

Genetic Analysis of Dog Congenital Deafness and Herding Behavior

Dissertation

to obtain the Ph. D. degree

in the International Ph. D. Program for Agricultural Sciences in Göttingen (IPAG) at

the Faculty of Agricultural Sciences,

Georg-August-University Göttingen, Germany

presented by

Fangzheng Xu

born in Shandong, China

Göttingen, June, 2021

Members of the Examination Board:

1. Referee: Prof. Dr. med. Ekkehard Schütz
2. Referee: Prof. Dr. Jens Tetens
3. Referee: Prof. Dr. Klaus Wimmers
4. Referee: Prof. Dr. Bertram Brenig
5. Referee: Prof. Dr. Michael Hölker

Date of dissertation: 02. June 2021

To my family

Table of Contents

Table of Contents	i
List of publications	ii
Summary	iv
Abbreviations	vii
Chapter 1 General introduction	10
1. <i>Hearing loss impact, classification and causes</i>	11
2. <i>Genetics of human non-syndromic hearing loss</i>	12
3. <i>Identifications of deafness genes in domestic animals</i>	13
4. <i>Genome-wide association analysis using modern dog breed standards as phenotypes</i>	16
5. <i>Genetic interpretations of dog behaviors</i>	17
6. <i>Genetic mapping of dog complex diseases such as neurological disorder and cancer</i>	19
7. <i>Effects of sperm membrane lipid on male fertility</i>	21
Chapter 2 A missense mutation in the <i>KLF7</i> gene is a potential candidate variant for congenital deafness in Australian Stumpy Tail Cattle Dogs	28
Chapter 3 Genome-wide Association Studies Reveal Neurological Genes for Dog Herding, Predation, Temperament and Trainability Traits	55
Chapter 4 Association of α/β-Hydrolase D16B with Bovine Conception Rate and Sperm Plasma Membrane Lipid Composition	91
Chapter 5 General discussion	119
1. <i>Dogs as biomedical models for human inherited hearing loss</i>	120
2. <i>Dog behavioral genetics studies contribute to the understanding of human mental disorders</i>	121
3. <i>Genetic mapping of dog diseases and morphological traits using whole genome re-sequences</i>	122
4. <i>Sperm lipidomics could be used to evaluate bull fertility</i>	123
5. <i>Prospects</i>	123
Acknowledgments	viii
Curriculum Vitae	x

List of publications

1-3 represent the main subjects of the presented thesis.

†: contributed equally. *: corresponding author.

- 1 **Xu, F.** †, Shan, S. †, Sommerlad, S., Seddon, J.M. and Brenig, B. *, 2021. A Missense Mutation in the KLF7 Gene Is a Potential Candidate Variant for Congenital Deafness in Australian Stumpy Tail Cattle Dogs. *Genes*, 12(4), p.467.
- 2 Shan, S., **Xu, F.** *, Brenig, B., 2021. Genome-wide Association Studies Reveal Neurological Genes for Dog Herding, Predation, Temperament and Trainability Traits. *Frontiers in veterinary science*, 8:693290.
- 3 Shan, S. †, **Xu, F.** †, Bleyer, M., Becker, S., Melbaum, T., Wemheuer, W., Hirschfeld, M., Wacker, C., Zhao, S., Schütz, E. and Brenig, B. *, 2020. Association of α/β -Hydrolase D16B with Bovine Conception Rate and Sperm Plasma Membrane Lipid Composition. *International journal of molecular sciences*, 21(2), p.627.
- 4 Shan, S., **Xu, F.**, Hirschfeld, M., Brenig, B. *, 2021. Sperm Lipid Markers of Male Fertility in Mammals. *International journal of molecular sciences*, 22(16), 8767.
- 5 Brenig, B. †*, Steingraber, L. †, Shan, S., **Xu, F.**, Hirschfeld, M., Andag, R., Spengeler, M., Dietschi, E., Mischke, R. and Leeb, T., 2019. Christmas disease in a Hovawart family resembling human hemophilia B Leyden is caused by a single nucleotide deletion in a highly conserved transcription factor binding site of the F9 gene promoter. *Haematologica*, 104(11), p.2307.
- 6 Ijaz, N., Liu, G., Jiang, X. *, Tesema, B., Wang, D., **Xu, F.**, Zhao, Q. and Huang, H., 2015. Genetic signature of strong recent positive selection at FSH β gene in goats. *Pakistan Journal of Agricultural Sciences*, 52(4).
- 7 **Xu, F.**, Jiang, X. *, Liu, G., Xie, G., Zhang, T., Wang, J. and Wang, R., 2015. Analysis on the origin of melanin in muscular tissue of black-boned goat. *Chinese Journal of Animal Science*, 51(11), pp.10-13.

Attended Conferences:

1. Poster: **F. Xu**, S. Shan, M. Bleyer, T. Melbaum, M. Hirschfeld, S. Zhao, E. Schuetz, and B. Brenig, 2021. "A novel genetic factor influencing male fertility in Holstein cattle." In *Reproduction in Domestic Animals*, vol. 56, pp. 17-17.
2. Presentation: **Xu F**, Shan S, Steingraber L, Hirschfeld M, Andag R, Spengeler M, Dietschi E, Mischke R, Leeb T, Brenig B. The hemophilia B Leyden disease in Hovawarts. The German Society for Animal Production (DGfZ) and the Society for Animal Sciences (GfT) Conference, Gießen, Germany (2019 September).

Summary

Strong artificial selections of canine morphological and behavioral traits lead to the formation of more than 400 modern dog (*Canis familiaris*, CFA) breeds within the past 300 years. Most dog breeds are derived from small numbers of founders, and this closed genetic pool within each breed results in the high frequency of occurrence of canine congenital disorders. The majority of these hereditary conditions share common clinical signs with corresponding human diseases. Therefore, dogs are appropriate spontaneous models for studying human diseases. Congenital deafness can cause both health and welfare problems in dogs, and it is quite prevalent among several dog breeds such as Dalmatian, Australian Cattle Dog, English Setter and Australian Stumpy Tail Cattle Dog (ASCD). However, hearing loss causative or associated genes in these dog breeds are not yet identified. The purpose of the study in **Chapter 2** was to identify congenital deafness related genes in ASCD. Three bilateral deaf and one normal hearing ASCDs were whole genome sequenced. The publicly available 722 canine whole genome sequences were also used to investigate potential causative mutations in this study. A case-control genome-wide association study (GWAS) was conducted by setting three deafness affected ASCDs as cases, and one unaffected ASCD and 43 additional herding group dogs were used as controls. The GWAS identified several loci on six chromosomes with potential canine deafness association (CFA3, 8, 17, 23, 28 and 37), and most (7 out of 13) of the significantly associated loci were located within CFA37. The private variants unique to three deaf ASCD were filtered by comparison to 722 canine controls of over 144 modern breeds. Subsequent annotation of these variants was performed, only potentially functional variants were filtered resulting in four remaining missense mutations. A missense mutation in the Kruppel-like factor 7 (*KLF7*) gene (NC_006619.3: g.15562684G>A; XP_022270984.1: p.Leu173Phe) on CFA37 could be emphasized to be associated considering the variant effect prediction and gene function. *KLF7* inner ear expression and a corresponding functional impact in development of inner ear and sensory neurons is known. Further genotyping of the

KLF7 variant in 28 affected and 27 normal hearing ASCDs still supported its association with ASCD congenital deafness.

Dogs have been selectively bred to intensify the performance abilities in regard to diverse tasks such as herding, hunting or companionship. Finally, modern dog breeds vary diversely in not only morphological but also behavioral traits. GWAS analysis of dog morphological traits using breed standard values have been well studied, and many auspicious genes were identified. However, due to the complexity of dog behavior traits, research progress on this topic is still limited. The study of **Chapter 3** was intended to elucidate the candidate genes underlying dog behavior traits including herding, predation, temperament and trainability. The phenotype information of these behavioral traits was obtained from American Kennel Club, which classified dog breeds into seven groups (Herding, Hound, Working, Terrier, Toy, Sporting and Non-sporting) based on the behavior, heritage and historical roles. 268 publicly available dog whole genome sequences of 130 modern breeds were used in this study. Four GWASs were performed to investigate potential candidate genes. Dogs with herding behavior were compared with the other dog categories by GWAS. Candidate neurological genes such as *THOC1*, *ASIC2*, *MSRB3*, *LLPH*, *RFX8* and *CHLI* were detected within or nearest to the significant loci of herding GWAS. Regarding dog predation behavior, herding behavior is the modified predatory behavior like repression of killing instinct, while hound dogs were selectively bred to enhance predation behaviors. We then use hound and herding group dogs in GWAS to analyze the dog predation behavior. Three neural genes *JAK2*, *MEIS1* and *LRRTM4* that were nearest to the significant loci of predation GWAS were revealed as candidates. In temperament GWAS, candidate neurological gene *ACSS3* was significantly associated with dog temperament trait. Dog behaviors were reported to be associated with body mass, so we repeated the four GWASs with incorporating dog breed standard body size as covariates. Similar results except for the significant associations of *ASIC2*, *JAK2* and *MEIS1* were observed, while these three candidate genes could contribute to dog behaviors through their effects on dog brain architecture. Linkage disequilibrium (LD) analysis of the herding GWAS significant associated

signals were also conducted. Promising neurological processes or cellular components were disclosed in GO analysis of potentially functional private genes of herding dogs.

In the study described in *Chapter 4*, one loss of function mutation in *ABHD16B* was identified to be associated with bull infertility. However, the exact gene function of *ABHD16B* remains unknown. Western blot was applied to locate ABHD16B protein expression, uncovering its occurrence in bull testis tissue but not in sperm cells. ABHD16B protein owns a function domain of α/β -hydrolase (ABHD) and several ABHD members are involved in lipid metabolism. It is assumed that *ABHD16B* could play roles in biosynthesis of sperm membrane lipids. Lipidomes of heterozygous and homozygous wild-type bull sperms were analyzed to explore potential aberrations. Several lipid components including PC, DAG, Cer, SM and PC were found significantly altered which verified our hypothesis. Therefore, the imbalanced lipid homeostasis of sperm membrane could be responsible for the bull infertility problem subjected in this study.

Abbreviations

ABHD16B	α/β -Hydrolase D16B
ACSS3	Acyl-CoA synthetase short chain family member 3
ADAMTSL3	ADAMTS like 3
AKC	American Kennel Club
ALOX15	Arachidonate 15-lipoxygenase
ASCD	Australian Stumpy-tail Cattle Dog
ASIC2	Acid sensing ion channel subunit 2
ATAC-seq	Assay for transposase-accessible chromatin using sequencing
ATG4D	Autophagy related 4D cysteine peptidase
ATGL	Adipose triglyceride lipase
ATP1B2	ATPase Na ⁺ /K ⁺ transporting subunit beta 2
ATXN1	Ataxin 1
BOR	Branchio-oto-renal
CAMKK2	Calcium/calmodulin dependent protein kinase kinase 2
CAPN1	Calpain 1
C-BARQ	Canine Behavioral Assessment and Research Questionnaire
CDH2	Cadherin 2
CDK5RAP2	CDK5 regulatory subunit associated protein 2
Cer	Ceramide
CFA	<i>Canis familiaris</i>
ChIP-Seq	Chromatin immunoprecipitation sequencing
CHL1	Cell adhesion molecule L1 like
CLDN14	claudin 14
CNTNAP1	Contactin associated protein 1
COL11A1	Collagen type XI alpha 1 chain
CTNNA2	Catenin alpha 2
DAG	Diacylglycerol
DENR	Density regulated re-initiation and release factor
DHA	Docosahexanoic acid
DIRAS1	DIRAS family GTPase 1
DMD	Dystrophin
DSCAM	DS cell adhesion molecule
FAM134B	Reticulophagy regulator 1
FBXW7	F-box and WD repeat domain containing 7
FGF4	Fibroblast growth factor 4
FGF5	Fibroblast growth factor 5
GDAP1	Ganglioside induced differentiation associated protein 1
GHR	Growth hormone receptor
GO	Gene Ontology
GRM1	Glutamate metabotropic receptor 1

GSDME	Gasdermin E
GTF2I	General transcription factor Iii
GTF2IRD1	GTF2I repeat domain containing 1
GWAS	Genome-wide association study
HMGA2	High mobility group AT-hook 2
HOXA1	Homeobox A1
HS6ST2	Heparan sulfate 6-O-sulfotransferase 2
HTC	Herding Trait Characterization
IFT122	Intraflagellar transport 122
IGF1	Insulin like growth factor 1
IGF1R	Insulin like growth factor 1 receptor
IMPG2	Interphotoreceptor matrix proteoglycan 2
ITPR1	Inositol 1,4,5-trisphosphate receptor type 1
JAK2	Janus kinase 2
KCNJ10	Potassium inwardly rectifying channel subfamily J member 10
KIT	KIT proto-oncogene, receptor tyrosine kinase
KITLG	KIT ligand
KLF7	Kruppel-like factor 7
LD	Linkage disequilibrium
LLPH	LLP homolog, long-term synaptic facilitation factor
LOXHD1	Lipoxygenase homology PLAT domains 1
LRIT3	Leucine rich repeat, Ig-like and transmembrane domains 3
LRRTM4	Leucine rich repeat transmembrane neuronal 4
MEIS1	Meis homeobox 1
MITF	Melanocyte inducing transcription factor
MSRB3	Methionine sulfoxide reductase B3
MYO7A	Myosin VIIA
OCD	Obsessive-compulsive disorder
P2RX7	Purinergic receptor P2X 7
PC	Phosphatidylcholine
PDE6B	Phosphodiesterase 6B
PFA	Polyunsaturated fatty acid
PGCP	Pepsinogen B, pseudogene
PITRM1	Pitrilysin metallopeptidase 1
PLA2G6	Phospholipase A2 group VI
PPP2R2B	Protein phosphatase 2 regulatory subunit Bbeta
PTPRQ	Protein tyrosine phosphatase receptor type Q
RAB24	RAB24, member RAS oncogene family
RBBP6	RB binding protein 6, ubiquitin ligase
RFX8	Regulatory factor X8
RSPO2	R-spondin 2
SBF2	SET binding factor 2
SEL1L	SEL1L adaptor subunit of ERAD E3 ubiquitin ligase
SLC12A6	Solute carrier family 12 member 6

SM	Sphingomyelin
SMAD2	SMAD family member 2
SNP	Single-nucleotide polymorphism
SNX14	Sorting nexin 14
SOD1	Superoxide dismutase 1
SP110	SP110 nuclear body protein
SPTBN2	Spectrin beta, non-erythrocytic 2
STC2	Stanniocalcin 2
STX8	Syntaxin 8
TAG	Triacylglycerol
THOC1	THO complex 1
TRPC6	Transient receptor potential cation channel subfamily C member 6
USP31	Ubiquitin specific peptidase 31
WGS	Whole genome sequencing

Chapter 1

General introduction

1. Hearing loss impact, classification and causes

Hearing loss is one common sensory defect with around 466 million affected people globally, and it is estimated to be the fourth highest human disability cause in the world [1]. It is predicted by World Health Organization that the number of people with disabled hearing ability will increase to 630 million by the year 2030 [2]. It is quite urgent to develop novel therapies to prevent the increasing trend. Understanding the molecular genetic mechanism that enhance or reduce susceptibility to hearing loss is fundamental for future deafness therapies. Impaired communication is the direct obstacle for hearing loss patients, which will reduce their life quality by affecting education and chances in professional career [3, 4]. Deaf people also suffer higher risk from mental diseases, like dementia [5] and depression [6]. Hearing loss inconveniences animals and increases their risk of mortality because auditory function is vital for responding to external environmental dangerous factors such as predators and cars. In domestic dogs, deafness compromises their health, welfare and utility. Working dogs with deafness may lose the ability to perform tasks [7].

Deafness can be classified following different parameters. It can be categorized mainly by three criteria in humans: (1) genetic or non-genetic cause; (2) syndromic association (combination with other abnormalities) or non-syndromic (with only hearing loss disorder); (3) prelingual or postlingual onset [8]. Deafness can also be classified by the number of affected ears: unilateral or bilateral hearing loss. Depending on the severity degree of hearing loss for the better-hearing ear, it can be classified into four hearing loss levels: mild, moderate, severe and profound. Regarding the affected sites of deafness, it can be classified into conductive or sensorineural hearing loss. External and/or middle ear dysfunctions belong to conductive hearing loss, and other defects from inner ear to brain cortical auditory centers are regarded as sensorineural deafness. Mixed hearing loss is recognized if conductive and sensorineural components were present at the same time [9].

Disfunctions in both peripheral and central parts of auditory systems could lead to hearing loss. The peripheral auditory part is composed by outer ear, middle ear and inner ear (cochlea). Inner ear is the site for transformation of physical sound waves into neural signals. Harmful noise exposure could cause physical damage into sensory hair cells of the inner ear, which is a quite common causative factor in acquired deafness [10, 11]. Besides physical causing factors, more than 400 diverse syndromes involving hearing impairment have been reported to be caused by genetic mutations [12], and more than 120 genes have been reported to cause human non-syndromic hearing loss [13]. Other factors such as aging [14], therapy drug side effects on auditory system [15, 16] and chronic diseases [17, 18] also contribute to hearing loss onset.

2. Genetics of human non-syndromic hearing loss

Genetics of hereditary deafness is highly heterogeneous in humans, as it can be caused by a single mutation (monogenic) or by a combination of different genes (polygenic). Environmental factors could also contribute to deafness pathogenesis, which will be considered as multifactorial hearing loss. The inheritance patterns of human deafness are also diverse being either autosomal recessive, autosomal dominant, X-linked or mitochondrial [19]. More than 120 genes have been identified to be associated with human non-syndromic deafness including 50 autosomal dominant, 77 autosomal recessive and five X-linked genes (Table 1) [20].

Table 1. Human non-syndromic hearing loss genes.

Hereditary Model	Genes
Autosomal Recessive	<i>GJB2, GJB6, MYO7A, MYO15A, SLC26A4, TMIE, TMCI, TMPRSS3, OTOF, CDH23, GIPC3, STRC, USH1C, OTOG, TECTA, OTOA, PCDH15, RDX, GRXCRI, GAB1, TRIOBP, CLDN14, MYO3A, WHRN, CDC14A, ESRRB, ESPN, MYO6, HGF, ILDR1, ADCY1, CIB2, MARVELD2, BDPI, COL11A2, PDZD7, PJVK, SLC22A4, SLC26A5, LRTOMT/COMT2, DCDC2, LHFPL5, S1PR2, PNPT1, BSND, MSRB3, SYNE4, LOXHD1, TPRN, GPSM2, PTPRQ, OTOGL, TBC1D24, ELMOD3, KARS, SERPINB6, CABP2, NARS2, MET, TSPEAR, TMEM132E, PPIP5K2, GRXCR2, EPS8, CLIC5, FAM65B/RIPOR2, DFNB32, EPS8L2, ROR1, WBP2, ESRP1, MPZL2, CEACAM16, GRAP, SPNS2, CLDN9, CLRN2, DIAPH1, KCNQ4, GJB3, IFNLRI, GJB2, GJB6, MYH14, CEACAM16, GSDME/DFNA5, WFS1, LMX1A, TECTA, COCH, EYA4, MYO7A, COL11A2, POU4F3, MYH9, ACTG1, MYO6, SIX1,</i>
Autosomal Dominant	<i>SLC17A8, REST, GRHL2, NLRP3, TMCI, COL11A1, CRYM, P2RX2, CCDC50, MIRN96, TJP2, TNC, SMAC/DIABLO, TBC1D24, CD164, OSBPL2, HOMER2, KITLG, MCM2, PTPRQ, DMXL2, MYO3A, PDE1C, TRRAP, PLS1, SCD5, SLC12A2, MAP1B, RIPOR2/FAM65B</i>
X-Linked	<i>PRPS1, POU3F4, SMPX, AIFM1, COL4A6</i>

Gene lists were obtained from: <https://hereditaryhearingloss.org> (accessed on 24 February 2021).

3. Identifications of deafness genes in domestic animals

Hearing loss in domestic animals can also be caused by several factors. Hereditary deafness contributes a significant amount in several domestic animal species such as pigs and dogs. Hereditary deafness can be classified into late onset and congenital deafness, and white pigmentation genes have been reported to be associated with hereditary deafness in several domestic animal species. In pigs, *de novo* mutation within non-regulatory region of the melanocyte-specific promoter of *MITF* gene causes hearing loss [21]. One missense mutation of *KIT* gene leads to both congenital bilateral severe sensorineural deafness and hypopigmentation in Bama miniature pigs. This disease resembles one common human syndromic hearing loss disease - the

Waardenburg syndrome [22]. In cattle, German White Fleckvieh syndrome is characterized by typical characters such as inherited bilateral deafness, colobomatous eyes, hypopigmentation and heterochromia irides. One missense mutation of *MITF* gene has been detected to be associated with the above-mentioned syndrome [23]. One approximate 63 kb deletion spanning exons 6-9 of the *MITF* gene is reported to be associated with splashed white depigmentation phenotype of American Paint horses, which is also identified as a risk variant of deafness in this horse breed [24].

Congenital deafness is quite prevalent in dogs, which has been reported in more than 100 dog breeds [25]. Several modern dog breeds were attested to have high hearing loss prevalence. Dalmatian dog breed was identified to have the most prevalent deafness rate with 29.9%. Congenital hearing loss is also quite common in Bull terrier, particularly in the white fur individuals. Other dog breeds such as English Setter, English Cocker Spaniel, Australian Cattle Dog and Australian Stumpy Tail Cattle Dog were also reported to show high hearing loss percentage [26-28]. Other dog breeds that were not investigated in detail to date might also bear a high deafness rate. Considering the great harmfulness and disadvantages of deafness, this problem should be subjected in dog breeding guidelines. The correlation between pigmentation and the occurrence of deafness has also been observed in dogs. White coat color seems to be associated with the deafness onset in some certain dog breeds like Bull Terrier [25]. Red coat color is positively associated with congenital deafness in Australian Stumpy Tail Cattle dogs when compared with blue ones [28]. On the other hand, blue eyes in dogs are likely to be associated with deafness onset, e.g. it was observed that Dalmatian dogs with blue eyes have higher risk for deafness. It is similar in English Setter and English Cocker Spaniel dog breeds: blue eyed dogs are more susceptible to congenital hearing loss [25]. However, not all dog breeds have shown a relationship between coat color and deafness. Investigations on breeds including Australian Cattle Dog showed no significant associations between coat color and deafness [27]. No significant differences of deafness associations were observed between black- and liver-spotted Dalmatians [25].

Congenital deafness in humans has been well studied. Because of the similarities between canine and human genetic disorders [29], many human deafness causative genes are considered as candidates for studying dogs with hearing loss. Human deafness candidate gene-based diagnosis has been performed in dogs, like exclusion of *PAX3*, *TMC1*, *TMIE*, *SILV*, *ESPN*, *MYO3A*, *SLC26A5* and *USH1C* as causative or associated genes for Dalmatian hereditary deafness [30-33]. The costs for canine SNP chips and whole genome sequencing (WGS) have declined in recent years. Several GWAS and WGS analysis were applied to detect the causative variants or genes for dog deafness. For example, seven quantitative trait loci were revealed by GWAS for congenital sensorineural deafness in Dalmatian dogs (235 Dalmatian dogs including 157 hearing and 78 deaf samples), five significantly associated loci on CFA6, 14, 27, 29 and 31 were detected near the well-known human or mouse deafness genes. These were *COL11A1* on CFA6, *GSDME* and *HOXA1* on CFA14, *GDAP1* on CFA29 and *CLDN14* on CFA31, while no causative genes were found in this study [34]. One recent GWAS of dog congenital deafness was conducted in three breeds (Dalmatian, Australian Cattle Dog and English Setter) with a high deafness prevalence, all of which have piebald coats. Though several loci exceeded the suggestive association threshold, no loci in common of these three breeds and none of these significant signals were near the piebald loci. It implies congenital pigment-associated deafness is a complex trait and larger sample size is required to figure out the causative genes [35]. Whole genome re-sequencing of hearing loss dogs has recently been proved to be one efficient way to identify causative genes. In Doberman Pinscher breed, *PTPRQ* and *MYO7A* have been detected to be associated with hearing loss [36, 37]. By combining GWAS with targeted next-generation sequencing, *USP31* and *RBBP6* showed strong associations with adult-onset deafness in Border Collies. These two genes are either involved in NF- κ B pathway or cochlear development [38]. Recently, a rare missense variant within in one human deafness gene *LOXHD1* was detected to be correlated with hearing loss in Rottweiler breed [39]. Congenital deafness in dogs is much more complex than just associated with white pigmentation, genes that are involved in sensory nervous and auditory system development should also be evaluated during genetic diagnosis.

4. Genome-wide association analysis using modern dog breed standards as phenotypes

Dogs were the first domesticated animal species, which are assumed to accompany humans since approximately 30,000 years according to archaeological records [40]. Recent studies found that modern dogs have originated from the now-extinct wolf population more than 11,000 years ago [41]. More than 400 documented dog breeds with various stable breed phenotypes were recognized around the world. The modern dog breeds were formed in less than three centuries with strong and continuous human-purpose based selective breeding. Diverse breed-specific morphological traits of dogs were formed such as body mass, skull shape and coat color. High morphological variations were present in different dog breeds, while the within-breed phenotypic difference is typically low [42]. This is due to the strict regulations for modern dog breed standards, and dogs of the same breed are required to have specific and unified morphological traits. The variation between dog breeds is around 27.5 percent, while genetic variation among human populations is only 5.4 percent [43]. Psychological characteristics of dogs are unique and diverse among modern dog breeds. For example, there might be a 50-fold difference in body size between Chihuahua and Saint Bernard breeds. Therefore, dogs can be good genetic models to study underlying molecular mechanism of these morphological traits. After the dog reference genome had been assembled in 2005 [44], GWAS with dog breed standards as phenotypes has become a powerful way to identify genetic factors underlying these traits. Meanwhile, extensive linkage disequilibrium (LD) (megabase scale) in dog genome makes it easier to get significant signals with smaller sample size [45]. Several convincing genes were pointed out in the first landmark study which was conducted in 2008 [46]. Notably, the body size associated genes were confirmed by subsequent researches. Many successful examples of genetic mapping of canine morphological traits using dog breed stereotypes have been conducted [47, 48]. Compared with humans, some complex traits

of purebred dogs can be explained by a rather small number of genes such as body size. Nearly 50% of dog breed standard weight variants could be explained by only six genes (*IGF1*, *GHR*, *SMAD2*, *STC2*, *HMGA2*, *IGF1R*), while 180 loci explained only around 10% of individual height in humans [49]. Other morphological traits that are shared across dog breeds can be used in case-control GWAS analysis such as ear shape. Highly significant signals were detected on CFA10 for drop ear and CFA12 for large and round ears [50]. In the same study, several other morphological traits like furnishing (*RSPO2* gene on CFA13, $p = 1.06 \times 10^{-68}$) and length of fur (*FGF5* gene on CFA32, $p = 4.71 \times 10^{-20}$) were also analyzed and convincing results were obtained. All of these traits were detected with single and very highly associated ($-\log(p\text{-value})$) signals. Moreover, a retrogene gene encoding fibroblast growth factor 4 (*FGF4*) on CFA18 has been identified to cause chondrodysplasia (short-legged phenotype in dog breeds such as Dachshund and Corgi) by GWAS in 2009 [51]. Eight years later, the same retrogene on CFA12 was demonstrated to be related with chondrodystrophy and intervertebral disc disease in dogs [52].

However, longer LD block of dog genome is a double-sided sword. It has disadvantages in fine mapping of causative variants. For instance, several studies have detected significantly associated signals within or around *HMGA2* gene indicating its determination role of small body size in dogs, while the causal mutation is still unknown [53].

5. Genetic interpretations of dog behaviors

During the dog breed formation, strong human selective breeding has also resulted in specific and diverse behavior traits among dog breeds. Different dog breeds own different characteristics such as aggressiveness and boldness. Using breed average values of behavioral traits as phenotypes, several GWASs were applied in detection of correlated genes across breeds. For example, dog behavioral stereotypes (herding,

pointing, boldness and trainability) have been firstly investigated as qualitative variables in 2008 by genome-wide mapping between modern dog breeds [46]. In another study, average values of five dog personality traits including aggressiveness, sociability, curiosity, chase-proneness and playfulness were used as phenotypes in GWAS mapping, and these data are obtained from the Swedish Kennel Club [48]. In the same study, boldness trait was also investigated. GWAS mapping of behavioral traits using breed-average values that were obtained from large scale questionnaires as phenotypes has been applied in several studies. For instance, Canine Behavioral Assessment and Research Questionnaire (C-BARQ) breed-average scores were applied in GWAS analysis to map dog personality traits [54]. GWAS of more than 100 dog breeds using breed-average C-BARQ scores were calculated from 29,656 pet dogs. MacLean et al, found that dog behavioral traits are highly heritable between breeds, and the significantly associated genes are involved in neurological development and express in brain. Genetic factors contributing to dog cognition traits including inhibitory control, communication, memory and physical reasoning were also analyzed [56]. GWAS using breed-average cognition values obtained from questionnaires has also revealed several neural genes [57]. A recent study of social skills such as communication with human beings in 375 8-week-old puppies have indicated that dog cognition is highly heritable, and genetic factors account for 40% of the variation of dog point-following abilities and attention to human faces [58]. American Kennel Club (AKC) is the most authoritative organization to register and classify modern purebred dogs in United States. Considering heritage, behavior, and physical attributes of dog breeds, 197 AKC recognized dog breeds were assigned into seven loosely defined groups: herding, hunting, terrier, toy, sporting, non-sporting and working [59]. Using AKC group classification information, genes involved in athletic success of sporting and hound dogs were identified by comparing with village and other group dogs [60].

Dog breed-specific behaviors are correlated with body sizes [61], which might be caused by the effect of brain mass differences between breeds with diverse body sizes [46, 54, 62]. Including body size effects into GWAS analysis could reveal residual

genes that are not explained by effects of brain size [55]. Thus, correcting body mass factors in the GWAS analysis of dog behaviors has both advantages and disadvantages [57].

6. Genetic mapping of dog complex diseases such as neurological disorder and cancer

Selective breeding programs of purebred dogs have resulted in an increased frequency of inherited diseases such as neurological diseases in dogs [63]. Genetic mapping of these neurological disorders has been conducted in several studies of various breeds, which could be applied to improve the health and welfare levels for dogs. Promising candidate genes including *CDH2*, *CTNNA2*, *ATXN1* and *PGCP* were identified for canine obsessive-compulsive disorders [64]. Canine myoclonic epilepsy is caused by one 4-bp deletion of *DIRAS1* gene in Rhodesian Ridgeback dogs, which could be used as naturally occurring models for human epilepsy [65]. Recently, an in-frame 6-bp deletion in *PITRMI* gene has been detected to be associated with epilepsy, mitochondrial dysfunction and neurodegeneration in Parson Russel Terriers [66]. Neuroaxonal dystrophy in Papillon dogs was shown to be associated with a missense variant within *PLA2G6* gene [67]. Mutations with *DMD* gene have been detected to be responsible for Duchenne muscular dystrophy disorder in 15 dog breeds (<https://omia.org/OMIA001081/9615/>). Progressive retinal atrophy resembling human retinitis pigmentosa was revealed to be associated with *IFT122* gene in Lapponian herders [68]. Mutations within human amyotrophic lateral sclerosis causative gene *SOD1* showed significant association with canine degenerative myelopathy, one modification risk gene *SP110* was also identified [69, 70]. Sensory neuropathy in the Border Collies was caused by an inversion within *FAM134B* [71]. Demyelinating polyneuropathy in Miniature Schnauzer is caused by a splicing variant of *SBF2* [72]. A missense variant of *CNTNAP1* gene was considered as one candidate mutation for canine laryngeal paralysis and polyneuropathy [73]. Using GWAS and WGS analysis,

a deletion in *LRIT3* was identified to be associated with canine congenital stationary night blindness [74]. One novel canine inherited neurodegenerative disorder (neurodegenerative vacuolar storage disease) was described in Lagotto Romagnolo dog breed, and *ATG4D* was identified as the potential causative gene [75]. Progressive retinal atrophy was related to *IMPG2* in Lhasa Apso dogs [76], while *PDE6B* was detected to be the candidate gene in Spanish Water Dogs [77]. Several causative genes including *ATP1B2* [78], *CAPN1* [79], *GRM1* [80], *ITPR1* [81], *KCNJ10* [82-84], *RAB24* [85], *SEL1L* [86], *SNX14* [87], *SPTBN2* [88] and *SLC12A6* [89] were identified for canine hereditary ataxia in various breeds.

Since dogs suffer from several cancers similar to humans, the genetics of dog cancer has recently been studied utilizing these valuable spontaneous models [90]. For example, GWAS and fine mapping using only 31 cases and 34 controls could identify the predisposing gene *KITLG* for canine squamous cell carcinoma of the digit in Standard Poodles [91]. Via combined GWAS with gene expression analysis, two risk genes (*TRPC6* and *STX8*) have been shown to contribute to B-cell lymphoma and hemangiosarcoma of Golden Retriever [92]. Risk variant of *DSCAM* has been detected to be significantly associated with mast cell tumor in Labrador and Golden Retrievers [93]. Candidate genes including *CDK5RAP2* have been identified for canine mammary tumors [94]. Three glioma susceptibility genes (*CAMKK2*, *P2RX7* and *DENR*) were detected to be significantly associated in across-breed GWAS using 39 dog glioma cases and 141 controls from 25 dog breeds [94]. GWASs of three hematopoietic cancers including histiocytic sarcoma, lymphoma and mast cell tumor in four dog breeds identified several common susceptible loci [95]. Mutations of *FBXW7* gene have been detected to be negatively associated with prognosis in canine B-cell lymphoma using exome sequencing of 71 affected dogs [96].

7. Effects of sperm membrane lipid on male fertility

Lipid homeostasis of sperm membrane is critical for male fertility. For example, the cholesterol/phospholipid ratio in sperm of unexplained infertile humans is about twice that of fertile men, so the ratio could be an indicator for male infertility [97]. It was suggested that lower phospholipid content in spermatozoa could be the cause of unexplained male reproduction disorders [97]. The cholesterol/phospholipid ratio was reported to be negatively associated with capacitation speed [98]. Reduction of cholesterol content of sperm membrane occurs during capacitation [99]. Abnormally high cholesterol content was observed in human sperms which failed to fertilize eggs *in vitro* [100]. *ATGL* is highly expressed in testis and it hydrolyses long-chain fatty acid triacylglycerol (TAG) to diacylglycerol (DAG). Impaired male fertility was observed in deficient *ATGL*^(-/-) mice. The deletion of *ATGL* gene could lead to defects in spermatogenesis in testis and also affect sperm maturation in epididymis, which is essential for sperms to acquire motility [101]. Lipid peroxidation and overproduction of reactive oxygen species have been reported to cause sperm defects in male infertile patients [102-105]. *ALOX15* and its inhibitor (6,11-dihydro[1]benzothiopyrano[4,3-b]indole) were reported to be involved in mouse and human sperm oxidative stress process [106, 107]. Increased *ALOX15* abundance was observed in sperms of infertile patients which could result in infertility by accelerating sperm membrane lipid peroxidation [108]. Several polyunsaturated fatty acids (PUFAs) of sperm membrane have been studied in male fertility. Reduced docosahexanoic acid (DHA) levels were detected in dysfunctional sperms with lower motility of boar (low motility) [109] and man (asthenozoospermia and oligozoospermia) [110]. Significantly elevated levels of phosphatidyl serine and some n-6 PUFAs with decreased amount of phosphatidyl ethanolamine and n-3 PUFAs were observed in sperms of infertile patients, indicating that disorders of sperm lipid metabolism may be the cause of male infertility [111].

References

1. Cunningham, L.L. and D.L. Tucci, Hearing Loss in Adults. *New England Journal of Medicine*, 2017. 377(25): p. 2465-2473.
2. Organization, W.H., Addressing the rising prevalence of hearing loss. 2018.
3. Kamil, R.J. and F.R. Lin, The Effects of Hearing Impairment in Older Adults on Communication Partners: A Systematic Review. *Journal of the American Academy of Audiology*, 2015. 26(2): p. 155-182.
4. Mick, P., I. Kawachi, and F.R. Lin, The Association between Hearing Loss and Social Isolation in Older Adults. *Otolaryngology-Head and Neck Surgery*, 2014. 150(3): p. 378-384.
5. Lin, F.R. et al., Hearing Loss and Incident Dementia. *Archives of Neurology*, 2011. 68(2): p. 214-220.
6. Li, C.M. et al., Hearing Impairment Associated With Depression in US Adults, National Health and Nutrition Examination Survey 2005-2010. *Jama Otolaryngology-Head & Neck Surgery*, 2014. 140(4): p. 293-302.
7. Strain, G.M., Aetiology, prevalence and diagnosis of deafness in dogs and cats. *British Veterinary Journal*, 1996. 152(1): p. 17-36.
8. Willems, P.J., Genetic causes of hearing loss. *New England Journal of Medicine*, 2000. 342(15): p. 1101-1109.
9. Petit, C., J. Leveilliers, and J.-P. Hardelin, Molecular genetics of hearing loss. *Annual review of genetics*, 2001. 35(1): p. 589-645.
10. Hammer, M.S., T.K. Swinburn, and R.L. Neitzel, Environmental Noise Pollution in the United States: Developing an Effective Public Health Response. *Environmental Health Perspectives*, 2014. 122(2): p. 115-119.
11. Carroll, Y.I. et al., Vital Signs: Noise-Induced Hearing Loss Among Adults - United States 2011-2012. *Mmwr-Morbidity and Mortality Weekly Report*, 2017. 66(5): p. 139-144.
12. Gettelfinger, J.D. and J.P. Dahl, Syndromic hearing loss: a brief review of common presentations and genetics. *Journal of pediatric genetics*, 2018. 7(1): p. 1.
13. Vona, B. et al., Non-syndromic hearing loss gene identification: A brief history and glimpse into the future. *Molecular and Cellular Probes*, 2015. 29(5): p. 260-270.
14. Li-Korotky, H.-S., Age-related hearing loss: quality of care for quality of life. *The Gerontologist*, 2012. 52(2): p. 265-271.
15. Duggal, P. and M. Sarkar, Audiologic monitoring of multi-drug resistant tuberculosis patients on aminoglycoside treatment with long term follow-up. *BMC Ear, Nose and Throat Disorders*, 2007. 7(1): p. 1-7.
16. Garinis, A.C. et al., The cumulative effects of intravenous antibiotic treatments on hearing in patients with cystic fibrosis. *Journal of Cystic Fibrosis*, 2017. 16(3): p. 401-409.
17. Cruickshanks, K.J. et al., Smoking, central adiposity, and poor glycemic control increase risk of hearing impairment. *Journal of the American Geriatrics Society*, 2015. 63(5): p. 918-924.
18. Bainbridge, K.E., H.J. Hoffman, and C.C. Cowie, Risk factors for hearing impairment among US adults with diabetes: National Health and Nutrition Examination Survey 1999–2004. *Diabetes care*, 2011. 34(7): p. 1540-1545.

19. Smith, R.J., J.F. Bale Jr, and K.R. White, Sensorineural hearing loss in children. *The Lancet*, 2005. 365(9462): p. 879-890.
20. Van Camp, G. and R.J. Smith, Hereditary hearing loss homepage. 2006.
21. Chen, L. et al., A de novo silencer causes elimination of MITF-M expression and profound hearing loss in pigs. *Bmc Biology*, 2016. 14.
22. Xu, C. et al., KIT gene mutation causes deafness and hypopigmentation in Bama miniature pigs. *American Journal of Translational Research*, 2020. 12(9): p. 5095-+.
23. Philipp, U. et al., A MITF Mutation Associated with a Dominant White Phenotype and Bilateral Deafness in German Fleckvieh Cattle. *Plos One*, 2011. 6(12).
24. Henkel, J. et al., Whole-genome sequencing reveals a large deletion in the MITF gene in horses with white spotted coat colour and increased risk of deafness. *Animal Genetics*, 2019. 50(2): p. 172-174.
25. Strain, G.M., Deafness prevalence and pigmentation and gender associations in dog breeds at risk. *Veterinary Journal*, 2004. 167(1): p. 23-32.
26. Strain, G.M., Hereditary deafness in dogs and cats: causes, prevalence, and current research. *strain*, 2003. 225: p. 578-9758.
27. Sommerlad, S.F. et al., Prevalence of congenital hereditary sensorineural deafness in Australian Cattle Dogs and associations with coat characteristics and sex. *Bmc Veterinary Research*, 2012. 8.
28. Sommerlad, S. et al., Congenital Sensorineural Deafness in Australian Stumpy-Tail Cattle Dogs Is an Autosomal Recessive Trait That Maps to CFA10. *Plos One*, 2010. 5(10).
29. Shearin, A.L. and E.A. Ostrander, Leading the way: canine models of genomics and disease. *Disease Models & Mechanisms*, 2010. 3(1-2): p. 27-34.
30. Brenig, B. et al., Analysis of the 5' region of the canine PAX3 gene and exclusion as a candidate for Dalmatian deafness. *Animal Genetics*, 2003. 34(1): p. 47-50.
31. Mieskes, K. and O. Distl, Elimination of TMC1 and TMIE as candidates for hereditary non-syndromic deafness in Dalmatian dogs. *Animal Genetics*, 2006. 37(5): p. 519-521.
32. Stritzel, S., A. Wohlke, and O. Distl, Elimination of SILV as a candidate for congenital sensorineural deafness in Dalmatian dogs. *Animal Genetics*, 2007. 38(6): p. 662-663.
33. Mieskes, K. and O. Distl, Evaluation of ESPN, MYO3A, SLC26A5 and USH1C as candidates for hereditary non-syndromic deafness (congenital sensorineural deafness) in Dalmatian dogs. *Animal Genetics*, 2007. 38(5): p. 533-534.
34. Kluth, S. and O. Distl, Congenital Sensorineural Deafness in Dalmatian Dogs Associated with Quantitative Trait Loci. *Plos One*, 2013. 8(12).
35. Hayward, J.J. et al., A genome-wide association study of deafness in three canine breeds. *Plos One*, 2020. 15(5).
36. Guevar, J. et al., Deafness and vestibular dysfunction in a Doberman Pinscher puppy associated with a mutation in the PTPRQ gene. *Journal of Veterinary Internal Medicine*, 2018. 32(2): p. 665-669.
37. Webb, A.A., A.L. Ruhe, and M.W. Neff, A missense mutation in MYO7A is associated with bilateral deafness and vestibular dysfunction in the Doberman pinscher breed. *Canadian Journal of Veterinary Research-Revue Canadienne De Recherche Veterinaire*, 2019. 83(2): p. 142-148.
38. Yokoyama, J.S. et al., Variation in Genes Related to Cochlear Biology Is Strongly Associated with Adult-Onset Deafness in Border Collies. *Plos Genetics*, 2012. 8(9).

39. Hytönen, M.K. et al., Missense Variant in LOXHD1 is Associated With Canine Nonsyndromic Hearing Loss. 2021.
40. Germonpré, M., M. Lázničková-Galetová, and M.V. Sablin, Palaeolithic dog skulls at the Gravettian Předmostí site, the Czech Republic. *Journal of Archaeological Science*, 2012. 39(1): p. 184-202.
41. Bergstrom, A. et al., Origins and genetic legacy of prehistoric dogs. *Science*, 2020. 370(6516): p. 557-563.
42. Bannasch, D.L., C.F. Baes, and T. Leeb, Genetic Variants Affecting Skeletal Morphology in Domestic Dogs. *Trends in Genetics*, 2020. 36(8): p. 598-609.
43. Ostrander, E.A., Genetics and the shape of dogs: studying the new sequence of the canine genome shows how tiny genetic changes can create enormous variation within a single species. *American Scientist*, 2007. 95(5): p. 406-413.
44. Lindblad-Toh, K. et al., Genome sequence, comparative analysis and haplotype structure of the domestic dog. *Nature*, 2005. 438(7069): p. 803-819.
45. Karlsson, E.K. et al., Efficient mapping of mendelian traits in dogs through genome-wide association. *Nat Genet*, 2007. 39(11): p. 1321-1328.
46. Jones, P. et al., Single-nucleotide-polymorphism-based association mapping of dog stereotypes. *Genetics*, 2008. 179(2): p. 1033-1044.
47. Boyko, A.R. et al., A Simple Genetic Architecture Underlies Morphological Variation in Dogs. *Plos Biology*, 2010. 8(8).
48. Vaysse, A. et al., Identification of Genomic Regions Associated with Phenotypic Variation between Dog Breeds using Selection Mapping. *Plos Genetics*, 2011. 7(10).
49. Rimbault, M. et al., Derived variants at six genes explain nearly half of size reduction in dog breeds. *Genome Research*, 2013. 23(12): p. 1985-1995.
50. Plassais, J. et al., Whole genome sequencing of canids reveals genomic regions under selection and variants influencing morphology. *Nature Communications*, 2019. 10.
51. Parker, H.G. et al., An Expressed Fgf4 Retrogene Is Associated with Breed-Defining Chondrodysplasia in Domestic Dogs. *Science*, 2009. 325(5943): p. 995-998.
52. Brown, E.A. et al., FGF4 retrogene on CFA12 is responsible for chondrodystrophy and intervertebral disc disease in dogs. *Proceedings of the National Academy of Sciences of the United States of America*, 2017. 114(43): p. 11476-11481.
53. Schoenebeck, J.J. and E.A. Ostrander, Insights into Morphology and Disease from the Dog Genome Project. *Annual Review of Cell and Developmental Biology*, Vol 30, 2014. 30: p. 535-560.
54. Zapata, I., J.A. Serpell, and C.E. Alvarez, Genetic mapping of canine fear and aggression. *Bmc Genomics*, 2016. 17.
55. MacLean, E.L. et al., Highly heritable and functionally relevant breed differences in dog behaviour. *Proceedings of the Royal Society B-Biological Sciences*, 2019. 286(1912).
56. Gnanadesikan, G.E. et al., Estimating the heritability of cognitive traits across dog breeds reveals highly heritable inhibitory control and communication factors. *Animal Cognition*, 2020. 23(5): p. 953-964.
57. Gnanadesikan, G.E. et al., Breed Differences in Dog Cognition Associated with Brain-Expressed Genes and Neurological Functions. *Integrative and Comparative Biology*, 2020. 60(4): p. 976-990.

58. Bray, E.E. et al., Early-Emerging and Highly-Heritable Sensitivity to Human Communication in Dogs. *bioRxiv*, 2021.
59. Club, A.K., *The complete dog book*. 2006: Complete Dog Book.
60. Kim, J. et al., Genetic selection of athletic success in sport-hunting dogs. *Proceedings of the National Academy of Sciences of the United States of America*, 2018. 115(30): p. E7212-E7221.
61. McGreevy, P.D. et al., Dog Behavior Co-Varies with Height, Bodyweight and Skull Shape. *Plos One*, 2013. 8(12).
62. Horschler, D.J. et al., Absolute brain size predicts dog breed differences in executive function. *Animal Cognition*, 2019. 22(2): p. 187-198.
63. Lequarre, A.S. et al., LUPA: A European initiative taking advantage of the canine genome architecture for unravelling complex disorders in both human and dogs. *Veterinary Journal*, 2011. 189(2): p. 155-159.
64. Tang, R.Q. et al., Candidate genes and functional noncoding variants identified in a canine model of obsessive-compulsive disorder. *Genome Biology*, 2014. 15(3).
65. Wielaender, F. et al., Generalized myoclonic epilepsy with photosensitivity in juvenile dogs caused by a defective DIRAS family GTPase 1. *Proceedings of the National Academy of Sciences of the United States of America*, 2017. 114(10): p. 2669-2674.
66. Hytonen, M.K. et al., In-frame deletion in canine PITRM1 is associated with a severe early-onset epilepsy, mitochondrial dysfunction and neurodegeneration. *Human Genetics*, 2021.
67. Raj, K. and U. Giger, Initial survey of PLA2G6 missense variant causing neuroaxonal dystrophy in Papillon dogs in North America and Europe. *Canine Medicine and Genetics*, 2020. 7(1): p. 1-4.
68. Kaukonen, M. et al., A missense variant in IFT122 associated with a canine model of retinitis pigmentosa. *Human Genetics*, 2021.
69. Awano, T. et al., Genome-wide association analysis reveals a SOD1 mutation in canine degenerative myelopathy that resembles amyotrophic lateral sclerosis. *Proceedings of the National Academy of Sciences of the United States of America*, 2009. 106(8): p. 2794-2799.
70. Ivansson, E.L. et al., Variants within the SP110 nuclear body protein modify risk of canine degenerative myelopathy. *Proceedings of the National Academy of Sciences of the United States of America*, 2016. 113(22): p. E3091-E3100.
71. Forman, O.P. et al., An Inversion Disrupting FAM134B Is Associated with Sensory Neuropathy in the Border Collie Dog Breed. *G3-Genes Genomes Genetics*, 2016. 6(9): p. 2687-2692.
72. Granger, N. et al., Charcot-Marie-Tooth type 4B2 demyelinating neuropathy in miniature Schnauzer dogs caused by a novel splicing SBF2 (MTMR13) genetic variant: a new spontaneous clinical model. *Peerj*, 2019. 7.
73. Letko, A. et al., A CNTNAP1 Missense Variant Is Associated with Canine Laryngeal Paralysis and Polyneuropathy. *Genes*, 2020. 11(12).
74. Das, R.G. et al., Genome-wide association study and whole-genome sequencing identify a deletion in LRIT3 associated with canine congenital stationary night blindness. *Scientific Reports*, 2019. 9.
75. Kyostila, K. et al., A Missense Change in the ATG4D Gene Links Aberrant Autophagy to a Neurodegenerative Vacuolar Storage Disease. *Plos Genetics*, 2015. 11(4).
76. Hitti-Malin, R.J. et al., A LINE-1 insertion situated in the promoter of IMPG2 is associated with autosomal recessive progressive retinal atrophy in Lhasa Apso dogs. *Bmc Genetics*, 2020. 21(1).

77. Winkler, P.A., H.D. Ramsey, and S.M. Petersen-Jones, A novel mutation in PDE6B in Spanish Water Dogs with early-onset progressive retinal atrophy. *Veterinary Ophthalmology*, 2020. 23(5): p. 792-796.
78. Mauri, N. et al., A SINE Insertion in ATP1B2 in Belgian Shepherd Dogs Affected by Spongy Degeneration with Cerebellar Ataxia (SDCA2). *G3-Genes Genomes Genetics*, 2017. 7(8): p. 2729-2737.
79. Forman, O.P., L. De Risio, and C.S. Mellersh, Missense Mutation in CAPN1 Is Associated with Spinocerebellar Ataxia in the Parson Russell Terrier Dog Breed. *Plos One*, 2013. 8(5).
80. Zeng, R. et al., A Truncated Retrotransposon Disrupts the GRM1 Coding Sequence in Coton de Tulear Dogs with Bandera's Neonatal Ataxia. *Journal of Veterinary Internal Medicine*, 2011. 25(2): p. 267-272.
81. Forman, O.P. et al., Spinocerebellar ataxia in the Italian Spinone dog is associated with an intronic GAA repeat expansion in ITPR1. *Mammalian Genome*, 2015. 26(1-2): p. 108-117.
82. Gilliam, D. et al., A Homozygous KCNJ10 Mutation in Jack Russell Terriers and Related Breeds with Spinocerebellar Ataxia with Myokymia, Seizures, or Both. *Journal of Veterinary Internal Medicine*, 2014. 28(3): p. 871-877.
83. Gast, A.C. et al., Genome-wide association study for hereditary ataxia in the Parson Russell Terrier and DNA-testing for ataxia-associated mutations in the Parson and Jack Russell Terrier. *BMC veterinary research*, 2016. 12(1): p. 1-7.
84. Van Poucke, M. et al., The novel homozygous KCNJ10 c. 986T> C (p.(Leu329Pro)) variant is pathogenic for the SeSAME/EAST homologue in Malinois dogs. *European Journal of Human Genetics*, 2017. 25(2): p. 222-226.
85. Agler, C. et al., Canine Hereditary Ataxia in Old English Sheepdogs and Gordon Setters Is Associated with a Defect in the Autophagy Gene Encoding RAB24. *Plos Genetics*, 2014. 10(2).
86. Kyostila, K. et al., A SEL1L Mutation Links a Canine Progressive Early-Onset Cerebellar Ataxia to the Endoplasmic Reticulum-Associated Protein Degradation (ERAD) Machinery. *Plos Genetics*, 2012. 8(6).
87. Fenn, J. et al., Genome sequencing reveals a splice donor site mutation in the SNX14 gene associated with a novel cerebellar cortical degeneration in the Hungarian Vizsla dog breed. *Bmc Genetics*, 2016. 17.
88. Forman, O.P. et al., Genome-wide mRNA sequencing of a single canine cerebellar cortical degeneration case leads to the identification of a disease associated SPTBN2 mutation. *Bmc Genetics*, 2012. 13.
89. Van Poucke, M. et al., Truncating SLC12A6 variants cause different clinical phenotypes in humans and dogs. *European Journal of Human Genetics*, 2019. 27(10): p. 1561-1568.
90. LeBlanc, A.K. and C.N. Mazcko, Improving human cancer therapy through the evaluation of pet dogs. *Nature Reviews Cancer*, 2020. 20(12): p. 727-742.
91. Karyadi, D.M. et al., A Copy Number Variant at the KITLG Locus Likely Confers Risk for Canine Squamous Cell Carcinoma of the Digit. *Plos Genetics*, 2013. 9(3).
92. Tonomura, N. et al., Genome-wide Association Study Identifies Shared Risk Loci Common to Two Malignancies in Golden Retrievers. *Plos Genetics*, 2015. 11(2).
93. Biasoli, D. et al., A synonymous germline variant in a gene encoding a cell adhesion molecule is associated with cutaneous mast cell tumour development in Labrador and Golden Retrievers. *Plos Genetics*, 2019. 15(3).

-
94. Melin, M. et al., Genome-Wide Analysis Identifies Germ-Line Risk Factors Associated with Canine Mammary Tumours. *Plos Genetics*, 2016. 12(5).
 95. Hedan, B. et al., Identification of common predisposing loci to hematopoietic cancers in four dog breeds. *Plos Genetics*, 2021. 17(4).
 96. White, M.E. et al., Consensus-based somatic variant-calling method correlates FBXW7 mutations with poor prognosis in canine B-cell lymphoma. *bioRxiv*, 2020.
 97. Sugkraroek, P. et al., Levels of Cholesterol and Phospholipids in Freshly Ejaculated Sperm and Percoll-Gradient-Pelleted Sperm from Fertile and Unexplained Infertile Men. *Fertility and Sterility*, 1991. 55(4): p. 820-827.
 98. Hoshi, K. et al., Variation in the Cholesterol Phospholipid Ratio in Human Spermatozoa and Its Relationship with Capacitation. *Human Reproduction*, 1990. 5(1): p. 71-74.
 99. Cross, N.L., Role of cholesterol in sperm capacitation. *Biology of Reproduction*, 1998. 59(1): p. 7-11.
 100. Benoff, S. et al., Fertilization Potential in-Vitro Is Correlated with Head-Specific Mannose Ligand-Receptor Expression, Acrosome Status and Membrane Cholesterol Content. *Human Reproduction*, 1993. 8(12): p. 2155-2166.
 101. Masaki, H. et al., Long-chain fatty acid triglyceride (TG) metabolism disorder impairs male fertility: a study using adipose triglyceride lipase deficient mice. *Molecular Human Reproduction*, 2017. 23(7): p. 452-460.
 102. Oborna, I. et al., Increased lipid peroxidation and abnormal fatty acid profiles in seminal and blood plasma of normozoospermic males from infertile couples. *Human Reproduction*, 2010. 25(2): p. 308-316.
 103. Aitken, R.J., J.S. Clarkson, and S. Fishel, Generation of Reactive Oxygen Species, Lipid-Peroxidation, and Human-Sperm Function. *Biology of Reproduction*, 1989. 41(1): p. 183-197.
 104. Aitken, R.J. et al., Causes and consequences of oxidative stress in spermatozoa. *Reproduction Fertility and Development*, 2016. 28(1-2): p. 1-10.
 105. Bromfield, E.G. et al., The impact of oxidative stress on chaperone-mediated human sperm-egg interaction. *Human Reproduction*, 2015. 30(11): p. 2597-2613.
 106. Bromfield, E.G. et al., Inhibition of arachidonate 15-lipoxygenase prevents 4-hydroxynonenal-induced protein damage in male germ cells. *Biology of Reproduction*, 2017. 96(3): p. 598-609.
 107. Walters, J.L.H. et al., Pharmacological inhibition of arachidonate 15-lipoxygenase protects human spermatozoa against oxidative stress. *Biology of Reproduction*, 2018. 98(6): p. 784-794.
 108. Walters, J.L.H. et al., Mechanistic Insight into the Regulation of Lipoxygenase-Driven Lipid Peroxidation Events in Human Spermatozoa and Their Impact on Male Fertility. *Antioxidants*, 2021. 10(1).
 109. Am-In, N. et al., Lipid profiles of sperm and seminal plasma from boars having normal or low sperm motility. *Theriogenology*, 2011. 75(5): p. 897-903.
 110. Zalata, A.A. et al., The fatty acid composition of phospholipids of spermatozoa from infertile patients. *Molecular Human Reproduction*, 1998. 4(2): p. 111-118.
 111. Gulaya, N.M. et al., Phospholipid composition of human sperm and seminal plasma in relation to sperm fertility. *Archives of Andrology*, 2001. 46(3): p. 169-175.

Chapter 2

A missense mutation in the *KLF7* gene is a potential candidate variant for congenital deafness in Australian Stumpy Tail Cattle Dogs

Fangzheng Xu[†], Shuwen Shan[†], Susan Sommerlad, Jennifer M. Seddon, Bertram Brenig^{*}

[†] These authors contributed equally to this work.

^{*}Corresponding author

The article was published online in *Genes*, 12(4), p.467. in March 2021. The full article can be found online at:

<https://doi.org/10.3390/genes12040467>

Author contributions:

Fangzheng Xu performed data analysis and wrote the original manuscript. He conducted whole genome re-sequence data analysis from raw fastq format to VCF files. GWAS was designed and performed by him. He also analyzed WGS data using recessive inherited model. Variant genotype information in 722 canine individuals was performed by him. He also conducted the variant effect prediction of the missense variants.

Abstract

Congenital deafness is prevalent among modern dog breeds, including Australian Stumpy Tail Cattle Dogs (ASCD). However, in ASCD, no causative gene has been identified so far. Therefore, we performed a genome-wide association study (GWAS) and whole genome sequencing (WGS) of affected and normal individuals. For GWAS, 3 bilateral deaf ASCDs, 43 herding dogs, and one unaffected ASCD were used, resulting in 13 significantly associated loci on 6 chromosomes, i.e., CFA3, 8, 17, 23, 28, and 37. CFA37 harbored a region with the most significant association ($-\log^{10}(9.54 \times 10^{-21}) = 20.02$) as well as 7 of the 13 associated loci. For whole genome sequencing, the same three affected ASCDs and one unaffected ASCD were used. The WGS data were compared with 722 canine controls and filtered for protein coding and non-synonymous variants, resulting in four missense variants present only in the affected dogs. Using effect prediction tools, two variants remained with predicted deleterious effects within the Heart development protein with EGF like domains 1 (*HEG1*) gene (NC_006615.3: g.28028412G>C; XP_022269716.1: p.His531Asp) and Kruppel-like factor 7 (*KLF7*) gene (NC_006619.3: g.15562684G>A; XP_022270984.1: p.Leu173Phe). Due to its function as a regulator in heart and vessel formation and cardiovascular development, *HEG1* was excluded as a candidate gene. On the other hand, *KLF7* plays a crucial role in the nervous system, is expressed in the otic placode, and is reported to be involved in inner ear development. 55 additional ASCD samples (28 deaf and 27 normal hearing dogs) were genotyped for the *KLF7* variant, and the variant remained significantly associated with deafness in ASCD ($p = 0.014$). Furthermore, 24 dogs with heterozygous or homozygous mutations were detected, including 18 deaf dogs. The penetrance was calculated to be 0.75, which is in agreement with previous reports. In conclusion, *KLF7* is a promising candidate gene causative for ASCD deafness.

Keywords: deafness; kruppel-like factor 7; genome wide association study; Australian stumpy tail cattle dog; brainstem auditory evoked response

Introduction

Deafness can cause several inconveniences for dogs (*Canis familiaris*, CFA), as more attention is required to avoid undetected danger. Deaf dogs are not suitable as working dogs because their training is more challenging than for normal hearing dogs. In addition, they are more likely to be startled and show more tendency to bite [1]. More than 100 modern dog breeds have been reported to be affected by congenital deafness [2]. Hence, deafness seems to be a common disorder among dogs, particularly in breeds such as the Dalmatian, Bull Terrier, English Setter, English Cocker Spaniel, and Australian Cattle Dog [3]. Hearing loss or deafness can be categorized mainly by five criteria in dogs: (1) Cause (genetic or nongenetic, inherited or acquired); (2) association with other diseases or phenotypes (syndromic or non-syndromic); (3) number of affected ears (unilateral or bilateral); (4) degree of loss (partial or total); and (5) site of pathology (peripheral or central) [4]. Peripheral deafness can also be classified as inherited or acquired, congenital or late onset, and sensorineural or conductive. In dogs, three classifications of deafness are commonly seen, including inherited congenital sensorineural, acquired later-onset sensorineural, and acquired later-onset conductive deafness [5].

In dogs, congenital sensorineural deafness is common, resulting in total deafness in young puppies that is either unilateral or bilateral. Sensorineural deafness results from dysfunction of cochlea or spiral ganglion. While it can be a degenerative process that relates to aging, noise trauma, exposure to therapeutic drugs that have ototoxic side effects, and chronic conditions [6], it is frequently inherited and so linked to one or more genetic mutations. Some morphological studies in dogs showed congenital sensorineural deafness manifested hypoplasia or aplasia of the sensory cells in the organ

of Corti, stria vascularis, macula sacculae, solidification, and calcification of tectorial membrane [7, 8]. Congenital sensorineural deafness is usually, but not always, related to pigmentation genes in some breeds [3].

Diagnosis of canine deafness usually consists of behavioral or electrodiagnostic testing. The behavioral testing is often unreliable, especially for the unilateral deafness or partial hearing impairment cases. The response of dogs may be affected by psychology (e.g., anxiety or loss of interest) and other senses (e.g., visual cues, vibration, or even air movement) [9]. The brainstem auditory evoked response (BAER) is the averaged record of the electrical activity of the auditory pathway in response to externally applied acoustic stimuli [10]. Compared with behavioral testing, the BAER test is an objective diagnostic method, with the advantages of being easy to record, noninvasive, safe, short test time, and giving reliable results [11].

The Australian Stumpy Tail Cattle Dog (ASCD) is a unique breed with a natural bob-tail, which should be distinguished from the Australian Cattle Dog breed. ASCD is alert, watchful and obedient, and talented in working and controlling cattle. It has been recognized as a standardized breed since 1988 by the Australian National Kennel Council. For a long time, general opinion held that the origins of the Australian Stumpy Tail Cattle Dog arose from European herding dogs and the Australian Dingo. However, recently it has been suggested that the ancestors of the Australian Stumpy Tail Cattle Dog and the Australian Cattle Dog, sharing a common origin, arrived in Australia with early free settlers, as their unidentified companions, between 1788 and c. 1800 (Clark, Noreen R. *A Dog for the Job*. (in prep. 2020)). Each pup should undergo a BAER test because this breed has a high deafness prevalence (<https://www.akc.org/dog-breeds/australian-stump-tail-cattle-dog/> (accessed on 24 March 2021)). A research study of 315 ASCDs showed the incidence of congenital sensorineural deafness was 17.8% [12]. There was no evidence that congenital sensorineural deafness in ASCD has

a left/right asymmetry or a sex-specific pattern, but there was a significant correlation between red (over blue) coat color and deafness [12].

No unique causative variants have been identified so far for any dog breeds, possibly in part due to the fact that deafness appears to be a comparatively heterogeneous disease as described above. In addition, there are several hypotheses about the inheritance pattern of congenital sensorineural deafness (reviewed by [1]). In Border Collies, for instance, Ubiquitin Specific Peptidase 31 (*USP31*) and RB Binding Protein 6 (*RBBP6*) have been associated with adult-onset deafness [13], whereas in the Doberman Pinscher, an insertion in Protein Tyrosine Phosphatase Receptor Type Q (*PTPRQ*) and a missense variant in Myosin VIIA (*MYO7A*) have been shown to be causative for a form of deafness that includes vestibular disease [14, 15]. Although chromosome 2 (CFA2), 6, 14, 17, 27, and 29 have been associated with hearing loss in Dalmatians, no causative variants have been identified so far [16].

In ASCD, congenital sensorineural deafness has been linked to a chromosomal region on CFA10 [12]. However, within a potential candidate gene Sry-related Hmg-box gene 10 (*SOX10*) located in this region, no causative alterations were detected. A recent genome-wide association study (GWAS) reported 14 chromosomes that were significantly associated with deafness in three canine breeds, and CFA3 was significantly associated with bilateral deafness in Australian Cattle Dogs [17]. In this study, three suggestive candidate genes near significantly associated regions were detected in these three dog breeds, including ATPase Na⁺/K⁺ Transporting Subunit α 4 (*ATP1A4*), Transformation/Transcription Domain Associated Protein (*TRRAP*), and Potassium Inwardly Rectifying Channel Subfamily J Member 10 (*KCNJ10*) [17]. However, none have been convincingly identified as causative mutations.

To extend the identification of potential candidate genes causing deafness in ASCD we performed a genome-wide association study and whole genome sequencing (WGS) in deaf ASCD. We identified a unique missense variant in Kruppel-like factor 7 (*KLF7*) gene significantly associated with deafness in ASCDs. This variant was absent in 722 dogs of bioproject PRJN448733 (see below). As *KLF7* plays an important role in the nervous system, is expressed in the inner ear, and seems to be involved in inner ear development [18, 19], it was a convincing candidate for ASCD deafness.

Materials and Methods

Ethical Statement

The collection of dog blood samples was done by S. Sommerlad at the time of BAER testing. The collection of samples was approved by the “Niedersächsisches Landesamt für Verbraucherschutz und Lebensmittelsicherheit” (33.19-42502-05-15A506) according to §8a Abs. 1 Nr. 2 of the TierSchG. All ASCDs were tested and sampled under approval of The University of Queensland’s Animal Ethics Committee.

Phenotyping and Samples

Fifty-nine Australian Stumpy Tail Cattle Dogs (Table S1) from a previous study [12] were used in this study. BAER testing was performed on 59 dogs [20], 28 were normal hearing dogs and 31 were diagnosed as deaf, of which 10 were bilateral deaf, 12 were left-sided deaf, and 9 were right-sided deaf (Table S1). Three bilaterally deaf ASCDs (#217, #253 and #330), and one control dog with normal hearing (#326) were used for next generation sequencing. Dog #326 was a littermate of #330. These four dogs were female and red in color; all but #330 had a speckled coat. DNA was extracted using a salting-out method as described [12]. All samples were pseudonymized using internal IDs. Furthermore, data from two repository were used in this study. One repository contain Variant Call Format (VCF) data of 722 canine individuals

(<https://www.ncbi.nlm.nih.gov/bioproject/PRJNA448733> (accessed on 24 March 2021)) [21]. It consists of 144 established breeds, 11 samples with mixed breed, 26 samples with unknown breed status, 104 village and feral dogs from different regions, and 54 wild canids from six species. An additional dataset consisted of 590 samples including 582 dogs from 126 breeds and 8 wolves (<https://www.ebi.ac.uk/ena/data/view/PRJEB32865> (accessed on 24 March 2021)) [22].

Next Generation Sequencing and Variant Calling

A total of 1.0 µg DNA per ASCD sample was used as input material for the DNA library preparations. Sequencing libraries were generated using NEBNext[®] DNA Library Prep Kit following manufacturer's recommendations and indices were added to each sample. The genomic DNA was randomly fragmented to a size of 350bp by shearing, then DNA fragments were end polished, A-tailed, and ligated with the NEBNext adapter for Illumina sequencing, and further PCR enriched by P5 and indexed P7 oligos. The PCR products were purified (AMPure XP system) and resulting libraries were analyzed for size distribution by Agilent 2100 Bioanalyzer and quantified using real-time PCR. For #217, #253, #326, #330, a total of 599,770,692, 723,624,660, 743,641,356, 620,101,998 raw reads were obtained, respectively. Corresponding coverages were around 40× (paired-end reads, 2 × 150 bp).

Raw sequence data were aligned to dog genome CanFam3.1 using BWA 0.7.17 [23]. SAMtools 1.9 were used for format change and sorting of sequences [24]. Duplicates were marked by PICARD (<http://broadinstitute.github.io/picard/> (accessed on 24 March 2021)). Variant calling was performed by GATK 4.1.3 with best practice pipeline [25].

Genome Wide Association Analysis

We used the VCF data obtained in the previous step for GWAS analysis. Three deaf dogs (#217, #253, #330) were used as cases. As ASCD is utilized for control and herding of cattle according to its breed standard (<http://www.fci.be/Nomenclature/Standards/351g01-en.pdf> (accessed on 24 March 2021)), VCFs of 43 herding dogs from 15 breeds (Australian Cattle Dog, Bearded Collie, Belgian Malinois, Belgian Sheepdog, Belgian Tervuren, Berger Blanc Suisse, Berger Picard, Border Collie, Bouvier des Flandres, Entlebucher Sennenhund, Finnish Lapphund, German Shepherd Dog, Pembroke Welsh Corgi, Shetland Sheepdog, Spanish Water Dog) were extracted from the publicly available 722 canine VCF repository (<https://www.ncbi.nlm.nih.gov/bioproject/PRJNA448733> (accessed on 24 March 2021)) [21]. Sample selection criteria were the same as described [21]. A total of 43 herding dogs and the normal hearing dog #326 were chosen as controls (Table S2). The VCF files of the 43 herding dogs and 4 ASCDs were merged by BCFtools 1.9 [24]. Filtering was done using VCFtools 0.1.13 with options `--max-alleles 2, --min-alleles 2, --min-meanDP 20, --minQ 20, --minGQ 20, --remove-indels, --max-missing 0.95, --maf 0.05, --hwe 0.001` [26]. After filtering, 857,343 variants remained and were further pruned by Linkage Disequilibrium with `-indep 1000 3 1` function in PLINK 1.90 [27]. The final data set consisted of 20,656 SNPs. Principal component analysis (PCA) was performed using EIGENSOFT package [28]. GEMMA 0.98 was used for association analysis by case–control setting (3 deaf cases vs. one normal hearing ASCD and 43 herding dogs as controls) [29]. A univariate linear mixed model with sex, 5 principal components, and relatedness of 47 dog individuals for corrections was applied for the association test. Bonferroni threshold $-\log_{10}P (0.01/20,656) = 6.32$ was utilized. Qqman package was used to generate Manhattan and quantile–quantile (QQ) plots [30]. The genomic inflation factor lambda was calculated with formula $\lambda = \text{median}(qchisq(1-p, 1))/qchisq(0.5, 1)$ where p is a vector of p values.

Next Generation Sequencing Data Analysis for Identification of Associated Variants

Data after variant calling were analyzed with SNP & Variation Suite 8.8.3 (Golden Helix Inc., Bozeman, MT, USA). SNPs and indels were set to missing with read depth ≤ 10 , genotype quality ≤ 15 , alt read ratios for Ref_Ref ≥ 0.15 , Ref_Alt outside 0.3 to 0.7, Alt_Alt ≤ 0.85 . Variants were analyzed using autosomal recessive and dominant models, respectively. In the autosomal recessive filtering model, 3 deaf ASCDs were set as Alt_Alt, control ASCD as Ref_Ref or Alt_Ref. In the autosomal dominant filtering model, the 3 deaf ASCDs were set as Alt_Alt or Alt_Ref and controls as Ref_Ref. To further narrow the range of candidate variants, we compared the common variants of deaf ASCDs with 722 canine genomes to identify private variants. The shared variants in the three deaf dogs were filtered by BCFtools 1.9 with 'isec' option. Private variants were annotated using SnpEFF software [31] to determine high (loss of function) and moderate (missense) impact variants (Ensembl transcripts release 101). These functional variants were further checked by Integrative Genome Viewer (IGV) software to obtain real high quality variants [32]. Variant effects were predicted by SIFT [33], PolyPhen-2 [34], and PROVEAN [35].

Genotyping of *KLF7* Variant in Australian Stumpy Tail Cattle Dogs

Targeted genotyping of the *KLF7* missense variant was performed in 59 ASCDs by PCR amplification using primers cfa_KLF7_Ex3_F (5'-AGACTCTCTCAGCCGTGGAT-3') and cfa_KLF7_Ex3_R (5'-GGCCAACTTGTACCACTACCT-3'), resulting in a 295 bp fragment. Genotyping of PCR products were implemented by RFLP analysis after cleavage with the restriction enzyme HinP1I (NEB). The wild type allele was cleaved into two fragments, 236 bp and 59 bp, while the homozygous mutant remained uncut. Frequency distribution for alleles and genotypes was calculated using Fisher's Exact Test in these 59 ASCDs. Allelic and genotypic odds ratios were calculated according to [36].

Investigation of Human Deafness Genes in 3 Deaf Australian Stumpy Tail Cattle Dogs

Human hearing loss or deafness genes were queried using online software GLAD4U with “hearing loss” and/or “deafness” as keywords [37]. After combining the three query results, 346 genes were chosen for further analysis (Table S3). The variants of these gene regions (including 1000 bp up- and downstream regions) were extracted by BCFtools from VCF files of the three deaf ASCDs and annotated by SnpEFF software. Variants with high (loss of function) and moderate (missense) impacts were selected for further analysis (Ensembl transcripts release 101). The genotype information of the chosen variants was further checked in 722 canines.

Results

Genome Wide Association Analysis

The analysis was done using three bilateral deaf female dogs from three different litters. The hearing status of the individuals determined using BAER is shown in Table 1 and Table S1.

Table 1. BAER (brainstem auditory evoked response) results of 4 Australian Stumpy Tail Cattle Dogs (ASCDs).

ID	Gender	Coat Colour	BAER Test Results
217	Female	Red speckled	Bilaterally Deaf
253	Female	Red speckled	Bilaterally Deaf
330	Female	Red	Bilaterally Deaf
326	Female	Red speckled	Normal Hearing

The three affected ASCDs were compared with 44 control dogs. 13 SNPs on 6 chromosomes (CFA3, 8, 17, 23, 28, 37) above the Bonferroni significance level were identified. The QQ-plot indicated that some associations might be due to population substructure. Associated SNPs are shown in Figure 1 and summarized in Table 2. The majority of the significantly associated SNPs (7/13) were located on CFA37 including

SNP chr37:44793 (position according to CanFam3.1) with the highest $-\log_{10}p$ -value = 20.02. A search for large structural variants (SVs) flanking the significantly associated regions on CFA3, 8, 17, 23, 28, and 37 using IGV was unsuccessful.

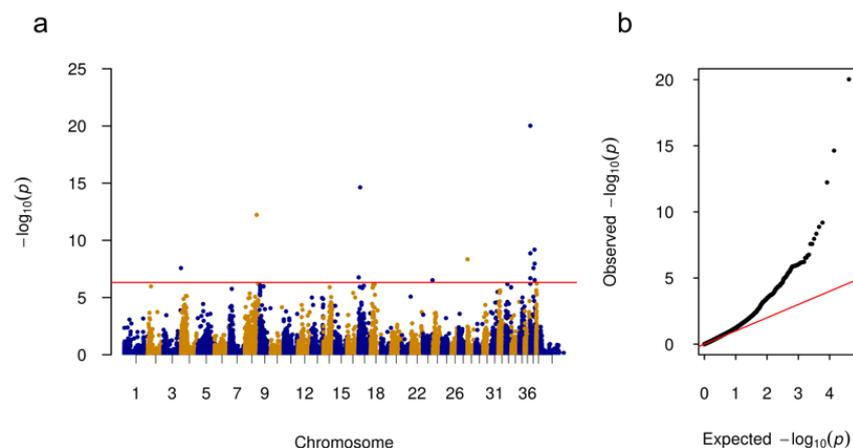


Figure 1. Manhattan and quantile–quantile (QQ) plots illustrating deafness associated chromosomal regions. **(a)** The Manhattan plot shows on the y -axis the negative log-base-10 of the p value for each of the polymorphisms in the genome (along the x -axis), when tested for differences in frequency between 3 bilateral deaf dogs (cases) and 44 controls ($1 \times$ ASCD, 43 herding dogs of 15 dog breeds). The red line indicates the Bonferroni significance threshold ($-\log_{10}(0.01/20,656) = 6.32$). **(b)** The QQ plot depicts the distribution of p -values of the genome-wide association study (GWAS) analysis and genomic inflation factor lambda is 1.20.

Table 2. Significantly associated SNPs above Bonferroni significance threshold (6.32).

CFA	Position	p -Value	Nearby Genes	Distance (bp)
3	90,987,932	2.67×10^{-8}	<i>LCORL</i>	186,575
8	62,032,863	5.93×10^{-13}	<i>DGLUCY</i>	19,884
17	1,977,343	1.73×10^{-7}	<i>EIPR1</i>	0
17	9,456,133	2.34×10^{-15}	<i>TRIB2</i>	204,307
23	50,096,314	3.04×10^{-7}	<i>KCNAB1</i>	0
28	21,516	4.50×10^{-9}	<i>PTPN20</i>	42,882
37	13,393	2.04×10^{-7}	<i>WDR75</i>	144,007
37	44,793	9.54×10^{-21}	<i>WDR75</i>	112,607
37	80,438	1.36×10^{-9}	<i>WDR75</i>	76,962
37	16,399,127	2.66×10^{-8}	<i>CRYGD</i>	25,757
37	22,102,392	6.48×10^{-10}	<i>ABCA12</i>	34,340
37	22,579,983	2.93×10^{-7}	<i>FNI</i>	57,573
37	22,711,697	1.10×10^{-8}	<i>FNI</i>	189,287

Whole Genome Sequencing Reveals Four Potential Variants

To further locate the candidate variants, next generation sequencing was performed in 3 deaf ASCDs (#217, #253, #330) and 1 normal hearing ASCD (#326). After quality control, a total of 4,208,002 SNPs and 2,298,760 indels were detected. According to previous deafness studies, sequence data were initially analyzed using a recessive model of inheritance. Using this model, 129,383 SNPs and 51,942 indels were detected. Using only variants that had been annotated and verified as mRNA transcripts (Ensembl release 101), 338 SNPs and 523 indels remained (Table S4). After filtering these variants against the 722 dog database, none of the homozygous Alt_Alt genotypes were exclusively present in the deaf ASCDs (Table S4). As there were no reports about such a high prevalence of deafness in the 722 control dogs and it can be assumed that the majority of the controls were hearing, these variants were presumably not causative.

As no associated variants were found using the recessive inheritance model, a dominant inheritance model was applied. In this analysis, private variants only present in the three deaf ASCDs (Alt_Alt and Alt_Ref) compared to the 722 controls (Ref_Ref) were filtered, resulting in 270,980 SNPs and 351,927 indels. After quality control and functional annotating, 167 protein-changing variants (58 SNPs and 109 indels) remained (Table S5). These variants were further filtered against #326 (normal hearing littermate of #330) assuming that this dog should be homozygous wild type under the supposed model. After this step, four missense variants remained as potential causative candidates (Table 3). Within the 722 control dogs, no homozygous Alt_Alt or heterozygous carriers were detected for these 4 missense variants. In an additional dataset consisting of 590 dog samples, only two heterozygous individuals (Brussels Griffon dogs) were determined for the Microtubule associated protein 6 (*MAP6*) gene variant. To deduce which of the variants could be causative for deafness, protein function prediction tools were used. As shown in Table 4 only the variants in Heart

development protein with EGF like domains 1 (*HEG1*) and *KLF7* were predicted to be deleterious by at least two of the prediction tools.

Table 3. Genotype information of four potential causative variants for ASCD deafness.

Chr	HGVS Genome Position ^(a)	Variant Type	Gene ^(b)	#217	#253	#330	#326
13	NC_006595.3:g.60805542 C>T	missense variant	<i>GC</i>	C_T	C_T	C_T	C_C
21	NC_006603.3:g.23019999 C>T	missense variant	<i>MAP6</i>	C_T	C_T	C_T	C_C
33	NC_006615.3:g.28028412 G>C	missense variant	<i>HEG1</i>	G_C	G_C	G_C	G_G
37	NC_006619.3:g.15562684 G>A	missense variant	<i>KLF7</i>	A_A	A_A	A_G	G_G

^(a) Positions according to CanFam3.1; ^(b) *GC*: GC vitamin D binding protein, *MAP6*: Microtubule associated protein 6, *HEG1*: Heart development protein with EGF like domains 1, *KLF7*: Kruppel-like factor 7.

Table 4. Variant effect predicted by SIFT, PolyPhen-2, and PROVEAN.

Gene	Amino acid exchange	SIFT	Polyphen-2	PROVEAN
<i>GC</i>	p.Gly389Rrg	Tolerated	Benign	Neutral
<i>MAP6</i>	p.Arg486Cys	Affect protein function	Benign	Neutral
<i>HEG1</i>	p.His531Asp	Affect protein function	Unknown	Deleterious
<i>KLF7</i>	p.Leu173Phe	Affect protein function	Possibly damaging	Neutral

SIFT: <https://sift.bii.a-star.edu.sg> (accessed on 24 March 2021), Polyphen-2: <http://genetics.bwh.harvard.edu/pph2/index.shtml> (accessed on 24 March 2021), PROVEAN: <http://provean.jcvi.org/index.php> (accessed on 24 March 2021).

To further confirm the causative possibilities of the two remaining variants, their amino acid conservation was analyzed in the same 7 species. The missense variant in *HEG1* gene (NC_006615.3: g.28028412G>C) resulted in an amino acid exchange of p.His531Asp (XP_022269716.1). In *KLF7* gene (NC_006619.3: g.15562684G>A), the variant led to an exchange of p.Leu173Phe (XP_022270984.1). Especially in *KLF7*, the amino acid position seems to be highly conserved across several different species, as shown in Figure 2.

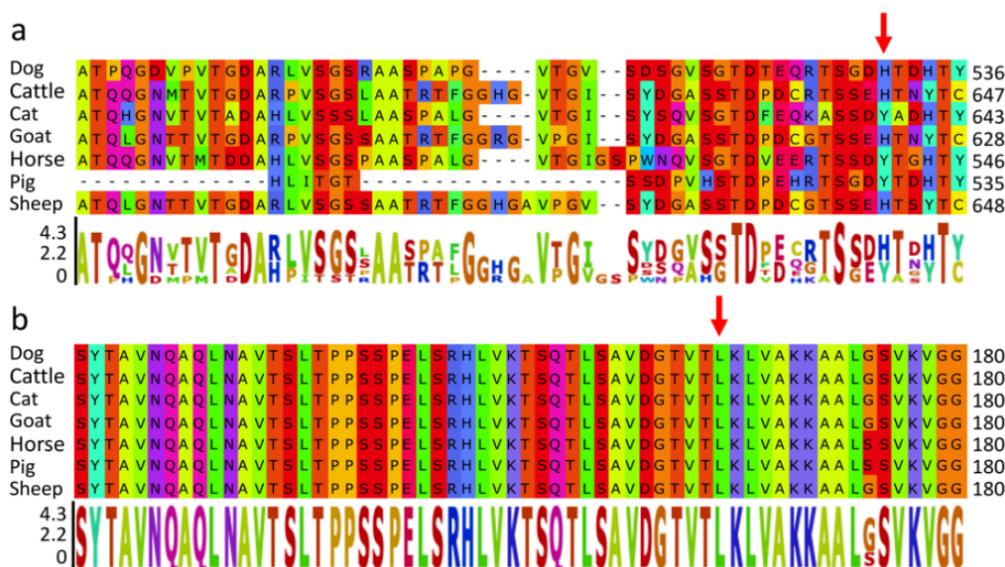


Figure 2. Cross-species comparison of variant amino acid positions in HEG1 and KLF7. Partial protein sequences of HEG1 (**A**) and KLF7 (**B**) flanking the variant amino acid positions were aligned using ClustalW (<https://www.ebi.ac.uk/Tools/msa/clustalo/> (accessed on 24 March 2021)). The variant positions are highlighted with a red arrow. Residual color scheme was referred from [38], sequence logos are shown according to [39].

Genotyping of *KLF7* Variant in ASCDs

To verify the association of the *KLF7* variant with ASCD congenital deafness, 27 normal hearing and 28 deaf ASCDs (21 unilaterally and 7 bilaterally deaf dogs) were used to investigate the *KLF7* variant genotype distribution. As summarized in Table 5, 59 ASCDs including the 4 whole genome sequenced dogs were used to check the association of the *KLF7* missense variant with ASCD deafness. Four dogs were homozygous carriers (A_A) and 14 heterozygous (A_G) among the 31 deaf ASCDs. Within the 28 normal hearing ASCDs, 5 heterozygous and one homozygous carrier were detected. The penetrance of ASCD deafness was calculated to be 0.75. As determined by Fisher's exact test, homozygosity for the *KLF7* variant was significantly associated with congenital deafness ($p = 0.014$). The odds ratio_{AA} = 6.8 (95% CI [0.68, 67.25]), i.e., homozygous carriers are 6.8 times more likely to be deaf than wild type.

Table 5. Genotype distribution of *KLF7* variant in 31 deaf and 28 normal hearing ASCDs dogs.

Phenotype	G_G	A_G	A_A	Total Number	P ^(c)
Unilaterally deaf	10	10	1	21	0.054
Bilaterally deaf	3	4	3	10	0.010
Deafness (uni ^(a) + bi ^(b))	13	14	4	31	0.014
Normal hearing	22	5	1	28	

^(a) uni: Unilaterally deaf; ^(b) bi: Bilaterally deaf; ^(c) *p*-value using Fisher's exact test.

Discussion

Deafness is a common disorder among dogs, and the observed prevalence is highest in Dalmatians (29.9%) [3] and 17.8% in ASCD [12]. Even selective breeding based on deafness phenotyping decreased the prevalence in Dalmatians only to 17.8% [40]. Several other dog breeds also show rather high prevalence rates (> 10%), e.g., Australian Cattle Dog and Bull Terrier [3]. To accelerate the decline of overall prevalence of congenital sensorineural deafness, it would be important to identify the genetic cause of the disorder to enable informed breeding.

We used four ASCD DNA samples from a previous study of deafness in Australian Stumpy Tail Cattle Dogs for GWAS and WGS analysis. The previous study used a genome screen with 325 microsatellite (290 were used for linkage mapping) to determine a significantly linked deafness region on CFA10 [12]. However, *SOX10*, the only potential candidate gene in this region, had to be excluded, as it did not harbor any causative variants. Another promising candidate in the CFA10 region, i.e., Trio- and f-actin-binding protein (*TRIOBP*), had also to be excluded. In the above mentioned study, deafness was reported to be autosomal recessive inherited with incomplete penetrance [12]. As shown before, GWAS with multiple breeds can improve the accuracy of causative variant mapping [41, 42]. Our analysis provided evidence for at least six highly associated chromosomal regions. However, due to the small number of affected dogs, some associated regions might have resulted from the close relationship of the

dogs. This can be seen in the QQ-plot which showed convincing evidence for an association with some indication of a population substructure. In our study, more than half of the significant associated SNPs (7 out of 13) were located on CFA37, including the most significantly related SNP (chr37:44793, $p = 9.54 \times 10^{-21}$). In a recent study of Dalmatian deafness, signals were also detected in this region [17]. However, there was no associated peak on CFA37 reported in the previous microsatellite-based study in ASCD. A possible explanation could be that there were only five microsatellite markers on CFA37, one of which had a low degree of polymorphism (3 alleles, PIC 0.5) [12]. This might have been insufficient to detect associations on this chromosome. An alternative explanation is that ASCD deafness may be heterogeneous. There may be more than one variant causing congenital deafness in this breed, and using limited family associations may reveal private mutations. Further genotyping analysis in a wider range of affected (28) and unaffected (27) ASCDs revealed that the *KLF7* missense variant was still significantly associated with congenital deafness (Table 5). Furthermore, the penetrance of deafness in ASCD calculated based on the *KLF7* variant was 0.75, which was in agreement with the previously calculated penetrance of 0.72 [12]. Altered allele (A) frequency is 24.58% (Table 5). If we take penetrance into consideration, the deafness frequency is $(24.58\% * 0.75) = 18.4\%$, which is also close to the previous investigation of 17.8% overall ASCD breed deafness frequency [12]. Several homozygous wild type individuals were detected among the deaf ASCDs suggesting additional genetic risk factors. This was not surprising, as canine congenital deafness seems to be a complex disorder and different regions were detected in other GWASs for deafness so far [17].

According to our GWAS, functional relationships with deafness of genes near the significantly associated loci on most chromosomes were unapparent (Table 2). Only the region on CFA37 was further supported by WGS. In the initial GWAS 651 variants on chromosome 37 (between CFA37:7217 to CFA37:30803691) were identified

(Figure 1). Variant CFA37:15503029 T>C with a p -value of 8.61×10^{-6} was only 12,534 bp distant from *KLF7*. To evaluate LD over-pruning and potential effects on resolution, we repeated the GWAS using less stringent pruning parameters (--indep 1000 5 4). This increased the number of associated variants to 60,746. In agreement with the previous analysis, a variant with $-\log_{10}p\text{-value} = 14.68$ at position CFA37:15463045 remained in the vicinity of *KLF7* (Table S6) and a significantly associated region spanning from CFA37:15463045 to CFA37:16433709 was detected harboring *KLF7* (CFA37:15515563-15607345). As expected, a further reduction of pruning stringency resulted in more chromosomal regions above the significant threshold (Figure S1). However, especially on CFA10, no significantly associated variants were identified.

In addition, we applied whole genome sequencing of the deaf dogs and used a large number of available canine whole genome sequence data as controls to improve the accuracy and efficiency of causative variant identification. Several GWAS of canine complex hereditary deafness failed to identify causative variants with the exception of two associated genes (*MYO7A*, *PTPRQ*) causative for a specific form of canine congenital bilateral deafness with vestibular disease [14, 15].

For next generation sequence analysis in the present study, functional variants within coding regions were primarily considered due to their direct impact on protein function [43]. We filtered all variants using an autosomal recessive model, however, no functional variants fulfilled this mode of inheritance. Again, the chromosomal region 1Mb up- and downstream of *SOX10* (CFA10:25680441-27690530) was checked using IGV, but no deafness associated variants including larger structural variants were identified. After WGS analysis and variant effect prediction, only two missense variants within *HEG1* and *KLF7* remained. *HEG1* is involved in cardiovascular development [44] and therefore seemed unlikely to be involved in the development of deafness.

However, the candidate variant (NC_006619.3: g.15562684G>A) in *KLF7* (CFA37:15515563-15607345) was close to the significantly associated SNP CFA37:16399127 ($p = 2.66 \times 10^{-8}$) (Table 2). *KLF7* is a zinc finger transcription factor and has been reported to play a role in the nervous system and is vital for neuronal morphogenesis that could function in axon outgrowth [18]. *KLF7* was suggested to have potential functions in neurogenesis of mice, like neuronal differentiation and maturation [45]. *KLF7* was also found to promote axon regeneration [46]. Furthermore, *KLF7* is required for the development of sensory neurons [47], and it has been reported to play roles in neurotransmission and synaptic vesicle trafficking [48]. These two processes have important influences on the auditory system, and therefore disruption of *KLF7* could lead to hearing impairment and dysfunction [49]. Indeed, *KLF7* was confirmed to be expressed in the otic placode which will develop into ears, indicating *KLF7* could have an effect on ear development [19]. *KLF7* was also detected to be a fibroblast growth factor (FGF) responsive factor in ear progenitor induction processes, which implies it may be involved in early ear induction [50]. *KLF7* has been considered as one high quality candidate gene for human branchio-oto-renal syndrome, which is an autosomal dominant disease with hearing loss as one clinical sign [51]. *KLF7* was the nearby gene (50,519 bp distance) of one significant signal in adult hearing difficulty GWAS [52]. One recent GWAS of hearing-related traits with up to 330,759 individuals (UK Biobank) revealed 31 significant genomic risk loci for adult hearing difficulty, *KLF7* was also detected to be significantly associated [53]. Furthermore, the protein sequence segments surrounding *KLF7* variant are much more conserved than that of HEG1 among the same 7 species (Figure 2). Recently, *KLF7* has been reported to directly regulate GATA Binding Protein 3 (*GATA3*) expression [54]. *GATA3* is expressed in the otic placode and is involved in inner ear development [55]. Though the interaction between *KLF7* and *GATA3* was reported in chicken adipogenesis, *KLF7* is quite conserved among several species (Figure 2). Knockdown of Paired Box Protein Pax-2 (*PAX2*) (inner ear development gene) led to a significant up-regulation of both

KLF7 and *GATA3* expression [19], which implies *KLF7* and *GATA3* are probably involved in the same pathway. Furthermore, *GATA3* is the causative gene for human hypoparathyroidism, deafness, and renal dysplasia (HDR) syndrome [56]. Therefore, *KLF7* could interact with *GATA3* during the development of inner ear, and defects in *KLF7* could affect *GATA3* normal expression patterns in otic placode. This may be a potential cause of hearing loss in ASCD cases. The incomplete penetrance presented by the *KLF7* variant in deafness may be related to its role as a transcription factor that is involved in a specific part of the hearing pathway. Our findings could provide clues for the functional analysis of the *KLF7* in inner ear development. Functional analysis of *KLF7* regarding ear development may provide further evidence for its role in deafness. Another intriguing possible pathway is suggested by the finding of a KLF binding site upstream of the M promoter of Microphthalmia-associated transcription factor (*M-MITF*) that induces gene expression changes in humans [57]. Although the aforementioned study was related to melanoma development, *M-MITF* has been identified as the locus responsible for white coat patterning in dogs [58]. Hereditary deafness has been reported to be associated with white pigmentation in several species, e.g., by affecting *M-MITF* isoform expression in pigs [59] and cows [60] as well as humans [61]. Canine deafness was also linked with white pigmentation due to the merle and piebald locus [62]. Congenital sensorineural deafness of English Bull Terrier is predominant in individuals with white coat color [63]. Similarly, congenital hereditary sensorineural deafness in the Australian Cattle Dog was negatively associated with bilateral facial masks, also individuals with pigmented body patches showed a lower risk of deafness [64]. An inverse association of pigmented head patches and congenital sensorineural deafness was also observed in Dalmatians, while on the other hand, a positive correlation was detected with blue irises [65-71]. In ASCD, congenital sensorineural deafness was moderately significant associated with red/blue coat color, but not with speckling and facial masks [12]. However, no functional alterations in genes related to coat color or pigmentation were detected after filtering for case-control

setting in the present study. Thus far, no causative variants within genes involved in pigmentation have been identified in canine deafness. Some pigmentation genes have actually been excluded as candidates in different dog breeds, e.g., c-Kit (*KIT*) and melanocyte protein 17 (*SILV*) [72, 73]. An alternative explanation is that deafness caused by dysfunctions of other biological processes may be more common, such as ear development and morphogenesis. This is highly relevant in the Gene Ontology (GO) category analysis of potential canine hereditary deafness genes [2]. In our study, *KLF7* was reported to participate in inner ear development processes [50]. There is good evidence here that the *KLF7* variant contributes to deafness, but the genotyping data supports the view that this is a multigene/multifactorial disease, and so this is one contributing mutation.

Conclusions

In summary, a missense variant within *KLF7* gene has been identified to be significantly associated with congenital deafness in Australian Stumpy Tail Cattle Dogs. As *KLF7* gene was reported to be expressed in the inner ear and associated with human hearing difficulties, our findings could provide clues for further elucidating novel genetic causes for human hearing loss.

Supplementary Materials: The following are available online at <https://www.mdpi.com/2073-4425/12/4/467/s1>.

Author Contributions: F.X.: Investigation, visualization, writing-original draft. S.S. (Shuwen Shan): Investigation, visualization, writing-original draft. S.S. (Susan Sommerlad): Resources, writing-review and editing. J.M.S.: Resources, writing-review and editing. B.B.: Conceptualization, resources, project administration, funding

acquisition, supervision, writing-original draft, review and editing. All authors have read and agreed to the published version of the manuscript.

Data Availability Statement: The data presented in this study are openly available in Center for Open Science (OSF) reference number <https://osf.io/z36ap> (accessed on 24 March 2021).

Acknowledgments: S. Shan and F. Xu were supported by a fellowship of the China Scholarship Council (CSC). The authors would like to thank Dogs Queensland and the participating Australian Stumpy Tail Cattle Dog breeders of Queensland and New South Wales Australia, for their cooperation.

Conflicts of Interest: The authors declare no conflict of interest.

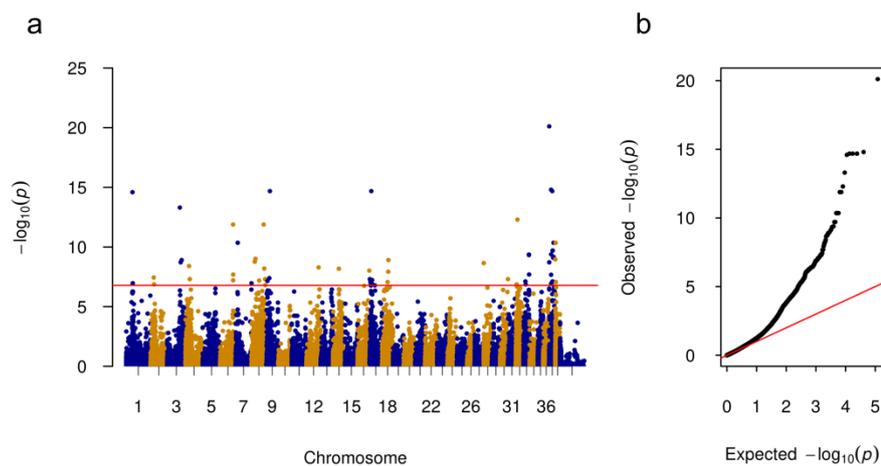


Figure S1. Manhattan and QQ plots of the Genome Wide Association Analysis (GWAS) for ASCD deafness. **(a)** Association of 60,746 variants with bilateral deafness in 47 dogs (3 cases and 44 controls). The plot shows the $-\log_{10}p$ -values for all variants. 60,746 were remained after pruning by Linkage Disequilibrium with parameters --indep 1000 5 4. The red horizontal line represents the Bonferroni genome-wide significance threshold of $-\log_{10}(0.01/60,746) = 6.78$. **(b)** Quantile-quantile (QQ) plot showed the observed $-\log_{10}p$ -values in the black curve, the red line indicated the distribution of expected $-\log_{10}p$ -values. λ is calculated to be 1.10.

Table S1. *KLF7* variant genotypes of 59 ASCDs.

Group	Sample	Deafness	Genotype of NC_006619.3:g.15562684 G>A
Normal	Sue143	normal	GG
	Sue265	normal	GG
	Sue75	normal	GG
	Sue137	normal	GG
	Sue264	normal	GG
	Sue144	normal	GG
	Sue266	normal	GG
	Sue426	normal	AG
	Sue145	normal	AG
	Sue194	normal	GG
	Sue286	normal	GG
	Sue287	normal	GG
	Sue211	normal	GG
	Sue326	normal	GG
	Sue288	normal	GG
	Sue113	normal	GG
	Sue116	normal	GG

	Sue251	normal	AA
	Sue317	normal	AG
	Sue393	normal	GG
	Sue433	normal	GG
	Sue252	normal	GG
	Sue318	normal	AG
	Sue395	normal	GG
	Sue434	normal	AG
	Sue139	normal	GG
	Sue394	normal	GG
	Sue435	normal	GG
Deaf L ear	Sue331	deaf L ear	AG
	Sue339	deaf L ear	AG
	Sue346	deaf L ear	AG
	Sue428	deaf L ear	AG
	Sue429	deaf L ear	GG
	Sue111	deaf L ear	AG
	Sue200	deaf L ear	GG
	Sue373	deaf L ear	GG
	Sue375	deaf L ear	GG
	Sue431	deaf L ear	AG
	Sue316	deaf L ear	AG
	Sue130	deaf L ear	GG
Deaf R ear	Sue332	deaf R ear	AG
	Sue424	deaf R ear	GG
	Sue284	deaf R ear	GG
	Sue427	deaf R ear	AG
	Sue81	deaf R ear	GG
	Sue106	deaf R ear	AA
	Sue112	deaf R ear	AG
	Sue342	deaf R ear	GG
	Sue380	deaf R ear	GG
Bilat deaf	Sue78	bilat deaf	AA
	Sue425	bilat deaf	GG
	Sue187	bilat deaf	AG
	Sue285	bilat deaf	AG
	Sue343	bilat deaf	GG
	Sue430	bilat deaf	AG
	Sue217	bilat deaf	AA
	Sue253	bilat deaf	AA
	Sue330	bilat deaf	AG
	Sue432	bilat deaf	GG

Reference

1. Rak, S.G. and O. Distl, Congenital sensorineural deafness in dogs: a molecular genetic approach toward unravelling the responsible genes. *The Veterinary Journal*, 2005. 169(2): p. 188-196.
2. Kelly-Smith, M. and G.M. Strain, STRING Data Mining of GWAS Data in Canine Hereditary Pigment-Associated Deafness. *Veterinary and Animal Science*, 2020: p. 100118.
3. Strain, G.M., Deafness prevalence and pigmentation and gender associations in dog breeds at risk. *The Veterinary Journal*, 2004. 167(1): p. 23-32.
4. Strain, G.M., The genetics of deafness in domestic animals. *Frontiers in veterinary science*, 2015. 2: p. 29.
5. Strain, G.M., Aetiology, prevalence and diagnosis of deafness in dogs and cats. *British Veterinary Journal*, 1996. 152(1): p. 17-36.
6. Cunningham, L.L. and D.L. Tucci, Hearing loss in adults. *New England Journal of Medicine*, 2017. 377(25): p. 2465-2473.
7. Hiraide, F. and M.M. Paparella, Histopathology of the temporal bones of deaf dogs. *Auris Nasus Larynx*, 1988. 15(2): p. 97-104.
8. Coppens, A., A. Resibois, and L. Poncelet, Bilateral deafness in a maltese terrier and a great pyrenean puppy: inner ear morphology. *Journal of Comparative Pathology*, 2000. 122(2-3): p. 223-228.
9. Strain, G.M., Hereditary deafness in dogs and cats: causes, prevalence, and current research. *strain*, 2003. 225: p. 578-9758.
10. Sims, M.H., Electrodiagnostic evaluation of auditory function. *Veterinary Clinics of North America: Small Animal Practice*, 1988. 18(4): p. 913-944.
11. Wilson, W.J. and P.C. Mills, Brainstem auditory-evoked response in dogs. *American journal of veterinary research*, 2005. 66(12): p. 2177-2187.
12. Sommerlad, S. et al., Congenital sensorineural deafness in Australian stumpy-tail cattle dogs is an autosomal recessive trait that maps to CFA10. *PLoS One*, 2010. 5(10).
13. Yokoyama, J.S. et al., Variation in genes related to cochlear biology is strongly associated with adult-onset deafness in border collies. *PLoS genetics*, 2012. 8(9).
14. Guevar, J. et al., Deafness and vestibular dysfunction in a Doberman Pinscher puppy associated with a mutation in the PTPRQ gene. *Journal of veterinary internal medicine*, 2018. 32(2): p. 665-669.
15. Webb, A.A., A.L. Ruhe, and M.W. Neff, A missense mutation in MYO7A is associated with bilateral deafness and vestibular dysfunction in the Doberman pinscher breed. *Canadian Journal of Veterinary Research*, 2019. 83(2): p. 142-148.
16. Kluth, S. and O. Distl, Congenital sensorineural deafness in Dalmatian dogs associated with quantitative trait loci. *PloS one*, 2013. 8(12).
17. Hayward, J.J. et al., A genome-wide association study of deafness in three canine breeds. *PLOS ONE*, 2020. 15(5): p. e0232900.
18. Laub, F. et al., Transcription factor KLF7 is important for neuronal morphogenesis in selected regions of the nervous system. *Molecular and Cellular Biology*, 2005. 25(13): p. 5699-5711.

19. Chen, J. et al., A systems-level approach reveals new gene regulatory modules in the developing ear. *Development*, 2017. 144(8): p. 1531-1543.
20. Strain, G., Brainstem auditory evoked response (BAER). *Deafness in dogs and cats*, 2011: p. 83-107.
21. Plassais, J. et al., Whole genome sequencing of canids reveals genomic regions under selection and variants influencing morphology. *Nature communications*, 2019. 10(1): p. 1-14.
22. Jagannathan, V. et al., A comprehensive biomedical variant catalogue based on whole genome sequences of 582 dogs and eight wolves. *Animal genetics*, 2019. 50(6): p. 695-704.
23. Li, H. and R. Durbin, Fast and accurate long-read alignment with Burrows–Wheeler transform. *Bioinformatics*, 2010. 26(5): p. 589-595.
24. Li, H. et al., The sequence alignment/map format and SAMtools. *Bioinformatics*, 2009. 25(16): p. 2078-2079.
25. McKenna, A. et al., The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome research*, 2010. 20(9): p. 1297-1303.
26. Danecek, P. et al., The variant call format and VCFtools. *Bioinformatics*, 2011. 27(15): p. 2156-2158.
27. Purcell, S. et al., PLINK: a tool set for whole-genome association and population-based linkage analyses. *The American journal of human genetics*, 2007. 81(3): p. 559-575.
28. Price, A.L. et al., Principal components analysis corrects for stratification in genome-wide association studies. *Nature genetics*, 2006. 38(8): p. 904-909.
29. Zhou, X. and M. Stephens, Genome-wide efficient mixed-model analysis for association studies. *Nature genetics*, 2012. 44(7): p. 821-824.
30. Turner, S.D., qqman: an R package for visualizing GWAS results using QQ and manhattan plots. *Biorxiv*, 2014: p. 005165.
31. Cingolani, P. et al., A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff: SNPs in the genome of *Drosophila melanogaster* strain w1118; iso-2; iso-3. *Fly*, 2012. 6(2): p. 80-92.
32. Thorvaldsdóttir, H., J.T. Robinson, and J.P. Mesirov, Integrative Genomics Viewer (IGV): high-performance genomics data visualization and exploration. *Briefings in bioinformatics*, 2013. 14(2): p. 178-192.
33. Kumar, P., S. Henikoff, and P.C. Ng, Predicting the effects of coding non-synonymous variants on protein function using the SIFT algorithm. *Nature protocols*, 2009. 4(7): p. 1073.
34. Adzhubei, I., D.M. Jordan, and S.R. Sunyaev, Predicting functional effect of human missense mutations using PolyPhen-2. *Current protocols in human genetics*, 2013. 76(1): p. 7.20. 1-7.20. 41.
35. Choi, Y. and A.P. Chan, PROVEAN web server: a tool to predict the functional effect of amino acid substitutions and indels. *Bioinformatics*, 2015. 31(16): p. 2745-2747.
36. Clarke, G.M. et al., Basic statistical analysis in genetic case-control studies. *Nat Protoc*, 2011. 6(2): p. 121-33.
37. Jourquin, J. et al., GLAD4U: deriving and prioritizing gene lists from PubMed literature. *BMC genomics*, 2012. 13(8): p. S20.

38. Taylor, W., Residual colours: a proposal for aminochromography. *Protein engineering*, 1997. 10(7): p. 743-746.
39. Schneider, T.D. and R.M. Stephens, Sequence logos: a new way to display consensus sequences. *Nucleic acids research*, 1990. 18(20): p. 6097-6100.
40. Lewis, T., J. Freeman, and L. De Risio, Decline in prevalence of congenital sensorineural deafness in Dalmatian dogs in the United Kingdom. *Journal of Veterinary Internal Medicine*, 2020.
41. Daetwyler, H.D. et al., Whole-genome sequencing of 234 bulls facilitates mapping of monogenic and complex traits in cattle. *Nature genetics*, 2014. 46(8): p. 858.
42. Raven, L.-A., B.G. Cocks, and B.J. Hayes, Multibreed genome wide association can improve precision of mapping causative variants underlying milk production in dairy cattle. *BMC genomics*, 2014. 15(1): p. 62.
43. Yan, J. and X. Wang, Detection of Disease-associated Mutations and Biomarkers Using Next-generation Sequencing. *Detection Methods in Precision Medicine*, 2020. 18: p. 119.
44. Donat, S. et al., *Heg1* and *Ccm1/2* proteins control endocardial mechanosensitivity during zebrafish valvulogenesis. *Elife*, 2018. 7: p. e28939.
45. Laub, F. et al., Developmental expression of mouse Krüppel-like transcription factor KLF7 suggests a potential role in neurogenesis. *Developmental biology*, 2001. 233(2): p. 305-318.
46. Blackmore, M.G. et al., Krüppel-like Factor 7 engineered for transcriptional activation promotes axon regeneration in the adult corticospinal tract. *Proceedings of the National Academy of Sciences*, 2012. 109(19): p. 7517-7522.
47. Lei, L. et al., The zinc finger transcription factor Klf7 is required for TrkA gene expression and development of nociceptive sensory neurons. *Genes & development*, 2005. 19(11): p. 1354-1364.
48. Kajimura, D. et al., Identification of genes regulated by transcription factor KLF7 in differentiating olfactory sensory neurons. *Gene*, 2007. 388(1-2): p. 34-42.
49. Akçimen, F. et al., A novel homozygous FBXO38 variant causes an early-onset distal hereditary motor neuronopathy type IID. *Journal of human genetics*, 2019. 64(11): p. 1141-1144.
50. Tambalo, M. et al., Enhancer activation by FGF signalling during otic induction. *Developmental biology*, 2020. 457(1): p. 69-82.
51. Brophy, P.D. et al., Genome-wide copy number variation analysis of a Branchio-oto-renal syndrome cohort identifies a recombination hotspot and implicates new candidate genes. *Human genetics*, 2013. 132(12): p. 1339-1350.
52. Wells, H.R. et al., GWAS Identifies 44 Independent Associated Genomic Loci for Self-Reported Adult Hearing Difficulty in UK Biobank. *The American Journal of Human Genetics*, 2019. 105(4): p. 788-802.
53. Kalra, G. et al., Biological insights from multi-omic analysis of 31 genomic risk loci for adult hearing difficulty. *PLoS genetics*, 2020. 16(9): p. e1009025.
54. Sun, Y. et al., GATA binding protein 3 is a direct target of Kruppel-like transcription factor 7 and inhibits chicken adipogenesis. *Frontiers in Physiology*, 2020. 11.
55. Lawoko-Kerali, G., M.N. Rivolta, and M. Holley, Expression of the transcription factors GATA3 and Pax2 during development of the mammalian inner ear. *Journal of Comparative Neurology*, 2002. 442(4): p. 378-391.

-
56. Van Esch, H. et al., GATA3 haplo-insufficiency causes human HDR syndrome. *Nature*, 2000. 406(6794): p. 419-422.
 57. Pierrat, M.-J. et al., Expression of microphthalmia-associated transcription factor (MITF), which is critical for melanoma progression, is inhibited by both transcription factor GLI2 and transforming growth factor- β . *Journal of Biological Chemistry*, 2012. 287(22): p. 17996-18004.
 58. Karlsson, E.K. et al., Efficient mapping of mendelian traits in dogs through genome-wide association. *Nature genetics*, 2007. 39(11): p. 1321-1328.
 59. Chen, L. et al., A de novo silencer causes elimination of MITF-M expression and profound hearing loss in pigs. *BMC biology*, 2016. 14(1): p. 1-15.
 60. Philipp, U. et al., A MITF mutation associated with a dominant white phenotype and bilateral deafness in German Fleckvieh cattle. *PloS one*, 2011. 6(12): p. e28857.
 61. Tassabehji, M., V.E. Newton, and A.P. Read, Waardenburg syndrome type 2 caused by mutations in the human microphthalmia (MITF) gene. *Nature genetics*, 1994. 8(3): p. 251-255.
 62. Strain, G.M., Canine deafness. *Veterinary Clinics: Small Animal Practice*, 2012. 42(6): p. 1209-1224.
 63. De Risio, L., J. Freeman, and T. Lewis, Prevalence, heritability and genetic correlations of congenital sensorineural deafness and coat pigmentation phenotype in the English bull terrier. *BMC veterinary research*, 2016. 12(1): p. 146.
 64. Sommerlad, S.F. et al., Prevalence of congenital hereditary sensorineural deafness in Australian Cattle Dogs and associations with coat characteristics and sex. *BMC Veterinary Research*, 2012. 8(1): p. 202.
 65. Strain, G.M. et al., Brainstem auditory-evoked potential assessment of congenital deafness in Dalmatians: Associations with phenotypic markers. *Journal of Veterinary Internal Medicine*, 1992. 6(3): p. 175-182.
 66. Greibrokk, T., Hereditary deafness in the Dalmation: relationship to eye and coat color. *Journal (USA)*, 1994.
 67. Famula, T., A. Oberbauer, and C. Sousa, A threshold model analysis of deafness in Dalmatians. *Mammalian genome*, 1996. 7(9): p. 650-653.
 68. Wood, J. and K. Lakhani, Prevalence and prevention of deafness in the Dalmatian—assessing the effect of parental hearing status and gender using ordinary logistic and generalized random litter effect models. *The Veterinary Journal*, 1997. 154(2): p. 121-133.
 69. Muhle, A.C. et al., Further contributions to the genetic aspect of congenital sensorineural deafness in Dalmatians. *The Veterinary Journal*, 2002. 163(3): p. 311-318.
 70. Juraschko, K. et al., Analysis of systematic effects on congenital sensorineural deafness in German Dalmatian dogs. *The Veterinary Journal*, 2003. 166(2): p. 164-169.
 71. Cargill, E. et al., Heritability and segregation analysis of deafness in US Dalmatians. *Genetics*, 2004. 166(3): p. 1385-1393.
 72. Metallinos, D. and J. Rine, Exclusion of EDNRB and KIT as the basis for white spotting in Border Collies. *Genome biology*, 2000. 1(2): p. 1-4.
 73. Stritzel, S., A. Wohlke, and O. Distl, Elimination of SILV as a candidate for congenital sensorineural deafness in Dalmatian dogs. *Animal genetics*, 2007. 38(6): p. 662-662.

Chapter 3

Genome-wide Association Studies Reveal Neurological Genes for Dog Herding, Predation, Temperament and Trainability Traits

Shuwen Shan, **Fangzheng Xu**^{*}, Bertram Brenig

Department of Animal Sciences, Faculty of Agricultural Sciences, Institute of Veterinary Medicine, University of Goettingen, Goettingen, Germany

^{*} Corresponding author

The article was published online in *Frontiers in Veterinary Science*, 8:693290. in July 2021. The full article can be found online at:

<https://doi.org/10.3389/fvets.2021.693290>

Abstract

Genome-wide association study (GWAS) using dog breed standard values as phenotypic measurements is an efficient way to identify genes associated with morphological and behavioral traits. As a result of strong human purposeful selections, several specialized behavioral traits such as herding and hunting have been formed in different modern dog breeds. However, genetic analyses on this topic are rather limited due to the accurate phenotyping difficulty for these complex behavioral traits. Here 268 dog whole genome sequences from 130 modern breeds were used to investigate candidate genes underlying dog herding, predation, temperament and trainability by GWAS. Behavioral phenotypes were obtained from American Kennel Club based on dog breed standard descriptions or groups (conventional categorization of dog historical roles). The GWAS results of herding behavior (without body size as covariates) revealed 44 significantly associated sites within 5 chromosomes. Significantly associated sites on CFA7, 9, 10 and 20 were either located in or near neuropathological or neuronal genes including *THOC1*, *ASIC2*, *MSRB3*, *LLPH*, *RFX8* and *CHL1*. *MSRB3* and *CHL1* genes were reported to be associated with dog fear. Since herding is a restricted hunting behavior by removing killing instinct, 36 hounds and 55 herding dogs were applied to analyze predation behavior. Three neuronal-related genes (*JAK2*, *MEIS1* and *LRRTM4*) were revealed as candidates for predation behavior. The significantly associated variant of temperament GWAS was located within *ACSS3* gene. The highest associated variant in trainability GWAS is located on CFA22, with no variants detected above the Bonferroni threshold. Since dog behaviors are correlated with body size, we next incorporate body mass as covariates into GWAS, and significant signals around *THOC1*, *MSRB3*, *LLPH*, *RFX8*, *CHL1*, *LRRTM4* and *ACSS3* genes were still detected for dog herding, predation and temperament behaviors. In humans these candidate genes are either involved in nervous system development or associated with mental disorders. In conclusion, our results imply that these neuronal or psychiatric genes might be involved in biological processes underlying dog herding, predation and temperament behavioral traits.

Keywords: dog behavior, GWAS, herding, predation, temperament, trainability, neurological genes

Introduction

Dogs are man's best friend and the first domesticated animal, originating from a now-extinct wolf population. Dogs have shared living space and food sources with humans, and have maintained this close relationship for more than 11,000 years (1). For only 200-300 years, humans have selectively bred dogs for excellence in herding, hunting and obedience, and created diverse breeds with a wealth of behaviors. At the same time, humans have also bred dogs for different morphological traits such as body types, sizes, skull shapes, coat colors and textures according to human preferences and needs. Two major bottlenecks in dog history: early domestication and the creation of modern breeds, have characterized long-range linkage disequilibrium (LD) within dog breeds, providing an excellent natural model for studying morphology, complex diseases and behaviors (2). Over the past two decades, scientists have attempted to explain the genetic basis of phenotypic variation among dog breeds. Many cross-breed researches were performed including morphologic traits (3-5), diseases (6), behavior or cognition (6-8) and athletic ability (9).

Dog behavior traits have been reported to be highly heritable, with a mean among-breed heritability (h^2) of 0.51 ± 0.12 (Standard Deviation) for 14 behavioral traits. Specifically, high h^2 values were observed for attachment and attention-seeking (0.56), chasing (0.62), stranger-directed aggression (0.68) and trainability (0.73) (7). However, the genetic mapping of behavior among dog breeds remains challenging. One reason is that behavior and cognition are complex traits, which are difficult to define and measure accurately (10). Therefore, different methods have been developed to classify and describe behavioral phenotypes. Behavioral studies across and within dog breeds have been explored and discussed. Using large SNP datasets and C-BARQ data of diverse breeds, dog fearlessness and aggression traits have been mapped to be associated with *GNAT3-CD36* (CFA18) and *IGSF1* (CFAX) loci (8). In the same study, variants within body size genes (*IGF1* and *HMGA2*) showed significant associations with dog behaviors such as dog rivalry, separation anxiety, touch-sensitivity and owner directed aggression (8). In one recent study, using breed-averaged C-BARQ data as phenotypes, 131 single nucleotide polymorphisms were demonstrated to be significantly associated with dog behavioral differences among 101 breeds, and the identified neurological candidate genes were highly expressed in brain (7). In addition, the among-breed

heritability of 14 behavioral traits was significantly higher than the heritability assessed in large within-breed samples (7). This study only did GWAS considering body size covariates, and they might have missed loci that affect both body size and behavior. Another GWAS of dog cognition (with and without body mass factors) using breed-averaged phenotypic values identified 5 SNPs significantly associated with breed differences in dog communication, memory, inhibitory control, and physical reasoning, and identified 188 genes related to breed cognitive differences (6).

Behavioral traits often exhibit complexity, polygenic control, and susceptibility to environmental influences. And they are inherited in linkage with other traits, for example, behavioral traits in dog are related to body size (11). For some behavioral studies, within-breed studies have shown good results and have been able to obtain more specific behavioral or cognitive locus. Recently, using the C-BARQ data as phenotypes, 11 SNPs within eight genomic regions were detected to be significantly related with six canine personality traits in Labrador retrievers (12). Two chromosome regions of CFA7:75-79Mb and CFA20:8-11Mb were investigated to be significantly associated with fearfulness in German Shepherd (13). Meanwhile, a locus of CFA11:12.8Mb was found to be significantly associated with fearfulness when investigated in Great Dane (14). These regions and the contained genes all correspond to the neuropsychiatric or neuronal gene regions in humans. In addition, human obsessive-compulsive disorder (OCD) has phenotypes similar to canine compulsive disorder (CCD), such as repetitive and time-consuming behaviors (15). Four CCD candidate genes: *CDH2*, *CTNNA2*, *ATXN1* and *PGCP* were mapped by case-control GWAS in Doberman pinschers and validated in high-risk breeds (16). Structural variants on CFA6 containing *GTF2I* and *GTF2IRD1* genes could contribute to behavioral differences (extreme sociability) between dogs and wolves, and these two genes are associated with human Williams–Beuren syndrome which is characterized by a happy and friendly disposition (17). Notably, *HS6ST2* gene was first reported to be associated with dog sociability behavior (8) and recently was detected to be significantly related to human neuroticism in GWAS of 405,274 UK Biobank samples (18). This indicates dogs could be good natural models for studying the molecular etiology of human neural disorders.

Herding dogs were bred to help people manage livestock, and they excel at controlling livestock movement. Herding derives from predatory behavior by amplifying some predatory instincts such as eye staring, stalking and chasing, while suppressing other instincts as crush, bite or kill the prey (19). Herding dogs are energetic, enthusiastic and eager to work. If they are not properly trained or assigned tasks, they even use the inclination to herd other creatures including human beings (20). They also exhibit characteristics such as agility, bravery, steadiness, and relatively low aggressiveness (21). The current study used breed specific behaviors and groupings from American Kennel Club (AKC), the most authoritative organization for the registration and classification of purebred dogs in the United States. The AKC recognizes and classifies 197 modern purebred dog breeds into seven loosely defined groups based on their breed features (heritage, physical attributes and behavior) and historical roles: Sporting, Hound, Working, Terrier, Toy, Non-sporting and Herding groups (22). The AKC group method has been successfully applied in identifications of genetic factors contributing to athleticism in sporting and hound dogs (9) as well as relationship investigations between artificial selection and human-directed play behavior (23). Genetic mapping of dog herding behavior has been firstly studied as qualitative variable in 148 dog breeds (24), and three other dog behaviors including pointing, boldness, and trainability were studied using cross-breed mapping.

Different dog breed specific traits are selected based on different human purposes, thus each dog breed has its unique temperament and trainability characteristic. Temperament is of great importance for dog breeding, especially in choosing good guide dogs (25). Pet owners are also interested in matching dog with suitable temperament (26). Among the genetic studies of temperament traits, dog activity-impulsivity endophenotype was first studied through the association analysis of candidate gene *DRD4* (27). Trainability levels were detected to have significant differences between 7 breed groups (conventional breed categories), which implies dog behavior traits such as trainability and boldness are partly owing to original function of breed. In the same study, scores of trainability, boldness, calmness and dog sociability were all detected significant differences among dog breeds (28). These breed-level behavioral differences can be used as phenotypes to study underlying genetic mechanisms, which will help us understand how these behaviors developed in dogs.

Significant brain neuroanatomical variations among breeds with different behavioral specialties, such as herding, hunting, guarding and companionship, are likely due to human selection for the behavior (29). It is reasonable to hypothesize that using a cross-breed research strategy could help us find loci that control significant behavioral variations between breeds. Therefore, this study used behavioral groupings provided by AKC to perform cross-breed GWAS to find genetic markers associated with behavioral differences among breeds. Incorporating body size factors into dog behavior GWAS can bring both merits and drawbacks as body-size related variants could also play roles in behaviors through their effects on brain architectures (30). While controlling body size factors could reveal genetics variants that are not explained by brain or body size (7). Inspired by Gnanadesikan et al. (6), significant signals identified in GWASs either with or without body mass corrections were regarded as candidates in our analysis. This study provides clues to the molecular genetic mechanisms underlying canine behaviors such as herding, predation, temperament and trainability. Understanding the formation of breed-specific behaviors in dogs will also pave the way for further elucidations of mechanisms underlying human neuropsychiatric disorders.

Materials and Methods

Samples and phenotypes

All 268 whole genome sequences of dogs that were used in this study have been extracted from vcf file data of 722 canine individuals (<https://www.ncbi.nlm.nih.gov/bioproject/PRJNA448733>), which is deposited by Dr Elaine A. Ostrander group of National Institutes of Health (NIH) (5). Many sources (n=128) such as NIH Intramural Sequencing Center are involved in the data generation with funds such as Intramural Program of the National Human Genome Research Institute. The 268 dog genomes consist of 130 established dog breeds (Supplementary Table 1) and the selection criterion is same as described in (5). For herding behavior, dogs were divided into cases and controls according to whether they belong to AKC herding group (conventional categorization) (<https://www.akc.org/dog-breeds/herding/>) or not. Forty-three herding group dogs were obtained, containing 15 modern dog breeds (Supplementary Table 1). In addition, extra 6 modern dog breeds with a herding phenotype (Rottweiler, Bernese Mountain Dog, Fomni's Dog, Lapponian Herder,

Samoyed and Swedish Lapphund) were selected according to the article (24), although these breeds were classified to working group in the AKC. Twelve cases were obtained in this step and finally fifty-five herding dogs were available as a case group. Herding dogs and hunting dogs are selected to meet different job requirements and therefore they have different degrees of prey-driven instincts. Thus, hounds and herding dogs can serve as good cases and controls for studying hunting behavior, such as aggressive behavior. In order to decipher this complex behavior, 36 hound group dogs were set as cases and 55 herding group dogs were regarded as controls for GWAS. Temperament and trainability traits were referred to the average scores of the AKC breed standard. When the ideal physical characteristics and temperament of a dog breed are specified in a written document, the breed becomes the standard breed. Therefore, different dog breeds have different levels of temperament (Outgoing, Friendly, Alert/Responsive, Reserved with Strangers and Aloof/Wary) (Figure 1A) and trainability (Eager to Please, Easy Training, Agreeable, Independent and May be Stubborn) (Figure 1B). Since kennel club group classifications are not the most accurate way to apply those phenotypes, we only set the top two levels as cases and last two levels as controls, and the middle levels (Agreeable and Alert/Responsive) were not included in GWAS analysis and considered as missing (NA). In total, 105 cases and 81 controls for the temperament analysis, 98 cases and 85 controls for trainability analysis were finally obtained (Table 1). Phenotype information of dog breed temperament and trainability traits were collected on 20 December 2020.

Genome wide association analysis (GWAS)

To obtain high-quality and only biallelic variants (Single nucleotide variants (SNVs) and small indels) for GWAS, vcf file of 722 dog genomes were firstly filtered by PLINK 1.90 with following functions (--max-alleles 2, --min-alleles 2, --minQ 20, --max-missing 0.9) (31). Then individual dogs for each GWAS were extracted from the above filtered vcf file and variants with missing value > 1% (--maf 0.01) were removed using PLINK 1.90 (31). After filtering, 14,489,548, 14,654,804, 14,984,476 and 14,853,066 biallelic variants were used for GWASs of herding, predation, temperament and trainability traits, respectively.

GWAS was conducted applying a univariate linear mixed model with sex and kinship (relatedness matrix) as covariates. The model is available in GEMMA 0.98 and two

steps of calculation were applied (32). A centered relatedness matrix was calculated in the first step, which was used as a covariate to adjust for sample structure after eigen-decomposition in the second step (32). Wald test was applied for the association significance assessment. Bonferroni thresholds ($P_{bon} = -\log(0.05/\text{number of analyzed variants})$) were used to identify significant association sites for herding ($P_{bon} = 8.46$) and predation ($P_{bon} = 8.47$) behavior. As no associated variants were above Bonferroni thresholds for temperament and trainability, suggestive thresholds ($P_{sug} = -\log(1/\text{number of analyzed variants})$) of temperament ($P_{sug} = 7.18$) and trainability ($P_{sug} = 7.17$) were applied. The suggestive threshold was first introduced by Lander and Kruglyak (33), which represents one false positive is expected per genome scan under the null hypothesis. Manhattan and quantile–quantile (QQ) plots were generated by qqman package (34). To account for body size factors in dog behavior, dog standard breed weights (SBW) and height (SBH) were further included in GWASs as covariates. Average values of body size were collected from (5). Only dog breeds that have SBW and SBH values were chosen for further analysis. The variant filtering conditions are the same as above. After filtering, 255 individuals with 14,416,697 variants, 88 dogs with 14,542,561 variants, 178 dogs with 14,829,902 variants and 177 dogs with 14,726,409 variants were analyzed in herding ($P_{bon} = 8.46$, $P_{sug} = 7.16$), predation ($P_{bon} = 8.46$, $P_{sug} = 7.16$), temperament ($P_{bon} = 8.47$, $P_{sug} = 7.17$) and trainability ($P_{bon} = 8.47$, $P_{sug} = 7.17$) GWAS, respectively. Bonferroni and suggestive thresholds were showed in the figures of GWAS results.

The genomic inflation factor lambda (λ) was calculated with the following formula: $\lambda = \text{median}(qchisq(1-p, 1)) / qchisq(0.5, 1)$ where p is a vector of P values in GWAS results. The lambda inflation factor indicates the rate of excess false positive and the extent of the bulk inflation. When values of $\lambda < 1.1$ are obtained, significant population stratification will not be considered, which was also observed in the GWAS of canine complex traits (35). The QQ plot shows the observed versus expected $-\log P$ values. The straight line in the QQ plot indicates the distribution of variant markers under the null hypothesis, and the skew at the right edge indicates those markers that are more strongly associated with the trait than would be expected by chance.

The detected associated signals were annotated by NCBI Canis lupus familiaris 3.1 Annotation Release 105. The positions were viewed by Genome Data Viewer with

CanFam3.1 reference genome
(<https://www.ncbi.nlm.nih.gov/genome/gdv/?org=canis-lupus-familiaris>).

Alternative allele frequencies of significantly associated variants

The allele frequencies of significantly associated sites were investigated in cases and controls for each GWAS trait setting separately using VCFtools 0.1.16 (36). The results of altered allele frequencies within these traits are shown in Table 2.

Linkage disequilibrium (LD) analysis of GWAS significant association signals

Linkage disequilibrium of each significantly associated site was analyzed by PLINK 1.90 (31) using following functions: `--ld-window-kb 5,000`, `--ld-window 99,999`, `--ld-window-r2 0.8`. Sites with r^2 value more than 0.8 were listed in Supplementary Table 2. Genes near or located around these LD sites were annotated by Genome Data Viewer.

Analysis of private variants in dogs with herding behavior

We next analyzed variants that were only present in 55 dogs with herding behavior. First, a total of 268 samples were quality controlled for all types of variants using VCFtools 0.1.16 (36). Only variants with minor allele frequency (MAF) > 0.05, genotype quality score (GQ) > 20 and mean depth values > 10x were selected. After separate filtering, 10,415,191 variants of 213 control dogs and 9,864,535 variants of 55 herding behavior dogs remained for further analysis. Private variants were analyzed by comparing the above filtered vcf files of 55 herding dogs and 213 controls using ‘`--diff-site`’ function in VCFtools 0.1.16 (36). The private variants were further annotated by SnpEff 5.0 with Ensembl genome 101 release (37). We have acquired 987,046 sites which were absent or rare (MAF < 0.05) in non-herding controls, and these variants were present in at least one herding dog. Variants within protein-coding genes were selected for further analysis. Variants with possible functions (high and moderate impact) in protein-coding genes were chosen, and 611 high impact variants within 270 genes and 6,740 moderate-impact variants within 2,133 genes were left. After merging genes of high and moderate impacts, 2,287 private genes remained. GO analysis was performed using these 2,287 genes with the online software WebGestalt (<http://www.webgestalt.org/>) (38). The top 10 significant biological process and cellular components were chosen for further analysis, WebGestalt applied FDR method to account for multiple testing.

The variant filtered quality conditions such as MAF and mean depth values could influence the variant content of filtered vcf files of cases and controls. For example, one variant has a MAF of 0.049 in 213 controls while its MAF is 0.051 in 55 cases, then it will be one private variant because it is absent in quality filtered vcf file of 213 controls due to $MAF < 0.05$. To prioritize the private candidate variants, these possibly functional private variants were further checked their altered allele frequencies in raw vcf files of 55 herding and 213 control dogs separately using ‘--freq’ function in VCFtools 0.1.16 (36). Variants present in more than one herding dog but not in controls, or variants with altered allele frequency differences greater than 0.1 between cases and controls are listed (Supplementary Table 3).

Investigate gene expressions of 10 candidate genes in online gene expression databases

Gene expressions of 10 candidate genes (*THOC1*, *ASIC2*, *MSRB3*, *LLPH*, *RFX8*, *CHL1*, *JAK2*, *MEIS1*, *LRRTM4*, *ACSS3*) were further examined by online database SCDevDB (<https://scdevdb.deepomics.org>) for single-cell atlas in the human neural developmental pathway (39). The cell types were oocyte, zygote, 2-cell, 4-cell, 8-cell, 16-cell, blastocyst, human embryonic stem cells (hESC), H1_24_wells, H1_96_wells, neural_D12 (neural cells generate from H1 cell line, 12 days after differentiation), neural_D26 (neural cells generate from H1 cell line, 26 days after differentiation), neural_D54 (neural cells generate from H1 cell line, 54 days after differentiation), neural_D80 (neural cells generate from H1 cell line, 80 days after differentiation). Cell details are available at <https://scdevdb.deepomics.org/data-summary/>, data information of neural cell lines was referenced in (40).

These genes were further investigated in Allen Brain Atlas Developing Mouse Brain atlas (<http://developingmouse.brain-map.org>) (41). Days of Embryonic (E) specimen age and postnatal (P) specimen age which is relative to birth (P0) are used to define the mouse brain development stages.

Results

Distribution and allele frequencies of GWAS associated sites of dog herding, predation, temperament and trainability traits

GWASs not including body size as covariates

We investigated 4 dog behavior trait phenotypes (herding, predation, temperament and trainability) using a univariate linear mixed model incorporating in GEMMA 0.98 (32). Sex and relatedness matrices (correcting for population stratification) were used as covariates to perform association tests on one single trait phenotype. For GWAS of dog herding behavior, 55 dogs with herding behavior and 213 control dogs were used. 44 significantly associated variants within regions of 5 chromosomes (CFA6, CFA7, CFA9, CFA10 and CFA20) were above Bonferroni threshold (Fig. 2). The most significantly associated region is located on CFA20 (16594598-16610335) including 16 associated sites, all of which were near or in one lncRNA: *LOC111091431*. Five of them were within *LOC111091431*, and one variant at position 16,607,008 was located in the exonic region of the lncRNA. Moreover, this variant (CFA20:16607008 A>T) was only present in dogs with herding behavior (Table 2). Another variant (CFA20:16603809 A>C) was located only 164 bp upstream of *LOC111091431*. More importantly, *LOC111091431* is located 159,124 bp upstream of the neural cell adhesion molecule L1-Like protein (*CHLI*), a neural-associated gene. On CFA7, one variant was 594 bp upstream of THO Complex 1 (*THOC1*) gene. One, two and eight significantly associated intron variants were detected in acid sensing ion channel subunit 2 (*ASIC2*), Methionine sulfoxide reductase B3 (*MSRB3*) and Regulatory factor X8 (*RFX8*) genes on chromosomes 9 and 10, respectively (Table 2). Genes such as *MSRB3* (42), *THOC1* (43), *ASIC2* (44) and *RFX8* (45) are reported to have either neuropathological or neuronal functions. These genes near significantly associated variants are indicated in Figure 2A. Other loci were located in genes that are not functionally annotated, or were located in intergenic regions and away from genes. For instance, two associated regions on CFA10 were located around 8.1 Mb and 8.6 Mb, and the closest genes in these regions were Long-term synaptic facilitation protein (*LLPH*) and *LOC111097584*.

Prey drive is the innate behavioral pattern of carnivores to pursue and capture prey, and it is a fundamental characteristic of herding dogs. Through selective breeding, humans have been able to reduce prey-driven behavior of herding dogs while maintaining their hunting skills (46). Therefore, we investigated the genetic difference between herding

and hunting dogs. Thirty-six hound group dogs and fifty-five herding group dogs were selected to study the predation differences between these two groups. This may provide further understanding of formation of herding behavior. Three chromosome regions on CFA1, 10 and 17 showed significant signals (Figure 3A). Three genes nearest to these regions were janus kinase 2 (*JAK2*) (about 1kp), meis homeobox 1 (*MEIS1*) (around 26kb) and Leucine rich repeat transmembrane neuronal 4 (*LRRTM4*) (approximately 313kb) (Table 2, Figure 3A).

To clarify potential genes that are associated with dog temperament and trainability traits, phenotypes based on breed-averaged measures were grouped as described on the AKC website (<https://www.akc.org>). The phenotypes were classified into five levels (Figure 1). The GWAS for dog temperament trait was based on 105 dogs of extraversion type and 81 dogs of aloof type. There was only one variant above suggestive threshold located in the intron region of Acyl-CoA synthetase short chain family member 3 (*ACSS3*) gene on CFA15 (Table 2 and Figure 4A). For trainability GWAS, 98 high and 85 low trainability level dogs were selected for analysis, and only one variant was detected near *LOC111091672* with a suggestive significant association for trainability (Table 2).

GWASs with body size as covariates

As body size has been reported to be related with dog behaviors, we then performed GWAS adding body size values into covariates. As shown in Figure 2-5, similar results were observed after incorporating SBW and SBH into analysis for herding, predation, temperament and trainability. In the new herding GWAS, the significantly associated chromosome regions were similar to the results without body size covariates, except for the associated site on CFA9. Although the p -values ($p = 4.03E-09$) for the CFA9 variants (CFA9:40067785 and CFA9:40068138) increased, they were still close to the Bonferroni threshold (Figure 2B). Bonferroni significantly associated signals of dog herding behavior remained around candidate genes like *THOC1*, *MSRB3*, *LLPH*, *RFX8* and *CHLI* (Figure 2B). For the predation GWAS analysis, only region of CFA17 remained significantly associated after incorporating body mass covariates into analysis, while variants near *MEIS1* gene on CFA10 were above the suggestive threshold (Figure 3B). In the new temperament GWAS analysis, the same variant of *ACSS3* showed a smaller p value ($1.92E-09$) above the Bonferroni threshold (Figure 4B). No significant

association was found for trainability after adding body size factors. GWAS QQ plots can be referred to Supplementary Figure 1.

One missense variant of *MSRB3* was in high LD level with herding GWAS associated sites

Causative variants are usually not directly detected by GWAS, and phenotypically based causal variants may be in linkage disequilibrium with GWAS-related markers. Linkage disequilibrium of each GWAS significant associated site for herding behavior was calculated by Plink 1.90 (31). The results are shown in Supplementary Table 2. Only sites with $r^2 > 0.8$ were chosen for further analysis. Except variants that already exceeded the significant threshold, 6 other variants within genes are shown in Table 3. A variant on CFA6 (CFA6:39,977,184 G>A) was located in the intron region of *PIGQ* gene, which has been reported to be the causative gene for human early-onset epilepsy (47). One missense variant (NC_006592.3:g.8037693G>A, XP_013972688.1:p.Gly179Ser) was detected in *MSRB3* gene, and the other 4 variants were located in exon regions of one lncRNA (*LOC111097584*) near *MSRB3* (38kb downstream). These 5 potentially functional variants may promote the development of dog herding behavior through directly or indirectly affecting the functions of *MSRB3* and *LOC111097584*. Ten species were chosen to analyze the *MSRB3* missense variant conservation. Six mammals have amino acid D and three species including dog, chicken and chimpanzee own G in this position (Supplementary Figure 2). This indicated that the missense variant is not conserved.

Neural development processes were highlighted in herding private genes with possible functions

Private functional variants that were only present in herding dogs could contribute to the herding behavior trait formation. Therefore, we analyzed the private functional genes of herding dogs in an attempt to find candidate genes. To obtain variants that were only present in 55 dogs with herding behavior, high quality variants of 55 herding and 213 control dogs were separately filtered. After comparing different sites between cases and controls, the variants that only existed in 55 herding dogs were annotated by SnpEff 5.0 software (37). The remained 7,351 private (611 high impact and 6,740 moderate) variants were chosen for further analysis, and these private variants were within in 2,287 protein coding genes (Supplementary Table 3). Considering that

functional variants can impact gene function, these 2,287 genes were used in GO analysis. Among the top 10 significantly enriched biological processes, 112 genes were enriched in nervous system development process (GO:0007399) and 54 genes were in neuron projection development process (GO:0031175) (Table 4). Moreover, 76 genes were enriched in neuron part (GO:0097458) within cellular component analysis. Details of the gene names and private functional variants are listed in Supplementary Table 3.

Nine candidate genes were highly expressed in different cell stages of neural development process

As these genes are related to nervous system or human mental disorders, ten candidate genes, i.e. *THOC1*, *ASIC2*, *LLPH*, *RFX8*, *MSRB3*, *CHL1*, *JAK2*, *MEIS1*, *LRRTM4* and *ACSS3*, were used for further analysis in a single-cell expression database of human neural developmental. Except *ASIC2*, 9 candidate genes were detected to be highly expressed in different early development stages of neural cells, which were generated after 12, 26, 54, 80 days' differentiation (Supplementary Figure 3). It is noted that *RFX8* gene showed unique high expression in neural cells of 12 days.

After checking these 10 genes in Allen Developing Mouse Brain Atlas, three genes including *ASIC2*, *CHL1* and *MEIS1* showed high expressions in mouse brain development stages (E11.5, E13.5, E15.5, E18.5, P4, P14 and P28). This suggests that *ASIC2* is also a neurodevelopmental gene.

Discussion

Research on the genetic mechanism of dog behaviors can help us understand dog domestication process and guide us on how to get along with dogs, which is important for dog welfare. Moreover, it could also provide clues to research of human behavior and health disorders. Dog genomes have undergone strong artificial selection with increased haplotype homozygosity and linkage disequilibrium (2). Therefore, compared with human studies, GWAS with smaller dog samples can even produce good results (2, 15). For example, GWAS with whole genome sequences across diverse

breeds has proved to be a powerful method to study canine morphological traits (5). Here we used genomic data from 268 modern dogs to perform GWAS for four behaviors and tried to find the genetic clues behind these phenotypes. In this study, phenotypes were based on dog breed standard values or group information from AKC, which is valid for revealing genomic regions and variants for several specific phenotypes such as dog fear, aggression, boldness, cognition and athleticism (4, 6, 7, 9, 24). Some dog behaviors have been reported to be highly heritable and higher than those assessed within breeds, and it is hypothesized that specific loci associated with behavioral differences between breeds can be found using across-breed genome-wide approach (7). Previous behavior or cognition GWAS were all performed with SNP chip data ($\leq 173\text{K}$), whereas we used nearly 15M variants of 130 dog breeds in this study, which were obtained by whole genome resequencing. It has a higher coverage of non-coding regions of the dog genome, which have important roles in dog behavioral traits such as differentiating dog from wolf (48). In this study, several promising candidate genes with neuronal or psychiatric were detected to be associated with breed differences of herding, predation, temperament and trainability traits.

Herding is a complex behavioral trait that requires dogs to be fearless and bold when facing large numbers of sheep or cattle. The genome-wide significant loci of fearless were mapped on CFA7:75-79 Mb and CFA20:8-11 Mb (13) and that of boldness was discovered on CFA10:6.8-8.8 Mb (4). In our herding GWAS results, nearby genomic regions of 67.1 Mb on CFA7 and 16.6 Mb on CFA20 were detected to be significantly associated (Table 2). Furthermore, two regions of 8-8.1 Mb and 8.6 Mb on CFA10, were also significantly related. These regions were either near or in the regions that were reported with dog behaviors before. The area of CFA10:8-8.6 Mb has been found to be associated with at least two morphological (ear type and body size) (3, 5, 24) and two behavioral (boldness and fear) traits (4, 7), including genes such as *MSRB3* and *HMG2*. *MSRB3* has been reported to be associated with human deafness (42, 49), brain morphology and late-onset Alzheimer's disease (50). It is also involved in stress resistance in *Drosophila* (51). Furthermore, according to GWAS Catalog database (<https://www.ebi.ac.uk/gwas/genes/MSRB3>), *MSRB3* was detected to be significantly associated with brain area volumes with the largest number of associations among the 31 reported traits. These reports suggest that *MSRB3* gene plays multiple roles in the nervous system. Except *MSRB3* region, we also identified a fragment downstream of

HMG2, which is closer to the *LLPH* gene. It has been reported that *LLPH* is involved in regulating neuronal development and synaptic transmission (52). Ear shape and body mass are two common targets of selection in domestic breeding, and selective breeding for specific traits in dogs may result in this region being selected. Also, body size was investigated to be correlated with dog behaviors (11), which were also observed in several genome-wide mapping of dog behaviors (4, 8, 24). One plausible explanation for these associations could be pleiotropy of these regions, which implies that genetic variants could affect both behavior and morphology traits in dogs. Alternatively, morphological and behavioral traits may have been co-selected due to genetic linkage (53).

We also localized another region on CFA10 (41.5 Mb) that was associated with herding behavior, which covered exons 9 and 10 of the *RFX8* gene (Table 2). This region is approximately 1.99 Mb apart from the top significantly associated site (CFA10:43493767) of dog rivalry behavior (7). It was suggested that *RFX8* could play roles in Schwann cell proliferation, as it was detected to be most prominently expressed in the Schwannoma cell line (45). Schwann cells are important for the nervous system as they direct the regeneration of peripheral axons (54). Meanwhile, *RFX8* has been identified as a candidate gene underlying human neurodevelopmental disorders (55). A significantly associated region on CFA20 covered the uncharacterized *lncRNA-LOC111091431*, the closest to which is a neural-associated gene, *CHLI*. Long non-coding RNAs are thought to be commonly but not absolutely involved in transcriptional regulation of nearby genes, and often function as cis, enhancer activity (56). Thus, it is assumed that *LOC111091431* may influence the formation of behavior through unknown interactions with *CHLI*, but its exact function remains to be verified. It was reported that *CHLI* could promote neurite outgrowth (57) and regulate cell migration during nerve regeneration (58). It is suspected that *CHLI* is also associated with intelligence (59), this could be an explanation of the higher learning ability of herding dogs. Meanwhile, *CHLI* was detected to be significantly associated with dog fear (7) and human 3p- syndrome mental impairment (60). Mice with *CHLI* deficiency demonstrated exploratory behavior changes in novel environments (61) and affected several behavioral parameters such as emotional reactivity (stress) and motor coordination (62). It was also supposed that *CHLI* could participate in nervous system development and signal transduction by regulating synaptic vesicles recycling (63).

In addition to requirements of courage, herding dogs have hunting instincts such as chasing. They are CCD-like behavioral traits that are manifested by dogs using pacing and circling to maintain and control the herd. Some CCD behaviors derive from predatory behavior, like tail chasing and fly snapping (15). The same study reported a strongly associated region of canine compulsive disorders between 61.83 and 63.87 Mb on CFA7, including *CDH2* gene (15, 16). It is noted in our findings that the region significantly associated with herding was localized between 67.13 and 67.16 Mb on CFA7, approximately 3.26 Mb from above-mentioned CCD interval. In addition, a significantly related variant CFA7:67137186 T>G was only 594 bp upstream of *THOC1* gene. However, the abovementioned variant is located within 27 Ts in a row, which suggests that it is unlikely to be regulatory. It is noted that *THOC1* gene is involved in presynaptic development and plays roles in dopamine neuron survival (43). It is also one causative gene for human late-onset hearing loss (64). Herding dogs have been selectively bred to detect and react to slight differences in whistle commands from a long distance nearly 1 kilometer and excellent hearing ability is necessary for herding tasks (65). Therefore, genes that are essential for auditory functions such as *MSRB3* and *THOC1* were detected in our herding GWAS analysis.

Significant associated regions of herding GWAS were also mapped on CFA9 containing *ASIC2* gene. *ASIC2* was reported to play roles in hippocampal neurons (44) and innating fear-like behaviors in mice (66). GO analysis revealed that *ASIC2* was detected in multiple neural cell components (Table 4). *ASIC2* was also detected among private genes of herding dog (Supplementary Table 3). *ASIC2* was detected high expressions in mouse brain development processes. However, significant signals were absent in the GWAS analysis including body mass factors (Figure 2b). Though gene functions of *LOC611691* and *OR28H03* detected on CFA6 were not related with neural function, one high LD site with the associated variant was located within *PIGQ* gene (Table 3). It has been reported that *PIGQ* is associated with the neurologic disorder of severe early-onset epilepsy (47). Overall, genes *MSRB3*, *LLPH*, *RFX8*, *CHL1*, *THOC1* and *ASIC2* are our top candidates based on herding GWAS and likely-functional variation in behavioral genes.

Hunting dogs exhibit higher prey-driven behavior in orientation, chasing, grab-bite and kill-bite (67). They usually show more excitement and aggression when hunting. However, herding dogs have higher abilities of eye-stalk and chase, but strongly inhibit the grasping, biting and killing instincts to prevent them from hurting livestock (19, 21). In a study of the behavioral interactions between dogs and livestock during herding, dog lip-licking and barking occurred less frequently, while stalking, crouching and chasing were more frequent. Moreover, not a single case of biting was observed (21). Different neurotransmitter have been detected among three dog breeds with distinct predatory behaviors: Border Collies, Siberian Huskies and Sharplaninatz (68). The GWAS between hound and herding dogs revealed three genes for prey-driven behavior (Figure 3A). The *JAK2* gene is located 1,193bp downstream of CFA1 association region (Table 2), which has been previously detected to be associated with dog snout ratio and curly tail (3, 4). One study found that dog chasing behavior has been significantly associated with skull shape. Specifically, hound or herding dog breeds tend to have long skulls as their historical roles in pursuit of potential prey animals or livestock, while companionship dogs such as toy group canines tend to have short skulls. It implies that skull shape is an indicator of hunting related behavior (11). Artificial selection based on morphological traits (like short skulls) could have affected dog behavior traits (like tendency to hunt). Meanwhile, *JAK2* is widely expressed and found to be potentially associated with dozens of traits by GWAS Catalog (<https://www.ebi.ac.uk/gwas/genes/JAK2>). Among these diverse roles, *JAK2* gene is involved in synaptic plasticity and has an essential role in the induction of NMDA-receptor dependent long-term depression (69). Inactivation of *JAK2* can cause memory loss in Alzheimer's disease (70). We found that *MEIS1* gene was detected as the nearest gene to the significantly associated region on CFA10 (Figure 3A, Table 2), and *MEIS1* was reported to be associated with Restless Legs Syndrome (71). Patients with this neurological disorder are irresistible to move the leg, which can affect sleep quality and even cause mood problems, like depression. Hyperactivity was also observed in heterozygous *MEIS1*-deficient mice suggesting its role in the specification of neuronal progenitors (72). Therefore, we propose that *MEIS1* may be associated with greater search and chase impulses in hounds when confronted with prey. The nearest gene to the significantly associated region on CFA17 for predation was *LRRTM4*. It has been reported that *LRRTM4* facilitates formation of excitatory synapse development on hippocampal dentate gyrus granule cells (73). More importantly, this gene was close to

the strongest associated signal in GWAS analysis of children aggressive behavior (74). Combined with the gene function and the report in humans, we suggest that *LRRTM4* may play a role in the differences in aggressive behavior between hounds and herding dogs. Moreover, only *LRRTM4* gene was left to be significantly associated with predation after correcting with breed standard body sizes.

Well-behaved dogs are appealing and conducive to establishing good interaction with humans. Temperament and trainability are the foundation of a dog's daily socialization or sports training, which are interesting traits for both dog owners and breeders. Dog fetching behavior has been detected to be suggestive associated with CFA22:32270336, which is 2.6 Mb away from our significantly associated signal CFA:34873149 (12). Fetching behavior has been proved to be the most efficient training method for building human-dog relationships, and it is a good indicator of trainability. The significantly associated gene of temperament GWAS has been detected to be *ACSS3* gene (Figure 4). Recently, *ACSS3* has been reported to be significantly associated with human depressive symptoms (75) and antidepressant response (76). Our results suggested that the *ACSS3* gene may contribute to the development of temperament in dogs. Different breeds of dogs have been strongly artificially selected to perform different tasks, accompanied by the production of multiple personalities. Increasing numbers of researches are focusing on the possibilities of dogs as models for studying neurological diseases (29, 77). Although the variant significantly associated with dog trainability is nearest to a *LOC111091672*, the nearest protein-coding genes upstream and downstream are *SPRY2* (distance of 1.50 Mb) and *SLITRK1* (distance of 1.45 Mb). *SPRY2* was detected to be highly expressed in the human brain, with the highest expression in the cerebellum (<http://biogps.org/#goto=genereport&id=10253>) (78). Variants in *SLITRK1* gene are associated with human psychiatric disorders such as Tourette's syndrome (79) and obsessive-compulsive disorder (80). MacLean et al. found that trainability had a very high heritability ($h^2= 0.73$) (7), indicating that the percentage of variance explained in the GWASs should be high. However, our top vs bottom GWAS designed based on AKC breed standard descriptions was underpowered. AKC written descriptions of dog breed temperament and trainability are not accurate enough for detecting variants controlling the behavioral differences among breeds. This may be one reason for the less signals obtained in GWAS of temperament and trainability.

The selection of genomic regulatory regions could contribute large effects on the formation of canine breed standards (81). Notably, epigenetic variations also play important roles in the behavioral formation (82-84). This might be due to the fact that gene coding regions are more conserved than non-coding regions, and protein-coding regions typically evolve at a slower rate. Behavioral selection for dog domestication might be caused by the regulation of gene expressions in hypothalamus (85). Several variants within lncRNAs or potential gene regulatory regions were detected in our studies, which implies they could play crucial roles in herding behavior formation through regulating gene expressions of the candidate neural genes.

Enrichment analysis was performed with candidate genes obtained from private variant analysis. Several processes or cellular components related to neurology function were obtained (Table 4). This indicates that changes in the regulation of neuron and nervous system development could contribute to herding behavior formation. These 7 candidate genes could be involved in the early neural system development (Supplementary Figure 3), which raise their possibilities of being regarded as candidate genes underlying dog behaviors. To increase the credibility of mapping, only variants above the Bonferroni genome-wide significance threshold were considered as candidates for herding and predation GWAS. Overall, 7 promising candidate genes were identified for dog herding (*THOC1*, *MSRB3*, *LLPH*, *RFX8* and *CHL1*), predation (*LRRTM4*) and temperament (*ACSS3*) between dog breeds after correcting with body mass in this study. Though associations of *ASIC2*, *JAK2* and *MEIS1* gene regions were not above significant threshold after controlling body size, they could still have potential roles on dog behaviors through effects on dog brain architectures which are related with body mass.

There are several limitations in this study. Specifically, herding group dogs are from different breeds that share herding behavior, but we were not able to determine if all herding dog breeds share a common ancestor. Phenotypic classification based on breed standard described by AKC is not robust enough to detect all the genetic variants between dog breeds, especially for trainability. Further studies using breed-average C-BARQ values could improve the accuracy. Even though GWASs using small numbers of dog individuals of very many breeds have proved to be powerful methods to

identifying variants influencing morphology (5), it is still prudent to simply apply GWASs to behavioral traits.

To fine-map the casual variants or genes for these behavioral traits accurately, professional behavioral scientists are required to perform accurate phenotypic dissections for those traits, which will be performed in Dog10K project (86). With more accurate phenotypic definitions of dog behavioral traits and more dog whole genome sequences released by Dog10k project, the understanding of genetic mechanisms underlying these behavioral traits will be significantly enhanced. In the following studies, accurate behavioral measurement methods such as Herding Trait Characterization (HTC) could be applied to evaluate large number of dogs from diverse breeds (87). Similar to this study, GWASs using breed-average scores of HTC questionnaire as phenotypes can be applied to identify genetic differences among dog breeds. To improve genetic mapping accuracy and reveal additional genes for these 4 dog behaviors, GWASs can be performed using genotype and phenotype data from the same canine individuals.

Cross-breed mapping approaches can effectively identify loci that may affect genetic differences between breeds that cannot be studied by segregation within breeds. The classic example is that the specific negative correlation between longevity and size is a strictly between-breed phenomenon and is difficult to conduct genetic analysis by within-breed studies (24). The herding behavior is also a clear between-breed behavior. Therefore, the method of classifying behaviors according to the historical roles of dogs and analyzing herding behaviors among dog breeds is reasonable. This was also reflected in the genetic mapping of herding, pointing, boldness and athleticism in dogs, and convincing genes appropriate to behaviors were obtained (4, 9, 24). Zapata et al. (77) performed a genome-wide scan of several dog behaviors of diverse breeds and also identified genes that overlap with human neurodevelopmental and psychopathological genes, implying that dogs and humans share some degree of common molecular mechanisms during neurological development. Hence, this study may provide genetic clues to further elucidate the formation of behavioral traits in dogs and provide potential models for studying complex neuropsychiatric disorders in humans.

Data Availability Statement

The original contributions presented in the study are included in the article/Supplementary Material, further inquiries can be directed to the corresponding author/s.

Author Contributions

SS: data analysis, writing-original draft, review, and editing. FX: conceptualization, data analysis, writing-original draft, review, and editing. BB: supervision, writing-review, and editing. All authors contributed to the article and approved the submitted version.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Acknowledgements

We are grateful to Dr. Ostrander and her co-authors for the availability of 722 canine genomes. SS and FX are fellows of the China Scholarship Council (CSC). Open Access Publications are supported by University of Göttingen.

Supplementary Material

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fvets.2021.693290/full#supplementary-material>

Figures

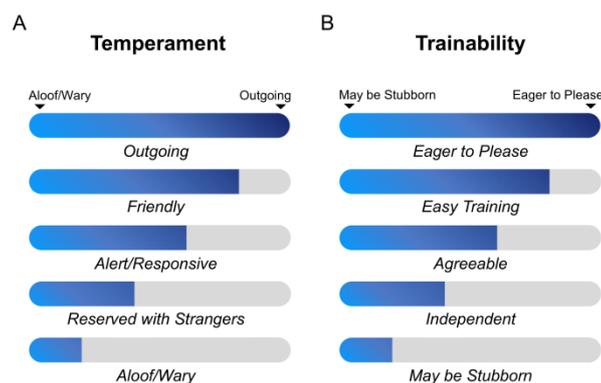


Figure 1: Classifications of trainability and temperament traits among modern dog breeds. Trait level information is obtained from AKC website (<https://www.akc.org/dog-breeds/>) (accessed on 20 December 2020), and each breed has a specific score for one of these five phenotype levels. **(A)** Aloof/wary, reserved with strangers, alert/responsive, friendly and outgoing were used to describe temperament character of each dog breed stereotype. **(B)** May be stubborn, independent, agreeable, easy training and eager to please were applied to describe trainability character of each dog breed stereotype.

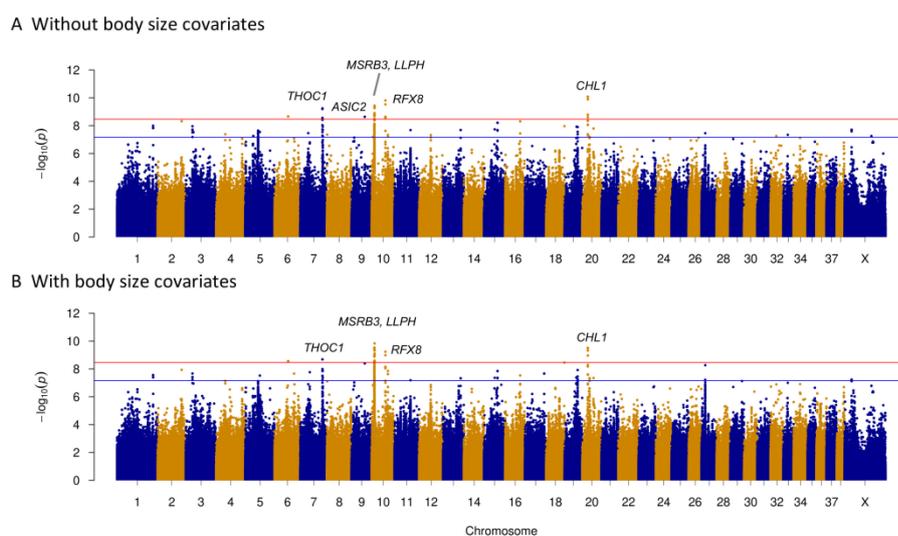


Figure 2: Manhattan plots of herding behavior Genome Wide Association Analysis (GWAS). The plots show the $-\log_{10} p$ -values for all variants of GWAS. Red horizontal line represents the Bonferroni genome-wide significance threshold, and blue horizontal line indicates suggestive significance threshold. **(A)** Without including body size as covariates. Candidate genes around significantly associated sites (above the Bonferroni threshold) were marked with red color in the Manhattan plot; they are *THOC1* of CFA7, *ASIC2* of CFA9, *MSRB3*, *LLPH* and *RFX8* of CFA10, *CHL1* of CFA20. **(B)** With including body size as covariates. Candidate genes (*THOC1*, *MSRB3*, *LLPH*, *RFX8* and *CHL1*) were marked in the plot.

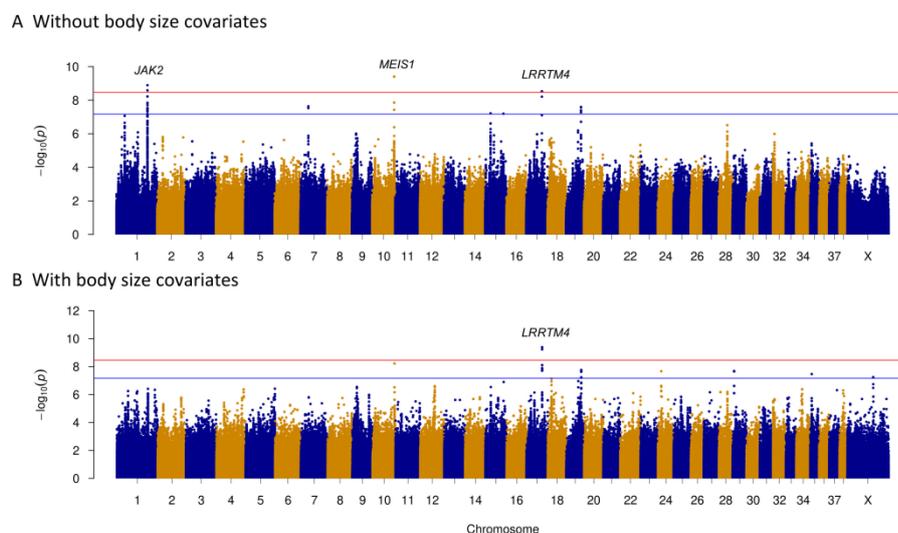


Figure 3: GWAS of dog predation analysis between hound and herding group dogs. Manhattan plots demonstrates the p -value distribution across all chromosomes. The Bonferroni and suggestive GWAS significance thresholds are indicated with the red and blue horizontal lines, respectively. **(A)** Without including body size as covariates. *JAK2*, *MEIS1* and *LRRTM4* were nearest genes to the significantly associated regions of CFA1, CFA10 and CFA17. **(B)** With including body size as covariates. Significantly associated region of CFA17 was remained after incorporating body size covariates into GWAS.

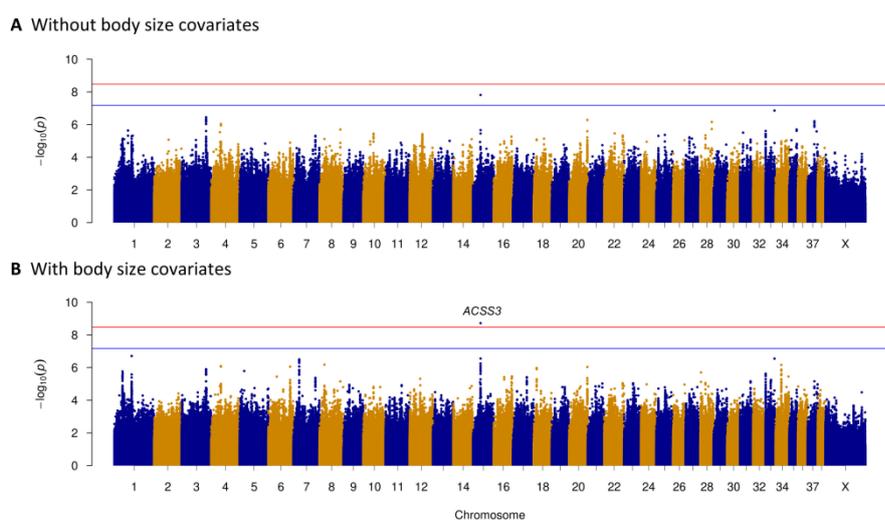


Figure 4: Temperament GWAS reveals an intron variant (CFA15:23,340,008 A>T) of *ACSS3* gene. Manhattan plots showing the association of whole genome variants with temperament levels in dogs. Bonferroni and suggestive thresholds are indicated with red and blue lines. **(A)** Without including body size as covariates. The *ACSS3* intron

variant is above the suggestive threshold. **(B)** With including body size as covariates. The ACSS3 intron variant is above the Bonferroni threshold.

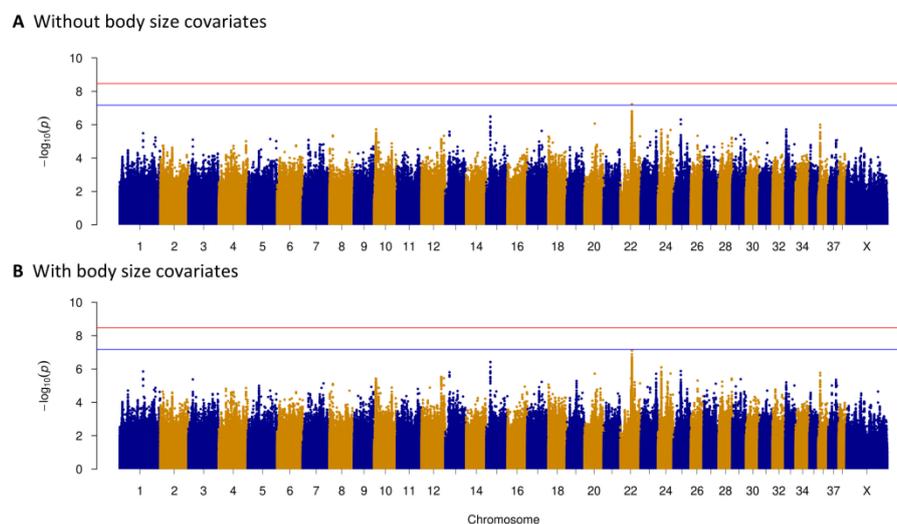


Figure 5: One variant is above the suggestive threshold of dog trainability GWAS. Manhattan plots showing the association of whole genome variants with trainability levels in dogs. Bonferroni and suggestive thresholds are indicated with red and blue lines. **(A)** Without including body size as covariates. One variant on CFA15 was slightly passed the suggestive threshold. **(B)** With including body size as covariates. No variants were detected above the suggestive threshold.

Tables
Table 1. Summary of dog behavioral phenotypes used in GWAS analysis.

Trait	Phenotype Levels	Numbers of dog	Group
Herding	Herding behavior	55	Case
	Non-herding	213	Control
Predation	Hound group	36	Case
	Herding group	55	Control
Temperament	Outgoing	19	Case
	Friendly	86	Case
	Alert/Responsive	65	NA
	Reserved with Strangers	76	Control
	Aloof/Wary	5	Control
Trainability	Eager to Please	63	Case
	Easy Training	35	Case
	Agreeable	72	NA
	Independent	68	Control
	May be Stubborn	17	Control

Table 2. GWAS significant associated variants of dog herding, predation, temperament and trainability behavior traits.

Trait	Variant	Ref	Alt	Alt_Freq_cases	Alt_Freq_controls	<i>P_value_a</i>	<i>P_value_b</i>	Nearest Gene Symbol	Distance to Gene (bp)
Herding	CFA6:40747205	A	C	0.1545	0	2.20E-09	2.75E-09	LOC611691	0
	CFA6:41114381	T	C	0.1545	0	2.20E-09	2.75E-09	OR28H03	13,451
	CFA7:67137186	T	G	0.1759	0.0024	5.72E-10	2.07E-09	THOC1	594
	CFA7:67155662	C	G	0.1455	0.0023	6.50E-10	2.07E-09	THOC1	19,070
	CFA7:67163810	A	T	0.1132	0.0023	2.78E-09	1.03E-08	THOC1	27,218
	CFA9:40067785	A	G	0.0545	0	2.33E-09	4.03E-09	ASIC2	0
	CFA9:40068138	T	C	0.0545	0	2.33E-09	4.03E-09	ASIC2	0
	CFA10:8016660	A	G	0.8182	0.331	3.29E-09	1.33E-09	MSRB3	0
	CFA10:8116174	T	G	0.7455	0.2379	2.04E-09	3.66E-10	LOC111097584	32,715
	CFA10:8116175	T	C	0.7455	0.2402	1.73E-09	3.00E-10	LOC111097584	32,716
	CFA10:8116176	C	G	0.7455	0.239	1.38E-09	3.00E-10	LOC111097584	32,717
	CFA10:8581163	C	T	0.5091	0.1056	4.40E-10	9.63E-10	LLPH	50,728
	CFA10:8583785	T	A	0.5	0.1056	1.26E-09	2.54E-09	LLPH	48,106
	CFA10:8589159	A	G	0.4909	0.1033	2.05E-09	3.87E-09	LLPH	42,732
	CFA10:8597348	G	A	0.4909	0.1033	1.52E-09	2.80E-09	LLPH	34,543
	CFA10:8601766	C	T	0.5	0.1056	5.50E-10	1.10E-09	LLPH	30,125
	CFA10:8604778	C	T	0.5091	0.1132	3.67E-10	1.49E-10	LLPH	27,113
	CFA10:8614536	CG	C	0.5545	0.1479	2.00E-09	4.86E-10	LLPH	17,355
	CFA10:8614872	AAGCTC	A	0.5545	0.1479	2.00E-09	4.86E-10	LLPH	17,091
	CFA10:8615480	G	A	0.5545	0.1479	2.00E-09	4.86E-10	LLPH	16,411
	CFA10:41504918	C	CCCTTT	0.1636	0.0236	2.98E-10	1.03E-09	RFX8	0
	CFA10:41505049	T	A	0.1545	0.0235	1.56E-10	5.89E-10	RFX8	0
	CFA10:41506217	A	G	0.1636	0.0258	2.70E-09	8.97E-09	RFX8	0
	CFA10:41506301	C	T	0.1636	0.0258	2.70E-09	8.97E-09	RFX8	0
	CFA10:41506568	C	T	0.1545	0.0235	2.31E-09	7.11E-09	RFX8	0
	CFA10:41506655	T	C	0.1636	0.0258	2.70E-09	8.97E-09	RFX8	0
	CFA10:41506849	C	T	0.1545	0.0235	2.31E-09	7.11E-09	RFX8	0
	CFA10:41507558	G	A	0.1636	0.0259	2.71E-09	8.99E-09	RFX8	0
	CFA20:16594598	C	T	0.1364	0	8.48E-11	3.16E-10	LOC111091431	9,375
	CFA20:16595519	T	C	0.1364	0	8.48E-11	3.16E-10	LOC111091431	8,454
	CFA20:16595717	G	A	0.1364	0	8.48E-11	3.16E-10	LOC111091431	8,256
	CFA20:16595938	C	T	0.1364	0	8.48E-11	3.16E-10	LOC111091431	8,035
	CFA20:16596248	A	AAAG	0.1364	0	8.48E-11	3.16E-10	LOC111091431	7,725
	CFA20:16596343	A	G	0.1296	0	2.48E-09	6.74E-09	LOC111091431	7,630
	CFA20:16596466	C	T	0.1273	0	1.65E-09	5.14E-09	LOC111091431	7,507
	CFA20:16596631	A	G	0.1364	0	8.48E-11	3.16E-10	LOC111091431	7,342
	CFA20:16597311	C	T	0.1455	0	1.30E-10	4.84E-10	LOC111091431	6,662
	CFA20:16598698	C	G	0.1364	0	8.48E-11	3.16E-10	LOC111091431	5,275
	CFA20:16603809	A	C	0.1364	0	8.48E-11	3.16E-10	LOC111091431	164
	CFA20:16604304	T	G	0.1364	0	8.48E-11	3.16E-10	LOC111091431	0

	CFA20:16607008	A	T	0.1364	0	8.48E-11	3.16E-10	LOC111091431	0
	CFA20:16607290	T	C	0.1364	0	8.48E-11	3.16E-10	LOC111091431	0
	CFA20:16610276	C	T	0.1364	0	8.74E-11	3.28E-10	LOC111091431	0
	CFA20:16610335	G	A	0.1364	0	8.48E-11	3.16E-10	LOC111091431	0
Predation	CFA1:93319503	C	T	0.1528	0.8273	2.59E-09	1.64E-06	JAK2	1,552
	CFA1:93319523	C	CATG	0.1528	0.8273	2.59E-09	1.64E-06	JAK2	1,532
	CFA1:93319862	T	C	0.1667	0.8364	1.28E-09	1.02E-06	JAK2	1,193
	CFA10:65924498	T	C	0.2639	0.9091	3.96E-10	6.02E-09	MEIS1	25,784
	CFA10:65924663	G	A	0.2639	0.9091	3.96E-10	6.02E-09	MEIS1	25,949
	CFA10:65924694	C	G	0.2639	0.9091	3.96E-10	6.02E-09	MEIS1	25,980
	CFA10:65924801	G	A	0.2639	0.9091	3.96E-10	6.02E-09	MEIS1	26,087
	CFA10:65925175	C	G	0.2639	0.9091	3.96E-10	6.02E-09	MEIS1	26,461
	CFA17:47109846	C	T	0.3056	0.8182	2.97E-09	4.16E-10	LRRTM4	312,739
	CFA17:47109848	C	T	0.3056	0.8182	2.97E-09	4.16E-10	LRRTM4	312,741
	CFA17:47109850	T	A	0.3056	0.8182	2.97E-09	4.16E-10	LRRTM4	312,743
	CFA17:47109882	C	T	0.3056	0.8182	2.97E-09	4.16E-10	LRRTM4	312,775
	Temperament	CFA15:23340008	A	T	0.7019	0.284	1.54E-08	1.92E-09	ACSS3
Trainability	CFA22:34873149	A	G	0.5941	0.1633	5.94E-08	7.92E-08	LOC111091672	19,895

Bold indicates significantly associated variants that were identified in both GWASs without/with body size as covariates. P_value_a :

GWASs without body size as covariates, P_value_b : GWASs with body size as covariates.

Table 3. Interesting LD sites of herding GWAS significant variants with $r^2 > 0.8$.

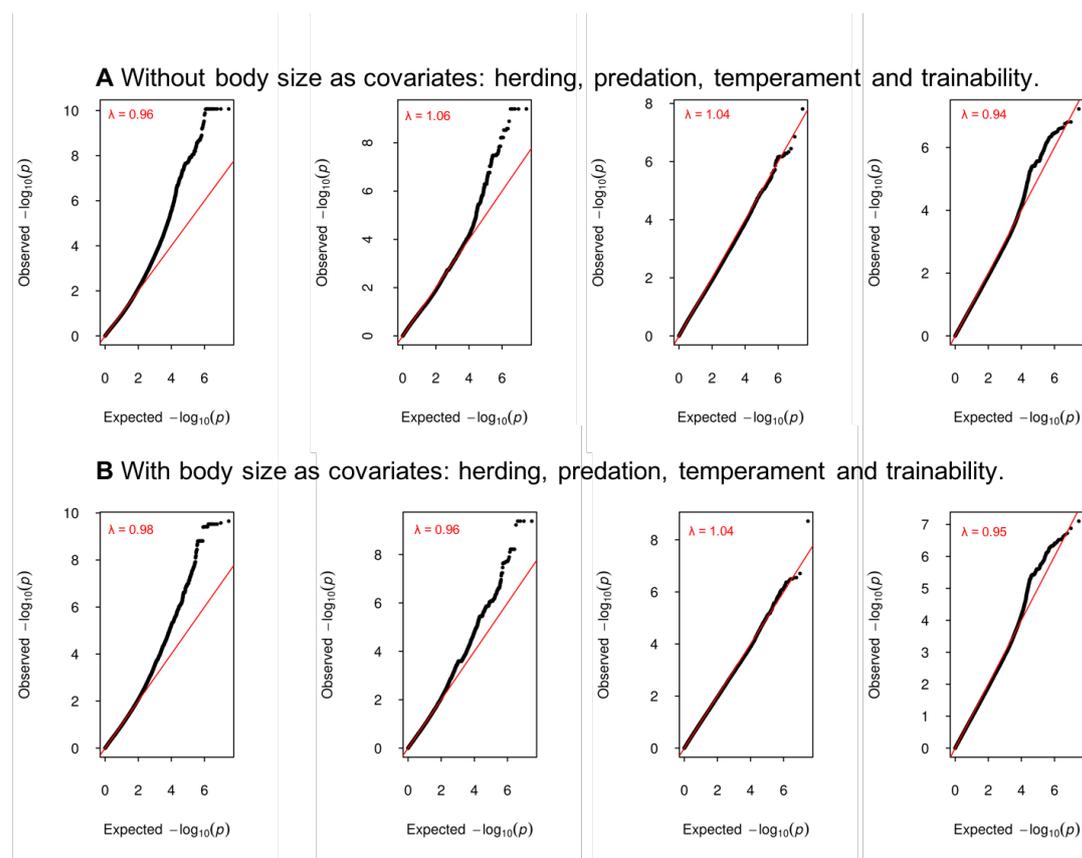
Chromosome	Position A	Position B	r^2	Gene	Gene Region	Gene type
6	40747205	39977184	0.805667	<i>PIGQ</i>	intron	Protein coding
10	8016660	8037693	0.934985	<i>MSRB3</i>	exon ^{a)}	Protein coding
10	8016660	8079815	0.908056	<i>LOC111097584</i>	exon	lncRNA
10	8016660	8079868	0.912205	<i>LOC111097584</i>	exon	lncRNA
10	8016660	8082492	0.891195	<i>LOC111097584</i>	exon	lncRNA
10	8016660	8083437	0.917264	<i>LOC111097584</i>	exon	lncRNA

a) One missense mutation within *MSRB3*: NC_006592.3:g.8037693 G>A, XP_013972688.1:p.Gly179Ser. Position

A is the position of herding GWAS significant association sites and Position B is the position of LD sites.

Table 4. Go analysis for potentially functional private genes of herding dogs.

GO category	GO ID	Description	<i>P</i> value	FDR <i>p</i> -value	Gene Counts
Biological Process	GO:0051239	Regulation of multicellular organismal process	3.49E-6	1.74E-2	167
	GO:0120036	Plasma membrane bounded cell projection organization	5.65E-6	1.74E-2	82
	GO:0050793	Regulation of developmental process	9.18E-6	1.89E-2	138
	GO:0030030	Cell projection organization	1.32E-5	2.04E-2	82
	GO:0045595	Regulation of cell differentiation	4.66E-5	5.75E-2	95
	GO:0007399	Nervous system development	5.86E-5	6.03E-2	112
	GO:2000026	Regulation of multicellular organismal development	7.00E-5	6.20E-2	109
	GO:0031175	Neuron projection development	9.99E-5	7.70E-2	54
	GO:0048869	Cellular developmental process	1.88E-5	0.11	193
	GO:0030154	Cell differentiation	1.95E-4	0.11	184
Cellular Component	GO:0044463	Cell projection part	1.40E-09	5.54E-7	73
	GO:0120038	Plasma membrane bounded cell projection part	1.40E-09	5.54E-7	73
	GO:0044459	Plasma membrane part	1.08E-08	2.85E-6	129
	GO:0042995	Cell projection	2.76E-08	5.11E-6	98
	GO:0005886	Plasma membrane	3.22E-08	5.11E-6	216
	GO:0120025	Plasma membrane bounded cell projection	4.09E-08	5.40E-6	96
	GO:0071944	Cell periphery	4.95E-08	5.61E-6	220
	GO:0098590	Plasma membrane region	2.52E-06	2.50E-4	55
	GO:0097458	Neuron part	3.22E-06	2.84E-4	76
	GO:0005887	Integral component of plasma membrane	5.92E-06	4.70E-4	71



Supplementary Figure 1: QQ plots of dog herding, predation, temperament and trainability GWASs without(A)/with(B) including body size as covariates.

House_mouse	G---KRYCINSASLSFTPADSSEAEAGSGIKESGSPAAADRAEL-----	181
Chimpanzee	SKKDLTRCCPS--ILDFLASRTVEPQ-----LIPSLYFGKVVLLWVLEAPAYHK	250
Pig	G---KRYCINSASLSFTPAESSAG--GSEGDGQA-QINRAEL-----	292
Cattle	G---KRYCINSASLSFSPAESSGAQ--GDSAGGSPAPCDKTEL-----	328
Sheep	G---KRYCINSASLSFTAESSGPQ--GDSAGGSLAPCDKTEL-----	191
Chicken	G---KRYCINSASLSFKPAEKNSAGE--GSGSTRPAQTGKAEL-----	208
Dog	G---KRYCINSASLAFTPAGGGTQ--GSSGPGGPAACGRAEL-----	183
Horse	G---KRYCINSASLSFTPADGLGA--GGSGAGSEGRADKAEL-----	209
Human	G---KRYCINSAALSFTPADSSGTAE--GGSGVASPAQADKAEL-----	185
Rhesus_monkey	G---KRYCINSAALSFTPADSSDAAE--GGSGVASPAQADKAEL-----	185

. * * . * * : : *

Supplementary Figure 2: Evolutionary conservation analysis results of MSRB3 missense mutation in 10 species using Clustal W (<https://www.ebi.ac.uk/Tools/msa/clustalo/>). The amino acid sequence accession numbers of the 10 species are as follows: House mouse XP_006513829.1, Chimpanzee XP_016778576.1, Pig XP_020947614.1, Cattle XP_024848255.1, Sheep XP_027823281.1, Chicken XP_015137097.1, Dog XP_013972688.1, Horse XP_023499721.1, Human XP_024304686.1 and Rhesus monkey NP_001244780.1.

References

1. Bergström A, Frantz L, Schmidt R, Ersmark E, Lebrasseur O, Girdland-Flink L, et al. Origins and genetic legacy of prehistoric dogs. *Science* (2020) 370:557-64. doi: 10.1126/science.aba9572.
2. Lindblad-Toh K, Wade CM, Mikkelsen TS, Karlsson EK, Jaffe DB, Kamal M, et al. Genome sequence, comparative analysis and haplotype structure of the domestic dog. *Nature* (2005) 438:803-19. doi: 10.1038/nature04338.
3. Boyko AR, Quignon P, Li L, Schoenebeck JJ, Degenhardt JD, Lohmueller KE, et al. A simple genetic architecture underlies morphological variation in dogs. *PLoS Biol* (2010) 8:e1000451. doi: 10.1371/journal.pbio.1000451.
4. Vaysse A, Ratnakumar A, Derrien T, Axelsson E, Pielberg GR, Sigurdsson S, et al. Identification of genomic regions associated with phenotypic variation between dog breeds using selection mapping. *PLoS Genet* (2011) 7:e1002316. doi: 10.1371/journal.pgen.1002316.
5. Plassais J, Kim J, Davis BW, Karyadi DM, Hogan AN, Harris AC, et al. Whole genome sequencing of canids reveals genomic regions under selection and variants influencing morphology. *Nat Commun* (2019) 10:1-14. doi: 10.1038/s41467-019-09373-w.
6. Gnanadesikan GE, Hare B, Snyder-Mackler N, Call J, Kaminski J, Miklósi Á, et al. Breed differences in dog cognition associated with brain-expressed genes and neurological functions. *Integr Comp Biol* (2020) 60:976-90. doi: 10.1093/icb/icaa112.
7. MacLean EL, Snyder-Mackler N, VonHoldt BM, Serpell JA. Highly heritable and functionally relevant breed differences in dog behaviour. *Proc Biol Sci* (2019) 286:20190716. doi: 10.1098/rspb.2019.0716.
8. Zapata I, Serpell JA, Alvarez CE. Genetic mapping of canine fear and aggression. *BMC Genomics* (2016) 17:572. doi: 10.1186/s12864-016-2936-3.
9. Kim J, Williams FJ, Dreger DL, Plassais J, Davis BW, Parker HG, et al. Genetic selection of athletic success in sport-hunting dogs. *Proc Natl Acad Sci U S A* (2018) 115:E7212-E21. doi: 10.1073/pnas.1800455115.
10. Chittka L, Rossiter SJ, Skorupski P, Fernando C. What is comparable in comparative cognition? *Philos Trans R Soc Lond B Biol Sci* (2012) 367:2677-85. doi: 10.1098/rstb.2012.0215.
11. McGreevy PD, Georgievsky D, Carrasco J, Valenzuela M, Duffy DL, Serpell JA. Dog behavior covaries with height, bodyweight and skull shape. *PloS One* (2013) 8:e80529. doi: 10.1371/journal.pone.0080529.
12. Iiska J, Haskell MJ, Blott SC, Sánchez-Molano E, Polgar Z, Lofgren SE, et al. Genetic characterization of dog personality traits. *Genetics* (2017) 206:1101-11. doi: 10.1534/genetics.116.192674.
13. Sarviaho R, Hakosalo O, Tiira K, Sulkama S, Salmela E, Hytonen MK, et al. Two novel genomic regions associated with fearfulness in dogs overlap human neuropsychiatric loci. *Transl Psychiatry* (2019) 9:18. Epub 2019/01/19. doi: 10.1038/s41398-018-0361-x. PubMed PMID: 30655508; PubMed Central PMCID: PMC636819.
14. Sarviaho R, Hakosalo O, Tiira K, Sulkama S, Niskanen J, Hytönen M, et al. A novel genomic region on chromosome 11 associated with fearfulness in dogs. *Transl Psychiatry* (2020) 10:1-10. doi: 10.1038/s41398-020-0849-z.

15. Dodman NH, Karlsson EK, Moon-Fanelli A, Galdzicka M, Perloski M, Shuster L, et al. A canine chromosome 7 locus confers compulsive disorder susceptibility. *Mol Psychiatry* (2010) 15:8-10. doi: 10.1038/mp.2009.111.
16. Tang R, Noh HJ, Wang D, Sigurdsson S, Swofford R, Perloski M, et al. Candidate genes and functional noncoding variants identified in a canine model of obsessive-compulsive disorder. *Genome Biol* (2014) 15:R25. doi: 10.1186/gb-2014-15-3-r25.
17. Shuldiner E, Koch IJ, Kartzinel RY, Hogan A, Brubaker L, Wanser S, et al. Structural variants in genes associated with human Williams-Beuren syndrome underlie stereotypical hypersociability in domestic dogs. *Sci Adv* (2017) 3:e1700398. doi: 10.1126/sciadv.1700398.
18. Luciano M, Davies G, Summers KM, Hill WD, Hayward C, Liewald DC, et al. The influence of X chromosome variants on trait neuroticism. *Molecular Psychiatry* (2021) 26:483-91. PubMed PMID: WOS:000613804500011.
19. Coppinger R, Coppinger L. Chapter 6. Behavioral Conformation. In: *Dogs: A Startling New Understanding of Canine Origin, Behavior & Evolution*. New York, NY: Scribner (2001). p. 189-224 p.
20. Renna CH. *Herding Dogs: Selection and Training the Working Farm Dog*. Allenhurst, NJ: Kennel Club Books (KCB) (2008).
21. Early J, Aalders J, Arnott E, Wade C, McGreevy P. Sequential Analysis of Livestock Herding Dog and Sheep Interactions. *Animals (Basel)* (2020) 10:352. doi: 10.3390/ani10020352.
22. Club AK. *The Complete Dog Book: 20th Edition*. New York, NY: Ballantine Books (2006).
23. Kolm N, Temrin H, Miklósi Á, Kubinyi E, Garamszegi LZ. The link between selection for function and human-directed play behaviour in dogs. *Biol Lett* (2020) 16:20200366. doi: 10.1098/rsbl.2020.0366.
24. Jones P, Chase K, Martin A, Davern P, Ostrander EA, Lark KG. Single-nucleotide-polymorphism-based association mapping of dog stereotypes. *Genetics* (2008) 179:1033-44. doi: 10.1534/genetics.108.087866.
25. Bray EE, Sammel MD, Cheney DL, Serpell JA, Seyfarth RM. Effects of maternal investment, temperament, and cognition on guide dog success. *Proc Natl Acad Sci U S A* (2017) 114:9128-33. doi: 10.1073/pnas.1704303114.
26. Jones AC, Gosling SD. Temperament and personality in dogs (*Canis familiaris*): A review and evaluation of past research. *Appl Anim Behav Sci* (2005) 95:1-53. doi: 10.1016/j.applanim.2005.04.008.
27. Hejjas K, Vas J, Topál J, Szántai E, Rónai Z, Székely A, et al. Association of polymorphisms in the dopamine D4 receptor gene and the activity-impulsivity endophenotype in dogs. *Anim Genet* (2007) 38:629-33. doi: 10.1111/j.1365-2052.2007.01657.x.
28. Turcsán B, Kubinyi E, Miklósi Á. Trainability and boldness traits differ between dog breed clusters based on conventional breed categories and genetic relatedness. *Appl Anim Behav Sci* (2011) 132:61-70. doi: 10.1016/j.applanim.2011.03.006.
29. Hecht EE, Smaers JB, Dunn WD, Kent M, Preuss TM, Gutman DA. Significant neuroanatomical variation among domestic dog breeds. *J Neurosci* (2019) 39:7748-58. doi: 10.1523/JNEUROSCI.0303-19.2019.

30. Horschler DJ, Hare B, Call J, Kaminski J, Miklósi Á, MacLean EL. Absolute brain size predicts dog breed differences in executive function. *Anim Cogn* (2019) 22:187-98. doi: 10.1007/s10071-018-01234-1.
31. Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MA, Bender D, et al. PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am J Hum Genet* (2007) 81:559-75. doi: 10.1086/519795.
32. Zhou X, Stephens M. Genome-wide efficient mixed-model analysis for association studies. *Nat Genet* (2012) 44:821-4. doi: 10.1038/ng.2310.
33. Lander E, Kruglyak L. Genetic dissection of complex traits: guidelines for interpreting and reporting linkage results. *Nat Genet* (1995) 11:241-7. doi: 10.1038/ng1195-241.
34. Turner SD. qqman: an R package for visualizing GWAS results using QQ and manhattan plots. *Biorxiv* (2014):005165. doi: 10.1101/005165.
35. Hayward JJ, Castelhana MG, Oliveira KC, Corey E, Balkman C, Baxter TL, et al. Complex disease and phenotype mapping in the domestic dog. *Nat Commun* (2016) 7:1-11. doi: 10.1038/ncomms10460.
36. Danecek P, Auton A, Abecasis G, Albers CA, Banks E, DePristo MA, et al. The variant call format and VCFtools. *Bioinformatics* (2011) 27:2156-8. doi: 10.1093/bioinformatics/btr330.
37. Cingolani P, Platts A, Wang LL, Coon M, Nguyen T, Wang L, et al. A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff: SNPs in the genome of *Drosophila melanogaster* strain w1118; iso-2; iso-3. *Fly (Austin)* (2012) 6:80-92. doi: 10.4161/fly.19695.
38. Liao Y, Wang J, Jaehnig EJ, Shi Z, Zhang B. WebGestalt 2019: gene set analysis toolkit with revamped UIs and APIs. *Nucleic Acids Res* (2019) 47:W199-W205. doi: 10.1093/nar/gkz401.
39. Wang Z, Feng X, Li SC. SCDevDB: a database for insights into single-cell gene expression profiles during human developmental processes. *Front Genet* (2019) 10:903. doi: 10.3389/fgene.2019.00903.
40. Yao Z, Mich JK, Ku S, Menon V, Krostag A-R, Martinez RA, et al. A single-cell roadmap of lineage bifurcation in human ESC models of embryonic brain development. *Cell Stem Cell* (2017) 20:120-34. doi: 10.1016/j.stem.2016.09.011.
41. Sunkin SM, Ng L, Lau C, Dolbeare T, Gilbert TL, Thompson CL, et al. Allen Brain Atlas: an integrated spatio-temporal portal for exploring the central nervous system. *Nucleic Acids Research* (2013) 41:D996-D1008. PubMed PMID: WOS:000312893300142.
42. Ahmed ZM, Yousaf R, Lee BC, Khan SN, Lee S, Lee K, et al. Functional null mutations of MSRB3 encoding methionine sulfoxide reductase are associated with human deafness DFNB74. *Am J Hum Genet* (2011) 88:19-29. doi: 10.1016/j.ajhg.2010.11.010.
43. Maeder CI, Kim J-I, Liang X, Kaganovsky K, Shen A, Li Q, et al. The THO complex coordinates transcripts for synapse development and dopamine neuron survival. *Cell* (2018) 174:1436-49. e20. doi: 10.1016/j.cell.2018.07.046.
44. Askwith CC, Wemmie JA, Price MP, Rokhlina T, Welsh MJ. Acid-sensing ion channel 2 (ASIC2) modulates ASIC1 H⁺-activated currents in hippocampal neurons. *J Biol Chem* (2004) 279:18296-305. doi: 10.1074/jbc.M312145200.

45. Sugiaman-Trapman D, Vitezic M, Jouhilahti E-M, Mathelier A, Lauter G, Misra S, et al. Characterization of the human RFX transcription factor family by regulatory and target gene analysis. *BMC Genomics* (2018) 19:1-15. doi: 10.1186/s12864-018-4564-6.
46. Siniscalchi M, Bertino D, d'Inge S, Quaranta A. Relationship between motor laterality and aggressive behavior in sheepdogs. *Symmetry* (2019) 11:233. doi: 10.3390/sym11020233.
47. Martin HC, Kim GE, Pagnamenta AT, Murakami Y, Carvill GL, Meyer E, et al. Clinical whole-genome sequencing in severe early-onset epilepsy reveals new genes and improves molecular diagnosis. *Hum Mol Genet* (2014) 23:3200-11. doi: 10.1093/hmg/ddu030.
48. Sahlén P, Yanhu L, Xu J, Kubinyi E, Wang G-D, Savolainen P. Variants That Differentiate Wolf and Dog Populations Are Enriched in Regulatory Elements. *Genome Biology and Evolution* (2021) 13:evab076.
49. Kwon T-J, Cho H-J, Kim U-K, Lee E, Oh S-K, Bok J, et al. Methionine sulfoxide reductase B3 deficiency causes hearing loss due to stereocilia degeneration and apoptotic cell death in cochlear hair cells. *Hum Mol Genet* (2014) 23:1591-601. doi: 10.1093/hmg/ddt549.
50. Hibar DP, Adams HH, Jahanshad N, Chauhan G, Stein JL, Hofer E, et al. Novel genetic loci associated with hippocampal volume. *Nat Commun* (2017) 8:1-12. doi: 10.1038/ncomms13624.
51. Lim D-H, Han JY, Kim J-R, Lee YS, Kim H-Y. Methionine sulfoxide reductase B in the endoplasmic reticulum is critical for stress resistance and aging in *Drosophila*. *Biochem Biophys Res Commun* (2012) 419:20-6. doi: 10.1016/j.bbrc.2012.01.099.
52. Yu N-K, Kim HF, Shim J, Kim S, Kim DW, Kwak C, et al. A transducible nuclear/nucleolar protein, mLLP, regulates neuronal morphogenesis and synaptic transmission. *Sci Rep* (2016) 6:1-11. doi: 10.1038/srep22892.
53. Webster MT, Kamgari N, Perloski M, Hoepfner MP, Axelsson E, Hedhammar Å, et al. Linked genetic variants on chromosome 10 control ear morphology and body mass among dog breeds. *BMC Genomics* (2015) 16:1-17. doi: 10.1186/s12864-015-1702-2.
54. Son Y-J, Thompson WJ. Schwann cell processes guide regeneration of peripheral axons. *Neuron* (1995) 14:125-32. doi: 10.1016/0896-6273(95)90246-5.
55. Coe BP, Stessman HA, Sulovari A, Geisheker MR, Bakken TE, Lake AM, et al. Neurodevelopmental disease genes implicated by de novo mutation and copy number variation morbidity. *Nat Genet* (2019) 51:106-16. doi: 10.1038/s41588-018-0288-4.
56. Kopp F, Mendell JT. Functional Classification and Experimental Dissection of Long Noncoding RNAs. *Cell* (2018) 172:393-407. PubMed PMID: WOS:000423447600003.
57. Hillenbrand R, Molthagen M, Montag D, Schachner M. The close homologue of the neural adhesion molecule L1 (CHL1): patterns of expression and promotion of neurite outgrowth by heterophilic interactions. *Eur J Neurosci* (1999) 11:813-26. doi: 10.1046/j.1460-9568.1999.00496.x.
58. Buhusi M, Midkiff BR, Gates AM, Richter M, Schachner M, Maness PF. Close homolog of L1 is an enhancer of integrin-mediated cell migration. *J Biol Chem* (2003) 278:25024-31. doi: 10.1074/jbc.M303084200.
59. Angeloni D, Wei M, Lerman M. Two single nucleotide polymorphisms (SNPs) in the CALL gene for association studies with IQ. *Psychiatr Genet* (1999) 9:165-7. doi: 10.1097/00041444-199909000-00011.

60. Angeloni D, Lindor NM, Pack S, Latif F, Wei MH, Lerman MI. CALL gene is haploinsufficient in a 3p- syndrome patient. *Am J Med Genet* (1999) 86:482-5. doi: 10.1002/(sici)1096-8628(19991029)86:5<482::aid-ajmg15>3.0.co;2-l.
61. Montag-Sallaz M, Schachner M, Montag D. Misguided axonal projections, neural cell adhesion molecule 180 mRNA upregulation, and altered behavior in mice deficient for the close homolog of L1. *Mol Cell Biol* (2002) 22:7967-81. doi: 10.1128/mcb.22.22.7967-7981.2002.
62. Pratte M, Rougon G, Schachner M, Jamon M. Mice deficient for the close homologue of the neural adhesion cell L1 (CHL1) display alterations in emotional reactivity and motor coordination. *Behav Brain Res* (2003) 147:31-9. doi: 10.1016/s0166-4328(03)00114-1.
63. Leshchyns'ka I, Sytnyk V, Richter M, Andreyeva A, Puchkov D, Schachner M. The adhesion molecule CHL1 regulates uncoating of clathrin-coated synaptic vesicles. *Neuron* (2006) 52:1011-25. doi: 10.1016/j.neuron.2006.10.020.
64. Zhang L, Gao Y, Zhang R, Sun F, Cheng C, Qian F, et al. THOC1 deficiency leads to late-onset nonsyndromic hearing loss through p53-mediated hair cell apoptosis. *PLoS Genet* (2020) 16:e1008953. doi: 10.1371/journal.pgen.1008953.
65. Yokoyama JS, Lam ET, Ruhe AL, Erdman CA, Robertson KR, Webb AA, et al. Variation in genes related to cochlear biology is strongly associated with adult-onset deafness in border collies. *PLoS Genet* (2012) 8:e1002898. doi: 10.1371/journal.pgen.1002898.
66. Price MP, Gong H, Parsons MG, Kundert JR, Reznikov LR, Bernardinelli L, et al. Localization and behaviors in null mice suggest that ASIC1 and ASIC2 modulate responses to aversive stimuli. *Genes Brain Behav* (2014) 13:179-94. doi: 10.1111/gbb.12108.
67. Mehrkam LR, Wynne CD. Behavioral differences among breeds of domestic dogs (*Canis lupus familiaris*): Current status of the science. *Appl Anim Behav Sci* (2014) 155:12-27. doi: 10.1016/j.applanim.2014.03.005.
68. Arons CD, Shoemaker WJ. The distribution of catecholamines and β -endorphin in the brains of three behaviorally distinct breeds of dogs and their F1 hybrids. *Brain Res* (1992) 594:31-9. doi: 10.1016/0006-8993(92)91026-b.
69. Nicolas CS, Peineau S, Amici M, Csaba Z, Fafouri A, Javalet C, et al. The Jak/STAT pathway is involved in synaptic plasticity. *Neuron* (2012) 73:374-90. doi: 10.1016/j.neuron.2011.11.024.
70. Chiba T, Yamada M, Sasabe J, Terashita K, Shimoda M, Matsuoka M, et al. Amyloid- β causes memory impairment by disturbing the JAK2/STAT3 axis in hippocampal neurons. *Mol Psychiatry* (2009) 14:206-22. doi: 10.1038/mp.2008.105.
71. Winkelmann J, Schormair B, Lichtner P, Ripke S, Xiong L, Jalilzadeh S, et al. Genome-wide association study of restless legs syndrome identifies common variants in three genomic regions. *Nat Genet* (2007) 39:1000-6. Epub 2007/07/20. doi: 10.1038/ng2099. PubMed PMID: 17637780.
72. Spieler D, Kaffe M, Knauf F, Bessa J, Tena JJ, Giesert F, et al. Restless Legs Syndrome-associated intronic common variant in *Meis1* alters enhancer function in the developing telencephalon. *Genome Research* (2014) 24:592-603. doi: 10.1101/gr.166751.113.
73. Siddiqui TJ, Tari PK, Connor SA, Zhang P, Dobie FA, She K, et al. An LRRTM4-HSPG complex mediates excitatory synapse development on dentate gyrus granule cells. *Neuron* (2013) 79:680-95. doi: 10.1016/j.neuron.2013.06.029.

74. Pappa I, St Pourcain B, Benke K, Cavadino A, Hakulinen C, Nivard MG, et al. A genome-wide approach to children's aggressive behavior: The EAGLE consortium. *Am J Med Genet B Neuropsychiatr Genet* (2016) 171:562-72. doi: 10.1002/ajmg.b.32333.
75. Arnau-Soler A, Macdonald-Dunlop E, Adams MJ, Clarke T-K, MacIntyre DJ, Milburn K, et al. Genome-wide by environment interaction studies of depressive symptoms and psychosocial stress in UK Biobank and Generation Scotland. *Transl Psychiatry* (2019) 9:1-13. doi: 10.1038/s41398-018-0360-y.
76. Hunter AM, Leuchter AF, Power RA, Muthén B, McGrath PJ, Lewis CM, et al. A genome-wide association study of a sustained pattern of antidepressant response. *J Psychiatr Res* (2013) 47:1157-65. doi: 10.1016/j.jpsychires.2013.05.002.
77. Zapata I, Hecht EE, Serpell JA, Alvarez CE. Genome scans of dog behavior implicate a gene network underlying psychopathology in mammals, including humans. *bioRxiv* (2021):2020.07.19.211078. doi: 10.1101/2020.07.19.211078.
78. Wu CL, Jin XF, Tsueng G, Afrasiabi C, Su AI. BioGPS: building your own mash-up of gene annotations and expression profiles. *Nucleic Acids Research* (2016) 44:D313-D6. PubMed PMID: WOS:000371261700043.
79. Abelson JF, Kwan KY, O'Roak BJ, Baek DY, Stillman AA, Morgan TM, et al. Sequence variants in *SLITRK1* are associated with Tourette's syndrome. *Science* (2005) 310:317-20. doi: 10.1126/science.1116502.
80. Wendland J, Kruse M, Murphy D. Functional *SLITRK1* var321, varCDfs and *SLC6A4* G56A variants and susceptibility to obsessive-compulsive disorder. *Mol Psychiatry* (2006) 11:802-4. doi: 10.1038/sj.mp.4001848.
81. Consortium EP. An integrated encyclopedia of DNA elements in the human genome. *Nature* (2012) 489:57-74. doi: 10.1038/nature11247.
82. Hammock EA, Young LJ. Microsatellite instability generates diversity in brain and sociobehavioral traits. *Science* (2005) 308:1630-4. doi: 10.1126/science.1111427.
83. Spady TC, Ostrander EA. Canine behavioral genetics: pointing out the phenotypes and herding up the genes. *Am J Hum Genet* (2008) 82:10-8. doi: 10.1016/j.ajhg.2007.12.001.
84. Janowitz Koch I, Clark MM, Thompson MJ, Deere-Machemer KA, Wang J, Duarte L, et al. The concerted impact of domestication and transposon insertions on methylation patterns between dogs and grey wolves. *Mol Ecol* (2016) 25:1838-55. doi: 10.1111/mec.13480.
85. Saetre P, Lindberg J, Leonard JA, Olsson K, Pettersson U, Ellegren H, et al. From wild wolf to domestic dog: gene expression changes in the brain. *Brain Res Mol Brain Res* (2004) 126:198-206. doi: 10.1016/j.molbrainres.2004.05.003.
86. Ostrander EA, Wang G-D, Larson G, Vonholdt BM, Davis BW, Jagannathan V, et al. Dog10K: an international sequencing effort to advance studies of canine domestication, phenotypes and health. *Natl Sci Rev* (2019) 6:810-24. doi: 10.1093/nsr/nwz049.
87. Arvelius P, Malm S, Svartberg K, Strandberg E. Measuring herding behavior in Border collie—effect of protocol structure on usefulness for selection. *Journal of Veterinary Behavior* (2013) 8:9-18.

Chapter 4

Association of α/β -Hydrolase D16B with Bovine Conception Rate and Sperm Plasma Membrane Lipid Composition

Shuwen Shan[†], **Fangzheng Xu[†]**, Martina Bleyer, Svenja Becker, Torben Melbaum, Wilhelm Wemheuer, Marc Hirschfeld, Christin Wacker, Shuhong Zhao, Ekkehard Schütz, and Bertram Brenig^{*}

[†] These authors contributed equally to this work.

^{*}Corresponding author

The article was published online in *International journal of molecular sciences*, 21(2), p.627. in January 2020. The full article can be found online at:

<https://doi.org/10.3390/ijms21020627>

Author contributions:

Fangzheng Xu carried out the experiments and participated in data analysis, mainly performed functional analysis of *ABHD16B* gene. He conducted FRET genotyping experiment and sperm lipidomic analysis. He was also involved in the manuscript writing and project discussion.

Acknowledgments:

Special thanks are given to Zentrum für molekulare Diagnostik (ZMD) for bovine BeadChip genotyping.

Abstract

We have identified a Holstein sire named Tarantino who had been approved for artificial insemination that is based on normal semen characteristics (*i.e.*, morphology, thermoresistance, motility, sperm concentration), but had no progeny after 412 first inseminations, resulting in a non-return rate (NR_{dev}) of -29. Using whole genome association analysis and next generation sequencing, an associated nonsense variant in the α/β -hydrolase domain-containing 16B gene (*ABHD16B*) on bovine chromosome 13 was identified. The frequency of the mutant allele in the German Holstein population was determined to be 0.0018 in 222,645 investigated cattle specimens. The mutant allele was traced back to Whirlhill Kingpin (born Feb. 13th, 1959) as potential founder. The expression of *ABHD16B* was detected by Western blotting and immunohistochemistry in testis and epididymis of control bulls. A lipidome comparison of the plasma membrane of fresh semen from carriers and controls showed significant differences in the concentration of phosphatidylcholine (PC), diacylglycerol (DAG), ceramide (Cer), sphingomyelin (SM), and phosphatidylcholine (-ether) (PC O-), indicating that *ABHD16B* plays a role in lipid biosynthesis. The altered lipid contents may explain the reduced fertilization ability of mutated sperms.

Keywords: Holstein cattle; male infertility; *ABHD16B*

Introduction

Fertility is an important economical productivity factor in animal breeding [1–4]. Indicators to assess male fertility can either be indirect (e.g., productivity of progeny, sire conception rate, non-return rate) or direct (e.g., semen characteristics, testis size) [5–10]. The latter parameters have the advantage that they can be easily measured and they provide an immediate answer; however, the heritabilities of scrotal circumference and semen traits vary extremely, ranging from 0.0 (*i.e.*, abnormal heads, bent tails, distal

cytoplasmic droplets) to 0.57 (*i.e.*, scrotal circumference) and, therefore, their use in selection is not always straightforward [11]. Alternative approaches were used to determine the differences between fertile and infertile bulls while using molecular tools. Transcriptome analyses for instance have shown that spermatozoa of high-fertility bulls show a higher concentration of specific transcripts for membrane and extracellular space protein locations [12,13]. In another study residual RNA content in spermatozoa of bulls with extreme non-return rates was analysed [14]. Low-fertile bulls showed a significantly increased amount of ribosomal and mitochondrial sequences, whereas high-fertile bulls exhibited transcripts of genes that are involved, for example, in metabolism, signal transduction, translation, and protein degradation [14]. From transcriptome and proteome studies, mainly in man, mouse, and rat, it is evident that differences between RNA and protein content, DNA methylation, posttranslational modifications between fertile and infertile individuals exist [15–18]. The use of these types of biomarkers in reproductive medicine is believed to bridge the gap between conventional semen analysis with limited clinical utility and biochemical pathways that regulate male fertility [19].

However, the assessment of mutational effects in candidate genes is normally challenging, especially when there are only subtle deviations in expression levels, due to the complex interactions of geno- and phenotypes in fertility traits [20,21]. With the advancement of high-throughput screening tools (DNA chip, next generation sequencing) and the availability of large datasets on fertility parameters of bulls, especially in Holstein cattle male fertility, can be practically implemented into genomic selection [22]. Genome-wide association studies have been conducted in Holstein bulls, identifying several fertility associated genomic regions [23]. A recent genome-wide association study has detected at least eight genomic regions, *i.e.*, on bovine chromosome 5 (BTA5), BTA9, BTA13, BTA21, and BTA25, in Holstein cattle associated with bull fertility while using Sire Conception Rate (SCR) as a parameter [24]. In a large multi-species comparative study 33 promising candidate genes have been identified for male fertility/infertility [25]. Recently, a whole exome sequencing of 24 high and low fertile bulls identified 484 SNPs that were significantly associated with fertility [26]. The second most significantly associated SNP in this study was located on BTA13 at position 53,691,419 within the *SIRPA* gene. Although these data point at a number of potential molecular targets only three causative mutations,

resulting in male sub- or infertility in cattle have been determined in the *FSHB*, *TMEM95*, and *ARMC3* gene hitherto [27–30].

Here, we report about the identification of nonsense variant in the bovine α/β -hydrolase D16B gene (*ABHD16B*) on BTA13 significantly associated with male subfertility in Holstein cattle. So far, nothing was known regarding the physiological or biochemical function of ABHD16B [31,32]. Our data provide evidence that ABHD16B is involved in lipid biosynthesis in testis and is crucial for fertilization.

Results

Conception Ability of Sires is Highly Associated with a Chromosomal Region on Bovine Chromosome 13

A Genome Wide Association Analysis (GWAS) was performed while using a cohort of 289 Holstein sires to determine chromosomal regions harboring associated causative genes for conception ability (NR_{dev}). The cohort consisted of 10 sires with a $NR_{dev} \leq -2$ (= cases) (Table 1) and 279 randomly chosen sires of the active breeding population (= controls).

Table 1. Sub- and infertile sires selected for genome-wide association analysis.

Sire/ID	NR_{dev} ^{a)}	No. of First Inseminations
Tarantino	-29	412
19_39644	-27	402
19_39643	-25	364
05_34345	-9	412
04_44565	-4	421
04_39067	-3	315
04_43327	-3	424
04_40476	-2	407
04_41962	-2	640
04_37666	-2	571

a) NR_{dev} : Non-return rate deviation.

Individual NR_{dev} values of the control sires were not available; however, they were assumed to be normal, as all of these sires were used in the current breeding population. As shown in Figure 1A, one genome-wide highly significant associated position on

BTA13 (ARS-BFGL-NGS-107931; position 63,500,701) was detected ($-\log_{10}P$ -value = 167.56). Seventeen additional regions above a Bonferroni threshold of $-\log_{10}P = 5.9$ ($p < 0.05$) with much lower significance were present on BTA1, 2, 3, 6, 7, 8, 10, 11, 14, 17, 18, 21, 22, 24, 25, 26, and 27. The QQ-plot clearly indicated a compelling evidence for an excess of association with no population substructure (Figure 1B). Whole-genome sequencing was performed while using Tarantino and his parents to determine which of the associated chromosomal regions harbored protein-altering variants that were causative for Tarantino's infertility.

Whole-Genome Sequencing Reveals Two Potential Protein-Altering Variants

Upstream the Associated Position on BTA13

Raw next generation sequencing data were quality filtered. Within the filtered 78,472 SNPs, only 20 resulted in a predicted loss of function, including 10 nonsense variants, five splice-donor variants, three splice acceptor-variants, and two initiator-codon variants. Two SNPs were located near the associated position on BTA13, *i.e.*, a nonsense variant at position 54,429,815 within the single exonic α/β -hydrolase D16B (*ABHD16B*) gene (AC_000170.1: g.54429815G>A, rs468948776) and a splice-acceptor variant at position 53,003,648 within the transmembrane channel-like protein 2 (*TMC2*) gene (AC_000170.1: g.53003648C>T, rs465702794). *TMC2* has been shown to be expressed in the inner ear and it is necessary for the mechanotransduction in cochlear hair cells [33,34]. *TMC2* was excluded as potential candidate due to this very specific function. On the other hand, *ABHD16B* has been shown in humans to be mainly expressed in testis, which suggested a potential role in Tarantino's infertility [35]. In addition, aberrant methylation patterns of *ABHD16B* have been shown to be associated with infertility in men [36].

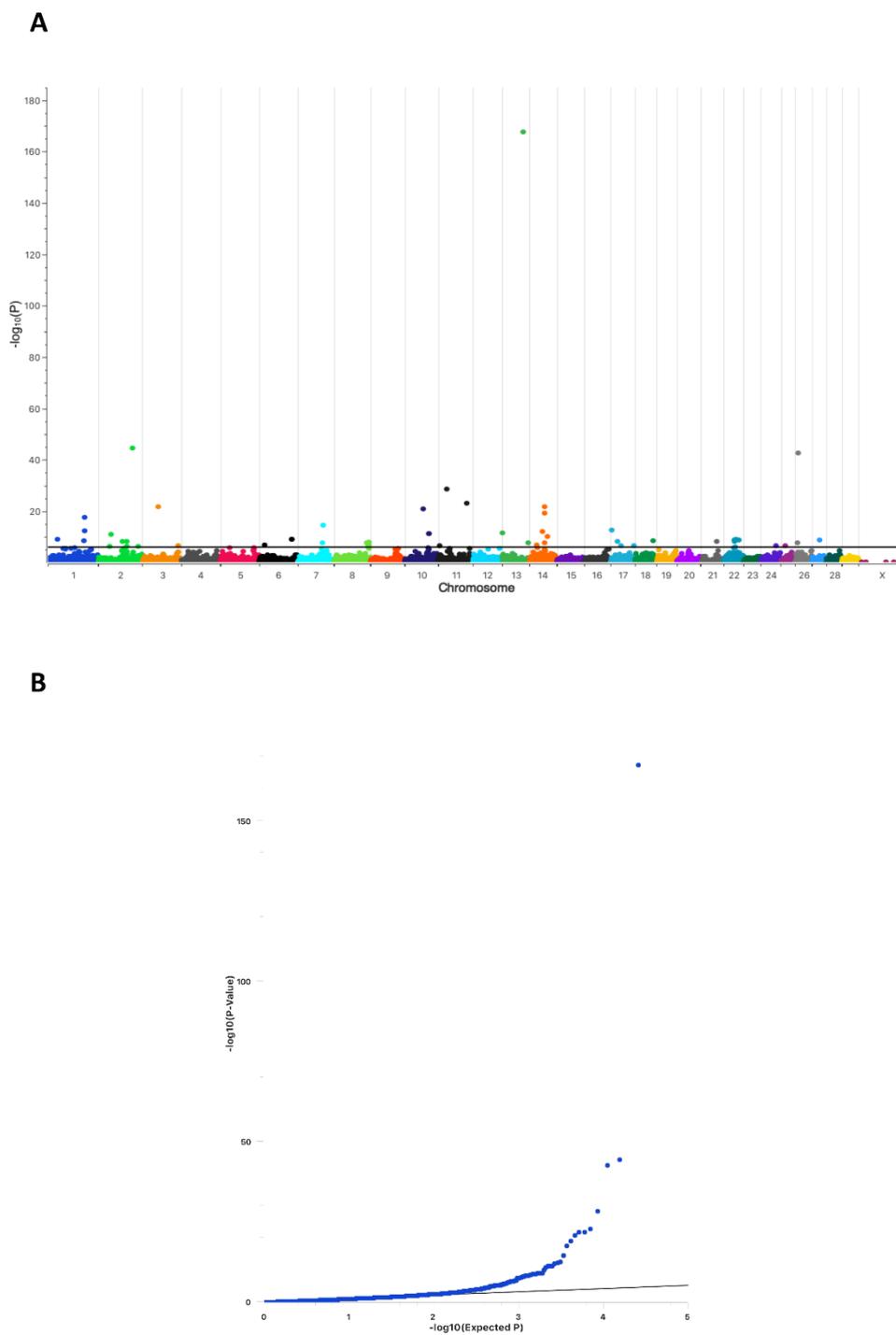


Figure 1. Manhattan plot of the Genome Wide Association Analysis (GWAS) ($n = 289$; 279 controls, 10 cases). **(A)** The plot shows the $-\log_{10}$ -transformed p -values for all SNPs. The black horizontal line represents the genome-wide significance threshold of $-\log_{10}P = 5.9$. **(B)** Quantile-quantile (QQ) plot of the GWAS.

Verification and Validation of the Nonsense Variant g.54429815G>A

(ABHD16B) in the Holstein Population

An initial set of 2072 randomly selected Holstein DNA samples were genotyped to verify and validate the presence of the detected variant in *ABHD16B*. In this set, 2052 wild type (G_G), 20 heterozygous (G_A), and no homozygous (A_A) carrier were detected (HWE $\chi^2= 0.05$). The results proved that the variant was present in the population at a very low frequency. Therefore, a larger cohort of 222,645 HF cattle (208,165 female, 14,480 male) was genotyped while using the bovinSNP50 BeadChip. In this cohort, 810 heterozygous (781 female, 19 male, 10 unknown sex) and no homozygous animals were identified, resulting in a frequency of the variant allele of 0.0018. According to Hardy-Weinberg equilibrium it was not unexpected that no homozygous individuals were detected (HWE $\chi^2= 0.73$). The low allele frequency further supported the data that the nonsense variant in *ABHD16B* was most likely the causative variant for Tarantino's infertility, because sires will be rapidly removed from the breeding population once a sub- or infertility would have been evident during routine fertility testing. Such a selection will efficiently reduce the transmission and spreading of the causative variant. The limited number of heterozygous individuals in the randomly chosen large Holstein cohort prompted us to determine the number of heterozygous sires in the available DNA samples of Tarantino's close male relatives in correlation with their conception ability (NR_{dev}). A total of 34 DNA samples were available and genotyped, resulting in 16 wild type and 18 heterozygous sires (HWE $\chi^2= 4.4$). Within the heterozygous sires, 15 had negative NR_{dev} values (-9 to < 0) and only three sires showed positive NR_{dev} values (0 to 2).

Expression and Tissue Distribution of ABHD16B

ABHD16B codes for a protein of 470 amino acids with a predicted α/β -hydrolase fold domain. The nonsense variant g.54429815G>A causes a premature stop at amino acid position 218 (glutamine residue), resulting in a truncation of 253 C-terminal amino acids and 53.8% of the protein (Figure 2). *In silico* protein sequence comparison of 11 mammals revealed that the glutamine residue (Q) is highly conserved. Due to the truncation 67.4% of the α/β -hydrolase fold domain is missing. Regarding the evolutionary appearance, it is interesting to note that *ABHD16B* first evolved in reptiles

performing internal fertilization. Species with external fertilization, e.g., fish and frogs, do not harbor an *ABHD16B* gene.

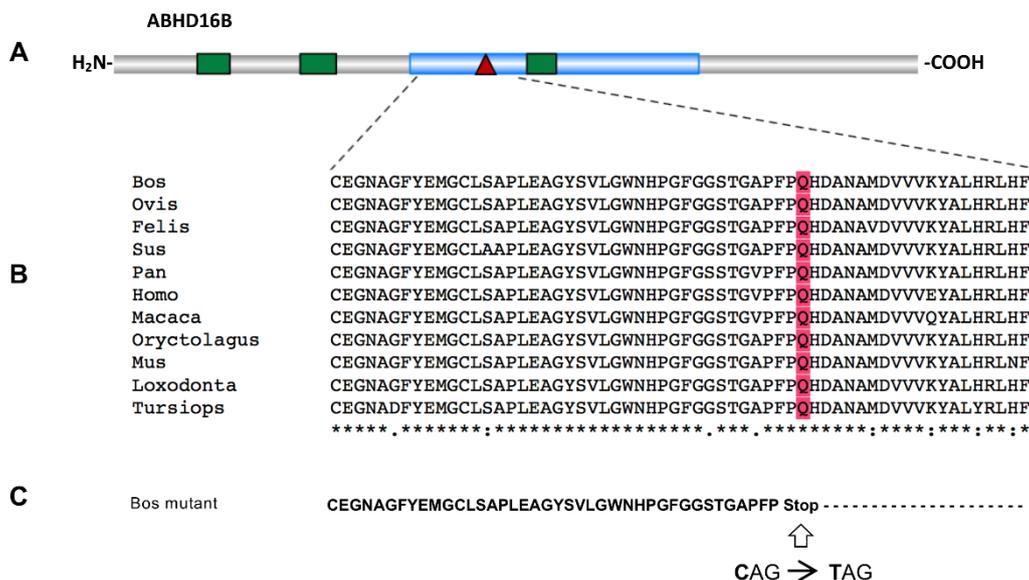


Figure 2. (A) Schematic representation of the ABHD16B protein structure indicates the position of the α/β -hydrolase fold domain (blue) and the transmembrane helices (green), predicted by NCBI Conserved Domains Database and TMbase, respectively. The amino acid position (218) of the nonsense variant leading in a premature stop is marked by a red triangle. (B) The comparative alignment of amino acid sequences of 11 mammals while using Clustal W (178-amino acid position 178–237) is shown. The amino acid position at the truncation site is indicated in red. NCBI protein sequence accession numbers are as follows: Bos (*Bos Taurus*) NP_001033630.1, Ovis (*Ovis aries*) XP_014955258.1, Felis (*Felis catus*) XP_003983341.3, Sus (*Sus scrofa*) XP_020933693.1, Pan (*Pan troglodytes*) XP_003317106.1, Homo (*Homo sapiens*) NP_542189.1, Macaca (*Macaca mulatta*) NP_001180656.1, Oryctolagus (*Oryctolagus cuniculus*) XP_008250767.2, Mus (*Mus musculus*) NP_899004.1, Loxodonta (*Loxodonta Africana*) XP_003421827.1, Tursiops (*Tursiops truncatus*) XP_019806804.1. (C) The amino acid sequence of bovine ABHD16B truncated protein with the stop-gain variant.

ABHD16B Is Expressed in Testis but not in Spermatozoa

Western blotting was used to detect ABHD16B in testis (wild type) and spermatozoa (wild type, heterozygous, and homozygous variant) extracts. Testes of heterozygous and/or homozygous carriers were unavailable due to the low genotype frequencies. However, a limited amount of deep-frozen semen samples of Tarantino and a further not directly related homozygous carrier (C_a) provided from the safety inventory of an AI station were included in the analysis. An ABHD16B specific band was detected in testis of wild type bulls at the expected size of approx. 70 kDa, as shown in Figure 3. Neither in wild type nor in heterozygous or homozygous variant spermatozoa extracts ABHD16B was detected (Figure 3B).

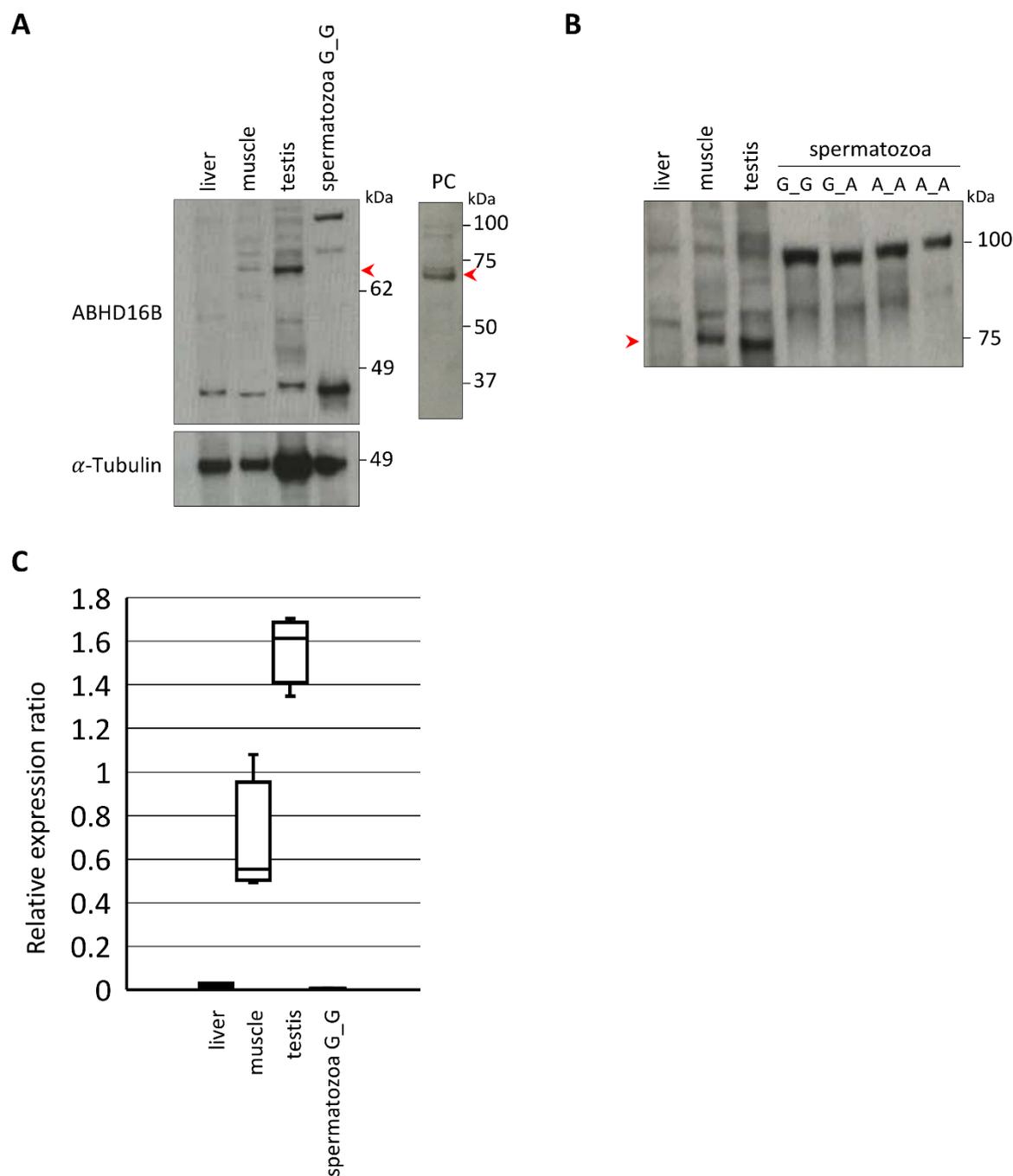


Figure 3. Western blot analysis of ABHD16B protein expression in tissues and spermatozoa. **(A)** ABHD16B protein band (red arrow) detected in wild type testis (G_G) but not in spermatozoa. Using liver as negative control, muscle and human ABHD16B over-expression lysate as positive controls (PC). α -Tubulin used as the loading control. **(B)** ABHD16B (approx. 70 kDa) is absent in spermatozoa of three genotypes (wild type (G_G), heterozygous (G_A) and homozygous carrier (A_A)). Spermatozoa of homozygous carriers were from Tarantino and C_a. **(C)** Box and Whisker plot of relative ABHD16B expression. Areas under curve were determined using ImageJ 1.52k software and relative expression ratios of ABHD16B (%) in liver, muscle, testis, and spermatozoa (G_G) were calculated while using α -Tubulin expression as internal standard. Horizontal lines within boxes indicate median values and whiskers show upper and lower extremes.

Immunohistochemical Analysis Revealed ABHD16B Expression in Testis and Epididymis

Sections of testicular and epididymal tissue samples that were collected at an abattoir were prepared for IHC. The *ABHD16B* genotype of the samples was tested prior to IHC and shown to originate from wild type sires. While using the PAC-ARK antibody, ABHD16B expression was detectable in testicular parenchyma, ductuli efferentes, as well as epididymal tail, body, and head, as shown in Figure 4. Specifically, there is ABHD16B expression in the nucleoplasm of Leydig cells, in the seminiferous tubules and, with variable intensity, in the epithelium of the ductus epididymis. These findings suggest that ABHD16B probably plays a role in spermatogenesis and sperm maturation.

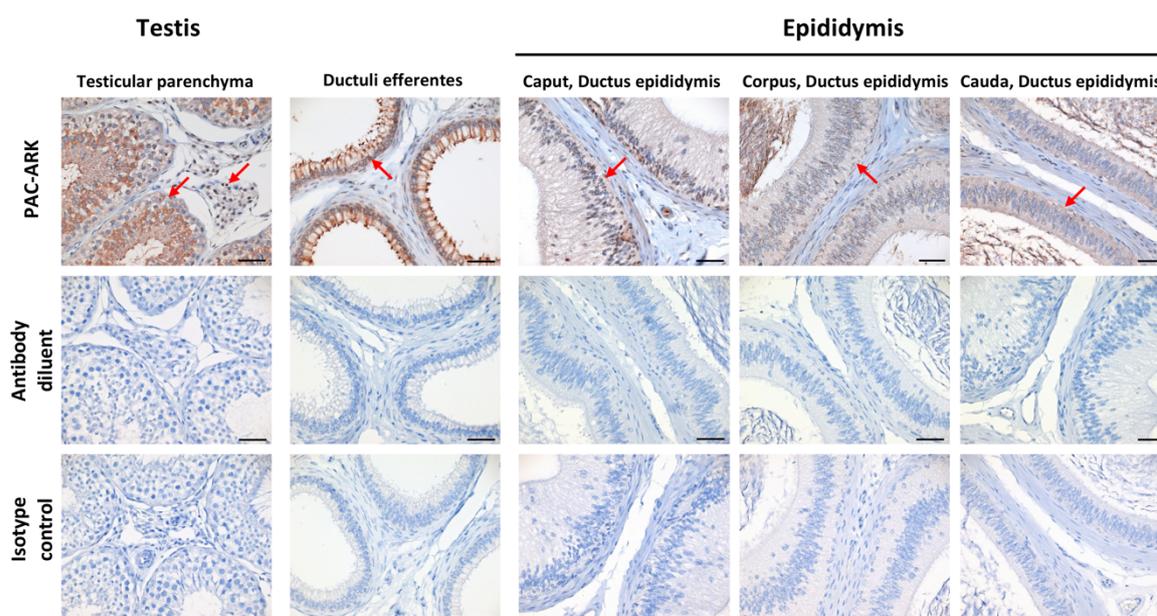


Figure 4. ABHD16B protein detection and localization in testis and epididymis of wild type bull by immunohistochemistry. PAC-ARK antibody was used as primary antibody. Positive staining is indicated with red arrows. When the primary antibody (PAC-ARK) was replaced with antibody diluent and isotype rabbit IgG at the same working dilution, no staining was observed in any of these tissues. Scale bars = 50 μ m.

ABHD16B Is Involved in Lipid Metabolism and Influences Sperm Plasma

Membrane Lipid Composition

We hypothesized that ABHD16B could be involved in plasma membrane lipid biosynthesis, as many members of the α/β -hydrolase superfamily of hydrolytic enzymes are involved in lipid metabolism. Sperm lipidomics of heterozygous and wild type semen samples was performed to interrogate this hypothesis. The number of available semen samples of Tarantino was limited and, therefore, it was decided not to include these valuable samples. However, if ABHD16B would have an effect on lipid biosynthesis during spermatogenesis, this should also be detectable in heterozygous samples. After normalization to 10^6 sperms per sample, no significant difference in the total lipid content between 15 wild type and 15 heterozygous samples was detected (Figure 5).

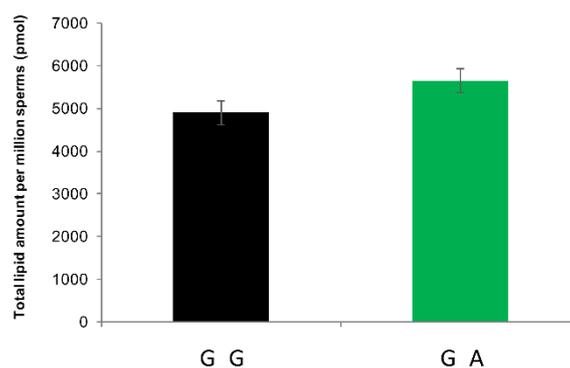


Figure 5. Total lipid amount of 10^6 sperms of each genotype revealed no significant difference after normalization. G_G: wild type; G_A: heterozygous carrier.

However, 10 out of 16 lipid classes showed significant differences (Figure 6A). The majority of different lipids belonged to the classes of diacylglycerols (DAG), glycerophosphocholines (PC, PC O-), ceramides (Cer), and sphingomyelins (SM). The sperms of heterozygous carriers showed significantly decreased amounts of SM and DAG, while PC, PC O-, and Cer were increased (Figure 6A). In total, 99 of 144 lipid species demonstrated significant differences between wild type and heterozygous sperm samples. Eight lipid species significantly decreased ($p_{(BH)} < 0.05$, $\log_2fc < -1$), six of them were DAGs. 25 lipid species significantly increased ($p_{(BH)} < 0.05$, $\log_2fc > 1$), almost half ($n=12$) of them were PCs (Figure 6B). Figure 6C shows the ten most significantly changed lipid species. An important indicator of cell membrane integrity is the PC:PE ratio. As shown in Figure 6D, heterozygous sperms have a significantly

increased the PC:PE ratio. Another sperm membrane structure criterion is the ratio between LPC 22:6 and PC 16:0_22:6, and it also significantly increased in the heterozygous samples (Figure 6E).

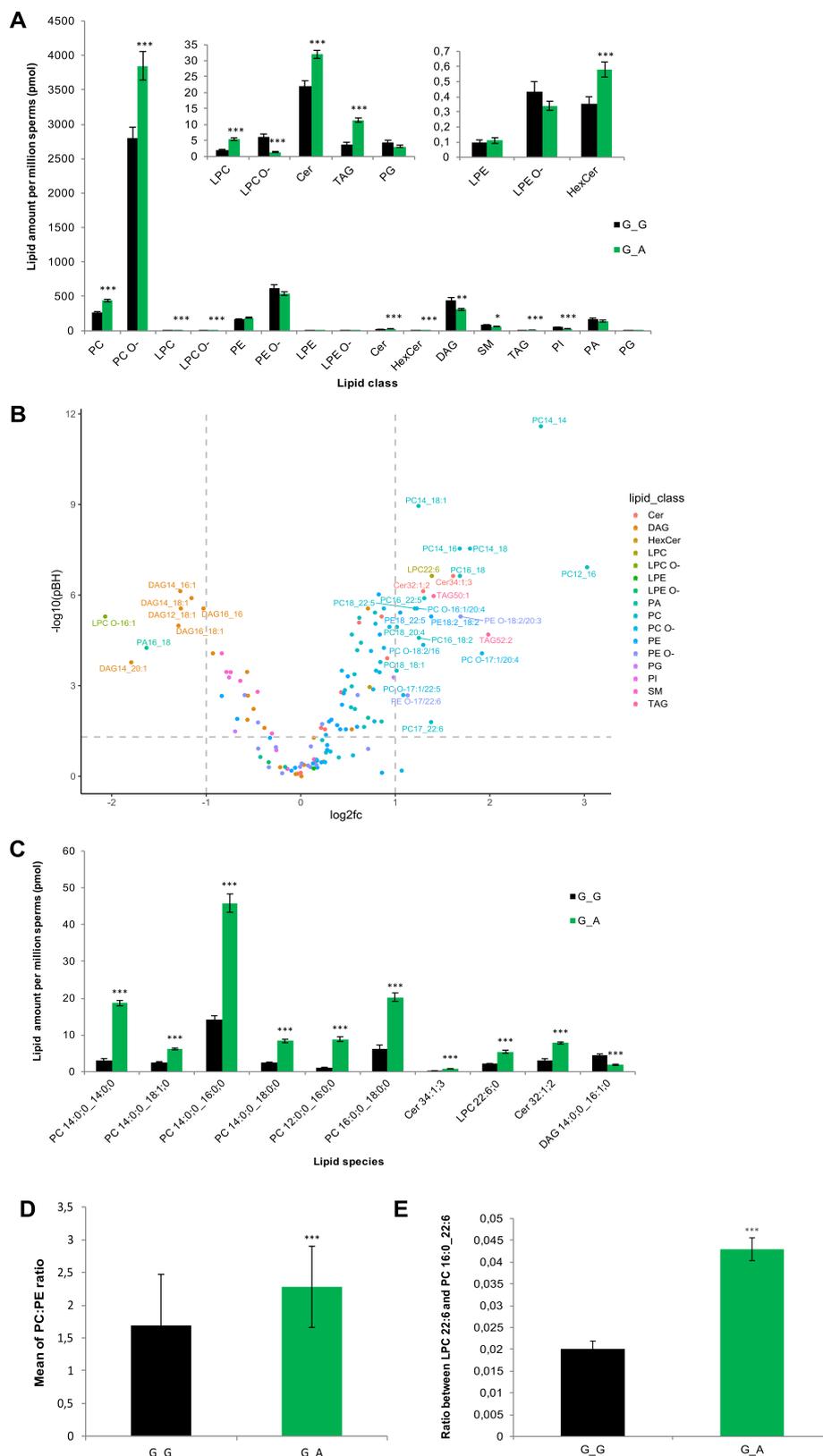


Figure 6. Lipid distribution and variance of spermatozoa in wild type and heterozygous sires. **(A)** Change profile of total lipid content for each lipid class of sperms in two genotypes. **(B)** $-\log_{10}$ of adjusted p -value ($p_{(BH)}$) and mean \log_2 -fold change (G_A vs G_G) of 144 lipid species were plotted, lipid species with $-\log_{10}p_{(BH)} > 1.3$ and $|\log_2\text{fold change}| > 1$ (dashed grey lines are included) were annotated with lipid feature names (simplified without the saturated acyl groups). **(C)** Top 10 most significantly changed lipid species between G_G and G_A bull sperm cells. **(D)** Bar chart shows the mean of PC:PE ratio in G_G and G_A bull sperms. **(E)** Bar chart shows the ratio between LPC 22:6 and PC 16:0_22:6 in two genotypes. G_G: wild type; G_A: heterozygous. Differential changes were tested by the Mann–Whitney U test or Student’s t-test. Data are presented as mean \pm SEM; $n = 15/\text{group}$; FDR adjusted p -values are indicated: * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

Discussion

Male infertility is a complex multifactorial idiopathic, congenital, or acquired heterogeneous disease [37,38]. In men, genetic factors predominantly cause idiopathic conditions contributing to 30–40% of male infertility [39]. However, up to now, for only three genes, *i.e.*, *NR5A1*, *DMRT1*, and *TEX11*, associations with male infertility have been evidenced in independent biological and functional studies [40]. The same number of genes has been identified in cattle causing bull sub- or infertility so far, *i.e.*, *FSHB* (BTA15, 61.7 Mb), *TMEM95* (BTA19, 27.6 Mb), and *ARMC3* (BTA13, 24.3 Mb) [27–30]. Although the exact chromosomal positions of these genes differ from precise infertility associated chromosomal regions that have been identified either by QTL studies, GWAS using SCR as parameter, or whole exome sequencing, they are located on the same chromosomes [24,26,41]. Regarding the location of *ABHD16B* on BTA13, it is noteworthy that not only *ARMC3* is located on the same chromosome, but also QTLs for percentage of normal sperms (68.18 cM), male fertility (43.76 cM), and non-return rate (EBV) (85.19 cM) have been mapped to BTA13 [42–44]. One region explaining roughly 0.6% of the genetic variance of SCR was detected on BTA13 from position 58,456,868–59,951,247 harboring two potential candidate genes for male fertility, *i.e.*, *CTCFL* and *SPO11* [45]. This region is located approximately 4 Mb downstream of *ABHD16B*. The closest SNP identified by whole exome sequencing was located on BTA13 at position 53,691,419 within the *SIRPA* gene only 738,396 bp

upstream of the nonsense variant in *ABHD16B*. In a further GWAS using a much larger dataset (11.5 k Holstein bulls) and higher density SNP chip (about 300 k), five markers with marked dominance effects were detected, one of them being located on BTA13 (13:g.60263194A>C; rs41701032) [46]. Hence, the molecular genetic data that are published elsewhere are well in agreement with our findings.

The identification of *ABHD16B* as an associated causative gene for bull infertility also allowed for us to elucidate its biochemical function. Except that *ABHD16B* belongs to a large protein superfamily of catalytic enzymes harboring an α/β -hydrolase domain and is predominantly expressed in Leydig cells of the testis, nothing was known regarding its biochemical or physiological function so far [31]. The methylation of *ABHD16B* was reported to be associated with chronic obstructive pulmonary disease (COPD) and aberrant methylation patterns were identified in infertile man [32,36]. In proteome studies, 11 members of the ABHD family (1, 2, 5, 6, 10, 11, 12, 13, 14B, 16A, 17A, 17B) have been detected in testis or spermatozoa [47,48]. Human *ABHD2* participates in sperm hyperactivation as a lipid hydrolase through depleting endocannabinoid 2-arachidonoylglycerol (2-AG), an inhibitor of sperm calcium channel (CatSper) [49]. However, in most cases, their exact role remains elusive.

Our data show that *ABHD16B* is involved in lipid biosynthesis of DAGs. According to the *ABHD16B* molecular structure, it is supposed to participate in lipid metabolism, like other ABHD family members, which could contribute to sperm maturation. Sperm lipidomics of two different genotypes was performed to confirm this. Sperm lipid composition changes during their maturation through the epididymis, the percentage of SM in spermatozoa increased [50,51]. SM is synthesized by the combination of Cer and phosphorylcholine from PC. During this reaction, DAG is produced as a by-product [52]. Our data demonstrated that SM and DAG were significantly decreased, while PC and Cer were significantly increased in heterozygous spermatozoa. This implies that *ABHD16B* might be involved in the lipid biosynthesis of DAG, which influences SM synthesis in the later process. On the other hand, increased levels of PCs in heterozygous sperms could also result from an inhibited degradation from PC to DAG and phosphorylcholine. Cer increased correspondingly without enough phosphorylcholine combined to synthesize SM.

The presence of ABHD16B in the epididymis, as shown by IHC, suggests a role in the lipid metabolism of DAG and SM during sperm maturation. DAG also influences the synthesis of 2-AG, which is an inhibitor of sperm calcium channel (CatSper) preventing sperm hyperactivation. DAG is hydrolyzed to 2-AG by diacylglycerol lipase (DAGL) and, hence, decreased DAG levels in homozygous carrier sperms could result in an insufficient amount of 2-AG leading to a premature capacitation [53]. Furthermore, this effect could be enhanced by the lack of SM. On the other hand, the accumulation of PC and Cer could also interfere with the fertilization capacity. For instance, increased PC concentrations in chicken sperms were reported to be negatively associated with fertility during aging [54]. It has also been observed that imbalanced lipid homeostasis of PC and SM caused sperm membrane instability and infertility in knockout mice [55]. The final sperm lipid composition is formed during epididymal maturation, which results in a decreased amount of cholesterol, PS, CL, PE, and PI, and an increase in PC and DAG. The amount of PI, PC, and DAG was significantly different between the wild type and heterozygous variant spermatozoa (Figure 6A), indicating a potential role of ABHD16B in sperm maturation.

Another impact of ABHD16B on lipid metabolism can be seen in the increased PC:PE ratio. Abnormal PC:PE ratios affect membrane permeability, fluidity, and integrity [56,57]. In cells that have abundant unsaturated fatty acids, such as spermatozoa, LPC is normally regarded as a marker of sperm membrane quality and oxidative stress. The increase of LPC content in the deteriorated membrane of spermatozoa indicates affected acrosome reaction, and an increased ratio between LPC 22:6 and PC 16:0/22:6 was observed in human spermatozoa with impaired membrane [58]. The ratio was also significantly enhanced in the heterozygous samples analysed here. Furthermore, LPC 22:6 is a reliable marker of spermatozoa lipid oxidation [59]. A significant increased concentration of LPC 22:6 was also found in heterozygous carrier samples. This could result in a higher oxidized state or membrane damaged level in contrast to wild type sperms. The lipidomics analysis clearly showed that the loss of ABHD16B function has a profound effect on sperm plasma membrane lipid composition. Therefore, in analogy with experiments in humans and mice, it can be hypothesized that the altered lipid composition of the *ABHD16B* homozygous carrier sperms interferes with the fertilization ability.

Materials and Methods

Ethical Statement

EDTA blood samples of cattle were taken for routine parentage control exclusively by local veterinarians. The Lower Saxony State Office for Consumer Protection and Food Safety approved the collection of samples (33.19-42502-05-17A196), according to §8a Abs. 1 Nr. 2 of the German Animal Protection Law.

Genome Wide Association Analysis (GWAS)

The conception ability of sires (non-return rate, NR_{dev}) was calculated based on the latest three daughter proven service-sire age groups (2019: A.I. sires born 2014–2016; 0% deviation). The NR_{dev} is expressed in %-deviation on the original non-return-rate scale. Sires in the breeding population with NR_{dev} values of approx. $\pm 2\%$ are scored as average (for more information see https://www.vit.de/fileadmin/DE/Zuchtwertschaetzung/Zws_Bes_eng.pdf). Data of NR_{dev} deviations of service-sires were provided by VIT (<https://www.vit.de/en/>).

For GWAS 279 sires of the current breeding population were randomly chosen as presumably fertile controls. As cases 10 sires (including Tarantino, $NR_{dev} = -29$) with NR_{dev} between -29 and -2 were selected (Table 1). The 289 samples were genotyped while using the Illumina BovineSNP50 or MD BeadChip. The chips were processed on a HiScan SQ and iScan System (Illumina GmbH, Munich, Germany) and raw data were converted using GenomeStudio Software (Illumina GmbH, Munich, Germany). Final reports were imported into SVS 8.8.3 for MacOSX (Golden Helix Inc. Bozeman, MT, USA). Prior to GWAS data were filtered while using a call rate < 0.95 , number of alleles > 2 , minor allele frequency (MAF) < 0.05 , and Fisher's HWE < 0.001 (based on controls) as marker dropping criteria. LD pruning was performed with a window size of 100 and increments of 5. R^2 -LD statistics with a threshold of 0.5 while using Cochran-Mantel-Haenszel (CHM) as computation method was applied. After filtering, 38,671 markers remained for further analysis. GWAS was done using a multi-locus mixed model (MLMM) while applying an additive genetic model with correction for male X-chromosomal hemizyosity [60,61]. The associations were regarded as

statistically significant above a Bonferroni threshold of $-\log_{10}P = 5.9$ ($p = 0.05$). The associations of markers ($-\log_{10}P$ -value, y-axis) were plotted against their chromosomal positions (UMD3.1.1, x-axis).

Next Generation Sequencing of Tarantino and Its Parents

Tarantino and its parents were sequenced on a HiSeq2500 System (Illumina GmbH, Munich, Germany), resulting in approx. 10^9 total reads per sample. Low quality (average phred quality < 15) and single reads were removed, resulting in approx. 9.4×10^8 per sample. Mapping to the bovine reference genome sequence (UMD3.1.1) was done while using BWA [62]. PCR duplicates were removed using Picard (<http://broadinstitute.github.io/picard/>). After read mapping, alignment and refinement approx. 7.8×10^8 reads remained per sample, corresponding to an average depth of coverage of approx. 46x (mean insert size 360 bp). A total of 9,315,126 SNPs and 1,439,972 indels were called using GATK Haplotype Caller [63]. SNP & Variation Suite 8.8.3 (Golden Helix Inc., Bozeman, MT, USA) was used for further analysis. SNPs and indels were set to missing with read depth ≤ 10 , genotype quality ≤ 15 , alt read ratios for Ref_Ref ≥ 0.15 , Ref_Alt outside 0.3 and 0.7, Alt_Alt ≤ 0.85 , and according to their inheritance pattern (Tarantino = Alt_Alt, parents Alt_Ref). After this filtering, 307,898 SNPs and 604 indels remained. A final filtering was done while using SNPs and indels only in annotated and verified mRNA transcripts, including splice donor and acceptor distances of 2 bp, splice region exonic distances of 3 bp and splice region intronic distances of 8 bp, resulting in 78,472 SNPs and 125 indels.

Genotyping of SNP rs468948776 (ABHD16B)

The nonsense variant in *ABHD16B* was genotyped while using fluorescence resonance energy transfer (FRET) analysis on a LightCycler 480 (Roche Life Science, Mannheim, Germany). The DNA concentrations were measured using NanoDrop ND-1000 spectrophotometer (PEQLAB Biotechnologie GmbH, Erlangen, Germany). Conventional PCR primers were designed using the online program Primer3 (<http://bioinfo.ut.ee/primer3-0.4.0/>). The FRET primers were designed with MeltCalc Software [64,65]. Table 2 lists FRET primers and probes.

Table 2. Fluorescence resonance energy transfer (FRET) primers and probes used for genotyping of *ABHD16B* variant.

Gene	Primer name	Sequence (5'→3')	Probe Name	Sequence (5'→3')
<i>ABHD16B</i>	ABHD16B_FRET_f	ACCCGGGCTTCGGGGGCAGC	ABHD16B_FRET_Pro	[Cy5]CGTTCCTCAGCATGATG[Phos]
	ABHD16B_FRET_r	GCGTACTTGACCACCACGTC	ABHD16B_FRET_Anc	GGGGCAGCACGGGCG[Fic]

SNP rs468948776 (*ABHD16B*) was amplified in a total volume of 25 μ L, including 20 ng DNA, 10 μ mol forward and reverse primer each, 10 μ mol probe and anchor (Sigma-Aldrich, Taufkirchen, Germany) each, 1 \times GC-RICH solution, 1 \times PCR reaction buffer (including 20 mM MgCl₂), 100 μ mol dNTPs and FastStart Taq Polymerase (1U; Qiagen, Hilden Germany) for 34 cycles at 95 °C for 15 S, 60 °C for 20 S, and 72 °C for 20 S. The melting curves were done using the following program: 95 °C for 30 S, 37 °C for 30 S, 95 °C continuous acquisition mode (2/°C), ramp rate 0.29 °C/S, followed by 37 °C for 30 S.

Western Blotting

Immunoblotting on cryopreserved semen specimens of one wild type (G_G), one heterozygous carrier (G_A), two homozygous affected (A_A; Tarantino, C_a), and testis, muscle, and liver samples of wild type bulls were prepared. The semen samples of sire C_a were provided from the safety-inventory of an AI station. Human ABHD16B over-expression lysate (NM_080622, OriGene, Rockville, Maryland, USA) was used as a positive control. Frozen semen samples were thawed at 37 °C in a water bath for 30 S., followed by 3 \times washes with phosphate-buffered saline (PBS; Invitrogen/ThermoFisher Scientific) and lysed in cold RIPA buffer (Sigma, R0278, St. Louis, MO, USA). Protease inhibitor (Roche, Cat. No.04693159001, Mannheim, Germany) and phosphatase inhibitor (Roche, Cat. No. 04906845001, Germany) were added to RIPA buffer in advance. The samples were incubated for 1h at 4 °C and centrifuged at 16,000 \times g for 20 min. at 4 °C. An additional homogenization of tissue samples with MagNA lyser green beads (Roche Life Science, Mannheim, Product No. 03358941001, Germany) was carried out followed by an incubation for 2h at 4 °C and then centrifuged at 16,000 \times g for 20 min. at 4 °C. Protein quantification was performed by Bradford method with the dye reagent concentrate (Bio-Rad, Cat. No. 5000006, Munich, Germany).

After denaturation (10 min. at 70 °C) in LDS sample buffer with 5% 2-mercaptoethanol, equal amounts of protein were loaded to SDS-PAGE (8% Bis-Tris Plus gel, ThermoFisher Scientific, Cat. No. NW00087BOX, USA). After electrophoresis at 15V for 1h, the proteins were transferred onto nitrocellulose membranes (Sigma, Cat. No. 10600098, Germany) with semi-dry blotter (Brenzel Bioanalytik, Lahntal, Germany). Membranes were blocked with 5% non-fat dry milk in TBS-T (0.1% Tween) overnight at 4 °C and then incubated with primary antibodies for 1 h at room temperature, followed by incubation with the secondary antibodies at room temperature for 1 h. Subsequently, the membranes were incubated with an ECL detection reagent (GE Healthcare, Product No. RPN2109, Little Chalfont, UK) and then exposed to X-ray films (GE Healthcare, Product No. 28906836, Tokyo, Japan) for detection.

A customized bovine ABHD16B primary antibody, affinity purification PAC-DFR (Davids Biotechnologie GmbH, Regensburg, Germany, 1 µg/mL dilution) was used. Goat Anti-Mouse IgG (H + L)-HRP (Bio-Rad, Munich, Germany; 1:10,000 dilution) and Goat Anti-Rabbit IgG (H + L)-HRP (Bio-Rad, Germany; 1:10,000 dilution) were the secondary antibodies. Anti- α -Tubulin (Sigma, T9026; 1: 2500 dilution) was used as the loading control.

The quantification of ABHD16B Western blots was done using ImageJ 1.52k software [66]. Areas under curve of ABHD16B specific bands were determined for liver, muscle, testis and spermatozoa (G_G). Relative expression ratios (%) were calculated with α -tubulin as the internal standard and plotted as Box and Whisker plot.

Immunohistochemistry of Testes

Testicular and epididymal tissues were obtained from freshly slaughtered wild type Holstein cattle and they were immediately fixed in 4% formaldehyde for 48 h. Immunohistochemistry (IHC) was performed on paraffin-embedded sections, including testicular parenchyma, as well as ductuli efferentes, epididymal head, corpus, and tail with efferent ducts and epididymal duct, respectively. The primary polyclonal antibody was directed against the PAC-ARK peptide and it was generated in the rabbit according to standard protocols (Davids Biotechnologie GmbH, Regensburg, Germany). IHC was

performed in an automated immunostaining system (Discovery XT, Roche Diagnostics GmbH, Mannheim, Germany) at a dilution of 1:1000 while using the SABC (streptavidin-biotin-complex) method, mild EDTA (ethylenediaminetetraacetic acid) pretreatment, and DAB (diaminobenzidine tetrahydrochloride) for signal detection (DAB Map Kit, Roche Diagnostics GmbH, Mannheim, Germany). A rabbit IgG isotype control (ABIN3023746, antibodies-online GmbH, Aachen, Germany) was included at the same concentration as the primary antibody for confirmation of primary antibody specificity. Additionally, pure antibody diluent instead of primary antibody was applied to the control sections for an evaluation of non-specific binding of the secondary antibody.

Lipidomics of Wild Type and Heterozygous Spermatozoa

Semen Collection for Lipidome Analysis

Wild type and heterozygous fresh semen samples were prepared for lipidome analysis. Three independent fresh ejaculates were collected from a heterozygous bull and five technical replicates were produced by dilution from each sample. Wild type semen samples were flushed from the epididymal tail of four unrelated bulls and a total of 15 technical replicates were generated by dilution. In the epididymal tail, spermatozoa are matured and the lipid composition is equivalent to ejaculated spermatozoa [67,68]. The samples were washed twice in Dulbecco's phosphate-buffered saline (D-PBS) without magnesium and calcium and centrifugated at $1000\times g$ for 5 min. at 4 °C. The cells were resuspended in D-PBS to a final concentration of approximately three million–eight million cells/mL. Cell density was determined in an improved Neubauer counting chamber (Marienfeld GmbH, Lauda-Königshofen, Germany).

Lipid Extraction for Mass Spectrometry Lipidomics

Mass spectrometry-based lipid analysis was performed by Lipotype GmbH (Dresden, Germany), as described [69]. The lipids were extracted while using a two-step chloroform/methanol procedure [70]. The samples were spiked with internal lipid standard mixture containing: cardiolipin 16:1/15:0/15:0/15:0 (CL), ceramide 18:1;2/17:0 (Cer), diacylglycerol 17:0/17:0 (DAG), hexosylceramide 18:1;2/12:0 (HexCer), lyso-phosphatidate 17:0 (LPA), lyso-phosphatidylcholine 12:0 (LPC), lyso-

phosphatidylethanolamine 17:1 (LPE), lyso-phosphatidylglycerol 17:1 (LPG), lyso-phosphatidylinositol 17:1 (LPI), lyso-phosphatidylserine 17:1 (LPS), phosphatidate 17:0/17:0 (PA), phosphatidylcholine 17:0/17:0 (PC), phosphatidylethanolamine 17:0/17:0 (PE), phosphatidylglycerol 17:0/17:0 (PG), phosphatidylinositol 16:0/16:0 (PI), phosphatidylserine 17:0/17:0 (PS), cholesterol ester 20:0 (CE), sphingomyelin 18:1;2/12:0;0 (SM), and triacylglycerol 17:0/17:0/17:0 (TAG). After extraction, the organic phase was transferred to an infusion plate and dried in a speed vacuum concentrator. First step dry extract was re-suspended in 7.5 mM ammonium acetate in chloroform/methanol/propanol (1:2:4, *V:V:V*) and second step dry extract in 33% ethanol solution of methylamine in chloroform/methanol (0.003:5:1; *V:V:V*). All liquid handling steps were performed while using Hamilton Robotics STARlet robotic platform with the Anti Droplet Control feature for organic solvents pipetting.

MS Data Acquisition

The samples were analyzed by direct infusion on a QExactive mass spectrometer (Thermo Scientific, Osterode am Harz, Germany) equipped with a TriVersa NanoMate ion source (Advion Biosciences, Ithaca, NY, USA). Samples were analyzed in both positive and negative ion modes with a resolution of $Rm/z=200=280,000$ for MS and $Rm/z=200=17,500$ for MSMS experiments, in a single acquisition. MSMS was triggered by an inclusion list that encompasses corresponding MS mass ranges scanned in 1 Da increments [71]. MS and MSMS data were both combined to monitor CE, DAG, and TAG ions as ammonium adducts; PC, PC O⁻, as acetate adducts; and, CL, PA, PE, PE O⁻, PG, PI, and PS as deprotonated anions. MS only was used to monitor LPA, LPE, LPE O⁻, LPI, and LPS as deprotonated anions; Cer, HexCer, SM, LPC, and LPC O⁻ as acetate adducts.

Data Analysis and Post-Processing

The data were analyzed with in-house developed lipid identification software based on LipidXplorer [72,73]. Data post-processing and normalization were performed while using an in-house developed data management system. Only lipid identifications with a signal-to-noise ratio >5, and a signal intensity five-fold higher than in corresponding blank samples were considered for further data analysis.

The total lipid amount occurring in each sperm sample were pre-tested to ensure that optimal amounts are used to achieve the greatest analysis quality and result comparability, despite the broad dynamic range of our analytical methods. Afterwards, the initially detected total lipid amount per sample was normalized to one-million sperms. A significant difference between normalized total lipid amount of wild type and heterozygous samples was analyzed.

A 70% occupational threshold was applied for data filter, valid data in more than 10 samples for each genotype were selected, NAs were replaced with zeros. Afterwards, lipid data that were present in both genotypes were chosen for further analysis. In total, 16 lipid classes with 144 lipid species were evaluated, and the data were analyzed in terms of lipid class and species separately. Shapiro–Wilk Test was used for normal distribution detection [74]. Significant difference analysis was performed with Mann–Whitney U-test or two-tailed t-test, depending on the normal distribution results by SPSS 16.0. PC:PE and (LPC 22:6):(PC 16:0_22:6) ratios were also checked for significant difference. Benjamini & Hochberg method was used for p -value adjustment of multiple testing [75] with R version 3.5.1, $p_{(BH)} < 0.05$ were considered to be statistically significant (Table S1). The comparison results of lipid classes and 10 most significantly changed lipid species between wild-type and rs468948776 heterozygous samples were demonstrated in histograms (data are presented with mean \pm standard error of mean). 144 lipid species were plotted with Y-axis of adjusted values ($-\log_{10}p_{BH}$) against X-axis of \log_2 fold change (heterozygous vs wild type).

Conclusions

We have identified a nonsense mutation in the bovine *ABHD16B* gene as a potential causative protein-altering variant for male infertility in Holstein cattle. This made it possible to elucidate the so far unknown physiological and biochemical role of ABHD16B in lipid biosynthesis, spermatogenesis, and fertilization. Our findings could also have implications on further elucidating a novel genetic cause for human male infertility, due to the fact that a number of deleterious variants, e.g., missense, frameshift, indels and one stop-gain variant in the human *ABHD16B* gene have been reported to the human ENSEMBL database.

Supplementary Materials: Supplementary materials can be found at <https://www.mdpi.com/1422-0067/21/2/627/s1>.

Author contributions: S.S., F.X., M.B., S.B., T.M., M.H. and C.W. conducted the experiments. W.W. collected samples. S.S., F.X., E.S. and B.B. analyzed the data. B.B. planned and supervised the project. B.B. and S.Z. supervised S.S. and F.X. S.S., F.X., M.B. and B.B. wrote the manuscript. All authors read and approved the final manuscript.

Funding: This research received no external funding.

Acknowledgments: The authors are grateful to S. Pach for expert technical assistance. Masterrind is thanked for providing semen samples of sires. The Association for Bioeconomy Research (FBF), German Livestock Association (BRS) and IT-Solutions for Animal Production (VIT) are thanked for support and pedigree data. S. Shan and F. Xu are fellows of the China Scholarship Council (CSC). We acknowledge support by the German Research Foundation and the Open Access Publications of the University of Goettingen.

Conflicts of interest: The authors declare no conflict of interest.

References

1. Sakaguchi, M. Practical aspects of the fertility of dairy cattle. *J. Reprod. Dev.* **2011**, *57*, 17–33.
2. Veerkamp, R.F.; Beerda, B. Genetics and genomics to improve fertility in high producing dairy cows. *Theriogenology* **2007**, *68* (Suppl. 1), S266–S273.
3. Lucy, M.C. Fertility in high-producing dairy cows: Reasons for decline and corrective strategies for sustainable improvement. *Soc. Reprod. Fertil. Suppl.* **2007**, *64*, 237–254.
4. Olynk, N.J.; Wolf, C.A. Economic analysis of reproductive management strategies on US commercial dairy farms. *J. Dairy Sci.* **2008**, *91*, 4082–4091.
5. Harstine, B.R.; Utt, M.D.; DeJarnette, J.M. Review: Integrating a semen quality control program and sire fertility at a large artificial insemination organization. *Animal* **2018**, 1–12. doi:10.1017/S1751731118000319
6. Yaniz, J.L.; Soler, C.; Alquezar-Baeta, C.; Santolaria, P. Toward an integrative and predictive sperm quality analysis in *Bos taurus*. *Anim. Reprod. Sci.* **2017**, *181*, 108–114.
7. Morrell, J.M.; Nongbua, T.; Valeanu, S.; Lima Verde, I.; Lundstedt-Enkel, K.; Edman, A.; Johannisson, A. Sperm quality variables as indicators of bull fertility may be breed dependent. *Anim. Reprod. Sci.* **2017**, *185*, 42–52.
8. Puglisi, R.; Gaspa, G.; Balduzzi, D.; Severgnini, A.; Vanni, R.; Macciotta, N.; Galli, A. Genomewide analysis of bull sperm quality and fertility traits. *Reprod. Domest. Anim.* **2016**, *51*, 840–843.
9. Kastelic, J.P. Understanding and evaluating bovine testes. *Theriogenology* **2014**, *81*, 18–23.
10. Amann, R.P.; DeJarnette, J.M. Impact of genomic selection of AI dairy sires on their likely utilization and methods to estimate fertility: A paradigm shift. *Theriogenology* **2012**, *77*, 795–817.
11. Kealey, C.G.; MacNeil, M.D.; Tess, M.W.; Geary, T.W.; Bellows, R.A. Genetic parameter estimates for scrotal circumference and semen characteristics of Line 1 Hereford bulls. *J. Anim. Sci.* **2006**, *84*, 283–290.
12. Feugang, J.M.; Rodriguez-Osorio, N.; Kaya, A.; Wang, H.; Page, G.; Ostermeier, G.C.; Topper, E.K.; Memili, E. Transcriptome analysis of bull spermatozoa: Implications for male fertility. *Reprod. Biomed. Online* **2010**, *21*, 312–324.
13. Selvaraju, S.; Parthipan, S.; Somashekar, L.; Kolte, A.P.; Krishnan Binsila, B.; Arangasamy, A.; Ravindra, J.P. Occurrence and functional significance of the transcriptome in bovine (*Bos taurus*) spermatozoa. *Sci. Rep.* **2017**, *7*, 42392.
14. Lalancette, C.; Thibault, C.; Bachand, I.; Caron, N.; Bissonnette, N. Transcriptome analysis of bull semen with extreme nonreturn rate: Use of suppression-subtractive hybridization to identify functional markers for fertility. *Biol. Reprod.* **2008**, *78*, 618–635.
15. Samanta, L.; Swain, N.; Ayaz, A.; Venugopal, V.; Agarwal, A. Post-Translational Modifications in sperm Proteome: The Chemistry of Proteome diversifications in the Pathophysiology of male factor infertility. *Biochim. Biophys. Acta* **2016**, *1860*, 1450–1465.
16. Mohanty, G.; Swain, N.; Samanta, L. Sperm Proteome: What Is on the Horizon? *Reprod. Sci.* **2015**, *22*, 638–653.

17. Legare, C.; Droit, A.; Fournier, F.; Bourassa, S.; Force, A.; Cloutier, F.; Tremblay, R.; Sullivan, R. Investigation of male infertility using quantitative comparative proteomics. *J. Proteome Res.* **2014**, *13*, 5403–5414.
18. Gan, H.; Cai, T.; Lin, X.; Wu, Y.; Wang, X.; Yang, F.; Han, C. Integrative proteomic and transcriptomic analyses reveal multiple post-transcriptional regulatory mechanisms of mouse spermatogenesis. *Mol. Cell. Proteom.* **2013**, *12*, 1144–1157.
19. Panner Selvam, M.K.; Agarwal, A. Update on the proteomics of male infertility: A systematic review. *Arab J. Urol.* **2018**, *16*, 103–112.
20. Abdollahi-Arpanahi, R.; Morota, G.; Penagaricano, F. Predicting bull fertility using genomic data and biological information. *J. Dairy Sci.* **2017**, *100*, 9656–9666.
21. Suchocki, T.; Szyda, J. Genome-wide association study for semen production traits in Holstein-Friesian bulls. *J. Dairy Sci.* **2015**, *98*, 5774–5780.
22. Taylor, J.F.; Schnabel, R.D.; Sutovsky, P. Review: Genomics of bull fertility. *Animal* **2018**, *12*, s172–s183.
23. Penagaricano, F.; Weigel, K.A.; Khatib, H. Genome-wide association study identifies candidate markers for bull fertility in Holstein dairy cattle. *Anim. Genet.* **2012**, *43* (Suppl. 1), 65–71.
24. Han, Y.; Penagaricano, F. Unravelling the genomic architecture of bull fertility in Holstein cattle. *BMC Genet.* **2016**, *17*, 143.
25. Ogorevc, J.; Dovc, P.; Kunej, T. Comparative genomics approach to identify candidate genetic loci for male fertility. *Reprod. Domest. Anim.* **2011**, *46*, 229–239.
26. Whiston, R.; Finlay, E.K.; McCabe, M.S.; Cormican, P.; Flynn, P.; Cromie, A.; Hansen, P.J.; Lyons, A.; Fair, S.; Lonergan, P.; et al. A dual targeted beta-defensin and exome sequencing approach to identify, validate and functionally characterise genes associated with bull fertility. *Sci. Rep.* **2017**, *7*, 12287.
27. Pausch, H.; Kolle, S.; Wurmser, C.; Schwarzenbacher, H.; Emmerling, R.; Jansen, S.; Trottmann, M.; Fuerst, C.; Gotz, K.U.; Fries, R. A nonsense mutation in TMEM95 encoding a nondescript transmembrane protein causes idiopathic male subfertility in cattle. *PLoS Genet.* **2014**, *10*, e1004044.
28. Pausch, H.; Venhoranta, H.; Wurmser, C.; Hakala, K.; Iso-Touru, T.; Sironen, A.; Vingborg, R.K.; Lohi, H.; Soderquist, L.; Fries, R.; et al. A frameshift mutation in ARMC3 is associated with a tail stump sperm defect in Swedish Red (*Bos taurus*) cattle. *BMC Genet.* **2016**, *17*, 49.
29. Fernandez-Fuertes, B.; Laguna-Barraza, R.; Fernandez-Gonzalez, R.; Gutierrez-Adan, A.; Blanco-Fernandez, A.; O'Doherty, A.M.; Di Fenza, M.; Kelly, A.K.; Kolle, S.; Lonergan, P. Subfertility in bulls carrying a nonsense mutation in transmembrane protein 95 is due to failure to interact with the oocyte vestments. *Biol. Reprod.* **2017**, *97*, 50–60.
30. Dai, L.; Xu, Y.; Yu, W.; Liu, S.; Gao, Y.; Zhang, L.; Yuan, B.; Chen, J.; Ma, T.; Zhang, J. Naturally occurring genetic mutations in the 5'-upstream regulatory region of bovine FSHB generate a novel cis-regulatory element that affects its expression. *Anim. Genet.* **2015**, *46*, 693–696.
31. Lord, C.C.; Thomas, G.; Brown, J.M. Mammalian alpha beta hydrolase domain (ABHD) proteins: Lipid metabolizing enzymes at the interface of cell signaling and energy metabolism. *Biochim. Biophys. Acta* **2013**, *1831*, 792–802.

32. Wan, E.S.; Qiu, W.; Baccarelli, A.; Carey, V.J.; Bacherman, H.; Rennard, S.I.; Agusti, A.; Anderson, W.H.; Lomas, D.A.; DeMeo, D.L. Systemic steroid exposure is associated with differential methylation in chronic obstructive pulmonary disease. *Am. J. Respir. Crit. Care Med.* **2012**, *186*, 1248–1255.
33. Kurima, K.; Ebrahim, S.; Pan, B.; Sedlacek, M.; Sengupta, P.; Millis, B.A.; Cui, R.; Nakanishi, H.; Fujikawa, T.; Kawashima, Y.; et al. TMC1 and TMC2 Localize at the Site of Mechanotransduction in Mammalian Inner Ear Hair Cell Stereocilia. *Cell Rep.* **2015**, *12*, 1606–1617.
34. Pan, B.; Geleoc, G.S.; Asai, Y.; Horwitz, G.C.; Kurima, K.; Ishikawa, K.; Kawashima, Y.; Griffith, A.J.; Holt, J.R. TMC1 and TMC2 are components of the mechanotransduction channel in hair cells of the mammalian inner ear. *Neuron* **2013**, *79*, 504–515.
35. Uhlen, M.; Fagerberg, L.; Hallstrom, B.M.; Lindskog, C.; Oksvold, P.; Mardinoglu, A.; Sivertsson, A.; Kampf, C.; Sjostedt, E.; Asplund, A.; et al. Proteomics. Tissue-based map of the human proteome. *Science* **2015**, *347*, 1260419.
36. Carrell, D.T.; Cairns, B.; Aston, K.I.; Jenkins, T.; Smith, A.D.; Uren, P.J.; Horsager, A. Methods of Identifying Male Fertility Status and Embryo Quality. 2019. Available online: https://patentscope.wipo.int/search/en/detail.jsf?docId=WO2017024311&_cid=P22-K5CGNG-22960-1. (accessed on 16.01.2020)
37. Moghbelinejad, S.; Mozdarania, H.; Ghoraeian, P.; Asadi, R. Basic and clinical genetic studies on male infertility in Iran during 2000–2016: A review. *Int. J. Reprod. Biomed. (Yazd)* **2018**, *16*, 131–148.
38. Krausz, C.; Riera-Escamilla, A. Genetics of male infertility. *Nat. Rev. Urol.* **2018**, *15*, 369–384.
39. Bracke, A.; Peeters, K.; Punjabi, U.; Hoogewijs, D.; Dewilde, S. A search for molecular mechanisms underlying male idiopathic infertility. *Reprod. Biomed. Online* **2018**, *36*, 327–339.
40. Tuttelmann, F.; Ruckert, C.; Ropke, A. Disorders of spermatogenesis: Perspectives for novel genetic diagnostics after 20 years of unchanged routine. *Med. Genet.* **2018**, *30*, 12–20.
41. Li, G.; Penagaricano, F.; Weigel, K.A.; Zhang, Y.; Rosa, G.; Khatib, H. Comparative genomics between fly, mouse, and cattle identifies genes associated with sire conception rate. *J. Dairy Sci.* **2012**, *95*, 6122–6129.
42. Blaschek, M.; Kaya, A.; Zwald, N.; Memili, E.; Kirkpatrick, B.W. A whole-genome association analysis of noncompensatory fertility in Holstein bulls. *J. Dairy Sci.* **2011**, *94*, 4695–4699.
43. Fortes, M.R.; Reverter, A.; Kelly, M.; McCulloch, R.; Lehnert, S.A. Genome-wide association study for inhibin, luteinizing hormone, insulin-like growth factor 1, testicular size and semen traits in bovine species. *Andrology* **2013**, *1*, 644–650.
44. Høglund, J.K.; Sahana, G.; Guldbbrandtsen, B.; Lund, M.S. Validation of associations for female fertility traits in Nordic Holstein, Nordic Red and Jersey dairy cattle. *BMC Genet.* **2014**, *15*, 8.
45. Nicolini, P.; Amarin, R.; Han, Y.; Penagaricano, F. Whole-genome scan reveals significant non-additive effects for sire conception rate in Holstein cattle. *BMC Genet.* **2018**, *19*, 14.
46. Nani, J.P.; Rezende, F.M.; Penagaricano, F. Predicting male fertility in dairy cattle using markers with large effect and functional annotation data. *BMC Genom.* **2019**, *20*, 258.
47. Wei, W.; Luo, W.; Wu, F.; Peng, X.; Zhang, Y.; Zhang, M.; Zhao, Y.; Su, N.; Qi, Y.; Chen, L.; et al. Deep Coverage Proteomics Identifies More Low-Abundance Missing Proteins in Human Testis Tissue with Q-Exactive HF Mass Spectrometer. *J. Proteome Res.* **2016**, *15*, 3988–3997.

48. Chalmel, F.; Lardenois, A.; Evrard, B.; Mathieu, R.; Feig, C.; Demougin, P.; Gattiker, A.; Schulze, W.; Jegou, B.; Kirchoff, C.; et al. Global human tissue profiling and protein network analysis reveals distinct levels of transcriptional germline-specificity and identifies target genes for male infertility. *Hum. Reprod.* **2012**, *27*, 3233–3248.
49. Miller, M.R.; Mannowetz, N.; Iavarone, A.T.; Safavi, R.; Gracheva, E.O.; Smith, J.F.; Hill, R.Z.; Bautista, D.M.; Kirichok, Y.; Lishko, P.V. Unconventional endocannabinoid signaling governs sperm activation via the sex hormone progesterone. *Science* **2016**, *352*, 555–559.
50. Rejraji, H.; Sion, B.; Prensier, G.; Carreras, M.; Motta, C.; Frenoux, J.-M.; Vericel, E.; Grizard, G.; Vernet, P.; Drevet, J.R. Lipid remodeling of murine epididymosomes and spermatozoa during epididymal maturation. *Biol. Reprod.* **2006**, *74*, 1104–1113.
51. Martínez, P.; Morros, A. Membrane lipid dynamics during human sperm capacitation. *Front. Biosci.* **1996**, *1*, d103–d117.
52. Merrill, A.; Sweeley, C.C. Sphingolipids: Metabolism and cell signalling. *New Compr. Biochem.* **1996**, *31*, 309–339.
53. Fonseca, B.; Costa, M.; Almada, M.; Correia-da-Silva, G.; Teixeira, N. Endogenous cannabinoids revisited: A biochemistry perspective. *Prostaglandins Other Lipid Mediat.* **2013**, *102*, 13–30.
54. Cerolini, S.; Kelso, K.; Noble, R.; Speake, B.; Pizzi, F.; Cavalchini, L. Relationship between spermatozoan lipid composition and fertility during aging of chickens. *Biol. Reprod.* **1997**, *57*, 976–980.
55. Bjorkgren, I.; Gylling, H.; Turunen, H.; Huhtaniemi, I.; Strauss, L.; Poutanen, M.; Sipila, P. Imbalanced lipid homeostasis in the conditional Dicer1 knockout mouse epididymis causes instability of the sperm membrane. *FASEB J.* **2015**, *29*, 433–442.
56. Min-Seok, R.; Kawamata, Y.; Nakamura, H.; Ohta, A.; Takagi, M. Isolation and characterization of ECT1 gene encoding CTP: Phosphoethanolamine cytidyltransferase of *Saccharomyces cerevisiae*. *J. Biochem.* **1996**, *120*, 1040–1047.
57. Li, Z.; Agellon, L.B.; Allen, T.M.; Umeda, M.; Jewell, L.; Mason, A.; Vance, D.E. The ratio of phosphatidylcholine to phosphatidylethanolamine influences membrane integrity and steatohepatitis. *Cell Metab.* **2006**, *3*, 321–331.
58. Glander, H.J.; Schiller, J.; Suss, R.; Paasch, U.; Grunewald, S.; Arnhold, J. Deterioration of spermatozoal plasma membrane is associated with an increase of sperm lysophosphatidylcholines. *Andrologia* **2002**, *34*, 360–366.
59. Fuchs, B.; Muller, K.; Goritz, F.; Blottner, S.; Schiller, J. Characteristic oxidation products of choline plasmalogens are detectable in cattle and roe deer spermatozoa by MALDI-TOF mass spectrometry. *Lipids* **2007**, *42*, 991–998.
60. Kang, H.M.; Sul, J.H.; Service, S.K.; Zaitlen, N.A.; Kong, S.Y.; Freimer, N.B.; Sabatti, C.; Eskin, E. Variance component model to account for sample structure in genome-wide association studies. *Nat. Genet.* **2010**, *42*, 348–354.
61. Vilhjalmsjon, B.J.; Nordborg, M. The nature of confounding in genome-wide association studies. *Nat. Rev. Genet.* **2013**, *14*, 1–2.
62. Li, H.; Durbin, R. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* **2009**, *25*, 1754–1760.

63. Van der Auwera, G.A.; Carneiro, M.O.; Hartl, C.; Poplin, R.; Del Angel, G.; Levy-Moonshine, A.; Jordan, T.; Shakir, K.; Roazen, D.; Thibault, J.; et al. From FastQ data to high confidence variant calls: The Genome Analysis Toolkit best practices pipeline. *Curr. Protoc. Bioinform.* **2013**, *43*, doi:10.1002/0471250953.bi1110s43.
64. Schutz, E.; von Ahsen, N. Spreadsheet software for thermodynamic melting point prediction of oligonucleotide hybridization with and without mismatches. *Biotechniques* **1999**, *27*, 1218–1222, 1224.
65. von Ahsen, N.; Oellerich, M.; Armstrong, V.W.; Schutz, E. Application of a thermodynamic nearest-neighbor model to estimate nucleic acid stability and optimize probe design: Prediction of melting points of multiple mutations of apolipoprotein B-3500 and factor V with a hybridization probe genotyping assay on the LightCycler. *Clin. Chem.* **1999**, *45*, 2094–2101.
66. Schneider, C.A.; Rasband, W.S.; Eliceiri, K.W. NIH Image to ImageJ: 25 years of image analysis. *Nat. Methods* **2012**, *9*, 671–675.
67. Johnson, L.A.; Pursel, V.G.; Gerrits, R.J. Total phospholipid and phospholipid fatty acids of ejaculated and epididymal semen and seminal vesicle fluids of boars. *J. Anim. Sci.* **1972**, *35*, 398–403.
68. Angrimani, D.S.R.; Nichi, M.; Losano, J.D.A.; Lucio, C.F.; Veiga, G.A.L.; Franco, M.V.M.J.; Vannucchi, C.I. Fatty acid content in epididymal fluid and spermatozoa during sperm maturation in dogs. *J. Anim. Sci. Biotechnol.* **2017**, *8*, 18.
69. Sampaio, J.L.; Gerl, M.J.; Klose, C.; Ejsing, C.S.; Beug, H.; Simons, K.; Shevchenko, A. Membrane lipidome of an epithelial cell line. *Proc. Natl. Acad. Sci. USA* **2011**, *108*, 1903–1907.
70. Ejsing, C.S.; Sampaio, J.L.; Surendranath, V.; Duchoslav, E.; Ekroos, K.; Klemm, R.W.; Simons, K.; Shevchenko, A. Global analysis of the yeast lipidome by quantitative shotgun mass spectrometry. *Proc. Natl. Acad. Sci. USA* **2009**, *106*, 2136–2141.
71. Surma, M.A.; Herzog, R.; Vasilj, A.; Klose, C.; Christinat, N.; Morin-Rivron, D.; Simons, K.; Masoodi, M.; Sampaio, J.L. An automated shotgun lipidomics platform for high throughput, comprehensive, and quantitative analysis of blood plasma intact lipids. *Eur. J. Lipid Sci. Technol.* **2015**, *117*, 1540–1549.
72. Herzog, R.; Schuhmann, K.; Schwudke, D.; Sampaio, J.L.; Bornstein, S.R.; Schroeder, M.; Shevchenko, A. LipidXplorer: A software for consensual cross-platform lipidomics. *PLoS ONE* **2012**, *7*, e29851.
73. Herzog, R.; Schwudke, D.; Schuhmann, K.; Sampaio, J.L.; Bornstein, S.R.; Schroeder, M.; Shevchenko, A. A novel informatics concept for high-throughput shotgun lipidomics based on the molecular fragmentation query language. *Genome Biol.* **2011**, *12*, R8.
74. Zorn, C. Shapiro-Wilk test. In *The SAGE Encyclopedia of Social Science Research Methods*; Sage Publications, Inc.: Thousand Oaks, CA, USA, 2004; pp. 1030–1031.
75. Benjamini, Y.; Hochberg, Y. Controlling the false discovery rate: A practical and powerful approach to multiple testing. *J. R. Stat. Soc. B* **1995**, *57*, 289–300.

Chapter 5

General discussion

In this dissertation, genetic factors contributing to congenital deafness in ASCDs and dog herding, predation, temperament and trainability behaviors in modern dog breeds were analyzed. The obtained data could be used to improve dog health and welfare by understanding genetic mechanism underlying dog diseases and behaviors. Expression and function analysis of bull infertility candidate gene *ABHD16B* were conducted to understand its role in reproduction.

1. Dogs as biomedical models for human inherited hearing loss

Dogs share the longest history with humans among domestic animals, and they are the only animals domesticated prior to the appearance of agriculture [1, 2]. Meanwhile, dogs are naturally susceptible to variety of interspecific Mendelian or complex diseases similar to humans. More than 480 canine genetic disorders have been reported as potential models for studying human diseases (<https://omia.org/home/>). In Chapter 2 of this thesis, *KLF7* gene was identified to be significantly associated with ASCD congenital deafness. Deafness has deleterious effects on dog health and welfare, so it is recommended to control the frequency of this risk factor in ACSD population. *KLF7* was reported to be a promising candidate gene for human Branchio-oto-renal (BOR) syndrome in which hearing loss is a symptom [3]. Notably, *KLF7* was detected to be the nearest gene of one significantly associated locus in human hearing difficulty GWAS using more than 250,000 samples [4]. *KLF7* was still the nearest gene to one of 31 human hearing difficulty risk loci in GWAS of larger populations ($n \leq 330,759$) [5]. As *KLF7* has been detected in several studies of human hearing defects, it is suggested to be one candidate gene for human hearing loss. Our findings in ASCD deafness further support this possibility and may provide clues for the deciphering molecular mechanisms of human deafness. On the other hand, other three recently identified canine deafness genes (*MYO7A*, *PTPRQ* and *LOXHD1*) were all identified as causative genes for human hearing loss. Therefore, dogs can be good naturally occurring animal models to study human deafness [6-8].

2. Dog behavioral genetics studies contribute to the understanding of human mental disorders

Diverse modern dog breeds have been strongly and selectively bred to perform various tasks such as herding, hunting or companionship depending on human purposes since Victorian era. Behavioral and morphological diversity is extremely high between dog breeds [9]. Large animals such as dogs with spontaneous and anomalous behaviors could serve as good research models for human complex psychiatric disorders [10]. For example, several potential genes involved in brain development and synapse formation were successfully mapped for canine compulsive disorders [11]. Candidate canine compulsive disorder genes were then applied into genetic analysis of human obsessive-compulsive disorder (OCD), at which four associated genes were identified [12]. The feasibility of using dogs as human OCD models was confirmed by one recent study on dog circling behavior. The study has identified two canine OCD risk genes (*PPP2R2B* and *ADAMTSL3*) which were reported to function in dendritic spines [13]. In chapter 3 of this thesis, promising candidate neural genes underlying dog herding, predation, temperament traits are pointed out. These results could not only help to understand the biological origins of these complex behavioral traits, but also could provide genetic clues for studying human mental disorders and motivational predispositions. For instance, hyper-social behavior is one unique character of dog domestication that is quite different from that of wolves. Structural variants within human Williams-Beuren syndrome associated genes, including *GTF2I* and *GTF2IRD1*, were reported to contribute to this special behavior in dogs [14]. It is noted that individuals with Williams-Beuren syndrome typically present a hyper-sociability personality that is outgoing, friendly and/or talkative. GWAS analysis of dog fearfulness behavior in German Shepherds and Great Danes has also detected candidate genes that are implicated in human neuropsychiatric diseases [15, 16]. *HS6ST2* was first reported to be associated with dog sociability behavior in GWAS of several hundred dogs [17], and

it was later detected to be significantly related with human neuroticism in GWAS of 405,274 UK Biobank samples [18]. Human behavior is complex and heterogeneous, and accurate phenotypic definitions are difficult. These conditions have hindered the research process of genetic analysis on this topic. In contrast, it is easier to obtain sufficient number of dog samples with personality traits, and dog behavioral genetics could shed light on human mental studies. For example, behavioral genetic studies of aggressiveness and fearfulness in dogs can be used to assist the study of anger and anxiety in human.

3. Genetic mapping of dog diseases and morphological traits using whole genome re-sequences

The declining costs for whole genome re-sequencing has made WGS become a prevalent way to identify causative genes for inherited diseases in dogs. A large size of whole genome re-sequencing data of dog from diverse breeds around the world has been available [19, 20]. WGS of analyzed dogs could generate millions of variants, and variant filtering is a critical step in the data analysis. The publicly available canine WGSs can be used as controls for variant filtering. By comparing affected dog genome sequences with those sequenced control canines, unique variants of the affected dogs can be obtained using variant filtering software such as VCFtools [21]. Usually, only a few hundred or a few dozen private variants of affected dogs remain to be analyzed, which significantly improves the efficiency of variant filtering. New dog reference genomes and annotations have been released recently, which will also improve the genetic diagnostic capabilities for inherited disorders [22-24]. Since whole genome sequences own higher coverage and much more dense variants than SNP chip data, WGS can also be efficiently applied in GWAS to identify genes for dog morphological traits between breeds [20]. Restricted breeding conditions within one dog breed reduce phenotypic and genotypic heterogeneity [25, 26], and lead to long range of LD regions within dog breeds [27, 28]. Thus, these specific genetic architectures are of great

advantages for identifying genes/variants that contribute to complex morphological traits and diseases. The identification of *KLF7* variant benefited from the WGS availability of large control canine samples that were applicable in both GWAS and variant filtering.

4. Sperm lipidomics could be used to evaluate bull fertility

Genomic analysis is one quite common way to identify markers for bull fertility [29]. Other omics assays such as transcriptomic [30] and proteomic analysis of spermatozoa have also been used to appraise bull fertility [31]. The balanced lipid components of sperms are vital for a successful reproduction process. Recently, sperm lipidomic analysis is being used to evaluate the quality of domesticated animal semen, including bull sperms [32]. Bull sperm lipidome studies have been performed to investigate the lipid profile differences for aging and cryo-tolerance traits, detecting several significantly different fatty acids associated with these two traits [33, 34]. Our *ABHD16B* function study also indicated that changes of sperm lipid contents might be related with bull infertility. Therefore, lipidomic analysis between high and low fertile sperms could be used to identify lipidomic biomarkers for bull fertilization competence. These additional markers could be incorporated into conventional semen analysis to improve the prediction accuracy of bull sperm fertility.

5. Prospects

More WGS from diverse dog breeds will be available with advanced process of Dog10K project, and the gene mapping accuracy of the dog complex traits and diseases will improve [35]. The current analysis of causative variants of dog diseases and morphological traits are mainly focused on protein coding-regions, but variants within non-coding regions of dog genome may also have large effects on phenotypic traits, disorders and domestication [36]. Novel variants or genes of interest for canine complex traits and diseases could be revealed after efforts working on the annotations of the

genome regulatory regions [37, 38]. Except WGS, multi-omics approach combining transcriptomes, chromatin immunoprecipitation sequencing (ChIP-Seq), and assays for transposase-accessible chromatin using sequencing (ATAC-seq) will be applied to the molecular interpretations of dog complex diseases such as cancer [39]. With all these incoming efforts, our understanding of molecular mechanisms underlying dog and human health will be further enriched.

Although lipidomic analysis indicated that *ABHD16B* gene participates in sperm lipid metabolisms, its exact function has not yet been elucidated. Further studies are necessary to identify the biochemical effects resulting from *ABHD16B* depletion in Holstein cattle.

References

1. Larson, G. and J. Burger, A population genetics view of animal domestication. *Trends in Genetics*, 2013. 29(4): p. 197-205.
2. Thalmann, O. and A.R. Perri, Paleogenomic inferences of dog domestication, in *Paleogenomics*. 2018, Springer. p. 273-306.
3. Brophy, P.D. et al., Genome-wide copy number variation analysis of a Branchio-oto-renal syndrome cohort identifies a recombination hotspot and implicates new candidate genes. *Human Genetics*, 2013. 132(12): p. 1339-1350.
4. Wells, H.R.R. et al., GWAS Identifies 44 Independent Associated Genomic Loci for Self-Reported Adult Hearing Difficulty in UK Biobank. *American Journal of Human Genetics*, 2019. 105(4): p. 788-802.
5. Kalra, G. et al., Biological insights from multi-omic analysis of 31 genomic risk loci for adult hearing difficulty. *Plos Genetics*, 2020. 16(9).
6. Webb, A.A., A.L. Ruhe, and M.W. Neff, A missense mutation in MYO7A is associated with bilateral deafness and vestibular dysfunction in the Doberman pinscher breed. *Canadian Journal of Veterinary Research-Revue Canadienne De Recherche Veterinaire*, 2019. 83(2): p. 142-148.
7. Guevar, J. et al., Deafness and vestibular dysfunction in a Doberman Pinscher puppy associated with a mutation in the PTPRQ gene. *Journal of Veterinary Internal Medicine*, 2018. 32(2): p. 665-669.
8. Hytönen, M.K. et al., Missense Variant in LOXHD1 is Associated With Canine Nonsyndromic Hearing Loss. 2021.
9. Parker, H.G. et al., Genomic Analyses Reveal the Influence of Geographic Origin, Migration, and Hybridization on Modern Dog Breed Development. *Cell Reports*, 2017. 19(4): p. 697-708.
10. Danek, M., J. Danek, and A. Araszkiwicz, Large animals as potential models of human mental and behavioral disorders. *Psychiatria Polska*, 2017. 51(6): p. 1009-1027.
11. Tang, R.Q. et al., Candidate genes and functional noncoding variants identified in a canine model of obsessive-compulsive disorder. *Genome Biology*, 2014. 15(3).
12. Noh, H.J. et al., Integrating evolutionary and regulatory information with a multispecies approach implicates genes and pathways in obsessive-compulsive disorder. *Nature communications*, 2017. 8(1): p. 1-13.
13. Cao, X. et al., Whole genome analyses reveal significant convergence in obsessive-compulsive disorder between humans and dogs. *Science Bulletin*, 2021. 66(2): p. 187-196.
14. vonHoldt, B.M. et al., Structural variants in genes associated with human Williams-Beuren syndrome underlie stereotypical hypersociability in domestic dogs. *Science Advances*, 2017. 3(7).
15. Sarviaho, R. et al., Two novel genomic regions associated with fearfulness in dogs overlap human neuropsychiatric loci. *Translational Psychiatry*, 2019. 9.
16. Sarviaho, R. et al., A novel genomic region on chromosome 11 associated with fearfulness in dogs. *Translational Psychiatry*, 2020. 10(1).
17. Zapata, I., J.A. Serpell, and C.E. Alvarez, Genetic mapping of canine fear and aggression. *Bmc Genomics*, 2016. 17.

18. Luciano, M. et al., The influence of X chromosome variants on trait neuroticism. *Molecular Psychiatry*, 2021. 26(2): p. 483-491.
19. Jagannathan, V. et al., A comprehensive biomedical variant catalogue based on whole genome sequences of 582 dogs and eight wolves. *Animal Genetics*, 2019. 50(6): p. 695-704.
20. Plassais, J. et al., Whole genome sequencing of canids reveals genomic regions under selection and variants influencing morphology. *Nature Communications*, 2019. 10.
21. Danecek, P. et al., The variant call format and VCFtools. *Bioinformatics*, 2011. 27(15): p. 2156-2158.
22. Field, M.A. et al., Canfam_GSD: De novo chromosome-length genome assembly of the German Shepherd Dog (*Canis lupus familiaris*) using a combination of long reads, optical mapping, and Hi-C. *Gigascience*, 2020. 9(4).
23. Wang, C. et al., A novel canine reference genome resolves genomic architecture and uncovers transcript complexity. *Communications Biology*, 2021. 4(1).
24. Halo, J.V. et al., Long-read assembly of a Great Dane genome highlights the contribution of GC-rich sequence and mobile elements to canine genomes. *Proceedings of the National Academy of Sciences of the United States of America*, 2021. 118(11).
25. Lindblad-Toh, K. et al., Genome sequence, comparative analysis and haplotype structure of the domestic dog. *Nature*, 2005. 438(7069): p. 803-819.
26. Freedman, A.H. et al., Demographically-Based Evaluation of Genomic Regions under Selection in Domestic Dogs. *Plos Genetics*, 2016. 12(3).
27. Sutter, N.B. et al., Extensive and breed-specific linkage disequilibrium in *Canis familiaris*. *Genome Research*, 2004. 14(12): p. 2388-2396.
28. Boyko, A.R. et al., A Simple Genetic Architecture Underlies Morphological Variation in Dogs. *Plos Biology*, 2010. 8(8).
29. Taylor, J.F., R.D. Schnabel, and P. Sutovsky, Review: Genomics of bull fertility. *Animal*, 2018. 12: p. S172-S183.
30. Feugang, J.M. et al., Transcriptome analysis of bull spermatozoa: implications for male fertility. *Reproductive Biomedicine Online*, 2010. 21(3): p. 312-324.
31. Viana, A.G.A. et al., Proteomic landscape of seminal plasma associated with dairy bull fertility. *Scientific Reports*, 2018. 8.
32. Evans, H.C. et al., Advancing Semen Evaluation Using Lipidomics. *Frontiers in Veterinary Science*, 2021. 8: p. 314.
33. Argov-Argaman, N. et al., Variation in lipid profiles within semen compartments the bovine model of aging. *Theriogenology*, 2013. 80(7): p. 712-721.
34. Evans, H.C. et al., Lipidomic markers of sperm cryotolerance in cattle. *Scientific Reports*, 2020. 10(1).
35. Ostrander, E.A. et al., Dog10K: an international sequencing effort to advance studies of canine domestication, phenotypes and health. *National Science Review*, 2019. 6(4): p. 810-824.
36. Sahlén, P. et al., Variants That Differentiate Wolf and Dog Populations Are Enriched in Regulatory Elements. *Genome Biology and Evolution*, 2021. 13(4): p. evab076.
37. Megquier, K. et al., BarkBase: Epigenomic Annotation of Canine Genomes. *Genes*, 2019. 10(6).
38. Wucher, V. et al., FEELnc: a tool for long non-coding RNA annotation and its application to the dog transcriptome. *Nucleic Acids Research*, 2017. 45(8).

39. Evans, J.M. et al., Multi-omics approach identifies germline regulatory variants associated with hematopoietic malignancies in retriever dog breeds. *PLoS genetics*, 2021. 17(5): p. e1009543.

Acknowledgments

PhD study is one special experience in my life, which is a valuable journey with hardship, fun and harvest. I really appreciate those people who helped me through the journey.

I would like to sincerely thank **Prof. Dr. Ekkehard Schütz** for his great support and specialized advice for my PhD study. I would like to express my cordial gratitude to **Prof. Dr. Bertram Brenig** for offering such a great opportunity to have both Shuwen and I to study in Göttingen. He not only provided excellent supervisions for the research projects, but also influence me with his positive attitude towards scientific research. I also appreciate his financial support to complete my PhD study after my scholarship ended. I am also grateful to **Prof. Dr. Jens Tetens** for being my thesis evaluator and supporting me with his valuable advice. Special thanks are also given to **Prof. Dr. Michael Hölker** for being my thesis defense member.

Thanks to the great colleagues in our lab, which allowed me to accomplish my PhD study in a comfortable and pleasant working environment. Firstly, I would like to express my sincere thanks to **Dr. Marc Hirschfeld** for his guidance and support in my PhD projects. His valuable discussions and feedback have significantly improved the efficiency of the projects. Thanks also go to **Dr. Wilhelm Wemheuer** for his kind help in sample collections. **Dr. Anne Hollmann** and **Christin Wacker** are thanked for teaching me the experimental skills. I would like to thank **Dr. Xuying Zhang** for helping me in getting familiar with the lab and life in Göttingen. Lilith Steingraber and Frederik Krull are thanked for their kind help in lab life. Research collaborators such as **Prof. Dr. Jennifer M. Seddon** and **Dr. Susan Sommerlad** of University of Queensland are appreciated for their kind support. Additionally, I also appreciate **Sabrina Pach, Ina Meyer, Melanie Scharfenstein** and **Delia Niehaus** for creating nice experimental conditions and ordering reagents.

I am also grateful for the financial support from China Scholarship Council.

The colorful life in Göttingen could not be that interesting without my friends, and I hope we could keep in touch in the future.

Finally, I would like to express my gratitude to my family. Especially for my parents **Xingshan Xu** and **Xuexiu Wang**, my wife **Shuwen Shan**, who supported me unconditionally with love and trust for my doctoral study.

Curriculum Vitae

Family Name: Xu
Given Name: Fangzheng
Gender: Male
Nationality: Chinese
Date of Birth: July 16.1990
Place of Birth: Linyi, Shandong, China

Education

2016.10 - Current Doctoral Studies
Department of Molecular Biology of Livestock and molecular
Diagnostics, University of Göttingen, *Göttingen, Germany*

2013.09 - 2016.06 Master Studies
Key Lab of Animal Genetics, Breeding and Reproduction, Huazhong
Agricultural University, *Wuhan, P.R. China*

2009.09 - 2013.06 Bachelor Studies
College of Animal Science and Technology, Qingdao Agricultural
University, *Qingdao, P.R. China*