133	3	0.0010617330	0.047626289

KEGG pathways that are significantly enriched (FDR < 0.05) for significant DMPs that are associated with anatomical neck lengths in chimpanzees, taking into account the differing number of probes per gene present on the EPIC array. Table include the identification numbers (KEGG ID) and pathways (KEGG Pathway) for each significantly enriched KEGG term, the total number of genes associated with each KEGG pathway (No. Genes Total), the number of genes with significant DMPs that are also associated with each KEGG pathway (No. Gene with DMPs), the p-value for over-representation of each KEGG pathway (P-Value), and the false discovery Rate for each KEGG pathway (FDR).

Differential Methylation and Inter-Specific Variation

In order to determine how methylation varies inter-specifically, significant DMPs were interrogated from 39,802 probes that were filtered using the alignment criteria and

shared among nonhuman primate species (Table 18, APPENDIX KK). As described above, species-specific DMPs were determined by first identifying significant DMPs between all possible pairwise comparisons of nonhuman primate species and isolating only those DMPs that were significant in all 4 pairwise comparisons containing the taxon of interest but not in any of the remaining pairwise comparisons (Hernando-Herraez et al. 2013). These methods identified 650 (1.63%) species-specific DMPs in baboons, 257 (0.65%) in macaques, 639 (1.61%) in vervets, 2,796 (7.02%) in chimpanzees, and 13,778 (34.62%) in marmosets (Table 18), which spanned 7,320 genes with an average number of 2 probes per gene (APPENDIX LL). Of these genes, 2,496 have at least 2 probes per gene. Additionally, these species-specific DMPs covered a range of locations with respect to genes and CpG islands (APPENDIX LL), indicating that these species-specific changes in methylation have a decent distribution throughout the genome.

Using various inter-specific $\Delta\beta$ cutoff thresholds decreases the final number of species-specific DMPs to varying degrees (Table 18, APPENDIX KK). For instance, using a $\Delta\beta \geq 0.1$ threshold decreased the final number of species-specific DMPs to 581 (1.46%) in baboons, 230 (0.58%) in macaques, 530 (1.33%) in vervets, 1572 (3.95%) in chimpanzees, and 12,394 (31.14%) in marmosets (Table 18), which spanned 6,673 genes with an average number of 2 probes per gene (APPENDIX LL). Of these genes, 2,276 have at least 2 probes per gene. Conversely, using a $\Delta\beta \geq 0.4$ threshold decreased the final number of species-specific DMPs even further to 15 (0.04%) in baboons, 23 (0.06%) in macaques, 54 (0.14%) in vervets, 89 (0.22%) in chimpanzees, and 473 (1.19%) in marmosets (Table 18), which spanned 489 genes with an average number of 1 probe per gene (APPENDIX LL). Of these genes, 33 have at least 2 probes per gene.

Counter to these decreases in numbers, though, species-specific DMPs cover a range of locations with respect to genes and CpG islands regardless of the Δβ cutoff threshold (APPENDIX LL).

Overall, across Δβ cutoff thresholds, more species-specific DMPs were found in the New World monkey marmosets, followed by the great ape chimpanzees, and lastly the Old World monkey baboons, macaques, and vervets. Additionally, the proportions of hypermethylated and hypomethylated species-specific DMPs within each taxon remain fairly constant across Δβ cutoff thresholds (Table 18). In baboons, macaques, vervets, and chimpanzees, more than half of all species-specific DMPs show patterns of hypermethylation. This pattern is only affected when no Δβ threshold is applied to the species-specific DMPs in chimpanzees. Lastly, in marmosets, more than half of all species-specific DMPs show patterns of hypomethylation regardless of the Δβ cutoff threshold. Nevertheless, the ability of this variation in methylation patterns at species-specific DMPs to cluster animals into taxonomic groups varies across different Δβ cutoff thresholds (Figure 39, Figure 40). For all thresholds, apes, Old World monkeys, and New World monkeys cluster into distinct groups. However, within Old World monkeys, vervets only cluster into a distinct species group when a $\Delta\beta \geq 0.2$ threshold is applied, and baboons and macaques only form a distinct clade when a $\Delta\beta \geq 0.3$ threshold is applied (Figure 39). Species-specific clustering of baboons, macaques, and vervets only occurs when a $\Delta\beta \geq 0.4$ threshold is applied (Figure 40).

Table 18. Number of Significant Species-Specific DMPs Identified in the EPIC Array Nonhuman Primate Inter-Specific Study.

	Baboon	Macaque	Vervet	Chimpanzee	Marmoset
Total Species-Specific DMPs Identified					
Significant DMPs	650	257	639	2,796	13,778
Hypermethylated	*282*	*76*	*250*	*1,370*	*10,335*
Hypomethylated	*314*	*146*	*327*	*1,210*	*3,342*
Mixture	*54*	*35*	*62*	*216*	*101*
Species-Specific DMPs with Average $\Delta\beta \geq 0.1$					
Significant DMPs	581	230	530	1,572	12,394
Hypermethylated	*245*	*67*	*213*	*503*	*9,408*
Hypomethylated	*308*	*139*	*277*	*1,039*	*2,966*
Mixture	*28*	*24*	*40*	*30*	*20*
Species-Specific DMPs with Average $\Delta\beta \geq 0.2$					
Significant DMPs	284	132	301	892	6,017
Hypermethylated	*76*	*36*	*85*	*245*	*5,401*
Hypomethylated	*198*	*90*	*205*	*643*	*608*
Mixture	*10*	*6*	*11*	*4*	*8*
Species-Specific DMPs with Average $\Delta\beta \geq 0.3$					
Significant DMPs	63	55	126	315	1,417
Hypermethylated	*24*	*8*	*30*	*66*	*1,271*
Hypomethylated	*38*	*46*	*94*	*249*	*144*
Mixture	*1*	*1*	*2*	*0*	*2*
Species-Specific DMPs with Average $\Delta\beta \geq 0.4$					
Significant DMPs	15	23	54	89	473
Hypermethylated	*4*	*2*	*4*	*19*	*414*
Hypomethylated	*11*	*20*	*50*	*70*	*59*
Mixture	*0*	*1*	*0*	*0*	*0*

Details on the number of species-specific DMPs identified for each nonhuman primate species in total, when the average absolute change in β values between each pairwise species comparison (Δβ) is greater than 0.1, greater than 0.2, greater than 0.3, or greater than 0.4. The pattern of methylation (hypermethylated, hypomethylated, or a mixture of both) of each species-specific DMP is also indicated.

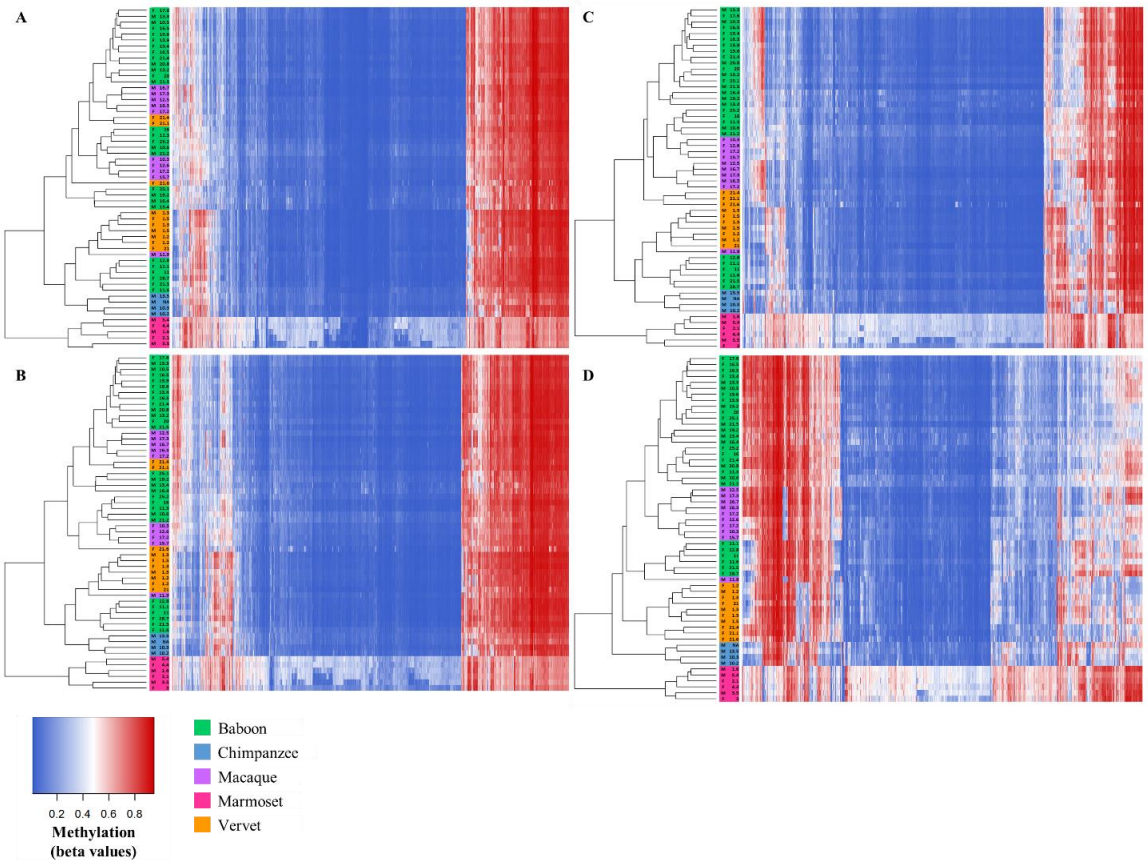

Figure 39. Methylation Levels at Species-Specific DMPs with Various Δβ Threshold Cutoffs Identified in the EPIC Array Nonhuman Primate Inter-Specific Study. Heatmap depicting (A) the DNA methylation levels (β values) of all species-specific DMPs (x-axis) in all nonhuman primate samples (n=58), (B) the DNA methylation levels (β values) of all species-specific DMPs with average absolute Δβ values greater than 0.1 between each taxonomic group (x-axis) in all nonhuman primate samples (n=58), (C) the DNA methylation levels (β values) of all species-specific DMPs with average absolute Δβ values greater than 0.2 between each taxonomic group (x-axis) in all nonhuman primate samples (n=58), and (D) the DNA methylation levels (β values) of all species-specific DMPs with average absolute Δβ values greater than 0.3 between each taxonomic group (x-axis) in all nonhuman primate samples (n=58). The sex and age of each

nonhuman primate are also provided (y-axis). Red indicates higher methylation at a DMP, while blue indicates lower methylation at a DMP. The dendrogram of all samples (y-axis) clusters individuals based on the similarity of their methylation patterns. Samples cluster into the large taxonomic groupings of New World monkeys, Old World monkeys, and apes, but Old World monkeys do not cluster by species for any of these filtering levels.

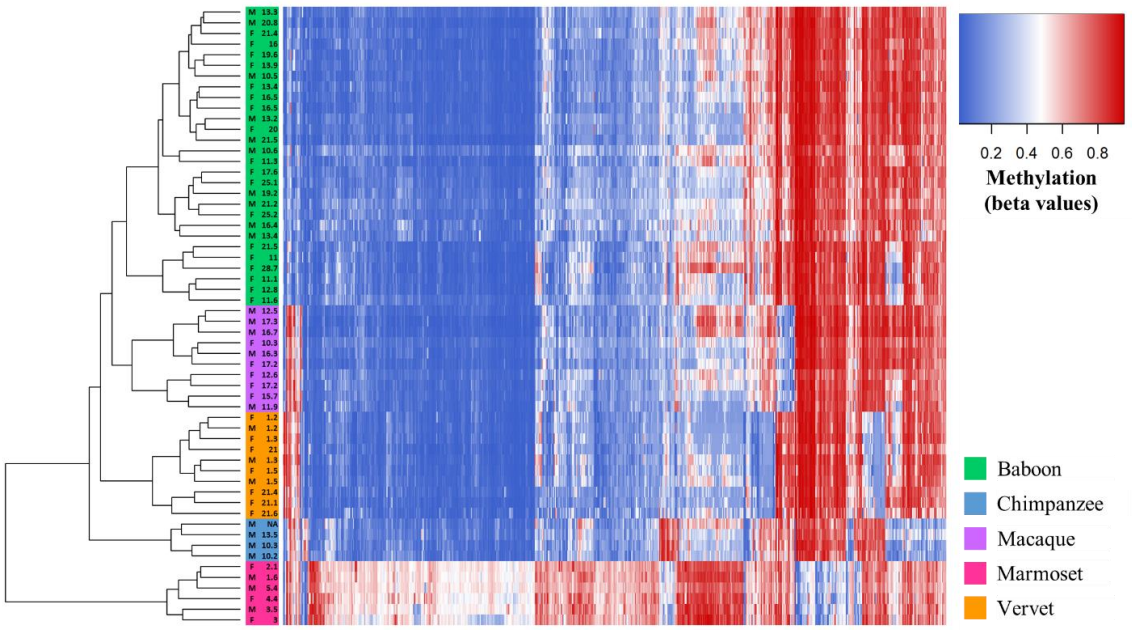

Figure 40. Methylation Levels at Species-Specific DMPs with Δβ≥0.4 Identified in the EPIC Array Nonhuman Primate Inter-Specific Study.

Heatmap depicting the DNA methylation levels (β values) of all species-specific DMPs with average absolute Δβ values greater than 0.4 between each taxonomic group (x-axis) in all nonhuman primate samples (n=58). The sex and age of each nonhuman primate are also provided (y-axis). Red indicates higher methylation at a DMP, while blue indicates lower methylation at a DMP. The dendrogram of all samples (y-axis) clusters individuals

based on the similarity of their methylation patterns. Samples cluster based on species-level taxonomic groupings and as predicted based on known species phylogenetic histories.

Global changes in methylation across all finalized 39,802 filtered probes and averaged within each species reveal that apes, Old World monkeys, and New World monkeys are phylogenetically distinct from one another, and these divergences are well supported (Figure 41). Similarly, when phylogenetic relationships are evaluated using the global changes in methylation of individual animals, apes, Old World monkeys, and New World monkeys form distinct lineages that are well supported (Figure 42). However, like the clustering of species-specific DMPs, baboons, macaques, and vervets do not form distinct lineages based on individual global methylation levels. Several of the branches within the Old World monkey clade are poorly supported, though. Phylogenetic separation of these Old World monkey species into distinct lineages is only possible when the methylation changes considered include species-specific DMPs with $\Delta\beta \geq 0.4$ (Figure 43).

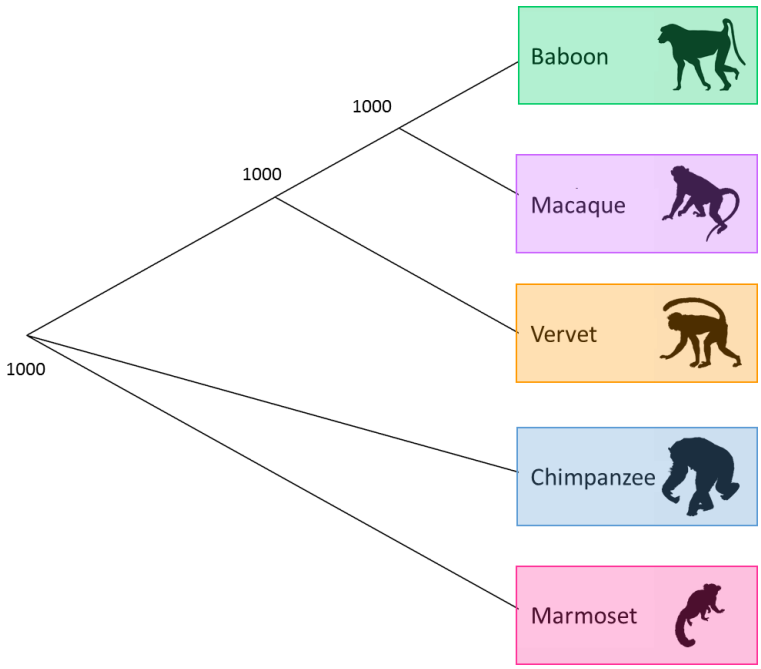

Figure 41. Phylogeny Based on Average Species-Level Global Changes in Methylation. Observed phylogenetic relationship among nonhuman primates when considering average species-level global changes in methylation. This tree was constructed using the methylation levels for all of the finalized 39,802 filtered probes. I averaged the β values per probe within a species, used Euclidean distances to calculate the difference between every two species, and estimated a neighbor joining tree using this distance matrix. For the resulting tree, 1000 bootstraps were performed to determine confidence values for each branch. The number provide at each node indicates the number of bootstrap replicates that support it out of the 1000 performed.

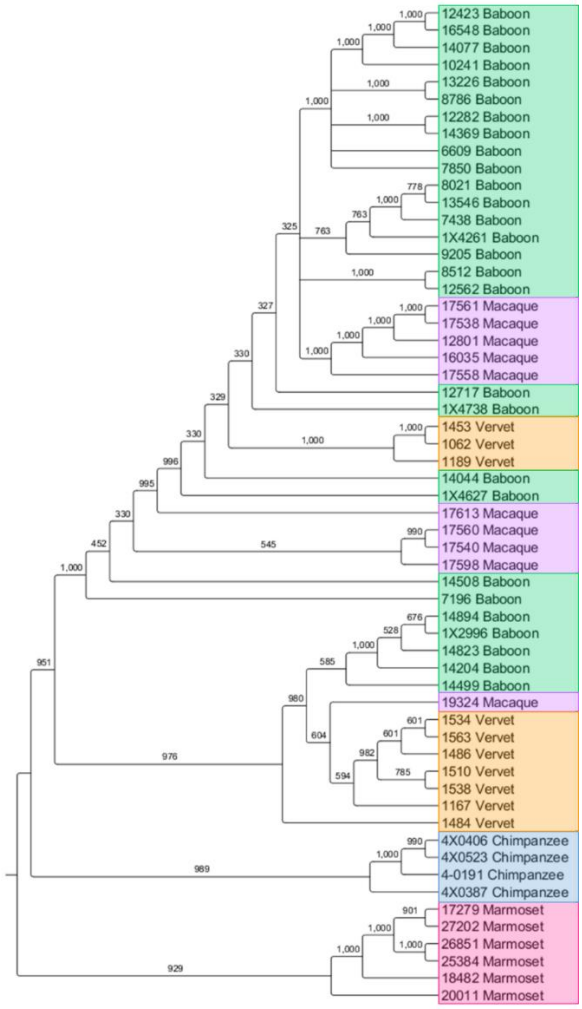

Figure 42. Phylogeny Based on Individual-Level Global Changes in Methylation. Observed phylogenetic relationship among nonhuman primates when considering individual-level global changes in methylation. This tree was constructed using the methylation levels for all of the finalized 39,802 filtered probes. I used Euclidean distances to calculate the difference between every two individuals, and estimated a neighbor joining tree using this distance matrix. For the resulting tree, 1000 bootstraps were performed to determine confidence values for each branch. The number provide at each node indicates the number of bootstrap replicates that support it out of the 1000 performed.

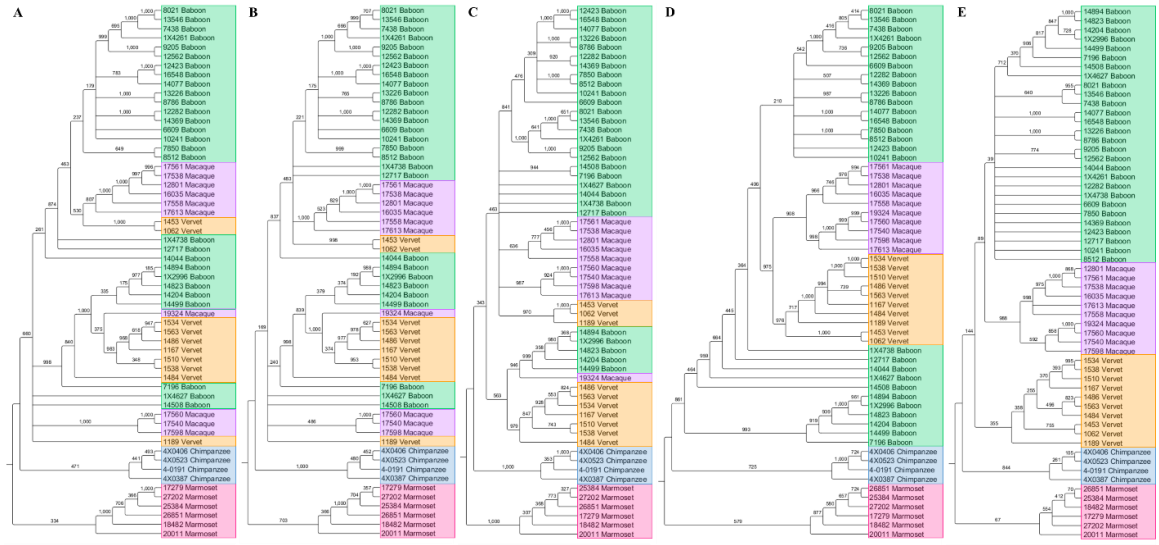

Figure 43. Phylogeny Based on Individual-Level Changes in Methylation at Species-Specific DMPs with Various Δβ Threshold Cutoffs.

Observed phylogenetic relationship among nonhuman primates when considering individual-level global changes in methylation. Trees were constructed using the methylation levels for (A) all species-specific DMPs, (B) species-specific DMPs with average absolute Δβ values greater than 0.1 between each taxonomic group, (C) species-specific DMPs with average absolute Δβ values greater than 0.2 between each taxonomic group, (D) species-specific DMPs with average absolute Δβ values greater than 0.3 between each taxonomic group, and (E) species-specific DMPs with average absolute Δβ values greater than 0.4 between each taxonomic group. I used Euclidean distances to calculate the difference between every two individuals, and estimated a neighbor joining tree using this distance matrix. For the resulting tree, 1000 bootstraps were performed to determine confidence values for each branch. The number provide at each node indicates the number of bootstrap replicates that support it out of the 1000 performed.

Species-specific DMPs also show associations with genes that have a wide array of GO biological processes (APPENDIX MM) and KEGG pathway functions (APPENDIX NN). Cellular adhesion is a primary GO function found to be highly enriched in species-specific DMPs from macaques, vervets, chimpanzees, and marmosets. Species-specific DMPs for chimpanzees and marmosets are also enriched for genes involved in the regulation of transcription and gene expression. Additionally, enrichment of genes involved in anatomical developmental processes is found in baboons, chimpanzees, and marmosets, and chimpanzees and marmosets show further enrichment of genes contributing to pattern specification processes, limb development, and skeletal system development. Moreover, marmoset species-specific DMPs are enriched for genes with functions very closely related to skeletal development, such as osteoblast differentiation and ossification, as well as genes involved in metabolism and the development of other organ systems including skeletal muscles, nerves, the brain, the heart, blood vessels, kidneys, eyes, and ears (APPENDIX MM). Several enriched pathways reinforce these molecular functions, and additional pathways related to cancers and other disease were also identified (APPENDIX NN).

Out of the species-specific DMPs identified, some were found to overlap with those previously identified as being differentially methylated among primates. These include *ARTN*, *COL2A1*, and *GABBR1* which have been found to be differentially methylated among modern humans and great apes (Hernando-Herraez et al. 2013), as well as *HOXD8*, *HOXD9*, and *HOXD10* which have been found to be differentially methylated among modern and ancient hominins (Gokhman et al. 2014). Specifically, 1 species-specific DMP in *ARTN* shows hypomethylation in marmosets, 3 species-specific

DMPs in *COL2A1* show hypermethylation in marmosets as compared to all other taxa, and while 2 species-specific DMPs in baboons and 3 species-specific DMPs in marmosets show hypermethylation in *GABBR1*, 2 species-specific DMPs show hypomethylation in marmosets (APPENDIX KK). Additionally, 3 marmoset species-specific DMPs show hypermethylation in *HOXD8*, 1 baboon and 2 marmoset species-specific DMPs show hypermethylation in *HOXD9*, 2 chimpanzee species-specific DMPs show hypomethylation in *HOXD9*, and 5 marmoset species-specific DMPs show hypermethylation in *HOXD10*. Of these *HOXD10* species-specific DMPs, 4 have Δβ between 0.2 and 0.3 and 1 has a Δβ < 0.1 (Figure 44, APPENDIX OO).

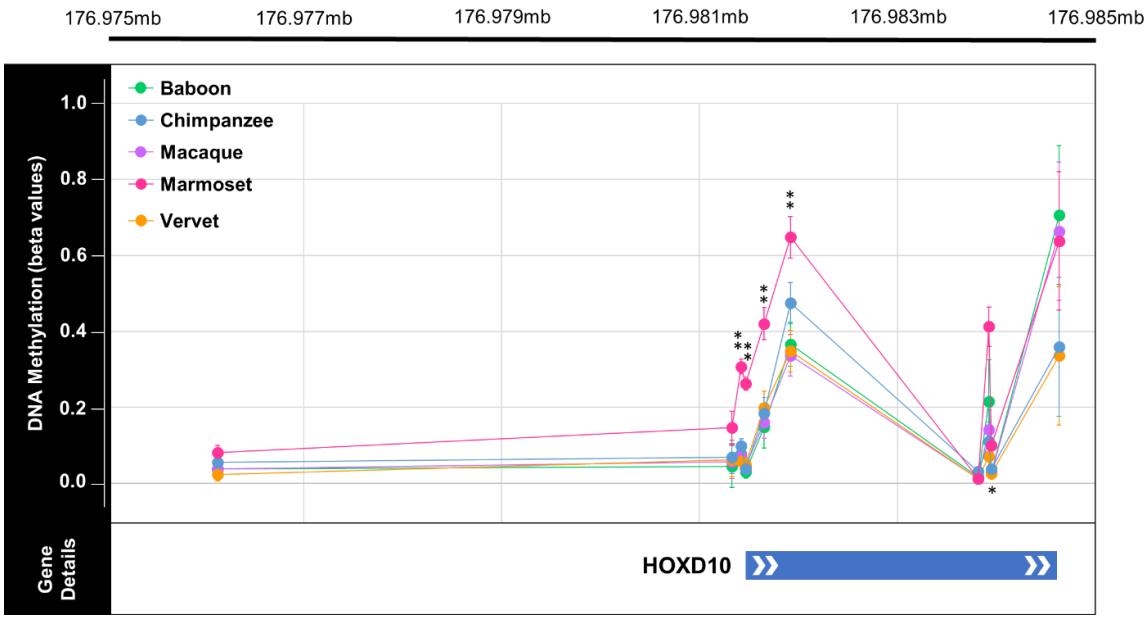

Figure 44. Genome-Wide Methylation Levels Across *HOXD10* in Nonhuman Primates. Plot of the methylation levels of significant DMPs across the *HOXD10* gene (hg19 chr2:176981492-176984670). Plot shows the average β values for each DMP with error bars indicating 1 standard deviation in each direction for each comparative group (teal =

baboon, orange = chimpanzee, purple = macaque, pink = marmoset, and light green = vervet). DMP chromosomal position in relation to the *HOXD10* gene is also depicted. This gene is of interest because it has been found to be differential methylation in ancient and modern hominin species (Gokhman et al. 2014). Of the sites depicted here, 5 DMPs were found to show significant species-specific methylation in marmosets. Of the 5 species-specific DMPs in the *HOXD10* gene of marmosets, 4 have $\Delta\beta$ between 0.2 and 0.3 (**) and 1 has a $\Delta\beta < 0.1$ (*). See Appendix OO for additional information.

Gene-Specific DNA Methylation Profiling and Analyses

Because of the high number of species-specific DMPs within *HOXD10*, methylation patterns across this gene were assessed at a higher resolution using gene-specific sequencing techniques. Regular and bisulfite sequences of several regions across the *HOXD10* were obtained from a subset of EPIC array samples – 3 baboons, 3 macaques, 3 vervets, 3 chimpanzees, and 3 marmosets (APPENDIX BB, APPENDIX PP, APPENDIX QQ, APPENDIX RR). Regular sequencing was very successful, while bisulfite sequencing was less successful, with several sequence reads uninterpretable.

Following the alignment of these sequences to the appropriate nonhuman primate references within the EPO whole-genome multiple alignments of several primate genomes [Ensembl Compara.8_primates_EPO] (Paten, Herrero, Beal, et al. 2008; Paten, Herrero, Fitzgerald, et al. 2008), the presence and absence of methylation across the *HOXD10* gene in each animal was determined (APPENDIX SS). Specifically, methylation was evaluated at human derived CpG sites. Methylation patterns at and around the positions targeted by EPIC array probes generally validate that study's results.

Additionally, these data reveal that across the *HOXD10* gene, nonhuman primates display generally low methylation with some clustered increased amounts of methylation upstream of the gene and at the start of the gene body (Figure 45).

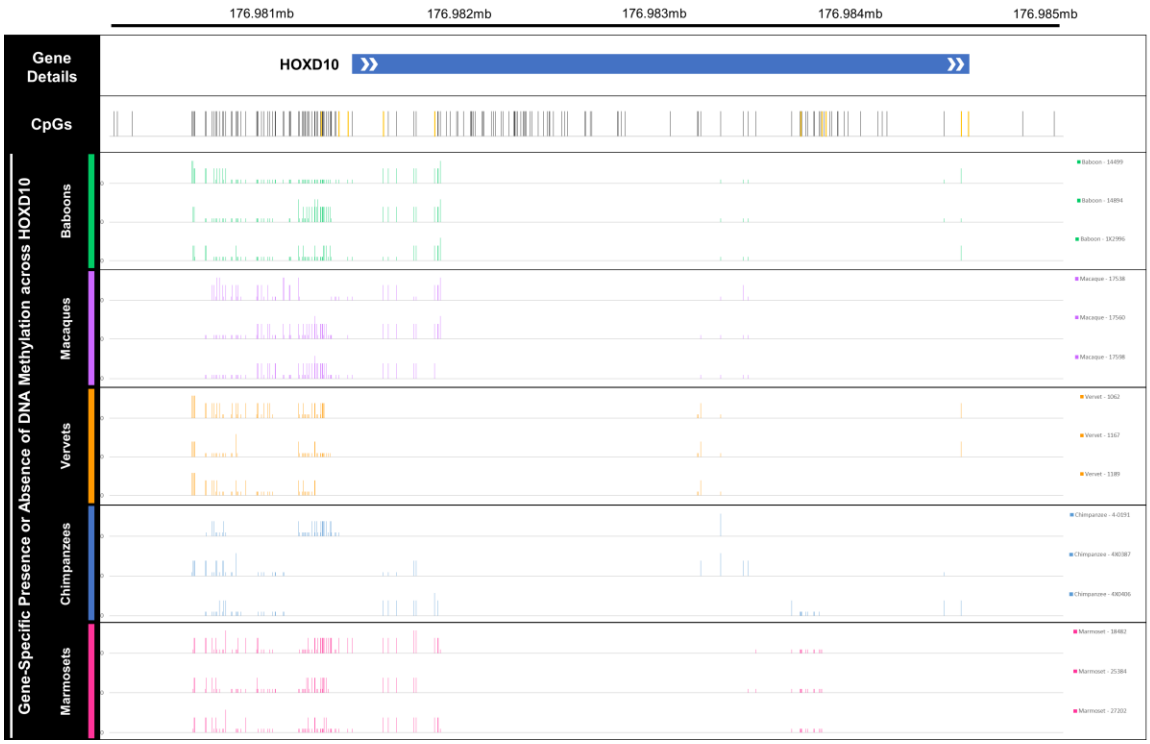

Figure 45. Gene-Specific Methylation Levels Across *HOXD10* in Nonhuman Primates. Bar plot of DNA methylation across the *HOXD10* gene (hg19 chr2:176981492-176984670), as well as upstream and downstream several hundred bases (hg19 chr2:176980532-176985117). Bars depict the presence (tall bar), partial presence (medium bar), or absence (low bar) of methylation at human derived CpG sites in 15 nonhuman primate samples – 3 baboons, 3 macaques, 3 vervets, 3 chimpanzees, and 3 marmosets. While regular sequencing was very successful, bisulfite sequencing was less successful, with several sequence reads uninterpretable. As such, nonhuman primate

methylation data is only available for a subset of the CpGs known in humans. Partial presence of methylation was called when sequencing fluorescence peaks for cytosine and thymine were both present at a particular site and one was at least half the size of the other. Overall, these data provide additional information regarding gene-specific methylation levels across *HOXD10* and confirm some of the EPIC array findings (Figure 44). CpG sites that were also targeted by the EPIC array are highlighted in yellow and include cg18115040 (chr2, position 176981328), cg25371634 (chr2, position 176981422), cg13217260 (chr2, position 176981469), cg03918304 (chr2, position 176981654), cg17489939 (chr2, position 176981919), cg26708100 (chr2, position 176983815), cg10393811 (chr2, position 176983927), cg08992581 (chr2, position 176983949), and cg06005169 (chr2, position 176984634). See Appendix SS for additional information.

Discussion

Here, I used the EPIC array to identify DNA methylation variation in femur bone tissues from five, phylogenetically diverse nonhuman primate species. This was done both to determine the effectiveness of this application for nonhuman primate DNA and to assess how these patterns vary intra- and inter-specifically.

I show that using the EPIC array is feasible for skeletal tissues in several nonhuman primates. *In silico* probe filtering methods (Hernando-Herraez et al. 2013; Ong et al. 2014) indicated that 39% of all human probes could be reliably mapped to the baboon, macaque, and vervet genomes and contained a CpG locus, 76% could be reliably mapped to the chimpanzee genome and contained a CpG locus, and 17% could be

reliable mapped to the marmoset genome and contained a CpG locus. These proportions were slightly lower than expected based on the previous findings of 44% retention for baboons using the 450K array (see Chapter 2). However, as previously described, this discrepancy may be due to the altered design of the EPIC array as compared to the 450K array (see Chapter 3). Additionally, in Old World monkeys, this number was lower than expected since previous researchers were able to use these same methods to reliably map 61% of the human probes on the 450K array to a different Old World monkey, the Cynomologus macaque genome (Ong et al. 2014). As previously discussed, this discrepancy in number may be due to the quality of each nonhuman primates' genome assembly (see Chapter 2). While all of these Old World monkeys have well annotated genomes, the average scaffold length (88,649,475) and contig length (86,040) of the Cynomologus macaque genome (Assembly: Macaca_fascicularis_5.0, Accession: GCF_000264685.2) are generally higher than those in the baboon, Rhesus macaque, and vervet genomes (Table 13). Additionally, the altered design of the EPIC array as compared to the 450K array may also contribute to these differences. Regardless, and as expected, species more closely related to humans (e.g., chimpanzees) have higher numbers of reliably mapped EPIC array probes than species more distantly related to humans (e.g., marmosets).

Subsequent *in silico* analyses based on sequence alignment criteria (Hernando-Herraez et al. 2013) and based on gene symbol criteria (Ong et al. 2014) retained similar numbers of probes for baboons, vervets, and marmosets, but different numbers for macaques and chimpanzees (Figure 35) which is due to the lack of gene information from Ensembl BioMart for these genome versions (Table 13). Nevertheless, again, species

more closely related to humans have higher numbers of retained probes than species more distantly related to humans. Additionally, in chimpanzees, the proportion of probes retained using the alignment criteria (72%) exactly matches the proportion previously found using the same filtering methods for the 450K array (Hernando-Herraez et al. 2013). Furthermore, these retained probes maintained wide and comparable distributions throughout the genome (APPENDIX GG). However, only a little more than half of the resulting probes for each filtering technique overlapped with one another for baboons, macaques, vervets, and marmosets, and for chimpanzees, almost all of the retained probes using the gene symbol criteria overlapped with those retained using the alignment filter criteria (Figure 36). This discrepancy is likely due to the incomplete nature of each nonhuman primate genome annotation, as described previously (see Chapter 2).

Fittingly, after applying the EPIC array to measure DNA methylation patterns of genomic material extracted from skeletal tissues collected from each nonhuman primate, I found that the hybridization efficiency of probes was significantly correlated with the alignment quality of each probe to the baboon genome, and thus, the degree of sequence conservation. The majority of filtered probes for both *in silico* methods passed quality controls and produced robust signals on the array, indicating that either filtering technique may be appropriate for future research. However, because the filtering method based on the alignment filter criteria retained a larger proportion of successfully hybridized probes than the method based on the gene symbol criteria (Figure 37) and because this method is less influenced by the degree of genome assembly annotation, I recommend that this alignment filter criteria method be preferentially used in subsequent nonhuman primate studies.

This work is an extension of previous work and uses the EPIC array to study DNA methylation in several nonhuman primate species. Like its 450K predecessor, the EPIC array is advantageous because it is cost efficient per sample and simultaneously measures a large number of CpG loci with a broad genomic representation (Michels et al. 2013). Similar to this study, previous researchers have used the 450K array to measure DNA methylation patterns in great apes (Hernando-Herraez et al. 2013), which have close evolutionary histories with humans, and in Cynomologus macaques (Ong et al. 2014), which represent another divergent nonhuman primate species. The current research samples primates with a broad range of phylogenetic proximities to humans, which further builds on previous studies. Specifically, I assess how DNA methylation patterns vary intra- and inter-specifically within the primate lineage.

With respect to intra-specific morphology, very few sites were found to be significantly differentially methylated. Specifically, DMPs were only identified in association with baboon bicondylar femur length, baboon maximum femur length, macaque proximal femur width, macaque medial condyle width, vervet superior shaft width, vervet inferior shaft width, vervet anatomical neck height, and chimpanzee anatomical neck length (Table 16). No DMPs were identified in association with marmoset morphologies. Additionally, most of these associations only identified one or a few DMPs. This limited number of associations may be due to the small sample sizes within each species or the small amount of variation identified in each morphology. This second point is supported by the fact that almost all of the morphologies found to have methylation associations also have the highest intra-specific variation in size as compared to morphologies with no methylation associations (Figure 38, APPENDIX CC,

APPENDIX JJ). Alternatively, it may be the case that DNA methylation variation does not have a large influence on nonpathological femur morphology within nonhuman primates.

Of the few associations observed between methylation and morphology, the functional and biological implications of these are questionable. Overall, they likely have no causative effects as research has shown that individual site-specific methylation changes are not readily associated with differential gene expression (Koch et al. 2011; Bork et al. 2010; Y. Chen et al. 2011). Rather, differential gene expression is made possible through the accumulation of several methylation changes within promotor regions (Suzuki and Bird 2008) or across the gene body (Singer et al. 2015). Thus, the individual sites identified as having differential methylation in baboons, macaques, and vervets, likely have no gene regulation or larger biological effects.

This is further supported by the degree of methylation variation observed among these DMPs. Sites that have an average change in mean methylation less than 10% ($\Delta\beta < 0.1$) are thought to have little biological relevance (Hernando-Herraez et al. 2013). Thus, those DMPs associated with morphological variation that have small changes in methylation likely have little to no biological function.

Furthermore, some of the correlations between morphology and methylation that show changes in methylation greater than 10% appear to be highly influenced by a small subset of the sample set. This can be seen in the case of macaque proximal width, where the majority of the change in methylation is due to two individuals (1 female and 1 male) that have extremely low methylation at cg19349877, as compared to all other macaque samples that have high methylation at this site (APPENDIX JJ). Additionally, in the case

of chimpanzee anatomical neck length which was associated with 273 DMPs, these correlations are largely affected by one individual which has an anatomical neck length longer than those of the other three chimpanzees. The validity of these associations could be determined by increasing sample sizes. Nevertheless, such patterns further support the claim that these DMPs have little to no biological effect on morphology.

Lastly, the limited enrichment of GO biological processes and KEGG pathway functions among intra-specific morphologies that had more than 2 significant DMP associations indicates the lack of common function among the genes containing DMPs. Overall, these finding suggest that while some DMPs appear to be associated with morphological variation in nonhuman primates, not enough evidence is present to support them having a functional role in the development and maintenance of this morphological variation. Larger samples sizes are necessary to validate this.

With respect to inter-specific variation, several sites were found to have significant species-specific methylation differences. Specifically, out of the 39,802 sites examined, 650 species-specific DMPs were identified in baboons, 257 in macaques, 639 in vervets, 2,796 in chimpanzees, and 13,778 in marmosets (Table 18), and these span 7,320 genes (APPENDIX LL). However, many of these DMPs had biologically insignificant changes in mean methylation ($\Delta\beta$) (Hernando-Herraez et al. 2013), so various $\Delta\beta$ cutoff thresholds were considered which reduced the overall number of species-specific DMPs among taxa. At the minimum, a 10% change in mean methylation ($\Delta\beta \geq 0.1$) identified 581 species-specific DMPs in baboons, 230 in macaques, 530 in vervets, 1,572 in chimpanzees, and 12,394 in marmosets (Table 18), and these span 6,673 genes (APPENDIX LL). Conversely, at the extreme, a 40% change in mean methylation

($\Delta\beta \geq 0.4$) identified only 15 species-specific DMPs in baboons, 23 in macaques, 54 in vervets, 89 in chimpanzees, and 473 in marmosets (Table 18), and these span only 489 genes (APPENDIX LL).

Regardless of $\Delta\beta$ cutoff threshold, more species-specific DMPs were found in the New World monkey marmosets, followed by the great ape chimpanzees, and lastly the Old World monkey baboons, macaques, and vervets. This trend in numbers of species-specific DMPs is expected given the known phylogenetic relationships between primates, with Old World monkeys more closely related to apes and New World monkeys more distantly related to both groups (Perelman et al. 2011; Rogers and Gibbs 2014). However, while the substantial discrepancy between the number of species-specific DMPs in marmosets as compared to all other taxa may simply be due to marmosets having more species-specific changes, aspects of the experimental design may also contribute to it. As described, marmosets are the only New World primates, so although some of the species-specific DMPs are marmoset specific, others may actually be specific to all New World monkeys. Comparably, DMPs specific to Old World monkeys and catarrhines (which include Old World monkeys and apes) may be cancelled out from this study, as such changes would be shared between all Old World monkeys and catarrhines, respectively. Second, since marmosets are the most phylogenetically distant from humans as compared to the other nonhuman primates included in this study (Perelman et al. 2011; Rogers and Gibbs 2014), the probe filtering steps may have biased downstream data in favor of finding significant results primarily in marmosets. Third, the marmoset data itself has a slightly different normalized data distribution than the other nonhuman primates, with more mean methylation levels of 50% than is usual (Figure 34). Although,

computationally, the filtered array probes examined here appear to hybridize sufficiently for marmosets (Figure 37), there may be other unknown biological or technical issues that impede proper DNA methylation analyses in marmosets using these arrays, and these may have inflated the overall number of marmoset specific DMPs.

The number of species-specific DMPs identified in this study is comparable to those identified in a previous study that assessed methylation patterns in blood from chimpanzees, bonobos, gorillas, and orangutans using the 450K array and similar alignment criteria filtering methods with a focus on sites that had at least a 10% change in mean methylation between species (Hernando-Herraez et al. 2013). This research used a final set of 99,919 probes that were shared across all great ape species and covered 12,593 genes with at least 2 probes per gene. Out of these, 2,284 species-specific DMPs were found in humans, 1,245 in chimpanzees and bonobos, 1,374 in gorillas, and 5,501 in orangutans (Hernando-Herraez et al. 2013). In the current study, when a 10% change in mean methylation ($\Delta\beta \geq 0.1$) was specified, only 1,572 sites were found to show methylation patterns specific to chimpanzees at the exclusion of all other nonhuman species (Table 18). This number is lower than expected given that previously 1,245 sites were found to distinguish chimpanzee from other great apes (Hernando-Herraez et al. 2013), which evolutionarily are more closely related to chimpanzees than are Old World and New World monkeys (Perelman et al. 2011; Rogers and Gibbs 2014). However, the total number of sites examined in the present study is approximately one-third of that examined in the prior great ape study. Thus, a 3-fold increase in the number of sites examined might identify a 3-fold increase in chimpanzee species-specific DMPs as

compared to Old World and New World monkeys, which would be closer to expectations given previous findings (Hernando-Herraez et al. 2013).

Additionally, the number of DMPs that distinguish species from one another in the current study is substantially smaller than the numbers of DMPs identified between different skeletal tissue types (bone vs. cartilage) (see Chapter 3), between different age cohorts (adults vs. juveniles) (see Chapter 4), and between individuals with different skeletal disease states (osteoarthritic vs. healthy) within a nonhuman primate species (see Chapter 3). This finding is to be expected since differences in DNA methylation regulate gene expression (Suzuki and Bird 2008; Singer et al. 2015) should increase or decrease with the degree of differences between cellular functions among comparative groups (Zhang et al. 2013). In the case of different tissues, substantial DNA methylation differences may enable some of the distinct gene regulation and expression that is necessary for the cells in these tissues to promote different tissue functions. In the case of different age cohorts, within the same skeletal tissue, slightly fewer DNA methylation differences may allow cells to emphasize efforts on growth and development in juveniles as compared to maintenance in adults without altering the general function of this tissue to anything other than bone-related functions. In the case of osteoarthritic disease states, even fewer regulatory changes are likely needed to initiate the dysregulation of tissue function than to promote the substantial functional differences between age ranges and tissues types. Lastly, in the cases of differences between adult, skeletally healthy, nonhuman primate species-specific differences, within the same skeletal tissue, only a small number of DNA methylation differences are present. The presence of more regulatory variation within species than between species has been observed in other

studies (Uebbing et al. 2016) and is likely due to the necessity of specific organs to perform specific functions at specific developmental stages regardless of species classification. Thus, it is reasonable for epigenetic and regulatory differences, which control and enable these age-dependent organ functions, to be greater within a species when comparing between tissue types, age cohorts, and possibly disease states, than when comparing between species.

When evaluating how well the methylation patterns at species-specific DMPs cluster animals into distinct species, it was observed that a $\Delta\beta$ cutoff threshold of greater than 0.4 is necessary to achieve this (Figure 40). Less stringent $\Delta\beta$ cutoff thresholds are able to separate apes, New World monkeys, and Old World monkeys into distinct groups, but the baboon, macaque, and vervet Old World monkeys do not form distinct species groups at these thresholds (Figure 39). Global changes in methylation across all finalized 39,802 filtered probes show similar phylogenetic patterns. While average species methylation patterns reveal a well-supported tree topology that reflects known phylogenetic relationships between taxa (Perelman et al. 2011; Rogers and Gibbs 2014) (Figure 41), global changes in methylation among individual animals do not separate the Old World monkey baboons, macaques, and vervets into distinct phylogenetic groups (Figure 42). As before, parsing down these global changes in methylation to only species-specific DMPs with $\Delta\beta \geq 0.4$ enables the phylogenetic separation of these Old World monkey species into distinct lineages (Figure 43). In previous studies, global changes in methylation were able to separate great apes into species-specific phylogenetic groups (Hernando-Herraez et al. 2013). In the current study, the need for a higher $\Delta\beta$ cutoff threshold to distinguish species is likely not due to the length of divergence times

between the great apes and the haplorhines included in this study. Rather, the slow rate of evolution previously observed within baboons and macaques or to methodological constraints of the current study may be the cause.

New World monkeys (platyrrhines) diverged from a common ancestor with catarrhines approximately 43.5 million years ago (MYA), and within catarrhines, Old World monkeys and apes diverged from one another 31.5 MYA. Within Old World monkeys, vervets are a part of the Cercopithecini tribe, while baboons and macaques are part of the Papionini tribe. These tribes diverged 11.5 MYA, and within Papionini, the common ancestors of baboons and macaques diverged 8.1 MYA. On the other hand, the great apes all shared a common ancestor 16.5 MYA, orangutans diverged 16.5 MYA, gorillas diverging 8.3 MYA, and humans and chimpanzees diverging from one another 6.6 MYA (Perelman et al. 2011). The divergence times between baboons, macaques, and vervets are comparable to those between humans, chimpanzees, gorillas, and orangutans. Therefore, the lengths of divergence times between species do not explain why global changes in methylation are unable to fully resolve species-specific phylogenetic clades in the current study as compared to previous studies. While nonhuman great apes have experienced higher rates of molecular evolution as compared to humans (Elango et al. 2006), baboons and macaques have slow rates of molecular evolution as compared to other Old World monkeys (Elango et al. 2009). This slower genetic rate of evolution in baboons and macaques may correspond to slower rates of epigenetic evolution in baboons and macaques, making these Old World monkeys appear more similar to vervets than expected based on divergence times. Correspondingly, this may make resolving the phylogenetic divergences between these taxa more difficult than that between the great

apes. Additionally, the number of sites included in the present study (39,802) as compared to previous studies (99,919) (Hernando-Herraez et al. 2013) may limit the ability of the present study to fully resolve species-specific lineages. This is supported by the low support of several branches in the phylogeny based on global changes in methylation across individual animals (Figure 42). On the other hand, sample size was likely not a contributing factor to the discrepancies between the current study (n=58) and previous studies (n=32) (Hernando-Herraez et al. 2013), as the current study has a slightly larger total sample size and larger average sample sizes per species (n=11.6 and 5.8, respectively).

Alternatively, the fact that global changes in methylation are unable to fully resolve Old World monkey species-specific phylogenetic clades in the current study, may instead indicate that not enough time has passed for Old World monkey species to evolve fixed epigenetic changes between taxa in this tissue. Additionally, it is possible that epigenetic variation at many of the sites examined in this study are under balancing selection in Old World monkeys which prevents these markers from accurately resolving the evolutionary divergences between these species. However, previous research of gene regulation differences between species has found that while some drastic deviations in gene expression may be under directional or balancing selection (Whitehead and Crawford 2006; Romero et al. 2012), most inter-specific regulatory differences appear to be under stabilizing selection or neutral evolution (Brawand et al. 2011; Gilad 2012; Romero et al. 2012).

The evolution of methylation changes along specific nonhuman primate lineages is associated with several functions that may contribute to species-specific phenotypic

differences (APPENDIX MM, APPENDIX NN). First, several skeletal tissue related functions are enriched in species-specific DMPs. Among almost all nonhuman primates, cellular adhesion functions are highly enriched. Cellular adhesion is necessary for cells to attach to other cells or to the extracellular matrix, which is a necessity for bone cells (Mbalaviele et al. 2006). Additionally, in baboons, chimpanzees, and marmosets, anatomical developmental processes are enriched, and in chimpanzees and marmosets, pattern specification processes, limb development, and skeletal system development are enriched. Lastly, in marmosets, specific skeletal functions, such as osteoblast differentiation and ossification are also enriched. Overall, these functions validate that most patterns of differential methylation relate to skeletal tissue function, regulation, development, and maintenance, as well as relate to larger anatomical developmental processes. Additionally, functions not specific to the skeletal system were identified. In chimpanzees and marmosets, transcription and gene expression regulatory functions were enriched. Further, in marmosets, functions related to the development of skeletal muscles, nerves, the brain, the heart, blood vessels, kidneys, eyes, and ears were also enriched. All together, these findings suggest that many species-specific changes in methylation may contribute to the regulation of complex phenotypic changes. While this was not the case for linear skeletal morphology, other skeletal traits may be related.

However, many of the genes associated with the described functions only contain an average of 1-2 differentially methylated sites. Previous studies have shown that individual site-specific methylation changes are not readily associated with differential gene expression (Koch et al. 2011; Bork et al. 2010; Y. Chen et al. 2011). Rather, differential gene expression is made possible through the accumulation of several

methylation changes within promotor regions (Suzuki and Bird 2008) or across the gene body (Singer et al. 2015). Therefore, the enriched functions identified are likely not true biological effects due to methylation differences on their own. Rather they hint at biological effects that may be the result of the combined effects of several genetic, epigenetic, and other regulatory processes.

Finally, some of genes containing species-specific DMPs overlap with those previously identified as being differentially methylated among primates. These include *ARTN*, *COL2A1*, and *GABBR1* which have been found to be differentially methylated in blood among modern humans and great apes (Hernando-Herraez et al. 2013), as well as *HOXD8, HOXD9,* and *HOXD10* which have been found to be differentially methylated among modern and ancient hominins (Gokhman et al. 2014). Specifically, the neurotrophic factor *ARTN* shows species-specific hypomethylation at 1 site in marmosets, and in previous work it shows species-specific hypermethylation at 3 sites in humans as compared to other great apes (Hernando-Herraez et al. 2013). Additionally, *COL2A1* which codes for a type of collagen found in cartilage, shows species-specific hypermethylation at 3 sites in marmosets, and in previous work it shows species-specific hypermethylation of 4 sites in humans as compared to other great apes (Hernando-Herraez et al. 2013). The neuronal receptor *GABBR1*, however, shows more complicated methylation patterns. In the current study, *GABBR1* shows species-specific hypermethylation at 2 sites in baboons, hypermethylation at 3 sites in marmosets, and hypomethylation at 2 sites in marmosets. In previous work, *GABBR1* shows relative hypomethylation in orangutans and relative hypermethylation in chimpanzees and bonobos (Hernando-Herraez et al. 2013). Lastly, 3 genes within the *HOXD* cluster

(*HOXD8*, *HOXD9*, and *HOXD10*), which are involved in limb development, also show complicated methylation patterns. In the current study, *HOXD8* shows species-specific hypermethylation at 3 sites in marmoset, *HOXD9* shows species-specific hypermethylation at 1 site in baboons and 2 sites in marmosets and species-specific hypomethylation at 2 sites in chimpanzees, and *HOXD10* shows species-specific hypermethylation at 5 sites in marmosets. In previous work, *HOXD8* shows hypomethylation in modern and archaic hominins, *HOXD9* shows hypermethylation in archaic hominins as compared to modern humans, and *HOXD10* shows hypermethylation in archaic hominins as compared to modern humans (Gokhman et al. 2014).

Of the *HOXD* cluster genes showing species-specific methylation differences, *HOXD10* contains the largest number of differences, has an active role in anatomical development, and has been found to be differentially methylated among hominins. Thus, it was selected for subsequent DNA methylation profiling and analysis at a higher resolution using gene-specific sequencing techniques. *HOXD10* specifically codes for a protein that functions as a sequence-specific transcription factor which is expressed in the developing limb buds and is involved in differentiation and limb development. In the current study, each nonhuman primate species shows low to intermediate methylation levels across the gene body, with 5 sites from the EPIC array showing slightly higher methylation levels in marmosets as compared to other taxa (Figure 44, APPENDIX OO). A similar pattern is observed in the gene-specific methylation data, but it further reveals that in general, *HOXD10* is not highly methylated in nonhuman primates. However, across all taxa, some clusters of higher methylation are found upstream of the gene and at

the start of the gene body, with marmosets on average displaying more methylation in the gene body than other taxa (Figure 45, APPENDIX SS).

In hominins, humans display hypomethylation across the *HOXD10* gene body, Neandertals displayed intermediate methylation levels, and Denisovans displayed high levels of methylation (Gokhman et al. 2014). The variation of methylation patterns in the gene body of this gene suggest that intermediate methylation levels may be a more ancestral epigenetic state for this region in the primate lineage, while the extreme hypermethylation of this region in Denisovans and the extreme hypomethylation of this region in humans may be derived epigenetic states. Previous work has proposed that methylation differences in *HOXD10* may be associated with phenotypic distinctions between modern human and archaic hominin limbs (Gokhman et al. 2014). However, the current study did not find substantial associations between methylation variation and aspects of femur morphology within nonhuman primate species. Further work to understand the role of differential methylation of *HOXD10* in promoting morphological changes of the limb should be explored.

In conclusion, I determined that the EPIC array can be used to measure genome-wide DNA methylation in skeletal tissues from several nonhuman primates. While significant associations were not readily found between methylation and nonpathological skeletal morphologies, several significant differences in methylation were observed inter-specifically. This is the first study to specifically assess DNA methylation using this method in skeletal tissues from a taxonomically diverse set of nonhuman primates. From an evolutionary perspective, the results of this study reveal DNA methylation variation in five species, as well as the degree to which intra-specific methylation variation relates to

skeletal morphology and the degree to which methylation variation is present inter-specifically. Methylation patterns are not substantially associated with nonpathological morphologies. However, among species, methylation variation is associated with genes that impact skeletal development and maintenance, and this may have direct downstream regulatory and phenotypic effects. Additionally, nonhuman primate species-specific DNA methylation patterns help provide a phylogenetic context in which to frame ancient hominin epigenetic studies.

Acknowledgements

Megann Phillips

School of Human Evolution & Social Change, Arizona State University, Tempe, AZ, USA

Anne Sheldrake, Jaydee Turner, Kara Peterson, Mel Carless, & Laura Cox

Department of Genetics, Texas Biomedical Research Institute, San Antonio, TX, USA

Funding

Leakey Foundation Research Grant for Doctoral Students to G. Housman

Wenner-Gren Foundation Dissertation Fieldwork Grant (Gr. 9310) to G. Housman

James F. Nacey Fellowship from the Nacey Maggioncalda Foundation to G. Housman

International Primatological Society Research Grant to G. Housman

Sigma Xi Grant-in-Aid of Research to G. Housman

Center for Evolution and Medicine Venture Fund (ASU) to G. Housman

Graduate Research and Support Program Grant (GPSA, ASU) to G. Housman

Graduate Student Research Grant (SHESC, ASU) to G. Housman

CHAPTER 6

CONCLUSION

The current research assesses how epigenetic patterns vary intra- and inter-specifically in primate taxa and in relation to skeletal phenotypes. Specifically, the overarching objective of this work was to identify genome-wide and gene-specific DNA methylation patterns in nonhuman primate skeletal tissues and assess variation within species by determining how patterns differ between tissue types, between age ranges, between skeletal disease states, between nonpathological skeletal morphologies, and among species.

Overall, more significantly differentially methylated positions (DMPs) were identified within species than between species. Within species, skeletal tissue types (bone vs. cartilage) (see Chapter 3) showed more DMPs than different age cohorts (adults vs. juveniles) (see Chapter 4), which showed more DMPs than different skeletal disease states (OA vs. healthy) (see Chapter 3) or variations in nonpathological morphologies of the femur (see Chapter 5). These finding are to be expected since other research has also observed more regulatory variation within species than between species (Uebbing et al. 2016). Additionally, within species, differences in DNA methylation and its regulation of gene expression (Suzuki and Bird 2008; Singer et al. 2015) should increase or decrease with the degree of differences between cellular functions among comparative groups (Zhang et al. 2013).

In Chapter 2, I used the Illumina Infinium Human Methylation450 BeadChip (450K array) to identify DNA methylation patterns in bone and cartilage of age-matched, adult female baboons, five with and five without knee OA. I validated that the

hybridization efficiency of 450K array probes is related to the degree of sequence similarity between the probes and the baboon genome. Additionally, approximately 44% of the 450K array probes reliably align to the baboon genome, contain a CpG site of interest, and maintain a wide distribution throughout the genome. I also found that filtering probes using alignment similarity criteria retains more efficiently hybridized probes than filtering probes using gene symbol similarity criteria. Both filtering methods identified significant DMPs between healthy and OA individuals in cartilage tissues. Specifically, these DMPs are associated with 8 genes, and the methylation patterns identified in 1 of these genes overlap with those previously identified in humans. Conversely, in bone tissues, no DMPs were found between disease states, and no DMPs were found between tissue types. In summary, I conclude that the 450K array can be used to measure genome-wide DNA methylation in baboon tissues and identify significant associations with complex traits and that within primates, while some DNA methylation patterns associated with OA are evolutionarily conserved across taxa, others are not.

In Chapter 3, I used the Illumina Infinium MethylationEPIC BeadChip (EPIC array) to further explore the evolution of OA epigenetics by identifying DNA methylation patterns in bone and cartilage of 56 pedigreed, adult baboons, 28 with and 28 without knee OA, and by assessing whether DNA methylation variation is associated with OA in baboons and in a manner similar to that observed in humans. Several significant DMPs were found between tissue types. Within cartilage tissue, many DMPs were also identified between healthy and OA individuals. Conversely, very few DMPs were identified between disease states in bone tissue. In summary, these finding provide some insight into the etiology of OA. Furthermore, some genes containing DMPs overlap with

and display methylation patterns similar to those previously identified in human OA studies, while others genes do not. These results provide insight into the evolutionary conservation of epigenetic mechanisms associated with OA. From an evolutionary perspective, these results provide evidence for DNA methylation variation in skeletal tissue from one primate species and two skeletal tissues. They also reveal the degree to which the common skeletal condition OA affects this variation.

In Chapter 4, I used the EPIC array to examine the evolution of aging epigenetics specifically within the primate lineage by identifying DNA methylation patterns in bone from 46 pedigreed baboons, 28 that were adults and 18 that were juveniles, and assessing whether DNA methylation variation is associated with aging in baboons and in a manner similar to that observed in humans. Several significant DMPs were found between these age cohorts, and similar to other animals, adult baboons display global hypomethylation as compared to juvenile baboons. The significant age-related DMPs identified are associated with genes involved in development processes and pathways related to the progression of diseases of aging. Additionally, while some of these age-related DMPs overlap with and display methylation patterns similar to those previously identified in human aging studies, the majority of previously identified age-related methylation loci in humans were not significantly differentially methylated in baboons. Nevertheless, methylation levels at these human loci are still able to differentiate baboon age cohorts. In summary, these results reveal how DNA methylation varies with respect to age in skeletal tissues from one primate species and provide insight into the evolutionary conservation of aging epigenetics within the primate lineage.

Lastly, in Chapter 5, I used the EPIC array and gene-specific methylation sequencing to identify DNA methylation patterns in bone from 28 baboons, 10 macaques, 10 vervets, 4 chimpanzees, and 6 marmosets in order to assess how these patterns vary intra-specifically in relation to nonpathological femur bone morphologies and inter-specifically. I validated that the hybridization efficiency of EPIC array probes is related to the degree of sequence similarity between the probes and each nonhuman primate genome and determined that approximately 39%, 39%, 39%, 76%, and 17% of the EPIC array probes reliably align to the baboon, macaque, vervet, chimpanzee, and marmoset genomes, respectively, contain a CpG site of interest, and maintain a wide distribution throughout the genome. I also found that filtering probes using alignment similarity criteria retains more efficiently hybridized probes than filtering probes using gene symbol similarity criteria. Additionally, significant differential methylation was identified in a subset of morphological variants within species. However, these significant DMPs likely do not have large biological effects and may be confounded by other variables associated with morphological variation. Furthermore, I found several sites that show species-specific methylation patterns. Higher resolution of methylation variation across a subset of these regions confirms these patterns and provides more insight into their evolution history. Finally, genome-wide DNA methylation patterns across all 39,802 sites examined produce a topology that reflects known phylogenetic relationships between taxa. In summary, from an evolutionary perspective, these findings give us an appreciation of DNA methylation variation in skeletal tissues within and among five nonhuman primate species. They also provide insight into the degree to which this

epigenetic variation relates to variation in skeletal morphology and taxonomic differences.

In conclusion, although this work is exploratory in nature and cannot make substantial claims regarding the mechanisms contributing to DNA methylation variation or the functional consequences of this variation within and among nonhuman species, it does identify intra- and inter-specific DNA methylation variation in nonhuman primate skeletal tissues and how this variation relates to complex traits. It also helps to inform our understanding of human evolution by examining the taxonomic diversity of primates more deeply and by exploring DNA methylation in novel tissues among nonhuman primates. With respect to complex skeletal phenotypes, it expands the range of traits that can be explored with respect to epigenetic mechanisms. Additionally, it builds on previous pathology focused studies to encourage an evolutionary medicine understanding and investigates novel nonpathological phenotypes. Finally, it begins to build a more robust phylogenetic context in which ancient epigenetic studies can be framed. All together this work begins to lay the foundation for future inquiries into the evolution of epigenetic changes within the primate lineage and their contribution to variations in complex phenotypes.

REFERENCES

Adkins, Ronald M, Fridtjof Thomas, Frances A Tylavsky, and Julia Krushkal. 2011. Parental Ages and Levels of DNA Methylation in the Newborn Are Correlated. *BMC Medical Genetics* 12 (March): 47. doi:10.1186/1471-2350-12-47.

Almeida, Maria Ines, Andreia Machado Silva, Daniel Marques Vasconcelos, Catarina Rodrigues Almeida, Hugo Caires, Marta Teixeira Pinto, George Adrian Calin, Susana Gomes Santos, and Mário Adolfo Barbosa. 2016. miR-195 in Human Primary Mesenchymal Stromal/Stem Cells Regulates Proliferation, Osteogenesis and Paracrine Effect on Angiogenesis. *Oncotarget* 7 (1): 7–22. doi:10.18632/oncotarget.6589.

Altschul, S F, T L Madden, A A Schäffer, J Zhang, Z Zhang, W Miller, and D J Lipman. 1997. Gapped BLAST and PSI-BLAST: A New Generation of Protein Database Search Programs. *Nucleic Acids Research* 25 (17): 3389–3402.

Alvarez-Garcia, Oscar, Kathleen M. Fisch, Nathan E. Wineinger, Ryuichiro Akagi, Masahiko Saito, Takahisa Sasho, Andrew I. Su, and Martin K. Lotz. 2016. Increased DNA Methylation and Reduced Expression of Transcription Factors in Human Osteoarthritis Cartilage: Differential DNA Methylation in OA Cartilage. *Arthritis & Rheumatology*, February, n/a-n/a. doi:10.1002/art.39643.

Ameye, Laurent G, and Marian F Young. 2006. Animal Models of Osteoarthritis: Lessons Learned While Seekiseeking the 'Holy Grail.' *Current Opinion in Rheumatology* 18 (5): 537–547.

Andrés, María C. de, Emmajayne Kingham, Kei Imagawa, Antonio Gonzalez, Helmtrud I. Roach, David I. Wilson, and Richard O. C. Oreffo. 2013. Epigenetic Regulation during Fetal Femur Development: DNA Methylation Matters. Edited by Nuno M. Neves. *PLoS ONE* 8 (1): e54957. doi:10.1371/journal.pone.0054957.

Ankel-Simons, Friderun. 2007. *Primate Anatomy*. 3rd ed. New York: Academic Press.

Aref-Eshghi, Erfan, Yuhua Zhang, Ming Liu, Patricia E. Harper, Glynn Martin, Andrew Furey, Roger Green, et al. 2015. Genome-Wide DNA Methylation Study of Hip and Knee Cartilage Reveals Embryonic Organ and Skeletal System Morphogenesis as Major Pathways Involved in Osteoarthritis. *BMC Musculoskeletal Disorders* 16 (1): 1–10. doi:10.1186/s12891-015-0745-5.

Aryee, M. J., A. E. Jaffe, H. Corrada-Bravo, C. Ladd-Acosta, A. P. Feinberg, K. D. Hansen, and R. A. Irizarry. 2014. Minfi: A Flexible and Comprehensive Bioconductor Package for the Analysis of Infinium DNA Methylation Microarrays. *Bioinformatics* 30 (10): 1363–69. doi:10.1093/bioinformatics/btu049.

Babbitt, C. C., O. Fedrigo, A. D. Pfefferle, A. P. Boyle, J. E. Horvath, T. S. Furey, and G. A. Wray. 2010. Both Noncoding and Protein-Coding RNAs Contribute to Gene Expression Evolution in the Primate Brain. *Genome Biology and Evolution* 2 (0): 67–79. doi:10.1093/gbe/evq002.

Barnett, Ross, and Greger Larson. 2012. A Phenol–Chloroform Protocol for Extracting DNA from Ancient Samples. In *Ancient DNA*, edited by Beth Shapiro and Michael Hofreiter, 13–19. Methods in Molecular Biology 840. Humana Press. doi:10.1007/978-1-61779-516-9_2.

Becer, Eda, Güldal Mehmetçik, Halin Bareke, and Nedime Serakıncı. 2013. Association of Leptin Receptor Gene Q223R Polymorphism on Lipid Profiles in Comparison Study between Obese and Non-Obese Subjects. *Gene* 529 (1): 16–20. doi:10.1016/j.gene.2013.08.003.

Bell, Christopher G., Gareth A. Wilson, Lee M. Butcher, Christian Roos, Lutz Walter, and Stephan Beck. 2012. Human-Specific CpG "beacons" Identify Loci Associated with Human-Specific Traits and Disease. *Epigenetics* 7 (10): 1188–99. doi:10.4161/epi.22127.

Bell, Christopher G., Yudong Xia, Wei Yuan, Fei Gao, Kirsten Ward, Leonie Roos, Massimo Mangino, et al. 2016. Novel Regional Age-Associated DNA Methylation Changes within Human Common Disease-Associated Loci. *Genome Biology* 17 (1): 193. doi:10.1186/s13059-016-1051-8.

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). doi:10.1371/journal.pgen.1002629.

Bendele, A. M. 2001. Animal Models of Osteoarthritis. *J Musculoskelet Neuronal Interact* 1 (4): 363–76.

Benjamini, Yoav, and Yosef Hochberg. 1995. Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. *Journal of the Royal Statistical Society. Series B (Methodological)* 57 (1): 289–300.

Berdyshev, G. D., G. K. Korotaev, G. V. Boiarskikh, and B. F. Vaniushin. 1967. [Nucleotide composition of DNA and RNA from somatic tissues of humpback and its changes during spawning]. *Biokhimiia (Moscow, Russia)* 32 (5): 988–93.

Blagojevic, M., C. Jinks, A. Jeffery, and K. P. Jordan. 2010. Risk Factors for Onset of Osteoarthritis of the Knee in Older Adults: A Systematic Review and Meta-Analysis. *Osteoarthritis and Cartilage* 18 (1): 24–33. doi:10.1016/j.joca.2009.08.010.

Blekhman, Ran, Alicia Oshlack, Adrien E. Chabot, Gordon K. Smyth, and Yoav Gilad. 2008. Gene Regulation in Primates Evolves under Tissue-Specific Selection Pressures. Edited by Gil McVean. *PLoS Genetics* 4 (11): e1000271. doi:10.1371/journal.pgen.1000271.

Bocker, Michael T., Isabelle Hellwig, Achim Breiling, Volker Eckstein, Anthony D. Ho, and Frank Lyko. 2011. Genome-Wide Promoter DNA Methylation Dynamics of Human Hematopoietic Progenitor Cells during Differentiation and Aging. *Blood* 117 (19): e182–89. doi:10.1182/blood-2011-01-331926.

Bocklandt, Sven, Wen Lin, Mary E. Sehl, Francisco J. Sánchez, Janet S. Sinsheimer, Steve Horvath, and Eric Vilain. 2011. Epigenetic Predictor of Age. *PLOS ONE* 6 (6): e14821. doi:10.1371/journal.pone.0014821.

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. doi:10.1016/j.mad.2008.12.003.

Bongers, Ernie M. H. F., Pascal H. G. Duijf, Sylvia E. M. van Beersum, Jeroen Schoots, Albert Van Kampen, Andreas Burckhardt, Ben C. J. Hamel, et al. 2004. Mutations in the Human TBX4 Gene Cause Small Patella Syndrome. *American Journal of Human Genetics* 74 (6): 1239–48. doi:10.1086/421331.

Bork, Simone, Stefan Pfister, Hendrik Witt, Patrick Horn, Bernhard Korn, Anthony D Ho, and Wolfgang Wagner. 2010. DNA Methylation Pattern Changes upon Long-Term Culture and Aging of Human Mesenchymal Stromal Cells. *Aging Cell* 9 (1): 54–63. doi:10.1111/j.1474-9726.2009.00535.x.

Bovée, Judith V. M. G., Pancras C. W. Hogendoorn, Jay S. Wunder, and Benjamin A. Alman. 2010. Cartilage Tumours and Bone Development: Molecular Pathology and Possible Therapeutic Targets. *Nature Reviews Cancer* 10 (7): 481–88. doi:10.1038/nrc2869.

Brand-Saberi, Beate. 2005. Genetic and Epigenetic Control of Skeletal Muscle Development. *Annals of Anatomy - Anatomischer Anzeiger* 187 (3): 199–207. doi:10.1016/j.aanat.2004.12.018.

Brawand, David, Magali Soumillon, Anamaria Necsulea, Philippe Julien, Gábor Csárdi, Patrick Harrigan, Manuela Weier, et al. 2011. The Evolution of Gene Expression Levels in Mammalian Organs. *Nature* 478 (7369): 343–48. doi:10.1038/nature10532.

Cáceres, Mario, Joel Lachuer, Matthew A. Zapala, John C. Redmond, Lili Kudo, Daniel H. Geschwind, David J. Lockhart, Todd M. Preuss, and Carrolee Barlow. 2003.

Elevated Gene Expression Levels Distinguish Human from Non-Human Primate Brains. *Proceedings of the National Academy of Sciences* 100 (22): 13030–13035.

Campisi, Judith, and Jan Vijg. 2009. Does Damage to DNA and Other Macromolecules Play a Role in Aging? If So, How? *The Journals of Gerontology Series A: Biological Sciences and Medical Sciences* 64A (2): 175–78. doi:10.1093/gerona/gln065.

Chen, Brian H., Riccardo E. Marioni, Elena Colicino, Marjolein J. Peters, Cavin K. Ward-Caviness, Pei-Chien Tsai, Nicholas S. Roetker, et al. 2016. DNA Methylation-Based Measures of Biological Age: Meta-Analysis Predicting Time to Death. *Aging*, September. doi:10.18632/aging.101020.

Chen, Yi-an, Sanaa Choufani, Jose Carlos Ferreira, Daria Grafodatskaya, Darci T. Butcher, and Rosanna Weksberg. 2011. Sequence Overlap between Autosomal and Sex-Linked Probes on the Illumina HumanMethylation27 Microarray. *Genomics* 97 (4): 214–22. doi:10.1016/j.ygeno.2010.12.004.

Chen, Yi-an, Mathieu Lemire, Sanaa Choufani, Darci T. Butcher, Daria Grafodatskaya, Brent W. Zanke, Steven Gallinger, Thomas J. Hudson, and Rosanna Weksberg. 2013. Discovery of Cross-Reactive Probes and Polymorphic CpGs in the Illumina Infinium HumanMethylation450 Microarray. *Epigenetics* 8 (2): 203–9. doi:10.4161/epi.23470.

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. doi:10.1371/journal.pgen.1000602.

Christiansen, Lene, Adam Lenart, Qihua Tan, James W. Vaupel, Abraham Aviv, Matt McGue, and Kaare Christensen. 2016. DNA Methylation Age Is Associated with Mortality in a Longitudinal Danish Twin Study. *Aging Cell* 15 (1): 149–54. doi:10.1111/acel.12421.

Clarke, B. 2008. Normal Bone Anatomy and Physiology. *Clinical Journal of the American Society of Nephrology* 3 (Supplement 3): S131–39. doi:10.2215/CJN.04151206.

Cooper, Cyrus, Shelagh Snow, Timothy E. McAlindon, Samantha Kellingray, Brenda Stuart, David Coggon, and Paul A. Dieppe. 2000. Risk Factors for the Incidence and Progression of Radiographic Knee Osteoarthritis. *Arthritis & Rheumatism* 43 (5): 995–1000. doi:10.1002/1529-0131(200005)43:5<995::AID-ANR6>3.0.CO;2-1.

Corner, Brian, Subhash Lele, and Joan Richtsmeier. 1992. Measuring Precision of Three-Dimensional Landmark Data. *Journal of Quantitative Anthropology* 3: 347–59.

Cox, L. A., A. G. Comuzzie, L. M. Havill, G. M. Karere, K. D. Spradling, M. C. Mahaney, P. W. Nathanielsz, et al. 2013. Baboons as a Model to Study Genetics and Epigenetics of Human Disease. *ILAR Journal* 54 (2): 106–21. doi:10.1093/ilar/ilt038.

Crabbe, Patricia, Stefan Goemaere, Hans Zmierczak, Inge Van Pottelbergh, Dirk De Bacquer, and Jean-Marc Kaufman. 2006. Are Serum Leptin and the Gln223Arg Polymorphism of the Leptin Receptor Determinants of Bone Homeostasis in Elderly Men? *European Journal of Endocrinology* 154 (5): 707–14. doi:10.1530/eje.1.02130.

Cross, Marita, Emma Smith, Damian Hoy, Sandra Nolte, Ilana Ackerman, Marlene Fransen, Lisa Bridgett, et al. 2014. The Global Burden of Hip and Knee Osteoarthritis: Estimates from the Global Burden of Disease 2010 Study. *Annals of the Rheumatic Diseases*, February, annrheumdis-2013-204763. doi:10.1136/annrheumdis-2013-204763.

Cucchiarini, Magali, Laura de Girolamo, Giuseppe Filardo, J. Miguel Oliveira, Patrick Orth, Dietrich Pape, and Pascal Reboul. 2016. Basic Science of Osteoarthritis. *Journal of Experimental Orthopaedics* 3 (1). doi:10.1186/s40634-016-0060-6.

Cushnaghan, J., and P. Dieppe. 1991. Study of 500 Patients with Limb Joint Osteoarthritis. I. Analysis by Age, Sex, and Distribution of Symptomatic Joint Sites. *Annals of the Rheumatic Diseases* 50 (1): 8–13. doi:10.1136/ard.50.1.8.

Dastani, Zari, Marie-France Hivert, Nicholas Timpson, John R. B. Perry, Xin Yuan, Robert A. Scott, Peter Henneman, et al. 2012. Novel Loci for Adiponectin Levels and Their Influence on Type 2 Diabetes and Metabolic Traits: A Multi-Ethnic Meta-Analysis of 45,891 Individuals. *PLoS Genetics* 8 (3). doi:10.1371/journal.pgen.1002607.

Del Campo, M., M. C. Jones, A. N. Veraksa, C. J. Curry, K. L. Jones, J. T. Mascarello, Z. Ali-Kahn-Catts, T. Drumheller, and W. McGinnis. 1999. Monodactylous Limbs and Abnormal Genitalia Are Associated with Hemizygosity for the Human 2q31 Region That Includes the HOXD Cluster. *American Journal of Human Genetics* 65 (1): 104–10. doi:10.1086/302467.

Delgado-Calle, Jesús, Agustín F. Fernández, Jesús Sainz, María T. Zarrabeitia, Carolina Sañudo, Raúl García-Renedo, María I. Pérez-Núñez, Carmen García-Ibarbia, Mario F. Fraga, and José A. Riancho. 2013. Genome-Wide Profiling of Bone Reveals Differentially Methylated Regions in Osteoporosis and Osteoarthritis. *Arthritis & Rheumatism* 65 (1): 197–205. doi:10.1002/art.37753.

Dhanoa, Bajinder S, Tiziana Cogliati, Akhila G Satish, Elspeth A Bruford, and James S Friedman. 2013. Update on the Kelch-like (KLHL) Gene Family. *Human Genomics* 7 (1): 13. doi:10.1186/1479-7364-7-13.

Dimitriou, Rozalia, Elena Jones, Dennis McGonagle, and Peter V. Giannoudis. 2011. Bone Regeneration: Current Concepts and Future Directions. *BMC Medicine* 9 (1): 1.

Dongen, Jenny van, Erik Ehli, Roderick Slieker, Meike Bartels, Zachary Weber, Gareth Davies, P. Slagboom, Bastiaan Heijmans, and Dorret Boomsma. 2014. Epigenetic Variation in Monozygotic Twins: A Genome-Wide Analysis of DNA Methylation in Buccal Cells. *Genes* 5 (2): 347–65. doi:10.3390/genes5020347.

Du, Pan, Xiao Zhang, Chiang-Ching Huang, Nadereh Jafari, Warren A. Kibbe, Lifang Hou, and Simon M. Lin. 2010. Comparison of Beta-Value and M-Value Methods for Quantifying Methylation Levels by Microarray Analysis. *BMC Bioinformatics* 11: 587. doi:10.1186/1471-2105-11-587.

Duong, Tarn. 2013. Local Significant Differences from Nonparametric Two-Sample Tests. *Journal of Nonparametric Statistics* 25 (3): 635–45. doi:10.1080/10485252.2013.810217.

Edge, Christopher, Clare Gooding, and Christopher W. J. Smith. 2013. Dissecting Domains Necessary for Activation and Repression of Splicing by Muscleblind-like Protein 1. *BMC Molecular Biology* 14 (December): 29. doi:10.1186/1471-2199-14-29.

Elango, Navin, Jeeyoung Lee, Zuogang Peng, Yong-Hwee E. Loh, and Soojin V. Yi. 2009. Evolutionary Rate Variation in Old World Monkeys. *Biology Letters* 5 (3): 405–8. doi:10.1098/rsbl.2008.0712.

Elango, Navin, James W. Thomas, NISC Comparative Sequencing Program‡§, and Soojin V. Yi. 2006. Variable Molecular Clocks in Hominoids. *Proceedings of the National Academy of Sciences* 103 (5): 1370–75. doi:10.1073/pnas.0510716103.

Enard, Wolfgang, Anne Fassbender, Fabian Model, Péter Adorján, Svante Pääbo, and Alexander Olek. 2004. Differences in DNA Methylation Patterns between Humans and Chimpanzees. *Current Biology* 14 (4): R148–R149.

Enjuanes, A., A. Supervía, X. Nogués, and A. Díez-Pérez. 2002. Leptin Receptor (OB-R) Gene Expression in Human Primary Osteoblasts: Confirmation. *Journal of Bone and Mineral Research: The Official Journal of the American Society for Bone and Mineral Research* 17 (6): 1135; author reply 1136. doi:10.1359/jbmr.2002.17.6.1135.

Fairbrother, Una L., László B. Tankó, Andrew J. Walley, Claus Christiansen, Philippe Froguel, and Alexandra I. F. Blakemore. 2007. Leptin Receptor Genotype at Gln223Arg Is Associated with Body Composition, BMD, and Vertebral Fracture in Postmenopausal Danish Women. *Journal of Bone and Mineral Research: The Official Journal of the American Society for Bone and Mineral Research* 22 (4): 544–50. doi:10.1359/jbmr.070114.

Farcas, R., E. Schneider, K. Frauenknecht, I. Kondova, R. Bontrop, J. Bohl, B. Navarro, et al. 2009. Differences in DNA Methylation Patterns and Expression of the CCRK Gene in Human and Nonhuman Primate Cortices. *Molecular Biology and Evolution* 26 (6): 1379–89. doi:10.1093/molbev/msp046.

Felson, D. T., and Y. Zhang. 1998. An Update on the Epidemiology of Knee and Hip Osteoarthritis with a View to Prevention. *Arthritis and Rheumatism* 41 (8): 1343–55. doi:10.1002/1529-0131(199808)41:8<1343::AID-ART3>3.0.CO;2-9.

Felson, David T. 2004. Risk Factors for Osteoarthritis: Understanding Joint Vulnerability. *Clinical Orthopaedics and Related Research*, no. 427 Suppl (October): S16-21.

Fernández-Tajes, Juan, Angel Soto-Hermida, Maria E. Vázquez-Mosquera, Estefania Cortés-Pereira, Alejandro Mosquera, Mercedes Fernández-Moreno, Natividad Oreiro, et al. 2014. Genome-Wide DNA Methylation Analysis of Articular Chondrocytes Reveals a Cluster of Osteoarthritic Patients. *Annals of the Rheumatic Diseases* 73 (4): 668–77. doi:10.1136/annrheumdis-2012-202783.

Flanagan, James M., Violeta Popendikyte, Natalija Pozdniakovaite, Martha Sobolev, Abbas Assadzadeh, Axel Schumacher, Masood Zangeneh, et al. 2006. Intra- and Interindividual Epigenetic Variation in Human Germ Cells. *American Journal of Human Genetics* 79 (1): 67–84.

Fleagle, John G. 1999. *Primate Adaptation and Evolution*. New York: Academic Press.

Fortin, Jean-Philippe, Aurélie Labbe, Mathieu Lemire, Brent W Zanke, Thomas J Hudson, Elana J Fertig, Celia MT Greenwood, and Kasper D Hansen. 2014. Functional Normalization of 450k Methylation Array Data Improves Replication in Large Cancer Studies. *Genome Biology* 15 (11). doi:10.1186/s13059-014-0503-2.

Fortin, Jean-Philippe, Timothy Triche, and Kasper Hansen. 2016. Preprocessing, Normalization and Integration of the Illumina HumanMethylationEPIC Array. *bioRxiv*, July, 065490. doi:10.1101/065490.

Fraga, Mario F., Ruben Agrelo, and Manel Esteller. 2007. Cross-Talk between Aging and Cancer. *Annals of the New York Academy of Sciences* 1100 (1): 60–74. doi:10.1196/annals.1395.005.

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. doi:10.1073/pnas.0500398102.

Fraga, Mario F., and Manel Esteller. 2007. Epigenetics and Aging: The Targets and the Marks. *Trends in Genetics* 23 (8): 413–18. doi:10.1016/j.tig.2007.05.008.

Fukuda, Kei, Kenji Ichiyanagi, Yoichi Yamada, Yasuhiro Go, Toshifumi Udono, Seitaro Wada, Toshiyuki Maeda, et al. 2013. Regional DNA Methylation Differences between Humans and Chimpanzees Are Associated with Genetic Changes, Transcriptional Divergence and Disease Genes. *Journal of Human Genetics* 58 (7): 446–54. doi:10.1038/jhg.2013.55.

Gama-Sosa, Miguel A., Rose Marie Midgett, Valerie A. Slagel, Sherwood Githens, Kenneth C. Kuo, Charles W. Gehrke, and Melanie Ehrlich. 1983. Tissue-Specific Differences in DNA Methylation in Various Mammals. *Biochimica et Biophysica Acta (BBA)-Gene Structure and Expression* 740 (2): 212–219.

García-Ibarbia, Carmen, Jesús Delgado-Calle, Iñigo Casafont, Javier Velasco, Jana Arozamena, María I. Pérez-Núñez, María A. Alonso, et al. 2013. Contribution of Genetic and Epigenetic Mechanisms to Wnt Pathway Activity in Prevalent Skeletal Disorders. *Gene* 532 (2): 165–72. doi:10.1016/j.gene.2013.09.080.

Garvie, Colin W., and Jeremy M. Boss. 2008. Assembly of the RFX Complex on the MHCII Promoter: Role of RFXAP and RFXB in Relieving Autoinhibition of RFX5. *Biochimica Et Biophysica Acta* 1779 (12): 797–804. doi:10.1016/j.bbagrm.2008.07.012.

Gates, Devika P., Leslie A. Coonrod, and J. Andrew Berglund. 2011. Autoregulated Splicing of Muscleblind-like 1 (MBNL1) Pre-mRNA. *The Journal of Biological Chemistry* 286 (39): 34224–33. doi:10.1074/jbc.M111.236547.

Geeleher, Paul, Lori Hartnett, Laurance J. Egan, Aaron Golden, Raja Affendi Raja Ali, and Cathal Seoighe. 2013. Gene-Set Analysis Is Severely Biased When Applied to Genome-Wide Methylation Data. *Bioinformatics (Oxford, England)* 29 (15): 1851–57. doi:10.1093/bioinformatics/btt311.

Gibbs, J. Raphael, Marcel P. van der Brug, Dena G. Hernandez, Bryan J. Traynor, Michael A. Nalls, Shiao-Lin Lai, Sampath Arepalli, et al. 2010. Abundant Quantitative Trait Loci Exist for DNA Methylation and Gene Expression in Human Brain. Edited by Jonathan Flint. *PLoS Genetics* 6 (5): e1000952. doi:10.1371/journal.pgen.1000952.

Gilad, Yoav. 2012. Using Genomic Tools to Study Regulatory Evolution. *Methods in Molecular Biology (Clifton, N.J.)* 856: 335–61. doi:10.1007/978-1-61779-585-5_14.

Giuliani, Cristina, Elisabetta Cilli, Maria Giulia Bacalini, Chiara Pirazzini, Marco Sazzini, Giorgio Gruppioni, Claudio Franceschi, Paolo Garagnani, and Donata Luiselli. 2016. Inferring Chronological Age from DNA Methylation Patterns of

Human Teeth. *American Journal of Physical Anthropology* 159 (4): 585–95. doi:10.1002/ajpa.22921.

Glyn-Jones, S., A. J. R. Palmer, R. Agricola, A. J. Price, T. L. Vincent, H. Weinans, and A. J. Carr. 2015. Osteoarthritis. *The Lancet* 386 (9991): 376–87. doi:10.1016/S0140-6736(14)60802-3.

Gokhman, David, Eitan Lavi, Kay Prüfer, Mario F. Fraga, José A. Riancho, Janet Kelso, Svante Pääbo, Eran Meshorer, and Liran Carmel. 2014. Reconstructing the DNA Methylation Maps of the Neandertal and the Denisovan. *Science (New York, N.Y.)* 344 (6183): 523–27. doi:10.1126/science.1250368.

Goldring, Mary B., and Kenneth B. Marcu. 2012. Epigenomic and microRNA-Mediated Regulation in Cartilage Development, Homeostasis, and Osteoarthritis. *Trends in Molecular Medicine* 18 (2): 109–18. doi:10.1016/j.molmed.2011.11.005.

Gómez-Bañuelos, Eduardo, Rosa Elena Navarro-Hernández, Fernanda Corona-Meraz, Perla Monserrat Madrigal-Ruíz, Beatríz Teresita Martín-Marquez, Oscar Enrique Pizano-Martinez, Jorge Aguilar-Arreola, et al. 2015. Serum Leptin and Serum Leptin/Serum Leptin Receptor Ratio Imbalance in Obese Rheumatoid Arthritis Patients Positive for Anti-Cyclic Citrullinated Peptide Antibodies. *Arthritis Research & Therapy* 17 (November): 335. doi:10.1186/s13075-015-0850-8.

Gonzalo, Susana. 2010. Epigenetic Alterations in Aging. *Journal of Applied Physiology* 109 (2): 586–97. doi:10.1152/japplphysiol.00238.2010.

Goodman, Frances R. 2002. Limb Malformations and the Human HOX Genes. *American Journal of Medical Genetics* 112 (3): 256–65. doi:10.1002/ajmg.10776.

Grönniger, Elke, Barbara Weber, Oliver Heil, Nils Peters, Franz Stäb, Horst Wenck, Bernhard Korn, Marc Winnefeld, and Frank Lyko. 2010. Aging and Chronic Sun Exposure Cause Distinct Epigenetic Changes in Human Skin. *PLOS Genetics* 6 (5): e1000971. doi:10.1371/journal.pgen.1000971.

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. doi:10.1016/j.molcel.2012.10.016.

Haygood, Ralph, Olivier Fedrigo, Brian Hanson, Ken-Daigoro Yokoyama, and Gregory A Wray. 2007. Promoter Regions of Many Neural- and Nutrition-Related Genes Have Experienced Positive Selection during Human Evolution. *Nature Genetics* 39 (9): 1140–44. doi:10.1038/ng2104.

Henriksen, Marius, Mark W. Creaby, Hans Lund, Carsten Juhl, and Robin Christensen. 2014. Is There a Causal Link between Knee Loading and Knee Osteoarthritis

Progression? A Systematic Review and Meta-Analysis of Cohort Studies and Randomised Trials. *BMJ Open* 4 (7): e005368.

Hernandez, Dena G., Michael A. Nalls, J. Raphael Gibbs, Sampath Arepalli, Marcel van der Brug, Sean Chong, Matthew Moore, et al. 2011. Distinct DNA Methylation Changes Highly Correlated with Chronological Age in the Human Brain. *Human Molecular Genetics* 20 (6): 1164–72. doi:10.1093/hmg/ddq561.

Hernando-Herraez, Irene, Javier Prado-Martinez, Paras Garg, Marcos Fernandez-Callejo, Holger Heyn, Christina Hvilsom, Arcadi Navarro, Manel Esteller, Andrew J. Sharp, and Tomas Marques-Bonet. 2013. Dynamics of DNA Methylation in Recent Human and Great Ape Evolution. *PLOS Genet* 9 (9): e1003763. doi:10.1371/journal.pgen.1003763.

Heyn, Holger, Sebastian Moran, Irene Hernando-Herraez, Sergi Sayols, Antonio Gomez, Juan Sandoval, Dave Monk, et al. 2013. DNA Methylation Contributes to Natural Human Variation. *Genome Research* 23 (9): 1363–72. doi:10.1101/gr.154187.112.

Ho, Anthony D., Wolfgang Wagner, and Ulrich Mahlknecht. 2005. Stem Cells and Ageing. *EMBO Reports* 6 (Suppl 1): S35–38. doi:10.1038/sj.embor.7400436.

Hollander, W den, Y F M Ramos, S D Bos, N Bomer, R van der Breggen, N Lakenberg, W J de Dijcker, et al. 2014. Knee and Hip Articular Cartilage Have Distinct Epigenomic Landscapes: Implications for Future Cartilage Regeneration Approaches. *Annals of the Rheumatic Diseases* 73 (12): 2208–12. doi:10.1136/annrheumdis-2014-205980.

Horvath, Steve. 2013. DNA Methylation Age of Human Tissues and Cell Types. *Genome Biology* 14 (10): R115. doi:10.1186/gb-2013-14-10-r115.

Horvath, Steve, Wiebke Erhart, Mario Brosch, Ole Ammerpohl, Witigo von Schönfels, Markus Ahrens, Nils Heits, et al. 2014. Obesity Accelerates Epigenetic Aging of Human Liver. *Proceedings of the National Academy of Sciences* 111 (43): 15538–43. doi:10.1073/pnas.1412759111.

Horvath, Steve, Paolo Garagnani, Maria Giulia Bacalini, Chiara Pirazzini, Stefano Salvioli, Davide Gentilini, Anna Maria Di Blasio, et al. 2015. Accelerated Epigenetic Aging in Down Syndrome. *Aging Cell* 14 (3): 491–95. doi:10.1111/acel.12325.

Horvath, Steve, Peter Langfelder, Seung Kwak, Jeff Aaronson, Jim Rosinski, Thomas F. Vogt, Marika Eszes, et al. 2016. Huntington's Disease Accelerates Epigenetic Aging of Human Brain and Disrupts DNA Methylation Levels. *Aging (Albany NY)* 8 (7): 1485–1504. doi:10.18632/aging.101005.

Horvath, Steve, and Andrew J. Levine. 2015. HIV-1 Infection Accelerates Age According to the Epigenetic Clock. *The Journal of Infectious Diseases* 212 (10): 1563–73. doi:10.1093/infdis/jiv277.

Horvath, Steve, Vei Mah, Ake T. Lu, Jennifer S. Woo, Oi-Wa Choi, Anna J. Jasinska, José A. Riancho, et al. 2015. The Cerebellum Ages Slowly according to the Epigenetic Clock. *Aging (Albany NY)* 7 (5): 294–306.

Horvath, Steve, and Beate R. Ritz. 2015. Increased Epigenetic Age and Granulocyte Counts in the Blood of Parkinson's Disease Patients. *Aging (Albany NY)* 7 (12): 1130–42.

Horvath, Steve, Yafeng Zhang, Peter Langfelder, René S Kahn, Marco PM Boks, Kristel van Eijk, Leonard H van den Berg, and Roel A Ophoff. 2012. Aging Effects on DNA Methylation Modules in Human Brain and Blood Tissue. *Genome Biology* 13 (10): R97. doi:10.1186/gb-2012-13-10-r97.

Houseman, Eugene Andres, John Molitor, and Carmen J. Marsit. 2014. Reference-Free Cell Mixture Adjustments in Analysis of DNA Methylation Data. *Bioinformatics* 30 (10): 1431–39. doi:10.1093/bioinformatics/btu029.

Huber, Wolfgang, Vincent J. Carey, Robert Gentleman, Simon Anders, Marc Carlson, Benilton S. Carvalho, Hector Corrada Bravo, et al. 2015. Orchestrating High-Throughput Genomic Analysis with Bioconductor. *Nature Methods* 12 (2): 115–21. doi:10.1038/nmeth.3252.

Iliopoulos, Dimitrios, Konstantinos N. Malizos, Pagona Oikonomou, and Aspasia Tsezou. 2008. Integrative MicroRNA and Proteomic Approaches Identify Novel Osteoarthritis Genes and Their Collaborative Metabolic and Inflammatory Networks. Edited by Sotirios Koutsopoulos. *PLoS ONE* 3 (11): e3740. doi:10.1371/journal.pone.0003740.

Iwamoto, I., T. Fujino, and T. Douchi. 2004. The Leptin Receptor in Human Osteoblasts and the Direct Effect of Leptin on Bone Metabolism. *Gynecological Endocrinology: The Official Journal of the International Society of Gynecological Endocrinology* 19 (2): 97–104.

Jacob, Robert A., Denise M. Gretz, Peter C. Taylor, S. Jill James, Igor P. Pogribny, Barbara J. Miller, Susanne M. Henning, and Marian E. Swendseid. 1998. Moderate Folate Depletion Increases Plasma Homocysteine and Decreases Lymphocyte DNA Methylation in Postmenopausal Women. *The Journal of Nutrition* 128 (7): 1204–1212.

Jaffe, Andrew E, and Rafael A Irizarry. 2014. Accounting for Cellular Heterogeneity Is Critical in Epigenome-Wide Association Studies. *Genome Biology* 15 (2): R31. doi:10.1186/gb-2014-15-2-r31.

Jeffries, Matlock A., Madison Donica, Lyle W. Baker, Michael E. Stevenson, Anand C. Annan, Mary Beth Humphrey, Judith A. James, and Amr H. Sawalha. 2016. Genome-Wide DNA Methylation Study Identifies Significant Epigenomic Changes in Osteoarthritic Subchondral Bone and Similarity to Overlying Cartilage: GENOME-WIDE METHYLATION IN OA SUBCHONDRAL BONE. *Arthritis & Rheumatology* 68 (6): 1403–14. doi:10.1002/art.39555.

Jeffries, Matlock A., Madison Donica, Lyle W. Baker, Michael E. Stevenson, Anand C. Annan, Mary Beth Humphrey, Judith A. James, and Amr H. Sawalha. 2014. Genome-Wide DNA Methylation Study Identifies Significant Epigenomic Changes in Osteoarthritic Cartilage. *Arthritis & Rheumatology* 66 (10): 2804–15. doi:10.1002/art.38762.

Johnson, Victoria L., and David J. Hunter. 2014. The Epidemiology of Osteoarthritis. *Best Practice & Research Clinical Rheumatology* 28 (1): 5–15. doi:10.1016/j.berh.2014.01.004.

Jordan, Joanne M., Charles G. Helmick, Jordan B. Renner, Gheorghe Luta, Anca D. Dragomir, Janice Woodard, Fang Fang, et al. 2007. Prevalence of Knee Symptoms and Radiographic and Symptomatic Knee Osteoarthritis in African Americans and Caucasians: The Johnston County Osteoarthritis Project. *The Journal of Rheumatology* 34 (1): 172–80.

Karere, Genesio M., Jeremy P. Glenn, Shifra Birnbaum, David L. Rainwater, Michael C. Mahaney, John L. VandeBerg, and Laura A. Cox. 2013. Identification of Candidate Genes Encoding an LDL-C QTL in Baboons. *Journal of Lipid Research* 54 (7): 1776–1785.

Karere, Genesio M., Jeremy P. Glenn, John L. VandeBerg, and Laura A. Cox. 2010. Identification of Baboon microRNAs Expressed in Liver and Lymphocytes. *Journal of Biomedical Science* 17 (1): 1.

Karere, Genesio M., Jeremy P. Glenn, John L. VandeBerg, and Laura A. Cox. 2012. Differential microRNA Response to a High-Cholesterol, High-Fat Diet in Livers of Low and High LDL-C Baboons. BMC Genomics 13 (1): 1.

Kasaai, B., M.-H. Gaumond, and P. Moffatt. 2013. Regulation of the Bone-Restricted IFITM-like (Bril) Gene Transcription by Sp and Gli Family Members and CpG Methylation. *Journal of Biological Chemistry* 288 (19): 13278–94. doi:10.1074/jbc.M113.457010.

Kaushal, Akhilesh, Hongmei Zhang, Wilfried JJ Karmaus, and Julie SL Wang. 2015. Which Methods to Choose to Correct Cell Types in Genome-Scale Blood-Derived DNA Methylation Data? *BMC Bioinformatics* 16 (Suppl 15): P7. doi:10.1186/1471-2105-16-S15-P7.

Kearse, Matthew, Richard Moir, Amy Wilson, Steven Stones-Havas, Matthew Cheung, Shane Sturrock, Simon Buxton, et al. 2012. Geneious Basic: An Integrated and Extendable Desktop Software Platform for the Organization and Analysis of Sequence Data. *Bioinformatics (Oxford, England)* 28 (12): 1647–49. doi:10.1093/bioinformatics/bts199.

Kim, Sun Mie, Seok Hyun Kim, Jung Ryeol Lee, Byung Chul Jee, Seung-Yup Ku, Chang Suk Suh, Young Min Choi, Jung Gu Kim, and Shin Yong Moon. 2008. Association of Leptin Receptor Polymorphisms Lys109Arg and Gln223Arg with Serum Leptin Profile and Bone Mineral Density in Korean Women. *American Journal of Obstetrics and Gynecology* 198 (4): 421.e1-8. doi:10.1016/j.ajog.2007.10.799.

Kimber, Wendy, Frank Peelman, Xavier Prieur, Teresia Wangensteen, Stephen O'Rahilly, Jan Tavernier, and I. Sadaf Farooqi. 2008. Functional Characterization of Naturally Occurring Pathogenic Mutations in the Human Leptin Receptor. *Endocrinology* 149 (12): 6043–52. doi:10.1210/en.2008-0544.

King, M. C., and A. C. Wilson. 1975. Evolution at Two Levels in Humans and Chimpanzees. *Science* 188 (4184): 107–16. doi:10.1126/science.1090005.

Koch, Carmen M., Christoph V. Suschek, Qiong Lin, Simone Bork, Maria Goergens, Sylvia Joussen, Norbert Pallua, Anthony D. Ho, Martin Zenke, and Wolfgang Wagner. 2011. Specific Age-Associated DNA Methylation Changes in Human Dermal Fibroblasts. *PLOS ONE* 6 (2): e16679. doi:10.1371/journal.pone.0016679.

Koch, Carmen M., and Wolfgang Wagner. 2011. Epigenetic-Aging-Signature to Determine Age in Different Tissues. *Aging* 3 (10): 1018–27. doi:10.18632/aging.100395.

Kothapalli, Kumar S.D., Joshua C. Anthony, Bruce S. Pan, Andrea T. Hsieh, Peter W. Nathanielsz, and J. Thomas Brenna. 2007. Differential Cerebral Cortex Transcriptomes of Baboon Neonates Consuming Moderate and High Docosahexaenoic Acid Formulas. Edited by Schahram Akbarian. *PLoS ONE* 2 (4): e370. doi:10.1371/journal.pone.0000370.

Kovacs, Christopher S. 2000. Calcium Metabolism during Pregnancy and Lactation. In *Endotext*, edited by Leslie J. De Groot, George Chrousos, Kathleen Dungan, Kenneth R. Feingold, Ashley Grossman, Jerome M. Hershman, Christian Koch, et al. South Dartmouth (MA): MDText.com, Inc. http://www.ncbi.nlm.nih.gov/books/NBK279173/.

Kumar, Sudhir, Glen Stecher, and Koichiro Tamura. 2016. MEGA7: Molecular Evolutionary Genetics Analysis Version 7.0 for Bigger Datasets. *Molecular Biology and Evolution* 33 (7): 1870–74. doi:10.1093/molbev/msw054.

Kuyinu, Emmanuel L., Ganesh Narayanan, Lakshmi S. Nair, and Cato T. Laurencin. 2016. Animal Models of Osteoarthritis: Classification, Update, and Measurement of Outcomes. *Journal of Orthopaedic Surgery and Research* 11 (February). doi:10.1186/s13018-016-0346-5.

Lampropoulou-Adamidou, Kalliopi, Pavlos Lelovas, Eleftherios V. Karadimas, Chrysoula Liakou, Ioannis K. Triantafillopoulos, Ismene Dontas, and Nikolaos A. Papaioannou. 2014. Useful Animal Models for the Research of Osteoarthritis. *European Journal of Orthopaedic Surgery & Traumatology* 24 (3): 263–71. doi:10.1007/s00590-013-1205-2.

Lee, Hee Jun, Hoon Kim, Seung-Yup Ku, Young Min Choi, Jong Hak Kim, and Jung Gu Kim. 2014. Association between Polymorphisms in Leptin, Leptin Receptor, and β-Adrenergic Receptor Genes and Bone Mineral Density in Postmenopausal Korean Women. *Menopause (New York, N.Y.)* 21 (1): 67–73. doi:10.1097/GME.0b013e31829366ed.

Lee, Y. H., S.-C. Bae, J.-H. Kim, Y. H. Seo, S. J. Choi, J. D. Ji, and G. G. Song. 2015. Meta-Analysis of SLC22A4 and RUNX1 Polymorphisms : Associations with Rheumatoid Arthritis Susceptibility. *Zeitschrift Fur Rheumatologie* 74 (4): 351–58. doi:10.1007/s00393-014-1447-3.

Leek, J. T., and J. D. Storey. 2008. A General Framework for Multiple Testing Dependence. *Proceedings of the National Academy of Sciences* 105 (48): 18718–23. doi:10.1073/pnas.0808709105.

Leek, Jeffrey T., W. Evan Johnson, Hilary S. Parker, Andrew E. Jaffe, and John D. Storey. 2012. The Sva Package for Removing Batch Effects and Other Unwanted Variation in High-Throughput Experiments. *Bioinformatics* 28 (6): 882–83. doi:10.1093/bioinformatics/bts034.

Leek, Jeffrey T., and John D. Storey. 2007. Capturing Heterogeneity in Gene Expression Studies by Surrogate Variable Analysis. *PLoS Genetics* 3 (9): e161. doi:10.1371/journal.pgen.0030161.

Leigh, Steven R., and Brian T. Shea. 1995. Ontogeny and the Evolution of Adult Body Size Dimorphism in Apes. *American Journal of Primatology* 36 (1): 37–60. doi:10.1002/ajp.1350360104.

Levine, Morgan E., Ake T. Lu, David A. Bennett, and Steve Horvath. 2015. Epigenetic Age of the Pre-Frontal Cortex Is Associated with Neuritic Plaques, Amyloid Load, and Alzheimer's Disease Related Cognitive Functioning. *Aging (Albany NY)* 7 (12): 1198–1211.

Li, Dan, Yulan Zhao, Changxing Liu, Xiaona Chen, Yanting Qi, Yue Jiang, Chao Zou, et al. 2011. Analysis of MiR-195 and MiR-497 Expression, Regulation and Role in Breast Cancer. *Clinical Cancer Research: An Official Journal of the American*

Association for Cancer Research 17 (7): 1722–30. doi:10.1158/1078-0432.CCR-10-1800.

Lin, Qiong, Carola I. Weidner, Ivan G. Costa, Riccardo E. Marioni, Marcelo R. P. Ferreira, Ian J. Deary, and Wolfgang Wagner. 2016. DNA Methylation Levels at Individual Age-Associated CpG Sites Can Be Indicative for Life Expectancy. *Aging (Albany NY)* 8 (2): 394–401.

Lindskog, Cecilia, Martin Kuhlwilm, Armaity Davierwala, Ning Fu, Geeta Hegde, Mathias Uhlén, Sanjay Navani, Svante Pääbo, and Fredrik Pontén. 2014. Analysis of Candidate Genes for Lineage-Specific Expression Changes in Humans and Primates. *Journal of Proteome Research* 13 (8): 3596–3606.

Ling, B. M. T., N. Bharathy, T.-K. Chung, W. K. Kok, S. Li, Y. H. Tan, V. K. Rao, et al. 2012. Lysine Methyltransferase G9a Methylates the Transcription Factor MyoD and Regulates Skeletal Muscle Differentiation. *Proceedings of the National Academy of Sciences* 109 (3): 841–46. doi:10.1073/pnas.1111628109.

Lister, Ryan, Mattia Pelizzola, Robert H. Dowen, R. David Hawkins, Gary Hon, Julian Tonti-Filippini, Joseph R. Nery, et al. 2009. Human DNA Methylomes at Base Resolution Show Widespread Epigenomic Differences. *Nature* 462 (7271): 315–22. doi:10.1038/nature08514.

Liu, Lin, Lin Chen, Yingxin Xu, Rong Li, and Xiaohui Du. 2010. microRNA-195 Promotes Apoptosis and Suppresses Tumorigenicity of Human Colorectal Cancer Cells. *Biochemical and Biophysical Research Communications* 400 (2): 236–40. doi:10.1016/j.bbrc.2010.08.046.

Liu, Yun, Martin J Aryee, Leonid Padyukov, M Daniele Fallin, Espen Hesselberg, Arni Runarsson, Lovisa Reinius, et al. 2013. Epigenome-Wide Association Data Implicate DNA Methylation as an Intermediary of Genetic Risk in Rheumatoid Arthritis. *Nature Biotechnology* 31 (2): 142–47. doi:10.1038/nbt.2487.

Loeser, Richard F., and Najia Shakoor. 2003. Aging or Osteoarthritis: Which Is the Problem? *Rheumatic Diseases Clinics of North America* 29 (4): 653–73.

Lönnstedt, Ingrid, and Terry Speed. 2002. Replicated Microarray Data. *Statistica Sinica* 12 (1): 31–46.

Loughlin, John, and Louise N. Reynard. 2015. Osteoarthritis: Epigenetics of Articular Cartilage in Knee and Hip OA. *Nature Reviews Rheumatology* 11 (1): 6–7. doi:10.1038/nrrheum.2014.189.

Lowe, Donna, Steve Horvath, and Kenneth Raj. 2016. Epigenetic Clock Analyses of Cellular Senescence and Ageing. *Oncotarget* 7 (8): 8524–31. doi:10.18632/oncotarget.7383.

Luis, Daniel A. de, Rocio Aller, Olatz Izaola, Rosa Conde, and Jose Eiros Bouza. 2013. Lys656Asn Polymorphism of Leptin Receptor Gene Is Related with Leptin Changes after a High Monounsaturated Fat Diet in Obese Patients. *Journal of Investigative Medicine: The Official Publication of the American Federation for Clinical Research* 61 (2): 286–90. doi:10.2310/JIM.0b013e31827c2e4e.

Macrini, T.E., H.B. Coan, S.M. Levine, T. Lerma, C.D. Saks, D.J. Araujo, T.L. Bredbenner, R.D. Coutts, D.P. Nicolella, and L.M. Havill. 2013. Reproductive Status and Sex Show Strong Effects on Knee OA in a Baboon Model. *Osteoarthritis and Cartilage / OARS, Osteoarthritis Research Society* 21 (6): 839–48. doi:10.1016/j.joca.2013.03.003.

Maegawa, Shinji, George Hinkal, Hyun Soo Kim, Lanlan Shen, Li Zhang, Jiexin Zhang, Nianxiang Zhang, Shoudan Liang, Lawrence A. Donehower, and Jean-Pierre J. Issa. 2010. Widespread and Tissue Specific Age-Related DNA Methylation Changes in Mice. *Genome Research* 20 (3): 332–40. doi:10.1101/gr.096826.109.

Maksimovic, Jovana, Belinda Phipson, and Alicia Oshlack. 2016. A Cross-Package Bioconductor Workflow for Analysing Methylation Array Data. *F1000Research* 5 (June): 1281. doi:10.12688/f1000research.8839.1.

Manichaikul, Ani, Xin-Qun Wang, Wei Zhao, Mary K. Wojczynski, Kyle Siebenthall, John A. Stamatoyannopoulos, Danish Saleheen, et al. 2016. Genetic Association of Long-Chain Acyl-CoA Synthetase 1 Variants with Fasting Glucose, Diabetes, and Subclinical Atherosclerosis. *Journal of Lipid Research* 57 (3): 433–42. doi:10.1194/jlr.M064592.

Marciniak-Czochra, Anna, Thomas Stiehl, and Wolfgang Wagner. 2009. Modeling of Replicative Senescence in Hematopoietic Development. *Aging (Albany NY)* 1 (8): 723–32.

Marioni, Riccardo E., Sonia Shah, Allan F. McRae, Brian H. Chen, Elena Colicino, Sarah E. Harris, Jude Gibson, et al. 2015. DNA Methylation Age of Blood Predicts All-Cause Mortality in Later Life. *Genome Biology* 16: 25. doi:10.1186/s13059-015-0584-6.

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. doi:10.1093/ije/dyu277.

Martin, David I. K., Meromit Singer, Joseph Dhahbi, Guanxiong Mao, Lu Zhang, Gary P. Schroth, Lior Pachter, and Dario Boffelli. 2011. Phyloepigenomic Comparison of Great Apes Reveals a Correlation between Somatic and Germline Methylation States. *Genome Research* 21 (12): 2049–57. doi:10.1101/gr.122721.111.

Martínez, Alfonso, Antonio Valdivia, Dora Pascual-Salcedo, Alejandro Balsa, Benjamín Fernández-Gutiérrez, Emilio De la Concha, and Elena Urcelay. 2006. Role of SLC22A4, SLC22A5, and RUNX1 Genes in Rheumatoid Arthritis. *The Journal of Rheumatology* 33 (5): 842–46.

Martino, David J., Meri K. Tulic, Lavinia Gordon, Megan Hodder, Tara R. Richman, Jessica Metcalfe, Susan L. Prescott, and Richard Saffery. 2011. Evidence for Age-Related and Individual-Specific Changes in DNA Methylation Profile of Mononuclear Cells during Early Immune Development in Humans. *Epigenetics* 6 (9): 1085–94. doi:10.4161/epi.6.9.16401.

Massart, Renaud, Zsofia Nemoda, Matthew J. Suderman, Sheila Sutti, Angela M. Ruggiero, Amanda M. Dettmer, Stephen J. Suomi, and Moshe Szyf. 2016. Early Life Adversity Alters Normal Sex-Dependent Developmental Dynamics of DNA Methylation. *Development and Psychopathology*, September, 1–14. doi:10.1017/S0954579416000833.

Mazerolle, Marc J. 2016. *AICcmodavg: Model Selection and Multimodel Inference Based on (Q)AIC(c)* (version 2.1-0). https://cran.r-project.org/web/packages/AICcmodavg/index.html.

Mbalaviele, Gabriel, Chan Soo Shin, and Roberto Civitelli. 2006. Perspective: Cell–Cell Adhesion and Signaling Through Cadherins: Connecting Bone Cells in Their Microenvironment. *Journal of Bone and Mineral Research* 21 (12): 1821–27. doi:10.1359/jbmr.060811.

McCarthy, Davis J., and Gordon K. Smyth. 2009. Testing Significance Relative to a Fold-Change Threshold Is a TREAT. *Bioinformatics* 25 (6): 765–71. doi:10.1093/bioinformatics/btp053.

McCartney, Daniel L., Rosie M. Walker, Stewart W. Morris, Andrew M. McIntosh, David J. Porteous, and Kathryn L. Evans. 2016. Identification of Polymorphic and off-Target Probe Binding Sites on the Illumina Infinium MethylationEPIC BeadChip. *Genomics Data* 9 (September): 22–24. doi:10.1016/j.gdata.2016.05.012.

McHenry, Henry M., and Robert S. Corruccini. 1978. The Femur in Early Human Evolution. *American Journal of Physical Anthropology* 49 (4): 473–87. doi:10.1002/ajpa.1330490407.

Michels, Karin B, Alexandra M Binder, Sarah Dedeurwaerder, Charles B Epstein, John M Greally, Ivo Gut, E Andres Houseman, et al. 2013. Recommendations for the Design and Analysis of Epigenome-Wide Association Studies. *Nature Methods* 10 (10): 949–55. doi:10.1038/nmeth.2632.

Moazedi-Fuerst, Florentine C., Manuela Hofner, Gerald Gruber, Andreas Weinhaeusel, Martin H. Stradner, Hannes Angerer, Daniela Peischler, et al. 2014. Epigenetic

Differences in Human Cartilage between Mild and Severe OA. *Journal of Orthopaedic Research* 32 (12): 1636–45. doi:10.1002/jor.22722.

Molaro, Antoine, Emily Hodges, Fang Fang, Qiang Song, W. Richard McCombie, Gregory J. Hannon, and Andrew D. Smith. 2011. Sperm Methylation Profiles Reveal Features of Epigenetic Inheritance and Evolution in Primates. *Cell* 146 (6): 1029–41. doi:10.1016/j.cell.2011.08.016.

Morris, Tiffany J., and Stephan Beck. 2015. Analysis Pipelines and Packages for Infinium HumanMethylation450 BeadChip (450k) Data. *Methods*, (Epi)Genomics approaches and their applications, 72 (January): 3–8. doi:10.1016/j.ymeth.2014.08.011.

Mugatroyd, Chris, Yonghe Wu, Yvonne Bockmühl, and Dietmar Spengler. 2010. The Janus Face of DNA Methylation in Aging. *Aging (Albany NY)* 2 (2): 107–10.

Niu, Chi-Chien, Song-Shu Lin, Wen-Jer Chen, Shih-Jung Liu, Lih-Huei Chen, Chuen-Yung Yang, Chao-Jan Wang, Li-Jen Yuan, Po-Han Chen, and Hsiao-Yang Cheng. 2015. Benefits of Biphasic Calcium Phosphate Hybrid Scaffold-Driven Osteogenic Differentiation of Mesenchymal Stem Cells through Upregulated Leptin Receptor Expression. *Journal of Orthopaedic Surgery and Research* 10 (July): 111. doi:10.1186/s13018-015-0236-2.

Numata, Shusuke, Tianzhang Ye, Thomas M. Hyde, Xavier Guitart-Navarro, Ran Tao, Michael Wininger, Carlo Colantuoni, Daniel R. Weinberger, Joel E. Kleinman, and Barbara K. Lipska. 2012. DNA Methylation Signatures in Development and Aging of the Human Prefrontal Cortex. *American Journal of Human Genetics* 90 (2): 260–72. doi:10.1016/j.ajhg.2011.12.020.

Oates, N. A., J. van Vliet, D. L. Duffy, H. Y. Kroes, N. G. Martin, D. I. Boomsma, M. Campbell, M. G. Coulthard, E. Whitelaw, and S. Chong. 2006. Increased DNA Methylation at the AXIN1 Gene in a Monozygotic Twin from a Pair Discordant for a Caudal Duplication Anomaly. *American Journal of Human Genetics* 79 (1): 155–62.

Oberdoerffer, Philipp, and David A. Sinclair. 2007. The Role of Nuclear Architecture in Genomic Instability and Ageing. *Nature Reviews Molecular Cell Biology* 8 (9): 692–702. doi:10.1038/nrm2238.

O'Connor, Mary I. 2006. Osteoarthritis of the Hip and Knee: Sex and Gender Differences. *The Orthopedic Clinics of North America* 37 (4): 559–68. doi:10.1016/j.ocl.2006.09.004.

Okada, Yukinori, Di Wu, Gosia Trynka, Towfique Raj, Chikashi Terao, Katsunori Ikari, Yuta Kochi, et al. 2014. Genetics of Rheumatoid Arthritis Contributes to Biology and Drug Discovery. *Nature* 506 (7488): 376–81. doi:10.1038/nature12873.

Ong, Mei-Lyn, Peck Yean Tan, Julia L MacIsaac, Sarah M Mah, Jan Paul Buschdorf, Clara Y Cheong, Walter Stunkel, et al. 2014. Infinium Monkeys: Infinium 450K Array for the Cynomolgus Macaque (Macaca Fascicularis). *G3: Genes|Genomes|Genetics* 4 (7): 1227–34. doi:10.1534/g3.114.010967.

Orozco, Gisela, Elena Sánchez, Miguel A. González-Gay, Miguel A. López-Nevot, Belén Torres, Dora Pascual-Salcedo, Alejandro Balsa, et al. 2006. SLC22A4, RUNX1, and SUMO4 Polymorphisms Are Not Associated with Rheumatoid Arthritis: A Case-Control Study in a Spanish Population. *The Journal of Rheumatology* 33 (7): 1235–39.

Osteoarthritis (OA) | Arthritis | CDC. 2016. Accessed November 12. http://www.cdc.gov/arthritis/basics/osteoarthritis.htm.

Ostrer, H, Di Wilson, and Na Hanley. 2006. Human Embryo and Early Fetus Research. *Clinical Genetics* 70 (2): 98–107. doi:10.1111/j.1399-0004.2006.00640.x.

Pai, Athma A., Jordana T. Bell, John C. Marioni, Jonathan K. Pritchard, and Yoav Gilad. 2011. A Genome-Wide Study of DNA Methylation Patterns and Gene Expression Levels in Multiple Human and Chimpanzee Tissues. *PLOS Genetics* 7 (2): e1001316. doi:10.1371/journal.pgen.1001316.

Palacios, Daniela, and Pier Lorenzo Puri. 2006. The Epigenetic Network Regulating Muscle Development and Regeneration. *Journal of Cellular Physiology* 207 (1): 1–11. doi:10.1002/jcp.20489.

Pandorf, C. E., F. Haddad, C. Wright, P. W. Bodell, and K. M. Baldwin. 2009. Differential Epigenetic Modifications of Histones at the Myosin Heavy Chain Genes in Fast and Slow Skeletal Muscle Fibers and in Response to Muscle Unloading. *AJP: Cell Physiology* 297 (1): C6–16. doi:10.1152/ajpcell.00075.2009.

Paradis, Emmanuel, Julien Claude, and Korbinian Strimmer. 2004. APE: Analyses of Phylogenetics and Evolution in R Language. *Bioinformatics (Oxford, England)* 20 (2): 289–90.

Paten, Benedict, Javier Herrero, Kathryn Beal, Stephen Fitzgerald, and Ewan Birney. 2008. Enredo and Pecan: Genome-Wide Mammalian Consistency-Based Multiple Alignment with Paralogs. *Genome Research* 18 (11): 1814–28. doi:10.1101/gr.076554.108.

Paten, Benedict, Javier Herrero, Stephen Fitzgerald, Kathryn Beal, Paul Flicek, Ian Holmes, and Ewan Birney. 2008. Genome-Wide Nucleotide-Level Mammalian Ancestor Reconstruction. *Genome Research* 18 (11): 1829–43. doi:10.1101/gr.076521.108.

Perelman, Polina, Warren E. Johnson, Christian Roos, Hector N. Seuánez, Julie E. Horvath, Miguel A. M. Moreira, Bailey Kessing, et al. 2011. A Molecular Phylogeny of Living Primates. Edited by Jürgen Brosius. *PLoS Genetics* 7 (3): e1001342. doi:10.1371/journal.pgen.1001342.

Perna, Laura, Yan Zhang, Ute Mons, Bernd Holleczek, Kai-Uwe Saum, and Hermann Brenner. 2016. Epigenetic Age Acceleration Predicts Cancer, Cardiovascular, and All-Cause Mortality in a German Case Cohort. *Clinical Epigenetics* 8: 64. doi:10.1186/s13148-016-0228-z.

Peters, Timothy J., Michael J. Buckley, Aaron L. Statham, Ruth Pidsley, Katherine Samaras, Reginald V Lord, Susan J. Clark, and Peter L. Molloy. 2015. De Novo Identification of Differentially Methylated Regions in the Human Genome. *Epigenetics & Chromatin* 8: 6. doi:10.1186/1756-8935-8-6.

Petronis, Arturas, Irving I. Gottesman, Peixiang Kan, James L. Kennedy, Vincenzo S. Basile, Andrew D. Paterson, and Violeta Popendikyte. 2003. Monozygotic Twins Exhibit Numerous Epigenetic Differences: Clues to Twin Discordance? *Schizophrenia Bulletin* 29 (1): 169–78.

Phipson, Belinda, Stanley Lee, Ian J. Majewski, Warren S. Alexander, and Gordon K. Smyth. 2016. Robust Hyperparameter Estimation Protects against Hypervariable Genes and Improves Power to Detect Differential Expression. *The Annals of Applied Statistics* 10 (2): 946–63. doi:10.1214/16-AOAS920.

Prendergast, James GD, Harry Campbell, Nick Gilbert, Malcolm G Dunlop, Wendy A Bickmore, and Colin AM Semple. 2007. Chromatin Structure and Evolution in the Human Genome. *BMC Evolutionary Biology* 7 (1): 72. doi:10.1186/1471-2148-7-72.

Provencal, N., M. J. Suderman, C. Guillemin, R. Massart, A. Ruggiero, D. Wang, A. J. Bennett, et al. 2012. The Signature of Maternal Rearing in the Methylome in Rhesus Macaque Prefrontal Cortex and T Cells. *Journal of Neuroscience* 32 (44): 15626–42. doi:10.1523/JNEUROSCI.1470-12.2012.

Purcell, Jamie, Julia C. Oddo, Eric T. Wang, and J. Andrew Berglund. 2012. Combinatorial Mutagenesis of MBNL1 Zinc Fingers Elucidates Distinct Classes of Regulatory Events. *Molecular and Cellular Biology* 32 (20): 4155–67. doi:10.1128/MCB.00274-12.

Rakyan, Vardhman K, and Stephan Beck. 2006. Epigenetic Variation and Inheritance in Mammals. *Current Opinion in Genetics & Development* 16 (6): 573–77. doi:10.1016/j.gde.2006.09.002.

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. doi:10.1101/gr.103101.109.

Rakyan, Vardhman K., Thomas Hildmann, Karen L. Novik, Jörn Lewin, Jörg Tost, Antony V. Cox, T. Dan Andrews, et al. 2004. DNA Methylation Profiling of the Human Major Histocompatibility Complex: A Pilot Study for the Human Epigenome Project. *PLOS Biology* 2 (12): e405. doi:10.1371/journal.pbio.0020405.

Ralston, Stuart H., and André G. Uitterlinden. 2010. Genetics of Osteoporosis. *Endocrine Reviews* 31 (5): 629–62. doi:10.1210/er.2009-0044.

Ramos, Yolande F. M., Wouter den Hollander, Judith V. M. G. Bovée, Nils Bomer, Ruud van der Breggen, Nico Lakenberg, J. Christiaan Keurentjes, et al. 2014. Genes Involved in the Osteoarthritis Process Identified through Genome Wide Expression Analysis in Articular Cartilage; the RAAK Study. *PLoS ONE* 9 (7): 1–12. doi:10.1371/journal.pone.0103056.

Rampersaud, G. C., G. P. Kauwell, A. D. Hutson, J. J. Cerda, and L. B. Bailey. 2000. Genomic DNA Methylation Decreases in Response to Moderate Folate Depletion in Elderly Women. *The American Journal of Clinical Nutrition* 72 (4): 998–1003.

Répásy, Judit, Szilvia Bokor, Éva Erhardt, and Dénes Molnár. 2014. Association of Gln223 Arg Polymorphism of the Leptin Receptor Gene with Indicators of Energy Expenditure in Obese Children. *Nutrition (Burbank, Los Angeles County, Calif.)* 30 (7–8): 837–40. doi:10.1016/j.nut.2014.01.012.

Reynard, Louise N., Catherine Bui, Catherine M. Syddall, and John Loughlin. 2014. CpG Methylation Regulates Allelic Expression of GDF5 by Modulating Binding of SP1 and SP3 Repressor Proteins to the Osteoarthritis Susceptibility SNP rs143383. *Human Genetics* 133 (8): 1059–73. doi:10.1007/s00439-014-1447-z.

Ritchie, Matthew E., Belinda Phipson, Di Wu, Yifang Hu, Charity W. Law, Wei Shi, and Gordon K. Smyth. 2015. Limma Powers Differential Expression Analyses for RNA-Sequencing and Microarray Studies. *Nucleic Acids Research*, January, gkv007. doi:10.1093/nar/gkv007.

Rivadeneira, Fernando, Unnur Styrkársdottir, Karol Estrada, Bjarni V Halldórsson, Yi-Hsiang Hsu, J Brent Richards, M Carola Zillikens, et al. 2009. Twenty Bone-Mineral-Density Loci Identified by Large-Scale Meta-Analysis of Genome-Wide Association Studies. *Nature Genetics* 41 (11): 1199–1206. doi:10.1038/ng.446.

Rodríguez-Rodero, Sandra, Juan Luis Fernández-Morera, Agustin F. Fernandez, Edelmiro Menendez-Torre, and Mario F. Fraga. 2010. Epigenetic Regulation of Aging. *Discovery Medicine* 10 (52): 225–33.

Rogers, Jeffrey, and Richard A. Gibbs. 2014. Comparative Primate Genomics: Emerging Patterns of Genome Content and Dynamics. *Nature Reviews Genetics* 15 (5): 347–59. doi:10.1038/nrg3707.

Romero, Irene Gallego, Ilya Ruvinsky, and Yoav Gilad. 2012. Comparative Studies of Gene Expression and the Evolution of Gene Regulation. *Nature Reviews Genetics* 13 (7): 505–16. doi:10.1038/nrg3229.

Rossignol, M, A Leclerc, F Allaert, S Rozenberg, J Valat, B Avouac, P Coste, E Litvak, and P Hilliquin. 2005. Primary Osteoarthritis of Hip, Knee, and Hand in Relation to Occupational Exposure. *Occupational and Environmental Medicine* 62 (11): 772–77. doi:10.1136/oem.2005.020057.

Rugg-Gunn, Peter J., Anne C. Ferguson-Smith, and Roger A. Pedersen. 2005. Human Embryonic Stem Cells as a Model for Studying Epigenetic Regulation during Early Development. *Cell Cycle* 4 (10): 1323–1326.

Rushton, Michael D., Louise N. Reynard, Matt J. Barter, Ramsay Refaie, Kenneth S. Rankin, David A. Young, and John Loughlin. 2014a. Characterization of the Cartilage DNA Methylome in Knee and Hip Osteoarthritis: Methylation Profile of OA Cartilage. *Arthritis & Rheumatology* 66 (9): 2450–60. doi:10.1002/art.38713.

———. 2014b. Characterization of the Cartilage DNA Methylome in Knee and Hip Osteoarthritis: Methylation Profile of OA Cartilage. *Arthritis & Rheumatology* 66 (9): 2450–60. doi:10.1002/art.38713.

Ryu, Hyo Sub, Young Hyun Joo, Sun Ok Kim, Kyoung Chan Park, and Sang Woong Youn. 2008. Influence of Age and Regional Differences on Skin Elasticity as Measured by the Cutometer®. *Skin Research and Technology* 14 (3): 354–58. doi:10.1111/j.1600-0846.2008.00302.x.

Saito, Taku, Atsushi Fukai, Akihiko Mabuchi, Toshiyuki Ikeda, Fumiko Yano, Shinsuke Ohba, Nao Nishida, et al. 2010. Transcriptional Regulation of Endochondral Ossification by HIF-2α during Skeletal Growth and Osteoarthritis Development. *Nature Medicine* 16 (6): 678–86. doi:10.1038/nm.2146.

Satterthwaite, F. E. 1946. An Approximate Distribution of Estimates of Variance Components. *Biometrics* 2 (6): 110–14.

Schellenberg, Anne, Qiong Lin, Herdit Schüler, Carmen M. Koch, Sylvia Joussen, Bernd Denecke, Gudrun Walenda, et al. 2011. Replicative Senescence of Mesenchymal Stem Cells Causes DNA-Methylation Changes Which Correlate with Repressive Histone Marks. *Aging (Albany NY)* 3 (9): 873–88.

Schlicker, Andreas, Francisco S. Domingues, Jörg Rahnenführer, and Thomas Lengauer. 2006. A New Measure for Functional Similarity of Gene Products Based on Gene Ontology. *BMC Bioinformatics* 7: 302. doi:10.1186/1471-2105-7-302.

Schultz, Adolph H. 1930. The Skeleton of the Trunk and Limbs of Higher Primates. *Human Biology* 2 (3): 303–438.

Schultz, Adolph H. 1937. Proportions, Variability and Asymmetries of the Long Bones of the Limbs and the Clavicles in Man and Apes. Human Biology 9 (3): 281–328.

Sharif, Jafar, Takaho A. Endo, Tetsuro Toyoda, and Haruhiko Koseki. 2010. Divergence of CpG Island Promoters: A Consequence or Cause of Evolution? *Development, Growth & Differentiation* 52 (6): 545–54. doi:10.1111/j.1440-169X.2010.01193.x.

Shelnutt, Karla P., Gail P. A. Kauwell, Jesse F. Gregory III, David R. Maneval, Eoin P. Quinlivan, Douglas W. Theriaque, George N. Henderson, and Lynn B. Bailey. 2004. Methylenetetrahydrofolate Reductase 677C→T Polymorphism Affects DNA Methylation in Response to Controlled Folate Intake in Young Women. *The Journal of Nutritional Biochemistry* 15 (9): 554–60. doi:10.1016/j.jnutbio.2004.04.003.

Simon, H. 1999. T-Box Genes and the Formation of Vertebrate Forelimb- and Hindlimb Specific Pattern. *Cell and Tissue Research* 296 (1): 57–66.

Simopoulou, T., K. N. Malizos, D. Iliopoulos, N. Stefanou, L. Papatheodorou, M. Ioannou, and A. Tsezou. 2007. Differential Expression of Leptin and Leptin's Receptor Isoform (Ob-Rb) mRNA between Advanced and Minimally Affected Osteoarthritic Cartilage; Effect on Cartilage Metabolism. *Osteoarthritis and Cartilage* 15 (8): 872–83. doi:10.1016/j.joca.2007.01.018.

Singer, Meromit, Idit Kosti, Lior Pachter, and Yael Mandel-Gutfreund. 2015. A Diverse Epigenetic Landscape at Human Exons with Implication for Expression. *Nucleic Acids Research*, March, gkv153. doi:10.1093/nar/gkv153.

Slieker, Roderick C., Steffan D. Bos, Jelle J. Goeman, Judith VMG Bovée, Rudolf P. Talens, Ruud van der Breggen, H. Eka D. Suchiman, et al. 2013. Identification and Systematic Annotation of Tissue-Specific Differentially Methylated Regions Using the Illumina 450k Array. *Epigenetics & Chromatin* 6: 26. doi:10.1186/1756-8935-6-26.

Smith, Rick W. A., Cara Monroe, and Deborah A. Bolnick. 2015. Detection of Cytosine Methylation in Ancient DNA from Five Native American Populations Using Bisulfite Sequencing. *PLoS ONE* 10 (5). doi:10.1371/journal.pone.0125344.

Smyth, Gordon K. 2004. Linear Models and Empirical Bayes Methods for Assessing Differential Expression in Microarray Experiments. *Statistical Applications in Genetics and Molecular Biology* 3: Article3. doi:10.2202/1544-6115.1027.

Smyth, Gordon K., Joëlle Michaud, and Hamish S. Scott. 2005. Use of within-Array Replicate Spots for Assessing Differential Expression in Microarray Experiments. *Bioinformatics* 21 (9): 2067–75. doi:10.1093/bioinformatics/bti270.

Spiers, Helen, Eilis Hannon, Leonard C. Schalkwyk, Rebecca Smith, Chloe C. Y. Wong, Michael C. O'Donovan, Nicholas J. Bray, and Jonathan Mill. 2015. Methylomic Trajectories across Human Fetal Brain Development. *Genome Research*, February, gr.180273.114. doi:10.1101/gr.180273.114.

Stein, Gary S., Jane B. Lian, Andre J. van Wijnen, Janet L. Stein, Martin Montecino, Amjad Javed, Sayyed K. Zaidi, Daniel W. Young, Je-Yong Choi, and Shirwin M. Pockwinse. 2004. Runx2 Control of Organization, Assembly and Activity of the Regulatory Machinery for Skeletal Gene Expression. *Oncogene* 23 (24): 4315–29. doi:10.1038/sj.onc.1207676.

Supek, Fran, Matko Bošnjak, Nives Škunca, and Tomislav Šmuc. 2011. REVIGO Summarizes and Visualizes Long Lists of Gene Ontology Terms. *PLOS ONE* 6 (7): e21800. doi:10.1371/journal.pone.0021800.

Suter, Catherine M., David I. K. Martin, and Robyn L. Ward. 2004. Germline Epimutation of MLH1 in Individuals with Multiple Cancers. *Nature Genetics* 36 (5): 497–501. doi:10.1038/ng1342.

Suzuki, Miho M., and Adrian Bird. 2008. DNA Methylation Landscapes: Provocative Insights from Epigenomics. *Nature Reviews Genetics* 9 (6): 465–76. doi:10.1038/nrg2341.

Takata, Yoichiro, Hiroshi Inoue, Aya Sato, Kazue Tsugawa, Katsutoshi Miyatake, Daisuke Hamada, Fumio Shinomiya, et al. 2008. Replication of Reported Genetic Associations of PADI4, FCRL3, SLC22A4 and RUNX1 Genes with Rheumatoid Arthritis: Results of an Independent Japanese Population and Evidence from Meta-Analysis of East Asian Studies. *Journal of Human Genetics* 53 (2): 163–73. doi:10.1007/s10038-007-0232-4.

Tam, Elisa M. S., Fiona W. P. Yu, Vivian W. Y. Hung, Zhen Liu, King Lok Liu, Bobby K. W. Ng, Simon K. M. Lee, Yong Qiu, Jack C. Y. Cheng, and Tsz-Ping Lam. 2014. Are Volumetric Bone Mineral Density and Bone Micro-Architecture Associated with Leptin and Soluble Leptin Receptor Levels in Adolescent Idiopathic Scoliosis?--A Case-Control Study. *PloS One* 9 (2): e87939. doi:10.1371/journal.pone.0087939.

Terzidis, Ioannis, Trifon Totlis, Efthymia Papathanasiou, Aristotelis Sideridis, Konstantinos Vlasis, and Konstantinos Natsis. 2012. Gender and Side-to-Side Differences of Femoral Condyles Morphology: Osteometric Data from 360 Caucasian Dried Femori. *Anatomy Research International* 2012 (August): e679658. doi:10.1155/2012/679658.

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. doi:10.1101/gr.103606.109.

Therneau, Terry M. 2015. *Coxme: Mixed Effects Cox Models* (version 2.2-5). https://cran.r-project.org/web/packages/coxme/index.html.

Therneau, Terry M., Schaid Daniel, Jason Sinnwell, and Elizabeth Atkinson. 2015. *kinship2: Pedigree Functions* (version 1.6.4). https://cran.r-project.org/web/packages/kinship2/index.html.

Thompson, Deborah L., Karen D. Lum, Sean C. Nygaard, Rolf E. Kuestner, Katherine A. Kelly, Jeffrey M. Gimble, and Emma E. Moore. 1998. The Derivation and Characterization of Stromal Cell Lines from the Bone Marrow of p53−/− Mice: New Insights into Osteoblast and Adipocyte Differentiation. *Journal of Bone and Mineral Research* 13 (2): 195–204. doi:10.1359/jbmr.1998.13.2.195.

Thompson, Reid F., Gil Atzmon, Ciprian Gheorghe, Hong Qian Liang, Christina Lowes, John M. Greally, and Nir Barzilai. 2010. Tissue-Specific Dysregulation of DNA Methylation in Aging. *Aging Cell* 9 (4): 506–18. doi:10.1111/j.1474-9726.2010.00577.x.

Triche, T. J., D. J. Weisenberger, D. Van Den Berg, P. W. Laird, and K. D. Siegmund. 2013. Low-Level Processing of Illumina Infinium DNA Methylation BeadArrays. *Nucleic Acids Research* 41 (7): e90–e90. doi:10.1093/nar/gkt090.

Tung, Jenny, Xiang Zhou, Susan C. Alberts, Matthew Stephens, and Yoav Gilad. 2015. The Genetic Architecture of Gene Expression Levels in Wild Baboons. *eLife* 4 (February): e04729. doi:10.7554/eLife.04729.

Uebbing, Severin, Axel Künstner, Hannu Mäkinen, Niclas Backström, Paulina Bolivar, Reto Burri, Ludovic Dutoit, et al. 2016. Divergence in Gene Expression within and between Two Closely Related Flycatcher Species. *Molecular Ecology* 25 (9): 2015–28. doi:10.1111/mec.13596.

Väänänen, H. K., and P. L. Härkönen. 1996. Estrogen and Bone Metabolism. *Maturitas* 23 Suppl (May): S65-69.

Van Neste, Dominique, and Desmond J Tobin. 2004. Hair Cycle and Hair Pigmentation: Dynamic Interactions and Changes Associated with Aging. *Micron*, Morphological and functional changes in aging skin, 35 (3): 193–200. doi:10.1016/j.micron.2003.11.006.

Vanyushin, B. F., L. E. Nemirovsky, V. V. Klimenko, V. K. Vasiliev, and A. N. Belozersky. 1973. The 5-Methylcytosine in DNA of Rats. Tissue and Age

Specificity and the Changes Induced by Hydrocortisone and Other Agents. *Gerontologia* 19 (3): 138–52.

Wagner, Wolfgang, Simone Bork, Günther Lepperdinger, Sylvia Joussen, Nan Ma, Dirk Strunk, and Carmen Koch. 2010. How to Track Cellular Aging of Mesenchymal Stromal Cells? *Aging (Albany NY)* 2 (4): 224–30.

Walker, Richard F., Jia Sophie Liu, Brock A. Peters, Beate R. Ritz, Timothy Wu, Roel A. Ophoff, and Steve Horvath. 2015. Epigenetic Age Analysis of Children Who Seem to Evade Aging. *Aging (Albany NY)* 7 (5): 334–39.

Wand, M.P., and M.C. Jones. 1994. *Kernel Smoothing*. Chapman and Hall. https://www.crcpress.com/Kernel-Smoothing/Wand-Jones/p/book/9780412552700.

Wang, K., D. Shi, P. Zhu, J. Dai, L. Zhu, H. Zhu, Y. Lv, B. Zhao, and Q. Jiang. 2010. Association of a Single Nucleotide Polymorphism in Tbx4 with Developmental Dysplasia of the Hip: A Case-Control Study. *Osteoarthritis and Cartilage* 18 (12): 1592–95. doi:10.1016/j.joca.2010.09.008.

Wang, Qingju, Ali Ghasem-Zadeh, Xiao-Fang Wang, Sandra Iuliano-Burns, and Ego Seeman. 2011. Trabecular Bone of Growth Plate Origin Influences Both Trabecular and Cortical Morphology in Adulthood. *Journal of Bone and Mineral Research* 26 (7): 1577–83. doi:10.1002/jbmr.360.

Warner, Lisa R., Courtney C. Babbitt, Alex E. Primus, Tonya F. Severson, Ralph Haygood, and Gregory A. Wray. 2009. Functional Consequences of Genetic Variation in Primates on Tyrosine Hydroxylase (TH) Expression in Vitro. *Brain Research* 1288 (September): 1–8. doi:10.1016/j.brainres.2009.06.086.

Weber, Michael, Ines Hellmann, Michael B. Stadler, Liliana Ramos, Svante Pääbo, Michael Rebhan, and Dirk Schübeler. 2007. Distribution, Silencing Potential and Evolutionary Impact of Promoter DNA Methylation in the Human Genome. *Nature Genetics* 39 (4): 457–66. doi:10.1038/ng1990.

Wei, Chuankui, Qifeng Luo, Xiaoguo Sun, Dengfeng Li, Hongming Song, Xiaoyu Li, Jialu Song, Kaiyao Hua, and Lin Fang. 2015. microRNA-497 Induces Cell Apoptosis by Negatively Regulating Bcl-2 Protein Expression at the Posttranscriptional Level in Human Breast Cancer. *International Journal of Clinical and Experimental Pathology* 8 (7): 7729–39.

Weksberg, Rosanna, Cheryl Shuman, Oana Caluseriu, Adam C. Smith, Yan-Ling Fei, Joy Nishikawa, Tracy L. Stockley, et al. 2002. Discordant KCNQ1OT1 Imprinting in Sets of Monozygotic Twins Discordant for Beckwith-Wiedemann Syndrome. *Human Molecular Genetics* 11 (11): 1317–25.

White, Tim, and Pieter Folkens. 2000. *Human Osteology*. 2nd ed. San Diego: Academic Press.

Whitehead, Andrew, and Douglas L. Crawford. 2006. Neutral and Adaptive Variation in Gene Expression. *Proceedings of the National Academy of Sciences of the United States of America* 103 (14): 5425–30. doi:10.1073/pnas.0507648103.

WHO | Chronic Rheumatic Conditions. 2016. *WHO*. Accessed November 12. http://www.who.int/chp/topics/rheumatic/en/.

Wilson, V. L., R. A. Smith, S. Ma, and R. G. Cutler. 1987. Genomic 5-Methyldeoxycytidine Decreases with Age. *Journal of Biological Chemistry* 262 (21): 9948–51.

Xiang, Ping, Chaoyu Lo, Bob Argiropoulos, C. Benjamin Lai, Arefeh Rouhi, Suzan Imren, Xiaoyan Jiang, Dixie Mager, and R. Keith Humphries. 2010. Identification of E74-like Factor 1 (ELF1) as a Transcriptional Regulator of the Hox Cofactor MEIS1. *Experimental Hematology* 38 (9): 798–798, 808.e1-2. doi:10.1016/j.exphem.2010.06.006.

Xin, Z., H. Soejima, K. Higashimoto, H. Yatsuki, X. Zhu, Y. Satoh, Z. Masaki, et al. 2000. A Novel Imprinted Gene, KCNQ1DN, within the WT2 Critical Region of Human Chromosome 11p15.5 and Its Reduced Expression in Wilms' Tumors. *Journal of Biochemistry* 128 (5): 847–53.

Ye, Xing L., and Chun F. Lu. 2013. Association of Polymorphisms in the Leptin and Leptin Receptor Genes with Inflammatory Mediators in Patients with Osteoporosis. *Endocrine* 44 (2): 481–88. doi:10.1007/s12020-013-9899-9.

Young, Matthew D., Matthew J. Wakefield, Gordon K. Smyth, and Alicia Oshlack. 2010. Gene Ontology Analysis for RNA-Seq: Accounting for Selection Bias. *Genome Biology* 11: R14. doi:10.1186/gb-2010-11-2-r14.

Zaghlool, Shaza B., Mashael Al-Shafai, Wadha A. Al Muftah, Pankaj Kumar, Mario Falchi, and Karsten Suhre. 2015. Association of DNA Methylation with Age, Gender, and Smoking in an Arab Population. *Clinical Epigenetics* 7: 6. doi:10.1186/s13148-014-0040-6.

Zamli, Zaitunnatakhin, Kate Robson Brown, John F. Tarlton, Mike A. Adams, Georgina E. Torlot, Charlie Cartwright, William A. Cook, Kristiina Vassilevskaja, and Mohammed Sharif. 2014. Subchondral Bone Plate Thickening Precedes Chondrocyte Apoptosis and Cartilage Degradation in Spontaneous Animal Models of Osteoarthritis. *BioMed Research International* 2014: 1–10. doi:10.1155/2014/606870.

Zeng, Jia, Genevieve Konopka, Brendan G. Hunt, Todd M. Preuss, Dan Geschwind, and Soojin V. Yi. 2012. Divergent Whole-Genome Methylation Maps of Human and

Chimpanzee Brains Reveal Epigenetic Basis of Human Regulatory Evolution. *The American Journal of Human Genetics* 91 (3): 455–65. doi:10.1016/j.ajhg.2012.07.024.

Zhang, Bo, Yan Zhou, Nan Lin, Rebecca F. Lowdon, Chibo Hong, Raman P. Nagarajan, Jeffrey B. Cheng, et al. 2013. Functional DNA Methylation Differences between Tissues, Cell Types, and across Individuals Discovered Using the M&M Algorithm. *Genome Research* 23 (9): 1522–40. doi:10.1101/gr.156539.113.

Zwetsloot, K. A., M. J. Laye, and F. W. Booth. 2009. Novel Epigenetic Regulation of Skeletal Muscle Myosin Heavy Chain Genes. Focus on "Differential Epigenetic Modifications of Histones at the Myosin Heavy Chain Genes in Fast and Slow Skeletal Muscle Fibers and in Response to Muscle Unloading." *AJP: Cell Physiology* 297 (1): C1–3. doi:10.1152/ajpcell.00176.2009.

APPENDIX A

NONHUMAN PRIMATE ILLUMINA INFINIUM METHYLATION ARRAY PROBE FILES

Consult the attached zipped file that contains detailed information on whether each probe on the 450K and EPIC arrays map to specific nonhuman primate genomes, and if so, how well each probe and the mapped region align and whether associated gene symbols for humans and nonhuman primates match. Specifically, the zipped file includes:

- a probe annotation file for baboons using the 450K array (Baboon_450K_Probe_Filter.csv)
- a probe annotation file for baboons using the EPIC array (Baboon_EPIC_Probe_Filter.csv)
- a probe annotation file for chimpanzees using the EPIC array (Chimpanzee_EPIC_Probe_Filter.csv)
- a probe annotation file for macaques using the EPIC array (Macaque_EPIC_Probe_Filter.csv)
- a probe annotation file for marmosets using the EPIC array (Marmoset_EPIC_Probe_Filter.csv)
- a probe annotation file for vervets using the EPIC array (Vervet_EPIC_Probe_Filter.csv)
- descriptions for the column headers in each probe annotation file (Probe_Filter_Key.csv)

The CSV files within the attached zipped file can be opened in Microsoft Excel, Notepad, or any other text editor.

Probe_Annotation_Filters.csv

APPENDIX B

NORMALIZED AND FILTERED 450K ARRAY METHYLATION VALUES FOR THE BABOON OSTEOARTHRITIS STUDY

Consult the attached zipped file that contains the normalized and filtered 450K array β-values (450K_Baboon_OA_Beta.csv) and M-values (450K_Baboon_OA_M.csv) for each sample and each retained probe that were used in the 450K array baboon osteoarthritis study. Each CSV can be opened in Microsoft Excel, Notepad, or any other text editor.

450K_Baboon_OA.zip

APPENDIX C

BABOON SAMPLE SET FOR 450K ARRAY OSTEOARTHRITIS STUDY

Consult the attached Excel file which can be opened in Microsoft Excel. Table outlines each sample's animal identification number (Animal ID), tissue type (Tissue Type), classification as healthy or having knee osteoarthritis (Disease Status), sex (Sex), and age in years (Age), as well as whether the tissue sample came from the right or left femur (Side) and the medial or lateral condyle (Condyle), the concentration of DNA extracted using a Nanodrop (Nanodrop) and a Qubit (Qubit), the amount of DNA run on the 450K array (DNA (ng)), and the identification number of the beadchip (Array ID) and array (Position) that each sample was run on.

450K_Baboon_OA_Samples.xlsx

APPENDIX D

GENOMIC DISTRIBUTION OF 450K ARRAY PROBES RETAINED FOR BABOONS

Consult the attached Excel file which can be opened in Microsoft Excel. Details on the gene associations, genomic locations, and proximity to CpG islands of all probes that successfully mapped to the baboon genome with e-values less than e^{-10}, had only unique BLAST hits, and targeted a CpG site (Total Mapped Probes), probes that fit the alignment filter criteria (Alignment Filter Probes), and probes that fit the gene symbol filter criteria (Gene Symbol Filter Probes). Number of Genes indicates the number of unique gene symbols associated with probes. Probes Per Gene indicates the average number of probes that target each associated gene. For genomic locations, TSS 200 and TSS 1500 indicate the transcription start site areas between the start of the gene and 200bp upstream or 1500bp upstream respectively, 5'UTR and 3'UTR indicate the untranslated regions of genes, 1st Exon indicates the first exon within a gene, and Gene Body indicates any area within the exons and introns of a gene. For proximity to CpG islands, Island indicates areas within CpG islands, North Shelf indicates areas 2-4kb upstream from a CpG island, North Shore indicates areas up to 2kb upstream from a CpG island, South Shelf indicates areas 2-4kb downstream from a CpG island, South Shore indicates areas up to 2kb downstream from a CpG island, and Open Sea indicates isolated CpG site in the genome.

450K_Baboon_ProbeDist.xlsx

APPENDIX E

ALIGNMENT PARAMETER CORRELATIONS OF 450K ARRAY PROBES RETAINED FOR BABOON OSTEOARTHRITIS STUDY

Consult the attached Excel file which can be opened in Microsoft Excel. Results of Spearman correlation tests between 450K array probe detection p-values and alignment quality parameters. These parameters included the alignment bitscores, e-values, and percent identity. For all probes that successfully mapped to the baboon genome with e-values less than e-10, had only unique BLAST hits, and targeted a CpG site (Total Mapped Probes), probes that fit the alignment filter criteria (Alignment Filter Probes), and probes that fit the gene symbol filter criteria (Gene Symbol Filter Probes), significant negative correlations were identified for alignment bitscores and percent identities, and significant positive correlations were identified for alignment e-values.

450K_Baboon_OA_ProbeCorr.xlsx

APPENDIX F

NORMALIZED AND FILTERED EPIC ARRAY METHYLATION VALUES FOR THE BABOON OSTEOARTHRITIS STUDY

Consult the attached zipped file that contains the normalized and filtered EPIC array β-values (EPIC_Baboon_OA_Beta.csv) and M-values (EPIC_Baboon_OA_M.csv) for each sample and each retained probe that were used in the EPIC array baboon osteoarthritis study. Each CSV can be opened in Microsoft Excel, Notepad, or any other text editor.

EPIC_Baboon_OA.zip

APPENDIX G

BABOON SAMPLE SET FOR EPIC ARRAY OSTEOARTHRITIS STUDY

Consult the attached Excel file which can be opened in Microsoft Excel. Table outlines each sample's animal identification number (Animal ID), species (Species), tissue type (Tissue Type), classification as healthy or having knee osteoarthritis (Disease Status), sex (Sex), age in years (Age), and adult steady state weight in kilograms (Weight), as well as the identification number of the beadchip (Array ID) and array (Position) that each sample was run on.

EPIC_Baboon_OA_Samples.xlsx

APPENDIX H

GENOMIC DISTRIBUTION OF EPIC ARRAY PROBES RETAINED FOR BABOONS

Consult the attached Excel file which can be opened in Microsoft Excel. Details on the gene associations, genomic locations, and proximity to CpG islands (based on the human genome hg19) of all probes that successfully mapped to the baboon genome with e-values < e-10, had only unique BLAST hits, and targeted a CpG site (Total Mapped Probes) and probes that fit the alignment filter criteria (Alignment Filter Probes). Number of Genes indicates the number of unique gene symbols associated with probes. Probes Per Gene indicates the average number of probes that target each associated gene. For genomic locations, TSS 200 and TSS 1500 indicate the transcription start site areas between the start of the gene and 200bp upstream or 1500bp upstream respectively, 5'UTR and 3'UTR indicate the untranslated regions of genes, 1st Exon indicates the first exon within a gene, ExonBnd indicates the gene body exons, and Gene Body indicates any area within the exons and introns of a gene. For proximity to CpG islands, Island indicates areas within CpG islands, North Shelf indicates areas 2-4kb upstream from a CpG island, North Shore indicates areas up to 2kb upstream from a CpG island, South Shelf indicates areas 2-4kb downstream from a CpG island, South Shore indicates areas up to 2kb downstream from a CpG island, and Open Sea indicates isolated CpG site in the genome.

EPIC_Baboon_ProbeDist.xlsx

APPENDIX I

ALIGNMENT PARAMETER CORRELATIONS OF EPIC ARRAY PROBES RETAINED FOR BABOON OSTEOARTHRITIS STUDY

Consult the attached Excel file which can be opened in Microsoft Excel. Results of Spearman correlation tests between EPIC array probe detection p-values and alignment quality parameters. These parameters included the alignment bitscores, e-values, and percent identity. For the probes that fit the alignment filter criteria, significant negative correlations were identified for alignment bitscores and percent identities, and significant positive correlations were identified for alignment e-values.

EPIC_Baboon_OA_ProbeCorr.xlsx

APPENDIX J

GENOMIC DISTRIBUTION OF SIGNIFICANT DMPS IDENTIFIED IN THE EPIC ARRAY BABOON OSTEOARTHRITIS STUDY

Consult the attached Excel file which can be opened in Microsoft Excel. Table showing the number of significant DMPs between comparative groups when only accounting for kinship (Kinship) and when accounting for kinship and using a $\Delta\beta \geq 0.1$ threshold ($\Delta\beta \geq 0.1$) and the total number of loci examined that occupy different genomic regions (based on the human genome hg19). Number of Genes indicates the number of unique gene symbols associated with DMPs. Probes Per Gene indicates the average number of DMPs within each associated gene. For genomic locations, TSS200 and TSS1500 indicate the transcription start site areas between the start of the gene and 200bp upstream or 1500bp upstream respectively, 5'UTR and 3'UTR indicate the untranslated regions of genes, 1st Exon indicates the first exon within a gene, ExonBnd indicates the gene body exons, and Gene Body indicates any area within the exons and introns of a gene. For proximity to CpG islands, Island indicates areas within CpG islands, North Shelf indicates areas 2-4kb upstream from a CpG island, North Shore indicates areas up to 2kb upstream from a CpG island, South Shelf indicates areas 2-4kb downstream from a CpG island, South Shore indicates areas up to 2kb downstream from a CpG island, and Open Sea indicates isolated CpG site in the genome.

EPIC_Baboon_OA_DMPDist.xlsx

APPENDIX K

GENE DETAILS OF SIGNIFICANT DMPS IDENTIFIED IN THE EPIC ARRAY BABOON OSTEOARTHRITIS STUDY

Consult the attached Excel file that can be opened in Microsoft Excel. Tables show the details on the significant DMPs between each comparative group – bone vs. cartilage in OA, bone vs. cartilage in healthy, OA vs. healthy in bone, OA vs. healthy in cartilage, and all four combinations of disease state and tissue type. Table includes the identification number of each significant DMP probe (EPIC Array Probe ID), as well as additional annotation information for each significant DMP probe (Human Gene Symbol, Baboon Gene Symbol, Baboon Chromosome, Baboon CpG Position), the average β values for each comparative group (Average β Values), and the difference in β values across groups (Absolute Difference in β Values). Results for the initial DMP analysis before accounting for kinship (DMP Analysis) includes the log fold difference in M values between each comparative group (Log Fold Change in M Values) and the p-values for each DMP after accounting for multiple testing (Adjusted P-Value). Results for the DMP analyses that account for kinship (Kinship Analysis) include the log likelihood values when the comparative group variable of interest was included in (Log Likelihood with Variable of Interest) or excluded from (Log Likelihood without Variable of Interest) the GLMM, the chi-square value (X2), and the p-value for these tests (P-Value). The table order is based on the kinship analysis p-values (smallest to largest) and divided such that those probes with the difference in β values greater than 0.1 and less than 0.1 are separated.

EPIC_Baboon_OA_DMP.xlsx

APPENDIX L

SIGNIFICANT DMRS IDENTIFIED IN THE EPIC ARRAY BABOON OSTEOARTHRITIS STUDY

Consult the attached Excel file that can be opened in Microsoft Excel. Tables show the details on the significant DMRs between each comparative group (adjusted p-value < 0.05) – bone vs. cartilage in OA, bone vs. cartilage in healthy, OA vs. healthy in bone, OA vs. healthy in cartilage, and all four combinations of disease state and tissue type. These analyses used a Gaussian kernel bandwidth of 1000 base pairs with a scaling factor of 2 as recommending in the DMRcate package in R. DMRs had to contain at least 2 CpG sites that were less than 1000 base pairs apart, and p-values were adjusted using the Benjamini-Hochberg method. Tables include the genomic location of each DMR based on the human (hg19) genome (Human Chromosome, Human DMR Start, Human DMR End), the length of each DMR in base pairs (DMR Length (bp)), the number of CpG sites constituting the significant region (No. CpGs), the minimum adjusted p-values from the CpGs constituting each DMR (Adjusted P-Value), the stouffer transformations of the groups of false detection rates for individual CpG sites as DMR constituents for each DMR (FDR), the maximum absolute beta fold change within each DMR (Max. Log Fold Change in β Values), the mean beta fold change within each DMR (Mean Log Fold Change in β Values), and lists of genes with promotor regions that overlap with each DMR (Overlapping Promotors).

EPIC_Baboon_OA_DMR.xlsx

APPENDIX M

GO BIOLOGICAL PROCESSES ENRICHED FOR SIGNIFICANT DMPS IDENTIFIED IN THE EPIC ARRAY BABOON OSTEOARTHRITIS STUDY

Consult the attached Excel file that can be opened in Microsoft Excel. Tables contain the GO biological process terms that are significantly enriched (FDR < 0.05) for significant DMPs that have Δβ values greater than 0.1 between comparative groups – bone vs. cartilage in OA, bone vs. cartilage in healthy, OA vs. healthy in bone, OA vs. healthy in cartilage, and all four combinations of disease state and tissue type – taking into account the differing number of probes per gene present on the EPIC array. Tables include the identification numbers (GO ID) and terms (GO Biological Process Term) for each significantly enriched GO term, the total number of genes associated with each GO term (No. Genes Total), the number of genes with significant DMPs that are also associated with each GO term (No. Gene with DMPs), the p-value for over-representation of each GO term (P-Value), and the false discovery Rate for each GO term (FDR). For OA versus healthy bone, no GO categories significant at 5% FDR were identified, so all GO functions with p-values < 0.05 are listed.

EPIC_Baboon_OA_GO.xlsx

APPENDIX N

KEGG PATHWAYS ENRICHED FOR SIGNIFICANT DMPS IDENTIFIED IN THE EPIC ARRAY BABOON OSTEOARTHRITIS STUDY

Consult the attached Excel file that can be opened in Microsoft Excel. Tables contain the KEGG pathways that are significantly enriched (FDR < 0.05) for significant DMPs that have Δβ values greater than 0.1 between comparative groups – bone vs. cartilage in OA, bone vs. cartilage in healthy, OA vs. healthy in bone, OA vs. healthy in cartilage, and all four combinations of disease state and tissue type – taking into account the differing number of probes per gene present on the EPIC array. Tables include the identification numbers (KEGG ID) and pathways (KEGG Pathway) for each significantly enriched KEGG term, the total number of genes associated with each KEGG pathway (No. Genes Total), the number of genes with significant DMPs that are also associated with each KEGG pathway (No. Gene with DMPs), the p-value for over-representation of each KEGG pathway (P-Value), and the false discovery Rate for each KEGG pathway (FDR).

EPIC_Baboon_OA_KEGG.xlsx

APPENDIX O

OVERLAP OF GENES WITH DIFFERENTIAL METHYLATION ASSOCIATIONS FROM HUMAN AND BABOON OSTEOARTHRITIS STUDIES

Consult the attached Excel file which can be opened in Microsoft Excel. Details on the previous findings of differential methylation associations with human OA that do overlap with the current findings of differential methylation association with baboon OA. Table outlines several genes that overlap between human and baboons (Gene), the reference sources in which human data are described (Reference), the number of significant DMPs for each gene that were identified as being hypermethylated or hypomethylated in baboon OA (No. Probes), and whether any particular EPIC array probe exactly overlapped between humans and baboons (Matching Probes). The baboon results described here are for the OA vs. healthy comparison in cartilage tissues (Cartilage tab) and bone tissues (Bone tab).

EPIC_Baboon_OA_CompareGene.xlsx

APPENDIX P

COMPARISON OF DIFFERENTIAL METHYLATION ASSOCIATIONS WITH OSTEOARTHRITIS IN HUMAN AND BABOON STUDIES

Consult the attached Excel file that can be opened in Microsoft Excel. Details on the previous findings of differential methylation associations with human OA that do and do not overlap with the current findings of differential methylation association with baboon OA.

The Overlapping Table outlines several genes with differential methylation in human and baboon OA studies (Gene). Details on previous findings in humans include the publications source of these data (Reference), the methods used to collect these data (Method), as well as specifics (when available) on the locus identification number that was differentially methylated (Array Probe ID), the average methylation (β Value for OA) for the particular OA phenotype evaluated in each study (Phenotype of Interest), the specific comparison made in each study (Comparison), and whether the phenotype was hyper- or hypo-methylated after each comparison was performed (OA Methylation Level). Details on the current findings of this study include the probe identification numbers (EPIC Array Probe ID) of DMPs that were significant after performing specific comparisons (Comparison) and accounting for kinship, whether the OA was hyper- or hypo-methylated after each comparison was performed (OA Methylation Level), whether the Δβ values identified were greater than 0.1 (Δβ Value > 0.1), and the significance level of the association (Adjusted P-Value). Additionally, the average β values for each comparative group and annotation information for each probe are provided. For the current study, when the comparison was tissue type x disease state (OA bone vs. healthy bone vs. OA cartilage vs. healthy cartilage), the OA methylation level was listed as hypermethylated when methylation in both tissues was higher for OA than that of

healthy, hypomethylated when methylation in both tissues was lower for OA than that of healthy, or mixed if either of the previous cases was not met.

The Non-Overlapping Table outlines several genes with differential methylation in human OA studies (Gene). Details on previous findings in humans include the publications source of these data (Reference), the methods used to collect these data (Method), as well as specifics (when available) on the locus identification number that was differentially methylated (Array Probe ID), the average methylation (β Value for OA) for the particular OA phenotype evaluated in each study (Phenotype of Interest), the specific comparison made in each study (Comparison), and whether the phenotype was hyper- or hypo-methylated after each comparison was performed (OA Methylation Level). Lastly, details on whether each gene or locus was or was not included in the current study after probe filtering (Tested in Current Study?) is provided.

The References Table provides details on the references included in the human-baboon OA methylation comparisons.

Abbreviations: NA = gene symbol not available, 27K = Illumina Infinium HumanMethylation27K BeadChip, 450K = Illumina Infinium HumanMethylation450K BeadChip, microarray = Agilent Human Promoter Microarray, OA = osteoarthritis, OP = osteoporosis.

EPIC_Baboon_OA_Compare.xlsx

APPENDIX Q

DMP ANALYSIS RESULTS FOR SPECIFIC GENES IN EPIC ARRAY BABOON OSTEOARTHRITIS STUDY

Consult the attached Excel file which can be opened in Microsoft Excel. Table describing the DMP analysis results for four genes - *TBX4*, *HOXD8*, *LEPR*, and *RUNX1*. Table includes the identification number of each DMP examined (EPIC Array Probe ID), as well as additional annotation information for each of these sites (Human Gene Symbol, Human Chromosome (hg19), Human CpG Position (hg19), Baboon Gene Symbol, Baboon Chromosome, Baboon CpG Position), whether each site was found to be significantly differentially methylated between OA and healthy baboons when examining cartilage or bone tissues and if so whether OA tissues were hypermethylated or hypomethylated as compared to healthy tissues (Significant DMP between OA and Healthy Baboons), and the average β values for each comparative group (Average β Values). Those significant DMPs that had $\Delta\beta \geq 0.1$ are shown in bold.

EPIC_Baboon_OA_DMPGenes.xlsx

APPENDIX R

NORMALIZED AND FILTERED EPIC ARRAY METHYLATION VALUES FOR THE BABOON AGING STUDY

Consult the attached zipped file that contains the normalized and filtered EPIC array β-values (EPIC_Baboon_Aging_Beta.csv) and M-values (EPIC_Baboon_Aging_M.csv) for each sample and each retained probe that were used in the EPIC array baboon aging study. Each CSV can be opened in Microsoft Excel, Notepad, or any other text editor.

EPIC_Baboon_Aging.zip

APPENDIX S

BABOON SAMPLE SET FOR EPIC ARRAY AGING STUDY

Consult the attached Excel file which can be opened in Microsoft Excel. Table outlines each sample's animal identification number (Animal ID), species (Species), age classification (Age Cohort), sex (Sex), and age in years (Age), as well as the identification number of the beadchip (Array ID) and array (Position) that each sample was run on.

EPIC_Baboon_Aging_Samples.xlsx

APPENDIX T

ALIGNMENT PARAMETER CORRELATIONS OF EPIC ARRAY PROBES RETAINED FOR BABOON AGING STUDY

Consult the attached Excel file which can be opened in Microsoft Excel. Results of Spearman correlation tests between EPIC array probe detection p-values and alignment quality parameters. These parameters included the alignment bitscores, e-values, and percent identity. For the probes that fit the alignment filter criteria, significant negative correlations were identified for alignment bitscores and percent identities, and significant positive correlations were identified for alignment e-values.

EPIC_Baboon_Aging_ProbeCorr.xlsx

APPENDIX U

GENE DETAILS OF SIGNIFICANT DMPS IDENTIFIED IN THE EPIC ARRAY BABOON AGING STUDY

Consult the attached Excel file that can be opened in Microsoft Excel. Table shows the details on the significant DMPs between adult and juvenile baboons. Table includes the identification number of each significant DMP probe (EPIC Array Probe ID), as well as additional annotation information for each significant DMP probe (Human Gene Symbol, Baboon Gene Symbol, Baboon Chromosome, Baboon CpG Position), the average β values for each comparative group (Average β Values), and the difference in β values between age cohorts (Absolute Difference in β Values). Results for the initial DMP analysis before accounting for kinship (DMP Analysis) includes the log fold difference in M values between each comparative group (Log Fold Change in M Values) and the p-values for each DMP after accounting for multiple testing (Adjusted P-Value). Results for the DMP analyses that account for kinship (Kinship Analysis) include the log likelihood values when the comparative group variable of interest was included in (Log Likelihood with Variable of Interest) or excluded from (Log Likelihood without Variable of Interest) the GLMM, the chi-square value (X2), and the p-value for these tests (P-Value). The table order is based on the kinship analysis p-values (smallest to largest) and divided such that those probes with the difference in β values greater than 0.1 and less than 0.1 are separated.

EPIC_Baboon_Aging_DMP.xlsx

APPENDIX V

GENOMIC DISTRIBUTION OF SIGNIFICANT DMPS IDENTIFIED IN THE EPIC ARRAY BABOON AGING STUDY

Consult the attached Excel file which can be opened in Microsoft Excel. Table showing the number of significant DMPs with $\Delta\beta \geq 0.10$ between adult and juvenile baboons when accounting for kinship and the total number of loci examined that occupy different genomic regions (based on the human genome hg19). Number of Genes indicates the number of unique gene symbols associated with DMPs. Probes Per Gene indicates the average number of DMPs within each associated gene. For genomic locations, TSS200 and TSS1500 indicate the transcription start site areas between the start of the gene and 200bp upstream or 1500bp upstream respectively, 5'UTR and 3'UTR indicate the untranslated regions of genes, 1st Exon indicates the first exon within a gene, ExonBnd indicates the gene body exons, and Gene Body indicates any area within the exons and introns of a gene. For proximity to CpG islands, Island indicates areas within CpG islands, North Shelf indicates areas 2-4kb upstream from a CpG island, North Shore indicates areas up to 2kb upstream from a CpG island, South Shelf indicates areas 2-4kb downstream from a CpG island, South Shore indicates areas up to 2kb downstream from a CpG island, and Open Sea indicates isolated CpG site in the genome.

EPIC_Baboon_Aging_DMPDist.xlsx

APPENDIX W

SIGNIFICANT DMRS IDENTIFIED IN THE EPIC ARRAY BABOON AGING STUDY

Consult the attached Excel file that can be opened in Microsoft Excel. Table shows the details on the significant DMRs between adult and juvenile baboons (adjusted p-value < 0.05). These analyses used a Gaussian kernel bandwidth of 1000 base pairs with a scaling factor of 2 as recommending in the DMRcate package in R. DMRs had to contain at least 2 CpG sites that were less than 1000 base pairs apart, and p-values were adjusted using the Benjamini-Hochberg method. Tables include the genomic location of each DMR based on the human (hg19) genome (Human Chromosome, Human DMR Start, Human DMR End), the length of each DMR in base pairs (DMR Length (bp)), the number of CpG sites constituting the significant region (No. CpGs), the minimum adjusted p-values from the CpGs constituting each DMR (Adjusted P-Value), the stouffer transformations of the groups of false detection rates for individual CpG sites as DMR constituents for each DMR (FDR), the maximum absolute beta fold change within each DMR (Max. Log Fold Change in β Values), the mean beta fold change within each DMR (Mean Log Fold Change in β Values), and lists of genes with promotor regions that overlap with each DMR (Overlapping Promotors).

EPIC_Baboon_Aging_DMR.xlsx

APPENDIX X

GO BIOLOGICAL PROCESSES ENRICHED FOR SIGNIFICANT DMPS IDENTIFIED IN THE EPIC ARRAY BABOON AGING STUDY

Consult the attached Excel file that can be opened in Microsoft Excel. Table contains the GO biological process terms that are significantly enriched (FDR < 0.05) for significant DMPs that have Δβ values greater than 0.1 between adult and juvenile baboons, taking into account the differing number of probes per gene present on the EPIC array. Tables include the identification numbers (GO ID) and terms (GO Biological Process Term) for each significantly enriched GO term, the total number of genes associated with each GO term (No. Genes Total), the number of genes with significant DMPs that are also associated with each GO term (No. Gene with DMPs), the p-value for over-representation of each GO term (P-Value), and the false discovery Rate for each GO term (FDR).

EPIC_Baboon_Aging_GO.xlsx

APPENDIX Y

KEGG PATHWAYS ENRICHED FOR SIGNIFICANT DMPS IDENTIFIED IN THE EPIC ARRAY BABOON AGING STUDY

Consult the attached Excel file that can be opened in Microsoft Excel. Table contains the KEGG pathways that are significantly enriched (FDR < 0.05) for significant DMPs that have Δβ values greater than 0.1 between adult and juvenile baboons, taking into account the differing number of probes per gene present on the EPIC array. Tables include the identification numbers (KEGG ID) and pathways (KEGG Pathway) for each significantly enriched KEGG term, the total number of genes associated with each KEGG pathway (No. Genes Total), the number of genes with significant DMPs that are also associated with each KEGG pathway (No. Gene with DMPs), the p-value for over-representation of each KEGG pathway (P-Value), and the false discovery Rate for each KEGG pathway (FDR).

EPIC_Baboon_Aging_KEGG.xlsx

APPENDIX Z

COMPARISON OF DIFFERENTIAL METHYLATION ASSOCIATIONS WITH AGING IN HUMAN AND BABOON STUDIES

Consult the attached Excel file that can be opened in Microsoft Excel. Details on the previous findings of differential methylation associations with human aging that overlap with the current findings of differential methylation associations with baboon aging. Tables outlines several CpG sites (CpG Comparison tab) and genes (Gene Comparison tab) with differential methylation in human and baboon aging studies. Details on previous findings in humans include the publications source of these data (Reference), as well as specifics on the locus identification number that was differentially methylated (Array Probe ID) and whether increased aging was associated with hyper- or hypo-methylation (Methylation Level Associated with Aging). Details on the current findings of this study include the probe identification numbers (EPIC Array Probe ID) of DMPs that were tested, whether aging was associated with hyper- or hypo-methylation (Methylation Level Associated with Aging), and the significance level of the association (Adjusted P-Value). Additionally, the average β values for each comparative group and annotation information for each probe are provided.

EPIC_Baboon_Aging_Compare.xlsx

APPENDIX AA

DMP ANALYSIS RESULTS FOR SPECIFIC GENES IN EPIC ARRAY BABOON AGING STUDY

Consult the attached Excel file which can be opened in Microsoft Excel. Table describing the DMP analysis results for two genes - *KCNQ1DN* and *MBNL1*. Table includes the identification number of each DMP examined (EPIC Array Probe ID), as well as additional annotation information for each of these sites (Human Gene Symbol, Human Chromosome (hg19), Human CpG Position (hg19), Baboon Gene Symbol, Baboon Chromosome, Baboon CpG Position), whether each site was found to be significantly differentially methylated between adult and juvenile baboons and/or had $\Delta\beta \geq 0.1$ (Baboon DMP Results), and the average β values for each comparative group (Average β Values).

EPIC_Baboon_Aging_DMPGenes.xlsx

APPENDIX BB

NONHUMAN PRIMATE SAMPLE SET FOR EPIC ARRAY INTRA- AND INTER-SPECIFIC STUDY

Consult the attached Excel file which can be opened in Microsoft Excel. Details of sample set. Table outlines each sample's animal identification number (Animal ID), species (Species), sex (Sex), age in years (Age), the identification number of the beadchip (Array ID) and array (Position) that each sample was run on, and whether downstream gene-specific analyses were performed on a sample (Gene-Specific Analyses).

EPIC_NHP_Study_Samples.xlsx

APPENDIX CC

DETAILS OF NONHUMAN PRIMATE MORPHOLOGICAL MEASUREMENTS

Linear morphological measurements used in GLMMs (Measurements), as well as triplicate measurements of linear morphologies and error calculations from these (Error). All measurements have units of millimeters. NA indicates that measurement could not be accurately collected. Error was calculated in each taxonomic group and equals the mean absolute difference divided by the mean (Corner et al. 1992; White and Folkens 2000). Only measurements with an error <5% were included in downstream analyses. This included all measurements except for intercondylar notch depth for macaques. NA indicates that measurement could not be accurately collected.

EPIC_NHP_IntraStudy_Morphology.xlsx

APPENDIX DD

NORMALIZED AND FILTERED EPIC ARRAY METHYLATION VALUES FOR THE NONHUMAN PRIMATE INTRA- AND INTER-SPECIFIC STUDY

Consult the attached zipped files that contains the normalized and filtered EPIC array β-values (*_Beta.csv) and M-values (*_M.csv) for each sample and each retained probe that were used in the EPIC array nonhuman primate intra- and inter-specific study. Each CSV can be opened in Microsoft Excel, Notepad, or any other text editor.

EPIC_NHP_IntraStudy_Baboon.zip

EPIC_NHP_IntraStudy_Chimpanzee.zip

EPIC_NHP_IntraStudy_Macaque.zip

EPIC_NHP_IntraStudy_Marmoset.zip

EPIC_NHP_IntraStudy_Vervet.zip

EPIC_NHP_InterStudy.zip

APPENDIX EE

NONHUMAN PRIMATE GENE-SPECIFIC *HOXD10* PRIMERS

Consult the attached Excel file which can be opened in Microsoft Excel. Table describing the primers used for the gene-specific amplification of the *HOXD10* gene using regular and bisulfite treated DNA. Details include the name of the *HOXD10* region being amplified (PCR Name), whether the primer was designed to amplify regular or bisulfite treated DNA (Primer Type), the species in which the primers were designed to amplify each *HOXD10* region (Species), the sequences for the forward and reverse primers in the 5' to 3' direction (Forward Primer Sequence (5'-3'), Reverse Primer Sequence (5'-3'), the length of amplicons in base pairs (Amplicon Length), and the optimized annealing temperature (Optimized Tm).

GeneSpecific_HOXD10_Primers.xlsx

APPENDIX FF

GENE-SPECIFIC *HOXD10* PCR ASSAY SPECIFICATIONS

Consult the attached Excel file which can be opened in Microsoft Excel. Specifications for gene-specific *HOXD10* PCR assays. The Temperatures tab describes the PCR amplification temperature conditions used for both regular and bisulfite treated DNA. The annealing temperature varied for primer pairs. See optimized temperature for each primer pair in Appendix EE. The Regular Reagents and Bisulfite Reagents tabs describe the reagent initial concentrations ([initial]), final concentrations ([final]), and amounts (Volume (uL)) used for PCR amplification of regular DNA and bisulfite converted DNA, respectively. The minimum initial concentration for DNA was 5ng/uL.

GeneSpecific_HOXD10_Specs.xlsx

APPENDIX GG

GENOMIC DISTRIBUTION OF EPIC ARRAY PROBES RETAINED FOR NONHUMAN PRIMATES

Consult the attached Excel file which can be opened in Microsoft Excel. Details on the gene associations, genomic locations, and proximity to CpG islands (based on the human genome hg19) of all probes that successfully mapped to each nonhuman primate genome with e-values less than e-10, had only unique BLAST hits, and targeted a CpG site (Total Mapped Probes), probes that fit the alignment filter criteria (Alignment Filter Probes), and probes that fit the gene symbol filter criteria (Gene Symbol Filter Probes). Number of Genes indicates the number of unique gene symbols associated with probes. Probes Per Gene indicates the average number of probes that target each associated gene. For genomic locations, TSS 200 and TSS 1500 indicate the transcription start site areas between the start of the gene and 200bp upstream or 1500bp upstream respectively, 5'UTR and 3'UTR indicate the untranslated regions of genes, 1st Exon indicates the first exon within a gene, ExonBnd indicates the gene body exons, and Gene Body indicates any area within the exons and introns of a gene. For proximity to CpG islands, Island indicates areas within CpG islands, North Shelf indicates areas 2-4kb upstream from a CpG island, North Shore indicates areas up to 2kb upstream from a CpG island, South Shelf indicates areas 2-4kb downstream from a CpG island, South Shore indicates areas up to 2kb downstream from a CpG island, and Open Sea indicates isolated CpG site in the genome.

EPIC_NHP_ProbeDist.xlsx

APPENDIX HH

ALIGNMENT PARAMETER CORRELATIONS OF EPIC ARRAY PROBES RETAINED FOR NONHUMAN PRIMATE INTRA- AND INTER-SPECIFIC STUDY

Consult the attached Excel file which can be opened in Microsoft Excel. Results of Spearman correlation tests between EPIC array probe detection p-values and alignment quality parameters. These parameters included the alignment bitscores, e-values, and percent identity. For all probes that successfully mapped to one of the nonhuman primate genomes with e-values < e-10, had only unique BLAST hits, and targeted a CpG site (Total Mapped Probes), probes that fit the alignment filter criteria (Alignment Filter Probes), and probes that fit the gene symbol filter criteria (Gene Symbol Filter Probes), significant negative correlations were identified for alignment bitscores and percent identities, and significant positive correlations were identified for alignment e-values.

EPIC_NHP_Study_ProbeCorr.xlsx

APPENDIX II

GENE DETAILS OF SIGNIFICANT DMPS IDENTIFIED IN THE EPIC ARRAY NONHUMAN PRIMATE INTRA-SPECIFIC STUDY

Consult the attached Excel file which can be opened in Microsoft Excel. Table showing details on all significant DMPs associated with a linear morphology in a nonhuman primate species. Table includes the identification number of each significant DMP probe (EPIC Array Probe ID), as well as the log fold difference in M values between each comparative group (Log Fold Change in M Values), the p-values for each DMP after accounting for multiple testing (Adjusted P-Value), and additional annotation information for each significant DMP probe (Human Gene Symbol, Baboon Gene Symbol, Baboon Chromosome, Baboon CpG Position).

EPIC_NHP_IntraStudy_DMPGene.xlsx

APPENDIX JJ

GENE DETAILS OF SIGNIFICANT DMPS IDENTIFIED IN THE EPIC ARRAY NONHUMAN PRIMATE INTRA-SPECIFIC MORPHOLOGY STUDY

Consult the attached Excel file that can be opened in Microsoft Excel. Tables show details on all significant DMPs associated with a linear morphology in a nonhuman primate species – baboon bicondylar femur length, baboon maximum femur length, macaque proximal width, macaque medial condyle width, vervet superior shaft width, vervet inferior shaft width, vervet anatomical neck height, and chimpanzee anatomical neck length. Tables includes the identification number of each significant DMP probe (EPIC Array Probe ID), as well as additional information about each sample associated with the significant DMP. Specifically, the animal identification number (Animal), sex (Sex), age in years (Age), morphological measurement in millimeters (Measurements), β value for each DMP (β Values), maximum absolute change in β values for each DMP (Max Δβ), and the realized change in β values between the nonhuman primate with the largest morphology measurement and the nonhuman primate with the smallest morphology measurement for each DMP (Realized Δβ) are provided.

EPIC_NHP_IntraStudy_DMP.xlsx

APPENDIX KK

GENE DETAILS OF SIGNIFICANT SPECIES-SPECIFIC DMPS IDENTIFIED IN THE EPIC ARRAY NONHUMAN PRIMATE INTER-SPECIFIC STUDY

Consult the attached Excel file that can be opened in Microsoft Excel. Tables show the details on the significant DMPs identified between taxonomic groups – baboons, macaques, vervets, chimpanzees, and marmosets – and showing species-specific methylation patterns. Table includes the identification number of each significant DMP probe (EPIC Array Probe ID), additional annotation information for each significant DMP probe (Human Gene Symbol, Human Chromosome, Human CpG Position, Nonhuman Primate Gene Symbol, Nonhuman Primate Chromosome, Nonhuman Primate CpG Position), the species-specific methylation observed for the species of interest (Nonhuman Primate Specific Methylation Pattern), the average β values for each taxonomic group (Average β Values), the difference in β values (Δβ) and p-values accounting for multiple testing (Adjusted P-Value) for each pairwise inter-specific comparison made for each DMP, and the average absolute difference in β values between the species of interest and each other taxonomic group (Average Absolute Difference in β Values). The table orders are divided such that those probes with the difference in β values less than 0.1, between 0.1 and 0.2, between 0.2 and 0.3, between 0.3 and 0.4, and greater than 0.4 are separated. Mixture indicates that pairwise DNA methylation comparisons between species have a combination of hypomethylation and hypermethylation. Human annotation information is for hg19.

EPIC_NHP_InterStudy_DMP.xlsx

APPENDIX LL

GENOMIC DISTRIBUTION OF SIGNIFICANT SPECIES-SPECIFIC DMPS IDENTIFIED IN THE EPIC ARRAY NONHUMAN PRIMATE INTER-SPECIFIC STUDY

Consult the attached Excel file which can be opened in Microsoft Excel. Details on the gene associations, genomic locations, and proximity to CpG islands (based on the human genome hg19) of all probes that successfully mapped to each nonhuman primate genome with e-values less than e-10, had only unique BLAST hits, targeted a CpG site, and fit the alignment filter criteria. Number of Genes indicates the number of unique gene symbols associated with DMPs. Probes Per Gene indicates the average number of DMPs within each associated gene. For genomic locations, TSS 200 and TSS 1500 indicate the transcription start site areas between the start of the gene and 200bp upstream or 1500bp upstream respectively, 5'UTR and 3'UTR indicate the untranslated regions of genes, 1st Exon indicates the first exon within a gene, ExonBnd indicates the gene body exons, and Gene Body indicates any area within the exons and introns of a gene. For proximity to CpG islands, Island indicates areas within CpG islands, North Shelf indicates areas 2-4kb upstream from a CpG island, North Shore indicates areas up to 2kb upstream from a CpG island, South Shelf indicates areas 2-4kb downstream from a CpG island, South Shore indicates areas up to 2kb downstream from a CpG island, and Open Sea indicates isolated CpG site in the genome.

EPIC_NHP_InterStudy_DMPDist.xlsx

APPENDIX MM

GO BIOLOGICAL PROCESSES ENRICHED FOR SIGNIFICANT SPECIES-SPECIFIC DMPS IDENTIFIED IN THE EPIC ARRAY NONHUMAN PRIMATE INTER-SPECIFIC STUDY

Consult the attached Excel file that can be opened in Microsoft Excel. Tables contain the GO biological process terms that are significantly enriched (FDR < 0.05) for species-specific DMPs – baboons, macaques, vervets, chimpanzees, and marmosets – taking into account the differing number of probes per gene present on the EPIC array and assessed in the current study. Tables include the identification numbers (GO ID) and terms (GO Biological Process Term) for each significantly enriched GO term, the total number of genes associated with each GO term (No. Genes Total), the number of genes with significant DMPs that are also associated with each GO term (No. Gene with DMPs), the p-value for over-representation of each GO term (P-Value), and the false discovery Rate for each GO term (FDR).

EPIC_NHP_InterStudy_GO.xlsx

APPENDIX NN

KEGG PATHWAYS ENRICHED FOR SIGNIFICANT SPECIES-SPECIFIC DMPS IDENTIFIED IN THE EPIC ARRAY NONHUMAN PRIMATE INTER-SPECIFIC STUDY

Consult the attached Excel file that can be opened in Microsoft Excel. Tables contain the KEGG pathways that are significantly enriched (FDR < 0.05) for species-specific DMPs – baboons, macaques, vervets, chimpanzees, and marmosets – taking into account the differing number of probes per gene present on the EPIC array and assessed in the current study. Tables include the identification numbers (KEGG ID) and pathways (KEGG Pathway) for each significantly enriched KEGG term, the total number of genes associated with each KEGG pathway (No. Genes Total), the number of genes with significant DMPs that are also associated with each KEGG pathway (No. Gene with DMPs), the p-value for over-representation of each KEGG pathway (P-Value), and the false discovery rate for each KEGG pathway (FDR).

EPIC_NHP_InterStudy_KEGG.xlsx

APPENDIX OO

DMP ANALYSIS RESULTS FOR *HOXD10* GENE IN EPIC ARRAY NONHUMAN PRIMATE INTER-SPECIFIC STUDY

Consult the attached Excel file which can be opened in Microsoft Excel. Table describing the differentially methylation position (DMP) analysis results for the *HOXD10* genes. Table includes the identification number of each DMP examined (EPIC Array Probe ID), as well as additional annotation information for each of these sites (Gene Symbols, Chromosomes, CpG Positions), whether each site was found to be significantly differentially methylated between taxonomic groups and if so whether the specified species was hypermethylated, hypomethylated, or had a mixture of hyper- and hypo-methylation as compared to other species (Species-Specific DMP), and the average β values for each taxonomic group (Average β Values). Of the 5 species-specific DMPs in the *HOXD10* gene of marmosets, 4 have Δβ between 0.2 and 0.3 (bold) and 1 has a Δβ < 0.1 (not bolded). Human information came from hg19.

EPIC_NHP_InterStudy_DMPGene.xlsx

APPENDIX PP

GENE-SPECIFIC *HOX10* RAW SEQUENCES

Consult the attached zipped file that contains the raw, unprocessed chromatogram files (*.ab1) and sequence files (*.seq) for all the regular and bisulfite gene-specific sequences of *HOXD10* used in the EPIC array nonhuman primate inter-specific study. These files can be opened in MEGA7 (Kumar et al. 2016), Geneious version 9.1.2 (Kearse et al. 2012), or other software compatible with reading these data types.

GeneSpecific_HOXD10_Sequences.zip

APPENDIX QQ

GENE-SPECIFIC *HOXD10* REGULAR SEQUENCE ALIGNMENTS

Consult the attached FASTA file that contains an alignment of the processed, regular gene-specific sequences of *HOXD10* used in the EPIC array nonhuman primate inter-specific study. These sequences are aligned to the regions surrounding and including *HOXD10* from several primates that were obtained from the EPO whole-genome multiple alignments of several primate genomes [Ensembl Compara.8_primates_EPO] (Paten, Herrero, Beal, et al. 2008; Paten, Herrero, Fitzgerald, et al. 2008). This file can be opened in MEGA7 (Kumar et al. 2016), Geneious version 9.1.2 (Kearse et al. 2012), or any other text editor.

GeneSpecific_HOXD10_AlignReg.fasta

APPENDIX RR

GENE-SPECIFIC *HOXD10* BISULFITE SEQUENCE ALIGNMENTS

Consult the attached FASTA file that contains an alignment of the processed, bisulfite gene-specific sequences of *HOXD10* used in the EPIC array nonhuman primate inter-specific study. These sequences are aligned to the regions surrounding and including *HOXD10* from several primates that were obtained from the EPO whole-genome multiple alignments of several primate genomes [Ensembl Compara.8_primates_EPO] (Paten, Herrero, Beal, et al. 2008; Paten, Herrero, Fitzgerald, et al. 2008) and bisulfited converted *in silico*. This file can be opened in MEGA7 (Kumar et al. 2016), Geneious version 9.1.2 (Kearse et al. 2012), or any other text editor.

GeneSpecific_HOXD10_AlignBS.fasta

APPENDIX SS

GENE-SPECIFIC SEQUENCING RESULTS FOR *HOXD10* GENE IN EPIC ARRAY NONHUMAN PRIMATE INTER-SPECIFIC STUDY

Consult the attached Excel file which can be opened in Microsoft Excel. Table describing the results for the gene-specific regular and bisulfite sequencing across the *HOXD10* gene (hg19 chr2:176981492-176984670), as well as upstream and downstream several hundred bases (hg19 chr2:176980532-176985117). Table includes the positions of human derived CpG sites in the *HOXD10* gene (Human CpG Position) based on the sequence alignments produced in this study (Appendix QQ and Appendix RR). CpG sites that were also targeted by the EPIC array are shown in bold, and probe information is also provided (EPIC Array Probe ID). For each nonhuman primate sample examined, the presence of nucleotide mutations (Mutation) at CpG sites, as well as the dinucleotide resulting from these mutations shown in parentheses, at CpG sites was determined based on regular sequences (Appendix QQ), and the presence of methylation (Methylation) at CpG sites was determined based on bisulfite sequences (Appendix RR). Abbreviations: no data collected (-), data quality too poor to call base pair (NA), methylation present at site indicated by cytosine in bisulfite sequence (yes), no methylation at site indicated by a lack of cytosine at site (no), partial methylation indicated by a partial cytosine signal at site (partial), CpG site 1bp upstream or downstream of that in humans (*), and CpG site 2bp upstream or downstream of that in humans (**). Overall, out of the 161 human derived CpG sites, 21 showed mutations in baboons except in 1X2996 which had one additional mutation likely due to intra-specific variation, 25 showed mutations in macaques except in 17538 which only had 19 mutations due to poor sequence quality and the inability to call variants at these sites, 21 showed mutations in vervets, 3 showed mutations in chimpanzees except in 4-0191 which only had 2 mutations likely due to intra-specific variation, and 28 showed mutations in marmosets except in 18482 which

only had 19 mutations due to poor sequence quality and the inability to call variants at these sites. None of the methylation present in these nonhuman primate samples is present at mutated CpG sites.

GeneSpecific_HOXD10_CpG.xlsx