

Scotland's Rural College

Phenome-wide study connects behavioral genetics of odor detection dogs with temperament traits

Eyre, Alexander W; Zapata, Isain; Fraire, Jasmin; Coty, Kevin D; Huson, Heather J; Evans, Katy M; Otto, Cynthia M; Serpell, James A; van Rooy, Diane; Wade, Claire M; Blott, Sarah C; Wiener, Pamela; Haskell, Marie J; Alvarez, Carlos E

Published in:
Scientific Reports

DOI:
[10.1038/s41598-026-42708-4](https://doi.org/10.1038/s41598-026-42708-4)

First published: 11/03/2026

Document Version
Publisher's PDF, also known as Version of record

[Link to publication](#)

Citation for pulished version (APA):

Eyre, A. W., Zapata, I., Fraire, J., Coty, K. D., Huson, H. J., Evans, K. M., Otto, C. M., Serpell, J. A., van Rooy, D., Wade, C. M., Blott, S. C., Wiener, P., Haskell, M. J., & Alvarez, C. E. (2026). Phenome-wide study connects behavioral genetics of odor detection dogs with temperament traits. *Scientific Reports*, 16, Article 13768. Advance online publication. <https://doi.org/10.1038/s41598-026-42708-4>

General rights

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal ?

Take down policy

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

Phenome-wide study connects behavioral genetics of odor detection dogs with temperament traits

Received: 17 September 2024

Accepted: 27 February 2026

Published online: 11 March 2026

Cite this article as: Eyre A.W., Zapata I., Fraire J. *et al.* Phenome-wide study connects behavioral genetics of odor detection dogs with temperament traits. *Sci Rep* (2026). <https://doi.org/10.1038/s41598-026-42708-4>

Alexander W. Eyre, Isain Zapata, Jasmin Fraire, Kevin D. Coty, Heather J. Huson, Katy M. Evans, Cynthia M. Otto, James A. Serpell, Diane Rooy, Claire M. Wade, Sarah C. Blott, Pamela Wiener, Marie J. Haskell & Carlos E. Alvarez

We are providing an unedited version of this manuscript to give early access to its findings. Before final publication, the manuscript will undergo further editing. Please note there may be errors present which affect the content, and all legal disclaimers apply.

If this paper is publishing under a Transparent Peer Review model then Peer Review reports will publish with the final article.

Phenome-Wide Study Connects Behavioral Genetics of Odor Detection Dogs with Temperament Traits

Alexander W. Eyre, Isain Zapata, Jasmin Fraire, Kevin D. Coty, Heather J. Huson, Katy M. Evans, Cynthia M. Otto, James A. Serpell, Diane van Rooy, Claire M. Wade, Sarah C. Blott, Pamela Wiener, Marie J. Haskell, Carlos E. Alvarez*

Alexander W. Eyre, Independent Researcher, USA, aweyre90@gmail.com

Isain Zapata, Department of Biomedical Sciences, Rocky Vista University
College of Osteopathic Medicine, Parker, CO, 80134, USA,
superisain@gmail.com

Jasmin Fraire, Independent Researcher, USA, jasminfraire97@gmail.com

Kevin D. Coty, Battelle, 505 King Avenue, Columbus, Ohio, 43201, USA,
cotyk@battelle.org

Heather J. Huson, Department of Animal Science, Cornell University, 201
Morrison Hall, 507 Tower Rd., Ithaca, NY 14853, USA, hjh3@cornell.edu

Katy M. Evans, The Seeing Eye, Inc., Morristown, NJ 07960,
USA, KEvans@seeingeye.org

Cynthia M. Otto, Penn Vet Working Dog Center, School of Veterinary
Medicine, University of Pennsylvania, Philadelphia, PA 19146 USA,
cmotto@vet.upenn.edu

James A. Serpell, Department of Clinical Sciences and Advanced Medicine,
School of Veterinary Medicine, University of Pennsylvania, Philadelphia,
PA 19104 USA, serpell@vet.upenn.edu

Diane van Rooy, School of Life and Environmental Sciences, University of
Sydney, RMC Gunn B19-301, University of Sydney NSW 2006, Australia,
currently Independent Researcher, drdianevanrooy@gmail.com

Claire M. Wade, School of Life and Environmental Sciences, University of
Sydney, RMC Gunn B19-301, University of Sydney NSW 2006, Australia,
claire.wade@sydney.edu.au

Sarah C. Blott, School of Veterinary Medicine and Science, University of Nottingham, Gateway Building, Sutton Bonington, Leicestershire, LE12 5RD, UK, sarah.blott@nottingham.ac.uk

Pamela Wiener, Roslin Institute, University of Edinburgh, Easter Bush, Midlothian, EH25 9RG, Scotland, UK, pam.wiener@roslin.ed.ac.uk

Marie J. Haskell, SRUC (Scotland's Rural College), UK, marie.haskell@sruc.ac.uk

Carlos E. Alvarez, Departments of Pediatrics and Veterinary Clinical Sciences, The Ohio State University Wexner Medical Center and The Ohio State University College of Veterinary Medicine, Columbus, OH 43210, USA, alvarez.73@osu.edu

*Corresponding author

Abstract

We conducted a Phenome-Wide Association Study (PheWAS) to investigate whether alleles previously shown to be associated with problem behaviors in Labrador Retrievers from the U.S. Transportation Security Administration's (TSA) odor detection program also show behavioral associations in other populations. The original TSA cohort (2002–2013) consisted of dogs from a former breeding program that drew from U.S. commercial sources and the Australian Customs Service. While those data included TSA testing results from the foster period, detailed behavior profiles and reasons for elimination from the program were not recorded. To extend and validate these genetic associations in populations with richer behavioral data, we analyzed three additional Labrador Retriever cohorts with both genotype and C-BARQ behavioral questionnaire data: (1) Australian pet dogs, (2) UK dogs from a mixed pet and working background (primarily gamebird retrieving), and (3) U.S. working guide dogs. This analysis identified a total of 15 associations between 12 behavioral traits and 8 markers at 6 genome loci. Notably, we found four types of aggression and one type of fear that are directed at familiar humans or dogs, but none directed at unfamiliar ones. Other problem traits identified include separation-related behaviors, excitability, and chasing small animals. Furthermore, we utilized whole genome sequencing to identify a functional candidate associated with "aggression when approached by a household dog at a favorite resting place". We propose this variant in an *ADAMTSL1* intron results in the loss of TCF7L1 protein binding, and we highlight the evolutionary history of that conserved

element, including the fixation of two mutations in the human lineage. Our PheWAS findings suggest relevance to working dog selection, breeding, and training, presenting opportunities to reduce costs while improving performance and resilience.

Keywords:

PheWAS, GWAS, canine, behavior, temperament, working dogs

Introduction

Detection dogs are widely recognized as the most effective tool for detecting and alerting to diverse and evolving odors critical to security and law enforcement [1]. The increasing use of detection dogs in the recent past has resulted in competition for the limited numbers of available dogs from a few breeds (mainly, Labrador Retriever, German Shorthaired Pointer, Belgian Malinois, and German Shepherd) [2]. There is also great interest in optimizing working dog selection, training, health, and career longevity [1]. A rapidly growing aspect of that is research of the behavioral, genetic, and environmental effects associated with detection dog success.

Beyond the interest in working dogs, the reduced genetic complexity of dog breeds [3] provides an exceptionally tractable model for understanding the genetic and environmental contributions to behavior. For instance, human complex traits have extremely high polygenicity and heterogeneity. Because many deleterious effects on behavior and cognition would have reduced evolutionary fitness, negative selection in human evolution has likely purged common variants with detrimental effects on these traits [4]. As a result, most behavioral and cognitive traits show few, if any, common genetic variants with sizable individual effects. Darwin noted that under domestication, artificial selection by humans dramatically relaxes the natural constraints on variation seen in the wild [5]. Domesticated species like dogs can accumulate countless deleterious genetic variants through several mechanisms: linkage to traits under artificial selection, insufficiently deleterious effects to be selected against in breeding contexts, or increases in frequency due to genetic drift or founder effects within subgroups of the population, such as geographic isolates or lineages from popular sires [3]. Thus, complex trait Genome Wide Association Studies (GWASs) in humans can yield from few to a great many variations that individually and collectively explain a small amount of the trait variance [4]. However, GWAS loci of the same traits in dogs can have moderate-to-large effects individually

and can collectively explain most of the variance [6]. As this GWAS effect applies to behavioral and cognitive traits [7-10], the hundreds of dog breeds offer unique and powerful approaches to dissecting brain development and function.

The field of canine behavioral genetics has relied heavily on behavioral questionnaires to measure phenotypes. The most popular questionnaire has been the C-BARQ, which was designed to capture all common problem behaviors as well as traits such as energy, excitability, and trainability [11]. There are several other questionnaires which capture similar domains [12]. Some investigators, however, have combined data from multiple instruments to assess a broader range of behaviors, including traits associated with specific breed groups, such as herding or pointing [13]. Genetic study designs have included interbreed mapping of breed averages of C-BARQ scores in unrelated cohorts of genotyped pedigree dogs [9, 10, 14] and single or multiple breed mapping in dogs with individual-level data for both C-BARQ and similar questionnaires and genome-wide genotypes [7, 15, 16]. A couple of genetic studies included both pedigree and mixed breed dogs with individual-level C-BARQ [17] or other questionnaire [13] data. The 14 C-BARQ factors have narrow sense heritabilities in the range of $h^2=0.27-0.77$ across breeds (mean 0.51 ± 0.12 s.d.) [14]. By contrast, a study of mostly pet Labrador Retrievers showed heritabilities of all 14 factors to be five times smaller than that on average [7]. While some of the highest heritabilities are shared in Labrador Retrievers and across breeds (e.g., stranger directed aggression and chasing), others are discrepant. For example, non-social fear ranks third for highest heritability in Labrador Retrievers but has the lowest heritability across breeds (ranked 14th). The heritability of trainability, a trait of great importance in working dogs, ranks ninth in Labrador Retrievers and first across breeds. The patterns of breed differences hint at selection under domestication, which is supported by $Q_{st} - F_{st}$ analysis showing all 14 C-BARQ factors are under strong positive selection [14].

The Labrador Retriever is the most popular registered dog breed in the world (Federation Cynologique Internationale). They are also one of the most popular breeds used as working dogs – most notably for odor detection, guiding, and assistance work [2]. They have been studied for C-BARQ behavioral traits [18] and heritability analysis and genetic mapping of the C-BARQ traits [7]. We recently performed a GWAS of elimination for behavioral reasons in Labrador Retrievers in the Transportation Safety Administration's (TSA) canine olfactory detection program [19]. That GWAS provided the SNPs used to conduct the present study. Briefly, we used historical data for a breeding population with founders from US commercial breeders of Labrador Retrievers selected for field work and the Australian Customs Service. This

program tested dogs four times over a one-year foster period that preceded selection for the training program or elimination for behavioral or medical reasons. The data included genome wide SNP genotypes and behavioral testing results, but not the basis for behavioral elimination. That work yielded six genome-wide significant and five suggestive loci. Haplotypes previously shown to be associated with problem behaviors across breeds and to commonly be present in pet Labrador Retrievers[9, 10] were absent in the population of our original GWAS [19]. This suggests that these behavior-associated variants may have been selectively bred against (i.e., individuals carrying them were less likely to be chosen as breeding stock) in both the U.S. hunting lines of the breed and in the Australian Customs Service, where dogs have been actively selected for odor detection work for at least 20 years [20]). For several loci that appeared to implicate a single gene, there was strong evidence of behavioral relevance for those genes in other species [19]. However, this cross-species relevance does not confirm the genetic associations identified in our study. In the second paragraph of the Discussion we note the challenges to replicating this study.

We performed separate studies of that same 2013 TSA cohort for associations of behavioral testing data and outcomes, specifically, success or elimination from the training program [21]. This showed that Machine Learning models could predict which dogs would succeed with reasonable accuracy but were less effective at identifying those likely to be eliminated. This aligns with other findings showing that training and performance data are more predictive of success, while temperament and personality assessments are better at identifying dogs likely to fail [21-23]. These findings highlight the need for working dog programs to systematically collect and integrate both types of data (work-related testing and temperament evaluations) to best understand and improve placement outcomes. For example, The Seeing Eye, an organization which breeds and trains guide dogs for visually impaired people across North America, successfully integrates both types of data – training records, C-BARQ and other behavioral assessments, and environmental factors – into their program.

The success of GWAS in identifying genetic variation associated with diverse traits presents opportunities to dissect molecular mechanisms [4]. It also allows a type of backwards analysis to identify additional phenotypes associated with GWAS haplotypes: Phenome-Wide Association Studies or PheWAS [24]. The approach starts with one or more genetic variants – often initially identified through a GWAS for a single trait – and tests their associations across a wide range of traits to better understand their broader physiological effects. These traits may be closely related to the context of

the original discovery or span a broad set selected by any criteria, provided appropriate statistical correction for multiple testing is applied. The strength of observed associations can reflect how directly a variant influences core biological processes, with stronger effects typically seen for traits tied to primary biochemical pathways and weaker effects for more indirect influences. Thus, PheWAS not only identifies pleiotropic effects of GWAS variants but also helps to stratify those variants based on the diversity of associated phenotypes and underlying biochemical pathways. Here, we use PheWAS to bridge a Labrador Retriever phenotype from the original TSA odor detection cohort (i.e., used in the initial GWAS) to three additional cohorts: Australian pet dogs, UK dogs with mixed pet and gamebird-retrieving (working) roles, and US guide (working) dogs. Specifically, we use temperament and personality traits captured by C-BARQ in the three PheWAS cohorts to find associations with loci mapped for elimination for behavioral reasons in odor detection dogs (for which there is no C-BARQ data [19]).

Results

Description of GWAS SNPs & Labrador Retriever Cohorts Used for PheWAS

The SNPs used in this study were identified in a prior GWAS on a 2002-2013 TSA breeding and pretraining program using elimination for behavioral reasons as the phenotype [19]. This cohort included 296 genotyped Labrador Retrievers with 267 successful and 29 eliminated dogs. Because a complete pedigree for the cohort was available, we performed a family-based GWAS using the *ROADTRIPS2* (version 2.0, https://faculty.washington.edu/tathornt/software/ROADTRIPS2/ROADTRIPS2_Documentation.pdf) software [25]. It reduces type I error due to relatedness by including the pedigree as a covariate in the association analysis and it increases power by leveraging familial relationships to impute genotypes for individuals with phenotype but no genotype data. That GWAS identified 21 SNPs. One additional SNP included here is from a GWAS of the same cohort but using GEMMA, which is part of separate ongoing work. GEMMA software generates and applies a relatedness matrix to its linear mixed model association analysis [26]. It was selected for inclusion in this PheWAS because it was the only finding for which both alleles were segregating in the two largest PheWAS cohorts (one of those cohorts was subsequently excluded after filtering for the minimum minor allele frequency). Both GWASs showed appropriate control of genomic inflation according to the calculated lambda values. In total, we used 22 SNPs in 11 loci here: nine

genome-wide significant after Bonferroni multiple testing correction, and two suggestive (chr3 and chr4; **Supplemental Table 1**).

Three Labrador Retriever cohorts with both genotype data and completed C-BARQ surveys were used for this PheWAS study. The first two were pet and pet/working (gamebird retrievers housed in kennels; see Methods), respectively, populations predominately comprised of privately owned dogs from Australia (AUS) [27] and the United Kingdom (UK) [7]. The third population was from The Seeing Eye, Inc. (TSE), which has a breeding and training program for guide dogs for blind people. Following quality control, a substantial number of SNPs within each cohort were removed due to fixed allele frequency (**Table 1**). Of the 22 SNPs previously identified in the 2002-2013 TSA GWAS, 13, 9, and 8 had both alleles and passed the minimum minor allele frequency of 2.5% for inclusion in the PheWAS for the AUS, UK, and TSE cohorts, respectively (**Supplemental Table 1**).

PheWAS and Intersection Union Test Meta-analysis

A PheWAS was performed to identify which dog behaviors were correlated with each previously identified SNP. These analyses involved testing each of the 22 SNPs for correlation with each C-BARQ question. Hits were considered significant if they passed a $p < 0.05$ Bonferroni adjusted threshold calculated for each cohort. **Figure 1** shows histograms of individual SNPs with C-BARQ scores coded by genotype. These tests resulted in a total of 15 hits spanning 8 of the 22 tested SNPs and 6 of the 10 defined SNP regions (**Table 2**). Seven of these hits were in the AUS cohort, five in the UK cohort, and one in TSE cohort. Additionally, a meta-analysis was conducted using the Intersection Union Test (IUT) to attempt to improve power by combining cohorts. This involved the calculation of the Fisher's combined p-value, and findings were only considered significant if C-BARQ questions both passed $p < 0.05$ for each individual cohort and their combined p-value passed a Bonferroni adjusted threshold $p < 0.05$. In all, the meta-analyses yielded a total of two hits for chr19.21,040,815 in the combined UK and TSE cohorts (**Table 2**).

The C-BARQ factor that contained the largest number of significant hits was aggression, which also accounted for the individual question with the strongest p-value (chr1.24,927,539 and agg_25 in the AUS cohort). The significant questions involving social behavior are predominantly ones directed at familiar humans and dogs, except for excitability when visitors arrive at the home. However, that latter chr19 locus has several other associations - the most of any SNP - and the p-value for aggression directed at household members for taking an object away was over three orders of

magnitude stronger. Chr6.76,632,282 showed associations with two questions in different factors but both questions had to do with anxiety-related behaviors, either when left alone or approached. Both SNPs in the chr36 locus had significant hits related to chasing small animals. The other two-SNP presumptive haplotype, on chr1, was also supported by both being associated with the same question.

Sequence analysis of the mapped haplotype at the ADAMTSL1 locus

As a follow-up to the original TSA GWAS and the loci supported by the present PheWAS – both conducted in Labrador Retrievers – we performed whole genome sequencing to identify candidate functional variants (Methods). This analysis was based on a random sample of seven Labrador Retrievers and seven German Shorthaired Pointers from the TSA odor detection program at the time of the study. These two breeds are currently the primary breeds used in the TSA detection program. Both were originally developed for hunting and are classified as Sporting Group dogs by the American Kennel Club. Among non-retriever breeds, German Shorthaired Pointers showed the third highest level of haplotype sharing with Labrador Retrievers out of 161 breeds, comparable to the second-ranking breed and to the Irish Water Spaniel, which is genetically classified in the Retriever Group [28]. While most mapped haplotypes are rare and absent in our small sample of randomly chosen dogs, we found evidence of the haplotype mapped on chr11. In this work, we showed it is associated with agg_33, aggression displayed when approached at a favorite resting place by a familiar dog. The risk allele of the peak SNP was present in seven carriers (four Labs and three German pointers). The two flanking SNPs in the mapping, with weaker p-values, were in imperfect LD as expected. We determined the haplotype interval shared by carriers to be chr11:38,564,807-38,745,391 on CanFam3.1. Most of it spans the gene *ADAMTSL1*, which encodes an extracellular matrix related protein.

We next conducted mining of genetic and genomic data to identify *ADAMTSL1* biology that may be pertinent to our study. High potential relevance to the mapped trait of elimination for behavior is implicated by human GWAS signals implicating *ADAMTSL1* for brain structure, cognitive traits, and drug use (GWAS Catalog [29]). The strongest p-values in human GWASs are bone density, height, and myopia (p-values of 1E-300, 6E-63, 4E-51, respectively). That is consistent with strong mRNA enrichment in adult mouse osteoblasts and eye (eyecup, ciliary body, iris, and retina) (BioGPS, tissue mRNA expression [30]). We also found enriched mRNA expression of *Adamts1* in the glomerular layer of the main olfactory bulb in adult mice

(Allen Brain Atlas, *in situ* mRNA hybridization [31]), which could be relevant to the associated social trait, agg_33. A human normalized atlas of single cell types across the body shows predominant mRNA expression in excitatory neurons, followed by 6.3-fold lower levels in inhibitory neurons (Human Cell Atlas, single cell RNA sequencing [32]). That is consistent with that resource's single cell atlas of the brain. Curiously, of the 26 excitatory neuron clusters, those with intermediate *ADAMTSL1* expression were *CUX2*-positive [33] but the lowest and by far the highest expressors were negative.

We compared the dog haplotype sequence to the matching sequence in humans, which is annotated for various evolutionary traits and deleteriousness prediction from diverse types of evidence, including human genetics (UCSC Genome Browser, hg38; see Methods). The strongest evidence of conservation at any variant sequence is in an *ADAMTSL1* intron, in the middle of a 17-base pair element (PhastCons lod=38; UCSC Genome Server, hg38). This variant position at canFam3.1 chr11:38,725,982 (C/T) has a CADD score for deleteriousness of 20.1, placing it in the top 1% of bases in the human genome (Combined Annotation Dependent Depletion [34]). In **Table 4**, we show detection of the C/T alleles in canids and their frequencies if those were available. Based on these numbers, the alternative dog allele T has a frequency of ~7% in wolves and is much higher frequency in ancient, village, indigenous, and modern pedigree dogs, ranging 22-35%. Follow-on studies are necessary to determine the origins and history of the canid T allele.

We analyzed the 17bp PhastCons element, and the dog variant within it, using the genome alignments '100 vertebrates', and 'Mutiz 470' and 'Cactus 447' mammals [35, 36] on the UCSC Genome Browser. This showed the conserved element and flanking sequence are specific to mammals, in the pattern presented in **Figure 2**. Most relevant to this work, there is segregation of the two alleles in dogs and wolves. We also observed fixation of single alternative variant alleles at the base of Anthropoidea and of Artiodactyla. Lastly, there was complete loss of a larger intronic sequence spanning the conserved element in all rodents (but not rabbits) and this frequently occurred in multiple marsupial species (opossums, possums, and wallabies each have species with and without the element and flanking sequence).

We next analyzed the conserved element for transcription factor binding predictions. We found a TCF7L1 binding site exclusive to the consensus mammalian sequence but not the Anthropoidea or Artiodactyla sequences, or the dog T allele (**Fig. 3A/C**). TCF7L1 and 2 binding site motifs are highly similar and both could be relevant [37]. Additionally, the variation of the

element conserved in Anthroidea was unique in matching another transcription factor, PHOX2B (**Fig. 3B/D**).

Lastly, we performed data mining of developmental and adult expression of *ADAMTSL1*, *TCF7L1*, and *PHOX2B* to demonstrate their theory building potential. To do so, we show the murine developmental expression of the two transcription factors in question (**Fig. 3D/E**) (while the rodent lineage has lost the element, we are not aware of a reason to expect that to have affected *Tcf7l1* expression) [31]. The reference atlas of brain development shows a rostro-caudal gradient of *Tcf7l1* expression (**Fig. 3F**). That pattern would not only be absent with loss of *Tcf7l1* binding, but *Phox2b* has an opposite pattern of predominantly caudal expression. There were no equivalent developmental data for mouse *Adamts1* but it is known to be expressed in the one and two cell conceptus [38] and in the neural ectoderm at embryonic day 8.5 [39], and it seems likely to be expressed throughout brain development.

Discussion

We report here the first phenome-wide association study (PheWAS) conducted in dogs. Using loci identified in a Labrador Retriever GWAS of behavioral elimination from the TSA odor detection program, we tested for associations with individual items from the C-BARQ questionnaire, which captures the most common problem behaviors in dogs. In three independent Labrador Retriever cohorts, the PheWAS identified six loci associated with 15 C-BARQ traits. Notably, the most frequently associated traits were four aggression items directed toward familiar dogs, overlapping with GWAS loci linked to program elimination in TSA dogs. Updated gene annotation of these GWAS/PheWAS loci revealed several novel candidate genes. To explore potential functional variation, we performed whole genome sequencing on a random sample of seven Labrador Retrievers and seven German Shorthaired Pointers currently in the TSA program. This enabled detailed analysis of a mapped haplotype spanning much of the *ADAMTSL1* gene. We identified a top candidate functional variant located within a putative enhancer in an *ADAMTSL1* intron, which displays a notable evolutionary trajectory across mammals. Transcription factor binding site analysis suggests this element is typically bound by the repressor TCF7L1/2 in most mammals, but by PHOX2B in simians. We hypothesize that the canine risk allele is hypomorphic. Finally, we present developmental gene expression data for these transcription factors in mice to support theory building. These findings highlight the tractability of the mouse model for functional testing of the behavioral effects of this locus and offer a testable hypothesis: that the variant alters

ADAMTSL1 expression during development through changes in transcriptional regulation.

The original GWAS presents two challenges [19]. First, the cohort was the result of a one-time (2002-2013) TSA breeding and training program that used Labrador Retrievers from the Australian Customs Service and commercial sources in the US. Additionally, the Australian dogs were part of a breeding and training program that began in 1993 and were selected for the same work as the TSA dogs [20]. Second, the mapped trait was elimination for behavioral reasons, without specific phenotype data to explain the reason for elimination. Thus, replication of the GWAS was not possible.

The present work serves as an alternative to a formal replication of the original GWAS. The initial mapping was supported by strong behavioral relevance at several loci implicating single genes (e.g., *CHD2*, *NRG3*, and *PDE1A*); however, such evidence is not sufficient to validate the findings. Similarly, while positive PheWAS results lend support to the original associations, the nonspecific nature of the GWAS trait warrants caution in interpreting these results as formal validation. Future studies, including ongoing work in current TSA dogs with behavioral questionnaire data (e.g., CTC-BARQ [40]), will allow for direct comparison with the 2013 TSA GWAS and the present PheWAS findings. We anticipate that some associations will be reinforced and that their effect sizes can be more precisely estimated, ultimately strengthening the case for specific haplotypes. Beyond replication of genetic association, it is important to identify the responsible genes and variants. Below we discuss this work's leveraging of the rich catalogs of human GWASs and of vertebrate genome sequences to nominate such functional variations. That can be followed by the creation of variant knock-ins in model species that have the relevant behaviors, possibly zebrafish or mice. Alternatively, with the owner's consent, it would be possible to conduct veterinary clinical trials in pet dogs to treat problem behaviors using any approach available. This is possible because, unlike humans, dogs tend to have many moderate to large effect variations associated with diverse traits - as we showed in our original GWAS that led to this work [19].

PheWAS Results

The original GWAS for behavioral elimination in a 2013 TSA detection dog cohort of Labrador Retrievers [19] identified 22 SNPs across 11 loci, which were subsequently screened in the PheWAS. Of those SNPs, only 13, 9, and 8 met inclusion criteria in the AUS, UK, and TSE cohorts, respectively. Given that all genotyping used the same SNP array platform, known to be reliable

across these cohorts, the most likely explanation for the reduced number of usable SNPs is insufficient allele frequency in some populations. This may also help explain the overall low number of significant associations observed in single-cohort PheWAS analyses. For example, only one of the 13 single-cohort associations appeared in the Seeing Eye (TSE) cohort, a population that has been under long-term selection for guide work and may therefore have experienced negative selection against alleles associated with problematic behaviors. In contrast, the Australian pet cohort, which likely underwent the least behavioral selection, showed more nominal associations, albeit with consistently higher p-values, possibly reflecting higher allele frequencies but weaker effect sizes or less phenotype specificity. The UK cohort, with a mix of pet and working (gamebird retrieval) dogs, showed intermediate patterns, consistent with partial selection. Together, these observations suggest that both allele frequency and behavioral selection history shape the PheWAS signal. The most informative future PheWAS cohorts for studying common genetic variation in the TSA Labrador Retriever population may be working/hunting-line Labrador Retrievers, which are currently being analyzed in our ongoing work.

The PheWAS netted a total of 15 C-BARQ associations with 8 SNPs at 6 genome loci. The strongest p-value, in the Australian cohort, was for chr1 and agg_25, defined as aggression to a household member staring at them, which was also significant for the second SNP at that locus. The same locus in the same cohort was also associated with agg_13, aggression when an object is taken away by a household member. That same agg_13 trait also showed the second strongest p-value, this time for association with chr19, again in the Australian cohort. That chr19 marker had the most associations, including two traits each in the UK cohort and UK-TSE meta-analysis. It is discussed further in the Biological annotation section below.

Early life stressors exert a disproportionate influence on later behavior and risk of psychopathology [41]. While trauma has a profound impact, psychological and physical stress can also arise from neglect or deprivation. This holds true for dogs, and it is especially important that owners and foster carers of working dogs remain vigilant about the social and environmental exposures that are essential for successful working careers and those that may increase the risk of problem behaviors [42]. Odor detection dogs, for instance, undergo rigorous training while housed in kennels and often work in demanding environments such as busy airport terminals; their early care must prepare them for these challenges. Our PheWAS results highlight aggression, particularly when directed at familiar dogs and humans. We previously showed that common genetic variants with large effect sizes on both social and nonsocial fear were also associated with aggression [9, 10].

However, distinct sets of variants were linked to aggression directed at familiar versus unfamiliar dogs and humans. Going forward, it will be important to determine whether this aggression reflects fear-based (reactive) behavior or non-fear-based (proactive) aggression, and to better understand its underlying causes. Identifying dogs at risk early in life could allow for tailored interventions to prevent the development of problematic behaviors. In contrast, the PheWAS finding related to food stealing is more difficult to interpret. While it may involve gene-environment interactions, it could also reflect innate survival behaviors, particularly those shaped by selection for scavenging during domestication [43].

The chr6 locus was associated with urination when approached, petted, or handled, a behavior that may indicate fear or stress sensitivity. This finding may be especially relevant for working dogs, as the original GWAS was based on dogs eliminated during or shortly after the foster period in the 2013 TSA cohort. Previous studies have shown that working dogs are prone to separation-related problems and hyperactivity or restlessness [17, 44], especially when left alone for extended periods during early development [44]. The new association suggests that carriers of the chr6 risk allele may be particularly sensitive to early-life stress, including during transitions between environments. Identifying such dogs early could enable targeted strategies to mitigate these vulnerabilities, particularly during the foster period and the transition into kennel-based training environments.

The paucity of PheWAS signal in the Seeing Eye cohort invites looking at the positive traits. Both the SNP and trait were uniquely positive in this cohort: chr13 and fear_51, fear when feet are towed by a member of the household. This may be interesting because it is the only fear trait identified. [Note: The C-BARQ item refers to a household member for dogs from a breeding and training program, like the 2013 TSA dogs in the original GWAS. That is because the questionnaires were completed by puppy raisers in the foster period.] Now it is possible to closely follow carriers of this allele. This begins with the foster period, which presents diverse grooming and health care opportunities to identify foot sensitivities and their contexts. The most common sensitivities of the feet are due to bad experiences with nail trimming. These can include stresses from restraint, unfamiliar/stress-associated environments, and extremely painful injury and associated medical complications. Another question that emerges is whether the stimulus for the dog includes aversion to towels for this association. This would be of interest to the TSA as some of their testing involves the use of scented towels. The two other signals in the Seeing Eye cohort were in the UK-TSE meta-analysis: exc_67, becoming excitable when visitors arrive in the home; and mis_74, chasing cats. These are common behaviors and are

associated with the chr19 locus, which we mentioned for having the most positive traits and discussed below.

This PheWAS focused on behavioral associations. However, the original GWAS cohort consisted of TSA odor detection dogs during their year-long foster period, during which they were tested every three months. While olfactory acuity was not directly assessed, some tasks, such as locating objects by scent, relied on olfactory abilities. Therefore, we cannot rule out the possibility that genetic variants associated with behavior may also be directly or indirectly linked to olfactory traits. While humans primarily navigate the world visually, olfaction is the most critical sense in dogs [45]. It was recently discovered that dogs are the only known species to have a direct and substantial connection between the olfactory bulb and the visual processing area of the brain (occipital lobe) [46]. The function of this connection is unknown, but it suggests direct integration of olfaction and vision. Despite these facts, we cannot yet determine to what extent smell is required to discern stimuli for traits directed at familiar vs. unfamiliar dogs and humans. Given the extreme importance of olfaction for canine discrimination, odor memory, and social interactions, it seems inconceivable that their sense of smell could be irrelevant. Conducting joint studies of olfaction and behavior will be important to answer these questions.

All four positively associated questions in the aggression factor and the one in the fear factor pertain to behaviors directed toward familiar people or dogs, typically members of the household. Consistent with that, the only positively associated item in the separation behavior factor also involves familiar individuals. Together, these items account for five of the six significant loci identified in this study. One additional significant association involved chasing small animals, which is distinct from human- or dog-directed behaviors. This finding is particularly relevant given the preferred use of retrievers and pointers as working dogs that must interact safely with the public. These breeds have a non-threatening, "floppy-eared" appearance, and there is essentially a zero-tolerance policy in the TSA detection dog program for aggression directed at anyone. Notably, in this study, aggression and fear were directed exclusively at familiar humans and dogs. This suggests a problem not with initial exposure to strangers, but with early social bonding, emotional regulation, or sensitivity in familiar contexts. These findings highlight the importance of early detection and tailored intervention during the first year of life, when dogs are developing core relationships and behavioral patterns [42]. Genotyping all dogs in working dog cohorts could allow identification of carriers of relevant variants, who might benefit from more individualized support during the foster period.

In addition to identifying significant hits, the C-BARQ questions were indicated as appearing or not in the new Working Dog and Canine Training Center versions of C-BARQ, referred to as WDC-/CTC-BARQ, which are slightly different from each other (**Table 2; bold**). Those are tailored to working dogs in general and the TSA CTC program more specifically. They were designed to monitor relevant problem behaviors and to remove questions that pertain to pets but not working dogs that are housed outside homes. The WDC-BARQ was validated in a survey of 1,117 working odor detection dogs [40]. Seven of the 12 C-BARQ questions significant here are captured in the WDC-/CTC-BARQ.

Biological Annotation

Except for the *ADAMTSL1* locus discussed below, we presented the biological annotation of candidate genes in the original GWAS paper [19]. The chr19 locus lacks a clear prioritization of candidate genes but merits re-examination given the new findings. This locus was the only one with signal in all three cohorts, had the second-strongest p-value, and was associated with the most C-BARQ phenotypes. Beginning with the strongest p-value, the chr19 PheWAS traits were aggression when an object is taken away by a household member (AUS), stealing food (UK), persistent barking when alarmed/excited (UK), aggression when stolen food is taken away by a household member (AUS), becoming excitable when visitors arrive at the home (UK-TSE), and chasing cats. It is possible there is a biological implication such as reactivity and excitability or excitatory-inhibitory imbalance, but further investigation is necessary.

Notably, there is no single gene implicated positionally in the chr19 locus. The nearest flanking mammalian genes are *PLEKHB2* ~190kb upstream and *HS6ST1* ~940kb downstream. These genes and others in their vicinity are enriched in the brain. We also noted previously ENSCAFG00000004229, an uncharacterized *GLUD1* retrogene ~345kb downstream [19]. Here we report its 5' end is a single open uninterrupted reading frame of 445 amino acids that account for all exons 1-11 and the 3' end is a poly-A sequence. It is present in the same genome location in the gray wolf and absent in the panda and cat (UCSC Genome Browser mCanLor1.2, ailMel1, and felCat9 assemblies, respectively). Based on the enhanced annotation mRNA tracks [47], this retro-*GLUD1* is widely expressed, including in the blood and brain. *GLUD1* or Glutamate Dehydrogenase 1 is a major regulator of glutamate levels in the central nervous system (we use the HUGO official gene name and symbol; note some authors use this symbol instead of the correct GRID1 to refer to glutamate ionotropic receptor delta type subunit 1). This

observation suggests the possibility of investigation of the retrogene in dogs and, if validated, testing if the mapped haplotype has altered expression and behavioral consequences.

Nomination of a Functional Variant at the Mapped ADAMTSL1 Locus and Theory Building

We identified the mapped haplotype overlapping *ADAMTSL1* (aka Punctin1) in whole genome sequence data from our ongoing study of current TSA dogs. It is related to the ADAMTS protein family but lacks the protease activity. It is an extracellular matrix protein with both structural and signaling roles in axon guidance, cell migration, and synapse organization known to have conserved functions with those of the worm *C. elegans* homolog MADD-4 [48, 49]. The discussion that follows regards the nomination of the most likely functional candidate and analysis of the potential biological relevance. It is important to understand these aspects are not intended to validate the mapping. Rather, these studies were done to develop testable theories.

The top-most evolutionarily conserved position that is variant in the mapped haplotype is in an *ADAMTSL1* intron. According to its CADD score for deleteriousness, it is in the top 1% most constrained bases in the human genome. It lies in the middle of a 17bp conserved element unique to mammals (PhastCons lod=38; UCSC Genome Server, hg38). We *hypothesize* that the mammalian conservation of this element is due to binding TCF7L1 or 2. This is based on our prediction of transcription factor binding sites, but it is possible that another transcription factor with an unknown binding motif binds that site. As TCF7L1 is a repressor [50], reduced or lost binding in the phyla mentioned - including in carriers of the dog risk allele - could increase *ADAMTSL1* expression. We further hypothesize the loss or reduced binding activity in Anthropoidea, Artiodactyla, rodents, marsupial species with the deletion, and T allele dogs resulted in a reduced threshold for some aggression or fear traits.

Our analysis of Human Protein Atlas [32] data showed that, on average, *TCF7L1* is expressed two-fold higher than *TCF7L2* in human inhibitory vs excitatory neuron mRNA single cell RNA sequencing clusters and vice versa. We searched for evidence of their co-expression with *ADAMTSL1* and found examples involving *TCF7L2* in both humans and mice. co-expression occurs in a neuronal single-cell RNA sequencing cluster (brain cluster 38), defined by a shared expression pattern of 321 genes (Human Protein Atlas [32]). In mice, co-expression was observed in glucokinase-positive (Gck⁺) neurons of the anterior paraventricular nucleus of the thalamus (aPVT), characterized using a genetically modified mouse model designed to study this specific

neuron type [51]. That study employed deep mRNA sequencing to compare Gck⁺ aPVT neurons with their Gck⁻ counterparts and conducted tract-tracing to map their connectivity. These neurons were shown to project strongly to forebrain regions involved in reward and stress processing, as well as hypothalamic regions critical for energy metabolism and feeding behavior.

The human and mouse neuron types mentioned above, as well as others, could be used to study whether TCF7L1/2 bind to the *ADAMTSL1* regulatory element identified in this study. This element, along with its flanking intronic sequence, is present in rabbits but was lost in rodents after their divergence. To test whether the dog variant has behavioral effects in rodents, the corresponding intronic sequence would need to be knocked in, introducing both the reference and alternative alleles in parallel, for functional comparison. In Anthroidea, the loss of TCF7L1 binding to the *ADAMTSL1* element may have been replaced by PHOX2B binding in subcortical regions, where its effect could be either activating or repressive depending on context [52]). These regulatory differences, along with the other variants discussed here, could be studied in mouse knock-in models to assess their effects on behavior and underlying mechanisms.

In the same way that coding variations are catalogued and ranked by deleteriousness in diverse species, the approach presented here could be applied to genome wide scans of canine non-coding variations. This information would be collected with breed, linkage disequilibrium/haplotype data, and available genetic associations with traits. For informative sites, this circumvents the great limitation of interpreting genome sequence data of, say, hundreds of mammals with one representative of each species. Highly conserved positions may be variant in one or more species but will not be feasible to study because their variation is unknown, or phenotyping is not tractable. As we found here, conserved elements with variation segregating in dogs could be prioritized for investigation - especially when genetically mapped for an important trait - and lead to animal models in model organisms.

Mining of *ADAMTSL1* biological relevance from traits mapped genetically and patterns of gene expression are consistent with our mapping for behavior in dogs. However, this evidence is so broad that it fails to rule out essentially any brain or sensory function. Traits genetically mapped in humans include visual acuity, behavior, cognition, drug use, psychopathology, and brain structure[29]. The highest mRNA expression in adult humans is in excitatory neurons of the brain, followed by inhibitory neurons. Adult mouse gene expression shows enrichment in the retina and other eye tissues, and the glomerular layer of the olfactory bulb. Transcript expression is present in the

conceptus and in neuroectoderm during embryonic brain development, but we did not find spatial localization studies.

We recently reported that brain imaging shows that as dog body/brain size decreases, the ratio of cortex to subcortex decreases[53]. We proposed there that increased behavioral reactivity with decreasing body and brain size in dogs could thus be explained by that, since the cortex is generally more inhibitory and the subcortex more excitatory. Given the predominant expression of *ADAMTSL1* in excitatory neurons in the brain, it will be interesting to know if this element is associated with variation in excitatory-inhibitory balance across mammals.

Conclusions

Major challenges in dog genetics include the lack of standardized phenotyping, interbreed structure, and controlled environments. Large, purpose-bred working dog cohorts can help overcome these limitations. Building on a previous GWAS of behavioral elimination in TSA odor detection Labradors, we conducted a PheWAS of mapped loci in three independent Labrador populations. The strongest signals involved four aggression traits and one fear trait, all directed at familiar people or dogs. This is notable because such behaviors are disqualifying in working dogs and may reflect issues in early development. The findings suggest that attention to behavior in familiar contexts, especially during fostering, could improve outcomes. We also highlight separation-related behaviors and excitability, traits known to influence working success. Finally, we demonstrate the utility of whole genome sequencing in identifying candidate variants and mechanisms. Together, these results extend a general GWAS of problem behavior into phenotype-specific associations that may improve prediction and prevention in working and pet dog populations.

Methods

Data Acquisition and Quality Control

For this PheWAS study, three cohorts of Labrador Retrievers with available genotype and C-BARQ data were used. In cohorts with multiple breeds, we only used the Labrador Retrievers. The Australian (AUS) dogs were predominantly pets. The United Kingdom (UK) dogs were mixed pets and working dogs. The working UK dogs were gamebird retrievers in the structured activity of 'shoots' in which dogs of other breeds such as spaniels 'flush' gamebirds out of hiding. Such dogs are generally housed in kennels and undergo training with great attention to socialization, exercise, and health care. Data from published studies of the UK and AUS cohorts were

obtained from those investigators, who are contributors here [7, 27]. Data and DNA for The Seeing Eye cohort were originally sourced by H. J. H. from The Seeing Eye, Inc. and provided to this program in an agreement between the Battelle Memorial Inst. and The Seeing Eye, Inc. The samples used in this study are from dogs privately owned by The Seeing Eye. Because the data and DNA samples were from The Seeing Eye's repository and a part of their normal data biological sampling routine, an IACUC protocol was not required to use samples from their repository. The Seeing Eye is a member organization of the International Guide Dog Federation (www.igdf.org.uk), meeting their standards for humane care of the dogs and veterinary standards. The samples used were previously taken by The Seeing Eye Vincent A. Stabile Canine Health Center, accredited by the American Animal Hospital Association (AAHA), for the purpose of health monitoring.

All genotyping in the original GWAS and this study was based on the Illumina CanineHD Beadchip Genotyping platform (the Embark Illumina version with ~220 k SNPs for the Seeing Eye cohort and the Illumina version with data for ~173 k SNPs for the others). Quality control of genomic data was then performed with the following settings in PLINK (version 1.9, <https://www.cog-genomics.org/plink/>): minor allele frequency cutoff of 2.5%, Hardy-Weinberg equilibrium filter of $1e-5$, maximum SNP missing rate of 10%, and maximum genotyping rate of 25%. The SNPs used in the PheWAS were identified in a GWAS of elimination from a 2002-2013 TSA pretraining program for behavioral reasons in Labrador Retrievers ($n=296$) [19]. The top 21 SNPs identified were extracted, if available, from each cohort following quality control. One additional marker, for a total of 22 SNPs, was added from a re-analysis of that cohort using GEMMA that is part of an ongoing study. The software was utilized by first calculating a centered relatedness matrix then performing a likelihood ratio test linear mixed model on the quality-controlled data. The lambda value showed proper control of genomic inflation, and the SNP was genome-wide significant. It was chosen because it was the only finding for which both alleles were segregating in the two largest PheWAS cohorts here. The 100 C-BARQ questions [14] graded on a 0-4 Likert scale were also quality controlled by first removing questions with greater than 25% missing data then removing dogs with greater than 25% missing data.

Phenome-Wide Association Studies (PheWAS) and Intersection Union Tests (IUT)

A PheWAS was performed for each cohort in R (version 4.2.2, <https://www.R-project.org/>) and the PheWAS package (version 0.99.5-5,

<https://github.com/PheWAS/PheWAS>). These results were uploaded into Microsoft Excel for IUT analysis using the Fisher's combined probability test to combine p-values across cohorts, and Bonferroni correction to determine an adjusted threshold for significance. Manhattan plots, volcano plots, and relative histograms for C-BARQ scores based on genotype were made using custom scripts written in Python (version 3.9.16, <https://www.python.org/>) using the seaborn package (version 0.12.2, <https://seaborn.pydata.org/>).

Whole Genome Sequencing and Analysis

In ongoing studies, we are collecting new data and DNA from dogs in the current TSA odor detection program, which are distinct from both the original TSA GWAS cohort and the three PheWAS populations. As part of this effort, a subset of fourteen TSA detection dogs was sampled using cheek swabs under the approved Institutional Animal Care and Use Committee (IACUC) protocol number 807033 titled "Defining multiple dimensions of the working dog phenotype in client owned dogs"; Principal Investigator: Cynthia Otto (**Table 3**). All these methods were performed in accordance with the guidelines and regulations. DNA was isolated using the QIAamp 96 DNA QIAcube HT Kit (QIAGEN) and concentration measured by NanoDrop spectrophotometry (Thermo Fisher Scientific). The quality of starting DNAs and sequencing libraries were assessed by capillary electrophoresis (High Sensitivity DNA Assay, Agilent 2100 Bioanalyzer) and library yields were measured by Qubit fluorometry (Thermo Fisher Scientific). Whole genome sequencing was performed as 2 x 150 bp paired end sequencing at 20-fold coverage (Illumina NovaSeq 6000; Genomics Shared Resource, The Ohio State University Comprehensive Cancer Center). The qualities of the reads both before and after mapping were assessed using FastQC (version 0.12.1, <https://www.bioinformatics.babraham.ac.uk/projects/fastqc/>) [54]. The samples had a uniform FastQC score of 37 or higher. That Phred quality score of 37 assesses the accuracy of each base call and translates to a base call accuracy of 99.98% (i.e., a 0.02% chance of an incorrect base call). Sequence data were mapped to the CanFam3.1 reference genome [47] using minimap2 (version 2.28, <https://github.com/lh3/minimap2>) with settings for short genomic paired end reads [55]. The GATK (version 4.5.0.0, <https://gatk.broadinstitute.org/hc/en-us>) [56] realigner tool was used to reduce the number of miscalls of indels. File conversions to a Variant Call Format (VCF) were performed using a combination of bcftools (version 1.17, <https://github.com/samtools/bcftools>) and samtools (version 1.19, <https://github.com/samtools>) [57]. Visualization and work with this data was performed with Integrative Genome Viewer (version 2.16.1, <https://github.com/igvteam/igv>) [58].

For analysis of the WGS, the sequence data were converted to the VCF format. VCF formatted WGS were analyzed using the Integrative Genomics Viewer (IGV) [58]. Canine and comparative genome analyses were performed using custom UCSC Genome Browser [59] tracks generated from the IGV VCF formatted data. Canine sequence analyses were primarily conducted on the UCSC CanFam3.1 assembly (Boxer breed; in some cases, compared builds and variations across the new build of the same Boxer individual, CanFam6, or the two other breeds available there, German Shepherd, Great Dane), with variants of interest checked for presence in three genome builds of Labrador Retrievers (NCBI Genome Assembly names, ASM1204487v1/ ASM1204501v1, Yella_v2, ROS_Cfam_1.0). Conversion of genome positions from dog to human and other species was performed using the LiftOver function on the UCSC browser. The following UCSC tracks are curated reference datasets that were loaded for analysis of the present WGS data on the human hg38 assembly: Vertebrate and mammalian phylogenetic alignments and conservation statistics called, Conservation (100 vertebrate species) [35], Cactus and Multiz (447 and 470 mammalian species, respectively) [35, 36]; and Combined Annotation Dependent Depletion [34] genome-wide annotation of deleteriousness of SNPs and insertion/deletion variants in the human genome.

Transcription Factor Binding Site Prediction

Transcription factor binding sites were predicted spanning variant positions prioritized high evolutionary conservation or other evidence of deleteriousness. Find Individual Motif Occurrences (FIMO) software (version 5.5.5, <https://meme-suite.org/meme/doc/fimo.html>) [60] and the JASPAR 2022 Core Vertebrate set of transcription factor binding site motifs [37] were used. FIMO scans DNA sequence with position specific scoring matrices, which define the binding site motifs for known transcription factors. The program's output is a p-value based on the log-likelihood ratio score at each base in a sequence, and a q-value, which estimates the false discovery rate. Both alleles at variant positions were analyzed using FIMO with its default threshold of $p < 1E-04$. Motifs present only in the wild type or variant sequence, or which varied significantly in p-value, were manually analyzed by both sequence logo alignment and phylogenetic alignment analysis on the UCSC Genome Browser. Motif information and sequence logo visualization were accessed from the JASPAR database for eukaryotic transcription factor binding profiles [37].

References

1. Bray, E.E., et al., *Enhancing the Selection and Performance of Working Dogs*. Front Vet Sci, 2021. **8**: p. 644431.
2. Leigh, P. *Dogs in Demand for Explosives Detection Work in U.S.* 2022; Available from: <https://www.akc.org/expert-advice/news/dogs-in-demand-explosives-detection-us/>.
3. Ostrander, E.A., et al., *Demographic history, selection and functional diversity of the canine genome*. Nat Rev Genet, 2017. **18**(12): p. 705-720.
4. Lappalainen, T., et al., *Genetic and molecular architecture of complex traits*. Cell, 2024. **187**(5): p. 1059-1075.
5. Darwin, C., *On the Origin of Species by Means of Natural Selection, or, the Presentation of Favoured Races in the Struggle for Life*. 1859, London: John Murray.
6. Hayward, J.J., et al., *Complex disease and phenotype mapping in the domestic dog*. Nat Commun, 2016. **7**: p. 10460.
7. Ilska, J., et al., *Genetic Characterization of Dog Personality Traits*. Genetics, 2017. **206**(2): p. 1101-1111.
8. MacLean, E.L., et al., *Highly heritable and functionally relevant breed differences in dog behaviour*. Proc Roy Soc Biol Sci, 2019. **286**(1912): p. 20190716.
9. Zapata, I., et al., *Genome scans of dog behavior implicate a gene network underlying psychopathology in mammals, including humans*. bioRxiv (non-reviewed preprint), 2021.
10. Zapata, I., J.A. Serpell, and C.E. Alvarez, *Genetic mapping of canine fear and aggression*. BMC Genomics, 2016. **17**: p. 572.
11. Alvarez, C.E., van Rooy, D., Wade, C. M., *Behavioral Genetics of Dog Breeds*, in *Encyclopedia of Animal Behavior*, G.J. Thompson, Editor. 2025, Elsevier.
12. Salonen, M., et al., *Prevalence, comorbidity, and breed differences in canine anxiety in 13,700 Finnish pet dogs*. Sci Rep, 2020. **10**(1): p. 2962.
13. Morrill, K., et al., *Ancestry-inclusive dog genomics challenges popular breed stereotypes*. Science, 2022. **376**(6592): p. eabk0639.
14. MacLean, E., N. Snyder-Mackler, and J. Serpell, *Highly Heritable and Functionally Relevant Breed Differences in Dog Behavior*. bioRxiv, 2019: p. 509315.
15. Dutrow, E.V., J.A. Serpell, and E.A. Ostrander, *Domestic dog lineages reveal genetic drivers of behavioral diversification*. Cell, 2022. **185**(25): p. 4737-4755 e18.
16. Sarviaho, R., et al., *Two novel genomic regions associated with fearfulness in dogs overlap human neuropsychiatric loci*. Transl Psychiatry, 2019. **9**(1): p. 18.
17. Zapata, I., et al., *Genetic testing of dogs predicts problem behaviors in clinical and nonclinical samples*. BMC Genomics, 2022. **23**(1): p. 102.
18. Lofgren, S.E., et al., *Management and personality in Labrador Retriever dogs*. Applied Animal Behaviour Science, 2014. **156**: p. 44-53.
19. Eyre, A.W., et al., *Genome scanning of behavioral selection in a canine olfactory detection breeding cohort*. Sci Rep, 2022. **12**(1): p. 14984.
20. Champness, K.A., *Development of a breeding program for drug detector dogs: based on studies of a breeding population of guide dogs*, in *Department of Agriculture and Resource Management*. 1996, The University of Melbourne.

21. Eyre, A.W., et al., *Machine learning prediction and classification of behavioral selection in a canine olfactory detection program*. Sci Rep, 2023. **13**(1): p. 12489.
22. Bray, E.E., et al., *Predictive Models of Assistance Dog Training Outcomes Using the Canine Behavioral Assessment and Research Questionnaire and a Standardized Temperament Evaluation*. Front Vet Sci, 2019. **6**: p. 49.
23. MacLean, E.L. and B. Hare, *Enhanced Selection of Assistance and Explosive Detection Dogs Using Cognitive Measures*. Front Vet Sci, 2018. **5**: p. 236.
24. Denny, J.C., et al., *PheWAS: demonstrating the feasibility of a phenome-wide scan to discover gene-disease associations*. Bioinformatics, 2010. **26**(9): p. 1205-10.
25. Thornton, T. and M.S. McPeck, *ROADTRIPS: case-control association testing with partially or completely unknown population and pedigree structure*. Am J Hum Genet, 2010. **86**(2): p. 172-84.
26. Zhou, X. and M. Stephens, *Genome-wide efficient mixed-model analysis for association studies*. Nat Genet, 2012. **44**(7): p. 821-824.
27. van Rooy, D. and C.M. Wade, *Association between coat colour and the behaviour of Australian Labrador retrievers*. Canine Genet Epidemiol, 2019. **6**: p. 10.
28. Parker, H.G., et al., *Genomic Analyses Reveal the Influence of Geographic Origin, Migration, and Hybridization on Modern Dog Breed Development*. Cell Rep, 2017. **19**(4): p. 697-708.
29. Welter, D., et al., *The NHGRI GWAS Catalog, a curated resource of SNP-trait associations*. Nucleic Acids Research, 2014. **42**(D1): p. D1001-D1006.
30. Wu, C., I. MacLeod, and A.I. Su, *BioGPS and MyGene.info: organizing online, gene-centric information*. Nucleic Acids Research, 2013. **41**(D1): p. D561-D565.
31. Thompson, C.L., et al., *A high-resolution spatiotemporal atlas of gene expression of the developing mouse brain*. Neuron, 2014. **83**(2): p. 309-323.
32. Karlsson, M., et al., *A single-cell type transcriptomics map of human tissues*. Sci Adv, 2021. **7**(31).
33. Weiss, L.A. and M. Nieto, *The crux of Cux genes in neuronal function and plasticity*. Brain Res, 2019. **1705**: p. 32-42.
34. Kircher, M., et al., *A general framework for estimating the relative pathogenicity of human genetic variants*. Nat Genet, 2014. **46**(3): p. 310-5.
35. Blanchette, M., et al., *Aligning multiple genomic sequences with the threaded blockset aligner*. Genome Res, 2004. **14**(4): p. 708-15.
36. Kuderna, L.F.K., et al., *Identification of constrained sequence elements across 239 primate genomes*. Nature, 2024. **625**(7996): p. 735-742.
37. Castro-Mondragon, J.A., et al., *JASPAR 2022: the 9th release of the open-access database of transcription factor binding profiles*. Nucleic acids research, 2022. **50**(D1): p. D165-D173.
38. Huang, Y., et al., *Stella modulates transcriptional and endogenous retrovirus programs during maternal-to-zygotic transition*. Elife, 2017. **6**.
39. Brunskill, E.W., et al., *A gene expression atlas of early craniofacial development*. Dev Biol, 2014. **391**(2): p. 133-46.
40. Hare, E., et al., *Development of a modified C-BARQ for evaluating behavior in working dogs*. Frontiers in Veterinary Science, 2024. **11**.

41. Humphreys, K.L. and C.H. Zeanah, *Deviations from the expectable environment in early childhood and emerging psychopathology*. Neuropsychopharmacology, 2015. **40**(1): p. 154-70.
42. Serpell, J., Duffy, D. L., Jagoe, J. A., *Becoming a dog: Early experience and the development of behavior*, in *The Domestic Dog: Its Evolution, Behavior and Interactions with People*, J. Serpell, Editor. 2016, Cambridge University Press. p. 93-117.
43. Serpell, J.A., *Commensalism or Cross-Species Adoption? A Critical Review of Theories of Wolf Domestication*. Front Vet Sci, 2021. **8**: p. 662370.
44. Foyer, P., et al., *Behaviour and experiences of dogs during the first year of life predict the outcome in a later temperament test*. Applied Animal Behaviour Science, 2014. **155**: p. 93-100.
45. Horowitz, A. and B. Franks, *What smells? Gauging attention to olfaction in canine cognition research*. Anim Cogn, 2020. **23**(1): p. 11-18.
46. Andrews, E.F., et al., *Extensive Connections of the Canine Olfactory Pathway Revealed by Tractography and Dissection*. J Neurosci, 2022. **42**(33): p. 6392-6407.
47. Hoepfner, M.P., et al., *An Improved Canine Genome and a Comprehensive Catalogue of Coding Genes and Non-Coding Transcripts*. PLoS ONE, 2014. **9**(3): p. e91172.
48. Cizeron, M., et al., *Specific heparan sulfate modifications stabilize the synaptic organizer MADD-4/Punctin at Caenorhabditis elegans neuromuscular junctions*. Genetics, 2021. **218**(4).
49. Taye, N., C. Redhead, and D. Hubmacher, *Secreted ADAMTS-like proteins as regulators of connective tissue function*. Am J Physiol Cell Physiol, 2024. **326**(3): p. C756-C767.
50. Bou-Rouphael, J. and B.C. Durand, *T-Cell Factors as Transcriptional Inhibitors: Activities and Regulations in Vertebrate Head Development*. Front Cell Dev Biol, 2021. **9**: p. 784998.
51. Gaspari, S., et al., *Structural and molecular characterization of paraventricular thalamic glucokinase-expressing neuronal circuits in the mouse*. J Comp Neurol, 2022. **530**(11): p. 1773-1949.
52. Dubreuil, V., et al., *The role of Phox2b in synchronizing pan-neuronal and type-specific aspects of neurogenesis*. Development, 2002. **129**(22): p. 5241-53.
53. Hecht, E.E., et al., *Neurodevelopmental scaling is a major driver of brain-behavior differences in temperament across dog breeds*. Brain Struct Funct, 2021. **226**(8): p. 2725-2739.
54. Andrews, S. *FastQC: A Quality Control Tool for High Throughput Sequence Data [Online]*. Accessed [April 17, 2024]; Available from: <https://www.bioinformatics.babraham.ac.uk/projects/fastqc/>.
55. Li, H., *Minimap2: pairwise alignment for nucleotide sequences*. Bioinformatics, 2018. **34**(18): p. 3094-3100.
56. Van der Auwera GA & O'Connor BD, *Genomics in the Cloud: Using Docker, GATK, and WDL in Terra (1st Edition)*. 2020: O'Reilly Media, Inc.
57. Danecek, P., et al., *Twelve years of SAMtools and BCFtools*. Gigascience, 2021. **10**(2).
58. Robinson, J.T., et al., *Integrative genomics viewer*. Nat Biotechnol, 2011. **29**(1): p. 24-6.

59. Navarro Gonzalez, J., et al., *The UCSC Genome Browser database: 2021 update*. Nucleic Acids Res, 2021. **49**(D1): p. D1046-D1057.
60. Grant, C.E., T.L. Bailey, and W.S. Noble, *FIMO: scanning for occurrences of a given motif*. Bioinformatics, 2011. **27**(7): p. 1017-8.
61. Gou, X., et al., *Whole-genome sequencing of six dog breeds from continuous altitudes reveals adaptation to high-altitude hypoxia*. Genome Res, 2014. **24**(8): p. 1308-15.
62. Plassais, J., et al., *Whole genome sequencing of canids reveals genomic regions under selection and variants influencing morphology*. Nat Commun, 2019. **10**(1): p. 1489.
63. Bergstrom, A., et al., *Origins and genetic legacy of prehistoric dogs*. Science, 2020. **370**(6516): p. 557-564.
64. Scornavacca, C., et al., *OrthoMaM v10: Scaling-Up Orthologous Coding Sequence and Exon Alignments with More than One Hundred Mammalian Genomes*. Mol Biol Evol, 2019. **36**(4): p. 861-862.

ARTICLE IN PRESS

Acknowledgements

We are grateful for the dog samples and data and thank all participating dog owners. We thank T. Jay and C. DeSanti for their project management and technical and logistical support, and J. Schuetter for statistical feedback. We thank the TSA CTC for making this work possible. We are grateful for critical feedback from officials of DHS S&T. We thank the University of Pennsylvania for supporting and hosting the C-BARQ behavioral assessment instrument used in this research.

Author Contributions

The study was designed by C.E.A., A.W.E., and I.Z. The analysis was performed by A.W.E., in coordination with I.Z. and C.E.A. The WGS, bioinformatics analyses, and annotation were performed by C.E.A. J.F. collected samples and data at the TSA CTC and K.D.C. isolated DNA and managed the samples. H.J.H. managed the genotyping operations and the data generated. The PheWAS genotype and C-BARQ data were generated by K.M.E. and H.J.H. (TSE cohort); S.C.B., P.W., and M.J.H. (UK cohort); and D.v.R. and C.M.W. (AUS cohort). J.A.S., C.M.O., K.M.E., M.J.H., P.W., D.v.R. C.M.W. provided C-BARQ and behavioral phenotype expertise. C.E.A., I.Z., A.W.E., H.J.H., K.M.E., D.v.R. C.M.W., and P.W. provided genetic expertise. C.E.A. and A.W.E. wrote the manuscript with the help of I.Z. All authors read and helped edit the manuscript.

Data Availability

The genomic sequencing data generated and analyzed during the current study are available in the European Variation Archive (EVA) at EMBL-EBI under accession number PRJEB81786, available at <https://www.ebi.ac.uk/eva/?eva-study=PRJEB81786>. The point of contact for data access is the corresponding author, C. E. A. (email: alvarez.73@osu.edu). The SNP data from the United Kingdom (UK) and Australian (AUS) cohorts are from published studies [7, 27]; and can be requested from P.W. and M.J.H., and C.M.W., respectively, at their institutions provided here. The Seeing Eye (TSE) SNP data must be requested from K.M.E. at the organization. A utility to explore the genotype data and results is available at: <https://aweyre7147-2013tsa-gwas--2013tsagwas-phewasresultsviewer-b4c5cg.streamlit.app/>

Ethics declarations

The authors declare no ethical concerns.

Competing interests

The authors declare no competing interests.

Consent for publication

Not applicable.

Funding

This work was supported by the Department of Homeland Security Science and Technology Directorate (S&T), Contract No.70RSAT19CB0000014 with Battelle Memorial Institute (Principal Investigators C.E.A., C.M.O., and J.A.S.; and A.W.E., I.Z., J.F., K.D.C., and H.J.H.). C.E.A. was also supported by grants from the American Kennel Club CHF (01660) and the Scottish Deerhound Club of America. The data was made available through a cooperative research and development agreement 14-EXD-002 between the Penn Vet Working Dog Center and the Department of Homeland Security Science and Technology Directorate

Cohort	Source	#SNPs	#PheWAS SNPs	#Questions
Australian (AUS)	D. van Rooy, C.M. Wade 2019	220,853	13	98
United Kingdom (UK)	J. Iliska, et al. 2017	108,829	9	100
Seeing Eye (SE)	Guide Dog Training	213,644	8	97

Table 1. Labrador Retriever cohorts used in this study. LR counts is the total number of Labrador Retrievers, #SNPs is the total number of SNPs in the cohort, #PheWAS SNPs is the number of SNPs from the prior GWAS study that passed QC parameters, and #Questions is the number of C-BARQ questions that passed QC parameters.

Cohort	P value							*	C-BARQ ID	C-BARQ Item
AUS		2.1E-04				5.7E-07			agg_13	Aggression when object taken away by household member
AUS	5.0E-05	1.5E-10							agg_25	Aggression to household member staring at them
AUS						3.8E-05			agg_31	Aggression when stolen food is taken away by household member
UK				2.2E-04					agg_33	Aggression when approached by a household dog at a favorite resting place
SE					3.1E-04				fear_51	Fear when feet are towed by a member of the household
AUS			2.1E-04						sep_56	Frequency of restlessness or pacing following separation-related behavior
UK						1.3E-05			exc_63	Frequency of persistent barking when alarmed or excited
UK-SE						1.9E-04			exc_67	Tendency to become excitable when visitors arrive at your home
UK-SE						3.3E-04			mis_74	Frequency of chasing cats given the opportunity
UK							1.5E-04	3.4E-04	mis_76	Frequency of chasing small animals
UK						3.3E-06			mis_83	Frequency of stealing food
AUS			7.8E-06						mis_87	Frequency of urinating when approached, petted, handled, or picked up
	chr1.22989459 (MC2R)	chr1.24927539 (MC2R)	chr6.76632282 (LRR7)	chr11.38609196 (ADAMTS1L)	chr13.59902870 (Many)	chr19.21040815 (HS6ST1)	chr36.25252101 (PDE1A)	chr36.25648690 (PDE1A)		

***C-BARQ Factors**

- Aggression
- Fear
- Separation related
- Excitability
- Miscellaneous

Table 2. PheWAS and Intersection Union Test results. Cohorts listed on the left with IUT hits indicated with a hyphen. The table displays results where the p-value (<0.05) was less than the Bonferroni corrected threshold calculated for each respective cohort or combination of cohorts with intensity. The gene(s) with the highest apparent biological relevance that overlap mapped intervals or are the closest to single SNPs are in parentheses below each SNP (see Suppl. Table S2 for full information). Asterisks (*) denote abbreviations of *PLEKHB2*, *GLUD1* retrogene, and *HS6ST1*. Questions in **bold** represent those that are present in the current versions of the WDC-/CTC-BARQ[40].

ID	Sex	Success	Elimination Reason	Breed
C608	F	Y	NA	GSP
E143	M	Y	NA	GSP
E179	M	NA	NA	GSP
E312	F	NA	NA	GSP
E419	F	N	Behavioral	GSP
E425	M	Y	NA	GSP
F228	F	NA	NA	GSP
E321	M	Y	NA	LR
E481	F	N	Behavioral	LR
F117	F	Y	NA	LR
F162	F	N	Behavioral	LR
F170	M	Y	NA	LR
F177	M	Y	NA	LR
F243	F	N	Environmental	LR

Table 3. Metadata for dogs included in the whole genome sequencing.

Canid	N	Dataset ¹	Alleles	MAF
Bat-eared fox	1	UCSC Cactus	C	NA
Raccoon dog	1	UCSC Cactus	C	NA
Arctic fox	1	UCSC Cactus	C	NA
Red fox	1	UCSC Multiz	C	NA
African wild dog	1	UCSC Cactus	C	NA
Golden Jackal	1	DogGD Browser	C	NA
Wolf ²	52	DogGD dataset 3	C/T	T:0.067
Dingo	1	DogGD Browser	C	NA
Ancient dog ³	27	DogGD dataset 4	C/T	T:0.35
Native dog ⁴	116	DogGD dataset 2	C/T	T:0.22
Dog ⁵	670	DogGD dataset 3	C/T	T:0.29

Table 4. Chr11:38,725,982 SNP presence and minor T allele frequencies.

¹References: DogGD Browser and dataset 2[61]; DogGD dataset 3[62]; DogGD dataset 4[63]. UCSC refers to Genome Browser tracks in hg38 assembly. Species with single samples are marked Not Applicable (NA) for Minor Allele Frequency (MAF).

²World wide sample

³27 ancient dog genomes from 100-10,900 years old from Europe, the Near East and Siberia (Median 4650 YO, 2601 s.d.)

⁴Diverse village dogs; Chinese indigenous dogs and breeds; a few dogs of other Asian and European breeds

⁵144 modern breeds; 54 wild canids; and 100 village dogs

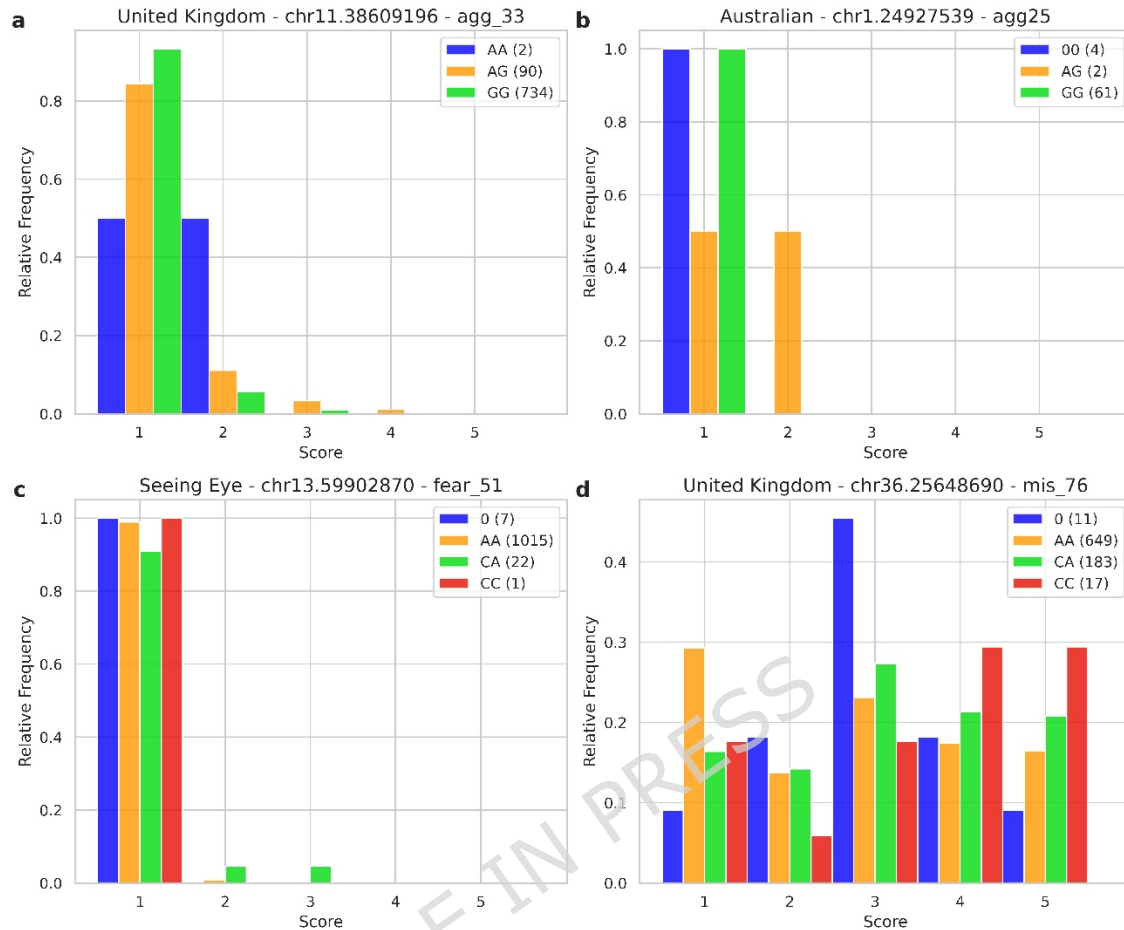


Figure 1. Relative histogram plots for select PheWAS significant hits. Relative proportion of dogs sorted by genotype and score for their respective question. (a) United Kingdom; chr38,609,196; agg_33 (b) Australian; chr1.24,927,539; agg_25 (c) Seeing Eye; chr13.59,902,870; fear_51 (d) United Kingdom; chr36.25,648,690.

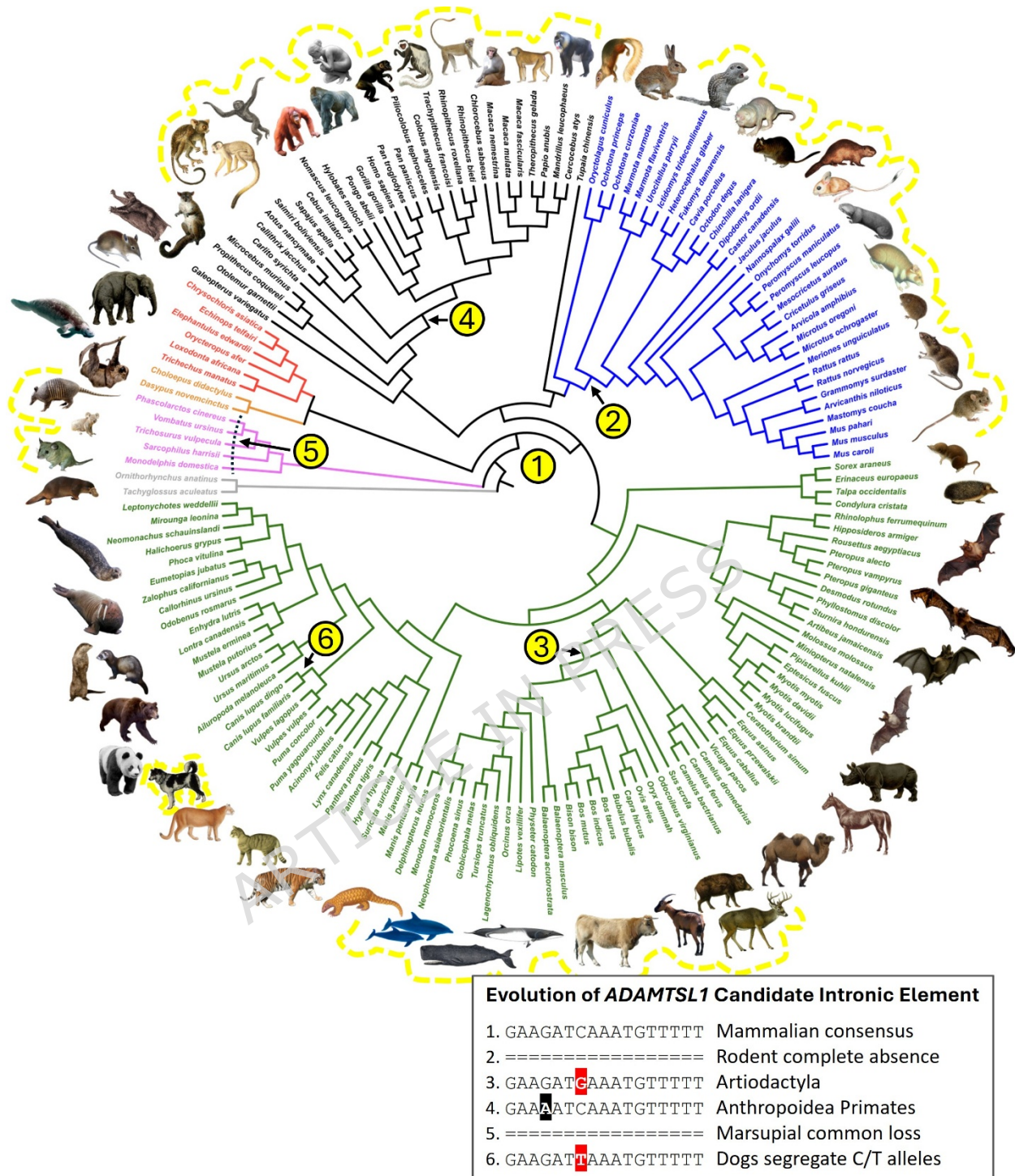


Figure 2. Phylogenetic Context of the Top-Ranked Whole Genome Sequence Candidate Functional Variant in the Chromosome 11 *ADAMTSL1* Locus. This locus was associated with TSA program elimination for behavior by GWA and C-BARQ “aggression_33” trait by PheWAS. The mammalian phylogenetic tree mainly focused on placental mammals (adapted from OrthoMaM [64], licensed under CC BY-NC 4.0) was used to

mark sequence variations in an *ADAMTSL1* intronic 17-bp conserved element (PhastCons lod=38, UCSC genome browser, hg38) affecting specific clades. The highlighted numbers correspond to the legend annotation and the yellow outline of phylogeny shows the animal groups affected. See Figure 3 for the sequence and transcription factor motif contexts.

ARTICLE IN PRESS

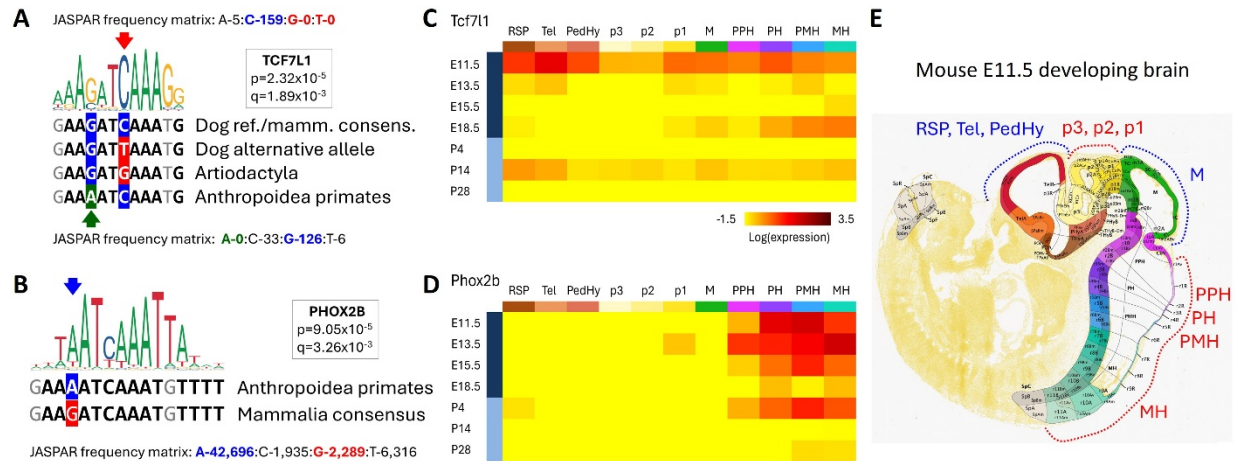
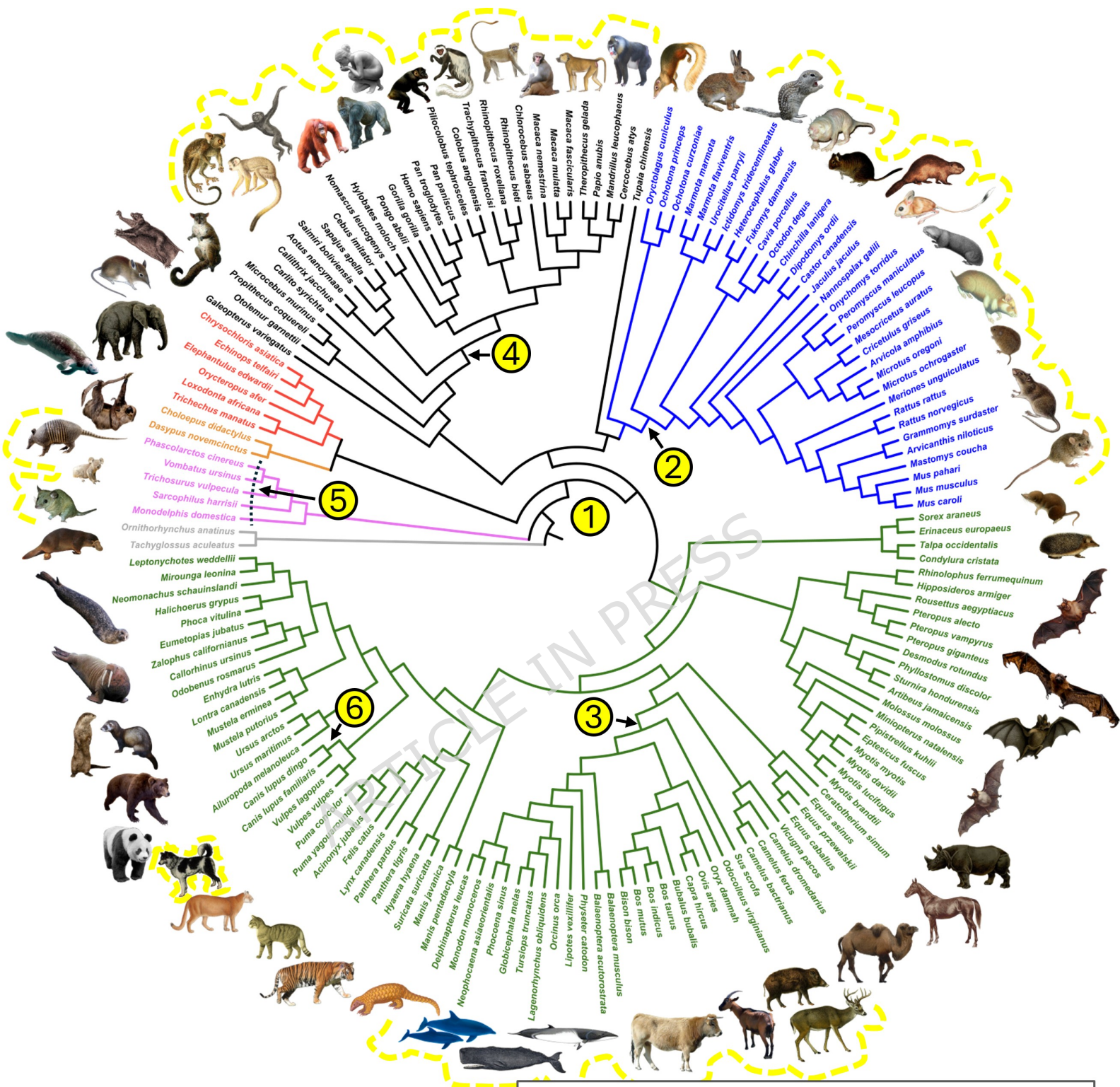


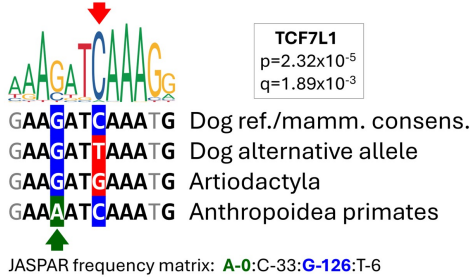
Figure 3. Transcription factor site prediction in highly evolutionarily conserved *ADAMTSL1* intronic element. (A/B) The dog variant and flanking DNA sequence identified by WGS are shown with alignments representing the phylogenetic findings in Figure 2 and the sequence logos for TCF7L1 and PHOX2B transcription factor binding sites. P and q values for the binding prediction were generated using FIMO [60]. (C/D) Time and tissue specific mRNA expression patterns of Tcf7l1 and Phox2b are shown for mouse brain development. Time points are in embryonic or postnatal days. The positions of brain regions are shown in (E), adapted from the Allen Brain Atlas [31].



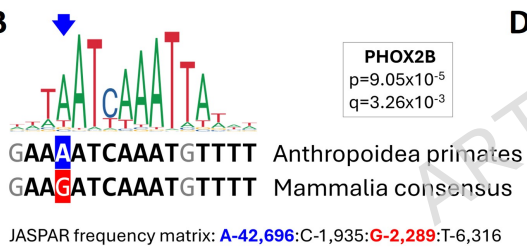
Evolution of *ADAMTSL1* Candidate Intronic Element

1. GAAGATCAAATGTTTTT Mammalian consensus
2. ===== Rodent complete absence
3. GAAGATCAAATGTTTTT Artiodactyla
4. GAAATCAAATGTTTTT Anthropeidea Primates
5. ===== Marsupial common loss
6. GAAGATTAAATGTTTTT Dogs segregate C/T alleles

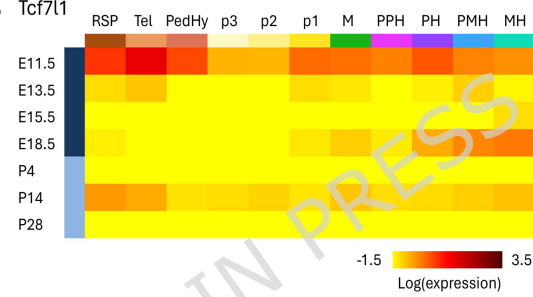
A JASPAR frequency matrix: A-5:**C-159**:**G-0**:**T-0**



B



C Tcf7l1

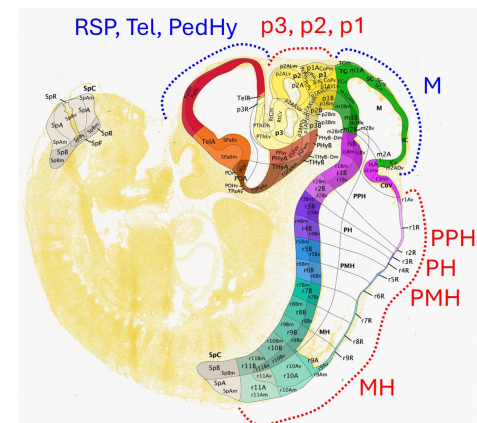


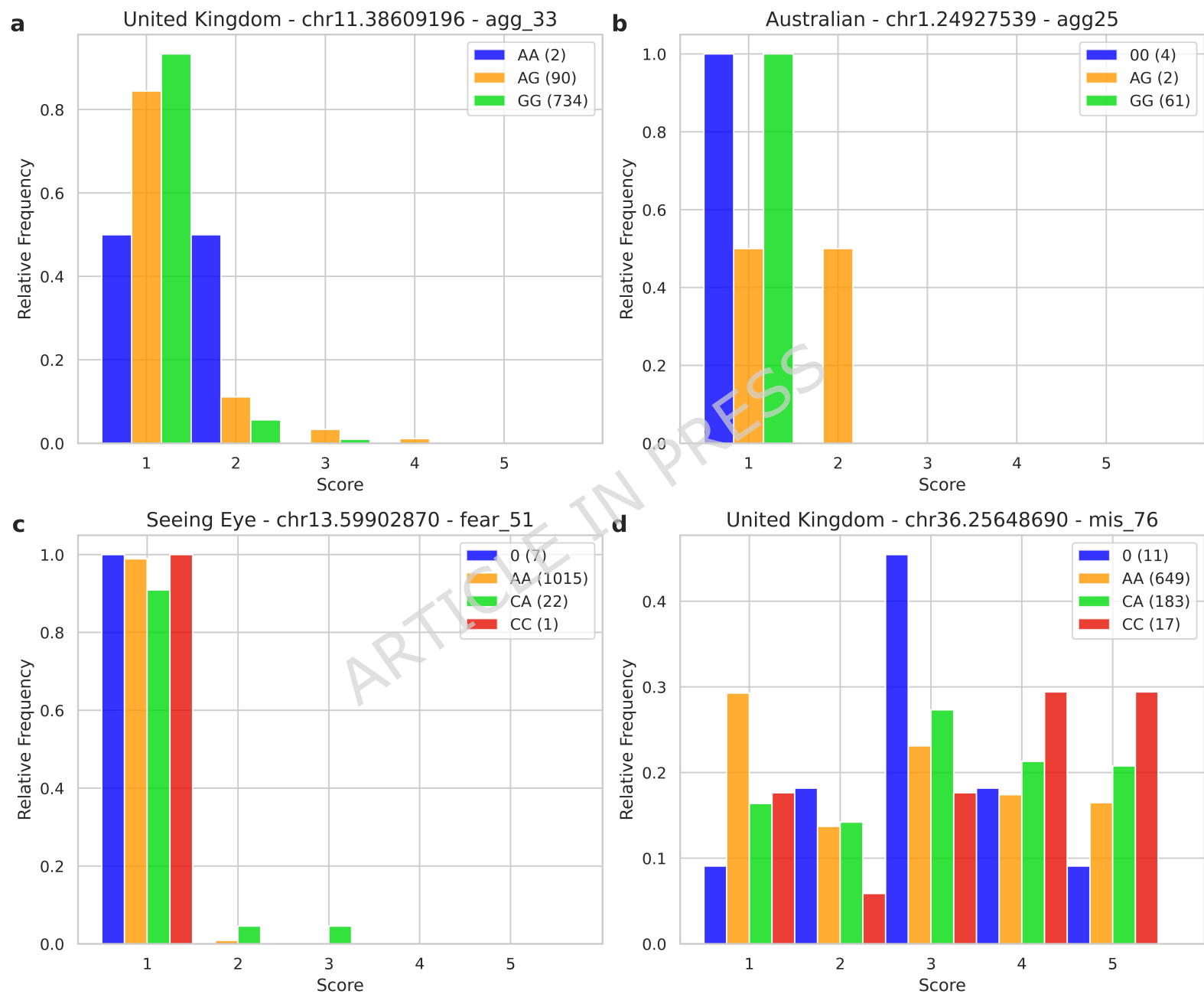
D Phox2b



E

Mouse E11.5 developing brain





Cohort	Source	#SNPs	#PheWAS SNPs	#Questions
Australian (AUS)	D. van Rooy, C.M. Wade 2019	220,853	13	98
United Kingdom (UK)	J. Ilska, et al. 2017	108,829	9	100
Seeing Eye (SE)	Guide Dog Training	213,644	8	97

Canid	N	Dataset ¹	Alleles	MAF
Bat-eared fox	1	UCSC Cactus	C	NA
Raccoon dog	1	UCSC Cactus	C	NA
Arctic fox	1	UCSC Cactus	C	NA
Red fox	1	UCSC Multiz	C	NA
African wild dog	1	UCSC Cactus	C	NA
Golden Jackal	1	DogGD Browser	C	NA
Wolf ²	52	DogGD dataset 3	C/T	T:0.067
Dingo	1	DogGD Browser	C	NA
Ancient dog ³	27	DogGD dataset 4	C/T	T:0.35
Native dog ⁴	116	DogGD dataset 2	C/T	T:0.22
Dog ⁵	670	DogGD dataset 3	C/T	T:0.29

ID	Sex	Success	Elimination Reason	Breed
C608	F	Y	NA	GSP
E143	M	Y	NA	GSP
E179	M	NA	NA	GSP
E312	F	NA	NA	GSP
E419	F	N	Behavioral	GSP
E425	M	Y	NA	GSP
F228	F	NA	NA	GSP
E321	M	Y	NA	LR
E481	F	N	Behavioral	LR
F117	F	Y	NA	LR
F162	F	N	Behavioral	LR
F170	M	Y	NA	LR
F177	M	Y	NA	LR
F243	F	N	Environmental	LR

Cohort	P value								*	C-BARQ ID	C-BARQ Item
AUS		2.1E-04				5.7E-07			Aggression	agg_13	Aggression when object taken away by household member
AUS	5.0E-05	1.5E-10								agg_25	Aggression to household member staring at them
AUS						3.8E-05				agg_31	Aggression when stolen food is taken away by household member
UK				2.2E-04						agg_33	Aggression when approached by a household dog at a favorite resting place
SE					3.1E-04				Fear	fear_51	Fear when feet are towed by a member of the household
AUS			2.1E-04						Separation related	sep_56	Frequency of restlessness or pacing following separation-related behavior
UK						1.3E-05			Excitability	exc_63	Frequency of persistent barking when alarmed or excited
UK-SE						1.9E-04			Excitability	exc_67	Tendency to become excitable when visitors arrive at your home
UK-SE						3.3E-04			Miscellaneous	mis_74	Frequency of chasing cats given the opportunity
UK							1.5E-04	3.4E-04	Miscellaneous	mis_76	Frequency of chasing small animals
UK						3.3E-06			Miscellaneous	mis_83	Frequency of stealing food
AUS			7.8E-06						Miscellaneous	mis_87	Frequency of urinating when approached, petted, handled, or picked up
	chr1.22989459 (MC2R)	chr1.24927539 (MC2R)	chr6.76632282 (LRRC7)	chr11.38609196 (ADAMTSL1)	chr13.59902870 (Many)	chr19.21040815 (HS6ST1)	chr36.25252101 (PDE1A)	chr36.25648690 (PDE1A)	*C-BARQ Factors - Aggression - Fear - Separation related - Excitability - Miscellaneous		