Science
AAAS

Supplementary Materials for

Ancestry-inclusive dog genomics challenges popular breed stereotypes

Kathleen Morrill *et al*.

Corresponding authors: Kathleen Morrill, kathleen.morrill@umassmed.edu; Elinor K. Karlsson, elinor.karlsson@umassmed.edu

Science **376**, eabk0639 (2022)
DOI: 10.1126/science.abk0639

The PDF file includes:

Materials and Methods
Supplementary Text
Figs. S1 to S34
Tables S1 to S15
References

Other Supplementary Material for this manuscript includes the following:

MDAR Reproducibility Checklist
Data S1 to S20

Table of contents

Materials and Methods

Enrollment

Upon enrollment in Darwin's Ark (https://darwinsark.org), owners were asked to provide consent for participation and information about their dog's approximate birth date, sex and spay/neuter status, suspected or known breed(s), purebred registration, and/or photograph. We assigned dogs to nine major regions of the contiguous United States based on participant zip code, and defined as urban, suburban, and rural by the U.S. Census Bureau Decennial Census of 2010 (www.census.gov) based on population density. We limited our analyses to data obtained until November 15th, 2019 ("data freeze" date). Dogs were prioritized for DNA sequencing by completeness of surveys, enrollment date, and distribution of survey responses.

Survey collection

This manuscript includes data from the initial 11 behavioral surveys added to the Darwin's Ark project (ten questions each) (see **Data and materials availability**), and one survey about physical characteristics (eight questions), for a total of 118 questions (table S1) (the Darwin's Ark site currently includes 22 surveys with 8-10 questions each). The surveys were offered to owners in a static order, two at a time, on the owner's main account page (dubbed "My Laboratory"), although participants could also opt to answer them in any order by selecting "View All". Participants can opt to retake surveys, and both the original and retake responses are stored. For all analyses described here, only the original answer is used.

When survey questions are offered to owners, the text of the question is automatically updated for their dog. The variable text is capitalized when questions are described in our data files. For example, the size question, "When DOG is standing next to someone of average height, how high are HIS shoulders?" would be modified to replace DOG with the dog's name, and HIS with the appropriately gendered pronouns based on the owner's report of the dog's sex. We include the dog's name in every survey question, and pronouns as needed. This personalization ensures that owners of multiple enrolled dogs answer the question for the correct dog.

All 110 behavioral questions used a 5-point Likert scale: (1) 81 questions had options of strongly agree, agree, neither agree nor disagree, disagree, or strongly disagree; (2) 29 had options of never, rarely, sometimes, often, or always. Responses saved were codified as values $\{0,1,2,3,4\}$.

We sourced 79 behavioral questions from published and validated canine behavioral and health surveys: (a) Dog Personality Questionnaire (DPQ / DPQL; 45 questions) (*37*); (b) Canine Health-related Quality of Life Survey (CHQLS; 11 questions) (*36*); (c) Dog Impulsivity Assessment Scale (DIAS; 18 questions, including one also in DPQ) (*34*); (d) Canine Cognitive Dysfunction Rating scale (CCDR; 6 questions) (*35*).

We also included 31 new behavior questions and eight about breed-defining aesthetic traits (table S1). We developed the new behavior questions with input from animal behavior professionals associated with the International Association of Animal

Behavior Consultants. Our goal was to identify behaviors that were both heritable and easy for owners to identify, and thus well suited for a community science behavioral genetics project. We first collated a list of 46 possible question topics based on the professionals' initial suggestions, and then asked them score each topic on eight criteria, along a 5 pt scale: (1) incidence in pet dog population (rare to common); (2) how easily behavior is observed by owners (hard to easy); (3) whether the behavior was quantitative or binary; (4) whether environment was likely to have a major affect (likely to unlikely); (5) how malleable the behavior is (prone to resistant); (6) whether the owner was likely assign a value judgment to the behavior, and thus to try and train or untrain it (Biased to Neutral 1-5); (7) whether the prevalence of behavior tended to differ between breeds, suggesting heritability (low to high); (8) whether spay/neuter status would have a strong effect (strong to weak). From this, we identified 31 questions that scored at least moderately highly on all criteria, and were not included in existing surveys. Our subsequent heritability analysis showed a disproportionately high number of the new questions in the top quantile of heritability (12 out of 31, or 39%; p= 0.04 on 1-sample proportions test without continuity correction).

We also included a survey with eight questions about physical characteristics, and these had more variable options, including one allowed for more than one answer to be selected (Q#122, "What color is DOG? Select all that apply.", with eight options) (table S1). Ticking coat pattern phenotypes were validated in dogs with photos through manual examination. Another four asked owners to select an answer based on an accompanying graphic showing a range of responses (fig. S3). When physical trait questions were analyzed as quantitative traits in subsequent analyses, answers of "I'm not sure", "I don't know", "Not sure" and, for question #125, "Surgically cropped ears", were set to missing (NA).

Phenotypes derived from survey data

Phenotypes were primarily defined directly as behavioral factor scores, responses to behavioral questions, and responses to physical trait questions. For quantitative analyses, scores were normalized by calculating the standard score; normalized scores are included in the shared survey data files (see **Data and materials availability**).

Birth dates and age

All responses to survey questions were timestamped in POSIX time format. Prior to July 3rd, 2018, age and birth dates for enrollment were collected as free response entries. In order to standardize these as birth dates in international format (YYYY-MM-DD) for estimation of age, the following steps were executed using a combination of functions from the *R* packages `*data.table*`, `*stringr*`, `*anytime*`, and `*lubridate*`:

1. For dogs with a parsable birth date, directly convert to YYYY-MM-DD.
2. For dogs with a year and month, assign birth date YYYY-MM-01.
3. For dogs with a year only, assign birth date YYYY-01-01.
4. For dogs with no parsable birth date but age given in years and/or months, parse into duration and subtract from date of earliest survey to estimate birth date.
5. After each of the above steps:
   a. Set all birth dates before January 1st, 1980 to NA

   b. Set all birth dates postdating survey response dates to NA
6. The remainder with age or birth date entries were parsed by hand, if an interpretable free response was given by the owner.
7. Otherwise, the birth date was rendered NA.

Sex and sterilization status

Half of dogs in our cohort (50.6%) were female, and 89.8% were spayed or neutered, higher than the ~70% reported in the most recent American Veterinary Medical Association demographic sourcebook (*18*). Sterilization rates for mutts in our data were 1.26x higher than purebred dogs, close to the 1.19x reported elsewhere (*108*). Sterilization status did not substantially change the effect of sex estimated for survey responses. The analysis of variance effect (ANOVA ges) of sex on behavioral survey reponses remained constant across intact and sterilized dogs (R_{pearson}=0.997, p=4.04x10^{-142}, N(intact)= 1,049, N(sterilized)= 13,278). In the final ANOVA analysis, we used four discrete sex categories: female (intact); female (sterilized); male (intact); male (sterilized).

Validation of size phenotypes

We validated the owner's responses to question #121, “When DOG is standing next to someone of average height, how high are HIS shoulders?” (see *Survey collection*) using both individual size measurements (done by both owner and non-owner), and using breed-average heights (data S1). **Validation set #1:** 337 dogs in which owners were provided with a measuring tape by mail and instructed to measure the height from their dog’s shoulder to the ground (fig. S2B,C). **Validation set #2:** 38 dogs recruited during the 2017 Somerville Dog Festival in Somerville, MA. Owners were asked to complete the Darwin’s Ark surveys, and responses were compared to the height to withers measurement done at the event by Darwin’s Ark staff members (fig. S2D). **Validation set #3:** For 2,025 purebred dogs from breeds with average heights given in (*109*), we compared owner reported size to the average heights for their breed (fig. S2E). Pearson's correlation was calculated for each validation set using `cor.test()` in *R* package `*stats*` (version 4.1.1).

Validation of behavioral questions

We validated a subset of behavioral questions derived from published questionnaires performed as expected by comparing question-question correlations reported by the The Dog Personality Questionnaire (DPQ) (*37*) to our survey data. The original DPQ reported question-question correlations across 2,556 dogs (Study 3). The DPQ used a 7-point Likert scale keyed from “Strongly Disagree” to “Strongly Agree”, though a subset of questions had responses reverse-keyed. The Darwin’s Ark implementation of 48 DPQ items used a 5-point Likert scale keyed from “Strongly Agree” to “Strongly Disagree” (reverse direction, no re-keying of responses). We matched question numbers from the original DPQ (Appendix G) to question numbers from Darwin’s Ark surveys and extracted the matrix of question-question correlations from the original DPQ (from Appendices E and F). We reversed original DPQ correlations (r = x-1) for question pairs for which one question was indicated as re-keyed. For the same 48 questions sourced from the DPQ in Darwin’s Ark across 10,253 dogs, we obtained the question-question correlation matrix after normalization. We converted correlations to correlation distances ($d = 1 - |r|$) and performed a Mantel's test between the original DPQ matrix and the Darwin's Ark DPQ matrix of correlation distances using *R* package `*ade4*` with 100,000 replicates and found a Mantel's *R* of 0.9466711 ($p = 1 \times 10^{-7}$). We were not able to do a similar analysis of the other questionnaires used to source questions, as the published data did not include question-question correlations.

We also validated the subset of the questions from two other studies shown to correlate with age. For all questions, we also see an age correlation in Darwin's Ark. Two questionnaires included age-related questions: six questions sourced or paraphrased from questions in the Canine Cognitive Dysfunction Rating (CCDR) scale (*110*) (table S11) and 11 questions from the Canine Health-related Quality of Life Survey (CHQLS) (*36*) (table S12). In total, 14 questions were assessed (three questions overlap between the surveys). The PPS scores for all 14 questions are significantly correlated with age, with an average Pearson correlation of 0.889 (SD:0.121; range:0.630-0.99), compared to 0.661 (SD:0.308; range=0.0375-0.994) for the other 96 behavioral questions. The age correlations for these 14 questions are significantly higher than for other behavioral questions (one-sided t-test; t=5.06; $p=4.0\times10^{-6}$; df = 43.7). The direction and magnitude of the change matched the previously published results.

Exploratory factor analysis

We performed exploratory factor analysis using the *R* package `*nFactors*` (*111*) on the behavioral survey questions using 10,253 dogs with answers for all 110 behavioral questions. All survey responses were first normalized to a 0-6 range using the *R* function normalize (method="standardize"). The number of factors to extract was estimated from several heuristic methods, namely Horn's Parallel Analysis (n = 20), Optimal Coordinates (n = 20), and Acceleration Factor (n = 2), and the optimal factor number of 20 was selected. After excluding questions with low pattern or structure loadings (< ±0.3), 19 factors were generated. A varimax orthogonal rotation was applied to generate a structure matrix with factor loadings for each question. The first 8 factors explained a cumulative 24.26% of variance (table S3), and were selected for additional analysis (fig. S4A,B). The remaining factors were discarded. An additional 6,269 dogs with less than 20% missing data had factor scores generated by filling missing data with randomly sampled values. The age of each dog for each factor was calculated as the mean age at survey response to questions included in that factor. Factor scores for 16,077 dogs enrolled before the data freeze date were retained.

We assigned names, adjectives to describe low and high scores, and short descriptions based on the questions captured by each factor. Factor 1 ("Human Sociability"), which captured questions about social interactions with people, explained 3.60% of variance. Factor 2 ("Arousal Level") included questions about a dog's reaction to excitement and explained another 4.06% of variance. Factor 3 ("Toy-directed Motor Patterns") included questions about engaging with toys and objects, which may represent underlying differences in canine motor patterns. Factor 4 ("Biddability") represented ease of training and amenability to trained behaviors. Factor 5 ("Agonistic Threshold") describes the conditions and contexts in which a dog may apply agonistic behavior.

Factor 6 ("Dog Sociability") highlighted dog-directed social interactions. Factor 7 ("Environmental Engagement") covers questions about a dog's responsiveness to their surroundings. Factor 8 ("Proximity Seeking") describes behaviors aimed at human contact and proximity. See table S4 for a breakdown of these factors.

Sample collection

We sent owners saliva collection kits (DNA Genotek PG-100 saliva swabs) to sample their dogs, and received and stored a total of 6,909 saliva samples. We preferentially selected dogs for sequencing based on survey completeness and enrollment date, as well as including dogs from several underrepresented breeds to expand the breed calling panel to include the 100 most common breeds in the U.S. Of 1,715 samples submitted for low-coverage DNA sequencing, 159 samples (7.4% of 2,155 dogs included in the genetic data set) had sequencing funded by owner donations to the Darwin's Ark Foundation, a 501(c)(3) non-profit organization (82-3942341).

Reference genome assembly

For all genetic analyses, we use the CanFam3.1 reference genome assembly (NCBI accession GCF_000000145.2). Chromosome Y was excluded because of the lack of a high quality assembly. For imputed data, chromosome X was excluded because of the potential for its smaller population size (~75% that of autosomes) to impact both imputation and selection analyses.

Whole genome sequencing and joint variant calling

We performed high-coverage whole genome sequencing (WGS) at an average 45.82x (SD: 9.76x) of 27 putatively mixed-breed dogs and completed the first genetic analysis of this population, dubbed the Mendel's Mutts cohort (data S2). We performed joint variant calling on 676 whole genomes of dogs, wolves, and other canids, including 22 of the 27 mutts, from publicly available sequencing data as well as sequencing data we released under BioProject PRJNA683923. All steps of variant calling were performed using the Genome Analysis Toolkit (GATK3, nightly version from June 24th, 2016). First, GATK Base Quality Score Recalibration (BQSR) ran with standard arguments and specified, known canid variants comprising 19,112,082 distinct single-nucleotide polymorphic sites compiled from the Dog Genome SNP Database (DoGSD), Broad Institute (BICF), and Axiom. Second, gVCFs were generated using GATK HaplotypeCaller. Two genotype quality bands were set with upper bounds of 20 and 100 (-GQB 20 -GQB 100). Paths with fewer than 2 supporting kmers were pruned (-minPruning 2). Since we were unsure if any PCR free samples were included, we set the PCR indel model to none (-pcrModel NONE). GATK GenotypeGVCFs performed joint variant calling of all samples together to produce a single VCF. The SNPs and indels were filtered separately and subsequently recombined. The SNPs were filtered using the following parameters: "QD < 2.0 || FS > 60.0 || MQ < 40.0 || MQRankSum < -12.5 || ReadPosRankSum < -8.0". Indel filtering used the following parameters: "QD < 2.0 || FS > 200.0 || InbreedingCoeff < -0.8 || SOR > 10.0 || ReadPosRankSum < -20.0". After joint calling, we removed samples that did not meet a minimum coverage of 10x and 0.8 calling rate. We split multiallelic sites into biallelic records and normalized and left-aligned insertions / deletions using the *BCFtools* norm function with the CanFam3.1 reference FASTA. We then performed haplotype phasing with *Beagle* (v5).

The final VCF, referred to as the Broad-UMass Canid Variants (available at https://data.broadinstitute.org/DogData/), contained 46,134,885 anchored variant records, with 34,191,821 SNPs and 11,943,064 insertions / deletions for 540 dogs of known breed ancestry distributed among ~133 breeds, 28 dogs of mixed breed ancestry (22 of which come from the deeply sequenced Mendel's Mutts cohort described below), 12 dogs of unknown ancestry, 62 worldwide indigenous or village dogs, 33 wolves, and 1 coyote. See data S4 for samples included in the variant call file. The other 5 mutts in the Mendel's Mutts cohort underwent calling for the same set of variants in the Broad-UMass VCF using HaplotypeCaller from GATK3 in genotyping mode `GENOTYPE_GIVEN_ALLELES`.

Unique and shared genetic variants

We selected 22,139,829 biallelic SNPs from autosomes and chrX with allele count $\geq$1 across all 27 mutts and 530 breed whole genomes representing 128 breeds; of these, 375,474 (1.7%) were unique to mutts and not observed in any sampled breeds; 11,599,379 (52.4%) were shared between mutts and breeds; and 4,577,549 (20.7%) were unique to individual breeds, among which each breed had, on average, 35,762 ± 68,719 unique SNPs. Another 5,587,427 (25.2%) SNPs were shared across 1+ breeds and unobserved in the mutts. To assess the frequencies of population-specific SNPs, we randomly sampled n=10 dogs from mutts and each of 13 breeds.

Cumulative variant discovery

We compared the rate of variant discovery by whole genome sequencing of individual purebred versus mutt dogs, using a randomly chosen chromosome (chr13) as a proxy for the whole genome. From among the 557 total dogs for which 30x WGS data were available, we considered three cohorts: one random dog per breed (N=128 possible dogs), mutts from Mendel's Mutts (N=27 dogs), and, separately, dogs of each of the four breeds for which WGS data were available from >27 individuals: golden retriever (N=36), Yorkshire terrier (N=56), Labrador retriever (N=31), and Leonberger (N=38). We computed the cumulative distribution of the fraction of the 619,031 chr13 variants discovered using all 557 dogs that were discovered using from one to ten dogs randomly chosen and ordered from within each cohort. We computed 95% CI from the distribution of values from random reorderings within each cohort.

Linkage disequilibrium

Analysis of linkage disequilibrium was done with *PLINK* v1.90b6.9. For each population in the Broad-UMass Canid Variants VCF, 25 dogs were chosen at random from all dogs in the population. We tested seven populations: (1) mixed breed dogs (Mendel's Mutts); (2) four breeds (golden retriever, Labrador retriever, Leonberger, and Yorkshire terrier); (3) village dogs; (4) wolves. We included mixed breed dogs and golden retrievers with WGS coverage > 30x; for other populations, coverage was >15x. We first filtered for autosomal, bi-allelic variants, and then selected 20,000 random SNPs. For each population, we assessed the extent of LD by measuring r^2 between each random variant and all variants within 100kb. We assessed tagging of variant sets by measuring r^2 between each random variant and variant sites included on genotyping arrays or in the low-pass sequencing GWAS dataset (171,882 for the Illumina HD, 1,011,992 for the Axion, and 10,355,966 for low pass sequencing). Within each population, we analyzed all SNPs with MAF > 0.025 (table S13).

Runs of homozygosity in mutts, breeds, and village dogs

We compared the lengths of detected runs of homozygosity in mutts, dog breeds, and village dogs using whole genome sequencing data for dogs in the Broad-UMass Canid Variants set (58,308,734 biallelic SNPs) and Mendel's Mutts cohort (56,797,766 biallelic SNPs). The settings used to detect ROH in the WGS data were as follows: minimum run length of 100kb (--homozyg-kb 100) and minimum SNP count of 100 SNPs (--homozyg-snp 100), at a density of 1kb per SNP (--homozyg-density 1), with no two SNPs more than 500kb apart (--homozyg-gap 500), and no heterozygous genotypes tolerated (--homozyg-window-het 0) (*112*). The remaining options given by the default settings implemented in *PLINK* v1.90b6.21 (19 Oct 2020). We then randomly sampled $n= 464$ runs, the mean number of ROH detected in each mutt, from the pool of ROH detected in mutts, purebred dogs, and village dogs, resampling $N= 100$ times.

Low-pass sequencing with imputation

1,715 dogs enrolled in Darwin's Ark were sequenced at coverages of 0.5x to 1.1x depth on the Gencove sequencing platform. Sequencing reads were processed into imputed autosomal variant calls through Gencove's *loimpute* software (*46*) using the copying model described in Li and Stephens, 2003 (*113*) , and an imputation reference panel containing publicly available whole genome sequence data (mean coverage 22.9x (SD: 14.2x)) for 435 canids and representing 287 dogs of known pure breed ancestry, 6 dogs of unknown ancestry, 100 worldwide indigenous or village dogs, 36 wolves, and 6 other wild canids, VCF provided by Elaine Ostrander of the National Human Genome Research Institute (all accessions in data S4). This generated 46,349,043 unfiltered call site records -- 32,438,672 SNPs and 13,910,371 indels -- mapping at a density 19.8 (SD: 6.9) variants of per kb. We used the impute genotype probability (GP) scores per genotype per call site per dog for filtering purposes. Filtered at GP >70%, each dog averaged 32,213,747 (SD: 141,060) SNPs and 13,603,537 (SD: 63,729) indels, including 98% of common variants (minor allele frequency > 1% in dogs with high-coverage WGS data). This data was merged and filtered according to the procedures listed below. See data S4 for samples in the imputation reference panel.

Genotyping arrays

440 dogs underwent genotyping on the Axiom Canine Genotyping Array Set A & B for 1,268,920 variant call sites (1,267,416 SNPs and 1,504 indels) before we adopted the low pass sequencing approach described above. Variant call data for these samples were subsequently imputed with *loimpute* against the same panel described above, also resulting in 46,349,043 unfiltered call sites with GP scores. Filtered at GP >70%, each dog averaged 32,006,290 (SD: 157,307) SNP and 13,497,132 (SD: 50,619) indels. This data was merged and filtered according to the procedures listed below.

Imputation performance and genotype concordance

We used two approaches to compare the performance of low-pass sequencing plus imputation to that of genotyping arrays in admixed dogs: (1) by running low coverage (1.0x ± 0.6x) re-sequencing and genotyping arrays on mutts with WGS data and (2) by down-sampling WGS data. For the truth set of genotypes, we selected SNPs at MAF >2% in 676 dogs and observed at least once in the 27 dogs from the Mendel's Mutts cohort. The sample genotype concordance between unimputed and imputed genotypes from genotyping arrays (9 dogs, 10 arrays) or imputed genotypes from low-pass sequencing (11 dogs, 14 sequencing runs) or downsampling (5 dogs) and whole genome sequencing data was evaluated using the *BCFtools* stats function.

Low-pass sequencing plus imputation provided a higher density of variant calls (19.8 ± 6.9 variants per kb, ~40x denser than the Axiom array). We found concordance between low-pass and WGS of 98.3 ± 0.7% (*n*=14 runs; ~7.7 million SNPs of MAF >2% across 676 dogs), similar but slightly lower than the Axiom array (99.3% ± 0.1%; *n*=10 arrays; 0.83 million SNPs), but better than imputed array calls (97.3 ± 0.3%; 7.6 million SNPs), with a lower discordance for imputing heterozygous genotypes (1.09% vs. 1.66%).

Merging and filtering

Prior to merging, samples processed by low-pass sequencing or genotyping arrays with imputation were filtered to remove genotypes with genotype probability below 70% (*BCFtools* `-e 'MAX(GP[*])<0.7'`). Subsequently, VCFs were merged (*BCFtools*) and converted to a *PLINK* data set. SNPs below a minor allele frequency of 2% and missing in over 20% of individuals were filtered out (*PLINK* `--maf 0.02 --geno 0.20`). Only biallelic SNPs with extreme deviation from Hardy-Weinberg equilibrium, given *p*-values below 1×10^{-20} in the exact test with mid-*p* adjustment and at observed/expected heterozygosity ratios under 0.25 or above 1.0, were excluded. After filtering data, 8,518,951 SNPs and 2,155 dogs remained with a total genotyping rate of 97.5%. Owner-reported sexes were encoded in the sample information file, confirmed by relative X-chromosome coverage for sequencing data (*SAMtools* `idxstats`) and the autosomal genotypes of X for genotyping data; in total, 1,084 males and 1,071 females. Six dogs flagged as unusual in their ratio of X chromosome coverage to autosomal coverage, and two of which came from dogs of the same household but could not be confirmed as sample mix ups. Variant IDs were assigned to include chromosome, position, reference allele, and alternate allele.

Homozygosity and inbreeding

We scanned for runs of homozygosity in the genetic data for the Darwin's Ark cohort. For the Darwin's Ark cohort, we scanned 32,742,462 biallelic SNP genotypes with genotype probability >70% using *PLINK* (v.1.9) with the following settings: minimum run length of 500kb (--homozyg-kb 500) and minimum SNP count of 100 SNPs (--homozyg-snp 100), at a density of 1kb per SNP (--homozyg-density 1), with no two SNPs more than 500kb apart (--homozyg-gap 500), and only 1 heterozygous genotypes tolerated per window (--homozyg-window-het 0) (*112*), performing scans without LD-based pruning (*114*). We then calculated the autosomal ROH-estimated coefficient of inbreeding (F_{ROH}) from the total ROH segment length divided by the total SNP-covered length across autosomes where ROH detection was possible (*115*, *116*).

Kinship and relatedness

We measured kinship *k* between pairs of individuals using the KING-robust kinship estimator (*117*) on the GWAS cohort (*N*= 2,097 with any phenotype data; 2,197,656 pairs). The majority (99%) of pairs were unrelated (*k*<0), and the average kinship score was -0.232 ±SD:0.298 (fig. S32). Only 28 pairs of dogs (0.0017%; 46 unique dogs) exceeded kinship of *k*=0.2 (just under half-siblings).

Breed name standardization

Owner-reported breed names were standardized to unify those referring to the same breed and to correct misspellings. In addition to analysis as separate breed populations, breeds were assigned higher level groupings to check whether ignoring certain features that may not necessarily distinguish dogs into separate breeds (e.g., `wirehaired dachshund` and `longhaired dachshund`; various sizes of poodle) impacted the relationship between breed and survey data.

Breed standards, stereotypes, and popularity

The breed standard values for six physical traits (body size, ear shape, number of coat colors, white spotting, coat length, and coat type) were obtained from the American Kennel Club (*53*) or, for the non-AKC breeds, breed-specific clubs (see **Data and materials availability**). We collected three-word stereotypes from the American Kennel Club breed resources at www.akc.org/dog-breeds (*52*) and assigned breeds to breed groups based on information from the AKC official breed standards (*53*). The following sets of descriptors were treated as the same stereotypes: (A) {bright, clever, intelligent, smart, very smart}, (2) {devoted, loyal}, and (3) {energetic, exuberant, lively, peppy}. The breed group “non-sporting” was excluded because it potentially encompasses a broad spectrum of working functions (it is defined as “A diverse group of multifunctional dogs not generally regarded to be game hunters.” (*53*)). We collated scores for each breed on ten behavioral characteristics from the “At a Glance” tables of the *Encyclopedia of Dog Breeds* (*19*). We used the mean number of American Kennel Club (AKC) registrations reported between 2000 and 2015 as a proxy for breed popularity in the United States (*39*).

Breed reference panel

We selected 871 purebred dogs representing 89 breeds with published genotypes for 185,805 SNPs from the Illumina CanineHD array (*118*). Owners submitted data for 12 purebred English shepherds with 214,220 SNPs genotyped on a custom high-density Illumina marker platform. Illumina CanineHD array data for 17 Golden Retrievers and 11 Leonbergers, for which we also have 30x whole genome sequencing data, were imputed to test the genotype concordance of imputed Illumina genotyping calls. We subsequently imputed genotypes for 32,438,672 SNPs from these arrays. An additional 109 dogs representing 43 breeds were genotyped at 1,011,254 SNPs on the Axiom Canine Genotyping Array Set A & B and imputed up to 32,438,672 SNPs. Among dogs enrolled in Darwin’s Ark, 115 registered purebred dog samples from 54 breeds had low-pass sequencing data considered for inclusion. A total of 380 purebred dog samples spanning

74 breeds with whole genome sequencing calls were considered for inclusion in the breed reference panel. Among these, 194 overlapped with samples in the imputation panel.

We performed nearest neighbor checks using *PLINK* and for most breeds, found that the 3 genetic nearest neighbors for each dog were of the same breed. We identified 14 dogs with nearest neighbors from an additional breed, including 8 American Pit Bull Terriers nearby Staffordshire Bull Terriers. Three samples of Greyhound were discovered to be duplicates.

For each of these 101 breeds, 12 dogs were selected to be included in the breed reference panel, prioritized by raw data density and genetic diversity within the breed.

We retained SNPs which were genotyped in over 80% of dogs and at a minor allele frequency of at least 5% across dogs. These 7,065,140 SNPs shared 2,528,037 positions with the recombination-based genetic map measured in free-living village dogs (*119*). For the remaining unshared positions, centimorgan map distance values were linearly interpolated using *R* `approx()`. In order to select dense ancestry-informative markers, SNPs within 5 kb of index SNPs above $F_{ST} = 0.15$ between reference breeds and with $r^2 > 0.9$ were clumped (as *PLINK* --clump considers values below threshold for indexing, $1-F_{ST}$ was supplied to the clumping function). The final set of local breed ancestry markers contained 2,468,442 SNPs (on average, spanning 110 SNPs per 100 kb) and 1,212 dogs (101 breeds with 12 dogs per breed) with a total genotyping rate of 99.44%, which we used to perform admixture simulations. Similarly, sparser markers for global ancestry inference were obtained by clumping with 50 kb, $F_{ST} > 0.15$, and $r^2 > 0.5$, yielding 688,060 global ancestry markers.

Admixture simulations

We used a Monte Carlo approach (GitHub repo: lindaboettger/ancestry_assignment) to generate simulated admixed individuals with known ancestry per haplotype, and then compared the known breed ancestry composition with global ancestry inferred via *ADMIXTURE* (*120*). To simulate admixed individuals, we performed N = 15 generations of admixtures according to the following procedure. N+1 random individuals from different breeds were selected to contribute to the admixture. With each iteration, recombination was simulated to incorporate a new individual. Recombination events were treated as a Poisson event occurring on average once every Morgan. The above simulation was run on 10 independently drawn data sets of 6 dogs per reference breed to create 1,000 admixed individuals of known ancestry. Genotype data from each set of simulated individuals was merged with the remaining 6 dogs per reference breed and filtered for SNPs in the global breed ancestry panel. We inferred global breed ancestry for simulated individuals using the supervised mode of *ADMIXTURE* (random seed = 43) supplied with reference population assignments. We correctly call breeds comprising >5% of simulated ancestry >90% of the time, and inferred ancestry correlates with true ancestry out to 12 generations of admixture ($R_{\text{pearson}}=0.94$; $p=2.2\text{x}10^{-16}$).

Global ancestry inference

We performed unsupervised admixture analysis for 1,073,779 LD-pruned (50 kb and $r^2 > 0.5$) biallelic SNPs genotyped in the Darwin's Ark genetic data set, using 23,613 SNPs on chromosome 38 for 5-fold cross-validation of $K = \{1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 10, 15, 20, 25, 50, 75, 100, 150, 200\}$ clusters to identify K between 50 and 100 as the optimal number of clusters for the Darwin's Ark genetic cohort (fig. S6). For the supervised admixture analysis, genotype data from all query dogs was merged with all reference breed data and filtered for SNPs in the global breed ancestry panel. Global ancestry from the 101 reference breeds were inferred using the supervised mode of *ADMIXTURE* (random seed = 43) supplied with reference population assignments. Population weights under 1% were discarded from individual ancestries.

Breed set assignments

We considered two scenarios in which a dog might present as purebred: (i) the dog was assigned a single owner-reported breed and (ii) the dog was assigned a single owner-reported breed and registered purebred status. Other dogs with no or two breeds reported were presumed to be non-purebred. For breeds included in the breed reference panel, we determined the rate at which owners correctly identified the predominant breed ancestry of their dog and the relationship between owner-reported breed and percent top breed (table S6). We assessed accuracy for the 880 dogs with only one reported breed, as well as for a subset of 304 dogs described as registered purebred and the remaining 1,186 presumably non-purebred dogs. Global ancestry inference assigned at least 85% of ancestry to a single breed in most (89.7%) registered purebred dogs, and therefore any dog with 85% or more breed ancestry was considered confirmed as purebreds by sequencing. Likewise, dogs with under 45% breed ancestry not matching their owner's report and not reported as a breed we cannot detect were considered mutts.

We defined three breed sets: (1) *confirmed purebred dogs* were either described as registered purebred by the owner, or confirmed by sequencing to have 85% or more ancestry from a single breed (3,637 dogs); (2) *candidate purebred dogs* included all confirmed purebred dogs and dogs with owner-reported ancestry from one breed not in disagreement with genetic breed ancestry, if available (9,009 dogs); (3) *mutts* were all other dogs including any dogs with <45% genetic breed ancestry in discordance with owner reported ancestry (9,367 dogs). Among 217 dogs with disagreement between owner reported breed and genetically inferred ancestry, 89 were reported as breeds absent from our reference panel and 90 were re-classified as mutts. Extrapolating from our analysis of dogs with genetic data, we expect that 89.7% of registered purebred and 58.2% of dogs with owner-reported ancestry from one breed would, if sequenced, have >85% ancestry called from their owner-defined breed. This is higher than the 3% expected by random chance from the mean probability of an owner guessing the breed correctly, for all dogs with breed confirmed by sequencing, given the assumption that owners guess breeds at rate corresponding to breed population frequency.

SNP-based heritability estimates

We used the Genome-wide Complex Trait Analysis (GCTA, version 1.92.3 beta 3) (*56*) software tool to calculate LD scores in 250kb regions using a block size of 10,000kb with an overlap of 5,000kb between blocks. In GCTA, we generated a genetic relationship matrix (GRM) for the data set of 2,155 dogs, as well as multiple GRMs calculated from SNPs stratified into LD score quartiles (*121*). We used restricted maximum likelihood (REML) analysis for all factor and question scores and 14 physical traits using the GRM for all SNPs to obtain estimates of h^2_{SNP} for the sample set of all dogs. We calculated the LD-stratified estimates of h^2_{SNP} for the same traits except when more than half of the variance components were constrained.

We also computed the first 10 principal components from the GRM of all SNPs using GCTA and performed heritability analyses in unrelated dogs (no kinship > 0.2) using the LD-stratified GRM(s) plus the first 10 PC eigenvectors as quantitative covariates. Heritabilities estimated in this manner had nearly perfect positive correlation with those estimated without PC covariates (R_{pearson}=0.974; p=1.1 x 10^{-58}) and an average fold change of 0.08 ± SD:0.43 with nearly perfect overlap of confidence intervals except for Q64 Circles before pooping. Additionally, we performed heritability analysis in unrelated, highly admixed mutts having no breed ancestry detected over 45%. Unsurprisingly (given the much smaller sample size) these correlated less tightly with estimates on the full set of dogs (R_{pearson}= 0.44, p=0.0004), with an average fold change of 0.53 ± SD:1.24, but the confidence intervals for most traits (92%) overlapped.

Analysis of variance

To assess how much of the variance in traits is explained by age, sex, size and breed, we ran an analysis of variance (ANOVA) 1−tailed analysis with all dogs > 1 year of age with age, sex, size, and breed information, using the anova_test function in *R* package `*rstatix*` 0.7.0, in *R* version 4.1.1. We tested both the candidate purebred and confirmed purebred classifications and included all breeds with ≥5 dogs. For the candidate dataset, we had 121 breeds and, on average, 5968±332 dogs per question/factor, and for the confirmed dataset, 78 breeds and 2251±119 dogs per question/factor. We used 2 different models to test how much of the variance in normalized factor and question scores was explained by age, sex, size and breed. For question 121 (the size question), which is the source of the size phenotypes, we used a model that excluded size: [value ~ breed + sex + age]. For all other questions we used [value ~ breed + sex + age + size]. We corrected for multiple testing using Benjamini-Hochberg procedure. To assess whether fur length explains any variance in Q21 (focus in distracting situations), we used the same procedure with the model [value ~ fur length + breed].

Predictive value of breed, age, and size

We calculated the improvement in prediction of behavioral traits afforded by information on the breed, age, and size of a given dog. To do so, we used a relative risk framework, asking, for example, by what fold information on any of these features could aid a prospective dog owner in identifying an individual dog with specified individual or combinations of behavioral traits, relative to a dog selected at random.

Interactive dashboard

We developed an interactive website for the public that illustrates the prevalence of particular traits in each breed. One version of the site shows the results for all eight behavioral factors, and the other, the results for the most heritable questions: Q17 H; Q54

(DOG retrieves objects); Q60 (DOG avoids getting wet); Q121 (When DOG is next to someone how high are HIS shoulders); Q123 (How much white fur does DOG have); Q124 (Is DOG's tail curly); Q125 (What is DOG's ear shape); Q127 (How long is DOG's fur on HIS back and sides). We grouped dogs in each breed into three groups based on their scores: "any" (all dogs); "negative" (dogs in the lower 25% quantile bound for all dogs; "positive" (dogs in the upper 25% quantile. We then empirically measured the frequency of all possible combinations of responses in each breed. We included both candidate and confirmed purebred dogs to maximize the number of breeds (minimum number of dogs / breed = 500). Input files for the website are included in data release (see **Data and materials availability**).

Population Peculiarity Scoring (PPS)

We used permutation testing to test whether populations of dogs, as defined by breed or age, differed significantly in their survey responses from randomly sampled groups on any question or factor. All dogs with any survey responses were included. We used *R* version 4.0 and custom scripts (see **Data and materials availability**). For each permutation, for a given sample size N (table S14), we calculated the mean (the observed test statistic) for each question/factor for N dogs sampled from among dogs of each type. For each permutation, we also calculated the mean for a random sampling of size N from the full dataset (the permuted test statistics). We counted how often the observed test statistics for each population were higher than the permuted test statistics.

We ran a total of 500,000 permutations and summed the counts. We calculated the one-tailed empirical *p*-value as the fraction of random samples where the permuted statistic is as large or larger than as the observed test statistic, and converted (1-p) into a Z-score, using the standard *R* function qnorm(), that matched the direction of the original survey score. We calculated 2-tailed *p*-values corrected for multiple testing by counting how often each observed test statistic exceeded the maximum of all permuted statistics (i.e., over all questions or factors), or was less than the minimum, for each of the replicates (*122*). This max(T) permutation approach preserves the correlational structure between questions, and thus is more appropriate than a Bonferroni correction, which assumes all tests are independent. When calculating permutation *p*-values, we added a 1 to the numerator and denominator to account for misestimation of the *p*-value (*123*).

We assessed whether the number of dogs available for sampling influenced the PPS results and found a small but discernible effect in the candidate purebred analysis. For the candidate purebred dogs, we set the minimum # dogs required to 25, matching the size of each sampling. Thus, for some breeds, each "random sampling" included the same dogs. While this allowed us to produce scores for many more breeds (up to 62, compared to maximum of 44 if the minimum is 50), breeds with smaller numbers of dogs available for sampling tended to have slightly more extreme z scores. In breeds with fewer than 50 dogs, 3.4±1.9% of traits have significant PPS scores, compared to 1.8±1.3% ($p_{\text{t-test}}$=0.00059) for breeds with more dogs. Number of dogs available for sampling is weakly inversely correlated with absolute z score = -0.14 (p=2.4x10^{-35}; N=7,482). We note that the correlation between number of dogs available for sampling and larger PPS z scores would have the effect of making breeds look more differentiated, rather than less, and thus the effect of sample size is unlikely to alter our conclusion that the behavioral differentiation between breeds is subtle.

We quantified the effect using an analysis of variance (data S9). The magnitudes but not the directions of the permutation results were affected by the use of less stringent candidate purebred breed labels. The permutation z scores for candidate breeds were strongly correlated with scores for confirmed breeds (for factors, Pearson correlation= 0.959, p= 5.95x10^{-18}; for questions, Pearson correlation = 0.934, p= 2.3x10^{-256}). On average, the z scores from the confirmed purebred sets are 0.82±0.61 larger (in either direction) than the candidate purebred z scores (fig. S13).

We used the PPS results for candidate purebred dogs to assess the validity of breed stereotypes determined from three different sources (see *Breed standards, stereotypes, and popularity* section). For stereotype groups (descriptive words and historic working roles), we used the *t_test* function in the *R* package `*rstatix*` version 0.7.0 (*R* version 4.1.1) to test the mean difference in PPS z scores between breeds in the group and not in the group. We excluded groups with fewer than 3 breeds from analysis. For quantitative stereotypes, we measured the spearman correlation between the PPS z scores and the stereotype score using the cor() function in the *R* package `*stats*` version 4.1.1. (*R* version 4.1.1). For both, we used the Benjamini-Hochberg procedure to correct for multiple testing.

MuttMix survey

To assess perceptions of breed ancestry in mixed-breed dogs by non-owner observers, we designed a survey hosted at www.MuttMix.org in which participants guess the three breeds detected at the largest percentage in each dog. The survey consisted of 30 mixed-breed dogs with ancestry assignments and one undeclared purebred dog (fig. S33). Owners of the 31 survey dogs were asked to provide a portrait and full body image, as well as a short video clip of the dog while in motion. In addition to the visual aids, owners indicated their dog's size relative to an average person via a visual graphic and noted other physical descriptions such as coat texture, markings or any concealed traits. The images and information provided by owners was consolidated and created into the MuttMix survey, where participants could guess the top three breeds.

Participants indicated whether they belong to the general public or are a dog professional and/or breeder. Participants were provided with 59 breed options to select as well as a "no choice" option for the third breed slot. Dogs were displayed in random order to participants. Participants were permitted to exit the survey at any time, return later, or leave the survey incomplete, but could not skip dogs. The survey launched on April 16th, 2018 and closed on June 16th, 2018, collecting responses from 26,639 people over a two month period. Survey data was compared to genetic breed ancestry, with any breed call below 5% removed; only breeds offered as survey options were examined.

To calculate the average total percentage of ancestry guessed correctly, we first calculated the percentage guessed correctly by each user for each dog individually by summing the real percent ancestry attributed to their three breed guesses. We then calculated the mean of all these values.

To assess the accuracy of user guesses of breed ancestry for MuttMix dogs, we first counted the number of breed guesses for a given dog that were among the top three breeds detected by our breed ancestry inference for that dog (or in the top two, if only two breeds comprised >5% ancestry). We find that the rate at which MuttMix users guessed a given breed ancestry is strongly correlated to the popularity of that breed (fig. S20), with no significant correlation to the position of a given breed in the guess dropdown menu. The expected rate of accurate guesses was determined using the overall population percent ancestry assigned to the breed in the "Darwin's Ark: Full Genetic Set" dataset (fig. S20). We calculated how often we expect to see each possible combination of breed guesses (N=32,509 for guesses for all 3 breeds; N=1,711 for guesses with just 2 breeds) by random chance if the guess rate for each breed is the overall population percent ancestry (table S2). For each dog, we then classified each set of breed guesses as 0 breeds correct, 1 breeds correct, 2 breeds correct or 3 breeds correct, and summed these to get the expected rate of guesses with 1+ breeds correct, 2+ breeds correct and 3 breeds correct. We then calculated the observed rate of guesses with 1+ breeds correct, 2+ breeds correct and 3 breeds correct for each dog, and then calculated the ratio of the observed rate to the expected rate (and did the same for guesses of just 2 breeds). We measured the significance of the difference between observed and expected correct guesses using the *chisq.test* function in the *R* package *`stats`* version 4.1.1 (*R* version 4.1.1), and applied Benjamini-Hochberg correction (table S15).

To assess the importance of individual physical traits in participants' breed choices, we compiled binary phenotypes for each of the above traits for each survey dog (see **Data and materials availability** for data files) and constructed a decision stump (a one-level decision tree) to assess how presence or absence of a given trait in a mutt impacted entropy among MuttMix guesses of the presence of a given breed ancestry. In particular, for each breed ancestry option, we calculated how well mutt phenotype for each of eight different traits (height, leg length, ear type, coat type, coat length, coat furnishings, white spotting, pigmentation) distinguished participant guesses of presence versus absence of that ancestry. For all breeds and all traits, we applied a leave-one-out analysis, omitting guesses for each mutt in series, to assess the impact of guesses for that mutt on entropy reduction. To calculate significance, we randomized trait assignments across mutts, then asked whether entropy reductions given true traits were greater than those expected from randomly assigned traits.

Linear mixed-effects regression models (LMERs)

We constructed linear mixed effects regression (LMER) models using the *R* package *`lme4qtl`* (*124*) (extending upon the package *`lme4`* (*125*)) (see **Data and materials availability** for script file) to measure the relationship of genetic breed ancestry with behavior question responses, normalized behavioral factor scores, and physical traits. Dogs with >45% ancestry from any one breed were excluded to eschew bias from recent breed admixture and purebred dogs. We included only breeds with a standard deviation ≥5% ancestry across dogs carrying that breed. Question and factor scores were treated as the independent variable. Breed ancestry from each of the remaining 68 breeds were scaled and treated as fixed effects. Three age brackets were considered as random effects: 2 years and under, between 2 and 12 years, and 12 years and older. The covariance matrix for genetic relationships between individual dogs was produced from the genetic relationship matrix (GRM) (see *SNP-based heritability estimates*). As the GRM was not positive definite, we generated a close positive definite version of the covariance matrix using the *R* package `*psych*` function cor.smooth() to provide the random effects of relatedness. A total of 1,002 dogs were included.

For each factor, we built a model with restricted maximum likelihood (REML) to obtain unbiased estimates, standard deviations, and Wald statistics (t.val) for the fixed effects of breed on factor score. We performed analysis-of-variance on each factor model to obtain the breed F-statistics. We also constructed each factor model using maximum likelihood (ML) with and without each breed, and performed an analysis of variance to obtain the likelihood ratio for each breed. We report *p*-values for the likelihood tests between the ML models, but not for REML models in which the derivation of *p*-values is not appropriate (*126*).

We obtained the proportion of variance explained by the fixed effects (breed ancestries) as the marginal Nakagawa's R^2 for each factor modeled with and without restricted maximum likelihood (*127*) (data S14). The conditional Nakagawa's R^2, which is the variance explained by the random effects, which are the age bracket and kinship covariance matrix, and fixed effects, which are the breed ancestries, could not be ascertained for all factors due to singularity in some models from the random effects structure being too complex. The source of structure complexity derived from the correlation between relatedness and breed ancestry, and upon removing the random effects of relatedness, only one factor model (factor 5) was singular.

Map of genes and open chromatin regions

We matched dog genes to human homologs with Ensembl gene annotation, GRCh38 (version 104.38) and CanFam3 (version 104.31), which gave us a set of 16,329 human gene-dog gene pairs for gene annotation.

For functional annotation of coding and non-coding regions, we derived 179,541 representative open chromatin regions (rOCRs) from 24 ATAC-seq datasets downloaded from BarkBase portal (http://www.barkbase.org/) (*102*). ATAC-seq datasets are processed using the standard ENCODE ATAC-seq pipeline (GitHub repo: ENCODE-DCC/atac-seq-pipeline). In short, trimmed FASTQ reads were aligned to the canine reference genome canFam3 using Bowtie2 (*128*). SAMTools and picard-tools were used for post-alignment filtering. MACS2 (*129*) and IDR (https://sites.google.com/site/anshulkundaje/projects/idr) were used for peak calling and filtering.

We used all samples with an IDR FRiP score greater than 0.05 (N=22). ATAC-seq IDR peaks were filtered based on signal (CPKM >10 percentile of all qualified samples), length (>150bp) and FDR ($<10^{-5}$). We also included peaks from a skeletal muscle and a occipital cortex sample with a more stringent FDR threshold ($<10^{-10}$) for better tissue diversity. A previously described (*101*) scheme was used to cluster ATAC-seq IDR peaks across samples, and the peak with the highest signal within each cluster was selected as the representative open chromatin region (rOCR) of the cluster.

Genome wide association studies (GWAS) using mixed linear models

We performed genome-wide mixed linear model-based associations on our Darwin's Ark cohort genetic data set (see **Data and materials availability**) using the "leave-one-chromosome-out" approach (MLMA-LOCO) implemented in GCTA (version 1.92.3 beta 3) by calculating multiple GRMs so that the model will exclude the random effects from the chromosome of each candidate SNP (fixed effect). Included were categorical covariates for sex and data type (genotyping or low-pass sequencing), as well as quantitative covariates for height and age for non-morphological traits. We used the human thresholds for genome-wide significance ($p= 5\text{x}10^{-8}$) and suggestive associations ($p= 1\text{x}10^{-6}$) owing to the similarity in linkage block length between humans and diverse dogs (*1*, *76*).

To define regions of association, SNPs in linkage disequilibrium ($r^2>0.2$ and $r^2>0.5$) and within range (250 kb) of associated index SNPs were clumped and annotated via *PLINK* --clump, referencing the gene map described above, and reported results in data S16. We also performed clumping for SNPs in linkage at $r^2>0.8$ within 250kb to compare the sizes of mapped regions (clumped for $r^2>0.8$ within 1Mb) with osteosarcoma-associated regions found by within-breed studies (*83*) in greyhounds (22, 22, 96, 90, 372, 82, 182, 72, 32, 198, 52, 105, 1304, and 72 kb), Rottweilers (1208, 330, 18, 19, 21, 487, 131, 16, 20, 26, 1988, 509, 54, 108, and 21 kb), and Irish wolfhounds (746, 1588, 1382, and 737 kb). We summarized the mean, median, and 25% and 75% quantiles for these osteosarcoma-associated regions and our suggestive ($p= 1\text{x}10^{-6}$) associated regions for physical traits and behavioral traits, and did not summarize any single-SNP associations that had no SNPs in linkage. We found that our regions extend to a median 5.6kb (Q25-75%=2.0-14kb, mean 16.8kb) around suggestive behavioral loci and 5.7kb at physical trait loci (1.4-22kb, mean 26.2kb), and that the intra-breed associations for osteosarcoma mapped at median ranges of 86kb (Q25-75%=57-162kb) in greyhounds, 54kb (21-409kb) in rottweilers and 1Mb (743kb-1.4Mb) in Irish wolfhounds.

The genomic inflation factors at 50th percentile for each study were calculated using the Python package `*scipy*`. Genomic inflation factor λ = np.nanmedian(chdtri(1, 1 - p_array)/chdtri(1, 1 - 0.5), where p_array is the array of GWAS *p*-values.

To assess how much phenotypic variance was explained by associated regions, we derived genetic relationship matrices for the set of SNPs clumped for suggestive association ($p= 1\text{x}10^{-6}$) with each trait and the set of all other SNPs, and estimated the partitioned heritability to measure the proportion of total heritability unattributed by discovered associations.

Just as with the heritability estimates, we also ran mixed linear model-based associations fitting the first 10 principal component eigenvectors as covariates and excluding 46 related individuals to achieve added control over population structure. Overall, we found small shifts in the top associated loci in this context. Out of 588 associations (1,388 SNPs of $p<1x10^{-6}$) detected in the original studies for behavioral question and factor scores, 71/109 SNPs (65.1%) were still significant (p<5x10-8) and 1,145/1,388 SNPs (82.5%) remained suggestive ($p<1x10^{-6}$). We report in data S16 when any associated locus did not replicate in this context ($p>1x10^{-6}$).

In the main text, we present the association results without PCs fitted as it is not clear that including PCs is more accurate and doing so may constitute an over-correction. The mixed linear model-based association (MLMA) analysis already controls for population stratification and cryptic relatedness as well or better than the corrections offered by principal components regression (*130*,*131*). Currently, PC-based approaches are used when MLMA is not computationally feasible (*132*). Principal component regression can be regarded as an approximation of a linear mixed model and may largely duplicate corrections already incorporated into the analysis (*133*). Furthermore, unlike using kinship, selecting the number of PCs to include is subjective and could thereby bias results (*134*).

Size prediction model

We used responses to question 121, "When DOG is standing next to someone of average height, how high are HIS shoulders?" with the answer options of ankle high or shorter ("0", 3% of dogs), calf-high ("1", 30%), knee-high ("2", 40%), thigh-high ("3", 24%), or hip-high or higher ("4", 3%), as phenotypes for building predictive models for body size.

A set of 1,730 adult dogs over 18 months old from the Darwin's Ark cohort was used as the training and testing set. To assess the power of the prediction model, we did ten-fold cross validation in which we split the data into 10 folds. At each round, 9 folds of dogs were taken as the training set, and the rest were taken as the testing set. The validation set includes a cohort of dogs recruited from the 2017 Somerville Dog Festival in Somerville, MA, with height to withers in inches measured at the event and a cohort of dogs in which owners were instructed to measure the height from their dog's shoulder to the ground (see *Validation of size phenotypes*). In total, 95 dogs were included in the validation set in addition to the training and testing set. We converted these measurements to the height question Likert scale by linear interpolation where {ankle-high: 0 = 4 in, calf-high: 1 = 10 in, knee-high: 2 = 18 in, thigh-high: 3 = 25 in, hip-high: 4 = 35 in from floor} according to body measurements of a 5'4" female human.

At each round, we performed a genome-wide association study on the training set, chose the predicted SNPs based on their strength of association, and built a random forest regression model using the selected SNPs. Once the model was built, we assessed prediction accuracy using the testing set to get the prediction accuracy and mean squared error. The final prediction accuracy and mean squared error were averaged across the ten testing sets. We iterated through different *p*-value cutoffs, from $1x10^{-8}$ to $1x10^{-3}$, when selecting predicted SNPs and chose the cutoff with the best performance on the testing set. We built the model with the *R* package `` `randomForest` `` and selected a *p*-value cutoff ($1x10^{-5}$) and mtry and ntree parameters (576 and 500, respectively, determined by iteration) to achieve the best model performance. We built a single model with 1,730 dogs and 2,733 SNPs and validated this model using our validation set of 95 dogs independent from the 1,730 adult dogs described above as the training and testing sets.

Our random forest regression model predicted the size to be numeric and we then rounded the decimal numbers to integers to match the survey measurement from 0 to 4. Besides building models based on GWAS-selected SNPs, we built models based on randomly selected SNPs as a control. We followed the same pipeline by splitting the data into ten folds. At each round, we randomly selected a matched number of SNPs to each *p*-value cutoff and built models on this. Prediction accuracy and mean squared error were calculated the same way described above.

Gene sets enrichment analysis

To assess the enrichment of sets of functionally related genes, we applied MAGMA (version 1.09) (*91*) on the GWAS summary statistics for 136 traits, including 14 physical traits, 114 behavioral questions, and 8 behavioral factors. MAGMA calculated gene-wide *p*-values by combining the *p*-values of all SNPs inside genes while accounting for gene size, number of SNPs in a gene, and LD. Using gene-based *p*-values, it tested for enrichment of association signals in genes belonging to the same set.

Gene sets were compiled from three main sources. First, curated neuropsychiatric genes. The autism spectrum disorder (ASD) gene set included 820 genes from the SFARI database which centers on genes implicated in autism susceptibility (*88*). The obsessive-compulsive disorder (OCD) gene set comprised 62 manually curated genes from OCDB, a database collecting genes, miRNAs and drugs for OCD (*87*). The schizophrenia (SCZ) gene set combined genes from two studies, the PGC2 GWAS in 2014 (*89*) and the UK CLOZUK GWAS in 2018 (*90*), which gave us a total of 371 genes. Second, to test whether dog size associated loci are enriched in GWAS for dog non-physical traits, we compiled a set of dog size-associated regions based on our size GWAS, the final set contains 75 regions with 4.6Mb in total. Third, top genes expressed in GTEx tissues (*85*). We compiled a GTEx gene set by choosing the top 100 expressed genes in 25 tissues and 13 brain subregions. For arteries, brain, cervix, colon, esophagus, heart, kidney, and skin that were sequenced at subregion level in GTEx, we merged the subregions into single tissue by taking the average as the expression level for each gene, then chose the top 100 genes. We reported raw *p*-values from MAGMA for each GWAS-gene set pair, as well as gene-wide *p*-values for significant enrichment (data S17).

Genetic differentiation of breeds

We defined a set of 27,674,130 SNPs genotyped in at least 80% of dogs and 50% of wolves. Samples from Darwin's Ark were assigned either to a breed population determined by ancestry inference ($\geq$ 85% breed) or labeled as `other_dog`. All available sequencing and imputed array data collected from purebred samples were assigned to their respective populations (data S4). We calculated genome-wide normalized population branch statistics (PBS, implemented in the Python package `*scikit-allele*`) using the Hudson estimator of fixation (F_{ST}) for each breed ($N \geq 12$ dogs) relative to dogs overall and wolves in sliding windows of 100kb by 10kb containing at least 10 SNPs (*92*, *135*, *136*). Across all windows, the genome-wide empirical *p*-values based on rank relative to other windows ((r+1)/(n+1)) were calculated for all test statistics (*122*). A summary of the per-breed population branch statistic is in data S18.

For each trait-associated locus ($p < 1 \times 10^{-6}$) from our genome-wide association studies, we permuted the region of combined overlapping LD-based clumps ~100,000 times on the same chromosome using BEDTools *shuffle* function, mapped with BEDTools *map* the max(PBS) for each real locus against the distribution of permuted loci for each breed (data S19), and generated z scores. We divided locus tests into physical trait, behavioral question, and behavioral factor associations, and performed a one-tailed Student's t-test for whether the observed max(PBS) in associated loci exceeded that which we expect by random chance, and corrected for multiple testing using the Benjamini-Hochberg procedure (data S20).

We also tested whether the allele frequencies at the most associated SNPs tended to differ more in breeds than at SNPs sampled at random across the whole genome, as might be expected if traits were breed-differentiated (fig. S34). For each SNP, we calculated the max F_{ST} observed between one of the top ten breeds and the full dog population. We measured whether SNPs associated with behavioral traits and physical traits tended to have higher max F_{ST} than 29,903 randomly sampled SNPs using a one-sided t-test (*R* package `` `stats` `` version 4.1.1). For physical traits, the maximum F_{ST} observed for the top SNP in each locus (1mb) was slightly higher than for random SNPs (mean = 0.150 vs. 0.145; t-test $p_{1\text{-}sided}$=0.026) but this was not true at behavioral trait loci (mean = 0.142 vs. 0.145; t-test $p_{1\text{-}sided}$=0.95). For physical traits, the most strongly associated loci (p<1e-12) tended to have higher breed F_{ST}s, consistent with breed differentiation (mean = 0.33 vs. 0.145; t-test $p_{1\text{-}sided}$=0.0023).

Supplementary Text

Marketing of DNA tests for breed ancestry

To determine whether breed testing was being marketed as a behavioral predictor, we searched the most popular genetic testing services and media articles for references to determining behavior, personality or trainability from a DNA test and found several examples:

A. Wisdom Panel:
"Consider learning more about their breed background with a Wisdom Panel™ DNA test. The insights you'll gain will allow you to tailor training and care to your dog's unique needs and preferences—helping you build your special bond."
Source: Why dogs have favorite people (and how to make sure your pup picks you). https://web.archive.org/web/20210705161645if_/https://www.wisdompanel.com/en-us/blog/why-dogs-choose-favorite-people (2021).

B. Amazon (marketing for Embark):
"Seeing the breeds your dog inherited will help you understand their personality along with what keeps them happy. Do they need a lot of exercise? Are they food-motivated? Do they like a good brain tease? Adapt your care routine based on their breed results."
Source: Embark | Dog DNA Test | Breed & Health Kit | Breed Identification & Canine Genetic Health

Screening. https://web.archive.org/web/20210705162556/https://www.amazon.com/dp/B01EINBA76

C. Orivet Genetic Petcare:
"Collect your own Dog's DNA with a simple cheek swab. Orivet will compare your dog's DNA to hundreds of genetic markers of the most common known breeds to uncover … Insights into Your Dog's Personality and Behaviour"
Source: Orivet Genetics.
https://web.archive.org/web/20210122163410/https://www.orivet.com/store/canine-mixed-breed-screen/dog-breed-identification-test.

D. DNA My Dog:
"DNA My Dog's simple cheek swab DNA test lets you learn the breeds in your dog and gain insight into the unique genetic background of your dog including the history of their breed, personality traits, exercise levels, and so much more!"
Source: DNA My Dog: Fast, easy and completely painless.
https://web.archive.org/web/20210503213607/https://dnamydog.com/.

E. VCA Pet Hospitals:
"Every purebred has certain established physical and personality traits that provide owners with an idea of the type of pet they are getting. Knowing the breeds that go into a mix can help the owners make better guesses about the size, temperament, energy levels, and exercise requirements their unique pet may have."
Source: Genetic (DNA) Testing.
https://web.archive.org/web/20200926224943/https://vcahospitals.com/know-your-pet/gene

<u>Overview of sample sizes and inclusion/exclusion criteria for all analyses</u>

The following describe the source of each data set and samples, the inclusion and exclusion criteria, and maximum final sample size for all analyses. Several analyses vary in sample size depending on how many dogs have survey responses for a given trait.

Data Set	Criteria	Sample Size	In Figure 1F
Darwin's Ark: Full Survey Set	Enrolled prior to November 15, 2019	18,385 dogs	surveyed
Darwin's Ark: Survey Set with Covariates	Dogs in "Darwin's Ark: Full Survey Set" phenotyped for age, sex, and height	14,327	sex, age, and size info
Phenotype Validation: Concordance with DPQ	Owner completed all survey questions derived from published Dog Personality Questionnaire	10,253 dogs	
Phenotype Validation: Owner-measured Height	Owner completed a size measurement worksheet and responded to size survey question	337 dogs	
Phenotype Validation: Staff-measured Height	Dog was measured at in-person event	38 dogs	
Phenotype Validation: Breed-average Height	Owner reported dog to be a purebred of a breed with average heights reported by kennel club (90 breeds)	2,025 dogs	

Exploratory Factor Analysis: Discovery Set	Owner answered the first 110 survey questions	10,253 dogs	
Exploratory Factor Analysis: Scored Set	Discovery set and any dog with responses to >80% of first 110 questions	16,522 dogs	
Mendel's Mutts	Dogs of unknown breed ancestry that underwent high coverage, whole genome sequencing	27 dogs	
Darwin's Ark: Full Genetic Set	Dogs genotyped or sequenced prior to November 15, 2019	2,155 dogs	genetic data
Darwin's Ark: Genotyped Set	Dogs genotyped on the Axiom Canine Genotyping Array Set A & B prior to adoption of low-pass sequencing approach	440 dogs	
Darwin's Ark: Low-pass Set	Dogs with low coverage genome sequencing on Gencove platform	1,715 dogs	
Heritability Analysis	In "Darwin's Ark: Full Genetic Set" and in "Darwin's Ark: Survey Set with Covariates", and phenotyped for trait	varies by trait, at most 1,967 dogs	
Breed Classification: Confirmed Purebred	Either described as registered purebred by the owner, or confirmed by sequencing to have 85% or more ancestry from a single breed	3,637 dogs	confirmed breed
Breed Classification: Candidate Purebred	All confirmed purebred dogs ("Breed Classification: Confirmed Purebred") as well as dogs with owner-reported ancestry from one breed	9,009 dogs	candidate breed
Breed Classification: Mutt	All dogs not classified as either confirmed or candidate purebred.	9,376 dogs	mutt
Breed Classification: Highly Admixed Mutt	All dogs in "Darwin's Ark: Full Genetic Set" with <45% ancestry from any single breed in admixture analysis.	1,071 dogs	
Analysis of Variance	In "Breed Classification: Candidate and Confirmed Purebred" and in "Darwin's Ark: Survey Set with Covariates" and in "Darwin's Ark: Full Genetic Set"	6,797 dogs	ANOVA pool
Analysis of Variance: Confirmed Set	In "Breed Classification: Confirmed Purebred" and in "Darwin's Ark: Survey Set with Covariates". Phenotyped for trait in question.	varies by trait, on average 2,251±119 dogs across 78 breeds	

Analysis of Variance: Candidate Set	In “Breed Classification: Candidate and Confirmed Purebred” and in “Darwin’s Ark: Survey Set with Covariates”. Phenotyped for trait in question.	varies by trait, on average 5,968±332 dogs across 121 breeds	
Relative Risk Analysis	In “Breed Classification: Candidate and Confirmed Purebred” and phenotyped for trait in question.	varies, at most 266 dogs for a breed x factor pair	
Population Peculiarity Scoring	All dogs in “Darwin’s Ark: Full Survey Set”. Groupings described in methods.	18,385 dogs	PPS pool
LMER Analysis	Dogs in “Darwin’s Ark: Survey Set with Covariates”, “Exploratory Factor Analysis: Scored Set”, and in “Breed Classification: Highly Admixed Mutt”	1,002 dogs	LMER pool
MuttMix Survey: Dogs	Dogs in “Darwin’s Ark: Low-pass Set” or “Mendel’s Mutts”. Owners did not know their dog’s ancestry and submitted photos and videos.	31 dogs	
MuttMix Survey: Participants	All participants	26,639 people	
Genome-wide Association Analysis: All Dogs Set	In “Darwin’s Ark: Full Genetic Set” and in “Darwin’s Ark: Survey Set with Covariates”. Phenotyped for the trait in question.	varies by trait, at most 1,967 dogs	GWAS pool
Genome-wide Association Analysis: Highly Admixed Set	In “Breed Classification: Highly Admixed Mutt” and in “Darwin’s Ark: Survey Set with Covariates”. Phenotyped for the trait in question.	varies by trait, at most 1,071 dogs	
Size Prediction Models: Training and Testing Set	In “Darwin’s Ark: Full Genetic Set” and phenotyped for height. Dogs younger than 18 months were excluded.	1,730 dogs	
Size Prediction Models: Validation set:	In “Darwin’s Ark: Full Genetic Set” and either "Phenotype Validation: Owner-measured Height" or "Phenotype Validation: Staff-measured Height". Dogs younger than 18 months were excluded.	95 dogs	

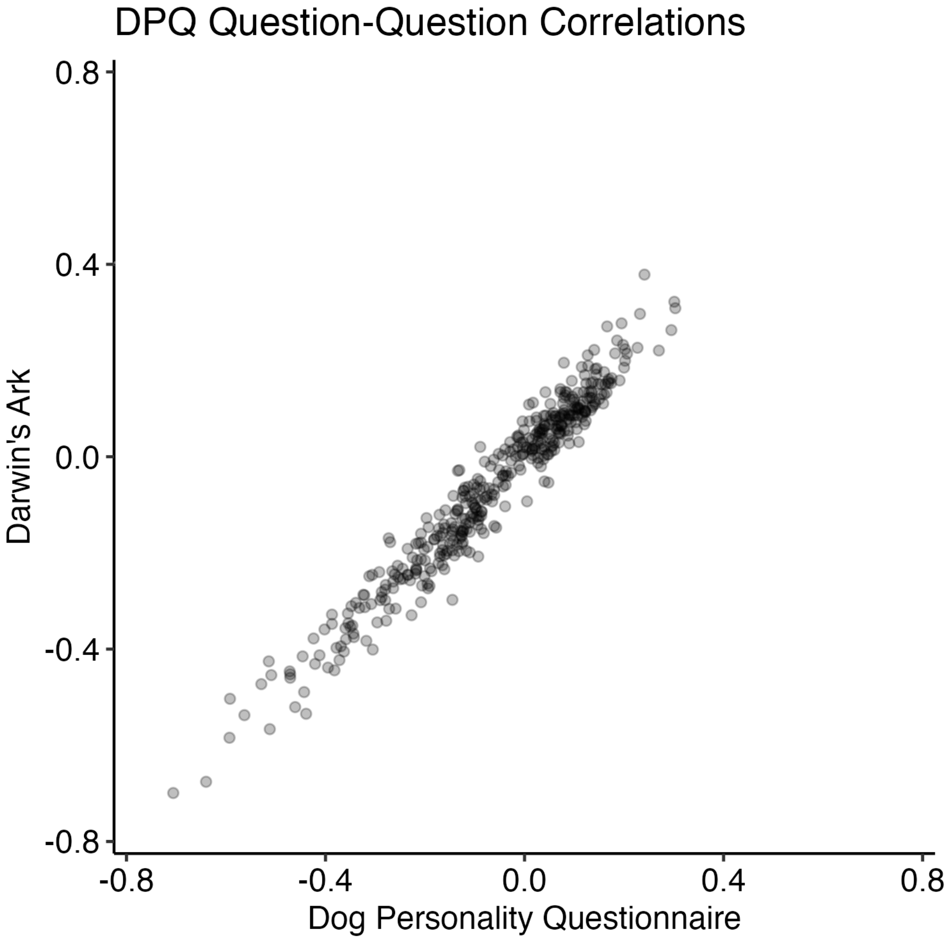


Fig. S1. Concordance between original Dog Personality Questionnaire (DPQ) and Darwin's Ark implementation of DPQ items. The Dog Personality Questionnaire (DPQ) published question x question correlation data from 2,556 dogs for 48 questions also included in Darwin's Ark (*37*). We extracted the correlation distances ($d = 1 - |r|$) for responses from 10,254 dogs in Darwin's Ark and performed a Mantel's test between the DPQ and the Darwin's Ark correlation distances using *R* package `ade4` with 100,000 replicates. The correlation was nearly perfect (Mantel's correlation=0.95; $p = 1 \times 10^{-7}$).

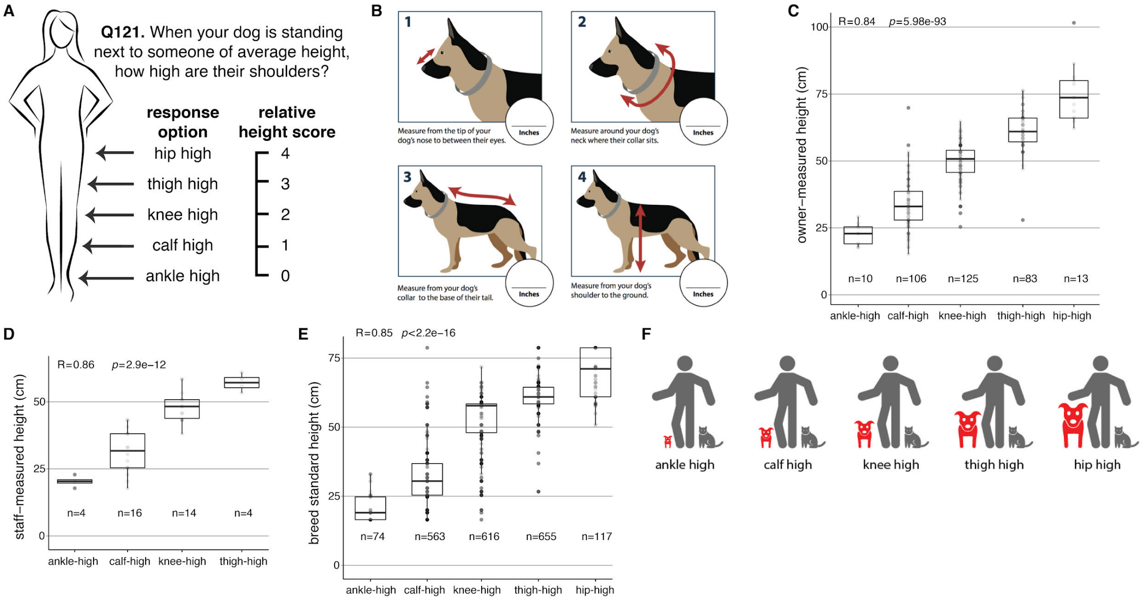


Fig. S2. Validation of surveyed size phenotypes against real measurements and reference data. (**A**) Darwin's Ark question 121 asks owners to select their dog's shoulder height relative to an average person on a five-step scale, with the options from "ankle-high or shorter" to "hip-high or higher". The survey response is converted into a relative height score ranging from 0 to 4. (**B**) For a subset of dogs, we asked owners to measure their dogs using a measuring tape and worksheet we provided. Panel 4 (height-to-withers) is the measurement coarsely captured by question 121. (**C**) Owner-measured height-to-withers is strongly correlated with owner responses on question 121 (*N*=337 dogs; R_{pearson}=0.84; t=28.885, df=335, *p*=$5.98\text{x}10^{-93}$). (**D**) Staff-measured height-to-withers for 38 dogs (collected at the Somerville Dog Festival in 2017) is strongly correlated with owner responses on question 121 (*N*=38 dogs; R_{pearson}=0.86; t=10.284, df=36, *p*=$2.9\text{x}10^{-12}$). (**E**) Breed average height, as defined in the AKC breed standard, is strongly correlated with owner responses on question 121 for purebred dogs (N= 2,025; R_{pearson}=0.85; t=71.268, df=2,023, *p*<$2.2\text{x}10^{-16}$). (**F**) For the 31 dogs in MuttMix, participants were given each dog's size on the Q121 scale, as illustrated using the graphic shown.

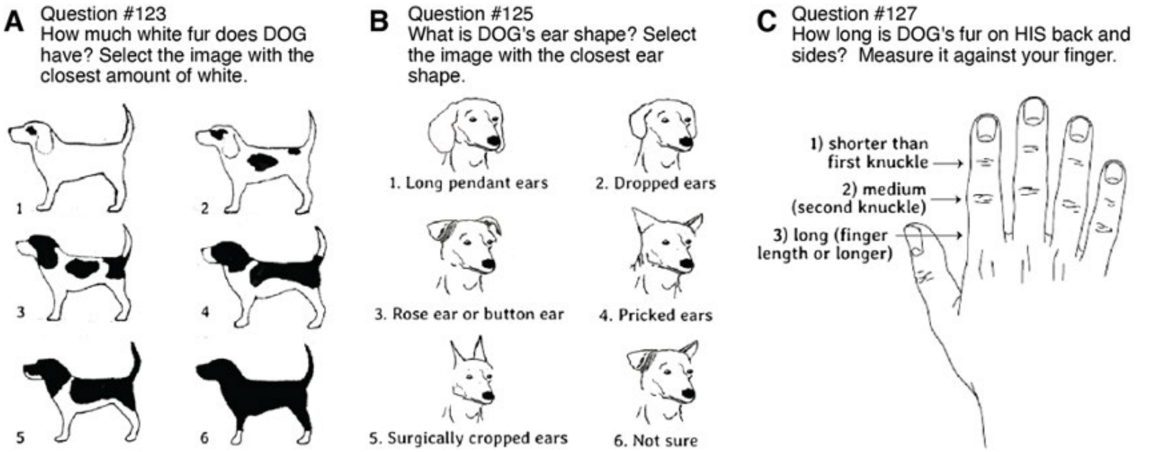
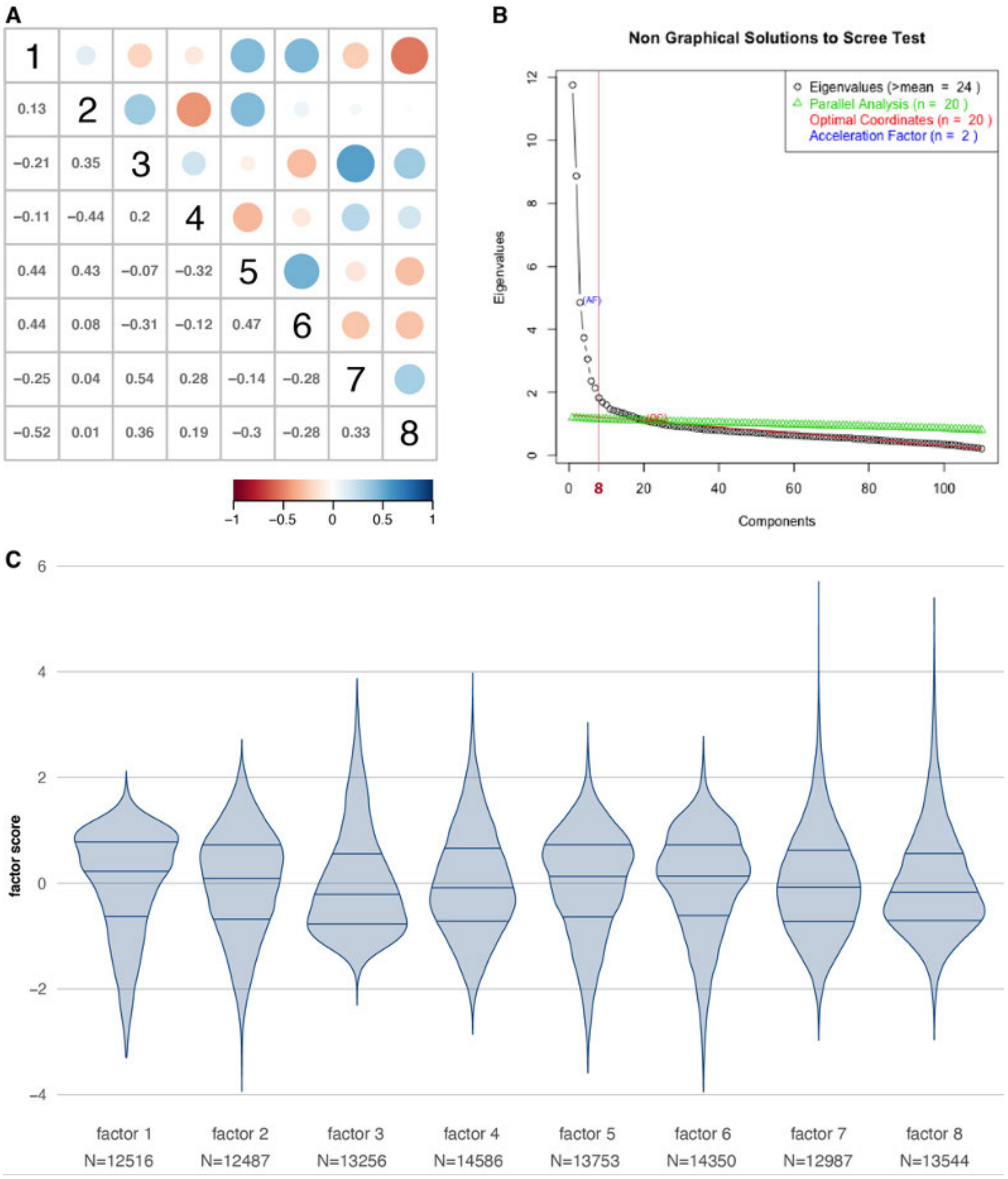


Fig. S3. For three physical trait questions other than size (table S1), owners were asked to choose from a range of possible options shown in an accompanying image. Questions were designed to be as easy to answer as possible**.** (**A**) For question 123 (white coat color), options range from all white to all black. Images were sourced from (*137*). (**B**) For question 125 (ear shape), options range from long and dropped to pricked. Two responses (surgically cropped ears and not sure) were excluded from subsequent analyses. (**C**) For fur length, owners are asked to respond using an easy-to-access measuring tool - their own finger. One other physical trait had multiple options but did not have an accompanying image. For coat color (not shown) owners chose one or more options from a list of colors including: (1) white; (2) red (from pale peach to dark red or liver colored); (3) yellow (from pale cream to orange, gold, fawn, or wheaten); (4) gray (slate, blue gray, charcoal, or silver); (5) chocolate brown; (6) pure black; (7) merle; (8) brindle.



1							
0.13	2						
−0.21	0.35	3					
−0.11	−0.44	0.2	4				
0.44	0.43	−0.07	−0.32	5			
0.44	0.08	−0.31	−0.12	0.47	6		
−0.25	0.04	0.54	0.28	−0.14	−0.28	7	
−0.52	0.01	0.36	0.19	−0.3	−0.28	0.33	8

Fig. S4. Exploratory factor analysis. To reduce the dimensionality of the survey data, and identify underlying associations between questions, we performed exploratory factor analysis on 1,127,830 total survey responses from 110 questions for 10,253 dogs. (**A**) Between-factor correlations show most are only weakly or moderately correlated. Factor numbers are listed along the diagonal, and Pearson's r values below the diagonal. (**B**) Scree plot for survey data used to predict optimal number of factors; test algorithms include eigenvalues, parallel analysis, optimal coordinates, and acceleration factor. The first 8 factors explained a cumulative 24.26% of variance. (**C**) Distribution of factor scores for each factor in Darwin's Ark: Full Survey Set, with the three horizontal lines marking quantiles 0.25, 0.5 and 0.75.

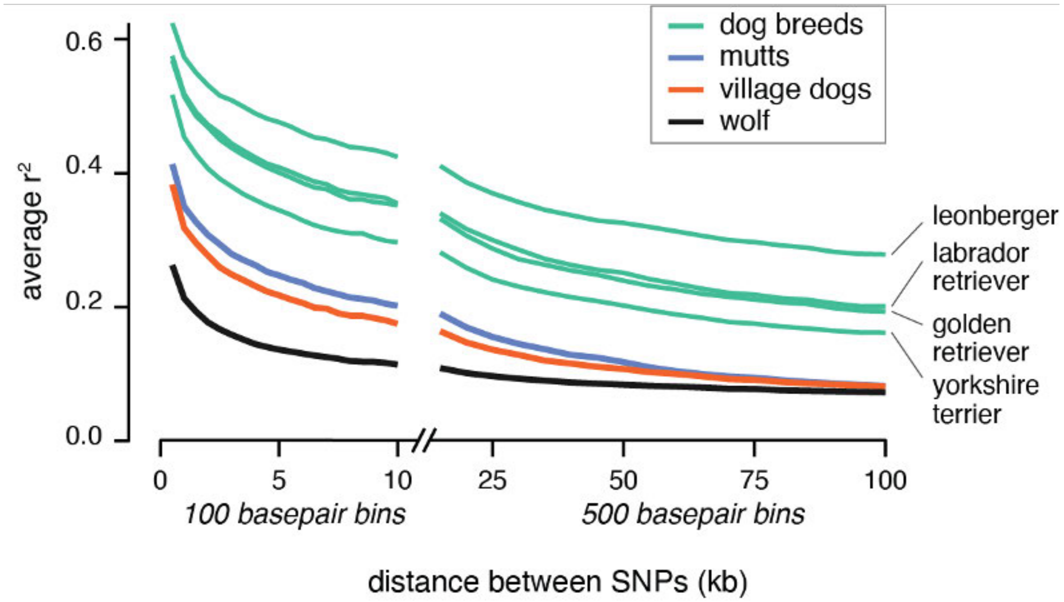
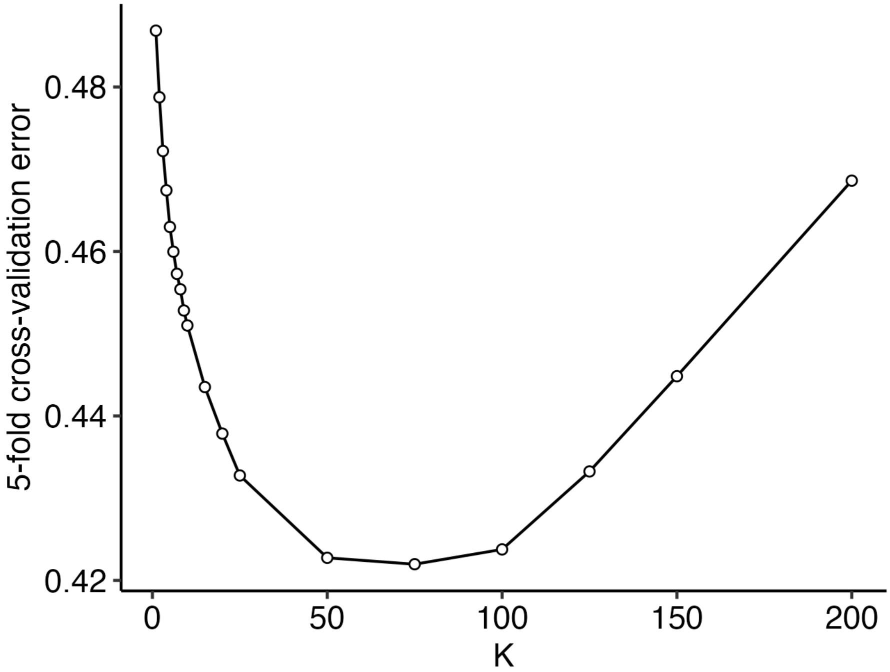


Fig. S5. Decay of linkage disequilibrium (LD) in dog populations with different histories and in wolves. LD is shorter in mutts (blue) than in dog breeds (green) but slightly longer than in outbred village dogs (red) and substantially longer than in wolves (black). For each population, 25 dogs were chosen at random from all dogs/wolves in the population. We assessed the extent of LD by measuring r^2 between each of 20,000 randomly selected SNP and all variants with minor allele frequency > 0.025 within 100kb of the index SNP.



unsupervised ADMIXTURE analysis on chromosome 38
with LD-based pruning (50kb, r2>0.5)

Fig. S6. Cross-validation of unsupervised admixture analysis in the Darwin's Ark cohort (N=2,155). We tested a wide range of cluster numbers (*K*) and achieved the lowest 5-fold cross validation error between *K*= 50 and 100.

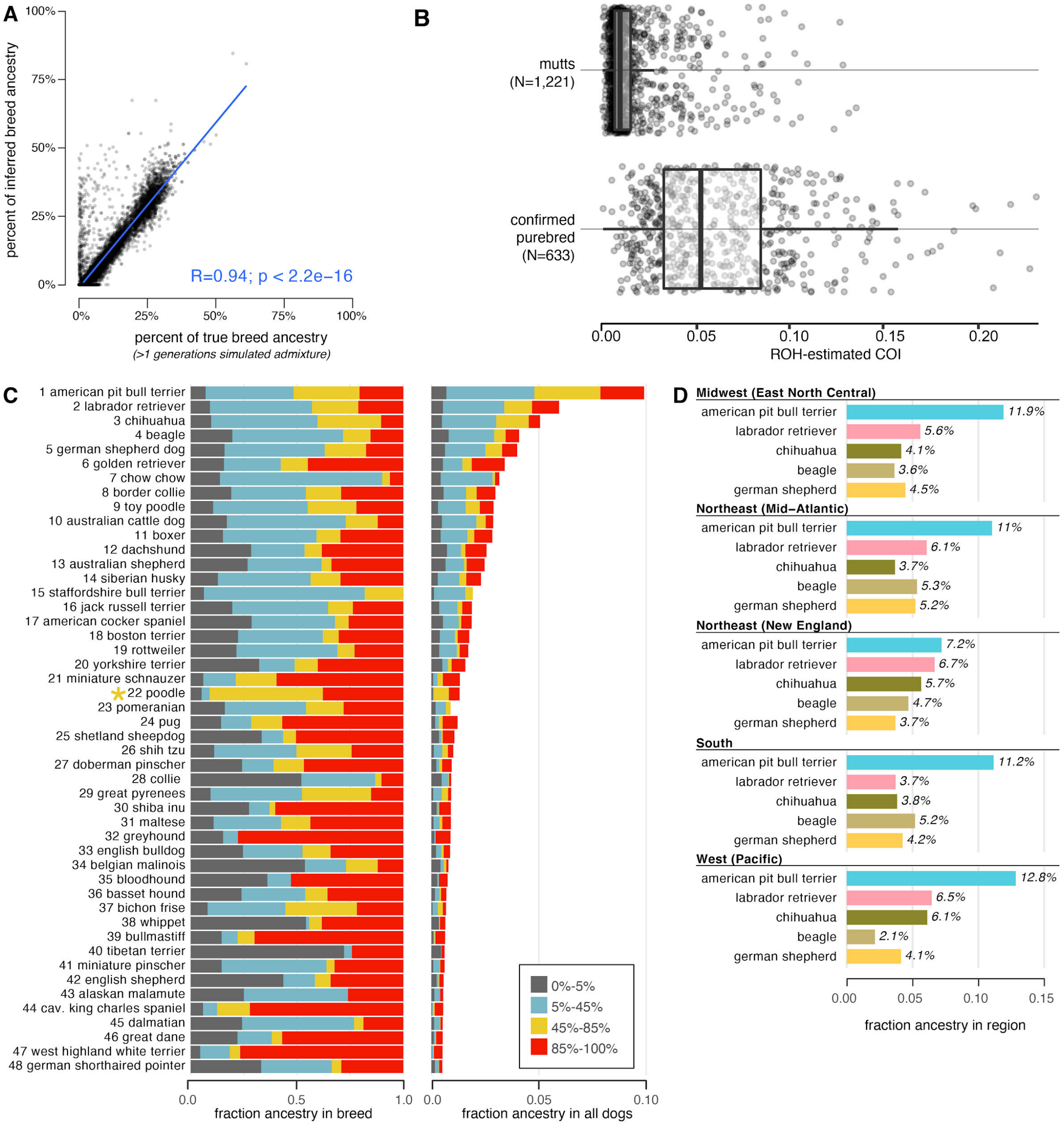


Fig. S7. Breed ancestry in mutts in the Darwin's Ark cohort. (**A**) Our breed calling pipeline accurately measures breed ancestry in simulated mixed breed dogs, with inferred ancestry strongly correlated with the known (simulated) breed ancestry. (**B**) Coefficients of inbreeding estimated from the total length of autosomal runs of homozygosity divided by SNP-covered total chromosome length run higher in confirmed purebred dogs than in mutts ($p_{\text{t-test}}$=1.7x10^{-122}; t=28.4, df=776.8). (**C**) The most common ancestry across all dogs is American pit bull terrier, followed by Labrador retriever and chihuahua. The proportion of the ancestry coming from purebred dogs vs. mutts varies by breed. An exceptionally high proportion of dogs with poodle ancestry (yellow asterisks) are 45-85% poodle, a range that includes F1 poodle crosses. (**D**) The proportion of ancestry by breed in the dog population varies by region of the United States, possibly reflecting differences in breed popularity in the past.

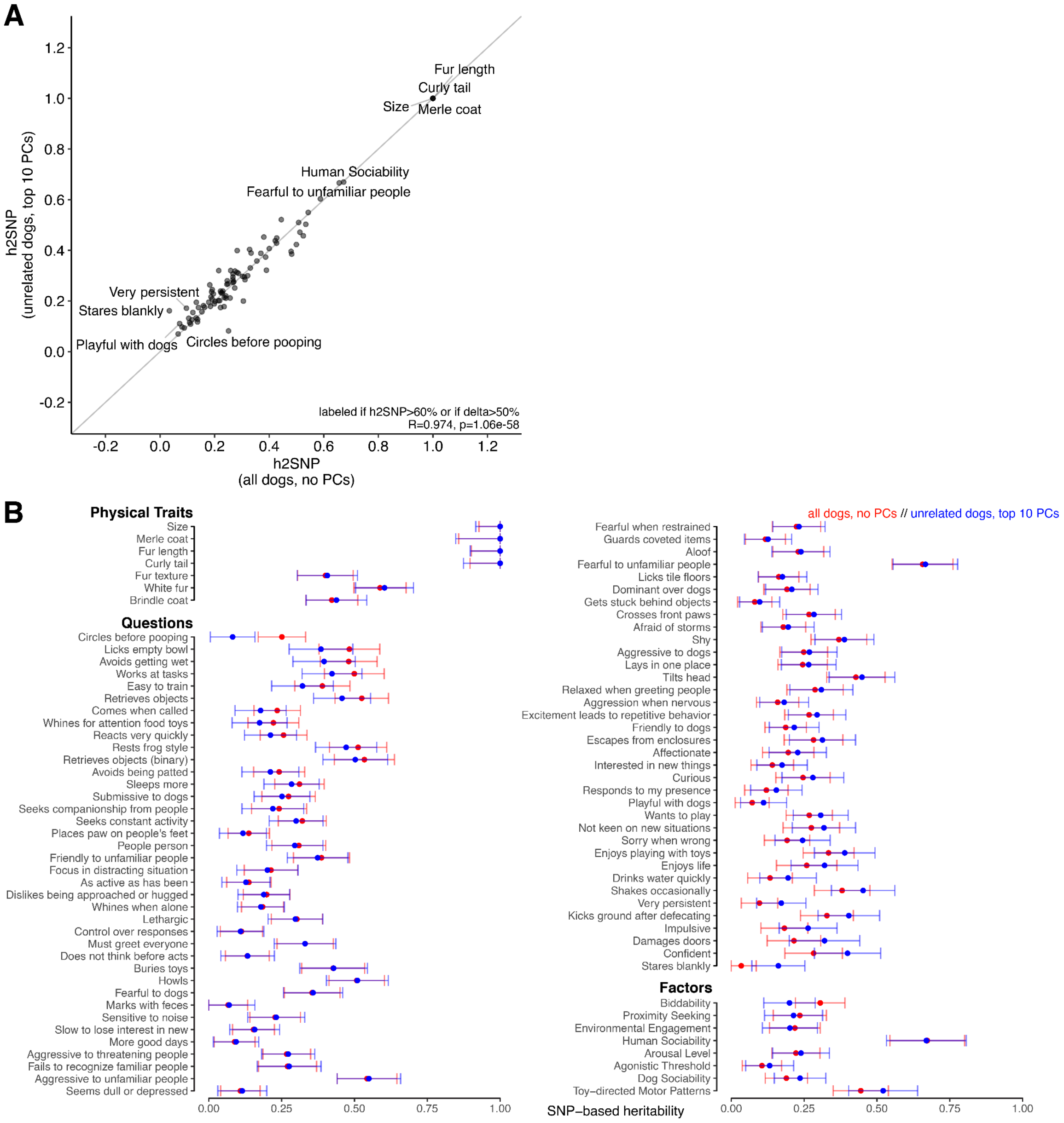


Fig. S8. Controlling for population structure via principal component analysis does not change SNP-based heritability estimates. We estimated heritability with standard errors using restricted maximum likelihood with LD score correction in two models: (1) all dogs with no principal component covariates (red) and (2) all dogs except 46 dogs from highly related pairs of kinship >0.2 with the top 10 principal components included (blue). (**A**) Heritabilities estimated by both models were highly correlated (R_{pearson}=0.974, p=1.06x10^{-58}). (**B**) The heritability estimates for physical traits did not differ between models, and the estimates for behavioral questions and factors rarely differed. The most dramatic shifts in heritability were observed for the question "Circles before pooping" and factor 4 (biddability).

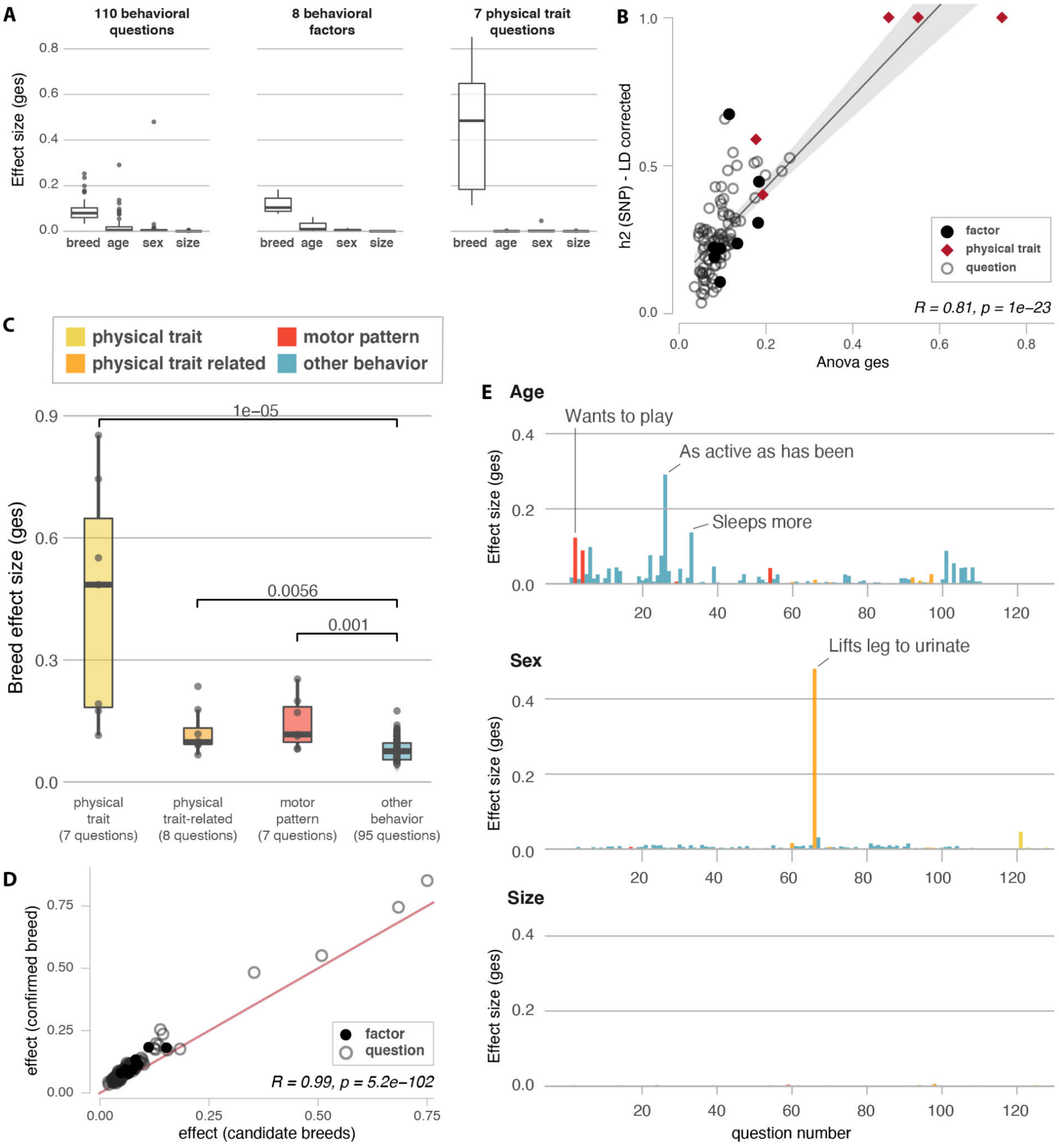


Fig. S9. Breed explains a larger fraction of the variance in behavior phenotypes than size, sex or age. (A) Breed has larger effect than age, sex or size, measured as generalized eta squared (ges), on both questions and factors, but the effect is still modest except for breed-defining physical traits. **(B)** The effect of breed is strongly correlated with the heritability of a trait in the pet dog population. **(C)** Breed effect sizes for physical traits, physical-trait related questions and motor patterns are significantly higher than for other behavioral questions (Wilcoxon test using alternative = "greater" with BH correction). **(D)** The variance explained by breed in ANOVA analysis of candidate purebred dogs (6,832 dogs in 122 breeds with >=5 dogs/breed) is strongly correlated with results from confirmed purebred set, suggesting that the less stringent classification adequately captures the effect of breed. **(E)** Age has ges>0.1 (labeled peaks) for three questions, sex for one question, and size for none.

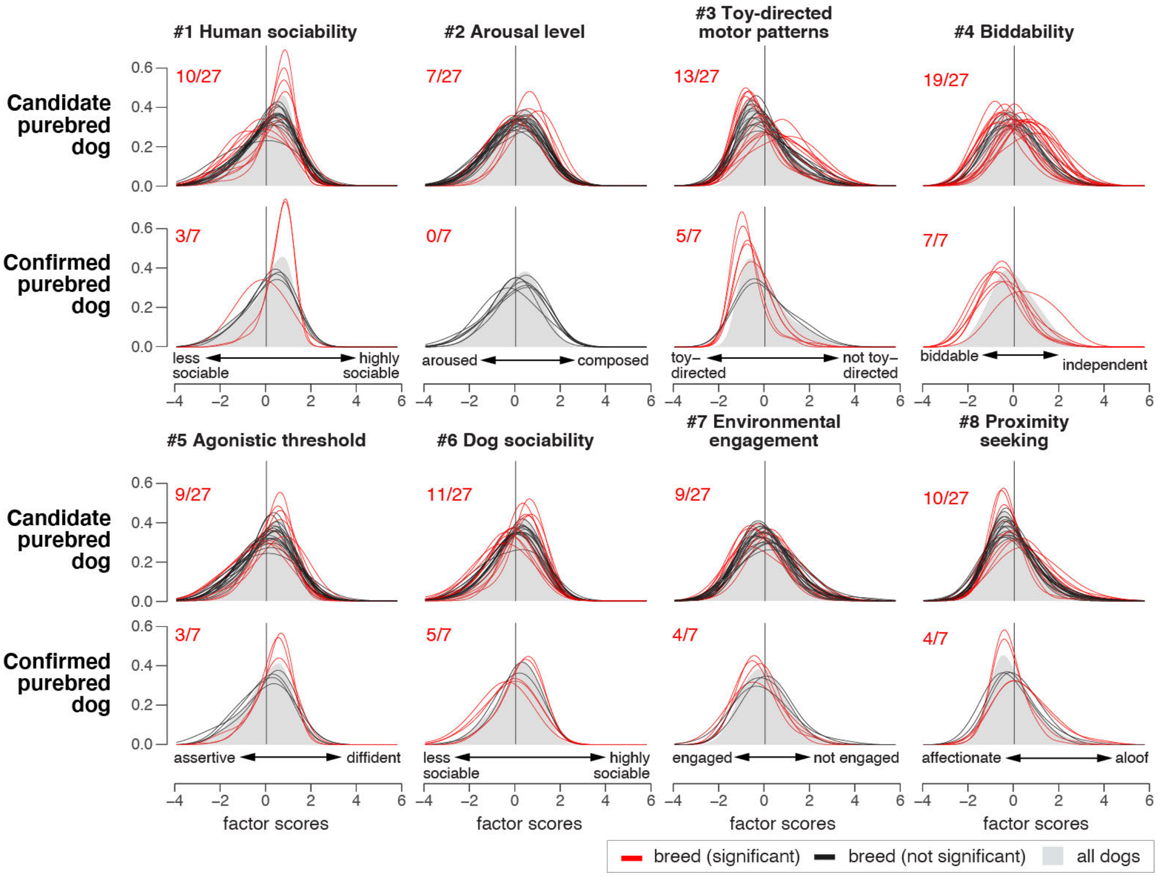


Fig. S10. Distribution of factor scores (based on owner surveys) in breeds differ somewhat from distribution in all dogs. For each factor, some breeds have distributions that are significantly different ($p_{\mathrm{BH}}<0.05$, red lines) from the distribution in all dogs (grey area), although most do not (black lines). Difference was measured using the Kolmogorov-Smirnov test, a nonparametric test of the equality of continuous distributions, with Benjamini-Hochberg FDR correction. Red text shows the number of breeds that are significantly different. All breeds with ≥50 dogs were included.

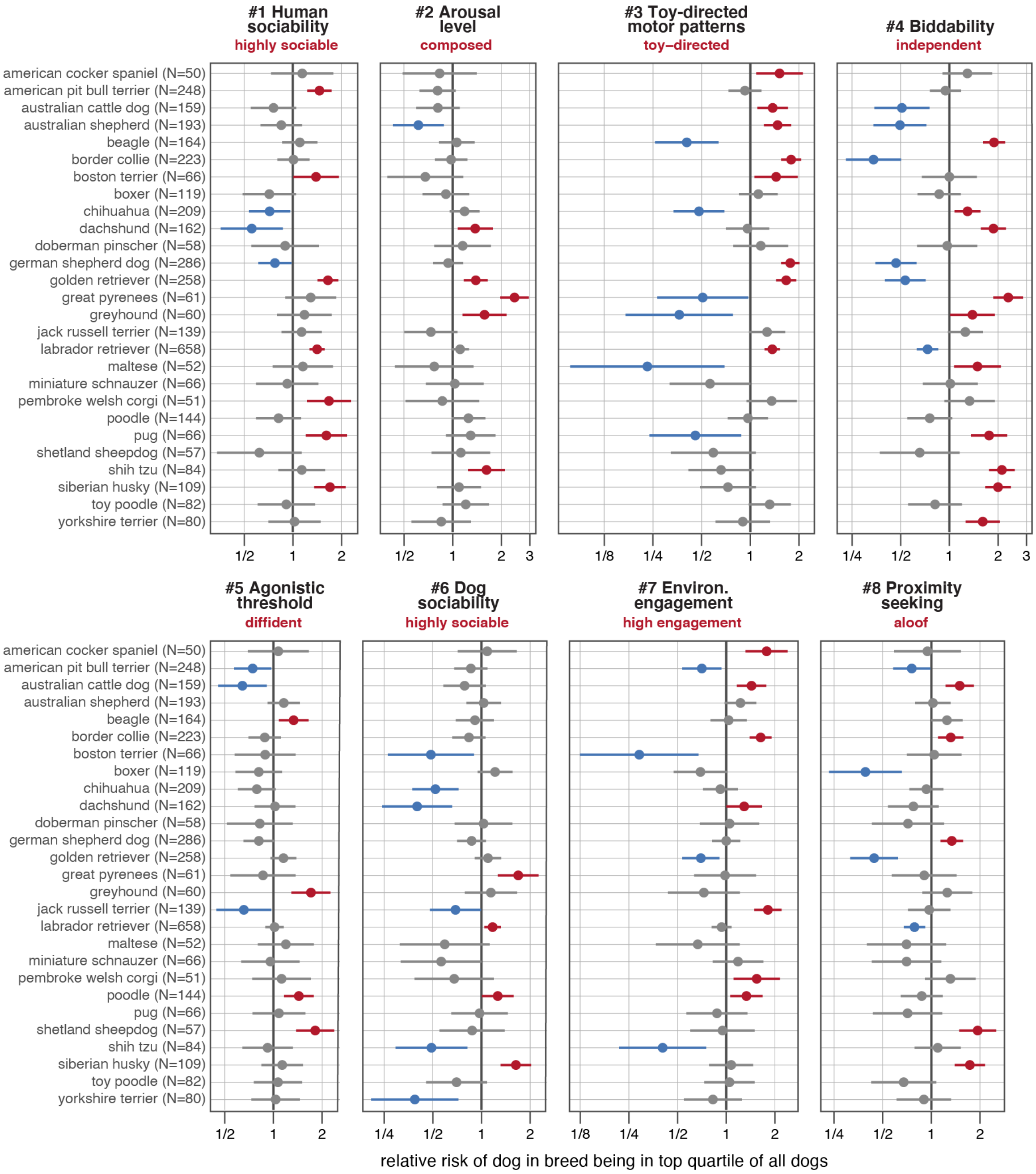


Fig. S11. Value of breed in predicting factors scores for individual purebred dogs. Fold difference (and 95% CI) in probability that a dog of a given breed will score in the top quartile for each factor, relative to a random dog. Red indicates fold probability significantly greater than 1; blue indicates significantly less than 1.

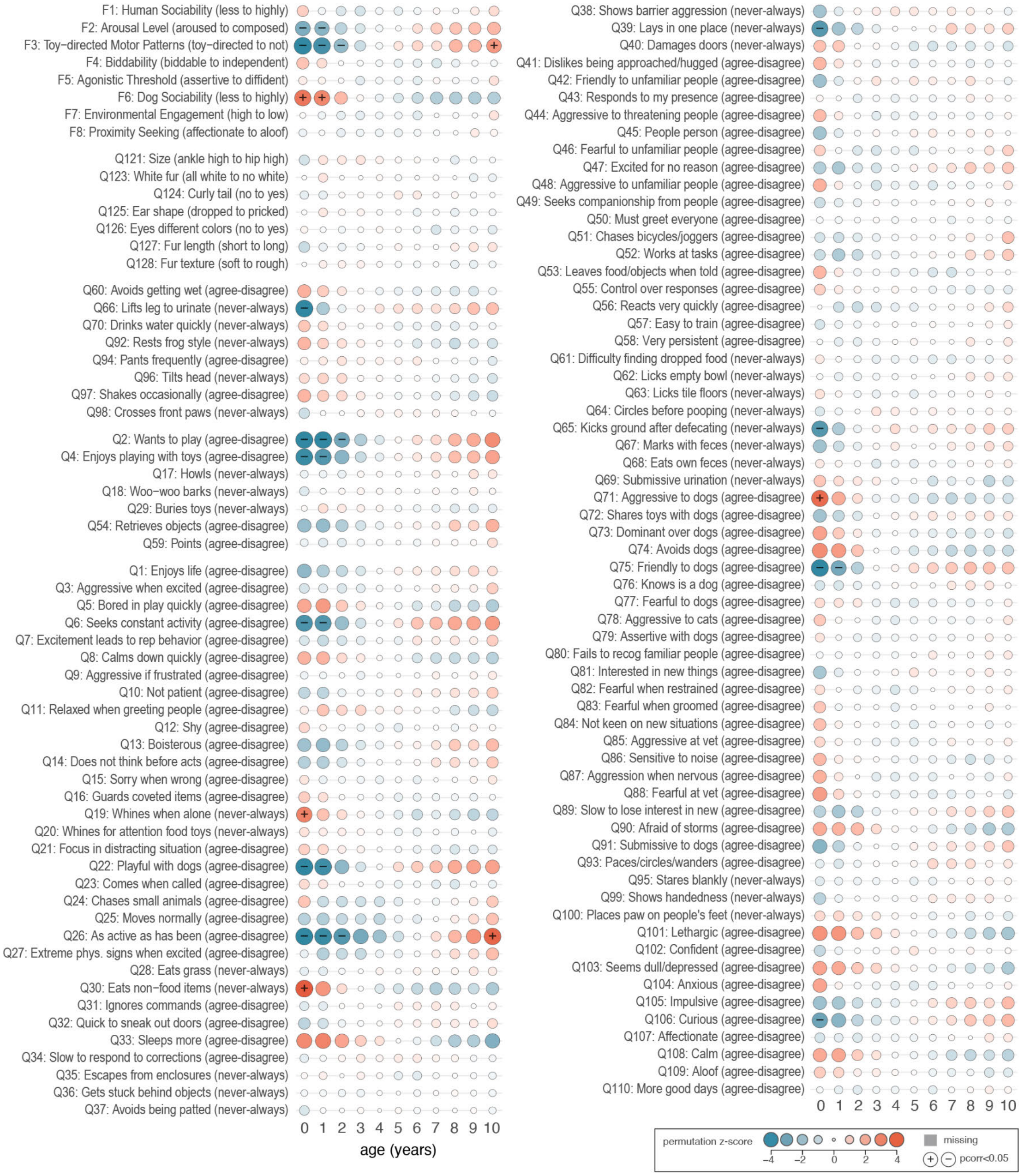


Fig. S12. Full set of population peculiarity scores (page 1/4). For each population tested, empirical z scores were measured over 500,000 random permutations. Circle size scales with abs(z) and color scales with z, with significant differences marked with "+" or "-". (**A**) Results comparing dogs of each year of age, rounded down (such that "Y01" is dogs between 0 and 1 years of age), to randomly sampled dogs, with sample size of 100.

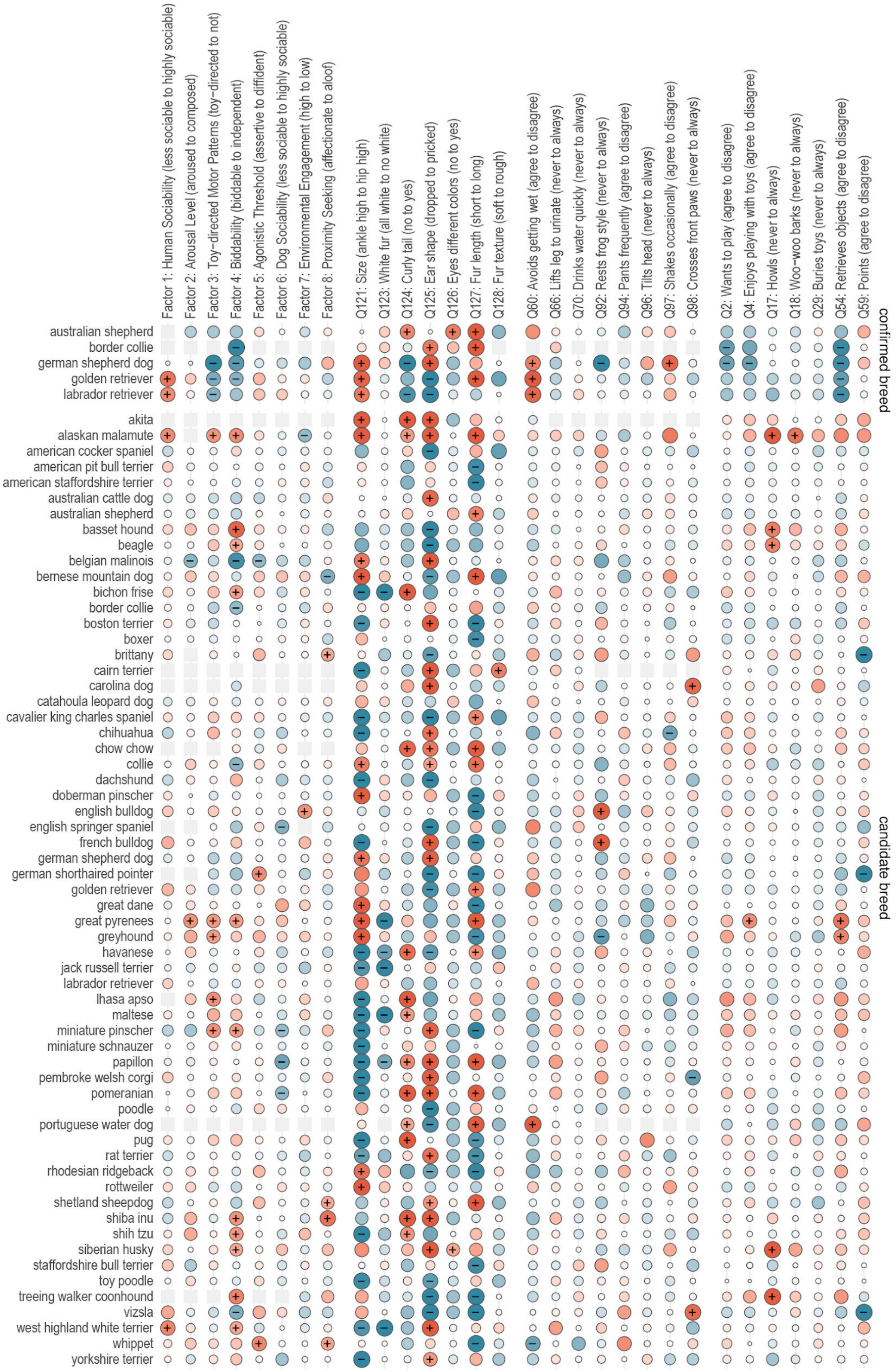


Fig. S12. Full set of population peculiarity scores (page 2/4). (B) Results comparing dogs from each breed to randomly sampled dogs.

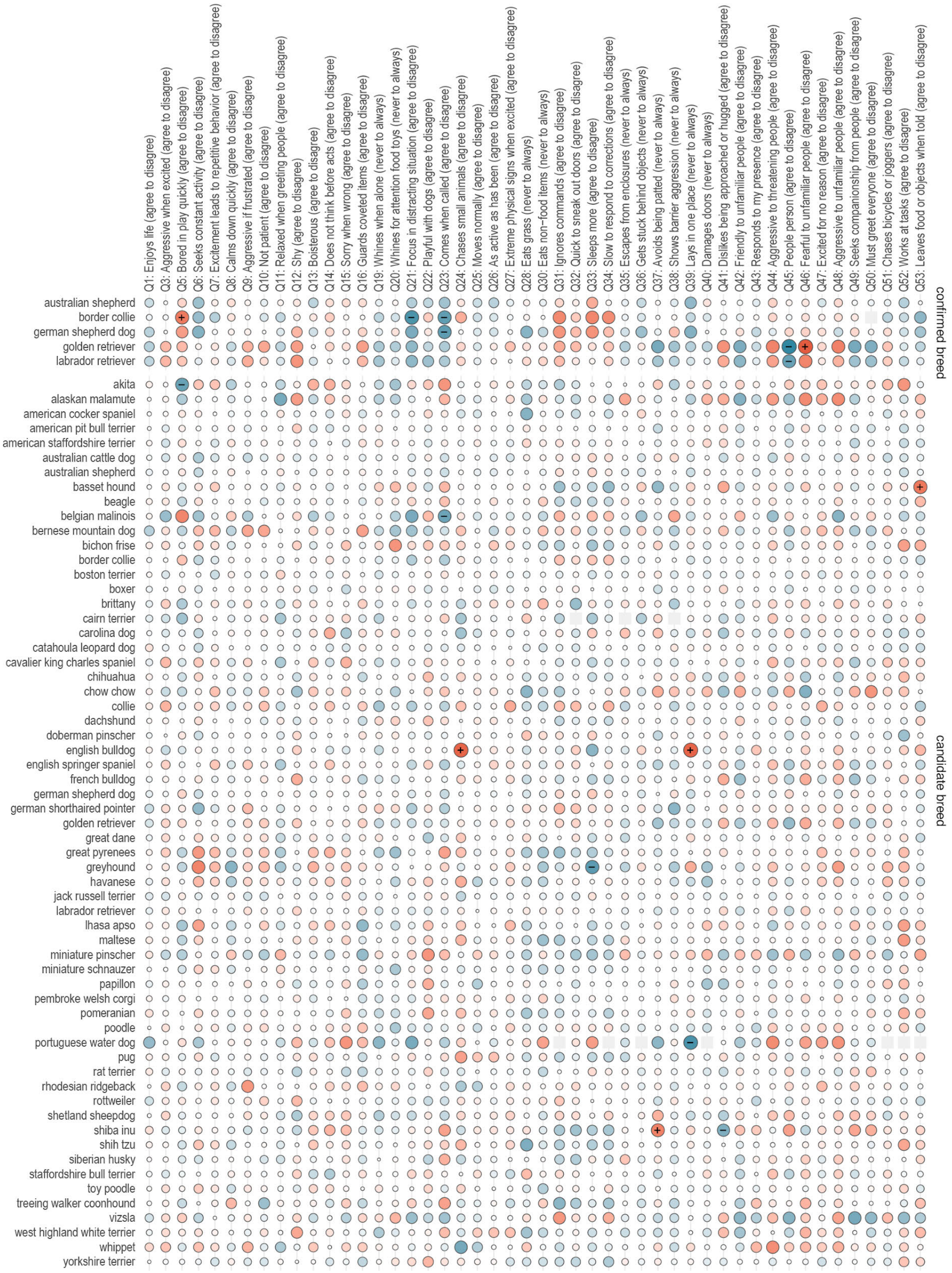


Fig. S12. Full set of population peculiarity scores (page 3/4). (B) Results comparing dogs from each breed to randomly sampled dogs.

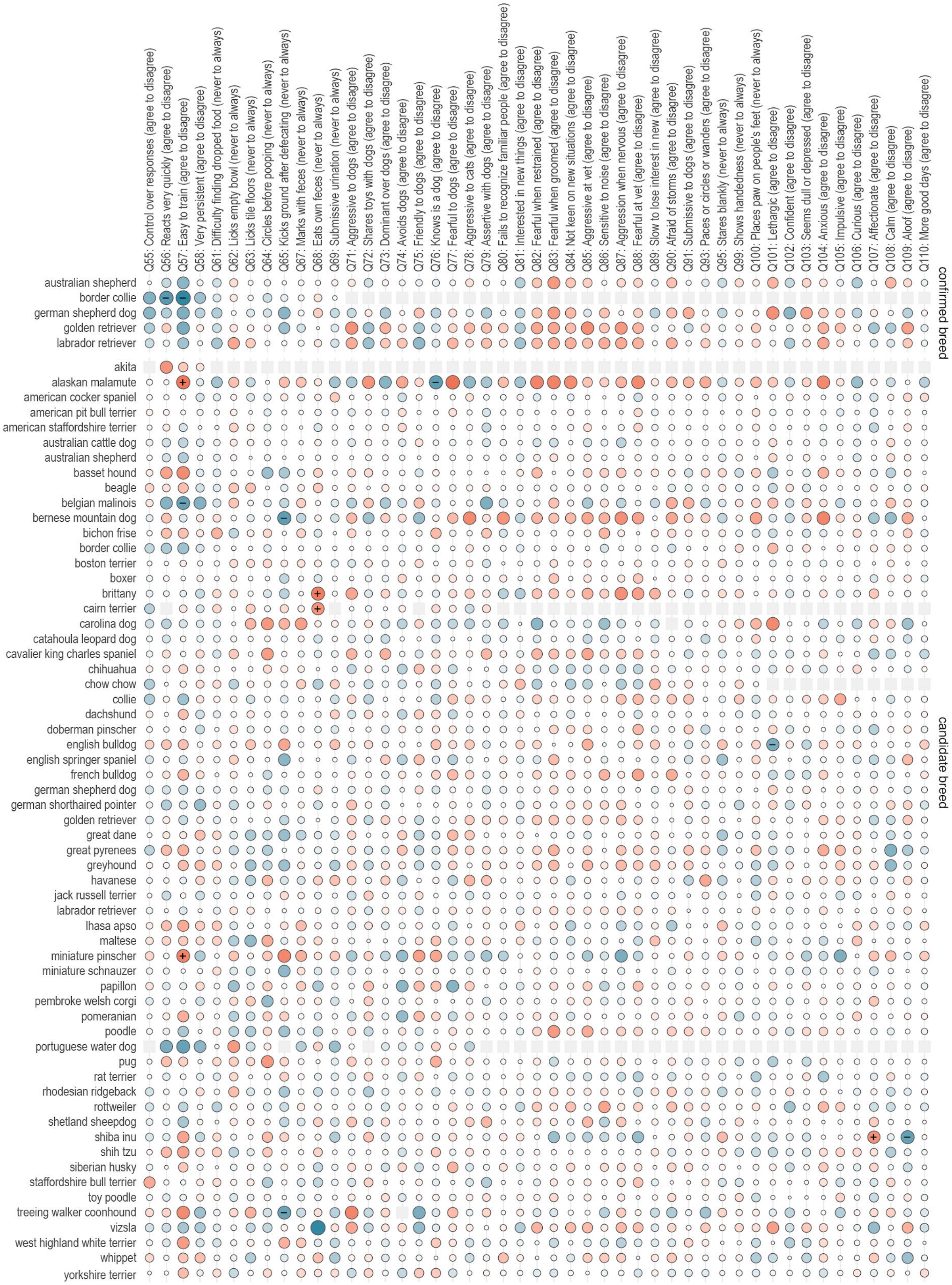


Fig. S12. Full set of population peculiarity scores (page 4/4). (B) Results comparing dogs from each breed to randomly sampled dogs.

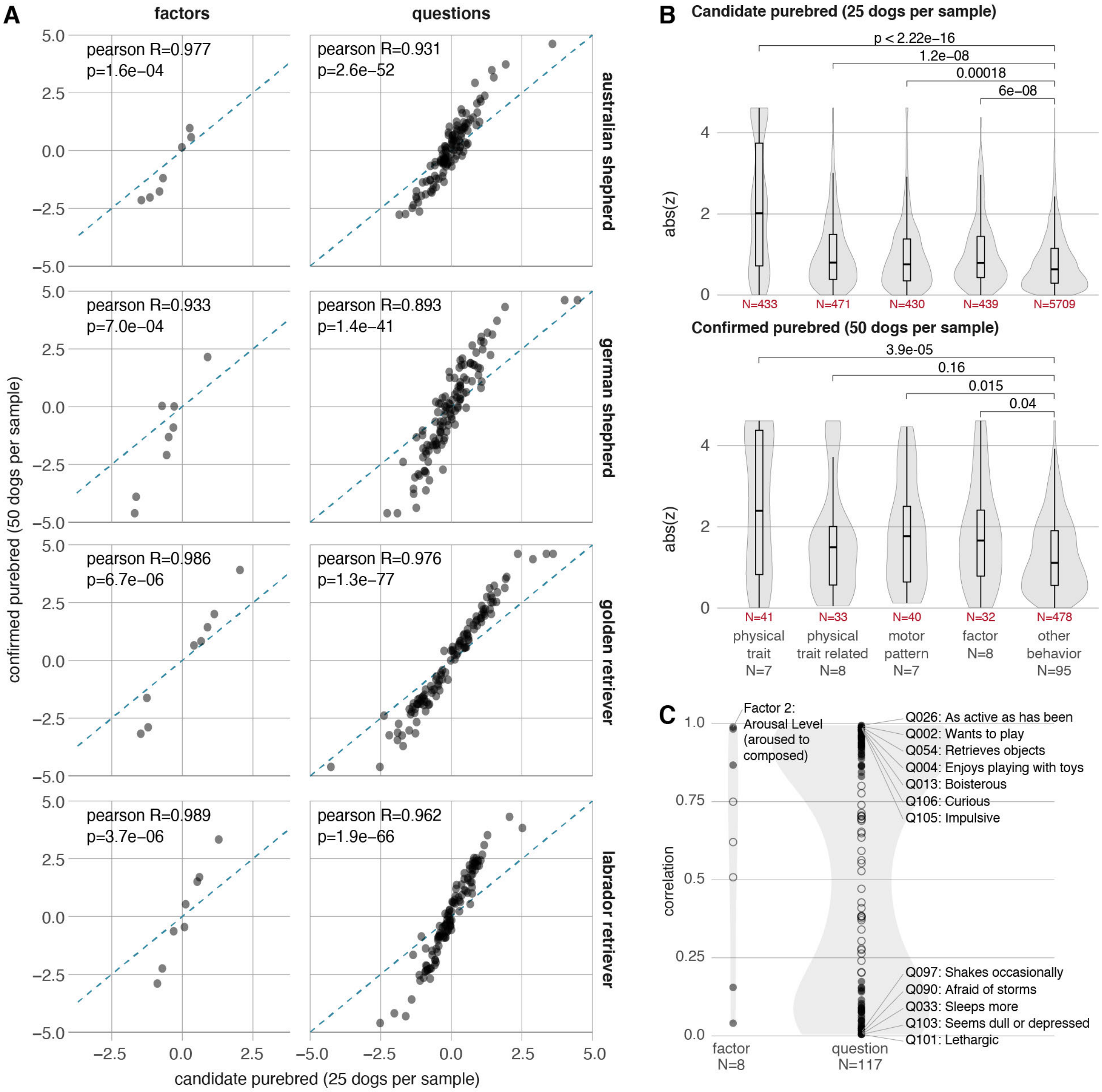


Fig. S13. Features of Population Peculiarity Scores (PPS). Permutation testing of breeds compared to random dogs yielded a 2-tailed z score reflecting how differentiated the breed is on each factor and question. (**A**) PPS z scores for candidate and confirmed purebred dogs are strongly correlated for both factors (left) and questions (right), with confirmed purebred almost always (91.4% of tests) yielding more extreme z scores (median z score change of 2.21-fold for factors and 2.15-fold for questions). Results shown for four breeds with >100 confirmed purebred dogs. (**B**) Comparison of absolute z scores for different question types shows breeds are most differentiated for physical traits, and that physical trait related questions, motor patterns, and factors score higher than other behavioral questions. Differences are more pronounced in the candidate purebred analysis (top), which includes up to 40 breeds per question/factor, than the confirmed purebred (bottom), with just 4-5 breeds. T-test with FDR corrected p-values (BH procedure). (**C**) Differentiation in dogs grouped by year of age shows scores for some questions/factors are highly correlated with age. Most correlated 10% of points are labeled.

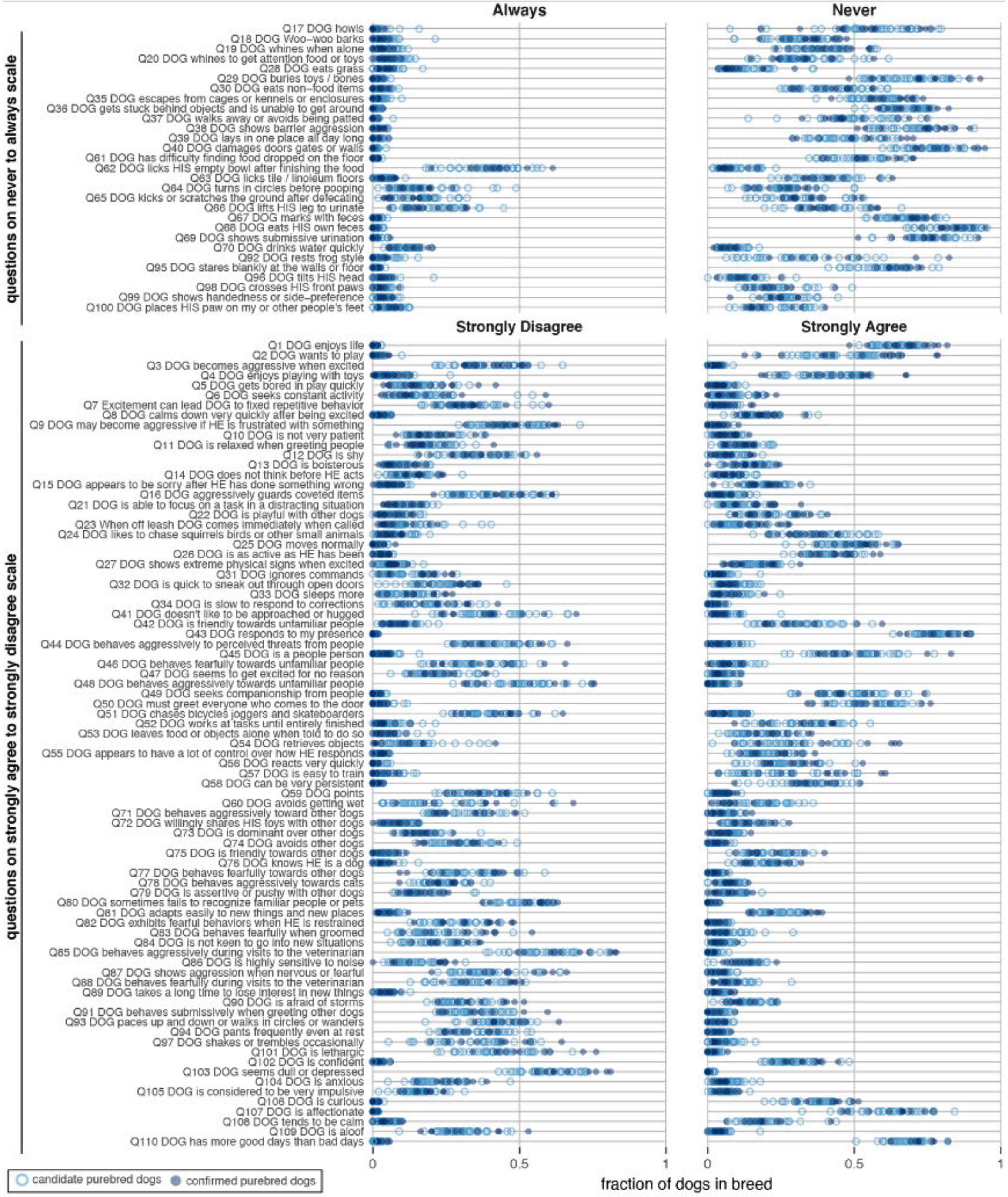


Fig. S14. No behaviors are exclusive to a subset of breeds. For breeds represented by ≥50 dogs (44 candidate breeds and 15 confirmed breeds), the fraction of owners choosing an answer at either end of the 5-level Likert scale (Never, Always, Strongly Agree, or Strongly Disagree) never reaches 100%.

Fig. S15. The 30 questions (representing top quartile) with responses most correlated (either negatively or positively) with year of age. PPS z score after 500,000 permutations is on y-axis, and Pearson correlation, with FDR corrected p−values (BH procedure), is printed in the lower left or upper right of each plot. Words in parentheses show direction of score. For example, "(agree to disagree)" means low scores indicate differentiation towards "agree" and high scores indicate differentiation towards "disagree".

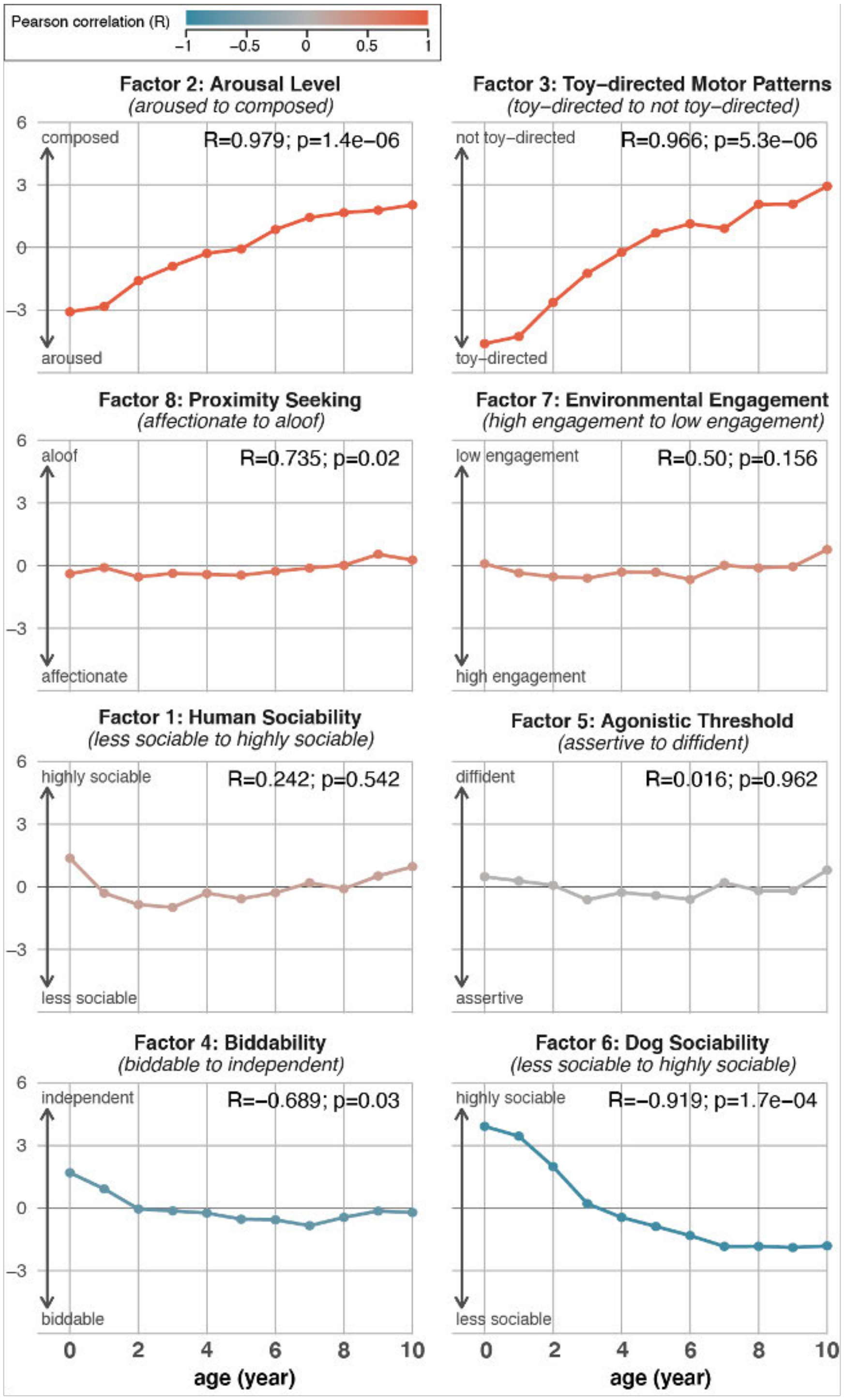


Fig. S16. PPS correlation with year of age for all eight behavioral factors. PPS z score after 500,000 permutations is on y-axis, and Pearson correlation, with FDR corrected *p*-values (BH procedure), is printed in the upper left of each plot. The direction of the differentiation is printed next to the y axis with a two-headed arrow. Six factors (all except factor 5 and factor 4) are correlated with age.

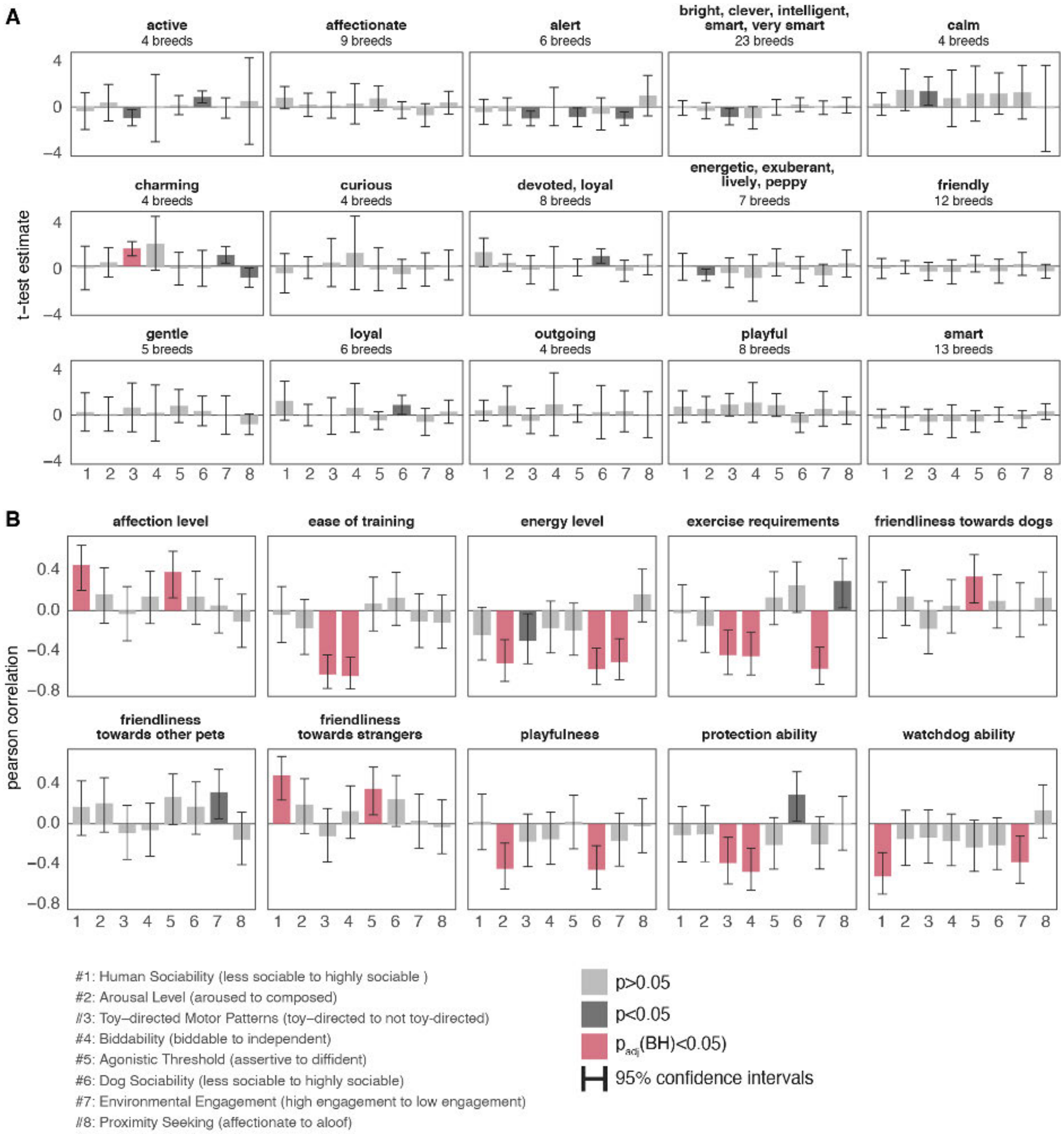
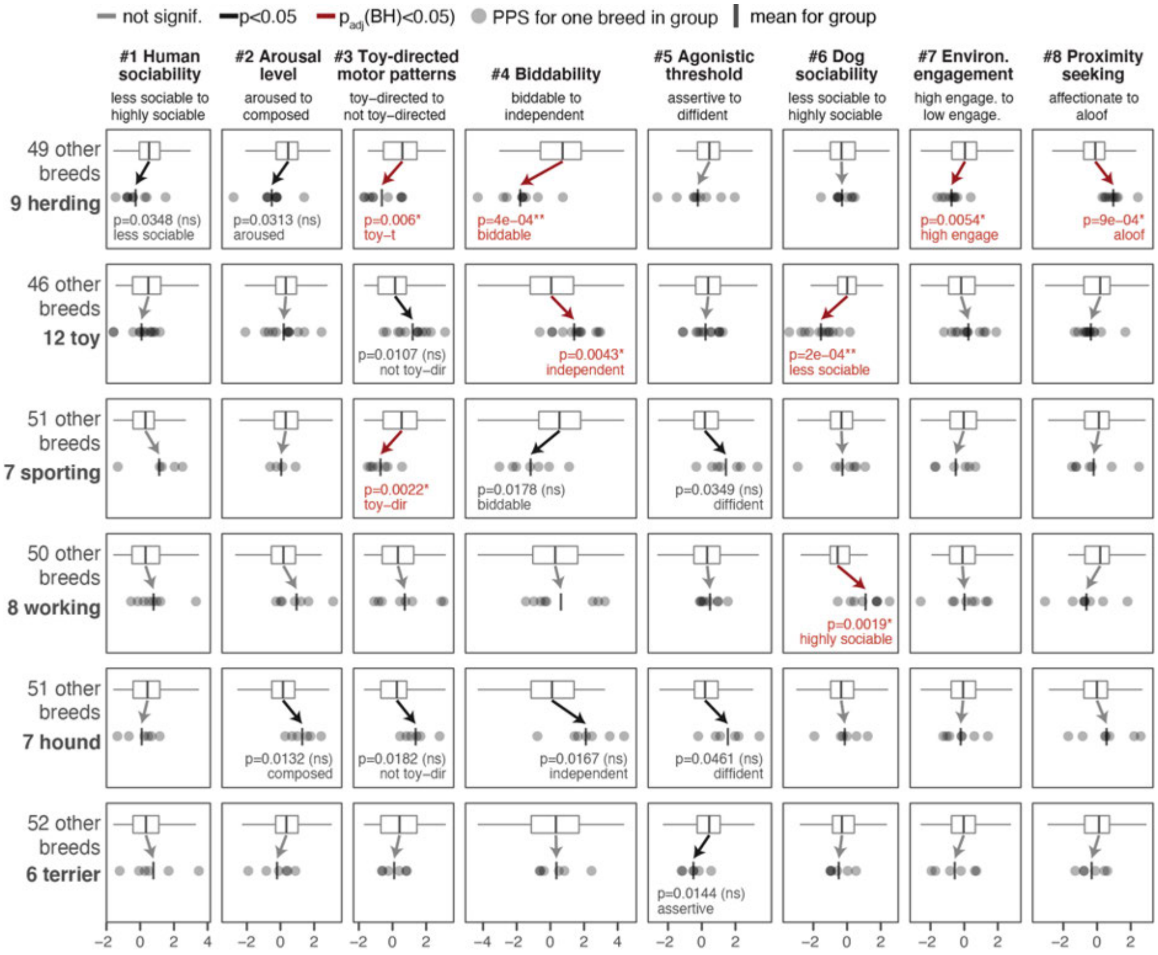


Fig. S17. Breed stereotypes versus breed peculiarity scores. (**A**) AKC's three descriptive words. Analysis of both single words and words grouped by shared meaning (all words and word groups describing ≥4 breeds included) using t-test estimates small difference in means for most words. Error bars show 95% confidence intervals around the t-test estimates. (**B**) Pearson correlation between behavior scores from "Encyclopedia of Dog Breeds" (*19*) and breed PPS.



herding: australian cattle dog, australian shepherd, belgian malinois, border collie, catahoula leopard dog, collie, german shepherd dog, pembroke welsh corgi, shetland sheepdog

hound: basset hound, beagle, dachshund, greyhound, rhodesian ridgeback, treeing walker coonhound, whippet

sporting: american cocker spaniel, brittany, english springer spaniel, german shorthaired pointer, golden retriever, labrador retriever, vizsla

terrier: american staffordshire terrier, jack russell terrier, miniature schnauzer, rat terrier, staffordshire bull terrier, west highland white terrier

toy: bichon frise, cavalier king charles spaniel, chihuahua, havanese, maltese, miniature pinscher, papillon, pomeranian, pug, shih tzu, toy poodle, yorkshire terrier

working: alaskan malamute, bernese mountain dog, boxer, doberman pinscher, great dane, great pyrenees, rottweiler, siberian husky

Fig. S18. Historic working role versus breed peculiarity scores. This plot visualizes the difference in mean breed PPS score on each behavioral factor between breeds assigned a particular historic working role (grey filled circles) and other breeds (box plot). The arrow highlights the difference in the mean PPS score between breeds with a particular historic working role and other breeds, and is red when the shift is significant. Vertical lines are the mean for group, box encloses 25-75% quartiles (Q25-Q75), and the horizontal line extends from 1.5 interquartile range (IQR) below Q25 to 1.5 IQR above Q75 + 1.5 IQR. Darker arrows are those reaching nominal significance ($p<0.05$). Red arrows are those significant after FDR correction (BH procedure). “(ns)” are not significant after FDR correction.

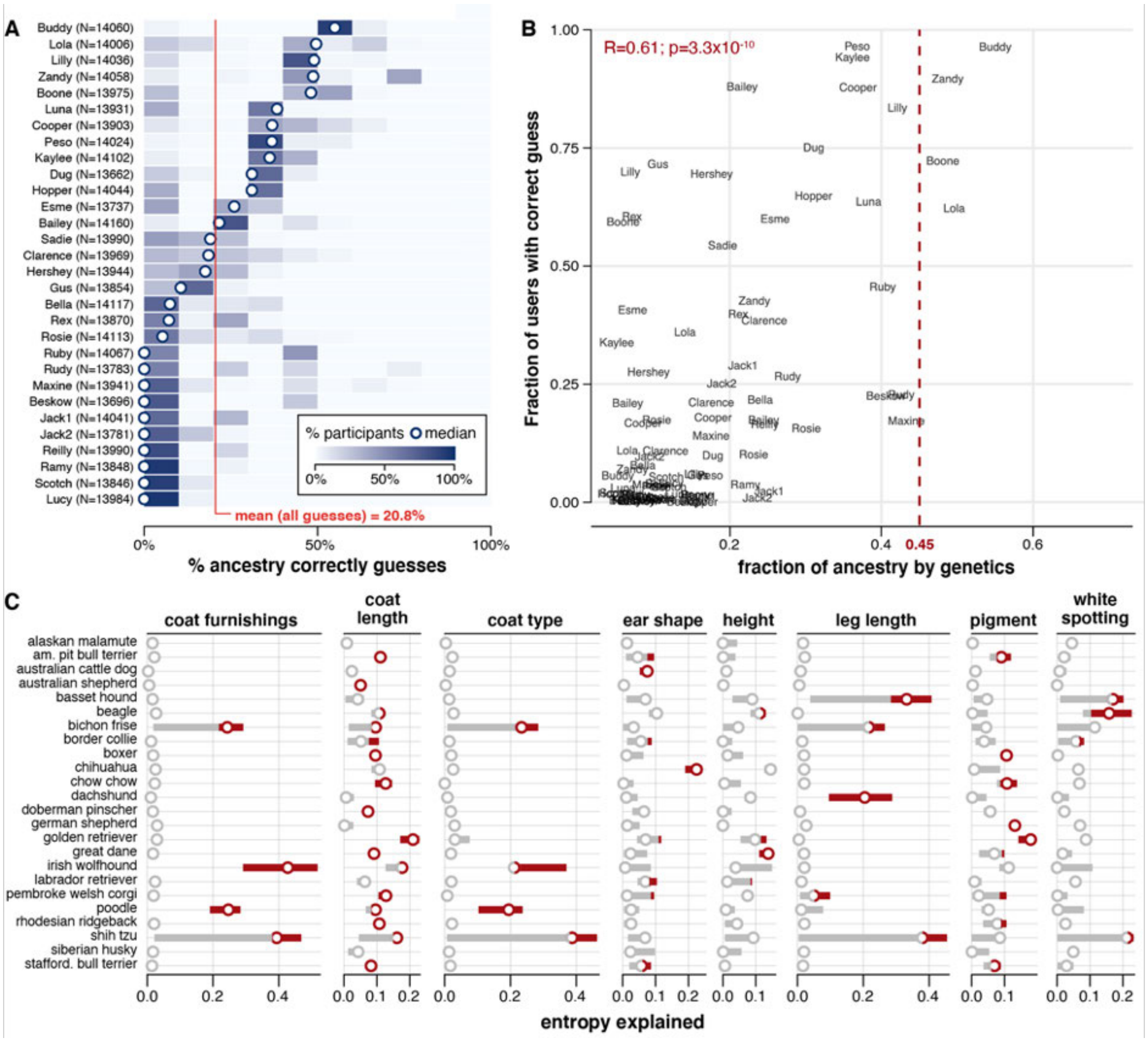


Fig. S19. Performance of Mendel's Mutts survey participants. (**A**) On average, participants correctly guessed breeds comprising a total of 20.8% of the breed ancestry in the mutts. (**B**) Breeds comprising a higher proportion of a mutt's ancestry are easier to guess, with all breeds > 45% guessed correctly by more than half of participants. Red text shows Pearson correlation. (**C**) Entropy decision stump analysis assessing what impact of presence or absence of an individual specific trait in a mutt has on probability that users would guess it to have ancestry from a given breed. Point indicates impact calculated for all 30 MuttMix mutts; bars span the range of values calculated when analysis was repeated iteratively, excluding each of the mutts in turn. Values in red indicate that a given trait significantly impacts the probability of users guessing ancestry from a given breed.

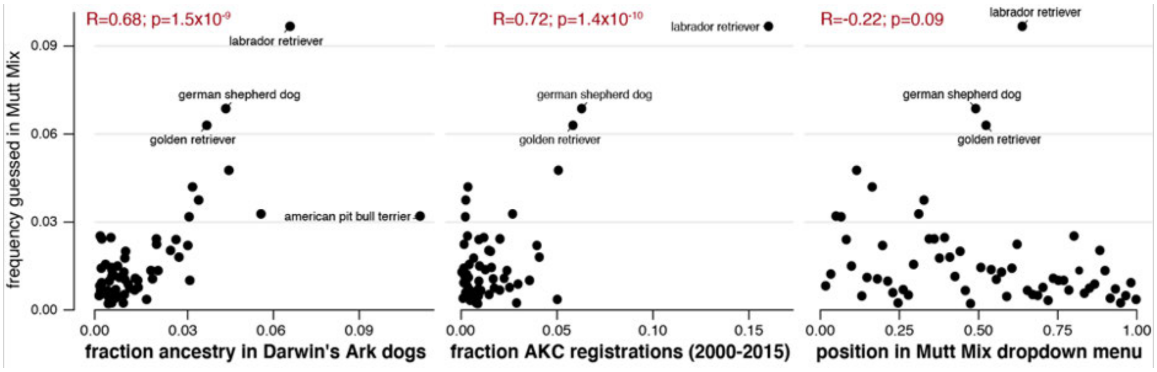


Fig. S20. MuttMix survey participants guessed popular breeds more often. The rate at which MuttMix users guessed a given breed ancestry is strongly correlated to the popularity of that breed, whether measured by fraction ancestry in Darwin's Ark (left) or by fraction AKC registrations (middle), and there is no significant correlation to the position of a given breed in the dropdown menu used by participants to guess breeds. Red text shows Pearson correlation.

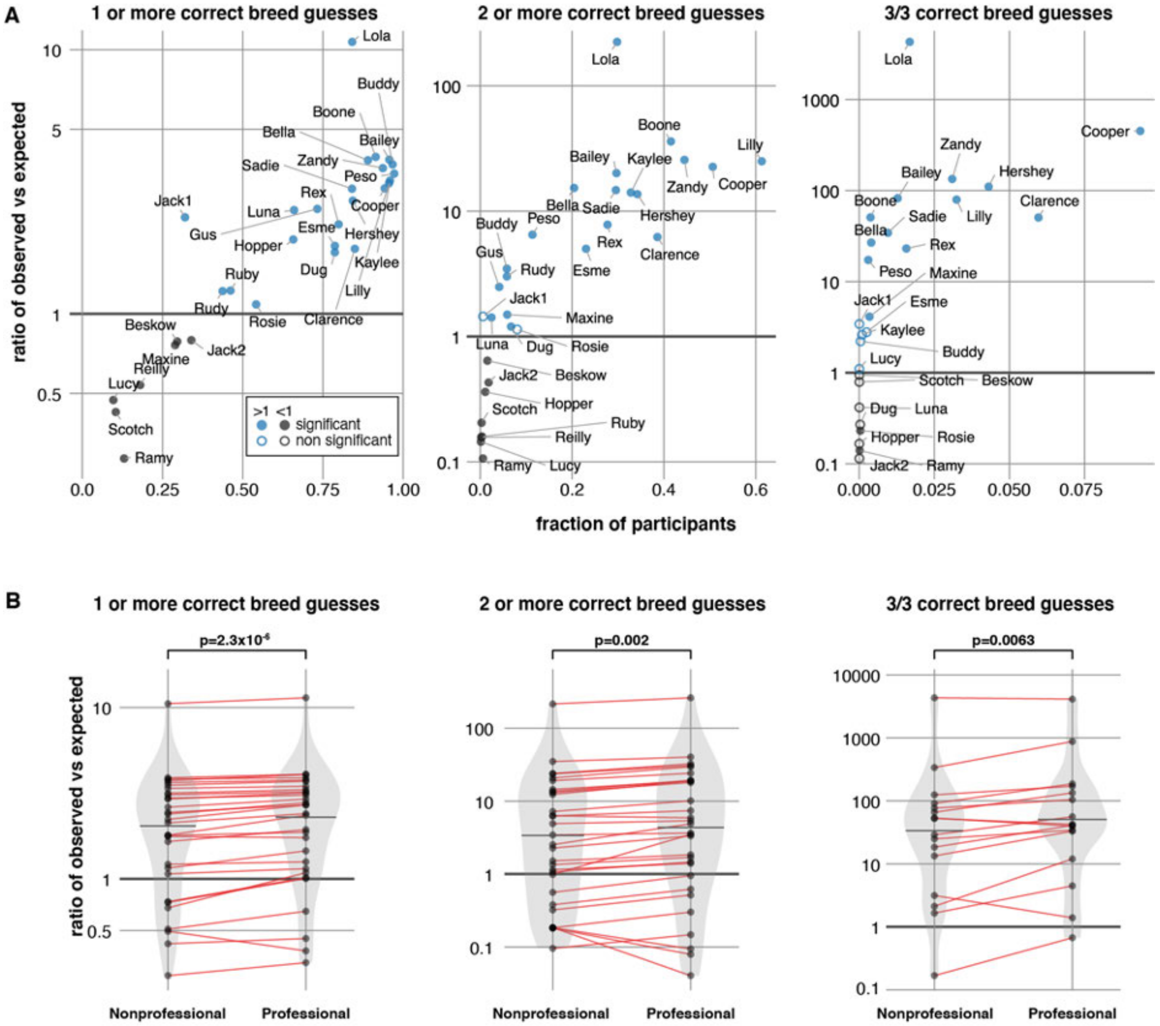


Fig. S21. When guessing breed ancestry in mutts, people are more accurate than random chance, but still often wrong. (**A**) Dogs with ancestry from more popular breeds tend to be guessed more accurately. Measuring the observed to expected ratio controls for this, and shows that people tend to be better than random chance (blue points; black is worse than random) at making 1+, 2+ and 3 correct guesses for most mutts (77%, 73% and 69% of mutts respectively), but their accuracy is still low, particularly when guessing more than one breed. Some dogs, like Cooper and Lilly, are more easily guessed than others, like Scotch. Significance was calculated using a chi-squared test to compare observed to expected. (**B**) Self-described dog professionals are slightly but significantly better than non-professionals, with higher observed/expected ratio for 1+, 2+ and 3 guesses for 93%, 87% and 75% of dogs respectively. Significance measured using paired two-sample Wilcoxon test. Horizontal black line is median, and black points are individual dogs. Red lines connect the points for the same dog. Statistical tests done using *R* package *stats* version 4.1.1 in R version 4.1.1.

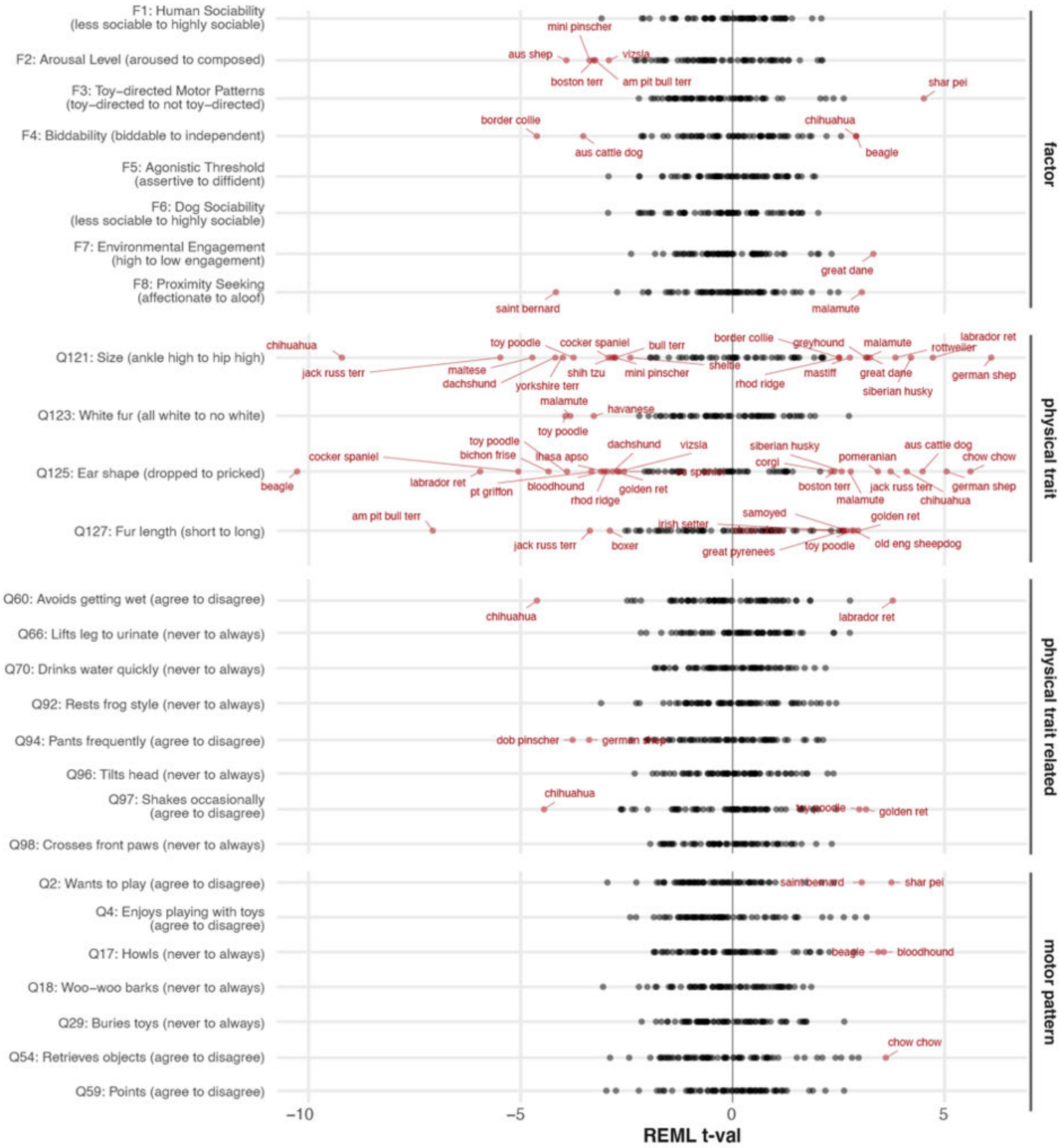


Fig. S22. Effect of breed ancestry in mutts. Results shown for all factors, physical traits, physical trait related behaviors, and motor patterns. Red dots (with labels) are statistically significant after FDR correction (BH procedure).

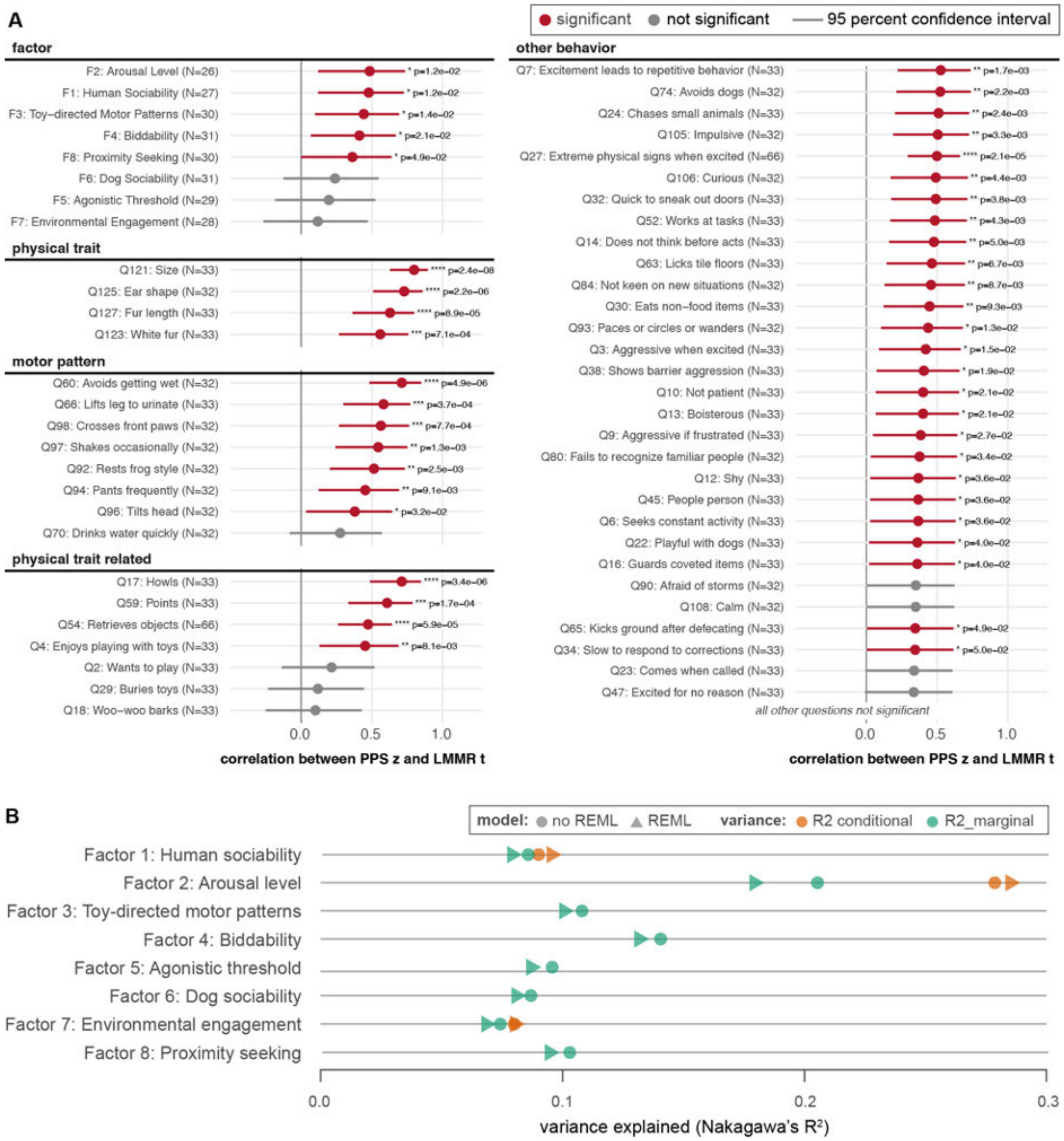


Fig. S23. Effects of breed ancestry in perspective. (**A**) Pearson correlation between effects of breed ancestry in mutts estimated by the LMER models and breed PPS in confirmed purebred dogs. Positive correlations indicate LMER is concordant with PPS. (**B**) Proportion of variance in factor score explained by the fixed effects of breed ancestry (the marginal R^2), as well as variance explained by the added random effects of age group (2 years and under, adult dogs, and dogs 12 years and older) and genetic kinship (the conditional R^2) in models not reaching singularity (tolerance = $1x10^{-5}$).

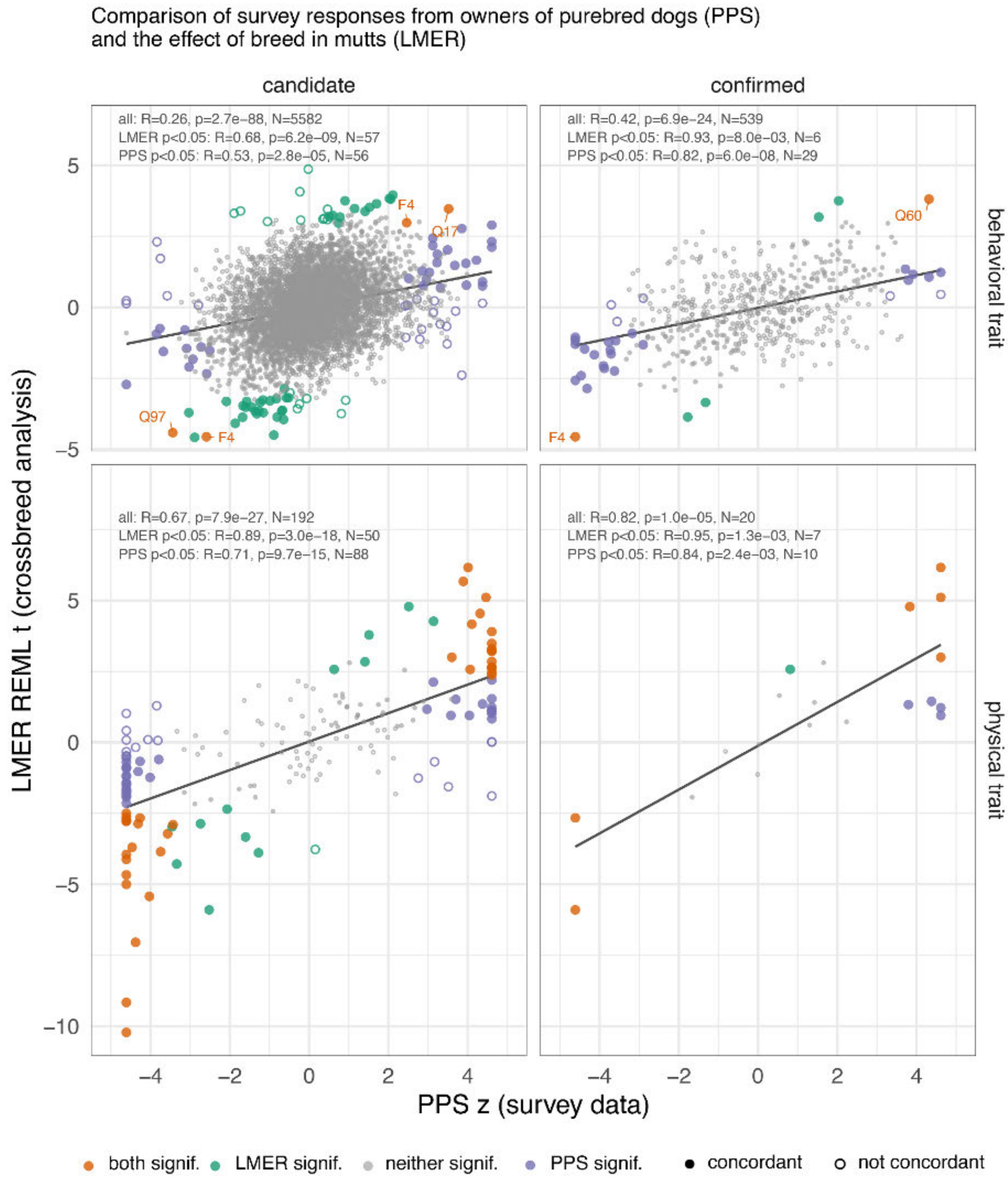
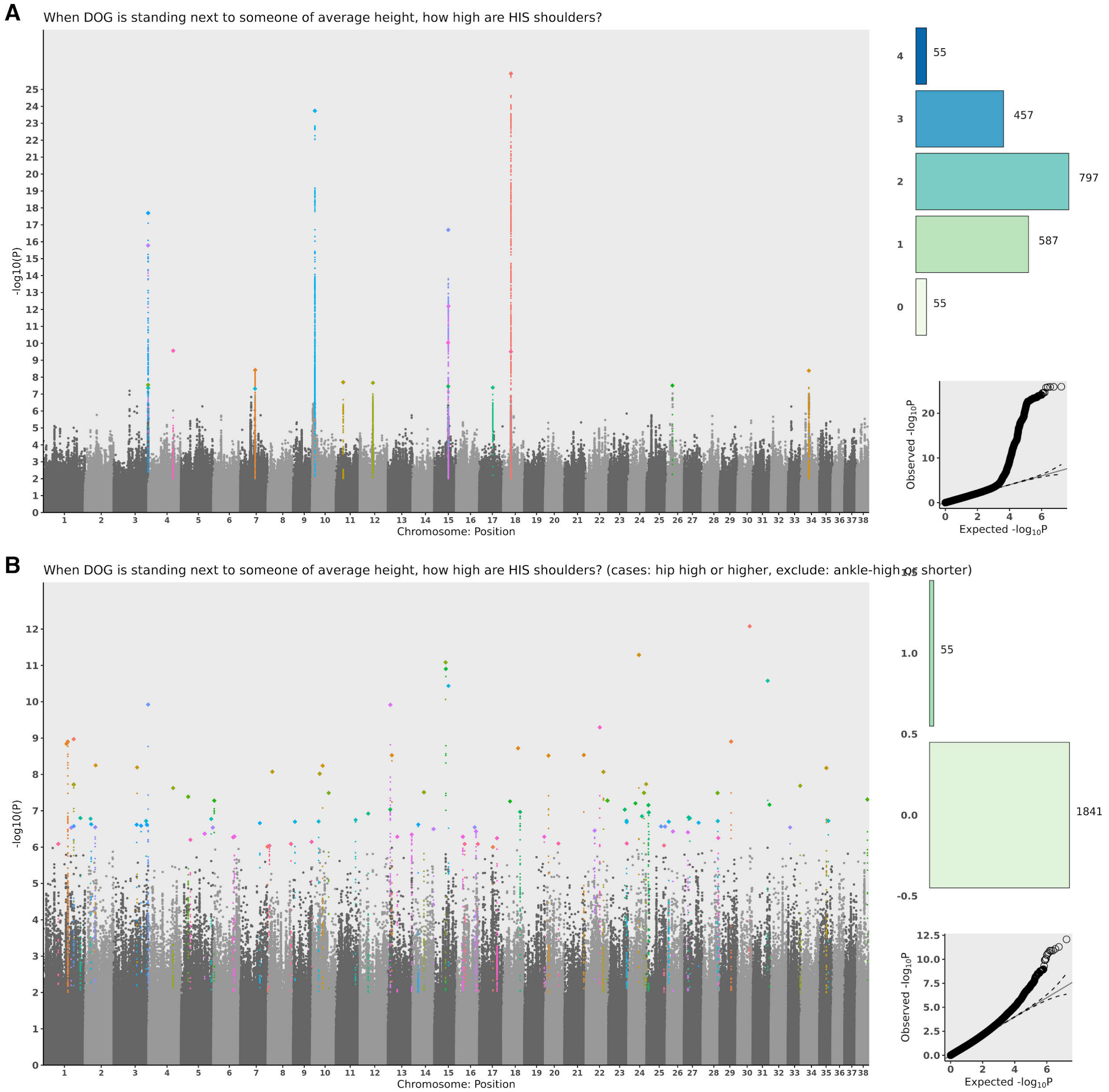
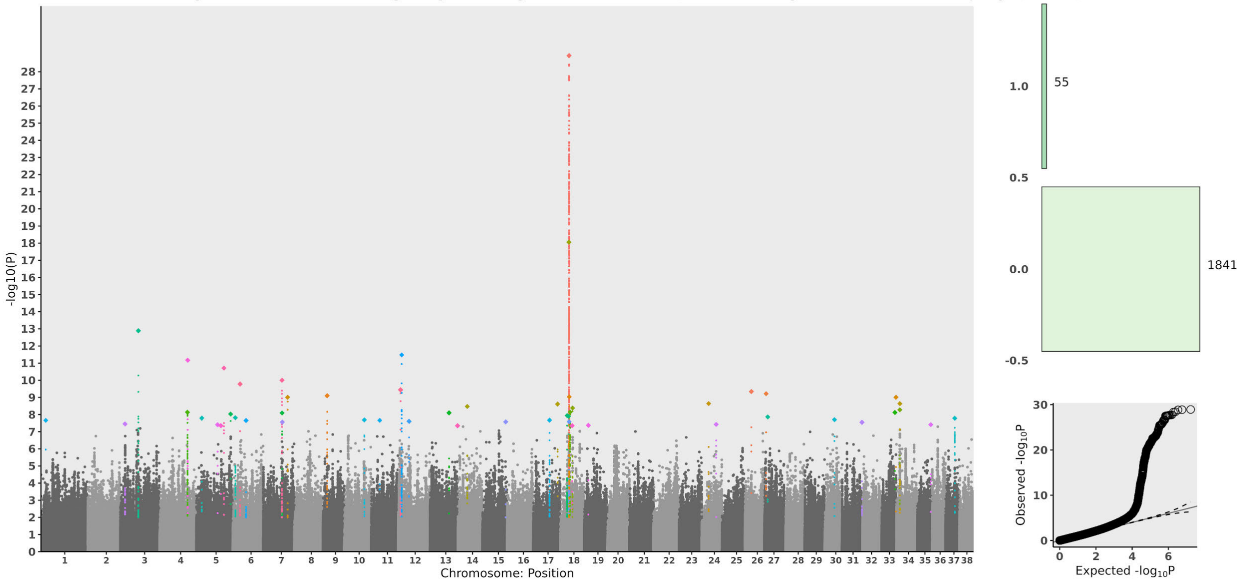


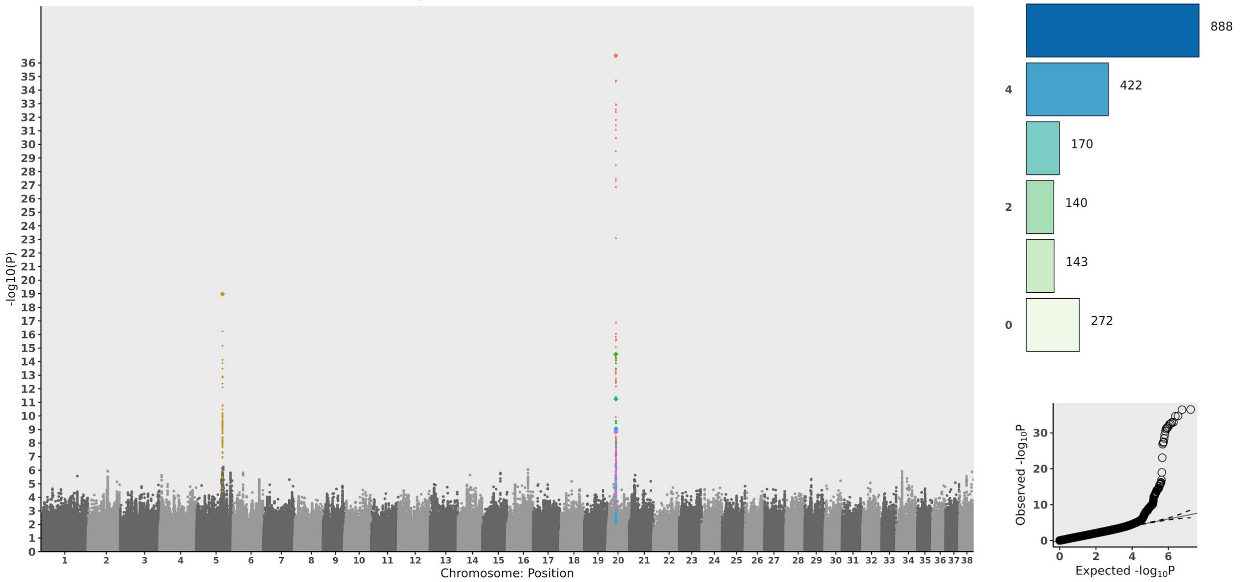
Fig. S24. Overview of the breed effects from the ascertained ancestry of mutts (LMER analysis) versus breed effects in owner surveys of reportedly purebred dogs (PPS analysis). Each point is a breed-trait pair. These points are solid if the directionality of the breed effect is the same in both mixed-breed and purebred dogs. Points are colored if the pair are both significant (orange), significant for breed ancestry effects (green), or significant for reported purebreds only (purple).



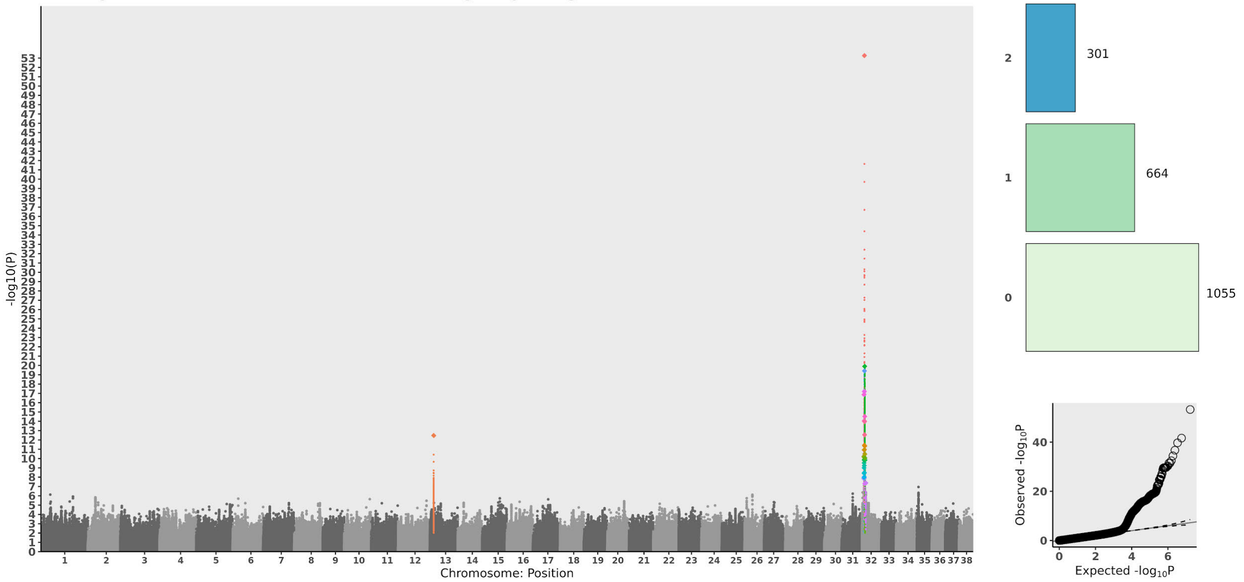
C When DOG is standing next to someone of average height, how high are HIS shoulders? (cases: ankle high or shorter, exclude: hip-high or taller)



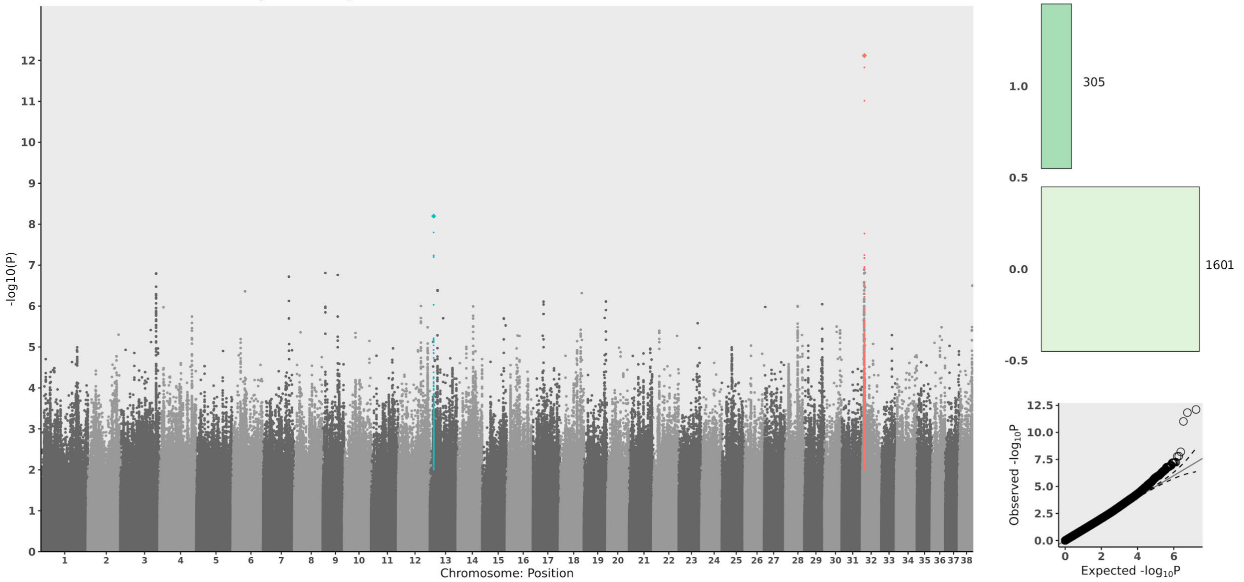
D How much white fur does DOG have? Select the image with the closest amount of white.

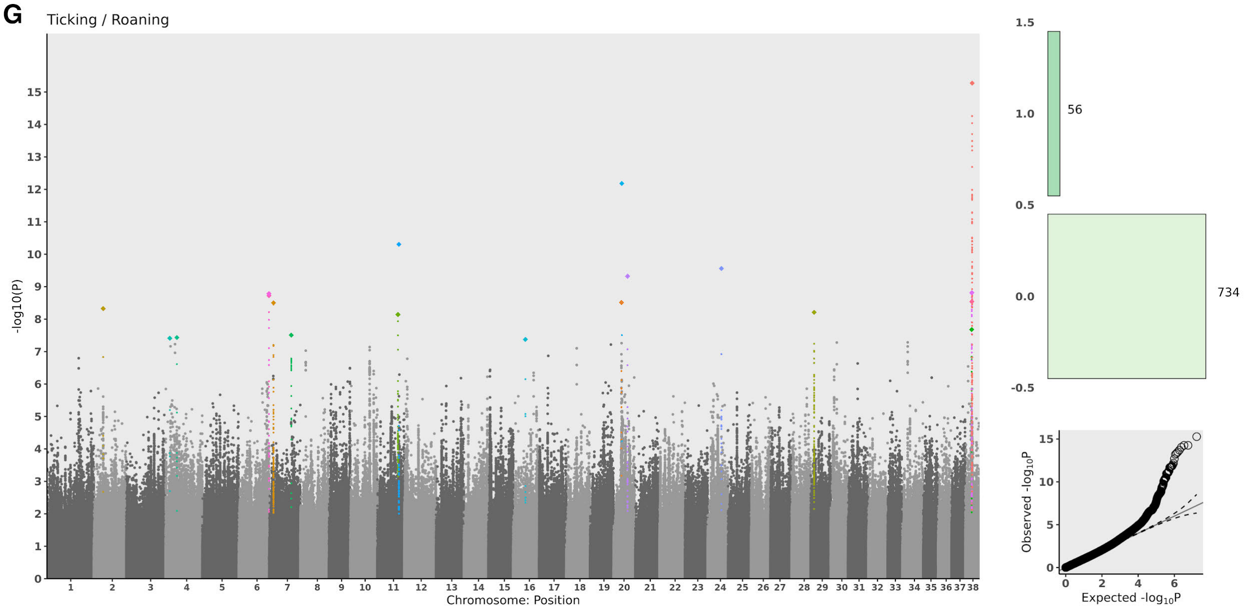


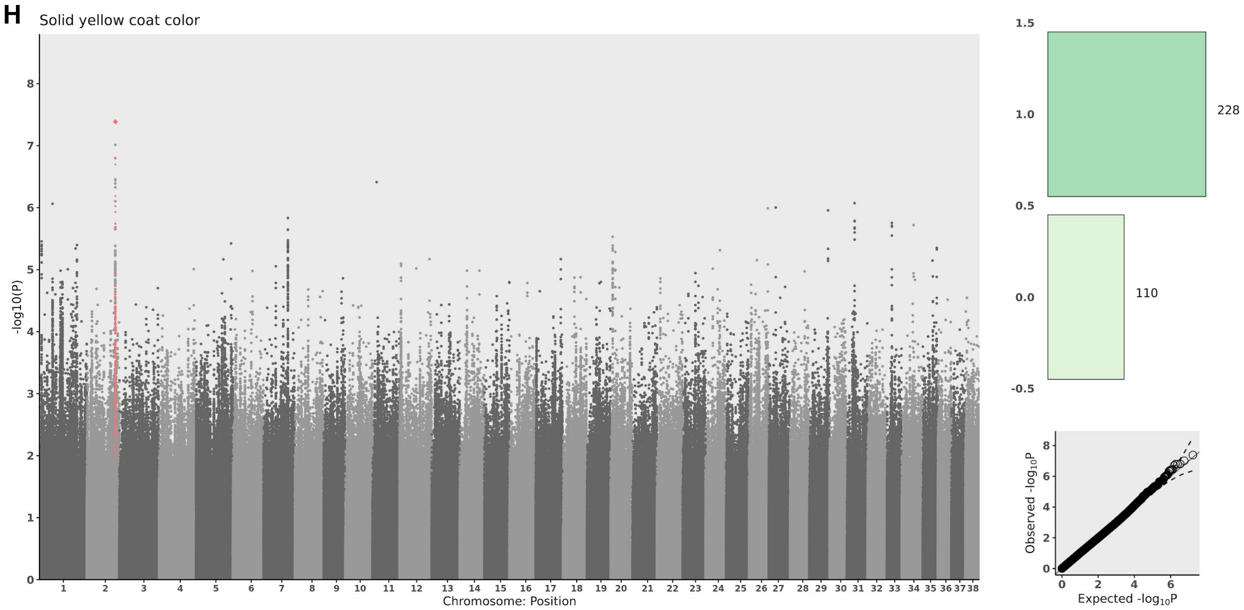
E How long is DOG's fur on HIS back and sides? Measure it against your finger.

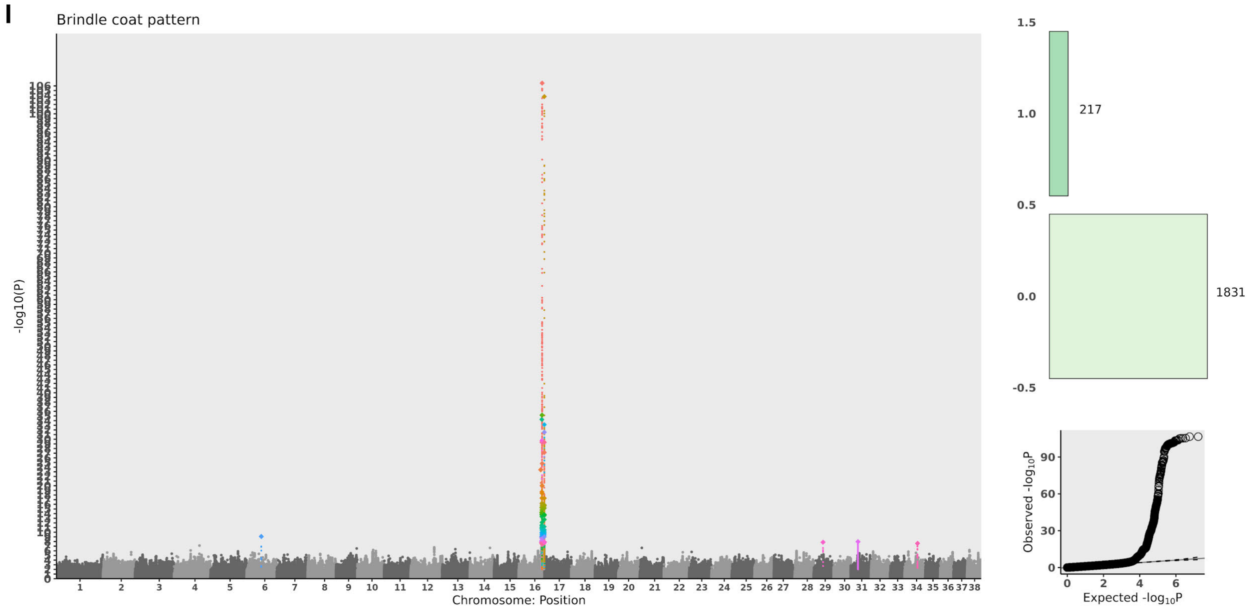


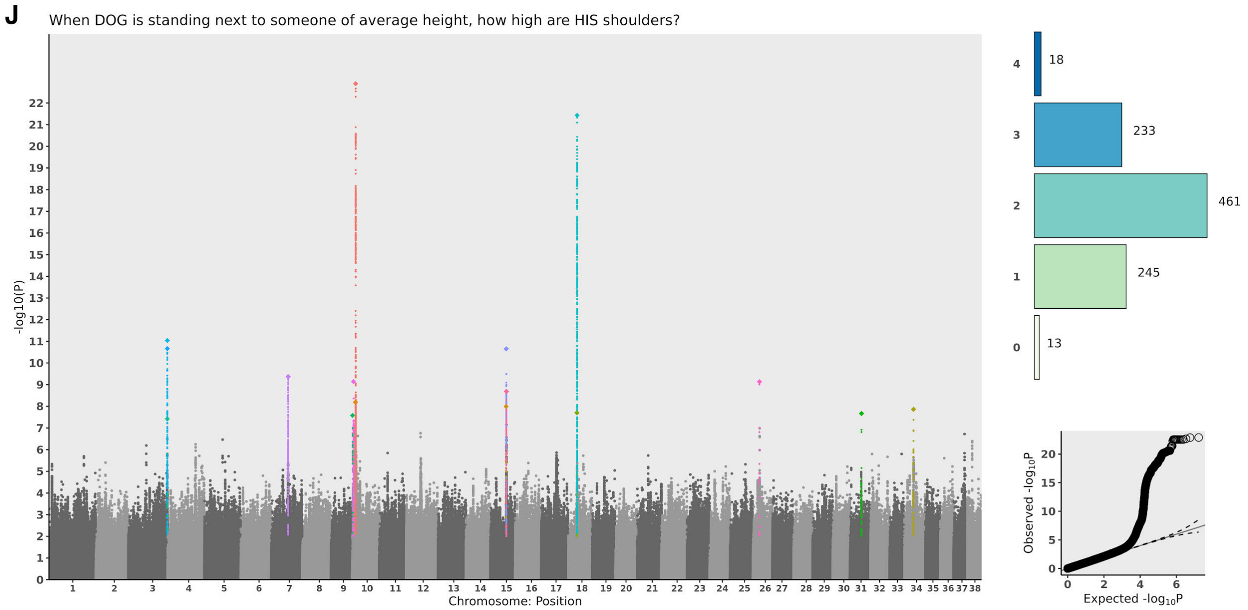
F Does DOG have soft fur or rough and bristly fur?



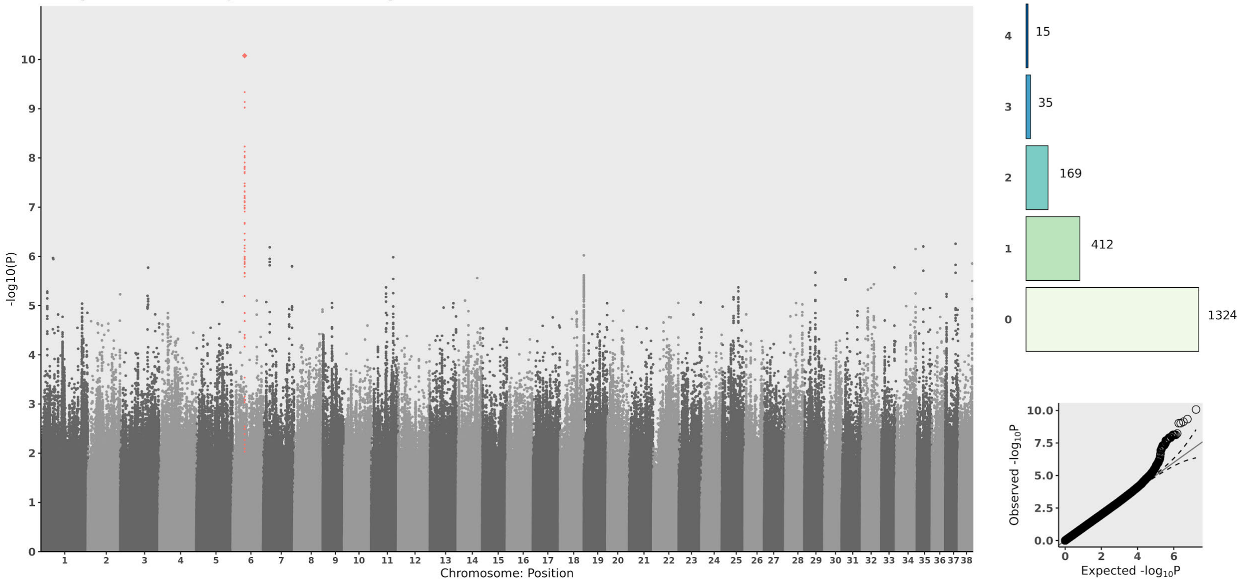

G
Ticking / Roaning
-log10(P)
0 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30 31 32 33 34 35 36 37 38
Chromosome: Position
1.5
1.0
0.5
0.0
-0.5
56
734
Observed -log$_{10}$P
Expected -log$_{10}$P
0 5 10 15
0 2 4 6


H
Solid yellow coat color
-log10(P)
0 1 2 3 4 5 6 7 8
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30 31 32 33 34 35 36 37 38
Chromosome: Position
1.5
1.0
0.5
0.0
-0.5
228
110
Observed -log$_{10}$P
Expected -log$_{10}$P
0 2 4 6 8
0 2 4 6

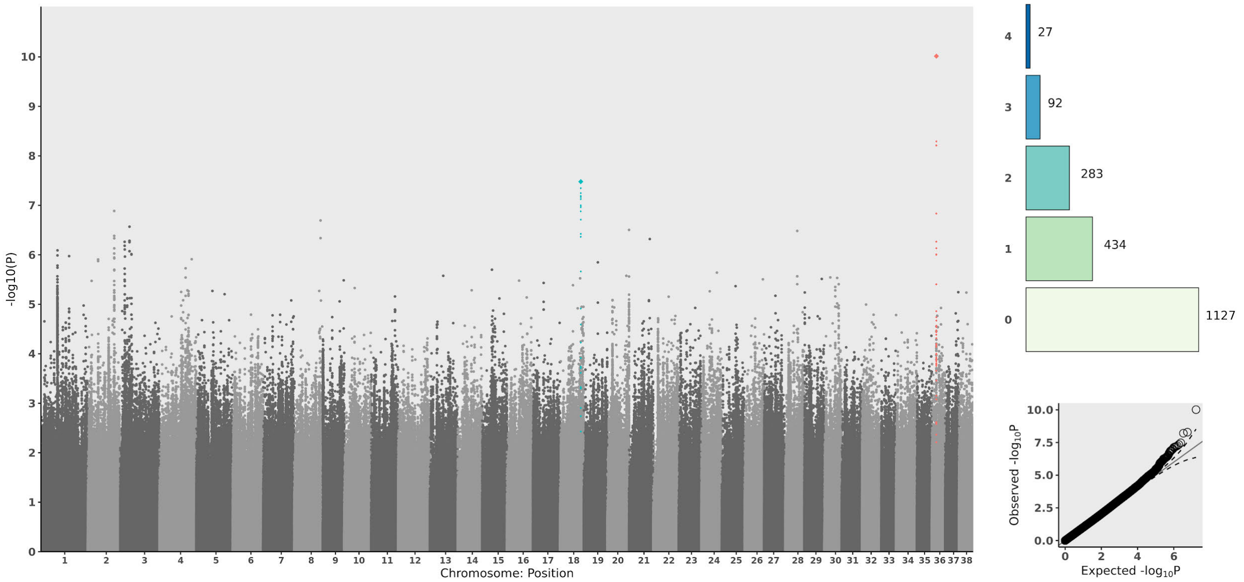

I
Brindle coat pattern
-log10(P)
106
0
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30 31 32 33 34 35 36 37 38
Chromosome: Position
1.5
1.0
0.5
0.0
-0.5
217
1831
Observed $-\log_{10}P$
90
60
30
0
Expected $-\log_{10}P$
0 2 4 6

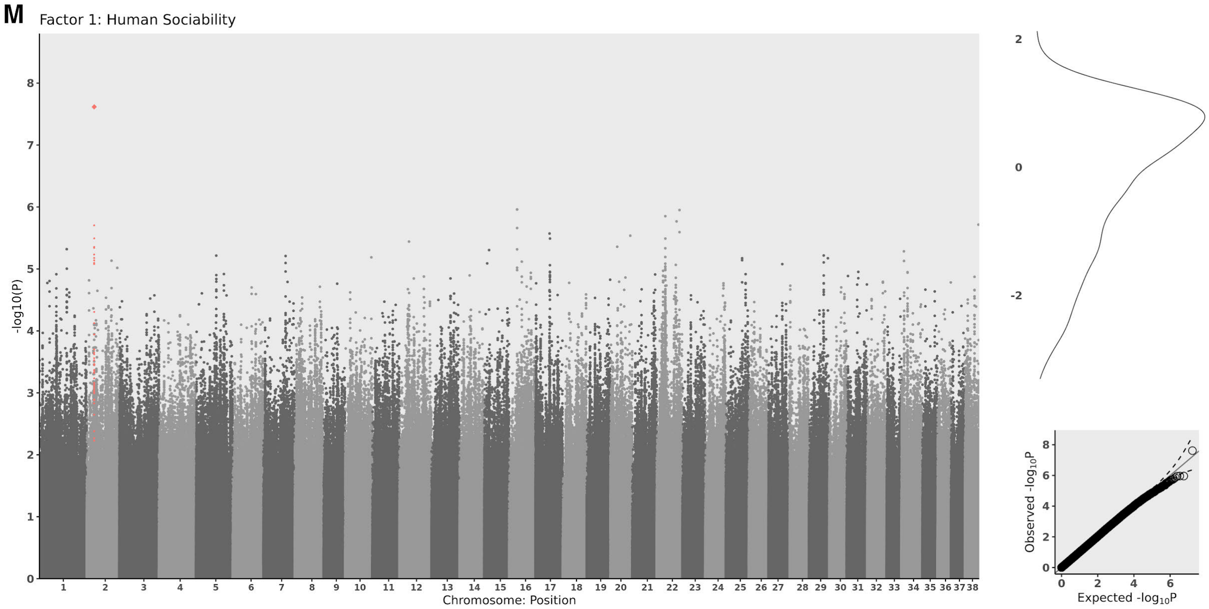

J
When DOG is standing next to someone of average height, how high are HIS shoulders?
-log10(P)
22 21 20 19 18 17 16 15 14 13 12 11 10 9 8 7 6 5 4 3 2 1 0
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30 31 32 33 34 35 36 37 38
Chromosome: Position
4
18
3
233
2
461
1
245
0
13
Observed $-\log_{10}P$
20
15
10
5
0
Expected $-\log_{10}P$
0 2 4 6

K DOG gets stuck behind objects and is unable to get around



L DOG howls





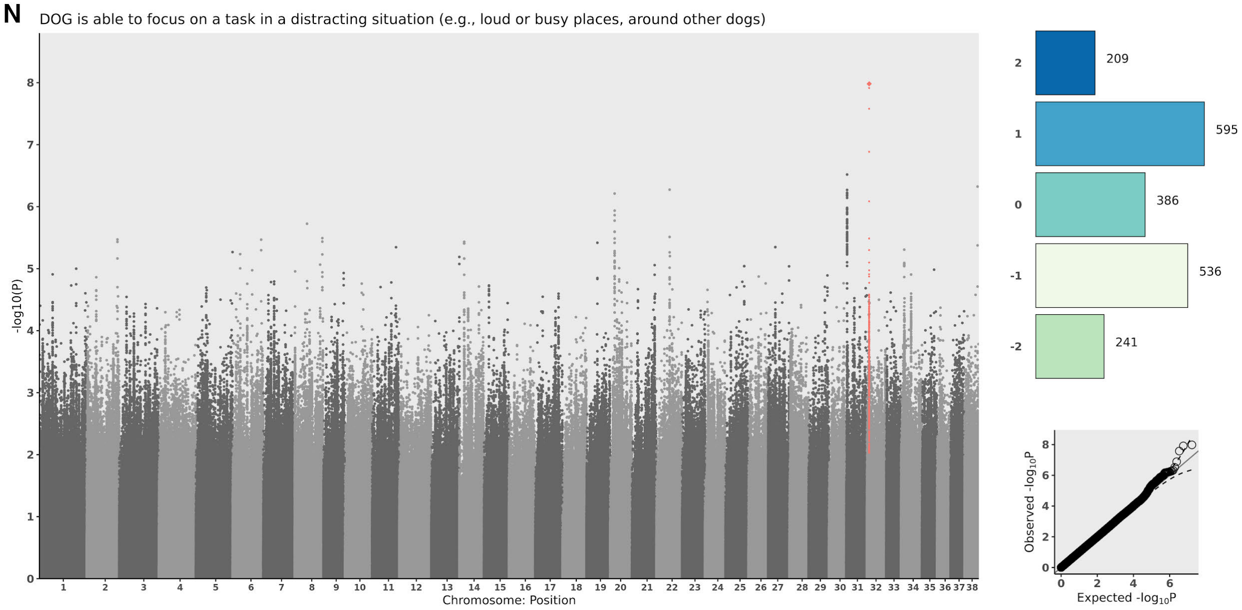


Fig. S25. Genome-wide association studies. All Manhattan plots with Q-Q plots and phenotype distributions for the genome-wide association studies for all traits have been published to Figshare (DOI: 10.6084/m9.figshare.16608793). Plots for the following phenotype studies shown are (**A**) body size as a quantitative trait in all dogs; (**B**) gigantism and (**C**) dwarfism as binary traits with average stature dogs as controls; several coat pigmentation phenotypes including (**D**) white spotting, (**E**) coat length, (**F**) coat texture, (**G**) roan and ticking coat patterns, (**H**) solid red coat color versus yellow coat color, (**I**) brindle coat pattern; (**J**) body size in highly admixed dogs with <45% breed ancestry; (**K**) Likert responses to whether dog "gets stuck behind objects and is unable to get around"; (**L**) whether dogs howl never, rarely, sometimes, often, or always; (**M**) human sociability factor scores; and (**N**) Likert responses to whether dog "is able to focus on a task in a distracting situation (e.g., loud or busy places, around other dogs)".

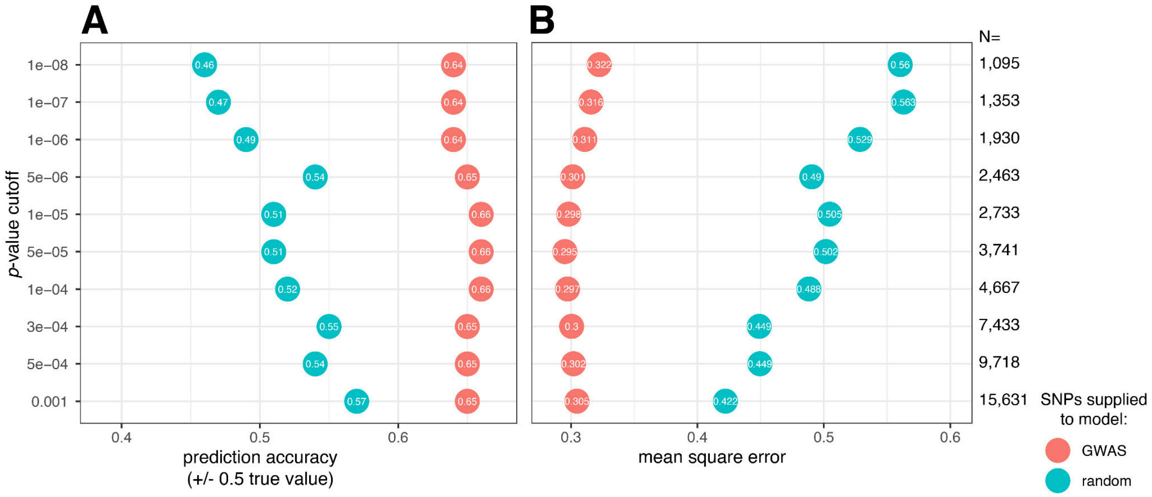


Fig. S26. Size prediction model performance. The optimal *p*-value cutoff is *p* from 1x10^{-5} to 1x10^{-4}, with (**A**) accuracy (defined as predicted value within +/- 0.5 of true value) falling at both higher and lower thresholds and (**B**) mean square error lowest at 5x10^{-5}. Random forest models drawing from GWAS-selected SNPs consistently outperform models drawing from a random pool of SNPs, which has an average prediction accuracy of 0.52 ± SD:0.05. Among models based on GWAS-selected SNPs, the most informative SNPs (defined as the top 100 ranked by Gini importance) carried absolute effect sizes of 0.239 ± SD:0.093 in the height GWAS, a ~1.68 fold-change compared to 0.142 ± SD:0.066, the effect sizes for all GWAS-selected SNPs used to build the model. The unselected models, too, derive predictive power from randomly drawn SNPs: the most informative SNPs carried absolute effect sizes of 0.044 ± SD:0.034 in the GWAS, a ~1.69 fold-change compared to 0.026 ± SD:0.024, effect sizes for all SNPs used to build the model.

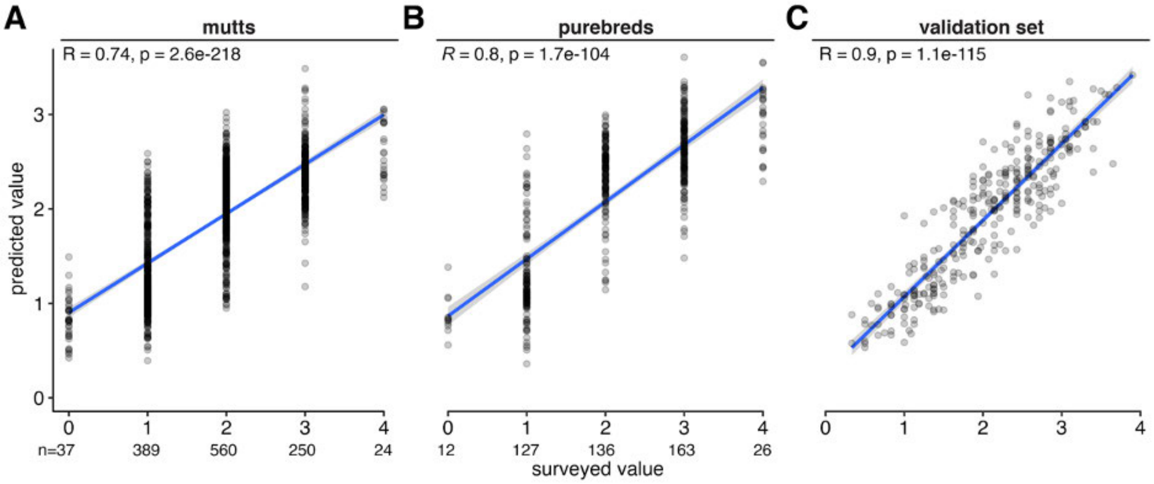


Fig. S27. Genetic predictions of size phenotypes perform well in mutts and purebred dogs and validate against real size measurements. Size predictions made for (**A**) mutts are just as accurate against their surveyed size as in (**B**) purebred dogs.

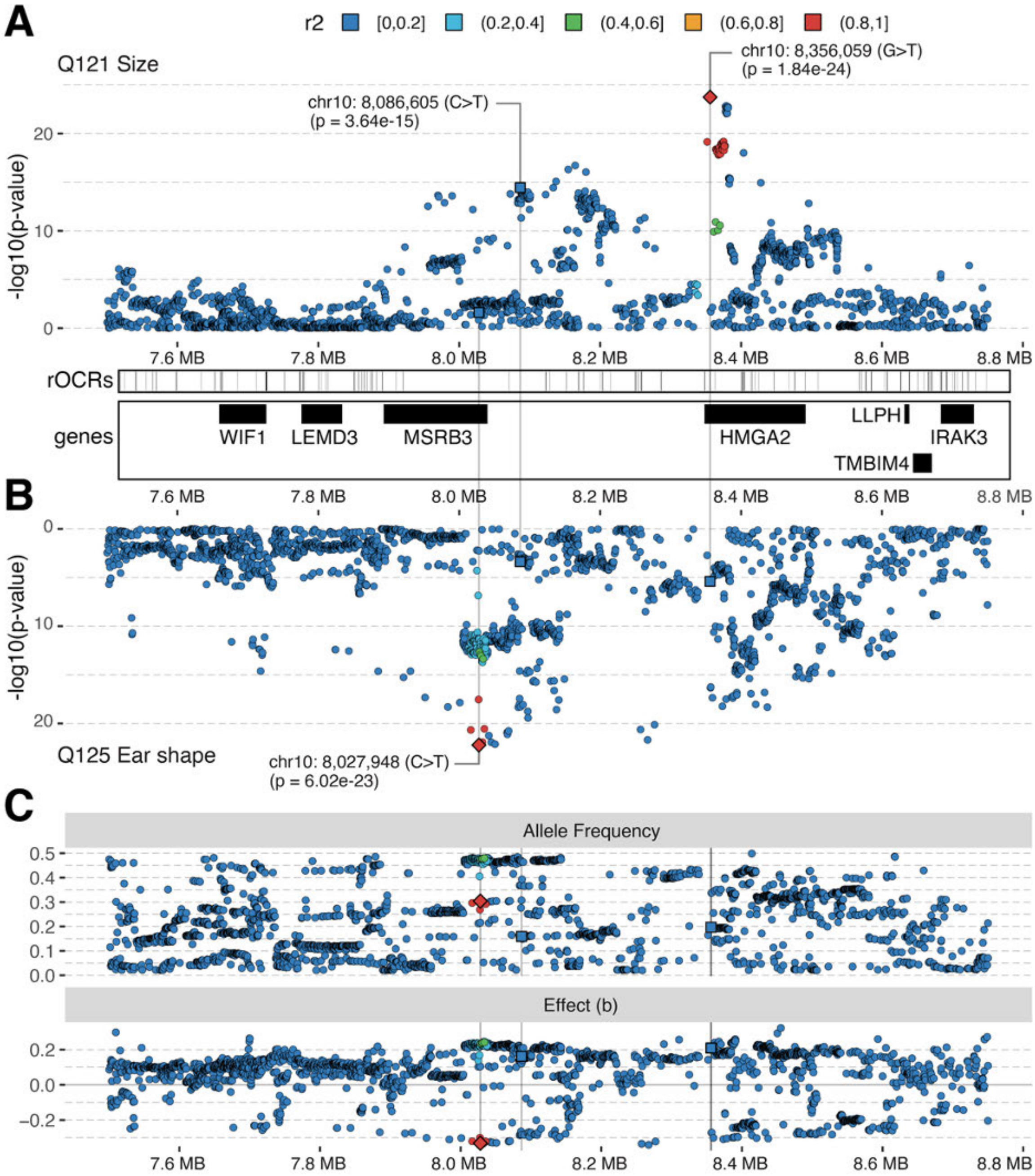


Fig. S28. Resolving the ear shape and body size locus at chromosome 10. Regional plots on chromosome 10 from 7.6 to 8.8 megabases (CanFam3.1) for association with (**A**) body size, colored by r^2 linkage with top SNP at 8,356,059 (diamond), and (**B**) ear shape, colored by r^2 linkage with top SNP at 8,027,948 (diamond). (**C**) The allele frequency and effect (beta) of the top ear shape-associated SNP.

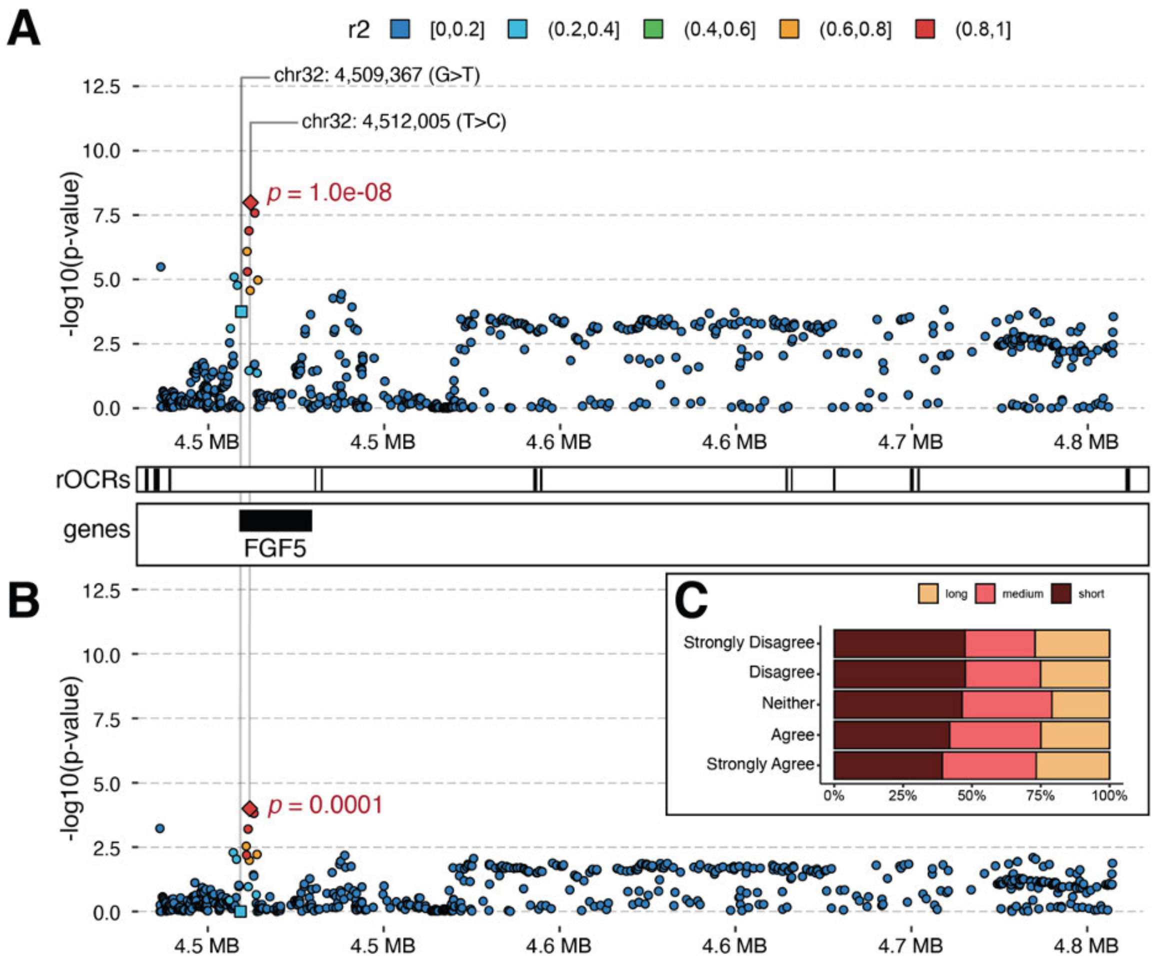


Fig. S29. Confounding effects of breed-defining physical traits on genome-wide associations. Regional association plots for (**A**) the locus on chromosome 32 associated with a dog's focus in distracting situations (Q21) with SNPs colored by r^2 linkage with the top associated SNP (diamond) and (**B**) the locus after inclusion of genotypes for the SNP associated with coat length differences (square) as a covariate. (**C**) Survey responses for Q21 against reported coat length.

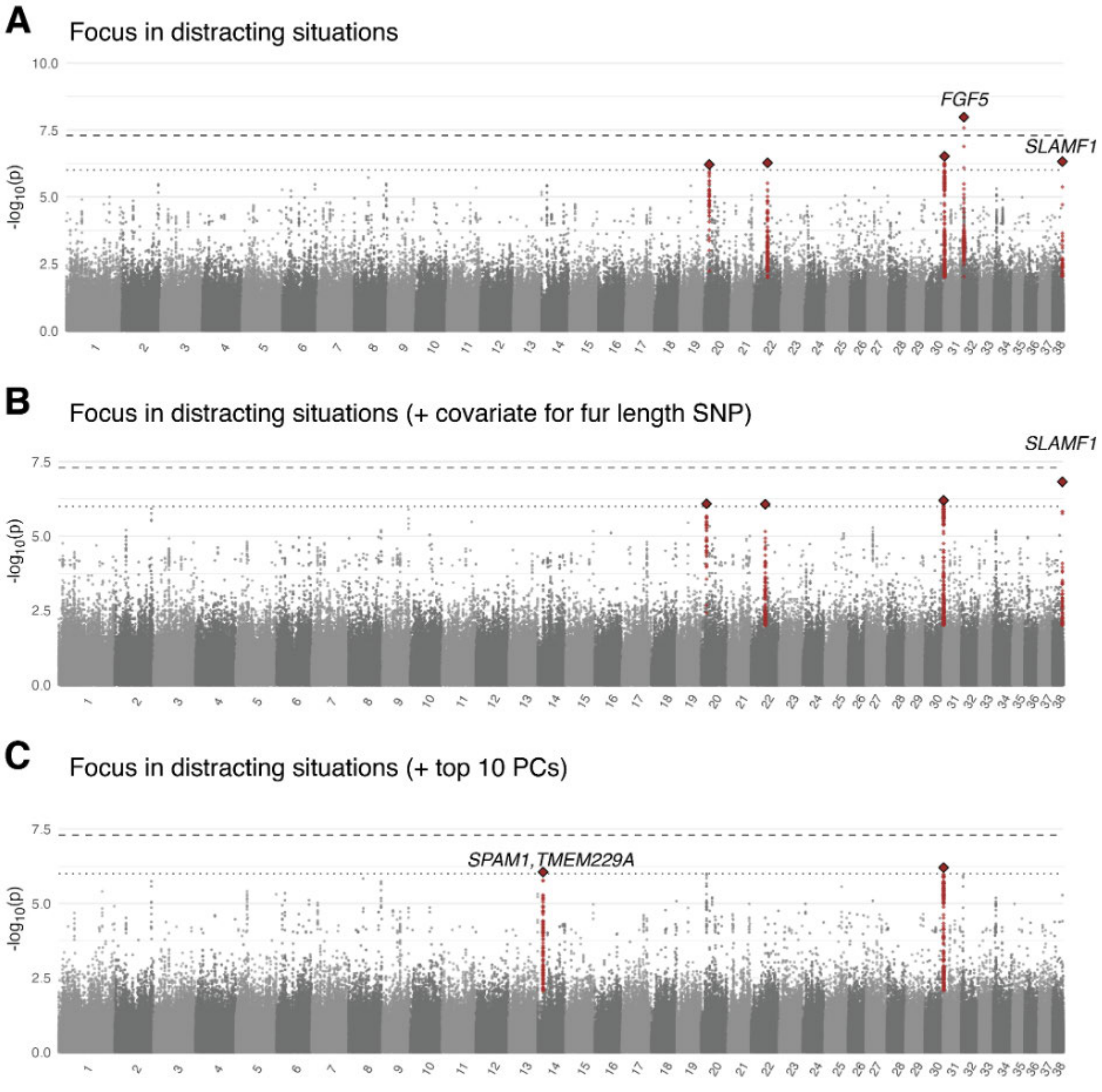


Fig. S30. Breed-confounding physical traits can be addressed by controlling for the traits themselves or by adjusting for population structure. (**A**) Before adding PCs to the mixed linear model, we identified one significant ($p < 5\text{x}10^{-8}$) peak at *FGF5* on chr32, a gene also associated with coat length, and 4 other suggestive ($p < 1\text{x}10^{-6}$) associations. (**B**) By including the fur length SNP in *FGF5* (chr32: 4,509,367), this association disappears (p=0.0001) but suggestive associations remain. (**C**) After inclusion of the first 10 PCs, the association with *FGF5* (p=$1.2\text{x}10^{-6}$) drops in strength, as do other suggestive associations.

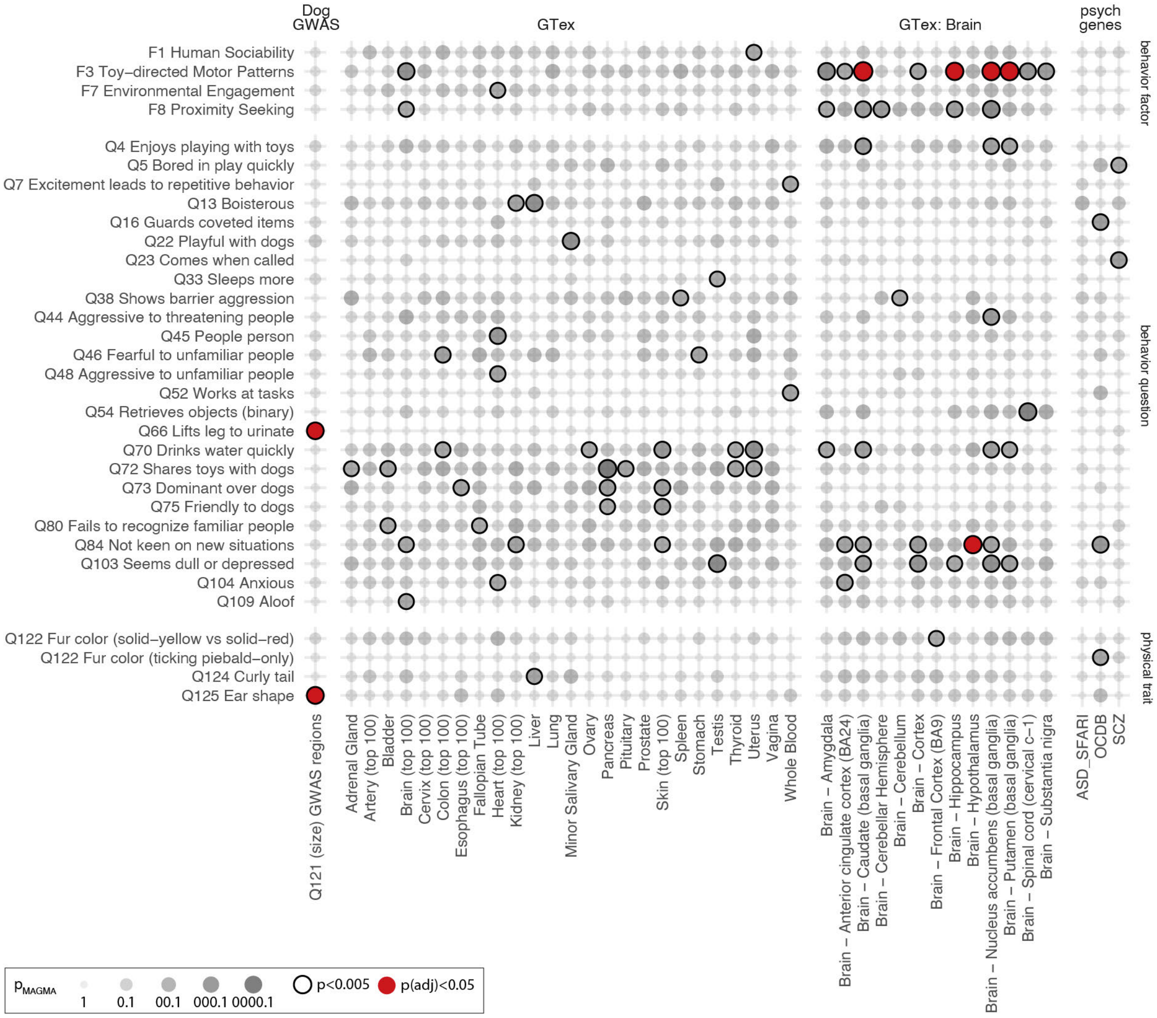


Fig. S31. Gene set enrichment of associated regions. The raw and FDR-adjusted (BH procedure) *p*-values for enrichment of trait associated SNPs upon (1) the canine body size GWAS regions, (2) top non-brain tissue-expressed genes from GTex, (3) top brain-expressed genes from GTex, and (4) psychiatric condition gene sets.

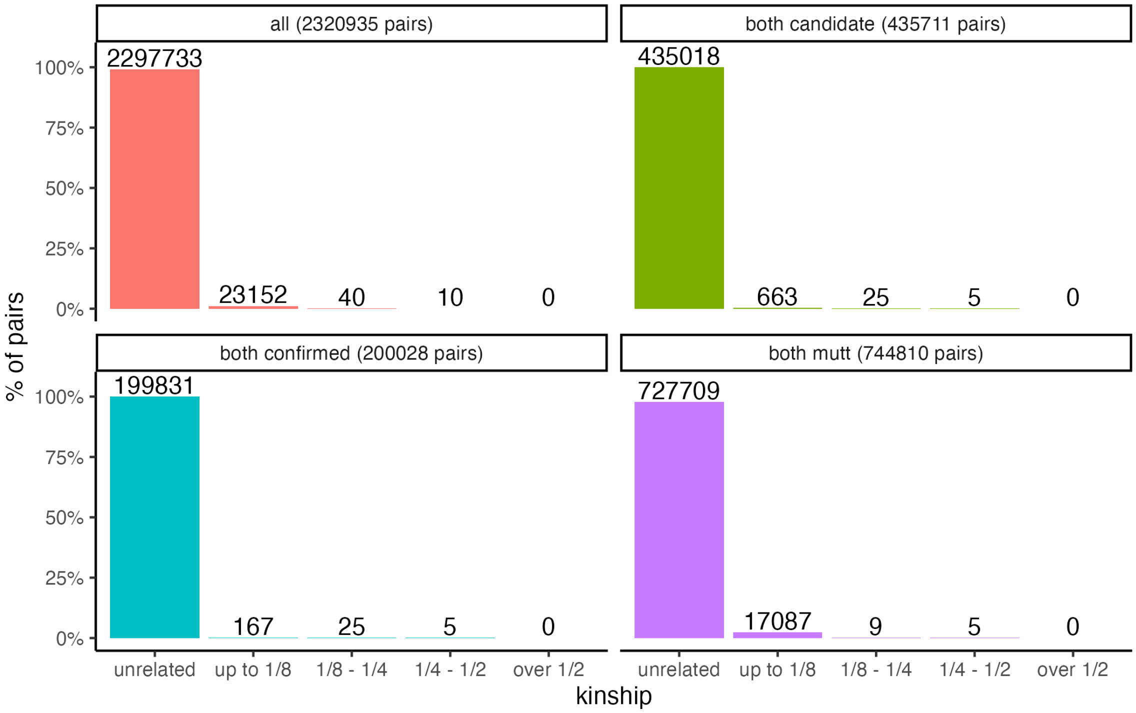


Fig. S32. Kinship between pairs of dogs in the genetic data set. We estimated kinship *k* using the KING-robust kinship estimator (*118*) and found only 28 pairs of dogs (0.0017%; 46 unique dogs) closely related (*k*>0.2).

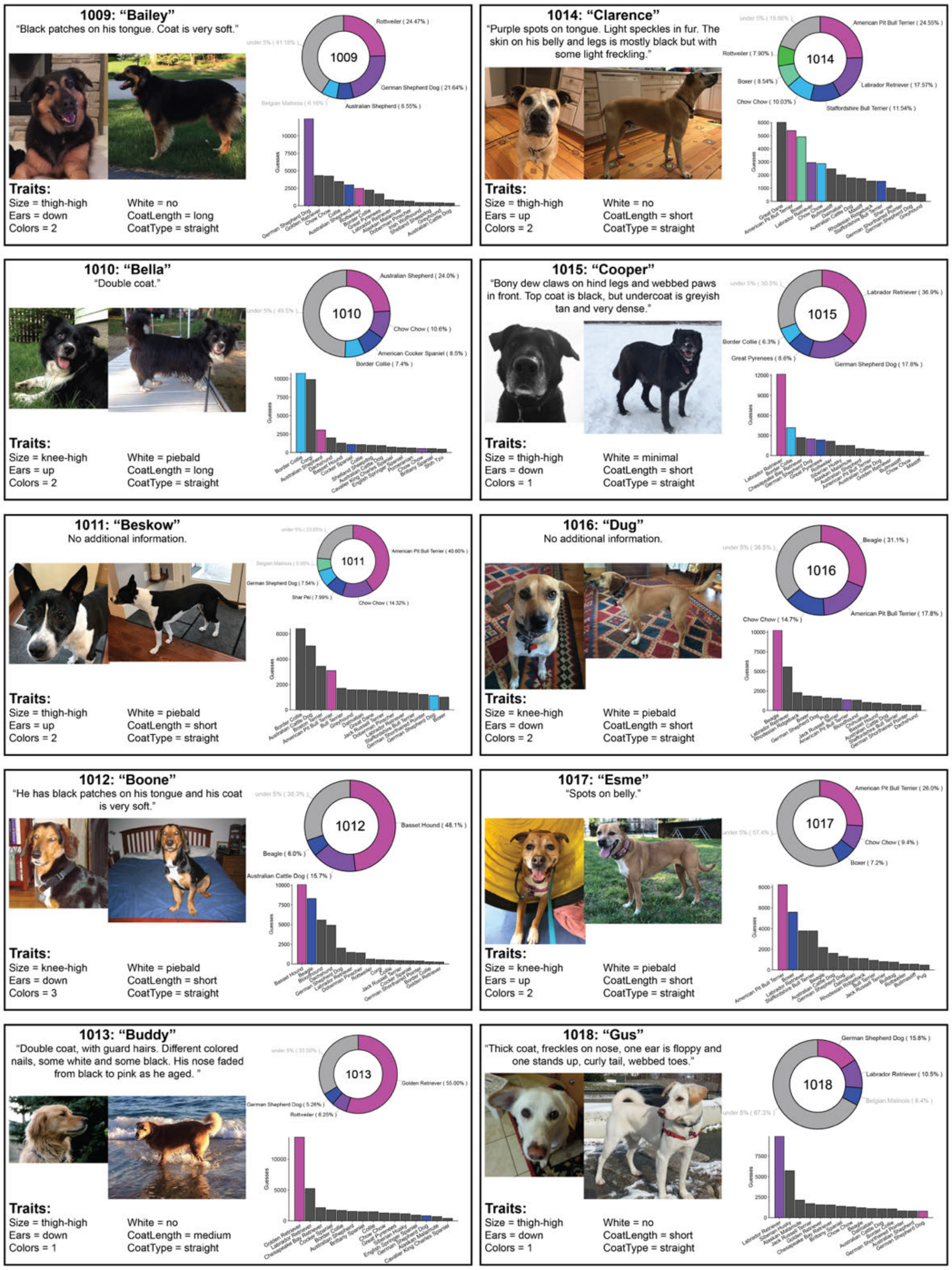
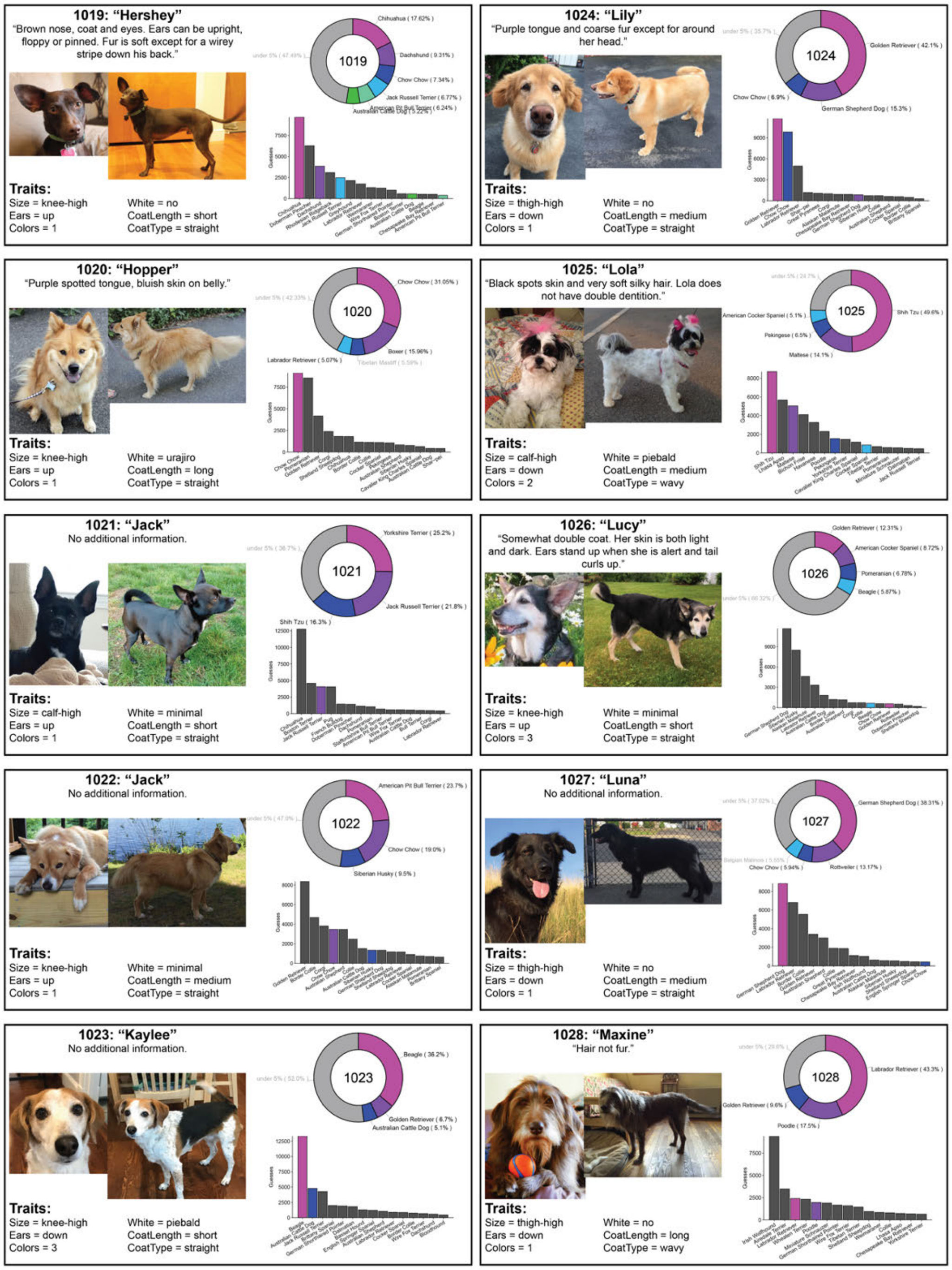


Fig. S33. Dogs included in the MuttMix survey (page 1/4). Front and side photographs and descriptions of the 30 mutts (black border) and single undeclared purebred (red). All photos and information were taken and provided by the dog owners for use in the public survey and research project.



1019: "Hershey"
"Brown nose, coat and eyes. Ears can be upright, floppy or pinned. Fur is soft except for a wirey stripe down his back."

Traits:
Size = knee-high
Ears = up
Colors = 1
White = no
CoatLength = short
CoatType = straight

1020: "Hopper"
"Purple spotted tongue, bluish skin on belly."

Traits:
Size = knee-high
Ears = up
Colors = 1
White = urajiro
CoatLength = long
CoatType = straight

1021: "Jack"
No additional information.

Traits:
Size = calf-high
Ears = up
Colors = 1
White = minimal
CoatLength = short
CoatType = straight

1022: "Jack"
No additional information.

Traits:
Size = knee-high
Ears = up
Colors = 1
White = minimal
CoatLength = medium
CoatType = straight

1023: "Kaylee"
No additional information.

Traits:
Size = knee-high
Ears = down
Colors = 3
White = piebald
CoatLength = short
CoatType = straight

1024: "Lily"
"Purple tongue and coarse fur except for around her head."

Traits:
Size = thigh-high
Ears = down
Colors = 1
White = no
CoatLength = medium
CoatType = straight

1025: "Lola"
"Black spots skin and very soft silky hair. Lola does not have double dentition."

Traits:
Size = calf-high
Ears = down
Colors = 2
White = piebald
CoatLength = medium
CoatType = wavy

1026: "Lucy"
"Somewhat double coat. Her skin is both light and dark. Ears stand up when she is alert and tail curls up."

Traits:
Size = knee-high
Ears = up
Colors = 3
White = minimal
CoatLength = short
CoatType = straight

1027: "Luna"
No additional information.

Traits:
Size = thigh-high
Ears = down
Colors = 1
White = no
CoatLength = medium
CoatType = straight

1028: "Maxine"
"Hair not fur."

Traits:
Size = thigh-high
Ears = down
Colors = 1
White = no
CoatLength = long
CoatType = wavy

Fig. S33. Dogs included in the MuttMix survey (page 2/4). Front and side photographs and descriptions of the 30 mutts (black border) and single undeclared purebred (red). All photos and information were taken and provided by the dog owners for use in the public survey and research project.

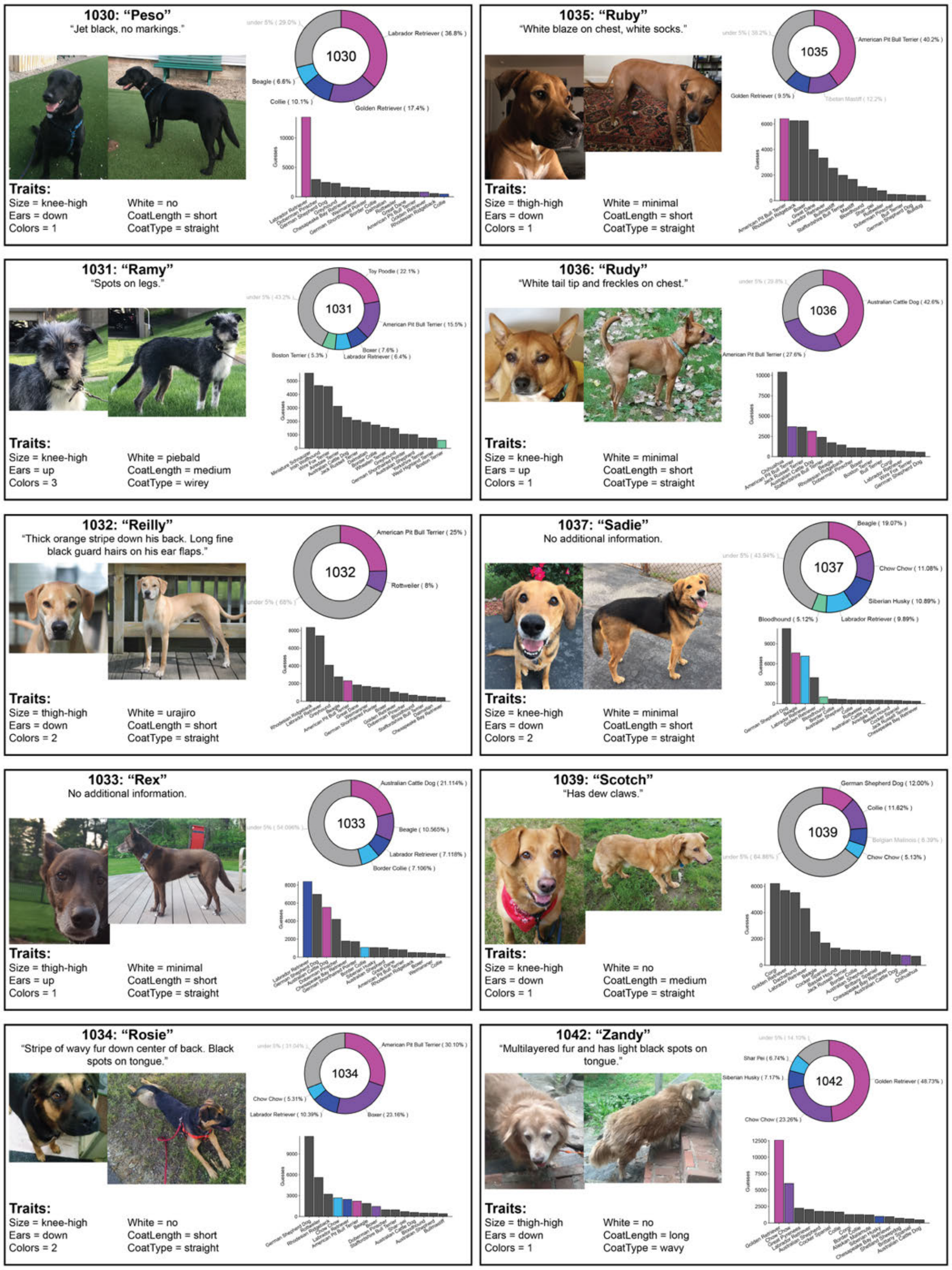


Fig. S33. Dogs included in the MuttMix survey (page 3/4). Front and side photographs and descriptions of the 30 mutts (black border) and single undeclared purebred (red). All photos and information were taken and provided by the dog owners for use in the public survey and research project.

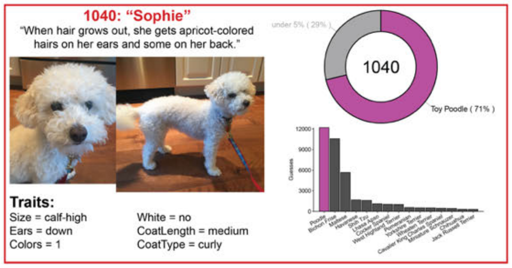


Fig. S33. Dogs included in the MuttMix survey (page 4/4). Front and side photographs and descriptions of the 30 mutts (black border) and single undeclared purebred (red). All photos and information were taken and provided by the dog owners for use in the public survey and research project.

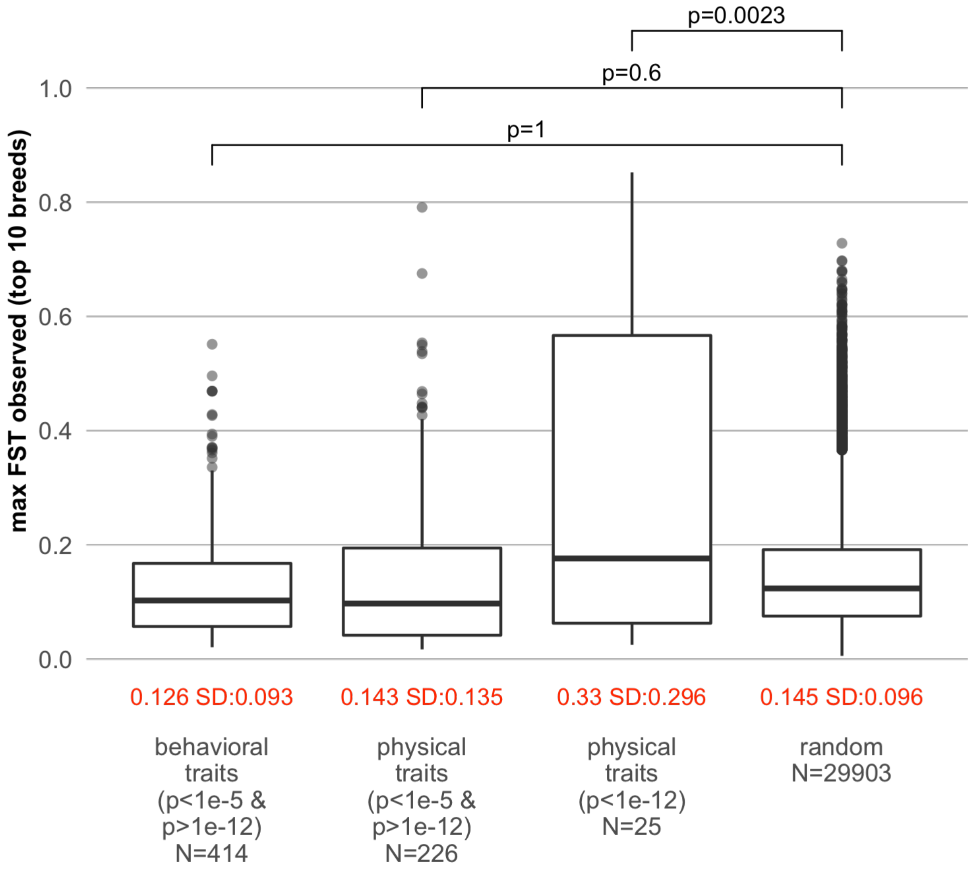


Fig. S34. SNPs strongly associated with physical traits only are more differentiated between breeds than random loci. For each SNP, we calculated the maximum F_{ST} observed between the one of the ten most common breeds in the GWAS dataset (American pit bull terrier, Australian cattle dog, beagle, border collie, Chihuahua, dachshund, German shepherd dog, golden retriever, Labrador retriever, toy poodle) and the full dog population. Brackets with p-values indicate whether each category of associated SNPs has higher F_{ST} values than randomly sampled SNPs (fourth column), measured using a one-sided t-test (*R* package `` `*stats*` `` version 4.1.1). Red text is the mean and standard deviation for each category. Boxplot: box encompasses the 25-75% of values with a thick horizontal line at the median; whiskers show the range of 5-95% of values; points represent values outside the 5-95% range.

Table S1. Survey questions in Darwin's Ark.

Survey	#	Source	Question type	Text	Shorthand	Range	Options*	# Responses	Note on question type
Play	1	CHQLS	other behavior	DOG enjoys life	Enjoys life	quantitative	Agree to disagree	17,637	
	2	CHQLS	motor pattern	DOG wants to play	Wants to play	quantitative	Agree to disagree	17,564	play is characterized by expression of motor patterns such as chase and grab-bite
	3	DIAS	other behavior	DOG becomes aggressive when excited	Aggressive when excited	quantitative	Agree to disagree	17,559	
	4	DPQ	motor pattern	DOG enjoys playing with toys	Enjoys playing with toys	quantitative	Agree to disagree	17,566	derivative of grab-bite motor pattern
	5	DPQ	other behavior	DOG gets bored in play quickly	Bored in play quickly	quantitative	Agree to disagree	17,524	
	6	DPQ	other behavior	DOG seeks constant activity	Seeks constant activity	quantitative	Agree to disagree	17,523	
	7	DIAS	other behavior	Excitement can lead DOG to fixed repetitive behavior	Excitement leads to repetitive behavior	quantitative	Agree to disagree	17,536	
	8	DIAS	other behavior	DOG calms down very quickly after being excited	Calms down quickly	quantitative	Agree to disagree	17,500	
	9	DIAS	other behavior	DOG may become aggressive if HE is frustrated with something	Aggressive if frustrated	quantitative	Agree to disagree	17,499	
	10	DIAS	other behavior	DOG is not very patient	Not patient	quantitative	Agree to disagree	17,127	
Communication	11	DPQ	other behavior	DOG is relaxed when greeting people	Relaxed when greeting people	quantitative	Agree to disagree	16,893	
	12	DPQ	other behavior	DOG is shy	Shy	quantitative	Agree to disagree	16,814	
	13	DPQ	other behavior	DOG is boisterous	Boisterous	quantitative	Agree to disagree	16,839	
	14	DIAS	other behavior	DOG does not think before HE acts	Does not think before acts	quantitative	Agree to disagree	16,796	
	15	DIAS	other behavior	DOG appears to be sorry after HE has done something wrong	Sorry when wrong	quantitative	Agree to disagree	16,790	
	16	DPQ	other behavior	DOG aggressively guards coveted items (e.g., stolen item, treats, food bowl)	Guards coveted items	quantitative	Agree to disagree	16,807	
	17	new	motor pattern	DOG howls	Howls	quantitative	Never to always	16,777	derivative of howling motor pattern
	18	new	motor pattern	DOG Woo-woo barks	Woo-woo barks	quantitative	Never to always	16,799	derivative of howling motor pattern
	19	new	other behavior	DOG whines when alone	Whines when alone	quantitative	Never to always	16,796	
	20	new	other behavior	DOG whines to get attention, food, or toys	Whines for attention food toys	quantitative	Never to always	16,389	
The Great Outdoors	21	DPQ	other behavior	DOG is able to focus on a task in a distracting situation (e.g., loud or busy places, around other dogs)	Focus in distracting situation	quantitative	Agree to disagree	16,519	
	22	DPQ	other behavior	DOG is playful with other dogs	Playful with dogs	quantitative	Agree to disagree	16,418	
	23	DPQ	other behavior	When off leash, DOG comes immediately when called	Comes when called	quantitative	Agree to disagree	16,413	
	24	DPQ	other behavior	DOG likes to chase squirrels, birds, or other small animals	Chases small animals	quantitative	Agree to disagree	16,413	
	25	CHQLS	other behavior	DOG moves normally	Moves normally	quantitative	Agree to disagree	16,452	
	26	CHQLS	other behavior	DOG is as active as HE has been	As active as has been	quantitative	Agree to disagree	16,414	
	27	DIAS	other behavior	DOG shows extreme physical signs when excited	Extreme physical signs when excited	quantitative	Agree to disagree	16,419	
	28	new	other behavior	DOG eats grass	Eats grass	quantitative	Never to always	16,413	
	29	new	motor pattern	DOG buries toys / bones	Buries toys	quantitative	Never to always	16,401	derivative of caching motor pattern
	30	new	other behavior	DOG eats non-food items	Eats non-food items	quantitative	Never to always	16,007	
Independence	31	DPQ	other behavior	DOG ignores commands	Ignores commands	quantitative	Agree to disagree	16,077	
	32	DPQ	other behavior	DOG is quick to sneak out through open doors, gates	Quick to sneak out doors	quantitative	Agree to disagree	15,969	
	33	CHQLS	other behavior	DOG sleeps more, is less awake	Sleeps more	quantitative	Agree to disagree	15,987	
	34	DPQ	other behavior	DOG is slow to respond to corrections	Slow to respond to corrections	quantitative	Agree to disagree	15,962	
	35	new	other behavior	DOG escapes from cages, kennels, or enclosures	Escapes from enclosures	quantitative	Never to always	15,979	
	36	CCDR	other behavior	DOG gets stuck behind objects and is unable to get around	Gets stuck behind objects	quantitative	Never to always	15,978	
	37	CCDR	other behavior	DOG walks away or avoids being patted	Avoids being patted	quantitative	Never to always	15,959	
	38	new	other behavior	DOG shows barrier aggression	Shows barrier aggression	quantitative	Never to always	15,997	
	39	CHQLS	other behavior	DOG lays in one place all day long	Lays in one place	quantitative	Never to always	15,986	
	40	new	other behavior	DOG damages doors, gates, or walls	Damages doors	quantitative	Never to always	15,584	

Survey	#	Source	Question type	Text	Shorthand	Range	Options*	# Responses	Note on question type
Socialization with Humans	41	DIAS	other behavior	DOG doesn't like to be approached or hugged	Dislikes being approached or hugged	quantitative	Agree to disagree	15,919	
	42	DPQ	other behavior	DOG is friendly towards unfamiliar people	Friendly to unfamiliar people	quantitative	Agree to disagree	15,822	
	43	CHQLS	other behavior	DOG responds to my presence	Responds to my presence	quantitative	Agree to disagree	15,836	
	44	DPQ	other behavior	DOG behaves aggressively in response to perceived threats from people (e.g., being cornered, having collar reached for)	Aggressive to threatening people	quantitative	Agree to disagree	15,853	
	45	new	other behavior	DOG is a people person	People person	quantitative	Agree to disagree	15,837	
	46	DPQ	other behavior	DOG behaves fearfully towards unfamiliar people	Fearful to unfamiliar people	quantitative	Agree to disagree	15,796	
	47	DIAS	other behavior	DOG seems to get excited for no reason	Excited for no reason	quantitative	Agree to disagree	15,817	
	48	DPQ	other behavior	DOG behaves aggressively towards unfamiliar people	Aggressive to unfamiliar people	quantitative	Agree to disagree	15,831	
	49	DPQ	other behavior	DOG seeks companionship from people	Seeks companionship from people	quantitative	Agree to disagree	15,825	
	50	new	other behavior	DOG must greet everyone who comes to the door	Must greet everyone	quantitative	Agree to disagree	15,428	
Hunting	51	DPQ	other behavior	DOG chases bicycles, joggers, and skateboarders	Chases bicycles or joggers	quantitative	Agree to disagree	15,342	
	52	DPQ	other behavior	DOG works at tasks (e.g., getting treats out of a Kong, shredding toys) until entirely finished	Works at tasks	quantitative	Agree to disagree	15,256	
	53	DPQ	other behavior	DOG leaves food or objects alone when told to do so	Leaves food or objects when told	quantitative	Agree to disagree	15,231	
	54	DPQ	motor pattern	DOG retrieves objects (e.g., balls, toys, sticks)	Retrieves objects	quantitative	Agree to disagree	15,269	derivative of chase/grab-bite motor pattern
	55	DIAS	other behavior	DOG appears to have a lot of control over how HE responds	Control over responses	quantitative	Agree to disagree	15,250	
	56	DIAS	other behavior	DOG reacts very quickly	Reacts very quickly	quantitative	Agree to disagree	15,271	
	57	DIAS	other behavior	DOG is easy to train	Easy to train	quantitative	Agree to disagree	15,286	
	58	DIAS	other behavior	DOG can be very persistent	Very persistent	quantitative	Agree to disagree	15,266	
	59	new	motor pattern	DOG points	Points	quantitative	Agree to disagree	15,249	derivative of eye-stalk motor pattern
	60	new	physical trait related	DOG avoids getting wet	Avoids getting wet	quantitative	Agree to disagree	14,882	potentially influenced by presence or absense of undercoat
The Ins and Outs	61	CCDR	other behavior	DOG has difficulty finding food dropped on the floor	Difficulty finding dropped food	quantitative	Never to always	15,110	
	62	new	other behavior	DOG licks HIS empty bowl after finishing the food	Licks empty bowl	quantitative	Never to always	15,016	
	63	new	other behavior	DOG licks tile / linoleum floors	Licks tile floors	quantitative	Never to always	15,038	
	64	new	other behavior	DOG turns in circles before pooping	Circles before pooping	quantitative	Never to always	15,031	
	65	new	other behavior	DOG kicks or scratches the ground after defecating	Kicks ground after defecating	quantitative	Never to always	15,026	
	66	new	physical trait related	DOG lifts HIS leg to urinate	Lifts leg to urinate	quantitative	Never to always	15,024	potentially related to hip/joint morphology or leg length
	67	new	other behavior	DOG marks with feces	Marks with feces	quantitative	Never to always	15,021	
	68	new	other behavior	DOG eats HIS own feces	Eats own feces	quantitative	Never to always	15,068	
	69	new	other behavior	DOG shows submissive urination	Submissive urination	quantitative	Never to always	15,023	
	70	new	physical trait related	DOG drinks water quickly	Drinks water quickly	quantitative	Never to always	14,677	potentially related to dog size and/or face shape
Socialization with Animals	71	DPQ	other behavior	DOG behaves aggressively toward other dogs	Aggressive to dogs	quantitative	Agree to disagree	14,903	
	72	DPQ	other behavior	DOG willingly shares HIS toys with other dogs	Shares toys with dogs	quantitative	Agree to disagree	14,798	
	73	DPQ	other behavior	DOG is dominant over other dogs	Dominant over dogs	quantitative	Agree to disagree	14,842	
	74	DPQ	other behavior	DOG avoids other dogs	Avoids dogs	quantitative	Agree to disagree	14,833	
	75	DPQ	other behavior	DOG is friendly towards other dogs	Friendly to dogs	quantitative	Agree to disagree	14,839	
	76	new	other behavior	DOG knows HE is a dog	Knows is a dog	quantitative	Agree to disagree	14,823	
	77	DPQ	other behavior	DOG behaves fearfully towards other dogs	Fearful to dogs	quantitative	Agree to disagree	14,818	
	78	DPQ	other behavior	DOG behaves aggressively towards cats	Aggressive to cats	quantitative	Agree to disagree	14,814	
	79	DPQ	other behavior	DOG is assertive or pushy with other dogs (e.g., if in a home with other dogs, when greeting)	Assertive with dogs	quantitative	Agree to disagree	14,834	
	80	CCDR	other behavior	DOG sometimes fails to recognize familiar people or pets	Fails to recognize familiar people	quantitative	Agree to disagree	14,434	

Survey	#	Source	Question type	Text	Shorthand	Range	Options*	# Responses	Note on question type
Coping with the Unfamiliar	81	DIAS, DPQ	other behavior	DOG is very interested in and adapts easily to new things and new places	Interested in new things	quantitative	Agree to disagree	14,599	
	82	DPQ	other behavior	DOG exhibits fearful behaviors when HE is restrained	Fearful when restrained	quantitative	Agree to disagree	14,503	
	83	DPQ	other behavior	DOG behaves fearfully when groomed (e.g., nails trimmed, brushed, bathed, ears cleaned)	Fearful when groomed	quantitative	Agree to disagree	14,546	
	84	DIAS	other behavior	DOG is not keen to go into new situations	Not keen on new situations	quantitative	Agree to disagree	14,494	
	85	DPQ	other behavior	DOG behaves aggressively during visits to the veterinarian	Aggressive at vet	quantitative	Agree to disagree	14,511	
	86	new	other behavior	DOG is highly sensitive to noise	Sensitive to noise	quantitative	Agree to disagree	14,526	
	87	DPQ	other behavior	DOG shows aggression when nervous or fearful	Aggression when nervous	quantitative	Agree to disagree	14,516	
	88	DPQ	other behavior	DOG behaves fearfully during visits to the veterinarian	Fearful at vet	quantitative	Agree to disagree	14,503	
	89	DIAS	other behavior	DOG takes a long time to lose interest in new things	Slow to lose interest in new	quantitative	Agree to disagree	14,523	
	90	new	other behavior	DOG is afraid of storms	Afraid of storms	quantitative	Agree to disagree	14,187	
Body Language	91	DPQ	other behavior	DOG behaves submissively (e.g., rolls over, avoids eye contact, licks HIS lips) when greeting other dogs	Submissive to dogs	quantitative	Agree to disagree	14,485	
	92	new	physical trait related	DOG rests frog style	Rests frog style	quantitative	Never to always	14,415	potentially related to hip morphology
	93	CCDR	other behavior	DOG paces up and down, walks in circles and/or wanders with no direction or purpose	Paces or circles or wanders	quantitative	Agree to disagree	14,376	
	94	CHQLS	physical trait related	DOG pants frequently, even at rest	Pants frequently	quantitative	Agree to disagree	14,395	potentially related to dog size and/or face shape
	95	CCDR	other behavior	DOG stares blankly at the walls or floor	Stares blankly	quantitative	Never to always	14,385	
	96	new	physical trait related	DOG tilts HIS head	Tilts head	quantitative	Never to always	14,424	potentially related to ear shape or skull shape
	97	CHQLS	physical trait related	DOG shakes or trembles occasionally	Shakes occasionally	quantitative	Agree to disagree	14,388	potentially related to dog size and capacity to thermoregulate
	98	new	physical trait related	DOG crosses HIS front paws	Crosses front paws	quantitative	Never to always	14,393	potentially influenced by bone & joint morphology
	99	new	other behavior	DOG shows handedness or side-preference	Shows handedness	quantitative	Never to always	14,415	
	100	new	other behavior	DOG places HIS paw on my or other people's feet	Places paw on people's feet	quantitative	Never to always	14,059	
Attitude	101	DPQ	other behavior	DOG is lethargic	Lethargic	quantitative	Agree to disagree	14,095	
	102	DPQ	other behavior	DOG is confident	Confident	quantitative	Agree to disagree	14,055	
	103	CHQLS	other behavior	DOG seems dull or depressed, not alert	Seems dull or depressed	quantitative	Agree to disagree	14,049	
	104	DPQ	other behavior	DOG is anxious	Anxious	quantitative	Agree to disagree	14,073	
	105	DIAS	other behavior	DOG is considered to be very impulsive	Impulsive	quantitative	Agree to disagree	14,037	
	106	DPQ	other behavior	DOG is curious	Curious	quantitative	Agree to disagree	14,059	
	107	DPQ	other behavior	DOG is affectionate	Affectionate	quantitative	Agree to disagree	14,079	
	108	DPQ	other behavior	DOG tends to be calm	Calm	quantitative	Agree to disagree	14,047	
	109	DPQ	other behavior	DOG is aloof	Aloof	quantitative	Agree to disagree	14,069	
	110	CHQLS	other behavior	DOG has more good days than bad days	More good days	quantitative	Agree to disagree	13,679	
Physical Traits	121	physical trait	physical trait	When DOG is standing next to someone of average height, how high are HIS shoulders?	Size	quantitative	ankle high or shorter | calf high | knee high | thigh high | hip high or higher	16,286	size
	122	physical trait	physical trait	What color is DOG? Select all that apply.	Fur color	multiple choice	white | red (from pale peach to dark red or liver colored) | yellow (from pale cream to orange, gold, fawn, or wheaten) | gray (slate, blue gray, charchoal, or silver) | chocolate brown | pure black | merle | brindle	16,579	pigmentation
	123	physical trait	physical trait	How much white fur does DOG have? Select the image with the closest amount of white.	White fur	quantitative	image options (*see figure S3A*)	16,436	pigmentation
	124	physical trait	physical trait	Is DOG's tail curly?	Curly tail	binary	no | yes | I don't know	16,081	tail shape
	125	physical trait	physical trait	What is DOG's ear shape? Select the image with the closest ear shape.	Ear shape	categorical	image options (*see figure S3B*)	16,105	ear shape
	126	physical trait	physical trait	Are DOG's eyes different colors?	Eyes different colors	binary	no | yes | I don't know	15,993	pigmentation
	127	physical trait	physical trait	How long is DOG's fur on HIS back and sides? Measure it against your finger.	Fur length	quantitative	short | medium | long | I'm not sure	16,117	fur type
	128	physical trait	physical trait	Does DOG have soft fur or rough and bristly fur?	Fur texture	binary	soft | rough (wiry) | I'm not sure	15,632	fur type

* Agree to Disagree: Strongly Agree | Agree | Neither Agree Nor Disagree | Disagree | Strongly Disagree

Table S2. Breed information, stereotypes, and reference panel.

			Breed Reference Panel							**Encyclopedia of Dog Breeds**									
Breed	**AKC Registrations (2000-2015)**	**% candidate purebred dogs (Darwin's Ark)**	**Illumina array**	**Axiom array**	**Low pass + imputation**	**30x WGS**	**Sources ***	**AKC Group**	**AKC Three-word Descriptors**	**energy level**	**exercise requirements**	**playfulness**	**affection level**	**friendliness towards**	**friendliness towards**	**friendliness towards**	**ease of training**	**watchdog ability**	**protection ability**
labrador retriever	15.18%	10.29%	0	0	0	12	1	sporting	active, friendly, outgoing	4	3	5	5	4	5	5	5	4	2
german shepherd dog	5.95%	4.86%	0	0	0	12	1	herding	confident, courageous, smart	3	4	2	3	1	3	2	5	5	5
golden retriever	5.52%	4.23%	0	0	0	12	1; 2	sporting	devoted, friendly, intelligent	3	3	5	5	5	5	5	5	3	2
chihuahua	2.54%	4.08%	8	0	1	3	3; 1; 4	toy	charming, graceful, sassy	5	1	2	2	1	3	1	2	5	1
american pit bull terrier	#N/A	3.93%	0	0	12	0	4												
border collie	0.33%	3.68%	0	0	5	7	1; 4	herding	affectionate, energetic, smart	5	5	4	3	3	1	2	5	4	3
australian shepherd	0.87%	3.02%	3	1	8	0	3; 4	herding	exuberant, smart, work-oriented	4	5	5	4	3	3	2	5	5	4
beagle	4.80%	2.59%	0	5	2	5	1; 4	hound	curious, friendly, merry	3	3	3	5	5	5	5	1	4	1
dachshund	3.85%	2.57%	3	1	3	5	3; 1; 4	hound	curious, friendly, spunky	4	2	3	3	3	2	1	2	5	2
poodle	3.37%	2.31%	10	0	2	0	3; 4	non-sporting	active, proud, very smart	3	4	4	3	3	4	3	5	5	4
jack russell terrier	0.14%	2.27%	7	0	2	3	3; 1; 4	terrier	alert, inquisitive, lively	5	4	5	3	2	1	3	3	5	2
australian cattle dog	0.20%	2.18%	3	2	4	3	3; 1; 4	herding	alert, curious, pleasant	5	5	4	3	3	2	2	5	5	4
boxer	3.73%	2.06%	10	0	1	1	3; 1; 4	working	active, bright, fun-loving	4	3	4	5	3	3	3	3	3	3
siberian husky	1.35%	1.70%	8	1	0	3	3; 1; 4	working	loyal, mischievous, outgoing	3	4	4	4	3	3	5	1	3	1
shih tzu	2.80%	1.53%	11	0	1	0	3; 4	toy	affectionate, outgoing, playful	3	2	4	4	4	4	4	2	3	1
yorkshire terrier	4.74%	1.44%	0	0	1	11	1; 4		affectionate, sprightly, tomboyish	4	1	4	3	2	2	3	2	5	1
toy poodle	#N/A	1.36%	10	1	1	0	3; 4	toy	agile, intelligent, self-confident	4	2	5	4	3	3	4	5	5	1
pug	1.90%	1.29%	7	1	4	0	3; 4	toy	charming, loving, mischievous	3	2	4	4	3	3	2	3	5	1
miniature schnauzer	2.38%	1.20%	10	1	1	0	3; 4	terrier	friendly, obedient, smart	3	3	4	4	3	3	3	3	5	1
boston terrier	1.57%	1.01%	6	5	0	1	3; 1; 4	non-sporting	amusing, bright, friendly	3	1	3	3	3	4	3	3	5	1
pomeranian	2.12%	1.01%	9	1	1	1	3; 1; 4	toy	bold, inquisitive, lively	4	1	4	1	1	3	1	1	5	1
pembroke welsh corgi	1.11%	0.97%	9	1	0	2	3; 1; 4	herding	affectionate, alert, smart	3	3	3	4	3	4	4	4	5	4
doberman pinscher	1.44%	0.94%	8	1	0	3	3; 1; 4	working	alert, fearless, loyal	4	3	3	3	1	3	1	5	5	5
shetland sheepdog	1.62%	0.94%	2	3	4	3	3; 1; 4	herding	bright, energetic, playful	3	3	3	3	4	4	1	5	5	1
cavalier king charles spaniel	0.86%	0.87%	10	0	2	0	3; 4	toy	affectionate, gentle, graceful	3	3	4	5	1	4	5	4	3	1
rottweiler	2.26%	0.87%	0	0	2	10	3; 1; 2	working	confident guardian, loving, loyal	3	3	2	2	1	2	1	3	5	5
greyhound	0.03%	0.86%	0	0	4	8	1; 4	hound	gentle, independent, noble	2	3	3	3	3	2	3	3	4	2
great pyrenees	0.25%	0.76%	9	2	0	1	3; 1; 4	working	calm, patient, smart	1	3	2	3	2	3	2	1	5	4
collie	0.61%	0.75%	9	3	0	0	3; 4	herding	devoted, graceful, proud	3	3	3	3	3	4	3	3	5	3
great dane	1.19%	0.75%	5	0	3	4	3; 1; 4	working	dependable, friendly, patient	2	3	2	4	3	3	4	4	4	3
maltese	1.37%	0.72%	9	1	2	0	3; 4	toy	charming, gentle, playful	4	1	4	3	3	3	1	3	5	1
staffordshire bull terrier	0.11%	0.72%	12	0	0	0	3	terrier	brave, clever, tenacious	3	3	5	4	2	3	3	3	3	3
shiba inu	0.24%	0.69%	3	6	1	2	3; 1; 4	non-sporting	active, alert, attentive	3	3	3	3	2	2	3	2	5	3
basset hound	0.90%	0.62%	9	2	1	0	3; 1; 4	hound	charming, low-key, patient	2	2	2	4	3	3	3	2	4	1
german shorthaired pointer	1.49%	0.48%	7	5	0	0	3; 1	sporting	friendly, smart, willing to please	5	5	4	4	3	2	3	3	5	3
american cocker spaniel	1.74%	0.82%	4	1	1	6	3; 1; 4	sporting	smart, happy, gentle	3	3	4	5	5	5	4	4	4	1
papillon	0.56%	0.64%	10	1	1	0	3; 4												
bichon frise	0.85%	0.62%	12	0	0	0	3												
french bulldog	0.81%	0.62%	8	1	2	2	3; 1; 4												
havanese	0.49%	0.62%	9	0	3	0	3; 4												
belgian malinois	0.12%	0.55%	4	1	1	6	3; 1; 5; 4												
vizsla	0.43%	0.54%	8	3	1	1	3; 1; 4												
brittany	0.87%	0.53%	11	0	1	0	3; 4												
bernese mountain dog	0.44%	0.51%	7	0	2	3	3; 1; 4												
english bulldog	2.73%	0.51%	9	1	1	1	3; 1; 4												
miniature pinscher	1.14%	0.51%	11	1	0	0	3; 4												
rhodesian ridgeback	0.31%	0.51%	7	0	1	4	3; 1; 4												
west highland white terrier	0.74%	0.51%	0	1	2	9	1; 4												
whippet	0.22%	0.46%	4	6	1	1	3; 1; 4												
english springer spaniel	0.97%	0.44%	10	0	0	2	3; 1												
cairn terrier	0.36%	0.36%	11	0	1	0	3; 4												
mastiff	0.78%	0.36%	10	0	0	2	3; 1												
lhasa apso	0.42%	0.35%	12	0	0	0	3												
alaskan malamute	0.26%	0.33%	9	0	1	2	3; 1; 4												
airedale terrier	0.29%	0.31%	8	0	1	3	3; 1; 4												

			Breed Reference Panel							Encyclopedia of Dog Breeds									
Breed	AKC Registrations (2000-2015)	% candidate purebred dogs (Darwin's Ark)	Illumina array	Axiom array	Low pass + imputation	30x WGS	Sources *	AKC Group	AKC Three-word Descriptors	energy level	exercise requirements	playfulness	affection level	friendliness towards	friendliness towards	friendliness towards	ease of training	watchdog ability	protection ability
akita	0.38%	0.31%	11	0	1	0	3; 4												
weimaraner	0.85%	0.31%	11	0	0	1	3; 1												
chow chow	0.22%	0.30%	3	7	1	1	3; 1; 4												
dalmatian	0.17%	0.30%	10	0	0	2	3; 1												
portuguese water dog	0.19%	0.30%	0	2	2	8	3; 1; 4												
newfoundland	0.39%	0.29%	11	0	0	1	3; 5												
italian greyhound	0.24%	0.26%	4	6	0	2	3; 1												
saint bernard	0.47%	0.26%	10	0	0	2	3; 1												
border terrier	0.11%	0.25%	8	0	0	4	3; 1												
nova scotia duck tolling retriever	0.07%	0.25%	11	0	0	1	3; 1												
bullmastiff	0.40%	0.24%	7	2	2	1	3; 1; 4												
english setter	0.09%	0.24%	7	0	2	3	3; 1; 4												
soft coated wheaten terrier	0.24%	0.24%	9	1	0	2	3; 1; 4												
english cocker spaniel	0.17%	0.23%	8	0	0	4	3; 1												
english shepherd	non-AKC	0.22%	10	0	2	0	6; 4												
samoyed	0.17%	0.22%	10	0	0	2	3; 1												
shar pei	0.41%	0.22%	12	0	0	0	3												
tibetan terrier	0.09%	0.22%	1	8	1	2	3; 1; 4												
chinese crested	0.24%	0.20%	10	0	0	2	3; 1												
basenji	0.12%	0.19%	10	0	0	2	3; 1												
scottish terrier	0.37%	0.19%	8	0	0	4	3; 1												
irish setter	0.18%	0.18%	9	2	0	1	3; 1; 4												
bloodhound	0.36%	0.17%	3	8	0	1	3; 1; 4												
bull terrier	0.21%	0.17%	5	0	0	7	3; 1												
chesapeake bay retriever	0.40%	0.17%	10	0	1	1	3; 1; 4												
irish wolfhound	0.12%	0.17%	3	0	2	7	3; 1; 4												
tibetan spaniel	0.06%	0.17%	12	0	0	0	3												
pekingese	0.44%	0.15%	11	0	0	1	3; 1												
borzoi	0.09%	0.14%	11	0	0	1	3; 1												
greater swiss mountain dog	0.09%	0.14%	7	3	1	2	3; 1; 4												
leonberger	N/A	0.14%	0	0	0	12	1; 2												
saluki	0.04%	0.14%	7	1	1	3	3; 1; 4												
bearded collie	N/A	0.13%	0	0	5	7	1; 4												
afghan hound	0.09%	0.12%	10	0	0	2	3; 1												
norwegian elkhound	0.09%	0.11%	10	0	0	2	3; 1												
schipperke	0.12%	0.11%	12	0	0	0	3												
wire fox terrier	0.12%	0.11%	12	0	0	0	3												
old english sheepdog	0.16%	0.10%	5	6	1	0	3; 1; 4												
belgian tervuren	0.06%	0.09%	2	0	0	10	3; 1												
chinook	N/A	0.08%	9	1	1	1	3; 1; 4												
wirehaired pointing griffon	0.07%	0.08%	11	0	1	0	3; 4												
tibetan mastiff	N/A	0.07%	0	1	0	11	1; 4												
gordon setter	0.09%	0.04%	10	0	0	2	3; 1												
norfolk terrier	0.04%	0.04%	12	0	0	0	3												
finnish spitz	0.01%	0.03%	12	0	0	0	3												
entlebucher	N/A	N/A	0	0	0	12	7; 2												

Sources:

1 NHGRI Dog Genome Project (Elaine Ostrander)(*40*)
2 Broad Institute (BioProject PRJNA683923)
3 Cornell Canine Dataset (Hayward et al 2016, doi:10.1038/ncomms10460) (*118*)
4 Darwin's Ark (BioProject PRJNA675863)
5 BarkBase (Megquier et al 2019, doi:10.3390/genes10060433) (*102*)
6 Embark Veterinary, Inc. (Darwin's Ark participant-submitted raw data)
7 National Entlebucher Mountain Dog Association (BioProject PRJNA683923)

Table S3. Descriptive statistics for the exploratory factor analysis (110 items, 10,252 dogs).

Statistic	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
Mean	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Std Dev	0.94	0.91	0.89	0.89	0.87	0.92	0.84	0.87	0.82	0.86	0.78	0.78	0.78	0.76	0.82	0.75	0.73	0.72	0.71	0.75
Range	4.98	6.01	5.46	5.73	5.86	6.2	7.24	7.18	6.81	5.83	7.12	5.26	5.82	5.86	6.54	7.64	6.75	6.04	7.22	7.88
Minimum	-3.08	-3.55	-2.05	-2.54	-3.18	-3.64	-2.48	-2.5	-3.14	-3.08	-3.24	-2.78	-2.19	-2.25	-3.07	-3.57	-3.77	-3.07	-2.46	-3.2
Maximum	1.9	2.46	3.41	3.19	2.68	2.56	4.76	4.68	3.67	2.75	3.88	2.48	3.63	3.61	3.47	4.07	2.98	2.97	4.76	4.68
25% Percentile	-0.58	-0.6	-0.68	-0.63	-0.54	-0.56	-0.59	-0.6	-0.55	-0.57	-0.5	-0.54	-0.54	-0.54	-0.56	-0.5	-0.49	-0.48	-0.49	-0.48
50% Percentile	0.23	0.09	-0.19	-0.07	0.12	0.13	-0.06	-0.14	0.01	0.04	0.07	0.08	-0.1	-0.04	-0.1	-0.01	-0.01	-0.01	-0.03	0.01
75% Percentile	0.74	0.67	0.5	0.59	0.64	0.68	0.53	0.49	0.56	0.6	0.57	0.58	0.41	0.51	0.46	0.48	0.47	0.48	0.46	0.47
Sum of Squared Loading	3.96	4.46	2.66	2.94	3.24	2.98	4.01	2.42	2.54	1.97	1.53	1.42	1.23	1.3	1.81	1.44	1.42	1.05	1.54	1.28
Proportion Variance	0.04	0.04	0.02	0.03	0.03	0.03	0.04	0.02	0.02	0.02	0.01	0.01	0.01	0.01	0.02	0.01	0.01	0.01	0.01	0.01
Cumulative Variance	0.04	0.08	0.1	0.13	0.16	0.18	0.22	0.24	0.27	0.28	0.3	0.31	0.32	0.33	0.35	0.36	0.38	0.39	0.4	0.41

Table S4. Factors discovered from Darwin's Ark survey data.

Pattern	Structure		Question	Scale
Factor 1: Human Sociability			{ - less sociable - | + highly sociable + }	
0.920	0.820	46	Fearful to unfamiliar people	Strongly agree to strongly disagree
-0.857	-0.774	42	DOG is friendly towards unfamiliar people	Strongly agree to strongly disagree
-0.632	-0.710	45	DOG is a people person	Strongly agree to strongly disagree
0.770	0.709	12	DOG is shy	Strongly agree to strongly disagree
0.434	0.569	41	DOG doesn't like to be approached or hugged	Strongly agree to strongly disagree
0.520	0.530	48	DOG behaves aggressively towards unfamiliar people	Strongly agree to strongly disagree
-0.527	-0.524	50	DOG must greet everyone who comes to the door	Strongly agree to strongly disagree
NA	0.471	84	DOG is not keen to go into new situations	Strongly agree to strongly disagree
NA	-0.451	81	DOG is very interested in and adapts easily to new things and new places	Strongly agree to strongly disagree
NA	-0.432	49	DOG seeks companionship from people	Strongly agree to strongly disagree
NA	-0.414	102	DOG is confident	Strongly agree to strongly disagree
NA	0.385	44	DOG behaves aggressively in response to perceived threats from people (e.g., being cornered, having collar reached for)	Strongly agree to strongly disagree
NA	-0.358	37	DOG walks away or avoids being patted	Never to always
NA	0.358	88	DOG behaves fearfully during visits to the veterinarian	Strongly agree to strongly disagree
NA	0.345	104	DOG is anxious	Strongly agree to strongly disagree
NA	-0.336	11	DOG is relaxed when greeting people	Strongly agree to strongly disagree
NA	0.312	82	DOG exhibits fearful behaviors when HE is restrained	Strongly agree to strongly disagree
Factor 2: Arousal Level			{ - aroused - | + composed + }	
-0.759	-0.707	108	DOG tends to be calm	Strongly agree to strongly disagree
0.742	0.668	105	DOG is considered to be very impulsive	Strongly agree to strongly disagree
0.684	0.596	14	DOG does not think before HE acts	Strongly agree to strongly disagree
-0.579	-0.587	8	DOG calms down very quickly after being excited	Strongly agree to strongly disagree
0.544	0.551	47	DOG seems to get excited for no reason	Strongly agree to strongly disagree
0.543	0.533	13	DOG is boisterous	Strongly agree to strongly disagree
0.561	0.531	10	DOG is not very patient	Strongly agree to strongly disagree
0.446	0.507	6	DOG seeks constant activity	Strongly agree to strongly disagree
-0.513	-0.449	55	DOG appears to have a lot of control over how HE responds	Strongly agree to strongly disagree
0.410	0.435	27	DOG shows extreme physical signs when excited	Strongly agree to strongly disagree
-0.383	-0.398	11	DOG is relaxed when greeting people	Strongly agree to strongly disagree
0.301	0.385	7	Excitement can lead DOG to fixed repetitive behavior	Strongly agree to strongly disagree
NA	0.330	104	DOG is anxious	Strongly agree to strongly disagree
NA	0.329	93	DOG paces up and down, walks in circles and/or wanders with no direction or purpose	Strongly agree to strongly disagree
NA	-0.303	21	DOG is able to focus on a task in a distracting situation (e.g., loud or busy places, around other dogs)	Strongly agree to strongly disagree
NA	0.303	3	DOG becomes aggressive when excited	Strongly agree to strongly disagree

Pattern	Structure		Question	Scale
Factor 3: Toy-directed Motor Patterns			**{ - MPs toy-directed - | + MPs not toy-directed + }**	
0.943	0.773	4	DOG enjoys playing with toys	Strongly agree to strongly disagree
0.751	0.721	2	DOG wants to play	Strongly agree to strongly disagree
0.719	0.651	54	DOG retrieves objects (e.g., balls, toys, sticks)	Strongly agree to strongly disagree
-0.554	-0.544	5	DOG gets bored in play quickly	Strongly agree to strongly disagree
NA	0.374	6	DOG seeks constant activity	Strongly agree to strongly disagree
NA	-0.317	33	DOG sleeps more, is less awake	Strongly agree to strongly disagree
NA	-0.316	101	DOG is lethargic	Strongly agree to strongly disagree
NA	0.308	1	DOG enjoys life	Strongly agree to strongly disagree
NA	0.305	89	DOG takes a long time to lose interest in new things	Strongly agree to strongly disagree
Factor 4: Biddability			**{ - biddable - | + independent + }**	
-0.778	-0.707	31	DOG ignores commands	Strongly agree to strongly disagree
0.675	0.639	23	When off leash, DOG comes immediately when called	Strongly agree to strongly disagree
0.613	0.622	57	DOG is easy to train	Strongly agree to strongly disagree
-0.612	-0.597	34	DOG is slow to respond to corrections	Strongly agree to strongly disagree
0.463	0.508	53	DOG leaves food or objects alone when told to do so	Strongly agree to strongly disagree
-0.467	-0.485	32	DOG is quick to sneak out through open doors, gates	Strongly agree to strongly disagree
0.360	0.422	21	DOG is able to focus on a task in a distracting situation (e.g., loud or busy places, around other dogs)	Strongly agree to strongly disagree
NA	0.315	55	DOG appears to have a lot of control over how HE responds	Strongly agree to strongly disagree
Factor 5: Agonistic Threshold			**{ - assertive - | + diffident + }**	
0.944	0.682	9	DOG may become aggressive if HE is frustrated with something	Strongly agree to strongly disagree
0.619	0.606	87	DOG shows aggression when nervous or fearful	Strongly agree to strongly disagree
0.750	0.581	3	DOG becomes aggressive when excited	Strongly agree to strongly disagree
0.550	0.551	44	DOG behaves aggressively in response to perceived threats from people (e.g., being cornered, having collar reached for)	Strongly agree to strongly disagree
0.431	0.497	48	DOG behaves aggressively towards unfamiliar people	Strongly agree to strongly disagree
0.521	0.454	16	DOG aggressively guards coveted items (e.g., stolen item, treats, food bowl)	Strongly agree to strongly disagree
0.331	0.428	85	DOG behaves aggressively during visits to the veterinarian	Strongly agree to strongly disagree
NA	0.382	71	DOG behaves aggressively toward other dogs	Strongly agree to strongly disagree
-0.383	-0.373	38	DOG shows barrier aggression	Never to always
Factor 6: Dog Sociability			**{ - less sociable - | + highly sociable + }**	
-0.914	-0.792	75	DOG is friendly towards other dogs	Strongly agree to strongly disagree
-0.847	-0.742	22	DOG is playful with other dogs	Strongly agree to strongly disagree
0.724	0.654	74	DOG avoids other dogs	Strongly agree to strongly disagree
0.591	0.563	71	DOG behaves aggressively toward other dogs	Strongly agree to strongly disagree
0.513	0.521	77	DOG behaves fearfully towards other dogs	Strongly agree to strongly disagree
-0.318	-0.337	72	DOG willingly shares HIS toys with other dogs	Strongly agree to strongly disagree

Pattern	Structure		Question	Scale
Factor 7: Environmental Engagement			**{ - high engagement - | + low engagement + }**	
-0.734	-0.467	101	DOG is lethargic	Strongly agree to strongly disagree
-0.692	-0.461	103	DOG seems dull or depressed, not alert	Strongly agree to strongly disagree
0.490	0.427	56	DOG reacts very quickly	Strongly agree to strongly disagree
0.456	0.403	61	DOG has difficulty finding food dropped on the floor	Never to always
0.423	0.373	36	DOG gets stuck behind objects and is unable to get around	Never to always
0.480	0.369	95	DOG stares blankly at the walls or floor	Never to always
0.493	0.345	39	DOG lays in one place all day long	Never to always
0.407	0.342	58	DOG can be very persistent	Strongly agree to strongly disagree
0.342	0.335	106	DOG is curious	Strongly agree to strongly disagree
-0.336	-0.320	93	DOG paces up and down, walks in circles and/or wanders with no direction or purpose	Strongly agree to strongly disagree
-0.486	-0.315	33	DOG sleeps more, is less awake	Strongly agree to strongly disagree
-0.347	NA	94	DOG pants frequently, even at rest	Strongly agree to strongly disagree
0.310	NA	25	DOG moves normally	Strongly agree to strongly disagree
Factor 8: Proximity Seeking			**{ - affectionate - | + aloof + }**	
0.732	0.646	107	DOG is affectionate	Strongly agree to strongly disagree
0.668	0.584	49	DOG seeks companionship from people	Strongly agree to strongly disagree
0.539	0.499	37	DOG walks away or avoids being patted	Never to always
-0.458	-0.439	109	DOG is aloof	Strongly agree to strongly disagree
0.446	0.424	45	DOG is a people person	Strongly agree to strongly disagree
-0.423	-0.409	41	DOG doesn't like to be approached or hugged	Strongly agree to strongly disagree
0.337	0.326	43	DOG responds to my presence	Strongly agree to strongly disagree

Table S5. Biallelic SNPs unique to sampled breed populations.

Population	random N	all SNPs	autosomal	X
mutt	10	375,474	361,663	13,811
belgian tervuren	10	19,018	18,165	853
border collie	10	51,279	48,421	2,858
entlebucher	10	27,627	26,465	1,162
german shepherd dog	10	21,211	19,127	2,084
golden retriever	10	35,055	33,513	1,542
greyhound	10	46,073	43,535	2,538
labrador retriever	10	43,346	41,462	1,884
leonberger	10	36,611	34,972	1,639
portuguese water dog	10	40,372	39,328	1,044
rottweiler	10	39,021	37,609	1,412
tibetan mastiff	10	651,551	629,712	21,839
west highland white terrier	10	30,757	29,579	1,178
yorkshire terrier	10	66,791	62,703	4,088
	breed mean:	85,286	101,875	4,138
	breed SD:	170,611	165,032	5,613

Table S6. Concordance of owner-reported breed with genetically inferred breed ancestry.

Classification	Group	Owner report	Reference breed	N	Top breed at 85%+	Breed matches top ancestry	Breed matches at 85%+
preliminary	non-purebred	0 or 2 breeds	either	1186	55	402	20
preliminary	single breed	1 breed	present	880	517	752	511
			absent	89	12	0	0
preliminary	registered purebred	1 breed & registered purebred	present	304	272	299	271
			absent	19	2	0	0
final	mutt	0 or 2 breeds, or <45% breed ancestry	either	1221	0	382	0
final	candidate purebred	1 breed	either	934	584	772	541
			present	814	541	772	531
			absent	120	43	0	0
final	confirmed purebred	1 breed reported & registered purebred, or 85%+ breed ancestry	either	633	584	559	531
			present	573	541	559	531
			absent	60	43	0	0

Table S7. The correlation of population peculiarity scores among traits.

	Confirmed Purebreds					Candidate Purebreds				
Comparison	**N**	**r**	**p**	**5% CI**	**95% CI**	**N**	**r**	**p**	**5% CI**	**95% CI**
behavioral vs. physical	3,790	0.045	0.0059	0.013	0.076	46,202	0.064	1.20E-43	0.055	0.073
behavioral trait questions	27,137	0.055	2.30E-19	0.043	0.066	353,925	0.137	0	0.134	0.141
physical trait questions	120	-0.101	0.275	-0.275	0.08	1,296	0.018	0.528	-0.037	0.072

Table S8. Validation of breed ancestry effects against PPS and breed standards.

Breed Standard Trait	**# breeds***	**% correctly predicted**	**Counts (by predicted effect)**	**Discordant Breed(s)**
size**	17	100%	9 ankle high; 8 hip high; 1 other	
white	3	100%	3 all white	
ear shape	22	96%	14 dropped; 8 pricked	jack russell terrier
fur length	9	100%	6 long; 3 short	
all	**51**	**98%**		

* breeds that are significant in LMER analysis

** bull terrier excluded because it comes in two sizes (miniature and standard), and LMER model predicted bull terrier ancestry contributes to smaller size.

Table S9. Breed propensities inferred only from the effects of breed ancestry in mutts.

Index	Class	Trait	Breed	Phenotype Direction	N	REML t	ML anova *p*-value (FDR adjusted)
110	behavior question	Q110: More good days	akita	disagree	24	4.77	6.88E-05
3	behavior factor	F3: Toy-directed Motor Patterns	shar pei	not toy-directed	93	4.57	1.61E-04
40	behavior question	Q40: Damages doors	chesapeake bay retriever	always	17	4.15	0.0011
8	behavior factor	F8: Proximity Seeking	saint bernard	affectionate	54	-4.11	0.0016
52	behavior question	Q52: Works at tasks	shar pei	disagree	28	3.89	0.0039
40	behavior question	Q40: Damages doors	shar pei	always	27	3.66	0.0059
2	behavior question	Q2: Wants to play	shar pei	disagree	28	3.78	0.0064
125	physical trait	Q125: Ear shape	bloodhound	dropped	62	-3.04	0.0100
17	behavior question	Q17: Howls	bloodhound	always	63	3.60	0.0109
121	physical trait	Q121: Size	mastiff	hip high	10	2.82	0.0160
31	behavior question	Q31: Ignores commands	shar pei	agree	27	-3.23	0.0204
35	behavior question	Q35: Escapes from enclosures	chesapeake bay retriever	always	17	3.46	0.0233
26	behavior question	Q26: As active as has been	english cocker spaniel	disagree	25	3.20	0.0319
127	physical trait	Q127: Fur length	old english sheepdog	long	29	2.90	0.0374
127	physical trait	Q127: Fur length	samoyed	long	28	2.72	0.0462
127	physical trait	Q127: Fur length	irish setter	long	18	2.64	0.0462

Table S10. Replication of physical trait associations found in literature for dogs and other animals.

Gene	Species	Published Trait	Published Associaton	Published *p*-value	Published effect size	DOI / Reference	GWAS Trait	GWAS Peak	GWAS *p*-value	GWAS effect size	Replicated canine locus?
FGF4 retrogene insertion to	*Canis lupus familiaris*	chondro-dysplasia	18:2031083-20742377	1.90E-111	N/A	10.1126/ science.1173275 (*67*)	Q121 Size	18:20423056:A:G	1.20E-26	-0.39	TRUE
FGF4 retrogene insertion to	Canis lupus familiaris	chondro-dysplasia	12:32413668-43115518	<5.00E-02	N/A	10.1073/ pnas.1709082114 (*70*)	Q121 Size	12:33887652:G:A	2.18E-08	-0.14	TRUE
HMGA2	*Canis lupus familiaris*	stature	10:8454499	7.06E–09	N/A	10.1101/gr.157339.113 (*69*)	Q121 Size	10:8356059:G:T	1.80E-24	-0.31	TRUE
	Homo sapiens	stature				10.1038/ng2121 (*138*)					
SMAD2	*Canis lupus familiaris*	stature	7:43865905	1.05E–06	N/A		Q121 Size	7:43868045:C:T	3.82E-09	-0.14	TRUE
GHR	*Canis lupus familiaris*	stature	4:39200720, 4:67026055	9.10E–09, 2.58E–07	N/A	10.1101/gr.157339.113 (*69*)	Q121 Size	4:67040898:C:T	2.80E-10	-0.15	TRUE
LCORL	*Canis lupus familiaris*	stature	3:91269525	4.04E-23	N/A	10.1038/ s41467-019-09373-w (*40*)	Q121 Size	3:91108467:T:C	2.00E-18	0.29	TRUE
	Homo sapiens	stature				10.1371/ journal.pgen.10004 09 (*139*)					
	Equus caballus	stature				10.1371/ journal.pone.00564 97 (*140*)					
		stature				10.1111/age.12031 10.1111/age.12031 (*141*)					
	Bos taurus	stature				10.1534/ genetics.110.12394 3 (*142*)					
		stature				10.1038/ s41588-018-0056-5 (*143*)					
IGF1	*Canis lupus familiaris*	stature	15:40446920-40447659	5.90 E–05	N/A	10.1073/ pnas.152333099 (*64*)	Q121 Size	15:41274602:G:T	2.00E-17	0.18	TRUE
		stature	15:34903589-34903860	2.80E-02	N/A	10.1101/gr.3712705 (*65*)					
		stature	15:41221438	2.00E-16	N/A	10.1126/ science.1137045 (*66*)					
IGF2BP2	*Canis lupus familiaris*	stature	34:20097018-212633271	1.26E-11	N/A	10.1038/ s41467-019-09373-w (*40*)	Q121 Size	34:18296868:C:T	4.10E-09	0.12	TRUE
SAR1B	*Homo sapiens*	metabolism				10.1186/1750-1172-6-7 8 (*72*)	Q121 Size	11:22737397:G:A	1.30E-13	0.32	FALSE
	Mus musculus	metabolism				10.1186/1750-1172-6-7 8 (*72*)					
MED13L	*Canis lupus familiaris*	stature	26:12796099-13004170	2.08E-11	N/A	10.1038/ s41467-019-09373-w (*40*)	Q121 Size	26:12838979:C:A	2.90E-37	-0.78	TRUE
ANAPC1	*Homo sapiens*	dwarfism				10.1016/ j.ajhg.2019.06.011 (*74*)	Q121 Size	17:36295546:C:T	4.10E-08	0.15	FALSE

Gene	Species	Published Trait	Published Associaton	Published *p*-value	Published effect size	DOI / Reference	GWAS Trait	GWAS Peak	GWAS *p*-value	GWAS effect size	Replicated canine locus?
IGF1R	*Canis lupus familiaris*	stature	3:41849479	3.6E–25, 1.4E–38	N/A	10.1007/s00335-012-9417-z (*68*)	Q121 Size (Tiny)	3:42107672:A:G	1.30E-13	0.07	TRUE
	Homo sapiens	stature				10.1159/000437324 (*144*)					
	Mus musculus	metabolism				10.1172/JCI42447 (*145*)					
LRIG3	*Caenorhabditis elegans*	BMP-mediated body size				10.1371/journal.pgen.10009 63 (*75*)	Q122 Merle coat	10:2677840:A:G	3.10E-08	-0.16	FALSE
ITGA11	*Ovis aries*	birth weight				10.3389/fgene.2020.00588 (*146*)	Q121 Size (Giant)	30:32302089:C:T	7.80E-13	0.09	FALSE
MITF	*Canis lupus familiaris*	white spotting	20:21836232-21836429	N/A	N/A	10.1093/jhered/esp029 (*147*)	Q123 White fur	20:21827323:C:T	2.90E-37	-0.78	TRUE
		white spotting	20:21839331-21839366	N/A	N/A	10.1038/ng.2007.10 (*14*)					
		coat color	20:21786368-21869849	5.99E-29	N/A	10.1038/s41467-019-09373-w (*40*)					
		white spotting		N/A	N/A	10.1371/journal.pone.01043 63 (*148*)					
	Bos taurus	white spotting				10.1111/age.12751 (*149*)					
	Galo galo	pigmentation				10.1080/00071668.2017.1379053 (*150*)					
MSRB3	*Canis lupus familiaris*	ear shape	10:8041973, 10:8053461, 10:8068217, 10:8087164, 10:8090498	< 1.00E-45	0.6464	10.1186/s12864-015-1702-2 (*71*)	Q125 Ear shape	10:8027948:C:T	1.40E-27	0.10	TRUE
	Ovis aries	ear shape				10.1111/age.12994 (*151*)					
	Sus scrofa domesticus	ear shape				10.1186/s12711-018-0442-6 (*152*)					

Gene	Species	Published Trait	Published Associaton	Published *p*-value	Published effect size	DOI / Reference	GWAS Trait	GWAS Peak	GWAS *p*-value	GWAS effect size	Replicated canine locus?
LEMD3	*Sus scrofa domesticus*	ear shape				10.1371/journal.pone.01020 85 (*153*)	Q125 Ear shape	10:8027948:C:T	6.00E-23	-0.33	FALSE
		ear shape				10.1016/S2095-3119(15)61173-X (*154*)					
FGF5	*Canis lupus familiaris*	coat length	32:4509367	3.08E-66	N/A	10.1111/j.1365-2052.2006.01448.x (*155*)	Q127 Fur length	32:4509367:G:T	5.50E-54	0.37	TRUE
		coat length	32:4528617-4528633, 32:4528621-4528621, 32:4528639-4528639	N/A	N/A	10.1111/age.12010 (*156*)					
		coat length	32:7473337	1.00E-157	N/A	10.1126/science.1177808 (*58*)					
	Felis catus	coat length				10.1111/j.1365-2052.2007.01590.x (*157*)					
RSPO2	*Canis lupus familiaris*	wire coat/furnishings	13:8568727-8694401	4.00E-292	N/A	10.1126/science.1177808 (*58*)	Q128 Fur texture	13:8491477:A:C	3.40E-13	0.27	TRUE
		"Improper coat" in PWD	13:8610419	N/A	N/A	10.1093/jhered/esq068 (*158*)					
USH2A	*Canis lupus familiaris*	ticking	38:11122467 and 38:11124294	1.70E-05	N/A	10.1111/age.13040 (*59*)	Q122 Ticking (piebald-only)	38:11165134:G:A	5.31E-16	0.20	TRUE
		ticking	38:11085443	3.6E−08	N/A	10.1371/journal.pone.02482 33 (*159*)					
CBD103	*Canis lupus familiaris*	brindle	16:43818953-57246829	N/A	N/A	10.1534/genetics.107.07423 7 (*63*)	Q122 Brindle coat	16:59013740:G:A	#######	0.43	TRUE
		brindle	16:58965448-58965450	< 1.00E-06	N/A	10.1126/science.1147880 (*61*)					
		brindle	16:58965448-58965450	N/A	N/A	10.1292/jvms.10-0439 (*62*)					
SILV/PMEL retrogene insertion to CFA10	*Canis lupus familiaris*	merle	10:292851			10.1073/pnas.0506940103 (*160*)	Q122 Merle coat	10:371299:T:G	6.40E-09	0.12	TRUE
		merle				10.1159/000491408 (*161*)					
		merle				10.1556/004.2019.018 (*162*)					

Gene	Species	Published Trait	Published Associaton	Published *p* -value	Published effect size	DOI / Reference	GWAS Trait	GWAS Peak	GWAS *p* -value	GWAS effect size	Replicated canine locus?
FOXP1	*Homo sapiens*	vitiligo				10.1038/ng.602 (*163*)	Q122 Ticking (piebald-only)	20:20888915:A:G	3.10E-09	0.23	FALSE
	Mus musculus	hair loss				10.1242/dev.097477 (*164*)					
RUNX3	*Canis lupus familiaris*	red intensity	2:74746906	1.48E-154	0.95	10.1371/ journal.pone.02505 79 (*60*)	Q122 Red intensity	2:74851797:T:C	2.00E-08	0.12	TRUE
	Homo sapiens	gray hairs				10.3892/etm.2019.7663 (*165*)					
TYRP1	*Canis lupus familiaris*	liver	0.4 Mb from 11:34516748-34517020	N/A	N/A	10.1186/1746-6148-1-1 (*166*)	Q122 Liver coat	11:33326685:C:T	5.30E-16	0.20	TRUE
		liver	11:33326719-33326719	1.08E−19	N/A	10.1111/age.12839 (*167*)					
		liver	11:33317810	N/A	N/A	10.1111/age.12337 (*168*)					
	Equus caballus	liver				10.1007/ s003350020017 (*169*)					

Table S11. Correlation of age-related questions in Darwin's Ark and the Canine Cognitive Dysfunction Rating scale.

Pearson correlation of PPS with year of age (8 yr +) **Darwin's Ark:**				**Spearman correlation of "% dogs that frequently perform behavior" with age** **Salvin et al 2011 (*35*):**						
Question	**# years**	**r**	**p**	**Question from CCDR**	**8-10 years**	**10-12 years**	**>12 years**	**Correlation**	**Direction of score**	**Match?**
#2 DOG wants to play (agree-disagree)	7	0.96	7.50E-04	Time spent playing	96%	95%	88%	-1	opposite	TRUE
#25 DOG moves normally (agree-disagree)	7	0.99	4.00E-05	Abnormal locomotion	13%	21%	39%	1	same	TRUE
#36 DOG gets stuck behind objects and is unable to get around (never-always)	7	0.77	0.044	Gets stuck behind objects	10%	11%	22%	1	same	TRUE
#39 DOG lays in one place all day long (never-always)	7	0.97	4.20E-04	Time spent active	52%	44%	29%	-1	opposite	TRUE
#61 DOG has difficulty finding food dropped on the floor (never-always)	7	0.97	3.20E-04	Difficulty finding dropped food	21%	26%	51%	1	same	TRUE
#95 DOG stares blankly at the walls or floor (never-always)	6	0.77	0.072	Stares blankly	10%	11%	22%	1	same	TRUE

Table S12. Correlation of age-related questions in Darwin’s Ark and Canine Health-related and Quality of Life Survey.

Darwin's Ark: Pearson correlation of PPS with year of age				**Lavan et al 2013 (*36*)**
Question	**N**	***p***	**r**	***p****
Q1 DOG enjoys life (agree-disagree)	15	4.60E-07	0.93	0.063
Q2 DOG wants to play (agree-disagree)	15	2.51E-09	0.97	0.0001
Q25 DOG moves normally (agree-disagree)	15	1.35E-07	0.94	0.0001
Q26 DOG is as active as HE has been (agree-disagree)	15	4.74E-09	0.97	0.0001
Q33 DOG sleeps more, is less awake (agree-disagree)	15	4.20E-10	-0.98	0.0001
Q39 DOG lays in one place all day long (never-always)	15	6.46E-07	0.93	0.024
Q43 DOG responds to my presence (agree-disagree)	15	8.43E-04	0.77	0.007
Q94 DOG pants frequently, even at rest (agree-disagree)	15	3.71E-05	-0.86	0.016
Q97 DOG shakes or trembles occasionally (agree-disagree)	14	2.69E-07	-0.95	0.135
Q103 DOG seems dull or depressed, not alert (agree-disagree)	14	3.50E-09	-0.97	0.002
Q110 DOG has more good days than bad days (agree-disagree)	14	4.42E-05	0.87	0.033
			Spearman's rank correlation rho=	0.633
			Spearman's rank correlation *p*=	0.037

* p from Kruskal–Wallis one-way ANOVA; this is the only statistical metric reported in Lavan et al 2013

Table S13. Set of SNPs with MAF>0.025 in each population randomly sampled.

Population	All	Illumina array SNPs	Axiom array SNPs	Low-coverage sequencing SNPs
wolves	18,216,839	133,193	678,689	7,944,474
village dogs	12,710,126	152,094	758,076	8,719,223
mutts	12,491,157	160,666	779,353	9,076,003
yorkshire terriers	8,983,416	141,619	646,509	7,140,504
labrador retrievers	8,407,324	137,550	619,205	6,806,734
golden retrievers	8,348,982	136,917	615,892	6,760,319
leonbergers	7,135,029	120,717	523,873	5,689,553

Table S14. Population peculiarity score datasets. Variable numbers of dogs available for random sampling has a small effect on the magnitude of PPS z scores in the candidate breed analysis.

					Effect of # dogs available for sampling on abs(z) scores (ANOVA)			
Set type	# tests	median # sets / trait (range)	# dogs per sample	# dogs available for sampling per set (range)	ges	DFd	F	p
year of age	1,375	11	100	639 - 1814	1.00E-03	1,249	1.25	0.26
candidate purebred	7,478	60 (51 - 62)	25	25 - 905	0.028	7,356	208	1.50E-46
	4,708	37 (28 - 44)	25	50 - 905	0.013	4,587	59.2	1.70E-14
confirmed purebred	604	5 (3 - 6)	50	100 - 316	0.01	498	4.8	0.028

ANOVA model: anova_test(abs(z) ~ ndogs_available+trait_id)

Table S15. Expected vs. observed rates of correct guesses in Mutt Mix.

			User guesses 2 breeds				**User guesses 3 breeds**					
			1+ correct		**2 correct**		**1+ correct**		**2+ correct**		**3 correct**	
ID	**Name**	**# guesses**	**exp**	**obs**	**exp**	**obs**	**exp**	**obs**	**exp**	**obs**	**exp**	**obs**
1009	Bailey	14,160	0.178	0.959	0.00541	0.16695	0.26	0.957	0.01606	0.31511	1.60E-04	1.30E-02
1010	Bella	14,117	0.169	0.881	0.00505	0.02006	0.247	0.892	0.01505	0.24285	1.50E-04	4.00E-03
1011	Beskow	13,696	0.281	0.194	0.00949	0.00494	0.395	0.316	0.02711	0.01761	9.20E-05	0.00E+00
1012	Boone	13,975	0.167	0.872	0.00429	0.23839	0.244	0.922	0.01281	0.44669	7.50E-05	3.80E-03
1013	Buddy	14,060	0.197	0.975	0.00681	0.02922	0.286	0.964	0.0201	0.06796	2.20E-04	4.80E-04
1014	Clarence	13,969	0.362	0.726	0.02428	0.18354	0.496	0.865	0.06671	0.41076	1.20E-03	6.00E-02
1015	Cooper	13,903	0.232	0.949	0.00864	0.37128	0.333	0.941	0.02519	0.53381	2.10E-04	9.40E-02
1016	Dug	13,662	0.351	0.774	0.02281	0.03458	0.483	0.79	0.06298	0.07397	1.30E-03	3.60E-04
1017	Esme	13,737	0.329	0.729	0.01835	0.12847	0.455	0.798	0.05122	0.24978	9.00E-04	2.50E-03
1019	Hershey	13,944	0.231	0.805	0.00968	0.20193	0.332	0.851	0.02821	0.37112	3.90E-04	4.30E-02
1020	Hopper	14,044	0.254	0.604	0.01171	0.00517	0.362	0.668	0.03381	0.01203	5.10E-04	0.00E+00
1021	Jack1	14,041	0.1	0.211	0.00165	0	0.149	0.351	0.00503	0.00795	2.70E-05	0.00E+00
1022	Jack2	13,781	0.319	0.213	0.01649	0.00284	0.443	0.358	0.04622	0.02047	7.30E-04	8.30E-05
1023	Kaylee	14,102	0.223	0.963	0.00923	0.27557	0.32	0.958	0.02696	0.3417	3.80E-04	9.90E-04
1024	Lilly	14,036	0.227	0.958	0.00963	0.57413	0.326	0.955	0.02811	0.62244	4.10E-04	3.20E-02
1025	Lola	14,006	0.054	0.772	0.00047	0.14896	0.082	0.852	0.00146	0.32036	3.90E-06	1.70E-02
1026	Lucy	13,984	0.148	0.028	0.00358	0	0.218	0.112	0.01074	0.00167	8.00E-05	0.00E+00
1027	Luna	13,931	0.192	0.595	0.00644	0.00315	0.279	0.67	0.01903	0.02794	2.00E-04	0.00E+00
1028	Maxine	13,941	0.285	0.206	0.01572	0.0288	0.402	0.309	0.04481	0.06574	8.40E-04	3.50E-03
1030	Peso	14,024	0.221	0.987	0.00757	0.0691	0.317	0.966	0.02217	0.13451	1.80E-04	3.00E-03
1031	Ramy	13,848	0.348	0.069	0.02199	0.00292	0.479	0.141	0.06082	0.00636	1.20E-03	1.70E-04
1033	Rex	13,870	0.273	0.766	0.01398	0.18317	0.386	0.806	0.04007	0.29694	6.80E-04	1.60E-02
1034	Rosie	14,113	0.379	0.323	0.02816	0.00114	0.516	0.573	0.07671	0.09173	1.80E-03	4.00E-04
1037	Sadie	13,990	0.206	0.728	0.00769	0.08899	0.298	0.864	0.02262	0.33852	2.80E-04	9.60E-03
1039	Scotch	13,846	0.176	0.044	0.00497	0.00051	0.256	0.114	0.01478	0.00312	1.10E-04	0.00E+00
1042	Zandy	14,058	0.192	0.94	0.00671	0.34548	0.28	0.936	0.01983	0.4682	2.30E-04	3.10E-02
1018	Gus	13,854	0.214	0.741	0.00628	0.01074	0.308	0.731	0.01848	0.04687		
1032	Reilly	13,990	0.246	0.102	0.00454	0	0.351	0.194	0.01315	0.00218		
1035	Ruby	14,067	0.279	0.372	0.00914	0.00147	0.394	0.477	0.02614	0.00416		
1036	Rudy	13,783	0.269	0.349	0.00767	0.01672	0.38	0.458	0.02204	0.06807		

References and Notes

1. K. Lindblad-Toh, C. M. Wade, T. S. Mikkelsen, E. K. Karlsson, D. B. Jaffe, M. Kamal, M. Clamp, J. L. Chang, E. J. Kulbokas III, M. C. Zody, E. Mauceli, X. Xie, M. Breen, R. K. Wayne, E. A. Ostrander, C. P. Ponting, F. Galibert, D. R. Smith, P. J. DeJong, E. Kirkness, P. Alvarez, T. Biagi, W. Brockman, J. Butler, C.-W. Chin, A. Cook, J. Cuff, M. J. Daly, D. DeCaprio, S. Gnerre, M. Grabherr, M. Kellis, M. Kleber, C. Bardeleben, L. Goodstadt, A. Heger, C. Hitte, L. Kim, K.-P. Koepfli, H. G. Parker, J. P. Pollinger, S. M. J. Searle, N. B. Sutter, R. Thomas, C. Webber, J. Baldwin, A. Abebe, A. Abouelleil, L. Aftuck, M. Ait-Zahra, T. Aldredge, N. Allen, P. An, S. Anderson, C. Antoine, H. Arachchi, A. Aslam, L. Ayotte, P. Bachantsang, A. Barry, T. Bayul, M. Benamara, A. Berlin, D. Bessette, B. Blitshteyn, T. Bloom, J. Blye, L. Boguslavskiy, C. Bonnet, B. Boukhgalter, A. Brown, P. Cahill, N. Calixte, J. Camarata, Y. Cheshatsang, J. Chu, M. Citroen, A. Collymore, P. Cooke, T. Dawoe, R. Daza, K. Decktor, S. DeGray, N. Dhargay, K. Dooley, K. Dooley, P. Dorje, K. Dorjee, L. Dorris, N. Duffey, A. Dupes, O. Egbiremolen, R. Elong, J. Falk, A. Farina, S. Faro, D. Ferguson, P. Ferreira, S. Fisher, M. FitzGerald, K. Foley, C. Foley, A. Franke, D. Friedrich, D. Gage, M. Garber, G. Gearin, G. Giannoukos, T. Goode, A. Goyette, J. Graham, E. Grandbois, K. Gyaltsen, N. Hafez, D. Hagopian, B. Hagos, J. Hall, C. Healy, R. Hegarty, T. Honan, A. Horn, N. Houde, L. Hughes, L. Hunnicutt, M. Husby, B. Jester, C. Jones, A. Kamat, B. Kanga, C. Kells, D. Khazanovich, A. C. Kieu, P. Kisner, M. Kumar, K. Lance, T. Landers, M. Lara, W. Lee, J.-P. Leger, N. Lennon, L. Leuper, S. LeVine, J. Liu, X. Liu, Y. Lokyitsang, T. Lokyitsang, A. Lui, J. Macdonald, J. Major, R. Marabella, K. Maru, C. Matthews, S. McDonough, T. Mehta, J. Meldrim, A. Melnikov, L. Meneus, A. Mihalev, T. Mihova, K. Miller, R. Mittelman, V. Mlenga, L. Mulrain, G. Munson, A. Navidi, J. Naylor, T. Nguyen, N. Nguyen, C. Nguyen, T. Nguyen, R. Nicol, N. Norbu, C. Norbu, N. Novod, T. Nyima, P. Olandt, B. O'Neill, K. O'Neill, S. Osman, L. Oyono, C. Patti, D. Perrin, P. Phunkhang, F. Pierre, M. Priest, A. Rachupka, S. Raghuraman, R. Rameau, V. Ray, C. Raymond, F. Rege, C. Rise, J. Rogers, P. Rogov, J. Sahalie, S. Settipalli, T. Sharpe, T. Shea, M. Sheehan, N. Sherpa, J. Shi, D. Shih, J. Sloan, C. Smith, T. Sparrow, J. Stalker, N. Stange-Thomann, S. Stavropoulos, C. Stone, S. Stone, S. Sykes, P. Tchuinga, P. Tenzing, S. Tesfaye, D. Thoulutsang, Y. Thoulutsang, K. Topham, I. Topping, T. Tsamla, H. Vassiliev, V. Venkataraman, A. Vo, T. Wangchuk, T. Wangdi, M. Weiand, J. Wilkinson, A. Wilson, S. Yadav, S. Yang, X. Yang, G. Young, Q. Yu, J. Zainoun, L. Zembek, A. Zimmer, E. S. Lander, Genome sequence, comparative analysis and haplotype structure of the domestic dog. *Nature* **438**, 803–819 (2005). doi:10.1038/nature04338 Medline

2. A. Bergström, L. Frantz, R. Schmidt, E. Ersmark, O. Lebrasseur, L. Girdland-Flink, A. T. Lin, J. Storå, K.-G. Sjögren, D. Anthony, E. Antipina, S. Amiri, G. Bar-Oz, V. I. Bazaliiskii, J. Bulatović, D. Brown, A. Carmagnini, T. Davy, S. Fedorov, I. Fiore, D. Fulton, M. Germonpré, J. Haile, E. K. Irving-Pease, A. Jamieson, L. Janssens, I. Kirillova, L. K. Horwitz, J. Kuzmanovic-Cvetković, Y. Kuzmin, R. J. Losey, D. L. Dizdar, M. Mashkour, M. Novak, V. Onar, D. Orton, M. Pasarić, M. Radivojević, D. Rajković, B. Roberts, H. Ryan, M. Sablin, F. Shidlovskiy, I. Stojanović, A. Tagliacozzo, K. Trantalidou, I. Ullén, A. Villaluenga, P. Wapnish, K. Dobney, A. Götherström, A. Linderholm, L. Dalén, R. Pinhasi, G. Larson, P. Skoglund, Origins and genetic legacy of prehistoric dogs. *Science* **370**, 557–564 (2020). doi:10.1126/science.aba9572 Medline

3. K. Lord, M. Feinstein, B. Smith, R. Coppinger, Variation in reproductive traits of members of the genus *Canis* with special attention to the domestic dog (*Canis familiaris*). *Behav. Processes* **92**, 131–142 (2013). doi:10.1016/j.beproc.2012.10.009 Medline

4. C. Hansen Wheat, W. van der Bijl, H. Temrin, Dogs, but not wolves, lose their sensitivity toward novelty with age. *Front. Psychol.* **10**, 2001 (2019). doi:10.3389/fpsyg.2019.02001 Medline

5. L. Moretti, M. Hentrup, K. Kotrschal, F. Range, The influence of relationships on neophobia and exploration in wolves and dogs. *Anim. Behav.* **107**, 159–173 (2015). doi:10.1016/j.anbehav.2015.06.008 Medline

6. M. Arendt, K. M. Cairns, J. W. O. Ballard, P. Savolainen, E. Axelsson, Diet adaptation in dog reflects spread of prehistoric agriculture. *Heredity* **117**, 301–306 (2016). doi:10.1038/hdy.2016.48 Medline

7. E. Axelsson, A. Ratnakumar, M.-L. Arendt, K. Maqbool, M. T. Webster, M. Perloski, O. Liberg, J. M. Arnemo, A. Hedhammar, K. Lindblad-Toh, The genomic signature of dog domestication reveals adaptation to a starch-rich diet. *Nature* **495**, 360–364 (2013). doi:10.1038/nature11837 Medline

8. J. de G. Mazzorin, A. Tagliacozzo, “Morphological and osteological changes in the dog from the Neolithic to the Roman period in Italy” in *Dogs Through Time: An Archaeological Perspective* (Archaeopress, 2000), pp. 141–161.

9. J. D. G. Mazzorin, A. Tagliacozzo, Dog remains in Italy from the Neolithic to the Roman period. *Anthropozoologica* **25**, 429 (1997).

10. M. Worboys, J.-M. Strange, N. Pemberton, *The Invention of the Modern Dog: Breed and Blood in Victorian Britain* (JHU Press, 2018).

11. H. Ritvo, Pride and pedigree: The evolution of the Victorian dog fancy. *Vic. Stud.* **29**, 227–253 (1986).

12. K. Lord, R. A. Schneider, R. Coppinger, in *The Domestic Dog: Its Evolution, Behavior and Interactions with People*, J. Serpell, Ed. (Cambridge Univ. Press, 2016), chap. 4, pp. 42–66.

13. C. D. Arons, W. J. Shoemaker, The distribution of catecholamines and β-endorphin in the brains of three behaviorally distinct breeds of dogs and their F1 hybrids. *Brain Res.* **594**, 31–39 (1992). doi:10.1016/0006-8993(92)91026-B Medline

14. E. K. Karlsson, I. Baranowska, C. M. Wade, N. H. C. Salmon Hillbertz, M. C. Zody, N. Anderson, T. M. Biagi, N. Patterson, G. R. Pielberg, E. J. Kulbokas III, K. E. Comstock, E. T. Keller, J. P. Mesirov, H. von Euler, O. Kämpe, A. Hedhammar, E. S. Lander, G. Andersson, L. Andersson, K. Lindblad-Toh, Efficient mapping of mendelian traits in dogs through genome-wide association. *Nat. Genet.* **39**, 1321–1328 (2007). doi:10.1038/ng.2007.10 Medline

15. E. K. Karlsson, K. Lindblad-Toh, Leader of the pack: Gene mapping in dogs and other model organisms. *Nat. Rev. Genet.* **9**, 713–725 (2008). doi:10.1038/nrg2382 Medline

16. N. B. Sutter, E. A. Ostrander, Dog star rising: The canine genetic system. *Nat. Rev. Genet.* **5**, 900–910 (2004). doi:10.1038/nrg1492 Medline

17. M. E. Gompper, Ed., *Free-Ranging Dogs and Wildlife Conservation* (Oxford Univ. Press, 2013).

18. American Veterinary Medical Association (AVMA), *United States Pet Ownership and Demographics Sourcebook* (AVMA, Veterinary Economics Division, 2018).

19. D. C. Coile, *Encyclopedia of Dog Breeds* (Barrons Educational Series, 1998).

20. American Kennel Club, *The New Complete Dog Book: Official Breed Standards and Profiles for Over 200 Breeds* (Fox Chapel Publishing, 2017).

21. J. P. Scott, J. L. Fuller, *Genetics and the Social Behavior of the Dog* (Univ. Chicago Press, 1965).

22. See the supplementary materials.

23. L. R. Mehrkam, C. D. L. Wynne, Behavioral differences among breeds of domestic dogs (*Canis lupus familiaris*): Current status of the science. *Appl. Anim. Behav. Sci.* **155**, 12–27 (2014). doi:10.1016/j.applanim.2014.03.005

24. K. Svartberg, Breed-typical behaviour in dogs—Historical remnants or recent constructs? *Appl. Anim. Behav. Sci.* **96**, 293–313 (2006). doi:10.1016/j.applanim.2005.06.014

25. H. J. Noh, R. Tang, J. Flannick, C. O'Dushlaine, R. Swofford, D. Howrigan, D. P. Genereux, J. Johnson, G. van Grootheest, E. Grünblatt, E. Andersson, D. R. Djurfeldt, P. D. Patel, M. Koltookian, C. M Hultman, M. T. Pato, C. N. Pato, S. A. Rasmussen, M. A. Jenike, G. L. Hanna, S. E. Stewart, J. A. Knowles, S. Ruhrmann, H.-J. Grabe, M. Wagner, C. Rück, C. A. Mathews, S. Walitza, D. C. Cath, G. Feng, E. K. Karlsson, K. Lindblad-Toh, Integrating evolutionary and regulatory information with a multispecies approach implicates genes and pathways in obsessive-compulsive disorder. *Nat. Commun.* **8**, 774 (2017). doi:10.1038/s41467-017-00831-x Medline

26. N. H. Dodman, E. K. Karlsson, A. Moon-Fanelli, M. Galdzicka, M. Perloski, L. Shuster, K. Lindblad-Toh, E. I. Ginns, A canine chromosome 7 locus confers compulsive disorder susceptibility. *Mol. Psychiatry* **15**, 8–10 (2010). doi:10.1038/mp.2009.111 Medline

27. K. L. Overall, Natural animal models of human psychiatric conditions: Assessment of mechanism and validity. *Prog. Neuropsychopharmacol. Biol. Psychiatry* **24**, 727–776 (2000). doi:10.1016/S0278-5846(00)00104-4 Medline

28. R. Sarviaho, O. Hakosalo, K. Tiira, S. Sulkama, J. E. Niskanen, M. K. Hytönen, M. J. Sillanpää, H. Lohi, A novel genomic region on chromosome 11 associated with fearfulness in dogs. *Transl. Psychiatry* **10**, 169 (2020). doi:10.1038/s41398-020-0849-z Medline

29. R. Sarviaho, O. Hakosalo, K. Tiira, S. Sulkama, E. Salmela, M. K. Hytönen, M. J. Sillanpää, H. Lohi, Two novel genomic regions associated with fearfulness in dogs overlap human neuropsychiatric loci. *Transl. Psychiatry* **9**, 18 (2019). doi:10.1038/s41398-018-0361-x Medline

30. J. Ilska, M. J. Haskell, S. C. Blott, E. Sánchez-Molano, Z. Polgar, S. E. Lofgren, D. N. Clements, P. Wiener, Genetic characterization of dog personality traits. *Genetics* **206**, 1101–1111 (2017). doi:10.1534/genetics.116.192674 Medline

31. R. Tang, H. J. Noh, D. Wang, S. Sigurdsson, R. Swofford, M. Perloski, M. Duxbury, E. E. Patterson, J. Albright, M. Castelhano, A. Auton, A. R. Boyko, G. Feng, K. Lindblad-Toh, E. K. Karlsson, Candidate genes and functional noncoding variants identified in a canine model of obsessive-compulsive disorder. *Genome Biol.* **15**, R25 (2014). doi:10.1186/gb-2014-15-3-r25 Medline

32. C. DeBoever, Y. Tanigawa, M. Aguirre, G. McInnes, A. Lavertu, M. A. Rivas, Assessing digital phenotyping to enhance genetic studies of human diseases. *Am. J. Hum. Genet.* **106**, 611–622 (2020). doi:10.1016/j.ajhg.2020.03.007 Medline

33. C. L. Hyde, M. W. Nagle, C. Tian, X. Chen, S. A. Paciga, J. R. Wendland, J. Y. Tung, D. A. Hinds, R. H. Perlis, A. R. Winslow, Identification of 15 genetic loci associated with risk of major depression in individuals of European descent. *Nat. Genet.* **48**, 1031–1036 (2016). doi:10.1038/ng.3623 Medline

34. H. F. Wright, D. S. Mills, P. Pollux, Development and validation of a psychometric tool for assessing impulsivity in the domestic dog, *Canis familiaris*. *Int. J. Comp. Psychol.* (2011). doi:10.1016/j.physbeh.2011.09.019 Medline

35. H. E. Salvin, P. D. McGreevy, P. S. Sachdev, M. J. Valenzuela, The canine cognitive dysfunction rating scale (CCDR): A data-driven and ecologically relevant assessment tool. *Vet. J.* **188**, 331–336 (2011). doi:10.1016/j.tvjl.2010.05.014 Medline

36. R. P. Lavan, Development and validation of a survey for quality of life assessment by owners of healthy dogs. *Vet. J.* **197**, 578–582 (2013). doi:10.1016/j.tvjl.2013.03.021 Medline

37. A. C. Jones, *Development and Validation of a Dog Personality Questionnaire* (Univ. Texas, 2008).

38. N. B. Sutter, D. S. Mosher, M. M. Gray, E. A. Ostrander, Morphometrics within dog breeds are highly reproducible and dispute Rensch's rule. *Mamm. Genome* **19**, 713–723 (2008). doi:10.1007/s00335-008-9153-6 Medline

39. J. Yordy, C. Kraus, J. J. Hayward, M. E. White, L. M. Shannon, K. E. Creevy, D. E. L. Promislow, A. R. Boyko, Body size, inbreeding, and lifespan in domestic dogs. *Conserv. Genet.* **21**, 137–148 (2020). doi:10.1007/s10592-019-01240-x Medline

40. J. Plassais, J. Kim, B. W. Davis, D. M. Karyadi, A. N. Hogan, A. C. Harris, B. Decker, H. G. Parker, E. A. Ostrander, Whole genome sequencing of canids reveals genomic regions under selection and variants influencing morphology. *Nat. Commun.* **10**, 1489 (2019). doi:10.1038/s41467-019-09373-w Medline

41. H. G. Parker, D. L. Dreger, M. Rimbault, B. W. Davis, A. B. Mullen, G. Carpintero-Ramirez, E. A. Ostrander, Genomic analyses reveal the influence of geographic origin, migration, and hybridization on modern dog breed development. *Cell Rep.* **19**, 697–708 (2017). doi:10.1016/j.celrep.2017.03.079 Medline

42. A. R. Martin, E. G. Atkinson, S. B. Chapman, A. Stevenson, R. E. Stroud, T. Abebe, D. Akena, M. Alemayehu, F. K. Ashaba, L. Atwoli, T. Bowers, L. B. Chibnik, M. J. Daly, T. DeSmet, S. Dodge, A. Fekadu, S. Ferriera, B. Gelaye, S. Gichuru, W. E. Injera, R. James, S. M. Kariuki, G. Kigen, K. C. Koenen, E. Kwobah, J. Kyebuzibwa, L. Majara, H. Musinguzi, R. M. Mwema, B. M. Neale, C. P. Newman, C. R. J. C. Newton, J. K. Pickrell, R. Ramesar, W. Shiferaw, D. J. Stein, S. Teferra, C. van der Merwe, Z. Zingela, NeuroGAP-Psychosis Study Team, Low-coverage sequencing cost-effectively detects

known and novel variation in underrepresented populations. *Am. J. Hum. Genet.* **108**, 656–668 (2021). doi:10.1016/j.ajhg.2021.03.012 Medline

43. S. Rubinacci, D. M. Ribeiro, R. J. Hofmeister, O. Delaneau, Efficient phasing and imputation of low-coverage sequencing data using large reference panels. *Nat. Genet.* **53**, 120–126 (2021). doi:10.1038/s41588-020-00756-0 Medline

44. R. M. Buckley, A. C. Harris, G.-D. Wang, D. Thad Whitaker, Y.-P. Zhang, E. A. Ostrander, Best practices for analyzing imputed genotypes from low-pass sequencing in dogs. *Mamm. Genome.* **33**, 213–229 (2022). doi:10.1007/s00335-021-09914-z Medline

45. J. H. Li, C. A. Mazur, T. Berisa, J. K. Pickrell, Low-pass sequencing increases the power of GWAS and decreases measurement error of polygenic risk scores compared to genotyping arrays. *Genome Res.* **31**, 529–537 (2021). doi:10.1101/gr.266486.120 Medline

46. K. Wasik, T. Berisa, J. K. Pickrell, J. H. Li, D. J. Fraser, K. King, C. Cox, Comparing low-pass sequencing and genotyping for trait mapping in pharmacogenetics. *BMC Genomics* **22**, 197 (2021). doi:10.1186/s12864-021-07508-2 Medline

47. D. H. Alexander, K. Lange, Enhancements to the ADMIXTURE algorithm for individual ancestry estimation. *BMC Bioinformatics* **12**, 246 (2011). doi:10.1186/1471-2105-12-246 Medline

48. J. Yang, B. Benyamin, B. P. McEvoy, S. Gordon, A. K. Henders, D. R. Nyholt, P. A. Madden, A. C. Heath, N. G. Martin, G. W. Montgomery, M. E. Goddard, P. M. Visscher, Common SNPs explain a large proportion of the heritability for human height. *Nat. Genet.* **42**, 565–569 (2010). doi:10.1038/ng.608 Medline

49. G. Athanasiadis, D. Speed, M. K. Andersen, E. V. R. Appel, N. Grarup, I. Brandslund, M. E. Jørgensen, C. V. L. Larsen, P. Bjerregaard, T. Hansen, A. Albrechtsen, Estimating narrow-sense heritability using family data from admixed populations. *Heredity* **124**, 751–762 (2020). doi:10.1038/s41437-020-0311-2 Medline

50. S. Luo, E. V. Feofanova, A. Tin, S. Tung, E. P. Rhee, J. Coresh, D. E. Arking, A. Surapaneni, P. Schlosser, Y. Li, A. Köttgen, B. Yu, M. E. Grams, Genome-wide association study of serum metabolites in the African American Study of Kidney Disease and Hypertension. *Kidney Int.* **100**, 430–439 (2021). doi:10.1016/j.kint.2021.03.026 Medline

51. R. Bakeman, Recommended effect size statistics for repeated measures designs. *Behav. Res. Methods* **37**, 379–384 (2005). doi:10.3758/BF03192707 Medline

52. American Kennel Club, Dog breeds; https://www.akc.org/dog-breeds/.

53. American Kennel Club, *The Complete Dog Book* (Ballantine Books, 2006).

54. D. L. Wells, D. J. Morrison, P. G. Hepper, The effect of priming on perceptions of dog breed traits. *Anthrozoos* **25**, 369–377 (2012). doi:10.2752/175303712X13403555186370

55. L. M. Gunter, R. T. Barber, C. D. L. Wynne, What’s in a name? Effect of breed perceptions & labeling on attractiveness, adoptions & length of stay for pit-bull-type dogs. *PLOS ONE* **11**, e0146857 (2016). doi:10.1371/journal.pone.0146857 Medline

56. J. Yang, S. H. Lee, M. E. Goddard, P. M. Visscher, GCTA: A tool for genome-wide complex trait analysis. *Am. J. Hum. Genet.* **88**, 76–82 (2011). doi:10.1016/j.ajhg.2010.11.011 Medline

57. M. Pirinen, P. Donnelly, C. C. A. Spencer, Efficient computation with a linear mixed model on large-scale data sets with applications to genetic studies. *Ann. Appl. Stat.* **7**, 369–390 (2013). doi:10.1214/12-AOAS586

58. E. Cadieu, M. W. Neff, P. Quignon, K. Walsh, K. Chase, H. G. Parker, B. M. Vonholdt, A. Rhue, A. Boyko, A. Byers, A. Wong, D. S. Mosher, A. G. Elkahloun, T. C. Spady, C. André, K. G. Lark, M. Cargill, C. D. Bustamante, R. K. Wayne, E. A. Ostrander, Coat variation in the domestic dog is governed by variants in three genes. *Science* **326**, 150–153 (2009). doi:10.1126/science.1177808 Medline

59. L. Brancalion, B. Haase, H. Mazrier, C. E. Willet, K. Lindblad-Toh, F. Lingaas, C. M. Wade, Roan, ticked and clear coat patterns in the canine are associated with three haplotypes near *usherin* on CFA38. *Anim. Genet.* **52**, 198–207 (2021). doi:10.1111/age.13040 Medline

60. A. J. Slavney, T. Kawakami, M. K. Jensen, T. C. Nelson, A. J. Sams, A. R. Boyko, Five genetic variants explain over 70% of hair coat pheomelanin intensity variation in purebred and mixed breed domestic dogs. *PLOS ONE* **16**, e0250579 (2021). doi:10.1371/journal.pone.0250579 Medline

61. S. I. Candille, C. B. Kaelin, B. M. Cattanach, B. Yu, D. A. Thompson, M. A. Nix, J. A. Kerns, S. M. Schmutz, G. L. Millhauser, G. S. Barsh, A β-defensin mutation causes black coat color in domestic dogs. *Science* **318**, 1418–1423 (2007). doi:10.1126/science.1147880 Medline

62. M. Oguro-Okano, M. Honda, K. Yamazaki, K. Okano, Mutations in the *melanocortin 1 receptor*, β-*defensin103* and *agouti signaling protein* genes, and their association with coat color phenotypes in Akita-inu dogs. *J. Vet. Med. Sci.* **73**, 853–858 (2011). doi:10.1292/jvms.10-0439 Medline

63. J. A. Kerns, E. J. Cargill, L. A. Clark, S. I. Candille, T. G. Berryere, M. Olivier, G. Lust, R. J. Todhunter, S. M. Schmutz, K. E. Murphy, G. S. Barsh, Linkage and segregation analysis of black and brindle coat color in domestic dogs. *Genetics* **176**, 1679–1689 (2007). doi:10.1534/genetics.107.074237 Medline

64. K. Chase, D. R. Carrier, F. R. Adler, T. Jarvik, E. A. Ostrander, T. D. Lorentzen, K. G. Lark, Genetic basis for systems of skeletal quantitative traits: Principal component analysis of the canid skeleton. *Proc. Natl. Acad. Sci. U.S.A.* **99**, 9930–9935 (2002). doi:10.1073/pnas.152333099 Medline

65. K. Chase, D. R. Carrier, F. R. Adler, E. A. Ostrander, K. G. Lark, Interaction between the X chromosome and an autosome regulates size sexual dimorphism in Portuguese water dogs. *Genome Res.* **15**, 1820–1824 (2005). doi:10.1101/gr.3712705 Medline

66. N. B. Sutter, C. D. Bustamante, K. Chase, M. M. Gray, K. Zhao, L. Zhu, B. Padhukasahasram, E. Karlins, S. Davis, P. G. Jones, P. Quignon, G. S. Johnson, H. G. Parker, N. Fretwell, D. S. Mosher, D. F. Lawler, E. Satyaraj, M. Nordborg, K. G. Lark, R. K. Wayne, E. A. Ostrander, A single *IGF1* allele is a major determinant of small size in dogs. *Science* **316**, 112–115 (2007). doi:10.1126/science.1137045 Medline

67. H. G. Parker, B. M. VonHoldt, P. Quignon, E. H. Margulies, S. Shao, D. S. Mosher, T. C. Spady, A. Elkahloun, M. Cargill, P. G. Jones, C. L. Maslen, G. M. Acland, N. B. Sutter, K. Kuroki, C. D. Bustamante, R. K. Wayne, E. A. Ostrander, An expressed *fgf4* retrogene is associated with breed-defining chondrodysplasia in domestic dogs. *Science* **325**, 995–998 (2009). doi:10.1126/science.1173275 Medline

68. B. C. Hoopes, M. Rimbault, D. Liebers, E. A. Ostrander, N. B. Sutter, The insulin-like growth factor 1 receptor (IGF1R) contributes to reduced size in dogs. *Mamm. Genome* **23**, 780–790 (2012). doi:10.1007/s00335-012-9417-z Medline

69. M. Rimbault, H. C. Beale, J. J. Schoenebeck, B. C. Hoopes, J. J. Allen, P. Kilroy-Glynn, R. K. Wayne, N. B. Sutter, E. A. Ostrander, Derived variants at six genes explain nearly half of size reduction in dog breeds. *Genome Res.* **23**, 1985–1995 (2013). doi:10.1101/gr.157339.113 Medline

70. E. A. Brown, P. J. Dickinson, T. Mansour, B. K. Sturges, M. Aguilar, A. E. Young, C. Korff, J. Lind, C. L. Ettinger, S. Varon, R. Pollard, C. T. Brown, T. Raudsepp, D. L. Bannasch, *FGF4* retrogene on CFA12 is responsible for chondrodystrophy and intervertebral disc disease in dogs. *Proc. Natl. Acad. Sci. U.S.A.* **114**, 11476–11481 (2017). doi:10.1073/pnas.1709082114 Medline

71. M. T. Webster, N. Kamgari, M. Perloski, M. P. Hoeppner, E. Axelsson, Å. Hedhammar, G. Pielberg, K. Lindblad-Toh, Linked genetic variants on chromosome 10 control ear morphology and body mass among dog breeds. *BMC Genomics* **16**, 474 (2015). doi:10.1186/s12864-015-1702-2 Medline

72. T. Okada, M. Miyashita, J. Fukuhara, M. Sugitani, T. Ueno, M.-E. Samson-Bouma, L. P. Aggerbeck, Anderson's disease/chylomicron retention disease in a Japanese patient with uniparental disomy 7 and a normal *SAR1B* gene protein coding sequence. *Orphanet J. Rare Dis.* **6**, 78 (2011). doi:10.1186/1750-1172-6-78 Medline

73. E. Levy, S. Spahis, C. Garofalo, V. Marcil, A. Montoudis, D. Sinnet, R. Sanchez, N. Peretti, J.-F. Beaulieu, A. Sane, Sar1b transgenic male mice are more susceptible to high-fat diet-induced obesity, insulin insensitivity and intestinal chylomicron overproduction. *J. Nutr. Biochem.* **25**, 540–548 (2014). doi:10.1016/j.jnutbio.2014.01.004 Medline

74. N. F. Ajeawung, T. T. M. Nguyen, L. Lu, T. J. Kucharski, J. Rousseau, S. Molidperee, J. Atienza, I. Gamache, W. Jin, S. E. Plon, B. H. Lee, J. G. Teodoro, L. L. Wang, P. M. Campeau, Mutations in *ANAPC1*, encoding a scaffold subunit of the anaphase-promoting complex, cause Rothmund-Thomson syndrome type 1. *Am. J. Hum. Genet.* **105**, 625–630 (2019). doi:10.1016/j.ajhg.2019.06.011 Medline

75. T. L. Gumienny, L. Macneil, C. M. Zimmerman, H. Wang, L. Chin, J. L. Wrana, R. W. Padgett, *Caenorhabditis elegans* SMA-10/LRIG is a conserved transmembrane protein that enhances bone morphogenetic protein signaling. *PLOS Genet.* **6**, e1000963 (2010). doi:10.1371/journal.pgen.1000963 Medline

76. Z. Chen, M. Boehnke, X. Wen, B. Mukherjee, Revisiting the genome-wide significance threshold for common variant GWAS. *G3* **11**, jkaa056 (2021). doi:10.1093/g3journal/jkaa056 Medline

77. J. J. Lee, R. Wedow, A. Okbay, E. Kong, O. Maghzian, M. Zacher, T. A. Nguyen-Viet, P. Bowers, J. Sidorenko, R. Karlsson Linnér, M. A. Fontana, T. Kundu, C. Lee, H. Li, R. Li,

R. Royer, P. N. Timshel, R. K. Walters, E. A. Willoughby, L. Yengo, M. Alver, Y. Bao, D. W. Clark, F. R. Day, N. A. Furlotte, P. K. Joshi, K. E. Kemper, A. Kleinman, C. Langenberg, R. Mägi, J. W. Trampush, S. S. Verma, Y. Wu, M. Lam, J. H. Zhao, Z. Zheng, J. D. Boardman, H. Campbell, J. Freese, K. M. Harris, C. Hayward, P. Herd, M. Kumari, T. Lencz, J. Luan, A. K. Malhotra, A. Metspalu, L. Milani, K. K. Ong, J. R. B. Perry, D. J. Porteous, M. D. Ritchie, M. C. Smart, B. H. Smith, J. Y. Tung, N. J. Wareham, J. F. Wilson, J. P. Beauchamp, D. C. Conley, T. Esko, S. F. Lehrer, P. K. E. Magnusson, S. Oskarsson, T. H. Pers, M. R. Robinson, K. Thom, C. Watson, C. F. Chabris, M. N. Meyer, D. I. Laibson, J. Yang, M. Johannesson, P. D. Koellinger, P. Turley, P. M. Visscher, D. J. Benjamin, D. Cesarini, 23andMe Research Team, COGENT (Cognitive Genomics Consortium), Social Science Genetic Association Consortium, 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). doi:10.1038/s41588-018-0147-3 Medline

78. J. E. Savage, P. R. Jansen, S. Stringer, K. Watanabe, J. Bryois, C. A. de Leeuw, M. Nagel, S. Awasthi, P. B. Barr, J. R. I. Coleman, K. L. Grasby, A. R. Hammerschlag, J. A. Kaminski, R. Karlsson, E. Krapohl, M. Lam, M. Nygaard, C. A. Reynolds, J. W. Trampush, H. Young, D. Zabaneh, S. Hägg, N. K. Hansell, I. K. Karlsson, S. Linnarsson, G. W. Montgomery, A. B. Muñoz-Manchado, E. B. Quinlan, G. Schumann, N. G. Skene, B. T. Webb, T. White, D. E. Arking, D. Avramopoulos, R. M. Bilder, P. Bitsios, K. E. Burdick, T. D. Cannon, O. Chiba-Falek, A. Christoforou, E. T. Cirulli, E. Congdon, A. Corvin, G. Davies, I. J. Deary, P. DeRosse, D. Dickinson, S. Djurovic, G. Donohoe, E. D. Conley, J. G. Eriksson, T. Espeseth, N. A. Freimer, S. Giakoumaki, I. Giegling, M. Gill, D. C. Glahn, A. R. Hariri, A. Hatzimanolis, M. C. Keller, E. Knowles, D. Koltai, B. Konte, J. Lahti, S. Le Hellard, T. Lencz, D. C. Liewald, E. London, A. J. Lundervold, A. K. Malhotra, I. Melle, D. Morris, A. C. Need, W. Ollier, A. Palotie, A. Payton, N. Pendleton, R. A. Poldrack, K. Räikkönen, I. Reinvang, P. Roussos, D. Rujescu, F. W. Sabb, M. A. Scult, O. B. Smeland, N. Smyrnis, J. M. Starr, V. M. Steen, N. C. Stefanis, R. E. Straub, K. Sundet, H. Tiemeier, A. N. Voineskos, D. R. Weinberger, E. Widen, J. Yu, G. Abecasis, O. A. Andreassen, G. Breen, L. Christiansen, B. Debrabant, D. M. Dick, A. Heinz, J. Hjerling-Leffler, M. A. Ikram, K. S. Kendler, N. G. Martin, S. E. Medland, N. L. Pedersen, R. Plomin, T. J. C. Polderman, S. Ripke, S. van der Sluis, P. F. Sullivan, S. I. Vrieze, M. J. Wright, D. Posthuma, Genome-wide association meta-analysis in 269,867 individuals identifies new genetic and functional links to intelligence. *Nat. Genet.* **50**, 912–919 (2018). doi:10.1038/s41588-018-0152-6 Medline

79. M. Lam, W. D. Hill, J. W. Trampush, J. Yu, E. Knowles, G. Davies, E. Stahl, L. Huckins, D. C. Liewald, S. Djurovic, I. Melle, K. Sundet, A. Christoforou, I. Reinvang, P. DeRosse, A. J. Lundervold, V. M. Steen, T. Espeseth, K. Räikkönen, E. Widen, A. Palotie, J. G. Eriksson, I. Giegling, B. Konte, A. M. Hartmann, P. Roussos, S. Giakoumaki, K. E. Burdick, A. Payton, W. Ollier, O. Chiba-Falek, D. K. Attix, A. C. Need, E. T. Cirulli, A. N. Voineskos, N. C. Stefanis, D. Avramopoulos, A. Hatzimanolis, D. E. Arking, N. Smyrnis, R. M. Bilder, N. A. Freimer, T. D. Cannon, E. London, R. A. Poldrack, F. W. Sabb, E. Congdon, E. D. Conley, M. A. Scult, D. Dickinson, R. E. Straub, G. Donohoe, D. Morris, A. Corvin, M. Gill, A. R. Hariri, D. R. Weinberger, N. Pendleton, P. Bitsios, D. Rujescu, J. Lahti, S. Le Hellard, M. C. Keller, O. A. Andreassen, I. J. Deary, D. C. Glahn, A. K. Malhotra, T. Lencz, Pleiotropic meta-analysis of cognition, education, and schizophrenia differentiates roles of early neurodevelopmental and adult synaptic

pathways. *Am. J. Hum. Genet.* **105**, 334–350 (2019). doi:10.1016/j.ajhg.2019.06.012 Medline

80. R. S. Smith, C. J. Kenny, V. Ganesh, A. Jang, R. Borges-Monroy, J. N. Partlow, R. S. Hill, T. Shin, A. Y. Chen, R. N. Doan, A.-K. Anttonen, J. Ignatius, L. Medne, C. G. Bönnemann, J. L. Hecht, O. Salonen, A. J. Barkovich, A. Poduri, M. Wilke, M. C. Y. de Wit, G. M. S. Mancini, L. Sztriha, K. Im, D. Amrom, E. Andermann, R. Paetau, A.-E. Lehesjoki, C. A. Walsh, M. K. Lehtinen, Sodium channel SCN3A (Na$_V$1.3) regulation of human cerebral cortical folding and oral motor development. *Neuron* **99**, 905–913.e7 (2018). doi:10.1016/j.neuron.2018.07.052 Medline

81. Y. F. Widmer, A. Bilican, R. Bruggmann, S. G. Sprecher, Regulators of long-term memory revealed by mushroom body-specific gene expression profiling in *Drosophila melanogaster*. *Genetics* **209**, 1167–1181 (2018). doi:10.1534/genetics.118.301106 Medline

82. M. Pelé, L. Tiret, J.-L. Kessler, S. Blot, J.-J. Panthier, SINE exonic insertion in the *PTPLA* gene leads to multiple splicing defects and segregates with the autosomal recessive centronuclear myopathy in dogs. *Hum. Mol. Genet.* **14**, 1417–1427 (2005). doi:10.1093/hmg/ddi151 Medline

83. E. K. Karlsson, S. Sigurdsson, E. Ivansson, R. Thomas, I. Elvers, J. Wright, C. Howald, N. Tonomura, M. Perloski, R. Swofford, T. Biagi, S. Fryc, N. Anderson, C. Courtay-Cahen, L. Youell, S. L. Ricketts, S. Mandlebaum, P. Rivera, H. von Euler, W. C. Kisseberth, C. A. London, E. S. Lander, G. Couto, K. Comstock, M. P. Starkey, J. F. Modiano, M. Breen, K. Lindblad-Toh, Genome-wide analyses implicate 33 loci in heritable dog osteosarcoma, including regulatory variants near *CDKN2A/B*. *Genome Biol.* **14**, R132 (2013). doi:10.1186/gb-2013-14-12-r132 Medline

84. A. R. Boyko, P. Quignon, L. Li, J. J. Schoenebeck, J. D. Degenhardt, K. E. Lohmueller, K. Zhao, A. Brisbin, H. G. Parker, B. M. vonHoldt, M. Cargill, A. Auton, A. Reynolds, A. G. Elkahloun, M. Castelhano, D. S. Mosher, N. B. Sutter, G. S. Johnson, J. Novembre, M. J. Hubisz, A. Siepel, R. K. Wayne, C. D. Bustamante, E. A. Ostrander, A simple genetic architecture underlies morphological variation in dogs. *PLOS Biol.* **8**, e1000451 (2010). doi:10.1371/journal.pbio.1000451 Medline

85. J. Lonsdale, J. Thomas, M. Salvatore, R. Phillips, E. Lo, S. Shad, R. Hasz, G. Walters, F. Garcia, N. Young, B. Foster, M. Moser, E. Karasik, B. Gillard, K. Ramsey, S. Sullivan, J. Bridge, H. Magazine, J. Syron, J. Fleming, L. Siminoff, H. Traino, M. Mosavel, L. Barker, S. Jewell, D. Rohrer, D. Maxim, D. Filkins, P. Harbach, E. Cortadillo, B. Berghuis, L. Turner, E. Hudson, K. Feenstra, L. Sobin, J. Robb, P. Branton, G. Korzeniewski, C. Shive, D. Tabor, L. Qi, K. Groch, S. Nampally, S. Buia, A. Zimmerman, A. Smith, R. Burges, K. Robinson, K. Valentino, D. Bradbury, M. Cosentino, N. Diaz-Mayoral, M. Kennedy, T. Engel, P. Williams, K. Erickson, K. Ardlie, W. Winckler, G. Getz, D. DeLuca, D. MacArthur, M. Kellis, A. Thomson, T. Young, E. Gelfand, M. Donovan, Y. Meng, G. Grant, D. Mash, Y. Marcus, M. Basile, J. Liu, J. Zhu, Z. Tu, N. J. Cox, D. L. Nicolae, E. R. Gamazon, H. K. Im, A. Konkashbaev, J. Pritchard, M. Stevens, T. Flutre, X. Wen, E. T. Dermitzakis, T. Lappalainen, R. Guigo, J. Monlong, M. Sammeth, D. Koller, A. Battle, S. Mostafavi, M. McCarthy, M. Rivas, J. Maller, I. Rusyn, A. Nobel, F. Wright, A. Shabalin, M. Feolo, N. Sharopova, A. Sturcke, J. Paschal, J. M. Anderson, E. L. Wilder, L. K. Derr, E. D. Green, J. P. Struewing, G. Temple, S.

Volpi, J. T. Boyer, E. J. Thomson, M. S. Guyer, C. Ng, A. Abdallah, D. Colantuoni, T. R. Insel, S. E. Koester, A. R. Little, P. K. Bender, T. Lehner, Y. Yao, C. C. Compton, J. B. Vaught, S. Sawyer, N. C. Lockhart, J. Demchok, H. F. Moore, The Genotype-Tissue Expression (GTEx) project. *Nat. Genet.* **45**, 580–585 (2013). doi:10.1038/ng.2653 Medline

86. E. L. MacLean, N. Snyder-Mackler, B. M. vonHoldt, J. A. Serpell, Highly heritable and functionally relevant breed differences in dog behaviour. *Proc. Biol. Sci.* **286**, 20190716 (2019). doi:10.1098/rspb.2019.0716 Medline

87. A. P. Privitera, R. Distefano, H. A. Wefer, A. Ferro, A. Pulvirenti, R. Giugno, OCDB: A database collecting genes, miRNAs and drugs for obsessive-compulsive disorder. *Database* **2015**, bav069 (2015). doi:10.1093/database/bav069 Medline

88. B. S. Abrahams, D. E. Arking, D. B. Campbell, H. C. Mefford, E. M. Morrow, L. A. Weiss, I. Menashe, T. Wadkins, S. Banerjee-Basu, A. Packer, SFARI Gene 2.0: A community-driven knowledgebase for the autism spectrum disorders (ASDs). *Mol. Autism* **4**, 36 (2013). doi:10.1186/2040-2392-4-36 Medline

89. Schizophrenia Working Group of the Psychiatric Genomics Consortium, Biological insights from 108 schizophrenia-associated genetic loci. *Nature* **511**, 421–427 (2014). doi:10.1038/nature13595 Medline

90. A. F. Pardiñas, P. Holmans, A. J. Pocklington, V. Escott-Price, S. Ripke, N. Carrera, S. E. Legge, S. Bishop, D. Cameron, M. L. Hamshere, J. Han, L. Hubbard, A. Lynham, K. Mantripragada, E. Rees, J. H. MacCabe, S. A. McCarroll, B. T. Baune, G. Breen, E. M. Byrne, U. Dannlowski, T. C. Eley, C. Hayward, N. G. Martin, A. M. McIntosh, R. Plomin, D. J. Porteous, N. R. Wray, A. Caballero, D. H. Geschwind, L. M. Huckins, D. M. Ruderfer, E. Santiago, P. Sklar, E. A. Stahl, H. Won, E. Agerbo, T. D. Als, O. A. Andreassen, M. Bækvad-Hansen, P. B. Mortensen, C. B. Pedersen, A. D. Børglum, J. Bybjerg-Grauholm, S. Djurovic, N. Durmishi, M. G. Pedersen, V. Golimbet, J. Grove, D. M. Hougaard, M. Mattheisen, E. Molden, O. Mors, M. Nordentoft, M. Pejovic-Milovancevic, E. Sigurdsson, T. Silagadze, C. S. Hansen, K. Stefansson, H. Stefansson, S. Steinberg, S. Tosato, T. Werge, D. A. Collier, D. Rujescu, G. Kirov, M. J. Owen, M. C. O'Donovan, J. T. R. Walters, GERAD1 Consortium, CRESTAR Consortium, Common schizophrenia alleles are enriched in mutation-intolerant genes and in regions under strong background selection. *Nat. Genet.* **50**, 381–389 (2018). doi:10.1038/s41588-018-0059-2 Medline

91. C. A. de Leeuw, J. M. Mooij, T. Heskes, D. Posthuma, MAGMA: Generalized gene-set analysis of GWAS data. *PLOS Comput. Biol.* **11**, e1004219 (2015). doi:10.1371/journal.pcbi.1004219 Medline

92. X. Yi, Y. Liang, E. Huerta-Sanchez, X. Jin, Z. X. P. Cuo, J. E. Pool, X. Xu, H. Jiang, N. Vinckenbosch, T. S. Korneliussen, H. Zheng, T. Liu, W. He, K. Li, R. Luo, X. Nie, H. Wu, M. Zhao, H. Cao, J. Zou, Y. Shan, S. Li, Q. Yang, P. Asan, P. Ni, G. Tian, J. Xu, X. Liu, T. Jiang, R. Wu, G. Zhou, M. Tang, J. Qin, T. Wang, S. Feng, G. Li, J. Huasang, J. Luosang, W. Wang, F. Chen, Y. Wang, X. Zheng, Z. Li, Z. Bianba, G. Yang, X. Wang, S. Tang, G. Gao, Y. Chen, Z. Luo, L. Gusang, Z. Cao, Q. Zhang, W. Ouyang, X. Ren, H. Liang, H. Zheng, Y. Huang, J. Li, L. Bolund, K. Kristiansen, Y. Li, Y. Zhang, X. Zhang, R. Li, S. Li, H. Yang, R. Nielsen, J. Wang, J. Wang, Sequencing of 50 human exomes

reveals adaptation to high altitude. *Science* **329**, 75–78 (2010). doi:10.1126/science.1190371 Medline

93. B. vonHoldt, Z. Fan, D. Ortega-Del Vecchyo, R. K. Wayne, *EPAS1* variants in high altitude Tibetan wolves were selectively introgressed into highland dogs. *PeerJ* **5**, e3522 (2017). doi:10.7717/peerj.3522 Medline

94. M. S. Sinding, S. Gopalakrishnan, J. Ramos-Madrigal, M. de Manuel, V. V. Pitulko, L. Kuderna, T. R. Feuerborn, L. A. F. Frantz, F. G. Vieira, J. Niemann, J. A. Samaniego Castruita, C. Carøe, E. U. Andersen-Ranberg, P. D. Jordan, E. Y. Pavlova, P. A. Nikolskiy, A. K. Kasparov, V. V. Ivanova, E. Willerslev, P. Skoglund, M. Fredholm, S. E. Wennerberg, M. P. Heide-Jørgensen, R. Dietz, C. Sonne, M. Meldgaard, L. Dalén, G. Larson, B. Petersen, T. Sicheritz-Pontén, L. Bachmann, Ø. Wiig, T. Marques-Bonet, A. J. Hansen, M. T. P. Gilbert, Arctic-adapted dogs emerged at the Pleistocene-Holocene transition. *Science* **368**, 1495–1499 (2020). doi:10.1126/science.aaz8599 Medline

95. P. E. Deane-Coe, E. T. Chu, A. Slavney, A. R. Boyko, A. J. Sams, Direct-to-consumer DNA testing of 6,000 dogs reveals 98.6-kb duplication associated with blue eyes and heterochromia in Siberian Huskies. *PLOS Genet.* **14**, e1007648 (2018). doi:10.1371/journal.pgen.1007648 Medline

96. J. S. Sanjak, J. Sidorenko, M. R. Robinson, K. R. Thornton, P. M. Visscher, Evidence of directional and stabilizing selection in contemporary humans. *Proc. Natl. Acad. Sci. U.S.A.* **115**, 151–156 (2018). doi:10.1073/pnas.1707227114 Medline

97. P. C. Sabeti, S. F. Schaffner, B. Fry, J. Lohmueller, P. Varilly, O. Shamovsky, A. Palma, T. S. Mikkelsen, D. Altshuler, E. S. Lander, Positive natural selection in the human lineage. *Science* **312**, 1614–1620 (2006). doi:10.1126/science.1124309 Medline

98. E. G. Atkinson, A. X. Maihofer, M. Kanai, A. R. Martin, K. J. Karczewski, M. L. Santoro, J. C. Ulirsch, Y. Kamatani, Y. Okada, H. K. Finucane, K. C. Koenen, C. M. Nievergelt, M. J. Daly, B. M. Neale, Tractor uses local ancestry to enable the inclusion of admixed individuals in GWAS and to boost power. *Nat. Genet.* **53**, 195–204 (2021). doi:10.1038/s41588-020-00766-y Medline

99. N. Barton, J. Hermisson, M. Nordborg, Population genetics: Why structure matters. *eLife* **8**, e45380 (2019). doi:10.7554/eLife.45380 Medline

100. G. L. Wojcik, M. Graff, K. K. Nishimura, R. Tao, J. Haessler, C. R. Gignoux, H. M. Highland, Y. M. Patel, E. P. Sorokin, C. L. Avery, G. M. Belbin, S. A. Bien, I. Cheng, S. Cullina, C. J. Hodonsky, Y. Hu, L. M. Huckins, J. Jeff, A. E. Justice, J. M. Kocarnik, U. Lim, B. M. Lin, Y. Lu, S. C. Nelson, S. L. Park, H. Poisner, M. H. Preuss, M. A. Richard, C. Schurmann, V. W. Setiawan, A. Sockell, K. Vahi, M. Verbanck, A. Vishnu, R. W. Walker, K. L. Young, N. Zubair, V. Acuña-Alonso, J. L. Ambite, K. C. Barnes, E. Boerwinkle, E. P. Bottinger, C. D. Bustamante, C. Caberto, S. Canizales-Quinteros, M. P. Conomos, E. Deelman, R. Do, K. Doheny, L. Fernández-Rhodes, M. Fornage, B. Hailu, G. Heiss, B. M. Henn, L. A. Hindorff, R. D. Jackson, C. A. Laurie, C. C. Laurie, Y. Li, D.-Y. Lin, A. Moreno-Estrada, G. Nadkarni, P. J. Norman, L. C. Pooler, A. P. Reiner, J. Romm, C. Sabatti, K. Sandoval, X. Sheng, E. A. Stahl, D. O. Stram, T. A. Thornton, C. L. Wassel, L. R. Wilkens, C. A. Winkler, S. Yoneyama, S. Buyske, C. A. Haiman, C. Kooperberg, L. Le Marchand, R. J. F. Loos, T. C. Matise, K. E. North, U. Peters, E. E.

Kenny, C. S. Carlson, Genetic analyses of diverse populations improves discovery for complex traits. *Nature* **570**, 514–518 (2019). doi:10.1038/s41586-019-1310-4 Medline

101. J. E. Moore, M. J. Purcaro, H. E. Pratt, C. B. Epstein, N. Shoresh, J. Adrian, T. Kawli, C. A. Davis, A. Dobin, R. Kaul, J. Halow, E. L. Van Nostrand, P. Freese, D. U. Gorkin, Y. Shen, Y. He, M. Mackiewicz, F. Pauli-Behn, B. A. Williams, A. Mortazavi, C. A. Keller, X.-O. Zhang, S. I. Elhajjajy, J. Huey, D. E. Dickel, V. Snetkova, X. Wei, X. Wang, J. C. Rivera-Mulia, J. Rozowsky, J. Zhang, S. B. Chhetri, J. Zhang, A. Victorsen, K. P. White, A. Visel, G. W. Yeo, C. B. Burge, E. Lécuyer, D. M. Gilbert, J. Dekker, J. Rinn, E. M. Mendenhall, J. R. Ecker, M. Kellis, R. J. Klein, W. S. Noble, A. Kundaje, R. Guigó, P. J. Farnham, J. M. Cherry, R. M. Myers, B. Ren, B. R. Graveley, M. B. Gerstein, L. A. Pennacchio, M. P. Snyder, B. E. Bernstein, B. Wold, R. C. Hardison, T. R. Gingeras, J. A. Stamatoyannopoulos, Z. Weng, ENCODE Project Consortium, Expanded encyclopaedias of DNA elements in the human and mouse genomes. *Nature* **583**, 699–710 (2020). doi:10.1038/s41586-020-2493-4 Medline

102. K. Megquier, D. P. Genereux, J. Hekman, R. Swofford, J. Turner-Maier, J. Johnson, J. Alonso, X. Li, K. Morrill, L. J. Anguish, M. Koltookian, B. Logan, C. R. Sharp, L. Ferrer, K. Lindblad-Toh, V. N. Meyers-Wallen, A. Hoffman, E. K. Karlsson, BarkBase: Epigenomic annotation of canine genomes. *Genes* **10**, 433 (2019). doi:10.3390/genes10060433 Medline

103. Zoonomia Consortium, A comparative genomics multitool for scientific discovery and conservation. *Nature* **587**, 240–245 (2020). doi:10.1038/s41586-020-2876-6 Medline

104. K. Morrill, J. Hekman, X. Li, J. McClure, B. Logan, L. Goodman, M. Gao, Y. Dong, M. Alonso, E. Carmichael, N. Snyder-Mackler, J. Alonso, H. J. Noh, J. Johnson, M. Koltookian, C. Lieu, K. Megquier, R. Swofford, J. Turner-Maier, M. E. White, Z. Weng, A. Colubri, D. P. Genereux, K. A. Lord, E. K. Karlsson, Associated data files and scripts for "Ancestry-inclusive dog genomics challenges popular breed stereotypes". Dryad (2022); https://doi.org/10.5061/dryad.g4f4qrfr0.

105. K. Morrill, J. Hekman, X. Li, J. McClure, B. Logan, L. Goodman, M. Gao, Y. Dong, M. Alonso, E. Carmichael, N. Snyder-Mackler, J. Alonso, H. J. Noh, J. Johnson, M. Koltookian, C. Lieu, K. Megquier, R. Swofford, J. Turner-Maier, M. E. White, Z. Weng, A. Colubri, D. P. Genereux, K. A. Lord, E. K. Karlsson, Associated data files and scripts for "Ancestry-inclusive dog genomics challenges popular breed stereotypes". Zenodo (2022); https://doi.org/10.5281/zenodo.5808330.

106. K. Morrill, Darwin's Ark - Data release 2022. Terra (2022); https://app.terra.bio/#workspaces/darwins-ark/Darwins%20Ark%20-%20Data%20Release%202022.

107. K. Morrill, J. Hekman, X. Li, J. McClure, B. Logan, L. Goodman, M. Gao, Y. Dong, M. Alonso, E. Carmichael, N. Snyder-Mackler, J. Alonso, H. J. Noh, J. Johnson, M. Koltookian, C. Lieu, K. Megquier, R. Swofford, J. Turner-Maier, M. E. White, Z. Weng, A. Colubri, D. P. Genereux, K. A. Lord, E. K. Karlsson, Additional figures for "Ancestry-inclusive dog genomics challenges popular breed stereotypes". FigShare (2022); https://doi.org/10.6084/m9.figshare.16608793.

108. R. Trevejo, M. Yang, E. M. Lund, Epidemiology of surgical castration of dogs and cats in the United States. *J. Am. Vet. Med. Assoc.* **238**, 898–904 (2011). doi:10.2460/javma.238.7.898 Medline

109. J. Plassais, M. Rimbault, F. J. Williams, B. W. Davis, J. J. Schoenebeck, E. A. Ostrander, Analysis of large versus small dogs reveals three genes on the canine X chromosome associated with body weight, muscling and back fat thickness. *PLOS Genet.* **13**, e1006661 (2017). doi:10.1371/journal.pgen.1006661 Medline

110. H. E. Salvin, P. D. McGreevy, P. S. Sachdev, M. J. Valenzuela, Growing old gracefully—Behavioral changes associated with "successful aging" in the dog, *Canis familiaris*. *J. Vet. Behav.* **6**, 313–320 (2011). doi:10.1016/j.jveb.2011.04.004

111. G. Raîche, T. A. Walls, D. Magis, M. Riopel, J.-G. Blais, Non-graphical solutions for Cattell's scree test. *Methodology* **9**, 23–29 (2013). doi:10.1027/1614-2241/a000051

112. F. C. Ceballos, S. Hazelhurst, M. Ramsay, Assessing runs of homozygosity: A comparison of SNP array and whole genome sequence low coverage data. *BMC Genomics* **19**, 106 (2018). doi:10.1186/s12864-018-4489-0 Medline

113. N. Li, M. Stephens, Modeling linkage disequilibrium and identifying recombination hotspots using single-nucleotide polymorphism data. *Genetics* **165**, 2213–2233 (2003). doi:10.1093/genetics/165.4.2213 Medline

114. R. Meyermans, W. Gorssen, N. Buys, S. Janssens, How to study runs of homozygosity using PLINK? A guide for analyzing medium density SNP data in livestock and pet species. *BMC Genomics* **21**, 94 (2020). doi:10.1186/s12864-020-6463-x Medline

115. R. McQuillan, A.-L. Leutenegger, R. Abdel-Rahman, C. S. Franklin, M. Pericic, L. Barac-Lauc, N. Smolej-Narancic, B. Janicijevic, O. Polasek, A. Tenesa, A. K. Macleod, S. M. Farrington, P. Rudan, C. Hayward, V. Vitart, I. Rudan, S. H. Wild, M. G. Dunlop, A. F. Wright, H. Campbell, J. F. Wilson, Runs of homozygosity in European populations. *Am. J. Hum. Genet.* **83**, 359–372 (2008). doi:10.1016/j.ajhg.2008.08.007 Medline

116. A. J. Sams, A. R. Boyko, Fine-scale resolution of runs of homozygosity reveal patterns of inbreeding and substantial overlap with recessive disease genotypes in domestic dogs. *G3* **9**, 117–123 (2019). doi:10.1534/g3.118.200836 Medline

117. A. Manichaikul, J. C. Mychaleckyj, S. S. Rich, K. Daly, M. Sale, W.-M. Chen, Robust relationship inference in genome-wide association studies. *Bioinformatics* **26**, 2867–2873 (2010). doi:10.1093/bioinformatics/btq559 Medline

118. J. J. Hayward, M. G. Castelhano, K. C. Oliveira, E. Corey, C. Balkman, T. L. Baxter, M. L. Casal, S. A. Center, M. Fang, S. J. Garrison, S. E. Kalla, P. Korniliev, M. I. Kotlikoff, N. S. Moise, L. M. Shannon, K. W. Simpson, N. B. Sutter, R. J. Todhunter, A. R. Boyko, Complex disease and phenotype mapping in the domestic dog. *Nat. Commun.* **7**, 10460 (2016). doi:10.1038/ncomms10460 Medline

119. A. Auton, Y. Rui Li, J. Kidd, K. Oliveira, J. Nadel, J. K. Holloway, J. J. Hayward, P. E. Cohen, J. M. Greally, J. Wang, C. D. Bustamante, A. R. Boyko, Genetic recombination is targeted towards gene promoter regions in dogs. *PLOS Genet.* **9**, e1003984 (2013). doi:10.1371/journal.pgen.1003984 Medline

120. D. H. Alexander, J. Novembre, K. Lange, Fast model-based estimation of ancestry in

unrelated individuals. *Genome Res.* **19**, 1655–1664 (2009). doi:10.1101/gr.094052.109 Medline

121. L. M. Evans, R. Tahmasbi, S. I. Vrieze, G. R. Abecasis, S. Das, S. Gazal, D. W. Bjelland, T. R. de Candia, M. E. Goddard, B. M. Neale, J. Yang, P. M. Visscher, M. C. Keller, Haplotype Reference Consortium, Comparison of methods that use whole genome data to estimate the heritability and genetic architecture of complex traits. *Nat. Genet.* **50**, 737–745 (2018). doi:10.1038/s41588-018-0108-x Medline

122. P. H. Westfall, S. Stanley Young, *Resampling-Based Multiple Testing: Examples and Methods for p-Value Adjustment* (Wiley, 1993).

123. B. Phipson, G. K. Smyth, Permutation P-values should never be zero: Calculating exact P-values when permutations are randomly drawn. *Stat. Appl. Genet. Mol. Biol.* **9**, e39 (2010). doi:10.2202/1544-6115.1585 Medline

124. A. Ziyatdinov, M. Vázquez-Santiago, H. Brunel, A. Martinez-Perez, H. Aschard, J. M. Soria, lme4qtl: Linear mixed models with flexible covariance structure for genetic studies of related individuals. *BMC Bioinformatics* **19**, 68 (2018). doi:10.1186/s12859-018-2057-x Medline

125. D. Bates, M. Mächler, B. Bolker, S. Walker, Fitting linear mixed-effects models using lme4. arXiv:1406.5823 [stat.CO] (2014).

126. D. Bates, lmer, p-values and all that (2006); https://stat.ethz.ch/pipermail/r-help/2006-May/094765.html.

127. S. Nakagawa, H. Schielzeth, A general and simple method for obtaining R^2 from generalized linear mixed-effects models. *Methods Ecol. Evol.* **4**, 133–142 (2013). doi:10.1111/j.2041-210x.2012.00261.x

128. B. Langmead, S. L. Salzberg, Fast gapped-read alignment with Bowtie 2. *Nat. Methods* **9**, 357–359 (2012). doi:10.1038/nmeth.1923 Medline

129. Y. Zhang, T. Liu, C. A. Meyer, J. Eeckhoute, D. S. Johnson, B. E. Bernstein, C. Nusbaum, R. M. Myers, M. Brown, W. Li, X. S. Liu, Model-based analysis of ChIP-seq (MACS). *Genome Biol.* **9**, R137 (2008). doi:10.1186/gb-2008-9-9-r137 Medline

130. K. Wang, X. Hu, Y. Peng, An analytical comparison of the principal component method and the mixed effects model for association studies in the presence of cryptic relatedness and population stratification. *Hum. Hered.* **76**, 1–9 (2013). doi:10.1159/000353345 Medline

131. J. Yang, N. A. Zaitlen, M. E. Goddard, P. M. Visscher, A. L. Price, Advantages and pitfalls in the application of mixed-model association methods. *Nat. Genet.* **46**, 100–106 (2014). doi:10.1038/ng.2876 Medline

132. L. Jiang, Z. Zheng, T. Qi, K. E. Kemper, N. R. Wray, P. M. Visscher, J. Yang, A resource-efficient tool for mixed model association analysis of large-scale data. *Nat. Genet.* **51**, 1749–1755 (2019). doi:10.1038/s41588-019-0530-8 Medline

133. E. Demidenko, *Mixed Models: Theory and Applications with R* (Wiley, 2013).

134. L. Ma, G. R. Wiggans, S. Wang, T. S. Sonstegard, J. Yang, B. A. Crooker, J. B. Cole, C. P. Van Tassell, T. J. Lawlor, Y. Da, Effect of sample stratification on dairy GWAS results. *BMC Genomics* **13**, 536 (2012). doi:10.1186/1471-2164-13-536 Medline

135. A.-S. Malaspinas, M. C. Westaway, C. Muller, V. C. Sousa, O. Lao, I. Alves, A. Bergström, G. Athanasiadis, J. Y. Cheng, J. E. Crawford, T. H. Heupink, E. Macholdt, S. Peischl, S. Rasmussen, S. Schiffels, S. Subramanian, J. L. Wright, A. Albrechtsen, C. Barbieri, I. Dupanloup, A. Eriksson, A. Margaryan, I. Moltke, I. Pugach, T. S. Korneliussen, I. P. Levkivskyi, J. V. Moreno-Mayar, S. Ni, F. Racimo, M. Sikora, Y. Xue, F. A. Aghakhanian, N. Brucato, S. Brunak, P. F. Campos, W. Clark, S. Ellingvåg, G. Fourmile, P. Gerbault, D. Injie, G. Koki, M. Leavesley, B. Logan, A. Lynch, E. A. Matisoo-Smith, P. J. McAllister, A. J. Mentzer, M. Metspalu, A. B. Migliano, L. Murgha, M. E. Phipps, W. Pomat, D. Reynolds, F.-X. Ricaut, P. Siba, M. G. Thomas, T. Wales, C. M. Wall, S. J. Oppenheimer, C. Tyler-Smith, R. Durbin, J. Dortch, A. Manica, M. H. Schierup, R. A. Foley, M. M. Lahr, C. Bowern, J. D. Wall, T. Mailund, M. Stoneking, R. Nielsen, M. S. Sandhu, L. Excoffier, D. M. Lambert, E. Willerslev, A genomic history of Aboriginal Australia. *Nature* **538**, 207–214 (2016). doi:10.1038/nature18299 Medline

136. C.-J. Rubin, M. C. Zody, J. Eriksson, J. R. S. Meadows, E. Sherwood, M. T. Webster, L. Jiang, M. Ingman, T. Sharpe, S. Ka, F. Hallböök, F. Besnier, O. Carlborg, B. Bed'hom, M. Tixier-Boichard, P. Jensen, P. Siegel, K. Lindblad-Toh, L. Andersson, Whole-genome resequencing reveals loci under selection during chicken domestication. *Nature* **464**, 587–591 (2010). doi:10.1038/nature08832 Medline

137. C. C. Little, *The Inheritance of Coat Color in Dogs* (Comstock Pub. Associates, 1957).

138. M. N. Weedon, G. Lettre, R. M. Freathy, C. M. Lindgren, B. F. Voight, J. R. B. Perry, K. S. Elliott, R. Hackett, C. Guiducci, B. Shields, E. Zeggini, H. Lango, V. Lyssenko, N. J. Timpson, N. P. Burtt, N. W. Rayner, R. Saxena, K. Ardlie, J. H. Tobias, A. R. Ness, S. M. Ring, C. N. A. Palmer, A. D. Morris, L. Peltonen, V. Salomaa, G. Davey Smith, L. C. Groop, A. T. Hattersley, M. I. McCarthy, J. N. Hirschhorn, T. M. Frayling, Diabetes Genetics Initiative, Wellcome Trust Case Control Consortium, A common variant of *HMGA2* is associated with adult and childhood height in the general population. *Nat. Genet.* **39**, 1245–1250 (2007). doi:10.1038/ng2121 Medline

139. U. Sovio, A. J. Bennett, I. Y. Millwood, J. Molitor, P. F. O'Reilly, N. J. Timpson, M. Kaakinen, J. Laitinen, J. Haukka, D. Pillas, I. Tzoulaki, J. Molitor, C. Hoggart, L. J. M. Coin, J. Whittaker, A. Pouta, A.-L. Hartikainen, N. B. Freimer, E. Widen, L. Peltonen, P. Elliott, M. I. McCarthy, M.-R. Jarvelin, Genetic determinants of height growth assessed longitudinally from infancy to adulthood in the northern Finland birth cohort 1966. *PLOS Genet.* **5**, e1000409 (2009). doi:10.1371/journal.pgen.1000409 Medline

140. J. Metzger, R. Schrimpf, U. Philipp, O. Distl, Expression levels of *LCORL* are associated with body size in horses. *PLOS ONE* **8**, e56497 (2013). doi:10.1371/journal.pone.0056497 Medline

141. J. Tetens, P. Widmann, C. Kühn, G. Thaller, A genome-wide association study indicates *LCORL/NCAPG* as a candidate locus for withers height in German Warmblood horses. *Anim. Genet.* **44**, 467–471 (2013). doi:10.1111/age.12031 Medline

142. J. E. Pryce, B. J. Hayes, S. Bolormaa, M. E. Goddard, Polymorphic regions affecting human height also control stature in cattle. *Genetics* **187**, 981–984 (2011). doi:10.1534/genetics.110.123943 Medline

143. A. C. Bouwman, H. D. Daetwyler, A. J. Chamberlain, C. H. Ponce, M. Sargolzaei, F. S. Schenkel, G. Sahana, A. Govignon-Gion, S. Boitard, M. Dolezal, H. Pausch, R. F. Brøndum, P. J. Bowman, B. Thomsen, B. Guldbrandtsen, M. S. Lund, B. Servin, D. J. Garrick, J. Reecy, J. Vilkki, A. Bagnato, M. Wang, J. L. Hoff, R. D. Schnabel, J. F. Taylor, A. A. E. Vinkhuyzen, F. Panitz, C. Bendixen, L.-E. Holm, B. Gredler, C. Hozé, M. Boussaha, M.-P. Sanchez, D. Rocha, A. Capitan, T. Tribout, A. Barbat, P. Croiseau, C. Drögemüller, V. Jagannathan, C. Vander Jagt, J. J. Crowley, A. Bieber, D. C. Purfield, D. P. Berry, R. Emmerling, K.-U. Götz, M. Frischknecht, I. Russ, J. Sölkner, C. P. Van Tassell, R. Fries, P. Stothard, R. F. Veerkamp, D. Boichard, M. E. Goddard, B. J. Hayes, Meta-analysis of genome-wide association studies for cattle stature identifies common genes that regulate body size in mammals. *Nat. Genet.* **50**, 362–367 (2018). doi:10.1038/s41588-018-0056-5 Medline

144. E. Gadelha Pereira Fontenele, M. E. Amaral de Moraes, C. Brasil d’Alva, D. P. Pinheiro, S. Aguiar Sales Pinheiro Landim, F. A. de Sousa Barros, E. Barbosa Trarbach, B. Bilharinho de Mendonca, A. A. Lima Jorge, Association study of GWAS-derived loci with height in Brazilian children: Importance of *MAP3K3*, *MMP24* and *IGF1R* polymorphisms for height variation. *Horm. Res. Paediatr.* **84**, 248–253 (2015). doi:10.1159/000437324 Medline

145. M. D. Mavalli, D. J. DiGirolamo, Y. Fan, R. C. Riddle, K. S. Campbell, T. van Groen, S. J. Frank, M. A. Sperling, K. A. Esser, M. M. Bamman, T. L. Clemens, Distinct growth hormone receptor signaling modes regulate skeletal muscle development and insulin sensitivity in mice. *J. Clin. Invest.* **120**, 4007–4020 (2010). doi:10.1172/JCI42447 Medline

146. Y. Cao, X. Song, H. Shan, J. Jiang, P. Xiong, J. Wu, F. Shi, Y. Jiang, Genome-wide association study of body weights in Hu sheep and population verification of related single-nucleotide polymorphisms. *Front. Genet.* **11**, 588 (2020). doi:10.3389/fgene.2020.00588 Medline

147. S. M. Schmutz, T. G. Berryere, D. L. Dreger, *MITF* and white spotting in dogs: A population study. *J. Hered.* **100**, S66–S74 (2009). doi:10.1093/jhered/esp029

148. I. Baranowska Körberg, E. Sundström, J. R. S. Meadows, G. Rosengren Pielberg, U. Gustafson, Å. Hedhammar, E. K. Karlsson, J. Seddon, A. Söderberg, C. Vilà, X. Zhang, M. Åkesson, K. Lindblad-Toh, G. Andersson, L. Andersson, A simple repeat polymorphism in the *MITF-M* promoter is a key regulator of white spotting in dogs. *PLOS ONE* **9**, e104363 (2014). doi:10.1371/journal.pone.0104363 Medline

149. S. Hofstetter, F. Seefried, I. M. Häfliger, V. Jagannathan, T. Leeb, C. Drögemüller, A non-coding regulatory variant in the 5′-region of the *MITF* gene is associated with white-spotted coat in Brown Swiss cattle. *Anim. Genet.* **50**, 27–32 (2019). doi:10.1111/age.12751 Medline

150. G. Wang, J. Liao, M. Tang, S. Yu, Genetic variation in the *MITF* promoter affects skin colour and transcriptional activity in black-boned chickens. *Br. Poult. Sci.* **59**, 21–27 (2018). doi:10.1080/00071668.2017.1379053 Medline

151. J. M. Paris, A. Letko, I. M. Häfliger, P. Ammann, C. Drögemüller, Ear type in sheep is associated with the *MSRB3* locus. *Anim. Genet.* **51**, 968–972 (2020). doi:10.1111/age.12994 Medline

152. C. Chen, C. Liu, X. Xiong, S. Fang, H. Yang, Z. Zhang, J. Ren, Y. Guo, L. Huang, Copy number variation in the *MSRB3* gene enlarges porcine ear size through a mechanism involving miR-584-5p. *Genet. Sel. Evol.* **50**, 72 (2018). doi:10.1186/s12711-018-0442-6 Medline

153. L. Zhang, J. Liang, W. Luo, X. Liu, H. Yan, K. Zhao, H. Shi, Y. Zhang, L. Wang, L. Wang, Genome-wide scan reveals *LEMD3* and *WIF1* on SSC5 as the candidates for porcine ear size. *PLOS ONE* **9**, e102085 (2014). doi:10.1371/journal.pone.0102085 Medline

154. J. Liang, N. Li, L.-C. Zhang, L.-G. Wang, X. Liu, K.-B. Zhao, H. Yan, L. Pu, Y.-B. Zhang, H.-B. Shi, Q. Zhang, L.-X. Wang, Porcine *LEM domain-containing 3*: Molecular cloning, functional characterization, and polymorphism associated with ear size. *J. Integr. Agric.* **15**, 1321–1329 (2016). doi:10.1016/S2095-3119(15)61173-X

155. D. J. E. Housley, P. J. Venta, The long and the short of it: Evidence that *FGF5* is a major determinant of canine 'hair'-itability. *Anim. Genet.* **37**, 309–315 (2006). doi:10.1111/j.1365-2052.2006.01448.x Medline

156. C. Dierks, S. Mömke, U. Philipp, O. Distl, Allelic heterogeneity of *FGF5* mutations causes the long-hair phenotype in dogs. *Anim. Genet.* **44**, 425–431 (2013). doi:10.1111/age.12010 Medline

157. C. Drögemüller, S. Rüfenacht, B. Wichert, T. Leeb, Mutations within the *FGF5* gene are associated with hair length in cats. *Anim. Genet.* **38**, 218–221 (2007). doi:10.1111/j.1365-2052.2007.01590.x Medline

158. H. G. Parker, K. Chase, E. Cadieu, K. G. Lark, E. A. Ostrander, An insertion in the *RSPO2* gene correlates with improper coat in the Portuguese water dog. *J. Hered.* **101**, 612–617 (2010). doi:10.1093/jhered/esq068 Medline

159. T. Kawakami, M. K. Jensen, A. Slavney, P. E. Deane, A. Milano, V. Raghavan, B. Ford, E. T. Chu, A. J. Sams, A. R. Boyko, R-locus for roaned coat is associated with a tandem duplication in an intronic region of *USH2A* in dogs and also contributes to Dalmatian spotting. *PLOS ONE* **16**, e0248233 (2021). doi:10.1371/journal.pone.0248233 Medline

160. L. A. Clark, J. M. Wahl, C. A. Rees, K. E. Murphy, Retrotransposon insertion in *SILV* is responsible for merle patterning of the domestic dog. *Proc. Natl. Acad. Sci. U.S.A.* **103**, 1376–1381 (2006). doi:10.1073/pnas.0506940103 Medline

161. B. C. Ballif, C. J. Ramirez, C. R. Carl, K. Sundin, M. Krug, A. Zahand, L. G. Shaffer, H. Flores-Smith, The *PMEL* gene and merle in the domestic dog: A continuum of insertion lengths leads to a spectrum of coat color variations in Australian shepherds and related breeds. *Cytogenet. Genome Res.* **156**, 22–34 (2018). doi:10.1159/000491408 Medline

162. Z. Pelles, A. Gáspárdy, L. Zöldág, X. Lénárt, N. Ninausz, L. Varga, P. Zenke, Merle allele variations in the Mudi dog breed and their effects on phenotypes. *Acta Vet. Hung.* **67**, 159–173 (2019). doi:10.1556/004.2019.018 Medline

163. Y. Jin, S. A. Birlea, P. R. Fain, C. M. Mailloux, S. L. Riccardi, K. Gowan, P. J. Holland, D. C. Bennett, M. R. Wallace, W. T. McCormack, E. H. Kemp, D. J. Gawkrodger, A. P.

Weetman, M. Picardo, G. Leone, A. Taïeb, T. Jouary, K. Ezzedine, N. van Geel, J. Lambert, A. Overbeck, R. A. Spritz, Common variants in *FOXP1* are associated with generalized vitiligo. *Nat. Genet.* **42**, 576–578 (2010). doi:10.1038/ng.602 Medline

164. E. Leishman, J. M. Howard, G. E. Garcia, Q. Miao, A. T. Ku, J. D. Dekker, H. Tucker, H. Nguyen, Foxp1 maintains hair follicle stem cell quiescence through regulation of Fgf18. *Development* **140**, 3809–3818 (2013). doi:10.1242/dev.097477 Medline

165. Y. Bian, G. Wei, X. Song, L. Yuan, H. Chen, T. Ni, D. Lu, Global downregulation of pigmentation-associated genes in human premature hair graying. *Exp. Ther. Med.* **18**, 1155–1163 (2019). doi:10.3892/etm.2019.7663 Medline

166. E. J. Cargill, T. R. Famula, R. D. Schnabel, G. M. Strain, K. E. Murphy, The color of a Dalmatian's spots: Linkage evidence to support the *TYRP1* gene. *BMC Vet. Res.* **1**, 1 (2005). doi:10.1186/1746-6148-1-1 Medline

167. H. E. Wright, E. Schofield, C. S. Mellersh, L. M. Burmeister, A novel *TYRP1* variant is associated with liver and tan coat colour in Lancashire heelers. *Anim. Genet.* **50**, 783 (2019). doi:10.1111/age.12839 Medline

168. L. V. Monteagudo, M. T. Tejedor, The b^c allele of *TYRP1* is causative for the recessive brown (liver) colour in German Shepherd dogs. *Anim. Genet.* **46**, 588–589 (2015). doi:10.1111/age.12337 Medline

169. S. Rieder, S. Taourit, D. Mariat, B. Langlois, G. Guérin, Mutations in the agouti (*ASIP*), the extension (*MC1R*), and the brown (*TYRP1*) loci and their association to coat color phenotypes in horses (*Equus caballus*). *Mamm. Genome* **12**, 450–455 (2001). doi:10.1007/s003350020017 Medline