Genome scans of dog behavior implicate a gene network underlying psychopathology in mammals, including humans

Authors

Isain Zapata[1,2], Erin E. Hecht[3], James A. Serpell[4], Carlos E. Alvarez[1,5,6]*

Affiliations

[1]Department of Veterinary Clinical Sciences, The Ohio State University College of Veterinary Medicine, Columbus, OH, 43210, USA.

[2]Department of Biomedical Sciences, Rocky Vista University College of Osteopathic Medicine, Parker, CO, 80134, USA.

[3]Department of Human Evolutionary Biology, Harvard University, Cambridge, MA, 02138, USA.

[4]Center for the Interaction of Animals and Society, School of Veterinary Medicine, University of Pennsylvania, Philadelphia, PA, 19104, USA.

[5]Center for Clinical and Translational Research, The Research Institute at Nationwide Children's Hospital, Columbus, OH, 43205, USA.

[6]Department of Pediatrics, The Ohio State University College of Medicine, Columbus, OH, 43210, USA.

*Corresponding Author: carlos.alvarez@nationwidechildrens.org

Abstract

Genetic studies show a general factor associated with all human psychopathology and strongly correlated with personality and intelligence, but its basis is unknown. We performed genome scans of 17 normal and problem behaviors in three multi-breed dog cohorts. 21 of 90 mapped loci were supported for the same, or a related, trait in a second cohort. Several of those loci were also associated with brain structure differences across breeds; and six of the respective top-candidate genes are also associated with human brain structure and function. More broadly, the geneset of canine behavioral scans is supported by enrichment for genes mapped for human behavior, personality, cognition, psychopathology and brain structure. The biology implicated includes, neurogenesis, axon guidance, angiogenesis, brain structure, alternative splicing, disease association, Hox-family transcription factors, and subiculum expression. Because body size and behavior are correlated in dogs, we isolated the effect of body size in the dog mapping and in the comparative human UK Biobank analyses. Our dog findings are consistent with pleiotropy of diverse brain traits with energy metabolism and growth, and suggest behavioral variations often affect neurogenesis. There is support for such pleiotropy in humans and well-powered genetic studies of human psychiatric traits consistently implicate neurogenesis. We propose a genetic network which underlies neuron birth and development throughout life is associated with evolutionary adaptation of behavior and the general psychopathology factor. This understanding has implications for genetic and environmental contributions to psychiatric disease. We discuss how canine translational models can further accelerate the study of psychopathology.

Author summary

We genetically mapped diverse normal and problem behaviors in dogs. The well-established approach we used is ideally suited for finding variation that is common across dog breeds and for pin-pointing the most likely gene candidates. Our analysis of the genes implicated at 90 genome regions shows they are

enriched for i) genes mapped for diverse brain functions and pathologies in humans; ii) genes involved in brain development throughout life; and iii) footprints of evolution in dogs, humans and other animals. We propose that is consistent with evolutionary conservation of the general genetic factor of mental health in humans, which is correlated with personality and intelligence. The implications are that this super-network of genes is preferentially targeted by evolutionary adaptation for behavior and that its dysregulation increases risk of mental health disorders.

Introduction

Beginning in 1916 and peaking in the last decade, twin and family studies have revealed a general genetic factor "p" that underlies risk for all human psychopathology[1-3]; and which is strongly correlated with personality and intelligence[4,5]. This understanding is consistent with more recent genome wide association studies (GWASs) in humans[2] and dogs[6,7]. Human cross-disorder psychiatric GWASs showed high levels of polygenicity and genetic-relatedness[8]. 109/146 of those mapped loci (~75%) were pleiotropic for psychiatric traits and were enriched for neurodevelopmental genes[8]. Those findings are consistent with cross-trait studies in dogs[6], and with human GWASs of individual traits involving diverse brain functions and structure. However, essentially all variants mapped for human complex traits have minute or small effect sizes. Thus, much of the current focus is on polygenic risk scores, and few individual risk variations have clinical or experimental utility. Biological understanding is more likely to be derived from rare large-effect variations. In contrast to human, canine behavioral pleiotropy manifests in many moderate-to-large effect variations that are common across breeds[6,7]. These present powerful and unique opportunities for biological dissection and medical translation[9,10].

Results

Interbreed genome scanning of behavior, body size and lifespan

Interbreed scanning has two great advantages: i) it allows mapping of variations that are commonly fixed in different breeds, and ii) it has the effect of fine mapping due to the breaking down of linkage disequilibrium (LD) on both sides of functional variations[7,11]. As we and others have done before[6,7], we performed GWASs using breed averages of 17 C-BARQ behavioral phenotypes (Suppl. Table S1; ref. [12]). Fourteen of these C-BARQ traits had average interbreed-heritabilities of 0.51 ±0.12 (SD; h^2 range =0.27 for non-social fear to 0.77 for trainability)[6]. We also mapped body mass and lifespan (Suppl. Text). We used genotypes from 3,752 dogs total in three cohorts of partially-overlapping breed make-up to conduct separate GWASs of each trait in each cohort (29 breeds, *n*=444; 37 breeds, *n*=423; and 50 breeds, *n*=2885)[13-15]. Cross-cohort analysis provides quasi-replication and reduces false positives due to population structure and latent variables. We measured single-marker association for each trait, correcting for population structure within each cohort by using linear mixed models in GEMMA[16]. The inflation factor (λ) had a narrow range with an average of 1.14, slightly above the 1.05-1.1 considered benign (Suppl. Text). In total, we found genome wide significant associations at 90 loci across all behavioral traits (Fig. 1A; Table 1; Suppl. Data 1, Tables S2-6)[13-15]. Eleven loci were supported for the same or a related trait in at least a second cohort – here or in our previous mapping[7] – or for different behaviors. Ten additional loci were supported by another interbreed GWAS[6]. That quasi-replication mitigates possible type I error (see Discussion for support from studies with individual-level phenotypes and genotypes[17,18]). The geneset analyses below suggest it is unlikely our mapping has a high false positive rate.

Effect direction and amount of variation explained

To determine effect direction, we built regression models using stepwise selection of loci with significant contributions to the model **(**Fig. 2A). Eigen decomposition was first used to cluster the markers correlated by linkage into one. Several loci showed the presence of multiple regions with different

effects on the variable. This has been reported for morphological and behavioral traits mapped to some of these loci (e.g., chr10 *MSRB3-HGMA2* locus[7,14]). The loci at chr10, chr15, chr18, chr20 and chrX were among the strongest findings. The directions of the effects are corroborated across the three cohorts (note Boyko appears opposite due to designation of minor allele; e.g., chr15/*IGF1* and body size). Many associations were significant after Bonferroni adjustment.

Genetic associations with brain structure

We performed joint analyses of behavioral loci and published brain imaging data from 62 dogs of both sexes from 33 breeds[19]. Magnetic resonance imaging (MRI) data was normalized for brain volume, and used to measure gray matter differences across breeds and sexes. We tested for association using allele frequencies at our quasi-replicated loci and the factor loading coefficients of each of the six reported independent component (IC) brain networks (Fig. 2B; Suppl. Text). Note IC2, IC3 and IC5 are strongly-significantly associated, while IC1 is marginally associated with total brain volume[19]. We evaluated each association all together for all IC's and all relevant loci in our analysis. Many associations were significant after Bonferroni adjustment. Of those, both IC2 and IC3 were associated with the chr15 and chr18 loci in all three cohorts, these associations reached individual Bonferroni significance. Multiple linear regression estimated all priority loci together account for 56%, 45% and 57% of the variance for IC2 in the three cohorts while they accounted for 66%, 67% and 57% for IC3 in the three cohorts. For IC6, two cohorts showed the variance explained was 60% and 29%. Supporting our findings, six top candidate genes at three of these loci have also been reported to be associated with differences in human brain structure: *MSRB3* and *HMGA2*, chr10 locus; *FOXP1* and *MITF*, chr20; *IGF1*, chr15; and *GNAT3*, chr18 (Table 1)[20-22]. Human alleles of those first four genes were associated with both brain structure and cognitive traits in the same study[21].

Trait associations, gene annotation and brain relevance of quasi-replicated loci

The four loci mapped for fear and aggression traits in multiple cohorts in our prior study[7] (chr10:8.1, chr15, chr18 and chrX:102.1) were confirmed here and additional trait associations were identified, including brain structure and body size for all four. As we previously reported, the gene annotation of those loci shows strong behavioral relevance (Suppl. Text)[7]. The new loci on chromosomes 20 and 24 have strong behavioral candidates (Suppl. Text). Both loci are under strong evolutionary selection, but the coat-pattern alleles presumed to be under selection cannot be discerned from the present SNP genotype data[23]. Two cohorts had chr20 associations with escaping and owner-directed aggression; and single cohorts had associations with chasing, separation anxiety and separation-related urination. The chr20 locus contains three transcription factor genes: *MITF*, *MDFIC2* and *FOXP1*. The chr24 locus was mapped in two cohorts for nonsocial fear, rivalry aggression and separation anxiety. The genes nearest the peak marker are *RALY*, *EIF2S2*, *ASIP* and *AHCY*. Both *MITF* and agouti/*ASIP* are well known for key roles in external pigmentation, but also have strong evidence of behavioral roles. Although *MITF* is at the same locus as the prominent behavioral gene *FOXP1*, math ability, educational attainment and brain structure traits have been specifically mapped to *MITF* in human GWASs[21,24,25]. Strikingly, the targeted null-mutation of *Asip* in a wild-derived mouse strain strongly increased tameness and reduced aggression in comparison to the wild type strain[26]. Domesticated strains of several mammalian species are deficient in *Asip* and are tamer than wild type (Suppl. Text)[27-30]. The most famous example is Belyaev's "Russian farm-foxes" which were selected for tameness for 50 generations beginning in 1959. Notably, the initial "silver foxes" tamed were silver- or black-colored Canadian red foxes later found to be null for *Asip*[31]. We found that *ASIP* mRNA is expressed in specific brain regions in humans and cattle, most prominently in the subiculum, subregions of the hippocampus, amygdala and hypothalamus, and cerebellum (Suppl. Text). These and the cited findings suggest *Asip* is one of the most important genes in domestication and the seminal tameness-variant in Belyaev's silver foxes. This presents a major

cautionary note given the vast majority of mammalian neuroscience research is conducted in C57BL/6J mice, which are null for *Asip* ("*a*" in mouse).

The following two quasi-replicated loci are among those supported by biological relevance (Suppl. Fig. S1; Suppl. Text). The haplotype on chr1, which implicates *RCL1*, was previously known to be associated with morphological and reproductive traits[14,15]. Here it was associated with chasing in two cohorts and anxiety traits in single cohorts. Human linkage and association studies showed a missense mutation in *RCL1* is associated with depression[32]. *ANGPT1* is very near the chr13 peak associated with excitability, nonsocial fear and body size in two cohorts and with dog aggression, separation anxiety, attachment and lifespan in single cohorts. This chr13 region has strong signals of evolutionary selection and interbreed differentiation in many breeds[14], but the basis for that is unknown. A SNP within human *ANGPT1* is associated with subjective well-being in Europeans[33].

Biological pathway and transcription factor relevance

We identified 127 candidate genes in the 90 loci mapped for any trait. For all geneset analyses, we only used genes with a human protein-coding ortholog (HUGO HGNC[34]). To evaluate the effect of body size, we compared our findings with a similar study that controlled for body weight. MacLean and colleagues mapped breed averages of 14 C-BARQ behaviors included in this work, but designed the association analysis to remove the effect of body weight as a proxy for body size[6]. They used two genotype cohorts, one in common with this work[13]. Our single-marker geneset of 108 and their gene-based mapping geneset of 715 share 19 genes, which makes them significantly overlapping (hypergeometric $P=6.91\times10^{-6}$). First, we tested for enrichment in diverse curated categories using Enrichr and DAVID[35,36]. Table 2 is a summary of the pathway and tissue analyses (Suppl. Tables S7-33). Many associations are shared by the two canine behavioral GWA genesets, including the top ranking of the biological pathway axon guidance and the annotation for tissue enrichment of brain. Both were significant for various aspects of

embryonic development and morphogenesis. Both were significant and similarly ranked for alternative splicing and disease mutation. Based on genome wide *in situ* hybridization data for mouse brain, the two genesets strongly overlapped for the subiculum and its subregions.

Approximately 100 transcription factor binding sites were significantly enriched for our geneset (Suppl. Table S25). The top 20 binding sites correspond to 26 transcription factors enriched for developmental processes, including 11 associated with both generation and differentiation of neurons (Table 3). Analysis of our geneset combined with MacLean et al.'s yielded 132 factors predicted by two algorithms to bind at 98% of mapped genes (Suppl. Table S34). With their overlap underlined, the top factors according to significance, beginning with the first, were POU6F1, LHX3, FOXA2, CDC5L, ALX1, PRX2 and MEF2A; and, according to fold-enrichment, IRF1, FOXO3, LHX3, FOXD3, FOXF2, POU6F1, FOXA2. Those are enriched for two classes of developmental transcription factor families: Fox and Hox (n=4, P_{Adj}=1.11x10^{-5}, and n=3, P_{Adj}=0.153, respectively; Enrichr).

Human GWA relevance

Table 4 summarizes results of the geneset analyses of human genetics. We first tested the compiled set of all GWASs with *n*>50,000 (ref. [37]). The MacLean geneset was enriched for the categories Activities, Environment, Cognition and Reproduction (Psychiatric was suggestive), of which Activities and Environment remained significant after Bonferroni correction (Suppl. Table S35). The small Social Interactions set (379 genes) was omitted from the main analysis, but was suggestive (*P*=7.95x10^{-3}). Our far lower-powered geneset was not significant for any, but had the smallest *P* value for the same category as MacLean's and combining the two improved the significance of all positive categories.

We tested the two dog genesets in parallel against three datasets of human genetics (Suppl. Tables S27-33). Enrichment testing of the GWAS Catalog (NHGRI-EBI) strongly reflected the difference in our study design vs. MacLean et al.'s, who sought to remove the effect of body size (Table 4). The traits

that were significant and top-ranked in both genesets were systolic blood pressure, QRS duration and pulse pressure. Analysis of the UK Biobank human GWAS dataset revealed 79 significant traits for our geneset and 181 for MacLean et al.'s. Both the UK Biobank and dbGAP analyses showed our geneset ranked higher than MacLean's for height-related traits and vice versa for traits related to metabolism and body mass index (BMI). These patterns suggest breed-average weights are good proxies for body size in dogs, and show much of that variation was removed in MacLean et al.'s GWASs (whereas BMI ranked #1). However, the MacLean geneset contained four of our nine genes/loci quasi-replicated for behavior and associated with body size (*MSRB3, ANGPT1, IGF2BP2, IGSF1*), and had strong signal for height in the UK Biobank GWAS and dbGAP (Bonferroni P=1.08x10^{-5} and P=5.63x10^{-8}, respectively).

Both canine behavior GWA genesets were significant for cardiovascular traits in all three human genetics datasets. Both were also significant for several other trait groups in the UK Biobank GWAS dataset: lung function, blood cell counts, skin color, behavior and demographics. Skin color associations are consistent with two of our behavioral loci which contain genes associated with coat color and pattern, but also expressed in the brain (*ASIP* and *MITF*). The brain traits significant in both genesets include some related to depression, demographics likely to be correlated with both socioeconomics and education, and reaction time. Six alcohol, tobacco and drug use traits were significant in MacLean's geneset, but only one of those was nominally significant in our much smaller geneset. Two anxiety traits and neuroticism scores were significant in MacLean's geneset, but not in ours.

We considered whether the comparative human GWA findings could be due to enrichment for body size/height. We asked if top human-GWAS traits associated with the canine genesets were also associated with height. Using the 162 complex-trait Polygenic Risk Scores (PRS) Atlas derived from UK Biobank GWA data[38], we performed phenome scans in that population (Suppl. Text and Table S35). The traits with top associations with our geneset but not MacLean et al.'s – forced vital capacity and height – were extremely-highly significant for height PRS's (both also for many anthropometric and diverse brain

traits). In contrast, the top traits common to both genesets did not show that pattern: Distance between home and workplace was only significant, with Bonferroni adjustment, for the Years of schooling PRS; in contrast, the Height PRS ranked #72, P=0.44. Also, pulse wave arterial stiffness index (defined as height divided by systole peak-to-peak time) was significantly, and, by very far, most strongly associated with the PRS of Blood pressure (a known correlation[39]). This suggests the top traits shared by the two dog studies are not primarily due to their correlation with body size in humans.

Analyses of hypothalamic single-cell expression data from juvenile mice

Based on data availability and tissue relevance[7], we queried single cell RNA sequencing (scRNA-seq) data from hypothalamus (detailed in Suppl. Text)[40]. That dataset is based on dissection of a vertical column spanning the rostrocaudal span from preoptic to tuberal areas from pooled hypothalami of 2-4 week old mice of both sexes. We focused on the 62 neurons that were defined as clusters of sequences. First we identified the clusters depleted (n=3) or enriched (n=6) for both our GWA geneset and MacLean et al.'s by applying a 1-SD threshold to equal numbers of differentiation-ranked genes for each cluster (using rank-sum P). Hypergeometric testing with Bonferroni correction showed all three depleted clusters and enriched Cluster 61 were significantly depleted or enriched, respectively (Fig. 3B). We also tested the nine clusters against datasets of human genes associated with brain traits having approximately 700 or more genes. The traits were neuroticism and wellbeing (weakly powered), educational attainment, alcohol and tobacco use, brain structure and autism. Genesets were from GWA and gene-based GWA, except for the more speculative autism set based on exome sequencing and network-based predicted genes. We observed a general pattern whereby enriched clusters were strongly enriched for all the human datasets by comparison to the depleted clusters. Clusters 55 and 61 consistently had the strongest effects.

None of the six priority scRNA-seq clusters have been mapped[40]. Five of the six, including #55 and #61, are glutamate neurons. While few of the 62 clusters were reported to be sex-associated, #55 was predominantly male and #61 female[40]. We noted #55 is among the most differentiated clusters for the Y chromosome gene *Ddx3y*, and #61 is the top-most for the X-inactivation gene *Xist* (rank-sum P=6.19x10^{-2}, P=8.38x10^{-4}, respectively). Further studies are necessary to rule out other explanations for these sex effects (e.g., systematic dissection differences by age or sex). To begin to understand the biology of the six priority neurons, we compiled their differential expression of 155 signaling genes (Suppl. Text and Table S36). Among the findings relevant to our dog behavioral mapping and the question of pleiotropy with body size/mass and metabolism, *Igf1*, *Igf1r, Igf2r* and *Igfbp5* are differentiating for Cluster 55 (as are *Abcc8* and *Kcnj11*, indicative of glucose-sensing), and *Igf1r* is for #61. We next evaluated the most differentiating-genes shared by the six priority clusters compared to the rest of the 62 hypothalamic neuron clusters (Suppl. Text; Table S37). Those include many genes prominent in brain biology, such as *Spock2*, *Chrna4*, *Gabbr1*, *Kcnc1*, *Nrxn1*, *Celsr2*, *Megf8* and *Rbfox1*. We analyzed the top 18 most differentiated genes shared by the six clusters. 17 of 18 are associated with 374 human traits based on GWA. 12 of 18 are associated with 120 brain traits, most commonly in the area of drug use, followed by psychopathology and intelligence. Of the other six, one is associated with circulating IGF1 levels and the remaining five are highly expressed in the brain. Of the 18 top genes, seven are associated with height and six with BMI – with two overlapping. Four of 18 are associated with longevity (overlap of one each for height and BMI).

Transcription factors that significantly differentiate priority clusters supported the predictions based on both dog GWA genesets (Suppl. Text). For example, the top four Table 3 predictions – *POU6F1*, *POU2F1/2*, and *MEF2A* – were corroborated in the priority clusters. Out of the 62 hypothalamic neuron clusters, those four were differentiating in nine, four, seven and six clusters, respectively. Of the six priority clusters, all five glutamatergic clusters (but not GABAergic #17) were supported by at least one

of those four transcription factors. All four were differentiating for Cluster 55 (as were the dog behavioral GWA genes *Pou6f2* and *Cux1*). Visual inspection of cluster differentiating genes revealed a neurogenesis and neurodevelopment profile for Cluster 55 (Suppl. Table S38). We thus tested all clusters for their content of dog behavioral GWA, neurogenesis and angiogenesis genes (Fig. 3A; Suppl. Text). This showed clear correlation between the mouse hypothalamic neuron cluster pattern and their overlap with dog GWA and mammalian neurogenesis genesets. Clusters 46-62 also correlated with angiogenesis. The restriction of *Lef1* and *Grin2b* expression to Clusters 51-62 suggests the implicated clusters are located more caudally, and are more immature and still undergoing synaptogenesis[41,42].

Testing for relevance to neurogenesis in human embryonic neocortex

We next tested for neurogenesis effects using open chromatin data from the developing neocortex of humans[43]. We found strong relevance of the combined canine GWASs for transcription factor motifs found in the neural precursor-enriched germinal zone *vs.* the neuron-enriched cortical plate (Fig. 4). That included transcription factors mapped and motifs predicted by the dog GWASs, and their paralogs. For example, the most strongly predicted transcription factor based on the dog mapping was POU6F1, which was supported by the mouse hypothalamus scRNA-seq data. It had the second strongest signal for germinal zone enrichment (P_{Adj}=2.96x10^{-92}). Two mapped dog genes, *CUX1* and *POU6F2*, were ranked eighth and eleventh, respectively (P_{Adj}=3.23x10^{-18} and 5.09x10^{-12}). The top neurogenic transcription factors in that study and those overlapping the dog GWASs are enriched for the Hox family (incl. POU, CUX and PAX subfamilies).

Detection of evolutionary footprints

We analyzed the genesets of the present and weight-corrected GWASs of dog behavior for relevance to selection (Table 5). Both were enriched for genes under positive selection in humans and domesticated

dogs. Relevance of selection in 15 Chinese indigenous dog breeds hints that mapped variants predate breed creation and thus represent standing variation in wolves. Other domesticated animals are difficult to test because selection regions tend to be very large and contain many candidate genes, but the combined dog genesets were nominally significant for 666 candidate genes implicated in at least two of cattle, goats, pigs or sheep. Consistent with each other, both genesets are enriched for loss-of-function intolerant genes in humans[44], and suggestive of depletion for genes with coding variation under selection in diverse wild and domesticated vertebrate species[45]. Both genesets are depleted for human single-trait associations and enriched for human accelerated-divergence regions[46]. In the preceding section we showed relevance to transcription factor motifs enriched in open chromatin in the germinal zone of developing human neocortex. That open chromatin is also enriched for human-gained enhancers, and their targeted genes are expressed in the progenitor-enriched laminae of the cortex[43].

Discussion

Our canine investigations of behavior, genetics, brain imaging and comparative genomics are core facets of the neurobiology and evolution approach begun by Darwin and now referred to as personality neuroscience[47]. By mapping across dog breeds in three separate cohorts, we implicated 127 genes at 90 loci for risk of diverse normal and problem behaviors. We mitigated population structure and latent variables in the association analysis by using multiple cohorts with different breed makeup. Many of our findings are quasi-replicated: supported by the same or related traits being mapped in another cohort here or in prior studies[6,7]. In the Supplementary Text, we discuss the strengths and weaknesses of canine interbreed genome scanning. The findings are supported by our unpublished study of pedigree and mixed breed dogs with individual-level genotype and C-BARQ phenotype data[18]. Further support comes from the geneset overlap between the combined interbreed C-BARQ studies and gene-based canine cognition GWASs using individual-level genotype and cognitive testing data[17] (hypergeometric

P=2.13x10^{-18}). While we can't rule out the possibility of some false-positive discovery, the biological relevance of our GWA geneset suggests most loci are likely valid – even those mapped in a single cohort (incl. behaviorally-relevant genes *SHISA6* and *SMOC2*; Suppl. Text).

This work takes the first step to move canine behavioral genetics into the realm of brain physiology. We showed several of the loci presently mapped in multiple cohorts in this study are associated with brain structure differences previously detected across dog breeds[19]. That is supported by geneset enrichment for human GWA genes mapped for differences in brain structure – which includes six priority genes mapped in multiple cohorts here. While geneset analyses ranked limbic regions highest, our brain structure association findings and deeper results of geneset analysis showed diverse other brain regions were also significant (incl. thalamic and cortical). In contrast to humans[48], complex genetic variations of dogs frequently have large effect sizes and thus have direct clinical and experimental utility[6,7,13].

It is widely accepted that domestication is dependent on reduced fear, and that this occurred in dogs[49]. It is also clear that a high range of body sizes was selected for during the creation of dog breeds. A great extent of haplotype sharing across breed groups is attributed to selection for morphological traits such as body size and coat color[50]. Historical documentation also describes breed admixture for selection of temperament in breed creation. However the genetic evidence of that is limited and confounded by correlations of behavioral and morphological traits[6,7]. Coat color and decreased sizes of brain and body are the most commonly reported traits in matched domesticated vs. wild animals, but they are not universal[51]. Belyaev's selection for tameness in foxes did not affect brain and body sizes. In contrast, Jensen and colleagues' eight-generation selection for tameness in Red Junglefowl – ancestors of domesticated chickens – resulted in increased body size and smaller brain relative to body weight ([52] and refs. within). There is thus strong evidence that selection for tameness does not necessarily affect body size, but that it can. Reduced fear was likely selected for in dog domestication, but it is unknown if

body size was also selected. In contrast, breed creation had strong purposed-selection for diverse body sizes and behavioral traits[14,15,49,50], and unintentionally selected for problem behaviors in smaller breeds[7,53-55].

The opposite body-size biases of the two canine behavioral GWA genesets allowed us to stratify associated behaviors and biology (discussed in Suppl. Text). The two dog studies were supported by UK Biobank GWA relevance; and our phenome analyses allowed us to isolate body size effects and show the top traits shared by the two dog GWASs were unlikely to be due to their genetic correlations with human height. However, we cannot rule that out due to the different genetic architectures of body size in the two species, and the known association of growth signaling and blood pressure (incl. human *IGF1* variation[56]). Canine body size is known to be strongly correlated with behavior based on behavioral and genetic evidence[6,7,53-55]. For instance, analysis of 32,000 C-BARQ-phenotyped dogs from 82 breeds showed that behavioral clustering of breeds was explained more by body size than breed relatedness[55]. In humans, height is genetically weakly-correlated with brain traits including neuroticism, risk tolerance and smoking cessation[57-59]. As intracranial volume is correlated with body size in dogs[19], it is with height in humans (phenotypic r=0.55; genetic r=0.26)[60]. Human brain volume is also genetically correlated with intelligence (r_g=0.24) and the PI3K growth signaling pathway was implicated for both these traits[60]. That supports the interpretation that correlations of canine behavior and body size are due to pleiotropy and not population structure (Suppl. Text). This is also supported by the significant overlap of mapped genes and behavioral relevance between our GWASs and MacLean et al.'s controlling for body size[6].

While both genesets were associated most strongly and consistently with the subiculum, ours weakly favored ventral aspects and the body size-controlled geneset predominantly implicated dorsal. The next-most implicated areas were hypothalamus and amygdala in ours and striatum in the body size-controlled geneset. The unique associations of tobacco and alcohol traits with the size-controlled geneset could be biological support for striatum relevance, but both those findings could simply reflect

the much higher power of that geneset. The top anatomical rankings for our geneset seem consistent with defense from threatening conspecifics and predators. We are not aware of human GWASs of those traits and did not detect signal for possibly overlapping traits such as anxiety. One UK Biobank trait uniquely associated with our geneset uncorrected for body size was household income, which is known to be associated with both genetics of intelligence and the impact of one's height on others that affects socioeconomic status[61]. Although the income signal was based on only four unlinked genes and could thus be spurious, all four genes in humans are associated with both height and intelligence or educational attainment, and are highly pleiotropic (median 122 ± 35 associations and 40 ± 22 traits (SD)). It is interesting that limbic regions and striatum are most highly implicated by our dog behavioral mapping. However, it is also possible other regions have similar effects, but rank lower due to a reduced density of relevant cells. Further studies are necessary to determine how robust and meaningful our anatomical enrichment findings are.

We also found associations with reproduction. Only the geneset uncorrected for body mass was enriched for uterus expression. This is consistent with the strong association of cattle stature with human height genetics and uterus gene expression[62], and with the bidirectional association of endometriosis and diverse psychiatric traits in women[63]. Body and litter sizes are strongly correlated in wolves and dogs, hinting the same genetic network could be involved in determining body, uterus and brain sizes and functions[64,65]. Mendelian Randomization studies in humans show causal relationships between neurobiological and puberty traits, and BMI[66,67]. Consistent with that, the higher-powered geneset of canine behavioral GWA corrected for body mass showed enrichment for puberty traits and BMI. In humans and wolves/dogs, reproductive traits are also correlated with both external pigmentation and longevity[64,65,68]. Body-wide analysis of gene expression for age of menarche GWA in humans was only significant for five brain tissues, three of which were prioritized in this work: anterior cingulate, hypothalamus and pituitary[67]. That study implicated 16 transcription factors. Of those, three

were mapped for dog behavior (*ELF1, NR2C2, MSX1*) or a close paralog was (human NR2F1 and SMAD3 implicated; dog *NR2F2* and *SMAD2/4* mapped); and two were significantly enriched in binding site predictions for both canine genesets (GATA, MSX1).

In addition to human genetics, the canine behavioral GWASs with and without control for body size were similarly supported by diverse geneset analyses (Suppl. Text). Both were most strongly enriched for the biological pathways axon guidance and morphogenesis, and gene annotation for alternative splicing and disease mutation. Analyses of evolutionary and genetic demographics show, or for some suggest, enrichment of genes that are pleiotropic, loss of function-intolerant, and under positive selection – including in humans. More specifically, there is also enrichment for human regions of highly accelerated evolution and containing human-specific enhancers that are active in neural progenitors.

Neurogenesis was strongly implicated in our analyses of dog GWA genes in both scRNA-seq data from juvenile mouse hypothalamus and for open chromatin data from embryonic human neocortex. Several transcription factor genes prioritized in this work are also known to have roles in adult neurogenesis, such as *Pou6f1*, *Pou3f2* and *Pax6* (Refs. [69-71]). Evidence for neurogenesis networks includes *Pax6*, which regulates the mapped gene *Cux1*, which in turn regulates the mapped gene *Pou6f2*[72,73]. There was strong overlap of neurogenesis and angiogenesis in the same hypothalamic scRNA-seq clusters, consistent with their intimate association[74]. That suggests a possible pathophysiological effect for our candidate behavioral gene *ANGPT1*. Among the possible mechanisms linking body size and neurodevelopment, Fox-family transcription factor binding sites were ranked near the top for the combined dog behavioral geneset. Fox-family members, including *FOXO3* and *FOXA2* nominated here, have key roles in growth/IGF1-signaling, longevity and neurogenesis[75].

Our scRNA-seq findings are supported by studies that used human GWA genesets from diverse traits to identify relevant cells in scRNA-seq data[76]. For instance, educational attainment and

schizophrenia showed a Spearman rank correlation of cell-type association of 0.94 because both were enriched for telencephalon projecting excitatory neurons out of the 39 cell types queried, and both associations were enriched for neurogenesis and synaptic processes genes. Other analyses in that work showed relevance of both excitatory and inhibitory neurons for those traits as well as bipolar disorder and BMI, whereas neural progenitors and neuroblasts were prioritized for intracranial volume (shown there to be associated with height) and depression, respectively. Those findings seem to stratify our enrichment analyses, which implicate the broadest definition of neurogenesis from neural stem cells to maturing neurons. In agreement with us, those investigators stated the shared biology of psychiatric and cognitive traits is consistent with a general psychopathology factor. However, they concluded this effect – and other neuron-type associations such as with BMI – is probably mediated at the level of shared cell types rather than genesets. We propose a more parsimonious interpretation (explained in Suppl. Text) is consistent with a heritable psychopathology factor[77,78]: that risk genes and their networks overlap across functional and structural brain traits. As in dogs, this network is associated with height[57-59,79], BMI[79] and lifespan[80] in humans. However, the large effect sizes in dogs seem likely to have resulted from body size selection under domestication. Further studies are necessary to assess that in other animals and in human "self-domestication".

In conclusion, genome scans of dog behavior implicate genes associated with human behavioral, cognitive and psychopathological traits. Both dog and human GWASs consistently and strongly implicate neurogenesis at, or near, the top of enriched biological pathways[6-8,25,57,59,81-86]. We propose that evolutionary adaptation of behavior in mammals, and probably vertebrates, preferentially targets a genetic network for neurogenesis and neurodevelopment throughout life. This could at least partially explain the molecular basis for a general genetic factor for human psychopathology that is correlated with personality and intelligence. Our findings suggest that network is integrated with energy metabolism, growth, longevity and reproduction. Accepting survival and reproduction are the

overarching priorities of natural selection, we believe the core evolutionary biology in play here is the balance of energy metabolism[87] and growth/development (Suppl. Text). In addition to several highly implicated transcription factors (incl. Fox and Hox subfamily members, NR2C2/F2, SMAD2/4 and MEF2A), there is supporting evidence to nominate mTOR and its regulation by the ancient LIN28/let-7 loop as a core mechanism of that pleiotropy (incl. growth factor signaling-PI3K-AKT upstream; Suppl. Text B.10). It is clear that connected and opposing traits must be balanced, and not difficult to imagine how tradeoffs could impact risks of psychiatric traits. While evolutionary selection and neurodevelopment have long been implicated in psychiatric genetics, we believe our integrated proposal to explain the "p" factor is novel (e.g., see review of "p" factor genetics[2] and discussion of possible mechanisms[78]). The cross-species conservation revealed here hints how genomics, neurophysiology and behavior could provide a scaffold for top-down descriptive and dimensional approaches to understanding the mind[88].

Methods

SNP datasets and phenotype data: Three previously published SNP datasets were used in this study. Since the breed average phenotype data was not available for all the breeds included in each of the datasets, only those for which phenotypes were available were kept. The Boyko et al. dataset contained 423 subjects from 37 breeds genotyped for ~45,000 SNPs (Affymetrix v.2 Canine array)[15]. The Vaysse et al. dataset contained 444 subjects from 29 dog breeds genotyped for ~175,000 SNPs (Illumina CanineHD array)[14]. The Hayward et al. dataset contained 2885 subjects from 50 dog breeds genotyped for ~160,000 SNPs (custom Illumina Canine HD array)[13]. Datasets were considered as independent, although the Vaysse dataset and the Hayward dataset used in this study share 192 subjects.

For the behavioral phenotype data, we used previously published C-BARQ data[7,89]. This data includes scores for the following problematic behaviors: stranger directed aggression, dog directed

aggression, owner directed aggression, dog rivalry, stranger oriented fear, dog oriented fear, nonsocial fear, touch sensitivity, separation-related anxiety, attachment and attention seeking, predatory chasing, excitability, energy, trainability, persistent barking, urination when left alone and escaping/roaming. For the additional traits of size and lifespan we used data compiled from several sources[14,15], the database curated by Dr. Kelly Cassidy at Washington State University available at http://users.pullman.com/lostriver/longhome.htm and from breed specifications published by the American Kennel Club (AKC) http://www.akc.org/ .

Genome Wide Association Analysis and gene annotation: Preparation of datasets, calculation of allele frequencies for each cohort, and removal of subjects with missing phenotypes were carried out in PLINK v.1.07(ref. [90]). To maximize reproducibility across different SNP platforms, no dataset trimming or LD clustering were performed on the SNP data. Association analysis were performed with GEMMA v.0.94.1 (ref. [16]). Population structure was removed by using the centered relatedness matrix correction. Association tests were performed using the univariate linear mixed model using the likelihood ratio test. Genome wide significance was based on Bonferroni adjusted *P*-value thresholds rounded down to the lowest order of magnitude: Vaysse dataset, 1×10^{-7}; Boyko dataset, 1×10^{-5}; Hayward dataset, 1×10^{-7}. The genomic inflation factor λ was calculated as the median χ^2(1 degree of freedom) association statistic observed across SNPs divided by the expected median under the null distribution[91]. Manhattan plots were generated with SAS v.9.4 from GEMMA outputs. Genome-wide significant hits were mapped on the CanFam3.1 genome assembly, UCSC Genome Browser[92] . Gene annotation was evaluated using the Broad Institute improved annotation V.1 track hub[93] and human assemblies hg19/39.

Effect direction and variance contribution regression modelling: To estimate effect direction and to determine the amount of variability that could be attributed to each relevant loci, we performed a regression modeling using a stepwise selection method. Since all hits detected were included, we performed a clustering step using Eigen decomposition on the first two dimensions. This decomposition

allowed us to select single markers to represent a cluster and remove all the correlated markers that would have caused collinearity issues in the model. Each cohort was evaluated independently. All this analyzes were performed on SAS v.9.4.

Brain-structure genetic associations: T2-weighted magnetic resonance imaging (MRI) morphometry data was derived from combined transversely- and sagittally-acquired images (using a 3.0 T MRI unit)[19]. Gray matter differences across breeds and sexes were measured as the degree of warping of per-subject maps to align with group-specific templates normalized by total brain volume (which was strongly correlated with body size). Independent Component (IC) analysis was performed using multivariate, source-based morphometry (GIFT software package for MATLAB[94])[19]. IC loadings for each individual were evaluated though regression modelling using a stepwise selection method against the same relevant loci list used to detect effect direction and variance contribution. Regression modeling was performed on SAS v.9.4.

Hypergeometric *P* value calculation: Of the 19,320 official protein coding genes in humans[34], we found 16,080 to have a mouse ortholog. We arbitrarily used 80% of the latter number – 12,864 – as a conservative estimate of the number of genes likely to be expressed in the brain or to be represented across different mammalian species and types of datasets used here. Calculations were performed on SAS v.9.4.

Geneset enrichment analyses: Geneset enrichment analyses were performed with Bonferroni multiple testing corrections on the Enrichr and DAVID algorithm and data servers[35,36]. Custom datasets were created for some enrichment analyses as follows. The hypothalamic scRNA-seq dataset was comprised of the top 2,000 genes, according to rank-sup *P* value, for each of the 62 neuron clusters reported[40]. Human genetics and evolutionary selection/genomic demographic genesets in Figure 3B and Table 5 are provided with sources in Supplementary Tables S39, S40. Hypergeometric testing (previous

section) was used to calculate *P* values for those, and Bonferroni multiple testing thresholds were provided.

Acknowledgements

We thank Marc Kent for providing opportunistically-collected brain imaging data from dogs, as described in ref. [19]. This work was made possible by high quality and publically available canine data. For that we thank the groups led by Adam Boyko[13,15] and Evan MacLean and Noah Snyder-Macklerand[6], and the consortium led by Matthew Webster[6,14]. We thank the dog owners who contributed with biological samples and those who contributed to the C-BARQ database of behavioral phenotypes. We thank Dr. Kim McBride for his technical advice for designing this study. We thank the Stanton Foundation (Cambridge, MA) and Elisabeth Allison for their support.

Competing Interests

The authors declare that they have no competing interests.

Authors Contributions

C.E.A. and I.Z. designed the study. I.Z. led the design of, and executed, the canine genetic analyses. C.E.A. led the design of, and performed, the comparative genetics and bioinformatics analyses, and the biological interpretations. J.A.S. collected and processed the C-BARQ data, and provided the behavioral expertise and, and E.E.H. collected and processed the brain imaging data, and provided that expertise. All authors contributed to the interpretation of results. C.E.A. drafted the manuscript with major contributions from I.Z., and participation from J.A.S. and E.E.H. All authors approved the final manuscript version.

References

1. Krueger, R.F. *et al.* Progress in achieving quantitative classification of psychopathology. *World Psychiatry* **17**, 282-293 (2018).
2. Smoller, J.W. *et al.* Psychiatric genetics and the structure of psychopathology. *Mol Psychiatry* **24**, 409-420 (2019).
3. Lahey, B.B. *et al.* Is there a general factor of prevalent psychopathology during adulthood? *J Abnorm Psychol* **121**, 971-7 (2012).
4. Oltmanns, J.R., Smith, G.T., Oltmanns, T.F. & Widiger, T.A. General Factors of Psychopathology, Personality, and Personality Disorder: Across Domain Comparisons. *Clin Psychol Sci* **6**, 581-589 (2018).
5. Harden, K.P. *et al.* Genetic associations between executive functions and a general factor of psychopathology. *Journal of the American Academy of Child & Adolescent Psychiatry* **59**, 749-758 (2020).
6. MacLean, E.L., Snyder-Mackler, N., vonHoldt, B.M. & Serpell, J.A. Highly heritable and functionally relevant breed differences in dog behaviour. *Proc Roy Soc Biol Sci* **286**, 20190716 (2019).
7. Zapata, I., Serpell, J.A. & Alvarez, C.E. Genetic mapping of canine fear and aggression. *BMC Genomics* **17**, 572 (2016).
8. Cross-Disorder Group of the Psychiatric Genomics Consortium. Electronic address, p.m.h.e. & Cross-Disorder Group of the Psychiatric Genomics, C. Genomic Relationships, Novel Loci, and Pleiotropic Mechanisms across Eight Psychiatric Disorders. *Cell* **179**, 1469-1482 e11 (2019).
9. Fenger, J.M., Rowell, J. L., Zapata, I., Kisseberth, W. C., London, C. A., Alvarez, C. E. Dog models of naturally occurring cancer. in *Animal Models for Human Cancer*, Vol. 69 (ed. Martic-Kehl, M.I., Mannhold, R., Kubinyi, H., Folkers, G.) 153-221 (Wiley-VCH Verlag GmbH & Co. KGaA, Weinheim, Germany, 2016).
10. Ostrander, E.A., Wayne, R.K., Freedman, A.H. & Davis, B.W. Demographic history, selection and functional diversity of the canine genome. *Nat Rev Genet* **18**, 705-720 (2017).
11. Jones, P. *et al.* Single-Nucleotide-Polymorphism-Based Association Mapping of Dog Stereotypes. *Genetics* **179**, 1033-1044 (2008).
12. Serpell, J.A. & Duffy, D.L. Dog breeds and their behavior. in *Domestic dog cognition and behavior* 31-57 (Springer, 2014).
13. Hayward, J.J. *et al.* Complex disease and phenotype mapping in the domestic dog. *Nat Commun* **7**, 10460 (2016).
14. Vaysse, A. *et al.* Identification of Genomic Regions Associated with Phenotypic Variation between Dog Breeds using Selection Mapping. *PLoS Genet* **7**, e1002316 (2011).
15. Boyko, A.R. *et al.* A Simple Genetic Architecture Underlies Morphological Variation in Dogs. *PLoS Biol* **8**, e1000451 (2010).
16. Zhou, X. & Stephens, M. Genome-wide efficient mixed-model analysis for association studies. *Nat Genet* **44**, 821-824 (2012).
17. Gnanadesikan, G.E. *et al.* Breed differences in dog cognition associated with brain-expressed genes and neurological functions. *Integr Comp Biol* (2020).
18. Zapata, I., Lilly, M.L., Herron, M.E., Serpell, J.A. & Alvarez, C.E. Genetic testing of dogs predicts problem behaviors in clinical and nonclinical samples. *bioRxiv DOI:10.1101/2020.08.13.249805 (non-reviewed preprint)* (2020).
19. Hecht, E.E. *et al.* Significant Neuroanatomical Variation Among Domestic Dog Breeds. *J Neurosci* **39**, 7748-7758 (2019).

20. Satizabal, C.L. *et al.* Genetic architecture of subcortical brain structures in 38,851 individuals. *Nat Genet* (2019).
21. Zhao, B. *et al.* Genome-wide association analysis of 19,629 individuals identifies variants influencing regional brain volumes and refines their genetic co-architecture with cognitive and mental health traits. *Nat Genet* **51**, 1637-1644 (2019).
22. Lyu, B. *et al.* Genetic signatures of human brain structure: A comparison between GWAS and relatedness-based regression. *bioRxiv*, 2020.08.07.239103 (2020).
23. Dreger, D.L. *et al.* Atypical Genotypes for Canine Agouti Signaling Protein Suggest Novel Chromosomal Rearrangement. *Genes (Basel)* **11**(2020).
24. Kichaev, G. *et al.* Leveraging Polygenic Functional Enrichment to Improve GWAS Power. *Am J Hum Genet* **104**, 65-75 (2019).
25. Lee, J.J. *et al.* Gene discovery and polygenic prediction from a genome-wide association study of educational attainment in 1.1 million individuals. *Nat Genet* **50**, 1112-1121 (2018).
26. Hirose, M. *et al.* CRISPR/Cas9-mediated genome editing in wild-derived mice: generation of tamed wild-derived strains by mutation of the a (nonagouti) gene. *Sci Rep* **7**, 42476 (2017).
27. Albert, F.W. *et al.* Genetic architecture of tameness in a rat model of animal domestication. *Genetics* **182**, 541-54 (2009).
28. Carola, V. *et al.* Modulation of social behavior by the agouti pigmentation gene. *Front Behav Neurosci* **8**, 259 (2014).
29. Hayssen, V. Effects of the nonagouti coat-color allele on behavior of deer mice (Peromyscus maniculatus): a comparison with Norway rats (Rattus norvegicus). *J Comp Psychol* **111**, 419-23 (1997).
30. Price, E.O. *Animal domestication and behavior*, (Cabi, 2002).
31. Vage, D.I. *et al.* A non-epistatic interaction of agouti and extension in the fox, Vulpes vulpes. *Nat Genet* **15**, 311-5 (1997).
32. Amin, N. *et al.* A rare missense variant in RCL1 segregates with depression in extended families. *Mol Psychiatry* **23**, 1120-1126 (2018).
33. Okbay, A. *et al.* Genetic variants associated with subjective well-being, depressive symptoms, and neuroticism identified through genome-wide analyses. *Nat Genet* **48**, 624-33 (2016).
34. Braschi, B. *et al.* Genenames.org: the HGNC and VGNC resources in 2019. *Nucleic Acids Res* **47**, D786-D792 (2019).
35. Kuleshov, M.V. *et al.* Enrichr: a comprehensive gene set enrichment analysis web server 2016 update. *Nucleic Acids Res* **44**, W90-7 (2016).
36. Huang da, W., Sherman, B.T. & Lempicki, R.A. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nat Protoc* **4**, 44-57 (2009).
37. Watanabe, K. *et al.* A global overview of pleiotropy and genetic architecture in complex traits. *Nat Genet* **51**, 1339-1348 (2019).
38. Richardson, T.G., Harrison, S., Hemani, G. & Davey Smith, G. An atlas of polygenic risk score associations to highlight putative causal relationships across the human phenome. *Elife* **8**(2019).
39. Sayed-Tabatabaei, F.A. *et al.* Heritability of the function and structure of the arterial wall: findings of the Erasmus Rucphen Family (ERF) study. *Stroke* **36**, 2351-6 (2005).
40. Romanov, R.A. *et al.* Molecular interrogation of hypothalamic organization reveals distinct dopamine neuronal subtypes. *Nat Neurosci* **20**, 176-188 (2017).
41. Lee, J.E., Wu, S.F., Goering, L.M. & Dorsky, R.I. Canonical Wnt signaling through Lef1 is required for hypothalamic neurogenesis. *Development* **133**, 4451-61 (2006).
42. Hall, B.J., Ripley, B. & Ghosh, A. NR2B signaling regulates the development of synaptic AMPA receptor current. *J Neurosci* **27**, 13446-56 (2007).

43. de la Torre-Ubieta, L. *et al.* The Dynamic Landscape of Open Chromatin during Human Cortical Neurogenesis. *Cell* **172**, 289-304 e18 (2018).
44. Lek, M. *et al.* Analysis of protein-coding genetic variation in 60,706 humans. *Nature* **536**, 285-91 (2016).
45. Moretti, S. *et al.* Selectome update: quality control and computational improvements to a database of positive selection. *Nucleic Acids Res* **42**, D917-21 (2014).
46. Doan, R.N. *et al.* Mutations in Human Accelerated Regions Disrupt Cognition and Social Behavior. *Cell* **167**, 341-354 e12 (2016).
47. Davis, K.L. & Panksepp, J. *The emotional foundations of personality: A neurobiological and evolutionary approach*, (WW Norton & Company, 2018).
48. Maier, R.M., Visscher, P.M., Robinson, M.R. & Wray, N.R. Embracing polygenicity: a review of methods and tools for psychiatric genetics research. *Psychol Med* **48**, 1055-1067 (2018).
49. Miklósi, Á. *Dog behaviour, evolution, and cognition*, (OUP Oxford, 2014).
50. Parker, H.G. *et al.* Genomic Analyses Reveal the Influence of Geographic Origin, Migration, and Hybridization on Modern Dog Breed Development. *Cell Rep* **19**, 697-708 (2017).
51. Lord, K.A., Larson, G., Coppinger, R.P. & Karlsson, E.K. The History of Farm Foxes Undermines the Animal Domestication Syndrome. *Trends Ecol Evol* **35**, 125-136 (2020).
52. Katajamaa, R. & Jensen, P. Tameness correlates with domestication related traits in a Red Junglefowl intercross. *Genes Brain Behav*, e12704 (2020).
53. McGreevy, P.D. *et al.* Dog Behavior Co-Varies with Height, Bodyweight and Skull Shape. *PLoS ONE* **8**, e80529 (2013).
54. Puurunen, J. *et al.* Inadequate socialisation, inactivity, and urban living environment are associated with social fearfulness in pet dogs. *Sci Rep* **10**, 3527 (2020).
55. Wilson, B., Serpell, J., Herzog, H. & McGreevy, P. Prevailing Clusters of Canine Behavioural Traits in Historical US Demand for Dog Breeds (1926(-)2005). *Animals (Basel)* **8**(2018).
56. Schutte, A.E., Volpe, M., Tocci, G. & Conti, E. Revisiting the relationship between blood pressure and insulin-like growth factor-1. *Hypertension* **63**, 1070-7 (2014).
57. Nagel, M. *et al.* Meta-analysis of genome-wide association studies for neuroticism in 449,484 individuals identifies novel genetic loci and pathways. *Nat Genet* **50**, 920-927 (2018).
58. Karlsson Linnér, R. *et al.* Genome-wide association analyses of risk tolerance and risky behaviors in over 1 million individuals identify hundreds of loci and shared genetic influences. *Nature Genetics* (2019).
59. Liu, M. *et al.* Association studies of up to 1.2 million individuals yield new insights into the genetic etiology of tobacco and alcohol use. *Nat Genet* (2019).
60. Jansen, P.R. *et al.* Genome-wide meta-analysis of brain volume identifies genomic loci and genes shared with intelligence. *Nat Commun* **11**, 5606 (2020).
61. Tyrrell, J. *et al.* Height, body mass index, and socioeconomic status: mendelian randomisation study in UK Biobank. *BMJ* **352**, i582 (2016).
62. Fang, L. *et al.* Comprehensive analyses of 723 transcriptomes enhance genetic and biological interpretations for complex traits in cattle. *Genome Res* **30**, 790-801 (2020).
63. Gao, M. *et al.* Psychiatric comorbidity among women with endometriosis: nationwide cohort study in Sweden. *Am J Obstet Gynecol* (2020).
64. Stahler, D.R., MacNulty, D.R., Wayne, R.K., vonHoldt, B. & Smith, D.W. The adaptive value of morphological, behavioural and life-history traits in reproductive female wolves. *J Anim Ecol* **82**, 222-34 (2013).
65. Borge, K.S., Tonnessen, R., Nodtvedt, A. & Indrebo, A. Litter size at birth in purebred dogs--a retrospective study of 224 breeds. *Theriogenology* **75**, 911-9 (2011).

66. Day, F.R. *et al.* Physical and neurobehavioral determinants of reproductive onset and success. *Nat Genet* **48**, 617-623 (2016).
67. Day, F.R. *et al.* Genomic analyses identify hundreds of variants associated with age at menarche and support a role for puberty timing in cancer risk. *Nat Genet* **49**, 834-841 (2017).
68. Hollis, B. *et al.* Genomic analysis of male puberty timing highlights shared genetic basis with hair colour and lifespan. *Nat Commun* **11**, 1536 (2020).
69. McClard, C.K. *et al.* POU6f1 Mediates Neuropeptide-Dependent Plasticity in the Adult Brain. *J Neurosci* **38**, 1443-1461 (2018).
70. Hashizume, K., Yamanaka, M. & Ueda, S. POU3F2 participates in cognitive function and adult hippocampal neurogenesis via mammalian-characteristic amino acid repeats. *Genes Brain Behav* **17**, 118-125 (2018).
71. Ninkovic, J. *et al.* The BAF complex interacts with Pax6 in adult neural progenitors to establish a neurogenic cross-regulatory transcriptional network. *Cell Stem Cell* **13**, 403-18 (2013).
72. Nieto, M. *et al.* Expression of Cux-1 and Cux-2 in the subventricular zone and upper layers II-IV of the cerebral cortex. *J Comp Neurol* **479**, 168-80 (2004).
73. Gray, L.T. *et al.* Layer-specific chromatin accessibility landscapes reveal regulatory networks in adult mouse visual cortex. *Elife* **6**(2017).
74. Kirby, E.D., Kuwahara, A.A., Messer, R.L. & Wyss-Coray, T. Adult hippocampal neural stem and progenitor cells regulate the neurogenic niche by secreting VEGF. *Proc Natl Acad Sci U S A* **112**, 4128-33 (2015).
75. Genin, E.C., Caron, N., Vandenbosch, R., Nguyen, L. & Malgrange, B. Concise review: forkhead pathway in the control of adult neurogenesis. *Stem Cells* **32**, 1398-407 (2014).
76. Bryois, J. *et al.* Genetic identification of cell types underlying brain complex traits yields insights into the etiology of Parkinson's disease. *Nat Genet* (2020).
77. Pettersson, E., Larsson, H. & Lichtenstein, P. Common psychiatric disorders share the same genetic origin: a multivariate sibling study of the Swedish population. *Mol Psychiatry* **21**, 717-21 (2016).
78. Rosenstrom, T. *et al.* Joint factorial structure of psychopathology and personality. *Psychol Med* **49**, 2158-2167 (2019).
79. Speed, M.S., Jefsen, O.H., Borglum, A.D., Speed, D. & Ostergaard, S.D. Investigating the association between body fat and depression via Mendelian randomization. *Transl Psychiatry* **9**, 184 (2019).
80. Marioni, R.E. *et al.* Genetic variants linked to education predict longevity. *Proc Natl Acad Sci U S A* **113**, 13366-13371 (2016).
81. Apple, D.M., Fonseca, R.S. & Kokovay, E. The role of adult neurogenesis in psychiatric and cognitive disorders. *Brain Res* **1655**, 270-276 (2017).
82. Luciano, M. *et al.* Association analysis in over 329,000 individuals identifies 116 independent variants influencing neuroticism. *Nat Genet* **50**, 6-11 (2018).
83. Turley, P. *et al.* Multi-trait analysis of genome-wide association summary statistics using MTAG. *Nat Genet* **50**, 229-237 (2018).
84. Savage, J.E. *et al.* Genome-wide association meta-analysis in 269,867 individuals identifies new genetic and functional links to intelligence. *Nat Genet* **50**, 912-919 (2018).
85. Hill, W.D. *et al.* A combined analysis of genetically correlated traits identifies 187 loci and a role for neurogenesis and myelination in intelligence. *Mol Psychiatry* **24**, 169-181 (2019).
86. Neumann, A. *et al.* Single Nucleotide Polymorphism Heritability of a General Psychopathology Factor in Children. *J Am Acad Child Adolesc Psychiatry* **55**, 1038-1045 e4 (2016).
87. Witting, L. The natural selection of metabolism and mass selects allometric transitions from prokaryotes to mammals. *Theoretical population biology* **117**, 23-42 (2017).

88. Marshall, M. The hidden links between mental disorders. *Nature* **581**, 19-21 (2020).
89. Serpell, J. & Duffy, D. Dog Breeds and Their Behavior. in *Domestic Dog Cognition and Behavior* (ed. Horowitz, A.) 31-57 (Springer Berlin Heidelberg, 2014).
90. Purcell, S. *et al.* PLINK: A Tool Set for Whole-Genome Association and Population-Based Linkage Analyses. *American journal of human genetics* **81**, 559-575 (2007).
91. Price, A.L., Zaitlen, N.A., Reich, D. & Patterson, N. New approaches to population stratification in genome-wide association studies. *Nat Rev Genet* **11**, 459-463 (2010).
92. Kent, W.J. *et al.* The Human Genome Browser at UCSC. *Genome Research* **12**, 996-1006 (2002).
93. Hoeppner, M.P. *et al.* An Improved Canine Genome and a Comprehensive Catalogue of Coding Genes and Non-Coding Transcripts. *PLoS ONE* **9**, e91172 (2014).
94. Calhoun, V.D., Adali, T., Pearlson, G.D. & Pekar, J.J. A method for making group inferences from functional MRI data using independent component analysis. *Hum Brain Mapp* **14**, 140-51 (2001).

Table 1. Canine interbreed GWA top confidence findings with human GWA relevance implicated in this work

				Human GWA annotation, n traits (n total hits)														
Peak marker (CanFam3)	Dog behavioral, body size & lifespan GWA traits (number of present cohorts or refs. for published support)	Peak *P*	Candidate dog genes[1]	Human GWA hits (n)	Behavior & personality	Psychiatric	alcohol & drug use	Cognitive	Brain structure	socio-demographic	Height & related	BMI & related	hair & external pigmentation	Cardio-vascular	Blood	Reproductive	Notable other human GWA	Other canine GWA phenotypes
chr1:42030177	Excitability (MacLean et al. 2019: Energy, DogFear)[2]	2.39E-09	**ESR1**	127		2(2)	2(2)	1(5)		2(5)	8(23)	3(5)		5(12)				Long legs, PMID: 30940804
chr1:60068123	Urination (MacLean: NonsocialFear)[2]	1.4832E-08	**TBC1D32** (& 1 "LOC#")	3				1(1)									Focal epilepsy	
chr1:62730770	StrangerFear, DogFear, Urination, StrangerAggression (MacLean et al. 2019: DogAggression)[2]	1.23E-09	**TRDN**, **NKAIN2**	12; 29	0; 1(1)	1(1); 3(3)	1(1); 2(2)	0; 2(1)	0; 2(2)	0; 1(1)		1(1); 1(3)		1(1); 1(1)		0; 1(1)	Amyloid depos., intraocular press.; sleep (2), Alzheimer's d.	
chr1:93,219,668	SeparationAnxiety, DogFear, Chasing[3] (MacLean et al. 2019: SeparationAnxiety)[2]	3.41E-10	**RCL1**, **JAK2**	43; 48	0; 1(1)						0; 1(1)				10(37); 9(17)		Alzheimer's disease	RCL1: litter size, PMID: 31263560; tail curl, 22022279
chr3:91114590	Excitability, Size, Lifespan[3] (Vaysse et al. 2011 and others: Size)[2]	4.88E-16	**LCORL**, **SLIT2**	111; 21	1(1); 0	1(1); 0	1(1); 1(1)	5(11); 0	0; 1(1)		12(62); 0	9(14); 3(8)		2(4); 2(3)		0; 3(3)	Sleep/chrono. (3), refract. Err., Parkinson's. d., Alzheimer's d.	LCORL: skull morphology, PMID: 28552356
chr4:39404339	Size (Boyko et al. 2010 and others: Size; MacLean et al. 2019: Chasing)[2]	8.66E-08	**STC2**	9					1(1)		3(4)	2(3)						Skull morphology, PMID: 28552356
chr5:35635596	Escaping, Chasing (MacLean et al. 2019: TouchSensitivity)[2]	2.52E-06	**SHISA6**	19	1(1)							1(1)		1(1)			Refractive error (9), sleep (3)	Skull shape, PMID: 28552356
chr7:43842285	Size (2); Lifespan	4.04E-13	**SMAD2**	7							1(1)	1(1)		2(2)	1(1)	1(1)		
chr7:46745071	OwnerAggression, Size, NonsocialFear, Chasing, Escaping (MacLean et al. 2019: OwnerAggression)[2]	1.69E-14	LINC01478, **SETBP1**	44			2(3)	2(1)			2(3)	3(6)	3(8)	3(10)	1(1)			(Under diversifying selection, PMID: 27233669)
chr10:2453907	TouchSensitivity (2), DogAggression (2), StrangerAggression, NonsocialFear, Exciability, Attachment, Urination, Size, Lifespan	3.21E-11	LRIG3, **ATP23** (& 3 "LOC#" genes)	4; 3				0; 3(1)						1(1); 0	1(1); 0		Alzheimer's disease; none	
chr10:8183593	Attach. (3), Barking (2), Chasing (2), DogAggress. (2), Escaping (3), Excitabil. (3), Lifespan (3), Nonsoc.Fear (2), OwnerAggress. (2), RivalryAggress. (2), Separat.Anxiety (2), Size (3), StrangerAggress. (2), TouchSensitiv. (3), Urinat. (2), Trainabil.	2.59E-39	**MSRB3**, **HMGA2**, LOC100507065	73; 108	2(2); 1(2)			0; 3(3)	12(49); 1(5)		1(2); 10(50)	1(1); 2(4)		0; 2(6)	0; 2(2)		Snoring (5); Alzheimer's disease, sleep	MSRB3: litter size, PMID: 31263560
chr11:26929946	DogAggression, Excitability, Size, DogFear, TouchSensitivity (MacLean et al. 2019: DogAggression)[2]	1.42E-09	**KIAA2026**	12					1(1)				2(2)					
chr12:21490122	Trainability (MacLean et al. 2019: Chasing)[2]	1.67E-08	LOC101927189, **MLIP**	15		1(1)					1(1)	2(2)		3(2)				
chr13:6822452	StrangerAggression (MacLean et al. 2019: Attachment, DogAggression, DogFear, DogRivalry)[2]	2.29E-06	ZFPM2-AS1, **ZFPM2**	99	3(3)	4(5)	1(1)	4(5)			1(3)	1(1)		10(20)	8(13)		Hearing impairment, glaucoma	
chr13:8391652	Excitability (2), NonsocialFear (2), Size (2), DogAggression, SeparationAnxiety, Attachment, Lifespan	3.09E-18	**ANGPT1**, RSPO2	31; 20	1(1); 0			1(1); 0			0; 2(2)	1(1); 0	0; 3(10)		4(4); 0		Intraoc. press./glauc. (17); refract. err.; Macular thickness	
chr15:41221438	Barking (2), DogFear (2), Excitabil. (3), OwnerAggress. (3), RivalryAggress. (2), Size (2), TouchSensitiv. (2), Urinat. (3), Attach., DogAggress., Lifespan, Nonsoc.Fear, SeparatAnxiety	3.06E-19	**IGF1**	23		1(1)			2(9)		1(3)	3(3)		1(1)			Refractive error	
chr18:20272961	Barking (2), DogAggression (2), DogFear (3), NonsocialFear (3), OwnerAggression (2), SeparationAnxiety (2), Size (3), StrangerAggression (3), StrangerFear (3), TouchSensitivity (3), Trainability (2), Attach., Chasing, Excitability, Lifespan, Urinat.	7.67E-43	**GNAT3**[5], **CD36**[5]	1; 39		0; 1(1)		0; 1(2)	1(1); 0		0; 1(1)	0; 1(1)			0; 6(15)		Hearing impairment; Hearing impairment	
chr20:21848176	Escaping (2), OwnerAggression (2), Chasing, Escaping, SeparationAnxiety, Urination	2.67E-15	**MITF**, **FOXP1**	19; 128	0; 4(8)	0; 4(9)	0; 4(10)	2(4); 8(20)	1(1); 2(2)	0; 4(4)	0; 1(4)	1(2); 3(4)	2(2); 2(2)	1(1); 3(3)	4(5); 6(9)		None; Sleep/chron. (5), oc. press. (5), refract. err., migraine	
chr20:37609601	Chasing (MacLean et al. 2019: Energy, Trainability)[2]	7.34E-08	**POC1A**	4							3(3)						SOFT syn., PMID: 22840363/4	
chr23:8387680	Urination (MacLean et al. 2019: Urination, DogFear, SeparationAnxiety)[2]	2.66E-10	SCN5A, **SCN10A**	63; 60				1(1); 1(1)						11(62); 15(58)				
chr24:32707621	NonsocialFear (2), RivalryAggression (2), SeparationAnxiety (2), Chasing, Energy, OwnerAggression, TouchSensitivity, Urination	2.12E-10	(EIF2S2[5]), **ASIP**, RALY	10; 38	0;1(1)			0;1(1)			0; 2(2)	0;2(4)	7(7); 13(13)	0; 1(1)	0; 1(1)		Alzheimer's D.; Myopia	Gestation length, PMID: 31263560
chr34:18559537	DogFear (1)[3]; DogAggression (1)[3], Excitability, Escaping, Excitability, DogAggression	2.95E-09	**IGF2BP2**, TRA2B				2(2); 2(2)	1(1); 0			5(18); 5(13)					0; 1(4)		
chrX:102927995	DogFear (2), SeparationAnxiety (2), Size (2), Barking, Chasing, Excitability, Lifespan, NonsocialFear, StrangerFear, TouchSensitivity	8.98E-13	**IGSF1**[4], (FIRRE, ARHGAP36)[6], STK26	3; 1	1(1); 0											2(2); 0		
chrX:104218823	Size (Shoenebeck & Ostrander 2014, PMID:25062362: Size; MacLean et al. 2019: Energy)[2]	5.92E-08	**GPC4** (may be in LD with preceeding)	none														

Dog genes with human protein coding orthologs were annotated for human GWA relevance using the NHGRI-EBI GWAS Catalog (none had such evidence for longevity traits). References cited in main article text given as name and year, and others given as PubMed ID

[1]Bold denotes favored by known biology and underlining denotes higher positional proximity; only annotated HUGO HGNC protein coding genes were evaluated for human GWA relevance

[2]Same haplotype associated with different traits in different cohorts

[3]Same gene mapped for behavior in this study and in a different cohort in Zapata et al. 2016 or MacLean et al. 2019 (or for size in refs. cited in this work)

[4]IGSF1 was shown to have the causative variant for size at this locus and is good candidate for behavioral traits (PMID: 28257443)

[5]EIF2S2 has two human GWA hits total: BMI, skin pigmentation

[6]FIRRE is a lncRNA; ARHGAP36 had no human GWA hits

Table 2. Pathway & tissue enrichment of canine behavioral GWA

	This study		Body-mass corrected[3]	
	Rank	Bonferroni *P*	Rank	Bonferroni *P*
Biol. Pathways (KEGG, hum)[1]				
Axon guidance	1	**1.76E-02**	1	**6.69E-04**
Biol. Pathways (gene ontology, GO)[2]				
Embryo development	1	**4.32E-03**	184	1.00E+00
Any term re:morphogenesis	1	**4.32E-03**	1	**1.02E-06**
Any term containing "epithelial"	3	7.84E-02	37	2.80E-01
Any term containing "proliferation"	3	7.84E-02	202	1.00E+00
Any term containing "growth"	10	7.47E-01	50	**2.62E-02**
Any term re:neuronal development	54	1.00E+00	4	**2.40E-05**
Annotation: keyword up (Swiss-Prot)[2]				
Alternative splicing	1	**4.23E-05**	1	**9.79E-36**
Disease mutation	2	**4.05E-02**	5	**4.10E-04**
Annotation: tissue up (Uniprot)[2]				
Brain	1	9.97E-01	1	**1.12E-11**
Tissue expression (hum. BioGPS enriched)[1]				
Uterus	1	**1.22E-04**	10	2.03E-01
Prefrontal Cortex	7	4.81E-01	1	**1.80E-04**
Fetal Brain	54	1.00E+00	2	**4.96E-03**
Brain region up (mouse Allen Brain Atlas)[1]				
molecular layer of Subiculum	1	**1.19E-02**	17	**4.53E-03**
Subiculum	2	**2.69E-02**	7	**8.81E-04**
nucleus of the stria terminalis (Mv)	3	**2.73E-02**	858	6.77E-01
superficial stratum of Subiculum	4	**2.80E-02**	25	**1.33E-02**
parastrial preoptic nucleus	5	**2.96E-02**	1233	8.63E-01
mantle zone of THyB-D	6	**3.15E-02**	1479	9.22E-01
mantle zone of Subiculum	7	**3.36E-02**	27	**1.40E-02**
intermediate stratum of the VAP	8	**3.60E-02**	734	5.80E-01
dorsal part of THyB	9	**3.87E-02**	1678	9.52E-01
Ventromedial hypothalamic nucleus	10	**4.04E-02**	1209	8.54E-01
layer 5 of PaS	11	**4.09E-02**	84	**6.57E-02**
Ventromedial hypothalamic nuc. (C)	12	**4.19E-02**	613	4.75E-01
Uvula (IX)	13	**4.58E-02**	1968	9.87E-01
Subiculum, dorsal part	33	9.10E-02	1	**9.05E-05**
Subiculum, dorsal part, molecular layer	111	1.72E-01	2	**4.88E-04**
Subiculum, dorsal part, stratum radiatum	100	1.67E-01	3	**5.06E-04**
Striatum dorsal region	1435	8.77E-01	4	**6.07E-04**
Caudoputamen	1752	9.73E-01	5	**7.59E-04**
Field CA1	255	2.67E-01	6	**7.81E-04**
striatum (corpus striatum)	1055	7.00E-01	8	**9.09E-04**
mantle zone of Striatum	1283	7.82E-01	9	**1.01E-03**
Anterior cingulate area, ventral part, 6b	475	3.74E-01	10	**1.01E-03**
TF-KO gene expression changes[1]				
LMX1B	1	**1.64E-03**	19	**1.94E-02**
MECP2	2	**2.72E-03**	1	**1.35E-04**

[1]Enrichr analysis, [2]DAVID analysis, [3]MacLean et al. 2019

Terms significant in both studies are italicized and underlined; P_{Bonf} <0.05 are in bold

Table 3. Transcription factor binding site prediction

TF binding site	This study			Body-mass corrected[1]		
	Rank	Enr.	Bonf. *P*	Rank	Enr.	Bonf. *P*
POU6F1	**1**	2.17	2.08E-15	**2**	2.02	8.66E-82
POU2F* 1/*2	**2**	2.03	1.98E-14	**15**	1.86	9.03E-70
MEF2A	**3**	1.96	7.90E-13	**13**	1.87	9.75E-75
PRX2	**4**	2.00	1.46E-12	**10**	1.92	5.98E-76
FOXA2	**5**	2.08	5.49E-12	**4**	2.04	2.79E-80
FOXL1	**6**	1.90	6.94E-12	**12**	1.85	2.82E-75
FOXQ1	**7**	1.98	1.16E-11	28	1.85	1.02E-62
GATA1-6	**8**	1.97	1.30E-11	43	1.74	7.99E-48
LHX3	**9**	2.17	1.71E-11	**3**	2.15	2.26E-80
CUX1	**10**	1.68	1.81E-11	26	1.60	6.77E-63
TBP*	**11**	1.77	3.51E-11	**5**	2.18	4.76E-79
CDC5L	**12**	1.93	9.18E-11	**7**	1.94	1.69E-78
POU3F2	**13**	1.67	1.98E-10	**18**	1.63	5.83E-66
CHX10	**14**	2.00	2.07E-10	**8**	2.00	2.41E-76
FOXI1	**15**	2.01	3.02E-10	**16**	1.96	5.71E-67
NKX2-5	**16**	1.60	3.12E-10	54	1.46	4.69E-42
PBX1	**17**	1.67	2.06E-09	32	1.62	5.17E-59
GFI1	**18**	1.90	2.11E-09	46	1.73	1.99E-45
FOXD1	**19**	1.85	3.98E-09	38	1.77	8.64E-55
FOXD3	**20**	2.11	6.42E-09	**20**	2.11	1.36E-65
ALX1	48	1.71	1.29E-06	**1**	1.90	6.10E-82

[1]MacLean et al. 2019. Full DAVID TF prediction datasets are provided in supplementary tables (note some TFs are official gene symbol here, but alias in suppl. tables; *TATA and TBP refer to same factor but appear separately: ranked 11 and 44 in this study, and 25 and 5 in MacLean et al. 2019, respectively). TF binding sites significant in both studies are in bold. Genes mapped in canine GWAS's are italicized and underlined.

Table 4. Human genetics relevance of canine behavioral GWA

Sources and representative traits	This study		Body-mass corrected[1]	
	Rank	Bonf. *P*	Rank	Bonf. *P*
Human GWA catalog				
Hip circumference adjusted for BMI*	1	**1.89E-03**	146	3.01E-01
Height*	2	**4.64E-03**	108*	2.49E-01
Systolic blood pressure	3	**6.96E-03**	2	**5.45E-07**
Menarche (age at onset)	4	**8.61E-03**	18	6.13E-02
PR interval	5	**1.08E-02**	169	3.23E-01
Birth weight*	6	**1.27E-02**	105*	2.44E-01
Infant length*	7	**1.65E-02**	804*	9.20E-01
Heart rate response to recovery post exercise	8	**3.03E-02**	375	7.03E-01
QRS duration	9	**4.28E-02**	13	**2.55E-02**
Pulse pressure	10	**4.31E-02**	5	**3.19E-03**
Human GWA, UK Biobank				
Distance_between_home_and_job_workplace_796_raw	1	**1.45E-15**	2	**3.14E-50**
Pulse_wave_Arterial_Stiffness_index_21021_raw	2	**4.75E-09**	3	**2.93E-47**
High_light_scatter_reticulocyte_percentage_30290_raw	3	**3.73E-08**	1	**1.72E-57**
Forced_vital_capacity_20151_raw	4	**2.76E-07**	75	**1.02E-04**
Standing_height_50_raw	6	**1.38E-06**	60	**1.08E-05**
Forced_expiratory_volume_in_1-second_(best_measure)_20150…	10	**2.75E-06**	83	**1.96E-04**
Whole_body_fat-free_mass_23101_raw	15	**1.27E-05**	16	**7.34E-14**
Monocyte_count_30130_raw	17	**2.55E-05**	7	**4.07E-23**
Weight_23098_raw	19	**3.30E-05**	41	**1.36E-07**
Longest_period_of_unenthusiasm/disinterest_5375_raw	23	**3.65E-05**	10	**9.56E-19**
Freq...needing_morning_drink_of_alcohol_after_heavy_drinking…	26	**6.29E-05**	6	**1.46E-25**
Basal_metabolic_rate_23105_raw	29	**6.72E-05**	18	**8.50E-13**
Inverse_distance_to_the_nearest_major_road_24012_raw	34	**1.58E-04**	11	**1.76E-18**
Hip_circumference_49_raw	35	**1.61E-04**	54	**4.55E-06**
Birth_weight_20022_raw	42	**5.25E-04**	163	**2.80E-02**
Lymphocyte_count_30120_raw	45	**5.72E-04**	4	**3.83E-41**
Duration_to_first_press_of_snap-button_in_each_round_404_raw	51	**1.19E-03**	12	**1.88E-18**
Longest_period_of_depression_4609_raw	52	**1.48E-03**	29	**3.82E-09**
Peak_expiratory_flow_3064_raw	54	**1.92E-03**	119	**2.91E-03**
Red_blood_cell_count_30010_raw	58	**5.60E-03**	99	**9.32E-04**
Home_large_urban_area_scotland_20118_11	61	**8.40E-03**	138	**8.88E-03**
Whole_body_fat_mass_23100_raw	62	**9.76E-03**	43	**4.26E-07**
Average_total_household_income_before_tax_738	63	**1.32E-02**	271	2.20E-01
Age_first_had_sexual_intercourse_2139_raw	64	**1.35E-02**	39	**1.31E-07**
Age_cataract_diagnosed_4700_raw	68	**2.23E-02**	193	**5.55E-02**
North_co-ordinate_of_birthplace_in_UK_129_raw	71	**3.16E-02**	81	**1.73E-04**
Skin_colour_1717	72	**3.42E-02**	162	**2.70E-02**
Body_mass_index_(BMI)_23104_raw	73	**4.24E-02**	52	**3.55E-06**
Birth_weight_of_first_child_2744	75	**4.29E-02**	342	3.84E-01
Malignant_melanoma_of_skin_C3_MELANOMA_SKIN	76	**4.31E-02**	216	**8.43E-02**
Right_intra-ocular_pressure_corneal-compensated_5254_raw	77	**4.54E-02**	72	**6.98E-05**
QRS_duration_during_ECG_12340_raw	78	**4.90E-02**	194	**5.68E-02**
Longest_period_spent_worried_or_anxious_20420_raw	176	3.73E-01	35	**1.31E-08**
Worrier/anxious_feelings_1980	271	9.62E-01	70	**4.80E-05**
Neuroticism_score_20127_raw	289	1.00E+00	69	**3.65E-05**
Never_smoked_20116_0	293	1.00E+00	127	**5.29E-03**
Alcohol_intake_frequency_1558	290	1.00E+00	157	**2.25E-02**
Genotype-phenotype (dbGAP)				
Body Height	1	**2.61E-03**	10	**5.63E-08**
Heart Function Tests	2	**1.15E-02**	36	**1.23E-03**
Body Mass Index	5	1.76E-01	1	**1.13E-24**
Hip	4	6.92E-02	7	**3.65E-08**
Echocardiography	6	2.11E-01	8	**3.86E-08**
Body Weight	7	3.14E-01	5	**2.85E-11**

[1]MacLean et al. 2019. Enrichment analysis performed with Inrichr. Traits significant in both studies are italicized and underlined, and P_{Bonf} <0.05 are in bold and those trending underlined.All significant GWA catalog traits for this study's geneset were included here. For UK Biobank, a subset were selected as representative. The full analysis results are provided as supplementary tables.

Table 5. Adaptive evolution and genomic demographics

Genesets compiled from the literature in this study[1]	Dog behavior GWA, this study, n=108		Mass-corrected dog GWA[2], n=715		All genes, both studies, n=804	
	Observed / expected	under *P* or over *Q*	Observed / expected	under *P* or over *Q*	Observed / expected	*under P or over Q*
Positive selection, dog breeds sampled in North Am. & Europe, n=1842	1.88	**4.57E-04**	1.35	**1.00E-04**	1.38	**7.40E-06**
Positive selection, 15 Chinese indigenous dog breeds, n=963	1.11	4.20E-01	1.77	**1.87E-08**	1.64	**4.01E-07**
Positive selection in at least two of cattle, goat, pig & sheep, n=666	1.61	1.07E-01	1.27	5.34E-02	1.3	2.88E-02
Positive selection, human, n=1412	1.27	2.03E-01	1.38	**3.03E-04**	1.36	**2.32E-04**
Loss of function intolerant, human, n=3154	1.28	6.02E-02	1.36	**2.15E-08**	1.34	**2.42E-08**
Human accelerated divergence regions, n=1623	2.64	**1.76E-08**	1.85	**2.49E-16**	1.93	**2.45E-21**
Protein-coding positive selection, vertebrates, n=550	0.433	1.54E-01	0.85	2.22E-01	0.785	1.05E-01
Differential brain gene expression in 3 wild v. domest. mamm., n= 121	1.02	6.41E-01	1.34	2.30E-01	1.32	2.24E-01
Disease associated genes (ClinVar), n=3047	1.17	1.85E-01	0.87	*2.82E-02*	0.89	*3.52E-02*
Haploinsufficient disease genes (ClinGen), n=294	3.65	**8.14E-04**	1.04	4.69E-01	1.31	1.09E-01
Nearest gene to GWAS peaks (MacArthur Lab), n=6288	0.97	4.02E-01	1.03	2.21E-01	1.01	3.99E-01
Single-trait pan-GWA meta-analysis (Watanabe et al.), n=1968	0.48	*1.06E-02*	0.62	**1.97E-06**	0.59	**5.31E-08**
Top pleiotropic pan-GWA meta-analysis (Watanabe et al.), n=1968	1.27	1.43E-01	1.10	1.40E-01	1.10	1.23E-01

Hypergeometric *P* and *Q* are lower and upper cumulative distribution functions. Bold surpass Bonferroni significance for 26 tests, 1.92×10^{-03}; italicized surpass unadjusted threshold of 0.05.

[1]Gene lists and references provided in Supplementary Tables S39/40.

[2]MacLean et al. 2019.

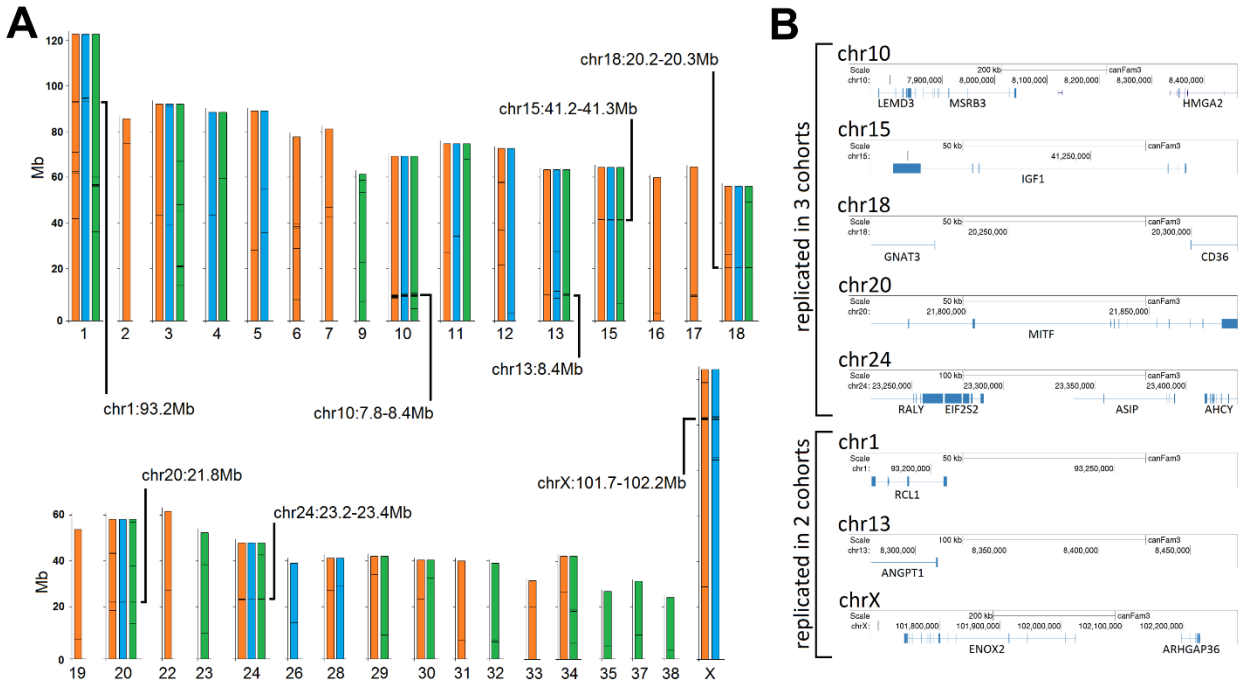


Figure 1. Canine behavioral mapping. (**A**) Summary of behavioral quantitative trait loci quasi-replicated in independent cohorts in this work. All loci were significant after Bonferroni adjustment. Only positive chromosomes are shown for each cohort: Vaysse et al. in orange, Boyko et al. in blue and Hayward et al. in green[13-15]. Genome coordinates are from CanFam3.1. (**B**) Mapped intervals and gene annotation for select quasi-replicated loci.

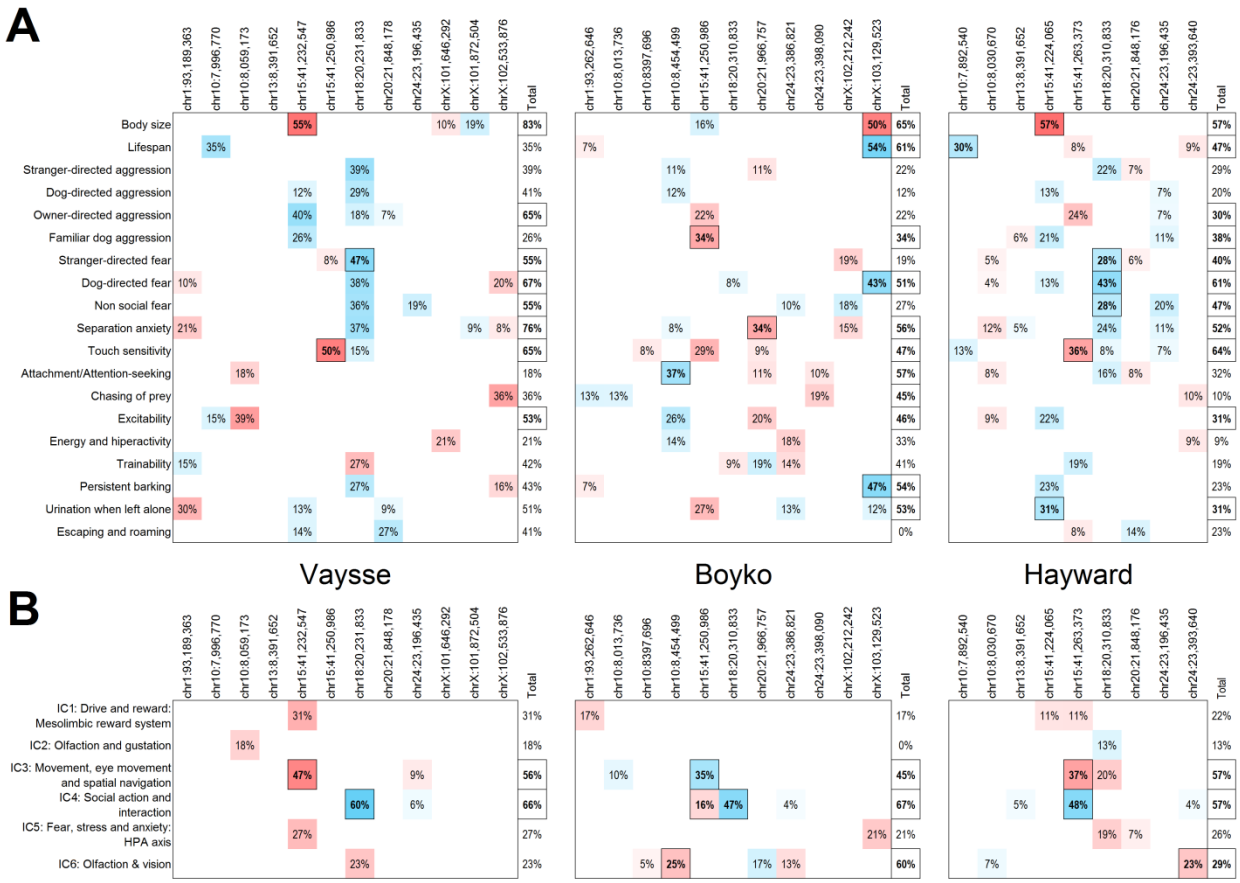
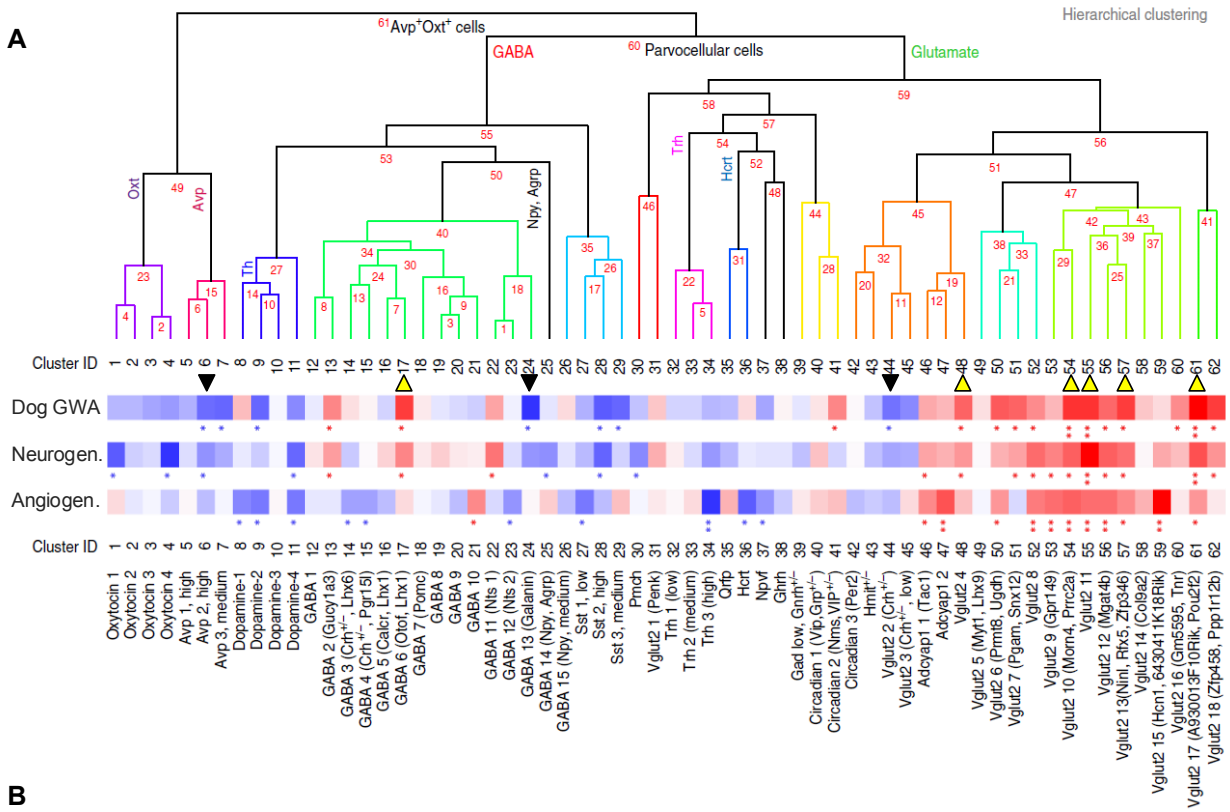


Figure 2. Effect directions and variance contributions. (**A**) Effect directions and variance contributions of relevant loci to behavioral traits estimated by linear regression. (**B**) Effect directions and variance contributions of relevant loci to Independent Component brain networks estimated by linear regression. Each relevant locus represents a compiled group of individual SNPs that are correlated grouped through Eigen decomposition of markers correlated by linkage. Cohorts correspond to Vaysse et al., Boyko et al. and Hayward et al., respectively[13-15]. Only significant associations are shown. Percentages correspond to R^2 values while total indicate the overall model R^2. Positive associations are indicated by a blue gradient while negative associations are indicated by a red gradient. Boxed and bolded values are Bonferroni significant for the cohort.



B

Mouse Hypothalamus scRNA-seq Neuron Cluster (of 62, including 15 GABA & 18 Glutamate)	Neurot., depress. & wellbeing (N 660)[2]		Educational attain. (N 1286)[3]		Tobacco & alcohol use (N 1117)[4]		Autism Spect. Dis. (N 1119)[5]		Brain structure (N 937)[6]		Dog behavior GWAS's (N 804)[1]	
	observed / expected	under *P* or over *Q*	observed / expected	under *P* or over *Q*	observed / expected	under *P* or over *Q*	observed / expected	under *P* or over *Q*	observed / expected	under *P* or over *Q*	observed / expected	*under P or over Q*
Dog behavior GWAS's (N 804)[1]	1.41	*5.10E-03*	1.58	**1.90E-08**	1.53	**3.88E-06**	1.48	**2.17E-05**	1.59	**4.23E-06**	NA	NA
Down Clust. 6, Avp (N 2k)	0.86	5.77E-02	1.13	*2.02E-02*	1.05	2.48E-01	0.91	9.87E-02	1.69	**1.08E-07**	0.71	**1.05E-04**
Down Clust. 24, GABA (Gal) (N 2k)	0.87	7.22E-02	1.02	4.15E-01	0.94	2.15E-01	0.85	*1.18E-02*	1.64	**7.16E-07**	0.63	**7.24E-07**
Down Clust. 44, Glut. (Crh) (N 2k)	0.74	*1.47E-03*	0.95	2.23E-01	1.07	1.74E-01	0.78	**4.09E-04**	1.78	**3.62E-09**	0.72	**1.61E-04**
Up Clust. 17, GABA (N 2k)	1.04	3.31E-01	1.67	**2.88E-24**	1.63	**4.57E-19**	1.33	**7.17E-07**	2.20	**2.05E-18**	1.22	*4.55E-03*
Up Clust. 48, Glut. (N 2k)	0.93	7.22E-02	1.67	**1.38E-24**	1.62	**1.80E-18**	1.16	*8.79E-03*	2.34	**5.49E-22**	1.16	*2.66E-02*
Up Clust. 54, Glut. (N 2k)	1.03	3.71E-01	1.65	**2.57E-23**	1.76	**1.44E-26**	1.25	**1.13E-04**	2.32	**1.59E-21**	1.24	*1.90E-03*
Up Clust. 55, Glut. (N 2k)	1.16	*4.18E-02*	1.88	**1.84E-39**	1.83	**5.91E-31**	1.34	**4.78E-07**	2.44	**7.51E-25**	1.24	*1.90E-03*
Up Clust. 57, Glut. (N 2k)	1.07	2.22E-01	1.68	**6.55E-25**	1.62	**1.80E-18**	1.25	**1.55E-04**	2.31	**4.56E-21**	1.22	*3.43E-03*
Up Clust. 61, Glut. (N 2k)	1.31	**4.82E-04**	1.96	**3.70E-46**	1.80	**1.01E-28**	1.37	**3.72E-08**	2.37	**6.34E-23**	1.32	**6.14E-05**

Figure 3. Relevance to specific neurons in the mouse hypothalamus and human neurogenetics. (**A**) Analysis of the canine behavioral GWA geneset vs. 62 neuron scRNA-seq clusters from mouse hypothalamus[40], and GO genesets for neurogenesis and angiogenesis. The labeled cluster image was taken directly from Romanov et al. 2017 (ref. [40]) and adapted. Yellow and black arrow heads mark the clusters depleted and enriched by 1SD from the mean in both the present and MacLean et al.'s (gene-based GWA corrected body mass)[6] canine behavioral scans. Blue signifies depletion and red enrichment, and the number of asterisks corresponds to one or two SD from the mean for the combined canine geneset. (**B**) The marked depleted and enriched clusters in "A" were compared to the combined canine geneset and several brain-related genesets from human genetics. Note *P* and *Q* values are hypergeometric probabilities for depletion and enrichment, respectively; bold denotes significant with Bonferroni correction ($P<8.47\times10^{-4}$ threshold for 59 tests). Blue signifies depletion, red enrichment. Human datasets are provided in Supplementary Tables S39/40.

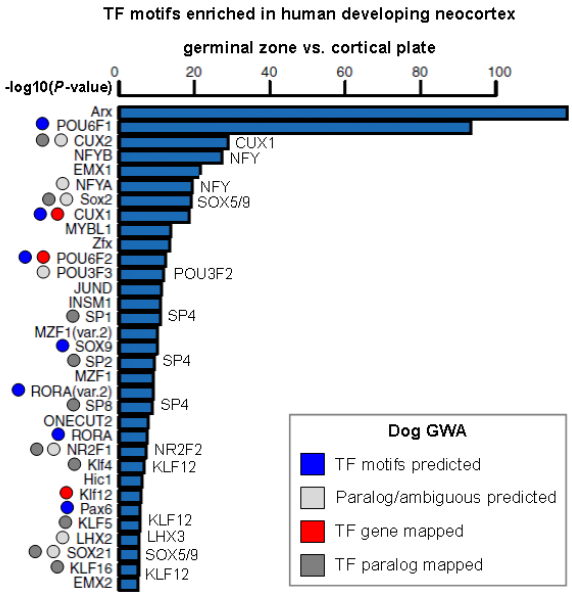


Figure 4. Relevance to neurogenesis in human embryonic neocortex. The combined canine behavioral GWA geneset from this study and MacLean et al.'s (gene-based GWA corrected body mass)[6] was used to compare enriched transcription factor binding motifs (labeled on the right) to those enriched in developing neocortex germinal zone vs. cortical plate (labeled on the left). The figure was adapted from de la Torre-Ubieta et al.[43].