






ARTICLE

Open Access

# A novel genomic region on chromosome 11 associated with fearfulness in dogs

R. Sarviaho<sup>1,2,3</sup>, O. Hakosalo<sup>1,2,3</sup>, K. Tiira<sup>1,2,3,4</sup> , S. Sulkama<sup>1,2,3</sup>, J. E. Niskanen<sup>1,2,3</sup> , M. K. Hytönen<sup>1,2,3</sup> , M. J. Sillanpää<sup>5</sup>  and H. Lohi<sup>1,2,3</sup> 

## Abstract

The complex phenotypic and genetic nature of anxieties hampers progress in unravelling their molecular etiologies. Dogs present extensive natural variation in fear and anxiety behaviour and could advance the understanding of the molecular background of behaviour due to their unique breeding history and genetic architecture. As dogs live as part of human families under constant care and monitoring, information from their behaviour and experiences are easily available. Here we have studied the genetic background of fearfulness in the Great Dane breed. Dogs were scored and categorised into cases and controls based on the results of the validated owner-completed behavioural survey. A genome-wide association study in a cohort of 124 dogs with and without socialisation as a covariate revealed a genome-wide significant locus on chromosome 11. Whole exome sequencing and whole genome sequencing revealed extensive regions of opposite homozygosity in the same locus on chromosome 11 between the cases and controls with interesting neuronal candidate genes such as *MAPK9/JNK2*, a known hippocampal regulator of anxiety. Further characterisation of the identified locus will pave the way for molecular understanding of fear in dogs and may provide a natural animal model for human anxieties.

## Introduction

Fear is an emotional state that is caused by any stimulus that is interpreted to predict or cause danger. Fear produces a behavioural reaction that aims for the survival of the animal<sup>1,2</sup>. However, increased and overscale fearfulness and anxiety are considered psychiatric conditions in humans and behavioural problems in dogs. In humans, anxiety disorders and specific phobias have estimated lifetime prevalences of over 15%<sup>1</sup> and 7.4%, respectively<sup>3</sup>. While these disorders are heritable, they are genetically complex, emphasising the need of physiologically relevant animal models to facilitate the discovery of molecular etiologies<sup>4–8</sup>.

Domestic dogs have emerged as important large animal models of disease, morphology, and behaviour due to their naturally occurring phenotypes, unique genetic architecture and physiological similarity, and shared living environment with humans<sup>9</sup>. Various anxieties are commonly reported in dogs, and can dramatically impair their welfare. For example, dogs showing fear towards strangers have significantly shorter lifespans when compared to non-fearful dogs<sup>10</sup>. Dogs' fearful reactions vary from withdrawing, staying close to the owner, or having a tail in a low position or placed between legs to behaviours such as barking, growling, or even biting, thus creating a serious health concern also for humans<sup>11</sup>. In dogs, fear can be categorised into social and non-social fearfulness. The social category includes fear of unfamiliar people and dogs while the non-social category includes fear of novel situations<sup>12</sup>. Our previous study suggested that fear of strange people is correlated both with fear of strange dogs and with fear of novel situations<sup>13</sup>. Nearly 60% of dogs showing fear of strangers were fearful in novel situations

Correspondence: H. Lohi ([hannes.lohi@helsinki.fi](mailto:hannes.lohi@helsinki.fi))

<sup>1</sup>Department of Veterinary Biosciences, University of Helsinki, 00014 Helsinki, Finland

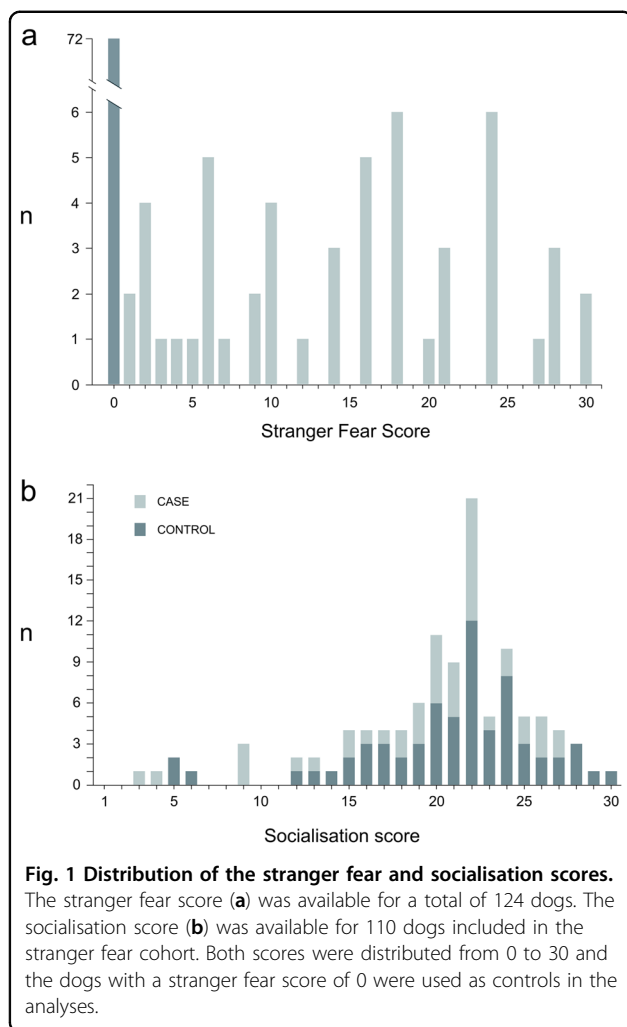
<sup>2</sup>Department of Medical and Clinical Genetics, University of Helsinki, 00014 Helsinki, Finland

Full list of author information is available at the end of the article

© The Author(s) 2020



**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.



and the two traits were considered together as aspects of general fearfulness in our previous genome-wide association study (GWAS)<sup>13,14</sup>.

The challenge of reliably measuring behaviour and its heritable component together with the complex genetic structure underlying behaviour<sup>15</sup> have slowed down the process of discovering loci or genetic variants for canine behaviour. We recently reported novel loci for canine fearfulness and noise sensitivity<sup>14</sup>. In addition, we have previously reported changes such as increased plasma glutamine levels in fearful dogs<sup>16,17</sup>. Candidate loci and genes have also been reported for example for canine fear and aggression<sup>18</sup>, social behaviour<sup>19,20</sup>, and obsessive-compulsive disorder<sup>21,22</sup>.

In addition to the genetic risk factors, environment has a prominent effect on fearful behaviour<sup>23,24</sup>. Pet dogs live their whole lives in human families and are closely monitored by their owners, enabling valid assessment of potentially important environmental factors. One factor that has a known, important role on a dog's behaviour is

the level of socialisation that the dog experiences during its puppyhood<sup>23,25,26</sup>. A limited amount of experiences from the surrounding world during this period is known to increase fearful behaviour in dogs<sup>13,26</sup>. Identification and incorporation of the environmental contributors such as socialisation into genetic analyses is important and may offer additional consolidation for a genetic signal<sup>24</sup>.

In this study, we have addressed the genetic background of fear of strangers in the Great Dane (GD) breed. We collected altogether 124 GD dogs with a phenotypic score using our validated behavioural survey. A GWAS and whole exome and genome sequencing revealed a novel locus on chromosome 11 that harbours interesting candidate genes.

## Materials and methods

### Study cohorts and phenotypic categorisation

A previously published and validated owner-completed behavioural survey focusing on fearfulness was used to recruit Finnish Great Dane dogs to the study<sup>13,27</sup>. The breed was chosen due to a large number of owner- and breeder-reported dogs that were fearful when meeting unfamiliar people.

Altogether 124 GD dogs were phenotyped (Fig. 1) using our validated behavioural questionnaire<sup>27</sup>. The dogs were categorised to cases and controls based on a categorical quantitative "stranger fear score" (SFS)<sup>13</sup>. The SFS describes the frequency and intensity of a fearful reaction towards unfamiliar people. The dogs with an SFS value of 0 showed no fearful reactions when meeting unfamiliar people and are referred to as controls. Additionally, it was required that in the questionnaire, the owners of the dogs included in the control cohort had chosen one of the following options when describing their dogs' behaviour when meeting a novel person: "greet strangers spontaneously (if given a chance)", "jumps, licks, is eager" or "sniffs, wags tail and is calm". Any dog with an SFS value above 0 showed at least a mild fearful reaction and is thus referred to as a case. The SFS was calculated as follows: (sum of fearful behavioural reactions when meeting unfamiliar people, where withdrawal was multiplied by 5) × (frequency of fear reactions when meeting unfamiliar people) (Supplementary Table 1). The range of possible SFS values was from 0 to 48 with intervals of 1, with the dogs that showed the most severe and frequent fear reactions gaining the highest stranger fear (SF) scores. Avoidance was emphasised in the scoring by a factor of 5 because it is a clear, common fear reaction that is easily recognised by owners<sup>14,28</sup>. The case or control status of a sub-cohort (15 cases and 13 controls) of the SF-cohort was validated by a behavioural test<sup>27</sup>. Canine personality has been shown to be stable from 1 year onwards<sup>29</sup>, while it has also been suggested that fearfulness as a personality trait could be reliably observed in dogs already at the age

of 6 months<sup>30</sup>. Thus, all control dogs were required to be over 1 year old while in the case-cohort, dogs slightly under 1 year old could be included if the dog had been through the behavioural test to confirm its status.

Based on the questionnaire answers, a socialisation score (SS) describing the extent and frequency of socialisation events during a dog's puppyhood was also calculated. The score was used as a covariate in the genetic analyses, as socialisation is known to have a large effect on fearfulness in dogs<sup>23,25</sup>. The SS was calculated as a sum of socialisation event frequencies during a developmental period when the dog was between 7 weeks and 3 months of age (Supplementary Table 2). SS was previously shown to have excellent repeatability when evaluating the test–retest repeatability of the behavioural questionnaire<sup>23</sup>. The range of possible SS values was from 0 to 35 with intervals of 1, with the dogs that had been exposed to the highest amount of socialisation gaining the highest SSs. The SS was not available for 14 dogs with an SFS. For analysis purposes, the SS was divided into four ordered sub-categories as follows: category 1: scores from 0 to 10; category 2: scores from 10 to 20; category 3: scores from 21 to 24; and category 4: scores from 25 to 35.

According to our previous study, 57.1% of dogs that showed fearfulness when meeting a stranger were also fearful in novel situations<sup>13</sup> and the two traits were studied together as general fearfulness in Sarviaho et al.<sup>14</sup>. In this study, we only focused on the fear of strangers. The score describing the fear in novel situations was not available for all 124 dogs included in this study (for details about the score, see Sarviaho et al.<sup>14</sup>). Out of the 72 controls in this study, a score that describes the level of fear they show in novel situations was available for 35 dogs. Among the 52 cases, 27 dogs had an available fear of novel situations score (Supplementary Table 3). Out of the 35 controls, three dogs showed fear in novel situations while out of the 27 cases, the equivalent number was 18. The fear in novel situations score was not used as an inclusion criterion in this study.

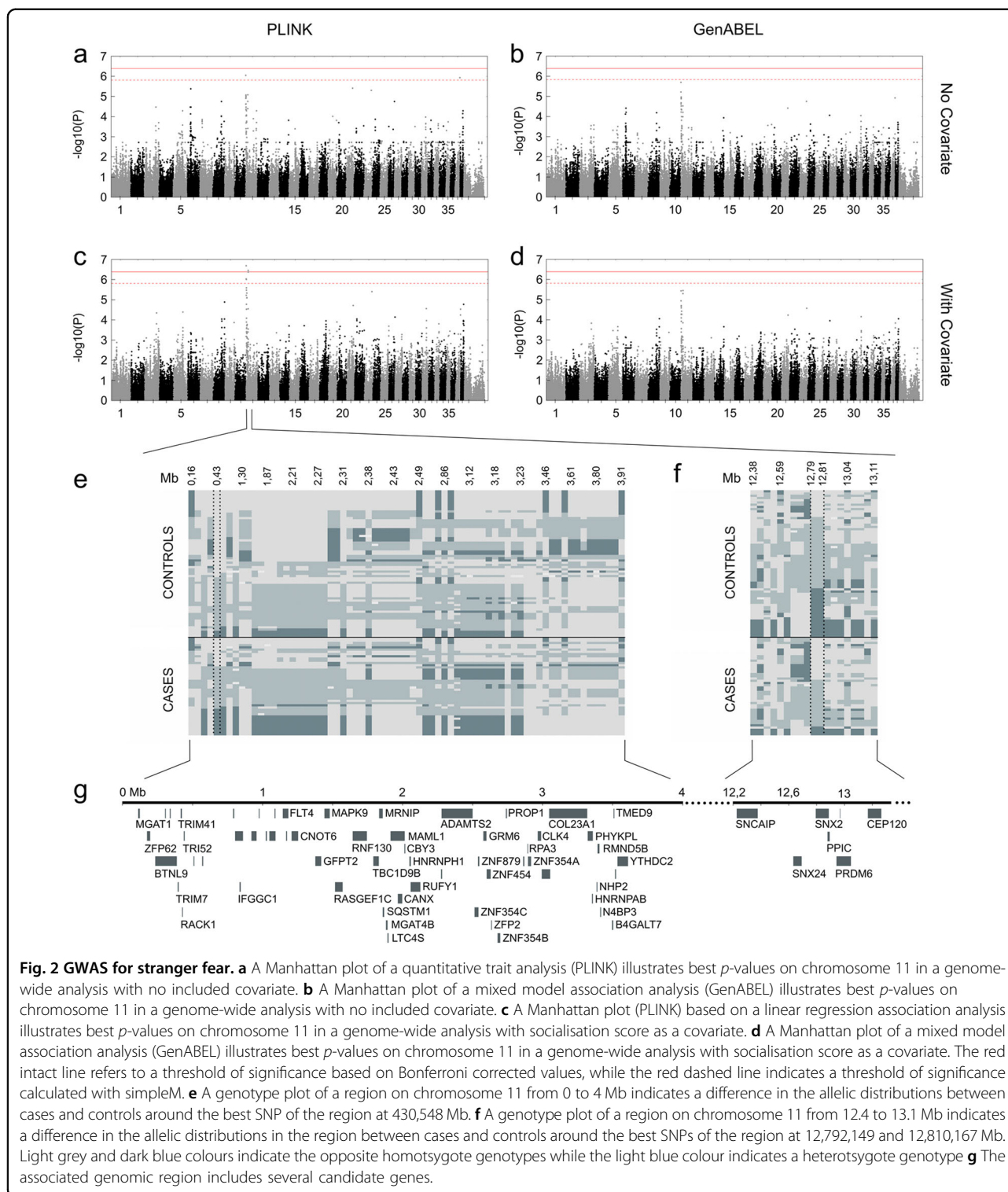
### Samples

Ethylenediamine tetraacetic acid blood samples were collected from 124 privately owned GD dogs for DNA. The blood samples were stored at  $-20^{\circ}\text{C}$  and genomic DNA was extracted by using a semi-automated Chemagen extraction robot (PerkinElmer Chemagen Technologie). A Qubit 3.0 Fluorometer (Thermo Fisher Scientific) and a NanoDrop ND-1000 UV/Vis Spectrophotometer were used to determine DNA concentration. Sample collection was ethically approved by the Animal Ethics Committee of State Provincial Office of Southern Finland, Hämeenlinna, Finland (ESAVI/6054/04.10.03/2012).

### Genome-wide association study

Altogether 124 GDs were genotyped using Illumina's Canine HD (173k) single nucleotide polymorphism (SNP) arrays (Fig. 1). The data was analysed by treating SF as a quantitative trait based on the categorical SF score (Fig. 1). The initial analysis cohort included 52 cases (29 females and 23 males) and 72 controls (44 females and 28 males), where cases had an SFS between 1 and 30 and all controls had a value of 0. In the analysis cohort in which the SS was included as an environmental covariate for SFS ("socialisation cohort"), the number of cases was 44 (23 females and 21 males) and of controls 66 (40 females and 26 males). The genotype data was filtered with a SNP call rate of  $>90\%$ , individual call rate of  $>90\%$ , minor allele frequency of  $>1\%$ , and by using a Hardy–Weinberg equilibrium test of  $P \geq 1 \times 10^{-8}$ . After the quality control, no individual dogs were removed and 122,529 SNPs remained for analysis. Genotyping data was analysed using a quantitative trait association analysis and linear regression association analysis in PLINK<sup>31</sup>, and a FASTA generalised linear mixed model approach implemented in the R package GenABEL<sup>32</sup>. The analyses were performed using a quantitative trait approach (Fig. 1). Multi-dimensional scaling and quantile–quantile plots were drawn to evaluate the existence of potential confounding signals due to cryptic relatedness and population stratification in the study populations (Supplementary Fig. 1).

In order to adjust the data for relatedness and population stratification, we used genomic control approach with PLINK<sup>33</sup> and a genomic control approach or a polygenic effect term with marker-estimated relationship matrix in GenABEL<sup>34</sup>. To account for multiple testing, Bonferroni correction was implemented to adjust the genome-wide significance threshold. As the Bonferroni method assumes independence between tests and is thus overly conservative for GWAS, the effective number of independent tests was estimated with simpleM<sup>35</sup> and the threshold of significance calculated as described by Mikola et al.<sup>36</sup>. The estimated effective number of independent tests was 34,271 for the whole cohort and 32,773 for the covariate cohort in the PLINK analyses. In the GenABEL analyses, the numbers were 34,097 and 32,678, respectively. The significance level after adjusting with Bonferroni correction was  $4.08 \times 10^{-7}$  (Fig. 2a–d). We also adjusted the significance level with the estimated number of independent tests, which resulted in a threshold of  $1.46 \times 10^{-6}$  for the PLINK analysis (Fig. 2a) and  $1.47 \times 10^{-6}$  for GenABEL (Fig. 2b) in the whole cohort analyses. For the covariate cohort, the significance level was  $1.53 \times 10^{-6}$  for both analyses (Fig. 2c, d). The linkage structure within the associated regions was determined by using PLINK's clump command. A window size of 4000 kb was used while other settings were used as default.



### Exome and whole genome sequencing

Dog exome libraries for eight GD cases and eight GD controls were generated from standard indexed Illumina libraries using a custom Roche/Nimblegen solution-based capture library (120705\_CF3\_Uppsala\_Broad\_EZ\_HX1) as

previously described by Elvers et al.<sup>37</sup>. The 16 dogs were chosen from among the GWAS cohort based on the best SNP from GWAS (chr11:430548G>A; BICF2P368048) being from opposite haplotypes, except for one case that was a heterozygote. The dogs represented medium to severe



SF phenotypes, with SF scores between 16 and 28. In addition, four cases and four controls were tested with a behavioural test as described in ref. <sup>27</sup>. The exome sequencing was performed in Science for Life Laboratory (Sweden). Out of the 16 exome sequenced dogs, four GD cases and two GD controls were whole genome sequenced using the Illumina HiSeqX platform at Novogene (Novogen (HK) Company Limited, China) and the data was further processed as previously described by Hytönen and Lohi<sup>38</sup>. For filtering the sequencing data, Genotype Query Tools<sup>39,40</sup> and our in-house analysis pipeline were used<sup>41</sup>. For the exome sequences, the sequencing coverage was on average 26–104× per sample and on average 105,843 variants per sample were revealed. For the whole genome sequences, the numbers were 26–33× and 6,574,191, respectively. While filtering the exome sequences, cases were filtered against the controls under a dominant model (which assumes that a variant shared by the cases is present either as heterozygous or homozygous), allowing a maximum of two cases to be wildtype homozygotes for each variant. Additionally, the controls were filtered against the cases, allowing two controls to be wildtype homozygotes for each variant. While filtering the whole genome sequences, the four cases were filtered against the two controls under a dominant model, allowing each variant to be missing from a maximum of one case. To predict the functional effects of the variants, Ensembl's Variant Effect Predictor was used<sup>42</sup>. Allele frequencies were calculated for the variants by utilising 1084 exome and whole genome sequences available through our database and collaborators in the Dog Biomedical Variant Database Consortium. Conservation scores (CS) were generated with Clustal Omega<sup>43</sup>. Mobile element insertions (MEI) and structural variants (SV), including insertions, deletions, duplications, and intra-chromosomal re-arrangements were identified as described in Dillard et al.<sup>44</sup> and by filtering the whole genome sequencing (WGS) cases against the controls under a dominant model.

Throughout the study, the CanFam 3.1 genome build was used as a reference. The short read data in BAM format for 16 WES and six WGS Great Dane samples is submitted to NCBI SRA with the BioProject ID: [PRJNA630029](https://www.ncbi.nlm.nih.gov/bioproject/PRJNA630029).

### Sanger sequencing

Samples from altogether 40 GD dogs, including 15 cases (with SF scores ranging from 3 to 25) and 25 controls were used in PCR and Sanger sequencing in order to study three candidate variants (chr11:1472008G>A (SNP1292825), chr11:3390145GGGGAA>G (rs851847854) and chr11:3416964G>T (rs851521203)) identified in WES and WGS. To amplify the variant region, forward and reverse primers were designed (see Supplementary Table 4) and used together with AmpliTaq Gold 360 Mastermix (Applied Biosystems, Life Technologies) or Biotools' DNA

Polymerase. After treatment with exonuclease I and shrimp alkaline phosphatase, the sequencing reactions were performed using an ABI 3730 capillary sequencer (Applied Biosystems, Life Technologies) at the Institute for Molecular Medicine Finland (FIMM). The sequence data were analysed using Sequencher 5.1 (GeneCodes).

### Results

Altogether 124 GD were phenotyped for fearfulness towards strange people. To evaluate the intensity of the fearful reactions, a quantitative categorical "stranger fear" score (SFS) was established. Examples of a fearful and a non-fearful dog in a behavioural testing situation can be seen in Supplementary Video 1. Additionally, a quantitative categorical score evaluating the amount of socialisation (SS) during the dogs' puppyhood was calculated and used as a covariate in the analyses. Due to the SS not being available for all 124 dogs, the covariate cohort included 110 dogs. The SFS values of cases varied between 1 and 30 while all control dogs had a value of 0. In both cohorts, the median value of SFS in cases was 16 and an average SFS value was 14.6 (Fig. 1a). The SS values varied between 3 and 30 with a median value of 21.5 and an average of 20.2 (Fig. 1b). The distribution of the SS, an environmental factor affecting SF, was rather similar between cases and controls, suggesting a genetic component for SF (Fig. 1b). For analysis purposes, the SS was further divided into four sub-categories, varying between 1 and 4, with a median of 3 and an average of 2.7. The genomic inflation factors ( $\lambda$ ) were 1.097 for the whole cohort and 1.072 for the covariate cohort, suggesting moderate levels of population stratification in both cohorts (Supplementary Figs. 1 and 2).

### SF maps to chromosome 11

In order to map loci for fear of strangers, altogether 124 GD dogs were genotyped. The initial cohort included 52 cases and 72 controls, while the cohort with the SS as a covariate included 44 cases and 66 controls. The PLINK and GenABEL analyses in both cohorts revealed an association on canine chromosome 11 (Fig. 2). When including socialisation as a covariate, the PLINK analysis assuming a linear model suggested a significant association both with a Bonferroni-based threshold of significance and when the effective number of independent tests was considered (Fig. 2a), while the PLINK analysis without a covariate suggested a significant signal only with the threshold of significance calculated based on simpleM. The GenABEL analyses yielded a signal which did not reach genome-wide significance (Fig. 2b). This may be a result of a known general drawback in mixed-model approaches, causing them to lose power (by over-correcting the structure) or possibly leading to false

**Table 1 Top 10 GWAS hits in association analyses without a covariate.**

| PLINK (quantitative trait association analysis) |     |            |                      |
|---|-----|------------|----------------------|
| SNP name  | CHR | Position   | p-Value (adjusted)   |
| BICF2P368048                                    | 11  | 430,548    | 8.846e-07            |
| BICF2S2294860                                   | 23  | 45,510,382 | 4.949e-06            |
| BICF2P757289                                    | 11  | 404,385    | 2.385e-05            |
| BICF2G630134982                                 | 37  | 29,628,498 | 1.164e-06            |
| BICF2S2363579                                   | 21  | 41,417,917 | 3.89e-06             |
| BICF2P1385199                                   | 6   | 20,120,837 | 4.158e-06            |
| BICF2P1130863                                   | 11  | 12,729,090 | 8.292e-06            |
| BICF2S23135525                                  | 11  | 1,177,541  | 8.494e-06            |
| BICF2P54557                                     | 11  | 1,871,910  | 9.305e-06            |
| BICF2P734487                                    | 11  | 1,848,161  | 1.047e-05            |
| GenABEL   |     |            |                      |
| SNP name  | CHR | Position   | p-Value              |
| BICF2P368048                                    | 11  | 430,548    | 1.99533519991599e-06 |
| BICF2P757289                                    | 11  | 404,385    | 6.07779244096143e-06 |
| BICF2P54557                                     | 11  | 1,871,910  | 1.07999611246649e-05 |
| BICF2P734487                                    | 11  | 1,848,161  | 1.17827498872651e-05 |
| BICF2G630134982                                 | 37  | 29,628,498 | 1.19426581999159e-05 |
| TIGRP2P144053                                   | 11  | 1,963,139  | 1.30896323580011e-05 |
| BICF2S2294860                                   | 23  | 45,510,382 | 1.76695001064685e-05 |
| BICF2S23135525                                  | 11  | 1,177,541  | 2.08872205163928e-05 |
| BICF2P331249                                    | 11  | 861,321    | 2.38961915938629e-05 |
| BICF2P1130863                                   | 11  | 12,729,090 | 2.98857919837111e-05 |

**Table 2 Top 10 GWAS hits in association analyses with a covariate.**

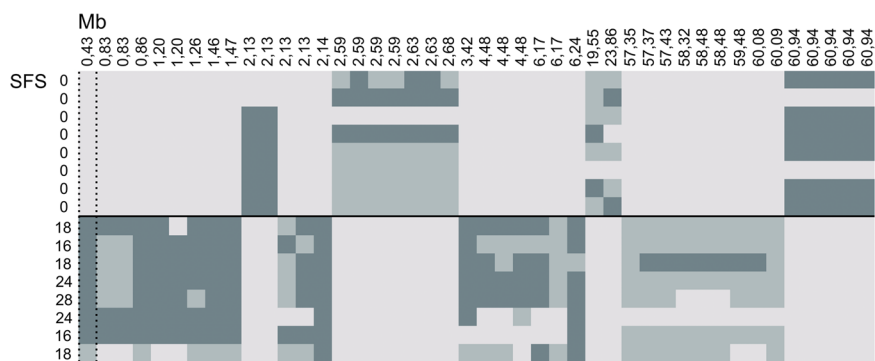
| PLINK (linear regression association analysis) |     |            |                      |
|--|-----|------------|----------------------|
| SNP name                                       | CHR | Position   | p-Value (unadjusted) |
| BICF2P368048                                   | 11  | 430,548    | 2.058e-07            |
| TIGRP2P147692                                  | 11  | 12,792,149 | 3.531e-07            |
| BICF2P365517                                   | 11  | 12,810,167 | 3.531e-07            |
| BICF2P1130863                                  | 11  | 12,729,090 | 4.251e-07            |
| BICF2P757289                                   | 11  | 404,385    | 9.224e-07            |
| BICF2S23135525                                 | 11  | 1,177,541  | 9.903e-07            |
| BICF2P54557                                    | 11  | 1,871,910  | 2.561e-06            |
| TIGRP2P144053                                  | 11  | 1,963,139  | 3.35e-06             |
| BICF2S2294860                                  | 23  | 45,510,382 | 3.937e-06            |
| BICF2S23445920                                 | 11  | 4,257,713  | 4.679e-06            |
| GenABEL  |     |            |                      |
| SNP name                                       | CHR | Position   | p-Value              |
| TIGRP2P147692                                  | 11  | 12,792,149 | 3.54156004567416e-06 |
| BICF2P365517                                   | 11  | 12,810,167 | 3.54156004567416e-06 |
| BICF2P368048                                   | 11  | 430,548    | 3.65165246465062e-06 |
| BICF2P1130863                                  | 11  | 12,729,090 | 5.04242375996489e-06 |
| BICF2P757289                                   | 11  | 404,385    | 1.14950271348944e-05 |
| BICF2P54557                                    | 11  | 1,871,910  | 1.98766237111194e-05 |
| BICF2S23135525                                 | 11  | 1,177,541  | 2.11438133432113e-05 |
| TIGRP2P144053                                  | 11  | 1,963,139  | 2.56387893916905e-05 |
| BICF2S2294860                                  | 23  | 45,510,382 | 2.67885122463861e-05 |
| BICF2S23445920                                 | 11  | 4,257,713  | 3.75198642012459e-05 |

findings when candidate SNPs are included in the calculation of genomic relationship matrix (see, for example, Würschum and Kraft<sup>45</sup>).

In three of the four association analyses, the best-associated SNP was BICF2P368048 located at 430,548 bp on chromosome 11. In the GenABEL analysis including a covariate, the best-associated two SNPs spanned a region on chromosome 11 from 12,792,149 to 12,810,167 bp (Tables 1 and 2). Based on the best-associated SNPs in both analyses, two regions on chromosome 11 are associated with SF: one spanning from 0.4 Mb and one around 12.8 Mb (Tables 1 and 2).

Genotype plots of the best-associated regions revealed differences between the cases and controls in the associated regions. Although no intact haplotypes segregating between cases and controls were seen,

there seemed to be signs of haplotype structure, with many studied markers in linkage. In the region at around 12.8 Mb, the two associated SNPs were in complete linkage. To study linkage disequilibrium within both loci, PLINK's clumping algorithm was used. For the locus at the beginning of chromosome 11, two regions of linkage were identified in the analysis without a covariate, the longer spanning from around 0.4 to 3.25 Mb (Supplementary Fig. 3a), while including the covariate resulted in only one region, spanning from roughly 0.4 to 2.4 Mb (Supplementary Fig. 3c). For the locus around 12.8 Mb, two regions were revealed in both analyses (Supplementary Fig. 3b, d). Both associated loci include several genes and the locus spanning from 0.4 Mb is largely syntenic to human 5q35.3 (Fig. 2e).



**Fig. 3 Whole exome variants between cases and controls on chromosome 11.** Altogether 44 variants were left on chromosome 11 when filtering eight cases against eight controls and the controls against the cases under a dominant model and allowing a maximum of either two cases or two controls to have a deviant genotype. Light grey and dark blue colours indicate the opposite homozygote genotypes while the light blue colour indicates a heterozygote genotype. Extensive regions of opposite homozygosity are observed in the region.

### Exome and whole genome sequencing identify extensive regions of homozygosity on chromosome 11

Eight GD case sequences were filtered under a dominant model against eight GD control exome sequences. The filtering resulted in 46 variants, of which 28 laid on chromosome 11, revealing extensive regions of opposite homozygosity (Fig. 3 and Supplementary Table 5). Out of the 28 variants that were left on chromosome 11 after filtering, 13 were located between 0 and 3.5 Mb, while no variants were left around the two associated SNPs around 12 Mb.

To further study the variants on chromosome 11, four case and two control GDs were whole genome sequenced (Supplementary Table 6). After filtering with a dominant model, 73,197 variants were left, of which 4021 laid on chromosome 11. Extensive homozygosity was seen in the GWAS associated regions. Between 0 and 3.5 Mb, 478 variants segregating between cases and controls were seen, including four predicted exonic variants. Altogether 45 variants laid between 12 and 13 Mb, including two predicted exonic variants. Altogether 23 variants within the areas of interest on chromosome 11 had an allele frequency of less than 0.05, while none of these 23 variants were predicted exonic. A total of 76 variants on chromosome 11 had a CS available. Four out of the 12 variants had a CS of over 500, suggesting high conservation, and were within the regions of interest (chr11: 430548G>A (SNP1292511; rs9103113), chr11:1472008G>A (SNP 1292825), chr11:3390145GGGGAA>G (rs851847854) and chr11:3416964G>T (rs851521203); Supplementary Table 6). Additionally, chr11:1472008G>A and chr11:3416964G>T were predicted exonic nonsynonymous variants within interesting neuronal candidate genes *MAPK9/JNK2* and *N4BP3*. While variant chr11:430548G>A was the best SNP in GWAS analyses and was thus not followed on further, the other three

variants were studied in a larger population of GDs included in GWAS. Furthermore, two ambiguous MEIs and six SVs were identified within the regions of interest (see Supplementary Table 7). One of the MEIs flagged as being ambiguous was predicted to lie within the exonic region of the immunity-related GTPase family M protein 1 gene *IRGM1*. However, the variant was not studied further due to the role of *IRGM1* in immune response. All other MEI and SVs were either intergenic or intronic and were therefore not considered for further validation.

### Validation of candidate variants in *MAPK9*, *N4BP3*, and *NHP2*

Three candidate variants were followed up by Sanger sequencing in a population 15 case and 25 control GDs. Sequencing the three variants, including two predicted exonic (in genes *MAPK9/JNK2* and *N4BP3*) and one predicted intronic variant (in the *NHP2* gene), did not reveal significant differences between cases and controls. The *p*-values were as follows: 0.79 for chr11:1472008G>A (*MAPK9/JNK2*), 0.09 for chr11:3390145GGGGAA>G (*NHP2*) and 0.95 for chr11:3416964G>T (*N4BP3*). The *p*-value was close to significance for chr11:3390145GGGGAA>G, but since the variant was not seen as homozygote in any of the tested case dogs, the difference between cases and controls did not follow the pattern seen in WGS, and thus the variant was not studied further. The results suggest that the studied three variants are likely not causal, which is in line with the fact that all three are previously known SNPs (although the allele frequencies for chr11:3390145GGGGAA>G and chr11:3416964G>T were rather low at 0.06 and 0.05; see Supplementary Table 6).

### Discussion

This study provides new insights into the genetics of fearfulness in dogs. We compared GD dogs that react

fearfully when meeting a stranger with non-fearful GDs and mapped a novel susceptibility locus on chromosome 11 with an extensive level of opposite homozygosity between the case and control groups. When socialisation was included as a covariate, the association signal remained on chromosome 11 while other background signals were reduced, supporting the finding. The identified gene-rich region includes potential neuronal candidate genes, although we failed to identify robust candidate causative variants by exome or whole genome sequencing.

The GWAS, exome and WGS data suggested an SF-related locus at the beginning of chromosome 11, which is largely syntenic with human 5q35.3. The locus is rich with genes, although for many of them the available functional information is limited. The most significant SNP of the locus hits a scaffolding protein called receptor of activated protein C kinase 1, encoded by *RACK1*, alternatively named *GNB2L1*. *RACK1* seems to have a function in regulating neural development<sup>46</sup>, but no clear behaviour-related function for the gene has been reported.

Another candidate gene, a mitogen-activated protein (MAP) kinase gene, *MAPK9*, also known as *JNK2*, lies at 1.4 Mb on chromosome 11. MAP kinases are divided into three families, extracellular signal-regulated (ERK), p38, and c-Jun N-terminal (JNK) kinases, *MAPK9* belonging to the latter family<sup>47</sup>. According to studies, JNKs are involved in, for example, synaptic plasticity and brain development<sup>48,49</sup> and are thought to have a role in, for example, Alzheimer's and Parkinson's diseases<sup>50</sup>. JNKs, along with the other MAPK families have also been linked to behaviour and anxiety. A study by Reinecke et al.<sup>48</sup> showed some indications of changes in behaviour in an anxiety-related testing situation in *Jnk2* knock-out mice while a study by Stefanoska et al.<sup>51</sup> suggested that JNKs are involved in an anxiety-modulating pathway regulated by p38 $\alpha$ . Additionally, Mohammad et al.<sup>52</sup> found that *JNK1* has a role in anxiety-related behaviour. Thus, *MAPK9* provides an interesting candidate gene for fearfulness in GDs. No segregating coding variants were found in *MAPK9* in the whole exome analyses.

At 3.4 Mb lies another candidate gene, a Fezzin family member, NEDD4-binding protein coding gene *N4BP3* that has been suggested to have a neurological function. A study by Schmeisser et al.<sup>53</sup> reported that *N4BP3* seems to have a role in the correct branching of neurites in developing neurons while a knock-out study by Kiem et al.<sup>54</sup> indicated that the protein has a function in early anterior neural development of vertebrates. Additionally, loci harbouring Fezzin family members have been linked to neurodevelopmental disorders such as autism and schizophrenia<sup>54–56</sup>.

The GWAS data suggests also another associated area on chromosome 11, with a two-SNP association around

12.8 Mb (Fig. 2). The two SNPs are in full linkage and no exonic variants were left in the area after filtering (Supplementary Tables 5 and 6). The association may be spurious and needs further replication. The two associated SNPs at ~12.8 Mb on chromosome 11 hit a sorting nexin gene *SNX24*. Sorting nexins are a protein family involved in endosomal sorting and signalling<sup>57</sup>, some of which have been linked to neurological functions<sup>58</sup>.

In our previous study using the same behavioural survey, we mapped loci for noise sensitivity and general fearfulness, which included fear of strangers and fear of novel situations, in the German Shepherd breed<sup>14</sup>. Both associated loci were different than the novel SF locus, which suggests that SF in GDs has a separate genetic background from the general fear in German Shepherds.

In addition to our study, Zapata et al.<sup>18</sup> have also studied the genetic background of stranger-oriented fear. The study identified a different association to ours, likely explained by their cross-breed setup with averaged phenotypes compared to our within-breed setup with individual behavioural scores. Thus, both studies offer a different view to the genetic background of human-directed fear. In another study, Ilska et al.<sup>15</sup> studied traits called "Human and Object Fear" and "Non-owner-directed Aggression", with the phenotypes only partly overlapping our phenotype and likely explaining different association results.

In summary, we discovered a novel fear locus with interesting candidate genes in dogs. Replicating the finding in a larger cohort, possibly narrowing down the region of interest and whole genome sequencing a larger set of cases and controls to identify candidate variants in the region will be prioritised in the future experiments. In this study, we were unable to find a risk variant for fearfulness in GDs, which likely reflects the complex genetic nature and unknown genetic mechanisms underlying the phenotype. Fear of strangers, a form of social phobia in dogs, may provide an important model for human anxieties for comparative studies.

#### Acknowledgements

The authors thank Petra Jaakonsaari, Ranja Eklund, Minna Virta, and Sini Karjalainen for technical assistance, Milla Salonen for editing the Supplementary Video, Lea Mikkola for giving valuable counsel, and all the dog owners for participating. The Dog Biomedical Variant Database Consortium (Gus Aguirre, Catherine André, Danika Bannasch, Doreen Becker, Brian Davis, Cord Drögemüller, Kari Ekenstedt, Katerie Faller, Oliver Forman, Steve Friedenberg, Eva Furrow, Urs Giger, Christophe Hitte, Marjo Hytönen, Vidhya Jagannathan, Tosso Leeb, Hannes Lohi, Cathryn Mellersh, Jim Mickelson, Leonardo Murgiano, Anita Oberbauer, Sheila Schmutz, Jeffrey Schoenebeck, Kim Summers, Frank van Steenbeck, Claire Wade) is acknowledged for providing whole-genome variant data to be utilised in the study. The study was supported by the European Research Council, ERC (Grant No. 260997 DOGPSYCH) (<https://erc.europa.eu/>); Academy of Finland (268019); ERA-NET Neuron Mental Disorders; the Jane and Aatos Erkko Foundation; the Sigrid Juselius Foundation; the Finnish Cultural Foundation; Jenny and Antti Wihuri, Ella and Georg Ehrnrooth Foundation, Alfred Kordelin Foundation, ILS Doctoral



Program of University of Helsinki, the Folkhälsan Research Center, DVM Doctoral Program of University of Helsinki, and University of Oulu.

#### Author details

<sup>1</sup>Department of Veterinary Biosciences, University of Helsinki, 00014 Helsinki, Finland. <sup>2</sup>Department of Medical and Clinical Genetics, University of Helsinki, 00014 Helsinki, Finland. <sup>3</sup>Folkhälsan Research Center, 00290 Helsinki, Finland. <sup>4</sup>Equine and Small Animal Medicine, University of Helsinki, Helsinki, Finland. <sup>5</sup>Department of Mathematical Sciences, Biocenter Oulu and Infotech Oulu, University of Oulu, Oulu, Finland

#### Author contributions

H.L. conceptualised the study and designed the research with the help from R. S., O.H., and K.T. R.S., O.H., K.T., and S.S. performed the experiments. R.S., O.H., J.E. N., M.J.S., M.K.H., and H.L. analysed the data. R.S., O.H., K.T., S.S., M.K.H., M.J.S., and H.L. contributed to materials and methods. R.S. and H.L. wrote the manuscript with the help from the other authors. All authors read the manuscript, contributed to comments, and approved the final version of the manuscript.

#### Conflict of interest

The authors declare that they have no conflict of interest.

#### Publisher's note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

**Supplementary Information** accompanies this paper at (<https://doi.org/10.1038/s41398-020-0849-z>).

Received: 4 October 2019 Revised: 5 May 2020 Accepted: 14 May 2020  
Published online: 28 May 2020

#### References

- Perusini, J. N. & Fanselow, M. S. Neurobehavioral perspectives on the distinction between fear and anxiety. *Learn. Mem.* **22**, 417–425 (2015).
- Carleton, R. N. Fear of the unknown: one fear to rule them all? *J. Anxiety Disord.* **41**, 5–21 (2016).
- Wardenaar, K. J. et al. The cross-national epidemiology of specific phobia in the World Mental Health Surveys. *Psychol. Med.* **47**, 1744–1760 (2017).
- Erhardt, A. & Spooemaker, V. I. Translational approaches to anxiety: focus on genetics, fear extinction and brain imaging. *Curr. Psychiatry Rep.* **15**, 417 (2013).
- Domschke, K. & Deckert, J. Genetics of anxiety disorders—status quo and quo vadis. *Curr. Pharm. Des.* **18**, 5691–5698 (2012).
- Van Houtem, C. M. H. H. et al. A review and meta-analysis of the heritability of specific phobia subtypes and corresponding fears. *J. Anxiety Disord.* **27**, 379–388 (2013).
- Torvik, F. A. et al. Longitudinal associations between social anxiety disorder and avoidant personality disorder: a twin study. *J. Abnorm. Psychol.* **125**, 114–124 (2016).
- Overall, K. L. Natural animal models of human psychiatric conditions: assessment of mechanism and validity. *Prog. Neuropsychopharmacol. Biol. Psychiatry* **24**, 727–776 (2000).
- Lindblad-Toh, K. et al. Genome sequence, comparative analysis and haplotype structure of the domestic dog. *Nature* **438**, 803–819 (2005).
- Dreschel, N. A. The effects of fear and anxiety on health and lifespan in pet dogs. *Appl. Anim. Behav. Sci.* **125**, 157–162 (2010).
- Blackwell, E. J., Bradshaw, J. W. S. & Casey, R. A. Fear responses to noises in domestic dogs: Prevalence, risk factors and co-occurrence with other fear related behaviour. *Appl. Anim. Behav. Sci.* **145**, 15–25 (2013).
- Svartberg, K. & Forkman, B. Personality traits in the domestic dog (*Canis familiaris*). *Appl. Anim. Behav. Sci.* **79**, 133–155 (2002).
- Tiira, K., Sulkama, S. & Lohi, H. Prevalence, comorbidity, and behavioral variation in canine anxiety. *J. Vet. Behav. Clin. Appl. Res.* **16**, 36–44 (2016).
- Sarviaho, R. et al. Two novel genomic regions associated with fearfulness in dogs overlap human neuropsychiatric loci. *Transl. Psychiatry* **9**, 18 (2019).
- Ilkka, J. et al. Genetic characterization of dog personality traits. *Genetics* **206**, 1101–1111 (2017).
- Puurunen, J., Tiira, K., Lehtonen, M., Hanhineva, K. & Lohi, H. Non-targeted metabolite profiling reveals changes in oxidative stress, tryptophan and lipid metabolisms in fearful dogs. *Behav. Brain Funct.* **12**, 7 (2016).
- Puurunen, J. et al. Fearful dogs have increased plasma glutamine and  $\gamma$ -glutamyl glutamine. *Sci. Rep.* **8**, 15976 (2018).
- Zapata, I., Serpell, J. A. & Alvarez, C. E. Genetic mapping of canine fear and aggression. *BMC Genomics* **17**, 572 (2016).
- Persson, M. E., Wright, D., Roth, L. S. V., Batakis, P. & Jensen, P. Genomic regions associated with interspecies communication in dogs contain genes related to human social disorders. *Sci. Rep.* **6**, 33439 (2016).
- vonHoldt, B. M. et al. Structural variants in genes associated with human Williams–Beuren syndrome underlie stereotypical hypersociability in domestic dogs. *Sci. Adv.* **3**, e1700398 (2017).
- Tang, R. et al. Candidate genes and functional noncoding variants identified in a canine model of obsessive-compulsive disorder. *Genome Biol.* **15**, R25 (2014).
- Dodman, N. H. et al. A canine chromosome 7 locus confers compulsive disorder susceptibility. *Mol. Psychiatry* **15**, 8–10 (2010).
- Tiira, K. & Lohi, H. Early life experiences and exercise associate with canine anxieties. *PLoS ONE* **10**, e0141907 (2015).
- Gratten, J., Wray, N. R., Keller, M. C. & Visscher, P. M. Large-scale genomics unveils the genetic architecture of psychiatric disorders. *Nat. Neurosci.* **17**, 782–790 (2014).
- Howell, T., King, T. & Bennett, P. Puppy parties and beyond: the role of early age socialization practices on adult dog behavior. *Vet. Med. Res. Rep.* **6**, 143 (2015).
- Scott, J. P. & Fuller, J. L. Genetics and the social behaviour of the dog. *Am. Biol. Teach.* **29**, 756 (1967).
- Tiira, K. & Lohi, H. Reliability and validity of a questionnaire survey in canine anxiety research. *Appl. Anim. Behav. Sci.* **155**, 82–92 (2014).
- Horwitz, D., Mills, D. S. & British Small Animal Veterinary Association. *BSAVA Manual of Canine and Feline Behavioural Medicine* (British Small Animal Veterinary Association, 2009).
- Fratkin, J. L., Sinn, D. L., Patall, E. A. & Gosling, S. D. Personality consistency in dogs: a meta-analysis. *PLoS ONE* **8**, e54907 (2013).
- Riemer, S., Müller, C., Virányi, Z., Huber, L. & Range, F. Individual and group level trajectories of behavioural development in *Border collies*. *Appl. Anim. Behav. Sci.* **180**, 78–86 (2016).
- Purcell, S. et al. PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am. J. Hum. Genet.* **81**, 559–575 (2007).
- Aulchenko, Y. S., Ripke, S., Isaacs, A. & van Duijn, C. M. GenABEL: an R library for genome-wide association analysis. *Bioinformatics* **23**, 1294–1296 (2007).
- Devlin, B. & Roeder, K. Genomic control for association studies. *Biometrics* **55**, 997–1004 (1999).
- Amin, N., van Duijn, C. M. & Aulchenko, Y. S. A genomic background based method for association analysis in related individuals. *PLoS ONE* **2**, e1274 (2007).
- Gao, X. Multiple testing corrections for imputed SNPs. *Genet. Epidemiol.* **35**, 154–158 (2011).
- Mikkola, L. I. et al. Novel protective and risk loci in hip dysplasia in German Shepherds. *PLoS Genet.* **15**, e1008197 (2019).
- Elvers, I. et al. Exome sequencing of lymphomas from three dog breeds reveals somatic mutation patterns reflecting genetic background. *Genome Res.* **25**, 1634–1645 (2015).
- Hytönen, M. K. & Lohi, H. A frameshift insertion in *SGK3* leads to recessive hairlessness in Scottish Deerhounds: a candidate gene for human alopecia conditions. *Hum. Genet.* **138**, 535–539 (2019).
- Layer, R. M., Kindlon, N., Karczewski, K. J. & Quinlan, A. R. Efficient genotype compression and analysis of large genetic-variation data sets. *Nat. Methods* **13**, 63–65 (2015).
- Arumilli, M., Layer, R. M., Hytönen, M. K. & Lohi, H. webGQT: a shiny server for genotype query tools for model-based variant filtering. *Front. Genet.* **11**, 152 (2020).
- Hytönen, M. K., Arumilli, M., Sarkiala, E., Nieminen, P. & Lohi, H. Canine models of human amelogenesis imperfecta: identification of novel recessive ENAM and ACP4 variants. *Hum. Genet.* **138**, 525–533 (2019).
- McLaren, W. et al. The Ensembl Variant Effect Predictor. *Genome Biol.* **17**, 122 (2016).

43. Sievers, F. et al. Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. *Mol. Syst. Biol.* **7**, 539 (2011).
44. Dillard, K. J. et al. Recessive missense LAMP3 variant associated with defect in lamellar body biogenesis and fatal neonatal interstitial lung disease in dogs. *PLoS Genet.* **16**, e1008651 (2020).
45. Würschum, T. & Kraft, T. Evaluation of multi-locus models for genome-wide association studies: a case study in sugar beet. *Heredity* **114**, 281–290 (2015).
46. Kershner, L. & Welshhans, K. RACK1 regulates neural development. *Neural Regen. Res.* **12**, 1036 (2017).
47. Morrison, D. K. MAP kinase pathways. *Cold Spring Harb. Perspect. Biol.* **4**, a011254 (2012).
48. Reinecke, K., Herdegen, T., Eminel, S., Aldenhoff, J. B. & Schifflholz, T. Knockout of c-Jun N-terminal kinases 1, 2 or 3 isoforms induces behavioural changes. *Behav. Brain Res.* **245**, 88–95 (2013).
49. Barnat, M. et al. Distinct roles of c-Jun N-terminal kinase isoforms in neurite initiation and elongation during axonal regeneration. *J. Neurosci.* **30**, 7804–7816 (2010).
50. Haeusgen, W., Boehm, R., Zhao, Y., Herdegen, T. & Waetzig, V. Specific activities of individual c-Jun N-terminal kinases in the brain. *Neuroscience* **161**, 951–959 (2009).
51. Stefanoska, K. et al. Neuronal MAP kinase p38 $\alpha$  inhibits c-Jun N-terminal kinase to modulate anxiety-related behaviour. *Sci. Rep.* **8**, 14296 (2018).
52. Mohammad, H. et al. JNK1 controls adult hippocampal neurogenesis and imposes cell-autonomous control of anxiety behaviour from the neurogenic niche. *Mol. Psychiatry* **23**, 362–374 (2018).
53. Schmeisser, M. J. et al. The Nedd4-binding protein 3 (N4BP3) is crucial for axonal and dendritic branching in developing neurons. *Neural Dev.* **8**, 18 (2013).
54. Kiem, L. M., Dietmann, P., Linnemann, A., Schmeisser, M. J. & Kühl, S. J. The Nedd4 binding protein 3 is required for anterior neural development in *Xenopus laevis*. *Dev. Biol.* **423**, 66–76 (2017).
55. King, D. A. et al. Mosaic structural variation in children with developmental disorders. *Hum. Mol. Genet.* **24**, 2733–2745 (2015).
56. Tabarés-Seisdedos, R. & Rubenstein, J. L. R. Chromosome 8p as a potential hub for developmental neuropsychiatric disorders: implications for schizophrenia, autism and cancer. *Mol. Psychiatry* **14**, 563–589 (2009).
57. Cullen, P. J. Endosomal sorting and signalling: an emerging role for sorting nexins. *Nat. Rev. Mol. Cell Biol.* **9**, 574–582 (2008).
58. Ha, C. M. et al. SNX14 is a bifunctional negative regulator for neuronal 5-HT<sub>6</sub> receptor signaling. *J. Cell Sci.* **128**, 1848–1861 (2015).