


Genome-wide association analysis uncovers variants for reproductive variation across dog breeds and links to domestication

Samuel P. Smith,^{1,2,3,†} Julie B. Phillips,^{1,4,†} Maddison L. Johnson,¹ Patrick Abbot,¹
 John A. Capra^{1,5,6} and Antonis Rokas^{1,5,6} 



¹Department of Biological Sciences, Vanderbilt University, Nashville, TN 37203, USA; ²Center for Computational Molecular Biology, Brown University, Providence, RI 02912, USA; ³Department of Ecology and Evolutionary Biology, Brown University, Providence, RI 02912, USA; ⁴Department of Biological Sciences, Cumberland University, Lebanon, TN 37087, USA; ⁵Department of Biomedical Informatics, Vanderbilt University, Nashville, TN 37203, USA and ⁶Vanderbilt Genetics Institute, Vanderbilt University Medical Center, Nashville, TN 37232, USA

*Corresponding author: Department of Biological Sciences, Vanderbilt University, Nashville, TN 37235, USA. Tel: 615-936-3892; E-mail: antonis.rokas@vanderbilt.edu

†These authors contributed equally to this work.

Received 16 December 2018; revised version accepted 9 May 2019

ABSTRACT

Background and objectives: The diversity of eutherian reproductive strategies has led to variation in many traits, such as number of offspring, age of reproductive maturity and gestation length. While reproductive trait variation has been extensively investigated and is well established in mammals, the genetic loci contributing to this variation remain largely unknown. The domestic dog, *Canis lupus familiaris* is a powerful model for studies of the genetics of inherited disease due to its unique history of domestication. To gain insight into the genetic basis of reproductive traits across domestic dog breeds, we collected phenotypic data for four traits, cesarean section rate, litter size, stillbirth rate and gestation length, from primary literature and breeders' handbooks.

Methodology: By matching our phenotypic data to genomic data from the Cornell Veterinary Biobank, we performed genome-wide association analyses for these four reproductive traits, using body mass and kinship among breeds as covariates.

Results: We identified 12 genome-wide significant associations between these traits and genetic loci, including variants near *CACNA2D3* with gestation length, *MSRB3* and *MSANTD1* with litter size, *SMOC2* with cesarean section rate and *UFM1* with stillbirth rate. A few of these loci, such as *CACNA2D3* and *MSRB3*, have been previously implicated in human reproductive pathologies, whereas

others have been associated with domestication-related traits, including brachycephaly (*SMOC2*) and coat curl (*KRT71*).

Conclusions and implications: We hypothesize that the artificial selection that gave rise to dog breeds also influenced the observed variation in their reproductive traits. Overall, our work establishes the domestic dog as a system for studying the genetics of reproductive biology and disease.

LAY SUMMARY The genetic contributors to variation in mammalian reproductive traits remain largely unknown. We took advantage of the domestic dog, a powerful model system, to test for associations between genome-wide variants and four reproductive traits (cesarean section rate, litter size, stillbirth rate and gestation length) that vary extensively across breeds. We identified associations at a dozen loci, including ones previously associated with domestication-related traits, suggesting that selection on dog breeds also influenced their reproductive traits.

KEYWORDS: life history; artificial selection; tradeoff; preterm birth; prematurity; pregnancy

INTRODUCTION

Mammals exhibit wide variation in traits associated with reproduction [1–3]. For example, gestation length can range from 12 days in the Gray dwarf hamster, *Cricetulus migratorius*, to 21 months in the African bush elephant, *Loxodonta africana* [4–6]; neonate size can range from <1 g in the shrew family (Soricidae), to more than a metric ton in the baleen whales (Balaenopteridae) [4, 6]; and neonates can be either precocial (e.g. cricetid rodents, rabbits and canids) or altricial (e.g. hystricomorph rodents, ungulates and cetaceans) [1]. This variation in reproductive traits also extends to methods of implantation [7], structure of the placenta [8, 9] and lactation strategies [10, 11]. Not surprisingly, many reproductive traits also exhibit substantial intraspecific variation [5]. For example, many mammals exhibit intraspecific variation in gestation length, including primates [12], rat and rabbits [13], as well as the domesticated cattle [14] and thoroughbred horses [15]. Similarly, body fat percentages, which are associated with the energetics of reproduction, vary greatly between wild and captive baboons, and intraspecific variation among captive lemurs can vary from 8 to 41% [16].

The existence of phenotypic variation in reproductive traits is well established, and can inform our understanding of the factors that shape patterns of survival and reproduction in both agricultural [17–20] and human populations [21]. Not surprisingly, most genome-wide association (GWAS) studies of reproductive traits focus on economically important traits in domesticated species, such as reproductive seasonality in rabbits [17], infertility in pigs [18] and dairy traits in cattle [19]. GWAS studies focused on understanding human reproductive biology and its associated pathologies have also shed light on the genetic basis of reproductive traits, including birth weight [22] and gestational duration or length [23, 24]. For example, maternal variation in six genomic loci (*ADCY5*, *AGTR2*, *EBF1*, *EEFSEC*, *RAP2C* and *WNT4*) is associated with gestational duration and preterm birth [24]. While these studies contribute to our understanding of the genetic architecture of reproductive traits, we still understand very little about the molecular pathways underlying this variation and are unable to explain the majority of the heritability in reproductive traits [25–27].

To address this challenge, we studied the genetics of reproductive traits using the domestic dog as a model system. The dog is well-suited to this question, because the domestication bottleneck followed by intense artificial selection and inbreeding imposed over the past 300 years has led to the generation of more than 340 recognized breeds that exhibit dramatic morphological variation [28–30]. Domestic dog breeds also show substantial variation in their reproductive traits. For example, Pomeranians and Norfolk Terriers typically have only 2 pups per litter, whereas Dalmatians and Rhodesian Ridgebacks typically sire 8–9 pups per litter [31]. Similarly, 80–90% of French Bulldogs and Boston Terriers are born via cesarean section due to cephalopelvic disproportion, whereas only 2–3% of Australian Shepherds and Shar Peis require cesareans [31]. Recent analyses have begun to study the genetic mechanisms that underlie the remarkable morphological variation between modern dog breeds in diverse traits such as snout length, ear erectness and tail curliness [32–35], as well as genetic disease [36].

To gain insight into the genetic basis of reproductive traits across domestic dog breeds, we collected phenotypic data for four reproductive traits, namely cesarean section rate, litter size, stillbirth rate and gestation length. We synthesized data from the primary literature and breeders' handbooks to obtain coverage of between 23 (gestation length) and 97 (cesarean section rate) dog breeds, as well as body mass data from 101 dog breeds. By matching our phenotypic data to genome-wide genotypic data from the Cornell Veterinary Biobank, we performed GWAS analyses and identified 12 genetic loci that are significantly associated with these reproductive traits (using log body mass as a covariate). Some of these variants are in or near genes previously implicated in human reproduction-related pathologies, whereas others are in or near genes associated with domestication-related traits. For example, we found that variation in a gene previously identified to be involved in brachycephaly is also significantly associated with rates of cesarean sections and that variation in genes previously linked to coat phenotypes, such as curliness, is also associated with litter size. These results suggest that selection for breed-specific morphological traits during dog domestication may have also

influenced variation in reproductive traits. More broadly, our results illustrate that the domestic dog is a promising and tractable system for studying the genetics of reproductive traits.

METHODS


Genotypic and phenotypic data

To identify single nucleotide polymorphisms (SNPs) that are significantly associated with reproductive traits, we used a previously published data set containing 160 727 SNPs from 4342 individual dogs across 162 breeds genotyped using the Illumina 173k CanineHD array that were downloaded from [http://datadryad.org/resource/doi: 10.5061/dryad.266k4](http://datadryad.org/resource/doi:10.5061/dryad.266k4) [32]. Following the original authors, SNPs with a genotyping rate (i.e. the proportion of genotypes per marker with non-missing data) below 95% and heterozygosity ratios (i.e. the ratio of the number of heterozygous SNPs divided by the number of non-reference homozygous SNPs) below 0.25 or above 1.0 were removed. To reduce potential error stemming from SNP misidentification in our analyses, we included only SNPs with a minor allele frequency (MAF) >0.05, since SNPs with very low MAFs are more prone to error due to the small number of samples that have the called nucleotide. Application of these two filters removed 10 893 SNPs, which resulted in a final data set that contained 149 834 variants.

Phenotypic reproductive trait data for litter size (number of pups), cesarean rate, stillbirth rate and gestation length across 128 breeds were collected from a variety of breeder's handbook and primary journal articles [31, 37–47] (see also [Supplementary File S1](#)). We also included body mass as a control trait. Each breed was assigned the average breed value for each phenotype; the full list of the values for all four reproductive traits and body mass across the 128 breeds is provided in [Supplementary Table S1](#). For the body mass control, our collected trait values overlapped with the genotypic data [32] for 101 breeds corresponding to 3384 individuals ([Table 1](#)). For the reproductive traits, our collected cesarean section rate trait values overlapped with the genotypic data for 95 breeds (3194 individuals), our litter size trait values for 60 breeds (2617 individuals), our stillbirth rate values for 56 breeds (2590 individuals) and our gestation length values for 23 breeds (1908 individuals) ([Table 2](#)). Finally, we note that while we replicate some of the signals for the body mass GWAS from Hayward *et al.* [32], where they use individual phenotype values, we instead assigned breed averages ([Supplementary Fig. S1](#)). We performed this analysis in order to validate that breed averages are a useful proxy for individual phenotype values when testing the genetic architecture of complex traits.

Genome-wide association analyses

To test SNPs for associations with the four reproductive traits of interest, we conducted a GWAS analysis for each individual trait

 **Table 1.** Numbers of breeds and individuals with overlapping phenotypes and genotypes included in our analysis


Trait	Number of overlapping breeds	Number of overlapping individuals
Body mass	101	3384
Cesarean section rate	97	3194
Litter size	60	2617
Stillbirth rate	57	2590
Gestation length	23	1908

using log body mass as a covariate, and accounting for kinship, as well as for body mass as a proof of concept. For each phenotype, we first log transformed all breed-average phenotype values. We then performed a Shapiro–Wilk test of normality ([Table 2](#)) for the distribution of log phenotypes for each phenotype and normalized using a Box–Cox transformation when necessary (cesarean section rate, litter size). All GWAS analyses were run using a linear-mixed model as implemented in the program GEMMA, version 0.94 [48]. Numerous studies have shown that the vast majority of morphological, ecological and physiological traits vary as a function of an organism's body mass [49–51] as well as a function of kinship [32, 33]. Most notably for the purpose of this study, body mass has been previously shown to be strongly correlated with litter weight [52–54], neonate weight [52–55] and gestation length [6, 52, 53, 56, 57].

To ensure our analysis reflected the reproductive trait of interest and not SNPs associated with body mass, we used log body mass as a covariate for all reproductive trait analyses. To be able to do so, we pruned our genotypic data so that they included only dog breeds (and individuals) for which we had both body mass and reproductive trait of interest values (see [Supplementary Table S1](#)).

To account for population stratification, we calculated a kinship matrix of the included breeds using GEMMA and included it as a random effect in each association analysis. Each value of a kinship matrix describes the probability that a particular allele from two randomly chosen individuals at a given locus is identical by descent [58]. Finally, to control for inflated *P*-value significance from the testing of multiple hypotheses, we used a significance threshold of $P = 3.3 \times 10^{-7}$ (Bonferroni cutoff of $\alpha = 0.05$, $N = 149\,834$) for all analyses. All reported *P*-values are Wald's *P*-values as calculated in GEMMA [48].

Finally, to validate variants significantly associated with at least one reproductive life history trait in the domestic dog, we performed a permutation analysis. For each trait, we randomly permuted the assignment of breed-specific phenotypes while holding body mass constant for each breed across 1000 permutations. In each permutation, we used the transformed phenotype

 **Table 2.** Shapiro–Wilk test of normality for the log transformed distribution of values for each phenotype

Trait	Shapiro–Wilk <i>P</i> -value
Cesarean section	0.026
Gestation length	0.491
Litter size	0.007
Stillbirth rate	0.145

value for each breed. Specifically, we used the log phenotype transformation for gestation length and stillbirth rate, and normalized the log phenotype transformations for cesarean section and litter size (Table 2). We then regressed the randomly assigned reproductive phenotypes onto log body mass and assigned the residual for each breed as the phenotype for all of the individuals of that breed. Next, using GEMMA [48], we performed an association test for each variant to obtain a single permuted *P*-value. The permutation *P*-value in Table 3 corresponds to the number of times that a single permuted *P*-value was less than the empirical *P*-value from the original analysis. Any variant with a permutation *P*-value >0.05 was acknowledged as a potential false positive association.

To gain insight into the genetic elements putatively involved with the traits of interest, we mapped all SNPs found to be significantly and marginally associated with each trait of interest using custom perl and R scripts to the CanFam3.1.87 dog genome assembly [59, 60]. Transcript IDs were mapped to gene names using bioconductor biomaRt interface to the ENSEMBL biomart [61]. If the significant SNP was outside gene boundaries, we reported the nearest upstream or downstream gene. Manhattan plots and quantile–quantile plots were generated using R 3.1.2 [62] with the qqman package [63]. Calculation of the λ inflation parameter, a metric of any existing systematic bias in the data set, was calculated using the GenABEL R package [64] and was used to interpret Type I error rate in the multiple testing of GWAS analyses [65].

RESULTS

To identify SNPs that are significantly associated with four reproductive traits in domestic dog breeds, we conducted across-breed GWAS analyses using a multivariate linear-mixed model implemented in the program GEMMA [48]. Number of individuals and distribution of breed varied with analysis (Supplementary Table S1). After filtering for MAF (MAF < 0.05; 10 804 SNPs were excluded), 149 834 .834 SNPs were included in the GWAS analysis for each reproductive trait. To control for inflated *P*-value significance from the testing of multiple hypotheses, we used a significance threshold of $P = 3.3 \times 10^{-7}$ (Bonferroni cutoff of $\alpha = 0.05$, $N = 149\,834$, 834) for all analyses. To validate our GWAS approach and analytical choices, we first used our collected values for body mass, a trait whose genetic associations have been previously extensively

studied in dogs [32]. However, whereas Hayward *et al.* [32] used individual body mass measurements in their analyses, we assigned the breed-average body mass as the phenotype for all individuals in a breed to test whether we could replicate their inferred associations for a complex trait when we used average breed value as a proxy. As expected, our analysis recovered the major genes associated with dog breed body mass variation, including *IGF1* ($P = 2.1 \times 10^{-31}$), *SMAD2* ($P = 1.2 \times 10^{-17}$) and *IGF2BP2* ($P = 5.1 \times 10^{-11}$) ($\times 10^{-11}$) (Supplementary Fig. S1 and Table S2).

Genetic loci that significantly associate with cesarean section rate

To examine whether there is variation in cesarean section rate among breeds, we first identified cesarean section rate values for a total of 97 of the 162 dog breeds with genotypic data (Supplementary Table S1). The cesarean section rate values were derived from a British survey across 151 breeds covering 13 141 bitches, which had whelped 22 005 litters over the course of a 10 year period [30]. The frequency of cesarean sections was estimated as the percentage of litters reported to be born by cesarean section. Among the 97 breeds with overlapping genetic data, the median cesarean section rate is 17.1%, with a minimum of 0% in Curly Coated Retrievers and Silky Terriers and a maximum of 92.3% in Boston Terriers (Supplementary Fig. S3A).

To identify SNPs that are significantly associated with the observed variation in cesarean section rate across domestic dog breeds, we conducted an across-breed GWAS analysis using 149 834 149, 834 SNPs and cesarean section values across 95 dog breeds (Fig. 1A, Supplementary Fig. S2A). As outlined in the ‘permutation analysis’ section of the Methods, we additionally performed a breed-specific permutation and present those variants with a permutation *P*-value >0.05 as suggestive associations. We identified four significant SNPs (Supplementary Table S3), all of which mapped to genes, namely sparc-related modular calcium-binding protein 2 (*SMOC2*, uncorrected $P = 2.0 \times 10^{-7}$, Perm $P = 0.011$), two linked SNPs (uncorrected $P = 3.08 \times 10^{-7}$, Perm $P = 0.001$ and uncorrected $P = 1.61 \times 10^{-7}$, Perm $P = 0.001$, respectively) that mapped to *MAT2B*, and a fourth that mapped to the intergenic region between the *CD36* glycoprotein and a lincRNA (uncorrected $P = 9.7 \times 10^{-8}$, Perm $P = 0.024$) (Fig. 1A, Table 3).

The first significantly associated SNP is in the intron between exons 13 and 14 of *SMOC2*, a gene that is associated with dog brachycephaly and whose variation accounts for 36% of facial length variation in dogs [35, 66]. In humans, *SMOC2* is highly expressed in endometrium as well as other reproductive tissues, including the fallopian tubes, ovaries and cervix (Fig. 2) [67]. The second and third SNPs are nearby the *MAT2B* gene. Although not previously associated with reproductive traits, this gene has been shown to play a major role in intramuscular preadipocyte differentiation in domestic pigs [68]. Finally, the fourth significant SNP



Table 3. Summary of genes that contain or are adjacent to the SNPs that are associated with variation in reproductive traits across dog breeds

Associated trait	Chromosome	Base pair	Mapped gene	Empirical P-value	Permutation P-value	Associated domesticated trait
Cesarean section	1	55983871	<i>SMOC2</i>	6.69e-09	0.002	Brachycephaly (Dogs)
Cesarean section	4	46570650	<i>MAT2B</i>	3.08e-07	0.001	Intramuscular preadipocyte differentiation (Pig)
Cesarean section	4	46592551	<i>MAT2B</i>	1.61e-07	0.001	Intramuscular preadipocyte differentiation (Pig)
Cesarean section	18	20272961	<i>CD36</i>	7.9e-11	0.017	Fertility and milk production (Cattle)
Gestation length	6	57457184	<i>HFM1</i>	1.16e-07	0.014	
Gestation length	20	35206774	<i>CACNA2D3</i>	3.13e-07	0.01	Blastocyst development (Cattle)
Gestation length	24	23382682		1.64e-08	0.08	
Gestation length	24	36399705	<i>SLC9A8</i>	6.75e-11	0.003	
Gestation length	24	36400728	<i>SLC9A8</i>	6.75e-11	0.003	
Gestation length	28	25752710	<i>PCGF5</i>	2.13e-07	0.003	
Litter size	1	93219668	<i>RCL1</i>	5.73e-08	0.001	
Litter size	3	61055458	<i>MSANTD1</i>	2.59e-07	0.001	
Litter size	3	61062626	<i>MSANTD1</i>	1.91e-08	0.001	
Litter size	3	61155415	<i>MSANTD1</i>	3.02e-10	0.001	
Litter size	3	61209777	<i>MSANTD1</i>	3.02e-10	0.001	
Litter size	3	61220065	<i>MSANTD1</i>	1.97e-09	0.001	
Litter size	10	7884978	<i>MSRB3</i>	3.065e-07	0.001	Ear erectness (Dog)
Litter size	10	7920882	<i>MSRB3</i>	1.27e-07	0.001	Ear erectness (Dog)
Litter size	10	8070103	<i>MSRB3</i>	1.61e-08	0.001	Ear erectness (Dog)
Litter size	10	8085469	<i>MSRB3</i>	3.55e-08	0.001	Ear erectness (Dog)
Litter size	10	8095399	<i>MSRB3</i>	5.41e-08	0.001	Ear erectness (Dog)
Litter size	10	8100127	<i>MSRB3</i>	5.41e-08	0.001	Ear erectness (Dog)
Litter size	10	8134640	<i>MSRB3</i>	2.12e-08	0.001	Ear erectness (Dog)
Stillbirth rate	25	42482266	<i>ENSCAFG00000010704</i>	1.03e-07	0.001	
Stillbirth rate	27	2539211	<i>UFM1</i>	2.01e-07	0.049	

is in the intergenic region between the *CD36* gene and a lincRNA (*ENSCAFG00000034312*). The protein product of *CD36* is the fourth major glycoprotein of the platelet surface and serves as a receptor for thrombospondin in platelets [69]. Other known functions include transport of long chain fatty acids [70]. However, we believe that this variant is more likely a replication of a previous association between the nearby *FGF4* retrotransposon and chondrodysplasia across breeds [71].

Genetic loci that significantly associate with litter size

To identify SNPs that are significantly associated with variation in litter size among breeds, we retrieved litter size data from 10 810 litters of 224 breeds registered in the Norwegian Kennel Club [37].

For these data, we were able to obtain average number of pups per litter values for 60 of the 162 dog breeds with overlapping genetic data ([Supplementary Table S1](#)). Among these 60 breeds, median litter size is 5.55 pups, with a maximum 8.9 in Rhodesian Ridgebacks and a minimum of 2.4 in Pomeranians ([Supplementary Fig. S3B](#)).

To identify SNPs that are significantly associated with the observed variation in litter size across domestic dog breeds, we conducted an across-breed GWAS analysis using 149 834 149, 834 SNPs and litter size data from 60 dog breeds ([Fig. 1B](#), [Supplementary Fig. S2B](#)). We identified three independently associated regions containing 13 SNPs ([Table 3](#), [Supplementary Table S4](#)) intersecting three genes, namely RNA Terminal Phosphate Cyclase-Like 1 (*RCL1*, uncorrected $P=2.6 \times 10^{-8}$, Perm $P=0.001$), Myb/SANT SNA Binding Domain Containing 1

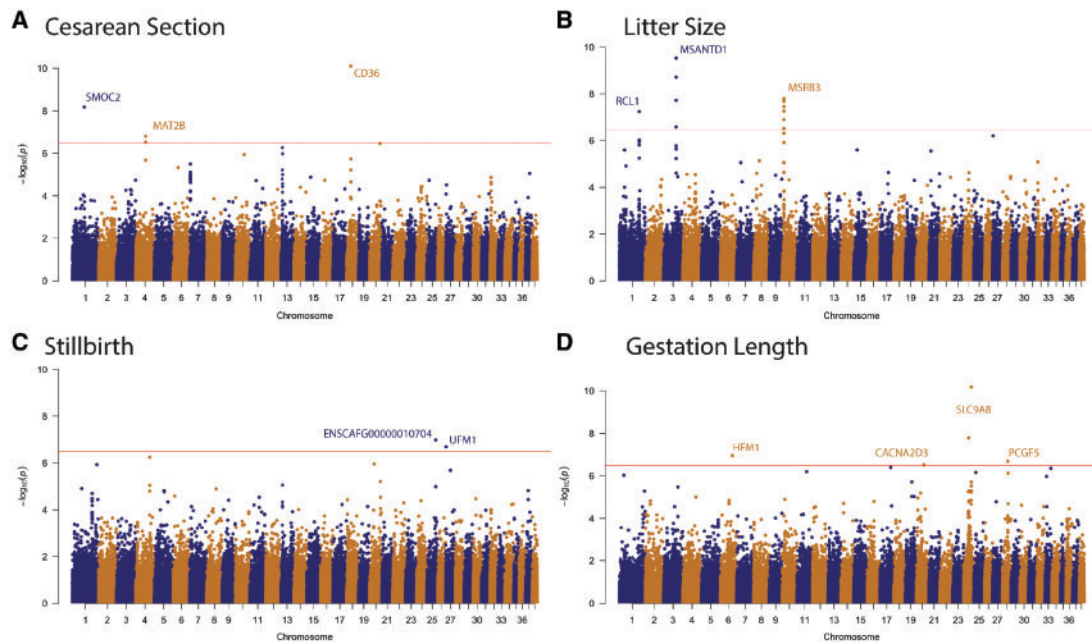


Figure 1. Genome-wide association results for reproductive traits in domestic dogs. Manhattan plots showing the statistical significance of each SNP as a function of genomic position for (A) cesarean section rate ($n=3,194$ individuals, $n=97$ breeds), (B) litter size ($n=2617$ individuals, $n=60$ breeds), (C) stillbirth ($n=2590$ individuals, $n=57$ breeds) and (D) gestation length ($n=1908$ individuals, $n=23$ breeds). Horizontal line indicates the significance threshold at $P=4.3 \times 10^{-7}$. Significant SNPs are labeled with the intersecting or nearest gene. Significant SNPs whose permutation P -values were above the 0.05 threshold. Plots were generated in R using the qqman package

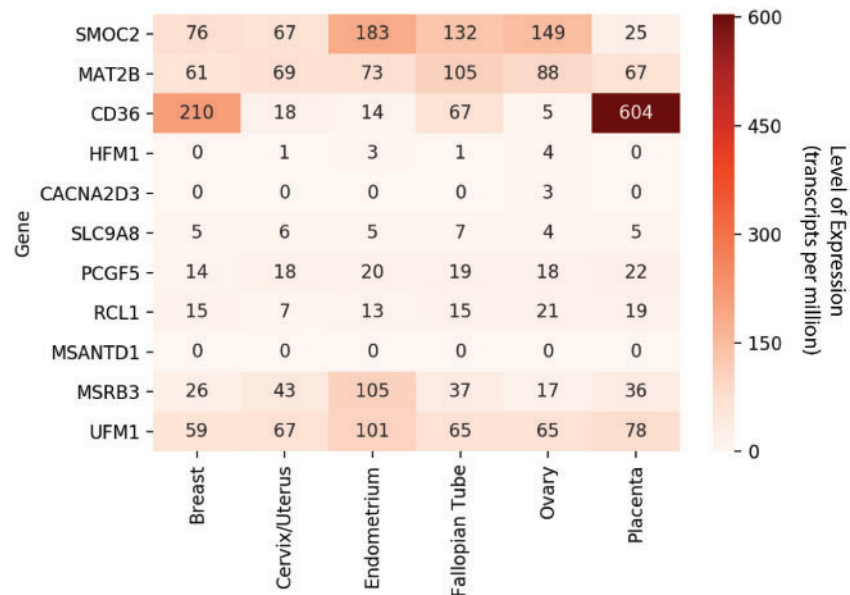


Figure 2. Gene expression in human female reproductive tissues of genes that contain or are adjacent to SNPs associated with reproductive traits in domestic dogs. Raw data were obtained from the Human Protein Atlas database [67]

(*MSANTD1*, see Table 3 for a full list of nearby linked SNPs and their association statistics) and methionine sulfoxide reductase B3 (*MSRB3*, see Table 3 for a full list of nearby linked SNPs and their association statistics).

The *RCL1* SNP is found in the intron between exons 7 and 8. *RCL1* functions in the maturation of 18s RNA [72] and is

associated with cervical cancer; one role of the gene in this cancer pathology is thought to involve the regulation of insulin receptors [72]. Additionally, a rare missense variation in *RCL1* was recently associated with depression [73]. The second SNP in the intergenic region downstream of *MSRB3*, whose protein product catalyzes the reduction of methionine-R-sulfoxides to methionine and

repairs oxidatively damaged proteins [74, 75]. In humans, *MSRB3* mutations are associated with deafness [76]. Epigenetic changes of *MSRB3* in the fetus during pregnancy may affect length of gestation, with increased DNA methylation correlated with increased gestational age [77, 78]. Furthermore, *MSRB3* shows an increase in mRNA expression in ripe (at term) versus unripe human uterine cervix, implying that *MSRB3* functions to ripen the cervix before the onset of labor [79]. In previous morphological studies in dogs, *MSRB3* is associated with ear erectness [33]. We additionally point out that this association is nearby the *HMGA2* gene that has been previously associated with body mass in several different organisms (domestic dog [32], pigs [80], mice [81, 82]). The last SNP that is significantly associated with litter size is located downstream of *MSANTD1*, which is part of a gene network believed to aid in cell-to-cell signaling and interaction, hematological system development and function and immune cell trafficking [83]. *MSANTD1* has been identified in two independent studies as a candidate gene for the determination of black coat color in goats [84, 85].

Genetic loci that significantly associate with stillbirth rate

To examine whether there are SNPs that are significantly associated with variation in stillbirth rate among breeds, we retrieved data for stillbirth rates for 57 of the 162 dog breeds (Supplementary Table S1). The data covers 10 810 litters of 224 breeds registered in the Norwegian Kennel Club and defines perinatal mortality as the sum of stillborn puppies and puppies that died during the first week after birth [38]. Among these 57 breeds with overlapping genomic data, the median stillbirth rate is 4.2 pups, with a maximum rate of 12.3% in Saint Bernards and a minimum of 0% in Basenjis and Italian Greyhounds (Supplementary Fig. S3C).

To test if any SNPs are significantly associated with the observed variation in stillbirth rate across domestic dog breeds, we conducted an across-breed GWAS analysis using 149 834 SNPs and stillbirth rate data from 56 dog breeds (Fig. 1C, Supplementary Fig. S2C). We identified two significantly associated SNPs (Table 3, Supplementary Table S5); both intersect genes, Ubiquitin fold modifier 1 (*UFM1*, uncorrected $P=2.01 \times 10^{-7}$, Perm $P=0.049$) and near a novel gene (ENSCAFG0000010704, uncorrected $P=1.03 \times 10^{-7}$, Perm $P=0.001$). The *UFM1* SNP has been previously associated with coat color variation in the domestic dog [86], but not with reproductive life history trait variation.

Genetic loci that significantly associate with gestation length

To examine whether there is variation in gestation length among breeds, we identified individual gestation length averages by breed predominantly in breeder handbooks. Utilizing breeders'

handbooks, we were able to identify gestation length means for a total of 23 of the 162 dog breeds that we had genotypic data for (Supplementary Table S1). Among these 23 breeds, the median gestation length is 62.2 days, with a maximum length of 65.3 in beagles and a minimum of 60.1 in the Alaskan Malamute (Supplementary Fig. S3D).

To identify SNPs significantly associated with the observed variation in gestation length across domestic dog breeds, we conducted an across-breed GWAS analysis using 149 834 SNPs and gestation length data from 23 dog breeds (Fig. 1D, Supplementary Fig. S2D). Our analysis identified five significantly associated SNPs (Table 3, Supplementary Table S6) that mapped to four genes, namely solute carrier family 9 (*SLC9A8*, uncorrected $P=3.7 \times 10^{-11}$, Perm $P=0.001$), calcium channel, voltage-dependent, alpha-2/delta Subunit 3 (*CACNA2D3*, uncorrected $P=3.1 \times 10^{-7}$, Perm $P=0.013$) and polycomb group ring finger 5 (*PCGF5*, $P=6.75 \times 10^{-11}$, Perm $P=0.003$).

The first SNP resides in intron 78 of *SLC9A8*, an integral transmembrane protein that exchanges extracellular Na^+ for intracellular H^+ . *SLC9A8* serves multiple functions [87] and is expressed ubiquitously (Fig. 2) [67]. Knockout male mice have impaired luteinizing hormone-stimulated cAMP production and are infertile, despite normal morphology of their reproductive system and normal behavior [88]. The second SNP is found in the intron between exons 26 and 27 of *CACNA2D3*. This gene is involved in regulating the influx of Ca^{2+} ions entering the cell upon membrane polarization [89], a critical process relevant to many functions, including fertilization and development [90]. In previous studies in humans, *CACNA2D3* is differentially methylated in the amnion between normal and pre-eclamptic pregnancies [91] and in blood between extreme preterm and term infants at birth [92, 93]. Additionally, *CACNA2D3* is one of four genes recently described as influencing cranial morphology in human populations [94]. In other domesticated animals, *CACNA2D3* is downregulated by Colony Stimulating Factor 2 (*CSF2*) in the trophoderm of pregnant cattle, which increases the ability of the preimplantation embryo to advance to the blastocyst stage [95]. In the closely related wolf, *CACNA2D3* is under diversifying selection associated with environmental adaptations to altitude [96–98]. Finally, the last SNP mapped to *PCGF5*, a gene previously shown to be necessary for neural differentiation in embryonic growth cell in humans [99].

DISCUSSION

Mammals exhibit a great deal of variation in their reproductive traits, yet remarkably little is known about the genetic basis of these traits. To begin to address this, we used GWAS analyses to examine the genetic basis of four reproductive traits (cesarean section rate, stillbirth rate, litter size and gestation length) across up to 97 domestic dog breeds. We identified several significant genetic associations for each trait (Fig. 1, Table 3).

Four of the 12 genetic regions that we found to be associated with reproductive trait variation have been previously identified to be involved in diverse traits associated with dog domestication (Table 3), such as brachycephaly and coat curl and color, suggesting that selection for signature traits of dog breeds may have also directly or indirectly influenced variation in reproductive traits. For example, one of the variants that we found to be associated with cesarean section rate is in an intron of *SMOC2*, a gene previously associated with brachycephaly in dogs [35, 66]. Brachycephaly, the shortening and widening of the muzzle and skull, is present in several 'fighting' breeds such as Boxer, Boston Terrier and Bulldog, and is thought to have been originally artificially selected on the basis that a shorter and wider cranial shape would enhance the dog's biting power [100]. Interestingly, one of the traits that associated with brachycephaly is cephalopelvic disproportion [31], a significant medical condition that can result in the death of both the litter and the bitch due to the inability of the pups to pass through the pelvic canal. The negative effects of cephalopelvic disproportion are alleviated by cesarean section, which not only allows these breeds to reproduce but also enables the continued application of artificial selection for the most extreme cranial morphology [66]. Whether the *SMOC2* variant identified directly influences parturition and birth timing in dogs (in humans, *SMOC2* is highly expressed in several reproductive tissues; see Fig. 2 and Ref. [67] or indirectly leads to adverse pregnancy outcomes (e.g. brachycephalic cranial morphology leading to cesarean section) remains unknown. It is highly likely, however, that the association between *SMOC2* and brachycephaly came first, paving the way for the subsequent association of both with cesarean section rate.

Several of the significantly associated genes that we identified in dogs appear to also be associated with reproductive phenotypes in humans. This suggests the possibility that the artificial selection that gave rise to dog breeds may have also contributed to the observed variation in their reproductive traits. For example, a member of the gene family for a subunit of the voltage-dependent calcium channel complex, *CACNA2D3*, which is associated with gestation length in our study, has been shown to be both differentially methylated in amnion between normal and pre-eclamptic human pregnancies [91], and in blood between extreme preterm and term infants at birth [92, 93]. Similarly, expression of *MSRB3*, which is associated with litter size in our study, is elevated in ripe (at term) versus unripe human uterine cervix and may be involved in the onset of labor [78], and *SMOC2* is also known to be expressed in human reproductive tissues (Fig. 2) [67].

There are several caveats to our analyses and results. To begin with, as is true of all GWAS studies, our findings will need to be replicated in other cohorts and our hypotheses about the variants and genes involved will need to be functionally validated. Furthermore, our set of reproductive phenotype data is breed averages rather than individual values, which would allow more precise association analyses. Additionally, it should be noted that some of the values of our

reproductive phenotypic data may not be always precise. For example, there is likely variability across breeds for total litter size and ultimately 1 week viability (stillbirth). Similarly, the cesarean section phenotype is largely determined by the owner and the veterinarian. In this way, it is an elective phenotype measurement but is informed by the breed-specific likelihood of a healthy delivery. However, cesarean section rates for some breeds are extremely high (e.g. 92.3% in Boston terrier), which seems to indicate that cesarean section procedures are necessary to preserve the health of both the mother and the pups. Therefore, we assume that these rates reflect veterinarians' information and knowledge of potential breed-specific birthing risks.

Previous analyses of traits, such as body mass, in the domestic dog have found that much of the variation can be explained by a few large-effect loci [101, 102]. While our results show that our approach and use of breed-specific average trait values are sufficiently powered to detect traits determined by a few causal variants of large effects, we still lack knowledge of the underlying genetic architecture of the four reproductive traits analysed in this study. The results of a recent GWAS of gestation length in humans [24] suggest that, at least in humans, variation in gestation length is likely explained by many small-effect variants; if that is the case, approaches based on breed averages may fail to detect variants of low effect due to the lack of sufficient phenotypic stratification. Finally, as is true for most complex traits in the domestic dog, the reproductive traits studied here are likely to have also been influenced by the rapid and strong artificial selection imposed during the recent breeding history of the domestic dog.

SUPPLEMENTARY DATA

Supplementary data is available at *EMPH* online.

ACKNOWLEDGEMENTS

We thank members of the Rokas lab and members of the March of Dimes Prematurity Research Center Ohio Collaborative for useful discussions of this work.

FUNDING

This research was supported by the March of Dimes through the March of Dimes Prematurity Research Center Ohio Collaborative to A.R., P.A. and J.A.C, by the Burroughs Wellcome Fund to J.A.C and A.R. and by the Guggenheim Foundation to A.R. M.L.J. received partial support through a Littlejohn Summer Research Scholarship at Vanderbilt University. The funders had no role in study design, data collection and analysis, decision to publish or preparation of the manuscript.

Conflict of interest: None declared.

REFERENCES

- Derrickson EM. Comparative reproductive strategies of altricial and precocial eutherian mammals. *Funct Ecol* 1992;**6**:57–65.
- Harrison RM. *Reproduction in Mammals: General Overview*. Chichester, UK: John Wiley & Sons, Ltd, 2001.
- Behringer RR, Eakin GS, Renfree MB. Mammalian diversity: gametes, embryos and reproduction. *Reprod Fertil Dev* 2006;**18**:99–107.
- Jones KE, Bielby J, Cardillo M *et al.* PanTHERIA: a species-level database of life history, ecology, and geography of extant and recently extinct mammals. *Ecology* 2009;**90**:2648.
- Kiltie RA. Intraspecific variation in the mammalian gestation period. *J Mammal* 1982;**63**:646–52.
- Martin RD, MacLarnon AM. Gestation period, neonatal size and maternal investment in placental mammals. *Nature* 1985;**313**:220–3.
- Cross JC, Werb Z, Fisher SJ. Implantation and the placenta: key pieces of the development puzzle. *Science* 1994;**266**:1508–18.
- Enders AC, Carter AM. What can comparative studies of placental structure tell us? A review. *Placenta* 2004;**25**:S3–9.
- Elliot MG, Crespi BJ. Phylogenetic evidence for early hemochorial placentation in *Eutheria*. *Placenta* 2009;**30**:949–67.
- Pond CM. The significance of lactation in the evolution of mammals. *Evolution* 1977;**31**:177–99.
- Lefèvre CM, Sharp JA, Nicholas KR. Evolution of lactation: ancient origin and extreme adaptations of the lactation System. *Ann Rev Genomics Hum Genet* 2010;**11**:219–38.
- Cross JF, Martin RD. Calculation of gestation period and other reproductive parameters for primates. *Dodo: J Jersey Wildl Preserv Trust. Durrell Wildlife Conserv Trust* 1981;**18**:30–43.
- Hudson R, Cruz Y, Lucio A *et al.* Temporal and behavioral patterning of parturition in rabbits and rats. *Physiol Behav* 1999;**66**:599–604.
- Burris MJ, Blunn CT. Some factors affecting gestation length and birth weight of beef cattle. *J Anim Sci* 1952;**11**:34–41.
- Davies Morel MCG, Newcombe JR, Holland SJ. Factors affecting gestation length in the thoroughbred mare. *Anim Reprod Sci* 2002;**74**:175–85.
- Dufour DL, Sauther ML. Comparative and evolutionary dimensions of the energetics of human pregnancy and lactation. *Am J Hum Biol* 2002;**14**:584–602.
- Carneiro M, Piorno V, Rubin C-J *et al.* Candidate genes underlying heritable differences in reproductive seasonality between wild and domestic rabbits. *Anim Genet* 2015;**46**:418–25.
- Sironen A, Uimari P, Nagy S *et al.* Knobbed acrosome defect is associated with a region containing the genes *STK17b* and *HECW2* on porcine chromosome 15. *BMC Genomics* 2010;**11**:699.
- Bolormaa S, Pryce JE, Hayes BJ *et al.* Multivariate analysis of a genome-wide association study in dairy cattle. *J Dairy Sci* 2010;**93**:3818–33.
- Maltecca C, Gray KA, Weigel KA *et al.* A genome-wide association study of direct gestation length in US Holstein and Italian Brown populations. *Anim Genet* 2011;**42**:585–91.
- Walker R, Gurven M, Hill K *et al.* Growth rates and life histories in twenty-two small-scale societies. *Am J Hum Biol* 2006;**18**:295–311.
- Horikoshi M, Beaumont RN, Day FR *et al.* Genome-wide associations for birth weight and correlations with adult disease. *Nature* 2016;**538**:248–52.
- Plunkett J, Doniger S, Orabona G *et al.* An evolutionary genomic approach to identify genes involved in human birth timing. *PLoS Genet* 2011;**7**:e1001365.
- Zhang G, Feenstra B, Bacelis J *et al.* Genetic associations with gestational duration and spontaneous preterm birth. *N Engl J Med* 2017;**377**:1156–67.
- Cassady JP, Johnson RK, Pomp D *et al.* Identification of quantitative trait loci affecting reproduction in pigs. *J Anim Sci* 2001;**79**:623–33.
- Langlois B, Blouin C, Chaffaux S. Analysis of several factors of variation of gestation loss in breeding mares. *Animal* 2012;**6**:1925–30.
- Johanson JM, Berger PJ, Tsuruta S *et al.* A Bayesian threshold-linear model evaluation of perinatal mortality, dystocia, birth weight, and gestation length in a Holstein herd. *J Dairy Sci* 2011;**94**:450–60.
- Neff MW, Rine J. A fetching model organism. *Cell* 2006;**124**:229–31.
- Beale H, Ostrander EA. Sizing up dogs. *Curr Biol* 2012;**22**:R315–6.
- Marsden CD, Ortega-Del Vecchyo D, O'Brien DP *et al.* Bottlenecks and selective sweeps during domestication have increased deleterious genetic variation in dogs. *Proc Natl Acad Sci USA* 2016;**113**:152–7.
- Evans KM, Adams VJ. Proportion of litters of purebred dogs born by caesarean section. *J Small Anim Pract* 2010;**51**:113–8.
- Hayward JJ, Castelhan MG, Oliveira KC *et al.* Complex disease and phenotype mapping in the domestic dog. *Nat Commun* 2016;**7**:10460.
- Boyko AR, Quignon P, Li L. *et al.* A simple genetic architecture underlies morphological variation in dogs. *PLoS Biol* 2010;**8**:e1000451.
- Vaysse A, Ratnakumar A, Derrien T *et al.* Identification of genomic regions associated with phenotypic variation between dog breeds using selection mapping. *PLoS Genet* 2011;**7**:e1002316. Public Library of Science;
- Marchant TW, Johnson EJ, McTeir L. *et al.* Canine brachycephaly is associated with a retrotransposon-mediated missplicing of *SMOC2*. *Curr Biol* 2017;**27**:1573–6.
- Karlsson EK, Lindblad-Toh K. Leader of the pack: gene mapping in dogs and other model organisms. *Nat Rev Genet* 2008;**9**:713–25.
- Borge KS, Tønnessen R, Nødtvedt A *et al.* Litter size at birth in purebred dogs—a retrospective study of 224 breeds. *Theriogenology* 2011;**75**:911–9.
- Tønnessen R, Borge KS, Nødtvedt A *et al.* Canine perinatal mortality: a cohort study of 224 breeds. *Theriogenology* 2012;**77**:1788–801.
- Chatdarong K, Tummaruk P, Sirivaidyapong S *et al.* Seasonal and breed effects on reproductive parameters in bitches in the tropics: a retrospective study. *J Small Anim Pract* 2007;**48**:444–8. Blackwell Publishing Ltd;
- Concannon P, Whaley S, Lein D *et al.* Canine gestation length: variation related to time of mating and fertile life of sperm. *Am J Vet Res* 1983;**44**:1819–21.
- Eilts BE, Davidson AP, Hosgood G *et al.* Factors affecting gestation duration in the bitch. *Theriogenology* 2005;**64**:242–51.
- Evans JM, White K. *The Book of the Bitch*. High Wycombe, UK: Interpet, 1997, 1.
- Gavrilovic BB, Andersson K, Linde Forsberg C. Reproductive patterns in the domestic dog—A retrospective study of the Drever breed. *Theriogenology* 2008;**70**:783–94.
- Okkens AC, Hekerman TWM, de Vogel JWA *et al.* Influence of litter size and breed on variation in length of gestation in the dog. *Vet Q* 1993;**15**:160–1.
- Son CH, Jeong KA, Kim JH *et al.* Establishment of the prediction table of parturition day with ultrasonography in small pet dogs. *J Vet Med Sci* 2001;**63**:715–21.

46. Kim BS, Son CH. Time of initial detection of fetal and extra-fetal structures by ultrasonographic examination in Miniature Schnauzer bitches. *J Vet Sci* 2007;**8**:289–93.
47. Linde Forsberg C, Wikström C, Lundeheim N. Differences between seasons of the year and breeds in mating frequency, gestation length and litter size in 13 breeds of dogs. *6th Intl. Symp. on Canine and Feline Reproduction*. 2008.
48. Zhou X, Stephens M. Genome-wide efficient mixed-model analysis for association studies. *Nat Genet* 2012;**44**:821–4.
49. Gould SJ. Allometry and size in ontogeny and phylogeny. *Biol Rev Camb Philos Soc* 1966;**41**:587–640.
50. Kleiber M. Body size and metabolism. *Hilgardia* 1932;**6**:315–53.
51. Shingleton A. Allometry: the study of biological scaling. *Nat Educ Knowl* 2010;**3**:2.
52. Sacher GA, Staffeldt EF. Relation of gestation time to brain weight for placental mammals: implications for the theory of vertebrate growth. *Am Nat* 1974;**108**:593–615.
53. Blueweiss L, Fox H, Kudzma V et al. Relationships between body size and some life history parameters. *Oecologia* 1978;**37**:257–72.
54. Tuomi J. Mammalian reproductive strategies: a generalized relation of litter size to body size. *Oecologia* 1980;**45**:39–44.
55. Ross C. The intrinsic rate of natural increase and reproductive effort in primates. *J Zool* 1988;**214**:199–219.
56. Kihlström JE. Period of gestation and body-weight in some placental mammals. *Comp Biochem Physiol* 1972;**43**:673–9.
57. Phillips JB, Abbot P, Rokas A. Is preterm birth a human-specific syndrome? *Evol Med Public Health* 2015;**2015**:136–48.
58. Lange K. *Mathematical and Statistical Methods for Genetic Analysis*. New York, NY: Springer Science & Business Media, 2002, 1.
59. Lindblad-Toh K, Wade CM, Mikkelsen TS et al. Genome sequence, comparative analysis and haplotype structure of the domestic dog. *Nature* 2005;**438**:803–19.
60. Hoepfner MP, Lundquist A, Pirun M et al. An improved canine genome and a comprehensive catalogue of coding genes and non-coding transcripts. *PLoS One* 2014;**9**:e91172.
61. Durinck S, Spellman PT, Birney E et al. Mapping identifiers for the integration of genomic datasets with the R/Bioconductor package biomaRt. *Nat Protoc* 2009;**4**:1184–91.
62. R Core Team. R: A Language and Environment for Statistical Computing. Vienna, Austria, 2013.
63. Turner SD. qqman: an R package for visualizing GWAS results using Q-Q and Manhattan plots. *bioRxiv* . 2014;005165.
64. Aulchenko YS, Ripke S, Isaacs A et al. GenABEL: an R library for genome-wide association analysis. *Bioinformatics* 2007;**23**:1294–6.
65. Rao TJ, Province MA. A framework for interpreting type I error rates from a product–term model of interaction applied to quantitative traits. *Genet Epidemiol* 2016;**40**:144–53.
66. Bannasch D, Young A, Myers J et al. Localization of canine brachycephaly using an across breed mapping approach. *PLoS One* 2010;**5**:e9632.
67. Uhlén M, Fagerberg L, Hallström BM et al. Tissue-based map of the human proteome. *Science* 2015;**347**:1260419.
68. Zhao C, Chen X, Wu W et al. MAT2B promotes adipogenesis by modulating SAME levels and activating AKT/ERK pathway during porcine intramuscular preadipocyte differentiation. *Exp Cell Res* 2016;**344**:11–21.
69. Tandon NN, Kralisz U, Jamieson GA. Identification of glycoprotein IV (CD36) as a primary receptor for platelet-collagen adhesion. *J Biol Chem* 1989;**264**:7576–83.
70. Febbraio M, Hajjar DP, Silverstein RL. CD36: a class B scavenger receptor involved in angiogenesis, atherosclerosis, inflammation, and lipid metabolism. *J Clin Invest* 2001;**108**:785–91.
71. Parker HG, VonHoldt BM, Quignon P et al. An expressed fgf4 retrogene is associated with breed-defining chondrodysplasia in domestic dogs. *Science* 2009;**325**:995–8.
72. Lyng H, Brøvig RS, Svendsrud DH et al. Gene expressions and copy numbers associated with metastatic phenotypes of uterine cervical cancer. *BMC Genomics* 2006;**7**:268.
73. Amin N, de Vrij FMS, Baghdadi M et al. A rare missense variant in RCL1 segregates with depression in extended families. *Mol Psychiatry* 2018;**23**:1120–26.
74. Brot N, Weissbach L, Werth J et al. Enzymatic reduction of protein-bound methionine sulfoxide. *Proc Natl Acad Sci USA* 1981;**78**:2155–8.
75. Kim H-Y, Gladyshev VN. Characterization of mouse endoplasmic reticulum methionine-R-sulfoxide reductase. *Biochem Biophys Res Commun* 2004;**320**:1277–83.
76. Ahmed ZM, Yousaf R, Lee BC et al. Functional null mutations of MSRB3 encoding methionine sulfoxide reductase are associated with human deafness DFNB74. *Am J Hum Genet* 2011;**88**:19–29.
77. Lee H, Jaffe AE, Feinberg JI et al. DNA methylation shows genome-wide association of NFIX, RAPGEF2 and MSRB3 with gestational age at birth. *Int J Epidemiol* 2012;**41**:188–99.
78. Schierding W, O’Sullivan JM, Derraik JGB et al. Genes and post-term birth: late for delivery. *BMC Res Notes* 2014;**7**:720.
79. Hassan SS, Romero R, Tarca AL et al. The transcriptome of cervical ripening in human pregnancy before the onset of labor at term: identification of novel molecular functions involved in this process. *J Matern Fetal Neonatal Med* 2009;**22**:1183–93.
80. Chung J, Zhang X, Collins B et al. High mobility group A2 (HMGA2) deficiency in pigs leads to dwarfism, abnormal fetal resource allocation, and cryptorchidism. *Proc Natl Acad Sci USA* 2018;**115**:5420–5.
81. Zhou X, Benson KF, Ashar HR et al. Mutation responsible for the mouse pygmy phenotype in the developmentally regulated factor HMGI-C. *Nature* 1995;**376**:771–4.
82. Anand A, Chada K. In vivo modulation of Hmgic reduces obesity. *Nat Genet* 2000;**24**:377.
83. Yu F, Shen X-Y, Fan L et al. Genome-wide analysis of genetic variations assisted by ingenuity pathway analysis to comprehensively investigate potential genetic targets associated with the progression of hepatocellular carcinoma. *Eur Rev Med Pharmacol Sci* 2014;**18**:2102–8.
84. Benjelloun B, Alberto FJ, Streeter I et al. Characterizing neutral genomic diversity and selection signatures in indigenous populations of Moroccan goats (*Capra hircus*) using WGS data. *Front Genet* 2015;**6**:107.
85. Wang X, Liu J, Zhou G et al. Whole-genome sequencing of eight goat populations for the detection of selection signatures underlying production and adaptive traits. *Sci Rep* 2016;**6**:38932.
86. Cadieu E, Neff MW, Quignon P et al. Coat variation in the domestic dog is governed by variants in three genes. *Science* 2009;**326**:150–3.

87. Xu H, Chen H, Dong J *et al.* Gastrointestinal distribution and kinetic characterization of the sodium-hydrogen exchanger isoform 8 (NHE8). *Cell Physiol Biochem* 2008;**21**:109–16.
88. Xu H, Chen H, Li J *et al.* Disruption of NHE8 expression impairs Leydig cell function in the testes. *Am J Physiol Cell Physiol* 2015;**308**:C330–8.
89. Jin Y, Cui D, Ren J *et al.* CACNA2D3 is downregulated in gliomas and functions as a tumor suppressor. *Mol Carcinog* 2017;**56**:945–59.
90. Berridge MJ, Bootman MD, Lipp P. Calcium—a life and death signal. *Nature* 1998;**395**:645–8.
91. Suzuki M, Maekawa R, Patterson NE *et al.* Amnion as a surrogate tissue reporter of the effects of maternal preeclampsia on the fetus. *Clin Epigenetics* 2016;**8**:67.
92. Cruickshank MN, Oshlack A, Theda C *et al.* Analysis of epigenetic changes in survivors of preterm birth reveals the effect of gestational age and evidence for a long term legacy. *Genome Med* 2013;**5**:96.
93. Eidem HR, Ackerman WE, McGary KL *et al.* Gestational tissue transcriptomics in term and preterm human pregnancies: a systematic review and meta-analysis. *BMC Med Genomics* 2015;**8**:27.
94. Paternoster L, Zhurov AI, Toma AM *et al.* Genome-wide association study of three-dimensional facial morphology identifies a variant in PAX3 associated with nasion position. *Am J Hum Genet* 2012;**90**:478–85.
95. Ozawa M, Sakatani M, Dobbs KB *et al.* Regulation of gene expression in the bovine blastocyst by colony stimulating factor 2. *BMC Res Notes* 2016;**9**:250.
96. Pilot M, Greco C, vonHoldt BM, *et al.* Genome-wide signatures of population bottlenecks and diversifying selection in European wolves. *Heredity* 2014;**112**:428–42.
97. Schweizer RM, vonHoldt BM, Harrigan R *et al.* Genetic subdivision and candidate genes under selection in North American grey wolves. *Mol Ecol* 2016;**25**:380–402.
98. Zhang W, Fan Z, Han E *et al.* Hypoxia adaptations in the grey wolf (*Canis lupus chanco*) from Qinghai-Tibet Plateau. *PLoS Genet* 2014;**10**:e1004466.
99. Yao M, Zhou X, Zhou J *et al.* PCGF5 is required for neural differentiation of embryonic stem cells. *Nat Commun* 2018;**9**:1463.
100. Ellis JL, Thomason J, Kebreab E *et al.* Cranial dimensions and forces of biting in the domestic dog. *J Anat* 2009;**214**:362–73.
101. Hoopes BC, Rimbault M, Liebers D *et al.* The insulin-like growth factor 1 receptor (IGF1R) contributes to reduced size in dogs. *Mamm Genome* 2012;**23**:780–90.
102. Webster MT, Kamgari N, Perloski M *et al.* Linked genetic variants on chromosome 10 control ear morphology and body mass among dog breeds. *BMC Genomics* 2015;**16**:474.