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,*]

[1]Department of Biological Sciences, Vanderbilt University, Nashville, TN, USA

[2]Center for Computational Molecular Biology, Brown University, Providence, Rhode Island, 02912, USA

[3]Department of Ecology and Evolutionary Biology, Brown University, Providence, RI 02912, USA

[4]Department of Biological Sciences, Cumberland University, Lebanon, TN, USA

[5]Department of Biomedical Informatics, Vanderbilt University, Nashville, TN 37203, USA

[6]Vanderbilt Genetics Institute, Vanderbilt University Medical Center, Nashville, TN 37232, USA

* These authors contributed equally

Short title: Probing the genetic basis of reproductive traits across dog breeds

Keywords: life history; artificial selection; trade-off; preterm birth; prematurity; pregnancy

***Corresponding Author:** antonis.rokas@vanderbilt.edu

Abstract

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. 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 co-variates. We identified 13 genome-wide significant associations between these traits and genetic loci, including variants near *CACNA2D3* with gestation length, *MSRB3* and *KRT71* with litter size, *SMOC2* with cesarean section rate, and *HTR2C* with stillbirth rate. Some of these loci, such as *CACNA2D3* and *MSRB3*, have been previously implicated in human reproductive pathologies, whereas others have been previously associated with domestication-related traits, including brachycephaly (*SMOC2*), coat curl (*KRT71*), and tameness (*HTR2C*). These results raise the hypothesis that the artificial selection that gave rise to dog breeds also shaped 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

Variation in reproductive traits across mammals has been extensively investigated and is well established, but the genetic contributors to this variation remain largely unknown. Here, we take advantage of the domestic dog, a powerful model for mammalian genetics, to gain insight into the genetic basis of reproductive traits. By examining the association between more than a hundred thousand genetic variants and four reproductive traits that vary extensively across dog breeds, we identified more than a dozen significant associations for cesarean section rate, litter size, stillbirth rate, and gestation length. Some of the variants that we identify are nearby genes previously implicated in human reproductive pathologies, whereas several others have been previously associated with domestication-related traits. Our results establish the domestic dog as a tractable system for studying the genetics of reproductive traits and underscore the potential for cryptic interactions between reproductive and other traits favored over the course of adaptation.

Introduction

Mammals exhibit wide variation in traits associated with reproduction (Derrickson 1992, Harrison 2001, Behringer et al. 2006). 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* (Jones et al. 2009, Kiltie 1992, Martin et al. 1985); neonate size can range from less than one gram in the shrew family (Soricidae), to more than a metric ton in the baleen whales (Balaenopteridae) (Jones et al. 2009, Martin et al. 1985); and neonates can be either precocial (e.g., cricetid rodents, rabbits, and canids) or altricial (e.g., hystricomorph rodents, ungulates, and cetaceans) (Derrickson 1992). This variation in reproductive traits also extends to methods of implantation (Cross et al. 1994), structure of the placenta (Enders et al. 2004, Elliot et al. 2009), and lactation strategies (Pond 1977, Lefèvre et al. 2011). Not surprisingly, many reproductive traits also exhibit substantial intra-specific variation (Kiltie 1982). For example, many mammals exhibit intraspecific variation in gestation length, including primates (Cross et al. 1981), rat and rabbits (Hudson et al. 1999), as well as the domesticated cattle (Burris et al. 1952) and thoroughbred horses (Davies et al. 2002). 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 – 41% (Dufour et al. 2002).

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 (Carneiro et al. 2015, Sironen et al. 2010, Bolormaa et al. 2010, Maltecca

et al. 2011) and human populations (Walker et al. 2006). 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 (Carneiro et al. 2015), infertility in pigs (Sironen et al. 2010), and dairy traits in cattle (Bolormaa et al. 2010). 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 (Horikoshi et al. 2016) and gestational duration or length (Plunkett et al. 2011, Zhang et al. 2015, Zhang et al. 2017). For example, maternal variation in six genomic loci (*ADCY5*, *AGTR2*, *EBF1*, *EEFSEC*, *RAP2C*, and *WNT4*) is associated with gestational duration and preterm birth (Zhang et al. 2017). 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 (Cassady et al. 2001, Langlois et al. 2012, Johanson et al. 2011).

To address this challenge, we studied the genetics of reproductive traits in a powerful new model system: the domestic dog. 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 (Neff et al. 2006, Beale and Ostrander 2012, Marsden et al. 2016). 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 (Evans et al. 2010). 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 (Evans and Adams 2010). 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 (Hayward et al. 2016, Boyko et al. 2010, Vaysse et al. 2011, Marchant et al. 2017), as well as genetic disease (Karlsson et al. 2008).

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 13 genetic loci that are significantly associated with these reproductive traits (using log body mass as a co-variate). Several of these variants are in or near genes previously implicated in human reproduction-related pathologies. The majority of the variants that we discovered to be significantly associated with reproductive trait variation are also 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 directly or indirectly influenced variation in reproductive traits. More broadly, our results establish the domestic dog as a tractable system for studying the genetics of reproductive traits and underscore the potential for cryptic interactions between reproductive and other traits favored over the course of adaptation.

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 (Zhou and Stephens, 2012). Number of individuals and distribution of breed varied with analysis (Supplementary Table 1). After filtering for MAF (MAF < 0.05; 10,804 SNPs were excluded) and linkage disequilibrium (34,240 additional SNPs were excluded), 115,683 SNPs were included in the GWAS analysis for each reproductive trait. 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 (36,37). 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* ($P = 5.1 \times 10^{-11}$) (Supplementary Figure 1, Supplementary Table 2).

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 1). 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 (Marsden et al. 2016). 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 Figure 3A).

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 115,683 SNPs and cesarean section values across 95 dog breeds (Figure 1A, Supplementary Figure 2A). 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 three significant SNPs (Supplementary Table S3), two of which mapped to genes, namely paralemmin 3 (*PALM3*, uncorrected $P = 1.4 \times 10^{-9}$, $Perm\ P = 0.001$) and sparc-related modular calcium-binding protein 2 (*SMOC2*, uncorrected $P = 2.0 \times 10^{-7}$, $Perm\ P = 0.011$), and a third that mapped to the intergenic region between the *CD36* glycoprotein and a lincRNA (uncorrected $P = 9.7 \times 10^{-8}$, $Perm\ P = 0.024$) (Figure 1A). Our GWAS analysis also recovered keratin 71 (*KRT71*, uncorrected $P = 2.9 \times 10^{-7}$, $Perm\ P = 0.217$), but its permutation p-value was above 0.05.

The first significantly associated SNP (chromosome 1: 55,983,871) is found in the intron between exons 13 and 14 of *SMOC2*, a gene that is associated with brachycephaly in dogs (Marchant et al. 2017, Bannasch et al. 2010); variation in *SMOC2* accounts for 36% of facial length variation in dogs (Marchant et al. 2017). In humans, *SMOC2* is highly expressed in endometrium as well as other reproductive tissues, including the fallopian tubes, ovaries and cervix (Figure 2) (Uhlén et al. 2015). The 3' intronic location of the SNP raises the possibility that it might be regulatory (Rose 2008).

The second SNP is found in the 3' UTR of *PALM3*, which is a member of the paralemmin gene family that also includes *PALM1*, *PALM2*, and *PALMD* (palmdelphin); members of this family are implicated in plasma membrane dynamics and as modulators of cellular cAMP signaling in the brain (Basile et al. 2006, Kutzleb et al. 1998). The function of *PALM3* may be slightly different from the rest of the genes in the family, with recent work suggesting that PALM3 is a binding protein of the single immunoglobulin IL-1 receptor-related molecule (SIGIRR), which is a negative regulator of Toll-Interleukin-1 receptor signaling (Chen et al. 2011). In humans, *PALM3* is primarily expressed in the membranes of the stomach, kidney, parathyroid gland and epididymis (Figure 2) (Uhlén et al. 2015). The SNP (chromosome 20: 48,454,259) that is significantly associated with cesarean section rate is found in the first intron of the *PALM3* gene, suggesting that it might be involved in regulatory actions typically observed in 5' introns (Rose 2008).

The third significant SNP (chromosome 18: 20,272,961) is found 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 (Tandon et al. 1989). Other known functions include transport of long chain fatty acids (Febbraio et al. 2001). However, we believe that this variant is more likely a replication of a previous association between the nearby *FGF4* retrotransposon and chondrodysplasia across breeds (Parker et al. 2009).

Genetic loci that significantly associate with litter size

To examine whether there are 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 (Borge et al. 2012). 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 1). 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 Figure 3B).

To identify SNPs, and genes proximal to them, that are significantly associated with the observed variation in litter size across domestic dog breeds, we conducted an across-breed GWAS analysis using 115,683 SNPs and litter size data from 60 dog breeds (Figure 1B, Supplementary Figure 2B). We identified two significant and one marginally significant SNPs (Supplementary Table S4) intersecting three genes, namely keratin 71 (*KRT71,* uncorrected $P = 2.2 \times 10^{-8}$, $Perm\ P = 0.017$), RNA Terminal Phosphate Cyclase-Like 1 (*RCL1*, uncorrected $P = 2.6 \times 10^{-8}$, $Perm\ P = 0.001$) and microphthalmia-associated transcription factor (*MITF*, uncorrected $P = 3.5 \times 10^{-7}$, $Perm\ P = 0.079$). The *KRT71*

SNP is the same variant that was marginally associated with variation in cesarean section rate described above, but in this instance its association with the trait remains significant after permutation analysis. Another three significant SNPs were found in intergenic regions; two were nearby genes *MSRB3* (methionine sulfoxide reductase B3, uncorrected $P = 1.3 \times 10^{-7}, Perm\ P = 0.001$) and *MSANTD1* (Myb/SANT DNA binding domain containing, uncorrected $P = 1.5 \times 10^{-9}, Perm\ P = 0.001$), respectively. The final variant was near an RNA of unknown function (ENSCAFG00000021196, uncorrected $P = 3.8 \times 10^{-10}, Perm\ P = 0.001$).

The *RCL1* SNP (chromosome 1: 93,189,363) is found in the intron between exons 7 and 8. *RCL1* functions in the maturation of 18s RNA (Lyng et al. 2006) and is associated with cervical cancer; one role of the gene in this cancer pathology is thought to involve the regulation of insulin receptors (Lyng et al. 2006). Additionally, a rare missense variation in *RCL1* was recently associated with depression (Amin et al. 2017).

The *KRT71* SNP (chromosome 27: 2,539,211) results in a missense mutation of exon 2 of the gene, which belongs to a family of keratin genes specifically expressed in the inner root sheath of hair follicles (Langbein et al. 2003). Prior analysis in dogs identified variation in gene *KRT71*, along with variation in genes *RSPO2* and *FGF5*, accounting for most coat phenotypes (Cadieu et al. 2009), such as curliness.

Another SNP (chromosome 10: 8,114,328) significantly associated with litter size is found in the intergenic region downstream of *MSRB3*, whose protein product catalyzes the

reduction of methionine-R-sulfoxides to methionine and repairs oxidatively damaged proteins (Brot et al. 1981, Kim et al. 2004). In humans, mutations in *MSRB3* are associated with deafness (Ahmed et al. 2011). Epigenetic changes of *MSRB3* in the fetus during pregnancy may affect length of gestation, with increased DNA methylation correlated with increased gestational age (Lee et al. 2012, Schierding et al. 2014). 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 (Hassan et al. 2009). In previous morphological studies in dogs, *MSRB3* is associated with ear erectness (Boyko et al. 2010).

The last SNP (chromosome 6: 61,062,626) 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 (Yu et al. 2014). *MSANTD1* has been identified in two independent studies as a candidate gene for the determination of black coat color in goats (Benjelloun et al. 2015, Wang et al. 2016).

Finally, following permutation analysis, the association between the *MITF* SNP (*Perm P* = 0.079) and litter size was just above the threshold for significance and can only be considered as suggestive. The *MITF* SNP (chromosome 20: 21,848,176) is found in the intron between exons 4 and 5. *MITF* plays an integral role in the development of neural crest-derived melanocytes and optic cup-derived retinal pigment epithelial cells. In human melanocytes, *MITF* is a regulator of *DIAPH1*, a member of the formin gene family whose

members are highly expressed in reproductive tissues and have been associated with a variety of reproductive phenotypes (Carreira et al. 2006, Lamm et al. 2018, Cruickshank et al. 2013, Elovitz et al. 2014, Montenegro et al. 2009). *DIAPH1* expression is increased in spontaneous term and preterm labor myometrial tissues (Lartey et al. 2007). In domesticated animals, *MITF* is a well characterized gene associated with coat color (Boyko et al. 2010, Wang et al. 2016). In humans, *MITF* is expressed in melanocytes, as well as reproductive tissues including the endometrium and cervix (Figure 2) (Uhlén et al. 2015).

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 1). 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 (Tønnessen et al. 2012). 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 Figure 3C).

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 115,683 SNPs and stillbirth rate data from 56 dog breeds (Figure 1C, Supplementary Figure 2C). We identified five significant and marginally significant SNPs (Supplementary Table S5); five intersecting 4 genes, 5-Hydroxtryptamine receptor 2C (*HTR2C*, uncorrected $P = 2.0 \times$

10^{-7}, $Perm\ P = 0.001$), keratin 71 (*KRT71*, uncorrected $P = 3.2 \times 10^{-9}$, $Perm\ P = 0.064$), and microphthalmia-associated transcription factor (*MITF*, uncorrected $P = 1.4 \times 10^{-7}$, $Perm\ P = 0.079$), and SP140 nuclear body protein (*SP140,* uncorrected $P = 2.76 \times 10^{-8}$, $Perm\ P = 0.001$) and one in an intergenic region near a snoRNA (ENSCAFG00000027305, uncorrected $P = 1.3 \times 10^{-7}$, $Perm\ P = 0.002$) of unknown function. The *KRT71* SNP associated with variation in stillbirth rate is the same one as that associated with litter size described above. Similarly, the *MITF* SNP associated with variation in stillbirth rate is the same as that associated with litter size. However, both of these associations have permutation p-values > 0.05 and thus can only be considered as tentative.

The *SP140* SNP (chromosome 25: 42,482,266) resides in the intro between exons 4 and 5. *SP140* is the lymphoid-restricted homolog of *SP100* expressed in mature B cells, as well as some T cells (Bloch et al. 1996). High levels of *SP140* mRNA are detected in human spleen and peripheral blood leukocytes, but not other human tissues (Bloch et al., 1996). *SP140* expression has been implicated in innate response to immunodeficiency virus type 1 (Madani et al. 2002). Finally, *SP140* was the gene showing the largest difference in expression level between normal and preeclamptic placentas (Heikkilä et al. 2005).

The *HTR2C* SNP (chromosome X: 87,378,551) is located in the intron between exons 3 and 4. *HTR2C* is one of the most important and extensively studied serotonin receptors (Drago et al. 2009). *HTR2C* has ten fixed SNP differences between dogs and wolves, and also belongs to the behavioral fear response (Li et al. 2013). Additionally, *HTR2C* is differentially

expressed in the brain between tame and aggressive mice and foxes (Kukekova et al. 2011), providing additional evidence for its involvement in the tame behaviors of domesticated dogs (Li et al. 2013).

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 1). 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 Figure 3D).

To identify SNPs, and genes proximal to them, that are significantly associated with the observed variation in gestation length across domestic dog breeds, we conducted an across-breed GWAS analysis using 115,683 SNPs and gestation length data from 23 dog breeds (Figure 1D, Supplementary Figure 2D). Our analysis identified six significantly associated SNPs (Supplementary Table S6) that mapped to 4 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 $P = 3.1 \times 10^{-7}$, $Perm\ P = 0.013$), microtubule associated tumor suppressor candidate 2 (*MTUS2*, uncorrected $P = 3.6 \times 10^{-7}$, $Perm\ P = 0.001$), and helicase family member 1 (*HFM1*, uncorrected $P = 4.0 \times 10^{-7}$, $Perm\ P = 0.013$), and two lincRNAs

(ENSCAFG00000037743, uncorrected $P = 4.4 \times 10^{-7}$, $Perm\ P = 0.001$, and ENSCAFG00000039067, uncorrected $P = 1.6 \times 10^{-7}$, $Perm\ P = 0.001$) whose functions are unknown.

The first significantly associated SNP (chromosome 24: 36,399,705) resides in intron 78 of *SLC9A8*, an integral transmembrane protein that exchanges extracellular Na+ for intracellular H+. *SLC9A8* serves multiple functions, including intracellular pH homeostasis, cell volume regulation, and electroneutral NaCl absorption in epithelia (Xu et al. 2008). Knockout male mice have impaired luteinizing hormone-stimulated cAMP production and are infertile, despite normal morphology of their reproductive system and normal behavior (Xu et al. 2015). *SLC9A8* is expressed ubiquitously (Figure 2) (Uhlén et al. 2015), an expression pattern suggestive of involvement in housekeeping functions.

The second SNP (chromosome 20: 35,206,774) is found in the intron between exons 26 and 27 of *CACNA2D3*. This gene is one of four members of the alpha-2/delta subunit three family of the voltage-dependent calcium (Ca2+) channel complex, regulating the influx of Ca2+ ions entering the cell upon membrane polarization (Jin et al. 2017). The regulation of calcium is a fundamental process relevant to life at fertilization, and subsequent control of development and differentiation of cells (Berridge et al. 1998). In previous studies in humans, *CACNA2D3* is differentially methylated in the amnion between normal and preeclamptic pregnancies (Suzuki et al. 2016) and in blood between extreme preterm and term infants at birth (Cruickshank et al. 2013, Eidem et al. 2015). Additionally, *CACNA2D3* is one of four genes recently described as influencing cranial morphology in human

populations (Paternoster et al. 2012). In other domesticated animals, *CACNA2D3* is downregulated by Colony Stimulating Factor 2 (*CSF2*) in the trophoectoderm of pregnant cattle, which increases the ability of the preimplantation embryo to advance to the blastocyst stage (Ozawa et al. 2016). In the closely related wolf, *CACNA2D3* is under diversifying selection associated with environmental adaptations to altitude (Pilot et al. 2014, Schweizer et al. 2016, Zhang et al. 2014).

The third significantly associated SNP (chromosome 25: 10,481,606) falls in a large intronic region of the *MTUS2* gene. The protein product of *MTUS2* is cardiac zipper protein (CAZIP), a member of a class of proteins that interact with angiotensin II receptor interacting proteins (ATIP) (Rodrigues-Ferreira et al. 2010). *MTUS2* plays a role in the development and function of the heart and nervous system in vertebrates (Puy et al. 2009).

The fourth SNP (chromosome 6: 57,457,184) is located in the 3' intron of *HFM1*, a DNA helicase that confers genome integrity in germline tissues (Tanaka et al. 2009). *HMF1* plays a role in meiotic recombination implying a major evolutionary role through the creation of diverse offspring. In mice, deletion of *HFM1* eliminates a major fraction of cross over events (Guiraldelli et al. 2013), whereas in cattle *HMF1* is associated with both fertility and milk production in Holstein cattle (Pimentel et al. 2011), as well as with alteration of global recombination rates in Holstein, Holstein-Friesian, Jersey, and crossbred individuals (Kadri et al. 2016).

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 (Figure 1).

Five of the 13 genetic variants 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 2), 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 (Bannasch et al. 2010, Marchant et al. 2017). 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 (Ellis et al. 2009). Interestingly, one of the traits that associated with brachycephaly is cephalopelvic disproportion (Evans and Adams 2010), 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 (Bannasch et al. 2010). Whether the *SMOC2* variant identified

directly influences parturition and birth timing in dogs (in humans, *SMOC2* is highly expressed in several reproductive tissues; see Figure 2 and Ref. (Uhlén et al. 2015) 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 preeclamptic human pregnancies (Suzuki et al. 2016), and in blood between extreme preterm and term infants at birth (Cruickshank et al. 2013, Eidem et al. 2015). Furthermore, 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 (Hassan et al. 2009). Finally, a few of the other genes significantly or marginally associated with reproductive traits (*SMOC2* and *MITF*) are also known to be expressed in human reproductive tissues (Uhlén et al. 2015)(Figure 2).

Methods

Genotypic and Phenotypic Data. To identify SNPs that are significantly associated with reproductive traits, we used a previously published data set containing 160,727 SNPs from 4,342 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 (Hayward et al. 2016). 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.

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 (Borge et al. 2011, Tønnessen et al. 2012, Evans and Adams 2010, Chatdarong et al. 2007, Concannon et al. 1983, Elits et al. 2005, Evans and White 1997, Gavrilovic et al. 2008, Okkens et al. 1993, Son et al. 2001, Kim et al. 2007, Linde-Forsberg et al.) (see also Supplementary File 1). 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 1. For the body mass control, our collected trait values overlapped with the genotypic data (Hayward et al. 2016) for 101 breeds corresponding to 3,384 individuals (Table 1). For the reproductive traits, our collected cesarean section rate trait values overlapped with the genotypic data for 95 breeds (3,194 individuals), our litter size trait values for 60 breeds (2,617 individuals), our stillbirth rate values for 56 breeds

(2,590 individuals), and our gestation length values for 23 breeds (1,908 individuals) (Table 1).

Genome Wide Association (GWAS) Analyses. To test SNPs for associations with the four reproductive traits of interest, we conducted a GWAS analysis for each individual trait using body mass as a covariate, and accounting for kinship, as well as for body mass as a proof of concept. All GWAS analyses were run using a linear-mixed model as implemented in the program GEMMA, version 0.94 (Zhou and Stephens 2012). Numerous studies have shown that the vast majority of morphological, ecological and physiological traits vary as a function of an organism's body mass (Gould 1966, Kleiber 1932, Shingleton 2010) as well as a function of kinship (Hayward et al. 2016, Boyko et al. 2010). Most notably for the purpose of this study, body mass has been previously shown to be strongly correlated with litter weight (Sacher and Staffeldt 1974, Blueweiss et al. 1978, Tuomi 1980), neonate weight (Sacher and Staffeldt 1974, Blueweiss et al. 1978, Tuomi 1980, Ross 1988), and gestation length (Martin et al. 1985, Sacher and Staffeldt 1974, Blueweiss et al. 1978, Kihlström 1972, Phillips et al. 2015).

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 1).

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 (Lange 2002). Finally, to control for inflated *P* value significance from the testing of multiple hypotheses, we used a significance threshold of $P = 4.3 \times 10^{-7}$ (Bonferroni cutoff of $\alpha = 0.05, N = 115{,}574$) for all analyses. All reported P values are Wald's P values as calculated in GEMMA (Zhou and Stephens 2012).

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 minor allele frequencies are more prone to error due to the small number of samples that have the called nucleotide. Furthermore, we pruned SNPs not in complete or near-complete linkage disequilibrium using a variance inflation factor of 10, using the PLINK command --indep 100 10 10 (Purcell et al. 2007).

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 1,000 permutations. 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 (Zhou and Stephens 2012), we performed an association test for each variant to obtain a single

permuted p-value. The permutation p-value in Table 2 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 greater than 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 (Lindblad-Toh et al. 2005, Hoeppner et al. 2014). Transcript IDs were mapped to gene names using bioconductor biomaRt interface to the ENSEML biomart (Durinck et al. 2009). 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 (R Core Team 2013) with the qqman package (Turner 2014). Calculation of the λ inflation parameter, a metric of any existing systematic bias in the data set, was calculated using the GenABEL R package (Aulchenko 2007) and was used to interpret Type I error rate in the multiple testing of GWAS analyses (Rao et al. 2016).

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 Statement

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) and by the Burroughs Wellcome Fund (to J.A.C and A.R.). MLJ 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.

TABLES

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	3,384
Cesarean Section Rate	97	3,194
Litter Size	60	2,617
Stillbirth Rate	57	2,590
Gestation Length	23	1,908

1 **Table 2. Summary of genes that contain or are adjacent to the SNPs that are associated with variation in reproductive**

2 **traits across dog breeds.**

Gene ID	Gene Name	Chr.	rs Number	Variant	Domestication-related Trait(s)	Reproductive Trait(s)	Empirical p-value	Permutation p-value
SMOC2	SPARC related modular calcium binding 2	1	rs21966904	Non-coding (Intron 13)	Brachycephaly (Dogs)	Cesarean section rate	2.0×10^{-7}	0.011
PALM3	paralemmin 3	20	rs22853767	Non-coding (3' UTR)	-	Cesarean section rate	1.4×10^{-9}	0.001
KRT71	keratin	27	rs23373415	Coding (exon 2)	Coat phenotypes (Dogs)	Cesarean section rate Litter size Stillbirth rate	2.9×10^{-7} 2.2×10^{-8} 3.2×10^{-9}	0.217 0.017 0.064
CD36	CD36 glycoprotein	18	rs22664051	Intergenic variant (downstrea	-	Cesarean section rate	9.7×10^{-8}	0.024

				m)				
RCL1	RNA terminal phosphate cyclase like 1	1	rs21894066	Non-coding (Intron 7)	-	Litter size	2.6×10^{-8}	0.001
MITF	melanogenesis associated transcription factor	20	21848176	Coding (exon 5)	Coat color (Dogs)	Litter size Stillbirth rate	3.5×10^{-7} 1.4×10^{-7}	0.079 0.091
MSRB3	methionine sulfoxide reductase B3	10	rs22060533	Intergenic variant (downstream)	Ear erectness (Dogs)	Litter size	1.3×10^{-7}	0.001
MSANTD1	Myb/SANT DNA binding domain containing	6	rs9084938	Intergenic variant (downstream)	Black coat color (Goats)	Litter size	1.5×10^{-9}	0.001

SP140	nuclear protein body SP140	25	rs8856304	Non-coding (intron 4)	-	Stillbirth rate	2.8×10^{-8}	0.001
HTR2C	5-hydroxytryptamine receptor 2	X	rs24622199	Non-Coding (intron 2)	Tameness (Dogs, Foxes, Mice)	Stillbirth rate	2.0×10^{-7}	0.001
SLC9A8	solute carrier family 9 member A8	24	rs23219089	Non-coding (intron 7)	-	Gestation length	3.7×10^{-11}	0.001
CACNA2D3	calcium voltage-gated channel auxiliary subunit alpha2delta 3	20	rs22853845	Non-coding (intron 9)	Blastocyst development (Cattle)	Gestation length	3.1×10^{-7}	0.013
MTUS2	microtubule	25	10481606	Non-coding	-	Gestation length	3.6×10^{-7}	0.001

	associated tumor suppressor candidate 2			(intron 6)				
HFM1	ATP dependent DNA helicase homolog	6	rs24306896	Non-coding (intron 4)	Fertility and milk production (Cattle)	Gestation length	4.0×10^{-7}	0.013

Figure Legends

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 = 2,617 individuals, n = 60 breeds), (C) stillbirth (n = 2,590 individuals, n = 57 breeds), and (D) gestation length (n =1,908 individuals, n = 23 breeds). Horizontal line indicates the significance threshold at $P = 4.3 \times 10^{-7}$. Significant SNPs are labels with the intersecting or nearest gene. Significant SNPs whose permutation p-values were above the 0.05 threshold are indicated by an asterisk (*). 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 (42).

19 **Supplementary Material**

20

21 **Supplementary Figure 1. Recapitulation of SNPs associated with body mass in 101**
22 **domesticated dog breeds.** (A) Body mass distribution for 101 breeds. (B) Manhattan plots
23 showing the statistical significance of each SNP as a function of genomic position for body
24 mass. Plot generated in R using the qqman package. (C) Quantile-quantile plot showing the
25 effectiveness of the stratification correction ($\lambda = 1.17$). Plot generated in R; inflation factor
26 was calculated using the GenABEL package implemented in R.

27

28 **Supplementary Figure 2. Distribution of phenotypic values of the four reproductive**
29 **traits examined in this study across dog breeds.** (A) cesarean section rate (n = 97
30 breeds), (B) litter size (n = 60 breeds), (C) stillbirth rate (n = 57 breeds), and (D) gestation
31 length (n = 23 breeds). Plots were generated in R using the gglplot2 package.

32

33 **Supplementary Figure 3. Quantile-quantile plots for the GWAS analyses of the four**
34 **reproductive traits.** The range for the inflation factor (λ) for all GWAS analyses is between
35 1.05 – 1.09, indicating the effectiveness of the stratification correction. (A) cesarean section
36 rate ($\lambda = 1.05$), (B) litter size ($\lambda = 1.05$), (C) stillbirth rate ($\lambda = 1.05$), and (D) gestation
37 length ($\lambda = 1.09$). Plots generated in R, and inflation factors were calculated using the
38 GenABEL package implanted in R.

39

40 **Supplementary Table 1. Summary of raw phenotypes for breeds included in analysis.**

41

Supplementary Table 2. Summary of top 100 SNPs associated with body mass.

Supplementary Table 3. Summary of top 100 SNPs associated with cesarean section rate.

Supplementary Table 4. Summary of top 100 SNPs associated with litter size.

Supplementary Table 5. Summary of top 100 SNPs associated with stillbirth rate.

Supplementary Table 6. Summary of top 100 SNPs associated with gestation length.

Supplementary File 1. Sources of phenotypic data.

References

Ahmed ZM, Yousaf R, Lee BC, Khan SN, Lee S, Lee K, et al. Functional null mutations of MSRB3 encoding methionine sulfoxide reductase are associated with human deafness DFNB74. Am J Hum Genet. 2011 Jan 7;88(1):19–29.

Amin N, de Vrij FMS, Baghdadi M, Brouwer RWW, van Rooij JGJ, Jovanova O, et al. A rare missense variant in RCL1 segregates with depression in extended families. Molecular Psychiatry. Nature Publishing Group; 2017 Mar 21;58:1323.

Aulchenko YS, Ripke S, Isaacs A, van Duijn CM. GenABEL: an R library for genome-wide association analysis. Bioinformatics. Oxford University Press; 2007 May 15;23(10):1294–6.

Bagnicka E, Wallin E, Łukaszewicz M, Ådnøy T. Heritability for reproduction traits in Polish and Norwegian populations of dairy goat. Small Ruminant Research. 2007 Apr;68(3):256–62.

Bannasch D, Young A, Myers J, Truvé K, Dickinson P, Gregg J, et al. Localization of Canine Brachycephaly Using an Across Breed Mapping Approach. Fairhead C, editor. PLoS ONE. Public Library of Science; 2010 Mar 10;5(3):e9632.

Basile M, Lin R, Kabbani N, Karpa K, Kilimann M, Simpson I, et al. Paralemmin interacts with D3 dopamine receptors: implications for membrane localization and cAMP signaling. Arch Biochem Biophys. 2006 Feb 1;446(1):60–8.

Beale H, Ostrander EA. Sizing up dogs. Current Biology. Elsevier; 2012 May 8;22(9):R315–6.

Behringer RR, Eakin GS, Renfree MB. Mammalian diversity: gametes, embryos and reproduction. Reprod Fertil Dev. 2006;18(2):99–9.

Benjelloun B, Alberto FJ, Streeter I, Boyer F, Coissac E, Stucki S, et al. Characterizing neutral genomic diversity and selection signatures in indigenous populations of Moroccan goats (Capra hircus) using WGS data. Front Genet. Frontiers Media SA; 2015;6:4447.

Berridge MJ, Bootman MD, Lipp P. Calcium--a life and death signal. Nature. 1998 Oct 15;:645–8.

Bloch DB, la Monte de SM, Guigaouri P, Filippov A, Bloch KD. Identification and characterization of a leukocyte-specific component of the nuclear body. J Biol Chem. 1996 Nov 15;271(46):29198–204.

Blueweiss L, Fox H, Kudzma V, Nakashima D, Peters R, Sams S. Relationships between body size and some life history parameters. Oecologia. Springer; 1978;37(2):257–72.

100 Bolormaa S, Pryce JE, Hayes BJ, Goddard ME. Multivariate analysis of a genome-wide
101 association study in dairy cattle. Journal of Dairy Science. Elsevier; 2010 Aug
102 1;93(8):3818–33.
103
104 Borge KS, Tønnessen R, Nødtvedt A, Indrebø A. Litter size at birth in purebred dogs--a
105 retrospective study of 224 breeds. Theriogenology. 2011 Mar 15;75(5):911–9.
106
107 Boyko AR, Quignon P, Li L, Schoenebeck JJ, Degenhardt JD, Lohmueller KE, et al. A Simple
108 Genetic Architecture Underlies Morphological Variation in Dogs. Hoekstra HE,
109 editor. Plos Biol. Public Library of Science; 2010 Aug 10;8(8):e1000451.
110
111 Brot N, Weissbach L, Werth J, Weissbach H. Enzymatic reduction of protein-bound
112 methionine sulfoxide. Proc Natl Acad Sci USA. National Academy of Sciences; 1981
113 Apr;78(4):2155–8.
114
115 Burris MJ, Blunn CT. Some factors affecting gestation length and birth weight of beef cattle.
116 J Anim Sci. American Society of Animal Science; 1952;11(1):34–41.
117
118 Carneiro M, Piorno V, Rubin C-J, Alves JM, Ferrand N, Alves PC, et al. Candidate genes
119 underlying heritable differences in reproductive seasonality between wild and
120 domestic rabbits. Animal Genetics. 2015 May 22;46(4):418–25.
121
122 Carreira S, Goodall J, Denat L, Rodriguez M, Nuciforo P, Hoek KS, et al. Mitf regulation of
123 Dia1 controls melanoma proliferation and invasiveness. Genes Dev. 2006 Dec
124 15;20(24):3426–39.
125
126 Cassady JP, Johnson RK, Pomp D, Rohrer GA, Van Vleck LD, Spiegel EK, et al. Identification
127 of quantitative trait loci affecting reproduction in pigs. J Anim Sci. 2001
128 Mar;79(3):623–33.
129
130 Chatdarong K, Tummaruk P, Sirivaidyapong S, Raksil S. Seasonal and breed effects on
131 reproductive parameters in bitches in the tropics: a retrospective study. J Small
132 Anim Pract. Blackwell Publishing Ltd; 2007 Aug;48(8):444–8.
133
134 Chen X, Wu X, Zhao Y, Wang G, Feng J, Li Q, et al. A novel binding protein of single
135 immunoglobulin IL-1 receptor-related molecule: Paralemmin-3. Biochem Biophys
136 Res Commun. 2011 Jan 28;404(4):1029–33.
137
138 Concannon P, Whaley S, Lein D, Wissler R. Canine gestation length: variation related to time
139 of mating and fertile life of sperm. Am J Vet Res. 1983 Oct;44(10):1819–21.
140
141 Cross JC, Werb Z, Fisher SJ. Implantation and the placenta: key pieces of the development
142 puzzle. Science. 1994 Dec 2;266(5190):1508–18.
143

144 Cross JF, Martin RD. Calculation of gestation period and other reproductive parameters for
145 primates. Dodo: Journal of the Jersey Wildlife Preservation Trust. Durrell Wildlife
146 Conservation Trust; 1981;18:30–43.
147
148 Cruickshank MN, Oshlack A, Theda C, Davis PG, Martino D, Sheehan P, et al. Analysis of
149 epigenetic changes in survivors of preterm birth reveals the effect of gestational age
150 and evidence for a long term legacy. Genome Med. BioMed Central; 2013;5(10):96.
151
152 Davies Morel MCG, Newcombe JR, Holland SJ. Factors affecting gestation length in the
153 Thoroughbred mare. Animal Reproduction Science. 2002 Dec 16;74(3-4):175–85.
154
155 Derrickson, E. M. "Comparative reproductive strategies of altricial and precocial eutherian
156 mammals." Functional Ecology (1992): 57-65.
157
158 Drago A, Serretti A. Focus on HTR2C: A possible suggestion for genetic studies of complex
159 disorders. American Journal of Medical Genetics Part B: Neuropsychiatric Genetics.
160 Wiley Subscription Services, Inc., A Wiley Company; 2009 Jul 5;150B(5):601–37.
161
162 Dufour DL, Sauther ML. Comparative and evolutionary dimensions of the energetics of
163 human pregnancy and lactation. Am J Hum Biol. Wiley-Blackwell; 2002 Aug
164 21;14(5):584–602.
165
166 Durinck S, Spellman PT, Birney E, Huber W. Mapping identifiers for the integration of
167 genomic datasets with the R/Bioconductor package biomaRt. Nat Protoc.
168 2009;4(8):1184–91.
169
170 Eidem HR, Ackerman WE, McGary KL, Abbot P, Rokas A. Gestational tissue transcriptomics
171 in term and preterm human pregnancies: a systematic review and meta-analysis.
172 BMC Med Genomics. BioMed Central; 2015 Jun 5;8(1):27.
173
174 Eilts BE, Davidson AP, Hosgood G, Paccamonti DL, Baker DG. Factors affecting gestation
175 duration in the bitch. Theriogenology. 2005 Jul;64(2):242–51.
176
177 Elliot MG, Crespi BJ. Phylogenetic Evidence for Early Hemochorial Placentation in Eutheria.
178 Placenta. Elsevier Ltd; 2009 Nov 1;30(11):949–67.
179
180 Ellis JL, Thomason J, Kebreab E, Zubair K, France J. Cranial dimensions and forces of biting
181 in the domestic dog. J Anat. Blackwell Publishing Ltd; 2009 Mar 1;214(3):362–73.
182
183 Elovitz MA, Brown AG, Anton L, Gilstrop M, Heiser L, Bastek J. Distinct cervical microRNA
184 profiles are present in women destined to have a preterm birth. Am J Obstet
185 Gynecol. Elsevier; 2014 Mar 1;210(3):221.e1–221.e11.
186
187 Enders AC, Carter AM. What Can Comparative Studies of Placental Structure Tell Us?—A
188 Review. Placenta. 2004 Apr;25:S3–S9.
189

190 Evans KM, Adams VJ. Proportion of litters of purebred dogs born by caesarean section. J
191 Small Anim Pract. Blackwell Publishing Ltd; 2010 Feb;51(2):113–8.
192
193 Evans JM, White K. The Book of the Bitch. Interpet; 1997. 1 p
194
195 Febbraio M, Hajjar DP, Silverstein RL. CD36: a class B scavenger receptor involved in
196 angiogenesis, atherosclerosis, inflammation, and lipid metabolism. J Clin Invest.
197 American Society for Clinical Investigation; 2001 Sep;108(6):785–91.
198
199 Gavrilovic BB, Andersson K, Linde Forsberg C. Reproductive patterns in the domestic dog—
200 A retrospective study of the Drever breed. Theriogenology. 2008 Sep;70(5):783–94.
201
202 Guiraldelli MF, Eyster C, Wilkerson JL, Dresser ME, Pezza RJ. Mouse HFM1/Mer3 is required
203 for crossover formation and complete synapsis of homologous chromosomes during
204 meiosis. Hawley RS, editor. PLoS Genetics. Public Library of Science; 2013
205 Mar;9(3):e1003383.
206
207 Gould SJ. Allometry and size in ontogeny and phylogeny. Biol Rev Camb Philos Soc. 1966
208 Nov;41(4):587–640.
209
210 Hassan SS, Romero R, Tarca AL, Nhan-Chang C-L, Vaisbuch E, Erez O, et al. The
211 transcriptome of cervical ripening in human pregnancy before the onset of labor at
212 term: identification of novel molecular functions involved in this process. J Matern
213 Fetal Neonatal Med. 2009 Dec;22(12):1183–93.
214
215 Harrison RM. Reproduction in Mammals: General Overview. Chichester, UK: John Wiley &
216 Sons, Ltd; 2001.
217
218 Hayward JJ, Castelhano MG, Oliveira KC, Corey E, Balkman C, Baxter TL, et al. Complex
219 disease and phenotype mapping in the domestic dog. Nature Communications.
220 Nature Publishing Group; 1AD;7:1–11.
221
222 Hoeppner MP, Lundquist A, Pirun M, Meadows JRS, Zamani N, Johnson J, et al. An Improved
223 Canine Genome and a Comprehensive Catalogue of Coding Genes and Non-Coding
224 Transcripts. Chadwick BP, editor. PLoS ONE. Public Library of Science; 2014 Mar
225 13;9(3):e91172.
226
227 Horikoshi M, Beaumont RN, Day FR, Warrington NM, Kooijman MN, Fernandez-Tajes J, et al.
228 Genome-wide associations for birth weight and correlations with adult disease.
229 Nature. 2016 Sep 28;538(7624):248–52.
230
231 Hudson R, Cruz Y, Lucio A, Ninomiya J, Martínez-Gómez M. Temporal and behavioral
232 patterning of parturition in rabbits and rats. Physiol Behav. 1999 Jun;66(4):599–
233 604.
234

Jin Y, Cui D, Ren J, Wang K, Zeng T, Gao L. CACNA2D3 is downregulated in gliomas and functions as a tumor suppressor. Mol Carcinog. 2017 Mar;56(3):945–59.

Johanson JM, Berger PJ, Tsuruta S, Misztal I. A Bayesian threshold-linear model evaluation of perinatal mortality, dystocia, birth weight, and gestation length in a Holstein herd. Journal of Dairy Science. 2011 Jan;94(1):450–60.

Jones KE, Bielby J, Cardillo M, Fritz SA, O'Dell J, Orme CDL, et al. PanTHERIA: a species-level database of life history, ecology, and geography of extant and recently extinct mammals. Ecology. Ecological Society of America; 2009;90(9):2648.

Kadri NK, Harland C, Faux P, Cambisano N, Karim L, Coppieters W, et al. Coding and noncoding variants in HFM1, MLH3, MSH4, MSH5, RNF212, and RNF212B affect recombination rate in cattle. Genome Research. Cold Spring Harbor Lab; 2016 Oct;26(10):1323–32.

Karlsson EK, Lindblad-Toh K. Leader of the pack: gene mapping in dogs and other model organisms. Nature Publishing Group. Nature Publishing Group; 2008 Sep;9(9):713–25

Kihlström JE. Period of Gestation and Body-Weight in Some Placental Mammals. Comp Biochem Physiol. 1972;43(NA3):673–&.

Kiltie RA. Intraspecific Variation in the Mammalian Gestation Period. Journal of Mammalogy. 1982;63(4):646–52.

Kim BS, Son CH. Time of initial detection of fetal and extra-fetal structures by ultrasonographic examination in Miniature Schnauzer bitches. J Vet Sci. The Korean Society of Veterinary Science; 2007 Sep;8(3):289–93.

Kim H-Y, Gladyshev VN. Characterization of mouse endoplasmic reticulum methionine-R-sulfoxide reductase. Biochem Biophys Res Commun. 2004 Aug 6;320(4):1277–83.

Kleiber M. Body size and metabolism. Hilgardia. 1932 Jan;6(January):315–53.

Kukekova AV, Johnson JL, Teiling C, Li L, Oskina IN, Kharlamova AV, et al. Sequence comparison of prefrontal cortical brain transcriptome from a tame and an aggressive silver fox (Vulpes vulpes). BMC Genomics. Third. BioMed Central; 2011 Oct 3;12(1):482.

Lamm KYB, Johnson ML, Baker Phillips J, Muntifering MB, James JM, Jones HN, et al. Inverted formin 2 regulates intracellular trafficking, placentation, and pregnancy outcome. Elife. eLife Sciences Publications Limited; 2018 Jan 8;7:742.

Lange K. Mathematical and Statistical Methods for Genetic Analysis. New York, NY: Springer Science & Business Media; 2002. 1 p.

Langlois B, Blouin C, Chaffaux S. Analysis of several factors of variation of gestation loss in breeding mares. Animal. 2012 Aug 9;6(12):1925–30.

Lartey J, Smith M, Pawade J, Strachan B, Mellor H, Bernal AL. Up-Regulation of Myometrial RHO Effector Proteins (PKN1 and DIAPH1) and CPI-17 (PPP1R14A) Phosphorylation in Human Pregnancy Is Associated with Increased GTP-RHOA in Spontaneous Preterm Labor. Biol Reprod. Oxford University Press; 2007 Jun 1;76(6):971–82.

Lee H, Jaffe AE, Feinberg JI, Tryggvadottir R, Brown S, Montano C, et al. DNA methylation shows genome-wide association of NFIX, RAPGEF2 and MSRB3 with gestational age at birth. International Journal of Epidemiology. 2012 Feb;41(1):188–99.

Lefèvre CM, Sharp JA, Nicholas KR. Evolution of Lactation: Ancient Origin and Extreme Adaptations of the Lactation System. http://dxdoiorg/101146/annurev-genom-082509-141806. Annual Reviews; 2010 Sep 7;11(1):219–38.

Li Y, vonHoldt BM, Reynolds A, Boyko AR, Wayne RK, Wu D-D, et al. Artificial selection on brain-expressed genes during the domestication of dog. Molecular Biology and Evolution. 2013 Aug;30(8):1867–76.

Lindblad-Toh K, Wade CM, Mikkelsen TS, Karlsson EK, Jaffe DB, Kamal M, et al. Genome sequence, comparative analysis and haplotype structure of the domestic dog. Nature. 2005 Dec 8;438(7069):803–19.

Linde Forsberg C, Wikström C, Lundeheim N. Matings by season.

Lyng H, Brøvig RS, Svendsrud DH, Holm R, Kaalhus O, Knutstad K, et al. Gene expressions and copy numbers associated with metastatic phenotypes of uterine cervical cancer. BMC Genomics. BioMed Central; 2006 Oct 20;7(1):268.

Madani N, Millette R, Platt EJ, Marin M, Kozak SL, Bloch DB, et al. Implication of the lymphocyte-specific nuclear body protein Sp140 in an innate response to human immunodeficiency virus type 1. J Virol. American Society for Microbiology (ASM); 2002 Nov;76(21):11133–8.

Maltecca C, Gray KA, Weigel KA, Cassady JP, Ashwell M. A genome-wide association study of direct gestation length in US Holstein and Italian Brown populations. Animal Genetics. 2011 May 6;42(6):585–91.

Marchant TW, Johnson EJ, McTeir L, Johnson CI, Gow A, Liuti T, et al. Canine Brachycephaly Is Associated with a Retrotransposon-Mediated Missplicing of SMOC2. Curr Biol. Elsevier; 2017 Jun 5;27(11):1573–6.

Marsden CD, Ortega-Del Vecchyo D, O'Brien DP, Taylor JF, Ramirez O, Vilà C, et al. Bottlenecks and selective sweeps during domestication have increased deleterious genetic variation in dogs. Proc Natl Acad Sci USA. 2016 Jan 5;113(1):152–7.

Martin RD, MacLarnon AM. Gestation Period, Neonatal Size and Maternal Investment in Placental Mammals. Nature. 1985;313(5999):220–3.

Montenegro D, Romero R, Kim SS, Tarca AL, Draghici S, Kusanovic JP, et al. Expression patterns of microRNAs in the chorioamniotic membranes: a role for microRNAs in human pregnancy and parturition. J Pathol. John Wiley & Sons, Ltd; 2009 Jan;217(1):113–21.

Neff MW, Rine J. A Fetching Model Organism. Cell. 2006 Jan;124(2):229–31.

Okkens AC, Hekerman TWM, de Vogel JWA, van Haaften B. Influence of litter size and breed on variation in length of gestation in the dog. Veterinary Quarterly. 1993 Dec;15(4):160–1.

Ozawa M, Sakatani M, Dobbs KB, Kannampuzha-Francis J, Hansen PJ. Regulation of gene expression in the bovine blastocyst by colony stimulating factor 2. BMC Res Notes. BioMed Central; 2016 Apr 29;9(1):250.

Parker, H. G., VonHoldt, B. M., Quignon, P., Margulies, E. H., Shao, S., Mosher, D. S., ... & Maslen, C. L. (2009). An expressed fgf4 retrogene is associated with breed-defining chondrodysplasia in domestic dogs. *Science*, *325*(5943), 995-998.

Paternoster L, Zhurov AI, Toma AM, Kemp JP, St Pourcain B, Timpson NJ, 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 Mar 9;90(3):478–85.

Phillips JB, Abbot P, Rokas A. Is Preterm Birth a Human-Specific Syndrome? Evolution, Medicine, and Public Health. 2015 Jun 14

Pilot M, Greco C, vonHoldt BM, Jędrzejewska B, Randi E, Jędrzejewski W, et al. Genome-wide signatures of population bottlenecks and diversifying selection in European wolves. Heredity. Nature Publishing Group; 2014 Apr;112(4):428–42.

Pimentel ECG, Bauersachs S, Tietze M, Simianer H, Tetens J, Thaller G, et al. Exploration of relationships between production and fertility traits in dairy cattle via association studies of SNPs within candidate genes derived by expression profiling. Animal Genetics. Blackwell Publishing Ltd; 2011 Jun;42(3):251–62.

Plunkett J, Doniger S, Orabona G, Morgan T, Haataja R, Hallman M, et al. An Evolutionary Genomic Approach to Identify Genes Involved in Human Birth Timing. Barsh GS, editor. PLoS Genetics. 2011 Apr 14;7(4):e1001365.

Pond CM. The Significance of Lactation in the Evolution of Mammals. Evolution. 1977 Mar;31(1):177–99.

Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MAR, Bender D, et al. PLINK: A Tool Set for Whole-Genome Association and Population-Based Linkage Analyses. The American Journal of Human Genetics. 2007 Sep;81(3):559–75.

Puy Du L, Beqqali A, Monshouwer-Kloots J, Haagsman HP, Roelen BAJ, Passier R. CAZIP, a novel protein expressed in the developing heart and nervous system. Dev Dyn. Wiley-Liss, Inc; 2009 Nov;238(11):2903–11.

R Core Team. R: A Language and Environment for Statistical Computing. Vienna, Austria; 2013.

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 Feb 1;40(2):144–53.

Rodrigues-Ferreira S, Nahmias C. An ATIPical family of angiotensin II AT2 receptor-interacting proteins. Trends Endocrinol Metab. 2010 Nov;21(11):684–90.

Rose AB. Intron-mediated regulation of gene expression. Curr Top Microbiol Immunol. 2008;326:277–90.

Ross C. The intrinsic rate of natural increase and reproductive effort in primates. Journal of Zoology. Wiley Online Library; 1988;214(2):199–219.

Sacher GA, Staffeldt EF. Relation of gestation time to brain weight for placental mammals: implications for the theory of vertebrate growth. Am Nat. JSTOR; 1974;963:593–615.

Schierding W, O'Sullivan JM, Derraik JGB, Cutfield WS. Genes and post-term birth: late for delivery. BMC Res Notes. BioMed Central; 2014 Oct 14;7(1):720.

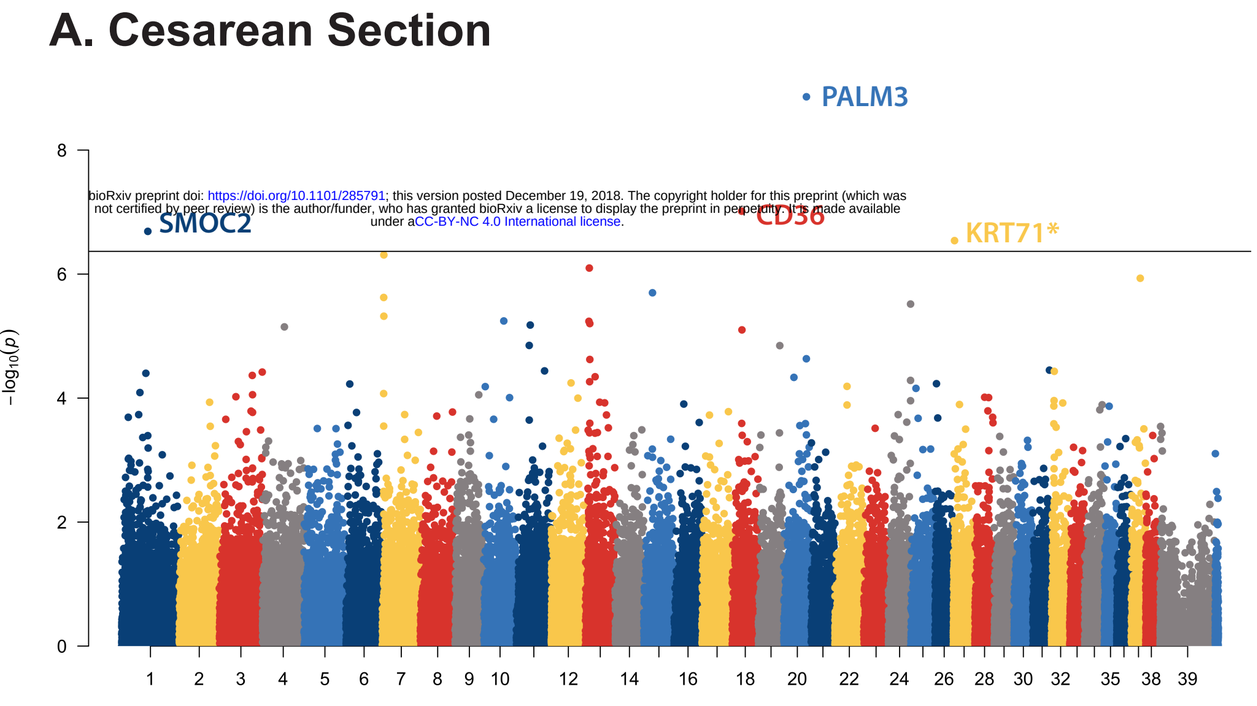
Schweizer RM, vonHoldt BM, Harrigan R, Knowles JC, Musiani M, Coltman D, et al. Genetic subdivision and candidate genes under selection in North American grey wolves. Mol Ecol. 2016 Jan;25(1):380–402.

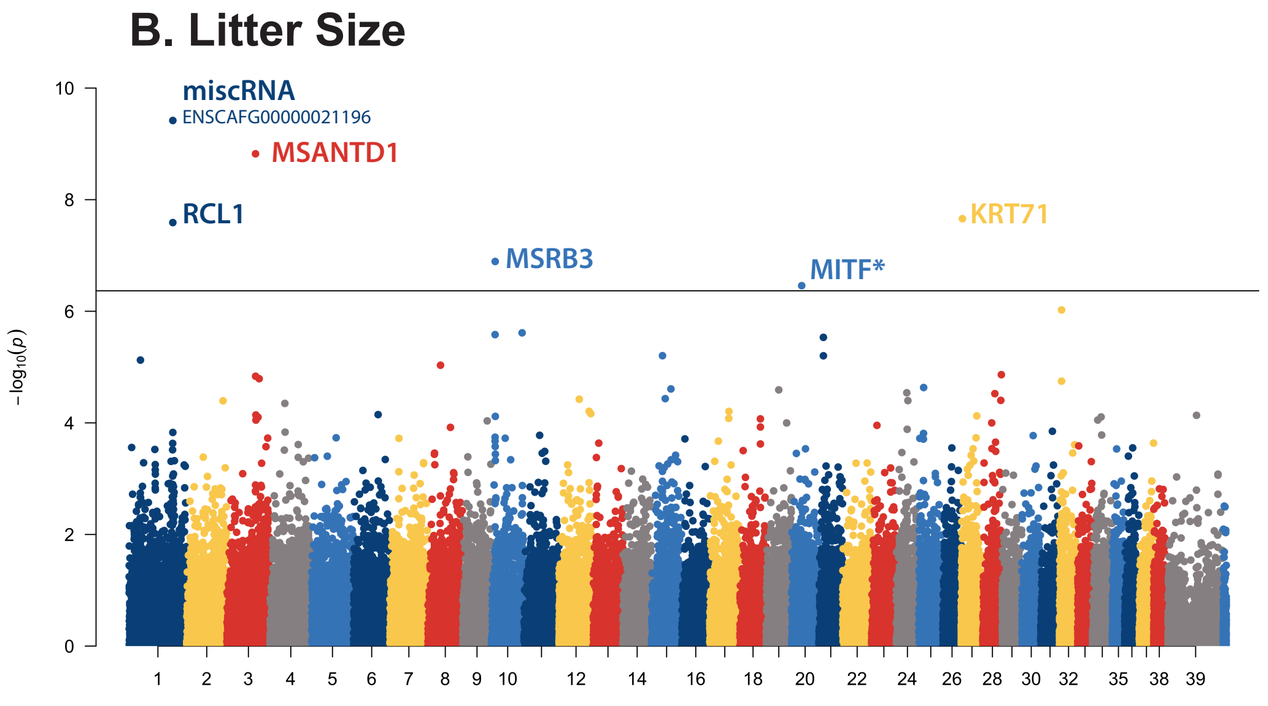
Shingleton A. Allometry: the study of biological scaling. Nature Education Knowledge. 2010;3(10):2.

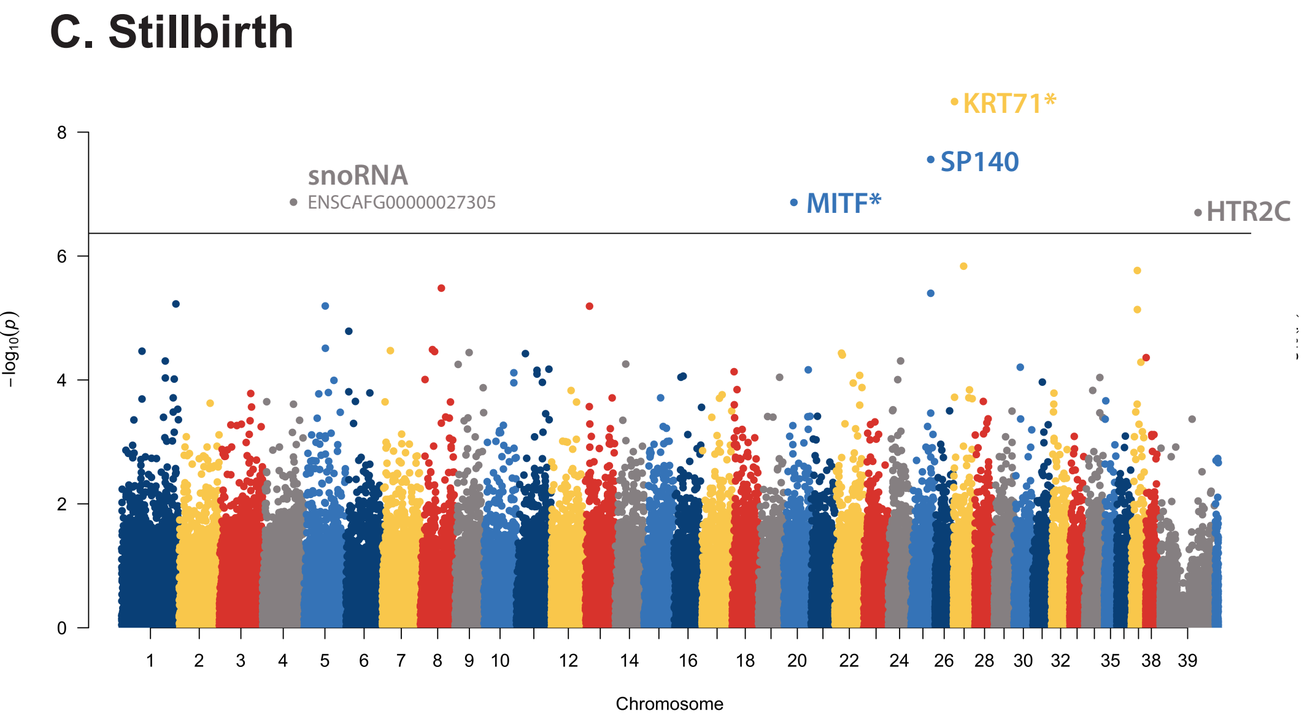
Sironen A, Uimari P, Nagy S, Paku S, Andersson M, Vilkki J. Knobbed acrosome defect is associated with a region containing the genes STK17b and HECW2 on porcine chromosome 15. BMC Genomics. BioMed Central; 2010 Dec 9;11(1):699.

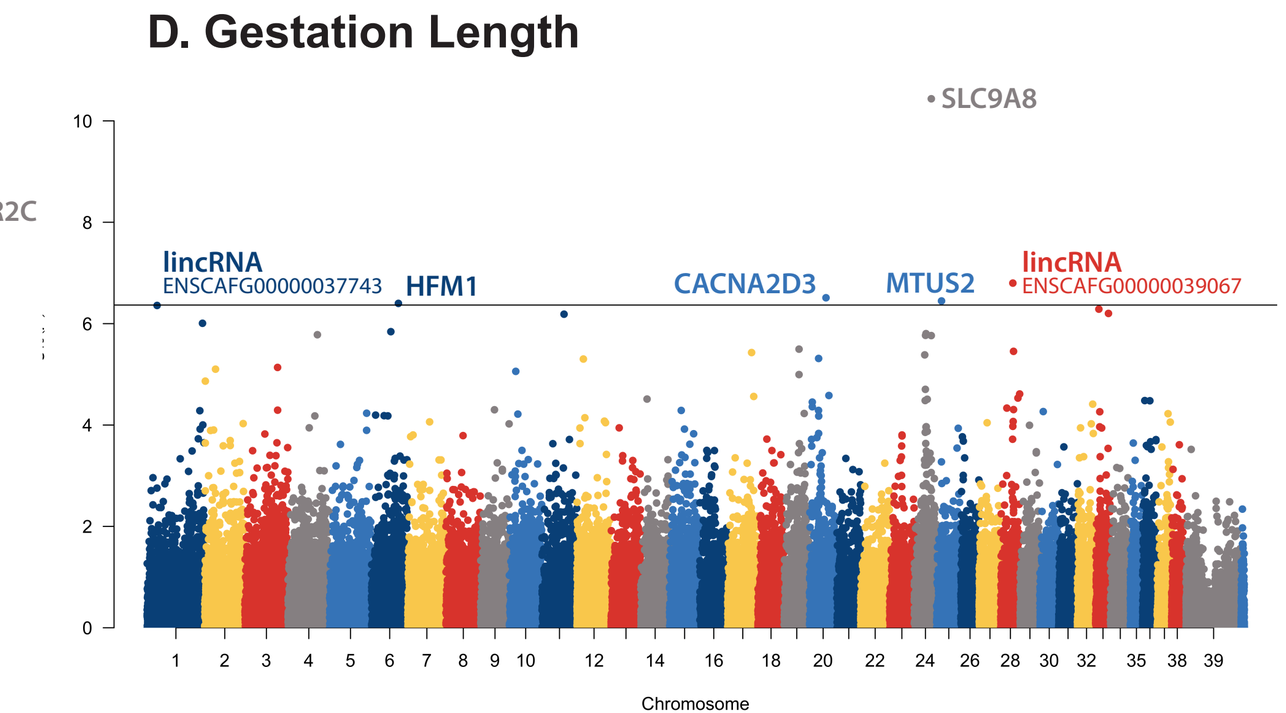
418 Son CH, Jeong KA, Kim JH, Park IC, Kim SH, Lee CS. Establishment of the prediction table of
419 parturition day with ultrasonography in small pet dogs. J Vet Med Sci. 2001
420 Jul;63(7):715–21.
421
422 Suzuki M, Maekawa R, Patterson NE, Reynolds DM, Calder BR, Reznik SE, et al. Amnion as a
423 surrogate tissue reporter of the effects of maternal preeclampsia on the fetus. Clin
424 Epigenetics. BioMed Central; 2016;8(1):67.
425
426 Tanaka K, Miyamoto N, Shouguchi-Miyata J, Ikeda J-E. HFM1, the human homologue of
427 yeast Mer3, encodes a putative DNA helicase expressed specifically in germ-line
428 cells. DNA Sequence. 2009 Jul 11;17(3):242–6.
429
430 Tandon NN, Kralisz U, Jamieson GA. Identification of glycoprotein IV (CD36) as a primary
431 receptor for platelet-collagen adhesion. J Biol Chem. 1989 May 5;264(13):7576–83.
432
433 Tønnessen R, Borge KS, Nødtvedt A, Indrebø A. Canine perinatal mortality: A cohort study
434 of 224 breeds. THE. Elsevier; 2012 Jun 1;77(9):1788–801.
435
436 Tuomi J. Mammalian reproductive strategies: a generalized relation of litter size to body
437 size. Oecologia. Springer; 1980;45(1):39–44.
438
439 Turner SD. qqman: an R package for visualizing GWAS results using Q-Q and manhattan
440 plots. bioRxiv. Cold Spring Harbor Laboratory; 2014 May 14;:005165.
441
442 Uhlén M, Fagerberg L, Hallström BM, Lindskog C, Oksvold P, Mardinoglu A, et al. Tissue-
443 based map of the human proteome. Science. American Association for the
444 Advancement of Science; 2015 Jan 23;347(6220):1260419–9.
445
446 Vaysse A, Ratnakumar A, Derrien T, Axelsson E, Pielberg GR, Sigurdsson S, et al.
447 Identification of Genomic Regions Associated with Phenotypic Variation between
448 Dog Breeds using Selection Mapping. Akey JM, editor. PLoS Genetics. Public Library
449 of Science; 2011 Oct 13;7(10):e1002316.
450
451 Walker R, Gurven M, Hill K, Migliano A, Chagnon N, De Souza R, et al. Growth rates and life
452 histories in twenty-two small-scale societies. Am J Hum Biol. 2006;18(3):295–311.
453
454 Wang X, Liu J, Zhou G, Guo J, Yan H, Niu Y, et al. Whole-genome sequencing of eight goat
455 populations for the detection of selection signatures underlying production and
456 adaptive traits. Sci Rep. Nature Publishing Group; 2016 Dec 9;:1–10.
457
458 Xu H, Chen H, Dong J, Lynch R, Ghishan FK. Gastrointestinal distribution and kinetic
459 characterization of the sodium-hydrogen exchanger isoform 8 (NHE8). Cell Physiol
460 Biochem. Karger Publishers; 2008;21(1-3):109–16.
461

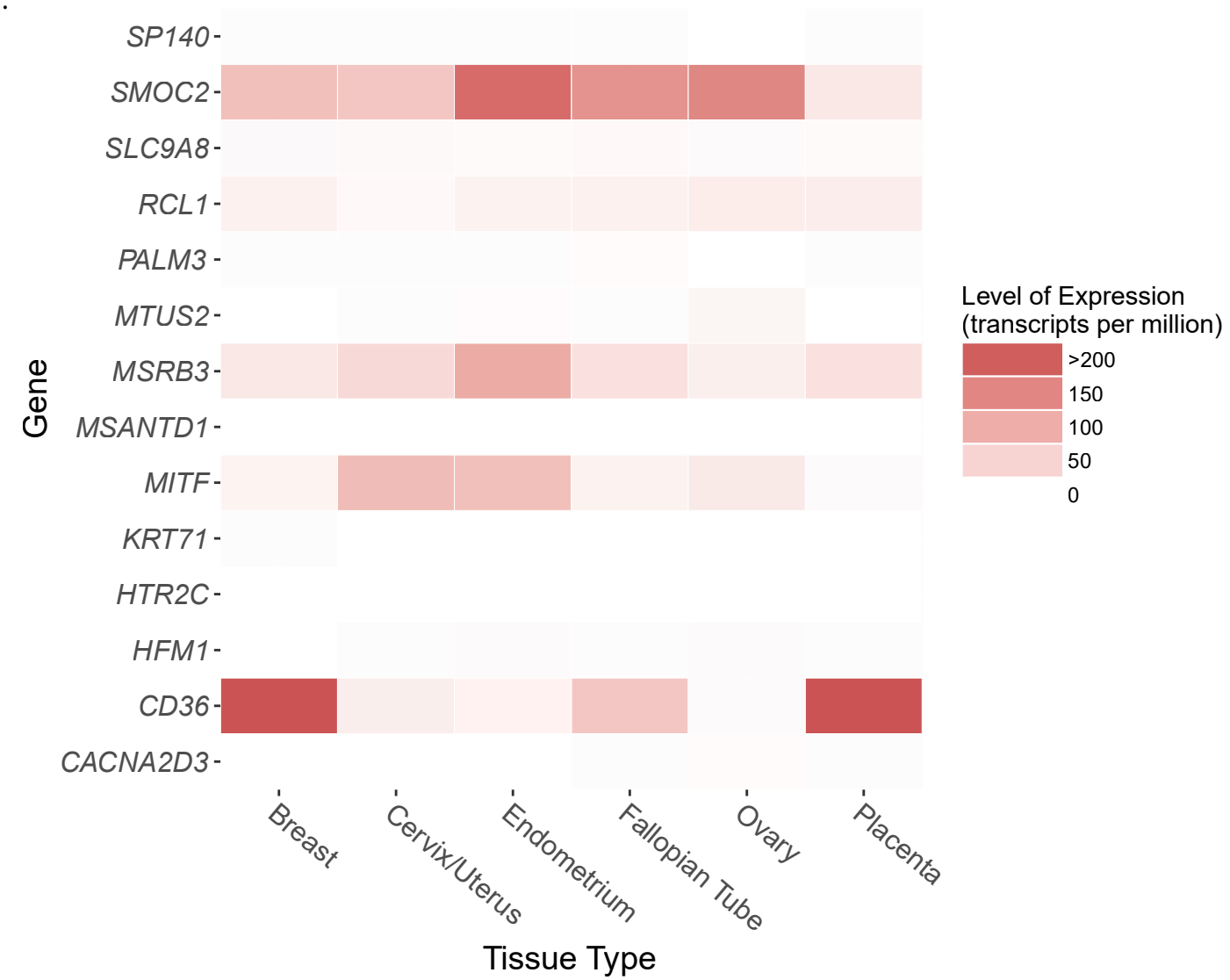
462 Xu H, Chen H, Li J, Zhao Y, Ghishan FK. Disruption of NHE8 expression impairs Leydig cell
463 function in the testes. American Journal of Physiology- Cell Physiology. American
464 Physiological Society; 2015 Feb 15;308(4):C330–8.
465
466 Yu F, Shen X-Y, Fan L, Yu Z-C. Genome-wide analysis of genetic variations assisted by
467 Ingenuity Pathway Analysis to comprehensively investigate potential genetic targets
468 associated with the progression of hepatocellular carcinoma. Eur Rev Med
469 Pharmacol Sci. 2014;18(15):2102–8.
470
471 Zhang G, Feenstra B, Bacelis J, Liu X, Muglia LM, Juodakis J, et al. Genetic Associations with
472 Gestational Duration and Spontaneous Preterm Birth. N Engl J Med. 2017 Sep
473 21;377(12):1156–67.
474
475 Zhang H, Baldwin DA, Bukowski RK, Parry S, Xu Y, Song C, et al. A genome-wide association
476 study of early spontaneous preterm delivery. Genet Epidemiol. 2015
477 Mar;39(3):217–26.
478
479 Zhang W, Fan Z, Han E, Hou R, Zhang L, Galaverni M, et al. Hypoxia adaptations in the grey
480 wolf (Canis lupus chanco) from Qinghai-Tibet Plateau. Akey JM, editor. PLoS
481 Genetics. 2014 Jul;10(7):e1004466.
482
483 Zhou X, Stephens M. Genome-wide efficient mixed-model analysis for association studies.
484 Nature Genetics. 2012 Jun 17;44(7):821–4.
485










SP140
SMOC2
SLC9A8
RCL1
PALM3
MTUS2
MSRB3
MSANTD1
MITF
KRT71
HTR2C
HFM1
CD36
CACNA2D3
Gene
Breast
Cervix/Uterus
Endometrium
Fallopian Tube
Ovary
Placenta
Tissue Type
Level of Expression
(transcripts per million)
>200
150
100
50
0