

UCLA

UCLA Previously Published Works

Title

The concerted impact of domestication and transposon insertions on methylation patterns between dogs and grey wolves

Permalink

<https://escholarship.org/uc/item/06p8k9fh>

Journal

Molecular Ecology, 25(8)

ISSN

0962-1083

Authors

Janowitz Koch, Ilana

Clark, Michelle M

Thompson, Michael J

et al.

Publication Date

2016-04-01

DOI

10.1111/mec.13480

Peer reviewed



Published in final edited form as:

Mol Ecol. 2016 April ; 25(8): 1838–1855. doi:10.1111/mec.13480.

The concerted impact of domestication and transposon insertions on methylation patterns between dogs and gray wolves

Ilana Janowitz Koch¹, Michelle M. Clark², Michael J. Thompson³, Kerry A. Deere-Machemer¹, Jun Wang⁴, Lionel Duarte², Gitanjali E. Gnanadesikan¹, Eskender L. McCoy⁵, Liudmilla Rubbi³, Daniel R. Stahler⁶, Matteo Pellegrini³, Elaine A. Ostrander⁷, Robert K. Wayne⁸, Janet S. Sinsheimer^{2,9}, and Bridgett M. vonHoldt¹

¹Ecology and Evolutionary Biology, Princeton University, Princeton, NJ 08544

²Department of Biostatistics, UCLA Fielding School of Public Health, University of California, Los Angeles, Los Angeles, CA 90095

³Molecular, Cell, and Developmental Biology, University of California, Los Angeles, Los Angeles, CA 90095

⁴Department of Biological Sciences, Wayne State University, Detroit, MI 48085

⁵Yale School of Management, Yale University, New Haven, CT, 06511

⁶Yellowstone Center for Resources, National Park Service, Yellowstone National Park, WY 82190

⁷National Human Genome Research Institute, National Institutes of Health, Bethesda, MD, 20892

⁸Ecology and Evolutionary Biology, University of California, Los Angeles, Los Angeles, CA 90095

⁹ Department of Human Genetics and Biomathematics, David Geffen School of Medicine at UCLA, Los Angeles CA 90095

Abstract

The process of domestication can exert intense trait-targeted selection on genes and regulatory regions. Specifically, rapid shifts in the structure and sequence of genomic regulatory elements could provide an explanation for the extensive, and sometimes extreme, variation in phenotypic

Bridgett vonHoldt, Princeton University, Princeton, NJ 08544; vonHoldt@princeton.edu.

Author Contributions

IJK and BMV designed the study, performed analyses, and wrote the manuscript. BMV and LR constructed the RRBS libraries. MJT analyzed RRBS data for quality, and assisted with mapping and computing cytosine methylation levels. MMC and JSS conducted heritability analyses and contributed to the discussion. LD assisted with heritability analyses. KDM and JW conducted neutrality analyses and contributed to the discussion; KDM also assisted with mapping samples to the dog genome. JSS provided statistical approaches to assess the neutrality analyses. GEG and ELM conducted initial analyses of differential methylation in dogs and wolves. DRS, EAO, and RKW provided blood samples and manuscript edits. MP, DRS, EAO, and RKW were an integral part of interpretation of analyses and discussion of results.

Data Accessibility

BS-Seeker2 log files, epigenotype variant files (used for ADMIXTURE, dendrograms, and PCA), and GO results/input files are accessioned onto Dryad (doi:10.5061/dryad.q2g6h). Raw Illumina Reads are accessioned onto NCBI's Sequence Read Archive (SRA) SRP065319; BioSample Accessions numbers 4217607-4217698. Bam files are accessioned onto SRA SRP065666: BioSample Accessions numbers 4217607-4217698.

traits observed in domesticated species. Here, we explored methylation differences from >24,000 cytosines distributed across the genomes of the domesticated dog (*Canis familiaris*) and the gray wolf (*C. lupus*). PCA and model-based cluster analyses identified two primary groups, domestic versus wild canids. A scan for significantly differentially methylated sites (DMSs) revealed species-specific patterns at 68 sites after correcting for cell heterogeneity, with weak yet significant hyper-methylation typical of purebred dogs when compared to wolves (59% and 58%, $p < 0.05$, respectively). Additionally, methylation patterns at eight genes significantly deviated from neutrality, with similar trends of hyper-methylation in purebred dogs. The majority (>66%) of differentially methylated regions contained or were associated with repetitive elements, indicative of a genotype-mediated trend. However, DMSs were also often linked to functionally relevant genes (e.g. neurotransmitters). Finally, we utilized known genealogical relationships among Yellowstone wolves to survey transmission stability of methylation marks, from which we found a substantial fraction that demonstrated high heritability (both H^2 and $h^2 > 0.99$). These analyses provide a unique epigenetic insight into the molecular consequences of recent selection and radiation of our most ancient domesticated companion, the dog. These findings suggest selection has acted on methylation patterns, providing a new genomic perspective on phenotypic diversification in domesticated species.

Keywords

Methylation; genome regulation; domestication; canid

Introduction

As a result of intense trait-targeted selection over a relatively small number of generations, the process of domestication may result in extreme phenotypic diversity (Darwin 1868; Clutton-Brock 1992; Clutton-Brock 1999; Diamond 2002). Recently, the molecular underpinnings of this diversification has been the subject of intense investigation (Ostrander & Wayne 2005; Dobney & Larson 2006; Wiener & Wilkinson 2011; Albert *et al.* 2012; Andersson 2013; Larson *et al.* 2014). Various approaches have been used to unravel the genetic changes that occur during domestication, most notably genome-wide SNP and haplotype analyses to map regions under selection and associate those regions with phenotypes (Andersson & Georges 2004; Doebly *et al.* 2006; Wiener & Wilkinson 2011; Wayne & vonHoldt 2012).

Changes in genomic regulatory elements or transcriptional machinery could explain a significant component of the dramatic evolution of phenotypic diversity in dogs and other domesticated species (Andersson & Georges 2004). Past studies are limited and have generally been transcriptome-based, focusing on differential expression analyses of brain tissues implicated in behavior and brain development (Saetre *et al.* 2004; Kukekova *et al.* 2011; Albert *et al.* 2012), as well as the impact of captivity on genes involved in the stress response (Kennerly *et al.* 2008). More recently, the use of epigenome-wide association studies (EWAS) have provided a new profile of the epigenomic changes that occur during domestication (Feeney *et al.* 2014).

Cytosine methylation serves as one mechanism for pre-transcriptional regulatory control and is involved in essential cellular processes, such as cell differentiation, X-inactivation, and genomic imprinting (Bernstein *et al.* 2007; Suzuki & Bird 2008; Richards *et al.* 2010). Methylated cytosines in 5' promoter regions act as transcriptional suppressors by inhibiting the transcriptional complex from binding to the promoter sequence (Bird 2007). New sequencing approaches can now interrogate the methylation state of each cytosine, specifically cytosines in CpG islands (CGIs) comprising putatively regulatory elements and representing a potentially functional fraction of the mammalian genome (Saxonov *et al.* 2006). In mammals, these islands are often found in association with promoter elements in genes (Deaton & Bird 2011) and epigenetic modifications most frequently occur within the CG dinucleotide motif (Bernstein *et al.* 2007).

Here, we investigate the methylation variation in the domesticated dog and its wild predecessor, the gray wolf. The dog is considered the oldest domesticated species, providing a rich system for investigation of the evolutionary changes that occur during the domestication process (Sutter & Ostrander 2004; Boyko 2011; Shearman & Wilton 2011; Wayne & vonHoldt 2012; Freedman *et al.* 2014). Remarkably, the extreme genetic and phenotypic variation observed within dog breeds arose in a short time period, initiating 15,000 to 40,000 years ago (Sablin & Khlopachev 2002; Savolainen *et al.* 2002; Germonpré *et al.* 2009; Boyko *et al.* 2010; Vonholdt *et al.* 2010; Ovodov *et al.* 2011; Larson *et al.* 2012; Freedman *et al.* 2014; Skoglund *et al.* 2015). Most modern dog breeds, however, have emerged from controlled breeding practices of the last 200 years involving narrow founding, genetic isolation, and strong phenotypic selection according to breed standards (Parker *et al.* 2004; Vonholdt *et al.* 2010). This unique history has created a genetic architecture that is extremely effective for unraveling quantitative phenotypic traits such as body size (Parker *et al.* 2004; Wayne & Ostrander 2007).

We begin by comparing domestic dogs to the pedigreed population of gray wolves in Yellowstone National Park (YNP) (Vonholdt *et al.* 2008). Recent studies have revealed many aspects of the Yellowstone population, from their genealogical history and trait evolution (Vonholdt *et al.* 2008; Anderson *et al.* 2009), to disease epidemiology (Almberg *et al.* 2009), life history (MacNulty *et al.* 2009a; MacNulty *et al.* 2009b; MacNulty *et al.* 2011; Stahler *et al.* 2013; Cubaynes *et al.* 2014), and population viability (Coulson *et al.* 2011). We build upon the rich literature on domestic dogs (e.g. heritability estimates, linkage analyses, genome-wide association studies, candidate gene approaches) through the addition of the canine epigenome, an ideal tool to investigate the molecular signature of recent rapid radiation and intense selective pressure. In addition, the samples we analyze include a wide selection of modern and ancient dogs breeds, as well as village dogs, which are dogs whose breeding is not well controlled and represent a mix of modern and ancient lineages (Boyko *et al.* 2010). Finally, using the Yellowstone wolf genealogy, we estimate the heritability of each methylation mark (Bell *et al.* 2012).

Methods

Samples and reduced representation bisulfite sequencing

Genomic DNA (1µg) was extracted from whole blood from 50 domestic dogs from 19 recognized American Kennel Club (AKC) purebred dog breeds (further referred to as purebred dogs), two Boz breed dogs from the Turkish Caucasus, two aboriginal island breeds (dingoes, n=2; New Guinea Singing Dogs, ngsd, n=4), and three village dogs. Similarly, we included 35 gray wolves (*C. lupus*) from YNP (Supplemental Table S1). All DNA was prepared using QIAamp DNA mini kits following manufacturer's protocol (Qiagen, DNeasy blood and Tissue kit). Whole blood is both a reliable proxy for organism-level transcriptional activity (Liew *et al.* 2006; Kaminsky *et al.* 2009; Rakyan *et al.* 2011) and methylation marks are preserved and stable over time in frozen blood, unlike RNA collections (Ferrer *et al.* 2008; Lofton-Day *et al.* 2008; Van der Auwera *et al.* 2009; Wiczorek *et al.* 2009). DNA was quantified using a Qubit 2.0 Fluorometer and checked on a 2% agarose gel for degradation.

We used reduced representation bisulfite sequencing (RRBS) to explore methylation patterns at CGIs, a single-base resolution methylation analysis from the sequencing of bisulfite converted DNA of CpG-rich regions (Meissner *et al.* 2005; Saxonov *et al.* 2006). We constructed RRBS genomic libraries for sequencing on an Illumina HiSeq2500. High purity genomic DNA was digested with *MspI*, ends were repaired with adenine on the 3' ends followed by ligation of adapters with unique 6mer barcode sequence tags using the TruSeq Sample Preparation kit's standard protocol (Illumina Inc). Library purification and size selection were completed using Agencourt AMPure XP beads (Beckman Coulter). We selected fragments between 100-400bp in size then treated them with bisulfite using an EpiTect kit (Qiagen) for conversion of unmethylated cytosines, following the manufacturer's procedure for low DNA input (1 ng-2 µg) with a double incubation time of 10 hours to ensure a high conversion rate. Our converted DNA was then subjected to 12 PCR cycles of amplification for enrichment of adapter-ligated fragments with MyTaq Mix (Bioline Inc). Library quality was assessed with an Agilent 2100 BioAnalyzer (Agilent) and concentration estimated using a Qubit 2.0 Fluorometer. Libraries were then standardized to 10nM, and then pooled on average four samples per lane for sequencing (minimum 2, maximum 6). To avoid generating redundant methylation information, single-end reads were chosen over paired-end reads for sequencing. Single-end DNA sequencing (1×101nt) was performed using the Illumina HiSeq 2500.

RRBS data quality, read mapping, and cytosine methylation levels

We used *bowtie2* in *BS-Seeker 2* to map our reads to the reference dog genome (canfam3.1, (Lindblad-Toh *et al.* 2005)) for read lengths bounded from 50-300 bp (Chen *et al.* 2010). Per barcode pool, there were 3.25 mismatches (min=3, max=5). Therefore, we deplexed pooled sequence data with a custom perl script with a maximum of two mismatches between expected and observed barcode sequence tags. FASTQ files were then trimmed for low quality reads and adapter sequences clipped using *Trim_Galore* v0.3.7, discarding reads that were <20 bp in length (Krueger 2013). Trimmed reads were mapped to the reference dog genome using *bowtie2* (Langmead 2010; Langmead & Salzberg 2012) and methylation

called using *BS-Seeker2* (Chen *et al.* 2010). Bowtie aligns C/T converted reads to C/T converted strands. During post processing, the number of user-defined mismatches is counted with the exception of those between read Ts and genomic Cs. Low-quality mappings with user-defined mismatches are then removed.

We calculated the methylation frequency (MF) as the proportion of methylated cytosines out of the total number of methylated and unmethylated cytosine (sequenced as thymine) reads per site (Chen *et al.* 2010). For a given frequency, we used a 95% confidence interval threshold value of 0.33 to calculate frequency confidence. Our confidence interval threshold (the maximum range we accept for the 95% confidence calculation) was a function of the number of counts at a particular position, so that for every frequency we used, we were 95% confident that our frequency was real within a window of 0.33 around that number. We also excluded cytosines within the CHG and CHH motifs (Zemach *et al.* 2010). Cytosines were annotated as being either intergenic or in a promoter (within 5Kb of transcriptional start site; (Agirre *et al.* 2015)), exon, or intron. Additionally, we used *RepeatMasker* (Smit *et al.* 1996) to classify cytosines within canine repeat elements (e.g. LTRs, simple repeats).

Conversion to single methylation polymorphism (SMPs)

To convert methylation frequency into a diploid genotype, we discretized methylation values into epigenotypes: sites with little or no methylation ($MF < 0.33$) were annotated as AA epigenotypes, sites with high or fixed methylation ($MF > 0.66$) were annotated as BB epigenotypes, and sites with intermediate methylation levels ($0.33 \leq MF \leq 0.66$) were annotated as AB epigenotypes (Heyn *et al.* 2013; Wang & Fan 2015). This allowed for the downstream application of traditional population genetic analyses (e.g. Principal Components Analysis, PCA, Bayesian-clustering, association scans) on methylation data through their conversion to diploid genotypes where each epiallele retains some broad level of methylation information (e.g. A epiallele is hypo-methylated; B epiallele is hyper-methylated).

Identification of outlier samples

To assess the samples for quality and general patterns, we employed two methods on our set of 96 samples: 1) hierarchical cluster analysis on methylation frequency data and 2) PCA of SMP data. First, we constructed a Gower pairwise dissimilarity matrix among all samples for autosomes and X chromosome separately with the R function *daisy* in the *cluster* package (Gower 1971; Struyf *et al.* 1997; Kaufman & Rousseeuw 2009). From this matrix, we then performed hierarchical clustering using Ward's method with the *hclust* function in R (Gordon 1999). Similarly, we conducted a principal component analysis (PCA). However, the nature of RRBS means that we sequence contiguous blocks of nucleotides, thus containing a high degree of redundant genetic information. To avoid violating the assumption of unlinked and independent loci when analyzing SMPs, we chose to use genotypic correlation as a measure of intra-locus linkage through non-random association (e.g. physical linkage). We used *PLINK* (Purcell *et al.* 2007) to prune SMPs in linkage disequilibrium (LD) with a genotype correlation of $r^2 > 0.2$ using the parameter flag `--indep-pairwise 50 5 0.2`. We used *flashPCA* (Abraham & Inouye 2014) for PCA as our primary method to identify outlier samples for exclusion in subsequent analyses.

Population genetics of SMPs

To explore the patterns of SMP variation, we analyzed all canid samples after outliers were removed. We surveyed structure using a clustering method with an explicit population genetics model based on Hardy-Weinberg expectations with the program *ADMIXTURE* and conducted a cross validation with the `--cv` flag (Alexander *et al.* 2009). With this method, each individual is assigned an ancestry proportion based on the number of assumed genetic clusters ($K=2-8$). Additionally, it assumes individuals are unrelated and loci are unlinked. We used *ADMIXTURE* to explore the relatedness of the YNP wolves. The purebred dogs are known to be unrelated at the grandparent level, while all family history data are lacking for the non-purebred dog samples. To identify loci that are associated with the domesticated dog phenotype, we conducted an association scan using *PLINK* and 1,307 unlinked SMPs across the purebred dogs ($n=47$) and YNP wolves ($n=33$) with the `--assoc` flag and a case/control phenotype. This is only a small fraction of the methylation dataset due to the nature of the RRBS approach to capture physical clusters of cytosines in CpG islands. Association p -values were converted into $-\log_{10}(p)$ scores and displayed as Manhattan plots using the *qqman* R function (Turner 2014).

Deviations from neutrality with SMP data

We estimated methylation-specific neutrality statistics with a recently developed method based on a Tajima's D framework with the specific application to SMP data. We first estimated the statistic D^m (Wang & Fan 2015) which detects methylation mutations likely under selection at a particular coding locus. For each gene body represented by sites located within exons and/or introns, we calculated a D^m test to assess potential selection on methylation mutations. $D^m = (\widehat{\theta}_\pi - \widehat{\theta}_s) / \text{standard deviation}(\widehat{\theta}_\pi - \widehat{\theta}_s)$. $\widehat{\theta}_\pi$ and $\widehat{\theta}_s$ are two estimators of the mutation parameter, $\widehat{\theta}(\widehat{\theta} = 4 * N * \mu)$. $\widehat{\theta}_\pi$ is the $\widehat{\theta}$ value estimated by the average methylation state difference per cytosine site. $\widehat{\theta}_s$ is the $\widehat{\theta}$ value estimated by the proportion of methylation segregation sites. When there is an excess of intermediate-frequency SMPs, $\widehat{\theta}_\pi$ will have a large value, and D^m tends to be positive. When there is an excess of low-frequency SMPs, $\widehat{\theta}_\pi$ will have a small value, and D^m tends to be negative. By theoretic analysis and simulation, it has been shown that directional selection in DNA methylation could generate an excess of intermediate-frequency epialleles (positive D^m) or rare epialleles (negative D^m) dependent on epimutation rate, whereas balancing selection in DNA methylation could generate an excess of intermediate-frequency epialleles (positive D^m). Namely, extreme positive D^m could suggest directional or balancing selection on epialleles, while extreme negative D^m may imply directional selection (Wang & Fan 2015).

Purebred dogs and YNP wolves ($n=47$ and 33, respectively) were analyzed separately for each gene body tested. In order to examine whether gene bodies in purebred dogs and YNP wolves were under different selection pressures, we calculated the difference between purebred dog and YNP wolves' D^m values (YNP wolves D^m - purebred dog $D^m = D^m \text{Diff}$) for each gene body with polymorphism in both groups ($n_{\text{both}}=38$). We then performed permutation tests to determine p -values for $D^m \text{Diff}$. Under the null hypothesis, there should be no difference between dogs and wolves, i.e. $D^m \text{Diff} = 0$. Therefore, we performed 10,000 permutations, randomly assigning dog and wolf labels without replacement while keeping

the vectors of gene bodies intact to preserve linkage disequilibrium. We then performed a gene ontology (GO) analysis as described below.

Finally, we examined the polymorphic gene bodies in purebred dogs to identify deviations from neutrality through a bootstrap analysis with 10,000 replicates, sampling with replacement while keeping the individual gene body vectors intact. This procedure allowed us to estimate the 99.5% bootstrap confidence interval of the D^m range for each gene body in purebred dogs ($n_{\text{gene_bodies}}=63$). We again conducted a GO enrichment analyses on genes that significantly deviated from neutrality as described below.

Detection of differential methylation between wolves and purebred dogs

To measure the genomic signals of purebred dog domestication, we analyzed methylation data from 33 YNP gray wolves and 47 AKC purebred dogs. We report MF in addition to Δ , the absolute difference between the MF of wolves and purebred dogs. All cytosines located within the mitochondrial genome were excluded, as it does not exhibit a strong signal of DNA methylation in vertebrates due to the underrepresentation of CpG motifs in mitochondrial genes (Cardon *et al.* 1994).

DNA methylation patterns can vary by biological processes, but also artificially by batch and cell heterogeneity (Christensen *et al.* 2009; Ji *et al.* 2010; Bock 2012). For example, blood samples consist of heterogeneous population of cell type in varying proportions, each potentially with a distinct DNA methylation signature that may confound association scans (Reinius *et al.* 2012). To address this issue, we used a reference-free correction approach in the program *EWASher*, implemented as part of the *FaST-LMM* Python v 0.2 package (Lippert *et al.* 2011; Zou *et al.* 2014). *EWASher* computes methylome similarity for every pair of samples. These similarities were then used as the covariance in a mixed model as an implicit proxy for cell-type composition. A reference-based cell-type corrected analysis was performed using the estimated cell-type composition as covariates in a linear regression, in addition to other covariates added to the analysis. The analysis begins by converting raw methylation data into β values. We considered a site to be constitutively methylated if the average probe β value across all samples (cases and controls) was $\text{MF} > 0.8$ and to be constitutively unmethylated when $\text{MF} < 0.2$. Linear regression was used to correct for all covariates, including batch, computing residuals used in the main association analyses. Using the similarity matrix as the covariance component in the LMM, an associated p -value was generated for each site. If the genomic control factor λ was still inflated (i.e. genome-wide type 1 error rate; (Devlin & Roeder 1999)), *EWASher* computed the PCs across all samples, included the top PC as a covariate, and then repeated the process. Parameters in the linear mixed model were estimated by maximum likelihood (ML). Association P values were obtained by using a 1 degree-of-freedom likelihood-ratio test and all Bonferroni-significant tests were performed at the $\alpha = 0.05$ significance level.

We observed only small deviations from the 95% error bars in the quantiles of the theoretical null distribution when plotted against the observed, concordant with our expectation that only a small proportion of methylation sites would be significantly different. To control for batch effects, we added batch as a covariate in our regression analysis. In addition, we included sex of the individual and breed as covariates in our analysis. We applied a

Bonferroni-adjusted significance threshold of $p < 0.005$. To survey a larger list of sites, we conducted a GO enrichment analysis on sites that had an adjusted $p < 0.05$ and were found in coding regions ($n=8$). Last, we compared the list of DMSs between the uncorrected SMP list of 1,307 sites ($p < 0.05$) and the corrected methylation (adjusted $p < 0.05$) association scan of 68 sites to determine if there was overlap between the two types of analyses.

Heritability analysis of methylation in a pedigreed wolf population

We utilized the known genealogical relationships among the 35 Yellowstone gray wolves to assess overall heritability of methylation marks and their mode of inheritance (Vonholdt *et al.* 2008). To limit the analysis to relevant members and thus reduce computation time, the pedigree was reduced to include only wolves with methylation values, along with their relatives by using the statistical genetics software Mendel's pedigree trimming option (Lange & Sinsheimer 2004; Lange *et al.* 2013). Wolves in singleton pedigrees post trimming were removed from the analysis (three wolves total). Thus, the final pedigree consisted of 63 wolves, 30 of which had methylation data.

To estimate heritability of methylation, we used a variance components analysis, which takes into account known genealogical relationships between wolves by partitioning the phenotypic variance (Lange 2002). The variance components considered in our analyses included additive genetic, dominance, and environmental variance. As discussed in detail in Lange (2002), the variance component model posits that a quantitative trait such as methylation levels is controlled by multiple loci (polygenic model). In this model, the additive genetic variance captures the effect of alleles at multiple loci as if they were acting independently on methylation. Dominance genetic variation results from deviation from this allelic independence. Environmental variation is any residual variation in the trait values after accounting for genetic and shared environmental variation.

Using the autosomal sites that remain after filtering, methylation levels were standardized and using Mendel, models with different sets of variance components were fit to the data (Lange *et al.* 2005; Lange *et al.* 2013). All models included batch and sex as covariates. In addition, the first 5 PCs from *EWASher* were included in our analysis to correct for cell-type composition. In order to parsimoniously detect evidence for the heritability of methylation marks, we built models in a stepwise fashion. In analysis 1 we fit the simplest model, which includes additive genetic and environmental variance components only. In analysis 2 we reanalyzed those methylation marks that had non-zero estimates of additive genetic variance from analysis 1 by including dominance genetic variance along with the additive genetic and environmental variances. Significance of heritability estimates for all analyses was determined using a 1-sided z-test. Results were termed genome-wide significant if the p -value was less than the Bonferroni-adjusted significance threshold of $p < 2.08 \times 10^{-6}$.

Gene ontology enrichment analysis

To test if any gene list was enriched for specific functional categories or ontological types, we used WebGestalt (Zhang *et al.* 2005; Wang *et al.* 2013b) on a list of genes that were found to be significant outliers. As a background for our GO analyses, we used all of our

unique genes from our RRBS data ($n=4580$). We applied a Benjamini-Hochberg FDR correction (Benjamini & Hochberg 1995).

Results

Reduced representation bisulfite sequencing

Four samples were dropped due to low coverage, or extremely high variation in coverage (Supplemental Table S1). Of the remaining 92 samples ($n_{\text{wolf}}=35$, $n_{\text{purebred}}=47$, $n_{\text{Boz}}=2$, $n_{\text{semi}}=6$, $n_{\text{village}}=2$), cytosine-specific coverage was calculated by pooling on average four samples per lane (average $27.5 \times \pm 4.8$ SD) having 17,024,508 $\pm 2,329,082$ cytosines, prior to any filtering (Supplemental Table S2). After conservative filtering, 24,205 sites remained ($n_{\text{Auto}}=23,997$; $n_{\text{Xchr}}=208$); 33% of the cytosines were located in intergenic regions, 40% were in coding regions and 26% were in promoters (Supplemental Table S3). Further, 19% ($n=4,688$) of sites were found within repeat elements, with the majority in LINEs (21%, $n=978$) and SINEs (43%, $n=2,002$) (Supplemental Table S3). Compared to the 34% TE distribution in the canine genome (Lindblad-Toh *et al.* 2005), the proportion of methylated loci within all repeat elements for our 24,205 sites was approximately 19% ($\chi^2=0.058$, $df=1$, $p=0.81$).

Outlier samples and methylation trends

The average methylation (\pm Standard Deviation; SD) was moderate across 92 samples ($MF_{\text{wolf}}=0.38 \pm 0.4$; $MF_{\text{purebred}}=0.34 \pm 0.4$; $MF_{\text{Boz}}=0.42 \pm 0.4$; $MF_{\text{semi}}=0.35 \pm 0.4$; $MF_{\text{village}}=0.40 \pm 0.4$). However, methylation was significantly higher on the X chromosome than on autosomes ($MF_{\text{autosomes}}=0.34 \pm 0.4$; $MF_{\text{Xchr}}=0.75 \pm 0.4$; 1-tailed *t*-test unequal variance $p < 0.0001$). To identify outlier samples, we conducted a PCA on 23,997 autosomal and 208 X-linked cytosines that were converted to SMP genotypes. Using a genotype correlation inclusion threshold of $r^2 < 0.2$, we retained 1,291 unlinked autosomal and 16 unlinked X chromosome SMPs, of which 8% were found in repeat regions ($n=106$, where 48 were found within LINEs and SINEs), 32% were within coding regions ($n=413$), 53% intergenic ($n=692$), and 15% regulatory ($n=202$). Across the genome, we identified seven samples that were subsequently excluded from all analyses due to their extremely polarizing effect on the PCA (Supplemental Fig. S1). After removal of outlier samples, the genome-wide average methylation (SD) remained moderate ($MF_{\text{wolf}}=0.34 \pm 0.4$; $MF_{\text{purebred}}=0.34 \pm 0.4$; $MF_{\text{Boz}}=0.39 \pm 0.4$; $MF_{\text{semi}}=0.36 \pm 0.4$; $MF_{\text{village}}=0.38 \pm 0.4$; Supplemental Table S4).

In addition, after removal of outlier samples, a hierarchical cluster analysis of autosomal SMPs ($n_{\text{SMPs}}=23,997$) identified three distinct groups: wolves, dogs, and a “mixed” cluster containing two subgroups of wolves, ancient dogs breeds, and aboriginal island dogs (dingoes removed as outliers) (Fig. 1A). Across the autosomes, the wolves formed a distinct group to that of dogs, with $>85\%$ accuracy of species assignment (29/33 wolves; 39/46 purebred dogs) (Fig. 1A). Although there was subdivision of sexes within each cluster, this was less absolute (11/15 female wolves; 10/18 male wolves; 17/25 female purebred dogs; 8/21 male purebred dogs). However, despite a smaller set of X-linked SMPs ($n_{\text{SMPs}}=208$) with variable branch lengths, overall cluster analysis produced five distinct groups: male dogs (15/21), female dogs (14/25), male wolves (18/18), female wolves (12/15), and a

“mixed” cluster of aboriginal island/village dogs and wolves (Fig. 1B). The primary axis of divergence for the X-linked SMPs was sex (26/40 females; 33/39 males), followed by species assignments (31/33 wolves; 32/46 purebred dogs). In both dendrograms, there were outlier assignments with wolves processed in a different batch (n=4), aboriginal island dogs (n=4), boz breed dog (n=1), village dogs (n=2), and some putatively ancient dog breeds (bsji, chcr, husk, peke, samo; n=6) (Fig. 1). PC1 also confirmed that the predominant divergence among the samples was a species-specific pattern explaining 7% of the variation (Fig. 2). Relative to the wolves, all dog samples displayed a wider spatial variation along axis 2 (3.8% variation explained).

We then evaluated population subdivision using the program *ADMIXTURE* for K=2-8 genetic groups for 1,307 unlinked SMPs (Alexander *et al.* 2009)(Supplemental Fig. S2; Supplemental Table S5). The best fit at K=2 clearly defined two epigenetic groups, gray wolves and all dogs (purebred, aboriginal island, and village). As K increased, subdivision was found within both groups, though it was not consistent with breed grouping. However, three main groups were distinct at higher K values: wolves, purebred dogs, and the cluster containing the aboriginal island and village dogs (Supplemental Fig. S2).

An association scan of 1,307 unlinked SMPs identified 36 significant sites associated with domestication ($p < 10^{-5}$, $-\log_{10}(p) > 5$) (Fig. 3; Supplemental Table S6A), of which 47% (17/36) were sites located within a LINE or SINE element, 67% (24/36) intergenic, 25% (9/36) coding, and 6% (2/36) promoter. Both LINES and SINES were significantly enriched compared to the 1,307 unlinked SMPs dataset (Supplemental Table S3). Eleven of the 36 significant sites were located within genes ($n_{\text{exons}}=1$; $n_{\text{intron}}=8$; $n_{\text{prom}}=2$), and the majority of the associated sites (27/36) were hyper-methylated in purebred dogs (average $\text{MF}_{\text{wolf}}=0.62$; $\text{MF}_{\text{purebred}}=0.75$). We found an identical pattern of hyper-methylation of transposons in purebred dogs (15/17 hyper-methylated in dogs; average $\text{MF}_{\text{wolf}}=0.61$; $\text{MF}_{\text{purebred}}=0.81$). Using the OMIM database (Hamosh *et al.* 2005), we annotated the significant DMSs associated with genes (Supplemental Table S6A).

Differential methylation between purebred dogs and wolves

EWASher retained 962 cytosines with $0.2 < \text{MF} < 0.8$ between 45 purebred dogs and 33 YNP wolves, of which 68 were significant (corrected $p < 0.05$; Supplemental Table S3) and 14 were highly significant ($p < 0.005$; Fig. 4, Supplemental Table S6B & S7; Supplemental Fig. S3). Out of the 68 significant cytosines, the majority was located within intergenic regions (n=53). The repeat elements were located predominantly in unknown locations (n=33). The majority of the associated sites (36/68) were significantly hyper-methylated in purebred dogs (adjusted $p < 0.05$; average $\text{MF}_{\text{wolf}}=0.58$; $\text{MF}_{\text{purebred}}=0.59$). The average methylation frequency difference between purebred dogs and gray wolves was 0.009 ($\text{SD} \pm 0.018$). Among these significantly associated DMSs, 51% (30/68) consisted of repetitive DNA of which 30 were transposons. We conducted a GO analysis of eight out of 68 sites that were non-overlapping and had an associated annotation. After Benjamini-Hochberg FDR correction, none of our GO categories under Biological Process, Molecular Function, and Cellular Component were significantly enriched (Supplemental Table S8). Further, we found 19 of 68 sites overlapped with the uncorrected SMP association scan, with the majority of

sites annotated as intergenic and located within SINES (Table 1). In addition, we found one out of our 14 highly significant, top-ranking sites that overlapped with the uncorrected SMP association scan (Table 1). Out of our 19 overlapping sites, three sites deviated from Hardy-Weinberg Equilibrium (two sites exhibited an excess of heterozygotes and a single site exhibited an excess of homozygotes; Table 1).

To obtain a perspective at the level of the gene and determine a possible correlation with methylation status, we asked if any repetitive element was within 3Kb of each DMS. We hypothesized that if proximal repetitive elements are epigenetically silenced, their methylation status may spread to nearby sites (Vonholdt *et al.* 2012). We found that 66% (n=45) of the 68 DMSs were located within or were proximal to repetitive elements. Of the 15 genic DMSs, the majority (67%, n=10) was associated with repeat elements.

Neutrality estimation

Estimating methylation-specific neutrality statistics using a D^m test revealed that out of 3,229 annotated gene bodies (i.e. the gene bodies in the genome covered by our methylome mapping approach), 121 ($n_{dog}=63$; $n_{wolf}=96$; $n_{shared}=38$) exhibited SMP variation (Supplemental Table S9A). The mean observed gene body D^m values we calculated for purebred dogs was -0.5598 and -0.144571 for YNP wolves, eight of which were significantly different ($p<0.05$) between dogs and wolves ($D^m Diff$ different from 0) as assessed by permutation (Table 2). Of these eight genes, SLC17A8, a glutamate transport gene, overlapped the list of significant outliers from both the uncorrected SMP association and corrected methylation scan (Table 1; Table 2). GO analyses revealed that none of our GO categories under Biological Process, Molecular Function, and Cellular Component were significantly enriched after Benjamini-Hochberg FDR correction (Supplemental Table S8). Further, 75% (6/8) of dog genes that deviated from neutrality contained repeat elements within their boundaries (Table 2).

We also assessed if gene body methylation in purebred dogs deviated from neutrality through a bootstrap analyses performed on polymorphic SMPs. We found that 11 of 63 gene bodies had 99.5% Bootstrap Confidence Intervals (BCIs) that did not overlap the mean D^m value (Supplemental Table S9B), consistent with substantial deviation from neutrality. From the list of 63 gene bodies, we again found neurotransmitter-related genes (GRBRB1, SLC17A8, and ADCY1). GO analyses did not yield significant results (Supplemental Table S8).

Heritability of methylation

In YNP wolves, pedigree-assisted estimates of heritability were calculated using 23,997 autosomal methylation sites. Analysis 1, which includes fixed effect covariates for batch, sex, and cell-type composition and incorporates variance components for additive genetic and environmental variance, resulted in 4,504 sites with non-zero variance component estimates, 2,282 of which exhibited significant narrow sense heritability after accounting for multiple testing ($p<2.08\times 10^{-6}$) (Table 3). Most of the sites exhibited low narrow sense heritability, 81.25% of the methylation marks exhibited virtually zero narrow sense heritability ($h^2<0.01$), whereas a sizable minority (8.89%) had extremely high narrow sense

heritability ($h^2 > 0.99$) with the remaining 2,367 sites showing moderate levels of heritability (median $h^2 = 0.536$) (Supplemental Fig. S4A; Supplemental Table S6). In addition, the methylation marks with the least variability (<5th percentile) had standard deviations ranging between 0.0004 and 0.0129. After adjusting for multiple testing ($p < 2.09 \times 10^{-6}$), 7% of these 1,200 least variable methylation marks were significantly heritable. When the 1,200 least variable sites were removed, 9.64% of the 22,797 remaining autosomal methylation marks were significantly heritable, demonstrating that the percentage of sites that were significantly heritable differed when considering non-variable sites compared to the rest of the autosomal sites ($p = 0.002$).

Because the wolf pedigree was large and highly detailed, we were able to further explore the broad sense heritability by looking for evidence of dominance genetic variance. In Analysis 2, we included only sites with non-zero additive genetic variance estimates from Analysis 1 ($n = 4,505$) to avoid building on a model in which the additive genetic variance was already estimated to be zero, and extended the model to include dominance variance. As a result, Analysis 2 identified 1,362 sites with significant broad sense heritability using the same fixed effect covariates and genome-wide significance threshold ($p = 2.08 \times 10^{-6}$) (Table 3). In addition, 309 of these methylation marks with significant broad sense heritability also had significant narrow sense heritability. As expected from the narrow sense heritability, the vast majority of methylation marks had near-zero broad sense heritability (81.26% with $H^2 < 0.01$). This percentage represented only two additional methylation marks than were observed to have near-zero narrow sense heritability. A sizable minority of methylation marks were estimated to have either extremely high narrow or broad sense heritability (14.71% with either $h^2 > 0.99$ or $H^2 > 0.99$), an additional 1,397 methylation marks over the number estimated when only allowing for additive genetic variance (Supplemental Fig. S4B). Some of these estimates of high heritability were likely due to the small sample size, making it difficult to accurately estimate the heritability; however, some of these estimates were likely due to methylation marks that were inherited in an essentially Mendelian fashion. The remaining 968 sites had moderate levels of broad sense heritability (median $H^2 = 0.50$). Manhattan plots confirmed that the methylation marks with significant narrow or broad sense heritability were distributed throughout the genome (Supplemental Fig. S5).

Discussion

From a genome-wide scan of methylation at CpG islands across 33 Yellowstone wolves and 45 American Kennel Club registered purebred dogs, we found that the regions exhibiting differential methylation are enriched for transposons and also contain several intriguing categories of functionally relevant genes (e.g. neurotransmitters). In order to minimize differences in methylation values resulting from cell composition, we used a correction-based approach that estimates the degree of genomic inflation (i.e. genome-wide type 1 error rate) with a subsequent correction for cell-type composition using a linear mixed model conjointly with PCA. In addition, we corrected our analyses for multiple covariates including batch, sex, and breed, when applicable. To measure methylation differences between purebred dogs and YNP wolves, we uncovered the signature of selection for polymorphic genes and determined trends in the overall heritability of methylation,

providing a revealing picture on the epigenetic changes that occurred during the domestication of the dog from the wolf.

Role of transposons and phenotype evolution

An association scan revealed that differentially methylated sites were generally hypermethylated in dogs and frequently found within repetitive DNA, primarily transposons. In canines, transposons make up 34% of the canine genome (Lindblad-Toh *et al.* 2005), but nearly half of other mammalian genomes (~45%) (Smit 1999; Whitelaw & Martin 2001). Transposons play a variety of roles for genome functioning and stability, which include chromosome breakage and rearrangement, in addition to epigenetic transcriptional regulators (Deininger *et al.* 2003; Slotkin & Martienssen 2007). Due to their large copy number variation and insertion site polymorphism, transposons can serve as the raw material for evolution and influence local transcriptional patterns as they are targeted by epigenetic mechanisms for silencing (McClintock 1984; Kidwell 2002; Deininger *et al.* 2003). Insertion sites may also disrupt gene function and alter linked phenotypes, serving as a mechanism that can rapidly generate phenotypic variation, an aspect often noted during species domestication. For example, miniature inverted-repeat transposons (MITEs) experienced a rapid amplification and sequence diversification over a relatively short period of time during rice domestication (Feschotte *et al.* 2002; Jiang *et al.* 2003; Naito *et al.* 2006). After transposition, newly inserted elements are often targeted for epigenetic silencing, which frequently suppresses transcription of linked genes (Kaasik & Lee 2004; Zhang *et al.* 2006; Lister *et al.* 2008; Hollister *et al.* 2011). In maize, a *Hopscotch* retro-element approximately 60Kb upstream of the *tb1* coding region is associated with increased *tb1* expression, resulting in apical dominance (i.e. changes in branch structure) in maize compared to teosinte. In addition, molecular dating of the *Hopscotch* retro-element provides suggestive evidence that it predates the domestication of maize and existed in standing genetic variation of teosinte prior to artificial selection (Studer *et al.* 2011).

Transposons, specifically in SINES, have been previously associated with genome evolution of the domestic dog (Wang & Kirkness 2005; Kirkness 2006). SINES were originally characterized as having a genome-wide distribution (Minnick *et al.* 1992; Wang & Kirkness 2005) with insertion mutation mapped to phenotypes: narcolepsy in Doberman Pinschers (Lin *et al.* 1999), autosomal recessive centronuclear myopathy in Labrador retrievers (Pele *et al.* 2005), and merle patterning in multiple breeds (Clark *et al.* 2006). There are ~170,000 SINEC_Cf elements, a subfamily of canine-specific SINES, with approximately half of all genes containing a SINEC_Cf insertion (Wang & Kirkness 2005). When these elements are transcribed, they can provide splice receptor sites that can result in the addition of novel exons to the genome.

A recent study revealed the evolutionary history of an intronic SINE insertion located within *IGF1* of purebred dogs was lacking in a variety of canid species, including the gray wolf (Gray *et al.* 2010). Taken together, these results provide a unique perspective of canine domestication where ancient transposons present at a low copy number in the ancestral genome could have experienced rapid amplification as a response to strong artificial selective pressures during domestication, and have since contributed towards functional

variation. The “stress hypothesis” can provide an intuitive context for transposon expansion and genome restructuring (Capy *et al.* 2000; Chénais *et al.* 2012). As it states, under certain conditions of stress that impact fitness, transposons may become activated as a potential genomic mechanism for novel phenotype production. Dogs, unlike their wild counterparts, recently experienced significant genomic stress due to strong artificial selection and line breeding (e.g. inbreeding; Capy *et al.* 2000). Physiological stress, for example, is easily imagined as changes in temperature, nutrition, pesticides, or infection. Genomic or molecular stress often include selection or inbreeding which elicits responses, either at the genotypic or phenotypic level (Capy *et al.* 2000; Bijlsma & Loeschcke 2005). This transposon expansion process, however, cannot continue uncontrolled. Methylation may have served as a flexible epigenetic control of transposition, with targeted methylation acting to suppress rapidly expanding transposon families and marking them for subsequent purging (Smit 1999; Matzke *et al.* 2007; Slotkin & Martienssen 2007). This speculative mechanism may have contributed to the rapid production of phenotypic variation evident in the history of the domestic dog.

Neutrality Estimation

Similar to single nucleotide changes at the genome level, spontaneous epigenetic mutations can occur at the level of the methylome (Becker *et al.* 2011; Schmitz *et al.* 2011; Schmitz *et al.* 2013) and contribute to fitness differences on which natural selection can act (Rapp & Wendel 2005; Geoghegan & Spencer 2012; Hirsch *et al.* 2012; Geoghegan & Spencer 2013). Specifically, these epialleles can accumulate in an evolutionary lineage in which the linked phenotype shows non-zero heritability and is thus expected to contribute toward individual fitness and the action of natural selection (Wang & Fan 2015).

We identified eight genes with neutrality estimates that are under different selection pressure or demography effects between purebred dogs and YNP wolves, indicating that both selection and demography have played a role in shaping species-specific methylation patterns. Of these eight genes, SLC17A8, a glutamate transport gene, was an outlier in multiple analyses, exhibiting high heritability and containing repetitive elements. In a surveillance of non-neutral methylation patterns within the purebred dog genome, we again found neurotransmitter-related genes (GRBRB1, SLC17A8, and ADCY1). ADCY1 is a neurotransmitter that is believed to play a role in memory acquisition and learning (Wang *et al.* 2004). Interestingly, ADCY1 is found within the same gene family as ADCY8, a gene previously identified to be under positive selection in the dog genome with a similar role in functional memory (Vonholdt *et al.* 2010).

Heritability of methylation marks

Epigenetic mechanisms, such as DNA methylation, can directly affect gene expression (Jaenisch & Bird 2003; Szyf *et al.* 2005; Ong & Corces 2011; Jones 2012). To date, few studies have investigated transgenerational epigenetic variation in natural animal populations (Peaston & Whitelaw 2006; Ho & Burggren 2010; Feil & Fraga 2011). Fitness-related phenotypes have been documented across generations in Yellowstone wolves, including reproductive success, fecundity, lifespan, inbreeding coefficients, behavior, and coat color (Coulson *et al.* 2011; Stahler *et al.* 2013; Hedrick *et al.* 2014). To account for inbreeding

effects that can result from a founding bottleneck, we used the Yellowstone wolf genealogy to estimate the heritability of each methylation mark (Bell *et al.* 2012). We did not have relatedness information for dogs; therefore, dogs were not included in heritability analyses.

More specifically, we examined the heritability of methylation marks using a variance component model, a classic approach used for many years with quantitative phenotype traits (Lange 2002). As with humans, a minority of the methylation marks in the wolves showed highly significant evidence of heritability (Bell *et al.* 2012; Grundberg *et al.* 2013; McRae *et al.* 2014; Shah *et al.* 2014). When using a simple model where the level of methylation was determined by multiple genes each acting independently, a fraction (9.51%) of 23,997 autosomal methylation marks was significantly heritable after adjusting for multiple testing ($p < 2.08 \times 10^{-6}$).

With the advantage of having an extensive and accurate pedigree for the Yellowstone wolves, we could further examine whether any of these methylation marks showed evidence of dominance genetic variance (broad sense heritability). We found 1,362 methylation marks (5.68%) that showed evidence of significant broad sense heritability. As our sample size for the heritability analysis was too small to reliably dissect the contribution of random household effects from genetic effects, the effects of common environment may inflate our estimates of the percentage of significantly heritable sites. However, they are similar to that observed in humans and, as in humans, these methylation marks seem to be uniformly distributed throughout the chromosomes.

While we detected transgenerational stability of differentially methylated sites, a necessary requirement for evolution, the scope of our study cannot address the functionality of these sites. However, we demonstrate the co-occurrence of such sites with repetitive elements. This relationship may indicate genotype-dependent epigenetic activity and, thus the patterns we describe may actually represent the underlying genotype that is both inherited and selected. The prevalence of repetitive elements found among the outlier sites support the notion that transposons, in particular, experienced a recent expansion during dog domestication with rapid and subsequent silencing through an epigenetic mechanism. Therefore, while we observe the stability of favorable epialleles, we cannot determine the cause of transmission or their association.

Role of neurotransmitters and domestication

In our study, we found both GABA (GABRB1) and glutamate (SLC17A8) to be significantly associated with the differences between dogs and wolves, both of which are essential inhibitory and excitatory neurotransmitters in the brain, respectively. More specifically, GABA receptors are a family of proteins involved in the GABAergic neurotransmission of the mammalian central nervous system (Johnston 1996; Bormann 2000). GABA-A receptors are the site of action of a number of important pharmacologic agents including barbiturates, benzodiazepines, and ethanol (Whiting *et al.* 1999). Glutamate transporters are located on the surface of the cells expressing them in the central nervous system, as well as peripheral organs and tissues whereby they control glutamate signaling, and affect a wide-array of functions including cognition, memory, learning, and cell migration, differentiation, and death (Danbolt 2001). Due to the fact that GABA is inhibitory

and glutamate is excitatory, both neurotransmitters conjointly control many processes, including brain excitation (Petroff 2002). A recent study found two glutamate receptors (*GRIK3* and *GRIK2*) that were differentially expressed in the frontal cortex of dogs and wolves, suggesting a role of increased excitatory plasticity during domestication (Li *et al.* 2014).

Alterations in neural signaling and pathways due to changes in density of receptors, biosynthetic pathway alteration, or gene expression differences are indicative of neural plasticity, or flexible changes in the neural system, as a result of recent adaptation, such as changes that may occur during the process of domestication (Popova *et al.* 1997; Cheng 2010; Kukekova *et al.* 2011). Behavioral selection is thought to be one of the primary factors that contributed to canid domestication, in which selection acted directly on the genes involved in behavioral development (Trut *et al.* 2004). Not surprisingly, changes in brain gene expression have been shown in studies of domestication. Specifically, wild canids show a distinct signature of gene expression of a number of genes potentially targeted during behavioral selection in the hypothalamus, prefrontal cortex, and other brain regions compared to domestic dogs (Saetre *et al.* 2004; Li *et al.* 2013).

Similarly, genetic polymorphisms in neurotransmitters have been observed both between wolves and dogs and within dog breeds (Hejjas *et al.* 2007). A study using whole genome sequencing in dogs and wolves uncovered 19 genes related to brain function, with approximately half of the genes associated specifically with nervous system development (Axelsson *et al.* 2013). Similarly, another study using whole genome sequencing generated a candidate list of genes thought to be under positive selection during dog domestication and found multiple genes categorized as being involved in neurological processes (Wang *et al.* 2013a).

Our results on neurotransmitter-related genes must be interpreted with caution, however, as we did not collect methylation values directly from brain tissue. Although other studies have shown that methylation in blood is a reliable proxy for methylation in brain tissue (Davies *et al.* 2012), future studies are needed to verify whether our aforementioned neurotransmitter-related genes are differentially methylated in brain-specific tissue, such as tissue originating from the cortex.

In conclusion, our analysis of DNA methylation provides novel insights into the processes (e.g. demography, selection) that shape the epigenetic patterns in the genome of wild and domesticated canids. We found distinct methylation patterns between dogs and wolves, with a prevalence of differentially methylated sites located in transposons. A substantial fraction of these methylated sites demonstrated high heritability. We develop a speculative hypothesis regarding canine methylome evolution that may have enabled phenotypic diversification. Future studies will be needed to measure the effect of specific transposons on gene functionality, particularly genes involved in neurotransmission.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgements

The authors would like to thank the purebred dog owners for providing blood samples, and Brian Peckinpaugh for providing Boz samples. In addition, we would like to thank Jennifer Listgarten for providing support for the program *FaST-LMM-EWASher*. We would also like to thank the Princeton University Computational Science & Engineering Support (CSES) group for providing computational assistance for multiple components of our work. This study was supported in part by the National Science Foundation grants DEB-0613730 and DEB-1245373, Yellowstone National Park, and many donors through the Yellowstone Park Foundation. This work was also partially funded through the Intramural Program of the National Human Genome Research Institute and these additional grants: AKC OAK (CHF #1822), NIH T32 HG002536, National Science Foundation DMS-1264153, and NIH GM053275.

References

- Abraham G, Inouye M. Fast principal component analysis of large-scale genome-wide data. *PLoS One*. 2014; 9:e93766. [PubMed: 24718290]
- Agirre X, Castellano G, Pascual M, et al. Whole-epigenome analysis in multiple myeloma reveals DNA hypermethylation of B cell-specific enhancers. *Genome Res*. 2015
- Albert FW, Somel M, Carneiro M, et al. A comparison of brain gene expression levels in domesticated and wild animals. *PLoS Genet*. 2012; 8:e1002962. [PubMed: 23028369]
- Alexander DH, Novembre J, Lange K. Fast model-based estimation of ancestry in unrelated individuals. *Genome Res*. 2009; 19:1655–1664. [PubMed: 19648217]
- Almberg ES, Mech LD, Smith DW, Sheldon JW, Crabtree RL. A serological survey of infectious disease in Yellowstone National Park's canid community. *PLoS One*. 2009; 4:e7042. [PubMed: 19756151]
- Anderson TM, vonHoldt BM, Candille SI, et al. Molecular and Evolutionary History of Melanism in North American Gray Wolves. *Science*. 2009; 323:1339–1343. [PubMed: 19197024]
- Andersson L. Molecular consequences of animal breeding. *Curr Opin Genet Dev*. 2013; 23:295–301. [PubMed: 23601626]
- Andersson L, Georges M. Domestic-animal genomics: deciphering the genetics of complex traits. *Nat Rev Genet*. 2004; 5:202–212. [PubMed: 14970822]
- Axelsson E, Ratnakumar A, Arendt ML, et al. The genomic signature of dog domestication reveals adaptation to a starch-rich diet. *Nature*. 2013; 495:360–364. [PubMed: 23354050]
- Becker C, Hagmann J, Muller J, et al. Spontaneous epigenetic variation in the Arabidopsis thaliana methylome. *Nature*. 2011; 480:245–249. [PubMed: 22057020]
- Bell JT, Tsai P-C, Yang T-P, et al. Epigenome-wide scans identify differentially methylated regions for age and age-related phenotypes in a healthy ageing population. *PLoS Genet*. 2012; 8:e1002629. [PubMed: 22532803]
- Benjamini Y, Hochberg Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society. Series B (Methodological)*. 1995:289–300.
- Bernstein BE, Meissner A, Lander ES. The mammalian epigenome. *Cell*. 2007; 128:669–681. [PubMed: 17320505]
- Bijlsma R, Loeschcke V. Environmental stress, adaptation and evolution: an overview. *Journal of evolutionary biology*. 2005; 18:744–749. [PubMed: 16033544]
- Bird A. Perceptions of epigenetics. *Nature*. 2007; 447:396–398. [PubMed: 17522671]
- Bock C. Analysing and interpreting DNA methylation data. *Nat Rev Genet*. 2012; 13:705–719. [PubMed: 22986265]
- Bormann J. The 'ABC' of GABA receptors. *Trends in Pharmacological Sciences*. 2000; 21:16–19. [PubMed: 10637650]
- Boyko AR. The domestic dog: man's best friend in the genomic era. *Genome Biol*. 2011; 12:216. [PubMed: 21338479]
- Boyko AR, Quignon P, Li L, et al. A simple genetic architecture underlies morphological variation in dogs. *PLoS Biol*. 2010; 8:e1000451. [PubMed: 20711490]

- Capy P, Gasperi G, Biémont C, Bazin C. Stress and transposable elements: co-evolution or useful parasites? *Heredity*. 2000; 85:101–106. [PubMed: 11012710]
- Cardon LR, Burge C, Clayton DA, Karlin S. Pervasive CpG suppression in animal mitochondrial genomes. *Proceedings of the National Academy of Sciences*. 1994; 91:3799–3803.
- Chen PY, Cokus SJ, Pellegrini M. BS Seeker: precise mapping for bisulfite sequencing. *BMC Bioinformatics*. 2010; 11:203. [PubMed: 20416082]
- Chénais B, Caruso A, Hiard S, Casse N. The impact of transposable elements on eukaryotic genomes: from genome size increase to genetic adaptation to stressful environments. *Gene*. 2012; 509:7–15. [PubMed: 22921893]
- Cheng HW. Breeding of tomorrow's chickens to improve well-being. *Poult Sci*. 2010; 89:805–813. [PubMed: 20308415]
- Christensen BC, Houseman EA, Marsit CJ, et al. Aging and environmental exposures alter tissue-specific DNA methylation dependent upon CpG island context. *PLoS Genet*. 2009; 5:e1000602. [PubMed: 19680444]
- Clark LA, Wahl JM, Rees CA, Murphy KE. Retrotransposon insertion in *SILV* is responsible for merle patterning of the domestic dog. *Proceedings of the National Academy of Sciences of the United States of America*. 2006; 103:1376–1381. [PubMed: 16407134]
- Clutton-Brock, J. A natural history of domesticated mammals. Cambridge University Press; 1999.
- Clutton-Brock J. The process of domestication. *Mammal Review*. 1992; 22:79–85.
- Coulson T, MacNulty DR, Stahler DR, et al. Modeling effects of environmental change on wolf population dynamics, trait evolution, and life history. *Science*. 2011; 334:1275–1278. [PubMed: 22144626]
- Cubaynes S, MacNulty DR, Stahler DR, et al. Density-dependent intraspecific aggression regulates survival in northern Yellowstone wolves (*Canis lupus*). *J Anim Ecol*. 2014
- Danbolt NC. Glutamate uptake. *Progress in neurobiology*. 2001; 65:1–105. [PubMed: 11369436]
- Darwin C. The variation of animals and plants under domestication. 1868
- Davies MN, Volta M, Pidsley R, et al. Functional annotation of the human brain methylome identifies tissue-specific epigenetic variation across brain and blood. *Genome Biol*. 2012; 13:R43. [PubMed: 22703893]
- Deaton AM, Bird A. CpG islands and the regulation of transcription. *Genes Dev*. 2011; 25:1010–1022. [PubMed: 21576262]
- Deininger PL, Moran JV, Batzer MA, Kazazian HH. Mobile elements and mammalian genome evolution. *Curr Opin Genet Dev*. 2003; 13:651–658. [PubMed: 14638329]
- Devlin B, Roeder K. Genomic Control for Association Studies. *Biometrics*. 1999; 55:997–1004. [PubMed: 11315092]
- Diamond J. Evolution, consequences and future of plant and animal domestication. *Nature*. 2002; 418:700–707. [PubMed: 12167878]
- Dobney K, Larson G. Genetics and animal domestication: new windows on an elusive process. *Journal of Zoology*. 2006; 269(2):261–71.
- Doebley JF, Gaut BS, Smith BD. The molecular genetics of crop domestication. *Cell*. 2006; 127:1309–1321. [PubMed: 17190597]
- Feeney A, Nilsson E, Skinner MK. Epigenetics and transgenerational inheritance in domesticated farm animals. *Journal of Animal Science and Biotechnology*. 2014; 5:48. [PubMed: 25810901]
- Feil R, Fraga MF. Epigenetics and the environment: emerging patterns and implications. *Nat Rev Genet*. 2011; 13:97–109. [PubMed: 22215131]
- Ferrer I, Martinez A, Boluda S, Parchi P, Barrachina M. Brain banks: benefits, limitations and cautions concerning the use of post-mortem brain tissue for molecular studies. *Cell and Tissue Banking*. 2008; 9:181–194. [PubMed: 18543077]
- Feschotte C, Jiang N, Wessler SR. Plant transposable elements: where genetics meets genomics. *Nat Rev Genet*. 2002; 3:329–341. [PubMed: 11988759]
- Freedman AH, Gronau I, Schweizer RM, et al. Genome sequencing highlights the dynamic early history of dogs. *PLoS Genet*. 2014; 10:e1004016. [PubMed: 24453982]

- Geoghegan JL, Spencer HG. Population-epigenetic models of selection. *Theor Popul Biol.* 2012; 81:232–242. [PubMed: 21855559]
- Geoghegan JL, Spencer HG. Exploring epiallele stability in a population-epigenetic model. *Theor Popul Biol.* 2013; 83:136–144. [PubMed: 23044385]
- Germonpré M, Sablin MV, Stevens RE, et al. Fossil dogs and wolves from Palaeolithic sites in Belgium, the Ukraine and Russia: osteometry, ancient DNA and stable isotopes. *Journal of Archaeological Science.* 2009; 36:473–490.
- Gordon, AD. Classification. Chapman & Hall/CRC Monographs on Statistics & Applied Probability; 1999.
- Gower JC. A general coefficient of similarity and some of its properties. *Biometrics.* 1971:857–871.
- Gray MM, Sutter NB, Ostrander EA, Wayne RK. The IGF1 small dog haplotype is derived from Middle Eastern grey wolves. *BMC Biol.* 2010; 8:16. [PubMed: 20181231]
- Grundberg E, Meduri E, Sandling JK, et al. Global analysis of DNA methylation variation in adipose tissue from twins reveals links to disease-associated variants in distal regulatory elements. *The American Journal of Human Genetics.* 2013; 93:876–890. [PubMed: 24183450]
- Hamosh A, Scott AF, Amberger JS, Bocchini CA, McKusick VA. Online Mendelian Inheritance in Man (OMIM), a knowledgebase of human genes and genetic disorders. *Nucleic Acids Res.* 2005; 33:D514–D517. [PubMed: 15608251]
- Hedrick PW, Stahler DR, Dekker D. Heterozygote advantage in a finite population: black color in wolves. *Journal of Heredity.* 2014; 105:457–465.
- Hejjas K, Vas J, Kubinyi E, et al. Novel repeat polymorphisms of the dopaminergic neurotransmitter genes among dogs and wolves. *Mamm Genome.* 2007; 18:871–879. [PubMed: 18049838]
- Heyn H, Moran S, Hernando-Herraez I, et al. DNA methylation contributes to natural human variation. *Genome Res.* 2013; 23:1363–1372. [PubMed: 23908385]
- Hirsch S, Baumberger R, Grossniklaus U. Epigenetic variation, inheritance, and selection in plant populations. *Cold Spring Harb Symp Quant Biol.* 2012; 77:97–104. [PubMed: 23619013]
- Ho DH, Burggren WW. Epigenetics and transgenerational transfer: a physiological perspective. *J Exp Biol.* 2010; 213:3–16. [PubMed: 20008356]
- Hollister JD, Smith LM, Guo Y-L, et al. Transposable elements and small RNAs contribute to gene expression divergence between *Arabidopsis thaliana* and *Arabidopsis lyrata*. *Proceedings of the National Academy of Sciences.* 2011; 108:2322–2327.
- Jaenisch R, Bird A. Epigenetic regulation of gene expression: how the genome integrates intrinsic and environmental signals. *Nat Genet.* 2003; 33(Suppl):245–254. [PubMed: 12610534]
- Ji H, Ehrlich LI, Seita J, et al. Comprehensive methylome map of lineage commitment from haematopoietic progenitors. *Nature.* 2010; 467:338–342. [PubMed: 20720541]
- Jiang N, Bao Z, Zhang X, et al. An active DNA transposon family in rice. *Nature.* 2003; 421:163–167. [PubMed: 12520302]
- Johnston GAR. GABA C receptors: relatively simple transmitter-gated ion channels? *Trends in Pharmacological Sciences.* 1996; 17:319–323. [PubMed: 8885697]
- Jones PA. Functions of DNA methylation: islands, start sites, gene bodies and beyond. *Nat Rev Genet.* 2012; 13:484–492. [PubMed: 22641018]
- Kaasik K, Lee CC. Reciprocal regulation of haem biosynthesis and the circadian clock in mammals. *Nature.* 2004; 430:467–471. [PubMed: 15269772]
- Kaminsky ZA, Tang T, Wang SC, et al. DNA methylation profiles in monozygotic and dizygotic twins. *Nat Genet.* 2009; 41:240–245. [PubMed: 19151718]
- Kaufman, L.; Rousseeuw, PJ. Finding groups in data: an introduction to cluster analysis. John Wiley & Sons; 2009.
- Kennerly E, Ballmann A, Martin S, et al. A gene expression signature of confinement in peripheral blood of red wolves (*Canis rufus*). *Mol Ecol.* 2008; 17:2782–2791. [PubMed: 18466232]
- Kidwell MG. Transposable elements and the evolution of genome size in eukaryotes. *Genetica.* 2002; 115:49–63. [PubMed: 12188048]
- Kirkness EF. 12 SINEs of Canine Genomic Diversity. *Cold Spring Harbor Monograph Archive.* 2006; 44:209–219.

- Krueger F, Trim Galore. A wrapper tool around Cutadapt and FastQC to consistently apply quality and adapter trimming to FastQ files, with some extra functionality for MspI-digested RRBS-type (Reduced Representation Bisulfite-Seq) libraries. 2013
- Kukekova AV, Johnson JL, Teiling C, et al. Sequence comparison of prefrontal cortical brain transcriptome from a tame and an aggressive silver fox (*Vulpes vulpes*). *BMC Genomics*. 2011; 12:482. [PubMed: 21967120]
- Lange K. Mathematical and statistical methods for genetic analysis. Springer Science & Business Media; 2002.
- Lange K, Papp JC, Sinsheimer JS, et al. Mendel: the Swiss army knife of genetic analysis programs. *Bioinformatics*. 2013; 29:1568–1570. [PubMed: 23610370]
- Lange K, Sinsheimer JS. The pedigree trimming problem. *Human heredity*. 2004; 58:108–111. [PubMed: 15711090]
- Lange K, Sinsheimer JS, Sobel E. Association testing with Mendel. *Genetic epidemiology*. 2005; 29:36–50. [PubMed: 15834862]
- Langmead B. Aligning short sequencing reads with Bowtie. *Current protocols in bioinformatics*. 2010:11.17, 11–11.17, 14. [PubMed: 20521243]
- Langmead B, Salzberg SL. Fast gapped-read alignment with Bowtie 2. *Nature methods*. 2012; 9:357–359. [PubMed: 22388286]
- Larson G, Karlsson EK, Perri A, et al. Rethinking dog domestication by integrating genetics, archeology, and biogeography. *Proceedings of the National Academy of Sciences*. 2012; 109:8878–8883.
- Larson G, Piperno DR, Allaby RG, et al. Current perspectives and the future of domestication studies. *Proc Natl Acad Sci U S A*. 2014; 111:6139–6146. [PubMed: 24757054]
- Li Y, Vonholdt BM, Reynolds A, et al. Artificial selection on brain-expressed genes during the domestication of dog. *Mol Biol Evol*. 2013; 30:1867–1876. [PubMed: 23660689]
- Li Y, Wang G-D, Wang M-S, et al. Domestication of the dog from the wolf was promoted by enhanced excitatory synaptic plasticity: a hypothesis. *Genome biology and evolution*. 2014; 6:3115–3121. [PubMed: 25377939]
- Liew CC, Ma J, Tang HC, Zheng R, Dempsey AA. The peripheral blood transcriptome dynamically reflects system wide biology: a potential diagnostic tool. *J Lab Clin Med*. 2006; 147:126–132. [PubMed: 16503242]
- Lin L, Faraco J, Li R, et al. The sleep disorder canine narcolepsy is caused by a mutation in the hypocretin (orexin) receptor 2 gene. *Cell*. 1999; 98:365–376. [PubMed: 10458611]
- Lindblad-Toh K, Wade CM, Mikkelsen TS, et al. Genome sequence, comparative analysis and haplotype structure of the domestic dog. *Nature*. 2005; 438:803–819. [PubMed: 16341006]
- Lippert C, Listgarten J, Liu Y, et al. FaST linear mixed models for genome-wide association studies. *Nat Meth*. 2011; 8:833–835.
- Lister R, O'Malley RC, Tonti-Filippini J, et al. Highly Integrated Single-Base Resolution Maps of the Epigenome in Arabidopsis. *Cell*. 2008; 133:523–536. [PubMed: 18423832]
- Lofton-Day C, Model F, Devos T, et al. DNA methylation biomarkers for blood-based colorectal cancer screening. *Clin Chem*. 2008; 54:414–423. [PubMed: 18089654]
- MacNulty DR, Smith DW, Mech LD, Eberly LE. Body size and predatory performance in wolves: is bigger better? *J Anim Ecol*. 2009a; 78:532–539. [PubMed: 19175444]
- MacNulty DR, Smith DW, Mech LD, Vucetich JA, Packer C. Nonlinear effects of group size on the success of wolves hunting elk. *Behavioral Ecology*. 2011; 23:75–82.
- MacNulty DR, Smith DW, Vucetich JA, et al. Predatory senescence in ageing wolves. *Ecol Lett*. 2009b; 12:1347–1356. [PubMed: 19780789]
- Matzke M, Kanno T, Huettel B, Daxinger L, Matzke AJM. Targets of RNA-directed DNA methylation. *Current Opinion in Plant Biology*. 2007; 10:512–519. [PubMed: 17702644]
- McClintock B. The significance of responses of the genome to challenge. *Science*. 1984; 226:792–801. [PubMed: 15739260]
- McRae AF, Powell JE, Henders AK, et al. Contribution of genetic variation to transgenerational inheritance of DNA methylation. *Genome Biol*. 2014; 15:R73. [PubMed: 24887635]

- Meissner A, Gnirke A, Bell GW, et al. Reduced representation bisulfite sequencing for comparative high-resolution DNA methylation analysis. *Nucleic Acids Res.* 2005; 33:5868–5877. [PubMed: 16224102]
- Minnick MF, Stillwell LC, Heineman JM, Stiegler GL. A highly repetitive DNA sequence possibly unique to canids. *Gene.* 1992; 110:235–238. [PubMed: 1537560]
- Naito K, Cho E, Yang G, et al. Dramatic amplification of a rice transposable element during recent domestication. *Proceedings of the National Academy of Sciences.* 2006; 103:17620–17625.
- Ong CT, Corces VG. Enhancer function: new insights into the regulation of tissue-specific gene expression. *Nat Rev Genet.* 2011; 12:283–293. [PubMed: 21358745]
- Ostrander EA, Wayne RK. The canine genome. *Genome Res.* 2005; 15:1706–1716. [PubMed: 16339369]
- Ovodov ND, Crockford SJ, Kuzmin YV, et al. A 33,000-year-old incipient dog from the Altai Mountains of Siberia: evidence of the earliest domestication disrupted by the Last Glacial Maximum. *PLoS One.* 2011; 6:e22821. [PubMed: 21829526]
- Parker HG, Kim LV, Sutter NB, et al. Genetic structure of the purebred domestic dog. *Science.* 2004; 304:1160–1164. [PubMed: 15155949]
- Peaston AE, Whitelaw E. Epigenetics and phenotypic variation in mammals. *Mamm Genome.* 2006; 17:365–374. [PubMed: 16688527]
- Pele M, Turet L, Kessler JL, Blot S, Panthier JJ. SINE exonic insertion in the PTPLA gene leads to multiple splicing defects and segregates with the autosomal recessive centronuclear myopathy in dogs. *Hum Mol Genet.* 2005; 14:1417–1427. [PubMed: 15829503]
- Petroff OAC. Book Review: GABA and Glutamate in the Human Brain. *The Neuroscientist.* 2002; 8:562–573. [PubMed: 12467378]
- Popova NK, Kulikov AV, Avgustinovich DF, Voitenko NN, Trut LN. [Effect of domestication of the silver fox on the main enzymes of serotonin metabolism and serotonin receptors]. *Genetika.* 1997; 33:370–374. [PubMed: 9244768]
- Purcell S, Neale B, Todd-Brown K, et al. PLINK: a tool set for whole-genome association and population-based linkage analyses. *The American Journal of Human Genetics.* 2007; 81:559–575. [PubMed: 17701901]
- Rakyan VK, Down TA, Balding DJ, Beck S. Epigenome-wide association studies for common human diseases. *Nat Rev Genet.* 2011; 12:529–541. [PubMed: 21747404]
- Rapp RA, Wendel JF. Epigenetics and plant evolution. *New Phytol.* 2005; 168:81–91. [PubMed: 16159323]
- Reinius LE, Acevedo N, Joerink M, et al. Differential DNA methylation in purified human blood cells: implications for cell lineage and studies on disease susceptibility. *PLoS One.* 2012; 7:e41361. [PubMed: 22848472]
- Richards CL, Bossdorf O, Pigliucci M. What Role Does Heritable Epigenetic Variation Play in Phenotypic Evolution? *BioScience.* 2010; 60:232–237.
- Sablin M, Khlopachev G. The earliest Ice Age dogs: evidence from Eliseevichi 11. *Current Anthropology.* 2002; 43:795–799.
- Saetre P, Lindberg J, Leonard JA, et al. From wild wolf to domestic dog: gene expression changes in the brain. *Brain Res Mol Brain Res.* 2004; 126:198–206. [PubMed: 15249144]
- Savolainen P, Zhang Y-p, Luo J, Lundeberg J, Leitner T. Genetic evidence for an East Asian origin of domestic dogs. *Science.* 2002; 298:1610–1613. [PubMed: 12446907]
- Saxonov S, Berg P, Brutlag DL. A Genome-Wide Analysis of CpG Dinucleotides in the Human Genome Distinguishes Two Distinct Classes of Promoters. *Proceedings of the National Academy of Sciences of the United States of America.* 2006; 103:1412–1417. [PubMed: 16432200]
- Schmitz RJ, Schultz MD, Lewsey MG, et al. Transgenerational epigenetic instability is a source of novel methylation variants. *Science.* 2011; 334:369–373. [PubMed: 21921155]
- Schmitz RJ, Schultz MD, Urich MA, et al. Patterns of population epigenomic diversity. *Nature.* 2013; 495:193–198. [PubMed: 23467092]
- Shah S, McRae AF, Marioni RE, et al. Genetic and environmental exposures constrain epigenetic drift over the human life course. *Genome Res.* 2014; 24:1725–1733. [PubMed: 25249537]

- Shearman JR, Wilton AN. Origins of the domestic dog and the rich potential for gene mapping. *Genet Res Int.* 2011; 2011:579308. [PubMed: 22567358]
- Skoglund P, Ersmark E, Palkopoulou E, Dalén L. Ancient Wolf Genome Reveals an Early Divergence of Domestic Dog Ancestors and Admixture into High-Latitude Breeds. *Current Biology.* 2015
- Slotkin RK, Martienssen R. Transposable elements and the epigenetic regulation of the genome. *Nat Rev Genet.* 2007; 8:272–285. [PubMed: 17363976]
- Smit AFA. Interspersed repeats and other mementos of transposable elements in mammalian genomes. *Curr Opin Genet Dev.* 1999; 9:657–663. [PubMed: 10607616]
- Smit AFA, Hubley R, Green P. RepeatMasker Open-3.0. 1996
- Stahler DR, MacNulty DR, Wayne RK, vonHoldt B, Smith DW. The adaptive value of morphological, behavioural and life-history traits in reproductive female wolves. *J Anim Ecol.* 2013; 82:222–234. [PubMed: 23043440]
- Struyf A, Hubert M, Rousseeuw PJ. Integrating robust clustering techniques in S-PLUS. *Computational Statistics & Data Analysis.* 1997; 26:17–37.
- Studer A, Zhao Q, Ross-Ibarra J, Doebley J. Identification of a functional transposon insertion in the maize domestication gene *tb1*. *Nat Genet.* 2011; 43:1160–1163. [PubMed: 21946354]
- Sutter NB, Ostrander EA. Dog star rising: the canine genetic system. *Nat Rev Genet.* 2004; 5:900–910. [PubMed: 15573122]
- Suzuki MM, Bird A. DNA methylation landscapes: provocative insights from epigenomics. *Nat Rev Genet.* 2008; 9:465–476. [PubMed: 18463664]
- Szyf M, Weaver IC, Champagne FA, Diorio J, Meaney MJ. Maternal programming of steroid receptor expression and phenotype through DNA methylation in the rat. *Front Neuroendocrinol.* 2005; 26:139–162. [PubMed: 16303171]
- Trut LN, Plyusnina IZ, Oskina IN. An experiment on fox domestication and debatable issues of evolution of the dog. *Russian Journal of Genetics.* 2004; 40:644–655.
- Turner SD. qqman: an R package for visualizing GWAS results using Q-Q and manhattan plots. *bioRxiv.* 2014
- Van der Auwera I, Elst HJ, Van Laere SJ, et al. The presence of circulating total DNA and methylated genes is associated with circulating tumour cells in blood from breast cancer patients. *Br J Cancer.* 2009; 100:1277–1286. [PubMed: 19367284]
- Vonholdt BM, Pollinger JP, Lohmueller KE, et al. Genome-wide SNP and haplotype analyses reveal a rich history underlying dog domestication. *Nature.* 2010; 464:898–902. [PubMed: 20237475]
- Vonholdt BM, Stahler DR, Smith DW, et al. The genealogy and genetic viability of reintroduced Yellowstone grey wolves. *Mol Ecol.* 2008; 17:252–274. [PubMed: 17877715]
- Vonholdt BM, Takuno S, Gaut BS. Recent retrotransposon insertions are methylated and phylogenetically clustered in japonica rice (*Oryza sativa* spp. japonica). *Mol Biol Evol.* 2012; 29:3193–3203. [PubMed: 22593226]
- Wang GD, Zhai W, Yang HC, et al. The genomics of selection in dogs and the parallel evolution between dogs and humans. *Nat Commun.* 2013a; 4:1860. [PubMed: 23673645]
- Wang H, Ferguson GD, Pineda VV, Cundiff PE, Storm DR. Overexpression of type-1 adenylyl cyclase in mouse forebrain enhances recognition memory and LTP. *Nat Neurosci.* 2004; 7:635–642. [PubMed: 15133516]
- Wang J, Duncan D, Shi Z, Zhang B. Web-based gene set analysis toolkit (WebGestalt): Update 2013. *Nucleic Acids Res.* 2013b; 41:W77–W83. [PubMed: 23703215]
- Wang J, Fan C. A Neutrality Test for Detecting Selection on DNA Methylation Using Single Methylation Polymorphism Frequency Spectrum. *Genome biology and evolution.* 2015; 7:154–171. [PubMed: 25539727]
- Wang W, Kirkness EF. Short interspersed elements (SINEs) are a major source of canine genomic diversity. *Genome Res.* 2005; 15:1798–1808. [PubMed: 16339378]
- Wayne RK, Ostrander EA. Lessons learned from the dog genome. *Trends Genet.* 2007; 23:557–567. [PubMed: 17963975]
- Wayne RK, vonHoldt BM. Evolutionary genomics of dog domestication. *Mamm Genome.* 2012; 23:3–18. [PubMed: 22270221]

- Whitelaw E, Martin DIK. Retrotransposons as epigenetic mediators of phenotypic variation in mammals. *Nat Genet.* 2001; 27:361–365. [PubMed: 11279513]
- Whiting PJ, Bonnert TP, McKernan RM, et al. Molecular and Functional Diversity of the Expanding GABA-A Receptor Gene Family. *Annals of the New York Academy of Sciences.* 1999; 868:645–653. [PubMed: 10414349]
- Wieczorek G, Asemussen A, Model F, et al. Quantitative DNA methylation analysis of FOXP3 as a new method for counting regulatory T cells in peripheral blood and solid tissue. *Cancer research.* 2009; 69:599–608. [PubMed: 19147574]
- Wiener P, Wilkinson S. Deciphering the genetic basis of animal domestication. *Proc Biol Sci.* 2011; 278:3161–3170. [PubMed: 21885467]
- Zemach A, McDaniel IE, Silva P, Zilberman D. Genome-wide evolutionary analysis of eukaryotic DNA methylation. *Science.* 2010; 328:916–919. [PubMed: 20395474]
- Zhang B, Kirov S, Snoddy J. WebGestalt: an integrated system for exploring gene sets in various biological contexts. *Nucleic Acids Res.* 2005; 33:W741–W748. [PubMed: 15980575]
- Zhang X, Yazaki J, Sundaresan A, et al. Genome-wide high-resolution mapping and functional analysis of DNA methylation in arabidopsis. *Cell.* 2006; 126:1189–1201. [PubMed: 16949657]
- Zou J, Lippert C, Heckerman D, Aryee M, Listgarten J. Epigenome-wide association studies without the need for cell-type composition. *Nat Methods.* 2014; 11:309–311. [PubMed: 24464286]

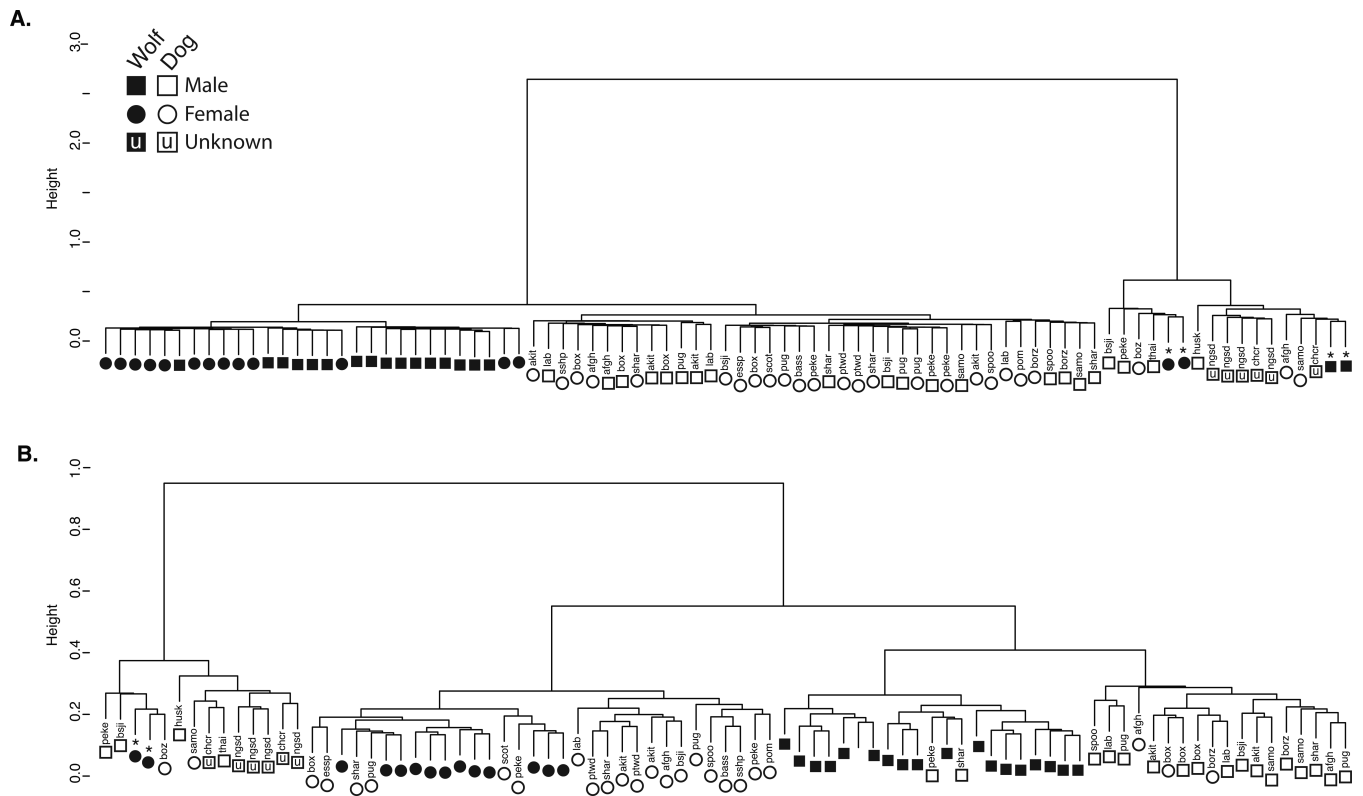


Figure 1. Hierarchical cluster analysis of SMP data from 86 samples (excluding outliers) using the Gower dissimilarity matrix and Ward's cluster method for the **A**) autosomes ($n_{SMPs}=23,997$); **B**) X chromosome ($n_{SMPs}=208$). (See Supplemental Table S1 for codes). Square symbols represent males, circles for females, while “u” are individuals with unknown sex. * indicates wolves that were processed in a different batch.

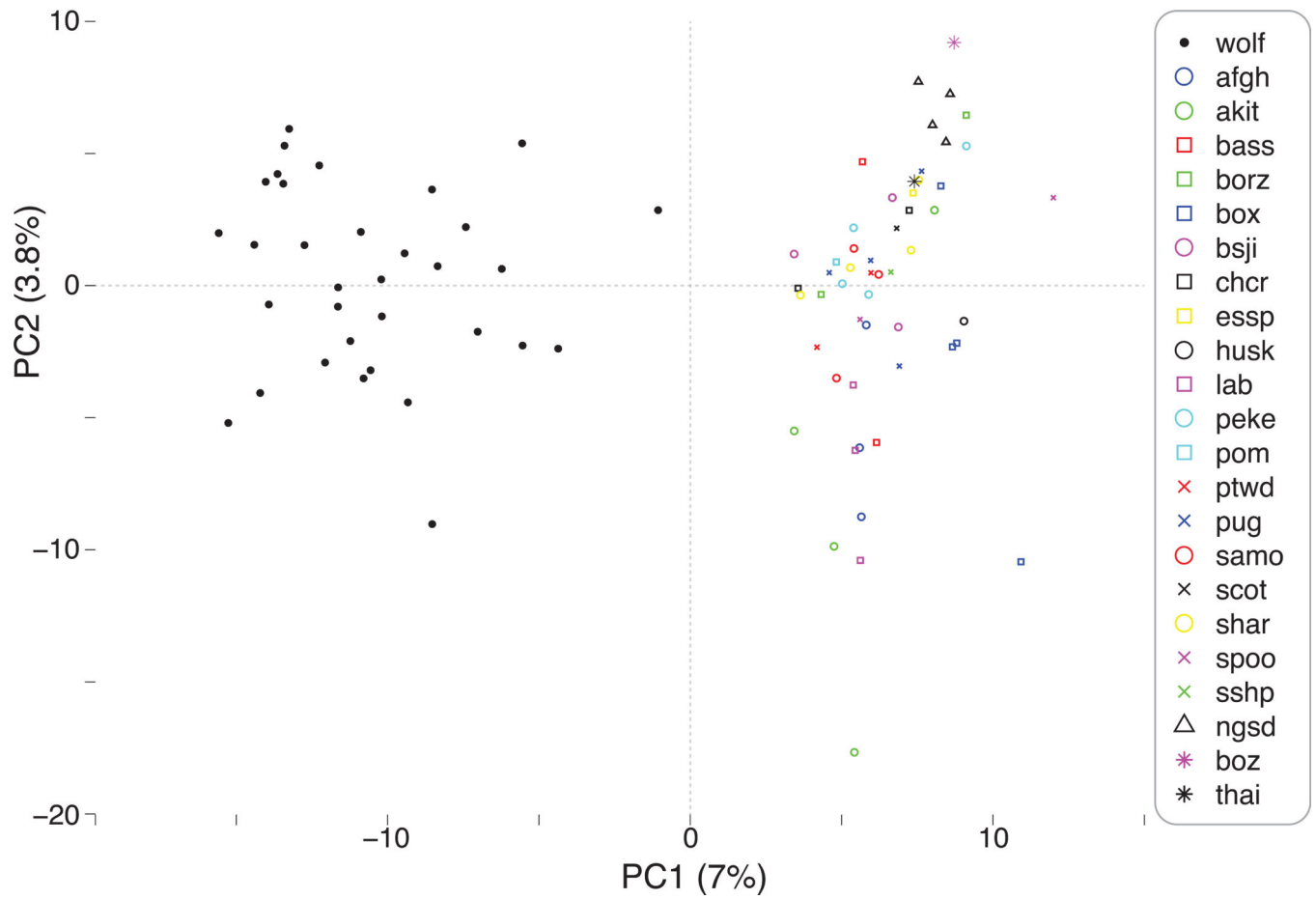


Figure 2. PCA (% variance explained) of 1,307 unlinked SMPs after excluding seven outlier samples. See Supplemental Table S1 for breed codes.

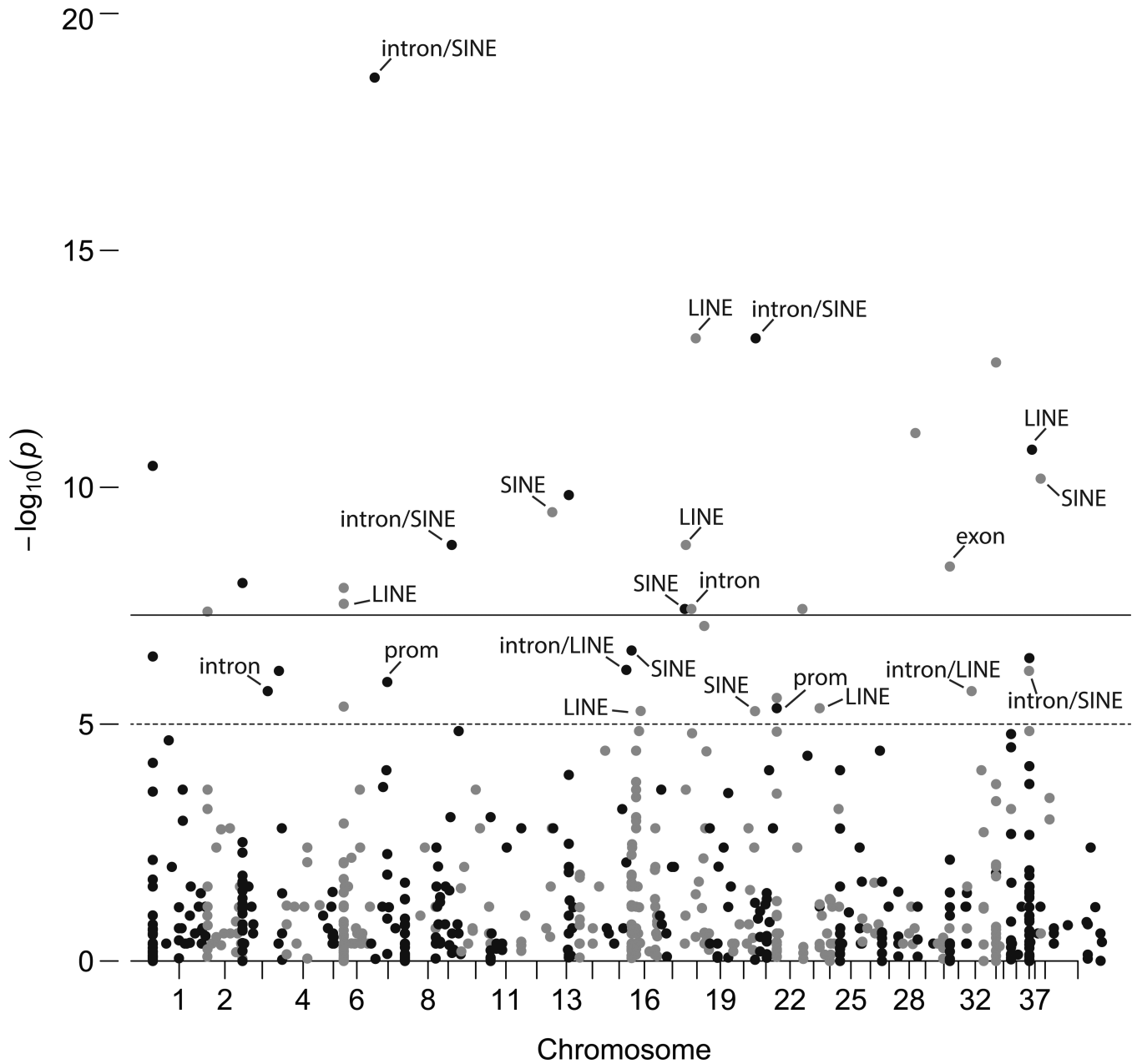


Figure 3.

An uncorrected association scan of 1,307 unlinked SMPs across YNP wolves and purebred dogs. Outliers ($-\log_{10}(p) > 5$, $n=36$) are intergenic unless otherwise noted (Abbreviation: prom, promoter). Dashed line ($p < 10^{-5}$) and solid lines ($p < 10^{-7}$) are indicated. All dots above the dashed line are significant outliers. Black dots indicate sites on odd-numbered chromosomes; gray for even-numbered chromosomes.

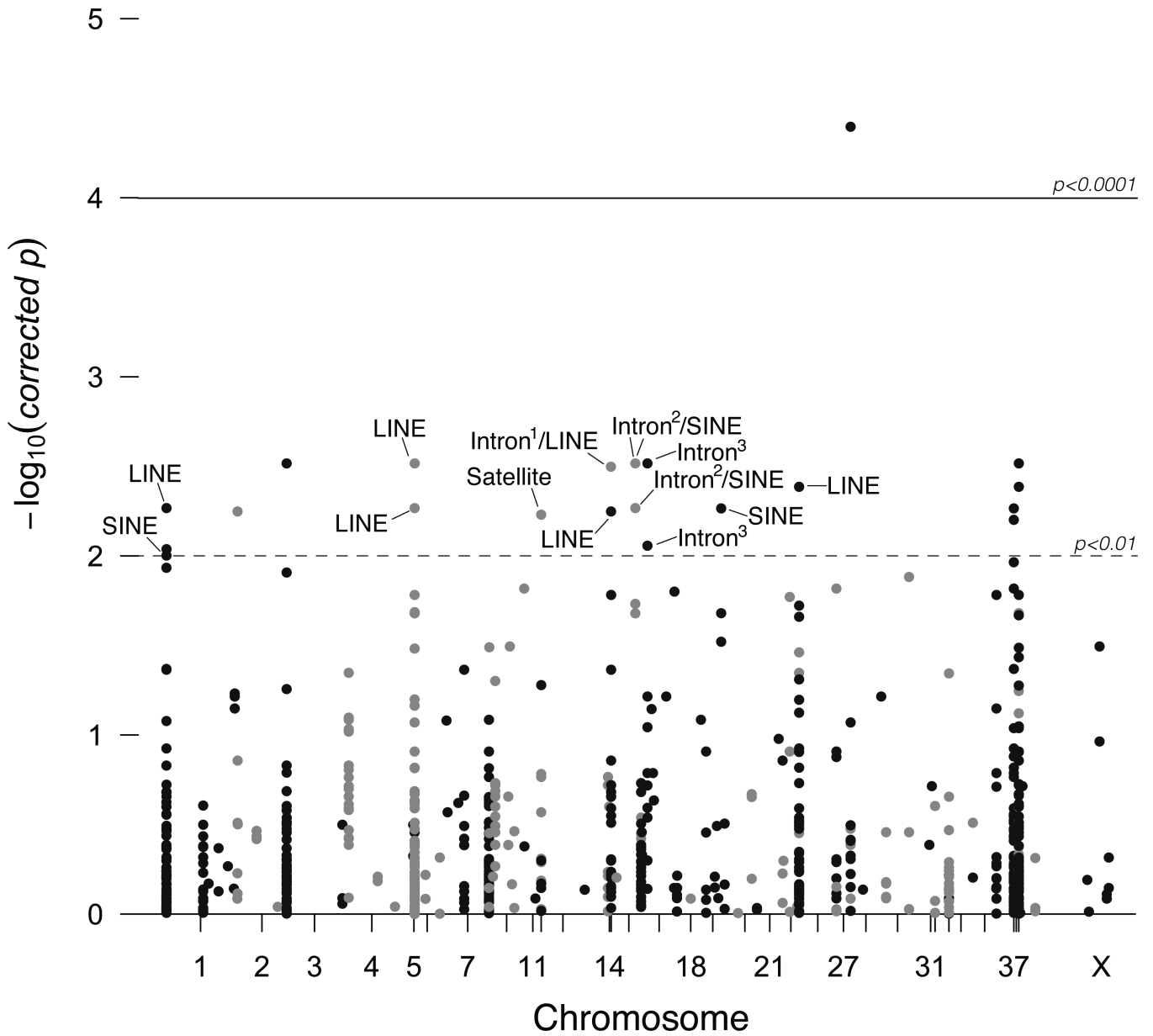


Figure 4.

Cell-type and covariate corrected association scan of methylation across YNP wolves and purebred dogs. Manhattan plot for 962 cytosines included in the association scan. Outliers are intergenic unless otherwise noted. Superscripts indicate Ensembl transcript identifiers: ¹ENSCAFT00000002993; ²ENSCAFT00000010950; ³ENSCAFT00000042812. All dots above the dashed line are significant outliers. Black dots indicate sites on odd-numbered chromosomes; gray for even-numbered chromosomes.

Table 1

Sites identified as significant outliers in both the uncorrected SMP association ($p < 0.05$) and corrected methylation scan ($p < 0.05$) between purebred dogs and YNP wolves. (Abbreviations: CFA, canine chromosome)

CFA	Position	Uncorrected p -value	Adjusted p -value	Gene name	Annotation	In repeat element?	Deviation from Hardy-Weinberg Equilibrium ($p < 0.05$)?
34	52849	<0.000	0.005		Intergenic		
6	49253	1.329×10^{-8}	0.017		Intergenic		
6*	172430	2.890×10^{-8}	0.003		Intergenic	LINE	
2	4609977	4.213×10^{-9}	0.006		Intergenic		
7	40156706	1.290×10^{-7}	0.043	DEGS1	Promoter		Dogs ($H_0=0.63; H_c=0.45$)
22	48102	2.803×10^{-6}	0.019		Intergenic		
6	49272	4.261×10^{-6}	0.005		Intergenic		
24	2471	4.614×10^{-8}	0.015		Intergenic	LINE	Dogs ($H_0=0.02; H_c=0.08$)
18	50124190	3.757×10^{-5}	0.030		Intergenic	SINE	Wolves ($H_0=0.16; H_c=0.12$)
1	202655	6.568×10^{-5}	0.005		Intergenic	LINE	
37	49946	7.712×10^{-5}	0.017		Intergenic	LINE	
1	72422	2.655×10^{-4}	0.043		Intergenic	LINE	
15	39451140	6.210×10^{-4}	0.021	SLC17A8	Intron	SINE	
9	34354850	9.100×10^{-4}	0.032	TUBD1	Exon	SINE	
11	22699	9.100×10^{-4}	0.006		Intergenic	Satellite	
6	168151	0.001	0.003		Intergenic		
37	28098	0.012	0.037		Intergenic		
37	105647	0.037	0.004		Intergenic		
37	21264	0.045	0.021		Intergenic		

* indicates a significant outlier in both uncorrected SMP association scan ($p < 0.05$) and top-ranking corrected methylation scan ($p < 0.005$).

Table 2

Genes identified through permutation analyses with significantly different neutrality estimates (D^m) between purebred dogs and YNP wolves ($D^w Diff$). The number of cytosines analyzed per gene is indicated in the parentheses following the gene name.

Ensembl gene(s) list	Description	Observed YNP wolf D^m	Observed purebred dog D^m	Observed $D^m Diff$	Two tailed p-value	M-F _{wolves}	M-F _{purebred}
SLC17A8 (6) [*]	Solute carrier family 17 (vesicular glutamate transporter), member 8	6.3571	1.3058	5.0512	0.0001	0.427	0.583
EPM2A (2) [*]	Epilepsy, progressive myoclonus type 2A, Lafora disease (laforin)	0.9911	-1.1272	2.1183	0.0007	0.878	0.975
MYRIP (5) [*]	Myosin VIIA and Rab interacting protein	1.2400	-0.8429	2.0828	0.0009	0.885	0.893
DCAF6 (2) [*]	DDB1 and CUL4 associated factor 6	0.8128	-1.0148	1.8277	0.0010	0.853	0.885
ANXA6 (4) [*]	Annexin A6	-1.2512	0.1042	-1.3554	0.0063	0.905	0.884
KLHL26 (2)	Kelch-like family member 26	-0.4590	1.0967	-1.5556	0.0209	0.207	0.214
ENSCAFT00000045887 (95) [*]	Uncharacterized protein	-0.9634	-2.1759	1.2124	0.0265	0.174	0.165
ENSCAFT00000012091 (11)	Uncharacterized protein	2.5544	-0.1056	2.6600	0.0472	0.905	0.888
Averaged Value						0.654	0.698

^{*} Gene contains repeat elements

Table 3

Heritability models (all analyses include covariates for batch, sex, and the first 5 PCs from *Fast-LMM-EWASher*). Asterisk indicates significance compared to Bonferroni threshold* based on total sites analyzed.

Analysis Model	Variance components	Total number of sites analyzed	Sites with non-zero variance components [#]	Sites with non-zero variance components and significant heritability*
Analysis 1	Additive Environmental	23997	4504	2282
Analysis 2	Additive Dominance Environmental	4504 (from Analysis 1)	1432	1362

[#]In analysis 1, both additive and environmental variance components are non-zero. In analysis 2, all three components (additive, dominance and environmental) are non-zero.

*The narrow sense heritability significance threshold from analysis 1 and the broad sense heritability significance threshold from analysis 2 equals $p < 2.08 \times 10^{-6}$.