**Supporting Information**

*Resource base influences genome-wide DNA methylation levels in wild baboons*

RRBS library construction, sequencing, and data processing 2

Testing for effects of resource base on DNA methylation levels 3

Estimating the relationship between sample size and power 4

Testing for cell type heterogeneity-related confounds 5

Defining genomic compartments to test for enrichment of differentially methylated sites 7

Transforming count data to normalized methylation proportions 8

Testing the degree to which DMRs occur more often than expected by chance 8

*PFKP* reporter constructs 9

Cell culture and transfection procedures 11

Using support vector machines to investigate the plasticity versus stability of DNA methylation 12

References 14

Figure S1. Flow chart describing data processing steps and main analyses 18

Figure S2. Genomic compartment annotations used in this study 19

Figure S3. Histone marks and genomic compartments associated with chromatin states 20

Figure S4. Differentially methylated sites are enriched near genes expressed in whole blood 21

Figure S5. Effect of resource base on DNA methylation levels at the pathway level 22

Figure S6. DMRs are observed in the real data set more often than expected by chance 23

Figure S7. RRBS enriches for putatively functional regions of the genome and recapitulates known patterns of DNA methylation across the genome 24

Figure S8. Power to detect differentially methylated sites increases with sample size 26

Figure S9. Magnitude of the effect of resource availability on DNA methylation levels in different genomic compartments 27

Figure S10. Cell type proportions did not significantly differ between wild-feeding and Lodge 28

Figure S11. Enrichment of resource base-associated sites is strongest near genes expressed in whole blood, compared to genes expressed in other tissues 29

Table S1. Information about males that switched between resource base conditions 30

Table S2. Baboon RRBS dataset sample characteristics and read mapping summary 31

Table S3. Differentially methylated regions 33

# RRBS library construction, sequencing, and data processing

To construct RRBS libraries, we extracted genomic DNA from whole blood samples using the DNeasy Blood and Tissue kit (QIAGEN), according to the manufacturer’s instructions. We then used 180 ng of genomic DNA per individual to generate adapter barcoded libraries for each sample. Libraries were pooled together in sets of 9-12 samples, subjected to sodium bisulfite conversion using the EpiTect Bisulfite Conversion kit (QIAGEN), and then PCR amplified prior to sequencing on the Illumina HiSeq 2000 platform. Each pooled set of libraries was sequenced in a single lane to a mean depth (± SD) of 27.25 ± 13.62 million reads per sample (173.85 ± 12.87 million reads per lane; Table S2). To assess the efficiency of the bisulfite conversion, 1 ng of unmethylated lambda phage DNA (Sigma Aldrich) was added to each 180 ng sample prior to library construction.

Following read trimming for adaptor contamination and base quality, we mapped all reads to a combined reference genome that included the olive baboon genome (*Panu 2.0*) and the lambda phage genome. We used sequences that mapped to the lambda phage genome to estimate the bisulfite conversion efficiency for each DNA sample. Specifically, for each sample, we summed (i) the number of reads that mapped to lambda phage CpG sites and were read as thymine (reflecting an unmethylated cytosine that was converted to thymine by sodium bisulfite); and (ii) the total number of reads that mapped to any lambda phage CpG site. Because all CpG sites in the lambda phage genome were originally unmethylated (and thus should have been converted to thymine), the ratio of these two sums provides an estimate of bisulfite conversion efficiency. Here, a ratio of 1 would represent perfect conversion of every unmethylated cytosine to thymine.

For data generated from baboon DNA, we performed several additional filtering steps before performing differential methylation analyses. First, following previous studies [1,2], we excluded sites that were constitutively hypermethylated (average DNA methylation level > 0.90) or hypomethylated (average DNA methylation level < 0.10). We also excluded sites in which the standard deviation of DNA methylation levels fell in the lowest 5% of the overall data set, and sites that fell in the lowest quartile of mean coverage depth (4.74 reads). The final filtered data set contained 535,996 CpG sites.

Importantly, although we mapped reads generated from yellow baboons to the olive baboon genome, biases associated with mapping to a heterospecific genome should be minimal in the filtered dataset. The primary source of potential bias comes from C/T SNPs, and specifically cases in which the olive baboon reference genome contains a CG site, but yellow baboons carry TG at the same location in the genome. In these cases, BSMAP would erroneously estimate the CpG site in our yellow baboon samples as completely unmethylated. However, because we filtered out all sites with average methylation levels below 10%, sites affected by this type of error should not be included in our final filtered dataset.

# Testing for effects of resource base on DNA methylation levels

We used the software MACAU [3] to test for differences in DNA methylation levels between wild-feeding and Lodge individuals at each CpG site in our filtered dataset. Specifically, we fit a binomial mixed effects model for each CpG site across all individuals (*i*):

where is the total read count for the *i*th individual; is the methylated read count for that individual; and is an unknown parameter that represents the true proportion of methylated reads for individual *i*. MACAU then uses a logit link to model as a linear function of several parameters:

where is a vector of covariates we wish to control for, as well as an intercept, and is an identically sized vector of corresponding coefficients. In our case, contains the following covariates: sex of the sampled individual, age of the sampled individual, age of the blood sample (measured in years since date of sampling), and bisulfite conversion rate (estimated from the lambda phage spike in). The ages of all individuals born in the Amboseli study population (n = 51 in our data set) are known to within a few days’ error. For animals that immigrate into the study population, ages are estimated from morphological features by trained observers (n = 18 animals in our data set). denotes the foraging group of the *i*th individual (Lodge or wild-feeding) and is the corresponding estimate of this effect. In addition, we included a random genetic effect (. The variance of is influenced by pairwise relatedness between individuals in the sample, given by the matrix . Pairwise relatedness values were calculated from previously collected microsatellite data (14 highly polymorphic loci) using the program COANCESTRY [5–7]. In this model, and represent genetic and environmental variance components for the DNA methylation trait, and denotes model error. For each model (i.e., CpG site), we extracted the p-value associated with the effect of foraging group on DNA methylation levels ().

# Estimating the relationship between sample size and power to detect resource base-associated sites

We used a subset of our data to estimate the relationship between sample size and the number of true positive, resource-base associated sites detected in our analyses. Specifically, for each CpG site on chromosome 1 (in our filtered dataset, n = 49,576 CpG sites), we estimated the p-value for the resource base effect on DNA methylation levels in our full data set (n = 61 lifelong wild-feeding or Lodge individuals, excluding switching individuals) and in 16 randomly subsampled data sets of n = 20, 30, 40, or 50 (four replicates for each sample size, balancing the number of Lodge versus wild-feeding individuals to reflect the composition of the original data set). We modeled resource base effects on DNA methylation levels in the subsampled data sets using the same approach described in the main text Methods and in the SI section on *Testing for effects of resource base on DNA methylation levels*.

Next, we modeled the relationship between sample size and p-value at each site using a power law distribution (as described in Mukherjee *et al.* 2003). For sites truly associated with resource base, we expect an inverse relationship between sample size and significance (the larger the sample, the lower the p-value). We considered sites that met this criterion as putative true positives, and estimated the p-values we would observe at these sites given data sets of increasing sample size. In contrast, we considered sites that showed no relationship between sample size and significance to be putative true negatives. Finally, we predicted our power to detect the total set of putative true positive sites at a 10% FDR in data sets of increasing sample size. The results of this analysis (Figure S8) show that we are underpowered to detect all true resource base-associated sites in our current data set at a 10% FDR. Specifically, they suggest that we detected ~9% of putative true positive sites with relatively large effects (sites that fall in the upper 25% of the effect size distribution) and probably none of the true positive sites with small effect sizes (bottom 25% of the effect size distribution).

# Testing for cell type heterogeneity-related confounds

Our analyses focused on DNA methylation levels in whole blood, which is composed of several cell types with distinct DNA methylation profiles [9]. Thus, if cell type composition consistently differed between blood samples collected from Lodge versus wild-feeding animals, differentially methylated sites could reflect these differences rather than changes in DNA methylation levels within cells. To investigate this possibility, we used two complementary approaches. First, we measured cell type proportions from wild-feeding animals (n = 25, including 6 animals for whom we also collected DNA methylation data) and Lodge group animals (n = 15, including 5 animals for whom we also collected DNA methylation data) based on manual counts of Giemsa-stained blood smears collected at the time of blood sample collection. These counts resulted in estimates of the proportional representation of eosinophils, basophils, neutrophils, lymphocytes, and monocytes for baboons in both resource conditions. We then used generalized linear models with a binomial link function (implemented in R via ‘glm’) to test the hypothesis that foraging group influenced these proportions, controlling for the effects of age, sex, and the identity of the person who counted the blood smear. We found that, for all cell types tested, there was no significant differences in abundance between wild-feeding and Lodge individuals (all p > 0.05, Figure S10).

Second, we drew on publicly available DNA methylation data generated from pure populations of CD4+ (helper) T cells, CD8+ (cytotoxic) T cells, natural killer cells, B cells, monocytes, and granulocytes from human whole blood [10]. Because these data were generated on Illumina 450K Infinium arrays, relatively few CpG sites overlapped with our data set. However, for the 7,425 CpG sites that were profiled in both the Jaffe data set and in our samples, we were able to ask whether sites associated with resource base in our analysis (at a 10% FDR) were more likely to be differentially methylated between cell types than expected by chance, using a Fisher’s Exact Test. We considered a CpG site to be differentially methylated by cell type if (i) a significant (10% FDR) difference among cell types was reported by Jaffe 2015 (using an ANOVA), and (ii) average DNA methylation levels differed by at least 10% between the two most prevalent cell types (granulocytes and CD4+ T cells, which together account for greater than two-thirds of nucleated human blood cells on average [11]. To explore the sensitivity of our findings to these cutoffs, we also repeated our analysis while varying both the FDR threshold used to identify CpG sites associated with resource base (from 0 to 20%) and the required methylation difference between the two most prevalent cell types (from 0 to 50%). The results of these analyses revealed that CpG sites that exhibit cell-type specific DNA methylation patterns in whole blood were not more likely to be associated with resource base in our analyses. Further, this result was robust across a range of cutoffs (p > 0.05 for 99.7% of tests).

# Defining genomic compartments to test for enrichment of differentially methylated sites

In the main text, we report analyses that investigated the enrichment of resource base- associated sites in functionally coherent genomic compartments (gene bodies, promoters, CpG islands and shores, enhancers, and regions with no known regulatory function, i.e., unannotated regions). Here, we provide information about how we defined those compartments.

We defined gene bodies as the regions between the 5’-most transcription start site (TSS) and 3’-most transcription end site (TES) of each gene using *Panu* 2.0 annotations from Ensembl [12] and gene promoters as the 2 kb region upstream of the TSS (following Deng *et al.* 2009; Shulha *et al.* 2013). CpG islands were annotated based on the UCSC Genome Browser track for baboon [15], and CpG island shores were defined as the 2 kb regions flanking either side of the CpG island boundary (following Gu *et al.* 2011; Rönn *et al.* 2013; Hernando-Herraez *et al.* 2013). Because no baboon-specific enhancer annotations are available, we defined putative enhancer regions based on H3K4me1 ChIP-seq data generated by ENCODE from human peripheral blood mononuclear cells [19]. Specifically, we took the list of all H3K4me1 ChIP-seq peak locations in the human genome, and projected these coordinates onto the olive baboon genome using the UCSC Genome Browser *liftover* tool [20]. Finally, we defined CpG sites that were not included in any of the above annotation categories as ‘unannotated’. Further information about the location and functional role of these genomic compartments is provided in Figure S2.

# Transforming count data to normalized methylation proportions

For our main association analyses, we directly modeled counts of methylated reads and unmethylated reads for each CpG site and individual. However, two of the analyses we performed (on DNA methylation levels near metabolism-related genes using the R package *GlobalTest*: [21]; and on methylation patterns in individuals that switched resource base using the R package *kernlab:* [22]) required continuous, normally distributed input data. For these analyses, we transformed our count data into normalized proportions as follows. First, we obtained the proportion of methylated reads for each individual and CpG site by dividing the number of methylated reads by the total read count for each individual-site combination. We then quantile normalized the resulting proportions for each CpG site across individuals to a standard normal distribution, and imputed any missing data using the K-nearest neighbors algorithm in the R package *impute* [23].

# Testing the degree to which DMRs occur more often than expected by chance

To identify resource base-associated DMRs in our data set, we focused on resource base-associated sites that had at least one other measured CpG site within either 1 kb upstream or downstream of the focal site. This criterion was met for 847 of the 1,014 resource base-associated sites detected at a 10% FDR. For these 847 sites, we counted (i) the absolute number of nearby (i.e., within 1 kb) sites that also exhibited evidence for differential methylation, and (ii) the proportion of nearby sites that also exhibited evidence for differential methylation. We considered a nearby site to show evidence for differential methylation if it was associated with resource base at a relaxed 20% FDR threshold.

To understand whether DMRs occurred more often than expected by chance in our real data, and to identify the cutoffs we ultimately used to define DMRs, we compared the distributions of the number and proportion of differentially methylated nearby sites to distributions created in an identical manner from permuted data. Specifically, we permuted resource base label (wild-feeding or Lodge) across individuals in our data set and re-ran our association analyses in MACAU. These permutations produced uniform p-value distributions, as expected: see Figure 2 in the main text. We then replicated the procedure for identifying DMRs using results obtained from the permuted data. Specifically, we (i) chose the 847 sites with the smallest p-values (again focusing only on CpG sites with at least 1 other CpG site within a 1 kb distance); and (ii) counted both the absolute number and proportion of nearby sites with evidence for differential methylation. We performed this procedure using p-values obtained from four independent permutations, and averaged the resulting distributions. To test whether DMRs were more likely to occur in the real versus permuted data sets, we then compared the distributions generated from the real versus permuted data sets using Kolmogorov-Smirnov tests.

We found that differentially methylated sites occurred in spatial clusters more often than expected by chance in the real data. CpG sites associated with resource base were surrounded by a higher number and higher proportion of nearby sites that also were associated with resource base, compared to chance expectations (Kolmogorov-Smirnov test for proportion of surrounding sites: D = 0.113, p = 3.76 x 10-5; for number of nearby sites, D = 0.123, p = 5.69 x 10-6 ; Figure S6). Together, these results suggest that changes in DNA methylation associated with resource base are consistently targeted to specific regions of the genome, and are thus more likely to have functional effects on gene expression [24–26]

# *PFKP* reporter constructs

We amplified 817 bp of the *PFKP* promoter region using the following primer pair: 5’-TCCAGATCTTAAGCTCTTTGGATGCGCGTATTTC-3’ (forward) and 5’-TCCATGGACTAAGCTGTGGGTTGGAGGACTCTGGT-3’ (reverse). These primers each contain a 15 base-pair tail that is complementary to the pCpGL vector, which is necessary for directional cloning via Gibson assembly. We amplified the *PFKP* promoter region by combining 1 U of AmpliTaq Gold, 1.5 mM MgSO4, 0.2 mM dNTPs, and 1 uM of the forward and reverse primers in a 1X buffer solution. Water was added to bring the total reaction volume to 25 uL. PCR cycling conditions were as follows: 94 °C for 3 minutes; followed by 35 cycles of 94 °C for 30 seconds, 53 °C for 30 seconds, and 72 °C for 1 minute; and a final 5 minutes at 72 °C. Following PCR amplification, we confirmed correct amplification of the target region by gel electrophoresis and Sanger sequencing.

We linearized the pCpGL vector with the restriction enzyme *HindIII* (New England BioLabs), following the manufacturer’s instructions. We then purified both the *PFKP* amplicon and the linearized plasmid via column cleanup (MinElute Reaction Cleanup Kit, QIAGEN), and ligated the two fragments together (Gibson Assembly Master Mix, New England BioLabs). The resulting plasmid DNA was chemically transformed into competent *E. coli* GT115 cells (InvivoGen), selected and grown on LB agar plates in the presence of Zeocin, and purified using the QIAprep Spin Miniprep Kit (QIAGEN).

We subjected the purified plasmid DNA (containing the pCpGL backbone and the *PFKP* promoter) to three different treatments: (i) *in vitro* methylation by *M.SssI* methyltransferase in the presence of the methyl donor S-adenosylmethionine (SAM), which results in methylation of all CpGs; (ii) *in vitro* methylation by *HhaI* methyltransferase in the presence of SAM, which results in methylation of the first cytosine (second base) in the sequence GCGC; and (iii) a mock treatment using only SAM and no methyltransferase. Below, we report the sequence for the *PFKP* promoter region we assayed, and show which CpGs were methylated in the 2 experimental treatment conditions. We confirmed that the *PFKP* promoter was methylated as expected by digesting each treated plasmid with *MspI*, a methylation insensitive enzyme that targets the sequence motif CCGG, and *HpaII*, an isoschizomer of *MspI* whose activity is blocked by DNA methylation, followed by gel electrophoresis to visualize the resulting fragments.

All restriction enzymes and methyltransferases were obtained from New England BioLabs.

*PFKP* promoter region sequence with color annotation for *M.SssI* condition: pink denotes CpGs methylated by *M.SssI*, yellow denotes primer sequence

CTTTGGATGCGCGTATTTCAGCACGTGGAGTTGGCCAGTGGAGCCGG CCGAGGTCGACACAGCGTACCGCAGCGCGCGTGTCCGTGGGGAGGAAGGG TCGTCCCGCCTCCCTCCCAGGCCAGTCCTGTTGCGCCTCCTCCTGCCCCA CTCTGACTTTCGCGCGCCCAGGTGTCTCTGAGGGTCGCGCCCCGATACCC TCCTCCACCCCAAGGCTTCTTCTCCCCAAGGGCCGGGCTGGGAGCCCAGG AGCGCCACAGCCCCCATGGGCCCAGGAAATTGCAGCTTCGTCATAGCTAA GCCCGCCGCGCTCAGCTTTTCCCACCACCTGAAGGATTCACTCTCCATCC GTCCTTCCCGACCTCAGGCTGCAGAGAAAACTTGGGGGCGGGGGTCGAGG AAGCCGCCCCCCGAGGCTCCCAGCGCAGCCCCAGGAGGAATCCTGCCGGC AACTAGAAGCTGTGCGCGTGGCCCGGCTCGGCGCCGACAACAGCCGGAGC GGCCGGGCCTGCAGCGAACCTCCGCATCCAGGTGGGGTCTCCGCGCCCCA ATTCCACCCCGCCCCGCCATCGGCACTCCCCGCAAGTGCGGGGTTTCCAC CCGCCCCGGCGTTGGCGCCCACTCACGGTCTCCCTCGCCCCCTCGGGAGA CGCTCCCCGCGTTCTATCCCGCCACGCCCTGGCGCGCCCCGCAAAAGATG TGACCCCGCCCCCCTGCGATCTCCGCTCCCCCGGACGCTGCCCCTGCCCC CACTTGGGGAGCTCCGGGCATCTCTAAAGCCCCCAACCAGAGTCCTCCAACCCAC

*PFKP* promoter region sequence with color annotation for *HhaI* condition: pink denotes CpGs methylated by *HhaI*, yellow denotes primer sequence

CTTTGGATGCGCGTATTTCAGCACGTGGAGTTGGCCAGTGGAGCCGG CCGAGGTCGACACAGCGTACCGCAGCGCGCGTGTCCGTGGGGAGGAAGGG TCGTCCCGCCTCCCTCCCAGGCCAGTCCTGTTGCGCCTCCTCCTGCCCCA CTCTGACTTTCGCGCGCCCAGGTGTCTCTGAGGGTCGCGCCCCGATACCC TCCTCCACCCCAAGGCTTCTTCTCCCCAAGGGCCGGGCTGGGAGCCCAGG AGCGCCACAGCCCCCATGGGCCCAGGAAATTGCAGCTTCGTCATAGCTAA GCCCGCCGCGCTCAGCTTTTCCCACCACCTGAAGGATTCACTCTCCATCC GTCCTTCCCGACCTCAGGCTGCAGAGAAAACTTGGGGGCGGGGGTCGAGG AAGCCGCCCCCCGAGGCTCCCAGCGCAGCCCCAGGAGGAATCCTGCCGGC AACTAGAAGCTGTGCGCGTGGCCCGGCTCGGCGCCGACAACAGCCGGAGC GGCCGGGCCTGCAGCGAACCTCCGCATCCAGGTGGGGTCTCCGCGCCCCA ATTCCACCCCGCCCCGCCATCGGCACTCCCCGCAAGTGCGGGGTTTCCAC CCGCCCCGGCGTTGGCGCCCACTCACGGTCTCCCTCGCCCCCTCGGGAGA CGCTCCCCGCGTTCTATCCCGCCACGCCCTGGCGCGCCCCGCAAAAGATG TGACCCCGCCCCCCTGCGATCTCCGCTCCCCCGGACGCTGCCCCTGCCCC CACTTGGGGAGCTCCGGGCATCTCTAAAGCCCCCAACCAGAGTCCTCCAACCCAC

# Cell culture and transfection procedures

K562 cells were maintained in RPMI 1640 medium (Sigma Aldrich) with 1% antibiotic-antimycotic and 10% fetal bovine serum. When cells were approximately 50% confluent, we washed and seeded them in pools of 25,000 cells for the transfection assays. Transient transfection was performed by adding 20 uL of OPTIMEM Reduced Serum Media (Gibco) containing the following reagents: (i) 100 ng of methylated, partially methylated, or unmethylated vector (4 replicates for each condition); (ii) 10 ng of Renilla control vector; (iii) 0.5 uL of Lipofectamine; and (iv) 0.1 uL of the PLUS reagent (from the Lipofectamine 2000 system, Life Technologies).

Cells were incubated for 24 hours following transfection, and subsequently assayed for transgene luciferase expression with the Dual Luciferase Assay kit (Promega). Firefly luciferase activity was normalized against Renilla activity to control for variation in the transfection efficiency or total number of cells in each experimental replicate.

**Using support vector machines to investigate the plasticity versus stability of DNA methylation levels**

To differentiate between the hypotheses presented in Figure 5, we used a machine learning approach. First, we built an SVM classifier that could distinguish between individuals that spent most or all of their lives in either a wild-feeding group or in Lodge group (n = 61 individuals) based on DNA methylation data alone. To do so, we used the *ksvm* function implemented in R package *kernlab*, using a linear kernel and setting the penalty constant, *C*, to 100 (Karatzoglou *et al.* 2004; note that our results were robust across several orders of magnitude of *C*). As predictive features for this model, we used the 334,840 CpG sites that were not associated with age, sex, bisulfite conversion rate, or sample age at a nominal p-value of 0.05. Because SVMs cannot work on binomially distributed count data, we used data that had been transformed to methylation proportions, transformed to a standard normal, and imputed to remove missing values (see Supporting Information: Transforming count data to normalized methylation proportions).

To assess the performance of the SVM model, we used leave-one-out cross-validation. Specifically, we iteratively (i) removed one individual from the data set, (ii) trained the SVM on DNA methylation data from equally sized samples of the remaining Lodge group individuals and wild-feeding individuals (to avoid biased estimates as a result of differences in class size); and (iii) used the resulting fitted model to predict the resource base of the originally removed test case. We repeated this procedure 610 times, so that DNA methylation data from every individual (n = 61) was used as the test case 10 times. We did not observe any consistent bias in class assignments during this procedure (49.02% of all misclassification events involved data from Lodge individuals, while 50.98% involved wild-feeding individuals).

Finally, to understand whether individuals that switched between resource bases more closely resembled their pre-switch or post-switch conspecifics, we repeated the same procedures described above, but used DNA methylation data from the 8 switching individuals as the test set. In this case, we used a fitted model trained on data (i.e., the n = 334,840 CpG sites not associated with any covariates) from the full set of Lodge individuals (minus switching individuals sampled in Lodge group), as well as an equally sized random sample of wild-feeding individuals (minus switching individuals sampled in a wild-feeding group). We used this fitted model to predict the resource base of the 8 individuals that dispersed between groups, using their DNA methylation data alone. We repeated this procedure 60 times to ensure that subsampling from the wild-feeding individuals did not bias our model predictions.

# References

1. Zou J, Lippert C, Heckerman D, Aryee M, Listgarten J. Epigenome-wide association studies without the need for cell-type composition. Nat Methods. 2014;11: 309–11. doi:10.1038/nmeth.2815

2. Lam LL, Emberly E, Fraser HB, Neumann SM, Chen E, Miller GE, et al. Factors underlying variable DNA methylation in a human community cohort. Proc Natl Acad Sci. 2012;109: 17253–60. doi:10.1073/pnas.1121249109

3. Lea A, Tung J, Zhou X. A flexible, efficient binomial mixed model for identifying differential DNA methylation in bisulfite sequencing data. bioRxiv. 2015; doi:http://dx.doi.org/10.1101/019562

4. Altmann J, Altmann S, Hausfater G. Physical maturation and age estimates of yellow baboons, Papio cynocephalus, in Amboseli National Park, Kenya. Am J Primatol. 1981;1: 389–399. doi:10.1002/ajp.1350010404

5. Alberts SC, Buchan JC, Altmann J. Sexual selection in wild baboons: from mating opportunities to paternity success. Anim Behav. 2006;72: 1177–1196. doi:10.1016/j.anbehav.2006.05.001

6. Buchan JC, Alberts SC, Silk JB, Altmann J. True paternal care in a multi-male primate society. Nature. 2003;425: 179–81. doi:10.1038/nature01866

7. Wang J. COANCESTRY: a program for simulating, estimating and analysing relatedness and inbreeding coefficients. Mol Ecol Resour. 2011;11: 141–5. doi:10.1111/j.1755-0998.2010.02885.x

8. Mukherjee S, Tamayo P, Rogers S, Rifkin R, Engle A, Campbell C, et al. Estimating dataset size requirements for classifying DNA microarray data. J Comput Biol. 2003;10: 119–142. doi:10.1089/106652703321825928

9. Reinius LE, Acevedo N, Joerink M, Pershagen G, Dahlén S-E, Greco D, et al. Differential DNA methylation in purified human blood cells: implications for cell lineage and studies on disease susceptibility. PLoS One. 2012;7: e41361. doi:10.1371/journal.pone.0041361

10. Jaffe AE. FlowSorted.Blood.450k: Illumina HumanMethylation data on sorted blood cell populations. R package version 1.5.1. 2015.

11. Jaffe AE, Irizarry R a. Accounting for cellular heterogeneity is critical in epigenome-wide association studies. Genome Biol. 2014;15: R31. doi:10.1186/gb-2014-15-2-r31

12. Cunningham F, Amode MR, Barrell D, Beal K, Billis K, Brent S, et al. Ensembl 2015. Nucleic Acids Res. 2014;43: D662–D669. doi:10.1093/nar/gku1010

13. Shulha HP, Cheung I, Guo Y, Akbarian S, Weng Z. Coordinated Cell Type–Specific Epigenetic Remodeling in Prefrontal Cortex Begins before Birth and Continues into Early Adulthood. Ren B, editor. PLoS Genet. 2013;9: e1003433. doi:10.1371/journal.pgen.1003433

14. Deng J, Shoemaker R, Xie B. Targeted bisulfite sequencing reveals changes in DNA methylation associated with nuclear reprogramming. Nat Biotechnol. 2009;27: 353–360. doi:10.1038/nbt.1530.Targeted

15. Karolchik D, Barber GP, Casper J, Clawson H, Cline MS, Diekhans M, et al. The UCSC Genome Browser database: 2014 update. Nucleic Acids Res. 2014;42: 764–770. doi:10.1093/nar/gkt1168

16. Hernando-Herraez I, Prado-Martinez J, Garg P, Fernandez-Callejo M, Heyn H, Hvilsom C, et al. Dynamics of DNA methylation in recent human and great ape evolution. PLoS Genet. 2013;9: e1003763. doi:10.1371/journal.pgen.1003763

17. Gu H, Smith ZD, Bock C, Boyle P, Gnirke A, Meissner A. Preparation of reduced representation bisulfite sequencing libraries for genome-scale DNA methylation profiling. Nat Protoc. 2011;6: 468–81. doi:10.1038/nprot.2010.190

18. Rönn T, Volkov P, Davegårdh C, Dayeh T, Hall E, Olsson AH, et al. A six months exercise intervention influences the genome-wide DNA methylation pattern in human adipose tissue. PLoS Genet. 2013;9: e1003572. doi:10.1371/journal.pgen.1003572

19. Dunham I, Kundaje A, Aldred SF, Collins PJ, Davis C, Doyle F, et al. An integrated encyclopedia of DNA elements in the human genome. Nature. 2012;489: 57–74. doi:10.1038/nature11247

20. Hinrichs A, Karolchik D, Baertsch R, Barber G, Bejerano G, Clawson H. The UCSC Genome Browser Database: update 2006. Nucleic Acids Res. 2006;34: D590–D598. doi:10.1093/nar/gkj144

21. Goeman JJ, Van de Geer S, De Kort F, van Houwellingen HC. A global test for groups fo genes: Testing association with a clinical outcome. Bioinformatics. 2004;20: 93–99. doi:10.1093/bioinformatics/btg382

22. Karatzoglou A, Smola A, Hornik K, Zeileis A. kernlab -- An S4 Package for Kernel Methods in R. J Stat Softw. 2004;11: 1–20.

23. Hastie T, Tibshirani R, Narasimhan B, Chu G. Impute: imputation for microarray data. R package version 1.42.0. 2015.

24. Jaffe AE, Murakami P, Lee H, Leek JT, Fallin MD, Feinberg AP, et al. Bump hunting to identify differentially methylated regions in epigenetic epidemiology studies. Int J Epidemiol. 2012;41: 200–209. doi:10.1093/ije/dyr238

25. Lister R, Pelizzola M, Dowen R, Hawkins R. Human DNA methylomes at base resolution show widespread epigenomic differences. Nature. 2009;

26. Hansen KD, Timp W, Bravo HC, Sabunciyan S, Langmead B, McDonald OG, et al. Increased methylation variation in epigenetic domains across cancer types. Nat Genet. Nature Publishing Group; 2011;43: 768–75. doi:10.1038/ng.865

27. Tung J, Zhou X, Alberts SC, Stephens M, Gilad Y. The genetic architecture of gene expression levels in wild baboons. eLife. 2015;4: 1–22. doi:10.7554/eLife.04729

28. Peng X, Thierry-Mieg J, Thierry-Mieg D, Nishida a., Pipes L, Bozinoski M, et al. Tissue-specific transcriptome sequencing analysis expands the non-human primate reference transcriptome resource (NHPRTR). Nucleic Acids Res. 2014;43: D737–D742. doi:10.1093/nar/gku1110

29. Jostins L, Ripke S, Weersma RK, Duerr RH, McGovern DP, Hui KY, et al. Host-microbe interactions have shaped the genetic architecture of inflammatory bowel disease. Nature. 2012;491: 119–124. doi:10.1038/nature11582.Host-microbe

30. Franke A, McGovern DPB, Barrett JC, Wang K, Radford-Smith GL, Ahmad T, et al. Genome-wide meta-analysis increases to 71 the number of confirmed Crohn’s disease susceptibility loci. Nat Genet. Nature Publishing Group; 2010;42: 1118–1125. doi:10.1038/ng.717

31. Estrada K, Krawczak M, Schreiber S, van Duijn K, Stolk L, van Meurs JBJ, et al. A genome-wide association study of northwestern Europeans involves the C-type natriuretic peptide signaling pathway in the etiology of human height variation. Hum Mol Genet. 2009;18: 3516–3524. doi:10.1093/hmg/ddp296

32. Vithana EN, Khor C-C, Qiao C, Nongpiur ME, George R, Chen L-J, et al. Genome-wide association analyses identify three new susceptibility loci for primary angle closure glaucoma. Nat Genet. 2012;44: 1142–1146. doi:10.1038/ng.2390

33. Estrada K, Styrkarsdottir U, Evangelou E, Hsu YH, Duncan EL, Ntzani EE, et al. Genome-wide meta-analysis identifies 56 bone mineral density loci and reveals 14 loci associated with risk of fracture. Nat Genet. 2012;44: 491–502. doi:10.1038/ng.2249

34. Soranzo N, Spector TD, Mangino M, Kühnel B, Rendon A, Teumer A, et al. A genome-wide meta-analysis identifies 22 loci associated with eight hematological parameters in the HaemGen consortium. Nat Genet. Nature Publishing Group; 2009;41: 1182–1190. doi:10.1038/ng.467

35. Gieger C, Kühnel B, Radhakrishnan a, Cvejic a, Serbanovic-Canic J, Meacham S, et al. New gene functions in megakaryopoiesis and platelet formation. Nature. Nature Publishing Group; 2011;480: 201–208. doi:10.1038/nature10659

36. Comuzzie AG, Cole S a., Laston SL, Voruganti VS, Haack K, Gibbs R a., et al. Novel Genetic Loci Identified for the Pathophysiology of Childhood Obesity in the Hispanic Population. PLoS One. 2012;7. doi:10.1371/journal.pone.0051954

37. Service SK, Verweij KJH, Lahti J, Congdon E, Ekelund J, Hintsanen M, et al. A genome-wide meta-analysis of association studies of Cloninger’s Temperament Scales. Transl Psychiatry. 2012;2: e116. doi:10.1038/tp.2012.37

38. Divaris K, Monda KL, North KE, Olshan a. F, Lange EM, Moss K, et al. Genome-wide Association Study of Periodontal Pathogen Colonization. J Dent Res. 2012;91: S21–S28. doi:10.1177/0022034512447951

39. Scuteri A, Sanna S, Chen WM, Uda M, Albai G, Strait J, et al. Genome-wide association scan shows genetic variants in the FTO gene are associated with obesity-related traits. PLoS Genet. 2007;3: 1200–1210. doi:10.1371/journal.pgen.0030115

# Figure S1. Flow chart describing data processing steps (light blue boxes) and main analyses (dark blue boxes).

# Macintosh HD:Users:amandalea:Dropbox:Amanda_files:alberts-tung lab:Lodge_DNAm_mar2014:May2015_final_analyses:diet_figures_part2.pdf

# Figure S2. Genomic compartment annotations used in this study. We tested for enrichment of resource-base associated sites in promoters, gene bodies, CpG islands, CpG island shores, enhancers, and unannotated regions of the genome. Below, we provide a cartoon depiction of these functional elements and their typical methylation status at an example gene. Methylated CpG sites are depicted as gray shaded lollipops, unmethylated CpG sites are depicted as white/unshaded lollipops, the gene body is depicted as a blue rectangle, and molecules that aid in transcriptional activation (e.g., transcription factors/activator proteins) are depicted as colored ovals. The promoter region is directly upstream of the gene body (defined as 2 kb upstream in our analyses), and is often associated with a CpG island (a dense cluster of CpG sites, usually unmethylated). CpG shores are defined as the 2 kb flanking CpG islands. Enhancer regions are short regions of DNA that often occur far from genes (although can also be found within or proximal to genes). Distal enhancers interact with promoter regions through DNA looping; they bind proteins (e.g., the green oval) that activate transcription. In our study, unannotated regions are defined as regions that do not fall into one of the five defined functional genomic compartments (promoters, gene bodies, CpG islands, CpG island shores, and enhancers). Such regions are generally hypermethylated.

# Figure S3. Histone marks and genomic compartments associated with individual chromatin states. (A) Histone mark data generated by the NIH Roadmap Epigenomics Project were used to define the 15 chromatin states used in this study (also produced by the Roadmap Epigenomics project). Each chromatin state is defined by the presence (dark blue square = strongly enriched; light blue square = weakly enriched) or absence (white square) of individual histone modifications (x axis labels). (B) We overlaid the Roadmap Epigenomics chromatin state annotations for human peripheral blood mononuclear cells onto the CpG sites tested in our data set and the genomic compartment annotations described in Figure S2. Here, we show the degree to which different chromatin states are more likely to occur in specific genomic compartments (dark purple square = strongly enriched; light purple square = weakly enriched; white square = not enriched).

****

**Figure S4. Differentially methylated sites are enriched near genes expressed in baboon whole blood.** In the main text, we report that differentially methylated sites (10% FDR) are more likely to occur in or near genes that are expressed in baboon whole blood, compared to genes that are unexpressed in this tissue (where “in or near” is defined as within the gene body or within 10 kb of the transcription start site, TSS, or transcription end site, TES). Here we report parallel results using alternative definitions for assigning CpG sites to genes, which result in analyses of different subsets of the data. These alternative definitions correspond to sites that occur: (i) 2 kb upstream, defined as <2 kb upstream of the TSS only (i.e., the putative promoter region); (ii) 10 kb upstream, defined as <10 kb upstream of the TSS only; (iii) within gene bodies only; or (iv) in CpG islands near genes, defined as in CpG islands within the gene body or within 10 kb of the gene TES or TSS. Below, we show the odds ratio from a Fisher’s exact test, asking whether differentially methylated sites are enriched near blood-expressed genes. Significant tests (p<0.05) are marked with a red asterisk, and the number of sites tested in each case is shown in parentheses.



# Figure S5. Effect of resource base on DNA methylation levels analyzed at the pathway level. We used the R package ‘GlobalTest’ [21] to test for a global effect of resource base on DNA methylation levels at CpG sites in or near genes in specific predefined pathways. We performed this test on 36 pathways related to the metabolism of food or to energy balance. Results of these analyses are shown here (blue = significant at a 10% FDR). For pathways that include CpG sites in or near *PFKP*, we also conducted a parallel analysis excluding all sites in or near this gene (denoted as red diamonds). When sites near *PFKP* are excluded from the analysis, the top three pathways do not show differential methylation at a 10% FDR threshold, and only the glycolysis and gluconeogenesis pathway remains significant at a nominal p-value of 0.05.

**Macintosh HD:Users:amandalea:Dropbox:Amanda_files:alberts-tung lab:Lodge_DNAm_mar2014:May2015_final_analyses:diet_figures.pdf**

# Figure S6. DMRs are observed in the real data set more often than expected by chance. We counted the number of sites that occur within 2 kb (i.e., ≤1 kb upstream or ≤1 kb downstream) of sites associated with resource base at a 10% FDR, considering only the 847 resource base-associated sites with at least 1 nearby site within the 2 kb window. The resulting distribution of sites is shown in blue; the distribution of values obtained from performing the same analysis on permuted data is shown in red. To obtain the permuted data distribution shown below, we analyzed four different permuted datasets and averaged the results.

Macintosh HD:Users:amandalea:Desktop:May2015_final_analyses:DMRs_nosingles_FDR20.pdf

**Figure S7. RRBS enriches for putatively functional regions of the genome and recapitulates known patterns of DNA methylation across the genome.** Here, we present quality control measures for RRBS data from the full sample set we analyzed (n=69 individuals; panels B, D, and F) and from the previously published data set (n=50 of the 69 total individuals; panels A, C, and E). **(A-B)** Proportion of total annotated features in the baboon genome for which a least one CpG site was analyzed. **(C-D)** Mean DNA methylation levels as a function of distance from the TSS, stratified by gene expression level quartiles obtained from whole blood RNA-seq for the same baboon population [27]. Only expressed genes (as identified by [27] were included in these analyses. As expected, more highly expressed genes exhibit lower levels of DNA methylation near the TSS. **(E-F)** Violin plots showing the distribution of average DNA methylation levels for CpG sites located in different genomic compartments. The white boxes indicate the interquartile range, and the black bars indicate the median DNA methylation level for each group of CpG sites. As expected, CpG islands, H3K3me1-marked enhancers and promoters tend to be lowly methylated, while gene bodies and the background set of all sites analyzed tend to be highly methylated. Note that the background set in RRBS data is highly biased towards functionally active regulatory elements, reducing mean/median methylation levels below true genome-wide values.

**Macintosh HD:Users:amandalea:Dropbox:Amanda_files:alberts-tung lab:Lodge_DNAm_mar2014:May2015_final_analyses:diet_figures_part2.pdf**

**Figure S8. Power to detect differentially methylated sites (between lifelong wild-feeding and Lodge individuals) increases with sample size.** Using data from individuals that spent all or the majority of their lives in one resource base condition (either wild-feeding or Lodge: n=61), we estimated the relationship between sample size and power to detect putative true positive sites (x-axis = sample size; y-axis = proportion of putative true positive sites detected at a 10% FDR). Results are stratified by quartiles of effect sizes (e.g., Q1 shows putative true positive sites with effect sizes in the the top 25% of our data set). There appear to be many true resource base-associated sites in our data set that do not pass genome-wide significance, especially for small effect sizes.



# Figure S9. Magnitude of the effect of resource availability on DNA methylation levels in different genomic compartments. The cumulative distribution function is shown for betas (effect sizes) generated by MACAU. Each line represents the distribution of betas associated with the effect of resource base on DNA methylation levels in a given genomic compartment. Only CpG sites with a significant effect of resource base (10% FDR) are shown. If the direction of the effect of resource base was random (i.e., methylation levels increased in Lodge versus wild animals with equal probability), we would expect all lines to pass through the intersection of the black dotted lines (at x = 0 and y = 0.5). Instead, promoters and enhancers are somewhat more likely to show increased methylation in Lodge animals, while all other regions are more likely to show decreased methylation in Lodge animals.

**Figure S10. Cell type proportions did not significantly differ between wild-feeding and Lodge individuals.** The distribution of cell-type proportions, obtained from manual counts of Giemsa-stained blood smears, are shown below (n = 15 Lodge individuals and 25 wild-feeding individuals). For the five major cell types we measured, cell type proportions did not differ between the two resource bases (p-values are from a generalized linear model with a binomial link function, controlling for age, sex, and the identity of the individual who scored the blood slide).Note that lymphocytes and neutrophils make up by far the largest proportion of blood cell types in whole blood.



Figure S11. Enrichment of resource base-associated sites is strongest near genes expressed in whole blood, compared to genes expressed in other tissues.We report in the main text that differentially methylated sites are enriched near genes expressed in whole blood. This result could be due to targeted regulation of blood-expressed genes, or arise simply because sites affected by diet fall near genes expressed across many tissues. Here, we used tissue specific RNA-seq data from olive baboons to identify genes expressed (FPKM > 1) or unexpressed (FPKM < 1) in a range of tissues [28]. We then asked, for each tissue, whether resource base-associated sites were enriched near expressed genes, using a Fisher’s Exact Test. We considered a CpG site to be near a gene if it fell in the gene body or within 10kb of the transcription start or end sites. The FET odds ratio is plotted for each tissue. Differentially methylated sites are significantly biased towards genes expressed in all tissues except skeletal muscle (at a nominal p-value of 0.05), but are most strongly biased for whole blood.

Macintosh HD:Users:amandalea:Desktop:May2015_final_analyses:tissue_exp_genes:odds_ratio.pdf

# Table S1. Information about males that switched between resource base conditions (n=8 baboons).

|  |  |  |  |
| --- | --- | --- | --- |
| Individual | Resource base (natal/adult)1 | Years in post-dispersal resource base condition2 | Certainty level for years in post-dispersal condition3 |
| AMB\_01 | L/W | 3.07 | Known |
| AMB\_05 | W/L | 4.90 | Lower bound |
| AMB\_17 | L/W | 2.22 | Known |
| AMB\_30 | W/L | 4.92 | Lower bound |
| AMB\_38 | W/L | 5.03 | Lower bound |
| AMB\_42 | W/L | 4.94 | Lower bound |
| AMB\_62 | W/L | 4.94 | Lower bound |
| AMB\_67 | L/W | 2.11 | Known |

1 L = Lodge group; W = wild-feeding group

2 Number of years the male resided in the post-dispersal resource base group prior to blood sample collection

3 For individuals that switched from the Lodge group to a wild-feeding group, the timing of dispersal events and group residency are known. For individuals that switched from a wild-feeding group to the Lodge group, early histories were inferred (see main text) and the precise timing of their switch from wild-feeding to Lodge is unknown. For these individuals, we provide the number of years they were directly observed in the Lodge group, which serves as a lower bound for the total number of years they experienced the Lodge resource base prior to blood sampling.

# Table S2. Baboon RRBS data set sample characteristics and read mapping summary.

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| Individual | Sex | Age of animal (years) | Bisulfite conversion rate1 | Sample age (years)2 | Total reads generated (in millions) | Uniquely mapped reads (in millions) | Resource base (natal/adult)3 |
| AMB\_01 | M | 11.29 | 0.9850 | 8.39 | 37.023 | 25.100 | L/W |
| AMB\_02 | F | 10.06 | 0.9994 | 24.30 | 33.072 | 22.820 | L/L |
| AMB\_03 | M | 7.67 | 0.9842 | 6.37 | 24.088 | 16.943 | W/W |
| AMB\_04 | M | 5.40 | 0.9988 | 20.20 | 14.729 | 10.458 | L/L |
| AMB\_05 | M | 18.01 | 0.9849 | 25.22 | 51.052 | 35.687 | W/L |
| AMB\_06 | M | 6.39 | 0.9847 | 25.16 | 21.887 | 14.800 | W/W |
| AMB\_07 | M | 6.85 | 0.9840 | 4.13 | 14.934 | 10.013 | W/W |
| AMB\_08 | M | 7.92 | 0.9988 | 25.21 | 32.612 | 22.532 | W/W |
| AMB\_09 | M | 5.16 | 0.9994 | 25.13 | 14.677 | 10.610 | W/W |
| AMB\_10 | M | 6.25 | 0.9837 | 25.21 | 35.170 | 23.064 | W/W |
| AMB\_11 | F | 14.56 | 0.9995 | 25.16 | 18.719 | 13.103 | L/L |
| AMB\_12 | M | 3.98 | 0.9837 | 25.14 | 26.056 | 17.660 | L/L |
| AMB\_13 | M | 6.01 | 0.9840 | 25.13 | 24.440 | 16.309 | W/W |
| AMB\_14 | M | 3.76 | 0.9989 | 25.16 | 20.660 | 14.073 | L/L |
| AMB\_15 | F | 9.53 | 0.9989 | 25.19 | 9.586 | 7.285 | L/L |
| AMB\_16 | F | 7.84 | 0.9994 | 22.64 | 18.432 | 12.718 | L/L |
| AMB\_17 | M | 11.01 | 0.9990 | 25.15 | 18.549 | 12.903 | L/W |
| AMB\_18 | M | 15.79 | 0.9990 | 6.29 | 36.645 | 25.193 | W/W |
| AMB\_19 | M | 3.04 | 0.9990 | 25.07 | 31.059 | 21.321 | W/W |
| AMB\_20 | M | 4.50 | 0.9990 | 25.13 | 29.389 | 20.758 | W/W |
| AMB\_21 | F | 6.71 | 0.9995 | 6.30 | 28.666 | 19.779 | W/W |
| AMB\_22 | F | 5.23 | 0.9994 | 25.17 | 16.784 | 12.084 | W/W |
| AMB\_23 | M | 9.79 | 0.9963 | 7.42 | 11.771 | 8.084 | W/W |
| AMB\_24 | M | 4.27 | 0.9987 | 20.20 | 24.483 | 16.747 | L/L |
| AMB\_25 | M | 6.00 | 0.9986 | 21.16 | 71.814 | 42.517 | W/W |
| AMB\_26 | M | 1.76 | 0.9987 | 25.09 | 15.461 | 10.784 | L/L |
| AMB\_27 | M | 5.98 | 0.9987 | 25.20 | 31.122 | 21.156 | W/W |
| AMB\_28 | M | 8.29 | 0.9980 | 25.20 | 35.575 | 24.680 | W/W |
| AMB\_29 | M | 4.80 | 0.9981 | 25.16 | 35.878 | 25.526 | L/L |
| AMB\_30 | M | 14.01 | 0.9980 | 25.21 | 15.382 | 10.708 | W/L |
| AMB\_31 | M | 2.90 | 0.9980 | 25.13 | 34.860 | 24.045 | L/L |
| AMB\_32 | M | 14.30 | 0.9980 | 8.34 | 21.900 | 16.169 | W/W |
| AMB\_33 | F | 5.03 | 0.9988 | 20.88 | 20.593 | 14.621 | L/L |
| AMB\_34 | F | 6.13 | 0.9963 | 25.16 | 39.121 | 27.385 | W/W |
| AMB\_35 | F | 3.96 | 0.9994 | 25.09 | 19.536 | 13.870 | W/W |
| AMB\_36 | M | 6.76 | 0.9978 | 25.18 | 39.791 | 27.011 | W/W |
| AMB\_37 | M | 6.11 | 0.9978 | 25.20 | 41.871 | 29.169 | W/W |
| AMB\_38 | M | 14.01 | 0.9978 | 25.13 | 23.945 | 18.063 | W/L |
| AMB\_39 | F | 8.10 | 0.9994 | 24.25 | 19.089 | 13.553 | L/L |
| AMB\_40 | F | 4.97 | 0.9995 | 25.18 | 22.715 | 15.673 | L/L |
| AMB\_41 | F | 3.49 | 0.9988 | 24.25 | 37.165 | 26.015 | L/L |
| AMB\_42 | M | 18.01 | 0.9977 | 25.22 | 30.003 | 21.484 | W/L |
| AMB\_43 | F | 4.69 | 0.9994 | 24.28 | 27.103 | 18.972 | L/L |
| AMB\_44 | M | 5.80 | 0.9990 | 25.22 | 23.953 | 16.974 | L/L |
| AMB\_45 | F | 16.44 | 0.9995 | 23.27 | 16.226 | 11.683 | L/L |
| AMB\_46 | F | 4.01 | 0.9964 | 25.13 | 53.669 | 37.032 | W/W |
| AMB\_47 | M | 3.64 | 0.9990 | 25.08 | 30.674 | 20.747 | W/W |
| AMB\_48 | M | 10.62 | 0.9991 | 6.30 | 37.266 | 26.408 | W/W |
| AMB\_49 | M | 11.86 | 0.9987 | 5.92 | 29.500 | 20.155 | W/W |
| AMB\_50 | M | 6.72 | 0.9988 | 24.30 | 79.784 | 54.079 | L/L |
| AMB\_51 | F | 4.92 | 0.9989 | 4.58 | 13.903 | 9.732 | W/W |
| AMB\_52 | F | 7.54 | 0.9988 | 25.16 | 30.061 | 22.546 | L/L |
| AMB\_53 | M | 2.15 | 0.9843 | 25.11 | 29.804 | 20.001 | L/L |
| AMB\_54 | F | 4.14 | 0.9964 | 25.13 | 16.493 | 11.545 | W/W |
| AMB\_55 | M | 7.43 | 0.9855 | 7.40 | 13.354 | 9.737 | W/W |
| AMB\_56 | M | 9.19 | 0.9955 | 8.30 | 11.922 | 8.275 | W/W |
| AMB\_57 | F | 5.95 | 0.9995 | 6.62 | 9.902 | 7.228 | W/W |
| AMB\_58 | F | 7.73 | 0.9966 | 4.15 | 33.865 | 23.028 | W/W |
| AMB\_59 | M | 5.44 | 0.9988 | 2.28 | 11.851 | 8.533 | W/W |
| AMB\_60 | M | 6.26 | 0.9990 | 4.96 | 29.918 | 20.598 | W/W |
| AMB\_61 | M | 2.59 | 0.9988 | 24.23 | 22.170 | 15.360 | W/W |
| AMB\_62 | M | 18.01 | 0.9981 | 25.22 | 36.660 | 25.248 | W/L |
| AMB\_63 | M | 4.50 | 0.9994 | 25.16 | 11.387 | 8.623 | W/W |
| AMB\_64 | M | 6.49 | 0.9977 | 5.64 | 17.274 | 11.996 | W/W |
| AMB\_65 | F | 9.24 | 0.9995 | 7.39 | 27.200 | 19.040 | W/W |
| AMB\_66 | M | 4.72 | 0.9990 | 25.18 | 61.599 | 41.574 | L/L |
| AMB\_67 | M | 7.89 | 0.9990 | 25.13 | 17.518 | 12.255 | L/W |
| AMB\_68 | F | 2.92 | 0.9995 | 24.29 | 18.857 | 13.200 | W/W |
| AMB\_69 | M | 7.58 | 0.9987 | 7.29 | 17.789 | 11.953 | W/W |
|  |  |  |  |  |  |  |  |
| Mean |  | 7.45 | 0.9965 | 19.53 | 27.2465 | 18.8260 |  |
| Standard deviation |  | 4.01 | 0.0051 | 8.38 | 13.6195 | 8.9361 |  |

1 To calculate bisulfite conversion rates, we mapped the sequencing data, for each individual separately, to the lambda phage genome. We then summed (i) the number of reads that mapped to lambda phage CpG sites and were read as thymine (reflecting an unmethylated cytosine converted to thymine); and (ii) the total number of reads that mapped to lambda phage CpG sites. Because all CpG sites in the lambda phage genome were completely unmethylated (and should thus have been converted to thymine), the ratio of these two sums gives us the efficiency of the bisulfite conversion. Here, a ratio of 1 would represent perfect conversion of every unmethylated cytosine to thymine.

2 Years from collection of blood sample to RRBS library construction

3 L = Lodge group; W = wild-feeding group

# Table S3. Differentially methylated regions.

|  |  |  |  |
| --- | --- | --- | --- |
| DMR coordinates | Differentially methylated sites (10% FDR) | Closest gene (within 100kb) | Phenotypes associated with genetic variation at this gene (in humans) |
| Chr1:  17489362 - 17491377 | 3 | KAZN | None |
| Chr1:  17497885 - 17499900 | 3 | KAZN | None |
| Chr1:  114729892 - 114731900 | 2 | DCLRE1B | Crohn’s disease [29] |
| Chr11:  128022967 - 128024967 | 1 | UBAP2L | None |
| Chr11:  129099353 - 129101353 | 1 | SCAMP3 | Crohn’s disease [30] |
| Chr11:  131300175 - 131302175 | 1 | GALNT9 | None |
| Chr11:  131420650 - 131422650 | 1 | GALNT9 | None |
| Chr11:  131484840 - 131486840 | 1 | GALNT9 | None |
| Chr12:  94220922 - 94876852 | 3 | ALPP/DIS3L2 | Height [31] |
| Chr13:  40548003 - 40550240 | 3 | GAPDH | None |
| Chr13:  54397960 - 54399960 | 1 | CLHC1 | None |
| Chr13:  79975255 - 79977255 | 1 | ANKRD11 | None |
| Chr15:  3447168 - 4033270 | 3 | RXRB/COL11A1/COL5A1 | Glaucoma [32]; Bone mineral density [33] |
| Chr15:  11338900 - 11341350 | 5 | CRB2 | None |
| Chr19:  1287252 - 1289259 | 2 | PLK5 | None |
| Chr19:  14706240 - 14708325 | 12 | TPM1 | Mean platelet volume [34]; Platelet count [35] |
| Chr19:  38896370 - 38898401 | 2 | NKPD1 | None |
| Chr19:  42063798 - 42065798 | 1 | BCAT2 | None |
| Chr5:  15905159 - 15907165 | 3 | KCNIP4 | Obesity-related traits [36] |
| Chr8:  2948895 - 2950895 | 1 | CSMD3 | Temperament [37]; Periodontal microbiota [38] |
| Chr8:  6806879 - 6808879 | 1 | ZNF705D | None |
| Chr9:  3192817 - 3194825 | 2 | PFKP | Obesity [39] |
| Chr9:  3209124 - 3213314 | 30 | PFKP | Obesity [39] |
| Chr10:  84590364 - 84592364 | 1 | KIAA1644 | None |
| Chr20:  67076699 - 67078699 | 1 | None | NA |
| Chr7:  157297300 - 157299300 | 1 | MEG9 | None |
| Chr6:  173378126 - 173380126 | 1 | RASGEF1C | None |