Downloaded from genome.cshlp.org on March 2, 2026 - Published by Cold Spring Harbor Laboratory Press

Analysis of canine gene constraint identifies new variants for orofacial clefts and stature

Reuben M Buckley[1], Nüket Bilgen[2], Alexander C Harris[1], Peter Savolainen[3], Cafer Tepeli[4], Metin Erdoğan[5], Aitor Serres Armero[1], Dayna L Dreger[1], Frank G van Steenbeek[6], Marjo K Hytönen[7,8,9], Heidi G Parker[1], Jessica Hale[1], Hannes Lohi[7,8,9], Bengi Çınar Kul[2], Adam R Boyko[10,11], Elaine A Ostrander[1]*

[1]National Human Genome Research Institute, National Institutes of Health, Bethesda, MD 20892, USA.

[2]Department of Animal Genetics, Faculty of Veterinary Medicine, University of Ankara, Ankara 06110, Türkiye

[3]KTH Royal Institute of Technology, School of Chemistry, Biotechnology and Health, Science for Life Laboratory, Stockholm, Sweden

[4]Department of Animal Science, University of Selcuk, Faculty of Veterinary Medicine, Konya, Türkiye

[5]Department of Veterinary Biology and Genetics, Faculty of Veterinary Medicine, Afyon Kocatepe University, Afyonkarahisar, Türkiye

[6]Utrecht University, Faculty of Veterinary Medicine, Dept. of Clinical Sciences, The Netherlands

[7]Department of Medical and Clinical Genetics, University of Helsinki, 00014 Helsinki, Finland

[8]Department of Veterinary Biosciences, University of Helsinki, 00014 Helsinki, Finland

[9]Folkhälsan Research Center, 00290 Helsinki, Finland

[10]Department of Biomedical Sciences, College of Veterinary Medicine, Cornell University, Ithaca, NY 14853, USA

[11]Embark Veterinary, Inc., Boston, MA 02210, USA

*Corresponding author

Downloaded from genome.cshlp.org on March 2, 2026 - Published by Cold Spring Harbor Laboratory Press

Author info:

Reuben M Buckley (RMB) (0000-0001-6334-3906) - reuben.buckley@nih.gov

Nüket Bilgen (NB) (0000-0003-2324-7965) - nbilgen@ankara.edu.tr

Alexander C Harris (ACH) (0000-0002-6868-5363)- alex.harris@nih.gov

Peter Savolainen (PS) (0000-0002-1495-8338) - savo@kth.se

Cafer Tepeli (CT) (0000-0003-2966-8330) - ctepeli@selcuk.edu.tr

Metin Erdoğan (ME) (0000-0003-0975-1989) - erdogan@aku.edu.tr

Aitor Serres Armero (ASA) (0000-0003-1679-811X) - aitor.serresarmero@nih.gov

Dayna L Dreger (DLD) (0000-0002-3218-6431) - dayna.dreger@nih.gov

Frank G van Steenbeek (FGVS) (0000-0001-5460-6540) - F.G.vanSteenbeek@uu.nl

Marjo K Hytönen (MKH) (0000-0003-1976-5874) - marjo.hytonen@helsinki.fi

Heidi G Parker (HGP) (0000-0002-9707-6380) - hgparker@nhgri.nih.gov

Jessica Hale (JH) (0009-0008-4953-9556) - jessica.hale@nih.gov

Hannes Lohi (HL) (0000-0003-1087-5532) - hannes.lohi@helsinki.fi

Bengi Çınar Kul (BCK) (0000-0002-8955-0097) - bengikul@gmail.com

Adam R Boyko (ARB) (0000-0002-9125-6874) - boyko@cornell.edu

Elaine A Ostrander (EAO) (0000-0001-6075-9738) - eostrand@mail.nih.gov

Downloaded from genome.cshlp.org on March 2, 2026 - Published by Cold Spring Harbor Laboratory Press

Abstract

Dog breeding promotes within-group homogeneity through conformation to strict breed standards, while simultaneously driving between-group heterogeneity. There are over 350 recognized dog breeds that provide the foundation for investigating the genetic basis of phenotypic diversity. Typically, breed standard phenotypes such as stature, pelage, and craniofacial structure are analyzed through genetic association studies. However, such analyses are limited to assayed phenotypes only, leaving difficult to measure phenotypic subtleties easily overlooked. We investigated coding variation from over 2,000 dogs, leading to discoveries of variants related to craniofacial morphology and stature. Breed-enriched variants were prioritized according to gene constraint, which was calculated using a mutation model derived from trinucleotide substitution probabilities. Among the newly found variants was a splice-acceptor variant in *PDGFRA* associated with bifid nose, a characteristic trait of Çatalburun dogs, implicating the gene's role in midline closure. Two additional *LCORL* variants, both associated with canine body size were also discovered: a frameshift that causes a premature stop in large breeds (>25 kg) and an intronic substitution found in small breeds (<10 kg), thus highlighting the importance of allelic heterogeneity in selection for breed traits. Most variants prioritized in this analysis were not associated with genomic signatures for breed differentiation, as these regions were enriched for constrained genes intolerant to nonsynonymous variation. This indicates trait selection in dogs is likely a balancing act between preserving essential gene functions and maximizing regulatory variation to drive phenotypic extremes.

Downloaded from genome.cshlp.org on March 2, 2026 - Published by Cold Spring Harbor Laboratory Press

Introduction

The genomes of domestic species hold immense promise for the discovery of novel genotype-phenotype associations (Andersson and Purugganan 2022; Buckley and Ostrander 2024). The domestic dog comprises over 350 breeds (Rogers and Brace 1995; Fogle 2000), many of which exhibit extreme levels of phenotypic diversity, including those associated with morphology, behavior, and disease susceptibility (Ostrander et al. 2017; Leeb et al. 2023). Identifying genetic variants that drive canine phenotypic variation will further define gene functions.

Genetic association studies typically require an initial and robust characterization of a phenotype, a practice that ultimately limits the scope of potential findings to only those that can be recognized and sufficiently defined. In this study, we explore how gene constraint can be used to identify variants with potential to impact phenotypes as trait-associated candidates, later determining association by investigating breed traits and characteristics of dogs carrying these alleles. Because variants are first prioritized according to functional impact, we can identify trait associations for variants often overlooked by traditional genome-wide association studies (GWASs).

The rigid population structure and extreme phenotypic diversity across breeds is an essential component of our analysis. By definition, dog breeds are closed populations, such that every registered dog is descended from a small number of historic founders and shares extensive genetic homogeneity (Mellanby et al. 2013; Lampi et al. 2020). Breed members have multigeneration pedigrees and are judged against morphologic standards that further enforce genetic homogeneity, resulting in a reduction in the locus heterogeneity that typifies human traits, where dozens or more genes may be foundational for a given trait. The ease of association studies in dogs is further enhanced by the fact that most breeds were developed within the last 200 years and, as such, there are minimal numbers of generations for meiotic recombination to occur (Worboys 2018). This unique population structure means that breed membership can be used as a surrogate when mapping breed-associated morphologic traits,

Downloaded from genome.cshlp.org on March 2, 2026 - Published by Cold Spring Harbor Laboratory Press

thus avoiding the time and expense of collecting individual metrics (Gough et al. 2018; Plassais et al. 2019; Dutrow et al. 2022; Morrill et al. 2022).

In humans, candidate trait and disease variants are generally identified using a combination of strategies including multiple species alignment, protein structure predictions, gene mutation tolerance, and machine learning (Ng and Henikoff 2001; Adzhubei et al. 2010; Havrilla et al. 2019; Rentzsch et al. 2019; Karczewski et al. 2020; Cheng et al. 2023; Gao et al. 2023; Sullivan et al. 2023). Most such tools benefit from extensive catalogues of human variation featuring large amounts of clinical and personal data (Landrum et al. 2014; Lek et al. 2016; Bycroft et al. 2018; All of Us Research Program Genomics 2024). Although efforts to catalogue trait-associated genetic variation exist for domestic mammals (Nicholas et al. 1995; Plassais et al. 2019; Nicholas 2021; Meadows et al. 2023), the resources for doing so are limiting. While mapping functional gene/genomic annotations over from humans can theoretically be used for improving variant annotation, it is estimated that ~13% of human pathogenic ClinVar variants are common in dogs, and thus likely benign (Gao et al. 2023). Systems such as the dog are therefore rife with lost opportunities for identifying genetic variants that play fundamental roles in mammalian biology. Nevertheless, the recent Dog10K release of whole genome sequence (WGS) data from over 2000 dogs (Ostrander et al. 2019; Meadows et al. 2023), inclusive of 321 breeds and populations, means the time is ripe to use measures of functional and evolutionary constraint to identify new variants with a high probability of impacting gene function.

This study focuses, specifically, on identifying genotypes that are rare overall, but common within a single breed or groups of related breeds that share a particular trait. We characterize coding variation across 2,291 canine genomes from over 200 phylogenetically informed breeds/subpopulations and develop a canine-specific mutation model to determine whether gene/site constraint metrics are predictive of variant fitness consequences. From this, we identify breed enriched variants predicted to cause loss of function in genes under

Downloaded from genome.cshlp.org on March 2, 2026 - Published by Cold Spring Harbor Laboratory Press

evolutionary constraint, identifying trait-causing alleles requiring additional study. Finally, we investigate the contribution of coding variation to breed differentiation, identifying the central role constrained genes play in shaping phenotypic differences between breeds. Collectively, our work provides a road map for performing constraint-based analysis of genetic variation in other domesticated species with similar genomic resources, such as cattle, swine, and horses.

Results

A framework for evaluating gene sequence constraint metrics

Genomic sequence constraint metrics are often used to predict whether specific mutations are likely to be deleterious. Characteristic breed traits are often phenotypic extremes that can be linked to deleterious health outcomes, such as in brachycephalic breeds (Marchant et al. 2019; O'Neill et al. 2020). Sequence constraint can therefore be used to prioritize likely trait-causing variants. To explore the utility of gene sequence constraint for this purpose, we assembled a dataset of over 2,291 high quality genomes and used phylogenetic clustering to group them into 200 breeds/groups of five or more (**Supplemental Fig. S1-S2**)(**Supplemental Table S1-S3**). Throughout the coding sequence (CDS) and splice sites, we captured 611,644 single nucleotide variants (SNVs), 28,970 small deletions, and 28,123 small insertions, including multi-allelic variants (**Methods**) (**Supplemental Fig. S3**) (**Supplemental Table S4**).

Three separate metrics for constraint are available in the dog: constrained CDS, human ortholog LOEUF (loss of function observed over expected upper fraction), and variant phyloP score. Constrained CDS and variant phyloP score both represent evolutionary conservation of canine genomic sequence derived from the Zoonomia consortium's alignment of 240 mammals. Variant phyloP score is the level of conservation of an individual nucleotide position. Constrained CDS is the fraction of nucleotides within a gene's CDS that have a phyloP score >2.27 (genome-wide FDR > 0.05) (Christmas et al. 2023; Sullivan et al. 2023). Human ortholog LOEUF indicates whether a human-dog ortholog is depleted of predicted loss of function (LoF)

Downloaded from genome.cshlp.org on March 2, 2026 - Published by Cold Spring Harbor Laboratory Press

variation in humans (Karczewski et al. 2020), where LoF here specifically represents nonsense, start loss, splice disruptor, and frameshift variant annotations (MacArthur et al. 2012). Importantly, these metrics do not measure canine coding constraint directly. Instead, evolutionary-based metrics, such as constrained CDS and variant phyloP score, identify sites and genes conserved across many species and are blind to lineage-specific effects. Similarly, because LOEUF values were calculated in human genes they are blind to lineage-specific selection in canids.

We evaluated the capacity of each constraint measure to predict gene tolerance to mutations within dogs by measuring their concordance with the ratio of observed to expected mutations in canine genes. Calculation of the observed mutation rates must only capture negative selection effects and not canine demographic history, which has been largely shaped by recurrent population bottlenecks and artificial selection. One approach for achieving this is to analyze recent spontaneous mutations whose presence in the genome is not reflective of population history. Singleton variants in dogs with highly represented ancestry represent the best candidates for this variant class, as variant scarcity alone is insufficient for determining how recently the variant arose and the likely demographic forces it was subjected to.

Canine ancestry representation was determined using individual singleton counts, where excess singletons per individual indicated either underrepresented ancestry or poor sequencing quality. Seventy percent of dogs carried <40 singletons each within their coding sequences, allowing this value to serve as a cutoff (**Supplemental Fig. S4A**). These dogs belonged to breeds with the greatest number of individual members in the dataset (greater Wilcoxon rank sum test; P-value < 10^{-15}), and larger fractions of their genomes were within runs of homozygosity (F_{ROH}) (greater Wilcoxon rank sum test; P-value < 10^{-15}), than was observed for individuals with >40 singletons (**Supplemental Fig. S4B**). Only 11.2% of all singletons were inferred to be recent spontaneous mutations by this method, and most low-frequency variants were contributed by a minority of individuals with poorly represented ancestry (**Supplemental**

Downloaded from genome.cshlp.org on March 2, 2026 - Published by Cold Spring Harbor Laboratory Press

Fig. S4C), mostly consisting of wolves, village dogs, and breeds of eastern origin, which have greater diversity than western dogs. (**Supplemental Fig. S2**) (**Supplemental Table S3**).

To evaluate the fit of the mutation model, we used the Pearson correlation coefficient between observed frequency of synonymous mutations and synonymous mutation probability per gene. From a total of 8,660 observations, the model achieved a modest correlation of 0.53, which was lower than anticipated and likely due to an insufficient number of observations. Sparsity in observations was addressed first by relaxing criteria for defining recent spontaneous mutations and, second, by grouping genes based on mutation probabilities. The 95% confidence intervals for expected fit were also calculated, allowing the mutation model to be evaluated as a departure from expected fit values. Although relaxing criteria led to increased correlations, the departure from expected model fit increased (**Fig. 1A**), as inclusion of additional variants to the analysis diluted the fraction of true recent spontaneous mutations. The alternative strategy of placing genes into groups based on mutation probability was more successful. For 1000 groups or less, correlations greater than 0.9 were reached and the model fit remained within expected ranges, unlike a competing model based on CDS length (**Fig. 1B**). These analyses indicate that accumulation of synonymous mutations follows expected mutation probabilities, demonstrating that this framework can be used to determine if genes have significantly fewer nonsynonymous mutations than expected, thus detecting negative selection effects.

The functional constraint of dog genes

Using 907 LoF and 14,947 missense singleton mutations, we tested whether constraint metrics were associated with negative selection effects (Methods). Gene-based constraint metrics, i.e., constrained CDS and human ortholog LOEUF, were significantly associated with intolerance to LoF mutations (**Fig. 2C**), indicating that a significant portion of the expected LoF mutations in the most constrained genes were deleterious. The strongest effects were observed in the top

Downloaded from genome.cshlp.org on March 2, 2026 - Published by Cold Spring Harbor Laboratory Press

human ortholog LOEUF decile, where only half the expected number of LoF mutations were observed (**Fig. 1C**). Conversely, no significant negative selection effects were observed for LoF mutations among the top phyloP constraint deciles, indicating the importance of sequence context when evaluating the fitness impacts of evolutionarily conserved sites. For example, most reference genomes used in multi-species alignments do not contain LoF variants, whereas population analyses show they are often tolerated (Lek et al. 2016).

All three constraint metrics were associated with intolerance to missense mutations, however selection effects for gene-based metrics were less pronounced than those observed for LoF mutations (**Fig. 1C**). By comparison, synonymous mutations demonstrated no significant negative selection effects amongst the top gene-based constraint deciles, as expected. Conversely, there was significant intolerance to synonymous substitutions for the top phyloP score deciles. Because such scores are based on genomic positions, this is likely due to synonymous mutations causing cryptic-splice sites, impacting mRNA stability, or inducing change in functionally relevant GC content, e.g., methylation sites (Sarkar et al. 2022; Gudkov et al. 2024).

We next tested whether constraint metrics were associated with known canine trait variants from the online mendelian inheritance in animals (OMIA) database (**Supplemental Table S5**), a catalogue of variants and genes associated with inherited disorders and traits in non-model organism species (Nicholas 2021; Nicholas 2024) (https://www.omia.org/home/). We incrementally increased the minimum threshold for defining constraint and measured enrichment of trait-associations for the remaining variants. To ensure these variants were segregating in dogs and did not represent sporadic cases, we only focused on OMIA variants with an allele count >2 within the combined cohort. Altogether, 31 out of 305 LoF and 50 out of 154 missense OMIA variants met these criteria (**Supplemental Table S5**). Importantly, supporting evidence for trait-associations varies widely throughout OMIA, ranging from candidate alleles found in a single individual, to mutations shown to be causative in functional

Downloaded from genome.cshlp.org on March 2, 2026 - Published by Cold Spring Harbor Laboratory Press

studies. Although, we detected instances of LoF and missense OMIA variant enrichment for evolutionary-based constraint metrics (**Supplemental Fig. S5**), increases to the minimum constraint threshold often led to a loss of OMIA enrichment (**Supplemental Table S6)**. Similarly, no enrichment of LoF OMIA variants were detected for human ortholog LOEUF constraint, indicating most OMIA LoF variants segregating within the combined cohort are likely not severely deleterious.

Detection of new genotype-phenotype associations using gene constraint.

We used constraint to investigate genotypes for their potential to impact canine breed phenotypes, focusing specifically on breed enriched genotypes. Analysis of variant breed specificity showed that single-breed enrichment of OMIA variants was highest at a minimum breed allele frequency of 20% (**Supplemental Fig. S6**). Therefore, we only considered variants with breed allele frequencies above this threshold in at least one breed. To counteract the overrepresentation of breeds with small population sizes and common variants, we also ensured that variants had a minimum allele count of three, a minimum breed allele number of six, and a maximum non-breed allele frequency of 5%.

For each constraint metric, potential LoF and missense candidates for canine traits were identified and ranked according to the number of breeds the variant was found in and the proportion of individuals in those breeds carrying the variant (**Supplemental Fig. S7**) (**Supplemental Table S7-S12**). Four variants with potential to impact breed-characteristic phenotypes were selected for further discussion (**Table 1**). These included a splice-acceptor mutation in *PDGFRA* in Çatalburun dogs as a candidate for bifid nose and orofacial clefts, a stop-gained mutation in *CSF1 fixed* in Chow Chows (Chow) as a candidate for their distinctive skull shape, a frame-shift in *LCORL* segregating in large breeds (>25 kg), and a frameshift in *ITGA11* fixed in small brachycephalic breeds.

Downloaded from genome.cshlp.org on March 2, 2026 - Published by Cold Spring Harbor Laboratory Press

Table 1: Breed genotypes and allele frequencies of select candidate variants.

Variant	Breed	AF[2]	Ref[3]	Het[4]	Hom[5]
Chr13:47440160, A/C, *PDGFRA*, splice-acceptor	Çatalburun	0.5	0	3	0
Chr6:42492671, C/A, *CSF1*, stop-gained	Chow Chow	1	0	0	6
	Elo	0.1	4	1	0
Chr3:91871030, AACAG/A, *LCORL*, frameshift variant[1]	Spanish Mastiff	1	0	0	6
	Berger Picard	0.77	0	5	6
	Briard	0.67	0	4	2
	Petit Bleu de Gascogne	0.58	1	3	2
	Bluetick Coonhound	0.5	0	2	0
	Bracco Italiano	0.5	3	0	3
	Spinone Italiano	0.43	2	4	1
	Pyrenean Mastiff	0.36	3	3	1
	Leonberger	0.32	9	8	2
	Beauceron	0.3	2	3	0
	Mastiff	0.28	5	3	1
	German Hound	0.25	3	3	0
	Griffon Nivernais	0.25	2	2	0
	Ariege Pointer	0.2	3	2	0
	Karakachan	0.2	3	2	0
	Picardy Spaniel	0.2	3	2	0
Chr30:32785671, ACT/A, *ITGA11*, frameshift variant[1]	Pug	1	0	0	3
	Japanese Chin	1	0	0	4
	Pekingese	1	0	0	4
	Shih Tzu	1	0	0	4
	Tibetan Spaniel	1	0	0	4
	Boston Terrier	1	0	0	6
	French Bulldog	1	0	0	6
	Lhasa Apso	1	0	0	6
	Chihuahua	0.36	4	1	2
	Biewer Terrier	0.3	2	3	0
	Portuguese Podengo Pequeno	0.25	1	1	0
	Russian Tsvetnaya Bolonka	0.25	3	3	0
	Chesapeake Bay Retriever	0.21	4	3	0

[1]Only breeds with AF ≥0.2 are displayed.
[2]Frequncy of the alternative allele within the breed.
[3]Number of individuals homozygous for the reference allele.
[4]Number of individuals with heterozygous genotype.
[5]Number of individuals homozygous for the alternate allele.

The *PDGFRA* splice-acceptor was heterozygous in all three whole genome sequenced Çatalburun dogs, a breed characterized by a split/bifid nose phenotype (**Fig 2A**) (Yilmaz et al. 2012; Özarslan et al. 2021). Mouse knockout models demonstrate that *PDGFRA* is essential for medial nasal process development and palatogenesis (He and Soriano 2013; Qian et al. 2017),

Downloaded from genome.cshlp.org on March 2, 2026 - Published by Cold Spring Harbor Laboratory Press

aligning with the hypothesis that the *PDGFRA* splice-acceptor variant causes the Çatalburun's bifid nose phenotype. The variant was genotyped in fourteen additional Çatalburun dogs. Although phenotypes were not available, thirteen dogs were heterozygotes and one was homozygous for the reference allele (**Supplemental Table S13**).

Eight additional bifid nose dogs, all lacking significant Çatalburun ancestry were recruited and underwent whole genome sequencing and candidate variant filtering (**Supplemental Fig. S8-S9**) (**Supplemental Methods**). Six of the dogs were unrelated to one another (**Supplemental Table S14**), while two were littermates and were analyzed in combination with an unaffected parent (**Supplemental Table S15**). Twelve heterozygous candidate variants were identified across six of the non-Çatalburun bifid nose dogs, either in evolutionarily constrained or orofacial cleft associated genes (**Supplemental Table S16**). One sample, 40931 (**Fig. 2B**), carried a heterozygous insertion that also disrupted a *PDGFRA* splice-acceptor and was the only dog other than the Çatalburun's to have a LoF variant in this gene (**Fig. 2C**), further supporting the gene as a candidate for orofacial clefts.

An additional breed, the Pachón Navarro, which originates in Spain, is also known for its bifid nose phenotype. Unfortunately, no samples from this breed were available for study. However, masonry on the outside of the 18th century Palace of Bertemati (Andalusia, Spain) shows a gargoyle-like dog with a bifid nose (Cobo 2017), attesting to the continuous selection for this trait in the Pachón Navarro breed (**Fig. 2D**).

Phylogenetic clustering of the Çatalburun haplotypes surrounding *PDGFRA* indicate a high degree of heterozygosity at this locus. Despite this, the haplotypes carrying the mutant allele clustered together amongst Spaniel and Setter haplotypes, appearing almost identical except for the splice-acceptor mutation (**Fig. 2E**). The depiction of a bifid nose dog in an 18th century Spanish statue, combined with observed haplotype similarity with Spaniels and Setters, indicates a Western European/Iberian origin for the mutation.

Downloaded from genome.cshlp.org on March 2, 2026 - Published by Cold Spring Harbor Laboratory Press

The identified frameshift in *LCORL* (c.1978_1981delACAG) causes a premature stop in the long isoform of the gene, much like the previously identified c.3768_2769insA variant (Plassais et al. 2019) (**Fig. 3A**). Importantly, the newly identified c.1978_1981delACAG variant is not detectable in GWASs for either breed standard height or weight, indicating that constraint-based approaches can identify otherwise overlooked trait-associated candidate variants (**Fig 3B**) (**Supplemental Table S17**). Association analysis of the *LCORL* loci using GEMMA also identified a third previously undescribed intronic variant common in small breeds (<40 cm in height and weighing <10 kg). This variant, c.525-14A>T, had a more significant association with breed standard height than c.3768_2769insA and was also associated with breed standard weight (**Supplemental Fig. S10**) (Zhou and Stephens 2012). All three variants were in close proximity but were not in linkage disequilibrium (LD) (**Fig. 3C**) (**Supplemental Fig. S11**). Conversely, since c.1978_1981delACAG has a similar impact on protein structure as c.3768_2769insA, it is found primarily in large breeds (>60 cm and >25 kg in height and weight, respectively) (**Fig. 3D**) (**Supplemental Fig. S11**), such as the Spanish mastiff, where the variant is fixed (**Table 1**). Haplotype analysis of the region showed an ancestral haplotype shared between dogs and wolves and a modern haplotype that is exclusive to dogs (**Fig. 4**). The large body size alleles are embedded within each of these haplotypes separately, with the c.1978_1981delACAG variant belonging to the ancestral haplotype and the c.3768_2769insA variant to the modern haplotype. Despite this, different combinations of the three variants segregate within multiple breeds, indicating that selection for these size alleles likely predates breed development (**Supplemental Table S18**). There are also instances of compound heterozygotes, further supporting this hypothesis (**Supplemental Fig. S11**).

Selection for breed traits is associated with constrained genes and not coding sequence variation.

Downloaded from genome.cshlp.org on March 2, 2026 - Published by Cold Spring Harbor Laboratory Press

Artificial selection for characteristic breed traits in dogs is associated with extended regions of homozygosity and strong signatures of population differentiation between breeds (Akey et al. 2010; Vaysse et al. 2011; Parker et al. 2017a; Morrill et al. 2022) (Boyko et al. 2010; Kim et al. 2018; Plassais et al. 2019). However, due to the large size of these regions it is difficult to identify specific variants that contribute to breed characteristic traits. To enhance detection of trait-associations, potentially functional variants in constrained genes were analyzed for their association with signatures of breed differentiation.

Breed-differentiated regions were defined as 50 kb genomic windows of extreme F_{ST} that overlapped 1Mb regions of breed homozygosity (**Methods**) (**Fig. 5A**). A total of 1,743 were identified, representing 2,085 genes (**Supplemental Table S19**). Importantly, 92.3% and 55.2% of breed-differentiated regions were within 1Mb and 100kb of each other, respectively, indicating that most breed-differentiated regions are likely part of larger selective sweeps. In addition, 17 of 25 autosomal trait-associated loci described by Plassais et al. (2019) are within 100kb of a breed-differentiated region, indicating genomic signatures for breed differentiation are associated with trait selection (**Fig. 5B**).

There were between 10 and 34 genes across all breed-differentiated regions that contained LoF mutations with potential functional consequences (**Fig. 5C**). Of these, only two were found in breeds that exhibited signatures of differentiation at these loci, specifically the *CSF1* premature stop codon found in Chows (**Table 1**), and the *SGK3* mutation causing hairlessness in American Hairless Terriers (Parker et al. 2017b). Although not significantly associated with breed differentiation, there were 115 variants within these regions that segregated in breeds with long stretches of homozygosity (**Supplemental Table S19**), indicating these potentially functional variants may be in LD with traits under selection.

Altogether, 70 genes contained potentially impactful missense mutations significantly associated with breed differentiation (**Fig. 5C**). The strongest result was for a missense variant in *RYR1* (**Supplemental Table S20**), a calcium ion channel primarily expressed in the

Downloaded from genome.cshlp.org on March 2, 2026 - Published by Cold Spring Harbor Laboratory Press

sarcoplasmic reticulum of skeletal muscle and involved with muscle contraction (Hernandez-Ochoa et al. 2015). The variant was almost exclusively found within Bull Terriers and American Staffordshire Terriers, two breeds with shared ancestry known for their musculature and agility.

Most breed-differentiated regions were not associated with nonsynonymous variants, indicating that breed selection may instead be acting upon noncoding variants, which is consistent with most mapping studies for breed-defining traits and breed/lineage differentiation (Dutrow et al. 2022; Morrill et al. 2022). Human ortholog LOEUF constraint was significantly overrepresented among genes within breed-differentiated regions (**Fig 5D**), suggesting that phenotypic breed differentiation is primarily driven by noncoding variation surrounding and likely regulating highly constrained genes. Further investigation is required to identify trait-associated noncoding variants and candidate gene regulatory regions.

Discussion

Artificial selection for desired traits has made the domestic dog one of the most phenotypically diverse mammals in the world (Wayne and vonHoldt 2012). Rigid population structure, enforced by breed standards, has promoted high levels of homozygosity within breeds while also maintaining isolation between breeds (Parker et al. 2004; Karlsson and Lindblad-Toh 2008). Together, these factors make the domestic dog an invaluable system for identifying genes related to breed characteristics, which are often morphological extremes (Ostrander et al. 2017; Buckley and Ostrander 2024). We therefore investigated the genetic basis of canine traits by using genomic constraint to identify variants with potential phenotypic impacts. This work provides significant advances to the field by: 1) defining a variant set whose presence in the genome is unaffected by canine population dynamics, thus making it possible to determine genes under negative selection; 2) evaluating three constraint metrics for prioritizing variants for potential phenotypic impacts; 3) identifying new variants with strong impacts on craniofacial

Downloaded from genome.cshlp.org on March 2, 2026 - Published by Cold Spring Harbor Laboratory Press

morphology and body size; and 4) applying genomic constraint to support a model of breed differentiation based on noncoding variation.

Using gene constraint, we identified strong candidates for phenotypic impacts as previously unreported LoF variants in *PDGFRA* and *LCORL*. Two *PDGFRA* variants were identified in dogs with bifid noses. This gene may therefore be one of a few for which functional mutations affecting midline closure occur, but are tolerated, as long as they are in the heterozygous state. In humans, a missense variant in this gene within a single family is associated with left non-syndromic cleft-lip and palate but is not fully penetrant (Yu et al. 2022). PDGFRA is a type III receptor tyrosine kinase that consists of an extracellular region with five immunoglobulin-like domains for ligand binding, a transmembrane helix, and an intracellular tyrosine kinase domain (Williams 1989; Chen et al. 2013). The Çatalburun variant occurs in the seventh intron of the gene within the fourth immunoglobulin-like domain, while the other variant in dog 40931 is further downstream within the protein kinase domain (**Fig 4C**). Both mutations are splice-acceptors, making it difficult to determine their impact on transcription. The affected exons may be skipped entirely, or an alternative splice site may be used, resulting in either partial exon retention or retention of nearby intronic sequence. Dogs carrying the *PDGFRA* mutations had no reported health concerns, which is surprising as *PDGFRA* is a broadly expressed developmental gene and strongly associated with gastrointestinal stromal tumor-plus syndrome in humans, with onset most likely to occur between 40 and 70 years of age (Manley et al. 2018; Huang et al. 2022).

Two other breeds share a similar phenotype as the Çatalburun, one is the Andean tiger hound, about which little is known, and the other is the Spanish Pachón Navarro. The Pachón Navarro, like the Çatalburun, is a pointing breed with only some members displaying the characteristic bifid nose phenotype. However, because Pachón Navarro dogs are susceptible to cleft palate, they have been used as surgical animal models for palate repair (Paradas-Lara et al. 2014). Evidence for bifid nose dogs in Spain dates back at least four hundred years, as the

Downloaded from genome.cshlp.org on March 2, 2026 - Published by Cold Spring Harbor Laboratory Press

trait is depicted in a 1623 painting of a hunting party titled “Diane et ses nymphes s'apprêtant à partir pour la chasse” (Diana and her nymphs getting ready to go hunting) by Pierre Paul Rubens and Jan Brueghel l'Ancien. There are also descriptions of bifid nose pointing dogs in the south of France that date back to the turn of the 20th century (Arkwright 1906). Although it is unknown whether this phenotype in the Çatalburun and Pachón Navarro breeds shares a common origin, Türkiye and Spain were both under the rule of the Umayyad Caliphate in the 8th century CE, providing a possible route for cultural exchange between the two regions. In addition to the recurrent literary and artistic depictions of bifid nose dogs throughout Spain and France, the haplotype similarity between Çatalburun dogs and Setters and Spaniels, whom have Spanish origins, indicates the phenotype likely arose in Western Europe or the Iberian Peninsula.

The newly discovered *LCORL* c.1978_1981delACAG allele was identified because of its high frequency in Spanish Mastiffs, Berger Picards, and Briards (**Table 1**). Like the previously described *LCORL* c.3768_2769insA variant, the c.1978_1981delACAG allele causes a frameshift and premature stop and loss of the DUF4553 DNA-binding domain from the long isoform of the protein, contributing to increased stature and weight (Plassais et al. 2019). We also identified a third variant in the gene that corresponded to smaller body size, indicating recurrent selection at this locus. Importantly, the *LCORL* locus is associated with stature across many domesticated mammals including goats (Saif et al. 2020; Graber et al. 2022), cattle (Bouwman et al. 2018; Chen et al. 2020), pigs (Rubin et al. 2012), and horses (Metzger et al. 2013; Tetens et al. 2013). As with dogs, large body size in most of these species is caused by truncation of the long isoform of the LCORL protein (Bai et al. 2025). Moreover, two additional candidate size variants upstream of *LCORL* have also been described based on size differences between Bull Terriers and Miniature Bull Terriers (Wade 2025), further indicating the complexity of the *LCORL* locus for controlling body size in dogs.

Downloaded from genome.cshlp.org on March 2, 2026 - Published by Cold Spring Harbor Laboratory Press

An additional variant revealed by this study that may correspond to a breed trait is the stop-gain at position p.Glu391* in CSF1 in Chows. The variant was homozygous in all six unrelated Chows, heterozygous in one of five Elos, and absent from two Eurasiers, the latter two breeds being the product of Chow admixture (**Table 1**), thus indicating a high degree of breed specificity. The Chow is a medium-sized spitz-type dog that traces its origins to China, and is characterized by its thick double coat, black pigmented tongue, straight hind legs, broad square skull and muzzle, and supernumerary third incisors (Yang et al. 2017). The genomic region containing *CSF1* is under strong positive selection in Chows (Yang et al. 2017), suggesting genes in this region contribute to the breed's characteristic phenotypes. The p.Glu391* mutation is located in exon six of the secreted isoform of the CSF1 protein (Cosman et al. 1988). Rodent knockout models implicate *CSF1* in skeletal and craniofacial development, specifically a domed skull, missing incisors, short thickened limb bones, fertility, and macrophage differentiation (Marks and Lane 1976; Yoshida et al. 1990; Dai et al. 2002; Van Wesenbeeck et al. 2002; Sehgal et al. 2021). While the precise function of the p.Glu391* mutation in dogs is unknown, the rodent phenotypes of a domed skull and altered number of incisors, suggest p.Glu391* may contribute to the Chow's broad square skull and two additional incisors, which contrast sharply with the Eurasier and Elo breed requirements. Similarly, the identified frameshift within *ITGA11* may also correspond to craniofacial structure (**Table 1**). The variant is fixed in small brachycephalic breeds, and mouse knockouts of the gene demonstrate its association with small size and craniofacial abnormalities, specifically the eruption of incisors (Popova et al. 2007). Consistent with this, small brachycephalic breeds (<5 kg) have delayed tooth eruption times compared to other breed sizes and skull types (Van den Broeck et al. 2022). Also, associations between the locus and brachycephaly have been described previously but were sensitive to corrections for canine size (Schoenebeck et al. 2012; Marchant et al. 2017). Thus, the variant could contribute to a range of phenotypes, including size, craniofacial structure, and development; however, functional analyses are needed.

Downloaded from genome.cshlp.org on March 2, 2026 - Published by Cold Spring Harbor Laboratory Press

One of the defining features of the canine system is that often few loci contribute to large phenotypic variation, particularly for morphological traits like breed standard weight and height or ear shape and position (Rimbault et al. 2013; Plassais et al. 2019), while in the human genome there are often large numbers of loci contributing to complex phenotypes (Yengo et al. 2022). Loci contributing to extreme phenotypic differences between breeds exhibit strong signatures of positive selection because of artificial selection for desired traits (Plassais et al. 2019; Morrill et al. 2022; Plassais et al. 2022). Our analysis shows that highly constrained genes are overrepresented in such regions and are therefore the likely targets of trait selection, driving phenotypic variation. This suggests an explanation for why CDS variation is rarely associated with morphological differences between breeds and more often associated with disease. Constrained genes are often essential for development and survival and are frequently implicated in highly penetrant genetic disorders and syndromes (Lek et al. 2016; Karczewski et al. 2020; Sun et al. 2024; Zeng et al. 2024). Trait selection in dogs is therefore a balancing act between preserving essential gene functions and maximizing regulatory variation of essential genes to drive phenotypic extremes. Future work should thus focus on assigning noncoding variants under positive selection to their target genes.

Central to the analysis is the identification of variants with phenotypic impacts, which was assessed according to evolutionary/functional constraint acting on genes/sites containing the variant in question. However, specific factors impact the interpretation of these results. For instance, the measured selection effects only capture negative selection acting on heterozygous variants with highly deleterious impacts on fitness (Fuller et al. 2019). Our analysis is therefore blind to recessive deleterious alleles and genes associated with late age of onset conditions such as breed-prevalent diseases like cancer. Since singletons are the input to the mutation model, depletion of variants with little to no deleterious impact in a heterozygous state remain unobservable, as do variants that produce no phenotype until after a dog's reproductive years have elapsed. This is relevant to canines where a high degree of inbreeding makes dogs

Downloaded from genome.cshlp.org on March 2, 2026 - Published by Cold Spring Harbor Laboratory Press

especially susceptible to genetic load and recessive disease allele burden (Marsden et al. 2016; Mooney et al. 2021; Donner et al. 2023). Moreover, OMIA variants segregating in canine populations likely persist because they do not have strong effects on fitness in a heterozygous state. In fact, many of the OMIA variants underlie traits like coat color or recessive and late-onset diseases, such as the many forms of progressive retinal atrophy (Miyadera et al. 2012; Downs et al. 2014; Bunel et al. 2019). Therefore, strong human ortholog LOEUF constraint lacks an association with OMIA variants because OMIA LoF variants have a low impact on fitness, and human ortholog LOEUF constraint indicates genes where heterozygous LoF mutations have a strong deleterious impact on fitness.

Another factor impacting interpretation of results is whether the unit for measuring constraint is gene or base-pair. The increased resolution provided by site-specific phyloP scores is only beneficial for analyzing missense mutations, which tend to have localized effects and primarily negligible impacts on biology unless interrupting important functional domains (Gerasimavicius et al. 2022). Conversely, LoF mutations, defined throughout as premature stop codons, frameshifts, start-losses, and splice-site disruptions, impact the entire downstream transcript. Therefore, in dogs, human ortholog LOEUF is the best of the three constraint metrics for identifying LoF mutations associated with phenotypes with severe impacts on fitness while variant phyloP scores are best for prioritizing missense mutations as having potential phenotypic impacts.

The dog genome project began with two goals and both remain a high priority today: improve the health of canine companions through informed genetics and translate those findings to advance human health. While many breeds feature largely healthy, long-lived animals, others are characterized by an excess of disease alleles and associated health problems (Axelsson et al. 2021). These breeds offer a unique opportunity to make use of the approaches and data provided here to identify underlying genetic flaws, particularly for diseases of relevance to human health. Of equal priority are studies of morphologic and behavioral

Downloaded from genome.cshlp.org on March 2, 2026 - Published by Cold Spring Harbor Laboratory Press

genetics, as identification of alleles that control extreme phenotypes associated with these traits can improve our understanding of human growth and development.

Methods

Defining a canine mutation model and callable gene set

We built a mutation model in dogs following the methods described in (Samocha et al. 2014). Briefly, trinucleotide substitution rates, normalized according to the generation mutation rate and trinucleotide frequencies, were applied to each possible substitution for each site within gene models. For each gene model, substitution probabilities were summed according to their consequences on coding sequence, such as synonymous, missense, stop gain, stop loss, start loss, etcetera. The result was per gene probabilities for each mutation type per generation (**Supplemental Fig. S12**). Intronic and intergenic Dog10K variants at least 20 bp from an exon boundary and within callable loci were used to calculate trinucleotide substation rates (Meadows et al. 2023). A canine generation mutation rate of 9.02×10^{-9} was used for normalization (Bergeron et al. 2023) (**Supplemental Fig. S13**). We calculated transcript-specific mutation probabilities using NCBI gene annotations from Dog10K, only considering sites that were callable loci, obviating the need to normalize probabilities by mean exon coverage (**Supplemental Table S21**). There was insufficient variant data to adjust mutation probabilities by evolutionary conservation as was performed in (Samocha et al. 2014). We classified genes according to intact start and stop codons, low-quality annotation, and partial annotation. Also, for each gene, a representative transcript was chosen based on reciprocal best hit BLAST results with protein sequences of human representative transcripts. For genes with no reciprocal best hit, we used the isoform annotated as “transcript variant X1” as the representative transcript. BLAST hits were also defined as dog-human orthologs and used to assign human LOEUF constraint metrics to dog genes. For phyloP constraint, we used the Zoonomia values mapped to UU_Cfam_GSD_1.0 as part of Dog10K, which were further summarized according to gene

Downloaded from genome.cshlp.org on March 2, 2026 - Published by Cold Spring Harbor Laboratory Press

models (**Supplemental Table S22**). Expected mutations per gene were calculated by sampling *n* gene names with replacement weighted by gene/gene group mutation probabilities, where *n* is the number of observations the expected rate is being compared to. Sampled gene names were tabulated to determine the expected per gene mutation count. This process was replicated 10,000 times to generate 95% expected ranges.

Preparing genotypes for analysis

Genotyping and filtering procedures are described in full in the supplemental methods. Briefly, a VCF comprised of SRA genomes was constructed using GATK to complement the publicly available Dog10K VCF (Poplin et al. 2017). All WGS data was mapped to the UU_Cfam_GSD_1.0 reference assembly (Wang et al. 2021). For each VCF, samples were filtered according to multiple genotype metrics (**Supplemental Table S23**), which included coverage statistics (**Supplemental Fig. S14**) and heterozygous allele balance (**Supplemental Fig. S15**) in addition to others (**Supplemental Fig. S16**). A total of 1,146 SRA and 1,778 Dog10K samples remained (**Supplemental Table S24**). Duplicates and first-degree relatives were removed from both datasets (**Supplemental Fig. S17**), leaving a total of 2,291 samples in the final cohort (**Supplemental Table S25**). Sites were filtered according to variant quality score recalibration (VQSR) using a common training set (**Supplemental Fig. S18**). Autosome and Chromosome X PAR sites (**Supplemental Table S26**) were analyzed separately from Chromosome X non-PAR sites (**Supplemental Table S27**). Sites that passed VQSR filtering in both datasets were used to create a merged VCF for population-scale analyses. Sites that passed filtering in either dataset and were within the CDS underwent further filtering (**Supplemental Fig. S19**) (**Supplemental Table S28**). Criteria for filtering CDS genotypes depended on chromosome, sex, and genotype (**Supplemental Table S29**). CDS indels and SNVs were reclassified and atomized in both datasets to improve harmonization (**Supplemental Fig. S20**) (**Supplemental Table S30**). Over 95% of SNVs and 80% of indels

Downloaded from genome.cshlp.org on March 2, 2026 - Published by Cold Spring Harbor Laboratory Press

passed filtering (**Supplemental Table S31**). Genotype quality was the criteria that failed most often (**Supplemental Table S32**). Sample genotype pass rates were higher for SNVs than indels (**Supplemental Table S33**). CDS filtering removed genotyping biases between each dataset for SNVs (**Supplemental Fig. S21**) and indels (**Supplemental Fig. S22**). Genotypes were merged and annotated with variant effect predictor (VEP) (McLaren et al. 2016). All VCF-based operations were performed with BCFtools (Danecek et al. 2021), PLINK (Purcell et al. 2007), VCFtools (Danecek et al. 2011), and vcflib (Garrison et al. 2022). LOEUF constraint values were obtained from gnomAD v4.1.0 and are based on expected rates of LoF variation from a combined set of over 800,000 human exomes and genomes (Chen et al. 2024). Zoonomia canine phyloP scores were obtained from Dog10K (Meadows et al. 2023). Details regarding sequencing of additional bifid nose dogs are presented in the Supplemental Methods.

Defining breeds and groups

Illumina CanineHD array genotypes from 2,156 samples were lifted over to the UU_Cfam_GSD_1.0 reference genome using LiftoverVCF (**Supplemental Table S34**). WGS sites were downsampled to match array positions and datasets were merged. After filtering at MAF > 0.01, genotyping rate > 0.9, and LD < 0.5, a total of 76,570 sites remained (Purcell et al. 2007). A consensus neighbor joining tree was built using PHYLIP with a Golden Jackal as the outgroup (Felsenstein 1993). Branch support was calculated from 10% jackknifing 10,000 times. Branches needed at least 50% support. Breed/group clusters were determined by optimizing for shared breed membership on the same branch. Samples that did not match the majority breed on a branch were assigned new breed membership, matching the majority breed. Branches that consisted of overlapping breeds were merged. Breeds that could not form a branch of at least three members were placed in the unassigned category.

Canine population statistics, breed differentiated loci, and haplotype structure

Downloaded from genome.cshlp.org on March 2, 2026 - Published by Cold Spring Harbor Laboratory Press

Population analyses were only performed on the 200 breeds/groups that had five or more members. Genome-wide population statistics were calculated across 200,000 randomly sampled sites. Nucleotide diversity within breeds was measured as pi, the average distance between two randomly sampled sequences in a breed (Nei and Li 1979). Breed differentiation was measured as F_{ST}, using the Weir and Cockerham estimator on a per-site basis (Weir and Cockerham 1984). F_{ROH} was calculated using the same procedures as specified in Meadows et al. (2023). We detected breed-differentiated regions by calculating F_{ST} between all 200 pairs of breeds with five or more members across 50kb windows using a 10kb step size. Windows with F_{ST} values > 0.8, which corresponded to ~0.1% of all pairwise comparisons, were extracted for further analysis. The density of high F_{ST} windows was then measured across the genome and the "peaks" function of IDPmisc, a peak-calling software R package, was used to define breed-differentiated regions (Locher 2024). For each region we extracted breeds with pairwise F_{ST} values > 0.8 within the region's peak window. Extracted breed pairs with F_{ST} < 0.05 were defined as sharing many of the same alleles and were assigned the same cluster ID. We also defined genomic segments of extended breed homozygosity if half of the 50kb windows along a megabase segment had pi values < 0.02. Breed-differentiated regions were defined as selection loci in individuals where there was overlap with extended homozygosity regions. All haplotype analyses were performed using the phased reference panel from Dog10K (Meadows et al. 2023). Only variants within 20kb upstream and downstream of the target variant and have an allele frequency of > 1% were included. Trees were built using the "ape" v5.8 R package (Paradis and Schliep 2019).

Software availability

Computer code generated in this analysis is deposited at https://github.com/ReubenBuck/constraint_dogs and within the Supplemental Code.

Downloaded from genome.cshlp.org on March 2, 2026 - Published by Cold Spring Harbor Laboratory Press

Data access

All whole genome sequencing data generated in this study has been submitted to the NCBI BioProject database (https://www.ncbi.nlm.nih.gov/bioproject/) under the accession numbers PRJNA726547 and PRJNA1185472. All genotyping array data generated in this study has been submitted to the Gene Expression Omnibus (GEO; https://www.ncbi.nlm.nih.gov/geo/) under accession number GSE278600. Individual accessions for all samples used in this study are listed in Supplemental Table S1 and Supplemental Table S34. Both tables indicate all newly released datasets associated with this study.

Acknowledgements

We thank Tatiana Feuerborn for her helpful advice and insight throughout the project and Gabby Spatola for assistance in recruiting samples. We also thank Zachary Lounsberry and Embark Veterinary for assistance in recruiting samples and Sruthi Hundi for bioinformatic assistance. For careful reading of the manuscript and providing helpful feedback, we also thank Christopher B. Kaelin and Wesley C. Warren. Most importantly, we thank all citizen scientists who donated samples for analysis, without whom this work would be impossible. EAO, RMB, ACH, DLD, ASA, and HGP are funded by the Intramural Program of the National Human Genome Research Institute at the National Institutes of Health (HG200377). HL is funded by the Jane and Aatos Erkko Foundation. The project was conceptualized by RMB and EAO. RMB and ASA were responsible for methodology. Software was written by RMB and ASA and validation experiments were performed by NB, FGVS, and MKH. RMB led the formal analysis and RMB, NB, MKH, HGP, and DLD performed the investigation. Resources were provided by NB, PS, CT, ME, FGVS, MKH, JH, HL, BCK, ARB, DLD, RMB, and EAO. RMB and ACH were responsible for data curation. Writing of the original draft was completed by RMB and EAO, while all authors contributed to review and editing. RMB created all visualizations and EAO was the primary supervisor. Project administration was carried out by RMB and EAO, while EAO was primarily

Downloaded from genome.cshlp.org on March 2, 2026 - Published by Cold Spring Harbor Laboratory Press

responsible for funding acquisition. All photos of individual pet dogs in Fig. 2B and Supplemental Fig. S9 are published with permission from their owners. NB is credited for the photo in Fig. 2A. We also thank Antonio Virgil Guzmán, who is credited with providing the photograph used in Fig. 2D.

Competing interest statement

A.R.B. is a co-founder and board member of Embark Veterinary, Inc., a private canine DNA testing company. All other authors declare no competing interests.

References

Adzhubei IA, Schmidt S, Peshkin L, Ramensky VE, Gerasimova A, Bork P, Kondrashov AS, Sunyaev SR. 2010. A method and server for predicting damaging missense mutations. *Nat Methods* **7**: 248-249.

Akey JM, Ruhe AL, Akey DT, Wong AK, Connelly CF, Madeoy J, Nicholas TJ, Neff MW. 2010. Tracking footprints of artificial selection in the dog genome. *Proc Natl Acad Sci U S A* **107**: 1160-1165.

All of Us Research Program Genomics I. 2024. Genomic data in the All of Us Research Program. *Nature* **627**: 340-346.

Andersson L, Purugganan M. 2022. Molecular genetic variation of animals and plants under domestication. *Proc Natl Acad Sci U S A* **119**: e2122150119.

Arkwright W. 1906. *The Pointer and His Predecessors: An Illustrated History of the Pointing Dog from the Earliest Times*. AL Humphreys.

Axelsson E, Ljungvall I, Bhoumik P, Conn LB, Muren E, Ohlsson A, Olsen LH, Engdahl K, Hagman R, Hanson J et al. 2021. The genetic consequences of dog breed formation-Accumulation of deleterious genetic variation and fixation of mutations associated with

Downloaded from genome.cshlp.org on March 2, 2026 - Published by Cold Spring Harbor Laboratory Press

myxomatous mitral valve disease in cavalier King Charles spaniels. *PLoS Genet* **17**: e1009726.

Bai F, Cai Y, Qiu M, Liang C, Pan L, Liu Y, Feng Y, Cao X, Yang Q, Ren G et al. 2025. *LCORL* and *STC2* variants increase body size and growth rate in cattle and other animals. *bioRxiv* doi:10.1101/2025.01.01.630985: 2025.2001.2001.630985.

Bergeron LA, Besenbacher S, Zheng J, Li P, Bertelsen MF, Quintard B, Hoffman JI, Li Z, St Leger J, Shao C et al. 2023. Evolution of the germline mutation rate across vertebrates. *Nature* **615**: 285-291.

Bouwman AC, Daetwyler HD, Chamberlain AJ, Ponce CH, Sargolzaei M, Schenkel FS, Sahana G, Govignon-Gion A, Boitard S, Dolezal M et al. 2018. Meta-analysis of genome-wide association studies for cattle stature identifies common genes that regulate body size in mammals. *Nat Genet* **50**: 362-367.

Boyko AR, Quignon P, Li L, Schoenebeck JJ, Degenhardt JD, Lohmueller KE, Zhao K, Brisbin A, Parker HG, vonHoldt BM et al. 2010. A simple genetic architecture underlies morphological variation in dogs. *PLoS Biol* **8**: e1000451.

Buckley RM, Ostrander EA. 2024. Large-scale genomic analysis of the domestic dog informs biological discovery. *Genome Res* doi:10.1101/gr.278569.123.

Bunel M, Chaudieu G, Hamel C, Lagoutte L, Manes G, Botherel N, Brabet P, Pilorge P, Andre C, Quignon P. 2019. Natural models for retinitis pigmentosa: progressive retinal atrophy in dog breeds. *Hum Genet* **138**: 441-453.

Bycroft C, Freeman C, Petkova D, Band G, Elliott LT, Sharp K, Motyer A, Vukcevic D, Delaneau O, O'Connell J et al. 2018. The UK Biobank resource with deep phenotyping and genomic data. *Nature* **562**: 203-209.

Chen PH, Chen X, He X. 2013. Platelet-derived growth factors and their receptors: structural and functional perspectives. *Biochim Biophys Acta* **1834**: 2176-2186.

Downloaded from genome.cshlp.org on March 2, 2026 - Published by Cold Spring Harbor Laboratory Press

Chen Q, Huang B, Zhan J, Wang J, Qu K, Zhang F, Shen J, Jia P, Ning Q, Zhang J et al. 2020. Whole-genome analyses identify loci and selective signals associated with body size in cattle. *J Anim Sci* **98**.

Chen S, Francioli LC, Goodrich JK, Collins RL, Kanai M, Wang Q, Alfoldi J, Watts NA, Vittal C, Gauthier LD et al. 2024. A genomic mutational constraint map using variation in 76,156 human genomes. *Nature* **625**: 92-100.

Cheng J, Novati G, Pan J, Bycroft C, Zemgulyte A, Applebaum T, Pritzel A, Wong LH, Zielinski M, Sargeant T et al. 2023. Accurate proteome-wide missense variant effect prediction with AlphaMissense. *Science* **381**: eadg7492.

Christmas MJ, Kaplow IM, Genereux DP, Dong MX, Hughes GM, Li X, Sullivan PF, Hindle AG, Andrews G, Armstrong JC et al. 2023. Evolutionary constraint and innovation across hundreds of placental mammals. *Science* **380**: eabn3943.

Cobo AA. 2017. El palacio de Bertemati. Iconografía y posible influencia en otros palacios de la zona. In *El Barroco: Universo de Experiencias*, pp. 227-248. Asociación Hurtado Izquierdo.

Cosman D, Wignall J, Anderson D, Tushinski J, Gallis B, Urdal D, Cerretti DP. 1988. Human macrophage colony stimulating factor (M-CSF): alternate RNA splicing generates three different proteins that are expressed on the cell surface and secreted. *Behring Inst Mitt*: 15-26.

Dai XM, Ryan GR, Hapel AJ, Dominguez MG, Russell RG, Kapp S, Sylvestre V, Stanley ER. 2002. Targeted disruption of the mouse colony-stimulating factor 1 receptor gene results in osteopetrosis, mononuclear phagocyte deficiency, increased primitive progenitor cell frequencies, and reproductive defects. *Blood* **99**: 111-120.

Danecek P, Auton A, Abecasis G, Albers CA, Banks E, DePristo MA, Handsaker RE, Lunter G, Marth GT, Sherry ST et al. 2011. The variant call format and VCFtools. *Bioinformatics* **27**: 2156-2158.

Downloaded from genome.cshlp.org on March 2, 2026 - Published by Cold Spring Harbor Laboratory Press

Danecek P, Bonfield JK, Liddle J, Marshall J, Ohan V, Pollard MO, Whitwham A, Keane T, McCarthy SA, Davies RM, Li H. 2021. Twelve years of SAMtools and BCFtools. *Gigascience* **10**.

Donner J, Freyer J, Davison S, Anderson H, Blades M, Honkanen L, Inman L, Brookhart-Knox CA, Louviere A, Forman OP, Chodroff Foran R. 2023. Genetic prevalence and clinical relevance of canine Mendelian disease variants in over one million dogs. *PLoS Genet* **19**: e1010651.

Downs LM, Hitti R, Pregnolato S, Mellersh CS. 2014. Genetic screening for PRA-associated mutations in multiple dog breeds shows that PRA is heterogeneous within and between breeds. *Vet Ophthalmol* **17**: 126-130.

Dutrow EV, Serpell JA, Ostrander EA. 2022. Domestic dog lineages reveal genetic drivers of behavioral diversification. *Cell* **185**: 4737-4755 e4718.

Felsenstein J. 1993. PHYLIP (phylogeny inference package), version 3.5 c. *Produced and distributed by author/Department of Genetics, University of Washington*.

Fogle B. 2000. *The new encyclopedia of the dog*. Dorling Kindersley Publishing, Inc., New York.

Fuller ZL, Berg JJ, Mostafavi H, Sella G, Przeworski M. 2019. Measuring intolerance to mutation in human genetics. *Nat Genet* **51**: 772-776.

Gao H, Hamp T, Ede J, Schraiber JG, McRae J, Singer-Berk M, Yang Y, Dietrich ASD, Fiziev PP, Kuderna LFK et al. 2023. The landscape of tolerated genetic variation in humans and primates. *Science* **380**: eabn8153.

Garrison E, Kronenberg ZN, Dawson ET, Pedersen BS, Prins P. 2022. A spectrum of free software tools for processing the VCF variant call format: vcflib, bio-vcf, cyvcf2, hts-nim and slivar. *PLoS Comput Biol* **18**: e1009123.

Gerasimavicius L, Livesey BJ, Marsh JA. 2022. Loss-of-function, gain-of-function and dominant-negative mutations have profoundly different effects on protein structure. *Nat Commun* **13**: 3895.

Downloaded from genome.cshlp.org on March 2, 2026 - Published by Cold Spring Harbor Laboratory Press

Gough A, Thomas A, O'Neill D. 2018. *Breed predispositions to disease in dogs and cats*. John Wiley & Sons.

Graber JK, Signer-Hasler H, Burren A, Drogemuller C. 2022. Evaluation of truncating variants in the LCORL gene in relation to body size of goats from Switzerland. *Anim Genet* **53**: 237-239.

Gudkov M, Thibaut L, Giannoulatou E. 2024. Quantifying negative selection on synonymous variants. *HGG Adv* **5**: 100262.

Havrilla JM, Pedersen BS, Layer RM, Quinlan AR. 2019. A map of constrained coding regions in the human genome. *Nat Genet* **51**: 88-95.

He F, Soriano P. 2013. A critical role for PDGFRalpha signaling in medial nasal process development. *PLoS Genet* **9**: e1003851.

Hernandez-Ochoa EO, Pratt SJP, Lovering RM, Schneider MF. 2015. Critical Role of Intracellular RyR1 Calcium Release Channels in Skeletal Muscle Function and Disease. *Front Physiol* **6**: 420.

Huang W, Yuan W, Ren L, Liang H, Du X, Sun X, Fang Y, Gao X, Fu M, Sun Y et al. 2022. Clinicopathological and therapeutic analysis of PDGFRA mutated gastrointestinal stromal tumor. *Pathol Res Pract* **239**: 154138.

Karczewski KJ, Francioli LC, Tiao G, Cummings BB, Alfoldi J, Wang Q, Collins RL, Laricchia KM, Ganna A, Birnbaum DP et al. 2020. The mutational constraint spectrum quantified from variation in 141,456 humans. *Nature* **581**: 434-443.

Karlsson EK, Lindblad-Toh K. 2008. Leader of the pack: gene mapping in dogs and other model organisms. *Nat Rev Genet* **9**: 713-725.

Kim J, Williams FJ, Dreger DL, Plassais J, Davis BW, Parker HG, Ostrander EA. 2018. Genetic selection of athletic success in sport-hunting dogs. *Proc Natl Acad Sci U S A* **115**: E7212-E7221.

Downloaded from genome.cshlp.org on March 2, 2026 - Published by Cold Spring Harbor Laboratory Press

Lampi S, Donner J, Anderson H, Pohjoismaki J. 2020. Variation in breeding practices and geographic isolation drive subpopulation differentiation, contributing to the loss of genetic diversity within dog breed lineages. *Canine Med Genet* **7**: 5.

Landrum MJ, Lee JM, Riley GR, Jang W, Rubinstein WS, Church DM, Maglott DR. 2014. ClinVar: public archive of relationships among sequence variation and human phenotype. *Nucleic Acids Res* **42**: D980-985.

Leeb T, Bannasch D, Schoenebeck JJ. 2023. Identification of Genetic Risk Factors for Monogenic and Complex Canine Diseases. *Annu Rev Anim Biosci* **11**: 183-205.

Lek M, Karczewski KJ, Minikel EV, Samocha KE, Banks E, Fennell T, O'Donnell-Luria AH, Ware JS, Hill AJ, Cummings BB et al. 2016. Analysis of protein-coding genetic variation in 60,706 humans. *Nature* **536**: 285-291.

Locher R. 2024. IDPmisc: Utilities of Institute of Data Analyses and Process Design (www.zhaw.ch/idp).

MacArthur DG, Balasubramanian S, Frankish A, Huang N, Morris J, Walter K, Jostins L, Habegger L, Pickrell JK, Montgomery SB et al. 2012. A systematic survey of loss-of-function variants in human protein-coding genes. *Science* **335**: 823-828.

Manley PN, Abu-Abed S, Kirsch R, Hawrysh A, Perrier N, Feilotter H, Pollett A, Riddell RH, Hookey L, Walia JS. 2018. Familial PDGFRA-mutation syndrome: somatic and gastrointestinal phenotype. *Hum Pathol* **76**: 52-57.

Marchant TW, Dietschi E, Rytz U, Schawalder P, Jagannathan V, Hadji Rasouliha S, Gurtner C, Waldvogel AS, Harrington RS, Drogemuller M et al. 2019. An ADAMTS3 missense variant is associated with Norwich Terrier upper airway syndrome. *PLoS Genet* **15**: e1008102.

Marchant TW, Johnson EJ, McTeir L, Johnson CI, Gow A, Liuti T, Kuehn D, Svenson K, Bermingham ML, Drogemuller M et al. 2017. Canine Brachycephaly Is Associated with a Retrotransposon-Mediated Missplicing of SMOC2. *Curr Biol* **27**: 1573-1584 e1576.

Downloaded from genome.cshlp.org on March 2, 2026 - Published by Cold Spring Harbor Laboratory Press

Marks SC, Jr., Lane PW. 1976. Osteopetrosis, a new recessive skeletal mutation on chromosome 12 of the mouse. *J Hered* **67**: 11-18.

Marsden CD, Ortega-Del Vecchyo D, O'Brien DP, Taylor JF, Ramirez O, Vila C, Marques-Bonet T, Schnabel RD, Wayne RK, Lohmueller KE. 2016. Bottlenecks and selective sweeps during domestication have increased deleterious genetic variation in dogs. *Proc Natl Acad Sci U S A* **113**: 152-157.

McLaren W, Gil L, Hunt SE, Riat HS, Ritchie GR, Thormann A, Flicek P, Cunningham F. 2016. The Ensembl Variant Effect Predictor. *Genome Biol* **17**: 122.

Meadows JRS, Kidd JM, Wang GD, Parker HG, Schall PZ, Bianchi M, Christmas MJ, Bougiouri K, Buckley RM, Hitte C et al. 2023. Genome sequencing of 2000 canids by the Dog10K consortium advances the understanding of demography, genome function and architecture. *Genome Biol* **24**: 187.

Mellanby RJ, Ogden R, Clements DN, French AT, Gow AG, Powell R, Corcoran B, Schoeman JP, Summers KM. 2013. Population structure and genetic heterogeneity in popular dog breeds in the UK. *Vet J* **196**: 92-97.

Metzger J, Schrimpf R, Philipp U, Distl O. 2013. Expression levels of LCORL are associated with body size in horses. *PLoS One* **8**: e56497.

Miyadera K, Acland GM, Aguirre GD. 2012. Genetic and phenotypic variations of inherited retinal diseases in dogs: the power of within- and across-breed studies. *Mamm Genome* **23**: 40-61.

Mooney JA, Yohannes A, Lohmueller KE. 2021. The impact of identity by descent on fitness and disease in dogs. *Proc Natl Acad Sci U S A* **118**.

Morrill K, Hekman J, Li X, McClure J, Logan B, Goodman L, Gao M, Dong Y, Alonso M, Carmichael E et al. 2022. Ancestry-inclusive dog genomics challenges popular breed stereotypes. *Science* **376**: eabk0639.

Downloaded from genome.cshlp.org on March 2, 2026 - Published by Cold Spring Harbor Laboratory Press

Nei M, Li WH. 1979. Mathematical model for studying genetic variation in terms of restriction endonucleases. *Proc Natl Acad Sci U S A* **76**: 5269-5273.

Ng PC, Henikoff S. 2001. Predicting deleterious amino acid substitutions. *Genome Res* **11**: 863-874.

Nicholas F, Tammen I, Sydney Informatics H. 1995. Online Mendelian Inheritance in Animals (OMIA). doi:10.25910/2amr-pv70. The University of Sydney.

Nicholas FW. 2021. Online Mendelian Inheritance in Animals (OMIA): a record of advances in animal genetics, freely available on the Internet for 25 years. *Anim Genet* **52**: 3-9.

Nicholas FW, Tammen, I., Sydney Informatics Hub. 2024. Online Mendelian Inheritance in Animals (OMIA). doi:https://doi.org/10.25910/2AMR-PV70.

O'Neill DG, Pegram C, Crocker P, Brodbelt DC, Church DB, Packer RMA. 2020. Unravelling the health status of brachycephalic dogs in the UK using multivariable analysis. *Sci Rep* **10**: 17251.

Ostrander EA, Wang GD, Larson G, vonHoldt BM, Davis BW, Jagannathan V, Hitte C, Wayne RK, Zhang YP, Dog KC. 2019. Dog10K: an international sequencing effort to advance studies of canine domestication, phenotypes and health. *Natl Sci Rev* **6**: 810-824.

Ostrander EA, Wayne RK, Freedman AH, Davis BW. 2017. Demographic history, selection and functional diversity of the canine genome. *Nat Rev Genet* **18**: 705-720.

Özarslan B, Atasoy F, Erdogan M, Özkul BY, Özbaser FT. 2021. Evaluation of some morphological and genetic characteristics of Catalburun dog breed. *Turk J Vet Anim Sci* **45**: 649-656.

Paradas-Lara I, Casado-Gomez I, Martin C, Martinez-Sanz E, Lopez-Gordillo Y, Gonzalez P, Rodriguez-Bobada C, Chamorro M, Arias P, Maldonado E et al. 2014. Maxillary growth in a congenital cleft palate canine model for surgical research. *J Craniomaxillofac Surg* **42**: 13-21.

Downloaded from genome.cshlp.org on March 2, 2026 - Published by Cold Spring Harbor Laboratory Press

Paradis E, Schliep K. 2019. ape 5.0: an environment for modern phylogenetics and evolutionary analyses in R. *Bioinformatics* **35**: 526-528.

Parker HG, Dreger DL, Rimbault M, Davis BW, Mullen AB, Carpintero-Ramirez G, Ostrander EA. 2017a. Genomic Analyses Reveal the Influence of Geographic Origin, Migration, and Hybridization on Modern Dog Breed Development. *Cell Rep* **19**: 697-708.

Parker HG, Harris A, Dreger DL, Davis BW, Ostrander EA. 2017b. The bald and the beautiful: hairlessness in domestic dog breeds. *Philos Trans R Soc Lond B Biol Sci* **372**.

Parker HG, Kim LV, Sutter NB, Carlson S, Lorentzen TD, Malek TB, Johnson GS, DeFrance HB, Ostrander EA, Kruglyak L. 2004. Genetic structure of the purebred domestic dog. *Science* **304**: 1160-1164.

Plassais J, Kim J, Davis BW, Karyadi DM, Hogan AN, Harris AC, Decker B, Parker HG, Ostrander EA. 2019. Whole genome sequencing of canids reveals genomic regions under selection and variants influencing morphology. *Nat Commun* **10**: 1489.

Plassais J, vonHoldt BM, Parker HG, Carmagnini A, Dubos N, Papa I, Bevant K, Derrien T, Hennelly LM, Whitaker DT et al. 2022. Natural and human-driven selection of a single non-coding body size variant in ancient and modern canids. *Curr Biol* **32**: 889-897 e889.

Poplin R, Ruano-Rubio V, DePristo MA, Fennell TJ, Carneiro MO, Van der Auwera GA, Kling DE, Gauthier LD, Levy-Moonshine A, Roazen D. 2017. Scaling accurate genetic variant discovery to tens of thousands of samples. *BioRxiv*: 201178.

Popova SN, Barczyk M, Tiger CF, Beertsen W, Zigrino P, Aszodi A, Miosge N, Forsberg E, Gullberg D. 2007. Alpha11 beta1 integrin-dependent regulation of periodontal ligament function in the erupting mouse incisor. *Mol Cell Biol* **27**: 4306-4316.

Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MA, Bender D, Maller J, Sklar P, de Bakker PI, Daly MJ, Sham PC. 2007. PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am J Hum Genet* **81**: 559-575.

Downloaded from genome.cshlp.org on March 2, 2026 - Published by Cold Spring Harbor Laboratory Press

Qian C, Wong CWY, Wu Z, He Q, Xia H, Tam PKH, Wong KKY, Lui VCH. 2017. Stage specific requirement of platelet-derived growth factor receptor-alpha in embryonic development. *PLoS One* **12**: e0184473.

Rentzsch P, Witten D, Cooper GM, Shendure J, Kircher M. 2019. CADD: predicting the deleteriousness of variants throughout the human genome. *Nucleic Acids Res* **47**: D886-D894.

Rimbault M, Beale HC, Schoenebeck JJ, Hoopes BC, Allen JJ, Kilroy-Glynn P, Wayne RK, Sutter NB, Ostrander EA. 2013. Derived variants at six genes explain nearly half of size reduction in dog breeds. *Genome Res* **23**: 1985-1995.

Rogers CA, Brace AH. 1995. *The international encyclopedia of dogs*. Howell Book House, New York.

Rubin CJ, Megens HJ, Martinez Barrio A, Maqbool K, Sayyab S, Schwochow D, Wang C, Carlborg O, Jern P, Jorgensen CB et al. 2012. Strong signatures of selection in the domestic pig genome. *Proc Natl Acad Sci U S A* **109**: 19529-19536.

Saif R, Henkel J, Jagannathan V, Drogemuller C, Flury C, Leeb T. 2020. The LCORL Locus is under Selection in Large-Sized Pakistani Goat Breeds. *Genes (Basel)* **11**.

Samocha KE, Robinson EB, Sanders SJ, Stevens C, Sabo A, McGrath LM, Kosmicki JA, Rehnstrom K, Mallick S, Kirby A et al. 2014. A framework for the interpretation of de novo mutation in human disease. *Nat Genet* **46**: 944-950.

Sarkar A, Panati K, Narala VR. 2022. Code inside the codon: The role of synonymous mutations in regulating splicing machinery and its impact on disease. *Mutat Res Rev Mutat Res* **790**: 108444.

Schoenebeck JJ, Hutchinson SA, Byers A, Beale HC, Carrington B, Faden DL, Rimbault M, Decker B, Kidd JM, Sood R et al. 2012. Variation of BMP3 contributes to dog breed skull diversity. *PLoS Genet* **8**: e1002849.

Downloaded from genome.cshlp.org on March 2, 2026 - Published by Cold Spring Harbor Laboratory Press

Sehgal A, Irvine KM, Hume DA. 2021. Functions of macrophage colony-stimulating factor (CSF1) in development, homeostasis, and tissue repair. *Semin Immunol* **54**: 101509.

Sullivan PF, Meadows JRS, Gazal S, Phan BN, Li X, Genereux DP, Dong MX, Bianchi M, Andrews G, Sakthikumar S et al. 2023. Leveraging base-pair mammalian constraint to understand genetic variation and human disease. *Science* **380**: eabn2937.

Sun KY, Bai X, Chen S, Bao S, Zhang C, Kapoor M, Backman J, Joseph T, Maxwell E, Mitra G et al. 2024. A deep catalogue of protein-coding variation in 983,578 individuals. *Nature* **631**: 583-592.

Tetens J, Widmann P, Kuhn C, Thaller G. 2013. A genome-wide association study indicates LCORL/NCAPG as a candidate locus for withers height in German Warmblood horses. *Anim Genet* **44**: 467-471.

Van den Broeck M, De Bels L, Duchateau L, Cornillie P. 2022. Time and sequence of emergence of the deciduous dentition in dogs and its applicability for age estimation. *Anat Histol Embryol* **51**: 640-657.

Van Wesenbeeck L, Odgren PR, MacKay CA, D'Angelo M, Safadi FF, Popoff SN, Van Hul W, Marks SC, Jr. 2002. The osteopetrotic mutation toothless (tl) is a loss-of-function frameshift mutation in the rat Csf1 gene: Evidence of a crucial role for CSF-1 in osteoclastogenesis and endochondral ossification. *Proc Natl Acad Sci U S A* **99**: 14303-14308.

Vaysse A, Ratnakumar A, Derrien T, Axelsson E, Rosengren Pielberg G, Sigurdsson S, Fall T, Seppala EH, Hansen MS, Lawley CT et al. 2011. Identification of genomic regions associated with phenotypic variation between dog breeds using selection mapping. *PLoS Genet* **7**: e1002316.

Wade CM. 2025. Insights into the genetics of body size in the Bull Terrier. *Anim Genet* **56**: e70000.

Downloaded from genome.cshlp.org on March 2, 2026 - Published by Cold Spring Harbor Laboratory Press

Wang C, Wallerman O, Arendt ML, Sundstrom E, Karlsson A, Nordin J, Makelainen S, Pielberg GR, Hanson J, Ohlsson A et al. 2021. A novel canine reference genome resolves genomic architecture and uncovers transcript complexity. *Commun Biol* **4**: 185.

Wayne RK, vonHoldt BM. 2012. Evolutionary genomics of dog domestication. *Mamm Genome* **23**: 3-18.

Weir BS, Cockerham CC. 1984. Estimating F-Statistics for the Analysis of Population Structure. *Evolution* **38**: 1358-1370.

Williams LT. 1989. Signal transduction by the platelet-derived growth factor receptor. *Science* **243**: 1564-1570.

Worboys M, Strange, J.M., Pemberton, N. . 2018. *The invention of the modern dog: breed and blood in Victorian Britain*. Johns Hopkins University Press, Baltimore.

Yang H, Wang G, Wang M, Ma Y, Yin T, Fan R, Wu H, Zhong L, Irwin DM, Zhai W, Zhang Y. 2017. The origin of chow chows in the light of the East Asian breeds. *BMC Genomics* **18**: 174.

Yengo L Vedantam S Marouli E Sidorenko J Bartell E Sakaue S Graff M Eliasen AU Jiang Y Raghavan S et al. 2022. A saturated map of common genetic variants associated with human height. *Nature* **610**: 704-712.

Yilmaz O, Ertugrul M, Wilson RT. 2012. The domestic livestock resources of Turkey: breed descriptions and status of guard and hunting dogs. In *63rd Annual Meeting of the EAAP*, pp. 27-31.

Yoshida H, Hayashi S, Kunisada T, Ogawa M, Nishikawa S, Okamura H, Sudo T, Shultz LD, Nishikawa S. 1990. The murine mutation osteopetrosis is in the coding region of the macrophage colony stimulating factor gene. *Nature* **345**: 442-444.

Yu Y, Alvarado R, Petty LE, Bohlender RJ, Shaw DM, Below JE, Bejar N, Ruiz OE, Tandon B, Eisenhoffer GT et al. 2022. Polygenic risk impacts PDGFRA mutation penetrance in non-syndromic cleft lip and palate. *Hum Mol Genet* **31**: 2348-2357.

Downloaded from genome.cshlp.org on March 2, 2026 - Published by Cold Spring Harbor Laboratory Press

Zeng T, Spence JP, Mostafavi H, Pritchard JK. 2024. Bayesian estimation of gene constraint from an evolutionary model with gene features. *Nat Genet* **56**: 1632-1643.

Zhou X, Stephens M. 2012. Genome-wide efficient mixed-model analysis for association studies. *Nat Genet* **44**: 821-824.

Downloaded from genome.cshlp.org on March 2, 2026 - Published by Cold Spring Harbor Laboratory Press

Figure 1: Rare canine variants are used to estimate mutational tolerance in constrained genes. A) Pearson correlation coefficient between observed rare variants and gene mutation probabilities. Variants are organized by the group of samples they were observed in and the maximum allele count threshold for defining rare variants, which is displayed near each datapoint. The grey band represents the 95% expected interval of the model fit based on simulated accumulation of synonymous mutations. **B)** Pearson correlation coefficient of rare variant frequency and gene mutation probability across gene groups. Genes were grouped according to their mutation probabilities. The x-axis shows the number of groups used to calculate correlation coefficients. The number of groups decreases as group sizes get larger. Two models are tested, one is the mutation probabilities calculated using the mutation model and the other is mutation probabilities according to gene CDS length. The 95% confidence intervals for the expected model fit are also displayed. **C)** Gene mutation frequency according to three different constraint metrics. The expected number of mutations was calculated following the mutation model. The observed number of mutations are from singletons in dogs with highly represented ancestry. Error bars indicate whether the observed mutation frequency is outside the 95% expected bounds.

Figure 2: *PDGFRA* LoF mutations are associated with bifid nose. A) Çatalburun dog with the distinctive bifid nose phenotype. **B)** A mixed-breed dog (Dog ID: 40931) with a bifid nose phenotype. The dog is heterozygous for an additional *PDGFRA* splice-acceptor variant. **C)** Canine *PDGFRA* transcript showing the locations of each splice acceptor variant found within dogs with orofacial clefts. The first variant is the Çatalburun variant and the second variant belongs to Dog 40931. **D)** Photograph of a statue depicting a dog with a bifid nose phenotype. The statue is located at the Palace of Bertemati in Andalusia, Spain and was constructed in the 18th Century. **E)** Haplotype analysis of the *PDGFRA* locus. A neighbor joining tree is shown alongside a matrix of phased haplotypes. Red diamonds on the tree show the locations of

Downloaded from genome.cshlp.org on March 2, 2026 - Published by Cold Spring Harbor Laboratory Press

Çatalburun haplotypes. The inset shows the Çatalburun haplotypes containing the *PDGFRA* splice-acceptor mutation. The position of the mutation within the haplotype is indicated in red.

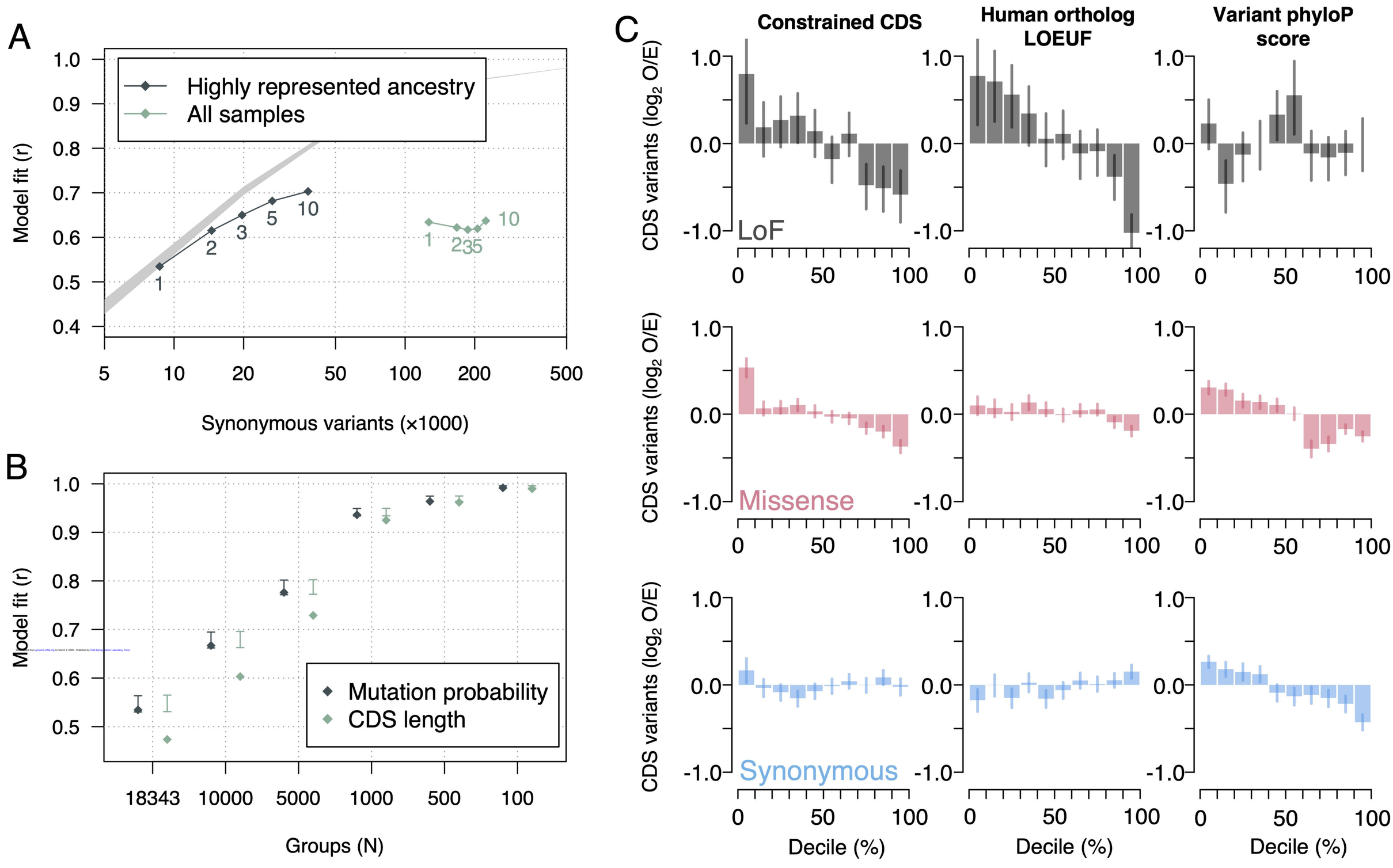
Fig. 3: Identification of new *LCORL* variants associated with canine body size. **A)** Mutations found in the long isoform of *LCORL*. Early termination of translation and loss of the DUF4553 domain caused by frameshift mutations is associated with larger dog size (>55 cm/>25 kg). **B)** Manhattan plots at the *LCORL* locus showing association with breed standard height for three different variants in separate linkage groups. **C - D)** Boxplots showing associations between homozygous *LCORL* size associated genotypes and breed standard height.

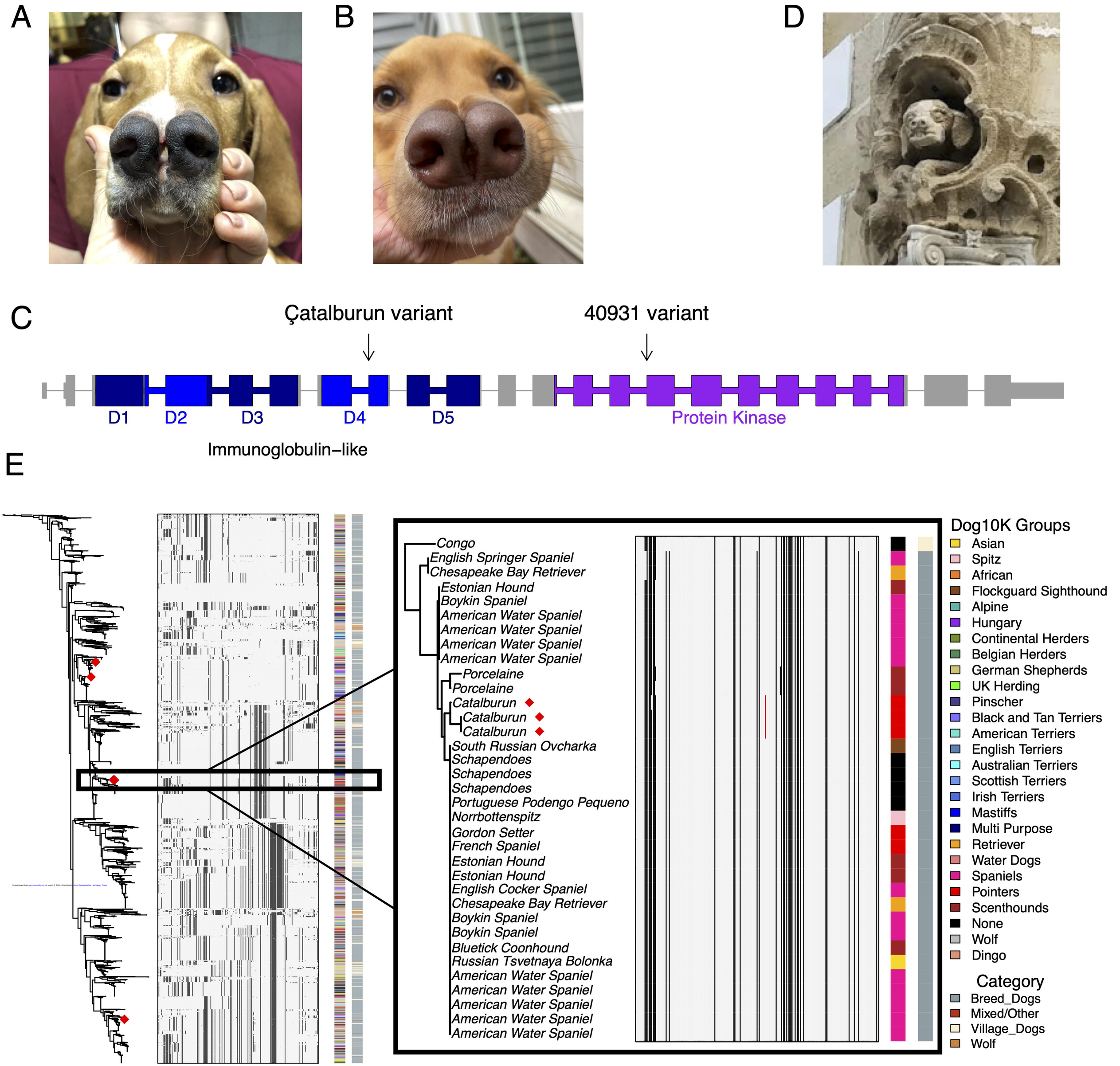
Fig 4: Haplotype analysis of *LCORL* locus. A neighbor joining tree is displayed alongside a matrix of phased haplotypes. Each of the *LCORL* size variants form distinct haplotypes with few recombinants. Depending on whether haplotypes are shared between wolves and dogs or are exclusive to dogs, they are classified as either ancestral or modern, respectively.

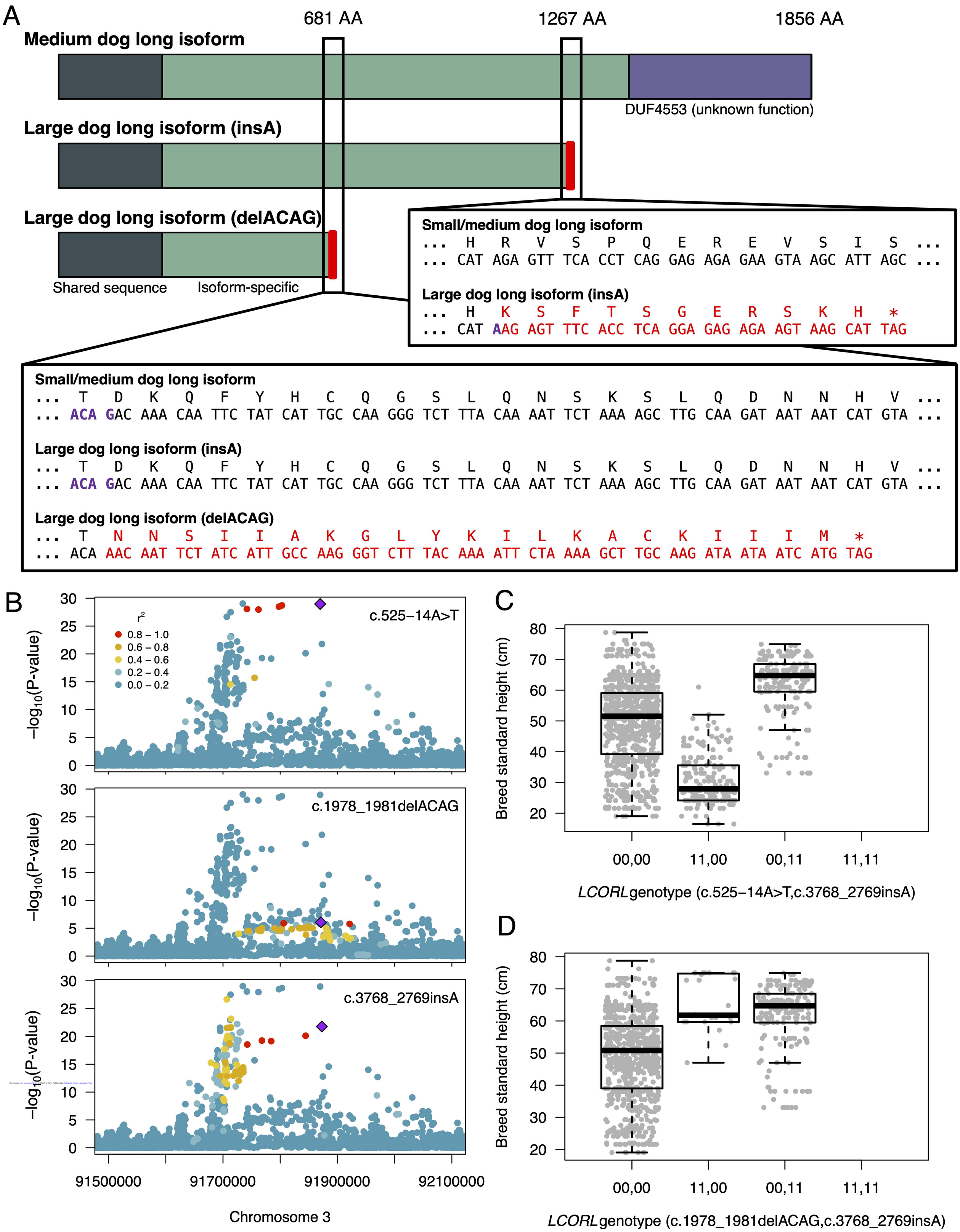
Fig 5: Noncoding variation at constrained genes drives breed differentiation. **A)** Identification of breed differentiated regions at Chromosome 15 *IGF1* locus. Genomic windows with F_{ST} > 0.8 between breed pairs were plotted across the genome. Peak F_{ST} regions were then classified as regions of breed differentiation. Top F_{ST} peaks are shown in the top panel and highlighted in red. The second panel shows genes found in the region. The third panel shows the density of regions of breed homozygosity that were identified as 1Mb genomic regions with breed nucleotide diversity <0.02. The final panel shows black bars as individual breed homozygosity stretches. **B)** Genome-wide distribution of breed-differentiated regions. Trait loci summarized in Plassais et al (2019) are displayed as lower-case characters. Trait loci in red are within 100kb of a breed-differentiated region. **C)** Number of genes in highly differentiated

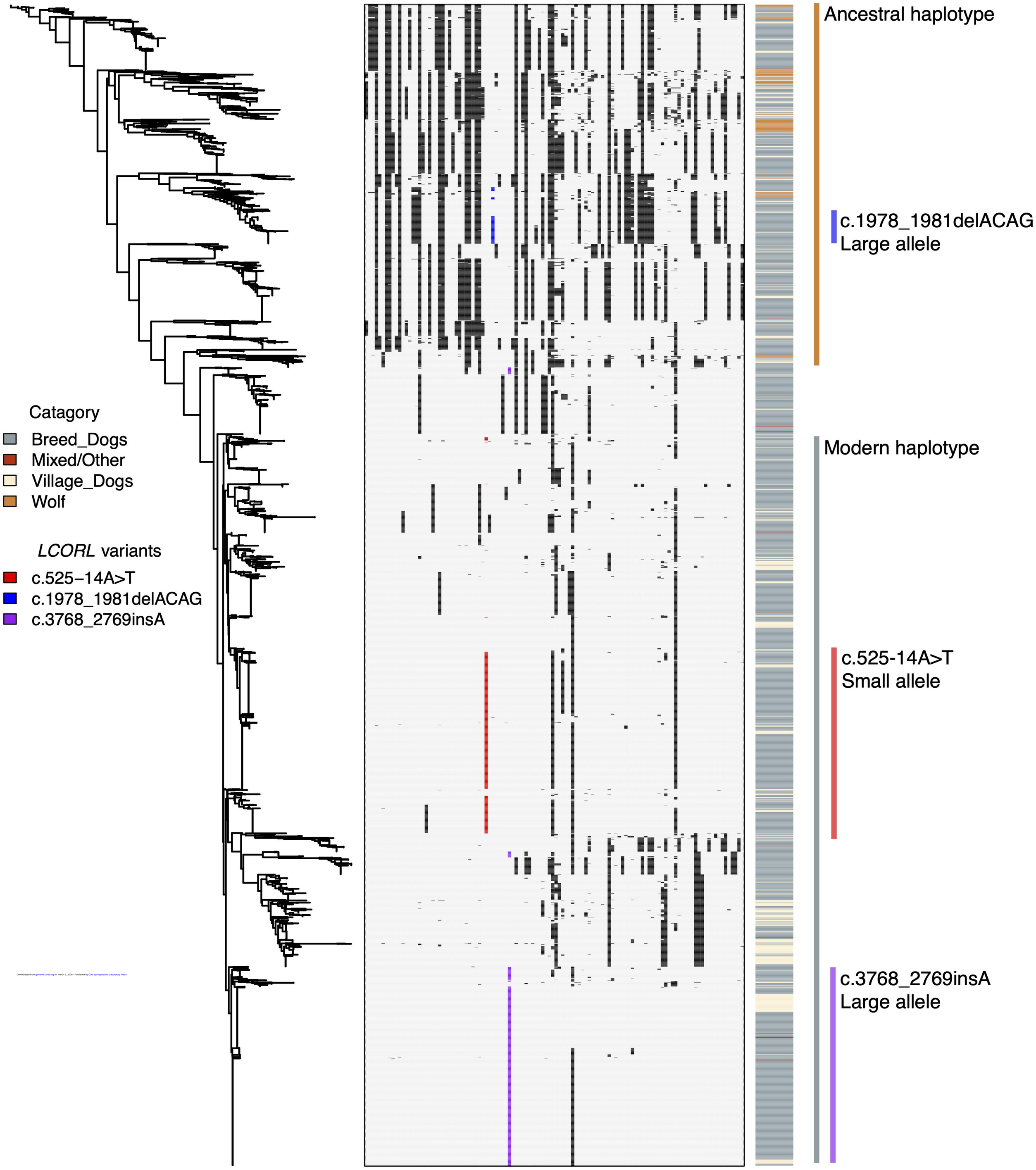
Downloaded from genome.cshlp.org on March 2, 2026 - Published by Cold Spring Harbor Laboratory Press

regions that contain potentially functional variants that pass minimum constraint thresholds. The "High F_{ST}" column is the number of genes that contain variants of interest. The "Low pi" column is the number of genes that contain variants in breeds with low nucleotide diversity across breed-differentiated regions. The final column is the number of genes that contain variants associated with breed differentiation (Fisher test, P-value < 0.01). **D)** Observed over expected frequency of constrained genes within breed-differentiated regions. The x-axis indicates constraint decile with 10 being the most constrained genes.


A
Highly represented ancestry
All samples
Model fit (r)
Synonymous variants (×1000)
B
Mutation probability
CDS length
Model fit (r)
Groups (N)
C
Constrained CDS
Human ortholog LOEUF
Variant phyloP score
CDS variants (log2 O/E)
LoF
Missense
Synonymous
Decile (%)

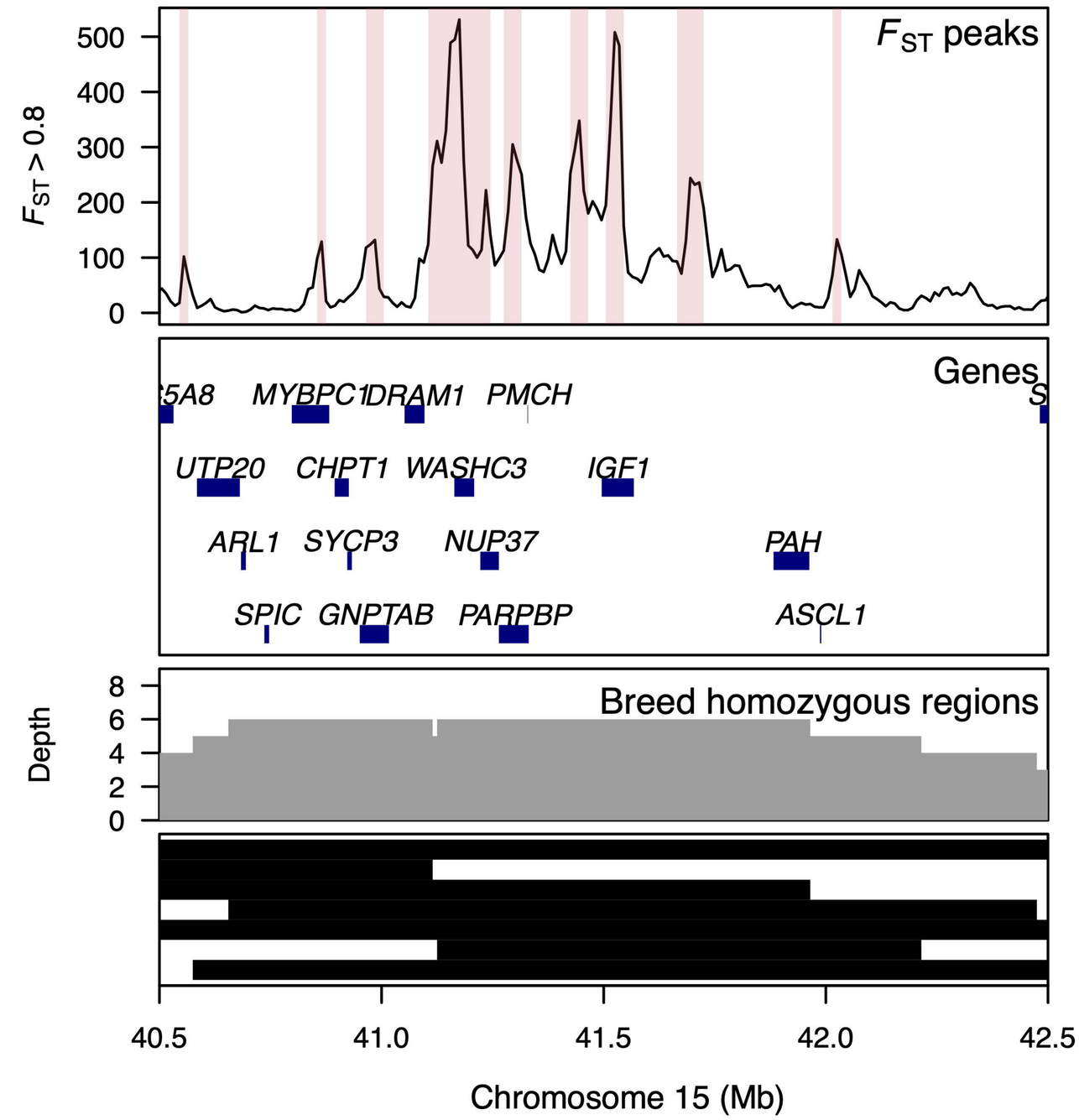

A
B
D
C
Çatalburun variant
40931 variant
D1
D2
D3
D4
D5
Protein Kinase
Immunoglobulin–like
E
Congo
English Springer Spaniel
Chesapeake Bay Retriever
Estonian Hound
Boykin Spaniel
American Water Spaniel
American Water Spaniel
American Water Spaniel
American Water Spaniel
Porcelaine
Porcelaine
Catalburun
Catalburun
Catalburun
South Russian Ovcharka
Schapendoes
Schapendoes
Schapendoes
Portuguese Podengo Pequeno
Norrbottenspitz
Gordon Setter
French Spaniel
Estonian Hound
Estonian Hound
English Cocker Spaniel
Chesapeake Bay Retriever
Boykin Spaniel
Boykin Spaniel
Bluetick Coonhound
Russian Tsvetnaya Bolonka
American Water Spaniel
American Water Spaniel
American Water Spaniel
American Water Spaniel
American Water Spaniel
Dog10K Groups
Asian
Spitz
African
Flockguard Sighthound
Alpine
Hungary
Continental Herders
Belgian Herders
German Shepherds
UK Herding
Pinscher
Black and Tan Terriers
American Terriers
English Terriers
Australian Terriers
Scottish Terriers
Irish Terriers
Mastiffs
Multi Purpose
Retriever
Water Dogs
Spaniels
Pointers
Scenthounds
None
Wolf
Dingo
Category
Breed_Dogs
Mixed/Other
Village_Dogs
Wolf


A
681 AA
1267 AA
1856 AA
Medium dog long isoform
DUF4553 (unknown function)
Large dog long isoform (insA)
Large dog long isoform (delACAG)
Shared sequence
Isoform-specific
Small/medium dog long isoform
... H R V S P Q E R E V S I S ...
... CAT AGA GTT TCA CCT CAG GAG AGA GAA GTA AGC ATT AGC ...
Large dog long isoform (insA)
... H K S F T S G E R S K H *
... CAT AAG AGT TTC ACC TCA GGA GAG AGA AGT AAG CAT TAG
Small/medium dog long isoform
... T D K Q F Y H C Q G S L Q N S K S L Q D N N H V ...
... ACA GAC AAA CAA TTC TAT CAT TGC CAA GGG TCT TTA CAA AAT TCT AAA AGC TTG CAA GAT AAT AAT CAT GTA ...
Large dog long isoform (insA)
... T D K Q F Y H C Q G S L Q N S K S L Q D N N H V ...
... ACA GAC AAA CAA TTC TAT CAT TGC CAA GGG TCT TTA CAA AAT TCT AAA AGC TTG CAA GAT AAT AAT CAT GTA ...
Large dog long isoform (delACAG)
... T N N S I I A K G L Y K I L K A C K I I I M *
... ACA AAC AAT TCT ATC ATT GCC AAG GGT CTT TAC AAA ATT CTA AAA GCT TGC AAG ATA ATA ATC ATG TAG
B
r²
0.8 – 1.0
0.6 – 0.8
0.4 – 0.6
0.2 – 0.4
0.0 – 0.2
c.525–14A>T
c.1978_1981delACAG
c.3768_2769insA
−log10(P-value)
91500000
91700000
91900000
92100000
Chromosome 3
C
Breed standard height (cm)
00,00
11,00
00,11
11,11
LCORL genotype (c.525–14A>T,c.3768_2769insA)
D
Breed standard height (cm)
00,00
11,00
00,11
11,11
LCORL genotype (c.1978_1981delACAG,c.3768_2769insA)

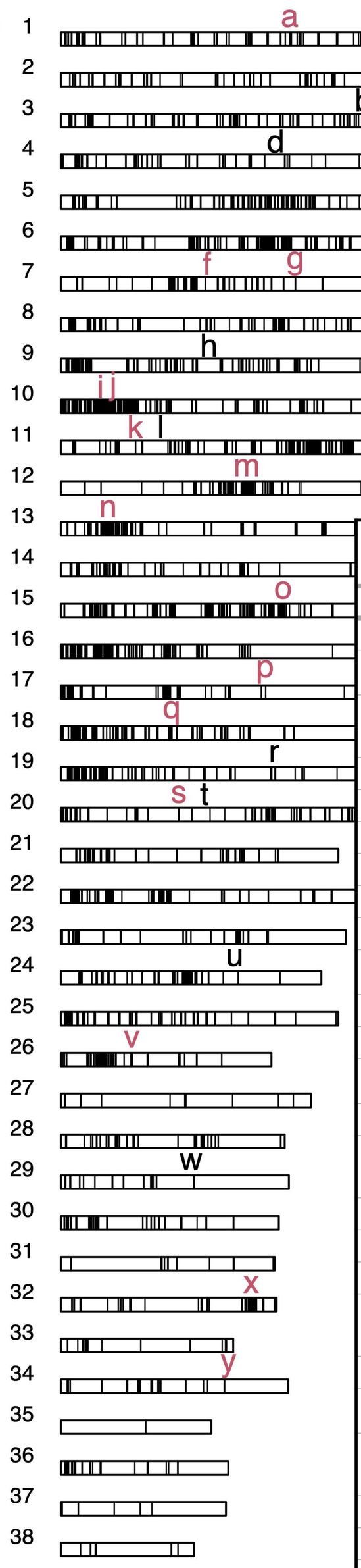


Downloaded from genome.cshlp.org on March 2, 2026 - Published by Cold Spring Harbor Laboratory Press

A



B



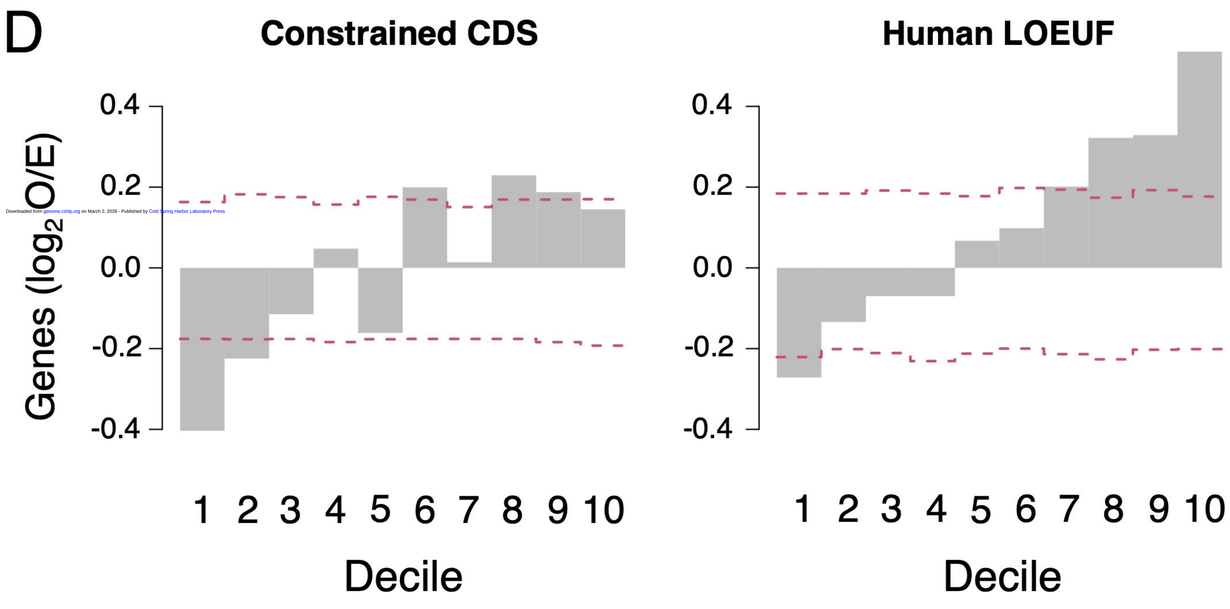
Trait loci summarized in Plassais et al. (2019)

Marker	Gene/locus	Trait
a	*ESR1*	Height
b	*ADAMTSL3*	Bulky
c	*LCORL*	Height, weight
d	*STC2*	Height
e	*GHR*	Height
f	*GORAB*	Curl tail
g	*SMAD2*	Height
h	*CA10*	Hairless
i	*MSRB3*	Drop ears
j	*HMGA2*	Height
k	*ZNF608*	Weight, bulky
l	*CHSY3*	Curl tail
m	*KCNQ5*	Large ears
n	*RSPO2*	Furnishing
o	*IGF1*	Height, weight
p	*FOXI3*	Hairless
q	*FGF4*	Height
r	*R3HDM1*	Weight, bulky
s	*MITF*	White chest, white head
t	*ADAMTS9*	Weight
u	*R3HDML*	Aggressive
v	*MED13L*	Height, weight
w	*HNF4G*	Bulky
x	*FGF5*	Length of fur
y	*IGF2BP2*	Weight

C

Variant type	Constraint	High F_{ST}	Low pi	P < 0.01
LoF	LOEUF	11	1	1
LoF	Con CDS	10	1	1
LoF	PhyloP score	34	7	1
Mis	LOEUF	123	45	25
Mis	Con CDS	70	18	9
Mis	PhyloP score	496	111	57

D



Downloaded from genome.cshlp.org on March 2, 2026 - Published by Cold Spring Harbor Laboratory Press

GENOME RESEARCH

Analysis of canine gene constraint identifies new variants for orofacial clefts and stature

Reuben M. Buckley, Nuket Bilgen, Alexander C. Harris, et al.

Genome Res. published online March 24, 2025
Access the most recent version at doi:10.1101/gr.280092.124

Supplemental Material http://genome.cshlp.org/content/suppl/2025/04/04/gr.280092.124.DC1

P<P Published online March 24, 2025 in advance of the print journal.

Accepted Manuscript Peer-reviewed and accepted for publication but not copyedited or typeset; accepted manuscript is likely to differ from the final, published version.

Open Access Freely available online through the *Genome Research* Open Access option.

Creative Commons License This manuscript is Open Access.This article, published in *Genome Research*, is available under a Creative Commons License (Attribution-NonCommercial 4.0 International license), as described at http://creativecommons.org/licenses/by-nc/4.0/.

Email Alerting Service Receive free email alerts when new articles cite this article - sign up in the box at the top right corner of the article or **click here.**

To subscribe to *Genome Research* go to:
https://genome.cshlp.org/subscriptions

Published by Cold Spring Harbor Laboratory Press