bioRxiv preprint doi: https://doi.org/10.1101/2024.05.01.592072; this version posted September 15, 2025. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

Genotypic and phenotypic consequences of domestication in dogs

Authors: Sweetalana[1], Shirin Nataneli[2], Shengmiao Huang[2], Jazlyn A Mooney[2,*] and Zachary A Szpiech[1,3,*]

[1]Department of Biology, Pennsylvania State University, University Park, PA, USA

[2]Department of Quantitative and Computational Biology, University of Southern California, Los Angeles, CA, USA

[3]Institute for Computational and Data Sciences, Pennsylvania State University, USA

*Correspondence: jazlynmo@usc.edu and szpiech@psu.edu

Keywords: runs of homozygosity, dogs, inbreeding, GWAS, complex traits

Abstract

Runs of homozygosity (ROH) are genomic regions that arise when two copies of identical haplotypes are inherited from a shared common ancestor. In this study, we leverage ROH to identify associations between genetic diversity and non-disease phenotypes in *Canis lupus familiaris* (dogs). We find significant association between the ROH inbreeding coefficient (F_{ROH}) and several phenotypic traits. These traits include height, weight, lifespan, muscled, white coloring of the head and chest, furnishings, and fur length. After correcting for population structure, we identified more than 45 genes across the examined quantitative traits that exceed the threshold for suggestive significance. We observe distinct distributions of inbreeding and elevated levels of long ROH in modern breed dogs compared to more ancient breeds, which aligns with breeding practices during Victorian era breed establishment. Our results highlight the impact of non-additive variation and of polygenicity on complex quantitative phenotypes in dogs due to domestication and the breed formation bottleneck.

bioRxiv preprint doi: https://doi.org/10.1101/2024.05.01.592072; this version posted September 15, 2025. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

Background

For over a century, scholars and dog-enthusiasts alike have sought to unravel the complex evolutionary history of man's best friend [1]. *Canis lupus familiaris* (dogs) have intrigued researchers, across scientific fields, due to their close knit ties to our species [2] and unique genetic features [3–5]. While the exact origins of domestication remain elusive, understanding the impact of this process on the phenotypic variance and evolutionary history of the species is important.

In the 19th century, Darwin suggested that the broad range of features exhibited in modern breeds is conducive to a descendancy from multiple canid species, including wolves and jackals. In his accounts on variation, he argues that a single ancestral origin from the gray wolf (*Canis lupus*) seems to contradict the historical and archeological evidence available. Given that many wild species of the genus Canis have the potential to interbreed, Darwin's proposal held fast for many years [1]. Advances in genome technology provided scientists with a new approach to studying dog ancestry, offering insights that complement and, in some cases, challenge the conclusions drawn from archaeological and historical evidence [6]. One commonly used approach is mitochondrial DNA (mtDNA) analysis which allows researchers to trace maternal ancestry. Using samples collected from dogs, wolves, and several other wild canid species, researchers found the similarity between dogs and the gray wolf to be significantly higher than that of dogs and any other wild canid species [7, 8]. In recent years, whole genome sequence (WGS) data has increasingly focused on the wolf-dog relationship, further supporting this finding and making it a primary area of investigation in contemporary studies [4, 9, 10].

bioRxiv preprint doi: https://doi.org/10.1101/2024.05.01.592072; this version posted September 15, 2025. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

While most agree that dogs were the first domesticated species, establishing their spatial and temporal roots has been a challenge [11]. The earliest depictions of human-canine ties come by means of cave paintings discovered in Saudi Arabia [12]. This cooperative hunting scene is believed to be 8000 to 9000 years old, but evidence gathered from burial sites predict an earlier origin of domestication [13]. The first dates back approximately 12,000 years ago to the Natufian in Northern Israel, where they discovered three sets of dog remains buried with a human [14, 15]. Additionally, a dog-like mandible was recovered at a burial site in Bonn-Oberkassel, Germany. This Northern European sample is believed to be about 14,000 years old [16, 17], however genetic work has added an additional layer of evidence for an even older wolf-dog divergence.

Genomic work, based on mtDNA and WGS data, currently offers two potential explanations for the origins of domestication. The first is that early dogs were a result of a single domestication event from gray wolves in East Asia between 16,000 and 100,000 years ago. The hypothesis stems from the limited haplotypic diversity in dogs, which may suggest a genetic bottleneck where domesticated dogs originated from a small-closely-related population of gray wolves [7]. Additionally, high levels of genetic diversity from the region south of the Yangtze River, indicate an older divergence than other areas in Europe and Asia. [18–20].

The second genomic narrative centers upon the theory that haplogroups entered the dog lineage at different times and places. Using nuclear genomic single nucleotide polymorphism (SNP) analysis, one study found evidence for allele sharing in Asian dog breeds and Asian wolves as well as European dog breeds and European wolves [21]. This work suggests that there were several founder events derived from separate wolf populations. Further WGS work has documented significant genetic differences in Eastern and Western Eurasian dog populations,

bioRxiv preprint doi: https://doi.org/10.1101/2024.05.01.592072; this version posted September 15, 2025. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

providing further support for multiple domestication events [22, 23]. Despite these findings, the origins of domestication in dogs remain a hot topic of debate, warranting the need for additional studies exploring the complexities of their history.

While the early evolution of dogs from wolves marked a significant shift between species, the more recent history of established dog breeds has shaped evolution within the species itself. During the Victorian era, the selective breeding of aesthetically desirable traits led to the emergence of hundreds of new breeds. Breed establishment typically involved a limited number of founding members, resulting in high levels of inbreeding and dramatic loss of genetic diversity [3, 24–28]. Founder effects, such as this, restrict the gene pool and result in long stretches of DNA sharing between individuals. Shared segments derived from a single common ancestor are said to be identical-by-descent (IBD). Segments of the genome inherited IBD, when found within an individual, are commonly referred to as runs of homozygosity (ROH). Artificial selection for extreme phenotypes during breed formation made much of the genome homozygous and simplified complex trait architecture in dogs, which has facilitated the study of phenotypic associations between breeds [3, 29, 30].

Studies of ROH in humans have revealed associations with numerous complex disease and non-disease phenotypes, which provide valuable insight on the impact of genetic diversity on health and evolutionary processes [31–36]. Similar work has been done in dogs, providing evidence of association between ROH and several disease phenotypes [3, 30, 37]. Additionally, previous studies have explored the influence of genetic variants on quantitative traits such as; leg length, wrinkled skin, coat color, hair length, and skeletal shape [38–42]. Despite these advances, associations with complex non-disease phenotypes in dogs are relatively unexplored [43].

bioRxiv preprint doi: https://doi.org/10.1101/2024.05.01.592072; this version posted September 15, 2025. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

In this study, we investigate the relationship between the genomic consequences of domestication and complex (non-disease) trait architecture by using the distribution of ROH to provide insights on complex phenotypic architecture. To examine the relationship between ROH and non-disease phenotypes in dogs, we analyzed 556 canid whole-genome sequences and characterized the genomic distribution of ROH. After accounting for breed structure, we test for associations between the total ROH coverage of the genome (as measured by F_{ROH}) among 13 breed groups with 13 phenotypes. Finally, we perform an ROH-mapping genome-wide association study and identify multiple genetic variants associated with phenotype, namely height, weight, and lifespan.

Results

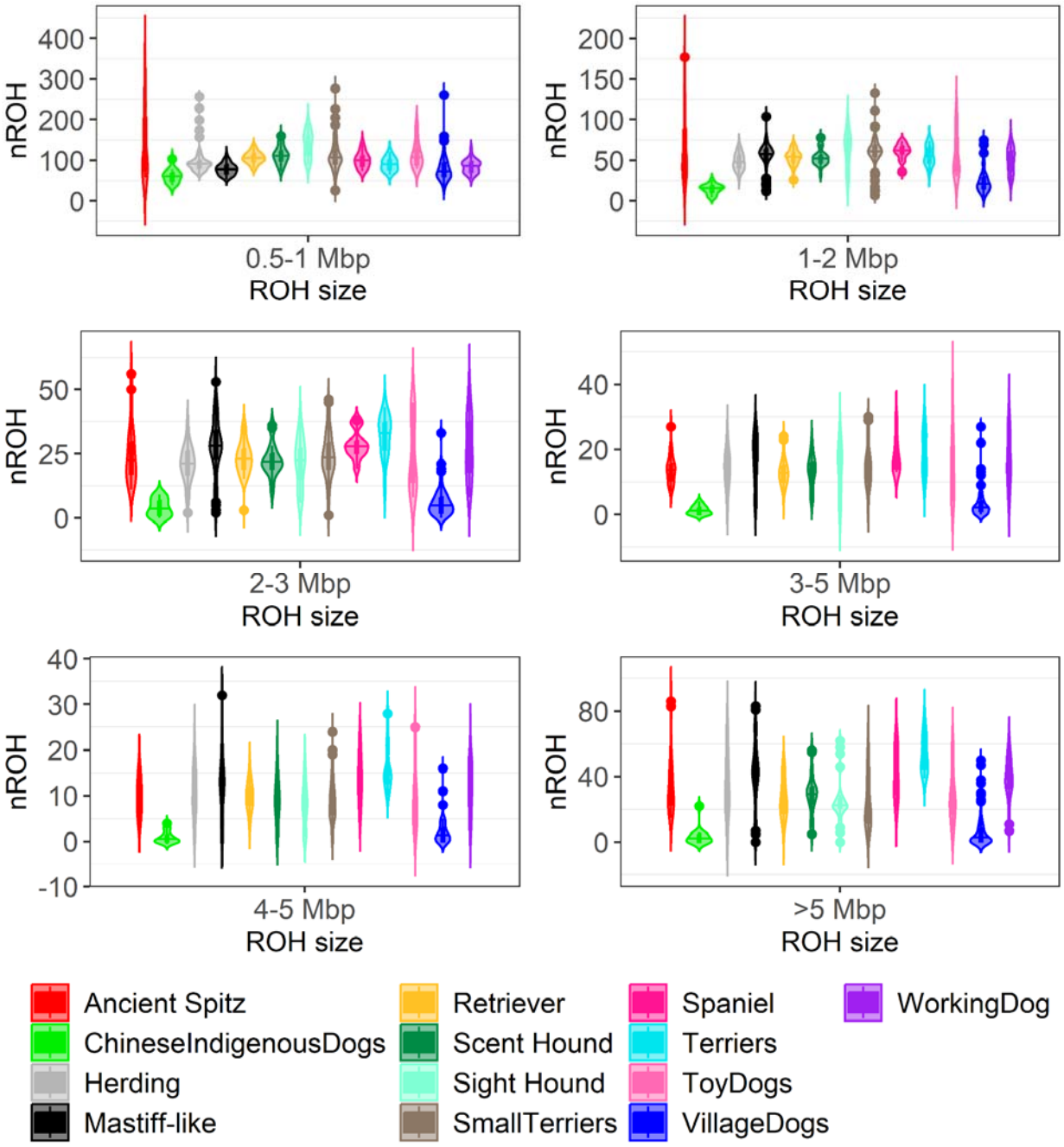
ROH distribution

ROH were categorized into 6 length classes: Class A (0.5 - 1 Mbp), Class B (1 - 2 Mbp), Class C (2 - 3 Mbp), Class D (3 - 4 Mbp), Class E (4 - 5 Mbp), and Class F (>5 Mbp). For our 556 samples, the mean number of ROH (n_{ROH}) was 37.06 and the mean length of ROH (s_{ROH}) was 94.17 Mbp. These parameters were also computed for each ROH length class: Class A (n_{ROH} = 99.44; s_{ROH} = 68.39 Mbp), Class B (n_{ROH} = 50.29; s_{ROH} = 70.36 Mbp), Class C (n_{ROH} = 21.60; s_{ROH} = 52.81 Mbp), Class D (n_{ROH} = 13.07; s_{ROH} = 45.23 Mbp), Class E (n_{ROH} = 9.21; s_{ROH} =41.15 Mbp), and Class F (n_{ROH} = 28.77; s_{ROH} = 287.05 Mbp) (Table 1, Table 2).

Domestic dog clades showed a distinct pattern of genetic diversity, with higher levels of long ROH (>5 Mbp) when compared with Chinese indigenous dogs (light green) and village dogs (royal blue) (Figs. 1 & 2). The highest total n_{ROH} (641) was observed in Dingo01 and the

bioRxiv preprint doi: https://doi.org/10.1101/2024.05.01.592072; this version posted September 15, 2025. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

highest total s_{ROH} (1643.98 Mbp) was observed in NGSD2, both individuals were from the Ancient Spitz breed group (Table S1). Terriers had the smallest mean n_{ROH} and the smallest mean s_{ROH} when compared with the remaining 11 breed groups (Tables 1 & 2).



bioRxiv preprint doi: https://doi.org/10.1101/2024.05.01.592072; this version posted September 15, 2025. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

Fig. 1: Relationship between the number of runs of homozygosity (n_{ROH}) and ROH size class in Mbp for all breed groups. *The x-axis represents the ROH size in Mbp. The y-axis represents the number of ROH. Each violin plot depicts the distribution of the total number of ROH within length classes 0.5-1 Mbp, 1-2 Mbp, 2-3 Mbp, 3-4 Mbp, 4-5 Mbp and >5 Mbp for each breed group.*

Table 1: n_{ROH} across all breed groups and ROH length classification. *This table details the range and mean ± standard deviation of the number of runs of homozygosity for each breed group and ROH length classification.*

Breed Group	**Mean n_{ROH}**	**Range n_{ROH}**	**Mean n_{ROH} for Class A (0.5 - 1 Mbp)**	**Mean n_{ROH} for Class B (1 - 2 Mbp)**	**Mean n_{ROH} for Class C (2 - 3 Mbp)**	**Mean n_{ROH} for Class D (3 - 4 Mbp)**	**Mean n_{ROH} for Class E (4 - 5 Mbp)**	**Mean n_{ROH} for Class F (> 5 Mbp)**
Ancient Spitz	47.10 ± 58.20	3 - 338	133.42 ± 85.63	65.13 ± 40.00	24.71 ± 11.23	13.90 ± 4.31	9.87 ± 3.97	35.58 ± 18.29
Retrievers	38.83 ± 34.58	0 - 138	106.42 ± 13.75	53.96 ± 9.05	23.98 ± 6.39	13.1 ± 4.31	9.94 ± 3.32	26.3 ± 11.77
Mastiff Like	38.37 ± 26.65	0 - 117	78.6 ± 14.12	54 ± 17.53	27.46 ± 11.19	16.91 ± 7.55	12.18 ± 6.53	41.08 ± 19.83
Herding	37.88 ± 33.97	0 - 257	97.76 ± 32.29	48.47 ± 10.17	21 ± 7.22	13.91 ± 5.79	10.8 ± 5.51	35.32 ± 19.08
Scent Hound	40.25 ± 37.52	1 - 160	113.26 ± 21.76	53.65 ± 9.85	22.7 ± 6.35	13.3 ± 4.87	9.09 ± 4.92	27.48 ± 12.44
Small Terrier	40.93 ± 41.19	0 - 277	113.58 ± 39.05	60.27 ± 17.74	24.08 ± 9.18	14.31 ± 6.32	8.98 ± 4.61	24.38 ± 16
Terriers	30.56 ±	9 - 122	90.75 ±	56.25 ±	31.81 ±	18.88 ±	15.88 ±	53.75 ±

bioRxiv preprint doi: https://doi.org/10.1101/2024.05.01.592072; this version posted September 15, 2025. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

	33.01		15.75	11.22	7.75	6.21	4.66	10.77
Spaniels	36.29 ± 35.23	5 - 144	101.18 ± 20.19	61 ± 9.23	28.73 ± 5.37	18.14 ± 5.35	14 ± 5.19	41.64 ± 14.62
Sighthound	45.29 ± 48.55	0 - 192	138.67 ± 31.05	63.33 ± 20.63	21.67 ± 8.98	13.86 ± 7.57	8.19 ± 4.76	26 ± 15.86
Toy Dogs	41.19 ± 40.77	1 - 187	114.56 ± 31.06	56.17 ± 25.99	24.06 ± 13.48	14.78 ± 9.78	9.44 ± 6.46	28.11 ± 15.34
Working Dogs	38.29 ± 27.43	1 - 129	85.18 ± 16.07	50.02 ± 15.37	27.36 ± 12.12	17 ± 7.86	11.82 ± 5.51	38.34 ± 13.11
Village Dogs	20.5 ± 31.83	0 - 261	80.12 ± 33.69	23.79 ± 13.69	6.33 ± 5.77	3.51 ± 4.76	2.09 ± 2.64	7.15 ± 11.03
Chinese Indigenous Dogs	14.36 ± 22.91	0 - 103	61.53 ± 16.63	14.67 ± 5.79	4 ± 3.31	1.27 ± 1.29	1 ± 1.32	3.67 ± 5.37

bioRxiv preprint doi: https://doi.org/10.1101/2024.05.01.592072; this version posted September 15, 2025. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

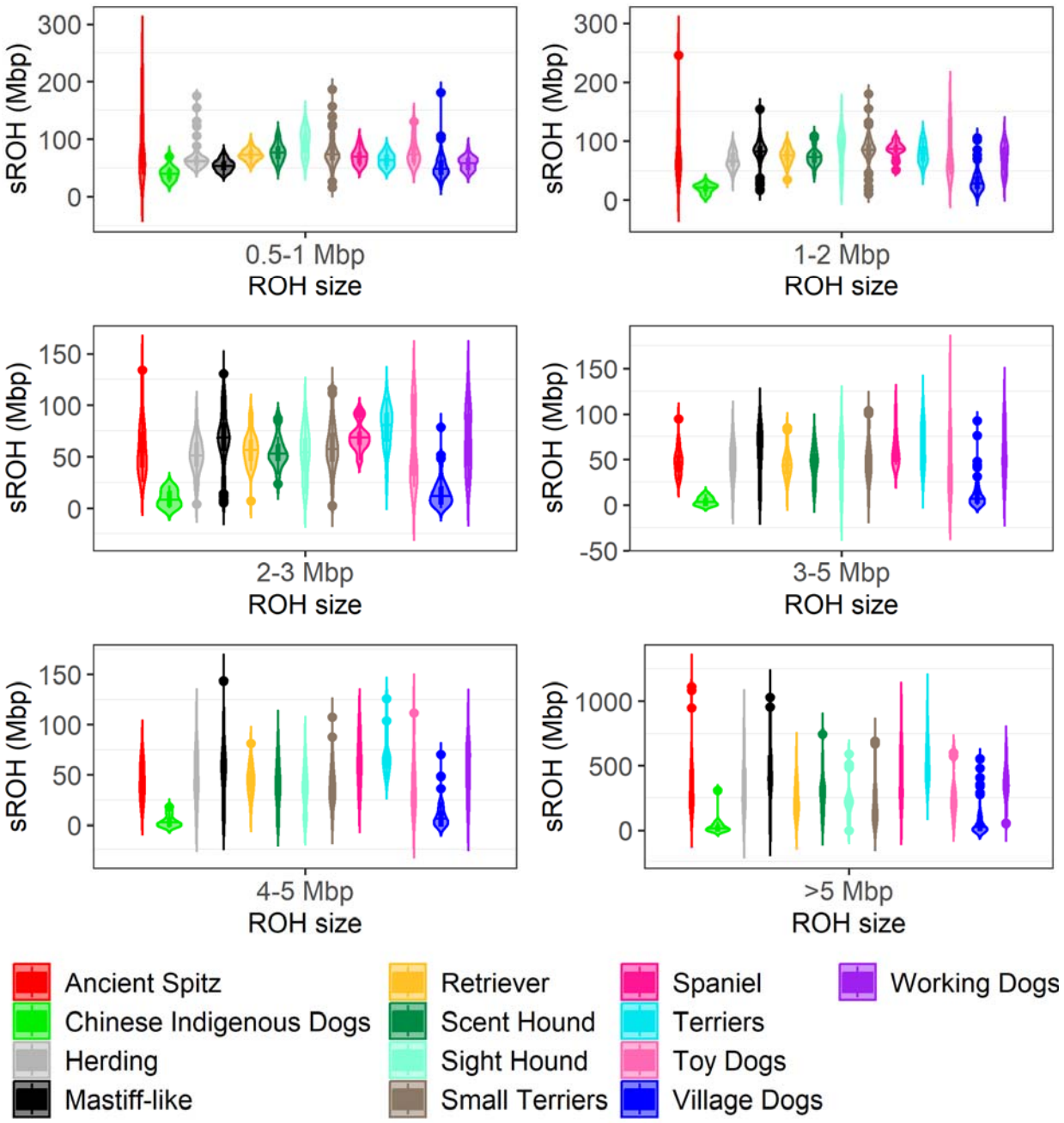


Fig. 2: Relationship between the sum of lengths of runs of homozygosity (s_{ROH}) and ROH size classes. *The x-axis represents our defined ROH size classes in Mbp. The y-axis represents s_{ROH}, which also has units of Mbp. Each violin plot depicts the distribution of the sum of ROH within length ranges 0.5-1 Mbp, 1-2 Mbp, 2-3 Mbp, 3-4 Mbp, 4-5 Mbp and >5 Mbp for each breed group.*

Table 2: s_{ROH} across all breed groups and ROH length classification. *The table contains the range and mean ± standard deviation of the sum of lengths of runs of homozygosity for each breed group and ROH length classification.*

bioRxiv preprint doi: https://doi.org/10.1101/2024.05.01.592072; this version posted September 15, 2025. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

Breed Group	**Mean S_{ROH} (Mbp)**	**Range S_{ROH} (Mbp)**	**Mean S_{ROH} for Class A (0.5 - 1 Mbp)**	**Mean S_{ROH} for Class B (1 - 2 Mbp)**	**Mean S_{ROH} for Class C (2 - 3 Mbp)**	**Mean S_{ROH} for Class D (3 - 4 Mbp)**	**Mean S_{ROH} for Class E (4 - 5 Mbp)**	**Mean S_{ROH} for Class F (>5 Mbp)**
Ancient Spitz	119.56 ± 162.09	13.73 - 1113.00	92.29 ± 59.41	90.67 ± 54.30	60.33 ± 27.20	48.30 ±	44.14 ± 17.51	381.60 ± 255.7
Retrievers	89.28 ± 86.63	0 - 598.02	72.96 ± 9.44	75.50 ± 13.25	57.37 ± 16.00	45.22 ± 15.07	44.47 ± 14.90	240.17 ± 125.98
Mastiff Like	126.47 ± 176.91	0 - 1029.28	54.77 ± 9.71	76.97 ± 25.95	67.21 ± 27.55	58.57 ± 25.92	54.38 ± 29.38	446.89 ± 247.25
Herding	105.25 ± 139.34	0 - 861.74	67.24 ± 21.71	67.53 ± 14.78	51.27 ± 17.64	48.23 ± 20.12	48.26 ± 24.73	348.94 ± 206.94
Scent Hound	99.19 ± 113.14	4.59 - 946.54	77.78 ± 14.73	74.84 ± 15.22	55.26 ± 15.67	45.90 ± 16.82	40.62 ± 21.78	300.75 ± 159.58
Small Terrier	89.41 ± 96.13	0 - 690.33	78.20 ± 26.38	84.24 ± 24.77	59.07 ± 22.76	49.63 ± 22.10	40.18 ± 20.51	225.15 ± 170.89
Terriers	75.86 ± 116.57	28.77 - 951.29	63.93 ± 10.90	79.69 ± 16.00	77.98 ± 19.72	65.84 ± 21.84	71.06 ± 20.79	541.71 ± 161.42

bioRxiv preprint doi: https://doi.org/10.1101/2024.05.01.592072; this version posted September 15, 2025. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

Spaniels	94.03 ± 127.54	23.23 - 856.91	70.13 ± 12.83	85.21 ± 12.79	70.39 ± 13.56	62.52 ± 18.65	62.55 ± 23.16	427.24 ± 190.19
Sighthound	94.65 ± 96.81	0 - 591.60	95.24 ± 21.50	86.87 ± 29.21	53.04 ± 22.07	47.98 ± 26.13	36.54 ± 21.32	248.26 ± 149.49
Toy Dogs	97.01 ± 107.57	4.12 - 602.43	78.92 ± 22.17	79.01 ± 36.61	58.31 ± 33.13	50.45 ± 33.72	42.33 ± 28.82	273.02 ± 162.01
Working Dogs	112.89 ± 130.13	4.74 - 658.42	58.67 ± 11.14	70.67 ± 22.08	66.54 ± 29.19	58.83 ± 27.51	52.76 ± 24.70	369.85 ± 139.08
Village Dogs	33.24 ± 59.75	0 - 554.84	54.07 ± 23.09	32.65 ± 19.59	15.30 ± 13.97	12.04 ± 16.34	9.35 ± 11.72	76.00 ± 127.67
Chinese Indigenous Dogs	19.82 ± 35.06	0 - 309.27	41.03 ± 11.34	20.00 ± 7.64	9.53 ± 7.68	4.27 ± 4.24	4.41 ± 5.87	39.67 ± 75.16

Next, we explored the relationship between the total number of ROH (n_{ROH}) and the total lengths of ROH (s_{ROH}) (Fig. 3). We observed significantly higher mean values for both n_{ROH} and s_{ROH} for the domesticated breed dogs relative to the domesticated non-breed Chinese indigenous dogs and village dogs. Breed dogs also tended to fall off of the y=x line when n_{ROH} was compared to s_{ROH}. The accumulation of long ROH within breed dogs in the recent time was most drastic in Terrier clade, where we also observed the largest F_{ROH} (Table 3). We observed the largest variance in nROH in the Small Terriers, which include breeds such as the Australian

bioRxiv preprint doi: https://doi.org/10.1101/2024.05.01.592072; this version posted September 15, 2025. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

Terrier, Cairn Terrier, Jack Russell Terrier, Norwich Terrier, Scottish Terrier, Tibetan Terrier, West Highland, White Terrier, and Yorkshire Terrier. We also observed quite a bit of variance in Ancient Spitz, Village dogs, Herding, Mastiff-like, and Toy dogs. Conversely, Chinese Indigenous dogs, Terriers, Spaniels, Retrievers, and Scent Hound cluster together quite tightly. The Spaniels fall somewhere between these groups.

We were also interested in whether any of ROH hotspots were shared across clades or between breed dogs and domesticated non-breed dog populations. We examined ROH coverage per site for each chromosome and plot sites where more than 50% of individuals within each clade have a ROH (Fig. S1-S16). There were clear hotspots within clades on every chromosome. Some of these hotspots corresponded with recombination rate, such as on chromosome 22, where we observed lower recombination rate in regions where there were shared ROH hotspots among all individuals. Conversely, on chromosome 12 we observed a high recombination rate, and very few shared ROH hotspots among individuals. Moreover, we observed very few shared ROH hotspots among different clades. One stand out was chromosome 28, where we observed a shared ROH hotspot among Ancient Spitz, Mastiff Like, Retrievers and Terriers (Fig. S13). Likewise on chromosome 32, we observed a shared ROH hotspot among Mastiff-Like, Scent Hound, Terriers and Toy Dogs (Fig. S14). The majority of hotspots were shared within a clade. This was best shown on chromosome 17, where we found multiple clade-specific ROH hotspots (Fig. S9). The same pattern was seen on chromosomes 31-38, with a few instances of sharing between Mastiff-like breeds and Terriers (Fig. S13 - S16). Lastly, we rarely observed any peaks for village dogs and Chinese Indigenous dogs (Fig. S3 - S16).

bioRxiv preprint doi: https://doi.org/10.1101/2024.05.01.592072; this version posted September 15, 2025. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

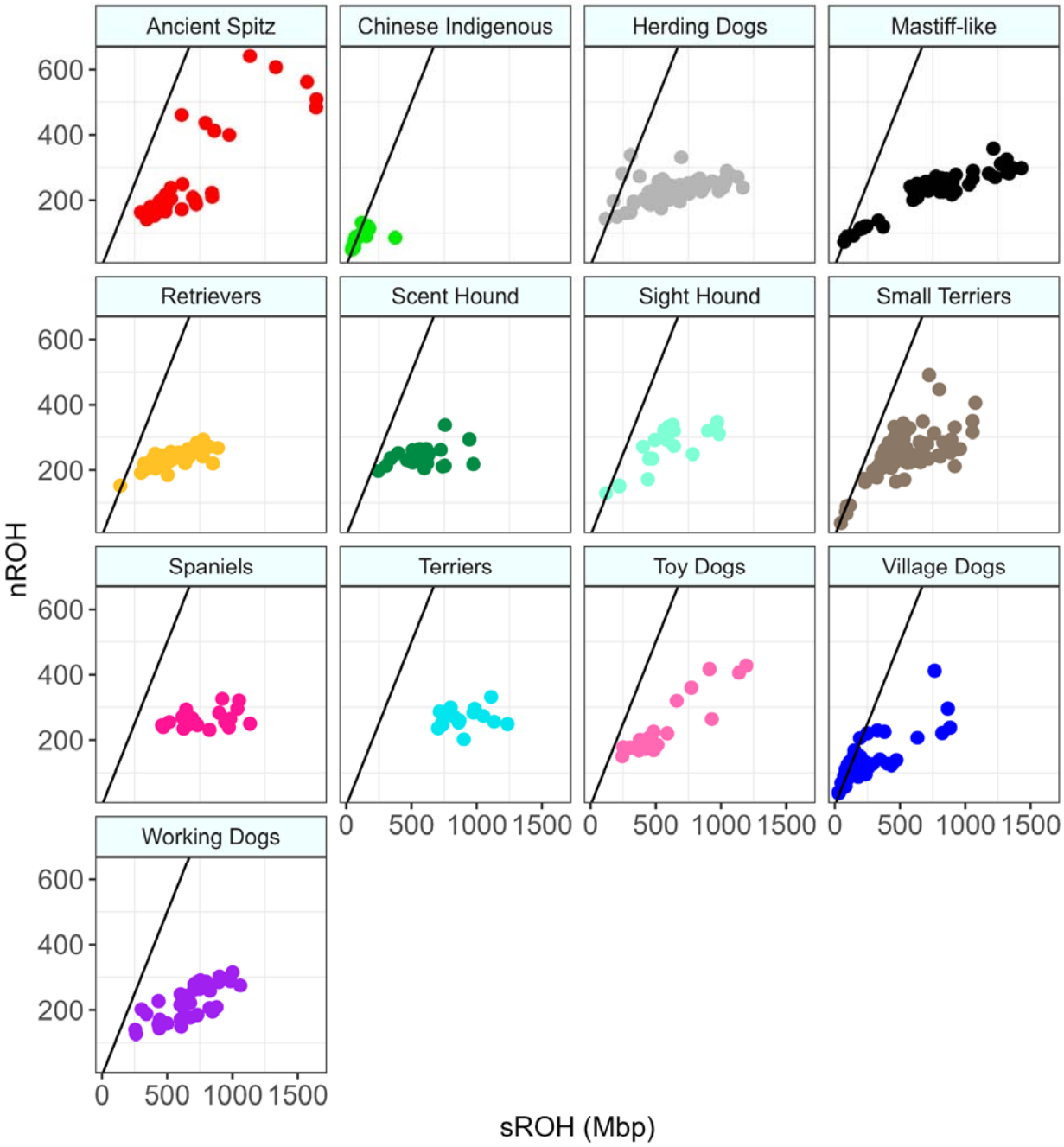


Fig. 3: Relationship between the sum of lengths of runs of homozygosity (s_{ROH}) and the number of runs of homozygosity (n_{ROH}) for each breed group. *The x-axis and the y-axis represent the sROH and nROH, respectively. The black line is y=x.*

bioRxiv preprint doi: https://doi.org/10.1101/2024.05.01.592072; this version posted September 15, 2025. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

Next, we examined the level of inbreeding within clades by computing the inbreeding coefficient, F_{ROH}, for each individual and each breed group (Tables 3 & S2). To account for the distribution of F_{ROH} across all breed groups, we computed mean F_{ROH} (Table 3). High levels of mean F_{ROH} were observed in breed dogs' clades. F_{ROH} mean values in breed dogs ranged from 0.41 ± 0.07 in Terriers to 0.24 ± 0.07 in Retrievers. The lowest mean values were observed in non-breed dogs, namely Chinese Indigenous Dogs (0.05 ± 0.04) and Village Dogs (0.09 ± 0.08). This result was expected and provides support for breed dogs sharing a more recent common ancestor and more background relatedness relative to village or Chinese indigenous dogs [3].

Table 3: F_{ROH} across all breed groups. *The table represents the mean ± standard deviation of F_{ROH} for each breed group.*

Breed Group	**Mean F_{ROH}**
Terriers	0.41 ± 0.07
Spaniels	0.35 ± 0.09
Mastiff-Like	0.34 ± 0.15
Ancient Spitz	0.33 ± 0.17
Working Dog	0.31 ± 0.09
Herding	0.29 ± 0.11
Scent Hound	0.27 ± 0.08
Toy Dogs	0.26 ± 0.13
Sight Hound	0.26 ± 0.10
Small Terrier	0.24 ± 0.10
Retriever	0.24 ± 0.07
Village Dogs	0.09 ± 0.08
Chinese Indigenous Dogs	0.05 ± 0.04

bioRxiv preprint doi: https://doi.org/10.1101/2024.05.01.592072; this version posted September 15, 2025. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

Association between F_{ROH} and phenotypic traits

Next, we wondered whether F_{ROH} could be used to detect non-additivity in quantitative (non-disease) phenotypes in dogs. Following an approach that was recently used in disease phenotypes in dogs [37] and quantitative and disease phenotypes in humans [32, 33, 44–48], we searched for non-additive effects by correlating F_{ROH} and 13 breed averaged phenotypes (bulky, drop ears, furnish, hairless, height, large ears, length of fur, lifespan, long legs, muscled, weight, white chest, and white head). The three quantitative phenotypes had significant associations with F_{ROH}: breed average height ($\beta = 0.31$ and $p = 4.05 \times 10^{-3}$), breed average weight ($\beta = 0.19$ and $p = 1.19 \times 10^{-2}$), and breed average lifespan ($\beta = -0.14$ and $p = 4.44 \times 10^{-2}$) (Fig. 4). In breed dogs, we found that as F_{ROH} increased by 1%; breed average height increased by 0.192 cm, breed average weight increased by 0.14 kg, and breed average lifespan decreased by 0.014 years. The statistical analysis, including normalized coefficient effect size (either β or log odds), p-values, and confidence intervals for quantitative traits (breed average height, breed average weight, and breed average lifespan) are shown in Table S3. Significant associations with F_{ROH} were present in 5 out of the 10 remaining qualitative phenotypes: muscled ($\beta = 7.96$ and $p = 9.68 \times 10^{-11}$), white chest ($\beta = 3.49$ and $p = 5.36 \times 10^{-5}$), white head ($\beta = 4.26$ and $p = 1.55 \times 10^{-4}$), length of fur ($\beta = -1.88$ and $p = 1.27 \times 10^{-2}$), and furnish ($\beta = -2.50$ and $p = 8.24 \times 10^{-3}$) (Fig. 5, Table S4). Across all breed groups, as F_{ROH} increased, white chest, white head and muscled phenotypes also increased, whereas length of fur and furnish phenotypes decreased.

bioRxiv preprint doi: https://doi.org/10.1101/2024.05.01.592072; this version posted September 15, 2025. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

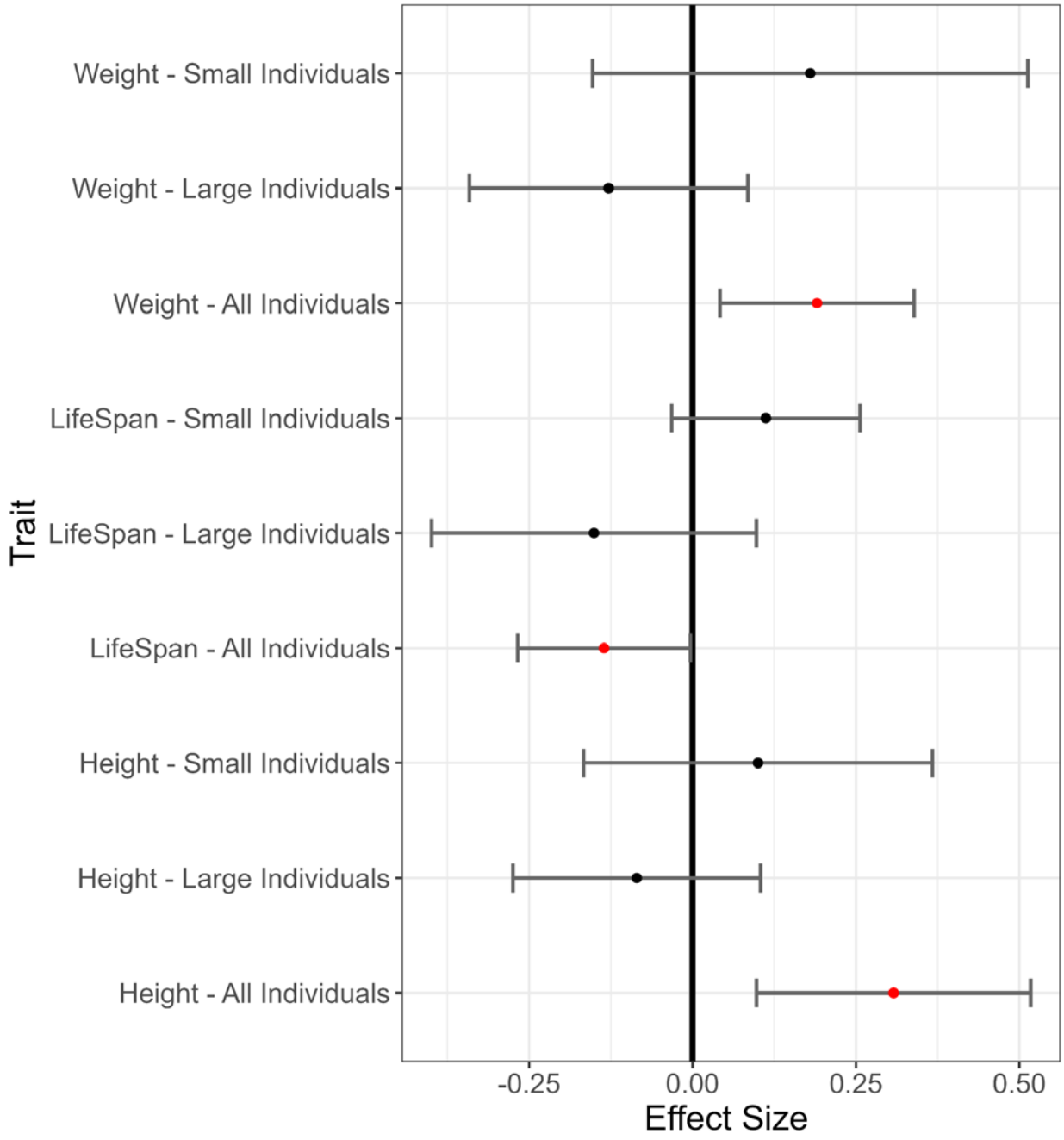


Fig. 4: Relationship between effect size and breed averaged phenotypic traits. *The y-axis represents the phenotypes for which the associations were tested. These are binned into 3 categories (small individuals, large individuals, and all individuals). The x-axis represents the normalized beta coefficients (β) for each trait. Significant effects of* F_{ROH} *on a phenotype (nominal* $p < 0.05$*) are indicated by a red point. An effect size larger than 0 indicates an increase of that trait value with respect to* $F_{ROH,}$ *and less than 0 represents a decrease in that trait value.*

bioRxiv preprint doi: https://doi.org/10.1101/2024.05.01.592072; this version posted September 15, 2025. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

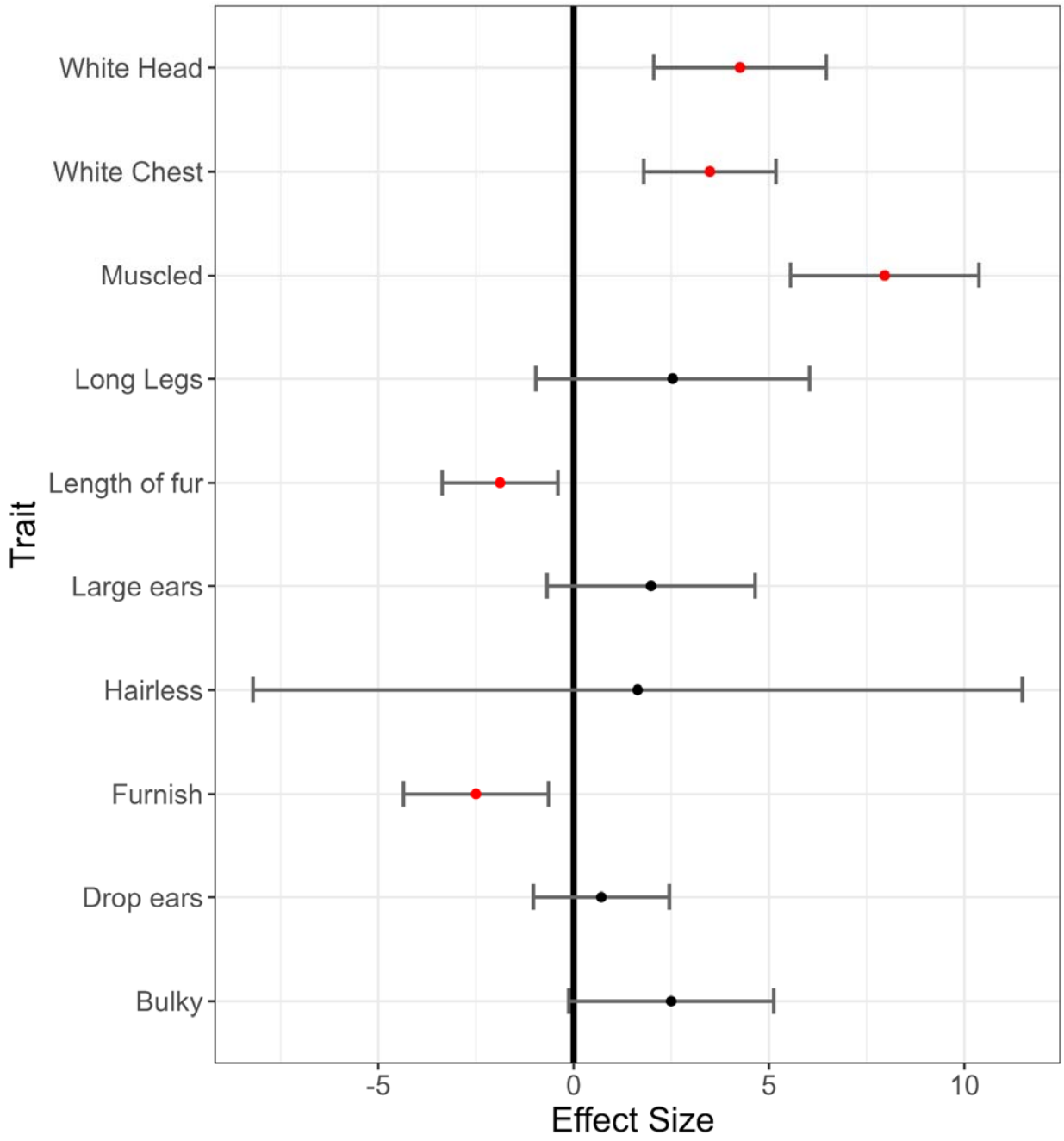


Fig. 5: Relationship between effect size and categorical phenotypic traits. *The y-axis represents the phenotypes for which the associations were tested. The x-axis represents the log odds for each trait. A significant effect of F_{ROH} on a phenotype (nominal $p < 0.05$) is indicated with a red point. An effect size larger than 0 indicates an increase of that trait with respect to $F_{ROH,}$ and less than 0 represents a decrease in that trait.*

bioRxiv preprint doi: https://doi.org/10.1101/2024.05.01.592072; this version posted September 15, 2025. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

ROH-mapping GWAS

Finally, we sought to identify the ROH-associated genomic regions that influence our quantitative traits of interest (weight, lifespan, or height). Thus, we performed a GLMM-based GWAS (Fig. 6, Fig. 7, Fig. 8, and Fig. S17-S25) using presence/absence of ROH across individuals as the phenotype predictor for each locus. For exact significant threshold values, see Tables S4 - S5.

For breed average height, we found 27 SNPs above the Suggestive-Wide Significance (SWS) threshold and 18 SNPs above the Genome-Wide Significance (GWS) threshold when testing for all individuals in our sample (Fig. 6A). Notably, a single SNP on chromosome 24 (position 22647289) had the strongest associated genomic signal with height ($p = 3.46 \times 10^{-6}$). This region corresponds with the gene *SNTA1*. Additionally, genes *CERS3* (7 SNPs) on chromosome 3, *CBFA2T2* (16 SNPs) on chromosome 24, and *SNTA1* (4 SNPs) on chromosome 24 were observed to be above the SWS threshold and contained multiple SNPs within their transcription regions (Table S6). SNPs linked to these three genes were also above the GWS threshold.

For small individuals, 34 SNPs were observed to be above the SWS threshold and 17 SNPs were observed to be above the GWS threshold when tested for association with height (Fig. 6B). A single SNP on chromosome 9 (position 45711401) had the strongest associated genomic signal ($p = 1.44 \times 10^{-6}$), which corresponds with the gene *PITPNA.* Across our small subgroup we identified additional genomic variants above the SWS threshold for weight: on chromosome 3, *CERS3* (3 SNPs), *IGF1R* (1 SNP), *LRRC28* (1 SNP), and *TTC23* (3 SNPs); on chromosome 9, *PITPNA* (4 SNPs), *SLC43A2* (2 SNPs), *VPS53* (2 SNPs), and *RPH3AL* (4

bioRxiv preprint doi: https://doi.org/10.1101/2024.05.01.592072; this version posted September 15, 2025. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

SNPs); on chromosome 18, *HGF* (2 SNPs); on chromosome 22, *COMMD6* (1 SNP); on chromosome 24, *CBFA2T2* (7 SNPs) and *SNTA1* (4 SNPs) (Table S6). Only SNPs associated with *TTC23*, *PITPNA*, *SLC43A2*, *RPH3AL*, CBFA2T2, and SNTA1 were above the GWS threshold. For large individuals, a single SNP on chromosome 11 (position 9943557), which is linked to *PRR16*, was observed to be above the SWS threshold for height (Fig. 6C). For an extended table containing all the observed genes associated with height refer to Table S6.

bioRxiv preprint doi: https://doi.org/10.1101/2024.05.01.592072; this version posted September 15, 2025. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

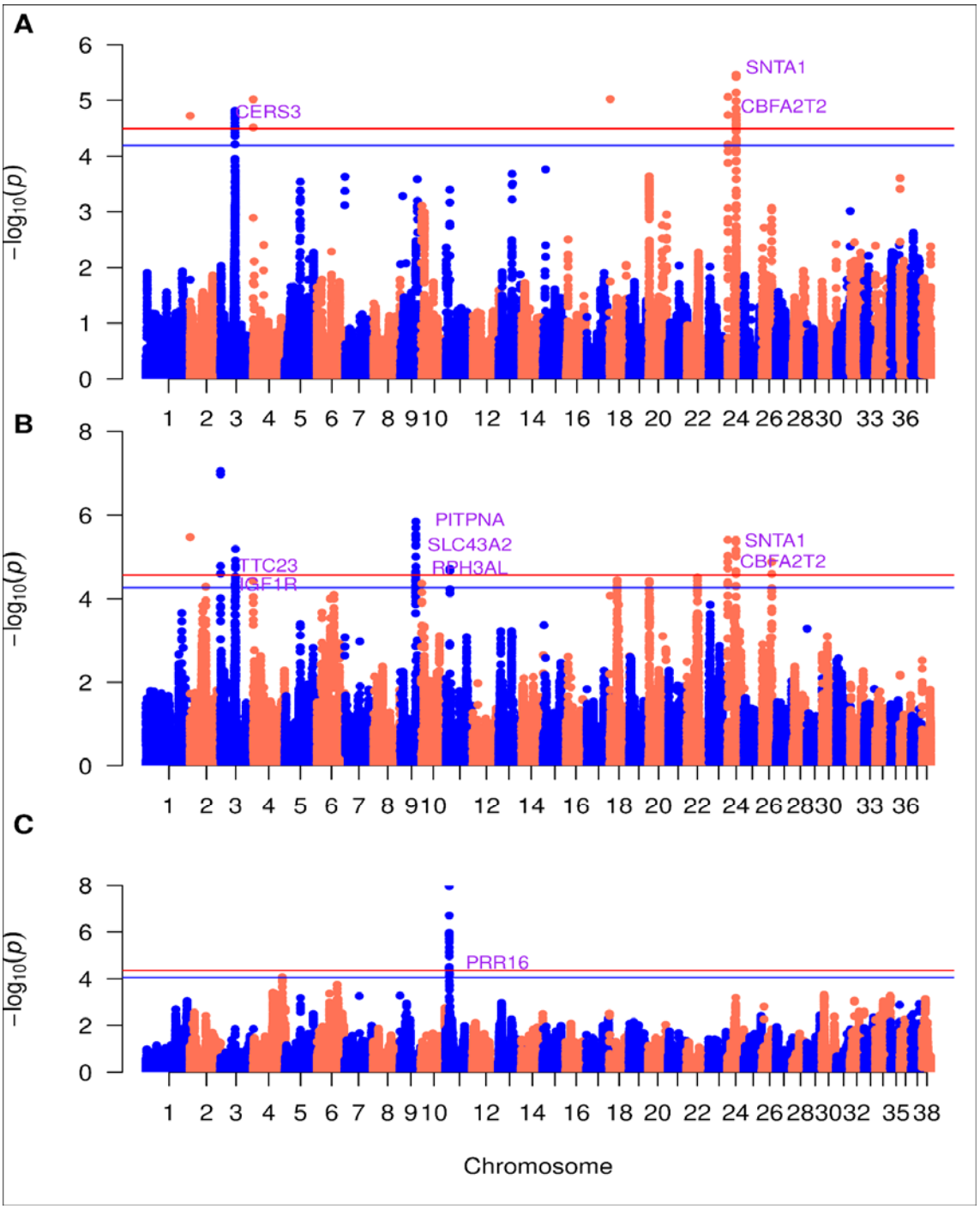


Fig. 6. Manhattan plots for GLMM-based GWAS using presence/absence of ROH for body height. *A. All individuals B. Small Individuals C. Large Individuals. The x-axis represents the genomic position. The y-axis represents the log10 base transformed p-values. Single nucleotide polymorphisms are represented by a single point. The red horizontal line indicates the genome wide significance (GWS) threshold, and the blue horizontal line indicates the suggestive wide significance (SWS) threshold. Only genes above the GWS threshold are labeled, with the exception of IGF1R. Marked genes are not drawn to scale.*

bioRxiv preprint doi: https://doi.org/10.1101/2024.05.01.592072; this version posted September 15, 2025. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

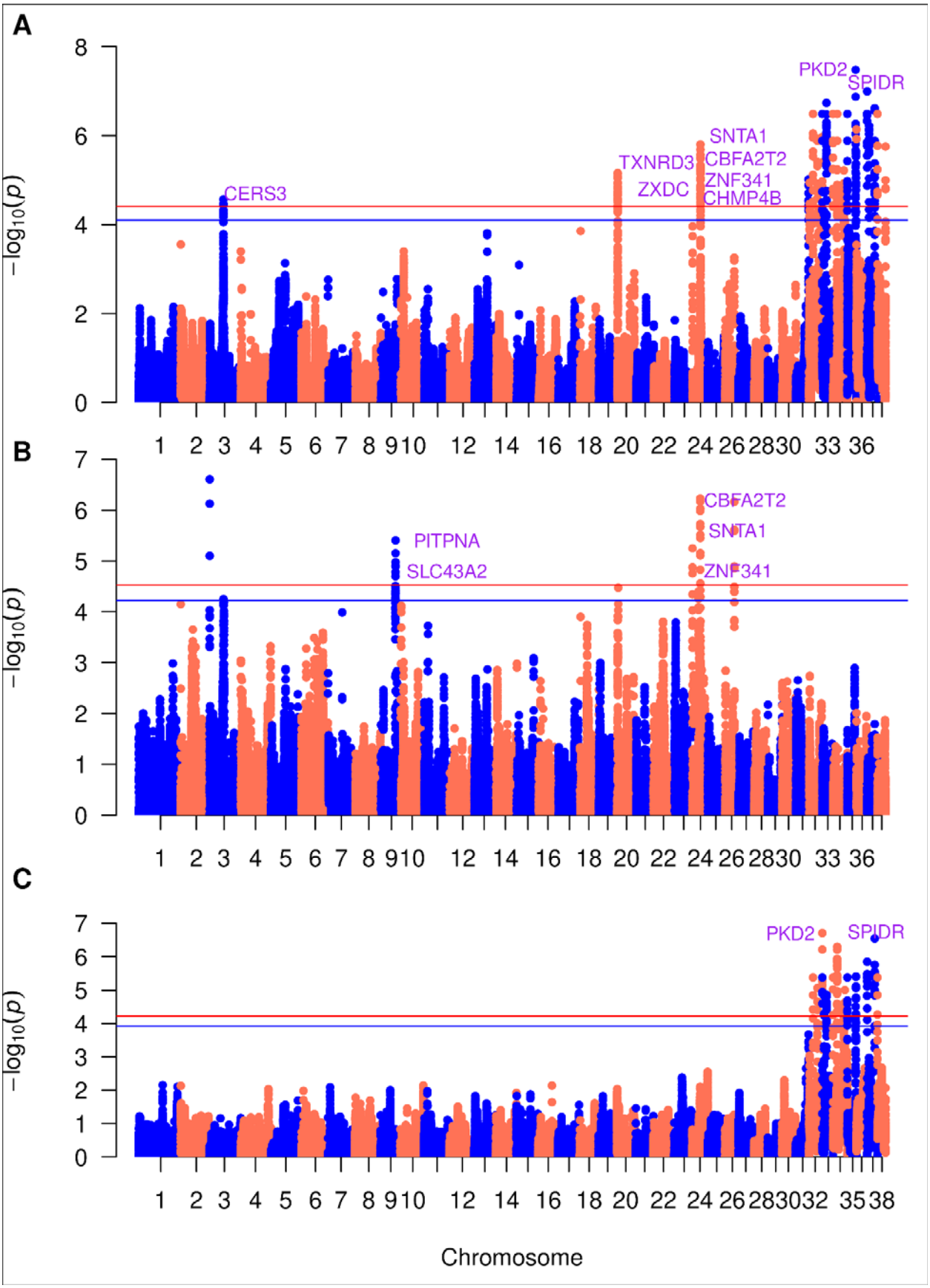
For breed average weight, 66 SNPs were observed above the SWS threshold and 45 SNPs above the GWS threshold, when including all individuals in our sample (Fig. 7A). Nearly half of the SNPs above the SWS threshold (30 SNPs) were found within genes on chromosome 24. A single SNP, on chromosome 32 (position 11413601) had the strongest association with weight ($p = 3.24 \times 10^{-7}$). This was observed within *PKD2*, a gene that encodes for protein kinase D2. Across all samples we identified several other SNPS to be above the SWS threshold for weight: on chromosome 3, *CERS3* (7 SNPs); on chromosome 20, *ZXDC* (3 SNPs) and *TXNRD3* (1 SNP); on chromosome 24, *SNTA1* (5 SNPs), *CBFA2T2* (19 SNPs), *ZNF341* (2 SNPs), and *CHMP4B* (4 SNPs); on chromosome 32, *SEPT11* (1 SNP), *CCNI* (1 SNP), *PKD2* (11 SNPs), *SGMS2* (4 SNPs) and *CYP2U1* (3 SNPs); on chromosome 35, *SPIDR* (5 SNPs) (Table S7). The genes on chromosomes 3, 20, 24, and 35 all had SNPs above the GWS threshold. The only gene from chromosome 32 with SNPs above the GWS threshold was *PKD2*.

For small individuals, 31 SNPs were observed above the SWS threshold and 19 SNPS above the GWS threshold (Fig. 7B). Nearly a third of those above the SWS threshold (i.e. 20 SNPs) belong to chromosome 24. A single SNP on chromosome 24 (position 22660667) had the strongest association with weight ($p = 5.88 \times 10^{-7}$) and is linked to the protein-coding gene *SNTA1* (Fig. 7B). For our subgroup of small individuals, we found SNPs above the SWS threshold for weight to be linked to the following genes: on chromosome 3, *CERS3* (1 SNP); on chromosome 9, *RPH3AL* (4 SNPs), PITPNA (4 SNPs), and SLC43A2 (2 SNPs); on chromosome 24, *ADAM33* (2 SNPs), *SIGLEC1* (2 SNPs), *HSPA12B* (1 SNP), *SNTA1* (4 SNPs), *CBFA2T2* (9 SNPs), *ZNF341* (1 SNP) and *CHMP4B* (1 SNP) (Table S7). SNPs on genes *PITPNA*, *SLC43A2*, *SNTA1*, *CBFA2T2*, and *ZNF341* were above the GWS threshold. In contrast, for the subgroup of large individuals, 10 SNPs were above the SWS threshold and 5 SNPs were above the GWS

bioRxiv preprint doi: https://doi.org/10.1101/2024.05.01.592072; this version posted September 15, 2025. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

threshold (Fig. 7C). Only two genes reached significance in large individuals: on chromosome 32, *PKD2* (4 SNPs) and on chromosome 35, *SPIDR* (6 SNPs). All the SNPs linked to these two genes were above the SWS threshold for weight (Table S7). SNPs on both of these genes were also above the GWS threshold. For an extended table containing all the observed genes associated with weight refer to Table S7.

bioRxiv preprint doi: https://doi.org/10.1101/2024.05.01.592072; this version posted September 15, 2025. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.



bioRxiv preprint doi: https://doi.org/10.1101/2024.05.01.592072; this version posted September 15, 2025. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

Fig. 7. Manhattan plots for GLMM-based GWAS using presence/absence of ROH for body weight. *A. All individuals B. Small Individuals C. Large Individuals. The x-axis represents the genomic position. The y-axis represents the log10 base transformed p-values. Single nucleotide polymorphisms are represented by a single point. The red horizontal line indicates the genome wide significance (GWS) threshold, and the blue horizontal line indicates the suggestive wide significance (SWS) threshold. Only genes above the GWS threshold are labeled. Marked genes are not drawn to scale.*

For breed average lifespan, 10 SNPs were observed above the SWS threshold and only 4 SNPs were above the GWS threshold when tested across all individuals (Fig. 8A). The strongest associated SNP ($p = 2.21 \times 10^{-5}$) was located on chromosome 20 (position 789204) within the gene *TXNRD3*. Across all samples we identified several other SNPS to be above the SWS threshold for lifespan: on chromosome 20, *ZXDC* (2 SNPs) and *TXNRD3* (1 SNP); on chromosome 32, *SEPT11* (1 SNP); on chromosome 33, *DRD3* (5 SNPs) and *TIGIT* (1 SNP) (Table S8). All, but *SEPT11*, had one SNP above the GWS threshold. For our subgroup of small individuals, we only found a gene on chromosome 12 *RUNX2* (4 SNPs), to be above the SWS threshold. (Fig. 8B and Table S8). For our subgroup of large individuals, we found 9 SNPs to be above the SWS threshold, all of which were also above the GWG threshold (Fig. 8C). The strongest associated SNP ($p = 1.17 \times 10^{-8}$) was located on chromosome 20 (position 789204) within the gene *TXNRD3*. All genes with SNPs above the significance thresholds (*TXNRD3* (1 SNP), *ZXDC* (3 SNPs), *CFAP100* (4 SNPs), and *SLC41A3* (1 SNP)) were located on chromosome 20 (Table S8). For an extended table containing all the observed genes associated with lifespan refer to Table S8.

bioRxiv preprint doi: https://doi.org/10.1101/2024.05.01.592072; this version posted September 15, 2025. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

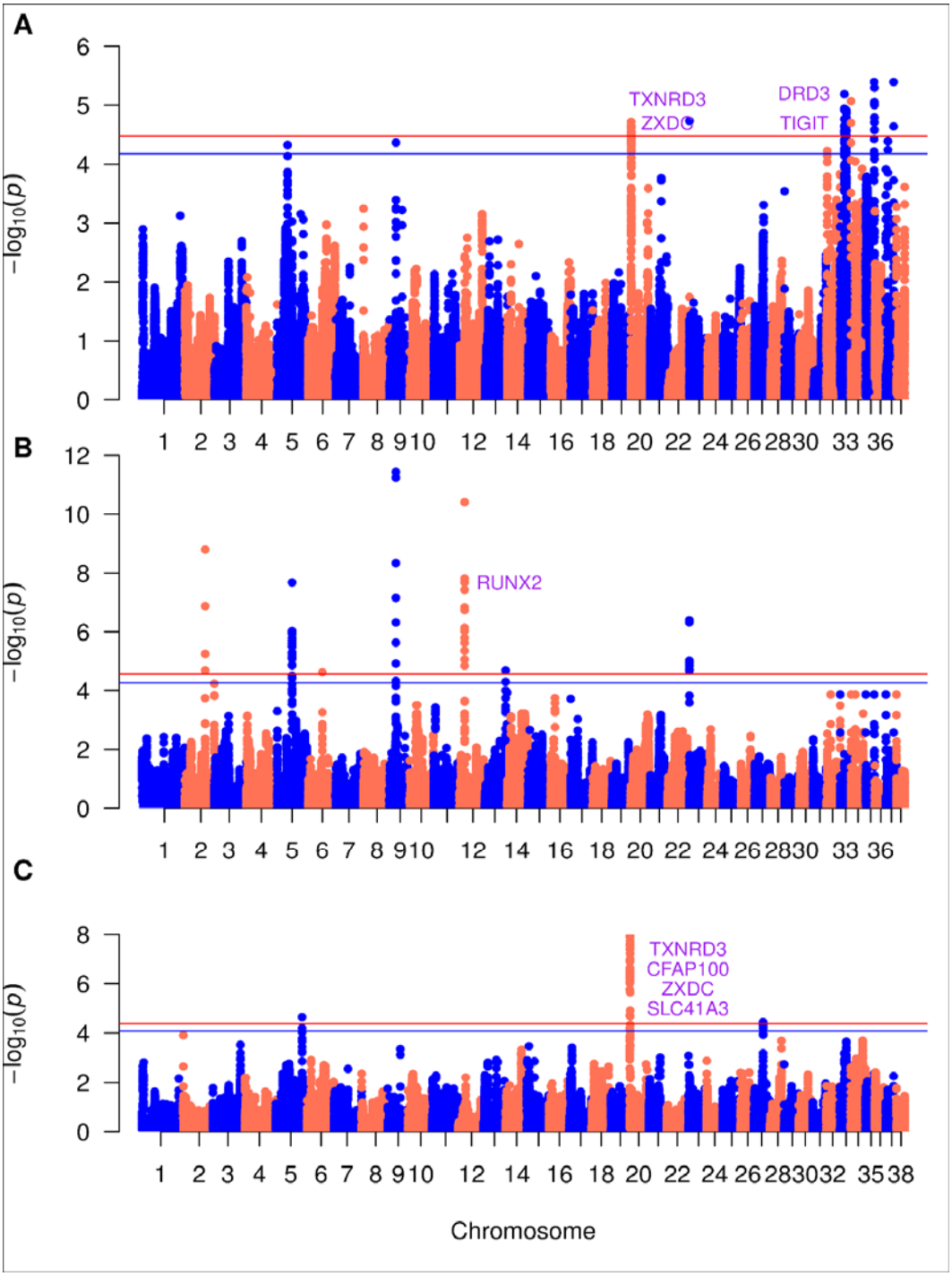


Fig. 8. Manhattan plots for GLMM-based GWAS using presence/absence of ROH for lifespan. *A. All individuals B. Small Individuals C. Large Individuals. The x-axis represents the genomic position. The y-axis represents the log10 base transformed p-values. Single nucleotide polymorphisms are represented by a single point. The red horizontal line indicates the genome wide significance (GWS) threshold, and the blue horizontal line indicates the suggestive wide*

bioRxiv preprint doi: https://doi.org/10.1101/2024.05.01.592072; this version posted September 15, 2025. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

significance (SWS) threshold. Only genes above the GWS threshold are labeled. Marked genes are not drawn to scale.

Conclusions

Our work provides in-depth analysis of patterns of runs of homozygosity (ROH) in domesticated breed dogs and non-breed domesticated village and Chinese Indigenous dogs, highlighting how the domestication and breed formation has shaped genetic diversity and trait architecture. We demonstrate that ROH, quantified as the total fraction of the genome within a run of homozygosity, F_{ROH}, can be used to uncover traits that are not fully additive. We focus on ROH because these genomic segments reflect recent demography and inbreeding [49, 50]. For breed dogs, there is a large fraction of the genome within long ROH because of the small number of individuals used for breed establishment. These long ROH are a result of very recent parental relatedness and are inextricably linked to this species' complex evolutionary, domestication, and breed formation history.

Previous research has shown that dogs originated from an isolated wolf population(s), and recent strong artificial selection drove breed emergence [3, 51]. Strong artificial selection during breed formation resulted in low genetic diversity and high homozygosity within breeds, alongside the large phenotypic variance between breeds [38]. Artificial selection, in the case of dogs, was achieved through inbreeding, and also led to the significant sharing of common genetic variants [52]. Artificial selection for specific phenotypic characteristics during breed establishment and subsequent breed standardization have resulted in homogeneity of both the phenotypic and genotypic variation within a breed. Further, when looking between breeds, previous work has shown that occasionally a dog with a desirable trait in one breed is also used to introduce the same phenotype in another breed, which creates a network of genetic relatedness

bioRxiv preprint doi: https://doi.org/10.1101/2024.05.01.592072; this version posted September 15, 2025. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

through shared common haplotypes between breeds [53]. In sum, because artificial selection was often achieved through inbreeding, and breed standards reduced the effective population size of each breed, we expect both phenomena to be reflected through shared haplotypic information between breeds and distinct patterns of homozygosity within breeds.

For example, we show that domesticated breed dogs have unique patterns of ROH in their genome, and on average carry more of their genome within long ROH than domesticated non-breed village and Chinese indigenous dogs (Table 2). Conversely, the village dogs and Chinese indigenous dogs which did not experience selection from breed formation have the lowest values for mean ROH distribution. Taken together our results correspond with the higher degree of relatedness among all domestic breed group individuals. This was expected given the known high levels of inbreeding during breed establishment. Our results also highlight village and Chinese indigenous dogs as more outbred populations given lower ROH proportion when compared to breed dogs. Specifically, we observe lower levels of mean n_{ROH} and mean s_{ROH} in Chinese indigenous dogs. This is in line with the origins of the Chinese indigenous dogs, which is highlighted by relaxed trait selection during establishment of the population [54]. The relaxed selection may have resulted in overall higher levels of genetic diversity. Additionally, village dogs also harbor lower values for mean n_{ROH} and mean s_{ROH} (Table 1 and Table 2), consistent with them having not experienced an additional bottleneck during breed formation [37].

Turning to ROH hotspots, we found that there was some relationship with recombination rate per chromosome and the density of ROH hotspots, though recombination certainly does explain some of the observed hotspot and coldspots, it does not fully explain all of the patterns that we observed (Fig. S1 & S4). For example, on many chromosomes, we observe at least one ROH hotspot (across breed dogs and village dogs) in a region where recombination rates peaks.

bioRxiv preprint doi: https://doi.org/10.1101/2024.05.01.592072; this version posted September 15, 2025. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

This is also consistent with the conclusions from previous studies which suggest recombination rates are not the only factor leading to ROH hotspots or coldspots [36, 37]. Our results show that the ROH patterns we observe are likely driven by demography (domestication), artificial selection (breed formation), and inbreeding. These processes have resulted in unique ROH patterns within breeds, sharing across breeds, and very little ROH sharing between breed dogs and non-breed village dogs.

We also examine the relationship between ROH and non-disease traits. We observe that the relationship between breed average height and F_{ROH} is positive when using all breed dogs (Fig. 4). This positive effect size opposes recently published work in human populations [55–57]. This may be a result of combining the effect sizes of multiple sizes of dogs (i.e. large and small dogs). When we partition by weight, our results correspond to previous studies on height in breed dogs [58]. As one might expect, we also find a positive β when testing for the relationship between F_{ROH} and weight. We also observe a negative β with respect to the relationship between F_{ROH} and lifespan, indicating an association with inbreeding and survival to old age. This result is consistent with the findings in dogs where an association with disease phenotypes was observed [37]. Thus, when we partition dogs by weight [59], we find a trend where large dog breeds have a shorter lifespan ($\beta = -0.15$) and small dogs have a longer lifespan ($\beta = 0.11$). However, our results were not significant (Table S3). Previous studies have linked size with longevity, and this remains an interesting area for future work [60]. When we include lifespan as a covariate when associating F_{ROH} and breed average height in all individuals, we still find a nominally significant, but larger p-value, of 0.034. Further, the effect size of F_{ROH} on height decreases from $\beta \sim 0.31$ to $\beta \sim 0.19$. We also observe a negative ($\beta \sim$ -0.83) and significant relationship between height and lifespan ($p = 6.41 \times 10^{-41}$).

bioRxiv preprint doi: https://doi.org/10.1101/2024.05.01.592072; this version posted September 15, 2025. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

Our GWAS using ROH identifies several significant associations between SNPs and the quantitative traits: height, weight, and lifespan (Table S6, Table S7, and Table S8). When utilizing data for all individuals, there were three hits *CERS3* (chromosome 3), *CBFA2T2* (chromosome 24) and *SNTA1* (chromosome 24) above the GWS threshold for height. In previous studies, these three genes were found to harbor genetic variants associated with human height [61, 62]. Additionally, human genetics studies have found *CERS3* to be linked with Body-Mass Index (BMI) [63]. We see that many of the significant genic hits within our subgroup of small individuals (*PITPNA*, *LRRC28*, *TTC23*, *HGF*, *SLC43A2*, and *VPS53*) are already associated with human height [62, 64]. We also observe a new relationship between *RPH3AL* and height. Though there is limited research on *RPH3AL*, its role in exocytosis and the secretion of growth hormones in humans has been noted [65]. Importantly, the signal within *IGF1R* replicated, and was above the SWS threshold. A mutation in *IGF1R* was linked with height in small dog breeds, like Chihuahuas [58]. Lastly for small individuals, *COMMD6* was shown to be associated with height. *COMMD6* has primarily been studied in relation to immune function, and to our knowledge, no studies to date have reported any associations between this gene and height [66]. In the subgroup of large individuals, the only gene above the SWS threshold was *PRR16* (chromosome 11), which has been previously associated with human height [62].

For weight, across all our samples, our GWAS results pinpoint *PKD2* (chromosome 32) to be the most strongly associated. Mouse models have shown the presence of *PKD2* has been linked to an increase in fat absorption, potentially promoting obesity [67]. Several other genes, with significant associations in our study, have been linked to weight in other species. *SNTA1* was found to be associated with the Live Body Weight (LBW) of goats, which is a factor often used to assess livestock health [68]. Another gene, *CBFA2T2*, was shown to regulate

bioRxiv preprint doi: https://doi.org/10.1101/2024.05.01.592072; this version posted September 15, 2025. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

adipogenesis in both humans and mice, which can affect obesity when its expression is altered [69–71]. *TXNRD3* influences adipocyte differentiation through its involvement in the Wnt signaling pathway, a key process in the regulation of body fat and energy storage [72–74]. *CERS3*, also associated with height, has been linked with Body Mass Index (BMI) in humans [63]. Additionally, the genes *SPIDR*, and *ZNF341* have been associated with total body fat and bone mineral content, respectively [75–77]. *CHMP4B* has been linked to the human birth weight, no studies have directly linked it to obesity and/or body mass. [78, 79]. Similarly, *ZXDC* has primarily been investigated in the context of immune system regulation and cancer biology. To our knowledge, this is the first association with weight thus far, which provides an additional avenue for future researchers to explore. In our subgroup of small individuals, we identify similar associations between weight and the genes *SNTA1*, *ZNF341*, and *CBFA2T2*. Additionally, we identify new associations specific to this subgroup with *PITPNA* (also associated with height) and *SLC43A2*. Both of these genes were shown to decrease body weight in knockout mice [80, 81]. In the subgroup of larger individuals, we find two genes to be associated with weight, *PKD2* and *SPIDR*, both of which are associated with body fat.

For lifespan, we highlight several associations of interest. When testing all samples, we find significant SNPs linked to the genes: *TXNRD3, DRD3, TIGIT, SEPT11* and *ZXDC*. *TXNRD3* has been previously associated with lower survival rates for various types of cancers [82]. However, to our knowledge, no previous studies have found an explicit connection between *TXNRD3* and lifespan. The gene that encodes the D3 dopamine receptor, *DRD3*, is associated with schizophrenia, which has too been linked with decreased life expectancy [83, 84]. We observe significant signals at *TIGIT* and *SEPT11*, which both have been tied to the promotion of tumor growth and reduced longevity [85–87]. *ZXDC* was previously associated with cervical

bioRxiv preprint doi: https://doi.org/10.1101/2024.05.01.592072; this version posted September 15, 2025. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

cancer metastasis, which has a significant negative impact on patient survival [88]. In small dogs, we saw an association with *RUNX2*, which has also been associated with multiple cancers and poor patient prognosis [89]. In large individuals, our GWAS revealed *SLC41A3* to be associated with lifespan. A relationship between this gene and liver hepatocellular carcinoma, which is the second primary contributor of carcinoma-associated death, has been established [90]. Another gene found in the group of large individuals is *CFAP100*, which to our knowledge has not previously been linked to lifespan.

Overall, our GWAS suggests that quantitative traits in dogs are polygenic with multiple variants associated with a trait, and some traits are not entirely additive. The newly identified genes, which have not been observed in previous studies, could be used as candidates in future functional exploration. The presence of these genes also highlights new avenues to explore complex trait architecture and non-additivity in domesticated species. In humans, there have been 25,551 associations with height, 2,233 associations with weight, and 664 associations with lifespan as of 2023 [91]. After correcting for inflation, we recover a total of 50 associations with height, 84 associations with weight, and 21 associations with lifespan. This corresponds to previous work, which has suggested that dogs have a more simplified complex trait architecture than humans. [38, 92].

Importantly, we conducted an additional test to validate our ROH based GWAS by attempting to replicate a previously identified peak that was known to be homozygous on chromosome 13 within RSPO2 for furnishings in dogs [29]. We can identify the same peak on chromosome 13 for RSPO2 (Fig. S26), though it did not meet genome-wide or suggestive significance after p-values were corrected (Fig. S27). In addition to this peak, we observe additional peaks that had not been previously identified. Replicating this peak suggests our

bioRxiv preprint doi: https://doi.org/10.1101/2024.05.01.592072; this version posted September 15, 2025. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

approach, which corresponds to different information, is powerful and worthwhile. Notably, similar power in human GWAS requires hundreds of thousands or even a million individuals [61, 62]. Here, we capture non-additivity when using ROH-mapping GWAS, and we can capture this information with less than 1000 dogs. This is due to the unique domestication and breed formation bottlenecks that dogs experienced that simplified complex trait architecture [29, 38, 43, 93–95].

There are some limitations in the work that we would like to highlight. First, the phenotypic values used in the analysis were calculated based on the breed average, restricting us to study associations between breeds. In other words, the within breed values would not yield any significant results as the expected association between the trait and F_{ROH} would be 0, due to using a breed average phenotype. Using the breed average has been previously validated [29, 38, 43, 93–95] but the assumption does come with caveats. For example, when using a breed average value, we are ignoring variance among individuals within a breed. This variance within the breed is quite small due to the strict standards of the AKC. To this end, we mathematically show that if the within breed variance is small, the effect size computed from individual level data will be equivalent to the effect size from using the average. However, bias can be introduced in some cases when aggregating across breeds. Secondly, breed group categorization based on both neighbor-joining trees and AKC could potentially still introduce errors when grouping some breeds. Additionally, we did observe that there was a concentration of peaks in the last few chromosomes (Fig. S8, Table S9). We believe that the large number of peaks on the smaller chromosomes could be due to stronger linkage disequilibrium in the smaller chromosomes, thus amplifying the strength of the signal. Alternatively, some chromosomes (such as chromosome 32) have a higher GC content when examining the CanFam3.1 assembly [96], which could

bioRxiv preprint doi: https://doi.org/10.1101/2024.05.01.592072; this version posted September 15, 2025. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

introduce noise into our results. Despite these limitations, the significant relationship between F_{ROH} and certain phenotypic traits suggest that inbreeding from domestication and artificial selection played a strong role in shaping complex non-disease traits within breed dogs.

In summary, our work has further elucidated the genetic structure of quantitative trait architecture in dogs. We also highlight how inbreeding, quantified in the form of ROH, occurred alongside selection for extreme phenotypes during breed formation. These ROH tag regions of the genome associated with complex non-disease phenotypes.

Methods

Data filtering and categorizations

In this study, we utilized whole genome sequences from 722 canids, which included various wild species, dingoes, and domestic dogs [29]. This data can be accessed via NCBI accession number PRJNA448733. All filtering was accomplished using BCFtools [97] where we retained only biallelic SNPs and sites of genotype quality score greater or equal to 20. Additionally, we removed any missing genotype rates above 10% and removed variants from sex chromosomes. Our filtered data consisted of 4,053,761 biallelic loci for all 38 pairs of autosomal chromosomes. Given the nature of our study, we removed all wild canids, mixed-breed individuals, and samples for which breed information was unavailable. Each remaining sample was categorized by common name or breed [29]. These classifications were informed by neighbor joining trees of domestic dogs [21] and American Kennel Club (AKC) groupings [98] . This resulted in 13 breed groups namely, Ancient Spitz (31 individuals and 11 breeds), Herding (76 individuals and 15 breeds), Mastiff-like (65 individuals and 15 breeds), Retrievers (50 individuals and 6 breeds),

bioRxiv preprint doi: https://doi.org/10.1101/2024.05.01.592072; this version posted September 15, 2025. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

Scent Hound (23 individuals and 10 breeds), Small Terrier (102 individuals and 8 breeds), Terriers (16 individuals and 7 breeds), Spaniels (22 individuals and 8 breeds), Toy Dogs (18 individuals and 7 breeds), Working Dogs (44 individuals and 10 breeds), Sight Hound (21 individuals and 9 breeds), Village Dogs (75 individuals and 13 regions), and Chinese Indigenous dogs (15 individuals). Overall, our working dataset included 558 individuals and 106 different breeds (Table S10).

Calling Runs of Homozygosity

To call runs of homozygosity (ROH) we used a likelihood-based inference method called GARLIC v1.6.0a [99]. This method uses a logarithm of odds (LOD) score measure of autozygosity, which is applied in a sliding window for the entire genome [100]. GARLIC requires input data with genotype and sample information in the form of TPED and TFAM files, which were generated using PLINK 2.0 [101] We generated a TGLS file to obtain per-genotype likelihoods. We employed genotype likelihood data in the form of GQ to account for errors in phred-scaled probability. Next, a window size of 100 SNPs was chosen based on SNP density (Fig. S28), with the window incrementally advancing by 10 SNPs at each step. GARLIC offers built-in ROH length classification, for which we defined 6 classes: <1 Mbp, 1-2 Mbp, 2-3 Mbp, 3-4 Mbp, 4-5 Mbp, and >5 Mbp. To account for sample size differences in allele frequency estimation, we set the number of resamples to 40. All other flags were set to default, and the following parameters were used: --auto-winsize --auto-overlap-frac --winsize 100 --centromere --size-bounds 1000000 2000000 3000000 4000000 5000000 --tped --tfam --tgls –gl-type GQ --resample 40 --out. To eliminate segments that were noisy, very short, and common, we filtered to retain ROH segments longer than 0.5 Mbp.

bioRxiv preprint doi: https://doi.org/10.1101/2024.05.01.592072; this version posted September 15, 2025. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

We conducted independent ROH calling for breeds with at least 10 individuals. Breeds consisting of fewer than 10 members were integrated into their respective breed groups (Table S11). Subsequently, ROH calls from all breeds, irrespective of individual numbers, were combined into their respective breed groups for further analysis. For each individual, we examined: 1) the relationship between coverage and F_{ROH} (Fig. S29) and 2) the relationship between the total number of runs of homozygosity (n_{ROH}) and total length of runs of homozygosity (s_{ROH}) (Fig. S30)).For each individual and breed group, we calculated the arithmetic mean and range of both total n_{ROH} and total s_{ROH}, while binning them based on ROH size classification. The longest total s_{ROH} (2141.22 Mbp) and the highest total n_{ROH} (1039) were identified in two individuals - PER00747 and PER00393, respectively - from the small terrier Yorkshire breed. PER00747 was sequenced at a low depth (~2x), which inflated s_{ROH}. To mitigate the effects of low depth PER00747 and PER00393 were removed from downstream analysis.

Inbreeding coefficient

For each sample and breed group, we computed the inbreeding coefficient, F_{ROH}, defined as the fraction of autosomal genome in ROH regions [102]:

$$F_{ROH} = \frac{s_{ROH}}{L_{Auto}}$$

Where s_{ROH} is the sum of the length of runs of homozygosity in an individual's genome and L_{Auto} is the length of autosomal genome in base pairs. An autosomal genome length of 2,203,764,842 base pairs was used since variants were called with the CanFam3.1 (NCBI RefSeq assembly: GCF_000002285.3) reference genome.

bioRxiv preprint doi: https://doi.org/10.1101/2024.05.01.592072; this version posted September 15, 2025. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

ROH Sharing Matrix

We generated an ROH sharing matrix, a square matrix that quantifies the sum of ROH overlaps between two individuals. This allowed us to analyze shared ROH patterns among dog breeds. To populate this matrix, we first generated individual ROH bed files for each sample. We then used BEDTools [103] to identify overlapping regions between pairs of ROH data. The following parameters were used: bedtools intersect -wao.

Phenotype data

We updated the phenotypic data [29] for 13 traits: bulky, drop ears, furnish, hairless, height, large ears, length of fur, lifespan, long legs, muscled, weight, white chest, and white head. For the three continuous traits, namely height, weight, and lifespan, we used average values obtained from AKC [98] (Table S12). We chose to use breed average values because this was previously shown to be a reliable measure for association tests [29, 38, 43, 93–95]. When sex-specific information was available, it was incorporated into the dataset; otherwise, we applied the same values for both males and females. We applied min-max scaling to normalize the three continuous traits and further categorized them into small and large groups based on the average weight (21.66kg) of all individuals (Tables S13-S15). For the remaining 10 traits, we averaged the values by the phenotypic data available for their respective breeds (Table S16). We used binary encoding to assign a value of 1 to individuals expressing the trait and 0 to those who did not.

bioRxiv preprint doi: https://doi.org/10.1101/2024.05.01.592072; this version posted September 15, 2025. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

Association tests

We computed the association between F_{ROH} and the 13 phenotypes using GMMAT, an R package that performs association tests with generalized linear mixed models (GLMM) [104]. To fit the GLMM, we used the built-in function glmmkin, which allowed us to examine the continuous traits (height, weight, and lifespan) as the quantitative phenotype traits. This also enabled us to include individual F_{ROH} as a covariate and use the ROH sharing matrix as a kinship matrix. We fit the model assuming a Gaussian distribution for the continuous phenotypes and used the identity link function. For associations between height, weight, and lifespan with F_{ROH}, we used the updated average phenotypic values of the breed for all 466 individuals (Fig. S31 and Supplementary Text). Based on our association tests and previous studies [105, 106], weight and F_{ROH} have been observed to have a significant association, thus we included them as covariates in our models. Weight and the interaction of weight and F_{ROH} were used as covariates across three data subsets (all individuals, small individuals, and large individuals) (Table S4). To fit the GLMM to the 10 binary traits, we specified the binomial distribution as the family and used the logit link function. Samples with no information available on the presence or absence of these phenotypes were removed from analysis. Due to lack of phenotype information, Chinese indigenous dogs and village dogs were excluded from the association tests.

ROH-mapping GWAS

To explore the biological mechanisms through which ROH-associated SNPs influence phenotypic traits, we conducted Genome-Wide Association Studies (GWAS). We generated a dataset that examined the occurrence of SNPs in ROH regions using BCFtools intersect and BCFtools subtract [97]. Any SNPs located outside of ROH regions were removed. Using

bioRxiv preprint doi: https://doi.org/10.1101/2024.05.01.592072; this version posted September 15, 2025. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

GMMAT, we fit GLMM models assuming a recessive genetic model for quantitative traits (height, weight, and lifespan), incorporating ROH-associated SNPs as covariates, an ROH sharing matrix for kinship structure, and a Gaussian distribution with an identity link function [104]. Additionally, given that breed structure has a particularly strong effect in dogs, we utilized a kinship matrix quantified by pairwise ROH sharing between individuals when associating traits with F_{ROH}. This procedure follows from an approach from a study on disease traits [37]. Lastly, we fit GLMM models using weight and the product of weight and F_{ROH} as covariates for the three data subsets (all individuals, small individuals, and large individuals) (Table S5). To obtain the effective sample size, adjusted for auto correlations between sites, we used the option effectiveSize within the coda R package [107]. To calculate the number of independent tests we used p-values for every SNP as the input, thus incorporating autocorrelation caused by LD. This approach was shown to be effective for correcting for structure previously [108]. To obtain the Genome-Wide Significance (GWS) and Suggestive-Wide Significance (SWS) thresholds, we used the Bonferroni correction with respect to effective sample size [109–111]:

$$\text{Genome-wide significance threshold} = -log10\left(\frac{0.05}{Effective\ Sample\ Size}\right)$$

$$\text{Suggestive-wide significance threshold} = -log10\left(\frac{0.1}{Effective\ Sample\ Size}\right)$$

To correct for population stratification, we calculated the genomic inflation factor, λ [112, 113]. We corrected the log-scaled p-values by dividing them by the genomic inflation factor (Table S5). To visualize the GWAS summary statistics, we used an R package, qqman to create Q-Q and Manhattan plots [114].

bioRxiv preprint doi: https://doi.org/10.1101/2024.05.01.592072; this version posted September 15, 2025. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

Accession Numbers

Whole genome sequence data is available on NCBI, accession number: PRJNA448733. The source phenotype data was obtained from [29]. All summary data are contained within the article and its supplementary information.

Declarations

The authors have nothing to declare.

Funding

This work was supported by the National Institute of General Medical Sciences of the National Institutes of Health under Award Number R35GM146926 (ZAS and S) and R35GM159982 (JAM, SN, and SH). This work was also supported by Eberly College of Science Startup Fund (ZAS and S). Computations for this research were performed using the Pennsylvania State University's Institute for Computational Data Sciences' Roar supercomputer. JAM, S, and SN were also supported by the startup funds from Dornsife College of Letters, Arts and Sciences through the Department of Quantitative and Computational Biology and the USC WiSE Gabilan Assistant Professorship.

Acknowledgements

Sweetalana would like to thank the entire Szpiech Lab for their invaluable support.

bioRxiv preprint doi: https://doi.org/10.1101/2024.05.01.592072; this version posted September 15, 2025. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

Availability of Data and Materials

All code is available at https://github.com/jaam92/QuantTraitInDogs and all data has been published.

Competing interests

The authors declare no competing interests.

Author Contributions

ZAS and JAM designed and supervised the research project. S performed the research and analyzed data. SN reviewed domestication literature for writing and editing of the manuscript. SH generated the proof and plots for the effect of using breed averages. S along with the assistance of SN, ZAS and JAM wrote the paper.

bioRxiv preprint doi: https://doi.org/10.1101/2024.05.01.592072; this version posted September 15, 2025. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

Bibliography

1. Darwin C. The variation of animals and plants under domestication. John Murray; 1868.

2. Nagasawa M, Mitsui S, En S, Ohtani N, Ohta M, Sakuma Y, et al. Oxytocin-gaze positive loop and the coevolution of human-dog bonds. Science. 2015;348:333–6. https://doi.org/10.1126/science.1261022.

3. Boyko AR. The domestic dog: man's best friend in the genomic era. Genome Biol. 2011;12:216. https://doi.org/10.1186/gb-2011-12-2-216.

4. Freedman AH, Gronau I, Schweizer RM, Vecchyo DO-D, Han E, Silva PM, et al. Genome sequencing highlights the dynamic early history of dogs. PLOS Genet. 2014;10:e1004016. https://doi.org/10.1371/journal.pgen.1004016.

5. Perri AR, Feuerborn TR, Frantz LAF, Larson G, Malhi RS, Meltzer DJ, et al. Dog domestication and the dual dispersal of people and dogs into the Americas. Proc Natl Acad Sci. 2021;118:e2010083118. https://doi.org/10.1073/pnas.2010083118.

6. Gray AP. Mammalian hybrids: a check-list with bibliography. 2d [rev.] ed. Slough: Commonwealth Agricultural Bureaux; 1972.

7. Vilà C, Savolainen P, Maldonado JE, Amorim IR, Rice JE, Honeycutt RL, et al. Multiple and ancient origins of the domestic dog. Science. 1997;276:1687–9. https://doi.org/10.1126/science.276.5319.1687.

8. Wayne RK. Molecular evolution of the dog family. Trends Genet. 1993;9:218–24. https://doi.org/10.1016/0168-9525(93)90122-X.

9. Meadows JRS, Kidd JM, Wang G-D, Parker HG, Schall PZ, Bianchi M, et al. Genome sequencing of 2000 canids by the Dog10K consortium advances the understanding of demography, genome function and architecture. Genome Biol. 2023;24:187. https://doi.org/10.1186/s13059-023-03023-7.

10. Ostrander EA, Wang G-D, Larson G, vonHoldt BM, Davis BW, Jagannathan V, et al. Dog10K: an international sequencing effort to advance studies of canine domestication, phenotypes and health. Natl Sci Rev. 2019;6:810–24. https://doi.org/10.1093/nsr/nwz049.

11. Larson G, Bradley DG. How much is that in dog years? The advent of canine population genomics. PLoS Genet. 2014;10:e1004093. https://doi.org/10.1371/journal.pgen.1004093.

12. Guagnin M, Perri AR, Petraglia MD. Pre-Neolithic evidence for dog-assisted hunting strategies in Arabia. J Anthropol Archaeol. 2018;49:225–36. https://doi.org/10.1016/j.jaa.2017.10.003.

13. Morey DF. Burying key evidence: the social bond between dogs and people. J Archaeol Sci. 2006;33:158–75. https://doi.org/10.1016/j.jas.2005.07.009.

bioRxiv preprint doi: https://doi.org/10.1101/2024.05.01.592072; this version posted September 15, 2025. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

14. Davis SJM, Valla FR. Evidence for domestication of the dog 12,000 years ago in the Natufian of Israel. Nature. 1978;276:608–10. https://doi.org/10.1038/276608a0.

15. Tchernov E, Valla FF. Two new dogs, and other Natufian dogs, from the Southern Levant. J Archaeol Sci. 1997;24:65–95. https://doi.org/10.1006/jasc.1995.0096.

16. Benecke N. Studies on early dog remains from Northern Europe. J Archaeol Sci. 1987;14:31–49. https://doi.org/10.1016/S0305-4403(87)80004-3.

17. Nobis G. Der älteste Haushund lebte vor 14.000 Jahren. Umsch Wiss Tech. 1979;79:610.

18. Pang J-F, Kluetsch C, Zou X-J, Zhang A -b., Luo L-Y, Angleby H, et al. mtDNA data indicate a single origin for dogs south of Yangtze River, less than 16,300 years ago, from numerous wolves. Mol Biol Evol. 2009;26:2849–64. https://doi.org/10.1093/molbev/msp195.

19. Savolainen P, Zhang Y, Luo J, Lundeberg J, Leitner T. Genetic evidence for an East Asian origin of domestic dogs. Science. 2002;298:1610–3. https://doi.org/10.1126/science.1073906.

20. Wang G-D, Zhai W, Yang H-C, Wang L, Zhong L, Liu Y-H, et al. Out of southern East Asia: the natural history of domestic dogs across the world. Cell Res. 2016;26:21–33. https://doi.org/10.1038/cr.2015.147.

21. vonHoldt BM, Pollinger JP, Lohmueller KE, Han E, Parker HG, Quignon P, et al. Genome-wide SNP and haplotype analyses reveal a rich history underlying dog domestication. Nature. 2010;464:898–902. https://doi.org/10.1038/nature08837.

22. Bergström A, Stanton DWG, Taron UH, Frantz L, Sinding M-HS, Ersmark E, et al. Grey wolf genomic history reveals a dual ancestry of dogs. Nature. 2022;607:313–20. https://doi.org/10.1038/s41586-022-04824-9.

23. Fan Z, Silva P, Gronau I, Wang S, Armero AS, Schweizer RM, et al. Worldwide patterns of genomic variation and admixture in gray wolves. Genome Res. 2016;26:163–73. https://doi.org/10.1101/gr.197517.115.

24. Bannasch D, Famula T, Donner J, Anderson H, Honkanen L, Batcher K, et al. The effect of inbreeding, body size and morphology on health in dog breeds. Canine Med Genet. 2021;8:12. https://doi.org/10.1186/s40575-021-00111-4.

25. Cecchi F, Paci G, Spaterna A, Ciampolini R. Genetic variability in *Bracco Italiano* dog breed assessed by pedigree data. Ital J Anim Sci. 2013;12:e54. https://doi.org/10.4081/ijas.2013.e54.

26. Mastrangelo S, Biscarini F, Auzino B, Ragatzu M, Spaterna A, Ciampolini R. Genome-wide diversity and runs of homozygosity in the “Braque Français, type Pyrénées” dog breed. BMC Res Notes. 2018;11:13. https://doi.org/10.1186/s13104-017-3112-9.

27. Vaysse A, Ratnakumar A, Derrien T, Axelsson E, Rosengren Pielberg G, Sigurdsson S, et al. Identification of genomic regions associated with phenotypic variation between dog breeds using selection mapping. PLoS Genet. 2011;7:e1002316.

bioRxiv preprint doi: https://doi.org/10.1101/2024.05.01.592072; this version posted September 15, 2025. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

https://doi.org/10.1371/journal.pgen.1002316.

28. Yordy J, Kraus C, Hayward JJ, White ME, Shannon LM, Creevy KE, et al. Body size, inbreeding, and lifespan in domestic dogs. Conserv Genet. 2020;21:137–48. https://doi.org/10.1007/s10592-019-01240-x.

29. Plassais J, Kim J, Davis BW, Karyadi DM, Hogan AN, Harris AC, et al. Whole genome sequencing of canids reveals genomic regions under selection and variants influencing morphology. Nat Commun. 2019;10:1489. https://doi.org/10.1038/s41467-019-09373-w.

30. Sams AJ, Boyko AR. Fine-scale resolution of runs of homozygosity reveal patterns of inbreeding and substantial overlap with recessive disease genotypes in domestic dogs. G3 GenesGenomesGenetics. 2019;9:117–23. https://doi.org/10.1534/g3.118.200836.

31. Bacolod MD, Schemmann GS, Wang S, Shattock R, Giardina SF, Zeng Z, et al. The signatures of autozygosity among patients with colorectal cancer. Cancer Res. 2008;68:2610–21. https://doi.org/10.1158/0008-5472.CAN-07-5250.

32. Clark DW, Okada Y, Moore KHS, Mason D, Pirastu N, Gandin I, et al. Associations of autozygosity with a broad range of human phenotypes. Nat Commun. 2019;10:4957. https://doi.org/10.1038/s41467-019-12283-6.

33. Johnson EC, Evans LM, Keller MC. Relationships between estimated autozygosity and complex traits in the UK Biobank. PLOS Genet. 2018;14:e1007556. https://doi.org/10.1371/journal.pgen.1007556.

34. Keller MC, Simonson MA, Ripke S, Neale BM, Gejman PV, Howrigan DP, et al. Runs of homozygosity implicate autozygosity as a schizophrenia risk factor. PLoS Genet. 2012;8:e1002656. https://doi.org/10.1371/journal.pgen.1002656.

35. Moreno-Grau S, Fernández MV, De Rojas I, Garcia-González P, Hernández I, Farias F, et al. Long runs of homozygosity are associated with Alzheimer's disease. Transl Psychiatry. 2021;11:142. https://doi.org/10.1038/s41398-020-01145-1.

36. Pemberton TJ, Szpiech ZA. Relationship between deleterious variation, genomic autozygosity, and disease risk: insights from the 1000 genomes project. Am J Hum Genet. 2018;102:658–75. https://doi.org/10.1016/j.ajhg.2018.02.013.

37. Mooney JA, Yohannes A, Lohmueller KE. The impact of identity by descent on fitness and disease in dogs. Proc Natl Acad Sci. 2021;118:e2019116118. https://doi.org/10.1073/pnas.2019116118.

38. Boyko AR, Quignon P, Li L, Schoenebeck JJ, Degenhardt JD, Lohmueller KE, et al. A simple genetic architecture underlies morphological variation in dogs. PLOS Biol. 2010;8:e1000451. https://doi.org/10.1371/journal.pbio.1000451.

39. Cadieu E, Neff MW, Quignon P, Walsh K, Chase K, Parker HG, et al. Coat variation in the domestic dog is governed by variants in three genes. Science. 2009;326:150–3.

bioRxiv preprint doi: https://doi.org/10.1101/2024.05.01.592072; this version posted September 15, 2025. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

https://doi.org/10.1126/science.1177808.

40. Candille SI, Kaelin CB, Cattanach BM, Yu B, Thompson DA, Nix MA, et al. A β-Defensin mutation causes black coat color in domestic dogs. Science. 2007;318:1418–23. https://doi.org/10.1126/science.1147880.

41. Parker HG, VonHoldt BM, Quignon P, Margulies EH, Shao S, Mosher DS, et al. An Expressed *Fgf4* retrogene is associated with breed-defining chondrodysplasia in domestic dogs. Science. 2009;325:995–8. https://doi.org/10.1126/science.1173275.

42. Sutter NB, Bustamante CD, Chase K, Gray MM, Zhao K, Zhu L, et al. A single *IGF1* allele is a major determinant of small size in dogs. Science. 2007;316:112–5. https://doi.org/10.1126/science.1137045.

43. Hayward JJ, Castelhano MG, Oliveira KC, Corey E, Balkman C, Baxter TL, et al. Complex disease and phenotype mapping in the domestic dog. Nat Commun. 2016;7:10460. https://doi.org/10.1038/ncomms10460.

44. Barton AR, Hujoel MLA, Mukamel RE, Sherman MA, Loh P-R. A spectrum of recessiveness among Mendelian disease variants in UK Biobank. Am J Hum Genet. 2022;109:1298–307. https://doi.org/10.1016/j.ajhg.2022.05.008.

45. Chundru VK, Zhang Z, Walter K, Lindsay SJ, Danecek P, Eberhardt RY, et al. Federated analysis of autosomal recessive coding variants in 29,745 developmental disorder patients from diverse populations. Nat Genet. 2024;56:2046–53. https://doi.org/10.1038/s41588-024-01910-8.

46. Fridman H, Khazeeva G, Levy-Lahad E, Gilissen C, Brunner HG. Reproductive and cognitive effects in carriers of recessive pathogenic variants. 2024. https://doi.org/10.1101/2024.09.30.615774.

47. Lynch MT, Maloney KA, Xu H, Perry JA, Center RG, Shuldiner AR, et al. Associations of genome-wide and regional autozygosity with 96 complex traits in old order Amish. BMC Genomics. 2023;24:134. https://doi.org/10.1186/s12864-023-09208-5.

48. Malawsky DS, Van Walree E, Jacobs BM, Heng TH, Huang QQ, Sabir AH, et al. Influence of autozygosity on common disease risk across the phenotypic spectrum. Cell. 2023;186:4514-4527.e14. https://doi.org/10.1016/j.cell.2023.08.028.

49. Purfield DC, McParland S, Wall E, Berry DP. The distribution of runs of homozygosity and selection signatures in six commercial meat sheep breeds. PLOS ONE. 2017;12:e0176780. https://doi.org/10.1371/journal.pone.0176780.

50. Thompson EA. Identity by Descent: Variation in Meiosis, Across Genomes, and in Populations. Genetics. 2013;194:301–26. https://doi.org/10.1534/genetics.112.148825.

51. Freedman AH, Wayne RK. Deciphering the origin of dogs: From fossils to genomes. Annu Rev Anim Biosci. 2017;5:281–307. https://doi.org/10.1146/annurev-animal-022114-110937.

bioRxiv preprint doi: https://doi.org/10.1101/2024.05.01.592072; this version posted September 15, 2025. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

52. Gregory TR. Artificial selection and domestication: Modern lessons from Darwin's enduring analogy. Evol Educ Outreach. 2009;2:5–27. https://doi.org/10.1007/s12052-008-0114-z.

53. Parker HG, Dreger DL, Rimbault M, Davis BW, Mullen AB, Carpintero-Ramirez G, et al. Genomic analyses reveal the influence of geographic origin, migration, and hybridization on modern dog breed development. Cell Rep. 2017;19:697–708. https://doi.org/10.1016/j.celrep.2017.03.079.

54. Yang Q, Chen H, Ye J, Liu C, Wei R, Chen C, et al. Genetic diversity and signatures of selection in 15 Chinese indigenous dog breeds revealed by genome-wide SNPs. Front Genet. 2019;10:1174. https://doi.org/10.3389/fgene.2019.01174.

55. Joshi PK, Esko T, Mattsson H, Eklund N, Gandin I, Nutile T, et al. Directional dominance on stature and cognition in diverse human populations. Nature. 2015;523:459–62. https://doi.org/10.1038/nature14618.

56. McQuillan R, Eklund N, Pirastu N, Kuningas M, McEvoy BP, Esko T, et al. Evidence of inbreeding depression on human height. PLoS Genet. 2012;8:e1002655. https://doi.org/10.1371/journal.pgen.1002655.

57. Swinford N, Gallagher B, Sheehama J, Lin M, Prall S, Scelza B, et al. Examination of runs of homozygosity in relation to height in an endogamous Namibian population. Am J Biol Anthropol. 2023;180:207–15. https://doi.org/10.1002/ajpa.24660.

58. Hoopes BC, Rimbault M, Liebers D, Ostrander EA, Sutter NB. The insulin-like growth factor 1 receptor (IGF1R) contributes to reduced size in dogs. Mamm Genome. 2012;23:780–90. https://doi.org/10.1007/s00335-012-9417-z.

59. Urfer SR, Wang M, Yang M, Lund EM, Lefebvre SL. Risk factors associated with lifespan in pet dogs evaluated in primary care veterinary hospitals. J Am Anim Hosp Assoc. 2019;55:130–7. https://doi.org/10.5326/JAAHA-MS-6763.

60. Galis F, Van Der Sluijs I, Van Dooren TJM, Metz JAJ, Nussbaumer M. Do large dogs die young? J Exp Zoolog B Mol Dev Evol. 2007;308B:119–26. https://doi.org/10.1002/jez.b.21116.

61. Kichaev G, Bhatia G, Loh P-R, Gazal S, Burch K, Freund MK, et al. Leveraging polygenic functional enrichment to improve GWAS power. Am J Hum Genet. 2019;104:65–75. https://doi.org/10.1016/j.ajhg.2018.11.008.

62. Yengo L, Vedantam S, Marouli E, Sidorenko J, Bartell E, Sakaue S, et al. A saturated map of common genetic variants associated with human height. Nature. 2022;610:704–12. https://doi.org/10.1038/s41586-022-05275-y.

63. Costa-Urrutia P, Colistro V, Jiménez-Osorio AS, Cárdenas-Hernández H, Solares-Tlapechco J, Ramirez-Alcántara M, et al. Genome-wide association study of body mass index and body fat in Mexican-Mestizo children. Genes. 2019;10:945. https://doi.org/10.3390/genes10110945.

64. Guo MH, Hirschhorn JN, Dauber A. Insights and implications of genome-wide association

bioRxiv preprint doi: https://doi.org/10.1101/2024.05.01.592072; this version posted September 15, 2025. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

studies of height. J Clin Endocrinol Metab. 2018;103:3155–68. https://doi.org/10.1210/jc.2018-01126.

65. RPH3AL gene information - The Human Protein Atlas. https://www.proteinatlas.org/ENSG00000181031-RPH3AL/summary/gene. Accessed 30 Oct 2024.

66. COMMD6 gene information - The Human Protein Atlas. https://www.proteinatlas.org/ENSG00000188243-COMMD6/summary/gene. Accessed 30 Oct 2024.

67. Trujillo□Viera J, El□Merahbi R, Schmidt V, Karwen T, Loza□Valdes A, Strohmeyer A, et al. Protein Kinase D2 drives chylomicron□mediated lipid transport in the intestine and promotes obesity. EMBO Mol Med. 2021;13:e13548. https://doi.org/10.15252/emmm.202013548.

68. Selionova M, Aibazov M, Mamontova T, Malorodov V, Sermyagin A, Zinovyeva N, et al. Genome-wide association study of live body weight and body conformation traits in young Karachai goats. Small Rumin Res. 2022;216:106836. https://doi.org/10.1016/j.smallrumres.2022.106836.

69. Ali AT, Hochfeld WE, Myburgh R, Pepper MS. Adipocyte and adipogenesis. Eur J Cell Biol. 2013;92:229–36. https://doi.org/10.1016/j.ejcb.2013.06.001.

70. Luo J, Dou L, Yang Z, Zhou Z, Huang H. CBFA2T2 promotes adipogenic differentiation of mesenchymal stem cells by regulating CEBPA. Biochem Biophys Res Commun. 2020;529:133–9. https://doi.org/10.1016/j.bbrc.2020.05.120.

71. Matulewicz N, Stefanowicz M, Nikołajuk A, Karczewska-Kupczewska M. Markers of adipogenesis, but not inflammation, in adipose tissue are independently related to insulin sensitivity. J Clin Endocrinol Metab. 2017;102:3040–9. https://doi.org/10.1210/jc.2017-00597.

72. Kang S, Bennett CN, Gerin I, Rapp LA, Hankenson KD, MacDougald OA. Wnt signaling stimulates osteoblastogenesis of mesenchymal precursors by suppressing CCAAT/Enhancer-binding protein α and peroxisome proliferator-activated receptor γ. J Biol Chem. 2007;282:14515–24. https://doi.org/10.1074/jbc.M700030200.

73. Kipp AP, Müller MF, Göken EM, Deubel S, Brigelius-Flohé R. The selenoproteins GPx2, TrxR2 and TrxR3 are regulated by Wnt signalling in the intestinal epithelium. Biochim Biophys Acta BBA - Gen Subj. 2012;1820:1588–96. https://doi.org/10.1016/j.bbagen.2012.05.016.

74. Martínez-Montes AM, Muiños-Bühl A, Fernández A, Folch JM, Ibáñez-Escriche N, Fernández AI. Deciphering the regulation of porcine genes influencing growth, fatness and yield-related traits through genetical genomics. Mamm Genome. 2017;28:130–42. https://doi.org/10.1007/s00335-016-9674-3.

75. Blake JA, Baldarelli R, Kadin JA, Richardson JE, Smith CL, Bult CJ, et al. Mouse Genome Database (MGD): Knowledgebase for mouse–human comparative biology. Nucleic Acids Res. 2021;49:D981–7. https://doi.org/10.1093/nar/gkaa1083.

bioRxiv preprint doi: https://doi.org/10.1101/2024.05.01.592072; this version posted September 15, 2025. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

76. Shapses SA, Sukumar D. Bone metabolism in obesity and weight loss. Annu Rev Nutr. 2012;32:287–309. https://doi.org/10.1146/annurev.nutr.012809.104655.

77. Tachmazidou I, Süveges D, Min JL, Ritchie GRS, Steinberg J, Walter K, et al. Whole-genome sequencing coupled to imputation discovers genetic signals for anthropometric traits. Am J Hum Genet. 2017;100:865–84. https://doi.org/10.1016/j.ajhg.2017.04.014.

78. CHMP4B gene information - The Human Protein Atlas. https://www.proteinatlas.org/ENSG00000101421-CHMP4B/summary/gene. Accessed 30 Oct 2024.

79. Warrington NM, Beaumont RN, Horikoshi M, Day FR, Helgeland Ø, Laurin C, et al. Maternal and fetal genetic effects on birth weight and their relevance to cardio-metabolic risk factors. Nat Genet. 2019;51:804–14. https://doi.org/10.1038/s41588-019-0403-1.

80. Alb JG, Cortese JD, Phillips SE, Albin RL, Nagy TR, Hamilton BA, et al. Mice lacking phosphatidylinositol transfer protein-α exhibit spinocerebellar degeneration, intestinal and hepatic steatosis, and hypoglycemia. J Biol Chem. 2003;278:33501–18. https://doi.org/10.1074/jbc.M303591200.

81. Guetg A, Mariotta L, Bock L, Herzog B, Fingerhut R, Camargo SMR, et al. Essential amino acid transporter Lat4 (*Slc43a2*) is required for mouse development. J Physiol. 2015;593:1273–89. https://doi.org/10.1113/jphysiol.2014.283960.

82. Wu W, Li D, Feng X, Zhao F, Li C, Zheng S, et al. A pan-cancer study of selenoprotein genes as promising targets for cancer therapy. BMC Med Genomics. 2021;14:78. https://doi.org/10.1186/s12920-021-00930-1.

83. Hjorthøj C, Stürup AE, McGrath JJ, Nordentoft M. Years of potential life lost and life expectancy in schizophrenia: a systematic review and meta-analysis. Lancet Psychiatry. 2017;4:295–301. https://doi.org/10.1016/S2215-0366(17)30078-0.

84. Kang Y, Zhang Y, Huang K, Wang Z. The genetic influence of the DRD3 rs6280 polymorphism (Ser9Gly) on functional connectivity and gray matter volume of the hippocampus in patients with first-episode, drug-naïve schizophrenia. Behav Brain Res. 2023;437:114124. https://doi.org/10.1016/j.bbr.2022.114124.

85. Chauvin J-M, Pagliano O, Fourcade J, Sun Z, Wang H, Sander C, et al. TIGIT and PD-1 impair tumor antigen–specific CD8+ T cells in melanoma patients. J Clin Invest. 2015;125:2046–58. https://doi.org/10.1172/JCI80445.

86. Fu L, Wang X, Yang Y, Chen M, Kuerban A, Liu H, et al. Septin11 promotes hepatocellular carcinoma cell motility by activating RhoA to regulate cytoskeleton and cell adhesion. Cell Death Dis. 2023;14:280. https://doi.org/10.1038/s41419-023-05726-y.

87. Johnston RJ, Comps-Agrar L, Hackney J, Yu X, Huseni M, Yang Y, et al. The immunoreceptor TIGIT regulates antitumor and antiviral CD8 + T cell effector function. Cancer Cell. 2014;26:923–37. https://doi.org/10.1016/j.ccell.2014.10.018.

bioRxiv preprint doi: https://doi.org/10.1101/2024.05.01.592072; this version posted September 15, 2025. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

88. Mao Y, Jiang X, Guo P, Ouyang Y, Chen X, Xia M, et al. ZXDC enhances cervical cancer metastasis through IGF2BP3-mediated activation of RhoA/ROCK signaling. iScience. 2023;26:107447. https://doi.org/10.1016/j.isci.2023.107447.

89. Zhao W, Yang H, Chai J, Xing L. RUNX2 as a promising therapeutic target for malignant tumors. Cancer Manag Res. 2021;Volume 13:2539–48. https://doi.org/10.2147/CMAR.S302173.

90. Chang Q, Xu Y, Wang J, Jing H, Rao L, Tang W, et al. SLC41A3 exhibits as a carcinoma biomarker and promoter in liver hepatocellular carcinoma. Comput Math Methods Med. 2021;2021:1–9. https://doi.org/10.1155/2021/8556888.

91. Sollis E, Mosaku A, Abid A, Buniello A, Cerezo M, Gil L, et al. The NHGRI-EBI GWAS Catalog: knowledgebase and deposition resource. Nucleic Acids Res. 2023;51:D977–85. https://doi.org/10.1093/nar/gkac1010.

92. Freedman AH, Lohmueller KE, Wayne RK. Evolutionary history, selective sweeps, and deleterious variation in the dog. Annu Rev Ecol Evol Syst. 2016;47:73–96. https://doi.org/10.1146/annurev-ecolsys-121415-032155.

93. Jones P, Chase K, Martin A, Davern P, Ostrander EA, Lark KG. Single-nucleotide-polymorphism-based association mapping of dog stereotypes. Genetics. 2008;179:1033–44. https://doi.org/10.1534/genetics.108.087866.

94. Plassais J, Rimbault M, Williams FJ, Davis BW, Schoenebeck JJ, Ostrander EA. 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. 2017;13:e1006661. https://doi.org/10.1371/journal.pgen.1006661.

95. Rimbault M, Beale HC, Schoenebeck JJ, Hoopes BC, Allen JJ, Kilroy-Glynn P, et al. Derived variants at six genes explain nearly half of size reduction in dog breeds. Genome Res. 2013;23:1985–95. https://doi.org/10.1101/gr.157339.113.

96. Wang C, Wallerman O, Arendt M-L, Sundström E, Karlsson Å, Nordin J, et al. A novel canine reference genome resolves genomic architecture and uncovers transcript complexity. Commun Biol. 2021;4:185. https://doi.org/10.1038/s42003-021-01698-x.

97. Danecek P, Bonfield JK, Liddle J, Marshall J, Ohan V, Pollard MO, et al. Twelve years of SAMtools and BCFtools. GigaScience. 2021;10:giab008. https://doi.org/10.1093/gigascience/giab008.

98. Dog Breeds - Types Of Dogs. American Kennel Club. https://www.akc.org/dog-breeds/. Accessed 14 Oct 2024.

99. Szpiech ZA, Blant A, Pemberton TJ. *GARLIC*: Genomic autozygosity regions likelihood-based inference and classification. Bioinformatics. 2017;33:2059–62. https://doi.org/10.1093/bioinformatics/btx102.

100. Pemberton TJ, Absher D, Feldman MW, Myers RM, Rosenberg NA, Li JZ. Genomic

bioRxiv preprint doi: https://doi.org/10.1101/2024.05.01.592072; this version posted September 15, 2025. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

patterns of homozygosity in worldwide human populations. Am J Hum Genet. 2012;91:275–92. https://doi.org/10.1016/j.ajhg.2012.06.014.

101. Chang CC, Chow CC, Tellier LC, Vattikuti S, Purcell SM, Lee JJ. Second-generation PLINK: rising to the challenge of larger and richer datasets. GigaScience. 2015;4:7. https://doi.org/10.1186/s13742-015-0047-8.

102. McQuillan R. Homozygosity, inbreeding and health in European populations. 2009.

103. Quinlan AR, Hall IM. BEDTools: a flexible suite of utilities for comparing genomic features. Bioinformatics. 2010;26:841–2. https://doi.org/10.1093/bioinformatics/btq033.

104. Chen H, Wang C, Conomos MP, Stilp AM, Li Z, Sofer T, et al. Control for population structure and relatedness for binary traits in genetic association studies via logistic mixed models. Am J Hum Genet. 2016;98:653–66. https://doi.org/10.1016/j.ajhg.2016.02.012.

105. Fareed M, Afzal M. Evidence of inbreeding depression on height, weight, and body mass index: A population□based child cohort study. Am J Hum Biol. 2014;26:784–95. https://doi.org/10.1002/ajhb.22599.

106. Kim K, Jung J, Caetano-Anollés K, Sung S, Yoo D, Choi B-H, et al. Artificial selection increased body weight but induced increase of runs of homozygosity in Hanwoo cattle. PLOS ONE. 2018;13:e0193701. https://doi.org/10.1371/journal.pone.0193701.

107. Plummer M, Best N, Cowles K, Vines K. CODA: convergence diagnosis and output analysis for MCMC. R News. 2006;6:7–11.

108. Shriner D, Adeyemo A, Rotimi CN. Joint Ancestry and Association Testing in Admixed Individuals. PLoS Comput Biol. 2011;7:e1002325. https://doi.org/10.1371/journal.pcbi.1002325.

109. Li M-X, Yeung JMY, Cherny SS, Sham PC. Evaluating the effective numbers of independent tests and significant p-value thresholds in commercial genotyping arrays and public imputation reference datasets. Hum Genet. 2012;131:747–56. https://doi.org/10.1007/s00439-011-1118-2.

110. Yang W, Guo Z, Huang C, Duan L, Chen G, Jiang N, et al. Combining high-throughput phenotyping and genome-wide association studies to reveal natural genetic variation in rice. Nat Commun. 2014;5:5087. https://doi.org/10.1038/ncomms6087.

111. Yuan J, Wang X, Zhao Y, Khan NU, Zhao Z, Zhang Y, et al. Genetic basis and identification of candidate genes for salt tolerance in rice by GWAS. Sci Rep. 2020;10:9958. https://doi.org/10.1038/s41598-020-66604-7.

112. Sofer T, Zheng X, Laurie CA, Gogarten SM, Brody JA, Conomos MP, et al. Variant-specific inflation factors for assessing population stratification at the phenotypic variance level. Nat Commun. 2021;12:3506. https://doi.org/10.1038/s41467-021-23655-2.

113. Van Den Berg S, Vandenplas J, Van Eeuwijk FA, Lopes MS, Veerkamp RF. Significance

bioRxiv preprint doi: https://doi.org/10.1101/2024.05.01.592072; this version posted September 15, 2025. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

testing and genomic inflation factor using high□density genotypes or whole□genome sequence data. J Anim Breed Genet. 2019;136:418–29. https://doi.org/10.1111/jbg.12419.

114. Turner SD. qqman: an R package for visualizing GWAS results using Q-Q and manhattan plots. 2014. https://doi.org/10.1101/005165.